

Władysław Marek Kolasa

Biblioteki cyfrowe - wstęp do problematyki

W literaturze na temat bibliotek cyfrowych często rozpoczyna się rozważania od definiowania i rozważania problemu literatury w środowisku internetowym. Literaturę tę na ogół zgodnie definiuje się jako synonim literatury dostępnej przez medium internetowe. Takie podejście wydaje się słusznie i zasadne, ale wyłącznie w dyskursie krytycznoliterackim lub wśród socjologów kultury. Nie wydaje się jednak, aby imponderabilia tego rodzaju mogły być przedmiotem rozważań bibliotekarskich. Z uwagi na skomplikowaną naturę pojęcia powinniśmy raczej poczekać, aż na łamach socjologii kultury czy też krytyki literackiej sprawy dojrzeją i jako bibliotekarze tę wiedzę wykorzystać.

Nieco więcej możemy powiedzieć na temat pojęcia nieco szerszego, czyli samego dokumentu elektronicznego. W tym zakresie panuje już pewien consensus: mamy normy ISO (w tym ich polskie lokalizację) oraz cały szereg innych aktów prawnych, podstawowych w polskim ustawodawstwie. Sądzę, że trafną definicję podaje norma ISO 690-2, która określa, że *dokumentem elektronicznym, jest dokument istniejący w postaci elektronicznej dostępny za pośrednictwem techniki komputerowej*¹. Zwraca się tutaj uwagę na dwa podstawowe aspekty dokumentu elektronicznego: 1. trwałość – czyli, aby był utwalony na materialnym nośniku (wprawdzie rodzi się wątpliwość co to znaczy materialny nośnik: czy rozproszony Internet jest materialny? sprawa wydaje się co najmniej dwuznaczna); 2. zdolność do odtwarzania w formie umożliwiającej percepcję - i to wydaje się być problemem kluczowym. Zatem w istocie nie jest ważny nośnik, lecz możliwość jego wielokrotnego odczytu.


























Sprawy terminologiczne są jednak z natury złożone. Wraz z rozwojem techniki rodzą się na nowo rozmaite problemy zarówno techniczne, jak i prawne. Te ostatnie znalazły nawet oddźwięk w specjalnej ustawie dotyczącej ujednoczenia terminologii informatycznej, gdzie na nowo definiuje się terminy: „informatyczny nośnik danych”, „dokument elektroniczny”, „system teleinformatyczny” oraz „środki komunikacji elektronicznej”²

Publikacje cyfrowe to nowa jakość. Przede wszystkim należy uświadomić sobie, że nośnik ten pozwolił pokonać wiele barier w stosunku do publikacji tradycyjnej. Dla publikacji

¹ PN-ISO 690-2 Informacja i dokumentacja. Zarządzanie dokumentami.

² Ustawa o zmianie ustaw w celu ujednoczenia terminologii informatycznej z dnia 4 września 2008. Dz. U. 2008, nr 171, poz. 1056.

elektronicznej tracą znaczenie takie pojęcia, jak: miejsce, nakład i po części czas. Właściwości te powodują daleko idące skutki. Ograniczmy się do kilku kluczowych: z punktu widzenia edukacji, jest to przede wszystkim umożliwienie indywidualizacji tempa lektury, gdyż mamy do czynienia z asynchronicznym przekazem multimedialnym; z punktu widzenia wolności przekazu jest to w praktyce zupełna odporność na cenzurę; publikacje elektroniczne to wreszcie doskonały nośnik spontanicznego rozszerzania się kultury. Dokument elektroniczny jednoznacznie sprzyja globalizacji (bez względu na ocenę tego zjawiska). Na potwierdzenie tego jak jest to międzynarodowe ilustruje zrzut ekranu z ostatniego miesiąca 25 najaktywniejszych użytkowników Małopolskiej Biblioteki Cyfrowej (w podziale na państwa i regiony).

Domeny/narodowość Internautów					
	Domeny/Kraje		Strony	Żądania	Bajtów
	Poland	pl	408246	1001709	46.51 GB
	Nieznane	ip	100190	209104	9.27 GB
	Network	net	39083	66499	3.13 GB
	Commercial	com	24739	48422	2.58 GB
	Czech Republic	cz	8399	41337	4.12 GB
	France	fr	6053	12201	747.50 MB
	Netherlands	nl	6003	7916	841.45 MB
	Germany	de	3805	8489	1.60 GB
	Sweden	se	2818	5869	668.10 MB
	Russian Federation	ru	2361	3748	508.61 MB
	Croatia	hr	2060	2313	140.65 MB
	Ukraine	ua	623	2019	58.20 MB
	United Kingdom	uk	463	2093	21.34 MB
	Norway	no	454	846	18.50 MB
	Australia	au	446	2335	151.56 MB
	Austria	at	411	2486	20.65 MB
	Lithuania	lt	359	1459	12.59 MB
	Canada	ca	325	4929	287.56 MB
	USA Educational	edu	307	525	7.64 MB
	Belarus	by	282	787	6.80 MB
	Belgium	be	269	697	9.09 MB
	Denmark	dk	258	448	9.48 MB
	Ireland	ie	228	846	7.83 MB
	Italy	it	203	1072	10.46 MB
	European Union	eu	201	758	8.92 MB

Rys. 1, Najaktywniejsi użytkownicy MBC [październik 2008]

Na wykresie widnieją niemal wszystkie państwa Europy, a gdybyśmy szukali korelacji to byłoby nasi sąsiedzi oraz państwa wysoko rozwinięte. To jedynie wierzchołek listy. Liczba państw, z których użytkownicy korzysta badanego serwisu jest ogromna, praktycznie są to

