

Перевод в цифровую форму для сохранения и доступа

Растущее беспокойство о сохранении многих типов документов вместе с желанием улучшить доступ к информации, которую они содержат, привели к резкому росту использования цифровых технологий архивами и библиотеками во всем мире.

Причины, вызывающие использование цифровых технологий, могут быть самыми разными. Некоторые учреждения хотят открыть свои коллекции более широкой аудитории и рассматривают оцифровывание и Интернет как средства достижения желаемой цели. Другие учреждения рассматривают оцифровывание как возможность предоставить доступ к хрупким документам не причиняя им вреда. Третья группа владеет документами, которые приближаются к пределу срока службы и оцифровывание, таким образом, является методом сохранения информации для будущих поколений.

Самая большая опасность заключается в поспешности оцифровывания. Процесс оцифровывания не ограничивается просто покупкой некоторой аппаратуры, подсоединением ее и началом работы. Необходимо четко себе представлять, для чего делаются цифровые копии документов.

- Делаются ли цифровые копии для доступа или для сохранения, а может быть и для того, и для другого? Это будет определять качество требуемых копий.
- Какие приоритеты будут установлены при производстве цифровых копий? Будут ли в первую очередь копироваться документы, находящиеся под угрозой исчезновения, или приоритет будет отдан тем документам, которые пользуются наибольшим спросом?

- Какая требуется скорость доступа? Нужна ли предполагаемая необходимость очень быстрого доступа, которую дает матричная дисковая подсистема (несколько секунд), средняя скорость доступа накопителя на магнитной ленте (несколько минут) или нужен медленный доступ, который обеспечивают носители, размещенные на традиционных библиотечных полках (несколько часов).

- Требуется ли доступ только на рабочем месте внутри учреждения или услуга предлагается "за пределы его стен"?

До тех пор пока эти и другие основные вопросы не будут тщательным образом обсуждены, а ответы на них согласованы, не следует принимать решение о необходимом оборудовании. Вполне может оказаться, что будет принято решение довольствоваться аналоговой технологией, такой как микрофильм, носитель, который еще не исчерпал своих возможностей. Микрофильм может предоставить прекрасного качества копию оригинальных документов. Эта технология хорошо опробована и все проблемы известны. Основным ограничением является то, что доступ может быть получен только при посещении центров, которые имеют копии пленок. Если принимается решение, что требуются цифровые копии материалов, которые ранее микрофильмировались, вовсе необязательно повторять весь процесс цифровой фиксации, поскольку существует оборудование, которое, при условии хорошего качества микрофильмированных изображений, может полуавтоматически перевести их в цифровую форму.

Как только получены ответы на основные вопросы, можно принимать реше-



Джордж Бостон

Член Технического комитета Международной ассоциации звуковых и аудиовизуальных архивов (IASA),
Докладчик подкомитета по технике Программы "Память мира",
Член Аудиотехнического общества (AES)

ние. Первым и самым очевидным является вопрос: "Какой выбрать формат внешнего устройства хранения?" Формат, выбираемый многими учреждениями, приступающими к оцифровыванию своих коллекций, это компакт-диск одноразовой записи или DVD. Чистые диски дешевы, записывающее устройство может находиться в стандартном ПК и существует также большой выбор программного обеспечения для работы с ним. Кроме того, базовые технологии являются широко распространенными и понятными. Записанные диски могут либо храниться на полках и выдаваться обычным образом, либо в устройстве автоматической смены дисков, позволяющем пользователю быть более самостоятельным в доступе к информации. Существует, однако, скрытая опасность в использовании CD и DVD для долговременного хранения данных. Когда впервые появились диски одноразовой записи, чистые диски стоили дорого, но были высокого качества. По мере того, как использование дисков распространялось, цены снизились до одной десятой первоначальной стоимости, но вместе со снижением цены, произошло и снижение надежности. Должен быть установлен строгий режим проверки только что записанных дисков и регулярные проверки хранящихся дисков, если необходимо обеспечить сохранность данных.

Лучшим решением является использование компьютерной системы хранения. Меньшие системы, требующие порядка 2 терабайт (2000 гигабайт) памяти обычно основаны на обслуживающих устройствах с матричной дисковой подсистемой. Они обеспечивают быстрое время доступа, но не могут предложить большого объема памяти по относительно низкой цене.

