

Ю.Ю.Юмашева

ТЕХНОЛОГИИ БАЗ ДАННЫХ И ИХ ПРИМЕНЕНИЕ В АРХИВНОЙ ПРАКТИКЕ

Слайд 1.

Тема сегодняшнего круглого стола – абсолютно неисчерпаема, поскольку технологии баз данных являются доминирующими технологиями организации и представления информации в электронной среде.

В своем выступлении я попытаюсь наглядно это продемонстрировать, а в его начале напомним, что 15 лет назад, 22 июля 2000 г. история человечества вступила в новую эпоху, поднявшись на качественно новый уровень развития. Для многих из нас это событие произошло довольно буднично и осталось практически незамеченным. Между тем в этот день лидерами большой «восьмерки» была подписана «Окинавской хартии глобального информационного общества», в которой официально признавалось не только существование этого нового миропорядка, возникшего в результате «цифровой революции», но и его глобальный характер.

Прошедшие с того момента полтора десятилетия наглядно подтвердили существенные отличия информационного общества от предыдущих эпох. Самое существенное отличие информационного общества выражено в известной формуле: «кто владеет информацией – владеет миром».

Информация и знания стали самым дорогим товаром информационного общества, одним из стратегических ресурсов государства, масштабы использования которого сопоставимы (а иногда и превосходят) с использованием материальных и природных ресурсов, а доступ к ним – одним из основных факторов социально-экономического развития.

Слайд 2.

Неслучайно один из самых известных деятелей цифровой эпохи создатель и владелец компании Микрософт Билл Гейтс сказал: «То, как человек собирает, управляет и использует информацию, будет определять, окажется ли он в выигрыше или в проигрыше».

Архивы являются собирателями, хранителями и распорядителями огромных массивов информации. И от того, насколько грамотно и успешно эти информационные массивы будут представлены архивистами в электронной среде, зависит и отношение общества к архивам, и перспективы архивного дела в целом.

Вместе с тем, следует отметить, что, как показывает опыт, многие архивисты увлеклись созданием больших информационных систем и информационных ресурсов, не обладая при этом даже базовыми знаниями в области информационных технологий, не представляя сложностей, специфики и, в определенной степени, коварства электронной среды, и положившись на «профессионализм» аутсорсинговых компаний, которые, в свою очередь, плохо знают предметную область архивного дела.

Все это ставит перед нами задачу пройти самый краткий курс на тему «введение в технологии баз данных», которые, как уже было сказано, в электронной среде являются основной формой организации и представления информации.

Слайд 3.

Начнем с определений.

База данных (информационная система) предназначается для информационного отображения (представления в информационной среде) модели определенной предметной области – части реального мира, совокупности реальных объектов, которые представляют интерес для пользователя, исследователя и т.п.

Модель – это упрощённое (структурированное) представление части реального мира и/или протекающих в ней процессов, явлений.

В свою очередь предметная область – это часть реального мира (например, архивная сфера в целом, конкретный архив; совокупность архивных фондов и т.д.).

Эти определения помогают понять, что, к примеру, программный комплекс «Центральный фондовый каталог» представляет собой модель реального Архивного фонда страны, в состав которого включены разные объекты, а массив данных, загружаемый каждым архивом в программный комплекс «Архивный фонд», является моделью состава фондов, описей и единиц хранения данного архива, и, одновременно, отражением всех действий, которые осуществляются в отношении любого из этих объектов.

Слайд 4.

Наиболее известные определения понятия базы данных представлены на экране и не нуждаются в комментариях:

БД – это *совокупность сведений* (материалов, данных) о конкретных объектах реального мира в какой-либо предметной области, представленных и *систематизированных* таким образом, чтобы эти сведения могли быть найдены и обработаны с помощью электронной вычислительной машины (компьютера).

БД – это организованная (*систематизированная*) в соответствии с определёнными правилами и поддерживаемая в памяти компьютера *совокупность данных*, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователей.

БД – это *целевая модель* предметной области, в которой находят отражение только те сведения о предметной области, которые необходимы для реализации этих целей.

Слайд 5.

У технологий баз данных, как и любой другой технологии, существующей в электронной среде, имеется ряд особенностей.

Наиболее важными являются:

Интегрированность и разделенность – под которыми подразумевается – логическое разделение информации между отдельными файлами с целью устранить избыточность информации и облегчить доступ к данным;

Структурированность и взаимосвязанность – отражение существенных свойств объектов реального мира в виде набора (структуры) взаимосвязанных характеристик (атрибутов, полей);

Независимость описания данных от прикладных программ.

Слайд 6.

Все представляемые в БД объекты описываются и хранятся в соответствии с определенной логической структурой – моделью представления данных (моделью данных).

Модель данных определяет способ организации данных, ограничения по их представлению и множество операций, допустимых над объектом.

Всего моделей пять. Все они перечислены на экране. И каждая в той или иной степени применялась или применяется и для создания электронных архивных ресурсов.

Слайд 7.

Самая простая – линейная модель — характеризуется тем, что записи об однородных объектах следуют одна за другой, и для каждой можно указать только одну предыдущую и только одну последующую запись. Такой тип связи между объектами называется «один к одному».

Пример: любой список объектов, составленный по определенному принципу (алфавитный список фамилий, перечень населенных пунктов и т.п.), в котором каждый объект описывается одним и тем же набором характеристик.

