

Володимир РУДЮК

АДАПТАЦІЯ КЛАСИФІКАЦІЇ ЕЛЕКТРОННИХ ДІЛОВИХ ДОКУМЕНТІВ ДЛЯ ВИКОРИСТАННЯ В МЕРЕЖІ ІНТЕРНЕТ

В Україні – так само, як і в решті країн світу – немає системи класифікації, створеної для впорядкування і використання електронних ділових* документів. На даний момент її відсутність ще не створює суттєвих проблем для вітчизняного електронного документообігу, оскільки він перебуває, по суті, ще в зародковому стані. Втім, уже зараз деякі фахівці інформаційно-документаційної сфери наголошують на необхідності створення класифікації, що сприяла б ефективному вирішенню питання “введення до обігу документальної інформації, оптимізації її використання, репрезентації можливостей архівних джерел”¹.

На жаль, в українському документознавстві – і не лише українському – поки що бракує уявлень про те, за якими характерними критеріями можна розрізнити електронні документи взагалі й ділові, зокрема. Так, на підставі розрізнення документів на структуровані, із слабкою структурою або неструктуровані неможливо створити практичну класифікацію, що відповідає б викладеним вище вимогам.

Із досвіду закордонних колег відомі невдалі спроби застосування традиційних систем класифікації, створених свого часу для роботи з документами з паперовою основою, для впорядкування масивів електронних документів та управління ними. Внаслідок того, що, зокрема, електронні документи мають особливу форму існування, а функціям електронних систем і способам дій їхніх користувачів властиві специфічні відмінності, такі спроби не ведуть до успіху. Серед причин фахівці зазначають нездатність аналітичних та універсальних класифікаційних систем пристосовуватися до зміни умов і спричинене цим їхнє тяжіння до неактуальності, “вавілонське скупчення систематизацій ключових слів” і почасти їх незрозумілість для користувачів, труднощі класифікування документів з гіперлінковими структурами тощо². Ці незручності не можна вважати прийнятними, якщо серед принципів електронного документообігу задекларовані простий пошук документів і швидкий доступ до них³.

У європейській фаховій літературі зустрічаються більш або менш зрозумілі критерії, за якими – принаймні у теорії – відбувається розрізнення електронних документів: за комунікаційними платформами (наприклад, електронне урядування), за комунікативними рівнями (зокрема, між бізнес-структурами – *Business-To-Business*), за сферами правових відносин (податки і виплати та ін.)⁴. Чи стало б можливим і доцільним використання цих критеріїв на будь-якому рівні класифікації з практичною метою – питання дискусійне. Досвід засвідчує, що ці критерії нині на практиці проігноровано. Так, Федеральний архів ФРН представляє власні фонди організаційно-розпорядчих документів, переведених в електронний формат, поділивши їх за історичними періодами, далі – за галузевими або адміністративними ознаками, насамкінець – за управлінськими структурами, і на цьому деталізація закінчується⁵. Ця система класифікації теж була створена багато років тому для зібрань документів з паперовою основою. Яким саме чином

класифікують свої документаційні фонди інші установи і підприємства, дізнатися важко. Проте в усіх випадках постає потреба поділу на класи на документному рівні, тобто для розрізнення окремих документів, причому наголошується на доцільності застосування єдиних принципів або стандартів класифікації електронних документів – для спрощення документальних комунікацій.

Попри викладені вище проблеми, можна навести успішний приклад стандартизованого класифікування електронних документів з метою забезпечення прямого доступу до них – приклад, вартий уваги, нехай навіть у цьому випадку йдеться не про ділову документацію. У сфері електронного бізнесу класифікаційні системи уже не відіграють ролі лише засобів упорядкування і структурування: “вони набувають більш широкого значення, перетворюючись на необхідний інструмент доступу до великих каталогів... Стандартизоване, однакове для всіх постачальників класифікування – це елементарна передумова для забезпечення глобального пошуку товарів на електронних ринках”⁶.

Подібний принцип має бути використаний для класифікації ділових електронних документів, адже наша мета – так само забезпечити оптимальний пошук і використання документів на електронних інформаційних ринках незалежно від того, хто є власником документів, електронних систем або постачальником документаційних послуг. При цьому важливо усвідомлювати, що класифікація має враховувати властивості електронних документів і принципи дії поширених пошукових систем; її структура та критерії пошуку мають бути зрозумілими для всіх.

