

Петро Марченко

ШЛЯХИ ВИРІШЕННЯ ПРОБЛЕМИ ЗБЕРІГАННЯ ЕЛЕКТРОННИХ ДОКУМЕНТІВ В АРХІВІ (аналіз Інтернет-ресурсів)

1. Про концепцію електронного документа (ЕД)

Розвиток інформаційних технологій та Інтернету ініціювали дискусії та дослідження щодо концепції ЕД, його відмінностей та особливостей порівняно з традиційним документом з паперовим носієм. Деякі американські архівісти¹ вважають що:

- ЕД або комбінація ЕД, створені фізичними або юридичними особами, є доказом дій або взаємодії та містить інформацію про зміст дій та взаємодії;
- ЕД відомий своїми метаданими ("метадані" (анг. - metadata)² означає "дані, що пов'язані або стосуються інших даних), тобто інформацією яка визначає його соціальну визначеність та розумінням (зміст, контекст, структура). Різноманітні набори метаданих мають стати юридичними, діловими, організаційними, процедурними доказами автентичності для авторів, користувачів та дослідників.

Як об'єкт архівного зберігання ЕД характеризується такими ознаками:

- ЕД є програмно-технічно залежним продуктом;
- ЕД мають широкий спектр інформаційного відображення (текстові, електронні таблиці, бази даних, мультимедійні);
- форма документа може бути відокремлена від змісту, зміст документа може бути фрагментованим (бази даних), тобто фізично документ може зберігатися в декількох різних файлах;
- ЕД може мати посилання, які не контролюються автором документа, наприклад, під час користування Інтернет-файлами, або корпоративними базами даних, що відрізняються життєвим циклом;
- ЕД зберігають на фізичних носіях інформації (магнітні, оптичні пристрої), що не можуть гарантувати довготривале збереження інформації (процес розмагнічування, механічне ушкодження, фізичне та моральне старіння програмно-технічних засобів).

Формати створених ЕД характеризуються тим, що мають відкриту або закриту специфікацію, рівень стандарту, або просто характеризуються обсягом розповсюдження та використання. Відзначається³, що в вирішенні проблеми довготривалого зберігання ЕД акцент має бути на створенні надійного сховища специфікацій форматів, документації та програм. Поширені такі формати електронних документів⁴:

- формати відкритого тексту (plain text): в діловодстві для створення ЕД широко використовуються формати програм Microsoft (Word, Excel), які характеризуються тим, що мають закриту специфікацію. В зв'язку з тим, що це не прийнятно в аспекті довготривалого зберігання ЕД, для архівів рекомендується використовувати формат відкритої специфікації - PDF фірми Adobe Systems Inc. програма "Acrobat", який визначений ISO (The International Organization for Standardization) як стандарт (ISO/DS 19005-1) довготривалого зберігання електронних документів;
- формати графічних зображень:
TIFF (Tag Image File Format): графічний формат, який рекомендується як стандарт високої якості зображень для широкого застосування і довготривалого зберігання графічних зображень;
GIF (Graphics Interchange Format): графічний формат рекомендований для передавання зображення каналами зв'язку, не рекомендується як формат довготривалого зберігання зображень;
JPEG (Joint Photographic Experts Group): графічний формат з ущільненням інформації з мінімальними її втратами, широко використовується в Інтернет, в діловодстві, ЕД в цьому форматі надходять і до архіву;

- інші формати (pict, bmp тощо) використовують в специфічних додатках і не рекомендовані для довготривалого зберігання зображень;
- формат структурованого тексту (текстові документи, які мають графічні та аудіо включення, Інтернет-посилання тощо): поширено використання формату SGML (Standard Markup Generalized Language). Цей формат структурованих документів має можливості подетального відкривання документа в зв'язку з чим зручний для довготривалого зберігання документів в архівах.

