

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

Женева, 15–17 июня 2015 года

Пункт 4 предварительной повестки дня

**Модернизация процесса подготовки статистических данных  
и услуг и управление, ориентированное на эффективность****От подготовки данных, ориентированной на предмет  
обследования, к подготовке данных, ориентированной  
на сам процесс****Записка Статистического управления Республики Словения***Резюме*

Несколько лет назад Статистическое управление Республики Словения начало модернизацию процесса подготовки статистических данных, начав крупный проект развития инфраструктуры. Проект опирался на Типовую модель производства статистической информации и несколько основных концепций, что должно было обеспечить разработку обобщенных, но все же довольно гибких решений, отвечающих требованиям большей части статистических обследований. Вместо создания единого общего решения для всего статистического процесса Управление решило избрать пошаговый подход. Постепенно будут разработаны и включены в процесс подготовки статистических данных обобщенные решения для конкретных стадий процесса, например редактирование данных, агрегирование, стандартные методы оценки. В настоящем документе описаны основные концепции проекта развития инфраструктуры, представлены достигнутые на настоящий момент результаты и излагаются дальнейшие планы развития.

Настоящий документ представлен для обсуждения на первом заседании семинара: "Модернизация процесса подготовки статистических данных и услуг и управление, ориентированное на эффективность" в ходе Конференции европейских статистиков.

GE.15-06708 (R) 070515 070515



\* 1 5 0 6 7 0 8 \*

Просьба отправить на вторичную переработку 

## I. Введение

1. Проведение статистических обследований – трудоемкий, затратный по времени и, следовательно, весьма дорогостоящий процесс. Это в первую очередь касается обследований в области официальной статистики, где надежность статистических материалов имеет решающее значение. Несомненно, объектами большинства официальных статистических обследований являются очень сложные, трудно поддающиеся изучению явления, что еще больше затрудняет работу официальных статистиков. С другой стороны, постоянная угроза сокращения финансирования приводит к тому, что официальной статистике все чаще приходится решать задачу подготовки статистических данных высокого качества при значительно меньших ресурсах.

2. Наблюдаемое в последние годы быстрое развитие информационных технологий (ИТ) открывает возможности абсолютно нового и более эффективного преобразования процесса проведения обследований. Вполне очевидно, что в последнее время большинство национальных статистических управлений прилагают на институциональном уровне значительные усилия для такой модернизации. В стратегических документах национальных статистических бюро весьма часто употребляются такие понятия, как реорганизация, модернизация, новая архитектура и т.д. Конечной целью, которая обычно указывается в таких документах, является использование широкого спектра новейших средств ИТ и прикладных программ в целях упрощения и снижения стоимости всего цикла подготовки данных.

3. Статистическое управление Республики Словения (далее – "СУРС") в этом отношении не является исключением. Работа по созданию новой модернизированной системы обработки данных ведется не первый год. Примерно десять лет назад СУРС начало осуществление плана модернизации своих процессов в целях разработки полностью интегрированных систем подготовки статистических данных. Внедрение подобной системы пошло не по плану, поэтому подход постепенно изменился. СУРС переключилось на разработку нескольких независимых типовых ИТ-решений, которые для каждого конкретного обследования будут объединяться в соответствующий статистический процесс.

4. В настоящее время СУРС осуществляет крупный проект развития инфраструктуры, направленный на воплощение идей модернизации на практике в рамках стандартной процедуры подготовки статистических данных. Постепенно будут разработаны и включены в процесс проведения статистических обследований обобщенные решения для отдельных этапов процесса, например редактирования данных, агрегирования и применения стандартных методов оценки. В настоящем документе описаны основные концепции этого проекта развития, представлены полученные на сегодняшний день результаты и кратко излагаются планы дальнейших нововведений.

## II. От полностью интегрированных решений к модульным

5. В 2007 году СУРС начало осуществление крупного проекта по созданию комплексной, полностью интегрированной системы подготовки статистических данных, которая охватывала бы весь цикл процесса подготовки данных для большей части статистических обследований. Она должна включать, среди прочего, типовую, построенную на метаданных систему для всего цикла обра-

ботки статистических данных. В отношении такой типовой системы по завершении проекта СУРС сделало вывод о том, что полученные результаты лишь частично оправдывали ожидания.

6. В итоге цель создать полностью интегрированную систему, которая объединила бы все отдельные инструменты, разработанные в ходе проекта, достигнута не была. После изучения результатов проекта возник следующий вопрос: действительно ли полностью интегрированные системы являются лучшим направлением дальнейших разработок? Другими словами, ведут ли такие интегрированные системы к модернизации системы подготовки статистических данных, обеспечивая ее гибкость и максимальную эффективность затрат. Все обсуждения и оценки в итоге привели к решению отказаться от идеи создания полностью интегрированной системы и вместо этого заняться разработкой несколько иных решений, сохраняющих при этом некоторые важные черты "прежней системы" (или решений, основанных на использовании метаданных). Было предложено предусмотреть некоторое дробление системы, что повысило бы ее гибкость при сохранении ее универсального характера.