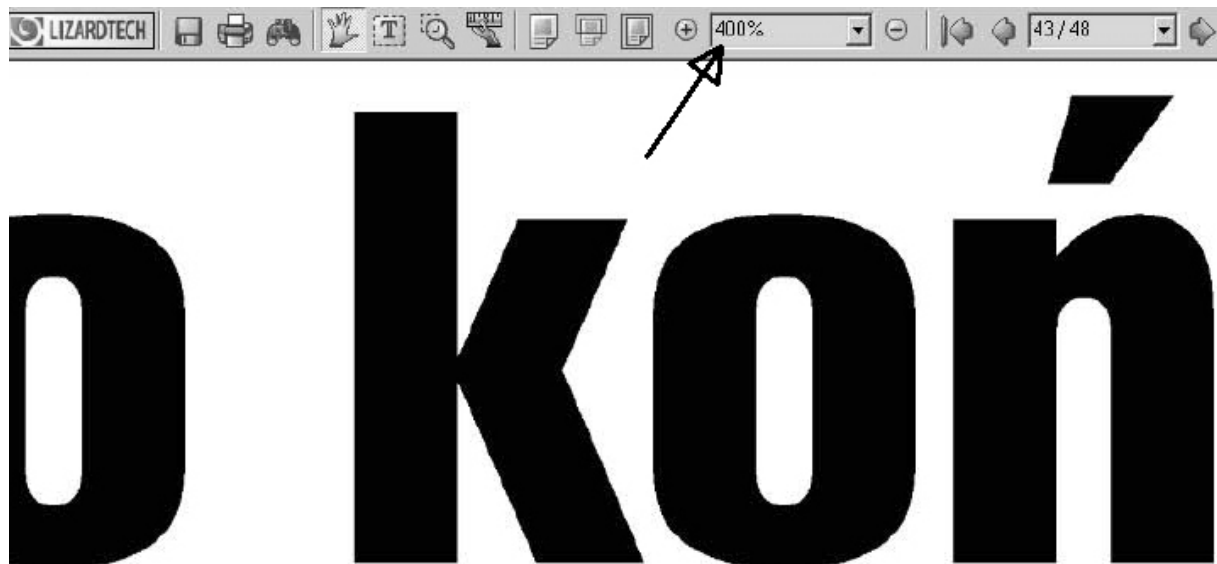
wszystkie duże państwa (w miesiącach szczytu do 190 domen), w tym nawet państwa najbardziej egzotyczne, (np. Niue – domena nu).

Teraz zwróć uwagę na kilka jakby podręcznikowych aspektów publikacji elektronicznych: po pierwsze - zwykle podkreśla się, że ten nośnik jest bardzo pojemny, więc jeśli uwzględnimy standardowe kodowanie ASCII, to teoretycznie na jednej małej (12 cm) płytce CD-ROM mieści się tyle co na półtora miliona stron tekstu (a to zapewne równowartość kilkunastu regałów z tradycyjną książką); po wtóre - publikacje elektroniczne umożliwiają wyszukiwanie pełnotekstowe, co jest swoistą wartością dodaną (szczególnie przydatną w selektywnym korzystaniu ze źródła); po trzecie - istnieje niebywała łatwość replikowania i kopiowania, co ważniejsze przy każdym replikowaniu i kopiowaniu ten nośnik samoczynnie się odświeża. W tym miejscu nieuchronnie pojawia się pytanie trwałość nośnika? i tu jest sprawa dyskusyjna (w szczególności jeżeli próbujemy zadać sobie pytanie jak trwały jest nośnik online – gdyż z uwagi na liczbę kopii nie jest on w sensie fizycznym obiektem jednostkowym, lecz rozproszonym). Po czwarte - publikacje elektroniczne umożliwiają łatwy sposób generowania rozmaitych statystyk lektury, a raczej dostępu - to jest kolejna wartość dodana w stosunku do tradycyjnych publikacji; po piąte – stosunkowo niski koszt utrzymania, którego skalę uświadamia porównanie kosztów utrzymania serwera (lub dzierżawy hosta) i tradycyjnego magazynu – a jest to rząd trudny do porównania na korzyść tych pierwszych. I wreszcie niebywale niski koszt dystrybucji - w przypadku publikacji elektronicznych koszt ów jest na tyle niski, że w praktyce można go zaniedbać.

Schodząc bliżej aspektom praktycznym myślę, że bibliotekarze, twórcy bibliotek cyfrowych powinniśmy raczej operować pojęciem e-książki, a nie dokumentu elektronicznego. Ostatnie pojęcie jest bowiem zbyt szerokie znaczeniowo i z uwagi na swoją pojemność grozi na utknięcie w ślepych zaułku (na potwierdzenie tej tezy można np. zadać przekorne pytanie: czy w bibliotekach cyfrowych należy gromadzić blogi czy posty internetowe?). Aby uniknąć takich problemów na potrzeby obecnego etapu rozwoju technologii biblioteki powinny zawęzić zakres gromadzonych obiektów. Najkorzystniej zatem jako synonim obiektu elektronicznego gromadzonego w bibliotece będzie operować pojęciem e-książki, rozumianej książkę tradycyjną na nośniku elektronicznym. Tak rozumiana publikacja ma w porównaniu z książką tradycyjną wiele cech wspólnych, w szczególności identyczny sposób tworzenia, identyczny sposób redakcji, recenzowania, składu; zatem etapie wydawniczym wszystkie cechy są identyczne – jedyna różnica to nośnik, a dokładnej brak papieru – jest ona zatem utrwalona na nośniku elektronicznym w specjalnym formacie.

Kolejnym ważnym przykazaniem, myślę, że ono powinno być szczególnie ważne dla tych, którzy tworzą zasoby elektroniczne jest bezwzględne wymaganie, aby serwowane obiekty były dostępne w jakości nie gorszej niż ich odpowiednik drukowany. Mamy bowiem liczne przykłady wartościowych treści, które są udostępniane w zbyt niskiej rozdzielczości. Nawet bardzo zdeterminowany czytelnik, zrezygnuje z lektury takiego dzieła po 3-4 min męczącej dla oczu lektury i zada sobie trud sięgnięcia do wersji papierowej. Jeśli zatem przygotowujemy dokument elektroniczny, spróbujmy choćby na chwilę wcielić się w rolę czytelnika i ocenić czy spełnia on minimalne wymogi higieny czytania. Przy odrobieniu trudu każdą publikację można przygotować w jakości porównywalnej z oryginałem, a nierzadko (gdy dysponujemy składem w wersji poligraficznej) nawet przewyższającej parametry drukowanego oryginału.

