

УСТНАЯ ИСТОРИЯ В ЦИФРОВОМ АРХИВЕ: ПРОБЛЕМЫ И ВОЗМОЖНОСТИ

Юозас Маркаускас

Резюме

Среди руководителей организаций или специалистов, не имеющих специфического опыта хранения коллекций иногда преобладаем мнение, что оцифровывание документов и обеспечение к ним доступа через интернет или при помощи иных электронных средств – это несложный способ решить все проблемы, возникающие при хранении традиционных документов. Однако цифровой архив – это не только оцифровывание документов и оборудование хранилищ информации большего или меньшего объема. Для того чтобы обеспечить долгосрочную защиту документов и доступ к ним, необходимо придерживаться ключевых правил, охватывающих сохранение целостности информации, обеспечение доступности цифровых документов и возможности ими пользоваться, создание и поддержку организационной инфраструктуры. Эти правила особенно важны для сохранения цифровых видео и аудиозаписей, которые так важны для фиксации устной истории.

Цифровой архив обычно состоит из оцифрованных традиционных аналоговых документов и материалов, созданных в цифровой среде (англ. born digital). К первой группе относятся копии документов, имеющих физическую форму, созданные при помощи сканеров, цифровых камер или другого оборудования: записи, рукописи, анкеты опросов, звуко- и видеозаписи на магнитных пленках и т. п. Ко второй группе относятся цифровые документы, у которых нет оригиналов в аналоговой форме и которые были созданы при помощи цифрового оборудования – фото и видеокамер, диктофонов, компьютерных программ. Для хранения документов обоих типов в электронной среде используются по сути те же самые форматы и стандарты файлов.

Для фиксации устной истории подходят те видео- и аудиозаписи, которые в некоторых случаях могут передать больше информации и контекст ее предоставления нежели текстовые документы. Несомненно то, что текстовые и аудиовизуальные документы могут прекрасно дополнять друг друга. Однако аудиовизуальные носители информации по своей специфике выделяются среди других цифровых документов, в которых хранятся тексты или которым не свойственно диахронное (англ. time-based) содержание.

Среди руководителей организаций или специалистов, не имеющих специфического опыта хранения коллекций иногда преобладаем мнение, что оцифровывание документов и обеспечение к ним доступа через интернет или при помощи иных электронных средств – это несложный способ решить все проблемы, возникающие при хранении нецифровых коллекций. Однако это не совсем так, поскольку цифровой архив – это не только оцифровывание документов и оборудование хранилищ информации большего или меньшего объема. Для обеспечения долговременной защиты цифровых документов и доступа к ним необходимо приложить немало усилий и такие усилия должны иметь непрерывный характер. По сравнению с традиционными архивами, эта область

отличается разнообразием стандартов и технологий, отсутствием единой практики и междисциплинарностью.

Для того чтобы сравнить обычный и цифровой архив, процесс сравнения можно начать со сравнения хранимых в них документов. Традиционные документы можно прочитать просто с помощью глаз, они не требуют какого-либо оборудования – даже микроплёнку можно просто поднять против солнечного света и текст можно будет прочитать при помощи самого обыкновенного увеличительного стекла. Для сохранения таких документов достаточно создать хорошие физические условия: контроль температуры и климата, противопожарная безопасность. Кроме того, на сегодняшний день накоплен большой опыт хранения таких материалов, выполняемые процедуры по сути не меняются на протяжении многих лет.

В то же самое время, одно из отличий цифровых документов – это то, что они нематериальны, для их фиксирования необходим физический носитель информации (жесткий диск, компакт-диск, USB-накопитель), который не имеет непосредственной связи с хранимым содержанием, т. е. то же самое содержание можно записать практически на любой носитель. Кроме физического носителя, для хранения содержания также необходим логический формат хранения данных (например, MS Word, Pdf или Jpeg). Как носитель, так и форматы стремительно устаревают (англ. *obsolescence*). Не всегда понятно, какое техническое или программное оборудование необходимо для просмотра того или иного документа – это особенно характерно для «устаревших» документов. Количество цифровых документов стремительно растёт – намного быстрее, чем количество традиционных документов.

Серьезную проблему представляет то, что в электронной среде нет четко определенного единого формата для хранения информации.

По сравнению с текстовыми документами или фотографиями, для видео- и аудиозаписей требуются хранилища с очень большой вместимостью.

Таким образом, для хранения традиционных документов достаточно создать надлежащие условия хранения и быть готовыми справиться с возможными стихийными бедствиями. Для хранения цифровых файлов этих мер недостаточно, поскольку в данном случае существует больше угроз, по причине которых информация может быть утеряна.

