

**PAN Biblioteka Kórnicka**

ul. Zamkowa 5  
62-035 Kórnik

Katarzyna Woźniak

tel.: 531 983 633

mail: katarzyna.wozniak@bk.pan.pl

**Dotyczy: Odpowiedzi na pytania do uczestników wstępnych konsultacji rynkowych.**

**Szanowni Państwo,**

w odpowiedzi na otrzymane pytania przekazujemy nasze odpowiedzi:

**1. WCAG 2.2 – technologie, narzędzia, doświadczenia**

Rekomendujemy, aby warstwa interfejsu użytkownika była projektowana w modelu AI-UI + klasyczny frontend. Część klasyczna może wykorzystywać nowoczesne frameworki (React / Angular / Vue) oraz spójny design system komponentów, natomiast warstwa AI-UI powinna zapewniać konwersacyjny dostęp do funkcji systemu (chat/voice), co może istotnie poprawiać dostępność dla osób z różnymi rodzajami niepełnosprawności. Komponenty warto definiować jako „accessible-by-design” (ARIA, nawigacja klawiaturą, responsywność, kontrast, focus management), a repozytorium design-systemu – generować i utrzymywać m.in. z użyciem rozwiązań takich jak GENESIS.ai oraz Cameleoo AI (AI Low-Code Platform).

W procesie zapewniania zgodności z WCAG 2.2 AA rekomendujemy traktowanie dostępności jako ciągłego, zautomatyzowanego procesu: testy dostępności (axe, Pa11y, WAVE) dobrze jest włączyć bezpośrednio w CI/CD, uzupełniając je o wykorzystanie modeli AI do automatycznego sugerowania opisów alternatywnych, wykrywania anty-patternów dostępności w kodzie oraz analizy nagrań z testów z udziałem osób z niepełnosprawnościami (np. identyfikacji miejsc, w których użytkownicy napotykały trudności). Dzięki takiemu podejściu dostępność może stać się stale monitorowanym parametrem jakości, a nie jednorazowym audytem.

**2. Integracja z Węzłem Krajowym (login.gov.pl / mObywatel)**

Rekomendujemy, aby bezpieczna integracja z Węzłem Krajowym była realizowana przy wykorzystaniu Keycloak jako brokera tożsamości, działającego w oparciu o standardy OpenID Connect / SAML 2.0. Aplikacje biblioteczne mogłyby komunikować się z Keycloak w spójny

**allclouds.pl sp. z o.o.**

ul. Jutrzenki 139, 02-231 Warszawa • www.allclouds.pl • office@allclouds.pl • phone: +48 22 100 43 80 • fax: +48 22 100 43 84  
NIP: PL5223052539 • REGON: 363597531 • KRS: 0000598708 • Akta prowadzone przez Sąd Rejonowy dla m.st. Warszawy w Warszawie  
XIV Wydział Gospodarczy • Kapitał zakładowy w wysokości 1 045 000 zł wniesiono w całości  
PN-EN ISO 9001 • PN-EN ISO/EIC 27001 • PN-EN ISO 14001 • PN-EN ISO/EIC 22301 • PN-EN ISO/EIC 27017 • PN-EN ISO/EIC 27018

sposób, natomiast integracja z login.gov.pl / mObywatel byłaby skoncentrowana w jednym, kontrolowanym komponentcie. W podejściu AI-by-design sugerujemy traktowanie SSO jako usługi platformowej, w której konfiguracje realmów, klientów i polityk bezpieczeństwa są opisywane deklaratorywnie oraz generowane z użyciem narzędzi takich jak GENESIS.ai / Cameleo AI – co ułatwia wersjonowanie, audyt i automatyczne odtwarzanie środowisk.

Dla zapewnienia wydajności i bezpieczeństwa rekomendujemy zastosowanie monitoringu wspomaganego technikami AI: modele analizy anomalii mogą śledzić statystyki logowań, nietypowe wzorce użycia i potencjalne ataki (credential stuffing, brute-force, nietypowe przepływy między IdP). Równoległe można analizować doświadczenie użytkownika – AI może agregować sygnały z logów, błędów i feedbacku, wskazując miejsca, w których proces logowania jest zbyt złożony lub nieintuicyjny. Sam proces uwierzytelniania sugerujemy traktować jako część AI-UI: wirtualny asystent może prowadzić użytkownika przez problemy z loginem, wyjaśniać komunikaty i pomagać w wyborze metody logowania, przy zachowaniu zgodności z WCAG 2.2.

