

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

Париж, 26–28 июня 2019 года

Пункт 6 а) предварительной повестки дня

**Новая роль национальных статистических управлений
в качестве служб статистики и данных****Заседание 1: новые возможности и вопросы систем данных****Новая функция при прежней роли? Последствия
датафикации для официальной статистики****Записка Евростата***Резюме*

В настоящем документе обсуждаются изменяющиеся условия доступа к данным, их обработки и распространения, а также их влияние на функционирование статистической системы. Авторы предполагают, что задача официальной статистики сохраняется, но средства и условия, при которых эта задача будет выполняться, изменятся. Это потребует системного подхода, при котором статистические управления должны будут переосмыслить свою модель производства и распространения данных.

Авторы предполагают, что статистические управления не смогут централизовать все данные и производство статистики. В силу этого задачи статистических управлений можно было бы расширить и распространить на предоставление рекомендаций по использованию, обработке и передаче статистических данных. В рамках государственных структур статистические управления могут взять на себя ведущую роль в качестве хранителей данных, используемых для статистических целей. Кроме того, в настоящем документе обсуждается идея сертификации данных, процессов и результатов, а также их последствия для статистических управлений. Растущий спрос на высококачественные статистические данные, которые удовлетворяют конкретные потребности и распространяются своевременно, требует активного и гибкого подхода к информированию различных групп общества в ходе широких обсуждений.

Настоящий документ представлен для обсуждения на семинаре Конференции европейских статистиков 2019 года на тему «Новая роль национальных статистических управлений в качестве служб статистики и данных», заседание 1 «Новые возможности и вопросы систем данных».



I. Введение

1. С момента появления Интернета в последнем десятилетии прошлого века технические достижения в области информационных и коммуникационных технологий положили начало глобальному процессу цифровизации всего общества. Ключевыми благоприятными факторами стали распространение Интернета и «всемирной паутины» в качестве общей информационной и прикладной платформы. С развитием мобильных телефонов доступ к Интернету вышел за рамки ПК, подключенных к проводной связи, и веб-услуги стали доступны любому человеку в любое время и в любом месте.

2. Миниатюризация этих устройств и насыщение их данными сделали их универсальными устройствами, которые сопровождают людей круглосуточно и без выходных. В то же время были разработаны услуги и платформы, которые заменили физическую активность или объекты их цифровыми аналогами. В результате объединения в общую сеть стали формироваться монополии для конкретных приложений, таких как поиск или сетевое взаимодействие. В последнее время все больше машин и приборов оснащаются информационными технологиями (ИТ), чтобы сделать их более интеллектуальными. Эти «умные устройства» способны общаться друг с другом, создавая так называемый «Интернет вещей». Для повышения их интеллектуальности в эти устройства или в системы, создаваемые сетью таких устройств, все чаще внедряется искусственный интеллект. Конечной целью этого является создание для пользователей интеллектуальных устройств или систем вместо товара той или иной услуги: например, вместо продажи автомобиля компании предлагают средства, обеспечивающие мобильность.

II. Датафикация как новая парадигма

3. Цифровизация – это условие для передачи данных, которое преобразует каждое действие и состояние в данные. Такие данные становятся объектом сбора, хранения, обработки, анализа и обмена. Масштабы объема, размерности, частоты, плотности и разнообразия этих данных на много порядков выше, чем масштабы сбора данных в доцифровую эпоху. Действительно, массивы данных, которые в прошлом собирались на основе дорогостоящих мероприятий, стали побочным продуктом датафикации. Сбор данных и предоставление информации больше не являются квазимонопольной деятельностью конкретных организаций, таких как статистические управления, а могут осуществляться любой организацией, обладающей необходимыми специалистами для преобразования данных в информацию.

