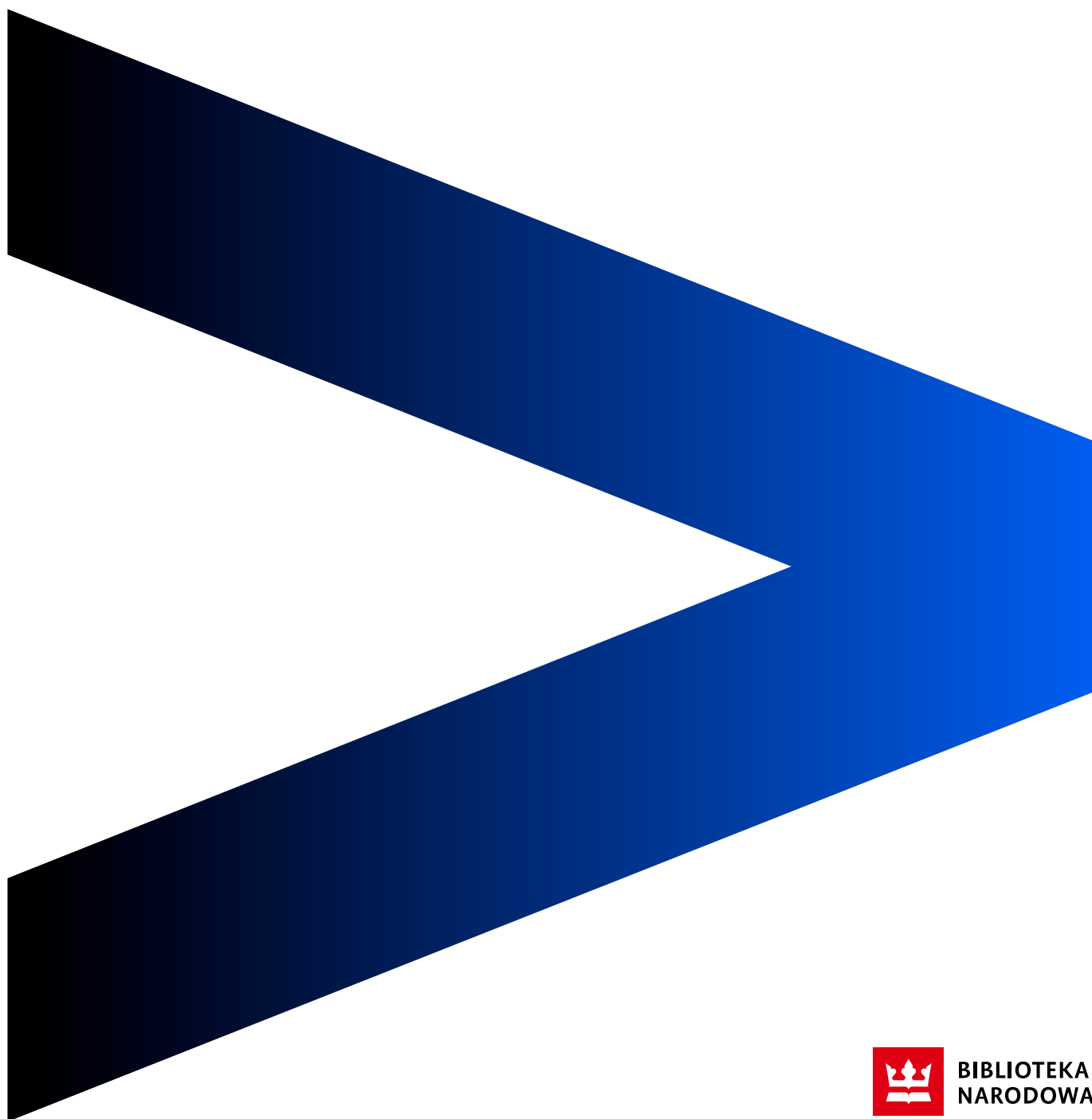


Digitalizacja piśmiennictwa

Aktualizacja



BIBLIOTEKA
NARODOWA

Opracowanie

Tymoteusz Barański
Natan Domel
Tomasz Gruszkowski
Justyna Król-Próba
Paweł Leleń
Dariusz Paradowski
Igor Rosa

Koordynator projektu

Tomasz Gruszkowski

Redakcja i weryfikacja

Małgorzata Kozłowska

Projekt graficzny i typograficzny

Pracownia Agata Muszalska

Publikację złożono krojami

Skolar Latin
Metronic Slab Narrow

ISBN 978-83-8259-498-0

Copyright © Biblioteka
Narodowa, Warszawa 2022

Licencja: Creative Commons
CC BY 4.0

Biblioteka Narodowa
Al. Niepodległości 213
02-286 Warszawa

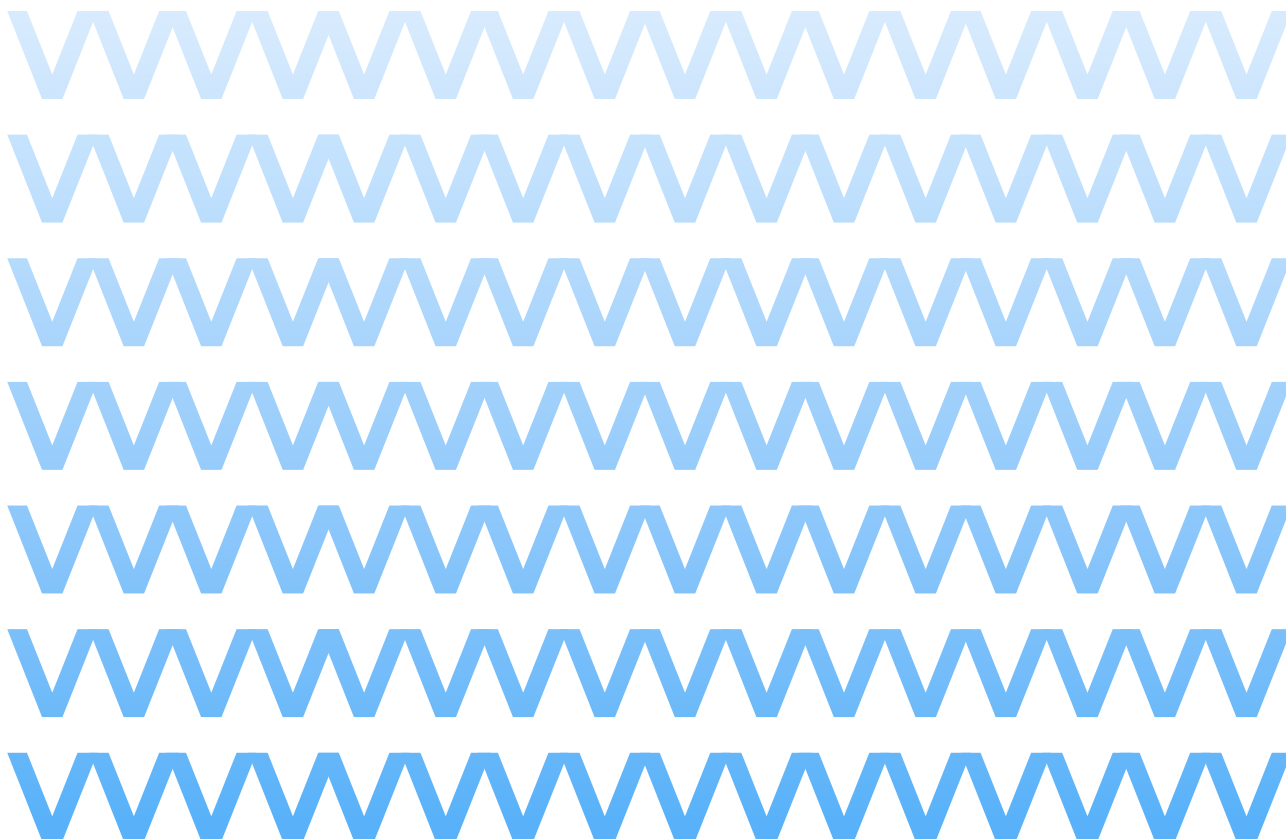
Zadanie

Sfinansowanie działalności
Centrum Kompetencji
w zakresie digitalizacji
materiałów bibliotecznych
w 2022 roku

Dofinansowano ze środków
Ministra Kultury i Dziedzictwa
Narodowego



Ministerstwo Kultury
i Dziedzictwa Narodowego



Spis treści

- 5 **Wprowadzenie**
- 7 **Zagadnienia strategiczne**
- 7 Zdefiniowanie celów digitalizacji
- 7 Określenie oczekiwanych wyników i terminów projektu
- 7 Ograniczenia i konieczności
- 8 Współpraca, otoczenie
- 8 Specyfika
- 8 Analiza strategiczna i zarządzanie projektem
- 9 Finansowanie
- 10 **Prawnoautorskie aspekty digitalizacji piśmiennictwa**
- 15 **Model danych**
- 15 Wstęp
- 15 Opis obiektu fizycznego a metadane obiektu cyfrowego
- 17 Standardy tworzenia opisów obiektów analogowych i cyfrowych
- 17 Dublin Core
- 18 EAD – *Encoded Archival Description*
- 18 METS – *Metadata Encoding and Transmission Standard*
- 19 ISBD i RDA
- 21 ISAD(G) – *General International Standard Archival Description*
- 22 Modele organizacji danych bibliograficznych
- 27 Format MARC
- 29 Języki informacyjno-wyszukiwawcze
- 31 Deskryptory Biblioteki Narodowej (DBN)
- 31 Wymiana danych
- 31 Zasady FAIR
- 33 Protokoły wymiany danych
- 34 Format RDF
- 35 Ontologie
- 36 Linked Open Data
- 37 Selekcja metadanych
- 41 Podsumowanie
- 42 **Transport między magazynem a pracownią digitalizacji**
- 42 Kwestie konserwatorskie
- 43 Aspekty prawne
- 43 Ubezpieczenie
- 44 Zabezpieczenie mienia
- 44 Środki transportu
- 44 Opakowania i skrzynie
- 45 Etapy transportu
- 45 Pobieranie obiektów z magazynu
- 45 Przygotowanie kwestii formalnych i akceptacja
- 45 Protokół
- 46 Pakowanie
- 46 Transport wewnątrz budynku
- 47 Transport na zewnątrz
- 47 Odbiór w miejscu digitalizacji
- 47 Droga powrotna
- 49 **Pomieszczenia**
- 49 Bezpieczeństwo zbiorów
- 50 Pomieszczenia pracowni digitalizacji

- 50 Wnętrza
- 50 Warunki klimatyczne: temperatura i wilgotność względna powietrza
- 51 Zanieczyszczenia powietrza
- 51 Oświetlenie. Szkodliwość dla zbiorów
- 52 Zakłócenia skanowania
- 53 Ergonomia pracy
- 54 **Działania konserwatorskie**
- 54 Prace konserwatorskie przed i w czasie digitalizacji
- 54 Doradztwo podczas planowania działań operacyjnych i na pozostałych etapach projektu
- 55 Ocena stanu zachowania i budowy technologicznej zbiorów, formułowanie zaleceń dotyczących konserwatorskiej ochrony zbiorów w trakcie projektu
- 55 Prowadzenie szkoleń dla reprografów, fotografów, skanerzystów i osób zajmujących się transportem zbiorów
- 55 Wykonywanie niezbędnych zabiegów konserwatorskich
- 56 Asystowanie podczas skanowania
- 56 Przekazywanie do skanowania
- 57 **Techniczne aspekty digitalizacji**
- 57 Sprzęt
- 58 Format
- 58 Rozdzielczość optyczna (częstotliwość próbkowania)
- 60 Głębia bitowa (głębia kolorów)
- 60 Wierność barwna
- 61 Gęstość optyczna
- 62 Profile kolorów
- 62 Szybkość i wydajność
- 62 Niezawodność
- 63 Inne cechy urządzeń digitalizujących
- 63 Oprogramowanie
- 64 Zalecenia
- 68 **Metody zapisu i przechowywania danych**
- 68 Cele cechy
- 69 Rozeznanie potrzeb
- 70 Infrastruktura
- 70 Przesyłanie
- 71 Nośniki i systemy zapisu
- 73 Zarządzanie zawartością
- 74 Bezpieczeństwo zapisanych danych
- 75 Długotrwałe przechowywanie
- 79 **Udostępnianie**
- 79 Dlaczego udostępniamy
- 79 Co udostępniamy
- 80 Komu udostępniamy
- 81 Jak udostępniamy
- 81 Wyszukiwarka
- 83 Prezentacja obiektu
- 84 Dostępność
- 85 Odnajdywalność
- 86 Dodatkowe funkcje biblioteki cyfrowej
- 86 Dostęp poprzez API
- 88 **Modelowy workflow procesu digitalizacji**

Wprowadzenie

Tomasz Gruszkowski

Konieczność rozwijania prac nad procesami digitalizacji materiałów bibliotecznych i przygotowania nowego opracowania aktualizującego te zagadnienia została zasugerowana już w roku 2010 przez autorów pierwszej edycji *Digitalizacji piśmiennictwa*. Rozwój wiedzy teoretycznej, obserwacja postępu technologicznego i doświadczenie w czasie realizacji projektów digitalizacyjnych nakłoniły autorów niniejszej publikacji – pracowników Biblioteki Narodowej oraz osoby współpracujące z Centrum Kompetencji w zakresie digitalizacji materiałów bibliotecznych w BN – do powrotu do tej tematyki.

Autorzy wyrażają nadzieję, że informacje zawarte w tym opracowaniu, aktualizującym niektóre z rozdziałów poprzedniego wydania, wskażą osobom zaczynającym działalność digitalizacyjną i planującym realizację projektów tego typu podstawę do skutecznego, racjonalnego działania.

Projekty digitalizacyjne podlegają ogólnym zasadom planowania i ich sukces zależy w istotnym stopniu od odpowiedniego przygotowania koncepcji i planu działania. W czasie opracowywania założeń planowanego przedsięwzięcia należy brać pod uwagę potencjał instytucji – zarówno pod względem materialnym, jak i ludzkim, możliwość pozyskania finansowania zewnętrznego, a także ewentualną współpracę innych instytucji.

W zależności od tych czynników oraz od skali projektu (zarówno pod względem zakresu czasowego realizacji przyjętego planu, jak i liczby przewidzianych do digitalizacji obiektów) można podjąć decyzję o przeprowadzeniu etapu planowania siłami własnymi lub za pośrednictwem specjalnie zatrudnionych specjalistów od zarządzania projektami.

Przedstawione w tym opracowaniu zagadnienia mają na celu ułatwienie podjęcia takich decyzji i zaplanowanie projektu w sposób, który pozwoli doprowadzić go do udanego finału. Należy pamiętać, że planując trzeba brać pod uwagę wszystkie istotne kwestie, wśród nich możliwość pojawienia się najrozmaitszych zdarzeń mogących powodować konieczność adaptacji planów do rzeczywistości.

Prezentowana publikacja obejmuje tylko część z zagadnień przedstawionych w *Digitalizacji piśmiennictwa*. Uznano bowiem, że znaczna część z zawartych w nim rozważań o charakterze teoretycznym jest aktualna, a nie jest wskazane szersze omawianie rozwiązań technologicznych – rozwój w tym zakresie jest szczególnie intensywny, a dzisiejsze zalecenia mogą zdezaktualizować się nadszpodziewanie szybko. Ponadto, wiele z rozważań teoretycznych, które

wydawały się nowatorskie przed dwunastoma laty, są w dzisiejszej praktyce oczywistościami. Wszystkich zainteresowanych digitalizacją nie tylko w wąskim, praktycznym zakresie, zachęcamy zatem do sięgnięcia do wieloaspektowego opracowania pod redakcją Dariusza Paradowskiego¹.

Zagadnienia dotyczące szeroko rozumianej digitalizacji zostaną w niniejszej pracy przedstawione w sposób ułatwiający stworzenie listy tematów, które należy wziąć pod uwagę w czasie pracy nad przygotowaniem i realizacją przyjętego projektu. Nie zostaną tu omówione kwestie związane z formalnymi i finansowymi aspektami projektów digitalizacyjnych – wykraczają one poza obszar zadań wyznaczonych Centrum Kompetencji w zakresie digitalizacji zbiorów bibliotecznych w Bibliotece Narodowej, które odpowiada za publikację.

Łącznie zakres obu opracowań obejmuje – w ocenie autorów – wszystkie zagadnienia istotne dla praktycznej realizacji projektu digitalizacyjnego, nawet w sytuacji, kiedy wychodzą one poza ramy czysto racjonalnie i schematycznie pojmowanych zadań czy kompetencji instytucji przygotowującej projekt. Digitalizacja jest na tyle młodą dziedziną działalności, że osoby zaangażowane w projekty digitalizacyjne muszą wciąż nabywać nowych umiejętności – poprzez lekturę, przez uczenie się i weryfikowanie błędów, poprzez udział w szkoleniach i warsztatach prowadzonych przez bardziej doświadczonych koleżanki i kolegów, a także korzystanie z zasobów internetowych takich jak materiały dostępne na stronie Centrum Kompetencji BN (<https://bn.org.pl/uslugi/centrum-kompetencji/edukacja>).

Zapraszamy do kontaktu: potrzeby i sugestie można przekazywać pod adresem e-mailowym – centrum.kompetencji@bn.org.pl.



1 *Digitalizacja piśmiennictwa*; oprac. i red. Dariusz Paradowski. Warszawa: Biblioteka Narodowa, 2010. Publikacja dostępna jest w księgarni Biblioteki Narodowej przy al. Niepodległości 213 w Warszawie oraz na stronie <https://ksiegarnia.bn.org.pl/201/Digitalizacja-pismienictwa.html>. Wersja elektroniczna – dostępna pod adresem: <https://www.bn.org.pl/uslugi/centrum-kompetencji/digitalizacja/digitalizacja-pismienictwa>.

Zagadnienia strategiczne

Tomasz Gruszkowski

Zdefiniowanie celów digitalizacji

Każda instytucja dziedzictwa kulturowego ma wyznaczone cele funkcjonowania – określają one jej codzienną działalność. Digitalizacja umożliwia lepsze – bardziej skuteczne, bezpieczniejsze, pozwalające na oszczędności finansowe – realizowanie niektórych z tych celów. Określenie potrzeb, których zaspokojenie będzie możliwe dzięki digitalizacji pozwala na precyzyjne zdefiniowanie celów digitalizacji, jej przewidywanego zakresu i czasu niezbędnego do realizacji przedsięwzięcia. Rozdział *Cele digitalizacji*, zawarty w publikacji *Digitalizacja piśmiennictwa*, omawia szerzej temat możliwych celów i konsekwencji wynikających z przyjęcia określonych założeń.

Określenie oczekiwanych wyników i terminów projektu

Po przyjęciu celów należy oszacować zakres ilościowy digitalizacji (liczbę i charakter obiektów) oraz określić jej wymaganą jakość (parametry techniczne odwzorowania). Niezależnie od tego, czy projekt ma mieć charakter jednorazowy, cykliczny czy ciągły, niezbędne jest zaplanowanie jego ram czasowych z jednoczesnym uwzględnieniem (posiadanych lub planowanych) zasobów instytucji. Przygotowany harmonogram winien być weryfikowany w czasie trwania projektu przy pomocy konsekwentnie prowadzonej sprawozdawczości, która umożliwi poprawne rozliczenie przedsięwzięcia (realizowanego na potrzeby wewnętrzne instytucji lub zewnętrzne w przypadku projektów finansowanych ze źródeł zewnętrznych).

Ograniczenia i konieczności

Zarówno na etapie przygotowania projektu, jak i jego realizacji konieczne jest wzięcie pod uwagę istnienia zarówno oczywistych ograniczeń (do których mogą należeć: liczba i cha-

rakter obiektów dostępnych do digitalizacji w zbiorach instytucji, ich forma fizyczna i stan zachowania, dostępne środki finansowe, posiadane urządzenia digitalizacyjne, dostępne pomieszczenia, zasoby ludzkie, ograniczenia powodowane koniecznością przestrzegania zapisów prawa autorskiego), jak i potencjalnych ryzyk i zagrożeń (np. takich, jakie pojawiają się w ostatnich latach – pandemia, zagrożenia wojenne, wzrost kosztów energii, inflacja). Wnikliwa analiza i wczesna ocena ryzyka ułatwią przygotowanie rozwiązań służących jego ograniczeniu.

Rozpoczęcie realizowania projektu (a dokładnie podpisanie umowy na jego realizację) – zwłaszcza związanego z przyjęciem finansowania z dotacji – powoduje konieczność wykonania zaplanowanych celów ilościowych i rozliczenia projektu, które może mieć wysoce sformalizowaną formę.

Współpraca, otoczenie

Przygotowując i planując projekt, warto zwrócić uwagę na możliwość współpracy z innymi instytucjami, dysponującymi zasobami czy wiedzą, których brakuje w instytucji macierzystej. Ponadto analiza otoczenia i wcześniej przeprowadzonych projektów może skierować uwagę na potencjalne ograniczenia, a także zagrożenia i przyczynić się do lepszego zaplanowania przedsięwzięcia.

Specyfika

Analiza innych projektów digitalizacyjnych może pozwolić na lepsze planowanie, należy jednak pamiętać, aby nie tracić z pola widzenia specyfiki własnej instytucji i jej zasobu. Te aspekty są kluczowe dla określenia konkretnych wymagań projektowych, dla zaplanowania sposobu wewnętrznej komunikacji i współpracy na potrzeby projektu. Istotne jest doprowadzenie do zapewnienia skutecznego gromadzenia i przekazywania informacji niezbędnej do prowadzenia projektu.

Analiza strategiczna i zarządzanie projektem

W przypadku planowania projektów na większą skalę, zwłaszcza przy założeniu pozyskania finansowania zewnętrznego, niezbędne będzie wykorzystanie narzędzi planowania strategicznego. Sformalizowanie podejścia do strategii może przynieść korzyści także w przypadku projektów digitalizacyjnych, w celu uporządkowania i skutecznego wykorzystania zgromadzonych informacji.

Istnieje wiele metod planowania strategicznego i trudno wskazać taką, która odpowiadałaby najlepiej specyfice digitalizacji. Wybór metody i sposób zaadoptowania wybranej do danych warunków warto pozostawić specjalistom.

Realizacja bardzo dużych projektów, szczególnie związanych z rygorystycznymi zasadami ich rozliczania, może wymagać kompetencji wykraczających poza dotychczasowe umiejętności pracowników, a nawet wymuszać zmiany dostosowujące strukturę organizacyjną instytucji do przyjętych założeń. Wiąże się to z ewentualnością nabycia nowych kompetencji przez obecnych pracowników lub zatrudnienia specjalistów w zakresie zarządzania projektami.

Finansowanie

Niezależnie od potencjalnego źródła finansowania projektu (ze środków własnych instytucji, z dotacji na realizację programów operacyjnych ministerstw czy innych instytucji państwowych, dofinansowania z funduszy europejskich) niezbędne jest precyzyjne określenie zakresu projektu, związane z możliwie dokładną analizą kosztów poszczególnych jego elementów i etapów. Dokonując wyliczeń, powinno się brać pod uwagę założenia dotyczące szczegółowych planów działania.

Przy tworzeniu kosztorysu należy uwzględnić zarówno etap realizacji przedsięwzięcia, jak i potrzebę (czasem konieczność) utrzymania projektu po zakończeniu digitalizacji.



Prawnoautorskie aspekty digitalizacji piśmiennictwa

Tymoteusz Barański

Procesy digitalizacji piśmiennictwa nie odbywają się w próżni prawnej. Co więcej, uwarunkowania prawne stanowią jeden z zasadniczych – oprócz technicznego – aspektów determinujących przedmiot i zakres masowych projektów digitalizacyjnych. Głównym, choć niejedynym czynnikiem prawnym, który wpływa na możliwość digitalizacji materiałów piśmiennych, w tym bibliotecznych, jest szeroka i restrykcyjna ochrona praw autorskich. Digitalizacja legalna musi zatem przebiegać z uwzględnieniem powszechnie obowiązujących w tym zakresie unormowań.

Ochrona prawa autorskiego ma charakter niemal powszechny. W zasadzie nie istnieją państwa, w których instytucja ta byłaby całkowicie nieobecna. Jest ona również w zasadniczych zrębach zharmonizowana niemal na całym świecie dzięki systemowi obowiązujących w tym zakresie umów międzynarodowych, zaś na poziomie Unii Europejskiej dodatkowo w szeregu dyrektyw. Polskie unormowania w obszarze prawa autorskiego datują się już od 1926 roku, kiedy to pod wpływem inicjatywy środowisk twórczych uchwalono pierwszą ustawę autorską, następnie zastąpioną ustawą z 1952 roku, zaś jeszcze później – obecnie obowiązującą i wielokrotnie nowelizowaną ustawą z roku 1994¹.

Prawo autorskie zalicza się do własności intelektualnej. Chroni ono dobro niematerialne, jakim jest utwór, który ustawodawca określa jako przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiegokolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia (art. 1 ust. 1 ustawy autorskiej). Utworem będzie więc wytwór ludzkiej myśli, charakteryzujący się dwoma przymiotami: indywidualności oraz twórczości. Przesłanki te mają charakter nieostry, zaś ich rozumienie wywołuje niekończące się spory w nauce prawa autorskiego. Równie niełatwym zadaniem bywa nierzadko rozstrzygnięcie, czy dany przejaw myśli spełnia te kryteria i w konsekwencji powinien być kwalifikowany jako utwór. W dużym uproszczeniu można powiedzieć, że cecha twórczości, jaką musi się cechować utwór, polega na jego oryginalności, a więc odróżnialności od tych przejawów ludzkiej myśli, które zostały już

1 Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz.U. z 2021 r. poz. 1062 z późn. zm.), dalej: „ustawa autorska”.

w przeszłości ustalone, czyli wyrażone w sposób dostrzegalny dla otoczenia. W rozstrzygnięciu wątpliwości może pomóc zabieg myślowy, polegający na zbudowaniu definicji negatywnej kryterium twórczości: twórczym nie będzie to, co jest odtwórcze. Dlatego właśnie reprodukcje i kopie nie stanowią samodzielnych utworów. Z kolei o indywidualności utworu można mówić wówczas, gdy da się racjonalnie założyć, że niemożliwe jest niezależne stworzenie przez inną osobę utworu identycznego (twórczość paralelna). Stąd też spod zakresu pojęcia utworu wyłącza się fotografie paszportowe czy legitymacyjne. Wymogi formalne, które musi spełnić taka fotografia, sprawiają, że wykonana przez dwóch różnych fotografów może wyglądać identycznie, zatem jej wykonawca nie może odcisnąć na niej piętna swojej indywidualności.

Utwór to wprawdzie niematerialny przejaw myśli ludzkiej, jednak musi być zakomunikowany otoczeniu. Ustawodawca wprost zastrzega, że ochronie podlega jedynie określony sposób wyrażenia tej myśli (art. 1 ust. 21 ustawy autorskiej). Ochroną prawa autorskiego nie są zatem objęte odkrycia, idee, procedury, metody i zasady działania oraz koncepcje matematyczne. Dla bytu utworu niezbędne jest jego ustalenie, które nie musi przybierać formy trwałej; będzie nim np. wykonanie piosenki, wyrecytowanie wiersza itp. Jeśli jednak utwór zostaje ustalony w jakimś materialnym nośniku, np. piosenka nagrana, czy wiersz zapisany, to wówczas mówimy o utrwaleniu utworu, zaś rzecz, za pomocą której do takiego utrwalenia doszło nazywamy nośnikiem utworu.

Model prawa autorskiego przyjęty przez polskiego ustawodawcę pod wpływem rozwiązań francuskich w okresie przedwojennym i konsekwentnie utrzymywany ma charakter dualistyczny. Oznacza to, że prawo autorskie chroni zarówno majątkowe interesy twórcy, których normatywnym wyrazem są autorskie prawa majątkowe, jak i niewygasalną więź niematerialną, łączącą twórcę i jego dzieło; więź ta jest przedmiotem autorskich praw osobistych, do których należą między innymi: prawo do autorstwa utworu, do oznaczenia utworu nazwiskiem lub pseudonimem albo do udostępniania go anonimowo, do nienaruszalności treści i formy utworu oraz jego rzetelnego wykorzystania, do decydowania o pierwszym udostępnieniu utworu publiczności oraz do nadzoru nad sposobem korzystania z utworu (art. 16 ustawy autorskiej). Istota autorskich praw majątkowych zbliża je do tradycyjnej własności, której przedmiotem są rzeczy. Oba te prawa mają charakter bezwzględny, skuteczny wobec wszystkich. Podobnie jak właściciel rzeczy ma wyłączne w zasadzie prawo do korzystania z niej i rozporządzania nią, twórcy przysługuje wyłączne prawo do korzystania z utworu i rozporządzania nim na wszystkich polach eksploatacji oraz do wynagrodzenia za korzystanie z utworu (art. 17 ustawy autorskiej).

Zasadniczym problemem, przed którym stoi digitalizacja piśmiennictwa jest zatem kwestia, która wynika z charakteru prawnego podlegających jej materiałów jako nośników utworów. Skoro to twórca ma wyłączne prawo do korzystania z utworu, zaś zwielokrotnienie utworu, także w formie cyfrowej stanowi postać takiej eksploatacji (art. 50 pkt 1 ustawy autorskiej), to czynność taka jest w zasadzie dopuszczalna tylko za zgodą twórcy albo podmiotu, na który przeszły autorskie prawa majątkowe; są one bowiem zbywalne i dziedziczne (art. 41 ust. 1 pkt 1 i 2 ustawy autorskiej). Również publiczne udostępnienie zdigitalizowanego materiału będącego nośnikiem utworu w sieciach cyfrowych jest postacią jego eksploatacji (art. 50 pkt 3 ustawy autorskiej), a zatem i takie działanie stanowi wkroczenie w monopol autorski i musi mieć wyraźną podstawę prawną.

W zasadzie zatem swobodnej digitalizacji – którą tu należy rozumieć jako przekształcenie danego materiału w postać cyfrową i publiczne udostępnienie w sieciach cyfrowych – mogą podlegać tylko te materiały piśmienne, które nie stanowią nośników chronionych utworów. Materiały te określa się jako należące do domeny publicznej. Zalicza się do nich przede wszystkim takie, których zawartość pod względem sposobu wyrażenia nie spełnia w ogóle kryteriów ogólnych utworu, a więc nieindywidualne i nietwórcze. Dalej trzeba mieć na względzie, że ustawodawca zdecydował się na wyłączenie określonych kategorii materiałów spod zakresu ochrony prawnoautorskiej. Należą do nich akty normatywne lub ich urzędowe projekty, urzędowe dokumenty, materiały, znaki i symbole, opublikowane opisy patentowe lub ochronne oraz proste informacje prasowe (art. 4 pkt 1–4 ustawy autorskiej).

