

Wiesław Wróbel

Biblioteka Uniwersytecka im. Jerzego Giedroycia

w.wróbel@uwb.edu.pl

ORCID: 0000-0002-4673-5761



Znaczenie digitalizacji w budowie zasobów Repozytorium Uniwersytetu w Białymstoku

Artykuł omawia znaczenie digitalizacji w budowie cyfrowych zasobów Internetu oraz roli Biblioteki Uniwersyteckiej im. Jerzego Giedroycia w tym procesie na przykładzie Repozytorium Uniwersytetu w Białymstoku. Digitalizację, w szerokim rozumieniu jej definicji, Biblioteka Uniwersytecka realizuje od ponad 15 lat. Już od 2006 roku współtworzy Podlaską Bibliotekę Cyfrową, na potrzeby której uruchomiono pierwszą w regionie Pracownię Digitalizacji Zbiorów. Zdobyte przez ten czas doświadczenia pozwoliły na świadomą i zorganizowaną digitalizację dokumentów na potrzeby Repozytorium. Dlatego w artykule omówiono przyjęte standardy digitalizacji zasobów przeznaczonych do udostępnienia w Repozytorium, metody zapewnienia czytelności, przeszukiwalności i dostępności plików (format) oraz płaszczyzny współpracy między Pracownią Zbiorów Specjalnych i Digitalizacji a redakcją i użytkownikami Repozytorium Uniwersytetu w Białymstoku.

Słowa kluczowe: biblioteki cyfrowe i repozytoria, digitalizacja, standardy digitalizacyjne, udostępnianie danych *on-line*, standard PDF

Skoro otacza nas przepastny świat cyfrowy, intuicyjnie domyślamy się, czym jest digitalizacja. Ale czy na pewno dobrze rozumiemy znaczenie tego terminu? W wyniku rewolucji, jaka dokonała się na przełomie XX i XXI wieku za sprawą Internetu, kultura człowieka stała się kulturą cyfrową. Dlatego,

kiedy mówimy o digitalizacji, zwykle myślimy po prostu o zasobach internetowych lub wszystkich danych cyfrowych bądź urządzeniach, które zapewniają do nich dostęp. Zwłaszcza, że im głębiej podążamy w XXI wiek, tym bardziej wersja cyfrowa staje się wersją podstawową wielu produktów czy utworów. Odchodzimy od dużych nakładów gazet i książek, filmy udostępniamy na platformach streamingowych, nie kupujemy płyt CD i DVD, zastępując to wszystko bezpośrednim dostępem *on-line*. Coraz mniej kserujemy, w zamian wykonując po prostu zdjęcia aparatem cyfrowym dostępnym w każdym smartfonie lub przesyłamy cyfrowe dokumenty w systemach EKD... Musimy jednak uzmysłwić sobie jedną i podstawową rzecz – dane cyfrowe nie rosną na drzewach i muszą być wytworzone w różnych okolicznościach i przy użyciu różnych technologii. Nie chodzi o kwestię zapisu takiej czy innej współczesnej twórczości wyłącznie w wersji cyfrowej, bo dziś jest to po prostu standard. Mam na myśli fakt, że nieprzebrane zasoby szeroko rozumianej kultury człowieka – archiwalia, rękopisy, książki, czasopisma, dzieła sztuki, utwory muzyczne czy filmowe – były od setek lat wytwarzane tylko w postaci tradycyjnej. Czyli takiej, którą nazwiemy *a contrario* wersji cyfrowej – analogową. Internet od początku był medium mającym w sobie potencjał otwarcia dostępu do wszystkich osiągnięć ludzkiej cywilizacji w sposób nieograniczony i nieskrępowany. Dzięki digitalizacji tak też się stało. Przyczyniły się do tego w sposób szczególny archiwa, biblioteki, muzea, ośrodki badawcze czy szkoły wyższe, które zaczęły masowo zwiększać zasoby Internetu przez digitalizację i udostępnienie w sieci swoich kolekcji. Ale nie luźno „wrzuconych” do sieci, tylko udostępnionych przy użyciu odpowiednio skonstruowanych programów prezentujących dane w sposób usystematyzowany i opisany, co w konsekwencji zrodziło pierwsze biblioteki i cyfrowe kolekcje. I tu dochodzimy do sedna sprawy.

