

Петро МАРЧЕНКО
СТАНДАРТИ КОДОВАНИХ
АРХІВНОГО ОПИСУ ТА КОНТЕКСТУ
(проблема впровадження)

Нині архівні інформаційні ресурси об'єднуються в глобальну мережу і питання розроблення та впровадження стандартів міжнародного призначення, які мають забезпечити доступ до архівної інформації незалежно від мовних чи галузевих кордонів, є актуальним для всіх учасників інформаційної спільноти. Міжнародна рада архівів тепер рекомендує в якості міжнародних стандартів кодованого описування архівних інформаційних ресурсів два стандарти метаданих: стандарт кодованого архівного опису – EAD (Encoded Archival Description)¹ та стандарт кодованого архівного контексту – EAC (Encoded Archival Context)². В роботах Л. Драгомірової³ та С. Єремеева⁴ уже було розглянуто науково-методологічні питання створення архівного опису та архівного контексту, пов'язаних з використанням стандартів EAD та EAC. Метою цієї статті є рекомендації щодо вирішення проблеми впровадження даних стандартів у вітчизняній архівній інформаційній системі з огляду на її інтеграцію в глобальну мережу.

Відповідно до інформаційної структури описуваних ресурсів розрізняють шість можливих типів метаданих⁵:

- метадані пошуку, ідентифікації ресурсу (“пошукові”);
- метадані правил доступу, вимог та умов використання ресурсу (“правила та умови”);
- метадані структури ресурсу та його користування (“структура”);
- метадані користування ресурсом (“історія користування”);
- метадані змістовної частини ресурсу (“зміст”);
- метадані аспектів створення та розвитку ресурсу (“контекст”).

EAD зарекомендував себе як стандарт описування ієрархічно структурованих архівних ресурсів, що містить метадані довідкового апарату архівного ресурсу, а також метадані змісту архівних об'єктів та процесів їх життєвого циклу (зберігання, виокремлення для переміщення, адміністрування тощо). Таким чином, EAD містить більшість вищезазначених типів метаданих. Специфікація EAD, підтримана Американським архівним товариством та Бібліотекою Конгресу США, нараховує 135 елементів, із яких 26 запозичено з Міжнародного стандарту архівного опису (загального) – ISAD(G) (General International Standard Archival Description).

EAC – стандарт архівного контексту, який надає інформацію про утворювачів архівного ресурсу та їхню діяльність (загальноісторичний,

політичний, економічний аспекти тощо). Стандарт підтримується робочою групою EAC Working Group⁶. Специфікація налічує 43 елемента, більшість із яких запозичено з Міжнародного стандарту архівних авторитетних даних на інституції, окремих осіб та родини – ISAAR(CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons, and Families).

Стандарти EAD та EAC, які є кодованими текстами метаданих, придатними для опрацювання програмними засобами, доцільно впроваджувати в уже наявній чи розроблюваній дескриптивній системі електронного архіву. Саме така система призначена для створення синтаксично та граматично правильних (англ. – well-formed) кодованих текстів метаданих (архівного опису, архівного контексту), які коректні (англ. – valid), тобто відповідають прийнятому стандарту. Кодування текстів метаданих, яке відповідає прийнятим стандартам, здійснюється за допомогою їх електронного аналога – схеми (англ. – schema)⁷. Схема, по суті, представляє собою еталонну модель метаданих інформаційного ресурсу, яка дає змогу звіряти створювані тексти метаданих на відповідність таким якостям як “well-formed” та “valid”. Цей процес називається верифікацією створених метаданих і, якщо кодовані метадані відповідають зазначеним якостям, то такі метадані відповідають прийнятому стандарту, а для їх опрацювання може бути застосовано уніфіковане Web-орієнтоване програмне забезпечення.

