

Nazwa dokumentu: Opis założeń projektu informatycznego pn. „FBC.AI: Cyfrowa dostępność i ponowne wykorzystanie zasobów nauki wspierane sztuczną inteligencją” - wnioskodawca: Minister Nauki i Szkolnictwa Wyższego, beneficjent: Instytut Chemii Bioorganicznej Polskiej Akademii Nauk — Poznańskie Centrum Superkomputerowo-Sieciowe					
Lp.	Organ wnoszący uwagi	Jednostka redakcyjna, do której wnoszone są uwagi	Treść uwagi	Propozycja zmian zapisu	Odniesienie do uwagi
1.	RA IT	Uwaga ogólna	Jako Beneficjenta wskazano Instytut Chemii Bioorganicznej Polskiej Akademii Nauk – Poznańskie Centrum Superkomputerowo-Sieciowe, podczas gdy Poznańskie Centrum Superkomputerowo-Sieciowe stanowi jedynie jednostkę organizacyjną Instytut Chemii Bioorganicznej Polskiej Akademii Nauk i nie posiada odrębnej osobowości prawnej.	Proszę o analizę i korektę opisu założeń.	Zmieniono nazwę Beneficjenta.
2.	RA IT	Uwaga ogólna	Federacja Bibliotek Cyfrowych agreguje dane z wielu bibliotek cyfrowych i repozytoriów, w tym licznych instytucji wykorzystujących system dLibra oraz innych podmiotów udostępniających zasoby w FBC. Proszę o wyjaśnienie, dlaczego współpracę partnerską w projekcie ograniczono akurat do Uniwersytetu Jagiellońskiego i Uniwersytetu Wrocławskiego oraz jaki będzie wpływ budowanego rozwiązania na pozostałe podmioty korzystające z systemu dLibra i/lub udostępniające zasoby za pośrednictwem FBC.	Proszę o analizę i korektę opisu założeń.	Wybór Uniwersytetu Jagiellońskiego i Uniwersytetu Wrocławskiego jako Partnerów projektu wynika z ich potencjału zasobowego, organizacyjnego i technologicznego oraz z zakresu wykorzystywanych przez nie środowisk digitalizacji i udostępniania danych. Wspólnie obie jednostki posiadają odpowiedniej wielkości, różnorodne i reprezentatywne zasoby cyfrowe, kompleksowe procesy digitalizacji oraz długoletnie doświadczenie w automatyzacji prac, podnoszeniu jakości danych i optymalizacji kosztów opracowania oraz udostępniania zbiorów. Uniwersytet Jagielloński prowadzi jedną z największych bibliotek cyfrowych w Polsce. Uniwersytet Wrocławski rozwija jedną z najstarszych bibliotek cyfrowych w Polsce. Obie jednostki posiadają zaawansowane platformy

					<p>przetwarzania i wzbogacania danych oraz heterogeniczne środowiska technologiczne - odpowiednio dLab UJ oraz IZZ UWr - umożliwiające przeprowadzenie pełnego procesu objętego projektem zgodnie z celami działania FERC 2.3.</p> <p>Budowane rozwiązanie zostanie opracowane w formie modułu rozszerzającego/integracyjnego FBC i udostępnione jako otwarte interfejsy komunikacyjne, umożliwiające podłączanie innych bibliotek i repozytoriów cyfrowych, niezależnie od wykorzystywanego przez nie oprogramowania, pod warunkiem dostosowania lokalnego środowiska do opublikowanej specyfikacji integracyjnej. W związku z powyższym integracja będzie możliwa do wykonania przez dowolną bibliotekę lub repozytorium cyfrowe.</p> <p>Modyfikacje systemów UWr i UJ zostaną wykonane wyłącznie w konkretnych, eksploatowanych przez UJ i UWr instalacjach, w zakresie niezbędnym do realizacji zaplanowanych przypadków użycia oraz w formie modułów rozszerzających. Rozwiązania te będą pełniły funkcję referencyjnych przypadków użycia. Na podstawie doświadczeń z ich wdrożenia zostaną opisane wzorce integracji, wymagania dotyczące danych, organizacja procesów oraz dobre praktyki stosowania sztucznej inteligencji w digitalizacji i przygotowaniu zasobów do ponownego wykorzystania. Planowane jest opracowanie katalogu dobrych praktyk digitalizacji wspieranej przez AI oraz upowszechnienie wypracowanego know-how wśród m.in. instytucji dostarczających zasoby do FBC.</p>
--	--	--	--	--	---

