

Державний комітет архівів України
Український науково-дослідний інститут архівної справи
та документознавства
(УНДІАСД)

КОНВЕРТУВАННЯ ЕЛЕКТРОННИХ ДОКУМЕНТІВ
ДЛЯ ПОСТІЙНОГО ЗБЕРІГАННЯ В АРХІВІ.
МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Схвалено

Протокол засідання Нормативно-
методичної комісії Укрдержархіву

30 листопада 2010 р., № 5

Київ 2011

СПИСОК АВТОРІВ

Старший відділу фізико-хімічних та
біологічних досліджень УНДІАСД, к.т.н.

П. М. Марченко

Старший науковий співробітник
відділу документознавства УНДІАСД

Ю. С. Ковтанюк

Завідувач сектора науково-технічної інформації
Науково-інформаційного відділу УНДІАСД

Р. В. Романовський

Зміст

Вступ.....	4
Терміни та визначення.....	5
Скорочення:.....	6
1 Загальні положення.....	7
2 Вимоги до форматів електронних документів постійного строку зберігання.....	8
3 Визначення переліку форматів даних електронних документів постійного строку зберігання.....	8
4 Виявлення електронних документів, які підлягають конвертуванню.....	8
5 Визначення засобу конвертування.....	9
6 Конвертування електронних документів.....	10
7 Створення АКЕД.....	10

Вступ

При зберіганні електронних документів (ЕД) відповідно до Закону України «Про електронні документи та електронний документообіг»¹ обов'язкове додержання таких вимог:

- 1) інформація, що міститься в ЕД, повинна бути доступною для її подальшого використання;
- 2) має бути забезпечена можливість відновлення ЕД у тому форматі, в якому він був створений, відправлений або одержаний;
- 3) у разі необхідності повинна зберігатися інформація, яка дає змогу встановити походження та призначення ЕД, а також дату і час його відправлення чи одержання.

Додержання цих вимог у довготривалому процесі зберігання ЕД є проблемою², адже існує моральне та фізичне старіння програмно-технічних засобів, яке призводить до втрати можливості відтворення змісту ЕД з формату, у якому ці дані були створені. Проблема відтворення інформації з ЕД постійного строку зберігання з оригінальних форматів в архіві може бути вирішена шляхом емуляції, але потребує створення достатньо складного програмного забезпечення, а таким чином і значних коштів. Уникнути вирішення цієї проблеми і зберегти доступ до інформації ЕД доцільно скориставшись завчасним конвертуванням ЕД в інші формати, які відповідають вимогам довгострокового зберігання.

Мета цієї розробки — надати методичні рекомендації Архівам, зокрема Центральному державному електронному архіву України (далі – ЦДЕА), щодо конвертування ЕД у формати, які забезпечують міжоперабельність та довгострокове зберігання ЕД.

Терміни та визначення

Архів – установа, яка здійснює постійне зберігання електронних документів.

Електронний документ – документ, інформація в якому зафіксована у вигляді електронних даних, включаючи обов'язкові реквізити документа згідно з ДСТУ 4163–2003: «Уніфікована система організаційно-розпорядчої документації» цілісність та автентичність якого підтверджена електронним цифровим підписом згідно з ДСТУ 4145:2002: «Криптографічний захист інформації: цифровий підпис, що ґрунтується на еліптичних кривих. Формування та перевіряння».

Архівний електронний документ – електронний документ, який містить метадані електронного документа та його файли в символічному кодуванні.

Інкапсульований електронний документ – електронний документ, файл якого розміщено в архівному електронному документі.

Дані електронного документа – змістовної частини електронного документа без електронного цифрового підпису.

Конвертування електронного документа – дія, результатом якої є перетворення електронного документа з існуючого формату в інший формат.

Конвертований електронний документ – дані, отримані в результаті конвертування, на які накладено електронний цифровий підпис.

Архівний конвертований електронний документ – архівний електронний документ, змістовна частина якого містить метадані та сам КЕД у символічному кодуванні.

Формат даних – спосіб представлення даних (кількість та порядок розміщення машинних кодів), які оформлені програмними засобами як одне ціле.