Архівісти та дослідники покладають великі надії на нещодавно представлений формат XML (Extensible Markup Language), прийнятий для створення та зберігання структурованих ієрархічних даних незалежно від характеристик програмно-технічних засобів. Формат створено W3C (World Wide Web Consortium)⁵. Носії інформації для запису ЕД класифіковано таким чином⁶:

- CD-ROM, CD-R носії: найпридатніші для архівного зберігання електронних документів, мають відносно значну ємність інформації для запису, захищені від випадкового стирання, досить широко використовуються і виготовляються (виробники обіцяють для своєї продукції термін безпечного зберігання інформації 100 і більше років). На практиці рекомендовано перезапис інформації на CD кожні 10 років при забезпеченні належних умов зберігання (інколи спостерігається руйнація алюмінієвої основи);
- DAT (Digital Audio Tape), DLT cassettes: порівняно з CD, записана інформація може бути випадково модифікована або стерта, але зручні при опрацюванні великої за об'ємом інформації (декілька десятків GB);
- DVD (Digital Versatile Disk): оптичний диск, має такі ж переваги як і CD і більше, оскільки може зберігати до 17 GB інформації і набувають широкого вжитку;
- ZIP-drive: носії за об'ємом в 100-250 MB інформації, не мають промислового стандарту і можливі у використанні за відсутності інших засобів;
- 1.44 MB floppy disk: найменш придатний для архівного зберігання: мала ємність, висока вразливість, приймаються тільки за погодженням з архівом.

2. Проблема архівного зберігання ЕД та стратегії її вирішення

Довготривале, довічне (тобто, по можливості найтриваліше) зберігання ЕД є однією з проблем інформаційної спільноти, важливість якої визначається все більшим проникненням інформаційних технологій в життя людини. Chen S.S.⁷ характеризує проблему довготривалого зберігання ЕД як "критичну, накопичувану слабкість нашої інформаційної інфраструктури". Відзначається, що проблема архівного зберігання є як технічною так і організаційною проблемою. Технічна проблема полягає в досить короткому життєвому циклі програмно-технічних засобів створення ЕД та носіїв інформації для їх зберігання (моральне та фізичне старіння). Організаційна проблема полягає в тому, що не завжди архівісти мають вплив на вирішення питань законодавчого та фінансового забезпечення організації створення та функціонування ЕД.

Міжнародна Рада Архівів (ІСА), Комітет електронних документів в 1997 р. визначив основні принципи довготривалого зберігання ЕД в архівах⁸.

Відповідно до цих рекомендацій архів повинен:

- дослідити життєвий цикл електронних систем, що створюють та зберігають архівні ЕД;
- упевнюватись, що ЕД, які створюють і зберігають, є автентичними і цілісними;
- здійснювати експертизу та контролювати процеси зберігання архівних ЕД;
- об'єднати вимоги до зберігання та користування ними, забезпечивши їх доступність.

Автентичність архівного ЕД - одне з вузлових питань. Загроза втрати автентичності ЕД зумовлена здатністю редагування програмними засобами а також необхідністю в конвертуванні та міграції документів (переміщення ЕД в наступне покоління програмно-технічних засобів).

В електронній публікації⁹ автентичність визначена як доказ того:

- що ЕД відповідає зазначеному;
- що ЕД, створений або надісланий зазначеним автором або організацією, що створили його;
- що зазначений час відповідає часу створення або надсилання.

Для забезпечення автентичності ЕД створювані метадані мають містити відомості про криптографічні перетворення документа (електронний цифровий підпис) або контрольна сума одиниць інформації для документа, а також інформацію, що автори мають визначені повноваження створення та передавання ЕД до архіву. У процесі здійснення процедур з ЕД електронний цифровий підпис або контрольна сума мають бути підтверджені.

До визначених стратегій вирішення проблеми довготривалого зберігання ЕД слід віднести¹⁰:

- стратегію консервації: базується на збиранні та зберіганні морально застарілих програмно-технічних засобів, що дозволяють відображати ЕД в конфігурації, в якій він був створений;
- стратегію емуляції - розроблення програм, які відтворюють дії морально застарілого програмного забезпечення так, що будуть отримані результати так само як і в конфігурації програмно-технічного забезпечення, в якому було створено ЕД;
- стратегію міграції - періодичне перенесення ЕД з існуючого в наступне покоління програмно-технічних засобів;
- стратегію інкапсуляції: базується на ідеї, що ЕД має інформаційну оболонку, що містить описання, яке слугує для організації процесів його зберігання та користування, тобто метадані.