3.	RA IT	Uwaga ogólna	<p>W związku z faktem, że podmioty będące Beneficjentami lub Partnerami projektu w obecnej chwili planują lub realizują projekty o podobnym charakterze, należy zweryfikować konieczność uspołnienienia opisu gromadzonych i przetwarzanych zasobów informacyjnych oraz wzajemnego wykorzystania danych i elementów infrastruktury, w celu zwiększenia ich przydatności i efektywności realizowanych przedsięwzięć.</p>	<p>Proszę o analizę i korektę opisu założeń.</p>	<p>Na etapie przygotowania projektu FBC.AI przeprowadzona została weryfikacja przedsięwzięć planowanych lub realizowanych przez Beneficjenta i Partnerów, obejmująca w szczególności zakres gromadzonych i przetwarzanych zasobów informacyjnych, rozwijane funkcjonalności, stosowane standardy oraz planowane zakupy i elementy infrastruktury. We wszystkich powiązanych przedsięwzięciach Beneficjent i Partnerzy kierują się zasadami komplementarności, interoperacyjności i wzajemnego wykorzystania rezultatów. Oznacza to stosowanie zgodnych standardów danych i metadanych, otwartych interfejsów, uzgodnionych mechanizmów wymiany informacji oraz technologii umożliwiających ponowne wykorzystanie danych i usług. Projekty nie będą powielały tych samych zasobów, funkcjonalności, zakupów ani elementów infrastruktury, jeżeli mogą one zostać wykorzystane wspólnie albo udostępnione pomiędzy przedsięwzięciami. Poszczególne przedsięwzięcia dotyczą odmiennych kolekcji, grup użytkowników, procesów digitalizacyjnych lub funkcji systemowych, natomiast ich rezultaty mogą się wzajemnie uzupełniać. Rozwiązania wypracowane w projekcie FBC.AI będą mogły wzbogacać inne projekty o komponenty związane ze sztuczną inteligencją, automatycznym opracowaniem zasobów, poprawą jakości danych, wyszukiwaniem i ponownym wykorzystaniem treści. Jednocześnie zasoby przygotowywane w innych przedsięwzięciach, w tym dotyczących specjalistycznych kolekcji, np. zasobów teatralnych, będą mogły – z uwzględnieniem warunków prawnych i jakościowych –</p>
----	-------	--------------	---	--	--

					<p>służyć do trenowania, dostrajania, testowania i oceny np. rozwiązań SI.</p> <p>Weryfikacja zakresów działań będzie prowadzona również w trakcie realizacji projektu, tak aby wyeliminować ryzyko dublowania prac i finansowania oraz identyfikować kolejne możliwości synergii pomiędzy przedsięwzięciami.</p>
4.	RA IT	Uwaga ogólna	<p>W dokumencie brakuje szerszych informacji dotyczących planowanego komponentu dot. AI</p> <p>Nie określono, jaki zakres funkcjonalny ma być realizowany z wykorzystaniem AI, jakie modele lub klasy modeli zostaną zastosowane oraz czy celem użycia AI będzie generowanie informacji, wspieranie decyzji, automatyzacja procesów, analiza treści, czy inny sposób przetwarzania danych. Czy będzie stosowany OCR, HTR, NLP czy RAG, jak będzie realizowane wyszukiwanie semantyczne, jakie będą zasady trenowania modeli, jakie dane będą używane do treningu modeli, czy w ogóle zastosowane modele będą trenowane.</p> <p>Brakuje również omówienia kluczowych aspektów związanych z wykorzystaniem AI, w szczególności: compliance, bezpieczeństwo, ochrona danych, jakość wyników oraz odpowiedzialność za rezultaty generowane przez system.</p> <p>Należy uzupełnić informacje dot. AI, m.in. w tabeli pkt. 1.1 oraz w liście systemów w pkt 7.1 w budowanym opisie systemu, który będzie zawierał komponent AI (w przypadku systemów</p>	Proszę o analizę i korektę opisu założeń.	Do uwagi odnosimy się dodatkowym wyjaśnieniem, umieszczonym pod tabelą (dla przejrzystości dokumentu), zgodnie z prośbą uwagodawcy (w OZPI brak miejsce na zaadresowanie uwag). W samym OZPI do Otoczenia prawnego dodajemy Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji – AI Act oraz Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 (RODO).

			modyfikowanych - w opisie zmiany), a także w widoku komponentów w pkt 7.2, jeśli rozwiązanie dot. AI będzie stanowił wyodrębniony komponent systemu. W przypadku braku miejsca w dokumencie dodatkowe pisemne wyjaśnienie.		
5.	RA IT	Uwaga ogólna	Planowane jest wdrożenie AI jako komponentu/modułu, czy rozważa się jego reużycie lub udostępnienie innym podmiotom?	Proszę o analizę i korektę opisu założeń.	Co do zasady planowane do wypracowania rozwiązania AI będą oparte o rozwiązania technologiczne udostępniane na wolnych/otwartych licencjach. Zatem reużycie lub udostępnienie innym podmiotom tych komponentów będzie co do zasady możliwe. Przy czym, naturalnie, szczegółowe warunki wynikać będą z licencji komponentów wybranych do budowy rozwiązań projektowych.
6.	RA IT	1.1. Identyfikacja problemu i potrzeb	Wymieniono system „dLibra Uniwersytetu Jagiellońskiego” – prawdopodobnie chodzi o „dLab UJ” (dLibra UJ to inna nazwa Biblioteki Cyfrowej UJ).	Proszę o analizę i korektę opisu założeń.	Wprowadzono poprawkę.
7.	RA IT	1.1. Identyfikacja problemu i potrzeb 2.4. Produkty końcowe projektu 7.1. Widok kooperacji aplikacji	W projekcie jest mowa o dLab UJ, dLibraUJ, Biblioteka UJ. Należy zweryfikować nazwy systemów i uspoźnić ich stosowanie w całym dokumencie. Nie należy używać synonimów.	Proszę o analizę i korektę opisu założeń	Nazwy uspoźniono.
8.	RA IT	1.1. Identyfikacja problemu i potrzeb	Choć problemy interesariuszy zostały opisane szeroko i wyczerpująco, zwłaszcza w odniesieniu do Beneficjenta i partnerów projektu, wątpliwości budzi stopień, w jakim opisane przedsięwzięcie rzeczywiście odpowiada na ich eliminację. zgodnie z dobrymi praktykami OZPI problemy interesariuszy powinny wynikać z braku lub niewystarczającego rozwoju systemów	Proszę o analizę i korektę opisu założeń.	Dokumentacja została zaktualizowana w taki sposób by problemy interesariuszy zostały bezpośrednio powiązane z elementami wskazanymi w pkt. 2.2-2.4, usuwając równocześnie aspekty zdefiniowane zbyt ogólnie.