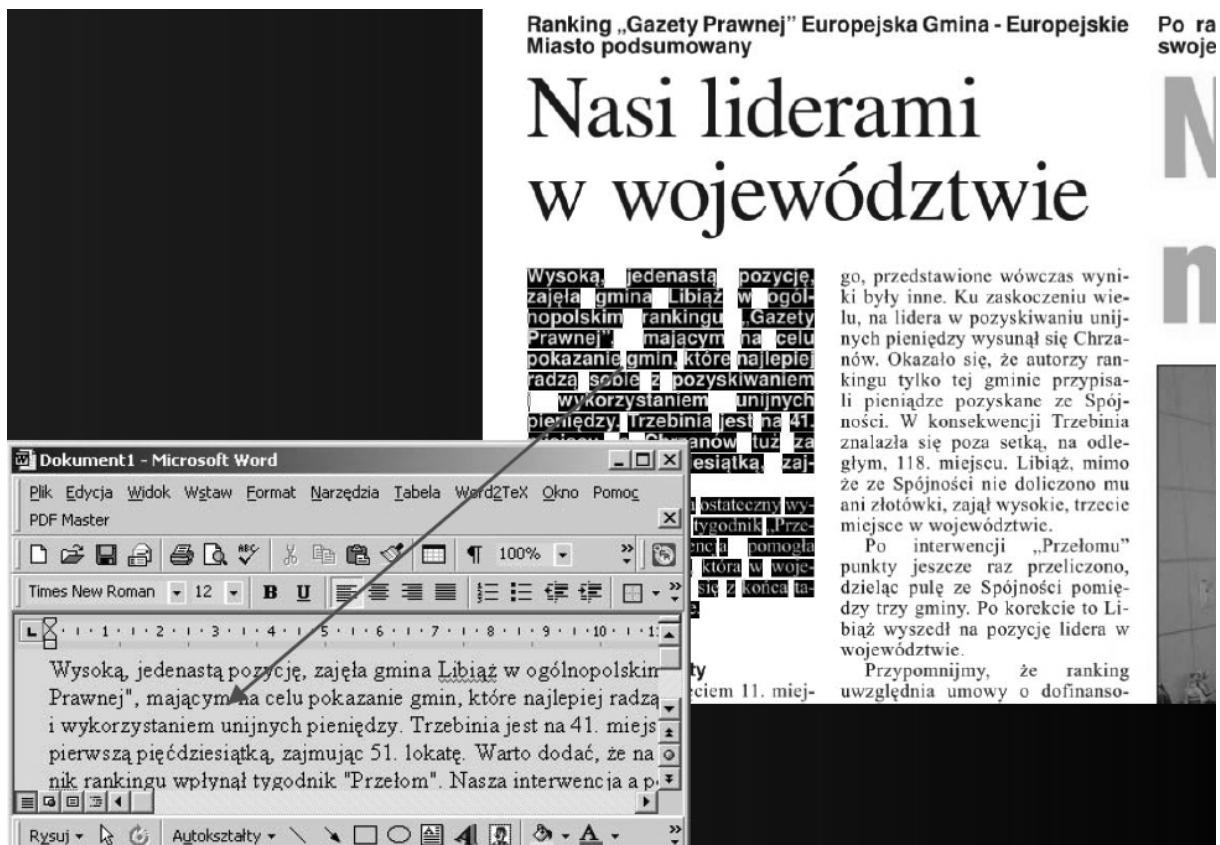
Publikacja elektroniczna względem drukowanej może oferować to pewne wartości dodane, w szczególności zoom i OCR. Jakkolwiek dobór maksymalnego zoomu jest sprawą sporną – bo nie wiadomo jak ile calu przekątnej będzie miał standardowy monitor przyszłości, warto jednak już teraz zadbać, o to, aby obiekt był serwowany w odpowiedniej rozdzielczości, gdyż ona delimituje granice powiększenia. Sądzę, że dla obiektów nowych (uzyskanych z plików wektorowych - minimalną wartością jest 600 ppi, zaś dla obiektów skanowanych 300-600 ppi).



Rys. 2, Publikacja w opcję zoom (400%, 600 ppi).

Kolejna wartość dodana to możliwość osadzania warstwy tekstowej. Z uwagi na łatwość i stosunkową dużą skuteczność programów OCR warto z tej możliwości korzystać i

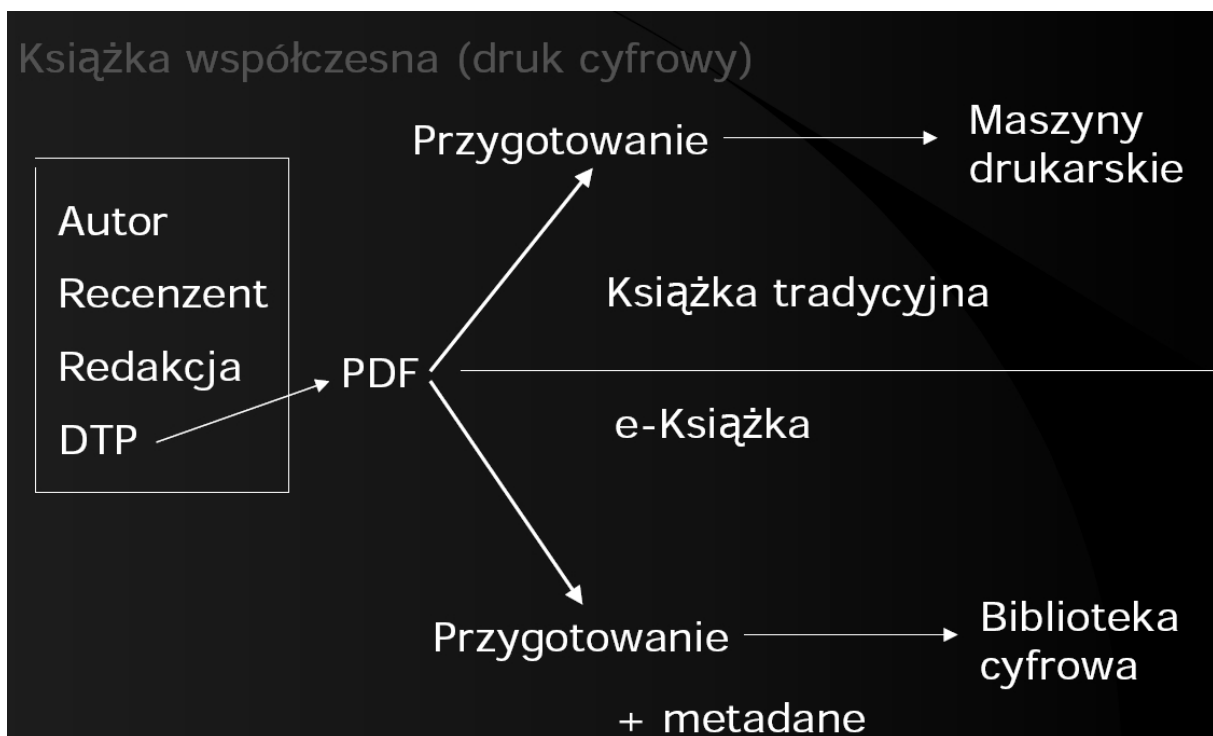
udostępniać obiekty wraz z osadzonym tekstem. Powody są dwa: zawarty w publikacji tekst może być indeksowany przez system i w sposób istotny wspomagać wyszukiwanie, w tym znaczeniu pełni funkcję pomocniczą wobec metadanych i umożliwia dotarcie do publikacji źle lub zbyt wąsko skatalogowanych (funkcja indeksująca); z drugiej strony – osadzony lub jawny tekst to ogromne udogodnienie dla użytkownika, który nie musi przepisywać wyszukanych fragmentów, lecz może je łatwo skopiować do własnego edytora (funkcja użytkowa). Problemu osadzania tekstu w każdym obiekcie nie należy jednak traktować dogmatycznie. Ocenia się bowiem, że już w niedalekiej przyszłości same przeglądarki będą w stanie wyzyskać tekst z dowolnego, to jednak – podkreślny - nie zaspokoi wskazanej wyżej funkcji indeksującej).



Rys. 3, Publikacja hybrydowa z osadzoną warstwą testową

Wywód warto uzupełnić porównaniem techniki powstawania książki tradycyjnej i elektronicznej, gdyż jest ona w wielu punktach tożsama, zaś różnice sprowadzają się do kilku elementów na końcu łańcucha produkcji. Poniższy schemat (rys. 4) pokazuje czym może się różnić lub czym się różni książka tradycyjna (przy zastosowaniu techniki druku cyfrowego)

od tej samej książki publikowanej na platformie biblioteki cyfrowej. Na etapie reakcji i składu proces jest identyczny; rozmaicie może być natomiast potraktowany produkt finalny, czyli PDF. Może być ona bowiem skierowany na maszyny drukarskie (i to jest podejście tradycyjne), ale może być równoległe lub wyłącznie skierowany do biblioteki cyfrowej i po odpowiednim przygotowaniu i skatalogowaniu metadanych, opublikowany być na platformie bibliotecznej. Nieco inaczej wygląda ów proces jedynie dla publikacji przygotowywanych w technice z użyciem klisz lub w przypadku pozyskiwania książki dawnej. W tym przypadku w łańcuchu produkcji w miejsce nieistniejącego PDF dochodzi etap skanowania.



Rys. 4, Etapy powstawania książki tradycyjnej i e-książki

Schemat obrazuje złożoność zadań, przed którymi staną twórcy bibliotek cyfrowych. Bibliotekarze cyfrowi powinni bowiem mieć wiedzę na temat zagadnień, które wcześniej były zarezerwowane wyłącznie dla znawców poligrafii (w szczególności chodzi o przetwarzanie tekstu, składu, konwersji, skanowania); dochodzi do tego pokaźna gama części zadań związanych z obsługą procesu wydawniczego (w szczególności regulacja praw autorskich, ale również może doraźnie chodzić o skład DTP); i wreszcie cała grupa zagadnień ściśle informatycznych, tam gdzie mamy do czynienia z usługami hostingowymi lub administrowaniem własną infrastrukturą serwerową. Na gruncie współczesnego kształcenia

bibliotekoznawczego sprawy w zasadzie nowe i czekają na metodyczne wdrożenie do programów studiów.

Na zakończenie uwag wstępnych warto jeszcze przytoczyć katalog pobudek i motywów, dla których tworzy się i rozwija biblioteki cyfrowe. Jest ich co najmniej pięć. Najszczytniejszą jest cel archiwizacji i ochrony zbiorów, istotny w szczególności dlatego, że duża część zbiorów od II połowy XIX wieku drukowana była na kwaśnym papierze i zbiory te ulegają samoistnej degradacji, zabezpieczenie dla potomnych jest zatem podstawowym imperatywem ich digitalizacji. Mniejszy zakres ma pobudka ochrony przed kradzieżami i uszkodzeniami, to jest jak gdyby zupełnie odrębny tor rozważań (istotny w przypadkach identyfikacji, konserwacji i restauracji). Trzeci powód to udostępnianie kopii zbiorów rzadkich, zatem takich, które występują w pojedynczych egzemplarzach (np. archiwalia, zbiory muzealne) lub zachowały się w nielicznych kopiach (staroduki, inkunabuły). Udostępnienie ich w sieci jest bodaj jedyną szansą na ich lekturę, gdyż na co dzień są przechowywane w specjalnych warunkach, a dostęp do nich możliwy jest tylko dla nielicznych i obwarowany wieloma ograniczeniami (nie wspominając konieczności osobistego wyjazdu do miejsca ich przechowywania). Powód czwarty związany jest z szerokim udostępnianiem dostępnych w domenie publicznej dzieł kanonu kulturowego (w tym literackiego); kierunek ten - podkreślmy – legło u podstaw tworzenia pierwszych bibliotek cyfrowych i jest obecnie już bardzo wyeksploatowany. I wreszcie pobudka piąta – udostępnienie szerokiej publiczności publikacji o zasięgu lokalnym; w tym przypadku dominuje tu funkcja promocyjna i połączona z procesem ich remediacji, czyli włączeniu do obiegu lekturowego. Kierunek ów jest aktualnie silnie eksploatowany przez większość polskich bibliotek cyfrowych. Na koniec katalogu motywów pojawia się pytanie o względy komercyjne. Wydaje się jednak, że względy zarobkowe nie powinny przyświecać bibliotekom, gdyż handel jest domeną księgarń. Egzemplifikacją tego skądinąd ciekawego problemu jest np. serwis Google Books przez niektórych postrzegany jako biblioteka cyfrowa. Należy podkreślić, że ani ten serwis, ani żaden podobnie działający, nie może być postrzegany jako biblioteka, gdyż całość, albo przeważająca część zasobu jest dostępna czasowo (biblioteki zaś udostępniają wieczyście), wydawca zaś ma prawo własny zasób w każdej chwili odpublikować. W efekcie zasobu dostępnego on-line w takich serwisach nie można ani w odpowiedzialny sposób cytować, ani polecać do celów edukacyjnych, tym bardziej do celów naukowych.

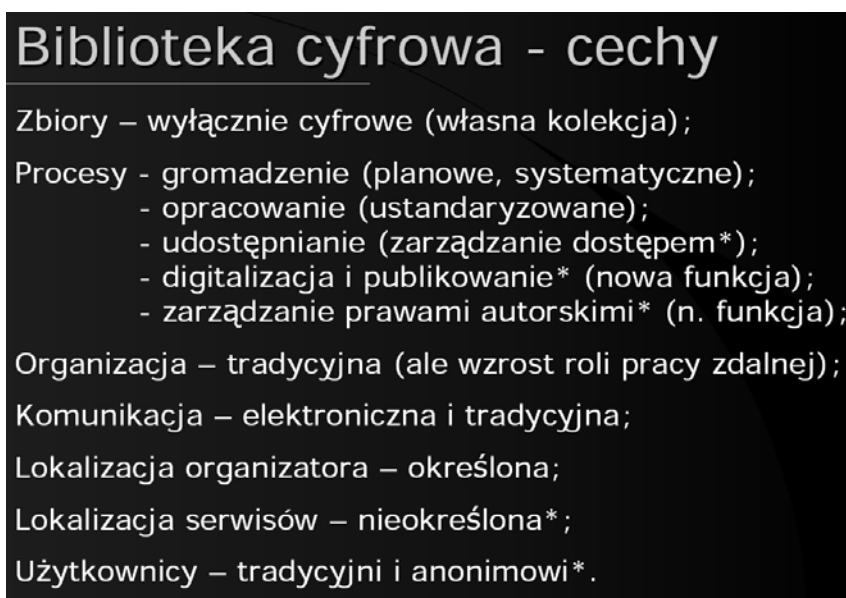
Wskazany wyżej argument uświadamia konieczność uporządkowania pojęcia biblioteki działającej w środowisku internetowym, gdyż zarówno w literaturze, jak i w potocznym obiegu funkcjonują bowiem obok siebie rozmaite pojęcia: biblioteka elektroniczna, cyfrowa, wirtualna, polimedialna, hybrydowa i in. Wieloznaczność owa jest w dużej mierze uwarunkowana historycznie. Uznaje się, że do szerokiego obiegu pojęcia te wprowadził w 1992 roku w jednym ze swoich artykułów M. Buchland, podobny podział został następnie w 1996 roku utrwalony przez P. Barkera w *Encyclopedia of Library Information Science*³. Także na polskim gruncie jeszcze na długo przed powstaniem pierwszych polskich bibliotek cyfrowych przez łamy wydawnictw przetoczyła się dyskusja na ten temat, głos zabrali m.in.: E. Chmielewska-Gorczyca, J. Czermiński i A. Radwański. Aktualnie rozważania te mają charakter wyłącznie historyczny, gdyż z uwagi na postęp technologiczny dyskutowane problemy straciły na aktualności. Wyczerpujący opis toczonych wówczas dyskusji zawiera monografia D. Grygrowskiego⁴.

