УДК 681.3:002.651.028(083.73)

А.О. Мелащенко, О.Л. Перевозчикова, О.С. Скарлат

СКЛАДОВІ СТЕНДА ВАЛІДАЦІЇ АРХІВНИХ ЕЛЕКТРОННИХ ДОКУМЕНТІВ

Чинна нормативно-правова база, що регулює електронний документообіг в Україні не регламентує повний життєвий цикл електронних документів, зокрема оцінювання відповідності електронних документів вимогам архівних електронних документів. Мета роботи є деталізація алгоритму валідації архівних електронних документів за допомогою стенда. Обґрунтовано потребу створення стенда валідації архівних електронних документів та розглянуто принципи його роботи. Стенд валідації архівних електронних документів із відповідними національними стандартами формату архівних електронних документів є основою створення підсистеми керування архівними електронними правомочними й достовірними фондами, що забезпечать довічне збереження.

Вступ

Архіви електронних документів (ЕД) є складовою прийняття рішень на всіх рівнях керування. Нажаль, нині не створено технологічних і організаційних засад побудови архівів електронних документів у розумінні Закону України «Про електронні документи та електронний документообіг».

Задачу побудови архівів ЕД доцільно розділити на: забезпечення фізичної цілісності електронних активів; доведення юридичної правомочності ЕД; коректність відтворення контенту.

Головною стратегією Державної служби архівів України у забезпеченні нормативно-методичною базою Національного архівного фонду України (НАФУ) є збереження документних і мультимедійних активів як національних інформаційних ресурсів загальнодержавного значення (ДСТУ 4163-2003, ДСТУ 3843-99, ДСТУ 3844-99). У роботі описано алгоритм валідації ЕД щодо вимог до архівних електронних документів (АЕД), запропонованих у [1] (далі вимоги), і розглянуто складові стенда валідації АЕД.

Передумови впровадження стенда валідації АЕД

За умови регламентації вимог до АЕД (організаційної та технологічної складової) із застосуванням механізмів Закону України «Про стандарти, технічні регламенти та процедури оцінки відповідності» [2], процедури оцінювання відповідності програмних продуктів нормам стандартів транзитивно оцінюють їх відповідність нормам законодавства, зокрема, Закону України «Про електронні документи і електронний документообіг». Для досягнення прийнятного рівня достовірності і економічної ефективності процедур оцінювання відповідності програмних продуктів доцільно застосувати набори програм для тестування і верифікації АЕД.

Згідно з [3] держава, бізнес і громадяни застосовують достовірні і правомочні дані, надані НАФУ. Згідно з порядком [4] електронні справи містять: документ, що підлягає збереженню, засвідчувальний напис справи і опис електронної справи. Впровадження PDF/A згідно з обґрунтуванням [1] як єдиного формату АЕД автоматизує керування АЕД і спрощує використання електронних справ [4]. Формат АЕД – самодостатній засіб забезпечення довготривалого збереження та спрощує семантичну модель НАФУ (рис. 1).

Контент АЕД, базований на PDF/A, містить набір метаданих, достатніх для ефективного пошуку та організації документів у підсистемі керування електронними фондами НАФУ.