			teleinformatycznych, infrastruktury, oprogramowania lub innych produktów projektu wskazanych w pkt 2.2–2.4. W obecnym kształcie część problemów wydaje się mieć charakter ogólny i organizacyjny, przy ograniczonym powiązaniu z konkretnymi produktami projektu oraz zakresem planowanych modyfikacji systemów.		
9.	RA IT	2.1. Cele i korzyści wynikające z projektu KPI	KPI 13-17 – brak wyodrębnienia wartości docelowych. Dwukrotne wskazanie wartości aktualnych.	Proszę o analizę i korektę opisu założeń.	Dokonano odpowiedniej korekty.
10.	RA IT	2.1. Cele i korzyści wynikające z projektu	Wśród podstaw strategicznych projektu dla celu 1 i 2 wskazano projekt dokumentu strategicznego „Sprawne i Nowoczesne Państwo”, który nie stanowi obowiązującego dokumentu polityki państwa, gdyż nie został przyjęty przez RM. Wątpliwości budzi również podana poz. 4 Należy zrewidować niniejsze tj. usunąć lub zastąpić obowiązującymi dokumentami strategicznymi i programowymi, w które wpisuje się cel projektu.	Proszę o analizę i korektę opisu założeń.	Punkt skorygowano zgodnie ze wskazaniem.
11.	RA IT	2.1. Cele i korzyści wynikające z projektu	Część celów projektu została sformułowana wokół technologii, funkcjonalności lub procesów, a nie produktów projektu. Tymczasem zgodnie z dobrymi praktykami OZPI cele projektu powinny odnosić się do docelowych produktów i rezultatów projektu oraz pozostawać spójne z przypisanymi KPI. Proszę o rozważenie przeformułowania celów, np. w kierunku: 1) „Zwiększenie dostępności oraz możliwości inteligentnego wyszukiwania i ponownego wykorzystania zasobów nauki agregowanych w Federacji Bibliotek Cyfrowych poprzez rozwój platformy FBC.AI”; 2) „Przygotowanie interoperacyjnych i przetwarzalnych maszynowo zasobów nauki w postaci AI-ready poprzez digitalizację, poprawę	Proszę o analizę i korektę opisu założeń.	Uwzględniono sugestie poprzez przeformułowanie celów.

			jakości danych oraz opracowanie skalowalnej metodyki przygotowania danych dla instytucji naukowych”		
12.	RA IT	2.1. Cele i korzyści wynikające z projektu	Nie wykazano wskaźników jakościowych pozwalających ocenić rzeczywiste efekty projektu, w szczególności jakość kopii cyfrowych, jakość OCR/HTR, jakość danych AI-ready, skuteczność wyszukiwania semantycznego, wykorzystanie warstwy konwersacyjnej, liczbę użytkowników korzystających z FBC.AI, liczbę zapytań realizowanych przez chatbot/warstwę konwersacyjną oraz stopień wykorzystania metodyki AI-ready. Wskaźniki mają głównie charakter obowiązkowy i ilościowy. Konieczność uzupełnienia.	Proszę o analizę i korektę opisu założeń.	Dodano wskaźnik Odsetek obiektów digitalizowanych w projekcie, których pliki wzorcowe spełniają wymagania poziomu 3-gwiazdkowego FADGI (Federal Agencies Digital Guidelines Initiative) dla materiałów dziedzictwa kulturowego..
13.	RA IT	2.1. Cele i korzyści wynikające z projektu	Dla wszystkich wskaźników przyjęto częstotliwość pomiaru wyłącznie na zakończenie projektu. W przypadku części KPI, w szczególności dotyczących np. digitalizacji zasobów taki sposób monitorowania wydaje się niewystarczający z punktu widzenia zarządzania postępem rzeczowym projektu. Część wskaźników zwłaszcza dotyczących procesów realizowanych etapowo powinna podlegać bieżącemu monitorowaniu.	Proszę o analizę i korektę opisu założeń	Uwagę uwzględniono, zwiększono częstotliwość pomiarów.

