



Европейская экономическая комиссия

Конференция европейских статистиков

Шестьдесят пятая пленарная сессия

Женева, 19–21 июня 2017 года

Пункт 4 предварительной повестки дня

Следующее поколение статистиков и ученых по данным

Навыки для нового поколения статистиков

Записка Статистического управления Финляндия и Евростата

Резюме

В настоящем документе анализируется профиль компетентности официальных статистиков с особым упором на новые умения, предусматриваемые наукой о данных. Модернизация официальной статистики будет зависеть от способности инкорпорировать новые источники данных и пользоваться «прорывными технологиями». Это потребует новых способностей, навыков и умений, которые могут и не быть частью набора традиционных навыков официальных статистиков.

Документ представляется семинару Конференции европейских статистиков на тему «Следующее поколение статистиков и ученых по данным» для обсуждения.



I. Введение

1. В основе модернизации официальной статистики будет лежать способность инкорпорировать новые источники данных и пользоваться «прорывными технологиями». Эти технологии, например умные счетчики, веб-технологии и платформы пользовательского опыта, потребуют новых видов навыков и умений, которые не были частью набора традиционных навыков официальных статистиков.

2. Существуют многочисленные примеры новых источников данных, которые имеют потенциал в этом смысле: административные отчеты и реестры, а также крупные источники цифровых данных, таких как дорожные датчики, данные сканирования или данные на базе Интернета. Эти крупные источники цифровых данных лучше известны в качестве «больших данных». Потенциальные новые технологии, с другой стороны, варьируются от алгоритмов веб-скрининга до возможностей использования связанных данных и от многодогового сбора данных до комбинированного использования данных обследований и административных данных. Этот список можно еще дополнительно продолжить.

3. За счет освоения новых источников данных и технологий национальные статистические управления (НСУ) могут быстрее подготавливать более точные и более полные статистические данные, адаптированные к пониманию более сложных и быстро меняющихся глобальных явлений. Короче говоря, новые источники данных служат одним из способов более эффективного удовлетворения потребностей пользователей. Это должно быть сделано таким образом, чтобы не поставить под угрозу признанную надежность и качество официальной статистики, но это укрепит наше конкурентное преимущество в быстро меняющейся информационной экосистеме.

4. Одним из ключевых факторов решения этих проблем является развитие и укрепление необходимых навыков и умений. Кроме того, статистическим организациям необходимо будет создать благоприятные условия для применения новых производственных методов и использования навыков, связанных с наукой о данных, на успешной основе. Это предполагает, например:

- формирование инновационной культуры, где распространены экспериментальные мероприятия;
- создание и поддержание действующих по принципу сотрудничества и носящих междисциплинарный характер даталогических групп;
- наем на работу отдельных ученых по данным и использование долгосрочного кадрового планирования;
- профессиональную подготовку и поддержку персонала в деле усвоения ими новых потребностей, касающихся компетентности;
- приверженность таким методам управления и руководства, которые делают возможными эти изменения.

5. Существуют многочисленные движущие факторы для освоения потенциала новых источников данных и технологий:

- во-первых, НСУ должны адаптироваться к меняющемуся цифровому миру и внедрять инновации, чтобы оставаться конкурентоспособными;
- во-вторых, сопутствующие расходы обследований и увеличение коэффициентов непредоставления ответов ставят статистические организации перед насущной необходимостью адаптировать свое делопроизводство и уже не опираться на «поточную» производственную систему;
- и последним, но не менее важным моментом является то, что новые данные и технологии открывают возможность для лучшей работы.

6. Вопросы статистической информации для принятия решений и проверки фактов. Статистика по сути способствует публичным дебатам и позволяет отдельным лицам, домашним хозяйствам, предприятиям и представителям директивных органов полагаться на что-то надежное. Поэтому важно постоянно совершенствовать методику, с помощью которой мы готовим статистические данные. Использование новых источников данных и технологий затрагивает весь производственный процесс. В начале процесса статистические организации могут продвигаться в направлении многомодового сбора данных (например, сочетания телефонных опросов, веб-обследования и датчиков) и использования множественных источников данных (например, «больших данных» и административных источников).

