

OPIS ZAŁOŻEŃ PROJEKTU INFORMATYCZNEGO

Tytuł projektu	Digitalizacja cennych zasobów nauki i udostępnienie danych genomicznych z wykorzystaniem infrastruktury Polskiego Węzła Europejskiego Archiwum Genomów		
Wnioskodawca	Minister Nauki i Szkolnictwa wyższego		
Beneficjent	Uniwersytet Łódzki		
Partnerzy	Pomorski Uniwersytet Medyczny w Szczecinie Instytut Chemii Bioorganicznej Polskiej Akademii Nauk - Poznańskie Centrum Superkomputerowo-Sieciowe Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie, Akademickie Centrum Komputerowe Cyfronet AGH		
Źródło finansowania	FERC, Priorytet FERC.02 Zaawansowane usługi cyfrowe, Działanie 02.03 Cyfrowa dostępność i ponowne wykorzystanie informacji; Budżet Państwa - część 27 - Informatyzacja		
Całkowity koszt projektu	30 579 170,00 zł		
Planowany okres realizacji projektu	01-2027 do 12-2029		
Osoba kontaktowa	Błażej Marciniak	blazej.marciniak@biol.uni.lodz.pl	600936417

1. POWODY PODJĘCIA PROJEKTU

1.1. Identyfikacja problemu i potrzeb

Planowane przedsięwzięcie odpowiada na rozbieżność pomiędzy obecnym stanem udostępniania zasobów poprzez Infrastrukturę Polskiego Węzła Europejskiego Archiwum Genomów (PL-FEGA). Obecnie zasoby nauki przechowywane są w postaci analogowej - próbek ludzkiego materiału biologicznego (krew, ślina, DNA itp.) (dalej zwane LMB) co uniemożliwia ich efektywne udostępnianie i wykorzystanie w badaniach naukowych w formie cyfrowej.

Dostęp do zasobów analogowych jest ograniczony a proces digitalizacji kosztowny i czasochłonny.

Transport LMB wymaga dochowania najwyższych standardów jakości i zabezpieczenia w trakcie transportu (czas, temperatura) aby nie uszkodzić cennych zbiorów.

Jedynie niewielka część (poniżej 10%) zasobów UŁ została zdigitalizowana i udostępniona w wersji cyfrowej

Obecnie funkcjonująca infrastruktura PL-FEGA hostowane jest przez Beneficjenta - pojedynczą instytucję co ogranicza zasoby oraz redundancję geograficzną do obrębu kampusu. Z uwagi na duże wolumeny ograniczeniem do pełnego wykorzystania danych są transfery pomiędzy PL-FEGA a lokalizacjami Interesariuszy oraz zasoby jakie należy przeznaczyć do analizy danych.

Polska jest jedynym krajem, który sygnował inicjatywy Miliona Genomów Europejskich (1+MG) co znacznie ogranicza dostęp polskim jednostkom naukowym do Europejskiej Infrastruktury Danych Genomicznych (GDI), również tej planowanej do wykorzystania w ramach EHDS.

Projekt zakłada unowocześnienie infrastruktury PL-FEGA do PL-FEGA 2.0, zwiększenie niezawodności i skalowalności poprzez rozproszenie infrastruktury pomiędzy 3 ośrodki. Integrację PL-FEGA z przestrzeniami analitycznymi RE minimalizując ruch sieciowy. Wdrożenie systemów IT pozwalających na optymalizację procesu zarządzania/katalogowania LMB. Dostarczenie dużego pakietu danych genomicznych i epigenomicznych. Uruchomienie RE dostępnych dla Interesariuszy na potrzeby prowadzenia analiz danych pochodzących z PL-FEGA 2.0. Taka infrastruktura mogła by zostać użyta do obsługi EHDS.

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>Pracownicy naukowci * sposób wykorzystania: - użycie RE do prowadzenia analiz danych genomicznych (wymagających szczególnej ochrony zgodnie z RODO) - wykorzystanie zdigitalizowanych danych do prowadzenia badań naukowych/ podstawowych w dziedzinie Genetyki i Epigenetyki populacji polskiej lub medycynie spersonalizowanej - wykorzystanie zdigitalizowanych danych na potrzeby prowadzenia zajęć edukacyjnych - użycie PL-FEGA 2.0 do deponowania i udostępniania w sposób bezpieczny i kontrolowany danych genomicznych w otwartym dostępie wygenerowanych w innych projektach naukowych zgodnie z wymogami redakcji lub sponsorów projektów np. NCN *sposób udostępniania - udostępnianie odbywać się będzie na podstawie aplikacji</p>	<p>- brak dostępnych w otwartym dostępie/ bezpłatnych wysokiej jakości danych genomicznych i epigenomicznych umożliwiających prowadzenie badań podstawowych w dziedzinie Genetyki i Epigenetyki populacji polskiej lub medycynie spersonalizowanej - ograniczona dostępność środowisk analitycznych (RE) dostępnych przy repozytoriach danych - niewystarczająca liczba centrów eksperckich wspierających procesy digitalizacji tego typu danych w Polsce - generowanie nowych danych podnosi koszty i wydłuża czas badań - brak repozytoriów pozwalających udostępnić dane genomiczne (wg. RODO wymagające szczególnej ochrony) w otwartym dostępie w sposób bezpieczny; otwarty dostęp wymagany jest na etapie publikacji wyników w czasopiśmie naukowych; otwarty dostęp jest też warunkiem rozliczenia projektów finansowanych z środków publicznych np. NCN</p>	<p>Pracownicy naukowci: 188 426 osób</p>

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>o dostęp do danych lub RE. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych (DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <ul style="list-style-type: none"> - dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane. - dostęp do RE odbywał się będzie na podstawie aplikacji o dostęp do środowiska. - wykorzystanie PL-FEGA 2.0 do deponowania i udostępniania danych rozliczane będzie umowami na korzystania z platformy oraz umowy powierzenia przetwarzania danych 		
<p>Personel B+R</p> <p>*sposób wykorzystania:</p> <ul style="list-style-type: none"> - wykorzystanie gotowych - 	<ul style="list-style-type: none"> - niewystarczająca liczba centrów eksperckich wspierających procesy digitalizacji tego typu danych w Polsce - brak dostępnych w otwartym dostępie/ darmowych wysokiej jakości danych 	305 000 osób

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>zdigitalizowanych danych do prowadzenia badań naukowych/ podstawowych w dziedzinie Genetyki i Epigenetyki populacji polskiej lub medycynie spersonalizowanej</p> <ul style="list-style-type: none"> - walidacja modeli/ metod/wyników w oparciu o istniejące dane gnomiczne <p>*sposób udostępniania - udostępnianie odbywać się będzie na podstawie aplikacji o dostęp do danych. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych (DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <ul style="list-style-type: none"> - dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane. 	<p>umożliwiających prowadzenie badań podstawowych w dziedzinie Genetyki populacji polskiej lub medycynie spersonalizowanej</p> <ul style="list-style-type: none"> - generowanie nowych danych podnosi koszty i wydłuża czas badań 	
Lekarze	Brak skalibrowanych danych dotyczących	166 515 osób