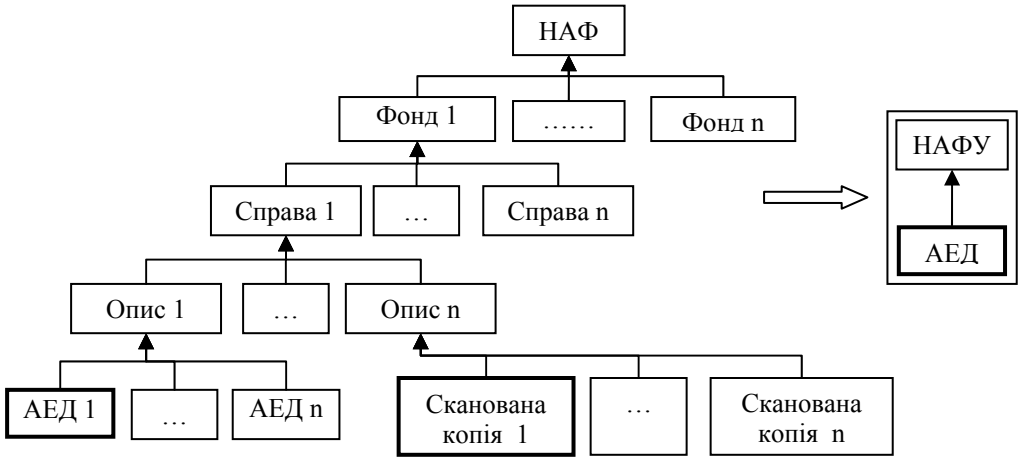

Рис. 1. Зміна семантичної моделі НАФУ з прийняттям формату АЕД

Валідація ЕД щодо вимог АЕД передбачає перевірку обов'язкових атрибутів згідно з чинною законодавчо-нормативною базою [3], структуру контенту і ЕЦП. Стенд АЕД проводить валідацію атрибутів і структури контенту документу, а також формату ЕЦП згідно з вимогами довготривалої валідації [5]. Складова ЕЦП проходить валідацію на тестовому стенді, розробленому для підтримки Технічного регламенту НСЕЦП [6].

Стандарти ETSI TS 102 778-4 і ISO 19005-1 [7] визначають основні вимоги до PDF-файлів, які мають відповідати PDF/A і підтримувати різновид ЕЦП PAdES-LTV. Стенд АЕД містить набір програм, які виконують тестування і верифікацію контенту та атрибутів ЕД на відповідність ISO 19005-1, який має наступні цілі [8]:

описати формат PDF/A, який забезпечує механізм подання ЕД щодо збереження їх вигляду незалежно від засобів і систем, використаних для створення, збереження і обробки ЕД;

забезпечити реєстрацію контексту і історії ЕД за допомогою метаданих у відповідних ЕД;

забезпечити подання логічної схеми та іншої семантичної інформації ЕД.

Деталі верифікації структури контенту АЕД

Згідно з ISO 32000-1 для розуміння синтаксису PDF, його слід розглядати з чотирьох частин: об'єкти, структуру файлу, структуру документа і потоків контенту. Документ PDF це структура даних, яка складається з набору базових типів **об'єктів даних** (булевих, числових Integer і Real, рядкових, об'єктів імен, масивів, довідників, потоків і null-об'єктів). **Структура файлу** визначає порядок зберігання, отримання доступу та оновлення об'єктів згідно з [9]. Вона не залежить від семантики об'єктів і її основна одиниця – довідник, асоціативна таблиця пар об'єктів, перший з яких є ключовим об'єктом, а другий – його значенням, яке може бути іншим об'єктом чи довідником. **Структура документа** визначає як базові типи об'єктів використовуються для представлення компонентів документа: сторінок, шрифтів. **Потоки контенту** містять послідовність інструкцій, що описують вигляд сторінок і інші графічні сутності.

Програми читання PDF файлів розпочинають роботу з заключної частини [10]. Для реалізації процедур оцінювання відповідності необхідно уточнити вимоги ISO 19005-1, оскільки цей стандарт не регламентує використання елементів, які явно не описані в ньому: «Оскільки PDF/A-1[i.3] за-

снований на форматі Adobe 1.4 PDF, а не на ISO 32000-1 [1], він не повністю підтримує всі його властивості для електронних підписів – а саме відсутні: вкладена інформація про анулювання і часовий штемпель. Однак, так як такі властивості явно не заборонені, ніщо не перешкоджає відповідній програмі запису PDF/A-1 розміщати ці розширені властивості у файл, але не слід очікувати на те, що відповідна програма читання PDF/A-1 правильно їх оброблятиме. Відповідна програма читання PDF/A-1 може реалізовувати інші функції, поза межами PDF/A-1. Це зауваження є суттєвим, оскільки ISO 32000-1 і ISO 19005-1 не підтримують довготривалі ЕЦП, специфіковані у ETSI TS 102 778-4, але явно їх не забороняють.

