

OPIS ZAŁOŻEŃ PROJEKTU INFORMATYCZNEGO

Tytuł projektu	Jeziora Danych Naukowych		
Wnioskodawca	Minister Nauki i Szkolnictwa Wyższego		
Beneficjent	Politechnika Gdańska		
Partnerzy	Nie dotyczy – projekt realizowany samodzielnie przez Politechnikę Gdańską		
Źródło finansowania	- Fundusze Europejskie na Rozwój Cyfrowy 2021-2027 (FERC), Oś priorytetowa 2. Zaawansowane usługi cyfrowe, Działanie 02.03 Cyfrowa dostępność i ponowne wykorzystanie informacji sektora publicznego – nauka. - Budżet państwa - cz. nr 27 Informatyzacja		
Całkowity koszt projektu	25 330 817,10 zł		
Planowany okres realizacji projektu	01-2027 do 06-2029		
Osoba kontaktowa	Andrzej Sobeczki	andrzej.sobeczki@pg.edu.pl	48583471744

1. POWODY PODJĘCIA PROJEKTU

1.1. Identyfikacja problemu i potrzeb

Badania naukowe coraz częściej opierają się na analizie dużych, heterogenicznych zbiorów danych. Ich wartość zależy od tego, czy są odnajdywalne, maszynowo interpretowalne, możliwe do ponownego wykorzystania oraz interoperacyjne w skali międzynarodowej. Polskie środowisko naukowe napotyka bariery systemowe. Projekt adresuje je poprzez rozszerzenie platformy MOST Wiedzy oraz uruchomienie nowej platformy analitycznej:

A. METADANE BEZ SEMANTYKI DZIEDZINOWEJ

1. Wartości metadanych (np. diagnozy, biomarkery) mają formę tekstu swobodnego, niepowiązanego z URI ontologii (np. UBERON, IEEE Thesaurus), co uniemożliwia maszynowe wnioskowanie i federacyjne wyszukiwanie w EOSC. Projekt dodaje do MOST Wiedzy moduł metadanych dziedzinowych powiązanych z URI ontologii.

2. Ocena jakości metadanych jest manualna. Brak FAIR Implementation Profiles (FIP), czyli deklaracji wdrażania FAIR. Projekt dodaje moduł wsparcia FIP z asystentem AI mapującym metadane do ontologii.

B. BARIERY W PRZETWARZANIU I ODTWARZALNOŚCI

3. Brak zarządzanego środowiska analitycznego powoduje, że naukowcy odtwarzają konfigurację lokalnie. Projekt buduje platformę analityczną z reużywalnymi workflowami i biblioteką modeli AI/ML.

4. Metadane opisują zbiór danych, lecz nie obejmują pełnego Research Object. Projekt dodaje do MOST Wiedzy moduł pakowania danych w formacie RO-Crate.

5. Brak integracji repozytorium z planami zarządzania danymi (DMP) i manualne deponowanie. Projekt udostępnia API MOST Wiedzy do maszynowego deponowania danych z aparatury.

C. DANE WRAŻLIWE

6. Brak procesów wnioskowania o dostęp, warunków wykorzystania i audytu. Projekt dodaje moduł kontrolowanego dostępu zgodny z CARE i TRUST.

D. KOMPETENCJE I EKOSYSTEM

7. Repozytoria nie mają wsparcia w certyfikacji CoreTrustSeal i integracji z EOSC. Brak jednolitej metodyki pracy z dużymi zbiorami. Projekt rozszerza Centrum Kompetencji Otwartej Nauki o

mentoring i materiały szkoleniowe.

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
Politechnika Gdańska (PG)	<ul style="list-style-type: none"> - ograniczona funkcjonalność platformy MOST Wiedzy (mostwiedzy.pl - instytucjonalna platforma otwartej nauki utrzymana i rozwijana przez Politechnikę Gdańską, szerzej w pkt 1.2) w zakresie semantyki dziedzinowej i automatycznej oceny jakości metadanych (manualna i nieskalowalna ocena) - brak komponentu analitycznego umożliwiającego naukowcom PG łatwy dostęp do uczelnianych zasobów obliczeniowych w celu analizy danych zdeponowanych na platformie - Ograniczona funkcjonalność repozytorium danych badawczych uniemożliwiająca rozwój usług przetwarzania i analizy danych 	1
Naukowcy i doktoranci Politechniki Gdańskiej (deponenci i użytkownicy danych)	<ul style="list-style-type: none"> - Brak łatwego dostępu do uczelnianych zasobów obliczeniowych w celu analizy danych badawczych - konieczność lokalnego odtwarzania środowisk ogranicza skalę analiz - Brak narzędzi umożliwiających opis danych z wykorzystaniem metadanych dziedzinowych powiązanych z ontologiami referencyjnymi (np. MeSH, UBERON, IEEE Thesaurus), co utrudnia odnajdywanie i ponowne wykorzystanie danych - Brak automatycznej walidacji jakości danych i zgodności z zasadami FAIR – ocena ma charakter manualny i nieskalowalny - Ograniczone wsparcie w bezpiecznym udostępnianiu danych wrażliwych (brak procesów wnioskowania o dostęp, definiowania warunków wykorzystania i audytu użycia) - Brak narzędzi AI wspomagających tworzenie i spełnianie FAIR Implementation Profiles (np. Doboru właściwych ontologii na podstawie metadanych) - Brak możliwości automatycznego deponowania danych z aparatury badawczej 	1200
Specjaliści data science i użytkownicy zaawansowanej analizy danych (AI/	<ul style="list-style-type: none"> - brak spójnej, otwartej platformy umożliwiającej tworzenie, uruchamianie i ponowne wykorzystanie workflowów oraz pipeline'ów analitycznych dla różnych 	2000

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
ML)	<p>dziedzin, zintegrowanej z repozytorium danych badanych</p> <ul style="list-style-type: none"> - Brak przenośności pełnego kontekstu badawczego: dane nie są pakowane wraz z kodem, workflowem i środowiskiem obliczeniowym (RO-Crate), co ogranicza odtwarzalność i transfer wyników - Brak jednolitej metodyki pracy z dużymi zbiorami danych – rozproszenie narzędzi i podejść utrudnia powtarzalność analiz - Brak integracji narzędzi analitycznych z repozytorium danych, co ogranicza możliwość pracy bezpośrednio na danych zdeponowanych. 	
Data stewardzi i osoby odpowiedzialne za zarządzanie danymi badawczymi)	<ul style="list-style-type: none"> - Brak dziedzinowych wytycznych i profili implementacyjnych FAIR Implementation Profiles (FIP) – kolektywnych, maszynowo czytelnych deklaracji wyborów technologicznych zasad FAIR - Brak narzędzi automatycznej walidacji jakości metadanych dostosowanych do specyfiki dziedzinowej - Brak automatycznych narzędzi AI wspierających opis i walidację metadanych (zgodność z politykami FIP) oraz analizę danych. - Ograniczone systemowe wsparcie w standaryzacji opisu i kontroli jakości danych zgodnie ze standardami międzynarodowymi 	100
Menadżerowie repozytoriów danych badawczych w Polsce	<ul style="list-style-type: none"> - brak dostępu do mentoringu technicznego w zakresie wdrażania standardów zarządzania danymi - trudności w implementacji dziedzinowych modeli metadanych, ontologii i identyfikatorów PID (Persistent Identifier – unikalny i trwały identyfikator zasobu) - brak wzorcowych rozwiązań i narzędzi do integracji z infrastrukturą europejską 	100
Naukowcy z innych uczelni i instytutów badawczych w Polsce (poza PG)	<ul style="list-style-type: none"> - ograniczona interoperacyjność metadanych między instytucjami - brak dostępu do ustandaryzowanych środowisk analitycznych - brak zaawansowanych narzędzi do wyszukiwania i analizy danych 	70000
Użytkownicy spoza środowiska	<ul style="list-style-type: none"> - trudności w odnajdywaniu i wykorzystaniu danych badawczych ze względu na 	3000

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
naukowego (przedsiębiorcy, administracja publiczna, organizacje społeczne)	niewystarczającą jakością i szczegółowością metadanych <ul style="list-style-type: none"> - brak przyjaznych narzędzi umożliwiających analizę danych pochodzących z różnych instytucji - ograniczony dostęp do danych wymagających szczególnych warunków udostępniania (brak przejrzystych procedur wnioskowania o dostęp oraz warunków ich wykorzystania) - ograniczone możliwości ponownego wykorzystania danych z uwagi na brak spójnego opisu i kontekstu ich powstania 	
Usługi (Systemy) działające w ramach europejskich i krajowych infrastruktur danych (np. EOSC)	<ul style="list-style-type: none"> - ograniczona interoperacyjność metadanych i standardów między repozytoriami z powodu niewystarczającej semantyki dziedzinowej metadanych - niepełna interoperacyjność usług polskiej nauki z innymi europejskimi infrastrukturami danych badawczych - brak Research Objects pakowanych standardowo (RO-Crate) ogranicza możliwość integracji wyników badań w środowiskach federacyjnych 	50
Ministerstwo Nauki i Szkolnictwa Wyższego	<ul style="list-style-type: none"> - ograniczona dostępność narzędzi wspierających realizację polityk otwartej nauki i zarządzania danymi badawczymi na poziomie krajowym - brak spójnych standardów i rozwiązań umożliwiających interoperacyjność danych badawczych między instytucjami 	1

1.2. Opis stanu obecnego

Politechnika Gdańska (PG) od 2017 r. utrzymuje i rozwija platformę MOST Wiedzy, zbudowaną w projektach MOST Wiedzy (POPC 2.3.1, 2016–2019) i MOST DANYCH (POPC 2.3.1, 2018–2021). Platforma jest największym instytucjonalnym repozytorium otwartej nauki w północnej Polsce. Zasoby (05-2026): 76 641 publikacji, 25 642 zbiory otwartych danych badawczych, 13 952 profili naukowców, 1 558 opisów projektów. Dane deponują 4 instytucje oraz indywidualni naukowcy spoza PG. Odnotowano 1,7 mln pobrań pełnych tekstów. Repozytorium Danych Badawczych posiada certyfikat CoreTrustSeal – pierwsze certyfikowane repozytorium instytucjonalne danych badawczych w Polsce. Otwartość i bezpłatność: >95% zasobów udostępnianych bezpłatnie na licencjach Creative Commons 4.0 i Open Database License, z trwałymi identyfikatorami DOI (Digital Object Identifier). Metadane rekordów eksponowane jako Linked Open Data – JSON-LD ze słownikiem schema.org, RDF (Resource Description Framework), URI każdego zasobu, linki do zewnętrznych grafów (DataCite, ORCID, ROR); poziom 5 gwiazdek w skali 5 Open Data dla struktury rekordu. Dane wrażliwe – w modelu kontrolowanego dostępu. Usługi: deponowanie i udostępnianie publikacji oraz danych katalogowanie i wyszukiwanie zarządzanie profilami naukowców.