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>sposób wykorzystania:</p> <ul style="list-style-type: none"> - użycie gotowych danych genomicznych dotyczących polskiej populacji w celu weryfikacji doniesień naukowych - wykorzystanie danych i RE do opracowywania lub kalibracji statystyk wariantów DNA występujących w populacji polskiej - weryfikacja w oparciu o zasoby PL-FEGA 2.0 (RE i dane) opracowań naukowych bazujących na populacjach z innym rozkładem wariantów niż populacja polska <p>*sposób udostępniania - udostępnianie odbywać się będzie na podstawie aplikacji o dostęp do danych lub RE. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych (DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <ul style="list-style-type: none"> - dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, 	<p>genetycznego ryzyka wystąpienia jednostki chorobowej dla polskiej populacji</p> <p>Brak danych pozwalających na weryfikację doniesień naukowych w odniesieniu do polskiej populacji</p>	

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane.</p> <p>- dostęp do RE odbywał się będzie na podstawie aplikacji o dostęp do środowiska.</p> <p>- wykorzystanie PL-FEGA 2.0 do deponowania i udostępniania danych rozliczane będzie umowami na korzystania z platformy oraz umowy powierzenia przetwarzania danych</p>		
<p>Przemysł farmaceutyczny</p> <p>*sposób wykorzystania:</p> <p>- dostęp do danych genomicznych w celu weryfikacji doniesień naukowych z badań prowadzonych w innych rejonów świata lub w oparciu o dane z innych rejonów świata w celu określenia aplikowalności w polskiej populacji</p> <p>- dostęp do danych genomicznych w celu oceny potencjalnych skutków ubocznych współczesnych zaawansowanych terapii.</p> <p>*sposób udostępniania - udostępnianie odbywać się będzie na podstawie aplikacji</p>	<p>brak dostępnych w otwartym dostępie wysokiej jakości danych umożliwiających prowadzenie badań</p> <p>brak danych służących do budowy modeli oddziaływania leków w populacji polskiej. niewystarczająca liczba centrów eksperckich wspierających procesy digitalizacji tego typu danych w Polsce</p> <p>Innowacyjne terapie opracowywane w innych rejonach świata dla populacji o innym rozkładzie wariantów genetycznych mogą okazać się mniej skuteczne dla polskich pacjentów; podobnie rzecz się ma, jeśli chodzi o skutki uboczne terapii - posiadając takie dane możliwa będzie dokładniejsza prognoza oddziaływania/skuteczności danej terapii.</p> <p>Posiadanie odpowiedniej ilości wysokiej jakości danych genomicznych o polskiej populacji pozwoli na wykorzystanie tej wiedzy przy projektowaniu nowoczesnych terapii przez krajowe podmioty</p>	697 podmiotów

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>o dostęp do danych. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych (DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <p>- dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane.</p>		
<p>Branża AI</p> <p>*sposób wykorzystania:</p> <p>- użycie danych zgromadzonych w FEGA 2.0 jako zbiorów danych testowych lub uczących dla opracowywanych modeli.</p> <p>*sposób udostępniania - udostępnianie odbywać się będzie na podstawie aplikacji o dostęp do danych. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych</p>	<p>Brak danych dokładnych danych genomicznych opisujących polską populację gotowych do wykorzystania w przygotowaniu algorytmów AI. Dostarczenie odpowiednich danych pozwoli na przygotowanie bardziej wiarygodnych modeli AI, mogących wesprzeć lekarzy lub system ochrony zdrowia, Dane te mogą stanowić zarówno komponent treningowy zwiększając różnorodność i reprezentacyjność danych wykorzystywanych do treningu modeli AI, jak również niezależny zbiór walidacyjny pozwalający na wiarygodną ewaluację modeli AI</p>	<p>13 000 podmiotów</p>

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>(DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <ul style="list-style-type: none"> - dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane. 		
<p>System ochrony zdrowia</p> <p>*Sposób wykorzystania:</p> <ul style="list-style-type: none"> - użycie danych genomicznych zgromadzonych w PL-FEGA 2.0 do planowania polityki ochrony zdrowia - wykorzystanie RE do analizy danych wrażliwych - wykorzystanie infrastruktury PL-FEGA 2.0 do deponowania danych genomicznych pochodzących badań diagnostycznych zgodnie z założeniami EHDS <p>*sposób udostępniania - udostępnianie</p>	<p>Brak danych dokładnych danych genomicznych opisujących polską populację.</p> <p>Statystyki wariantów genetycznych występujących w populacji mogą stać się cennym narzędziem w planowaniu profilaktyki</p> <ul style="list-style-type: none"> - dla przykładu znajomość częstości występowania i rozkładu występowania wariantów predysponujących do ostrego przebiegu COVID pozwoliłaby odpowiednio zaplanować rozmieszczenie miejsc hospitalizacji; <p>Podobnie w przypadku występowania zwiększonego ryzyka poligenicznego wystąpienia danej jednostki chorobowej na jakimś obszarze pozwoli na dobór odpowiednich działań profilaktycznych,</p> <p>Ponadto znajomość częstości występowania wariantów pozwala określić aplikacyjność testów molekularnych przed ich wdrożeniem np. test genetyczny szacujący predyspozycję do danej jednostki chorobowej mający zastosowanie w innych populacjach może być</p>	<p>305 620 osób</p>

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
<p>odbywać się będzie na podstawie aplikacji o dostęp do danych lub RE. Aplikacja rozpatrywana będzie przez Komitet Dostępu do Danych (DAC). Na podstawie oceny zaprezentowanego celu, zgodności ze zgodami dawców i jakości badań DAC podejmuje decyzję o przyznaniu dostępu do zasobów.</p> <p>*rozliczalność</p> <ul style="list-style-type: none"> - dostęp do zbioru danych poprzedzony jest podpisaniem umowy Transferu danych (DTA), które reguluje m.in. zasady bezpieczeństwa, odpowiedzialności za dane, sposób wykorzystania danych oraz prawa własności intelektualnej wyników/dzieł powstałych w oparciu o dane. - dostęp do RE odbywał się będzie na podstawie aplikacji o dostęp do środowiska. - wykorzystanie PL-FEGA 2.0 do deponowania i udostępniania danych rozliczane będzie umowami na korzystania z platformy oraz umowy powierzenia przetwarzania danych 	<p>nieskuteczny w populacji PL ze względu na niską częstość występowania danego wariantu</p>	

1.2. Opis stanu obecnego

digitalizacja LMB odbywa się w trakcie procesu sekwencjonowania - generowane są cyfrowe zapisy odczytanych sekwencji DNA. Następnie w trakcie analiz bioinformatycznych generowany jest produkt docelowy - sekwencja pełnego genomu WGS. PL-FEGA umożliwi publikację WGS w otwartym dostępie w sposób bezpieczny i zgodny z RODO. PL-FEGA zintegrowana jest z Centralnym węzłem EGA (CEGA) za pomocą interfejsów API i brokerów wiadomości RabbitMQ. Wrażliwe dane dotyczące polskiej populacji przechowywane są w PL-FEGA w formie szyfrogramu algorytm Crypt4GH. Zarządzanie użytkownikami i przechowywanie niewrażliwych w rozumieniu RODO metadanych realizowane jest w CEGA, odpowiedzialnym za dostarczenie oprogramowania do zarządzania dostępem do danych - Portal Komitatu Dostępu do Danych (DAC Portal). PL-FEGA jest dostępna produkcyjnie od końca 2023 roku. Skuteczność infrastruktury może potwierdzić historia pobrań danych a bezpieczeństwo przeprowadzone w 2025 roku testy penetracyjne. Obecnie dane stanowią podstawę rozwoju nowoczesnej gospodarki, dane genomiczne nowoczesnej medycyny i profilaktyki zdrowia. UE zwiększa innowacyjności (raport Draghiego). W tym obszarze uruchamia projekty GDI, "Genome of Europe" (GoE) - jako część większego programu 1+MG, którego cel to milion WGS mieszkańców Europy. Zgodnie z dokumentacją GoE aby należycie opisać polską populację należy dysponować 50-70 tys. Genomów. Obecnie w otwartym dostępie z wykorzystaniem PL-FEGA udostępniono 400 WGS. Dane wygenerowane w projektach naukowych nie stają się publicznie dostępne. Dla porównania UK dysponuje 500 tys a ZEA 800 tys WGS. W Polsce równolegle trwa projekt G4PL produkujący również dane genomiczne - projekt zawiera komponent komercyjny i inny zakres danych i wytworzenie narzędzi i usług. Choć obie inicjatywy się dopełniają, w tym projekcie kładziony jest nacisk na nową technologię i unikatowe obecnie dane WGS+Epigenom.

