Организация Объединенных Наций ECE/CES/GE.41/2016/12

Экономический и Социальный Совет

Distr.: General
29 August 2016

Original : English and Russian
English and Russian only

Европейская экономическая комиссия

Конференция европейских статистиков
Восемнадцатое совещание
Женева, 28−30 сентября 2016 года
Пункт 5 предварительной повестки дня
Методы оценки качества и пригодности для использования регистров и административных источников

Обеспечение взаимодействия инфосистем для проведения переписи на базе регистров

Записка Департамента статистики Эстонии

Резюме

В Эстонии сложилась ситуация, когда данные, имеющиеся в регистрах, покрывают почти все обязательные для переписи признаки и планируется подготовить инфотехнологическое решение, которое позволило бы максимально автоматизировать процесс сбора данных для переписи. Предпосылкой для автоматического получения данных является формат данных и их качество.

Доклад рассматривает возможность обеспечения взаимодействия инфосистем в процессе сбора данных для проведения переписи на базе регистров. Упорядочивание баз данных и данных регистров позволяет экономить в коммерческом процессе, не выполняется вручную в виде файла, если в отношении каждой базы данных данный процесс больше не является особенным.

GE.16-14029 (R)
1614029

Просьба отправить на вторичную переработку

I. Введение

1. За последнее десятилетие автоматизация коммерческой деятельности с помощью инфосистем обусловила накопление большого объема данных, или т.н. революцию данных. Актуальным стал вопрос о создании таких совместимых и способных к взаимодействию инфосистем, которые помогают собирать и находить данные, обрабатывать их и организованно представлять.

2. Настоящий доклад рассматривает необходимость совместимости инфосистем для организации и проведения в Эстонии переписи населения, основанной на базе регистров. Совместимость определяется как способность диспарантных и других различных организаций общаться между собой с целью обмена знаниями и информацией во имя достижения общих значимых целей[1].

3. В докладе рассматриваются возможности решения задачи по совместимости различных инфосистем для достижения поставленных целей, а также ключевая роль Департамента статистики при проведении предстоящей переписи на базе регистров. Кроме того, будут рассмотрены препятствующие факторы для ее проведения в Эстонии и представлена стратегия, позволяющая методически решить поставленную задачу.

4. Опыт двух пробных переписей в Эстонии (в 2014 и 2016 годах) показал, что для организации переписи на базе регистров необходимо в сотрудничестве с партнерами использовать стандарты, улучшать качество данных и развивать семантическую составляющую. Обеспечить совместимость инфосистем проще, если в ведомственных инфосистемах будут использованы созданные со стороны государства системы, например, система адресных данных и система классификаторов.

II. Какие возможности совместимых инфосистем необходимы для проведения переписи на базе регистров?

5. Новые технологии быстро доходят до потребителей, а объем и значение информации растут во многих сферах деятельности. Совместимость инфосистем в статистике следует рассматривать в связи с интеграцией процессов, чтобы подчеркнуть относительное увеличение в инфообществе объемов информации и ее обработки[2]. Стоящая перед инфотехнологиями задача – обрабатывать растущий объем информации, передавать ее быстро и без потерь, при этом обеспечивая ее целостность, доступность для обработки и качество – стала также показателем возможностей к совмещенности на уровне организаций[3]. Тематика необходимости в совместимой работе систем выходит на первый план в целом в двух случаях:

[1] Валлнер, У. (2009). Совместимость регистров. Доклад на XXI Конференции Общества статистики Эстонии.

[2] Klischevski, R. (2004). Information Integration or Process Integration? How to Achive Interoperability in Administration. Electronic Goverment Volume 3183 of the series "Lecture Notes in Computur Science. Pp. 57–65.

[3] Валлнер, У. (2009). Совместимость регистров. Доклад на XXI Конференции Общества статистики Эстонии.

(a) обмен данных на базе информационно-коммуникационных технологий[4];

(b) гетерогенность инфосистем[5].

6. В статистике совместимость инфосистем позволяет создавать более эффективные и экономичные решения. При проведении переписей наибольшие затраты связаны именно со сбором данных, и одной из возможностей по их сокращению служит использование готовых данных, имеющихся в регистрах.

7. В эстонском государстве используются централизованные инфосистемы и регистры, позволяющие получить данные о лицах и жилых помещениях.[6] К предстоящей переписи населения, которая пройдет в 2020 году, и на дальнейший период следует обеспечить совмещенность инфосистем.