Autorskie prawa majątkowe mają charakter terminowy. Oznacza to, że wygasają z upływem określonego terminu, którym co do zasady jest siedemdziesiąt lat od śmierci twórcy (art. 36 pkt 1 ustawy autorskiej). Zasadniczym problemem praktycznym bywa tu możliwość ustalenia dat życia twórcy, co w odniesieniu do autorów mało znanych może nastroić istotne trudności. Dla legalnej digitalizacji materiału stanowiącego nośnik takiego utworu konieczne jest bowiem jednoznaczne stwierdzenie, w którym roku doszło do wygaśnięcia autorskich praw majątkowych. Samo podjęcie w tym zakresie starań w dobrej wierze, jak również poczynienie błędnych ustaleń nie zwalnia od odpowiedzialności za naruszenie autorskich praw majątkowych. Wyjątkiem jest instytucja utworu osieroconego, która umożliwia instytucjom dziedzictwa kulturowego digitalizację materiałów z własnych zbiorów po przeprowadzeniu sformalizowanej procedury tzw. starannego poszukiwania oraz umieszczeniu utworu w bazie utworów osieroconych (art. 355 i n. ustawy autorskiej). Stopień skomplikowania powyższej procedury oraz jej kosztochłonność sprawia, że nie znajduje ona szerszego zastosowania w procesach masowej digitalizacji.

Oprócz materiałów, których zawartość nie spełnia kryteriów ogólnych utworu, materiałów będących nośnikami wytworów wyraźnie wyłączonych spod zakresu ochrony oraz stanowiących nośniki utworów, do których autorskie prawa majątkowe już wygasły, do domeny publicznej zalicza się także materiały obejmujące utwory powstałe przed powstaniem instytucji praw autorskich, jednak i ta kategoria nie jest wolna od kontrowersji. Po pierwsze rozstrzygnięcie, od kiedy można mówić o obowiązywaniu w tym zakresie powszechnie obowiązujących przepisów prawa, choćby na ziemiach polskich, nie jest oczywiste, zaś po drugie istnieje problem ochrony autorskich praw osobistych do utworów dowolnie dawnych, który ma istotny wpływ na możliwość digitalizacji materiałów rękopiśmiennych. O problemie tym będzie jeszcze skrótowo mowa.

Materiały nienależące do domeny publicznej, a więc podlegające ochronie prawnoautorskiej, mogą być digitalizowane jedynie na wyraźnej podstawie prawnej. Może być nią nabycie autorskich praw majątkowych przez podmiot dokonujący digitalizacji albo pozyskanie licencji od uprawnionego podmiotu. Jednakże w procesach masowej digitalizacji czasochłonność, a także kosztochłonność pozyskiwania takich uprawnień sprawia, że jedynie niektóre podmioty mogą sobie pozwolić na korzystanie z takiej formy legalnej digitalizacji piśmiennictwa.

Choć przedmiotem swobodnej digitalizacji mogą być materiały z domeny publicznej, ustawodawca umożliwia niektórym podmiotom, zaliczanym do instytucji dziedzictwa kul-

turowego (m.in. placówkom oświatowym, uczelniom, bibliotekom, szkołom i jednostkom prowadzącym działalność archiwalną), na zwielokrotnianie utworów znajdujących się w zbiorach własnych oraz ich udostępnianie w wewnętrznych sieciach takich jednostek (art. 28 ust. 1 pkt 2 i 3 ustawy autorskiej). Jest to jedna z form tzw. dozwolonego użytku publicznego chronionych utworów. Przykład procesu digitalizacyjnego, w którym wykorzystywana jest ta licencja ustawowa, stanowi cyfrowa wypożyczalnia międzybiblioteczna książek i czasopism naukowych „Academica” – administrowana przez Bibliotekę Narodową.

Istotną barierą prawną w procesach digitalizacji bywa to, że ten sam materiał, rozważany jako kandydat do digitalizacji, może mieć niejednorodną strukturę prawnoautorską. Może bowiem zawierać szereg różnych elementów, z których pewne w ogóle nie stanowią utworów, inne stanowią utwory objęte wygasłymi prawami majątkowymi, zaś jeszcze inne – utwory, których okres ochrony wciąż trwa. Przykładem może być druk zwarty (książka), zawierający okładkę stanowiącą utwór graficzny, główny utwór słowny oraz utwory towarzyszące (wstęp, posłowie). Każdy z tych elementów może mieć odmienny status prawnoautorski. Tymczasem, aby móc zwielokrotnić i rozpowszechnić całość takiego materiału konieczne jest dysponowanie stosowną podstawą prawną w stosunku do wszystkich jego elementów. Przy jej braku podmiot dokonujący digitalizacji staje przed dylematem, czy dokonać fragmentacji materiału i zdigitalizować tylko te jego elementy, co do których jest to prawnie dopuszczalne, czy też zaniechać digitalizacji w ogóle.

Konieczność przestrzegania autorskich praw majątkowych przysługujących twórcy bądź jego następcom prawnym nie wyczerpuje całokształtu potencjalnych barier prawnoautorskich procesów digitalizacji piśmiennictwa. Oprócz praw majątkowych ustawodawca chroni bowiem wskazane wyżej autorskie prawa osobiste, zaś wśród nich – prawo do decydowania o pierwszym udostępnieniu utworu publiczności. Prawa te mają charakter niewygasalny, co oznacza, że utrzymują się także po śmierci twórcy oraz po upływie siedemdziesięcioletniego terminu ochrony praw majątkowych. Zatem wygaśnięcie tych ostatnich w odniesieniu do utworu nierozpowszechnionego nie będzie tożsame z jego przejściem do domeny publicznej, zaś digitalizacja ucieleśniającego go materiału będzie wymagała zgody określonych w ustawie członków rodziny twórcy (art. 78 ust. 3 w zw. z ust. 2 ustawy autorskiej). W praktyce digitalizacji problem ten manifestuje się w odniesieniu do materiałów rękopiśmiennych obejmujących utwory nierozpowszechnione. Pomimo wygaśnięcia autorskich praw majątkowych, brak zgody twórcy na publikację w powiązaniu z brakiem możliwości zidentyfikowania osób uprawnionych do udzielenia takiej zgody, sprawia, że digitalizacja takich materiałów rodzi ryzyko naruszenia autorskich praw osobistych. Co więcej, sposób ukształtowania ochrony tych praw, w powiązaniu z tzw. normami intertemporalnymi, a więc regulującymi stosunki prawne powstałe pod rządami dawnego prawa, sprawia, że mówi się o ochronie utworów dowolnie dawnych. Teoretycznie zatem nie istnieje cezurą czasową umożliwiającą – odmiennie niż w odniesieniu do praw majątkowych – przyjęcie, że utwory powstałe przed momentem, który ją wyznacza, nie są chronione prawami osobistymi. Oznacza to – dalej, teoretycznie – że gdyby obecnie odnaleziono niepublikowane rękopisy obejmujące utwory twórców zmarłych przed wiekami, to legalna digitalizacja wymagałaby zgody ustawowo określonych członków ich rodzin. Pogląd taki był wyrażany w nauce prawa np. w odniesieniu do twórczości C. K. Norwida.

Zarysowane tu problemy nie wyczerpują całokształtu zagadnień prawnych związanych z digitalizacją piśmiennictwa. W każdym razie procesy masowej digitalizacji piśmiennictwa zawsze wiążą się z procesami ewaluacji prawnej materiałów, która musi poprzedzać ich kwalifikację do zwielokrotnienia i publicznego udostępnienia w postaci cyfrowej. Ewaluacja ta nie zawsze prowadzi do jednoznacznych konkluzji, co sprawia, że podmiot digitalizujący musi nierzadko samodzielnie oceniać ryzyko prawne związane z danym materiałem. Z uwagi na absolutny niemal charakter odpowiedzialności za bezprawne naruszenie autorskich praw majątkowych i osobistych – ocena ta powinna być wnikliwa i ostrożna.



Model danych

Natan Domel, Paweł Lelęć

Wstęp

Zdigitalizowany obiekt fizyczny nie może spełniać swojego zadania bez odpowiednio dobranych i prawidłowo przypisanych metadanych, dlatego coraz częściej w kontekście digitalizacji dostrzega się nie tylko potrzebę zapewnienia właściwego miejsca archiwizacji dla obiektów cyfrowych, ale też konieczność przechowywania odpowiednich informacji o nich. Nie ulega wątpliwości, że zanim obiekt zostanie zdigitalizowany, powinien posiadać opis, choćby uproszczony. Praktyką nie do przyjęcia, powodującą błędy w wyszukiwaniu i utrudniającą użytkownikom korzystanie z bibliotek cyfrowych, jest upublicznianie obiektów z niejednoznacznymi, niepoprawnymi bądź zdawkowymi opisami. Oczywiście sporządzenie pełnego opisu obiektu bywa czasochłonne, więc gdyby oczekiwać pełnych zestawów metadanych wszystkich obiektów jeszcze zanim zostaną zdigitalizowane, mogłoby to mocno spowolnić proces digitalizacji. Dlatego optymalne wydaje się rozwiązanie, w którym przed wykonaniem kopii cyfrowych każdy obiekt posiada przynajmniej skrócony opis. Metadane powinny pozwolić na jednoznaczną identyfikację obiektu i ograniczyć możliwość potencjalnej dezinformacji. Obiekty fizyczne i cyfrowe bez prawidłowych opisujących je metadanych mogą stać się praktycznie bezużyteczne, ponieważ w przypadku instytucji, które posiadają bardzo liczne zbiory, obiekty staną się niezwykle trudne do odnalezienia. W rozdziale skupiono się na omówieniu metadanych opisujących obiekty cyfrowe i analogowe w kontekście ich digitalizacji oraz standardów, które ustalają zasady ich powstawania.

Opis obiektu fizycznego a metadane obiektu cyfrowego

Często mianem „opisu” określa się dane o obiekcie analogowym, natomiast w kontekście informacji o obiekcie cyfrowym traktowanym jako dane komputerowe mówi się o „metadanych”. Istnieje pewien rozdźwięk między uproszczoną definicją metadanych określającą je po prostu jako „dane o danych” a tą szerszą, według której są to informacje, które opisują, lokaliz-

zują, pomagają w przeszukiwaniu zbioru informacji. Dla uproszczenia można jednak przyjąć, że termin „dane obiektu” odnosi się zarówno do obiektów analogowych, jak i cyfrowych¹.

W każdej sytuacji powinno być możliwe wskazanie obiektu fizycznego, któremu odpowiada obiekt cyfrowy. Dane, które zostały zawarte w opisie dla obiektu fizycznego są przejmowane dla obiektu cyfrowego w praktycznie niezmienionej formie. Dodatkowo metadane obiektu cyfrowego są uzupełniane o dane charakterystyczne dla tego typu obiektu, m.in. format czy rozmiar pliku.

Opis obiektu analogowego można podzielić na dwie zasadnicze części²:

- opis formalny, zawierający informacje odnoszące się do cech fizycznych obiektu i informacji identyfikacyjnych, np. tytuł, autora/autorów, adres wydawniczy, objętość,
- opis przedmiotowy, zawierający informacje dotyczące treści wyrażone np. deskryptorami lub symbolami Uniwersalnej Klasyfikacji Dziesiętnej.

Wśród metadanych obiektów cyfrowych można wskazać trzy podstawowe rodzaje:

- metadane opisowe, takie jak tytuł, autor i słowa kluczowe, które opisują obiekt cyfrowy oraz umożliwiają jego identyfikację i wyszukiwanie,
- metadane strukturalne, odnoszące się do relacji między obiektami i ich elementami,
- metadane administracyjne, które dostarczają informacji pomagających w zarządzaniu obiektem, np. informacji technicznych.

W celu umożliwienia sprawnego odnalezienia właściwych obiektów cyfrowych muszą do nich być przypisane odpowiednie metadane, które dostarczą wystarczającą liczbę punktów dostępu oraz będą w sposób jednoznaczny identyfikowały dany obiekt. Uwzględniając ogromną liczbę zdigitalizowanych obiektów, która przyrasta w bardzo dużym tempie³, odnalezienie zdigitalizowanego obiektu bez odpowiedniego pakietu metadanych byłoby praktycznie niemożliwe. Pośród metadanych nie może również zabraknąć sygnatury obiektu fizycznego, który został zdigitalizowany, dzięki czemu możliwe będzie jednoznaczne wskazanie, który obiekt został zdigitalizowany w przypadku, gdy w zbiorach instytucji znajduje się więcej niż jeden egzemplarz⁴. Należy unikać sytuacji, w których wyszukiwanie publikacji jest tak skomplikowane, że wyłącznie pracownicy danej instytucji potrafią prawidłowo używać narzędzi wyszukiwawczych. Znaczący wpływ na oczekiwania użytkowników wywierają powszechnie wykorzystywane wyszukiwarki internetowe, które przyzwyczajają do otrzymywania prawie natychmiast wyników po wpisaniu jedynie słów kluczowych⁵.

1 *Digitalizacja piśmiennictwa*, opracowanie i redakcja Dariusz Paradowski, Warszawa 2010.

2 Tamże.

3 A. Dąbrowski, *Digitalizacja w procesie komercjalizacji dziedzictwa lokalnego*, w: *Rozwój lokalny i strategiczne zarządzanie rozwojem gminy w oparciu o dziedzictwo kulturowe*, Warszawa 2021, s. 100–106.

4 R. Derewenda, *Opracowanie, digitalizacja i udostępnianie fotografii związanych z działalnością ks. Franciszka Blachnickiego w Archiwum Głównym Ruchu Światło-Życie*, „Archiwa Bibliotek i Muzea Kościelne” 2021, t. 117, s. 49–66; M. Salamonowicz, *Digitalizacja utworów w działalności bibliotek publicznych i naukowych*, w: *Własność intelektualna a dziedzictwo kulturowe: księga jubileuszowa dedykowana Profesorowi Wojciechowi Kowalskiemu*, 2020, s. 232–242.

5 T. Gross, A. G. Taylor, D. N. Joudrey, *Still a lot to lose. The role of controlled vocabulary in keyword searching*, „Cataloging & Classification Quarterly” 2015, t. 53, nr 1, s. 1–39.

Standardy tworzenia opisów obiektów analogowych i cyfrowych

W wyniku dążenia do ujednoczenia opisów obiektów analogowych i cyfrowych powstały liczne standardy, których zadaniem jest przede wszystkim dostarczenie zbioru zasad tworzenia opisów. Pośród znanych i stosowanych standardów można wymienić m.in.⁶:

- Dublin Core (DC) lub Dublin Core Metadata Element Set (DCMES),
- Encoded Archival Description (EAD),
- Metadata Encoding and Transmission Standard (METS),
- International Standard Bibliographic Description (ISBD),
- Resource Description and Access (RDA),
- Międzynarodowy Standard Opisu Archiwalnego ISAD(G) (ang. *General International Standard Archival Description*).

Dublin Core

Dublin Core to otwarty, wymienny standard metadanych do opisu zasobów sieciowych. Powstał w 1995 roku z inicjatywy grupy informatyków i bibliotekarzy szukających sposobu na usprawnienie wyszukiwania informacji w Internecie. Opracowano wówczas podstawowy, 15-elementowy zestaw metadanych – DCMES (Dublin Core Metadata Element Set). Powołano również do istnienia podlegającą OCLC (Online Computer Library Center) organizację DCMI (Dublin Core Metadata Initiative), która obecnie odpowiada za rozwój standardu DC⁷.

W Dublin Core można wyróżnić dwa poziomy opisu: *Simple Dublin Core* – prosty, składający się z 15 elementów metadanych (takich jak tytuł, twórca, temat, format czy identyfikator), stanowiących podstawę opisu dokumentu, oraz *Qualified Dublin Core* – złożony, pozwalający uszczegółwić opis, zawierający trzy dodatkowe elementy oraz tzw. kwalifikatory, które zawierają semantykę podstawowych elementów⁸.

Główną zasadą, która przyświecała twórcom DCMES, było dążenie do prostoty, stąd początkowa uproszczona lista jedynie 15 elementów. Wkrótce jednak podjęto decyzję o uszczegóławianiu ich znaczenia przez dołączanie kwalifikatorów. DCMES miał być też zrozumiały i łatwy do wykorzystania – tak, aby każdy twórca zasobów sieciowych, nawet nie przechodząc żadnego specjalistycznego szkolenia ani nie zdobywając fachowej wiedzy bibliotekarskiej, mógł opisać własny dokument⁹. Gramatykę Dublin Core charakteryzuje prostota leksykalna – obiekt można opisać jedynie za pomocą elementów (rzeczowniki) i ich atrybutów (przymiotniki) – oraz semantyczna – wynikiem łączenia terminów jest nieskomplikowany układ twierdzeń¹⁰. Elementy w Dublin Core, zgodnie z założeniami twórców, są opcjonalne i powta-

6 M. Seikel, T. Steele, *Comparison of key entities within bibliographic conceptual models and implementations: definitions, evolution, and relationships*, „Library Resources and Technical Services” 2020, vol. 64, iss. 2, s. 62–71.

7 <https://www.bn.org.pl/dla-bibliotekarzy/normy,-formaty,-standardy/metadane/dublin-core> [dostęp: 14.10.2022].

8 <https://www.dublincore.org/specifications/dublin-core/usageguide/> [dostęp: 19.11.2022].

9 M. Nahotko, *Czym są metadane?*, „Fides – Biuletyn Bibliotek Kościelnych” 2004, nr 1/2, s. 59–67.

10 A. Filipek, *Dublin Core, czyli metadane w nowej formie*, „Zagadnienia Informatyki Naukowej” 2006, nr 2, s. 50–58.

rzalne, a ich kolejność jest dowolna. Są też międzynarodowe i niezależne od żadnej dyscypliny w odniesieniu do zakresu i zastosowania. Zestaw elementów jest rozszerzalny, dzięki czemu może być rozbudowywany w zależności od potrzeb dziedziny zastosowań.

W roku 2007 podjęto się ujednolicenia zapisu elementów metadanych Dublin Core z zaleceniami Konsorcjum 3W zapisanymi w RDF (*Resource Description Framework*). Obecnie istnieją dwa opracowania metadanych Dublin Core:

- **DCMES v. 1.1**¹¹ – podstawowy 15-elementowy zestaw metadanych Dublin Core,
- **DCMI Metadata Terms**¹² – zestaw złożony z DCMES uzupełnionego o kilkadziesiąt dodatkowych elementów.

Problematicznym zagadnieniem jest konwersja danych między Dublin Core a innymi formatami. Twórcy formatów metadanych dążą do tego, by zapewnić możliwość mapowania do DCMES jako powszechnie znanego i stosowanego standardu. Konwersja jest też możliwa między Dublin Core a starszymi formatami, w szczególności formatem MARC 21, choć ze względu na bogatą strukturę MARC 21 nastrocza to pewne problemy¹³. Przy konwersji MARC 21 na Dublin Core część danych jest nieuchronnie tracona ze względu na pomijanie lokalnych etykiet. Przy konwersji w drugą stronę opisy zwykle trzeba rozwinąć, by zapewnić zgodność z obowiązującymi standardami. Problem stanowi również fakt, że w Dublin Core nie można kontrolować haseł – jest to możliwe dopiero po konwersji na MARC 21.

EAD – Encoded Archival Description

EAD jest standardem służącym do kodowania pomocy archiwalnych. Początkowo był oparty na SGML, z czasem stał się także zgodny z XML¹⁴.

Składają się na niego cztery grupy elementów:

- I. **Encoded Archival Description** – nagłówek
- II. **EAD Header** – metadane o pomocy archiwalnej
- III. **Front Matter** – strona tytułowa z wybranymi elementami
- IV. **Archival Description** – właściwa treść

METS – Metadata Encoding and Transmission Standard

METS to zaawansowany standard kodowania metadanych opisowych, administracyjnych i strukturalnych wyrażony w języku XML¹⁵. Opis w standardzie METS zawiera następujące sekcje:

11 <https://www.dublincore.org/specifications/dublin-core/dces/> [dostęp: 19.11.2022].

12 <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> [dostęp: 19.11.2022].

13 M. Nahotko, *Czym są metadane?*, wyd. cyt.; <https://www.loc.gov/marc/marc2dc.html> [dostęp: 19.11.2022]; <https://www.loc.gov/marc/dccross.html> [dostęp: 19.11.2022].

14 <https://www.loc.gov/ead/> [dostęp: 14.10.2022].

15 <https://www.loc.gov/standards/mets/> [dostęp: 14.10.2022].

- I. METS Header
- II. Descriptive Metadata
- III. Administrative Metadata
- IV. File Section
- V. Structural Map
- VI. Structural Links
- VII. Behavior

ISBD i RDA

Międzynarodowy Znormalizowany Opis Bibliograficzny ISBD (*International Standard Bibliographic Description*) to podstawowy standard powszechnej rejestracji bibliograficznej, pozwalający tworzyć dane bibliograficzne dla wszystkich typów publikacji, zapisane w formie elektronicznej przez społeczność międzynarodową. Podstawowy cel ISBD to zapewnienie spójności udostępnianych informacji bibliograficznych¹⁶. Potrzeba normalizacji w zakresie struktury opisów bibliograficznych oraz formatu ich wymiany zaczęła się wyraźnie zaznaczać w latach 60. i 70. XX wieku, kiedy to Biblioteka Kongresu uruchomiła program współkatalogowania (*Shared Cataloging Program*) i katalogowania w trakcie procesu wydawniczego (*CIP – Cataloging-in-Publication*) oraz rozwijała program UBC (*Universal Bibliographic Control*), a w pracach katalogowych w wielu krajach na świecie coraz powszechniej stosowano komputery, co zapoczątkowało międzynarodową wymianę danych bibliograficznych. Dzięki ujednoczeniu przepisów współpraca w zakresie wymiany rekordów miała stać się efektywniejsza i intensywniejsza¹⁷. Pierwszy ze standardów ISBD, opublikowany w roku 1971 ISBD(M) (*International Standard Bibliographic Description for Monographic Publications*), dotyczył opisu publikacji zwartych. Po trzech latach ogłoszono poprawiony tekst ISBD(M), a następnie sukcesywnie publikowano zasady opisu dla kolejnych typów publikacji – ISBD(CR), ISBD(NBM), ISBD(CM), ISBD(PM), ISBD(A), ISBD(ER) – i ogólne zasady opisu bibliograficznego – ISBD(G) – oraz ich nowelizacje.

Intencją IFLA nie było utworzenie przepisów katalogowania, lecz zbioru szczegółowych zasad, które mogą stanowić podstawę do opracowania takich przepisów w poszczególnych krajach¹⁸. Tak też się stało – zasady ISBD posłużyły do stworzenia przepisów i norm opisu bibliograficznego w krajach niemieckojęzycznych, Stanach Zjednoczonych, Wielkiej Brytanii, Francji, Związku Radzieckim, a w kolejnych latach także w innych państwach, w tym w Polsce.

16 <https://www.bn.org.pl/dla-bibliotekarzy/normy,-formaty,-standardy/katalogowanie/isbd> [dostęp: 18.11.2022].

17 Prace IFLA nad międzynarodowymi zasadami katalogowania. D. Chłopkowska, *Od Paris Principles, przez International Standard Bibliographic Description i Functional Requirements for Bibliographic Records do International Cataloging Principles*, „Przegląd Biblioteczny” 2012, z. 3, s. 319–336.

18 <https://www.ifla.org/references/best-practice-for-national-bibliographic-agencies-in-a-digital-age/resource-description-and-standards/bibliographic-control/international-standard-bibliographic-description-isbd/> [dostęp: 11.10.2022].

Prace nad ISBD zintensyfikowano w pierwszych latach XXI wieku, kiedy to m.in. pojawiła się konieczność uzgodnienia ich z wdrażanymi właśnie wymaganiami FRBR (omawianymi w kolejnym podrozdziale). Efektem tych wysiłków było opublikowanie wydania skonsolidowanego ISBD, którego końcowa wersja ukazała się w 2011 roku.

Zasadnicza cecha ISBD to standaryzacja budowy opisu bibliograficznego. Jest on podzielony na dziewięć stref:

- **strefa 0** – strefa typu zawartości i typu mediów,
- **strefa 1** – strefa tytułu i oznaczenia odpowiedzialności,
- **strefa 2** – strefa wydania,
- **strefa 3** – strefa specjalna,
- **strefa 4** – strefa adresu wydawniczego,
- **strefa 5** – strefa opisu fizycznego,
- **strefa 6** – strefa serii,
- **strefa 7** – strefa uwag,
- **strefa 8** – strefa międzynarodowego znormalizowanego numeru i sposobu uzyskania dokumentu.

Poszczególne elementy opisu zgodnie z ISBD mają ustalony porządek i są oddzielone znakami umownymi¹⁹.

RDA (*Resource Description and Access*) stanowi standard katalogowania, który zastąpił drugie wydanie zasad katalogowania AACR2 (*Anglo-American Cataloging Rules*). Pierwotnie zasady AACR2 miały zostać zrewidowane i zastąpione trzecią edycją – AACR3. W trakcie prac nad nią Joint Steering Committee for the Revision of AACR zdecydował jednak o całkowitej zmianie podejścia. Uświadomiono sobie, że poprzednie wersje AACR projektowane były w świecie zdominowanym przez katalogi kartkowe, natomiast teraz potrzebne są standardy przeznaczone dla środowiska cyfrowego, odpowiadające na wyzwania i możliwości cyfrowego świata.

Cel, który przyświecał twórcom RDA, to przygotowanie znormalizowanych wytycznych i instrukcji prezentujących spójne i elastyczne zasady opisu formy oraz zawartości różnego typu publikacji (także dokumentów elektronicznych, odgrywających coraz większą rolę w obiegu informacji w nauce), które będą zgodne z międzynarodowymi zasadami, modelami oraz standardami (pierwotna wersja RDA opierała się na modelach FRBR i FRAD, a obecna czerpie z IFLA LRM). W efekcie tych prac w 2010 powstał nowy standard – RDA, obejmujący wytyczne i instrukcje dotyczące opisu i dostępu do wszystkich zasobów cyfrowych i analogowych.

Jedną z kluczowych cech RDA to uniezależnienie prezentacji opisu od konkretnego formatu danych bibliograficznych oraz systemu informacyjno-wyszukiwawczego, który służy do udostępniania i przechowywania tych danych²⁰. RDA zapewniło: (1) elastyczną strukturę, która umożliwia opis zasobów dowolnego typu; co więcej, może być rozszerzona na potrzeby nowych typów, (2) dane, które można łatwo dostosować do nowych i powstających struktur

19 <https://www.bn.org.pl/dla-bibliotekarzy/normy,-formaty,-standardy/katalogowanie/isbd> [dostęp: 18.11.2022].

20 J. Franke, *RDA (Resource Description and Access) – geneza i koncepcja*, „Przegląd Biblioteczny” 2015, R. 83, z. 1, s. 7–61; <https://www.bn.org.pl/dla-bibliotekarzy/normy,-formaty,-standardy/katalogowanie/rda> [dostęp: 11.10.2022].

baz danych, (3) dane zgodne z rekordami istniejącymi już w katalogach bibliotecznych. To wszystko stwarza możliwość zastosowania tego standardu w różnych środowiskach cyfrowych (jak choćby Internet czy Web OPAC)²¹. Co również istotne, mimo że zalecenia RDA kierowane są głównie do świata bibliotekarskiego, to mogą być zaadaptowane także w innych środowiskach, jak archiwa czy muzea.

Zasadniczo standard ISBD jest w znacznej mierze związany z praktykami w katalogowaniu, w odniesieniu do opisu bibliograficznego konkretnego egzemplarza konkretnej materializacji, natomiast w standardzie RDA przyjęto znaczne rozszerzenie opisów o elementy, które wykraczają poza konkretną materializację i konkretny egzemplarz²².

ISAD(G) – *General International Standard Archival Description*

Międzynarodowy Standard Opisu Archiwalnego ISAD(G), opracowany przez Międzynarodową Radę Archiwów i opublikowany w roku 1994, składał się z 26 elementów pogrupowanych w 6 blokach²³. Są to:

- I. Blok identyfikujący – z podstawowymi informacjami o obiekcie,
- II. Blok proveniencji i archiwizacji – z informacjami na temat proveniencji i przechowywania obiektu,
- III. Blok opisu zawartości i układu akt – z informacjami o treści i stanie uporządkowania obiektu,
- IV. Blok udostępniania – z informacjami o możliwościach wykorzystania lub udostępniania obiektu,
- V. Blok źródeł uzupełniających – z informacjami o innych materiałach archiwalnych, które są w ważny sposób powiązane z opisywanym obiektem,
- VI. Blok uwag – z informacjami specjalnymi bądź takimi, które nie mogły zostać ujęte w pozostałych blokach.

W powstałej w 1999 roku kolejnej wersji ISAD(G) znalazł się dodatkowo:

- VII. Blok kontroli opisu – z informacjami o tym, jak, kiedy i kto przygotował opis obiektu.

Zgodnie z dokumentem wydanym przez Komisję Ad Hoc Międzynarodowej Rady Archiwów ISAD(G) „dostarcza ogólnych wskazówek dla przygotowania opisu archiwalnego. Należy go stosować/używać w połączeniu z istniejącymi krajowymi standardami lub jako podstawę dla opracowania takiego standardu”²⁴.

21 <https://www.librarianshipstudies.com/2017/07/resource-description-and-access-rda.html> [dostęp: 19.11.2022].

22 J. Kalinowski, *Podstawowe elementy opisu dzieła i realizacji w standardach RDA (Resource Description and Access) i ISBD (International Standard Bibliographic Description) i ich zapis w formacie MARC 21*, „Rocznik Biblioteki Narodowej” 2013, vol. 44, s. 327–336.

23 https://agad.gov.pl/?page_id=862 [dostęp: 14.10.2022].

24 https://agad.gov.pl/wp-content/uploads/2015/01/isad_ISADv2PL.pdf [dostęp: 14.10.2022].