III. Изменение привычек и ожиданий пользователей

4. В дополнение к данным в качестве конечного продукта в рамках соответствующей бизнес-модели сбор данных осуществляется с помощью интернет-платформ. В обмен на предложение услуг или других стимулов пользователи делятся личной и иной информацией, которая повторно используется как для разработки новых услуг, так и для получения дохода, например посредством рекламы, с тем чтобы предоставлять населению услуги платформ бесплатно. В дополнение к прямым услугам для пользователей интернет-платформы и приложения используют дополнительные стимулы для привлечения пользователей к своим услугам. Это может быть сделано на основе игрофикации, в рамках которой предлагаются типичные элементы игры, такие как начисление очков или конкуренция с другими пользователями. Еще одним примером являются схемы общественных наград, в рамках которых поощряется создание сообществ для достижения общей цели, что предусматривает вознаграждение как отдельных лиц, так и сообщества в целом. Примером может служить поощрение энергосберегающих мероприятий в пределах микрорайона, которые, помимо прочего, могут включать подходы на основе игрофикации. Еще одним стимулом для участия в предоставлении цифровых услуг и данных может стать общественное признание.

5. Для получения данных от граждан также используются финансовые или псевдофинансовые стимулы. Одним из примеров псевдофинансовых стимулов является начисление баллов на потребительские карты, которые можно использовать для покупки товаров. Иногда достаточно пообещать какие-то призы, чтобы убедить граждан поделиться своими личными данными. Граждане одновременно являются потребителями услуг и производителями данных. Они все больше осознают наличие таких механизмов и обмениваются своими данными для получения личных услуг или других бонусов. Отношение граждан, оказавшихся в схожих ситуациях в различных контекстах, таких как предоставление данных для статистических управлений, представляющих общественное благо, изменится в связи с описанными механизмами, с которыми они сталкиваются на повседневной основе.

6. В плане предоставления услуг и стимулирования граждан к тому, чтобы поделиться личными данными, важным фактором является персонализация. В качестве примера можно привести передачу данных индивидуальными устройствами слежения за состоянием здоровья платформе поставщика, когда сбор данных и снабжение ими осуществляются в фоновом режиме с использованием автоматической передачи данных. В качестве компенсации пользователи получают информацию о состоянии своего здоровья, персонализированные рекомендации в отношении поведения и могут сравнивать себя с другими пользователями или создавать группы, например для совместного занятия спортом.

7. Широкая доступность информации и услуг изменила ожидания пользователей. Информация и услуги должны быть доступны точно в срок, круглосуточно и без перерывов. С помощью поисковых систем и онлайн-энциклопедий информация доступна в момент поступления запроса. Любой товар можно приобрести в любое время суток с доставкой вскоре после оплаты. Объем информации корректируется в соответствии со спросом. Доставка обеспечивается в целом незамедлительно.

8. Автоматическая обработка данных обеспечивает возможность получения информации практически в режиме реального времени. Собранные данные могут быть обработаны, проанализированы и агрегированы за очень короткое время. Время между сбором данных и предоставлением информации может быть резко сокращено; в некоторых случаях может быть предложена информация в режиме реального времени. Например, в интеллектуальной энергосистеме можно отслеживать производство и потребление энергии в режиме реального времени; в таком же режиме собираются и обрабатываются сведения о движении транспорта в целях предоставления информации о пробках и выработки рекомендаций по оптимизации маршрута.

9. В Интернете отсутствуют органы, которые подтверждали бы достоверность информации или качество товара или услуги. В целях создания атмосферы доверия были созданы механизмы на основе подхода «снизу вверх». Следуя характеристикам Интернета как сети, поставщики услуг пытаются повысить достоверность информации, связываясь с другими узлами сети. Количество ссылок на источник информации повышает актуальность веб-поиска. Платформы собирают информацию об оценке товаров лицами, которые покупали этот конкретный товар в прошлом. На некоторых платформах эта концепция получила дальнейшее развитие путем сопоставления обзоров продавцов и потребителей услуг и создания «сети доверия». В таких случаях репутация становится показателем надежности. Потребителям рекомендуется оценивать качество потребляемых ими товаров или услуг.