7. Основным отличием этого подхода является то, что СУРС решило разбить статистический процесс на ряд более мелких вспомогательных процессов и начало разработку модульных решений для каждого из этих вспомогательных процессов. Эти модульные решения должны быть разработаны таким образом, чтобы обеспечить легкую и гибкую связь между вводимыми и выходящими данными отдельных компонентов и статистическим процессом в целом. Эти компоненты (также именуемые "структурными элементами") должны представлять собой типовые программы для определенных стадий процесса подготовки статистических данных, и при их разработке должна быть предусмотрена возможность их автономного функционирования. Основные характеристики этих структурных элементов можно кратко описать следующим образом:

a) такие элементы должны разрабатываться на основе согласованных, прозрачных и общепринятых методологических принципов, установленных до непосредственного создания конкретного структурного элемента;

b) такие элементы должны быть настолько открытыми, чтобы эти структурные элементы могли использоваться в различных базах данных, функционирующих на различных платформах (например, на платформе "Оракл", SAS) при условии, что базы данных используют основополагающие правила организации данных;

c) такие элементы разрабатываются как системы, полностью построенные на метаданных, т.е. информация, определяющая параметры выполнения операций по обработке данных для конкретного обследования и в конкретный учетный период, должна предоставляться вне основного компьютерного кода. Любая информация, касающаяся проведения конкретного обследования, должна быть исключена из общего программного кода и должна предоставляться профильным специалистом в специальных таблицах метаданных;

d) метаданные о процессах могут также предоставляться в других базах данных для каждого обследования в других средах, но каждая из этих баз (мета)данных должна быть построена с учетом строгих правил в отношении структуры баз данных (таблицы и переменные).

8. Типовые модульные решения разрабатывались в два этапа. В 2007–2010 годах была создана первая версия новой обобщенной системы. Цель этого первого этапа разработки заключалась в создании программ, основанных на платформе SAS, которые бы могли осуществлять конкретные этапы

процесса для различных обследований, при этом не меняя саму программу SAS, а лишь адаптируя метаданные о процессе (правила обработки), которые являются вводимым элементом общей программы. Наиболее важным нововведением стало значительное изменение методов осуществления каждого конкретного этапа статистического процесса.

9. До внедрения новой системы СУРС применяло "классическую вертикальную схему подготовки данных", в рамках которой средства и решения по подготовке данных "зависели от обследования". Организация статистических процессов (например, редактирование и условный расчет данных, оценка ошибок выборки, контроль над раскрытием статистической информации и составление таблиц) во многом зависела от групп специалистов, проводящих конкретное обследование. От специфики обследований зависели разрабатывавшиеся в первую очередь решения в области программного обеспечения. Такие прикладные программы в основном разрабатывались специально под требования конкретного обследования. Специалисты по информационным технологиям использовали "открытые инструкции" статистиков для разработки программного обеспечения и затем (при необходимости) адаптировали его для каждого конкретного элемента процесса осуществления в ходе обследования. При такой системе на этапе разработки требовалось проведение большого объема работ в области ИТ, из-за чего сопровождение программ было крайне трудоемким.

10. Основным достижением новых решений стало то, что теперь программы для отдельных этапов процесса стали разрабатываться один раз и затем адаптироваться для конкретных условий использования только с помощью метаданных о процессе. Эти метаданные должны вводиться в систему статистиками, занимающимися проведением обследований, которые теперь управляют всей системой. Специалисты в области ИТ и общей методологии лишь предоставляют поддержку при возникновении проблем.

11. Первое общее решение было весьма гибким и открытым. Общие программы могли подключаться к различным средам баз данных для получения микроданных и метаданных о процессах, если эти данные и метаданные были организованы согласно заранее установленным правилам. Такая крайне обобщенная и открытая система действительно очень гибка и является подходящим средством построения статистического процесса. Тем не менее такая открытая система обладает некоторыми очевидными недостатками. Эти недостатки в основном касаются процедур управления метаданными о процессах. Как было указано выше, база данных для метаданных о процессах имеет четко определенную структуру, но для каждого отдельного обследования она может размещаться в различных базах данных и даже в различных средах (например, "Оракл", MS Access, SAS). В действительности в случае большинства обследований метаданные о процессах хранились в базах данных MS Access. Это связано главным образом с тем, что профильные специалисты, которые в основном занимаются управлением такими метаданными, предпочитают эту среду ввиду ее простоты и удобства использования.

