

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

Женева, 10–12 июня 2013 года

Пункт 4 а) предварительной повестки дня

**Факторы, обуславливающие потребность в доступе
к микроданным****Обмен микроданными и проблемы, связанные
с открытыми данными и транспарентностью****Записка, подготовленная Соединенным Королевством
и Организацией экономического развития и сотрудничества***Резюме*

В настоящем документе рассматриваются проблемы использования открытых данных для обмена микроданными. В целях обслуживания исследовательских кругов многие статистические управления создали национальные архивы данных, лаборатории данных на местах и средства дистанционного доступа. Однако в связи с обслуживанием широкой общественности и формирующегося сообщества предпринимателей информационного сектора возникает иной комплекс проблем. Производители официальной статистики должны учитывать давление, обусловленное проблемой анонимизации открытых данных. Статистическим управлениям необходимо будет найти способы решения целого ряда проблем, связанных, в частности, с лабораторными средствами/средствами дистанционного доступа, файлами, используемыми в научных целях, файлами общего пользования и информационными продуктами, относящимися к открытым данным. В последнее десятилетие качество услуг, оказываемых исследовательскому сообществу, существенно повысилось, и эта тенденция будет продолжаться. Однако прогресс в области разработки и доведения файлов общего пользования до стандартов для открытых данных с целью включения их в поток "больших данных" был значительно меньшим. В этой связи в настоящем документе рассматриваются упомянутые выше вопросы, которые могут стать важной областью работы в предстоящее десятилетие.

I. Введение

1. Настоящий документ основан на опыте Национального статистического управления Соединенного Королевства, однако его нельзя рассматривать как заявление, в котором излагается политика этого Статистического управления или правительства Соединенного Королевства. В нем обсуждаются некоторые программные и законодательные инициативы национальных статистических учреждений по открытым данным в контексте микроданных. Движущие факторы и варианты решений для открытых данных в составе микроданных отличаются от соответствующих факторов и решений в контексте обмена конфиденциальными микроданными, однако проблемы и в той, и в другой области деятельности не могут решаться без понимания особенностей обеих областей.

2. Для целей настоящего доклада "открытые данные" означают данные (наборы данных), которые:

- a) являются доступными для всех и каждого – в идеальном случае через Интернет;
- b) имеются в цифровой машиночитаемой форме, обеспечивающей функциональную совместимость с другими данными;
- c) предоставляются за плату, не выше или даже ниже стоимости воспроизводства;
- d) свободны от ограничений на использование или повторное использование.

3. Для целей настоящего доклада:

- a) микроданные означают данные, отражающие в количественном выражении наблюдения или факты по конкретной единице отчетности;
- b) статистические данные означают сочетания данных, которые могут указывать на закономерности в наблюдениях и фактах.

4. Таким образом, для целей обсуждения открытых данных соответствующие микроданные означают отдельные записи прямых наблюдений и фактов, полученные с помощью инструментов статистического исследования (например, обследований), из регистров или административных источников. Перед распространением в данные вносятся изменения лишь тогда, когда необходимо соблюсти обязательство в отношении конфиденциальности перед субъектами наблюдений. Такие данные можно назвать открытыми микроданными. Цель настоящего документа заключается в изучении предположений относительно того, что обмен микроданными должен во все большей степени основываться на принципах работы с открытыми данными, и проблем, возникающих перед разработчиками официальной статистики.

II. Факторы, обуславливающие потребность в открытых данных

5. К факторам, побуждающим к разработке микроданных как открытых данных, относятся, в частности, принятие *научных принципов*, поддержка демократии, стимулирование социального и экономического роста и реакция на законодательство и государственную политику.

A. Научные принципы

6. Открытые данные поддерживают статус статистических данных как элементов научной дисциплины.

7. Основные принципы официальной статистики¹ требуют, чтобы решения относительно методов распространения официальных статистических данных принимались на основе *научных принципов*. Когда данные представлены в соответствии с *научными стандартами*, пользователям проще толковать их.