Opis powstający w ramach ISAD(G) jest wielopoziomowy i podlega czterem podstawowym regułom:

1. Opis od ogółu do szczegółu
2. Informacje istotne dla poziomu opisu
3. Łączenie opisów
4. Niepowtarzanie tych samych informacji

Uzupełnienie ISAD(G) stanowi Międzynarodowy standard archiwalnych haseł wzorcowych ciał zbiorowych, osób i rodzin – ISAAR(CPF) (ang. *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*), w którym określa zasady tworzenia haseł wzorcowych.

Patrząc z perspektywy międzynarodowej wymiany informacji, za zasadnicze – takie, które obowiązkowo należy ująć w opisie – uznano 6 elementów:

- kod kraju, archiwum, zespołu, sygnatura,
- tytuł (nazwa),
- daty,
- poziom opisu,
- rozmiary,
- aktotwórca.

Modele organizacji danych bibliograficznych

FRBR (*Functional Requirements for Bibliographic Records*) stanowi model konceptualny, którego celem jest zwiększenie funkcjonalności i użyteczności danych bibliograficznych. W 1998 roku został on opublikowany przez Międzynarodową Federację Stowarzyszeń i Instytucji Bibliotekarskich (IFLA) jako model organizacji danych bibliograficznych. Była to reakcja na gwałtowny rozwój technologii informacyjnych i informatycznych, który – jeśli biblioteki miały zmierzać do integracji z cyfrowym środowiskiem informacyjnym – zmuszał do zmiany myślenia o działalności bibliotekarskiej oraz celach i metodzie opracowania bibliograficznego²⁵.

Model FRBR oparty jest na informatycznym (stosowanym wcześniej do porządkowania relacyjnych baz danych) modelu jednostka-relacja²⁶, opracowanym metodą analizy jednostek i relacji, w której operuje się trzema typami elementów: jednostkami (*entities*) i ich atrybutami (*attributes*) oraz relacjami (*relationships*)²⁷. Dzięki umieszczeniu danych w złożonej struk-

25 J. Pacek, *Nowoczesne standardy metadanych*, „Bibliotekarz” 2015, nr 7/8, s. 4–9.

26 E. F. Codd, *A relational model of data for large shared data banks*, „Communications of the ACM” 1970, 13(6), s. 377–387; P. P. Chen, *The entity-relationship model – toward a unified view of data*, „ACM Transactions on Database Systems” 1976, 1(1), s. 9–36.

27 A. Filipek, *Model opisu dokumentu według FRBR (Functional Requirements for Bibliographic Records)*, „Zagadnienia Informacji Naukowej” 2015, nr 2, s. 13–30.

turze połączeń semantycznych możliwe jest uzyskanie szerszego kontekstu i przetworzenie informacji w wiedzę²⁸.

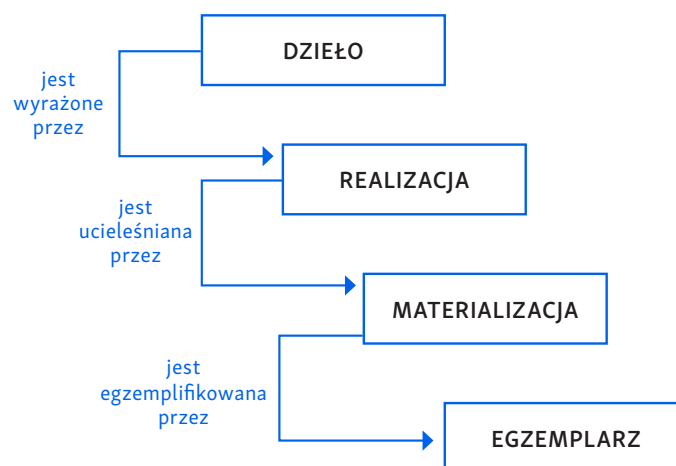
W FRBR jednostki (encje) bibliograficzne podzielono na trzy grupy:

- Grupa 1 to Dzieło (*Work*), Realizacja (*Expression*), Materializacja (*Manifestation*) i Egzemplarz (*Item*).
- Grupa 2 to Osoba (*Person*) i Ciało zbiorowe (*Corporate body*).
- Grupa 3 to Pojęcie (*Concept*), Rzecz (*Object*), Zdarzenie (*Event*) i Miejsce (*Place*).

Zgodnie z modelem FRBR Dzieło jest odrębnym wytworem działalności intelektualnej lub artystycznej. Na tym poziomie w opisie bibliograficznym powinny znaleźć się informacje o tytule, autorze oraz dacie powstania dzieła rozumianego jako wytwór działalności intelektualnej bądź artystycznej w formie abstrakcyjnej. Realizacja to intelektualne lub artystyczne wyrażenie Dzieła w formie zapisu alfanumerycznego, choreograficznego, dźwiękowego bądź cyfrowego. Może to również być dowolna kombinacja tych form. Różnymi Realizacjami tego samego Dzieła mogą być np. tekst, zapis dźwiękowy i film. Materializację stanowią wszystkie obiekty fizyczne o tych samych cechach – zarówno pod względem zawartości intelektualnej, jak i formy fizycznej. Mogą to być np. kolejne wydania Dzieła. Ten poziom modelu odpowiada informacjom przede wszystkich z takich stref opisu bibliograficznego jak strefa wydania, adresu wydawniczego, opisu fizycznego, identyfikatora i sposobu uzyskania. Najniższy poziom to Egzemplarz, to znaczy pojedynczy obiekt w ramach tej samej Materializacji²⁹. Encje te powiązane są ze sobą następującymi relacjami (przedstawionymi na rys. 1)³⁰:

- Dzieło jest wyrażone przez Realizację,
- Realizacja jest ucieleśniana przez Materializację,
- Materializacja jest egzemplifikowana przez Egzemplarz.

Rys. 1. Relacje między encjami Grupy 1³¹



28 J. Pacek, *Nowoczesne standardy metadanych*, wyd. cyt.

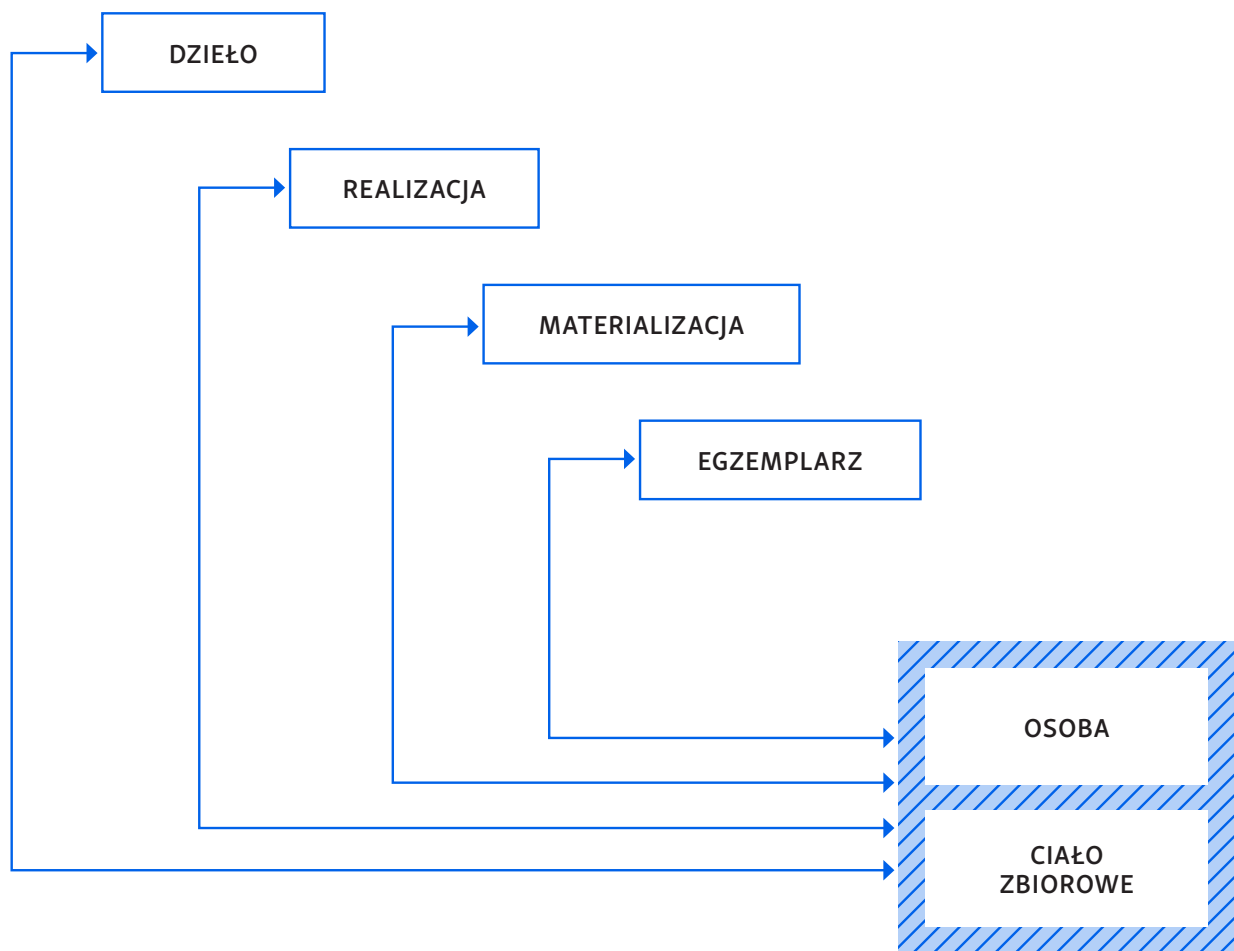
29 D. Paleczna, *FRBR jako model danych bibliograficznych dla bibliotekarza i użytkownika końcowego*, „Acta Universitatis Lodzianensis. Folia Librorum” 2014, 1(18), s. 129–141; <https://www.bn.org.pl/dla-bibliotekarzy/normy,-formaty,-standardy/katalogowanie/modele-frbr> [dostęp: 11.10.2022].

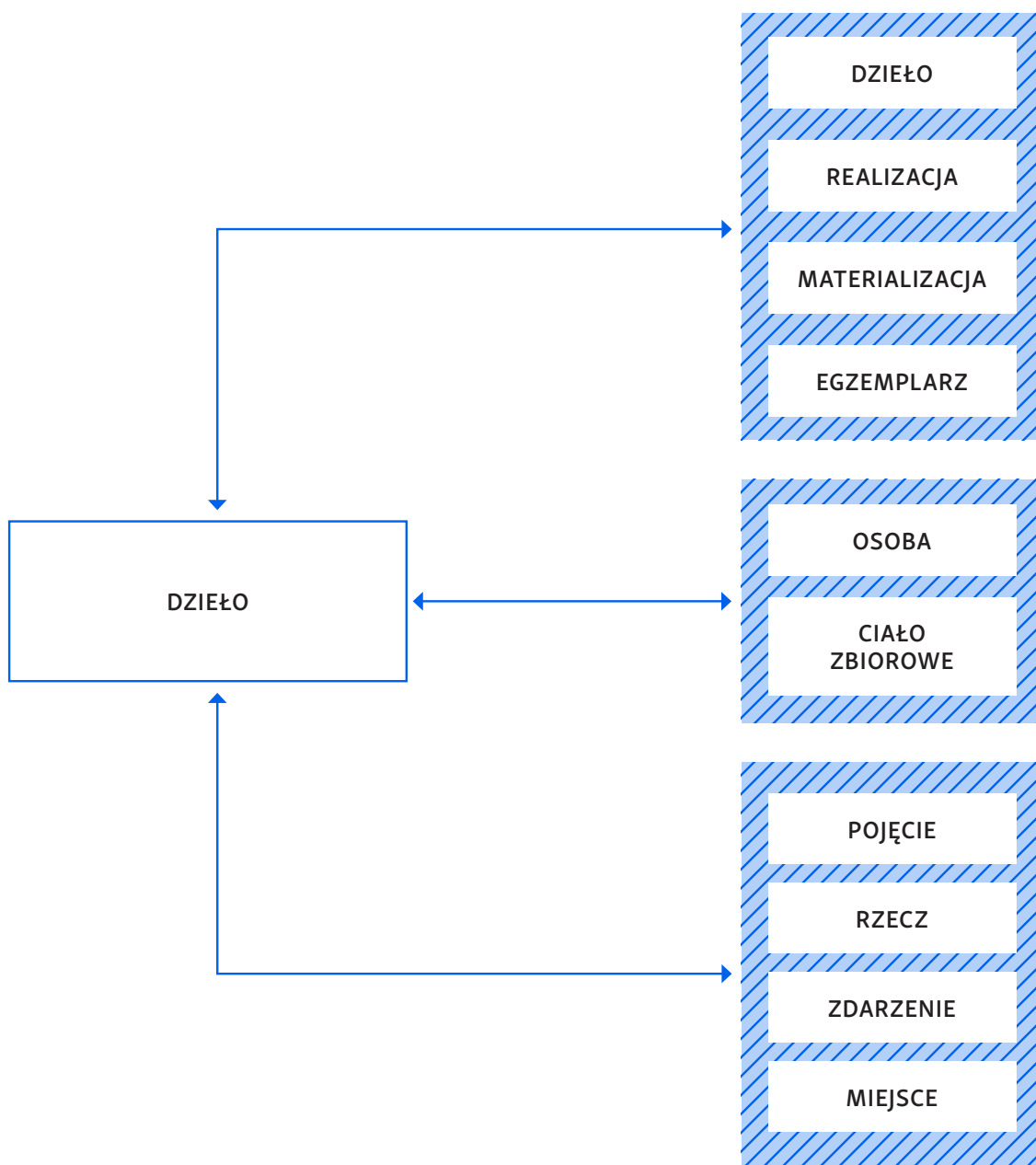
30 E. Golec-Nycz, *FRBR (eferberyzacja) – nowa filozofia katalogowania*, „Bibliotekarz” 2004, nr 9, s. 9–13.

31 Opracowanie własne na podstawie: E. Golec-Nycz, *FRBR (eferberyzacja)...*, wyd. cyt.

Ponadto z jednostkami z Grupy 1 powiązane są te z Grup 2 i 3. Osoby lub Ciała zbiorowe są podmiotami, które tworzą, publikują, zachowują encje z Grupy 1. Osobami mogą być np. autorzy, kompozytorzy, artyści, recenzenci, tłumacze, reżyserzy itp., natomiast Ciała zbiorowe to organizacje lub grupy osób i/lub organizacji działających wspólnie. Z kolei encje z Grupy 3 odnoszą się do tematów opisujących encje z Grupy 1. Mogą to być choćby dziedziny bądź dyscypliny wiedzy, teorie, techniki (Pojęcia), obiekty materialne – ożywione lub martwe (Rzeczy), wydarzenia historyczne, epoki, okresy (Zdarzenia), lokalizacje ziemskie i pozaziemskie, historyczne i współczesne, geograficzne i geopolityczne (Miejsca), ale jako temat mogą występować też jednostki dwóch wcześniejszych grup, to znaczy Dzieło może dotyczyć innych Dzieł, Realizacji, Materializacji, Egzemplarzy, Osób lub Ciał zbiorowych. Relacje te ilustrują rys. 2 i 3.

Rys. 2. Relacje między encjami Grup 1 i 2³²



Rys. 3. Pełny schemat modelu FRBR³³

Takie pogrupowanie encji i relacji między nimi pozwala na tzw. eferberyzację, tzn. uporządkowanie katalogów bibliotecznych tak, by zebrać w jednym miejscu informacje o obiektach, pozostających ze sobą w pewnej relacji semantycznej. Stwarza to użytkownikowi możliwość odnalezienia zbioru informacji, który jest uporządkowany logicznie, hierarchicznie i łatwy do intuicyjnego przeszukiwania³⁴.

Według twórców FRBR rekordy bibliograficzne tworzone zgodnie z tym modelem mają służyć czterem zasadniczym celom realizowanym przez użytkowników poszukujących informacji bibliograficznych. Te cztery zadania to:

33 Opracowanie własne na podstawie: tamże.

34 J. Pacek, *Nowoczesne standardy metadanych*, wyd. cyt.

- wyszukiwanie (*find*) jednostek zgodnych z kryteriami wyszukiwawczymi ustalonymi przez użytkownika (np. wszystkich publikacji na dany temat lub o danym tytule),
- identyfikacja (*identify*) odpowiedniej jednostki (np. rozróżnienie dokumentów o takim samym tytule lub upewnienie się, że dany rekord odnosi się do dokumentu rzeczywiście poszukiwanego przez użytkownika),
- wybór (*select*) jednostki spełniającej wymagania użytkownika (np. wybór tekstu w odpowiednim dla użytkownika języku bądź wybór odpowiedniej wersji programu komputerowego),
- uzyskanie dostępu (*obtain*) do danej jednostki (np. jej zakup, wypożyczenie lub skopiowanie)³⁵.

W 2009 i 2010 roku IFLA ogłosiła kolejne raporty poświęcone wymaganiom funkcjonalnym – najpierw FRAD (*Functional Requirements for Authority Data*), dotyczący danych wzorcowych, potem FRSAD (*Functional Requirements for Subject Authority Data*), dotyczący danych wzorcowych wyłącznie opisu rzeczowego. Modele te stanowią rozwinięcie FRBR i wspólnie z nim tworzą rodzinę trzech modeli konceptualnych IFLA dotyczących wymagań funkcjonalnych.

FRAD zawiera analizę jednostek wykorzystywanych w bazach bibliograficznych jako dane wzorcowe, nazw, pod którymi znane są te jednostki oraz ujednoczonych punktów dostępu dla nich tworzonych, a także opisuje atrybuty tych jednostek i relacje między nimi. Model ten składa się z trzech elementów powiązanych ze sobą relacjami. W grupie jednostek bibliograficznych do tych ustanowionych już przez model FRBR dodano encję Rodzina. Drugą grupę stanowią Nazwy, za pomocą których encje bibliograficzne są rozpoznawane oraz Identyfikatory przyznane tym encjom. Natomiast najniższy poziom to Ujednoczone punkty dostępu, które bazują na Nazwach i Identyfikatorach uznanych za dane wzorcowe.

Celem FRSAD natomiast było zbudowanie modelu konceptualnego opisującego encje będące wyrażeniem przedmiotu dzieła i relacji między nimi. Kluczowa zmiana to wprowadzenie pojęć Thema i Nomen, które wykorzystano do określenia dowolnej encji używanej jako przedmiot dzieła oraz dowolnego znaku lub ciągu znaków, za pomocą których Thema jest wyrażona i rozpoznawana. Znaczenie tych terminów można streścić następująco: Dzieło ma za przedmiot jedno lub wiele Thema, a jedno Thema może być przedmiotem jednego lub więcej Dzieł. Zarazem Thema może mieć za nazwę jedno lub wiele Nomen, a jedno Nomen może stanowić wyrażenie jednego lub wielu Thema. Należy przy tym zaznaczyć, że Thema jest klasą nadrzędną dla encji Pojęcie – zawiera wszystkie encje wyróżniane przez FRBR i FRAD. Także Nomen nie odpowiada encji Nazwa obecnej w modelu FRAD – jest klasą szerszą, ponieważ zawiera wszystkie encje z tego modelu.

W 2017 IFLA skonsolidowała FRBR, FRAD i FRSAD, powołując do życia IFLA LRM (*IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*), który miał być udoskonaloną, pozbawioną pewnych nieścisłości wersją swoich poprzedników. W rezultacie stworzono pojedynczy model, który jest usprawniony i logicznie spójny, obejmuje wszystkie aspekty danych bibliograficznych i zawiera ustrukturyzowany opis ontologii. Jest też dosto-

35 Tamże.

sowane do nowych realiów, w których istotna stała się kwestia wykorzystania danych bibliotecznych poza środowiskiem bibliotecznym – w aplikacjach sieci semantycznej³⁶.

W kontekście FRBR mówi się o „nowej filozofii katalogowania”³⁷, zupełnej zmianie myślenia o informacji bibliograficznej. Innowacyjność tego konceptu polegała na zastąpieniu płaskiego sposobu budowania relacji bibliograficznych opisem wielopoziomowym i wielorelacyjnym³⁸. W FRBR nie rejestruje się już jedynie katalogowanego egzemplarza, ale też jego relacje z innymi elementami uniwersum bibliograficznego – zarówno relacje z twórcami czy poprzednimi wydaniem, ale też z dowolnymi innymi elementami, które mogą w jakiś sposób okazać się użyteczne z perspektywy użytkownika.

Należy również podkreślić, że rodzina modeli FRBR nie jest powiązana z żadnym ze standardów katalogowania, dzięki czemu charakteryzuje się uniwersalnością, jednak trzeba przy tym pamiętać, że nie stanowi ona instrukcji czy też zbioru reguł katalogowania. Są to modele koncepcyjne, które wyjaśniają strukturę informacji oraz w jaki sposób jej poszczególne elementy są ze sobą powiązane. W gruncie rzeczy stanowią więc one punkt wyjścia do opracowywania modeli danych na podstawie właściwego im modelowania poziomu abstrakcyjnego³⁹.

Format MARC

MARC (Machine-Readable Cataloging) został opracowany w latach sześćdziesiątych XX wieku w Stanach Zjednoczonych, natomiast w roku 1997 w wyniku połączenia wersji amerykańskiej (US MARC) i kanadyjskiej (CAN MARC) stworzono format MARC 21. MARC 21 stanowi format dla opisu bibliograficznego tworzonego przy użyciu komputerów. Jest to format, który służy do zapisu oraz wymiany uporządkowanych i ujednoczonych danych. Rekordy powstające zgodnie z formatem MARC 21 zawierają pola o wyróżnikach cyfrowych, które jednoznacznie identyfikują zawartość pól. Pola te są dzielone na podpola, które są identyfikowane za pomocą jednoznakowych etykiet. Istnieje również MARC XML, który stanowi reprezentację MARC 21 wyrażoną za pomocą języka XML⁴⁰.

Co istotne, format MARC jest jedynie swego rodzaju „matrycą” do rejestrowania i przechowywania danych – nie daje konkretnych wytycznych na temat katalogowania, czyli doboru i reguł formułowania informacji bibliograficznych. Format MARC 21 przystosowany jest wyłącznie do opisu obiektów bibliotecznych, natomiast nie jest przydatny do tworzenia wielopoziomowych opisów obiektów archiwalnych. Opis bibliograficzny w formacie MARC 21 szczegółowo odzwierciedla cechy formalne obiektu oraz umożliwia przedstawienie tematu treści publikacji, a także opisanie relacji pomiędzy poszczególnymi obiektami.

36 Library Reference Model: krok w kierunku Sieci Semantycznej: <https://babin.bn.org.pl/?p=5205> [dostęp: 18.11.2022].

37 E. Golec-Nycz, *FRBR (eferberyzacja)...*, wyd. cyt.

38 J. Kalinowski, *Podstawowe elementy opisu dzieła...*, wyd. cyt.

39 M. Cichoń, *Zależności pomiędzy FRBR, RDA a zasadami katalogowania w kontekście integracji metadanych*, „Annales Universitatis Paedagogicae Cracoviensis. Studia ad Bibliothecarum Scientiam Pertinentia” 2017, vol. 15, s. 39–54.

40 *Digitalizacja piśmiennictwa...*, wyd. cyt.

Kluczowe dla formatu MARC są:

- struktura – określona postać fizyczna rekordu na nośniku elektronicznym,
- znaczniki treści – elementy identyfikujące i charakteryzujące dane zawarte w rekordzie, umożliwiające prawidłową ich interpretację (etykiety pól, wskaźniki, kody podpól),
- zawartość – konkretne dane (informacje) sformułowane zgodnie z przyjętymi zasadami, które należy stosować razem z dokumentacją formatu (zalecenia krajowe i międzynarodowe są treścią osobnych opracowań, np. ISBD).

Do głównych cech rekordów w formacie MARC należą zmienna długość rekordu i jego elementów, podział na jednoznacznie identyfikowane pola i podpola, obecność wskaźników oraz powtarzalność pól i podpól. Istotną zaletę formatu MARC 21 stanowi jego jednoznaczność. Należy jednak zauważyć, że nie jest ona pełna, ponieważ istnieją braki w precyzji niektórych podpól. Dzięki ustandaryzowanemu zapisowi danych w formacie MARC 21 jest możliwa wymiana danych pomiędzy różnymi systemami bibliotecznymi. Może to być realizowane przy wykorzystaniu protokołów opisanych w rozdziale 6.7.2.

Pomimo ogromnej popularności tego formatu, coraz częściej w literaturze pojawia się stwierdzenia, że aby zwiększyć elastyczność katalogów bibliotecznych oraz ułatwić stosowanie nowoczesnych rozwiązań technologicznych format MARC powinien zostać zastąpiony⁴¹.

Powszechna dostępność danych w Internecie obnażyła wiele niedoskonałości formatu MARC, szczególnie fakt, że brak znajomości zasad formatu MARC przez większość użytkowników Internetu wpływa na trudności w wyszukiwaniu. Zasadniczą wadą i ograniczeniem formatu MARC jest jego skupienie na dokumencie a nie na informacji⁴². Ponadto proces wprowadzania danych w formacie MARC 21 wymaga dobrej znajomości zasad sporządzania opisów bibliograficznych dla poszczególnych typów publikacji, co wpływa na dużą pracochłonność i kosztowność tego procesu. Efektywne korzystanie z katalogów bibliotecznych opartych na tym formacie wymaga najczęściej przynajmniej elementarnej znajomości języka wyszukiwawczego, co może generować trudności przy korzystaniu z katalogów przez mniej doświadczonych użytkowników.

Szczegółowe omówienie formatu MARC, w tym poszczególnych pól, nie jest celem niniejszego opracowania, szczegóły można znaleźć między innymi na stronach Biblioteki Kongresu (USA)⁴³.

41 M. Krynicka, *Charakterystyka formatu MARC 21 – stan obecny i perspektywy rozwoju*, „Zagadnienia Informatyki Naukowej: studia informacyjne” 2012, t. 50, nr 2 (100), s. 89–99; A. J. Kroeger, *The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC future*, „Cataloging & Classification Quarterly” 2013, 51, nr 8, s. 873–890; X. Amanda, H. Kirk, L. Akerman, *From MARC to BIBFRAME 2.0: crosswalks*, „Cataloging and Classification Quarterly” 2018, vol. 56, iss. 2–3, s. 224–250.

42 D. Siwecka, *Najnowsze kierunki w udostępnianiu danych bibliograficznych (Resource Description Framework oraz Linked Open Data)*, „Praktyka i Teoria Informatyki Naukowej i Technicznej” 2014, t. 22, nr 4, s. 22–32.

43 <https://www.loc.gov/marc/> [dostęp 18.11.2022].

Języki informacyjno-wyszukiwawcze

Przy tworzeniu opisów obiektów dużą rolę odgrywa stosowany język informacyjno-wyszukiwawczy. Ważne jest, aby w jak najlepszy sposób realizowane były dwie podstawowe funkcje: metainformacyjna, polegająca na odwzorowaniu cech informacji, oraz wyszukiwawcza, umożliwiająca odnalezienie informacji spełniającej warunki zapytania⁴⁴. Języki informacyjno-wyszukiwawcze można zdefiniować jako sztuczne języki, które pozwalają stworzyć sformalizowaną reprezentację dokumentów i zapytań interesujących pewną grupę użytkowników w celu wyszukania dokumentów, które odpowiadają na zapytania⁴⁵. Pośród powszechnie stosowanych języków informacyjno-wyszukiwawczych można wyróżnić:

- klasyfikacje dokumentów, w tym mono- i polihierarchiczne,
- języki słów kluczowych, w tym języki unitermowe,
- języki haseł przedmiotowych,
- języki deskryptorowe,
- kody semantyczne.

W ostatnim czasie w bardzo szybkim tempie swoją popularność zdobywają języki deskryptorowe. Języki deskryptorowe można zdefiniować jako typ języków informacyjno-wyszukiwawczych, w których słownictwo jest zaczerpnięte z języka naturalnego, elementarne jednostki nazywane są deskryptorami, słownik zaś najczęściej – tezaurem, a gramatykę stanowią reguły indeksowania współrzędnego. Indeksowanie współrzędne daje możliwość wyszukiwania dokumentów przy użyciu swobodnie zestawionych wyrażen⁴⁶. W tego typu językach eliminowana jest wieloznaczność przy zachowaniu podobieństwa do języków naturalnych. Zastosowanie języków deskryptorowych daje możliwość wykorzystania wyszukiwania fasetowego, które stanowi bardzo wygodne dla użytkowników narzędzie umożliwiające zawężanie lub poszerzanie zakresu poszukiwań w zależności od potrzeb. Informacje różnych typów są rozdzielane, co wpływa na zwiększenie liczby punktów dostępu do danych o katalogowanych materiałach⁴⁷. Fasety stanowią niezależne właściwości lub wymiary, dzięki którym może zostać sklasyfikowany obiekt. Fasety bardzo pomagają w kategoryzacji treści⁴⁸. Zgodnie z inną definicją, fasetą stanowi jedną z kategorii przedmiotów lub ich nazw, stosowanych w danej dziedzinie nauki, techniki lub działalności praktycznej⁴⁹. Fasety pozwalają na organizowanie treści i wyszukiwanie informacji zgodnie z potrzebami różnych użytkowników, nie ograniczają możliwości prezentacji treści do z góry określonych⁵⁰.

44 J. Sadowska, T. Turowska, *Języki informacyjno-wyszukiwawcze: katalogi rzeczowe*, Warszawa 1990, s. 71–75.

45 Y. Courrier, *Analyse et langage documentaire*, „Documentaliste – sciences de l’information” 1976, 13, s. 178–189.

46 J. Sadowska, T. Turowska, *Języki informacyjno-wyszukiwawcze...*, wyd. cyt.; *Język deskryptorowy*, w: *Podręczny słownik bibliotekarza*, oprac. G. Czapnik, Z. Gruszka przy współpracy H. Tadeusiewicz, Warszawa 2011, s. 129.

47 S. Skórka, *Fasety na nowo odkryte. Integrowanie systemów nawigacji i organizowania informacji*, „Zagadnienia Informatyki Naukowej” 2014, t. 52, nr 2, s. 92–109.

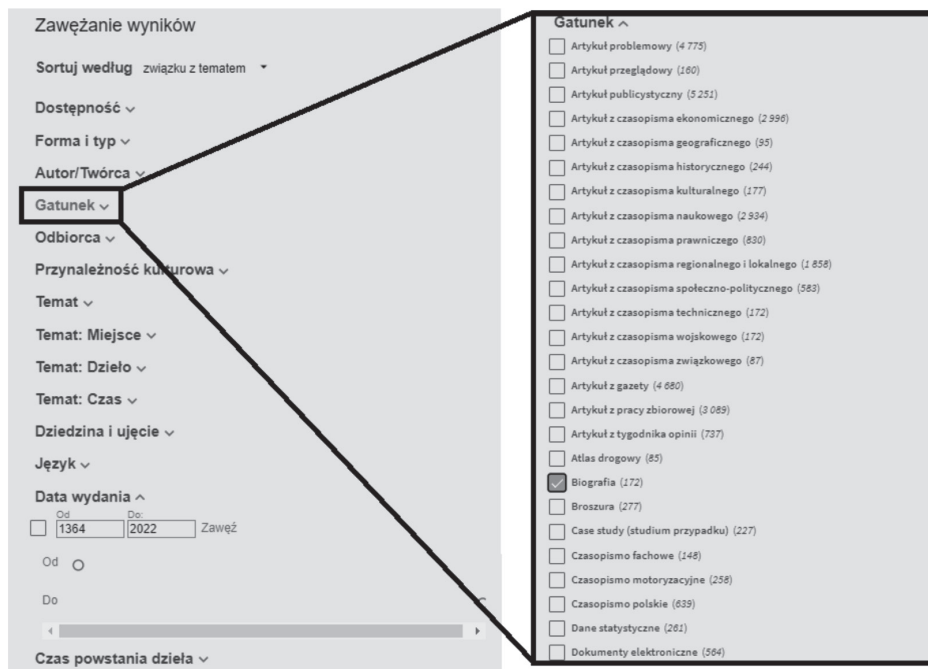
48 Tamże.

49 A. I. Czerny, *Teoria wyszukiwania informacji*, Warszawa 1981, s. 257.

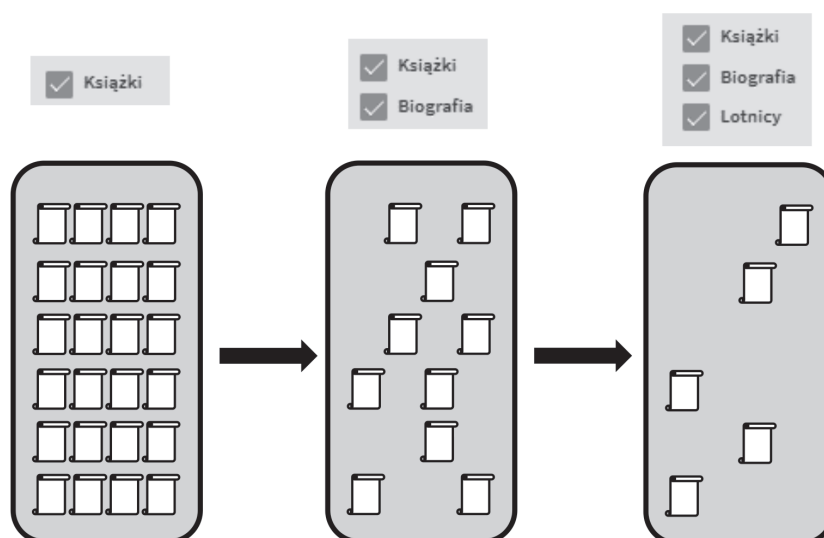
50 S. Papa, *The faceted navigation and search revolution*, <https://www.kmworld.com/Articles/White-Paper/Article/The-Faceted-Navigation-and-Search-Revolution-15378.aspx> [dostęp: 18.11.2022].

Nawigacja fasetowa opiera się na interakcji interfejsu z użytkownikiem. Użytkownik może stopniowo zmniejszać liczbę obiektów w zbiorze wyników poprzez klikanie w określoną fasetę, reprezentującą określoną szczegółową kategorię (rys. 5). Przykładowo użytkownik, który chce spośród wyników wyszukiwawczych wyświetlić wyłącznie opisy bibliograficzne książek stanowiących biografie lotników, powinien zaznaczyć z kategorii głównej Forma i typ fasetę Książki oraz w kategorii Gatunek fasetę Biografia, a następnie z kategorii Temat fasetę Lotnicy. W każdej chwili może zrezygnować z tego zawężenia wyników, odznaczając niektóre lub wszystkie fasety.

Rys. 4. Przykład faset w katalogu internetowym Biblioteki Narodowej⁵¹



Rys. 5. Zasada działania nawigacji fasetowej⁵²



51 <https://katalogi.bn.org.pl/> [dostęp: 18.11.2022].

52 Opracowanie własne na podstawie: S. Skórka, *Fasety na nowo odkryte...*, wyd. cyt.

Deskryptory Biblioteki Narodowej (DBN)

Deskryptory Biblioteki Narodowej (DBN) to język oparty na słownictwie stosowanego wcześniej Języka Haseł Przedmiotowych Biblioteki Narodowej (JHP BN). DBN stanowią wyraz bądź grupę wyrazów będących preferowaną nazwą danego elementu rzeczywistości, a także jednoznacznie identyfikują ten element. W ten sposób stanowią ujednolicone i kontrolowane punkty dostępu.

Deskryptory Biblioteki Narodowej wykorzystuje się zarówno w indeksowaniu cech formalnych zbiorów bibliotecznych, jak i w opisie przedmiotowym. Deskryptory są ze sobą nawzajem powiązane relacjami semantycznymi, tworząc bazę rekordów wzorcowych.

Typy DBN:

- deskryptor osobowy,
- deskryptor korporatywny,
- deskryptor imprezy,
- deskryptor tytułu,
- deskryptor chronologiczny,
- deskryptor przedmiotowy,
- deskryptor geograficzny,
- deskryptor formy/rodzaju/gatunku,
- deskryptor ujęciowy.

Istotną cechą Deskryptorów Biblioteki Narodowej są atrybuty, czyli cechy wyróżniające osoby lub rzeczy spośród innych.

Wymiana danych

Zasady FAIR

W opublikowanej w 2016 roku pracy zespół autorów wskazał podstawowe cechy, które powinny posiadać udostępniane dane⁵³. Zasady te zostały nazwane FAIR, tj.: *findable*, *accessible*, *interoperable*, *reusable*. Zgodnie z tymi założeniami, udostępniane dane powinny być:

- łatwe do odnalezienia (*findable*), powinny być udostępniane w taki sposób, aby było możliwe ich proste i szybkie przeszukiwanie oraz odnajdowanie. Udostępniane dane powinny być łatwe do znalezienia nie tylko przez ludzi, ale również przez komputery w ramach odczytu i przetwarzania maszynowego. Dane powinny posiadać trwale przypisane identyfikatory, które powinny być unikalne. Ważne, aby określone dane mogły być zawsze odnajdowane za pomocą tego samego identyfikatora. Pośród takich

53 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg i in., *The FAIR Guiding Principles for scientific data management and stewardship*, „Scientific Data” 2016, 3 (160018).

unikalnych identyfikatorów można wymienić cyfrowy identyfikator dokumentu elektronicznego (DOI), ORCID oraz odnośniki bezpośrednie (permalinki). Istotne jest również prawidłowe i trwałe powiązanie danych z danymi o tych danych (metadanymi). Dane powinny być w odpowiedni sposób indeksowane, aby mogły być wyszukiwane i były widoczne w Internecie;

- dostępne dla wszystkich (*accessible*), ważne jest, aby istniały jasne zasady, na których wymagane dane są dostępne dla użytkowników. Dane powinny być dostępne przy wykorzystaniu jednego ze znormalizowanych protokołów komunikacyjnych. Dane mogą być udostępniane w ramach otwartego, bezpłatnego protokołu dostępnego i stosowanego w większości krajów (np. HTTP, FTP, SMTP) lub protokołu, który wymaga uwierzytelnienia i autoryzacji (np. HTTPS, FTPS). Dostęp nie zawsze musi być otwarty i bezpłatny, jednak ważne jest, aby zasady dostępu były zawsze jednoznacznie określone, a sposób uzyskania danych był intuicyjny i nie wymagał wykonywania nadmiernej liczby czynności, najlepiej aby mógł być w pełni zautomatyzowany. Zakłada się, że metadane będą dostępne znacznie dłużej od samych danych, których przechowywanie może z czasem okazać się zbyt kosztowne lub z innych przyczyn niezasadne;
- interoperacyjne (*interoperable*), oznacza to, że dane powinny być łatwe do integracji z innymi danymi, a także powinny być łatwo przetwarzalne, szczególnie pod kątem wykorzystania ich przez aplikacje komputerowe. Pomiędzy poszczególnymi metadanymi powinien istnieć szereg istotnych powiązań. Język danych powinien być powszechnie znany oraz posiadać jasne i jednoznaczne reguły. Istotne jest, aby stosowane były powszechnie używane słowniki kontrolowane, ontologia oraz tezaury. Coraz częściej wykorzystywany jest model RDF;
- udostępnione, w sposób umożliwiający ich wielokrotne użycie przy zachowaniu postanowień licencyjnych (*reusable*). Dane powinny być opisane poprzez szereg etykiet, co powinno wpłynąć na znacznie łatwiejsze odnajdowanie i ponowne wykorzystywanie danych. Między innymi warto umieścić takie informacje jak: cel wygenerowania lub zebrania danych, ograniczenia związane z danymi, wyjaśnienia wszystkich zmiennych, datę wygenerowania lub zebrania danych, a także inne informacje związane z przygotowaniem danych. Postanowienia licencyjne powinny być przedstawione w sposób klarowny i jednoznaczny. Musi również być podana informacja, w jaki sposób osoby wykorzystujące dane powinny na to wskazywać (cytowania itp.). Dostawca danych powinien dostarczać jak największą ich liczbę, nie ograniczając się wyłącznie do danych, które w jego ocenie mogą być przydatne użytkownikowi, gdyż trudno jest przewidzieć wszystkie potrzeby obecnych i przyszłych użytkowników⁵⁴.

54 B. Mons, C. Neylon, J. Velterop i in., *Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud*, „Information Services & Use” 2017, vol. 37, iss. 1, s. 49–56; <https://www.go-fair.org/fair-principles/> [dostęp: 15.11.2022].

Protokoły wymiany danych

Pośród powszechnie stosowanych lub zdobywających swoją popularność sposobów wymiany danych można wymienić następujące:

- Z39.50 – stanowi powszechnie stosowany protokół, umożliwiający wymianę danych bibliograficznych pomiędzy poszczególnymi systemami bibliotecznymi, został zaproponowany w 1984 roku, natomiast od 1988 roku zaznaczył się jego silny rozwój. Od tego czasu powstały liczne rozszerzenia i udoskonalenia protokołu, jednak – ze względu na swoje ograniczenia, szczególnie trudności w prawidłowym indeksowaniu przez wyszukiwarki internetowe⁵⁵ – coraz częściej jest on zastępowany przez inne protokoły.
- OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) – założenia zostały opracowane przez organizację Open Archives Initiative. Pierwszą wersję przedstawiono w 2001 roku, kolejną w 2002, natomiast w następnych latach wdrażano kolejne aktualizacje. Zgodnie z założeniami protokołu OAI-PMH jest w pełni zgodny z Dublin Core⁵⁶.
- REST API (*Representational State Transfer Application Programming Interface*) – nazywany również aplikacyjnym interfejsem programistycznym zgodnym ze stylem architektury REST⁵⁷. Jako jedna z największych zalet REST API podawana jest wysoka elastyczność i uniwersalność. Do komunikacji wykorzystywane są żądania HTTP, dzięki którym możliwe jest korzystanie z podstawowych funkcjonalności baz danych m.in. tworzenie, odczytywanie, aktualizowanie i usuwanie rekordów w zasobie. W ramach tego sposobu wymiany danych istnieje możliwość stosowania większości znanych języków programowania, a także bardzo licznych formatów danych. Prawidłowo zaprojektowane interfejsy REST API powinny być zgodne z podstawowymi założeniami architektury REST, tj:
 - ▶ interfejs powinna charakteryzować jednolitość, każdy element danych powinien mieć przypisywany unikalny identyfikator,
 - ▶ dostęp do danych przez użytkownika (klienta) powinien być niezależny od wykorzystywanych aplikacji zarówno przez dostawcę danych, jak i użytkownika (klienta),
 - ▶ bezstanowość, co oznacza, że po stronie dostawcy danych nie powinny być przechowywane dane, które dotyczą poszczególnych żądań użytkownika (klienta),
 - ▶ umożliwienie buforowania danych zarówno po stronie użytkownika (klienta), jak i po stronie serwera udostępniającego dane,
 - ▶ budowa warstwowa, jednak użytkownik (klient) nie powinien posiadać informacji, czy komunikuje się z warstwą (elementem) pośredniczącą, czy aplikacją końcową⁵⁸.

55 PN-ISO 23950:2002 – wersja polska. Informacja i dokumentacja – Pobieranie informacji (Z39.50) – Definicja usługi aplikacyjnej i specyfikacja protokołu, 2002; <http://www.loc.gov/z3950/#about> [dostęp: 16.11.2022].

56 <http://www.openarchives.org/pmh/> [dostęp: 19.11.2022].

57 <https://www.ibm.com/pl-pl/cloud/learn/rest-apis> [dostęp: 16.11.2022].

58 M. Śliwa, B. Pańczyk, *Porównanie wydajności interfejsów programistycznych na przykładzie REST API, GraphQL i gRPC*, „Journal of Computer Sciences Institute” 2021, vol. 21, s. 356–361; S. P. Ong, S. Cholia, A. J. M. Brafman

Format RDF

Aby sprostać rosnącym wymaganiom związanym z wymianą danych, w tym wymaganiom FAIR, z coraz większym powodzeniem jest wykorzystywany model/format RDF (*Resource Description Framework*)⁵⁹. Założenia RDF zostały opracowane na przełomie XX i XXI wieku przez W3C (*World Wide Web Consortium*). Jako główne przeznaczenie tego formatu wskazano środowisko internetowe. RDF jest oparty na składni języka XML (*Extensible Markup Language*), lecz nałożono znaczne ograniczenia na stosowane znaczniki, co zgodnie z założeniami powinno wpłynąć na ułatwienie maszynowego przetwarzania danych w tym formacie. Format ten z założenia jest zrozumiały również dla komputerów, a nie tylko dla człowieka, co znacznie ułatwia automatyczne przetwarzanie danych oraz naprowadzanie przez komputery na konkretne powiązania, istotne z punktu widzenia określonego wyszukiwania⁶⁰. Wskazuje się, że format RDF stanowi istotny element w dążeniu do powstania sieci semantycznych (*Semantic Web*), w których komputery przy przetwarzaniu danych będą w stanie rozróżniać homonimy, a także dokonywać prawidłowego wnioskowania na podstawie tych danych⁶¹. Zasób każdego rodzaju, w tym dokumenty tekstowe, pliki audiowizualne, zdjęcia, jest opisywany za pomocą zdania/stwierdzenia (*statement*) wyrażonego w języku RDF, składającego się z trzech podstawowych elementów, tj.: podmiotu (*subject*), orzeczenia (predykatu/własności; *predicate*) i dopełnienia (obiektu/wartości; *object*). Te trzy elementy nazywane są trójkami RDF, które opisują zasoby zgodnie z analogią do tworzenia zdań. Pośród sposobów reprezentacji RDF można wskazać pliki XML, grafy oraz wyliczenia poszczególnych trójek. Fakt, że format ten umożliwia zapisywanie danych w formie grafu skierowanego wpływa na znaczną poprawę czytelności danych. Przy takiej reprezentacji wierzchołki przedstawiają podmiot i dopełnienie, natomiast krawędzie skierowane od podmiotu do dopełnienia reprezentują orzeczenia⁶². Dzięki zastosowaniu RDF możliwe jest dwustronne powiązanie informacji bibliograficznych z danymi różnych typów, w tym filmami, obrazami, dokumentami tekstowymi itp. Umożliwia to odsyłanie użytkowników od metadanych do konkretnych obiektów cyfrowych, ale również od tych obiektów cyfrowych do konkretnych metadanych bibliograficznych⁶³. Rozszerzenie formatu RDF stanowi format OWL (*Web Ontology Language*), który w porównaniu z RDF jest bardziej rozbudowany, posiada zasobniejszy słownik oraz mocniejszą składnię.

Dzięki konwersji danych zapisanych w formacie MARC na RDF/OWL, która jest stosowana przez liczne biblioteki na całym świecie, znacznie wzrasta dostępność tych danych w Internecie oraz łatwość ich wzajemnego powiązywania⁶⁴. W celu określenia szczegółowej ścieżki

i in., *The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles*, „Computational Materials Science” 2015, vol. 97, s. 209–215.

59 E. Miller, *An introduction to the Resource Description Framework*, „D-Lib Magazine” 1998, vol. 4, iss. 5, s. 14–25; A. Sprochi, *Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model*, „International Information and Library Review” 2016, vol. 48, iss. 2, s. 129–136.

60 <https://www.w3.org/TR/rdf-schema/> [dostęp: 17.11.2022].

61 T. Berners-Lee, J. Hendler, O. Lassila, *The semantic web*, „Scientific American” 2001, vol. 284, iss. 5, s. 34–43.

62 A. Zygmunt, J. Koźlak, *Metody efektywnego zarządzania wiedzą reprezentowaną w postaci ontologii*, w: *Strategie informatyzacji i zarządzanie wiedzą*, red. Zdzisław Szyjewski, Jerzy S. Nowak, Janusz K. Grabara, Warszawa 2004, s. 373–383.

63 D. Siwecka, *Najnowsze kierunki w udostępnianiu danych bibliograficznych...*, wyd. cyt.

64 <https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> [dostęp: 17.11.2022].

przejścia/konwersji dla formatu MARC do technik opartych na Linked Data została zainicjowana przez Bibliotekę Kongresu inicjatywa BIBFRAME (*Bibliographic Framework*)⁶⁵.

Dla plików RDF stosowany jest dedykowany język zapytań i protokół SPARQL (*SPARQL Protocol And RDF Query Language*)⁶⁶. Z punktu widzenia użytkowników szczególnie istotne są tzw. punkty końcowe (*SPARQL endpoints*), które umożliwiają przesyłanie zapytań do grafowej bazy danych. Grafowe bazy danych skupiają się na magazynowaniu i pobieraniu relacji, w przeciwieństwie do relacyjnych, które skupiają się na magazynowaniu i pobieraniu rzeczywistych encji. W grafowych bazach danych graf stanowi reprezentację danych, reprezentacja ta składa się z dwóch oddzielnych elementów – wierzchołków i krawędzi. Wierzchołki stanowią reprezentację koncepcji lub encji w danych, natomiast krawędzie – reprezentację relacji lub połączeń pomiędzy dwoma wierzchołkami (elementami danych)⁶⁷.

Ontologie

Jednym z istotnych elementów koncepcji sieci semantycznych jest ontologia⁶⁸. Określenie zakresu znaczeniowego pojęcia ontologii jest bardzo skomplikowane i istnieje szereg jej definicji⁶⁹. Thomas R. Gruber zaproponował definicję ontologii jako formalnej, jawnej specyfikacji wspólnej konceptualizacji, lecz bez precyzowania samej konceptualizacji⁷⁰. Przedstawił również pięć kryteriów projektowania ontologii⁷¹. Najważniejszym celem ontologii w odniesieniu do wiedzy jest dzielenie się zasobami wiedzy dziedzinowej w taki sposób, aby była ona zrozumiała zarówno dla ludzi, jak i komputerów⁷². Ontologię powinna charakteryzować czytelność dla maszyn cyfrowych, co oznacza, że musi być ona wyrażana w języku formalnym, którego składnia i semantyka są precyzyjnie określone. Ontologia konkretnej dziedziny powinna się charakteryzować kompletnością i całościowym ujęciem, nie powinno być w niej bytów ukrytych lub niejawnych. Wiedza, którą obejmuje ontologia, powinna być powszechnie uznawana, np. przez środowisko specjalistów z tej dziedziny. Ontologia stanowi model abstrakcyjny określonej dziedziny, który identyfikuje istotne elementy tej dziedziny. Dostępna w Internecie ontologia z określonej dziedziny może stanowić podstawę do tworzenia

65 <https://www.loc.gov/bibframe/> [dostęp: 17.11.2022].

66 <https://www.w3.org/TR/rdf-sparql-protocol/> [dostęp: 16.11.2022].

67 D. Gosnell, M. Broecheler, *Dane grafowe w praktyce. Jak technologie grafowe ułatwiają rozwiązywanie złożonych problemów*, Gliwice 2021.

68 T. Berners-Lee, J. Hendler, O. Lassila, *The semantic web*, „Scientific American” 2001, vol. 284, iss. 5, s. 34–43.

69 W. Gliński, *Ontologie. Próba uporządkowania terminologicznego chaosu*, w: *Od informacji naukowej do technologii społeczeństwa informacyjnego: praca zbiorowa pod redakcją Barbary Sosińskiej-Kalaty i Marii Przystek-Samokowej przy współpracy Andrzeja Skrzypczaka*, Warszawa 2005, s. 163–175.

70 T. R. Gruber, *A translation approach to portable ontology specifications*, „Knowledge Acquisition” 1993, vol. 5, iss. 2, s. 199–220.

71 T. R. Gruber, *Toward principles for the design of ontologies used for knowledge sharing?*, „International Journal of Human-Computer Studies” 1995, vol. 43, iss. 5–6, s. 907–928.

72 W. Gliński, *Ontologie...*, wyd. cyt.

w pełni współpracujących ze sobą systemów informatycznych, często wytworzonych przez zupełnie niezależne podmioty. Ta cecha jest szczególnie istotna dla realizacji idei sieci semantycznych⁷³. Istnieje szereg metod i metodologii tworzenia ontologii. Należy również zwrócić uwagę, że jedną dziedzinę może modelować wiele ontologii, lecz na różne sposoby. Ontologie mogą być budowane od postaw lub przy wykorzystaniu już istniejących ontologii⁷⁴. Wśród typów klasyfikacji ontologii można wyróżnić klasyfikację pod względem formalizacji. Uwzględniając to kryterium, ontologie dzielą się na:

- ontologie nieformalne,
- ontologie formalne.

Z punktu widzenia otwartej wymiany danych większą rolę odgrywają ontologie formalne. W ramach udostępniania danych bibliograficznych największą rolę odgrywają ontologie wyrażone trójkami RDF oraz, wyrażone w dialektach logiki opisowej, ontologie OWL.

Innym podziałem ontologii jest podział ze względu na rolę, którą odgrywają w systemie informatycznym:

- ontologie wysokiego poziomu (np. Dublin Core, DOLCE), które zawierają pojęcia bardzo ogólne, które występują powszechnie. Ujednolicenie tych pojęć jest niezbędne z punktu widzenia pełnej współpracy aplikacji w sieciach semantycznych,
- ontologie dziedzinowe, które zawierają pojęcia i relacje specyficzne dla określonej dziedziny, przy wykorzystaniu pojęć i relacji z ontologii wysokiego poziomu,
- ontologie aplikacyjne, które są związane z konkretnymi zastosowaniami w konkretnym systemie informatycznym. Najczęściej ontologie aplikacyjne opierają się na ontologiach dziedzinowych oraz pośrednio na ontologiach wysokiego poziomu. Najczęściej ontologie aplikacyjne są wykorzystywane wyłącznie na potrzeby aplikacji, dla której zostały opracowane⁷⁵.

Linked Open Data

W celu propagowania jak największej dostępności danych oraz ich powiązań powstała idea LOD (*Linked Open Data*), w której założono uwolnienie i powiązanie danych w Internecie w oparciu o format RDF/OWL. LOD stanowi otwarte środowisko, w którym możliwe jest eksplorowanie, tworzenie i łączenie danych w całym Internecie. Zgodnie z tą ideą dane powinny być udostępnione w taki sposób, aby możliwe było ich wykorzystanie bez żadnych ograni-

73 K. Goczyła, *Ontologie w systemach informatycznych*, Warszawa 2011, s. 25–26.

74 W. Gliński, *Wybrane metodologie i metody budowania ontologii*, w: *Informacja w sieci: problemy, metody, technologie*, praca zbiorowa pod redakcją Barbary Sosińskiej-Kalaty, Ewy Chuchro i Włodzimierza Daszewskiego, Warszawa 2006, s. 157–207.

75 K. Goczyła, *Ontologie w systemach informatycznych*, wyd. cyt., s. 26–30.

czeń⁷⁶. Powinno być możliwe pełne wykorzystanie, modyfikowanie, a także rozpowszechnianie danych w dowolnym celu, jedynie pod warunkiem uznania autorstwa lub udostępnienia na tej samej licencji⁷⁷.

W ramach koncepcji otwartych danych zostały zaproponowane w 2006 roku przez Tima Bernersa-Lee⁷⁸ cztery zasady:

1. Stosowanie URI (*Uniform Resource Identifier*) jako nazw dla zasobów.
2. Stosowanie protokołów HTTP w celu lokalizacji identyfikatorów URI.
3. Stosowane odniesienia do URI muszą dostarczać informacji o zasobie za pomocą standardów RDF, SPARQL.
4. Stosowanie powiązań do innych URI⁷⁹.

Aby możliwa była realizacja tej koncepcji, podmioty udostępniające dane powinny:

- udostępniać dane o zasobach na zasadach otwartych licencji,
- udostępniać dane w sposób, który pozwoli na realizację odpowiednich powiązań i przetwarzanie maszynowe, bardzo dobrze sprawdza się format RDF,
- udostępniać zasoby, z których każdy powinien posiadać unikalny, stały identyfikator URI,
- udostępniać dane online przy wykorzystaniu protokołu internetowego HTTP⁸⁰.

Selekcja metadanych

Jako najważniejszy element identyfikujący dany obiekt, zarówno fizyczny, jak i cyfrowy można wskazać jego tytuł. Jest on definiowany (standard RDA, FRBR) jako słowo, znak lub grupę słów i znaków, pod którymi dzieło jest znane. Zgodnie z przepisami RDA główny tytuł dzieła jest elementem obowiązkowym.

Należy dążyć, aby obiekt przed digitalizacją posiadał już pełny zakres przypisanych danych – kompletny opis bibliograficzny. Niejednokrotnie zdarza się jednak, że digitalizacja jeszcze nieskatalogowanych obiektów jest realizowana w bardzo wysokim tempie i utworzenie pełnych opisów bibliograficznych dla każdego z nich spowodowałoby nadmiernie

76 A. Iwaniak, I. Kaczmarek, M. Strzelecki, J. Łukowicz, *Publikowanie danych przestrzennych jako Linked Open Data*, „Roczniki Geomatyki” 2014, t. 12, z. 1 (63), s. 67–79.

77 <https://opendefinition.org/> [dostęp: 17.11.2022].

78 <https://www.w3.org/DesignIssues/LinkedData> [dostęp: 19.11.2022].

79 P. Ciccurese, M. Ocana, L. J. Garcia Castro i in., *An open annotation ontology for science on web 3.0*, „Journal of Biomedical Semantics”. 2011, vol. 2, iss. 2. <https://link.springer.com/content/pdf/10.1186/2041-1480-2-S2-S4.pdf> [dostęp: 16.11.2022].

80 G. Alemu, B. Stevens, P. Ross, J. Chandler, *Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models*, „New Library World” 2012, vol. 113, iss. 11, s. 549–570.

spowolnienie tempa tego procesu. W takich sytuacjach dopuszcza się utworzenie uproszczonego opisu bibliograficznego, który jednak zawsze musi zawierać szereg niezbędnych do identyfikacji danego obiektu danych, natomiast pozostałe dane powinny być uzupełnione jak najszybciej po wykonaniu prac digitalizacyjnych.

Na to, jakie metadane powinny zostać wybrane, wpływ może mieć wiele czynników. Częściowo wybór metadanych jest determinowany specyfiką instytucji – m.in. rodzajem gromadzonych przez nią obiektów. Czynniki, na które należałoby również zwrócić uwagę, to np. dostosowanie opisów do obiektów fizycznych i cyfrowych, cele digitalizacji, a także sposób przechowywania metadanych.

Pośród podstawowych elementów opisu bibliograficznego należy wymienić:

- tytuł,
- autora/twórcę,
- adres wydawniczy, w tym miejsce wydania, produkcji lub dystrybucji, nazwę wydawcy, producenta lub dystrybutora, datę wydania, produkcji lub dystrybucji,
- opis fizyczny,
- numer inwentarzowy/sygnatura⁸¹.

Powyższy zestaw danych, o ile tylko wskazane informacje są możliwe do ustalenia, powinien towarzyszyć każdemu obiektowi fizycznemu przed procesem digitalizacji.

Na potrzeby współpracy innych instytucji z Biblioteką Narodową opracowano poniższą tabelę wskazującą informacje, które muszą zostać zawarte w przekazywanym zestawie metadanych, odpowiednio dla grup obiektów.

81 J. Kalinowski, *Podstawowe elementy opisu dzieła...*, wyd. cyt.

Minimalny zestaw metadanych opisowych

Opis	MARC21	książki	rękopisy	mapy / atlasy	grafika / rysunki	fotografie / pocztówki	nuty	czasopisma / prasa	druki ulotne	artykuły
Autor/Współautor	100, 700 (110, 710, 111, 711)	x	x	x	x	x	x		x	x
Redaktor/instytucja sprawcza	700, 710							x		
Tytuł	245	x	x	x	x	x	x	x	x	x
Wydanie	250a	x		x	x	x	x	x	x	
Częstotliwość	310a							x		
Data powstania dokumentu	008:11-14	x	x	x	x	x	x	x	x	
Data (przedział)	045, 046	x	x	x	x	x	x	x	x	x
Adres wydawniczy	260	x	x	x	x	x	x	x	x	x
Opis fizyczny	300abce	x	x	x	x	x	x	x	x	x
Kraj	008 znak 15-17, pole 044	x	x	x	x	x	x	x	x	x
Język	pole 041adh lub pole 008 znak 35-37	x	x	x	x	x	x	x	x	x
Numer wydawniczy (ISBN/ISSN/ISMN)	(020/022/024)	x		x			x	x	x	x
Sygnatura lub numer inwentarza (jeśli identyfikuje egzemplarz)	różne pola	x	x	x	x	x	x	x	x	
Prawa (domena publiczna, licencja itp.)		x	x	x	x	x	x	x	x	x
Źródło		x	x	x	x	x	x	x	x	x
Opis/uwagi	pola z obszaru 500	x	x	x	x	x	x	x	x	x
Hasła przedmiotowe	610, 655	x	x	x	x	x	x	x	x	x
Forma	380	x	x	x	x	x	x	x	x	x
Gatunek/rodzaj	655	x	x	x	x	x	x	x	x	x
Grupa odbiorców	385	x	x	x	x	x	x	x	x	x
Przynależność kulturowa	386	x	x	x	x	x	x	x	x	x
Zapis muzyczny	254						x			
Kartograficzne dane matematyczne	255			x						
Zawiera się w:	773									x
Iteracja (oznaczenie numeru czasopisma)								x		

W przypadku opisu archiwalnego pośród podstawowych elementów, które powinien zawierać, należy wymienić⁸²:

- sygnaturę (kod kraju, skrót lub kod archiwum, zespołu),
- tytuł (nazwę) i język,
- twórcę (aktotwórcę),
- daty skrajne,
- rozmiary opisywanego obiektu, nośnik, charakterystykę stanu fizycznego,
- poziom opisu.