Основний фонд електронних документів – фонд для зберігання архівних електронних документів та відповідних їм останніх та передостанніх архівних конвертованих електронних документів.

Скорочення:

АЕД – архівний електронний документ;

АКЕД – архівний конвертований електронний документ;

ЕД – електронний документ;

ЕПК – експертно-перевірна комісія;

ЕЦП – електронний цифровий підпис;

ІТ-фахівець – фахівець з інформаційних технологій;

КЕД – конвертований електронний документ;

Перелік – перелік форматів даних електронних документів постійного строку зберігання;

ДЕД – дані електронного документа;

XML – Extensible Markup Language (розширювана мова розмітки).

1 Загальні положення

1.1 Конвертуванню підлягають ЕД постійного строку зберігання, формати яких не зазначені в Переліку.

1.2 Організація роботи з конвертування електронних документів постійного зберігання передбачає:

- визначення Переліку;
- визначення засобу конвертування;
- виявлення ЕД або відповідних їм КЕД, які підлягають конвертуванню;
- конвертування ЕД або КЕД в КЕД відповідного формату за Переліком;
- створення АКЕД.

1.2.1 Перелік визначає Держкомархів, як спеціалізована установа у даній сфері діяльності.

1.2.2 Контроль за організацією роботи з конвертування ЕД та КЕД здійснює відділ забезпечення збереженості та обліку документів Архіву.

1.2.3 Виявлення ЕД або відповідних їм КЕД, у разі наявності останніх, які підлягають конвертуванню, здійснюється шляхом пошуку ЕД та КЕД, формати даних яких не містяться в Переліку.

1.2.3.1 Даний вид роботи виконується в плановому порядку надходження ЕД до Архіву, або при зміні Переліку.

1.2.3.2 Пошук ЕД та КЕД при зміні Переліку здійснюється серед АЕД та АКЕД основного фонду ЕД.

1.2.3.2.1 Якщо для АЕД ще не створювався АКЕД, тоді аналізується формат ЕД. Якщо для АЕД вже створювалися АКЕД, тоді аналізується формат останнього КЕД

1.2.3.3 Результати пошуку ЕД та КЕД, які підлягають конвертуванню, оформлюється у вигляді доповідної записки начальнику відділу забезпечення збереженості та обліку документів Архіву. Доповідна записка є підставою для складання плану робіт з конвертування виявлених ЕД та КЕД, який подається на затвердження директору Архіву.

1.2.3.4 Розглядає та затверджує перелік ЕД та КЕД, які потребують конвертування, ЕПК Архіву.

1.2.4 Визначення засобу конвертування для кожного з форматів виявлених ЕД та КЕД у формати відповідного типу за Переліком здійснюється Держкомархівом.

1.2.4.2 Якщо Архів, який самостійно займається конвертуванням ЕД та КЕД, виявив формат, по якому не проводилося визначення засобу конвертування у відповідний формат за Переліком, на ім'я директора ЦДЕА надсилається лист із проханням визначити відповідний засіб для виявленого формату.

1.2.5 Конвертування файлів ЕД та КЕД здійснюється за рішенням ЕПК Архіву і за умови наявності відповідного програмного забезпечення в Архіві, яке дозволяє автоматизувати процедури витягання ЕД та КЕД із відповідних їм АЕД та АКЕД, виокремлення ДЕД від ЕЦП ЕД та КЕД, конвертування ДЕД у формат за Переліком, створення КЕД та створення АКЕД.

1.2.5.1 Створення КЕД відбувається в присутності членів ЕПК Архіву в ізолюваному інформаційному середовищі і завершується накладанням ЕЦП членів ЕПК Архіву.

1.2.5.2 Створення АКЕД відбувається в присутності членів ЕПК Архіву в ізолюваному інформаційному середовищі і завершується накладанням ЕЦП членів ЕПК Архіву.

2 Вимоги до форматів електронних документів постійного строку зберігання

2.1 Довговічність – формат даних ЕД, КЕД повинен мати широке застосування, підтримуватись розробником протягом щонайменше десяти років в усіх версіях програм його створення, бути стандартизованим і мати відкриту специфікацію.

