



Петр Марченко, Андрей Чекатков

УНИФИЦИРОВАННЫЙ ИНФОРМАЦИОННЫЙ ОБЪЕКТ ДЛЯ ОБМЕНА И ХРАНЕНИЯ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Предложен проект унифицированного информационного объекта для обмена и хранения электронных документов, в основе формирования структуры и содержания которого лежит концепция жизненного цикла «records continuum» с XML-инкапсуляцией файлов.

Ключевые слова: архивный электронный документ, долговременное хранение, XML-инкапсуляция.

Сейчас в сфере описывания и управления электронными документами в мире разработан ряд национальных и международных стандартов¹, среди которых есть и базирующиеся на принципе «records continuum»², ориентированном на решение проблемы их долговременного хранения. Эти стандарты отличаются тем, что в начале жизненного цикла архивного электронного документа создается информационный объект, в котором он сохраняется в инкапсулированном виде³. Инкапсуляция используется на практике мировыми авторитетами в сфере электронного документооборота и архивного дела – США и Австралией. В США NARA (U.S. National Archives and Records Administration) принята программа перспективного развития электронных архивов на основе использования POP (Persistent Object Preservation)⁴, в Австралии администрация электронного документооборота штата Виктория использует стандарт PROS 99/007⁵ с аналогичным информационным объектом VEO (Victorian Encapsulated Object). Для этих информационных объектов характерным является следующее:

- электронный документ долговременного хранения, его метаданные и электронная цифровая подпись (ЭЦП) представляются одним информационным объектом;

- XML (eXtensible Markup Language)⁶ используется как формат объединения метаданных, электронного документа и электронных цифровых подписей в один информационный объект.

Учитывая то, что решение проблемы долговременного и постоянного хранения электронных документов для Украины является насущным вопросом в связи с переходом к электронному правительству, авторы предлагают разработку унифицированного информационного объекта (УИО) для обмена и сохранения электронных документов, которая позволит реализовать на протяжении их жизненного цикла следующее:

- поиск и отображение необходимой документальной информации, проверка целостности и аутентичности электронных документов;

- документировать информацию о правах доступа и миграции электронных документов на другие носители информации и форматы данных с течением времени;

- документировать события жизненного цикла электронных документов с целью обеспечения их аутентичности и доступа к информации;

- пользоваться единым стандартизованным информационным объектом для обмена и хранения электронных документов в системах электронного документооборота и электронных архивах.

1. Определение УИО

1.1. Терминологическое определение

В эталонной модели открытой архивной информационной системы OAIS (Open archival information system) ISO 14721:2003⁷ информационный объект определен как совокупность объекта данных с репрезентативной информацией, которая характеризует информационную среду его функционирования. Объект данных представляет собой совокупность данных, содержащих информацию, ради которой он был создан.

Здесь УИО определим как информационный объект, который представляет собой электронный XML-документ (ЭЦП, согласно с украинским законодательством, – **обязательный реквизит электронного документа**), составляющими которого являются файлы электронного документа, метаданные, характеризующие информационную среду его функционирования на протяжении его жизненного цикла, и ЭЦП должностного лица, которая обеспечивает его целостность и аутентичность.

1.2 Требования к модели УИО

Модель УИО должна удовлетворять потребности делопроизводителей, архивистов и пользователей документной информации в выполнении их работы с помощью электронных архивов. Основные функциональные требования к УИО:

- классификация электронных документов;
- пригодность для документирования и использования (создание соответствующих для это-



го метаданных УИО и электронных документов) процессов жизненного цикла во времени;

– импорт и экспорт электронных документов в/из УИО;

– межоперабельность, то есть пригодность для межсистемного обмена;

– масштабируемость, то есть пригодность как для систем электронного документооборота, так и электронных архивов;

– пригодность для администрирования процессов жизненного цикла: смена носителей информации и форматов данных, прав доступа к документной информации, аудита целостности и аутентичности электронных документов;

– наличие объектного структурирования, которое позволяет представить УИО для пользователей в соответствии с их полномочиями относительно действий, которые могут ими выполняться.