Osobiście sędzę, że adekwatnym pojęciem dla bibliotek działających w środowisku internetowym jest pojęcie *biblioteka cyfrowa* (por. rys. 5), która: 1. podobnie jak biblioteka tradycyjna opiera się na własnej kolekcji (to różni ją od wirtualnej); 2. zachodzą w niej te same procesy, co w bibliotece tradycyjnej (gromadzenie, opracowanie, udostępnianie), nadto pojawiły się nowe (digitalizacja, zarządzanie prawami autorskimi); 3. jest zorganizowana podobnie jak biblioteka tradycyjna; 4. opiera się na komunikacji mieszanej (elektronicznej i tradycyjnej, z wyraźnym wzrostem znaczenia pracy zdalnej); 5. przedstawicielstwo ma określoną lokalizację (wynika to z uwarunkowań prawnych); 6. usługi i serwisy mogą nie mieć określonej lokalizacji; 7. występują dwie grupy użytkowników: tradycyjni (zarejestrowani) i anonimowi. Dodatkowego rozwinięcia wymaga punkt drugi, czyli specyficzne procesy biblioteczne, które mają określony porządek, literaturę i tradycję. Należy w szczególności wymagać, aby także w bibliotekach cyfrowych gromadzenie było planowane i systematyczne, zaś opracowanie – ustandaryzowane. Warto pamiętać aby to co najlepsze w tradycyjnym bibliotekarstwie stosować także w bibliotekarstwie nowej generacji. Wyliczenie warto uzupełnić jeszcze kilkoma uwagami natury jakościowej. W szczególności odpowiedzieć na pytanie: *czym się różni biblioteka cyfrowa od zwykłych stron www?* Myślę, że prócz wymogu wysokiej jakości zasobów w sensie technicznym, czym wspomniałem już na wstępie, należy także wymagać, aby serwowane zasoby odpowiadały wysokim wymogom

³ Encyclopedia of Library and Information Science. Vol. 59, New York 1996, s. 145-148

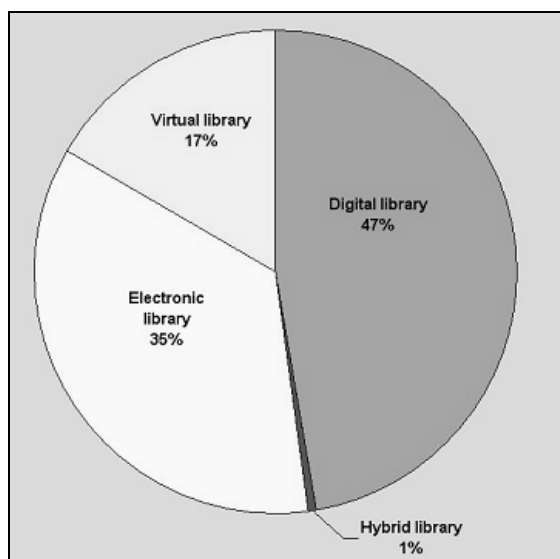
⁴ D. Grygowski, Dokumenty nieksiążkowe w bibliotece, Warszawa 2001 s. 157-164

jakości merytorycznej, z którymi powszechnie jest kojarzona biblioteka i informacja drukowana. A zatem w bibliotekach cyfrowych należy gromadzić: po pierwsze - informacje rzetelne (dotyczy w szczególności publikacji, przygotowanych z pominięciem procesu redakcyjnego – tu ciężar kontroli spada na barki bibliotekarza); po wtóre - winny być to publikacje o trwałej wartości, nie zaś twory publikowane czasowo; po trzecie – należy gwarantować trwałość linku i zasobu, gdyż publikację ulotną trudno odpowiedzialnie cytować w publikacji (postulat ten musi być absolutnie przestrzegany, to immanentna cecha biblioteki cyfrowej); po czwarte wreszcie – publikowane zasoby winny mieć uregulowany status prawny i ciężar tej kontroli spada na barki bibliotekarza.



Rys. 5, Cechy biblioteki cyfrowej

Z drugiej strony problem nazewnictwa weryfikuje praktyka. Z przeprowadzonych przeze mnie w ciągu dwóch ostatnich lat eksperymentów na słownictwem indeksowanym przez z Google wynika, że 37-47% twórców bibliotek sieciowych określa się jako *biblioteka cyfrowa* (*digital library*) lub 17-20% *biblioteka wirtualna* (*virtual library*) – razem (54-67%); niezmienną liczbę trafień 35-38% uzyskało pojęcie *biblioteka elektroniczna* (*electronic library*) – która wszelako nie jest synonimem biblioteki cyfrowej, traci zaś systematycznie określenie *biblioteka hybrydowa* (*hybrid library*) – 5% w 2007 i tylko 1% w 2008.



Rys. 6, Określenia stosowane przez twórców bibliotek sieciowych (2008 r.)

Rozwój bibliotek cyfrowych następował w kilku okresach, a pierwszych śladów owej idei możemy się doszukać jeszcze na długo przed powstaniem komputera. Za protoplastę idei fikcyjnej biblioteki cyfrowej możemy uznać literacką wizję H.G.Wellsa utrwaloną w jednym z jego esejów w z 1937 roku. Nieco dojrzalszą wizję (teoretyczna maszyna Memex) przedstawił w 1945 roku V.Bush; zaś na pierwszy eksperyment przeprowadzony przez D. C. Engelbarta należało poczekać do 1968 roku (system Augment).

