

**Европейская экономическая комиссия****Конференция европейских статистиков****Шестьдесят первая пленарная сессия**

Женева, 10–12 июня 2013 года

Пункт 5 предварительной повестки дня

**Работа Группы высокого уровня по модернизации  
статистического производства и услуг****Каково значение "больших данных" для официальной  
статистики?****Записка секретариата***Резюме*

На Семинаре высокого уровня "Модернизация процесса производства статистических данных и услуг", состоявшемся в Санкт-Петербурге 3–5 октября 2012 года, участники просили подготовить "документ, поясняющий вопросы, связанные с использованием "больших данных" в сообществе официальной статистики". Они заявили о желании располагать документом стратегической направленности, ориентированным на глав или руководителей высшего звена статистических организаций.

Для выполнения этого поручения Группа высокого уровня по модернизации статистического производства и услуг учредила неформальную Целевую группу<sup>1</sup>, в состав которой вошли национальные и международные эксперты и работа которой координируется секретариатом Европейской экономической комиссии Организации Объединенных Наций. Группа высокого уровня хотела бы получить замечания по содержанию этого документа и соответствующие рекомендации.

Этот документ и другую информацию о работе Группы высокого уровня можно найти в Интернете по адресу: [www1.unece.org/stat/platform/display/hlgbas](http://www1.unece.org/stat/platform/display/hlgbas)

<sup>1</sup> В состав этой Целевой группы входят: Майкл Глассон (Австралия), Жюли Трепанье (Канада), Винченцо Патруно (Италия), Пит Дас (Нидерланды), Михаил Скалиотис (Евростат) и Анджум Кан (Европейская экономическая комиссия Организации Объединенных Наций).

## I. Введение

1. В нашем современном мире все больше данных генерируется в Сети и производится датчиками на все большем числе электронных устройств, окружающих нас. Объем данных и частота, с которой их производят, привели к появлению понятия "большие данные". Большие данные представляют собой наборы данных растущего объема, скорости и разнообразия. Большие данные часто являются в значительной степени неструктурированными, что означает, что они не имеют какой-то определенной заранее модели и/или не вписываются в традиционные реляционные базы данных. Помимо создания новых торговых возможностей в частном секторе большие данные также могут представлять значительный интерес с точки зрения их использования в качестве исходных данных для официальной статистики: либо в отдельности, либо в сочетании с более традиционными источниками данных, такими как выборочные обследования и административные регистры. Тем не менее извлекать информацию из больших данных и использовать ее в процессе статистического производства нелегко. Исходя из этого, в настоящем документе будут рассматриваться два основных вопроса: *какие данные и каким образом*.

а) *Какому поднабору больших данных национальные статистические организации (НСО) должны уделять повышенное внимание исходя из роли официальной статистики?*

б) *Каким образом НСО могут использовать большие данные и преодолевать связанные с ними трудности?*

2. Ответ на вопрос "какие данные?" содержится в документе под названием "Большие данные в целях развития", подготовленном в рамках проекта Организации Объединенных Наций "Глобальный пульс"<sup>2</sup>.

*(стр. 6) На самом общем уровне проанализированные надлежащим образом большие данные позволяют получать "мгновенные снимки" уровня благосостояния населения с высокой периодичностью, большой степенью детализации и с самых различных углов зрения, сокращая при этом необходимое время и пробелы в знаниях. Анализ этих данных может фактически помочь обнаружить то, что "Глобальный пульс" назвал "цифровыми сигналами тревоги" – аномальные изменения в доступе общин к услугам, которые могут служить косвенными показателями изменений в собственно благосостоянии людей.*

*(стр. 12) Официальная статистика будет продолжать генерировать релевантную информацию, однако революция цифровых данных (стр. 9 – "Действия, решения и предпочтения людей в повседневной жизни, поддающиеся отслеживанию или хранению в цифровом формате") открывает громадные возможности с точки зрения получения данных для всестороннего и углубленного изучения жизни людей, которые могут дополнять показатели, по которым уже ведется сбор данных.*

3. Большие данные обладают потенциалом разработки более релевантной и своевременной статистики по сравнению с традиционными источниками официальной статистики. Официальная статистика практически полностью основывается на сборе данных в рамках обследований и получении административных данных от правительственных программ, что часто является прерогативой

<sup>2</sup> <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.

