

Nazwa dokumentu: WIEDZ@ - Węzły Integracji Edukacji i Danych Zasobów Akademickich

Lp.	Organ wnoszący uwagi	Jednostka redakcyjna, do której wnoszone są uwagi	Treść uwagi	Propozycja zmian zapisu	Odniesienie do uwagi
1	MC	Uwaga ogólna	Projekt przewiduje budowę federacyjnej platformy integrującej zasoby naukowe, archiwalne i dziedzictwa kulturowego, obejmującej m.in. agregację metadanych, wyszukiwanie semantyczne, API oraz udostępnianie zasobów cyfrowych (bez repozytorium zapasowego). Zakres ten może częściowo pokrywać się z funkcjonalnościami rozwijanymi w innych projektach finansowanych lub planowanych do finansowania w ramach FERC, w szczególności projekcie KRONIK@ 2.0.	Prosimy o przedstawienie analizy komplementarności projektu względem innych przedsięwzięć realizowanych lub planowanych do realizacji w ramach FERC, w szczególności projektu KRONIK@ 2.0, wraz ze wskazaniem zakresów funkcjonalnych, które nie będą się dublować, sposobu zapewnienia interoperacyjności oraz mechanizmów eliminujących ryzyko podwójnego finansowania podobnych przedsięwzięć.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.
2	MC	Rozdział 4.2 „Wykaz poszczególnych pozycji kosztowych”	Z przedstawionej struktury budżetu wynika, że dominującą pozycję kosztową stanowi budowa i rozwój oprogramowania (26,39 mln zł z 35,85 mln zł całkowitego budżetu projektu). Jednocześnie projekt jest zgłaszany w ramach działania FERC 02.03 ukierunkowanego na cyfrową dostępność i ponowne wykorzystanie informacji. W obecnym kształcie dokumentacji nie jest możliwa ocena proporcji pomiędzy wydatkami przeznaczonymi na digitalizację zasobów a wydatkami związanymi z budową systemów informatycznych, integracją oraz komponentami AI.	Prosimy o przedstawienie szczegółowego podziału kosztów pomiędzy działania digitalizacyjne, udostępnianie zasobów, integrację źródeł danych, rozwój oprogramowania, komponenty AI oraz pozostałe działania projektowe.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.
3	MC	Rozdział 2.3 „Udostępnione informacje sektora publicznego i zdigitalizowane zasoby” oraz rozdział 2.1 „Cele i KPI”	W dokumentacji jako jednostki pomiaru wykorzystywane są jednocześnie strony dokumentów, klatki mikrofilmowe, rekordy metadanych, publikacje, tłumaczenia, rekordy analityczne oraz pakiety ruchu sieciowego. W związku z tym nie jest możliwa ocena porównywalności danych	Prosimy o przedstawienie szczegółowej metodologii liczenia wskaźników oraz wskazanie zasad agregowania różnych kategorii obiektów w ramach wspólnych wskaźników projektu.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.

			wykorzystywanych do wyliczenia wskaźników projektu.		
4	MC	Rozdział 2.1 – KPI nr 3 „Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne on-line”	Wartość docelowa wskaźnika została określona na poziomie 4 podmiotów, podczas gdy projekt przewiduje integrację danych pochodzących z licznych krajowych i zagranicznych systemów, repozytoriów i platform. Sposób określenia wartości docelowej wskaźnika nie został wyjaśniony.	Prosimy o wskazanie podmiotów uwzględnionych przy kalkulacji wskaźnika oraz przedstawienie metodologii określenia jego wartości docelowej.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.
5	MC	Rozdział 2.2 „Udostępnione e-usługi”	Wszystkie planowane e-usługi zostały zakwalifikowane jako usługi o poziomie dojrzałości „Transakcja”. Z przedstawionych opisów wynika jednak, że usługi obejmują głównie wyszukiwanie, analizę i pobieranie danych oraz metadanych.	Prosimy o przedstawienie uzasadnienia dla zakwalifikowania wszystkich usług do poziomu dojrzałości „Transakcja” oraz wskazanie elementów świadczących o spełnieniu kryteriów tego poziomu dojrzałości.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.
6	MC	Rozdział 2.3 „Udostępnione informacje sektora publicznego i zdigitalizowane zasoby”	Dokumentacja nie pozwala jednoznacznie określić, jaka część wskazanych zasobów będzie faktycznie podlegała digitalizacji, a jaka stanowi zasoby już istniejące w postaci cyfrowej, które będą wyłącznie agregowane, integrowane lub ponownie udostępniane.	Prosimy o przedstawienie zestawienia rozdziałającego zasoby podlegające digitalizacji od zasobów już zdigitalizowanych, które będą wyłącznie integrowane lub udostępniane za pośrednictwem platformy WIEDZ@.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.
7	MC	2.3. Udostępnione informacje sektora publicznego i zdigitalizowane zasoby	Zwracamy uwagę na konieczność stosowania zasad „otwartości w fazie projektowania” oraz „otwartości domyślnej” przy budowaniu i modernizowaniu infrastruktury informatycznej. W związku z planowanym udostępnieniem zasobów danych należy dokonać analizy możliwości ich udostępnienia również w portalu dane.gov.pl.	W przypadku uwzględnienia uwagi prosimy o uzupełnienie dokumentu o informacje dotyczące planowanej publikacji danych w portalu dane.gov.pl.	Odniesienie do uwagi, ze względu na jego zakres, zostało przedstawione poniżej tabeli.

ODPOWIEDŹ OPI PIB 1:

Uwaga nr 1.

Odnosząc się do uwagi dotyczącej możliwego pokrywania się zakresu projektu WIEDZ@ z projektem KRONIK@ 2.0, należy podkreślić, że samo występowanie w obu przedsięwzięciach zagadnień związanych z digitalizacją, opracowaniem, metadanymi, udostępnianiem lub ponownym wykorzystaniem zasobów cyfrowych nie oznacza automatycznie dublowania funkcjonalności ani ryzyka podwójnego finansowania. W obszarze zasobów nauki, archiwaliów i dziedzictwa kulturowego naturalne jest istnienie wielu systemów, repozytoriów, baz danych, archiwów cyfrowych i usług udostępniania prowadzonych przez różne instytucje publiczne, uczelnie, instytuty badawcze, biblioteki, archiwa, muzea oraz inne podmioty posiadające własne zasoby, prawa, kompetencje i obowiązki w zakresie ich opracowania. Istnienie projektu KRONIK@ 2.0 nie wyłącza możliwości realizacji przez inne podmioty, w tym uczelnie publiczne, działań polegających na digitalizacji, opracowaniu, wzbogacaniu metadanych i udostępnianiu własnych lub powierzonych im zasobów, o ile działania te dotyczą innego zakresu rzeczowego i nie są finansowane podwójnie.

KRONIK@ 2.0, czyli Krajowe Repozytorium Obiektów Nauki i Kultury definiuje siebie jako system teleinformatyczny prowadzony przez Ministerstwo Cyfryzacji, który pozwala na zebranie i udostępnianie w jednym miejscu cyfrowych zasobów polskiego dziedzictwa narodowego z zakresu nauki, kultury i administracji. Przyjęty przez twórców projektu KRONIK@ cel koncentruje się na udostępnieniu obiektów dziedzictwa narodowego. Cel ten jest realizowany konsekwentnie, co potwierdza analiza zawartości katalogu portalu KRONIK@.

- Spośród 16 438 471 obiektów wykazywanych 11 czerwca 2026 r. w katalogu portalu KRONIK@ 85% (14 032 180) obiektów miało charakter archiwalny, na co przede wszystkim składa się dokumentacja aktowa.
 - Pozostałe obiekty to głównie wydawnictwa (w tym czasopisma, książki, atlasy, zbiory nut, druki ulotne), oraz obiekty multimedialne (w tym głównie nagrania filmowe i dźwiękowe).
- Do tej funkcji – katalogowania i udostępniania dziedzictwa narodowego – dostosowany są wypracowane przez autorów portalu standardy zarządzania metadanymi.

Cele i narzędzia zaplanowane do realizacji celu Portalu WIEDZ@ są zasadniczo różne od celów i narzędzi portalu KRONIK@.

1. Celem projektu WIEDZ@ jest faktyczne udostępnienie **danych badawczych**, których analiza prowadzić ma do budowania wiedzy, a przez to do przyspieszenia rozwoju społeczno-gospodarczego Polski.
2. Produkty projektu WIEDZ@ - standard metadanych, silnik indeksujący, moduł generujący metadane oraz moduł wspierający budowanie wiedzy – zbudowane zostaną **na potrzeby przetwarzania danych badawczych**, a nie danych aktowych i innego rodzaju archiwaliów. Prawdą jest, że archiwalia i dane aktowe mogą stanowić dane badawcze, ale głównie w dziedzinie nauk humanistycznych. Standard metadanych i katalog Portalu WIEDZ@ obejmować będzie dane badawcze z dziedzin nauki inżynieryjno-technicznych, nauk medycznych i nauki o zdrowiu, nauk rolniczych, nauk społecznych, nauk ścisłych i przyrodniczych.

Materialnym efektem zwiększenia dostępności danych badawczych będą m.in. nowe publikacje w postaci artykułów, rozdziałów bądź monografii, z których wiele prawdopodobnie w przyszłości zostanie skatalogowanych lub udostępnionych w pełnej postaci w portalu KRONIK@. Portal WIEDZ@ pełnić będzie zatem rolę poprzednika portalu KRONIK@ w systemie generowania i udostępniania wiedzy. Specyficznym składnikiem projektu WIEDZ@ jest część związana z cyfryzacją największego znajdującego się poza Polską zbioru dokumentów dot. historii Polski, czyli polskich archiwaliów Instytutu Hoovera. Archiwalia te po zdigitalizowaniu zasila katalog zarówno Platformy WIEDZ@, jak i – dzięki interoperacyjności platformy osiągniętej otwartym API w standardzie OAI-PMH – również inne katalogi, w tym być może katalog platformy KRONIK@. Portal KRONIK@ nie zajmuje się jednak bezpośrednio cyfryzacją archiwaliów, a wyłącznie ich katalogowaniem. Ponownie więc Portal WIEDZ@ w procesie budowania i udostępniania wiedzy stoi wcześniej, niż KRONIK@, która potrzebuje takich projektów jak WIEDZ@, by móc udostępniać publikacje i archiwalia.

Projekt WIEDZ@ nie finansuje tych samych produktów, które są finansowane w projekcie KRONIK@ 2.0. Zakres AGH dotyczy określonych zasobów przewidzianych do opracowania w ramach WIEDZ@ oraz komponentów umożliwiających ich przetworzenie, opisanie, wzbogacenie metadanymi i włączenie do federacyjnej platformy. Jeżeli w ramach projektu powstanie repozytorium, baza lub komponent udostępniania po stronie AGH, będzie on dotyczył konkretnego zasobu i konkretnej funkcji w architekturze WIEDZ@, a nie budowy równoległego krajowego repozytorium obiektów nauki i kultury konkurencyjnego wobec KRONIK@ 2.0.

Z punktu widzenia eliminacji ryzyka podwójnego finansowania istotne jest zatem wykazanie, że:

- zasoby objęte zakresem AGH nie są finansowane w tym samym zakresie w projekcie KRONIK@ 2.0;
- produkty AGH są przypisane do zadań, produktów i kamieni milowych WIEDZ@;
- koszty AGH dotyczą digitalizacji, opracowania, przetwarzania, metadanych, komponentów AI i integracji z WIEDZ@ dla określonych zasobów;
- ewentualne funkcje repozytoryjne lub udostępniające AGH mają charakter zasobowy, dziedzicowy lub komponentowy, a nie zastępują projektu KRONIK@ 2.0;
- interoperacyjność z innymi systemami, w tym z KRONIK@ 2.0, będzie zapewniana przez standardy metadanych, API, identyfikatory, linkowanie i mechanizmy wymiany danych;
- projekt WIEDZ@ będzie korzystał z zasady komplementarności, a nie wyłączności jednego systemu na określony typ zasobów.

W związku z powyższym projekt WIEDZ@ nie powinien być oceniany jako dublujący KRONIK@ 2.0 wyłącznie dlatego, że obejmuje digitalizację, metadane, repozytorium lub udostępnianie wybranych zasobów archiwalnych. Są to typowe funkcje występujące w wielu projektach cyfrowych, a o ewentualnym dublowaniu można mówić dopiero wtedy, gdy finansowany jest ten sam zasób, ten sam produkt, ta sama funkcjonalność i ten sam zakres rzeczowy. W przypadku WIEDZ@ oraz udziału AGH taka sytuacja nie zachodzi.

Jednocześnie projekt WIEDZ@ deklaruje gotowość do zapewnienia interoperacyjności i komplementarności względem KRONIK@ 2.0 oraz innych przedsięwzięć FERC, w szczególności przez uzgodnienie standardów metadanych, wykorzystanie API, niedublowanie tych samych zasobów i prowadzenie ewidencji źródeł danych oraz zakresów integracji. Celem WIEDZ@ nie jest ograniczenie roli innych systemów ani zastępowanie KRONIK@ 2.0, lecz stworzenie federacyjnej platformy umożliwiającej odnajdywanie, analizę i ponowne wykorzystanie rozproszonych zasobów nauki, w tym zasobów opracowanych przez partnerów projektu.

Działanie FERC 2.3 ze swojej istoty dopuszcza realizację projektów obejmujących digitalizację, opracowanie, udostępnianie i ponowne wykorzystanie zasobów cyfrowych przez różne podmioty. Nie oznacza to jednak dublowania finansowania, o ile projekty dotyczą odrębnych zasobów, produktów i zakresów rzeczowych. W przypadku WIEDZ@ nie są udostępniane dane ani kolekcje objęte zakresem KRONIK@ 2.0, lecz inne zasoby, w szczególności zasoby archiwalne opracowywane przez AGH oraz zasoby naukowe integrowane w ramach platformy federacyjnej.

Uwaga nr 2.

W odpowiedzi na uwagę dotyczącą struktury budżetu projektu oraz proporcji pomiędzy wydatkami przeznaczonymi na digitalizację zasobów, udostępnianie, integrację, rozwój oprogramowania i komponenty AI wyjaśniamy, że budżet projektu WIEDZ@ został przygotowany na poziomie szczegółowości właściwym dla Opisu Założeń Projektu Informatycznego oraz zgodnie z Dobrymi praktykami przygotowywania OZPI. Na etapie OZPI budżet służy wykazaniu, że projekt posiada realne i uzasadnione założenia finansowe, powiązane z zakresem rzeczowym, produktami i kamieniami milowymi. OZPI nie zastępuje jednak wniosku o dofinansowanie, studium wykonalności ani szczegółowego kosztorysu projektu. Szczegółowa kalkulacja kosztów, obejmująca przypisanie wartości do zadań, partnerów, kategorii wydatków, produktów i kamieni milowych, zostanie przedstawiona we wniosku o dofinansowanie oraz w studium wykonalności, zgodnie z wymaganiami dokumentacji naboru FERC 2.3. Jednocześnie, dla zapewnienia przejrzystości, poniżej przedstawiamy funkcjonalny podział kosztów projektu według głównych obszarów realizacji:

- ✓ udostępnianie zasobów - 9 258 097,00 zł
- ✓ rozwój oprogramowania - 9 024 820,00 zł
- ✓ działania digitalizacyjne - 7 229 725,00 zł
- ✓ integracja źródeł danych - 4 321 467,00 zł
- ✓ komponenty AI - 3 901 411,00 zł
- ✓ pozostałe działania projektowe - 2 116 480,00 zł

Dominująca pozycja kosztowa dotycząca budowy i rozwoju oprogramowania wynika z charakteru projektu WIEDZ@. Projekt nie polega wyłącznie na wykonaniu digitalizacji pojedynczego zbioru, lecz na stworzeniu platformy umożliwiającej federacyjne wyszukiwanie, integrację, indeksowanie, semantyczne przetwarzanie, udostępnianie i ponowne wykorzystanie zasobów pochodzących z wielu źródeł. Wydatki na oprogramowanie obejmują zatem nie tylko klasyczny development systemu, ale także mechanizmy integracji, API, indeksowania, zarządzania metadanymi, wyszukiwania semantycznego, komponenty AI oraz funkcje konieczne do realizacji celu działania FERC 2.3.

Należy również podkreślić, że działanie FERC 2.3 jest ukierunkowane na cyfrową dostępność i ponowne wykorzystanie informacji. Osiągnięcie tego celu w projekcie WIEDZ@ wymaga nie tylko digitalizacji i opracowania wybranych zasobów, ale również budowy narzędzi pozwalających na ich efektywne wyszukiwanie, łączenie, indeksowanie, analizowanie, udostępnianie przez API i ponowne wykorzystanie przez użytkowników oraz systemy zewnętrzne. Z tego względu znaczący udział kosztów oprogramowania jest uzasadniony zakresem projektu i pozostaje bezpośrednio związany z celami działania.

Projekt WIEDZ@ nie finansuje danych, kolekcji ani obiektów cyfrowych objętych zakresem projektu KRONIK@ 2.0. Wydatki dotyczą odrębnych zasobów, w szczególności zasobów archiwalnych opracowywanych przez AGH oraz zasobów naukowych integrowanych w ramach platformy federacyjnej. Ewentualne podobieństwo ogólnych kategorii funkcjonalnych, takich jak digitalizacja, metadane, API czy udostępnianie, nie oznacza dublowania finansowania, o ile przedmiotem projektu są inne zasoby, inne produkty, inne funkcjonalności szczegółowe i inny zakres rzeczowy.

Szczegółowa ocena kwalifikowalności, racjonalności i efektywności poszczególnych wydatków będzie dokonywana na etapie oceny wniosku o dofinansowanie oraz studium wykonalności przez właściwą instytucję odpowiedzialną za nabór FERC 2.3, w tym przez ekspertów wyłonionych do oceny projektów. Na etapie opiniowania OZPI właściwe jest wykazanie zasadności zakresu rzeczowego, architektury, produktów, interoperacyjności oraz braku dublowania funkcjonalności z innymi projektami. W tym zakresie projekt WIEDZ@ wykazuje komplementarność wobec innych przedsięwzięć FERC, w tym KRONIK@ 2.0, oraz brak finansowania tych samych zasobów i produktów.

Uwaga nr 3.

W odpowiedzi na uwagę dotyczącą stosowania w dokumentacji różnych jednostek pomiaru, tj. stron dokumentów, klatek mikrofilmowych, rekordów metadanych, publikacji, tłumaczeń, rekordów analitycznych oraz pakietów ruchu sieciowego, wyjaśniamy, że jednostki te nie są agregowane automatycznie do jednego wspólnego wskaźnika bez zastosowania metodologii przeliczeniowej lub przypisania do właściwego typu wskaźnika.

W projekcie WIEDZ@ stosowane są dwa poziomy pomiaru:

1. poziom wskaźników projektowych i programowych, zgodnych z SZOP dla działania FERC 2.3 oraz dokumentacją konkursową, wykorzystywany do monitorowania osiągnięcia celów projektu;
2. poziom jednostek operacyjnych i technicznych, wykorzystywany do zarządzania pracami projektowymi, digitalizacją, integracją danych, przetwarzaniem treści, indeksowaniem oraz monitorowaniem działania platformy.

Różne jednostki pomiaru pojawiające się w dokumentacji pełnią różne funkcje i nie zawsze stanowią samodzielny wskaźnik rezultatu lub produktu. Przykładowo strony dokumentów i klatki mikrofilmowe służą do planowania oraz rozliczania prac digitalizacyjnych, rekordy metadanych służą do monitorowania integracji i indeksowania zasobów, publikacje mogą stanowić typ obiektu naukowego, tłumaczenia są wynikiem

przetwarzania treści, rekordy analityczne służą do monitorowania wykorzystania platformy, natomiast pakiety ruchu sieciowego są jednostką techniczną związaną z monitorowaniem przepływu danych i wydajności systemu.

Wskaźniki wskazane w OZPI zostały przygotowane jako wskaźniki zgodne z SZOP dla działania FERC 2.3 oraz dokumentacją konkursową i są spójne z Dobrymi praktykami przygotowywania OZPI. Szczegółowa metodologia ich wyliczenia, wraz z przypisaniem wartości do zadań, partnerów, źródeł danych i produktów, zostanie przedstawiona we wniosku o dofinansowanie oraz w studium wykonalności i będzie podlegała ocenie na etapie oceny projektu.

Dla zapewnienia przejrzystości przyjmuje się następujące zasady metodologiczne:

Kategoria obiektu / jednostki	Funkcja w projekcie	Sposób ujęcia we wskaźnikach
Strony dokumentów	Jednostka służąca do planowania i rozliczania digitalizacji dokumentów papierowych	Mogą być wykorzystywane do wyliczenia wolumenu zdigitalizowanych zasobów, po przeliczeniu na obiekty cyfrowe lub wolumen danych, zgodnie z metodologią przyjętą we WoD
Klatki mikrofilmowe	Jednostka digitalizacji materiałów utrwalonych na mikrofilmach	Nie są sumowane bezpośrednio ze stronami dokumentów; mogą być przeliczane według odrębnej reguły na obiekty cyfrowe, pliki lub wolumen danych
Rekordy metadanych	Jednostka opisująca zasób, obiekt, dokument, publikację, kolekcję lub inny element indeksowany w platformie	Wykorzystywane do monitorowania integracji źródeł danych, indeksowania i udostępniania metadanych
Publikacje	Typ zasobu naukowego objętego integracją lub udostępnieniem	Liczone jako odrębna kategoria zasobów naukowych, o ile odpowiada definicji danego wskaźnika
Tłumaczenia	Wynik przetwarzania treści lub metadanych	Nie są traktowane jako odrębne zasoby, jeżeli dotyczą tego samego obiektu; służą do pomiaru zakresu przetwarzania i wzbogacenia zasobów
Rekordy analityczne	Dane służące do monitorowania wykorzystania platformy, aktywności użytkowników, zapytań, pobrań i użycia API	Wykorzystywane pomocniczo do monitorowania KPI dotyczących wykorzystania usług, a nie jako zasoby udostępniane
Pakiety ruchu sieciowego	Jednostka techniczna związana z monitorowaniem przepływu danych, wydajności i dostępności usług	Nie są agregowane do wskaźników digitalizacji; służą do testów technicznych, wydajnościowych i eksploatacyjnych

Wskaźniki dotyczące digitalizacji i udostępnienia zasobów będą liczone według zasady nierozdzielania jednego zasobu na wiele odrębnych osiągnięć wskaźnikowych. Oznacza to, że ten sam zasób nie będzie liczony wielokrotnie tylko dlatego, że posiada stronę, plik, rekord metadanych, tłumaczenie, wynik OCR/HTR lub opis semantyczny. Poszczególne warstwy opracowania tego samego zasobu będą wykazywane jako element jakościowego wzbogacenia lub technicznego przetworzenia, natomiast nie będą powodować wielokrotnego naliczania tego samego obiektu w ramach wskaźnika, jeżeli definicja wskaźnika tego nie dopuszcza.

W projekcie zostanie zastosowana zasada przypisania każdej jednostki pomiaru do jednej z następujących kategorii:

- jednostka zasobu – np. dokument, publikacja, obiekt cyfrowy, kolekcja, rekord metadanych;
- jednostka digitalizacji – np. strona, klatka mikrofilmowa, plik wynikowy;
- jednostka przetwarzania – np. wynik OCR/HTR, tłumaczenie, ekstrakcja informacji, opis semantyczny;
- jednostka udostępniania – np. rekord udostępniony przez API, zasób dostępny w indeksie, metadane dostępne do ponownego wykorzystania;
- jednostka techniczna – np. pakiet ruchu sieciowego, wywołanie API, log systemowy, zdarzenie analityczne.

Agregacja danych będzie następować wyłącznie w ramach kategorii porównywalnych. Oznacza to, że strony dokumentów nie będą sumowane bezpośrednio z rekordami metadanych, publikacjami, tłumaczeniami lub pakietami ruchu sieciowego. Jeżeli dana jednostka operacyjna będzie wykorzystywana do wyliczenia wskaźnika programowego, zostanie zastosowana reguła przeliczeniowa opisana we wniosku o dofinansowanie i studium wykonalności.

W szczególności:

- dla wskaźników dotyczących liczby zdigitalizowanych lub udostępnionych zasobów liczone będą zasoby spełniające definicję przyjętą dla danego wskaźnika, a jednostki pomocnicze, takie jak strony lub klatki mikrofilmowe, będą wykorzystywane do ustalenia zakresu prac digitalizacyjnych;
- dla wskaźników dotyczących wolumenu danych liczone będą dane w jednostkach pojemności, np. MB/GB/TB, a nie liczba stron, publikacji lub rekordów;
- dla wskaźników dotyczących API liczone będą interfejsy, usługi, wywołania lub udostępnione źródła, zgodnie z definicją wskaźnika;
- dla wskaźników dotyczących użytkowników liczone będą zdarzenia lub użytkownicy według zasad określonych w dokumentacji konkursowej i metodologii monitorowania projektu;
- dla wskaźników dotyczących metadanych liczone będą rekordy metadanych spełniające minimalne wymagania jakościowe i techniczne, w tym przypisanie do źródła, identyfikatora, typu zasobu i statusu udostępnienia.

