

METADANE W ARCHIWIZACJI DOKUMENTÓW ELEKTRONICZNYCH

Justyna Adamus
Instytut Bibliotekoznawstwa
i Informacji Naukowej
Uniwersytetu Śląskiego w Katowicach

Archiwizacja dokumentów elektronicznych, metadane, Dublin Core, AGLS, system ERM, zarządzanie dokumentami

Archiwizacja jest tym etapem w funkcjonowaniu kancelarii¹, który zapewnia, że dokument zostaje zachowany w niezmienionej postaci dla celów bieżących administracji, a także dla przyszłych użytkowników tych dokumentów, zwłaszcza po przekazaniu materiałów archiwalnych do archiwów państwowych. Zgodnie z zapisami *Ustawy o narodowym zasobie archiwalnym i archiwach z dnia 14 lipca 1983 r.* (Dz. U. nr 38, poz. 173) materiałami archiwalnymi wchodzącymi do narodowego zasobu archiwalnego są wszelkiego rodzaju akta i dokumenty, korespondencja, dokumentacja finansowa, techniczna i statystyczna, mapy i plany, fotografie, filmy i mikrofilmy, nagrania dźwiękowe i wideofonowe, dokumenty elektroniczne w rozumieniu przepisów ustawy z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz. U. nr 64, poz. 565) oraz inna dokumentacja, bez względu na sposób jej wytworzenia, mająca znaczenie jako źródło informacji o wartości historycznej o działalności Państwa Polskiego, jego poszczególnych organów i innych państwowych jednostek organizacyjnych oraz o jego stosunkach z innymi państwami. Dokumentacja w tym rozumieniu świadczy o rozwoju życia społecznego i gospodarczego, o działalności organizacji o charakterze politycznym, społecznym i gospodarczym, zawodowym i wyznaniowym, o organizacji i rozwoju nauki, kultury i sztuki, a także o działalności jednostek samorządu terytorialnego i innych samorządowych jednostek organizacyjnych.

Metadane

Metadane towarzyszą każdemu obiektowi cyfrowemu, odsyłają do niego oraz dostarczają odpowiednich informacji opisowych, strukturalnych, prawnych

¹ Kancelaria – „komórka organizacyjna lub zespół komórek organizacyjnych w urzędzie (instytucji) zajmujących się przyjmowaniem, rejestracją i rozdzielaniem pism wpływających, przygotowywaniem, rejestracją i wysyłaniem pism wychodzących oraz przechowywaniem akt danego urzędu” (...). *Polski Słownik Archiwalny*. Pod red. W. Maciejewskiej. Warszawa 1974, s. 40.

itp.² W archiwizacji dokumentów metadane stanowią pewien klucz, gdyż opisują instytucjonalny lub administracyjny kontekst, w którym dokument powstał. Wskazują kluczowe daty, włączając w to nie tylko datę powstania, ale także datę transmisji (przekazania) wewnątrz systemu lub na zewnątrz i potwierdzenie odbioru przez adresata. Metadane określają wytwórcę dokumentu, jego autora, jego sygnaturę, link do podpisu, tytuł, streszczenie treści, związek z innymi dokumentami odnoszącymi się do tej samej sprawy, odnoszą się także do tego wszystkiego, co charakteryzuje logiczną strukturę dokumentu. Dodatkowo metadane zawierają elementy zarządzania, takie jak długość przechowywania i okres udostępniania, uwzględniają też elementy techniczne, które doprowadziły do powstania dokumentu i odnosiły się do zarządzania nim, takie jak: oprogramowanie, format danych, identyfikacja fizycznego nośnika. Metadane uwzględniają wszystkie techniczne i organizacyjne operacje na dokumencie po jego wytworzeniu i w całym cyklu jego życia.

Problematyka wytworzenia odpowiednich metadanych skupia się na metodyce opisu dokumentu, w oderwaniu od jego nośnika. Mając na uwadze to, że pierwotna forma z czasem może zostać bezpowrotnie utracona, przyjmuje się, że najważniejsza jest treść dokumentu i sposób jej prezentacji³.

W kontekście archiwalnym oraz zarządzania dokumentami metadane są ustrukturalizowanymi lub częściowo ustrukturalizowanymi informacjami, które dokumentują tworzenie, zarządzanie i użycie dokumentów w czasie oraz wewnątrz dziedzin i zagadnień. Metadane utrzymania dokumentów mogą identyfikować, uwierzytelniać i wskazywać kontekst dokumentów oraz ludzi, procesy i systemy, które tworzą, organizują i wykorzystują dokumenty.

Dokument elektroniczny w środowisku archiwalnym

Dokument elektroniczny funkcjonuje w określonym środowisku informatycznym, zwanym systemem elektronicznego obiegu dokumentów (SOD). Automatyzacja procesów biznesowych i administracyjnych, w całości lub w części, sprawia iż dokumenty, informacje lub zadania są przekazywane od jednego uczestnika do następnego, według odpowiednich procedur zarządczych (ang. *workflow*). Podczas obiegu dokumentów system musi być zdolny do przypisywania elementów metadanych automatycznie dla dokumentów, w momencie gdy są one przyjmowane do systemu i prowadzone są operacje na dokumentach. Ponadto system musi pozwalać wartościować metadane, czyli prowadzić taką ich strukturalizację, która pozwoli na ich wyszukiwanie i przechowywanie w tabelach przeglądowych w odniesieniu do innych aplikacji oprogramowania. Twórcy dokumentów mają możliwość wprowadzania w sposób manualny metadanych dla dokumentów, które nie mogą być przyjmowane automatycznie. W systemie jest prowadzona walidacja metadanych, które są wprowadzone przez użytkownika lub też nawet metadanych, które są importowane z innych systemów. Innym aspektem funkcjonowania metadanych w środowisku archi-

² M. Nahotko: *Metadane. Sposób na uporządkowanie Internetu*. Kraków 2004, s. 22.

³ W. Sylwestrzak: *XML... a informatyzacja archiwów: Po co informatyzować archiwa?* „Archeion” 2004, t. 107, s. 205-218.

walnym jest logiczne przywiązanie danych do dokumentów, jednostek i klas, dzięki czemu użytkownicy mogą przeglądać informacje zapisane w metadanych, a następnie przeszukać dokumenty, które zawierają odpowiednie dane.

