*Załącznik –* ***Uwagi Ministra Kultury i Dziedzictwa Narodowego (MKiDN) do o****pisu założeń projektu informatycznego (OZPI) pn.* ***„KRONIK@ 2.0”*** *(projekt nr P341)– wnioskodawca: Minister Cyfryzacji*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Nazwa dokumentu:*** *Opis założeń projektu informatycznego pn.* ***„KRONIK@ 2.0”*** *– wnioskodawca: Minister Cyfryzacji, beneficjent: Ministerstwo Cyfryzacji.* | | | | | |
| **Lp.** | **Organ wnoszący uwagi** | **Jednostka redakcyjna, do**  **której wnoszone są uwagi** | **Treść uwagi** | **Propozycja zmian zapisu** | **Odniesienie do uwagi** |
|  | **MKiDN** | * 1. Identyfikacja problemu  i potrzeb   oraz   * 1. Opis stanu obecnego   7.1. Widok kooperacji aplikacji | Niejasna jest liczba 10 instytucji, które zostaną przyłączone poprzez API. Czy będą to podmioty spośród 20 wymienionych, czy nowe? Jak wynika z informacji w OZPI sam proces wyboru tych instytucji jest dopiero planowany na połowę roku 2026, co utrudnia znacznie rzetelne oszacowanie zasobów oraz kosztów.  Ponadto lista przepływów wymienia metodę REST API. Skoro zasilanie, w większości przypadków, odbywa się zdalnie (harvester), w jaki sposób realizowany będzie zatem docelowo standard komunikacji pomiędzy systemami?  W świetle wymienionych w OZPI 20 instytucji pojawia się pytanie o te pozostałe, spośród wymienionych w OZPI liście przepływów 13. | Proszę o analizę i korektę opisu założeń w przede wszystkim w zakresie instytucji  i ich zasobów, przygotowania do nich API, ponowną ocenę kosztów. | Planowane do przyłączenia w ramach projektu będą nowe podmioty/instytucje, które obecnie nie udostępniają swoich danych w portalu. Wybór tych instytucji został zaplanowany jako część działań projektowych i zostanie przeprowadzony w trakcie realizacji przedsięwzięcia.  Docelowym standardem komunikacji pomiędzy systemami będzie REST API, z wykorzystaniem dedykowanych konektorów (adapterów) dostosowanych do różnego typu oprogramowania zarządzającego zasobami w instytucjach. Takie podejście zapewni płynną i automatyczną aktualizację danych w systemie KRONIK@, a jednocześnie pozwoli na integrację także z tymi systemami, które obecnie udostępniają dane w modelu *harvestingu*.  Przyjęcie jednolitego mechanizmu REST API pozwoli na uproszczenie procesu integracji, zapewni kompatybilność oraz umożliwi rzetelne szacowanie zasobów i kosztów. |
|  | **MKiDN** | 1.2. Opis stanu obecnego | Na podstawie ankiety oszacowano liczbę zainteresowanych instytucji oraz liczbę obiektów, które można udostępnić w ramach projektu. | Proszę o przedstawienie do wglądu ankiety. | W dniach 7–8 sierpnia 2024 r. został przeprowadzony mailing z ankietą (pytaniami) skierowany do muzeów znajdujących się w wykazie prowadzonym przez MKiDN. Jego celem było pozyskanie informacji niezbędnych do oszacowania potencjalnych i zainteresowanych instytucji współpracą i udostępnianiem zasobów w portalu KRONIK@.  W wiadomości zawarto trzy pytania:   1. Czy instytucja dysponuje własnym API umożliwiającym integrację z API KRONIK@, co pozwoliłoby na szybkie i bezpośrednie przesyłanie zasobów? 2. Jaka jest przybliżona liczba obiektów cyfrowych znajdujących się w instytucji, które mogłyby zostać udostępnione w portalu KRONIK@? 3. Jaki jest szacowany wolumen danych (w GB) możliwych do zdeponowania w repozytorium?   Odpowiedzi przesłało 45 instytucji na podstawie których została o oszacowana liczba instytucji które można podłączyć oraz ilość zasobów które mogą zostać udostępnione. 20 instytucji odpowiedziało, że jest zainteresowana udostępnianiem zasobów w portalu KRONIK@. Po analizie i możliwościach zespołu projektowego podjęto decyzję o przyłączeniu 10 instytucji w ramach projektu  Ostateczny wybór instytucji został zaplanowany jako część działań projektowych i zostanie przeprowadzony w trakcie realizacji projektu. Dodatkowo proces wyboru 10 instytucji zostanie oparty o obiektywne kryteria (m.in.: gotowość techniczna, liczba i unikalność zasobów, potencjał zwiększenia atrakcyjności portalu). |