Dane wykorzystywane do monitorowania wskaźników będą pochodzić z dokumentacji odbiorowej, raportów partnerów, rejestru źródeł danych, systemu zarządzania metadanymi, logów platformy, API, narzędzi analitycznych oraz protokołów potwierdzających wykonanie digitalizacji, integracji i udostępnienia zasobów. Dla każdego wskaźnika zostanie wskazane źródło danych, częstotliwość pomiaru, jednostka miary, odpowiedzialny partner oraz sposób walidacji.

W celu zapewnienia spójności i porównywalności danych projekt będzie stosował następujące mechanizmy kontrolne:

- słownik typów zasobów i jednostek pomiaru;
- standard metadanych WIEDZ@;
- rejestr źródeł danych i integracji;
- identyfikatory źródłowe i informacje o pochodzeniu danych;
- deduplikację rekordów metadanych i zasobów;
- rozdzielenie jednostek digitalizacji, jednostek zasobu, jednostek przetwarzania i jednostek technicznych;
- protokoły odbioru potwierdzające wykonanie digitalizacji, przetworzenia, integracji i udostępnienia;
- raporty okresowe z realizacji wskaźników.

Tym samym dokumentacja projektu nie zakłada niekontrolowanego sumowania różnych typów jednostek. Różnorodność jednostek wynika z wielowarstwowego charakteru projektu, który obejmuje zarówno digitalizację i opracowanie zasobów, jak i ich przetwarzanie, wzbogacenie metadanymi, indeksowanie, udostępnianie przez API oraz monitorowanie wykorzystania platformy. Szczegółowe wyliczenie wartości wskaźników oraz ich uzasadnienie zostanie przedstawione we wniosku o dofinansowanie i studium wykonalności, natomiast na etapie OZPI wskazano wskaźniki zgodne z SZOP dla działania FERC 2.3, dokumentacją konkursową oraz Dobrymi praktykami przygotowywania OZPI.

Uwaga nr 4.

W odpowiedzi na uwagę dotyczącą KPI nr 3 „Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne on-line” wyjaśniamy, że wartość docelowa wskaźnika została określona zgodnie z definicją wskaźnika przewidzianą dla działania FERC 2.3 oraz dokumentacją konkursową.

Wskaźnik ten nie mierzy liczby systemów, repozytoriów, baz danych, kolekcji, platform, API ani źródeł danych integrowanych w projekcie, lecz liczbę podmiotów, które w wyniku realizacji projektu udostępnią informacje sektora publicznego lub dane on-line. Oznacza to, że jeden podmiot może udostępniać wiele zbiorów danych, wiele baz, wiele kolekcji lub wiele interfejsów API, natomiast na potrzeby tego wskaźnika liczony jest jako jeden podmiot.

Wartość docelowa wskaźnika na poziomie 4 została określona na podstawie liczby podmiotów, które w ramach projektu będą bezpośrednio odpowiedzialne za udostępnienie danych, metadanych lub zasobów on-line w sposób możliwy do wykazania i rozliczenia w projekcie. Do kalkulacji wskaźnika nie wliczano automatycznie wszystkich krajowych i zagranicznych systemów, repozytoriów, platform lub baz danych, które mogą być integrowane, indeksowane, linkowane lub wykorzystywane referencyjnie przez Platformę WIEDZ@.

Przyjęta metodologia zakłada, że do wskaźnika wliczany jest podmiot, który spełnia łącznie następujące warunki:

- jest uczestnikiem projektu lub podmiotem objętym zakresem integracji w sposób pozwalający na jednoznaczne przypisanie efektu udostępnienia do projektu;
- udostępnia informacje sektora publicznego, dane, metadane lub zasoby on-line w wyniku realizacji projektu;
- udostępnienie ma charakter mierzalny i możliwy do potwierdzenia dokumentacją projektową, np. protokołem odbioru, raportem z integracji, rejestrem źródeł danych, dokumentacją API lub raportem z udostępnienia zasobów;
- udostępnienie nie polega wyłącznie na technicznym wykorzystaniu istniejącego zewnętrznego źródła danych, lecz stanowi efekt rzeczowy projektu po stronie danego podmiotu.

W ramach kalkulacji wskaźnika uwzględniono następujące podmioty:

Lp.	Podmiot	Zakres uwzględnienia we wskaźniku
1	OPI PIB	Udostępnienie metadanych, danych i usług w ramach Platformy WIEDZ@, w tym mechanizmów API, indeksowania, wyszukiwania i ponownego wykorzystania danych
2	AGH	Udostępnienie zasobów archiwalnych objętych zakresem AGH, metadanych oraz wyników przetwarzania treści, w tym efektów digitalizacji, OCR/HTR, wzbogacania metadanych i analizy semantycznej

3	PCSS	Udostępnienie danych, metadanych lub usług integracyjnych związanych z zakresem realizowanym przez PCSS w projekcie, w szczególności w obszarze infrastruktury, integracji lub federacji zasobów
4	PP	Udostępnienie danych, metadanych lub zasobów on-line w zakresie przypisanym do danego partnera/podmiotu w projekcie

Jednocześnie należy podkreślić, że projekt może integrować większą liczbę systemów, repozytoriów, baz danych i platform niż liczba podmiotów wykazana w KPI nr 3. Nie oznacza to niespójności wskaźnika, ponieważ liczba źródeł danych nie jest tożsama z liczbą podmiotów udostępniających dane. Przykładowo jeden podmiot może udostępniać kilka baz danych lub kilka API, a z kolei jedno zewnętrzne repozytorium może być wykorzystywane przez Platformę WIEDZ@ jedynie jako źródło referencyjne, bez wliczania jego właściciela do wskaźnika.

Do wskaźnika nie będą wliczane podmioty, których zasoby są jedynie indeksowane, linkowane, wykorzystywane pomocniczo lub dostępne w ramach istniejących publicznych usług, jeżeli projekt WIEDZ@ nie powoduje nowego, mierzalnego udostępnienia danych po stronie tego podmiotu. Takie źródła będą ujmowane w rejestrze źródeł danych i integracji, natomiast nie będą automatycznie zwiększać wartości KPI nr 3.

Metodologia określenia wartości docelowej opiera się zatem na ostrożnym i weryfikowalnym podejściu: do wskaźnika zaliczono wyłącznie te podmioty, dla których projekt będzie w stanie wykazać bezpośredni efekt w postaci udostępnienia informacji sektora publicznego, danych, metadanych lub zasobów on-line. Szczegółowe przypisanie podmiotów, zakresów danych, produktów i źródeł weryfikacji zostanie przedstawione we wniosku o dofinansowanie oraz studium wykonalności.

W konsekwencji wartość docelowa wskaźnika na poziomie 4 nie jest sprzeczna z planowaną integracją licznych systemów, repozytoriów i platform. Integracja wielu źródeł danych będzie monitorowana odrębnie, m.in. poprzez rejestr źródeł danych, liczbę integrowanych baz, API, rekordów metadanych i zasobów objętych indeksowaniem, natomiast KPI nr 3 pozostaje wskaźnikiem odnoszącym się do liczby podmiotów udostępniających informacje lub dane on-line.

Uwaga nr 5.

Poziom planowanych e-usług został określony zgodnie z definicją wskazaną w Dobrych praktykach dotyczących przygotowywania OZPI. Kwalifikacja usług do poziomu dojrzałości „Transakcja” wynika z tego, że usługi nie polegają wyłącznie na biernym udostępnieniu informacji, lecz obejmują pełną elektroniczną realizację procesu: zainicjowanie żądania przez użytkownika lub system zewnętrzny, automatyczne przetworzenie żądania przez Platformę WIEDZ@ oraz zwrócenie wyniku w postaci danych, metadanych, rekomendacji, zestawienia, pliku, linku referencyjnego lub odpowiedzi API.

W przypadku federacyjnego wyszukiwania użytkownik formułuje zapytanie, wybiera parametry lub filtry, a system przetwarza żądanie, przeszukuje indeks metadanych i zwraca wynik dopasowany do zapytania. W przypadku asystenta wiedzy użytkownik inicjuje interakcję, a system dokonuje analizy semantycznej i generuje rekomendacje lub odpowiedzi. W przypadku API do ponownego wykorzystania danych użytkownik lub system zewnętrzny przesyła żądanie, a system zwraca dane lub metadane w ustrukturyzowanej postaci możliwej do dalszego wykorzystania.

Usługi te są zatem realizowane w pełni online i nie wymagają przejścia na kanał papierowy ani kontaktu poza systemem. Jednocześnie nie jest konieczne, aby usługa transakcyjna miała charakter płatności lub klasycznego postępowania administracyjnego. Wystarczające jest, że użytkownik wykonuje określoną czynność elektroniczną, system przetwarza ją automatycznie i zwraca rezultat możliwy do wykorzystania cyfrowo.

Dla uniknięcia wątpliwości opisy e-usług zostaną doprecyzowane poprzez wskazanie elementów świadczących o poziomie „Transakcja”, tj. sposobu inicjowania usługi, mechanizmu przetwarzania żądania, rodzaju zwracanego wyniku oraz możliwości jego ponownego wykorzystania.

Uwaga nr 6.

W odpowiedzi na uwagę dotyczącą rozdzielenia zasobów podlegających digitalizacji od zasobów już istniejących w postaci cyfrowej wyjaśniamy, że projekt WIEDZ@ obejmuje dwa odrębne typy działań w zakresie zasobów:

1. **digitalizację i opracowanie wybranych zasobów analogowych lub częściowo zdigitalizowanych**, które wymagają wykonania prac digitalizacyjnych, OCR/HTR, opracowania metadanych, przetworzenia treści i przygotowania do udostępnienia;
2. **integrację, agregację, indeksowanie i ponowne udostępnianie zasobów już istniejących w postaci cyfrowej**, które nie wymagają ponownej digitalizacji, lecz wymagają ujednolicenia metadanych, integracji z Platformą WIEDZ@, indeksowania, przygotowania API lub zapewnienia możliwości ponownego wykorzystania.

Na etapie OZPI wskazano zakres zasobów i kierunki integracji w stopniu właściwym dla opisu założeń projektu informatycznego. Szczegółowa inwentaryzacja zasobów, obejmująca przypisanie konkretnych kolekcji, zbiorów, źródeł danych i wolumenów do kategorii „digitalizacja” albo „integracja/ponowne udostępnienie”, zostanie przedstawiona w studium wykonalności oraz we wniosku o dofinansowanie.

Dla zapewnienia przejrzystości przyjmuje się następujące rozróżnienie:

Kategoria zasobu	Charakter prac w projekcie WIEDZ@	Sposób ujęcia w projekcie
Zasoby analogowe	Wymagają wykonania digitalizacji, opracowania technicznego, OCR/HTR, przygotowania plików wynikowych oraz metadanych	Będą wykazywane jako zasoby podlegające digitalizacji
Zasoby częściowo zdigitalizowane	Wymagają uzupełnienia digitalizacji, poprawy jakości, przetworzenia OCR/HTR, wzbogacenia metadanych lub przygotowania do udostępnienia	Będą wykazywane jako zasoby podlegające digitalizacji lub ponownemu opracowaniu, w zależności od faktycznego zakresu prac
Zasoby już zdigitalizowane, ale nieudostępnione lub trudno dostępne	Nie wymagają wykonania digitalizacji od podstaw, ale wymagają opracowania, ujednolicenia metadanych, indeksowania, integracji lub publikacji przez API	Będą wykazywane jako zasoby przeznaczone do integracji i ponownego udostępnienia
Zasoby cyfrowe dostępne w systemach źródłowych partnerów	Wymagają integracji z Platformą WIEDZ@, mapowania metadanych, indeksowania, deduplikacji i udostępnienia w warstwie federacyjnej	Będą wykazywane jako zasoby integrowane, a nie digitalizowane
Zasoby cyfrowe dostępne w zewnętrznych repozytoriach lub systemach referencyjnych	Mogą być indeksowane, linkowane lub integrowane poprzez API, metadane albo identyfikatory źródłowe	Będą ujmowane w rejestrze źródeł danych i integracji, bez zaliczania do zasobów digitalizowanych w projekcie

W szczególności zasoby archiwalne objęte zakresem AGH będą stanowiły kategorię zasobów wymagających digitalizacji, opracowania, OCR/HTR, wzbogacenia metadanymi i przygotowania do udostępnienia lub ponownego wykorzystania w Platformie WIEDZ@. Z kolei zasoby już istniejące w postaci cyfrowej, pochodzące z krajowych lub zagranicznych systemów, repozytoriów, baz danych i platform, będą co do zasady przedmiotem integracji, generowania i agregacji metadanych, indeksowania, linkowania, udostępniania przez API lub ponownego wykorzystania, bez wykazywania ich jako zasobów nowo digitalizowanych. Projekt będzie stosował zasadę rozdzielania wskaźnikowego i ewidencyjnego obu kategorii zasobów. Oznacza to, że zasoby faktycznie digitalizowane w projekcie będą wykazywane odrębnie od zasobów już cyfrowych, które zostaną jedynie zintegrowane, zaindeksowane lub ponownie udostępnione. Ten sam zasób nie będzie liczony jednocześnie jako zasób zdigitalizowany i jako zasób wyłącznie integrowany, chyba że dokumentacja szczegółowa wykaże odrębne, mierzalne zakresy prac, np. digitalizację brakującej części kolekcji oraz integrację wcześniej istniejących metadanych.

W celu zapewnienia prawidłowego przypisania zasobów do właściwych kategorii w projekcie zostanie przygotowany rejestr zasobów i źródeł danych, obejmujący co najmniej:

1. nazwę zbioru, kolekcji, bazy danych lub systemu źródłowego;
2. podmiot odpowiedzialny za zasób;
3. status zasobu: analogowy, częściowo zdigitalizowany, zdigitalizowany, dostępny online, dostępny przez API, wymagający integracji;
4. zakres prac przewidzianych w projekcie: digitalizacja, OCR/HTR, opracowanie metadanych, integracja, indeksowanie, linkowanie, API, ponowne udostępnienie;
5. planowany sposób udostępnienia lub integracji z Platformą WIEDZ@;
6. jednostkę pomiaru właściwą dla danego zasobu;
7. przypisanie do produktu, kamienia milowego i partnera odpowiedzialnego za realizację danego zakresu.

Takie podejście pozwoli na jednoznaczne rozróżnienie zasobów, które będą faktycznie digitalizowane w projekcie, od zasobów już istniejących w postaci cyfrowej, ale dla rzeczywistego udostępnienia wymagających zintegrowania źródeł, wzbogacenia o metadane, zagregowania metadanych do standardu WIEDZ@, zaindeksowania i udostępniania za pośrednictwem Platformy WIEDZ@.

Pełne zestawienie zasobów, wraz z ich wolumenem, statusem cyfrowym, zakresem prac i przypisaniem do wskaźników, zostanie przedstawione w studium wykonalności oraz we wniosku o dofinansowanie. Na etapie OZPI przedstawiono natomiast podział kategorii zasobów oraz metodologię ich kwalifikowania, adekwatnie do celu tego dokumentu.

Uwaga nr 7.

W odpowiedzi na uwagę dotyczącą konieczności stosowania zasad „otwartości w fazie projektowania” oraz „otwartości domyślnej” wyjaśniamy, że uwaga zostanie uwzględniona w dalszych pracach nad projektem. Projekt WIEDZ@ zakłada udostępnianie zasobów, metadanych i usług w sposób zgodny z zasadami otwartości, interoperacyjności i ponownego wykorzystywania informacji. Oznacza to, że już na etapie projektowania architektury systemu zostaną uwzględnione mechanizmy umożliwiające publikację i wymianę danych w otwartych, maszynowo odczytywalnych formatach, w szczególności przez API, standard metadanych, trwałe identyfikatory, informacje o pochodzeniu danych, licencje oraz mechanizmy eksportu metadanych.

Jednocześnie należy wskazać, że nie wszystkie zasoby objęte projektem muszą być publikowane bezpośrednio w portalu dane.gov.pl. Część zasobów może być udostępniana w systemach źródłowych partnerów, część za pośrednictwem Platformy WIEDZ@, a część może być udostępniana w portalu dane.gov.pl w postaci zbiorów danych, metadanych, linków referencyjnych, opisów kolekcji lub informacji o interfejsach API.

Ostateczny sposób publikacji będzie zależał od charakteru zasobu, statusu prawnego, praw do danych, ograniczeń licencyjnych, jakości metadanych, formatu danych oraz możliwości technicznej integracji.

W projekcie zostanie przyjęta zasada, że dla każdego zasobu lub zbioru danych zostanie przeprowadzona kwalifikacja pod kątem możliwości otwartego udostępnienia, obejmująca co najmniej:

1. ustalenie dysponenta lub właściciela zasobu;
2. określenie statusu prawnego danych i metadanych;

- ocenę możliwości publikacji w otwartych formatach;
- wskazanie warunków licencyjnych i zasad ponownego wykorzystywania;
- ocenę, czy dane mogą być publikowane bez ograniczeń, czy wyłącznie w zakresie metadanych lub informacji referencyjnych;
- wskazanie właściwego kanału udostępnienia: Platforma WIEDZ@, system źródłowy partnera, API, portal dane.gov.pl lub kombinacja tych kanałów.

W przypadku zasobów spełniających wymagania prawne, techniczne i jakościowe dla publikacji w portalu dane.gov.pl przewiduje się możliwość udostępnienia ich w tym portalu bezpośrednio albo przez publikację metadanych, opisów zbiorów, linków do zasobów, informacji o API lub innych mechanizmów umożliwiających ponowne wykorzystanie danych. Dotyczyć to będzie w szczególności tych danych i metadanych, które będą mogły zostać udostępnione jako informacje sektora publicznego w sposób otwarty, maszynowo odczytywalny i możliwy do ponownego wykorzystania.

WIEDZ@ będzie projektowana z uwzględnieniem zasad „otwartości w fazie projektowania” oraz „otwartości domyślnej”. Na etapie realizacji projektu możliwość publikacji danych lub metadanych w portalu dane.gov.pl zostanie przeanalizowana dla poszczególnych kategorii zasobów

KORESPONDECJA OPI PIB – MC za pośrednictwem KdSC nr 2

Uwagi MC nr 2 z 19.06.2026

- Odpowiedź beneficjenta nie rozwiewa wątpliwości wskazanych w uwadze MC. Beneficjent skoncentrował się przede wszystkim na wykazaniu, że projekt WIEDZ@ będzie obejmował inne zasoby niż projekt KRONIK@ 2.0, w szczególności inne kolekcje podlegające digitalizacji lub opracowaniu. Nie to było jednak zasadniczym przedmiotem uwagi Ministerstwa Cyfryzacji.

MC nie kwestionuje możliwości prowadzenia działań digitalizacyjnych przez różne podmioty ani nie wskazuje, że każdy projekt obejmujący digitalizację, metadane, repozytorium lub udostępnianie zasobów cyfrowych automatycznie dubluje KRONIK@ 2.0. Wątpliwość dotyczy przede wszystkim zasadności finansowania ze środków publicznych budowy kolejnego portalu/platformy o zbliżonych funkcjonalnościach do rozwiązań rozwijanych lub planowanych w ramach innych projektów FERC, w szczególności KRONIK@ 2.0.

Z dokumentacji projektu WIEDZ@ wynika, że w ramach przedsięwzięcia planowana jest budowa federacyjnej platformy obejmującej m.in. agregację metadanych, indeksowanie, wyszukiwanie semantyczne, API, mechanizmy integracji źródeł danych, udostępnianie zasobów cyfrowych oraz komponenty wspierające ponowne wykorzystanie danych. Są to funkcjonalności, które mogą pozostawać zbliżone do funkcji rozwijanych w innych systemach finansowanych ze środków FERC. Kluczowe znaczenie ma zatem nie tylko odrębność zbiorów, ale również odrębność finansowanych funkcjonalności, architektury usług, mechanizmów udostępniania, katalogowania, wyszukiwania, integracji i ponownego wykorzystania zasobów.

W tym kontekście odpowiedź beneficjenta pozostaje niewystarczająca. Nie przedstawiono szczegółowej analizy porównawczej funkcjonalności WIEDZ@ i KRONIK@ 2.0, która pozwalałaby jednoznacznie ocenić, które elementy oprogramowania, usług i mechanizmów technicznych są komplementarne, a które mogą się pokrywać. Brakuje w szczególności zestawienia pokazującego różnice między planowanymi funkcjami portalu WIEDZ@ a funkcjami KRONIK@ 2.0 w obszarach takich jak: agregacja metadanych, wyszukiwanie, API, udostępnianie zasobów, repozytorium/warstwa przechowywania, indeksowanie, integracje z systemami zewnętrznymi oraz mechanizmy ponownego wykorzystania danych.

Dodatkowo argumentacja beneficjenta opiera się na zbyt daleko idącym rozdzieleniu zakresu danych badawczych i zasobów udostępnianych w portalu KRONIK@. Należy zauważyć, że w portalu KRONIK@ znajdują się już dane badawcze przekazane m.in. przez Przemysłowy Instytut Automatyki i Pomiarów PIAP, który wchodzi w skład Sieci Badawczej Łukasiewicz. Ponadto standard metadanych stosowany w portalu KRONIK@ został w toku jego przygotowania pozytywnie zaopiniowany przez Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, a więc przez podmiot będący beneficjentem projektu WIEDZ@. Powyższe okoliczności wskazują, że rozróżnienie między KRONIK@ jako systemem właściwym wyłącznie dla dziedzictwa narodowego a WIEDZ@ jako systemem właściwym dla danych badawczych nie jest tak jednoznaczne, jak wynikałoby z odpowiedzi beneficjenta.

Należy również zauważyć, że w projekcie WIEDZ@ dominującą pozycję budżetową stanowi budowa i rozwój oprogramowania, której wartość wynosi ponad 26 mln zł. Wątpliwość MC dotyczy więc przede wszystkim racjonalności finansowania budowy nowego rozwiązania informatycznego o szerokim zakresie funkcjonalnym, a nie samego faktu digitalizacji określonych zasobów. Odpowiedź beneficjenta nie wykazuje, dlaczego dla realizacji zakładanych celów konieczne jest tworzenie odrębnego portalu/platformy, zamiast wykorzystania, rozbudowy lub integracji z rozwiązaniami już istniejącymi albo finansowanymi lub planowanymi do finansowania w ramach FERC.

W szczególności prosimy o wyjaśnienie czy część zadań planowanych do realizacji w ramach projektu WIEDZ@ — zwłaszcza w zakresie agregacji metadanych, udostępniania zasobów cyfrowych, wyszukiwania, API, integracji źródeł danych oraz ponownego wykorzystania informacji — nie mogłaby zostać zrealizowana przez istniejące systemy lub przez ich rozbudowę, zamiast przez budowę nowej platformy. Prosimy także o wskazanie, które funkcjonalności rzeczywiście wymagają stworzenia nowego rozwiązania, a które mogłyby zostać zapewnione przez integrację z systemami już działającymi w administracji publicznej lub rozwijanymi ze środków FERC.

W związku z powyższym prosimy o uzupełnienie wyjaśnień o szczegółową analizę komplementarności funkcjonalnej WIEDZ@ względem KRONIK@ 2.0 oraz innych projektów FERC. Analiza powinna obejmować nie tylko zakres zasobów, lecz także zakres finansowanego oprogramowania, funkcjonalności portalu, mechanizmy wyszukiwania i agregacji metadanych, API, udostępnianie zasobów cyfrowych, integracje zewnętrzne oraz mechanizmy eliminujące ryzyko finansowania podobnych komponentów informatycznych ze środków publicznych.

2. Odpowiedź beneficjenta nie usuwa wątpliwości wskazanych w uwadze MC. Przedstawiony podział kosztów ma nadal charakter mocno ogólny i nie pozwala ocenić rzeczywistych proporcji między działaniami digitalizacyjnymi, budową oprogramowania, integracją źródeł danych, komponentami AI, udostępnianiem zasobów oraz pozostałymi działaniami projektowymi. Beneficjent wskazuje, że rozwój oprogramowania obejmuje kwotę 26 389 000 zł, integracja źródeł danych 3 154 100 zł, udostępnianie zasobów 634 500 zł, a pozostałe działania projektowe 5 674 400 zł. Jednocześnie największa pozycja kosztowa, tj. rozwój oprogramowania, nie została rozbita na poszczególne komponenty, produkty, partnerów, funkcjonalności ani kategorie wydatków. Nie wiadomo zatem, jaka część tej kwoty dotyczy budowy Platformy WIEDZ@, AGH Digital AI System, modyfikacji NAVOICA, API, silnika indeksującego, komponentów AI, licencji, testów, dokumentacji, integracji, mechanizmów udostępniania zasobów czy innych elementów technicznych.