Infrastruktura wykorzystywana w projekcie: autorskie oprogramowanie utrzymywane przez Centrum Usług Informatycznych PG; istniejąca własna infrastruktura serwerowa PG – środowisko zwirtualizowane VMware i klaster Kubernetes; serwerownia PG klasy TIER III/IV. Integracje: DataCite, ORCID, OpenAIRE (OAI-PMH), OJS, Indico, EOSC, federacja Edugain. Publiczne, udokumentowane interfejsy API REST/JSON i OAI-PMH dla maszynowego pobierania zasobów.

Standardy i bezpieczeństwo: dostępność WCAG 2.0 AA, zgodność z KRI, polityka OWASP ASVS, regularne testy penetracyjne. Projektowanie zorientowane na użytkownika (User-Centered Design) stosowane od początku rozwoju platformy.

Co zmieni projekt – patrz pkt 2.1 (cele) i 2.4 (produkty).

2. EFEKTY PROJEKTU

2.1. Cele i korzyści wynikające z projektu

Cel - 1	Zwiększenie możliwości ponownego wykorzystania danych badawczych poprzez ich udostępnianie w postaci ustandaryzowanych obiektów badawczych.
Cel strategiczny	Strategia Produktywności 2030, obszary: Wiedza - Wzrost intensywności wykorzystania wiedzy i technologii w gospodarce; Dane - Rozwój algorytmicznej gospodarki opartej na danych, czyli np. rozwój sztucznej inteligencji.
Korzyść:	Zapewnienie odtwarzalności i transferu wyników badań między ośrodkami, Na wysoką jakość naszych danych będzie miało wpływ ich udostępnienie wraz z kontekstem badawczym, metadanymi dziedzinowymi oraz mechanizmami zgodnymi z zasadami FAIR i CARE.
KPI:	KPI-1: Liczba zasobów udostępnionych w postaci RO-Crate KPI-2: WLWK-RCR011 – Użytkownicy nowych i zmodernizowanych publicznych usług, produktów i procesów cyfrowych KPI-3: WLWK-PLR0010 – Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne on-line KPI-4: Liczba pobrań zasobów on-line udostępnionych w postaci RO-Crate
Wartość aktualna i docelowa KPI:	KPI-1: 0 KPI-2: 0 KPI-3: 0 KPI-4: 0 KPI-1: 5000 KPI-2: 1000 KPI-3: 1 KPI-4: 10000
Metoda pomiaru KPI	KPI-1: zliczanie rekordów z typem RO-Crate w systemie repozytoryjnym. Źródło: system monitoringu platformy MOST Wiedzy. Częstotliwość: kwartalna. Termin pomiaru wartości docelowej: 06-2029. KPI-2: Zliczanie unikalnych użytkowników nowych i zmodernizowanych usług platformy; ten sam użytkownik korzystający wielokrotnie w roku liczony jednokrotnie; Źródło danych: Baza użytkowników platformy analitycznej.

	<p>Termin: kwartalnie w okresie 07-2029 do 06-2030; termin pomiaru wartości docelowej: 06-2030 (12 miesięcy po zakończeniu projektu).</p> <p>KPI-3: Politechnika Gdańska udostępnia ORD on-line w wyniku interwencji. Źródło: system monitoringu platformy MOST Wiedzy. Częstotliwość: kwartalna. Termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-4: Zliczanie pobrań rekordów z typem RO-Crate w systemie. Źródło: system monitoringu platformy MOST Wiedzy. Częstotliwość: roczna. Termin pomiaru wartości docelowej: 06-2029</p>
Cel - 2	Podniesienie semantycznej interoperacyjności dziedzinowych metadanych badawczych zgodnie z zasadami FAIR.
Cel strategiczny	Strategia Produktywności 2030, obszary: Wiedza - Wzrost intensywności wykorzystania wiedzy i technologii w gospodarce; Dane - Rozwój algorytmicznej gospodarki opartej na danych, czyli np. rozwój sztucznej inteligencji.
Korzyść:	Usprawnienie wyszukiwania zasobów polskiej nauki oraz spójnej interpretacji metadanych między instytucjami. Poprawa wymiany danych badawczych między instytucjami badawczymi w Polsce i za granicą.
KPI:	KPI-PROG-FERC-P013 Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne
Wartość aktualna i docelowa KPI:	KPI- 0 KPI- 1500
Metoda pomiaru KPI	Zliczanie udostępnionych on-line dokumentów (datasetów) zawierających informacje sektora publicznego (dane badawcze), opisanych wyszukiwalnymi semantycznymi metadanymi dziedzinowymi powiązаныmi z URI ontologii, rejestrowanych w zewnętrznych bazach (OpenAIRE). Każdy taki dataset stanowi udostępnioną on-line ISP. Źródło: baza danych platformy. Częstotliwość: kwartalna. Termin pomiaru wartości docelowej: 06-2029.
Cel - 3	Ułatwienie naukowcom przetwarzania danych badawczych z wykorzystaniem metod AI/ML oraz uproszczenie dostępu do zasobów obliczeniowych uczelni.
Cel strategiczny	Strategia na rzecz Odpowiedzialnego Rozwoju; cel szczegółowy III: Skuteczne państwo i instytucje służące wzrostowi oraz włączeniu społecznemu i gospodarczemu. Strategia Produktywności 2030, obszary: Wiedza - Wzrost intensywności wykorzystania wiedzy i technologii w gospodarce; Dane - Rozwój algorytmicznej gospodarki opartej na danych, czyli np. rozwój sztucznej inteligencji.
Korzyść:	Eliminacja konieczności lokalnego odtwarzania konfiguracji obliczeniowych i zwiększenie skali analiz danych badawczych; Zwiększenie efektywności wykorzystania istniejącej infrastruktury badawczej i obliczeniowej na Politechnice Gdańskiej; Usprawnienie tworzenia procesów przetwarzania danych badawczych, w tym w oparciu o metody ML oraz otwarte modele AI.
KPI:	KPI-1: PROG-FERC-P015 - Liczba uruchomionych systemów teleinformatycznych. KPI-2: WLWK-PLR0160 – Liczba rozwiązań wykorzystujących informacje sektora publicznego/dane prywatne.

	<p>KPI-3: WLWK-RC0013 - Wartość usług, produktów i procesów cyfrowych opracowanych dla przedsiębiorstw</p> <p>KPI-4: WLWK-PLR0008 – Liczba podmiotów wspartych w zakresie rozwoju usług, produktów i procesów cyfrowych</p> <p>KPI-5: Liczba pojedynczych wykonań workflow naukowych na Platformie Analitycznej.</p>
<p>Wartość aktualna i docelowa KPI:</p>	<p>KPI-1: 0</p> <p>KPI-2: 0</p> <p>KPI-3: 0</p> <p>KPI-4: 0</p> <p>KPI-5: 0</p> <p>KPI-1: 1</p> <p>KPI-2: 7</p> <p>KPI-3: 2 436 000 PLN</p> <p>KPI-4: 1</p> <p>KPI-5: 5000</p>
<p>Metoda pomiaru KPI</p>	<p>KPI-1: Weryfikacja uruchomienia produkcyjnego Platformy Analitycznej - protokół odbioru końcowego; Źródło danych: protokół odbioru technicznego + dokumentacja wdrożenia; Częstotliwość + termin: jednorazowo (przy wdrożeniu produkcyjnym); termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-2: protokół odbioru technicznego + dokumentacja wdrożenia + zrzut produkcyjny CI TASK; Na wartość docelową (7) składają się rozwiązania cyfrowe wykorzystujące ISP, wytworzone w projekcie: (1) moduł metadanych dziedzinowych powiązanych z URI ontologii; (2) moduł pakowania danych w standardzie RO-Crate 1.1; (3) asystent AI FAIR / FAIR Implementation Profiles; (4) moduł kontrolowanego dostępu zgodny z zasadami CARE i TRUST; (5) API maszynowego deponowania danych z aparatury laboratoryjnej; (6) serwer Model Context Protocol; (7) Platforma Analityczna. Częstotliwość + termin: jednorazowo (przy wdrożeniu produkcyjnym); termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-3: Zgodnie z definicją Instrukcji wypełniania wniosku o dofinansowania dla konkursu nr FERC.02.03-IP.01-005/26 wartość „wydatków kwalifikowalnych w projekcie, przeznaczonych na wytworzenie produktów, usług, procesów cyfrowych. Nie są wliczane wydatki na wynagrodzenia, szkolenia, informację i promocję oraz koszty pośrednie.”; Źródło danych: dane systemu księgowego, faktury, potwierdzenia przelewów; Częstotliwość + termin: kwartalnie; termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-4: Dokumentacja rozliczenia funduszy projektowych dla beneficjenta projektu. Źródło danych: Dokumentacja projektowa i odbiór projektu; Częstotliwość: Na koniec projektu; termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-5: Dane statystyczne Platformy Analitycznej. Źródło danych: Monitoring systemu “Platforma Analityczna”. Częstotliwość + termin: Kwartalnie, termin pomiaru wartości docelowej: 06-2029.</p>
<p>Cel - 4</p>	<p>Zwiększenie poziomu bezpieczeństwa i kontroli dostępu do danych badawczych.</p>
<p>Cel strategiczny</p>	<p>Strategia na rzecz Odpowiedzialnego Rozwoju do roku 2020 (z perspektywą do 2030); cel szczegółowy III: Skuteczne państwo i instytucje służące</p>