Если требуются большие системы хранения, предпочтительными могут оказаться кассеты с магнитной лентой, но заплатить придется более медленным доступом. Эти устройства используются большими генераторами данных, которые к тому же требуют низких уровней ошибки в сохраняемых данных. Пользователи таких систем – это банковский сектор и исследовательские институты, включая, среди прочих, космические агентства и правительственные структуры.

Более медленный доступ, обеспечиваемый с помощью лентопротяжных устройств, вызвал появление систем, которые сочетают магнитную ленту с жесткими дисками, и используются внутри системы управления иерархической памятью (HSM). В то время как все файлы хранятся на магнитных лентах, файлы, которые регулярно используются, продублированы на жестком диске, чтобы обеспечить быстрый доступ. Когда спрос на использование файла с жесткого диска идет на убыль, он может быть заменен другим, более популярным файлом.

Когда формат сохранения выбран, следующей проблемой становится стоимость. Стоимость процесса цифровой фиксации документов складывается из двух составляющих: стоимости оборудования и стоимости человеческих ресурсов. Первая включает оборудование, необходимое для сбора, обработки и хранения данных, которые должны быть переведены в цифровую форму – текстов, звука или изображений. Здесь необходимо заметить, что стоимость сканеров, персональных компьютеров, программного обеспечения и цифровых запоминающих устройств устойчиво снижается и, похоже, будет продолжать падать. Вторая составляющая связана со стоимостью персонала

для подготовки и осуществления процесса цифровой фиксации, а также для документирования работы. Эта стоимость, в лучшем случае, не меняется. Стоимость персонала, занимающегося подготовкой документа и работой с аппаратурой, в своей основе, та же самая, что и при микрофильмировании. Действительно, большое число учреждений делают цифровые копии и микрофильмы параллельно – и общая стоимость складывается из затрат на одну группу, осуществляющую подготовку (более дорогая часть) и две группы, работающие с аппаратурой (более дешевая часть).

Процесс цифровой фиксации требует также ответа на два не технических вопроса. Первый – это авторское право. В большинстве стран материалы, находящиеся под действием авторского права, не могут копироваться архивами без согласия владельцев авторского права – даже с целью сохранения информации. В зависимости от материала, который должен быть переведен в цифровую форму, это может быть большим или маленьким препятствием.

Второй не технический вопрос касается метаданных, которые часто определяются как "данные о данных". В аналоговом мире, это детальное и специфическое расширение обычного процесса каталогизирования. Однако, будучи связанными с цифровыми коллекциями, они становятся необходимой частью использования и контроля хранящейся информации. Набор метаданных для сохранения является оператором информации, который, как предполагают, потребуется для управления сохранением цифровой коллекции. Он дополняет и группирует описательные, управляющие и структурные наборы метаданных, собранных для цифровой коллекции. Сохранение

метаданных будет основным компонентом сохранения любой цифровой коллекции и управления ею и должно предусматривать поддержку будущих стратегий сохранения.

Метаданные могут сохраняться внутри ресурса, который они описывают (например, форматы файлов, которые поддерживают описательные заголовки), отдельно от ресурса (например, внешний каталог) либо отдельно, но быть связанными с ресурсом (например, файл, связанный с цифровым объектом в структуре репозитория). Каждая стратегия имеет свои достоинства и недостатки. Возможно, а фактически, желательно, использовать эти стратегии параллельно.

Создание цифровых копий оригинальных документов, однако, не свидетельствует об окончании процесса. Так же как и аналоговые носители информации, цифровые носители нуждаются в обслуживании, чтобы быть уверенными, что данные остаются в исправном состоянии и точными. Кроме того, технологическое устаревание аппаратуры, нужной для доступа к информации, потребует регулярной модернизации и периодической миграции данных на новые системы. Будучи цифровыми, такие миграции не должны требовать дорогостоящего наблюдения человеком.

Более подробная информация может быть найдена в Руководстве по производству и сохранению цифровых объектов, под редакцией Кевина Брэдли, ISBN 8799030918. Это руководство рекомендовано подкомитетом по технике Программы "Память мира" и может быть заказано в Интернете на сайте Международной ассоциации звуковых и аудиовизуальных архивов:

<http://www.iasa-web.org/tc04/index.htm>.