



**ЭКОНОМИЧЕСКИЙ
И СОЦИАЛЬНЫЙ СОВЕТ**

Distr.
GENERAL

CES/2005/19
24 March 2005

RUSSIAN
Original: ENGLISH

**СТАТИСТИЧЕСКАЯ КОМИССИЯ и
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ**

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Пятьдесят третья пленарная сессия
(Женева, 13-15 июня 2005 года)

**eSTATISTIK.core: ЭЛЕКТРОННАЯ СИСТЕМА СБОРА ДАННЫХ
ИЗ СИСТЕМ ERP**

Вспомогательный документ, представленный Федеральным
статистическим управлением Германии*

ВВЕДЕНИЕ

1. В сотрудничестве с ведущими производителями программного обеспечения и респондентами германские статистические органы разработали единую систему ввода первичных данных в режиме он-лайн через интернет - eSTATISTIK.core. Главная цель этого проекта состоит в разработке эффективных методов подготовки и представления статистических данных из систем планирования ресурсов предприятия (системы ERP). Она опирается на системный подход, направленный на снижение нагрузки на компании по представлению отчетности и одновременно с этим представляет собой значительный шаг вперед в деле улучшения процедур сбора данных. Система eSTATISTIK.core также интегрирована в национальную инициативу в области электронного правительства "Германия онлайн" и была представлена для целей разработки общенациональных стандартов.

* Автор: Михаэль Шефер.

2. Система eSTATISTIK.core состоит из нескольких инфраструктурных и программных компонентов: единого пункта доставки в сети расположенного на центральном сервере сбора данных (CORE.server), стандартизованных и общих для всех обследований форматов документов на основе XML¹, электронных описаний обследований и бесплатного программного обеспечения, а именно библиотеки программного обеспечения CORE.connect и автономного приложения CORE.reporter. Система eSTATISTIK.core обеспечивает отправку первичных данных с центрального сервера сбора данных статистическому органу, указанному в качестве адресата такой информации, и доступность таких данных для процесса проведения обследования в надлежащем формате. Система eSTATISTIK.core также обеспечивает контроль достоверности входящих данных и генерирует протоколы на основе XML, пригодные для автоматизированной обработки данных респондентами.

3. Оценочная версия eSTATISTIK.core используется с начала 2004 года. В первом квартале 2005 года будет опубликована производственная версия, предназначенная для регулярного использования в рамках отдельных видов обследований.

МОТИВАЦИЯ И НАЧАЛО РАБОТЫ

4. Сокращение нагрузки на респондентов - одна из первоочередных задач политической повестки дня. К сожалению, эта многогранная задача не имеет простых решений. С точки зрения процесса статистического производства возможными путями ее решения в некоторых случаях являются стандартизация и упрощение вопросников и терминологии, уменьшение количества вопросов и периодичности обследований, а также использование уже имеющихся данных. В процессе фактического сбора данных данная задача сводится к оказанию помощи респондентам в представлении информации. В этом отношении во многих статистических органах были разработаны основанные на использовании Интернета решения, которые позволяют упростить доступ к заполнению вопросника, а также отправку файлов с первичной информацией. Такие решения нередко предусматривают предоставление ручного интерфейса для отправки данных, который является достаточным для физических лиц и малых предприятий при использовании общих, имеющих разумный размер вопросников, и отправке небольших объемов данных.

5. В рамках обследований предприятий от относительно небольшого числа источников поступают весьма значительные объемы первичных данных. Как правило, такими источниками являются либо крупные компании, представляющие данные от своего собственного имени, либо поставщики услуг, представляющие данные в качестве третьей стороны от имени клиентов. В целом поставщики услуг представляют значительное количество малых и средних предприятий.

6. Многие крупные компании и поставщики услуг пользуются для управления корпоративными данными сложными системами ERP, в основном нескольких основных разработчиков, либо другими ИТ-системами. В прошлом функции разработки статистической информации реализовывались, хотя и не во всех случаях, разработчиками программного обеспечения ERP или самими предприятиями. Однако многие из этих систем обладают рядом недостатков: они предназначены для конкретных видов обследований и требуют доработки даже в случае небольших изменений в обследовании; разнообразие интерфейсов данных обследования затрудняет многократное использование программного обеспечения, причем данные интерфейсы практически исключают возможность автоматизации процедуры представления отчетности. Кроме того, федеральная структура Статистической системы Германии часто обуславливает необходимость представления предприятиями данных в рамках одного и того же обследования одновременно нескольким статистическим органам. Данное положение может быть улучшено по целому ряду направлений, и в этой связи промышленность Германии, представленная ее корпоративными органами, предложила разработать современные, унифицированные и основанные на использовании Интернета процедуры представления отчетности, которые обеспечат автоматизированную разработку и передачу статистической отчетности из систем ERP.

