

Różne aspekty opisu dokumentów ikonograficznych w Bibliotece Cyfrowej Politechniki Warszawskiej na przykładzie kolekcji fotografii Henryka Poddębskiego

Istnieje wiele definicji charakteryzujących dokumenty ikonograficzne. Najpopularniejsza z nich określa tę grupę materiałów bibliotecznych jako dokument graficzny, którego treść jest przedstawiona w postaci obrazu. Biblioteka Narodowa, opierając się na francuskiej normie, definiuje dokument ikonograficzny jako „dwuwymiarowy, nieruchomy, nieprzeznaczony do projekcji obraz utrwalony dowolną techniką na przenośnym podłożu”¹. Zapewne podobnie określa ten dokument nie zatwierdzona jeszcze polska norma. Na skonfrontowanie tej informacji musimy jednak jeszcze trochę poczekać.

Jest wiele rodzajów dokumentów ikonograficznych, zaliczamy do nich m. in.: grafiki, rysunki, dzieła malarskie, fotografie (pozytywy, negatywy), matryce graficzne, plakaty, zaproszenia, kartki pocztowe, medale i wiele innych.

Biblioteka Cyfrowa Biblioteki Głównej Politechniki Warszawskiej jako jeden z pierwszych w/w dokumentów wprowadziła do swoich zbiorów fotografie.

Digitalizacja stworzyła nowe możliwości w archiwizacji i udostępnianiu zbiorów fotograficznych. Powstało w związku z tym kilka inicjatyw, m.in. projekt TAPE w ramach unijnego programu Kultura 2000, dotyczący zabezpieczania kolekcji audiowizualnych jako istotnej części dziedzictwa kulturowego Europy. Skierowany był w szczególności do tych instytucji, w których materiały audiowizualne nie stanowią głównej części zbiorów, a jednym z jego celów było promowanie świadomości potrzeby zabezpieczenia i ochrony kolekcji audiowizualnych. Również program operacyjny „Dziedzictwo Kulturowe”, realizowany przez Ministerstwo Kultury i Dziedzictwa Narodowego uwzględnia znaczącą rolę digitalizacji fotografii jako elementu cyfryzacji zasobów bibliotecznych i archiwalnych. Stosowanie, wybór standardu do opracowania zbiorów fotograficznych stanowiło problem od dawna, zarówno dla bibliotek, jak i archiwów czy muzeów. W latach 2000-2003 w ramach projektu SEPIA, realizowanego w ramach programu „Kultura 2000”, 9 instytucji (Biblioteki, archiwa, muzea i European Commission on Preservation and Access) wypracowało rekomendację SEPIADES (Sepia Data Element Set) - dla opracowania zbiorów fotograficznych. Standard został opracowany, dołączono darmowe oprogramowanie, a jednak SEPIADES nie stał się powszechnie stosowanym rozwiązaniem. Środowisko bibliotekarskie wykorzystuje przede wszystkim standard MARC, a archiwa standard ISAD, który właściwie stosuje się do opisu zbioru archiwalnego (dopasowanego teczki zawierającej akta). Do opisu samej fotografii archiwa używają standard EAD, wspierany przez Bibliotekę Kongresu USA. W związku z rozwojem fotografii cyfrowej na pocz. XXI w. i masowym ich wykorzystaniem przez agencje prasowe potrzebny był ustandaryzowany opis, pozwalający na szybkie opisanie zdjęć. MARC był zbyt skomplikowany i związany z fizycznym nośnikiem, ISAD wymagał wskazania do jakiego zbioru należy fotografia. Możliwość zapisania metadanych bezpośrednio w plikach

¹ Format MARC 21 rekordu bibliograficznego : dokumenty ikonograficzne : wersja 2008. Oprac. Agata Pietrzak. Warszawa 2008 s.1.