14.	RA IT	2.1. Cele i korzyści wynikające z projektu	Dla celu 2 wartości KPI dotyczące liczby zdigitalizowanych i udostępnionych dokumentów nie są w pełni spójne z liczbą zasobów wskazanych w pkt 2.3. OZPI (po zsumowaniu 23 950)	Proszę o analizę i korektę opisu założeń	Wartość poprawiono na 23950
15.	RA IT	2.3 Udostępnione informacje sektora publicznego i zdigitalizowane zasoby	W odniesieniu do czasopism oraz kart katalogowych UJ nie wynika jednoznacznie, czy zasób jest już zdigitalizowany, czy będzie digitalizowany w projekcie, ani na czym konkretnie będzie polegać „podniesienie jakości” zasobów. Proszę o doprecyzowanie, które zasoby będą digitalizowane, które będą jedynie udostępniane, a które będą podlegały poprawie jakości np. poprzez dodanie warstwy OCR, wraz ze wskazaniem zakresu tej poprawy.	Proszę o analizę i korektę opisu założeń	Dodano informację, że te grupy zasobów są już zdigitalizowane. Poprawa jakości ich udostępnienia realizowana będzie w zakresie wytworzenia warstw OCR i HTR, których te zasoby obecnie nie posiadają.
16.	RA IT	2.4 Produkty końcowe projektu	Produkty końcowe wymagają uporządkowania chronologicznego.	Proszę o analizę i korektę opisu założeń	Kolejność skorygowano.
17.	RA IT	3. Kamienie milowe	W harmonogramie kamieni milowych w obszarze podnoszenia jakości i publikacji danych wskazano głównie kamień dotyczący digitalizacji 100% zasobu. Brakuje wyraźnych kamieni milowych dotyczących poprawy jakości danych, przygotowania danych AI-ready, wdrożenia produkcyjnych usług FBC.AI dla użytkowników końcowych oraz zakupu infrastruktury do digitalizacji i opracowania zasobów.	Proszę o analizę i korektę opisu założeń	Wprowadzono dodatkowe kamienie dot. poprawy jakości danych, przygotowania danych AI-ready, oraz zakupu infrastruktury do digitalizacji i opracowania zasobów oraz zmodyfikowano istniejące kamienie, aby wskazywały na wdrożenie produkcyjnych usług FBC.AI oraz zakup infrastruktury.
18.	RA IT	5.1 Ryzyka wpływające na realizację projektu	Brakuje ryzyk związanych z wykorzystaniem modeli AI, w szczególności dotyczących ochrony danych, transparentności modeli, ryzyka halucynacji, jakości generowanych odpowiedzi. Z uwagi na charakter projektu niniejsze jest konieczne do uzupełnienia.	Proszę o analizę i korektę opisu założeń	Wprowadzono dodatkowe ryzyka związane z wykorzystaniem SI.
19.	RA IT	6. Otoczenie prawne	W przedstawionym otoczeniu prawnym projektu nie uwzględniono części kluczowych regulacji związanych z wykorzystaniem technologii AI oraz	Proszę o analizę i korektę opisu założeń	Brakujące regulacje uzupełniono.

			przetwarzaniem danych, w szczególności Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (AI Act) oraz Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 (RODO).		
20.	RA IT	7.1. Widok kooperacji aplikacji	<p>Lista systemów z diagramu kooperacji nie jest odzwierciedlenia w Liście systemów (powinno być 1:1 zgodne)</p> <p>W Liście systemów wykazuje się wszystkie systemy teleinformatyczne z diagramu kooperacji aplikacji odrębnie, z:</p> <ul style="list-style-type: none"> ● uwzględnieniem systemów będących produktami projektu, ● pominięciem grup systemów <p>W Liście systemów w kolumnie:</p> <ul style="list-style-type: none"> ● „Nazwa systemu” należy zaprezentować skrótowiec lub ew. nazwę systemu teleinformatycznego zgodnie z wykazaną na diagramie kooperacji ● „Gestor systemu” należy wprowadzić nawę podmiotu będącego właścicielem systemu ● „Opis systemu” <p>Opisy systemów powinny zawierać:</p> <ul style="list-style-type: none"> ○ Pełną nazwę systemu (na początku pełna nazwa potem sformułowanie „to system wspierający...”) ○ Cel utworzenia systemu ○ Informacja o prowadzonych w systemie rejestrach publicznych, jeśli nie jest to zawarte w celu ○ Opis głównych grup funkcjonalności (może być lista modułów z krótkimi opisami) ○ Informacja o integracjach z innymi systemami krajowymi lub zagranicznymi 	Proszę o analizę i korektę opisu założeń	Uspójniono diagram kooperacji z Listą systemów (są zgodne 1:1). Niespójności wynikały z faktu, że do poprawnych diagramów, wypracowanych podczas spotkań z pracownikami KdSC, błędnie dodano opisy z wcześniejszych etapów prac nad opisem, powodując wystąpienie niespójności.

			<ul style="list-style-type: none"> ○ w przypadku grup systemów wstawiamy na początku "System wielokrotny" ○ Opis: <ul style="list-style-type: none"> ▪ powinien być sformułowany w czasie teraźniejszym, niezależnie od tego czy jest planowany, istniejący czy modyfikowany, ▪ nie powinien zawierać odniesień do jakichkolwiek projektów czy przyszłych lub zrealizowanych działań, ▪ nie powinien zawierać odniesień do wyjątkowości systemu. ● „Status systemu” należy prezentować wartości zgodne z prezentowanymi na diagramie: <ul style="list-style-type: none"> ○ planowane (na diagramie: planowane w projekcie, planowane w innym projekcie) ○ modyfikowane (na diagramie: modyfikowane w projekcie, modyfikowane w innym projekcie) ○ istniejące (na diagramie: istniejące) ● „Krótki opis ewentualnej zmiany” należy: <ul style="list-style-type: none"> ○ wypełnić pole tylko dla systemów „modyfikowanych” w projekcie ○ wskazać nazwy modułów budowanych i włączanych do modyfikowanego systemu budowanego/modyfikowanego modułu, o których mowa w pkt 1.1 i w pkt 2.4, jako uzupełnienie nazwy głównego produktu ○ opisać budowane/modyfikowane w projekcie funkcjonalności/podmoduły tych modułów 		
--	--	--	---	--	--

21.	RA IT	7.1. Widok kooperacji aplikacji	<p>Występują niespójności pomiędzy diagramem kooperacji aplikacji, listą systemów oraz listą przepływów danych.</p> <p>Należy stosować te same nazwy systemów na diagramie i tabeli. Muszą być opisane wszystkie przepływy, które są przedstawione na diagramie. Część systemów wskazanych w tabeli nie została przedstawiona na diagramie (m.in. Węzeł Krajowy), natomiast część systemów obecnych na diagramie nie została jednoznacznie odzwierciedlona w tabelach lub występuje pod inną nazwą, np. na diagramie wskazano GRAPHIA, podczas gdy w tabeli systemów figuruje SSH Knowledge Graph.</p> <p>Wątpliwości budzi również przedstawienie na diagramie Biblioteki Cyfrowej UJ i Biblioteki Cyfrowej UWr jako odrębnych systemów, podczas gdy z opisu tabelarycznego wynika, że są to instancje/komponenty systemu dLibra. Status systemu Kronika jest ponadto niespójny pomiędzy diagramem a tabelą systemów. Na liście systemów brakuje również opisanych systemów udostępniających dane zdigitalizowane oraz Europeany.</p>	Proszę o analizę i korektę opisu założeń	Uspójniono diagram kooperacji z Listą systemów (są zgodne 1:1). Niespójności wynikały z faktu, że do poprawnych diagramów, wypracowanych podczas spotkań z pracownikami KdSC, błędnie dodano opisy z wcześniejszych etapów prac nad opisem, powodując wystąpienie niespójności.
22.	RA IT	7.1. Widok kooperacji aplikacji	W przepływach danych wskazano system FBC.AI, który został częściowo rozbity na komponenty, mimo że na diagramie kooperacji występuje wyłącznie FBC. Zgodnie z informacją z p. 1.1. OZPI „FBC.AI”. FBC.AI nie jest odrębnym systemem, lecz zestawem komponentów wykorzystujących sztuczną inteligencją (SI).	Proszę o analizę i korektę opisu założeń	Dokonano korekt listy przepływów.
23.	RA IT	7.2 Diagram komponentów	Diagram częściowo miesza funkcjonalności biznesowe systemów z warstwą integracyjną, co utrudnia jednoznaczną interpretację architektury rozwiązania. W diagramie komponentów dla	Proszę o analizę i korektę opisu założeń	Dokonano korekty zgodnie ze wskazaniem, poprzez usunięcie komponentów, które odnosiły się do warstwy integracyjnej oraz dodano komponenty już istniejące.