Форма представления – таблица или список.

На экране приведены примеры таких баз данных: слева – База данных, созданная архивом Хабаровского края по материалам Главного бюро по делам российских эмигрантов в Манчжурской империи и представляющая собой перечень фамилий эмигрантов. Справа – одна из экранных страниц Книги Памяти Челябинской области.

Слайд 8.

Следующая модель – сетевая. В ней записи об однородных объектах являются узлами, связанными с произвольным числом других записей об объектах. Такая связь между объектами называется «многие ко многим».

Сетевые модели создавались в 1960–1970-е годы для малоресурсных ЭВМ. Это достаточно сложные структуры, и в настоящее время практически не используемые. Насколько они сложны, можно проиллюстрировать высказыванием одного из разработчиков операционной системы UNIX, который сказал, что «сетевая база – это самый верный способ потерять данные».

Слайд 9.

Одна из самых ранних по дате создания и до сих пор весьма распространенная – иерархическая модель представления данных.

Модель характеризуется тем, что у каждой записи есть только одна предыдущая и произвольное число последующих записей. Тип связей «один ко многим».

На основе данной модели в конце 1960 – начале 1970 гг. фирмой IBM была создана первая профессиональная Система управления базами данных – программная оболочка IMS.

Думаю, что примеры такой структуры аудитория легко приведет сама: так графически можно представить структуру организации фондов в любом архиве, осуществляющем фондовое хранение документов;

Слайд 10.

Одна из самых известных икон, написанных великим в XVII веке иографом Оружейной палаты Симоном Ушаковым «Древо государей Российских», представляет не что иное как генеалогическое древо русских князей и царей.

Ниже – фрагмент генеалогического древа английской династии Тюдор, представленный на сайте «Государственные документы он-лайн», созданном Национальным архивом Великобритании и Британской библиотекой и представляющем электронные копии важнейших документов Англии XVI–XVIII веков.

Внешнее представление большинства сайтов тоже организуется в виде иерархической структуры (к примеру, структура главного меню официального сайта Росархива), как и базы данных по генеалогии (к примеру, «Потомственное дворянство Казанской губернии 1785–1917», проект, осуществленный Национальным архивом Республики Татарстан), библиографические классификаторы: УДК, ББК, и Единый классификатор документной информации и т.д.

Иерархические структуры присутствуют в полном объеме или в виде отдельных элементов используются практически во всех информационных ресурсах.

Существенным недостатком этих моделей является сравнительная ограниченность поисковых процедур: т.е. для успешного поиска

пользователь должен помнить, по какой «ветке» ему нужно «спуститься» для того, чтобы найти конечный объект.

Слайд 11.

Наиболее популярной и распространенной в настоящее время является реляционная модель представления данных, предложенная математиком Эдгаром Коддом в 1970 г.

Под термином «реляционная база данных» подразумевается хранилище данных, организованных в виде взаимосвязанных двумерных таблиц. По английский «relation» – «отношение». Основное отличие этой модели от других заключено в том, что модель представляет собой комплект таблиц, в которых описываются характеристики объектов. Комплект, как правило, состоит из основной таблицы, где перечислены все описываемые объекты и их одинарные («атомарные») характеристики (например, название документа и его дата), и набора подчиненных таблиц, в которых представлены так называемые «множественные» (или «переменные») характеристики (например, авторы документа, языки, на котором он написан и т.п.). Каждой такой характеристике посвящается отдельная подчиненная таблица.

Для того чтобы в подчиненных таблицах не повторять уникальный идентификатор объекта (к примеру, его имя, которое, кстати, может и не быть уникальным), каждому объекту в основной таблице присваивается уникальный идентификатор (ключ), по которому с ним связываются относящиеся к данному объекту данные из других таблиц.

Большинство баз данных, создаваемых и используемых в архивной сфере, созданы по этой логике.

Слайд 12.

На слайде сверху справа представлена примерная схема реляционной базы данных, а внизу слева – знакомая всем, кто использует ПК АФ,

процедура установления взаимосвязи «отношения» между конкретной записью в таблице «единица хранения» (карточкой «единицы хранения») и соответствующей данному объекту записью в таблице, представляющей список Актов.

Слайд 13.

При сравнительной простоте и логичности данной модели для ее успешной реализации и функционирования необходимо выполнить ряд довольно жестких условий.

Первое условие заключается в том, что при создании реляционной БД необходимо иметь полное представление обо всей совокупности описываемых объектов, их разновидностях, особенностях и специфике работы с ними, а также о целях создания БД, поскольку именно эти сведения определяют набор граф в таблицах базы данных и набор полей на «карточке» для описания каждого объекта.

Отсутствие такого понимания приводит к трудно преодолимым проблемам. К примеру, первые версии ПК АФ проектировались исключительно для представления в БД фондового учета «бумажных» документов, а сама структура базы данных проектировалась на шаблоне архивных описей, ставшем обязательным в середине 1960-х годов. В результате, с огромным трудом и ценой больших усилий удалось в версиях 4.3. и старше реализовать возможность нефондового учета, а также учета кино, фото, видео, (отчасти) фонодокументов и НТД. При этом так и осталась нерешенной задача отражения в ПК АФ информации из исторических форм учетной архивной документации.

