

**Статистическая комиссия****Сорок пятая сессия**

4–7 марта 2014 года

Пункт 3(j) предварительной повестки дня\*

**Вопросы для обсуждения и принятия  
решения: большие данные и модернизация  
статистических систем****Большие данные и модернизация статистических систем****Доклад Генерального секретаря***Резюме*

В настоящем докладе, подготовленном в соответствии с решением 2013/235 Экономического и Социального Совета, дается оценка текущей практики использования больших данных для подготовки официальной статистики. В докладе содержится обзор последних мероприятий в рамках сообщества специалистов, занимающихся официальной статистикой, и представлены результаты глобальной оценки использования больших данных для подготовки официальной статистики, которая была проведена Статистическим отделом в 2013 году.

В докладе предлагается план дальнейшей работы, предусматривающий создание рабочей группы на глобальном уровне, в развитие существующих региональных инициатив по обмену методическими разработками, передовой практикой решения стратегических вопросов и возможностями обучения. Рабочая группа будет также содействовать налаживанию международного партнерства в целях использования больших данных для передачи технологии развивающимся странам и для поддержки разработки программы развития на период после 2015 года. Статистической комиссии предлагается высказать свои мнения по предложению, изложенному в заключительном разделе доклада.

\* E/CN.3/2014/1.



## I. Введение

1. Темой пятничного семинара по наметившимся вопросам в 2013 году была «Роль больших данных в формировании политики, развитии и подготовке официальной статистики»<sup>1</sup>. Выступившие на семинаре представители частного сектора наряду с руководителями статистических органов высказали свои мнения по вопросу об актуальном значении больших данных для формирования политики и о роли, которую надлежит играть национальным статистическим системам в плане использования новых источников информации. Встреча проходила под председательством руководителя Австралийского статистического бюро. На данном мероприятии был сделан вывод о том, что большие данные представляют собой источник информации, который не может игнорироваться официальной статистикой, и о том, что специалисты, занимающиеся официальной статистикой, должны сорганизоваться и принять срочные меры к тому, чтобы эффективным образом использовать имеющиеся возможности и ограничить возникающие проблемы.

2. Сразу же по окончании этого параллельного мероприятия представитель Австралии обратился на сорок четвертой сессии Статистической комиссии с просьбой провести оценку использования больших данных для подготовки официальной статистики и представить Комиссии доклад по этому вопросу на ее сорок пятой сессии. Настоящий доклад представляется в ответ на эту просьбу.

3. В результате всемирного использования электронных устройств и повсеместного генерирования и наличия цифровой информации кардинально изменился характер данных, которые генерируются теперь постоянно и в огромных количествах и которые мы называем большими данными. Эти данные обладают важными отличительными свойствами, которые отличают их от данных, получаемых из традиционных источников. Данные, получаемые из этих инновационных источников, обладают исключительно широким диапазоном распределения, неорганизованной структурой, огромным объемом и часто поступают в масштабе реального времени. Большие данные являются источником данных, которые можно описать следующим образом: «данные огромного объема, скорости прироста и многообразия, требующие затратоэффективных новаторских видов обработки и предназначенные для углубленного аналитического исследования и принятия решений».

4. В эпоху снижающейся доли респондентов, отвечающих на обследования домашних хозяйств и предприятий, большие данные могут обеспечивать директивные органы фактологической информацией в масштабе реального времени в таких областях, как цены, занятость, объем производства, экономическое развитие и динамика населения<sup>2</sup>. Большие данные могут в потенциале обеспечивать подготовку более актуальных и своевременных статистических данных по сравнению с такими традиционными источниками официальной

---

<sup>1</sup> Данный семинар, проведенный в качестве параллельного мероприятия в ходе сорок четвертой сессии Статистической комиссии, состоялся в пятницу, 22 февраля 2013 года. Подробнее см. [http://unstats.un.org/unsd/statcom/statcom\\_2013/seminars/Big\\_Data/default.html](http://unstats.un.org/unsd/statcom/statcom_2013/seminars/Big_Data/default.html).

<sup>2</sup> Organization for Economic Cooperation and Development (OECD), “Exploring data-driven innovation as a new source of growth: mapping the policy issues raised by ‘big data’” (DSTI/ICCP(2012)9/FINAL), 2013.

статистики, как обследования и административные источники данных. Источники больших данных сосредоточены в основном в частном секторе, и в большинстве стран еще не введены в действие законы об использовании больших данных для целей официальной статистики. Поэтому необходимо проведение дальнейшего диалога на широкой основе для рассмотрения не только вопросов, связанных с правовой базой, конфиденциальностью и отношением общественности, но и вопросов своевременного, надежного и достоверного использования больших данных. Включив источники больших данных в свой процесс подготовки официальной статистики, национальные, региональные и международные статистические организации могли бы укрепить свои позиции в плане более своевременного и экономически эффективного получения данных официальной статистики по секторам экономики, социальной сферы и окружающей среды с меньшими затратами ресурсов.

5. В свете технического прогресса и бурного роста количества и разнообразия информации, поступающей в масштабе реального времени, Генеральный секретарь Организации Объединенных Наций приступил в 2009 году к реализации инициативы «Глобальный пульс» в целях задействования научно-технических новшеств в области цифровой информации и быстрого сбора и анализа данных для того, чтобы принимающие решения руководители могли в масштабе реального времени получать полное представление о том, как кризисы влияют на положение уязвимых слоев населения. Подчеркнув, что большие данные не являются панацеей от всех бед, авторы официального документа, подготовленного в рамках инициативы «Глобальный пульс», вместе с тем заявляют, что «большие данные открывают перед нами историческую возможность для наращивания нашей общей способности поддерживать и защищать людские сообщества путем уяснения информации, постоянно генерируемой такими сообществами в цифровой форме»<sup>3</sup>.

6. В рекомендациях в адрес Генерального секретаря, вынесенных созданной им Группой видных деятелей высокого уровня по программе развития на период после 2015 года в ее докладе, озаглавленном «Новое глобальное партнерство: искоренение бедности и преобразование экономик посредством устойчивого развития», содержится призыв к осуществлению информационной революции. В докладе признается, что большие данные являются важным элементом информационной революции, который мог бы в потенциале способствовать повышению качества статистики в некоторых ее аспектах, например в плане своевременности и полноты данных, без ущерба для значимости, беспристрастности и методологической состоятельности статистики. В докладе Группы высокого уровня говорится:

«Более достоверная информация и статистические данные помогут органам государственной власти отслеживать результаты проделанной работы и гарантируют принятие ими решений на достоверной основе; кроме того, с их помощью можно будет наладить более эффективную систему отчетности. Это касается не только органов государственной власти, но и международных организаций, организаций гражданского общества и представителей частного сектора. Настоящая информационная революция позволит с помощью существующих и новых источников данных полностью интегрировать статистику в процесс принятия решений, пропаган-

<sup>3</sup> Global Pulse, “Big data for development: challenges and opportunities”, May 2012.

