



## ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

Распространение: Общее  
1 мая 2012 г.  
Оригинал: английский

### Европейская экономическая комиссия ООН

Конференция европейских статистиков

#### 60-е пленарное заседание

Париж, 6-8 июня 2012

#### 2010 раунд переписи – инновации и уроки

### Перепись 2011 года в Республике Чехии: новый опыт

#### Записка Статистического управления Чехии

##### *Резюме*

В данном документе описано проведение Переписи населения и жилого фонда в Республике Чехии для получения основной информации о населении, домах, жилых помещениях и домохозяйствах. В документе описаны новые подходы и методы проведения переписи, в частности, организация полевой работы, модернизация ИТ систем, мониторинг жизненного цикла переписных листов, сканирование и валидация данных. Наконец, в документе приведены предварительные результаты переписи, которая проводилась в соответствии с национальным законом о Переписи населения и жилого фонда №. 296/2009 Sb от 22 июля 2009 г. и Регламентом ЕС №. 763/2008.

### **I. Подготовка данных для проведения полевых работ по переписи**

1. Перепись представляет собой типичную задачу по массовой обработке данных, что означало применение перфокарт и магнитных лент в 1970-х и 1980-х годах и использование сканирования и других ИТ технологий на рубеже веков, за чем последовало использование электронных переписных листов. Решение по электронному распространению и сбору переписных листов было одной из отличительных черт Переписи населения и жилого фонда 2011 года в Республике Чехии. Такая форма переписи была предложена в качестве альтернативы заполнению и возврату бумажных форм.

2. Решение, принятое Статистическим управлением Чехии (ЧСУ) о том, что ЧСУ не будет осуществлять самостоятельно распространение и сбор переписных листов, а поручит это другой организации, было существенным для проведения этого раунда переписи. В качестве подрядчика для проведения полевых работ по переписи в ходе открытого конкурса была выбрана «Почта Чехии».

3. Это означало, что должны были взаимодействовать две разные информационные системы - подрядчика и ЧСУ. ЧСУ должно было подготовить огромный объем исходной информации, которая должна была быть консолидирована, обработана электронными средствами и подготовлена в соответствующей технической форме для предоставления подрядчику. Эта информация также использовалась министерствами, которые принимали участие в Переписи, и всеми муниципальными органами, которые предоставляют информацию населению, а также некоторыми

другими организациями, например Государственной типографии ценных бумагах.

#### **А. Подготовка данных из административных источников**

4. Относительно обработки данных было выбрано классическое решение по использованию реляционной базы данных, которое было основано на имевшихся возможностях ЧСУ. Таким образом, были разработаны две основные базы данных: одна для подготовки данных до начала переписи (подготовка переписи), а вторая – для обработки собранных данных (обработка данных).

5. База данных для подготовки переписи была названа Регистрационной базой данных (RDB). Она была создана для того, чтобы поддерживать центральный регистр всех единиц и объектов переписи в актуальном виде. Таким образом, эта база данных, в первую очередь, содержала информацию о лицах, которые обязаны участвовать в переписи<sup>1</sup>, переписываемых домах и жилых помещениях, сотрудниках переписи, бумажных и электронных переписных листах и их жизненном цикле - от создания до обработки.

#### **В. Территориальная подготовка к Переписи**

6. Во время территориальной подготовки проводились так называемые рабочие встречи полевого  
эком состоянии  
е данных

## **A. Внутренний и внешний информационный портал**

11. Внутренний и внешний порталы образовали пару порталных решений для развертывания приложений Переписи населения и жилого фонда 2011 года или предоставления доступа к ним для пользователей Статистического управления Чехии (внутренний портал) или третьих сторон, а именно «Почты Чехии»<sup>4</sup> (внешний портал).

