

Семинар по Модернизации Официальной Статистики

Нижний Новгород, Российская Федерация, 10-12 Июня 2014

Типовая модель статистической информации (ТМСИ)

Сессия 1: Стандарты и инструменты для модернизации статистической продукции и услуг

Подготовлен Группой высокого уровня по модернизации статистического производства и услуг

Резюме

В настоящем документе вниманию сотрудников статистических организаций предлагается Типовая модель статистической информации. В нем рассказывается о достоинствах этой модели, а также о том, как ее внедрение может отразиться на сотрудниках статистических организаций. В документе рассматривается также взаимодействие Типовой модели статистической информации и других рамочных инициатив и стандартов, таких как Типовая модель производства статистической информации, Инициатива в области документирования данных и Стандарт обмена статистическими данными и метаданными. Более подробные сведения об этой информационной модели можно найти в документе, содержащем спецификацию Типовой модели статистической информации, выдержках, приводимых в приложении, и на вики-странице¹, которая может оказаться полезной для организаций, внедряющих эту информационную модель в качестве корпоративного стандарта.

Настоящий документ предлагается Конференции европейских статистиков для информации.

¹ <http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model>.

Содержание

	<i>Стр.</i>
I. Введение	3
II. Охват.....	4
III. Что представляет собой Типовая модель статистической информации?.....	4
IV. Выгоды Типовой модели статистической информации для организации в целом.....	8
A. Долгосрочные выгоды.....	8
B. Непосредственные выгоды	8
V. Типовая модель статистической информации и Типовая модель производства статистической информации	9
VI. Какое значение это имеет?	10
A. С точки зрения производственного процесса.....	10
B. С точки зрения информационных технологий	12
VII. Стандарт обмена статистическими данными и метаданными, Инициатива в области документирования данных и другие стандарты	13
Приложение – Выдержки из спецификации Типовой модели статистической информации	
A. Информация в статистическом производственном процессе	16
B. основополагающая информация.....	30

I. Введение

1. Статистические организации во всем мире осуществляют схожую деятельность, хотя и используя при этом разные процедуры. В ходе этой деятельности используется и генерируется схожая информация (например, организации используют классификации, создают наборы данных и распространяют информацию). Хотя используемая статистическими организациями информация в основе своей является идентичной, все организации склонны описывать ее несколько по-своему (нередко эта информация описывается по-разному в рамках одной и той же организации). В прошлом единого способа описания используемой информации не существовало. Это осложняло четкую связь как в рамках организаций, так и между ними, а без этой связи невозможно представить себе углубленное сотрудничество, стандартизацию или обмен инструментами и методиками.

2. Типовая модель статистической информации (ТМСИ) является первой одобренной на международном уровне справочной моделью статистической информации. Эта рамочная концептуальная модель будет играть важную роль в обновлении, оптимизации и согласовании стандартов, а также в производстве официальной статистики на национальном и международном уровнях.

3. ТМСИ является справочной моделью информационных объектов, обеспечивающей типовые описания процедур, определения, управления и использования данных и метаданных на протяжении всего процесса статистического производства. Она содержит набор стандартизированных, последовательно определяемых информационных объектов, являющихся как исходным, так и конечным продуктом в процессе разработки и генерирования статистической информации. В качестве справочной модели ТМСИ помогает объяснить значимые связи между субъектами, принимающими участие в статистическом процессе, и может быть использована в качестве руководства при разработке и использовании согласованных имплементационных стандартов и спецификаций.

4. ТМСИ является одним из тех краеугольных камней, которые позволяют модернизировать официальную статистику и отказаться от узко отраслевого принципа разработки данных. Эта модель лежит в основе стратегического подхода, сформулированного Группой высокого уровня по модернизации статистического производства и услуг (ГВУ) и одобренного Конференцией европейских статистиков².

5. Модернизация статистического процесса нужна для того, чтобы деятельность статистических организаций могла оставаться актуальной и гибкой в динамичной и конкурентной информационной среде. Хотелось бы надеяться, что статистические организации будут брать на вооружение и использовать ТМСИ и предлагаемый ею единый язык. В то же время модель сама по себе не может трансформировать организацию или порядок ее работы. ТМСИ призвана способствовать максимально широкому применению инновационных подходов в

² См. www1.unece.org/stat/platform/display/hlgbas.

сфере статистического производства. Она является одним из главных компонентов Единой архитектуры статистического производства³ – коллективной инициативы, направленной на разработку общих и взаимозаменяемых услуг со стандартными интерфейсами в поддержку процесса стандартизации и модернизации. При этом ТМСИ поддерживает и текущие подходы к разработке статистики.

6. Настоящий документ служит введением в ТМСИ для довольно широкой статистической аудитории, обобщая ее главные особенности. Более подробную информацию технического плана можно найти в документе со спецификациями и смежных материалах, размещенных на веб-сайте ЕЭК ООН⁴.

II. Охват

7. ТМСИ предлагает модель информационных объектов для всех статистических процессов подобных тем, которые описываются в Типовой модели производства статистической информации (ТМПСИ)⁵, присваивая информационным объектам согласованные наименования, давая им определение, описывая их важнейшие характеристики и указывая их связь с другими объектами информации. В то же время в модели не высказывается никаких предположений о том, на основе каких стандартов или технологий должно обеспечиваться ее применение.

8. ТМСИ не содержит информационных объектов, относящихся к бизнес-функциям организации, таким как кадровые, финансовыми ресурсы или юридические функции, за исключением тех случаев, когда такая информация непосредственно используется в процессе статистического производства.

III. Что представляет собой Типовая модель статистической информации?

9. ТМСИ содержит объекты, которые специфицируют информацию о реальном мире – "информационные объекты", например данные и метаданные (такие как классификации), а также правила и параметры, необходимые для функционирования производственных процессов (например, правила редактирования данных). В ТМСИ определяются порядка 110 информационных объектов, которые сгруппированы в 4 основные группы и более подробно рассматриваются в спецификационной документации.

³ См. <http://www1.unece.org/stat/platform/display/CSPA>.

⁴ См. [http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+\(GSIM\)](http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+(GSIM)).

⁵ См. www.unece.org/stats/gsbpm, document number (ECE/CES/2014/1).

Диаграмма 1

Основные группы информационных объектов ТМСИ



10. Четыре основные группы описываются ниже:

а) *производственная группа* используется для регистрации разработки и планирования статистических программ и процесса их осуществления. Речь идет об определении статистических потребностей, производственных процессах, применительно к которым сформулирована статистическая программа, и их оценке;

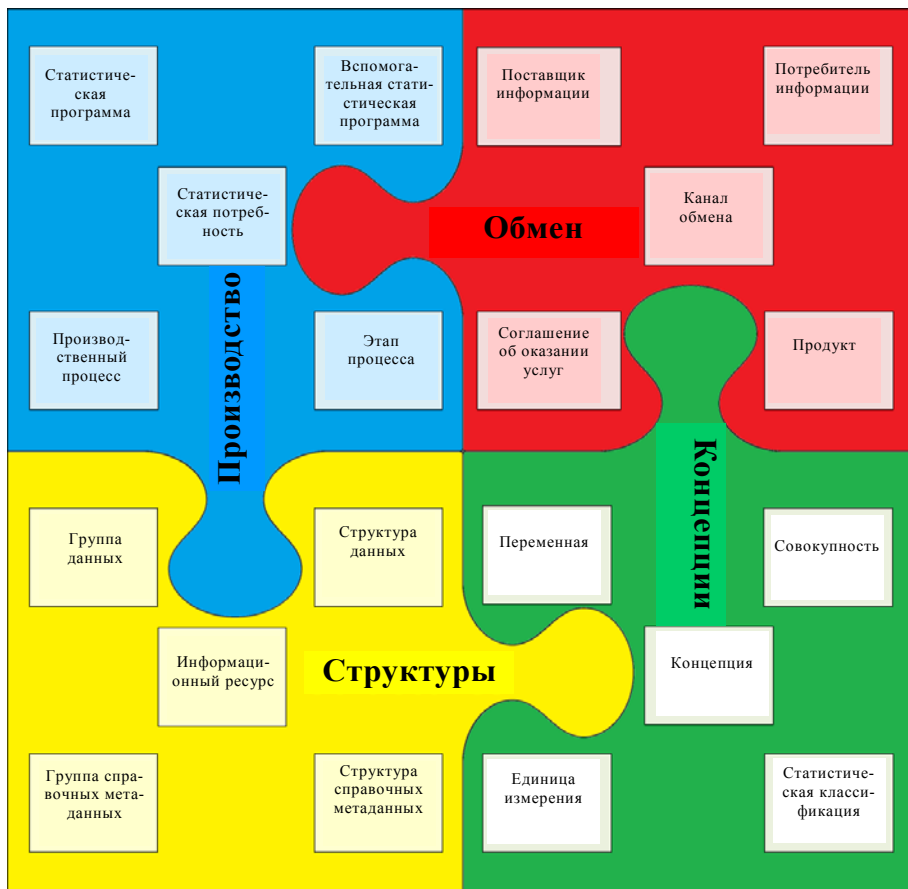
б) *группа обмена* используется для каталогизации информации, поступающей в статистическую организацию и выходящей из нее по каналам обмена. К ней относятся объекты, описывающие сбор и распространение информации;

в) *группа концепций* используется для определения смысла данных, обеспечивая понимание того, что измеряют те или иные данные;

г) *группа структур* служит цели описания и определения терминов, используемых применительно к информации, и ее структуры.

11. Диаграмма 2 дает упрощенное представление об информационных объектах ТМСИ. На ней пользователи могут найти примеры объектов, относящихся к каждой из четырех основных групп.

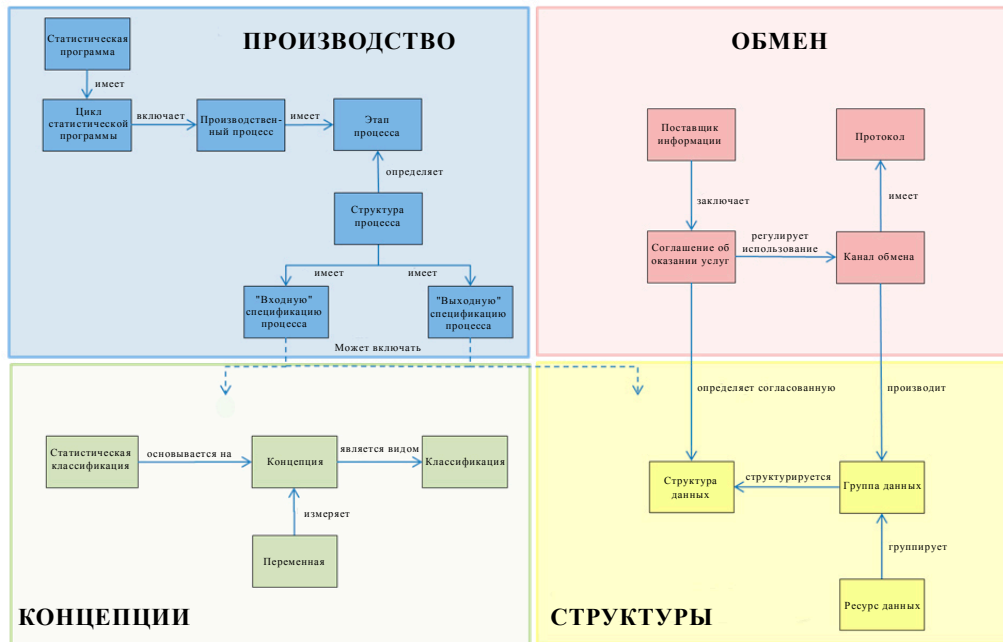
Диаграмма 2
Упрощенный взгляд на информационные объекты ТМСИ



12. На диаграмме 3 ТМСИ изображена в ином ракурсе, несколько более техническом, но, тем не менее, также рассчитанном на сравнительно широкую аудиторию. Диаграммы 2 и 3 могут предлагаться вниманию тех пользователей, которых интересуют примеры объектов и взаимосвязей в ТМСИ.

Диаграмма 3

Альтернативный упрощенный взгляд на информационные объекты ТМСИ



13. Диаграмма 3 является примером информационных объектов ТМСИ, дающих определенное представление о том, какая информация важна для статистической организации. Информационные объекты в ТМСИ выделены курсивом.

"Статистическая организация инициирует *статистическую программу*. *Статистическая программа* представляет собой одну из текущих работ, например обследование или составление рядов данных, и имеет свой *программный цикл* (она может проводиться ежеквартально или ежегодно).

Цикл статистической программы включает в себя набор производственных процессов. Производственный процесс состоит из ряда этапов, которые определяются в схеме процесса. Эта схема содержит спецификации данных "на входе" и "на выходе". Зачастую спецификации представляют собой набор сведений концептуального и структурного характера (статистические классификации, переменные, совокупности, структура данных и наборы данных).

Например, если *производственный* процесс связан со сбором данных, должен существовать *поставщик информации*, который согласен снабжать ею статистическую организацию (в рамках *соглашения об оказании услуг*). В таком *соглашении об оказании услуг* должны определяться *структура данных* и *канал обмена* поступающей информацией. Таким *каналом обмена* может быть *вопросник* или *административный регистр*. Информация должна поступать в него через конкретный механизм (*Протокол*), такой как опрос или обмен файлами данных.

Набор данных, поступивший из *канала обмена* будет храниться в *архиве информационных ресурсов* в структурированном виде в соответствии с требованиями к *структуре данных*".

IV. Выгоды Типовой модели статистической информации для организации в целом

14. Предполагается, что организации будут пользоваться ТМСИ в разной степени. В некоторых случаях организации могут ограничиваться ссылками на нее во внутренних процедурах или при взаимодействии с другими организациями в целях уточнения отдельных моментов. В других случаях организация может принять решение о внедрении ТМСИ в качестве информационной модели, определяющей ее операционную среду. Все возможные сценарии использования ТМСИ имеют под собой основание, хотя на максимальную отдачу могут рассчитывать те организации, которые используют модель в максимальной степени.

A. Долгосрочные выгоды

15. ТМСИ содержит набор стандартных информационных объектов, служащих как исходным материалом, так и конечным результатом в процессе планирования и производства статистических работ. Определяя объекты, которые являются общими для процесса производства статистики вне зависимости от ее тематической направленности, ТМСИ позволяет статистическим ведомствам переосмыслить организацию своей деятельности, с тем чтобы сделать ее более эффективной.