została w 2001 r. dodana w oprogramowaniach Adobe Acrobat w standardzie XMP. Opis zawartości pliku zgodny z XMP można zakodować w pliku PDF, a także w formatach graficznych typu JPEG, GIF czy TIFF. Opis może być dokonany w dowolnym standardzie, również Dublin Core czy MARC. Powiązanie opisu z samym plikiem graficznym pozwala uzyskać informacje o zawartości pliku użytkownikowi, który skopiuje go na własny dysk, ale ograniczenie opisu tylko do opisu „wewnętrznego” przy utracie takiego pliku powoduje, że równocześnie traci się informację o tym co zostało utracone. Tego rodzaju rozwiązanie jest nie do przyjęcia dla bibliotek czy archiwów. Czy powstanie ujednoczony standard? Wygląda na to, że niekoniecznie. „Zespół do spraw digitalizacji” powołany przez Ministra Kultury w 2006 r. wstępnie wskazywał na „branżowe” rozwiązanie: MARC dla bibliotek i EAD (zakodowany ISAD) dla archiwów oraz Dublin Core, jako standard dopuszczalny do uporządkowania informacji dla opracowania zbiorów powstałych w wyniku digitalizacji.

W opracowaniu dokumentów ikonograficznych można zastosować dwie metody, grupową/zespołową oraz opracowanie jednostkowe/indywidualne. Pierwszą metodę stosuje się dla utworzonej w sposób sztuczny grupy dokumentów zebranych w całość ze względu na jej unikatowość, wspólną proveniencję, zasady przechowywania, opracowania bądź np. wspólny temat. W przypadku Biblioteki Cyfrowej BGPW, ze względu na dużą wartość historyczną zdjęć Henryka Poddębskiego, przyjęliśmy zasadę opracowania jednostkowego.

Katalogowanie poprzedza zmuszony proces gromadzenia informacji niezbędnych do stworzenia opisu bibliograficznego. Źródła z których należy czerpać dane określa norma i przepis.

Wszystkie materiały prezentowane do tej pory w Bibliotece Cyfrowej Politechniki Warszawskiej opracowywane są wg ściśle określonych procedur wypracowanych we współpracy z Narodowym Uniwersalnym Katalogiem Centralnym NUKAT. Podobny tryb chcemy zachować w przypadku dokumentów ikonograficznych, chociaż trudności stwarza tu brak opublikowanej normy oraz przepisów bibliograficznych opracowanych przez Centrum NUKAT.

Procedura wygląda następująco: praca rozpoczyna się na oprogramowaniu Virtua. W pierwszej kolejności zakładane są rekordy wzorcowe dla wszystkich haseł, które będą wykorzystane w opisie katalogowym. Następnym krokiem jest stworzenie dwóch odrębnych opisów bibliograficznych z wykorzystaniem formatu MARC 21 dla różnych rodzajów dokumentu, dla dokumentu ikonograficznego i dla dokumentu elektronicznego o dostępie zdalnym. Po zatwierdzeniu rekordów przez Administratorów Centrum NUKAT bibliotekarz przystępuje do stworzenia opisu bibliograficznego w formacie Dublin Core na oprogramowaniu dLibra. Opis można wykonać metodą konwersji lub ręcznie przenosić dane. Każda z tych metod ma swoje wady i zalety ale tak na prawdę nie ma to zupełnie znaczenia na jaki sposób się zdecydujemy. Konwersja wymaga poprawienia każdego atrybutu w formacie Dublin Core. Ręczne przenoszenie danych przez kopiowanie, może wydawać się pracochłonne, za to nie tracimy czasu na modyfikację.

Ten sposób tworzenia opisów bibliograficznych w bibliotece cyfrowej przyjęto z myślą o wszystkich użytkownikach bibliotek oraz Internetu. Zapewnia on spójność zarówno haseł jak

i opisów bibliograficznych z Narodowym Uniwersalnym Katalogiem Centralnym co ma istotne znaczenie w procesie wyszukiwania informacji przez użytkownika.

