

Software Engineering Forum



**Конверсия, определяемая правилами.  
О конвертации данных с помощью  
продуктов с открытым кодом**

Алексей Коренев. ЭПAM Системз  
Дмитрий Хусаинов. ЭПAM Системз

# О чем доклад?

- ✓ Что такое конверсия
- ✓ Где возникают задачи, связанные с конверсией
- ✓ Какие есть решения и связанные с ними проблемы
- ✓ Инструментарий преобразований, построенный на основе правил
- ✓ Демонстрация

# Что такое конверсия?

Конверсия = ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ

- ✓ Изменение семантического смысла информации
- ✓ Добавление новой смысловой нагрузки
- ✓ Реорганизация информации
- ✓ Изменение формата представления

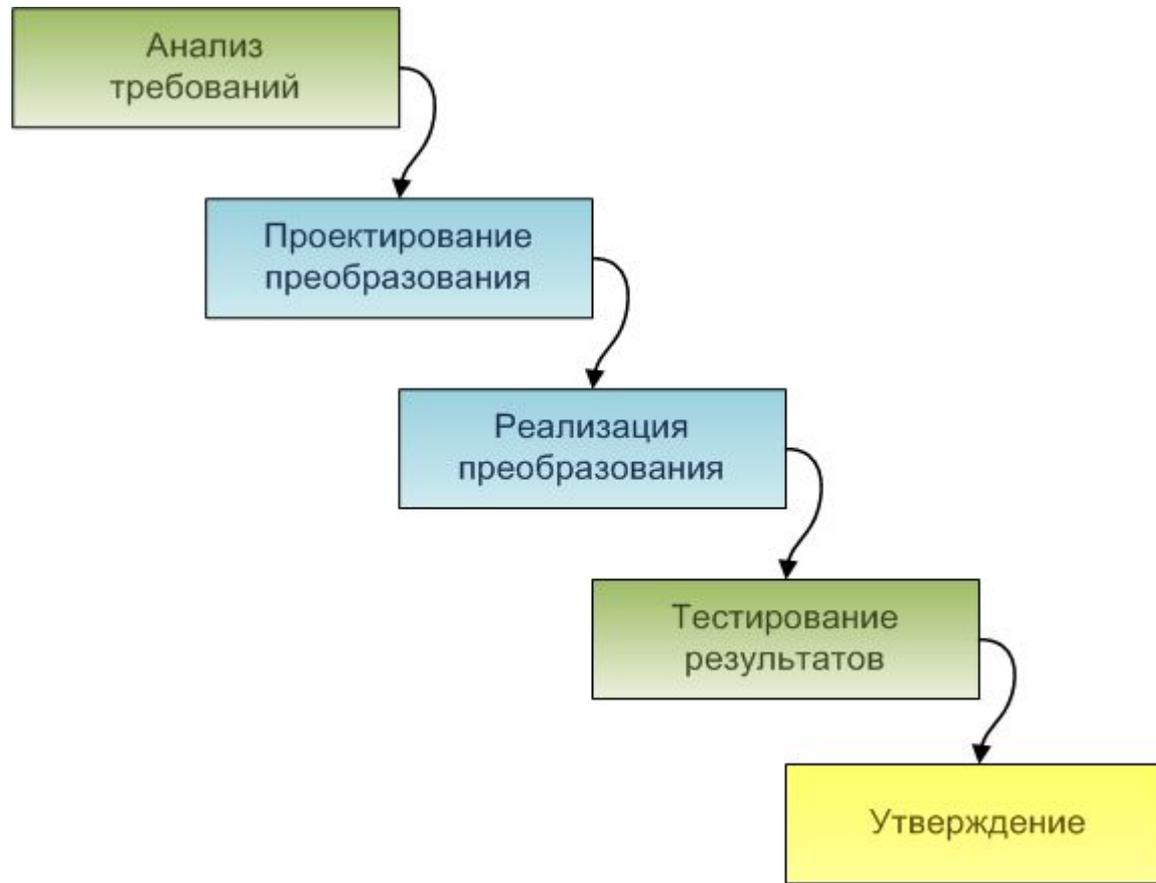
# Где и как возникают задачи преобразования информации?