7. Одновременно статистические органы Германии уже предприняли первые шаги в направлении модернизации процессов сбора данных, которые включали в себя разработку DatML/RAW, представляющего собой типовой документ в формате XML для сбора первичных данных во всех видах обследований; а также концепции для его использования в качестве унифицированного интерфейса первичных данных, на основе которого могут быть созданы автоматизированные общие серверные процедуры сбора данных. Предложение промышленности Германии послужило благоприятной возможностью для начала обсуждения этих концепций с пользователями в целях их дальнейшего развития и доработки, а также внедрения в практику.

8. В начале 2003 года была создана рабочая группа с целью обсуждения возможных усовершенствований процедур сбора данных с участием предприятий и организаций, а также разработки экспериментальной версии для отдельных видов статистики оплаты труда. В состав этой рабочей группы входят представители статистических органов промышленности Германии, и она функционирует главным образом в формате AWV - *Arbeitsgemeinschaft für wirtschaftliche Verwaltung* (Рабочей группы по совершенствованию методов управления). Деятельность РГСМУ направлена на совершенствование взаимоотношений между хозяйственными субъектами и государственными органами. В качестве членов-респондентов в ее состав входят компании-производители услуг и

в общей сложности 70 таких крупных и влиятельных разработчиков программного обеспечения. В качестве примера можно назвать такие компании, как "Люфтганза", Datev, SAP, Oracle и UBM.

9. eSTATISTIK.coге не является каким-то обособленным проектом. В рамках общих усилий по стандартизации на центральном уровне этот проект интегрирован в общую стратегическую программу статистических органов Германии, называемую "*Генеральным планом*", которая преследует две основные цели: повышение эффективности статистической системы Германии и снижение нагрузки на респондентов. Генеральный план направлен на расширение сотрудничества и обеспечение координации мероприятий по модернизации, многие из которых связаны с термином eSTATISTIK. eSTATISTIK представляет собой стратегическую инициативу, которая направлена на перевод стратегических услуг в онлайн-режим и модернизацию систем сбора, разработки и распространения статистических данных. Все мероприятия, проводимые eSTATISTIK, опираются на тщательный анализ соответствующего процесса и применение стандартов в ходе его осуществления.

ЦЕЛИ

10. Хотя главная цель рабочей группы состояла в поиске путей снижения нагрузки на респондентов, вскоре выяснилось, что фундамент для любого решения должен закладываться в рамках статистической системы и что это должно предусматривать модернизацию внутренних процедур сбора данных. По этой причине цели, намеченные рабочей группой, учитывают потребности как респондентов, так и статистической системы.

11. Предлагаемое решение предусматривает внесение значительных усовершенствований в следующих областях: создание и проверка достоверности клиентами передаваемых файлов первичных данных в целом, а также стандартизация и унификация форматов в качестве необходимого предварительного условия для снижения расходов по внедрению и обслуживанию; электронная передача файлов первичных данных в статистические органы, а также подтверждений и протоколов - респондентам и отправителям данных; ориентированные на потребности заказчиков серверы.

12. Нагрузку по представлению стратегических данных необходимо уменьшить в максимально возможной степени. Это предусматривает устранение ручных операций за счет автоматизации процедур представления отчетности, снижение затрат за счет стандартизации и многократного использования программного обеспечения, а также улучшение инфраструктуры сбора данных.

13. Автоматизация и унификация процедур сбора данных позволит повысить эффективность статистической системы. Дополнительные выгоды для статистических органов будут заключаться в получении предварительно выверенных данных высокого качества и повышении своевременности и точности их материалов.

14. Ожидается, что сотрудничество между компаниями, экономическими организациями и статистическими органами обеспечит внедрение и использование новых процедур сбора данных разработчиками программного обеспечения и респондентами, а также будет содействовать укреплению репутации статистических органов как поставщиков современных и ориентированных на потребности пользователей услуг и информации.

ТРЕБОВАНИЯ И ОРГАНИЧЕНИЯ

15. Современные Интернет-технологии позволяют передавать данные затратноэффективным, оперативным и безопасным способом. Поэтому система eSTATISTIK.core опирается на использование Интернета. Предприятия-респонденты должны иметь доступ к Интернету, что на сегодняшний день является обычной практикой. Для передачи объемных документов лучше иметь широкополосный доступ.

16. Респонденты смогут пакетировать любое количество отчетов в любом сочетании обследования и адресата в единый файл и отправлять его в единый пункт сбора информации. Кроме того, респонденты также получают возможность запрашивать подтверждение и получать доступ к электронным протоколам и обрабатывать их для отслеживания деятельности по предоставлению отчетности и выявлять и устранять ошибки, возникающие в ходе этого процесса.

17. Во избежание роста, связанных с внедрением и обслуживанием, особенно в средне- и долгосрочной перспективе, следует свести к минимуму потребность адаптации к специфике конкретных обследований. Для решения этой задачи следует разрабатывать и использовать как можно больше общих программных компонентов.

18. Статистические органы представляют метаданные, управляющие eSTATISTIK.core. Они должны гарантировать точность и достоверность метаданных, а также их способность обеспечивать поддержку инфраструктуры сбора данных.

19. Особое значение придается надежности интерфейсов данных и метаданных, или форматам документов, а также их пригодности для процесса общей обработки результатов обследований различных типов. Для обеспечения совместимости необходимо

управление версиями, причем нейтральность платформы является обязательным условием с учетом разнообразия аппаратных и программных платформ, на которых работают системы ERP и системы статистического производства. Расширяемый язык разметки (XML) является современной технологией для разработки типов документов, удовлетворяющих этим требованиям. В наличии имеется большое количество (зачастую бесплатного) программного обеспечения для создания приложений на основе XML, которые нередко пишутся на языке Java, обеспечивающем особенно эффективную поддержку обработки XML и возможность использования на многих платформах. Следовательно, все документы, которыми обмениваются респонденты и статистические органы, будут являться XML-документами.

20. Именно статистические органы должны проделать основной объем работы по совершенствованию процедур и услуг в области сбора данных. Вместе с тем любые их разработки будут ограничиваться лишь той частью процесса, которая предшествует интеграции искомым элементов в конкретные программные и аппаратные средства респондента; однако они способны облегчить решение этой задачи.

21. Система eSTATISTIK.core способна обеспечить значительную экономию средств. Возможно, эта система в большинстве случаев будет экономичной, однако для каждого отдельного респондента это будет зависеть от многих факторов, выходящих за рамки данного решения, и едва ли поддается прогнозу. Однако проект eSTATISTIK.core занимает центральное место в модернизации статистической системы, в связи с чем важно оценить его влияние. В настоящее время проводится исследование, результаты которого позволят составить четкую картину воздействия системы eSTATISTIK.core на снижение нагрузки по представлению отчетности.

ОСНОВНЫЕ СОСТАВЛЯЮЩИЕ ДАННОГО РЕШЕНИЯ

22. Исходя из вышеупомянутых требований и ограничений участвующие в проекте партнеры пришли к выводу о том, что в системе eSTATISTIK.core основное внимание должно уделяться совершенствованию интерфейсов данных и программного обеспечения, а сама система должна обладать следующими основными характеристиками:

- общие типовые документы на основе XML для первичных данных, определения обследований, протоколы проверки достоверности и подтверждения; в отношении первичных данных будет использоваться типовой документ DatML/RAW; кроме того, будут разработаны типовые документы для определений обследований - DatML/SDF и подтверждений и протоколов - DatML/RES;

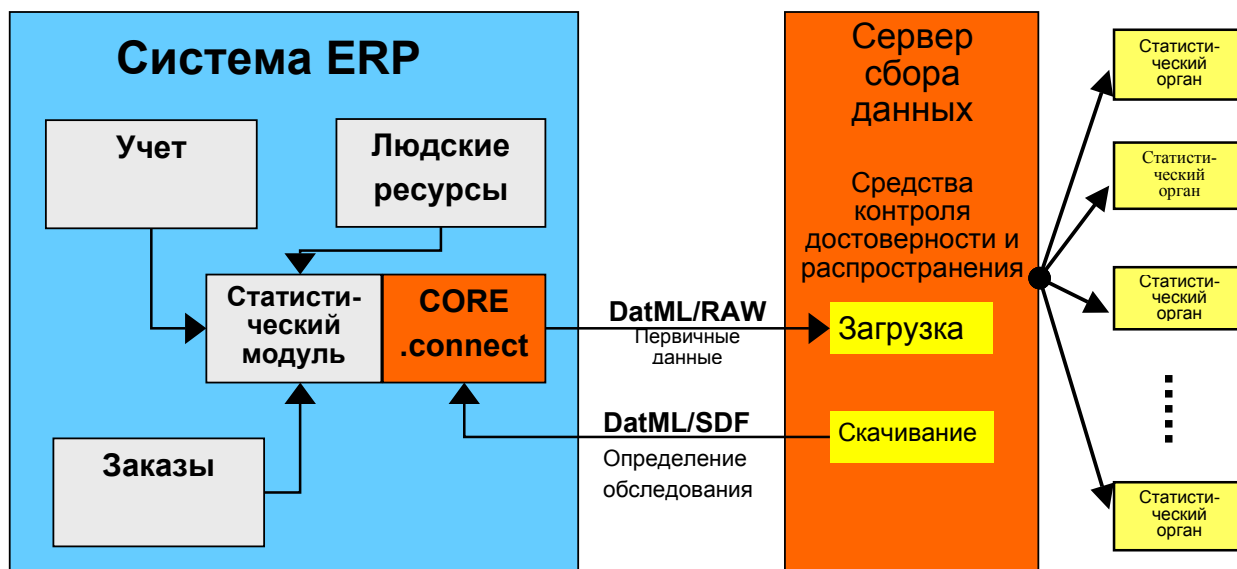
- библиотека программного обеспечения CORE.connect, реализованная с помощью различных общих языков программирования, для реализации клиентами следующих функций: проверка и загрузка наборов первичных данных, а также загрузка определений обследований и протоколов проверки;
- CORE.reporter - автономное приложение, разработанное на основе CORE.connect, с графическим интерфейсом пользователя и инструментами для управления отчетами и для преобразования данных из существующих источников, как, например, двумерные файлы, в типовые модели данных обследования;
- *так называемые статистические модули*, реализованные разработчиками систем ERP, которые обеспечат извлечение первичных данных из систем ERP, генерирование документа DatML/RAW и его загрузку с помощью CORE.connect на сервер сбора данных. Такие модули могут быть реализованы в виде общих модулей с использованием вместо жесткого кодирования связи между первичными данными и переменными определений обследований DatML/SDF и гибких механизмов преобразования данных;
- центральный Интернет-сервер для отправки файлов первичных данных на единый Интернет-адрес и для загрузки определений обследований и протоколов;
- пригодность для работы в полностью автоматизированном режиме.

23. Первая версия eSTATISTIK.coge не поддерживает генерирование документов в формате DatML/RAW. Приложения создают документ DatML/RAW, направляют его CORE.connect для проверки и загружают на сервер сбора данных.

24. Помимо описанного выше чисто технического подхода, было также принято решение о согласовании статистической и корпоративной терминологии в целях достижения более высокой степени концептуального соответствия и сведения к минимуму необходимости адаптации корпоративных данных к статистическим концепциям.

АРХИТЕКТУРА eSTATISTIK.core

25. Обзор:



26. Вышеприведенная диаграмма иллюстрирует концепцию архитектуры eSTATISTIK.core: а) в системе ERP статистический модуль извлекает первичные данные из таких подсистем, как людские ресурсы и учет, и компилирует их в документ формата DatML/RAW; б) документ DatML/RAW проходит контроль на достоверность с использованием схемы XML и определения обследования DatML/SDF и при положительном результате с) загружается на центральный сервер сбора данных, на котором д) происходит его декомпиляция на отдельные файлы первичных данных, которые затем е) направляются в соответствующий статистический орган.

27. Обработка данных клиентом отчасти зависит от способа реализации статистического модуля. Как правило, статистический модуль обеспечивает загрузку определения обследования DatML/SDF текущего обследования, которое используется для извлечения необходимых первичных данных. В нынешней версии eSTATISTIK.core статистический модуль должен генерировать весь документ DatML/RAW, который затем передается CORE.connect для проверки на достоверность и загрузки на сервер сбора данных.

28. На сервере сбора данных входящие документы проходят последовательные этапы обработки, каждый из которых реализуется в отдельном модуле.

29. Модуль *Selector* осуществляет проверку типа документа. Документы допустимых типов, помимо DatML/RAW, обрабатываются "традиционно", т.е. загружаются в подсистему сбора данных, обрабатывающую документы не формата XML. Если документ является документом DatML/RAW, то он проходит дополнительную проверку на достоверность по схеме XML.
30. Модуль *Inspector* осуществляет проверку каждого файла с первичными данными в документе DatML/RAW по совпадающему определению обследования, загружаемому из документа DatML/SDF.
31. Модуль *RawFilter* компилирует документ DatML/RAW в один или более документов DatML/RAW; количество итоговых документов зависит прежде всего от количества файлов первичных данных, которые содержались в входящем документе, и конфигурации модуля RawFilter. К примеру, конфигурация модуля RawFilter может быть таковой, что каждый итоговый документ будет содержать один файл первичных данных или все файлы входящего конкретного статистического органа или обследования.
32. Модуль *Raw2Flat* преобразует документы DatML/RAW в простые файлы различных форматов (EBCDIC, ASCII, CSV). Этот модуль необходим, поскольку в большинстве обследований в качестве входных в производственном процессе используются форматы двумерного файла. При преобразовании документа DatML/RAW модуль Raw2Flat осуществляет загрузку файла с определением набора данных, который содержит описание структуры набора данных, а также инструкции по составлению набора данных на основе отдельного файла первичных данных.
33. Модуль *Forwarder* отвечает за передачу документов, созданных модулем RawFilter и/или Raw2Flat, в соответствующий статистический орган. В нем используется приложение FTP/SFTP или DVE, разработанное статистическими органами для стандартизованного обмена наборами данных.
34. Модуль *Protocol* создает протоколы подтверждения достоверности DatML/RES. Модули, участвующие в создании протокола проверки (в настоящее время: Selector и Inspector), используют протокол API² для хранения набора данных в специально отведенных для этого узлах файловой системы. Через определенные интервалы времени сервер модуля сканирует эти узлы файловой системы; если при сканировании выявляется узел, обозначенный как полный, на основе содержания этого узла создается документ DatML/RES, после чего узел удаляется. В будущем Protocol API сможет использовать вместо файловой системы базу данных для хранения временных данных протокола.

ТИПЫ ДОКУМЕНТОВ XML

35. DatML/RAW является общим, т.е. не зависимым от вида обследования типом документа для передаваемых файлов первичных данных. Впервые этот формат был выпущен в 2001 году и используется в более 10 обследованиях. Документ DatML/RAW содержит, по крайней мере, одно *сообщение*, содержащее, по крайней мере, один *отчет*. Его гибкая структура позволяет обрабатывать любое количество отчетов с первичными данными в любом сочетании обследования, отчетного периода, респондента и статистического органа, ведущего сбор данных. Внутри сообщения в отчетах могут использоваться общие метаданные. DatML/RAW обеспечивает хранение первичных данных, относящихся к конкретному обследованию, в общей структуре элементов, представляющих записи, переменные и группы переменных. Эта общая структура варьируется от обследования к обследованию по количеству и месту расположения этих элементов и по присвоенным им именам.

36. DatML/RES представляет собой тип документа для подтверждений и протоколов проверки достоверности. Протокол проверки содержит информацию о входном документе и описывает параметры проверки и результаты проверки на уровне документа, сообщения и отчета, включая тип и местоположение ошибки, и затрагиваемую переменную, если таковая имеется.

37. Документ DatML/SDF содержит ключевые метаданные, необходимые для использования eSTATISTIK.core применительно к конкретному обследованию. Он описывает, как определяется обследование, его такие общие характеристики, как отчетный период и периодичность, его модель данных, т.е. переменные и группы переменных, и их характеристик и зависимости величин, а также то, каким образом эта модель преобразуется в логическую структуру передаваемого файла первичных данных.

УПРАВЛЕНИЕ МЕТАДААННЫМИ

38. Для обеспечения эффективности системы eSTATISTIK.core необходим значительный объем метаданных, который возрастает с каждым новым обследованием и с каждым изменением, которое вносится в ресурсы существующих метаданных. Кроме того, недействительные метаданные могут поставить под удар всю систему. После ввода в эксплуатацию такая система не сможет работать на метаданных сомнительного качества. Поэтому в целях создания метаданных для eSTATISTIK.core используются разнообразные инструменты, а также строится новая система управления метаданными.

