

7-8 декабря 2011 г. в штаб-квартире Статистического института Азии и Тихого океана (СИАТО) в г.Тиба (Япония) состоялся 10-й семинар высокого уровня для руководителей национальных статистических служб региона ЭСКАТО, в работе которого приняла участие делегация Росстата. Среди участников семинара присутствовали руководители статистических служб Австралии, Бангладеша, Бутана, Вануату, Вьетнама, Гонконга, Грузии, Ирана, Индии, Индонезии, Камбоджи, Лаоса, Малайзии, Мьянмы, Монголии, Науры, Новой Зеландии, Пакистана, Самоа, Тимор-Лесте, Фиджи, Филиппин, Шри-Ланки, Южной Кореи и Японии, а также представители ЭСКАТО и ряда других международных организаций.

В рамках повестки дня семинара были рассмотрены актуальные вопросы развития современной официальной статистики: модернизация статистических информационных систем, повышение качества статистической информации, удовлетворение потребностей пользователей в интегрированных статистических данных по социальной, экономической и природоохранной статистике и др. Особое внимание было уделено вопросам наращивания потенциала официальной статистики в регионе ЭСКАТО с учетом новых вызовов глобальной статистической системы.

Основной доклад, подготовленный Главным статистиком Австралии г-ном Брайаном Пинком, публикуется ниже.

### МОДЕРНИЗАЦИЯ СТАТИСТИКИ И МЕЖДУНАРОДНОЕ СОТРУДНИЧЕСТВО

**Брайан Пинк,**

*Главный статистик Австралии*

Очевидно, что цифровая информация создается и хранится в обществах XXI века на совершенно феноменальном уровне, хотя никто из нас, вероятно, в полной мере не может оценить эту феноменальность. Цифровые данные поступают из всех возможных источников информации, причем не только на основе процессов делового обмена, исследовательских и административных систем государственных и частных организаций, но также благодаря различным измерениям, специальным инструментам, компьютерным программам и моделям. Неимоверные объемы информации перемещаются по коммуникационным, транспортным, защищенным и логистическим сетям. И разумеется, в наше время социальные сети, такие, как Facebook и Twitter, вырабатывают огромные объемы цифровых данных с камер, мобильных телефонов и других персональных цифровых устройств. Учитывая их общее количество, этот экспоненциальный рост цифровых данных превосходит не только способность многих существующих систем и инструментов, созданных для их упорядочения, анализа и хранения, но также указывает на вызовы, связанные с обучением нового поколения бизнесменов, политиков, ученых и людей искусства навыкам того, как понимать и использовать эти данные. Мы, официальные статистики, всегда будем испытывать подобного рода давление.

За последние годы все, что связано с информацией, связью и вычислительной техникой, изменилось кардинальным образом. И становится очевидным, что мы в действительности находимся лишь в начале «информационной эры». Поэтому хотелось бы привести пример того, насколько быстро меняется мир данных, и процитировать журнал Economist, прошлогодний февральский номер. Если мы посмотрим на отрасль, с ко-

торой многие из нас неплохо знакомы, а именно на розничную торговлю, то Walmart, крупнейшая розничная сеть США, ежечасно обслуживает более миллиона потребительских операций, наполняя базу данных, которая, по оценкам, «весит» более 2,5 петабайт ( $2^{50}$ ). Google же, по данным Economist, как будто бы обрабатывает 1 петабайт информации в час. CISCO оценивает, что объем трафика данных, перемещающихся в Интернете, достигает примерно 670 экзабайт ( $2^{60}$ ).

До сих пор мы довольно неплохо справлялись с многочисленными проблемами, связанными с данными, и можем полагать, что все, что работало на нас раньше, будет работать на нас и в дальнейшем. В связи с этим рассмотрим тенденции изменения ситуации от 1980-х годов до наших дней.

**Традиция: что привело нас из 80-х и 90-х в наши дни.** Для сообщества национальных статистических ведомств период 1980-х - 1990-х годов характеризовался следующими факторами:

- обособленная деятельность по сбору статистических данных в рамках наших собственных организаций;
- преимущественно неупорядоченный и случайный обмен знаниями и системами/инструментарием;
- развитие традиционных систем управления информацией;
- статистические процессы, сосредоточенные на бумаге, с выходом на печатные издания;
- «робкие набег» в Интернет для сбора данных, распространения и использование формата pdf.

На протяжении последних 10 лет мы наблюдали различные тенденции в области технологий, с которыми процессы и инструменты, разработанные нами для управления статистическими системами, уже не справляются:

- получение данных из Интернета (перепись, обследования предприятий, обследования домашних хозяйств);

- быстро эволюционирующие источники данных (административные данные - здравоохранение, образование, налогообложение и т. д.; данные об операциях - со считывающих устройств, кредитных карт и т. д.).