2. EFEKTY PROJEKTU

2.1. Cele i korzyści wynikające z projektu

Cel - 1	PL-FEGA 2.0 - Utworzenie rozproszonej i skalowalnej infrastruktury IT pozwalającej na udostępnienie danych o Ludzkim DNA z terytorium RP w bezpieczny sposób
Cel strategiczny	<p>Polityka naukowa Państwa – budowa infrastruktury udostępniania danych (repozytoria)</p> <p>- Krajowej Strategii Rozwoju Regionalnego 2030 - wzmacnianie potencjałów badawczych poszczególnych regionów.</p> <p>Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2025/327 z dnia 11 lutego 2025 r. w sprawie europejskiej przestrzeni danych dotyczących zdrowia oraz zmiany dyrektywy 2011/24/UE i rozporządzenia (UE) 2024/2847 - EHDS - ponowne (secondary use) wykorzystanie danych</p>
Korzyść:	<p>Budowa gospodarki opartej na wiedzy</p> <p>Skalowalna infrastruktura udostępniania danych</p> <p>Obniżenie kosztów przechowywania i zarządzania danymi w poszczególnych jednostkach</p>
KPI:	<p>1 - Instytucje publiczne otrzymujące wsparcie na opracowywanie usług, produktów i procesów cyfrowych</p> <p>2 - Liczba podmiotów wspartych w zakresie rozwoju usług, produktów i procesów cyfrowych</p> <p>3 - Liczba platform udostępniania dokumentów zawierających informacje sektora publicznego/dane prywatne</p> <p>4 - Liczba uruchomionych systemów teleinformatycznych</p> <p>5 - Użytkownicy nowych i zmodernizowanych publicznych usług, produktów i</p>

	procesów cyfrowych 6(?) - ilość pobrań dokumentów/rok
Wartość aktualna i docelowa KPI:	<p>KPI 1: - wartość aktualna: 0 - wartość docelowa: 1</p> <p>KPI 2: - wartość aktualna: 0 - wartość docelowa: 4</p> <p>KPI 3: - wartość aktualna: 0 - wartość docelowa: 1</p> <p>KPI 4: - wartość aktualna: 0 - wartość docelowa: 2</p> <p>KPI 5: - wartość aktualna: 0 - wartość docelowa: 10</p> <p>KPI 6: - wartość aktualna: 0 - wartość docelowa: 4 800</p>
Metoda pomiaru KPI	<p>KPI 1: Pomiar na podstawie zawartych umów o dofinansowanie projektu. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</p> <p>KPI 2: Pomiar na podstawie zawartych umów o dofinansowanie projektu i umów partnerstwa. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</p> <p>KPI 3: Pomiar na podstawie protokołów uruchomienia/oddania do użytku zmodyfikowanej platformy PL-FEGA 2.0. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</p> <p>KPI 4: Pomiar na podstawie protokołów uruchomienia/oddania do użytku systemów IT. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</p> <p>KPI 5: Pomiar na podstawie zawartych umów DTA - udostępnienia danych zgromadzonych w PL-FEGA 2.0. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub 12 miesięcy od zakończenia projektu</p>
Cel - 2	Digitalizacja i udostępnienie zasobów nauki, sektorów B+R i AI, systemu ochrony zdrowia i lekarzy
Cel strategiczny	<p>Polityka naukowa Państwa: otwarte dane, otwarta nauka</p> <p>Polityka rozwoju sztucznej inteligencji w Polsce do 2030r.:dostęp do danych</p> <p>Krajowa Strategia Rozwoju Regionalnego 2030: wzmocnienie potencjałów badawczych poszczególnych regionów.</p>

	<p>Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2025/327 z dnia 11 lutego 2025 r. w sprawie europejskiej przestrzeni danych dotyczących zdrowia oraz zmiany dyrektywy 2011/24/UE i rozporządzenia (UE) 2024/2847 – EHDS: ponowne (secondary use) wykorzystanie danych</p> <p>Zdrowa Przyszłość. Ramy strategiczne rozwoju systemu ochrony zdrowia na lata 2021–2027, z perspektywą do 2030 r.: rozwój nowoczesnego i efektywnego systemu ochrony zdrowia.</p> <p>Strategia na rzecz Odpowiedzialnego Rozwoju do roku 2020, z perspektywą do 2030 r.: skuteczne państwo i instytucje, w tym cyfrowy rozwój kraju.</p>
Korzyść:	<p>Budowa gospodarki opartej na wiedzy, Zwiększenie dostępności wysokiej jakości danych na potrzeby interesariuszy Zwiększenie potencjału badawczego polskich uczelni, obniżenie kosztów prowadzenia prac badawczo rozwojowych, badań podstawowych, możliwość doboru skuteczniejszych leków/terapii, rozwój biotechnologii, zwiększenie dostępności danych służące rozwojowi sztucznej inteligencji w obszarze zdrowia/medycyny</p>
KPI:	<p>7 - Rozmiar zdigitalizowanych informacji sektora publicznego/danych prywatnych [TB]; 8 - Rozmiar udostępnionych on-line informacji sektora publicznego/danych prywatnych [TB]; 9 - Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne [szt.]; 10 - Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne [szt.]</p>
Wartość aktualna i docelowa KPI:	<p>KPI 7: - wartość aktualna: 0 - wartość docelowa: 69 [TB] KPI 8: - wartość aktualna: 0 - wartość docelowa: 69 [TB] KPI 9: - wartość aktualna: 0 - wartość docelowa: 1 600 KPI 10: - wartość aktualna: 0 - wartość docelowa: 1 600 .</p>
Metoda pomiaru KPI	<p>KPI 7: Pomiar na podstawie protokołów odbioru produktów projektu. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu KPI 8: Pomiar na podstawie raportów publikacji dokumentów. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu KPI 9: Pomiar na podstawie protokołów odbioru produktów projektu. Narastająco</p>

	<p>zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</p> <p>KPI 10: Pomiar na podstawie raportów publikacji dokumentów. Narastająco zgodnie z harmonogramem DIP, Pomiar końcowy po osiągnięciu wskaźnika (na podstawie pomiarów częściowych) lub na koniec projektu</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)
<p>Kolekcja materiału biologicznego Generatio-A jest to kolekcja zebrana (jest w posiadaniu Beneficjenta) przy okazji realizowanego badania populacyjnego mieszkańców miasta Aleksandrów Łódzki. Kolekcja jest jedną z nielicznych w Polsce kolekcji rodzinnych. Dzięki czemu możliwe jest prowadzenie badań nad dziedziczeniem poszczególnych genów. Z uwagi na planowaną unikatową jeszcze w skali kraju i europy metodę digitalizacji zasobów Genom + Epigenom, możliwe będzie prowadzenie badań powiązanych z chorobami neurodegeneracyjnymi, badanie interakcji pomiędzy genomem i epigenomem, zasób będzie nieoceniony jako niezależny zestaw danych walidacyjnych dla</p>	30-06-2028	400 Genomów (WGS) - zasoby przeznaczone do zdigitalizowania i udostępnienia