16. ТМСИ может быть использована для ориентирования будущих инвестиций на те сферы статистической деятельности, где в них ощущается наибольшая общая потребность. Модель может также стимулировать определенную специализацию в международном статистическом сообществе. Например, одни организации могли бы специализироваться на сезонных факторах, анализе временных рядов или валидации данных, а другие – пользоваться их опытом.

17. Внедрение ТМСИ одновременно с ТМПСИ позволит получить еще большие преимущества. ТМСИ способна:

- создать условия для повторного использования методик, компонентов и процессов и обмена ими;
- обеспечить возможность контролировать процесс, исходя из установленных правил, тем самым ограничивая роль человеческого фактора в производственном процессе;
- содействовать получению экономии масштаба за счет разработки общих инструментов сообществом статистических организаций.

B. Непосредственные выгоды

18. Важным преимуществом использования ТМСИ является то, что она предлагает единый язык для улучшения взаимодействия на разных уровнях:

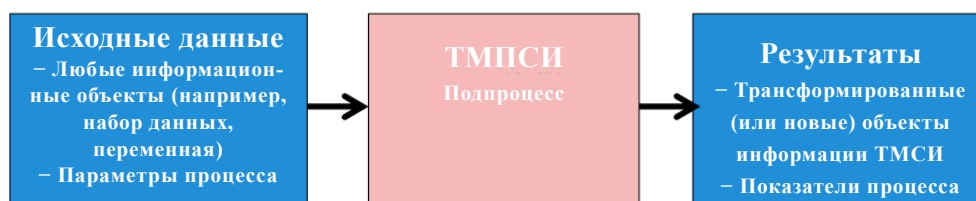
- между участниками статистического процесса, играющими разные роли (эксперты по производственным вопросам и по информационным технологиям);
- между разными тематическими областями статистического процесса;

- между статистическими организациями на национальном и международном уровнях.
19. Улучшение взаимодействия позволит повысить эффективность обмена данными и метаданными как в рамках статистических организаций, так и между ними, а также с внешними пользователями и поставщиками.
20. ТМСИ может уже сегодня использоваться организациями в следующих целях:
- для укрепления кадрового потенциала благодаря использованию ТМСИ в качестве учебного пособия, упрощающего понимание сложной информации и предлагающего четкие определения;
 - для валидации существующих информационных систем и их сопоставления с появляющимися передовыми международными образцами с целью заимствования там, где это возможно, лучшего международного опыта;
 - для придания необходимой направленности процессу разработки или обновления международных или местных стандартов, с тем чтобы они обеспечивали удовлетворение самых широких потребностей международного статистического сообщества.

V. Типовая модель статистической информации и Типовая модель производства статистической информации

21. ТМСИ и ТМПСИ являются взаимодополняющими моделями для производства статистической информации и управления ею. ТМПСИ моделирует статистический производственный процесс, определяя те виды деятельности производителей официальной статистики, которые производят статистические материалы. Эта деятельность разбивается на подпроцессы, такие как "импутация" и "расчет агрегированных показателей". Как видно на диаграмме 6, ТМСИ помогает описать подпроцессы ТМПСИ, определяя объекты информации, которые движутся между ними, создаются в их рамках и используются для производства официальной статистики.

Диаграмма 4
Связь между ТМСИ и ТМПСИ



22. Эффективность использования ТМСИ может быть повышена, если эта модель будет применяться в сочетании с ТМПСИ. Аналогичным образом отдача от ТМПСИ будет выше, если она будет использоваться совместно с ТМСИ. В то же время использование одной модели без другой является возможным, хотя и не идеальным вариантом. Аналогично тому, как в отдельных статистических производственных процессах не используются все подпроцессы, описанные в ТМПСИ, весьма маловероятно, что в любом конкретном статистическом произ-

водственном процессе могут найти применение все информационные объекты ТМСИ.

23. Эффективное управление метаданными имеет важнейшее значение для эффективного функционирования статистического производственного процесса. Созданные с нуля, обновленные или взятые в неизменном виде с предыдущих этапов, метаданные присутствуют на каждом этапе ТМПСИ. В контексте ТМПСИ главное внимание в процессе управления метаданными, которому отводится важнейшая роль, уделяется созданию, обновлению и повторному использованию метаданных. Таким образом, стратегии и системы управления метаданными, использованию которых помогает ТМСИ, играют жизненно важную роль в функционировании ТМПСИ.

24. Одновременное применение ТМСИ и ТМПСИ (или аналогичной модели, используемой конкретной организацией) может:

- способствовать созданию эффективных систем сбора, обработки и распространения информации, построенных вокруг метаданных;
- содействовать гармонизации инфраструктур статистических расчетов.

25. ТМСИ закрепляет последовательный подход к метаданным, помогая метаданным играть свою главную роль, предусмотренную в части А "Статистические метаданные в корпоративном контексте"⁶, Общей системы метаданных, которая заключается в *уникальном и формальном определении содержания и связей между объектами и процессами в статистической информационной системе.*

VI. Какое значение это имеет?

А. С точки зрения производственного процесса

26. ТМСИ поможет улучшить взаимоотношения с коллегами (как на местном, так и на международном уровне).

27. Обмен содержательной информацией по разным темам зачастую налажен плохо, что осложняет обмен концепциями, переменными и структурными компонентами, требуя сложного планирования. ТМСИ может стать тем общим языком, который облегчит взаимодействие между:

- специалистами по конкретным темам, методологиям и информационным технологиям;
- статистиками, работающими в разных подразделениях статистических организаций;
- статистиками из разных организаций.

28. ТМСИ поможет в разработке и понимании процедур (в том числе с точки зрения информации на входе и выходе).

29. В рамках производственного цикла статистик может принять решение о том, какой должна быть информация на входе и на выходе и что должно происходить с ней в промежутке. В случае использования ТМСИ информационные объекты на входе и выходе могут носить структурный и концептуальный харак-

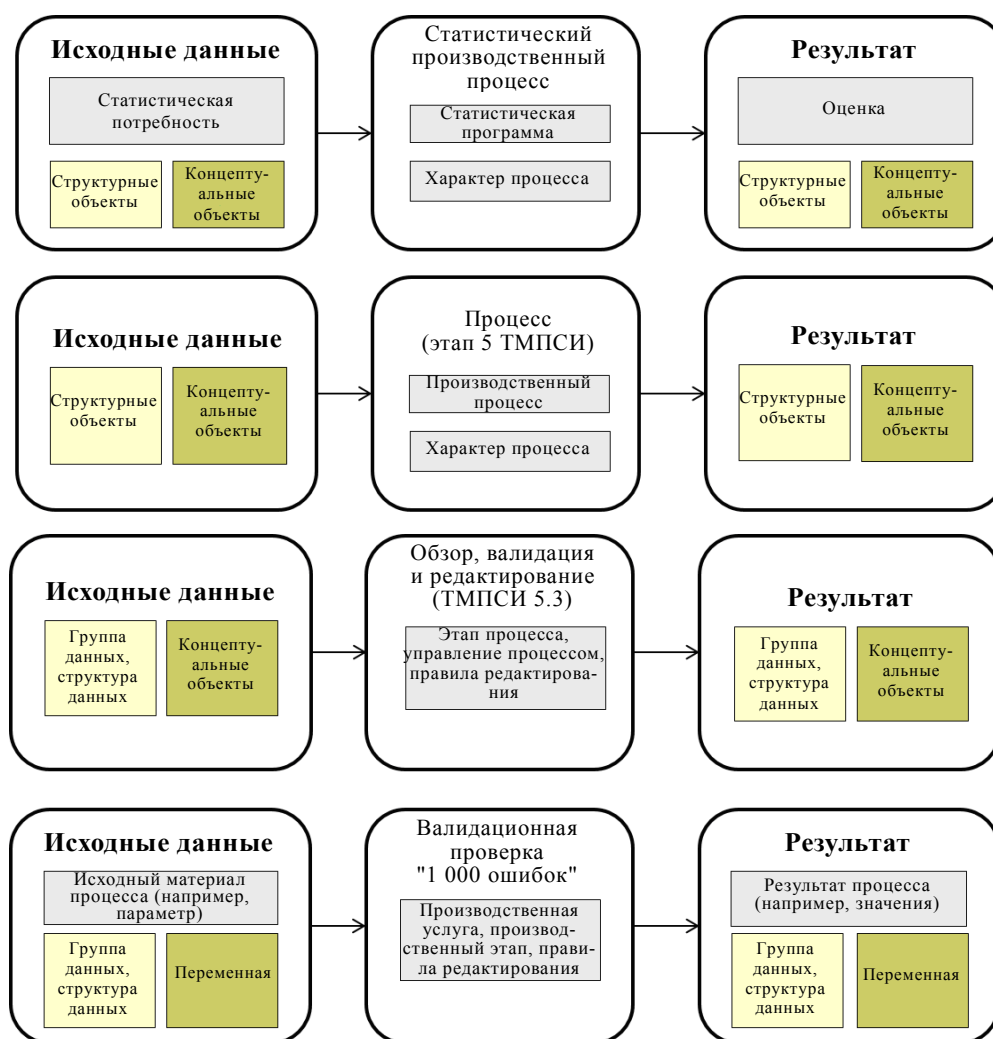
⁶ <http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework>.

тер, а в промежутке – производственный характер. Структурные и концептуальные объекты формулируют специалисты по конкретным темам.

30. Как видно на диаграмме 4, если за модель статистического производственного процесса взять ТМПСИ, ее первый уровень можно считать эквивалентным статистическому производственному процессу в целом. Следующий уровень соответствует одному из этапов статистического производственного процесса (например, "процессуальному" этапу ТМПСИ). Третий уровень характеризует часть того или иного процесса (например, подпроцесс 5.3 ТМПСИ – обзор и валидация). Четвертый уровень состоит из отдельных "кирпичиков", из которых складывается тот или иной подпроцесс (определение финансовых показателей, которые могут выражаться не в единицах, а в тысячах единиц).

Диаграмма 5

Информационные объекты ТМСИ в контексте ТМПСИ



31. Серьезные проблемы перед статистиками встают в связи с использованием разовых компонентов, которые нередко приходится воссоздавать или модифицировать для каждого нового производственного цикла. Облегчая унифицированное описание информационных объектов на входе и выходе, ТМСИ позволяет структурировать разработку статистических процессов. Эта модель способствует разработке, спецификации и внедрению унифицированных методов и стандартных технологий в интересах создания единообразной системы статистического производства.
32. Применение ТМСИ позволит гибко выстраивать статистический процесс из "кирпичиков" многоразового пользования, с тем чтобы получать конечную продукцию разной степени сложности, более широкого ассортимента и полнее удовлетворяющую меняющиеся требования клиентов.
33. Использование ТМСИ сделает работу проще, поскольку многие процессы после необходимой адаптации можно будет использовать повторно. В результате меньше времени будет уходить на однообразную работу и больше – на инновационную.
34. В долгосрочном плане ТМСИ снизит зависимость статистиков от информационных технологий.
35. Сегодня статистиков очень волнуют вопросы применимости, пригодности и стабильности их методик и технических инструментов. При "вертикальном" подходе к статистическому процессу тематическая работа сильно зависит от специалистов по информационным технологиям, которые проектируют, выстраивают и обслуживают статистическую систему.
36. Усиление контроля над разработкой статистических процессов позволит статистикам стать более самостоятельными в планировании и осуществлении статистической деятельности.
37. В основу производственного процесса лягут стандартные процедуры, более устойчивые к изменениям и к кадровым проблемам. Расширение использования стандартных процедур, применимых в разных областях, облегчит статистикам работу сразу по нескольким направлениям.

В. С точки зрения информационных технологий

38. С точки зрения специалистов по информационным технологиям, главная проблема кроется в дублировании усилий по причине "вертикальной" организации статистического процесса. Непостоянные и различные требования, предъявляемые в рамках такого подхода, становятся причиной индивидуализированных одновременных решений, в то время как высокая текучесть специалистов по информационным технологиям ведет к появлению недокументированных нестандартных процедур.
39. Внедрение ТМСИ на национальном и международном уровне уже сегодня может оказать непосредственное позитивное влияние на работу специалистов по информационным технологиям. Эта модель предоставляет в их распоряжение единый язык для общения с клиентами и коллегами как на местном, так и на международном уровне.
40. На национальном уровне статистики станут более самостоятельными в планировании (см. диаграмму 6) и осуществлении своей статистической деятельности. Возможность повторного использования и адаптации унифицированных компонентов ТМСИ для решения новых задач сделает статистические

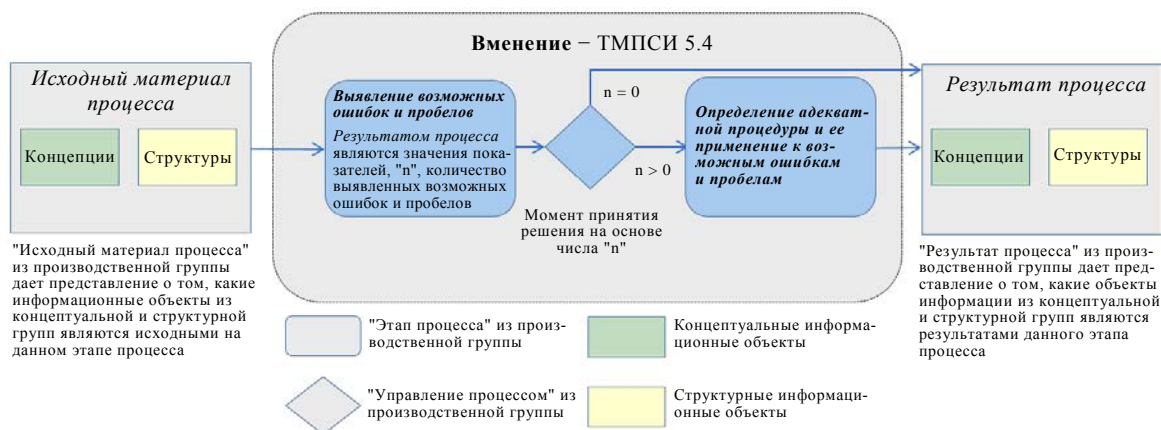
системы более гибкими и модульными. В основу производственного процесса лягут стандартные процедуры, более устойчивые к изменениям и к текучести специалистов по информационным технологиям. Расширение использования стандартных процедур, применимых в разных областях, облегчит статистикам работу сразу по нескольким направлениям.

41. Использование ТМСИ сделает работу проще, поскольку многие процессы после необходимой адаптации можно будет использовать повторно. В результате меньше времени будет уходить на однообразную работу и больше на инновационную.