8. Каковы же эти *научные принципы* и *научные стандарты* распространения и представления данных? Название "Основные принципы" предполагает, что они уже какое-то время существуют и хорошо известны.

9. В июне 2012 года Королевское общество опубликовало свой доклад, озаглавленный "Наука как открытое предприятие"². Наука в нем рассматривается как самокорректирующийся процесс. Теории могут независимым образом подтверждаться, отвергаться или совершенствоваться. Выводы и подтверждающие их данные доводятся до сведения как можно более широкой аудитории в целях дальнейшего развития соответствующих теорий.

10. Разработчикам официальных статистических данных следует задать себе вопрос о том, могут ли публикуемые ими статистические данные – их выводы относительно закономерностей в наблюдениях и фактах – независимым образом подвергаться, отвергаться или совершенствоваться. Применить этот научный принцип к официальной статистике позволяют открытые данные. Опубликованная методология допускает лишь абстрактные возражения за исключением тех случаев, когда данные, используемые разработчиками статистики, имеют также у другой независимой стороны. В какой-то степени это может быть достигнуто путем использования каналов контролируемого доступа, например независимых статистических обзоров, или доступа к данным в исследовательских целях. Однако максимальная транспарентность достигается тогда, когда для независимого анализа статистических данных не создается никаких препятствий в виде ограничений, отбора или давления (будь то реального или вымышленного).

11. Доверие общественности может быть подорвано, если отсутствует свободный доступ к данным, лежащим в основе научной информации и статистики, а также к информации об использовавшихся методах. Например, доверие общественности к публикациям Группы по климатологическим исследованиям Университета Восточной Англии было подорвано, когда общественности был закрыт доступ к некоторым данным и информации о методах. Было проведено два обследования, после которых правительство рекомендовало полностью раскрыть исходные данные, а также необходимые компьютерные программы и методологии представления результатов.

"Раскрытие исходных данных и достаточно подробной информации о компьютерных программах имеет крайне важное значение для побуждения людей к критическому обсуждению научных разработок традиционным способом, к опровержению результатов проделанной работы, в результате чего появляется возможность ее валидации и выдвижения новых гипотез".

¹ <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>

² http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

- Реакция правительства на первый доклад Комитета по науке и технике о работе его сессии 2010–2012 годов³.

В. Неотъемлемый элемент демократии

12. Согласно Основным принципам официальной статистики официальные статистические данные являются "... неотъемлемым элементом информационной системы демократического общества".

13. Открытое правительственное партнерство представляет собой инициативу, направленную на усовершенствование органов управления. Оно было основано в сентябре 2011 года 8 членами правительства, и в настоящее время его состав увеличился до 55 членов правительства. В своей декларации Партнерство взяло на себя обязательство расширять объем имеющейся информации о деятельности правительства. Партнеры взяли на себя обязательство:

"... действуя на упреждение, своевременно предоставлять имеющую высокую ценность информацию, включая исходные данные, в формах, которые облегчают поиск, понимание и использование этой информации общественностью, а также ее повторное использование".

- Открытая декларация правительства, Открытое правительственное партнерство⁴.

14. Применяя Основные принципы, большинство национальных статистических систем поставят перед собой предусмотренные законом цели, в которых будет упоминаться поддержка тщательного анализа государственной политики. Главная цель Статистического комитета Соединенного Королевства и его исполнительного органа – Национального статистического управления (НСУ) – формулируется следующим образом:

"Комитет должен преследовать цель содействия разработке и опубликованию официальных статистических данных, которые идут на благо общественности, и обеспечения гарантий этого процесса. Формулировка "идут на благо общественности" означает, в частности, информирование общественности о социальных и экономических вопросах и оказание содействия в разработке и оценке государственной политики".

- Закон о статистической и регистрационной службе 2007 года⁵.

С. Открытые данные для обеспечения экономического и социального роста

15. Согласно оценке, содержащейся в исследовании Викери⁶, проведенном Европейской комиссией, прямые и косвенные выгоды от того, что информация государственного сектора является открытой и может повторно использоваться, составляют порядка 140 млрд. евро.