Powyższe budzi dodatkową wątpliwość co do realności założeń digitalizacyjnych projektu. Z jednej strony projekt zakłada bardzo wysokie wartości wskaźników dotyczących digitalizacji i udostępnienia zasobów, w tym 5 000 000 zdigitalizowanych dokumentów lub obiektów źródłowych, 5 000 000 dokumentów lub obiektów udostępnionych on-line oraz 50 TB zdigitalizowanych informacji. Z drugiej strony z przedstawionego podziału kosztów wynika, że zdecydowana większość budżetu przeznaczona jest na budowę i rozwój oprogramowania, natomiast nie wskazano odrębnej, czytelnej pozycji kosztowej odpowiadającej za właściwe działania digitalizacyjne.

W konsekwencji powstaje wątpliwość, czy założony wolumen digitalizacji jest możliwy do osiągnięcia przy tak ograniczonej i niejednoznacznie wyodrębnionej części budżetu przeznaczonej na digitalizację. Jeżeli działania digitalizacyjne mają być finansowane w ramach innych pozycji, prosimy o jednoznaczne wskazanie, w których pozycjach budżetu zostały ujęte, jaka jest ich wartość oraz jakie prace obejmują, w szczególności: przygotowanie materiałów, skanowanie lub digitalizację mikrofilmów, OCR/HTR, kontrolę jakości, opracowanie techniczne plików, przygotowanie metadanych, przechowywanie plików wynikowych oraz udostępnienie zasobów.

W obecnym kształcie odpowiedź beneficjenta nie pozwala ocenić, czy projekt jest projektem przede wszystkim digitalizacyjnym i udostępniającym zasoby, czy w praktyce projektem budowy nowej platformy informatycznej, w którym digitalizacja stanowi jedynie uzasadnienie dla wytworzenia kosztownego oprogramowania. Ma to szczególne znaczenie w kontekście działania FERC 02.03, którego celem jest cyfrowa dostępność i ponowne wykorzystanie informacji, a nie samo finansowanie budowy nowych systemów informatycznych.

Prosimy o uzupełnienie dokumentacji o bardziej przejrzyste zestawienie najważniejszych grup kosztów związanych z digitalizacją i udostępnieniem zasobów. W szczególności prosimy o wskazanie, jaka część budżetu została przeznaczona na właściwe prace digitalizacyjne, opracowanie i przygotowanie metadanych, przetwarzanie OCR/HTR lub inne przetwarzanie treści, kontrolę jakości, przechowywanie danych oraz techniczne przygotowanie zasobów do udostępnienia. Takie zestawienie jest niezbędne do oceny, czy deklarowane wskaźniki digitalizacji i udostępnienia zasobów mają realne pokrycie w strukturze budżetu projektu.

3. Odpowiedź beneficjenta częściowo porządkuje sposób rozumienia różnych jednostek pomiaru stosowanych w projekcie, jednak nie usuwa w pełni wątpliwości wskazanych w uwadze MC. Beneficjent wyjaśnia, że strony dokumentów, klatki mikrofilmowe, rekordy metadanych, publikacje, tłumaczenia, rekordy analityczne oraz pakiety ruchu sieciowego nie będą automatycznie agregowane do jednego wspólnego wskaźnika. Wskazuje również na rozróżnienie między poziomem wskaźników projektowych i programowych a poziomem jednostek operacyjnych i technicznych.

Wyjaśnienie to ma jednak w dużej mierze charakter deklaracyjny. W odpowiedzi nie przedstawiono konkretnych reguł przeliczeniowych ani jednoznacznego przypisania poszczególnych jednostek do konkretnych KPI wskazanych w OZPI. Przykładowo beneficjent wskazuje, że strony dokumentów mogą być wykorzystywane do wyliczenia wolumenu zdigitalizowanych zasobów po przeliczeniu na obiekty cyfrowe, a klatki mikrofilmowe mogą być przeliczane według odrębnej reguły na obiekty cyfrowe lub pliki. Nie wskazano jednak, jakie reguły przeliczeniowe zostaną zastosowane, jakie będą jednostki bazowe dla poszczególnych wskaźników oraz w jaki sposób zapewniona zostanie porównywalność danych między różnymi kategoriami zasobów.

Wątpliwość pozostaje szczególnie istotna w odniesieniu do KPI dotyczących digitalizacji i udostępniania zasobów. Projekt zakłada osiągnięcie wartości 5 000 000 zdigitalizowanych dokumentów lub obiektów źródłowych oraz 5 000 000 dokumentów lub obiektów udostępnionych on-line. Bez jednoznacznego wskazania, co w praktyce będzie liczone jako „dokument”, „obiekt źródłowy”, „obiekt cyfrowy”, „plik wynikowy”, „strona” lub „klatka mikrofilmowa”, nie jest możliwa ocena realności i porównywalności tych wskaźników. Przykładowo, z obecnej odpowiedzi nie wynika, czy w przypadku udostępnienia w portalu jednej książki będzie ona liczona na potrzeby KPI jako jeden dokument/obiekt, czy też do wskaźnika zostanie zaliczona każda jej strona jako odrębny obiekt.

Nie jest wystarczające ogólne odesłanie do wniosku o dofinansowanie i studium wykonalności. Na obecnym etapie przygotowania projektu metodologia liczenia wskaźników powinna być już opracowana, przynajmniej w zakresie podstawowych założeń niezbędnych do oceny OZPI. Jeżeli szczegółowa metodologia została przygotowana w studium wykonalności, prosimy o przedstawienie jej syntetycznego ujęcia w odpowiedzi oraz odpowiednie odzwierciedlenie najważniejszych zasad w OZPI.

Prosimy o uzupełnienie dokumentacji o jednoznaczne przypisanie jednostek pomiaru do poszczególnych KPI oraz wskazanie podstawowych zasad przeliczania jednostek operacyjnych na wartości wskaźników projektowych. W szczególności prosimy o doprecyzowanie, jak będą liczone wartości KPI dotyczące liczby zdigitalizowanych dokumentów lub obiektów oraz liczby udostępnionych dokumentów lub obiektów. Prosimy również o wskazanie, które jednostki będą miały charakter wyłącznie pomocniczy lub techniczny i nie będą wpływać na wartości wskaźników produktu lub rezultatu.

4. Odpowiedź beneficjenta częściowo wyjaśnia sposób rozumienia KPI nr 3 „Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne on-line”, wskazując, że wskaźnik ten odnosi się do liczby podmiotów, a nie do liczby systemów, baz danych, repozytoriów, kolekcji, API ani źródeł danych integrowanych w projekcie. Beneficjent wskazał również cztery podmioty uwzględnione w kalkulacji wskaźnika: OPI PIB, AGH, PCSS oraz PP.

Należy jednak zauważyć, że z części 7 OZPI wynika znacznie szersza skala przepływów danych i integracji systemowych niż wartość KPI nr 3 określona na poziomie 4 podmiotów. W dokumentacji wskazano m.in. integrację Platformy WIEDZ@ z systemami partnerów projektu, NAVOICA, ZSUN II/RAD-on, Ludzie Nauki, AGH Digital AI System oraz wybranymi zewnętrznymi źródłami metadanych i danych. W tabeli przepływów danych opisano również zakres danych przekazywanych lub wykorzystywanych przez Platformę WIEDZ@, w tym dane o osobach, osiągnięciach, profilach badaczy, podmiotach nauki, działalności naukowej, projektach, publikacjach, ewaluacji, kursach oraz zasobach edukacyjnych. W związku z tym problemem nie jest wyłącznie sama wartość KPI nr 3, lecz brak jednoznacznego powiązania tego wskaźnika z opisem integracji i przepływów danych przedstawionym w części 7 OZPI. Dokumentacja wskazuje liczne systemy źródłowe i przepływy danych, natomiast nie wyjaśnia w wystarczający sposób, które z nich są traktowane jako nowe, mierzalne udostępnienie danych przez podmiot w rozumieniu KPI nr 3, a które pełnią jedynie funkcję integracyjną, techniczną, referencyjną lub pomocniczą.

Prosimy o uzupełnienie wyjaśnień poprzez powiązanie KPI nr 3 z opisem przepływów danych i systemów źródłowych wskazanych w części 7 OZPI. W szczególności prosimy o wskazanie, które systemy i przepływy danych opisane w części 7 są przypisane do czterech podmiotów wykazywanych w KPI nr 3, które pozostają poza tym wskaźnikiem oraz z jakiego powodu. Prosimy również o wyjaśnienie, w jaki sposób monitorowana będzie faktyczna skala integracji źródeł danych, skoro KPI nr 3 obejmuje wyłącznie liczbę podmiotów bezpośrednio udostępniających dane online.

5. Przyjmujemy wyjaśnienia beneficjenta dotyczące kwalifikacji e-usług do poziomu dojrzałości „Transakcja”. Prosimy jedynie, aby na dalszym etapie prac opisy poszczególnych e-usług jednoznacznie wskazywały czynność inicjowaną przez użytkownika, sposób automatycznego przetworzenia żądania oraz rezultat zwracany użytkownikowi lub systemowi zewnętrznemu.
6. Odpowiedź beneficjenta częściowo porządkuje rozróżnienie między zasobami podlegającymi digitalizacji a zasobami już istniejącymi w postaci cyfrowej. Beneficjent wskazuje, że projekt obejmuje zarówno digitalizację i opracowanie wybranych zasobów analogowych lub częściowo zdigitalizowanych, jak również integrację, agregację, indeksowanie i ponowne udostępnianie zasobów już cyfrowych. Przedstawione rozróżnienie ma jednak charakter ogólny i typologiczny.

Odpowiedź nie pozwala nadal jednoznacznie ocenić, jaka część zasobów wskazanych w OZPI będzie faktycznie podlegała nowej digitalizacji, a jaka część stanowi zasoby już istniejące w postaci cyfrowej, które będą jedynie integrowane, indeksowane, agregowane lub ponownie udostępniane za pośrednictwem Platformy WIEDZ@. Brakuje w szczególności syntetycznego przypisania konkretnych kategorii zasobów, kolekcji lub wolumenów do jednej z tych grup.

Wątpliwość ta jest szczególnie istotna w kontekście relacji między wskaźnikami dotyczącymi digitalizacji i udostępnienia. Projekt zakłada osiągnięcie 5 000 000 zdigitalizowanych dokumentów lub obiektów źródłowych oraz 5 000 000 dokumentów lub obiektów udostępnionych on-line. Takie wartości mogą sugerować, że cały wolumen zasobów udostępnionych on-line będzie odpowiadał wolumenowi zasobów nowo zdigitalizowanych w ramach projektu. Jednocześnie z odpowiedzi beneficjenta wynika, że część zasobów ma już istnieć w postaci cyfrowej i ma podlegać jedynie integracji, indeksowaniu, agregacji lub ponownemu udostępnieniu.

W związku z tym powstaje pytanie, jak zasoby już cyfrowe, które nie będą podlegały digitalizacji, zostały ujęte w wartościach wskaźników. Jeżeli mają być one również udostępniane za pośrednictwem Platformy WIEDZ@, to wartość wskaźnika dotyczącego udostępnienia powinna co do zasady obejmować szerszy zakres niż wartość wskaźnika dotyczącego nowej digitalizacji albo powinno zostać wyjaśnione, dlaczego obie wartości są identyczne. Jeżeli natomiast wskaźnik udostępnienia obejmuje wyłącznie zasoby nowo zdigitalizowane, należy wskazać, w jaki sposób będzie mierzony efekt integracji i ponownego udostępnienia zasobów już cyfrowych.

Bez takiego rozróżnienia nie jest możliwa ocena czy wskaźniki zostały przypisane prawidłowo oraz czy właściwie odzwierciedlają rzeczywisty zakres projektu. Obecna odpowiedź nie pozwala ustalić, czy zasoby już cyfrowe będą wliczane do KPI dotyczących udostępnienia, czy będą ujmowane wyłącznie w innych wskaźnikach, np. dotyczących API, baz danych, integracji, metadanych lub wykorzystania usług.

Prosimy o uzupełnienie wyjaśnień o syntetyczne wskazanie, jaka część zasobów będzie nowo digitalizowana, jaka będzie jedynie integrowana lub ponownie udostępniana oraz jak każda z tych kategorii wpływa na poszczególne KPI. W szczególności prosimy o wyjaśnienie relacji między wskaźnikiem 5 000 000 zdigitalizowanych dokumentów lub obiektów a wskaźnikiem 5 000 000 dokumentów lub obiektów udostępnionych on-line, z uwzględnieniem zasobów już istniejących w postaci cyfrowej.

7. MC dziękuje za szczegółowe wyjaśnienia dot. uwagi nr 7. W związku z przesłanymi wyjaśnieniami uznajemy uwagę za uzgodnioną.

ODPOWIEDZI OPI PIB NR 2 z 19.06.2026

AD 1. W związku z powyższym prosimy o uzupełnienie wyjaśnień o szczegółową analizę komplementarności funkcjonalnej WIEDZ@ względem KRONIK@ 2.0 oraz innych projektów FERC. Analiza powinna obejmować nie tylko zakres zasobów, lecz także zakres finansowanego oprogramowania, funkcjonalności portalu, mechanizmy wyszukiwania i agregacji metadanych, API, udostępnianie zasobów cyfrowych, integracje zewnętrzne oraz mechanizmy eliminujące ryzyko finansowania podobnych komponentów informatycznych ze środków publicznych.

ODPOWIEDŹ:

Beneficjent przyjmuje, że istotą uwagi Ministerstwa Cyfryzacji nie jest wyłącznie odrębność digitalizowanych kolekcji, ale również odrębność finansowanych funkcjonalności, architektury usług oraz mechanizmów udostępniania, katalogowania, wyszukiwania, integracji i ponownego wykorzystania zasobów. W związku z tym doprecyzowujemy, że odrębność projektu WIEDZ@ ma charakter nie tylko zasobowy, ale przede wszystkim funkcjonalny i architektoniczny.

WIEDZ@ nie jest kolejnym repozytorium treści ani portalem służącym wyłącznie do przechowywania kopii obiektów cyfrowych. Zasadniczym produktem projektu jest federacyjna warstwa pośrednia, obejmująca silnik indeksujący, indeks metadanych federacji, standard metadanych, mechanizmy integracji źródeł danych, wyszukiwanie semantyczne wspierane AI oraz API umożliwiające ponowne wykorzystanie danych. Dane właściwe pozostają w rozproszonych repozytoriach i systemach źródłowych, natomiast platforma WIEDZ@ przechowuje i udostępnia metadane oraz niezbędne informacje techniczne. Taki model odróżnia projekt od rozwiązań typu biblioteka cyfrowa albo portal treści, których funkcją jest przede wszystkim gromadzenie i udostępnianie zdigitalizowanych obiektów we własnym repozytorium.

WIEDZ@ realizuje cel działania FERC 2.3 w odniesieniu do zasobów nauki, edukacji, danych badawczych oraz wybranych zasobów kulturowych wspierających ponowne wykorzystanie informacji. Platforma ma zapewnić agregację i wzbogacanie metadanych, wyszukiwanie semantyczne, możliwość zadawania zapytań w języku naturalnym, interoperacyjność z europejskimi infrastrukturami otwartej nauki, w tym poprzez standard OAI-PMH, oraz powiązanie zasobów nauki z komponentem edukacyjnym NAVOICA. Nie jest to funkcjonalność tożsama z udostępnianiem zdigitalizowanego dziedzictwa w ramach jednego repozytorium, lecz budowa krajowego punktu dostępu do rozproszonych zasobów nauki, edukacji i danych badawczych.

W obszarze architektury WIEDZ@ opiera się na modelu federacyjnym: dane pozostają u właścicieli i w systemach źródłowych, a platforma integruje metadane oraz buduje indeks federacyjny. W rozwiązaniach typu portal lub biblioteka cyfrowa zasadniczą funkcją jest natomiast przechowywanie i serwowanie kopii zdigitalizowanych obiektów.

W obszarze metadanych WIEDZ@ zakłada harvesting, mapowanie i ujednolicanie metadanych pochodzących z wielu heterogenicznych źródeł nauki, takich jak RAD-on, Ludzie Nauki, PBN, SSH Open Marketplace, GoTriple, EUNICE i inne. W obszarze wyszukiwania projekt przewiduje wyszukiwanie hybrydowe: leksykalne, semantyczne, wektorowe i kontekstowe, w tym z wykorzystaniem komponentów AI. W obszarze API projekt zakłada utworzenie otwartych interfejsów, w tym OAI-PMH, umożliwiających włączenie zasobów do obiegu EOSC/OpenAIRE oraz integrację z narzędziami sektora nauki. W obszarze ponownego wykorzystania danych WIEDZ@ wspiera re-use danych badawczych zgodnie z zasadami FAIR Data, integrację z komponentem edukacyjnym oraz wykorzystanie danych dla polityki nauki.

Odnosząc się do argumentu, że w portalu KRONIK@ znajdują się wybrane dane badawcze, a standard metadanych KRONIK@ był opiniowany przez OPI PIB, należy wskazać, że okoliczności te nie przesądzają o tożsamości funkcjonalnej obu projektów. Przeciwnie — potwierdzają możliwość komplementarności i interoperacyjności rozwiązań. Fakt, że określone zasoby mogą mieć charakter graniczny, naukowo-kulturowy albo badawczo-archiwalny, nie oznacza, że jedna platforma wyczerpuje potrzeby całego ekosystemu. WIEDZ@ została zaprojektowana tak, aby istniejące zasoby cyfrowe, w tym potencjalnie także zasoby KRONIK@, mogły być indeksowane, odnajdywane i ponownie wykorzystywane bez ich powielania.

Nie można także zgodzić się z sugestią, że część funkcji WIEDZ@ mogłaby zostać prosto przejęta przez istniejące systemy. Żaden z istniejących systemów nauki, takich jak POL-on, PBN, RAD-on czy Ludzie Nauki, nie pełni obecnie funkcji federacyjnego, semantycznego punktu dostępu do rozproszonych danych badawczych i edukacyjnych, z re-use, API, integracją z europejskimi infrastrukturami otwartej nauki oraz warstwą AI. Systemy te są systemami źródłowymi, referencyjnymi albo raportowymi, które WIEDZ@ integruje, a nie zastępuje.

Tam, gdzie jest to możliwe i racjonalne, projekt nie buduje funkcji od nowa, lecz wykorzystuje istniejące komponenty. Dotyczy to m.in. uwierzytelniania poprzez System Centralnego Logowania, Węzeł Krajowy i ORCID, obsługi zgłoszeń poprzez ZSUN HelpDesk, komponentu edukacyjnego poprzez modyfikację istniejącej platformy NAVOICA, a także źródeł danych poprzez indeksowanie istniejących repozytoriów krajowych i europejskich zamiast tworzenia ich kopii. Ogranicza to ryzyko dublowania już na poziomie architektury i potwierdza, że finansowane oprogramowanie dotyczy funkcji nieobecnych w dotychczasowym ekosystemie.

PORÓWNANIE FUNKCJONALNOŚCI PORTALI WIEDZ@ I KRONIK@ W ZAKRESIE ROZWIĄZYWANIA PROBLEMY NISKIEJ FAKTYCZNEJ DOSTĘPNOŚCI DANYCH BADAWCZYCH

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
1	Standard metadanych dostosowany do precyzyjnego opisu danych	NIE	Standard metadanych KRONIK@ jest pomyślany przede wszystkim jako uniwersalny model agregacji zasobów nauki i kultury , ale jego rdzeń koncepcyjny jest silnie osadzony w logice dziedzictwa kulturowego: CIDOC-CRM, Dublin Core, EAD, LIDO, FRBR/FRBROO, obiekty kultury, archiwalia, biblioteki, muzea i opisy obiektów informacyjnych. Wprowadzie dokument przewiduje typ DC.Type.Dataset, ale zestaw danych mapuje jako szczególny przypadek klasy E31 Document, czyli jako dokument/opisany obiekt informacyjny, a nie jako pełnoprawny wynik procesu badawczego z metodologią, aparaturą, zmiennymi, pomiarami, wersjonowaniem, jakością i warunkami ponownego użycia.

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
	badawczych z dziedzin nauk inżyniersko-technicznych, nauk ścisłych i przyrodniczych, nauk medycznych i nauki o zdrowiu, nauki rolniczych, nauk społecznych, w ramach których tworzona jest większość danych badawczych wykorzystywanych w działalności B+R+I.		Największy problem polega na tym, że standard dobrze nadaje się do opisania, zidentyfikowania, powiązania i wyszukania zasobu, ale jest za mało precyzyjny do naukowego opisu danych badawczych. Dla danych badawczych nie wystarcza informacja typu: tytuł, twórca, data, temat, opis, format, identyfikator, relacja, prawa. Potrzebne są również informacje o tym, jak dane powstały, z jaką dokładnością, przy użyciu jakich metod, jakich narzędzi, na jakiej populacji lub próbie, w jakich warunkach, z jakimi ograniczeniami i w jakim zakresie można je ponownie wykorzystać. Dlatego standard metadanych przyjętych przez twórców portalu KRONIK@ nie jest wystarczający z poniższych powodów do realizacji celu portalu WIEDZ@ w postaci zmiany stanu, w którym repozytoria danych badawczych służą głównie do deponowania danych na stan, w którym repozytoria danych badawczych są źródłem danych, na których budowana jest wiedza naukowa i innowacje biznesowe: 1. Zbyt dokumentacyjny sposób traktowania datasetu W części dotyczącej DC.Type.Dataset standard wskazuje, że cały rekord Dublin Core jest mapowany do klasy E31 Document, a zestaw danych jest traktowany jako szczególny przypadek dokumentu. To rozwiązanie jest wygodne dla agregatora metadanych, ale redukuje dataset do obiektu podobnego do publikacji, raportu albo opisu archiwalnego. Dla danych badawczych to za mało, ponieważ dataset powinien być opisywany również jako: - wynik określonego procesu badawczego, - zbiór obserwacji, pomiarów, symulacji albo odpowiedzi, - struktura zmiennych i wartości, - obiekt posiadający wersje, zależności, pochodzenie i ograniczenia jakościowe, - zasób powiązany z metodą, aparaturą, protokołem, projektem, finansowaniem, publikacją i kodem analitycznym. Standard opisuje raczej metadane katalogowe datasetu, a nie jego metadane badawcze. 2. Zbyt słabe wykorzystanie DataCite, CERIF i OpenAIRE Standard powołuje się na DataCite, CERIF i OpenAIRE: DataCite definiuje jako schemat metadanych dla publikacji i cytowania danych badawczych, CERIF jako standard informacji o działalności badawczej, a OpenAIRE jako profil dla repozytoriów publikacji, archiwów danych i CRIS. Wskazuje też, że wytyczne OpenAIRE dla archiwów danych są oparte na DataCite, a dla systemów CRIS na CERIF. Słabość polega na tym, że dokument nie buduje konsekwentnego profilu aplikacyjnego dla danych badawczych opartego na DataCite/OpenAIRE/CERIF. W praktyce oznacza to ryzyko, że dane mogą być formalnie opisane, ale nie będą w pełni interoperacyjne z ekosystemem otwartej nauki: repozytoriami danych, agregatorami, systemami CRIS, mechanizmami cytowania datasetów i wymogami FAIR. 3. Brak silnych metadanych metodologicznych W dziedzinach nauk inżyniersko-technicznych, nauki ścisłych i przyrodniczych, nauk medycznych i nauki o zdrowiu, nauki rolniczych, nauk społecznych kluczowa jest nie tylko zawartość zbioru danych, ale również opis procesu ich uzyskiwania. Brak tego opisu jest podawany przez badaczy jako jeden z istotnych powodów niskiego poziomu powtórnego wykorzystania danych przez naukowców. Standard KRONIK@ przewiduje metadane opisowe, administracyjne, strukturalne, techniczne i kontekstowe, ale ich ujęcie pozostaje ogólne. Brakuje pól lub profili dla takich informacji jak: - metoda pozyskania danych, - protokół badawczy, - projekt eksperymentu, - kryteria doboru próby, - warunki pomiaru, - parametry eksperymentu, - sposób czyszczenia i przetwarzania danych, - procedura anonimizacji, - znane błędy i ograniczenia, - poziom niepewności, - kompletność danych, - zmiennie, jednostki, skale i kody braków danych. Bez ww. danych zestaw danych pozostaje trudny do odnalezienia, a jego ponowne użycie naukowe będzie dla badacza obarczone ryzykiem poważnych błędów metodologicznych. 3. Niewystarczające wsparcie dla danych tabelarycznych, pomiarowych i obliczeniowych

