

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

## Załącznik Nr 1b do SIWZ Szczegółowy opis przedmiotu zamówienia

(Znak postępowania: K.331-1/2020)

Lp.	Opis produktu	Ilość
1.	Wykonanie i wdrożenie systemu do opracowywania, katalogowania i obsługi OCR zeskanowanych numerów zasobów bibliotecznych oraz prezentowania ich w chmurze	1

### **Specyfikacja wykonania i wdrożenia systemu do opracowywania, katalogowania i obsługi OCR zeskanowanych numerów zasobów bibliotecznych oraz prezentowania ich w chmurze**

System ma służyć do opracowywania, katalogowania i ocr-owania zeskanowanych w odrębnym procesie numerów czasopism i artykułów czasopisma, książek oraz prezentowania ich w chmurze dostępnej dla wszystkich internautów. Obiekty mogą być łączone przez administratorów w grupy (wydarzenia, tematy, etc.). Każdy z obiektów może należeć do wielu grup. Zarówno opis bibliograficzny czasopisma, książki, jak i opis bibliograficzny artykułu muszą być oparte na formacie MARC21. System ma umożliwiać zintegrowanie aplikacji z istniejącą bazą opisów bibliograficznych w formacie MARC21 w celu pobierania z niej opisów oraz tworzenie, uzupełnianie i modyfikowanie kolejnych rekordów bibliograficznych również w oparciu o format MARC21.

System ma mieć możliwość grupowego przetwarzania nowych obiektów (skanów) – tworząc repozytorium cyfrowe, które będzie dostępne dla katalogerów w celu opracowania poszczególnych obiektów.

Przewidywana ilość obiektów podpiętych do pierwszej bazy – 2200 obiektów (jako jeden obiekt jest traktowany jeden numer czasopisma). Przewidywany rozmiar obiektu (od 250 do 900 MB), ilość skanów w obiekcie – 5-8. Szacowana ilość wyświetleń obiektów z zasobu – 1500-2000 na rok w odniesieniu do każdego rocznika oddzielnie.

Przewidywana ilość wszystkich obiektów podpiętych do drugiej bazy – 30558 (jako jeden obiekt jest traktowany jeden rekord, zawierający opis bibliograficzny zdigitalizowanego numeru czasopisma lub książki). Przewidywany rozmiar całej bazy wynosi 3 TB, ilość skanów 345210. W bazie tej przewidywana ilość obiektów będących czasopismami to 23644 (jako jeden obiekt jest traktowany jeden rekord, zawierający opis bibliograficzny zdigitalizowanego numeru czasopisma). Przewidywany rozmiar obiektów, będących czasopismami, umieszczonych w tej bazie to 1.5 TB, ilość skanów 170912.

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

**Podsumowując przewidywana wielkość dla pierwszej bazy ok. 1,9 TB. Natomiast wielkość dla drugiej bazy 3 TB, w tym 1,5 TB rozmiar w części dotyczącej czasopism.**

Obiekty będą pogrupowane w poszczególne roczniki, ale system ma umożliwić dodanie w terminie późniejszym również innych tytułów.

System musi obejmować bezterminowych 10 licencji na stanowiska katalogerów oraz 3 administratorów systemu. Licencje obejmują pięcioletnią gwarancję oraz prawo do aktualizacji do najnowszej wersji systemu wraz z asystą techniczną.

System musi umożliwiać import danych (opisów bibliograficznych) z innych systemów posiadanych przez Zamawiającego.

Przeglądanie katalogu musi być dostępne dla wszystkich zainteresowanych internautów, z tym że część skatalogowanych zbiorów może być dostępna tylko w intranecie Zamawiającego.

System musi być dostępny przez 99,9% czasu w skali roku, dopuszczalne są 2-3 godzinne przerwy serwisowe w wymiarze nie częstszym niż raz w miesiącu, po uzgodnieniu z Zamawiającym

### **Wymagania techniczne**

Po podpisaniu umowy Wykonawca uruchomi system w chmurze

System ma umożliwiać udostępnianie części zbiorów tylko na określonych komputerach rozpoznawanych np. poprzez adres IP.

System ma mieć zaimplementowany mechanizm OCR, automatycznie dodający warstwę tekstową do plików pdf.

System ma mieć możliwość zmigrowania na infrastrukturę wskazaną przez zamawiającego – możliwość migracji bezpłatna.

System ma mieć zaimplementowany protokół OAI-PMH, zgodny z formatem Dublin Core.

System ma działać przez przeglądarki internetowe (Firefox, Opera, Chrome) na wszystkich najpopularniejszych systemach operacyjnych.