42. Все это позволит высвободить кадровые ресурсы и направить их на разработку более надежных ИТ-процедур и на поиск новых путей более полного удовлетворения потребностей статистических организаций и клиентов в целом. Больше времени останется на разработку надежных, модульных, унифицированных и хорошо документированных процедур, отвечающих требованиям Единой архитектуры статистического производства.

Диаграмма 6

Разработай свою собственную процедуру выполнения отсутствующих данных



43. На международном уровне появятся новые возможности для совместного планирования и разработки общих компонентов, в основу которых лягут точно сформулированные требования более широкого круга пользователей. Разработчики ИТ-процедур смогут взаимодействовать друг с другом в рамках более широкого профессионального сообщества, используя один и тот же язык для описания своей статистической информации.

VII. Стандарт обмена статистическими данными и метаданными, Инициатива в области документирования данных и другие стандарты

44. В качестве базовой модели информационных объектов ТМСИ имеет также связь с такими стандартами, как стандарты электронного обмена статистическими данными и метаданными (ОСДМ) и Инициатива в области документирования данных (ИДД), которые традиционно используются для представления статистических данных и метаданных и обмена ими.

45. Информационные объекты ТМСИ носят концептуальный характер; они не предусматривают физического представления данных. Можно привести следующий упрощенный пример: название организации можно представить в рамках одной и той же концепции вне зависимости от того, содержится ли информация в базе данных, в крупноформатной электронной таблице, в CSV-файле (значения, разделенные запятыми), в XML-файле (расширяемый язык разметки) или на листке бумаги в рукописном виде.
46. ТМСИ позволяет организациям начать применение единого языка для описания данных и метаданных, используемых на протяжении всего процесса статистического производства. Для этого информационные объекты ТМСИ были соответствующим образом преобразованы в формат ОСДМ и ИДД.
47. Это поможет статистическим организациям описывать статистическую информацию и работать с ней, используя единый язык, и при этом, на системном уровне, представлять информацию и обмениваться ей в соответствующем стандартном техническом формате.
48. Хотя информационные объекты ТМСИ можно отразить в формате ОСДМ и ИДД (использование этих стандартов сулит серьезные выгоды), ТМСИ не требует применения этих стандартов. Некоторые производители и потребители статистики могут в своих конкретных целях воспользоваться иными стандартами. В других случаях производители статистики могли бы использовать ОСДМ и/или ИДД, однако модернизировать их традиционные информационные системы для того, чтобы перейти на эти стандарты, не представляется рентабельным.
49. Описание статистической информации с использованием такой единой базы, как ТМСИ, помогает пользователям устанавливать связь между двумя массивами статистической информации, представленной по-разному с технической точки зрения.
50. Например, статистик может получить часть данных в формате ИДД, а часть в каком-то ином местном формате. Статистик может перевести их в формат ТМСИ, что позволит ему определить, какие различия являются чисто техническими, а какие носят более глубокий концептуальный характер.
51. После того, как характер и степень различий понятны, часто бывает несложно перевести информацию в какой-либо единый технический формат (например, в ОСДМ или ИДД), что позволит объединить ее в единое целое и начать ее анализ. В рамках такого подхода результаты технического преобразования данных в единый формат остаются хорошо понятными и полноценными с концептуальной точки зрения.
52. Использование ТМСИ в качестве единой базы и применение таких стандартов, как ОСДМ и ИДД для отображения информации, сулят синергический эффект, который можно сделать максимальным благодаря правильному планированию работы.
53. Например, при решении вопроса о том, какой набор определений следует использовать применительно к информационным объектам в рамках ТМСИ, главным источником справочной информации служили существующие стандарты и модели. Хотя ни один из существующих источников не имел того же назначения и той же сферы применения, что и ТМСИ – справочной модели информационных объектов, охватывающих весь статистический процесс, – работа с каждым из них предполагала анализ и учет конкретных потребностей и сценариев, имеющих отношение к конкретным видам статистических данных и метаданных.

54. Таким образом, затраты времени на анализ, моделирование, проверку и доработку этих стандартов и моделей с целью доведения их до нынешней степени совершенства пошли на пользу ТМСИ. Кроме того, можно сказать, что с точки зрения используемых терминов и определений ТМСИ не отличается "без веских на то оснований" от существующих стандартов и моделей. Когда такие отличия существуют, они объясняются тем, что существующие стандарты и модели не отличаются внутренней последовательностью, не согласуются друг с другом и/или когда, по мнению статистиков, их профессиональным потребностям в большей степени отвечают иные термины и определения. Прямым следствием этого стал пересмотр Невшательской модели классификаций для приведения ее в соответствие и интеграции с ТМСИ.

Приложение

Выдержки из спецификации Типовой модели статистической информации⁷

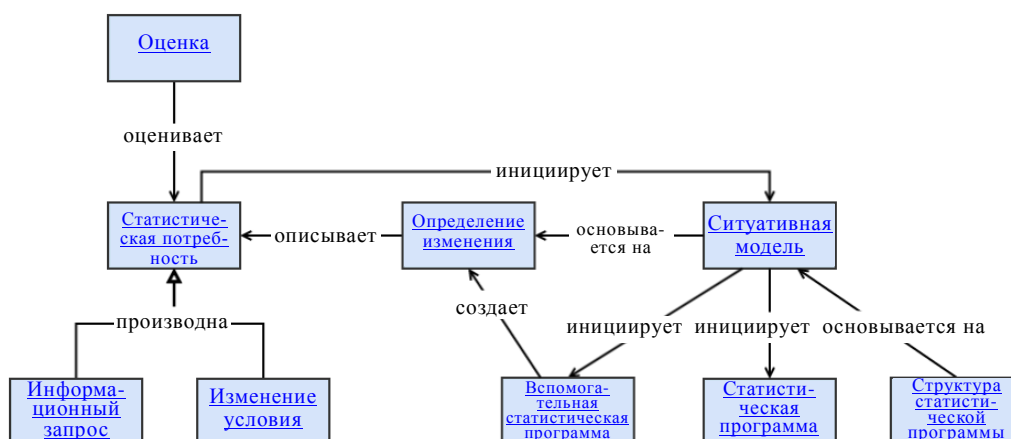
А. Информация в статистическом производственном процессе

1. В настоящем разделе рассматриваются различные способы использования информационных объектов в статистическом производственном процессе. В нем анализируются восемь разных сценариев, определяются используемые информационные объекты и связи между ними.

1. Определение и оценка статистических потребностей

Диаграмма 1

Определение и оценка статистических потребностей



2. Организация реагирует на самые разные потребности и меняется соответствующим образом. Для нее статистические потребности могут возникать в связи с изменением внешних условий или информационным запросом.

3. Изменение внешних условий должно происходить под влиянием извне. Для конкретной организации это может проявиться в сокращении бюджета, в новых требованиях заинтересованных сторон либо в более масштабных изменениях, например в появлении новых методологий или технологий.

⁷ Примечание: в настоящем приложении воспроизводится не весь текст спецификации ТМСИ, а лишь выдержки из этого документа.

4. Когда организация получает информационный запрос, это означает, что среди пользователей тому или иному лицу или организации для той или иной цели требуется та или иная конкретная информация. Круг пользователей может включать в себя пользователей как в самой организации, так и за ее пределами. Например, группе, занимающейся ведением национальных счетов, может потребоваться новая процедура ввода в расчеты новых видов данных. В подобных случаях обычно создается новая тематическая область в зависимости от того, что хочет измерить пользователь. После этого проводится поиск на предмет наличия необходимых данных. Найти такие группы данных может помочь поиск в концепциях и классификациях. Каждая из этих операций представляет собой отдельный этап процесса.

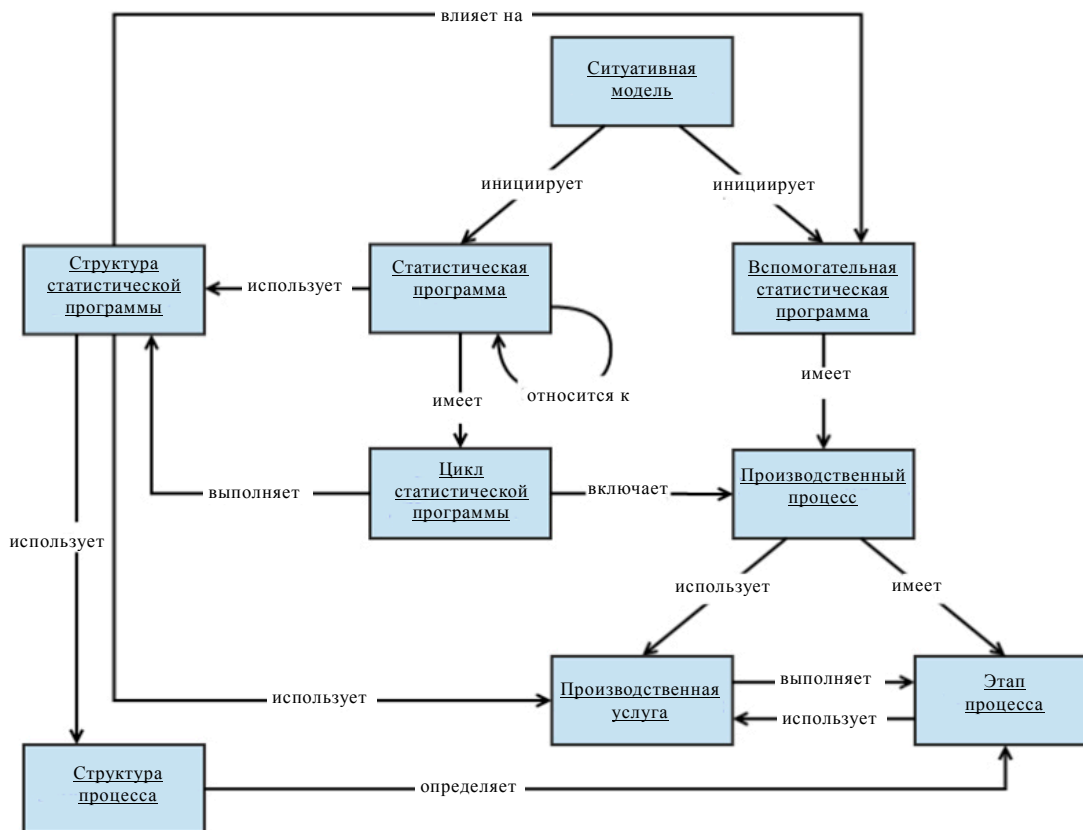
5. Статистическая потребность, возникающая в результате либо информационного запроса, либо изменения внешних условий, формально закрепляется в определении изменения, которое обычно формулируется в рамках вспомогательной статистической программы ("программа статистических изменений"). В определении изменения отражается конкретный характер изменения с точки зрения его последствий для организации, конкретной статистической программы или вспомогательной статистической программы. Такое определение изменения вводится в ситуативную модель. В итоге должна появиться либо новая статистическая программа, либо новая вспомогательная статистическая программа, при помощи которой можно будет разработать новую модель применения уже существующей статистической программы.

6. Статистическая потребность может носить и внутренний характер. На любом этапе статистического процесса организация может оценить полезность или эффективность его самого или его исходных параметров и конечных результатов. Предметом оценки могут стать любые объекты модели – ресурсы, процессы и результаты. В рамках такой оценки должен быть проведен и анализ пробелов в ситуативной модели, а также анализ того, насколько результаты статистического процесса отвечают потребностям, для удовлетворения которых он был разработан.

2. Разработка статистических программ и управление ими

Диаграмма 2

Разработка статистических программ и управление ими



7. Статистическая организация реагирует на возникшую, по ее мнению, статистическую потребность, создавая ситуативную модель. Для этого необходимо сделать одно из трех: разработать новую вспомогательную статистическую программу, создать новую статистическую программу или усовершенствовать структуру существующей статистической программы, которую предстоит использовать.

8. При помощи вспомогательных статистических программ статистические организации осуществляют программы статистических изменений, программы управления данными, программы управления метаданными, программы методологических исследований и т.д. Наглядным примером может служить программа управления классификациями.

9. Статистическими именуется те программы, которые организация осуществляет с целью производства статистической информации (например, обследование розничной торговли). Статистические программы являются циклическими, т.е. они выполняют циклы сбора, производства и распространения статистики. Каждый цикл представлен объектом цикла статистической программы.

Цикл статистической программы представляет собой повторяющийся вид работ по производству статистики в конкретный момент времени (например, обследование розничной торговли за март 2012 года).

10. Для того чтобы статистические программы позволяли решать поставленные задачи, они должны быть должным образом структурированы. Такая структура должна быть свойственна всем запланированным видам деятельности, в первую очередь на уровне производственного процесса. Цикл статистической программы обычно состоит из нескольких производственных процессов. Их можно приравнять к процессам и подпроцессам Типовой модели производства статистической информации (ТМПСИ). Производственный процесс может быть повторен в рамках одного цикла. Каждый такой повтор может состоять из многочисленных видов деятельности, которые могут быть как одноплановыми, так и разноплановыми. Например, в рамках одного цикла Статистической программы можно три раза повторить операцию по сбору и обработке данных, затем проанализировать эти данные и распространить созданные статистические продукты. Каждую такую операцию можно считать самостоятельным производственным процессом.

11. Структура статистической программы определяет порядок осуществления производственных процессов, в том числе использования статистических услуг многоразового пользования (которые статистическая организация может получать извне), или разработки и применения более традиционных процедур. Во втором случае для описания разных этапов процедуры используются структурные объекты. (Статистические услуги многоразового пользования также определяются структурой и разными процессуальными этапами, однако они уже существуют и не требуют новой работы над структурой статистической программы.)

