Организация Объединенных Наций ECE/CES/GE.41/2017/8

Экономический и Социальный Совет

Distr.: General
12 July 2017

Original: Russian
English and Russian only

Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения и жилищного фонда

Девятнадцатое совещание
Женева, 4–6 октября 2017 года
Пункт 3 предварительной повестки дня
Инновации в технологии переписи

Вопросы обеспечения защиты первичных данных при реализации технологических инноваций переписи населения

Записка Федеральной службы государственной статистики, Российская Федерация

Резюме

При переписи 2020 года в России будет использована комбинированная технология сбора сведений о населении: самостоятельное заполнение респондентами электронных вопросников в сети Интернет, опрос и заполнение переписчиком электронных вопросников на планшетных компьютерах, опрос и традиционное заполнение переписчиком бумажных вопросников. Инновации создают новые вызовы по защите конфиденциальности данных, внедряя информатизацию не только на этап обработки собранных данных, но и на все этапы подготовки и проведения переписи на полевом уровне. Этот оказывает существенное влияние на стоимость переписи в целом, удорожая ее технологическую составляющую.

GE.17-11637 (R)
1711637

Просьба отправить на вторичную переработку

I. Практика Всероссийской переписи населения 2010 года

1. Перепись 2010 года проводилась традиционным методом обхода жилых помещений и сбора сведений о населении путем опроса. Данные заносились специально обученными переписчиками в машиночитаемые бланки переписных листов, которые впоследствии сканировались, распознавались, верифицировались и объединялись в общий информационный массив по регионам и по стране в целом.

2. Федеральный закон «О Всероссийской переписи населения» включает норму о том, что сведения о населении в переписных листах являются информацией ограниченного доступа, не подлежат разглашению или распространению и используются только в целях формирования официальной статистической информации. На машиночитаемых переписных листах была надпись: “КОНФИДЕНЦИАЛЬНО (гарантируется получателем информации)”.

3. По закону о переписи в обязательном порядке обеспечивается организационная защита информации: лица, которые имеют доступ к переписным листам, несут ответственность в соответствии с законодательством за утрату, разглашение этих сведений и за их фальсификацию. Обязанность не разглашать конфиденциальную информацию переписи населения, предусматривается договорами, заключаемыми с работниками переписи.

4. В связи с необходимостью контроля корректности собранных переписчиками данных и объединения данных по персонам в домохозяйства, при опросе населения помимо машиночитаемых бланков переписных листов составлялись немашиночитаемые Списки проживающих в каждом помещении. По этим спискам на этапе сбора и контроля данных можно было идентифицировать в каждом помещении, кто из респондентов переписан в домохозяйстве под номером 1, 2 и т.д., кто проживает в одном домохозяйстве, а кто – в разных.

5. Для деперсонификации данных, собираемых при переписи населения, в бланках переписных листов не предусмотрены строки для заполнения фамилии, имени, отчества, а лишь номера из Списка проживающих. Это позволяет при обработке переписных листов обезличивать данные, отделив машиночитаемые переписные листы от немашиночитаемых Списков проживающих.

6. Обработка переписных листов осуществляется в условиях, обеспечивающих их защиту от несанкционированного доступа и предотвращение их хищения, утраты, подделки или иного искажения. Для этого применяются организационно-технологические решения при создании инфраструктуры переписи и сертификация этой инфраструктуры уполномоченными государственными органами по защите информации.

7. При проведении Всероссийской переписи населения в 2010 году мы пострадали от появления клона информационного сайта переписи. В первый день Всероссийской переписи населения в Интернете появился фальшивый сайт переписи. Этот сайт-дублер копировал официальный сайт Всероссийской переписи населения, транслируя материалы в упрощенном или искаженном виде. Кроме того, при переходе с любой страницы на начало сайта появлялась примитивная мобильная версия – текстовая и сильно упрощенная. Кроме стиля оформления доменное имя фальшивого сайта схоже с именем настоящего. Однако представители Росстат сразу же объявили, что домен сайта-дублера PERIPIS-2010.RU не имеет никакого отношения к официальному домену PEREPIS-2010.RU.

8. Специалисты, ответственные за разработку официального сайта Всероссийской переписи населения, предупреждали пользователей о том, чтобы они не заходили по подозрительным ссылкам, и призывали их быть осторожными, не скачивать информацию с сайта, пока они не убедятся в его подлинности.

