

Глава 3. Учреждения памяти и положение дел в этой области

Считается, что Музеион (храм муз) в Александрии был предшественником наших музеев, и что Писистратус создал первую публичную библиотеку в Афинах в шестом веке до нашей эры. Архивы, предположительно, возникли в первых государствах, обладавших письменностью, и находились там же, где находилась власть.¹ В них должны были сниматься копии с государственных законов, которые хранились у высшего должностного лица.

Такое тройственное разделение мест хранения коллективной памяти (музеи, библиотеки и архивы) просуществовало на протяжении многих веков и, распространившись на всю цивилизацию и соединив деятельность государственной власти и частные инициативы, приобрело универсальный характер. Нас здесь в первую очередь интересуют закономерности, которые определяют область применения возможных методов сохранения коллективной памяти.²

Подчеркнув экспансионистский характер цифровых технологий в отношении социальной сферы, и, в значительной степени, в отношении большинства способов выражения людьми своих мнений, можно говорить о том, что на каждую из этих важнейших функций памяти сохранение цифрового наследия оказывает воздействие в разной степени. Фундаментальное различие определяется природой документа, находящегося в соответствующем учреждении, и степенью оцифровывания на каждой стадии обращения документа.

Существует целая шкала нюансов в оценке документов, начиная с тех из них, цифровые копии которых могут рассматриваться как полные заменители оригиналов, т. е. когда оригинал не отличается от копии, до документов, которые претерпевают более или менее серьезное

¹ Ж. Деррида. Болезнь архива. Париж, Галилей, 1995 (пер. Эрика Преновица: Архивная лихорадка. Издательство Университета Чикаго, 1996).

² Следовало бы добавить сюда монументальное наследие, но это может рассматриваться как еще один тип музея, тогда как памятники народного творчества, передаваемые в устной форме, рассматриваются как еще один вид библиотеки. «Когда умирает старый человек, это то же самое, как если бы дотла сгорела библиотека», как сказал Амаду Ампате Бэ.

ухудшение качества, и документов, природа которых принципиально чужда цифровой форме. Их копии будут не более, чем упрощенное изображение оригинала.

3.1. Коллекционирование цифровых объектов

Виртуальные музеи и музеи виртуальных объектов

Музеи часто сохраняют уникальные или редкие объекты (до XVIII века они обычно назывались кунсткамерами) и всегда представляют большую ценность. Весь художественный рынок основывается на оценке физических объектов. Однако мы слабо себе представляем экономическую модель, которая должна лежать в основе оценки нематериальных объектов. Объекты, воспроизводимые механическим способом, такие как фотографии, несколько искусственно возродили концепцию датированных и/или помеченных оригиналов и таким образом нашли свое место на художественном рынке.

Вообще говоря, музейные экспонаты имеют более двух измерений, и понятие оригинала является определяющим. Цифровые технологии пока еще не могут предложить такого метода, который позволил бы вообразить, что виртуальный цифровой объект может полностью заменить реальный, трехмерный физический объект. В особенности это справедливо по отношению к монументальному наследию.

Цифровые технологии, таким образом, затрагивают музеи только в минимальной степени, главным образом, это касается функции распространения, и иногда, когда какой-то объект погибает, сохранения содержания, на что требуются относительно небольшие усилия. Цифровые технологии являются незаменимыми для распространения работ и демократизации доступа к ним. Они позволяют значительно расширить возможности представления работы, либо в режиме он-лайн, либо на мультимедийных носителях, когда работа может сопровождаться критическими отзывами. Иногда такое представление осуществляется в том музее, где она хранится, в случае, когда состояние работы не позволяет выставлять ее, хотя это и не является представлением собственно работы.

Более того, цифровые технологии все в большей степени присутствуют у истоков различных форм художественного творчества: автор-

ских рукописей, архивов архитектурных и дизайнерских проектов, работ фотографов, эскизов сценических и кинематографических декораций, костюмов, архивов рисунков для двух- и трехмерных мультипликационных фильмов, макетов, используемых в градостроительстве, графики и других художественных работ, в процессе создания которых используется весь спектр возможностей обработки изображений.

Как правило, все это подготовительные или промежуточные стадии работы, которая станет, соответственно, либо памятником, либо напечатанным на бумаге документом, либо цифровым или видеофильмом, либо промышленным объектом..., либо Web-сайтом, и все они являются формами публикаций, сохранение которых возможно с помощью известных методов.

Но что необходимо сделать в отношении всех оригинальных материалов, следов происхождения работы, что вызывает такую озабоченность в наше время? Эти оригинальные материалы попадают в ситуацию «между двумя стульями»: пока они остаются в пределах замкнутого цикла создания работы, они являются отдельными объектами, но вследствие их цифрового характера, это бесконечно размножаемые объекты. Их происхождение в качестве объектов частной собственности приведет к огромному разнообразию используемых компьютерных форматов, а это сделает их сохранение для будущих поколений еще более проблематичным.

Электронные архивы

Архивы, поскольку они являются результатом деятельности общественных и частных организаций, формируют общественное и историческое наследие первоочередного порядка, что делает возможным передачу будущим поколениям следов деятельности, как государственной власти, так и народов и отдельных групп. Являясь одновременно инструментами подтверждения деятельности, важными для функционирования современного общества, и привилегированными носителями коллективной памяти, архивы, как правило, находятся в ведении специализированных учреждений или служб. Их обязанности заключаются в собирании и сохранении всех документов, научных работ, отчетов, статистических данных и других объектов... всех носителей,

фиксирующих дела и факты в письменной форме, чтобы информировать живущие и будущие поколения о современной деятельности и деятельности предшествующих поколений.

Эти собрания, как только они должным образом отобраны, классифицированы, каталогизированы, сохранены и обработаны в соответствии с методами, которые были испытаны и проверены веками — разве не возникли архивы одновременно с появлением письменности в Месопотамии? — будут тихо занимать метр за метром на полках хранилищ, ожидая того момента, когда читатель к ним обратится.

В настоящее время эта последовательность операций, до сего дня четко контролировавшаяся, не является такой уж неизменной. Предпосылки к таким изменениям наблюдались, уже начиная с 1960-х гг., когда большие организации начали использовать компьютерные ресурсы для обработки информации, особенно для обработки больших массивов данных, таких как демографические данные. Движение в этом направлении ускорилось за последние десять лет: компьютерные технологии проникли не только во все сферы управленческой деятельности, но организации стали использовать их для внутренней коммуникации и связи с внешним миром.

Действительно, если посмотреть на увеличение количества производимой информации, вызванное как увеличением обрабатывающей способности компьютеров, например, при увеличении числа экономических показателей, так и легкостью написания и воспроизводства, которая порождает высокую степень избыточности при ухудшении качества оригинала, и если рассмотреть в то же время чрезвычайную неустойчивость этой информации — электронная почта, заметки, отчеты фиксируются на материальном носителе только в том случае, если отправитель или получатель сам этого захочет — легко можно увидеть те изменения функций, которые испытывает архив под воздействием цифровых технологий.¹

Мы все еще далеки от «безбумажного» общества. Эта идея, похоже, отступает все дальше обратно пропорционально степени компь-

¹ Как сообщают, многие десятки миллионов сообщений электронной почты были помещены на хранение в NARA: the U. S. National Archive and Records Administration — Управление Американским государственным архивом и документами — в конце правления Администрации Клинтона.

ютеризации общества. Для современных архивов, парадоксальным образом, наиболее конкретный эффект компьютеризации скажется в существенном увеличении потребления бумаги, что будет способствовать сохранности архива. Это произойдет вследствие чрезвычайного увеличения производства текстов, которое стало возможным благодаря офисной автоматизации.

Как мы уже видели (см. Главу 1), эта фаза компьютеризации подходит к завершению. За последние несколько лет электронная почта с присоединением файлов все шире используется в административных и управленческих системах. Более того, законодательные органы легализовали системы удостоверения подлинности электронных документов, передав им одну из последних важнейших функций, которая до сих пор сохранялась для работы с бумажными документами.

