UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

ECE/CES/BUR/2013/OCT/2
2 October 2013

CONFERENCE OF EUROPEAN STATISTICIANS

Second Meeting of the 2013/2014 Bureau
Geneva (Switzerland), 21-22 October 2013

For discussion and recommendations

Item 2(a) of the Provisional Agenda

IN-DEPTH REVIEW OF BIG DATA

Prepared by the temporary Task Team on Big Data and the UNECE Secretariat

The current note provides the basis for the in-depth review of "Big data" by summarising the international statistical activities, identifying issues and challenges, and making recommendations how the international statistical community could tackle the issues. The Bureau is invited to discuss the issues raised in the in-depth review, and consider the recommendations made in section VIII of this note and the attached project proposal on the role of Big Data in the modernisation of statistical production.

I. EXECUTIVE SUMMARY

1. This in-depth review considers the response of the official statistics community to the "Big Data" phenomenon. It defines Big Data and summarizes the activities in this area of various national and international statistical organizations. It outlines the main issues and challenges identified so far, and concludes that it would be more efficient for the international statistical community to tackle these issues in a collaborative way, rather than each organization seeking its own solution. The review concludes with the following three recommendations:

(a) Specify the key priority areas relating to Big Data to be tackled as a collaborative activity by the international statistical community. The project proposal in the Annex provides the thinking so far in this area at the expert level. The views of the CES Bureau on whether this proposal covers the right activities would be very welcome;

(b) As knowledge and experience of using Big Data are gained, a mechanism for sharing that information is needed. The inventory of Big Data activities started by the informal Task Team should be consolidated and expanded as a resource for the whole statistical community;

(c) The two activities above should be overseen by the High-Level Group on the Modernisation of Statistical Production and Services, to ensure a sufficiently strategic focus.

II. INTRODUCTION

2. The Bureau of the Conference of European Statisticians (CES) regularly reviews selected statistical areas in depth. The aim of the reviews is to improve coordination of

statistical activities in the UNECE region, identify gaps or duplication of work, and address emerging issues. The review focuses on strategic issues and highlights concerns of statistical offices of both a conceptual and a coordinating nature. The current paper provides the basis for the review by summarising the international statistical activities in the area of Big Data, identifying issues and problems, and making recommendations on possible follow-up actions.

3. In our modern world more and more data are generated on the web and produced by sensors in the ever growing number of electronic devices surrounding us. The amount of data and the frequency at which they are produced have led to the concept of “Big Data”. The term “Big Data” is used to describe data sets of increasing volume, velocity and variety; the three V's. Sources described as ‘Big Data’ are often largely unstructured, meaning that they have no pre-defined data model and/or do not fit well into conventional relational databases. Apart from generating new commercial opportunities in the private sector, Big Data is also potentially very interesting as an input for official statistics; either for use on its own, or in combination with more traditional data sources such as sample surveys and administrative registers. However, harvesting the information from Big Data and incorporating it into a statistical production process is not easy.

4. Big Data has the potential to produce more relevant and timely statistics than traditional sources of official statistics. Official statistics have been based almost exclusively on survey data collections and acquisition of administrative data from government programmes. But this is not the case with Big Data where most data are readily available or with private companies. As a result, the private sector may take advantage of the Big Data era and produce more and more statistics that attempt to beat official statistics on timeliness and relevance. It is unlikely that statistical organizations will lose the "official statistics" trademark but they could slowly lose their reputation and relevance unless they get on board. One big advantage that statistical organizations have is the existence of infrastructures to address the accuracy, consistency and interpretability of the statistics produced. By incorporating relevant Big Data sources into their official statistics process, statistical organizations are best positioned to measure their accuracy, ensure the consistency of the whole systems of official statistics and provide interpretation while constantly working on relevance and timeliness. The role and importance of official statistics will thus be protected.

5. However, the topic of Big Data is still rather new for many statistical organizations, and there is uncertainty about what it really means for official statistics, and how best to react.

III. SCOPE/DEFINITION OF THE STATISTICAL AREA COVERED

6. The topic of Big Data cuts across all statistical activities, and could be relevant for all statistical domains. In terms of the Classification of Statistical Activities (Rev. 1, October 2009[1]), Big Data probably fits best in activity 4.3 (Data Sources), but it does not readily fit into any of the more specific subject areas under this heading, unless a very broad definition of “administrative sources” is used, such that Big Data could be included in subject area 4.3.5 (Other administrative sources).