В ISO червні 2011 року вийшов PDF/A-2 ISO 19005-2, заснований на ISO 32000-1. Стандарт ETSI TS 102 778 визначає серію профілів, що описують використання ЕЦП у PDF для забезпечення розширених ЕЦП для підписання PDF документів. З повною підтримкою ЕЦП з розширеними властивостями ETSI TS 102 778 і PDF/A-2 доцільно використовувати для надійного довготривалого архівування підписаного в електронній формі та заснованого на PDF електронного контенту.

Для гарантування довготривалої валідації (Long-Term Validation – PAdES-LTV) формат ЕЦП АЕД має відповідати ETSI TS 102 778-4. PAdES-LTV визначає нові довідники:

безпечне сховище документа DSS (Document Security Store), призначений для збереження даних валідації підписів по мірі їх надходження;

додаткову інформацію валідації (VRI –Validation Related Information), яка пов'язує дані валідації з конкретним підписом;

довідник часового штемпеля документа (Document Time Stamp), що гарантує довічну правомочність ЕД, тобто LTV.

Довідники DSS містять покажчики на «фонд» даних валідації для всіх ЕЦП і часових штемпелів, накладених на АЕД. Часовий штемпель документа захищає DSS, пов'язуючи його з тілом документа. На рівні структури контенту документу, довідник часового штемпеля є стандартним довідником підпису зі змінами, визначеними у ETSI TS 102 778-4, таблиця. Це допомагає використати уніфіковані механізми валідації АЕД.

На рис. 2 показано послідовність отримання даних для PAdES підписів для забезпечення LTV. У разі підпису PKCS#1, сертифікат розташовано у довіднику підпису у Cert. У разі PKCS#7 сертифікат розташовано у самому бінарному DER-закодованому об'єкті ЕЦП.

Таблиця. Відповідність між довідниками стандартного ЕЦП і часового штемпелю згідно з ISO 32000-1 і ETSI TS 102 778-4

Стандартний довідник підпису	Довідник часового штемпеля	Стандартний довідник підпису	Довідник часового штемпеля
Type	**Опціональний**	M	Може бути відсутнім, дублювання
Filter	-	Location	Може бути відсутнім, дублювання
SubFilter	**Обов'язковий**	Reason	Може бути відсутнім, дублювання
Contents	**Обов'язковий**	ContactInfo	Може бути відсутнім, дублювання
Cert	Обов'язково відсутній	R	Обов'язково відсутній
ByteRange	-	**V**	**Опціональний**
Reference	Обов'язково відсутній	Prop_Build	-
Changes	Обов'язково відсутній	Prop_AuthTime	Обов'язково відсутній
Name	Може бути відсутнім, дублювання	Prop_AuthType	Обов'язково відсутній

Довідники DSS и VRI показано на рис. 3. При зміненні PDF файлу перезаписування всього попереднього контенту не відбувається, тобто початкові тіло, таблиця перехресних посилань і заключна частина залишаються без змін, а в кінець файлу приєднуються нові тіло, таблиця перехресних посилань і заключна частина з своїми довідниками.