7. Это, с другой стороны, требует применения методов и умений для увязки, редактирования и условного расчета данных. В эпоху новых источников данных, представленных в цифровой форме, НСУ должны обладать способностью для эффективного использования огромных объемов неструктурированных данных. Примеры новых статистических методологий, которые подлежат изучению, включают коррекцию избирательности в «больших данных», распознавание образов, снижение размерности и обработку высокочастотных данных.

8. Использование новых источников данных и связанных с ними технологий будет влиять на широкий диапазон возможностей статистической организации. Это не только потребует обновления производственных систем. Статистические организации смогут предоставлять пользователям в большем объеме более адресные услуги по анализу данных, включая составление наборов показателей, конкретно ориентированных на потребности пользователей, в целях контекстуализации информации или обеспечения доступа к инструментам извлечения данных, которые позволяют пользователю получить максимальную отдачу от наборов статистических данных.

9. Вопрос состоит не только в том, как готовить статистические данные, но и в том, какого рода конечный продукт необходим. Такое развитие в сфере официальной статистической деятельности потребует создания информационно-технологической (ИТ) среды, способствующей интеграции различных источников данных (виртуализация данных). Последствия применения новых источников данных в рамках статистической организации можно наглядно проследить с использованием модели деловых возможностей Европейской статистической системы (ЕСС). Модель описывает высокоуровневые функции статистической организации и позволяет применить систематический и структурированный подход к модернизации потенциала в области официальной статистики (см. диаграмму 1).

Диаграмма 1

Новые источники данных – влияние на основные возможности и навыки**Новые источники данных: воздействие на основные возможности и навыки**

10. В ней определяются функции статистических организаций, которые должны быть расширены или воссозданы. Каждое увеличение потенциала будет обеспечиваться за счет надлежащего объединения кадровых ресурсов, бизнес-процессов, методов, инструментов информационной технологии (ИТ) и информации.

11. Чтобы расширить эти возможности, НСУ придется пересмотреть навыки и умения сотрудников. Модернизация требует не только усовершенствовать конкретные профили, а 4 группы профилей: методы и статистическую обработку; управление информацией; управление ИТ; и человеческий потенциал (руководство, творческий подход и коммуникация).

12. На диаграмме 1 приводится описание профиля профессиональных навыков, связанных с каждой возможностью, на которую влияет использование новых источников данных. В прошлом эти профили, как правило, были отделены друг от друга, и это обычно было отражено в структуре организации. В настоящее время одна из задач заключается в том, что эти профили должны быть объединены синергичным образом в рамках небольших групп, приспособленных к динамичному развитию. Эта эволюция проливает свет на новый тип ключевых функций в области официальной статистики, которые возлагаются на статистика или ученого по данным, чей профиль можно в общих чертах определить по линии пересечения этих 4 традиционных профилей. Следовательно, статистические управления должны рассмотреть вопросы о том: 1) как набрать персонал с необходимыми профилями компетентности; 2) как модернизация влияет на методы работы и организационную культуру в статистических управлениях; и 3) как практика управления и руководства в статистических управлениях может способствовать предстоящим изменениям.

13. В следующей главе анализируются основные требования, касающиеся использования новых источников данных и технологий. После этого в документе будут обсуждаться необходимые навыки и умения будущего статистика в статистических организациях. И наконец, в документе будут предложены пути приобретения этих навыков и умений. В заключительной части документа бу-

дут сформулированы выводы и рекомендации относительно дальнейших действий.

II. Проблемы освоения новых источников данных и технологий

14. Новые источники данных влияют не только на этап сбора данных, но и на весь процесс подготовки официальной статистики.

15. Получение доступа к источникам «больших данных» и административным отчетам требует учета правовых и этических аспектов, обеспечения конфиденциальности и налаживания стабильного процесса предоставления данных с поставщиками данных. В настоящем документе мы будем рассматривать доступ к новым источникам данных как объективную действительность.

16. Для того чтобы иметь возможность использовать новые данные, необходимо быть в состоянии решать следующие методологические проблемы:

- увязка данных из различных источников (обследования, административные отчеты, регистры и «большие данные») и методов (телефон и веб-вопросники);
- редактирование и условный расчет данных со значительным объемом и скоростью;
- обработка неструктурированных данных и выделение признаков;
- выбор соответствующей информации из огромных объемов данных (большая размерность, высокая частотность или и то и другое);
- решение вопросов расчетного времени (возможные алгоритмы, выбор программного обеспечения);
- вопросы выборности «больших данных» (каким образом обеспечить репрезентативность и охват «больших данных»);
- распознавание образов;
- интеграция множественных источников;
- оценка качества.

17. Чаще всего для успешного решения этих вопросов потребуется использовать технику моделирования и методы, позволяющие преодолеть нелинейность. Среди них можно отметить методы искусственной нейронной сети и другие методы машинного обучения, байесовские подходы, в основе которых лежат методы имитации и бутстрепинга, а также анализ сетевых данных. Применение таких методов подразумевает освоение алгоритмов оптимизации, а также навыков креативного и инновационного моделирования. Аналогичным образом, интеграция множественных источников потребует также расширения использования статистики, основанной на модели. И наконец, в целях поддержания признанных стандартов качества официальной статистики необходимо будет разрабатывать индикаторные методы для оценки качества выходных статистических данных.

18. В конце производственного процесса, при выпуске новых статистических продуктов и проведении индивидуализированного анализа данных, должны быть учтены следующие вопросы:

- описание многоаспектных явлений, которые перекрывают социальные, экономические и экологические области;
- интеграция качественных доменных знаний в рамках аналитической среды;

- контекстуализация статистической информации на основе селективного статистического анализа (выявление соответствующих поднаборов данных);
- представление и толкование новых показателей, вытекающих из методов, используемых для подготовки данных;
- статистическая оценка наборов показателей;
- использование новых данных в целях поощрения разработки конкретно заданной и индивидуализированной статистики;
- визуализации данных и способы дополнения традиционных методов распространения информации инфографикой, видеоматериалами и другими изображениями.

19. Доменные экспертные знания и статистические характеристики конкретных наборов показателей будут влиять на то, каким образом можно структурировать информацию для пользователей. Статистической информации должны быть присущи контекстуальная информация и структура, которая позволяет гибко подходить к этапу распространения данных. Кроме того, при использовании методов моделирования или байесовских подходов могут генерироваться, например, некоторые новые категории информации, в частности информация, касающаяся полных функций плотности или состояния наблюдаемой системы. Благодаря такой информации пользователи могут получить в свое распоряжение количественные сценарии, адаптированные для толкования и прогнозирования более сложных и неопределенных явлений.

III. Профиль компетентности будущего статистика

20. Регулирование компетентности играет ключевую роль в решении будущих задач официальной статистики. В этой главе будут рассмотрены следующие два главных вопроса: 1) какие умения, навыки и отличительные черты мы хотим видеть у будущего статистика; и 2) как могут быть приобретены эти навыки и умения?

A. Навыки и умения будущего статистика

21. С учетом ожидаемых изменений в статистическом производстве, о которых говорится в настоящем документе, потребуются по крайней мере умения и навыки, перечисленные в нижеследующей таблице. Этот перечень не является исчерпывающим, а скорее указывает на некоторые примеры таких областей компетентности.