	wzrostowi oraz włączeniu społecznemu i gospodarczemu, Obszar: E-Państwo.
Korzyść:	Umożliwienie bezpiecznego udostępniania danych wymagających szczególnej ochrony (medycznych, dotyczących osób, środowiskowych).
KPI:	Liczba zbiorów danych udostępnianych w modelu kontrolowanego dostępu.
Wartość aktualna i docelowa KPI:	0 50
Metoda pomiaru KPI	Zliczanie rekordów oznaczonych jako kontrolowany dostęp z przypisaną polityką wykorzystania w systemie repozytoryjnym. Źródło: baza danych platformy MOST Wiedzy. Częstotliwość: kwartalna. Termin pomiaru wartości docelowej: 06-2029.
Cel - 5	Umożliwienie maszynowego deponowania danych do repozytorium z automatyczną normalizacją do przyjętego standardu.
Cel strategiczny	Strategia Produktywności 2030, obszary: Wiedza - Wzrost intensywności wykorzystania wiedzy i technologii w gospodarce; Dane - Rozwój algorytmicznej gospodarki opartej na danych, czyli np. rozwój sztucznej inteligencji.
Korzyść:	Zwiększenie dostępności surowych danych badawczych i powiązanie repozytorium z planami zarządzania danymi (DMP) wymaganymi przez politykę NCN, Horizon Europe; Uproszczenie procesu deponowania danych badawczych z istniejącej i przyszłej infrastruktury badawczej.
KPI:	KPI-1: PROG-FERC-P014 - Liczba utworzonych API KPI-2: Liczba wygenerowanych kluczy API
Wartość aktualna i docelowa KPI:	KPI-1: 0 KPI-2: 0 KPI-1: 1 KPI-2: 10
Metoda pomiaru KPI	KPI-1: Weryfikacja udostępnienia interfejsu API z dokumentacją OpenAPI/ Swagger; Źródło danych: dokumentacja techniczna API + protokół odbioru końcowego komponentu deponowania + endpoint produkcyjny; Częstotliwość + termin: jednorazowo (przy wdrożeniu produkcyjnym); termin pomiaru wartości docelowej: 06-2029. KPI-2: Odczyt liczby wygenerowanych kluczy API dla platformy analitycznej. Źródło: rejestr użytkowników platformy analitycznej. Częstotliwość: kwartalna. Termin pomiaru wartości docelowej: 06-2029.
Cel - 6	Rozwój kompetencji polskich repozytoriów instytucjonalnych w zakresie zarządzania danymi badawczymi i integracji z infrastrukturami europejskimi.
Cel strategiczny	Strategia Produktywności 2030, obszary: Organizacja i instytucje - Stymulowanie mechanizmów współpracy pomiędzy podmiotami gospodarczymi; Wiedza - Wzrost intensywności wykorzystania wiedzy i technologii w gospodarce; Dane - Rozwój algorytmicznej gospodarki opartej na danych, czyli np. rozwój sztucznej inteligencji.
Korzyść:	Zwiększenie liczby polskich repozytoriów otrzymujących wsparcie techniczne (przygotowanie do certyfikacji CoreTrustSeal, zwiększenie świadomości zasad FAIR, integracji z innymi usługami i podobne).

KPI:	KPI-1: FERC.02.03: PROG-FERC-P016 (pracownicy nie-IT szkolnie) KPI-2: PROG-FERC-P019 (pracownicy IT szkolnie);
Wartość aktualna i docelowa KPI:	KPI-1: 0 KPI-2: 0 KPI-1: 600, K/M 390/210 KPI-2: 100, K/M 35/65
Metoda pomiaru KPI	<p>KPI-1: Zliczanie uczestników szkoleń, warsztatów i programów mentoringu CKON (Centrum Kompetencji Otwartej Nauki) 2.0 (z rozróżnieniem osób objętych mentoringiem technicznym repozytoriów oraz osób przeszkolonych z metodyki pracy z dużymi zbiorami danych) z podziałem na reprezentowane repozytoria. Źródło: rejestry CKON 2.0 i system ewidencji szkoleń i/lub certyfikat/dyplom/dokument poświadczający ukończenie szkolenia. Częstotliwość: po ukończeniu danego rodzaju wsparcia, nie rzadziej niż raz na kwartał. Termin pomiaru wartości docelowej: 06-2029.</p> <p>KPI-2: Zliczanie uczestników szkoleń, warsztatów i programów mentoringu CKON (Centrum Kompetencji Otwartej Nauki) 2.0 (z rozróżnieniem osób objętych mentoringiem technicznym repozytoriów oraz osób przeszkolonych z metodyki pracy z dużymi zbiorami danych). Źródło: rejestry CKON 2.0 i system ewidencji szkoleń i/lub certyfikat/dyplom/dokument poświadczający ukończenie szkolenia z podziałem na reprezentowane repozytoria. Częstotliwość: po ukończeniu danego rodzaju wsparcia, nie rzadziej niż raz na kwartał. Termin pomiaru wartości docelowej: 06-2029.</p>

2.2. Udostępnione e-usługi

Lp.	Nazwa e-usługi	Typ	Zakres oddziaływania	Poziom dojrzałości e-usługi

2.3. Udostępnione informacje sektora publicznego i zdigitalizowane zasoby

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
Udostępnione nowe zbiory otwartych danych badawczych zdeponowane w ramach projektu.	30-06-2029	1500
Udostępnione istniejące zbiory danych badawczych z repozytorium Open Research Data Catalog w postaci RO-	30-06-2029	5000

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
Crate		

Czy wszystkie zdigitalizowane zasoby objęte projektem będą udostępniane bezpłatnie?

TAK/NIE

2.4. Produkty końcowe projektu

Nazwa produktu	Planowana data wdrożenia
Raport z inicjalnego testu prywatności.	01-2027
Studium wykonalności	01-2027
Dokumentacja analityczna projektu	06-2027
Infrastruktura teleinformatyczna – węzły obliczeniowe GPU/AI, rozszerzenie storage, sprzęt komputerowy zespołu projektowego. Dostarczana etapowo: wstępna infrastruktura na potrzeby pilotażu (PoC) Platformy Analitycznej – KM3 (11-2027); infrastruktura docelowa (pełne skalowanie węzłów GPU/AI i storage) – odbiór w KM6 (12-2028).	12-2028
Modyfikacja systemu teleinformatycznego MOST Wiedzy w zakresie modułów: metadanych dziedzinowych, pakowania danych w formacie ROcrate, wsparcia FAIR Implementation Profiles (w tym asystenta AI wspomagającego dobór odpowiednich ontologii na podstawie metadanych), wyszukiwania semantycznego (RAG) i mapowania metadanych do ontologii dziedzinowych, kontrolowanego dostępu do danych wrażliwych zgodnego z zasadami CARE i TRUST.	06-2029
Interfejs API systemu MOST Wiedzy w zakresie maszynowego deponowania danych z aparatury laboratoryjnej	09-2028
Interfejs API systemu MOST Wiedzy zgodny ze standardem Model Context Protocol	09-2028
Platforma analityczna - system dostarczający naukowcom możliwość deponowania danych z aparatury badawczej oraz przeprowadzania analiz na tych danych.	06-2029
Materiały szkoleniowe i e-learningowe Centrum Kompetencji Otwartej Nauki 2.0.	02-2029
Raport z weryfikacyjnego testu prywatności	02-2029
Raport z badań UX, w tym zgodności z WCAG 2.2	06-2029
Pozytywny raport z testów bezpieczeństwa	06-2029
Pozytywny raport z testów wydajności	06-2029
Materiały informacyjno-promocyjne projektu	06-2029

3. KAMIENIE MIŁOWE

Kamienie milowe	Planowany termin osiągnięcia
KM1 - Zamknięcie Fazy I (Inicjacja): przeprowadzony inicjalny test prywatności (DPIA) i odebrany raport z testu.	2027-01-31
KM2 - Zamknięcie Fazy II (Analiza i projektowanie): odebrany projekt architektoniczny rozwiązania oraz zakończone mapowanie standardów metadanych dziedzinowych, ontologii i słowników kontrolowanych.	2027-06-30
KM3 - Zamknięcie Fazy III, etap pilotażowy (budowa infrastruktury): przygotowana wstępna infrastruktura serwerowa i uruchomiona pilotażowa (PoC) wersja Platformy Analitycznej, weryfikująca architekturę przed skalowaniem docelowym.	2027-11-30
KM4 - Zamknięcie Fazy IV (Rozwój modułów MOST Wiedzy): wdrożone FAIR Implementation Profiles dla wybranych dyscyplin, moduł katalogowania z metadanymi dziedzinowymi oraz moduł pakowania danych w formacie RO-Crate.	2028-04-30
KM5 - Zamknięcie Fazy V (AI/ML i integracje): wdrożony moduł automatycznego deponowania z aparatury, udostępniony moduł AI/ML, wyszukiwania semantycznego i asystent AI wspierający FIP, zrealizowana integracja z EOSC oraz uruchomiony serwer MCP - zgodnie z komponentem „Moduł AI/ML i asystent semantyczny FAIR” (pkt 7.2) i Celem 3.	2028-09-30
KM6 - Odbiór końcowy infrastruktury teleinformatycznej (Faza III, etap docelowy): odebrana docelowa infrastruktura (węzły obliczeniowe GPU/AI, rozszerzenie storage) zamykająca produkt „Infrastruktura teleinformatyczna” (12-2028) - protokół odbioru.	2028-12-31
KM7 - Zamknięcie Fazy VI (Rozwój kompetencji): odebrane materiały szkoleniowe CKON 2.0, przeprowadzone pilotażowe szkolenia oraz weryfikacyjny test prywatności.	2029-02-28
KM8 - Zamknięcie Fazy VII (Wdrożenie produkcyjne): uruchomiony produkcyjnie system Platforma Analityczna oraz moduł kontrolowanego dostępu do danych wrażliwych zgodny z CARE i TRUST.	2029-04-30
KM9 - Odbiór końcowy i zamknięcie projektu: uzyskane pozytywne raporty z testów bezpieczeństwa, wydajności i UX (WCAG 2.2), wytworzone materiały informacyjno-promocyjne, zakończona rzeczowa realizacja - odebrane wszystkie produkty oraz udostępnione zasoby i interfejsy API.	2029-06-30

4. KOSZTY

4.1. Koszty ogólne projektu wraz ze sposobem finansowania

Całkowity koszt projektu (netto oraz brutto), w tym	Netto 24 566 768,32 zł Brutto 25 330 817,10 zł	
Procent dofinansowania ze środków UE (brutto)	79,71%	
Procent środków z budżetu państwa (brutto)	20,29%	
Podział całkowitego kosztu projektu na poszczególne lata (netto oraz brutto)	2027	Netto 9 560 491,99 zł Brutto 9 866 111,50 zł
	2028	Netto 10 050 201,09 zł Brutto 10 355 820,60 zł
	2029	Netto 4 956 075,24 zł Brutto 5 108 885,00 zł

4.2. Wykaz poszczególnych pozycji kosztowych

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
Oprogramowanie	Produkty końcowe powiązane: Produkt nr 3 – Dokumentacja analityczna projektu; Produkt nr 5 – Modyfikacja systemu teleinformatycznego o MOST Wiedzy (moduły: metadanych dziedzinowych, RO-Crate, FAIR Implementation Profiles, kontrolowany dostęp CARE/TRUST); Produkt nr 6 – Interfejs API MOST Wiedzy: maszynowe deponowanie danych z aparatury laboratoryjnej; Produkt nr 7 – Interfejs API MOST Wiedzy: Model	14 454 101,00 zł	Główny składnik projektu (ok.57% budżetu). Stawki personelu wg zestawienia średniego wynagrodzenia pracowników PG na analogicznych stanowiskach (aktualne 04.2026); zastosowano mnożnik wynagrodzenia 1,2114 (narzuty pracodawcy ZUS/PPK). W kolejnych latach realizacji zastosowano wzrost 7% r/r (rok 2: ×1,07; rok 3: ×1,07 ²) na podstawie średniej inflacji GUS 2021-2025. Uwzględniono kwoty dodatkowego wynagrodzenia rocznego. Liczebność i struktura zespołu adekwatna do skali wytworzenia 5 nowych modułów funkcjonalnych i rozbudowy istniejącej platformy MOST Wiedzy/MOST DANYCH.