Digitalizacja nie oznacza tylko, że coś przybiera postać cyfrową, czyli digitalną. Termin ten możemy rozumieć bowiem dwojako, w zależności od kontekstu. *Sensu stricto* digitalizacja, zgodnie z encyklopedyczną definicją, to czynność przetwarzania danych utrwalonych w wersji analogowej do postaci cyfrowej (dodajmy, że przy zachowaniu wiernego odwzorowania). To zjawisko jest najbardziej powszechne i zasadniczo dziś każdy prędzej czy później coś zdigitalizuje (kto nie przegrywał kaset VHS na komputer albo nie skanował zdjęć z rodzinnego albumu?). Ale z punktu widzenia instytucji podejmujących się działań digitalizacyjnych termin ten ma znacznie szersze

znaczenie. Digitalizacja to proces, a czynność przekształcania tradycyjnych zasobów w ich cyfrowe odwzorowanie jest tylko jednym z wielu etapów. Dlatego *sensu largo* digitalizacja będzie ciągiem czynności obejmujących świadomą i z góry zaplanowaną działalność, w której kluczowe znaczenie ma ustalenie celów digitalizacji (naukowe, popularyzatorskie, zachowawcze), dobór zasobu przeznaczonego do cyfryzacji (z uwzględnieniem stanu prawno-autorskiego) i jego odpowiednie przygotowanie (np. konserwacja), zaś po wykonaniu kopii cyfrowej – opracowanie metadanych, publikacja w wersji cyfrowej *on-line* w przeznaczonym do tego serwisie internetowym, zapewnienie trwałości, dostępności, aktualizacja oprogramowania służącego do udostępniania lub po prostu monitorowanie statystyk wykorzystania. Digitalizacja jest więc całością działań i procesów zmierzających do powstania wersji cyfrowej obiektów przeznaczonych zarówno do długoterminowego przechowywania, jak i do udostępnienia za pomocą systemów zarządzania bibliotekami cyfrowymi/repozytoriami/cyfrowymi kolekcjami¹. Tak sformułowana definicja odnosi się więc głównie do digitalizacji zinstytucjonalizowanej, będącej częścią działalności różnego rodzaju placówek czy organizacji, których celem jest upowszechnianie oraz zapewnianie dostępu do posiadanych informacji czy zasobów. W naszym przypadku przykładem będzie Uniwersytet w Białymstoku i jego ogólnouczelniana jednostka organizacyjna – Biblioteka Uniwersytecka im. Jerzego Giedroycia.

Może stwierdzenie, że bez digitalizacji nie byłoby świata cyfrowego byłoby przesadą, ale niewątpliwie poprawnie rozumiana digitalizacja miała (i nadal ma) ogromne znaczenie w przebiegu rewolucji cyfrowej oraz przekształcaniu świata w globalną wioskę. Jej znaczenia po prostu nie da się przecenić. Przez pryzmat tak przedstawionej szerokiej definicji digitalizacji spróbujmy spojrzeć na ten segment dotychczasowej działalności Biblioteki Uniwersyteckiej im. Jerzego Giedroycia ze szczególnym uwzględnieniem Szanownego Jubilata – Repozytorium Uniwersytetu w Białymstoku.

¹ Przykładem teoretycznego przedstawienia schematu prac digitalizacyjnych może być *workflow* zaproponowany w publikacji: *Digitalizacja piśmiennictwa*, oprac. i red. D. Paradowski, Warszawa 2010, s. 208–213. Obejmuje on rozbudowane działania: zaplanowanie digitalizacji danego obiektu, sprawdzanie dostępności i wydanie z magazynu, transport, przygotowanie do digitalizacji (w tym konsultacja konserwatorska), digitalizacja, zwrot obiektu oraz równoległe udostępnienie wersji cyfrowej i archiwizacja.

Współcześnie digitalizacja staje się jedną z kluczowych usług świadczonych przez biblioteki. Skanuje się cymelia i zbiory specjalne, wypożycza e-booki, prenumeruje dostępy do cyfrowych księgozbiorów czy periodyków. Nikt już się nie dziwi powszechnemu zjawisku, w którym biblioteka przyjmuje rolę gracza drugiego planu i staje się pośrednikiem w dostępie do cyfrowych wersji części swoich zbiorów w celu zabezpieczenia oryginałów i ułatwienia do nich dostępu, a w przypadku bibliotek akademickich także w celu wsparcia dydaktyki czy prac naukowo-badawczych. Te cele przyświecały praktycznie wszystkim bibliotekom w chwili podjęcia działań z zakresu digitalizacji, które obejmowały inwestycje w budowę własnych odpowiednio wyposażonych pracowni digitalizacji oraz uruchomienie bibliotek cyfrowych. Biblioteka Uniwersytecka w globalną i krajową działalność digitalizacyjną włączyła się stosunkowo wcześniej, bo już w 2006 roku, wraz z uruchomieniem Podlaskiej Biblioteki Cyfrowej (PBC).