Kazimierz Schmidt wyróżnił następujące rodzaje informacji, które powinny być zawarte w metadanych dla dokumentów elektronicznych:

- twórcy i współtwórcy dokumentu,
- rozmiar dokumentu (ilość znaków, ilustracji, rozmiar w bajtach itp.),
- data i czas powstania dokumentu,
- status dokumentu (roboczy, wstępnie zatwierdzony, zatwierdzony),
- data i czas zatwierdzenia dokumentu oraz osoba/instytucja upoważniona do zatwierdzenia,
- tytuł dokumentu,
- powiązania/relacje z innymi dokumentami,
- tematyka dokumentu,
- język dokumentu,
- cel powstania dokumentu (w szczególności adresaci dokumentu),
- dostępność treści dokumentu (ograniczenia dotyczące czasu lub osób upoważnionych do odczytania treści),
- prawa autorskie (czy dokument jest chroniony prawem autorskim, ewentualnie ograniczenie dotyczące czasu lub obszaru eksploatacyjnego praw autorskich),
- typ dokumentu (np. rozporządzenie, sprawozdanie, kwestionariusz, raport), kategoria archiwalna dokumentu,
- informacja o długoterminowym przechowywaniu (obowiązkowy czas przechowywania dokumentu wraz z kompletem metadanych, czas przekazania do archiwum państwowego),
- unikatowy numer dokumentu nadany w systemie teleinformatycznym,
- numer kancelaryjny dokumentu, określony zgodnie z obowiązującą w instytucji publicznej instrukcją kancelaryjną⁴.

Funkcje i rodzaje metadanych

Występuje wiele rodzajów metadanych. Marek Nahotko podzielił je ze względu na cele, jakim służą na:

- metadane administracyjne – używane w zarządzaniu i administrowaniu zasobami informacyjnymi,
- metadane opisowe – stosowane do opisu lub identyfikacji zasobów informacyjnych,
- metadane zasad i warunków – opisują one reguły użytkowania obiektów,
- metadane oceny treści – są opisem atrybutów obiektu,
- metadane proveniencji – to dane definiujące źródło lub pochodzenie treści obiektów, np. opisujące określony fizyczny obiekt, z którego pobrano treść,

⁴ Internetowe Forum Archiwalne. K. Schmidt: *Co powinno się znaleźć w rozporządzeniu dotyczącym struktury*. [online]. [dostęp 16.10.2007]. Dostępny w World Wide Web: <<http://www.lublin.ap.gov.pl/ifar/ifarforum2/index.php?topic=31.0> – doświadczenia australijskie i brytyjskie>.

- metadane o odnośnikach – relacjach i treści obiektów, często mających związek z innymi obiektami,
- metadane dla ochrony – związane z ochroną zasobów informacyjnych,
- metadane techniczne – związane ze sposobem funkcjonowania systemu lub działaniem metadanych,
- metadane strukturalne – służące głównie dla przechowywania obiektów i ich prezentacji – nawigacji, np. spis treści dla dokumentu tekstowego,
- metadane użytkownika⁵.

Elementy metadanych mogą pełnić różne funkcje, np. identyfikować zasady katalogowania wykorzystywane przy tworzeniu opisów, mogą wywoływać zachowania w określonym systemie lub aplikacji, mogą też być wykorzystywane jako terminy w zapytaniach użytkowników. Występują także elementy metadanych wyrażające daty i okresy, służą one zazwyczaj czynnościom wykonywanym automatycznie, takim jak np. sortowanie.

Metadane mogą być kontrolowane za pomocą słowników lub kartotek, ale mogą to być także metadane niekontrolowane. W systemie komputerowym przyjmują one formę nazwy lub rodzaju stwierdzenia oraz zestawu parametrów opisujących nazwę.

W kontekście archiwizacji dokumentów elektronicznych kluczową funkcją metadanych jest zapewnienie integracji i autentyczności, dzięki czemu zachowana jest wartość intelektualna, realizowana poprzez zachowanie nośnika i technologii zapewniającej ochronę struktury intelektualnej informacji w takiej postaci, w jakiej została zapisana przez autora⁶.

Integralność obiektu cyfrowego według Nahotki polega na badaniu jego wewnętrznej zawartości dla potwierdzenia autentyczności. Pewność co do integralności obiektu ograniczona jest do pewności autentyczności i integralności skrótu. Autentyczność z kolei to weryfikacja stwierdzeń związanych z obiektem, tj. stwierdzenie faktu, że obiekt jest rzeczywiście tym, czym twierdzi, że jest. Kolejna weryfikacja dotyczy kontroli dostępu, czyli tego, kto i jakie ma uprawnienia w zakresie źródła informacji. Kontrola taka odbywa się w oparciu o metody identyfikacji użytkownika. Autentyczność jest także potwierdzana przez kontrolę treści, czyli odpowiedź na pytanie o znaczenie źródła, jego proveniencję i w związku z tym opiera się na metodach identyfikacji i weryfikacji źródła. Proveniencja ukazuje pochodzenie obiektu, charakterystykę i historię pochodzenia oraz przedstawia kolejnych zarządzających obiektem i jego relacje z innymi obiektami.

Metadane są wykorzystywane przede wszystkim do:

- zapisywania informacji technicznej, wspomagającej decyzje i czynności związane z archiwizacją,
- dokumentowania zrealizowanych czynności archiwizacyjnych takich jak kopiowanie lub emulacja,
- zapisywanie efektów strategii archiwizacji,
- zapewnienie autentyczności zasobów cyfrowych bez względu na upływ czasu,
- zapisywanie informacji o zarządzaniu zbiorami i o prawach własności intelektualnej⁷.

⁵ M. Nahotko: op. cit., s. 24-26.

⁶ Ibidem, s. 39.

⁷ Ibidem, s. 52.