вой НСО, вытекающей из законодательства. Однако с большими данными дело обстоит иначе, поскольку большинство таких данных находится в открытом доступе или у частных компаний. В результате частный сектор может использовать эру больших данных для разработки все большего объема статистических данных, которые способны превзойти официальную статистику по своевременности и релевантности. НСО вряд ли утратят свое фирменное наименование "поставщиков официальной статистики", однако если они останутся в стороне, то могут постепенно утратить свою репутацию и релевантность. Одним из больших преимуществ НСО является наличие инфраструктуры для обеспечения точности, непротиворечивости и интерпретируемости производимых статистических данных. Задействуя соответствующие источники больших данных в своем процессе подготовки официальной статистики, НСО способны лучше других оценивать их точность, обеспечивать непротиворечивость всей системы официальной статистики и надлежащее толкование данных и при этом постоянно работать над их релевантностью и своевременностью. Таким образом удастся сохранить роль и значимость официальной статистики.

## II. Определения

4. Определение официальной статистики можно вывести из Основных принципов официальной статистики, принцип 1<sup>3</sup> которых гласит:

"Официальная статистика является необходимым элементом информационной системы демократического общества, обеспечивая правительство, экономические круги и общественность *данными об экономическом, демографическом, социальном и экологическом положении*".

5. В основе официальной статистики лежит идея о том, что официальная статистика является статистикой, которая описывает определенную ситуацию, создавая *картину* страны, ее экономики, ее населения и т.д. При использовании больших данных в качестве дополнительного источника информации эту *картину* необходимо учитывать.

6. Большие данные можно определить на основе определения, используемого компанией "Гартнер"<sup>4</sup>:

«Большие данные представляют собой источники данных, которые в общем можно описать как "данные больших объема, скорости и разнообразия, требующие затратоэффективных и инновационных видов обработки для совершенствования анализа и процесса принятия решений"».

## III. Источники

7. Такое явление, как большие данные, заставляет нас осознать, что наш мир сейчас полон данных. Этого нельзя игнорировать, чем и обусловлен к ним интерес со стороны официальной статистики. В настоящее время существуют главным образом два различных способа разработки НСО и международными организациями (МО) статистических данных: на основе выборочных обследований и на основе административных источников, включая регистры. Необходимо ответить на следующий вопрос: каким образом большие данные могут

<sup>3</sup> <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>.

<sup>4</sup> <http://www.gartner.com/it-glossary/big-data/>.

помочь более точно и своевременно измерять экономические, социальные и экологические явления?

8. В целом источники больших данных можно разделить на следующие группы:

a) административные (являющиеся результатом реализации государственной или иной программы), например электронные медицинские карты, посещения больниц, досье страховых компаний, банковские данные, производственные банки и т.д.;

b) торговые или операционные (являющиеся результатом операции между двумя субъектами), например операции по кредитным картам, онлайн-вые операции (в том числе через мобильные устройства) и т.д.;

c) датчиковые, например спутниковые изображения, дорожные датчики, климатические датчики и т.д.;

d) отслеживания, например данные отслеживания с мобильных телефонов, Глобальной навигационной системы (GPS) и т.д.;

e) поведенческие, например запросы в онлайн-поисковых системах (для получения информации о продукте, услуге или любого другого типа информации), просмотр веб-страниц и т.д.;

f) оценочные, например комментарии в социальных сетях и т.д.

9. Административные данные являются одним из основных источников данных, используемых НСО для статистических целей. Административные данные регулярно собираются статистическими управлениями и используются для подготовки официальной статистики. Традиционно эти данные получались часто из органов государственного управления, обрабатывались, хранились, велись и использовались НСО в хорошо структурированной форме. Можно ли считать административные данные "большими" согласно вышеуказанному определению? На сегодняшний день вероятным ответом будет "нет". Административные данные могут стать "большими" при увеличении скорости их появления, например при широком использовании административных данных, когда данные вместо обычного их сбора раз в год или раз в месяц собираются ежедневно или еженедельно.