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			Standard wspomina o metadanych technicznych, takich jak format pliku, oprogramowanie, sprzęt, suma kontrolna, podpis cyfrowy, trwałość i możliwość odtworzenia pliku. To ważne, ale dotyczy głównie utrzymania obiektu cyfrowego. W przypadku danych badawczych potrzeba więcej: opisu struktury plików, schematu tabel, typów zmiennych, jednostek miary, wartości dopuszczalnych, słowników kodowych, relacji między tabelami, pipeline'u przetwarzania, środowiska obliczeniowego, wersji skryptów, zależności programistycznych i sposobu reprodukcji wyników. Dla nauk inżynieryjno-technicznych i ścisłych jest to szczególnie istotne, bo zestaw danych może obejmować np. dane sensoryczne, wyniki symulacji, dane CAD/CAE, logi maszyn, obrazy pomiarowe, szeregi czasowe, dane laboratoryjne albo duże macierze wyników. 4. Niedostateczne rozróżnienie typów danych badawczych Standard ma typ Dataset, ale nie widać w nim wystarczająco rozbudowanej typologii danych badawczych. Tymczasem inaczej powinno się opisywać: • dane obserwacyjne, • dane eksperymentalne, • dane symulacyjne, • dane ankietowe, • dane kliniczne, • dane obrazowe, • dane genomyczne, • dane geograficzne, • dane sensorowe, • dane jakościowe, • dane administracyjne wykorzystywane wtórnie w badaniach. Bez takiego rozróżnienia metadane będą zbyt płaskie. Dwa zestawy danych mogą mieć podobny opis katalogowy, choć ich wymagania dokumentacyjne, prawne i techniczne są zupełnie inne. 5. Zbyt ogólne podejście do słowników kontrolowanych Dokument słusznie podkreśla znaczenie słowników kontrolowanych, tezaursów, taksonomii i ontologii, wskazując, że są one warunkiem spójnego opisu, indeksowania, wyszukiwania i interoperacyjności. Problem polega na tym, że warstwa słownikowa jest opisana głównie ogólnie albo z silnym nachyleniem ku dziedzictwu kulturowemu. Wśród interfejsów słowników pojawiają się m.in. Getty, RISM, LIDO i Iconclass, czyli źródła właściwe dla sztuki, muzyki, muzealnictwa i ikonografii. Dla traktowania standardu jako dostosowanego do wymogów B+R+I konieczne są słowniki i ontologie dziedzinowe, np. dla nauk: • inżynieryjno-technicznych: słowniki materiałów, procesów technologicznych, urządzeń, norm, parametrów pomiarowych, • nauki ścisłych i przyrodniczych: taksonomie biologiczne, słowniki chemiczne, geologiczne, środowiskowe, astronomiczne, • nauk medycznych i nauki o zdrowiu: MeSH, ICD, SNOMED CT, terminologie badań klinicznych, klasyfikacje procedur i substancji, • nauki rolniczych: AGROVOC, taksonomie organizmów, słowniki upraw, gleboznawstwa, fitopatologii, • nauk społecznych: klasyfikacje zmiennych społecznych, metody badań, typy prób, narzędzia ankietowe, klasyfikacje terytorialne. 6. Słabe wsparcie dla aspektów etycznych i prawnych danych wrażliwych W naukach medycznych, naukach o zdrowiu i społecznych dane często dotyczą ludzi. Standard mówi o metadanych praw, licencjach i zarządzaniu prawami, ale to nie wystarcza do opisu danych badawczych zawierających dane osobowe, dane wrażliwe, dane kliniczne lub dane pochodzące z badań ankietowych. Potrzebne byłyby pola dotyczące m.in.: • podstawy prawnej przetwarzania, • statusu anonimizacji lub pseudonimizacji, • ograniczeń dostępu, • zgód uczestników, • zgody komisji bioetycznej,

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			- warunków wtórnego wykorzystania, - poziomu ryzyka reidentyfikacji, - populacji badanej i kryteriów włączenia/wyłączenia, - ograniczeń publikacji danych. Bez tego standard może być niewystarczający dla danych medycznych, psychologicznych, socjologicznych, pedagogicznych i ekonomicznych. 7. Ryzyko słabej porównywalności między dziedzinami Standard kładzie nacisk na interoperacyjność semantyczną, mapowania i grafową reprezentację danych. Jednak dla danych badawczych interoperacyjność wymaga nie tylko mapowania ogólnych elementów, ale także spójnego opisu jednostek, zmiennych, metod, populacji, instrumentów i jakości danych. Przykładowo pole „coverage” może oznaczać zasięg przestrzenny, czasowy, tematyczny albo populacyjny. W danych społecznych może chodzić o populację respondentów; w medycynie o grupę pacjentów; w rolnictwie o obszar upraw; w naukach przyrodniczych o lokalizację próbkowania; w inżynierii o zakres pracy urządzenia. Bez doprecyzowania semantycznego te informacje będą trudne do automatycznego porównywania. 8. Zbyt mało informacji o jakości danych Standard mówi o jakości metadanych jako założeniu oraz o walidacji, ale nie rozwija metadanych jakościowych właściwych dla danych badawczych. Dla danych badawczych (szczególnie ważne w medycynie, naukach przyrodniczych, inżynierii i rolnictwie, gdzie ponowne użycie danych bez informacji o jakości może prowadzić do błędnych wniosków) potrzebne są pola takie jak: - kompletność danych, - dokładność pomiaru, - precyzja, - niepewność, - błąd systematyczny, - liczba braków danych, - sposób imputacji, - poziom agregacji, - ograniczenia interpretacyjne, - kontrola jakości, - certyfikacja aparatury, - walidacja instrumentu badawczego. 9. Podsumowanie Standard metadanych systemu KRONIK@ może być dobrym standardem agregacyjnym i katalogowym: pozwala identyfikować zasoby, mapować metadane, budować graf powiązań, korzystać ze słowników i udostępniać dane przez API. Jego słabością w zastosowaniu do danych badawczych jest jednak to, że nie tworzy wystarczająco mocnego profilu naukowego zestawu danych badawczych, czyli nie odpowiada na problem, którego rozwiązanie jest istotą projektu WIEDZ@ - problem niskiej rzeczywistej dostępności danych badawczych, gdzie przez rzeczywistą dostępność beneficjent rozumie taki opis danych, aby naukowcy mogli je w sposób nie wymagający niewspółmiernego wysiłku (m.in. przez samodzielne ustalanie wszystkich metadanych) wykorzystać w procesie budowania wiedzy.
	Silnik indeksujący metadane badawcze dostosowany do specyfiki nauk inżyniersko-technicznych, nauki ścisłych i przyrodniczych, nauk medycznych i	NIE	Metoda działania portalu KRONIK@ nie odpowiada w pełni na zapotrzebowanie badaczy potrzebujących jednego miejsca dostępu do danych badawczych, umożliwiającego nie tylko ich odnalezienie, ale także ocenę przydatności i rzeczywiste ponowne wykorzystanie w pracy naukowej. Wynika to zarówno z przyjętej architektury pozyskiwania danych, jak i z charakteru standardu metadanych, w którym dane są opisywane przede wszystkim jako zasoby informacyjne podlegające identyfikacji, agregacji, wyszukiwaniu i długoterminowemu przechowywaniu, a nie jako pełnoprawne obiekty badawcze wymagające szczegółowej dokumentacji metodologicznej, technicznej, prawnej i jakościowej. Wirtualny harvester KRONIK@, instalowany w infrastrukturze instytucji indeksującej, zakłada rozproszoną odpowiedzialność za proces przygotowania, indeksowania i przekazywania danych. Taki model może być użyteczny dla agregacji heterogenicznych zasobów instytucjonalnych, zwłaszcza z obszaru kultury, archiwów, bibliotek, muzeów i ogólnych zasobów naukowych. Nie tworzy jednak jednolitego, centralnie kontrolowanego mechanizmu standaryzacji danych badawczych pochodzących z wielu repozytoriów. W efekcie jakość, kompletność i porównywalność metadanych zależą w dużym stopniu od praktyk poszczególnych instytucji oraz od tego, jakie informacje udostępniają one w swoich lokalnych systemach. Dla badacza oznacza to, że portal może wskazać istnienie zasobu, ale nie musi dostarczyć metadanych wystarczających do świadomej decyzji, czy dane mogą zostać wykorzystane w nowym badaniu.

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
	nauki o zdrowiu, nauki rolniczych, nauk społecznych,		Zapotrzebowanie badaczy jest szersze niż potrzeba katalogowego odnalezienia zbioru. Badacz poszukujący danych badawczych musi wiedzieć, jak dane zostały wytworzone, w jakim celu, na jakiej próbie, przy użyciu jakiej metody, aparatury, oprogramowania lub procedury, z jaką dokładnością, w jakich warunkach i z jakimi ograniczeniami. W naukach inżyniersko-technicznych istotne są m.in. parametry pomiaru, konfiguracja urządzeń, normy techniczne, modele, środowiska symulacyjne i wersje oprogramowania. W naukach medycznych i o zdrowiu kluczowe są informacje o populacji badanej, kryteriach włączenia i wyłączenia, statusie anonimizacji, zgodach etycznych i ograniczeniach dostępu. W naukach ścisłych i przyrodniczych znaczenie mają jednostki, niepewność pomiaru, warunki eksperymentu, protokoły laboratoryjne, taksonomie i lokalizacja próbkowania. W naukach rolniczych konieczny jest opis warunków glebowych, klimatycznych, odmian, zabiegów agrotechnicznych i sezonowości. W naukach społecznych niezbędne są informacje o metodzie badania, operacje losowania, konstrukcji próby, kwestionariuszu, zmiennych, kodowaniu, wagach statystycznych i zasadach dostępu do danych wrażliwych. Standard KRONIK@ nie został zaprojektowany przede wszystkim jako profil metadanych dla danych badawczych z tych dziedzin. Jego siłą jest elastyczne opisanie i powiązanie różnych typów zasobów, natomiast jego słabością — z punktu widzenia ponownego wykorzystania danych — jest zbyt ogólny poziom opisu zbioru danych. Dane badawcze są wynikiem procesu badawczego i powinny być opisane poprzez powiązanie z projektem, metodą, aparaturą, osobami, instytucjami, finansowaniem, publikacjami, kodem analitycznym, wersjami oraz ograniczeniami prawnymi i etycznymi. Bez tych informacji użytkownik może uzyskać dostęp do opisu zasobu, lecz nie uzyska wystarczającego kontekstu, aby odpowiedzialnie użyć danych w analizie naukowej. Portal WIEDZ@ odpowiada na tę lukę inną metodą działania. Zakładany silnik indeksujący nie ogranicza się do przyjęcia danych uprzednio zaindeksowanych w standardzie instytucji przekazującej, lecz harvestuje dane badawcze przez API z repozytoriów danych badawczych, a następnie przekształca je do indeksu zbudowanego według standardu metadanych zoptymalizowanego pod konkretne obszary nauki. Oznacza to przejście od modelu agregacji zasobów do modelu aktywnej normalizacji i wzbogacania metadanych badawczych. Dzięki temu możliwe jest zbudowanie jednego punktu dostępu, w którym dane z różnych repozytoriów są nie tylko zgromadzone, ale również opisane w sposób porównywalny, dziedzinowo adekwatny i użyteczny dla ponownego wykorzystania. Różnica między KRONIK@ a WIEDZ@ jest więc różnicą celu obu rozwiązań. • KRONIK@ odpowiada przede wszystkim na pytanie: gdzie znajduje się zasób i jak powiązać go z podstawowymi metadanymi. • WIEDZ@ ma odpowiadać na pytanie: czy dane nadają się do wykorzystania w konkretnym procesie badawczym, jakie mają parametry, ograniczenia i kontekst oraz jak można je porównać z innymi danymi z tej samej lub pokrewnej dziedziny. Dla badaczy ta druga funkcja jest decydująca. Samo odnalezienie zestawu danych jest istotne, ale oznacza jeszcze, że dane mogą zostać ponownie użyte. Warunkiem ponownego wykorzystania jest zrozumienie danych. Z tego powodu metoda działania KRONIK@, choć wartościowa dla integracji szerokiego spektrum zasobów nauki i kultury, nie zaspokaja specyficznej potrzeby środowiska badawczego w zakresie dostępu do danych badawczych. Nie zapewnia wystarczająco silnego, dziedzinowego i metodologicznego opisu danych, nie gwarantuje jednolitej normalizacji metadanych z wielu repozytoriów oraz nie tworzy indeksu zoptymalizowanego pod praktyczne scenariusze ponownego użycia danych. Portal WIEDZ@ jest uzasadniony jako odrębne rozwiązanie, ponieważ koncentruje się nie na katalogowaniu zasobów jako takich, lecz na budowie infrastruktury wyszukiwania, oceny i wykorzystania danych badawczych w naukach inżyniersko-technicznych, ścisłych i przyrodniczych, medycznych i o zdrowiu, rolniczych oraz społecznych.
	Generator metadanych	NIE	Generator metadanych stanowi kluczowy element projektu WIEDZ@, ponieważ rozwiązuje jeden z podstawowych problemów ograniczających rzeczywiste ponowne wykorzystanie danych badawczych: istnienie istotnych informacji opisujących dane wyłącznie wewnątrz plików, poza warstwą metadanych dostępną dla wyszukiwarek, agregatorów, systemów analitycznych i użytkowników końcowych. W wielu repozytoriach dane badawcze są formalnie zdeponowane, ale ich opis jest ograniczony do metadanych wprowadzonych ręcznie przez autora: tytułu, twórcy, daty, krótkiego opisu, słów kluczowych, licencji lub formatu pliku. Tymczasem znaczna część informacji potrzebnych do oceny wartości danych znajduje się bezpośrednio w plikach, np. 1. Metadane identyfikacyjne - opisują co to jest za plik i skąd pochodzi a. nazwa pliku / identyfikator eksperymentu b. unikalny identyfikator próbki (Sample ID) c. identyfikator projektu / kampanii badawczej d. wersja pliku lub danych 2. Metadane dotyczące aparatury - Opisują jakim urządzeniem wykonano pomiar a. typ urządzenia (np. SEM, NMR, LC-MS, spektrofotometr) b. producent i model c. numer seryjny aparatu d. wersja firmware / oprogramowania e. konfiguracja sprzętowa (np. detektory, moduły dodatkowe) 3. Parametry eksperymentu - kluczowe dla interpretacji i reprodukowalności a. ustawienia pomiarowe: i. napięcie, prąd, temperatura

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			ii. długość fali, zakres pomiarowy iii. czas ekspozycji / integracji b. tryb pracy urządzenia i. częstotliwość próbkowania ii. kalibracja (data i parametry) iii. użyte standardy / referencje 4. Informacje o próbce a. typ próbki (np. biologiczna, chemiczna, materiałowa) b. skład chemiczny / biologiczny c. przygotowanie próbki (np. filtracja, suszenie, barwienie) d. warunki przechowywania e. historia próbki (pochodzenie, batch) 5. Metadane czasowe i przestrzenne a. data i czas pomiaru (często z dokładnością do sekund lub ms) b. strefa czasowa c. lokalizacja eksperymentu (laboratorium, współrzędne GPS) d. identyfikator sesji pomiarowej 6. Dane o operatorze i procesie a. imię/nazwisko lub ID operatora b. jednostka / laboratorium c. procedura (SOP – Standard Operating Procedure) d. komentarze użytkownika e. notatki eksperymentalne 7. Metadane jakości danych do oceny wiarygodności wyników a. poziom szumu / sygnału (SNR) b. błędy pomiarowe / niepewność c. flagi jakości (np. "valid", "outlier") d. informacje o brakujących danych e. logi błędów aparatury 8. Struktura i format danych umożliwiające interpretację pliku a. format pliku (np. CSV, HDF5, NetCDF, FITS) b. struktura datasetu (kolumny, zmienne) c. jednostki fizyczne (np. °C, nm, Hz) d. schemat kodowania / kompresji 9. Metadane przetwarzania (jeśli dane były przetwarzane) a. zastosowane algorytmy (np. filtracja, normalizacja, baseline correction) b. wersja pipeline'u analitycznego c. parametry przetwarzania d. historia transformacji danych (provenance) 10. Metadane zgodności i zarządzania umożliwiające zarządzanie danymi – np. zgodność ze standardami ISO Udostępnienie danych w sensie naukowym wymaga natomiast, aby badacz mógł dane odnaleźć, zrozumieć, porównać z innymi zbiorami, ocenić ich przydatność, sprawdzić ograniczenia i zastosować je w nowym procesie badawczym. Jeżeli informacje zawarte w plikach nie zostaną wyekstrahowane i przeniesione do warstwy metadanych, pozostają niewidoczne dla wyszukiwania maszynowego. System może wiedzieć, że istnieje plik z danymi, ale nie wie, jakie zmienne zawiera, jakiego typu są to pomiary, jakiego okresu dotyczą, jakie mają parametry techniczne ani czy są porównywalne z innymi zbiorami. Problem braku dostępu do technicznie istniejących metadanych ma charakter systemowy. Naukowcy deponujący dane badawcze zwykle koncentrują się na zakończeniu projektu, publikacji wyników i spełnieniu formalnego obowiązku zdeponowania danych. Ręczne wydobywanie metadanych z plików, ich porządkowanie, mapowanie do standardu, uzupełnianie słowników i opisywanie struktury danych jest procesem czasochłonnym, wymagającym dodatkowych kompetencji oraz niewynagradzanym w typowych mechanizmach oceny dorobku naukowego. W praktyce oznacza to, że nawet dobrze przygotowane pliki danych trafiają do repozytoriów z metadanymi szczątkowymi.

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			Generator metadanych w systemie WIEDZ@ usuwa tę barierę przez automatyzację procesu, którego nie da się skutecznie oprzeć wyłącznie na ręcznej pracy deponentów. Mechanizm przeglądający poszczególne typy plików może identyfikować i wydobywać informacje z ich wewnętrznej struktury, a następnie zapisywać je w indeksie metadanych badawczych zgodnym ze standardem zoptymalizowanym również dla nauk inżynierjno-technicznych, ścisłych i przyrodniczych, medycznych i o zdrowiu, rolniczych oraz społecznych. Oznacza to, że metadane przestają być wyłącznie deklaracją autora, a stają się wynikiem systematycznej analizy samych danych. WIEDZ@ nie tylko przejmuje opis dostarczony przez repozytorium, ale wzbogaca go o informacje, które w przeciwnym razie pozostałyby ukryte wewnątrz plików. Znaczenie tego rozwiązania jest szczególnie duże w przypadku danych badawczych, ponieważ ich użyteczność zależy od precyzyjnego kontekstu technicznego i metodologicznego. W naukach inżynierjno-technicznych metadane wydobyte z plików mogą obejmować parametry pomiarów, konfiguracje urządzeń, formaty danych sensorowych, wersje modeli, cechy symulacji lub informacje o środowisku obliczeniowym. W naukach ścisłych i przyrodniczych mogą dotyczyć jednostek, współrzędnych, czasu obserwacji, parametrów aparatury, sekwencji pomiarów lub właściwości próbek. W naukach medycznych i o zdrowiu mogą pomagać w identyfikacji typu danych, struktury plików, dat, parametrów urządzeń diagnostycznych lub zakresu zmiennych, oczywiście z zachowaniem mechanizmów ochrony danych wrażliwych. W naukach rolniczych mogą ujawniać lokalizację, sezonowość, parametry środowiskowe, serie pomiarowe i dane obrazowe. W naukach społecznych mogą wspierać rozpoznawanie struktury arkuszy, zmiennych, kodów odpowiedzi, fał badania i relacji między plikami. Dzięki generatorowi metadanych portal WIEDZ@ może przekształcić repozytoria danych badawczych z rozproszonych magazynów plików w źródło danych możliwych do maszynowego przeszukiwania i porównywania. Badacz nie będzie zmuszony otwierać kolejnych repozytoriów, pobierać plików, ręcznie sprawdzać ich zawartości i dopiero wtedy oceniać, czy zawierają użyteczne dane. System będzie mógł udostępnić wyszukiwanie według cech faktycznie wynikających z zawartości plików: typu danych, zmiennych, parametrów, zakresu czasowego, geolokalizacji, formatu, jednostek, typu aparatury, struktury tabelarycznej lub rodzaju pomiaru. To zasadniczo zmienia jakość dostępu. Użytkownik nie odnajduje jedynie opisu zbioru, lecz otrzymuje możliwość wyszukiwania po cechach, które decydują o wartości danych dla konkretnego badania. Mechanizmy KRONIK@ nie zapewniają tego efektu, ponieważ ich podstawową funkcją jest agregacja i import metadanych przygotowanych przez jednostkę oraz ich mapowanie do przyjętego standardu. Taki model jest wartościowy dla porządkowania i wyszukiwania szerokiego spektrum zasobów nauki i kultury, ale nie rozwiązuje problemu niskiej podaży metadanych w wyszukiwarce (mimo ich istnienia wewnątrz plików danych badawczych). Jeżeli instytucja lub autor nie wyekstrahuje informacji z plików i nie umieści ich w warstwie metadanych, portal oparty na przejmowaniu przygotowanych metadanych nie będzie miał do nich efektywnego dostępu. W rezultacie dane pozostaną formalnie obecne w systemie, ale w praktyce niewidoczne dla zaawansowanego wyszukiwania i nieporównywalne z innymi zbiorami. WIEDZ@ wprowadza mechanizm, który odpowiada na realne zachowania naukowców i realne ograniczenia repozytoriów. Nie zakłada, że każdy autor samodzielnie przygotowuje pełny, dziedzinowo poprawny i maszynowo użyteczny opis danych. Zakłada natomiast, że system powinien aktywnie wspierać wydobywanie, normalizację i indeksowanie metadanych. To podejście jest konieczne, jeśli celem projektu jest nie tylko formalna dostępność danych, ale ich faktyczna używalność w dalszych badaniach. Generator metadanych staje się więc elementem infrastruktury naukowej, który przekształca zdeponowane pliki w zasoby możliwe do odnalezienia, zrozumienia, selekcji, porównania i ponownego wykorzystania. Z tego względu generator metadanych należy traktować jako jeden z centralnych komponentów projektu WIEDZ@. Bez niego portal byłby jedynie kolejnym punktem agregacji opisów dostarczanych przez repozytoria, obciążonym tymi samymi ograniczeniami jakości i kompletności metadanych, które występują obecnie. Z generatorem metadanych WIEDZ@ może natomiast pełnić funkcję aktywnej warstwy wzbogacającej dane badawcze: odkrywać informacje zapisane w plikach, przenosić je do indeksu, porządkować według standardu dziedzinowego i udostępniać w sposób przydatny dla maszynowego wyszukiwania oraz naukowego ponownego użycia danych. To właśnie ta funkcja odróżnia rzeczywiste udostępnienie danych badawczych od samego ich zdeponowania.
	Moduł budowania wiedzy z danych	NIE	Moduł budowania wiedzy z danych stanowi kluczowy element portalu WIEDZ@, ponieważ przesuwa punkt ciężkości z samego odnalezienia danych badawczych na ich rzeczywiste przygotowanie do ponownego użycia. W praktyce badacz nie potrzebuje jedynie informacji, że dany zbiór istnieje i może zostać pobrany. Potrzebuje odpowiedzi na pytania, czy dane nadają się do wykorzystania w jego badaniu, jakie zawierają ograniczenia, jakie ryzyka interpretacyjne należy uwzględnić, czy występują w nich powtórzenia, niespójności lub braki oraz jakie działania należy wykonać, aby praca ze zbiorem była efektywna i metodologicznie poprawna. Samo udostępnienie danych w repozytorium nie rozwiązuje tego problemu. Dane badawcze są często złożone, wieloplikowe, częściowo opisane, niespójne, zawierają duplikaty, różne warianty tych samych obserwacji, odmienne formaty, niejednoznaczne nazwy zmiennych albo ograniczenia wynikające z metody pozyskania. Badacz, który trafia na taki zbiór, musi zwykle samodzielnie wykonać czasochłonną pracę wstępną: przejrzeć strukturę danych, rozpoznać ich zawartość, ocenić jakość, zidentyfikować ograniczenia, usunąć duplikaty i zdecydować, jakie operacje przygotowawcze są konieczne. Bez wsparcia systemowego bariera ponownego użycia danych pozostaje wysoka, nawet jeśli dane są formalnie otwarte. Moduł budowania wiedzy w portalu WIEDZ@ odpowiada na problem pracochłonności pierwotnego rozpoznania znaczenia danych i ich oceny. Jego zadaniem jest wstępna analiza danych znalezionych przez naukowca, wskazanie najważniejszych informacji o zbiorze, zidentyfikowanie ograniczeń, przeprowadzenie deduplikacji oraz zaproponowanie dalszych działań porządkujących i analitycznych. Dzięki temu użytkownik nie otrzymuje jedynie listy wyników wyszukiwania, lecz pierwszą warstwę interpretacji danych. System wspiera go w przejściu od etapu „znalazłem potencjalnie interesujący zbiór” do etapu „rozumiem, co zawiera ten zbiór, jakie ma ograniczenia i jak mogę z nim pracować”. Znaczenie tego rozwiązania jest szczególne w naukach inżynierjno-technicznych, ścisłych i przyrodniczych, medycznych i o zdrowiu, rolniczych oraz społecznych, ponieważ w każdej z tych dziedzin ponowne użycie danych wymaga oceny kontekstu, jakości i struktury zbioru. W danych pomiarowych istotne są powtórzenia, dokładność i warunki pozyskania. W danych medycznych i społecznych kluczowe są ograniczenia metodologiczne, zakres zmiennych, poziom anonimizacji i możliwość łączenia danych. W danych rolniczych i przyrodniczych znaczenie mają lokalizacja, czas, sezonowość i porównywalność obserwacji. Moduł AI może wstępnie uporządkować te informacje i wskazać badaczowi elementy wymagające uwagi.

