



UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

ESA/STAT/AC.289/25
11 May 2015

**Meeting of the Expert Group on
International Statistical Classifications
New York, 19-22 May 2015**

**Results of the UNSD/UNECE Survey on organizational context and individual
projects of Big Data**

UNSD, UNECE

Statistical Commission

Forty-sixth session

3 – 6 March 2015

Item 3(a) (iii) of the provisional agenda

**Items for discussion and decision: Data in support of
the Post-2015 Development Agenda: Big Data**

Background document

Available in English only

Results of the UNSD/UNECE Survey on
organizational context and individual projects of Big Data

Prepared by the Statistics Divisions of
UN/DESA and UN Economic Commission for Europe

(February 2015)

Table of Contents

Big Data Project Survey	3
1. Introduction.....	3
2. Main findings – Organizational context.....	3
(i) Big Data strategy	4
(ii) Governance.....	4
(iii) Quality.....	4
(iv) Privacy.....	5
(v) Skills.....	5
3. Main findings – Individual projects	6
(i) Status and objective	6
(ii) Statistical area of use.....	7
(iii) Partnerships.....	7
(iv) Data sources: General.....	8
(v) Data sources: Granularity and frequency.....	9
(vi) Data sources: Privacy concerns.....	10
(vii) Technical tools	10
Annex 1: Big Data Strategy.....	12
Annex 2: Big Data Governance and management structure.....	20
Annex 3: Big Data Quality.....	25
Annex 4: Big Data Privacy and confidentiality	28
Annex 5: Big Data Skills	33
Annex 6: Big Data Projects by main area of use.....	37

Big Data Project Survey

UNSD / UNECE joint survey – Fall 2014

1. Introduction

The United Nations Statistics Division (UNSD) and the United Nations Economic Commission for Europe (UNECE) teamed up to conduct a survey on Big Data projects for Official Statistics and provide herewith an overview of active Big Data projects, which could help to facilitate a more informed discussion within the community at large and develop the programmes of the international working groups.

The survey was conducted from mid-September to mid-October 2014 among the participants in the UNECE Big Data Project and/or the UN Global Working Group on Big Data for Official Statistics.

The questionnaire consisted of two parts: one part focussing on the organizational aspects of executing a Big Data project and another part with questions specifically for Big Data projects (even if in the planning phase). For this questionnaire we adopted a fairly wide definition of what "Big Data" is, namely:

- *Big Data are data sources with a high volume, velocity and variety of data, which require new tools and methods to capture, curate, manage, and process them in an efficient way.*

In total, the survey was sent to 78 national statistical offices and 28 international organizations. Replies to the questionnaire on the organizational context were provided by 32 countries and 3 organizations, while 24 countries and 3 organizations were able to report on a total of 57 Big Data projects. This background document gives a summary of the results in parts 2 and 3 below and provides details in the annexes. Countries and organizations are only mentioned if consent was given by the institute. Whereas the summary of the results includes all replies, the more detailed annexes reveal only the results as permitted by the respondents.

2. Main findings – Organizational context

While most respondents recognize the challenges related to IT, skills, legislation and methodology, most argue that the biggest challenge for most Big Data projects is the limited access to potential datasets. Big Data are, to a large extent, owned by the private sector (e.g. on-line companies, mobile phone operators, banks, etc.), thus, it is very important to build close collaboration with the private sector. Many of these players are global companies; hence the global statistical community should use collective bargaining power to obtain the access to these data sets.

Many respondents argue that the various aspects can be tackled at different levels; i.e. national, regional and global. Some of these aspects, such as quality, confidentiality and partnerships, are

most efficiently handled on the global level. The international community should facilitate the global partnerships and coordinate the development of quality and confidentiality frameworks. This is already undergoing in the UNECE Big Data project.

(i) Big Data strategy

- ✓ What is your long-term vision for the usage of Big Data in your organization?
- ✓ Have you defined business processes for integrating Big Data sources and results into your work?
- ✓ Are there specific Big Data partnerships and/or data sources you are attempting to obtain?

Only a few countries have developed a long-term vision for the usage of Big Data, while a number of countries are currently on the brink of formulating a Big Data strategy. However, most countries do not yet have a Big Data strategy. A number of countries have established internal labs/task teams/working groups to carry out pilot projects to determine if and how Big Data could be used as a source of Official Statistics. Most countries have not yet defined business processes for integrating Big Data sources and results into their work and do not have a defined structure for managing Big Data projects.

In many countries the key issue at the moment is to build partnerships in order to explore the opportunities of Big Data. In order to minimize the financial risks of exploring these new technologies, countries have decided to actively participate in regional collaborations, such as the UNECE Sandbox project and Eurostat's Big Data Task force. Others are exploring strategic partnerships with other governmental agencies or commercial enterprises.

(ii) Governance

- ✓ Do you have a defined structure for managing Big Data projects?
- ✓ Do you limit yourselves to certain types of data and/or partnerships?

About half of the responding institutes have formally assigned staff to work on Big Data projects. The other half has not yet engaged in Big Data projects. Of those offices which have staff assigned, only a handful has established a dedicated team or dedicated structure for Big Data management. In the other offices with assigned staff, Big Data projects are either managed by an ad-hoc team of experts from IT and Methodology Departments, or within the existing corporate structure.

(iii) Quality

- ✓ How do you assess the statistical quality of Big Data sources and/or of the output of analysis based on Big Data sources?
- ✓ If you have frameworks for assessment discuss that here.

More than two thirds of the organizations explained that they do not yet have defined a quality assessment framework for Big Data sources or the output of analysis.

A few countries have established a quality assurance framework to assess, for example, the usability of Big Data for statistics production within the general framework for the determination

of the input quality of secondary data sources that has been developed in the BLUE-ETS project. Others use statistical and machine learning methods to produce quantified and defensible measures of quality, or use correlations with other sources of statistics to assess the quality of the results under the traditional approach of official statistics.

In reference to a quality framework, many countries refer to their participation in the development of a quality framework within the UNECE Big Data project.

(iv) Privacy

- ✓ Do you have a framework for dealing with privacy issues?
- ✓ Please include any other comments regarding privacy and confidentiality.

Most survey respondents mention that the same privacy framework that applies to traditional statistics also applies to Big Data, and very few countries have a privacy framework only for Big Data.

Many organizations stress how they work to protect privacy and confidentiality when dealing with Big Data, even beyond what is strictly required by law, in order to ensure their public image. These organizations only undertake studies of Big Data sources in cases where the public good is clearly evident and outweighs the privacy invasion.

Some organizations avoid the issue of privacy and confidentiality by having all the testing and manipulation of the data at the location done by the Big Data owner, which then transmits only aggregates to the NSO.

Some organizations argue that while strict data protection regulation is needed, it can sometimes be a barrier to gaining access to additional datasets for the NSO.

(v) Skills

- ✓ How have you addressed and/or are planning to address skill shortages in Big Data?
- ✓ What skills do you consider to be most important for staff working with Big Data?
- ✓ Are you actively hiring data scientists?
- ✓ Are you training existing staff?
- ✓ What has been the most helpful?

Overwhelmingly the survey respondents indicate that they so far have relied more on providing training to existing staff rather than hiring of a new type of staff (“data scientists”). There are two different trends within this pattern. Some respondents mention that they provide training to statisticians/methodologists in IT/statistical programmes such as Java, SAS, SQL, SPSS, Stata and R, as well as on the data governance and standards to process large volumes of data. Other respondents mention that they have reserved IT staff for Big Data and have started to train them on statistical and methodological issues.

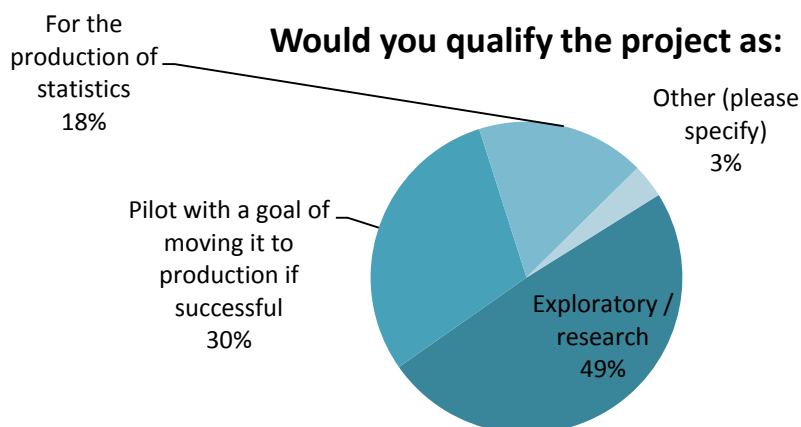
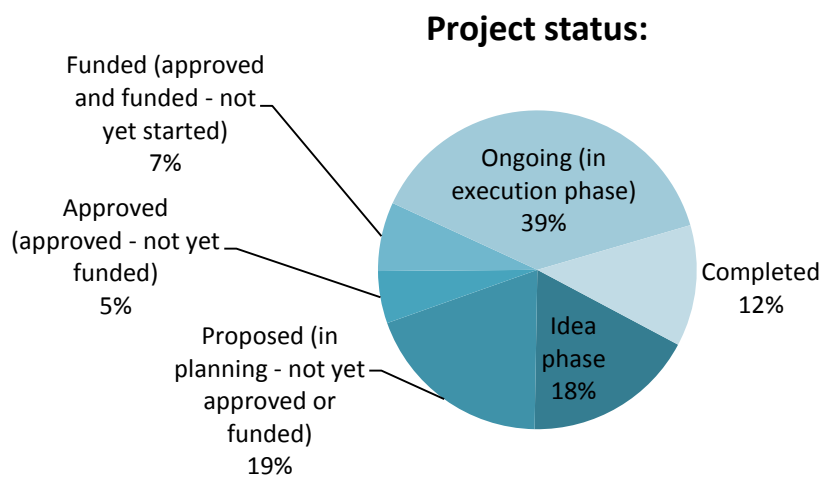
However, most agree that it is important to attract talents and teams in the development and use of Big Data. The mastery of multiple disciplines, such as humanities, sciences and arts, and a thorough understanding of the running track of society and human behaviour contribute to

realization of the Big Data concept. Due to budget constraints in a number of countries, hiring of data scientists is currently not planned and therefore further training of the current staff is a more realistic step for the future.