Быстрое расширение объема данных, о которых всего несколько лет назад большинство людей не имели представления, сегодня стало реальностью нашей повседневной жизни. И хотя многие организации сами не накапливают цифровые данные в таком объеме, проблемы, с которыми тем не менее сталкиваются эти организации в XXI столетии, заставляют их признать, что имеет место бурный рост объема цифровых данных, и решить вопрос о том, как использовать в полной мере возможности, которые им предоставляются.

**Возрастающий спрос против уменьшающихся ресурсов.** В то время как национальные статистические организации стремятся удовлетворить возрастающий спрос на данные, их ресурсы уменьшаются. При этом мы сталкиваемся с возрастающими ожиданиями наших пользователей в таких областях, как:

- интеграция данных - увязка источников данных, комплексные информационные решения;

- распространение данных через Интернет, услуги веб-сайтов, стыковка компьютеров, запросы данных в режиме он-лайн;

- коллективное использование данных: через веб-сайты правительственных организаций, через семантические вебы.

Бурный рост объема цифровых данных, иногда описываемый как эра «больших данных», - это то, с чем я столкнулся, работая в качестве руководителя Австралийского бюро статистики (АБС). Правительственные органы Австралии нуждаются в точной информации для выработки политики и стратегии, оценки своих инвестиций и точной разработки своих программ. Объем информации, которая требуется для правительственных органов, возрос в значительной степени, и также ускорила срочность этой информации, вызванная комплексным характером и взаимосвязями в мире. Некоторые из критических областей связаны с признанием обществом насущных социальных проблем, несостоятельностью финансовых систем, глобализацией, терроризмом, изменением климата и новым акцентом на устойчивое развитие. Деловые круги и широкая общественность также нуждаются в надежной информации для того, чтобы принимать основанные на ней решения, проводить исследования и делать оценки, а также понимать происходящие изменения.

Комплексные проблемы обусловлены множеством причин и требуют различных подходов к представлению данных. Однако различные представления нуждаются в согласованности для того, чтобы избежать путаницы и поддержать принятие эффективных решений, основанных на информации. АБС рассматривало вопросы сво-

его будущего и новых возможностей, которые ему потребуются, с тем, чтобы решить новые задачи и соответственно построить свою работу (в частности, какие действия необходимо предпринять сейчас). Моя точка зрения заключается в том, что эффективной национальной статистической организации будущего, такой, как Австралийское бюро статистики, будет необходимо не только продолжать обеспечивать стабильную, пользующуюся доверием, регулярную и ключевую национальную и международную статистическую информацию. Кроме того, нам будет необходимо быстро комбинировать данные, получаемые из официальных статистических источников и широкого круга других источников для того, чтобы разрабатывать последовательную информацию по текущим и новым возникающим вопросам, представляющим национальный интерес. Мы должны быстрее реагировать и быть более подвижными в сборе новой информации для того, чтобы заполнять имеющиеся пробелы и отвечать на новые вопросы. Мы должны максимизировать информацию, используя инновационные статистические методы и инструменты, позволяющие иметь большие объемы данных, которые можно будет хранить, комбинировать и анализировать, тем самым обеспечивая их безопасность и соблюдение требований конфиденциальности.

В моем мире статистики я уже вижу революцию в скорости получения новых информационных продуктов и услуг и путей комбинирования и использования информации. Альтернативные источники информации, которые можно обновлять в реальном времени, имеются в наличии с меньшими ограничениями и большей детализацией, нежели информация, традиционно предоставляемая АБС. В дополнение к данным, разрабатываемым государственными и частными организациями, сейчас появляется все большее количество данных, которые можно получить, используя компьютеры, различные персональные устройства, сенсорные системы и инструменты. Примерами здесь могут быть сканеры, используемые в розничной торговле, научное оборудование, системы обработки изображений, транспортные системы, телекоммуникационные сети и даже данные, собираемые о людях, использующих Интернет. В результате этого происходит размывание границ официальной статистики, которая является основным продуктом АБС.

Например, в Австралии наблюдается все больший рост количества организаций, довольно часто публикующих экономические показатели, которые широко используются в дополнение к официально разрабатываемым показателям или ранее, чем они публикуются. Я привел в качестве примера опыт АБС для того, чтобы показать, что хотя имеются большие возможности для такой организации получить выгоду от бурного роста количества данных, нам необходимо определить свою позицию сейчас для того, чтобы выжить и успешно работать в этом развивающемся мире. Пока мы этого не сделаем, мы не можем быть уверены в том,

что АБС останется в предстоящие десятилетия сердцевиной официальной информации для нашего общества. Хорошей аналогией здесь является то, что сейчас происходит со многими газетами - старой модели газетного бизнеса бросает вызов цифровая революция, и многие газеты исчезают.

