

# XML i nowoczesne metody zarządzania treścią

Wykład 12: Zarządzanie dokumentami

Maciej Ogrodniczuk

MIMUW, 17 grudnia 2009

# Dlaczego zarządzanie dokumentami jest ważne?

Bo dokumenty są ważne.

90% zasobów informacyjnych firm jest przechowywanych w dokumentach, a nie w bazach danych.

(Deloitte and Touche)

# Dlaczego zarządzanie dokumentami jest ważne?

Bo dokumenty są ważne.

90% zasobów informacyjnych firm jest przechowywanych w dokumentach, a nie w bazach danych.

(Deloitte and Touche)

Rodzaje i odmiany systemów do zarządzania dokumentami:

- Web Content Management System – do zarządzania zawartością witryny internetowej,
- Enterprise Content Management System – zarządzanie dokumentami biznesowymi organizacji,
- system obiegu dokumentów kancelaryjnych,
- system publikacyjny,
- portal korporacyjny,
- system do pracy grupowej,
- elektroniczne archiwum,
- ...

„Tradycyjne” metody zarządzania dokumentami:

- obieg dokumentów papierowych (szafy, segregatory, asystentka, goniec...),
- poczta elektroniczna,
- dyskietki (brrr!),
- pen-drive'y,
- dyski sieciowe,
- ...

„Tradycyjne” metody zarządzania dokumentami:

- obieg dokumentów papierowych (szafy, segregatory, asystentka, goniec...),
- poczta elektroniczna,
- dyskietki (brrr!),
- pen-drive'y,
- dyski sieciowe,
- ...

Problemy (i rodzące się potrzeby):

- redundancja (ta sama informacja powtórzona w wielu miejscach) wobec możliwości wielokrotnego wykorzystania (ang. *reuse*),
- nieaktualne informacje,
- kłopoty ze znalezieniem właściwej informacji,
- trudności z koordynacją prac zespołu autorów/redaktorów,
- brak możliwości publikacji (i to wielomedialnej),
- brak personalizacji.

# Jeszcze więcej problemów z dokumentami

Co robić, gdy dokumenty mają:

- dużą objętość?
- spory stopień komplikacji?
- wysoką wartość?
- długi cykl życia?
- dużą częstość aktualizacji?

a ponadto funkcjonują w organizacji:

- rozproszonej geograficznie,
- wieloosobowej,
- o dużej specjalizacji?

# Jeszcze więcej problemów z dokumentami

Co robić, gdy dokumenty mają:

- dużą objętość?
- spory stopień komplikacji?
- wysoką wartość?
- długi cykl życia?
- dużą częstość aktualizacji?

a ponadto funkcjonują w organizacji:

- rozproszonej geograficznie,
- wieloosobowej,
- o dużej specjalizacji?

Rozwiązaniem są:

- systemy zarządzania dokumentami (CMS/DMS, ang. *Content/Document Management Systems*),
- systemy wyszukiwawcze (IA, ang. *Information Access*),
- systemy zarządzania wiedzą (KMS)

(albo ich prostsze odpowiedniki).

## Typowe funkcje:

- centralne składowanie dokumentów,
- lokalne kopie, synchronizowane z repozytorium,
- blokowanie dokumentów do edycji i zwalnianie blokady po edycji,
- wersjonowanie dokumentów,
- możliwość równoległej edycji dokumentów przez wiele osób i scalanie dokumentów.

## Typowe funkcje:

- centralne składowanie dokumentów,
- lokalne kopie, synchronizowane z repozytorium,
- blokowanie dokumentów do edycji i zwalnianie blokady po edycji,
- wersjonowanie dokumentów,
- możliwość równoległej edycji dokumentów przez wiele osób i scalanie dokumentów.

## Najpopularniejsze:

## Typowe funkcje:

- centralne składowanie dokumentów,
- lokalne kopie, synchronizowane z repozytorium,
- blokowanie dokumentów do edycji i zwalnianie blokady po edycji,
- wersjonowanie dokumentów,
- możliwość równoległej edycji dokumentów przez wiele osób i scalanie dokumentów.

## Najpopularniejsze:

- CVS (Concurrent Versions System),
- SVN (Subversion).

Czyli strony internetowe edytowalne „przez każdego”.

Cechy charakterystyczne:

- każdy użytkownik ma prawo do edycji,
- rozwiązanie działa bez dodatkowych wtyczek, bezpośrednio w przeglądarce,
- do tworzenia treści wykorzystuje się uproszczoną składnię znacznikową.

Przedstawiciele:

Czyli strony internetowe edytowalne „przez każdego”.

Cechy charakterystyczne:

- każdy użytkownik ma prawo do edycji,
- rozwiązanie działa bez dodatkowych wtyczek, bezpośrednio w przeglądarce,
- do tworzenia treści wykorzystuje się uproszczoną składnię znacznikową.

Przedstawiciele: MediaWiki, MoinMoin, TiddlyWiki...

- 1 repozytorium – scentralizowana, neutralna pula zasobów,
- 2 aplikacja:
  - funkcjonalność biznesowa,
  - obieg dokumentów, obsługa procesów (ang. *workflow*),
  - wyszukiwanie,
  - prezentacja/publikacja,
- 3 interfejs użytkownika:
  - system nawigacji,
  - system redakcyjny.

## Repozytorium:

- dokumentów:
  - możliwość zapisu dowolnych typów dokumentów,
  - wersjonowanie,
  - blokowanie dokumentów do edycji:
    - pesymistyczne – unikamy konfliktów za wszelką cenę, zakładamy blokadę już przy pobieraniu danych do edycji,
    - optymistyczne – konflikty nie są częste, więc możemy chronić samą operację modyfikacji,
  - specjalne wsparcie dla XML-a,

## Repozytorium:

- dokumentów:
  - możliwość zapisu dowolnych typów dokumentów,
  - wersjonowanie,
  - blokowanie dokumentów do edycji:
    - pesymistyczne – unikamy konfliktów za wszelką cenę, zakładamy blokadę już przy pobieraniu danych do edycji,
    - optymistyczne – konflikty nie są częste, więc możemy chronić samą operację modyfikacji,
  - specjalne wsparcie dla XML-a,
- metadanych (informacji o dokumencie – autorach, dacie publikacji, numerze wersji...):
  - zapis metadanych zwykle poza dokumentami – konieczność synchronizacji,
  - najchętniej: możliwość dowolnej konfiguracji metadanych (nazw, typów, etykiet, warunków prezentacji w interfejsie, warunków synchronizacji, ...)
  - czasem metadane strukturalne (listy, drzewa).

Chodzi o „automatyzację procesów biznesowych, w całości lub w części, podczas której dokumenty, informacje lub zadania są przekazywane od jednego uczestnika do następnego według odpowiednich procedur zarządczych”.

Workflow Management Coalition, [www.wfmc.org](http://www.wfmc.org)

Chodzi o „automatyzację procesów biznesowych, w całości lub w części, podczas której dokumenty, informacje lub zadania są przekazywane od jednego uczestnika do następnego według odpowiednich procedur zarządczych”.

Workflow Management Coalition, [www.wfmc.org](http://www.wfmc.org)

Dwa składniki:

- działania osób sterują procesem,
- proces steruje działaniami osób.

Chodzi o „automatyzację procesów biznesowych, w całości lub w części, podczas której dokumenty, informacje lub zadania są przekazywane od jednego uczestnika do następnego według odpowiednich procedur zarządczych”.

Workflow Management Coalition, [www.wfmc.org](http://www.wfmc.org)

Dwa składniki:

- działania osób sterują procesem,
- proces steruje działaniami osób.

Zdefiniowanie procesu to określenie co najmniej:

- kolejnych faz pracy nad dokumentami (stanów, ang. *states*),
- dozwolonych przejść między stanami,
- ról użytkowników uprawnionych do pracy nad dokumentem w danym stanie.

## 1 zarządzanie treścią:

- wszystkie zasoby dostępne dla (uprawnionych) użytkowników,
- użytkownik decyduje, z których zasobów w danej chwili korzysta,
- typowy sposób dostępu:
  - przeglądanie katalogów,
  - wyszukiwanie,

## 1 zarządzanie treścią:

- wszystkie zasoby dostępne dla (uprawnionych) użytkowników,
- użytkownik decyduje, z których zasobów w danej chwili korzysta,
- typowy sposób dostępu:
  - przeglądanie katalogów,
  - wyszukiwanie,

## 2 zarządzanie procesem:

- ściśle zdefiniowane role i kompetencje,
- użytkownik wykonuje zadania wskazane przez system,
- system zarządza przekazywaniem dokumentu kolejnym osobom,
- typowy sposób dostępu: lista zadań do wykonania.

Najczęściej spotykane elementy interfejsu:

- struktura drzewiasta  
(drzewo katalogów, kategorii, metadanych...)
- lista metadanych,
- dokumenty powiązane,
- wyszukiwarka,
- filtry wyników,
- ...

## Konfiguracja:

- typów dokumentów,
- katalogów i ich dopuszczalnej zawartości,
- metadanych,
- typów dowiązań,
- schematu przepływu prac,
- uprawnień,
- pluginów,
- ...

To nie tylko hiperlinki – to w ogólności wszystkie rodzaje powiązań między dokumentami i treścią.

Rodzaje linków:

- OO: między całymi dokumentami,
- CO: do całych dokumentów (hiperlinki, dołączanie dokumentów podrzędnych),
- CC: do fragmentów dokumentów (hiperlinki, zarządzanie wersjami i wariantami),
- jedno- lub dwukierunkowe,
- o dwóch lub więcej końcach (kotwicach),
- opisane metadanymi.

To nie tylko hiperlinki – to w ogólności wszystkie rodzaje powiązań między dokumentami i treścią.

Rodzaje linków:

- OO: między całymi dokumentami,
- CO: do całych dokumentów (hiperlinki, dołączanie dokumentów podrzędnych),
- CC: do fragmentów dokumentów (hiperlinki, zarządzanie wersjami i wariantami),
- jedno- lub dwukierunkowe,
- o dwóch lub więcej końcach (kotwicach),
- opisane metadanymi.

Miejsce przechowywania:

- pełna informacja w dokumencie,
- identyfikatory w dokumencie, pełna informacja w bazie,
- komplet informacji w bazie danych (ze ścieżkami do fragmentów dokumentu).

Cel: możliwość powrotu do wcześniejszej wersji.

Wielopoziomowe tworzenie wersji:

- tworzone automatycznie w momencie zapisu dokumentu (ang. *revisions*):
  - przy każdym zapisie,
  - w momencie zwolnienia blokady,
- wydania tworzone na żądanie (ang. *releases*):
  - w dowolnym (przełomowym) momencie życia dokumentu,
  - w momencie publikacji – wszystkie składniki dokumentu są zwykle „zamrażane”.

Cel: możliwość powrotu do wcześniejszej wersji.

Wielopoziomowe tworzenie wersji:

- tworzone automatycznie w momencie zapisu dokumentu (ang. *revisions*):
  - przy każdym zapisie,
  - w momencie zwolnienia blokady,
- wydania tworzone na żądanie (ang. *releases*):
  - w dowolnym (przełomowym) momencie życia dokumentu,
  - w momencie publikacji – wszystkie składniki dokumentu są zwykle „zamrażane”.

Ważne: nie tylko dokumenty, ale także:

- metadane,
- linki,
- ...

Warianty to dokumenty różniące się między sobą w niewielki sposób i będące zwykle związane jakąś relacją semantyczną.

Przykłady:

- nowelizacje aktów prawnych,
- dokumentacja kolejnych wersji urzędnienia.

Idea: zapobieganie redundancji części dokumentu wspólnych dla wszystkich wariantów.

Do 17 grudnia 2009:

...

```
<artykuł nr="212">  
  Przywóz wody nie jest  
  objęty cłem.</artykuł>
```

...

Po 17 grudnia 2009:

...

```
<artykuł nr="212">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>
```

...

## Warianty treści: przykład

Do 17 grudnia 2009:

```
...  
<artykuł nr="212">  
  Przywóz wody nie jest  
  objęty cłem.</artykuł>  
...
```

Po 17 grudnia 2009:

```
...  
<artykuł nr="212">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>  
...
```



Dokument wynikowy:

```
...  
<wariant id="a3819"/>  
...
```

## Warianty treści: przykład

Do 17 grudnia 2009:

```
...  
<artykuł nr="212">  
  Przywóz wody nie jest  
  objęty cłem.</artykuł>  
...
```

Po 17 grudnia 2009:

```
...  
<artykuł nr="212">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>  
...
```

→

Dokument wynikowy:

```
...  
<artykuł nr="212"  
  do="2009-12-17">  
  Przywóz wody nie jest  
  objęty cłem.  
</artykuł>  
<artykuł nr="212"  
  od="2009-12-18">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>  
...
```

Czy system zarządzania dokumentami powinien być jednocześnie systemem publikacyjnym (zawierać moduł publikacyjny)?

- + przecież publikacja jest celem!
- + po co przechowywać informację zwrotną z systemu składu (np. informacji o łamaniu stron) poza CMS-em?
- repozytorium jest niezależne, skład powinien odbywać się w wyspecjalizowanym systemie,
- procesy zarządzania dokumentami nie powinny być podporządkowane kształtowi przyszłej publikacji.

## Zarządzanie:

- strukturą witryny (serwisami, kolekcjami stron, odnośnikami do innych stron),
- zawartością tekstową witryny (artykułami, stronami),
- wyglądem witryny (szablony stron).

## Specyfika:

- dostępność w postaci aplikacji webowej,
- zawartość: zwykle fragmenty (X)HTML-a/XML-a i obrazki,
- łatwość tworzenia i publikacji treści,
- proces obsługi publikacji,
- funkcja podglądu nieopublikowanych treści,
- serwowanie dynamiczne lub eksport do struktur statycznych.

Obsługa kancelarii – jednostki organizacyjnej odpowiedzialnej za przyjmowanie, dystrybucję wewnętrzną i wysyłkę dokumentów.

Specyfika:

- proces zgodny z instrukcją kancelaryjną – szczegółowym opisem procedur obowiązujących w danej instytucji,
- klasyfikacja dokumentów zgodna z jednolitym rzeczowym wykazem akt (JRWA),
- wsparcie informatyczne w dwóch wariantach:
  - tradycyjny obieg dokumentów papierowych, podczas gdy system przechowuje metadane dokumentów (metryczki):
    - identyfikacja za pomocą kodów kreskowych, RFID, ...
    - w systemie zapisane jest miejsce przechowywania dokumentów papierowych (regał/półka),
  - obieg dokumentów elektronicznych:
    - wytwarzane dokumenty powstają w formie elektronicznej,
    - dokumenty papierowe są skanowane (i ew. OCR-owane), a następnie zapisywane w systemie.

## Specyfika:

- proces zgodny z instrukcją archiwizacyjną – szczegółowym opisem procedur archiwizacji dokumentów,
- rejestracja dokumentów na podstawie spisów zdawczo-odbiorczych,
- klasyfikacja dokumentów zgodna z jednolitym rzeczowym wykazem akt (JRWA),
- kategorie archiwalne:
  - A – dokument o trwałej wartości archiwalnej, przeznaczony do przekazania do archiwum państwowego,
  - B $n$  – dokument mający czasowe znaczenie praktyczne, przechowywany w archiwum przez  $n$  lat (np. B50 – 50 lat),
  - BE $n$  – dokument podlega ekspertyzie po upływie  $n$  lat.
- kontrolowane udostępnianie zasobów archiwalnych,
- brakowanie = kontrolowane usuwanie dokumentów bez wartości archiwalnej.

## Potrzeby klienta:

System do zarządzania zasobami archiwalnymi od roku 1952 do chwili obecnej.