8. В Эстонии планируется подготовить такое инфотехнологическое решение, которое позволило бы максимально автоматизировать процесс сбора данных. Для этого необходимо выполнить следующие условия:

(a) эстонские адреса, имеющиеся во всех базах данных, исходят из системы адресных данных Земельного департамента;

(b) все лица, имеющиеся в базе данных, идентифицированы личным кодом (в отношении физических лиц) или кодом Коммерческого регистра, недоходных объединений и целевых учреждений (Коммерческого регистра) в отношении юридических лиц.

(c) политика безопасности данных в государстве предполагает, что в охвате данных используется защищенный обмен данными X-путем (X-tee).

9. Департамент статистики предоставил собственникам/держателям регистров список необходимых действий, которые предстоит осуществить, исходя из оценки уровня качества баз данных, а также указал сроки, когда недостатки должны быть устранены. Департамент статистики проинформировал, что в соответствии с достигнутыми соглашениями для проведения переписи на базе данных регистров собираются данные, состояние которых должно отвечать требованиям статистических работ.

10. Тематика совместимости инфосистем для Департамента статистики связана прежде всего с архитектурой IT-систем, программным обеспечением, синтаксисом, семантикой и набором правил в связи данных, а также с ожиданием, что совместимость инфосистем будет жизнеспособной более продолжительный период и для большинства статистических работ, а не только для проведения переписи населения. В достижении данной цели помогут договоренности различных сторон в следующих аспектах[7]:

- правила,
- стандарты,

[4] Министерство экономики и коммуникаций (2011). Riigi Infosüsteemide koosvõime raamistik. [http://www.riso.ee/et/koosvõime/raamistik] 20.06.2016.

[5] Ouksel, A., Iqbal, A. (1999). Ontologies are not the Panacea in Data Integration: A Flexible. Coordinator to Mediate Context Construction. Distributed and Parallel Database, 7, pp. 1-29.

[6] Тийт, Е-М. (2015). Подготовка к переписи на базе регистров. Eesti Statistika Kvartalikiri nr 3.

[7] Zhao, K. & Xia, M. (2014). Forming Interoperability Through Interorganizational Systems Standards. Journal of Management Information Systems /Spring 2014, Vol. 30, No.4.

- процедуры,
- методы,
- единые или согласованные инфосистемы,
- сети сотрудничества.

III. Интеграция как вызов номер один

11. Инфосистему можно рассматривать как средство, структура которого базируется на правилах, а ее работа направлена на логику сбора количества информации, ее сохранение и выдачу[8]. Человек при этом является частью инфосистемы и для совместного действия необходимо обеспечить архитектуру организации, логическую совместимость с инфосистемами и связанность с производственным процессом[9]. Если любой из компонентов в инфосистеме изменить, может измениться организация, в которой люди инфосистему используют, поэтому необходимо оценить показатели качества инфосистемы относительно ее доступности, постоянства, единства, правильности и своевременности[10]. Наряду с оцениванием качества инфосистем с точки зрения статистических работ по переписи также важны своевременность, точность и правильность показателей[11]. Низкое качество данных может оказывать влияние на совместимость различных инфосистем, хотя при этом необязательно могут возникать проблемы в их технической соединяемости[12]. Так, например, немодернизированные метаданные в статистике переписи могут обусловить перебои в обработке данных, но передача данных из одной инофсистемы в другую при этом осуществляется гладко. Из анкетного опроса, проведенного рабочей группой в 2016 году, выяснилось, что вопрос качества данных занимает центральное место для обеспечения совместимости систем. Можно смело утверждать, что при проведении переписи показатели качества данных оказывают влияние на качество инфосистем и наоборот.

12. Если для переписи на базе данных регистров целесообразно иметь между инфосистемами возможность обмена данными, то следует исходить из возможности, что система является базой работающих объектов во имя общих целей и имеет информацию об операциях ввода и вывода, а общение систем между собой происходит посредством соединения[13].

13. В таком случае при перемещении данных важно обеспечить качество данных во времени: их своевременность и правильность на момент проведения переписи. Данные должны быть своевременными и достоверными, а также

[8] Мерсте, У. (1987). Системное применение: О системном образе мышления при рассмотрении экономических явлений. Таллинн. «Валгус», 264.
[9] Хийе, И. (2014). Содержание IT-систем. Материал лекций. Таллиннский унивреситет.
[10] Alberts, D. S., Garstka, J. J., Hayes, R. E., Siguari, D. A. (2001). Undrestanding Information Age Warfare. CCP, 312.
[11] United Nations Economic Commission for Europe (UNECE) (2006:17-90). CES Recommendations for the 2010 Censuses of Population and Housing. New York and Geneva: United Nations. [www]http://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf
[12] McGinnes,S., Kapros, E. (2015). Conceptual independence. A design principle for the construction of adaptive information systems. Information Systems. Vol. 47, pp. 33-50.
[13] Рава, К. (2000). Предпосылки для успешности проектов инфосистем.. A & A, nr 1.