³ <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/496/496.pdf>

⁴ <http://www.opengovpartnership.org/open-government-declaration>

⁵ <http://www.legislation.gov.uk/ukpga/2007/18/section/7>

⁶ http://ec.europa.eu/information_society/policy/psi/index_en.htm

16. Рекомендация Совета Организации экономического сотрудничества и развития (ОЭСР) о более широком доступе к информации государственного сектора и ее более эффективном использовании 2008 года⁷ содержит ряд соответствующих рекомендаций относительно открытых данных в статистических микроданных, направленных на "увеличение экономических и социальных благ, в частности, за счет более эффективного распределения (информации), более активных инноваций и внедрения новых видов использования".

17. Странам-членам рекомендуется принять соответствующие принципы, которые включают положения, имеющие актуальное значение с точки зрения статистических микроданных:

а) *открытость*. Обеспечение максимального наличия данных с открытым доступом в качестве действующего по умолчанию правила;

б) *транспарентные условия повторного использования*. Недискриминационные условия, устраняющие эксклюзивные договоренности и необязательные ограничения;

в) *перечни активов*. Публикуемые в онлайн-режиме реестры открытых данных;

г) *качество и безупречность*. Использование наилучших методов подготовки данных и защиты от представления неверных данных;

е) *плата*. Бесплатное предоставление или же взимание минимальной платы, позволяющей лишь возместить издержки;

ф) *международные аспекты*. Облегчение трансграничного использования и интероперабельности.

18. Эти рекомендации создают концептуальную основу для проведения национальными статистическими учреждениями (НСУ) оценки целесообразности принятия и распространения на микроданные принципов, касающихся открытых данных.

D. Законодательство и государственная политика в области открытых данных

19. Существуют две основные формы законодательства и политики в области открытых данных. Во-первых, это те законы и государственная политика, которые возлагают на государственный сектор обязанность "толкать" открытые данные. Во-вторых, это те законы и государственная политика, которые дают пользователям право "вытягивать" открытые данные из государственного сектора.

1. Законодательство и политика, направленные на "толкание" данных

20. Деятельность многих НСУ подпадает под действие таких положений законодательства и политики, которые вселяют надежду на то, что они будут "толкать" открытые данные в направлении общественности в соответствии с определенным регулятивным режимом.

21. Например, в Европейском союзе регулирование рынка информации государственного сектора началось с принятия Директивы о повторном использова-

⁷ <http://www.oecd.org/internet/interneteconomy/40826024.pdf>

нии информации государственного сектора⁸. В настоящее время эта директива пересматривается с целью укрепления ее положений и учета реалий цифрового мира. Строго говоря, эта директива не предусматривает обязанности готовить документы общего пользования, а содержит лишь позитивную нормативную основу, предлагаемую для закрепления в законодательстве государств-членов, исходя из того, что информация государственного сектора должна быть открытой при определенных разумных ограничениях. Например, в директиве закрепляется принцип, согласно которому общий доход от предоставления данных и выдачи разрешения на их повторное использование не должен превышать затрат на сбор, обработку и распространение данных, включая разумную прибыль на инвестиции.

22. Выполнение этой директивы было обеспечено во всех государствах-членах, при этом некоторые государства как в своих законах, так и в национальной политике пошли дальше ее требований. Например, Соединенное Королевство приняло национальные принципы, регулирующие общедоступные данные⁹, которые являются обязательными для государственного сектора и предусматривают обязанность активно поощрять использование и повторное использование открытых данных министерств.

23. В нормы, регулирующие авторские права в государственном секторе Соединенного Королевства ("Королевские авторские права"), были внесены изменения с целью закрепления принципа "толкания" данных. Существует действующая по умолчанию лицензия на использование данных государственного сектора Соединенного Королевства, которая называется открытая правительственная лицензия (ОПЛ)¹⁰. ОПЛ является позитивной и проактивной лицензией на "толкание" данных:

"[Правительство] предоставляет вам во всемирных масштабах бесплатную, бесконечную и неэксклюзивную лицензию на использование информации... Вы можете копировать, публиковать, распространять и передавать информацию; адаптировать информацию; использовать информацию в коммерческих целях, например путем ее объединения с другой информацией или путем включения ее в ваш собственный продукт или приложение".