3. Main findings – Individual projects

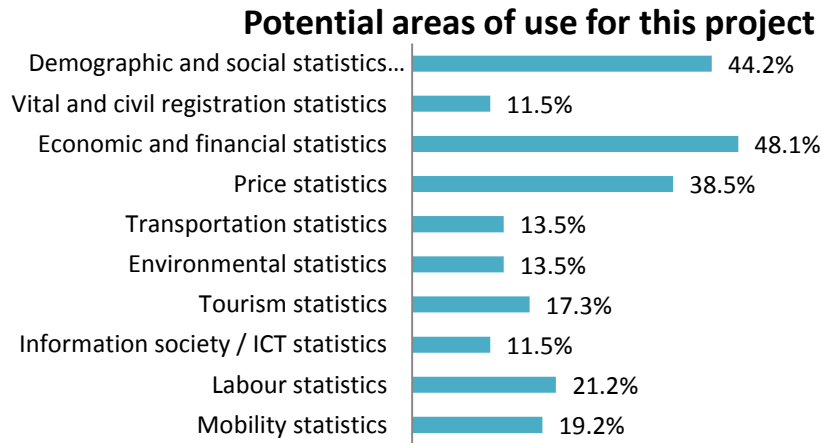
(i) Status and objective

Of the 57 projects of this survey, about half are in the initial phases of proposal, approval and getting funded. The other half of the projects are mostly ongoing, with a total of 12% being completed. In general, these figures indicate that most institutes are at an early stage of implementation of Big Data projects. Moreover, only half of all the projects are intended to be taken into production (of which about a third have actually been taken into production). The other half of the projects are for exploration and research.



(ii) Statistical area of use

Most of the projects are related to economic and financial statistics, demographic and social statistics, and price statistics.

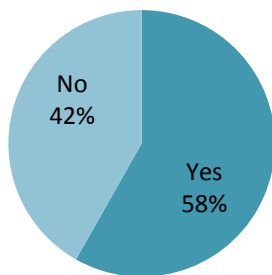


(iii) Partnerships

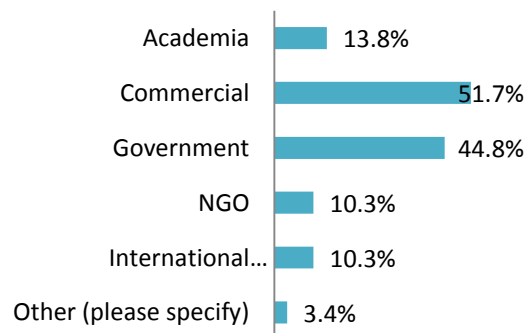
Just more than half of the projects have partnerships with other organizations or data providers. The most common partner is commercial enterprises, followed by other government agencies. The most common type of partnership is with a data provider, followed by analytical partners. A few partnerships are with data consumers, design partners and technology partners. Access to data is the main reason for engaging in partnerships and, for the projects in this survey, the NSOs had the necessary IT infrastructure to implement them.

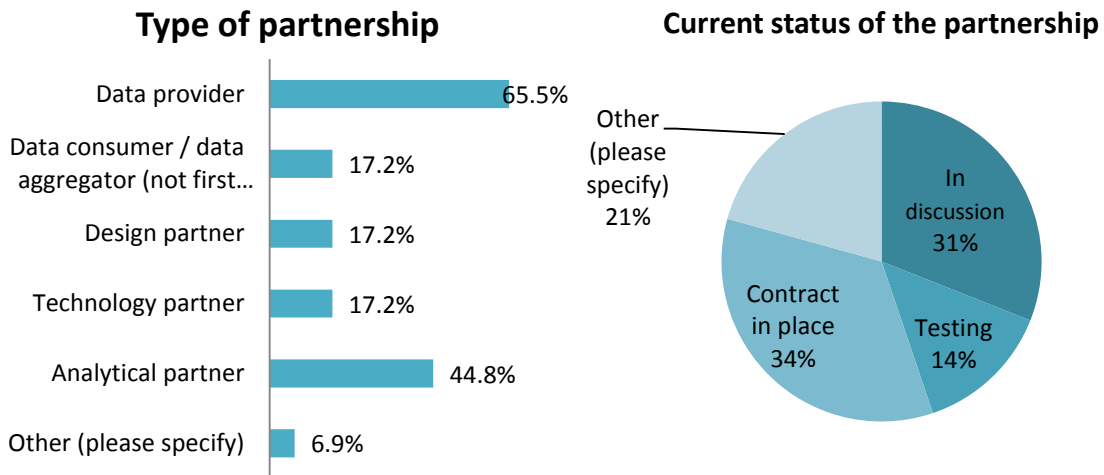
Among the partnerships, one third have a contract in place, while 44% are either still in discussion or prototyping/testing. For half of the projects with partners, there are no payments or financial arrangements, while 39% of the partners have agreements which include some form of payment.

Do you have any partnerships with other organizations or data providers on this project?



Type of partner organization

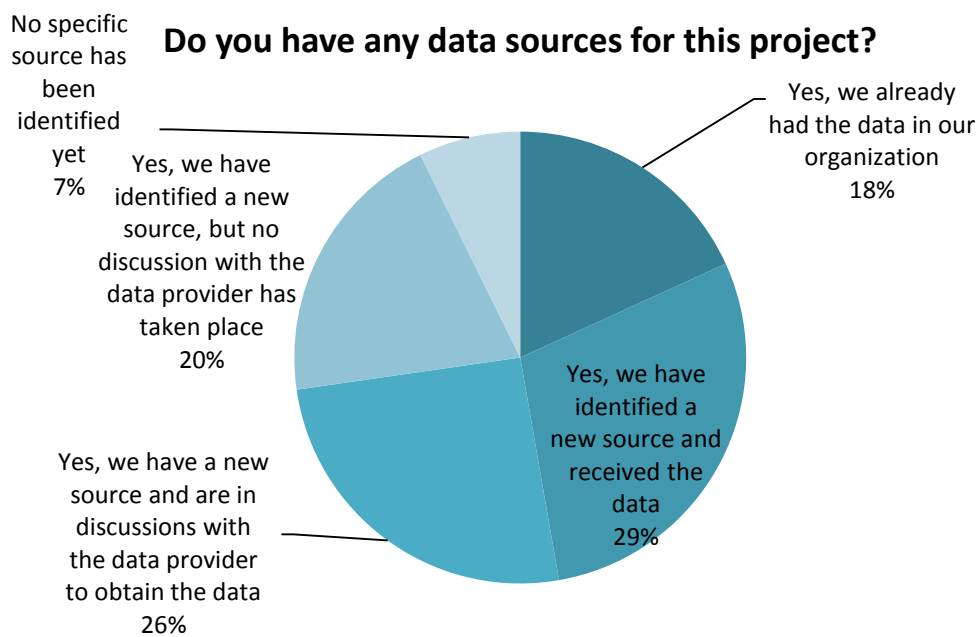




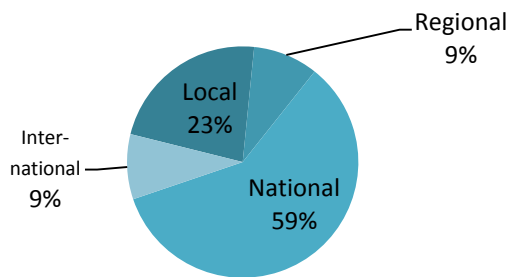
(iv) Data sources: General

For almost all projects, a specific data source has been identified. For around half the projects access to the data source is secured, while the rest of the projects still require further discussions with the data provider. Access to the data source can become the principal risk factor to the success or failure of a Big Data project.

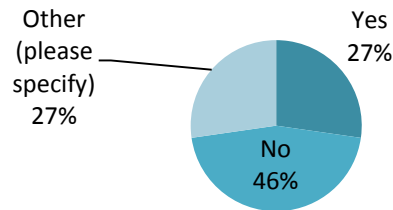
More than half of the data sources are national in scope, while only very few have an international scope, which could indicate that obtaining access to a Big Data source with an international dimension is challenging.



What is the geographical scope of the data source?



Have you established automatic links for transmitting this data source (e.g. API, automatic file download)?

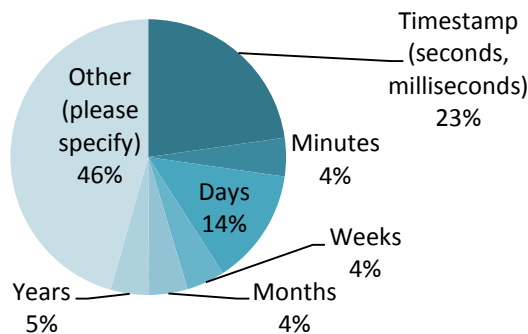


While the method of data receipt was not mentioned often, responses cited that it was primarily determined by the partner providing the data. Data was transferred via APIs, encrypted hard drives, ad-hoc methods, and web scraping.

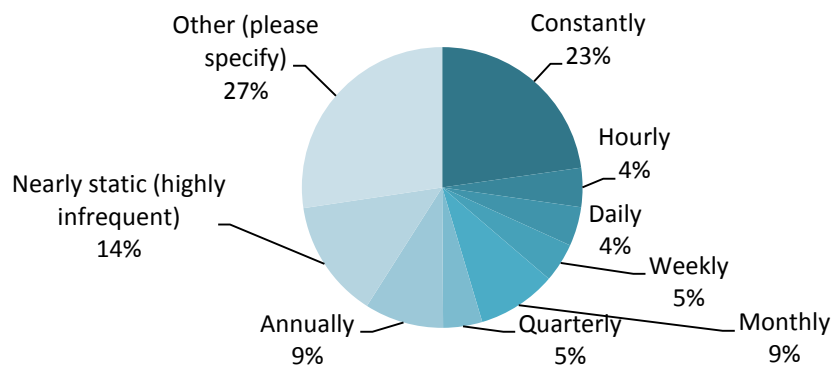
(v) Data sources: Granularity and frequency

The data sources showed a great variety in granularity and frequency of updating. About a quarter of the data sources have an accurate time stamp (of seconds or milliseconds) and about a quarter are updated constantly.

How granular is the information in the data source? This should correspond to unit of time used to mark individual records.



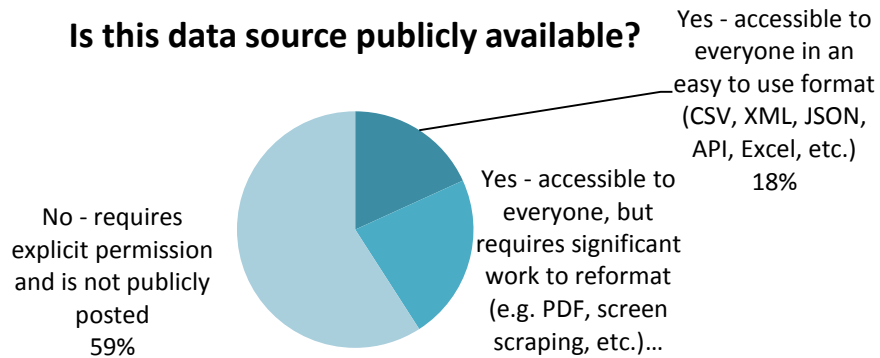
How frequently are data source updates made available?



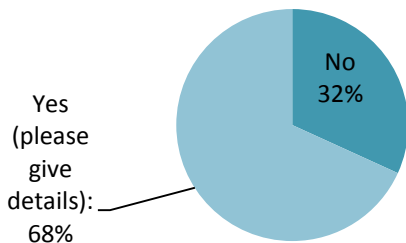
(vi) Data sources: Privacy concerns

Whereas about half of the data sources used have privacy concerns, which need to be properly addressed, the other half of the data sources are publicly available and pose less problems in that respect. It is also interesting to note that in more than half of the projects, the Big Data sources are mixed with existing traditional data sources in the office.