[1] http://www1.unece.org/stat/platform/display/disaarchive

7. This review focuses on Big Data as a source, rather than an output of statistical organizations, as statistical outputs do not (yet) meet the criteria of volume, velocity and variety to be truly considered as Big Data.

IV. OVERVIEW OF INTERNATIONAL STATISTICAL ACTIVITIES IN THE AREA

A. UNECE

8. The UNECE, together with Rosstat, the Russian Federal State Statistics Service, organized a High-level Seminar on Modernization of Statistical Production and Services, in St. Petersburg in October 2012[2]. Following some discussion of Big Data, one of the conclusions of the Seminar was:

"Big Data is an increasing challenge. The official statistical community needs to better understand the issues, and develop new methods, tools and ideas to make effective use of Big Data sources. This includes closer integration with geographical data and standards."

9. As a follow up activity, it was proposed that the High-Level Group for the Modernization of Statistical Production and Services (HLG) should provide "a document explaining the issues surrounding the use of Big Data in the official statistics community". The HLG convened a group of leading international experts, facilitated by the UNECE, to prepare a paper to address this requirement. The resulting paper, "What does Big Data mean for official statistics?", was published by the HLG in March 2013[3], and presented to the Conference of European Statisticians three months later at their 2013 Plenary Session.

10. This paper provoked further discussion on the topic of Big Data in official statistics, including at the joint UNECE / Eurostat / OECD / UN-ESCAP meeting on the Management of Statistical Information Systems (Paris and Bangkok, April 2013)[4]. This meeting decided that Big Data is a key issue for official statistics, and noted the following key points:

- It is an ideal time to start collaborating on Big Data, as organizations typically did not have systems in place yet, and could develop them collaboratively;
- Each organization faces common issues in relation to using Big Data, so it could be more efficient to work together to find common solutions. This should be a priority for the HLG;
- Statistical organizations have traditionally focused on producing consistent time series, but increasingly there is also a need for short-lived measures that address a phenomenon in a country when it happens;
- Experience gained in the use of administrative sources may be helpful for Big Data;
- It is important to take a multidisciplinary approach to Big Data, currently different groups are all looking at this issue from their own perspectives;
- Agreeing a common classification of the different types of Big Data should be an early priority;
- A concrete project to produce specific statistics from Big Data, and to find real solutions would be useful;

[2] http://www.unece.org/stats/documents/2012.10.hls.html
[3] http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170614
[4] http://www1.unece.org/stat/platform/display/msis/MSIS+2013

- A virtual task team should be set up to define the issues and formulate a clear project proposal, which would be passed to the HLG.

11. In May 2013, a temporary task team[5] was set up to identify the key issues with using Big Data for official statistics, determine priority actions and formulate a project proposal. The team worked virtually, by Wiki and web conferencing, during May and June, and produced the following outputs:

- A draft project proposal consisting of three major work strands: an exploration of strategic and methodological issues; analysis of Big Data sources and international replication of outputs using a shared computing environment; and training and dissemination activities (see Annex);
- A draft classification of types of Big Data[6];
- A specification for an inventory of Big Data sources and projects, based on the above classification. The inventory has subsequently been launched, and populated with information about projects in several countries[7], and will be developed further in the context of UNECE work on the modernisation of statistics.

12. Following the presentation of the annual report of the HLG at the 2013 Plenary Session of the CES, delegates were asked to identify priorities for HLG projects for 2014. Most speakers recognised the need for international collaboration activities in the area of facilitating the use of Big Data for official statistics.

13. The HLG will discuss and decide the key priorities for 2014 with representatives of expert groups at the annual Workshop on the Modernisation of Statistical Production and Services, to be held in Geneva on 25-27 November 2013.

B. Eurostat

14. Eurostat is investigating the potential use of Big Data for official statistics in areas such as price statistics (using Internet price data) and information and communication technology (ICT) usage statistics.

15. Eurostat has contributed actively to the work of the Task Team facilitated by the UNECE, and the development of the resulting project proposal. Eurostat staff have also prepared several papers on this topic for international conferences.