IV. Задача официальной статистики

10. Статистические управления предоставляют обществу высококачественную статистическую информацию о состоянии и тенденциях в обществе, экономике и окружающей среде. Для выполнения этой задачи статистическая система создала условия, основанные на правовых рамках.

11. Кроме того, надежность статистических данных обеспечивается благодаря этическим кодексам поведения, таким как Кодекс норм европейской статистики и рамки качества, предусматривающие систематический сбор информации

(метаданных) для описания соответствующего процесса и статистического продукта. Соблюдение принципов этического кодекса осуществляется на основе проверок. В случае Европейских статистических систем (ЕСС) механизм безопасности ЕСС обеспечивает гарантии и сертификацию в отношении ИТ-безопасности и проведение коллегиальных обзоров для установления соответствия Кодексу норм европейской статистики. В силу дефицитности и дороговизны данных основные усилия направлялись на разработку наиболее оптимальных методов их сбора. Системы качества определяют и описывают критерии, процедуры и показатели для сбора и оценки качества статистических информационных продуктов. В заключение следует отметить, что статистическая система была разработана на основе подхода «сверху вниз», начиная с ее задачи и заканчивая статистическими продуктами.

12. Несмотря на то, что эта общая задача официальной статистики еще не выполнена, средства ее выполнения должны измениться, отчасти в результате описанных выше изменений в методах сбора, обработки и извлечения информации.

V. Последствия для официальной статистики

13. Официальная статистика подвергается критике за утрату способности точно отображать мир¹. Это также может быть обусловлено изменением привычек в отношении поиска информации и ожиданий граждан в плане потребления информации в результате повсеместного распространения Интернета. Наряду с изменением привычек меняются и ожидания. Пользователи требуют, чтобы те возможности, которыми они пользуются в одной области поиска информации и предоставления услуг, имелись и в других областях. Кроме того, изменения в ожиданиях обусловлены необходимостью принятия основанных на данных решений, которые требуют подтверждающих сведений, основанных на информации, до их принятия. Спрос на это наличествует не только в деловом секторе, но и в секторе государственном, особенно в рамках политических обсуждений в обществе. Принятие решений на основе данных или разработка политики с опорой на имеющиеся сведения окажут огромное влияние на статистику, поскольку спрос на надежные и высококачественные данные будет расти в случае серьезного отношения к нему.

14. Наличие новых источников данных и меняющиеся привычки в плане использования информации влияют на производство статистических данных, а также на то, как потребляется статистическая информация.

A. Вопросы вводимых данных

15. Как правило, для целей официальной статистики новые данные не собираются. В силу этого для получения статистики их необходимо перепрофилировать. Часто эти данные отличаются детализацией по сравнению с микроданными, которые обрабатываются официальной статистикой. Например, они не содержат информации о состоянии здоровья человека, но приводят данные о кровяном давлении или сердечном ритме, полученные от устройств, отслеживающих состояние здоровья. Кроме того, их временная частота, как правило, намного выше, чем частота микроданных статистических управлений. Эти данные должны быть проанализированы и преобразованы в значимую информацию, подходящую для официальной статистики. В отличие от микроданных, используемых в официальной статистике, мы называем эти данные глубокими данными или наноданными.

16. Объемы этих данных намного выше, чем объемы данных, обрабатываемых статистическими управлениями. Использование всех или только части этих данных может выходить за рамки возможностей их обработки статистическими управлениями. По этой причине такие данные должны быть предварительно обработаны для их преобразования из наноданных в микроданные, с тем чтобы ими

¹ William Davies (2017): How statistics lost their power; The Guardian; <https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy>.

можно было манипулировать и интегрировать их в рамках процесса статистического производства. Такую передачу функции, связанной с вычислением, держателям данных необходимо также рассмотреть с точки зрения их безопасности. Централизация всей возможной информации в одном месте резко повышает риск утечки данных, так как центральное хранилище данных будет очень привлекательным для нападений. С целью распределения рисков данные должны храниться там, где они были получены, а технические средства обработки, агрегации и увязки данных должны включать в себя технологии сохранения конфиденциальности до того этапа, когда полученные результаты будут отражать уже не личные, а совокупные показатели.

