

# Новые технологии сбора данных с помощью электронной переписи населения, часть 4: разработка спецификации электронной анкеты

*Выбор тем в международных переписях населения<sup>1</sup>*

Выпущено в октябре 2021 года

## ВВЕДЕНИЕ

Необработанные данные содержат ошибки, которые необходимо исправлять до публикации или анализа наборов данных. В рамках переписи населения это делается на этапе редактирования данных. Редактирование данных — это процедура обнаружения и исправления ошибок или логических несоответствий. Редактирование данных также помогает выявить проблемы, связанные с отсутствующими и неправильными ответами. Обычно эти проблемы обычно решаются с помощью подстановки данных. Разработка подробных требований к личным опросам с использованием компьютера (CAPI, в соответствии с английским акронимом) важна для сохранения исходных данных и сведения к минимуму количества отредактированных и подставленных ответов в наборах данных переписи населения. В настоящей технической записке по отдельным темам международных переписей (STIC, в соответствии с английским акронимом) содержатся руководящие указания для национальных статистических служб (NSO, в соответствии с английским акронимом) в отношении методов разработки требований к электронным анкетам.

## ВВОДНАЯ ИНФОРМАЦИЯ

Процесс редактирования данных переписи населения со временем изменился. Первоначально для редактирования анкет требовалось большое число сотрудников, которые делали это вручную. Задача была трудоемкой, сложной и подверженной ошибкам. Сложные корреляции между переменными затрудняли для сотрудников редактирование всех расхождений в данных. Разные сотрудники, а зачастую один и тот же, могли по-разному толковать правила (United Nations, 2020).

Внедрение компьютеров для редактирования данных сделало процесс менее трудоемким и более точным, а применение правил редактирования стало более упорядоченным. Компьютеры также помогают вносить изменения в оцифрованные записи благодаря автоматизации.

<sup>1</sup> Настоящая техническая записка является одной из серии «Избранные темы международных переписей населения» (STIC), в которой рассматриваются вопросы, представляющие интерес для международного статистического сообщества. Бюро переписи населения США помогает странам совершенствовать национальные системы статистики путем содействия в устойчивом развитии статистических компетенций.

Еще одним важным достижением стало внедрение электронного сбора данных. Использование персональных компьютеров и планшетов для ввода ответов респондентов позволяет свести к минимуму ошибки, в том числе благодаря тому, что CAPI дает возможность предварительно запрограммировать порядок следования вопросов с учетом предыдущих ответов, а также проверять корректность данных, вводимых в поле.

При проведении переписей с помощью бумажных анкет необходимо обучать сотрудников оценивать целесообразность включения того или иного вопроса либо раздела для конкретного лица или домохозяйства. Кроме того, сотрудникам приходится сверять ответы с предыдущими или последующими для определения их достоверности, логичности и последовательности. Перепись населения — объемный и сложный процесс, потому ошибки неизбежны. По этой причине после переписи необходимо выполнять редактирование данных. При использовании компьютера объем редактирования данных можно уменьшить за счет проверок на согласованность, иначе называемых проверками на достоверность, эффективность которых зависит от качества разработки спецификации редактирования.

## ПОРЯДОК ИСПОЛЬЗОВАНИЯ СРЕДСТВ CAPI ПРИ ПРОВЕДЕНИИ ЭЛЕКТРОННОЙ ПЕРЕПИСИ

Переход от бумажных методов сбора данных к CAPI влияет на сроки подготовки окончательного варианта спецификации редактирования и на состав группы экспертов, которые их разрабатывают. Качественную спецификацию необходимо подготовить на ранней стадии CAPI, а в группу по их разработке должны входить как специалисты в предметной области, так и программисты. В поле 1 приведены более подробные сведения по этим вопросам из документа STIC с руководящими указаниями для национальных статистических служб по планированию электронной переписи населения (U.S. Census Bureau, 2016a).

Поле 1.

