



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.

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

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.

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



у схеми. Кожна схема ідентифікується унікальним простором імен URI і містить довільну кількість властивостей. Простори імен URI можуть і не вказувати на конкретний ресурс – вони є просто унікальними ідентифікаторами для деяких сутностей, які використовує XMP.

XMP специфікація включає понад 12 визначених схем і сотні властивостей для звичних документів і зображень. Схема, що є найбільш використовуваною – Dublin Core (dc)⁴. Вона включає найпоширеніші властивості: назва, автор, предмет і опис. У XMP включено можливість визначити додаткові схеми для виконання вимог конкретної сфери. XMP реалізовано у всіх Adobe продуктах. Цей формат підтримується багатьма виробниками програмного забезпечення і гру-

пами користувачів. PDF/A-1 в ISO/IEC 19005-1:2005⁵ вимагає XMP для ідентифікації відповідних файлів і підтримує метадані користувачів по XMP схемам розширень. PDF згідно з ISO 32000-1:2008⁶ повністю підтримує XMP метадані. Dublin Core – розповсюджена схема XMP метаданих, зокрема, популярна серед архівістів.

Доцільно для імплементації машинозчитувального подання 32-ох реквізитів уточнити їхній формат із застосуванням міжнародного стандарту загальних типів даних (ISO/IEC 11404:2007) та XML Schema Part 2: Datatypes Second Edition. Уточнене подання сформує XML схему, яку доцільно інтегрувати у XMP метадані PDF/A файлу. Приклад подання реквізитів у PDF/A файлі відображено на рисунках 1–4.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core
3 4.2.1-
4 c043 52.372728, 2009/01/18-15:08:04 <>
5 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
6 <rdf:Description rdf:about="" xmlns:pdf="http://ns.adobe.com/pdf/1.3/"
7 </rdf:Description>
8 <rdf:Description
9 rdf:about="" xmlns:dc="http://purl.org/dc/elements/1.1/"
10 <dc:creator><rdf:Seq><rdf:li>Olena</rdf:li></rdf:Seq></dc:creator>
<dc:title>
11 <rdf:Alt><rdf:li xml:lang="x-default"><Contract</rdf:li>
12 </rdf:Alt>
13 </dc:title>
14 </rdf:Description>
15 <rdf:Description rdf:about=""
16 xmlns:mysch="e:\myschema1.xsd">
17 <mysch:назва_організації>Інститут кібернетики
18 </mysch:назва_організації>
19 <mysch:код_організації>3222
20 </mysch:код_організації>
21 <mysch:код_форми_документа>02 011 325
22 </mysch:код_форми_документа>
23 </rdf:Description>
24 </rdf:RDF>
25 </x:xmpmeta>
```

Рис. 1. Вміст файлу метаданих Dublin Core.



На рис. 2 подано XSD-схему, що використовується у файлі XMP для подання специфічних для українського діловодства атрибутів.

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <xs:schema
3 xmlns:xs="http://www.w3.org/2001/XMLSchema"
4 <xs:element name="mysch" type="mysch"/>
5 <xs:complexType name="mysch">
6 <xs:sequence>
7 <xs:element name="назва_організації" type="xs:string"/>
8 <xs:element name="код_організації" type="xs:string"/>
9 <xs:element name="код_форми_документа" type="xs:string"/>
10 </xs:sequence>
11 </xs:complexType>
12 </xs:schema>
```

Рис. 2. XML-схема із описом додаткових атрибутів.