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>Context Protocol (MCP); Produkt nr 8 – System teleinformatyczny: Platforma Analityczna. KM powiązane: KM2 – projekt architektoniczny + mapowanie standardów; KM3 – pilotażowa Platforma Analityczna; KM4 – FAIR IP + moduł katalogowania + moduł RO-Crate; KM5 – moduł deponowania, AI/ML, asystent FAIR, integracja EOSC, serwer MCP; KM7 – wdrożony moduł kontrolowanego dostępu + Platforma Analityczna + zakończenie realizacji. Zasoby cyfrowe powiązane: Zasób/ISP nr 1 (Udostępnione nowe zbiory otwartych danych badawczych zdeponowane w ramach projektu - 1 500 obiektów), Zasób/ISP nr 2 (Udostępnione istniejące zbiory danych badawczych z repozytorium Open Research Data Catalog w postaci RO-Crate - 5 000 obiektów). Główne kategorie kosztów</p>		

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Wynagrodzenia personelu merytorycznego CUI PG (Centrum Usług Informatycznych Politechniki Gdańskiej) i BPG (Biblioteki Politechniki Gdańskiej) (zespół ok. 35 osób na 30 m-cy w 3 latach budżetowych): lider techniczny, architekci systemu i baz danych, 8 programistów (Java, React, Python/AI, PHP), programista SQL/BI, analityk wymagań, analityk systemowy, ekspert AI/ML, ekspert RO-Crate / workflow CWL, ekspert ds. trwałych identyfikatorów (PID INST / RAiD / DOI / ORCID), 2 DevOps, lider naukowy, lider zespołu Biblioteki PG, redaktorzy metadanych (6 osób na 0,5 etatu), 3 data stewardzi dziedzinowi, ekspert FAIR/CARE/TRUST. • Licencje narzędzi developerskich (IDE, repo, monitoring, analiza statyczna kodu). • Wartości niematerialne i 		

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	prawne - integracja z PID/DOI/RAiD (DataCite, RAiD).		
Infrastruktura	<p>PRODUKTY POWIĄZANE:</p> <ul style="list-style-type: none"> • Produkt nr 4 – Infrastruktura teleinformatyczna (węzły obliczeniowe GPU/ AI, rozszerzenie storage, sprzęt komputerowy zespołu projektowego) – 12-2028 – <p>POZYCJA BEZPOŚREDNIO REALIZUJE TEN PRODUKT. POZYCJA WSPIERA RÓWNIEŻ:</p> <ul style="list-style-type: none"> • Produkt nr 5 – modyfikacja MOST Wiedzy (hostowanie modułów) • Produkt nr 8 – System teleinformatyczny platformy analitycznej • Produkt nr 6, 7 – interfejsy API (hostowanie) <p>KAMIENIE MIŁOWE POWIĄZANE:</p> <ul style="list-style-type: none"> • KM3 (30-11-2027) – przygotowana infrastruktura serwerowa + pilotaż Platformy <p>Pozycja stanowi techniczne zaplecze hostujące Zasoby cyfrowe ISP nr 1 i 2 z OZPI pkt 2.3 (łącznie 6 500 obiektów</p>	1 211 000,00 zł	<p>Wzrost wolumenu danych badawczych wymaga rozbudowy posiadanego storage i mocy obliczeniowej. Środowisko GPU jest niezbędne do działania asystenta AI dla FAIR i wyszukiwania semantycznego w systemie MOST Wiedzy Produkt nr 5). Dodatkowo infrastruktura posłuży jako zasoby obliczeniowe dla Platformy Analitycznej (Produkt 7) do wykorzystania przez interesariuszy. Wykorzystujemy istniejącą infrastrukturę CUI/CI TASK PG jako bazę - minimalizuje to koszty zakupu nowego sprzętu. UWAGA - kryterium 21: udział infrastruktury w kosztach kwalifikowanych aktualnie wynosi ok. 5% (1,21 mln / 25,33 mln) - PONIŻEJ progu 10% (kryterium 21 naboru FERC.02.03-IP.01-005/26 spełnione, 6 pkt).</p>

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>ORD).</p> <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Rozbudowa infrastruktury serwerowej PG/CI TASK (Centrum Informatyczne TASK) - storage klasy PB dla rosnącego repozytorium danych badawczych. • Środowiska obliczeniowe GPU/ AI dla asystenta FAIR Implementation Profile. • Oprogramowanie systemowe i narzędziowe związane z infrastrukturą (zg. wytyczną rozdz. 11 Dobrych praktyk OZPI). 		
Koszty UX i grafiki	<p>PRODUKTY POWIĄZANE (z OZPI 2.4):</p> <ul style="list-style-type: none"> • Produkt nr 11 – Raport z badań UX, w tym zgodności z WCAG 2.2 (06-2029) – POZYCJA BEZPOŚREDNIO REALIZUJE TEN PRODUKT. POZYCJA ZAPEWNIĄ JAKOŚĆ INTERFEJSÓW DLA: • Produkt nr 5 (modyfikacja MOST Wiedzy – interfejsy 	1 274 881,00 zł	<p>Pozycja OBOWIĄZKOWA przy rozwoju oprogramowania (rozdz. 11 Dobrych praktyk OZPI). Pełna implementacja standardu WCAG 2.2 na poziomie AA wymagana ustawą z 04.04.2019 o dostępności cyfrowej. Metodyka User-Centered Design stosowana z sukcesem w projektach MOST Wiedzy i MOST DANYCH. Stawki personelu wg zestawienia średniego wynagrodzenia pracowników PG na analogicznych stanowiskach (aktualne 04-2026); zastosowano mnożnik wynagrodzenia 1,2114 (narzuty pracodawcy ZUS/PPK). W kolejnych latach realizacji zastosowano wzrost 7% r/r (rok 2: $\times 1,07$; rok 3: $\times 1,07^2$) na</p>

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>modułów: deponowania, katalogowania metadanych, panel wnioskowania o dostęp do danych wrażliwych);</p> <ul style="list-style-type: none"> • Produkt nr 8 (frontend Platformy Analitycznej); <p>KAMIENIE MIŁOWE POWIĄZANE:</p> <ul style="list-style-type: none"> • KM7 (30-06-2029) – uzyskane pozytywne raporty z badań UX (w tym zgodności z WCAG 2.2) <p>Pełna implementacja standardu WCAG 2.2 dla wszystkich interfejsów udostępnianych użytkownikom platformy. Główne kategorie kosztów: wynagrodzenia personelu CUI PG (UX Designer, UI Designer, Grafik (DTP / wizualizacja), Ekspert ds. zgodności WCAG) realizujący wewnętrznie projektowanie interfejsów, badania UX z użytkownikami, prototypowanie i testy użyteczności (User-Centered Design); audyt WCAG 2.2 zewnętrzny (etap 1 i 2 – 50 000 zł).</p>		<p>podstawie średniej inflacji GUS 2021-2025. Uwzględniono kwoty dodatkowego wynagrodzenia rocznego.</p>
Bezpieczeństwo	PRODUKTY	1 145 729,00 zł	Pozycja OBOWIĄZKOWA - system

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>POWIĄZANE:</p> <ul style="list-style-type: none"> • Produkt nr 1 – Raport z inicjalnego testu prywatności (01-2027) – POZYCJA BEZPOŚREDNIO REALIZUJE; • Produkt nr 10 – Raport z weryfikacyjnego testu prywatności (02-2029) – POZYCJA BEZPOŚREDNIO REALIZUJE; • Produkt nr 12 – Pozytywny raport z testów bezpieczeństwa (06-2029) – POZYCJA BEZPOŚREDNIO REALIZUJE. <p>POZYCJA WSPIERA RÓWNIEŻ:</p> <ul style="list-style-type: none"> • Produkt nr 5 – modyfikacja MOST Wiedzy (audyty bezpieczeństwa). • Produkt nr 4 – Infrastruktura teleinformatyczna – węzły obliczeniowe GPU/AI, rozszerzenie storage • Produkt nr 6 – Interfejsu API systemu MOST Wiedzy w zakresie maszynowego deponowania danych z aparatury laboratoryjnej • Produkt nr 7 – Interfejs API systemu MOST Wiedzy zgodny ze 		<p>będzie podlegał rygorom KRI (§20 rozporządzenia RM ws. KRI) - wymagany monitoring zgodności i okresowe audyty. Inicjalny test prywatności (02-2027) i weryfikacyjny przed dostarczeniem systemu - obowiązkowe kamienie milowe wg KRMC. Certyfikacja CoreTrustSeal wzmacnia wiarygodność repozytorium w społeczności badawczej. Stawki audytów i certyfikacji na poziomie analogicznym do podobnych zrealizowanych na PG z indeksacją 7%. Stawki personelu wg zestawienia średniego wynagrodzenia pracowników PG na analogicznych stanowiskach (aktualne 04.2026); zastosowano mnożnik wynagrodzenia 1,2114 (narzuty pracodawcy ZUS/PPK). W kolejnych latach realizacji zastosowano wzrost 7% r/r (rok 2: $\times 1,07$; rok 3: $\times 1,07^2$) na podstawie średniej inflacji GUS 2021-2025. Uwzględniono kwoty dodatkowego wynagrodzenia rocznego.</p>

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>standardem Model Context Protocol</p> <ul style="list-style-type: none"> • Produkt nr 8 – Interfejs API systemu MOST <p>Wiedzy zgodny ze standardem Model Context Protocol</p> <p>ZAKRES</p> <p>MERYTORYCZNY:</p> <p>DPIA art. 35 RODO; analiza zgodności z AI Act (opinia EROD 28/2024); zgodność z Data Governance Act.</p> <p>KAMIENIE MIŁOWE</p> <p>POWIĄZANE:</p> <ul style="list-style-type: none"> • KM1 (31-01-2027) – przeprowadzony inicjalny test prywatności i raport; • KM6 (28-02-2029) – przeprowadzony weryfikacyjny test prywatności; • KM7 (30-06-2029) – uzyskane pozytywne raporty z testów bezpieczeństwa. <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Wynagrodzenia personelu: specjalista ds. bezpieczeństwa/ KRI (30 m-cy). • Audyty bezpieczeństwa / testy penetracyjne (3 cykle: rok 1, 2, 3). • Certyfikacja CoreTrustSeal (utrzymanie wymaga certyfikacji) 		

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>repozytorium).</p> <ul style="list-style-type: none"> • Analiza statyczna kodu, testy podatności, ocena DPIA (RODO art. 35). 		
Wydajność rozwiązań	<p>PRODUKTY POWIĄZANE (z OZPI 2.4):</p> <ul style="list-style-type: none"> • Produkt nr 13 – Pozytywny raport z testów wydajności (06-2029) – POZYCJA BEZPOŚREDNIO REALIZUJE POZYCJA ZAPEWNIĄ ODPOWIEDNIĄ WYDAJNOŚĆ DLA: • Produkt nr 8 (System Platforma Analityczna – obliczeniowe workflow CWL/ Snakemake, Apache Spark, Jupyter Hub, SLA dla zapytań); • Produkt nr 5 (modyfikacja MOST Wiedzy – wyszukiwanie facetowe po metadanych, SLA dla deponowania); • Produkt nr 6 (API maszynowego deponowania – SLA dla aparatury laboratoryjnej); • Produkt nr 7 (Serwer MCP – SLA dla zapytań agentów AI). <p>KAMIENIE MIŁOWE POWIĄZANE:</p> <ul style="list-style-type: none"> • KM7 (30-06-2029) – uzyskane pozytywne raporty 	1 118 115,00 zł	<p>Szybkość działania platformy decyduje o jej popularności wśród naukowców oraz o możliwości włączenia jej do europejskich sieci otwartej nauki, dlatego wraz z rosnącą liczbą deponowanych danych konieczne są regularne testy wydajności. Pozycja OBOWIĄZKOWA przy projektach udostępniających informacje/zasoby (rozdz. 11 Dobrych praktyk OZPI). Platforma obsługuje duże wolumeny danych badawczych i zaawansowane zapytania facetowe - krytyczna wydajność.</p> <p>Stawki personelu wg zestawienia średniego wynagrodzenia pracowników PG na analogicznych stanowiskach (aktualne 04.2026); zastosowano mnożnik wynagrodzenia 1,2114 (narzuty pracodawcy ZUS/PPK). W kolejnych latach realizacji zastosowano wzrost 7% r/r (rok 2: $\times 1,07$; rok 3: $\times 1,07^2$) na podstawie średniej inflacji GUS 2021-2025. Uwzględniono kwoty dodatkowego wynagrodzenia rocznego.</p>

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>z testów wydajności</p> <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Wynagrodzenia 3 testerów oprogramowania CUI PG (1 etat × 30 m-cy każdy). • Testy wydajnościowe zewnętrzne (load, stress, soak). • Audyt i optymalizacja indeksów wyszukiwarki (Elasticsearch/ Solr). 		
Szkolenia	<p>Produkty końcowe powiązane (z OZPI pkt 2.4): Produkt nr 9: Materiały szkoleniowe i e-learningowe Centrum Kompetencji Otwartej Nauki 2.0 (Data wdrożenia: 02-2029).</p> <p>Kamień milowy powiązany: Kamień milowy nr 6 (Opracowane materiały szkoleniowe (...)) 28-02-2029).</p> <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Konferencje krajowe i międzynarodowe, wizyty studyjne i szkoleniowe (udział, prelekcje) np. CERN, 	1 100 000,00 zł	<p>Centrum Kompetencji 2.0 - kluczowy element trwałości projektu. Bez podniesienia kompetencji kadry PG i data stewardów efekty projektu nie będą trwałe. Pakiet szkoleniowy stanowiący rozwój i kontynuację komponentu szkoleniowego projektu MOST DANYCH.</p> <p>UWAGA: wynagrodzenia data stewardów (stanowiska merytoryczne realizujące standardy/walidację) wykazane w pozycji nr 1 Oprogramowanie zgodnie z rozdz. 11 Dobrych praktyk.</p>

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>Uniwersytet DELFT, DCC (UK)</p> <ul style="list-style-type: none"> • Utworzenie i realizacja programu szkoleniowego Centrum Kompetencji Otwartej Nauki 2.0 (rozszerzony program dla data stewardów PG). • Szkolenia dla kadry naukowej PG (FAIR, CARE, TRUST, CoreTrustSeal). • Cross-financing - szkolenia wyjazdowe międzynarodowe, konferencje szkoleniowe. • Szkolenia wewnętrzne dla zespołu projektowego (technologie, standardy). • Materiały szkoleniowe i e-learningowe. 		
Działania informacyjno-promocyjne	<p>PRODUKTY POWIĄZANE (z OZPI 2.4):</p> <ul style="list-style-type: none"> • Produkt nr 14 – Materiały informacyjno-promocyjne projektu (06-2029) – POZYCJA BEZPOŚREDNIO REALIZUJE TEN PRODUKT POZYCJA PROMUJE: • Wszystkie Produkty (nr 1-14 z OZPI 2.4); • Zasoby cyfrowe 	550 000,00 zł	<p>Promocja platformy w środowisku naukowym - kluczowa dla osiągnięcia wskaźników korzyści (liczba pobrań, liczba użytkowników). Bez aktywnej promocji repozytorium pozostanie nieznane. Zgodnie z wytycznymi w zakresie informacji i promocji programów spójności 2021-2027. Budżet na promocję na poziomie analogicznym do projektów o podobnej skali np. MOST DANYCH (550 tys. zł), proporcjonalnie do skali projektu.</p>

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>(Zasób/ISP nr 1 i 2 z OZPI 2.3) wśród społeczności naukowej krajowej i europejskiej;</p> <ul style="list-style-type: none"> • Adresacja docelowych użytkowników (osiągnięcie WLWK-RCR011 = 1 000 użytkowników/rok). <p>KAMIENIE MIŁOWE POWIĄZANE:</p> <ul style="list-style-type: none"> • KM7 (30-06-2029) – wytworzone materiały informacyjno-promocyjne projektu. <p>Promocja kluczowa dla osiągnięcia wskaźników korzyści projektu - liczby pobrań, deponowanych zbiorów, zapytań API od zewnętrznych usług oraz adopcji platformy w ekosystemie EOSC.</p> <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Działania informacyjno-promocyjne: kampanie, materiały, media społecznościowe. • Konferencje Open Research Data (po kluczowych etapach projektu) (analogicznie do MOST DANYCH). • Materiały drukowane, 		

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	gadżety, tablice pamiątkowe.		
Koszty zarządzania i wsparcia (w tym wynagrodzenia personelu wspomagającego)	<p>PRODUKTY POWIĄZANE (z OZPI 2.4):</p> <ul style="list-style-type: none"> • Produkt nr 2 – Studium wykonalności (05-2026) – POZYCJA BEZPOŚREDNIO REALIZUJE • Produkt nr 3 – Dokumentacja analityczna projektu (06-2027) – POZYCJA BEZPOŚREDNIO REALIZUJE (w części dot. dokumentacji projektowej i sprawozdań) POZYCJA WSPOMAGA REALIZACJĘ WSZYSTKICH PRODUKTÓW (nr 1-14 z OZPI 2.4) oraz udostępnienie Zasobów cyfrowych (z OZPI 2.3): • Kierownictwo projektu (Kierownik Projektu, Kierownik Techniczny, Owner Produktu). • Koordynacja merytoryczna i techniczna. • Doradztwo prawne (RODO, Data Governance Act, AI Act, otwarte licencje). • Helpdesk dla użytkowników platformy. • Obsługa 	4 476 991,10 zł	<p>Personel merytoryczny został wykazany w innych pozycjach zgodnie z rozdz. 11 Dobrych praktyk OZPI (zakaz wykazywania pracowników merytorycznych w kategorii zarządzania). Funkcja kierownika projektu wzorowana na dobrych praktykach zrealizowanych przez PG projektów. Helpdesk - wsparcie użytkowników (telefoniczne, mailowe), opracowywanie instrukcji. Koszty pośrednie naliczane ryczałtem 15% kwalifikowalnych BEZPOŚREDNICH KOSZTÓW PERSONELU, zgodnie z Wytocznymi dotyczącymi kwalifikowalności wydatków 2021-2027.</p> <p>Stawki personelu wg zestawienia średniego wynagrodzenia pracowników PG na analogicznych stanowiskach (aktualne 04.2026); zastosowano mnożnik wynagrodzenia 1,2114 (narzuty pracodawcy ZUS/PPK). W kolejnych latach realizacji zastosowano wzrost 7% r/r (rok 2: $\times 1,07$; rok 3: $\times 1,07^2$) na podstawie średniej inflacji GUS 2021-2025. Uwzględniono kwoty dodatkowego wynagrodzenia rocznego.</p>

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
<p>administracyjna i sprawozdawczość do CPPC.</p> <ul style="list-style-type: none"> • Koszty pośrednie (15% × personel – art. 54 lit. b rozp. (UE) 2021/1060). <p>KAMIENIE MIŁOWE POWIĄZANE: wszystkie KM1-KM7 (zarządzanie projektowe i sprawozdawczość obejmuje całość realizacji).</p> <p>Pozycja niezbędna dla terminowej i zgodnej z metodyką realizacji projektu (kryterium 17 oceny FERC).</p> <p>Główne kategorie kosztów wykazanych w pozycji:</p> <ul style="list-style-type: none"> • Wynagrodzenia personelu zarządczego i wsparcia (zg. rozdz. 11 Dobrych praktyk OZPI - WYŁĄCZNIE kierownictwo i wsparcie): kierownik projektu (1 etat × 30 m-cy), 2 pracowników Helpdesk (technik informatyk). • Konsultacje merytoryczne ekspertów zewnętrznych (RO-Crate, FAIR, CARE). • Usługi prawne / RODO / open licensing. • Delegacje krajowe i zagraniczne (cross-financing dla zarządczych). 		