Другой проблемой является постоянное изменение целей создания и функционирования ПК АФ и расширение функционала этой программы: от первоначально запланированной системы учета, представляющей количество и состав документов архивного фонда и формирующей только отчетно-

статистическую документацию, до полномасштабной информационно-справочной системы публичного доступа, реализованной в виде ЦФК. Черты такой системы ПК АФ стал приобретать уже в версии 4.0.

Цели создания и функционирования любого программного обеспечения вообще и баз данных в частности определяет набор характеристик (полей), по которым описываются объекты, и правила их (полей) заполнения. Поэтому изменение целей использования баз данных, когда программные оболочки уже внедрены и активно заполняются, приводит к необходимости увеличивать количество полей и создает дополнительные требования как к уже введенной информации, так и к тем, кто ее вводит, что резко замедляет процесс заполнения баз данных. Здесь, кстати следует отметить один существенный «недостаток» реляционных моделей, о котором, как правило, в первую очередь говорят люди, далекие от технологий БД: большое количество «пустых» (незаполняемых) полей. Их присутствие на экранах карточек объектов мотивировано тем, что БД создаются для всех архивов и всех типов объектов, и если в вашем конкретном случае информация для данного поля отсутствует, то у «соседнего» архива может обнаружиться объект, который потребует наличия такого поля.

Третья существенная проблема связана с тем, что реляционная модель, обладая широкими возможностями по формированию статистических отчетов, выборок, сводных данных, сортировке и фильтрации информации, требует для реализации всего этого функционала обязательного выполнения правил однотипного заполнения полей, т.е. унификации описания. К примеру, для того чтобы грамотно (с точки зрения компьютерных технологий) представить информацию о крайних датах фонда, в БД нужно предусмотреть особую систему, состоящую из 16 полей. К сожалению, вплоть до 2013 г. для ПК АФ на уровне Росархива не существовало никаких инструкций по данному вопросу, что привело к крайне неприятным

последствиям, когда массивы данных, собранные со всей страны, были объединены в рамках ЦФК.

В 2013 г. был разработан проект «Единого порядка заполнения полей ЕАИС», и теперь всем архивам предстоит пройти трудный путь исправления ошибок в своих заполненных ПК АФ. Этот проект является первым шагом в направлении создания отечественного стандарта архивного описания, предназначенного для использования в информационных системах.

Завершая разговор о моделях представления данных, следует отметить, что существуют и иные модели (постреляционная и объектно-ориентированная), однако, на сегодняшний день – это слишком сложные модели, используемые чаще всего в исследовательских проектах или в сфере облачных вычислений.

Слайд 14.

Рассмотрев наиболее популярные логические модели представления данных, необходимо сказать о том инструментарии, с помощью которого они реализуются в электронной среде.

Комплекс программных и языковых средств, необходимых для создания баз данных, проведения всех рабочих операций, поддержания баз данных в актуальном состоянии и организации поиска в них необходимой информации, называется система управления базами данных (СУБД).

Самой известной и широко распространенной СУБД является встроенная в стандартный пакет Windows Office СУБД Access. Это очень удобное программное обеспечение, которое в умелых руках может эффективно решать многие задачи.

Наиболее популярными СУБД являются MySQL и Oracle. С точки зрения набора функциональных возможностей Oracle несколько превосходит MySQL. Однако нужно иметь в виду то, что, в отличие от Oracle, MySQL является бесплатным программным обеспечением, распространяемым либо

по универсальным общедоступным лицензиям, либо по условно-платным лицензиям.

Лицензионная политика компании Oracle более сложная, и цена лицензий на СУБД довольно существенна. Кроме того, поскольку разработку и поддержание работоспособности программ, написанных на СУБД Oracle, могут осуществлять только сертифицированные Oracle специалисты, то цена на подобные продукты весьма значительна.

В качестве примера: на MySQL разработан ПК АФ, на Oracle – программное обеспечение КАИСА-Архив и «Документы советской эпохи».

Слайд 15.

Распространение сетевых технологий привело к тому, что крупнейшие производители СУБД, создали следующее поколение программного обеспечения, основанное на архитектуре «клиент-сервер». Смысл работы этого ПО прост: сервер баз данных принимает запросы по сети от компьютера-клиента и возвращает информацию, соответствующую запросу.

Наиболее известны и широко распространены Microsoft SQL Server и Oracle, а также немецкое программное обеспечение Саперион, эксклюзивным владельцем прав на которое на территории РФ является компания Элар, что создает определенные проблемы при использовании программных продуктов, реализованных на данном ПО.

Слайд 16.

Теперь рассмотрим общетеоретические подходы к собственно проектированию информационных систем вообще, и баз данных в частности. Их довольно много, но в рамках проблемы информатизации архивного дела имеет смысл рассмотреть только три.

Первый называется источником-ориентированным. Он основан на принципе точного переноса и представления в БД полного объема

информации, содержащейся в историческом источнике, без купюр и изменений.

Приведенный на экране пример наглядно иллюстрирует такой подход.

Первая-третья версия ПК АФ была спроектирована в рамках данного подхода. Он абсолютно точно соответствовал задаче, сформулированной в начале заполнения ПК АФ: точное воспроизведение в БД всей информации учетной документации архивов. При этом опись рассматривалась не столько как документ учета, или основной элемент системы НСА, а как исторический источник, любая информация которого должна быть в неизменном виде перенесена в систему.

Однако при этом подходе утрачиваются многие возможности и преимущества технологии баз данных. К примеру, невозможно проводить операции сортировки и фильтрации информации в рамках одного и того же поля, некорректно работают поисковые механизмы и т.п.

Слайд 17.

Нужно отметить, что источниково-ориентированный подход традиционно используется, во-первых, в исследовательских и справочных целях, а во-вторых, он имеет смысл только в случае работы с текстом документа, а не со вторичной документацией, которой являются, например, описи.

Так, его активно применяют филологи, представляя в электронной среде полнотекстовые публикации средневековых рукописей (к примеру, проект «Манускрипт»), его используют в базах данных, где принципиально сохранение точного содержания исторического источника, включая все ошибки и опечатки. Например, портал Министерства обороны «Подвиг народа», где приведены списки награжденных орденами и медалями в годы Великой Отечественной войны. На этом же портале представлен перечень

географических объектов упомянутых на военных картах времен Великой Отечественной.

В качестве примера можно упомянуть и исследовательский проект «Служилые иноземцы в России в первой трети XVII в.», выполненный на основе сложного комплекса документов РГАДА (документы Посольского приказа: в том числе челобитные иноземцев, отписки полковых воевод, переписка приказов, памяти и т.д.). В рамках проекта была разработана специальная методика, которая позволяла учитывать противоречивую информацию (разночтения) об одном и том же человеке, имеющуюся в различных источниках.

Слайд 18.

Следующий подход – проблемно-ориентированный – наиболее востребован в архивной среде. Он основан на представлении в БД тематически связанных документов и/или данных, извлеченных из документов и предназначенных для решения определенных задач. При этом структура БД и набор полей обусловлены содержанием решаемой задачи, а заполнение полей максимально унифицировано.

Т.е. реализацию данного подхода на практике можно разделить на два варианта: первый (назовем его условно «архивным») чаще всего отражает информацию о местах хранения документов, или основывается на уже созданных архивных тематических указателях. В электронной среде – это базы данных, информационные ресурсы, выставочные проекты, тематические подборки документов.

Наиболее характерным примером такой базы данных, спроектированной с позиций проблемно-ориентированного подхода, является база данных о местах хранения дел по личному составу, которая в настоящий момент создается в Справочно-информационном центре федеральных архивов, а

также большинство ресурсов, созданных, в том числе, и в архивах Южного федерального округа.

Второй вариант – сугубо исследовательский – основан на создании базы данных на основе содержания документов. При этом структура полей создается для отражения не всего содержания документа, а только тех характеристик, которые позволяют изучить поставленную исследователем проблему.

Слайд 19.

На этом слайде приведены примеры наиболее известных проблемно-ориентированных баз данных, созданных и функционирующих в отрасли. Среди них стоит особо отметить Государственный реестр уникальных документов – базу данных, которая наравне со справочно-информационным предназначением в некотором смысле выполняет и охранные функции. БД рассекреченных дел и документов федеральных архивов в режиме он-лайн информирующую пользователей о работе по рассекречиванию документов. Развивающийся и активно пополняемый межархивный проект «Документы Советской эпохи».

Проекты федеральных архивов: РГАСПИ «Интебригады», в котором была предпринята попытка не только опубликовать документы, но и дать пользователям простейший исследовательский инструмент для работы с ресурсом; недавний тематический проект РГАЛИ, посвященный Ивану Алексеевичу Бунину.

Отдельного упоминания заслуживают проблемно-ориентированные ресурсы по генеалогии (БД «Ревизские сказки» (Государственного исторического архива Чувашской Республики) и БД, вводящие в оборот документы органов региональной и муниципальной исполнительной власти. Последние БД, пользуются невероятным спросом и повышающие авторитет архивов (например, БД ГАКО).

Особо следует отметить БД, создаваемые в архивах Министерства обороны и МИД. На экране – БД «Документы Второй мировой войны» и «СССР и союзники».

Слайд 20.

Первоначально в 1970–начале 1990-х гг. проблемно-ориентированный подход был доминирующим в исследовательской практике. При этом создаваемые на основе архивных материалов БД не предназначались к публикации. Впоследствии многие созданные в то время базы данных стали основой для крупных архивных проектов. Так, созданная в 1993 г. аспиранткой истфака МГУ база данных по личным делам бакинских нефтепромысловых рабочих «Товарищества нефтяного производства "Братья Нобель" (начало XX в.)». БД легла в основу международного проекта фонда братьев Нобель.

Другим примером может служить созданная в Оксфордском университете межархивная БД с названием «Эпоха просвещения», в которой собраны, описаны и представлены свыше 64 тысяч писем и прочих материалов переписки 6,4 тысячи корреспондентов – выдающихся ученых эпохи Просвещения (Вольтера (18 тысяч писем!), Лейбница, Руссо, Дидро, Ньютона, Франклина и др.). Доступ к ресурсу платный.

Такое же «исследовательское» начало послужило отправной точкой для создания и последующей публикации на сайте Национального архива Великобритании БД по завещаниям жителей Англии. БД включает в себя более 40 млн электронных копий текстов завещаний и развитую поисковую систему, позволяющую найти аналоги для того или иного прецедента, разбираемого в суде, и тем самым значительно облегчить юридическую практику. Доступ и просмотр бесплатный, получение заверенной копии – платное.

На экране также представлена польская база по узникам нацистских лагерей и просопографическая БД по аристократическим родам средневекового Китая.

Слайд 21.

В настоящее время самыми популярными и востребованными проблемно-ориентированными базами данных, основанными на архивных документах, являются базы данных по вспомогательным историческим дисциплинам. К примеру, созданная в Австрии межархивная база данных (указатель) филиграней средневековой бумаги (в ней принимал участие ГИМ), аналогичная база данных, сформированная на основе каталога Пикара и размещенная на сайте Штутгартского архива, проект по средневековой западноевропейской палеографии (Национальный архив Великобритании и Оксфордский университет), и, наконец, база данных по средневековой геральдике, созданная в национальном архиве Финляндии.

Слайд 22.

Еще один подход к созданию баз данных, о котором вскользь уже упоминалось – объектно-ориентированный. Он основан на описании каждого объекта методами специального языка программирования.

Достоинство подхода: отказ от «пустых» полей реляционных БД. Недостаток – «штучное» описание каждого объекта методами языка программирования, при котором оператор становится своего рода «программистом».

На основе такого подхода в середине 1990-х гг. сотрудниками Мосгорархива был реализован проект «Лишение избирательных прав в Москве в 1920—1930-е годы. Новые архивные материалы и методы обработки». На основе СУБД KLEIO была создана объектно-

ориентированная БД, проведена ее статистическая обработка, и в 1998 г. опубликована монография.

Широкого распространения данный подход не получил и сохранился в практике создания БД на основе археологических коллекций.

Слайд 23.

Завершив рассмотрение общетеоретических вопросов, рассмотрим типы базы данных, используемых в практике архивного дела и создаваемых самими архивистами или по их заказу.

На слайде перечислены основные типы архивных баз данных, выделенные по целевому предназначению.

Во избежание недоразумений хочу особо обратить внимание, что все информационные системы, большинство информационных ресурсов (включая ресурсы, представленные в интернет), а также большинство применяемого в отрасли программного обеспечения функционирует на основе технологий баз данных. Поэтому в дальнейшем термины «база данных» и «информационная система» будут использоваться на равных.

Слайд 24.

Самые известные в архивной сфере – учетные базы данных. На экране представлены экранные формы отраслевого ПК Архивный фонд (версии 4.3. и выше), разработанного по заказу Росархива, а также используемые в архивах программы КАИСА-Архив, Элар-Архив, и информационная система архивов РАН.

К сожалению, вынуждена констатировать, что с точки зрения собственно процедуры учета ни одна из этих информационных систем собственно учетной не является.

Слайд 25.

Что имеется в виду? Поясню на практике.

У баз данных, выполняющих функции учета историко-культурных ценностей (архивных документов в том числе), имеется ряд особенностей.

Первое. Базы данных должны иметь так называемый «закрытый» характер. Это означает, что любая информация вносится в соответствующие поля один раз. Правильность заполнения полей проверяется комиссией; после чего поля закрываются для дальнейшего редактирования. Т.е. должен быть реализован принцип неизменяемости учетной информации, применяемый при создании бумажных описей.

На сегодняшний день ни одна из перечисленных систем такую функцию не реализует. Технологически и технически для осуществления подобного функционала нет никаких препятствий. Препятствия заключаются в качестве содержания описей, точности переноса информации в информационную систему (особенно это касается результатов деятельности аутсорсинговых компаний), в нашем неизбывном желании редактировать, т. е. – в человеческом факторе.

Второе. Учетная система предполагает исключительно автоматизированный режим получения отчетов, справок, сводных таблиц и т.п.

Третье. Учет всех видов и типов объектов, включенных в состав АФ РФ.

Четвертое. Возможность ведения всех видов учета: фондового, нефондового, музейного, библиотечного и т.п.

Пятое. Ввод информации в соответствии с определенными правилами. (Здесь возникает методологическая проблема разницы между содержанием описей и отражением этого содержания в информационных системах. Существует известная дилемма, основанная на двойственной природе описей и выраженная в ответе на вопрос о цели создания автоматизированной учетной системы: учетная система для получения сведений «о количестве и составе» (Регламент гос. учета), т.е., прежде всего, количественных

показателях, или для точного отражения сведений описи? В ПК АФ предусмотрен механизм решения этой проблемы: поля системы заполняются по оговоренным правилам, а электронная копия имеющейся описи, сохраняющая все отличия имеющиеся в написании и оформлении текста, прикрепляется в качестве гиперссылки к соответствующей записи. При этом, ПК АФ дает возможность хранить любое количество подобных ссылок, а сами электронные образы не интегрируются в систему, а хранятся в специальных каталогах).

Шестое. Учетная система – это инструмент постоянной работы, который предполагает отражение всех действий с единицами учета в режиме «реального времени» (ввод актов в систему).

Седьмое. В соответствии с «Временным порядком ведения автоматизированного учета»... автоматизированный учет ведется на основе уже созданной бумажной документации. В настоящий момент все острее ставится задача о ведении учета изначально в системе с последующей распечаткой готовых описей и их заверением установленным порядком. Технически и технологически для реализации этого подхода препятствий нет. Препятствием является действующая нормативная база отрасли, в частности, «Правила организации ...» и «Временный порядок», которые не предполагают такую возможность.