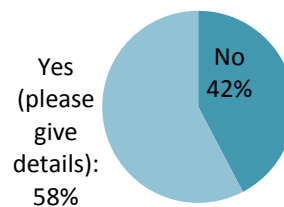
Is this data source publicly available?



Are there any privacy and confidentiality issues related to this data source?



Do you integrate traditional data sources with the new "Big Data" source discussed above?



(vii) Technical tools

Overwhelmingly, survey participants chose to use internal hosting solutions rather than purchase external hosting services. Privacy concerns were cited by a number of participants as the primary reason for internal hosting. External hosts and partners were used more often when the data were deemed to already be in the public domain. A lack of access to partners perceived to be able to securely host sensitive data will certainly slow the adoption of complex Big Data tools, such as Hadoop, relative to the private sector. It is much less costly to begin projects using external hosts even if in the long-run in-house hosting might become cost-effective. Having to

invest significant resources in a cluster before beginning a project represents a significant hurdle to gaining familiarity with a set of tools that are still changing rapidly.

Data analysis was mainly conducted using R and SAS. However, a large variety of tools were used in the projects, making it clear that no significant standard has emerged. RHadoop and SAS clusters were occasionally used to facilitate parallel processing. Python, Java, and Scala were also used to do processing on Hadoop. Amongst projects actively handling data, R+Hadoop was the most common setup. Also a set of projects used RDF tools to produce graph models, R+SPARQL for analysis, and Virtuoso for the data store. A few projects also used JavaScript libraries such as D3 and Cross-filter for interactive display of results.

Projects in planning were less likely to use tools generally associated with “Big Data”. Often this decision was made due to a lack of familiarity with new tools or a deficit of secure “Big Data” infrastructure (e.g., parallel processing no-SQL data stores such as Hadoop). This deficit illustrates a major hurdle to the scalability of many projects. While pilot projects can often be run using traditional methods, full scale solutions generally will require significant infrastructure investment and training if disaggregated data is to be kept in-house.

	Data transfer is:		
	In planning	Active / Complete	Total
Analysis Tools			
C	0	0	1
Elastic Search	0	1	1
GIS	1	1	2
Hadoop (PIG/HIVE)	0	2	2
Java	0	1	1
Partner	0	2	2
Python	1	1	2
R	9	6	15
SAS	8	3	11
Scala	0	1	1
SPSS	0	2	2
Stata	2	0	2
Weka	1	0	1
N/A	13	12	25
Storage			
Flat files	1	1	2
Hadoop	4	4	8
RDBMS*	3	1	4
RDF**	0	3	3
Teradata	0	1	1
N/A	23	18	41
Internal/External			
Internal (fully)	7	20	27
External (any)	0	2	2
N/A	23	5	28

NOTE: Specific software was not always specified. Many projects used more than one analysis tool.

* RDBMS: Relational database management systems such as Oracle, SQL Server, MySQL, and MS Access.

** RDF: Resource Description Framework is a W3C standard for denoting relationships between objects, e.g. graphs. It is often used to give semantic meaning to elements in a format understandable by computers and is a building block of the “semantic web”.

Annex 1: Big Data Strategy

- What is your long-term vision for the usage of Big Data in your organization?
- Have you defined business processes for integrating Big Data sources and results into your work?
- Are there specific Big Data partnerships and/or data sources you are attempting to obtain?

China

National Bureau of Statistics of China lays out -its Big Data work according to the following general ideas: designing from the overall perspective, taking the lead in research, tackling easy issues first and making breakthroughs from a professional perspective. Currently, we have signed strategic cooperation agreements regarding Big Data with 17 commercial enterprises, for the main purpose of obtaining Internet data, e-commerce transaction data, part of the department data, as well as business data, and applying these data to government statistics.

Denmark

Statistics Denmark has a long tradition of using data from administrative registers in the production of statistics. As an organization we take a big interest in Big Data and the possible usage for official statistics. We are on the brink of formulating a new strategy, and Big Data will have a place as a strategic point in this new strategy. We have one Big Data project in relation to scanner data for Price statistics.

Finland

Our vision could be: we are utilizing successfully new data sources, increasing relevance and timeliness of official statistics. However, as being part of the European Statistical System, we share European visions and strategies towards that. Statistics Finland is a member of the ESS Taskforce on Big Data and Official Statistics. We have also an internal Big Data working group that is preparing a big data strategy for Statistics Finland. At this moment, we have not defined business process for integrating Big Data sources. The traffic sensor data of Finnish Transport Agency can be defined as "Big Data". We have been using that in research and development for better drive time estimation in accessibility statistics as part of the Social Statistics Data Warehouse. We are interested in new Big Data sources and looking forward of using such.

Italy

1. Istat planned to have Big Data as sources for statistical production.
2. The activity of defining business processes for integrating Big Data sources is in progress.
3. Istat is investigating partnerships with Big Data providers, in particular with telecommunication companies, and supermarket chains.

Japan

We haven't discussed this issue. We recognize that we should consider it in the future, as necessary.

Portugal

The first steps in "Big Data" are provided by the Administration Board in the Plan of activities for 2015. Being not yet defined are business processes for integration of sources and not existing at this time are specific partnerships. However it is intended to explore and promote the use of geospatial data and, if possible, Big Data as alternatives to traditional survey sources.

Increase cooperation with stakeholders, establishing strategic alliances with public and private partners holders of information, including Big Data, susceptible of appropriation for statistical purposes.

Romania

We are now in the process of defining the NSI's strategy for 2014-2020. Our strategy will include a section about integrating Big Data sources into the modernization process of official statistics in Romania. At this moment we are investigating the possible Big Data sources that we can use in the statistical production.

Serbia

There is an idea to start using Big Data in SORS. However, it is still on the level of idea, but we plan to develop the Strategy and activity plan for implementing it.

Sweden

The vision is that Big Data sources will be incorporated in the production in the future. As of now, we use scanner data for the CPI and the processes for that are in place. Partnerships with suppliers run smoothly. Other sources are at a very early stage of planning. For electricity consumption, we are discussing with relevant organizations and suppliers.

UK

Like many other National Statistics Institutes (NSIs) the Office for National Statistics (ONS) recognises the importance of understanding the impact that big data and associated technologies may have on our statistical processes and outputs. A 15 month Big Data Project (which is due to complete at the end of March 2015) has been established to investigate the benefits alongside the challenges of using Big Data and associated technologies within official statistics. The key deliverable from the project will be an ONS strategy and implementation plan for big data. The long term vision is that ONS moves away from using survey and Census data sources and uses more administrative and potentially Big Data sources to produce official statistics. Ultimately Big Data and associated technologies have the potential to allow ONS to produce more relevant and more timely statistics in a more efficient way, for example:

- outputs currently based on “traditional” sources could be produced using Big Data
- Big Data sources could be used to produce entirely new outputs
- statistics produced using Big Data sources could complement official statistics (based on “traditional sources”) either by filling in any gaps or being used as an auxiliary variable in a model or being used for quality assurance purposes
- Big Data sources could be used to improve operational processes such as field operations
- Big Data technologies could be used to improve data storage, processing and analysis – not just for Big Data sources

However, as well as potential benefits Big Data also bring new challenges with their use within official statistics, specifically:

- around the technologies needed to store, process, analyse them
- the statistical and methodological techniques needed to analyse them
- the ethical and legal issues that arise from their use
- the commercial issues that are associated with accessing these data and
- capability issues, the skills that are required to handle and analyse the data and support these new technologies

The ONS Big Data Project was established to investigate these potential benefits alongside the challenges of using Big Data and associated technologies within official statistics and how these might fit within the overall palette of sources, methods and technologies used to support the production of official statistics. Since we are only still within a research/feasibility stage we have not yet formally defined business processes for integrating Big Data sources and results into production.

A key component of the Big Data Project has been to include some practical applications of Big Data to both assess the role they may have within official statistics and also to help understand the methodological and technical issues that may arise when handling them. Four pilot projects have been chosen covering both economic and social themes. Each pilot uses a key Big Data source, namely Internet price data, Twitter messaging, smart meter data and mobile phone positioning data.

In addition to the practical element of the project we have also undertaken significant activities around stakeholder engagement. In particular we have looked to develop partnerships with the following groups:

- International (e.g., UN, UNECE, OECD, Eurostat, European Central Bank, NSOs) Like the ONS many NSOs are interested in the potential impact of Big Data and associated technologies. Many key issues will be common across NSOs. Engagement facilitates knowledge sharing and collaborative working.
- Academia (e.g., ESRC, RSS, Universities) Specific universities have expertise on data science and hence through engagement there may be opportunities to commission pieces of work within the project or to work collaboratively with academics. In addition there may be opportunities to recruit graduates or take on placement students with a data science background hence enhancing our understanding of the skills required for this type of work. Engagement with overarching bodies such as the ESRC Economic and Social Research Council and RSS Royal Statistical Society will help to make these links with academics and raise the profile of the project. There may also be opportunities for jointly funded work such as a study on public perception.
- Private Sector (e.g., Mobile network operators, Billion Prices, Twitter, Supermarkets) The key motivation behind engagement with the private sector is to acquire their data for use within the pilot projects.
- 'Big Data' Companies (e.g., Google, Growth intelligence) Where companies already have extensive experience with Big Data/data science it may be fruitful to work collaboratively
- Technology providers (e.g., MongoDB, Cloudera Hortonworks, Amazon Elastic Compute Cloud (EC2), Openstack, Canonical) The motivation behind engagement with technology providers is to acquire tools/technologies for use within the pilot projects and to support the project in general
- Government (e.g., Cabinet office, Government Digital Service (GDS), Open Data Institute, GSS, other Government Departments) ONS is not the only UK government department interested in Big Data and hence there is a need to engage bilaterally (e.g. with Cabinet Office, Bank of England, other departments directly) and also through

cross-Government initiatives and community discussions (e.g. Community of Interest, Government Conferences).

- Privacy (e.g., Privacy groups, Information Commissioners Office (ICO)) The use of Big Data sources often raises concerns around security and privacy. It is therefore within the scope of the project to consider ethical issues around public perception this requires engagement with the public in general and those representing privacy issues.

Eurostat

The mission of the Task Force is to lead and co-ordinate developments within the ESS (European Statistical System) and the Commission with regard to maximizing the potential of Big Data for Official Statistics and evidence-based policy making. In particular, a new ESS Big Data strategy, along the lines of the Scheveningen Memorandum adopted by the ESSC on September 2013, has been recently presented to the ESSC accompanied by an action plan and roadmap. The proposed action plan and roadmap may be seen as part of the implementation strategy of the ESS Vision 2020, in particular with regard to how the ESS will respond to the data revolution and integrate new data sources into the official statistical production.

ESCAP

Big data have a great potential to assist NSOs and other producers of official statistics to produce new or complement existing data to serve the needs of policy makers, including the realization of the new data revolution and facilitating the measurement of the post-2015 development agenda. Big data hold a particular potential for enhancing the availability and use of localized data that is useful for planning purposes and rapid response to emergencies, in addition to progress monitoring. They can also facilitate the measurement of social phenomena at low levels of geographical disaggregation and for smaller groups of population.