Podlaska Biblioteka Cyfrowa powstała jako szósta biblioteka cyfrowa w kraju. Była efektem działalności założonego w 2004 roku Konsorcjum Bibliotek Naukowych Miasta Białegostoku². Prace przy budowie platformy zainicjowano w 2005 roku. Politechnika Białostocka zakupiła wówczas zyskujące na popularności oprogramowanie dLibra, opracowane przez Poznańskie Centrum Superkomputerowo-Sieciowe, które zainstalowano na serwerach Miejskiej Sieci Komputerowej BIAMAN. Oficjalne uruchomienie platformy nastąpiło w grudniu następnego roku. Platformę i zasady jej funkcjonowania wzorowano na polskich bibliotekach cyfrowych uruchomionych nieco wcześniej. Zadaniem PBC była (i jest nadal) rozbudowa zasobów Internetu poprzez cyfryzację zbiorów posiadanych przez współorganizatorów tego przedsięwzięcia oraz zapewnienie nieskrępowanego do nich dostępu. Oprócz cennych dokumentów historycznych, starodruków, zbiorów kartograficznych, wydawnictw regionalnych, muzykaliów czy archiwaliów zdecydowano się na digitalizację najnowszych i najczęściej wykorzystywanych skryptów uczelnianych, podręczników czy publikacji, zwłaszcza regionalnych. Miało to stanowić nie tylko promocję potencjału badawczego i naukowego uczelni zrzeszonych

² W jego skład weszły biblioteki: Uniwersytetu w Białymstoku, Uniwersytetu Medycznego, Archidiecezjalnego Wyższego Seminarium Duchownego, Uniwersytetu Muzycznego Fryderyka Chopina w Warszawie – Wydziału Instrumentalno-Pedagogicznego w Białymstoku i Politechniki Białostockiej oraz Książnica Podlaska.

w Konsorcjum, ale również stanowić pomoc w procesach dydaktycznych³. W ówczesnych realiach przedsięwzięcie to było w regionie działaniem pionierskim i pozwoliło włączyć tę część Polski w proces szybko postępującej budowy krajowej sieci kolekcji cyfrowych. Miarą sukcesu było nadanie w 2006 roku PBC tytułu Podlaskiej Marki Roku w kategorii „Pomysł roku”. Koordynatorem przedsięwzięcia nadal pozostaje Biblioteka Uniwersytecka. Przez lata grono współtwórców platformy urosło do dwunastu, a lista współpracowników stale się wydłuża i liczy obecnie ponad 30 instytucji, które cyfryzują i udostępniają swoje zasoby.

Fundamentalne znaczenie dla organizacji PBC miała budowa pracowni digitalizacyjnej, stanowiącej warunek *sine qua non* działania serwisu. Dzięki funduszom unijnym zorganizowano w Bibliotece Uniwersyteckiej Pracownię Digitalizacji Zbiorów (przekształconą w 2019 roku w Pracownię Zbiorów Specjalnych i Digitalizacji), która jako pierwsza tego typu jednostka w mieście i regionie wykonywała kopie cyfrowe dokumentów na potrzeby PBC, zarówno z zasobu instytucji macierzystej, jak i większości partnerów. Podlaska Biblioteka Cyfrowa była pierwszym krokiem Biblioteki Uniwersyteckiej wykonanym w kierunku urzeczywistnienia coraz głośniejszych postulatów ruchu Otwartego Dostępu (*Open Access*), polegającego w głównej mierze na zapewnianiu dostępu do zasobów naukowych i informacyjnych bez żadnych barier i ograniczeń. Stanowiła także cenną lekcję z zakresu technik digitalizacyjnych, sposobów opracowywania metadanych czy współpracy z otoczeniem oraz użytkownikami. Zebrane przy tej okazji doświadczenia można było wykorzystać przy organizacji Repozytorium Uniwersytetu w Białymstoku.

Na pierwszy rzut oka niewiele różni Podlaską Bibliotekę Cyfrową i Repozytorium Uniwersytetu w Białymstoku. Przecież obie platformy mają wspólny cel, jakim jest udostępnianie bez ograniczeń cyfrowej wersji dokumentów za pośrednictwem Internetu. Ktoś mógłby więc zapytać – po co tworzyć dwie biblioteki cyfrowe, skoro ich zadania są analogiczne? Ale gdy przyjrzymy się im bliżej okaże się, że znacznie więcej je dzieli niż łączy. Podstawową różnicą jest fakt, że Repozytorium ma charakter instytucjonalny, a więc jest ono organizowane przez konkretną instytucję (w przypadku PBC współtwórców

³ E. Bezzubik, *Rusza Podlaska Biblioteka Cyfrowa*, „Biuletyn EBIB” 2006, nr 4(74), <http://www.ebib.pl/2006/74/bezzubik.php> (dostęp: 12.06.2023).