39. *STATSPEZ*³ - основанное на метаданных средство спецификации и создания программ-табуляторов. Один из ее компонентов - *data set designer*, который выдает определения наборов данных, включая преобразование переменных, управляющих преобразователем Raw2Flat.

40. Модуль *data edit designer* ("Редактор-PL") был разработан в первую очередь для спецификации процедур редактирования данных, но также представляет собой средство, с помощью которого создаются определения переменных.

41. Модуль *survey definition editor* ("Редактор-SDF") был разработан для создания документов с определением обследования DatML/SDF. В нем используются определения переменных из разработчика процедур редактирования данных.

42. В контексте eSTATISTIK.core основными ресурсами метаданных являются определения обследований и определения наборов данных (для преобразования документов DatML/RAW в двумерные файлы). Вместе с тем в контексте организации полного обследования возникает потребность в задействовании гораздо большего объема ресурсов метаданных. По этой причине в настоящее время создается новая система управления метаданными, которая призвана обеспечить управление метаданными в течение всего цикла обследования и в последующий период. В этой системе используются три концептуальных элемента для организации метаданных: статистика, обследование и ресурс. При появлении экземпляра одного из этих элементов ему присваивают уникальный идентификатор. Этот идентификатор управляет ресурсами в рамках обследования и включает в себя функцию управления версиями.

ПЕРСПЕКТИВЫ

43. Будущая версия библиотеки CORE.connect будет содержать программные интерфейсы, которые устроят необходимость решения вопросов форматов документов. Вместо создания документа DatML/RAW соответствующие программы смогут только передавать данные и, опираясь на CORE.connect, создавать приемлемый документ.

44. В дополнение к протоколу HTTPS система CORE.connect будет поддерживать OSCI⁴ - протокол, который поддерживает ключевые инфраструктуры государственного и частного секторов и который был стандартизован для приложений электронного правительства (eGovernment).

45. Протоколы проверки DatML/RES будут указывать местоположение ошибки в документе DatML/RAW с помощью XPath⁵ - языка создания семантических указателей местоположения в документах XML. С помощью XPath содержащие ошибку данные

смогут извлекаться непосредственно из документа DatML/RAW и, например, представляться для исправления вместе с информацией, взятой из определений переменных в соответствующем документе DatML/SDF.

46. В протоколах проверки DatML/RES будет указываться местоположение ошибки в документе DatML/RAW с помощью выражения XPath (XPath - язык для составления выражений, указывающих местоположение в документах XML). Ошибочные данные могут быть затем извлечены непосредственно из документа DatML/RAW и, например, представлены для исправления вместе с информацией, взятой из определения переменной в соответствующем определении обследования DatML/SDF.

47. Для обеспечения успеха eSTATISTIK.core статистические органы, AVW и фирмы, участвующие в пилотных проектах, а также разработчики программного обеспечения реализуют ряд мероприятий по маркетингу и рекламе.

ВЫВОДЫ

48. eSTATISTIK.core представляет собой новый подход к совершенствованию системы сбора данных от предприятий. Ее основными компонентами являются стандартизованные и основанные на XML интерфейсы первичных данных и объекты метаданных. Они позволяют создавать общие, автоматизированные процедуры представления отчетности и сбора данных, которые обладают значительным потенциалом с точки зрения снятия значительной части нагрузки с респондентов и в то же время повышение эффективности статистической системы, а также качества и своевременности ее материалов.

49. Вместе с тем одних лишь стандартизованных интерфейсов первичных данных и метаданных не достаточно для обеспечения успеха eSTATISTIK.core. Для статистических органов высокое качество метаданных и надежное управление метаданными являются необходимым условием продуктивной эксплуатации системы, которая также невозможна без участия и поддержки со стороны целевых пользователей.

* * * * *

¹ Расширяемый язык разметки; см. <http://www.w3.org/XML>.

² Прикладной программный интерфейс.

³ Statistische Tabellenspezifikation; см. <http://www.statspez.de>.

⁴ Онлайн-сервисный компьютерный интерфейс; см. <http://www.osci.de>.

⁵ XML Path Language; см. <http://www.w3.org/TR/xpath>.