Powyższe rodzaje informacji dotyczą dwóch cech funkcjonalnych, tj.:

- 1) dostarczają osobom obsługującym archiwizację wiedzy do podjęcia odpowiednich działań w celu obsługi ciągu bitów obiektu cyfrowego w długim czasie,
- 2) zapewniają, że treść obiektów archiwalnych może być odpowiednio obsłużona i zinterpretowana, pomimo przyszłych zmian w technologiach dostępu.

Schemat metadanych

Schemat metadanych, który znajdujemy u Nahotki, służących archiwizacji, powinien być:

- kompletny – metadane dla archiwizacji powinny dotyczyć wszystkich aspektów procesu archiwizacji cyfrowej, począwszy od umieszczenia obiektu w archiwum, do prowadzenia udostępniania informacji dla użytkownika archiwum;
- ustrukturalizowany – schemat metadanych powinien umożliwiać szczegółowy opis podstawowych elementów i procesów funkcjonalnych systemu archiwizacji cyfrowej (jest to rozwinięcie pierwszego wymogu);
- szerokiego zastosowania – umożliwiać stosowanie schematu metadanych dla dużej ilości rodzajów obiektów cyfrowych, różnych działań związanych z archiwizacją i różnych instytucji.

Międzynarodowa Organizacja Normalizacji (ISO) stworzyła w 1999 r. model OAIIS (Open Archival Information System)⁸. Model ten określa funkcje, które mają być stosowane we wszystkich archiwach, dzięki czemu ma być możliwa realizacja archiwizacji, począwszy od pozyskania materiału do archiwizacji poprzez magazynowanie, zarządzanie i administrowanie danymi, aż po ich rozpowszechnianie na zewnątrz archiwum.

Wojciech Sylwestrzak opisał w swoim artykule standard XML (eXtensible Markup Language)⁹. Jest to formalny i faktyczny standard zewnętrzny wymiany danych, niezależny od platformy i wolny od ograniczeń licencyjnych. Standard XML jest językiem opisu dokumentów zawierających ustrukturalizowaną informację, przy czym struktura informacji dotyczy zarówno samej treści, jak i roli, w jakiej występuje ona w dokumencie. W uproszczeniu, jest to formalny język pozwalający na zdefiniowanie składni zapisywanej informacji. Format ten wywodzi się z SGML (ISO 8879), stworzonego do obsługi publikacji elektronicznych. XML powstał w 1996 r., zaprojektowała go grupa robocza XML, działająca przy Konsorcjum WWW (W3C) pod kierunkiem Jona Bosaka. XML, w przeciwieństwie do HTML, nie określa zbioru dostępnych znaczników. Format ten może służyć zarówno do zapisu składni samego dokumentu, jak też składni charakteryzujących go metadanych. Jako przykład implementacji standardu XML autor podaje standard opisu EAD. Innym przykładem jest standard RDF (Resource Description Framework), który jest formalnym standardem dla reprezentacji informacji o zasobach w sieci. Wykorzystując prosty model relacyjny, pozwala na mieszanie oraz przekazywanie ustrukturalizowanych danych między różny-

⁸ Reference Model for an Open Archival Information System (OAIIS). [online]. [dostęp 17.10.2007]. Dostępny w World Wide Web: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>.

⁹ W. Sylwestrzak: op. cit.

mi aplikacjami. Metadane RDF mogą być szeroko używane do opisu zasobów na różnych płaszczynach zastosowań, np. do opisu praw własności intelektualnej, kolekcji zdjęć, do tworzenia katalogu biblioteki, opisu listy nagrań muzycznych czy po prostu do zwiększenia skuteczności wyszukiwania w Internecie. RDF umożliwia integrację WWW na poziomie znaczeniowym, co pozwala na komputerowe przetwarzanie informacji w powiązaniu z ich treścią. RDF jest rutynowo używany do reprezentacji metadanych w postaci XML, na przykład w Dublin Core.

Dublin Core

Dublin Core to jedna z najistotniejszych inicjatyw w zakresie tworzenia i rozwijania standardów metadanych¹⁰. Prace zostały rozpoczęte w 1995 r. i są kontynuowane na corocznych warsztatach gromadzących ekspertów bibliotekarzy, informatyków, specjalistów w zakresie sieci komputerowych i bibliotecznych oraz osób pracujących nad tekstami (np. wydawców).

Elementy Dublin Core mogą być podzielone na trzy klasy (Zawartość – Content, Własność Intelektualna – Intellectual Property, Dookreślenie – Instantiation) w następujący sposób:

Title (Tytuł) – element opisujący nazwę nadaną dokumentowi; z reguły nazwę, pod którą dane zagadnienie (dokument) jest powszechnie znane;

Creator (Twórca lub Autor) – osoba, organizacja bądź usługa odpowiedzialna za stworzenie dokumentu; najczęściej używane – imię i nazwisko autora;

Subject (Opis rzeczowy) – temat poruszanego w dokumencie zagadnienia; z reguły temat będzie opisany słowami kluczowymi, kluczowymi frazami bądź kodami klasyfikacyjnymi (planuje się użycie kontrolowanych słowników i schematów klasyfikacyjnych);

Description (Opis) – element opisujący dokument i jego treść; może zawierać odniesienia do spisu treści, graficznej jego prezentacji itp.;

Publisher (Wydawca) – element przedstawiający osobę, organizację bądź usługę, która opublikowała dokument;

Contributor (Współtwórca) – osoba lub organizacja niezamieszczona w elemencie Twórca, która posiada istotny wkład intelektualny w powstanie dokumentu, lecz wkład ten jest wtórny w stosunku do osoby lub organizacji określonej w elemencie Twórca (np. redaktor, tłumacz lub ilustrator);

Date (Data) – data udostępnienia dokumentu w obecnej formie; data odnosząca się do ważnego wydarzenia w dziejach dokumentu; z reguły będzie to data powstania dokumentu, zapisana w formacie zgodnym ze standardem ISO-8601 (W3CDTF), czyli: RRRR-MM-DD (rok-miesiąc-dzień);

Type (Typ) – rodzaj dokumentu, taki jak strona domowa, powieść, poemat, dokument roboczy, raport techniczny, słownik;

Format (Format) – format danych w dokumencie, wykorzystywany do identyfikacji oprogramowania oraz czasem sprzętu potrzebnego do wyświetlenia

¹⁰ *Dublin Core Metadata Initiative*. [online]. [dostęp 6.05.2009]. Dostępny w World Wide Web: <<http://dublincore.org/>>.

i działania na dokumencie; podobnie jak Typ, Format także będzie wybierany z listy.