jest wielu), zaś zasób i struktura serwisu odzwierciedla strukturę instytucji, która je tworzy. Ponadto odmienny jest zasób przeznaczony do udostępnienia. W przypadku Podlaskiej Biblioteki Cyfrowej głównym materiałem poddanym digitalizacji były zasoby znajdujące się w domenie publicznej bądź takie, których udostępnienie nastąpiło na zasadzie udzielonych licencji, przy czym kryterium ich doboru jest najczęściej fakt wytworzenia w regionie lub odnoszenie się do regionu. Tymczasem Repozytorium Uniwersytetu w Białymstoku zostało powołane w 2013 roku przez Rektora Uniwersytetu w Białymstoku w jednym, podstawowym celu – bezterminowego gromadzenia oraz upowszechniania dorobku naukowego i dydaktycznego pracowników, doktorantów i studentów Uczelni. Platforma ma więc za zadanie zwiększyć powszechność i zapewnić łatwy dostęp do wyników prac naukowych prowadzonych w naszej placówce. Podlaska Biblioteka Cyfrowa może w każdej chwili wycofać udostępnione dokumenty, podczas gdy materiały zdeponowane w Repozytorium posiadają przypisane do nich na stałe unikalne identyfikatory URI i DOI, mające za zadanie trwale identyfikować konkretny zasób nawet w sytuacji jego niedostępności w sieci (w przeciwieństwie do adresu URL). Wreszcie w celu zwiększenia efektywności udostępniania dorobku naukowego Repozytorium zostało zarejestrowane w światowych katalogach repozytoriów OpenDOAR oraz ROAR, a także w Agregatorze CEON i portalu prac doktorskich DART Europe. Od momentu otwarcia dane z repozytorium są pobierane przez światowe wyszukiwarki m.in: Google Scholar, BASE, WordCat, Yandex. Od 2016 roku metadane większości czasopism akademickich Uniwersytetu w Białymstoku, znajdujących się w Repozytorium Uniwersytetu w Białymstoku, są importowane do wybranych baz. Takiego rejestrowania danych nie realizuje się w bibliotekach cyfrowych. Z założenia nie mają one zadań związanych z możliwie najszerszym upowszechnianiem zawartych w nich danych, natomiast w przypadku Repozytorium kwestia upowszechnienia ma fundamentalne znaczenie oraz przekłada się na różnego rodzaju dane parametryzacyjne i bibliometryczne, ostatecznie służące do określania pozycji Uniwersytetu w krajowych i światowych rankingach bądź też stymulowania rozwoju Uczelni i jej kadry. Kolejną kluczową różnicą między PBC i RUB jest sposób deponowania dokumentów. W przypadku biblioteki cyfrowej robi to oddelegowany do tego pracownik instytucji współtworzącej, który jednocześnie odpowiada za dobór dokumentów przeznaczonych do udostępnienia i opracowanie ich metadanych. W RUB deponowanie jest samodzielne – osoby zarejestrowane umieszczają na platformie

własne prace naukowe i uzupełniają podstawowe metadane dotyczące każdego obiektu, udzielając tym samym odpowiedniej licencji. Wreszcie, zgodnie z zarządzeniami rektorskimi, w RUB są deponowane wszystkie uniwersyteckie czasopisma oraz obligatoryjnie obronione na Uczelni prace doktorskie. Za poprawną działalność Repozytorium odpowiada zespół bibliotekarzy-redaktorów pracujących w wydzielonej sekcji w ramach Biblioteki Uniwersyteckiej.

W chwili powołania do życia Repozytorium Uniwersytetu w Białymstoku w sposób naturalny rozszerzono także zakres działalności Pracowni Zbiorów Specjalnych i Digitalizacji, która dotychczas realizowała prace przede wszystkim na potrzeby PBC. Pomijając kwestie techniczne związane z implementacją oprogramowania dSpace i uruchomieniem Repozytorium należy zauważyć, że przystępując do budowy nowej platformy Biblioteka nie dysponowała w ogóle (lub w bardzo ograniczonym stopniu) wersjami cyfrowymi publikacji kadry Uczelni. Trzeba je było dopiero pozyskać bezpośrednio u wydawców lub – co oczywiste – poprzez retrospektywną digitalizację dokumentów dostępnych tylko w wersji tradycyjnej. W związku z tym Pracownia Digitalizacji Zbiorów, jako jednostka wyspecjalizowana w zakresie procesów digitalizacyjnych, musiała przyjąć na siebie ciężar przedsięwzięcia odnośnie do cyfryzacji dokumentów na zlecenie Repozytorium Uniwersytetu w Białymstoku oraz nieodpłatnie *on demand*, czyli na żądanie na rzecz pracowników Uczelni, którzy nie byli w stanie samodzielnie zdigitalizować prac przeznaczonych do zdeponowania. Przy czym pewna część dokumentów możliwych do udostępnienia w Repozytorium została już wcześniej poddana digitalizacji na potrzeby Podlaskiej Biblioteki Cyfrowej, co w pewnym stopniu ułatwiło początkowe prace nad budową zasobu Repozytorium.