Первая угроза – это потеря битов (англ. *bit rot*), когда по причине различных обстоятельств теряются «единицы» и «ноли», составляющие основу любой цифровой информации.

Вторая угроза – устаревание (англ. *obsolescence*) программного обеспечения. Многие ли из нас сегодня имеют возможность просматривать документы, сохраненные в программе PageMaker или первых версиях MS Word? Возможно, эти файлы можно тем или иным образом открыть, однако останутся ли в них литовские символы? Будет ли сохранен оригинальный формат текста?

Устаревание технического оборудования ограничивает возможности доступа к документам, сохраненным на старых устройствах. Легко ли будет найти устройство, подходящее для чтения 5-дюймового диска или магнитно-оптического диска? Вероятно, даже если мы и найдем такое оборудование, оно не будет совместимо с сегодняшними операционными системами, да и сами данные могут быть с утерянными битами.

Также возможны поломки носителей информации. После хранения на протяжении 7–9 лет, около 10 % данных, записанных на компакт-дисках или DVD дисках, становятся недоступными. Международные ассоциации архивистов единогласны в том, что ни компакт-диски, ни DVD диски не подходят для долговременного хранения данных. Если в большинстве домашних фонотек или видеотек диски с таким сроком хранения могут использоваться, это еще не означает, что все данные прекрасно сохранились. Дело в том, что большинство проигрывателей могут воспроизвести видео- или аудиозапись и недостатки будут не всегда заметны. Однако это не приемлемо в контексте хранения данных.

Серьезную угрозу для сохранения цифровых документов представляют ненадлежащие метаданные или недостаток таких данных. Никакой пользователь не может найти неописанные документы. В отсутствие метаданных, тысячи часов видеоархивов представляют из себя всего лишь терабайт или большую кучу битов, и только благодаря метаданным эти часы могут стать ценным источником информации.

Документы необходимо описать даже в тех случаях, когда может показаться, что в этом нет никакой надобности. В бостонском архиве национальной вещательной компании США хранится фотография, на которой изображен студент, принимающий участие в акции протеста.

Только благодаря тому, что она была надлежащим образом описана, сегодня мы можем без труда увидеть, как выглядел теперешний президент США в свои студенческие годы.

Для того чтобы создать надежный цифровой архив, необходимо следовать по крайней мере трем правилам:

Обеспечение целостности информации и предотвращение потери информации.

Обеспечение постоянного доступа к хранимому цифровому файлу и пригодности файла к использованию.

Создание организационной инфраструктуры, ответственной за работу цифрового архива. Это включает в себя физических или юридических лиц и применимые правила.

Для выполнения первого правила необходима продуманная стратегия хранения данных. Необходимо дублировать хранимые данные – лучше всего использовать физические носители разных типов, а также удаленные друг от друга географические точки. Необходимо постоянно осуществлять проверки целостности хранимых данных и

своевременно устанавливать появившиеся проблемы. С совершенствованием технологий необходимо своевременно обновлять носители, на которых хранятся данные.

Одним из ключевых факторов, определяющих доступность цифровых данных и возможность их использовать, являются метаданные. При создании метаданных важно стремиться к тому, чтобы они отвечали на следующие вопросы:

Где находятся данные?

Что входит в эти данные?

Кто их создал?

Могу ли я их использовать?

Как их надлежащим образом показать?

Как их сохранить?

Откуда они?

Несмотря на то что это не всегда просто по причине ограниченности финансовых и технических ресурсов, для хранения всего цифрового содержания следует выбирать самое высокое качество сохранения данных и самую последнюю версию формата файла.

Не рекомендуется использовать какие-либо проприетарные форматы файлов. Для просмотра или создания таких форматов зачастую необходимо дополнительное платное программное обеспечение, они не получают широкого распространения, ограничен выбор программного обеспечения с открытым кодом для работы с ними.

С выходом новых форматов или версий файлов, необходимо обеспечить миграцию хранимых данных.

В архивах должна быть продуманная стратегия хранения. Такую стратегию несложно подготовить по созданной НАСА модели OAIS, получившей широкое применение во всем мире.

Для обеспечения долгосрочного хранения и доступа к документам необходимо использовать стандарты – например, схемы описания метаданных. Стандартизированные метаданные позволяют упростить процессы обмена данными.

Важно вести постоянное наблюдение за технологическими изменениями, чтобы решения по обновлению носителей или миграции данных были актуальными и своевременными.

Эти правила следует выполнять как большим учреждениям, так и частным архивам. Для того чтобы им следовать не всегда нужны большие инвестиции – достаточно обеспечивать надлежащий подход и найти самые приемлемые инструменты.

Давайте не будем рисковать нашей цифровой коллективной памятью и не будем разбрасываться возможностями воспользоваться преимуществами цифровых технологий.