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
innych badań naukowych.		
Kolekcja Materiału biologicznego POPULOUS (jest w posiadaniu Beneficjenta). Digitalizacji podlegać będzie fragment jednej z większych kolekcji materiału biologicznego w kraju. Kolekcja odzwierciedla strukturę zasiedlenia kraju z uwzględnieniem podziału na województwa, tereny miejskie, wiejskie i duże aglomeracje.	31-12-2029	1000 Genomów (WGS) - zasoby przeznaczone do zdigitalizowania i udostępnienia
Onkologiczna kolekcja LMB - istniejący zasób PUM. Kolekcja zawiera próbki materiału biologicznego pobranego od pacjentów onkologicznych w formie zdrowa tkanka i tkanka nowotworowa - sekwencjonowanie obu materiałów. Zdigitalizowana para zasobów tworzy jeden dokument.	30-06-2029	200 WGS zasoby przeznaczone do zdigitalizowania i udostępnienia

Czy wszystkie zdigitalizowane zasoby objęte projektem będą udostępniane bezpłatnie?
TAK/NIE

2.4. Produkty końcowe projektu

Nazwa produktu	Planowana data wdrożenia
Rozproszona infrastruktura Polskiego Węzła Europejskiego Archiwum Genomów PL-FEGA 2.0	12-2029
BIMSAR (Nazwa robocza)- system informatyczny służący katalogowaniu i zarządzaniu analogowymi zasobami poddawany procesowi digitalizacji. System będzie wykorzystywał elementy rozszerzonej rzeczywistości AR dla lepszej ergonomii pracy i usprawnienia procesu digitalizacji. System będzie udostępniany nieodpłatnie tym jednostkom, które będą chciały deponować dane na platformie PL-FEGA 2.0.	06-2028
Raport z zewnętrznych testów bezpieczeństwa systemu BIMSAR	07-2028
Raport z zewnętrznych testów bezpieczeństwa PL-FEGA 2.0	02-2028
Uaktualniona platforma PL-FEGA 2.0	01-2028

Nazwa produktu	Planowana data wdrożenia
Materiały promocyjne i szkoleniowe	02-2028
Raporty potwierdzające jakość wygenerowanych dokumentów (WGS)	12-2029
studium wykonalności projektu	01-2027

3. KAMIENIE MIŁOWE

Kamienie milowe	Planowany termin osiągnięcia
Opracowana koncepcji architektury rozproszonego krajowego węzła PL-FEGA 2.0 wraz z analizą wymagań technologicznych, bezpieczeństwa i interoperacyjności oraz przygotowaniem infrastruktury pod wdrożenie w PCSS i Cyfronet	2027-03-31
Wdrożone i uruchomione preprodukcyjne instancje środowisk usługowych i narzędziowych PL-FEGA w PCSS i Cyfronet, obejmujących środowiska testowe oraz komponenty niezbędne do walidacji procesów deponowania, przechowywania, katalogowania i kontrolowanego udostępniania danych	2027-09-30
Produkcyjne uruchomione platformy PL-FEGA w PCSS i Cyfronet, poprzedzone testami funkcjonalnymi, integracyjnymi, wydajnościowymi i bezpieczeństwa oraz przygotowaniem dokumentacji technicznej i procedur operacyjnych	2027-12-31
Zsynchronizowane pierwotne zasoby danych krajowego węzła PL-FEGA 2.0 do środowisk w PCSS i Cyfronet, potwierdzone testami walidacji kompletności, integralności i zgodności metadanych	2028-03-31
Wdrożone i zwalidowane mechanizmy automatycznej replikacji danych, wysokiej dostępności oraz procedur awaryjnego przełączania usług pomiędzy ośrodkami PCSS i Cyfronet	2028-09-30
Wyłoniony dostawca na odczynniki do digitalizacji zasobów	2027-06-30
Zdigitalizowane próbki kolekcji GEN-A (WGS + Epigenom)	2028-06-30
Zdigitalizowane próbki kolekcji POPULOUS (WGS)	2029-12-31
Przeprowadzony inicjalny test prywatności	2027-01-31
Zdigitalizowane zasoby kolekcji onkologicznej - 200 WGS	2029-06-30
przeprowadzony weryfikacyjny test prywatności	2027-09-30
Uruchomiony system BIMSAR	2028-06-30

4. KOSZTY

4.1. Koszty ogólne projektu wraz ze sposobem finansowania

Całkowity koszt projektu (netto oraz brutto), w tym	Netto 28 063 671,24 zł Brutto 30 579 170,00 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 14 087 282,53 zł Brutto 15 876 575,00 zł
	2028	Netto 9 373 613,71 zł Brutto 10 053 820,00 zł
	2029	Netto 4 602 775,00 zł Brutto 4 648 775,00 zł

4.2. Wykaz poszczególnych pozycji kosztowych

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
Oprogramowanie	System informatyczny przeznaczony do katalogowania i zarządzania analogowymi kolekcjami przeznaczonymi do digitalizacji, licencje oprogramowania do digitalizacji genomów	1 242 300,00 zł	Systemy niezbędne do udostępniania danych zostały zaimplementowane w trakcie realizacji poprzedniego projektu w ramach Programu Operacyjnego Polska Cyfrowa. W niniejszym projekcie planowany jest zakup systemu wspierającego katalogowanie i zarządzanie zasobami przeznaczonymi do digitalizacji - BIMSAR. Dzięki m.in. implementacji elementów rozszerzonej rzeczywistości (AR) wspierających pracę personelu system usprawni proces digitalizacji i zarządzania zasobami - Szacowany koszt: 996300 Dragen - komercyjne oprogramowanie wraz z dedykowaną platformą sprzętową jest wykorzystywane w procesie digitalizacji przez Beneficjenta odpowiedzialne za składanie WGS - szacowany koszt licencji obliczeniowej: 246 00.

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
Infrastruktura	Zakup serwera realizującego backup w lokalnej przestrzeni składowania danych	492 000,00 zł	Od momentu zsekwencjonowania danych do wytworzenia gotowego do udostępniania produktu cyfrowego/dokumentu niezbędne jest wykonanie ciągu analiz. Dodatkowo w docelowym repozytorium umieszczane będą kompletne paczki danych. Do czasu ulokowania danych w docelowym repozytorium należy je zabezpieczyć w lokalnej infrastrukturze. Wycena zawiera rezerwę z uwagi na aktualną wzrostową dynamikę cen infrastruktury IT
Koszty UX i grafiki			
Bezpieczeństwo	Kontrakty serwisowe na urządzenia wykorzystywane w procesie digitalizacji, audyty bezpieczeństwa produktów projektu - platforma FEAGA 2.0 oraz BIMSAR	3 089 040,00 zł	Pozyskanie pełnogenomowej sekwencji DNA jest procesem kosztownym i czasochłonnym. Beneficjent i Partnerzy minimalizują ryzyko w sposób wcześniej sprawdzony. Mianowicie poprzez zakup kontraktu serwisowego. Takie rozwiązanie przenosi znaczną część ryzyka na dostawcę technologii. W ramach takiego zabezpieczenia dostawca nie tylko odpowiedzialny jest za naprawę urządzenia, ale również za zwrot materiałów zużywalnych i odczynników jeśli sekwencjonowanie nie powiodło się w wyniku awarii. Niezależne audyty WCAG 2.0 i Bezpieczeństwa systemów.
Wydajność rozwiązań			
Szkolenia			
Działania informacyjno-promocyjne	Działania pinformacyjno-promocyjne	615 000,00 zł	Przekazanie interesariuszom informacji o funkcjonowaniu i zasadach korzystania z Repozytorium jest kluczowa dla korzystania z zasobu przez grupy odbiorców. W ramach dziania przewidziano m.in. udział w konferencjach naukowych i branżowych w celu prezentacji funkcjonalności, organizację