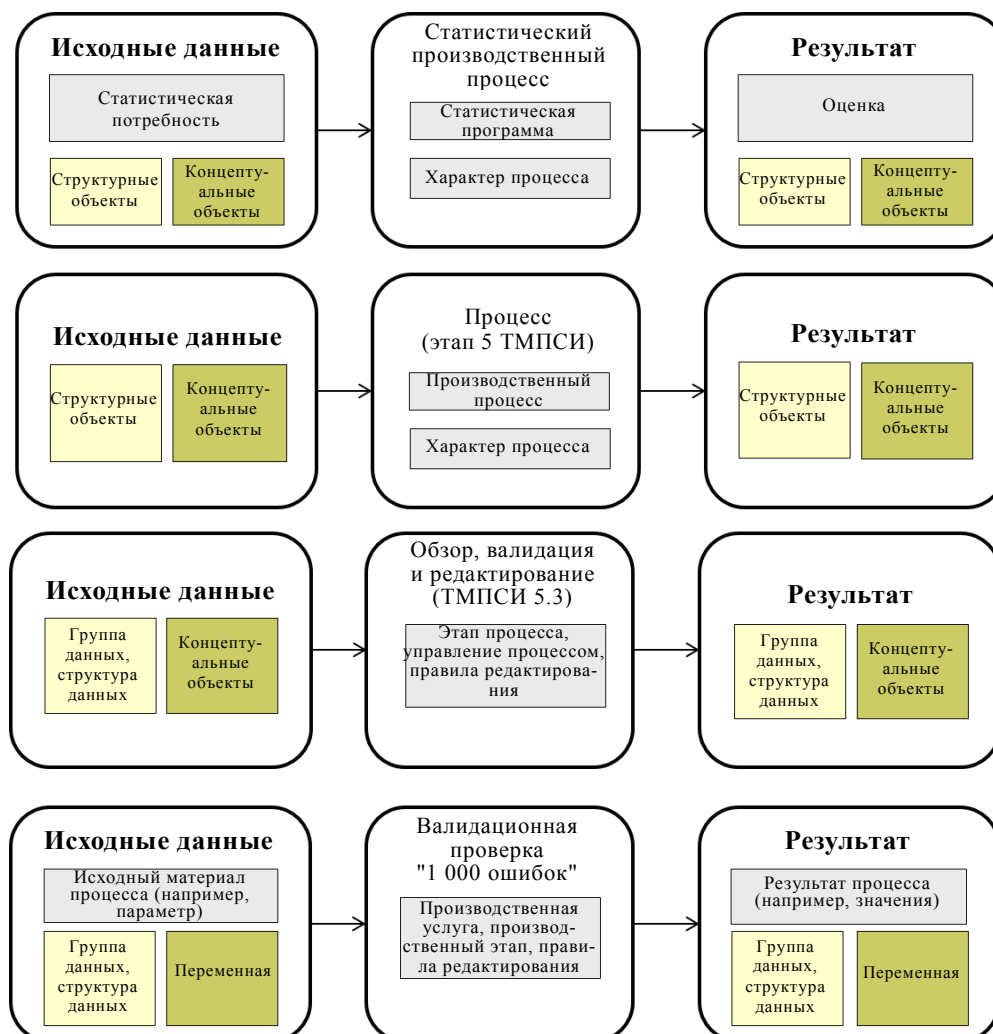
12. Следует отметить, что в структуре статистической программы определяется лишь то, какие этапы производственного процесса потребуются разработать и какие статистические услуги будут использоваться. В то же время эта структура не является тем документом, который подробно описывал бы порядок работы на разных этапах процесса и использования статистических услуг. Эти спецификации можно найти в описании структуры отдельных процедур.

3. Структурирование различных этапов процесса

13. Прежде чем перейти к анализу объектов, используемых в ТМСИ для структурирования отдельных этапов процесса, необходимо рассмотреть характер этих процессов в целом. Разные виды объектов в ТМСИ выполняют свои конкретные функции. В ТМСИ производственные процессы разбиты на этапы процесса. Каждый такой этап может быть настолько "крупным" или "мелким", насколько пожелает разработчик конкретного производственного процесса (см. диаграмму ниже).

Диаграмма 3

Этапы процесса могут быть настолько "крупными" и "мелкими", насколько это необходимо



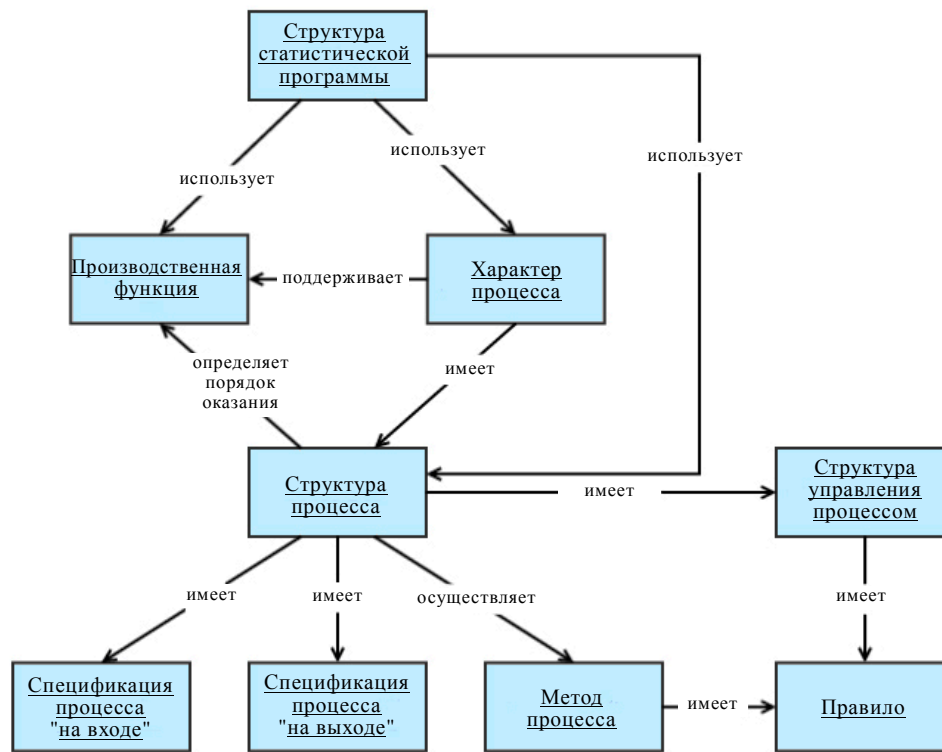
14. Этапы процесса могут состоять из "подэтапов", эти "подэтапы" из еще более мелких "подэтапов" и так далее. Результаты одного этапа процесса обычно становятся исходным материалом для следующего этапа. В зависимости от вводимых параметров или условий, которым удовлетворяют результаты предшествующего этапа, последовательность этапов процесса при определенных условиях можно рассматривать в динамике.

15. Таким образом, можно считать, что в структуре каждого этапа процесса используются элементы других этапов процесса и даже другие статистические производственные услуги, которые уже были разработаны для многократного использования. При более традиционном сценарии этап процесса вначале структурируется, а затем осуществляется. Ожидается, что в будущем статистические услуги, разрабатываемые и предоставляемые другими внешними организациями, будут приобретать все большую популярность. Об этих двух сценариях подробно рассказывается в следующих разделах.

А. Структурирование этапов процесса

Диаграмма 4

Структурирование этапов процесса



16. Структура статистических программ ассоциируется с этапами процесса высшего уровня, структура которых включает все подэтапы и динамические связи, необходимые для запуска статистической программы. Каждый этап процесса статистического производства служит определенной цели: для каждой статистической функции характерен свой этап процесса. Примером статистической функции может служить "восполнение недостающих данных". Для выполнения этой статистической функции требуется процедура импутации, которая в свою очередь должна быть должным образом структурирована.

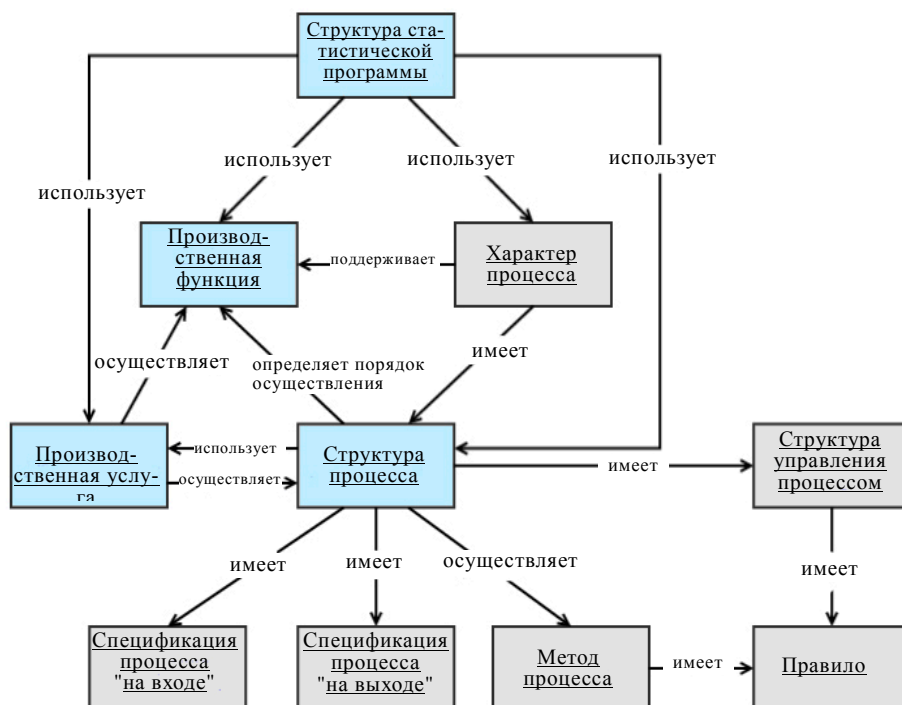
17. В соответствии с характерным для ТМСИ принципом разделения структуры процесса авторы ТМСИ исходят из того, что этапы процесса должны определяться в процессе работы над структурой. После того, как запланированный статистический производственный процесс разделен на отдельные этапы, следующим этапом должно стать структурирование каждого этапа процесса. Целью такого структурирования является описание порядка осуществления каждого этапа. Эта работа может вестись на базе плана процесса, описывающего структуру этапов процесса в динамическом ключе (контрольные процедуры), т.е. с учетом возможности многократного использования.

18. Структура процесса выполняет ряд функций: она определяет различные виды исходных и конечных продуктов, описываемых "входными" и "выходными" спецификациями процесса. Примерами исходных продуктов являются данные, метаданные, например статистические классификации, правила импутации и редактирования, параметры и т.д. К числу конечных продуктов можно отнести различные виды отчетов (результаты обработки данных, отчеты о результатах валидации и качестве данных и т.д.), отредактированные группы данных, новые группы данных, новые или отредактированные метаданные и т.д.
19. Если продолжить этот пример, то разработчик процесса во "входной" спецификации процесса должен указать среди исходных продуктов правила импутации и ту группу данных, к которой их следует применять. "Выходная" спецификация процесса должна включать группу отредактированных данных со вмененными значениями, а также отчет о том, какие данные были исчислены условно.
20. В структуре процесса отражена логическая цепочка управления связями, т.е. последовательность и зависимость отдельных динамических промежуточных процессов (этапов процесса). Такая последовательность динамических связей описывается структурой управления процессом. В ходе структурирования процесса именно структура управления дает информацию о том, что должно произойти на следующем этапе. Иногда за одним этапом процесса вне зависимости от обстоятельств может следовать другой строго оговоренный этап и никакой иной. В таких случаях в структуре управления процессом попросту указывается, какой этап является следующим. Однако иногда между последующими этапами существует выбор. В этом случае в структуре управления процессом указывается набор возможных "следующих этапов" и критерии, позволяющие выбрать один или несколько из них.
21. Структура процесса применительно к тому или иному этапу задает метод осуществления статистической производственной функции, связанной с этим этапом. Например, если статистическая функция заключается в "восполнении недостающих данных путем импутации", тогда метод осуществления этого процесса может быть сформулирован как "подстановка ближайшего значения".
22. Метод процесса представляет собой метод, который должен применяться и увязываться с набором применимых правил. Например, использование метода "подстановка ближайшего значения" регулируется (параметрическим) правилом определения ближайшего значения. В рассматриваемом примере правило должно быть математическим (например, основанным на формуле). Правила могут также быть логическими (например, если условия 1 и 2 не соблюдаются, требуется импутация, во всех остальных случаях импутация не требуется).
23. Полученные таким образом объекты структуры процесса и структуры управления процессом (наряду с соответствующими спецификациями процесса "на входе" и "на выходе") будут использоваться в реализации этапов процесса.

В. Услуги статистического производства многоразового пользования

Диаграмма 5

Использование Услуг статистического производства многоразового пользования



24. В рамках статистической программы далеко не всегда существует необходимость в структурировании этапов процесса с самого начала. Единая архитектура статистического производства (ЕАСП) описывает, каким образом статистические организации могут разрабатывать статистические услуги, которые могут легко заимствоваться другими статистическими организациями. С точки зрения ТМСИ, статистическая услуга является производственной услугой. Производственная услуга представляет собой средство выполнения производственной функции (наличие у организации соответствующего потенциала, обычно выражаемого в общих чертах на достаточно высоком уровне и требующего сочетания организационных, кадровых, производственных и технологических факторов).

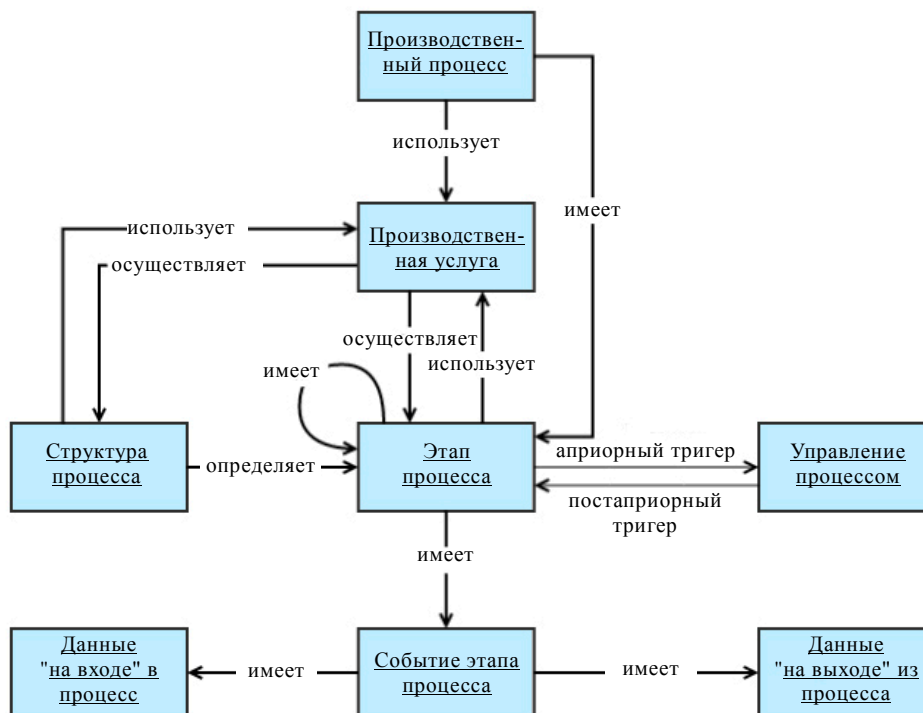
25. Расширение обмена производственными услугами и их многоразового использования позволяет существенно уменьшить потребность в ресурсах для удовлетворения спроса на новую статистическую продукцию и время ее подготовки. Для того чтобы содействовать этому, в ЕАСП вводится концепция каталога статистических услуг, в который различные статистические организации могут внести разработанные ими статистические услуги с целью обмена ими с другими статистическими организациями.