17. Использование таких принципов ведет к переносу акцента в производстве данных с разработки порядка их сбора на проектирование обработки данных и с доступа к данным на получение окончательной или промежуточной статистической информации, что является конечной целью статистических управлений. Основное внимание будет уделяться не обмену данными, а алгоритмам и методологии совместного использования для получения статистических (конечных или промежуточных) результатов. Такой подход потребует тесного партнерства между держателями данных и статистическими управлениями и использования стандартных или согласованных методологий для обработки данных у источника.

18. Кроме того, механизмы отбора источников определенного качества обеспечивают проверку алгоритмов обработки, предпочтительно в виде открытого исходного кода, и гарантируют то, что согласованные программы выполняются на основе данных из согласованных источников. В идеале полученные результаты должны быть полезны не только статистическим управлениям, но и держателям данных для формирования беспроблемной ситуации. Такая ситуация значительно облегчила бы сотрудничество между держателями данных и статистическими управлениями. Вместе с тем помимо предоставления данных или информационных услуг могут существовать и другие стимулы.

19. Необходимо будет широко использовать системы стимулирования для содействия получению данных из частных источников или от отдельных лиц. Этими стимулами могут быть игрофикация, вовлечение сообществ, общественное признание, персонализированные услуги, денежное или псевдо-денежное вознаграждение для отдельных лиц. Для предприятий такими стимулами могли бы стать вклад в общественное благо, предоставление информации, адаптированной для компании, создание возможности получения информационных услуг со статистическим управлением в качестве третьей стороны, что требует интеграции данных от конкурентов, например получение информации о том, сколько компаний подвергаются нападениям со стороны хакеров, без раскрытия того, какая компания действительно подвергается таким нападениям, или объединение данных для получения совместных статистических результатов. Одним из способов стимулирования рынков данных может стать внедрение моделей минимально-бесплатных услуг.

20. В контексте корпоративных данных важно напомнить, что новые данные, как правило, информируют не о конкретных характеристиках компании, а о том, что субъектами данных являются третьи лица. В силу этого вопрос о бремени отчетности следует толковать по-иному. В случае использования стимулов для сбора данных их предоставление может не восприниматься как дополнительное время.

В. Обработка данных в статистических управлениях

21. Мы предполагаем, что статистические управления по каналам расчетов, обеспечивающим неприкосновенность частной жизни, будут получать промежуточные данные, которые могут быть подвергнуты дальнейшей обработке для получения окончательных статистических результатов. Часто эти новые данные могут использоваться для различных целей, т. е. для создания статистики в различных областях. Так, данные мобильных сетей могут быть использованы для улучшения

статистики народонаселения, миграции, туризма или платежного баланса. В силу этого данный процесс должен быть гибким, чтобы обеспечить многоцелевое использование. В то же время какая-то область статистики может получать данные из другого источника, например статистика туризма – из обследований, данных мобильных сетей, счетчиков движения и т. д. Архитектура инфраструктуры их производства должна поддерживать принятие множества видов входных данных для получения разнообразных результатов.

22. Кроме того, какой-то новый источник данных, скорее всего, не сможет полностью заменить традиционный метод сбора данных, однако в совокупности источники данных улучшат различные аспекты конечного продукта, такие как временная или пространственная детализация.

23. Эта новая архитектура производства данных увеличит потребность в координации между статистическими областями в рамках статистических управлений. Изменения в производственном процессе могут отразиться на нескольких конечных продуктах. Возможности новых источников данных способны положительно повлиять на различные области статистики.

24. В связи с техническими характеристиками новых источников данных в статистических управлениях может возрасти потребность в дополнительных специалистах-инженерах. Для того чтобы успешно согласовать и спроектировать вычислительный процесс, персонал должен обладать специальными знаниями в различных областях новых источников данных, например связи, транспорта, электроники, инженерии.