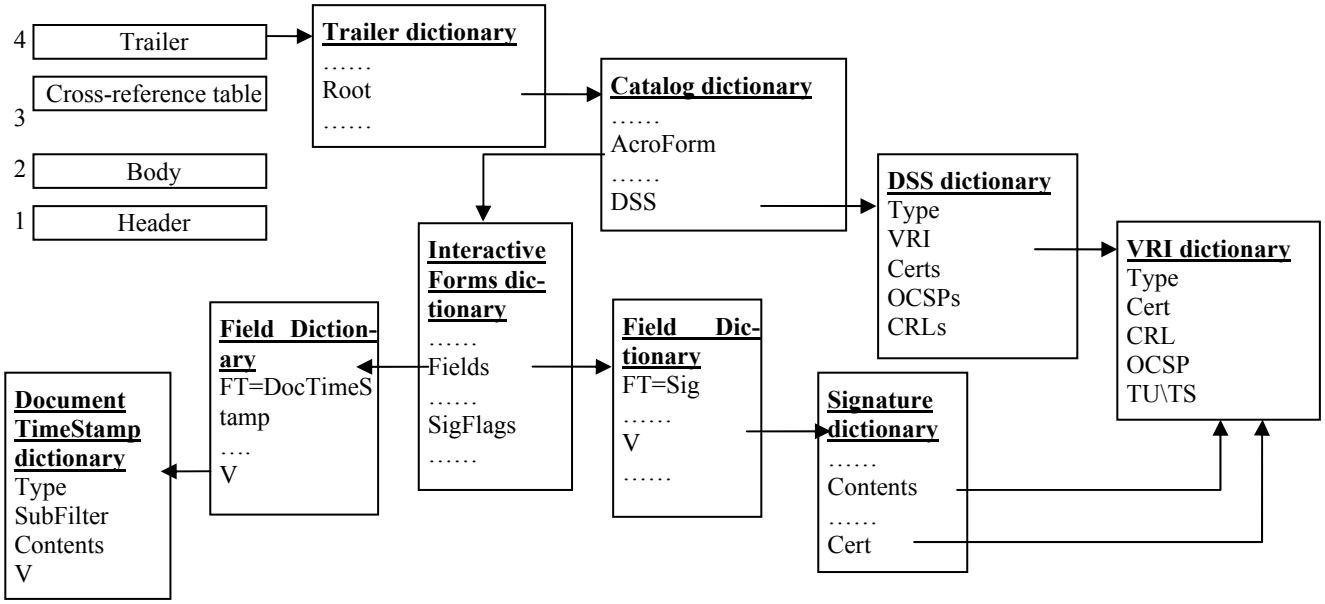

Рис. 2. Взаємозв'язок довідників ЕЦП з довідниками LTV

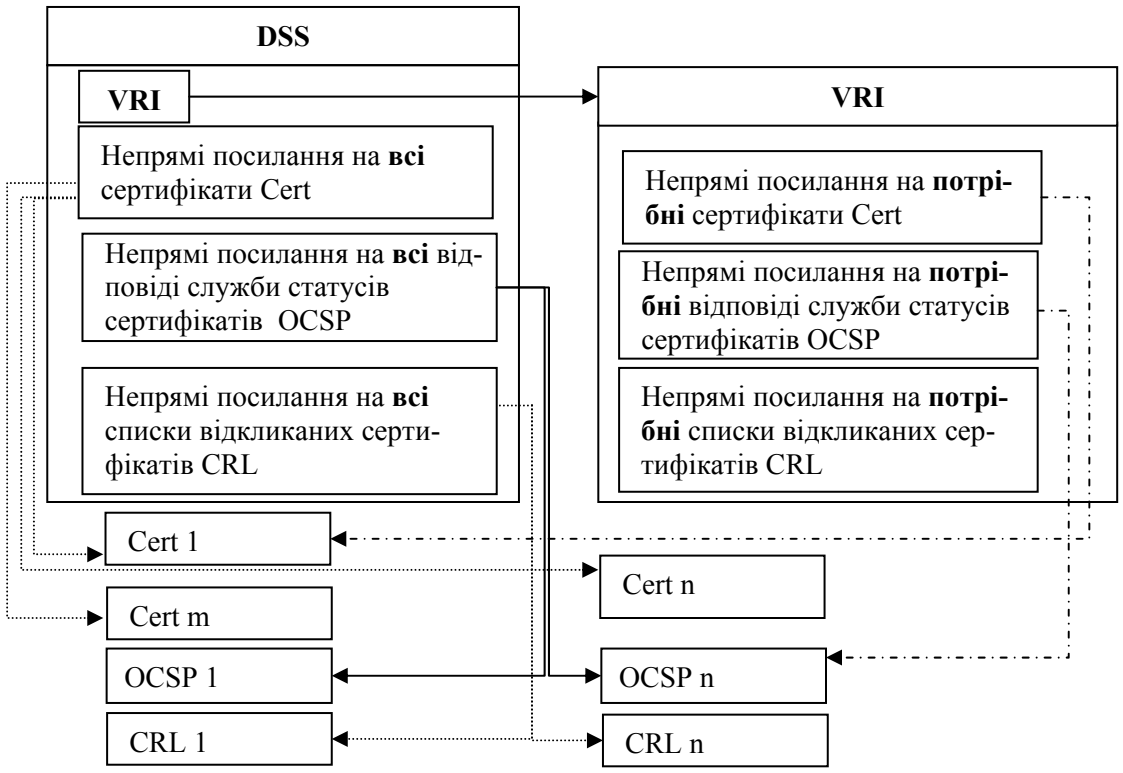

Рис. 3. Зв'язок довідників DSS і VRI, які забезпечують довготривалу валідацію підписів