2.2 Міжоперабельність – здатність поширених програмно-технічних засобів на різних платформах (операційних системах і технічних засобах) відтворювати дані ЕД, КЕД певного формату.

2.3 Самодостатність – вся інформація, необхідна для відтворення змісту ЕД, КЕД в часі у незмінному вигляді, впроваджена у ЕД та КЕД.

2.4 Ізолюваність – дані ЕД та КЕД не повинні бути якимось чином зв'язані із зовнішнім джерелом по відношенню к ЕД, КЕД: використовувати зовнішні джерела для відтворення змісту ЕД, КЕД або містить посилання на такі джерела.

2.5 Незмінність – для запобігання будь-якої автоматичної зміни інформації ЕД, КЕД за допомогою програмних засобів, формат даних не повинні містити будь-якого програмного коду або команд для запуску програмних засобів.

2.6 При аналізі за п.п. 2.1–2.5 перевага надається форматам, які підтримуються програмним забезпеченням з відкритим кодом, розповсюджуваними за ліцензіями, які не встановлюють обмежень щодо використання та копіювання програмних продуктів.

3 Визначення переліку форматів даних електронних документів постійного строку зберігання

3.1 Перелік визначається Держкомархівом (див. п. 1.2.1) і уточнюється за потребою шляхом аналізу існуючих у світовій практиці форматів даних ЕД³ на відповідність вимогам п. 2.

3.2 Моніторинг існуючих форматів на відповідність п. 2 здійснює відділ збереженості та обліку документів ЦДЕА.

3.2.1 У разі наявності змін до існуючого Переліку, ЦДЕА подає до Держкомархіву доповідну записку з обґрунтуванням необхідності внесення змін до існуючого Переліку.

4 Виявлення електронних документів, які підлягають конвертуванню

4.1 Виявлення ЕД та КЕД, які підлягають конвертуванню, здійснюється послідовно для кожного АЕД або АКЕД шляхом порівняння формату даних з

Переліком після приймання-передавання АЕД до Архіву або в основному фонді ЕД при зміні Переліку.

4.2 Виявлення ЕД та КЕД, які підлягають конвертуванню, доцільно здійснювати за метаданими ЕД або КЕД, які представлені в АЕД або АКЕД.

4.3 Результати виявлення ЕД та КЕД для конвертування документуються у формі доповідної записки.

4.3.1 Доповідна записка повинна містити назви файлів виявлених ЕД та КЕД разом з їх розширеннями.

4.3.2 Для кожного ЕД та КЕД вказуються назви файлів АЕД та АКЕД, у які інкапсульовані виявлені ЕД та КЕД.

4.3.3 Файли ЕД та КЕД групуються у доповідній записці у розділи за своїм типом: текст, растрова графіка, векторна графіка, аудіо та відео.

4.3.3.1 Відповідний розділ створюється за наявності файлів відповідного типу. Заголовком розділу слугує назва відповідного формату із Переліку.

4.3.3.2 Файли одного типу групуються у відповідному розділі за форматами, із зазначенням для кожної групи кількості файлів.

4.3.4 Кожен розділ завершується записом про загальну кількість файлів у розділі, а доповідна записка завершується записом про загальну кількість файлів виявлених ЕД та КЕД, які підлягають конвертуванню.

4.4 Рекомендується здійснювати пакетне конвертування файлів ЕД та КЕД. Пакети створюються із груп файлів одного формату, сформованих у доповідній записці.

5 Визначення засобу конвертування

5.1 Попередній вибір засобу конвертування ЕД та КЕД серед існуючих здійснюється ІТ-фахівцями ЦДЕА з орієнтацією на існуючий рівень автоматизації такого процесу в Архівах.

5.2 Остаточний вибір засобу конвертування здійснюється шляхом його апробації у такій послідовності:

5.2.1 Для вхідного формату ЕД або відповідного йому останнього КЕД визначається сумісний за ознакою призначення формату з Переліку.

5.2.2 Виконується витягання ЕД або КЕД із відповідного АЕД або АКЕД.