Загалом дескриптивна система електронного архіву, яка використовує схему метаданих архівного ресурсу, має такі інтерфейси:

- інтерфейс адміністратора системи, який відкриває можливість керувати процесами створення метаданих та адміністрування доступу до метаданих;
- інтерфейс створювача метаданих, який дозволяє заповнювати інформацією розроблені шаблони метаданих, перевіряти їх на відповідність представлений схемі на сервері та передавати створені метадані для користування;
- інтерфейс користувача, який представляє собою Web-сторінку із елементами реалізації пошуку архівної інформації для подальшого її опрацювання за визначеною користувачем метою.

Кожен із згаданих вище інтерфейсів є певним набором Web-орієнтованих програмно-технічних засобів для вирішення завдань за його призначенням.

Інтерфейси дескриптивної системи зорієнтовані на використання розширеної мови описування XML (Extensible Markup Language)⁸ та відповідного XML-інструментарію – програм, які дають можливість створювати тексти метаданих за XML шаблонами. XML – мовний засіб описування інформаційних ресурсів, запропонований Інтернет-консорціумом W3C, який уже широко використовується. XML дає



Дескриптивна система електронного архіву

зможу подати метадані інформаційного ресурсу у вигляді звичайного тексту, тобто у вигляді XML-документа. Для подання схеми опису інформаційного ресурсу використовують чинні специфікації XML-схем: DTD (Document Type Definition)⁹ або XML Schema¹⁰.

Метою впровадження EAD може бути¹¹:

- конвертація наявного електронного довідкового апарату (ДА) у формат XML EAD-описів;
- створення електронного ДА у форматі XML EAD-описів для архівних об'єктів, які не мають такого;
- використання EAD як стратегії створення дескриптивної системи з єдиним web-порталом на уже розробленому та апробованому програмному забезпеченні.

Те саме можна екстраполювати і для EAC.

В українській архівній галузі вже існують напрацювання у створенні власної дескриптивної системи¹²: на основі міжнародного стандарту архівного описування (загального) ISAD(G) розроблено стандарт ДСТУ 4331:2004 :“Правила описування архівних документів (ПАО)”. Зважаючи на це, метою впровадження EAC та EAD в архівній інформаційній мережі України доцільно вибрати останній пункт із переліку, тобто використати EAD та EAC як приклади стратегії створення дескриптивних систем. Така доцільність додатково підтверджується позитивним досвідом їх апробації в США, Франції, Німеччині, Нідерландах та Великобританії¹³. У такому випадку ПАО, як дескриптивний стандарт, цілком може слугувати основою для розроблення власного стандарту кодового архівного опису (КАО), при цьому КАО за повнотою опису має відповідати як вітчизняному стандарту ПАО, так і EAD та EAC. Маючи на увазі, що в КАО можна використати англійські позначення елементів та атрибутів, запозичених із специфікацій EAD чи EAC і таким

чином досягти рівня сумісності, який значно полегшить розпізнавання пошуковими засобами, розробленими для EAD чи EAC, а також вирішувати питання конвертації КАО-описів в EAD/EAC-описи.

Для реалізації визначеного варіанту впровадження EAD та EAC доцільно рекомендувати і відповідне рішення, а саме формат XML DOM (Document Object Model)¹⁴, який може представляти гнучкіші, динамічні моделі XML-документів. XML DOM в залежності від потреб користувача, надає можливість формувати певні деревовидні структури XML-документа. Таким чином, використовуючи формат кодованого архівного опису XML Schema DOM, в подальшому можна спростити отримання EAD/EAC-описів у результаті виокремлення та використання певних елементів, гілок деревовидної структури КАО.

Розглянемо структуру EAD та визначимо завдання, які потрібно вирішити при розробленні стандарту КАО.

EAD має ієрархічну структуру, верхній рівень складається з трьох блоків:

Блок <eadheader> містить інформацію про існуючий довідковий апарат (finding aids), які пропонуються користувачеві (бази даних, файлові сервери тощо) та складається з ідентифікаторів, описів файлів, описів параметрів користування, описів оновлення засобів. В стандарті ПАО цей блок відсутній. Таким чином, він потребує в майбутньому стандарті КАО розроблення. В процесі розроблення КАО корисно створювати крос-таблиці наявності елементів КАО/EAD.