У стандарті ETSI TS 102 778-4 процес валідації стосується всіх часових штемпелів і безпечних сховищ даних, присутніх у документі. Валідація "найпізнішого" часового штемпеля може проводитись в поточний час з даними валідації, зібраними в поточний час. Валідація "внутрішнього" часового штемпеля може проводитись за попереднім часовим штемпелем з даними валідації, присутніми в попередньому DSS (і з часовим штемпелем для хронологічно послідовних вкладених часових штемпелів). Валідація підпису та часового штемпеля підпису може проводитись за часом найпізнішого найбільш внутрішнього LTV часового штемпеля, використовуючи дані валідації, збережені в DSS і заштамповані часовим штемпелем (для хронологічно послідовних вкладень часових штемпелів), рис. 4. Таке послідовне накладання часових штемпелів надає документу юридичної правомочності та достовірності на довготривалий термін. Валідація АЕД без часових штемпелів не розглядається, так як це виходить за межі стандарту.

ISO 19005-1 [8] не регламентує заборони використання цих додаткових довідників, тому можна стверджувати про відсутність перешкод для відповідної програми запису PDF/A-1 вкладати ці розширені властивості в файл. Тривалість захисту може бути продовжена за межі останнього накладеного часового штемпеля документа способом додавання подальшої інформації DSS для валідації попереднього часового штемпеля разом з новим часовим штемпелем. Відповідна програма запису PDF/A файлів згідно з схемою на рис. 5 буде поміщати дані ЕЦП відповідно ETSI TS 102 778-4.

Засоби обробки підписів виділяють із всього набору даних валідації лише необхідні, відповідно до вказаних покажчиків. Стенд валідації АЕД має наступні складові:

1. Набір програмних тестів верифікації структури контенту згідно з ISO 19005-1.

2. Верифікація метаданих згідно з ДСТУ 4163-2003.

3. Перевірка наявності довідників, що забезпечують довготривалу валідацію згідно з ETSI TS 102 778-4. За їх наявності необхідно:

а) здійснити послідовний розбір, тобто виділити довідники DSS і часових штемпелів, що забезпечують LTV;

б) здійснити процес валідації довідників LTV згідно з рис. 4;

в) передати керування тестовому стенду НСЕЦП, що верифікує підписи і часові штемпелі по даним, що зібрані з довідників LTV.