Powyższe wymagane elementy są zgodne z ISAD(G), a także stanowią zasadnicze elementy z punktu widzenia międzynarodowej wymiany informacji. Ze względu na bardzo szybko postępującą digitalizację zasobów archiwalnych i trudności w nadążeniu z powstawaniem opisów przed digitalizacją, dopuszcza się digitalizację materiałów archiwalnych posiadających jedynie szczątkowy opis (np. wyłącznie sygnatura, numer ewidencyjny). Uzupełnienie pozostałych elementów opisu powinno nastąpić jak najszybciej po wykonaniu digitalizacji⁸³.

Docelowo metadane zdigitalizowanego materiału archiwalnego powinny zawierać⁸⁴:

- źródło pochodzenia obiektu archiwalnego,
- datę i sposób powstania,
- cel i funkcję, którą pełnił obiekt przed znalezieniem się w archiwum,
- wskazówki, dotyczące odczytywania,
- warunki udostępniania,
- historię zmian od czasu wytworzenia,
- historię migracji dokonywanych w celu ich zabezpieczenia,
- powiązania z oprogramowaniem służącym do zarządzania materiałami cyfrowymi.

Bardzo ważne jest, aby wszystkie metadane obiektów fizycznych przypisane do odpowiadających im obiektów cyfrowych były ze sobą zgodne. Konieczne jest, aby istniała pełna jednoznaczność, która kopia cyfrowa odnosi się do którego obiektu fizycznego. Porządek w tym zakresie pozwoli również na uniknięcie powstawania dubletów opisów.

82 J. Bednarek, *Zakres standaryzacji opisu archiwalnego w „Cyfrowym Archiwum” Instytutu Pamięci Narodowej*, w: *Standaryzacja opisu archiwalnego* pod red. Jerzego Bednarka i Pawła Perzyny, Warszawa; Łódź 2016, s. 131–147; ISAD(G): Międzynarodowy Standard Opisu Archiwalnego. Część ogólna. Wersja 2. https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_PL.pdf [dostęp: 18.11.2022].

83 P. Perzyna, *Standaryzacja opisu archiwalnego a digitalizacja*, w: *Standaryzacja opisu archiwalnego...*, wyd. cyt., s. 17–41.

84 A. Sobczak, *Między tradycyjnym a wirtualnym archiwum. Narodziny cyfrowej tożsamości archiwów w Niemczech*, Szczecin; Warszawa 2014, s. 53–54.

Podsumowanie

W procesie digitalizacji obiektów fizycznych należy zadbać o odpowiedni zestaw metadanych, który będzie przypisany zarówno do digitalizowanego obiektu fizycznego, jak i do obiektu cyfrowego. Zestaw metadanych (w przypadku bibliotek jest to opis bibliograficzny) powinien pozwalać na jednoznaczną identyfikację wszystkich obiektów. W prawidłowym doborze metadanych pomocne są opracowane standardy, z których wybrane zostały przedstawione w niniejszym rozdziale. Metadane dla obiektu cyfrowego stanowią, uzupełnione o dane charakterystyczne dla obiektów cyfrowych, metadane powstałe dla obiektu fizycznego. W praktyce to właśnie metadane dla obiektów fizycznych powinny być podstawą do utworzenia metadanych dla ich cyfrowych odpowiedników. Procesy związane z tworzeniem oraz gromadzeniem metadanych obiektów cyfrowych powinny odbywać się w sposób jak najbardziej zautomatyzowany.



Transport między magazynem a pracownią digitalizacji

Justyna Król-Próba

Opisane poniżej wytyczne dotyczą wszystkich etapów transportu obiektów, zarówno wewnątrz instytucji, jak i pomiędzy instytucjami. Stosowane procedury powinny uwzględniać przede wszystkim kwestie konserwatorskie, zapewnić bezpieczeństwo, ułatwić pakowanie i rozpakowywanie, przemieszczanie obiektów i postępowanie z nimi. W przypadku outsourcingu wszystkie istotne aspekty transportu zbiorów poza instytucję do siedziby wykonawcy digitalizacji powinny zostać uzgodnione i regulowane zapisami umowy zawartej między stronami. Wskazane w rozdziale 17 *Digitalizacji piśmiennictwa* procedury i wzory raportów można stosować zgodnie z potrzebami własnej instytucji.

Kwestie konserwatorskie

Transportowanie obiektów niesie ze sobą ryzyko ich uszkodzenia. Transportować należy obiekty w stanie stabilnym, zabezpieczone w odpowiednich opakowaniach ochronnych oraz skrzyniach jeśli zbiory przemieszczane są między instytucjami. Stan obiektów powinien być oceniony przez konserwatora przed realizacją transportu i digitalizacji. Przemieszczanie zbiorów powinno przebiegać z uwzględnieniem zaleceń konserwatorskich, a w wypadku digitalizacji obiektów cennych – konserwator powinien wchodzić w skład zespołu. Dla projektów o mniejszym znaczeniu wskazana jest konsultacja konserwatorska z zewnętrznym ekspertem.

Konserwator powinien ustalić stan i konieczność konserwacji każdego obiektu i określić wszystkie niezbędne kryteria konserwatorskie przed skierowaniem obiektu do digitalizacji. Chodzi zwłaszcza o zapewnienie, aby stan obiektów nie pogorszył się pod wpływem warunków otoczenia i pozwalał na ich digitalizację. Konserwator powinien wskazać zalecany czas digitalizacji i sposób jej prowadzenia. Jednocześnie – biorąc pod uwagę, że gwałtowne i częste zmiany warunków otoczenia są szkodliwe dla obiektów – należy zapewnić odpowiedni poziom temperatury i wilgotności, a także czas na aklimatyzację transportowanych obiektów.

Aspekty prawne

W zależności od kategorii zbiorów (biblioteczne, muzealne) zasady przewozu i ochrony zbiorów regulują różne akty prawne. W przypadku części kategorii zbiorów, brak jest ogólnie obowiązujących regulacji, które określają zasady transportu. W takich przypadkach zaleca się stosowanie zapisów aktów prawnych określających zasady transportu dla kategorii o podobnym charakterze.

Zasady ochrony transportów zbiorów regulują:

- Ustawa o ochronie osób i mienia z 22 sierpnia 1997 r. (Dz. U. 2021 poz. 1995),
- Rozporządzenie Ministra Kultury i Dziedzictwa Narodowego z dnia 15 maja 2008 r. w sprawie warunków, sposobu i trybu przenoszenia muzealiów (Dz. U. 2008 nr 91 poz. 569),
- Obwieszczenie Ministra Kultury nr 77 w sprawie nadania Regulaminu ogólnych warunków i trybu wykonywania ochrony muzeów i innych jednostek organizacyjnych podlegających obowiązkowej ochronie, podległych, podporządkowanych lub nadzorowanych przez Ministra Kultury (Dz. Urz. Min. Kult. z 31 października 2002 nr 5 poz. 77),
- Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 7 września 2010 r. w sprawie wymagań, jakim powinna odpowiadać ochrona wartości pieniężnych przechowywanych i transportowanych przez przedsiębiorców i inne jednostki organizacyjne (Dz. U. 2016 poz. 793 z późn. zm.),
- Rozporządzenie Ministra Kultury i Dziedzictwa Narodowego z dnia 19 września 2014 r. w sprawie zabezpieczenia zbiorów w muzeach przed pożarem, kradzieżą i innym niebezpieczeństwem grożącym zniszczeniem lub utratą (Dz. U. 2014 poz. 1240).

Ubezpieczenie

Wskazane jest ubezpieczenie obiektów cennych na czas transportu i digitalizacji prowadzonej poza instytucją macierzystą. Należy odpowiednio wcześniej ustalić kwestie zakresu odpowiedzialności ubezpieczyciela (szczególnie istotne są sprawy związane z nieprawidłowym postępowaniem z obiektem, niedbałością lub brakiem należytej staranności), kwoty ubezpieczenia, która ze stron pokrywa jego koszty. W umowie można rozważyć wskazanie sądu lub organu arbitrażowego, do którego strony zwrócą się, gdy zajdzie potrzeba rozstrzygnięcia kwestii spornych.

Zabezpieczenie mienia

Mając na uwadze bezpieczeństwo zbiorów, należy z wyprzedzeniem zorganizować bezpieczny transport i ochronę obiektów. Transportowi zbiorów poza instytucję powinien towarzyszyć pracownik merytoryczny lub konserwator – w zależności od charakteru i potrzeb zbiorów. Osoba ta powinna być przeszkolona i upoważniona do podejmowania podczas transportu decyzji i działań, mających na celu ochronę zbiorów przed zniszczeniem. Do towarzyszenia jej aż do miejsca digitalizacji zobowiązany jest przedstawiciel instytucji przejmującej obiekty.

Obiekty o szczególnej wartości zabytkowej, cenne dla dziedzictwa narodowego lub wysokiej wartości materialnej powinny podlegać ochronie fizycznej realizowanej przez pracowników ochrony podczas wszystkich czynności wykonywanych w procesie digitalizacji, gdy znajdują się poza odpowiednio zabezpieczonym technicznie miejscem przechowywania (np. zamkniętym magazynem lub szafą pancerną). W ramach prowadzonego projektu digitalizacji powinny zostać określone procedury ochrony digitalizowanych zbiorów.

Środki transportu

Zbiory biblioteczne i archiwalne są wrażliwe na wpływ temperatury i wilgotności względnej powietrza. Pewne typy obiektów mogą być wrażliwe na drgania. Przewożenie zbiorów należy realizować środkiem transportu zapewniającym stałe warunki klimatyczne i, jeśli to konieczne, tłumienie drgań. Skrzynie w przestrzeni ładunkowej muszą być unieruchomione. Ochrona transportu powinna być zorganizowana zgodnie z przepisami podanymi powyżej w punkcie *Aspekty prawne*.

Opakowania i skrzynie

Zalecenia dotyczące opakowań ochronnych dla zbiorów bibliotecznych i archiwalnych podają Polska Norma ISO 11799 oraz Raport Techniczny ISO/TR 19814:2021-12 (Załącznik A). Obiekty powinny być transportowane w opakowaniach ochronnych dostosowanych do ich rozmiaru i ciężaru. Obiekty nie powinny przesuwac się w opakowaniu. Na opakowaniu powinno znajdować się oznaczenie pozwalające zidentyfikować obiekt.

Skrzynie transportowe powinny być skonstruowane w taki sposób, aby zapewnić zbiorom ochronę przed zniszczeniem lub uszkodzeniem, a także być dostosowane rozmiarem do zbiorów i pozwalać na ułożenie obiektów w pozycji horyzontalnej. Obiekty w opakowaniach ochronnych trzeba układać w skrzyniach tak, aby największe i najcięższe egzemplarze znajdowały się na dnie. Należy zwrócić uwagę na liczbę i ciężar ułożonych na sobie obiektów: zbyt duży ciężar i nacisk może spowodować zgniecenie i uszkodzenie znajdujących się najniżej

opakowań z obiektami. W razie wątpliwości co do bezpiecznego i prawidłowego ułożenia zbiorów w skrzyni trzeba skonsultować się z konserwatorem. Puste przestrzenie w skrzyni należy wypełnić materiałami, które zapobiegają przesuwaniu się opakowań z obiektami podczas przemieszczania.

Skrzynie powinny mieć oznaczenia informujące o prawidłowym sposobie transportowania (np. etykiety logistyczne: „Góra”, „Unikać wilgoci”, „Nie piętrować”).

Etapy transportu

Przygotowanie kwestii formalnych i akceptacja

W przypadku transportu zbiorów w związku z ich przemieszczeniem pomiędzy instytucjami, powinna zostać sporządzona umowa określająca obowiązki każdej ze stron, w tym wymagania techniczne w celu zapewnienia warunków klimatycznych i szczegóły ochrony transportu.

Kierownik jednostki, lub osoba przez niego upoważniona, udziela pozwolenia na transport danego obiektu (po konsultacji konserwatorskiej). W przypadku transportu obiektów o szczególnej wartości zabytkowej lub materialnej, decyzja o transporcie powinna być podejmowana po określeniu warunków transportu przez osoby do tego wyznaczone (opiekun zbiorów, osoba odpowiedzialna za ochronę zbiorów, konserwator).

Pobieranie obiektów z magazynu

Na podstawie listy obiektów przeznaczonych do digitalizacji, opiekun zbiorów lub upoważniony pracownik pobiera z magazynu obiekty w stabilnym stanie zachowania, przygotowane do digitalizacji pod względem merytorycznym i konserwatorskim oraz zabezpieczone w odpowiednich opakowaniach ochronnych. Pobranie egzemplarzy powinno być udokumentowane i przeprowadzone zgodnie z procedurami obowiązującymi w instytucji. Należy przy tym uwzględnić możliwości transportu. Pobiera się egzemplarze, które w danym cyklu zostaną przetransportowane do pracowni digitalizacyjnej, zgodnie z wcześniej nadanymi priorytetami wyznaczającymi kolejność skanowania i zamówieniem reprografa.

Pobranie egzemplarzy powinno być udokumentowane i przeprowadzone zgodnie z procedurami obowiązującymi w instytucji.

Protokół

Przekazanie zbiorów do digitalizacji następuje na podstawie protokołu zdawczo-odbiorczego. Wzór protokołu oraz procedura przekazania zbiorów między komórkami w obrębie instytucji powinny być regulowane stosownymi przepisami wewnętrznymi. W przypadku transportu zbiorów między instytucjami pracownik merytoryczny przygotowuje protokół zdawczo-odbiorczy zawierający:

- miejsce i datę,
- nazwę i dane kontaktowe instytucji wydającej zbiory,
- nazwę komórki organizacyjnej, z której pochodzą zbiory,
- nazwę i dane kontaktowe wykonawcy digitalizacji,
- numer umowy podpisanej z wykonawcą digitalizacji,
- spis obiektów z ich numerami inwentarza lub sygnaturami, opisem, liczbą kart oraz liczbą opakowań i skrzyń, w których są umieszczone,
- imię, nazwisko i dane kontaktowe kuriera,
- podpisy osób upoważnionych ze strony instytucji oraz wykonawcy digitalizacji.

Oryginał protokołu podpisywany jest:

- w siedzibie instytucji przy wydaniu obiektów,
- przy odbiorze obiektów w miejscu digitalizacji,
- w miejscu digitalizacji przy wydaniu obiektów,
- przy odbiorze obiektów w siedzibie instytucji.

Na każdym z tych etapów przedstawiciele instytucji oraz wykonawcy digitalizacji sprawdzają liczbę obiektów i ich zgodność z protokołem.

Zbiory zabytkowe i podlegające wieczystemu przechowaniu powinny być objęte procedurą kontroli stanu zachowania na każdym etapie ich przekazania, potwierdzonego podpisaniem protokołu zdawczo-odbiorczego. Wyniki kontroli należy odnotować w formie raportu (przykładowy formularz raportu znajduje się poniżej), na arkuszu sporządzonym dla każdego obiektu. Arkusze raportu stanu zachowania powinny być przygotowane razem z protokołem zdawczo-odbiorczym, dla wszystkich obiektów wymienionych w tym protokole, i być integralną częścią protokołu. Procedura kontroli stanu zachowania i uzupełnienia raportów może znacząco wydłużyć realizację transportu zbiorów. Należy wziąć to pod uwagę podczas planowania harmonogramów.

Umieszczenie obiektów w skrzyniach transportowych powinno odbyć się zgodnie z procedurami obowiązującymi w instytucji lub określonymi w umowie z uwzględnieniem zaleceń konserwatorskich.

Pakowanie

Pakuje się lub zleca pakowanie obiektów zgodnie z zaleceniami konserwatorskimi oraz procedurami przyjętymi w instytucji lub określonymi w umowie.

Transport wewnątrz budynku

Zbiory powinny być przemieszczane w obrębie budynku zgodnie z procedurami obowiązującymi w instytucji.

Transport na zewnątrz

Zamówienie na transport i ochronę składa opiekun zbiorów na formularzu przewidzianym w Obwieszczeniu lub Zarządzeniu Ministra Kultury, w zależności od kategorii zbiorów.

Szef ochrony lub osoba odpowiedzialna za ochronę instytucji – na podstawie zadeklarowanej wartości obiektów – przydziela, w przypadku, gdy w instytucji funkcjonuje wewnętrzna służba ochrony, odpowiednią liczbę konwojentów oraz określa rodzaj ich uzbrojenia lub zleca ochronę transportu przedsiębiorcy, który uzyskał koncesję na prowadzenie działalności gospodarczej w zakresie usług ochrony osób i mienia, w uzgodnieniu z opiekunem zbiorów podejmuje też decyzję o czasie i środku transportu.

W przypadkach określonych w obwieszczeniu lub zarządzeniu ministra kultury, kierownik instytucji lub osoba przez niego upoważniona przekazuje do instytucji kultury (obecnie Narodowego Instytutu Muzealnictwa i Ochrony Zbiorów) informacje zawierającą dane o organizacji transportu.

Załadunek skrzyń ze zbiorami do środka transportu i ich późniejszy rozładunek powinien odbywać się pod zadaszeniem, z udziałem przeszkolonego personelu wyposażonego w niezbędne narzędzia i sprzęt BHP. Kurier biorący udział w transporcie powinien być przygotowany do sporządzenia notatki w przypadku zdarzenia stanowiącego zagrożenie dla obiektów i opisanie ich stanu, jeśli ulegnie pogorszeniu.

Samochód powinien być wyposażony w udogodnienia gwarantujące obiektom bezpieczeństwo i odpowiednie warunki transportu, zgodne z zaleceniami konserwatorskimi.

Odbiór w miejscu digitalizacji

Otworzenie skrzyń i wypakowanie obiektów powinno się odbyć w wydzielonym pomieszczeniu o odpowiednich zabezpieczeniach i warunkach klimatycznych. Wyjęcie obiektów ze skrzyń transportowych powinno odbyć się zgodnie z procedurami obowiązującymi w instytucji lub określonymi w umowie z uwzględnieniem zaleceń konserwatorskich.

Droga powrotna

Droga powrotna zbiorów po digitalizacji odbywa się analogicznie i zgodnie z opisanymi procedurami.

Przykładowy raport stanu zachowania

Pieczęć instytucji:		RAPORT STANU ZACHOWANIA Informacja o projekcie digitalizacyjnym, numer umowy podpisanej z wykonawcą digitalizacji:		
Nazwa i dane kontaktowe instytucji wydającej zbiory: Nazwa komórki organizacyjnej, z której pochodzi obiekt:		Numer katalogowy/sygnatura obiektu: Opis obiektu (autor, tytuł, technika wykonania, data wykonania/wydania): Wymiary obiektu i liczba kart:		
Data sporządzenia raportu:				
Opis i lokalizacja uszkodzeń:				
Imię, nazwisko i podpis osoby sporządzającej raport:				



Pomieszczenia

Tomasz Gruszkowski

Bezpieczeństwo zbiorów

Troska o bezpieczeństwo wszystkich materiałów jest podstawową zasadą obowiązującą także w czasie prowadzenia digitalizacji. Należy wziąć pod uwagę konieczność przestrzegania zasad i procedur dotyczących obiektów o szczególnej wartości zabytkowej dla dziedzictwa narodowego lub wysokiej wartości materialnej, co może wymagać zapewnienia ochrony fizycznej, realizowanej przez pracowników ochrony podczas wszystkich czynności wykonywanych w procesie digitalizacji poza odpowiednio zabezpieczonym technicznie miejscem przechowywania.

Istotne jest odpowiednie dokumentowanie przekazywania obiektów między poszczególnymi osobami, a także przenoszenia ich do innych pomieszczeń.

Wyposażenie pracowni oraz magazynu podręcznego w środki zabezpieczeń odpowiednie do wartości zbiorów: drzwi, zamki (zwrócenie uwagi na trwałość ścian), systemy kontroli dostępu (od najprostszycy typu domofon ograniczający niekontrolowane przemieszczanie się ludzi, przez systemy rejestracji wejść i czujki ruchu, po systemy monitoringu wizyjnego itp.).

Wprawdzie zasadą powinno być jak najszybsze przeprowadzanie digitalizacji i zwracanie obiektów do magazynu, w którym są na stałe przechowywane, konieczne może być wyposażenie pracowni lub magazynu tymczasowego w zamykane szafy lub sejfy.

W wypadku większych instytucji i licznych pracowni, warto ograniczyć dostęp do nich i do magazynu tymczasowego dodatkowymi drzwiami z systemem kontroli wejść.

Niezbędne jest ograniczenie do minimum możliwości wystąpienia zalań i zawilgoceń czy pożaru.

Pomieszczenia pracowni digitalizacji

Magazyn tymczasowy i pracownia digitalizacji należy ulokować w pomieszczeniach znajdujących się możliwie najbliżej magazynu stałego przechowywania, na tym samym piętrze. Jeśli pomieszczenia znajdują się na innym piętrze niż magazyn, należy zapewnić możliwość korzystania z windy do transportu zbiorów. Ciągi komunikacyjne i windy powinny umożliwiać bezpieczne i wygodne przemieszczanie zbiorów na wózkach ręcznych. Ani pracownia, ani magazyn nie mogą stanowić części miejsca pracy służącego do realizacji innych zadań.

W przypadku, gdy obiekty będą transportowane z innych lokalizacji, ważny jest wygodny dostęp do środków transportu (rampy itp.), jednak – ze względu na bezpieczeństwo klimatyczne i fizyczne – bez możliwości wejścia bezpośrednio z zewnątrz.

Wielkość pracowni powinna być zgodna z ogólnymi przepisami BHP i uwzględniać konieczność umieszczenia w niej blatów roboczych, z których będą korzystał skanerzyści podczas pracy z obiektami. Liczbę i wielkość blatów należy dostosować do wielkości obiektów, co umożliwi ich wygodne i bezpieczne układanie. Ponadto trzeba zapewnić odpowiednio dużą przestrzeń wokół skanerów czy stanowisk aparatów fotograficznych. Konserwator powinien na bieżąco weryfikować potrzebę zapewnienia skanerzystom odpowiedniego wyposażenia pomocniczego do pracy z obiektami (np.: kołyska, kliny, obciążniki, podpórki, szpatułki do przytrzymywania kart itp.). Na podstawie rekomendacji konserwatora zakup wyposażenia realizuje właściwa komórka.

Magazyn tymczasowy należy wyposażać w meble o odpowiedniej do przechowywania zbiorów konstrukcji i optymalnym rozmiarze. Wytrzymałość regałów powinna być dostosowana do ciężaru obiektów. Obiekty nie mogą wystawać poza krawędzie półek. Przy ustawianiu mebli w pomieszczeniu należy pamiętać o zachowaniu odpowiednich odległości, pozwalających na bezpieczne przemieszczanie się osób, wygodne wyjmowanie i układanie obiektów oraz usuwanie kurzu z podłóg i półek.

W pracowni nie wolno wykonywać czynności mogących zagrozić obiektom (gotowanie wody, jedzenie), dlatego należy zapewnić pracownikom dostęp do pobliskiego pomieszczenia socjalnego.

Wnętrza

Warunki klimatyczne: temperatura i wilgotność względna powietrza

Warunki klimatyczne w pomieszczeniach, w których podczas projektu digitalizacji znajdują się cenne zbiory (np. magazyn tymczasowy, pracownia digitalizacyjna), powinny być jak najbardziej zbliżone do warunków panujących w magazynie macierzystym. Wymagania dotyczące optymalnych warunków długoterminowego przechowywania materiałów bibliotecznych i archiwalnych oraz zarządzania klimatem określają: norma PN-ISO 11799:2006P i Raport Techniczny PKN-ISO/TR 19815:2022-03P.

Przy podejmowaniu decyzji o parametrach warunków klimatycznych w pomieszczeniach należy przestrzegać norm BHP dotyczących minimalnej temperatury pracy oraz zaleceń konserwatorskich opartych na analizie stanu i budowy technologicznej obiektów.

Konieczne jest monitorowanie temperatury i wilgotności względnej powietrza w pomieszczeniach, w których znajdują się zbiory. Należy w tym celu wyposażyć pomieszczenia w termohigrometry i regularnie kontrolować ich wskazania. Osoby pracujące w pomieszczeniu powinny systematycznie sprawdzać, czy odczyty termohigrometrów mieszczą się w zaleconych przez konserwatora zakresach. W przypadku stwierdzenia rozbieżności należy zawiadomić konserwatora, który zaproponuje metody i środki optymalizacji klimatu. Jeśli pomieszczenia nie są objęte systemem pełnej klimatyzacji można w nich zastosować przenośne klimatyzatory. Poziom wilgotności względnej powietrza można modyfikować, stosując nawilżacze lub osuszacze powietrza.

Należy dążyć do uzyskania stabilnych warunków klimatycznych przez minimalizowanie fluktuacji temperatury i wilgotności względnej powietrza. Częste wahania parametrów klimatu stanowią zagrożenie dla zbiorów bibliotecznych i archiwalnych, mogą być przyczyną deformacji, a nawet uszkodzeń mechanicznych podłoży.

Urządzenia digitalizacyjne i reprograficzne nie powinny nagrzewać się i wydzielać gazów (np. ozonu).

Zanieczyszczenia powietrza

Czystość powietrza jest istotna dla trwałości zbiorów bibliotecznych i archiwalnych, a szczególnie dla obiektów wrażliwych jak np. zabytkowe fotografie czy negatywy. Należy dbać o utrzymanie czystości w pomieszczeniach, aby zapobiec akumulacji cząsteczek kurzu i sporów pleśni na opakowaniach ochronnych i meblach. Powinno się unikać malowania ścian farbami wydzielającymi lotne związki organiczne (LOC) i zachować po malowaniu odpowiedni czas przed wprowadzeniem zbiorów do pomieszczenia. Wymagania dotyczące dopuszczalnych stężeń zanieczyszczeń powietrza atmosferycznego podczas długoterminowego przechowywania materiałów archiwalnych i bibliotecznych podaje norma PN-ISO 11799:2006P.

Oświetlenie.

Szkodliwość dla zbiorów

Promieniowanie elektromagnetyczne we wszystkich zakresach – światło widzialne, promieniowanie ultrafioletowe i podczerwone – ma niszczący wpływ na zbiory biblioteczne i archiwalne. Ekspozycja na światło może prowadzić do zmian chemicznych (np. blaknięcie partii barwnych) lub fizycznych (kruchłość i pękanie materiałów organicznych). Zniszczenia wywołane przez światło są kumulatywne i nieodwracalne.

Materiały organiczne mają zróżnicowaną światłotrwałość, czyli odporność na zmiany wywołane ekspozycją na światło. Zabytkowe zbiory biblioteczne i archiwalne zaliczają się zwykle do kategorii wrażliwej. Szczególnie szkodliwe dla nich jest wysokoenergetyczne promieniowanie UV, które katalizuje i przyspiesza niszczące procesy chemiczne w materiałach takich jak papier, pergamin, skóra, atramenty, warstwy malarskie, tkaniny. Najlepszą ochroną przed negatywnymi skutkami ekspozycji na światło jest opracowanie strategii i procedur ograniczających intensywność i czas ekspozycji do minimum niezbędnego dla prawidłowej i bezpiecznej realizacji poszczególnych zadań w projekcie digitalizacji.

Podczas digitalizacji zbiorów zabytkowych, cennych i unikatowych należy:

- odciąć dostęp światła słonecznego (np. zasłonięcie okien tkaniną wyciemniającą lub roletami)
- ograniczyć natężenie i czas działania oświetlenia sztucznego poprzez:
 - ▶ usunięcie zbędnych źródła światła,
 - ▶ zastosowanie oświetlenia aktywowanego ruchem;
- wyeliminować promieniowanie UV poprzez:
 - ▶ zastosowanie lamp LED,
 - ▶ wykonanie pomiarów udziału promieniowania UV w pozostałych rodzajach źródeł światła i zastosowanie w nich filtrów eliminujących UV o długości fali krótszej niż 400nm;
- wyeliminować promieniowanie IR poprzez zastosowanie lamp LED.

Obiekty należy przechowywać w ciemności: przykrywać, umieszczać w pudłach ochronnych lub szafach.

Informacje o niszczącym wpływie światła na obiekty biblioteczne i archiwalne oraz wskazówki dotyczące strategii zapobiegania szkodliwym skutkom działania światła zawarte są w Raporcie Technicznym PKN-ISO/TR 19815:2022-03P.

Zakłócenia skanowania

Oświetlenie w pomieszczeniach, w których odbywa się digitalizacja i ocena jakości jej produktów powinno być maksymalnie stabilne w czasie pracy. Oznacza to, że należy zminimalizować wpływ (bezpośredniego) światła słonecznego poprzez zainstalowanie na oknach pracowni neutralnych kolorystycznie (szarych albo czarnych) rolet lub żaluzji. Dodatkową ich zaletą będzie ograniczenie nagrzewania pomieszczeń.

Niedopuszczalne jest umieszczenie większej liczby urządzeń wykorzystujących światło błyskowe w jednym pomieszczeniu, należy także dostatecznie daleko od siebie rozstawić skanery stosujące inne oświetlenie lub oddzielić urządzenia przegrodami wysłaniającymi źródła światła, tak, żeby na powierzchnie skanowane nie padało światło innych urządzeń.