12. Проблема подобной разобщенной системы метаданных о процессах заключается в невозможности создания общей действенной программы для управления и контроля над введенными метаданными. Как было отмечено в аналитических отчетах по итогам первого этапа использования дезинтегрированной системы, наибольшие проблемы были вызваны значительным количеством ошибок в метаданных о процессах. Поскольку сегменты для введения правил в настоящий момент являются полностью открытыми, большая доля

этих ошибок касалась синтаксиса правил (например, ошибки в расстановке скобок) или согласования правил и переменных.

13. В целях модернизации системы управления метаданными о процессах и их перемещения СУРС решило провести некоторую реорганизацию всей системы. Целью такой реорганизации является, конечно, не повторное создание полностью интегрированной системы, как было первоначально задумано, а осуществление только таких мер, которые, с одной стороны, позволили бы создать общий инструмент управления, а, с другой стороны, обеспечивали бы высокую гибкость системы. В этих целях было решено предпринять следующие шаги:

а) создать одну единую базу метаданных о процессе. Эта база данных будет создана на платформе "Оракл" и будет управляться приложением .NET, что обеспечит удобство в управлении метаданными о процессах;

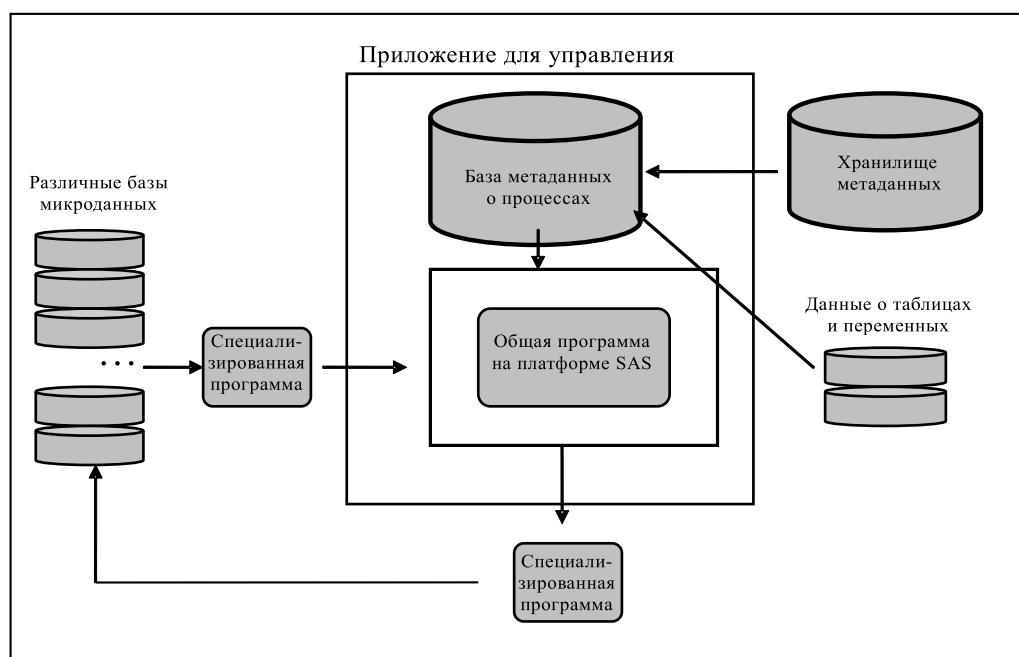
б) соединить систему с хранилищем метаданных, в котором хранятся данные об обследованиях и отдельных элементах обследований;

в) предусмотреть в программе для управления доступ к данным о переменных во вводимых таблицах. Эти данные могут вводиться в программу по двум различным каналам: из реестра переменных, если в ходе обследования ее переменные были зарегистрированы именно там, или из какого-либо другого места, где хранятся структуры таблиц данных.

14. В следующей таблице представлена упрощенная схема планируемой реорганизованной системы:

Таблица 1

**Схема реорганизованной системы**



15. Разработка этой новой системы является основной целью крупного проекта развития инфраструктуры, о котором упоминалось в начале документа. СУРС уже успешно завершило этап, касающийся проверки достоверности данных, автоматического исправления данных и условного расчета данных, и мы в настоящее время постепенно внедряем этот модуль программы в стандартный процесс подготовки данных. Также весьма далеко продвинулась работа над разработкой структурных элементов, связанных с агрегированием данных, контролем над раскрытием данных, вычислением среднеквадратической погрешности и составлением таблиц. СУРС планирует начать внедрение этих модулей в стандартный процесс подготовки данных в конце этого года.