### 3. MARC21, Kronik@, integracja z repozytoriami

Rekomendujemy utrzymywanie MARC21 jako formatu źródłowego, przy jednoczesnym traktowaniu go jako punktu wyjścia do wielowarstwowego, wspieranego przez AI przetwarzania metadanych. Warstwę integracyjną sugerujemy realizować w oparciu o narzędzia takie jak GENESIS.ai / Cameleo AI, które mogą dostarczyć łańcuchy przetwarzania do konwersji MARC21 ⇔ MARCXML ⇔ RDF ⇔ JSON-LD oraz inne profile (np. EDM dla Europeany). Pipeline’y można opisywać w sposób deklaratorywny (reguły mapowania, profile kolekcji, rozszerzenia lokalne), a moduły AI wykorzystywać do automatycznej klasyfikacji tematów, uzupełniania haseł wzorcowych, wykrywania niespójności oraz semantycznego wzbogacania rekordów (ontologie, powiązania autorów, miejsc, instytucji).

Integrację z Kronik@ i podobnymi repozytoriami rekomendujemy traktować jako element szerszego ekosystemu usług semantycznych. W warstwie wyszukiwania i eksploracji zasobów można zastosować wyszukiwanie semantyczne oparte na wektorach oraz podejście RAG (Retrieval-Augmented Generation), co umożliwi użytkownikom zadawanie pytań w języku naturalnym (np. o zdarzenia, osoby, miejsca), a systemowi – łączenie odpowiedzi z rekordami MARC, materiałami cyfrowymi i publikacjami naukowymi.

### 4. OCR – model lokalny czy chmurowy?

Rekomendujemy rozważenie hybrydowego modelu AI-by-design, w którym różne rodzaje zasobów obsługiwane są różnymi wariantami technologii OCR:

- dla materiałów wrażliwych, unikatowych i objętych restrykcyjnymi politykami RODO i archiwizacji – lokalne (on-premises) silniki OCR oraz lokalnie wdrożone modele LLM,

- dla materiałów standardowych, o niższym poziomie wrażliwości – rozwiązania chmurowe, o ile spełniają wymagania formalne (RODO, umowy powierzenia, lokalizacja danych).

W szczególności można rozważyć wdrożenie on-premises modeli LLM wspierających proces OCR, pod warunkiem zapewnienia odpowiedniej infrastruktury (serwery lub klastry serwerów GPU, odpowiednia moc obliczeniowa, pamięć masowa, chłodzenie, zasilanie awaryjne). Tego typu modele mogą realizować zadania post-processingu (korekta błędów językowych w języku polskim, rozpoznawanie układu strony, identyfikacja nazw, dat, miejsc, klasyfikacja dokumentów), a także wspierać automatyczne generowanie metadanych.

Niezależnie od wyboru wariantu OCR rekomendujemy traktowanie go jako elementu całego łańcucha przetwarzania z użyciem AI: klasyfikacja dokumentu, rozpoznanie struktury, OCR, korekta LLM, detekcja jakości, generacja metadanych oraz zapis do formatów PDF/A i ALTO XML. Zgodność z RODO może być wspierana dzięki jasnemu podziałowi ról (trenowanie, inferencja, archiwizacja), kontroli lokalizacji przetwarzania danych oraz pełnemu logowaniu przebiegu procesu. Przy kolekcjach historycznych warto rozważyć dostrajanie (fine-tuning) modeli na wybranych zasobach, wykorzystując tę samą infrastrukturę GPU, która obsługuje inne moduły AI biblioteki.