Перша стратегія є не досить практичною: технічні засоби та програмне забезпечення виробляється та виготовляється незалежними виробниками, тому у разі припинення виробництва, відновлення працездатності обладнання потребує значних фінансових ресурсів.

Друга стратегія - емуляції, потребує зберігання значного обсягу документації стосовно програмно-технічних морально застарілих засобів, що дозволить в майбутніх поколіннях здійснити конвертацію або розробку програмного забезпечення для відтворення і подальшого зберігання оригіналів ЕД. Ця стратегія досить обтяжлива, а тому також потребує великих витрат.

Періодична міграція електронних документів із покоління в покоління програмно-технічних засобів є одним з перспективних рішень проблеми довготривалого зберігання ЕД. Успішне використання цієї стратегії залежить від метаданих, пов'язаних з електронним документом, що містять історію міграції документа і забезпечує розуміння середовища, в якому було створено ЕД.

Стратегія інкапсуляції притягує увагу архівістів, наприклад Австралії¹¹, як можливість перенести початок традиційного життєвого циклу для ЕД на момент до його створення, або водночас зі створенням, так звана концепція "records continuum". Це, на думку дослідників, створює усі умови для успішного вирішення проблеми довготривалого зберігання ЕД. Ця стратегія передбачає включення до метаданих певної частини архівного описування і передавання до архіву ЕД у вигляді інкапсульованого інформаційного об'єкта.

3. Огляд зарубіжних проектів та ініціатив

Основні зарубіжні проекти та ініціативи щодо вирішення проблеми довготривалого зберігання ЕД виконані у співпраці асоціацій бібліотекарів, архівістів та керуючих документацією.

Архівісти та керуючі документацією на початку концентрували свою увагу на створенні ЕД в автоматизованих системах діловодства. Одним з найперших таких досліджень був Піттсбургський проект "PITT"¹². Результатом цього проекту була розробка прийнятних функціональних вимог до систем діловодства. Головною вимогою було те, що системи

мають забезпечувати приймання, використання та як можливо у більш продовженому часі здатність відтворення ЕД. В проекті була також запропонована базова модель метаданих ЕД для ділових зв'язків (Business Acceptable Communications), яка включає унікальний ідентифікатор документа та назву ресурсу для його відкриття, умови користування ЕД, структура ЕД, функції перевірки документа під час користування ним. Метадані документа мали нести всю необхідну інформацію, що дозволяла користування документом в умовах навіть коли комп'ютерна система, або інформаційний стандарт за яким його було створено, вже відсутні. Специфікація метаданих була прийнята до міжнародного стандарту архівного описування ISAD(G) (General International Standard Archival Description)¹³.

Значний внесок у вирішення зазначеної проблеми отримав проект PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), розроблений співробітниками Національної бібліотеки Австралії¹⁴. В ньому була запропонована модель (діаграма) зв'язків та відповідних метаданих для інформаційних об'єктів, що також важливо для функціонування ЕД.

Інший значний проект, реалізований в Австралії, це проект - "Стратегія електронних документів штату Вікторія" або VERS (Victorian Electronic Records Strategy), що базується на стратегії інкапсуляції ЕД (the VERS Encapsulated Object) із використанням формату XML (Extensible Markup Language). Специфікації цього проекту визначені як чинний стандарт діловодства і передавання ЕД до архіву PROV (Public Record Office of Victoria)¹⁵.

Заслугує на увагу проект OAIS (Open Archival Information System - Відкритий проект архівної інформаційної системи)¹⁶, виконаний за ініціативою ISO Консультативним комітетом систем космічних даних CCSDS, у складі якого представники десятих країн.

Проект визначив загальну структуру архівних інформаційних систем. Слово у назві "відкритий" означає його відкритість в аспекті використання, адже він запропонований як базова модель у контексті вимог архівістів, тобто наявності довготривалого зберігання електронних даних за умови змін технологій, носіїв інформації, форматів даних, складу користувачів.