У властивостях відкритого файлу засобами, наприклад Adobe Reader 10, відображається таке:

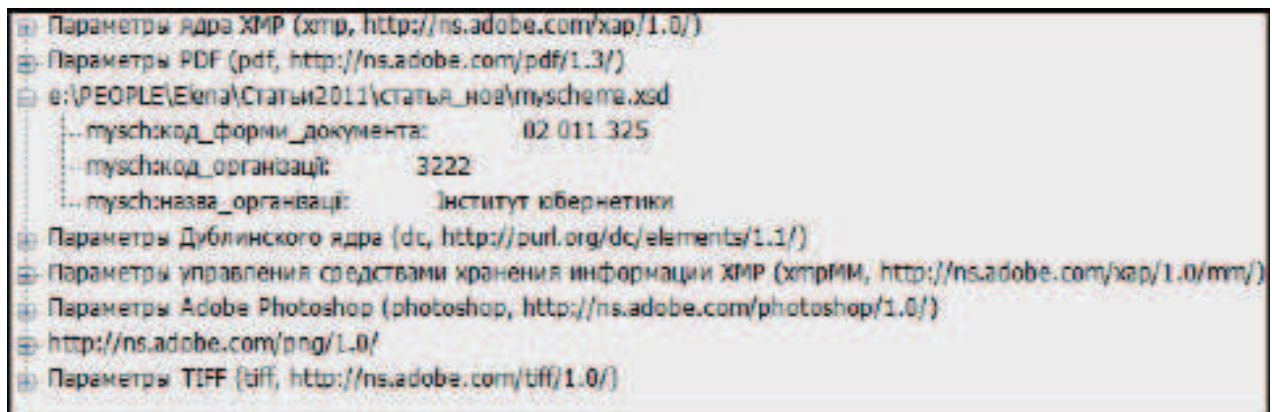


Рис. 3. Відображення нових атрибутів у Adobe Reader 10.