Z technicznego punktu widzenia przedsięwzięcie nie było trudne. W przypadku digitalizacji na potrzeby PBC należało każdorazowo uwzględniać aspekty techniczne niezbędne do bezpiecznego i poprawnego cyfryzowania dokumentów o najróżniejszym stanie zachowania, wielkości, formacie czy technice wykonania. Inaczej bowiem należy skanować starodruk, inaczej mapę, jeszcze inaczej wydawnictwa nowsze. Służą do tego celu różnego typu skanery – płaskie, planetarne zaopatrzone w stoły szalkowe lub kołyski bądź stoły reprograficzne. W przypadku dokumentów udostępnianych w Repozytorium digitalizacją obejmowano w zasadzie wyłącznie druki wydane po 1997 roku, a więc od chwili przekształcenia Filii Uniwersytetu Warszawskiego w Białymstoku w samodzielny Uniwersytet, aczkolwiek

z uwzględnieniem w pewnych przypadkach także druków wcześniejszych. W każdym razie digitalizacją objęto tylko obiekty współczesne, wydane przy użyciu najnowszych technik drukarskich. Stan ten znacząco ułatwiał prace digitalizacyjne, gdyż mogły być one realizowane przy użyciu skanerów płaskich znajdujących się na wyposażeniu Pracowni Digitalizacji Zbiorów, bez względu na ryzyko uszkodzenia bloku książki w czasie nadmiernego docisku do płyty skanera. Często digitalizacja dotyczy tylko wybranych fragmentów całych publikacji (artykuły, rozdziały itp.), dlatego pośrednim rezultatem tego stanu jest przede wszystkim skrócenie czasu pracy niezbędnego do wykonania kopii cyfrowej, ale także – co równie ważne z punktu widzenia całego procesu – zmniejszenie przestrzeni dyskowej koniecznej do przechowywania cyfrowych wersji dokumentów.

Fizyczna praca przy cyfryzacji dokumentów na skanerach to jedno. Drugie to ustalenie standardów digitalizacji. Nie ulegało wątpliwości, że wersję cyfrową dokumentów należy udostępniać przy użyciu formatu PDF. Jest on aktualnie najbardziej rozpowszechnionym formatem służącym do prezentacji, przenoszenia oraz drukowania treści tekstowych i graficznych. Twórcą formatu jest firma Adobe, ale od wielu lat jest on rozwijany i utrzymywany przez Międzynarodową Organizację Normalizacyjną (ISO). Dzięki „uwolnieniu” i ustandaryzowaniu tego formatu na rynku istnieje wiele darmowych programów i systemów czytających pliki PDF oraz umożliwiających eksport do tego formatu. Wreszcie możliwość bezpośredniego wyświetlania plików PDF posiadają praktycznie wszystkie przeglądarki internetowe w wersji komputerowej czy mobilnej. W przypadku Repozytorium, do którego dostęp odbywa się przy użyciu przeglądarek internetowych, wybór był więc oczywisty. Trzeba jednak przyznać, że wynikał on także w pewnym stopniu z doświadczeń zdobytych w czasie tworzenia zasobów Podlaskiej Biblioteki Cyfrowej. W przypadku PBC od początku przyjęto PDF jako główny format udostępnianych plików, ale wiele innych instytucji organizujących własne biblioteki cyfrowe zdecydowało się na prezentację dokumentów w formacie DJVU. Różnice między tymi formatami są dwójakiej natury. Po pierwsze, odczyt pliku PDF za pośrednictwem aplikacji w przeglądarce wymaga załadowania pliku w całości, natomiast w przypadku DJVU odpowiednia aplikacja (współcześnie większość przeglądarek jest w stanie sobie bez niej poradzić, ale jakiś czas temu wielu użytkowników miało z tym ogromny problem) łąduje każdą stronę dokumentu niezależnie. Jeśli zdecydujemy się pobrać tekst udostępniony