12. Оба решения использовали технологию Oracle WebLogic Portal 11g, и каждое из них было разработано для выделенной аппаратной инфраструктуры для того, чтобы разделить домены безопасности. Используемые приложения были либо развернуты непосредственно на сервере, либо - из соображений безопасности - функционировали на другом сервере, и портал предоставлял им функциональность «резервного прокси-сервера». В таком режиме портал отвечает на запрос пользователя следующим образом: посылает запрос целевому серверу, берет ответ с сервера и доставляет его пользователю.

## **B. База знаний**

13. База знаний представляла собой систему, основная цель которой состояла в консолидации имеющейся информации по переписи и предоставления ее внешним пользователям, а также внутренним пользователям в ЧСУ. База знаний играла важную роль в распространении и сборе бумажных переписных листов и электронных форм, когда использовалась сотрудниками Контактного центра переписи. ыть до м

14. Функционирование Контактного центра переписи было обеспечено подрядчиком («Почтой Чехии»), включая услуги связи (телефонные операторы, электронная почта, факс). Однако при этом информационная база была предоставлена ЧСУ

## **C. Администрирование запросов**

15. В ходе Переписи населения и жилого фонда 2011 года на стадии распространения и сбора переписных листов было необходимо обеспечить связь между сотрудниками переписи. Стандартный сервис электронной почты был полезен, но его было недостаточно, особенно в следующих ситуациях:

- (a) Когда сообщения по рассматриваемому вопросу нужно было контролировать, и
- (b) Когда сообщения должны были быть доказуемы.

регистрации и мониторинга запросов и управления уведомлениями. Эта система использовалась для связи

для работы не требуются исходные персональные данные, обработке подвергаются только производные данные. Даже в этом случае существует возможность того, что данные об отдельных лицах или конкретных группах населения могут быть раскрыты из агрегированных данных. ЧСУ применяет большое количество методов защиты информации в рамках обеспечения конфиденциальности первичных и вторичных данных;

(с) Отслеживание жизненного цикла переписного листа включает методы и процедуры, которые необходимы для контроля обработки форм и первичных носителей информации. Хотя эти действия были связаны с разработанной политикой безопасности, они оказались важными для сбора данных переписи, что будет показано ниже.

#### **A. Регистрационная база данных**

19. Данные хранились и обрабатывались в так называемой Регистрационной базе данных (RBD). Содержание базы данных было разделено на три области следующим образом:

(a) Интеграция данных – информация и структуры для обмена данными, а именно обмена с подрядчиком по проведению полевых работ<sup>5</sup> и другими системами для поддержки Переписи населения и жилого фонда 2011 года

(b) Интерфейс ввода-вывода для административных источников данных – информация и структуры для поиска данных из административных источников и из систем ЧСУ;

(с) Регистрационные данные – включают регистрационные данные Переписи (лица, дома, жилые помещения, переписные листы, территориальное деление, статистическое деление, характеристики данных единиц).

#### **B. Жизненный цикл переписного листа**

20. Обмен данными о статусе бумажных и электронных переписных листов был очень важен для операций переписи. Для отслеживания жизненного цикла бумажного переписного листа был необходим обмен данными с подрядчиком по проведению полевых работ (через систему ISTEP- информационную систему полевых работ). Подрядчик по проведению полевых работ осуществлял ввод данных с переписных листов. С технологической точки зрения обмен данными выполнялся на уровне Регистрационной базы данных, где осуществлялся механизм пакетного обмена данными для интегрирования данных в такой же среде, как на стороне ISTEP. Использовалось зашифрованное соединение, организованное как туннель в Интернете.

21. После того, как обмен данными был завершен, логика разрешенных состояний и их переходов проверялась на основе предопределенного процесса распространения и сбора переписных листов. Информация о переписных листах, которые не были в предполагаемых состояниях и переходах, отправлялась обратно в ISTEP для определения проблемы. В связи с большим количеством печатных форм, обработка информации о передаче данных и логистике была очень трудоемкой задачей; она выполнялась каждый день с полуночи до утра.