В целом эти требования может удовлетворять модель УИО, составными частями которой является электронный документ со всеми его отдельно документированными реквизитами и приложениями (далее – компоненты) и соответствующие метаданные в межоперабельных форматах данных, и которая функционирует в информационной среде, представленной на рис. 1.

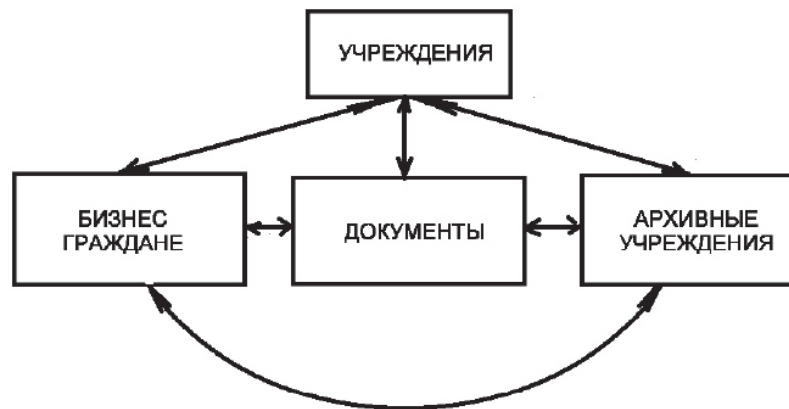


Рис. 1. Информационная среда функционирования УИО.

1.1. Модель УИО

Разработанная нами модель УИО, которая соответствует вышеуказанным требованиям, представлена на рис. 2.

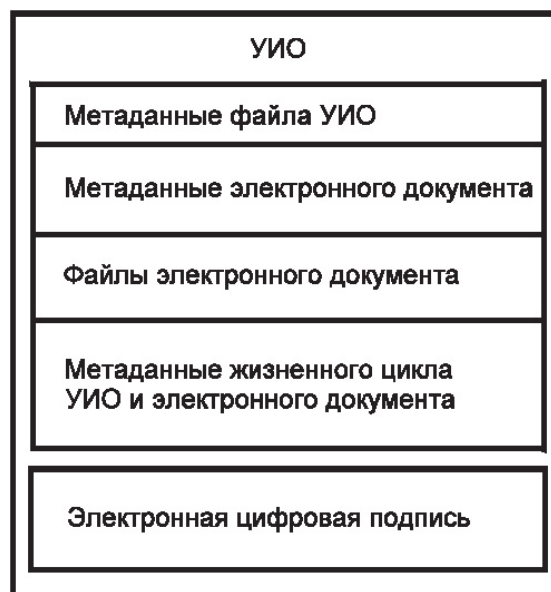


Рис. 2. Модель УИО.



Наиболее приемлемым форматом файла УИО является XML, который позволяет представить УИО обычным размеченным текстом с набором символов UNICODE⁸, имеет XML DOM⁹ объектную модель и возможность импорта, экспорта файлов электронного документа в содержательную часть УИО – инкапсуляцию с конвертированием оригинальных форматов документов в текстовый формат BASE64¹⁰ и наоборот.

Преимущество такого способа хранения электронного документа – уменьшение риска потери метаданных и других его компонентов в сравнении с раздельным способом их хранения.

Метаданные УИО:

– метаданные файла УИО представляют структурированную информацию об авторе УИО, дате создания файла, дате его авторизованного редактирования;

– метаданные электронного документа представляют структурированную информацию об авторе документа, поисковые данные, структуру документа с его компонентами в виде отдельных файлов;

– метаданные жизненного цикла УИО, которые отображают историю событий во времени, а именно: архивное описывание, перемещение в пределах архивного фонда, конвертирование, изменение прав доступа для пользователей, аудит целостности и аутентичности электронных документов.