## 5. Zastosowanie AI – chatbot, rekomendacje, automatyzacja

Rekomendujemy traktowanie AI jako domyślnego sposobu interakcji z systemem bibliotecznym (AI-UI), a nie tylko dodatkowego modułu. Kluczowym elementem mogłoby być uruchomienie Wirtualnego Asystenta Czytelnika opartego na dużych modelach językowych (LLM), zintegrowanych z zasobami biblioteki poprzez RAG oraz wyszukiwanie semantyczne. Asystent taki mógłby wspierać użytkowników w wyszukiwaniu publikacji, budowaniu list literatury, zrozumieniu kontekstu naukowego, obsłudze konta czy zrozumieniu regulaminów i usług biblioteki, a także prowadzić ich krok po kroku przez poszczególne procesy.

System rekomendacyjny polecamy oprzeć na połączeniu klasycznych metod (analiza zachowań użytkowników, współwystępowanie wypożyczeń) z embeddingami i modelami AI rozumiejącymi treść publikacji (abstrakty, pełne teksty, metadane). Całość mogłaby zostać zrealizowana na platformach takich jak GENESIS.ai / Cameleo AI, które ułatwiają generowanie i utrzymanie mikroservisów AI (asystent, rekomendacje, automatyzacja procesów katalogowych) w sposób spójny, wersjonowany i przenaszalny między środowiskami.

Z doświadczeń sektora kultury i nauki wynika, że warto szczególnie zadbać o: kontrolę jakości odpowiedzi (logowanie, możliwość weryfikacji źródeł, testy halucynacji), jasne zasady odpowiedzialności człowiek–AI (AI jako asystent, nie decydent) oraz transparentną komunikację z użytkownikami, że korzystają z wirtualnego asystenta.

### allclouds.pl sp. z o.o.

ul. Jutrzenki 139, 02-231 Warszawa • www.allclouds.pl • office@allclouds.pl • phone: +48 22 100 43 80 • fax: +48 22 100 43 84  
NIP: PL5223052539 • REGON: 363597531 • KRS: 0000598708 • Akta prowadzone przez Sąd Rejonowy dla m.st. Warszawy w Warszawie  
XIV Wydział Gospodarczy • Kapitał zakładowy w wysokości 1 045 000 zł wniesiono w całości  
PN-EN ISO 9001 • PN-EN ISO/EIC 27001 • PN-EN ISO 14001 • PN-EN ISO/EIC 22301 • PN-EN ISO/EIC 27017 • PN-EN ISO/EIC 27018

## 6. Koszty i ryzyka

Rekomendujemy, aby w podejściu AI-by-design koszty były dzielone na cztery główne grupy:

1. licencje i usługi (silniki OCR, komponenty AI, ewentualne biblioteki komercyjne),
2. prace wytwórcze i integracyjne (analiza, projekt, implementacja modułów, integracje SSO, repozytoria, MARC21),
3. warstwa AI-UI i moduły AI (Wirtualny Asystent, rekomendacje, automatyzacja MARC21, łańcuchy przetwarzania OCR+LLM),
4. utrzymanie i ewolucja (SLA, aktualizacje bezpieczeństwa, rozwój modeli, utrzymanie infrastruktury, w tym GPU).

W obszarze ryzyk rekomendujemy zwrócić szczególną uwagę na: niedoszacowanie złożoności integracji AI z istniejącym systemem, ryzyko lock-in technologicznego wobec dostawców OCR/AI, aspekty bezpieczeństwa i poufności danych (logi zapytań, dane treningowe) oraz zarządzanie zachowaniem modeli (halucynacje, zmiany jakości po aktualizacjach). Jako środki zaradcze sugerujemy: wybór rozwiązań opartych na otwartych standardach, architekturę modułową, etapowe wdrażanie (pilotáže, rollout modułowy), formalne przeglądy architektury oraz wprowadzenie mechanizmów zarządzania wykorzystaniem AI (AI governance) (polityki, rejestr modeli, procedury walidacji).

W zakresie testów bezpieczeństwa rekomendujemy zarówno klasyczne testy (testy penetracyjne, SAST, DAST, skanowanie komponentów), jak i ukierunkowane na AI: testy typu red-team (prompt-injection, próby wyprowadzenia danych, obchodzenie kontroli dostępu poprzez asystenta) oraz testy odporności na błędne lub złośliwe dane wejściowe.