A number of initiatives are taking place to develop solutions in using Big Data to produce official statistics. It is important that perspectives for less developed statistical systems in the region be taken into account in the work of these global mechanisms. NSOs in developing countries are also facing the need to use Big Data for their work, but their business needs, sources and challenges are different. Their interest is also in answering different questions, such as measuring development indicators, e.g. poverty and inequality. Big Data solutions developed by organizations from more developed NSSs might not be immediately useful to less developed ones and less developed systems do not have adequate financial resources to engage in such projects themselves.

The statistics community needs to embrace Big Data to maintain its relevance as the primary producer of official statistics. Big Data require new and effective public-private sector collaboration, underpinned by new codes of conduct to ensure that confidentiality and privacy of individuals is respected. At the same time they pose extra challenges for statistical institutions and add pressure on their already limited resource base. Taking into consideration the importance of Big Data, the possibilities they bring for the measurement of the post-2015 development agenda, and the challenges that the statistical community has to face, UNESCAP has initiated discussions in the region with the objective of defining a regional strategy on how Big Data can be best used to support economic and social development in Asia and the Pacific. Big Data, together with the debate on the new data revolution in delivering the post-2015 development agenda, and the implications for statistical development in the region, will be explored by the Expert Group Meeting in December 2014.

Other countries or institutes

1. The office is currently developing a Big Data Strategy. The office's long term vision for Big Data is to build a world class national statistical capability to transform large, diverse sources of rich data into purposeful new insights for informed decision-making by government, business, academia and the public. In articulating this vision, the Big Data Strategy seeks to:
 - Enhance the reach, visibility and utility of the office's statistics
 - Ensure that the privacy of individuals is protected
 - Maintain public trust in the integrity of the office
 - Position the office at the forefront of statistical practice
 - Align the office's planning with whole of government directions

At the heart of the Big Data Strategy is an integrated multifaceted capability for systematically exploiting the potential value of Big Data for official statistics. The dimensions of this capability are:

- A skilled and innovative workforce able to interpret information needs and communicate the insights gleaned from rich data
 - Advanced methods, tools and infrastructure to represent, store, manipulate, integrate and analyse large, complex data sets
 - Routine acquisition of a diversity of government, private and open data sources for statistical purposes
 - Safe and appropriate public access to micro-data sets and statistical solutions derived from an array of data sources
 - Strong multidisciplinary partnerships across government, industry, academia and the statistical community.
2. For the office, Big Data is a new implementation concept. In the framework of official statistics development, the office considers the use of Big Data as an important tool for improving quality statistics. In addition Big Data will include wider range of users which correspond on increasing request for official statistics. We haven't defined yet business processes for Big Data Sources integration and the office has not yet any partnership for Big Data
 3. Big Data is explicitly cited as a scope of duties in our midterm work program. In analogy to the Vision 2020 it has been recognized as an important field and will be included in the process of discussion of the internal strategy 2020. Although we do not have defined business processes concerning Big Data yet we are involved in national as well as international projects (i.e. the UNECE project on Big Data in Official Statistics). Concerning current projects with Big Data reference we are attempting to obtain data especially in the fields of mobile phone data, scanner data and road pricing data.
 4. The strategies and projects of Big Data are still been developed in our institute. Also, the solutions adopted are in a trial mode. Therefore we can say that we believe that Big Data will be used in a long-term vision. Yes, there are specific partnerships that we are trying to obtain.

5. The office views Big Data as an evolutionary force in modern statistics. We continue to explore new tools, techniques and data sources as they become available and as needs arise. By and large Big Data projects are considered when they fall within our mandate to provide our citizens with accurate and timely information that matters. Indeed the use of what is effectively administrative data, albeit of increased volume, velocity and of much greater variety, is subject to the same rigorous data quality assessments as those emanating from more traditional sources. The office does not currently have a formally defined business process specifically geared towards "Big Data". However, there are well defined legislations (Section 13 of the Statistics Act), directives and guidelines that specifically address the integration of "administrative data" sources and their results into our work. Note that at our office, "administrative data" is defined as information relating to individual persons, businesses or organizations that is collected by other organizations and departments for their own purposes. Hence it includes other alternative data sources such as Big Data. In other words, as an evolution in statistics, Big Data sources, techniques and tools are not considered disparate entities. While these data may sometimes demand idiosyncratic software and hardware solutions, they are nonetheless, in the end, data. The office continues to find some solutions within their existing relationship with the SAS Institute, having successfully tested a distributed computing system (SAS GRID) on the analysis of Smart Meter data. SAS is also able to integrate the R open source Data Analytics software providing access to an ever growing repository of data mining, visualization routines and a host of other Big Data tools. We continue to negotiate the acquisition of various new data sources such as retailer store scanner data and high resolution satellite images.
6. We think that Big Data usage will be a must for our statistical office in the future. However, we have not defined business processes for integrating Big Data Sources and results into our production yet. We are currently looking into the possibility to obtain mobile phone data as this might be useful for some statistics, such as tourism and telecom statistics.
7. Enhanced services. New services and business partnership opportunities. Improved policy development. Protection of privacy. Leveraging of governments investment ICT.
8. Our office has not developed a long-term strategy on the usage of Big Data. However, we understand the necessity to diversify data sources. We think that more activities should be carried out to involve Big Data sources into the statistical business processes and measures should be taken for assuring their quality.
9. According to our long-term vision, implementation of Big Data (BD) projects is foreseen. As members of the UNECE Sandbox Task Team, we are currently gathering experience and knowledge with BD and the related developments. As member of the Task Team, we are involved in the testing of BD IT tools. Our office does not have well-established procedures for the integration of BD sources and the processing of BD data in the office's business processes. We are currently building partnerships with concerned partners in our country. One example is our partnership with the National Tax and Customs Administration (NTCA). We are currently in consultation with NTCA to receive monthly data of online cash registers to be used for retail trade statistics and possibly for other purposes. Based on the outcome of the current and future discussions, the source will be a real BD source if more detailed (e.g. daily) datasets are received in the future. We also have partnership with the National Toll Payment Services Plc. As an outcome of this cooperation, our office receives

traffic webcam data. This dataset is used for tourism statistics to measure incoming and outgoing traffic at the borders of the country.

10. My office has tackled this issue in its strategy for period 2014-2016. The strategy sets long term vision to use the Big Data as possible data source for official statistics.
11. We are evaluating possibilities to use Big Data in our country. Currently we haven't defined any business processes for integrating Big Data sources and results into our work. There aren't specific Big Data partnerships and/or data sources we are attempting to obtain, only preliminary investigation on that.
12. We are currently working with Big Data on an experimental basis, and our exercises are a proof of concept. We want to acquire experience to get a clearer picture of the extent to which Big Data can be used for official statistics. Once we fully understand the potential of Big Data we could develop a long-term view about its use as parte our regular work program.
13. Big Data has been identified as a strategic priority by the Board of Directors. A corporate Big Data programme (a “roadmap”) has been agreed on. The roadmap is monitored and is updated every six months. Apart from a few statistics that already use Big Data, the roadmap lists eight statistics that may be based on Big Data in the future, two of which have first priority and will result in publications within half a year. The programme is supported by a methodological research program and an experimental IT infrastructure for Big Data. Partnerships are sought in academia, with other parts of general government, with IT providers, with data providers and with the international community of official statistics. One of the forms this can take is forging partnerships in the context of the EU Horizon 2020 research programme or other international initiatives.
14. Our office's strategy emphasize the need for access to new data sources, and specifically the use of statistical methodology and tools on Big Data to extract new and/or relevant statistics. No, we have not defined business processes for integrating Big Data sources and results into your work. We have started to look at mobile positioning data for use in tourism statistics, and card transaction data for Survey of consumer expenditure. We recognize the need to develop a Big Data strategy.
15. Our long term vision for the usage of Big Data is to make step from the classical statistical surveys and usage of administrative data to usage new data sources which will enable SURS to complement existing data sources with the BD sources. In some cases usage of BD will allow us to calculate new type of statistics, to validate existing statistics, to calculate the flash statistics and publishing a more timely statistics. One of the consequences could also be reducing a burden of reporting units and reducing a cost of surveys. The goal is to define:
 - the possible data sources of Big Data and their possible usage
 - legal aspects for the acquisition of BD
 - communication strategy with the data owners
 - communication strategy with the public and government institutions
 - establishing the partnership with the stakeholders (data owners, academia,...)
 - privacy issues
 - set up the IT infrastructure for storage and manipulation of BD

In order to define the business process for integrating BD sources in our work we have started carrying out the Pilot projects on Big Data (mobile data, price data,...) which will give us the picture how to achieve this goal. Specific public partnerships are in that moment partnerships with the data owners (mobile data, price data,...)

16. Member countries of the organization have adopted a postal Big Data vision and a postal Big Data plan. The organization consolidates real-time tracking data of more than 200 postal operators designated by their government to ensure the interconnectivity between postal networks all over the world. Tracking data is related to international postal shipments. More than 100 billion records are consolidated every year. Overall, more than 10 trillion records are available through postal and express delivery networks once domestic shipments information is added. The organization's vision and plan aim at developing greater operational efficiency of postal networks worldwide, economic and market intelligence as well as governance and strategic intelligence for the members thanks to one of the largest big physical data in the world. The plan consists in integrating all relevant internal and external Big Data sources, develop the most relevant data analytics frameworks and deliver key data products and insights in real-time to our membership.

Annex 2: Big Data Governance and management structure

- Do you have a defined structure for managing Big Data projects?
- Do you limit yourselves to certain types of data and/or partnerships?

China

The labor division of China's National Bureau in Big Data applications is as follows: Research Institute of Statistical Sciences is responsible for refining the related work that our bureau promote, communicating and negotiating with corporations, leading in the signing of strategic cooperation agreements with related enterprises; Department of Policies and Legislation is responsible for improving relevant legal system; Department of Statistical Design and Management is responsible for communicating and coordinating with different departments to promote the sharing of administrative information through data exchange; other professional departments are responsible for promoting Big Data applications work in their own direction according to their actual work; currently, there is no clear limitation about data and collaborators.

Denmark

The project on scanner data for Price statistics is embedded in the ordinary management structure i.e. it is conducted by the Price statistics unit which is part of the department of economic statistics.

Finland

Big data is still very much an open question. We do not have Big Data at this moment (other than the traffic sensor data as given in the previous answer).

Italy

1. ISTAT has one permanent statistical methodological unit and two IT units that are involved in Big Data projects. ISTAT also set up a two-year Technical Commission on Big Data with internal and external members.
2. ISTAT is investigating different alternative data sources and partnerships.
3. ISTAT set up research protocols with Italian Universities and is involved in International activities (UNECE Big Data Project) and European ones (Eurostat Task Force).

Japan

We have no problem regarding organizational management for Big Data projects, because we haven't discussed this issue. We recognize that we should consider this issue in the future, as necessary.

Portugal

We are now being evaluated to define a structure for the management of Big Data projects.

Romania

There are no Big Data projects at this moment in our NSI. For the future we intend to start 1-2 pilot projects using data from private sources and administrative data. We are evaluating the possibility of e regional collaboration in this area. Our NSI is an active member of the ESS Task Force of Big Data.