Identifier (Identyfikator) – ciąg znaków lub numer używany do indywidualnej identyfikacji dokumentu; przykładami dla zasobów sieciowych są URL i URN, innymi powszechnie stosowanymi identyfikatorami są ISBN i ISSN;

Source (Źródło) – ciąg znaków lub numer służący do jednoznacznej identyfikacji dokumentu, z którego bieżący dokument pochodzi; np. wersja PDF powieści w elemencie Źródło może zawierać ISBN powieści w formie książkowej, na podstawie której stworzono wersję PDF;

Language (Język) – język lub języki, w których przedstawiona została intelektualna treść dokumentu; dostępna jest pełna lista kodów języków;

Relation (Relacja) – relacje z innymi dokumentami; element ten ma służyć wyrażaniu relacji istniejących pomiędzy dokumentami, które jednak istnieją samodzielnie, np. obrazy (ilustracje) w dokumencie, rozdziały książki lub części pliku;

Coverage (Miejsce i czas) – czasowe i/lub przestrzenne charakterystyki dokumentu;

Rights (Własność) – opis praw autorskich, copyright, lub odesłanie do serwisu dostarczającego informacji o warunkach dostępności dokumentu¹¹.

Marek Nahotko wskazał kilka pozytywnych cech Dublin Core, które dają nadzieję na jego dalszy szybki rozwój:

- prostota – jest prosty nawet dla nieprzygotowanego użytkownika;
- spójność – dostarcza spójne kategorie metadanych dla różnych typów dokumentów;
- konsensus – DC ma charakter międzynarodowy, jest coraz powszechniej stosowany na wszystkich kontynentach;
- elastyczność – może służyć do tworzenia zarówno prostych, jak i złożonych opisów;
- dostosowawczość – wpisuje metadane w znane już i powszechnie zrozumiałe systemy, a więc może pracować w środowiskach już wcześniej powstałych i działających (takich jak tradycyjne biblioteki czy przeszukiwarki internetowe).

Standardy narodowe

W wielu krajach powstają standardy narodowe metadanych dla archiwizacji dokumentów elektronicznych; do najstarszych z nich należy standard AGLS, który powstał w Australii, w 1999 r.

Kolejnym przykładem standardu metadanych jest brytyjski e-GMS (e-Government Metadata Standard). Standard ten stworzony jest również w oparciu o Dublin Core (ISO 15838), zawiera 25 elementów podstawowych.

Według Schmidta, podstawowe założenia standardu e-GMS to:

– całkowita niezależność od oprogramowania (projekt nie może opierać się na jakiegokolwiek aplikacji czy narzędziach do budowy aplikacji);

¹¹ Ibidem.

Lista podstawowych elementów w opisie według standardu AGLS

Element	Co oznacza	Obligatoryjność
twórca	nazwa osoby lub instytucji odpowiedzialnej za stworzenie dokumentu	tak
wydawca	nazwa osoby lub instytucji odpowiedzialnej za udostępnienie dokumentu	tak
inna osoba odpowiedzialna (współtwórca)	nazwa osoby lub instytucji współodpowiedzialnej za stworzenie dokumentu, ale pełniącej drugorzędną rolę	nie
prawa	stwierdzenie praw albo wskazanie miejsca, gdzie można dowiedzieć się o prawach do dokumentu	tak
tytuł	tytuł dokumentu	tak
temat (słowa kluczowe)	słowa kluczowe według ustalonego słownika	tak
opis	swobodny opis tekstowy zawartości	nie
źródło	informacja o pochodzeniu dokumentu	nie
język	język narodowy, w jakim sporządzono dokument	nie
relacje	kopie, dokumenty podrzędne, nadrzędne, wersje itp., także powiązania z innymi dokumentami, np. czego jest częścią	nie
zasięg	obszar zainteresowania (geograficzny, administracyjny)	nie
funkcja	funkcje, jakie spełniają powiązane z dokumentem instytucje (w formie płaskiej listy określeń), np. wynagrodzenia, zarządzanie strategiczne, umowy, itp.	nie
data	data powstania dokumentu	tak
typ	określenie typu np. dokument tekstowy, obraz, dźwięk, wideo, aplikacja, obiekt fizyczny	nie
format	format danych (np. pdf, html)	nie
identyfikator	unikatowy identyfikator	tak
dostępność	informacja o możliwości uzyskania dokumentu (czy można otrzymać, w jakiej formie, do kogo się zwrócić)	tak (jeśli nie występuje identyfikator)

Źródło: K. Schmidt: *Prace przygotowawcze do archiwizacji dokumentów elektronicznych – doświadczenia australijskie i brytyjskie*. „Archeion” 2004, t. 107, s. 137.

- zapewnienie łatwości implementacji przez dokładnie opisane zasady, jakie należy stosować podczas „opatrywania” dokumentów metadanymi;
- uwzględnianie już istniejących standardów (międzynarodowych i europejskich);
- stabilność (podkreślono, że każda zmiana standardu to duże koszty implementacji);
- możliwość rozbudowy;
- potrzeba uwzględnienia wielu istniejących schematów metadanych w taki sposób, aby zminimalizować potrzebę ich restrukturyzacji;