Nieprzerwaną historię bibliotek cyfrowych rozpoczyna natomiast dopiero w 1971 roku inicjatywa M. Harta – Gutenberg Project. W latach 70. i 80. pojawiło się jeszcze kilka serwisów (np. Perseus Digital Library – w 1987) oraz kilkadziesiąt w pierwszej połowie lat. 90 (np. Bartleby – 1993; The Humanities Text Initiative - 1994)⁵; zaś ich uwieńczeniem jest serwis American Memory założony w roku październiku 1994. Okres 1971-1995 można by nazwać pionierskim, gdyż tworzone repozytoria ciągle eksperymentują, współpracują firmami nad ich udoskonalaniem sprzętu, jednocześnie technologie internetowe zmieniają się w błyskawicznym tempie. Aby obrazowo przedstawić ten okres rozwoju bibliotek cyfrowych można się posłużyć żartobliwą metaforą – „czas elektronicznych inkunabułów”.

W rozwoju dedykowanych rozwiązań niebagatelną rolę należy przypisać komercyjnym serwisom dystrybucji czasopism elektronicznych (np. Tulip, Swetscan, Galen), a nade wszystkim projektowi American Memory, na którego potrzeby przeznaczono duże kwoty na badania rozwojowe. Ostatecznie zatem realny, masowy rozwój bibliotekarstwa

⁵ Pełny wykaz bibliotek cyfrowych (do 1995 roku) – zob. Timeline of the Open Access Movement: <http://www.earlham.edu/~peters/fos/timeline.htm> [2009.04.13].

cyfrowego należy datować na 1995 - zatem trawa on niespełna 15 lat. Jest to okres zbyt krótki na rzetelną refleksję naukową, stąd wiedza teoretyczna na ten temat wciąż jest powijakach. Wciąż przeważa praktyka i eksperymenty i stan ów zapewne nieszybko się zmieni.

Od 1995 roku (acz jest to data umowna) rozpoczyna się okres bibliotek cyfrowych pierwszej generacji. W sensie technicznym były to twory nieskomplikowane. Zwykle tworzono je jako proste witryny HTML zawierające wykazy hiperłączy do publikacji (tzw. model „strona www”), które z czasem, w miarę powiększania się zbiorów, wyposażono w mechanizmy wyszukiwawcze i system metadanych (tzw. model „katalog obiektów cyfrowych”). Wadami tych rozwiązań był jednak brak mechanizmów konserwacyjnych oraz brak kontroli dostępu i zarządzania zasobem i metadanymi. Obydwa modele odzwierciedlały filozofię pierwszych bibliotek cyfrowych (np. Projekt Gutenberg [1971] czy American Memory [1994], a nawet Polska Biblioteka Internetowa [2001]); charakterystycznym rysem wspomnianych projektów było programowe ukierunkowanie na udostępnienie kanonu kulturowego (klasyka literacka, dziedzictwo historyczne) oraz założenie, że dostęp do zbiorów winien być nielimitowany. W efekcie biblioteki I generacji realizowały przede wszystkim cel archiwalny (nie biblioteczny); z drugiej zaś strony udostępniane zasoby nie wymagały stosowania żadnych mechanizmów zabezpieczeń i kontroli dostępu, gdyż były z natury swej wolne od ograniczeń wynikających z ochrony praw autorskich.

Cechy bibliotek I generacji

Cechy bibliotek I generacji:

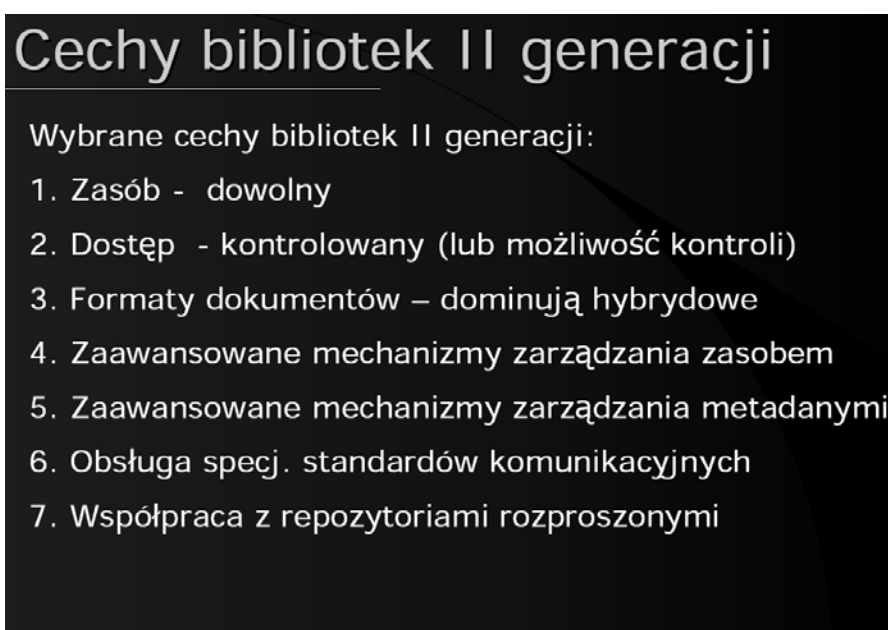
1. Zasób - dzieła z domeny publicznej
2. Dostęp - nielimitowany
3. Formaty dokumentów (tekstowe lub graficzne)
4. Brak kontroli dostępu i zarządzania zasobem
5. Słabe mechanizmy zarządzania metadanymi

Dwa modele:

- tzw. model „strona www”,
- tzw. model „katalog obiektów cyfrowych”

Rys. 7, Cechy bibliotek cyfrowych I generacji

Takie usytuowanie celów nie sprzyjało rozwojowi bibliotek cyfrowych. Było bowiem niewystarczające do realizacji funkcji bibliotecznych, gdyż stało w sprzeczności z jedną z zasadniczych cech biblioteki, zakładającej, że proces gromadzenia jest permanentny (stały, ciągły, powtarzalny), a także otwarty na nowe formy dokumentów. Z drugiej zaś strony istniała uzasadniona niechęć wydawców do nielimitowanego udostępniania, gdyż nie zabezpieczało to ich interesu ekonomicznego. Stało się to powodem doskonalenia platformy udostępniania i pracy nad nowymi formatami. Już od około 2002 roku zaczęły powstawać serwisy nowego typu – umownie nazywam je bibliotekami II generacji.