Serbia

The structure is still not defined but it will be the part of the strategy. The certain types of data and the plan to use as much as possible data sources will also be a part of long term strategy. The partnerships will be very useful, both with institutions already using the Big data (NSIs and similar) and data producers (private sector, financial institutions etc.).

Sweden

No defined structure for Big Data projects in particular. Our limitations are mainly due to limited resources within the organization.

UK

The overarching ONS Big Data Project has a clear governance and management structure. We have established an internal Project Board and Steering Committee and are running the project using standard project management principles. Each of the pilot projects is run as a work package within the overarching project. We have not limited ourselves to specific datasets; in fact the four pilots were chosen to cover a range of different sources, applications and to address different challenges around the use of Big Data within official statistics.

Each of the pilots does involve an element of collaboration with external parties:

- Smart meter pilot -initial work on this data source was undertaken by academics at Southampton University;
- Mobile phone pilot - working with mobile network operators to purchase the data, also colleagues across UK Government with an interest in this data source;
- Internet Price Scraping pilot - working collaboratively with an academic from Huddersfield University to undertake analysis of the price data scraped from the internet around the methodological issues, also engaging with MySupermarket to potentially purchase this data. In addition we have engaged with Statistics Netherlands, learning from their experiences around price scraping within a NSI. This was immensely useful and probably gave us a six month head start.
- Twitter pilot - shared progress with academics from Cardiff University, also working with academics from Southampton University to support analysis of the data.

All pilots report on progress on a 3 monthly basis, versions of the reports are published on the external webpage. At the end of the project all pilots will produce a detailed technical report (conclusions from which will be taken forward into the overarching ONS strategy for Big Data). External experts (from NSOs, UK Government or academia) have been identified to quality assure each of these pilot reports.

Eurostat

The mission will be realized through a number of concrete operational objectives, activities and projects in partnership with NSIs, other Commission DGs, international bodies and research institutions. NSIs are expected to contribute to the execution of the action plan by participating in Big Data pilots and working on horizontal issues in a collaborative way related to the identified topics. Eurostat would contribute to the required investments by means of setting up ESSnets, centers of excellence, defining research programs or contracting pieces of work to specialists.

ESCAP

The Strategic Advisory Body for Modernization of Statistical Production and Services in Asia and the Pacific (SAB-AP) was created under the auspices of the Committee on Statistics and

represents the regional governance structure for all modernization work, including Big Data. The primary objective of the SAB-AP is to drive and support changes towards the modernization of statistical production and services in the Asia-Pacific region. Its members shall provide strategic direction to this area of work, develop regional strategy, and mobilize financial and human resources. At the moment the SAB-AP has nine high-level members from NSOs in Asia-Pacific and is chaired by Australian Bureau of Statistics. Regional discussions have identified an inadequate level of awareness and sense of urgency on the part of NSOs in relation to the modernization of official statistics and the use of Big Data. In response, the SAP-AP identified Big Data as one of its priority areas of work. Future Big Data projects may be taken up by the Modernization Working Group (MWG) established by the SAB-AP. Members of such MWG are technical level staff from NSOs in the region that are interested to participate. This group could also closely collaborate with other organizations in the region, mainly the UN Global Pulse Lab in Jakarta, and possibly other private sector players.

Other countries or institutes

1. The Methodology Flagship Reference Group was established to oversee the progress and delivery of outcomes for all Flagship Projects, of which the Big Data and Semantic Web / Data Visualization projects are two. These two Big Data related projects report to the Reference Group on a quarterly basis. In addition the Analysis Steering Committee is a Deputy Statistician level committee that oversees and approves the entire work program for Analytical Community. While we may be interested in social media data in the future, we have made a conscious decision to focus upon structured data such as satellite data, the Linked Employer Employee Datasource and the Semantic Web technologies in the early stages of our Big Data Flagship project work. In deciding what data sources we harness, we have a guiding principle that the outcomes must be driven by clear and well supported internal business objectives, which determine the most suitable data sources to explore.
2. We don't have a defined structure.
3. We have a small task team consisting of two people from the methodology department and one person from our core unit for analysis. This team was founded to get an overview on national as well as on international level and we take part at international initiatives like the UNECE Big Data project. In general we don't have constraints concerning potential Big Data projects but from a midterm perspective we focus on machine-generated and transactional data.
4. We do not have a specific structure for managing Big Data projects only. Mainly regional collaborations are involved as partnerships.
5. Our nascent Big Data efforts are all effectively managed within our current governance structure which relies on an Integrated Strategic Planning Process, a Departmental Project Management Framework and Integrated Risk Management. They, like all other data from an external source, are considered "administrative data". As such, they are subject to all of the existing data confidentiality provisions as outlined in the Statistics Act and the Privacy Act. We do not, a priori, dismiss any data source. All data sources must conform to our Quality Assurance Framework which includes the dimensions of relevance, accuracy, timeliness,

accessibility, interpretability, and coherence. When applied, these criteria effectively gauge the data's fitness for use, but the use itself must pass the test of relevance which is paramount as the office strives to continuously reflect the country's most important information needs. It should be noted that many new initiatives are still at the exploration phase. Hence, they are carried through as pilots or as research projects with the goal to determine a proof of concept for the support or production of official statistics before engaging further.

6. We have not defined structure for managing Big Data yet.
7. No, we don't have any specific structure or any project done till now
8. The office has not created any specific structure for managing Big Data projects. Nowadays we limit ourselves to administrative data received from several government ministries and agencies (e.g. Public registry, Civil registry, Tax officials etc.), local authorities etc. In our opinion, it is desirable to attract resources of external cooperation projects.
9. The office has already established a small team with experts from IT and Methodology Departments. When actual work with BD sources begins, the team will focus on a few (1-2) types of data. The office currently uses free data sources or data sources managed by governmental agencies (based on existing contracts between the concerned agencies and the office). We are considering to use other sources and tools (such as web scraping) and constantly looking for other partnerships to ease the reception of additional data and to extend the use of possible BD sources.
10. Currently we have not defined the structure (organizational) for managing Big Data projects. The office has identified that Big Data can be used to enhance population statistics (migration), tourism statistics and workday labour mobility.
11. We are only on idea phase so we haven't defined structure for managing Big Data projects. We are looking for all possible types of data and data providers.
12. We have set up an internal interdisciplinary task force for our experimental work. To conduct this work we have established an informal relationship with the University of Pennsylvania to learn about the methodological aspects for the exploitation of social networks' raw data, and other universities for technological and the operational aspects. As for the data sources, our current work relies on publicly available Twitter information which is being downloaded to build a data base. In the future we could use other sources, provided they are easily available and that we do not violate the persons' privacy rights.
13. The use of Big Data in statistics is a decentralised responsibility, but in most cases actual production is preceded by research and innovation, which are both organised at the corporate level. At the level of the Board of Directors, the Deputy Chief Statistician has overall responsibility for the implementation of the Big Data strategy.
14. Not at the moment * No * Not at the moment.

15. Governance of BD is undertaken in the form of pilot projects. At the office we have standard management structure for running projects which have infrastructural meaning or aim of major revision of the Statistical Survey. We try to include BD in existing structure of projects which aim is broader than BD. When we launch a new project which aim is revision of some Survey we try to include possibility of usage of new data sources (such as BD). We are also carrying out the projects which aim is to research potential usage of BD (e.g. mobile data). Big Data sources which are involved in Pilot projects are maintained by private companies (partnership).
16. The governance of the project is ensured by two projects groups dealing with postal economics and statistics issues. New proposals will be put forward during the congress. A cooperation agreement between the office and the United Nations Global Pulse (UN Executive Office of the Secretary General) was signed in September 2014 and postal Big Data has started being shared between both organizations to conduct a number of data analytics for development projects. Data sharing often requires the formal agreement of the postal operator ready to share the data with UN Global Pulse, particularly when data is provided at the disaggregated level and/or not normalized.

Annex 3: Big Data Quality

- How do you assess the statistical quality of Big Data sources and/or of the output of analysis based on Big Data sources?
- If you have frameworks for assessment discuss that here.

China

Currently there is no relevant assessment work ongoing.

Denmark

Very varied.

Finland

When the traffic sensor data is in question, specific assessment can be sent separately. Otherwise we do not have Big Data yet.

Italy

ISTAT is contributing to the definition of the quality framework within the UNECE Big Data project. We have some pilots experimenting: (i) quality of estimation based on mobile phone data as compared to the one based on administrative data; (ii) quality of estimation based on Internet data as compared to survey-based estimation.

Japan

We haven't discussed this issue. Generally speaking, we think that most Big Data is different from the target coverage of the statistical surveys. On the other hand, it will be possible to improve Official Statistics, using Big Data to complement existing statistics, and to develop new big-data-based statistics which have never been available from traditional data sources. In those cases, sufficient discussion about accuracy assurance of statistics will be required.

Portugal

The quality of the sources is for us a very important aspect. So far we have no experience in the study of Big Data sources. Statistics Portugal has a large experience with administrative sources, and a major concern with the quality of these sources, that experience and some of the procedures in the administrative sources, will be very useful in the analysis of quality of Big Data sources. At this moment we have not yet a framework for evaluating Big Data sources.

Romania

The quality of Big Data sources will be primarily assessed using the existing statistical quality framework for the data sources. Once we will start the pilot projects we will develop a specific quality framework for Big Data sources.

Serbia

Not applicable for now.

Sweden

No framework in place. We participate in the UNECE Big Data project, the AAPOR task force on Big Data, and we have an application for Horizon 2020. All of these involve work on quality frameworks for Big Data. In addition, we have long experience from using administrative registers and we see similarities.

UK

To date we are only within a research and feasibility phase of the pilot projects so have not yet defined a quality assessment framework for Big Data sources or the output of analysis. However, we are undertaking standard quality checks on the data as one would with administrative data before analysis.

Eurostat

The provision of high quality information is one of the corner stones of official statistics. A new framework for quality assessment of official statistics using Big Data sources in the ESS should be developed and this has been included as one of the medium-term (by 2020) aims for the Task Force.

ESCAP

We do not have a framework for assessment of data quality at this moment, however, the important criteria for the assessment of quality are: relevance, accessibility, representativeness and sustainability.

Other countries or institutes

1. The office, in partnership with participating agencies to the UNECE Big Data Quality Project, is developing a Big Data Quality Framework for assessing the quality of Big Data sources and statistical products derived from those sources. We are also exploring to use of semantic web technologies for assessing conceptual relationships between concepts implicit in the data, and visualizing those relationships. This should help reveal quality issues. In addition, we intend to use statistical and machine learning methods to produce quantified and defensible measures of quality.
2. As we don't have data yet it is too early for us to assess quality.
3. It depends. In case of statistics data use, we do not see problems of quality in an integration process. But if the Big Data system inputs administrative records, quality could be a concern.
4. We currently use the earlier mentioned Quality Assurance Framework as a guide. We have also recently developed tools to assess the quality of administrative data at the input stage. In parallel, we are currently collaborating on the UNECE HLG Big Data Project task force to develop a framework that may include some added hyper dimensions (such as "source", "metadata" and "data") that could help guide the assessment process with regards to some specific types of Big Data when used for Official Statistics.
5. We have not taken any measures yet.
6. The quality of Big Data (Administrative data) is, in general, acceptable. After receiving them a thorough validation and editing processes are carried out to assure the highest possible quality.