Lista podstawowych elementów opisu e-GIF

Element	Zawartość	Obligatoryjność
data	data powstania dokumentu	tak
dostęp	dla kogo dokument jest dostępny	tak
adresat	do kogo jest adresowany	opcjonalny
agregacja	miejsce w hierarchii	opcjonalny
publiczność	kategoryzacja użytkowników	opcjonalny
współtwórca	osoba współodpowiedzialna za treść dokumentu, nie autor, jest to np. pracownik organizacji	opcjonalny
zasięg	może dotyczyć konkretnego obszaru geograficznego miasta, okręgu wyborczego	zalecany
twórca	może to być nazwa organizacji lub nazwisko autora albo osoby odpowiedzialnej	tak
opis	swobodny opis treści (pełnymi zdaniami)	opcjonalny
sygnatura cyfrowa	nie zdecydowano jeszcze	
dyspozycja	oznakowanie pozwalające na automatyzację postępowania z dokumentem – np. przekazanie do archiwum, automatyzacja itp. – coś na kształt kategorii archiwalnej	opcjonalny
format	opis formatu pliku z podaniem aplikacji, za pomocą której utworzono, jeśli jest; także opis postaci fizycznej	opcjonalny
identyfikator	nie ustalono jeszcze, jaki ma to być identyfikator, ale już napisano, że nie powinien to być adres URL, gdyż ulega on zmianie	tak, jeżeli istnieje
język	język narodowy, w jakim sporządzono dokument	zalecany
lokalizacja	gdzie przechowywany jest dokument	opcjonalny
podstawa prawna	na jakiej powstał podstawie	opcjonalny
zabezpieczenie	informacja o długoterminowym przechowywaniu i sposobie zabezpieczenia treści; zawiera także informacje o dokonanych transferach przestarzałych formatów	opcjonalny
wydawca	kto posiada prawa do udostępniania (publikacji)	tak, jeżeli istnieje
relacje	kopie, dokumenty podrzędne, nadrzędne, wersje itp., także powiązania z innymi dokumentami, np. czego jest częścią	opcjonalny
prawa autorskie	informacja o prawach do zasobu (dokumentu)	opcjonalny
źródła	informacja o pochodzeniu dokumentu	opcjonalny
status	np. draft wersja 1, draft wersja 2, wersja 1 do publikacji itp.	opcjonalny
temat	podstawowy przewodni temat zawartości – lista tematów to opisany poniżej standard GCL	tak
tytuł	tytuł dokumentu	tak
typ	np. sprawozdanie, kwestionariusz	opcjonalny

Źródło: K. Schmidt, op. cit., s. 131-132.

– zapewnienie możliwości łatwego przeszukiwania zgromadzonych danych¹².

Dla standardu została przygotowana lista tematyczna, obowiązująca w elemencie *subject*. Jest to słownik zawierający kilkaset haseł zebranych w teaurusie (12 tematów najwyższego poziomu), rozwiniętych przeważnie do poziomu trzeciego, obejmujący hasła przedmiotowe, geograficzne i nazwy własne. Tak przygotowany tezaurus pełni ważne funkcje porządkowania i wyszukiwania informacji.

Klasyfikacja dokumentów

W schematach metadanych przyjmuje się pewne rozwiązania dla klasyfikacji dokumentów, czyli systematycznej identyfikacji i organizacji dokumentów wewnątrz kategorii, stosownie do przyjętej logicznej struktury, metod i zasad proceduralnych reprezentowanych w schemacie klasyfikacji. Schemat klasyfikacji, czasami również nazywany planem teczek, jest wykresem, tabelą lub inną reprezentacją kategoryzowanych dokumentów, zazwyczaj przez klasy hierarchii oraz stosownie do systemu kodowania wyrażonego w alfabetycznych, numerycznych bądź alfanumerycznych symbolach. Uczestnicy dobrego schematu klasyfikacji: 1) prowadzą połączenia pomiędzy pojedynczymi dokumentami; 2) zapewniają, że dokumenty są nazywane w jednakowy sposób niezależnie od czasu; 3) wspomagają wyszukiwanie wszystkich powiązanych dokumentów według poszczególnych działań; 4) zakreślają właściwy czas przechowywania dokumentów; 5) określają bezpieczeństwo właściwe dla grup dokumentów; 6) określają pozwolenie użytkownika na dostęp lub działania dla poszczególnych grup dokumentów; 7) rozdzielają odpowiedzialność za zarządzanie poszczególnymi grupami dokumentów.

Jest wiele typów schematów klasyfikacji. WERM jest rekomendowany schemat oparty na wyrażeniu funkcji i transakcji organizacji, wyprowadzony z analizy procesów gospodarczych. Schemat klasyfikacji gospodarczej reprezentuje i opisuje funkcje, procesy gospodarcze, transakcje i inne elementy oraz pokazuje ich relacje. Wiele poziomów wewnątrz schematu może być zróżnicowanych w zależności od wymaganego stopnia szczegółowości. Struktura schematu jest hierarchiczna, rozwija się od ogółu do szczegółu. Każda funkcja ma procesy biznesowe i każdy proces (przypisany do funkcji) ma kategorie transakcji, które mu odpowiadają. Innymi słowy: schemat klasyfikacji bazuje na ścisłej klasyfikacji procesów gospodarczych, co oznacza, że dokumenty są klasyfikowane według przyczyny ich istnienia (ich funkcje lub sprawy gospodarcze, które powodują, że dokumenty zaczynają istnieć) bardziej niż według treści (ich przedmiot). System musi wspierać i być kompatybilny ze schematem klasyfikacji organizacji. Gdy schemat klasyfikacji nie istnieje lub jest tylko częściowo skonstruowany, lub gdy tworzymy nowy system, zaleca się, aby schemat klasyfikacji był oparty na procesach biznesowych i identyfikacji transakcji biznesowych, które tworzą dokumenty. System musi automatycznie oznaczać odpowiednie metadane kla-

¹² K. Schmidt, op. cit., s. 129.

syfikacji dla dokumentów i teczek oraz dla klas wewnątrz schematu klasyfikacji w punkcie tworzenia i przejmowania do systemu. System powinien również zapewniać autoryzację klasyfikowania, dodawania czy usuwania lub inną kontrolę i monitorowanie wszelkich modyfikacji schematu klasyfikacji.

Metadane w systemie zarządzania dokumentami

W rozważaniach nad metadanymi w archiwizacji dokumentów elektronicznych nie można również pominąć obszernej specyfikacji pod tytułem Model Requirements for the Management of Electronic Records (Moreq). Jest to zestaw wymagań dla zarządzania elektronicznymi dokumentami, utworzony dla programu IDA (Interchange of Data between Administration) Komisji Europejskiej przez Cornwell Management Consultants¹³.