Rekordy w katalogu NUKAT i w Bibliotece Cyfrowej BGPW, mimo wspomnianej spójności bardzo się od siebie różnią. Powodem tego są odmienne formaty w jakich zostały zapisane. Format MARC² określa sposób opisu danych bibliograficznych. Natomiast Dublin Core jest standardem metadanych wykorzystywanych do opisu zasobów cyfrowych. Metadane to wg M. Nahotki „ustrukturyzowane, czytelne maszynowo dane zawierające charakterystykę cyfrowych obiektów informacyjnych służącą ich efektywnemu oraz trafnemu wyszukiwaniu, szczególnie w wielkich zasobach informacyjnych w Internecie, zarządzaniu nimi i ich wartościowaniu”³. Wikipedia definiuje je jako "dane o danych"⁴. Często nazywane są też informacją o informacji.

Dublin Core⁵ definiuje 15 atrybutów. Do opisu dokumentów ikonograficznych wykorzystujemy ok. 10 elementów formatu DC w zależności od ilości danych pozyskanych z i spoza katalogowanego dokumentu.

Oto kilka elementów formatu DC i sposób ich wypełniania w BC BGPW:

Tytuł (Title) - Dane do tego pola powinno się podawać w formie występującej na dokumencie. Może to być tekst umieszczony przez autora na awersie, na rewersie, na oryginalnej oprawie, opakowaniu dokumentu. Dopuszcza się tekst napisany odręcznie przez autora oraz uzyskany m. in. np. z negatywu. Jeżeli nie mamy takich danych szukamy napisów dokonanych przez kolejnych właścicieli. Często zawierają one opisy osób i obiektów lub miejsce przedstawionych na fotografii. Tytułów szukamy też na dokumentach towarzyszących lub innych egzemplarzach, jeżeli takie istnieją. W dalszej kolejności zaglądamy do literatury specjalistycznej m. in. katalogów, bibliografii, słowników biograficznych, encyklopedii, Internetu. W ostateczności określamy tytuł na podstawie obrazu utrwalonego na fotografii. Pomijamy informacje o pochodzeniu tytułu, nie dodajemy określenia typu dokumentu oraz nie stosujemy nawiasów kwadratowych. Inaczej wygląda sytuacja w katalogu NUKAT. Za rzecz bardzo istotną uważa się wyraźne odróżnienie danych przejętych z dokumentu od danych przejętych z innych źródeł oraz pochodzących od katalogującego. W rekordzie bibliograficznym wprowadzanym do katalogu NUKAT informacje do poszczególnych stref opisu przejęte spoza źródeł uznanych za podstawowe, ujmują się w nawiasy kwadratowe i uzupełnia odpowiednią uwagą w polu 500. W bibliotece cyfrowej atrybut Źródło (Source) odsyła użytkownika do opisu w Formacie MARC21 w katalogu NUKAT. Nie ma więc potrzeby powtarzania wszystkich informacji w formacie DC. W przypadku fotografii Henryka Poddębskiego, na których nie ma żadnego tekstu, każda informacja zawarta w polach opisu

² Marc Standards. W: [on-line]. Dostępny w World Wide Web: <http://www.loc.gov/marc/>

³ Nahotko Marek: Metadane. Kraków 2004

⁴ Wikipedia, wolna encyklopedia: W: [on-line]. Dostępny w World Wide Web <http://pl.wikipedia.org/wiki/Metadane>

⁵ Dublin Core. W: [on-line]. Dostępny w World Wide Web: <http://dublincore.org/documents/1999/07/02/dces/>

bibliograficznego będzie ujęta w nawiasy kwadratowe i uzupełniona o uwagi w polach typu 5XX. Tytuł utworzony przez katalogującego powinien w pełni charakteryzować zawartość lub temat fotografii. Do każdego tytułu dodajemy nazwę geograficzną.

Twórca (Creator) - W tym polu podajemy nazwę autora fotografii. Hasło autorskie przejmujemy w formie ustalonej w kartotece haseł wzorcowych NUKAT.