дировать открытый доступ и использование информации и расширить поддержку систем сбора статистики» (глава 4, стр. 28).

7. В более широком плане имеются веские основания утверждать, что большие данные могут в потенциале изменить многие аспекты жизни общества. В своем недавнем докладе Комиссия по большим данным при Фонде «ТехАмерика» заявила, что «большие данные могут в потенциале трансформировать сферу государственного управления и само общество»<sup>4</sup>. В производимых сегодня данных огромного объема, многообразия и скорости прироста скрыта новая информация, факты, взаимосвязи, показатели и индикаторы, которые в прошлом либо не могли быть обнаружены практическим путем, либо просто не существовали<sup>5</sup>. Эта новая информация, эффективно собранная, систематизированная и проанализированная, способна радикальным образом повысить эффективность работы органов государственного управления<sup>6</sup>.

8. В разделе II настоящего доклада дается общий обзор деятельности, связанной с большими данными и модернизацией статистических систем. В разделе III обобщены источники больших данных и проблемы, возникающие в связи с их использованием. В разделе IV представлены результаты глобальной оценки использования больших данных для подготовки официальной статистики, которая была проведена Статистическим отделом в период с июля по октябрь 2013 года. В разделе V изложен ряд тематических исследований по вопросу об использовании больших данных. В заключительной части доклада содержится предлагаемый план дальнейшей работы, который согласуется с инициативами организаций-партнеров и делает основной упор на вопросах, касающихся использования больших данных для подготовки официальной статистики развивающимися странами и разработки программы развития на период после 2015 года.

## **II. Общий обзор деятельности, связанной с большими данными, которая осуществлялась в последнее время**

9. Правительства все большего числа стран признают важность больших данных и создают сообщества специалистов-практиков и рабочие группы для изучения вопроса об их использовании и получении от них потенциальной отдачи. Статистическое сообщество постепенно осознает, что назревает качественный скачок. Национальный институт статистики Италии признает, что «для того, чтобы завладеть (потенциально несметными) богатствами, которые сверкают сквозь горы больших данных, национальным статистическим институтам придется покорить эти горы»<sup>7</sup>. Статистическое бюро Нидерландов допускает,

<sup>4</sup> Первое предложение краткого резюме к докладу Комиссии по большим данным при Фонде «ТехАмерика», озаглавленному [“Demystifying big data: a practical guide to transforming the business of government”](#) (Washington, D.C., October 2012).

<sup>5</sup> Ibid., второе предложение.

<sup>6</sup> Ibid., третье предложение.

<sup>7</sup> Monica Scannapieco, Antonino Virgillito and Diego Zardetto, “Placing big data in official statistics: a big challenge?”, 21 December 2012.

что сообщество специалистов, занимающихся официальной статистикой, делает только первые шаги в исследовании возможностей, которые таят в себе большие данные<sup>8</sup>.

10. Европейская экономическая комиссия (ЕЭК), Статистическое бюро Европейской комиссии (Евростат), Организация экономического сотрудничества и развития (ОЭСР) и Экономическая и социальная комиссия для Азии и Тихого океана (ЭСКАТО) организовали Совещание по вопросам управления статистическими информационными системами, которое проходило одновременно в Париже и Бангкоке 23–25 апреля 2013 года<sup>9</sup>. Главный вывод, сделанный на этом совещании, заключался в том, что статистические организации должны объединить усилия для решения общих вопросов, связанных с использованием больших данных, и вести поиск таких общих решений, как подготовка статистических данных с коротким сроком службы, применение междисциплинарного подхода к использованию больших данных и согласование общей классификации различных видов больших данных.

11. На пятьдесят девятом Всемирном статистическом конгрессе Международного статистического института (МСИ), состоявшемся в Китае, специальный административный район Гонконг, 25–30 августа 2013 года, большое внимание было уделено тематике больших данных: были проведены одно заседание по документу, подготовленному приглашенным специалистом, по теме больших данных; одно заседание по специальной теме «Потенциал Интернета, больших данных и органических данных для официальной статистики»; и вспомогательное мероприятие по теме «Большие данные и расчет статистических показателей». На заседании по специальной теме Джон Данн из Центрального статистического управления Ирландии представил документ, озаглавленный «Big data coming soon ... to a national statistical institute near you» («Смотрите скоро в близлежащем национальном статистическом институте — большие данные»)<sup>10</sup>, в котором были описаны некоторые источники больших данных, например ведомости оплаты за мобильную телефонную роуминг-связь, потребление электроэнергии или пользование электронными сетями, и представлены возможные пути обработки таких больших объемов своевременно получаемых данных третьими сторонами по методу облачных вычислений с использованием электронно-вычислительной техники. Большие данные будут также важной темой научной программы на шестидесятом Всемирном статистическом конгрессе, который состоится в Рио-де-Жанейро, Бразилия, 26–31 июля 2015 года<sup>11</sup>. Конгрессы МСИ предоставляют возможность для сотрудничества и обмена знаниями между статистиками, работающими в официальных статистических органах, учебных заведениях и деловых компаниях.

12. Евростат изучает вопрос о потенциальном использовании больших данных для подготовки официальной статистики в таких областях, как статистика цен (с использованием данных о ценах из Интернета) и статистика пользования информационно-коммуникационными технологиями (ИКТ). Кроме того, одно из заседаний на ежегодном совещании генеральных директоров европейских

<sup>8</sup> Piet Daas and Mark van der Loo, “Big data (and official statistics)”, рабочий документ, подготовленный для Совещания по вопросам управления статистическими информационными системами (МСИС-2013), Париж и Бангкок, 23–25 апреля 2013 года.

<sup>9</sup> См. <http://www1.unece.org/stat/platform/display/msis/MSIS+2013>.

<sup>10</sup> Имеется по адресу <http://www.statistics.gov.hk/wsc/STS018-P3-S.pdf>.

<sup>11</sup> См. <http://www.isi2015.ibge.gov.br>.

национальных статистических институтов в сентябре 2013 года в Гааге было посвящено теме больших данных<sup>12</sup>. Результатом этого заседания стало принятие Схевенингенского меморандума по большим данным и официальной статистике (утвержденного Комитетом Европейской статистической системы 27 сентября 2013 года), в котором членам Европейской статистической системы рекомендуется разработать стратегию использования больших данных, обмениваться опытом и поддерживать взаимодействие на уровне Европейской статистической системы и за ее пределами. К середине 2014 года должны быть приняты план и порядок действий, которые будут учтены в программе работы Евростат.