### **III. Трудности, возникающие при внедрении абсолютно новой системы**

16. Когда началось внедрение нового программного обеспечения (разработанного на основе абсолютно нового подхода) в стандартный процесс подготовки данных, СУРС регулярно получало комментарии от статистиков, которые начали использовать новый способ обработки данных. Основные, по их мнению, преимущества и недостатки нового подхода кратко представлены ниже.

Основные преимущества:

- Профильные специалисты получили большую независимость от отдела ИТ, который до этого обеспечивал техническое выполнение процессов.
- Правила обработки данных могут быть очень быстро изменены с помощью централизованной системы метаданных о процессах. Это придает всему циклу обработки данных дополнительную гибкость.
- Поскольку пользователь может выполнять процедуры несколько раз за короткий промежуток времени, теперь намного легче осуществлять проверку эффективности различных методов обработки данных.

Основные недостатки:

- В процессе введения метаданных высока вероятность ошибок синтаксической структуры. В результате программа не может использоваться вообще либо при ее использовании неверно задаются параметры.
- Профильным специалистам необходимо обучаться новым навыкам, что на практике иногда вызывает трудности в связи с крайней трудоемкостью процесса подготовки данных.
- В случае возникновения ошибок в ходе выполнения процедуры необходимо связываться с техническими специалистами, и, если они недоступны, процесс выполнения может на некоторое время остановиться.

17. Разработка абсолютно новой системы подготовки статистических данных действительно является большим достижением для СУРС. Мы твердо уверены в том, что результаты проекта помогут нам разработать новую, модернизированную систему управления процессами обработки статистических данных в различных статистических системах. Конечной целью является переход от вертикальных к централизованным методологическим и ИТ-решениям. Центральным элементом обновленной системы является программа использования метаданных, которая, с одной стороны, достаточно гибка для работы в различных средах микроданных и, с другой стороны, предполагает централизованное управление метаданными о процессах.

18. Внедрение такого общего, основанного на метаданных программного обеспечения для обработки данных неизбежно влечет за собой изменения на общем институциональном уровне, когда речь идет о разработке и проведении статистических обследований. На основе полученного опыта ключевые изменения можно кратко описать следующим образом:

а) работа между профильными специалистами, специалистами по общей методологии и экспертами в области ИТ распределена в целом по-другому. При старой системе за каждым профильным статистиком был закреплен "свой программист" и "свой специалист по общей методологии", которые использовали конкретные инструкции профильного специалиста для разработки и применения специализированных процедур для каждого обследования. Теперь же специалисты по общей методологии и эксперты по ИТ могут действовать только как "группа поддержки" при возникновении ошибок программного обеспечения или в случае, если процесс не дает ожидаемых результатов. Это означает, что профильные специалисты становятся более независимыми от отдела информационных технологий и методологического отдела;

б) изменение роли профильных статистиков в статистическом процессе также привело к смене ожиданий относительно их квалификации и возможностей. Раньше от них ожидалось глубокое знание профильного предмета и способность составлять письменные инструкции (открытого формата) для выполнения конкретных этапов процесса (например, условный расчет, агрегация). Теперь они должны проходить специальную подготовку и уметь составлять эти правила на математическом языке программирования;

в) вся структура работы отдела ИТ и отдела общей методологии должна быть переориентирована с предмета обследования на сам процесс. Такая реорганизация означает изменение общего представления о структуре учреждения и распределении работы и в связи с этим является весьма сложной задачей для статистической организации. СУРС в настоящее время находится на начальном этапе решения этой задачи;

г) указанный выше переход от концепции подготовки данных, ориентированной на (конкретную) область, к концепции подготовки данных, ориентированной на сам (общий) процесс, также должен быть осуществлен на уровне экспертов в области ИТ и методологии. Разработка и поддержка таких типовых программ требует наличия экспертов, способных работать на более общем уровне, воспринимая отдельное обследование как одну из составляющих единого статистического процесса.

## Литература

Dolenc, D., Krek, M., Seljak, R. (2011), "Editing Process in the Case of Slovenian Register-based Census", доклад, представленный на рабочем заседании ЕЭК ООН по вопросам редактирования статистических данных, Любляна, Словения, 9–11 мая 2011 года;

Seljak, R. (2009), "Integrated statistical systems and their flexibility – How to find the balance?", представлен на конференции "Новые техники и технологии в статистике", Брюссель, Бельгия, 5–7 марта 2013 года;

Seljak, R., Blazic, P. (2011), "Sampling error estimation – SORS practice", представлен на втором Европейском семинаре корпоративной статистики, Невшатель, Швейцария, 12–14 сентября 2011 года;

Seljak, R. (2014), “Metadata driven application for data processing – from local toward global solution”, доклад, представленный на рабочем заседании ЕЭК ООН по вопросам редактирования статистических данных, Париж, Франция, 28–30 апреля 2014 года.

---