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			Mechanizmy portalu KRONIK@ nie zapewniają takiego efektu, ponieważ ich zasadniczą funkcją jest agregacja, import, mapowanie, wzbogacanie słownikowe i łączenie metadanych. Są to funkcje ważne dla budowy katalogu zasobów i poprawy ich wyszukiwalności, ale nie zastępują analitycznego wsparcia badacza w pracy z konkretnym zbiorem danych. WIEDZ@ pozwoli nie tylko odnaleźć dane badawcze, ale również znacząco obniżyć barierę ich wstępnej analizy - pomóc zrozumieć dane, ocenić ich użyteczność oraz przygotować je do dalszej pracy badawczej. Z tego powodu moduł budowania wiedzy należy traktować jako jeden z elementów przesadzających o jakości rzeczywistego udostępnienia danych badawczych. Bez niego portal byłby przede wszystkim narzędziem dostępu do opisów i plików. Z nim staje się środowiskiem wspierającym naukowca w pierwszej, najbardziej pracochłonnej fazie ponownego wykorzystania danych: rozpoznaniu, ocenie, oczyszczeniu i zaplanowaniu dalszej pracy ze zbiorem.
	Zintegrowana warstwa edukacyjna	NIE	Warstwa edukacyjna stanowi istotny element Platformy WIEDZ@, dane badawcze stają się realnie użyteczne dopiero wtedy, gdy użytkownik potrafi je właściwie zinterpretować, ocenić ich jakość, rozpoznać ograniczenia, dobrać odpowiednie metody analizy oraz świadomie zastosować je w nowym kontekście badawczym, edukacyjnym, gospodarczym lub społecznym. Sama obecność danych w repozytorium, nawet dobrze opisanych metadanymi, nie gwarantuje jeszcze ich ponownego wykorzystania. Dla wielu użytkowników barierą jest brak kompetencji pozwalających przejść od znalezienia zestawu danych do jego odpowiedzialnego użycia. Platforma WIEDZ@ ma odpowiadać na potrzeby zróżnicowanych grup odbiorców: doświadczonych badaczy, młodych naukowców, doktorantów, przedsiębiorców, przedstawicieli administracji publicznej. W przypadku użytkowników mniej doświadczonych szczególnie ważne jest zapewnienie wsparcia, które nie ogranicza się do prezentacji wyników wyszukiwania, lecz pomaga zrozumieć, z jakim typem danych mają do czynienia i jakie działania należy wykonać, aby dane te mogły zostać wykorzystane prawidłowo. Integracja Platformy WIEDZ@ z systemem NAVOICA tworzy w tym zakresie istotną wartość dodaną, ponieważ pozwala powiązać dostęp do danych z dostępem do kontekstowych materiałów edukacyjnych. Kontekstowość warstwy edukacyjnej jest kluczowa. Użytkownik, który odnajduje określony zbiór danych, nie powinien otrzymywać wyłącznie ogólnej listy kursów z zakresu analizy danych. Powinien otrzymać podpowiedź dopasowaną do charakteru zbioru danych, dziedziny nauki, poziomu złożoności, formatu plików oraz potencjalnych problemów metodologicznych. Inne wsparcie wymaga użytkownik analizujący dane ankietowe z nauk społecznych, innego osoba pracująca z danymi obrazowymi, innego badacz korzystający z danych medycznych, danych sensorowych, danych środowiskowych, danych rolniczych, wyników eksperymentów laboratoryjnych albo danych symulacyjnych. Materiały edukacyjne dobrane do konkretnego typu zestawu danych mogą wyjaśniać, jak czytać strukturę danych, jak interpretować zmienne, jak oceniać kompletność zbioru, jak rozpoznawać błędy, jak wykonywać podstawowe operacje czyszczenia, jak uwzględniać ograniczenia metodologiczne i jak dobrać adekwatne narzędzia analityczne. W ten sposób Platforma WIEDZ@ jest środowiskiem wspierającym cały pierwszy etap pracy z danymi: od wyszukiwania, przez zrozumienie, po przygotowanie do analizy. Warstwa edukacyjna wzmacnia działanie modułów odpowiedzialnych za generowanie metadanych, budowę indeksu, analizę zbiorów danych, wskazywanie ograniczeń i rekomendowanie działań przygotowawczych. Jeżeli system wskaże użytkownikowi, że zbiór zawiera dane tabelaryczne, zmienne ilościowe, braki danych, duplikaty albo zróżnicowane jednostki miary, może jednocześnie zaproponować materiał edukacyjny wyjaśniający, czym są te problemy i jak należy z nimi postępować. Jeżeli użytkownik trafia na dane medyczne lub społeczne, system może wskazać szkolenia dotyczące anonimizacji, etyki, interpretacji danych wrażliwych, ograniczeń próby badawczej lub zasad wtórnego wykorzystania danych. Jeżeli użytkownik korzysta z danych przyrodniczych albo rolniczych, może otrzymać materiały dotyczące pracy z danymi przestrzennymi, sezonowością, jednostkami pomiarowymi i zmiennością środowiskową. Tak rozumiana edukacja jest elementem warunkującym skuteczność całego modelu ponownego wykorzystania danych badawczych. W praktyce wiele danych pozostaje niewykorzystanych nie dlatego, że są niedostępne, lecz dlatego, że potencjalni użytkownicy nie wiedzą, jak z nimi pracować. Dotyczy to szczególnie przedsiębiorców, którzy mogliby wykorzystywać dane badawcze do tworzenia produktów, usług, analiz rynkowych, rozwiązań technologicznych lub innowacji społecznych, ale często nie posiadają kompetencji akademickich pozwalających samodzielnie ocenić wartość naukową datasetu. Dotyczy to również młodych naukowców, którzy dopiero uczą się praktyk otwartej nauki, zarządzania danymi, statystyki, metodologii i reprodukowalności badań. Warstwa edukacyjna zmniejsza próg wejścia, skracając czas potrzebny na rozpoczęcie pracy z danymi i ogranicza ryzyko błędnej interpretacji. Z punktu widzenia polityki otwartej nauki integracja WIEDZ@ z NAVOICA wzmacnia także zasadę odpowiedzialnego udostępniania danych. Otwartość nie powinna oznaczać wyłącznie publikacji plików. Powinna oznaczać zapewnienie takich warunków, w których dane mogą być odnalezione, zrozumiane, ocenione i wykorzystane zgodnie z ich przeznaczeniem oraz ograniczeniami. Warstwa edukacyjna wspiera każdy z tych etapów. Pomaga użytkownikowi zrozumieć, że dataset nie jest neutralnym zbiorem liczb lub plików, lecz wynikiem określonego procesu badawczego, obciążonym określonymi założeniami, metodą pozyskania, zakresem, błędami, brakami i ograniczeniami. Dzięki temu zmniejsza się ryzyko powierzchownego, przypadkowego lub nieuprawnionego użycia danych. Istotne jest również to, że warstwa edukacyjna może być powiązana z konkretnymi scenariuszami pracy użytkownika. Użytkownik nie musi z góry wiedzieć, jakiego kursu potrzebuje. To Platforma WIEDZ@, na podstawie typu odnalezionych danych, może zasugerować właściwy materiał: wprowadzenie do pracy z danymi tabelarycznymi, podstawy analizy statystycznej, czyszczenie danych, deduplikację, pracę z metadanymi, analizę danych przestrzennych, podstawy wizualizacji, wykorzystanie danych w badaniach społecznych, zasady interpretacji danych medycznych, przygotowanie danych do uczenia maszynowego albo ocenę jakości datasetu. Taki model jest bardziej efektywny niż tradycyjna biblioteka kursów, ponieważ edukacja pojawia się w momencie faktycznej potrzeby, w bezpośrednim związku z problemem, który użytkownik próbuje rozwiązać. Warstwa edukacyjna ma także znaczenie dla wyrównywania szans w dostępie do danych badawczych. Bez niej największą korzyść z Platformy WIEDZ@ odnoszą przede wszystkim użytkownicy już posiadający wysokie kompetencje analityczne, metodologiczne i techniczne. Integracja z NAVOICA pozwala rozszerzyć krąg odbiorców i umożliwić korzystanie z danych także tym grupom, które nie dysponują zapleczem dużych zespołów badawczych, wyspecjalizowanych analityków lub jednostek wsparcia danych. Dzięki temu Platforma WIEDZ@ może pełnić funkcję demokratyzującą dostęp do wiedzy: nie tylko udostępnia dane, ale również pomaga użytkownikom rozwijać kompetencje potrzebne do ich wykorzystania. Z perspektywy jakości badań warstwa edukacyjna może ograniczać liczbę błędów wynikających z niewłaściwego użycia danych. Użytkownicy mniej doświadczeni często mogą nie rozpoznawać problemów takich jak niereprezentatywność próby, nieporównywalność zmiennych, braki danych, różne definicje pojęć, ograniczenia licencji, konieczność anonimizacji, efekt doboru próby, brak kontroli nad zmiennymi zakłócającymi albo różnice między korelacją a zależnością przyczynową. Materiały edukacyjne zaproponowane w odpowiednim momencie

LP.	Funkcjonalność Platformy WIEDZ@	Czy funkcjonalność istnieje w systemie KRONIK@	Porównanie działania Portalu WIEDZ@ i portalu KRONIK@ w zakresie omawianej funkcjonalności
			mogą pełnić funkcję ostrzegawczą i instruktażową. Nie zastąpią samodzielnej oceny naukowej, ale mogą znacząco podnieść jakość pierwszego kontaktu z danymi i pomóc użytkownikowi uniknąć podstawowych błędów. Integracja z NAVOICA wzmacnia również trwałość efektów projektu. Platforma technologiczna umożliwia znalezienie i analizę danych, natomiast warstwa edukacyjna buduje kompetencje użytkowników, które pozostają z nimi także poza pojedynczą sesją pracy. Użytkownik, który dzięki kontekstowym szkoleniom nauczy się oceniać jakość danych, rozpoznawać ograniczenia datasetu i stosować właściwe procedury przygotowawcze, będzie lepiej korzystać z kolejnych zbiorów, tworzył lepsze analizy i bardziej świadomie deponował własne dane. W dłuższej perspektywie może to podnieść ogólną kulturę pracy z danymi badawczymi w Polsce: zarówno po stronie odbiorców danych, jak i ich wytwórców. Warstwa edukacyjna jest także istotna dla przedsiębiorców i użytkowników spoza sektora nauki. Otwarte dane badawcze mogą stanowić podstawę innowacji, analiz technologicznych, produktów cyfrowych, usług eksperckich, modeli predykcyjnych i decyzji inwestycyjnych. Jednak użytkownicy gospodarczy często potrzebują szybkiej orientacji, czy dane są wiarygodne, aktualne, kompletne i możliwe do połączenia z innymi źródłami. Kontekstowe szkolenia mogą wyjaśniać, jak oceniać przydatność datasetu w zastosowaniach praktycznych, jak rozumieć ograniczenia wynikające z metody badawczej, jak przygotować dane do analizy i kiedy konieczna jest konsultacja ekspercka. Dzięki temu WIEDZ@ może stać się nie tylko narzędziem naukowym, ale również elementem infrastruktury transferu wiedzy do gospodarki. Warstwa edukacyjna uzupełnia również działanie modułu budowania wiedzy opartego na AI. Moduł AI może wstępnie przeanalizować zbiór, wskazać ograniczenia, wykryć duplikaty i zaproponować dalsze działania, ale użytkownik nadal powinien rozumieć sens tych rekomendacji. Edukacja kontekstowa pozwala wyjaśnić, dlaczego system rekomenduje deduplikację, standaryzację jednostek, uzupełnienie braków, analizę rozkładu zmiennych, sprawdzenie wartości odstających albo weryfikację jakości metadanych. W ten sposób AI nie działa jak czarna skrzynka wydająca polecenia, lecz staje się elementem szerszego środowiska wspierającego świadomą pracę z danymi. Użytkownik nie tylko wykonuje rekomendowane działania, ale uczy się, dlaczego są one ważne. W tym sensie warstwa edukacyjna wzmacnia zaufanie do Platformy WIEDZ@. Użytkownik otrzymuje nie tylko wynik działania systemu, ale również ścieżkę zrozumienia tego wyniku. Ma możliwość pogłębienia wiedzy, sprawdzenia podstaw metodologicznych i lepszego przygotowania się do samodzielnej pracy. Jest to szczególnie ważne w systemie, który ma wspierać użytkowników o różnym poziomie kompetencji. Zaufanie do danych i narzędzi nie wynika wyłącznie z jakości technologii, lecz także z przejrzystości procesu, zrozumiałości rekomendacji i możliwości podnoszenia kompetencji przez użytkownika. Warstwa edukacyjna odróżnia Platformę WIEDZ@ od rozwiązań, które ograniczają się do funkcji agregacji, indeksowania i wyszukiwania. Portal, który pokazuje użytkownikowi wyłącznie listę datasetów, odpowiada na pytanie: gdzie znajdują się dane. Platforma wyposażona w kontekstową warstwę edukacyjną odpowiada na pytania znacznie ważniejsze z punktu widzenia ponownego użycia: czym są te dane, jak można z nimi pracować, czego trzeba się nauczyć, aby ich użycie było poprawne, jakie ryzyka należy uwzględnić i jakie kompetencje są potrzebne do wykonania dalszej analizy. Dzięki temu WIEDZ@ realizuje pełniejszą definicję dostępności danych badawczych — dostępności rozumianej nie tylko jako możliwość pobrania pliku, lecz także jako możliwość świadomego, kompetentnego i odpowiedzialnego wykorzystania danych. Z tego względu integracja z NAVOICA powinna być traktowana jako jeden z kluczowych komponentów Platformy WIEDZ@. Umożliwia ona połączenie infrastruktury danych z infrastrukturą edukacyjną, a tym samym tworzy warunki do rzeczywistego upowszechnienia ponownego wykorzystania danych badawczych. WIEDZ@ nie będzie wyłącznie miejscem, w którym użytkownik znajduje dane. Będzie środowiskiem, w którym użytkownik może rozpocząć pracę z danymi, zrozumieć ich charakter, rozpoznać własne braki kompetencyjne i skorzystać z materiałów edukacyjnych dobranych do konkretnego datasetu. Taki model odpowiada na realne potrzeby młodych naukowców, naukowców-amatorów, przedsiębiorców i wszystkich użytkowników, którzy chcą korzystać z danych badawczych, ale potrzebują wsparcia w przejściu od dostępu technicznego do rzeczywistego, merytorycznego wykorzystania danych.

Ad. 2 Prosimy o uzupełnienie dokumentacji o bardziej przejrzyste zestawienie najważniejszych grup kosztów związanych z digitalizacją i udostępnieniem zasobów. W szczególności prosimy o wskazanie, jaka część budżetu została przeznaczona na właściwe prace digitalizacyjne, opracowanie i przygotowanie metadanych, przetwarzanie OCR/HTR lub inne przetwarzanie treści, kontrolę jakości, przechowywanie danych oraz techniczne przygotowanie zasobów do udostępnienia. Takie zestawienie jest niezbędne do oceny, czy deklarowane wskaźniki digitalizacji i udostępnienia zasobów mają realne pokrycie w strukturze budżetu projektu.

Ad 2. Przejrzystość podziału kosztów oraz realność założeń digitalizacyjnych

Beneficjent uzupełnia wcześniejsze wyjaśnienia dotyczące struktury kosztów projektu. Jednocześnie ponownie podkreśla, że szczegółowa ocena zasadności, racjonalności i kwalifikowalności wydatków należy do etapu oceny wniosku o dofinansowanie, studium wykonalności i załączników finansowych przez właściwe instytucje oraz ekspertów CPPC. OZPI nie jest dokumentem służącym do ostatecznego przesądzania kwalifikowalności poszczególnych kategorii kosztów, lecz opisem założeń projektu informatycznego.

Z danych zawartych w dokumentacji projektowej wynika, że koszty projektu obejmują w szczególności: personel projektu odpowiedzialny za wytworzenie, integrację, digitalizację, testy i komponenty AI, koszty pośrednie, usługi zewnętrzne, wartości niematerialne i prawne oraz środki trwałe/dostawy. Dominacja kosztów personelu wynika z charakteru projektu: WIEDZ@ jest przedsięwzięciem wytwórczym, integracyjnym, badawczo-technicznym i wysokospecjalistycznym. Kluczowe rezultaty — silnik indeksujący, standard metadanych, komponenty AI, API, digitalizacja i opracowanie zasobów — powstają przede wszystkim jako praca interdyscyplinarnych zespołów, a nie jako zakup gotowego produktu.

Nie można zatem utożsamiać kosztów budowy i rozwoju oprogramowania wyłącznie z budową „kolejnego portalu”. W tej kategorii mieszczą się różne komponenty techniczne i funkcjonalne, w tym koncepcja biznesowo-techniczna, architektura, model danych, standard metadanych, analiza potrzeb, UX/UI/WCAG, onboarding i integracja źródeł danych, silnik indeksujący, generator metadanych, indeks i logika federacyjnego wyszukiwania, infrastruktura wdrożeniowa, framework węzłów brzegowych, bezpieczeństwo, tożsamość, komponenty AI, portal użytkownika, rekomendacje, API, komponent edukacyjny, testy, audyty, odbiory, wdrożenie i upowszechnienie.

W ramach projektu zaplanowano 11 skonsolidowanych zadań:

Nr	Zadanie	Lider	Uczestnicy	Szacowany koszt zadania do wniosku o dofinansowanie
1	Zarządzanie projektem, nadzór merytoryczny, koordynacja partnerstwa i sprawozdawczość	OPI	wszyscy partnerzy	862 318,00 zł
2	Opracowanie koncepcji biznesowo-technicznej, architektury rozwiązania, modelu danych, standardu metadanych i design systemu	OPI	PP, PCSS, Ł-AI, AGH	1 075 450,00 zł
3	Analiza potrzeb grup docelowych, badania użytkowników, badanie bazowe i końcowe satysfakcji użytkowników, UX/UI/WCAG oraz przełożenie wyników na wymagania systemowe	OPI	Ł-AI, PP, PCSS, MNiSW	1 954 390,00 zł
4	Inwentaryzacja, przygotowanie i onboarding zasobów partnerów do udostępnienia i re-use oraz integracja pierwszych źródeł	OPI	PP, PCSS, AGH	856 150,00 zł
5	Budowa silnika indeksującego, generatora metadanych, indeksu metadanych i logiki federacyjnego wyszukiwania	PCSS	PP, OPI, AGH, Ł-AI	2 958 393,00 zł
6	Budowa infrastruktury wdrożeniowej, środowisk, frameworku węzłów brzegowych oraz mechanizmów bezpieczeństwa i federacyjnej tożsamości	PP (w uzgodnieniu z PCSS)	PCSS, AGH, OPI	3 100 420,00 zł
7	Budowa modułu przygotowania / budowania wiedzy z danych	Ł-AI	PCSS, PP, OPI	3 489 620,00 zł
8	Budowa Portalu WIEDZ@: wyszukiwarka, warstwa użytkownika, konta, uwierzytelnianie, rekomendacje szkoleń	OPI	Ł-AI, PCSS, PP, AGH	5 366 750,00 zł
9	Budowa repozytorium, digitalizacja, OCR/HTR, tłumaczenia, opracowanie metadanych i semantyczne przetwarzanie zasobów archiwalnych AGH	AGH	OPI	4 263 960,00 zł
10	Integracja komponentów AGH i usług AI z architekturą WIEDZ@, brokerem, wyszukiwarką, API i mechanizmami dostępu	AGH	OPI, PCSS	2 375 300,00 zł
11	Budowa komponentu edukacyjnego, Brzegowej NAVOICA, testy końcowe, audyty, odbiory, wdrożenie i upowszechnienie	OPI	Ł-AI, MNiSW, PP, PCSS, AGH	6 521 970,00 zł

Działania digitalizacyjne nie są ukryte w kosztach oprogramowania – stanowią wyodrębniony blok realizowany przez AGH w zadaniach 9 i 10. Łączna wartość kosztów bezpośrednich tych zadań wynosi ok. 6 519 260 zł, a wraz z przypisanymi kosztami pośrednimi AGH w wysokości 952 710 zł digitalizacja i opracowanie zasobów obejmują łącznie ok. 7,3 mln zł.

W ramach tej kwoty mieszczą się prace, o które pyta Ministerstwo, w szczególności:

- przygotowanie materiałów i organizacja procesu digitalizacji, w tym udział osób w digitalizacji w Instytucie Hoovera – 100 000 zł,
- usługi zewnętrzne oraz delegacje digitalizacyjne AGH – 463 100 zł;
- skanowanie oryginałów w wysokiej rozdzielczości oraz przetwarzanie kopii mikrofilmowych, z wykorzystaniem sprzętu do digitalizacji ujętego jako środki trwałe AGH;
- OCR/HTR, w tym rozpoznawanie pisma ręcznego, oraz tłumaczenia maszynowe realizowane w ramach AGH Digital AI System, obejmujące moc obliczeniową, składowanie danych oraz opłaty za LLM;
- kontrola jakości, opracowanie techniczne plików oraz przygotowanie i wzbogacanie metadanych zgodnie ze standardem WIEDZ@;
- przechowywanie plików wynikowych w repozytorium AGH oraz techniczne przygotowanie zasobów do udostępnienia online.

Wysoki wolumen wskaźników, tj. 5 mln zdigitalizowanych obiektów oraz 50 TB danych, jest osiągalny przy tej strukturze budżetu, ponieważ digitalizacja dotyczy w przeważającej części materiałów już częściowo przygotowanych, w tym zmikrofilmowanych – 1 599 rolek. Ponadto proces zostanie oparty na zautomatyzowanym potoku OCR/HTR i Computer Vision w AGH Digital AI System, co istotnie obniża jednostkowy koszt przetworzenia strony lub obiektu. Koszt nie jest zatem zdominowany przez ręczne skanowanie, lecz przez specjalistyczne opracowanie oraz automatyczne przetwarzanie zasobów.

Poniżej przedstawiono zestawienie liczby polskich zbiorów Instytutu Hoovera, tak jak ujęto je w Studium Wykonalności. Są to wartości szacunkowe, określone na podstawie tabeli inwentarzowej i opisów grup zasobów.

Wartość zbiorcza przyjęta w projekcie wynosi ok. 5 mln obiektów liczonych jako pojedyncze strony dokumentów lub obrazy cyfrowe, tj. jednostki skanowania. Jest to wartość przyjęta dla wskaźników digitalizacji i udostępnienia, tj. KPI 11 i KPI 12.

Grupa zasobów	Szacunkowa liczba
---------------	-------------------

I. Archiwalia rządowe i dyplomatyczne	ok. 300–400 pudeł archiwalnych, tj. ok. 1,5 mln stron
II. Archiwalia wojskowe i zeznania deportowanych	ok. 200 pudeł; sama kolekcja Andersa: ok. 1,5 mln stron; 33 tys. zeznań deportowanych; 15 tys. certyfikatów zwolnienia z obozów NKWD, tj. ok. 200 tys. stron
III. Wydawnictwa i prasa podziemna	ponad 3 tys. periodyków, broszur i ulotek; łącznie kilkanaście tysięcy jednostek
IV. Materiały fotograficzne i ikonograficzne	kilkadziesiąt tysięcy obiektów, w tym fotografie, plakaty i pocztówki
V. Dokumenty życia społecznego i materiały audiowizualne	kilkaset nagrań oraz tysiące dokumentów prywatnych, tj. ok. 1 mln stron/obiektów

Powyższe zestawienie potwierdza, że założony wolumen digitalizacji nie ma charakteru deklaratywnego, lecz wynika z rozpoznanej struktury zasobów oraz przyjętej metody liczenia jednostek digitalizacyjnych. Jednocześnie należy podkreślić, że wskaźniki digitalizacji i udostępnienia dotyczą tego samego korpusu zasobów nowo digitalizowanych, natomiast zasoby już istniejące w postaci cyfrowej będą mierzone odrębnymi wskaźnikami dotyczącymi integracji, API, baz danych i ponownego wykorzystania informacji.