7. No such framework exists in the office. According to our limited experiences, the quality of the data received might significantly differ over time: at the time of the introduction of online cash registers, the collected data is less accurate compared to the dataset collected at later stage, when the use of online cash registers is more established.
8. The office has not made quality assessment of Big Data sources as well as output. The usage of Big Data is long term vision.
9. We haven't assessed the statistical quality of Big Data sources and/or of the output of analysis based on Big Data sources because we are only on idea phase.
10. In terms of content, we have obtained information that otherwise would almost be impossible to get with a survey, because people are freely expressing their sentiments and emotions without being asked in a questionnaire how they feel. For an assessment of subjective well-being this is a very valuable source of data. In terms of timeliness, this source has proved to be very good when the data are used for other experiments we are undertaking such as internal tourism flows on certain holidays. No other source could provide the same timeliness on a broad topic. In terms of the output, we will perform correlations with other sources of statistics to assess the quality of the results under the traditional approach of official statistics.
11. The usability of Big Data for statistics production is assessed with the general framework for the determination of the input quality of secondary data sources that has been developed in the BLUE-ETS project. Where possible, statistical methods are chosen from the repository of validated statistical methods. New methods have also been developed for some specific Big Data sources. Because this is a new and uncharted terrain, the acceptance of methods and statistical results also depend on professional judgement. There are also cases where known relationships with other sources or statistics can be taken into account.
12. We have not yet made a detailed assessment of the data quality of different Big Data sources yet.
13. First experiences with the price data (BD from the supermarket chains) show us that the quality of the Big Data sources is excellent. We assess that BD sources have a potential to be at least the complementary (together with the administrative and/or survey data) source in some Surveys. Regarding the quality framework we are the member of the UNECE BD Quality Team which will prepare the quality framework for Big Data. In this framework we adopted the similar concepts as in case of quality framework of administrative data (Input, Throughput and Output Quality Dimensions).
14. 80 to 90 percent of our data analytics work is spent into data management where the consistency and quality of Big Data sources are checked very carefully with a variety of statistical methods. Moreover, member countries are also trained in order to improve the real-time collection of postal shipments tracking data, more particularly regarding the scanning operations triggered by any postal shipment at the various stages of the international postal process, from origin to destination including transportation.

Annex 4: Big Data Privacy and confidentiality

- Do you have a framework for dealing with privacy issues?
- Please include any other comments regarding privacy and confidentiality.

China

There is no framework dealing with privacy issues. If there is no network security, there is no national security. We should accelerate the formulation of basic laws in the process of collecting, using and releasing data involving personal privacy, trade secrets and government confidential issues in order to protect national information security and safeguard the legitimate rights and interests of all parties.

Denmark

We have an extended framework for dealing with privacy issues; i.e., a policy decided by the Board.

Finland

Yes.

Italy

1. ISTAT is contributing to the definition of a privacy framework within Eurostat Big Data TF.
2. The activity of defining a framework for dealing with privacy issues is in progress.
So far, we have reached an agreement with the National Privacy Authority to deal with anonymised mobile phone data.

Japan

We haven't discussed this issue. Generally speaking, we think that this issue may be covered by the Personal Information Protection Law (e.g. ensuring confidentiality), because personal information is included in Big Data in large quantities. Moreover, although it is not related to the use of Big Data directly, the related organizations are currently exploring the best way for private sectors to utilize personal information in their various business areas while maintaining the protection of privacy.

Portugal

SP is able to deal with issues of privacy and confidentiality, so we think it will not be much different from what we already do with other sources. But, we will give special attention to Big Data sources, when in practice and we begin acquiring knowledge.

Romania

The privacy issues will be treated according to the national legislation regarding data access.

Serbia

Not applicable.

Sweden

Not for Big Data in particular.

UK

The ONS is committed to protecting the confidentiality of all the information it holds. In order to produce statistics using Big Data sources we are only interested in trends or patterns that can be observed not data about individuals. However, we recognise that accessing data from the private sector or from the internet may raise concerns around security and privacy. We have therefore committed to only access publically available, anonymous or aggregated data within the Big Data Project and this data will only be used for statistical research purposes. In addition all of our work will fully comply with legal requirements and our obligations under the Code of Practice for Official Statistics.

Part of the research within the ONS project will consider the ethical issues associated with using these types of data sources within official statistics. To date we have engaged with a number of privacy groups to provide an overview of the project, with particular focus on the pilot projects and the privacy challenges they bring. We have had positive feedback about this early engagement from the groups who also provided advice on handling, communications, policy in this area and future directions with the pilots.

In addition we have begun discussions with the Economic and Social Research Council to explore the possibility of jointly funding research into public attitudes of the use of big data for research/official statistics. One of the key deliverables from the ONS Big Data Project in this initial feasibility phase has been an ONS Policy on Big Data that sets out operational principles and guidance for how ONS will move forward with Big Data to ensure we maximise the statistical benefits while maintaining our reputation for protecting confidential information and trusted statistics. Our Code of Practice for Official Statistics does not make specific reference to Big Data and there are several areas where new policy and guidance is needed. The main gaps are around ethical use of personal data and use of public data sourced from the web. The main focus of the policy is on new forms of data from commercial organisations containing information about individuals and the use of these data for statistical purposes. A proposal has been put forward to establish an Ethical Committee to support the implementation of this policy.

Eurostat

Legislation plays a crucial role in determining the framework conditions for accessing, processing and disseminating statistics derived from Big Data sources. It refers partly to protection of personal information and privacy of natural and legal persons and to intellectual property right. In general, directives define minimum requirements that can be further refined at national level while regulations are directly applicable at national level. Depending on the type of legislation, this might have consequences as regards harmonisation of Big Data processing within the ESS.

ESCAP

We do not have a framework for dealing with privacy and confidentiality at this moment, however, privacy and confidentiality is of utmost importance. In many countries in Asia-Pacific the NSO and other agencies in the NSS have a strong reputation. This reputation is based, among others, in trust and independence. Hence, the NSOs should ensure that the data they use remain confidential and will be used exclusively for statistical purposes.

Other countries or institutes

1. The statistical office is legally mandated by the Census and Statistics Act 1905 and the country's Statistics Act 1975, and related determinations and policies as well as other Commonwealth acts of parliament such as the Crimes Act 1914 and the Privacy Act 1988. At

the heart of our legislative basis is the requirement to protect the confidentiality of the individuals and organisations that provide the office with information used for statistical purposes. Privacy and Confidentiality are key tenants to the statistical office's core values and the country's Public Service Code of Conduct and is paramount to the high level of trust and reputation the country's public have in the statistical office, and the high quality of the data our providers supply us with. The use of Big Data amplifies the 'mosaic effect': distinct pieces of data that pose no privacy risk when released independently may reveal personal information when they are combined. If there is a disclosure risk then the office does not disclose – irrespective of the analytical utility of the data set. In terms of internal governance, the statistical office has a well-defined and long established set of processes for acquiring external data sources under the legislation and a Disclosure Review Board (DRB) which advises on whether datasets proposed for release or micro level statistical outputs are suitably confidentialised under the terms of our legislation. DRB advice is required before any micro level data can be released externally. In addition the Data Integration Steering Committee assesses and approves proposals for the linking of multiple datasets, with regard to the feasibility of linking the datasets with a level of accuracy that meets the office's standards, the relevance to well justified business requirements and any privacy issues the public may have in relation to the linking being undertaken. The statistical office believes that this policy framework is sufficient for dealing with the privacy and confidentiality issues relevant to Big Data sources.

2. We have the legal framework which ensures the privacy and confidentiality principles.
3. The same strict rules based on legal acts (e.g. the country's Federal Statistics Act) and internal guidelines for statistical disclosure apply also for any other statistical procedure.
4. Yes we have. Privacy issues are treated according to legal patterns existing in our country. As well, confidentiality too.
5. The notions of privacy and confidentiality are legislatively and institutionally ingrained into the fabric of the statistical office. Through the Statistics Act, and the Privacy Act, the office continues to enforce strict directives in the protection of individual privacy and confidentiality. The purpose of the Privacy Act is to ensure the protection of the privacy of individuals with respect to personal information about themselves held by a government institution and the provision that individuals have a right of access to that information. In support of the Privacy Act, a Privacy Impact Assessment (PIA) is an evaluation process which allows those involved in the collection, use or disclosure of personal information to assess and evaluate privacy, confidentiality or security risks associated with these activities, and to develop measures intended to mitigate or eliminate identified risks. These are reflected throughout the process from the stage of acquisition right through to dissemination and are governed by a number of internal policies and directives. As an example, considering that "administrative data" are often linked to other data holdings, and because record linkages are invasive by their nature, the office undertakes them only in cases where the public good is clearly evident and outweighs the privacy invasion. The process via which the office evaluates the merits of a record linkage is laid out in our internal Directive on Record Linkage.

6. We do not have a framework yet. However, we think that various legal issues have to be defined and solved concerning the privacy of the respondents and the legal framework for accessing these data from the owners, such as telecoms etc.
7. We have laws that punish anyone leaking data. The office can't disseminate Personal data.
8. The privacy and confidentiality of data are protected by the Law on Official Statistics and some other legal acts.
9. No such framework exists in our office. Strict data protection regulations are in force in our country. On one hand, these regulations are safeguards, ensuring that data managed by the country's statistical service are strongly protected. On the other hand, it is also a barrier for our office to get access to additional datasets.
10. The office has great experience with data from administrative sources, so the primary data protection will not create a security treats. Protection measures for statistical output are not analysed at the moment.
11. Currently we haven't had any framework for dealing with privacy issues. Only standard framework which deals with a standard data. We will try to adopt it to the needs of Big Data.
12. The office does have a legal mandate to preserve the respondents' privacy and confidentiality for all statistical undertakings. For our experimental work, we deleted all the information that could make it possible to identify individuals or their location. Nevertheless, from a broader point of view, we recognize it will be necessary to accommodate the privacy frameworks to Big Data projects, considering that Big Data, in general, could effectively undermine privacy and therefore damage the perception of peoples' confidence in national statistical offices.
13. For privacy and confidentiality, the regional legislative framework applies. In addition, the country's statistics law and the country's general law on privacy apply. In general, the framework allows using Big Data for research purposes. Also, the statistics law defines a procedure to bring the production of specific statistics under the law, including data collection provisions. In practice, this procedure has not been applied to the use of Big Data yet. Furthermore, public image considerations make us more restrictive in respect of privacy and confidentiality than strictly required by law. Whenever a Big Data source is going to be studied, a checklist is filled in to assess the privacy and confidentiality considerations.
14. Yes. We use the same frameworks as we do in the production of all types of statistics that involve sensitive data.
15. The office has a very strict policy regarding privacy issues. This policy follows the regional Statistics Code of Practice (Principle 5: The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes must be absolutely guaranteed). Experiences with the price BD show that the office must be very careful with sensitive data. We are not allowed to share this BD data with other institutions and we also have to secure IT infrastructure for the manipulation of this data. Possible strategy in order to insure this

policy could be to have different approaches regarding the sensitivity of data. If Big Data is not sensitive we could store the data in our own database or in the "common" cloud. In the case of sensitivity of BD, our office will offer two options to the data owner: detailed specification of the agreement (contract) with the owners of the data where should be stated mode of acquisition of data, usage of the data, persons who have access to the data, possible linkage with other data sources, mode of erasing the data, etc. –The office receives the testing set of BD. The office then does all the testing and manipulation with the data then does all statistical process at the location of the BD owner. In this case only aggregates will be transmitted to our office.