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	• Koszty pośrednie 7% kosztów bezpośrednich (ryczałt FERC) - obsługa administracyjna projektu.		

4.3. Koszty ogólne utrzymania wraz ze sposobem finansowania (okres 5 lat)

Całkowity koszt utrzymania trwałości projektu (brutto)	12 076 551,92 zł		Źródło finansowania
Podział całkowitego kosztu utrzymania trwałości projektu na poszczególne lata (netto oraz brutto)	2029	1 050 000,00 zł (brutto) (965 853,66 zł netto)	Środki własne Politechniki Gdańskiej.
	2030	2 173 500,00 zł (brutto) (1 999 317,07 zł netto)	Środki własne Politechniki Gdańskiej.
	2031	2 325 645,00 zł (brutto) (2 139 269,27 zł netto)	Środki własne Politechniki Gdańskiej.
	2032	2 488 440,16 zł (brutto) (2 289 018,13 zł netto)	Środki własne Politechniki Gdańskiej.
	2033	2 662 630,96 zł (brutto) (2 449 249,39 zł netto)	Środki własne Politechniki Gdańskiej.
	2034	1 376 335,80 zł (brutto) (1 266 037,11 zł netto)	Środki własne Politechniki Gdańskiej.

4.4. Planowane koszty ogólne realizacji (w przypadku projektu współfinansowanego – wkład krajowy z budżetu państwa) oraz koszty utrzymania projektu:

- zostaną pokryte w ramach budżetów odpowiednich dysponentów części budżetowych bez konieczności występowania o dodatkowe środki z budżetu państwa
- ~~- będą powodować konieczność przyznania dodatkowych kwot~~

5. GŁÓWNE RYZYKA

5.1. Ryzyka wpływające na realizację projektu

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Przekroczenie harmonogramu realizacji projektu	Średnia	Średnie	<ol style="list-style-type: none"> 1) Kamienie milowe co 5 miesięcy z przeglądem postępu. 2) Bufor czasowy w harmonogramie - odbiór końcowy 3 miesiące przed końcem projektu. 3) Tygodniowy stand-up (krótkie, regularne spotkanie statusowe) zespołu CUI; aktualizacja Rejestru Ryzyk co kwartał. 4) Iteracyjna implementacja (np. Agile/ Scrum) z możliwością priorytetyzacji.
Niezrealizowanie zamówień publicznych w terminie	Średnia	Średnie	<ol style="list-style-type: none"> 1) Wczesne przygotowanie dokumentacji zamówieniowej (równoległe z fazą analityczną, I kwartał). 2) Wykorzystanie niektórych umów ramowych PG na usługi/dostawę IT. 3) Bufor czasowy 4-6 tygodni na proces PZP. 4) Stosowanie kar umownych i procedur ponownego wyboru wykonawcy.
Brak wystarczającego zainteresowania naukowców deponowaniem danych.	Średnia	Niskie	<ol style="list-style-type: none"> 1) Program zachęt finansowych dla naukowców za przygotowanie i zdeponowanie zbiorów danych. 2) Włączenie deponowania danych do polityki otwartego dostępu PG. 3) Szkolenia i warsztaty w ramach Centrum Kompetencji 2.0.
Brak wystarczających środków na realizację projektu (wzrost cen, inflacja)	Duża	Średnie	<ol style="list-style-type: none"> 1) Budżet zindeksowany rocznie wskaźnikiem 7% (na podstawie historycznych danych GUS 2021-2025: 7,62% zaokr. w dół do 7%). 2) Klauzule waloryzacyjne w umowach z dostawcami (sprzęt, licencje, usługi). 3) Zakup kluczowego sprzętu w I roku projektu (zabezpieczenie cen przed dalszymi wzrostami).
Brak możliwości zatrudnienia osób o odpowiednich kompetencjach IT	Duża	Średnie	<ol style="list-style-type: none"> 1) Stawki bazowe zgodne ze średnimi wynagrodzeniami pracowników zatrudnionych aktualnie na PG - aktualne na 04.2026 r. (np. samodzielny programista 9 659 zł, lider techniczny 13 362 zł). 2) Premia regulaminowa 40% gwarantowana w umowie - łączna stawka brutto-brutto na poziomie

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			<p>konkurencyjnym dla rynku IT.</p> <p>3) Wykorzystanie zespołu CUI PG i CI TASK; PG kształci specjalistów IT na wszystkich poziomach.</p> <p>4) Buforowanie ról krytycznych - co najmniej 2 osoby na każdą rolę, elastyczność w wymiarach etatów pracowników oddelegowanych.</p>
Niewystarczająca dostępność kadry projektowej w wymaganych terminach	Duża	Niskie	<p>1) Plan obsadowy sumarycznie ok. 35 etatów na przestrzeni 30 miesięcy zatwierdzony przez Kanclerza PG.</p> <p>2) Współpraca z Biurem Projektów Strategicznych PG - rotacja kadry między projektami.</p> <p>3) Rezerwa na zatrudnienie zewnętrznych ekspertów (250 tys. zł) w sytuacjach awaryjnych.</p>
Nieosiągnięcie wskaźników produktu/celu (5 Star Open Data, % bezpłatnego udostępniania)	Duża	Niskie	<p>1) Plan rezerwy: poziom 4 gwiazdek dla części zasobów; mechanizmy podnoszenia jakości metadanych przez 3 data stewardów dziedzinowych.</p>
Niepełne zidentyfikowanie zależności z innymi projektami i standardami (RDA, EOSC, WDS)	Średnia	Średnie	<p>1) Etap PoC w I roku projektu - weryfikacja integracji z RO-Crate v1.1, CWL, OAI-PMH.</p> <p>2) Współpraca ze społecznością EOSC; udział w grupach roboczych RDA (Research Data Alliance), WDS (World Data Systems).</p> <p>3) Dedykowani eksperci ds. RO-Crate i workflow CWL w zespole projektowym (po 1 etacie x 30 mcy).</p>
Zaostrzenie lub zmiana wymogów regulacyjnych w trakcie realizacji	Średnia	Średnie	<p>1) Monitoring legislacyjny prowadzony przez radcę prawnego (cykliczny przegląd kwartalny).</p> <p>2) Zgodność projektu z Dyrektywą Open Data 2019/1024 i ustawą z 11.08.2021 o otwartych danych.</p> <p>3) Modułowa architektura systemu pozwalająca na szybkie adaptacje do zmian prawnych i mechanizmy konfiguracji polityk licencyjnych.</p>
Cyberatak na platformę (DDoS, próba włamania, ransomware)	Duża	Średnie	<p>1) Zgodność z KRI §20 i polityką bezpieczeństwa PG; hardening serwerów, monitoring SIEM 24/7.</p> <p>2) Zatrudnienie specjalisty ds.</p>

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			bezpieczeństwa/KRI (1 etat × 30 mcy); audyty bezpieczeństwa i testy penetracyjne (rok 1, 2 i 3).
Naruszenie ochrony danych osobowych w trakcie budowy i wdrażania platformy (RODO)	Duża	Niskie	Polityka prywatności, klauzule informacyjne, umowy o powierzeniu przetwarzania (art. 28 RODO) opracowane z radcą prawnym PG i IOD PG; szkolenia RODO dla zespołu projektowego.
Halucynacje i niska jakość odpowiedzi asystenta AI	Średnia	Średnie	Architektura RAG z obowiązkowym cytowaniem źródeł; nadzór człowieka - mapowania metadanych zatwierdza data steward; progi ufności i mechanizm eskalacji do człowieka; cykliczna ewaluacja jakości na zbiorze referencyjnym (golden set).
Dryf modelu AI i degradacja jakości odpowiedzi w czasie / niska jakość źródeł	Średnia	Średnie	Monitoring metryk jakości (skuteczność, % poprawnych cytowań, satysfakcja) z progami alarmowymi; okresowa re-ewaluacja i aktualizacja modeli; wersjonowanie modeli, promptów i danych; kontrola jakości źródeł zasilających RAG.
Brak transparentności i rozliczalności wyników AI oraz ochrona danych w procesie dostrajania modeli	Duża	Niskie	Dokumentacja modeli (model cards) i rejestrowanie wersji modelu/promptów; dostrajanie wyłącznie na danych zanonimizowanych; zgodność z AI Act (system ograniczonego ryzyka) i RODO; nadzór człowieka nad decyzjami; odpowiedzialność za rezultaty po stronie operatora (PG).

5.2. Ryzyka wpływające na utrzymanie efektów

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Brak środków na utrzymanie platformy po zakończeniu projektu.	Duża	Niskie	1) Zabezpieczenie środków w budżecie PG na 5-letni okres trwałości - decyzja Rektora przed podpisaniem umowy o dofinansowanie (uchwała Senatu). 2) Optymalizacja kosztów utrzymania przez wykorzystanie infrastruktury PG 3) Poszukiwanie kolejnych źródeł

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			finansowania (np. Horyzont Europa, EOSC, kolejne nabory FERC). 4) Możliwość partnerstw z innymi uczelniami (model IaaS).
Brak osób o kompetencjach do utrzymania systemu Platformy Analitycznej	Średnia	Niskie	1) Transfer wiedzy w trakcie projektu - dokumentacja techniczna i szkolenia wewnętrzne. 2) Zespół CUI PG posiada wieloletnie doświadczenie w utrzymaniu platformy MOST Wiedzy. 3) Stała współpraca z absolwentami PG. 4) Modułowa architektura - niewielki skład zespołu utrzymaniowego (4-5 osób).
Nieosiągnięcie zaplanowanych korzyści w zakresie adopcji przez naukowców.	Średnia	Średnie	1) Kontynuacja działań Centrum Kompetencji 2.0 po zakończeniu projektu. 2) Integracja z politykami otwartości uczelni i wymogami grantów (Horizon Europe, NCN), oraz polityką naukową MNiSW. 3) Indeksacja w EOSC EU Node i międzynarodowych agregatorach (OpenAIRE, B2FIND, re3Data, Data Citations Index). 4) Monitoring KPI kwartalnie (liczba pobrań, cytowań, użytkowników). 5) Integracja platformy z wewnętrznym systemem zarządzania projektem
Starzenie się technologii i konieczność migracji infrastruktury	Średnia	Średnie	1) Stosowanie technologii open source o szerokim wsparciu społeczności (Java/ JEE LTS, PostgreSQL, Elasticsearch). 2) Architektura modułowa pozwalająca na wymianę komponentów bez przebudowy całego systemu. 3) Rezerwa na drobne prace rozwojowe w utrzymaniu (200 tys. zł/rok); roczne przeglądy stosu technologicznego.
Niezgodność z nowymi standardami (CoreTrustSeal, FAIR, CARE, TRUST)	Średnia	Średnie	1) Certyfikacja CoreTrustSeal w ramach projektu - utrzymanie certyfikatu przez 5 lat trwałości. 2) Stała współpraca z RDA, EOSC, World Data Systems; regularne audyty zgodności (raz na 2 lata). 3) Roczna aktualizacja FAIR Implementation Profiles (FIP) dla wybranych dyscyplin.