22. Каждой бумажной форме соответствовала электронная форма, которую можно было заполнить в качестве альтернативы. Таким образом, нужно было также отслеживать жизненный цикл электронных форм. Эта задача была, однако, более простой во многих аспектах в связи с более простой моделью состояний и переходов и отсутствием коммуникаций с третьей стороной.

23. Все бумажные формы, которые поступали в сканеры для оцифровки, на первом этапе сканирования были идентифицированы по штрих-коду, и эта информация передавалась в систему. Если бы пришлось ждать доставки переписных листов в конвертах, то они бы поступали на сканирование со значительной задержкой. Решение состояло в том, чтобы вынуть переписные листы из конвертов и поместить в архивные коробки перед доставкой их на сканирование, штрих-коды «складированных» переписных листов считывались считывающим устройством. Эта информация передавалась и загружалась в ISTEP и затем передавалась в информационную систему ЧСУ. Таким образом, ЧСУ получало информацию о возвращенных переписных листах еще до того, как они были отсканированы, и могло в рамках

---

<sup>5</sup> Подрядчик по проведению полевых работ обеспечивал распространение и сбор бумажных переписных листов. «Почта Чехии» была выбрана в ходе открытого тендера.

определенных законом сроков провести расследование и уведомить граждан, не выполняющих свой определенный законом долг об обязательном участии в переписи.

### **С. Сотрудники, задействованные в переписи**

24. Поддерживался реестр лиц (кроме респондентов), которые могут контактировать с переписными листами. Большую часть составляли переписчики, занимающиеся распространением и сбором переписных листов. Переписчики были сотрудниками «Почты Чехии» (ок. 11 тысяч), ЧСУ (ок. 2 тысяч) и некоторых министерств, например министерства внутренних дел, министерства обороны, министерства иностранных дел и министерства юстиции Республики Чехии. Они выполняли задания, связанные с переписью, в областях своей ответственности, определенной законом.

25. Одним из условий, которое должны были выполнять в соответствии с законом ЧСУ, города и муниципальные образования, состояло в требовании в раскрытии для широкой публики списков имен переписчиков и их идентификационных карточек. Цель состояла в том, чтобы дать гражданам возможность проверить личность сотрудника переписи.

## **IV. Сканирование и валидизация данных**

26. В ходе процесса сканирования были созданы электронные изображения бумажных переписных листов как исходных документов для следующих этапов обработки. Операции по сканированию включали:

- (a) Размещение переписных листов;
- (b) Подавление цветного фона;
- (c) Считывание штрих-кода.

27. Приложение, в котором отслеживается жизненный цикл переписного листа, получало данные о штрих-коде. Таким образом, информацию о том, что переписной лист прошел сканирование, получали оперативно. В противном случае нужно было бы ждать, пока переписные листы пройдут валидизацию. Изображения переписных листов создавались в разрешении 300 dpi. Формат TIFF позволяет сохранять несколько страниц изображений в одном файле. В связи с большим размером файлов и «ценой» этих цифровых изображений, изображения в формате TIFF непрерывно сохранялись в качестве файлов в медиа-архивах.

28. При сканировании переписные листы должны обрабатываться таким образом, чтобы обеспечить отслеживание их жизненного цикла и предотвратить потерю. Переписные листы забирались из центрального склада переписных листов ЧСУ в транспортировочных коробках «Почты Чехии» (переписные листы, собранные переписчиками) или в архивных коробках (формы, присланные на абонентский ящик Переписи).

### **А. Распознавание, верификация и валидизация**

29. Распознавание информации, внесенной в переписные листы, и ее верификация относительно данных регистра, особенно адресов и имен. Верификация данных, в основном, проводилась путем поиска данных в регистре по принципу схожести, при этом было использовано приложение Netrics Matching Engine<sup>6</sup>. В ходе обработки изображений процедура верификации была заменена на другие алгоритмы с использованием более подходящих типов совпадения данных для целей верификации.