## Подготовка к мобильной фиксации данных

### Расписание переписи населения

Чтобы обеспечить успех переписи населения, ее расписание следует адаптировать в соответствии с потребностями в реализации API. Программирование анкеты нельзя начинать до того, как будет разработана спецификация анкеты, однако учитывать задачи программирования лучше еще на ранних стадиях разработки анкеты. При мобильной фиксации данных процессы, которые ранее были отдельными, могут быть объединены или перенесены на более раннюю стадию цикла переписи населения. Например, при мобильной фиксации данных сбор, фиксация и редактирование данных могут выполняться одновременно. Однако в общем случае на разработку и тестирование приложения, настройку систем передачи и обработки данных, а также закупку, программирование и тестирование мобильных устройств уходит больше времени. Если в приложении реализуются проверки редактирования, их программирование необходимо завершить до начала переписи — не следует программировать их отдельно, уже после ее завершения. Кроме того, следует выделить больше времени на обучение переписчиков, поскольку в программу обучения необходимо включить курс по работе с мобильными устройствами. Соответственно, в числе важнейших задач — определить все этапы настройки системы мобильной фиксации данных и выделить на это достаточное время перед проведением переписи в ее расписании.

### Отличия процесса разработки анкеты

Как и в случае с бумажной анкетой, процесс разработки электронной анкеты является итеративным. Ее необходимо разработать, подвергнуть тестированию, доработать и снова протестировать, повторяя цикл до тех пор, пока анкета не начнет функционировать, как задумано. В случае электронной анкеты помимо содержания необходимо также тестировать и дорабатывать технические особенности приложения. После того, как специалисты в предметной области выработают окончательный вариант содержания анкеты, необходимо подготовить спецификацию, которая станет руководством для программистов, реализующих приложение. Кроме того, необходимо составить документацию надлежащего качества по инструменту анкетирования. Отсутствие бумажной документации затрудняет обсуждение анкеты с различными заинтересованными лицами, а также приводит к тому, что тестирование длится дольше и становится более подверженным ошибкам ввиду меньшей очевидности шаблонов пропуска. Кроме того, электронная анкета может иметь дополнительные функции, отсутствующие у бумажной, такие как проверка достоверности данных и сообщения об ошибках. Требования к этим функциям необходимо описать таким образом, чтобы программисты могли реализовать приложение согласно замыслу. При проектировании электронной анкеты специалисты в предметной области могут допускать ошибки, связанные с формулировками, макетом и архитектурой инструмента. Поэтому специалисты в предметной области должны работать в тесной связке с программистами, чтобы обеспечить обмен точными сведениями о содержании, макете и архитектуре анкеты, проверке достоверности данных и других требованиях. Важны также наличие четкого графика работ; доступность сведений об изменениях, вносимых в содержание анкеты по ходу разработки; обеспечение защищенности данных и контроля качества.

Источник: U.S. Census Bureau, 2016a.

Поскольку электронная анкета позволяет выполнять автоматические проверки достоверности данных и не имеет ограничений по размеру, вопросы содержания и структуры такой анкеты будут иными, чем в случае с бумажной. В поле 2 описано, каким образом процедуры проверки достоверности данных встраиваются в электронную анкету на основе документа STIC по разработке таких анкет (U.S. Census Bureau, 2016b).

## ПРОВЕРКИ СОГЛАСОВАННОСТИ

Ошибки при переписи населения могут возникать в связи с охватом и содержанием. Ошибки охвата возникают из-за пропуска или двойного учета людей при переписи, тогда как ошибки содержания появляются из-за того, что респонденты предоставляют неверную информацию, переписчики некорректно задают вопрос или неправильно понимают ответ, либо переписчики выбирают неверный вариант ответа.

Ошибки охвата обнаружить непросто. Один из способов — составить список всех домохозяйств и их членов на территории переписи еще до ее проведения, а затем сравнить информацию из такого списка со сведениями, собранными при переписи. Ошибки содержания обнаруживаются проще. Правильно разработанное приложение API позволяет значительно снизить число ошибок содержания за счет автоматических проверок согласованности.

Назначение таких проверок в API — определить наличие внутренних несоответствий в данных. Проверки согласованности помогают переписчикам обнаруживать ошибки во время работы на местах и сразу же исправлять их. Например, необходимо запрограммировать проверку согласованности, которая не позволит переписчику ввести большее число в поле «Количество лет проживания», чем число, указанное в поле «Возраст».