System ma spełniać wymagania zawarte w:

- Rozporządzeniu Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności, w zakresie minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej
  - WCAG (Web Content Accessibility Guidelines) 2.0 na poziomie AA

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

- Normie PN/ISO 27001-2007 z zastrzeżeniem, że w systemie nie będą przetwarzane żadne dane osobowe
- Skali „5 Star Open Data” dla poziomu 5 gwiazdek

## Dostęp dla czytelników

### Wyszukiwanie

Wszystkie poniższe metody wyszukiwania odnoszą się zarówno do przeszukiwania dokumentów cyfrowych (pełnych skanów czasopisma), jak i opisów poszczególnych artykułów z czasopism. W obu przypadkach czytelnik będzie mógł wybrać również czy chce przeszukiwać dokumenty pełnotekstowo.

System umożliwia następujące metody wyszukiwania:

- **Wyszukiwanie proste** – jedno pole tekstowe umożliwiające formułowanie prostych zapytań odnoszących się do jednej lub dwóch metadanych (np. autor, tytuł; autor, słowo kluczowe; tytuł, data). Przy wyszukiwaniu prostym czytelnik może dokonać wyboru czy przeszukuje opisy bibliograficzne czy chce również przeszukać dokumenty po treści.
- **Wyszukiwanie zaawansowane** – czytelnik ma możliwość wyszukiwania po konkretnych metadanych (tytuł, autor, miejsce wydania, słowo kluczowe, wydawca, częstotliwości wydawania (dla czasopism), sygnatury). Podczas wpisywania system podpowiada czytelnikowi dostępne hasła. Czytelnik może dodawać kolejne kryteria do wyszukiwania i łączyć je ze sobą za pomocą podstawowych operatorów logicznych: ORAZ (służącego do bezwzględnego łączenia poszukiwanych metadanych) i LUB (pozwalającego na uzyskiwanie w wynikach wyszukiwania zarówno kombinacji poszukiwanych fraz, jak i ich wystąpień odrębnych). W przypadku dat mogą również wystąpić operatory OD i DO, pozwalające na określenie okresu czasu interesującego czytelnika.
- **Filtrowanie wyników za pomocą faset/deskryptorów** – wyniki wyszukiwania można filtrować korzystając z różnych faset, np.: autor, źródło, wydawca (nazwa wydawnictwa lub imię i nazwisko drukarza), język, częstotliwość (pozwala filtrować czasopisma ze względu na częstotliwość ukazywania się, np. dzienniki, tygodniki), gatunek, słowa kluczowe oraz czas. Faseta czas służy do filtrowania wyników według czasu powstania dokumentu. Aby to zrobić, należy samodzielnie wpisać wybrany przedział czasowy. Zakres dat można wpisać z dokładnością do dnia, ale nie jest to konieczne; w przypadku wpisania tylko daty rocznej, wartości dni i miesięcy

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

domyślnie ustawią się na 1 stycznia. Okienko filtrowania według czasu znajduje się na samym dole panelu filtrów fasetowych.

## **Prezentacja rekordów**

Podstawową formą prezentacji dokumentów cyfrowych jest widok „kafelków”, czyli okładek/pierwszych stron czasopism z podpisem zawierającym tytuł wraz numerem i datą wydania.

Podstawową formą prezentacji artykułów jest przypis bibliograficzny z charakterystycznymi elementami: Nazwisko i imię autora artykułu. Tytuł artykułu. "Tytuł czasopisma" rok, numer czasopisma, numery stron zajęte przez artykuł.

Przy prezentacji pojedynczego dokumentu cyfrowego poniżej pojawiają się artykuły zawarte w tym numerze czasopisma, autorzy, którzy napisali do niego artykuły oraz hasła przedmiotowe użyte w opisach artykułów: osoby (pole 600 MARC21), instytucje/organizacje (pole 610), imprezy (pole 611). Autorzy mogą być prezentowani w postaci fotografii. Po wybraniu autora czytelnik widzi tylko artykuły tego autora napisane do wcześniej wybranego czasopisma. Analogicznie jak z autorzy wyświetlane będą hasła przedmiotowe.

Czytelnicy będą mogli zmienić formę prezentacji poszczególnych danych na formy przygotowane przez administratorów.

W systemie dostępne będą również kolekcje, czyli zbiory zebrane tematycznie. Kolekcje będą prezentowane na stronie serwisu. Daje to możliwość prezentowania zasobów, popularyzacji wybranych fragmentów dziedzictwa kulturowego Podlasia. Przykładami kolekcji mogą być: artykuły dotyczące ważnych wydarzeń, instytucji itp.