На сьогодні відповідні дослідження у сфері електронних документів або не ведуться, або є маловідомими. Втім, потрібні рішення можуть підказати результати інших дослідних проектів, що мають безпосереднє або непряме відношення до класифікації документів. В українській фаховій літературі є публікації, де піддано аналізу перебіг і результати закордонних проектів з розроблення інтернаціональних моделей архівних документів, структури їх метаданих, стратегій зберігання документів в електронних архівах (австралійський – проект *VERS*, американський – *PITT* та ін.)⁷. Один із проектів Німецької національної бібліотеки присвячений аналізу й модернізації записів десяткової класифікації бібліотечних фондів відповідно до специфіки електронних пошукових систем⁸. Колеги з ФРН, що працюють над запровадженням і вдосконаленням технологій електронного урядування, запропонували схему класифікації відповідних електронних послуг з метою їх упорядкування⁹.

На основі аналізу зазначених вище публікацій та проектів на 30-й щорічній конференції Німецького класифікаційного товариства, що відбулася 8–10 березня 2006 р. у Берліні, нами було запропоновано три класифікаційні схеми, за якими може відбуватися вибір критеріїв класифікування електронних документів з урахуванням специфічних способів їх ідентифікації, пошуку і доступу. Ці схеми ґрунтуються на розрізненні за такими ознаками:

- види електронно-документаційних послуг;
- критерії інформаційного пошуку в електронних системах (*Information*

Retrieval);

- метадані електронних документів.

Схеми можуть застосовуватись окремо або доповнювати одна одну – залежно від того, які документаційні фонди підлягають класифікації (загальні або спеціальні, наприклад веб-ресурси або бази даних), який ступінь деталізації потрібен, які способи пошуку є найбільш зручними для користувачів.

1. Класифікування за видами електронно-документаційних послуг

Класифікаційна схема послуг електронного урядування у ФРН, про яку йшлося вище, не призначена для забезпечення прямого доступу до ресурсів. Проте передбачене нею укладення класів за принципом простого впорядкування “уможлиблює першу стадію пошуку потрібних послуг електронного урядування”¹⁰. Оскільки платформа електронного урядування в широкому розумінні забезпечує електронну комунікацію між адміністративними установами, організаціями і громадянами¹¹, видається цілком слушним застосувати запропоновану німецькими колегами класифікаційну схему не лише у сфері державного управління. Водночас, у цьому контексті викликають інтерес електронні послуги, у рамках яких відбувається обмін електронними діловими документами.

Кожна послуга – якщо також ідеться про інформацію і документацію – передбачає наявність виконавчих структур у відповідних сферах діяльності. Відтак розрізнення за електронно-документаційними послугами можна вважати першим логічним кроком для поділу неструктурованої маси електронних документів (зокрема, розміщених у веб-ресурсах) за сферами діяльності (приблизно так, як учинив Федеральний архів ФРН з власними електронно-документаційними фондами). Відмінність полягає в тому, що у випадку нашої класифікації йтиметься про групування не архівних ділових документів (точніше, не лише архівних), але більшою мірою активно використовуваних; інша відмінність – в тому, що сфера діяльності, зафіксована в назві класу, позначатиме не просто відповідний тематичний розділ фонду архівних документів, але вказуватиме шлях доступу до гомогенної онлайн-бази даних, що містить потрібні для розв’язання певної проблеми документи.

Аргументи на користь запропонованої схеми є такими: а) поділ за сферами діяльності подібний до поділу за галузями знань, застосовуваного відомими міжнародними класифікаційними системами (наприклад, десятковою), а відтак не повинен стати проблемою при укладанні міжнародної класифікації електронних документів; б) більшість “пересічних” користувачів електронних пошукових систем віддають перевагу інтуїтивному пошуку документів, формулюючи початкові запити за потрібною їм сферою діяльності, і групування документів за цим принципом забезпечить комфортний пошук на верхніх рівнях класифікації для всіх категорій користувачів; в) групування за сферами діяльності дасть змогу визначити номенклатуру електронних ділових документів, використовуваних у кожній сфері, та укласти відповідні реєстри. Останній аргумент є особливо актуальним для України, оскільки укладання таких реєстрів занесено до пріоритетних перспективних завдань спеціального документознавства¹².