13. Большие данные были также одним из предметов обсуждений на Семинаре ЕЭК по вопросам сбора статистических данных, состоявшемся в Женеве 25–27 сентября 2013 года<sup>13</sup>. На этом Семинаре Петер Стрэйс и Пит Дас из Статистического бюро Нидерландов представили рабочий документ (№ 31), озаглавленный “Big data, big impact?” («Большие данные — большое воздействие?»), в котором говорилось, что большие данные окажут, со всей очевидностью, большое воздействие на статистическое сообщество и на официальную статистику. Конкретные аспекты такого воздействия станут проявляться лишь постепенно, однако некоторые особенности уже видны или вполне предсказуемы. Национальные статистические системы будут испытывать более острую конкуренцию со стороны действующих сторон, не входящих в их сообщество. Им придется скорректировать свои методы подготовки статистики и найти новый баланс, с использованием новых возможностей, которые заключают в себе большие данные. Для этого может потребоваться качественный переход от статистической деятельности, ориентированной на проведение обследований, к деятельности, имеющей большую направленность на использование вторичных данных и предполагающей широкое применение методов моделирования. Вступая в эпоху больших данных, статистические управления должны — в порядке общего правила, а не в порядке исключения — иметь в своем штате ученых по информатике, и соответствующее изменение организационной культуры не заставит себя ждать. Заранее признав возможности и проблемы, связанные с использованием больших данных, сообщество специалистов, занимающихся официальной статистикой, будет лучше готово к тому, чтобы быстро переключиться на выполнение своей новой роли — предоставление информации, собираемой из этих новых источников.

14. На своем втором совещании, состоявшемся 21–22 октября 2013 года, Бюро Конференции европейских статистиков 2013/2014 годов, являющееся руководящим органом ЕЭК в области статистики, провело углубленное изучение темы больших данных<sup>14</sup>. По результатам этого обзора были вынесены следующие основные рекомендации: международному статистическому сообществу следует совместными усилиями определить ключевые приоритетные области использования больших данных и взяться за освоение этих областей; и

<sup>12</sup> См. <http://www.cbs-events.nl/dgins2013/>.

<sup>13</sup> См. <http://www.unece.org/stats/documents/2013.09.coll.html>.

<sup>14</sup> См. ECE/CES/BUR/2013/OCT/2.

следует создать механизм для обмена информацией о знаниях и опыте использования больших данных. Бюро одобрило также предлагаемый проект<sup>15</sup>, посвященный большим данным и имеющий следующие цели:

а) выявить основные возможности, которые заключают в себе большие данные, и подготовить методические указания для статистических организаций, а также выработать скоординированные меры по решению основных стратегических и методологических вопросов, которые возникают в секторе официальной статистики в связи с использованием больших данных;

б) продемонстрировать целесообразность эффективной подготовки как новой статистической продукции, так и «традиционной» официальной статистики с использованием источников больших данных и возможность копирования этих подходов в разных национальных условиях;

в) содействовать обмену знаниями, техническим опытом, инструментами и методами между организациями в целях подготовки статистики с использованием источников больших данных.

15. Вопрос о больших данных и модернизации статистических систем был включен также в повестку дня нескольких недавних мероприятий в Азии. Статистический отдел Организации Объединенных Наций и Национальное статистическое бюро Китая организовали, при поддержке ЭСКАТО, международный семинар по вопросам модернизации официальной статистики, который проходил 24–26 октября 2013 года в Тяньцзине, Китай. На этом семинаре, посвященном модернизации статистических организаций, было сделано несколько докладов по вопросу о больших данных. В центре внимания на одиннадцатом Семинаре по вопросам управления для руководителей национальных статистических ведомств стран Азии и Тихого океана<sup>16</sup>, который был организован Статистическим институтом для Азии и Тихого океана (СИАТО) ЭСКАТО 21 и 22 ноября 2013 года в Тибете, Япония, находились инициатива в области информационной революции и потенциальная роль больших данных в программе развития на период после 2015 года.

### **III. Большие данные: источники и проблемы**

16. Неофициальная целевая группа, учрежденная Группой высокого уровня по модернизации статистической продукции и услуг и координируемая ЕЭК, обобщила источники больших данных и проблемы, связанные с использованием больших данных, в своем докладе, озаглавленном “What does big data mean for official statistics” («Каково значение больших данных для официальной статистики») (10 марта 2013 года). Доклад был представлен на шестьдесят первой пленарной сессии Конференции европейских статистиков, состоявшейся 10–12 июня 2013 года. Выводы, сделанные в этом докладе, рассматриваются ниже. Разбивки, предложенные целевой группой ЕЭК в этом докладе, были использованы при подготовке опросника для всемирного обследования, проведенного Статистическим отделом в третьем квартале 2013 года (см. раздел IV).

<sup>15</sup> Данный проект является частью инициативы по модернизации статистической продукции в рамках Группы высокого уровня по модернизации статистической продукции и услуг, координируемой ЕЭК.

<sup>16</sup> См. [http://www.unsiap.or.jp/training/6\\_Leading\\_MQS/6\\_ms11.php](http://www.unsiap.or.jp/training/6_Leading_MQS/6_ms11.php).

17. Сегодня данные автоматически и непрерывно производятся в цифровой форме многими различными способами. Эти различные источники данных представляют потенциальный интерес с точки зрения их использования в официальной статистике для более точного и своевременного измерения определенных социальных, экологических, финансовых или экономических тенденций. В целом, источники возникновения больших данных можно классифицировать следующим образом:

- источники данных, связанные с осуществлением программы, будь то государственной или иной, например электронные медицинские карты, ведомости приема клиентов больничными учреждениями, учетные страховые документы, учетные банковские документы и продовольственные банки;
- коммерческие или операционные источники данных, связанные с совершением операций между двумя сторонами, например операции по кредитным карточкам и онлайн-операции (в том числе совершаемые с помощью мобильных устройств);
- источники данных, связанные с работой сенсорных сетей, например данные с изображений, полученных со спутников, данные с автодорожных датчиков и метеорологические данные с измерительных устройств;
- источники данных, связанные с работой регистрирующих устройств, например регистрация данных из сети мобильной телефонной связи и из Глобальной системы определения координат (GPS);
- источники данных, связанные с поведением пользователей, например данные поиска в Интернете (по тому или иному продукту, услуге или по любому другому виду информации) и данные о просмотрах веб-страниц;
- источники данных, связанные с выражением пользователями своих мнений, например данные из комментариев в социальных сетях.

18. Данные из административных документов являются одним из главных источников информации для подготовки официальной статистики национальными статистическими системами. Такие данные, получаемые от государственно-административных органов, традиционно носят сильно структурированный характер и затем обрабатываются, хранятся, систематизируются и используются статистическими ведомствами. Данные из административных документов в настоящее время не являются источником больших данных, однако они могут стать таковым в случае увеличения скорости прироста и физического объема, например когда статистические ведомства начнут шире пользоваться данными из административных документов благодаря их сбору в масштабе реального времени или на ежедневной или еженедельной основе, а не раз в год или раз в месяц, как это обычно делается.

19. Использование больших данных в официальной статистике представляет множество проблем, которые подразделяются на следующие основные категории:

- юридические, то есть связанные с доступом к данным и их использованием;



- связанные с неприкосновенностью частной жизни, то есть пользование общественным доверием и получение согласия на вторичное использование данных и их увязку с другими источниками;
- финансовые, то есть потенциальные затраты на извлечение данных из источника в сопоставлении с получаемыми выгодами;
- управленческие, например политика и директивы по вопросам управления данными и обеспечения их защиты;
- методологические, то есть качество данных и пригодность статистических методов;
- технологические, то есть вопросы, связанные с информационными технологиями.