Specyfikacja wskazuje na potrzebę ewidencji wszystkich działań towarzyszących dokumentom. Działania te obejmują tworzenie dokumentów, umieszczanie w systemie, przechowywanie, konserwację i udostępnianie. System ERM (Electronic Records Management) musi spełniać warunek zgodności z wymaganiami prawnymi, administracyjnymi oraz ze standardami międzynarodowymi.

Podstawowe założenie dla systemu zarządzania dokumentami jest takie, że dokument ma być ukazany przez zestawienie: zawartości, struktury, kontekstu, prezentacji (formy).

Zawartość jest prezentowana w jednym lub wielu fizycznych i/lub elektronicznych dokumentach, które przekazują informacje z akt. Dokumenty są magazynowane w taki sposób, aby przyszli użytkownicy mogli je właściwie odczytać i zrozumieć ich kontekst. Wszelkie kontrole proceduralne są prowadzone w celu uniemożliwienia wprowadzania niepożądanych zmian do dokumentów.

Opracowanie Moreq jest przeznaczone zarówno dla publicznych, jak i prywatnych organizacji. Specyfikacja skupia się na wymaganiach funkcjonalnych. Zarządzanie aktami polega na układaniu spraw w sposób strukturalny, odzwierciedlający funkcje gospodarcze. Budowa strukturalna przyjmuje układ hierarchiczny, tak powstaje schemat klasyfikacji, który może być wspierany przez tezaursus. Dokumenty mogą być łączone w sprawy na każdym poziomie w hierarchii klasyfikacji.

Każdy dokument, tworzony lub otrzymywany w działalności gospodarczej, staje się właściwym dokumentem dla systemu w momencie, gdy jest przejmowany do systemu ERM, wtedy też zostaje sklasyfikowany, czyli oznaczony kodem klasy, do której jest przyporządkowany, pozwalającym zarządzać tym dokumentem w systemie. Innymi słowy, umieszczanie dokumentów w systemie (ang. *capture*) jest to proces rejestrowania dokumentów, który polega na wybieraniu klas, do których dokumenty mają być przydzielone w hierarchii uporządkowania. Kolejne zadanie realizowane przez system w trakcie rejestrowania dokumentu to dodawanie metadanych. Taki dokument może być już bezpiecznie przechowywany w systemie ERM.

¹³ *Model Requirements for the Management of Electronic Records 2001*. [online]. [dostęp 22.04.2009]. Dostępny w World Wide Web: <<http://www.cornwell.co.uk/edrm/moreq.asp#moreqdownload>>.

System musi selekcjonować dokumenty według wszystkich zdefiniowanych funkcji i działań, proces ten może następować w sposób automatyczny lub ręczny. Typy procesów biznesowych powinny determinować metody tworzenia. Najlepszym sposobem jest automatyczna segregacja dokumentów, bez interwencji człowieka. Może to być wykonane przez proces biznesowy lub narzędzia obiegu dokumentów (ang. *workflow*).

Konieczne jest nadawanie wszystkich elementów metadanych określonych w procesie projektowania systemu oraz przywracanie ich wraz z dokumentami, do których są przypisane. Ważne jest zapewnienie, że dokumenty są związane ze schematem klasyfikacji, a także związane z odpowiednią liczbą elektronicznych jednostek. Jednostki te mogą być zdefiniowane prosto jako zestaw elektronicznych dokumentów. Jednostka (teczka) jest grupą dokumentów zebranych i przechowywanych razem ze względu na ich przynależność do tego samego przedmiotu, działania lub transakcji. Innymi słowy, jest to związek lub relacja pomiędzy dokumentami w jednostce (teczce). Elektroniczna jednostka nie musi istnieć realnie, często teczki są wirtualne i istnieją, ponieważ atrybuty metadanych przypisanych do dokumentów i aplikacja oprogramowania pozwala użytkownikom oglądać i zarządzać folderami tak, jakby fizycznie zawierały dokumenty do nich przypisane.

Rejestrowanie dokumentów w systemie odbywa się przez oznaczenie ich unikatowym numerem oraz przez datę i czas, kiedy weszły one do systemu przechowywania dokumentów. Identyfikator jest atrybutem, który oddziela indywidualne dokumenty lub jednostki wewnątrz systemu¹⁴. Zaleca się, aby identyfikator był zestawem numerów, które są generowane automatycznie. System musi utrzymywać lokalne relacje pomiędzy dokumentami oraz transakcjami, jakie one dokumentują. Możliwe powinno być także traktowanie połączonych dokumentów jako jeden dokument. Niektóre dokumenty elektroniczne, takie jak strony internetowe z grafiką lub wiadomości mailowe z załącznikami, są zbudowane z więcej niż jednego składnika. System powinien wychwytywać wszystkie te składniki oraz utrzymywać je jako jeden dokument. Oznacza to utrzymywanie relacji pomiędzy składnikami dla zapewnienia przyszłego odszukiwania i zarządzania tymi dokumentami. Czasami dokumenty mają więcej niż jedną wersję, co musi zostać zapisane. Zezwala się zarówno traktować wszystkie wersje tak jak jeden dokument lub każdą wersję jako osobny dokument. W drugim przypadku numer wersji powinien być dodany do metadanych. System ERM powinien być zdolny do przejmowania dokumentów transakcji generowanych przez inne systemy, także w przypadku dokumentów dużych rozmiarów tworzonych lub utrzymywanych w tych systemach. Dokumenty transakcji takie jak faktury czy formularze zamówień mogą często być przejmowane jako gotowy plik z innych systemów, dlatego powinny istnieć metody zapewnienia integracji danych i przejmowania wszystkich najważniejszych metadanych. W niektórych przypadkach przejmowanie wszystkich typów dokumentów może być niepraktyczne, przed zbudowaniem systemu należy więc zdecydować, które typy dokumentów muszą być przejmowane. Istotna jest także integracja (połączenie) z systemem poczty elektronicznej. Proces przejmowania musi być niezawodny, dokumenty

¹⁴ Ibidem, s. 9.

nie mogą być zagubione lub zmienione. Odpowiednie procedury powinny zapobiegać jakimkolwiek działaniom użytkowników mogącym prowadzić do zmiany treści lub miejsca dokumentów znajdujących się w aktach.

