

# Миграция данных при модернизации ИТ-среды предприятия

“Как Богу удалось в шесть дней создать мир?” Ответ: “Ему не приходилось иметь дело со старыми инсталляциями”. Эта старая ИТ-шутка напоминает ИТ-экспертам о надежности их мечты начать работу над формированием ИТ-ландшафта своего предприятия с “белого листа”, “с нуля”, с самого начала. Там, где сегодня вводятся новые информационные системы, неизбежно предстоит иметь дело с интеграцией старых функций и данных для дальнейшего их использования. Сложность этой “само собой разумеющейся” задачи очень часто недооценивается, объем кроющихся затрат становится очевидным только на поздних стадиях проекта, когда бюджет уже на исходе и сроки крепко поджимают. В статье обобщается многолетний опыт, приобретенный сотрудниками фирмы ECS GmbH в различных проектах, связанных с модернизацией ИТ-среды предприятий.

## Обзор проблем

Первый шаг к успеху в процессе переноса (миграции) данных из устаревших систем в новую ИТ-среду состоит в том, что учет затрат на “ремонт” данных и оценка миграционного риска производится уже на стадии планирования проекта. В случаях, когда миграция проводится в виде краткосрочной операции накануне сдачи новой системы, она способна попросту “взорвать” все границы – как бюджетные, так и временные.

Миграцию данных рекомендуется планировать и проводить параллельно с введением новой системы. Это, как правило, означает, что начать придется с полного обновления переносимых данных, так как использовать их в том виде, в ко-

тором они представлены в старой системе, практически невозможно.

Причина не только в том, что организация базы данных в старой системе отличается от таковой в новой, и поэтому требуется реорганизация в структуре классов, отношений и атрибутов. Существует ряд и других “недугов”, некоторые из которых перечислены ниже:

- ▶ описание идентичных обстоятельств десятком различных способов;
- ▶ различные допустимые значения для одних и тех же атрибутов в различных системах;
- ▶ неполнота данных старой системы с точки зрения новой;
- ▶ различная степень обработки документов и процессов, хранящихся в различных источниках, что дополняется наличием “мертвых душ”, то есть данных в БД, не обрабатывавшихся годами;
- ▶ работа с различными форматами данных одной и той же прикладной программы, используемой в различных филиалах предприятия. Эти различия возникают на основе недокументированных отклонений от когда-то определенной, но устаревшей схемы.

При миграции старых данных необходимо быть готовыми практически ко всему. Особенно это касается метаданных – по неопытности здесь часто исходят из того, что в описанных документах нет отклонений от правил форматирования, индексирования и наименования файлов. Но оригиналы (документы в Word-формате, САД-чертежи со стандартными текстовыми полями и т.д.) терпеливы в меньшей степени, чем бумага. С другой стороны, метаданные, которые удается извлечь из старых документов (например, из переписки, заметок-на-

поминаний, технических чертежей), часто представляют собой ценный актив для предприятия.

Таким образом, систематическое извлечение и очистка старых данных, включая проверку на достоверность различных источников, имеют смысл по двум причинам: во-первых, для обеспечения максимально безошибочной обработки данных в новой системе и, во-вторых, для вскрытия до сих пор неиспользованного информационного потенциала.

Большую помощь в этой ситуации окажет систематическая процедура вскрытия и протоколирования ошибок в данных, позволяющая позднее практически безошибочно произвести очистку и необходимые изменения в их структуре.

## Рекомендация первая: промежуточный шаг

Наиболее важными критериями качества миграции и “ремонта” данных являются, наряду с их безошибочностью, также их воспроизводимость и изменимость. Предпосылкой для достижения этих целей является процесс документирования с точно определенными шагами и промежуточными результатами, которые можно подвергнуть перепроверке.

Организацию процесса миграции можно наглядно сравнить со строительством моста, который должен обеспечить переход от старой к новой модели данных, ассоциируемых с противоположными берегами одной реки. В то время как о ситуации на противоположном берегу (новой модели данных и их форматах) имеется достаточно четкое представление, “почва” на исходном берегу имеет неустойчивую и не вполне ясную структуру.

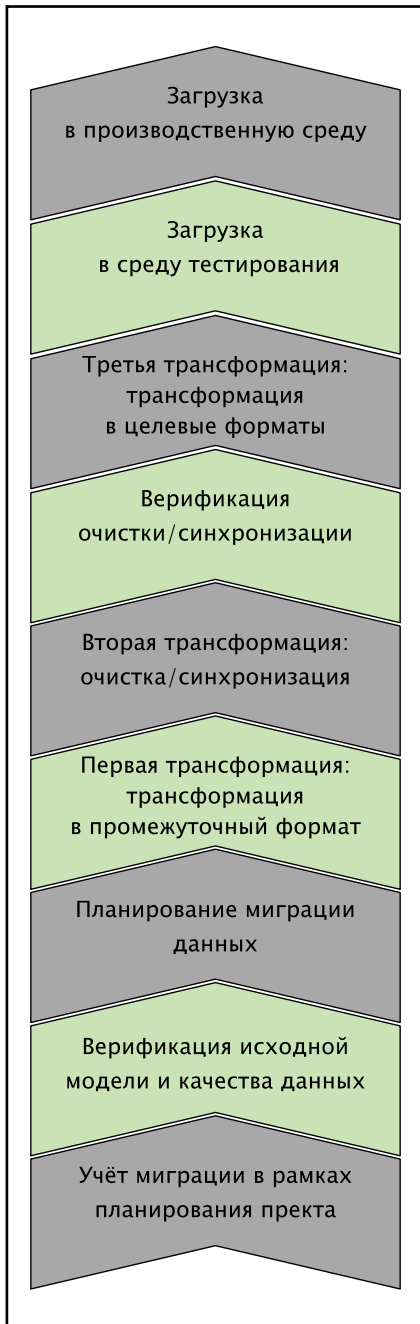


Рис. 1. Основные этапы миграции старых данных

Для понимания имеющейся модели данных требуются инвестиции в ее реконструкцию. Опросы пользователей и программистов могут привести к успеху только после многократного повторения цикла **допущение – опрос – тест – новое допущение ...**

Непосредственно переводить старые данные в формат новых не рекомендуется, так как при этом трудно оценить затраты, необходимые на очистку. К тому же при таком подходе может быть потеряна потенциальная возможность унификации процессов и инстру-

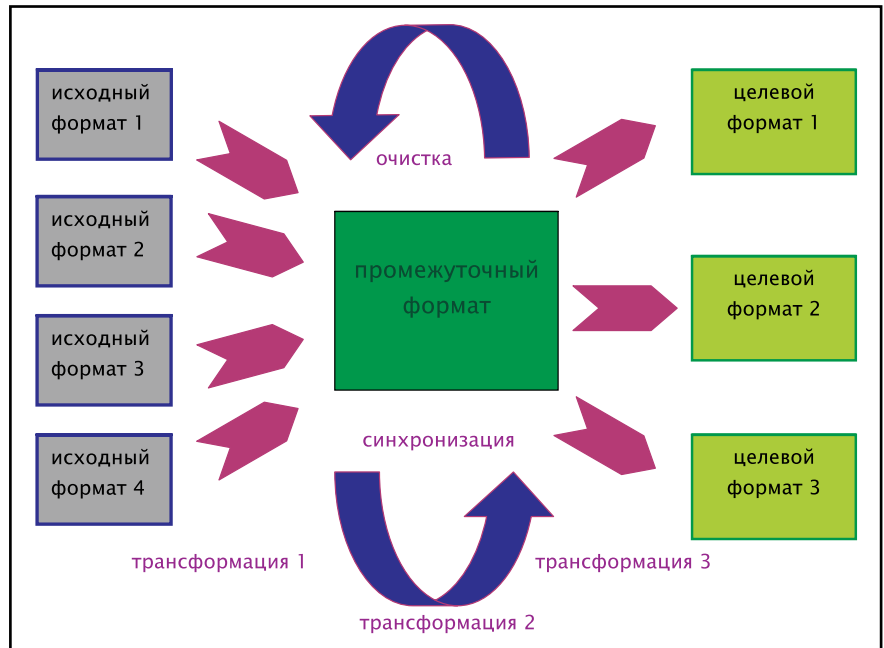


Рис. 2. Нейтральный промежуточный формат – базис для применения стандартизированных процедур и инструментов к очистке и синхронизации старых данных

ментов. Разумнее использовать промежуточный формат (используя аналогию с мостом, – промежуточную опору), опираясь на который, будет легче перейти к целевому формату, шаг за шагом применяя стандартизированные процессы и инструменты для очистки и реорганизации данных (рис. 1 и 2).

### Рекомендация вторая: целенаправленная автоматизация

Следующее требование заключается в необходимости автоматизировать обработку больших объемов информации за достаточно короткое время. Обработать большие объемы данных вручную, например, используя команду **найти-заменить**, означает подвергнуться риску нереконструируемых ошибок. К тому же очевидно, что скорость таких процедур и их функциональность в части распознавания и замены текстовых шаблонов явно неудовлетворительны.

Альтернативой является последовательная автоматизация работы с данными, то есть использование программируемых, параметризуемых и конфигурируемых инструментов. Достаточно эффективными показали себя скриптовые языки, содержащие все известные функции обработки текстов и объемный

интерфейс ко всем современным операционным системам. В соответствии с опытом авторов статьи, на сегодня это Perl и Python, являющиеся полноценными языками программирования с усложненными типами данных (например, string, список, словарь), а также с функциями для распознавания и замены образцов текста.

### Рекомендация третья: использование языка Python

Язык Python характеризуется доступностью, мобильностью, надежной обработкой текстов, интегрируемостью с языками C/C++. Этот язык легок в освоении, доступен для Windows, различных производных от UNIX и Linux, для систем, базирующихся на BSD-UNIX, IBM, а также для встраиваемых систем QNX и Vx Works. К этому перечню следует добавить поддержку платформ Apple и Handheld с Palm OS, Eros или Windows CE. Однотипные интерфейсы к системным функциям позволяют использовать идентичные Python-программы, например, как в среде Unix, так и Windows. Python можно интегрировать с приложениями и библиотеками, написанными на C/C++ или Java. Вызов функций (методов) возможен в обоих направлениях.

Качественному проектированию программного обеспечения способствует четкая концепция языка: логичный синтаксис, разделение кода на семантические и синтаксические единицы, образующие ступенчатую структуру, комбинация простых семантических понятий, а также мощные стандартные типы данных. Все это позволяет в короткое время создавать сжатые по форме, легко читаемые программы для решения относительно сложных задач.

Что касается возможностей по обработке текстов, то Python позволяет так организовать регулярные выражения для анализа текстов и манипуляций с ними (в том числе и в Unicode), что написание и доработка кода, работающего с этими выражениями, значительно упрощается. Кроме того, исключаются типичные ошибки, возникающие при изменениях в регулярных выражениях.

Небольшое число и простота правил, а также прозрачность кода позволяют овладеть языком Python в значительно более короткое время, чем этого требуют конкурирующие языки программирования. Python по этой причине является не только языком для "одноразовых" скриптов, но и для универсальных инструментов, таких, например, как конфигурируемый фильтр для таблиц данных, который применим для очистки от устаревшей системы обозначений в старых данных. В этом случае необходимые замены данных в каждом поле определяются с помощью таблиц или регулярных выражений.

## Заключение

Перенос данных из устаревших систем в новые – это трудоемкий, кропотливый процесс. Наиболее важными в нем являются следующие три шага:

- ▶ признание на уровне руководства предприятия серьезности вопроса миграции данных;
- ▶ учет вопроса миграции данных при планировании проекта;
- ▶ тщательный анализ состояния старых данных с целью получения четкого представления об их структуре.

Опыт специалистов компании ECS GmbH в разработке этих решающих трех шагов привел к созданию собственной методики организации и проведения проектов.

Вполне возможно, что дальнейшие действия потребуют также немалых затрат и усилий, но при условии, что перечисленные выше начальные шаги проработаны тщательно и корректно, на последующих этапах практически исключен фактор непредвиденного риска. Здесь центр ответственности смещается в сторону soft-инструментария, то есть набора средств (методик, языков программирования), с помощью которых компетентный эксперт в обозримое время может завершить работу. Нарботанный специалистами ECS GmbH собственный soft-инструментарий достаточно отшлифован и продолжает совершенствоваться в новых проектах, связанных с модернизацией IT-среды.

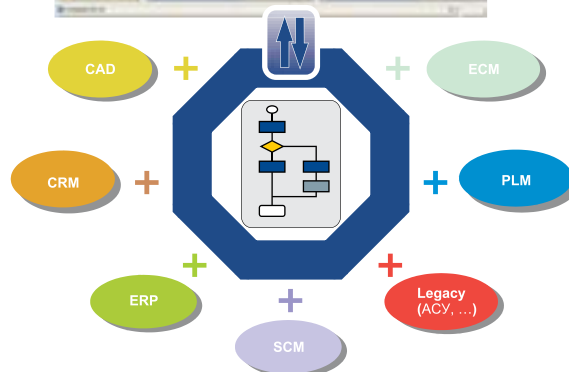
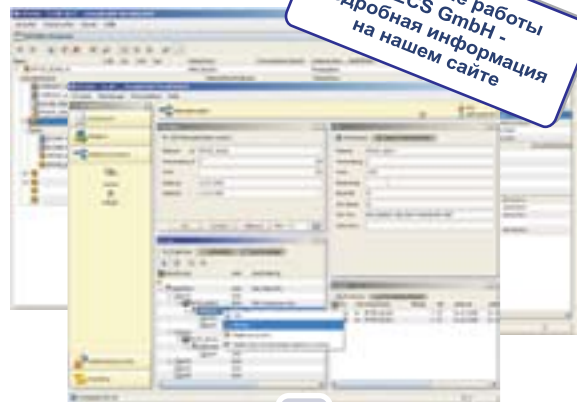
**Р. Фишбах, И. Лахманн, М. Виннемут,**  
компания ECS GmbH

# ecs

**Ваш партнер в PDM/PLM-проектах:**

- современные интеграционные решения
- консалтинг
- тренинг

Дипломные работы  
у ECS GmbH -  
подробная информация  
на нашем сайте



**Наших сотрудников отличают:**

- многоязычие: в палитре языков, на которых говорят наши специалисты, представлен и русский
- опыт внедрения PDM/PLM-систем, приобретенный в многочисленных международных проектах
- опыт переноса данных из устаревших систем в новую IT-среду
- опыт интеграции PLM- и ERP-систем на уровне общих процессов
- знание и использование современных IT-технологий
- опыт создания и внедрения интеграционных платформ

Для ряда сотрудников ECS русский язык является родным

# ecs

Engineering Consulting & Solutions GmbH  
Holzgartenstraße 8  
D-92318 Neumarkt - Germany

Tel : +49 9181 47 64 -0  
Fax: +49 9181 47 64 -50

Mail: [ecs@ecs-gmbh.de](mailto:ecs@ecs-gmbh.de)  
<http://www.ecs-gmbh.de>