#### **IV. Результаты всемирного обследования по вопросу об использовании больших данных**

20. Взяв за основу определения, приведенные в докладе целевой группы ЕЭК, Статистический отдел разработал опросник, посвященный использованию больших данных для подготовки официальной статистики. Опросник состоит из трех основных частей: источники, проблемы и области использования. Вопросы, касающиеся источников и проблем, охватывают все вышеупомянутые аспекты с более подробной проработкой управленческих аспектов, которые были определены как «вопросы управления, например соблюдение новых директив и правил и развитие людских ресурсов, обладающих необходимым набором навыков и знаний». Эта всемирная оценка была проведена с целью получения информации о национальных приоритетах, событиях и опыте в том, что касается текущего или планируемого использования больших данных для подготовки официальной статистики.

21. Опросник был разослан статистическим ведомствам более 200 стран и территорий в период с июля по сентябрь 2013 года: опросник рассылался на английском языке в июле, на испанском языке в августе, и на французском языке в сентябре. Его можно было заполнять либо через Интернет, либо в формате PDF. По состоянию на 2 ноября 2013 года было получено 107 ответов. В полном виде полученные результаты представляются Статистической комиссии в качестве справочного документа. В обобщенном виде полученные результаты изложены ниже.

##### **A. Источники больших данных**

22. Вопрос, касающийся источников больших данных, был сформулирован следующим образом: «Пожалуйста, укажите, какие из нижеперечисленных источников больших данных будут, по всей вероятности, использоваться в течение следующих 12 месяцев Вашим управлением или другими учреждениями, являющимися частью национальной статистической системы Вашей страны». В случае положительного ответа респондентам предлагалось пояснить, какие конкретные источники данных они собираются использовать. Если более 50 процентов стран и территорий сообщили, что они собираются использовать

административные источники в качестве источников больших данных, то по каждому из других пяти источников данных показатель использования составил около 25 процентов. Несколько стран подняли вопрос о том, являются ли объемные административные документы источником больших данных или нет. В более общем плане данные из административных источников служат основой для подготовки многих видов статистической продукции, однако вопрос о том, следует ли рассматривать их в том же контексте, что и большие данные, нуждается в обсуждении.

23. В целом, многие страны положительно ответили на вопрос о том, собираются ли они использовать один или несколько источников больших данных. Однако дальнейший анализ показал, что некоторые из источников, указанные в этих положительных ответах, не совместимы с определением больших данных в том виде, как оно приведено в упомянутом докладе целевой группы ЕЭК. Чтобы придать большую весомость выводам, сделанным по результатам этого обследования по вопросу об использовании больших данных, полученные ответы были тщательнейшим образом проанализированы, и была выделена та группа положительных ответов, которую можно отнести к категории «надлежащей практики». В эту категорию не вошли все те положительные ответы, которые либо были даны без пояснений, либо сопровождались весьма ограниченными или не относящимися к делу пояснениями. В результате, вместо первоначальных 25–50 процентов положительных ответов скорректированные показатели составили почти 20 процентов для административных источников данных, около 10 процентов для операционных источников данных и источников, связанных с работой сенсорных сетей и регистрирующих устройств, и менее 5 процентов для источников, связанных с поведением пользователей и выражением ими своих мнений. Подробные результаты будут помещены на веб-сайте Статистического отдела и будут содержать отдельную колонку по всем странам, которые дали положительные ответы, и отдельную колонку по тем странам, которые применяют «надлежащую практику».

24. Одним из наиболее часто упоминаемых источников административных данных являются данные, поступающие из управлений, занимающихся сбором подоходных налогов и налогов на добавленную стоимость (НДС). Другим источником являются записи актов гражданского состояния и реестры предприятий. К числу упомянутых операционных источников данных относятся данные сканирования товаров в универсамах, данные из сети телефонной связи, операции по кредитным карточкам и данные о сборе платы за проезд по платным дорогам. Что касается сенсорных или спутниковых источников, то несколько стран сообщили о том, что они пользуются данными с изображений, полученных со спутников, при разработке своих обследований. Данные с изображений, полученных со спутников, были названы также источником, используемым при изучении статистики растениеводства. Данные с автодорожных датчиков используются как при подготовке статистики транспорта (индекс дорожного движения в Нидерландах), так и при подготовке статистики окружающей среды (замеры качества воздуха в Соединенном Королевстве Великобритании и Северной Ирландии).

25. В нескольких странах используются источники данных, связанные с работой регистрирующих устройств и системы GPS. Например, данные о местонахождении мобильных телефонов используются для совершенствования статистики туризма в Эстонии, как это показано в тематическом исследовании,

представленном в следующем разделе. В Канаде данные GPS, получаемые с автогрузового, железнодорожного, водного и воздушного транспорта, используются для дополнения статистики транспорта. Хотя данные, связанные с поведением пользователей, используются в небольшом числе стран, некоторые страны планируют начать использование этого источника данных. Например, в Италии данные запросов в Интернете, возможно, будут использоваться: а) в качестве вспомогательной информации в целях сокращения затрат, связанных с проведением обследований; б) для целей прогнозирования; и с) для повышения уровня географической детализации оценочных показателей (особенно показателей рынка труда).

26. Равным образом, в весьма незначительном числе стран используются источники больших данных, связанные с выражением пользователями своих мнений или с работой социальных сетей. Нидерланды указали, что в стране проводится определенная исследовательская работа, основанная на общедоступных сообщениях в социальных сетях (см. тематическое исследование в разделе IV). В Сингапуре проводится исследование по вопросу о том, как можно использовать этот источник больших данных для заблаговременного выявления проблем, связанных с безопасностью продуктов питания.

## **В. Проблемы, связанные с использованием больших данных**

27. Вопрос, относящийся к данному разделу, был сформулирован следующим образом: «Представляет ли [тот или иной вопрос] серьезную проблему для национальной статистической системы в Вашей стране?». Были предложены следующие возможные ответы: «Нет» (не представляет проблемы), «Нет мнения на этот счет» (этот вопрос не обсуждался) или «Да» (представляет проблему). По всем шести категориям проблем большинство стран ответили «Да» (представляет проблему), следующим по частотности был ответ «Нет мнения на этот счет» (то есть эти вопросы еще не были предметом тщательного обсуждения), и небольшое число стран ответили «Нет» (не представляет проблемы). В целом, чаще всего указывались методологические, информационно-технологические и управленческие проблемы, за которыми с небольшим отставанием следовали юридические проблемы и проблемы, связанные с неприкосновенностью частной жизни.