Temat i słowa Kluczowe (Subject and Keywords) – Ten element formatu DC jest przeznaczony na opis rzeczowy w formie słów kluczowych. Na opracowanie rzeczowe w BGPW składa się: klasyfikacja wg symboli UKD, słowa kluczowe oraz kod ustawienia dokumentu na półce. W bibliotece cyfrowej do stematowania dokumentu używamy wyłącznie swobodnych słów kluczowych. Przejmujemy hasła z kartoteki haseł wzorcowych tworzonych w BGPW ale pozwalamy sobie też na pewną dowolność. Nasza kartoteka słów kluczowych budowana jest zgodnie z metodyką budowy tezaurusów, ale wykorzystano tylko metodę tworzenia jednostek leksykalnych. Słowa kluczowe posiadają formę rzeczowników i najczęściej są to terminy używane w literaturze fachowej. Jeżeli to możliwe zachowujemy naturalny szyk wyrazów. Ponieważ dokument ikonograficzny nie posiada tekstu przeszukiwalnego staramy się wpisywać w ten atrybut jak najwięcej fraz aby użytkownik otrzymał pożądaną informację. W przypadku książek każde słowo z tekstu wchodzi do słownika. Natomiast dokument ikonograficzny wymaga innej formy odtworzenia jego treści. Uznaliśmy, że będą to słowa kluczowe, często wzbogacane o propozycje naszych użytkowników. Część haseł, które nie wchodzi w skład kartoteki haseł wzorcowych BGPW dobierana jest w taki sposób aby nie odbiegała leksykalnie od pozostałych. Zawsze jednak są to słowa, które najlepiej opisują zawartość fotografii. Wśród nich spotkać można dużo nazw własnych oraz nazwy osobowe we wszystkich możliwych formach (np. autor obiektu). Im więcej słów kluczowych tym łatwiejsze zidentyfikowanie cyfrowej postaci dokumentu bibliotecznego. W przypadku fotografii stosujemy kilkanaście a nawet więcej słów kluczowych mających ułatwić nie tylko wyszukanie zdjęcia użytkownikom, ale i analizę jego zawartości. Stosowane w przypadku fotografii Poddębskiego słowa kluczowe, tworzone były na podstawie m. in. słów występujących w Katalogu Zabytków Sztuki, encyklopedii PWN, Słownika Terminologicznego Sztuk Pięknych pod red. Kozakiewicza, częściowo również intuicyjnie. Wiele bibliotek cyfrowych, opracowujących dokumenty ikonograficzne, nie docenia tego atrybutu. Słowa kluczowe bywają bardzo skromne w związku z tym nie spełniają swojej roli jaką jest rozszerzenie dostępu do informacji o tych dokumentach. Istnieje bardzo małe prawdopodobieństwo, że użytkownik będzie poszukiwał np. fotografii lub pocztówki wg tytułu, który został sformułowany przez katalogującego. W przypadku dokumentu ikonograficznego atrybut „Subject and Keywords jest najważniejszym elementem opisu.

Opis (Description) – Wg twórców formatu to element przeznaczony na wyliczenie zawartości treści źródła. W przypadku Biblioteki Cyfrowej BGPW atrybut Description, zawiera informacje, które w formacie MARC 21 zaliczamy do uwag i umieszczamy w polach o etykietach 5XX. Są to na ogół dane historyczne, daty budowy obiektów, informacje o autorze obiektu (architekt) i inne.

Współtwórca (Contributor) - Atrybut zawiera nazwy współtwórców fotografii a więc zakładów fotograficznych, drukarzy, autorów dokumentów towarzyszących, redaktorów, twórców kolekcji itd. Nazwy te są zgodne z formą ustaloną w kartotece haseł wzorcowych.

Data (Date) – W ten element DC wpisujemy datę wydania oryginału. Datę odbitki podajemy w polu opis.

Źródło (Source) - Pole zawiera link do pierwotnego źródła, z którego pochodzi opisywany w bibliotece cyfrowej dokument czyli do opisu bibliograficznego w katalogu lokalnym Aleph BGPW. Jest to wierna kopia rekordu z katalogu NUKAT.

Powiązania/Relacja (Relation) - W tym polu na ogół umieszczamy nazwy serii lub podkolekcji. W przypadku fotografii Henryka Poddębskiego jest to sztucznie utworzona seria: Fotografia polska przed 1939 r.