Klasyfikacja w systemie zarządzania dokumentami elektronicznymi

Klasyfikacja, jako systematyczna identyfikacja i organizacja dokumentów wewnątrz kategorii, stosownie do przyjętej logicznej struktury, metod i zasad proceduralnych reprezentowanych w schemacie klasyfikacji jest najważniejszym elementem systemu zarządzania dokumentami¹⁵. Schemat klasyfikacji, nazywany również planem teczek, jest wykresem, tabelą lub inną reprezentacją kategoryzowanych dokumentów; ma zazwyczaj układ hierarchiczny kodowany w alfabetycznych, numerycznych bądź alfanumerycznych symbolach. Dzięki dobrze zaprojektowanym schematom klasyfikacji jest możliwe:

- 1) prowadzenie relacji pomiędzy pojedynczymi dokumentami;
- 2) zapewnienie, że dokumenty są nazywane w jednakowy sposób niezależnie od czasu;
- 3) wyszukiwanie wszystkich powiązanych dokumentów przy uwzględnieniu poszczególnych działań;
- 4) określenie właściwego czasu przechowywania dokumentów;
- 5) określenie bezpieczeństwa właściwego dla grup dokumentów;
- 6) określenie dostępu użytkowników lub działań dla poszczególnych grup dokumentów;
- 7) przydzielenie odpowiedzialności za zarządzanie poszczególnymi grupami dokumentów¹⁶.

Jest wiele typów schematów klasyfikacji. WERM jest rekomendowany schemat oparty na wyrażeniu funkcji i transakcji organizacji wyprowadzonych z analizy procesów gospodarczych. Schemat klasyfikacji gospodarczej reprezentuje i opisuje funkcje, procesy gospodarcze, transakcje i inne elementy oraz ukazuje relacje między nimi. Wiele poziomów wewnątrz schematu może być zróżnicowanych zależnie od wymaganego poziomu skomplikowania działań oraz ich szczegółowości. Struktura schematu jest hierarchiczna, rozwija się od ogółu do szczegółu. Każda funkcja ma procesy biznesowe i każdy proces (przypisany do funkcji) ma kategorie transakcji, które mu odpowiadają. Innymi słowy, schemat klasyfikacji bazuje na ścisłej klasyfikacji procesów gospodarczych, co oznacza, że dokumenty są klasyfikowane na podstawie przyczyny ich powstania (ich funkcje lub sprawy gospodarcze, które powodują, że dokumenty zaczynają istnieć). Klasyfikacja skupia się na „kontekście” tworzenia i użycia dokumentów bardziej niż na zawartości samego dokumentu. System musi być kompatybilny ze schematem klasyfikacji organizacji. Gdy schemat klasyfikacji nie istnieje lub jest tylko częściowo skonstruowany, lub gdy tworzymy nowy system, zaleca się, aby schemat klasyfikacji był oparty na procesach biznesowych i identyfikacji

¹⁵ Ibidem, s. 6.

¹⁶ Ibidem, s. 16-20.

transakcji biznesowych, które tworzą dokumenty. W systemie powinny być automatycznie oznaczane odpowiednie metadane klasyfikacji dla dokumentów i teczek oraz dla klas wewnątrz schematu klasyfikacji w punkcie tworzenia i przejmwania. Autoryzacja przy klasyfikowaniu, dodawaniu, kasowaniu lub innej modyfikacji schematu klasyfikacji jest dokładnie kontrolowana i monitorowana.

Weryfikacja autentyczności i wiarygodności dokumentów elektronicznych

Kolejnym problemem zarządzania dokumentami jest zapewnienie ich autentyczności i wiarygodności. W archiwach i zarządzaniu dokumentami cel ten jest spełniony, gdy dokument jest prawdziwy¹⁷. Dla zapewnienia autentyczności dokumentu systemy, które tworzą, przejmują i zarządzają dokumentami elektronicznymi, muszą utrzymywać je w niezmienionej postaci, chronić przed przypadkowymi lub nieautoryzowanymi zmianami oraz przed ich usunięciem z pamięci. W tym celu należy podjąć następujące kroki.

Zapewnić należy odpowiedni stopień bezpieczeństwa i nienaruszalności dokumentów w systemie, ich zawartości oraz metadanych, które potwierdzają zawartość, kontekst i strukturę. Zabezpieczeniu podlega nienaruszalność dokumentu i wszystkie składniki dokumentu, co jest absolutnie konieczne dla udowodnienia jego autentyczności. System musi dokumenty chronić, a więc kontrolować dostęp do dokumentów oraz zapisywać każde wejście do systemu. W niektórych przypadkach dokument może być zmieniany i jest to część procesu biznesowego. Modyfikacje tego typu mogą być realizowane wyłącznie poprzez kontrolowane działania. Można także przenieść dokument do ERM przed jego ostatecznym zamknięciem.

Specyfikacja Moreq wskazuje, że dokumenty nie mogą być zniszczone w jakikolwiek sposób oprócz wyraźnego nakazu i zapisu tego zdarzenia. Ponadto system musi podlegać regularnym i systematycznym kontrolom dla weryfikacji integralności. Dotyczy to także oprogramowania i zabezpieczeń punktów dostępu.

Kontrola dostępu do metadanych

W systemie zalecane jest prowadzenie kontroli (śledzenia) wszelkich działań na dokumentach (ang. *Audit Trails*). Kontrola działań jest dokumentacją śladów operacji prowadzonych w systemie. W szczególności dokumentuje działania dokonane na dokumentach oraz ich metadane, od utworzenia aż po brakowanie. Te działania mogą być zapoczątkowane przez użytkowników, administratorów systemu lub przez sam system, czyli w procesie automatycznym. Kontrola dokumentuje działanie w zakresie tworzenia, migracji i zabezpieczanie poprzez transfer lub przenoszenie dokumentów, modyfikacje, usuwanie, definiowanie dostępu oraz historii użycia. Kontrola taka obejmuje także udostępnianie dokumentów.

¹⁷ Ibidem, s. 28.