26. Производственные услуги уже разрабатываются с использованием всех обычных исходных и выходных данных, структуры управления процессом и других оговоренных параметров. Таким образом, производственная услуга может выполнять функцию, аналогичную функции этапа процесса, разработанного в рамках организации, не требуя при этом усилий, которых не избежать при традиционном сценарии.

4. Осуществление процесса

Диаграмма 6

Осуществление процесса



27. Для выполнения производственных функций статистическая программа должна осуществлять определенные процессы. Этого можно добиться двумя путями: производственный процесс может либо предусматривать непосредственное осуществление того или иного этапа процесса, либо использование производственной услуги многократного пользования в качестве непосредственного активатора этапа процесса.

28. Для того чтобы понять, как работает эта схема, можно описать характер этапов процесса более подробно. Этапы процесса представляют собой ресурсы, которые предусмотрены структурой процесса и которые могут быть использованы несколько раз. Этапы процесса могут существовать на разных уровнях детализации и могут предусматривать использование других этапов процесса в качестве "подпроцессов". Последовательность осуществления подпроцессов задается в процессе управления процессом в рамках применения структуры управления процессом.

29. Разовое исполнение этапа процесса рассматривается как событие процесса. Именно на этом уровне задействуются конкретные исходные и выходные данные. В структуре процесса определяются используемые виды исходных и

выходных данных (входная и выходная спецификация процесса), а фактические исходные и выходные данные специфичны для события процесса и представлены объектами процесса "на входе" и "на выходе". Исходной может быть любая информация – правила, параметры, группы данных, разные разновидности метаданных и т.д. Совершенно разными могут быть и разновидности выходных данных, которые нередко принимают форму параметров процесса и различных отчетов, а также данных и метаданных.

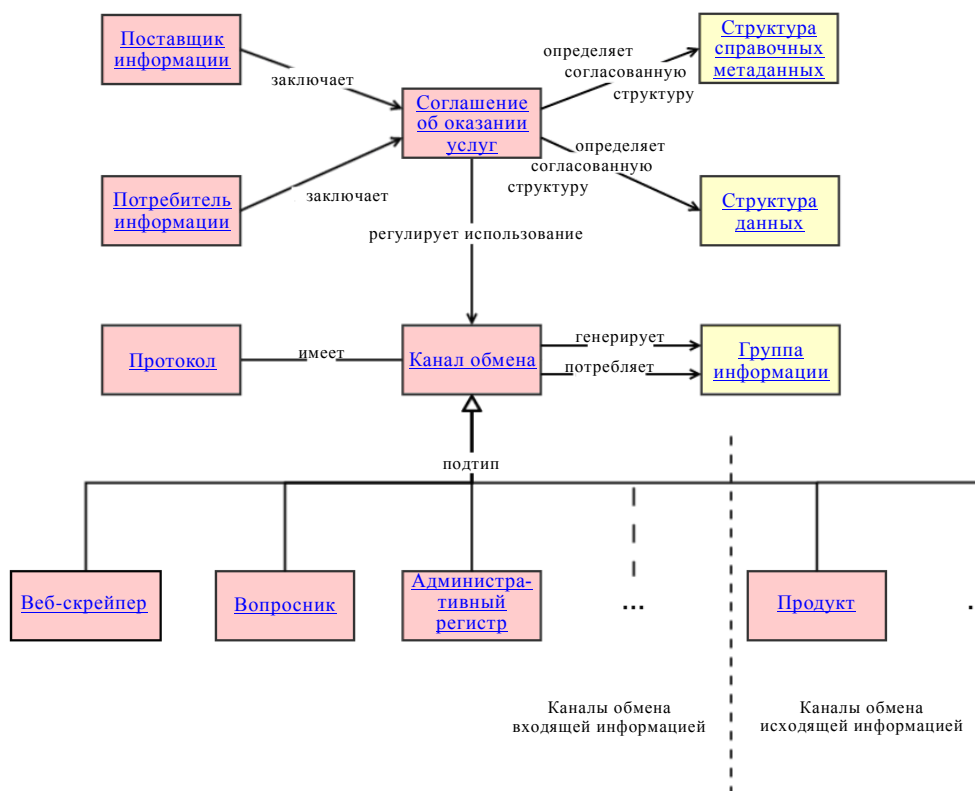
30. В процессе применения структуры процесса должно быть обеспечено применение и выбранных методики и правил. В рамках этой структуры может быть выбрана производственная услуга, которая будет обеспечивать применение метода процесса. Предоставление производственной услуги обеспечивается той или иной компьютерной программой (см. раздел выше) или физическим лицом. Размещение публикации на веб-сайте статистического института или сдача анкет с ответами в общий накопитель для последующей обработки являются примерами таких производственных услуг.

31. Следует отметить, что эта модель допускает использование как автоматизированных, так и ручных процессов, а также процессов, состоящих из "под-процессов" обоих видов.

D. Обмен информацией

Диаграмма 7

Каналы обмена



32. Статистические организации получают данные и справочные метаданные от поставщиков информации, например от респондентов обследования или от органов, ведущих административные регистры, и распространяют информацию среди таких потребителей, как государственные учреждения, коммерческие предприятия и общественность. Каждый обмен данными и справочными метаданными происходит через канал обмена, имеющий свой порядок получения (сбора) или отправки (распространения) данных. Поставщиками и потребителями информации могут быть организации и физические лица как внутри самой статистической организации, так и за ее пределами.

33. Для сбора и распространения информации используются разные каналы. Примерами каналов сбора информации могут служить вопросники, веб-скрейперы и административные регистры. Единственным примером канала для распространения информации, предусмотренным на данный момент в ТМСИ, является статистический продукт. В зависимости от своих потребностей организации могут добавить и другие каналы обмена информацией.

34. Порядок использования канала обмена информацией регулируется соглашением об оказании услуг между статистической организацией и поставщиком (сбор) или потребителем (распространение) информации. Соглашение об оказании услуг, которое может быть явно выраженным или подразумеваемым, служит правовой или иной основой для обмена данными между двумя сторонами. Стороны используют соглашения об оказании услуг и для согласования структуры данных и справочных метаданных, которые подлежат обмену.

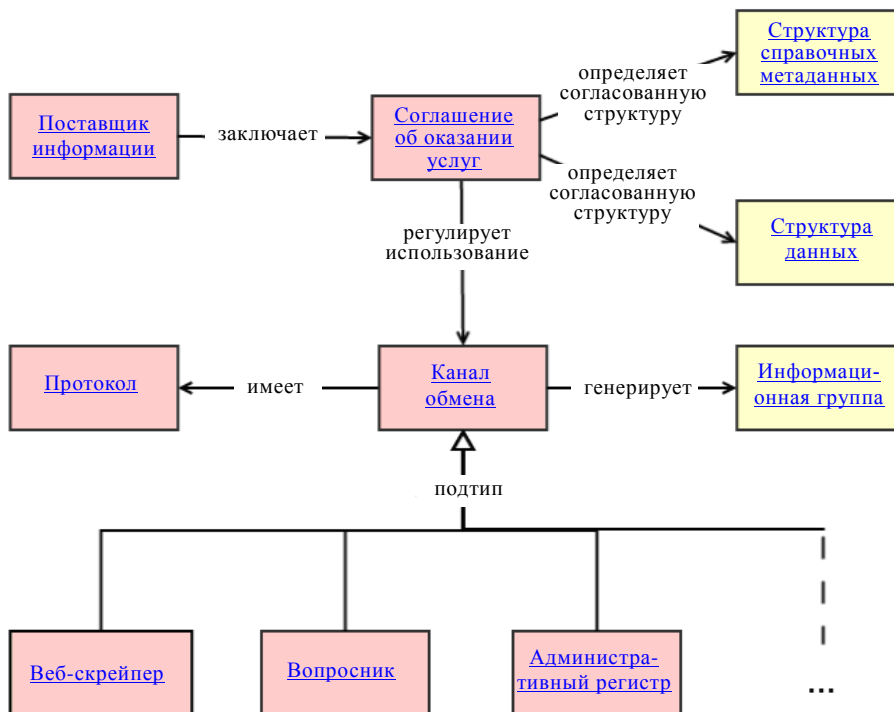
35. Механизм обмена информацией через канал обмена определяется протоколом (например, веб-услуга на базе ОСДМ, обмен файлами данных, личный опрос).

36. В процессе сбора информации статистическая организация в соответствии с протоколом и соглашением об оказании услуг получает от поставщиков данные и справочные метаданные, которые через канал обмена превращаются в информационный пакет. При распространении данных информационный пакет поступает в канал обмена и в соответствии с протоколом и соглашением об оказании услуг доводится до потребителей. Более подробную информацию о процедуре сбора и распространения данных можно найти в следующих разделах.

5. Сбор информации

Диаграмма 8

Каналы обмена, используемые для сбора информации



37. В ТМСИ содержатся три примера возможных каналов обмена информацией: вопросник, веб-скрейпер и административный регистр. Каждый из них подробно описывается в приложении А. Статистические организации могут собирать данные и справочные метаданные у поставщиков информации и через дополнительные каналы обмена, например обмениваясь файлами, используя веб-услуги и сканируя данные. С учетом своих потребностей статистические организации могут вводить в ТМСИ и другие дополнительные каналы.

38. Порядок использования канала обмена для сбора информации регулируется соглашением об оказании услуг между статистической организацией и поставщиком информации. Обе стороны используют соглашение для согласования структуры подлежащих обмену данных и справочных метаданных. Механизм сбора информации через канал обмена определяется протоколом (личный опрос, обмен файлами данных, веб-робот). Собирающая информацию организация использует ее для подготовки информационного пакета, который может содержать данные или справочные метаданные.

7. Обработка и анализ информации

39. В ТМСИ процедуры обработки и анализа информации описываются очень гибко.

40. Процесс статистического производства можно рассматривать под углом зрения данных⁸. Статистические организации стараются готовить качественные и точные данные, которые, для того чтобы сделать их максимально полезными, подкрепляются метаданными. В этой связи уместно говорить об эволюции данных по мере прохождения ими через все этапы статистического процесса. Многие операции связаны с метаданными, однако в конце производственного процесса метаданные становятся всего лишь вспомогательным ресурсом с точки зрения данных – конечного статистического продукта. Эту связь между данными и метаданными важно хорошо понимать.

41. Собранные данные поступают в статистическую организацию по каналам обмена. Вне зависимости от того, как собирались данные и каково их происхождение, они представляют собой ресурс, трансформирующийся в процессе прохождения многочисленных этапов. Исходные данные определяются как группа данных, имеющих соответствующую структуру. Группы данных хранятся в организованном виде в виде ресурсов данных. Они являются первичным материалом на входе и на выходе разных этапов статистического процесса, осуществляемых в рамках статистической программы.

42. В процессе того, как статистическая организация трансформирует сырые исходные данные во все более точную информацию, на каждом этапе такой переработки в ресурсах данных появляются все новые группы данных. Методика этого процесса может быть самой разной. В рамках статистической организации он проходит через различные этапы процесса.

43. В определенный момент (который может наступать на разных этапах производственного процесса в зависимости от того, какому виду редактирования подвергаются данные) данные подвергаются анализу с целью производства статистических продуктов. Анализ данных можно рассматривать как использование групп данных из ресурса данных в качестве исходного материала таких процедур, как составление доверительных интервалов или объяснения данных. Выполняемые в ходе анализа операции различаются в зависимости от того, какими должны быть конечные продукты – точечными данными в пределах доверительного интервала или агрегированными показателями и таблицами, используемыми в конкретных политических целях. Несмотря на различные виды анализа, суть процесса остается той же – преобразование информационного ресурса.

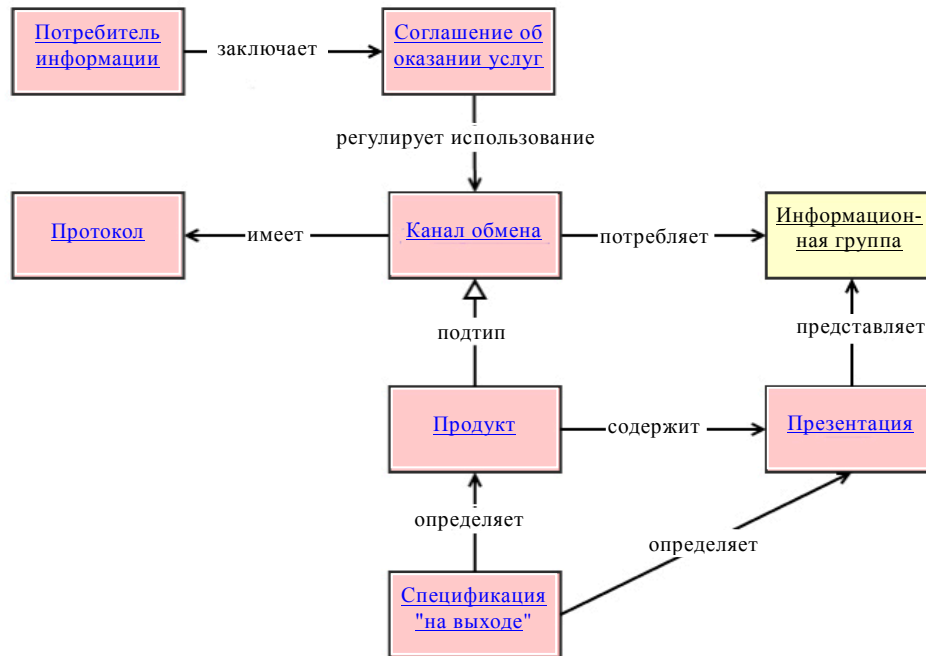
44. В прошлом считалось, что за сбором данных обязательно следует их обработка, анализ и распространение статистического продукта. Такая процедура требует больших затрат времени и ресурсов. Один из способов повысить эффективность работы статистической организации заключается в том, чтобы повторно использовать данные для производства новых запрашиваемых продуктов, тем самым сокращая расходы и продолжительность процесса. В этом смысле ресурсы данных можно рассматривать как актив организации, который следует беречь и использовать максимально продуктивно.