Prawa własności (Rights Management) - pole zawiera informacje o prawach własności do źródła czyli oryginału. Czasami prawa własności oprócz BGPW posiada również spadkobierca. Tak się stało w przypadku zdjęć Henryka Poddębskiego. Prawa własności do jego spuścizny posiada córka Pani Krystyna Kukieła. W przypadku cyfrowej postaci eksponatów muzealnych prawa zachowało Muzeum PW.

Zdjęcia, do celów archiwizacji zostały zdigitalizowane w formacie TIFF (rozdzielczość 600dpi), a do publikacji w bibliotece cyfrowej skonwertowane do formatu JPEG (300 dpi). Skanowanie zdjęć wykonywane było w skali szarości o głębi 256 (8bit/px). Do zdjęć niezbyt wyraźnych w nielicznych przypadkach stosowana była maska wyostrzająca, nieprzekraczająca 100%, promień wyostrzania nieprzekraczający 1,3 pikseli), w kilku przypadkach zdjęcia były rozjasniane. Miniatury, zostały zeskanowane w rozdzielczości 200 dpi i stanowią ok. 10% wielkości podstawowego obrazka (jpg). Na zdjęciach dodajemy znak wodny, a w zasadzie nadruk, w prawym dolnym rogu, o treści „Copyright © 2008, BCPW”, na tyle mały, żeby nie przeszkadzał odbiorcy. Nie jest to oczywiście wystarczające zabezpieczenie fotografii przed użyciem przez osoby, które zamierzają je dalej wykorzystać, tylko małe utrudnienie, mogące zniechęcić poszukiwaczy zdjęć pasujących im do własnych serwisów internetowych.

Dokumenty ikonograficzne, ze względu na ich specyficzne cechy, m.in. brak informacji wymaganych do stworzenia opisu bibliograficznego, zalicza się do najtrudniejszych do skatalogowania materiałów bibliotecznych. Na tego typu dokumentach jest bardzo mało tekstu lub wcale. Osoba katalogująca musi przetworzyć informacje wizualne na poszczególne strefy opisu bibliograficznego. Podstawę do stworzenia rekordu stanowią dane wydobyte z dokumentu, zinterpretowane z warstwy wizualnej, wzbogacone o informację spoza podstawowego źródła.

Z takimi problemami będą się borykać bibliotekarze tworzący kolekcje ikonograficzne w Bibliotekach Cyfrowych. Nie zmieni tego ani ukazanie się tak wyczekiwanej polskiej normy ani przepisy standaryzujące opis bibliograficzny. Brak danych zawsze będzie stwarzał

możliwość błędnej interpretacji warstwy wizualnej dokumentu ikonograficznego co może spowodować zwiększoną ilość dubletów. Być może zapobiegnie temu powołanie do grona twórców bibliotek cyfrowych odpowiednich fachowców (np. historyków sztuki) przygotowujących informacje bibliograficzne dla katalogera. Ponadto jest to jeszcze jeden powód aby przekształcić nasze biblioteki cyfrowe w wersje 2.0. Wkład internautów w uwiarygodnianiu informacji bibliograficznych będzie nieoceniony.