Rys. 8, Cechy bibliotek cyfrowych II generacji

W odróżnieniu projektów starych ich funkcjonalność pozwala pokonać wspomniane wyżej bariery i zagwarantować bezpieczny rozwój. W bibliotekach system oferuje funkcjonalność zbliżoną do tradycyjnej biblioteki, w szczególności: 1. pozwala na pełną kontrolę udostępniania zarówno na poziomie użytkownika, jak i zasobu (autoryzacja, autentykacja, połączenia szyfrowane), zaś z punktu widzenia użytkownika jest samoobsługowy; 2. obsługuje dokumenty w dowolnych formatach (tekstowe, graficzne, aplikacje) i strukturze (proste, złożone) oraz różne typy metadanych (DublinCore, EAD DTD, MARC XML i in.); 3. oferuje rozbudowane mechanizmy wyszukiwawcze (OPAC) i jest otwarty na alternatywne metody dostępu (wyszukiwarki, współpraca z system rozproszonymi) oraz umożliwia wyszukiwanie pełnotekstowe; 4. ponadto m.in.: pozwala na

pracę grupową; gwarantuje stałość punktu dostępu (hiperłącza) do publikacji; wspiera obsługę wydawnictw publikowanych wyłącznie w wersji elektronicznej (mechanizm wersjonowania); zachowuje informację o zasobach zmienianych i usuniętych; obsługuje najnowsze standardy komunikacyjne i informatyczne (np. RDF); pozwala na personalizację ustawień; ma wbudowane mechanizmy informacyjne (RSS, powiadomienia e-mail) i statystyczne oraz zintegrowany system zarządzania zasobem (moduł administracyjny i redakcyjny). Jest ponadto przystosowany do pracy na różnych platformach zarówno na poziomie serwera, jak i klienta.

Kolejny etap w rozwoju bibliotek cyfrowych – to próba scalania wspólnym interfejsem projektów na bazie rozmaitych standardów komunikacyjnych, w szczególności OAI-PMH. Proces ten rozpoczął się od ok. 2004. Największymi serwisami tego typu są: OAISTER, OpenDOAR, DOAJ i Institutional Archives Registry

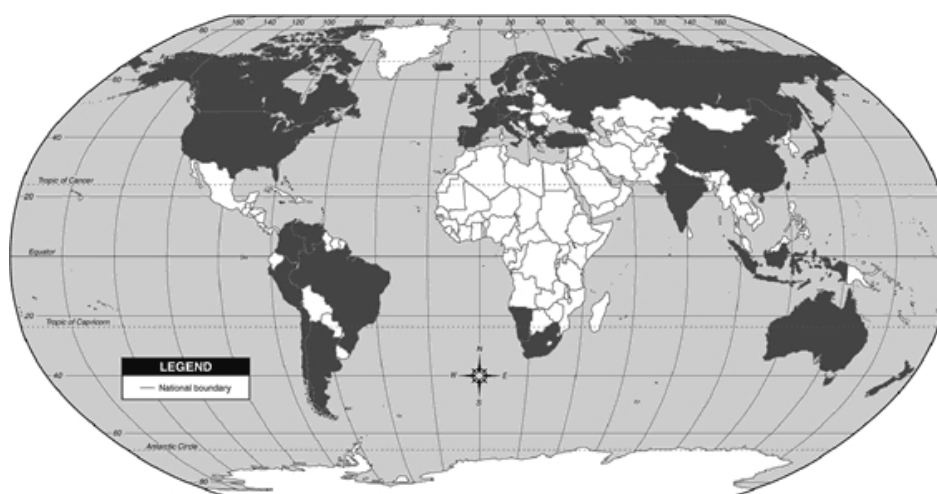
Aktualnie działa kilkanaście bibliotek cyfrowych o zasięgu globalnym. Listę największych otwiera: Internet Archive (IA) - jedno z największych światowych repozytoriów uniwersalnych oferujących ponad 1 mln publikacji; odrębne miejsce zajmuje Google Books (9,7 mln publikacji), który jakkolwiek jest serwisem przede wszystkim handlowym, oferuje ponad 817 tys. w pełnym dostępie z domeny publicznej oraz – co istotne – integruje swoje zasoby z globalnym serwisem WorldCat. Dużą rolę odgrywają też: ogólnoświatowe repozytorium tekstów należących do domeny publicznej - Project Gutenberg Consortia Center (75 tys.), będący nieformalnym następcą Gutenberga. Serwis ten ściśle współpracuje z kolejnym dużym komercyjnym serwisem o zasięgu globalnym - WorldLibrary.net (ok. 500 tys.). Liczące się zasoby oferuje też OCLC – serwis NetLibrary (ok. 100 tys.) oraz powołany niedawno serwis Europeana.

Bibliotek o mniejszym zasięgu jest znacznie więcej. Wnikliwa analiza danych zaczerpniętych z serwisu OAISTER⁶ pozwala nawet pewne uogólnienia. Można bowiem stwierdzić, że w prawie w każdym rozwiniętym kraju istnieje: przynajmniej 1 biblioteka o aspiracjach narodowych⁷; przynajmniej 1 udostępniająca dziedzictwo historyczne.; przynajmniej 1 serwująca teksty klasyki literackiej; nadto liczne biblioteki czasopism historycznych; liczne biblioteki regionalne (stanowe); liczne biblioteki uniwersyteckie (np. szczególnie w Holandii i Niemczech) oraz ogromna liczba bibliotek instytucjonalnych. Rolę

⁶ OAISTER - <http://www.oaister.org/> [2009.04.13]

⁷ Wybór najważniejszych projektów narodowych: http://mbc.malopolska.pl/dlibra/text?id=dlib_world [2009.04.13]

integracyjną dla większości serwisów odgrywają multiwyszukiwarki globalne (repozytoria rozproszone). Największy zasięg ma OAISTER – w listopadzie 2008 roku z serwisem współpracowało 1034 bibliotek z całego świata (w tym wszystkie polskie biblioteki oparte na systemie dLibra); serwis umożliwiał dostęp do ponad 18,3 mln rekordów. Z punktu widzenia integracji kluczową rolę odgrywają tu standardy, w szczególności standard metadanych DublinCore oraz standard komunikacyjny OAI-PMH. Nie oznacza, to że nie istnieją inne konsorcja; blisko z wymienionym serwisem współpracuje sieć bibliotek The Digital Library Federation (DLF) – 62 biblioteki oferujące razem 2,6 mln rekordów oraz sieć bibliotek wspierających standard MODS - 20 bibliotek oferujących 624 tys. rekordów. Nieco inną rolę pełnią specjalistyczne wyszukiwarki repozytoriów, np. OpenDOAR - zorientowany europejsko portal akademickich repozytoriów naukowych Open Access (1300 repozytoriów); Institutional Archives Registry - identyczny portal zorientowany na dostawców amerykańskich (955 repozytoriów), czy DOAJ - największe światowe repozytorium czasopism naukowych udostępnianych bezpłatnie (zawartość ponad 3740 tyt.).