⁸ Хотя данные пункты посвящены данным, все, что в них говорится, применимо и к справочным метаданным.

8. Распространение информации

Диаграмма 9

Канал обмена для распространения информации



45. Статистическая организация направляет статистическую информацию потребителю.

46. Потребитель информации получает доступ к группе данных через продукт (или потенциально через другой канал обмена), в котором содержится одна или несколько презентаций данных. Каждая такая презентация определяет или описывает структуру представляемых данных и, возможно, справочных метаданных в текстовом формате, например в виде отчетов о качестве.

47. Презентация может принимать разные формы: например, данные могут быть представлены как графическое отображение таблицы данных на экране в виде HTML-страницы, загружаемый файл в формате PDF, а также как ОСДМ-файл в формате XML.

48. В спецификации конечного продукта определяется содержание презентации. Продукт, в который упакованы презентации, может представлять собой стандартный специфичный конечный результат работы организации, который можно найти в:

- регулярно выходящем статистическом бюллетене (например, в ежемесячной публикации индекса розничных цен);

- в динамическом статистическом пакете, созданном после получения запроса от потребителя информации, желающего получать данные через ИПП (интерфейс прикладного программирования); или
- в той или иной программе поиска данных, которая может быть встроена в веб-сайт статистической организации.

49. В спецификации конечного продукта определяется также, какие данные из информационной группы требуются для презентации. Спецификации часто формулируются в рамках той или иной внутренней процедуры организации, по результатам которой должен появиться конкретный статичный продукт (например, упомянутый выше статистический бюллетень). Что касается динамических продуктов, те или иные аспекты спецификации могут определяться потребителем информации во время ее просмотра в рамках интерфейса машина–машина, свидетельством чему может служить упоминавшийся выше сценарий ИПП. В любом случае, получение запроса ведет к появлению спецификации конечного продукта, определяющей, какая информация из группы данных и/или какие справочные метаданные войдут в презентацию.

50. Механизм представления продукта определяется протоколом (SDMX-ML, DDI XML, PDF и т.д.). Такое форматирование информации является составной частью составления спецификации конечного продукта, на основе которой готовится продукт и его презентация в соответствующем формате.

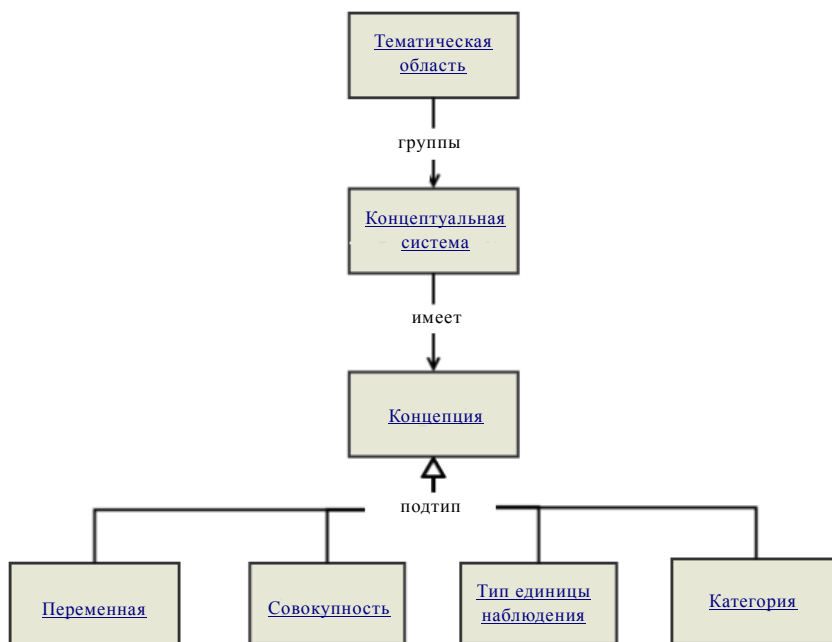
51. В зависимости от сценария запроса потребители информации могут принимать самые разные формы. Таким потребителем может быть физическое лицо, заходящее на веб-сайт статистической организации и просматривающее содержимое веб-страницы, или компьютерная программа, запрашивающая информацию через ИПП путем отправки ОСДМ-запроса. Доступ потребителя к информации регламентируется соглашением об оказании услуг, в котором установлены условия получения и использования информации. Пользователь может дать свое согласие с условиями использования данных в пассивной форме при посещении веб-сайта или пройти регистрацию, если ему требуется более подробная информация, которую можно получить только через ИПП.

В. Основополагающая информация

52. Концептуальные и структурные группы ТМСИ содержат объекты информации, которые являются основополагающими для процесса статистического производства. Иными словами, эти объекты носят концептуальный и структурный характер и используются в статистическом процессе как "на входе", так и "на выходе". Концептуальная область ТМСИ содержит группы информационных объектов, описывающих и определяющих те термины, которыми статистики пользуются, рассматривая реальные события. Структурная область включает группы информационных объектов, относящихся к данным и справочным метаданным и их структуре. Те объекты, о которых пойдет речь в настоящем разделе документа, дают пользователям данных и метаданных информацию, которая помогает им понять результаты производственного процесса и статистических программ.

1. Концепции

Диаграмма 10
Концепции



53. На абстрактном уровне концепция определяется как "понятие, отличающееся своими характеристиками". На протяжении статистического цикла жизни концепции используются по-разному, и каждая роль концепции описывается при помощи своего информационного объекта (объекты являются подтипами концепции). Концепция может использоваться в этих ситуациях следующим образом:

а) как характеристика. Концепция используется в переменной для описания конкретной измеряемой характеристики совокупности. Например, для измерения концепции пола применительно к совокупности взрослых жителей Нидерландов переменная увязывает эту концепцию с "типичным" человеком;

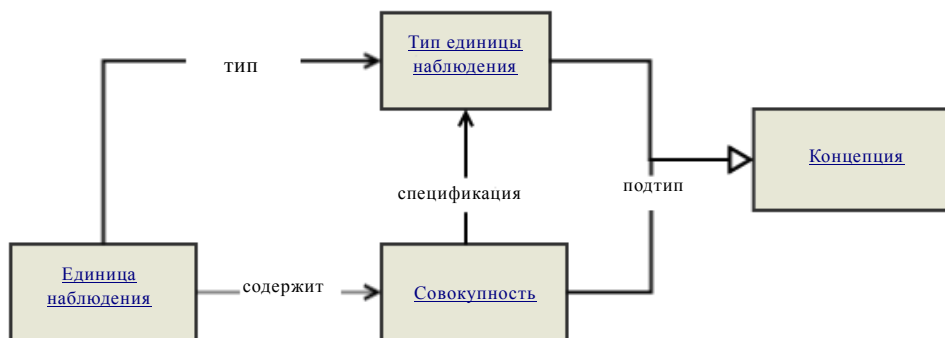
б) как типичная единица наблюдения или как совокупность для описания группы объектов, о которых планируется собрать информацию в ходе статистического обследования. Например, совокупность взрослых жителей Нидерландов как "типичных" людей;

в) как категория для дальнейшей детализации концепции. Например, "мужчина" и "женщина" для концепции пола. Коды могут быть увязаны с категорией через узловые понятия (кодовый или классификационный номер), используемые, например, в списке кодов или в статистической классификации;

54. Концептуальные системы представляют собой наборы концепций, структурированные с учетом связей между отдельными концепциями. Концептуальные системы образуют тематические области в конкретных сферах знаний (например, рынок труда, туризм).

2. Совокупность

Диаграмма 11
Совокупности и статистические единицы

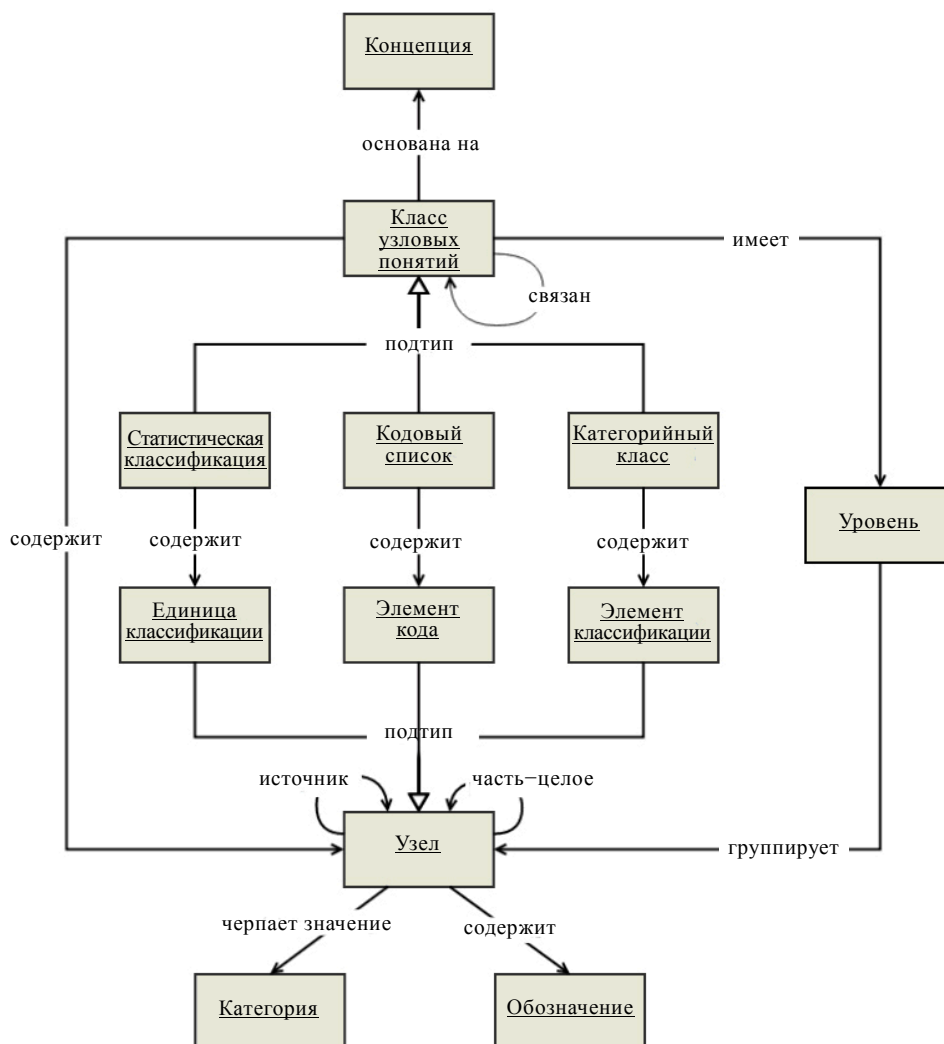


55. Существует несколько видов совокупностей в зависимости от того, на каких процессуальных этапах они используются. Например, статистическая организация может говорить о целевой, обследуемой, базовой или анализируемой совокупности. Представляющим интерес объектом в ходе статистического процесса является статистическая единица наблюдения (например, физическое лицо или предприятия). Именно о них собираются данные. В модели единица наблюдения может быть определена двояко. Единицу наблюдения можно определить как индивидуальный субъект, связанный с той совокупностью, о которой может быть получена информация. Тип единицы (например, физические лица или предприятия) позволяет определить абстрактный вид единицы наблюдения, которую измеряет та или иная переменная.

3. Узловое понятие и классы узловых понятий

Диаграмма 12

Узловое понятие и классы узловых понятий



56. Категория представляет собой конкретный вид концепции, чья роль заключается в определении характеристики. Существует три вида использования категорий. В ТМСИ они определяются как три подвида узловых понятий: элемент категории, элемент кода и элемент классификации. В зависимости от возможного использования категории объединяются в классы узловых понятий. Существует три подвида таких групп (классов узловых понятий): категорийные классы, кодовые списки и статистические классификации.

57. Категорийный класс представляет собой класс категорийных элементов, несущий в себе значение категории без соответствующего представления. "Мужской пол, женский пол" является примером категорийного класса.

58. В кодовом списке кодовые позиции не только несут в себе значение категорий, но и представлены соответствующим образом. Можно привести следующий пример кодового списка: 1. мужской пол, 2. женский пол.

59. Статистическая классификация схожа с кодовым списком. Она сочетает значение категории с кодовым представлением. В то же время содержание статистической классификации должно удовлетворять определенным критериям и иметь определенный статус. Элементы классификации должны быть взаимоисключающими и в сумме своей исчерпывающими на том уровне, на котором они представлены в статистической классификации. Пример статистической классификации: 1. мужской пол, 2. женский пол, 3. интерсекс.

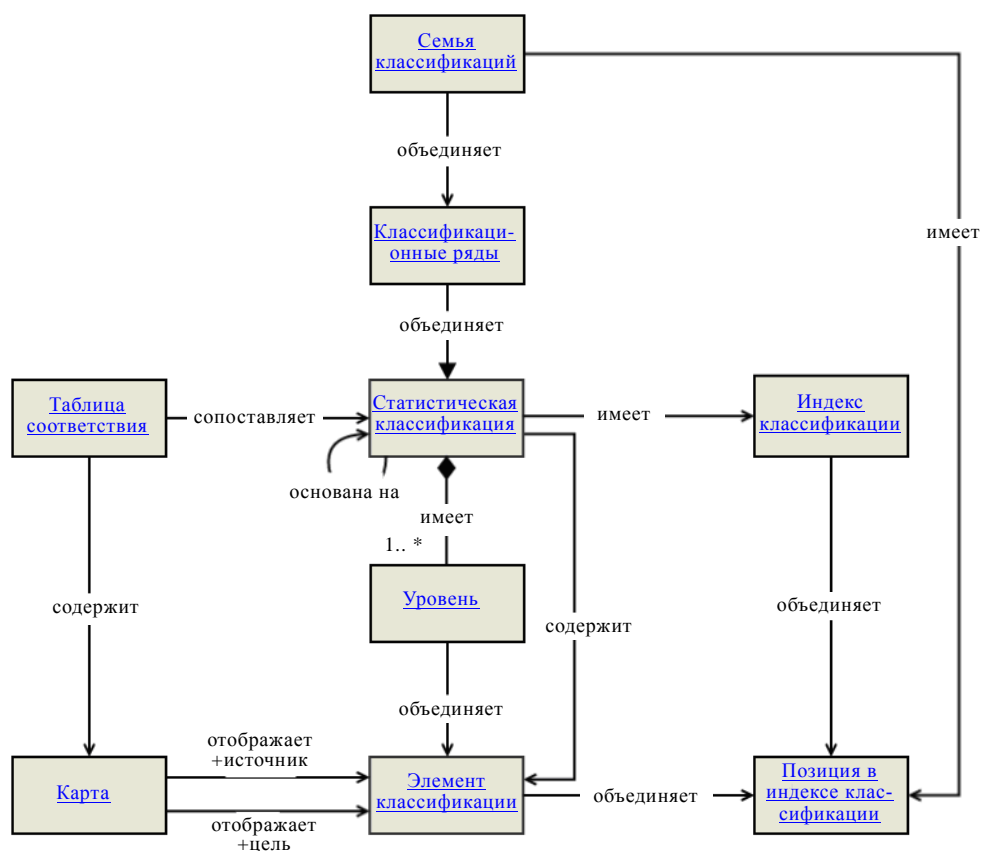
60. Кодовый список не обязательно должен удовлетворять тем же критериям, что и статистическая классификация. Кодовый список может содержать и дополнительные элементы с учетом его конкретного предназначения, например восполнения недостающих значений.