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Naruszenie praw autorskich lub przepisów RODO przez deponowane dane badawcze	Duża	Niskie	1) Polityki licencyjne CC i CC0 dla deponowanych zbiorów; weryfikacja praw przez data stewardów przed publikacją; zasady CARE dla danych wrażliwych. 2) Procedury notice-and-takedown; umowa o powierzeniu przetwarzania danych z deponentami. 3) Cykliczne szkolenia dla deponentów i data stewardów z zakresu praw autorskich i RODO (raz w roku).
Awaria, uszkodzenie infrastruktury sprzętowo-programowej w trakcie utrzymania	Średnia	Niskie	1) Redundancja sprzętowa, procedury postępowania z infrastrukturą (zg. doświadczeń MOST DANYCH). 2) Bieżący monitoring infrastruktury sprzętowo-programowej; kwartalne testy DRP (Disaster Recovery Plan). 3) Backupy 3-2-1, plan ciągłości działania (BCP), wsparcie techniczne CI TASK/CUI.
Naruszenie ochrony danych osobowych w okresie eksploatacji platformy (RODO)	Duża	Niskie	Coroczne aktualizacje polityki prywatności, klauzul informacyjnych i umów o powierzeniu przetwarzania (art. 28 RODO) z radcą prawnym PG i IOD PG; procedury reagowania na incydenty bezpieczeństwa danych (notyfikacja UODO w 72h zg. art. 33 RODO); cykliczne szkolenia RODO dla zespołu utrzymaniowego.

6. OTOCZENIE PRAWNE

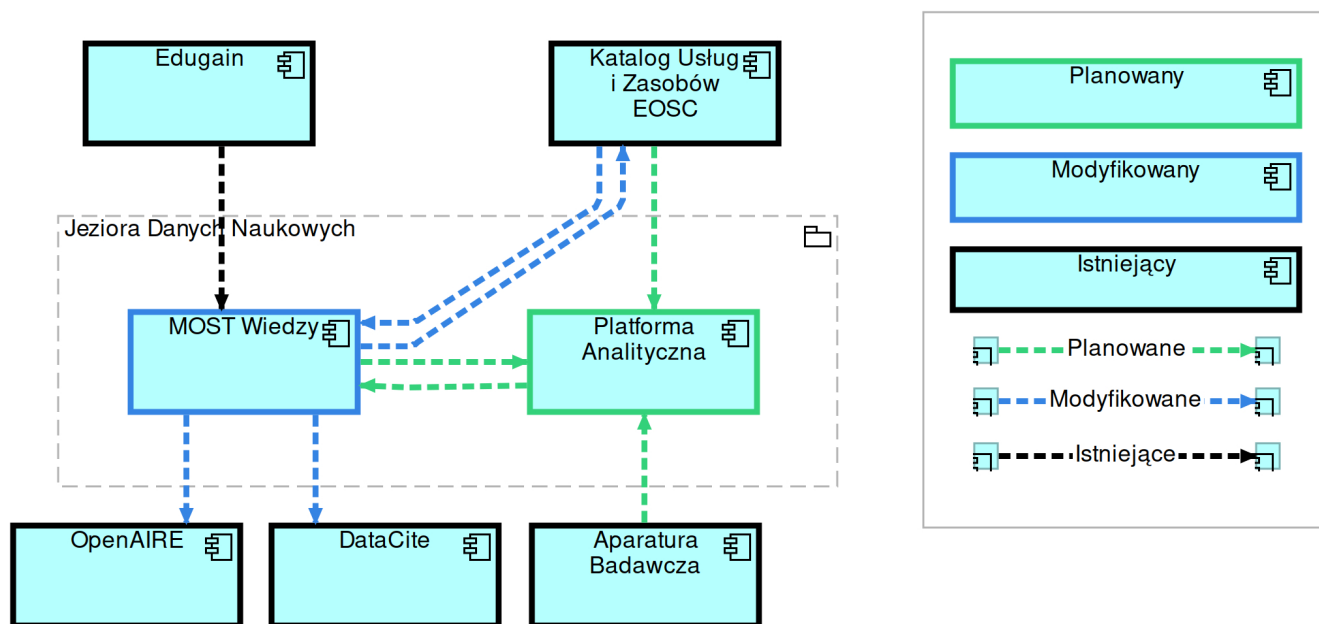
Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
1	Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz.U. 2005 nr 64 poz. 565 z późn. zm.)	TAK/NIE		
2	Rozporządzenie Rady Ministrów z dnia 21 maja 2024 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w	TAK/NIE		

Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
	postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych (Dz.U. 2024 poz. 773)			
3	Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego (Dz.U. 2021 poz. 1641)	TAK/NIE		
4	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 (RODO)	TAK/NIE		
5	Ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa (Dz.U. 2018 poz. 1560 z późn. zm.)	TAK/NIE		
6	Ustawa z dnia 4 kwietnia 2019 r. o dostępności cyfrowej stron internetowych i aplikacji mobilnych podmiotów publicznych (Dz.U. 2019 poz. 848)	TAK/NIE		
7	Ustawa z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1668 z późn. zm.)	TAK/NIE		
8	Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych (Dz.U. 2001 nr 128 poz. 1402 z późn. zm.)	TAK/NIE		
9	Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych (Dz.U. 2018 poz. 1000 z późn. zm.)	TAK/NIE		
10	Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz.U. 1994 nr 24 poz. 83 z późn. zm.)	TAK/NIE		
11	Ustawa z dnia 18 lipca 2002 r. o świadczeniu usług drogą elektroniczną (Dz.U. 2002 nr 144 poz. 1204 z późn. zm.)	TAK/NIE		
12	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2023/2854 z dnia 13 grudnia 2023 r. w sprawie zharmonizowanych przepisów dotyczących sprawiedliwego dostępu do danych i ich wykorzystywania oraz w sprawie zmiany rozporządzenia (UE) 2017/2394 i dyrektywy (UE) 2020/1828 (Akt o danych / Data Act) (Dz.U. UE L 2023/2854)	TAK/NIE		
13	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/868 z dnia 30 maja 2022 r. w sprawie europejskiego zarządzania danymi i zmieniające rozporządzenie (UE) 2018/1724 (Akt w sprawie zarządzania danymi / Data Governance Act) (Dz.U. UE L 152 z 3.6.2022)	TAK/NIE		

Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
14	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji / AI Act) (Dz.U. UE L 2024/1689)	TAK/NIE		

7. ARCHITEKTURA

7.1. Widok kooperacji aplikacji



Lista systemów wykorzystywanych w projekcie

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
1	MOST Wiedzy	Politechnika Gdańska	MOST Wiedzy to serwis udostępniający publikacje naukowe, otwarte dane badawcze (Open Research Data Catalog), profile naukowców i opisy	Modyfikowany	Rozbudowa o: moduł metadanych dziedzinowych (MeSH, IEEE), katalogowanie

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			projektów badawczych. Stworzony w celu zwiększenia dostępności polskiego potencjału naukowego w sieci Internet.		oraz wyszukiwanie z ontologiami dziedzinowymi; rozszerzenie wyszukiwania o tryb semantyczny (RAG) - hybrydowe łączenie istniejącego indeksu pełnotekstowego z warstwą wektorową; Asystent AI FIP, interfejs API, integracja z platformą analityczną, obsługa PID INST (identyfikatory aparatury), rozszerzone modele licencjonowania, komponent etyczny zarządzania danymi.
2	Platforma Analityczna	Politechnika Gdańska	Platforma Analityczna to środowisko chmurowe do przetwarzania i analizy otwartych danych badawczych. Stworzone w celu ułatwienia analizy danych badawczych oraz deponowania ich w serwisie MOST Wiedzy. Wykorzystuje otwarte narzędzia i technologie uruchomiona on-premise na infrastrukturze PG i zintegrowane z systemem MOST Wiedzy. Integracja z oprogramowaniem takimi jak Jupyter Hub ,	Planowany	