Zalecany kolorem ścian, sufitu oraz posadzki jest kolor szary. Ściany i sufit powinno się malować matową szarą farbą akrylową, winylową lub emulsją akrylową o współczynniku odbicia światła mniejszym niż 60%. Farby nie powinny być łatwopalne, nie powinny też emitować pyłów, przyciągać ani zatrzymywać kurzu.

Oświetlenie otoczenia w pracowniach powinno mieć temperaturę barwową około 5000K (zbliżoną do światła słonecznego), a jego CRI (wskaźnik oddania barw) powinien być wyższy niż 90. Zaleca się oświetlenie o natężeniu nieprzekraczającym 35–70 luksów (ISO 3664).

Ergonomia pracy

Przy stosowaniu powyższych zaleceń, należy jednocześnie brać pod uwagę komfort pracy operatorów. Nie sprzyja mu długotrwała praca w ciemnym pomieszczeniu, dlatego konieczne jest stosowanie zasad dotyczących długości czasu pracy i przerw – zgodne z przepisami BHP (Dz. U. z 1998 r., Nr 148, poz. 973).



Działania konserwatorskie

Justyna Król-Próba

Prace konserwatorskie przed i w czasie digitalizacji

W projekcie digitalizacji – obejmującym zabytkowe, cenne, unikatowe zbiory piśmiennictwa podlegające wieczystemu przechowaniu – konieczny jest udział konserwatorów specjalizujących się w opiece nad obiektami bibliotecznymi i archiwalnymi.

Fundament właściwej ochrony zbiorów stanowi opracowanie i wdrożenie pisemnych procedur, określających zasady postępowania i zakres odpowiedzialności pracowników całej instytucji. Jeśli w planie ochrony zbiorów brakuje procedur związanych z digitalizacją, należy je opracować w porozumieniu z konserwatorem i przeszkolić personel zaangażowany w projekt.

Konsultacja konserwatorska jest wskazana także wtedy, gdy instytucja nie zatrudnia konserwatora na stałe. Można wówczas ocenić, które z przewidzianych do digitalizacji obiektów powinny być traktowane w szczególnie ostrożny sposób oraz czy nie ma wśród nich obiektów narażonych podczas realizacji projektu na uszkodzenia. Projekty, w których digitalizacji mają podlegać obiekty mniej ważne z punktu widzenia ochrony dziedzictwa kulturowego, wymagają standardowo stosowanego sposobu traktowania zbiorów – tak, aby ich stan nie uległ pogorszeniu.

Poniżej przedstawiono modelowe rozwiązania, które należy wprowadzić w wypadku projektów obejmujących cenne zbiory.

Doradztwo podczas planowania działań operacyjnych i na pozostałych etapach projektu

Konserwator formułuje rekomendacje w zakresie:

- szacowania nakładu czasu i środków koniecznych do konserwatorskiego przygotowania obiektów do digitalizacji;
- opracowania procedur i instrukcji dotyczących pracy z obiektami;
- wyposażenia pomieszczeń tj. pracowni reprograficznej i magazynu tymczasowego w urządzenia monitorujące klimat, meble do przechowywania zbiorów, wyposażenie pomocnicze do pracy z obiektami.

Osoby pracujące ze zbiorami w trakcie realizacji projektu powinny mieć możliwość bieżących konsultacji z konserwatorem w razie pojawienia się jakichkolwiek wątpliwości lub problemów związanych z ochroną konserwatorską skanowanych obiektów.

Ocena stanu zachowania i budowy technologicznej zbiorów, formułowanie zaleceń dotyczących konserwatorskiej ochrony zbiorów w trakcie projektu

Przeprowadzenie przez konserwatora przeglądów stanu zachowania i budowy technologicznej zbiorów sprzyja płynnej realizacji zadań w projekcie digitalizacji. Im wcześniej i wnikliwiej zostaną wykonane przeglądy, tym łatwiej będzie zidentyfikować potencjalne problemy, przewidzieć środki zaradcze i dostosować procedury lub harmonogramy.

Jeśli przeglądy stanu zachowania zbiorów nie mogą być wykonane przed rozpoczęciem projektu, konieczne będzie ich organizowanie równoległe z procesem digitalizacji. Pracownicy merytoryczni selekcjonujący obiekty do skanowania dokonują wówczas pierwszej oceny stanu obiektów. Egzemplarze w bardzo dobrym stanie można przekazać bezpośrednio do digitalizacji. Opinia konserwatora jest niezbędna w przypadkach, w których stan zachowania lub budowa technologiczna obiektu budzi wątpliwości i może stanowić przeszkodę dla jego bezpiecznego i skutecznego zeskanowania. Przeglądając zbiory, konserwator powinien formułować wnioski i zalecenia dotyczące bezpiecznej digitalizacji każdego egzemplarza. Wskazówki i uwagi powinny być dołączone do obiektu w formie pisemnej lub w systemie komputerowym. Na podstawie rekomendacji konserwatora należy ustalić dalszą ścieżkę postępowania z obiektem (np. przekazanie do konserwacji, digitalizacja z asystą konserwatorską, digitalizacja z użyciem konkretnego rodzaju urządzeń).

Na podstawie obowiązujących przepisów i norm, planu ochrony zbiorów instytucji i wiedzy specjalistycznej konserwator powinien określić optymalne parametry klimatyczne i oświetleniowe w pomieszczeniach, w których znajdować się będą zbiory. Stworzenie właściwych warunków jest kluczowe dla minimalizowania ryzyka uszkodzenia obiektów. Konserwator może zaproponować zastosowanie właściwych urządzeń, rozwiązań lub procedur dla osiągnięcia oczekiwanych efektów. Powyższe parametry powinny być zweryfikowane przed rozpoczęciem digitalizacji i kontrolowane w trakcie prac. Więcej informacji na ten temat można znaleźć w rozdziale *Pomieszczenia*.

Prowadzenie szkoleń dla reprografów, fotografów, skanerzystów i osób zajmujących się transportem zbiorów

Konserwatorzy powinni zrealizować szkolenia personelu przed rozpoczęciem skanowania i – stosownie do potrzeb – w trakcie projektu prowadzić kolejne, uzupełniając materiał szkoleniowy o zagadnienia pojawiające się podczas realizacji zadań. Zakres szkoleń powinien być dostosowany do charakteru zbiorów i potrzeb uczestników, ich zadań w projekcie i doświadczenia.

Wykonywanie niezbędnych zabiegów konserwatorskich

Część zbiorów, ze względu na stan zachowania, może wymagać poddania zabiegom konserwatorskim przed rozpoczęciem digitalizacji w dwóch celach:

- ochrony obiektu przed dalszymi uszkodzeniami lub rozproszeniem (np. reperacje rozdarć, połączenie oderwanych fragmentów kart);
- poprawienia czytelności przesłoniętych fragmentów informacji (np. rozprostowanie zagięć kart).

Należy mieć na uwadze, że prace konserwatorskie mogą być czasochłonne i trwać nawet kilka miesięcy. Dlatego niezwykle istotne jest jak najwcześniejsze wykonanie przeglądów pozwalających ocenić stan kolekcji i potrzeby związane z jej konserwacją.

Asystowanie podczas skanowania

Obiekty o szczególnej wartości zabytkowej, cenne egzemplarze posiadające unikatową formę bądź nietypową budowę, mogą wymagać asysty konserwatora w trakcie procesu digitalizacji. Konserwator pomaga wówczas skanerzyście w bezpiecznej manipulacji obiektem.

Przekazywanie do skanowania

Obiekty składuje się w magazynie tymczasowym. Stąd są wydawane skanerzystom lub dostarczane do pracowni digitalizacyjnych – wraz z kartami skanowania, jeśli takie są tworzone. Transport z i do magazynu podręcznego po zakończeniu skanowania winien odbywać się zgodnie z procedurami obowiązującymi w instytucji. Skanerzyści potwierdzają otrzymanie obiektów na dokumentach przekazania lub w systemie.



Techniczne aspekty digitalizacji

Tomasz Gruszkowski

Sprzęt

Wybór urządzeń najbardziej odpowiadających potrzebom projektu jest bardzo trudny ze względu na to, że istnieje wielka różnorodność skanerów, aparatów cyfrowych oraz urządzeń łączących ich cechy, a także sprzętu wspomagającego. Jakość i przydatność tych urządzeń należy oceniać, posługując się wieloma uniwersalnymi parametrami i biorąc pod uwagę specyficzne cechy funkcjonalne istotne w digitalizacji wybranych obiektów. Ważność tych parametrów zależy od cech obiektów, które mają zostać odwzorowane oraz od wartości tych obiektów i ich podatności na uszkodzenia. Ocena taka musi zawsze odnosić się do celu digitalizacji i uwzględniać aspekty ekonomiczne.

Niezbędne jest wzięcie pod uwagę co najmniej następujących czynników:

- cechy fizyczne obiektów wybranych do digitalizacji – ich format, liczba obiektów o podobnym formacie, stan zachowania obiektów, uwzględniając konieczność zapewnienia bezpieczeństwa zbiorów,
- jakość oczekiwana od urządzeń, która wynika ze zdefiniowanego celu digitalizacji: niezbędna rozdzielczość przestrzenna, jaką urządzenie zapewnia, wymagana wierność oddania kolorów, właściwości stosowanego oświetlenia, brak wad optycznych itp.,
- szybkość działania i ergonomia pracy urządzenia (biorąc pod uwagę oprogramowanie),
- koszt – samego urządzenia, koszt pracy operatora, a także koszt utrzymania (serwisowania) sprzętu.

Dwie główne grupy urządzeń digitalizacyjnych to aparaty cyfrowe i skanery. W związku z nieustannymi zmianami technologicznymi (zarówno w zakresie sprzętu, jak i oprogramowania) nie wydaje się sensowne formułowanie jednoznacznych zaleceń dotyczących rodzaju sprzętu.

Należy jednak podkreślić, że w przypadku prowadzenia postępowań zakupowych, niezbędne jest zapewnienie kontroli parametrów podawanych przez producentów. Wiele z nich można obiektywnie skontrolować za pomocą pomiarów wykonywanych z użyciem wzorni-

ków, część może zostać oceniona wyłącznie wizualnie przez specjalistów z zakresu fotografii reprodukcyjnej.

Format

Format urządzenia określa maksymalną wielkość obiektu, jaki można przy jego pomocy zdigitalizować bez konieczności łączenia plików fragmentarycznych, co jest pracochłonne i czasochłonne. Jeśli w planowanym projekcie bardzo duże obiekty występują wyjątkowo, uzasadniony może być zakup urządzenia o mniejszym formacie powierzchni skanowania i wykonanie odwzorowań takich obiektów poza instytucją lub przez łączenie części skanów.

W przypadku aparatów cyfrowych wielkość pola skanowania zależy od odległości matrycy od obiektu i długości ogniskowej użytego obiektywu. Należy zwrócić uwagę na to, że o ile w wypadku skanerów możliwa do uzyskania rozdzielczość jest ustalona przez producenta (w niektórych modelach można ją zmieniać ustawieniami urządzenia), to w przypadku aparatów cyfrowych rozdzielczość jest zmienna dla różnych wielkości pola roboczego i zależna także od wielkości matrycy w pikselach.

Rozdzielczość optyczna (częstotliwość próbkowania)

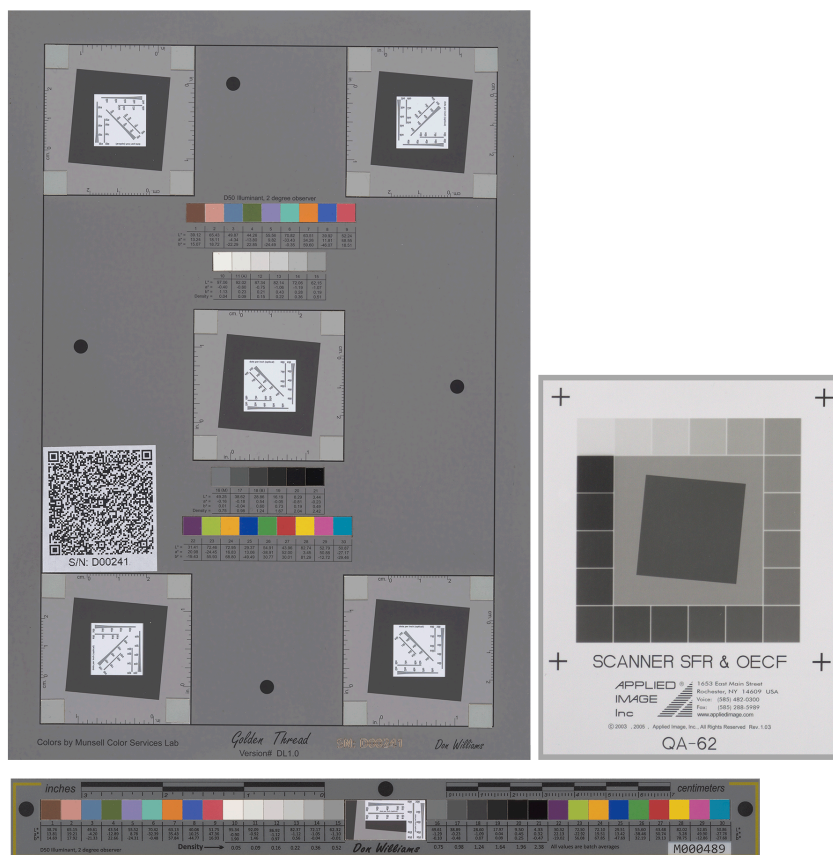
Rozdzielczość optyczna urządzenia określa jego zdolność do oddawania szczegółów odwzorowanego obiektu, potocznie zaś – to maksymalna liczba punktów na jednostkę długości obiektu, jaką może zarejestrować to urządzenie. Ogólnie mówiąc – im większa rozdzielczość, tym drobniejsze szczegóły mogą zostać odwzorowane. Jednakże wysoka rozdzielczość nie jest gwarancją wysokiej jakości zapisywanej informacji.

Rozdzielczość przestrzenna urządzenia digitalizującego wyrażana jest liczbą punktów na cal – ppi (ang. pixels per inch, czasem także punkty na cal – dpi) w postaci dwu liczb (np. 600x1200 ppi) oznaczających rozdzielczość w dwu wymiarach. Gdy podano tylko jedną liczbę, najczęściej oznacza to, że obie wartości są identyczne, powinno się jednak to sprawdzać. W niektórych typach skanerów uzyskiwana wartość rozdzielczości zależy od wielkości skanowanego obszaru.

Istotne jest, czy podawana rozdzielczość wynika z fizycznej charakterystyki urządzenia, czy też uzyskiwana jest poprzez interpolację przeprowadzaną w oprogramowaniu. Nie obrazuje ona możliwości odwzorowania rzeczywistych szczegółów obiektu, ale możliwość sztucznego podwyższenia rozdzielczości uzyskiwanego obrazu przez odpowiednie programy działające w urządzeniu skanującym lub w komputerze. Wstawiają one między dwa punkty obrazu kolejne punkty, których jasność i kolor są wyliczane (interpolowane) na podstawie punktów sąsiednich. Nie należy brać pod uwagę tego parametru, lecz omówioną wyżej rozdzielczość fizyczną. Często dane dotyczące rozdzielczości, jako parametr istotny marketingowo, są zawyżane przez producentów (np. poprzez podanie wartości interpolowanej), co jest mylące, ale przede wszystkim przekształcenia takie nie służą zwiększeniu wierności odwzorowania oryginału.

Nie ma bezpośredniego związku między deklarowanymi liczbami opisującymi rozdzielczość sprzętu a jego zdolnością odwzorowania szczegółów obiektu, na którą wpływa: jakość optyki, wielkość pikseli układu światłoczułego, stabilność i ewentualne drgania układu, po-

ziom szumu cyfrowego czy algorytmy stosowane przy przetwarzaniu danych. Uznaje się, że miarą realnej rozdzielczości urządzenia digitalizującego jest funkcja przenoszenia modulacji (MTF – Modulation Transfer Function), opisująca jego zdolność do oddawania kontrastu w zależności od rosnącej gęstości szczegółów. Do zmierzenia jej wartości niezbędne jest wykonanie cyfrowego odwzorowania wzorników o cechach pozwalających na dokonanie pomiarów skuteczności urządzenia digitalizującego pod tym względem, a następnie skorzystanie z oprogramowania pozwalającego na przeprowadzenie analizy pliku cyfrowego. Obok oprogramowania komercyjnego dostępne są strony internetowe pozwalające na poznanie wartości realnych rozdzielczości oraz skuteczności próbkowania urządzenia¹.

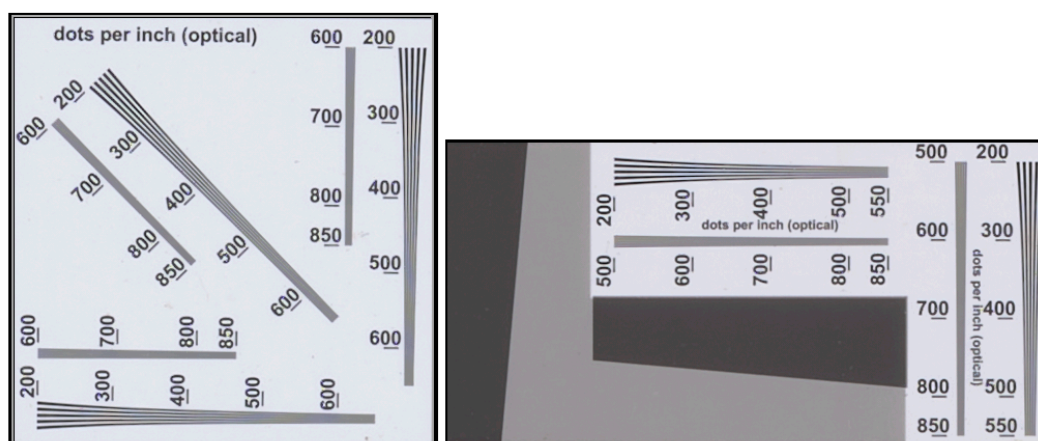


Il. 1.

Wzornik QA-62 pozwala na dokonanie analizy rozdzielczości (dzięki pomiarowi wartości pikseli znajdujących się na ukośnych krawędziach ciemniejszego kwadratu na jaśniejszym tle) i wartości funkcji konwersji optoelektronicznej (poprzez pomiary wartości szarych pól).

Wzorniki Golden Thread (większy – do pomiaru i kalibracji parametrów urządzenia i mniejszy wykorzystywany jako wzornik do poszczególnych sesji digitalizacji). Wzorniki te posiadają zarówno pola przeznaczone do pomiaru rozdzielczości, jak i pola szare i barwne przeznaczone do kontroli wierności barwnej.

1 Dwie nieodpłatne strony pozwalające, po zarejestrowaniu się, na analizę niektórych z wzorników tego typu to: <https://deltae.picturae.com> i <https://imagezebra.com> [dostęp: 15.12.2022].



Il.2.

Wzorniki tego typu oprócz kwadratów służących do analizy rozdzielczości posiadają także kliny pozwalające na wizualną kontrolę zdolności urządzenia do rejestracji szczegółów z zachowaniem kontrastu, jak na ilustracji 2 przedstawiającej fragmenty wzorników w skali 1:1.

Głębina bitowa (głębina kolorów)

Parametr ten opisuje, ile poziomów szarości lub kolorów urządzenie może zarejestrować i przetworzyć. Zalecana dla materiałów bibliotecznych głębina 8 bitów na składową oznacza, że każdy z pikseli (w trzech barwach: czerwonej, niebieskiej, zielonej) może przybierać 256 wartości, co daje ponad 16 milionów kolorów – wartość pozwalającą na uzyskanie dobrych jakościowo skanów, wystarczających do niemal wszystkich celów wyznaczanych w projektach. W wyjątkowych przypadkach można korzystać z urządzeń zapewniających jeszcze szerszy zakres pozyskiwanych i zapisywanych barw, głębinię 48-bitową, po 16 bitów na kanał barwny, co łączy się z potrzebą zapisu znacznie większych plików i używania innej przestrzeni barwnej (ECI-RGB lub ProPhoto).

Wierność barwna

Ze względu na to, że postrzeganie barw jest różne u różnych ludzi, opracowano obiektywny sposób pomiaru różnic między kolorami oryginału i reprodukcji cyfrowej w oparciu o pomiar wartości liczbowych barw w ramach zstandaryzowanej przestrzeni kolorów. Niezbędne do takiego pomiaru jest zastosowanie wzornika z polami o różnych (ale znanych) wartościach barwnych poszczególnych pól, zdigitalizowanie go i wyznaczenie wartości odchylenia. Miarą tych różnic jest wartość Delta E (ΔE) – różnica między wartościami pól oryginalnego wzornika koloru a wartościami odczytanymi z cyfrowej reprodukcji tegoż wzornika w przestrzeni kolorów CIE $L^*a^*b^*$. Jest ona wyznaczana zarówno dla poszczególnych pól wzornika, jak i jako średnia wartość dla wszystkich pól.

Do obliczenia tej wartości niezbędna jest znajomość wartości barwnych pól stosowanego wzornika (podana przez producenta lub zmierzona) i oprogramowanie pozwalające na zmierzenie i uśrednienie wartości odpowiednich pól reprodukcji cyfrowej².

2 Oprogramowanie komercyjne, a także dwie platformy internetowe wskazane w poprzednim przypisie.

Wartość średnia dla wszystkich pól ΔE poniżej 2, jak wynika z badań nad ludzką percepcją kolorów, jest niedostrzegalna lub stanowi „ledwo dostrzegalną różnicę”. Według najostrożniejszej wersji zaleceń FADGI³ wartość graniczna średniej różnicy pozwalająca na uznanie wierności reprodukcji wynosi właśnie 2, wyższe (gorsze) wartości odchylenia mogą być przyjmowane zależnie od założeń projektu lub uznawane za zbyt duże.



Il. 3.

Przykładowe wzorniki do pomiaru wierności barwnej zawierają od kilkudziesięciu do ponad stu pól i są produkowane z dbałością o utrzymanie wartości barwnej poszczególnych pól. Ich reprodukcje cyfrowe mogą być wykorzystane do kalibracji skanera lub aparatu fotograficznego (np. wzornik ColorChecker Digital SG), a zawierające mniejszą liczbę pól wzorniki typu ColorChecker Classic w różnych rozmiarach mogą być wykorzystywane jako wzornik towarzyszący poszczególnym sesjom digitalizacji.

Gęstość optyczna

Gęstość optyczna (dynamika) zapewniana przez urządzenie opisuje jego zdolność do prawidłowego oddawania tonów oryginału, szczególnie w jego najciemniejszych i najjaśniejszych częściach. Rejestrowanie przez urządzenie szczegółów ciemnych fragmentów bez zniekształceń powodowanych przez szum własny czujnika wskazuje na wysoką jakość urządzenia.

3 Dostępne w języku angielskim na stronie: <https://www.digitalizationguidelines.gov/guidelines/digitize-technical.html> [dostęp: 15.12.2022]

Profile kolorów

Aby móc osiągnąć poprawne i stabilne działanie sprzętu pod kątem reprodukcji barw, niezbędne jest regularne przeprowadzanie jego kalibracji, a następnie wytwarzanie odpowiadającemu mu profilu barwnego, zapisującego sposób odtwarzania przez urządzenie barw, który zapisywany jest w pliku i pozwala na poprawne mapowanie barw w systemie zarządzania kolorem, jeśli będzie on stosowany.

W wypadku skanerów kalibracja jest na ogół przewidziana przez producenta i pozwala na korektę jego ustawień na podstawie skanu wzorca (w optymalnym rozwiązaniu posiadającego zmierzone wartości co najmniej 100 pól, w innych rozwiązaniach wzorca zawierającego pola szare, czasem wyłącznie na podstawie białej karty). W wyniku kalibracji zmienia się sposób, w jaki urządzenie zapisuje kolory (odcienie szarości).

Po przeprowadzeniu kalibracji wskazane jest profilowanie urządzenia i zapisanie wyniku w profilu ICC. Profil taki nie zmienia wartości zapisywanych przez urządzenie kolorów, ale pozwala na przekazanie wiedzy o tej charakterystyce kolejnym urządzeniom (np. monitorom, drukarkom) mogącym działać w ramach systemów zarządzania kolorem i odpowiednie traktowanie zapisanych wartości.

Najczęściej używane w zastosowaniach związanych z reprodukowaniem dzieł należących do dziedzictwa kulturowego są następujące profile:

sRGB, Adobe RGB 1998, ECI-RGB, ProPhoto (dwa ostatnie szczególnie w przypadku stosowania głębi kolorów o liczbie 16 bitów na składową). W przypadku digitalizowania w odcieniach szarości zalecanym profilem ICC jest Gamma 2.2.

Szybkość i wydajność

Na szybkość skanowania wpływ mają rodzaj obiektów, ustawienia koloru, rozdzielczości, obszaru skanowania, a także czas poświęcony na przygotowanie i zdanie obiektu.

Dla projektów digitalizacyjnych kluczowym parametrem jest wydajność długoterminowa, określana jako liczba skanów możliwych do wykonania w pewnym okresie, np. dziennie lub miesięcznie. Wartość ta w praktyce nie ma jednoznacznego związku z czasem wykonywania pojedynczego skanu, gdyż oprócz powyżej wyliczonych czynników należy także brać pod uwagę walory ergonomiczne urządzenia i indywidualną wydajność operatora.

Niezawodność

Jakość urządzeń, wyznaczona przez parametry plików wytwarzanych w procesie digitalizacji, jest bardzo istotna przy wyborze aparatury, ale niezwykle ważnym aspektem jest określenie jej niezawodności – wynikającej na ogół z jakości wykonania. Jako że jest to parametr odgrywający rolę w całym okresie funkcjonowania urządzenia, można brać pod uwagę takie wskaźniki jak deklarowane przez producenta czy dostawcę MTBF (średni czas bezawaryjnej pracy) i MTTR (średni czas naprawy). Jednak w praktyce decydujące będą: długość okresu gwarancji, a także godziny i dni przyjmowania zgłoszeń o awarii oraz realny czas skutecznej reakcji serwisu.

Inne cechy urządzeń digitalizujących

Z rozmaitych względów wiele parametrów jest pomijanych w opisach technicznych urządzeń, niektóre są trudne do zmierzenia czy zaobserwowania w inny sposób. Należą do nich między innymi:

- wierność odzwierciedlenia rozmiarów obiektu (w obu wymiarach),
- nieobecność aberracji chromatycznej i zniekształceń geometrycznych,
- niewprowadzanie zafałszowania barwnego (związane z prawidłowym balansem bieli),
- szum cyfrowy,
- jednolitość oświetlenia przestrzeni roboczej skanera.

Wiele cech urządzeń jest istotnych w przypadku zbiorów cennych i wymagających delikatnego postępowania ze strony operatora. Trzeba brać pod uwagę (w niektórych przypadkach w konsultacji z konserwatorem) następujące właściwości maszyn:

- delikatność postępowania z obiektem (zwłaszcza cechy części ruchomych wchodzących w kontakt z obiektem – szyby dociskowe, szalki, na których spoczywa obiekt),
- sposób regulowania docisku szyby i możliwość zdjęcia (odsunięcia) szyby dociskowej,
- rodzaj stosowanego oświetlenia (ograniczenie promieniowania UV i IR),
- jasność oświetlenia i czas ekspozycji obiektu na światło,
- minimalny kąt otwarcia (książek i innych obiektów ciasno oprawnych),
- dopuszczalny ciężar całkowity obiektu.

Oprogramowanie

Obecnie w zasadzie wszyscy producenci sprzętu zapewniają oprogramowanie pozwalające na wykonywanie podstawowych działań prowadzących do zrealizowania serii poprawnych skanów/fotografii cyfrowych. Należy podkreślić, że warto dążyć do uzyskania obrazu o właściwych parametrach bez potrzeby lub przy możliwie niewielkiej ingerencji w jego wygląd w procesie postprodukcji – jest ona czasochłonna i może spowodować utratę informacji zawartej w oryginalnym pliku.

Oprogramowanie skanera czy aparatu powinno pozwalać na: kalibrację sprzętu, przeprowadzanie profilowania, dołączanie profili barwnych ICC, określenie wymiarów i miejsca powierzchni skanowanej, ustawianie parametrów sprzętu przy wykonywaniu zdjęcia/skanu, przycinanie i obracanie plików (to ostatnie w wielokrotnościach 90°, nie powodując utraty informacji), zapis plików w wybranym, zalecanym formacie (w razie potrzeby także wytworzenie formatów pochodnych, prezentacyjnych) z właściwą nazwą w odpowiednim miejscu w hierarchii folderów, w wybranej przestrzeni kolorów, dodawanie metadanych do skanowanych plików i obiektów, tworzenie profili zadaniowych i profili użytkownika.

W wypadku prowadzenia digitalizacji obiektów szczególnie trudnych do działania automatycznego, bez ingerencji operatora, takich jak dagerotypy, negatywy fotograficzne, przezrocza, niezbędne jest korzystanie z oprogramowania do obróbki graficznej plików rastrowych. W takim przypadku zalecenie dopuszczania się minimalnej ingerencji w pliki pochodzące z urządzenia digitalizującego nie może być stosowane, gdyż bez takiej ingerencji nie jest możliwe uzyskanie odwzorowania możliwie bliskiego oryginałowi.

Zalecenia

Wychodząc z założenia, że podstawowym celem digitalizacji jest możliwie wierne oddanie obecnej zawartości obiektu przy zagwarantowaniu jego bezpieczeństwa, a następnie długoterminowe zapewnienie dostępu do cyfrowych plików, zaleca się prowadzenie digitalizacji w sposób zapewniający wysoką jakość plików zapisywanych następnie w otwartych formatach, do których dostęp w przewidywalnej przyszłości nie będzie ograniczony zmianami technologii komputerowej (zarówno urządzeń, jak i oprogramowania) lub restrykcjami wprowadzanymi przez jej producentów (np. patenty czy restrykcyjne licencje).

Biblioteka Narodowa zaleca do zapisu plików wzorcowych (źródłowych) stosowanie formatu TIF (aktualnie w wersji 6.0) z opcjonalnym wykorzystaniem bezstratnej kompresji (np. LZW). Pojedynczy plik może zawierać jeden obraz rastrowy, będący wynikiem reprodukcji (części) obiektu za pomocą urządzenia digitalizacyjnego po ewentualnej konwersji z pliku pochodzącego bezpośrednio z urządzenia (pliku RAW).