Чтобы учесть всю сложность процесса сбора данных при переписи, необходимо, чтобы в разработке проверок на согласованность помимо специалистов по обработке данных участвовали эксперты в предметной области и специалисты по проведению опросов. Правильно разработанное приложение САРІ должно предусматривать достаточное количество проверок на согласованность, чтобы предотвратить появление противоречивых данных. В то же время оно должно позволить переписчикам оперативно вводить данные, не создавая для них помех. Наличие проверок данных переписи на согласованность в рамках САРІ не отменяет необходимости в редактировании данных после фиксации, но правильно реализованные проверки уменьшают объем такого редактирования.

## Типы проверок

Проверки данных на согласованность могут быть двух типов — жесткими и мягкими. Жесткие проверки не позволяют переписчику перейти к следующему вопросу до разрешения проблемы, а мягкие проверки только предупреждают о возможной некорректности данных. Жесткие проверки необходимо применять только при абсолютной уверенности в том, что допущена ошибка, например, возраст вступления в первый брак не может быть больше возраста респондента. Наихудшая ситуация с жесткой проверкой — когда она неверно запрограммирована и переписчик не может продолжить работу, введя корректный ответ; в этом случае переписчик может быть вынужден ввести неверные данные, чтобы продолжить опрос.

Когда переписчик сталкивается с мягкой проверкой, он должен проверить ответ, чтобы выяснить, следует ли ему продолжить опрос или исправить ответ. Мягкие проверки оптимальны для ситуаций, когда чаще всего ответ верный, однако иногда могут быть исключения. Например, можно реализовать мягкую проверку, чтобы предупреждать переписчиков, когда они вводят ответ, находящийся в допустимых пределах, но являющийся нетипичным, как в случае указания возраста более 100 лет. При наличии сомнений разработчики инструмента САРІ должны добавлять мягкую проверку для предупреждения переписчика, а не жесткую.

Таблица 1.

### Время выполнения проверок на согласованность различных типов

Характеристика	Зависимость	Время выполнения	Пример
Единичные	Зависят только от ответа на конкретный вопрос. Обычно это простой контроль попадания в интервал.	Могут выполняться сразу после получения ответа.	Находится ли ответ в списке допустимых кодов?
Уровня вхождения	Зависят от других ответов, полученных в разделе для одного вхождения.	Могут выполняться после того, как введены все зависимые данные или после того, как введены все данные для вхождения.	Является ли возраст лица при вступлении в первый брак меньшим или равным возрасту лица?
Уровня раздела	Зависят от других ответов, полученных в разделе для всех вхождений.	Могут выполняться после того, как введены все зависимые данные для всех вхождений в разделе или после того, как введены все данные для всех вхождений.	Указано ли единственное лицо в качестве главы домохозяйства?

Источник: U.S. Census Bureau.

Поле 2.

## Разработка электронной анкеты

### Проверка достоверности данных

Одно из преимуществ электронной анкеты в том, что она позволяет проверять достоверность данных по мере того, как переписчик вводит ответы в мобильное устройство. Для этого правила проверки данных на достоверность должны быть написаны специалистами в предметной области на этапе разработки анкеты, чтобы их можно было запрограммировать в приложении. Именно специалисты в предметной области должны писать правила проверки, поскольку они обладают обширными знаниями по вопросам переписи и возможным ответам.

Источник: U.S. Census Bureau, 2016b.

## Зависимость проверок

Обычно проверку данных желательнее выполнять как можно быстрее, чтобы переписчик и респондент могли исправить ошибку во время опроса. Проверки согласованности также можно классифицировать в зависимости от времени выполнения в анкете, как единичные, уровня вхождения<sup>2</sup> и уровня раздела.<sup>3</sup> В таблице 1 приведены зависимости и время выполнения проверок.

<sup>2</sup> Вхождение — это появление одного и того же элемента данных в анкете несколько раз. Например, вопросы «кем приходится», «пол» и «возраст» будут заданы о каждом члене домохозяйства. В домохозяйстве с пятью членами будет по пять вхождений каждого из этих элементов данных. «На уровне вхождения» означает, что проверку необходимо выполнять для конкретного вхождения. В приведенном примере данные по каждому члену домохозяйства будут составлять вхождение.