- ✓ Финансовые компании
- ✓ Поддержка white-label брендов и продуктов
- ✓ Travel industry
- ✓ Поддержка web сайтов для мобильных устройств

# Существующие подходы и решения

- ✓ Компании, специализирующиеся на преобразовании информации (INNODATA, OpenWorld Data, Unisoft Datatech)
- ✓ Платные продукты (Omnimark, Microsoft BizTalk Mapper, EMC Documentum Content Transformation Services)
- ✓ Использование существующих технологий преобразования данных (XSLT, Smooks), не требующих платного лицензирования

# Как работают существующие подходы



## Чего не хватает существующим решениям?

Одного или нескольких из следующего списка:

- ✓ Предоставления гибкой возможности преобразовывать информацию бизнес-экспертам.
- ✓ Обеспечения непрерывной и параллельной работы всем участникам процесса
- ✓ Повышения надежности процесса преобразования при помощи автоматического тестирования, обеспечивающего соответствие требованиям по качеству и сервисным соглашениям
- ✓ Отсутствия зависимостей от сторонних коммерческих приложений и компонентов

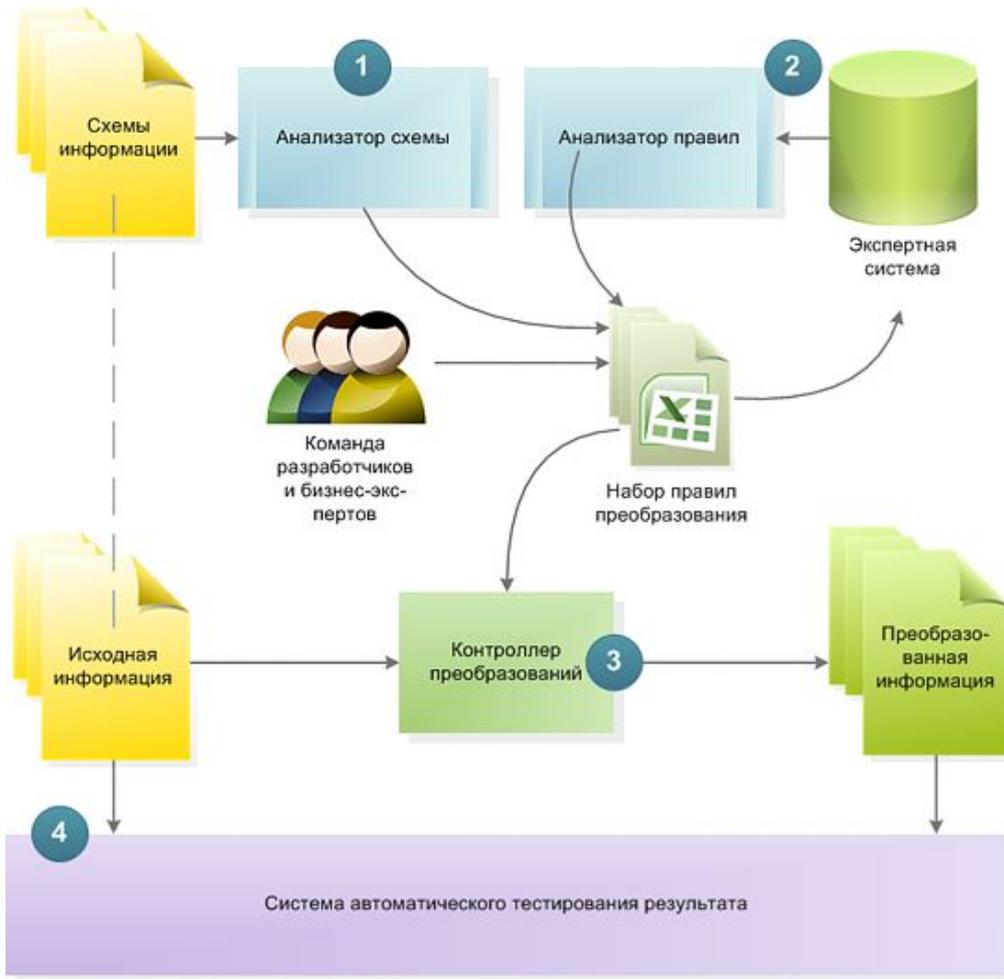
# Задачи, решаемые инструментарием преобразований

- ✓ Автоматизация процесса анализа информации, базируясь на основе исходной и результирующей схем определения информации
- ✓ Создание и поддержка экспертной системы правил преобразования информации
- ✓ Возможность создавать и модифицировать правила преобразования с помощью интуитивно-понятного пользовательского интерфейса
- ✓ Поддержка автоматического тестирования преобразованной информации

# Как работает наш подход



# Из чего состоит инструментарий?



- 1. Анализатор схемы** автоматизирует процесс анализа информации, определяя набор потенциальных преобразований.
- 2. Экспертная система и анализатор правил** являются базой знаний, используемой для принятия решений и выводов во время преобразования.
- 3. Контроллер преобразований** представляет собой технологическое ядро системы, обеспечивающее процесс преобразования.
- 4. Система автоматического тестирования результата** проверяет согласованность, целостность и семантический смысл информации, гарантируя правильность преобразования.

# Используемые технологии



**А также:**

**Apache  
Velocity**

**TestNG**

**JAXB**

**JExcelAPI**

**Apache Maven**

**Log4J**

**XMLUnit**

**AspectJ**

# Анализатор схемы

Позволяет определить набор потенциальных преобразований, анализируя модели исходной и результирующей схем определения информации. Механизм определения ограничений, накладываемых схемами, имеет модульно-расширяемую структуру.