## IV. Трудности

10. Использование больших данных в официальной статистике связано с многочисленными трудностями. В этом разделе рассматриваются основные трудности, относящиеся к следующим категориям:

a) законодательные, с точки зрения доступа к данным и их использования;

b) связанные с конфиденциальностью, т.е. обеспечение доверия общественности и получение ее согласия на многократное использование данных и их увязку с другими источниками;

c) финансовые, т.е. потенциальные издержки получения данных из сторонних источников в сопоставлении с выгодами;

d) управленческие, например политика и директивы в области управления данными и защиты данных;

- e) методологические, т.е. качество данных и применимость статистических методов;
- f) технологические, т.е. вопросы, связанные с информационной технологией.

## **А. Трудности законодательного порядка**

11. Законодательство некоторых стран (например, Канада) может предоставлять право доступа к данным как государственным, так и негосударственным организациям, в то время как в других странах (например, в Ирландии) право доступа к данным предоставляется только государственным органам. Это может привести к ограничению доступа к некоторым видам больших данных, которые описывались в пункте 8.

12. Признано<sup>5</sup>, что:

"Право доступа НСО к административным данным, установленное в принципе законом, часто не подкрепляется надлежащим образом конкретными обязательствами для держателей данных".

13. Даже если законодательство предусматривает доступ ко всем видам данных, в различных странах может понадобиться в разной степени обосновывать статистическую цель получения доступа к этим данным.

## **В. Связанные с конфиденциальностью трудности**

14. Определения могут быть различными в зависимости от страны, однако обычно конфиденциальность определяется как *право людей контролировать или регулировать содержание касающейся их информации, которая может быть раскрыта*. Можно провести параллель с компаниями, которые желают защитить свою конкурентоспособность и своих потребителей. Право на конфиденциальность является одной из основ демократии. Проблема, связанная с большими данными, заключается в том, что пользователи услуг и устройств, генерирующих данные, скорее всего, не знают, что они это делают и/или для чего эти данные могут использоваться. Увязка данных приведет к росту их объема, а также озабоченности по поводу сохранения конфиденциальности.

## **С. Финансовые трудности**

15. Как представляется, есть цена, которую НСО должны платить за приобретение больших данных, особенно больших данных, держателем которых является частный сектор, в частности, если законодательство обходит молчанием финансовые условия приобретения внешних данных. Поэтому НСО должны принимать взвешенные решения, учитывающие, с одной стороны, требования качества (которые включают в себя релевантность, своевременность, точность, целостность, доступность и интерпретируемость), а с другой стороны, издержки и сокращение нагрузки на респондентов. Издержки могут даже быть существенными для НСО, однако потенциальные выгоды могут быть значительно их перевешивать, поскольку из больших данных можно извлекать информацию, которая способна повысить эффективность государственных программ (напри-

<sup>5</sup> <http://essnet.admindata.eu/WorkPackage?objectId=4251> (p41).

мер, системы здравоохранения). В этом случае также могут применяться правила государственных закупок.

16. Один из выводов доклада, подготовленного Федеральной комиссией по вопросам больших данных организации "ТехАмерика" Соединенных Штатов<sup>6</sup>, заключается в том, что успех перехода на большие данные зависит от

"понимания главных профессиональных задач потребностей конкретного учреждения, правильной постановки вопросов и понимания искусства возможного, а также предпринятия первоначальных шагов по обслуживанию набора ряда четко определенных видов использования".

17. Такой подход, безусловно, можно использовать и в работе НСО.

#### **D. Управленческие трудности**

18. Использование больших данных в целях официальной статистики означает поступление в НСО все большего объема информации, которая должна отвечать требованиям политики и директив в области управления и защиты информации, которые НСО должны соблюдать.

19. Еще одна управленческая трудность связана с людскими ресурсами. Наука о данных<sup>7</sup> применительно к большим данным, появляющимся в частном секторе, по всей видимости, пока еще не установила связей с сообществом официальной статистики. Возможно, НСО придется проводить внутренние и национальные опросы (научных кругов, государственного и частного секторов) для выявления "ученых по данным" и подключения их к сфере официальной статистики.

#### **E. Методологические трудности**

20. Основной проблемой, связанной с большими данными, является репрезентативность. Трудность определения целевой совокупности, обследуемой совокупности и основы выборки ставит под угрозу традиционное представление официальных статистиков о целевой (и конечной) совокупности и их статистический анализ. При традиционном обследовании статистики определяют целевую/обследуемую совокупность, создают основу выборки для охвата этой совокупности, формируют выборку, собирают данные и т.д. Затем они строят таблицу и заполняют ее данными в строго структурированной форме. В случае больших данных сначала появляются сами данные, на основе которых официальным статистикам предстоит построить таблицу. В связи с этим возникает вопрос: *неужели это единственный способ построения целостной и комплексной национальной системы официальной статистики?* Возможно пришло время выйти за рамки таблицы?

<sup>6</sup> <http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf>.

<sup>7</sup> "Википедия" определяет науку о данных как науку, охватывающую различные элементы и опирающуюся на технические приемы и теории из различных областей, включая математику, статистику, инженерию данных, теорию распознавания образов, передовые информационно-вычислительные системы, визуализацию, моделирование погрешностей, хранение данных и обработку данных с помощью высокопроизводительных компьютеров, с целью извлечения смысла из данных и создания информационных продуктов.

21. Еще одна проблема связана с информационной технологией (ИТ) и самой методологией. При анализе все большего объема данных традиционные статистические методы, разработанные для крайне тщательных анализов небольших выборок, становятся неэффективными; в самых простых случаях они оказываются просто недостаточно быстрыми. В связи с этим возникла потребность в таких новых методах и инструментах, как:

а) методы быстрого извлечения информации из массивов имеющихся данных, такие как методы визуализации и методы извлечения данных, текста и потокового извлечения знаний, которые способны "ужимать большие данные". Увеличение вычислительной мощности компьютеров является первоначальной мерой поддержки на этом этапе;

б) методы, способные интегрировать информацию, обнаруженную в статистическом процессе, такие как увязка в массовом масштабе, макро/мезоинтеграция, а также статистические методы, конкретно подходящие для больших наборов данных. Необходимо разработать методы, которые позволили бы быстро получать надежные результаты при работе с весьма большими наборами данных.

22. Использование больших данных в официальной статистике, безусловно, порождает потребность в новых технических приемах. Эти технические приемы связаны со следующими методологическими проблемами, которые необходимо решить:

а) показатели качества результатов, полученных из внешних поставок данных, с трудом поддающихся регулированию. Зависимость от внешних источников ограничивает набор показателей, которые могут указываться при сравнении с результатами, полученными при использовании целевых методов сбора информации;

б) ограниченная сфера применения и ценность данных, полученных из внешних источников;

в) сложности, связанные с интеграцией информации из различных источников для получения высококачественных продуктов;

г) трудность выявления предлагаемых преимуществ в отсутствие замкнутой обратной связи, существующей в коммерческих организациях.

## **Ф. Технологические трудности**

23. Как указывалось в пункте 9, увеличение скорости доступа к административным данным предполагает также активное использование стандартных интерфейсов прикладного программирования (ИПП) или (иногда) исполнение ИПП в потоковом режиме для доступа к данным. Таким образом, можно связать приложения для сбора данных и обработки данных напрямую с административными данными. Действительно, сбор данных в режиме реального времени и режиме, близком к реальному времени, максимально увеличивает потенциал данных, открывая новые возможности для увязки административных данных с высокоскоростными данными, поступающими из других различных источников, таких как:

а) коммерческие данные (операции по кредитным картам, онлайнновые операции, продажи и т.д.);

- b) отслеживающие устройства (сотовые телефоны, GPS, камеры наблюдения, приложения) и физические датчики (метеорологические и энергетические датчики, датчики движения и загрязнения окружающей среды и т.д.);
- c) социальные сети ("Твиттер", "Фейсбук" и т.д.) и поисковые системы (поиск в Интернете, просмотр веб-страниц);
- d) общинные данные (сообщения граждан или краудсорсинговые данные).

24. В эпоху больших данных это изменение парадигмы сбора данных позволяет собирать и интегрировать множественные виды данных из многочисленных различных источников. Комбинирование источников данных для получения новой информации – это еще одна интересная задача на ближайшее будущее. Комбинирование "традиционных" источников данных, таких как обследования и административные данные, с новыми источниками данных, а также новых источников данных между собой позволяет описывать модели поведения "разумных" сообществ. Это – еще не исследованная область, которая способна открыть новые возможности.