Зміст елементів блоку <eadheader> (eadheader — позначення кореневого елемента схеми):

<eadid> – елемент з атрибутами, який унікально ідентифікує дану версію пошукового засобу;

<filedesc> – елемент з атрибутами, що містить інформацію про файл пошукового засобу (назва, автор, вихідні дані);

<profiledesc> – елемент з атрибутами, що містить інформацію про набір параметрів, що характеризують користування пошуковим засобом (мова, версія);

<revisiondesc> – елемент з атрибутами, що містить інформацію про зміни та доповнення параметрів пошукового засобу, на які потрібно звернути увагу.

По цьому блоку при розробленні КАО необхідно вирішити питання визначення унікальних ідентифікаторів як для елементів, так і для файлів (галузевий стандарт або методичні рекомендації). Також слід визначитися з інтерфейсом користувача (у вирішенні цього питання бажана уніфікація рішення за спорідненими ознаками архівних ресурсів та пошукових засобів довідкового апарату).

Блок <frontmatter (головна сторінка)> містить інформацію, яка стосується опису архіву і складається з двох елементів з атрибутами <div>, <titlepage>. Зміст елементів цього блоку:

<div> – текст у вигляді титульного аркуша з передмовою та інструкціями;

<titlepage> – інформація про пошуковий засіб існуючого електронного довідкового апарату.

В стандарті ПАО даний блок відсутній.

Блок <archdesc> – архівний опис фонду. Блок складається з десяти елементів з такими атрибутами:

<add> – додаткова інформація (додатки до документів, додаткові надходження);

<admininfo> – інформація про доступність документів;

<arrangement> – інформація про схему систематизації документів;

<bioghist> – біографічна та історична інформація про особи або заклад;

<controllaccess> – інформація, що стосується рівнів доступу;

<did> – ідентифікатор архівного опису;

<dsc> – архівний опис фонду, справи або документа;

<note> – примітки;

<odd> – інші описові дані, які не відносяться до жодного з елементів опису;

<organization> – інформація про рух документів архівного фонду;

<scopecontent> – інформація щодо складу документів архівного фонду.

У стандарті ПАО цей блок, як і в EAD, складає основну його частину. Елементи цього блоку в EAD відповідають елементам архівного описування ISAD(G), який взято за основу і в розробленні ПАО. Проблема наповнення КАО відповідниками ЕАС буде полягати у їх терміно-лексичному узгодженні. При вирішенні питання позначення елементів відповідників різних стандартів доцільно скористатися рекомендаціями розробників EAD. Вони звертають увагу на те, що ідентифікатори елементів та атрибутів в EAD унікальні для даного стандарту, що забезпечує однозначність зв'язків, але використовуються і запозичення типу “encoding analogs” із уже наявних позначень у цьому стандарті, а також запозичення типу “relatedencoding”, що мають кодовані аналоги суміжних дескриптивних стандартів, наприклад, MARC 21, Dublin Core тощо. Отже, у створенні КАО для забезпечення сумісності та полегшення конвертації описів вбачається можливим запропонувати атрибути – кодовані аналоги елементів EAD та ЕАС, що забезпечить відповідну конвертованість описів ПАО, EAD та ЕАС.

ЕАС розроблявся одночасно з другою версією стандарту ISAAR(CPF) і є засобом он-лайнного обміну стандартизованими машиночитаними ISAAR-сумісними описами створювачів документів. Стандарт, по суті, відноситься до групи стандартів метаданих типу Dublin Core, MARC XML. ЕАС, як і EAD, має ієрархічну структуру. Вищий рівень пред-

ставлений двома блоками елементів: блок <eachheader> – титульний аркуш та блок <condesc> – зміст контексту. Склад блоків наступний:

<eachheader> – блок елементів про створювача змісту контексту та метаданих самого опису контексту.

