

PDF/A – универсальный формат для хранения информации

АРХИВАЦИЯ. Издательства, рекламные агентства и типографии работают с большими объемами цифровых данных, которые (после отправки заказа в производство) должны архивироваться. Международная организация по стандартизации ISO предложила использовать для архивации цифровых данных формат PDF/A. Весной в Нидерландах Competence Center, международный центр поддержки PDF/A, организовал конференцию, посвященную этому стандарту.

В конце 2005 г. международная организация по стандартизации ISO утвердила в качестве стандарта для архивации информации (ISO 19005-1) формат PDF/A. В настоящее время особенно большой интерес к PDF/A наблюдается в Европе, так как североамериканская ассоциация АИМ пока еще не утвердила этот формат в качестве стандарта.

В настоящее время предлагается большая номенклатура программных инструментов для работы с PDF/A. К сожалению, внедрение нового формата тормозится низким уровнем информированности о

его возможностях и трудностью реорганизации документооборота. В полиграфии ситуация иная – здесь уже много лет успешно используются различные разновидности формата PDF, в частности PDF/X.

АРХИВАЦИЯ. Одной из проблем, с которой сталкиваются крупные промышленные предприятия, государственные учреждения, страховые компании, издательства и архивы, является надежная архивация больших объемов данных. Традиционные способы, например, хра-



В ходе первой международной конференции, посвященной формату PDF/A, пользователи и разработчики обсуждали стратегии и технические аспекты внедрения нового стандарта.

нение информации на бумаге или микрофильмах, безнадежно устарели, так как они не обеспечивают компактности данных, а также возможностей эффективного поиска и передачи информации. Поэтому в последние годы для архивации используются цифровые форматы представления данных, чаще всего – TIFF. Однако формат TIFF в свою очередь имеет существенные недостатки – невозможность поиска текстовой информации без применения средств распознавания знаков (OCR), а также потребность в больших объемах памяти для хранения TIFF-файлов. Формат PDF лишен этих недостатков – файлы имеют компактный размер, возможен поиск по тексту, обеспечивается высокое качество визуализации как графических, так и текстовых данных. По этой причине стандартизировать PDF/A было предложено еще в 2002 г.

ЧТО ТАКОЕ PDF/A? Стандартизированный формат PDF/A является подмножеством формата PDF, обеспечивающим аутентичное представление данных спустя годы и десятилетия после их создания. При этом гарантируется независимость пользователя от разработчиков программного и аппаратного обеспече-



Доклад Томаса Мерца, руководителя компании PDFlib GmbH, был посвящен использованию метаданных.



Бетси Фаннинг из североамериканской ассоциации AIIM была приятно удивлена интересом к PDF/A в Европе.

ния, а также полная сохранность информации.

Спецификация PDF/A определяет содержание PDF-файлов. Так, например, файлы PDF/A обязательно должны содержать информацию о шрифтах и управле-

нии цветом, обеспечивающую корректное представление данных. С другой стороны, в файлы PDF/A не допускается включение элементов, которые могут затруднять визуализацию или предполагать неоднозначную трактовку, например,

«прозрачностью», слоев, встроенных видеоданных и скриптов JavaScript. При этом встроенная защита не должна препятствовать открытию PDF-файлов. Более подробную информацию о спецификациях PDF/A можно найти на Интернет-сайте www.pdfa.org.

А И В. Разработано две версии стандарта PDF/A: 1a и 1b. Если PDF/A-1b отвечает минимальным требованиям к точности визуализации документа, то PDF/A-1a соответствует наивысшим требованиям. В частности, в файлах PDF/A-1a, благодаря использованию тегов (Tagged PDF), сохраняется информация о логической структуре документа. Преимуществом такого решения является возможность извлечения и дальнейшей обработки содержащихся в файле данных. Это весьма актуально, например, для официальной документации, так как обеспечение доступности содержащейся в ней информации законодательно закреплено в США и рекомендовано в странах ЕС.

PDF/A И PDF/X. Нашедший широкое применение в полиграфии стандарт PDF/X имеет много общего с PDF/A, что неудивительно, так как в обоих случаях должна обеспечиваться повторяе-

Сравнение спецификаций PDF/A и PDF/X

	PDF/A	PDF/X
Цель	Долгосрочное архивирование с обеспечением точной визуализации документа (PDF/A-1b) и его структуры (PDF/A-1a)	Формат обмена данными для полиграфии
Информация о цвете	Управление цветом необходимо при использовании аппаратно-зависимых цветовых моделей	Необходимо внедрение выходного профиля
Шрифты	Внедряются в документ (за исключением невидимого текста)	Внедряются в документ
Установки переполнения	Не требуются	Необходимы
Параметры страницы	Не требуются	Необходимы
Прозрачности	Недопустимы	Необходимы
Слои	Недопустимы	Необходимы
Метаданные	Допустимо использование XMP-метаданных	Допустимы
Структура документа	Необходимо в PDF/A-1a	Не требуется
Акции	Недопустимы	Недопустимы
JavaScript	Недопустимы	Недопустимы
Аудио- и видеоданные	Недопустимы	Недопустимы
Цифровые сигнатуры	Допустимы	Допустимы
Защита и кодирование данных	Недопустимы	Недопустимы

мость воспроизведения данных. Документ может отвечать требованиям обоих стандартов, то есть быть совместимым и с PDF/A, и с PDF/X. Конечно PDF/X – это, прежде всего, формат обмена данными, но ведь и информация, подготовленная к полиграфическому воспроизведению, на определенном этапе технологического процесса должна архивироваться.

