

**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 проблемы систем данных****К сертификации новых общественно значимых
статистических данных: почему и как?****Записка Национального института статистики и экономических
исследований (Франция)***Резюме*

В настоящем документе представлены анализ и предложение по признанию в качестве общественно значимых новых статистических данных во Франции. В нем описывается нынешняя ситуация во Франции в области сертификации статистических данных, производимых за рамками национальной статистической системы, будь то государственных или частных, а также вопросы и проблемы, связанные с их появлением, поскольку эти статистические данные могут дополнять, а иногда и бросать вызов официальной статистике. Авторы предлагают трехуровневый механизм, на первом уровне ограничивающийся самодокументированием статистики в соответствии с заранее определенным шаблоном, на втором уровне предусматривающий одобрение такой документации и на третьем уровне состоящий из одобрения документации, а также методологии и выходных материалов.

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. Во Франции под государственной статистикой понимаются статистические данные статистических обследований или административных источников, произведенные не только национальной статистической системой (НСС), но и государственными учреждениями или частными органами, которым поручено выполнение задач государственной службы. Поэтому потенциально она представляет собой очень обширный массив статистических данных, которые являются общественным благом и предоставляются всем потенциальным пользователям.
2. Появление открытых данных и цифровизация экономики расширяют сферу охвата новых государственных или частных наборов данных и статистики. Таким образом, производство данных или статистики стало доступным для новых субъектов, которые имеют свои собственные источники данных, созданные в рамках их деятельности, или повторно используют публичные источники благодаря растущему предложению открытых (необработанных) данных. Эти новые статистические данные могут пересекаться со сферой государственной статистики, предоставляя дополнительные сведения или создавая помехи, если сообщаемые данные отличаются, а также охватывают новые области наблюдения.
3. Это явление не является чем-то новым: например, всегда существовали обследования, проводившиеся профессиональными органами, которые готовили и публиковали статистические данные, широко освещаемые в средствах массовой информации. С цифровой революцией изменилась общая возможности субъектов частного сектора и гражданского общества производить массивы данных и получать новые статистические данные, представляющие общий интерес.
4. НИСЭИ в своей стратегии на 2025 год наметил провести перспективный анализ, направленный на «позиционирование государственной статистической службы в качестве органа по сертификации данных». В условиях, когда феномен «лже-новостей» набирает все большую силу, этот подход представляется тем более легитимным. Как развить функцию сертификации этих новых статистических данных, с тем чтобы привлечь внимание к тем из них, которые имеют дополнительную полезность? Какую область она должна охватывать? Каковы возможные методы реализации этой новой функции?
5. Для ответа на эти вопросы была запрошена миссия Генеральной инспекции НИСЭИ, основные выводы которой представлены в настоящем документе.

II. Обзор положения во Франции

6. Национальная статистическая система (НСС) уже осознала необходимость сертификации статистических данных, лежащих в основе общественных дебатов. Статистический орган, таким образом, присваивает определенный знак части статистической продукции организаций, на которые возложены задачи государственной службы. Эта процедура обеспечивает соблюдение принципов Кодекса норм европейской статистики: профессиональная независимость при разработке, производстве и распространении государственной статистики, а также принципов объективности, беспристрастности, релевантности и качества производимых показателей.
7. Другие статистические данные, производимые частными субъектами или организациями гражданского общества, «калибруются» в соответствии с критериями, принятыми Национальным советом по статистической информации (НССИ). Однако эта процедура весьма обременительна: она включает в себя не только оценку документации, но и проверку методологии. В результате число «откалиброванных» статистических данных остается очень низким по сравнению с данными, производимыми частным сектором или организациями гражданского общества, в то время как некоторые из них находят широкий отклик в средствах массовой информации или в ходе общественных дебатов.

8. Статистическая система уже воспользовалась некоторыми возможностями, предоставляемыми новыми государственными или частными источниками, вместо того чтобы игнорировать их и придерживаться «обычной практики», что повлекло бы за собой явный риск маргинализации, поскольку репутация статистической системы и доверие к ней не способны компенсировать стремление к инновациям (Ricciatio et al., 2018). Многие национальные статистические институты (НСИ) разработали планы использования больших данных в качестве вводимых ресурсов для своего процесса статистического производства. В мае 2018 года НИСЭИ создал инновационную лабораторию (SSPLab) в качестве ресурсного и стимулирующего центра прикладных исследований и экспериментальных разработок в целях содействия инновациям и использованию новых источников данных, технологий и методов интеллектуальной обработки данных. SSPLab участвует в работе Целевой группы Евростата по большим данным. Например, SSPLab проводит эксперименты с использованием данных мобильной телефонии для подготовки демографических оценок.