w formacie PDF na dysk własnego urządzenia, ma on postać pojedynczego pliku. Natomiast dokument typu DJVU składa się z wielu plików połączonych przy użyciu pliku indeksacyjnego, a *gros* całości stanowią po prostu zeskanowane pojedyncze strony. Po drugie, format DJVU został zaprojektowany tak, aby zminimalizować konieczność ładowania dużych plików. Wyraża się to właśnie poprzez wyświetlanie kolejno pojedynczych stron oraz użycie silnej kompresji prezentowanych skanów. Skutkiem tego jest najczęściej niska jakość graficzna dokumentów w tym formacie. Mniejszy kłopot, jeśli udostępniono w ten sposób dokument tekstowy. Znacznie gorzej, jeśli jest to grafika, fotografia czy dokument kartograficzny. Oczywiście format PDF daje możliwość zastosowania różnych stopni kompresji, ale z punktu widzenia technik digitalizacyjnych zdecydowanie lepsze efekty przynoszą odpowiednio przygotowane pliki wglądowe, scalone następnie w plik PDF bez stosowania jakichkolwiek działań optymalizacyjnych, które mogą „uszkodzić” skan. Takie rozwiązanie ma swoje wady i zalety. Najpoważniejszą wadą jest fakt, że jeśli chcielibyśmy udostępnić za pomocą tak skonstruowanego pliku PDF publikację, w której kluczowe znaczenie ma jakość obrazu (np. starodruki, atlasy, albumy), to przy skanowaniu w wysokiej rozdzielczości pojedynczy plik PDF będzie miał bardzo duże rozmiary. Nie wszystkie łącza internetowe są w stanie obsłużyć tego typu dokumenty, a jednocześnie nie wszystkie urządzenia dadzą sobie radę z ich otwarciem czy poprawnym wyświetleniem. Jednym ze stosowanych rozwiązań są implementacje w bibliotekach cyfrowych systemów pozwalających na wyświetlanie pojedynczych skanów przy wykorzystaniu otwartoźródłowych technologii pozwalających na przeglądanie plików graficznych o wysokiej rozdzielczości w czasie rzeczywistym (IIIF⁴). Trzeba jednak zauważyć, że dylematy te zasadniczo nie dotyczą Repozytorium, gdyż – jak już wspomniano – udostępnia się w nim przede wszystkim wydawnictwa współczesne, najnowsze oraz w przeważającej mierze *stricte* naukowe, a więc składające się głównie z tekstu, w mniejszym zaś stopniu z grafik. Co więcej, współcześnie wszystkie publikacje przygotowywane do druku są eksportowane do formatu PDF, a zacieśniona współpraca między redakcją Repozytorium a Wydawnictwem Uniwersytetu w Białymstoku pozwala na udostępnianie gotowych plików bez konieczności digitalizacji retrospektywnej. Takie dokumenty mają znacznie

⁴ International Image Interoperability Framework, <https://iiif.io/> (dostęp: 12.06.2023).

mniejsze rozmiary (przy zastosowaniu odpowiedniego poziomu kompresji ewentualnych grafik), gdyż cała warstwa tekstowa i dołączone do niej obiekty, np. wykresy, mają postać wektorową, a nie rastrową. *Summa summarum* format PDF jest wciąż najlepszą opcją w przypadku udostępniania materiałów za pośrednictwem Repozytorium.

Fakt użycia plików w formacie PDF, a także digitalizacja w większości zasobów współczesnego piśmiennictwa naukowego, zdecydowała również o przyjęciu podstawowych parametrów cyfryzowania dokumentów na potrzeby Repozytorium. Bazując na standardach digitalizacji obiektów bibliotecznych określone przez Bibliotekę Narodową, przyjęto dla tekstów monochromatycznych rozdzielczość skanowania na poziomie 600 DPI przy 1 bicie (czerni i bieli), z wyjątkiem sytuacji, gdy w tekście pojawiają się obiekty w skali szarości lub kolorowe (fotografie, wykresy, plany itp.) – wówczas stosujemy przy rozdzielczości 600 DPI 8-bitową skalę szarości oraz 24-bitową głębię kolorów. Decyzja ta wynikała z konieczności ograniczenia wielkości plików. Skany realizowane przy użyciu takich parametrów są odpowiedniej jakości, niezbędnej do czytania przez użytkowników, a jednocześnie ograniczają ilość zapisanych w nich informacji do absolutnego minimum. Konsekwencją takiego działania, w połączeniu z faktem scalenia skanów w plik w formacie PDF, była także decyzja o sposobie przetwarzania każdego dokumentu przy użyciu systemu rozpoznawania tekstu (optical character recognition – OCR), dostarczonego w oprogramowaniu Adobe Acrobat Professional lub ABBYY Fine Reader (oba te programy są stosowane w Pracowni Zbiorów Specjalnych i Digitalizacji). Nikogo nie trzeba przekonywać o konieczności rozpoznawania tekstowego, gdyż jest to jeden z kluczowych czynników zwiększających dostępność treści danego obiektu – jest on odnajdywany przez wyszukiwarki internetowe nie tylko dzięki metadansom zawartym w systemie dSpace, ale także dzięki treści całego dokumentu. Przy czym OCR w przypadku produktu firmy Adobe umożliwia stosowanie specjalnego trybu przetwarzania o nazwie ClearScan⁵. Zasadą działania tego trybu jest nie tylko rozpoznanie warstwy tekstowej, ale również przekształcanie obrazu rastrowego w wektorowy, co w znaczącym stopniu redukuje wielkość pliku, a więc ułatwia jego pobieranie

⁵ R. Borstein, *Better PDF OCR: ClearScan is Smaller, Looks Better*, <https://acrobatusers.com/tutorials/better-pdf-ocr-clearscan-smaller-looks-better/> (dostęp: 21.06.2023).

i przeglądanie. Dzięki temu na pierwszy rzut oka trudno jest odróżnić retrospektywnie zdigitalizowany dokument od plików wyeksportowanych z programów do *desktop publishing* i przeznaczonych do druku.