Zaleca się, aby system automatycznie zapisywał kontrolę działań oraz automatycznie gromadził odpowiednie informacje. Dane te nie mogą być modyfikowane w żaden sposób. Kontrola taka musi być utrzymywana co najmniej do czasu, aż dokumenty, dla których jest prowadzona, zostaną zniszczone, a nawet później, niektóre dane kontrolne muszą być zachowane. Zapis działań na dokumentach musi być logicznie powiązany z dokumentami, użytkownicy mogą więc przeglądać informacje kontrolne, gdy wyszukują dokumenty. Dotyczy to również przypadku przechowywania dokumentów w różnych systemach. Dane te muszą być dostępne ale w granicach autoryzowanego użycia. Jest to niezbędne dla umożliwienia sprawdzenia prawidłowości działań systemu. System musi utrzymywać główną dokumentację systemu oraz kontrolę modyfikacji systemu tak długo, jak będzie to wymagane dla kontynuowania dostępu do dokumentów. Wszystkie zmiany dokonane na oprogramowaniu i sprzęcie systemu oraz parametrach administracyjnych, takich jak zmiana praw dostępu użytkownika, muszą być prowadzone w rejestrze działań. Instytucja określa, co kontrola ma na celu oraz jak ma być zorganizowana.

Niektóre dane kontroli mogą być usuwane, gdyż dokumenty, do których się odnoszą, są niszczone, zgodnie z regułami niszczenia dokumentów. System zarządzania dokumentami może decydować, że poszczególne działania nie muszą być zapisane, ale wszystkie inne muszą być zapisane w dokumentacji systemu.

W systemie ERM metadane, są ustrukturalizowanymi lub częściowo ustrukturalizowanymi informacjami, które potwierdzają tworzenie, zarządzanie i użycie dokumentów w czasie oraz wewnątrz działań¹⁸. Metadane utrzymania dokumentów mogą identyfikować, uwierzytelniać i wskazywać kontekst dokumentów oraz ludzi, procesy i systemy, które tworzą, organizują i wykorzystują dokumenty. Dla sprawności systemu zalecane są takie rozwiązania, które zapewnią zdolność systemu do wnioskowania i rejestrowania elementów metadanych automatycznie dla dokumentów, gdy są one przejmowane do systemu. Metadane muszą być wartościowane, aby były odnajdywane i przejmowane z tabeli przeglądowych w odniesieniu do innych aplikacji oprogramowania¹⁹.

Twórcy dokumentów mogą też metadane dla dokumentów wprowadzać manualnie, zwłaszcza, gdy nie mogą być nadawane automatycznie. W systemie powinna istnieć możliwość wspierania walidacji metadanych, wprowadzanych przez użytkownika lub importowanych z innych systemów. Podczas modyfikacji lub przekształcenia metadanych musi być restrykcyjnie przestrzegana autoryzacja przed dokonaniem zmian. W systemie tylko autoryzowani pracownicy mogą tworzyć, przejmować i aktualizować metadane. Kontrolowany jest dostęp do dokumentów, zgodnie z dobrze zdefiniowanymi kryteriami. Użytkownik nigdy nie może przeglądać informacji, na które nie ma pozwolenia.

W systemie zawarte są wskazówki postępowania w planie zarządzania, obejmujące m.in. automatyczne niszczenie niepotrzebnych dokumentów, zgodnie z procedurami i przepisami prawnymi oraz z autoryzowanymi i zatwierdzonymi zasadami zapisu i przechowywania dokumentów. Administrator może wpro-

¹⁸ Ibidem, s. 7.

¹⁹ Ibidem, s. 87-100.

wadzać zgodnie z tymi zasadami zmiany czy poprawki do wskazówek powiązanych z poszczególnymi grupami dokumentów w każdym punkcie życia dokumentu.

Bezpieczeństwo w systemie musi gwarantować, że²⁰ dokumenty, składniki dokumentów, kontrola operacji, metadane, linki do metadanych lub do jednostek oraz schematy klasyfikacji będą zmieniane lub przenoszone na nowy sprzęt, oprogramowanie i nośniki danych bez utraty wartości informacyjnej. Wszystkie dokumenty, które zostały eksportowane, muszą być zapisane w poprzednim miejscu zanim zatwierdzone zostanie powodzenie procesu transferu. Wszystkie braki podczas konwersji lub transferu dokumentów, muszą być odnotowane w specjalnym raporcie. Automatyczne procedury, powinny zapewniać regularne kopiowanie oraz odzyskiwanie dokumentów, jednostek, metadanych oraz schematów klasyfikacji, ale kopiowanie nie może zastępować odpowiedniej konserwacji dokumentów²¹.

Procedury systemu muszą być dobrze zdefiniowane i dokumentowane. Dokumentacja taka pomaga zapewnić dostęp do dokumentów wewnątrz systemu oraz może być użyta do weryfikacji autentyczności dokumentu. Administratorzy systemu odpowiadają za utrzymanie polityki i procedur zarządzania dokumentacją. Dokumentacja systemu powinna też uwzględniać specyfikację sprzętową i oprogramowanie. Niezawodność sprzętu i oprogramowania systemu musi być regularnie testowana. System powinien obsługiwać zarówno dokumenty tradycyjne jak i elektroniczne. W Polsce wciąż istnieje potrzeba dostosowania standardów archiwalnych do archiwizacji dokumentów elektronicznych. Przyjęte zostały rozporządzenia wykonawcze do ustawy o narodowym zasobie archiwalnym i archiwach. Standard metadanych dla dokumentów elektronicznych określa Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie niezbędnych elementów struktury dokumentów elektronicznych (Dz. U. 2006 nr 206 poz. 1517)²².

Summary

The article describes the role of metadata in electronic documents' archivisation. Types of information to be included in e-documents' metadata are characterized, as well as functions and types of metadata. Present metadata standards (Dublin Core, AGLS, E-GMS) are presented, with particular attention put on the Model Requirements for the Management of Electronic Records (Moreq).

²⁰ Ibidem, s. 82.

²¹ Więcej zob. Ibidem s. 83-86.

²² J. Adamus: *Archiwizacja dokumentów elektronicznych w administracji publicznej*. „Bibliotekarz” 2005, nr 11, s. 20-22.