## V. Каким образом большие данные могут использоваться в сообществах официальной статистики?

25. В этом разделе обсуждается несколько примеров проводимых или планируемых исследований в области больших данных. Первые два примера касаются Нидерландов.

26. **Статистика дорожного движения и транспортных потоков.** В Нидерландах ежедневно поступает около 80 млн. данных с дорожных петлевых детекторов. Эти данные могут использоваться в качестве источника информации для статистики дорожного движения и транспортных потоков, а также в целях статистики, описывающей другие экономические явления. Эти данные поступают на очень подробном уровне. Конкретно, более 10 000 петлевых детекторов на голландских дорогах ежеминутно регистрируют количество проезжающих машин по разным категориям в зависимости от их длины. Категории длины позволяют проводить различие между легковыми автомобилями и грузовиками. Недостаток этого источника заключается в том, что он страдает серьезной неполнотой охвата и избирательностью. Данные о количестве регистрируемых автомобилей недоступны по каждой минуте, и пока еще не на всех (важных) голландских дорогах установлены петлевые детекторы. На наиболее подробном уровне, т.е. на уровне отдельных петлевых детекторов, количество регистрируемых транспортных средств демонстрирует (высоко) волатильную динамику, а значит, требуется применение более строгого статистического подхода. Извлечение большого объема информации из данных представляет серьезную проблему для статистики. Использование этой информации в полном объеме могло бы позволить получать более оперативную и надежную статистику о дорожном движении, а также более подробную информацию о передвижении крупных транспортных средств, служащем индикатором изменений в экономическом развитии.

27. **Статистика социальных сетей.** В Нидерландах в общедоступных социальных сетях ежедневно появляется около 1 млн. сообщений. Доступ к этим сообщениям имеет любое лицо, подключенное к Интернету. Социальные сети



способны служить источником данных, поскольку люди добровольно обмениваются в них информацией, обсуждают интересующие их темы и общаются с семьей и друзьями. Для выяснения того, можно ли считать социальные сети интересным источником данных для статистики, Статистическое управление Нидерландов исследовало сообщения из нидерландских социальных сетей с двух точек зрения: с точки зрения их содержания и с точки зрения их эмоциональной окраски. Исследование содержания сообщений из нидерландского "Твиттера" (главная общедоступная социальная сеть для обмена сообщениями Нидерландов на момент проведения исследования) показало, что почти 50% сообщений состоят из "беспредметной болтовни". В остальных сообщениях обсуждаются главным образом занятия в свободное время (10%), работа (7%), СМИ (телевидение и радио – 5%) и политика (3%). Использование этих более серьезных сообщений ограничивалось менее серьезной "болтовней". Последнее также негативно повлияло на подходы, основанные на глубинном анализе текста. Определение эмоциональной окраски сообщений из социальных сетей позволило выявить одно крайне интересное потенциальное направление использования этого источника данных в целях статистики. Оказалось, что эмоциональная окраска сообщений в нидерландских социальных сетях сильно коррелируется с уровнем потребительского доверия в Нидерландах, в частности с оценкой экономической ситуации. Последняя зависимость оставалась стабильной на ежемесячной и еженедельной основе. Однако ежедневные данные характеризовались высокой волатильностью. Это означает возможность расчета еженедельных показателей потребительского доверия, а также их расчета в первый рабочий день после исследуемой недели, что свидетельствует о способности выдавать быстрые результаты.

28. Ниже перечислены исследования, запланированные к проведению в рамках программы работы Евростата, а также исследования по изучению потенциала использования больших данных в целях официальной статистики.

29. **Статистика цен:** использование и анализ данных о ценах, собранных в Интернете. Этот одногодичный проект, запущенный в январе 2013 года, направлен на разработку передового "очищающего" программного обеспечения с открытым исходным кодом, призванного помочь специалистам, занимающимся индексом потребительских цен (ИПЦ), в автоматическом сборе информации о ценах в Интернете. Он обладает схожими чертами с проектом "Миллиард цен" Массачусетского технологического института (МТИ) и будет тестироваться в пяти странах с целью решения технологических и методологических проблем. Программное обеспечение, которое будет разработано в рамках этого проекта, будет по запросу предоставляться "как есть" другим статистическим организациям в рамках лицензирования открытого ПО на основе Общественной лицензии Европейского союза (ОЛЕС).