Podsumowując, Pracownia Digitalizacji Zbiorów odegrała kluczową rolę w budowaniu zasobów Repozytorium dzięki doświadczeniom wynikającym z prac przy Podlaskiej Bibliotece Cyfrowej. Jako standard przyjęto skanowanie w rozdzielczości 600 DPI i stosowanie 1-bitowej głębi kolorów dla tekstów monochromatycznych, 8-bitowej dla obiektów w skali szarości oraz 24-bitowej dla obiektów kolorowych. Do udostępnienia przeznacza się wielostronicowe pliki w formacie PDF, poddane rozpoznawaniu tekstowemu z zastosowaniem trybu ClearScan. Decyzje dotyczące standardów miały na celu zoptymalizowanie jakości udostępnianych dokumentów i wielkości plików, tak aby każdy użytkownik miał do nich nie tylko swobodny dostęp, ale także równy pod względem technologicznym.

Jak wspomniano, praca przy digitalizacji zasobów na potrzeby Repozytorium obejmuje w przytłaczającej większości digitalizację retrospektywną, a więc polegającą na przetworzeniu dokumentu istniejącego tylko wersji drukowanej do wersji cyfrowej. Przy czym, w przeciwieństwie do Podlaskiej Biblioteki Cyfrowej, w Repozytorium udostępniane są tylko dokumenty, których deponowanie odbywa się na zasadach obligatoryjności określonej w zarządzeniach rektorskich, bądź decyzji samego autora. Dlatego większość działań cyfryzacyjnych na rzecz Repozytorium w ciągu minionej dekady była realizowana w trybie *on demand*. Najczęściej kierowane są one przez redakcję Repozytorium, ale nie należą do rzadkości przypadki, gdy autorzy zwracali się bezpośrednio do Pracowni z prośbą o wsparcie techniczne. Liczba takich prośb zwiększyła się w czasie pandemii koronawirusa, gdy wielu depozytariuszy zdecydowało się na upowszechnienie swego dorobku w czasie, gdy praca i dydaktyka były realizowane zdalnie. Bywało też, że redakcja Repozytorium wnioskowała do Pracowni o przeskanowanie danego dokumentu, gdyż plik przesłany przez autora podczas samodzielnego deponowania nie spełniał minimalnych standardów digitalizacyjnych. W początkowym okresie funkcjonowania Repozytorium Pracownia wykonywała także inne usługi związane z udostępnianymi dokumentami, np. podział pliku PDF wydawnictwa zbiorowego lub tomu czasopisma na pojedyncze części i artykuły. W związku z nieregularnością prowadzonych prac nie były one też stale monitorowane pod względem kwantytatywnym. Z punktu widzenia

funkcjonowania Repozytorium znacznie większe znaczenie ma liczba zdeponowanych dokumentów niż wykonanych skanów. Najogólniej rzecz ujmując, w Pracowni wykonuje się na potrzeby Repozytorium do kilku tysięcy skanów rocznie (przykładowo w 2018 roku było to nieco ponad 5000 stron tekstu). Nie chodzi przecież o imponowanie wysokimi statystykami, ale o jakość wykonywanych działań. Zresztą, im bliżej współczesności, tym więcej dokumentów spływa do Repozytorium od razu w wersji cyfrowej (*born digital*), często będącej także podstawą wydania dzieła w tradycyjnej formie drukowanej.

Nie mamy jednak złudzeń, że w którymś momencie Pracownia Zbiorów Specjalnych i Digitalizacji nie będzie miała co robić! Digitalizacja jest działaniem ciągłym, wciąż bowiem zdarzają się sytuacje, w których autorzy publikując w innych wydawnictwach w kraju i za granicą nie są w stanie pozyskać od wydawcy oryginalnego pliku w wersji cyfrowej. Wówczas do akcji wkracza Pracownia Digitalizacji Zbiorów... Należy też pamiętać o spuściznie naukowej Filii Uniwersytetu Warszawskiego w Białymstoku, której cyfryzacja i udostępnienie *on-line* jest aktualnie dyskutowane. Pracy na pewno nam nie zabraknie.