5.2.3 Виконується виокремлення ДЕД від ЕЦП ЕД або КЕД.

5.2.4 Аналізуються наявні параметри засобу та пропонуються їх встановлення відповідно вимогам наведеним у п. 2.

5.2.5 Здійснюється конвертування пакету файлів, який нараховує не менше десяти файлів виявлених ЕД та КЕД певного формату.

5.2.6 Здійснюється відтворення ДЕД вхідного та вихідного форматів ЕД або КЕД, проводиться їх аналіз на ідентичність із зазначенням в протоколі апробації отриманого результату для кожного із вхідних та вихідних форматів та обраного засобу конвертування.

5.2.7 Ідентичність ЕД або вхідного КЕД та нового КЕД встановлюється за правилами архівної практики та діловодства, тобто за результатами аудіо-візуального та посимвольного аналізу відтворених ЕД та КЕД.

5.2.8 У разі наявності одиночних невідповідностей вихідних ДЕД вхідним ДЕД переходять до дій за п. 5.1.

5.2.9 За результатами повної відповідності вихідного ДЕД вхідному ДЕД ЕПК ЦДЕА приймає рішення подати випробуваний засіб на перевірку та сертифікацію до спеціалізованої установи у даній сфері діяльності.

5.2.10 Після успішної перевірки та сертифікації визначеного засобу ЕПК ЦДЕА рекомендує випробуваний засіб Держкомархіву для його затвердження як засобу конвертування існуючого формату у відповідний формат за Переліком в архівній галузі..

6 Конвертування електронних документів

6.1 Робота з конвертування має плановий характер, який відповідає обсягу робіт з конвертування виявлених ЕД та КЕД (див. п. 1.2.3.3).

6.2 Виявлені ЕД та КЕД конвертуються в інший формат за Переліком в пакетному режимі (див. п. 4.4) опрацювання ДЕД певного формату ЕД та КЕД, за схемою зазначеною у пп. 5.2.1-5.2.7.

6.4 Якщо під час конвертування будуть виявлені невідповідності вихідних ДЕД вхідним ДЕД, ЕПК Архіву складає звіт про наявні помилки конвертування, який є підставою для вибору нового засобу конвертування для цих форматів ЕД та КЕД.

6.5 Рекомендується для індексації КЕД використовувати імена відповідних ЕД, що дозволить встановити однозначний зв'язок оригінального та конвертованого документів.

6.6 Цілісність та автентичність файлу КЕД підтверджується накладанням ЕЦП осіб, визначених для цього ЕПК (див. п. 1.2.5.1).

6.6.1 Формат ЕЦП повинен відповідати формату ЕЦП з повним набором даних для перевірки, що відповідає вимогам довгострокового зберігання.

6.7 За результатами планового конвертування ЕД та КЕД ЕПК складає загальний звіт про конвертування, який подає у відділ збереженості та обліку документів Архіву.

6.7.1 Форма звіту про конвертування така ж сама, як доповідної записки (див. п. 4.3). У звіті зазначається, файли яких АЕД та АКЕД були конвертовані і у які формати, і окремо вказується які файли конвертувати не вдалося із зазначенням причини.

7 Створення АКЕД

7.1 Порядок створення АКЕД визначається окремим нормативним документом Держкомархіву, який рекомендується розробити за ГСТУ⁴.

7.2 Правила іменування файлів АЕД та АКЕД визначаються визначаються окремим нормативним документом Держкомархіву⁵.

7.3 Створення АКЕД завершується накладанням ЕЦП осіб, визначених для цього ЕПК.

¹ Про електронні документи та електронний документообіг (22 травня 2003 р. № 851-IV) // <http://www.archives.gov.ua/Law-base/Laws/index.php?2003-851-4>.

² П. Марченко. Проект концепції архівного зберігання електронних документів // Студії з архівної справи та документознавства. – 2005. – Т. 13. – С. 122-129.

³ Recommended Data Formats for Preservation Purposes in the Florida Digital Archive // <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.

⁴ ГСТУ Метадані архівного електронного документа: структура та зміст.

⁵ Правила найменування файлів архівних електронних документів та архівних конвертованих копій //