Хотя движение в направлении оцифровывания имеет необратимый характер, очень небольшое число учреждений задается вопросом об организации памяти таким образом, чтобы она была доступна будущим поколениям. В прошлом эта ответственность была делегирована учреждениям сохранения информации, которые в силу своей близости к структурам, в которых информация создавалась, собирали документы, как только их практическая ценность (период административного использования) заканчивалась. Документы отбирались в соответствии с твердо установленными критериями, и затем из них формировались полностью идентифицированные коллекции, в которых учитывалось их органическое строение. Эти действия, не имевшие отношения к процессу создания документов, позволяли последним приобрести статус архива.¹ Однако с использованием цифровых технологий, оказывалось все более и более необходимым, чтобы действия, направленные на сохранение документа, были включены в производственную цепочку уже на стадии его создания, и не только с точки зрения классификации содержания, как это делается в настоящее время, но также с точки зрения носителя и формата. Иначе, если этого не сделать, ничего от документа не останется и не будет архива.

Таким образом, учреждения, которые создают информацию, должны сами принимать необходимые меры для ее сохранения, осуществ-

¹ К. Дерен. Электронные архивы. Париж, Французский государственный архив, 2002.

для при этом тесное сотрудничество с учреждениями наследия, задачей которых будет сохранение этой информации для будущих поколений.

Электронные библиотеки

Кардинальные изменения в процессе написания текстов, а точнее — изменения на всех стадиях этого процесса, от акта создания до цифрового распространения текстов в сетях, глубоко затронут библиотеки. Мы, конечно, говорим здесь о функциях библиотек по сохранению и предоставлению в пользование объектов, изданных большими тиражами, на любых носителях, а не об их музейных функциях, которые также часто осуществляются библиотеками (хранение инкунабул, печатей и монет, подлинных рукописей и т. д.)

Большие библиотеки уже начали работу по осуществлению программ перевода в цифровую форму своих старых коллекций, но, помимо этого, работа библиотек (собирающие, сохраняющие и предоставляющие) претерпевает глубокие изменения вследствие существования книг и газет, «цифровых по происхождению», которые прочитываются в Web.

Во многих странах в правовые акты об обязательном экземпляре включаются базы данных и мультимедийные работы на физических носителях. Однако большинство таких актов, до недавнего времени, не включали сохранение он-лайн-документов¹.

Некоторые документы, опубликованные в цифровой форме и хранящиеся в библиотеках, не вызывают особых проблем с их хранением: аудио-CD, в частности, могут быть с легкостью скопированы без потери качества, благодаря их цифровому выходному сигналу.

Видео-DVD, с другой стороны, постепенно будут вызывать все большие трудности из-за встроенных устройств, препятствующих копированию. Введение таких устройств начинает узакониваться под воздействием законов об авторском праве.

Мультимедийные продукты являются еще более сложными. Они не могут существовать без платформ аппаратного обеспечения (мик-

¹ Дания, Финляндия, Норвегия и Швеция включили это в свои законодательства, Франция также собирается вскоре это сделать.

рокомпьютеров или игровых приставок) и программного обеспечения, соответствующего этим платформам. Существует большой риск того, что в игру нельзя будет больше играть, как только платформа или даже версия операционной системы, установленной на данную платформу, изменится. Это уже случилось с некоторыми самыми ранними видеоиграми, приставки для которых уже исчезли или больше не работают (Амига, Атари и т. п.).

Одним из решений этой проблемы является эмуляция¹ старой операционной системы на современный компьютер. Пока еще находятся ностальгирующие разработчики, которые добровольно создают программы-эмуляторы, чтобы добиться работы этих старых игр и программ. Однако весьма сомнительно, будет ли такое решение жизнеспособным в долговременной перспективе и, более того, если библиотекам придется приспособливаться к осуществлению такой функции на постоянной основе, не окажется ли стоимость этого чрезмерно высокой.

В более общем плане, можно говорить о том, что вопрос о сохранении компьютерных языков «в рабочем состоянии», с их исходными редакторами систем кодирования, библиотеками и компилирующими программами и т. д. в настоящее время может быть обращен только к производителям программного обеспечения, многие из которых быстро исчезают, либо к нескольким любителям-коллекционерам.

В то время как коллекционирование распространяемых мультимедийных продуктов может быть сведено к известным категориям содержания, связанным с носителями, используемыми для массового производства, развитие он-лайн-электронных библиотек полностью изменяет методы собирания и сохранения.

С одной стороны, такой метод собирания характерен для объектов цифрового происхождения, и поэтому отпадает необходимость оциф-

¹ Эмуляция — выполнение на одном компьютере программ, написанных для компьютера другого типа. Эмуляция заключается в точной имитации функционирования всех частей одного компьютера на другом. Она может быть выполнена на программном или аппаратном уровнях. Используется для систем, написанных для уже неиспользуемых компьютеров. Эмулятор — микропрограмма, с помощью которой компьютер может выполнять программы, написанные для другой машины.

ровывания, уменьшаются емкости запоминающих устройств, требуемых для сохранения объектов, так как документы уже существуют в исконно присущем им режиме, а не в режиме изображения. Для структурированных текстовых документов такой способ собирания обеспечивает возможность более точной автоматизированной каталогизации. К тому же, в этой области издатели также стремятся разработать устройства, препятствующие незаконному копированию (загружаемая музыка в потоковом режиме, и т. п.), поэтому принцип коллекционирования кардинально меняется и библиотеки должны проявлять активность там, где они раньше просто ждали, когда работа поступит на хранение. И последнее, библиотеки вынуждены научиться работать с он-лайнowymi цифровыми материалами и всем остальным, что для этого требуется, с точки зрения материально-технического обеспечения.

Тогда как последнее легко может быть осуществлено большими библиотеками, в которых работа с коллекциями он-лайнowych материалов сочетается с процессом оцифровывания старых документов в больших базах данных, это коренным образом меняет характер инвестиций, необходимых для маленьких библиотек и поднимает вопрос о концепциях предоставления кредитов и о методах консультирования.

Вначале консультирование по вопросам цифровых коллекций может быть организовано в форме компьютерных сетей внутри библиотеки, что, естественно, должно развиваться в сторону виртуальной библиотеки. Принципиальное ограничение для документов, находящихся вне рамок общественного достояния, состоит в решении вопросов, касающихся издательских и авторских прав. В конечном итоге сама концепция библиотеки для чтения размывается и превращается в нечто, что может стать новыми методами финансирования общественного чтения, основанными, с одной стороны, непосредственно на издательских сайтах, а с другой, — на сайтах некоторых крупных центральных библиотек. Таким образом, должна быть создана совершенно новая система государственного участия, с точки зрения демократизации чтения.

Что касается сохранения, то аудиовизуальный сектор всегда был настоящей головной болью для учреждений, работающих в сфере наследия. Они должны были сохранять в приемлемом состоянии не только множество носителей, подверженных быстрому физическому разрушению, — фильмов и магнитных лент самых разных форматов и

стандартов, но также и целый спектр читающих устройств, которые обеспечивали бы чтение с этих носителей. При этом следовало также учитывать невосполнимые, все возрастающие потери качества сигнала, происходившие при каждом цикле копирования. В качестве меры, противостоящей этой недолговечности с точки зрения хранения, предложенный цифровой способ находится полностью в сфере дематериализации: как только сообщение отделено от своего материального носителя, оно может быть перенесено любым способом на однородный базовый носитель, формат которого будет определен самим сохраняющим учреждением. Если мы еще добавим к этому тот факт, что скорости передачи компьютерных архивов не идут ни в какое сравнение со временем аналогового копирования в реальном времени, то аудиовизуальный сектор, возможно впервые в своей истории, получил надежду управлять своей памятью. Хотя цена этого будет высокой.

Web и поток информации

Нематериальная, неустойчивая, изменчивая, мульти- и гипермедийная, незавершенная и бесконечная сеть Web ставит пугающие вопросы, касающиеся науки о наследии, методы и практическая деятельность которой последовательно формировались на основе коллекционирования и сохранения устойчивых материальных носителей.

Тем не менее, инициативы в области сохранения наследия сегодня множатся и, хотя не все проблемы пока решены, контуры ответов показывают нам, каким путем следует продвигаться, чтобы создать долгосрочную память сети Web.

Управление массами данных

Одним из препятствий, на которое чаще всего обращают внимание, является масса, объем данных, которые считаются такими огромными, что их невозможно контролировать. На самом деле, несмотря на то, что Web быстро расширяется, она становится устойчивой. Реакция учреждений наследия на эти массы данных, в соответствии с их областью компетенции или центром интереса, заключается в расчленении Web, путем создания подмножеств по признакам территории, темы или языка. Большие национальные библиотеки и архивные учреждения, на-

пример, намереваются сохранять ту часть Web, которая находится в рамках их национального наследия.

Создание гиперхархива

Деятельность по расчленению Web должна обеспечивать возможность взаимодействия систем, т. е. возможность доступа стандартным образом ко всем существующим коллекциям. Этот шаг является необходимым, если мы хотим, чтобы Web была воссоздана в полном объеме. Это тем более необходимо, поскольку одной из важнейших характеристик Web является гипертекстовость. Она смещает границы документа, который может быть открыт с помощью бесконечно большого числа способов чтения, таким образом, что направление продвижения читателя к документу определяется рядом переходов от ссылки к ссылке, от метки к метке либо внутри сайта, либо от одного сайта к другому. В качестве реакции на гипертекстовость Web нам следует подумать об обобщенной структуре гиперархива.

Вместилище изменчивого содержания

Еще одной проблемой является изменчивость данных: Интернет представляет собой среду существования потоков данных, характеризующихся кратковременностью существования. Web проявляет исключительную неустойчивость, когда речь идет о содержании: 70 % страниц Web живут не более 4 месяцев и только 10 % всех .com сайтов остаются стабильными более одного месяца. Решение проблемы такой нестабильности содержания может быть найдено в процедуре и инструментах, используемых для коллекционирования. Так, при проведении операций по выявлению изменений, поисковые роботы могут определить любые посещенные страницы сайтов, которые были изменены или обновлены. Такая информация способствует выработке плана поиска изменений, в соответствии с которым контролируется процесс коллекционирования в зависимости от скорости изменения страниц: чем более нестабильным является сайт, тем более высокая частота коллекционирования будет установлена, и наоборот, чем более стабильным является сайт, тем ниже будет частота посещений.

Девальвация совместимости

Совместимость носителей является еще одной характеристикой Интернета, вызывающей серьезные проблемы. Так как все типы информации, будь то тексты, изображения, звуки или аудиовизуальные продукты могут существовать в виде цифровых данных, с которыми можно производить любые действия, они, естественно, могут распространяться в Web, которая вмещает все формы записи, представленные на более ранних носителях, в некоторых случаях продолжающие существовать в Web-телевидении или Web-радио. Совместимость носителей является не просто сосуществованием с более ранними носителями, но скорее их интеграцией в систему взаимообусловленных связей, в которой проблемы смысла, полноты информации решаются с помощью способа взаимодействия одной формы выражения с другой. Такое совмещение, связанное гипертекстовыми ссылками, превращает сам Интернет в своего рода носитель, а не просто среду существования более ранних носителей.

Состояние текущей практики и идей способствует началу процесса выработки решения; в частности, этому способствует необходимость дифференцировать подходы при работе с документами в зависимости от типа носителя. Так, в процессе коллекционирования, автономные окна коллекционирования будут посвящены непрерываемой записи потоковых аудио или аудиовизуальных данных в соответствии с принципами, похожими на те, которые использовались при коллекционировании программ радио и телевидения для целей наследия.

Но чтобы можно было воспроизвести динамическую природу этого содержания во время предоставления этих записей, избегая при этом огромной избыточности данных, каждый объект будет обрабатываться с двух точек зрения, отличающих содержание от структуры. Так, в процессе коллекционирования должно быть сделано различие между информационным содержанием, которое будет способствовать документальному описанию, и его организационной структурой, т. е. тем, как оно было написано и отформатировано, и как оно связано со всем своим окружением, что позволит описывать его структуру. Это пока еще предмет экспериментирования с большими объемами данных, но целью этого экспериментирования должно быть содействие процессу воспроизведения динамической реальности Web.

От незавершенности объекта к единству документа

В универсуме незавершенности вопрос единства документа приобретает особую остроту.

Одни учреждения наследия считают, что с технической точки зрения архивный объект определяется той страницей или тем файлом, которые могут быть локализованы с помощью данных URL, и датируется в хронологическом порядке, начиная с первой публикации. Другие учреждения считают, что единицей документа является сайт или группа сайтов, например, портал, которые должны быть описаны и датированы. Следует, однако, иметь в виду, что с точки зрения восстановления, новые уровни детализации, которые создаются самими пользователями, будут накладываться на объекты, таким образом заданные, например, в соответствии с разбиением на тематические группы по ключевым словам или по отношениям схожести. Поэтому необходимо устанавливать системы восстановления, способные воспроизвести все первоначальное содержание архива.

Средства компенсации технологической нестабильности

Технологическая нестабильность часто представляет собой серьезное препятствие на пути создания памяти Web. Действительно, сама жизнестойкость сети, где одна программа вытесняет другую и где новое приложение раздвигает границы возможного, вызывает серии постоянных изменений в стандартах кодирования и форматах написания, что еще больше уменьшает возможность сохранения цифровых архивов. И в этой сфере сами учреждения наследия, которые поняли степень возникающих трудностей, предложили решения: миграцию, эмуляцию и инкапсуляцию¹, которые позволяют предусматривать

¹ Миграция — массовое перемещение. Эмуляция — точное выполнение на некоторой ЭВМ программы, написанной для другой ЭВМ. Инкапсуляция (пакетирование) — в сетевых технологиях инкапсуляцией называется метод, используемый многоуровневыми протоколами. Суть этого метода заключается в том, что каждый уровень добавляет заголовок в модуль данных протокола, например, в терминах Интернета пакет содержит заголовок физического уровня, за которым следует заголовок сетевого уровня, далее заголовок транспортного уровня, за которым располагаются данные прикладных протоколов.

долговременное сохранение данных. Никаких сомнений не вызывает необходимость решительных действий по определению стандартизированного формата сохранения для всех архивных институтов.

Сохранение и передача нематериального

И, наконец, создание архива Web означает фиксирование ее содержания на массовом запоминающем устройстве¹, и здесь мы сталкиваемся с еще одной проблемой, связанной с Web, ее нематериальностью. Однако, и это обстоятельство возвращает нас к хорошо знакомой проблеме относительных достоинств и недостатков массовых запоминающих устройств, выбор запоминающего устройства и архитектуры приводит к новым ограничениям, в частности, к ограничениям в выборе методов доступа и восстановления первоначального состояния. Действительно, мы должны будем представить себе такую организацию архива, которая обеспечивает доступ, как к его исторической глубине, так и к его пространственному измерению, делая возможным динамичное перемещение среди взаимосвязанных объектов прошлого. Включение такого пространственно-временного измерения в организацию архива будет трудным делом и потребует новых исследований и дополнительных усилий по совершенствованию, которые будут определяться, главным образом, использованием этого нового наследия.

Некоторые аспекты истории цифровых объектов

Цифровое общество сводит воедино представленную как последовательность разрозненных сведений информацию о людях, получая полную совокупность знаний. В самом начале эта последовательность формировалась, главным образом, как результат компьютеризации административных данных, таких как персональные данные, данные о доходах и налогах, досье преступников, данные о социальном обеспечении, о предметах, изучавшихся в школе, и т. п. Затем к ним добавлялись все данные о связях, которые поддерживал человек, выступая

¹ Массовое запоминающее устройство — вспомогательное запоминающее устройство, работающее под управлением центрального процессора и способное хранить большие массивы информации, значительно превышающие емкости обычных вспомогательных запоминающих устройств.

в качестве клиента, с различными поставщиками, то есть данные о структуре потребления и счетах, о банковских счетах, о полисах страхования, и т. д.

Первая волна ограничивалась, таким образом, управленческими данными. Чтобы сохранить личные свободы и тайну частной жизни, многие государства создали барьеры, препятствующие взаимосвязи этих различных архивов, из страха оказаться государством, выступающим в роли «большого брата», страха, который неотступно преследовал их население.

Современные цифровые данные позволяют получить гораздо больше информации. Они могут сообщить обо всех контактах, осуществляемых посредством обмена сообщениями по электронной почте, о персональных Web-страницах, о переходах с одного сайта на другой, о набранных телефонных номерах, о стоимости и продолжительности разговоров. В будущем к этому могут добавиться еще и записи самих телефонных разговоров, данные видеонаблюдения и т. п., поскольку цифровое общество становится способным к регистрации всего, что бы человек ни делал.

Но цифровые данные могут носить и еще более интимный характер, касающийся полной информации о теле человека. Вспомним о возрастающем использовании цифрового сканирования тела в медицинских целях. Развитие геномной инженерии и получение медицинских цифровых видеоданных в значительной мере увеличат количество «цифровых следов», оставляемых человеком.

В настоящее время большинство этих данных не сохраняется либо рассредоточено. Они не собраны воедино и не приносят выгоды. Они пока еще не формируют цифровые архивы отдельных людей. Но технически этому ничто не препятствует. Какими должны быть ограничения на «оцифровывание» человека, включая все компоненты его тела, на создание цифрового двойника человека на протяжении всей его жизни? Начиная с генетического кода до любовных писем, фотографий, сделанных в отпуске, и Web-видео — вся жизнь человека может быть записана на один и тот же цифровой носитель, которым чрезвычайно легко оперировать и который очень просто восстановить.

В этом случае мы имеем потенциально все элементы полной хроники человечества, от нашего общего генетического наследия, до каждого его конкретного проявления. Чтобы не позволить мертвым хватать

живых, чтобы избежать вероятности быть похороненным под лавинообразной перегрузкой памяти, совершенно необходимо для эпохи информации создать новые правила проведения траурных мероприятий, новые ритуалы похорон.

3.2. Перевод в цифровую форму

Целый ряд учреждений, музеи, библиотеки, архивные центры, научные центры и культурные институты, уже готовы, в разной степени, встать на путь долговременного сохранения.

Одни из них опубликовали свои ориентиры и рекомендации ¹, и мы обсудим их в общих чертах ниже. Другие — принимали участие в проведении экспериментов и в исследовательских программах. Все согласны признать необходимость учитывать цели сохранения с момента появления цифрового документа — превентивное сохранение — и поддерживать заинтересованность в сохранении на каждой стадии его существования.

Выработка стратегии

Общая долговременная стратегия с учетом необходимости постоянного сохранения должна определяться всякий раз, когда цифровая информация создается, обновляется или предоставляется. Это должно делаться по следующим причинам:

- разнообразие аппаратных и программных ресурсов должно сохраняться, чтобы обеспечить доступ к содержанию, позволяя предоставлять это содержание по прошествии длительного периода времени, после того как закончится обычный период его текущего использования (например, по причине правового или исторического характера);
- устройства сохранения должны гарантировать читаемость данных, какой бы материальной обработке эти данные не подвергались;

¹ См., например, «Руководящие принципы по использованию электронной информации» (http://europa.eu.int/information_society) и «Принципы сохранения цифрового наследия» (ЮНЕСКО, 2003).

- предоставление или обновление данных может осуществляться совсем другими людьми, не имеющими отношения к происхождению информации;
- должна быть обеспечена возможность многократного предоставления данных в связи с наличием других источников.

К выработке этой стратегии должны быть привлечены все заинтересованные стороны: хранители, архивариусы, библиотекари, инженеры-компьютерщики, специалисты и т. д. Одно из возможных решений состоит в том, чтобы учредить с самого начала междисциплинарную группу для определения и поддержания такой стратегии. При этом необходимо иметь в виду следующие задачи:

- отбор и идентификация цифровых данных, подлежащих сохранению;
- установление правил организации и классификации, в частности, определение единицы документа и категорий, соответствующих этому определению;
- определение стандартов и спецификаций, которые будут использоваться для обеспечения независимости данных от носителей, и служить гарантией их постоянства;
- установление ответственности на каждой стадии;
- определение потребностей пользователей;
- решение правовых вопросов относительно собственности коллекций, их распространения и условий использования;
- осуществление плана по контролю выполнения программы и устранению ошибок;
- создание наблюдательного технологического центра для изучения и последующего внедрения новых разработок, касающихся систем;
- определение политики в области подготовки специалистов и внутренней информации;
- обмен информацией и изучение передового опыта родственных учреждений.

На этой стадии очень важно определить процедуры и правила для материальной и документальной обработки цифровых объектов, с тем чтобы впоследствии иметь возможность следить за ними на каждой стадии их существования.

Каждый документ должен быть точно идентифицирован и описан с точки зрения всех основных данных, которые составляют его.

Техника перевода в цифровую форму

Существует два принципиальных решения для перевода бумажного документа в цифровой формат.

Во-первых, режим изображения, который представляет собой сканирование документа для получения его изображения с более или менее высоким разрешением; современные сканеры обеспечивают разрешение от 300 до 600 dpi (dots per inch — точек на дюйм). Во-вторых, буквенно-цифровой режим, который представляет собой кодирование уже отсканированного документа в электронный формат с помощью механизма оптического распознавания символов (OCR — Optical Character Recognition).

Существуют также механизмы векторизации (преобразования данных в векторную форму), которые используются, главным образом, для обработки графики.

В режиме изображения размер файла, подлежащего сохранению, гораздо больше, порядка 50 килобайт на страницу, по сравнению с несколькими килобайтами буквенно-цифрового кодирования. Одним из возможных решений является сочетание этих двух режимов: сохранение изображений в режиме изображения и обработка текстов в системе OCR.

В случае использования OCR, оператор всегда должен производить проверку с помощью программ исправления ошибок.

Неподвижные изображения также обрабатываются с помощью сканера, с использованием надлежащего стандарта разрешения в соответствии с требуемым качеством и предполагаемым использованием; после этого файл может быть подвергнут сжатию для уменьшения его размера и, следовательно, для облегчения сохранения или распространения в сети.

В случае неподвижных изображений, анимированных изображений и аудио, проблемы возникают в связи со сжатием цифрового сигнала, получающегося непосредственно в результате перевода в цифровую форму (270 мегабит в секунду в случае видео). Сжатие считается не наносящим повреждений, только если оно полностью обратимо, т. е. если существует возможность возвращения из сжатого архива к первоначальному состоянию файла, каким оно было до сжатия.

Каким бы не был носитель, все это особенно справедливо для аудиовизуальных записей. Прежде чем предпринимать операцию по переводу в цифровой формат, необходимо задать вопрос о цели оцифро-

ывания записи, для определения коэффициента разрешения, необходимого для процесса оцифровывания, и коэффициент сжатия. Будет ли целью:

- изготовление дублирующей копии старого аналогового носителя, который может исчезнуть вследствие недолговечности носителя. Идеальным решением в таких случаях будет выбор максимально возможного разрешения и нулевого сжатия, или, по крайней мере, нулевого деструктивного сжатия (в случае видео, это применение цифрового Бета стандарта, то же самое и для аудио, так как потоки сжимаются от 50 мегабит в секунду до 1,5 килобит в секунду). Если следовать этому принципу, часто, по финансовым причинам, возникает необходимость принятия компромисса между количественными и качественными характеристиками оригинала, которые могут не заслуживать необходимых для этой операции вложений. Действительно, стандарт MPEG 2, работающий с потоком информации, превышающим 8 мегабит в секунду, обеспечивает приближение к качеству прямого вещания.

Эта стадия дает возможность продолжать исправления или восстановить оригинальный носитель.