Несколько слов о перспективах ПК АФ. Говорить об этом сложно, т.к. за последние 3 года Росархив не объявлял конкурс ни на техническую поддержку ПК АФ, ни на модернизацию комплекса. Вернее, в конце прошлого года такой конкурс был объявлен, но результаты выполнения работ по этому конкурсу мне неизвестны.

Слайд 26.

В прошлом году на НМС, проходившем во Владикавказе, говорилось о том, что в перспективе предполагается создание на базе ПК АФ

полномасштабной многофункциональной портално-модульной системы автоматизации внутриархивной деятельности, ядром которой должен стать модуль автоматизированного учета документов Архивного фонда. Схема такой модернизации представлена на слайде.

Слайд 27.

В конце 2014 г. ВНИИДАД выполнил научно-исследовательскую работу по теме «Концепция информатизации (автоматизации) деятельности государственного архива», в которой центральное место заняло описание инструмента эффективного функционирования государственного архива в современном информационном мире – корпоративной информационной системы, созданной на платформе ПК АФ и реализованной в форме внутриархивного портала. Основное предназначение системы – обеспечение взаимодействия специалистов архива при реализации их функций по всем основным направлениям работ государственного архива, предоставление актуальной информации для принятия решений в управлении основной деятельностью и обеспечение сотрудников подразделений спектром информационно-телекоммуникационных услуг.

Информатизация собственно архивной деятельности в рамках этого портала должна обеспечиваться функционированием 9 взаимосвязанных специализированных подсистем и 20 программных модулей, в том числе предназначенных для управления базовыми архивными технологиями (автоматизации внутриархивной деятельности):

- Комплектованием;
- Ведением архивного учета;
- Осуществлением научно-справочной работы;
- Управлением логистикой архивохранилищ;
- Обеспечением и мониторингом сохранности документов;

–Созданием, хранением и управлением информационными ресурсами (включая оцифровку фондов и учетных документов архива);

–Использованием документов (включая автоматизацию работы читального зала, обеспечение удаленного доступа к информационным ресурсам и электронную коммерцию);

–Осуществлением межведомственного взаимодействия при оказании государственных услуг;

–Взаимодействием с внешними информационными системами архивной отрасли (Фондовый каталог, Центральный фондový каталог, Государственный реестр уникальных документов);

–Обеспечением интеграционного взаимодействия с отраслевыми порталами, органами управления архивным делом субъектов Федерации и/или Федеральным архивным агентством (Росархив) и др.

Текст Концепции представлен на сайте Росархива. Однако уверенно сказать, что данная концепция является стратегией информатизации государственных архивов, я не решусь.

Слайд 28.

Третий тип баз данных, используемых в практике архивного дела и создаваемых самими архивистами, или по их заказу, являются справочно-информационные базы данных.

Упомяну общеотраслевые: Центральный фондový каталог и БД «Путеводители по фондам», размещенные на портале «Архивы России».

ЦФК – полностью индексируемая общеотраслевая государственная информационная система, представляющая собой информационную модель Архивного фонда Российской Федерации и одна из главных форм представления научно-справочного аппарата архивов. На 01 мая 2015 г. в нее загружено более 700 тыс. описаний фондов.

БД «Путеводители по фондам» включает в себя 135 справочников по федеральным региональным архивам, опубликованные в период с 1987 по 2011 годы. База данных содержит описания около 230 тыс. архивных фондов.

Очевидно, что и по объему информации и по уровню информационного сервиса БД «Путеводителей» объективно проигрывает ЦФК, что сказывается и на количестве посещений. Но для людей, привыкших к логике «бумажной» системы НСА, такое «послойное» представление системы: отдельно БД описаний фондов на основе путеводителей, отдельно БД описей, БД заголовков единиц хранения более привычно. Другой вопрос, что пользователей, уже привыкших к большим, наполненным, многоуровневым и всеобхватным информационным системам, это давно не устраивает.

Слайд 29.

Наиболее распространенной формой представления элементов научно-справочного аппарата в сети Интернет являются различные виды информационных систем, создаваемых архивами на основе описей и размещаемые ими на собственных сайтах.

В результате анализа информационных ресурсов, созданных на основе описей и размещенных на сайтах архивов в Интернет, были выявлены наиболее типичные формы представления описей. Их три:

1. информация описей, представленная в виде полностью индексированных информационно-поисковых систем. Информация описей в этих АИС может быть систематизирована либо по архивным шифрам, либо по тематике собрания архива (как, например, информационные ресурсы, представленные на сайтах ГА РФ, ЦГАСО (ЦГА Самарской области), РГАЭ). Реализация этой формы представления описей является наиболее дорогостоящей, но при этом – и максимально эффективной, и перспективной. К тому же, она может быть реализована в качестве выгрузки информации из

ПК «Архивный фонд», поэтому именно к этой форме представления информации описей стоит стремиться.

2. Описи, представленные в виде электронных текстов в формате doc, rtf, размещенные на сайтах архивов или органов управления архивным делом: например, Государственного архива Тульской области, или Агентства по делам архивов Астраханской области.