# Правила преобразования – основные компоненты экспертной системы



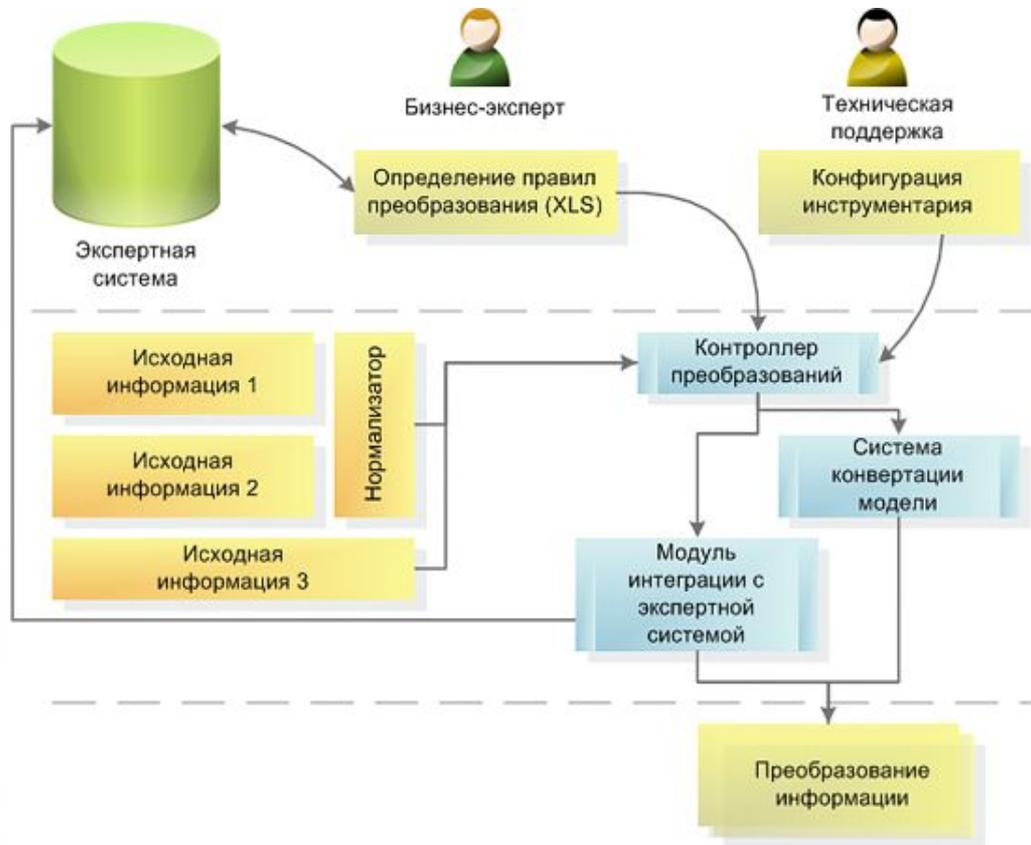
Исходное имя	Условие	Новое имя	Обернуть	Добавить атрибуты
addressbook		persons		
employee		person		
name		name		
tel	\$type == "mobile"	mobile-telephone	contact	
tel	\$type == "home"	home-telephone	contact	
addr		address	contact	\$_primary = "y"

Инструментарий предоставляет возможность бизнес - эксперту определять правила преобразования в виде электронной таблицы (Excel).

Правило состоит из:

1. Условия или условий, представляющих собой ограничения, накладываемые на состояние элементов исходной модели информации .
2. Параметризованных команд преобразования, т.е. модулей, определяющих логику преобразований элементов модели, структура которых может быть расширена в случае необходимости.

# Контроллер преобразований



- 1. Нормализирует** исходную информацию в виде правильно сформированной модели
- 2. Исполняет** преобразование, применяя правила, определенные в Excel спецификации к модели
- 3. Конвертирует** преобразованную модель в результирующий формат информации

# Стратегия тестирования

Стратегия автоматического тестирования фокусируется на верификации и контроле

- ✓ Различных численных характеристик результирующей информации (числа файлов, числа связей и зависимостей)
- ✓ Согласованности информации
- ✓ Индивидуальных правил
- ✓ Правильности реорганизации информации
- ✓ Корректной последовательности представления результирующей информации, включая семантику

# Когда преобразование информации завершено?

Преобразование считается законченным, когда:

- ✓ Полностью определена спецификация преобразования (Excel правила)
- ✓ Не было ошибок во время выполнения
- ✓ Все автоматические тесты были настроены для покрытия заданных требований
- ✓ Созданы тестовые сценарии, базирующиеся на правилах
- ✓ Все тесты были запущены и подготовлен соответствующий отчет
- ✓ Все противоречивости, найденные при помощи инструмента автоматического тестирования преобразования были прояснены

# Демонстрация