## 7. Harmonogram wdrożenia

Rekomendujemy harmonogram oparty na etapowym wdrożeniu AI-by-design, który mógłby wyglądać następująco:

1. Analiza przedwdrożeniowa i warsztaty przypadków użycia AI (ok. 2–3 miesiące) – inwentaryzacja systemu, procesów i danych, identyfikacja scenariuszy AI o najwyższej wartości, opracowanie architektury docelowej (AI-UI, backend, integracje).
2. Projektowanie i prototypowanie (ok. 2–3 miesiące) – makiety interfejsów (UI + AI-UI), prototypy w GENESIS.ai / Cameleoo AI (np. pierwsza wersja Wirtualnego Asystenta Czytelnika), definicja łańcuchów przetwarzania danych pod OCR i AI.
3. Implementacja i integracje modułowe (ok. 6–9 miesięcy) – rozwój frontendu, backendu i API, wdrożenie OCR + post-processing AI, integracje z Węzłem Krajowym, repozytoriami (Kronik@), systemem bibliotecznym, budowa i integracja modułów AI (asystent, rekomendacje, automatyzacja MARC21).
4. Testy i przygotowanie organizacyjne (ok. 2–3 miesiące) – testy funkcjonalne, wydajnościowe, bezpieczeństwa i WCAG (w tym testy AI-UI), walidacja jakości

### allclouds.pl sp. z o.o.

ul. Jutrzenki 139, 02-231 Warszawa • www.allclouds.pl • office@allclouds.pl • phone: +48 22 100 43 80 • fax: +48 22 100 43 84  
NIP: PL5223052539 • REGON: 363597531 • KRS: 0000598708 • Akta prowadzone przez Sąd Rejonowy dla m.st. Warszawy w Warszawie  
XIV Wydział Gospodarczy • Kapitał zakładowy w wysokości 1 045 000 zł wniesiono w całości  
PN-EN ISO 9001 • PN-EN ISO/EIC 27001 • PN-EN ISO 14001 • PN-EN ISO/EIC 22301 • PN-EN ISO/EIC 27017 • PN-EN ISO/EIC 27018

odpowiedzi AI, przygotowanie dokumentacji i szkolenia (bibliotekarze, administratorzy, redaktorzy treści).

5. Wdrożenie produkcyjne i stabilizacja (ok. 1–2 miesiące) – uruchomienie modułów w środowisku produkcyjnym, intensywny monitoring zachowania modeli AI i interfejsu, zdefiniowanie roadmapy dalszego rozwoju na podstawie feedbacku użytkowników.

## 8. Wymagana infrastruktura

Rekomendujemy zaprojektowanie infrastruktury w sposób wspierający zarówno klasyczny system biblioteczny, jak i warstwę AI-by-design. Proponowany model obejmuje:

- warstwę aplikacyjną i API – serwery (fizyczne lub VM/kontenery) dla backendu, frontendu i serwisów integracyjnych (SSO, MARC21, repozytoria),
- warstwę danych – wydajne, redundantne macierze dla baz danych, indeksów wyszukiwania, plików PDF/A, ALTO XML oraz logów, z mechanizmami backupu i długoterminowej archiwizacji,
- warstwę AI (compute) – węzły CPU/GPU przeznaczone do inferencji modeli (asystent, rekomendacje, post-processing OCR) oraz okresowego trenowania / dostrajania modeli,
- warstwę platformową – orkiestrację kontenerów (np. Kubernetes), centralny monitoring (metryki, logi, trace), system zarządzania konfiguracją i tajemnicami, system SIEM do obsługi bezpieczeństwa.

Rekomendujemy również rozważenie architektury hybrydowej (on-prem / cloud), w której poszczególne komponenty – zwłaszcza modele AI i repozytoria danych treningowych – mogą być przenoszone między środowiskami w zależności od wymogów RODO i polityk instytucji, bez konieczności zmiany warstwy logiki biznesowej.