Zalecenia dotyczące wybranych parametrów cyfrowych reprodukcji płaskich nieruchomych obiektów dziedzictwa kulturowego:

Wymagania minimalne	Grupa A	Grupa B	Grupa C	Grupa D	Grupa E	Grupa F	Grupa G
	teksty drukowane	teksty drukowane z ilustracjami	rysunki i grafiki monochromatyczne, rękopisy, odbitki fotograficzne cz/b	materiały fotograficzne: negatywy i przezrocza	mikrofilmy	malarstwo, odbitki fotograficzne barwne, małe obiekty muzealne	plakaty, duże mapy, duże obiekty muzealne
Rozdzielczość (ppi)	400	300	nie mniej niż 3000 pikseli na dłuższym wymiarze	nie mniej niż 3000 pikseli na dłuższym wymiarze	jak dla mikrofilmowanego oryginału w granicach przenoszenia jego cech przez mikrofilm	nie mniej niż 3000 pikseli na dłuższym wymiarze	300
Średnia wierność odwzorowania barwnego (ΔE_{2000})	≤ 3,5	≤ 3,5	≤ 3,5	–	–	≤ 3,5	≤ 3,5
Bity na piksel	8-bitowa skala szarości	8-bitowa skala szarości lub 8 bitów na kolor (24-bit RGB)	8-bitowa skala szarości lub 8 bitów na kolor (24-bit RGB)	8-bitowa skala szarości lub 8 bitów na kolor (24-bit RGB)	8-bitowa skala szarości	8 bitów na kolor (24-bit RGB)	8 bitów na kolor (24-bit RGB)
Profile barwne	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2	Adobe RGB 1998	Adobe RGB 1998

Wymagania zalecane	Grupa A	Grupa B	Grupa C	Grupa D	Grupa E	Grupa F	Grupa G
	teksty drukowane	teksty drukowane z ilustracjami	rysunki i grafiki monochromatyczne, rękopisy, odbitki fotograficzne cz/b	materiały fotograficzne: negatywy i przezrocza	mikrofilmy	malarstwo, odbitki fotograficzne barwne, małe obiekty muzealne	plakaty, duże mapy, duże obiekty muzealne
Rozdzielczość (ppi)	400	400	nie mniej niż 5000 pikseli na dłuższym wymiarze	nie mniej niż 5000 pikseli na dłuższym wymiarze	jak dla mikrofilmowanego oryginału w granicach przenoszenia jego cech przez mikrofilm	nie mniej niż 5000 pikseli na dłuższym wymiarze	300
Średnia wierność odwzorowania barwnego (ΔE_{2000})	≤ 2,0	≤ 2,0	≤ 2,0	–	–	≤ 2,0	≤ 2,0
Bity na piksel	8-bitowa skala szarości	8/16-bitowa skala szarości; 8 lub 16 bitów na kolor (48-bit RGB)	8/16-bitowa skala szarości; 8 lub 16 bitów na kolor (48-bit RGB)	8/16-bitowa skala szarości; 8 lub 16 bitów na kolor (48-bit RGB)	8-bitowa skala szarości	8 lub 16 bitów na kolor (24/48-bit RGB)	8 lub 16 bitów na kolor (24/48-bit RGB)
Profile barwne	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2 lub Adobe RGB 1998	Gray Gamma 2.2	Adobe RGB 1998 lub lepszy	Adobe RGB 1998 lub lepszy

Grupa A

książki, gazety, czasopisma bez ilustracji; rysunki; mapy monochromatyczne; rysunki techniczne (z wyraźnym kontrastem); nuty; dokumenty urzędowe (normy, monitory, rozporządzenia itp.); maszynopisy; prace licencjackie, magisterskie, doktorskie

Grupa B

książki, gazety, czasopisma z ilustracjami w odcieniach szarości (ewentualnie pojedyncze strony jako uzupełnienie grupy A); rysunki techniczne, druki, mapy, nuty (z niskim kontrastem)

Grupa C

odbitki fotograficzne czarno-białe; rysunki i grafiki czarno-białe lub w odcieniach szarości

Grupa D

materiały fotograficzne (tylko negatywy i przezrocza), negatywy szklane o typowej wielkości

Grupa E

mikrofilmy i mikrofiszce

Grupa F

odbitki fotograficzne barwne, ew. monochromatyczne (sepia itp.); rysunki i grafiki kolorowe; mapy; obrazy; miedzioryty; drzeworyty; rękopisy; inkunabuły; starodruki (gdy użycie koloru jest uzasadnione szczególnymi cechami dokumentów, to także książki, czasopisma); obiekty muzealne małe i średniej wielkości

Grupa G

mapy wielkoformatowe; atlasy; plakaty; duże obrazy; duże obiekty muzealne; malarstwo naścienne

Powyższa tabela odwołuje się do zaleceń sformułowanych w publikacji *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*, pod redakcją Grzegorza Płoszajskiego, Warszawa 2008, Biblioteka Główna Politechniki Warszawskiej, rozdział 7 *Propozycje zaleceń* i *Technical Guidelines for Digitizing Cultural Heritage Materials* (Wytyczne dotyczące digitalizacji obiektów dziedzictwa kulturowego), wydanych przez amerykańską FADGI. Zmiany wynikają zarówno z rozwoju technologii (m.in. rozwoju sprzętu digitalizacyjnego i znacznego spadku kosztów związanych z przechowywaniem danych), jak i doświadczeń w prowadzeniu dużych projektów digitalizacyjnych.

Zalecenia należy regularnie weryfikować i w miarę potrzeby korygować.



Metody zapisu i przechowywania danych

Dariusz Paradowski

Dziedzina zapisu i przechowywania danych jest bardzo rozległa i wymaga specjalistycznej wiedzy z zakresu technologii informatycznych. Powstało na ten temat wiele fachowych opracowań, których analiza i podsumowanie zdecydowanie wykracza poza cel tego rozdziału. Wyboru konkretnych rozwiązań należy zawsze dokonywać w konsultacji z odpowiednimi specjalistami, a następnie takim specjalistom powierzać wdrożenie projektu. W tej części publikacji zebrano tylko podstawowe, poglądowe informacje ułatwiające orientację osobom poszukującym rozwiązań i wykonawców.

Cele i cechy

Zapis danych następuje w wielu etapach procesu digitalizacji, każdy z nich ma inną dynamikę przepływu informacji, a zatem inne wymagania wobec parametrów infrastruktury realizującej zapis. W zależności od skali i oczekiwanej wydajności może być konieczne przygotowanie specjalizowanych urządzeń dla każdego z tych etapów lub może się okazać, że wszystkie realizuje się w miarę dobrze za pomocą jednego urządzenia.

Poniższa tabela obrazuje typowe wymagania wobec zapisu w głównych elementach cyklu życia obiektu cyfrowego.

Tabela 1

	Przyjmowanie plików z zawartością obiektów (tzw. skany) z urządzeń digitalizujących	Przetwarzanie – wytwarzanie pochodnych: PDF, OCR itp.	Udostępnianie – np. biblioteka cyfrowa udostępniająca zasoby online	Archiwizacja – długotrwałe przechowywanie
Szybkość zapisu i odczytu	duża – warto szybko odblokować komputer obsługujący skanowanie, jednak, jeśli skala jest niewielka, traci to na znaczeniu	bardzo duża – z reguły oczekuje się wyników przetwarzania (np. miniaturki) niemal natychmiast	średnia – z reguły przedmiotem prezentacji są pliki o zmniejszonym rozmiarze stąd szybkość mniejsza niż dla plików źródłowych	mała – archiwizacja może zachodzić asynchronicznie (całą dobę, w weekendy), ważne, by umożliwiła archiwizację miesięcznego wpływu w tym samym okresie
Trwałość zapisu	duża – trzeba zabezpieczyć pliki operatora, który mógł po przekazaniu usunąć je ze swojego komputera	niska – w razie potrzeby pliki można ponownie pobrać z miejsca, w którym zostały zapisane przy pobieraniu lub zarchiwizowane	średnia – w razie potrzeby pliki można pobrać z archiwum (może być konieczne ponowne wytworzenie niektórych plików pochodnych)	bardzo duża – archiwizacja jest podstawą bezpieczeństwa zgromadzonych zasobów
Pojemność	mała – rzędu dobowej produkcji	mała – rzędu dobowej produkcji	duża – cały zasób prezentacyjny	bardzo duża – cały zasób źródłowy, często również prezentacyjny

Warto pamiętać, że szybkość może być rozumiana na dwa sposoby: jako przepustowość, np. gdy potrzeba w ciągu kilku godzin przesłać wiele terabajtów danych, lub jako czas dostępu, gdy oczekiwany jest natychmiastowy dostęp do licznych niewielkich plików (jak w typowym przypadku udostępniania). Kolejne parametry odnoszące się do powyższej tabeli, to rozróżnienie wymagań zapisu i odczytu. Takie zagadnienia w projekcie małej skali nie spowodują zwykle ekonomicznie uzasadnionego różnicowania infrastruktury, zaś w dużym projekcie powinny być szczegółowo przeanalizowane, by precyzyjnie dobrać rozwiązania do konkretnych warunków.

Rozeznanie potrzeb

W tab. 1 użyto względnych określeń wielkości, by pokazać proporcje między wartościami. Zasadniczym zadaniem osób przygotowujących projekt digitalizacyjny jest możliwie precyzyjne oszacowanie zarówno potrzeb bieżących, jak i zmian przewidywalnych w kolejnych latach.

Punktem wyjścia powinna być konkretna decyzja co do wielkości zbioru obiektów, które są planowane do digitalizacji (będzie to niezbędne nie tylko dla przygotowania podsystemów zapisu, ale i wielu innych elementów procesu digitalizacji).

Jeśli zbiór jest na tyle różnorodny, że ma podlegać digitalizacji z różnymi parametrami, należy wyznaczyć licznosc obiektów w każdej z grup (np. ile plakatów będzie skanowanych w niższej rozdzielczości, map w wyższej, a mikrofilmów w wysokiej, ale bez kolorów), a następnie średnią liczbę skanów przypadającą na jeden obiekt, w przypadku piśmiennictwa – z reguły odpowiadającą liczbie stron (plus okładki, ew. grzbiety).

Dla każdej z tych grup trzeba określić: planowaną rozdzielczość digitalizacji w ppi, średnią powierzchnię strony (wysokość razy szerokość w calach) i dokładność odwzorowania kolorów (w bitach). Mnożąc te trzy wartości uzyskuje się dla danej grupy szacowaną sumaryczną wielkość plików źródłowych, uzyskiwanych z urządzenia digitalizującego, przy założeniu, że są to pliki typu TIFF (co jest standardem w wielu projektach digitalizacyjnych).

Jeśli w ramach planowanej digitalizacji zdecydowano pobierać z urządzenia digitalizującego inne typy plików, należy eksperymentalnie ustalić średnią proporcję wielkości uzyskiwanych plików do teoretycznej wielkości plików tiff (pliki JPG będą mniejsze od plików TIFF). Jeśli urządzenie digitalizujące pozwala zapisywać zarówno tiff, jak i wybrany inny standard można tę proporcję wyznaczyć wprost, porównując wielkości uzyskiwanych plików.

Sumaryczna wielkość plików źródłowych (w miarę możliwości wyznaczona dla pierwszego i kolejnych lat projektu), uzyskana po uwzględnieniu typu plików i zsumowaniu dla wszystkich grup zbiorów, jest punktem wyjścia do planowania systemów zapisu i przechowywania we współpracy ze specjalistami IT.

W kolejnych etapach trzeba będzie ustalić, jakiego typu pliki zostaną wybrane do prezentacji, a jakie do archiwizacji oraz przewidzieć, jakim zainteresowaniem czytelników będą cieszyły się zdigitalizowane zbiory, np. ile obiektów dziennie będzie przeglądanych, co pozwoli wyznaczyć parametry podsystemów zapisu dla etapów wymienionych w tab. 1.

Infrastruktura

Przesyłanie

Tylko w amatorskich i eksperymentalnych warunkach można sobie wyobrazić realizowanie wszystkich elementów procesu digitalizacji w jednym urządzeniu. Realia profesjonalnego podejścia dyktują użycie specjalizowanych urządzeń do zapisu i przechowywania. Urządzenia te mogą być podłączone różnymi sposobami, które będą miały równie duży wpływ na parametry wydajnościowe (i stabilność), co same urządzenia zapisujące.

- a) Sieć LAN (Local Area Network), czyli popularna sieć wewnętrzna – jeśli jest dobrze zaplanowana, to wystarcza w większości zastosowań,
- b) Sieć SAN (Storage Area Network) – sieć dedykowana do przesyłania danych do urządzeń zapisujących, generalnie wydajniejsza i stabilniejsza od LAN, ale również droższa przy zakupie i utrzymaniu,

- c) Sieć WAN (Wide Area Network) – w uproszczeniu oznacza przesyłanie przez Internet danych do zapisu w urządzeniach pracujących poza budynkiem, gdzie następuje digitalizacja, często w pomieszczeniach i/lub urządzeniach należących do innych instytucji lub firm. Osiągnięcie podobnej wydajności będzie bardziej kosztowne niż w wyżej wymienionych metodach, jednak głównym atutem jest możliwość pokonywania praktycznie dowolnej odległości.

Oczywiście w każdej z wymienionych metod wydajność i stabilność zależy od wielu czynników, takich jak klasa i stan urządzeń oraz okablowania.

Nośniki i systemy zapisu

Nośnik w ścisłym rozumieniu to materiał, w którym następuje zapis danych. Jednak by nośnik spełniał swoją rolę, z reguły potrzebuje urządzenia zapisującego, tworząc razem system zapisu. W tym podrozdziale zestawione są systemy zapisu czasem składające się z wielu urządzeń, które warto brać pod uwagę przy zapisie i przechowywaniu danych w digitalizacji.

- a) Dyski optyczne lub optomagnetyczne (pochodne standardów CD i MD) są medium, które (mimo udanych prób stworzenia bardzo trwałego nośnika) stopniowo odchodzi w przeszłość i nie należy go brać pod uwagę w digitalizacji.
- b) Pamięci flash usb (flashdrive, nie mylić z dyskami SSD) mogą mieć tylko znaczenie pomocnicze, np. służyć przenoszeniu danych w projektach bardzo małej skali.
- c) Eksperymentalne technologie zapisu, np. określane jako 3D i 5D, choć obiecujące, są i jeszcze długo pozostaną w sferze badań bez praktycznego zastosowania¹.
- d) Pamięć RAM, inaczej mówiąc pamięć operacyjna, to często nieuświadomiane miejsce pierwszego zapisu danych płynących z digitalizacji. Ze względu na właściwości, tj. bardzo dużą prędkość i bardzo małą trwałość, warto również wziąć pod uwagę pamięć operacyjną jako nośnik w czasie przetwarzania (do zapisu plików źródłowych i wytwarzanych pochodnych) oraz nośnik wspomagający dla prezentacji (mogą być tam zapisywane – skopiowane najczęściej udostępniane obiekty cyfrowe oraz baza danych obsługująca system wsparcia digitalizacji).
- e) Systemy dyskowe. Obecnie w pewnym stopniu ulega rozmyciu tradycyjny podział na macierze dyskowe z dyskami magnetycznymi, macierze z bardzo szybkimi dyskami ssd, klastrowe lub rozproszone systemy plików i – podobne programowo – definiowane platformy pamięci masowej (software defined storage). W jednej macierzy nieraz umieszcza się wiele rodzajów dysków i przenoszenie danych między nimi może odbywać się automatycznie. Niektóre macierze działają jak klastrowe systemy plików, a rozproszone systemy plików dostarczają funkcji takich jak macierze. Systemy dyskowe umożliwiają dobranie urządzenia, jego elementów i konfiguracji tak, by uzyskać zasoby o pożądanym pojemnościach i wydajnościach

¹ Porównaj: https://en.wikipedia.org/wiki/3D_optical_data_storage, https://en.wikipedia.org/wiki/5D_optical_data_storage [dostęp: 10.12.2022].

odpowiednich do wszystkich etapów digitalizacji pokazanych w tab. 1.

Wraz z rosnącą skalą projektu warto wziąć pod rozwagę zakres użycia magazynu danych typu blokowego (ang. block storage) i typu obiektowego (ang. object storage). Choć ogólne porównania akcentują różnice (nieraz oczywiste – jak wybór zapisu blokowego dla bazy danych oraz lepsza skalowalność zapisu obiektowego), wiele zależy od specyfiki danego projektu.

Wybór właściwych systemów dyskowych, często kosztownych, wymaga współpracy ze specjalistami IT, powinien być poprzedzony planowaniem systemu zarządzania zawartością, bazować nie tylko na dostępności sprzętu, ale również kadry do jego dalszej obsługi, uwzględniać przewidywany balans między środkami inwestycyjnymi dostępnymi na początku i bieżącymi (klasycznym przykładem jest tu zakup kosztownego systemu wraz z opłaconym z góry wsparciem technicznym vs. instalacja bezpłatnego oprogramowania wraz z koniecznością zapewnienia wsparcia we własnym zakresie).

- f) Taśmy i biblioteki taśmowe. Kasety z taśmą, na której cyfrowe dane zapisywane są przez odpowiednie napędy nazywane streamerami, to typowe nośniki do przechowywania danych w ramach wykonywania kopii bezpieczeństwa (backupu), szczególnie predestynowane do archiwizacji danych. Oferują najtańszą pojemność, dużą trwałość i praktycznie nie wymagają zasilania w czasie przechowywania danych (z wyjątkiem odświeżania zapisu np. raz na pięć lat), kosztem względnie długiego oczekiwania na dostęp do danych. Przy małej skali, gdy roczny przyrost danych nie przekracza pojemności kilku kaset, można sobie wyobrazić ręczne umieszczanie kaset w napędzie. Jednak, gdy kasety używane są do backupu lub skala archiwizacji jest nieco większa, warto skorzystać z biblioteki taśmowej, która automatycznie umieszcza kasety w napędzie – modele obsługujące około 10 kaset są obecnie niewiele droższe niż sam napęd.
- g) Szczególnym rodzajem systemu umożliwiającego zapis i przechowywanie danych są usługi przechowywania danych dostępne przez Internet. Firmy dostarczające tych usług z reguły oferują również inne zintegrowane usługi (łącznie określane jako usługi chmurowe), takie jak udostępnianie serwerów, gotowych systemów i aplikacji, co pozwala umieszczać w ich infrastrukturze całe systemy informatyczne. Korzystając z takich możliwości, trzeba uważnie zaprojektować przepływy danych i odpowiednie parametry zapisu. Dużymi zaletami takiego podejścia są: przeniesienie kosztów inwestycyjnych i części potrzeb w zakresie kompetencji IT na usługodawcę oraz możliwość płacenia tylko za zużyte zasoby. Trzeba jednak zwracać uwagę na ogromną zależność ceny od sposobu użycia wybranych usług. Przykładowo, usługa przechowywania danych, w której roczny koszt za jednostkę pojemności bywa niższy niż przy zakupie dysku, może okazać się wielokrotnie droższa, gdy okaże się, że trzeba kilka razy dziennie pobierać dane. Konieczna jest też szczegółowa analiza zapisów umowy regulaminu usługi itp. Nierzadko problemem jest zakładanie, że usługi chmurowe z definicji gwarantują bezpieczeństwo, gdy w praktyce wymaga to wykupienia dodatkowych opcji

gwarancyjnych, podobnie z przepustowością i innymi parametrami. W przypadku globalnych korporacji świadczących takie usługi (np. Amazon Web Services, Google Cloud Platform, Microsoft Azure) trzeba odnaleźć potrzebne punkty wyłącznie w gotowej (choć szerokiej) ofercie, przy czym większe szanse na dostosowanie do specyficznych potrzeb dają oferty dostawców z rynku krajowego. W szczególności mniejsi dostawcy chętniej poddają się certyfikacji wspomagającej całościową ocenę bezpieczeństwa (takiej jak Tier Classification System²), co oczywiście nie znaczy, że najwięksi dostawcy nie gwarantują co najmniej analogicznych poziomów bezpieczeństwa.

Zarządzanie zawartością

Urządzenia nie wystarczą do prowadzenia digitalizacji – potrzebne jest oprogramowanie, dotyczy to również aspektu zapisu i przechowywania danych. Tylko w projektach bardzo małej skali można pokusić się o ręczne wykonywanie zapisu danych za pomocą standardowych narzędzi systemowych i będzie to działanie czasochłonne, wrażliwe na pomyłki. Racjonalne podejście do procesu digitalizacji dyktuje konieczność wykorzystania systemu zarządzania zawartością. Taki system, nazywany repozytorium, steruje nie tylko zapisem danych, ale całością przepływów danych i często również ich przetwarzaniem, a zarazem odpowiada za integralność obiektów cyfrowych.

System repozytorium można zbudować samodzielnie, jednak jest to inwestycja wymagająca czasu i pieniędzy, a także wiedzy i doświadczenia. Warto więc rozważyć wykorzystanie już działających rozwiązań. Repozytoria o bardzo różnym zakresie działania, od skupiających się na archiwizacji po wspomagające pracę przy tworzeniu i przetwarzaniu obiektów cyfrowych, od silników wykonujących podstawowe operacje po całościowe rozwiązania – dostępne są w postaci oprogramowania do samodzielnego zainstalowania, ale też jako internetowe usługi wraz z infrastrukturą.

Wśród oprogramowania warto wyróżnić systemy dostępne jako nieodpłatne, często o otwartym kodzie (Open Source) np.:

- a) DSpace – <https://dspace.lyrasis.org/>
- b) Fedora – <https://duraspace.org/fedora/>
- c) Archivematica – <https://www.archivematica.org/en/>
- d) EPrints – <https://www.eprints.org/uk/>
- e) iRODS – Integrated Rule – Oriented Data System – <https://irods.org/>

Wśród repozytoriów dostępnych jako gotowe internetowe usługi wraz z infrastrukturą trzeba wspomnieć o Repozytorium Cyfrowym Biblioteki Narodowej, które dla polskich in-

2 <https://uptimeinstitute.com/tier-certification> [dostęp: 10.12.2022].

stytucji kultury jest dostępne nieodpłatnie i pozwala po wprowadzeniu przez Internet plików dostarczanych przez urządzenia digitalizujące, wytwarzać formy pochodne potrzebne do udostępniania, tworzyć obiekty cyfrowe, uzupełniać metadane, prezentować je w usłudze biblioteki cyfrowej z akcentowaniem proveniencji zbiorów oraz bezpiecznie archiwizować: <https://pdb.polona.pl/>.

Bezpieczeństwo zapisanych danych

Bezpieczeństwo danych i bliskie mu znaczeniowo bezpieczeństwo systemów informatycznych, jest szeroką dziedziną wiedzy usystematyzowaną normami (jak ISO/IEC 27001:2013 and ISO/IEC 27002:2013) i dobrymi praktykami.

Podstawowe tematy, które warto poddać rozważeniu i konsultacji przy okazji digitalizacji:

- a) Ochrona przed nieuprawnionym odczytem danych ma znaczenie głównie wtedy, gdy przedmiotem digitalizacji są utwory chronione prawem autorskim oraz w odniesieniu do danych osobowych (regulacje RODO). Realizuje się ją metodami technicznymi (np. firewalle i separacja sieci, szyfrowanie, reguły tworzenia oprogramowania, mocne hasła), ale nie mniej ważne jest budowanie świadomości zasad bezpieczeństwa wśród osób mających dostęp do danych.
- b) Ochrona przed utratą danych, która może następować w wyniku błędów lub celowego działania człowieka, ale też jako proces naturalnie związany z upływem czasu. Prócz zagadnień wymienionych w punkcie a) podstawowym podejściem po zabezpieczeniu danych jest zapisywanie wielu kopii na różnych nośnikach, przyjmujące postać replikacji – gdy skopiowane dane są w przybliżeniu jednakowo dostępne jak źródłowe oraz backupu – gdy kopia wykonywana jest poza standardowe nośniki, a w celu wykorzystania trzeba ją najpierw odtworzyć. Elementarnym przykładem zwielokrotnienia jest zapis danych w macierzy dyskowej (patrz RAID³), w czasie którego te same dane mogą być zapisywane na dwóch dyskach. Zwykle stosuje się tańsze choć wolniejsze rozwiązanie, w którym – mimo że dane zajmują mniej miejsca niż zwykła kopia – dzięki obliczeniom można po utracie jednego nośnika odtworzyć całość. Bardzo cenne dane bywają zapisywane w dwu i więcej macierzach dyskowych umieszczonych w różnych lokalizacjach. Oczywistą zaletą replikacji jest natychmiastowa dostępność danych, wadą również szybka propagacja ewentualnych błędów w danych. Standardową metodą zabezpieczenia danych jest tworzenie kopii bezpieczeństwa nazywanej backupem. Najczęściej wykonywana jest na kasety z taśmą magnetyczną (choć istnieją systemy backupu dyskowego) za pomocą specjalistycznego oprogramowania pozwalającego wybrać m.in. jak często ma być wykonywana i jak długo przechowywana oraz określić, czy ma być szybsza

3 <https://www.techtarget.com/searchstorage/definition/RAID> [dostęp: 10.12.2022].

do odtworzenia, czy ma zajmować mniej miejsca. Istotne jest, że backup może przechowywać wiele wersji danych z różnych momentów w czasie. Zaletą jest tani nośnik oraz wysoki poziom bezpieczeństwa, wadą – wolniejszy dostęp do danych. Ważne jest okresowe testowanie backupów.

Długotrwałe przechowywanie

W tytule rozdziału celowo wymieniono dwa pojęcia, by podkreślić znaczenie różnic w problemach zapisu i przechowywania. Szczególnie wyraźnie widać to w odniesieniu do przechowywania cyfrowych danych przez długi czas, które rozwinęło się w osobną dziedzinę określaną w angielskojęzycznej literaturze jako Digital Preservation lub Long Time [Digital] Preservation zaś po polsku bywa często określane w kontekście zbiorów cyfrowych jako długotrwałe przechowywanie, archiwizacja (a nawet wieczysta archiwizacja), konserwacja cyfrowa (pojawiają się również inne kombinacje bliskich znaczeniowo terminów). Współcześnie nieraz akcentuje się cel takiego przechowywania, używając nazwy długotrwały dostęp.

Za długotrwałe przechowywanie, według niektórych definicji, uważa się przechowywanie, w czasie którego może nastąpić zmiana technologii. W praktyce warto myśleć w kategoriach długotrwałego przechowywania, gdy przewidywany czas ważności danych to kilka, np. pięć, lat.

Długotrwałe przechowywanie to zestaw działań organizacyjnych i technicznych mających na celu zapewnienie dostępu do danych pomimo występowania takich zagrożeń jak:

- degradacja nośników – niszczenie zapisu z upływem czasu pod wpływem czynników: mechanicznych (np. wibracja, wyginanie, rysowanie), termicznych (np. odkształcanie, przyspieszenie rozkładu), elektromagnetycznych (np. rozmagnesowanie przez pole magnetyczne Ziemi lub pola od urządzeń elektrycznych, rozpad pod wpływem światła), chemicznych (korozja) lub biologicznych (grzyby, bakterie),
- utrata możliwości odczytu plików z przestarzałych nośników – brak dostępnych zgodnych urządzeń (w ciągu kilkunastu lat zmieniają się popularne metody zapisu i z rynku znikają urządzenia pozwalające odczytać zapisany dawniej nośnik. Nawet posiadanie takiego urządzenia nie gwarantuje odczytu – po latach urządzenie może nie dać się uruchomić lub nie będzie możliwości podłączenia – przestarzałe złącze),
- utrata możliwości interpretacji plików w przestarzałych formatach – brak zgodnego oprogramowania (nawet jeśli jest dostępne stare oprogramowanie, jego uruchomienie w aktualnym systemie może być bardzo utrudnione),
- utrata integralności obiektu (np. brak metadanych, uniemożliwiający odnalezienie potrzebnego obiektu lub identyfikację znalezionych plików, pliki pomieszane między obiektami itp.)
- błędy człowieka (np. słaba dokumentacja powodująca narastanie strefy braku wiedzy, wpływ czasu zwiększający szanse wystąpienia nawet mało prawdopodobnego błędu).

wać dla niewielkiego projektu trudny do przyjęcia. Jest to jednak bardzo uniwersalny model i nawet jeśli w danym przypadku miałyby okazać się zbyt wymagający do pełnej realizacji (co jest rekomendowane i powinno skończyć się uzyskaniem certyfikacji, patrz Audyt i certyfikacja⁵), to jest bardzo pożyteczną bazą tematów i problemów, które warto przeanalizować, myśląc o długotrwałym przechowywaniu. Na początek pomocne będą publikacje omawiające ten dokument, np.: Wprowadzenie do modelu referencyjnego OAIS⁶.

Celem tego rozdziału nie jest omawianie standardu OAIS, ale już przedstawiony wyżej poglądowy diagram pozwala zauważyć, że postać danych przekazywanych do archiwizacji (paczka przekazywana – Submission Information Package – SIP) jest niekoniecznie identyczna z tymi, które podlegają archiwizacji (paczka archiwalna Archival Information Package – AIP), a także tymi, które będą w przyszłości udostępniane (paczka udostępniana Dissemination Information Package – DIP). Przykładowo, dane wejściowe powinny w paczce archiwalnej być uzupełnione o informacje potrzebne do potwierdzenia autentyczności, identyfikatory, metadane opisujące wszelkie przemiany, którym podlegają w czasie przetwarzania (tu standardem de facto jest PREMIS, patrz: definicje⁷ oraz webinaria wprowadzające do PREMIS⁸).

Punktem wyjścia do długotrwałego przechowywania jest **określenie jego celu, a więc ustalenie dla kogo mają być dostępne przechowywane dane**. Nie zwalnia z tego obowiązku, naturalne w instytucjach publicznych, stwierdzenie „zbiory są dla wszystkich”, ponieważ tylko dzięki temu możliwy jest wybór istotnych cech przechowywanych obiektów. Ma to podstawowe znaczenia już dla selekcji obiektów do przechowywania, wyboru formatów danych przenoszących te istotne cechy (oba te procesy mają wpływ na koszt archiwizacji), ale też dla prawdopodobnie koniecznych w czasie długotrwałego przechowywania decyzji o zmianie formatu. Zadaniem archiwum jest dostarczenie odbiorcy obiektów z zachowaniem wybranych istotnych cech. W prostym przypadku użytkownik będzie spodziewał się otrzymać niezmiennie bitowo pliki z zawartością obiektów cyfrowych – temu służą przechowywane w archiwum sumy kontrolne i informacje pozwalające korygować ewentualne uszkodzenia. Ze względu na zmiany technologiczne w długiej perspektywie i opisane wcześniej kwestie zgodności, może być konieczne przetworzenie przechowywanych obiektów do nowego formatu – ważne, by te zmiany były precyzyjnie udokumentowane i zachowywały istotne cechy obiektu.