**Решения: совместное развитие и индустриализация.** За редким исключением, каждое национальное статистическое ведомство самостоятельно строит все свои статистические системы.

Так как же обрести способность наиболее эффективно отвечать на все вызовы, которые предлагает нам эта революция цифровой информации? И должны ли мы снова браться за решение этих проблем в основном каждый самостоятельно? Или все же существует реальная возможность совместно работать над модернизацией наших статистических систем?

Сегодня международное сотрудничество в области информационных и коммуникационных технологий (ИКТ), а также автоматизации процессов и статистических методов в основном осуществляется среди специалистов. При этом обычно ставятся цели поделиться опытом и, время от времени, провести исследования и продемонстрировать инновации. Доказано, однако, что взаимодействие и общее восприятие результатов в этой области весьма затруднительно. Хотя специалисты и могут приходить к соглашениям, у них нет достаточной власти, чтобы инициировать существенные изменения в своих организациях.

Если мы сможем распространить это развитие на многие национальные статистические ведомства, то это позволит добиться значительной экономии: ключевым условием здесь является соглашение о стандартах и передача комплексных решений, а не «конъюнктурный» обмен.

Наши процессы могут быть индустриализованы. Это будет достигнуто не теми способами, которыми Генри Форд изменил мир, а более тонкими методами. В распоряжении г-на Форда не было других действительно мощных средств, кроме человеческих ресурсов. Его автоматизированная фабрика состояла из длинных цепочек людей, делающих простые шаги в общем процессе.

Сегодня имеются гораздо более изощренные инструменты, и индустриализация означает цепь очень мощных инструментов и специально написанных программ, которые обеспечивают промышленное производство статистического продукта. Человеческий труд здесь ограничен и сосредоточен в сфере проектирования, где он, используя стандартизованные средства и методологии, создает «статистическую фабрику» для заданного продукта. И именно в сфере проектных разработок накапливаются человеческие знания.

Краеугольным камнем развивающегося нового подхода к модернизации статистической инфраструктуры служит ТСМБП - типовая статистическая модель биз-

нес-процесса (GSBPM, the Generic Statistical Business Process Model), находящаяся в настоящее время под управлением ЕЭК ООН, контролирующей, в том числе, и вносимые в нее изменения, а также ТСИМ - типовая статистическая информационная модель (GSIM, the Generic Statistical Information Model), разрабатываемая Статистической сетью ЕЭК ООН (Statistical Network) и проектом CORE Европейской статистической системы (ESS). В практической области стандартизованные методы и технологии позволят нам добиться большей рационализации. Если мы сможем это осознать, то мы сможем снизить разнородность методов, решений в области информационных технологий, бизнес-концепций и информационных концепций, создавая тем самым необходимые условия для гармонизации и стандартизации нашей отрасли.

Это позволит нам превратить то, что сегодня преимущественно напоминает кустарную промышленность, в совокупность хорошо оснащенных индустриальных процессов.

**Общность: процессы, архитектура, стандарты, технология.** В распространении инструментов наблюдается некоторый прогресс: Blaise (система обработки данных обследований), OECD.stat (база данных), PC-AXIS (система разработки статистических таблиц), инструменты SDMX<sup>1</sup>, но этого недостаточно. Остаются значительные барьеры на пути заимствования из-за адаптационных требований. Нам необходимо принять общие:

- процессы (типовая статистическая модель бизнес-процесса - ТСМБП);
- архитектуру (типовая статистическая информационная модель - ТСИМ);
- стандарты (DDI<sup>2</sup>, SDMX).

Это создаст базовую общность, которая позволит нам создавать и обмениваться приложениями. Сегодня существует ряд форумов, на которых эти принципы обсуждались: MSIS<sup>3</sup>, METIS, SAB, HLG-BAS и др.

**Австралийское бюро статистики заинтересовано в переходе от теории к практике.** АБС, вместе с единомышленниками из других национальных статистических ведомств, крайне заинтересовано в проработке идеи сотрудничества, направленного на модернизацию нашей статистической инфраструктуры в течение следующих нескольких лет. Первые шаги:

- Новая Зеландия - инновации в области распространения данных;
- Канада - сбор данных через Интернет;
- Норвегия - редактирование данных;
- Швеция - конфиденциальность;
- Австралия - ввод в действие Структуры управления общими метаданными/информацией.

Эксперименты с использованием DDI (инициатива по документированию данных) и SDMX (обмен статистическими данными и метаданными) в качестве форматов данных/метаданных. Недостаточное использова-

<sup>1</sup> Международный стандарт обмена статистическими данными и метаданными.

<sup>2</sup> Международный стандарт распространения данных.

<sup>3</sup> Проект Евростата и ОЭСР по разработке и внедрению единых стандартов данных.

ние метаданных является главным тормозом для нахождения нужных данных, их согласования и оценки качества, а также препятствует экономической интеграции данных из различных источников.