Rys. 9, Biblioteki cyfrowe świata w rozkładzie terytorialnym⁸

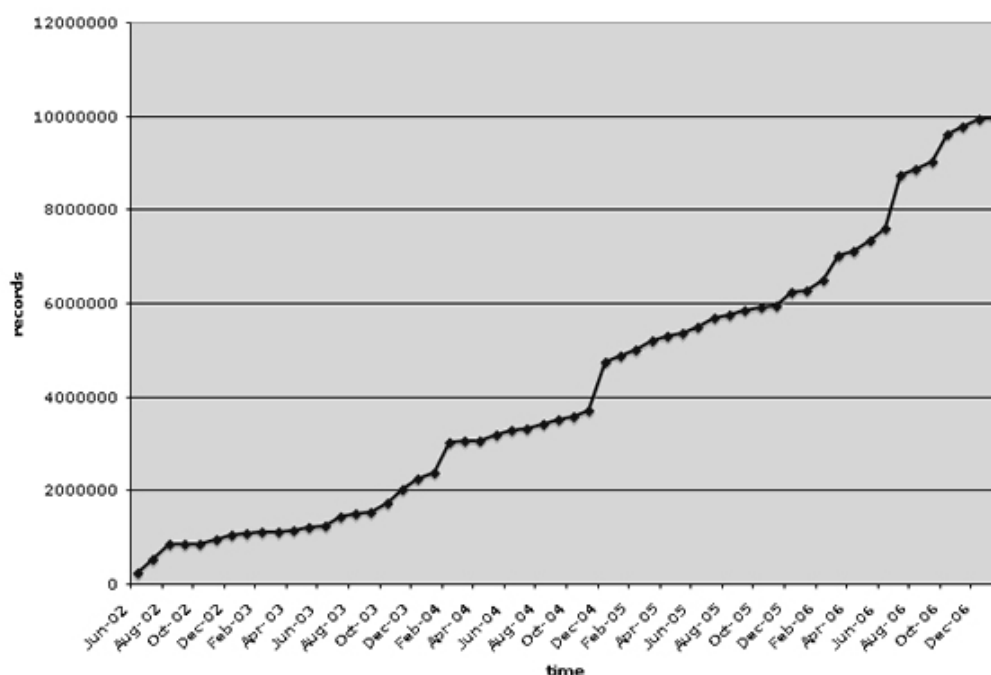
Abstrahując wyliczenia projektów czy ich oceny można się pokusić o kilka dalszych wniosków i spróbować oszacować globalną liczbę bibliotek cyfrowych i udostępnianych obiektów. Wnikliwa analiza repozytoriów zrzeszonych we wszystkich serwisach rozproszonych pozwala na następujące szacunki:

- ok. 500 repozytoriów dużych – ponad 10 tys. poz.

⁸ Źródło: OAISTER : <http://www.oaister.org/stats.html> [2009.04.13]

- ok. 500 repozytoriów średnich – 5-10 tys.
- ok. 2000-4000 repozytoriów małych – 1-5 tys.

Łącznie zatem działa na świecie między 3000 a 5000 repozytoriów (o zbiorach ponad 1000 jednostek każda), które udostępniają 20-30 mln obiektów (w tym ok. 30% w dostępie Open Access); odrębnie należy potraktować zasoby Google Books – ok. 9,7 mln (z czego 817 tys. w pełnym dostępie), co daje ostateczną liczbę - ok. 30-40 mln jednostek.



Rys. 10, Nowe obiekty w serwisie OAISTER (2002-2006) – źródło: OASTER⁹

Zbiory bibliotek cyfrowych rozwijają się jednaka bardzo szybko i stabilnie. Przekonują o tym zarówno statystyki globalne, skąd wynika, że o okresie 2002-2006 następował wzrost o progresie zbliżonym do liniowego (rys. 10), jak i dynamika wzrostu polskich zasobów cyfrowych, które rosły w podobnym tempie (rys. 11).

⁹ <http://www.oaister.org/stats.html> [2009.0413].



Rys. 10, Nowe obiekty w Federacji Bibliotek Cyfrowych (2007-2009)¹⁰

Wstęp do problematyki bibliotek cyfrowych winien jeszcze obejmować wiele innych zagadnień – które jednak z uwagi na ograniczenia objętości publikacji nie może być zaprezentowany w całości. Wśród problemów, na które warto zwrócić uwagę znalazły się: 1. zagadnienia formatów dokumentów, 2. standardów metadanych, 3. funkcjonalności systemów do obsługi bibliotek cyfrowych i 4. zagadnienia prawne.

Problemy 1-3 częściowo rozwijam w prezentacji załączonej do niniejszej książki [slajdy 82-139] (dostępnej też w Śląskiej Bibliotece Cyfrowej¹¹); sprawom formatów poświęciłem referat dostępny w Małopolskiej Bibliotece Cyfrowej¹², zaawansowanym użytkownikom warto polecić książkę G. Płoszajskiego dostępną w BC Politechniki Warszawskiej¹³; zagadnienia metadanych rozwijają kompetentnie publikacje M. Nahotki¹⁴; zaś problematyce prawnej warto polecić liczne publikacje S. Stanisławskiej-Kloc i B. Szczepańskiej¹⁵.

¹⁰ Źródło: <http://fbc.pionier.net.pl/owoc/owoc-stats> [2009.04.13].

¹¹ W. M. Kolasa: Biblioteki cyfrowe - wstęp do problematyki - <http://www.sbc.org.pl/publication/11865> [2009.04.13].

¹² W. M. Kolasa: Formaty hybrydowe w bibliotekach cyfrowych : kontekst i praktyka - <http://mbc.malopolska.pl/publication/14696> [2009.04.13]

¹³ G. Płoszajski: Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego - <http://bcpw.bg.pw.edu.pl/publication/1113> [2009.04.14]

¹⁴ Pełny wykaz pkt 9 - <http://ultra.up.krakow.pl/~wmkolasa/digital/index.html> [2009.04.13]

¹⁵ Pełny wykaz pkt 11 - <http://ultra.up.krakow.pl/~wmkolasa/digital/index.html> [2009.04.13]

Literatura

- PN-ISO 690-2 Informacja i dokumentacja. Zarządzanie dokumentami.
- Encyclopedia of Library and Information Science. Vol. 59, New York 1996, s. 145-148
- D. Grygowski, Dokumenty nieksiążkowe w bibliotece, Warszawa 2001 s. 157-164
- W. M. Kolasa: Biblioteki cyfrowe : wstęp do problematyki - <http://www.sbc.org.pl/publication/11865> [2009.04.13].
- W. M. Kolasa: Formaty hybrydowe w bibliotekach cyfrowych : kontekst i praktyka - <http://mbc.malopolska.pl/publication/14696> [2009.04.13]
- G. Płoszajski: Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego - <http://bcpw.bg.pw.edu.pl/publication/1113> [2009.04.14]