61. Сходства статистических классификаций, кодовых списков и категориальных классов объясняются их связью (как подтипов) с классом узловых понятий. Аналогично этому, три вида элементов каждой группы (соответственно элемент классификации, элемент кода и элемент категории) являются подтипами узлового понятия.

4. Статистическая классификация

62. В настоящем разделе статистическая классификация и связанные с ней объекты управления описываются с точки зрения того конкретного раздела ТМСИ, который конкретно посвящен классам узловых понятий. Более подробные сведения о статистических классификациях можно найти в модели статистической классификации ТМСИ.

Диаграмма 13
Статистические классификации



63. Приводимая выше диаграмма дает общее представление об объектах статистических классификаций.

64. Семья классификаций представляет собой группу, состоящую из серий классификаций, объединяемых общей концепцией (например, экономическая деятельность). Серия классификаций представляет собой одну или несколько классификаций, в основе которых лежит общая концепция. Статистические классификации в рамках серии связаны между собой как версии или обновленные варианты одной классификации. Такие статистические классификации, как правило, имеют общее название (например, Международная стандартная отраслевая классификация всех видов экономической деятельности (МСОК) или Международная стандартная классификация занятий (МСКЗ)).

65. Статистическая классификация представляет собой набор категорий, которые могут быть связаны с одной или несколькими представляемыми переменными, используемыми для производства и распространения статистической информации. Категории на каждом уровне структуры классификации должны быть взаимоисключающими и в совокупности своей исчерпывающими, включая все объекты/единицы наблюдения в рассматриваемой совокупности. Одним из примеров статистической классификации является четвертый пересмотренный вариант МСОК.

66. Категории определяются со ссылкой на одну или несколько характеристик интересующей совокупности. Статистическая классификация может иметь плоскую, линейную структуру или быть выстроена по иерархическому принципу, в соответствии с которым все категории на низшем уровне являются подкатегориями категорий следующего уровня.

67. Статистическая классификация включает категории, представленные элементами классификации. Эти элементы классификации выстроены по уровням по иерархическому принципу. Уровень представляет собой группу взаимоисключающих и в сумме своей исчерпывающих концепций. Примером уровня могут служить раздел, подраздел, группа и подгруппа четвертого пересмотренного варианта МСОК.

68. Элемент классификации должен объединять в себе значение категории, ее представление (например, код) и дополнительную информацию, с тем чтобы отвечать критериям статистической классификации (например, "А – сельское хозяйство, лесоводство и рыболовство"), а также сопровождаться пояснительным текстом, т.е. сведениями о том, что входит, а что не входит в его охват.

69. Статистические классификации могут составляться как версии и как разновидности. В основе разновидностей статистических классификаций лежат их версии. В разновидности классификации категории той или иной версии могут быть разукрупнены, укрупнены или перегруппированы, с тем чтобы дополнить стандартный порядок и структуру исходной статистической классификации и предложить ей альтернативу.

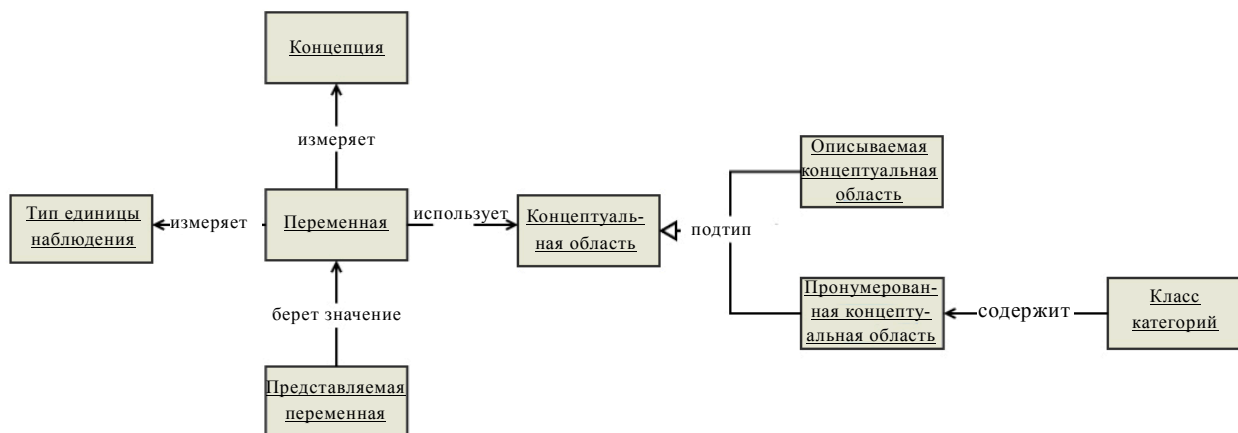
70. Таблица соответствия представляет собой набор карт. Эти карты увязывают элемент классификации в статистической классификации с соответствующим элементом классификации в другой статистической классификации через концепцию, которая является общей для обоих элементов классификации. Например, в таблице соответствия, отражающей связь между четвертым пересмотренным вариантом МСОК и Североамериканской системой отраслевой классификации продуктов (НАИКС, 2007 год (США)), "0112 – выращивание риса" в МСОК, Rev.4, соотносится с "111160 – рисоводство" в НАИКС через общую концепцию "выращивание риса".

71. Индекс классификации указывает на связь текста в источниках статистических данных (ответы на вопросники, административные регистры) с одной или несколькими статистическими классификациями. Индекс классификации может быть использован для присвоения кодов элементов классификации наблюдениям в статистических программах.

72. Отдельные позиции в индексе классификации представляют собой отдельно взятое слово или краткий текст (название населенного пункта, вид экономической деятельности или занятие), описывающие вид концепции, к которой относится элемент классификации, наряду с кодом соответствующего элемента классификации. Каждая позиция в индексе классификации обычно соответствует одному элементу статистической классификации. Хотя позиции в индексе классификации могут быть связаны с элементами классификации на любом уровне статистической классификации, они, как правило, соотносятся с элементами классификации самого нижнего уровня.

5. Переменная

Диаграмма 14
Переменная



73. При использовании в рамках производственного процесса тип статистической единицы, определяющей совокупность, соотносится с определенной характеристикой. Связка типа статистической единицы и концепции, играющей роль характеристики, именуется "переменной" (см. диаграмму 14). Например, если совокупность представляет собой всех взрослых жителей Нидерландов, то соответствующей переменной может быть концепция уровня образования типа статистической единицы "человек".

74. Переменная (уровень образования человека) не содержит никакой информации о том, как может быть представлено соответствующее ей значение. Эта информация (область значений) имеет связь с понятием представленной переменной. Данное различие позволяет несколько раз использовать одно и то же определение переменной в тех случаях, когда объект измерения остается неизменным с концептуальной точки зрения, но представлен иначе.

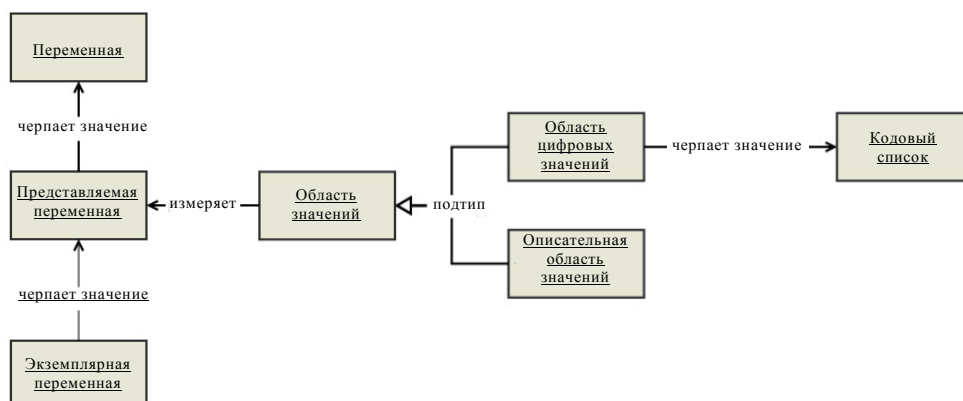
75. В ходе этапа процесса с применением процессового метода к одному или нескольким исходным параметрам (переменным) создается производная переменная. Такая переменная процесса является результатом осуществления этапа процесса.

76. С переменной соотносится концептуальная область. Существует два подвидов концептуальной области: описательный и пронумерованный. Пронумерованная концептуальная область в сочетании с категориальным классом содержат информацию о семантике используемых в переменной категорий.

6. Представление переменной

77. ТМСИ помогает пользователям лучше понять как смысл объекта, так и конкретную форму его представления. Соответственно, для разграничения объектов, используемых для концептуального описания информации и ее представления, в ТМСИ проводится различие между концепцией модели и ее представлением.

Диаграмма 15
Представление переменной



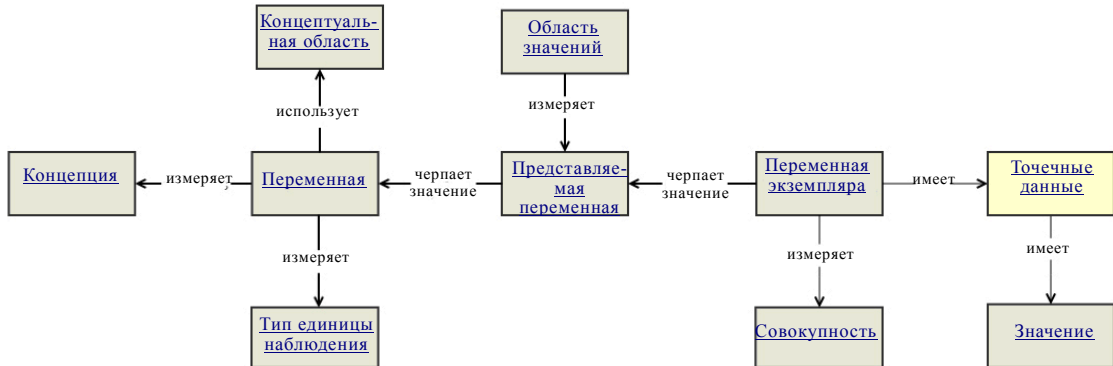
78. Представляемая переменная (см. диаграмму 15) дает дополнительную информацию, описывающую, как полученные результаты можно отобразить через их связь с областью значений. В отличие от концептуальных областей, связанных с переменными, области значений соотносятся с представляемыми переменными. Различие между этими областями проводится в силу того, что в ТМСИ семантический аспект (концептуальная область) отделен от представления информации (область значений).

79. Как цифровая, так и описательная область значений (два подтипа области значений) дают информацию о том, как отображается представляемая переменная. Цифровая область значений выполняет эту функцию совместно с кодовым списком, в то время как область описательных значений описывает, как формируется значение, а не перечисляет их.

80. Область значений содержит информацию о типе данных и единицах измерения. Информация о типе данных дает представление о том, какие вычислительные операции можно проводить с данными (номинальные, порядковые, интервальные и иные данные), а информация о единицах измерения (тонны, доллары, и т.д.) уточняет количественные данные области значений. Например, гендерные коды позволяют получить номинальные статистические данные, а возрастные данные в годах – интервальные данные.

6. Переменная экземпляра

Диаграмма 16
Переменная экземпляра

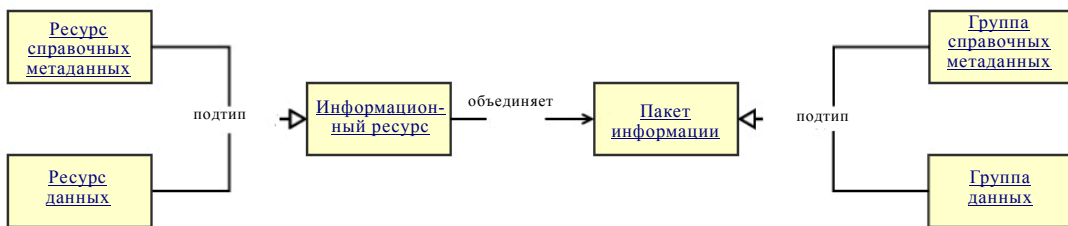


81. Переменная экземпляра (см. диаграмму 16) представляет собой представляемую переменную, связанную с группой данных. Она может соответствовать столбцу данных в базе данных. Например, "возраст всех президентов США в данный момент времени (если они живы) или в момент смерти" является столбцом данных, описываемых переменной экземпляра, сочетающей в себе представляемую переменную "возраста человека" и область значений "десятичные натуральные числа (в годах)".

82. Значение содержится в группе данных в специально отведенном для него месте (точке). Оно может выражаться в единицах измерения из области значений, соотносимых с описывающей переменной экземпляра, и быть связано либо с единичным наблюдением (точечные данные), либо с совокупностью (размерные данные).