9. Таким образом, национальные статистические системы быстро отреагировали на необходимость инвестирования в область больших данных и расширения их использования для производства официальной статистики. Однако нельзя исключить, что чисто частная статистика, полученная с помощью новых технологий, бросает вызов официальной статистической продукции. Если говорить в более позитивном плане, то частная статистика могла бы также дополнять официальную статистику, поскольку в этом случае она должна признаваться в качестве заслуживающей внимания информации наряду с другими источниками.

10. Например, некоторые частные компании готовят отчеты о занятости в качестве побочного продукта своей деятельности по начислению заработной платы для частных компаний или используют методы извлечения парсинга контента для сбора вырезок из печати о планах по сокращению штатов, найму или местным инвестициях. Их бизнес-модель, лежащая в основе этой деятельности, заключается не в коммерциализации агрегированных статистических данных, а в привлечении внимания и завоевании доверия средств массовой информации в целях продажи коммерческих продуктов, таких как подробная информация для частных клиентов.

11. Однако основные производители неофициальной статистики принадлежат к государственному сектору: различные информационные центры, финансируемые из государственных фондов или фондов местных органов власти, например регионов или крупных городов. Они производят статистические данные, которые могут дополнять государственную статистику по конкретной области (либо секторальную, либо географическую).

III. Вопросы и проблемы

12. Первый вопрос заключается в предоставлении общественности соответствующей информации посредством сопроводительных статистических данных, подготовленных с использованием соответствующих метаданных. Первая и наиболее фундаментальная цель статистических метаданных состоит в том, чтобы помочь пользователям интерпретировать, понимать и анализировать статистические данные. Уровень детализации метаданных зависит от навыков и потребностей пользователей. Например, журналист может не иметь времени переварить большой объем подробных метаданных, и ему понадобится минимум информации, чтобы избежать неправильного толкования. И наоборот, социолог может даже захотеть поставить под сомнение предположения, сделанные первоначальным составителем статистики, и вывести новые статистические результаты, основанные на альтернативных гипотезах. Последний тип пользователей должен иметь доступ ко всем соответствующим допущениям и другим условиям сбора, подготовки и оценки данных процесса, сконструированным и используемым статистическим органом (Dippo, Sundgren, 2000).

13. Помимо метаданных, важно значение имеет качество произведенных данных. Кодекс норм европейской статистики (КНЕС) регулирует производственный процесс государственной статистической системы Франции и других государств-членов. Он обеспечивает, чтобы производимые статистические данные основывались на принципах качества – профессиональной независимости, релевантности, качества процессов и продуктов – и сопровождалась соответствующими метаданными.

14. Статистические данные, производимые за рамками Национальной статистической системы, не всегда имеют необходимые метаданные и не всегда отражают обязательства в отношении качества. Частные компании представляют некоторые элементы описания методологии, но часто неполные. Например, во многих случаях отсутствуют данные об охватываемой области или возможных колебаниях в сборе информации, которые могут повлиять на временную сопоставимость. Даже описания самой концепции и способа ее измерения могут быть неполными, что ведет к неправильному толкованию.

15. Поэтому статистическая грамотность имеет первостепенное значение для понимания пользователями, как производятся статистические данные и каково их качество. Например, в работе Radermacher (2018) утверждается, что статистики должны инициативно заниматься развитием навыков работы с данными в обществе, чтобы обучить граждан более эффективному использованию официальной или иной статистики.

IV. Предлагаемый механизм

16. В условиях, когда феномен «лже-новостей» приобретает все больший размах, журналисты ежедневно сталкиваются с необходимостью проверять имеющуюся у них информацию. Во французских СМИ, как и во всем мире, все больше и больше редакций новостей имеют сегодня в своем штате редакторов, специализирующихся на проверке фактов. Основой журналистики является перекрестная проверка источников и контроль их достоверности (Léchenet, 2015). Так, журналисты предпочитают проводить свои исследования в одиночку на основе имеющихся источников, а не «руководствуясь» интересами одной из заинтересованных сторон. Исходя из этого, НСИ должны проявлять осторожность при «сертификации» внешних источников, рискуя быть обвиненными в защите своей монополии с использованием механизма «припечатывания к позорному столбу».