<sup>3</sup> Раздел — это группа связанных элементов, например, данные об образовании, экономические данные и сведения о фертильности. Раздел может охватывать все вхождения. Например, данные об образовании указываются для всех лиц в домохозяйстве старше определенного возраста; сведения о фертильности указываются для всех женщин (т.е. для всех вхождений данных о женщинах) в домохозяйстве. В приложении САРІ может быть предусмотрен механизм, принуждающий к вводу всех данных в этом разделе перед тем, как разрешить переход к следующему.

## Неизвестные и частично известные ответы

Важно также определиться, следует ли разрешить ответ «отсутствующий» или «неизвестный» в качестве корректного в анкетах САPI. Для гибкости и возможности фиксации максимального объема информации для некоторых вопросов можно разрешить «частично известные» ответы. Например, вместо категории «Возраст неизвестен» может быть полезным создать следующие категории: «Возраст неизвестен, но ребенок еще не ходит в школу», «Возраст неизвестен, но респондент — взрослый трудоспособного возраста» или «Возраст неизвестен, но вероятно, больше 50 лет». Частично известные ответы позволяют переписчикам собирать частичные данные вместо того, чтобы оставлять ответ незаполненным. Частично известные ответы можно будет заменить на определенные ответы из известных пределов значений после сбора данных.

## Область проверок

При разработке проверок на последовательность важно иметь в виду, что проверки ограничивают свободу ввода данных переписчиками. В некоторых случаях проблемы лучше устранять после фиксации, а не во время сбора данных, чтобы избежать добавления чрезмерного количества проверок. До проведения переписи населения необходимо выполнить подробный анализ всех проверок на последовательность. Такой анализ проводят главным образом на этапах предварительного тестирования анкеты и в ходе пилотной переписи населения.

Результаты предварительного тестирования анкеты и пилотной переписи позволяют:

- проанализировать, какие проверки срабатывали чаще всего;

Таблица 2.

## Спецификация электронной анкеты

Характеристика	Описание
Номер	Номер соответствующего вопроса в бумажной анкете.
Переменная	Укажите имя переменной для файла данных САPI (напр., P01_ИМЯ_1, P01_ИМЯ_2, ...).
Подпись	Укажите сообщение, которое появится вверху на экране планшета.
Текст вопроса	Точная формулировка вопроса, который будет прочитан вслух переписчиком.
Инструкции для сотрудника	Любые конкретные указания для сотрудника.
Заполнение	Укажите, как будет отображаться заранее загруженный элемент, например, [ИМЯ].
Область применения	Укажите лиц или объекты, к которым применим данный вопрос (например, жители в возрасте от 3 лет).
Ответы	Укажите все возможные ответы в зависимости от типа вопроса.
Маршрутизация	Укажите, какой вопрос следует задать следующим, и кому. Не забудьте включить особые маршруты для ответов вроде «Другое (укажите)».
Проверки	Для каждого вопроса укажите необходимые проверки на согласованность и полноту.
Сообщения об ошибке	Укажите, какие сообщения следует выдавать для конкретной проверки на согласованность или попадание в интервал.
Инструкции для программиста	Изложите особые указания для программистов.
Меню подсказки	Если программа должна содержать меню подсказки, его состав также необходимо указать.
Журнал изменений	Опишите вкратце внесенные изменения. Укажите дату внесения изменений и фамилию программиста, который это сделал.

Источник: U.S. Census Bureau.

- выяснить наличие явно противоречивых данных, которые следует проверять в будущих версиях инструмента САPI;
- узнать у переписчиков, мешали ли проверки нормальному процессу сбора данных;
- выяснить, какой объем данных был подвергнут мягким проверкам и определить, следует ли сохранить ту или иную мягкую проверку, не нужно ли изменить пределы проверки и не следует ли изменить тип проверки на жесткий.

## Разработка спецификации редактирования

При составлении спецификации указываются проверки, которые должны выполняться приложениями для переписей населения САPI. Спецификация служит четырем основным целям: (1) эксперты в предметной области могут описывать в ней необходимые проверки для программистов и других участвующих специалистов, (2) программисты могут пользоваться спецификацией при реализации инструмента, (3) эксперты в предметной области могут проверять проверки на точность и полноту, (4) пользователи данных после их фиксации могут выяснять сведения о параметрах, которые учитывались при сборе данных.