Таблица 1
Навыки и умения для будущих статистиков

<p>Математика и статистика</p> <ul style="list-style-type: none"> • статистическое моделирование • байесовский подход • машинное обучение • выявление и использование множественных источников данных • способность различать сигнал от шума • потенциал для решения вопросов, связанных с массовыми неструктурированными данными • методы увязки, редактирования и условного расчета данных 	<p>Доменные знания и профессиональные качества</p> <ul style="list-style-type: none"> • увлеченность изучением явления/существа дела • стремление удовлетворить потребности клиентов • интерес к данным • решение проблем • коллективная работа • инициативность, склонность к сотрудничеству, новаторству и творческой деятельности
<p>Программирование и базы данных</p> <ul style="list-style-type: none"> • основы информатики • скриптовый язык • языки для статистических вычислений и графики • ЯСЗ • концепции «Map Reduce» • «Hadoop» 	<p>Коммуникация и визуализация</p> <ul style="list-style-type: none"> • сотрудничество с заинтересованными сторонами • навыки ведения переговоров • повествовательные навыки • оформление и дизайн • инструменты визуализации

22. Некоторые из перечисленных выше функций присущи традиционной статистической деятельности на протяжении многих лет. Таким образом, наука о данных влияет на вес различных областей компетентности в профиле компетентности будущего статистика.

23. Наличие наборов новых цифровых массовых данных изменяет традиционную парадигму статистического производства. В ближайшем будущем статистическое производство будет чаще начинаться с того, что ученые по данным станут собирать и анализировать наборы неструктурированных данных и продумывать потенциальные пути их использования. Такая последовательность действий противоположна традиционному подходу, когда сначала определяются потребности, а затем происходит сбор данных. Поэтому статистикам необходимо обладать способностью применять гибкий и инициативный подход в зависимости от данных для обнаружения информационного содержания новых источников данных. В этом контексте использование концепции «хакерского склада ума» наглядно показывает, каким образом статистическая работа требует повторной оценки проблем и непрерывного изучения пределов того, что является возможным.

24. Еще одним примером смещения акцентов в плане компетенции служат повествовательные навыки. Будущие статистики должны иметь способность представлять свои цифры в определенном контексте и распространять статистическую информацию в том или ином визуальном формате. Следовательно, речь идет уже не совсем о том, чтобы просто давать числа и предоставлять исследователям данные для их дальнейшего анализа. Будущий статистик должен быть в состоянии обеспечивать добавочный результат путем анализа, оценки и контекстуальной привязки сообщаемых показателей. В качестве подспорья для этого более важное значение будут также иметь навыки оформления и дизайна и использования различных инструментов визуализации.

25. Статистики в статистических организациях наиболее часто обладают способностью к познанию и применению методов науки о данных на экспериментальной основе. Вместе с тем для решения новых задач в области статистического производства также потребуются следующие навыки:

- способность объединять информацию из различных источников в целях подготовки новых информационных продуктов. Это требует существенных знаний о явлениях, которые должны быть описаны, и хорошего понимания потребностей клиентов;
- эффективное и действенное коммуникационное взаимодействие с другими заинтересованными сторонами в целях разработки и опробования новых методов, в частности в контексте сотрудничества с различными доменными экспертами и академическими учреждениями;
- навыки программирования для создания робюстного и масштабируемого инструментария обработки и анализа данных и использования скриптовых статистических языков, которые позволяют применять новые методы.

В. Средства для приобретения услуг квалифицированных статистиков и поддержки работы, движимой наукой о данных

26. Наука о данных требует многочисленных навыков, которые бывают редко воплощены в одном работнике. Таким образом, подлинная наука о данных начинает процветать в рамках построенных на сотрудничестве и носящих многодисциплинарный характер даталогических групп, которые состоят из экспертов с различными профилями компетентности и специализированными знаниями. Чтобы эти даталогические группы были действенными, им следует обеспечить возможность работать по горизонтали при наличии определенного уровня гибкости в разрезе структуры организации и ее руководства. Однако обязанности и функции групп должны быть совместно согласованными и понятными. Это относится также к управленческой деятельности: распределение обязанностей и функций между директорским эшелоном, начальниками и руководителями групп должно быть четко прописано и реализовано на практике. Таким образом, расширение коллективной составляющей в работе статистических управлений ставит перед организациями задачу поиска путей создания и поддержания групп, которые охватывают различные звенья деятельности организации, используя передовую практику и впитывая опыт друг у друга.