17. Таким образом, предлагаемый механизм представляет собой трехуровневый процесс сертификации статистических данных с растущим градиентом требований. Первый уровень должен требовать самодокументирования статистических данных в соответствии со стандартизированным шаблоном, который надлежит определить. Второй уровень должен заключаться в одобрении («регистрации») документации на добровольной основе. Это не означает одобрения методологии или выходных материалов, а лишь признание того, что документация правильно заполнена, с тем чтобы предоставлять необходимую информацию пользователям. Третий уровень должен предполагать оценку качества и релевантности выходных материалов.

18. Эта процедура будет касаться в первую очередь разработчиков статистики, которым поручены задачи государственной службы, которые потенциально гораздо чаще придерживаются такой схемы, чем частные производители. Предполагается, что это способно оказать положительное влияние даже в частной сфере.

19. Сначала НССИ примет решение о приемлемости заявки на оценку документации статистического продукта с учетом его релевантности в плане информирования общественности. После уведомления о приемлемости оценка документации будет поручена техническому комитету. В состав этого комитета войдут эксперты, работающие как в рамках национальной статистической системы, так и за ее пределами. Данный процесс изучения сопряжен с переменными расходами, которые будет нести в основном комитет: одобрение документации статистики может потребовать обменов информацией с заявителем. Для снижения этих затрат

необходимо уточнить стандарт и условия его применения. Заключение комитета будет касаться только качества документации, без одобрения качества продукта.

20. Этот механизм предполагает определение шаблона для документирования статистики. Для установления такого стандарта предлагается создать рабочую группу в составе официальных статистиков и внешних экспертов, а также потенциальных производителей из частного и государственного секторов, с целью предоставления пользователям соответствующей информации в шаблоне.

21. Документация будет одобряться на ограниченный, возобновляемый период, а онлайн-хранилище будет содержать описание перечня одобренных статистических данных.

22. Третий уровень сертификации соответствует присвоению знака государственной статистики: поэтому она будет предназначена для статистических данных, подготавливаемых организациями, выполняющими задачи государственной службы. В настоящее время он включает в себя оценку релевантности выходных материалов и используемой методологии, а также проверку соблюдения основных принципов Кодекса норм европейской статистики. Поскольку НИСЭИ и статистические отделы министерств, которые образуют НСС, взаимодействуют со многими партнерами, занимающимися подготовкой государственной статистики, члены НСС могли бы играть активную роль в сертификации источников данных своих партнеров. Это добавило бы восходящий подход, который мог бы привести к расширению сферы охвата сертифицированных статистических данных.

Диаграмма 1

Трехуровневый механизм сертификации статистических источников



23. Однако процедура одобрения останется неработоспособной, если потенциальным кандидатам не будет известно о ней и если пользователи не будут проводить различия между разными уровнями качества. Пирамида, состоящая из возрастающих уровней требований к качеству, с НСС в ее верхней части и ее обязательством соблюдать Кодекс норм европейской статистики, должна быть представлена в визуальном и образовательном формате.

24. Наконец, одним из способов контроля или ограничения аудитории неофициальной статистики является более убедительное информирование о наших обязательствах и достижениях с точки зрения качества. Поэтому рекомендуется развивать коммуникацию, в большей степени ориентированную на пользователя, по вопросам качества статистических данных, в частности на веб-сайте НИСЭИ. Например, Статистический комитет Соединенного Королевства принял к использованию ориентированный на пользователя подход к коммуникации, сгруппировав свой кодекс норм в области статистики по трем основным принципам (надежность, качество и ценность), с тем чтобы пользователи были четко осведомлены об этом.

V. Выводы и рекомендации

На семинаре КЕС можно было бы обменяться мнениями о различных национальных ситуациях и о возможности выработки единого подхода к сертификации неофициальной статистики.

VI. Справочные материалы

Dippo, C, Sudgren, B. (2000). «The Role of Metadata in Statistics», paper presented at the International Conference on Establishment Surveys II, Buffalo, New York. <https://www.bls.gov/osmr/abstract/st/st000040.htm>.

Léchenet, A. (2015). Le journalisme de données. Synthèse des débats. Cafés de la statistique, soirée du 14 avril 2015. Société Française de statistique.

Radermacher, W. (2018). Official statistics in the era of big data opportunities and threats. International Journal of Data Science and Analytics. 6:225-231. Springer International Publishing AG.

Ricciato, R, et alii. (2018). Towards a reference architecture for trusted smart statistics. Eurostat Task Force on Big Data. Papier présenté à la Conférence européenne sur la qualité dans les statistiques officielles 2018.