16. A number of the office's regulations protect personal data as well as commercially sensitive data. However these regulations were developed before the rise of Big Data and will require a number of adaptations so that a new balance is found between privacy on the one hand and the opportunity to improve postal services for citizens on the other hand.

Annex 5: Big Data Skills

- How have you addressed and/or are planning to address skill shortages in Big Data?
- What skills do you consider to be most important for staff working with Big Data?
- Are you actively hiring data scientists?
- Are you training existing staff?
- What has been the most helpful?

China

We need compound talents and teams in the development and use of Big Data. The mastery of multiple disciplines such as humanities, sciences and arts and a thorough understanding of the running track of society and human behavior contribute to realization of Big Data concept. Although China's National Bureau of Statistics has not hired data scientists at present, the education centers and other departments have organized Big Data training for staff at different levels in order to raise Big Data awareness and promote relevant work gradually.

Denmark

We are aware of the potential skill shortages but we do at the moment not address them.

Finland

First, we need Big Data. But generally speaking the use of Big Data requires strong statistical and IT skills together.

Italy

1. Machine learning, design of algorithms, parallel computing, statistical learning, text analysis.
2. We are not hiring data scientists due to shortage of resources.
3. We have not yet training programmes.

Japan

We recognize that this issue will be raised if we adopt a serious stance against the production of the Official Statistics using Big Data, and that we should consider this issue in the future, as necessary.

Portugal

SP has a great autonomy and expertise in information technology. Many of the tools such as SAS, SPSS and R are dominated by a large number of technicians. On the other hand, there is also a wide experience in ETL processes. Our physical technology infrastructure is already implemented on cluster nodes and this will be an advantage. We planned to do some training and attendance of workshops on Map Reduce, Hadoop, Hive and Pig.

Romania

We consider that the most important skills in this area are: IT skills: programming languages (C/C++, Java, etc.), databases manipulation (SQL), software products for statistical data analysis (R, SPSS, Matlab), machine learning techniques; maths and statistics skills; Starting with this year we have a training course regarding "R and Big Data Analysis" (one-day course). One university from our country has a master program for official statistics (it is part of the EMOS

program). Hiring data scientists is very difficult because we have to compete with private companies that offer greater salaries.

Serbia

The most important skills will be understanding and operability with complex information system, high educated methodologist with the vision and creativity for getting the most out of the velocity of data and statisticians capable for data mining and complex data analysis. For now SORS engages 5 PhDs and encourages young staff for further education in the university. Also, few of the employees are already familiar with the term of Big Data, through several workshops organized by Eurostat or other relevant institutions. We are of the opinion that workshops, conferences and available papers are the most important especially when the materials are shared with other colleagues.

Sweden

No training or active hiring yet, but this is something we are discussing.

UK

Within the ONS Big Data Project our approach has been to train existing staff. We have established a team with a mixture of backgrounds; methodological, social researchers, IT who are working together on different aspects of the projects. Many of the team have undertaken relevant training courses, in particular to learn new Big Data technologies or programming languages, quite often through online courses. As covered above we are working in partnerships with external bodies and have brought in consultants to support some of the technical aspects of the project but have not hired in data scientists. For a longer term strategy we are engaging with universities offering courses in data science/analytics with one angle being to investigate the opportunities for attracting new graduates of these courses to join ONS, in particular we are looking at possibility of taking on placement students from these courses.

Eurostat

The preparation of a definition, an inventory and a strategy for acquiring the necessary skills for the ESS will be essential for success and is part of the action plan for the short-term (by the end of 2016).

ESCAP

UNESCAP role is mainly to coordinate and facilitate the work on Big Data on the regional/global level. This means that the required data skills have to be owned by our partners, the NSOs and other partner organizations, which will need to build or acquire human resources with data skills.

Other countries or institutes

1. The most important skills for staff working with Big Data are a wide ranging knowledge and understanding of mathematical and statistical methods that are relevant to solving Big Data problems. The office is carrying out a multi-pronged strategy in developing and acquiring technical skills through:
 - Developing internal staff through on the job training, courses (where relevant) and presentation of papers at conferences and workshops,
 - Bringing in staff with specialist skills through contractual arrangements and transferring those skills to existing staff,

- Engagement and collaborations with other government agencies and academic research institutions.
 - Collaborations with international agencies such as the UNECE Big Data Project.
2. The following skills are considered to be important for the staff working with Big Data; Mathematics, Economics, Econometrics and IT.
 3. We don't see a big skill shortage at the moment. Relevant skills are distributed among different units, especially in the methodology unit. We take part at conferences, international initiatives and courses like the UNECE Big Data Project Sandbox training. We also plan to take part at ESTP-courses covering Big Data.
 4. We are training our personnel, but we consider the institutional integration the most important characteristic to be applied.
 5. From our observations thus far, imagination and adaptability are perhaps the single two most important attributes with regards to Big Data. Their unconventional form and size can sometimes overwhelm and obfuscate what is possible with current technology. To foster the internal growth of Big Data science, a Big Data Community of Practice has been in existence for 2 years, a modest "Big Data lab" where new emerging software can be assessed and Massive Open Online Courses (MOOCs) can be followed, has recently been made available. Training opportunities regarding Big Data topics have also been shared on an internal Wiki dedicated to Big Data for the benefit of all employees. While the office is not specifically recruiting "data scientists", post-secondary institutions in the country are adapting their curriculum to cover the mathematical, statistical, theoretical and technical components of Big Data "science". Corporate level HR planning is exploring various ways, including partnerships with postsecondary institutions, to ensure a source of supply for needed skills. In addition, as the office continues to foster a professional work environment where innovations and research are valued, encouraged and applauded, there is no doubt the number of "data scientist" in the office will organically grow.
 6. Methodological skills should be very important as well as IT skills dealing with unstructured data and large datasets. We are not hiring data scientists or training existing staff.
 7. The most important skills for staff working with Big Data are Customer-driven, Curiosity and Intuition. Yes, we are hiring data scientist and training existing staff. Both of them help in the development of work.
 8. In our opinion, statistical skills like methodology for processing Big Data, standards for processing, strong computer skills etc. are most important together with other skills like creativity, communication, initiative etc. At this moment we are not hiring data scientists and are not training the staff.
 9. Hiring data scientists is currently not planned and foreseen in the office. Due to financial constraints, this situation is not likely to change in the near future. Further training of the current staff is a more realistic step for the future. Subject-matter statisticians need further training, in general, in IT skills (such as: programming). In general, office staff needs

substantive expertise in the regard of BD management (possibly from private sector or from other partnerships).

10. The office has strong Mathematical support unit which has experience using record matching techniques which somehow relates to Big Data analysis. We are planning to build on the existing experience by providing adequate training to the staff.
11. We think that most important for staff working with Big Data is integrated knowledge of IT and statistics which can provide new methods and adopt new tools for working with Big Data.
12. We have created task force with persons with different profiles that complement each other: statisticians, IT professionals and economists. Ideally the persons should be strong in IT and statistics or applied mathematics, but this profile should be complemented with someone who has an overview of the social, economic and environmental phenomena we intent to study through the use of Big Data.
13. The need for training and perhaps hiring data scientists is recognised, but no policy has been drafted yet. In the meantime, the research and innovation programmes provide training in practice, since they contain a number of Big Data subjects. Moreover, a group of researchers has organised its own training by means of free internet courses, discussing books on the topic, attending meet-ups, etc. There are also many contacts with academia and other statistical organisations. For instance, we participate in the UNECE Big Data Sandbox experiments, for which special training courses were held in Rome and in the Netherlands.
14. Statistical Methodology * No * No * N/A
15. In the office, we are aware that there is a need for the new type of statisticians (Data Scientist) which could manage the Big Data in the most efficient way. Currently we collaborate in international BD projects (UNECE, Eurostat) where we share experiences with the statisticians who gather IT, statistical and subject matter knowledge. The office will launch in short time period education programs which aim is to educate such statisticians. We will be also in contact with the Academia in order to get knowledge about this new type of statisticians.
16. Five people were recently trained in Big Data analytics or Hadoop. Data analytics skills and IT engineering are essential but are usually not available from one staff member only. Collaborations with universities and other international organizations are also developed so as to enable the necessary knowledge transfer.

Annex 6: Big Data Projects by main area of use

General/Cross-cutting

Title	Organization	Overview
Developing a Curriculum and Training Modules on Using Big Data for Official Statistics	UNESCAP	To use "Big Data" in producing official statistics, statisticians and managers in national and local statistical systems need to closely examine and understand the potential use of various types of Big Data as well as the issues and limitations of their usage for producing official statistics and indicators. Statisticians and managers also need to gain and/or improve their knowledge and skills for working with such data and integrating such work in the standard statistical business processes. This project aims to develop a training curriculum to address capacity-building requirements on understanding and assessing the potentials for utilizing Big Data for official statistics, particularly in developing statistical systems of Asia and the Pacific, based on an assessment of knowledge and skills levels of their human resources. The curriculum will serve as an integrating framework for developing and conducting training modules focusing on Big Data utilization in specific domains of official statistics.
Inclusive infrastructure for sustainable development	UNESCAP	The link between broadband access and socio-economic development has been examined from a variety of approaches. However, most approaches do not incorporate consideration of the delivered broadband quality as experienced at the consumer level. Because the quality and performance of internet connectivity varies greatly based on local conditions, a robust analytical approach should incorporate these nuances into research exercises. This research activity draws upon large-scale user data to derive industry standard indicators, such as jitter, lag, packet loss and delivered speeds. These indicators enable the researcher to draw up on this quantitative data to gain a more detailed picture of ICT connectivity, which enhances the accuracy of analytical exercises.
Capacity Building in using Big Data as sources for public statistics	Cameroon NIS	Capacities building to develop national skills in processing Big Data for official statistics. Share experience and benchmarking with countries ahead on the topic.
Capacity building for the use of Big Data for statistical purposes in Cameroon	Cameroon NIS	As developing country, Cameroon needs to build capacities and skills on this new domain in statistics. We are exploring these opportunity in line to learn and adopt methodologies and processing. We expect it be cost less than classic surveys. But challenges seem to be huge in term of coverage and public acceptance to collaborate.