27. Хотя даталогические группы, состоящие из различных экспертов, будут, вероятно, одним из наиболее типичных способов организации работы, движимой наукой о данных, может возникнуть необходимость в найме на работу отдельных ученых по данным для выполнения, например, следующих задач:

- осуществление планов, направленных на расширение использования новых цифровых массовых данных в статистическом производстве;
- выявление новых источников данных и оценка их потенциала;
- разработка процессов, методов и ИТ-решений;
- профессиональная подготовка и консультирование, обмен опытом по вопросам, связанным с наукой о данных.

28. Возможности найма на работу ученых по данным, подобно всем видам деятельности по набору кадров, которые ассоциируются с брендом статистической организации, имеют важное значение и должны быть привлекательными для будущих сотрудников. Еще более важно то, что при найме отдельных ученых или экспертов по данным в даталогические группы организациям необходимо иметь реалистичное понимание задач и профилей компетентности таких сотрудников. Долгосрочное кадровое планирование помогает выяснить потребность в экспертах-датолагах и штат будущих должностей. На ученых по дан-

ным будет большой спрос, и их удержание будет нелегким делом. Поэтому необходимо обеспечить, чтобы задачи, возложенные на ученого по данным, соответствовали его/ее компетенции. Удержать ученых по данным может быть сложно в любом случае, особенно если они принимаются на работу на должности, на которых они не смогут использовать весь свой потенциал.

29. Одним из путей подбора ученых по данным, который частично связан с наймом персонала, является сотрудничество с университетами. Статистические управления могут брать стажеров или дипломников, имеющих знания в области информатики. Предоставление учащимся доступа к реальным наборам данных и обеспечение им возможностей практической работы позволяет статистическим управлениям пропагандировать себя в качестве интересного места для работы. Кроме того, сотрудничество с университетами помогает понять потенциал науки о данных и навыков, которые имеют отношение к решению проблем статистического производства.

30. Если рассматривать различные альтернативы для приобретения услуг компетентных специалистов за пределами организации, то найм на работу является не единственным вариантом. Статистическим управлениям не обязательно иметь у себя в штате специалистов, обладающих необходимыми умениями и навыками. Они также могут пользоваться их компетентностью, полагаясь на их сети или отдавая решение некоторых задач на внешний подряд. Поэтому статистическому сообществу будет необходимо налаживать эффективные партнерские связи с исследовательскими и даталогическими сетями, а также с соответствующими экспертами и пользователями статистических данных. Сотрудничество в идеале должно открывать новые возможности для того, чтобы статистические организации поддерживали свою компетентную базу на современном уровне.

31. Одним из наиболее традиционных способов развития компетентности является профессиональная подготовка. После того как организации определяют потребности и существующие пробелы в сфере компетентности, они могут по внутренним или внешним каналам организовать учебные курсы, практикумы и программы для устранения конкретных пробелов в сфере компетентности. Однако одной лишь профессиональной подготовки недостаточно для приобретения компетентности; это, скорее, дополнительный инструмент. Также важно помнить о том, что нынешний персонал, возможно, уже имеет реальную заинтересованность и потенциал в решении вопросов, связанных с наукой о данных. Этот потенциал может проявиться только в тех случаях, когда предоставляется возможность ознакомления с новыми наборами данных и опробования новых методов. В целях поддержки способностей персонала организациям следует экспериментировать с новыми наборами данных и выделять определенное время для этого. Опять-таки, как и при укреплении группового взаимодействия, здесь необходимы решения, определение приоритетов и поддержка со стороны директоров и руководителей.