иметь приемлемую форму. Для этого следует разделить подготовительные работы на две категории: работа с данными и работы с системой[14].

14. Для проведения переписи на базе регистров необходимо сотрудничество с собственниками или держателями регистров, находящимися в зоне ответственности министерств.

- Министерство внутренних дел: RR – Регистр народонаселения; ETR – Регистр видов на жительство и разрешений на работу, KIR – Государственный регистр заключённых и лиц, находящихся под криминальным надзором; KMAIS.
- Министерство образования и науки: EHIS – Инфосистема образования Эстонии.
- Министерство финансов: EMTA – Регистр налогообязанных лиц; TÖR – Регистр трудовой деятельности, KPR (KOPIS) – Регистр обязательной накопительной пенсии; SAP);
- Министерство экономики и коммуникаций: EHR – Государственный регистр строений;
- Министерство юстиции: KR – Крепостная книга; ARIREG – Коммерческий регистр; KIR – Государственный регистр заключённых и лиц, находящихся под криминальным надзором; (в будущем: RKOARR – Государственный регистр государственных учреждений и органов местного самоуправления; э-досье);
- Министерство социальных дел: STAR – Регистр данных социальных услуг и пособий; PKR – Регистр пенсионного страхования;
- Больничная касса Эстонии: KIRST – Регистр медицинского страхования;
- Министерство обороны: KVKR – Регистр военнообязанных;
- Касса страхования по безработице: EMPIS – Регистр безработных лиц, соискателей работы и оказания услуг в области рынка труда;
- Министерство окружающей среды (ADS - система адресных данных Земельного департамента);
- Департамент шоссейных дорог (Регистр дорожного движения).

15. Факторы, препятствующие стабильной передаче данных, в 2016 году были следующими:

(a) документация по подсоединению инфосистем была недостаточно детальной (вопреки мнению статистических деятелей);

(b) описания состава данных слишком объемно, в том числе много времени требует их согласование;

(c) вместо данных передаются выдержки базы данных регистров (например, Регистр строений), что требует позднее селектирования информации в соответствии с необходимостью;

(d) изменения (обновления) требуют постоянного мониторинга и мануального вмешательства;

[14] (Ibid).

(e) выявлены технические ограничения объема (Инфосистема образования Эстонии) при передаче данных посредством x-tee.

16. Таким образом, для улучшения ситуации следует делать больший упор на работу с информацией на организационном уровне. Те, кто передает данные, должны получить ясное представление о том, в какой форме данные нужно представить получателю. Процесс передачи данных в целях интеграции следует выделить как самостоятельный компонент[15], а затем идентифицировать самостоятельные компоненты и проанализировать их в аспекте производственной способности (например, методика LEAN.)

17. Следует организовать сотрудничество между несколькими учреждениями, и критическое значение имеет здесь движение информации и коммуникативная интеграция на социальном уровне. Проблемы интеграции инфосистем вытекают из истории создания регистров. Регистры Эстонии существуют относительно недолгий срок и создавались они не с целью проведения статистической переписи.

18. Необходимая для проведения переписи интеграция заключается в способности систем функционировать совмещено, передавая для этого как информацию и содержание, так и формат. Подобные возможности систем предполагают от держателей регистров совместимости в семантической зоне[16]. Если будет достигнут автоматический охват данных, будет также создана способность к совместной функциональной работе систем.

19. Упорядочивание баз данных и данных регистров позволяет экономить в коммерческом процессе, не выполняется вручную в виде файла, если в отношении каждой базы данных данный процесс больше не является особенным. В статистике направленность на перепись, опирающуюся на данные регистров, предполагает создание на уровне организации такой среды, которая позволила бы добиться большей производительности, использования многообразных динамичных источников информации, доступа к точной и релевантной информации.

20. Таким образом, цель интеграции – обеспечение потока данных и доступ к структурированной информации – сложная, поскольку производственная среда становится более комплексной, гетерогенной, охватывает имеющиеся системы, семантику, технологии и т.д. Ключевым вопросом в новой ситуации становится управление гетерогенностью, соответствующей совместимостью, обеспечивая это на семантическом, синтаксическом, структурном уровне системы.