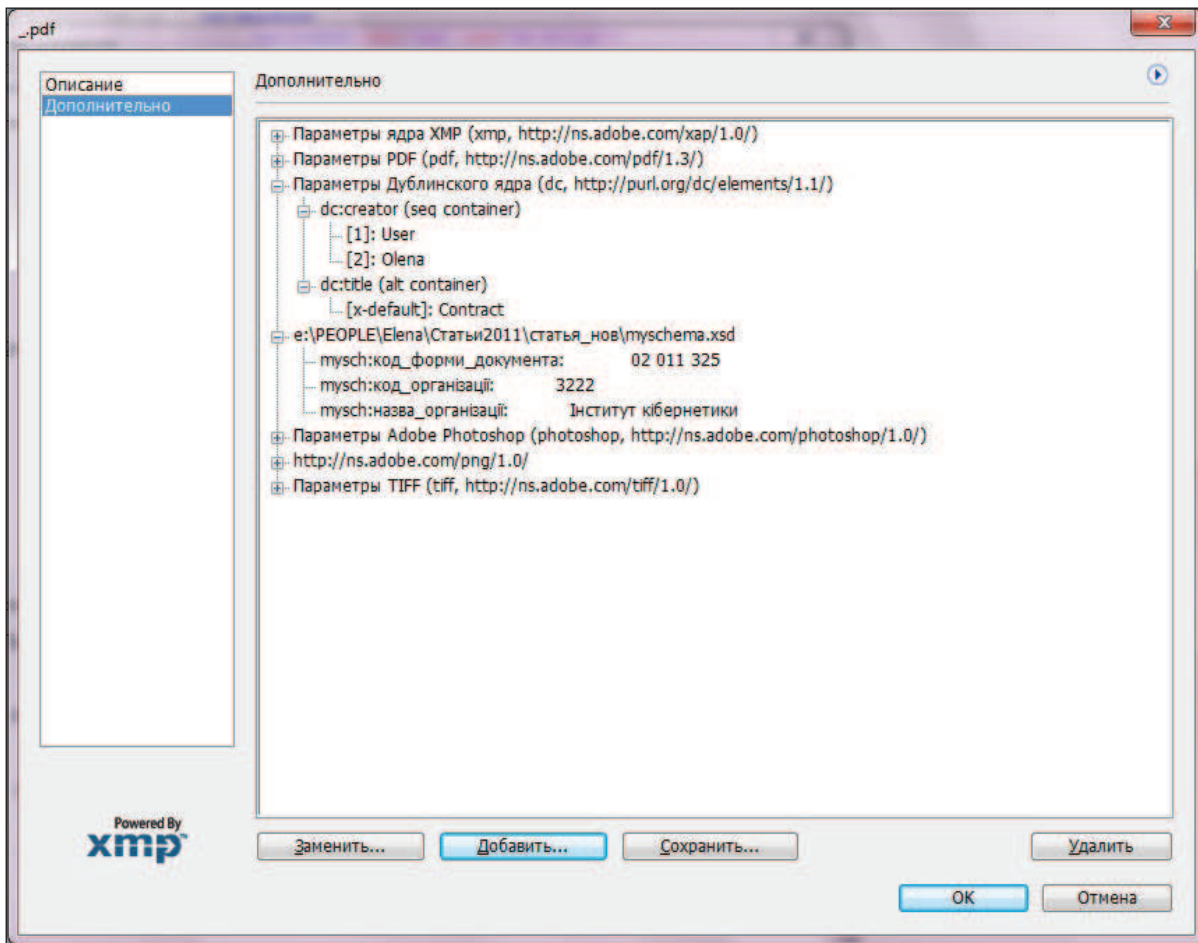


Рис. 4. Відображення нових атрибутів у Adobe Reader 10.

XMP властивості зберігаються як XML, а саме – згідно із стандартом RDF. Серіалізація властивостей метаданих у файлі – це представ-

лення властивостей метаданих, асоційованих із документом. На рис. 5: масив із двох елементів, кожний має кваліфікатор властивості «роль».

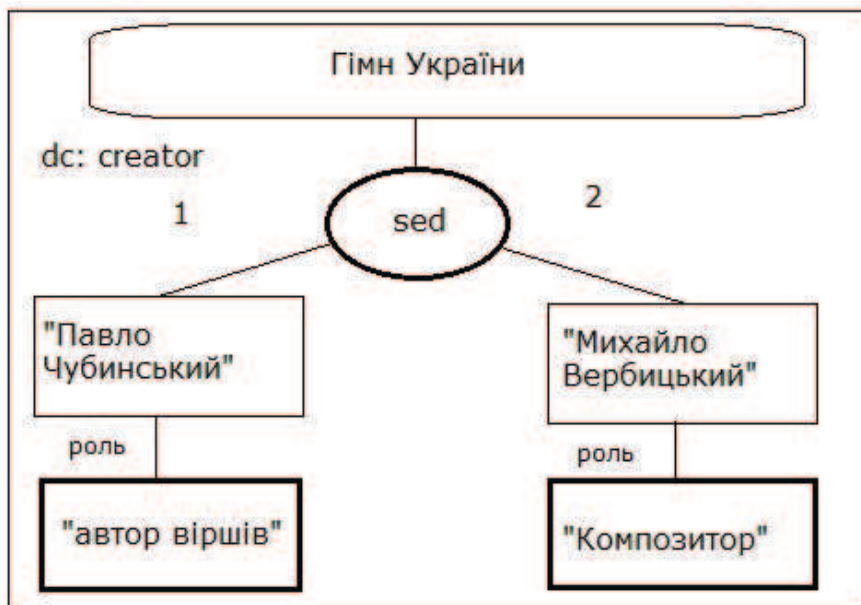


Рис. 5. Приклад властивостей метаданих.



Серіалізація відбувається таким чином (рис. 6):

```
<x:xmpmeta xmlns:x='adobe:ns:meta/'>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:creator>
      <rdf:Seq>
        <rdf:li>
          <rdf:Description>
            <rdf:value>Павло Чубинський</rdf:value>
            <role>автор віршів</role>
          </rdf:Description>
        </rdf:li>
        <rdf:li>
          <rdf:Description >
            <rdf:value>Михайло Вербицький</rdf:value>
            <role>композитор</role>
          </rdf:Description>
        </rdf:li>
      </rdf:Seq>
    </dc:creator>
  </rdf:RDF>
</x:xmpmeta>
```

Рис. 6. Серіалізація метаданих.