30. **Статистика туризма:** исследование по изучению возможностей использования данных мобильных устройств о местоположении в целях статистики туризма. Рассчитанный на 15 месяцев проект должен начаться в январе 2013 года. В рамках этого исследования будет изучаться полезность применения данных мобильных устройств о местоположении для целей статистики туризма (и связанных с ней областей) и будет проведена оценка преимуществ и недостатков. Это исследование будет охватывать такие вопросы, как доступ (и непрерывность доступа), доверие (производителей и пользователей статистики), расходы, концепции (в контексте преобразования существующей концепции статистики туризма применительно к новому источнику данных), а также другие методологические темы (например, репрезентативность, формирование выборки из очень большого количества наблюдений). Способность обрабатывать

большие файлы данных, имеющиеся у мобильных операторов, считается важнейшим препятствием, которое необходимо преодолеть для успешной реализации этого проекта. Включение этого проекта в рабочую программу обусловлено, среди прочего, обнадеживающими результатами исследований, которые были получены в ряде стран.

31. **Статистика использования информационно-коммуникационных технологий (ИКТ):** исследование по изучению возможностей использования потоков Интернет-трафика для сбора статистических данных об информационном обществе. В рамках этого проекта Евростат планирует проверить и оценить практические возможности использования подходов к измерению, "ориентированных на пользователей" и "ориентированных на Сеть", с множественных точек зрения, включая технические, методологические, ценовые, правовые и социально-политические аспекты. В числе результирующих материалов этого проекта, которые могут быть полезными для национальных и международных статистических агентств, необходимо отметить три запланированных доклада: i) "Как разработать процедуру аккредитации"; ii) "Руководство по методологии и реализации процесса для НСО"; и iii) «Тестирование концепции "интегрированных открытых данных"». Эта концепция "интегрированных открытых данных" представляет собой аналог так называемых "открытых данных" правительств (или дополнение к ним). Речь идет о совместно используемом поднаборе больших данных субъектов частного сектора, которые будут "открыты" для использования НСО.

32. Хотя НСО находятся на начальном этапе изучения потенциала больших данных для целей официальной статистики, первые результаты указывают на три основные области для проведения экспериментов:

- a) комбинирование больших данных с официальной статистикой;
- b) замещение официальной статистики большими данными;
- c) заполнение новых пробелов в данных, т.е. разработка новых "основанных на больших данных" показателей для изучения новых явлений (которые ранее не были известны или к которым нельзя применять традиционные подходы).

33. Потенциал комбинирования больших данных с официальной статистикой обладает некоторыми схожими чертами с тем, что делалось в последние несколько десятилетий в области использования административных данных в целях официальной статистики. Однако, как представляется, небольшое, но потенциально привлекательное отличие заключается в возможности более широкого применения статистического моделирования для комбинирования этих двух видов данных. Благодаря этому можно получить такие оценки, которые могут сохранять высокие качественные характеристики официальной статистики и улучшать их за счет полученных на основе больших данных измерений в режиме, близком к реальному времени.

## VI. Выводы/рекомендации

34. В этом разделе обсуждаются сделанные выводы и предлагаются рекомендации в качестве возможных последующих мер. **Очевидно, что в течение двух последующих лет необходимо будет определить небольшое число пилотных проектов для тестирования концепции** (подобно тем, которые обсуждались

в разделе V) с участием нескольких стран-партнеров. Полученные результаты затем можно будет представить ГВУ.