9. Прямого вреда данные действия не нанесли, но было подорвано доверие некоторых граждан к безопасности проведения переписи. Если такая ситуация повториться при проведении переписи с использованием интернета, то граждане, не сильно разбирающиеся в современных технологиях, могут передать свою информацию мошенникам. Возникнет недоверие к переписи.

II. Инновации переписи населения 2020 года

10. При переписи 2020 года будут использованы переписные электронные и бумажные вопросники (идентичные по составу вопросов) и 3 способа сбора сведений о населении:

- самостоятельное заполнение респондентами электронных вопросников в сети Интернет;
- опрос и заполнение переписчиком электронных вопросников на плашетных компьютерах;
- опрос и традиционное заполнение переписчиком бумажных вопросников.

11. С учетом международного опыта первый этап обычно предполагает проведение переписи населения через Интернет. Это связано с тем, чтобы у статистических органов имелась возможность установить какая именно часть населения и население каких помещений и домохозяйств переписались через Интернет. После этого стартует второй этап – классический опрос населения либо с использованием бумажных носителей, либо с использованием планшетных компьютеров. Планируемое соотношение счетных участков с планшетными компьютерами и с бумажными переписными листами – 45 к 55.

12. С точки зрения экономической целесообразности, такая комбинированная технология позволит существенно сократить расходы на изготовление бумажных бланков, закупку оборудования для сканирования бланков, оплату труда персонала для ввода МЧД, кодирования и проверки качества информации, а также технологического сопровождения этих процедур.

Таблица 1

Сравнение общей стоимости работ по ВПН-2020 при применении разных технологических способов сбора данных
(миллион рублей)

Варианты проведения ВПН-2020	*Тираж МЧД (% к вар.1)*	*ГПД (полевые работники и автоматизаторы) (% к вар.1)*	*Разработка ПО, закупка оборудования и лицензий, обработка (% к вар. 1)*	*Прочие расходы (ИРР, экипировка, МТО, обучение) (% к вар. 1)*	*Итог (% к вар.1)*	*Превышение над стоимостью минимального по затратам варианта 1 (% к вар.1)*
1. Три метода сбора (ПК, МЧД и Интернет)	371.7 (100.0%)	33 153.3 (100.0%)	9 360.3 (100.0%)	9 364.8 (100.0%)	52 250.1 (100.0%)	–
Оптимальный вариант проведения переписи: 10% Интернет-перепись, 40% на планшетных компьютерах и 50% – МЧД						
2. Традиционный (МЧД)	743.4 (200.0%)	39 957.1 (120.5%)	6 454.6 (69.0%)	9 364.8 (100.0%)	56 519.9 (108.2%)	8.2%
От варианта 1:Тираж МЧД для 100% населения (больше в 2 раза), больше нагрузка на все категории работников (ГПД больше на 20%), ПО – минимальные расходы, но больше рабочих станций на районном и региональном уровнях для обработки 100% МЧД						
3. Два метода сбора (ПК, МЧД)	371.7 (100.0%)	36 627.4 (110.5%)	8 909.2 (95.2%)	9 364.8 (100.0%)	55 273.1 (105.8%)	5.8%
От варианта 1: 60% – переписчики с МЧД, 40% – переписчики с планшетными компьютерами, больше нагрузка на полевых работников (ГПД больше на 10%), нет затрат на ПО Интернет-переписи, дополнительные рабочие станции для обработки 10% МЧД						
4. Два метода сбора (МЧД и Интернет)	743.4 (200.0%)	38 292.2 (115.5%)	6 642.3 (71.0%)	9 364.8 (100.0%)	55 042.7 (105.3%)	5.3%
От варианта 1: 90% населения – переписчики с МЧД, тираж МЧД больше в два раза, ГПД больше на 8% – больше нагрузки на переписчиков и автоматизаторов, больше рабочих станций на районном и региональном уровнях для обработки дополнительного количества 100% МЧД						

13. Предполагаемая последовательность проведения ВПН-2020: 1–10 октября 2020 года проходит Интернет-перепись.

14. Люди, изъявившие желание пройти перепись через Интернет, не должны быть потеряны или учтены дважды. В этой связи каждый участник переписи должен быть идентифицирован. Для решения этой задачи предполагается использование уже сложившихся способов идентификации личности на Портале Госуслуг через Единую Систему Идентификации и Аутентификации (ЕСИА).

15. ЕСИА представляет собой надежный государственный ресурс, агрегирующий персональные данные населения из различных ведомственных источников данных. Особенность его в том, что в этой системе содержатся данные лишь тех людей, которые захотели в ней зарегистрироваться. Но это как раз те люди, которые предпочитают использовать электронные государственные услуги взамен традиционного личного общения с госорганами.