			<p>większości systemów partnerskich wskazano nowe komponenty w postaci „Wymiany danych”.</p> <p>Jednocześnie z opisów systemów zawartych w tabelach 7.1 oraz 4.2 OZPI wynika, że zakres planowanych modyfikacji polega głównie na integracji z komponentami FBC.AI oraz rozbudowie systemów o API udostępniające dane lub treści. Wątpliwości budzi zatem, czy „Wymiana danych” stanowi rzeczywiście nowy komponent biznesowy systemów, czy raczej mechanizm integracyjny i/lub interfejs API służący komunikacji pomiędzy systemami wdrażany w ramach modyfikacji już istniejących komponentów biznesowych wskazanych systemów.</p> <p>Ponadto wątpliwości budzi, czy systemy Omeka, IZZ, dLab oraz biblioteki cyfrowe (które wydają się stanowić komponenty systemu dLibra) rzeczywiście posiadają wyłącznie pojedynczy komponent biznesowy związany z udostępnianiem lub wymianą danych. Czy może systemy te mają więcej niż jeden komponent.</p> <p>Konieczna jest ponowna analiza i korekta diagramu komponentów oraz doprecyzowanie charakteru modyfikacji poszczególnych systemów zgodnie z dobrymi praktykami OZPI.</p>		
24.	RA IT	7.3 Przyjęte założenia technologiczne	<p>Tabela nie zawiera żadnych założeń, co jest niezbędne w przypadku produktów systemowych, w szczególności modyfikowanych.</p> <p>Dotyczy to również standardów wymiany danych, o których m.in mowa w rozporządzeniu KRI i ustawie o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego</p>	Proszę o analizę i korektę opisu założeń	Tabela została uzupełniona wg wskazań.

Odpowiedź na uwagę nr 4:

Elementy SI, w tym przede wszystkim dostęp konwersacyjny (FBC Czatbot), realizowane będą w architekturze RAG opartej na otwartych modelach (open-weight) hostowanych na infrastrukturze wewnętrznej konsorcjum. Dotyczy to zarówno modeli embeddingowych (reprezentacja wektorowa, retrieval), jak i generatywnych modeli językowych (synteza odpowiedzi ugruntowanych w źródle), a także specjalizowanych modeli przetwarzania dokumentów (OCR/HTR, analiza układu (OLR), NER). Przyjęcie modeli otwartych i hostowania lokalnego zapewnia, że zapytania użytkowników oraz przetwarzane treści nie są przekazywane do zewnętrznych dostawców ani poza obszar instytucji, co istotnie ogranicza ryzyka związane z ochroną danych, bezpieczeństwem i transferem danych do jednostek zewnętrznych. Zgodnie z zasadą neutralności technologicznej projekt nie wiąże się z konkretnym dostawcą ani wersją modelu; przewiduje się wykorzystanie wiodących otwartych modeli językowych klasy produkcyjnej (przykładowo z rodzin Qwen, DeepSeek lub europejskich modeli otwartych, w tym Mistral, oraz ich nowszych lub równoważnych wersji), dobieranych do konkretnego zadania, jakości wyników dla języka polskiego i materiałów wielojęzycznych oraz możliwości dostępnej infrastruktury obliczeniowej. Zakres zadań przewidzianych w projekcie nie wymaga wykorzystania największych dostępnych na rynku modeli językowych. Dostęp konwersacyjny realizowany jest w architekturze RAG, w której model generatywny pełni funkcję syntezy odpowiedzi ugruntowanej w wyszukanych fragmentach źródłowych, a nie samodzielnego źródła wiedzy ani narzędzia otwartego, złożonego wnioskowania. Ciężar trafności odpowiedzi spoczywa przede wszystkim na warstwie wyszukiwania oraz jakości przygotowanych danych, a nie na rozmiarze modelu i jego wiedzy parametrycznej. W konsekwencji do realizacji zadań wystarczające są otwarte modele klasy produkcyjnej o umiarkowanej wielkości, a docelowy rozmiar modelu zostanie dobrany po przeprowadzeniu niezbędnych benchmarków. Podejście to jest korzystne również z punktu widzenia kosztów, wymagań infrastruktury obliczeniowej i możliwości uruchomienia modeli lokalnie.