35. Использование больших данных выдвигает ряд серьезных вызовов и задач для национальных и международных статистических организаций, которые касаются главным образом методологических, технологических, управленческих, правовых и профессиональных проблем. **Исходя из этого статистическим организациям рекомендуется заняться в официальном порядке вопросами больших данных в рамках своих ежегодных и многолетних рабочих программ путем проведения исследований и пилотных проектов в определенных областях и выделения надлежащих ресурсов для этих целей.**

36. Использование громадных массивов данных – нелегкая задача. Уже сам размер больших данных затрудняет их изучение и обеспечение качества. Этап исследования больших данных может быть намного более продолжительным, чем в случае других, часто более структурированных источников данных большого объема. В связи с этим **возникает потребность в "новых" методах исследования и анализа.** Слово "новые" здесь закавычено по той причине, что многие из этих методов уже существуют и используются, но являются новыми для сферы официальной статистики. Три таких метода были отмечены в качестве особо полезных: *методы визуализации, интеллектуальный анализ текстов и высокопроизводительные вычисления.*

37. Хотя ограниченное количество НСО активно занимаются технологическими аспектами больших данных, основную работу по аналитическим инструментам и решениям, связанным с большими данными, проводит главным образом частный сектор. Для адаптации связанных с большими данными аналитических инструментов и систем к официальной статистике, безусловно, потребуется задействовать НСО. **Международное статистическое сообщество следует ознакомить с успешными примерами их использования.**

38. Синергизм между НСО и частным сектором не ограничивается лишь технологическими вопросами. **Взаимодействие НСО с частными владельцами источников данных имеет крайне важное значение** и связано с такими деликатными вопросами, как конфиденциальность, доверие и корпоративная конкурентоспособность, а также с нормативно-правовой основой деятельности НСО. **Сбор примеров национальной практики в этой области, иллюстрирующих некоторые вопросы предоставления НСО привилегированного доступа к частным источникам больших данных, должен рассматриваться в качестве одного из приоритетов.**

39. Для использования больших данных статистикам необходимы новые навыки и иной склад ума. Обработка все больших и больших объемов данных для целей официальной статистики требует статистически грамотных людей с аналитическим складом ума и хорошими познаниями в области информационных технологий (например, навыки программирования) и решимостью извлекать ценные "знания" из этих данных. Этим так называемых "ученых по данным" можно найти в различных научных дисциплинах. Нидерланды имеют хороший опыт использования для этих целей ученых (имеющих степень кандидата наук), работающих в таких областях, как математика, физика, (био)химия и экономика/эконометрия, и обладающих хорошими познаниями в области информационных технологий.

40. Если в долгосрочной перспективе вопросы приобретения профессиональных навыков, необходимых статистикам по большим данным и в целом "ученым по большим данным", будут решаться посредством соответствующей

адаптации учебных программ в университетах (некоторые университеты уже начали предлагать соответствующие курсы), то в краткосрочной и среднесрочной перспективах НСО следует создать **необходимый внутренний аналитический потенциал** путем организации специализированной подготовки. **Международное сотрудничество в этом отношении способно принести большую пользу сообществу официальной статистики.**

41. ГВУ также следует в среднесрочной перспективе подумать о разработке руководящих указаний/принципов эффективного использования больших данных в целях официальной статистики. Необходимо также подумать о новой роли, которую могли бы играть НСО в будущем, с точки зрения классификации и сертификации статистики, получаемой на основе больших данных и используемой в целях государственной политики. Вследствие проникновения больших данных в многочисленные аспекты жизни все более насущной становится потребность в наличии пользующейся доверием третьей стороны. Должны ли НСО выполнять эту роль в одиночку или же в качестве членов какого-то независимого междисциплинарного органа?

42. В рамках сообщества официальной статистики существует ряд различных инициатив, связанных с большими данными. К ним относятся: вынесенное на сорок четвертой сессии Статистической комиссии Организации Объединенных Наций предложение о проведении по этой теме осенью 2013 года международной конференции, проведение специальных заседаний на Всемирном статистическом конгрессе Международного статистического института в августе 2013 года, обсуждение этих вопросов на запланированном на осень 2013 года заседании "ГДНСИ" для глав организаций, включающих Европейскую статистическую систему, а также проведение рабочего совещания по вопросам практического применения больших данных по предложению Евростата/ЕЭК ООН и в конце 2013 года или начале 2014 года. Запланировано также несколько мероприятий по вопросам использования геопространственных данных (которые часто рассматриваются в качестве больших данных) и их интеграции в официальную статистику. **ГВУ следует обеспечить, чтобы результаты этих и схожих мероприятий эффективно координировались и доводились до сведения на стратегическом уровне.**

---