Недоліками схеми видаються: необхідність укласти спершу переліки

електронних сервісів, лише після чого можна вдатися до відбору і групування потрібних у нашому випадку електронно-документаційних послуг; неможливість прямого доступу до документів. Коментар до першого зауваження такий: німецькі колеги, що працюють над проектами електронного урядування, уже вдалися до спроби укласти переліки відповідних послуг, розробити критерії їхньої оцінки для занесення до цих переліків, визначити типи електронних послуг електронного урядування, сформулювати принципи пошуку (навігації), а це добрий матеріал для аналізу¹³. Щодо другого “недоліку”: на цьому рівні класифікації групування відбувається з метою формування гомогенних баз даних і надання шляху доступу до них. А вже там можна буде здійснювати цілеспрямований пошук документів за критеріями інформаційного пошуку в базах даних. Саме на цих критеріях ґрунтується наступна класифікаційна схема.

2. Класифікування за критеріями інформаційного пошуку в електронних системах.

Класифікаційна система здатна забезпечити оптимальне розташування і легке віднайдення електронних документів в онлайн-базах даних лише за умови її створення з урахуванням специфіки моделей інформаційного пошуку в таких базах та орієнтування на спосіб дій їхніх користувачів. Як засвідчує аналіз, найпопулярніша в даному випадку модель – це так звана *Boolesche Modell*, за якою зіставляються зміст запиту (тут можливі комплексні запити) і документа. Тому, якщо переслідується мета створення загальнозрозумілої практичної класифікації, навряд чи є вагомими підстави зневажати принципи дії пошукових систем, які використовують дану модель.

Сучасні пошукові системи на базі *Boolesche Modell* зазнають модернізації за допомогою так званих факторів ранжування (“ranking factors”), котрі умовно можна поділити на дві групи: на залежні від запиту (“query dependent factors”, або “on-the-page criteria”) і незалежні від нього (“query independent factors”, або “off-the-page criteria”)¹⁴. Фактори другої групи призначені для підтвердження справжності й релевантності документа незалежно від змісту запиту. У нашому випадку слід виходити з того, що відбір ділових документів до баз даних відбуватиметься за жорсткими критеріями відповідності, й підтвердження їхньої релевантності не знадобиться. Натомість потрібно застосувати залежні фактори ранжування, тісно пов’язані зі змістом документа. З-поміж цих факторів вирізняються: позиція ключового слова у тексті документа, виділення ключового слова за допомогою елементів HTML-коду, тексти внутрішніх гіперлінків, територіальна прив’язка документа (*Geo-Targeting*); за ними можна ініціювати поділ електронних документів на релевантні для користувачів класи. Так, використання фактора “позиція ключового слова” дасть змогу групувати документи за видами (наприклад, скарга або податкова декларація), оскільки відповідне ключове слово є назвою документа. Фактор “територіальна прив’язка” уможливить упорядкування документів за територіальним принципом, полегшуючи електронну комунікацію з потрібними для вирішення конкретних питань органами або підприємствами. За допомогою фактора “текст гіперлінка” можна розв’язати проблему класифікування електронних документів, що містять гіперпосилання, запропонувавши розрізнення на самостійні доку-

менти або пакети взаємопов'язаних і потрібних для вирішення одного питання документів.

Чим більше документів містить база даних (тобто якщо користувач не в змозі ознайомитися з усім фондом за допомогою мінімального набору класів), тим більше факторів ранжування слід застосувати у класифікаційній системі.

До найсуттєвіших переваг цієї схеми належать специфічність запропонованих критеріїв класифікування для електронних систем обробки даних, а також прийнятність критеріїв для користувачів найпоширеніших сучасних онлайн-баз даних.

Тепер щодо недоліків. Можливо, доведеться дещо змінити структуру електронних документів для того, щоб функціональність побудованої за такою схемою класифікації була якнайвищою. Втім, тут Україна перебуває у кращому становищі, ніж, наприклад, Німеччина: оскільки вітчизняний електронний документообіг ще не розвинений і форми електронних документів не усталені, експерименти з їхньою структурою протягом певного часу відбуватимуться безболісно або навіть залишатимуться непоміченими. Інший недолік пов'язаний із специфікою *Boolesche Modell*: деякі автори стверджують, що використовуваний нею для впорядкування і пошуку ресурсів так званий контрольований лексикон малозрозумілий для користувачів¹⁵. Чи залишиться це зауваження справедливим для нашої класифікаційної системи, – разом із підтримуваними нею способами доступу до документів, – залежатиме від формулювання назв класів.