16. One session of the annual DGINS (Director Generals of national statistical organizations) meeting, held in The Hague, Netherlands, in September 2013[8], was devoted to the topic of Big Data. This included presentations from national statistical organizations and private companies. It resulted in the Scheveningen Memorandum on Big Data and Official Statistics, which encourages members of the European Statistical System to develop a Big Data strategy, share experiences, and collaborate at the level of the European Statistical System and beyond. An action plan and roadmap should be adopted by mid-2014, and integrated into the Eurostat work programme.

[5] http://www1.unece.org/stat/platform/display/msis/Members+of+the+task+team
[6] http://www1.unece.org/stat/platform/display/msis/Classification+of+Types+of+Big+Data
[7] http://www1.unece.org/stat/platform/display/msis/Big+Data+Inventory
[8] http://www.cbs-events.nl/dgins2013/

17. Eurostat is planning to hold a workshop or similar event on the topic of Big Data, in cooperation with the UNECE, in spring 2014.

C. OECD

18. The OECD is currently investigating the use of Big Data in the areas of innovation indicators, quality of Internet connections, and well-being / better life indicators.

19. OECD has contributed actively to the work of the Task Team facilitated by the UNECE, and the development of the resulting project proposal. OECD staff have also prepared papers on this topic for international conferences. The OECD has also released a policy paper "Exploring data-driven innovation as a new source of growth: Mapping the Policy Issues Raised by Big Data"[9]

D. United Nations Statistical Division

20. The United Nations Statistical Division organized a one-day side-event to the 2013 meeting of the Statistical Commission, with the title "Big Data for Policy, Development and Official Statistics"[10]. This included presentations from national statistical agencies, international organizations and private companies.

21. The provisional agenda for the 2014 meeting of the Statistical Commission includes an item on "Big Data and modernization of statistical systems"[11].

E. World Bank

22. The World Bank has organized several events on the topic of Big Data, including an event "Turning Big Data into Big Impact"[12] in October 2012 and a live webcast "What happens when Big Data meets official statistics?"[13] in December 2012.

F. Other

23. There are many other events and discussions on the topic of Big Data outside the area of official statistics. The numbers of Big Data consultants, and software tools specifically designed for handling Big Data are growing rapidly. It is clear that in developing a response to the emergence of Big Data, the international official statistics community should actively follow external developments and determine how they might apply to our activities.

V. COUNTRY PRACTICES

24. The Task Team on Big Data identified a wide range of activities relating to the use of Big Data in participating countries. Most of these activities are at the planning or

[9] http://www.oecd-ilibrary.org/science-and-technology/exploring-data-driven-innovation-as-a-new-source-of-growth_5k47zw3fcp43-en
[10] http://unstats.un.org/unsd/statcom/statcom_2013/seminars/Big_Data/default.html
[11] Chapter 1B of the report of the 2013 meeting: http://unstats.un.org/unsd/statcom/doc13/2013-Report-E.pdf
[12] http://www.linkedin.com/groups/World-Bank-Event-Turning-Big-137043.S.177763767
[13] http://live.worldbank.org/what-happens-when-big-data-meets-official-statistics-live-webcast

experimental stage, aiming to determine the feasibility of using Big Data sources for statistical production. However, a few countries are starting to take the next step, and move towards regular data production using Big Data.

25. One issue identified was the lack of a mechanism for sharing information on current and planned activities. This resulted in the proposal to develop an inventory of Big Data projects and resources.

VI. IMPACT OF CRISES ON THE STATISTICAL AREA

26. The financial crisis starting in 2009 has strongly encouraged statistical organizations to look for ways to increase efficiency and cut data costs. Traditionally data collection has been one of the most cost-intensive parts of the statistical production process, so the interest in alternative data sources, including Big Data, is growing.

VII. ISSUES AND CHALLENGES

27. The following issues and challenges were identified in the HLG paper “What does Big Data mean for official statistics?”

A. Legislative

28. Legislation in some countries may provide the right to access data from both government and non-government sources while in other countries, legislation may provide the right to access data from public authorities only. This can result in limitations for accessing certain types of Big Data.

29. Even if legislation has provision to access all types of data, the statistical purpose for accessing the data might need to be demonstrated to an extent that may be different from country to country.