Спецификация редактирования должна включать как минимум элементы, перечисленные в таблице 2 (полное руководство по составлению спецификации электронной анкеты см. в приложении). В таблице 3 приведен полный набор требований к гипотетическому вопросу «Кем вам приходится глава домохозяйства?»

Таблица 3.

**Образец требований к вопросу о степени родства с главой домохозяйства**

Требование	Образец
Номер	A03
Имя переменной	A03_РОДСТВО
Подпись	Степень родства с главой домохозяйства.
Текст вопроса	Кем [ИМЯ] приходится главе домохозяйства?
Заполнение	A01_ИМЯ
Тип данных	Числовой
Тип фиксации	Кнопка-переключатель
Ширина поля	2
Инструкции для сотрудника	Заполнение этого поля является обязательным
Область применения	Все члены домохозяйства в списке.
Ответы	00- Глава 01- Супруг или партнер 02- Сын или дочь 03- Зять или невестка 04- Пасынок или падчерица 05- Внук или внучка 06- Родитель 07- Тесть или теща 08- Бабушка или дедушка 09- Брат или сестра 10- Племянник или племянница 11- Другой родственник 12- Не является родственником
Маршрутизация	При появлении сообщения об ошибке (E1) вернуться к A01_ИМЯ и правильно указать главу домохозяйства. Иначе перейти к A04_SEX.
Проверки на согласованность	1. В каждом домохозяйстве должен быть ровно 1 глава. 2. Проверить, является ли первое лицо в списке главой домохозяйства.
Сообщение об ошибке	E1: Первое лицо в списке должно быть главой домохозяйства.
Указания для программиста	Мягкая проверка.
Меню подсказки	1. Глава домохозяйства — это лицо в возрасте от 12 лет, которое считается другими членами домохозяйства его главой. Для исключительных случаев следует разрешить указание возраста главы домохозяйства меньше 10 лет. 2. У главы домохозяйства может быть более одной супруги.
Журнал изменений	Неправильное написание в поле «Текст вопроса» исправлено 25.03.2018 by John Doe. Проверка на согласованность добавлена 28.05.2019 by Dr. Jane Doe.

Источник: U.S. Census Bureau.

## ЗАКЛЮЧЕНИЕ

Процесс сбора данных подвержен ошибкам содержания. Инструменты САРІ имеют преимущества перед бумажными анкетами, так как позволяют встраивать заранее запрограммированные операции редактирования, повышающие качество данных на этапе сбора. Для разработки качественной спецификации инструмента САРІ необходимо взаимодействие специалистов в предметной области, специалистов по опросам на местах и программистов. Разработка спецификации инструмента САРІ имеет критически важное значение, поскольку данный документ позволяет экономить время и ресурсы и служит ценным источником сведений о принятых решениях. Цель состоит в том, чтобы улучшить качество данных в источнике, уменьшив объем редактирования и подстановки данных после сбора.

## ЛИТЕРАТУРА

United Nations Statistics Division, “Handbook on the Management of Population and Housing Censuses, Revision 2,” New York, 2016.

United Nations Statistics Division, “Principles and Recommendations for Population and Housing Censuses, Revision 3,” United Nations Publications, New York, 2017.

United Nations Statistics Division, “Handbook on Population and Housing Census Editing, Revision 2,” New York, 2020.

U.S. Census Bureau, “New Technologies in E-Census Data Collection Part 1: Planning for Mobile Data Capture, 2016a,” <[www.census.gov/library/working-papers/2016/demo/mobile-data-1.html](http://www.census.gov/library/working-papers/2016/demo/mobile-data-1.html)>, accessed July 6, 2021.

U.S. Census Bureau, “New Technologies in E-Census Data Collection Part 2: Developing an Electronic Questionnaire, 2016b,” <[www.census.gov/library/working-papers/2016/demo/mobile-data-2.html](http://www.census.gov/library/working-papers/2016/demo/mobile-data-2.html)>, accessed July 6, 2021.