Файлы электронного документа – это файлы электронного документа, его компонентов и возможных отделенных от файлов электронных документов ЭЦП лиц, которые подписали электронный документ и его компоненты, в текстовом формате BASE64.

Эталонную модель структуры и содержания УИО целесообразно представить в виде XML-схемы¹¹.

1.1. Описание структуры и содержания УИО

Структуру и содержание XML-схемы УИО (англ. UIO) целесообразно для наглядности представить в виде текстового документа, который имеет разделы, подразделы, пункты, подпункты как структурные составляющие. При этом в скобках обозначим следующие опции:

Е – XML-элемент;

А – XML-атрибут;

С – XML-контейнер, то есть XML-элемент, который содержит в себе другие XML-элементы;

О – обязательно присутствующий;

НО – необязательно присутствующий;

П – повторяющийся.

1. **UIO (C,O)** – наименование XML-схемы UIO

1.1. **AccessRights (E,O)** – для метаданных, которые определяют уровень доступа и использования информации UIO

1.2. **AgentUIO (C,O)** – для метаданных учреждения - владельца UIO

1.2.1. **DepartmentUIO (C,O)** – для метаданных подразделения учреждения, ответственного за электронный документооборот

1.2.2. **Address (E,O)** – для метаданных адреса подразделения

1.2.3. **Digital Signature (E,O)** – для метаданных ЭЦП

1.2.4. **AccessHistory (C,O)** – для метаданных истории пользования UIO

1.3. **Record (C,O)** – для метаданных ЭД и инкапсулированных файлов ЭД

1.3.1. **AuthorRecord (C,O, П)** – для метаданных авторов ЭД

1.3.2. **AgentRecord (C,O)** – для метаданных учреждения – автора ЭД

1.3.2.1. **Address (E,O)** – для адреса учреждения

1.3.2.2. **OfficerIPersons (E,O, П)** – для метаданных должностных лиц – подписантов ЭД

1.3.2.3. **Digital Signature (E,O)** – для метаданных ЭЦП

1.3.3. **Persons (C,O, П)** – для метаданных физического лица – подписанта ЭД

1.3.3.1. **Address (E,O)** – для метаданных адреса

1.3.3.2. **Digital Signature (E,O)** – для метаданных ЭЦП

1.3.4. **ClassRecord (E,O)** – для метаданных классификации ЭД

1.3.5. **Rights (E,O)** – для метаданных о праве на доступ и пользование ЭД

1.3.6. **RecordIdentification (C,O)** – для метаданных идентификации ЭД

1.3.6.1. **SearchData (E,O)** – для поисковых метаданных

1.3.7. **RKK (C,O)** – для РКК ЭД

1.3.8. **RecordFiles (C,O)** – для файлов ЭД и внешних файлов ЭЦП

1.3.8.1. **RecordFile (E,O)** – для инкапсулированного ЭД

1.3.8.2. **SignatureFiles (E,O)** – для внешних файлов ЭЦП

1.3.9. **AppendixFiles (E,O)** – для файлов ЭД (UIO) – приложений ЭД

1.3.10. **Resolution (E,O)** – для файлов ЭД «Резолюция»

1.3.11. **RecordkeepingMetadata(E,O)** – для метаданных делопроизводства учреждения



1.4. **Preservation (C,O)** – для метаданных жизненного цикла архивного хранения ЭД

1.4.1. **AggregationMetadata (E,O)** – для метаданных архивного описания

1.4.2. **LocationHistory (E,O)** – для метаданных истории размещения UIO в архиве

1.4.3. **VerifyingSignature (C,O)** – для метаданных проверки ЭЦП, их текущих и предыдущих результатов

1.4.3.1. **VerifyingAgent (E,O)** – для метаданных учреждения, осуществившего проверку ЭЦП

1.4.3.1.1. **AddressAgent (E,O)** – для адреса учреждения

1.4.3.1.2. **OfficerPerson (E,O)** – для метаданных должностного лица

1.4.4. **VerifyingResultat (E,O)** – для результатов проверки ЭЦП

1.4.5. **ConversionHistory (C,O)** – для метаданных конвертирования ЭД UIO и конвертируемых файлов ЭД