Базова модель OAIS:

- пропонує базову модель довготривалого зберігання і доступу до електронних даних;
- надає концепцію, що дозволяє усім не архівним організаціям бути активними учасниками в процесі зберігання документів;
- описує структуру архіву (термінологія, архітектура засобів, експлуатація) для існуючих та майбутніх архівів;
- слугує підставою для порівняльного аналізу та дискусій з питань функціональної та інформаційної моделей довготривалого зберігання електронних даних;
- пропонує специфікації функцій прийому, доступу, адміністрування та зберігання електронних даних.

У Великобританії 1998 р. консорціум "The Consortium of University Research Libraries (Консорціум університетських дослідницьких бібліотек)" разом з організацією UKOLN (The UK Office for Library and Information Networking) приступили і виконали протягом трьох років в співробітництві з Національним архівом PRO (Public Records Office) проект типового зразку електронного архіву CEDARS (the CURL Exemplars in Digital ArchiveS)¹⁷.

У даному проекті значну увагу приділено метаданим ЕД у вирішенні проблеми довготривалого збереження з використанням стратегій емуляції та міграції. Важливим є також вирішення питання створення самих метаданих, їх збереження та управління метаданими.

Підсумки

Викладене дозволяє зробити висновок, що в Україні на початковому етапі вирішення проблеми архівного зберігання ЕД необхідним є розв'язання таких питань:

- визначення стратегії довготривалого зберігання, формату ЕД, носіїв інформації;

- розроблення структури та змісту метаданих ЕД;
- розроблення архітектури програмно-технічних засобів для реалізації архіву ЕД.

Попередньо можливо відзначити перспективні напрями - це стратегія інкапсуляції ЕД, яка на початку реалізується в системах автоматизованого діловодства і має метадані, семантична модель яких має відтворювати вітчизняні нормативні вимоги до організації діловодства та роботи архівів.

Архітектура програмно-технічних засобів для реалізації архіву електронних документів може базуватися на широко застосовуваній платформі Microsoft. Такі пропозиції, наприклад, зроблені фахівцями Російської Федерації в проекті "Электронный документооборот и делопроизводство. Решения Microsoft в области документооборота для российских органов государственной власти и местного самоуправления"¹⁸.

1. D. Bearman and J. Trant. "A Report from Archives Community" // URL: <http://www.asis.org/Bulletin/Feb-98/bearman.html>.
2. Bearman. D. "Research issues in metadata. Working Meeting on Electronic Records Research, Pittsburgh" // URL: <http://www.sis.pitt.edu/~cerar/s4-db.htm>
3. Lawrence G. W., Kehoe W. R. & others. "Risk Management of Digital Information: a file format investigation" // URL: <http://www.clir.org/pubs/abstract/pub93abst.html>
4. URL: <http://www.nationalarchives.gov.uk/default.htm> / The National Archives Records Management
5. "Основи XML" // URL: <http://www.microsoft.com/ru/ru/default.aspx>
6. URL: <http://www.nationalarchives.gov.uk/default.htm> / The National Archives Records Management/storage of semi-current records.html; <http://www.nationalarchives.gov.uk/recordsmanagement/default.htm>
7. M. Day. "Integrated Metadata Schema Registries"
8. International Council on Archives / Committee on Electronic Records: Guide for Managing Electronic Records from an Archival Perspective, ISA Studies, - Paris, 1997 // URL: <http://www.ica.org/en/node/30019>
9. Dr. Laura Millar. "Authenticity of Electronic Records: a Report prepared for UNESCO and International Council of Archives" // URL: http://www.ica.org/biblio/cds/Study13_2Erev.pdf
10. NARA, 18-th Preservation Annual // URL: <http://www.archives.gov/preservation/conferences/papers-2003/chapman.html>)
11. O. Shea G. "Keeping electronics records: issues and strategies" // URL: <http://www.netpac.com/provenance/vol1/no2/features/erecs1a.htm>

James Williams, Richard J. Cox, David Bearman. "Functional Requirements for Evidence in Recordkeeping", University of Pittsburgh, School of Information Sciences // URL: