

Marek Nahotko
Energoprojekt Kraków S.A.

Zasady tworzenia bibliotek cyfrowych

W ostatnich latach w sposób czasem dość żywiołowy powstają w naszym kraju biblioteki cyfrowe, organizowane przez różne instytucje. Wiele z tych inicjatyw krytykowanych jest za brak stosowania zasad, które od zawsze stanowiły podstawę działalności bibliotecznej. Amerykańska NISO (National Information Standards Organization) przedstawiła przewodnik tworzenia bibliotek cyfrowych^[1], który ma realizować dwa cele:

- przedstawienie głównych składników i czynności związanych z tworzeniem ?dobrych? zasobów cyfrowych,
- przedstawienie zasad identyfikacji, organizacji i stosowania istniejącej wiedzy i źródeł w celu promowania lokalnych działań służących tworzeniu i zarządzaniu ?dobrymi? zasobami cyfrowymi.

Wyjaśnienia wymaga zastosowanie określenia dobre zasoby. Co innego oznacza bowiem ?dobre? w początkowym etapie digitalizacji zbiorów, a co innego w okresie realizacji projektów dojrzałych. W początkowym okresie, w którym znajduje się ciągle większość polskich projektów, mogą one być oceniane jako faza testowa czy eksperymentalna, służąca rozwojowi metod i technologii, eksperymentom technicznym lub organizacyjnym albo po prostu zdobyciu nowej wiedzy. W takim przypadku biblioteka cyfrowa może być uznana za dobrą, jeśli spełnia te oczekiwania, nawet, gdy jest ona tylko w minimalnym stopniu przydatna dla użytkowników, albo nawet przestaje funkcjonować po zakończeniu projektu.

Dojrzałe projekty znajdują się w następnej fazie. Mają one zaspokajać dobrze określone potrzeby odpowiednich grup użytkowników. W tym przypadku dobre zasoby są użyteczne i dostępne dla właściwie określonych grup użytkowników.

Na świecie tworzenie zasobów cyfrowych przeszło do trzeciej fazy, w której już nawet efektywne usługi informacyjne nie są wystarczające. Wynika to z rozwoju samego środowiska cyfrowego, charakteryzującego się wzrostem poziomu integracji i zaufania. Integrację wielkiej ilości rozproszonej informacji zapewnia stosowanie standardów i technologii Web, dzięki czemu użytkownicy otrzymują wszelkie potrzebne informacje w jednym miejscu. Jednocześnie coraz większa ilość informacji utrudnia wyselekcjonowanie informacji wiarygodnej ? długotrwałej i o potwierdzonej autentyczności. Obiekty cyfrowe, dotyczące ich metadane i zawierające je biblioteki cyfrowe, nie tylko muszą być postrzegane w kontekście projektów, którym służą, ale także jako elementy do wykorzystania w innych projektach. W związku z tym dobry zasób musi umożliwiać współdziałanie, wielokrotne stosowanie, niezmiennosc, weryfikację, udokumentowanie i obsługę praw autorskich.

W dalszej części artykułu przedstawione zostaną czynniki, które według autorów przewodnika NISO wpływają na ocenę bibliotek cyfrowych jako ?dobrych?. Są to:

- Zasoby (zorganizowane grupy obiektów),
- Obiekty (zdigitalizowane materiały),
- Metadane (informacja dotycząca obiektów),
- Projekty (inicjatywy służące tworzeniu lub zarządzaniu zasobami).

Jak piszą autorzy przewodnika, pominięto czynniki wpływające na jakość usług, gdyż przyjęto, że warunkiem usług dobrej jakości są wartościowe zasoby, obiekty i metadane. Należy jednak podkreślić, że nie ma jedynie słusznych zasad tworzenia dobrych zasobów, obiektów i metadanych. Każdy projekt jest wyjątkowy, realizuje własne potrzeby i cele. Można zapewne stworzyć tyle sposobów kategoryzowania zasobów, ile jest związanych z nimi projektów. Kluczem do sukcesu konkretnego projektu nie jest ściśle przestrzeganie wytyczonych sposobów postępowania, ale dalekosiężne planowanie strategiczne i dokonywanie właściwych wyborów spośród wielu narzędzi i procesów realizujących specyficzne cele i potrzeby każdego zasobu.

Zasoby cyfrowe zawierają obiekty cyfrowe, wyselekcjonowane i zorganizowane w celu umożliwienia ich udostępniania i wykorzystania. Dobre zasoby cyfrowe zawierają metadane służące ich opisowi i zarządzaniu nimi. Metadane te mogą być tworzone dla zasobu jako całości i/lub dla poszczególnych pozycji w zasobie. Oprócz obiektów i metadanych biblioteka cyfrowa powinna posiadać przynajmniej jeden interfejs pozwalający na dostęp do niej. Interfejs ten często określa sposób wyszukiwania obiektów, przeszukiwania ich kategorii, udostępniania i wykorzystywania. Zasoby tworzone są przez organizacje lub grupy współpracujących organizacji, często jako efekt realizacji projektu.

Oto zasady tworzenia dobrych zasobów cyfrowych:

Zasada 1: *Dobry zasób cyfrowy jest tworzony na podstawie jasno określonej polityki jego rozwoju, uzgodnionej i udokumentowanej przed rozpoczęciem digitalizacji.*

Rozwój zasobów zależy przede wszystkim od własnych celów organizacji, będącej jego twórcą. Polityka rozwoju zasobu powinna wpływać na realizację misji organizacji. Należy określić celowego odbiorcę zasobów (aktualnego i przyszłego) i sposób zaspokajania jego potrzeb przez wyselekcjonowane materiały. Nie należy również zapominać o tym, aby zasoby cyfrowe nie pozostawały w izolacji od dokumentów tradycyjnych; zasoby te powinny uzupełniać się wzajemnie.

Często bezpodstawnie zakłada się, że digitalizacja materiałów spowoduje gwałtowny wzrost ich wykorzystania. Nie zawsze tak jest; należy określić stopień wykorzystania materiałów w postaci tradycyjnej (jeśli istnieją) i zastanowić się nad przyczynami ewentualnego wzrostu wykorzystania po zdigitalizowaniu.

Błędy dotyczące braku realizacji tej zasady są często spotykane w polskich przedsięwzięciach, co świadczy o ich niedojrzałości. Przykładem może być Polska Biblioteka Internetowa^[2].

Zasada 2: *Zasoby powinny być opisane w sposób pozwalający użytkownikowi uzyskać charakterystyki zasobów, w tym ich zakres, format, ograniczenia dostępu, własność i inne informacje istotne dla określenia autentyczności, integralności i interpretacji zasobów.*

Zasoby opisujemy za pomocą metadanych (zob. dalej). Opis taki służy dwóm celom: pozwala na uzyskanie informacji o istnieniu zasobów i pomaga użytkownikowi zorientować się w charakterze otrzymanych danych. Opis zasobów w katalogu czy rejestrach pozwala także na kontrolę autorytatywną tych treści.

Należałoby zastanowić się nad możliwością tworzenia opisów na poziomie zasobów w narodowych katalogach centralnych, takich jak NUKAT czy KaRo. Polskie biblioteki cyfrowe nie są nigdzie opisywane jako całość. Na Zachodzie informacje o tamtejszych bibliotekach cyfrowych znaleźć można w takich serwisach, jak OCLC czy RLIN.

Zasoby można także rejestrować. Rejestry mogą być tworzone przez nie-katalogerów, często zawierając prostą, mało ustrukturyzowaną informację opisową. W Polsce rejestr taki prowadzony jest w ramach EBIB-u (<http://ebib.oss.wroc.pl/linki/wirtua.php>). Pod hasłem ?biblioteki wirtualne? znajduje się ich wykaz z podziałem najpierw na projekty polskie i zagraniczne, a potem na inicjatywy narodowe i lokalne. Natomiast serwis EBIB ?Baza bibliotek polskich?, w którym można wyszukać biblioteki digitalizujące zbiory (<http://ebib.oss.wroc.pl/linki/searchnew.php>), pozwala na wyszukanie 30 takich bibliotek.

Zasada 3: *Zasób powinien być rozwijany w czasie. Należy tworzyć plany dalszego rozwoju po zakończeniu finansowania, szczególnie gdy wykorzystywane są fundusze specjalne (granty).*

Rozwój w czasie na poziomie zasobu łączy się z koniecznością zapewnienia niezmienności na poziomie obiektu. Strategie archiwizacji na obu poziomach powinny być ze sobą związane. Zarządzający zasobami zawierającymi materiały o długotrwałym znaczeniu powinni podejmować działania, dzięki którym obiekty cyfrowe będą zachowane w użytecznej formie przez długi czas oraz utrzymywany będzie dostęp do materiałów na poziomie zasobu.

Wynika z tego potrzeba stałego zaangażowania organizacyjnego w bieżące zarządzanie zasobami. Obejmuje ono takie funkcje, jak: posiadanie informacji o aktualnej lokalizacji, zapewnienie użyteczności aplikacji, wprowadzanie danych i ich modyfikacja, obsługa logowań i gromadzenie przyrastających statystyk oraz obsługa użytkownika końcowego. Powinny się tam znaleźć także funkcje administrowania systemem w przypadku okresowych upgrade'ów serwera i oprogramowania operacyjnego, zarządzanie bezpieczeństwem serwera i tworzenie odpowiednich kopii bezpieczeństwa.

Zasada 4: *Dobry zasób jest szeroko dostępny, m.in. dzięki usuwaniu zbędnych przeszkód w jego użytkowaniu. Zasoby powinny być dostępne dla osób niepełnosprawnych dzięki efektywnemu wykorzystywaniu technologii adaptacyjnych.*

Obecnie Web jest najpopularniejszym narzędziem udostępniania zasobów. Powinien on więc być stosowany do udostępniania zasobów przy użyciu technologii znanych w środowisku użytkowników.

Często występuje sprzeczność pomiędzy funkcjonalnością a ogólną użytecznością; wdrażanie nowych funkcji powinno być powiązane z uwzględnieniem zdolności potencjalnych użytkowników do wykorzystania tych technologii i upewnieniem się, czy nie zostaną one uznane za barierę. Należy wziąć pod uwagę takie zagadnienia jak przepustowość łącz, gdyż może ona stanowić barierę dla stosowania niektórych formatów plików lub interfejsów. Częścią opisu zasobu powinna być co najmniej informacja o wersji przeglądarki i minimalnej przepustowości sieci.

Zaleca się testowanie stron Web, metadanych i sposobu wyświetlania obiektów za pomocą różnych przeglądarek (np. Netscape, IE, Opera), ich wersji i rozdzielczości ekranu. Podobnie należy testować przynajmniej Windows, Mac i Linux w bieżącej i poprzednich wersjach (przynajmniej trzy lata wstecz). Do testowania wybierać należy szczególnie trudne obiekty, o zróżnicowanej kolorystyce i niełacińskim zestawie znaków.

Zasada 5: *W dobrych zasobach przestrzega się zasad prawa autorskiego. Zarządzający biblioteką cyfrową powinni tworzyć zapisy o właścicielach praw autorskich i uzyskanych pozwoleniach na korzystanie z zasobów.*

Prawa własności intelektualnej powinny być rozpatrywane z kilku punktów widzenia: praw właściciela oryginalnego źródła; praw lub pozwoleń na digitalizację i udostępnianie uzyskanych przez twórców zasobów; uprawnień do zasobów cyfrowych posiadanych przez właścicieli biblioteki cyfrowej; praw lub pozwoleń niezbędnych użytkownikom dla korzystania z materiałów. Bez względu na punkt widzenia, prawa autorskie rzadko stanowią zagadnienia jednoznacznie uregulowane, a polityka zarządzania nimi związana jest częściej z określeniem akceptowanego poziomu ryzyka niż z jednoznacznym określeniem, co jest dobre, a co złe. O zagadnieniach tych pisała m.in. Sybilla Stanisławska-Kloc^[3].

Zasada 6: *Dobre zasoby zawierają mechanizmy dostarczania danych o poziomie ich wykorzystania i pozwalających na stosowanie standardowych miar użyteczności.*

Efektywne zarządzanie biblioteką cyfrową opiera się na standardowych miarach użyteczności zasobów, pozwalających pozyskiwać z określoną częstotliwością jednolite dane o wykorzystaniu zasobów, dzięki czemu oszacować można dostosowanie zasobów do potrzeb użytkowników, a także na miarach pozwalających na bieżącą ocenę zasobów z punktu widzenia realizacji misji organizacji. Stosowanie miar jest także ważne ze względu na możliwość wykazania powodzenia projektu, co ułatwia zapewnienie stałego wsparcia instytucjonalnego.

Badanie użyteczności zasobów może być realizowane różnymi metodami, takimi jak: obserwacja, ekspertyza, wywiad, eksperyment, analiza logów transakcji. Każda metoda ma cechy dodatnie i ujemne. Na przykład logi transakcji mogą dostarczyć informacji o częstotliwości korzystania, o tym, jakie materiały były wykorzystywane i kto ich używał (dokument był używany przez ?x? użytkowników z ?y? różnych domen). Daje to jednak ograniczone pojęcie o rzeczywistej użyteczności zasobów dla użytkowników. Niezbędne jest więc jednocześnie stosowanie wielu różnych miar.

Ponieważ miary powinny być stosowane przez dłuższy czas oraz wymagają pewnych nakładów, ich wybór powinien być związany z celami organizacji lub sponsora projektu. Zazwyczaj chodzi o wyjaśnienie powodów włączenia materiałów do zasobów za pomocą wskaźnika poziomu ich wykorzystania przez wszystkich użytkowników lub ich określoną grupę. Innymi celami mogą być: doskonalenie polityki rozwoju zasobów lub poprawa funkcjonalności systemu wyszukiwawczego i udostępniania.

Zasada 7: *Prace w zakresie bibliotek centralnych na poziomie lokalnym powinny być zgodne z inicjatywami ogólnokrajowymi i międzynarodowymi.*

Podczas określania intelektualnej zawartości zasobów ich twórcy powinni brać pod uwagę ogólniejsze priorytety narodowe i międzynarodowe, takie jak najnowsze kierunki badań naukowych. Dzięki temu wzrasta wykorzystanie zasobów cyfrowych oraz łatwiejsze staje się pozyskiwanie odpowiednich środków.

Twórcy zasobów powinni także zwrócić uwagę na problem współdziałania, szczególnie w zakresie wymiany metadanych. Należy śledzić i uwzględniać wyniki prac o podobnym zakresie i uwzględniać wyznaczniki jakości treści i metadanych. Wśród narzędzi ułatwiających współdziałanie można wyróżnić:

- Dublin Core Metadata Initiative. Oprócz zestawu elementów metadanych, DCMI prowadzi także rejestr elementów metadanych służący współdziałaniu. Zob. <http://dublincore.org/>.
- METS (Metadata Encoding and Transmission Standard), standard utworzony przez bibliotekarzy dla wymiany obiektów cyfrowych. Zob. <http://www.loc.gov/standards/mets/>.
- MXF (Materials Exchange Format), standard ułatwiający wymianę materiałów audiowizualnych między aplikacjami i organizacjami. Zob. <http://www.broadcastpapers.com/siqdis/Snell&WilcoxMXF01.htm>.
- OAI (Open Archives Initiative), w której utworzono protokół służący włączaniu lokalnych metadanych do większych repozytoriów metadanych. Zob. <http://www.openarchives.org/>. W Polsce wykorzystanie OAI planuje sieć PIONIER^[4].

Na Zachodzie od dawna powstają biblioteki cyfrowe będące rodzajem portali dziedzinowych. Zawierają one informacje z określonych gałęzi wiedzy i są bramkami do istniejących stron Web i innych zasobów, zarządzanych wspólnie przez grupę instytucji, dbających o jakość informacji z wybranej poddziedziny. Tworząc własną bibliotekę, należy uwzględnić ich istnienie. W Polsce tego typu miejsca w sieci są dopiero tworzone.

Obiekty

Przewodnik NISO uwzględnia dwa rodzaje obiektów cyfrowych: obiekty digitalizowane tworzone jako surogaty materiałów tradycyjnych (książek drukowanych, rękopisów, obiektów muzealnych, taśm wideo) oraz obiekty oryginalnie powstałe jako cyfrowe, w formie czytelnej maszynowo (e-książki, naukowe bazy danych, fotografie cyfrowe, strony Web itp.).

Obiekt może zawierać się w jednym pliku (np. raport wydany jako plik pdf) lub składać się z wielu plików połączonych ze sobą (np. strona HTML i powiązane z nią obrazki) lub też może on obejmować wiele plików oraz metadanych strukturalnych, niezbędnych do powiązania ich w całość (np. książka zdigitalizowana jako obrazy stron). W takim sensie obiekty są konceptualnymi odpowiednikami jednostek w zbiorach bibliotecznych, kolekcjach muzealnych i zasobach archiwalnych.

Zasoby zawierają obiekty, nie można jednak jednoznacznie wyznaczyć granicy pomiędzy zasobem a obiektem. Przedstawiona definicja obiektu obejmuje obiekty złożone, takie jak cyfrowo przeformatowane książki, ale nie zasoby (które w tym przypadku powinny się składać z wielu cyfrowo przeformatowanych książek). Obiekt cyfrowy może należeć do więcej niż jednej biblioteki cyfrowej.

Można przedstawić następujące problemy z określeniem, czym jest dobry obiekt:

- Dzięki ograniczeniu do minimum wymogów dla dobrych obiektów organizacje tworzące te obiekty mogą ograniczyć ryzyko związane z ich produkcją i zarządzaniem.
- Ponieważ dobre obiekty powinny zaspokajać znane bieżące i przewidywane przyszłe potrzeby, organizacje mogą bezpiecznie inwestować w ich tworzenie, wiedząc, że nie będą musiały powtórnie tworzyć obiektów w przyszłości.
- Będzie wzrastało zaufanie użytkowników do dobrych obiektów, ponieważ zawierać one będą minimalny poziom dobrze znanych i jednolitych atrybutów i będą wspomagać zaspokajanie wielu różnych potrzeb.
- Dzięki określeniu konsensusu co do charakterystyk dobrych obiektów organizacje będą mogły zawierać lepsze kontrakty z dostawcami tworzącymi obiekty, porównywać ceny różnych dostawców oraz definiować sposoby archiwizacji polegającej na migracji lub emulacji obiektów w przyszłości.

Dla obiektów wyróżniono następujące zasady:

Zasada 1: *Dobry obiekt cyfrowy produkowany jest w sposób zapewniający realizację priorytetów zasobów, a poprawa jego jakości zwiększa współdziałanie i możliwości wielokrotnego wykorzystania.*

Decyzje dotyczące sposobów produkcji i opisu obiektów powinny wynikać z powodów, dla których (np. dla kogo, do jakich celów) został on wyprodukowany. Dlatego też przedstawione wcześniej zasady dotyczące dobrych zasobów dotyczą także tworzenia dobrych obiektów.

Zasada 2: *Dobry obiekt jest niezmienny. Oznacza to, że znana jest osoba lub instytucja, która dba o to, aby obiekt pozostawał dostępny przez długi czas bez względu na dokonywane zmiany technologiczne.*

Informacja cyfrowa z natury jest krótkotrwała. ?Okres życia? mediów cyfrowych, platform sprzętu i oprogramowania oraz formatów plików cyfrowych jest krótki. Można oczekiwać, że okres stosowania określonych mediów cyfrowych może trwać około 10 lat, ale tylko niewiele platform sprzętowych czy oprogramowania przetrwa ten czas.

Testowanych jest wiele strategii archiwizacji obiektów cyfrowych. Najczęściej wymieniane są migracja i emulacja. Migracja polega na transformacji obiektów w taki sposób, aby można je było przenosić pomiędzy różnymi technologiami i występuje pomiędzy wszystkimi poziomami:

- pomiędzy nośnikami (np. z dyskietek do CD, a z CD na dysk optyczny),
- pomiędzy oprogramowaniem (np. wersjami edytorów lub z jednej aplikacji do innej),
- pomiędzy formatami (np. z SGML do XML lub z jpeg do jpeg2000).

Czasami lepiej jest, ze względu na koszty i utratę danych, emulować na nowym systemie środowisko komputerowe, w którym obiekty cyfrowe były pierwotnie tworzone i używane (np. środowisko DOS w Windows). Strategie emulacji mogą być szczególnie przydatne dla złożonych obiektów multimedialnych, takich jak interaktywne pomoce dydaktyczne.

Zasada 3: *Dobry obiekt jest digitalizowany w formacie ułatwiającym jego aktualne i przyszłe wykorzystywanie lub kopiowanie. W efekcie dobry obiekt można wymieniać pomiędzy platformami, jest on ogólnie dostępny i zdigitalizowany zgodnie ze znanymi standardami.*

Prawie zawsze istnieje bezpośredni związek pomiędzy jakością digitalizacji obiektu a łatwością i elastycznością korzystania z niego, wielokrotnego użycia i migracji pomiędzy platformami. Wysokiej jakości obiekt zdigitalizowany pozostanie dłużej użyteczny i dostępny. Nie wszystkie obiekty zachowują swoją wartość przez długi czas. W projekcie należy określić wartość obiektów cyfrowych w zasobach i podjąć odpowiednie decyzje dotyczące ich niezmienności i współdziałania.

Zasada 4: *Nazwa dobrego obiektu służy jako niezmienny, unikalny identyfikator. Nie powinna to być nazwa związana z nazwą pliku lub adresem (takim jak URL), gdyż nazwy i adresy często się zmieniają. Bieżącemu adresowi powinien być przydzielony niezmienny identyfikator.*

W idealnej sytuacji lokalnie przydzielany identyfikator powinien być zgodny z odpowiednimi narodowymi i międzynarodowymi standardami. Niestety, większość identyfikatorów albo odnosi się do określonych klas obiektów (jak ISBN), albo mogą one być wyznaczane tylko przez wytypowane agencje lub za rejestrację obiektu trzeba płacić. Dla większości zasobów cyfrowych identyfikatory obiektów muszą być przydzielane lokalnie na podstawie lokalnie utworzonego schematu. Nie stanowi to problemu, o ile schemat jest udokumentowany i dokumentacja ta jest dostępna. Można także stworzyć schemat lokalny na podstawie standardowych identyfikatorów.

Identyfikatory funkcjonujące dla obiektów dostępnych poprzez Internet powinny stosować rejestry (?resolver?), które mapują statyczny, niezmienny identyfikator na bieżącą lokalizację obiektu. Pomimo tego, że tablice mapowania muszą być aktualizowane za każdym razem, gdy obiekt zmienia swoje miejsce, tego typu działania ułatwiają zarządzanie obiektami, gdyż należy zaktualizować lokalizację tylko raz w centralnym rejestrze, bez względu na częstotliwość występowania identyfikatora w odsyłaczach. Istnieją dwie aplikacje rejestrowe: Persistent URL (PURL) i Handle. Bez względu na zastosowane technologie ważne jest, aby identyfikatory pozostawały niezmiennie. Istotne jest także, aby miały podbudowę instytucjonalną.

Zasada 5: *Autentyczność dobrego obiektu może być określona na przynajmniej trzy sposoby. Po pierwsze użytkownik powinien mieć możliwość określenia pochodzenia obiektu, jego struktury i kolejnych wersji. Po drugie użytkownik powinien mieć możliwość zbadania, czy obiekt jest tym, czym powinien być. Po trzecie użytkownik powinien móc stwierdzić, że obiekt nie został uszkodzony lub zmieniony w sposób niedozwolony.*

Możliwość określenia autentyczności obiektu jest ważna z kilku powodów. Badania naukowe są prowadzone na podstawie weryfikowalnych faktów. Nauczanie i uczenie się oraz inne formy działalności kulturalnej także oparte są na możliwości zapewnienia użytkownika o wiarygodności, poprawności i autentyczności obiektu.

Informacja dokumentująca pochodzenie obiektu i historię jego modyfikacji jest nazywana ?cyfrową proveniencją? i stanowi metadane zapisywane wraz z obiektem.

Metadane

Jednym z najtrudniejszych aspektów środowiska cyfrowego jest identyfikacja zasobów dostępnych w Web. Istnienie wyszukiwawczych metadanych opisowych zwiększa szanse, że zdigitalizowane treści zostaną odnalezione i wykorzystane. Tutaj zajmiemy się metadanymi dla opisu poszczególnych obiektów oraz zestawów obiektów w zasobach.

Metadane są ustrukturyzowaną informacją dotyczącą obiektu, mającą na celu umożliwienie jego wyszukania, opisu, wykorzystania, administrowania i/lub zarządzania. Metadane mogą być dołączone na dowolnym etapie funkcjonowania obiektu cyfrowego.

Najczęściej wyróżnia się metadane trojakiego rodzaju:

- metadane opisowe pozwalające użytkownikowi odnaleźć obiekt, odróżnić od siebie różne obiekty i zrozumieć przedmiot lub treść obiektu;
- metadane administracyjne pomagają zarządzającym zasobami kontrolować obiekty dla takich celów, jak zarządzanie plikami, obsługa praw autorskich i archiwizacja;
- metadane strukturalne opisują relacje pomiędzy obiektami, np. związki pomiędzy artykułami, zeszytami i tomami czasopism lub stronami i rozdziałami książek.

Główną przyczyną tworzenia bibliotek cyfrowych jest zwiększenie dostępności zasobów posiadanych przez organizację. Wykorzystanie szeroko dostępnych metadanych opisowych jest sposobem na maksymalizację udostępniania dla aktualnych użytkowników i przyciągnięcie nowych. W ciągu ostatnich lat utworzono wiele schematów metadanych służących opisowi różnych obiektów cyfrowych. Te różnorodne schematy są jednak na tyle spójne, że umożliwiają współdziałanie.

W Polsce stosuje się praktycznie dwa schematy metadanych: Dublin Core i MARC21. Na problemy z właściwym ich stosowaniem w bibliotekach cyfrowych wskazywał m.in. Henryk Hollender^[5]. Poniższe zasady pomogą we właściwym wyborze i implementacji standardów metadanych.

Zasada 1: *Dobre metadane powinny być odpowiednio dobrane do materiałów w zasobach, do użytkowników tych zasobów oraz planowanego, aktualnego i możliwego przyszłego wykorzystania obiektu cyfrowego.*

Istnieje wiele opublikowanych schematów metadanych przydatnych dla obiektów cyfrowych, stron Web i e-zasobów. Zazwyczaj można znaleźć więcej niż jeden schemat możliwy do zastosowania dla danego zasobu. Wybór powinien zależeć od poziomu zasobów przeznaczonych do opracowania przy pomocy metadanych, doświadczenia twórców metadanych, sposobu wykorzystania zasobów, przygotowania ich użytkowników i podobnych czynników. Należy także sprawdzić, jakie schematy są najczęściej używane w podobnych organizacjach; wykorzystanie tych samych schematów zwiększy poziom współdziałania.

Zasada 2: *Dobre metadane ułatwiają współdziałanie.*

Edukacja i nauka funkcjonują obecnie w rozproszonym środowisku sieciowym. Wyszukanie zasobów rozproszonych pomiędzy bibliotekami, archiwami i instytucjami badawczymi może stwarzać poważne problemy. Można ich uniknąć dzięki projektowaniu przez instytucje tego typu metadanych umożliwiających współdziałanie.

Celem współdziałania jest pomoc użytkownikowi w odnalezieniu i uzyskaniu dostępu do obiektów informacyjnych rozproszonych w domenach i instytucjach. Stosowanie standardowych schematów metadanych ułatwia współdziałanie przez umożliwienie wymiany rekordów metadanych pomiędzy systemami stosującymi podobne schematy. Innym sposobem osiągnięcia współdziałania jest konwersja elementów z jednego schematu do innego.

Kolejnym sposobem zwiększenia stopnia współdziałania jest zastosowanie protokołu indeksowania Open Archives Initiative (OAI). System zgodny z protokołem OAI może udostępnić swoje metadane narzędziom indeksującym, co pozwala na włączenie tych metadanych do wielkich baz danych i ich wykorzystywanie przez zewnętrzne serwisy wyszukiwawcze.

Jeszcze innym sposobem zwiększenia współdziałania jest zastosowanie protokołów dla wyszukiwania pomiędzy systemami; nosi to nazwę metawyszukiwania. W takim przypadku metadane pozostają w zasobach źródłowych, ale lokalny system wyszukiwawczy akceptuje zapytania z zewnętrznych systemów wyszukiwawczych. Najlepiej znanym protokołem tego typu jest standard międzynarodowy Z39.50.

Zasada 3: *Dobre metadane stosują kontrolę autorytarną i narzędzia kontroli treści, takie jak słowniki kontrolowane, spełniające wymagania użytkowników co do opisu treści obiektów i łączenia ze sobą obiektów podobnych.*

Tam, gdzie jest to możliwe, atrybuty obiektów powinny być opisywane za pomocą słowników kontrolowanych. Dotyczy to nazw osobowych, korporacyjnych, nazw geograficznych oraz haseł przedmiotowych. Istnieje wiele tezaursów, taksonomii i kartotek autorytarnych, brak natomiast jednego rozwiązania zaspokajającego wszystkie potrzeby. Wybór narzędzia zależy m.in. od wybranego schematu metadanych i rodzaju zasobów. Za wyjątkiem bardzo prostych i krótkich list autorytarnych nie należy planować ich stosowania przez osoby nieprzeszkolone (np. autorów).

Zasada 4: *Dobre metadane zawierają jasno określone warunki i zasady użytkowania obiektu cyfrowego.*

Zasady i warunki użytkowania obejmują prawa własności, pozwolenia i ograniczenia dotyczące pozwoleń. Użytkownik powinien być poinformowany o sposobie uzyskania pozwoleń na ograniczone stosowanie oraz sposobu cytowania materiałów. Dane tego typu mogą dotyczyć także całego zasobu i wówczas są odnotowane na tym poziomie.

Zasada 5: *Dobre metadane ułatwiają długoterminowe zarządzanie obiektami w zasobach.*

Metadane administracyjne są informacją ułatwiającą zarządzanie zasobami. Dotyczy to danych typu: jak i kiedy obiekt został utworzony, kto jest odpowiedzialny za kontrolę dostępu lub archiwizację, jakie działania kontrolne lub przetwarzania były podejmowane i jakie są ograniczenia użytkowania. Metadane techniczne, takie jak format fizyczny, rozmiar pliku, suma kontrolna itp., mogą być również niezbędne dla zapewnienia stałej dostępności obiektu lub jego rekonstrukcji po uszkodzeniu.

Metadane archiwizacji są częścią metadanych administracyjnych, służąc długoterminowemu zachowaniu obiektów cyfrowych. Mogą one zawierać szczegółowe metadane techniczne, a także informacje dotyczące zarządzania prawami, historię funkcjonowania i modyfikacji obiektu.

Metadane strukturalne pozwalają na powiązanie ze sobą różnych odrębnych części informacji. Jednym z najczęściej stosowanych schematów metadanych strukturalnych jest wspomniany już METS.

Zasada 6: *Rekordy dobrych metadanych także są obiektami, powinny więc być tworzone zgodnie z zasadami tworzenia dobrych obiektów, takimi jak autorytatywność, autentyczność, archiwizacja, niezmienność i jednoznaczna identyfikowalność.*

Niezbędne jest stworzenie kontroli autorytatywnej metadanych, gdyż zawierają one informacje

świadczące o pochodzeniu, integralności i autorytatywności obiektu. Takie meta-metadane, a raczej zapisana informacja o metadanych, powinna zawierać identyfikację tworzącej ją instytucji oraz określenie standardów kompletności i jakości przyjętych podczas tworzenia. Instytucja powinna dostarczyć informację wystarczającą do weryfikacji metadanych przez użytkownika, w tym: sposób tworzenia (manualny czy automatyczny) oraz wykorzystane standardy i słowniki.

Problem nieautentycznych i niepoprawnych metadanych jest rzeczywisty i poważny. Wiele wyszukiwarek internetowych nie wykorzystuje metadanych włączonych do stron HTML z powodu uciążliwych problemów ze spoofingiem (jedna organizacja dostarcza błędne metadane o zasobach należących do innej organizacji) i spammingiem (sztucznie powtarzane słowa kluczowe dla wymuszenia wyższej pozycji w rankingu). W stosunku do metadanych można stosować te same techniki, które służą weryfikacji integralności i autentyczności dokumentów cyfrowych (np. cyfrowe podpisy).

Projekty

Projekty służące tworzeniu bibliotek cyfrowych mogą być realizowane przez użytkowników, takich jak naukowcy czy studenci, lub przez pracowników instytucji kultury (kustoszy muzealnych, bibliotekarzy, archiwistów). Wszystkie projekty mają swoje własne cele i określony czas trwania. Planowanie projektu dotyczy także przyszłości zasobów cyfrowych po zakończeniu realizacji projektu. Należy przygotować plany zapewnienia dalszej dostępności do zasobów i zarządzania zasobami, wraz z jasnym określeniem odpowiedzialności za ich realizację.

Projekty związane z bibliotekami cyfrowymi często realizowane są przez osoby reprezentujące różne dyscypliny i instytucje. Podczas kompletowania zespołu kierownik projektu powinien brać pod uwagę kwalifikacje, kompetencje, niezawodność i umiejętności współpracy członków zespołu.

W razie potrzeby kierownik projektu powinien mieć możliwość zażądania przygotowania dodatkowych ekspertyz stworzonych przez zewnętrznych konsultantów.

Zasada 1: *Ważnym elementem dobrego projektu tworzenia biblioteki cyfrowej jest odpowiednie projektowanie i planowanie.*

Planowanie projektu ma zasadnicze znaczenie dla zakończenia projektu sukcesem. Dotyczy ono wszystkich aspektów projektu, od rozpracowywania planów działania do opracowania ostatecznego wyglądu strony Web biblioteki cyfrowej. Planujący powinni jak najwcześniej określić celowego odbiorcę treści biblioteki cyfrowej i przeprowadzić przegląd wymagań służących realizacji potrzeb funkcjonalnych tych użytkowników. Następnie na piśmie przygotować można plan projektu, który obejmuje wszystkie jego istotne aspekty: krótko- i długoterminowe cele, ograniczenia projektu (czasowe, dotyczące zasobów, czynniki polityczne), zasady selekcji i digitalizacji, zagadnienia copyrightu, metadane i dostęp, zarządzanie, rozpowszechnianie i ocena.

Zasada 2: *Dobry projekt posiada plan jego oceny.*

Plan oceny obrazuje stopień zaangażowania projektu w realizację jego celów. Ocena może skupiać się na procesach i wynikach. Ocena procesu może zawierać oszacowanie poprawności wykonywanych operacji, takich jak zatrudnienie i zarządzanie, przebiegi czynności i procedury. Ostatnio częściej stosowana jest ocena na podstawie wyników. Wyszczególnienie celów projektu powinno pomóc zespołowi realizującemu projekt określić pożądane wyniki.

Często oceniane są także produkty, usługi i wpływ projektu. Podczas oceny produktu projektu tworzenia biblioteki cyfrowej można mierzyć rozmiary zasobów cyfrowych, ich jakość i użyteczność. Dodatkowo dobrymi wskaźnikami projektu są użyteczność strony Web biblioteki, doświadczenia użytkowników dotyczące zasobów i usług oraz wpływ zasobów na użytkowników. Dane mogą być zbierane za pomocą ekspertyz, badania grup celowych, wywiadów, badania logów transakcji i studiów przypadków.

Zasada 3: *W dobrym projekcie tworzony jest raport z jego realizacji, w którym rozpowszechniana jest informacja o procesach projektu i jego wynikach.*

Realizacja dobrego projektu przewiduje tworzenie dokumentacji, która wyjaśnia czytelnikom jego procesy i wyniki. Jeżeli efektem realizacji projektu są jakieś modele, narzędzia lub prototypy, powinny one stać się łatwo dostępne, aby mogły być zastosowane gdzie indziej. Jeżeli projekt ma znaczenie lokalne, regionalne lub narodowe, powinno ono być opisane w publikacjach, prezentacjach, mediach i innych źródłach.

Zasada 4: *Dobry projekt dotyczy całego cyklu życia zasobu cyfrowego i związanych z nim usług dostępnych w wyniku realizacji projektu.*

Zasoby, takie jak: personel, sprzęt, oprogramowanie i poziom wysiłków niezbędnych do planowania i tworzenia biblioteki cyfrowej zazwyczaj są bardzo różne od tych, które służą późniejszemu zarządzaniu i długoterminowemu utrzymaniu. Podczas planowania należy zaprojektować wykorzystywanie zasobów w długim okresie czasu i niezbędne w tym okresie aktualizacje zasobu i strony Web projektu.

Podczas implementacji projektu często niezbędni są wykwalifikowani specjaliści, tacy jak: menedżerowie projektu, twórcy metadanych, technicy digitalizacji. Zasoby nierozbudowywane lub rozwijające się powoli powinny być umieszczone w stałych planach pracy organizacji. Bardzo istotne jest oszacowanie i zaplanowanie tych permanentnych wysiłków.

Dobry projekt biblioteki cyfrowej powinien w efekcie zapewniać zasoby i usługi, które staną się ważnym i wiarygodnym elementem działań informacyjnych organizacji, w związku z czym muszą podlegać tym samym standardom, które organizacja stosuje do innych zasobów i usług.

Zasady tworzenia bibliotek cyfrowych, sformułowane przez NISO, mogą wydawać się znajome bibliotekarzom, również pracującym w bibliotekach tradycyjnych. Wiele tych zasad realizowanych jest na co dzień w tradycyjnych bibliotekach, chociaż oczywiście uwzględniają one specyfikę zasobów i obiektów cyfrowych oraz stosowanych metadanych. To podobieństwo wynika z faktu, że biblioteka cyfrowa to także jest biblioteka. Wydaje się, że kłopoty i słabości wielu polskich projektów bibliotek cyfrowych wynikały z pomijania tego niby oczywistego faktu. Efektem takiego podejścia były próby tworzenia bibliotek cyfrowych bez udziału bibliotekarzy i praktyki bibliotekarskiej, traktując je głównie, jeśli nie wyłącznie, jako przedsięwzięcia informatycznego. Szczęśliwie wydaje się, że praktyka zmusiła do powrotu do zasad, które w sposób uporządkowany i usystematyzowany przedstawia omówiony tu dokument NISO.

Przypisy:

[1] *A Framework of Guidance for Building Good Digital Collections (2004)* [on-line]. Wyd. 2. Bethesda: NISO Framework Advisory Group [dostęp 6 marca 2006]. Dostępny w World Wide Web: <http://www.niso.org/framework/framework2.pdf>.

[2] FRANKE, J. Warianty rozwoju cyfrowych kolekcji. In GRYGROWSKI, D., ZYBERT, E. (red.) *Bibliologia, biblioteki, bibliotekarze*. Warszawa: Wydaw. SBP, 2005, s. 84-87.

[3] STANISŁAWSKA-KLOC, S. Prawo autorskie a biblioteka cyfrowa. In Barbara Szczepańska (red.) *III Konferencja Internet w bibliotekach. Zasoby elektroniczne: podaż i popyt. Wrocław, 12-14 grudnia 2005 roku* [on-line]. [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich, K[omisja] W[ydawnictw] E[lektronicznych], 2006 [dostęp 28 marca 2006]. EBIB Materiały Konferencyjne. Dostępny w World Wide Web: http://ebib.oss.wroc.pl/konferencje/iwb3/pdf/stanislawski_kloc.pdf.

[4] HELIŃSKI, M. [et al.]. Biblioteka cyfrowa jako otwarte, interaktywne repozytorium publikacji. In Barbara Szczepańska (red.) *III Konferencja Internet w bibliotekach. Zasoby elektroniczne: podaż i popyt. Wrocław, 12-14 grudnia 2005 roku* [on-line]. [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich, K[omisja] W[ydawnictw] E[lektronicznych], 2006 [dostęp 28 marca 2006]. EBIB Materiały Konferencyjne. Dostępny w World Wide Web: http://ebib.oss.wroc.pl/konferencje/iwb3/pdf/mazurek_in.pdf.

[5] Hollender, H. Biblioteki cyfrowe w Polsce: warunki i zagrożenia. In *Biuletyn EBIB* [on-line]. 2005 nr 4 [dostęp 6 marca 2006]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2005/65/hollender.php>.

Tekst jest modyfikacją wykładu przedstawionego przez autora podczas seminarium ?Cyfrowy świat bibliotek?, Warszawa 17-18 styczeń 2006 r., z organizowane przez Centrum Promocji Informatyki. Przygotowany został on na podstawie A Framework of Guidance for Building Good Digital Collections [on-line].). Wyd. 2. Bethesda: NISO Framework Advisory Group, 2004 [dostęp 6 marca 2006]. Dostępny w World Wide Web: <http://www.niso.org/framework/framework2.pdf>.