“Втім, досвід використання контрольованого лексикону дає право зробити корисний висновок про те, що самі лише слова тексту не здатні забезпечити досконалий опис документа. Серед фахівців у сфері пошукових машин відбулася дискусія щодо розширення можливостей пошуку завдяки використанню метаданих; далі слід розмірковувати над розширенням можливостей власне документів за рахунок описових даних”¹⁶. Ця цитата наближає нас до третьої класифікаційної схеми.

3. Класифікування за метаданими.

Якщо перша схема уможлиблює поділ різнорідної маси електронних документів на гомогенні бази даних, а друга дає змогу структурувати їхні ресурси, завдання третьої класифікаційної схеми полягає в деталізації на найнижчому, документному рівні, тобто в розрізненні окремих документів. По-перше, такий намір ґрунтується на уявленні про метадані як такі, що містять відомості про зміст, контекст, структуру електронного документа та в різних наборах “мають стати юридичними, діловими, організаційними, процедурними доказами автентичності для авторів, користувачів та дослідників”; по-друге, розроблення структури і змісту метаданих належить до пріоритетних завдань “на початковому етапі вирішення проблеми архівного зберігання електронних документів”²⁰. У такому разі розроблення метаданих відбуватиметься з оглядом на можливість їх використання при класифікації документів. Проте спершу слід визначити, чи можливе взагалі створення функціональної класифікації на документному рівні на основі метаданих, а якщо так, то які саме метадані варто використовувати як критерії класифікування.

Прикладом такого аналізу можуть слугувати атрибути заголовної

частини (*Header*) електронного самоописового контейнерного документа, що використовується, зокрема, для визначення його типу (*Document Type Definition*); ці атрибути представлені звичайно у популярному нині форматі XML і виконують функцію метаданих, забезпечуючи доступ до документа та його каталогізацію²¹.

За обсягом і значенням інформації атрибут “відомості про документ і про користування ним” винесемо на чільне місце. Відомості про документ – це у даному випадку його приналежність до інших об’єктів (наприклад, розташованих вище класів), дані про версії, заміну інших документів шляхом “логічного виключення” тощо. Ще одна важлива деталь: цей атрибут дає змогу простежити зв’язок електронного документа з паперовим оригіналом. Приклади відомостей про користування документом – це дані про його укладачів, цільову групу, статус розміщеної в ньому інформації, допустимі операції з документом та ін. Не виключено, що цей атрибут може стати корисним для розрізнення на будь-якому рівні класифікації.

Атрибут “відомості про захищеність документа” містить дані про право доступу і може бути використаний для класифікації документів конфіденційного характеру. Так, класифікаційна схема сервісів електронного урядування у ФРН передбачає виділення групи так званих індивідуальних послуг, коли доступ до відповідних електронних документів мають лише певні особи.

