

О некоторых способах преобразования электронных документов в формат PDF/A и их последующей проверке на соответствие стандартам PDF/A

В целях выполнения пункта 37 Инструкции о порядке работы с электронными документами в государственных органах, иных организациях, утвержденной постановлением Министерства юстиции Республики Беларусь от 6 февраля 2019 № 19 (далее – Инструкция № 19), организации, являющиеся источниками комплектования государственных архивов, после завершения подготовки проекта электронного документа (далее – ЭД), для которого в номенклатуре дел предусмотрен срок хранения свыше 10 лет (в том числе постоянно) или с отметкой «ЭПК», а также организации, не являющиеся источниками комплектования государственных архивов, при подготовке к подписанию ЭД, предназначенных для направления в государственные органы и подчиненные им (входящие в состав, систему) организации, должны преобразовывать общую часть ЭД (включая приложения) в формат PDF/A-1 или

Необходимость такого преобразования обусловлена тем, что файлы в иных форматах по прошествии длительного промежутка времени могут отображаться некорректно либо не открываться вовсе. В то же время, формат PDF/A стандартизирован Международной организацией по стандартизации (далее – ISO) и установлен в качестве формата для долгосрочного хранения ЭД.

Долгосрочное хранение ЭД в данном формате возможно благодаря тому, что в файлы, составляющие ЭД, внедряется вся информация (шрифты, цветовые профили, изображения и др.), необходимая для его воспроизведения в будущем точно таким же образом с использованием различного программного обеспечения (далее – ПО). Также стандарты на формат PDF/A устанавливают некоторые ограничения для файлов в данном формате: запрет на шифрование файлов, запрет на вставку исполняемого кода, аудио- и видеоконтента.

Поскольку визуально отличить файлы в формате Portable document format (далее – PDF) и PDF/A затруднительно (оба файла будут иметь расширение .PDF), то при подготовке проектов ЭД нужно убедиться в том, что форматами файлов, составляющих ЭД, являются PDF/A-1 или PDF/A-2, как того требует Инструкция № 19.

Для определения наиболее оптимальных способов преобразования файлов ЭД в формат PDF/A и проверки форматов файлов ЭД, полученных в результате преобразования, на соответствие требованиям *стандартов*¹ на формат PDF/A было проведено небольшое исследование.

В исследовании приняли участие известные текстовые процессоры

1

ISO 19005-1:2005 Управление документами. Формат файлов электронных документов для долгосрочного хранения. Часть 1. Использование ISO 32000-1 (PDF/A-1);
ISO 19005-2:2005 Управление документами. Формат файлов электронных документов для долгосрочного хранения. Часть 2. Использование ISO 32000-1 (PDF/A-2);

ISO 19005-3:2012 Управление документами. Формат файлов электронных документов для долгосрочного хранения. Часть 3. Использование ISO 32000-1 с поддержкой вложенных файлов (PDF/A-3).

(бесплатное ПО), **LibreOffice Writer** (бесплатное ПО) и редактор PDF **Adobe**

Основным отличием текстовых процессоров от редактора PDF является то, что в них создаются файлы в отличных от PDF (PDF/A) форматах, после чего возможен экспорт данных файлов в формат PDF (PDF/A). В редакторе PDF мы работаем изначально с файлами в формате PDF.

Исследовав графический пользовательский интерфейс данного ПО, мы установили следующее.