Other projects

Representation of statistical information using semantic technology

Big Data Visualisation

Automated Coding based on content extraction from textual data

Automization of data recovering and integration

Use of web activity data for the production of flash estimates

Agricultural statistics

Other projects

Harvest statistics based on satellite images

Satellite and Ground Sensor Data for Agricultural Statistics

Demographic and social statistics (including subjective well-being)

Title	Organization	Overview
counts of individuals held on a major commercial marketing database comparison to Census data estimates	UK Office for National Statistics	Counts of individuals by age band and sex were obtained from the data provider Experian. The counts were based on their commercial marketing database - a foundation of edited electoral roll plus various other data sources including large scale continuous surveys fielded by Experian. The counts were compared with Census data.
Smartmeter type data for household structure/size and occupancy	UK Office for National Statistics	This is exploratory research, commissioned out to academia, into the potential of electricity smartmeter type data to identify household structure and size. A second objective is to research models to see if probability of occupancy by time of day might be derived. Smartmeter data will be collected on all households in England by 2020. The minimum specification is energy usage every 30 minutes per meter. Data will be centralised and might be available for research (details/legislation still to be formally agreed). This research is being conducted on data from trials of energy use.
Smartmeter data potential for detecting unoccupied dwellings	UK Office for National Statistics	Very much exploratory research. ONS has acquired electricity smartmeter data from trials of energy usage. This data has various potential uses within official statistics - the focus for our work is currently on occupancy. Another objective is to familiarise ONS with the methods and technologies needed to handle this type of data. Awareness of long-term vacant properties would help within survey sampling - knowing which areas of the country have high levels of unoccupied housing would be of benefit to fieldforce logistics as well as improved sample designs. Extension of this research,

		collaboratively across government, might be to compare estimates derived from smartmeter data with the alternative official sources of data on vacant properties: to highlight if smartmeter data has any advantage in producing in terms of cost, timeliness, geography or accuracy.
Tweet analysis	INEGI Mexico	Since tweets are publicly available, since a number of academic projects are looking into how to use them, and since a continuous 1% georeferenced sample can be obtained for free, we decided to use this data source in order to get first-hand experience regarding technological and methodological requirements, including human and material resources, and their capabilities, for tackling Big Data projects for the production of official statistics. All of the above while assessing tweet's usefulness in three specific application areas: subjective well-being, tourism and border mobility. For the first we partnered with a university to get students to classify content. For the other two, counting change of location rather than content is the issue.
Other projects		
Official statistics use of mobile phone positioning data with a particular application to building population grids		
Production of short-term indicators of household budgets		
Population Estimates		

Economic and financial statistics

Title	Organization	Overview
Big Data Enterprise Statistical Indicator Ten-day Report	National Bureau of Statistics of China	The project is mainly for research purpose, offering references for researches in economic situation and statistical production.
Other projects		
Online cash register data from retail stores; acquired from the National Tax and Customs Administration		
Sentiment indicator (based on social media messages)		

Environmental statistics

Other projects		
A Big Data Pilot Project - With Smart Meter Data		
Improvements in the CO2 emissions calculations and developments in emission indicators for air transport		

House market statistics

Title	Organization	Overview
Non-Residential Buildings Inventory: Feasibility Study	Statistics Canada	We completed a feasibility study for the development of a Canadian inventory of non-residential building (commercial, industrial, government and institutional buildings). Two prototypes were developed (one for a municipality; the second for a typology of buildings). In its most complete form, it is envisioned that the inventory will be used for storage and query, analysis and dissemination, with a specific attention to open data considerations and integration of Big Data in a spatial framework. It is expected that the inventory can be used to generate statistics from Big Data sources.
Other projects		
House market indicators (based on website information)		

Information society / ICT statistics

Title	Organization	Overview
Internet as a Data Source for ICT Usage by Enterprises and Public Institutions	ISTAT	<ol style="list-style-type: none"> 1. This experimental project considers the Istat sampling survey on “ICT in enterprises”, that aims at producing information on the use of ICT and in particular on the use of Internet by Italian enterprises for various purposes (e-commerce, e-recruitment, advertisement, e-tendering, e-procurement, e-government). 2. The experiment investigates the possibility to collect a subset of the data collected by means of the questionnaire, also by directly accessing to the websites owned or used by the enterprises. The aim is twofold: (i) from a technological point of view, verify the capability to access the websites indicated by enterprises participating to the sampling survey, and physically collect all the relevant information, (ii) from a methodological point of view, use the information collected from the Internet in order to predict the characteristics of the websites not only for surveyed enterprises, but for the whole population of reference, in order to produce estimates with a higher level of accuracy (to be evaluated). 3. Once this alternative approach will be proved to offer a quality of obtainable estimates higher than that of the traditional approach, the new process could become an important part of the survey on “ICT usage by enterprises and public institutions”. It will also be possible to consider not only to increase the accuracy of already available estimates, but also to produce new estimates related to additional information currently not covered by the survey. 4. Purpose: production of statistics based on Big Data

<p>Analysis of methodologies for using the Internet for the collection of information society and other statistics</p>	<p>Eurostat</p>	<p>The aim of this project was to assess the feasibility of employing modern and enhanced methodologies and indicators for collecting high quality statistics from non-traditional data sources such as the Internet (IaD) or Big Data Sources. Concrete objectives were:</p> <ul style="list-style-type: none"> - To elaborate an up-to-date conceptual framework of ICT statistics and related indicators from data collected using Internet as a data source or Big Data repositories. The framework should take into account methodological considerations such as definition of the universe, definition of observation and statistical units, selection of observation units and definition of characteristics or indicators - To exploit the feasibility of extending Eurostat's methods and indicators for collecting ICT official statistics from non-traditional sources. The feasibility of collection of data directly through the user should be explored and the collection of statistical data from enterprises web-sites should be elaborated. - To implement and test the methods and indicators developed for the two different approaches as a proof of concept and to collect information on a possible large scale implementation. - To exploit the feasibility of using Big Data Repositories as official data sources and examine their potential of either supplementing official statistics or even completely replacing official statistics indicators, especially to describe in the form of use cases five data repositories related to statistical domains and assess their feasibility across the technical, organizational, methodological, cost-benefit, legal and socio-political dimension - To elaborate an accreditation procedure this will analyse and assess the required quality aspects of a board range of possible resources in order to qualify statistical data as official statistics. The accreditation procedure should be in-line with the principles and guidelines of Eurostat related to the quality assurance as well as the European Statistics code of practice
------------------------------------------------------------------------------------------------------------------------	-----------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Labour statistics

Other projects

Estimate the job vacancies based on information from job portals, using web scraping tools.

Linked Employer Employee Database

Use of online job postings data for data exploration and validation

Mobility statistics

Title	Organization	Overview
Aggregated Mobile Phone data to identify commuting patterns	UK Office for National Statistics	To source aggregated data from one of the main mobile phone providers for comparison with worker flow estimates from Census. The aggregated data will be based on the movement patterns of the provider's customers. Specific areas of the country will be selected for comparison - and segmentation by age, sex and main mode of transport will be requested. If comparison is successful, then further research will probably be required to assess the potential of using such data for more timely estimates of worker flows.
Other projects		
A set of indices of accessibility and remoteness		

Price statistics

Title	Organization	Overview
Online Price Changes of Means of Production in Circulation Area in Shandong Zhuochuang	National Bureau of Statistics of China	The project is for research purposes at present and it has not applied to the statistical production.
Multi-purpose consumer price statistics, sub-project Scanner Data	Eurostat (European Commission)	Eurostat supports a number of projects aimed at integrating different price statistics and using collected prices for multiple purposes. The projects fall under a broader project 'Multi-purpose consumer price statistics'. In particular one sub-project is relevant in this context. This is the work done on obtaining and implementing the use of scanner data in consumer price statistics in the Member States. Eurostat supports Member States in this work. Eurostat also supports further methodological development on the use of this data source. Scanner data is transaction data generated by retailers in point-of-sales terminals. This data is considered Big Data as it describes all supermarket transactions for a given retailer and period of time, the data is very detailed and delivered frequently (often data is supplied per week). Several Member States use scanner data in their consumer price statistics. These are The Netherlands, Sweden, Norway and Switzerland. Eurostat currently supports a further 17 Member States in obtaining and testing scanner data. Ten of these receive scanner data in some form. Eurostat is responsible, together with the Member States, to produce the HICP (Harmonised Index of Consumer Prices). Given this responsibility, Eurostat is keen to ensure that the

		introduction of scanner data does not jeopardise the comparability of the national HICPs. To this end Eurostat is in the process of drafting guidelines on obtaining and using scanner data.
Use of scanner data for consumer price index	ISTAT	1. Use of scanner data from large retailers for replacing the traditional survey data collection. 2. Main purpose: production
Using scanner data	Romania National Statistics Institute	We intend to use scanner data for improvement of price statistics and other economic statistics indicators. The project is in the conception phase and the results will be used for developing new statistical techniques, monitoring new products, making comparisons between different regions.
Assessing use of scanner data for compiling the Consumer Price Index	Statistics South Africa	Assessing the transactional data of large retail chains with the aim of determining their suitability for transforming into data for the Consumer Price Index. They will also be assessed for suitability for the generation of sales values for statistics on Retail trade sector.
Other projects		
Price collection with scanner data		
Estimation of the customer price index based on webpage information of online retail shops, using web scraping tools.		
Price information (based on scanner data and website information)		
Price statistics, tourism statistics		
Modernisation of consumer price collection and compilation		
CPI Scanner Data Initiative		
CPI Web Scraping Project		

Tourism statistics

Title	Organization	Overview
Persons and Places: Mobility Estimates based on Mobile Phone Data	ISTAT	1. The project focuses on the production of the origin/destination matrix of daily mobility for purpose of work and study at the spatial granularity of municipalities. 2. In particular, it aims to produce statistics on the so-called city users, that is Standing resident, Embedded city users, Daily city users and Free city users. 3. The project has the objective to compare two approaches to mobility profiles estimation, namely: (i) Estimation based on mobile phone data and (ii) Estimation based on administrative archives. 4. Production purposes.
Feasibility study on the use of mobile positioning data for tourism	Eurostat	The aim of the current study was to assess the feasibility of using mobile positioning data for generating statistics on domestic, outbound and inbound tourism flows, and to address the strengths and weaknesses related to access,

statistics

trust, cost, and the technological and methodological challenges inherent in the use of such a new data source. The international consortium that conducted the study concentrated on the various aspects involved in the use of mobile positioning data in terms of tourism statistics and other domains: an overview of the situation involving the use of mobile data; accessibility to the data from the legal, technological, financial and business aspects, including possible cost and burden implications; methodological principles of statistical data collection and compilation, including evaluation by using different quality aspects and comparing the results against existing traditional methods; opportunities offered by, as well as limitations inherent in, the use of the data source. The outcomes can to a large extent be generalised beyond tourism statistics (i.e. applicable for other areas of statistics) and beyond the source of mobile positioning data (i.e. applicable for other 'Big Data' sources) – in particular the comprehensive discussion on feasibility of access.

Other projects

Tourism statistics produced from camera data to estimate the incoming and/or outgoing traffic at the borders

Daytime population and tourism (based on mobile phone location data)

Transportation statistics

Other projects

Traffic indices (based on road sensor data)

True Origin/Destination data in air transport statistics.