|  | **MKiDN** | 1.2. Opis stanu obecnego | W świetle zapisu w OZPI o zastosowaniu zaawansowanych mechanizmów semantyki  i ontologii w wyszukiwarce, umożliwiających wyszukiwanie obiektów z różnych dziedzin budowa klastra wyposażonego w potężny akcelerator GPU na potrzeby zaawansowanego systemu sztucznej inteligencji stosowanego na wielkich zbiorach (LLM), (wskazana  w założeniach technologicznych 7.3.), wydaje się nieuzasadniona i generująca niepotrzebnie koszty. Nasuwa się zatem pytanie, w jakim celu, i w jaki sposób wdrożone będą LLM?  Nigdzie w opisie nie pojawia się LMM np. w rozwoju wyszukiwarki –w kontekście GPU  w Centrum Przetwarzania Danych. | Proszę  o szczegółową analizę oraz wprowadzenie odpowiednich korekt oraz uszczegółowienia zapisów lub usuniecie propozycji. | Wyszukiwarka systemu KRONIK@ w każdym przypadku wymaga wsparcia zaawansowanych mechanizmów deep learningu, aby umożliwić wyszukiwanie semantyczne i wielodziedzinowe – zgodnie z zapisami OZPI.  Możliwe są dwa podejścia technologiczne:   1. Komponenty do ElasticSearch / OpenSearch – rozszerzone o modele semantyczne i embeddingi, które pozwalają na lepsze odwzorowanie znaczenia zapytań i obiektów niż klasyczne indeksowanie pełnotekstowe. 2. Architektura typu RAG (Retrieval-Augmented Generation) – łącząca klasyczne wyszukiwanie dokumentów z wykorzystaniem dużych modeli językowych (LLM), które umożliwiają:    * kontekstualizację wyników,    * ich streszczanie,    * odpowiadanie na złożone pytania w języku naturalnym.   Dodatkowo, integracja architektury RAG z modelami dobrze rozumiejącymi język polski (np. Plum, Bielik) pozwoli na znacznie lepsze rozumienie zapytań sformułowanych w języku naturalnym. Tradycyjne mechanizmy ontologiczne i semantyczne nie radzą sobie w takim stopniu z wieloznacznością i złożonością języka.  Oba podejścia – zarówno rozszerzony ElasticSearch, jak i architektura RAG – wymagają akceleracji GPU, aby możliwe było:   * szybkie trenowanie i uruchamianie modeli embeddingowych, * obsługa LLM na wielkich zbiorach danych, * zapewnienie odpowiedniej wydajności i krótkiego czasu odpowiedzi dla użytkowników końcowych.   Warto podkreślić, że tego typu rozwiązania będą musiały stopniowo pojawiać się również w innych repozytoriach danych, a ich włączenie do systemu KRONIK@ poprzez API stworzy możliwość wsparcia i integracji także z innymi systemami wyszukiwania w obszarze nauki i kultury.B  Budowa klastra z GPU nie jest kosztem nadmiarowym, lecz koniecznym elementem infrastruktury, aby sprostać wymaganiom dotyczącym wyszukiwarki semantycznej i integracji mechanizmów sztucznej inteligencji w projekcie. |
|  | **MKiDN** | 2.1. Cele i korzyści wynikające  z projektu | Dobre praktyki dotyczące OZPI stanowią, że „nazwy wskaźników powinny być tak sformułowane, aby ich wartość aktualna/bazowa wynosiła „0” (np. zamiast „liczba…” wskazane jest użycie sformułowania typu „zwiększenie liczby..”).”” Wskaźniki KPI zawierają sformułowania przeznaczone, jak dla nowobudowanych systemów. Przy kontynuacji wskaźnik powinien w jasny sposób wskazywać na **zwiększenie liczby**. | Proszę o analizę  i korektę opisu założeń | Zostały użyte nazwy wskaźników z programu FERC. Dodatkowo skonsultowane z sekretariatem KRMC w ramach oceny przygotowanego OZPI oraz z IZ FERC. |
|  | **MKiDN** | 3. Kamienie milowe  4.2. Wykaz poszczególnych pozycji kosztowych | Zapisy OZPI przewidują rozbudowanie Portalu KRONIK@ o 3 nowe funkcjonalności. | Proszę  o wymienienie tych funkcjonalności  w OZPI. | W ramach projektu zostanie wdrożonych 6 nowych funkcjonalności, które zostały wymiennie w pozycji 4.2 OZPI (oprogramowanie) oraz powtórzone w pozycji 7.1 lista systemów przy portalu KRONIK@. |