W konsekwencji niezasadne jest twierdzenie, że projekt jest w praktyce wyłącznie projektem budowy kosztownej platformy, a digitalizacja stanowi jedynie jego uzasadnienie. WIEDZ@ jest projektem cyfrowym, integracyjnym i usługowym, którego efektem ma być zarówno opracowanie i udostępnienie zasobów, jak również zapewnienie trwałych mechanizmów ich wyszukiwania, integracji, wzbogacania, udostępniania oraz ponownego wykorzystania. W działaniu FERC 2.3 wartość projektu nie wynika wyłącznie z liczby zeskanowanych obiektów, ale z zapewnienia cyfrowej dostępności i re-use informacji w sposób interoperacyjny, skalowalny i zgodny z potrzebami użytkowników.

Ad 3. Prosimy o uzupełnienie dokumentacji o jednoznaczne przypisanie jednostek pomiaru do poszczególnych KPI oraz wskazanie podstawowych zasad przeliczania jednostek operacyjnych na wartości wskaźników projektowych. W szczególności prosimy o doprecyzowanie, jak będą liczone wartości KPI dotyczące liczby zdigitalizowanych dokumentów lub obiektów oraz liczby udostępnionych dokumentów lub obiektów. Prosimy również o wskazanie, które jednostki będą miały charakter wyłącznie pomocniczy lub techniczny i nie będą wpływać na wartości wskaźników produktu lub rezultatu.

Beneficjent uzupełnia wyjaśnienia w zakresie metodologii pomiaru wskaźników poprzez jednoznaczne przypisanie jednostek pomiaru do poszczególnych KPI oraz wskazanie podstawowych zasad przeliczania jednostek operacyjnych na wartości wskaźników projektowych. Doprecyzowanie to zostanie odzwierciedlone w dokumentacji projektu, w szczególności w OZPI oraz w dokumentach powiązanych ze Studium Wykonalności.

Na potrzeby projektu przyjęto rozróżnienie pomiędzy:

- jednostkami wskaźnikowymi, które bezpośrednio wpływają na wartość KPI;
- jednostkami operacyjnymi, które służą do bieżącego monitorowania postępu prac;
- jednostkami technicznymi i pomocniczymi, które nie są samodzielnie agregowane do wartości wskaźników produktu lub rezultatu, chyba że metodologia danego KPI wyraźnie to przewiduje.

Wskaźniki projektowe nie będą liczone poprzez automatyczne sumowanie wszystkich jednostek występujących w procesie technicznym. Oznacza to, że strony dokumentów, klatki mikrofilmowe, rekordy metadanych, publikacje, tłumaczenia, pakiety ruchu sieciowego czy rekordy analityczne nie będą agregowane do jednego wspólnego wskaźnika. Każda kategoria będzie przypisana wyłącznie do właściwego KPI zgodnie z funkcją, jaką pełni w projekcie.

3.1. Przypisanie jednostek pomiaru do poszczególnych KPI

Nazwa wskaźnika	Jednostka pomiaru	Zasada liczenia
Instytucje publiczne otrzymujące wsparcie na opracowywanie usług, produktów i procesów cyfrowych	podmiot	Liczone są instytucje publiczne spełniające definicję wskaźnika; w projekcie: MNiSW jako partner projektu i bezpośredni interesariusz rezultatów platformy.
Liczba podmiotów wspartych w zakresie rozwoju usług, produktów i procesów cyfrowych	podmiot	Liczone są podmioty uczestniczące w projekcie i objęte wsparciem w zakresie rozwoju usług, produktów lub procesów cyfrowych.
Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne online	podmiot	Liczone są podmioty, które w wyniku projektu udostępnią dane online; wskaźnik nie obejmuje liczby systemów, baz, API ani przepływów danych.
Wartość usług, produktów i procesów cyfrowych opracowanych dla przedsiębiorstw	PLN	Wartość ustalana na podstawie kosztorysu projektu, bez wydatków na wynagrodzenia, szkolenia, informację/promocję oraz koszty pośrednie.
Liczba rozwiązań wykorzystujących informacje sektora publicznego/dane prywatne	rozwiązanie / e-usługa	Liczone są 3 wdrożone e-usługi wykorzystujące informacje sektora publicznego lub dane prywatne.
Użytkownicy nowych i zmodernizowanych publicznych usług, produktów i procesów cyfrowych	unikalny użytkownik	Liczba unikalnych użytkowników w okresie 12 miesięcy od zakończenia projektu; użytkownik liczony jednokrotnie w okresie pomiarowym.
Liczba pracowników IT objętych wsparciem szkoleniowym	osoba	Liczba pracowników IT objętych wsparciem szkoleniowym w ramach projektu.
Liczba pracowników niebędących pracownikami IT objętych wsparciem szkoleniowym	osoba	Liczba pracowników nie-IT objętych wsparciem szkoleniowym.
Liczba baz danych udostępnionych online poprzez API	baza danych	Liczone są bazy danych udostępnione online poprzez API, zgodnie z zakresem projektu.

Nazwa wskaźnika	Jednostka pomiaru	Zasada liczenia
Liczba platform udostępniania dokumentów zawierających informacje sektora publicznego/dane prywatne	platforma	Liczona jest Platforma WIEDZ@ jako jeden punkt dostępu do rozproszonych zasobów.
Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne	dokument / obiekt cyfrowy	Liczone są dokumenty lub obiekty źródłowe przetworzone z postaci analogowej lub mikrofilmowej do postaci cyfrowej, zgodnie z przyjętą jednostką obiektową.
Liczba udostępnionych online dokumentów zawierających informacje sektora publicznego/dane prywatne	dokument / obiekt cyfrowy	Liczone są zdigitalizowane dokumenty lub obiekty faktycznie udostępnione online użytkownikom końcowym.
Liczba uruchomionych systemów teleinformatycznych	system	Liczona jest Platforma WIEDZ@ jako uruchomiony system teleinformatyczny.
Liczba utworzonych API	API	Liczone są nowe interfejsy API utworzone, uruchomione, udokumentowane i gotowe do wykorzystania przez użytkowników lub systemy zewnętrzne.
Rozmiar udostępnionych online informacji sektora publicznego/danych prywatnych	TB	Liczony jest sumaryczny rozmiar plików i zasobów udostępnionych online w ramach platformy.
Rozmiar zdigitalizowanych informacji sektora publicznego/danych prywatnych	TB	Liczony jest sumaryczny rozmiar plików cyfrowych powstałych w wyniku digitalizacji dokumentów źródłowych z postaci analogowej lub mikrofilmowej.
Liczba przeprowadzonych badań satysfakcji użytkowników	Szt.	Wskaźnik własny, liczony jako liczba przeprowadzonych badań satysfakcji użytkowników oraz analizie opinii użytkowników dotyczących usług, produktów lub procesów cyfrowych realizowanych w ramach projektu. Badania zostaną przeprowadzone w formie ankietowej: przed wdrożeniem systemu oraz po wdrożeniu systemu. Ankieta obejmie ocenę m.in. łatwości obsługi, szybkości działania, funkcjonalności rozwiązania oraz ogólnego poziomu satysfakcji użytkowników.
Liczba wygenerowanych kluczy API	klucz API	Wartość docelowa 0; projekt nie przewiduje dedykowanych kluczy API jako odrębnego mechanizmu komunikacji backend-backend. Udostępnienie API będzie realizowane przez REST API i mechanizmy autoryzacji OAuth2/SCL.

3.2. Doprecyzowanie metodologii KPI 11 i KPI 12

W odniesieniu do KPI 11 i KPI 12, tj. wskaźników dotyczących liczby zdigitalizowanych oraz udostępnionych online dokumentów lub obiektów, przyjęto zasadę, że jednostką bazową jest pojedyncza strona dokumentu lub jeden obraz cyfrowy odpowiadający jednostce skanowania.

Oznacza to, że:

Jednostka operacyjna / techniczna	Reguła przeliczeniowa	Wpływ na KPI
Strona dokumentu, w tym akta i druki	1 strona = 1 obiekt cyfrowy	KPI 11 oraz KPI 12 po udostępnieniu
Klatka mikrofilmu	1 klatka = 1 obiekt cyfrowy / 1 plik wynikowy	KPI 11 oraz KPI 12 po udostępnieniu
Fotografia / plakat / pocztówka / obiekt ikonograficzny	1 obiekt = 1 obraz cyfrowy	KPI 11 oraz KPI 12 po udostępnieniu
Dokument lub jednostka archiwalna wielostronicowa	Liczona według liczby stron/skanów, a nie jako jeden obiekt zbiorczy	KPI 11 oraz KPI 12 po udostępnieniu
Rekord metadanych zasobu już cyfrowego	Nie przelicza się na obiekt zdigitalizowany	Nie wpływa na KPI 11/12; może być monitorowany w ramach KPI dotyczących API, baz danych, integracji lub ponownego wykorzystania
Publikacja, tłumaczenie, rekord analityczny, pakiet ruchu sieciowego	Jednostki techniczne lub procesowe	Nie są agregowane do KPI 11/12 ani do jednego wspólnego KPI
Rozmiar plików wynikowych	Sumaryczny rozmiar w TB	KPI 15 i KPI 16

W praktyce oznacza to, że w przypadku jednej książki, teczki aktowej, jednostki archiwalnej lub mikrofilmu do KPI 11 nie będzie zaliczany cały zbiór jako jeden obiekt, lecz poszczególne strony, skany, klatki lub obrazy cyfrowe. Taka metoda zapewnia porównywalność między różnymi kategoriami materiałów, w tym aktami, mikrofilmami, fotografiami, drukami, plakatami i dokumentami życia społecznego.

KPI 11 będzie obejmował dokumenty lub obiekty źródłowe przetworzone z postaci analogowej lub mikrofilmowej do postaci cyfrowej. KPI 12 będzie obejmował te spośród nich, które zostaną faktycznie udostępnione online użytkownikom końcowym za pośrednictwem Platformy WIEDZ@. Wartości docelowe obu wskaźników wynoszą po 5 000 000 szt., ponieważ odnoszą się do tego samego korpusu zasobów nowo digitalizowanych: najpierw obiekt zostaje zdigitalizowany, a następnie ten sam obiekt zostaje udostępniony online.

3.3. Uwzględnienie metodologii AGH dla zasobów Instytutu Hoovera

Założona wartość 5 000 000 obiektów wynika z rozpoznanej struktury zasobów polskich Instytutu Hoovera, ujętej w Studium Wykonalności oraz w materiałach AGH. Wartość ta obejmuje obiekty liczone jako pojedyncze strony dokumentów, klatki mikrofilmowe lub obrazy cyfrowe.

Zasoby te obejmują w szczególności:

Grupa zasobów	Szacunkowa liczba
I. Archiwalia rządowe i dyplomatyczne	ok. 300–400 pudeł archiwalnych, tj. ok. 1,5 mln stron
II. Archiwalia wojskowe i zeznania deportowanych	ok. 200 pudeł; sama kolekcja Andersa: ok. 1,5 mln stron; 33 tys. zeznań deportowanych; 15 tys. certyfikatów zwolnienia z obozów NKWD, tj. ok. 200 tys. stron
III. Wydawnictwa i prasa podziemna	ponad 3 tys. periodyków, broszur i ulotek; łącznie kilkanaście tysięcy jednostek
IV. Materiały fotograficzne i ikonograficzne	kilkadziesiąt tysięcy obiektów, w tym fotografie, plakaty i pocztówki
V. Dokumenty życia społecznego i materiały audiowizualne	kilkaset nagrań oraz tysiące dokumentów prywatnych, tj. ok. 1 mln stron/obiektów

Wartość zbiorcza przyjęta w projekcie, tj. ok. 5 mln obiektów, jest więc wartością szacunkową wynikającą z liczby stron, klatek i obrazów cyfrowych, a nie z liczby kolekcji, pudeł archiwalnych, publikacji lub zbiorczych jednostek archiwalnych. Wskaźnik nie będzie zatem sztucznie zawyżany przez liczenie tych samych zasobów w kilku kategoriach, ale oparty na jednolitej jednostce obiektowej.

3.4. Jednostki pomocnicze i techniczne niewpływające na wartości KPI produktu lub rezultatu

Dla uniknięcia wątpliwości beneficjent wskazuje, że następujące jednostki będą miały charakter pomocniczy lub techniczny i nie będą bezpośrednio wpływać na wartości KPI 11 i KPI 12:

- rekordy metadanych zasobów już istniejących w postaci cyfrowej;
- rekordy analityczne generowane w procesie przetwarzania danych;
- tłumaczenia maszynowe;
- pakiety ruchu sieciowego;
- publikacje rozumiane jako jednostki bibliograficzne, jeżeli są wielostronicowe;
- zbiory danych już istniejące w postaci cyfrowej;
- integracje systemowe;
- wywołania API;
- konta użytkowników, sesje, zapytania i logi techniczne.

Jednostki te mogą służyć do monitorowania postępu prac technicznych, jakości procesu, skali integracji albo wykorzystania usług, ale nie będą automatycznie przeliczane na liczbę zdigitalizowanych dokumentów lub obiektów. Jeżeli będą wpływać na inne KPI, zostanie to wskazane w metodologii danego wskaźnika, np. w odniesieniu do KPI 9, KPI 14, KPI 15 lub KPI 16.

3.5. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych

Beneficjent dodatkowo wskazuje, że zasoby już istniejące w postaci cyfrowej, które będą integrowane, indeksowane, agregowane lub ponownie udostępniane przez Platformę WIEDZ@, nie będą wliczane do KPI 11, ponieważ nie są nowo digitalizowane w ramach projektu.

Zasoby już cyfrowe będą ujmowane w innych wskaźnikach, właściwych dla ich charakteru, w szczególności:

- KPI 9 – liczba baz danych udostępnionych online poprzez API;
- KPI 10 – liczba platform udostępniania dokumentów;
- KPI 14 – liczba utworzonych API;
- KPI 15 – rozmiar udostępnionych online informacji, jeżeli zasoby będą udostępniane online w ramach platformy;
- KPI 5 – liczba rozwiązań wykorzystujących informacje sektora publicznego lub dane prywatne.

Takie podejście zapewnia rozdzielenie efektu nowej digitalizacji od efektu integracji i ponownego wykorzystania zasobów już cyfrowych oraz eliminuje ryzyko podwójnego liczenia.

Podsumowując, metodologia pomiaru wskaźników zostanie doprecyzowana w dokumentacji poprzez wskazanie jednostek bazowych, jednostek technicznych i jednostek pomocniczych dla poszczególnych KPI. Dla KPI 11 i KPI 12 podstawową jednostką będzie pojedyncza strona dokumentu, klatka mikrofilmu albo obraz cyfrowy. Jednostki takie jak rekord metadanych, publikacja, tłumaczenie, pakiet ruchu sieciowego, rekord analityczny lub integracja systemowa będą miały charakter pomocniczy lub techniczny i nie będą automatycznie wpływać na wartości wskaźników produktu lub rezultatu.

Ad 4. Prosimy o uzupełnienie wyjaśnień poprzez powiązanie KPI nr 3 z opisem przepływów danych i systemów źródłowych wskazanych w części 7 OZPI. W szczególności prosimy o wskazanie, które systemy i przepływy danych opisane w części 7 są przypisane do czterech podmiotów wykazywanych w KPI nr 3, które pozostają poza tym wskaźnikiem oraz z jakiego powodu. Prosimy również o wyjaśnienie, w jaki sposób monitorowana będzie faktyczna skala integracji źródeł danych, skoro KPI nr 3 obejmuje wyłącznie liczbę podmiotów bezpośrednio udostępniających dane online.

ODPOWIEDŹ:

Beneficjent podtrzymuje przyjętą metodologię kalkulacji KPI nr 3 „Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne online”. Zgodnie z definicją wskaźnika, jednostką miary jest podmiot, a nie system teleinformatyczny, baza danych, repozytorium, kolekcja, API, przepływ danych ani źródło danych integrowane w projekcie.

Zgodnie z definicją wskaźnika, do jego wartości wliczane są podmioty sektora publicznego lub prywatnego, które w wyniku interwencji udostępniły online informacje sektora publicznego lub dane prywatne. Oznacza to, że dla ustalenia wartości KPI nr 3 decydujące znaczenie ma liczba podmiotów udostępniających dane, a nie liczba systemów technicznych, za pośrednictwem których dane są integrowane, przetwarzane, indeksowane, udostępniane lub ponownie wykorzystywane.

W konsekwencji większa liczba przepływów danych, systemów źródłowych, API, baz danych lub repozytoriów wskazanych w części 7 OZPI nie oznacza automatycznie konieczności zwiększenia wartości KPI nr 3. Część 7 OZPI opisuje architekturę integracji i przepływy danych w projekcie, natomiast KPI nr 3 mierzy wyłącznie liczbę podmiotów, które udostępnią online informacje sektora publicznego lub dane prywatne w wyniku realizacji projektu.

Wartość docelowa KPI nr 3 została określona na poziomie 4, ponieważ w ramach projektu dane lub zasoby będą udostępniać następujące podmioty uczestniczące w partnerstwie:

Podmiot	Sposób powiązania z KPI nr 3
OPI PIB	Udostępnienie danych i metadanych za pośrednictwem Platformy WIEDZ@ oraz zapewnienie jednego punktu dostępu do zasobów integrowanych w projekcie.
AGH	Udostępnienie zasobów opracowywanych i digitalizowanych w ramach AGH Digital AI System, w szczególności zasobów Instytutu Hoovera / AGH.
PCSS	Udostępnienie zasobów, metadanych lub źródeł integrowanych w projekcie, w tym zasobów powiązanych z infrastrukturami i źródłami naukowymi.
Politechnika Poznańska	Udostępnienie własnych zasobów lub danych wnoszonych do projektu, w tym zasobów/źródeł wskazanych w części 7 OZPI jako przypisane do PP.

Jednocześnie należy podkreślić, że systemy i przepływy danych wskazane w części 7 OZPI pełnią różne funkcje i nie wszystkie stanowią odrębne udostępnienie danych przez nowy podmiot w rozumieniu KPI nr 3. W szczególności można wyróżnić trzy kategorie systemów i przepływów:

- systemy i zasoby przypisane do podmiotów wykazywanych w KPI nr 3 – są one uwzględniane w kalkulacji wskaźnika poprzez podmiot, który udostępnia dane;
- systemy referencyjne lub źródłowe – są integrowane w celu wzbogacenia metadanych, zapewnienia kontekstu, interoperacyjności lub poprawy wyszukiwalności, ale nie oznaczają automatycznie nowego udostępnienia danych przez odrębny podmiot;
- systemy techniczne i pomocnicze – służą m.in. uwierzytelnianiu, obsłudze zgłoszeń, integracji technicznej, komunikacji systemowej albo autoryzacji i nie wpływają na wartość KPI nr 3.

W tym ujęciu systemy takie jak NAVOICA, ZSUN II/RAD-on, Ludzie Nauki, POL-on, PBN, System Centralnego Logowania, Węzeł Krajowy, ORCID czy ZSUN HelpDesk nie są traktowane automatycznie jako odrębne podmioty zwiększające wartość KPI nr 3. Ich rola w projekcie może być źródłowa, referencyjna, integracyjna albo techniczna, ale nie każda integracja systemowa jest równoznaczna z udostępnieniem danych przez kolejny podmiot w rozumieniu definicji wskaźnika.

Analogicznie, jeżeli jeden podmiot udostępnia więcej niż jedną bazę danych, kolekcję, repozytorium lub API, nadal jest liczony w KPI nr 3 jako jeden podmiot. Definicja wskaźnika nie przewiduje mnożenia wartości KPI przez liczbę baz danych, systemów, kolekcji lub przepływów danych przypisanych do danego podmiotu. Takie elementy są monitorowane innymi wskaźnikami projektu, właściwymi dla skali integracji i udostępnienia technicznego.

W szczególności rzeczywista skala integracji źródeł danych będzie monitorowana poprzez inne KPI, w tym:

Zakres monitorowania	Właściwy KPI
Liczba podmiotów udostępniających dane online	KPI nr 3
Liczba rozwiązań wykorzystujących informacje sektora publicznego lub dane prywatne	KPI nr 5
Liczba baz danych udostępnionych online poprzez API	KPI nr 9
Liczba platform udostępniania dokumentów	KPI nr 10
Liczba utworzonych API	KPI nr 14
Rozmiar danych udostępnionych online	KPI nr 15

Tym samym KPI nr 3 nie służy do mierzenia całej złożoności architektury integracyjnej projektu. Jego funkcją jest pomiar liczby podmiotów, które w wyniku interwencji udostępnią online informacje sektora publicznego lub dane prywatne. Natomiast liczba systemów, baz danych, API, przepływów danych i zasobów integrowanych w projekcie jest mierzona i opisywana odrębnie, przez wskaźniki właściwe dla danego zakresu projektu oraz przez część 7 OZPI.

Beneficjent uzupełni dokumentację o jednoznaczne powiązanie KPI nr 3 z opisem przepływów danych w części 7 OZPI poprzez wskazanie, które systemy i przepływy danych są przypisane do podmiotów wykazywanych w KPI nr 3, a które pełnią funkcję referencyjną, integracyjną, techniczną lub pomocniczą i pozostają poza kalkulacją tego wskaźnika.

Podsumowując, przyjęta metodologia jest zgodna z definicją wskaźnika. Wartość KPI nr 3 została określona na podstawie liczby podmiotów udostępniających dane online, a nie na podstawie liczby systemów lub przepływów danych. Liczne integracje opisane w części 7 OZPI nie podważają wartości KPI nr 3, lecz pokazują architekturę techniczną projektu oraz skalę integracji, która jest monitorowana innymi wskaźnikami.

Ad 5. Przyjmujemy wyjaśnienia beneficjenta dotyczące kwalifikacji e-usług do poziomu dojrzałości „Transakcja”. Prosimy jedynie, aby na dalszym etapie prac opisy poszczególnych e-usług jednoznacznie wskazywały czynność inicjowaną przez użytkownika, sposób automatycznego przetworzenia żądania oraz rezultat zwracany użytkownikowi lub systemowi zewnętrznemu.

ODPOWIEDŹ:

Beneficjent przyjmuje uwagę Ministerstwa Cyfryzacji dotyczącą sposobu opisu e-usług. Skoro kwalifikacja e-usług do poziomu dojrzałości „Transakcja” została co do zasady przyjęta, na dalszym etapie prac opisy poszczególnych e-usług zostaną doprecyzowane tak, aby jednoznacznie wskazywały: czynność inicjowaną przez użytkownika, sposób automatycznego przetworzenia żądania oraz rezultat zwracany użytkownikowi lub systemowi zewnętrznemu.

Dla e-usługi dotyczącej dostępu do zasobów nauki i archiwów na podstawie zapytania czynnością użytkownika będzie złożenie zapytania, w tym zapytania w języku naturalnym. Automatyczne przetworzenie będzie obejmowało wyszukiwanie leksykalne i semantyczne w indeksie federacji oraz ranking wyników. Rezultatem będzie zwrócenie listy zasobów wraz z metadany oraz odesłaniem do źródła.

Dla e-usługi dotyczącej analizy zasobów cyfrowych czynnością użytkownika będzie wskazanie zasobu albo zbioru do analizy. Automatyczne przetworzenie będzie obejmowało przetwarzanie w wydzielonym środowisku, ocenę jakości oraz przygotowanie wyników lub wizualizacji. Rezultatem będzie zwrócenie użytkownikowi raportu albo wyników analizy.

Dla e-usługi dotyczącej pobrania danych i metadanych przez API czynnością użytkownika lub systemu zewnętrznego będzie wywołanie API z parametrami. Automatyczne przetworzenie będzie obejmowało złożenie i realizację żądania, kontrolę uprawnień oraz przygotowanie odpowiedzi. Rezultatem będzie zwrócenie danych lub metadanych do systemu zewnętrznego w celu ich ponownego wykorzystania.

Powyższe doprecyzowania zostaną odzwierciedlone w opisach e-usług w Studium wykonalności.

AD 6. Prosimy o uzupełnienie wyjaśnień o syntetyczne wskazanie, jaka część zasobów będzie nowo digitalizowana, jaka będzie jedynie integrowana lub ponownie udostępniana oraz jak każda z tych kategorii wpływa na poszczególne KPI. W szczególności prosimy o wyjaśnienie relacji między wskaźnikiem 5 000 000 zdigitalizowanych dokumentów lub obiektów a wskaźnikiem 5 000 000 dokumentów lub obiektów udostępnionych on-line, z uwzględnieniem zasobów już istniejących w postaci cyfrowej.

ODPOWIEDŹ:

Beneficjent uzupełnia wyjaśnienia dotyczące rozróżnienia zasobów nowo digitalizowanych oraz zasobów już istniejących w postaci cyfrowej. Projekt obejmuje dwie odrębne grupy zasobów.