U.S. Census Bureau, “New Technologies in E-Census Data Collection Part 3: Timeline Impacts, 2019,” <[www.census.gov/library/working-papers/2019/demo/new-technologies-data-collection-part3.html](http://www.census.gov/library/working-papers/2019/demo/new-technologies-data-collection-part3.html)>, accessed July 6, 2021.

Приложение  
**Спецификация электронной анкеты**

Требование	Описание
Номер	Номер соответствующего вопроса в бумажной анкете.
Переменная	Укажите имя переменной для файла данных API (напр., P01_ИМЯ_1, P01_ИМЯ_2,...).
Подпись	Укажите сообщение, которое появится вверху на экране планшета.
Текст вопроса	Точная формулировка вопроса, который будет прочитан вслух переписчиком. Если бумажная анкета была выполнена в формате таблицы, возможно, понадобится написать подробный вопрос (пример: «Перечислите имена всех людей, которые ночевали в домохозяйстве в ночь переписи населения, начиная с главы домохозяйства»). Укажите заранее заполненные вопросы и вариант формулировки (например, «Где родился [ИМЯ]?») (См. графу «Заполнение» ниже).
Инструкции для переписчика	Любые указания для переписчика. Укажите, что они не предназначены для прочтения респонденту. Укажите на необходимость использования другого шрифта или цвета шрифта, чтобы отличать инструкции от вопросов. Здесь также необходимо указать точные формулировки наводящих вопросов.
Заполнение	Укажите, как будет отображаться заранее загруженный элемент, например, [ИМЯ]. Таким образом обозначаются заполнения. Укажите источник входных данных (например, имена переменных, внешний источник данных).
Область применения	Укажите лиц или объекты, к которым применим данный вопрос (например, жители в возрасте от 3 лет).
Ответы	Для категоричных ответов укажите все возможные ответы, которые могут быть даны на поставленный вопрос. Укажите допустимые пределы для числовых полей с возможностью ввода произвольных чисел (для контроля попадания в интервал). Укажите, должна ли проверка попадания в интервал быть мягкой или жесткой. Укажите формат ввода для дат, номеров телефона и т. д. Укажите, должен ли быть выбран только один ответ или разрешена возможность выбора нескольких. Укажите графический элемент управления для ответа (например, кнопка-переключатель, флажок и т. д.). Укажите максимально допустимое количество символов (т.е. ширину поля) для ответов с произвольным числом символов. Укажите, допустимы ли ответы «Не знаю», «Отказываюсь отвечать» и незаполненный ответ. Если не предусмотреть возможность оставить поле незаполненным, CSPro не примет такой ответ и не позволит переписчику перейти к следующему вопросу.
Маршрутизация	Укажите, какой вопрос должен быть задан следующим и кому. Не забудьте указать особые маршруты для ответов вроде «Другое (укажите)».
Проверки	Для каждого вопроса укажите необходимые проверки на согласованность и полноту. Указывайте конкретные переменные или ответы, участвующие в проверке. Укажите, должна ли проверка быть мягкой или жесткой.
Сообщения об ошибке	Укажите, какие сообщения следует отображать для конкретной проверки на согласованность или попадание в интервал. Укажите маршрутизацию для случая, когда проверка не пройдена. Обеспечьте возможность продолжения опроса в случае, если проблему разрешить невозможно.
Инструкции для программиста	Изложите особые указания для программистов. Укажите требования к циклам. Укажите, сколько раз программа должна воспроизвести серию вопросов (например, количество мест работы, которое следует выяснить для вопроса о трудовой биографии).
Меню подсказки	Если программа должна содержать меню подсказки, его состав также необходимо указать.
Журнал изменений	Опишите вкратце внесенные изменения. Укажите дату внесения изменений и имя программиста, который это сделал.

Источник: U.S. Census Bureau.



**USAID**  
 FROM THE AMERICAN PEOPLE



Документы серии «Избранные темы международных переписей населения» (STIC) публикуются в рамках Международных программ Отдела по народонаселению Бюро переписи населения США. Агентство США по международному развитию финансирует выпуск серии документов STIC и оказание двусторонней поддержки статистическим организациям, предоставляющим сведения авторам. Фонд Организации Объединенных Наций в области народонаселения оказывает содействие в подготовке содержания и распространении документов серии STIC, способствуя их доступности для широкой аудитории.