B. Privacy

30. Privacy is generally defined as the right of individuals to control or influence what information related to them may be disclosed. Privacy is a pillar of democracy. The problem with Big Data is that the users of services and devices generating the data are most likely unaware that they are doing so, and/or what it can be used for. The data would become even bigger if they are pooled, as would the privacy concerns.

C. Financial

31. There is likely to be a cost to statistical organizations to acquire Big Data, especially from the private sector, particularly if legislation is silent on the financial modalities surrounding acquisition of external data. As a result, statistical organizations have to balance quality (which encompasses relevance, timeliness, accuracy, coherence, accessibility and interpretability) against costs and reduction in response burden. Costs may be significant, but the potential benefits may far outweigh the costs, with Big Data potentially providing information that could increase the efficiency of government programmes (e.g. health systems). Rules around procurement in the government may come into play as well.

D. Management

32. Big Data for official statistics may mean more information coming to statistical organizations that is subject to policies and directives on management and protection of information.

33. Another management challenge relates to human resources. The data science associated with Big Data that is emerging in the private sector does not seem to have connected yet with the official statistics community. Statistical organizations may have to perform in-house and national scans (academic, public and private sector communities) to identify where data scientists are and connect them to the area of official statistics.

E. Methodological

34. Representativeness is a fundamental issue with Big Data. The difficulty in defining the target population, survey population and survey frame jeopardizes the traditional way in which official statisticians think and do statistical inference about the target (and finite) population. With a traditional survey, statisticians identify a target/survey population, build a survey frame to reach this population, draw a sample, collect the data etc. They will build a box and fill it with data in a very structured way. With Big Data, the data come first and the reflex of official statisticians would be to build a box! This raises the question is this the only way to produce a coherent and integrated national system of official statistics? Is it time to think outside of the box?

35. Another issue is both technological and methodological in nature. When more and more data are analysed, traditional statistical methods, developed for the very thorough analysis of small samples, run into trouble. In the most simple case they are just not fast enough. New methods and tools are needed, for example:

(a) Methods to quickly uncover information from massive amounts of data available, such as visualisation methods and data, text and stream mining techniques, that are able to ‘make Big Data small’. Increasing computer power is a way to assist with this step at first;

(b) Methods capable of integrating the information uncovered in the statistical process, such as linking at massive scale, data integration, and statistical methods specifically suited for large datasets. Methods need to be developed that rapidly produce reliable results when applied to very large datasets.

36. The use of Big data for official statistics triggers a need for new techniques. Methodological issues that these techniques need to address are:

(a) Measures of quality of outputs produced from hard-to-manage external data supply. The dependence on external sources limits the range of measures that can be reported when compared with outputs from targeted information-gathering techniques;

(b) Limited application and value of externally-sourced data;

(c) Difficulty of integrating information from different sources to produce high-value products;

(d) Difficulty of identifying a value proposition in the absence of the closed loop feedback seen in commercial organizations.

F. Technological

37. New tools are needed to connect applications for data capturing and data processing directly with data sources. Collecting data in real time or near real time can maximize the potential of data, opening new opportunities for using data from high-velocity sources, such as:

(a) Commercial data (credit card transactions, on line transactions, sales, etc.);

(b) Tracking devices (cellular phones, global positioning systems, surveillance cameras, 'apps') and physical sensors (traffic, meteorological, pollution, energy, etc.);

(c) Social media (Twitter, Facebook, etc.) and search engines (online searches, online page views);

(d) Community data (Citizen Reporting or Crowd-sourced data).

38. In the era of Big Data this change of paradigm for data collection presents the possibility to collect and integrate many types of data from many different sources. Combining traditional data sources, such as surveys and administrative data, with Big Data could provide new challenges and opportunities.

VIII. CONCLUSIONS AND RECOMMENDATIONS

39. It is clear that the official statistical community is just starting to explore the potential issues and benefits of Big Data. If each organization does this on its own, this will lead to inefficiency within the statistical system at the global level. The main recommendations of this in-depth review are therefore:

(a) Specify the key priority areas relating to Big Data to be tackled as a collaborative activity by the international statistical community. The project proposal in the Annex provides the thinking so far in this area at the expert level. The views of the CES Bureau on whether this proposal covers the right activities would be very welcome;

(b) As knowledge and experience of using Big Data are gained, a mechanism for sharing that information is needed. The inventory of Big Data activities started by the informal Task Team should be consolidated and expanded as a resource for the whole statistical community;

(c) The two activities above should be overseen by the High-Level Group on the Modernisation of Statistical Production and Services to ensure a sufficiently strategic focus.