8. Информационные ресурсы

Диаграмма 17
Информационные ресурсы



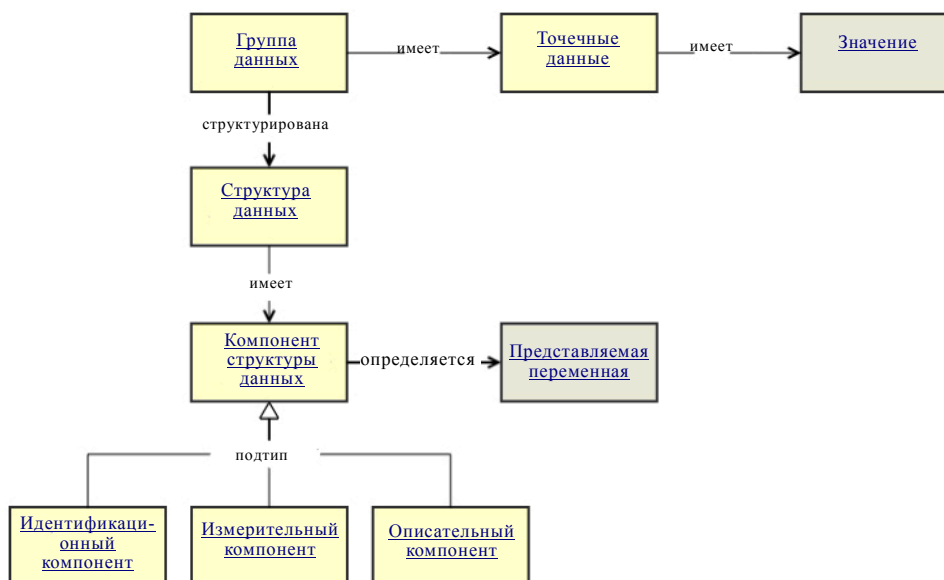
83. Статистические организации собирают, обрабатывают, анализируют и распространяют пакеты информации в форме либо данных (группы данных), либо справочных метаданных (группы справочных метаданных).

84. Каждая группа данных должна быть организована в соответствии со структурой данных (например, структурой платежного баланса, структурой народонаселения, туристических потоков, образовательных услуг и т.д.). Аналогичным образом группа справочных метаданных должна быть составлена в соответствии со структурой справочных метаданных (например, системой контроля качества организации).

85. Информационные ресурсы содержат пакеты информации. Главная задача информационного ресурса заключается в том, чтобы помочь в поиске и использовании пакетов информации, указывая их местоположение и предоставляя другие сведения, облегчающие выполнение этой задачи. Информационный ресурс может быть двух видов. Ресурс данных содержит группы данных, а ресурс справочных метаданных – группу справочных метаданных.

9. Группы данных

Диаграмма 18
Группы данных



86. Данные группы имеют свои "точки", т.е. конкретно отведенные для них места (например, ячейка в таблице), в которых отражаются значения данных, представляющие собой, например, единицы фактической информации, полученные в результате измерения или производственного процесса. Структура данных описывает структуру группы данных при помощи компонентов структуры данных (идентификационные компоненты, измерительные компоненты и описательные компоненты). Все они представляют собой представляемые переменные, выполняющие конкретные функции.

87. Группы данных могут существовать в разных формах, например в форме административных регистров, временных рядов, данных многомерных рядов, данных выживаемости и т.д. От типа группы данных зависят группа конкретных атрибутов, которые необходимо описать, вид требующейся структуры данных (структура точечных данных или структура размерных данных) и метод, применимый к данным.

88. Например, административный регистр имеет структуру точечных данных, обладающих такими атрибутами, как цель первоначальной записи или дата ее последнего обновления. Он содержит переменную, позволяющую идентифицировать ту или иную запись, и может использоваться для определения совокупности, используемой в качестве основы выборки в целях замены или дополнения существующих обследований или в качестве вспомогательного источника информации для импутации. Примером метода, специально применяемого к регистрам, может служить метод увязки данных.

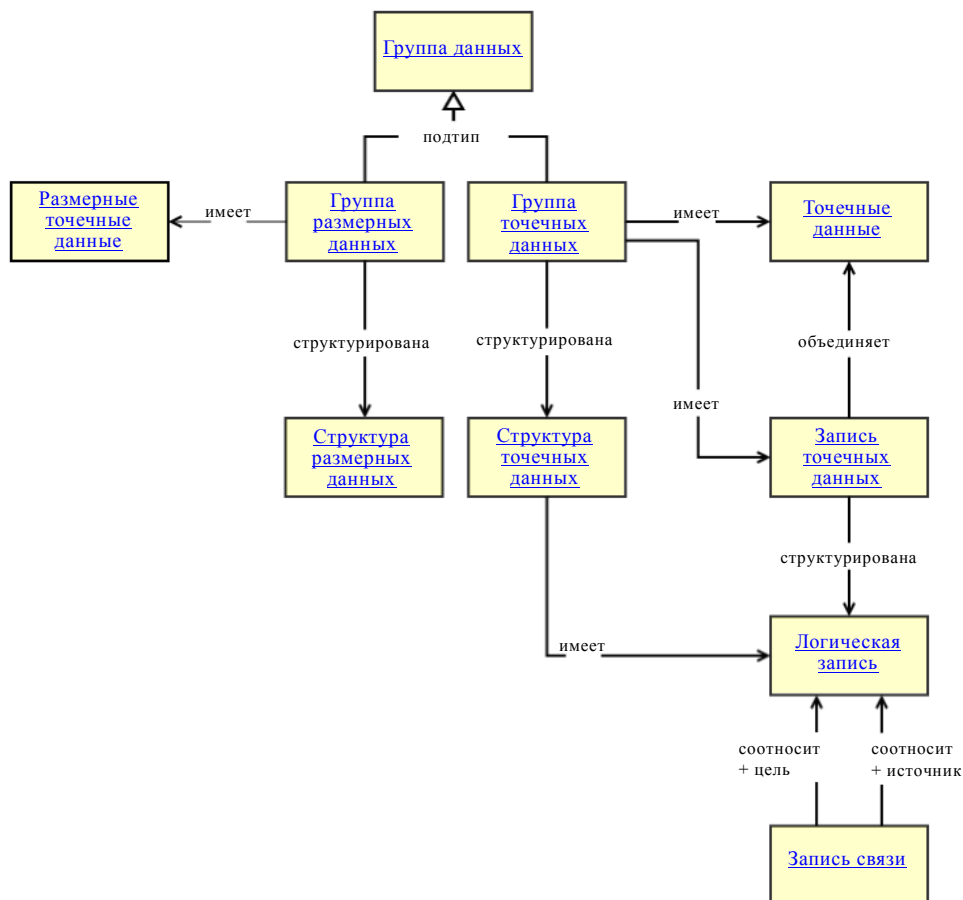
89. Примером группы данных, имеющей размерную структуру, являются временные ряды. Они имеют конкретные атрибуты, например частотность и тип агрегирования во времени, а также метод агрегирования, такой как сезонная корректировка, и должны содержать временную переменную.

90. Вопрос о том, являются ли данные точечными или размерными, является во многом субъективным. Некоторые группы данных можно считать и теми, и другими, хотя это не является общим правилом. Иногда данные могут считаться размерными одной организацией (например, национальным статистическим управлением) и точечными другой (например, Евростатом, для которой единицей является государство-член). Тот или иной конкретный набор данных совершенно необязательно считать либо одним, либо другим. Взгляд на данные во многом связан концептуальным подходом. В ТМСИ разница заключается в том, что группа точечных данных содержит данные о статистических единицах, а группа размерных данных – данные либо о единицах, либо о совокупности.

91. ТМСИ предполагает, что все группы данных должны иметь свою структуру. В то же время возможны случаи, когда группа данных структуры не имеет, поскольку эти данные не были зарегистрированы, были потеряны или о них ничего не известно. В будущем статистические организации будут, по всей видимости, сталкиваться с подобными случаями все чаще. Для того чтобы статистическая организация могла пользоваться такими данными, они должны пройти этап структурирования. Например, в случае поиска новых потенциальных источников данных для удовлетворения новых или изменившихся статистических потребностей потребуется проанализировать эти данные, с тем чтобы определить их содержание и структуру. Лишь после этого можно будет приступить к описанию этой новой группы данных при помощи объектов структуры данных. Неструктурированные данные сегодня описываются в ТМСИ лишь как исходные данные процесса. Организации могут распространить действие ТМСИ и на подобные случаи, создав новый подтип объекта информационной группы.

10. Структура размерных и точечных данных

Диаграмма 19
Структуры данных



92. Структура размерных данных описывает структуру группы размерных данных при помощи представляемых переменных, выполняющих конкретные функции.

93. Сочетание размеров, закрепленных в структуре размерных данных, служит ключом или идентификатором измеряемых величин. Например, страна, индикатор, единица измерения, частота и временные параметры вместе идентифицируют ячейки межстрановых временных рядов со множественными показателями (например, валовой внутренний продукт, общая сумма внутреннего долга), имеющими разные единицы измерения (например, разные валюты, изменения в процентах) и разную частотность измерения (ежегодно, ежеквартально). Ячейки такой многомерной таблицы содержат наблюдаемые значения.

94. Величина представляет собой переменную, являющуюся оболочкой для наблюдаемых значений. Ее семантика связана с подгруппой размеров в структуре размерных данных. В предыдущем примере источниками такой семантики можно считать индикатор и единицу измерения, в то время как частотность и время являются временными измерениями, а страна – географическим. Примером величины, помимо простого "наблюдаемого значения", может служить "значение, наблюдаемое до прерывания временного интервала", когда речь идет

о временных рядах. Размерные данные, как правило, ассоциируются с представляемыми переменными с кодовыми областями значений (цифровые области значений), а величины – с представляемыми переменными с областями значений, не снабженных кодами (описательные области значений).

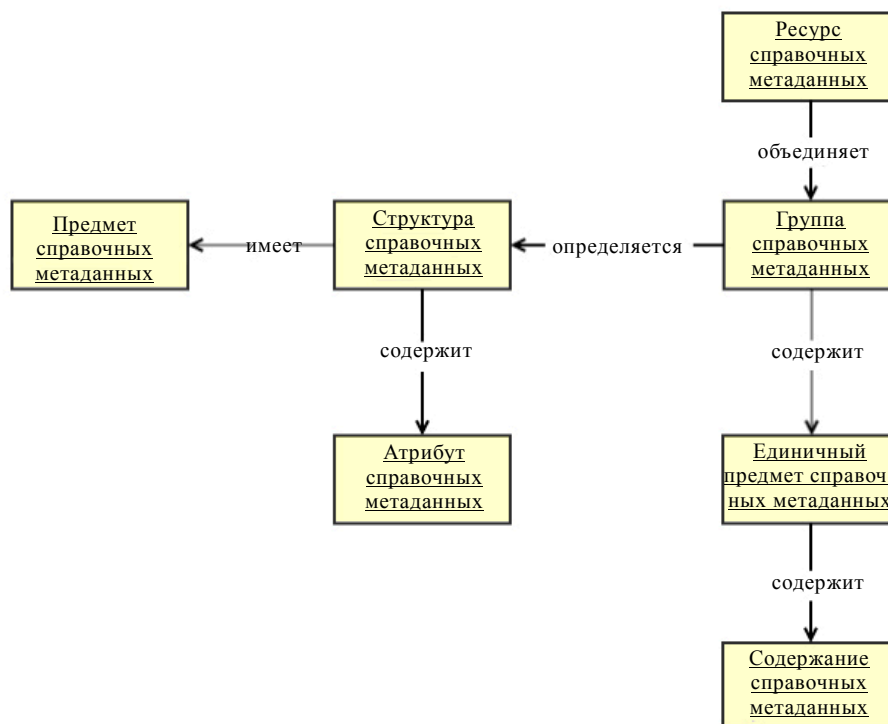
95. Структура точечных данных описывает структуру группы точечных данных при помощи представляемых переменных, выполняющих конкретные функции, при этом проводятся различия между логической и физической структурой группы данных. Группа точечных данных может включать данные нескольких видов, имеющих свой особый порядок регистрации.

96. Логические записи описывают структуру соответствующих видов данных вне зависимости от физических особенностей. Для этого используются представляемые переменные, которые могут предусматривать идентификацию наблюдаемой единицы (например, номер домохозяйства). Под функциональной зависимостью записей понимается связь между логическими записями типа "исходное значение – целевой показатель".

11. Группы справочных метаданных

Диаграмма 20

Группы справочных метаданных



97. Информация, описывающая характеристики статистических данных, именуется "справочными метаданными". Эти метаданные могут быть широкими и описывать всю статистическую программу или узкими, касающимися отдельных точечных данных. Ресурсы справочных метаданных, представляющие собой особый вид информационных ресурсов, являются высшим уровнем хранения справочных метаданных.

98. Группа справочных метаданных систематизирует справочные метаданные, структура которых определяется структурой справочных метаданных. Структура справочных метаданных определяет предмет справочных метаданных, которому могут быть посвящены справочные метаданные, и содержит структурированный перечень атрибутов справочных метаданных, которые могут быть воспроизведены или разработаны применительно к конкретному предмету справочных метаданных.

99. Этими предметами могут являться любые типы объектов ТМСИ или любые точечные данные или группы точечных данных, сформированные на базе структуры данных.

- Пример типа объектов ТМСИ как предмета справочных метаданных: продукт, для которого имеется свой перечень из области значений. В области значений перечисляются реальные продукты, применительно к которым могут быть воспроизведены или разработаны справочные метаданные с использованием структуры справочных метаданных.
- Примерами атрибутов справочных метаданных являются статус, охват, методологическое описание и качественные показатели.

100. Группа справочных метаданных содержит реальные воспроизводимые или специально разработанные метаданные. Предмет справочных метаданных определяет реальный объект, т.е. реальный продукт (например, платежный баланс, позиция международных инвестиций, Австралия, июнь 2013 года) или реальные точечные данные (например, точечные данные для одного региона в рамках группы данных по всем регионам страны).

101. Каждый элемент справочных метаданных имеет свое фактическое наполнение – содержание справочных метаданных. Содержание каждого элемента справочных метаданных содержит справочные метаданные, касающиеся одного из атрибутов справочных метаданных, определенных в структуре справочных метаданных.

Таблица 1

Пример использования объектов справочных метаданных ТМСИ

Объект ТМСИ	Статистический бюллетень УНС: государственные финансы, октябрь 2013 года: Таблица 1
<i>Структура справочных метаданных</i>	Имплицитная
<i>Предмет справочных метаданных</i>	Компонент структуры данных
<i>Атрибут справочных метаданных</i>	Табличная сноска
<i>Группа справочных метаданных</i>	Сноски
<i>Элемент предмета справочных метаданных</i>	<i>Компонент структуры данных: £миллиард; текущий бюджет госсектора; текущий государственный бюджет исключая ex-APF</i>
<i>Элемент контента справочных метаданных</i>	Текст сноски