**ЭСКАТО - осознание общности проблем и интересов.** С учетом финансовых ограничений, насыщенных рабочих программ и устаревающей инфраструктуры, а также вызовов стремительно изменяющейся среды национальным статистическим ведомствам для достижения успеха необходимо работать вместе. В связи с этим им нужно сделать две важные вещи для более оперативного реагирования на изменения, а именно: повысить универсальность статистической инфраструктуры, стандартов и классификаций по отношению к современным проблемам и выработать подходы к более быстрому развитию, обмену и увязке технологических инфраструктур. Таким образом, ключевой вопрос звучит так: разделяем ли мы общие проблемы? И если это так, то заинтересованы ли Вы в сотрудничестве?

**Важность согласованных ключевых стандартов.** Однако если мы намерены сотрудничать продуктивно, то очень важно достичь гораздо большей степени взаимной стандартизации, чем та, которую мы наблюдаем сегодня. А это значит, что нам придется забрать у наших работников что-то из того, что они ценят очень высоко, - часть их свободы в принятии решений о том, как им делать их индивидуальную работу.

Для нашего будущего принципиально важно сократить разнородность и сложность там, где они не могут быть оправданы. Более того, сами инструменты и подходы, которые хорошо работали в прошлом, возможно, уже не будут так хорошо служить нам в будущем. Основные усилия в нашем технологическом развитии были сконцентрированы на поддержке наших индивидуальных расчетных моделей. И хотя мы и участвуем в международных структурах, в обмене информацией о подходах и иногда делимся технологией, большинство наших систем специально разработаны для наших собственных целей. Наши ограниченные ресурсы для инвестирования в технологию также растрачиваются в циклах развития и попытках модернизации.

Если мы поднимемся над нашим собственным миром статистики, то сможем увидеть, что многие другие отрасли и источники данных используют различные метаданные и стандарты данных, однако в рамках отдельных областей интересов они чаще унифицированы в мировом масштабе. Например, чтобы обеспечивать проведение банковских операций для потребителей по всему миру, существуют стандарты, используемые многими банками. Большинство производителей делового оборудования, такого, как кассовые аппараты и сканеры, действующих на мировом рынке, используют единые стандарты, такие, как штриховые коды и RFID. Это справедливо также для производителей оборудования в области телекоммуникаций, путешествий и другого оборудования, используемого в обществе (как, например, светофоры) и в космических средствах

передачи информации (включая спутниковую информацию, карты Google и данные GPS).

Полезно было бы использовать коллективный подход к определению лучших путей, которыми НСО могут интегрировать эти данные в информационные банки своих стран. Также полезной была бы дискуссия о том, как отбирать данные, представляющие информационную ценность в будущем и подлежащие хранению.

Использование стандартов и структур будет оставаться важным для обеспечения согласованности наших статистик. Однако нам также нужен способ быстрого внедрения новых разделов и представлений данных, не требующий перестройки всей базы. Наряду с поддержкой международных стандартов, НСО приходится работать в различных условиях с несколько различными областями особого внимания. Пользователи наших данных должны иметь возможность адаптировать или дополнять стандартные классификации, чтобы они лучше соответствовали задачам их анализа, ситуации, среде или позволяли сравнивать эти данные с информацией, созданной ранее другими. Классификации должны стать более динамичными, поддерживаться автоматически средствами быстрого преобразования данных, таких, как автоматические кодировщики.

Автоматизированные подходы важны, потому что они позволяют получать множество представлений данных без существенного увеличения стоимости и затрат времени. Сюда включается быстрое перекодирование существующих массивов данных, кодирование больших объемов некодированных данных, множественное кодирование данных или перекодирование данных по запросам. В АБС предприняты исследования некоторых новых многообещающих методик, использующих машинное обучение. Поскольку методы кодирования имеют широкое применение и разнообразные функции, АБС предпочитает сотрудничать в области развития новых средств кодирования, основанных на этих методах. Использование общих устройств во множестве организаций также могло бы обеспечить согласованность кодирования и более высокое качество данных.

В этом докладе я в очень общем виде изложил некоторые возможности и проблемы, которые, как я уверен, встают перед всеми национальными статистическими ведомствами, по мере того как мы пытаемся определить наше место в быстро развивающейся информационной среде, сформировавшейся вокруг нас за последнее десятилетие. Я надеюсь, что дискуссии, которые нам предстоят в будущем, будут побуждать всех нас к осознанию некоторых вопросов, поднятых мною в этом докладе, и положат начало поискам возможности разделить расходы по адаптации наших собственных статистических инфраструктур, процессов и систем к условиям меняющегося мира большой информации. Или можно сказать иначе: в цифровом мире нам, официальным статистикам, надо ориентировать себя не просто на выживание, но и на процветание.