- изготовление издательской копии, т. е. копии, предназначенной для нового использования (издание печатной копии на бумаге, видео, аудио, телевидение, радио и т. д.). Выбранное качество будет зависеть от обычного диапазона качества, используемого в соответствующей издательской области.
- изготовление копии с целью предоставления ее в пользование. Выбор здесь может быть широким, в зависимости от того, предназначена ли эта копия для локальной среды, где объемы могут быть высокими (до нескольких мегабит в секунду в формате MPEG 1 или MPEG 2 видео), или для Web, где объемы не превышают нескольких десятков килобит в секунду при телефонном соединении или нескольких сотен килобит в секунду в широкополосных сетях (MP3, MPEG 4, Real audio или video, Windows Media, QuickTime).

Во время процесса оцифровывания различные типы копий будут часто изготавливаться параллельно. Каждая копия будет храниться на носителе, выбранном в соответствии с планируемым использованием. Носители очень часто различаются (жесткие диски, магнитные ленты, оптические диски) в зависимости от типа копии (дублирующие, издательские, для предоставления).

После обработки данных необходимо провести контроль качества и проверить запись и связанную с ней информацию.

Если для перевода в цифровую форму необходимо приобретать аппаратное и программное обеспечение, рекомендуется включить в контракт отдельный пункт на их приобретение, обязывающий поставщиков предоставлять средства восстановления данных, полученных с помощью их оборудования, чтобы содействовать их долговременному сохранению и использованию. Очевидно, что приобретение оборудования, основанного на стабильных, общедоступных стандартах является лучшей гарантией безотказной работы.

В случае заключения субдоговоров, перечень условий должен гарантировать, что люди, которые будут заниматься сохранением данных, будут обеспечены аппаратной частью, программной частью и необходимой документацией для долговременного использования.

Форматы файлов

При существующем количестве и разнообразии стандартов и норм рекомендуется с самого начала выбрать область стандартов. Учреждения, насколько это возможно, должны придерживаться этого принципа. Кроме того, выбранные форматы должны всегда основываться на общедоступных, многоплатформных международных стандартах. Ниже представлены категории форматов в соответствии с типами объектов, с которыми они могут работать:

- для неподвижных изображений в точечном представлении наиболее широко распространены следующие стандарты: формат TIFF, который согласуется с данными, полученными с помощью сканера; GIF, присутствующий в приложениях Интернет; JPEG, формат сжатия, который становится основным в Интернете ¹;

¹ TIFF: Tag Image Archive Format — файловый формат тега для изображений; GIF: Graphics Interchange Format — формат графического обмена, обеспечивает сжатие данных без потерь; JPEG: Joint Photography Experts Group — разработанный данной группой метод сжатия изображений и соответствующий графический формат, часто используемый в WWW, характеризуется компактностью файлов и более быстрой передачей, чем GIF, но медленным декодированием и «потерей» деталей изображения.

- среди графических форматов, которые позволяют сохранять структуру графики, можно назвать, например, CGM¹, стандарт для векторной графики, который обеспечивает хорошую устойчивость;
- видеоформаты, включающие, в частности, стандарты MPEG-1, предназначенный для CD-ROM;
- MPEG-2, для цифрового телевидения и DVD;
- MPEG-4, который лучше всего подходит для интерактивных приложений, связанных с мультимедиа и Web, а также MPEG-7, который является стандартом, предназначенным для исследовательских целей; AVI, QuickTime и RealVideo² — форматы, находящиеся в частной собственности, используемые, главным образом, для распространения в Интернете;
- аудиоформаты, в частности MP3, который очень широко используется в Web, Wav (Wave audio) и MPEG используются благодаря их аудиосоставляющей;
- текстовые форматы, которые интерпретируют структуру, знаки и представление текста. Стандарт UNICODE, например, имеет дело с алфавитами и идеограммами и постепенно вытесняет старый формат ASCII. Формат RTF³ делает возможным сохранение расширенной версии текста, оставаясь, теоретически, полностью обратимым во все другие форматы. В этой области существует также большое число форматов, находящихся в частной собственности, новые версии которых быстро вытесняют старые. Ни один формат не может действительно гарантировать постоянство. Следует заметить, что формат PDF⁴ позволяет просматривать документ на разных платформах. Хотя это и находящийся в частной собственности формат (им владеет Adobe), он является общедоступным, имею-

¹ CGM: Computer Graphics Metafile — метафайл компьютерной графики.

² AVI: Advanced Visual Interfaces — расширенные визуальные интерфейсы; QuickTime — разработанный фирмой «Apple» компактный формат файла для цифрового видео и анимации; RealVideo — стандарт де-факто на формат видеопотока в Интернете фирмы «RealNetworks».

³ ASCII: American Standard Codes for Information Interchanges — американский стандартный код обмена информацией; RTF: Rich Text Format — расширенный текстовый формат.

⁴ PDF: Portable Document Format — переносимый формат документа.

щим тенденцию к распространению в качестве *de facto* стандарта для представления громоздких текстовых документов, что привело к его использованию многими библиотеками и web-сайтами. Существуют также стандарты для описания и логической обработки текстовых документов, которые предоставляют языки разметки для международных стандартов, таких как SGML, его расширение XML, наряду с HTML ¹, который получил широкое распространение в Интернете.

Запоминающие устройства

Существуют два семейства запоминающих устройств для сохранения цифровых данных:

- Накопитель на магнитной ленте, самый старый и самый недолговечный тип накопителя. Накопители этого типа представляют собой магнитные дискеты, картриджи, кассеты DAT и DLT ². Главным достоинством накопителя на магнитной ленте является их накопительная емкость (300 гигабайт в 2003 г.) и самая низкая цена одного килобайта. Их главным недостатком является медленный доступ к запоминающему устройству. В самом деле, они не позволяют читать информацию непосредственно, для этого требуется, воспользовавшись буфером, скопировать файл на жесткий диск и только потом использовать его. Эти накопители используются либо для осуществления дублирующей функции, либо вместе с накопителями, обеспечивающими более быстрый доступ к данным в роботизированных системах, которые допускают дифференцированное время доступа, или созданы для удовлетворения потребности в очень большой емкости накопителя, например, для аудиовизуальных записей. В области высокеемких накопителей, в настоящее время все внимание сконцентрировано на трех предложениях: S-DLT фирмы Quantum, LTO от IBM и Hewlett Packard и S-AIT от

¹ SGML: Standard Generalized Markup Language — стандартный обобщенный язык разметки и его расширение XML: Extended Markup Language — расширенный язык разметки, HTML: HyperText Markup Language — язык разметки гипертекста.

² DAT: Digital Audio Tape — цифровые аудиокассеты; DLT: Digital Linear Tape — лента для цифровой записи с последовательным доступом.

Sony. Эти три стандарта предлагают накопители, емкость которых превышает 100 гигабайт.

- Оптические накопители, вообще оптические диски, которые позволяют осуществить прямой доступ к информации, но у которых меньшая накопительная емкость. Существовало много решений, основанных на использовании накопителей в виде оптических дисков, обычно известных как DOD¹. Эти решения являются нестандартизованными, и так как эти диски занимают только свою нишу на рынке, они не обладают преемственностью. С другой стороны, варианты аудио-CD, выпускавшиеся «Филипсом» и «Сони» в 1980 г., пользовались значительным успехом:

- это CD-форматы CD-ROM (уплотненные), CD-WORM и перезаписываемые оптические диски, которые соединяют две накопительных технологии, чтобы обеспечить скорость доступа, плотность и возможность перезаписи данных,
- и DVD форматы, видеoverсии которых быстро завоевали широкую популярность.

DVD-ROM, которые обладают емкостью в восемь раз большей, чем CD-ROM, существуют в различных форматах, допускающих перезапись (DVD-R, DVD-RAM, DVD-RW и т. д.) и битва производителей за формат сейчас в самом разгаре.

Помимо использования этих внешних накопителей, жесткие диски также начинают применяться для сохранения данных, поскольку постоянное снижение их стоимости делает их надежными накопителями.

Выбор запоминающего устройства в большой степени связан с характером данных, подлежащих сохранению, и, в частности, с его требуемой емкостью. Выбор также зависит от предполагаемого использования и выполнения функций представления и доступа. В любом случае, рекомендуется выбирать один тип носителя, а если это не удастся, ограничиваться как можно меньшим их числом, чтобы коллекция была как можно более однородной.