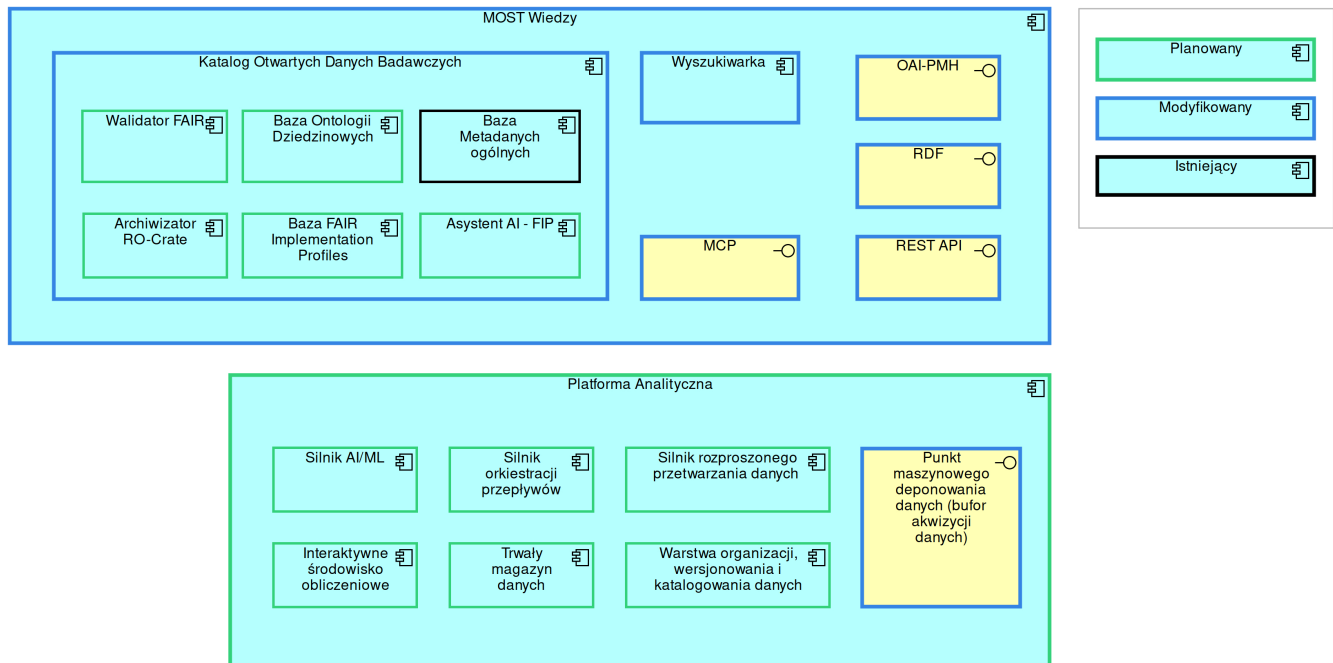
Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			Galaxy lub innymi narzędziami analizy danych. Obsługa wersjonowania i rejestracji workflow. Umożliwia analizę i przetwarzanie danych wykorzystaniem technik ML oraz AI. Eksport i udostępnianie wyników analizy w serwisie MOST Wiedzy.		
3	Usługi EOSC (European Open Science Cloud)	EOSC EU Node	Katalog Usług i Zasobów EOSC to katalog usług oraz zasobów dostępnych w ramach Europejskiej Chmury Otwartej Nauki utrzymywany w ramach węzła EOSC EU Node. Wspiera federację repozytoriów i usług. MOST Wiedzy eksportuje metadane do EOSC. Inne podmioty zrzeszone w ramach EOSC mają dostęp do zasobów z wykorzystaniem tego katalogu.	Istniejący	
4	OpenAIRE	OpenAIRE AMKE	OpenAIRE to Europejska infrastruktura agregująca metadane publikacji i danych badawczych. MOST Wiedzy eksportuje rekordy przez protokół OAI-PMH.	Istniejący	
5	DataCite	DataCite e.V.	DataCite to międzynarodowy rejestr identyfikatorów DOI wspierający cytowanie oraz wykorzystywanie danych badawczych. MOST Wiedzy nadaje DOI wszystkim zdeponowanym zbiorom danych z wykorzystaniem integracji z DataCite.	Istniejący	
6	Edugain	GEANT	Edugain to federacyjny system uwierzytelniania	Istniejący	

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			dla instytucji akademickich. Zapewnia logowanie użytkowników spoza PG do platformy.		
7	Aparatura Badawcza	Politechnika Gdańska	<p>Aparatura badawcza to ogół czujników i przyrządów pomiarowych służących do rejestrowania i przetwarzania wielkości fizycznych w trakcie badań dostępnych w laboratoriach Politechniki. Z perspektywy architektury aplikacji nie jest częścią projektowanej platformy, lecz systemem zewnętrznym leżącym poza granicą systemu i poza naszą kontrolą. Ze względu na heterogeniczny charakter aparatury (różne protokoły, formaty, część bez cyfrowego wyjścia), nie jest możliwa bezpośrednia integracja w każdym przypadku, konieczne jest stworzenie warstwy adaptacyjnej (bramy), jednakże ze względu na wspomnianą mnogość protokołów i formatów jak i również lokalizacji (część czujników będzie wymagać użycia sieci GSM, lub LoRa oraz pośredników o niskim zużyciu energii) jest to zadanie sędowane na gestorów tych urządzeń i poza przykładami demonstracyjnymi nie stanowi zakresu tego projektu.</p>	Istniejący	

Lista przepływów

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
1	EduGAIN	MOST Wiedzy	Dane uwierzytelniające	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	Protokół SAML
2	MOST Wiedzy	OpenAIRE	Metadane danych badawczych	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	Protokół OAI-PMH
3	MOST Wiedzy	DataCite	Metadane danych badawczych	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	Usługa REST (HTTP)
4	Usługa REST (HTTP)	Platforma Analityczna	Dane badawcze wraz z metadanymi	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	HTTP, protokół S3
5	MOST Wiedzy	Katalog Usług i Zasobów EOSC	Dane badawcze wraz z metadanymi	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	Protokół OAI-PMH, HTTP, protokół S3
6	Katalog Usług i Zasobów EOSC	Platforma Analityczna	Dane badawcze i/ lub metadane	Kopiowanie danych (§13 ust. 3)	Realizowalny inną metodą	Różne (wymaga doprecyzowania)
7	Platforma Analityczna	MOST Wiedzy	Nowe dane badawcze	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	HTTP, protokół S3
8	Aparatura Badawcza	Platforma Analityczna	Dane pomiarowe	Kopiowanie danych (§13 ust. 3)	Krytyczny dla sukcesu projektu	Usługa REST (HTTP), protokół MQTT, protokół AMQP lub inne

7.2. Kluczowe komponenty architektury rozwiązania



7.3. Przyjęte założenia technologiczne

Lp.	Obszar	Założenie technologiczne
1.	Infrastruktura	Infrastruktura teletechniczna serwerowni podstawowej spełniająca wymagania co najmniej standardu TIER III/IV. Środowisko uruchomieniowe oparte o zwirtualizowaną, prywatną chmurę usług z orkiestracją kontenerów, wdrożenia i konfiguracja w modelu Infrastructure-as-Code (ansible, terraform). Zasoby obliczeniowe alokowane na żądanie (ad hoc), ze skalowaniem automatycznym zależnym od obciążenia.
2.	Sieć i bezpieczeństwo	Backup i przywracanie: regularne kopie zapasowe (pełne i przyrostowe), wersjonowanie danych umożliwiające powtarzalność i wycofywanie zmian. Szyfrowanie: szyfrowanie transmisji w oparciu o uznane standardy kryptograficzne oraz szyfrowanie danych wrażliwych w spoczynku. Kontrola dostępu: szczegółowa kontrola dostępu na poziomie ról oraz uwierzytelnianie każdego urządzenia i usługi (mechanizmy oparte o certyfikaty/klucze) z zasadą najmniejszych uprawnień. Komunikacja publiczna szyfrowana przy użyciu certyfikatów. Segmentacja sieci: logiczny podział na strefy i wzajemna izolacja komponentów. Ochrona styku z siecią zewnętrzną i kontrolowana brama wejściowa. Kopie zapasowe przechowywane w odrębnej lokalizacji geograficznej.
3.	Standardy wymiany danych	Protokoły MQTT, HTTP/REST lub gRPC (w zależności od potrzeb), AMQP, API dostępu do danych zgodne z S3, SAML, OAI-PMH.
4.	Systemy operacyjne	Otwarte, serwerowe systemy operacyjne wspierające

Lp.	Obszar	Założenie technologiczne
	serwerowe	wirtualizację i konteneryzację (przykładowo Linux Debian). Po stronie urządzeń brzegowych dopuszcza się systemy klasy komputerów jednopłytkowych (przykładowo Raspberry Pi OS) oraz dedykowane oprogramowanie układowe mikrokontrolerów. Szczegółowe wersje oraz utwardzenie (hardening) zostaną doprecyzowane na etapie realizacji Projektu.
5.	Bazy danych	Warstwa danych obejmuje bazę relacyjną (SQL) oraz bazy nierelacyjne (noSQL), w tym bazę zoptymalizowaną pod dane szeregów czasowych, magazyn obiektowy dla danych, a także bazy wektorowe do wyszukiwania semantycznego. Wszystkie pracują w trybie wysokiej dostępności (HA).
6.	Serwery aplikacji	Usługi aplikacyjne uruchamiane jako konteneryzowane, skalowalne usługi, z równoważeniem obciążenia. Warstwa obejmuje usługi pośredniczące i brokery komunikacyjne, obsługę sieci radiowej dalekiego zasięgu oraz wysokowydajne klastry obliczeniowe (przetwarzanie rozproszone i zadania uczenia maszynowego) alokowane ad hoc.
7.	Portale	Dla Platformy Analitycznej i Mostu Wiedzy dostęp poprzez przeglądarki internetowe (m. In. Chrome i Firefox) na platformach Desktop (Windows, Mac, Linux, inne) jak i Mobilnych (Android, iOS, inne).
8.	Inne	Komponenty platformy oparte wyłącznie o rozwiązania o otwartym kodzie źródłowym (open source). Mechanizmy pośredniczące (bramy) akwizycji danych realizowane na urządzeniach klasy komputerów (w tym jednopłytkowych) oraz mikrokontrolerów, łączność lokalna (przewodowa/bezprzewodowa) i zdalna (krótkiego zasięgu oraz radiowa dalekiego zasięgu -LoRa). Dane utrwalane w otwartych, niezastrzeżonych formatach. Repozytorium kodu (zarówno platformy jak i IoC) realizowane poprzez system kontroli wersji git. Realizacja funkcji wykorzystujących ML/AI w oparciu o istniejące otwarte modele uruchamiane na infrastrukturze własnej Politechniki Gdanskiej. Wykorzystane modele AI nie będą poddawane dodatkowemu wstępnemu trenowaniu; możliwe dodatkowe dotrenowanie modeli w późniejszych etapach projektu na podstawie wstępnie zgromadzonych danych. W celu ochrony prywatności i bezpieczeństwa danych rozwiązania AI będą uruchamiane jedynie na własnej infrastrukturze.

7.4. Opis zasobów danych przetwarzanych w planowanym rozwiązaniu

Czy nowy system będzie tworzył zasoby danych o charakterze rejestru publicznego?

TAK/NIE

Czy nowy system będzie przetwarzał (używał, zmieniał) zawartość innych rejestrów publicznych?

TAK/NIE

7.5. Bezpieczeństwo

Planowany poziom zapewnienia bezpieczeństwa (w rozumieniu przepisów §20 rozporządzenia Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności [...] (Dz. U. 2012, poz. 526 z późn. zm.) w zakresie dot. systemu zarządzania bezpieczeństwem informacji:

- ~~-system nie podlega rygorom KRI – należy wyjaśnić czy istnieją inne normy bezpieczeństwa, które będą spełnione przez system zgodnie z wymogami KRI~~
- ~~-dodatkowe zabezpieczenia powyżej wymogów KRI: należy wskazać uzasadnienie~~