40. **The Bureau is invited to discuss the issues raised in this in-depth review and comment on the recommendations above, including the attached project proposal on the role of Big Data in the modernisation of statistical production.**

Annex

Draft Project Proposal on the Role of Big Data in the Modernisation of Statistical Production

I. Background

1. The importance of the relationship of Big Data to the official statistics industry has been raised in a number of arenas during recent years. At a High-Level Seminar on Streamlining Statistical Production and Services, held in St Petersburg, 3-5 October 2012, participants called for a strategically-focused document aimed at heads and senior managers of statistical organizations, outlining the issues, challenges and opportunities that Big Data poses for official statistics. The resulting paper[14] discussed definitions and sources, and identified challenges in the areas of legislation, privacy, financial aspects, management, methodology and technology. It suggested that there are a great many opportunities for the use of Big Data, broadly dividing these opportunities into three categories: combining Big data with official statistics; replacing official statistics by Big data; and filling new data gaps.

2. Subsequently the April 2013 meeting of the UNECE Expert Group on the Management of Statistical Information Systems (MSIS) once again identified Big Data as a key challenge for official statistics, and called for the High-Level Group for the Modernisation of Statistical Production and Services (HLG) to focus on the topic in its plans for future work. A temporary task team composed of representatives of 13 national and international statistics organizations was convened to formulate the present project proposal.

3. This project is important for the HLG's broad programme of modernisation of statistical production. As a component of the modernisation programme, it will contribute to the goals of international harmonisation and collaborative approaches to new challenges, improved efficiency of statistical production, and the modification of products and production methods to meet changing user needs. The HLG's strategy document[15] states that "products and services must become easier to produce, less resource-intensive, and less burdensome on data suppliers" and that "new and existing products and services should make use of the vast amounts of data becoming available, to provide better measurements of new aspects of society". The project is aligned with these aspirations since it focuses on new sources, new methods, new outputs, and ways to tackle the issues surrounding these.

4. This project outline includes the objectives, scope and content of this project, as well as some practical project management issues.

II. Project objectives

5. The project has three main objectives:

[14] *What does Big Data mean for Official Statistics?* available at http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170622
[15] *Strategy to Implement the Vision of the HLG* available at http://www1.unece.org/stat/platform/display/hlgbas/HLG+Strategy

- To identify the main possibilities offered by Big Data and provide guidance for statistical organizations, and to develop a coordinated response to the main strategic and methodological issues that Big Data poses for the official statistics industry
- To demonstrate the feasibility of efficient production of both novel products and 'mainstream' official statistics using Big Data sources, and the possibility to replicate these approaches across different national contexts
- To facilitate the sharing across organizations of knowledge, expertise, tools and methods for the production of statistics using Big Data sources.

III. Scope

6. This project concerns the role of Big Data in the modernisation of official statistical production. It will tackle strategic and practical issues that are multi-national in nature, rather than those that are specific to individual organizations or national sources. It will not attempt to identify a comprehensive list of all possible sources or uses of Big Data, nor can it hope to ascertain all the issues and challenges, let alone solve them, since these are broad-ranging and constantly evolving. Whilst the project involves a practical component and a consideration of methodological issues, its aim is not to focus on the technical details of analysis of Big Data, unless these are sufficiently cross-cutting to be of concern internationally.

7. By including representatives of many national and international statistical organizations in the task team that formulated this project proposal, and by continuing to consult with these and other partners throughout, the project aims to be complementary to other initiatives and to avoid duplication of efforts. It also aims to be as relevant as possible to organization-specific needs and concerns. The project itself will endeavour to include inputs from academia and the private sector in addition to the official statistics community, in order to maximise learning across fields and avoid 'reinventing the wheel'.

IV. Contents

8. This project comprises the four work packages outlined below. As a precursor to the project, the following activities have been undertaken by the temporary task team and the UNECE secretariat:

- Formulation of a classification scheme for Big Data sources and identification of the attributes of these sources that are relevant to their use in the production of official statistics
- Development of a repository with examples of sources being used, products being created and other activities being undertaken by statistical organizations, according to the classification and attributes identified above. This repository can be viewed as a repository of case studies for organizations intending to use similar sources or undertake similar projects
- Initial specification of the 'sandbox' environment described under work package 2 below.