|  | **MKiDN** | 2.1. Cele i korzyści wynikające  z projektu  7.3. Przyjęte założenia technologiczne | Pojemność infrastruktury IT Centrum Przetwarzania Danych zakłada7 PB. Na jakiej podstawie określono jej wysokość? Wielkość ta wydaje się niedoszacowana wobec chociażby zbiorów audiowizualnych. | Proszę  o wyjaśnienie założeń pomiaru. | Wartość docelowa wskaźnika Pojemność infrastruktury IT w PB w wysokości 7 PB została określona na podstawie:  a) analizy obecnych i prognozowanych wolumenów danych deponowanych w portalu KRONIK@ (w tym kopii zapasowych w ramach Recovery Data Center),  b) przewidywanego przyrostu danych w wyniku podłączenia 10 nowych instytucji poprzez API oraz udostępniania zasobów przez instytucje realizujące projekty digitalizacyjne w ramach FERC,  c) uwzględnienia bufora na potrzeby rozwoju portalu i przyszłe zwiększenie liczby instytucji korzystających z usługi hostingu.  Instytucje posiadające obszerne zasoby audiowizualne (FINA, WFDiF, TVP), które mają realizować projekty digitalizacyjne w ramach FERC w swoich OZPI wskazywały portal KRONIK@ jako miejsce udostępniania swoich zasobów w postaci plików prezentacyjnych wraz z metadanymi. W związkiu z tym na obecnym etapie nie było planowane w projekcie przechowywanie kopii zapasowych ich zasobów audiowizualnych w repozytorium portalu KRONIK@. |
|  | **MKiDN** | 2.4. Produkty końcowe projektu  5.1. Ryzyka wpływające na realizację projektu | Wdrożenie zmodyfikowanego Portalu KRONIK@ zostało zaplanowane na listopad 2028 r.  W świetle wcześniejszego zaimplementowania API instytucji wraz z Portalem KRONIK@ (wrzesień 2028 r.) kolejność wdrożenia tych produktów wydaje się być odwrotna. Efekt końcowy – Portal powinien poprzedzać jedną z funkcjonalności – API, aby uniknąć ryzyka związanego z jej nieprawidłowym działaniem i dodatkowych kosztów.  Zresztą problemy z integracją API instytucji kultury z API portalu KRONIK@ obarczone są dużą siłą oddziaływania, zaś zapis o prowadzeniu migracji próbnych na etapie budowy API świadczyłby raczej o ich uruchomieniu już na gotowym Portalu. | Proszę o analizę  i korektę zapisów. | Podłączenie API instytucji z API KRONIK@ jest elementem niezależnym od wdrażania wskazanych w OZPI nowych funkcjonalności (zmodyfikowanego portalu) i nie ma wpływu na bieżące działanie systemu. Są to dwa niezależne działania stąd zostały rozdzielone w kamieniach milowych. API KRONIK@ powstało z ramach poprzedniego projektu (nie będzie zmieniane), a w nowym projekcie planowana jest budowa API integracyjnego (adaptera) na potrzeby integracji API instytucji z API portalu KRONIK@. To API (adapter) będzie pełniło rolę dedykowanego rozwiązania integracyjnego, dostosowanego do specyfiki systemów źródłowych każdej z podłączanych instytucji. Jego wcześniejsze przygotowanie umożliwi:   * prowadzenie testów i migracji próbnych danych, * zapewnienie stabilności i kompatybilności z różnymi systemami instytucji kultury, * ograniczenie ryzyka błędów i nieprzewidzianych kosztów.   Portal KRONIK@ jest warstwą prezentacyjną i użytkową, która korzysta z zasobów dostarczanych przez API. Dlatego jego uruchomienie zostało zaplanowane po przygotowaniu i przetestowaniu API, tak aby zapewnić stabilność, spójność danych i bezpieczeństwo całego rozwiązania. |
|  | **MKiDN** | 5.1. Ryzyka wpływające na realizację projektu  5.2. Ryzyka wpływające na utrzymanie efektów | Brak zainteresowania odbiorców produktami projektu, zważywszy chociażby na przerwy  w funkcjonowaniu obecnego Portalu KRONIK@, są dużym problemem. Sama bieżąca komunikacja za pośrednictwem mediów społecznościowych wydaje się niewystarczająca. Na czym ma polegać dostosowanie narzędzi komunikacyjnych do potrzeb informacyjnych interesariuszy? | Proszę  o wyjaśnienia i ewentualne zmiany zapisów. | Obecny projekt ma rozwiązać problem z przerwami w dostępnie do zasobów jakie wyniknęły z korzystania z usługi hostingowej od dostawców zewnętrznych. Uruchomienie stabilnej usługi hostingowej na serwerach państwowych zagwarantuje jej ciągłość, a także umożliwi rozwój portalu oraz zwiększenie zaufania instytucji udostępniających swoje zasoby w portalu oraz nowych podmiotów które dopiero planujących skorzystanie z portalu KRONIK@. W ramach działań promocyjnych w projekcie oprócz przeprowadzenia kampanii w internecie zaplanowano również video blog, cykliczne artykułu promocyjne (1 na kwartał), dwie konferencje skierowane do potencjalnych podmiotów/instytucji oraz obecnie udostępniających zasoby w portalu. |