Budowa i uruchomienie Repozytorium odbyło się między innymi przy znaczącym wsparciu funduszy resortowych. W latach 2013–2014 Uniwersytet w Białymstoku był beneficjentem programu Działalność Upowszechniająca Naukę, w ramach którego udało się zrealizować projekt „Poszerzenie zasobów naukowych Biblioteki Uniwersyteckiej im. Jerzego Giedroycia w Białymstoku poprzez rozwój repozytorium instytucjonalnego Uniwersytetu w Białymstoku” (umowa nr 677/P-DUN/2014). Należy zaznaczyć, że działania wpisane w projekt nie obejmowały prac digitalizacyjnych, które były wykonywane doraźnie i na zlecenie redakcji Repozytorium, będącej zobligowaną do udostępnienia pierwszych 900 dokumentów do końca trwania projektu. Dopiero realizowany w latach 2021–2023 projekt finansowany ze środków Ministra Edukacji i Nauki w ramach programu Społeczna Odpowiedzialność Nauki (moduł wsparcia bibliotek) pod tytułem „Rozbudowa otwartych zasobów naukowych Repozytorium Uniwersytetu w Białymstoku” przewidywał odstępstwo od nieregularnej digitalizacji *on demand*. Założeniem projektu było udostępnienie w Repozytorium 1200 różnego typu dokumentów wydanych w Wydawnictwie Uniwersytetu w Białymstoku i Wydawnictwie Temida 2 w latach 1994–2015, które w zdecydowanej większości wymagały wykonania cyfrowych odwzorowań z wersji drukowanych. Mimo dużych doświadczeń kadry Pracowni i wypracowanych od dawna zasad, digitalizacja w ramach

projektu była nie lada wyzwaniem, gdyż w ciągu zaledwie pół roku należało wykonać 40 tys. skanów (a więc średnio ponad 330 skanów dziennie, nie wliczając w to czasu poświęconego na postprodukcję, czyli odpowiednią obróbkę skanów, scalenie w plik PDF, poddanie procesowi rozpoznania tekstu i zapisanie pod nazwą umożliwiającą identyfikację). Ostatecznie w ramach projektu Pracownia wykonała ponad 45 tys. skanów, które łącznie dały wymaganą sumę 1200 dokumentów. Idziemy za ciosem i w 2023 roku na biurko Ministra Edukacji i Nauki trafił kolejny wniosek o dofinansowanie projektu będącego kontynuacją działań w zakresie poszerzania otwartych zasobów naukowych Repozytorium Uniwersytetu w Białymstoku⁶.

Nie ulega więc wątpliwości, że rola digitalizacji w budowie Repozytorium Uniwersytetu w Białymstoku ma istotne znaczenie i ma swój udział w sukcesie tego przedsięwzięcia. Dziś nie jesteśmy już tylko odbiorcami rewolucji cyfrowej – my ją współtworzymy. W takim ujęciu jakość techniczna deponowanych dokumentów odzwierciedla stopień zaangażowania kadry Pracowni w działania cyfryzacyjne oraz właściwe, szerokie rozumienie tego pojęcia. Retrospektywna digitalizacja w przypadku Repozytorium ma bowiem cel bardziej wyspecjalizowany niż w przypadku Podlaskiej Biblioteki Cyfrowej. Dobrze wykonana kopia cyfrowa dzieł będących efektem pracy naukowo-badawczej Uniwersytetu w Białymstoku, udostępniona w formie poprawnie wyświetlającego się pliku o czytelnej warstwie tekstowej i graficznej, jest bez wątpienia jednym z elementów – może nie najważniejszym, ale ważnym – wypełniania misji naszej Uczelni. A jest nią działalność na rzecz dobra społecznego poprzez tworzenie i upowszechnianie wiedzy. Dziś oczywiście nie inaczej, jak tylko za pośrednictwem wersji cyfrowej.

Niech więc szanowny Jubilat – Repozytorium Uniwersytetu w Białymstoku – przyjmie od Kierownika Pracowni Digitalizacji Zbiorów najlepsze życzenia dalszego pomyślnego rozwoju i kolejnych sukcesów w realizacji ambitnych celów! W czym, mam nadzieję, jeszcze niejednokrotnie pomogą Koledzy od digitalizacji.

⁶ 27 czerwca 2023 roku wniosek o numerze BIBL/SP/0040/2023/01 uzyskał pozytywną ocenę i został zakwalifikowany do finansowania.

- Bezzubik E., *Rusza Podlaska Biblioteka Cyfrowa*, „Biuletyn EBIB” 2006, nr 4(74), <http://www.ebib.pl/2006/74/bezzubik.php> (dostęp: 12.06.2023).
- Borstein R., *Better PDF OCR: ClearScan is Smaller, Looks Better*, <https://acrobatusers.com/tutorials/better-pdf-ocr-clearscan-smaller-looks-better/> (dostęp: 21.06.2023).
- Digitalizacja piśmiennictwa*, oprac. i red. D. Paradowski, Warszawa 2010.
- International Image Interoperability Framework*, <https://iiif.io/> (dostęp: 12.06.2023).