С. Вопросы получаемых результатов

25. Ранее мы уже описывали изменение привычек пользователей при извлечении информации из Интернета. Статистическим управлениям следует уделять больше внимания способам передачи статистической информации различным группам пользователей, в идеале предоставляя им информацию в нужное время в нужном количестве и в легко усваиваемом виде.

26. Пользователи ожидают, что для получения запрашиваемой информации будут применяться различные режимы взаимодействия. Все чаще люди пользуются цифровыми помощниками, которые взаимодействуют с ними через естественную речь. Это может породить ожидания, что информация может быть получена с помощью функции распознавания речи.

27. В целом растут ожидания в отношении детализации статистической информации. Средние показатели и суммы, агрегированные для более крупных территорий, могут дать представление обо всем обществе в целом, но могут не отражать саму реальность или ее восприятие с точки зрения конкретных групп общества и более мелких регионов. Растет потребность в том, чтобы статистические данные лучше отражали реальное положение более мелких социальных групп или частей территории.

28. Статистическая информация должна быть немедленно доступна в форматах и количествах, соответствующих потребностям в такой информации.

29. Восприятие своевременности информации будет меняться в связи с ее непосредственным наличием и появлением систем, действующих в режиме реального времени. Люди могут больше не принимать в качестве свежей информацию более чем годичной давности.

VI. Производство официальной статистики третьими сторонами

30. Спрос на статистическую информацию растет с учетом расширения предложения оцифрованных данных и возникновения парадигмы решений,

основанных на данных. Новые источники данных, скорее всего, не смогут полностью заменить существующие методы сбора данных. В силу этого подход, заключающийся в интеграции новых данных в процесс подготовки официальной статистики, скорее всего, не приведет к сокращению ресурсов, а в лучшем случае потребует выполнения дополнительных задач при сопоставимых ресурсах.

31. По-прежнему будет существовать необходимость в механизме определения приоритетов для принятия решений об инвестициях в конкретные области. Кроме того, акцент в работе будет перенесен со сбора данных на проектирование процесса извлечения информации из исходных данных. Мы предполагаем, что в будущем на данный процесс будет расходоваться больше всего ресурсов в рамках производства статистики, включая, например, разработку методологии, алгоритмов, ИТ-инфраструктур, рамок для методологии, показателей качества и метаданных. Это те элементы, которые в дальнейшем будут использоваться для документирования процессов производства данных и качества конечных результатов при условии, что весь производственный процесс или промежуточные этапы выполняются третьими сторонами. Эти элементы были бы также пригодны для внедрения сертификации, в ходе которой процессы или результаты подтверждаются на предмет соответствия заранее определенным рамкам.

32. Типичный процесс подготовки сертификации может включать разработку соответствующих рамок, для которой лучше всего подходят технические специалисты, действующие совместно со статистиками. Такие технические специалисты, как инженеры, будут предоставлять конкретные знания об источнике данных, что необходимо для преобразования исходных данных в промежуточные форматы, которые могут служить основой для статистических приложений. Полученные таким образом рамки будут использоваться для документирования обработки данных и описания характеристик результатов обработки. В зависимости от желаемого уровня качества продукта некоторые элементы могут быть выбраны для документирования. Третья сторона, не обязательно статистическое управление, может подтверждать достоверность документации. Пригодность полученных данных для конкретных статистических приложений может быть оценена путем сопоставления с ее документацией.

33. Учитывая то, что большая часть ресурсов будет израсходована в процессе разработки методологии и что соответствующую деятельность будут осуществлять статистические управления, сертификация не будет совместима со значительным расширением общего предложения статистической информации без увеличения ресурсов статистических управлений. Кроме того, необходимо будет повторно сертифицировать новые источники данных, поскольку они имеют тенденцию к неустойчивости вследствие технологических изменений или изменений в поведении пользователей. В заключение следует отметить, что внедрение сертификации не освободит статистические управления от необходимости инвестировать в людские ресурсы для осуществления этого процесса.