30. Валидизация данных, полученных в процессе распознавания изображений, явилась самой масштабной ручной операцией в обработке результатов переписи. Обычно эта работа проводилась в две смены, а в периоды пиковой нагрузки – круглые сутки. В каждой смене работали 60 человек. У каждого был выделенный персональный компьютер с установленным приложения eFlow для выполнения двух видов операций:

- (a) Валидизация отдельных символов – многоэлементная матрица с идентичными символами, идентифицированными во многих формах в ходе автоматического распознавания, показывается валидизатору (проверяющему). Человеческий мозг способен на восприятие всего набора символов, когда простой взгляд

---

<sup>6</sup> [http://www.netrics.com/netrics\\_matching\\_platform.php](http://www.netrics.com/netrics_matching_platform.php)

помогает определить неправильный символ, например «8» вместо «В». Затем валидизатор исправляет такую ошибку;

(b) Валидизация слов в контексте – если в ходе автоматического распознавания символов не удалось идентифицировать символы в слове, то слово было представлено валидизатору в контексте соответствующей ячейки переписного листа. Цель состоит в исправлении нераспознанных символов – если валидизатор ясно определил слово «Prada», а в соответствии с контекстом очевидно, что это слово должно быть «Praha», то будет оставлен первый вариант. Корректировка данных с точки зрения смысла проводилась на первом этапе обработки – при кодировании.

31. Валидизация данных в формате предложений, которые соответствуют информации, записанной в соответствующих формах. Эти данные передавались посредством определенного интерфейса для дальнейшей обработки статистического характера. Каждое предложение данных передавалось в обоих состояниях: до и после валидизации.

## **V. Предварительные результаты Переписи 2011 года в Республике Чехии**

32. На дату переписи численность населения Республики Чехии составляла 10 562 214 человек, на 332 тысячи человек больше, чем по переписи 2001 года.

33. В структуре населения Чехии выросла доля неграждан; их численность достигла почти 450 тысяч человек. За десять лет, прошедших с предыдущей переписи, численность тех, кто не является гражданами Республики Чехии, выросла более чем в 3,5 раза. Почти весь прирост населения с предыдущей переписи произошел за счет увеличения численности неграждан. Одновременно выросла концентрация проживания неграждан по сравнению с 2001 годом, так как более половины неграждан в стране обосновались в столице страны Праге (где проживает 2/5 из них) и в Центральном регионе. В Праге неграждане составляют уже около 14% населения.

### **A. Старение населения**

34. Сравнение структуры населения по результатам Переписей 2001 и 2011 гг. подтверждает тенденцию постепенного старения населения. Доля населения в трудоспособном возрасте (15-64 лет) заметно не изменилась с прошлой Переписи и составила почти 70% от общей численности населения. Численность населения старше трудоспособного возраста (64+ лет) в 2011 году, однако, превысила численность детей (0-14 лет). В Республике Чехии на 26 марта 2011 года было 1 674 тысяч человек в возрасте 65 лет и старше, что составляло почти 16% населения. Численность детей младше 14 лет составляла 1 528 тысяч человек, то есть 14,5%. За десять лет численность населения в трудоспособном возрасте увеличилась почти на 200 тыс. чел., а численность населения старше 65 лет – на 260 тыс. чел. Доля и абсолютная численность детей младше 14 лет монотонно снижалась, начиная с переписи 1980 года. В период между 2001 г. и 2011 г. это снижение было не столь заметным как в предыдущие десятилетия, но тенденция к снижению сохраняется.

35. Структура населения по брачному состоянию различается для мужчин и женщин. Если в 2001 году доля лиц, состоящих в браке, составляла самую большую долю населения (48 и 45% соответственно), то в 2011 г. похожее распределение было верным только для женщин, где самую большую долю составляли замужние женщины (41%), а затем незамужние (35%). Что касается мужчин, то самую большую группу составляли неженатые мужчины (45%), а доля женатых была несколько ниже (43%). И для мужчин, и для женщин численность и доля лиц в браке снизились за десять лет. Напротив, доля одиноких и разведенных выросла. Эти результаты не вызывают удивления в свете данных о количестве браков и разводов за последние двадцать лет. Количество вдов превышает количество вдовцов в пять раз в связи с большей продолжительностью жизни женщин.