## Archiwum przed wdrożeniem:

- 3 km dokumentów papierowych,
- archiwum zdjęć,
- materiały audio/video (setki godzin nagrań).

## Rozwiązanie:

- system szyty na miarę (na bazie istniejących komponentów),
- specyficzny interfejs.

## Potrzeby klienta:

System do zarządzania zasobami archiwalnymi od roku 1952 do chwili obecnej.

## Archiwum przed wdrożeniem:

- 3 km dokumentów papierowych,
- archiwum zdjęć,
- materiały audio/video (setki godzin nagrań).

## Rozwiązanie:

- system szyty na miarę (na bazie istniejących komponentów),
- specyficzny interfejs.

## Cechy systemu:

- pełna obsługa procesów archiwalnych (rejestracja → zarządzanie → udostępnianie → ekspertyza → brakowanie),
- dostęp z poziomu przeglądarki,
- możliwość wymiany danych z systemami zewnętrznymi (obieg dokumentów).

## Moduły:

- repozytorium dokumentów i metadanych,
- wirtualna biblioteka,
- proste wyszukiwanie (względem metadanych),
- zaawansowane raportowanie.

## W zasadzie dwa projekty:


- **2003-2004:** implementacja i wdrożenie systemu,
- **2005-2009:** skanowanie i masowy import 7 mln stron dokumentów.

## Architektura:

- 2 x Windows 2003 Server (wcześniej: Solaris),
- PostgreSQL (wcześniej: Oracle),
- JBoss,
- macierz dyskowa 12 TB (wcześniej: jukebox DVD, 100 slotów).

Archiwum Prezydenta RP - Windows Internet Explorer

Ark Edycja Wydruk Ułubione Narzędzia Pomoc

 **Archiwum Prezydenta RP**  
oparte na technologii eXms

PRZYJMIOWANIE UODSTĘPIANIE BRAKOWANIE EKSPERTYZA PRZEKAZYWANIE INDEKSY RAPORTY ADMINISTRACJA

WPROWADZANIE I WERYFIKACJA PRZEDPOŁA EWIDENCJA

> Szczegóły spisu zdawczo-odbiorczego <

Szczegóły spisu zdawczo-odbiorczego  
„Przejdź wyżej...”

**Numery spisu**  
Numer spisu: 176  
Zespół: Kancelaria Prezydenta RP  
Jednostka organizacyjna: PVL  
Daty skrajne: 1996-1997  
Data przyjęcia akt: 2005-07-12  
Data sporządzenia spisu: 1997-09-15  
Wybrakowano:   
Przekazano do AAN:   
Uwagi:   
Zapisz zmiany

**Porządek spisu zdawczo-odbiorczego**

| Lp. | Znak teckki | Tytuł   | Daty skrajne | Kategoria archiwalna | Liczba teczek (kaset, fotografii) | Położenie (regal/półka) | Ilość m.b. | Uwagi | Status |
|-----|-------------|---|--------------|----------------------|-----------------------------------|-------------------------|------------|-------|--------|
| 1   | PVL_040     | Zaproszenia   | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 2   | PVL_045     | Kongresy, zjazdy jubileuszowe, sympozja organizowane przez różne instytucje i organizacje | 1996-1997    | A                    | 1                                 | /                       | 0.0        |       |        |
| 3   | PVL_060     | Podania i skargi  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 4   | PVL_070     | Współpraca z organami administracji państwowej  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 5   | PVL_079     | Współpraca z innymi instytucjami  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 6   | PVL_301     | Współpraca z organizacjami międzynarodowymi   | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 7   | PVL_31      | Współpraca z Ministerstwem Spraw Zagranicznych  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 8   | PVL_32      | Ambasadorowie i inni przedstawiciele RP   | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 9   | PVL_37      | Wyjazdy zagraniczne pracowników Kancelarii Prezydenta RP                                  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |
| 10  | PVL_355     | Przyjazdy, delegacje, inne  | 1997         | A                    | 1                                 | /                       | 0.0        |       |        |

zalogowany jako: rag ostatnio zalogowany: 2005-10-08 18:33:18