– Открытая правительственная лицензия Соединенного Королевства.

24. Любой государственный орган Соединенного Королевства, желающий получить в порядке исключения разрешение на возмещение минимальных издержек, должен запрашивать аккредитацию при Программе для справедливых торговцев информацией¹¹, представлять технико-экономическое обоснование и проходить проверку, которая проводится Управлением информации государственного сектора.

25. Можно ожидать, что в будущем политика и структуры "толкания" данных будут получать дальнейшее развитие, в особенности, в связи с проведением общественных кампаний¹². В настоящее время Соединенное Королевство вводит в действие в качестве государственно-частного партнерства Институт по

⁸ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:EN:NOT>

⁹ <http://www.data.gov.uk/library/public-data-principles>

¹⁰ <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

¹¹ <http://www.nationalarchives.gov.uk/information-management/ifts.htm>

¹² <http://www.freeourdata.org.uk>

открытым данным¹³. Сопредседателями Института по открытым данным (ИОД) являются Найджел Шедболт и Тим Бернерс Ли. Он был создан специально для содействия "вытаскиванию" открытых данных из государственного сектора:

"...Речь идет о сотрудничестве между нашими ведущими предприятиями и предпринимателями, университетами и исследователями, правительством и гражданским обществом, направленном на стимулирование предпринимательства и создание социальных ценностей за счет использования громадного объема открытых правительственных данных, к которым в настоящее время открывается соответствующий доступ".

– Информация об ИОД размещена в Интернете по следующему адресу: www.theODI.com

2. Законодательство и политика, направленные на "вытягивание" данных

26. На примере многих НСУ можно также показать, что существуют законы и политика, которые предоставляют общественности право "вытягивать" данные из государственного сектора. Здесь следует упомянуть действующие во многих странах нормативные положения об экологической информации и закон о свободе информации. Обычно статистические данные подпадают под действие положений этих законов и политики, предоставляющих право на получение информации за исключением тех случаев, когда можно доказать, что данные носят конфиденциальный характер или планируются к опубликованию в запрошенной форме. Влияние таких законов на статистические данные может быть значительным:

a) как правило, законы, направленные на "вытягивание" данных, содержат положение о крайнем сроке выполнения просьбы или задания. В Соединенном Королевстве НСУ обязаны в течение 20 дней либо предоставить запрошенные данные, либо разъяснить, почему применяются положения об исключении из правил предоставления данных;

b) в сочетании с открытой правительственной лицензией (или эквивалентным инструментом) никакие ограничения на использование и повторное использование вводиться не могут, равно как и не могут устанавливаться какие-либо условия в этой связи;

c) обычно заявитель имеет право обжаловать решение об отказе или лишь частичном выполнении просьбы о предоставлении данных. Как правило, апелляция рассматривается судом по вопросам информации или другим подобным нестатистическим органом. Хотя статистические управления имеют возможность представить свое обоснование того, почему соответствующие данные не должны раскрываться (например, на основании конфиденциальности), решение выносится судом по вопросам информации;

d) как правило, в заявлении на предоставление данных цель подачи заявления не указывается. Заявитель не должен разъяснять, почему запрашиваются данные или в каких целях они будут использованы;

e) часто положение об исключении из правил, регулирующих предоставление данных, к данным плохого качества не применяется;

¹³ <http://www.theodi.org/people/nrs>

f) предоставление данных в соответствии с законодательством, регулирующим "вытягивание" данных, часто является основанием для установления прецедента;

g) за исключением случаев, когда тщательно соблюдаются все правила, предоставление данных согласно законодательству, направленному на "вытягивание" данных, выглядит как привилегированный доступ;

h) представление наборов данных может вести к нарушению конфиденциальности в результате "мозаичной атаки";

i) вопросы контроля за раскрытием данных часто недоступны для понимания непосвященных.