A. Work Package 1: Issues and Methodology

9. This work package involves an analysis of the major strategic questions posed by the emergence of Big Data. It will require, first of all, more concrete definitions of the various

terms. The work package will require very broad inputs from across the statistical community and hence will begin with gathering input through electronic consultation and virtual meetings.

10. The work package will expand on, and seek to address the major challenges listed in the HLG paper '*What does Big Data mean for Official Statistics?':*

- Legislative: how to access and use data?
- Privacy: how to manage public trust and acceptance of data re-use and linking to other sources?
- Financial: what are the potential costs and benefits of using Big Data?
- Management: what policies are necessary for the effective management and protection of the data?
- Methodological: how to manage data quality? Are current statistical methods and models suitable for Big Data?
- Technological: what are the issues related to information technology?

11. It will also address a variety of issues and questions identified by the task team, including (*but not limited to*) the following:

- How can we assess the suitability of Big Data sources for the production of official statistics?
- How can we effectively capitalise upon the promise of massively increased timeliness offered by many Big Data sources?
- Can we identify best practices or guidelines for the major methodological issues relating to Big Data? E.g.:
    - Methods for reducing data volume
    - Methods for noise reduction
    - Methods for ensuring confidentiality and avoiding inadvertent disclosure
    - Methods for obtaining information on statistical concepts (text mining, classification methods, etc.)
    - Methods for determination of population characteristics, e.g. determining the population of users of social media services through analysis of words or phrases that are highly correlated with certain demographic characteristics
    - Assessing the applicability of models
- Should Big Data be treated as homogeneous, or do they require different treatment according to the role they play in the production of official statistics?
    - Experimental uses
    - Complementing existing statistics e.g. benchmarking and validity checking;
    - Supplementing existing sources, permitting the creation of entirely new statistics;
    - Replacing existing sources and methods
- Are there 'quick wins', applicable beyond Big Data, such as data storage, technology, advanced analytics, methods and models which could transform our thinking in relation to the production of official statistics more generally?
- How should statistical organizations react to the novel idea that in a Big Data world there are no 'bad' data (they all tell us something)?
- How can organizations mitigate the risk of a data source ceasing to exist, or changing substantially, when it is outside the control of the organization?

- How can Big Data be combined with survey data? And relatedly, how can the transition from statistical data production based entirely on surveys to production based substantially on Big Data be managed?
- Do we need a research question before exploring a Big Data source, or should we just experiment and innovate to see what is possible?
- What becomes of the time series in a world where data sources and uses may become more transient?
- What is the demand for new types of statistical information, given the new possibilities?
- How should statistical organizations approach the need to 'educate' (or re-educate) staff and users?
- How will institutional structures need to change in order to support the use of Big Data and ensure its quality and the quality of resulting outputs?

12. The output from this work package will take the form of recommendations, good practices and guidelines, developed through broad consultation of experts throughout the official statistics community, and coordinated by expert task teams. The material will be collated in an electronic environment such as a wiki. Such an environment will allow the guidelines to function as a 'living document', permitting timely updating as circumstances change. The task of maintaining the content after its initial formulation will be overseen by the HLG's Modernisation Committee on Products and Sources.

B. Work Package 2: Shared computing environment ('sandbox') and practical application

13. This work package will form the practical element of the project, aimed at proving concepts in two related strands:

(a) Statistics:
- The possibility of producing valid and reliable statistics from novel sources, including the ability to produce statistics which correspond in a predictable and systematic way with existing 'mainstream' products, such as price statistics
- The cross-country applicability of new analytical techniques and sources, such as the analysis of data from social networking websites. This will be done by attempting to reproduce the results of a national project in other countries

(b) Tools:
- The efficiency of various software tools for large-scale processing and analysis
- The applicability of the Common Statistical Production Architecture (CSPA – under development) to the production of statistics using Big Data sources.