28. Респонденты согласились с тем, что одной из серьезнейших проблем является решение методологических вопросов, и указали на необходимость дополнительной исследовательской работы в этой области. Качество, репрезентативность, изменчивость и размерность данных: все эти вопросы представляют методологические трудности с точки зрения освоения источников больших данных. Информационно-технологические вопросы также были названы серьезной проблемой во многих странах, поскольку существующие системы обработки и анализа данных не могут быть масштабированы до уровня больших данных. Характер больших данных (особенно их физический объем, а также скорость прироста и многообразие) создает серьезные проблемы при разработке необходимых информационно-технических инструментов и структуры информационно-технологической системы, в частности в странах развивающегося мира. Вопросы управления, например соблюдение новых директив и правил и развитие людских ресурсов, обладающих необходимым набором навыков и знаний, считаются серьезной проблемой в большинстве стран.

29. Систематическое использование неправительственных источников данных сопряжено с юридическими проблемами для правительств, желающих воспользоваться возможностями, которые заключают в себе большие данные. Если в одних странах уже введено в действие законодательство, обеспечивающее доступ как к правительственным, так и к неправительственным источникам данных, то в других странах отсутствие законодательных актов, гарантирующих доступ к неправительственным источникам, представляет проблему. Вопросы, связанные с неприкосновенностью частной жизни, также являются серьезной проблемой во многих странах, когда речь идет об использовании больших данных. Защита персональных данных в этой области является исключительно важным вопросом; следовательно, необходимо добиться оптимального баланса между защитой персональных данных и их использованием в статистических целях.

30. Проблематичными для ряда стран представляются и финансовые вопросы, в частности затраты на извлечение данных из источника, которые были названы серьезной проблемой при использовании больших данных. Кроме того, финансовые ресурсы, необходимые для приобретения новейших аппаратных средств и программного обеспечения, предназначенных для обработки больших данных, могут оказаться недостаточными или вообще отсутствовать. Затем, частные компании, как представляется, придают все большую значимость имеющимся в их распоряжении данным и поэтому запрашивают за них все более высокие цены.

31. Были названы и другие серьезные проблемы, связанные с использованием больших данных в национальных статистических системах. В частности, многие респонденты поднимали вопрос о методах интеграции больших данных в статистические системы. Как можно встроить этот новый источник данных в существующую статистическую программу? Большинство сходится в том, что большие данные не заменят собой официальную статистику, а, скорее, будут дополнять ее, позволяя, например, очень быстро получать статистические данные и бета-показатели в ключевых областях статистики. Поскольку большие данные представляют собой сравнительно новое понятие в сфере официальной статистики, они нуждаются в активном пропагандировании в кругах специалистов, занимающихся официальной статистикой.

### **С. Существующие и потенциальные области использования**

32. В третьей и заключительной части всемирного обследования респондентам предлагалось указать области использования (или изучения вопроса об использовании) больших данных в течение следующих 12 месяцев. На выбор были предложены следующие области: «Демографическая и социальная статистика», «Статистика естественного движения населения и записей актов гражданского состояния», «Экономическая и финансовая статистика», «Статистика цен», «Статистика транспорта», «Статистика окружающей среды» и «Другие области официальной статистики». Респондентам были даны два варианта ответа: «Нет» или «Да», где ответ «Да» нужно было сопровождать пояснением. Среди областей, в которых используются большие данные, чаще всего указывались следующие: «Демографическая и социальная статистика» и «Экономическая и финансовая статистика». Однако, как и в случае использования источников больших данных, положительные ответы в этой части обследования не-

обходимо было повторно проанализировать, выделив ту группу ответов, которую можно отнести к категории «надлежащей практики». После проведения такого повторного анализа оказалось, что надлежащая практика использования больших данных имеет место в области статистики цен и экономической и финансовой статистики. В частности, несколько стран указали, что они используют данные сканирования и/или методы просеивания веб-страниц для расчета часто обновляемых индексов цен, используемых в дополнение к стандартному индексу потребительских цен. В целом, надлежащая практика использования больших данных по областям статистики составляет в процентном отношении чуть более 10 процентов в области статистики цен, экономической и финансовой статистики и демографической и социальной статистики и около 5 процентов в каждой из других областей использования.

33. В области демографической и социальной статистики несколько стран упомянули о том, как данные из сети мобильной телефонной связи могут использоваться для выявления маршрутов путешествий и для целей туризма, настроения пользователей социальных сетей — для определения уровня уверенности потребителей, а имеющие в Интернете данные о вакансиях — для анализа занятости, и о потенциальных видах использования ресурса Trends в поисковой системе Google. Среди других примеров было названо сотрудничество с операторами мобильной телефонной связи для расчета статистики маятниковой миграции с конкретной географической привязкой. В области экономической и финансовой статистики было упомянуто изучение вопроса об использовании данных с изображений, полученных со спутников, и данных с наземных датчиков для оценки урожайности сельскохозяйственных культур и использование данных регистрации местонахождения мобильных устройств для подготовки статистики мобильности населения в сфере услуг. Несколько стран пояснили, как они используют большие данные в области статистики транспорта. Подробные данные об оплате проезда по платным автомагистралям и данные с автодорожных датчиков используются для подготовки статистики дорожного движения.

## **V. Большие данные и модернизация статистических систем**

34. Для того чтобы решить проблемы, указанные в ответах, полученных в ходе всемирного обследования, необходимо будет модернизировать статистические системы. По мнению респондентов, необходимо проводить более широкую исследовательскую работу для преодоления методологических трудностей, препятствующих использованию источников больших данных. С учетом характера больших данных вопросы репрезентативности и охвата населения могут вызвать проблемы. Кроме того, многообразие и потенциальный временный характер данных (например, вопрос о том, как долго люди будут продолжать пользоваться сетью Facebook) создают проблемы с точки зрения официальной статистики ввиду того, что основными качественными характеристиками статистических данных являются их сопоставимость, непрерывность и согласованность. Изменения в методологии могут включать более частое использование метода моделирования и могут потребовать более тесного сотрудничества с научными и учебными учреждениями. Необходимо проводить более широкие аналитические и экспериментальные исследования для изучения потенциаль-

ных видов использования больших данных в официальной статистике, и исследовательская работа должна, пожалуй, стать частью процесса подготовки статистики.

35. Существующая информационно-техническая инфраструктура и возможности нуждаются в наращивании, с тем чтобы они могли поддерживать обработку наборов данных большого объема и высокой скорости прироста, поступающих из различных источников. Кроме того, следует рассмотреть вопрос о стандартизации методов и инструментов, применяемых национальными статистическими управлениями. Необходимо приобретать новейшую технологию (программное обеспечение, аппаратные средства и кадровые ресурсы), чтобы постоянно отвечать технологическим требованиям, связанным со сбором, обработкой и использованием больших данных. Управлениям, возможно, придется рассмотреть возможность применения метода облачных вычислений вместо попыток перемещения огромных объемов данных на свои серверы. В качестве альтернативного варианта органу, где хранится источник больших данных, может быть предложено произвести первоначальные вычисления и представить статистическому управлению данные с более проработанной структурой и с более высоким уровнем агрегирования. Необходимо будет выделить достаточный объем ресурсов для создания требуемой информационно-технической инфраструктуры для частого извлечения данных из соответствующих источников, их увязки и анализа и для представления их в формате, подходящем для пользователей.