Dokumenty cyfrowe zawierające całe numery czasopism (w formacie pdf) oraz wybrane artykuły będą wyświetlane w różnych trybach – tryb pełnoekranowy, wyświetlanie pojedynczej strony, wyświetlanie dwóch stron obok siebie, wyświetlanie miniatur stron, wyświetlanie opisu bibliograficznego obok obiektu. W trybie pełnoekranowym czytelnik będzie miał możliwość płynnego powiększania przeglądanych obiektów, przesuwania i powiększania dużego obrazu o wysokiej rozdzielczości lub dużej kolekcji obrazów przy skróconym czasie potrzebnym na początkowe obciążenie, poprzez pobieranie tylko oglądanego regionu lub tylko w rozdzielczości, w jakiej jest wyświetlany.

## **Aplikacja czytelnika**

Wyszukiwanie i przeglądanie obiektów musi być również dostępne za pomocą dostarczonych aplikacji na urządzenia działające pod kontrolą systemów operacyjnych Android w wersji 7.0 i wyższej oraz iOS w wersji 12.0 i wyższej.

## **Dostęp dla redaktora**

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

Dostęp do systemu dla redaktora musi być udostępniony pod innym adresem niż dostęp dla czytelnika. Każdy użytkownik (redaktor) może mieć nadaną jedną z trzech kategorii uprawnień:

- Administrator – ma dostęp do wszystkich tytułów i roczników. Może dodawać nowych użytkowników i zmieniać im uprawnienia. Może edytować dostępność poszczególnych tytułów i roczników. Może być kilku administratorów w jednej instancji.
- Moderator – ma dostęp do wszystkich tytułów i roczników i może poprawiać wszystkie rekordy. Nie może edytować użytkowników i ich uprawnień. Może zarządzać tytułami i rocznikami i ich dostępnością.
- Kataloger – może dodawać rekordy tylko do określonych przez administratora tytułów i roczników.

Uprawnienia są stopniowane. Każdy moderator ma uprawnienia katalogera, a każdy administrator ma wszystkie uprawnienia.

### **Dodawanie nowych rekordów**

Podstawową metodą dodawania nowych dokumentów cyfrowych, nowych artykułów i nowych haseł są formularze. Formularz powinno się dać wypełnić bez odrywania rąk z klawiatury (do czasu dodawania plików cyfrowych). Formularz będzie się składał z trzech kroków:

1. Dane podstawowe
2. Hasła z kartotek
3. Dokumenty cyfrowe

Krok 1 to wpisanie podstawowych informacji, potrzebnych do poprawnego wypełnienia rekordu. W komentarzach poniżej każdego pola są informacje, gdzie wpisane informacje mają trafić w formacie MARC21. W tym kroku występują dwa pola powiązane z Kartotekami: autor i współautor. Gdy użytkownik zaczyna wpisywać wartość, system filtruje na bieżąco wyniki. Przy artykułach, w tym miejscu wiążemy rekord z konkretnym dokumentem cyfrowym.

Krok 2 to dodawanie haseł z kartotek. Kartoteki są prowadzone dla pól grupy 600 formatu MARC21. Gdy użytkownik zaczyna wpisywać wartość, system filtruje na bieżąco wyniki.

Krok 3. W tym kroku dodajemy dokumenty cyfrowe.

Pozostałymi metodami dodawania nowych rekordów jest importowanie metadanych z zewnętrznych źródeł, dodawanie w formularzu MARC21 oraz dodawanie grupowe, gdzie wskazujemy na dokument cyfrowy oraz na plik z metadanymi w formacie mrc.

Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Podlaskiego na lata 2014-2020.

Użytkownicy mają możliwość wyszukiwania wcześniej dodanych elementów i edytowania ich.

## Statystyki

Użytkownicy mają dostęp do następujących statystyk:

- Liczba dostępnych dokumentów cyfrowych, wprowadzonych artykułów i dodanych haseł przedmiotowych w zadanym okresie
- Liczba dodanych dokumentów cyfrowych, wprowadzonych artykułów i dodanych haseł przedmiotowych w zadanym okresie
- Liczba wyszukiwań prostych i zaawansowanych w zadanym okresie
- Liczba unikalnych odwiedzających oraz liczba sesji w zadanym okresie
- Liczba wyświetlonych dokumentów cyfrowych, wprowadzonych artykułów i dodanych haseł przedmiotowych w zadanym okresie z możliwością ograniczenia do określonych tytułów i roczników
- Lista najpopularniejszych dokumentów cyfrowych, wprowadzonych artykułów i dodanych haseł przedmiotowych w zadanym okresie z możliwością ograniczenia do określonych tytułów i roczników
- Liczba wyświetleń poszczególnych dokumentów cyfrowych, wprowadzonych artykułów i dodanych haseł przedmiotowych w zadanym okresie