Pierwszą grupę stanowią zasoby nowo digitalizowane w ramach projektu, w szczególności zbiory Instytutu Hoovera / AGH, obejmujące akta, mikrofilmy, fotografie, druki, dokumenty życia społecznego i inne materiały opracowywane w ramach AGH Digital AI System. Zasoby te będą ujmowane we wskaźniku dotyczącym liczby zdigitalizowanych dokumentów lub obiektów, a po publikacji również we wskaźniku dotyczącym liczby dokumentów lub obiektów udostępnionych online. Ich rozmiar będzie ujmowany we wskaźnikach dotyczących rozmiaru zdigitalizowanych i udostępnionych informacji.

Drugą grupę stanowią zasoby już istniejące w postaci cyfrowej, które będą integrowane, indeksowane, agregowane lub ponownie udostępniane za pośrednictwem Platformy WIEDZ@. Do tej grupy należą m.in. RAD-on, Ludzie Nauki, PBN, POL-on, NAVOICA, SSH Open Marketplace, GoTriple, EUNICE, Otwarte Dane Poznań oraz zbiory ruchu sieciowego. Zasoby te nie będą wliczane do wskaźników dotyczących liczby nowo zdigitalizowanych dokumentów lub obiektów ani liczby udostępnionych online dokumentów lub obiektów nowo zdigitalizowanych. Ich efekt będzie ujmowany w odrębnych wskaźnikach dotyczących baz danych udostępnionych przez API, liczby API, rozwiązań wykorzystujących dane oraz liczby podmiotów udostępniających dane online.

Relacja między wskaźnikiem 5 000 000 zdigitalizowanych dokumentów lub obiektów a wskaźnikiem 5 000 000 dokumentów lub obiektów udostępnionych online jest następująca: obie wartości odnoszą się do tego samego korpusu zasobów nowo digitalizowanych w ramach projektu. Obiekt najpierw zostaje zdigitalizowany, a następnie ten sam obiekt zostaje udostępniony online. Identyfikacja wartości obu wskaźników nie oznacza zatem, że cały wolumen zasobów dostępnych przez platformę odpowiada wyłącznie zasobom nowo digitalizowanym. Oznacza jedynie, że wskaźniki digitalizacji i udostępnienia zostały przyjęte dla tego samego, wyodrębnionego korpusu zasobów nowo digitalizowanych.

Zasoby już cyfrowe, które będą integrowane, indeksowane i ponownie udostępniane, nie będą liczone w tych dwóch wskaźnikach, aby uniknąć podwójnego liczenia. Ich efekt będzie mierzony wskaźnikami dotyczącymi integracji, API, baz danych, rozwiązań wykorzystujących dane oraz liczby podmiotów. Takie rozróżnienie zapewnia przejrzystość metodologiczną i pozwala oddzielić efekt nowej digitalizacji od efektu integracji i ponownego wykorzystania zasobów już cyfrowych.

LP	Nazwa zestawu danych	Jednostka liczenia obiektów	Liczba obiektów	Czy zestaw digitalizowany w projekcie	Czy zestaw udostępniany w projekcie
	Bazy ofert technologicznych (TRL), usług badawczych, infrastruktury (IQ/OQ/PQ) oraz własności intelektualnej w Portalu Innowacji.	rekord	1 000	NIE	TAK
	Dokumenty aktowe Ministerstwa Spraw Zagranicznych RP oraz polskich placówek dyplomatycznych przechowywane w Hoover Institution, obejmujące m.in. dokumentację ambasad i konsulatów. Obiektem jest pojedyncza strona dokumentu lub jeden obraz cyfrowy odpowiadający jednostce skanowania.	obiekt	1 800 000	TAK	TAK
	Zdigitalizowane i udostępnione materiały Ministerstwa Informacji i Dokumentacji z lat 1939 –1945, obejmujące dokumenty administracyjne, materiały informacyjne i dokumentację działalności rządu emigracyjnego. Przyjęto, że obiektem jest jedna strona dokumentu po wykonaniu digitalizacji i podstawowej obróbki obrazu.	sztuka	650 000	TAK	TAK
	Mikrofilmy historyczne, obejmujące wcześniej wykonane odwzorowania zbiorów polskich. Obiektem jest jedna klatka mikrofilmowa po przekształceniu do postaci obrazu cyfrowego oraz przypisaniu podstawowych metadanych technicznych i opisowych.	obiekt	1 200 000	NIE	TAK
	Udostępniony wykaz realizowanych i zrealizowanych w Polsce projektów naukowych zawierające tytuły, streszczenia, okres realizacji, dyscyplinę, jednostkę realizującą, źródło finansowania i wysokość przyznanych środków.	rekord metadanych projektu	9 500	NIE	TAK
	Wykaz patentów i wzorów użytkowych zawierający m.in. tytuł, opis patentu, nr patentu, datę udzielenia, klasyfikację IPC, dane autorów, afiliacje autorów.	rekord metadanych patentu	12 500	NIE	TAK
	Udostępnione referencyjne zasoby informacyjne systemu RAD-on (m.in. instytucje, publikacje, patenty, projekty naukowe, wyniki ewaluacji).	terabajt danych	14,7	NIE	TAK
	Ustrukturyzowane profile naukowców oraz metadane dorobku naukowego z portalu Ludzie Nauki (w ramach budowy interfejsu programistycznego API).	profil naukowca	218 000	NIE	TAK
	Wykaz publikacji (artykuły, monografie, książki pod redakcją, rozdziały w książkach) raportowanych przez instytucja sektora nauki od 2015 roku jako dorobek naukowy podlegający ewaluacji.	Rekord metadanych publikacji	3 200 000	NIE	TAK
	Udostępniony wykaz zbiorów danych, publikacji i workflows, narzędzi, usług i materiałów szkoleniowych, dla nauk humanistycznych i społecznych (SSH Open Marketplace).	sztuki	5 000	NIE	TAK
	Metadane publikacji, zbiorów danych, profili badaczy i opisów projektów z obszaru nauk społecznych i humanistycznych, dostępne za pośrednictwem platformy GoTriple. Zasób obejmuje metadane agregowane przez GoTriple z wielu źródeł, m.in. Isidore (Integration de Services, Interconnexion de DONnees de la Recherche et de l'Enseignement), OpenAIRE (Open Access Infrastructure for Research in Europe), BASE – Bielefeld Academic Search Engine, DOAJ (Directory of Open Access Journals) oraz EconStor (repozytorium dyscyplinarne dla ekonomii i studiów biznesowych). W projekcie WIEDZ@ bezpośrednim punktem integracji będzie GoTriple, z którego metadane będą pobierane przez REST API / OAI-PMH. Nie przewiduje się bezpośrednich przepływów integracyjnych z każdym ze źródeł pierwotnych zasilających GoTriple.	sztuka	22 000 000	NIE	TAK
	Metadane recenzowanych książek od ponad 600 wydawców (Directory of Open Access Books)	sztuka	103 500	NIE	TAK
	Pełne teksty monografii w otwartym dostępie z dziedziny nauk humanistycznych i społecznych.	sztuka	24 000	NIE	TAK
	Wykaz czasopism, monografii dostępnych w otwartym dostępie, wydarzeń i blogów naukowych.	sztuka	600	NIE	TAK
	Udostępnione informacje o publikacjach oraz projektach realizowanych w ramach Uczelni Partnerskich konsorcjum European University for Customised Educatio (EUNICE).	sztuka	280 000	NIE	TAK

LP	Nazwa zestawu danych	Jednostka liczenia obiektów	Liczba obiektów	Czy zestaw digitalizowany w projekcie	Czy zestaw udostępniany w projekcie
	Dane zarejestrowanego ruchu sieciowego generowanego przez urządzenia IoT w obszarze technologii miejskich i obronnych rejestrowanych przez środowiska testowe Politechniki Poznańskiej.	pakiet danych	500 000 000	NIE	TAK
	Dane badawcze opisujące rzeczywistość społecznoekonomiczną Poznania, rejestrowanych przez służby miejskie Poznania.	zbiór danych tematycznych	34	NIE	TAK
	Metadane kursów edukacyjnych MOOC, szkoleń oraz sylabusów z platformy edukacyjnej NAVOICA.	rekord	1000		
	Kolekcje prywatne osób życia publicznego i kulturalnego, w tym m.in. Jana Karskiego, Ignacego J. Paderewskiego, Tadeusza Bora-Komorowskiego, Stefana Korbońskiego oraz Augusta Zaleskiego. Na potrzeby przeliczenia przyjęto, że obiektem jest pojedyncza karta korespondencji, notatki, rękopisu, fotografii lub dokumentu osobistego.	sztuka	350 000	TAK	TAK
	Druki podziemne, materiały z okresu II wojny światowej, Powstania Warszawskiego, pierwszych lat powojennych oraz epoki „Solidarności”, przekształcone do postaci umożliwiającej ich przeszukiwanie. Obiektem podlegającym digitalizacji jest pojedyncza strona publikacji, ulotki, biuletynu lub gazety.	sztuka	900 000	TAK	TAK
	Wydawnictwa emigracyjne oraz materiały propagandowe, w tym druki przygotowane na konferencję wersalską oraz publikacje z lat 1939 – 1989. Przyjęto, że obiektem jest pojedyncza strona publikacji lub jedno odwzorowanie obiektu	sztuka	450 000	TAK	TAK
	Baza metadanych opisowych dla dokumentów i jednostek archiwalnych, obejmująca opisy zespołów, serii, jednostek oraz dokumentów, wzbogacona o słowa kluczowe, identyfikatory, daty, nazwy geograficzne i osoby.	rekord	5 000 000	NIE	TAK
	Tłumaczenia dokumentów na język polski i angielski dla zasobów objętych projektem. Obiektem jest pojedynczy rekord tłumaczenia przypisany do jednej strony dokumentu lub jednego logicznego segmentu tekstu.	obiekt	3 000 000	NIE	TAK
	Zbiory danych do ponownego wykorzystania, obejmujące wyodrębnione metadane, indeksy pełnotekstowe, słowniki bytów oraz rekordy analityczne wspierające wyszukiwanie semantyczne. Obiektem jest pojedynczy rekord danych udostępniany poprzez interfejs systemowy lub eksport danych.	obiekt	5 000 000	NIE	TAK

KORESPONDECJA OPI PIB – MC za pośrednictwem KdSC nr 3

UWAGI MC NR 3 z 23.06.2026

Ad 1. Komplementarność projektu WIEDZ@ względem KRONIK@ 2.0

Ministerstwo Cyfryzacji przyjmuje wyjaśnienia beneficjenta w zakresie komplementarności projektu WIEDZ@ względem KRONIK@ 2.0. Prosimy o konsekwentne odzwierciedlenie tych wyjaśnień w dokumentacji projektu, w szczególności poprzez jednoznaczne wskazanie, że w ramach projektu WIEDZ@ nie będą finansowane rozwiązania informatyczne analogiczne lub podobne do komponentów wytwarzanych lub rozwijanych w ramach KRONIK@ 2.0.

Ad 2. Struktura kosztów digitalizacji i udostępnienia zasobów

Ministerstwo Cyfryzacji przyjmuje wyjaśnienia beneficjenta dotyczące struktury kosztów projektu.

Ad 3. Metodologia pomiaru KPI dotyczących digitalizacji i udostępnienia

Ministerstwo Cyfryzacji przyjmuje wyjaśnienia beneficjenta dotyczące metodologii pomiaru KPI, w szczególności w zakresie rozróżnienia zasobów nowo digitalizowanych od zasobów już istniejących w postaci cyfrowej, które będą integrowane, indeksowane, agregowane lub ponownie udostępniane przez Platformę WIEDZ@, z zastrzeżeniem uwag zgłoszonych w Ad 6.

Ad 4. KPI nr 3 i przepływy danych

Ministerstwo Cyfryzacji przyjmuje wyjaśnienia beneficjenta dotyczące KPI nr 3, zgodnie z którymi wskaźnik ten mierzy liczbę podmiotów udostępniających dane online, a nie liczbę systemów, baz danych, API, kolekcji lub przepływów danych.

Ad 5.

Wcześniej już zaakceptowaliśmy odpowiedź.

Ad 6. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych

Ministerstwo Cyfryzacji podtrzymuje zastrzeżenie dotyczące sposobu ujęcia wskaźnika odnoszącego się do 5 000 000 zdigitalizowanych dokumentów lub obiektów.

Z przedstawionych wyjaśnień oraz tabeli wynika, że liczby obiektów oznaczonych jako digitalizowane w projekcie dają sumę 4 150 000, a nie 5 000 000. Prosimy zatem o wyjaśnienie tej rozbieżności albo odpowiednią korektę wartości wskaźnika.

Dodatkowo, po przedstawionych wyjaśnieniach zasadne jest doprecyzowanie samego opisu KPI. Obecne sformułowanie odnoszące się do liczby zdigitalizowanych dokumentów lub obiektów może wprowadzać w błąd, ponieważ sugeruje, że w projekcie zostanie zdigitalizowanych 5 000 000 odrębnych obiektów. Z wyjaśnień wynika natomiast, że w rzeczywistości chodzi o 5 000 000 stron różnych publikacji, klatek mikrofilmowych, pojedynczych kart lub innych jednostek technicznych digitalizacji.

Taka metodologia liczenia nie powinna być opisywana jako liczba obiektów. Przykładowo pojedyncza karta korespondencji nie może być uznawana za odrębny obiekt w sytuacji, gdy kilka kart składa się na jeden list. Przyjęcie przeciwnego założenia prowadziłoby do wniosku, że jeden list jest kilkoma obiektami jednocześnie. Analogiczny problem dotyczy gazet, druków i innych wydawnictw, w których liczba stron nie jest tożsama z liczbą odrębnych egzemplarzy, tytułów, dokumentów lub obiektów.

W związku z powyższym prosimy o zmianę KPI na liczbę zdigitalizowanych stron, klatek mikrofilmowych, pojedynczych kart lub innych jednoznacznie zdefiniowanych jednostek technicznych digitalizacji. Alternatywnie prosimy o podanie liczby obiektów zgodnej ze stanem faktycznym, tj. odnoszącej się do liczby pojedynczych egzemplarzy gazet, druków, listów, dokumentów i innych wydawnictw, a nie do ich łącznej liczby stron lub kart.

ODPOWIEDŹ OPI PIB NR 3 z 23.06.2026

Ad 6. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych oraz wartość wskaźnika dotyczącego liczby zdigitalizowanych dokumentów

Beneficjent dziękuje za ponowne zwrócenie uwagi na sposób prezentacji danych dotyczących wolumenu zasobów przewidzianych do digitalizacji oraz na potrzebę doprecyzowania metodologii liczenia wskaźnika dotyczącego liczby zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne.

Po ponownej weryfikacji zestawienia potwierdzamy, że w przekazanej wcześniej tabeli wystąpiła omyłka pisarska polegająca na nieuwzględnieniu wśród zasobów przeznaczonych do digitalizacji części materiałów mikrofilmowych. W konsekwencji przedstawiona wcześniej suma nie odzwierciedlała pełnej populacji zasobów objętych procesem digitalizacji.

Beneficjent przekazuje zaktualizowaną tabelę, w której ujęto również materiały mikrofilmowe przeznaczone do digitalizacji. Po uwzględnieniu tej kategorii szacowana populacja zasobów objętych digitalizacją wynosi **5 350 000 szt.**, a zatem przekracza wartość docelową wskaźnika określoną na poziomie **5 000 000 szt.**

W związku z powyższym beneficjent podtrzymuje wartość docelową KPI na poziomie **5 000 000 szt.** Wartość ta została określona ostrożnościowo, z uwzględnieniem charakteru zasobów archiwalnych, mikrofilmowych, ikonograficznych i dokumentacyjnych oraz faktu, że ostateczna liczba jednostek digitalizacji zostanie potwierdzona dopiero w toku realizacji prac, w szczególności po digitalizacji, opracowaniu, kontroli jakości i odbiorze zasobów.

Jednocześnie beneficjent doprecyzuje w dokumentacji sposób opisu jednostki wskaźnika. W miejsce sformułowań mogących sugerować liczbę „obiektów” w znaczeniu inwentarzowym, muzealnym, bibliotecznym lub kolekcjonerskim, stosowane będzie określenie zgodne z jednostką miary wskaźnika, tj. „**sztuka**”, oraz opis odnoszący się do **dokumentów lub części dokumentów** w rozumieniu definicji wskaźnika programowego.

Beneficjent podkreśla, że przyjęta metodologia jest zgodna z definicją wskaźnika „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne”. Zgodnie z tą definicją dokument oznacza dowolną treść, niezależnie od jej nośnika, albo dowolną część tej treści. Definicja ta wprost obejmuje zatem zarówno całe dokumenty, jak i części treści, jeżeli to one są jednostkami faktycznie przetwarzanymi z postaci analogowej lub mikrofilmowej do postaci cyfrowej.

W przypadku zasobów objętych projektem jednostkami wykazywanymi we wskaźniku będą więc dokumenty lub części dokumentów przetworzone z postaci analogowej lub mikrofilmowej do postaci cyfrowej, w szczególności strony, karty, klatki mikrofilmowe, fotografie, plakaty, pocztówki oraz inne jednoznacznie identyfikowalne jednostki treści. Takie podejście odpowiada definicji wskaźnika programowego i specyfice zasobów archiwalnych, mikrofilmowych, ikonograficznych oraz dokumentacyjnych.

Nie oznacza to natomiast liczenia technicznych kopii tego samego dokumentu ani produktów pośrednich procesu digitalizacji. Do wartości wskaźnika nie będą wliczane duplikaty techniczne, miniatury, wersje robocze, rekordy metadanych, tłumaczenia maszynowe, indeksy pełnotekstowe ani inne warstwy pomocnicze, o ile nie stanowią one odrębnego dokumentu lub części dokumentu w rozumieniu definicji wskaźnika.

Odnosząc się do propozycji Ministerstwa Cyfryzacji dotyczącej zmiany KPI na liczbę zdigitalizowanych stron, klatek mikrofilmowych, pojedynczych kart lub innych jednostek technicznych digitalizacji, beneficjent wskazuje, że nie przyjmuje tej propozycji jako zmiany wskaźnika. Wskaźnik „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” jest wskaźnikiem programowym produktu, wynikającym z dokumentacji konkursowej, a nie wskaźnikiem własnym zdefiniowanym przez beneficjenta na potrzeby OZPI. Beneficjent nie powinien zatem zastępować go innym wskaźnikiem opartym wyłącznie na technicznych jednostkach digitalizacji.

Jednocześnie beneficjent doprecyzuje metodologię liczenia wskaźnika, wskazując, że strony, karty, klatki mikrofilmowe, fotografie, plakaty, pocztówki i inne jednoznacznie identyfikowalne jednostki treści są uwzględniane nie jako odrębny, nowy typ KPI, lecz jako dokumenty lub części dokumentów w rozumieniu definicji wskaźnika programowego. Tak szerokie ujęcie definicyjne pozwala na stosowanie wskaźnika do zasobów archiwalnych, mikrofilmowych, ikonograficznych i dokumentacyjnych, w których jednostką faktycznej digitalizacji jest często strona, karta, klatka mikrofilmu lub pojedynczy obraz cyfrowy.

Beneficjent nie przyjmuje również alternatywy polegającej na podaniu liczby obiektów rozumianych wyłącznie jako pojedyncze egzemplarze gazet, druków, listów, dokumentów lub wydawnictw. Takie podejście zawęziłoby pojęcie „dokumentu” wyłącznie do całych egzemplarzy lub jednostek inwentarzowych, pomijając wyrażnie wskazaną w definicji możliwość liczenia także części treści. W przypadku zasobów archiwalnych i mikrofilmowych prowadziłoby to do nieadekwatnego przedstawienia rzeczywistego zakresu digitalizacji oraz do oderwania metodologii od definicji wskaźnika przyjętej w dokumentacji konkursowej.

W związku z powyższym beneficjent nie zmienia KPI na wskaźnik liczby stron, kart lub klatek mikrofilmowych oraz nie ogranicza go do liczby zbiorczych obiektów inwentarzowych. Beneficjent podtrzymuje stosowanie wskaźnika programowego w brzmieniu wynikającym z dokumentacji konkursowej, przy jednoczesnym doprecyzowaniu, że wartość **5 000 000 szt.** odnosi się do liczby dokumentów lub części dokumentów w rozumieniu definicji wskaźnika, a nie do liczby technicznych plików ani do liczby zbiorczych jednostek inwentarzowych.

Należy również wskazać, że na etapie OZPI i wniosku o dofinansowanie wartości wskaźników muszą być prezentowane jako konkretne wartości docelowe, a nie jako wartości przybliżone. Przy tak znacznym wolumenie zasobów precyzyjne określenie liczby jednostek digitalizacji co do jednej strony, karty, klatki mikrofilmowej lub obrazu cyfrowego na etapie przygotowania OZPI nie jest możliwe. Z tego względu wartość 5 000 000 szt. została przyjęta jako wartość docelowa oszacowana na podstawie rozpoznanego wolumenu zasobów, danych inwentarzowych oraz planowanego zakresu digitalizacji.

Osiągnięcie wskaźnika będzie monitorowane i rozliczane po stronie beneficjenta w toku realizacji projektu, na podstawie dokumentacji digitalizacji, protokołów odbioru, raportów partnera odpowiedzialnego za digitalizację oraz danych z systemu zarządzania zasobami. Beneficjent ma świadomość, że wykonanie wskaźnika będzie przedmiotem rozliczenia projektu, dlatego wartość docelowa została określona ostrożnościowo — poniżej szacowanej populacji zasobów przeznaczonych do digitalizacji, która po aktualizacji tabeli wynosi 5 350 000 szt.

W związku z powyższym, informujemy, że wcześniejsza rozbieżność wynikała z omyłki w tabelarycznym ujęciu zasobów, a nie z błędnego założenia metodologicznego. Po aktualizacji tabeli populacja zasobów przewidzianych do digitalizacji wynosi **5 350 000 szt.**, czyli więcej niż wartość docelowa KPI. Beneficjent podtrzymuje zatem wartość wskaźnika na poziomie **5 000 000 szt.**, przy jednoczesnym doprecyzowaniu, że chodzi o dokumenty lub części dokumentów w rozumieniu definicji wskaźnika programowego, a nie o obiekty w znaczeniu inwentarzowym.