## 9. Przewidywany podział kosztów

Rekomendujemy model podziału kosztów oparty na blokach funkcjonalno-technologicznych, z wyraźnym wydzieleniem komponentów AI-by-design, obejmujący m.in.:

- licencje (moduły systemu, OCR, komponenty AI, biblioteki komercyjne),
- prace analityczne i projektowe (w tym identyfikacja use-cases AI, projekt AI-UI, architektura integracyjna),
- prace UX / UCD (projekt interfejsu klasycznego i konwersacyjnego, testy z użytkownikami, prototypy),
- szkolenia (obsługa systemu, modułów AI, zasady zarządzania wykorzystaniem AI (AI governance)),
- frontend / warstwa interfejsu (implementacja design systemu, integracja z AI-UI),

- backend, API, integracje (systemy biblioteczne BK PAN, Węzeł Krajowy, repozytoria, MARC21).

W obszarze Wirtualnego Asystenta Czytelnika sugerujemy dodatkowy podział na: moduł NLP / LLM, moduł wyszukiwania i eksploracji zasobów (RAG, wektory), moduł interfejsu użytkownika, moduł raportowania i eksportu danych, moduł integracji z systemem bibliotecznym, moduł administracji dla bibliotekarzy oraz moduł automatycznego wspomaganie tworzenia rekordów MARC21 (ekstrakcja z PDF/OCR + sugestie AI).

Dalsze kategorie kosztów, które rekomendujemy uwzględnić, to: moduł OCR (silnik, pipeline, integracje, korekta AI), moduł Konserwatora (workflow cyfrowej konserwacji), moduł AI (usługi wspólne – embeddingi, RAG, klasyfikacje), trenowanie modułu AI (przygotowanie danych, eksperymenty, walidacja), integracja z Węzłem Krajowym i obsługa SSO, moduł Analityczny (statystyki, raportowanie), migracja i integracje MARC21, testy funkcjonalne, testy WCAG, audyt bezpieczeństwa (w tym testów typu red teaming ukierunkowanych na komponenty AI), utrzymanie roczne systemu (helpdesk, rozwój, ewolucja modeli) oraz utrzymanie roczne infrastruktury (serwery, storage, backup, monitoring, licencje platformowe).

Taki model może ułatwiać Zamawiającemu elastyczne skalowanie zakresu projektu – np. rozpoczęcie od nowego interfejsu i Wirtualnego Asystenta, a następnie sukcesywne rozszerzanie o OCR, automatyzację MARC21 i bardziej zaawansowane analityki.

### III. Dodatkowe rekomendacje allclouds.pl

allclouds.pl może wnieść kompetencje w zakresie projektowania architektury integracyjnej (API, SSO, integracje z repozytoriami), projektowania i wdrażania rozwiązań AI dla sektora regulowanego (RAG, chatboty, wyszukiwanie semantyczne) oraz zapewnienia zgodności z wymaganiami WCAG 2.2 i standardami bezpieczeństwa. Z perspektywy wykonawcy rekomendujemy podejście iteracyjne, w którym najpierw dostarczane są moduły o najwyższej wartości dla użytkowników (np. nowy interfejs i wyszukiwanie), a następnie kolejne funkcje (OCR, AI, integracje z repozytoriami) są sukcesywnie dokładane w oparciu o wyniki testów i feedback użytkowników.

### Z poważaniem

Wojciech Biernat

e-mail: [office@allclouds.pl](mailto:office@allclouds.pl)

#### allclouds.pl sp. z o.o.

ul. Jutrzenki 139, 02-231 Warszawa • [www.allclouds.pl](http://www.allclouds.pl) • [office@allclouds.pl](mailto:office@allclouds.pl) • phone: +48 22 100 43 80 • fax: +48 22 100 43 84  
NIP: PL5223052539 • REGON: 363597531 • KRS: 0000598708 • Akta prowadzone przez Sąd Rejonowy dla m.st. Warszawy w Warszawie  
XIV Wydział Gospodarczy • Kapitał zakładowy w wysokości 1 045 000 zł wniesiono w całości  
PN-EN ISO 9001 • PN-EN ISO/EIC 27001 • PN-EN ISO 14001 • PN-EN ISO/EIC 22301 • PN-EN ISO/EIC 27017 • PN-EN ISO/EIC 27018