21. Возможно ли формализовать интеграционный процесс? Вызовом служит учет влияния, исходящего из причины, по которой регистры создавались. Следует учитывать, что в отношении переписи речь идет о многочисленных источниках информации, которые находятся в постоянном развитии в связи с обновлением технического оборудования и платформ программного обеспечения. При обеспечении совместимости систем для переписи важен социально-культурный аспект процессов, способность к гибкости и понимание необходимости в развитии партнеров по сотрудничеству.

[15] Словарь по инфотехнологии, Eesti Standard EVS-ISO/IEC 2382

[16] Ouksel, M., Sheth, A. (1999). Semntic Interoperability in Global Information Systems. Abrief introduction to the research area and the special section. SIGMOD Record, Vol. 28, No1, pp. 5-12.

IV. Интеграция процесса

22. Центральным пунктом процесса интеграции стоит его продуктивность и способность создавать новые услуги. Решения должны быть жизнеспособны и поддерживать увеличение объема производства текущей статистики на базе регистров, оптимизировать методы работы и технологии с целью создания данных для переписи в два раза быстрее, чем при проведении предыдущей, в т.ч. собрать и опубликовать все данные в течение одного года.

23. Процесс интеграции предполагает управление услугами. Чтобы уменьшить сложность управления при подсоединении систем и интерфейсов, поставщикам услуг необходимо дополнительно взять на вооружение новые средства, которые помогли бы различать содержательную сложность коммерческого процесса от функциональной. Целесообразно отслеживать возможности трактования услуг и их направленность.

Требования к присоединениям инфосистем могут быть универсальными:

- все присоедиения инфосистем должны быть реализованы на уровне приложений как веб-услуга;
- все обращения к присоединению инфосистем должны быть безопасно аутентифицированы и авторизованы;
- присоединения должны быть задокументированы в соответствии с действующей моделью;
- внешние присоединения следует поставлять вместе с описанием услуг WADL или WSDL.

24. Стратегия процесса интеграции могла бы следовать стратегии информационной интеграции, которая в Эстонии связана с одной из мер стратегического плана учреждения статистики по ускорению публикации статистики и снижению административной нагрузки. На первом этапе, в период 2016-2018 годов, целью является создание интеграционной возможности для проведения переписи.

V. Автоматический охват данных как показатель совместимости государственных инфосистем

25. В процессе сбора данных для проведения переписи на базе регистров обеспечение взаимодействия инфосистем предполагает фиксирование определенных принципов, на основании которых можно получать данные из регистров.

26. Автоматическое получение данных может функционировать на основе двух методов (pull или push), предполагая универсальный формат данных. Однако опыт пробной переписи указывает на то обстоятельство, что, несмотря на тестирование, следует также оставить возможность передачи (получения) данных вручную.

27. На техническом уровне до получения данных следует описать характер данных переписи относительно каждого регистра в базе метаданных и описать показатели регистра, которые покрывают все охватываемые данные. Показатели регистра следует увязать с классификаторами, индексами, словарями. Вторым шагом создается XSD по push-методу, записывается и присваивается уникаль-

ный URI. Метод позволяет автоматически получить обратную связь качеству данных, т.е. об ошибках. Применение метода pull было бы противоположным.

28. Способы отправки данных различны. Департамент статистики исходит из архитектурной модели, состоящей из трех частей, в которой учитывается функция, концепция и форма данных[17].

29. Для держателей регистров разработаны требования по планированию дальнейших действий в четырех категориях:

(a) Требование первичного охвата данных: все данные должны быть представлены в виде метаинформации, в т.ч. с кодами классификаторов.

(b) Требования охвата обновлений

(i) Все предоставляемые данные должны быть с метаинформацией, в т.ч. с кодами классификаторов.

(ii) Согласно соглашению между Департаментом статистики и ответственным или уполномоченным обработчиком данных, Департамент статистики получает обновления посредством созданной ответственным или уполномоченным обработчиком данных X-tee услуги или делает это с использованием услуги X-tee Департамента статистики сам ответственный или уполномоченный обработчик базы данных.

(iii) Данные предоставляются в формате XML и описание данных предоставляет создатель услуги X-tee в виде XSD. При получении данных Департамент статистики использует описание ответственного или уполномоченного обработчика базы данных, предоставленного в формате XSD.

(iv) Обновления должны содержать время предоставления; обновления должны быть способны охватывать время предоставления начиная с последнего.

(c) Требования к идентифицикации объекта.

(i) Все адреса Эстонии, имеющиеся в базе данных, должны исходить из системы адресных данных Земельного департамента или быть связаны с описаниями ADS Земельного департамента (идентификатор адресного объекта ADS_OID и идентификатор адреса ADR_ID).