27. В настоящее время законы и политика, направленные на "толкание" и "вытягивание" данных, позволяют предпринимателям информационного сектора – от финансовых аналитиков до разработчиков программного обеспечения – с доверием относиться к экспертам по использованию данных. Формирующаяся таким образом индустрия оказывает свое собственное влияние на владельцев данных государственного сектора.

III. Проблемы подготовки открытых (микро)данных

28. Как национальные статистические управления реагируют на проблемы, связанные с открытыми данными? Судя по их статистическим данным, мы могли бы утверждать, что эта реакция является вполне положительной. Основные методы практического представления и распространения статистических данных по своему характеру совместимы с принципами распространения открытых данных. Однако в тех случаях, когда проблемы, связанные с открытыми данными, затрагивают "исходные" микроданные национальных статистических управлений, ответ на данный вопрос является менее убедительным. Обычно статистическое управление подразделяет свои активы данных на три категории:

a) "статистические данные" для опубликования или использования в государственном ведомстве – обычно это агрегированные данные или визуализированная информация;

b) "наборы исследовательских данных" для различного использования для статистических исследований при контролируемых условиях – обычно это микроданные, как правило, с пометкой "конфиденциально";

c) "производственные данные" – микроданные или агрегированные данные в статистических производственных системах, которые, как предполагается, не будут использоваться ни для какой другой цели, кроме разработки статистических данных и наборов исследовательских данных.

29. "Файлы общего пользования" – знакомое понятие, но в современной практике такие файлы обычно относят к категории "наборов исследовательских данных". Обычно такая классификация является корректной, поскольку большинство файлов общего пользования нельзя назвать открытыми данными, так как установлены определенные условия использования последних и существуют вопросы своевременности, полноты, детализации, формата и интероперабельности. Ни НСУ, ни пользователи файлов общего пользования в настоящее время не ожидают, что такие файлы можно будет использовать для тиражирования статистических данных, подготовленных НСУ – не в последнюю очередь потому, что сами НСУ не используют файлы общего пользования для разработ-

ки статистических данных. Таким образом, через файлы общего пользования с учетом того, каким образом они в настоящее время составляются, реализовать научный принцип независимого подтверждения, проверки или усовершенствования статистики невозможно.

30. Признание открытых данных в качестве категории микроданных и отнесение информационных активов к этой категории для распространения на упомянутых выше условиях – это необычные явления для современной практики. Это, возможно, объясняется многочисленными проблемами подготовки открытых микроданных.

А. Конфиденциальность

31. Проблемы конфиденциальности чаще всего приводятся в качестве довода в пользу того, чтобы не разрабатывать открытые микроданные. Очень часто этот довод является совершенно оправданным. Однако НСУ обычно рассматривает все свои неопубликованные данные как конфиденциальные либо автоматически, либо с учетом их статуса как "данных, хранимых для целей официальной статистики". Такая практика вряд ли сохранится при применении к открытым данным рычагов законодательства и политики, направленных на "толкание" или "вытягивание" данных. Во-первых, в рамках политики НСУ следует признать, что в связи с некоторыми массивами данных возникновение проблем, связанных с конфиденциальностью, потенциально менее вероятно, чем в связи с другими данными – например, данные о бюджете и расходах государственного сектора или данные о ценах можно определить как микроданные, но не как конфиденциальные данные, поскольку информация о них уже имеется в общем доступе. Во-вторых, в рамках НСУ должна быть установлена практика проведения более систематического и критического анализа того, действительно ли какой-либо массив данных разглашает конфиденциальные сведения. Для этого может использоваться так называемый "тест на проникновение", для проведения которого доверенной стороне предоставляется подлежащий проверке набор данных и разрешается выяснить, можно ли в результате проведения разумного теста почерпнуть из этих данных какую-либо личную или конфиденциальную информацию об обозначенных лицах. Если получение такой личной информации возможно, то НСУ получает независимое доказательство для представления в суд по вопросам информации или аналогичный орган, помогая им тем самым принять положительное для НСУ решение по предъявленному к нему иску в связи с непредоставлением данных какому-либо заявителю. Если же такую личную информацию с помощью "теста на проникновение" получить невозможно, то НСУ, вероятно, имеет потенциальный набор открытых микроданных.

32. Инспекторы по защите данных признают, что отличить личную информацию от неличной весьма сложно. Справиться с задачей определения того, какие данные являются или не являются личными, НСУ помогает заключение рабочей группы по статье 29, касающееся понятия личных данных¹⁴. Недавно комиссар по вопросам информации Соединенного Королевства при содействии со стороны Национального статистического управления и многих других субъектов опубликовал свод правил по анонимизации личных данных и регулированию рисков в области защиты данных¹⁵.

¹⁴ http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

¹⁵ http://www.ico.gov.uk/news/latest_news/2012/~media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

В. Качество и репутация данных

33. Из всех факторов, влияющих на доверие общественности к официальной статистике, наиболее важным, возможно, является качество данных. Завоевать репутацию источника доброкачественной статистики тяжело, а потерять легко. Вполне понятно, что открытые данные, особенно открытые микроданные, рассматриваются как фактор угрозы для репутации доброкачественной статистики. Эта угроза может проистекать из двух сценариев. Во-первых, данные вполне могут иметь низкое качество. Если данные "вытягиваются", то они могут быть разглашены до прохождения контроля качества и поступления в НСУ дополнительных исходных данных. Во-вторых, данные могут иметь высокое качество, но при этом они могут иначе и даже неправильно быть истолкованы другой стороной. НСУ, возможно, должно будет пояснить эти различия и обосновать их для определенной аудитории, которая может предпочесть вариант, при котором правой окажется неправительственная сторона. Эти проблемы решаются путем представления высококачественных метаданных, изъясления готовности прокомментировать качество анализа, проведенного другой стороной, и проведения кампаний по разъяснению данной аудитории вопросов качества и анализа. Проблема культурного характера может быть решена посредством ссылки на доброкачественную научную практику: повторное использование исходных данных в контексте применения открытых данных не представляет собой никакой угрозы для высококачественных статистических методов.

С. Стандарты распространения

34. Соблюдение стандартов распространения открытых микроданных является сложной задачей. Лишь некоторые НСУ смогут подняться до соблюдения ожидаемых стандартов в ближайшее время. Настоящим вызовом для государственного сектора Соединенного Королевства является принятие пятизвездочной программы по открытым данным¹⁶. В частности, серьезным вызовом является использование незапатентованных форматов и единых идентификаторов ресурсов (ЕИР) во всех случаях разглашения открытых микроданных, к чему текущие пользователи, возможно, даже отнесутся безо всякого энтузиазма.

Д. Использование веб-сайтов, предназначенных для других целей

35. По сравнению с агрегированными статистическими данными открытые микроданные могут являться очень крупными по объему и иметь формат, не совместимый с существующими каналами распространения, такими как веб-сайты НСУ. Переделка веб-сайтов НСУ под открытые данные может рассматриваться как низкоприоритетная задача, особенно если существуют серьезные ограничения, связанные, например, с полосой пропускания и объемом файла, а также ограничения, связанные с запатентованными форматами. Одним из вариантов решения этой задачи могло бы являться принятие конкретно разработанных альтернативных каналов распространения. Например, Соединенное Королевство создало в качестве совместного инструмента обслуживания всех учреждений государственного сектора сайт www.data.gov.uk. Следует иметь в виду, что к нему как к сайту открытых данных ограничения по стандартам безопасности правительственной информации не применяются, что позволяет использовать каналы распространения частного сектора. Одним из примеров сайта, на

¹⁶ <http://5stardata.info>

котором в настоящее время размещены открытые данные из ОЭСР, Евростата, МВФ, Всемирного банка, ФСУГ и Бюро по проведению переписей Соединенных Штатов, является сайт Google Public Data. Хотя Google Public Data используется в первую очередь для агрегирования данных, распространяемых по другим каналам, он мог бы использоваться в качестве канала первого распространения открытых данных.

Е. Аутентичность и атрибуция данных

36. НСУ являются признанным изначальным источником открытых микроданных, но они могут вполне обоснованно испытывать озабоченность по поводу изменений, вносимых в содержание данных, которые влияют на результаты анализа, но которые остаются незарегистрированными и/или неразъясненными пользователями. Иными словами, аутентичность данных утрачивается, а обозначение источника "Источник: НСУ" остается. Решение в таком случае заключается в использовании разработчиком открытых данных "стойких идентификаторов". Каждый массив открытых данных должен обозначаться с помощью уникального прямого предметного идентификатора и содержать единообразное название ресурса в сочетании с постоянным универсальным указателем местоположения ресурса и уникальной ссылкой в случае электронных изданий, использующих стандарт ИСО 690-2 (Международная организация по стандартизации). Это позволит НСУ и пользователям легко находить и получать оригинальные и неискаженные данные из источника.

Ф. Метаданные

37. Наиболее актуальным с точки зрения национальной статистики стандартом в области метаданных является стандарт, касающийся обмена статистическими данными и метаданными (ОСДМ), который, однако, был оптимизирован не для открытых микроданных, а для агрегированных статистических данных. Если НСУ не располагает экспертными знаниями в области стандартов метаданных, применяемых к открытым микроданным, то оно может попросить помощи в преодолении этого препятствия у партнеров, таких как архивы данных, в частности членов Совета европейских архивов данных по социальным наукам (СЕАДСН).

Г. Соблюдение предусмотренных Сводом правил обязательств относительно разглашения данных по определенному графику и равенства доступа к ним и предусмотренных стандартами для открытых данных обязательств в отношении временных рамок

38. НСУ пожелают обеспечить, чтобы статистическим данным, намеченным к разглашению в будущем в соответствии с оптимальной практикой установления предсказуемых и заранее объявляемых сроков разглашения, не наносился ущерб аналогичными статистическими данными, появившимися в результате повторного использования открытых микроданных. Когда речь идет о "толкании" открытых данных, этого можно достичь путем правильного установления сроков разглашения тех и других данных. Однако когда речь идет о "вытягивании" открытых микроданных, то между сроками их разглашения может возникнуть коллизия. Решить эту проблему можно путем переноса, когда это возмож-

но, на более ранний период намеченных сроков официального разглашения или же заблаговременного разъяснения потенциального наличия неаутентичных и неавторизованных статистических данных из других источников до официального опубликования.

39. Когда открытые микроданные "вытягиваются" из НСУ, важно соблюдать принцип равного доступа. Этот принцип является важным принципом статистики, но в то же время он является важным принципом работы с открытыми данными. Если открытые данные или открытые микроданные предоставляются (например) в приложениях к электронному посланию какому-либо заявителю, то это может выглядеть как предоставление привилегированного доступа. У пользователя может возникнуть впечатление, что использование данных является "незаконным". НСУ на своих веб-сайтах следует вести журнал раскрытия данных¹⁷ с целью информирования о внеплановых разглашениях открытых данных, которые, возможно, были получены конкретным пользователем, но и одновременно являются доступными для всех. Пользователю, получившему данные, URL (универсальный указатель ресурса) должен сообщаться только тогда, когда информационный ресурс становится доступным для всех.