### **B. Повышение уровня образования**

36. Тенденция к постепенному повышению уровня образования населения продолжилась. В 2011 году самую большую группу (почти 3 млн. чел) составляли лица со средним образованием без экзаменов уровня А (включая добровольное обучение), хотя численность и доля этой группы снизились за последнее десятилетие. Доля лиц с высшим образованием увеличивалась в долгосрочном периоде. В 2001 году доля лиц со средним образованием с экзаменами уровня А и с высшим профессиональным образованием увеличилась на 15% по сравнению с 2001 г., и их численность составила почти 0 2,8 млн. чел. Это

составляло приблизительно одну треть населения в возрасте старше 15 лет. Рост числа лиц с высшим образованием был еще значительнее и превысил 40%, а их численность составила более 1,1 млн. чел или более 12% населения.

### **С. Национальный состав (этническая структура)**

37. По сравнению с результатами Переписи 2001 года, доля лиц, которые относят себя к основной национальности – чехам – значительно сократилась. В 2001 году приблизительно 9,25 млн. чел. или более 90% населения Республики Чехии, называли себя чехами. В 2011 году национальность «чех» указали в переписных листах 6,73 млн. чел, то есть около двух третей населения. Однако это снижение было только частично в пользу других национальностей. Основной причиной этого снижения явилось то, что значительная часть населения воспользовались юридически предоставленной возможностью не отвечать на вопрос о национальности и религиозной принадлежности. Более 2,74 млн. чел, почти 26% населения, вообще не указали свою национальность. В Переписи 2001 года на вопрос о национальности не было получено ответа только от 1,7% населения. Чешскую национальность указали 92% респондентов в 2001 году и только 86% - в 2011 году от всех ответивших на этот вопрос. Снижение очевидно.

38. Численность населения традиционных национальных меньшинств, таких как словаки, немцы и поляки, также продемонстрировала снижение в той же степени, что и чехов. В 2011 году только 5,2 тысячи человек указали в графе национальность «цыган»; 7,1 тыс. указали две национальности: цыгане и чехи. Всего было около 13 тыс. чел., которые указали национальность «цыган» (включая совмещение). По сравнению с 2001 годом большее количество человек указало некоторые другие национальности, например, «украинцы», «вьетнамцы» и «фрусские».

### **Д. Жилищные условия**

39. На 26 марта 2011 года было зарегистрировано 2,1 миллион домов, то есть строений для проживания (например, частных домов, многоквартирных домов, пансионатов для престарелых и проч.) или служащих для других целей, но имеющих как минимум одно жилое помещение (например, здание школы с жилым помещением), или не принадлежащих ни к одной из перечисленных групп, но тех, которые люди указали как место своего постоянного проживания (например, муниципальные органы власти с офисом уведомления резидентов).

40. В период с 2001 по 2011 более чем для 200 тысяч заселенных домов были отмечены периоды строительства и реконструкции. Новое строительство было явно сконцентрировано в группе семейных домов, часто имеющих одно жилое помещение. Что касается территориального распределения, то доля жилых домов, построенных после 2001 года, была выше средней в Центральном регионе, где она составляла более 16% от всех жилых домов. В других регионах эта доля колебалась вокруг 10-11%. Самая низкая доля нового строительства была в Устецком крае.

### **Е. Состояние жилого фонда**

41. На дату переписи было зарегистрировано 3 894 тыс. жилых помещений, в которых проживали люди. Структура жилого фонда и ее изменения фактически повторяли тенденции и изменения в формах собственности на дома. Преобладание жилых помещений в частных семейных домах соответствовало преобладанию частных домов в жилом фонде. После временного снижения в 1991 году количество жилых помещений в частных домах увеличивалось и в 2011 году превысило 1,4 млн. (последний раз такое наблюдалось в 1980 г.).