(ii) Все лица, данные которых имеются в базе данных, должны быть идентифицированы по личному коду (в отношении физических лиц) или коду Коммерческого регистра, недоходных объединений и целевых учреждений (в отношении юридических лиц).

(iii) Требования к классификаторам, имеющимся в базе данных, и таблицам соответствия предусматривают, что они согласованы с Департаментом статистики в соответствии с требованиями системы администрирования государственных инфосистем (RIHA).

(d) Требования по обеспечению качества данных

(i) Базы данных должны учитывать результаты периодического оценивания качества, проводимого со стороны Департамента статистики.

[17] Кютт, А. (2014). Образцы архитектуры по передаче данных. Земельный департамент. (2013). Справочник ADS.

(ii) Данные, получаемые из баз данных, должны соответствовать признакам, согласованным с Департаментом статистики, быть полными и корректными.

(iii) Базы данных должны контролировать на своем уровне внутреннюю слаженность (например, обнаруживать и исправлять двойные описания и ошибки, возникшие под влиянием различных признаков и т.п.).

30. Действия, необходимые для выполнения требований, для всех баз данных следующие: централизованное исправление (улучшение) качества данных и их отслеживание; использование данных ADS-системы вместо адресов из других источников; накопление данных с классификаторами и создание т.н. X-пути или использование услуги X-пути Департамента статистики.

31. Исполнители статистических работ в объеме деятельности по переписи должны обеспечить:

(a) техническую документацию состава базы данных и применение мер информационной безопасности;

(b) заключение договоров по части технических условий между ответственным обработчиком и уполномоченными обработчиками баз данных;

(c) сотрудничество между инфосистемой и базами данных по техническим вопросам с инфосистемами государства среднего уровня и с базами данных.

32. Выполнение требований по охвату данных и претворению в жизнь конкретных действий обеспечит совместимость государственных инфосистем для проведения переписи, основанной на базе регистров.

VI. Итоги

33. Предпосылками для проведения переписи на базе регистров служит наличие данных в регистрах и возможность связать эти данные между собой.

34. В Эстонии сложилась ситуация, когда данные, имеющиеся в регистрах, покрывают почти все обязательные для переписи признаки, за исключением места работы и должности. Имеется сеть инфосистем для проведения переписи лиц и жилых помещений. Функционирует система личных кодов и внедряется система адресных данных, сделаны шаги по использованию защищенного обмена информацией через X-путь.

35. Использование сформировавшихся стандартов в инфраструктуре государственных инфосистем служит предпосылкой для перехода на проведение переписи на базе регистров. Важно обеспечить получение данных из регистров, а также обновление данных. Если сбор данных будет происходить автоматически, можно говорить о совместимости государственных инфосистем. Данные для переписи можно получить из регистров в течение пары дней, останутся только некоторые исключения, которые обусловлены фиксированием определенных событий во времени постфактум, например, рождения и смерти.

36. Предпосылкой для автоматического получения данных является техническая среда: формат данных и их качество. Для того, чтобы иметь возможность работать с данными различных регистров, в дальнейшем следует делать ставку именно на этот аспект во имя обеспечения совместимости работы систем.

Охват, базирующийся на связях уникальных данных, снизил бы нагрузку при получении и сохранении данных. Следует разработать критерии качества для оценивания получаемых данных (например, количество технических ошибок, недостающие показатели и т.д.). Систематический контроль качества на этапе сбора данных способствует сокращению числа ошибок в данных регистров. При этом необходимо разработать методику, которая позволит лицам, проводящим статистические работы, внести свой вклад в повышение качества данных в регистрах, а держатель базы данных сможет упорядочивать данные регистра без нарушения производственных действий.

37. В Эстонии необходимо провести подготовку для проведения переписи на базе регистров, в ходе которой данные будут получены из различных баз данных в соответствии с регуляциями защиты данных и статистической безопасности. Разработанные для баз данных требования будут достаточными для обеспечения совместимости государственных инфосистем, если:

- все предоставляемые данные будут аннотированы метаинформацией, в т.ч. кодами классификаторов;
- получение и обновление данных происходят посредством X-пути;
- данные предоставляются в формате XML и описание данных предоставляет создатель услуги X-tee в виде XSD, а обновления содержат время момента предоставления.

38. С учетом стратегических расчетов Департамента статистики относительно методов получения данных, следует реорганизовать сбор данных из регистров – другой возможности для более эффективной организации процесса сбора данных и сокращению расходов нет.