Текстовый процессор Microsoft Word 2016 позволяет экспортировать созданный в нем файл в формат PDF/A-1. Текстовый процессор Hancor Office 2020 позволяет экспортировать созданный в нем файл в формат PDF/A (в какой части PDF/A будет экспортированный файл не указывается). Текстовый процессор OpenOffice Writer позволяет экспортировать созданный в нем файл в формат PDF/A-1. Текстовый процессор LibreOffice Writer позволяет экспортировать созданный в нем файл в формат PDF/A-1, PDF/A-2 или PDF/A-3. Редактор PDF Adobe Acrobat Pro DC 2021 позволяет сохранить созданный в нем файл в формате P

D Для того, чтобы проверить функции экспорта или сохранения файлов в формате PDF/A в данных программах мы использовали ПО **veraPDF** и / веб-сервис **PDF Tools Online – Validate PDF**, предназначенные для проверки файлов в формате PDF/A на соответствие требованиям стандартов на формат PDF/A, а также функцию проверки соответствия в редакторе PDF Adobe Acrobat 1

На основе информации, представленной на официальных сайтах ПО veraPDF и jPDFPreflight, веб-сервиса PDF Tools Online – Validate PDF, редактора PDF Adobe Acrobat Pro DC 2021, мы исходили из того, что данное ПО и веб-сервис действительно проверяют файлы в формате PDF/A на соответствие требованиям вышеупомянутых стандартов.

ПО veraPDF – это продукт объединения таких некоммерческих организаций, как Open Preservation Foundation, PDF Association и Digital Preservation Coalition.

Объединение финансируется Европейской комиссией в рамках проекта Европейского союза *PREFORMA*. Цель данного проекта – решение проблемы внедрения в работу организаций стандартизированных форматов файлов для сохранения цифровых объектов в долгосрочной перспективе, предоставление библиотекам и архивам полного контроля над принятием на хранение и управлением цифровыми объектами в электронных хранилищах [1].

veraPDF – ПО с открытым исходным кодом, лицензия на которое позволяет бесплатно его использовать. К плюсам данной программы можно отнести наличие графического пользовательского интерфейса. Перед установкой VeraPDF на компьютере должна быть установлена Java 8 или более поздней версии. Также есть демонстрационная веб-версия данного ПО, доступная по адресу:

ПО jPDFPreflight – это продукт компании Qoppa Software, которая является членом PDF Association. ПО для работы с PDF, разработанное Qoppa Software, используют компании из рейтинга Fortune Global 500, а также государственные учреждения в Европе и США [2]. Плюсы программы: наличие графического пользовательского интерфейса, демонстрационная версия является бесплатной и неограниченной по времени использования. Единственное ее ограничение – добавление водяных знаков в отчет с результатами проверки.

Веб-сервис PDF Tools Online - Validate PDF – продукт компании PDF Tools, которая насчитывает более 5 000 компаний и организаций в 70 странах среди своих клиентов. Помимо обширной сети клиентов можно отметить тот факт, что компания PDF Tools принимала участие в разработке стандартов ISO на формат

Однако данный бесплатный веб-сервис перестанет работать в начале 2023 года. Вместо него на сайте компании предлагается приобрести ПО **PDF Validator**. К его минусам можно отнести отсутствие графического пользовательского интерфейса. Взаимодействие с программой осуществляется посредством интерфейса командной строки. Поэтому перед работой с данным ПО придется ознакомиться с руководством пользователя.

Редактор PDF Adobe Acrobat Pro DC 2021 – это продукт компании Adobe, которая в 1993 году и представила миру формат PDF, предназначенный прежде всего для отображения полиграфической продукции в электронном виде.

Результаты проверки форматов файлов, полученные с использованием вышеупомянутых инструментов, приведены в таблице ниже.

Условия проверки	ПО				
Заявленная часть формата PDF/A, в которую возможен экспорт файла или в которой возможно его создание		PDF/A (без указания части)		PDF/A-1 PDF/A-2 PDF/A-3	
Часть формата PDF/A выявленная по результатам проверки	1			PDF/A-1 PDF/A-2 PDF/A-3	
Проверка форматов в	Пройдена	Пройдена	Пройдена	Пройдена	Пройдена

Проверка форматов в	Пройдена	Пройдена	Пройдена	Пройдена	Пройдена
Проверка ф	Пройдена	Не пройдена	Пройдена	Пройдена	Пройдена
Проверка форматов в	Пройдена	Файл в формате PDF/A-1 по результатам проверки не соответствовал требованиям стандарта на данный формат	Пройдена	Пройдена	Не пройдена Файлы в форматах PDF/A-1 и PDF/A-2 по результатам проверки не соответствовали требованиям стандартов на данные форматы

Проверка показала, что текстовые процессоры Microsoft Word 2016, OpenOffice Writer и LibreOffice Writer экспортируют созданные в них файлы в заявленные форматы. Текстовый процессор Nancom Office 2020 – в формат PDF/A-1. Редактор PDF Adobe Acrobat Pro DC 2021 создает файлы в заявленных форматах.

Проверка форматов файлов на соответствие требованиям стандартов на формат PDF/A в Adobe Acrobat Pro DC 2021, VeraPDF, jPDFPreflight и PDF Tools Online – Validate PDF прошла успешно для файлов, экспортированных из экспортированный из Nancom Office 2020, не прошел проверку в jPDFPreflight и PDF Tools Online – Validate PDF. Файлы, созданные в Adobe Acrobat Pro DC 2021 не прошли проверку в PDF Tools Online – Validate PDF.

Однако ситуация, при которой различные инструменты проверки файлов в формате PDF/A дают разные результаты, является допустимой, если недостающая информация, из-за которой возникают ошибки соответствия, не используется при визуализации файлов в формате PDF/A [4].

Таким образом, все исследованное ПО позволяет экспортировать или создавать файлы ЭД, как минимум, в формате PDF/A-1. LibreOffice Writer и Adobe Acrobat Pro DC 2021 предусматривают экспорт и создание файлов также и в форматах PDF/A-2, PDF/A-3. По результатам исследования для получения файлов ЭД в форматах, установленных Инструкцией № 19, рекомендуется использовать текстовые процессоры Microsoft Word 2016, OpenOffice Writer и LibreOffice Writer, так как файлы в формате PDF/A, полученные в них, прошли проверку во всех валидаторах PDF/A, рассмотренных в статье. Тем не менее, использование иного ПО (например, Nancom Office 2020, Adobe Acrobat Pro DC 2021) не запрещается.

Для проверки файлов в формате PDF/A можно использовать редактор PDF Online – Validate PDF. При использовании веб-сервисов для проверки форматов файлов следует помнить о том, что они, как правило, не обеспечивают конфиденциальность переданной им информации.

В заключение отметим, что, помимо выполнения требований пункта 37 Инструкции № 19 о преобразовании файлов ЭД в формат PDF/A-1 или PDF/A-2, организациям при внедрении у себя автоматизированных систем документационного обеспечения управления рекомендуется выяснить у разработчиков данных систем информацию о наличии в них функции проверки форматов файлов ЭД на соответствие требованиям стандартов на формат PDF/A-1 или PDF/A-2. Так как в случае отсутствия в упомянутых системах данной проверки существует риск того, что на постоянное хранение будут переданы ЭД в отличных от PDF/A-1 или PDF/A-2 форматах, что может привести к проблемам с воспроизведением данных ЭД в будущем.

Источники:

1. About veraPDF [Electronic resource] // veraPDF. – Mode of access: <https://verapdf.org/home>. – Date of access: 12.07.2022.
2. About Qoppa Software [Electronic resource] // Qoppa Software. – Mode of access: <https://www.qoppa.com/company/about-us>. – Date of access: 04.08.2022.
3. About PDF Tools [Electronic resource] // PDF association. – Mode of access: <https://www.pdfa.org/member/pdf-tools-ag>. – Date of access: 12.07.2022.
4. PDF/A validation tools give different results [Electronic resource] // OPF Knowledge Base Wiki. – Mode of access: <https://wiki.opf-labs.org/display/SPR/PDFA+Validation+tools+give+different+results>. – Date of access: 12.07.2022.

*Илья Дорошевич,
стажер младшего научного
сотрудника отдела методологии
управления электронными
документами БелНИЦЭД*