3. электронные копии (изображения) страниц описей в форматах pdf, jpeg (как интегрированные в информационно-поисковые системы в виде гиперссылок на соответствующем уровне описания информации, так и представленные непосредственно на сайтах) (например, разделы на сайтах РГИА, РГАДА, Государственного архива Калининградской области).

Мотивация выбора формы электронного представления информации описей обусловлена несколькими причинами. Основные – это качество самих «бумажных» описей, наличие специализированных знаний в области информатики у архивистов и объемы финансирования, выделяемые для реализации этой работы.

Вместе с тем в этом вопросе нельзя обойти вниманием тему качества представленной информации, которая тесно связана с процессом усовершенствования описей. К сожалению, включение в состав электронных ресурсов описей, не прошедших процедуру усовершенствования, резко снижает ценность информационных систем и наносит объективный вред авторитету конкретных архивов и архивному сообществу в целом.

Особого внимания заслуживает вопрос о переводе в электронный вид и предоставлении доступа к такой форме НСА как каталоги и картотеки, которые являются объектами творческой деятельности коллективов архивов. Многие из них создавались в служебных целях. Кроме того, их перевод в электронную форму и представление в сети Интернет требуют значительных финансовых затрат, а также серьезной адаптации к специфике представления в информационной среде, без которой невозможно их эффективное

использование. Поэтому создание информационных ресурсов на основе каталогов и картотек каждый раз требует индивидуального подхода. А представление созданных ресурсов в сети Интернет, вероятно, целесообразно организовывать на условиях платного доступа.

Слайд 30.

Зарубежный опыт представления научно-справочного аппарата в сети Интернет на основе технологии баз данных значительным образом отличается от российского. Здесь стоит обратить внимание на то, что за рубежом (имеются в виду крупнейшие западноевропейские страны, США, Канада, Австралия, Япония), во-первых, отсутствует понятие объединенного в масштабах страны национального архивного фонда. А во-вторых, принято подокументное описание архивных документов на основе международных стандартов архивного описания ISAD (g), EAD, EAC. Последнее обстоятельство значительно облегчает решение проблемы объединения описаний.

К примеру, NARA ведет проект Online Public Access, в который каждый архив или частное лицо-владелец отдельных архивных документов может добавить записи о хранящихся у него документах, если эти записи сделаны на основе формата EAD.

Аналогичные проекты реализованы в Италии и Новой Зеландии.

В Италии специалисты университета г. Рима и итальянской компании GAP разработали специальную информационную оболочку, в которой объединены все когда-либо применявшиеся в исследовательской практике стандарты и форматы электронных описаний оригиналов объектов историко-культурного наследия (включая архивные документы) и их электронных копий. Система SINAPSI позволяет создавать полномасштабные археографические описания любых типов и видов исторических источников (письменных, вещественных, изобразительных, кино, фото, фоно,

фольклорных, этнографических, устных и т. п.), самостоятельно определяя полноту и степень их детализации и обращая особое внимание на описание специфических особенностей материальных носителей и способов фиксации информации в подлиннике; хранить электронные копии, независимо от технических параметров, сопровождая их полноценными метаданными (техническим описанием).

Слайд 31.

Четвертый тип баз данных, применяемых в архивной сфере, – тематические базы данных. Они весьма разнообразны и специально говорить о них не стоит. Тем более, что мы подробно рассмотрели их на примере проблемно-ориентированного подхода.

Слайд 32.

И, наконец, последний, пятый тип баз данных – исследовательские. Не секрет, что на сегодняшний момент этот тип баз данных для архивов является чрезвычайно редким. Исследовательские базы в архивах создаются либо в рамках индивидуальной научной деятельности сотрудников, либо в рамках выполнения научно-исследовательских работ, осуществляемых по грантам, либо в сотрудничестве с высшими учебными заведениями. Сотрудничество с ВУЗами всегда складывается непросто. Тем не менее, можно упомянуть несколько совместных проектов. К примеру, Базу данных и информационный ресурс «Из истории региональной политической рекламы», созданные совместно сотрудниками Пермского ГАНИ и Пермского государственного университета. Сложные «сборные» проекты, реализованные сотрудниками Пермского Гос. университета на основе архивных документов, хранящихся у разных фондодержателей: «Колчаковские газеты» и «Первая мировая война в Пермской губернской периодике». В настоящий момент, насколько мне известно, начинается

совместный проект Управления архивами Свердловской области и Научной лаборатории «Международный центр демографических исследований» Уральского федерального университета по созданию источниково-ориентированных баз данных на основе архивных материалов – Церковно-приходских книг; Ревизских сказок, первичных материалов переписей (Партийная перепись 1922 г.) Уральского региона.

Одним из наиболее интересных примеров подобного сотрудничества архивистов и исследователей, является работа группы историков Стэнфордского и Оксфордского – Mapping the Republic of Letters (Картография Республики писем). Используя представленную в Интернет базу данных электронных копий архивных документов Electronic Enlightenment (Эпоха Просвещения), и, применив метод визуализации траектории перемещения писем (подобно тому, как это происходит в современных GPS-системах), изучив динамику развития этой глобальной эпистолярной сети во времени, исследователи смогли не только проследить географическое происхождение и пункт назначения каждого письма, но и доказать, что места зарождения и пути и механизмы распространения идей Просвещения значительно отличаются от общепринятой точки зрения.

Слайд 33.