Za przykład wpływu określenia odbiorcy i istotnych cech może służyć porównanie kolekcji starodruków przechowywanej z myślą o dowolnym odbiorcy, ale szczególnie o badaczach (tu w uzasadnionych przypadkach wybrane zostaną bezstratne formaty źródłowe) oraz skryptów akademickich, gdzie może się okazać, że ważna jest tylko zawartość tekstowa.

Wybierając (uzgadniając z ich dostawcą) formaty danych (ale również szeroko rozumiane nośniki), warto preferować: otwarte (niezastrzeżone), o znanej i dobrze udokumentowa-

5 <https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification> [dostęp: 10.12.2022].

6 B. Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*, <https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file> [dostęp: 10.12.2022].

7 <https://www.loc.gov/standards/premis/> [dostęp: 10.12.2022].

8 Introduction to PREMIS, Karin Bredenberg, Micky Lindlar Eld Zierau, <https://www.youtube.com/embed/GiQCN-gw-HOE>, https://www.youtube.com/embed/LhRMF6vP_PU [dostęp: 10.12.2022].

nej strukturze, szeroko rozpowszechnione i będące w użyciu dłużej czas (jednak unikając schyłkowych) (porównaj formaty rekomendowane przez Bibliotekę Kongresu USA⁹).

Dążenie do zapewnienia czytelności długotrwanie przechowywanych obiektów (zwłaszcza że w czasie gdy użytkownicy będą chcieli z nich skorzystać, twórca archiwum może nie być dostępny) dyktuje możliwie **szerokie wykorzystanie standardów**. Dotyczy to również sposobu organizowania zawartości obiektów i różnorodnych metadanych niezbędnych dla długotrwałego przechowywania – czyli tworzenia paczek. Co najmniej dwa standardy powinny tu zostać wymienione:

Metadata Encoding and Transmission Standard – METS¹⁰ pozwala w jednym pliku XML zebrać i uporządkować wszelkie metadane oraz informacje o lokalizacji plików zawartości obiektu (a nawet zakodować w tym pliku XML pliki zawartości, co ma sens tylko dla niewielkich obiektów),

Oxford Common File Layout Specification – OCFL¹¹ – standard porządkujący strukturę plików obiektu cyfrowego oraz metadanych.

Konieczność stosowania znormalizowanych metadanych jest w świecie bibliotecznym oczywistością, którą kontekst długotrwałego przechowywanie dodatkowo wzmacnia i podkreśla wagę użycia unikalnych, trwałych i standardowych identyfikatorów.

Nawet najlepsze standardy ewoluują, technologie zmieniają się, a dane na nośnikach (zwłaszcza cyfrowych) ulegają stopniowemu zanikowi. Dlatego istotą długotrwałego przechowywania jest nieustanna i ujęta w procedury troska, wyznaczająca z jednej strony konieczność prowadzenia systematycznych przeglądów wykorzystanych standardów, sprawdzania dostępności urządzeń, weryfikacji poprawności i jakości zapisu, a z drugiej – gotowość do precyzyjnie dokumentowanych zmian metodycznych, migracji zasobów do nowych standardów (ewentualnie stosowania emulacji) oraz szybkiej korekty danych lub odświeżania zapisu zanim korekta okaże się potrzebna.



9 <https://web.archive.org/web/20221205160658/https://www.loc.gov/preservation/resources/rfs/data.html> [dostęp: 10.12.2022].

10 <https://www.loc.gov/standards/mets/> [dostęp: 10.12.2022].

11 <https://ocfl.io/1.1/spec/> [dostęp: 10.12.2022].

Udostępnianie

Igor Rosa

Dlaczego udostępniamy

Jednym z głównych celów digitalizacji dziedzictwa jest udostępnianie zdigitalizowanych zbiorów. Nawet jeśli działania digitalizacyjne miały na celu ochronę, archiwizację lub zbadanie zasobów, efekty digitalizacji powinny zostać udostępnione możliwie najszerszemu gronu odbiorców. Udostępnienie zbiorów w sieci może być sporym wyzwaniem – rozdział ten ma na celu przedstawienie odbiorców, przedmiotu i sposobu udostępniania zdigitalizowanych obiektów.

Co udostępniamy

Digitalizacja materiału bibliotecznego wiąże się z wykonaniem cyfrowej reprezentacji fizycznego obiektu oraz załączeniem do niej informacji o tym obiekcie – jego metadanych. W zależności od zidentyfikowanych potrzeb, środków oraz rodzaju i specyfiki zasobów należy wybrać optymalne techniki digitalizacji oraz schematu metadanych, o czym traktują poprzednie rozdziały.

Udostępnienie cyfrowych wersji jest również sposobnością do zwiększenia dostępności zbiorów poprzez dołączenie do obiektu dodatkowych informacji, których odwiedzający instytucję kultury nie odnajdzie w tradycyjnym katalogu bądź między regałami czytelnia. Dostęp do tych informacji wynika z wyjątkowego kontekstu bibliotek cyfrowych – możliwości operowania na ogromnych zbiorach danych.

Dodatkowe informacje można uzyskać poprzez maszynową analizę zdigitalizowanego obiektu lub prace redakcyjne. Część z tych informacji może zostać udostępniona szerokiemu gronu użytkowników, a inne pozostają do dyspozycji pracowników instytucji lub specjalistów.

Do dodatkowych informacji o zdigitalizowanych obiektach można zaliczyć:

- Wyekstrahowaną automatycznie warstwę tekstową publikacji – OCR. Obecnie dostarczaniem oprogramowania do ekstrakcji zajmuje się wiele firm oferujących

komercyjne rozwiązania. Istnieją również liczne open source'owe alternatywy w postaci gotowych programów (np. Tesseract) oraz bibliotek do języków programowania. Zarówno w przypadku otwartych, jak i komercyjnych rozwiązań, jakość wyekstrahowanej warstwy tekstowej może pozostawiać wiele do życzenia i wymagać ręcznej korekty. Kluczowa jest odpowiednia selekcja materiałów kierowanych do procesu OCR, aby wzbogacić zasoby o wysokiej jakości warstwę tekstową. Odczytanie warstwy OCR obiektów nieprzeznaczonych (nieposiadających tekstu, zapisanych odręcznie lub krojem pisma nieobsługiwanych przez program) może zaśmiecić bazę danych wyszukiwarki pełnotekstowej. Do specjalistycznych zastosowań warto wybrać metody ekstrakcji przeznaczone do innego rodzaju zapisów – np. muzycznego (OMR) lub pisma odręcznego.

- Efekty przetwarzania obiektów innymi metodami – np. informacje o ilustracjach i ich zawartości dodane przez redaktorów lub wygenerowane automatycznie, wyniki analizy NER (Named-entity Recognition – wydobywanie z tekstu nazw własnych i ich klasyfikacja), otoczenie semantyczne obiektu – czyli powiązania metadanych. Z kolei rozwój sieci neuronowych sprawia, że coraz prostsze staje się automatyczne przypisywanie słów kluczowych do tekstu.
- Obok rozpoznawania znaków istnieje dziedzina rozpoznawania układu tekstu na stronie, która umożliwia np. wydzielenie ze skanu gazety poszczególnych artykułów, wskazywanie elementów graficznych. W połączeniu z wyżej opisanymi metodami wzbogacania treści metoda ta pozwala dostarczać użytkownikom skuteczne narzędzia do przeszukiwania dużych zbiorów danych.

Komu udostępniamy

Udostępniając zdigitalizowane obiekty, należy pamiętać o grupach odbiorców i odpowiednio dostosować do nich interfejs i zawartość cyfrowej biblioteki.

Wyniki ankiety przeprowadzonej przez Bibliotekę Narodową w ramach projektu Patrimonium w 2019 roku pokazały, że największa grupa użytkowników odwiedza Cyfrową Bibliotekę Narodową POLONĄ w celach hobbystycznych lub rozrywkowych – 39%, w celach naukowych i badawczych – prawie 25%, w celach zawodowych – 19%, edukacyjnych – niecałe 13%.

Wśród odbiorców można spodziewać się sporej reprezentacji specjalistów z dyscyplin, których przedmiotem zainteresowania są prezentowane przez nas zbiory. Dobierając udostępniane informacje o zbiorach oraz organizując sposób ich prezentacji, warto pamiętać o potrzebach tych grup, jednocześnie nie zapominając, że mimo wszystko większość odwiedzających będzie posiadała przeciętne kompetencje cyfrowe i biegłość w posługiwaniu się zarówno wyszukiwarką, jak i warstwą prezentującą obiekty cyfrowe. Należy więc unikać hermetycznej terminologii.

Niezmiernie ważną kwestią jest wysoka dostępność serwisów internetowych prezentujących zdigitalizowane zbiory. Osoby, które mogą skorzystać z wysokiej dostępności to wszyscy użytkownicy, nie tylko niepełnosprawni (osoby niewidome, niedowidzące, ale również te,

które muszą korzystać z alternatywnych peryferiów – innych niż mysz i klawiatura). W budowaniu zrozumiałych i przejrzystych interfejsów wsparciem są wymagania dostępności (WCAG). Warto zaznaczyć, że dostosowanie serwisów np. do specjalistycznych urządzeń peryferyjnych nie musi oznaczać konieczności ich nadmiernego uproszczenia czy zubożenia – w wielu przypadkach wystarczy zapewnienie jasnej informacji np. o ograniczonej dostępności danego zasobu czy usługi oraz możliwości łatwego wyjścia z tych miejsc. Więcej uwag na ten temat znajduje się w podrozdziale *Dostępność*.

W celu skutecznego dotarcia do grup odbiorców, które mogą być niezaznajomione ze zbiorami instytucji należy również zadbać o dobrą indeksację serwisu w wyszukiwarkach internetowych, a dla odbiorców zagranicznych – o co najmniej angielską wersję językową interfejsu. Jeśli udostępniane zbiory prezentują tematykę związaną z konkretną grupą językową warto dodać wersję językową serwisu, która odpowie na potrzeby osób zajmujących się tą dziedziną. Cyfryzacja zbiorów jest również szansą na zapoznanie się z dziełami kultury dla osób mieszkających daleko od instytucji posiadających zbiory lub niemogących korzystać z nich w godzinach otwarcia tej instytucji.

Jak udostępniamy

Wyszukiwarka

Digitalizacja jest często działaniem masowym. Biblioteki narodowe wielu krajów posiadają zbiory cyfrowe liczące wiele milionów publikacji. Mniejsze repozytoria uniwersyteckie i lokalne często udostępniają dziesiątki lub setki tysięcy obiektów. Przeglądanie takich wolumenów danych jest niemożliwe bez skutecznych i dobrze zaplanowanych opcji wyszukiwania, sortowania i filtrowania zbiorów.

Ważne jest zrozumienie różnych rodzajów zapytań, typów pól wprowadzania tekstu, sposobów łączenia kryteriów zapytania i obciążeń z nimi związanych. Należy pamiętać, że różne metadane przyjmują różne typy wartości (unikalne, słownikowane, boolowskie – logiczne tak/nie itd.). W zależności od typu wartości metadana powinna być wykorzystywana w inny sposób w wyszukiwaniu. Podstawowe typy metadanych przedstawiają się następująco:

- Wartości unikalne lub prawie unikalne. Najważniejszym przykładem są wartości określające tytuł. Instytucja udostępniająca kilkaset tysięcy publikacji będzie w swoich zbiorach posiadała podobną liczbę tytułów (nie oznacza to jednak, że publikacja powinna być identyfikowana przy pomocy tytułu). Tytuł często niesie ze sobą również informację o treści. Projektując wyszukiwarkę tytułów, należy więc pamiętać o tym, że użytkownicy będą wyszukiwali fragmentów tytułów, słów zamienionych kolejnością, słów z tytułów odmienionych przez przypadki, nieraz wyrazów synonimicznych. W przypadku tytułów wyszukiwarka powinna udostępniać wieloznaczne opcje *wildcard*, wyszukiwania dokładnego, niedokładnego, lematyzacji, być powiązana ze słownikami itp. Nie ma najmniejszego sensu filtrowanie zbiorów przy pomocy tytułów, ponieważ filtrowanie przy pomocy wartości tytułu zwróci niewiele wyników, a faseta zawierająca

wartość filtrowania będzie zawierała ogromną liczbę wartości. Jednak sortowanie przy pomocy tytułów jest jak najbardziej instruktywne.

- Metadane, które mogą być słownikowane, czyli posiadają wartości z zamkniętej puli, np. autor, miejsce wydania, temat publikacji opisujące publikację powinny być wyszukiwane jako słowa kluczowe indeksu wyszukiwania. Jest to taka forma zapisywania informacji w bazie danych, która pozwala generować w łatwy sposób listy dostępnych wartości, udostępniać filtrowanie przy pomocy tych wartości i pozwalać łączyć kryteria zapytania. O ile w przypadku tytułu użytkownicy operują fragmentem metadanej wyszukiwanego obiektu, w przypadku metadanych słownikowanych mamy do czynienia z wyszukiwaniem lub zawężaniem zbioru obiektów o określonych cechach – nie próbą trafienia na konkretny obiekt. W tym przypadku należy udostępnić użytkownikom przeszukiwalną listę wartości określonej metadanej (np. poprzez sugerowanie i autouzupełnianie pola wyszukiwania) oraz filtrowanie wyniku wyszukiwania przy pomocy tych metadanych. Sortowanie przy pomocy tych pól jest mniej przydatne i może być pominięte.
- Atrybuty obiektu, takie jak: sposób dostępu do obiektu (ograniczony prawem autorskim, publiczny, tylko metadane), dostępność warstwy OCR, dostępność różnych formatów plików. Są to atrybuty funkcyjne, na podstawie których użytkownik jest w stanie zdecydować, czy interesuje go odnalezienie danego obiektu. Powinny być one przeszukiwalne w analogiczny sposób jak metadane słownikowane.

Alternatywnym sposobem wyszukiwania zasobów cyfrowych jest sieć semantyczna – ustandaryzowana struktura opisu zasobów i ich relacji. Sieć semantyczna posługuje się ideą grafu – przedstawia wszelkie relacje (tak zwane trójki) w postaci podmiotu (często nazywanego tematem), która łączy się przy pomocy orzeczenia (predykatu) z dopełnieniem (obiektem). Za przykład niech posłuży metadana autor dla pewnego obiektu, będącego zdigitalizowaną kopią Pana Tadeusza. Przedstawiona w postaci zwyczajnej metadanej jest w formie:

Autor = Adam Mickiewicz

W sieci semantycznej ta relacja ma postać:

Podmiot: Adam Mickiewicz orzeczenie: jest autorem dopełnienie: Pan Tadeusz

W sieci semantycznej podmioty, orzeczenia i dopełnienia należą do ustrukturyzowanych ontologii. Ontologie te są rozumiane przez maszyny i mogą być stosowane wspólnie. Celem jest dostarczenie maszynom narzędzi do wnioskowania na temat zasobów.

Sieć semantyczna daje więc możliwość tworzenia dalece bardziej skomplikowanych zapytań oraz umożliwia udostępnienie zasobów w sposób, który może zostać wykorzystany ponownie przez agregatory nie tylko jako zbiór metadanych, ale jako zrozumiały dla maszyn magazyn wiedzy.

Przy implementacji sieci semantycznej należy dokładnie zaplanować mapowania i wyselekcjonować odpowiednie ontologie do opisu zasobów oraz przygotować zaplecze techniczne w postaci grafowej bazy danych¹.

1 Podstawową pozycją omawiającą tę tematykę jest: J. Powell, M. Hopkins, *A Librarian's Guide to Graphs, Data and the Semantic Web*, 2015.

Prezentacja obiektu

Stwierdzenie, że digitalizowana wersja materiału bibliotecznego musi być prezentowana w sposób inny niż fizyczny egzemplarz tego obiektu jest truizmem. Należy jednak pochylić się nad sposobami w jakie możemy prezentować różne rodzaje obiektów.

Jednym z najpopularniejszych sposobów udostępniania zdigitalizowanych materiałów bibliecznych jest prezentowanie ich skanów w postaci plików graficznych. Prezentując skany należy zadbać o ich odpowiednią jakość oraz optymalizację ładowania zasobu. Udostępnienie wysokiej jakości skanów, które nie obciążają łącza internetowego użytkownika jest możliwe przy pomocy obrazów piramidalnych. Jest to technologia polegająca na generowaniu wielu rozmiarów oryginalnego skanu² i prezentowaniu użytkownikowi końcowemu tylko tych wielkości obrazu, które są mu niezbędne. Aby w pełni obsłużyć funkcjonalność piramidalnych obrazów należy również zaimplementować viewer obrazów, który pozwala na strumieniowanie tylko tych fragmentów skanu, które są potrzebne w danym momencie użytkownikowi. Oznacza to, że przybliżając wysokiej jakości skan, użytkownik otrzymuje tylko te fragmenty obrazu, które są przybliżone – zmniejszając wykorzystanie danych nawet kilkudziesięciokrotnie. W obiekcie cyfrowym składającym się nierzadko z setek skanów, viewer powinien obsługiwać możliwość przeglądania poszczególnych skanów oraz podglądu galerii skanów, w którym prezentowane są miniatury stron danego obiektu. Podstawowe informacje o obiekcie powinny być dostępne na tej samej stronie, aby użytkownik, który trafi do widoku viewera wiedział, w jakim kontekście się znajduje.

Popularnym środkiem prezentowania obiektów są również pliki PDF. Z reguły obiekty w tym formacie są eksponowane przy pomocy wtyczki osadzonej w serwisie lub udostępniane do pobrania – wtedy ciężar otwarcia pliku spoczywa na przeglądarce użytkownika lub innym oprogramowaniu na jego urządzeniu. Pliki PDF ładowane są z reguły w całości. Jako że zdigitalizowana książka może zajmować kilkaset megabajtów przestrzeni dyskowej, załadowanie pozycji w formacie PDF powinno być poprzedzone świadomą akcją użytkownika, a linki do takiego pliku nie powinny być dostępne z poziomu wyszukiwarek internetowych. Pliki tego typu najlepiej udostępniać z poziomu strony prezentacyjnej obiektu, zawierającej również informacje o nim, w tym o jego rozmiarze.

Należy unikać wykorzystywania znaków wodnych lub nakładek na obrazy uniemożliwiających pobranie plików, blokad zrzutów ekranu itp. Nie jest to zabezpieczenie chroniące przed nieupoważnionym pobraniem i udostępnieniem plików. Zdeterminowany, zaawansowany użytkownik jest w stanie pobrać zarówno zabezpieczony obraz, jak również usunąć znak wodny z pliku, jest to zarazem utrudnienie i zmniejszenie wygody użytkowania serwisu dla większości użytkowników oraz wyraz braku zaufania do nich.

Nawet jeśli obiekty są udostępniane w jeden wybrany sposób, warto rozważyć udostępnienie opcji pobierania obiektu w innych formatach:

- Paczki ze skanami – warto w tym przypadku rozważyć limit wielkości paczki/liczby skanów możliwych do pobrania jednocześnie.

2 Wśród technologii wspierających tego typu generowanie obrazu należy wymienić: Cantaloupe, IIP Image Server, Loris oraz nakładki na różne frameworki Flask-IIIIF, Shimmy itp.

- Pliku PDF – należy zastanowić się nad wyborem jednego z dwóch modeli pobierania obiektów: generowania plików PDF na żądanie lub przechowania ich we własnych bazach. Generowanie plików zwalnia znaczne zasoby pojemności dyskowej, ale jest operacją wymagającą większej mocy obliczeniowej i pamięci RAM.
- Pliku MOBI, EPUB – jeśli dostępna jest dobrej jakości warstwa tekstowa.
- Innych specyficznych dla danego obiektu formatów.

Dostępność

Standardy dostępności stały się powszechnym elementem współczesnych praktyk wytwarzania oprogramowania³. Zgodność ze standardem WCAG 2.1 jest nie tylko wymogiem w stosunku do stron rządowych i tworzonych z pomocą środków publicznych, ale również staje się standardem w licznych komercyjnych serwisach. Na wielu stronach internetowych możemy znaleźć przełączniki podnoszące kontrast kolorów lub zwiększające rozmiar czcionki. Są to efekty błędnego rozumienia zasad dostępności i nie ułatwiają one pracy osobom niepełnosprawnym ani nie umożliwiają przejścia audytów dostępności. Standard WCAG reguluje nie tylko sferę wizualną interfejsu, ale również jego funkcjonalność, logikę projektu interfejsu, poprawność kodu i dostępność zawartych w nim treści.

Warto więc wspomnieć, że pomiędzy poprawnie stworzoną stroną internetową a spełnianiem standardów dostępności następuje swoiste sprzężenie zwrotne – poprawnie stworzony serwis łatwo dostosować do tych standardów, a spełnienie ich podnosi często poprawność kodu takiego serwisu.

Strony podmiotów publicznych są również zobowiązane do publikowania deklaracji dostępności. Dokument ten pomaga osobom polegającym na standardach dostępności zorientować się w sposobach dostępu do treści. Aby stworzyć taką deklarację należy dokonać audytu serwisu. Przy samodzielnym audycie można skorzystać z licznych darmowych narzędzi: czytników, np. NVDA, walidatorów takich jak WAVE – Web Accessibility Evaluation Tool, W3C Markup Validation Service i narzędzi testujących zawartość strony, np. ARC Toolkit lub Color Contrast Analyser.

Regulacje dotyczące dostępności zobowiązują podmioty publiczne również do zapewnienia dostępności swoich zasobów. Działanie takie byłoby karkołomne w przypadku zdigitalizowanych obiektów dziedzictwa, dlatego dla takich treści istnieje wyłączenie z obowiązku zapewnienia dostępności⁴. Należy jednak pamiętać, że wyłączenie dotyczy jedynie samego materiału bibliotecznego, archiwalnego bądź zabytkowego – nie ścieżki do tego zasobu w ramach strony internetowej i strony go udostępniającej.

3 Jako glossę warto zamieścić informację, że standard WCAG dotyczący stron internetowych jest tylko jednym z kilku obowiązujących. Istnieją również inne standardy, np. ATAG – regulujący zasady dostępności przy tworzeniu narzędzi CMS i UAAG – zasady dostępności przy tworzeniu aplikacji klienckich. Ponadto istnieją standardy tworzenia dokumentów i treści multimedialnych w sposób dostępny.

4 Ustawa z dnia 4 kwietnia 2019 r. o dostępności cyfrowej stron internetowych i aplikacji mobilnych podmiotów publicznych. Art. 3 ust 7.

Odnajdywalność

Zdigitalizowane zasoby powinny być odnajdywalne nie tylko z poziomu biblioteki cyfrowej, ale również z poziomu wyszukiwarek internetowych. Wielu potencjalnych użytkowników nie wie, w której instytucji znajdują się poszukiwane przez nich zasoby. Inni przestają na przeszukaniu tylko kilku największych bibliotek cyfrowych. Pewnym rozwiązaniem są agregatory, które jednak nie dostarczają zawsze perfekcyjnych wyników wyszukiwania i muszą upraszczać schemat danych, aby zachować spójność wielu instytucji (co powoduje nieuwzględnianie pewnych metadanych lub wspólne traktowanie różnych metadanych), ponadto wielu użytkowników nie wie o istnieniu takich serwisów. Większość korzystających z Internetu potrafi jednak posługiwać się wyszukiwarkami internetowymi.

Istnieje szereg dobrych praktyk pozwalających na zapewnienie bibliotece cyfrowej zgodności z wyszukiwarkami internetowymi. Podobnie jak w przypadku standardów dostępności, jedną z najważniejszych zasad jest poprawna konstrukcja strony internetowej. Odpowiednie stosowanie znaczników HTML (w szczególności tytułowych, tzw. headerów) pozwala wyszukiwarkom automatycznie zinterpretować zawartość strony. Warto zapewnić odpowiednie znaczniki meta dla każdej podstrony – mogą one wskazywać na tytuł publikacji, krótki opis zawierający słowa kluczowe i metadane obiektu. Najbardziej optymalne jest podawanie informacji o sposobie dostępu do zasobu (w szczególności wskazanie, że jest darmowy – co może być kluczowe dla użytkownika wyszukiwarki internetowej), formatu publikacji i podstawowych metadanych. W ramach strony zawierającej opis obiektu ogromne znaczenie może mieć zaimplementowanie danych uporządkowanych⁵ – znaczników stojących obok metadanych, które wskazują maszynom, jaki rodzaj informacji niesie ze sobą dana wartość.

Przy tworzeniu nowej wersji biblioteki cyfrowej warto zadbać o przekierowania ze starych zasobów do nowego adresu. Pozwala to na zachowanie aktualności zasobu, utrzymanie funkcjonowania przypisów do zasobów cyfrowych i utrzymanie zaufania do instytucji. Przekierowanie z kodem HTTP 301 pozwala na zachowanie tej spójności oraz podtrzymanie przez wyszukiwarki internetowe stron zasobów w nienaruszonym rankingu. Maszyny indeksujące (crawlers) co jakiś czas odwiedzają strony udostępnionego w sieci serwisu i dodają/aktualizują opis w wyszukiwarce. Z tego powodu udostępniony obiekt może przez długi czas nie figurować w popularnych przeglądarkach lub mogą istnieć różnice w wynikach między przeglądarkami. Możliwym rozwiązaniem tego problemu jest zaimplementowanie mapy serwisu. W przypadku bibliotek cyfrowych zawierających tysiące publikacji taka mapa powinna zostać udostępniona zgodnie ze specyfikacją – najczęściej w formacie XML.

W celu monitorowania zachowań użytkowników, popularności serwisu i jego funkcjonalności warto zaimplementować narzędzie do monitorowania statystyk przeglądania np. Google Analytics, lub stworzyć własny moduł statystyk.

5 Informacje na ten temat znajdują się w dokumentacji wyszukiwarki Google. <https://developers.google.com/search/docs/appearance/structured-data/intro-structured-data> [dostęp: 17.11.2022].

Dodatkowe funkcje biblioteki cyfrowej

Biblioteka cyfrowa nie musi być wyłącznie repozytorium zasobów instytucji. Istnieje kilka sposobów, dzięki którym udostępnianie zbiorów może stać się bardziej atrakcyjne i przystępne dla użytkowników. Jedną z najprostszych możliwości organizowania zasobów jest tworzenie redagowanych kolekcji. Instytucja udostępniająca swoje zasoby jest z reguły najlepiej zorientowana w ich tematyce, dlatego warto rozważyć tworzenie kolekcji tematycznych, dzięki którym dostęp do informacji staje się o wiele prostszy. Zwiększaniu widoczności, zwłaszcza zbiorów graficznych, sprzyjają umiejętnie prowadzone profile w mediach społecznościowych, a użytkownicy zainteresowani wnikliwszym poznaniem konkretnych zagadnień docenią obszerniejsze formy oferujące pogłębiony wgląd w zasoby, np. bloga.

Pewne kategorie zasobów są osadzone w szerszych strukturach, np. czasopisma, serie wydawnicze, powiązane utwory. Warto przemyśleć odwzorowanie tych powiązań w ramach biblioteki cyfrowej. Każde z tych powiązań powinno posiadać specyficzny interfejs. Dla gazet codziennych warto stworzyć widok kalendarza, dla czasopism naukowych spis artykułów. Serie i powiązane publikacje warto prezentować w kontekście samego obiektu.

Użytkownicy powinni mieć możliwość wchodzenia w interakcje z zasobami. Tworzenie notatek do obiektów i kolekcji zasługuje na przemyślenie. Kolejnym krokiem może być zaimplementowanie treści społecznościowych – zgodnie z duchem serwisów Web 2.0 (w których użytkownik nie jest tylko konsumentem treści jak w sieci Web 1.0, ale również ich producentem). Wiedza i ekspertyza użytkowników przekazane w postaci treści społecznościowych mogą być dodatkową warstwą opisu zasobów. Podejście to wymaga moderacji i zwiększenia odpowiedzialności instytucji. Użytkownicy mogą współtworzyć zasoby serwisu również poprzez crowdsourcing danych, czyli uzupełnianie i podawanie pewnych informacji o zasobach, z którymi wchodzi w interakcję. Można zachęcać do poprawiania i uzupełniania metadanych obiektów, zaznaczania i opisywania ilustracji, tagowania treści. Organizowanie spotkań z użytkownikami – edytorów, w celu uzupełniania informacji również może skokowo podnieść jakość opisu zasobów.

Dostęp poprzez API

Udostępnianie zdigitalizowanych obiektów nie powinno ograniczać się wyłącznie do umieszczenia zasobu w serwisie internetowym – cenne są również alternatywne sposoby dostępu. Poza zapewnieniem plików pochodnych, warto pochylić się nad udostępnieniem zasobów poprzez publiczne API. Wiele projektów badawczych, agregatorów informacji i baz danych publicznych korzysta z interfejsów programistycznych do pobierania zasobów z innych serwisów. Z kolei nowoczesna architektura serwisów internetowych również rozdziela warstwę prezentacyjną od warstwy logiki (backendu). W takim wypadku udostępnienie części funkcji poprzez API, staje się nie tylko relatywnie łatwym zadaniem, ale również może drastycznie zwiększyć dostępność zasobów.

Obecnym standardem otwartych API jest architektura REST. Poprzez API możliwe jest udostępnianie wyszukiwarki, poszczególnych zasobów, a także dodatkowych funkcjonalności. Warto rozważyć udostępnianie metadanych obiektów poprzez API w jednym z popularnych standardów, np. DC Terms. W przypadku zawartości obiektu (skanów, plików pdf

itp.) standardem są manifesty IIIF Presentation API dla obiektu, jego metadanych i struktury oraz IIIF Image API dla pojedynczego skanu. Standard IIIF jest wspierany przez rosnącą liczbę wiodących instytucji⁶.

Dla wszelkiego rodzaju harvesterów warto również udostępnić swoje zasoby poprzez protokół OAI-PMH.



6 <https://iiif.io/>.

Modelowy workflow procesu digitalizacji

Paweł Mazur, Dariusz Paradowski, Jakub Kostynowicz