14. A web-accessible environment for the storage and analysis of large-scale datasets will be created and used as a 'sandbox' for collaboration across participating institutions. One or more free or low-cost, internationally-relevant datasets will be obtained and installed in this environment, with the goal of exploring the tools and methods needed for statistical production and the feasibility of producing Big Data-derived statistics and replicating outputs across countries. Simple configurations with tools and data will be specified so that partners will be able to test them within their own technical environments. The details of

the sandbox will be specified in a separate annex to this proposal, following a study of alternative scenarios and a consideration of criteria by a task team of experts.

C. Work Package 3: Training and dissemination

15. This work package will ensure that the conclusions reached in the two preceding work packages are shared broadly throughout the statistical world and beyond. This will be done through a variety of means, including:

(a) Establishing and maintaining a central location and online infrastructure for documentation and information-sharing on the UNECE wikis, including detailed documentation arising from work packages 2 and 3;

(b) Preparation of electronic demonstrations of tools and results, for example in the form of Webex presentations and Youtube videos which can be disseminated widely. Identification of existing electronic resources and online training materials is also included in this strand;

(c) A workshop in which the results of work package 2 will be presented to members of the various of expert groups involved in the HLG's modernisation programme. This would be held back-to-back with the annual workshop on modernisation of statistics at which all these expert groups are represented (likely to be November 2014).

D. Work Package 4: Project management and coordination

16. This work package comprises the necessary project management activities to ensure the successful delivery of the other three work packages.

V. Definition of success

17. Overall, this project will be successful if it results in an improved understanding within the international statistical community of the opportunities and issues associated with using Big Data for the production of official statistics. Success criteria for the individual work packages are:

- Work package 1: a consistent international view of Big Data opportunities, challenges and solutions, documented and released through a public web site;
- Work package 2: recommendations on appropriate tools, methods and environments for processing and analysing different types of Big Data, and a report on the feasibility of establishing a shared approach for using Big Data sources that are multi-national or for which similar sources are available in different countries;
- Work package 3: exchange of knowledge and ideas between interested organizations and a set of standard training materials;
- Work package 4: the project is completed on schedule, and delivers results that are of value to the international statistical community.

VI. Expected costs

18. The following table shows an estimate of the minimum resources and other costs needed to deliver the different work packages. Each organization involved in the project

will be expected to cover the costs of their participation (including wages and any travel expenses for participants).

Work Package	**Estimated resources**	**Source of resources**	**Other costs** (in US Dollars)
1: Issues and methodology	8 person months	Volunteer NSOs plus UNECE Secretariat	Possible travel costs if a workshop or sprint session is needed
2: Shared computing environment & practical applications	12 person months	Volunteer NSOs plus UNECE Secretariat	Costs associated with renting a shared space and acquiring data and tools (max $10,000?) Possible travel costs if a workshop or sprint session is needed
3: Training & dissemination	4 person months	Volunteer NSOs plus UNECE Secretariat	Up to $1,000 for costs associated with preparing and disseminating training materials and running workshop for expert groups
4: Project management	6 person months	A project manager working in the UNECE Secretariat. Input from Executive Board and HLG members (in their role as project sponsors)	Up to $500 for telecommunications and other incidentals Travel costs for project events
Total	**30 person months**	**UNECE Secretariat (9 person months)** **NSO / International organization staff (21 person months)**	**Up to $11,500 total costs as described above, plus:** • **possible consultancy costs** • **travel costs of experts**

VII. Timetable

19. The project will aim to complete the activities described by the end of 2014. There are, however, various unknowns which may affect the timetable:

- The availability of resources from national and international statistical organizations to support this project – if the necessary resources are not available, either the timetable will need to be extended, or the outputs will need to be re-defined (in terms of quality or quantity or both);
- The availability of project management and support resources in the UNECE Secretariat – to meet the resource requirements of this project will require the continuation of the current extra-budgetary post in the UNECE secretariat, through additional donor funding. As above, if this is not forthcoming, either the timetable will need to be extended, or the outputs will need to be re-defined.

20. All four work packages will run throughout the year, though substantial work should be completed by mid-November so that outcomes can be reported and demonstrated at the HLG Workshop.

VIII. Project governance

21. The project sponsor is the HLG. This is the group that has ultimate responsibility for signing off the project deliverables. In practice, this responsibility will be delegated to the Executive Board.

22. A project manager will have day-to-day responsibility for the running of the project, providing regular updates and signalling any issues to the Executive Board as necessary.

* * * * *