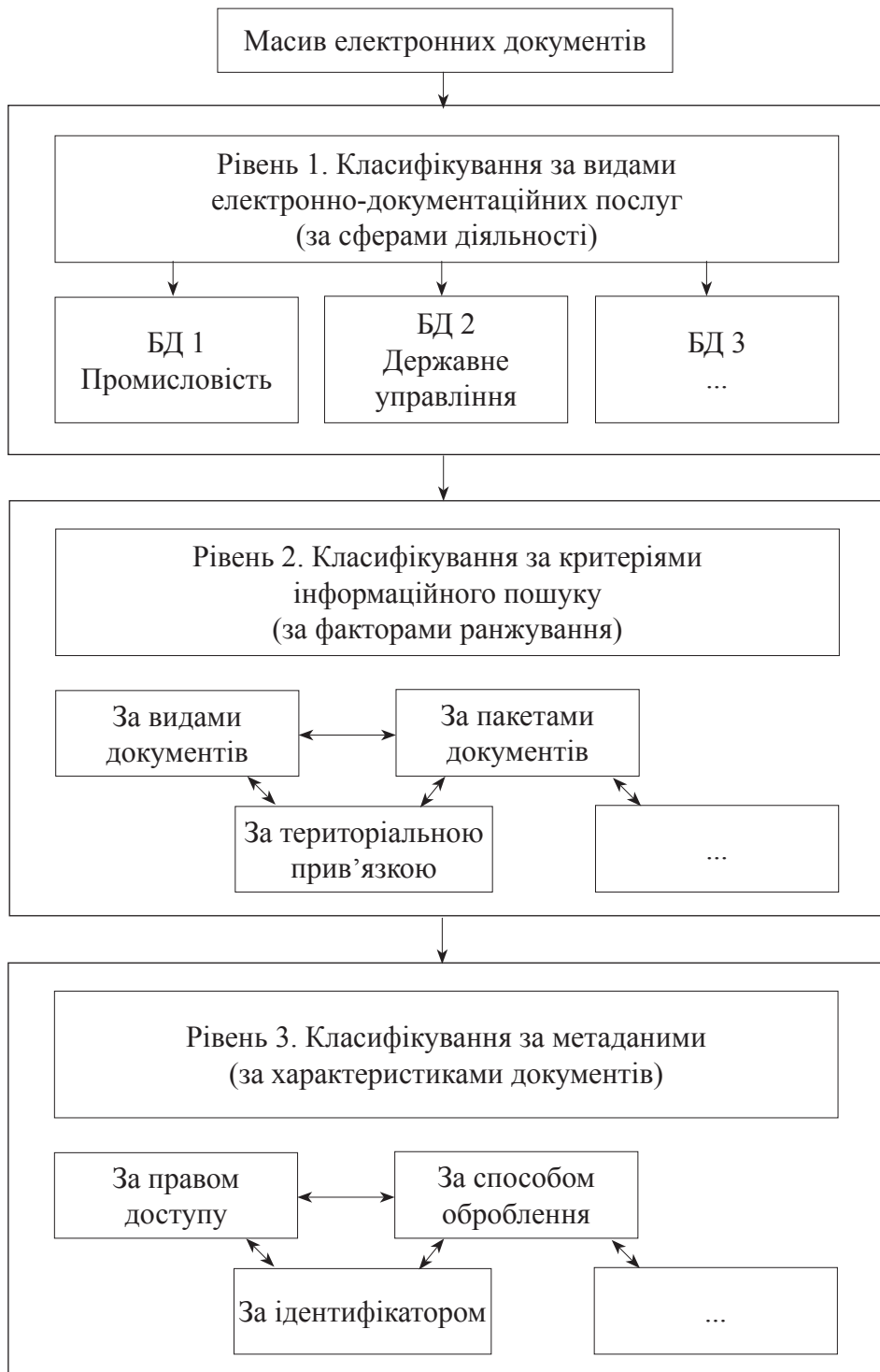
“Ідентифікація об’єкта”: цей атрибут діє звичайно за допомогою так званого універсального ідентифікатора (*Unique Identifier*), використовуваного для забезпечення доступу до об’єктів і наділеного інформацією про місце й дату їхнього створення. Завдяки цьому атрибуту можливе розрізнення документів за виконавчими структурами та їх підрозділами – для подальшої “територіальної” деталізації або навіть як альтернатива застосуванню фактора ранжування “територіальна прив’язка” у попередній класифікаційній схемі.

Інформацію про види, кількість і структуру компонентів змісту електронного самоописового документа надає атрибут “компоненти”. Якщо введений до попередньої схеми фактор ранжування “текст гіперлінка” уможливіло виділення документів з гіперлінковими структурами, то цей атрибут дає змогу розрізняти такі документи за видами додатків – для того, щоб користувач мав уявлення про потрібні йому документи та отримав доступ до пакету. Це, наприклад, відповідає вимогам якісного надання послуг на платформі електронного урядування ФРН²².

Три запропоновані класифікаційні схеми цілком вкладаються у типовий перебіг електронної комунікації між адміністративною установою та громадянином²³. Перша схема дає змогу здійснити за допомогою класифікації пошук потрібної установи, точніше – її сервера або відповідної бази даних. Друга – забезпечує вибір електронно-документаційної послуги на сервері або типу документів у базі. Третя схема дає можливість вибирати документи або цілі пакети документів за змістом, авторством та іншими відмінностями. Це можна вважати аргументом на користь застосування всіх трьох класифікаційних схем для укладання триступеневої поліієрархічної класифікації (табл. 1), що враховуватиме властивості електронних документів та неоднозначні відношення між ними, принципи дії електронних систем та уподобання їхніх користувачів.

Табл. 1.