Чтобы обеспечить физическое сохранение накопителей, необходимо соблюдать определенные правила, касающиеся поддержания необходимой температуры и влажности. Учреждения стандартизации, такие как Международная организация по стандартизации, Нацио-

¹ DOD: Digital Optical Disk — цифровой оптический диск.

нальный институт стандартизации США и Международный совет по архивам рекомендуют производить перезапись примерно каждые 10 лет.

3.3. Новые законы сохранения

Постоянный перенос на новые носители

Давайте заявим еще раз, что в аналоговую эру сохранность работы зависела от сопротивляемости физическому старению компонентов, из которых она состояла: носителей, чернил, магнитных покрытий..., но какой бы ни была эта сопротивляемость, все носители неизбежно погибали. На это требовалось гораздо больше времени для очень твердых, однородных материалов, на которых надписи выполнялись грубым способом (гранитные стелы, пирамиды, глиняные таблички), чем, например, для видеокассет. Но даже если носитель будет периодически восстанавливаться, он погибнет, так как циклы аналогового восстановления постепенно стирают надпись.

В цифровую эру было бы очень рискованно делать ставку на долговечность того или иного носителя, отмеченного, как и подавляющее большинство современных потребительских товаров, коротким сроком службы и ускоренным технологическим устареванием. С другой стороны, двоичные цифровые сигналы обладают высокой устойчивостью при копировании. Сохранение компьютерного файла будет, таким образом, зависеть от организации систематических циклов переноса информации с одного носителя на другой, а в качестве дополнительной гарантии, создание дублирующих копий и их хранение в разных местах.

Цифровая информация может, таким образом, приобрести свойство существовать бесконечно долго, но при условии, что мы постоянно будем заботиться об этом. Это коренным образом меняет процесс хранения: раньше основными фазами сохранения были коллекционирование, сохранение и затем предоставление информации. В настоящее время сохранение становится ключевой функцией, так как для сохранения любого архива требуется создание искусственных условий внешней среды, контролируемых с точки зрения гигрометрии, температуры и освещенности, а, главное, процесс сохранения нуждается в постоянном управлении. Существуют, однако, и позитивные моменты,

которые способствуют выполнению циклов переноса информации: с развитием технологий стоимость устройств памяти постоянно снижается, а скорость переноса данных с одного носителя на другой постоянно возрастает.

В этой области мы находимся только в самом начале использования тех возможностей, которые предоставляют технологии. Но уже сейчас комплексные системы и автоматическое сохранение цифровой информации включают факультативное наблюдение за ухудшением состояния носителей (анализ скорости пропадания знаков или разрядов на магнитном покрытии) и автоматический перенос на новый носитель при превышении определенного порога.

Еще одним достоинством компьютеризации, обеспечивающим сохранность информации, является избыточность. Мы уже говорили об этом, когда упоминали о дублирующих копиях, но компьютеризация объединила эти системы благодаря появлению программ исправления ошибок, частичной или полной избыточности на жестких дисках и сохранению данных в их отдаленных физических секторах.

В цифровой области постоянство будет, таким образом, поддерживаться благодаря скорости: более быстрый оборот информации, более высокие скорости доступа к информации, более быстрый анализ и исправление ошибок, более быстрое переформирование данных для сохранения их целостности.

В аналоговую эпоху сохранение означало паузу, время перерыва для поддержания срока службы объекта; в цифровой век все наоборот, так как сохранение означает теперь непрерывное движение объекта.

То, что общая стоимость сохранения цифровой информации несоизмеримо выше, чем аналоговой, является установленным фактом. Технологическая конкуренция влияет существенным образом на снижение стоимости многих стадий этого процесса, но структура стоимости изменилась коренным образом, вызвав глубокие изменения в методах работы большинства учреждений. Хранение аналоговых материалов допускало некоторую небрежность в отношении к этому процессу, так как оно заключалось просто в размещении материалов где-нибудь на полках, тогда как компьютеризация не допускает такого подхода.

Новый ряд проблем, сопутствующих управлению архивами, обусловил появление новых программных решений, которые производители компьютеров и программного обеспечения называли «Управление цифровыми ценностями».

Позволим форматам развиваться: перенос, эмуляция, пакетирование

Мы уже убедились, что самые большие трудности, возникающие в результате развития цифрового общества, связаны с множественностью логических форматов, часто находящихся в частной собственности, — операционных систем, файлов для записи данных, исполняемых программ — и очень высокой скоростью их изменений.

Однако какими бы эти трудности не были, решения могут быть найдены, и современное техническое развитие позволяет определить, каким образом может быть достигнуто долговременное архивирование цифровых форматов. Решения, которые уже используют учреждения наследия, могут быть трех типов: перенос, эмуляция и пакетирование. Эти решения могут применяться по отдельности или в сочетании.

Решение, которое используется наиболее часто, это — миграция или перенос, процесс, во время которого биты файла или программы изменяются таким образом, чтобы их могла читать новая операционная система или новая версия приложения. Старые данные заново обрабатываются, чтобы сделать их совместимыми с новой конфигурацией.

Это решение может осуществляться в том случае, когда данные либо физически переносятся с носителя, имеющего другой логический формат данных, либо на уровне операционной системы или форматов файла, либо даже на уровне формата данных или приложения. Любая миграция может, таким образом, выполняться посредством изменения части системы кодирования первоначального содержания, чтобы оно могло использоваться современными системами.

При данной технологической неустойчивости, на которую мы уже обращали внимание, систематический и регулярный перенос всех документов должен планироваться при появлении каждой новой версии. Например, Национальная библиотека Франции осуществила перенос 15 миллионов отпечатанных страниц, переведенных в цифровой формат в режиме изображения, с целью их представления на сайте Интернета Gallica. Этот перенос данных объемом примерно 1 терабайт без изменения их формата, а просто с его модификацией в соответствии с изменением типа носителя, занял примерно 10 дней, с учетом всех тех-

нических трудностей, которые могли возникнуть в процессе обработки данных ¹.

Следующее решение, эмуляция, состоит в создании программы, имитирующей компьютерную конфигурацию, необходимую для использования данных на современной платформе, после того как оборудование, на котором осуществлялась работа с ними, устарело. Применяя мини-программу, так называемый эмулятор, добиваются того, что имитируется работа системы, которой больше не существует. Если миграция преобразовывает способ хранения фактических данных, то эмуляция применяется на стадии их воссоздания. Этот метод имеет то преимущество, что позволяет сохранять в нетронутом виде первоначальные биты архива или программы, но он не всегда делает возможным воссоздание оптимальных условий для их использования. Например, может быть изменено форматирование страниц или может быть нарушена синхронизация аудио и видео. Хотя, теоретически, такое решение выглядит менее разрушительным, оно все еще требует проведения важных исследований, в частности, осуществления таких исследовательских проектов, которые должны привести к созданию эмуляторов будущего. Кроме того, такое решение вызывает появление новой проблемы, заключающейся в определении того, каким образом разные уровни эмуляторов будут сохраняться в каскаде.

Следующим перспективным решением, над которым работает объединение библиотек и архивных центров по всему миру, является инкапсуляция (пакетирование), известная также как эталонная модель. Этот способ появился в результате последующего развития модели OAIS ², которая была создана для сохранения космических данных Консультативным комитетом по системам космических данных и которая совсем недавно была выбрана в качестве стандарта Международной организацией по стандартам ³.

¹ К. Люповичи. Технические и организационные принципы представления документов.

² OAIS: Open Archival Information Systems — Системы общедоступной архивной информации.

³ К. Хук. Постоянное хранение электронных документов.
<http://www.archivesdefrance.culture.gouv.fr>, 2001

Модель OAIS, первоначально созданная в рамках работы группы по стандартизации, была предназначена для обмена спутниковыми данными, но постепенно превратилась в общую модель, определяющую концептуальные рамки, обеспечивающие идентификацию функциональных компонентов цифровых архивов, и устанавливающую концепции и терминологию для описания моделей конфигурации и данных.