Powstaje też pytanie czy nasze wysiłki w kwestii wzbogacania opisu bibliograficznego w dużą ilość słów kluczowych są potrzebne i czy potrzebne jest kontrolowane słownictwo, słownik, tezaurus. Użytkownika wchodzącego z Internetu tak naprawdę nie interesuje jaki standard opisu jest stosowany. Zazwyczaj chce po prostu znaleźć zdjęcie swojej ulicy, miejscowości z której pochodzi, drużyny piłkarskiej, jeśli jest kibicem, czy np. zdjęcie gen. Hallera jeśli jest historykiem. Informacje, które go interesują to przede wszystkim data wykonania zdjęcia, autor zdjęcia, ewentualnie właściciel praw autorskich. W dalszej kolejności może zastanawiać się nad dotarciem do oryginału czy opisem stanu fizycznego zdjęcia. Nie oznacza to, że mamy zaniechać wprowadzania innych elementów opisu. Jeśli chodzi o słowa kluczowe to wśród propozycji wysuwanych np. na forum „Biblioteka 2.0” padł pomysł wykorzystania zapytań kierowanych przez użytkowników w wyszukiwarkach Bibliotek Cyfrowych, przy pomocy mechanizmu zbierającego zapytania ze wszystkich instalacji BC, z uwzględnieniem form błędnych, które można by umieszczać w słowniku synonimów. Alternatywą jest implementacja jakiegoś profesjonalnego tezausa, który byłby jednak pomocą głównie dla bibliotekarzy, mniej dla użytkowników. Zapytania użytkowników mogłyby pełnić rolę pomocniczą np. w wiązaniu synonimów. Przyszłość należy pewnie do struktur taksonomicznych i sieci semantycznej. Użytkownicy nie wykorzystują praktycznie w ogóle terminów ze słowników, którymi posługują się bibliotekarze – jak pokazała statystyka udostępniona przez Bibliotekę Uniwersytetu Wrocławskiego, z której wynika też, że tylko 4 % użytkowników bibliotek cyfrowych wyszukuje z wykorzystaniem ograniczonym do atrybutu „Słowa kluczowe”.

Przeprowadzenie badań zapytań użytkowników, nie tylko zresztą bibliotek cyfrowych oraz zestawienie ich ze statystyką określającą częstotliwość wykorzystywania przez czytelników słów kluczowych tworzonych przez bibliotekarzy i istniejących w opisach mogłoby dać materiał pozwalający określić preferencje i profile czytelników.

Kolekcja ikonograficzna w Bibliotece Cyfrowej PW jest w fazie rozwoju. Referat powstał na bazie naszych dotychczasowych doświadczeń z prac nad jej tworzeniem.

Bibliografia

1. Format MARC 21 rekordu bibliograficznego : dokumenty ikonograficzne : wersja 2008. Oprac. Agata Pietrzak. Warszawa 2008
2. Nahotko Marek: Metadane. Kraków 2004

3. Wikipedia, wolna encyklopedia: W: [on-line]. Dostępny w World Wide Web <http://pl.wikipedia.org/wiki/Metadane> [cytowany listopad 2008]
4. Marc Standards. W: [on-line]. Dostępny w World Wide Web: <http://www.loc.gov/marc/> [cytowany listopad 2008]
5. Dublin Core. W: [on-line]. Dostępny w World Wide Web: <http://dublincore.org/documents/1999/07/02/dces/> [cytowany listopad 2008]
6. Polska Norma PN-N-01152-4: Opis bibliograficzny. Dokumenty ikonograficzne (nie opublikowana)
7. Sanetra Krystyna: Format MARC21 rekordu bibliograficznego dla dokumentu ikonograficznego. Warszawa 2008 (nie opublikowana).
8. Forum Biblioteka 2.0, *Tezaurus dla bibliotek cyfrowych*;
9. Forum Biblioteka 2.0, *Hasła formalne, przedmiotowe, rzeczowe*
10. Lista 20 najczęściej zadawanych pytań w systemie dLibra. Statystyka ze strony Biblioteki Uniwersyteckiej we Wrocławiu:
<http://www.bu.uni.wroc.pl/bcredaktor/statystyki.php>
11. - Piotr Chojnacki, Radosław Morawski, *Digitalizacja i opracowanie fotografii z zasobu IPN w systemie ZEUS, ?* Konferencja naukowa „Fotografia w nowoczesnym archiwum”, Warszawa 27.10.2008
12. - Kazimierz Schmidt, *Standardy tak, ale jakie i czy za wszelką cenę?* Konferencja naukowa „Fotografia w nowoczesnym archiwum”, Warszawa 27.10.2008
13. - Górny Mirosław, Skutecki Jakub, *Problemy udostępniania zbiorów ikonograficznych w bibliotekach cyfrowych*, <http://www.inib.uj.edu.pl/konfer/ki/index.html>

Słowa kluczowe: Katalogowanie, Katalogi komputerowe, Zbiory specjalne, Metadane, Dokumenty elektroniczne, Dokumenty ikonograficzne