Samo trenowanie modeli nie stanowi celu ani priorytetu projektu FBC.AI. Przyjęta architektura nie wymaga dodatkowego trenowania modeli - zarówno generatywne modele językowe, jak i modele embeddingowe wykorzystywane są w postaci gotowej (pre-trained), a trafność i ugruntowanie odpowiedzi wynikają z wyszukiwania w kontrolowanym korpusie danych przygotowanych w komponentach AI4Data i Data4AI, a nie z trenowania modelu na danych projektu. Takie podejście ogranicza ryzyka związane z danymi treningowymi (w tym prawne i dotyczące ochrony danych) oraz upraszcza utrzymanie i aktualizację komponentów.

Jedynym obszarem, w którym przewiduje się możliwość trenowania lub dostrajania (fine-tuning) modeli, jest rozpoznawanie tekstu (OCR/HTR). Wynika to z heterogeniczności materiału źródłowego. Dla części tych materiałów gotowe modele OCR/HTR mogą okazać się niewystarczające, a uzyskanie użytecznej jakości warstwy tekstowej może wymagać dostrojenia lub wytrenowania wyspecjalizowanych modeli na materiale danej kolekcji. Warto podkreślić, że ewentualne dostrajanie będzie prowadzone wyłącznie na modelach otwartych przeznaczonych do konwersji pisma odręcznego lub drukowanego do tekstu

oraz modeli typu Vision-Language Reasoning Model do prowadzenia tzw. post-OCR/HTR correction. Projekt nie przewiduje tworzenia modeli od podstaw. Dane wykorzystywane do dostrajania będą pochodzić wyłącznie z zasobów objętych projektem (w domenie publicznej) oraz z transkrypcji i anotacji referencyjnych wytwarzanych w projekcie z udziałem ekspertów dziedzinowych. Projekt nie przewiduje wykorzystania do trenowania danych osobowych użytkowników ani zewnętrznych zbiorów danych o nieuregulowanym statusie prawnym.

Wykorzystanie SI w projekcie będzie realizowane zgodnie z obowiązującymi przepisami dotyczącymi systemów AI, ochrony danych osobowych, praw autorskich, cyberbezpieczeństwa i dostępności cyfrowej. Planowane rozwiązania nie będą służyły do profilowania użytkowników ani do podejmowania automatycznych decyzji wywołujących skutki prawne lub istotnie wpływających na osoby fizyczne. Ich funkcją będzie wspieranie wyszukiwania, analizy, przetwarzania, wzbogacania i udostępniania treści cyfrowych. Na etapie projektowania i wdrażania przeprowadzona zostanie ocena zastosowań SI, związanych z nimi ryzyk oraz wymagań prawnych i organizacyjnych. Bezpieczeństwo i ochrona danych zostaną zapewnione przede wszystkim poprzez lokalne hostowanie modeli i przetwarzanie danych w infrastrukturze konsorcjum, kontrolę dostępu, separację środowisk, rejestrowanie zdarzeń, aktualizację komponentów oraz testy bezpieczeństwa. Zapytania użytkowników i dane dotyczące korzystania z systemu nie będą wykorzystywane do trenowania modeli. Zakres ich przetwarzania i okres przechowywania zostaną ograniczone do minimum niezbędnego do zapewnienia działania, bezpieczeństwa i monitorowania jakości usług. Jakość wyników będzie oceniana odrębnie dla poszczególnych zastosowań SI. W przypadku OCR/HTR, analizy układu i NER stosowane będą właściwe dla danego zadania miary jakości oraz zbiory referencyjne. Dla mechanizmów RAG oceniane będą m.in. trafność wyszukanych fragmentów, zgodność odpowiedzi z materiałami źródłowymi, kompletność oraz częstotliwość występowania odpowiedzi nieugruntowanych. Odpowiedzi FBC Czatbota będą, w miarę możliwości, wskazywały wykorzystane źródła, a użytkownik zostanie poinformowany, że treść została wygenerowana automatycznie i może wymagać weryfikacji. Rezultaty generowane przez SI będą miały charakter wspierający i nie będą traktowane jako informacje autorytatywne ani zastępować oceny eksperckiej.

Dodatkowo, w otoczeniu prawnym dodano powiązane regulacji tj. AI Act i dot. RODO.