<condesc> – блок елементів, які подають зміст контексту (дані про створювача архівного об'єкта), а саме:

<eacid> – унікальний ідентифікатор архівного об'єкта в межах використовуваної створювачем системи їх ідентифікації як візуального, так і машиночитаного сприйняття;

<mainhist> – елемент, який має один або декілька піделементів типу <mainevent>, який вказує, що опис стосується якогось із процесів створення, тобто, зберігання, передавання, внесення змін або виділення до знищення. Кожен із блоків <mainevent> містить дані про фізичну або юридичну особу, дату, опис певного процесу, для позначення яких використовується атрибут “maintype”.

<languagedecl> – елемент, що визначає використовувані мови опису контексту;

<ruledecl> – елемент з атрибутами, який представляє правила описування змісту процесів;

<sourcedecl> – елемент з атрибутами, які представляють джерела інформації стосовно правил описування процесів;

<authdecl> – елемент з атрибутами, які подають використану інформацію в описуванні контексту.

Виходячи з переахованого вище, ЕАС представляє цінні метадані, які можуть бути використані архівістами для організації архівних фондів, а користувачами в пошуку необхідної інформації.

Таким чином, проблема впровадження міжнародних стандартів EAD та ЕАС зводиться до вирішення наступних завдань:

- розроблення узгодженого із стандартами EAD та ЕАС власного стандарту кодованого архівного опису КАО та його XML-схеми;
- освоєння технології створення КАО-описів як XML-документів, що відповідають поданій XML-схемі;
- розроблення та впровадження Web-орієнтованих інтерфейсів створення, адміністрування та користування метаданими;
- розроблення та впровадження XML-орієнтованих баз даних КАО-описів.

В результаті реалізації вищезазначеного електронні архіви отримають міжоперабельну дескриптивну систему, що дозволить інтегрувати архівні ресурси в інформаційну систему будь-якого рівня.

¹ <<http://www.ifla.org/documents/libraries/cataloging/metadata/bead1.txt>>.

² <http://xml.coverpages.org/EAC-DTD200203.html>.

³ Драгомірова Л. Оптимальна модель організації документів національного архівного фонду: методологія визначення і технологія створення. http://www.scau.kiev/Publicat/Studii/Studii_2004.11_05.php.

⁴ *Єремєєв С.* Американський дескриптивний стандарт EAD: історія, структура, термінологія, можливості використання // Студії з архівної справи та документознавства. – Т. 13. – К., 2005. – С. 104–110.

⁵ David Bearman and Ken Sochats: Metadata Requirements for Evidence. <<http://www.lis.pitt.edu/~nhprc/BACartic.html>>.

⁶ <http://www.library.yale.edu/eac/>.

⁷ <http://www.microsoft.com/rus/msdn/activ/article/xml/using.asp>.

⁸ <http://wdvl.com/Authoring/Languages/XML/>.

⁹ <http://wdvl.com/Authoring/HTML/Validation/DTD.html>.

¹⁰ <http://www.w3.org/2001/XMLSchema>.

¹¹ http://www.cdlib.org/inside/diglib/resources/local1_ead.html/.

¹² *Христова Н.* Стандартизація в архівній справі: проблеми та перспективи // Студії з архівної справи та документознавства. – Т. 12. – К., 2004. – С. 75–78.

¹³ Международные стандарты для электронных архивов. Третья Европейская конференция EAD, EAC и METS. Берлин, 24–26 апреля 2007 г. <http://www.vniidad.ru/news1/example>.

¹⁴ *Jonathan Robie.* What is the Document Object Model? <http://www.w3.org/TR/REC-DOM-Level-1/cover.html>.

P. MARCHENKO

Standards encoded the archived description and the context (the problem of implementation)

The implementation problem of the international standards encoded the archival description and the context (EAD, EAC) is considered. The decision of the problem demands creation of the certain descriptive system of electronic archive. The primary goals, which are necessary for solving at creation of descriptive system, and also approaches to their decision are outlined.