Основой подобных проектов является выявление и описание конкретных архивных документов, а создание развернутых описаний – задача чрезвычайно трудоемкая, требующая участия профессионалов. В настоящий момент информационные технологии дают возможность компенсировать отсутствие специалистов в числе сотрудников архивов, привлекая профессионалов в режиме «удаленного доступа» (так называемые проекты co-laboratories) или используя волонтеров, работающих либо в специальной технологической среде, либо непосредственно в фондах. Широко известны проекты Национального архива Великобритании: по генеалогии – Find my

past, основанный на технологии краудсорсинга¹ и привлечении «волонтеров» («друзей архива») для изучения и описания архивных документов. Описание производится в созданных в архиве базах данных с легко трансформируемой структурой, адаптируемой для добавления новых характеристик, необходимых для более точного описания исторического источника. Представленный в сети Интернет ресурс позволяет любому жителю Соединенного королевства в режиме он-лайн присоединиться к процессу обработки архивных источников, почувствовать себя исследователем и внести свою лепту в изучение прошлого. Другой проект англичан – публикация дневников солдат, сражавшихся на фронтах Первой мировой войны, реализованный с помощью системы удаленного доступа Zooneverse, где были представлены электронные копии документов, и каждый желающий мог их расшифровать и предложить свой вариант прочтения рукописных текстов. На этой же платформе все желающие могут принять участие в описании и расшифровке коллекции оцифрованных античных папирусов из Оксфордского собрания египетских древностей.

Привлечение максимально широкого круга заинтересованных лиц и использование преимуществ сетевых технологий (в том числе, технологии Web 2.0) являются современными формами решения задач представления копий исторических источников в цифровой среде. По этому же пути в настоящий момент идут многие страны, к числу которых относятся и Польша, архивисты которой реализуют проект «Твои фото. Общая память», предлагая своим согражданам принять участие в атрибутировании опубликованных в сети электронных копий архивных фотографий, подлинники которых хранятся в Национальном архиве.

¹ Crowdsourcing (англ.) – передача некоторых производственных функций неопределённому кругу лиц, решение задач силами добровольцев, часто координирующих при этом свою деятельность с помощью информационных технологий, при описании и атрибутировании архивных документов, а также комментарии пользователей к уже опубликованным документам.

Очевидно, что данный метод имеет свои отрицательные стороны, самой существенной из которых является организация экспертизы содержания предложенных описаний, требующая больших усилий профессиональных архивистов и источниковедов. Но, с другой стороны, и польза от этих проектов может быть значительна.

Эти или аналогичные им методы нашли свое применение в последний год и в России. Так на базе технологии баз данных и подхода «народный архив» сформировано несколько значимых информационных ресурсов, существующих только в виртуальной среде: «Бессмертный полк», «Победа», «Живая память» и др. Опыт участия в одном из этих проектов показал, что в российском обществе при соприкосновении с историческим прошлым и историческими артефактами присутствует высокая степень моральной и этической ответственности, позволяющая на достойном уровне осуществлять подобные проекты.

Слайд 34.

Последняя тема, которую необходимо затронуть в рамках данного выступления. Тема сложная, неоднозначная, требующая отдельного подробного разговора. Я коснусь ее только вкратце: авторское право на БД.

Согласно ст. 1261 IV части Гражданского кодекса: Программы для ЭВМ «Авторские права на все виды программ для ЭВМ ..., охраняются так же, как авторские права на произведения литературы...».

Автор программы для ЭВМ и базы данных обладает рядом исключительных имущественных и личных неимущественных авторских прав, в том числе пользуется полным объемом авторских прав на «осуществленные им подбор или расположение материалов, представляющие результат творческого труда (составительство). Составитель пользуется авторским правом при условии соблюдения им прав авторов каждого из произведений, включенных в составное произведение».

Это означает, что если БД создана в рамках выполнения служебных обязанностей, всеми имущественными правами на нее обладает та организация, сотрудником которой является автор (архив), неимущественными (право на имя) – ее автор (составитель). Если БД создана в инициативном порядке – все права принадлежат ее автору.

Стандартная регистрация программы ЭВМ и регистрация базы данных производится Федеральной службой по интеллектуальной собственности (ФИПС – Роспатент (<http://www.rupto.ru>)). Срок оформления – не более 2-х месяцев. Проблемы с оформлением (если БД является оригинальным произведением) мне неизвестны.

Консультацию по вопросам оформления авторских прав можно получить на сайтах Роспатента и copyright.ru.

Режим доступа пользователей к любым БД определяет владелец имущественных прав.

Пожалуй, единственной существенной проблемой является вопрос поддержания БД в рабочем состоянии и обслуживание информационных ресурсов. Здесь необходимо обязательно предусматривать отдельное и весьма существенное финансирование. В противном случае первичные расходы на их создание могут быть затрачены зря.

Подводя итог краткому рассмотрению технологий баз данных и обзору их применения в архивной практике, следует, с одной стороны, отметить всеобъемлющий характер этих технологий, а с другой – еще раз подчеркнуть необходимость системного подхода в вопросах повышения квалификации архивистов в информационных технологиях, без знания и уверенного владения которыми в современных условиях уже не обойтись.

В заключение хочу поблагодарить всех архивистов Ростовской области, помогавших мне своими вопросами при подготовке данного выступления.