36. Серьезную проблему будет представлять как развитие, так и удержание на службе кадровых ресурсов, обладающих необходимыми навыками в области применения передовых методов анализа данных и вычислительной информатики. Освоение новых источников больших данных потребует реорганизации процесса сбора и обработки данных. Большинство национальных статистических управлений не располагает кадровыми ресурсами, обладающими соответствующими навыками, и поэтому, чтобы иметь возможность эксплуатировать такие источники данных, они должны будут наращивать свой потенциал и проводить интенсивное обучение людских ресурсов. Имеющихся специалистов, обладающих классическими статистическими знаниями, будет недостаточно для обеспечения надлежащей обработки этого нового типа данных, и поэтому необходимо будет выделить ассигнования на цели найма и обучения достаточного числа новых профессиональных работников (например, ученых в области информатики).

37. Для подготовки официальной статистики с использованием больших данных требуется принятие дополнительных законодательных мер. Во многих странах существующие нормативно-правовые основы статистики не охватывают вопросы доступности и использования больших данных, как имеющихся в государственном секторе, так и поступающих из частного сектора. Поэтому будет особенно трудно получить доступ к большим данным, сбор и хранение которых осуществляются другими сторонами. Кроме того, требуется комплекс правовых норм, обеспечивающих неприкосновенность частной жизни, который устанавливал бы базовые правила, регулирующие возможные методы объединения, защиты, передачи, обнародования, анализа и сохранения наборов больших данных. Это позволило бы решить важный вопрос о недопущении злоупотребления общественным доверием в процессе надлежащего использования правительственными органами персональных данных о частных лицах. Важно

сохранять общественное доверие: частные лица должны быть уверены в том, что их личная информация будет должным образом защищена и не станет объектом разглашения или злоупотребления. Например, одним из потенциальных источников больших данных являются данные о местонахождении, получаемые из сети мобильной телефонной связи, однако даже при сокрытии личности абонентов люди все равно будут крайне обеспокоены возможностью передачи таких сведений компаниями, предоставляющими мобильную телефонную связь, другим сторонам для любых целей и возможностью обмена такой информацией. Аналогичным образом, компаниям, предоставляющим связь через мобильные устройства, требуются гарантии того, что при передаче ими своих данных правительству права на неприкосновенность частной жизни не будут нарушены.

38. Реакция общественности на использование больших данных для подготовки официальной статистики является одним из важнейших факторов, определяющих будущие перспективы использования больших данных. Правительство и, в еще большей мере, частный сектор должны не злоупотреблять, а, наоборот, заручиться доверием общественности и вселить в нее уверенность в том, что персональные данные будут использоваться надлежащим образом. Государственному сектору необходимо будет направить значительные усилия на создание санкционирующей правовой основы, которая устанавливала бы правила доступа к автоматически генерируемым данным в цифровом формате, содержащим элементы личной информации, и правила пользования такими данными и обеспечивала бы гласность в том, что касается функций органов, занимающихся хранением данных, и функций соответствующих государственных ведомств. Только при таких условиях могут быть гарантированы неприкосновенность частной жизни и конфиденциальность личной информации и может быть обеспечено доверие общественности в том, что касается использования больших данных в официальной статистике.

## **VI. Тематические исследования**

### **A. Австралия: использование изображений, полученных со спутников, в статистике сельского хозяйства**

39. Изображения, полученные со спутников, используются при определении видов земного покрова и в исследовательской программе, в рамках которой параметры сельскохозяйственного землепользования применяются для оценки конкретных видов сельскохозяйственных культур. Полученные пространственные параметры вводятся в качестве исходных данных в статистические модели, созданные для произведения расчетов при составлении статистики сельского хозяйства. В потенциале они могут применяться и при составлении статистики рыбного и лесного хозяйства, а также статистики окружающей среды. Исходные данные хранятся и обрабатываются на облачном сервере в системе поставщика данных, вне электронно-вычислительной среды статистической организации. Алгоритмы, применяемые при анализе изображений для извлечения пространственных параметров землепользования, разрабатываются на основе

совместных исследований с участием сотрудничающих университетов. По линии этого сотрудничества разрабатываются алгоритмы, модели пространственных и временных рядов данных, системы геокодирования и процессы статистической оценки.

## **В. Бразилия: использование данных высокой частотности, получаемых с контрольных станций, для составления счетов водных ресурсов**

40. Соглашением, подписанным в 2012 году Национальным институтом географии и статистики (статистический орган), Национальным агентством водного хозяйства (федеральный орган по водным ресурсам) и Секретариатом по водным ресурсам и городской экологии при министерстве охраны окружающей среды (орган, координирующий политику в области использования водных ресурсов), был создан комитет, которому поручено разработать счета водных ресурсов в Бразилии. Национальное агентство водного хозяйства осуществляет контроль за количеством и качеством водных ресурсов через Национальную гидрометеорологическую сеть, включающую почти 22 000 контрольных станций, которые собирают данные высокой частотности об объеме водотоков, уровне дождевых осадков, качестве воды, скорости испарения и образовании отложений. Иными словами, Национальное агентство водного хозяйства каждый день собирает большой объем данных о водных ресурсах, обрабатывает эти данные и размещает их в свободном доступе на институциональном веб-сайте Агентства<sup>17</sup> или через Национальную информационную систему по водным ресурсам<sup>18</sup>. Информационная система представляет собой всеобъемлющую систему сбора, обработки, хранения и предоставления данных и информации, связанных с рациональным использованием водных ресурсов. Данные, полученные через Информационную систему, объединяются Национальным институтом географии и статистики с данными, взятыми из его обычного набора данных экономической статистики и статистики окружающей среды, в целях составления счетов водных ресурсов Бразилии.

## **С. Бутан: использование технологии мобильной связи при расчете индекса потребительских цен**

41. В Бутане сбор данных о потребительских ценах поддерживается и облегчается использованием современной технологии связи. Индекс потребительских цен рассчитывается ежемесячно на основе данных, которые собираются в различных округах местными работниками статистического ведомства с использованием мобильной платформы телефонной связи Android. Работники собирают данные по всем наименованиям товаров и услуг, входящих в корзину индекса потребительских цен, в соответствующих магазинах в своих округах и автоматически пересылают данные на сервер головного отделения Национального статистического бюро в Тхимпху. Головное отделение объединяет данные, поступающие из различных округов, и производит ежемесячный расчет индекса потребительских цен и показателей инфляции на национальном уровне.

<sup>17</sup> <http://www2.ana.gov.br>.

<sup>18</sup> <http://www.snirh.gov.br/telemetria/consultardadoshidrovisualizacaogeo/inicial.action>.