Инкапсуляция, таким образом, оказывается средством объединения вокруг содержания информации, включающей указания, необходимые для декодирования в будущем ее битов любой системой. Она содержит ряд уровней указаний, внешний из которых несет текст, читаемый любыми системами. В этом тексте описывается содержание инкапсулированного элемента и способы его использования, а внутренний уровень содержит требования к программному обеспечению, операционной системе и к характеристикам оборудования, которые должны быть воспроизведены, чтобы обеспечить возможность чтения самого объекта. Инкапсуляция делает содержание и информацию, присоединенную к нему, автономными, и, похоже, сам этот метод становится относительно жизнеспособным, в частности, для долговременного сохранения текстовых файлов. Пока все же существуют сомнения в отношении документов других типов, которые являются жертвами технологического разброса и чрезмерной распространенности новых типов программного обеспечения, систем сжатия и форматов, появляющихся на рынке каждый год.

В конечном итоге, решение, возможно, находится в определении формата сохранения, предписанного учреждениями архивирования, которые возьмут на себя обязательства использовать данный формат на протяжении длительного времени. Однородность форматов позволит тогда следующим поколениям применять автоматическое сохранение.

3.4. Документация

Мы уже упоминали о коренных изменениях в области документации, вызванных цифровыми технологиями. Результатом такого значительного увеличения количества документов, создаваемых в информа-

ционном обществе, является беспрецедентное расширение доступа к ним через их содержание.

- С документарной точки зрения документа, аналоговый вариант был бесполезен. Он ничего не говорил нам о себе, не давал никакой информации. Вся информация о документах должна была создаваться самыми разными специалистами, которые обеспечивали каталогографическую, библиографическую и архивную информацию и предоставляли доказательные данные. Это позволяло формировать структуру коллекции пассивно и создавать «ключ доступа» для читателей. При заданном количестве информации, предназначенной для обработки, которое быстро росло уже в аналоговую эру, это часто превращалось в сизифов труд, никогда не кончающийся, оставляющий большие, не до конца обработанные области, даже в лучших каталогах и, в конечном итоге, дающий читателю весьма скромную информацию.

При использовании цифровых технологий можно не только мгновенно написать резюме документа, не прибегая к помощи хранителя, который должен был бы пойти и найти документ на полке, но его содержание, кроме всего прочего, напрямую понятно компьютеру, а сам первичный документ, с точки зрения информационного поиска, это не более, чем добавление последовательности ключей доступа. Такие важные достижения определили судьбу поисковых систем.

Подобные возможности цифровых технологий в настоящее время используется только применительно к текстовой информации. Но уже существуют несколько новаторских исследовательских программ, которые делают возможным преобразовать произнесенное слово в текстовую информацию, даже в такой информационно избыточной области, как радио. Существует также некоторое число исследовательских программ по анализу неподвижных и движущихся изображений, и остается только надеяться, что в один прекрасный день огромные базы знаний приблизятся к человеческой семантике и нашей способности распознавать, обозначать, приводить в согласие концепции и устанавливать между ними ассоциативные связи.

- Проблема, таким образом, больше не в том, что документы немногословны, а скорее в том, что существует избыток, чрезмерное избытие информации. И каждый из нас обеспокоен тем, что большин-

ство ответов на запросы, посылаемые поисковым системам, становится невозможно использовать. Это явление привело к тому, что основатели Web советуют пользоваться семантической Web ¹, в которой данные будут более структурированы, причем не только данные, относящиеся непосредственно к документу, но и данные, связанные с ним (известные как метаданные). Неизбежное развитие потребует большего участия со стороны производителей данных, с точки зрения их качества.

Не касаясь деталей, необходимо упомянуть об интенсивной деятельности в области стандартизации, что влияет на процесс структурирования документа. Текстовые документы сначала были крайне структурированы с использованием формата SGML; формат HTML, стандартный формат Web, был уже не так сильно структурирован, поскольку это наложило бы слишком много ограничений. Но развитие Web и распространение в настоящее время стандарта XML опять возвращает нас к этим проблемам. В области движущихся изображений, стандарты MPEG-7 и MPEG-21 в значительной степени предназначены для стандартизации различных типов метаданных (документальной, правовой и технической информации, данных об использовании и т. п.), которые присоединяются к данным содержания и остаются прикрепленными к содержанию, расширяя его на протяжении всего времени его существования.

3.5. Пилотные проекты для обеспечения постоянного сохранения

К началу 1990-х гг. цифровые технологии проникли во все формы выражения и человеческого творчества. С этого времени программы экспериментов и исследований, которые прежде обычно выполнялись группами научно-исследовательских и архивных учреждений и библиотек, начали посвящаться изучению вопроса о том, каким образом новая ситуация оказывает влияние на выработку политики в области архивирования и сохранения.

¹ Т. Бернерс-Ли. Семантическая Web. Доклад, представленный на Международной конференции по World Wide Web, Амстердам, 2000.

Пионером в этой области был Архив Интернета, американский фонд, учрежденный в 1996 г. в качестве частной, некоммерческой кампании. Он занимался коллекционированием всех страниц Web, находящихся в свободном доступе по всему миру. Эта коллекция насчитывает более 10 миллиардов страниц, что в пять раз больше количества документов, хранящихся в библиотеке Конгресса. В октябре 2001 г. фонд организовал «машину времени», чтобы обеспечить свободный доступ к его данным с помощью Web ¹.

Область, которая должна была исследоваться в ходе этой работы — изучение практики и потребностей, состояние искусства с точки зрения инструментов, оценка технических и процедурных решений, разработка моделей — была и остается чрезвычайно обширной. И хотя первые результаты этой работы уже появляются, все еще остается много вопросов и неопределенностей, с которыми сталкиваются при выполнении программ исследований и экспериментов, а также при использовании этих результатов в конкретной практической деятельности.

Первые шаги многочисленны и каждый заслуживает рассмотрения, чтобы вынести уроки и максимум преимуществ из приобретенных знаний. Мы просто обратим внимание на некоторые пилотные программы из-за их институционального и научного положения, в частности, потому что они обращались к международной экспертизе и потому, что их работа получила широкое одобрение, что обуславливает желание изучить их опыт.

- **Pandora (Preserving and Accessing Networked Documentary Resources of Australia** — сохранение и доступ к сетевым документальным ресурсам Австралии) ² представляет собой одну из первых широкомасштабных программ, которая была поддержана австралийскими властями. Целью программы было внедрение системы архивирования для электронных публикаций и австралийских Web-сайтов. Вслед за подготовительной стадией, которая была начата в 1996 г. Национальной библиотекой Австралии, был опубликован ряд рекомендаций. Они, в частности, определяли принципы и методы отбора и способы организации архива и управления им.

¹ <http://www.archive.org>

² <http://pandora.nla.gov.au/index.html>

Публикация в них определяется в самом общем виде: все опубликованное в Интернете рассматривается как публикация, причем только внутренняя документация решительно исключается из этого числа. Сайты, отобранные для сохранения, должны иметь отношение к Австралии или к рассмотрению вопросов, имеющих большую важность для страны, и должны быть созданы австралийцами.

Отбор производится на основании содержания, а ранжирование по приоритету осуществляется для публикаций, которые являются официальными и имеют долговременную ценность для исследования.

Вторая стадия работы, названная «PADI»: (Preserving Access to Digital Information — сохранение доступа к цифровой информации) расширила число партнеров за счет привлечения к проекту широкого круга международных экспертов и австралийских государственных библиотек, наряду с Национальным кино- и звуковым архивом, Национальным музеем Австралии, профессиональными ассоциациями, работающими в области ИКТ, исследовательскими центрами, австралийскими университетами и т. д. Целью этой стадии работы является создание и воплощение национальной модели общего архива, включающего электронные данные и австралийскую часть Web.

Кроме того, Государственный архив и Национальный архив Австралии расширили задачу управления электронным архивом, включив в него правительственные Web-сайты (общественные и сайты внутренних сетей) и разработали принципы управления с учетом передового опыта. Национальный архив привлек внимание к тому факту, что документы, находящиеся на Web-сайтах не всегда оцениваются как архив. Такие сайты, тем не менее, должны быть строго организованы. К тому же в мире Интернета идентификация и организация документального материала включают принятие ответственности и соблюдения определенных процедур.