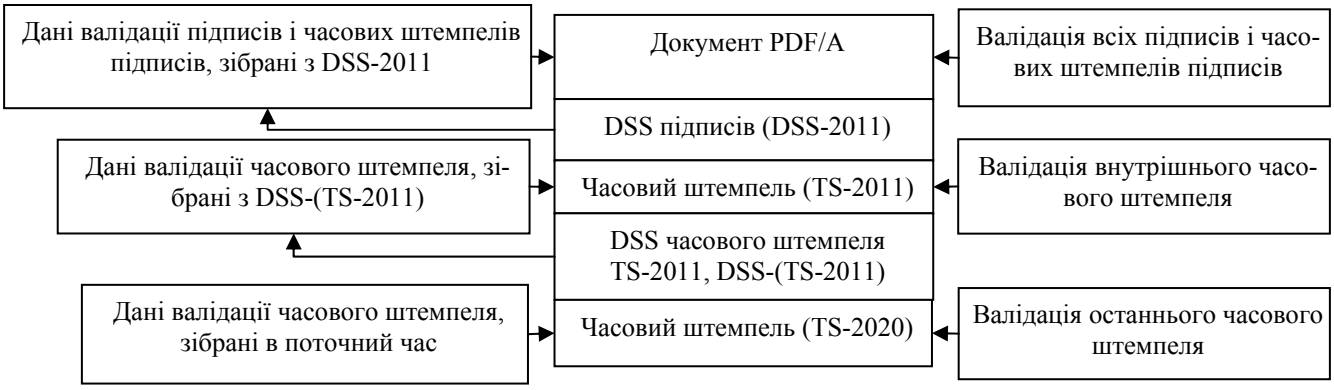

Рис. 4. Порядок процесу валідації довідників LTV

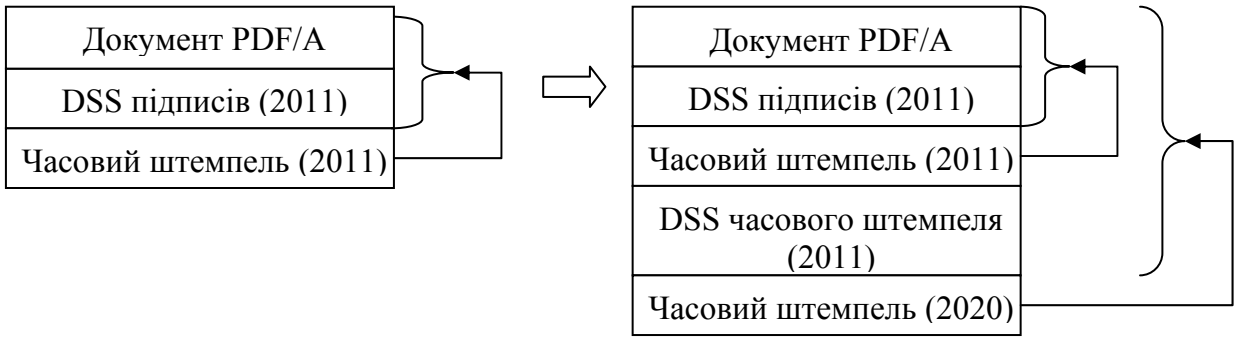

Рис. 5. Порядок додавання довідників забезпечення LTV

Приклад верифікації PAdES підписів

Верифікація ЕЦП – останній етап процесу оцінювання файлу на відповідність PDF/A. Розглянемо приклад верифікації PAdES підписів [5]. Для детальної візуалізації взаємодії довідників DSS, VRI і часового штемпеля документа з рис. 5 слід уявити сам PDF документ, який підписаний декількома PAdES підписами згідно ETSI TS 102 778-3:

– *початковий документ, який слід підписати;*
– *довідник 1-го підпису (Signature Dictionary);*
– *довідник 2-го підпису;*
– *....*
– *довідник N-го підпису.*

Цей документ потрапляє до відповідного засобу обробки підписів, який після верифікації збирає весь матеріал валідації. Виходячи з того, що збереження ЕЦП має тривати мінімум п'ять років, засіб обробки підписів генерує потрібні довідники:

– *Початковий документ, який слід підписати;*
– *довідник 1-го підпису;*
– *довідник 2-го підпису;*
– *....*
– *довідник N-го підпису;*
– **DSS (покажчик на VRI_1, покажчик на VRI_2, … , покажчик на VRI_N і на весь «фонд» Даних валідації);**
– **VRI_1 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 1-го підпису);**
– **VRI_2 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 2-го підпису);**
– **…**
– **VRI_N (покажчики на ті елементи Даних валідації, що необхідні для верифікації N-го підпису);**
– **Дані валідації.**

Після валідації всіх ЕЦП, засіб обробки підписів отримує новий часовий штемпель накладає його на документ у довіднику Document Time Stamp:

– *Початковий документ, який слід підписати;*
– *довідник 1-го підпису;*
– *довідник 2-го підпису;*
– *....*
– *довідник N-го підпису;*
– DSS (покажчик на VRI_1, покажчик на VRI_2,… , покажчик на VRI_N і на весь «фонд» Даних валідації);
– VRI_1 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 1-го підпису);
– VRI_2 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 2-го підпису);
– …
– VRI_N (покажчики на ті елементи Даних валідації, що необхідні для верифікації