1. Серіалізовані дані заключаються в пакети має заголовок, серіалізовані дані, заповнення для включення у файли (рис. 7). Кожний пакет і трейлер:

Header	<?xpacket begin='■' id='W5M0MpCehiHzreSzNTczkc9d'?>
Serialized XMP	<x:xmpmeta xmlns:x='adobe:ns:meta/'> <rdf:RDF xmlns:rdf= ...> ... </rdf:RDF> </x:xmpmeta>
Padding	... XML whitespace as padding ...
Trailer	<?xpacket end='w'?>

Рис. 7. Вміст пакета даних для включення у файли.



2. Пакети зберігаються у файлах.

3. Зовнішнє зберігання метаданих – як зберігати ХМР дані окремо від документа, з яким його асоційовано.

ХМР метадані необхідно включати у файл, який вони описують (ХМР пакети), однак це не є обов'язковою вимогою. Бувають випадки, коли таке включення неможливе або не є придатним, наприклад: для моделі зберігання в базах даних, великий розмір файлів, проблеми формату або доступу. Якщо метадані зберігаються окремо від файлу, є ризик їх втрати. Виникає питання щодо того, як асоціювати метадані з файлом, що містить контент.

Програмні застосування мають:

– записати зовнішній файл як повний, сформований XML документ, включаючи XML декларування;

– розширення має бути .xmp;

– якщо вимагається тип MIME, необхідно використувати application/rdf+xml;

– записати зовнішні метадані так, ніби їх було вкладено і мають ХМР пакети, що видобуті та з'єднані постпроцесором;

– якщо можливо, необхідно помістити екземпляр ID, що використовується в атрибуті rdf:about в файлі, який ХМР описує, щоб відповідні формату застосунки мали змогу впевнитись, що мають правильні метадані;

– ХМР файл необхідно помістити в той же каталог, назвати тим же ім'ям, що й документ, але з розширенням .xmp.

В Adobe PDF схема використовує URI: <http://ns.adobe.com/pdf/1.3/>, префікс: pdf. Для PDF файлів, ХМР пакет включається в потік метаданих, що містяться в PDF об'єкті (починаючи з версії PDF 1.4). Приклад показано на рисунку 8:

```
1152 0 obj
<< /Type /Metadata /Subtype /XML /Length 1706 >>
stream
<?xpacket begin=" id='W5M0MpCehiHzreSzNTczkc9d'?">
<!-- The serialized RDF goes here. It has been removed
for brevity. -->
<?xpacket end='w'?">
endstream
endobj
```

Рис. 8. Приклад включення ХМР пакета.

PDF файли, які поступово зберігали по мірі редагування, можуть мати безліч пакетів метаданих, і всі виглядають як «головні» ХМР метадані. При поступовому збереженні нові дані (включаючи ХМР пакети) записуються в кінець файлу без видалення старих. Високорівнені PDF довідники також перезаписуються, і застосунки PDF перевіряють довідники і знаходять лише нові пакети метаданих.

На прикладі метаданих архівного PDF, як формату юридично правомочного самодостатнього електронного документа, продемонстровано створення машинозчитувального подання реквізитів документа поряд із візуальним поданням.

¹ ISO 16684-1:2012. Graphic technology – Extensible metadata platform (XMP) specification. Part 1: Data model, serialization and core properties [Electronic resource]. – Mode of access: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57421. – Title from screen.

² Уніфікована система організаційно-розпорядчої документації. Вимоги до оформлювання документів: ДСТУ 4163-2003 – [Чинний від 2003-09-01]. – К.: Держспоживстандарт України, 2003. – 22 с. – (Національний стандарт України).

³ XMP Specification [Electronic resource]. – Mode of access: <http://partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf>. – Title from screen.

⁴ ISO 15836:2009. Information and documentation – The Dublin Core metadata element set [Electronic resource]. – Mode of access: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52142. – Title from screen.

⁵ ISO/IEC 19005-1:2005. Document management – Electronic document file format for long-term preservation. Part 1: Use of PDF 1.4 (PDF/A-1) [Electronic resource]. – Mode of access: http://www.iso.org/iso/catalogue_detail?csnumber=38920. – Title from screen.