32. Очень часто необходимые изменения, такие как учреждение групп по всей организации или создание необходимых управленческих или рабочих условий, идут вразрез с нынешними «поточными» подходами, используемыми в статистических управлениях. Таким образом, эти изменения требуют реализации способности и готовности сделать иначе, мыслить творческим и новаторским образом и предпринять смелые шаги вперед. Например, использование новых наборов данных требует хорошего уровня воображения; зачастую сотрудник не может определить полезность набора данных до проведения его анализа, и необходимо создать условия, которые позволяют применять творческий подход и разрабатывать новые информационные продукты. Статистические организации должны стремиться к инновационной культуре, при которой экспериментальные мероприятия становятся обычным явлением и допущение неудачи является неотъемлемой частью культуры.

33. Экспериментальная культура ставит вызовы настоящему, тому, каким образом подготавливались статистические данные в течение длительного време-

ни. Профессиональный облик статистиков зиждется на подготовке объективной и надежной статистики хорошего качества. Традиция имеет большое значение для статистиков, которые привыкли работать в качестве «памяти нации». Это остается важной задачей даже тогда, когда вызовы той среды, в которой мы действуем, требуют от статистиков перемен. Заботясь о производстве высококачественных статистических данных, мы должны также позаботиться о профессиональных статистиках. Как статистиков нас также необходимо поощрять и поддерживать в приспособлении к новым умениям.

34. Статистическое сообщество уже рассматривает новые пути и методы для подготовки статистических данных. Это должно быть связано с тем, какие необходимы умения и как они могут быть приобретены. Факт состоит в том, что все мы являемся официальными статистиками, чья первейшая неизменная задача состоит в том, чтобы сделать хорошие статистические данные еще более качественными. Кроме того, вопрос заключается не в том, как статистик может стать ученым по данным, а как статистическая организация может интегрировать «науку о данных» в свой набор инструментальных средств и как необходимые умения и навыки могут быть воплощены в профессиональном облике официальных статистиков.

IV. Выводы и рекомендации

35. В этом документе излагается несколько идей на будущее, с тем чтобы адаптировать статистические организации к эпохе цифровых технологий. Ниже подытоживаются основные моменты и рекомендации:

- использование новых источников данных и разработка нового инструментария анализа данных в сфере официальной статистики является одной из ключевых задач для сохранения конкурентоспособности и обеспечения того, чтобы хорошие статистические данные были еще более качественными;
- новые методы и изменения в системе подготовки статистической информации влияют на потребности в компетентности персонала; решающее значение имеет развитие навыков, связанных с наукой о данных;
- сотрудников статистических управлений следует поощрять и поддерживать в деле усвоения ими новых требований компетентности и новых методов работы;
- необходимые умения и навыки могут быть приобретены, например, путем подготовки нынешнего персонала, найма на работу отдельных ученых по данным и сотрудничества с университетами и другими сетями;
- совершенствование навыков, связанных с наукой о данных, является необходимым, но также важно изменить методы работы и организационную культуру статистических управлений;
- будущие методы работы должны больше опираться на сотрудничество и должны носить более горизонтальный и междисциплинарный характер; работа, движимая наукой о данных, должна быть построена через посредство даталогических групп, охватывающих различные звенья функционирования организации, чтобы задействовать необходимые навыки;
- статистические управления должны впитать инновационную культуру, где экспериментальные мероприятия являются обычным явлением; это имеет важное значение, поскольку будущая статистическая работа требует творческого мышления, а также наличия способности и готовности использовать иной подход;
- управление и руководство играют решающую роль в реализации предстоящих изменений, которые требуют, в частности, долгосрочного кадрового планирования и налаживания и поддержания групповой работы, при

обеспечении достаточного времени и пространства для проведения экспериментальных мероприятий и построенных на участии механизмов на пути формирования общего будущего в лоне статистических организаций.

V. Библиография

C. F. Jeff Wu: *Statistics=Data Science?* University of Michigan, Ann Arbor. www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf

Methodology architecture: A roadmap for new methodological directions in the Australian bureau of statistics. Statistical Journal of the IAOS, vol. 30, no. 4, pp. 371-375, 2014

ONS Strategy 2013–2023, Trusted Statistics – Understanding the UK. <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/what-we-do/corporate-documentation/strategies-and-policies/ons-strategy-2013-23/index.html>