Zaktualizowana tabela:

LP	Nazwa zestawu danych	Jednostka liczenia obiektów	Liczba obiektów	Czy zestaw digitalizowany w projekcie	Czy zestaw udostępniany w projekcie
1.	Bazy ofert technologicznych (TRL), usług badawczych, infrastruktury (IQ/OQ/PQ) oraz własności intelektualnej w Portalu Innowacji.	rekord	1 000	NIE	TAK
2.	Dokumenty aktowe Ministerstwa Spraw Zagranicznych RP oraz polskich placówek dyplomatycznych przechowywane w Hoover Institution, obejmujące m.in. dokumentację ambasad i konsulatów. Sztuką jest pojedyncza strona dokumentu lub jeden obraz cyfrowy odpowiadający jednostce skanowania.	sztuka	1 800 000	TAK	TAK
3.	Zdigitalizowane i udostępnione materiały Ministerstwa Informacji i Dokumentacji z lat 1939–1945, obejmujące dokumenty administracyjne, materiały informacyjne i dokumentację działalności rządu emigracyjnego. Przyjęto, że sztuką jest jedna strona dokumentu po wykonaniu digitalizacji i podstawowej obróbki obrazu.	sztuka	650 000	TAK	TAK
4.	Mikrofilmy historyczne, obejmujące wcześniej wykonane odwzorowania zbiorów polskich. Sztuką jest jedna klatka mikrofilmowa po przekształceniu do postaci obrazu cyfrowego oraz przypisaniu podstawowych metadanych technicznych i opisowych.	sztuka	1 200 000	TAK	TAK
5.	Udostępniony wykaz realizowanych i zrealizowanych w Polsce projektów naukowych zawierające tytuły, streszczenia, okres realizacji, dyscyplinę, jednostkę realizującą, źródło finansowania i wysokość przyznanych środków.	rekord metadanych projektu	9 500	NIE	TAK
6.	Wykaz patentów i wzorów użytkowych zawierający m.in. tytuł, opis patentu, nr patentu, datę udzielenia, klasyfikację IPC, dane autorów, afiliacje autorów.	rekord metadanych patentu	12 500	NIE	TAK
7.	Udostępnione referencyjne zasoby informacyjne systemu RAD-on (m.in. instytucje, publikacje, patenty, projekty naukowe, wyniki ewaluacji).	terabajt danych	14,7	NIE	TAK
8.	Ustrukturyzowane profile naukowców oraz metadane dorobku naukowego z portalu Ludzie Nauki (w ramach budowy interfejsu programistycznego API).	profil naukowca	218 000	NIE	TAK

LP	Nazwa zestawu danych	Jednostka liczenia obiektów	Liczba obiektów	Czy zestaw digitalizowany w projekcie	Czy zestaw udostępniany w projekcie
9.	Wykaz publikacji (artykuły, monografie, książki pod redakcją, rozdziały w książkach) raportowanych przez instytucja sektora nauki od 2015 roku jako dorobek naukowy podlegający ewaluacji.	rekord metadanych publikacji	3 200 000	NIE	TAK
10.	Udostępniony wykaz zbiorów danych, publikacji i workflows, narzędzi, usług i materiałów szkoleniowych, dla nauk humanistycznych i społecznych (SSH Open Marketplace).	sztuki	5 000	NIE	TAK
11.	Metadane publikacji, zbiorów danych, profili badaczy i opisów projektów z obszaru nauk społecznych i humanistycznych, dostępne za pośrednictwem platformy GoTriple. Zasób obejmuje metadane agregowane przez GoTriple z wielu źródeł, m.in. Isidore (Integration de Services, Interconnexion de DONnees de la Recherche et de l'Enseignement), OpenAIRE (Open Access Infrastructure for Research in Europe), BASE – Bielefeld Academic Search Engine, DOAJ (Directory of Open Access Journals) oraz EconStor (repozytorium dyscyplinarne dla ekonomii i studiów biznesowych). W projekcie WIEDZ@ bezpośrednim punktem integracji będzie GoTriple, z którego metadane będą pobierane przez REST API / OAI-PMH. Nie przewiduje się bezpośrednich przepływów integracyjnych z każdym ze źródeł pierwotnych zasilających GoTriple.	sztuka	22 000 000	NIE	TAK
12.	Metadane recenzowanych książek od ponad 600 wydawców (Directory of Open Access Books)	sztuka	103 500	NIE	TAK
13.	Pełne teksty monografii w otwartym dostępie z dziedzin nauk humanistycznych i społecznych.	sztuka	24 000	NIE	TAK
14.	Wykaz czasopism, monografii dostępnych w otwartym dostępie, wydarzeń i blogów naukowych.	sztuka	600	NIE	TAK
15.	Udostępnione informacje o publikacjach oraz projektach realizowanych w ramach Uczelni Partnerskich konsorcjum European University for Customised Education (EUNICE).	sztuka	280 000	NIE	TAK
16.	Dane zarejestrowanego ruchu sieciowego generowanego przez urządzenia IoT w obszarze technologii miejskich i obronnych rejestrowanych przez środowiska testowe Politechniki Poznańskiej.	pakiet danych	500 000 000	NIE	TAK
17.	Dane badawcze opisujące rzeczywistość społecznoekonomiczną Poznania, rejestrowanych przez służby miejskie Poznania.	zbiór danych tematycznych	34	NIE	TAK
18.	Metadane kursów edukacyjnych MOOC, szkoleń oraz sylabusów z platformy edukacyjnej NAVOICA.	rekord	1000		
19.	Kolekcje prywatne osób życia publicznego i kulturalnego, w tym m.in. Jana Karskiego, Ignacego J. Paderewskiego, Tadeusza Bora-Komorowskiego, Stefana Korbońskiego oraz Augusta Zaleskiego. Na potrzeby przeliczenia przyjęto, że sztuką jest pojedyncza karta korespondencji, notatki, rękopisu, fotografii lub dokumentu osobistego.	sztuka	350 000	TAK	TAK
20.	Druki podziemne, materiały z okresu II wojny światowej, Powstania Warszawskiego, pierwszych lat powojennych oraz epoki „Solidarności”, przekształcone do postaci umożliwiającej ich przeszukiwanie. Sztuką podlegającym digitalizacji jest pojedyncza strona publikacji, ulotki, biuletynu lub gazety.	sztuka	900 000	TAK	TAK
21.	Wydawnictwa emigracyjne oraz materiały propagandowe, w tym druki przygotowane na konferencję wersalską oraz publikacje z lat 1939 – 1989. Przyjęto, że sztuką jest pojedyncza strona publikacji lub jedno odwzorowanie obiektu	sztuka	450 000	TAK	TAK

LP	Nazwa zestawu danych	Jednostka liczenia obiektów	Liczba obiektów	Czy zestaw digitalizowany w projekcie	Czy zestaw udostępniany w projekcie
22.	Baza metadanych opisowych dla dokumentów i jednostek archiwalnych, obejmująca opisy zespołów, serii, jednostek oraz dokumentów, wzbogacona o słowa kluczowe, identyfikatory, daty, nazwy geograficzne i osoby.	rekord	5 000 000	NIE	TAK
23.	Tumaczenia dokumentów na język polski i angielski dla zasobów objętych projektem. Sztuką jest pojedynczy rekord tłumaczenia przypisany do jednej strony dokumentu lub jednego logicznego segmentu tekstu.	sztuka	3 000 000	NIE	TAK
24.	Zbiory danych do ponownego wykorzystania, obejmujące wyodrębnione metadane, indeksy pełnotekstowe, słowniki bytów oraz rekordy analityczne wspierające wyszukiwanie semantyczne. Sztuką jest pojedynczy rekord danych udostępniany poprzez interfejs systemowy lub eksport danych.	sztuka	5 000 000	NIE	TAK

KORESPONDECJA OPI PIB – MC za pośrednictwem KdSC nr 4

UWAGI MC NR 4 z 25.06.2026

Ad 6. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych oraz wartość wskaźnika dotyczącego liczby zdigitalizowanych dokumentów

Ministerstwo Cyfryzacji podtrzymuje uwagę dotyczącą sposobu opisu i metodologii liczenia KPI odnoszącego się do liczby zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne.

Ministerstwo Cyfryzacji nie kwestionuje, że wskaźnik „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” jest wskaźnikiem programowym produktu wynikającym z dokumentacji FERC. Nie oczekujemy zatem zastąpienia go innym wskaźnikiem własnym. Istotne jest natomiast, aby opis KPI w OZPI i we wniosku o dofinansowanie jednoznacznie wskazywał, co faktycznie będzie liczone w projekcie.

Najistotniejsze jest to, że metodologia zaproponowana przez beneficjenta nie znajduje potwierdzenia w dokumentacji właściwej dla tego naboru. W „Instrukcji wypełniania wniosku o dofinansowanie” dla Priorytetu FERC.02, Działania FERC.02.03, typu projektu „Cyfrowe udostępnienie zasobów nauki”, naboru nr FERC.02.03-IP.01-005/26, w sekcji „Lista mierzalnych wskaźników projektu”, w części „Lista programowych (dodatkowych) wskaźników produktu w naborze z definicjami”, przy wskaźniku „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” wskazano:

„Nazwa wskaźnika: Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne.

Jednostka miary: szt.

Typ wskaźnika: wskaźnik programowy produktu.

Definicja: Liczba dokumentów stanowiących informację sektora publicznego/dane prywatne, które w wyniku wsparcia, zostaną przetworzone z postaci analogowej do postaci cyfrowej.

"Dokument" oznacza: dowolną treść, niezależnie od jej nośnika (papier lub forma elektroniczna lub zapis dźwiękowy, wizualny lub audiowizualny); lub dowolną część tej treści.

Podstawa prawna: art. 2, Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1024 dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego (Dz. Urz. UE L 172 z 26.06.2019).

Jeden utwór, w rozumieniu ustawy o prawie autorskim i prawach pokrewnych (Dz.U.2022.2509), powinien być zliczany jako jeden dokument, niezależnie od liczby części (plików), w których został zdigitalizowany.”

Powyższe oznacza, że powołanie się przez beneficjenta na programowy charakter KPI nie usuwa wątpliwości. Przeciwnie, bezpośrednio z instrukcji wypełniania wniosku dla tego naboru wynika, że jeden utwór powinien być zliczany jako jeden dokument niezależnie od liczby części lub plików, w których został zdigitalizowany. Jeżeli zatem jeden dokument, publikacja, jednostka archiwalna albo inny utwór został zdigitalizowany w wielu częściach, plikach lub skanach, nie powinien być liczony jako tyle dokumentów, ile wykonano skanów, plików, stron, kart lub klatek mikrofilmowych.

Z odpowiedzi beneficjenta wynika natomiast, że wartość 5 000 000 szt. nie odnosi się do 5 000 000 dokumentów lub obiektów źródłowych, lecz do 5 000 000 technicznych jednostek digitalizacji, tj. pojedynczych skanów, obrazów cyfrowych, stron, kart lub klatek mikrofilmowych. Taki sposób liczenia może zostać opisany w dokumentacji projektu, ale nie powinien być przedstawiany jako liczba zdigitalizowanych dokumentów lub obiektów źródłowych. Dlatego opis KPI powinien zostać zmieniony w taki sposób, aby wprost wskazywał, że wartość docelowa 5 000 000 dotyczy skanów / obrazów cyfrowych / stron / kart / klatek mikrofilmowych, a nie 5 000 000 dokumentów, publikacji, jednostek archiwalnych lub innych obiektów źródłowych.

Należy przypomnieć, że we wcześniejszych wyjaśnieniach beneficjent wskazywał, iż zasoby lub metadane powstające w ramach WIEDZ@ będą mogły zasilać także inne systemy, w tym KRONIK@, oraz pozostawać z nimi interoperacyjne dzięki standardom metadanych, API, identyfikatorom, linkowaniu i mechanizmom wymiany danych. Skoro zatem projekt WIEDZ@ ma pozostawać interoperacyjny względem KRONIK@, opis zasobów i metodologia liczenia obiektów nie powinny abstrahować od logiki standardu metadanych KRONIK@.

Należy również zauważyć, że Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, jako podmiot uczestniczący w projekcie WIEDZ@, brał udział w opiniowaniu standardu metadanych KRONIK@. Standard ten został przed publikacją pozytywnie zaopiniowany m.in. przez OPI PIB jako jedno z Centrów Kompetencji wskazanych przez właściwe ministerstwa. Oznacza to, że logika standardu KRONIK@, w tym rozróżnienie obiektu od plików/skanów składających się na jego cyfrowe odwzorowanie, powinna być beneficjentowi znana i powinna zostać uwzględniona przy opisie metodologii KPI w projekcie WIEDZ@.

Standard metadanych KRONIK@ nie utożsamia obiektu z pojedynczym plikiem lub pojedynczym skanem. Wskazuje, że w skład jednego obiektu może wchodzić dowolna liczba plików. Oznacza to, że pojedyncze skany, strony, karty lub klatki mikrofilmowe mogą być plikami składowymi albo częściami cyfrowego odwzorowania jednego obiektu fizycznego, dokumentu, publikacji lub jednostki archiwalnej. Nie powinny być więc automatycznie opisywane jako odrębne obiekty źródłowe.

Takie rozumienie potwierdzają również referencyjne wytyczne Centrów Kompetencji właściwych dla digitalizacji zasobów archiwalnych i bibliotecznych.

W odniesieniu do materiałów archiwalnych Narodowe Archiwum Cyfrowe, pełniące funkcję Centrum Kompetencji w zakresie digitalizacji materiałów archiwalnych, udostępnia „Podręcznik Digitalizatora. Katalog dobrych praktyk i standardów digitalizacji obiektów archiwalnych”. Wytyczne te rozróżniają jednostkę archiwalną lub inwentarzową od kolejnych skanów wykonywanych w jej obrębie. Skan występuje jako techniczny element odwzorowania danej jednostki, np. jako kolejny skan w obrębie jednostki archiwalnej, a nie jako samodzielny obiekt źródłowy.

Analogiczny wniosek wynika z wytycznych Biblioteki Narodowej, pełniącej funkcję Centrum Kompetencji w zakresie digitalizacji materiałów bibliotecznych. W opracowaniach BN dotyczących digitalizacji piśmiennictwa projekt digitalizacyjny planuje się przez określenie liczby i charakteru obiektów oraz parametrów technicznych ich odwzorowania. Wytyczne BN odróżniają obiekt fizyczny, odpowiadający mu obiekt cyfrowy, metadane obiektu oraz pliki powstające w procesie digitalizacji. Wprost wskazują również, że jeden obiekt cyfrowy może składać się z wielu, a nawet setek skanów. Tym samym skan jest elementem odwzorowania obiektu bibliotecznego, a nie automatycznie odrębnym dokumentem lub obiektem źródłowym.

W konsekwencji, jeżeli KPI ma być rozumiany jako liczba zdigitalizowanych dokumentów lub obiektów źródłowych, powinien być liczony na poziomie dokumentu, publikacji, jednostki archiwalnej, obiektu fizycznego albo innej logicznej jednostki źródłowej podlegającej digitalizacji. Nie powinien być liczony na poziomie pojedynczych skanów, stron, kart, klatek mikrofilmowych ani obrazów cyfrowych, chyba że dany skan stanowi pełne odwzorowanie samodzielnego, jednoelementowego obiektu, np. pojedynczej fotografii, plakatu lub pocztówki.

Jeżeli natomiast beneficjent zamierza monitorować techniczny wolumen prac digitalizacyjnych, tj. liczbę skanów, stron, kart, klatek mikrofilmowych lub obrazów cyfrowych, powinien wykazać tę wartość odrębnie, np. jako dodatkowy wskaźnik własny albo informacyjny dotyczący liczby technicznych jednostek digitalizacji. Nie powinien jednak utożsamiać tej wartości z liczbą zdigitalizowanych dokumentów w ramach wskaźnika programowego. Wskaźnik „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” powinien być liczony zgodnie z metodologią FERC, tj. z uwzględnieniem zasady, że jeden utwór powinien być zliczany jako jeden dokument niezależnie od liczby części lub plików, w których został zdigitalizowany.

Dodatkowo odpowiedź beneficjenta potwierdza niespójność w zakresie wartości docelowej KPI. Beneficjent wskazał, że po aktualizacji tabeli „szacowana populacja zasobów objętych digitalizacją wynosi 5 350 000 szt., a zatem przekracza wartość docelową wskaźnika określoną na poziomie 5 000 000 szt.” Jednocześnie beneficjent podtrzymuje wartość docelową KPI na poziomie 5 000 000 szt., mimo że sam wskazuje wyższą wartość wynikającą z aktualnego zliczenia.

Jest to niespójne z dalszym stwierdzeniem beneficjenta, że: „na etapie OZPI i wniosku o dofinansowanie wartości wskaźników muszą być prezentowane jako konkretne wartości docelowe, a nie jako wartości przybliżone”. Skoro beneficjent po ponownej weryfikacji wskazuje wartość 5 350 000 szt., to utrzymywanie wartości 5 000 000 szt. jako wartości „ostrożnościowej” wymaga jednoznacznego doprecyzowania, że jest to konkretna wartość docelowa przyjęta do rozliczenia projektu, a nie szacunkowa liczba wszystkich jednostek możliwych do digitalizacji. W przeciwnym razie opis wskaźnika może sugerować, że wartość 5 000 000 szt. jest przybliżeniem całkowitego wolumenu zasobów przewidzianych do digitalizacji.

W związku z powyższym prosimy o poprawienie opisu KPI oraz metodologii jego liczenia w OZPI i we wniosku o dofinansowanie zgodnie z metodologią wynikającą z „Instrukcji wypełniania wniosku o dofinansowanie” dla naboru nr FERC.02.03-IP.01-005/26. W obecnej formie OZPI nie jest zgodne z założeniami FERC, ponieważ opis wskaźnika może sugerować, że wartość docelowa 5 000 000 szt. odnosi się do liczby zdigitalizowanych dokumentów, podczas gdy z wyjaśnień beneficjenta wynika, że faktycznie chodzi o liczbę skanów, obrazów cyfrowych, stron, kart lub klatek mikrofilmowych. Liczba technicznych jednostek digitalizacji może zostać wykazana odrębnie, natomiast wskaźnik programowy powinien być liczony zgodnie z zasadą, że jeden utwór powinien być zliczany jako jeden dokument niezależnie od liczby części lub plików, w których został zdigitalizowany. Do czasu dokonania tej korekty uwaga Ministerstwa Cyfryzacji pozostaje aktualna.

ODPOWIEDŹ OPI PIB nr 4 z 25.06.2026:

Ad 6. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych oraz wartość wskaźników dotyczących liczby zdigitalizowanych i udostępnionych dokumentów

Beneficjent dziękuje za doprecyzowanie uwagi Ministerstwa Cyfryzacji dotyczącej metodologii liczenia wskaźników: „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” oraz „Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne”.

Po ponownej analizie uwagi oraz po dodatkowej weryfikacji danych przekazanych przez partnera odpowiedzialnego za digitalizację, beneficjent przyjmuje, że w dokumentacji projektu należy rozdzielić liczbę dokumentów źródłowych od technicznego wolumenu prac digitalizacyjnych, obejmującego skany, strony, karty, klatki mikrofilmowe lub obrazy cyfrowe.

W związku z tym beneficjent dokona korekty wartości dwóch wskaźników programowych:

- „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” — z 5 000 000 szt. na 900 000 szt.;
- „Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne” — z 5 000 000 szt. na 900 000 szt.

Korekta ta wynika z przyjęcia metodologii zgodnej z definicją wskaźnika programowego oraz z zasadą, że jeden utwór w rozumieniu przepisów o prawie autorskim i prawach pokrewnych powinien być liczony jako jeden dokument, niezależnie od liczby części, plików, skanów lub obrazów cyfrowych, w których został zdigitalizowany.

Jednocześnie beneficjent podkreśla, że zmiana wartości wskaźników nie oznacza zmniejszenia zakresu rzeczowego digitalizacji. Projekt nadal obejmuje bardzo szeroki wolumen prac digitalizacyjnych, szacowany na ok. **5 000 000–5 350 000 skanów, stron, kart, klatek mikrofilmowych lub obrazów cyfrowych**. Wartość ta będzie ujmowana w dokumentacji jako parametr opisujący techniczną skalę digitalizacji, a nie jako wartość wskaźnika programowego.

W konsekwencji zmianie ulegnie również **tabela 2.3 w OZPI**, dotycząca udostępnionych informacji sektora publicznego i zdigitalizowanych zasobów. Tabela zostanie zaktualizowana w taki sposób, aby opisy zasobów i wartości liczbowe odnosiły się do liczby dokumentów źródłowych, a nie do pojedynczych stron, kart, klatek mikrofilmowych lub skanów.

Zaktualizowana tabela będzie obejmowała m.in. następujące kategorie zasobów:

Lp.	Zakres zasobów	Liczba dokumentów
1	Zdigitalizowane i udostępnione dokumenty aktowe Ministerstwa Spraw Zagranicznych RP oraz polskich placówek dyplomatycznych przechowywane w Hoover Institution, obejmujące m.in. dokumentację ambasad i konsulatów.	280 000 szt.
2	Zdigitalizowane i udostępnione materiały Ministerstwa Informacji i Dokumentacji z lat 1939–1945, obejmujące dokumenty administracyjne, materiały informacyjne i dokumentację działalności rządu emigracyjnego.	70 000 szt.
3	Udostępnione cyfrowo mikrofilmy historyczne, obejmujące wcześniej wykonane odwzorowania zbiorów polskich.	90 000 szt.
4	Zdigitalizowane i udostępnione kolekcje prywatne osób życia publicznego i kulturalnego, w tym m.in. Jana Karskiego, Ignacego J. Paderewskiego, Tadeusza Bora-Komorowskiego, Stefana Korbońskiego oraz Augusta Zaleskiego.	120 000 szt.
5	Zdigitalizowane i udostępnione druki podziemne, materiały z okresu II wojny światowej, Powstania Warszawskiego, pierwszych lat powojennych oraz epoki „Solidarności”, przekształcone do postaci umożliwiającej ich przeszukiwanie.	240 000 szt.
6	Zdigitalizowane i udostępnione wydawnictwa emigracyjne oraz materiały propagandowe, w tym druki przygotowane na konferencję wersalską oraz publikacje z lat 1939–1989.	100 000 szt.
Razem		900 000 szt.

Beneficjent doprecyzuje jednocześnie, że wskazane dokumenty zostaną zdigitalizowane i udostępnione online, natomiast wolumen skanów, stron, kart, klatek mikrofilmowych i obrazów cyfrowych będzie wykazywany wyłącznie jako informacja pomocnicza pokazująca skalę prac technicznych.

Podsumowując, beneficjent przyjmuje uwagę Ministerstwa Cyfryzacji w zakresie potrzeby korekty sposobu prezentacji wartości wskaźników. W OZPI i we wniosku o dofinansowanie zostaną skorygowane wartości dwóch wskaźników programowych z **5 000 000 szt. na 900 000 szt.**, a także odpowiednio zaktualizowana zostanie **tabela 2.3 OZPI**.

KORESPONDECJA OPI PIB – MC za pośrednictwem KdSC nr 5

UWAGI MC NR 5 z 26.06.2026 24

Dzień dobry,

z poniższą uwagą DZD uznaje wyjaśnienia beneficjenta za przyjęte:

Ad 6. Rozróżnienie zasobów nowo digitalizowanych i zasobów już cyfrowych oraz wartość wskaźników dotyczących liczby zdigitalizowanych i udostępnionych dokumentów

Ministerstwo Cyfryzacji akceptuje wyjaśnienia beneficjenta dotyczące sposobu liczenia wskaźników „Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne” oraz „Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne”, w szczególności korektę wartości docelowych tych wskaźników z 5 000 000 szt. na 900 000 szt. oraz rozdzielenie liczby dokumentów/obiektów źródłowych od technicznego wolumenu prac digitalizacyjnych, obejmującego skany, strony, karty, klatki mikrofilmowe lub obrazy cyfrowe.

Ministerstwo Cyfryzacji rozumie, że w wyniku tej korekty jako 1 szt. na potrzeby wskaźnika liczony będzie jeden dokument/obiekt źródłowy, w tym jeden utwór w rozumieniu przepisów o prawie autorskim i prawach pokrewnych, niezależnie od liczby skanów, stron, kart, plików lub innych technicznych jednostek digitalizacji, w których zostanie utrwalony jego cyfrowy obraz.

Jednocześnie Ministerstwo Cyfryzacji zauważa, że równość wartości wskaźnika dotyczącego liczby zdigitalizowanych dokumentów oraz wskaźnika dotyczącego liczby dokumentów udostępnionych on-line oznacza, że projekt zakłada udostępnienie w internecie wszystkich obiektów objętych digitalizacją. W tym kontekście powstaje wątpliwość, czy wszystkie obiekty liczone jako 1 szt. mogą zostać zgodnie z prawem upublicznione w internecie.

Wątpliwość ta dotyczy w szczególności zasobów, których opis wskazuje na datę wytworzenia „do 1989 r.”. Taki zakres czasowy sugeruje, że co najmniej część obiektów może nadal podlegać ochronie na podstawie ustawy o prawie autorskim i prawach pokrewnych. W konsekwencji konieczne jest ustalenie, czy są to materiały znajdujące się w domenie publicznej albo czy beneficjent posiada stosowne licencje lub zgody autora, jego następców prawnych albo spadkobierców, umożliwiające publiczne udostępnienie tych treści w internecie.

W związku z powyższym Ministerstwo Cyfryzacji wskazuje, że w przygotowywanym studium wykonalności powinna zostać doprecyzowana liczba obiektów, które zgodnie z obowiązującymi przepisami mogą zostać upublicznione w internecie. Prosimy również o wskazanie sposobu weryfikacji statusu prawnego zasobów objętych digitalizacją oraz odróżnienia obiektów możliwych do pełnego udostępnienia online od obiektów, których udostępnienie będzie ograniczone ze względu na ochronę praw autorskich lub praw pokrewnych.

ODPOWIEDŹ OPI PIB nr 5 z 03.07.2026

Beneficjent dziękuje za przyjęcie wyjaśnień dotyczących korekty wartości wskaźników oraz sposobu liczenia dokumentów objętych digitalizacją i udostępnieniem online. Wniosek w raz z wymaganymi załącznikami został wysłany w wymaganym dla naboru terminie 30.06.2026 r.

Jednocześnie beneficjent wskazuje, że w Studium Wykonalności ujęto planowany model prawny cyfrowego udostępnienia zasobów oraz prawne możliwości i ograniczenia ich ponownego wykorzystania. W dokumencie wskazano, że zasoby objęte projektem mogą podlegać zróżnicowanym reżimom prawnym, w tym obejmującym domenę publiczną, licencje Creative Commons, dzieła osierocone oraz dostęp ograniczony.

Beneficjent zakłada, że przed pełnym udostępnieniem treści online zostanie zweryfikowany status prawny poszczególnych grup zasobów. Pełne udostępnienie online będzie dotyczyło wyłącznie tych dokumentów, dla których istnieje odpowiednia podstawa prawna, w szczególności dokumentów znajdujących się w domenie publicznej albo objętych stosownymi licencjami lub zgodami. W przypadku materiałów, których status prawny będzie wymagał ograniczenia dostępu, udostępnione zostaną metadane, opisy kolekcji albo dostęp warunkowy, zgodnie z przyjętym modelem prawnym.

Ewentualne dalsze doprecyzowania w tym zakresie zostaną przedstawione na etapie oceny wniosku o dofinansowanie oraz w toku realizacji projektu, zgodnie z uwagami ekspertów i wymogami właściwej instytucji.