Спецификации PDF/A и PDF/X различаются, например, определением параметров страниц, которые не требуются в PDF/A, но необходимы в PDF/X. Также есть различия в представлении информации о цвете. Краткое сравнение спецификаций PDF/A и PDF/X ДАНО в таблице.

РАБОЧИЙ ПОТОК PDF/A. В настоящее время предлагается множество программных утилит для работы с PDF/A. Есть пакеты для создания, конвертации, отображения и проверки файлов PDF/A. Обзор программных решений представлен на сайте <http://www.pdfa.org/doku.php?id=pdfa>. К сожалению, несмотря на обилие утилит, многим предприятиям элементарно не хватает знаний для организации документооборота на базе PDF/A.

ЦЕНТР ПОДДЕРЖКИ. В 2006 г. для помощи по внедрению PDF/A в документооборот был создан центр поддержки – PDF/A Competence Center. Техническая рабочая группа центра занимается разработкой инструкций по внедрению стандарта (см. <http://www.pdfa.org/doku.php?id=pdfa:techdok>).

PDF/A Competence Center – независимое международное объединение, созданное с целью содействия внедрению стандарта ISO 19005: PDF/A. Членами объединения являются около 70 организаций, в том числе разработчики программного обеспечения, крупные пользователи и консалтинговые фирмы.

10 – 11 апреля PDF/A Competence Center провел в Амстердаме (Нидерланды) международную конференцию, в работе которой приняли участие около 200 специалистов из Европы, США и Японии. В обсуждении аспектов внедрения PDF/A, помимо экспертов, участвовали и представители крупных пользователей – компаний Airbus, Rhätische Bahn, а также европейского патентного бюро.

ПРОВЕРКА И УТВЕРЖДЕНИЕ. Одним из центральных вопросов конференции стала проверка файлов PDF/A на их соответствие требованиям стандарта. На практике реализуются два способа организации такой проверки: для каждого

файла по отдельности и выборочная проверка отдельных файлов. Первый способ предпочтителен, когда файлы для архива собираются из разных источников, второй можно рекомендовать, когда все файлы создаются в одном месте с использованием одних и тех же программных средств.

В настоящее время на рынке представлен целый ряд программных решений для проверки файлов PDF/A, но, к сожалению, многие из них интерпретируют спецификации PDF/A по-разному. Как результат: файл, утвержденный одной утилитой, может не пройти проверку другой утилитой. Для решения этой проблемы рабочая группа Technical Working Group центра PDF/A Competence Center разработала тестовый программный пакет – Isartor-Suite, – который впервые был представлен на конференции и в последствии будет доступен на сайте www.pdfa.org. Этот пакет включает тестовые документы, с помощью которых можно проверить правильность интерпретации утилитой разных аспектов стандарта.

МЕТАДАнные XMP. Важным аспектом архивации документов является возможность поиска необходимой информации. Формат PDF позволяет сохранять текстовые данные и выполнять контекстный поиск, но существуют и другие возможности. Уже довольно длительное время многие архивы сопровождают каждый документ метаданными, которые содержат дополнительную информацию, упрощающую поиск. Спецификации PDF позволяют включать в файл XMP-метаданные. Этот формат используется в полиграфии и рекомендован к применению различными отраслевыми объединениями, например, консорциумом AdsML для рекламной индустрии и международным советом по прессе и телекоммуникациям IPTC (International Press Telecommunications Council) для газетного производства.

Спецификации PDF/A-1 предусматривают использование XMP для идентификации одинаковых данных и позволяют применять определяемые пользователем метаданные, которые сохраняются в файлы с расширением XMP. Указание на XMP-файл должно содержаться в документе PDF/A (дополнительную информацию можно найти по ссылке <http://www.pdfa.org/de/developer/xmp-metadata>). Все это означает, что при архивировании файлов PDF/A не обязательно использовать ранее определенные шаблоны, но можно задавать метаданные согласно собственным критериям.

ПЕРСПЕКТИВЫ. Одной из ключевых проблем, замедляющих внедрение PDF/A, является сомнение пользователей в том, что этот стандарт принят действительно «на века» и не будет пересмотрен через несколько лет или десятилетий, что затруднит работу с ранее созданными архивами. Тем не менее, в Европе число пользователей PDF/A растет достаточно динамично, чему способствуют, в том числе, усилия центра PDF/A Competence Center.

Ирина Ляйхзенринг

Ирина Ляйхзенринг – сотрудница фирмы Marketing PDFlib GmbH, автор многочисленных публикаций по проблемам электронного документооборота, редактор журнала Mac Profiler.

Программное обеспечение для работы с PDF/A

Решения PDFlib GmbH ■ Мюнхенская компания предлагает комплексные решения для организации документооборота на базе формата PDF/A.

Семейство PDFlib: библиотеки PDFlib, PDFlib+PDI и PDFlib

Personalization Server позволяют реализовать динамическое создание файлов PDF/A-1b и PDF/A-1a с поддержкой XMP-метаданных.

PDFlib Plop: возможна оптимизация совместимых с PDF/A данных, а также внедрение в готовые PDF/A-файлы метаданных.

PDFlib Plop DS: расширенная версия PDFlib Plop, которая позволяет внедрять в PDF/A-файлы цифровые подписи.

PDFlib P-Cos: инструмент для извлечения из PDF-документов XMP-метаданных. Может поставляться как отдельный продукт или как модуль для других утилит.

PDFlib TET: патентованный инструмент для извлечения из PDF-документов текстовой информации.

PDFlib TET PDF I-Filter: основанный на движке TET инструмент для извлечения из PDF-файлов текстовых данных, метаданных, а также меток страниц. Совместим с такими серверными решениями, как Microsoft Office Sharepoint Server, Exchange Server и SQL Server.