42. Даже несмотря на то, что нынешняя система, действующая в Бутане, не связана с получением больших объемов данных, развитие применения мобильных устройств и беспроводной связи могло бы продолжаться, и они могли бы использоваться для сбора гораздо большего объема данных о ценах. Применяемый в Бутане подход, начинающийся со сбора электронных данных в небольших масштабах, мог бы служить отправной точкой и для других развивающихся стран. Автоматизация ввода цифровых данных и беспроводная передача информации на централизованный сервер являются важными первыми шагами в направлении освоения потенциальных возможностей, которые включают в себе источники больших данных.

#### **D. Колумбия: использование изображений, полученных со спутников, в статистике сельского хозяйства и транспорта**

43. В Колумбии изображения, полученные со спутников, используются в качестве источника больших данных в целом ряде областей. Одним из примеров в этой связи является экспериментальная проверка, основанная на использовании изображений, полученных со спутников, в дополнение к работе, которую выполняют проводящие опрос сотрудники в ходе национальной сельскохозяйственной переписи. Цель диагностического этапа проверки состоит в том, чтобы выявить те виды переменных величин, которые могут быть зафиксированы путем получения изображений со спутников и которые могут использоваться для подтверждения информации, собираемой в ходе проведения переписи. Другой пример связан с использованием изображений, полученных со спутников, Управлением Организации Объединенных Наций по наркотикам и преступности и министерством внутренних дел и юстиции для количественной оценки и мониторинга посевов коки в Колумбии через комплексную систему мониторинга посевов незаконных культур.

44. Национальный автодорожный институт Колумбии использует данные GPS для улучшения движения транспорта и в качестве исходной информации в статистике транспорта. При применении данного метода автомобилям не нужно останавливаться у пунктов оплаты проезда; вместо этого на каждом автомобиле устанавливается электронное регистрирующее устройство, которое считывается при приближении автомобиля к пункту оплаты. Такая динамическая система оплаты проезда по платным дорогам обеспечивает высокий уровень удобства, безопасности и обслуживания. Регистрирующее устройство содержит также всю информацию о данном автомобиле, которая дополняет собой информацию, имеющуюся в Национальном едином транзитном реестре. На сегодняшний день этот новый метод опробован на 10 станциях оплаты проезда в Колумбии и уже дал положительные результаты в плане усиления контроля за транспортными потоками и укрепления статистики транспорта.

## **Е. Эстония: использование данных о местонахождении из сети мобильной связи для подготовки статистики международных поездок**

45. Совместными усилиями Центрального банка Эстонии, Тартуского университета и компании «Позициум ЛБС» разработана методология составления статистики международных поездок, которая основывается на использовании данных о местонахождении, пассивно поступающих из сети мобильной связи. Компания «Позициум» занимается сбором и обработкой анонимных данных, поступающих от операторов мобильной связи, с помощью статистических моделей и строит на их основе общие ряды данных. Данная методология позволяет получать достоверную картину движения через эстонскую границу физических лиц, выезжающих за рубеж (выезд из страны) или въезжающих в Эстонию (въезд в страну). Данные о местонахождении, поступающие из сети мобильной связи, являются анонимными. Они обрабатываются, анализируются и распространяются с использованием методологии, которая полностью исключает применение каких бы то ни было методов идентификации номеров телефонов или имен их владельцев.

46. Центральный банк пользуется статистикой международных поездок для расчета объема импорта и экспорта услуг по организации поездок при составлении счета текущих операций в платежном балансе. В соответствии с методологией расчета платежного баланса место жительства пассажира при использовании данных о местонахождении, поступающих из сети мобильной связи, определяется по критерию постоянного места жительства, независимо от того, гражданином какой страны является данное лицо. Предположительное место жительства владельцев телефонов определяется путем использования телефонного роуминга и местных сетей мобильной связи. Число, продолжительность и характер посещений Эстонии лицами, являющимися и не являющимися ее жителями, определяются по динамике пользования мобильными телефонами в привязке к конкретным населенным пунктам, в зависимости от страны и региона пользования телефонной связью.

47. Статистика выездов из страны отражает поездки ее жителей в зарубежные страны на один день или с ночевкой. Выезды жителей Эстонии из страны могут включать поездки в несколько стран. Страной назначения считается наиболее удаленная от Эстонии страна или страна (страны), где совершающее поездку лицо проводит не меньше двух дней подряд. Другие страны, в которых совершающее поездку лицо не останавливается на ночевку и которые не отвечают критериям страны назначения, относятся к категории стран транзита. Число и продолжительность посещений Эстонии лицами, являющимися и не являющимися ее жителями, рассчитываются с помощью статистических моделей и добавляются в общие ряды данных, при этом в расчет принимаются и другие официальные источники данных (количество пассажиров в Таллинском морском порту, количество пассажиров в Таллинском аэропорту, количество пересечений восточной границы Эстонии, официальная гостиничная статистика).

48. Анонимные данные о местонахождении, поступающие из сети мобильной связи, используются с соблюдением эстонского законодательства и законодательства Европейской комиссии, касающихся защиты информации и защиты персональных данных (Закон Эстонии о защите персональных данных; директива 95/46/ЕС Европейского парламента и Совета Европейского союза от

24 октября 1995 года о защите физических лиц при обработке персональных данных и о свободной передаче таких данных; и директива 2002/58/ЕС Европейского парламента и Совета Европейского союза от 12 июля 2002 года, касающаяся обработки персональных данных и защиты неприкосновенности частной жизни в сфере электронных коммуникаций). Методология определения местонахождения с использованием сети мобильной связи разработана и с 2006 года применяется в различных сферах деятельности в Эстонии, в сотрудничестве с Инспекцией по защите информации при решении вопросов, касающихся защиты персональных данных.

## **Е. Нидерланды: социальные сети как потенциальный источник данных для официальной статистики**

49. Все большее число людей активно пользуется социальными сетями, где они добровольно обмениваются информацией, обсуждают интересующие их темы и контактируют с членами семьи и друзьями. В Нидерландах социальные сети используются значительной частью населения; около 70 процентов сообщили о том, что они являются активными пользователями одной или нескольких социальных медиа-платформ. Поскольку доля респондентов, отвечающих на опросники Статистического бюро Нидерландов, продолжает снижаться, был изучен вопрос о возможности использования сообщений, которые добровольно оставляют пользователи социальных сетей и которыми они добровольно обмениваются, в качестве источника данных для подготовки официальной статистики. Для этого были проанализированы сообщения, имеющиеся в публичном доступе в социальных сетях, которые создаются на различных социальных медиа-платформах, таких как Twitter и Facebook, а также публичные сообщения, помещаемые на новостных сайтах, веб-форумах и в блогах. Эти сообщения были предоставлены коммерческой компанией, которая регулярно собирает все имеющиеся в публичном доступе сообщения на голландском языке, появляющиеся в голландскоязычном Интернете.