Принципова схема триступеневої полієрархічної класифікації електронних документів



- ¹ *Бездрабко В.* Класифікація архівних документів: проблеми і пошук. http://www.archives.gov.ua/Publicat/Researches/MaterialyConfer_11.php#Bezdrabko
- ² Buder, M.; Rehfeld, W.; Seeger, T. (Hrsg.). Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die fachliche Informationsarbeit. – 5., völlig neu gefasste Ausg. – München: Saur, 2004; Lewandowski, D. Web Information Retrieval. Technologien zur Informationssuche im Internet. <http://www.durchdenken.de/lewandowski/web-ir>
- ³ Dokumentenmanagement: Was ist elektronisches Dokument? <http://www.lexikon-definition.de/Dokumentenmanagement.html>. Pernul, G., Dridi, F. Electronic Government. <http://www.uni-essen.de/unikate/pdf/18-Pernul.pdf>; Eikner, S., Schwichtenberg, H. Bits und Bytes statt Papier. <http://www.uni-essen.de/unikate/pdf/18-Eicker.pdf>
- ⁴ Pernul, G., Dridi, F.
- ⁵ Online-Findmittel des Bundesarchivs. http://www.bundesarchiv.de/bestaende_findmittel/findmittel_online/index.html
- ⁶ Dorloff, F.-D. Standards – die neue Sprache für das E-Business? <http://www.uni-essen.de/unikate/pdf/18-Dorloff.pdf>; Naumann, S., u.a. Such- und Klassifizierungsstrategien in elektronischen Produktkatalogen. <http://iss.umwelt-campus.de/download/suchundklassifikation.pdf>
- ⁷ *Марченко П.* Проект концепції архівного зберігання електронних документів. http://www.archives.gov.ua/Publicat/Studii/Studii_2005.13.03-06.php#3-2
- ⁸ Lösse, M. Vom Projekt zur Praxis: Wie geht es weiter nach dem Erscheinen der DDC Deutsch? // Advances in Data Analysis. 30th Annual Conference of the German Classification Society. – Freie Universität Berlin, March 8-10.2006
- ⁹ E-Government-Handbuch. Klassifikationsschema. http://www.bsi.bund.de/fachthem/egov/download/2_Klass.pdf, 2005.
- ¹⁰ E-Government-Handbuch. Klassifikationsschema.
- ¹¹ E-Government. [http://www.lexikon-definition.de/E-Government-\(im-weiteren-Sinn\).html](http://www.lexikon-definition.de/E-Government-(im-weiteren-Sinn).html); Особливості “електронного уряду” і фактори, що сприяють успіху http://www.kmu.gov.ua/control/uk/publish/article?art_id=10268985&cat_id=2251559
- ¹² *Кулешов С.* Перспективні напрямки наукових досліджень у спеціальному документознавстві. http://www.archives.gov.ua/Publicat/Studii/Studii_2004.12_07.php#Kuleshov
- ¹³ Fortschrittsanzeiger der Initiative BundOnline. Dienstleistungen. http://www.kbst.bund.de/cln_006/nn_836958/Content/Egov/DL/dl.html__nnn=true
- ¹⁴ Lewandowski, D. http://www.durchdenken.de/lewandowski/web-ir/?61_Rankingfaktoren.html
- ¹⁵ Cooper, W. S. Getting Beyond Boole. Information Processing & Management 24 (3), 1988.
- ¹⁶ Lewandowski, D. http://www.durchdenken.de/lewandowski/web-ir/?52_Kontrolliertes_Vokabular.html
- ²⁰ *Марченко П.* Шляхи вирішення проблеми зберігання електронних документів в архіві (аналіз Інтернет-ресурсів). http://www.archives.gov.ua/Publicat/Studii/Studii_2004.12_05.php#Marchenko
- ²¹ Dokumentenmanagement: Selbstbeschreibende elektronische Dokumente. http://www.lexikon-definition.de/Elektronisches_Dokumentenmanagement#Selbstbeschreibende_elektronische_Dokumente
- ²² E-Government-Handbuch: Ausgangsszenario. http://www.bsi.bund.de/fachthem/egov/download/4_Qualit.pdf
- ²³ E-Government-Handbuch: Typischer Ablauf (online) eines Bürger-Behörde-Kontakts (Antragsverfahren). http://www.bsi.bund.de/fachthem/egov/download/4_Qualit.pdf