16. После входа на страницу Интернет-переписи респонденту потребуется выбрать свое жилое помещение из списка адресов, подготовленного Росстатом при переписном районировании. Один из постоянно проживающих в жилом помещении может заполнить электронные переписные листы на себя и осталь-

ных жителей этого помещения, которые уполномочили его передать сведения о себе для целей переписи населения.

17. Электронный переписной лист должен быть также дополнен вопросом о согласии респондента о передаче данных через Интернет.

18. ЕСИА представляет возможность предзаполнить часть ответов электронного переписного листа на основании хранящихся в ЕСИА персональных данных респондента, а именно:

- пол;
- дата рождения;
- место рождения;
- гражданство.

19. Указанные сведения могут быть перенесены в электронный переписной лист автоматически с согласия респондента. При этом в процессе прохождения федерального статистического наблюдения у респондента будет возможность их проверки и корректировки.

20. Во время Интернет-переписи и еще 4 дня после ее завершения (до 15 октября 2020 года) Росстат актуализирует электронные списки адресов, отмечая адреса жилых помещений, в которых население прошло Интернет-перепись, и внося число переписанных мужчин и женщин, число домохозяйств.

21. С 16 по 31 октября 2020 года переписчики обходят все помещения своих счетных участков, фиксируют подтверждение прохождения переписи по Интернету и заполняют электронные или бумажные переписные вопросники на население, не прошедшее Интернет-перепись.

22. Респонденты, прошедшие Интернет-перепись будут получать коды подтверждения, которые также будут вноситься в информацию по соответствующим помещениям в хранилище адресов. При посещении жилых помещений, по которым имеются коды подтверждения прохождения переписи населения, переписчик будет сверять коды у респондентов с кодами в его адресном списке и, при необходимости, корректировать информацию (например, в случае, если один из членов домохозяйства не переписался по Интернету). Коды подтверждения будут содержать в себе в зашифрованном виде информацию из электронного переписного листа о реквизитах переписного листа, идентифицирующих домохозяйство, поле респондента и его родственных отношениях с членами домохозяйства. Это обеспечит возможность контроля информации по членам домохозяйства, переписанным по Интернету и при опросе переписчиком. ФИО респондента ни в переписных листах, ни в списках адресов фиксироваться не будут.

III. Новые вызовы защиты конфиденциальности при внедрении инноваций переписи населения 2020 года

23. Главная задача защиты конфиденциальной информации при проведении переписи населения с использованием портативных компьютеров и в сети Интернет – это защита первичных данных переписи с целью максимально обезопасить их от несанкционированного доступа.

24. С технологической точки зрения встает вопрос о размещении программного обеспечения Интернет-переписи: его можно встраивать в инфраструктуру

самого Портала Госуслуг, а можно – на специально разработанном сайте Интернет-переписи, куда человек будет переадресовываться с Портала Госуслуг.

25. Основными плюсами размещения программного обеспечения Интернет-переписи на Портале Госуслуг является то, что данный портал уже обладает техническими средствами обеспечения защиты конфиденциальности и персональных данных, защиты от DoS-атак, а также достаточной производительностью. А при размещении программного обеспечения на специально разработанном сайте Интернет-переписи необходимы дополнительные вложения для обеспечения защиты конфиденциальности данных, защиты сайта от DoS-атак и обеспечения необходимой производительности программного обеспечения.

26. Основными плюсами размещения программного обеспечения на специально разработанном сайте Интернет-переписи является то, что нет технических ограничений по внешнему виду и функциональности электронных переписных листов, а также по алгоритмам контроля ответов на вопросы переписных листов. Практика реализации программного обеспечения для проведения федеральных статистических наблюдений показывает, что часть правил интерактивного контроля заполнения ответов на вопросы изменяются непосредственно в процессе проведения федерального статистического наблюдения, когда реально поступающие данные отличаются от предположений разработчиков. Процесс согласования изменений и реализации изменений в электронные формы, размещенные на Портале Госуслуг, затрагивает многих участников и занимает продолжительный период времени. Подобные задержки могут быть критичными при проведении переписи.

27. Таким образом, мы выбираем вариант специально разработанного сайта Интернет-переписи с использованием ЕСИА как средства доступа.

28. Для обеспечения безопасности и конфиденциальности сбора данных в ходе Интернет-переписи планируется применять трехфазную авторизацию респондентов:

- авторизация респондента по учетной записи ЕСИА – проверка полномочий респондента по его доступу к полному набору государственных услуг в электронной форме;
- прохождение респондентом теста Тьюринга (CAPTCHA) – защита от автоматизированного заполнения переписных листов в электронной форме;
- идентификация по уникальным номерам помещения и респондента – установление связи между респондентом и всеми переписными листами в электронной форме, которые он заполнит, и Списком адресов помещений для проведения Всероссийской переписи населения 2020 года.

29. Возможно также применение следующих способов защиты:

- персональный идентификационный код, позволяющий заполнять онлайн переписной лист в течении нескольких интернет-сессий (более чем за один раз);
- секретный вопрос, создаваемый домохозяйством и позволяющий прервать и продолжить заполнение онлайн переписного листа.

30. При предоставлении ответов через Интернет будет разработан целый ряд методов контроля качества данных и корректировки собранных данных: автоматическое определение последовательности соответствующих вопросов, интерактивное редактирование в целях увязки проблемных ответов и использование выпадающих списков для обеспечения соответствия вводимых данных до-

пустимым вариантам. В отдельных случаях используются другие методы, такие как автоматизированное кодирование.

31. При использовании планшетных компьютеров необходимо заранее решить несколько вопросов связанных с безопасностью:

- вид хранения данных (на устройстве или мгновенная передача в центральную систему);
- способы передачи данных с персонального компьютера (ПК) на рабочие станции регионального уровня, а также с регионального на общий сервер;
- безопасность данных, хранящихся в ПК.

32. На ПК переписи населения 2020 планируется устанавливать операционную систему Android – широко распространенную бесплатную ОС, успешно применяемую Росстатом в обследованиях с 2012 года.

33. При этом согласно национальным требованиям по защите мобильных технических средств, применяемых в государственной информационной системе, на планшетных компьютерах, кроме основной ОС Android, будет установлена отдельная защищенная операционная система ОС Android:

(a) Не допускающая установку стороннего, неразрешенного ПО без согласования с Росстатом;

(b) Обеспечивающая работу с отключенными каналами связи и беспроводными сетями – Bluetooth и WiFi;

(c) Обеспечивающая конфиденциальность информации о населении.

34. Вопрос безопасности данных, хранящихся на устройстве, остро встает в случае потери или кражи ПК.

35. В этих случаях применяются два способа защиты.

36. Первый способ – шифрование данных. Все данные находящиеся на устройстве находятся в зашифрованном виде, и для просмотра требуется “ключ” дешифрования. Возможен вариант, когда после заполнения анкеты шифруются так, что даже переписчик не может внести туда исправления.

37. Второй способ – пароли. Доступ к программному обеспечению осуществляется по паролю переписчика соответствующего счетного участка. Альтернативой паролям является устройство считывания отпечатков пальцев, которое обеспечивает высокую степень защиты, но значительно повышает конечную стоимость устройства.

38. Для сохранности информации, собранной на ПК в течение периода переписи, ПК оснащаются съемными microSD-картами памяти, на которые производится резервное копирование информации и которые переписчики должны периодически сдавать контролерам, заменяя на чистую карту памяти. Информация с планшетных компьютеров будет передаваться в систему с использованием проводного (usb) подключения к рабочей станции в закрытом защищенном аттестованном контуре или на sd-картах с использованием аттестованных ноутбуков.

39. Защита каналов связи, используемых для соединения оборудования внутри автоматизированной системы, реализуется средствами системы криптографической защиты информации.

40. Для переписи населения 2020 года потребуется доработка автоматизированной системы переписи 2010 года с обязательной апробацией ее при пробной

переписи населения в 2018 году. При этом помимо обновления существующих программных и технических средств предстоит разработка совершенно новых блоков и модулей:

- автоматизация функций полевого уровня (уровень переписных, инструкторских и счетных участков) и уровня Интернет-сбора данных переписи;
- мультимедийные средства обучения, в том числе и для респондентов Интернет-переписи;
- обеспечение сбора и обработки данных Интернет-переписи;
- аппаратно-программные средства защиты информации.

41. Таким образом, инновационные методы проведения переписей населения создают новые вызовы по защите конфиденциальности данных переписи, внедряя информатизацию не только на этап обработки собранных данных, как это было при традиционной переписи, но и на все этапы подготовки и проведения переписи на полевом уровне. Этот фактор оказывает существенное влияние на стоимость переписи в целом, удорожая ее технологическую составляющую: