

УПРАВЛЕНИЕ ЖИЗНЕННЫМ ЦИКЛОМ ОФИЦИАЛЬНЫХ ДОКУМЕНТОВ В АВСТРИЙСКИХ ГОСУДАРСТВЕННЫХ ОРГАНАХ

Бертольд Конрат (Berthold Konrath)

Австрийский государственный архив

Ноттендорфер Гассе 2, 1030 Вена, Австрия (Nottendorfer Gasse 2, 1030 Wien,
Austria)

ТЕЗИСЫ

Австрийский государственный аппарат строит интегрированную систему управления «от создания до уничтожения» электронными официальными документами и архивным процессом в целях обеспечения корректного управления электронными официальными документами на протяжении всего их жизненного цикла.

В рамках проекта уже реализовано управление официальными документами федеральных ведомств посредством системы ELAK (Elektronischer Akt), а также внедрена спецификация формата межведомственного обмена документами под названием EDIAKT. Последний регламентирует, в частности, предоставление информации в государственные архивы и таким образом, в сущности, определяет формат сдаточных информационных пакетов (SIP, Submission Information Package), которые могут вводиться в архивную систему.

В настоящее время Австрийская федеральная канцелярия финансирует поставку такой централизованной архивной системы, а также приобретение генеральной лицензии, которая позволит всем австрийским государственным органам воспользоваться преимуществами внедряемой технологии архивирования и обеспечения сохранности информации. Конкурс на поставку системы, в которой используется средство цифрового архивирования Safety Deposit Box (SDB) фирмы Tessella, выиграла компания Siemens IT Solutions & Systems (Atos).

Указанная система обеспечит прием сдаточных информационных пакетов (в формате EDIAKT) в архив длительного хранения, а также все необходимые функции доступа, управления данными, их сохранности и администрирования. Система будет

введена в эксплуатацию в 2011 году; первым ее пользователем станет Австрийский государственный архив (Österreichische Staatsarchiv).

Названный проект послужит основой для выработки требований по хранению электронных официальных документов, которые в будущем станут источниками для исторических исследований. Это позволит сохранить историческое наследие Австрии для грядущих поколений.

1. ВВЕДЕНИЕ

Влияние современной информационной техники и технологий (компьютеров, мобильных телефонов, Интернета и т. п.) на нашу повседневную жизнь ощущается уже в течение довольно длительного времени. Оно проявляется и в работе органов государственной власти и, в том числе, в обработке государственных официальных документов. Под этим влиянием появились новые понятия, такие как «электронное правительство», которое в некоторой степени стало синонимом современного государства. В то же время эта, безусловно, захватывающая тенденция ставит и проблемы, вынуждая, в частности, делопроизводителей и архивариусов внедрять новые процессы.

Уже на протяжении нескольких лет различные ведомства используют ИТ для выполнения своих обязанностей, причем в последнее время темпы внедрения соответствующих инструментов растут. Еще в 1985 году Австрийский республиканский архив (подразделение Австрийского государственного архива [1], отвечающее за официальные документы, созданные начиная с 1918 года) ввел в эксплуатацию систему управления электронными документами. В 2003 году Австрийский государственный архив стал одним из первых федеральных ведомств страны, перешедших на использование исключительно электронных дел.

В Австрии за долгосрочное хранение цифровых данных, поступающих как из систем управления электронными документами, так из других систем, а также за прием традиционных бумажных документов отвечает Республиканский архив. Специалисты архива с самого начала были вовлечены во внедрение «безбумажного офиса» (использование федеральных электронных дел), создание системы управления электронными делами (Document Lifecycle Management — управление жизненным циклом документа) и создание электронного интерфейса (EDIAKT II) между различными системами управления электронными документами федерального правительства Австрии.

Несколько позднее, в 2006-2007 году, было проведено исследование осуществимости создания цифрового архива длительного хранения, в рамках которого

были определены требования, возможные решения, условия и категории такого архива. Одним из основных требований стало соответствие системы основным международным стандартам, в особенности стандарту Открытой архивной информационной системы (OAIS, Open Archival Information System). Это продиктовало очевидную необходимость сотрудничества между Австрийским государственным архивом и Австрийской федеральной канцелярией. Создаваемая система получила название Digitale Langzeitarchivierung im Bund (DigLAimBund), основанная на указанных требованиях.

Был организован открытый конкурс, который выиграла компания Siemens IT Solutions & Services вместе со своим партнером Tessella, выпускающим программное обеспечение под названием Safety Deposit Box (SDB), которое соответствует стандарту OAIS. Эта система будет использоваться Австрийским государственным архивом и другими ведомствами для обеспечения сохранности электронных официальных документов для будущих поколений.

2. УПРАВЛЕНИЕ ОФИЦИАЛЬНЫМИ ДОКУМЕНТАМИ ФЕДЕРАЛЬНОГО ПРАВИТЕЛЬСТВА

Часто случается так, что бумажные дела теряются, оказываются не на своем месте, их неправильно заносят в картотеку или забывают на дальней полке архива. В связи с этим одной из важнейших разработок Электронного правительства стала электронная система управления официальными документами под названием ELAK, обеспечивающая непрерывную коммуникацию между органами государственной власти и государственными службами.

В 2001 году Министерство иностранных дел и Австрийская федеральная канцелярия начали использовать систему управления официальными документооборотом ELAK на ведомственном уровне. Затем систему ELAK стали внедрять на уровне государства, а теперь этот процесс затронул и органы местного самоуправления.

Преимущества обработки электронных официальных документов очевидны. Система ELAK существенно сокращает время обработки заявлений, ведь с ней нет необходимости пересылать бумаги из министерств в органы государственной власти и обратно. Удобство заключается в том, что вся обработка осуществляется в режиме онлайн. Процессы стандартизированы и могут идти параллельно друг другу. Запросы осуществляются непосредственно с рабочего стола. Исполнение рабочего процесса полностью прозрачно. Практически одним легким нажатием кнопки можно в любое время суток узнать степень обработки файла. Более того, нет никаких проблем, связанных с изменением формата файла (печатные экземпляры, скан-версии), потому что система

ELAK основана на стандартизированной системе с едиными пользовательскими интерфейсами. Сочтены дни традиционной обработки бумажных дел. Сегодня ей на замену приходят автоматизированные бизнес-процессы.

Выполняя функции управления документами и рабочими процессами, системы осуществляют внутреннюю работу в электронном виде. Таким образом, эти системы становятся своего рода узлами данных, интегрирующими различные приложения и источники данных с тем, чтобы можно было избежать изменения в формате носителя. В системе управления электронными официальными документами федерального правительства можно выделить следующие наиболее важные интерфейсы и системы публичного администрирования:

- Сервер форм. Благодаря отображению форм в графических интерфейсах пользователя, данный интерфейс становится наиболее важным для жителей страны. Заявления, подаваемые в Веб-форме, можно напрямую обрабатывать в системе ELAK за счет стандартизированной структуры данных и XML-синтаксиса.
- Электронная доставка. Чтобы передать некоему лицу информацию, уведомления или документы, необходимо доставить их через службу доставки с использованием описанных методов.
- Интерфейсы к другим приложениям. Нередки случаи, когда для выполнения отдельных процедур требуется получить от граждан определенную информацию. Однако по разным причинам, ввиду определенных трудностей, случается так, что они не могут ее предоставить. Вместо того чтобы гонять людей по различным инстанциям, требуя те или иные сведения, необходимо сделать так, чтобы система ELAK сама могла автоматически получать доступ к нужным данным из таких приложений органов государственной власти, как реестры, системы SAP или службы каталогов. Коммуникация происходит через определенные интерфейсы, которые поддерживают стандартизированный обмен данными.

Согласно требованиям стратегии Электронного правительства Австрии сотрудничество между Федеральным правительством, провинциями и муниципалитетами должно, в частности, проявляться в активном участии в создании интерфейсов, стандартизированных для всех органов государственной власти, и составлении спецификаций, эффективных в масштабах всей страны. В этом процессе рабочие группы должны по возможности основываться на международных нормах и стандартах или использовать их в качестве модели. Типичные компоненты Электронного правительства, необходимые в административных процессах и процессах бэк-офиса, вместе создают общую картину, которая включает в себя, наряду с индивидуальными приложениями,

модули для онлайн-приложений и компоненты концепции карточки гражданина. Протоколы, используемые в архитектуре коммуникации, образно говоря, функционируют как раствор, цементирующий строительные блоки.

3. ОПРЕДЕЛЕНИЕ СДАТОЧНЫХ ИНФОРМАЦИОННЫХ ПАКЕТОВ С ИСПОЛЬЗОВАНИЕМ EDIAKT

Правительственные учреждения могут использовать электронные официальные документы и управлять ими. Но этого достаточно только для начала, ведь существенную роль играет обмен информацией. Конечно, во всех правительственных учреждениях есть системы управления документами (системы электронного документооборота), которые работают с электронными официальными документами, бизнес-решениями и бизнес-разделами, содержащими документы, а также иные различные информационные объекты, которые не подчиняются единому стандарту и разнятся от одного производителя программного обеспечения к другому. Вследствие этого был разработан EDIAKT [2] – формат стандартизации коммуникации между различными государственными учреждениями (властями, судами общей юрисдикции, бизнесами).

В процессе доработки системы EDIAKT, а также ввиду широкого распространения системы ELAK, стандарт был усовершенствован до его сегодняшнего формата – EDIAKT II.

В этом стандарте данные упаковываются как объекты EDIAKT, состоящие из:

- мета-данных, которые описывают официальный документ, бизнес-досье или раздел, неофициальный документ;
- данных о процессе для экземпляров и действий процесса в соответствии с стандартом XPDL Коалиции по управлению потоками работ (WfMC);
- контента официального документа, бизнес-досье (раздела), неофициального документа;
- данных по конкретной процедуре, которые можно прикрепить к объекту.

С целью выполнения требований учреждений, пользующихся системой ELAK, в EDIAKT применяется иерархическая структура с четырьмя уровнями. Внизу располагается неофициальный документ, содержащий файл в исходном формате. В случае если формат, в котором сохранен файл, не стандартный, необходимо дополнительно приложить документ в стандартном формате. Один неофициальный документ или более упаковывается в бизнес-раздел, который представляет собой самый маленький объект в EDIAKT II. Далее такой бизнес-раздел вместе с другими разделами может быть упакован в досье, находящееся уровнем выше.

Даже если у ведомства нет собственной системы ELAK, оно, тем не менее, имеет возможность читать пакеты EDIAKT с помощью бесплатной программы для просмотра EDIAKT Viewer. С помощью последней версии можно:

- отображать все мета-данные, включая данные о процессе,
- показывать документы-вставки,
- подтверждать подлинность цифровых подписей.

EDIAKT II – больше чем просто интерфейс между разными системами электронного документооборота. Его также можно использовать для внутреннего обмена данными между специальными приложениями и архивными системами. При условии ввода стандартного документа формата PDF/A, EDIAKT II вместе с EDIAKT Viewer и EDIAKT Creator формируют прочное основание для длительного архивного хранения официальных документов. В будущем этот формат мог бы сыграть главную роль в процессе подачи оригиналов официальных документов, что часто требуется в суде.

4. ОБЗОР DigLAimBUND

По истечении срока службы государственных официальных документов в стенах создавшего их учреждения (и других смежных учреждений или преемников) необходимо провести экспертизу ценности, чтобы понять, есть ли необходимость в продлении срока их хранения. Некоторые документы, срок хранения которых был продлен в результате такой экспертизы, в итоге будут отобраны для постоянного хранения в Австрийском государственном архиве.

Для выполнения задачи постоянного хранения официальных документов Австрийскому государственному архиву нужна система, дающая возможность приема, хранения, управления, обеспечения долговременной сохранности и предоставления доступа к этим официальным документам. Именно такую систему и представляет собой *Digitale Langzeitarchivierung im Bund (DigLAimBund)*.

Основой DigLAimBund стала уже существующая технология SDB4. Названная система – это серверное Веб-приложение на Java, работающее с реляционной базой данных (в данном случае, Oracle). Вместе с аппаратным обеспечением, системой управления физическими накопителями и системой авторизации и аутентификации, система DigLAimBund способна выполнить все функции, требуемые стандартом OAIS для открытых архивных информационных систем. Далее рассмотрим по очереди каждое из функциональных требований стандарта OAIS, иллюстрируя, как им соответствует DigLAimBund.

В качестве одной из ключевых функций системы является «активное обеспечение долговременной сохранности» (“Active Preservation”), благодаря которому под контролем Технического реестра (“Technical Registry”) происходит автоматизированное, верифицируемое цифровое сохранение. Технический реестр – еще одно серверное Веб-приложение на Java с реляционной базой данных. Он основан на Planets Core Registry [3], который, в свою очередь, основан на национальной архивной системе Великобритании PRONOM [4,5]. Технический реестр содержит не только фактическую информацию (например, список форматов, известное программное обеспечение, пути миграции и т.д.), но также и политики (например, какие форматы или комбинация форматов и свойств принимаются во внимание при признании дела устаревшим и тем самым требующим проведения действий по обеспечению его долговременной сохранности; как измерить эти характеристики, какие действия по обеспечению сохранности нужно произвести в каких обстоятельствах и т.д.). Такая информация сохраняется в машиночитаемом виде, что способствует автоматизации процесса обеспечения долговременной сохранности.

В дополнение к описанным выше функциям в DigLAimBund встроена система управления потоком работ (workflow-система) на базе Drools Flow (разработка на основе «открытого программного кода»). Это позволяет сравнительно легко создавать конфигурируемые рабочие потоки для каждого процесса предусмотренного стандартом OAIS. Каждый рабочий поток состоит из серии шагов. Шаги могут быть как автоматизированными, так и предполагать участие пользователя (через Веб-форму). Каждый шаг автоматически записывается, и так ведется контрольный след (протокол аудита).

5. DigLAimBUND: ПРИЕМ НА ХРАНЕНИЕ

В процессе приема документов на хранение первая функция, которую согласно требованиям стандарта OAIS должна выполнять архивная система, заключается в способности принять поступающий документ. Разумеется, это предполагает, что государственное учреждение передает в архив сдаточный информационный пакет в формате EDIAKT II.

5.1. Прием поступающего документа

Как только произошла физическая передача, пакет должен войти в границы системы DigLAimBUND. Здесь первым шагом, требующим выполнения, является перевод пакета из формата EDIAKT в формат понятный для архивной системы. Так сдаточные информационные пакеты (SIP) конвертируются в архивные информационные пакеты (AIP, Archival Information Package). При этом используются стандартные шаги рабочих

потоков, которые уже присутствуют в Safety Deposit Box, такие как прием, хранение, доступ, управление и обеспечение сохранности данных. Это ограничивает количество требуемых по проекту доработок лишь теми аспектами, которые нужны для особых локальных конфигураций или оптимизаций.

В Safety Deposit Box используется схема метаданных под названием XIP (которая покрывает: SIP – сдаточные информационные пакеты, AIP – архивные информационные пакеты и DIP – дистрибутивные информационные пакеты). Такая схема определяет структурные метаданные, которые необходимы для связывания официальных документов с файлами в данных проявлениях. В случае с EDIAKT II это играет особенно важную роль, так как подается обычно и оригинал и нормализованная форма каждого официального документа (например, документы Word и PDF). В дополнение к этому, схема XIP определяет технические метаданные, что было специально разработано с целью достижения автоматизации цифрового сохранения (см. ниже). Однако XIP не запрещает описательные метаданные, а наоборот позволяет производить описание с использованием любой подходящей схемы метаданных, которая может быть встроена в XIP.

5.2. Контроль качества

В соответствии с требованиями стандарта OAIIS, далее выполняются шаги по контролю качества (“Perform Quality Assurance”). Они включают в себя проверку на вирусы, верификацию соответствия схемам, проверку по контрольным параметрам и проверку уникальности идентификатора для каждого элемента

5.3. Создание архивного информационного пакета (AIP)

Следующий шаг – создание архивного информационного пакета (AIP). Выше отмечалось, что XIP позволяет сделать это путем поэтапной детализации. Но одним из ключевых шагов при этом является характеристика.

Характеризация в Safety Deposit Box сама по себе уже многоэтапный процесс. Она включает в себя и техническую, и концептуальную характеристику. Это полностью автоматизированный процесс, который специально разработан с целью обеспечения в будущем полной автоматизации процессов по обеспечению сохранности.

5.3.1. Техническая характеристика

Техническая характеристика предполагает выявление свойств самих файлов. Цель – обнаружение свойств, которые могут определить, представляет ли файл собой устаревшую технологию и требует ли проведения действий по обеспечению его сохранности:

- В первую очередь производится попытка идентифицировать формат каждого файла с помощью приложения DROID [6]. Важно отметить, что файл связывается

с идентификатором формата. А это можно использовать с целью автоматизации процесса принятия решений, основанных на информации в Техническом реестре, относительно политики по обеспечению сохранности.

- Вслед за первичной идентификацией формата происходит его валидация. По результатам первичной идентификации формата определяется, каким инструментом лучше всего воспользоваться далее (например, JPEG-файлы обрабатывает Jhove [7]). (Решение по политике использования того или иного инструмента в зависимости от итогов первичной идентификации формата сохраняется в машиночитаемом виде в Техническом реестре. Таким образом, данный процесс также может быть автоматизирован). Валидация формата может привести к обновлению результатов его идентификации. Например, на сегодняшний день DROID не различает форматы TIFF3, TIFF4, TIFF5 и TIFF6, в то время как Jhove может валидировать каждый из этих форматов, и, тем самым оставляя только один из четырех результатов первичной идентификации.

- Затем из каждого файла с помощью определенного инструмента вырезаются ключевые свойства. Какой инструмент использовать и какие свойства вырезать зависит от формата и определяется политикой в Техническом реестре. И снова важно отметить, что каждое свойство связано с идентификатором Реестра с целью достижения автоматического применения любой из политик, имеющих отношение к этим свойствам.

- Объекты-вставки, по возможности, вырезаются из каждого файла и по очереди характеризуются. Важность этого процесса заключается в том, что объект-вставка может быть устаревшим, даже если контейнерный файл таковым не является. (Опять же. выбор инструмента для осуществления этой вырезки также зависит от формата контейнерного файла и основывается на машиночитаемой политике, хранящейся в Реестре).

5.3.2. Концептуальная характеристика

В качестве предмета концептуальной характеристики выступают концептуальные единицы, которым необходимо обеспечить сохранность. Такие единицы называют «компонент». Не всегда компоненты являются эквивалентами файлов, так как будучи концептуальными они не зависят от технологии. Например, «Веб-страница» может представлять собой одну компоненту. Хотя в текущей технологии 2011 года, скорее всего, в ее состав будет входить несколько файлов (HTML, CSS, GIF и т.д.). Объединяясь, эти файлы создают концептуальную единицу, сохранность которой требуется обеспечить. Однако нет никаких гарантий, что с развитием технологий физическая структура останется той же.

После идентификации концептуальных компонент необходимо измерить их независимые от технологий свойства, выступающие в качестве «значимых свойств» компоненты. При хорошей миграции такие свойства должны оставаться инвариантными. Сохранность официальному документу обеспечивает сохранность всех его компонент со всеми их свойствами и отношениями между ними.

Фактически, конечно, концептуальные свойства нужно измерять в технологии сдаточного информационного пакета (SIP), тогда свойства компонент будут тесно связаны с техническими свойствами, измеряемыми для отдельных файлов (или их агрегации). Однако, даже если соответствие между файлом и компонентой в текущей технологии взаимно-однозначно, забывать о различии между ними не стоит. Файл характеризуется зависящими от технологии свойствами, сохранность которых, может быть, пока и не требуется обеспечивать. Свойства компонент не зависят от технологии, и поэтому их сохранность должна быть обеспечена. Более подробно рассмотрим этот вопрос далее, в разделе, посвященном сохранности документов.

5.3.3. К вопросу о контроле качества

На практике, шаги создания архивного информационного пакета (AIP) накладываются на шаги проведения контроля качества. Так, в ходе контроля качества ограничения по разрешенным форматам или допускаемым свойствам файлов (например, зашифрованные PDF-файлы не принимаются на хранение) могут вступить в силу только после проведения технической характеристики. Или же выполнение шагов, перечисленных в Разделе 5.2 (например, проверка на вирусы), приводит к добавлению в архивный информационный пакет (AIP) определенных метаданных (таких как информация о программе проверки на вирусы).

5.4. Создание описательной информации

В Австрийском государственном архиве ведется каталог описательной информации по всем фондам архива, как на традиционных, так и на электронных носителях. Следовательно система DigLAimBund должна уметь делать снимки описательной информации, требуемой системе каталога, и давать последний доступ к этим снимкам. Осуществляется это с помощью протокола для сбора метаданных OAI-PMH.

5.5. Координация обновлений

Последним шагом приема документов на хранение, по определению стандарта OAIS, является отправка архивного информационного пакета (AIP) на хранение в реляционную базу данных и в систему массового хранения файлов. Рассмотрим это более подробно в следующих двух разделах.

6. DigLAimBUND: ХРАНЕНИЕ

6.1. Получение и предоставление данных

В качестве системы массового хранения в DigLAimBund используется EMC Centera. Интерфейсы Safety Deposit Box к таким системам как система массового хранения реализованы посредством серийных API (интерфейсов программного взаимодействия), что позволяет изолировать (и сделать независимыми) изменения в системе хранения от изменений программного обеспечения репозитория. Разные интерфейсы позволяют по-разному хранить и извлекать контент (например, со снимком метаданных или без них; сохраняя контентные файлы независимо или внутри пакета; подписывая пакет или нет и т.д.) Обсудим по очереди каждое из этих решений.

Метаданные хранятся в базе данных, поэтому может показаться, что при наличии надлежащей резервной копии не нужно хранить снимок метаданных. Однако если добавить снимок, то в случае невозможности восстановления поврежденной базы данных в системе хранения будет достаточно информации, чтобы восстановить официальный документ. С другой стороны, не стоит забывать, что база данных содержит последний набор информации об официальном документе (например, информацию о случаях доступа), поэтому при утере базы данных некоторая информация будет тоже утеряна. Несомненно, можно также хранить информацию в системе массового хранения, регулярно обновляя снимки. Но это будет тяжело для системы хранения. Система DigLAimBund разумно оптимизирована и выбрана «золотая середина»: снимок метаданных добавляется в систему хранения, но обновляется, только если совершается действие по обеспечению сохранности, но не в случаях доступа или обновления описательных метаданных. В то же время любое обновление описательных метаданных сохраняется в системе каталога. Так они в любом случае получают независимую резервную копию. Такой компромисс предполагает, что историческая информация о том, кто и когда получал доступ к официальному документу, будет утеряна в случае невозможности восстановления базы данных, но он кажется вполне приемлемым.

DigLAimBund также хранит архивные информационные пакеты (AIP) в виде пакетов (один архивный информационный пакет (AIP) на один полученный сдаточный информационный пакет (SIP)). Происходит это частично в связи с принятым решением о политике, частично вследствие характеристик системы хранения, которая в случае хранения большого количества маленьких файлов (например, хранение Веб-сайта) будет терять немалую часть дорогостоящего объема памяти.

В заключение на пакетах ставится подпись XadES, которая позже гарантирует подлинность пакета.

6.2. Управление хранилищем

Согласно требованиям стандарта OAIS система должна уметь управлять иерархией хранилища, по требованию заменять носители, производить восстановление в аварийных ситуациях. Все эти функции входят в состав стандартного набора функциональных возможностей EMC Centera.

6.3. Обнаружение ошибок

EMC Centera предоставляет встроенные функции проверки каждого файла в сопоставлении с его контрольными параметрами на предмет искажений. Кроме того, в Safety Deposit Box имеется функция постоянной проверки целостности (основывается на алгоритме наиболее давно проверенных), которая делает то же самое по всем имеющимся адаптерам для систем хранения (это позволяет, например, сохранить вторую копию в другой технологии хранения). Проверка в Safety Deposit Box также позволяет убедиться, что список известных базе данных дел идентичен реально сохраненным (а контрольные параметры в системе хранения идентичны тем, которые сохранены в метаданных).

7. DigLAimBUND: УПРАВЛЕНИЕ ДАННЫМИ

7.1. Получение обновлений базы данных

Обновления базы данных (запросы на прием документов на хранение либо запросы на обновление, в зависимости от разных причин, описанных далее) поступают в базу данных Safety Deposit Box и обрабатываются системой как единицы хранения из модели данных и сохраняются в соответствующих таблицах базы данных. Вся информация в таблицах хранится в виде фрагментов XML, но там где требуется обеспечить возможность быстрого доступа и быстрой обработки запроса, некоторая информация денормализуется в стандартные поля реляционной базы данных.

7.1.1. Обновления после приема документа на хранение

Safety Deposit Box позволяет обновлять описательные метаданные. Однако метаданные могут также обновляться и в системе каталога. Следовательно, очень важен обмен информацией между системами (через протокол для сбора метаданных OAIPMH).

В дополнение DigLAimBund поддерживает несколько специальных сценариев:

- Одобрить перемещение официальных документов в новую коллекцию.
- Одобрить оценку официальных документов после приема на хранение и, при необходимости, последующий экспорт (в формате EDIAKT) в другую систему или полное удаление.
- Одобрить удаление официальных документов в соответствии с постановлением суда.

Действия по оценке и удалению осуществляются по принципу «четыре глаза» (т.е. рабочий процесс требует, чтобы супервизор одобрил первичную оценку). Удаление данных по предписанию суда (в очень редких случаях) являет собой тщательную операционную процедуру. Для поддержки такой процедуры в Safety Deposit Box также есть возможность «обратимого удаления» (позволяет немедленно сделать официальный документ невидимым для обычных пользователей, не затрагивая при этом весь поток работ).

Помимо прочего, обновления могут стать результатом действий по обеспечению сохранности (см. ниже) или повторной характеристики (проведение характеристики еще раз с целью воспользоваться преимуществами более совершенных инструментов).

7.2. Выполнение запросов

Весь доступ к базе данных осуществляются через Hibernate [8]. Это обеспечивает независимость системы от какой-либо конкретной технологии ядра базы данных (хотя в настоящее время и используется Oracle 11). Соответственно, все запросы в системе осуществляются на HQL, а не на SQL. Все запросы, необходимые для работы системы в нормальных условиях, уже встроены в Safety Deposit Box.

7.3. Создание отчета

В DigLAimBund отчетность может быть внутренняя и внешняя. Для создания внутреннего отчета Safety Deposit Box использует открытое программное обеспечение Jasper Reports (что требует некоторых программистских навыков, но позволяет создавать такие отчеты которые встраиваются в приложение). Внешняя отчетность осуществляется с помощью инструмента отчетности Pentaho (формируются простые отчеты менее техничным способом). В обоих случаях предоставляется полный доступ к единицам хранения базы данных, в том числе и к контрольному следу (протоколу аудита) и истории рабочих потоков. Таким образом, можно получить отчет по полному происхождению каждой единицы хранения.

7.4. Администрирование базы данных

Здесь используются стандартный инструментарий базы данных, предоставляемый Oracle.

8. DigLAimBUND: ДОСТУП

8.1. Координация действий доступа

8.1.1. Требования запроса

DigLAimBund предоставляет пользователям возможность:

- просматривать архитектуру архива («тектонику» архива), чтобы найти интересующий официальный документ;
- искать официальные документы с помощью простого поиска (по всей имеющейся информации) и расширенного поиска (т.е. по соответствующим полям), включая возможность полнотекстового поиска по контенту документа. Каждый поиск выявляет соответствующие критериям официальные документы в порядке релевантности и (в случае полнотекстового поиска) выделяет компоненты в официальном документе, по которым произошло совпадение.

Доступ к официальным документам предоставляется только при наличии у пользователя права на их просмотр. Как только пользователь нашел нужный официальный документ, он может просмотреть все известные о нем метаданные (включая его место в архитектуре архива («тектонике» архива) и описательные метаданные). Также у него есть возможность увидеть список файлов, хранящихся вместе со снимком файла (для обычных форматов).

Сотрудникам архива доступна вся информация, хранящаяся в блоке метаданных, включая:

- Список возможных проявлений, доступный к загрузке.
- По каждому проявлению: список файлов и список составляющих их компонент (обнаруженных при концептуальной характеристике).
- По каждому файлу: все имеющиеся технические метаданные.
- Полный контрольный след (протокол аудита) по каждой единице хранения.

8.1.2. Заказы

Существует два способа заказа контента авторизованными пользователями: "ad hoc", т.е. случайный конкретный заказ (немедленная загрузка или создание образа одного выбранного файла) или заказ официального документа на базе события.

8.2. Создание дистрибутивного информационного пакета (DIP) и доставка ответа

Когда в систему поступает заказ, из хранилища извлекается соответствующий контент и формируется дистрибутивный информационный пакет (DIP). Данный процесс включает в себя: извлечение необходимых файлов из системы хранения, проверку их целостности, упаковку файлов в пакет (например, в ZIP-файл). Затем полученный результат доставляется конечному клиенту. В случае заказов на базе события получить документ можно несколькими способами: по электронной почте, путем загрузки или путем размещения контента в предварительно заданном месте и информировании конечного пользователя текстовым сообщением).

9. DigLAimBUND: ДОЛГОВРЕМЕННАЯ СОХРАННОСТЬ

9.1. Планирование деятельности по обеспечению долговременной сохранности согласно стандарту OAIS

Большая часть деятельности по обеспечению долговременной сохранности, согласно стандарту «Открытая архивная информационная система» (OAIS), связана с планированием, нежели с непосредственным осуществлением конкретных действий по обеспечению долговременной сохранности. В основном, это деятельность, для которой требуется человеческое суждение.

Несомненно, такая деятельность очень важна. Однако одной из главных функциональных возможностей Safety Deposit Box (а, следовательно, и DigLAimBund) является функция «Активное обеспечение долговременной сохранности» (автоматизированный способ осуществления действий по обеспечению сохранности). Рассмотрим этот вопрос далее.

9.2. Политика

Технический реестр, помимо прочего, содержит информацию о форматах (и их технических свойствах) и маршрутах миграции. Такая информация помогает выработать политику относительно признаков, по которым определяется устаревший файл, и действий, осуществляемых для миграции файлов в новый формат. Решение может основываться на абсолютных показателях (например, положение о том, что любой файл, сохраненный в данном формате, считается устаревшим) или на оценке рисков (например, несколько критериев риска в совокупности, при превышении определенного порога служат причиной того, что файл считается устаревшим). Разная политика может быть сформулирована в Реестре в зависимости от ситуации (например, устаревание сохранного экземпляра определяется политикой отличной от той, которая определяет устаревание экземпляра-представления).

9.3. Определение дел и официальных документов под угрозой риска

Критерии политики, определяющие устаревание, хранятся в машиночитаемом виде. Соответственно, можно автоматически сопоставить с ними технические характеристики, полученные на этапе приема документа на хранение (или при последующей повторной характеристике), и тем самым выявить, какие дела требуют совершения определенных действий. Затем можно выяснить, какие проявления каких официальных документов в репозитории требуют какого-то рода действий по обеспечению сохранности. Чтобы идентифицировать, какие проявления релевантны,

каждое проявление официального документа классифицируется по типам (например, «сохранение» или «представление»).

Такой процесс может произойти в любое время, поэтому во избежание повторных миграций к каждому проявлению прикрепляется активный флажок. Таким образом, гарантируется, что только одно активное проявление каждого типа разрешено в определенный момент времени и только активные проявления типа, который соответствует причине миграции (и тем самым сохраненной политике), рассматриваются для миграции.

9.4. Ссылки на измененные официальные документы

Установив, какие проявления официального документа требуют внимания, система расширяет миграцию до всех остальных официальных документов внутри ветви архитектуры архива («тектоники» архива). Так, например, родительский официальный документ того официального документа, который требует внимания, должен иметь возможность быть доставленным в полном объеме в новом проявлении. Соответственно, очень важно, чтобы система проверила, что новое проявление родительского документа (который будет включать в себя файлы дочернего официального документа) когерентно. Возможно такая проверка приведет к дополнительным миграциям файла, возможно, нет. Но она обязательно повлечет за собой дополнительные верификационные проверки, если официальные документы связаны ссылками. Приведем следующий пример: родительский официальный документ – это Веб-сайт, дочерний официальный документ – отчет, размещенный на этом Веб-сайте. Если отчет мигрировал, скажем, из Word в PDF, что привело к изменению расширения в названии файла, то необходимо внести изменения и в html-страницу Веб-сайта, чтобы появилась ссылка на новый файл.

9.5. Миграция

Определив требуемую степень миграции, система выявляет все компоненты официальных документов, обнаруженные при концептуальной характеристике (описано выше). Некоторые из этих компонент будут содержать файлы, которые должны мигрировать (по причине устарелости или эффекта домино, как в вышеописанном случае с Веб-страницей). Это атомные единицы миграции, так как эти единицы и их свойства и отношения – это то, что сохраняется в процессе миграции, несмотря на возможные изменения физической структуры файлов, которая их проявляет.

Основываясь на описанной выше политике, каждая такая компонента проходит по пути миграции, что определяет, какой(ие) инструмент(ы) миграции использовать. Так набор файлов, входящих в компоненту, подвергается миграции в новый набор файлов. Вновь созданные файлы затем проходят техническую характеристику, а новое

проявление компонент концептуальную характеристику. Последний шаг определяет свойства и отношения нового проявления компонент. Они, в свою очередь, должны быть идентичны свойствам и отношениям оригинала (подлежит любым допускам, разрешенным в политике и обусловленным, возможно, допустимой погрешностью округления или предполагаемой деградацией: например, создание для долговременной сохранности образа с меньшим разрешением). Вся эта информация хранится в XML, так что снова этот процесс может протекать автоматически.

По завершении процесса система может агрегировать проявления компонент в проявления официальных документов и ввести их в репозиторий. Система также обеспечивает то, что замещающее проявление становится неактивным и больше мигрировать не может.

10. DigLAimBUND: АДМИНИСТРИРОВАНИЕ

DigLAimBund включает тщательно разработанные функции администрирования для осуществления функций требуемых стандартом OAIS а именно:

- Управлять конфигурацией системы. Сюда входит проведение стандартного администрирования ИТ системы (например, осуществлять контроль за резервными копиями, производительностью базы данных и т.д.).
- Устанавливать стандарты и процедуры, конфигурируя рабочие потоки для приема на хранение, доступа, хранения, управления и обеспечения сохранности данных. Рабочие потоки могут запускаться вручную, через равные промежутки времени или в ответ на контролируемые события, такие как поступление сдаточного информационного пакета (SIP) в определенное место. Если какой-то шаг требует отдельного внимания, соответствующий пользователь оповещается по электронной почте. При входе в систему каждый пользователь видит список любого рода действий, ожидающих его участия. Также можно получить отчет о ходе потока работ или просмотреть, что произошло в потоках работ, которые были завершены в любой момент в прошлом.
- Контролировать права доступа.
- Одобрять обновление архивной информации (например, редактирование метаданных, удаление или оценку, как описано выше).
- Проверять информацию (например, позволить пользователям сообщать о содержании архива или авторизованному пользователю просматривать контрольный след (протокол аудита) любой единицы системы).

- Согласовывать договор о переуступке прав. Договор о переуступке (включая ограничения по размерам сдаточного информационного пакета, разрешенных форматов и т.д.) может быть создан и автоматически верифицирован в процессе приема документа на хранение.

11. ЗАКЛЮЧЕНИЕ

Австрийский государственный аппарат направляет значительные инвестиции в управление электронными официальными документами и архивный процесс. В результате, государственные учреждения уже перешли на систему управления электронными официальными документами (ELAK) и используют формат межведомственной передачи материала (EDIAKT).

На сегодняшний день, программа также включает в себя архивную систему, которая будет введена в эксплуатацию в 2011 году.

12. СПРАВОЧНЫЕ МАТЕРИАЛЫ

- [1] Австрийский государственный архив, <http://www.oesta.gv.at>
- [2] EDIAKT Viewer/Creator, <http://www.ag.bka.gv.at/index.php/EDIAKT-Viewer>
- [3] Домашняя страница Planets project, <http://www.planetsproject.eu>
- [4] А. Браун: Автоматизация сохранения: новые разработки в службе PRONOM, RLG DigiNews 9(2) (2005)
- [5] Домашняя страница PRONOM, <http://www.nationalarchives.gov.uk/pronom>
- [6] Домашняя страница DROID, <http://droid.sourceforge.net>
- [7] Домашняя страница Jhove, <http://hul.harvard.edu/jhove>
- [8] Домашняя страница Hibernate, <http://www.hibernate.org/>