50. Сообщения были проанализированы как по содержанию, так и по настроению, выраженному их авторами. Анализ содержания сообщений, распространенных через голландскоязычный Twitter — социальную медиа-платформу, на которой создается большинство имеющихся в публичном доступе сообщений на голландском языке, показал, что почти 50 процентов сообщений содержали «пустую болтовню». Остальная часть сообщений касалась преимущественно обсуждений свободного времяпрепровождения (10 процентов), работы (7 процентов), средств массовой информации (телевидение и радио) (5 процентов) и политики (3 процента). Использованию этих более серьезных сообщений мешало наличие менее серьезных, «пустых» сообщений. Анализ настроений, выраженных во всех сообщениях, созданных на всех имеющихся платформах, выявил весьма интригующую возможность для потенциального использования этих данных при подготовке статистики. Было установлено, что настроения, выраженные в этих сообщениях, имеют высокую коррелятивную связь с уровнем уверенности потребителей в Нидерландах, в частности с настроениями, связанными с экономическим положением в стране. Такая корреляция с экономическими настроениями оставалась стабильной и на ежемесячной, и на еженедельной основе. В то же время ежедневно рассчитываемые цифры дают крайне неустойчивую зависимость, что указывает на возможность

подготовки ежемесячных и еженедельных показателей настроения пользователей социальных сетей, сопоставимых с уровнем потребительской уверенности. Такие показатели могут быть рассчитаны уже в первый рабочий день по прошествии анализируемой недели, что свидетельствует о способности быстрого получения результатов.

## **VII. Выводы и план дальнейшей работы**

51. Потенциальные возможности использования больших данных для подготовки официальной статистики признаны статистическим сообществом. Большие данные могут помочь более эффективному выполнению задачи своевременной подготовки согласованных статистических данных по экономике, социальной сфере и экологии для целей принятия решений, проведения исследований и общественных обсуждений. Кроме того, по мере совершенствования технологии геокодирования статистических единиц в экономической, социальной и экологической сферах и увеличения объема имеющейся информации на самом нижнем уровне географической детализации, ожидается скорейший прогресс в области использования больших данных в развитых и развивающихся странах. Поэтому для того, чтобы воспользоваться преимуществами, которые заключают в себе большие данные, между региональными инициативами внутри мирового статистического сообщества необходимо наладить обмен методическими разработками, передовой практикой решения стратегических вопросов и возможностями обучения, в том числе в деле решения вопросов, связанных с законодательной базой, неприкосновенностью частной жизни, финансами, управлением, методологией и технологиями. При решении этих вопросов, связанных с использованием больших данных, правительствам и корпоративному сектору необходимо будет создать условия, при которых могут быть гарантированы неприкосновенность частной жизни и конфиденциальность личной информации для целей официальной статистики и может быть обеспечено доверие общественности в том, что касается использования больших данных в официальной статистике.

52. В ходе глобальной оценки использования больших данных для подготовки официальной статистики был выявлен ряд положительных примеров, связанных в первую очередь с использованием данных, поступающих из операционных источников, с регистрирующих устройств, из сенсорных источников и со спутников. Статистика цен рассчитывалась с использованием информации, полученной со сканирующих устройств, потребительские расходы — с использованием электронных операций по карточкам, статистика поездок и транспорта — с использованием регистрирующих устройств системы GPS, статистика благосостояния населения — с использованием данных из социальных сетей, а статистика сельского хозяйства — с использованием изображений, полученных со спутников. В ходе всемирного обследования было выявлено также множество потенциальных проблем, связанных прежде всего с методологическими трудностями, возникающими при составлении больших данных, с информационно-техническими требованиями и с вопросами, касающимися неприкосновенности частной жизни.

53. Очевиден тот факт, что решение проблем, связанных с использованием больших данных, предполагает модернизацию статистической системы. Например, необходимо будет укрепить исследовательский сектор статистической

системы, наращивать информационно-технологические возможности, провести структурную реорганизацию людских ресурсов за счет привлечения ученых по информатике и наладить партнерские отношения с частным сектором в области обмена автоматически генерируемой информацией в цифровом формате в условиях, при которых соблюдаются принципы неприкосновенности частной жизни и соблюдаются правила конфиденциальности.

54. Большие данные и модернизация статистических систем создают для большинства стран весьма схожие проблемы и возможности. Поэтому между национальными статистическими ведомствами может быть налажен аналогичный обмен опытом, практическими методами и решениями. На региональном уровне уже начато осуществление ряда инициатив в этом направлении. В контексте Конференции европейских статистиков был одобрен предлагаемый проект ЕЭК, посвященный развитию роли больших данных в деле модернизации производства статистической продукции. Аналогичные инициативы осуществляются в настоящее время по линии ЭСКАТО в Азиатско-Тихоокеанском регионе. Кроме того, рассматривается вопрос о начале реализации проектов аналогичной направленности в других регионах.

55. Предполагается, что региональные инициативы и начинания в рамках глобальной программы работы будут дополнены деятельностью глобальной рабочей группы по вопросам использования больших данных. Данная программа обеспечит возможности для обмена практическими методами работы между региональными комиссиями. Если региональные комиссии могли бы заниматься реализацией конкретных проектов с участием меньшего числа стран на региональном уровне, то глобальная рабочая группа на совместной основе занималась бы решением вопросов в целях достижения соглашения по аспектам, связанным с методологией, передовой практикой решения стратегических вопросов, обучением и использованием общих платформ. В рамках своей программы работы глобальной рабочей группе следует подготовить веские доводы, доказывающие, что национальные статистические системы обладают сравнительными преимуществами по сравнению с другими сторонами с точки зрения их роли в плане подтверждения состоятельности использования больших данных для целей выработки политики или в плане объединения источников больших данных с существующими источниками официальной статистики.

56. Глобальная рабочая группа могла бы заниматься дальнейшим рассмотрением конкретных вопросов, касающихся роли больших данных в программе развития на период после 2015 года, и содействовать налаживанию партнерских отношений с другими учреждениями Организации Объединенных Наций и международными учреждениями, организациями гражданского общества и другими участниками глобального партнерства по реализации формирующейся инициативы в области информационной революции. Одна из претензий, которая высказывалась применительно к показателям достижения целей в области развития, сформулированных в Декларации тысячелетия, была связана с несвоевременностью их подготовки. Большие данные в большинстве случаев образуются в масштабе реального времени и в потенциале могут обеспечить подготовку весьма своевременных показателей.

57. Для того чтобы глобальная рабочая группа могла начать свою работу, во втором или третьем квартале 2014 года будет организована международная конференция, на которой будет выработана программа работы по вопросам общей методологии, обучения, платформ и по другим вопросам, связанным с использованием больших данных для подготовки официальной статистики.

## **VIII. Вопросы для обсуждения**

58. Статистической комиссии предлагается высказать свои мнения в отношении:

а) необходимости дальнейшего изучения источников, проблем и областей использования больших данных для подготовки официальной статистики на глобальном уровне, особенно с учетом условий, в которых находятся развивающиеся страны, и в свете увязки с программой развития на период после 2015 года и с инициативой в области информационной революции;

б) создания глобальной рабочей группы по вопросам использования больших данных для подготовки официальной статистики, деятельность которой дополняла бы работу, проводимую региональными комиссиями, и которая занималась бы решением вопросов, имеющих общерегиональную значимость.

---