Н. Распределение ресурсов

40. У открытых микроданных профиль ресурсопотребления отличается от профиля ресурсопотребления файлов, используемых для исследований. Первоначальные затраты на ресурсы высоки, несмотря на то, что исходная информация по определению уже существует. Затраты на анализ риска раскрытия, подготовку метаданных, сведение данных в открытый формат и т.д. являются накладными расходами. Доходы от использования открытых данных по определению не могут превышать суммы расходов плюс суммы разумной прибыли на инвестиции в их производство. Однако после того, как открытые микроданные произведены, текущие расходы на них являются ничтожно малыми, особенно если канал их распространения представляет собой совместный ресурс или если расходы по нему покрываются другой стороной. В отличие от этого профиль ресурсопотребления в случае наборов исследовательских данных изначально является низким, поскольку эти данные практически остаются неизменными по сравнению с их состоянием в момент производства. В случае наборов исследовательских данных расходы на ресурсы велики по таким позициям, как поддержание безопасной исследовательской среды, аккредитация исследователей и исследовательских проектов и проверка результатов. Бюджеты НСУ, как правило, ориентированы на покрытие текущих расходов и обеспечивают упрощенное ассигнование средств, даже если накапливающееся за несколько лет давление на ресурсы со стороны наборов исследовательских данных является более значительным по сравнению с инвестициями в открытые микроданные, имеющие эквивалентную ценность для общества.

IV. Реагирование на проблемы – проект рекомендаций

41. В настоящем докладе рассматриваются побудительные факторы и проблемы обмена микроданными в условиях действия разрабатываемых в настоящее время стандартов для открытых данных.

¹⁷ <http://www.ons.gov.uk/ons/about-ons/what-we-do/FOI/foi-requests/index.html>

42. Совершенно ясно, что преимущества обеспечения соблюдения стандарта для открытых данных в случае официальных микроданных являются весьма значительными:

- a) использование микроданных в качестве открытых данных позволяет применять *научные принципы* подтверждения, проверки достоверности и совершенствования официальных статистических данных, почерпнутых из тех же самых источников;
- b) открытыми микроданными можно обмениваться, избегая высоких затрат и бюрократических административных препятствий;
- c) открытые микроданные могут использоваться с любой целью, включая цели, о которых никто не предполагал на стадии их разработки;
- d) открытые микроданные позволяют НСУ и третьим сторонам из всех секторов осуществлять всестороннее сотрудничество и взаимодействие на базе совместно используемого информационного ресурса;
- e) для НСУ открытые микроданные представляют собой дополнительную категорию информационных активов, придавая НСУ статус важного и (как можно надеяться) ценного партнера в государственном секторе в интересах будущего современных информационных обществ;
- f) открытые микроданные способствуют развитию сообщества предпринимателей в информационном секторе и усилению экономического и социального роста.

43. Препятствия и проблемы также являются значительными:

- a) необходимо изучить риски и концепции, связанные с конфиденциальностью;
- b) возникают логистические проблемы, которые бросают вызов бизнес-архитектуре НСУ;
- c) должны быть решены вопросы аутентичности и идентификации активов;
- d) стандарты, которые, как ожидается, будут установлены для открытых данных, высоки.

A. Рекомендации

- НСУ следует применять стандарты для открытых данных в своей повседневной статистической работе. Это обеспечит ознакомление НСУ с проблемами, связанными с открытыми данными, прежде чем начнут решаться конкретные проблемы, связанные с открытыми микроданными.
- НСУ следует включить в свои реестры информационных активов категорию "открытые микроданные".
- НСУ следует взаимодействовать с целью распределения расходов на разработку методик создания категории открытых микроданных.
- НСУ следует изучить альтернативные каналы распространения открытых микроданных, если существующая архитектура является неподходящей.

- НСУ следует работать в тесном контакте с их национальным инспектором по защите данных над стандартами анонимизации и концепцией личных данных.
 - НСУ следует заранее изучить вопрос о том, как можно поддержать Свод правил статистической деятельности, когда начнут разрабатываться открытые микроданные.
 - НСУ следует использовать навыки и опыт кураторов данных, специалистов по вычислительной технике, специалистов по управлению знаниями и информацией, архивы данных и национальные библиотеки для оказания содействия в подготовке и распространении открытых микроданных.
 - НСУ следует разработать совместно с основными пользователями статистики и общественностью коммуникационную стратегию для пропаганды инновационных продуктов, связанных с открытыми данными.
-