1.4.5.1. **ConversionAgent (E,O)** – для метаданных учреждения, осуществившего конвертирование

1.4.5.1.1. **AddressAgent (E,O)** – для метаданных адреса учреждения, осуществившего конвертирование

1.4.5.1.2. **OfficerPersons (E,O)** – для метаданных должностного лица, ответственного за конвертирование

1.4.5.1.3. **Digital Signature (E,O)** – для метаданных ЭЦП

1.4.5.1.4. **ConversionResultat (E,O)** – для файлов преобразования

1.5. **EXT (C,O)** – другое

2. Межоперабельность UIO

Межоперабельность UIO обеспечивается использованием для его создания языка разметки XML и кодированием символов стандарта UNICODE, а также нормированным использованием UIO для документирования: созданием электронных документов с использованием XML-схемы UIO как его эталонной модели с использованием форматов данных, которые имеют следующие характеристики:

- массовость использования;
- открытая спецификация;
- качественное отображение документной информации.

К таким форматам возможно отнести следующие:

– для текстово-графических документов – PDF/A (формат спецификации ISO 19005–1);

– для графических документов – TIFF (формат спецификации ISO 12639);

– для аудиодокументов – FLAC (формат спецификации FLAC Format Specification, 2008);

– для видеодокументов – Matroska (формат спецификации The standard of multimedia container formats Matroska (04/11/2011)).

3. Схема внедрения UIO на практике

1. Разработка XML-схемы UIO и программного средства создания UIO с использованием XML-схемы с реализацией импорта и экспорта файлов электронного документа и его компонентов, условное название – «Uniform_Information_Object».

2. Аprobация XML-схемы UIO и программного средства «Uniform_Information_Object» в условиях реального электронного документооборота системы государственных архивов Украины.

3. Уточнение XML-схемы UIO и программного средства «Uniform_Information_Object» по результатам апробации.

4. Разработка ГСТУ «Унифицированный информационный объект для обмена и хранения электронных документов».

5. Доработка существующих нормативно-правовых актов Украины в сфере электронного документооборота и архивного хранения электронных документов.

¹ *Kampffmeyer U., Wasniewski A.* Records Management: Prinzipien, Standards & Trends [Electronic resource]. – Hamburg, Januar 2012. – 100 s. – Mode of access: http://www.project-consult.net/files/RM_Kff_20120130.pdf. – Title from screen.

² Australian ideas and management models. The Records Continuum [Electronic resource]. – Mode of access: <http://john.curtin.edu.au/society/index.html>. – Title from screen.

³ VERS – Victorian Electronic Records Strategy – Forever Digital [Electronic resource]. – Mode of access: <http://210.8.122.120/vers/standard>. – Title from screen; *Марченко П.* Архівний електронний документ: XML інкапсуляція // Студії з архів. справи та документознавства. – 2006. – Т. 14. – С. 61–63.

⁴ *Thibodeau K.* Preservation and Migration of Electronic Records: The State of the Issue [Electronic resource]. – Mode of access: http://www.interpares.org/display_file.cfm?doc=ip1_dissemination_cp_thibodeau_ica_2000.pdf. – Title from screen.

⁵ PROS 99/007. Standard for the management of electronic records, v. 1.2 [Electronic resource]. – Mode of access: <http://www.prov.vic.gov.au/vers/standards/pros9907.htm>. – Title from screen.

⁶ XML – W3C Recommendation [Electronic resource]. – Mode of access: <http://www.w3.org/TR/REC-xml>. – Title from screen.

⁷ CCSDS 650.0–B–1. (2002). Reference model for an Open Archival Information System (OAIS). The Consultative



Committee for Space Data Systems [Electronic resource]. – Mode of access: <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>. – Title from screen.

⁸ What is Unicode? [Electronic resource]. – Mode of access: <http://www.unicode.org/standard/WhatIsUnicode.html>. – Title from screen.

⁹ XML DOM Tutorial [Electronic resource]. – Mode of access: <http://www.w3schools.com/dom/default.asp>. – Title from screen.

¹⁰ Base64? [Electronic resource]. – Mode of access: <http://www.google.com>. – Title from screen.

¹¹ XML Schema – W3C Recommendation [Electronic resource]. – Mode of access: <http://www.w3.org/XML/Schema>. – Title from screen.

Запропоновано проект уніфікованого інформаційного об'єкта для обміну та зберігання електронних документів, в основі формування структури та змісту якого використовується концепція життєвого циклу «records continuum» з XML-інкапсуляцією файлів.

Ключові слова: архівний електронний документ, довгострокове зберігання, XML-інкапсуляція.

The authors propose a project of unified information object for the exchange and storage of electronic documents based on the formation its structure and content using the concept of life cycle «records continuum» with XML-encapsulated files.

Key words: archival electronic document, long term preservation, XML-encapsulation.

УДК 681.3:002.651.028(083.73)

Андрій Мелашенко, Олена Скарлат

МОДЕЛЬ ЗБЕРІГАННЯ МЕТАДАНИХ ЕЛЕКТРОННИХ ДОКУМЕНТІВ, ПРИДАТНИХ ДЛЯ АРХІВНОГО ЗБЕРІГАННЯ

Деталізовано механізми застосування інструментарію додавання метаданих до електронних документів формату PDF/A. На прикладі метаданих архівного PDF, як формату юридично правомочного самодостатнього електронного документа, продемонстровано створення машиночитувального подання реквізитів документа поряд із візуальним поданням.

Ключові слова: електронний документ, метадані, XML документ, реквізити документа.

Для повноцінного застосування електронного документа (далі – ЕД), як складової електронного діловодства суб'єкта господарювання, потрібно розв'язати два завдання, пов'язані з поданням ЕД: їхнім середньостроковим зберіганням відповідно до чинного законодавства та забезпеченням міжсистемної взаємодії. Для вирішення останнього доцільно орієнтуватися на світові практики в архівній справі, а саме: при застосуванні рекомендованого формату файлу PDF/A використовувати і відповідний стандарт метаданих. Нами подано основні принципи роботи з метаданими на основі платформи Extensible Metadata Platform (XMP), що вже 2012 р. було імплементовано у міжнародному стандарті ISO 16684-1¹.

Чинні ДСТУ з діловодства розглядають документ, відповідно і ЕД, як самодостатню множину атрибутів, що сприймається людиною, із визначеними типами даних, серед яких наявний обов'язковий атрибут – електронний цифровий підпис. Іншими словами, ЕД є контейнером, який містить необхідні реквізити та подання. В електронному діловодстві ЕД і архівний ЕД, придатний для архівного збереження, повинні бути однією сутністю. У такому випадку всі ЕД

матимуть єдиний формат, який доцільно унормувати із застосуванням національних стандартів, які гармонізовані з міжнародними. Тут використовуємо поняття саме «електронне діловодство» для конкретизації предметної області, яку охоплено.

Згідно з ДСТУ 4163-2003 «Уніфікована система організаційно-розпорядчої документації. Вимоги до оформлювання документів»² документ складається з 32-ох можливих атрибутів, отже у формат ЕД, окрім візуального подання, доцільно інтегрувати їхнє машиночитувальне подання для спрощення організації обміну ЕД. Використання реквізитів можливе при застосуванні інструментарію додавання до PDF-файлів метаданих: XMP-формат файлу метаданих, який застосовує необхідну для конкретних цілей XML-схему³. В PDF існує можливість вкладення необхідних метаданих за допомогою платформи XMP. Згідно із специфікацією, метадані транспортуються разом із файлом і можуть використовуватись у файлах форматів PDF, TIFF і JPEG. Властивості метаданих формуються

© Андрій Мелашенко, Олена Скарлат, 2012