⁶ ISO 32000-1:2008. Document management – Portable document format. Part 1: PDF 1.7 [Electronic resource]. – Mode of access: http://www.iso.org/iso/catalogue_detail.htm?csnumber=51502. – Title from screen.



Детализирован механизм применения инструментария добавления метаданных к электронным документам формата PDF/A. На примере метаданных архивного PDF, как формата юридически правомочного самостоятельного электронного документа, продемонстрировано создание машиночитываемого представления реквизитов документа наряду с визуальным представлением.

Ключевые слова: электронный документ, метаданные, XML документ, реквизиты документа.

In the article are depicted in detail the mechanisms of using tools for adding metadata to electronic documents in format PDF/A. By the example of metadata of the archival PDF file, a format for legally sustainable all-sufficient electronic document, is demonstrated formation of the machine-readable representation of properties of the document, along with a visual representation.

Key words: electronic document, metadata, XML document, properties of the document.

УДК 930.253:004(574.42)

Инесса Носова

«ЭЛЕКТРОННЫЙ АРХИВ» ВОСТОЧНО-КАЗАХСТАНСКОЙ ОБЛАСТИ

Освещены вопросы организации и создания информационной системы «электронные архивы Восточно-Казахстанской области».

Ключевые слова: информатизация архивной отрасли, электронный архив, Восточно-Казахстанская область.

Проблема возрастания объемов документов сейчас является актуальной для всех архивных учреждений. Современные информационно-коммуникационные технологии слабо внедряются в архивную отрасль, вследствие чего огромный информационный потенциал Национального архивного фонда минимально используется для удовлетворения возрастающего спроса граждан, государства и общества на информацию. Наиболее эффективным решением этой проблемы является организация и создание «электронного архива».

В государственных архивах Восточно-Казахстанской области сейчас хранится более 2 млн. дел на бумажной основе, из них около 50 тыс. являются особо ценными. Отсюда задача – грамотно управлять огромным объемом документов.

Архивные учреждения Восточно-Казахстанской области одними из первых в Республике Казахстан начали работу по созданию «электронного архива». В соответствии с планом мероприятий по реализации Государственной программы формирования «электронного правительства» в Республике Казахстан на 2005–2007 гг., утвержденным постановлением Правительства Республики Казахстан от 8 декабря 2004 г. № 1286, управление архивов и документации Восточно-Казахстанской области разработало и согласовало с акиматом региона план мероприятий по созданию информационной системы «электронные архивы Восточно-Казахстанской области» на 2007–2010 гг. Из областного бюджета было выделено 54,3 млн. тенге для ее создания.

Перед архивными учреждениями Восточно-Казахстанской области сейчас стоят две основ-

ные задачи в области информационных технологий: ввести в систему массив имеющихся в архиве значимых документов и обеспечить возможность оперативного полнотекстового доступа к электронным документам.

Информатизация невозможна без соответствующей технической инфраструктуры. В настоящее время в государственных архивах области имеется следующее оборудование: для возможности перевода документов с бумажных носителей в электронный вид приобретены три цветных книжных сканера ЭЛАР ПланСкан формата А3; для хранения электронных документов установлены три сервера; для обеспечения совместной работы сотрудников архива над единой базой электронных документов созданы локальные вычислительные сети, установлены мини АТС; для чтения микрофиш и микропленок (16/35 мм) приобретены два специальных аппарата Indus 4601-11, цветные планшетные сканеры формата А3 и А4, цветные сканеры формата А4 со встроенным слайд-адаптером, профессиональные видеокамеры, цветные лазерные принтеры; для показа мультимедийных презентаций приобретены проекторы и экраны, ноутбуки; для всех архивных учреждений приобретены компьютеры необходимой мощности, установлены антивирусные программы; все архивные учреждения подключены к сети Интернет.

Важнейшим этапом работ по созданию «электронного архива» является сканирование документов, их обработка и создание баз дан-