- **The PRESERV programme**¹: эта программа выполняется Группой исследовательских библиотек — Research Libraries Group (RLG), Федерацией цифровых библиотек — Digital Library Federation (DLF) и Он-лайнovým компьютеризованным библиотечным центром — On-line Computer Library Center (OCLC). Эта программа призвана следить за различными исследованиями, обзорами и экс-

¹ <http://www.oclc.org/research/preserv>

периментами, проводимыми в области библиотечного дела. В настоящее время библиотеки сталкиваются с целым рядом проблем, которые вызваны, с одной стороны, увеличением объемов электронных публикаций, а с другой — с необходимостью сохранять цифровые коллекции, за которые они несут ответственность. В ходе выполнения программы исследуются, как с методологической, так и с практической точек зрения, условия архивирования электронных публикаций, цифровых архивов и документов, полученных в результате проведения политики, направленной на перевод в цифровую форму. Она пытается, в частности, определить отличительные черты цифрового архива относительно целей исследования, прояснить концепцию коллекционирования однородных данных, а также вопросы сертификации, структурирования метаданных и использования метаданных в рамках политики долговременного сохранения.

- **NEDLIB (Networked European Deposit Library** — Сетевая европейская депозитная библиотека) ¹ является программой, которая поддерживалась Европейской комиссией с 1998 по 2001 г. Она объединяет восемь крупнейших европейских библиотек: Королевскую библиотеку Нидерландов, Национальные библиотеки Франции, Германии, Норвегии, Португалии и Швейцарии, библиотеку Флоренции и библиотеку Хельсинкского университета. В программе также принимали участие издательства «Эльсевьер Сайнс», «Клювер Академик» и «Шпрингер Ферлаг». Программа предназначалась для создания модели функциональных и технических спецификаций, имеющих отношение к электронным документам, опубликованным на физических носителях или в Web. Она рассматривала все функции, от коллекционирования посредством создания опытного образца, документации и каталогизирования до сохранения, включая полный обзор методов переноса и эмуляции.
- **The CEDARS programme** ² разрабатывалась консорциумом британских (Оксфорд, Лидс и Кембридж) и Американской (Мичиганский университет) университетских исследовательских библиотек. Учрежденная в апреле 1998 г., она предназначалась для исследова-

¹ <http://www.kb.nl.coop/nedlib>

² <http://www.leeds.ac.uk/cedars>

ния вопросов, касающихся приобретения электронных публикаций, их сохранения, описания и, наконец, доступа к ним. Целью этой программы было предоставить в распоряжение общинных библиотек специальное руководство. В нем должен был бы подробно описываться, и с методологической и с эксплуатационной точек зрения, передовой опыт по объединению в единое целое существовавших коллекций и собрания текстов в цифровом виде. В октябре 1999 г. в рамках программы была выдвинута новая инициатива, целью которой было дальнейшее изучение проблем, возникающих в связи с методами эмуляции. Эта новая программа, названная CAMILEON (Creative Archiving At Michigan and Leeds Emulating the Old on the New — создание архивов в Мичигане и Лидсе с помощью имитации старого на новом), должна была продолжаться до сентября 2003 г. В рамках этой программы проводилось исследование того, как возможности эмуляции, применяемой в качестве метода сохранения, могут противостоять устареванию систем, с помощью которых создавались цифровые данные, и обеспечивать возможность доступа к содержанию в условиях использования той конфигурации, при которой эти данные первоначально создавались. Цель программы заключалась в том, чтобы разработать и испытать механизмы эмуляции и выработать ряд рекомендаций, касающихся как миграции, так и эмуляции.

- **The InterPARES programme (International Research on Permanent Authentic Records in Electronic Systems)** — международные исследования по проблемам долговременных аутентичных записей в электронных системах)¹ является в настоящее время одной из самых важных исследовательских и разработческих программ, посвященных долговременному хранению цифровых данных. Она объединила группы исследователей, архивные учреждения, производителей содержания и изготовителей из более чем двадцати стран и осуществляется под руководством Университета Британской Колумбии, Канада. Эта программа проводилась в два этапа: InterPARES-1 началась в 1999 г. и закончилась в 2001 г. публикацией теоретических моделей и методологических разработок в этой области. За этим последовала InterPARES-2, которая, как планиро-

¹ <http://www.interpares.org>

валось, должна была продолжаться до конца 2006 г. Ее целью было определение и анализ физических и интеллектуальных параметров, влияющих на сохранение цифровых данных, а также идентификация новых методов и процедур архивирования и сохранения, полученных в результате исследования. Программой предусматривалось предложение стандартов, организационных принципов и рекомендаций по использованию, которые должны быть внедрены правительствами, учреждениями культуры и наследия и производителями, для обеспечения сохранности цифровых данных.

Наряду с этими программами международного сотрудничества, многие национальные учреждения науки, культуры и наследия взяли на себя обязательства провести исследования и содействовать внедрению методов и средств архивирования цифровых данных, находящихся в сфере их ответственности. Так, например, Австралийская Национальная библиотека объединила усилия с Техническим университетом Вены для разработки системы коллекционирования и сохранения цифровых данных; Национальная библиотека Китая осуществляет программу учреждения цифровой библиотеки; в Нидерландах университеты Амстердама, Тилбурга и Твента проводят исследования и осуществляют разработки в сотрудничестве с IBM, в значительной мере основанные на модели OAIS; главные учреждения культуры и науки в Южной Африке исследуют осуществимость учреждения цифрового архива этого южно-африканского государства. В Тайване девять национальных институтов задействованы в проекте, связанном с государственными документами в цифровой форме; в Соединенных Штатах Библиотека Конгресса отвечает за проведение и выполнение Национальной цифровой программы, стоимостью 100 миллионов долларов. В Канаде канадская информационная сеть по вопросам наследия, при поддержке федерального правительства, содействует тому, чтобы культурные и образовательные учреждения имели возможность пользоваться руководствами, советами и ресурсами при осуществлении деятельности по переводу в цифровую форму и сохранению своих коллекций. В Европе осуществление различных проектов с получением конкретных результатов проводятся в Норвегии, Финляндии, Швеции, Великобритании и Франции. Эти инициативы в значительной степени поддерживаются Европейским Сообществом, Совет которого

опубликовал в мае 2002 г. резолюцию «о сохранении цифрового содержания»¹.

Как можно увидеть, общество и организации постепенно становятся все более обеспокоены угрожающей им потерей памяти, хотя универсальное лекарство, которое помогло бы им справиться с этой проблемой, пока не найдено. Именно поэтому представляется необходимым продолжать исследования и осуществлять разработки в этой сфере, организовывать сотрудничество и взаимный обмен знаниями и средствами. Проблемы, связанные с новым наследием, вынуждают нас признать, что новое наследие распространяется во всемирном масштабе и его сохранение требует также всемирного участия.

В то время как не подлежит сомнению тот факт, что большие успехи уже достигнуты на пути долговременного сохранения цифровых документов, все еще остаются большие темные пятна, связанные с условиями, гарантирующими долговременность содержания и доступ к нему.

Так, сохранение цифрового наследия в настоящее время остается неизвестной сферой для многих учреждений. Тогда как ответственность в этой области официально возлагается на них, им еще только предстоит адаптировать свои организационные структуры и перераспределить рабочую силу для выполнения этой задачи.

В самом деле, нематериальная природа цифровой продукции заставит их освоить новые способы сохранения наследия, заново определить новые процедуры и методы коллекционирования, хранения, индексирования и предоставления их коллекций, которые будут в большей мере соответствовать киберсодержанию, создаваемому в условиях величайшей технологической нестабильности.

Таким образом, очень важно, чтобы механизмы, обеспечивающие сохранение памяти о документе, разрабатывались одновременно с созданием физической среды для организации канала передачи данных.

¹ <http://europa.eu.int>