N-го підпису);
– Дані валідації;
– ***1-ий часовий штемпель LTV.***

Через деякий термін з'являється необхідність накласти новий часовий штемпель. У цьому випадку відповідний засіб обробки підписів має зібрати всі дані валідації попереднього часового штемпеля, додати їх до PDF документа та додати новий довідник з новим часовим штемпелем:

– *Початковий документ, який слід підписати;*
– *довідник 1-го підпису;*
– *довідник 2-го підпису;*
– ....
– *довідник N-го підпису;*
– DSS (покажчик на VRI_1, покажчик на VRI_2, … , покажчик на VRI_N і на весь «фонд» Даних валідації);
– VRI_1 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 1-го підпису);
– VRI_2 (покажчики на ті елементи Даних валідації, що необхідні для верифікації 2-го підпису);
– …
– VRI_N (покажчики на ті елементи Даних валідації, що необхідні для верифікації N-го підпису);
– Дані валідації;
– *1-ий часовий штемпель LTV;*
– **DSS (покажчик на дані валідації 1-го часового штемпеля);**
– **Дані валідації (для 1-го часового штемпеля);**
– **2-ий часовий штемпель LTV.**

Висновки

Розглянуто основні складові компоненти стенда валідації АЕД. Написання специфікацій програмних тестів на основі нормативних документів, насамперед гармонізованих ДСТУ, і подальші їх реалізації – необхідні заходи щодо становлення архівної системи, забезпечуючи довготривале зберження правомочних АЕД. Валідація АЕД функціонально поєднує стенд АЕД і тестовий стенд НСЕЦП.

1. *Мелащенко А.О., Перевозчикова О.Л., Скарлат Е.С.* Формат долгосрочного хранения электронных документов // Компьютерная математика. – 2011. – № 1. – С. 55–64.
2. *Закон* України № 3164 від 01.12.2005 «Про стандарти, технічні регламенти та процедури оцінки відповідності».
3. *Закон* України № 3814 від 24.12.1993 «Про Національний архівний фонд та архівні установи».
4. *Наказ* Державного комітету архівів України 25.04.2005 № 49 Зареєстровано в Міністерстві юстиції України 7 червня 2005 р. за № 627/10907 «Порядок зберігання електронних документів в архівних установах».
5. *ETSI* TS 102 778-4 V1.1.2 (2009-12) Electronic Signatures and Infrastructures (ESI); PDF Advanced Electronic Signature Profiles; Part 4: PAdES Long Term - PAdES-LTV Profile.
6. *Мелащенко А.О., Перевозчикова О.Л.* Організація кваліфікованої інфраструктури відкритих ключів. – К.: Наукова думка, 2010.
7. *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).
8. *Мелащенко А.О., Перевозчикова О.Л., Скарлат О.С.* Формат файлів для довгострокового збереження електронних документів. // Матеріали VI Міжнар. наук.-практ. конф. "Наука і соціальні проблеми суспільства: інформатизація і інформаційні технології". – Харків: ХНУРЕ, 2011. – С. 211–212.
9. *ISO* 32000-1:2008 Document management.- Portable document format. Part 1: PDF 1.7.
10. *Мелащенко А.О., Скарлат О.С.* Принципи функціонування тестового стенду валідації архівних електронних документів. // Матеріали V Міжнар. наук.-техн. конф. "АПКТ 2011", – Хмельницький: ХНУ, 2011. – Т.1. – С. 139–145.

Отримано 26.06.2011

Про авторів:

Мелащенко Андрій Олегович,
кандидат фізико-математичних наук,
науковий співробітник,

Перевозчикова Ольга Леонідівна,
член-кореспондент НАН України,
доктор фізико-математичних наук,
професор,
завідуюча відділом,

Скарлат Олена Сергіївна,
аспірантка,
молодший науковий співробітник.

Місце роботи авторів:

Інститут кібернетики імені В.М. Глушкова
НАН України,
03187, Киев-187,
проспект Академіка Глушкова, 40.
Тел.: 526 3603,
Email: dep145@gmail.com