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
			hackathonów z wykorzystaniem dostępnych danych, produkcję materiałów informacyjnych i szkoleniowych
Koszty zarządzania i wsparcia (w tym wynagrodzenia personelu wspomagającego)	Pozycja zawiera koszty personelu merytorycznie odpowiedzialnego za digitalizację zasobów, budowanie i migrację infrastruktury, synchronizację danych i prace nad rozwojem PL-FEGA 2.0 i środowiskiem RE; materiałów zużywalnych i odczynników; koszty aparatury niezbędnej do digitalizacji; koszty usług obliczeniowych realizowanych w środowisku chmurowym dostawcy technologii oraz koszty pośrednie	25 140 830,00 zł	<p>Materiały zużywalne i odczynniki są niezbędne do przeprowadzenia procesu digitalizacji - szacunkowy koszt: 7 698 526,00 - oszacowano na podstawie aktualnych cen zakupowych ostateczne ceny mogą być niższe z uwagi na rabat uwzględniający wolumen zakupu Sprzęt laboratoryjny niezbędny w procesie digitalizacji - szacunkowy koszt: 430 500,00 usługi obliczeniowe realizowane w środowisku chmurowym dostawcy technologii - szacunkowy koszt: 189 000,00 koszt wynagrodzeń - szacunkowy koszt: 14 892 800,00 - wynagrodzenia personelu laboratoryjnego i bioinformatycznego przeprowadzającego proces cyfryzacji zasobów, personel IT odpowiedzialny za budowę infrastruktury PL-FEGA 2.0 utrzymanie kompatybilności protokołów komunikacyjnych z istniejącymi infrastrukturami oraz budowę RE, Personel zarządzający danymi implementacja standardów opisu danych aby zachować maksymalną kompatybilność z równoległymi projektami, innymi infrastrukturami, kontrola jakości. koszty pośrednie - szacunkowy koszt: 2 233 944,00</p> <p>Dużą część budżetu stanowią koszty personelu, jest to wynikiem niskich kosztów infrastruktury, która została przygotowana w trakcie realizacji wcześniejszego projektu. Zostanie ona teraz rozproszona</p>

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
			<p>między akademickie centra komputerowe tak aby uzyskała większą niezawodność i skalowalność. Personel projektu składa się z ekspertów biotechnologicznych i informatycznych. Poziom wynagrodzeń starają się odzwierciedlać stawki rynkowe, aby wysokiej klasy specjaliści nie opuszczali środowiska akademickiego. Chodzi o zminimalizowanie ryzyka odejścia kluczowego personelu przed końcem realizacji projektu i zapewnienie sprawnej jego realizacji. Z drugiej strony jest zgodne z Polityką Naukową Państwa - ochrona środowiska naukowego przed drenażem mózgow.</p>

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

Całkowity koszt utrzymania trwałości projektu (brutto)	1 960 000,00 zł		Źródło finansowania
Podział całkowitego kosztu utrzymania trwałości projektu na poszczególne lata (netto oraz brutto)	2030	372 000,00 zł (brutto) (372 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2031	382 000,00 zł (brutto) (382 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2032	392 000,00 zł (brutto) (392 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2033	402 000,00 zł (brutto) (402 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2034	412 000,00 zł (brutto) (412 000,00 zł netto)	krajowe środki publiczne - budżet państwa

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
Niewzględnienie w analizie istotnych wymagań funkcjonalnych	Średnia	Niskie	Unikanie ryzyka - Potrzeby w zakresie funkcjonalności są dobrze zbadane i określone na podstawie funkcjonowania obecnej infrastruktury węzła FEGA. Poznane zostały słabe i mocne strony rozwiązania a także oczekiwania w zakresie zmian przez użytkowników końcowych. Personel Beneficjenta na bieżąco monitoruje dostępne technologie i weryfikuje je z zebranym zestawem wymagań funkcjonalnych.
Niedoszacowanie trudności realizacji poszczególnych funkcjonalności	Średnia	Niskie	Unikanie ryzyka - Wstępne oszacowanie pracochłonności pod kątem kosztów i czasochłonności w oparciu o analizę obecnie dostępnych technologii i rozwiązań. Ciągłe monitorowanie postępów i dokonywanie korekt w planach rozwoju.
Poważna zmiana w dostępnych technologiach	Duża	Niskie	Redukowanie ryzyka - Śledzenie tendencji rozwoju technologii i standardów. Projekt realizowany jest w oparciu o popularne i dobrze ugruntowane technologie otwartego oprogramowania. Personel beneficjenta na bieżąco monitoruje dostępne technologie i weryfikuje je z zebranym zestawem wymagań funkcjonalnych.
Problemy integracyjne z istniejącym środowiskiem infrastrukturalnym	Duża	Średnie	Unikanie ryzyka - Wczesne testy integracyjne, przygotowanie środowiska testowego oraz stała współpraca zespołów odpowiedzialnych za utrzymanie infrastruktury. System działa w oparciu o technologię Kubernetes,

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			która jest elastyczna i zbudowana z myślą o skalowalności, niemniej całkowita zmiana warstwy infrastruktury sprzętowej jest dużym ryzykiem. Mimo najlepszego przygotowania istnieje prawdopodobieństwo wystąpienia problemów. Ryzyko zostało sklasyfikowane jako średnie aby zespół projektowy miał na uwadze możliwość wystąpienia tej klasy ryzyk, a z uwagi na wysoką wartość ryzyka zostanie objęta monitoringiem do czasu jej obniżenia.
Problemy z integracją mechanizmów replikacji danych pomiędzy partnerami infrastrukturalnym i	Duża	Średnie	Unikanie ryzyka - Wspólne uzgodnienie architektury replikacji w oparciu o standardy i dobre praktyki oraz przeprowadzenie testów interoperacyjności. Mimo najlepszego przygotowania istnieje prawdopodobieństwo wystąpienia problemów. Ryzyko zostało sklasyfikowane jako średnie aby zespół projektowy miał na uwadze możliwość wystąpienia tej klasy ryzyk, a z uwagi na wysoką wartość ryzyka zostanie objęta monitoringiem do czasu jej obniżenia - najprawdopodobniej po przeprowadzonych testach interoperacyjności
Opóźnienie łańcucha dostaw w wyniku niestabilnej sytuacji geopolitycznej	Średnia	Średnie	Redukowanie ryzyka - Zawarcie długofalowych umów na dostawy (sukcesywna dostawa), przeniesienie części ryzyka na dostawcę
Zmiana ceny usług, środków trwałych i materiałów niezbędnych do realizacji projektu	Mała	Średnie	Redukowanie ryzyka – przy zakupie materiałów i usług strategicznych z punktu widzenia realizacji projektu Benefcjent stosuje umowy gwarantujące stałą cenę dla sukcesywnych dostaw lub dokonuje jednorazowego zakupu.
Ryzyka kursowe USD/PLN - ceny odczynników i kontraktów serwisowych wyrażone są w USD.	Duża	Średnie	Budżet projektu planowany jest z marginesem błędu, kontrakty z dostawcami zawierane będą w PLN - przeniesienie części ryzyka na dostawców, w wycenach starano się zachować odpowiednie marginesy

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Przekroczenie harmonogramu realizacji projektu	Duża	Niskie	Unikanie - zaplanowany koniec projektu przypada dwa miesiące przed końcem perspektywy finansowej - możliwości przedłużenia projektu są ograniczone. Dlatego Beneficjent będzie dążył do unikania tego ryzyka. Harmonogram realizacji projektu będzie systematycznie monitorowany i weryfikowany z rzeczywistym postępem prac. Sam harmonogram został utworzony przez doświadczony personel Beneficjenta i Partnerów.
Brak wystarczających środków na realizację projektu	Średnia	Niskie	Akceptacja - realizacja projektu zakłada 100% finansowanie kosztów, które zostały dobrze oszacowane bazując na wiedzy i doświadczeniu personelu Beneficjenta i Partnerów. W przypadku wystąpienia okresowych zaburzeń płatności Beneficjent i Partnerzy gotowi są prefinansować wydatki projektu
Brak możliwości zatrudnienia osób o odpowiednich kompetencjach	Średnia	Średnie	Unikanie - zaplanowano wynagrodzenia na poziomie pozwalającym konkurować z sektorem prywatnym. W przypadku niepowodzenia tej strategii do realizacji projektu alokowany będzie aktualny personel Beneficjenta i Partnerów.
Brak wystarczających zasobów kadrowych do realizacji projektu	Duża	Niskie	Unikanie - zaplanowano wynagrodzenia na poziomie pozwalającym konkurować z sektorem prywatnym. W przypadku niepowodzenia tej strategii do realizacji projektu alokowany będzie aktualny personel Beneficjenta i Partnerów. Realizatorzy mogą liczyć ponadto na rekrutację do projektu z pośród własnych studentów
Nieosiągnięcie wskaźników produktu/celu oraz celu projektu	Duża	Znikome	Unikanie - wartość wskaźników zostały zaplanowane biorąc pod uwagę możliwości produkcyjne Beneficjenta i Partnerów. Postępy w realizacji wskaźników będą regularnie monitorowane. W przypadku opóźnień będą na bieżąco wprowadzane plany naprawcze

5.2. Ryzyka wpływające na utrzymanie efektów

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Zmiana przepisów prawa uniemożliwiająca udostępnianie danych genomicznych na obecnych zasadach	Duża	Niskie	Akceptacja ryzyka - Obserwacja aktualnego otoczenia prawnego, konsultowanie aktów prawnych w procesie legislacyjnym na etapie konsultacji społecznych
Problemy z kompatybilnością rozwiązań po aktualizacjach istniejącej infrastruktury lub infrastruktury partnerskiej	Średnia	Średnie	Unikanie ryzyka - Stosowanie otwartych standardów i kompatybilnych technologii, wersjonowania komponentów oraz przeprowadzanie testów kompatybilności przed wdrożeniem aktualizacji w środowisku produkcyjnym.
Przerwy w dostępności usług wynikające z awarii infrastruktury lub usług zależnych	Duża	Znikome	Redukcja ryzyka - Wdrożenie mechanizmów redundancji i monitoringu usług, regularne testy procedur odtworzeniowych oraz utrzymywanie kopii zapasowych. Ponadto każdy z Partnerów implementuje redundancje na poziomie własnego centrum danych. Właśnie ograniczenie tego ryzyka spowodowało podjęcie realizacji tego projektu.
Utrata synchronizacji danych pomiędzy węzłami redundantnymi	Duża	Niskie	Unikanie ryzyka - Stałe monitorowanie procesów replikacji, automatyczne alertowanie o błędach synchronizacji oraz okresowe testy odtwarzania danych.
Niewystarczająca wydajność mechanizmów replikacji przy rosnącym wolumenie danych	Średnia	Średnie	Redukcja ryzyka - Regularne monitorowanie wydajności mechanizmów replikacji, dostosowywanie konfiguracji synchronizacji danych do aktualnego obciążenia systemu.
Brak możliwości zatrudnienia osób o odpowiednich kompetencjach niezbędnych do utrzymania efektów projektu	Średnia	Niskie	Unikanie - Zakładane jest zatrzymanie kluczowych dla utrzymania efektów projektu pracowników z zespołu projektowego. W ramach projektu wytworzone zostaną rozwiązania automatyzujące utrzymanie Infrastruktury dzięki czemu potrzebna będzie mniejsza ilość personelu oraz kompetencje pracowników będą mogły być niższe niż na etapie realizowanego

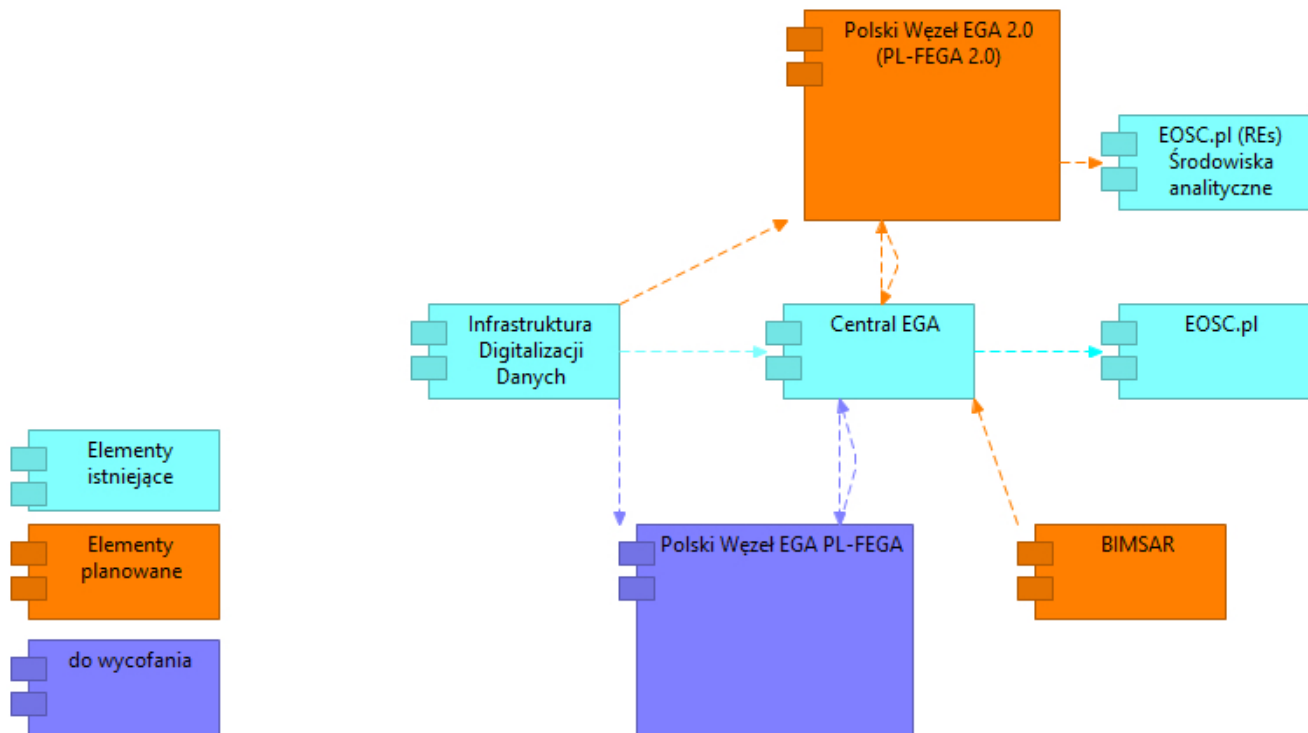
Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			projektu. Dlatego ustawiono wartość prawdopodobieństwa wystąpienia na niską.
Brak wystarczających środków na utrzymanie efektów projektu	Duża	Niskie	Unikanie - m.in. właśnie w celu uniknięcia tego ryzyka podjęto decyzję o realizacji projektu. Rozproszenie infrastruktury PL-FEGA pomiędzy 3 jednostki znacznie obniża szanse zakłócenia funkcjonowania Węzła w przypadku wystąpienia problemów po stronie któregoś z Partnerów
Brak wystarczających zasobów kadrowych do utrzymania efektów projektu	Średnia	Niskie	Unikanie - Zakładane jest zatrzymanie kluczowych dla utrzymania efektów projektu pracowników z zespołu projektowego. Realizatorzy mogą liczyć ponadto na rekrutację do projektu z pośród własnych studentów. W ramach projektu wytworzone zostaną rozwiązania automatyzujące utrzymanie Infrastruktury dzięki czemu potrzebna będzie mniejsza ilość personelu oraz kompetencje pracowników będą mogły być niższe niż na etapie realizowanego projektu. Dlatego ustawiono wartość prawdopodobieństwa wystąpienia na niską.
Brak wystarczających środków na utrzymanie efektów projektu	Duża	Znikome	Akceptacja - podejmując decyzję o realizacji projektu uwzględniono koszty utrzymania infrastruktury. Zostały ujęte w planach finansowych Beneficjenta i Partnerów.
Nieosiągnięcie wszystkich zaplanowanych korzyści	Duża	Niskie	Unikanie - Beneficjent zaplanował akcje promocyjne, które pozwolą dotrzeć z informacją do zidentyfikowanych grup interesariuszy w celu zachęcenia ich do korzystania z infrastruktury węzła i pozyskiwania zdigitalizowanych zasobów.

6. OTOCZENIE PRAWNE

Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
1	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)	TAK/NIE		
2	Ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa	TAK/NIE		
3	Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego	TAK/NIE		
4	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2025/327 z dnia 11 lutego 2025 r. w sprawie europejskiej przestrzeni danych dotyczących zdrowia oraz zmiany dyrektywy 2011/24/UE i rozporządzenia (UE) 2024/2847.	TAK/NIE		
5	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/868 z dnia 30 maja 2022 r. w sprawie europejskiego zarządzania danymi (Akt w sprawie zarządzania danymi)	TAK/NIE		
6	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 (Akt w sprawie danych)	TAK/NIE		
7	ustawa o dostępności cyfrowej stron internetowych i aplikacji mobilnych podmiotów publicznych,	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	Centralny Węzeł Europejskiego Archiwum Genomów	Europejski Instytut Bioinformatyki i Centrum Regulacji Genomicznych w Barcelonie	European Genome Archive – repozytorium danych genetycznych przechowujących i udostępniających dane dotyczące ludzi. Jako centralny element federacyjnej infrastruktury FEGA odpowiada za zarządzanie użytkownikami systemu, uprawnieniami do zasobów, jest magazynem metadanych oraz dostarcza silnik wyszukiwania. Integracje z innymi systemami: EOSC.pl poprzez API PL-FEGA (i innymi węzłami krajowymi) protokołem AMQP PL-FEGA 2.0 protokołem	Istniejący	Brak modyfikacji

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			AMQP		
2	EOSC	EOSC Association / Polski Węzeł EOSC	Europejska Chmura Danych - system udostępniający dane badawcze lub metadane je opisujące w celu łatwiejszej identyfikacji zasobów przez środowisko naukowe. Integracja: Europejskie Archiwum Genomów przez API	Istniejący	Brak modyfikacji
3	Polski Węzeł Europejskiego Archiwum Genomów (PL-FEGA)	Uniwersytet Łódzki	Polski Węzeł Europejskiego Archiwum Genomów (PL-FEGA) - System odpowiedzialny za przetwarzanie danych genomicznych populacji polskiej. Dane genetyczne to dane wymagające szczególnej ochrony wg. RODO. W infrastrukturze PL-FEGA są one przechowywane w formie szyfrogramu nad którym dostęp do danych czuwa Komitet Dostępu do Danych (DAC). Jeden DAC może obsługiwać tylko jeden jak i wiele zbiorów danych - decyzję podejmują gestorzy zbiorów. Dzięki przetwarzaniu na terenie kraju w krajowej infrastrukturze dane podlegają wyłącznie polskiej jurysdykcji. Integracje: CEGA przez protokół AMQP BIMSAR Planowany do wycofania	Modyfikowany	Istniejąca infrastruktura Polskiego Węzła zostanie rozproszona pomiędzy Partnerów projektu - powstaną trzy lokalizacje w których zostaną składowane zasoby niezbędne do działania systemu. Dzięki temu rozwiązaniu Polski Węzeł uzyska większą niezawodność i skalowalność. Zmodyfikowana zostanie architektura infrastruktury tak aby mogła - zostać rozproszona pomiędzy 3 lokalizacje, - zachodzić synchronizacja

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
					<p>danych pomiędzy poszczególnymi i lokalizacjami Ponadto zmodyfikowany zostanie sposób udostępniania plików wykorzystujący tworzenie dedykowanych zasobów chmurowych - zostanie utworzone TRE co pozwoli kopiować dane z repozytorium do TRE w oparciu o sieć lokalną.</p> <p>Wprowadzone modyfikacje nie zmienią dotychczasowej interoperacyjności systemu.</p>
4	BIMSAR	Uniwersytet Łódzki	System informatyczny obsługujący zarządzanie i katalogowanie zasobów przeznaczonych do digitalizacji. System wykorzystywał będzie elementy rozszerzonej rzeczywistości (AR) aby zwiększyć ergonomię obsługi systemu i usprawnić procesy związane z zarządzaniem zasobami do digitalizacji - materiałem biologicznym przechowywanym w dedykowanych	Planowany	Nowy system

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			<p> pomieszczeniach laboratoryjnych m.in. w zamrażarkach niskotemperaturowych (-80st.C) lub ciekłym azocie. System obsługiwał będzie kluczowe procesy takie jak przyjęcie czy wydanie materiałów, System wprowadzi automatyzację w przygotowaniu metadanych opisujących deponowane w PL-FEGA dane. System udostępniany będzie nieodpłatnie podmiotom chcącym deponować dane w PL-FEGA </p>		

Lista przepływów

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
1	Centralny Węzeł Europejskiego Archiwum Genomów	Polski Węzeł Europejskiego Archiwum Genomów PL-FEGA	<p> Po pozytywnej weryfikacji przez Komitet Dostępu do Danych żądania dostępu do danych, przekazywane jest żądanie nadania uprawnień do zbioru danych konkretnemu użytkownikowi. Przekazywana jest nazwa użytkownika i hash hasła w celu </p>	Tryb odwołań bezpośrednich	Realizowany inną metodą	Protokół AMQP

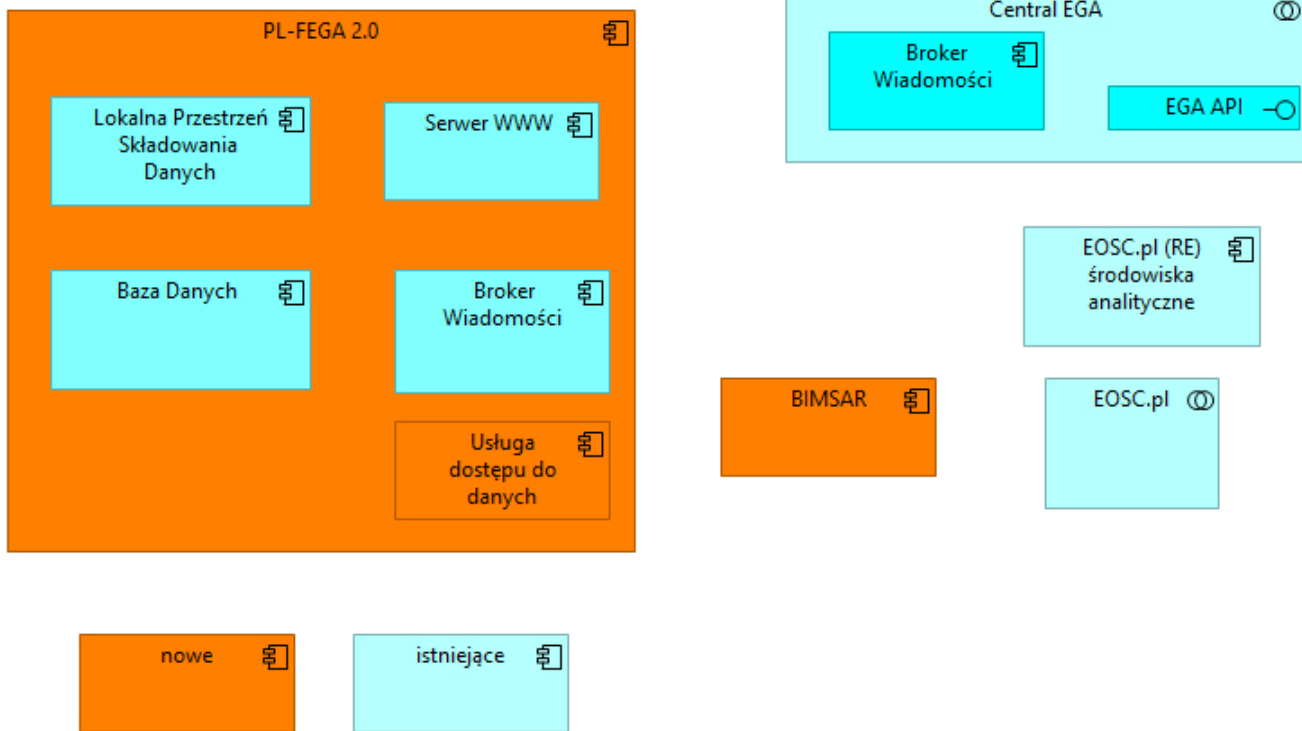
Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			weryfikacji poprawności logowania do usługi pobierania danych, klucz publiczny w celu przekodowania danych Po zakończonym procesie umieszczania danych w Węźle, CEGA przesyła unikalny stały identyfikator zbioru danych.			
2	Polski Węzeł Europejskiego Archiwum Genomów PL-FEGA	Centralny Węzeł Europejskiego Archiwum Genomów	Potwierdzenie poprawnego zdeponowania przez użytkownika danych w Polskim Węźle Europejskiego Archiwum Genomów. Zakres przesyłanych danych to Nazwa pliku, Suma kontrolna Login użytkownika odpowiedzialnego za umieszczenie danych Opcjonalnie sumy kontrolne plików.	Tryb Odwołań Bezpośrednich	Realizowany inną metodą	Protokół AMQP
3	Centralny Węzeł Europejski	EOSC	Metadane opisujące zdeponowane	Kopiowanie	Realizowany inną metodą	API

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
	ego Archiwum Genomów		w Polskim Węźle Europejskiego Archiwum Genomów zbiory danych. Zakres danych: Trwały (persistent) Identyfikator zbioru danych, url, opis zbioru			
4	Infrastruktura Digitalizacji Danych	PL-FEGA	sekwencje WGS	kopiowanie danych	realizowana inną metodą	SFTP
5	Infrastruktura Digitalizacji Danych	Central EGA	Metadane opisujące WGS, protokół i odczynniki wykorzystane do digitalizacji,	tryb odwołań bezpośrednich	Krytyczne dla sukcesu projektu	API
6	Infrastruktura Digitalizacji Danych	PL-FEGA 2.0	sekwencje WGS	kopiowanie danych	Krytyczny dla sukcesu projektu	HTTPS/S3
7	PL-FEGA 2.0	Central EGA	Po pozytywnej weryfikacji przez Komitet Dostępu do Danych żądania dostępu do danych, przekazywane jest żądanie nadania uprawnień do zbioru danych konkretnemu użytkownikowi. Przekazywana jest nazwa użytkownika i hash hasła w celu	tryb odwołań bezpośrednich	Krytyczny dla realizacji projektu	Protokół AMQP

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			weryfikacji poprawności logowania do usługi pobierania danych, klucz publiczny w celu przekodowania danych Po zakończonym procesie umieszczania danych w Węźle, CEGA przesyła unikalny stały identyfikator zbioru danych.			
8	Central EGA	PL-FEGA 2.0	Po pozytywnej weryfikacji przez Komitet Dostępu do Danych żądania dostępu do danych, przekazywane jest żądanie nadania uprawnień do zbioru danych konkretnemu użytkownikowi. Przekazywana jest nazwa użytkownika i hash hasła w celu weryfikacji poprawności logowania do usługi pobierania danych, klucz publiczny w celu przekodowania	Tryb odwołań bezpośrednich	Krytyczny dla sukcesu projektu	Protokół AMQP

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			danych Po zakończonym procesie umieszczania danych w Węźle, CEGA przesyła unikalny stały identyfikator zbioru danych.			
9	BIMSAR	Central EGA	Metadane opisujące dane w zakresie wiek, rozpoznanie ICD-10, płeć itp.	Kopiowanie danych	realizowany inną metodą	HTTPS
10	PL-FEGA 2.0	EOSC.pl RE (research environment)	Sekwencje WES	Kopiowanie Danych	realizowany inną metodą	Protokoły QCG lub ONE DATA

7.2. Kluczowe komponenty architektury rozwiązania



7.3. Przyjęte założenia technologiczne

Lp.	Obszar	Założenie technologiczne
1.	Infrastruktura	W projekcie zostanie wykorzystana w całości już istniejąca infrastruktura informatyczna FEGA. W ramach projektu istniejąca instancja Polskiego Węzła FEGA zostanie rozproszona pomiędzy trzy jednostki, zyskując tym samym większą odporność, skalowalność i niezawodność. Partnerzy projektu są jednostkami publicznymi PCSS i Cyfronet są dużymi publicznie dostępnymi Data Center realizujące krajowe projekty infrastruktury takich jak: KMD, Fabryki AI,
2.	Sieć i bezpieczeństwo	Komunikacja z planowanymi do wdrożenia aplikacjami webowymi odbywać się będzie z wykorzystaniem protokołu TLS, SSH lub SFTP/FTPS
3.	Standardy wymiany danych	Wymiana danych odbywać się będzie w oparciu o standardy wypracowane przez Europejskie Archiwum Genomów
4.	Systemy operacyjne serwerowe	Preferowane do stosowania przez Beneficjenta, Partnerów i twórców rozwiązania są systemy operacyjne klasy Linux
5.	Bazy danych	Obecnie stosowany system wykorzystuje silnik bazy danych PostgreSQL, Partnerzy projektu pragną pozostać przy rozwiązaniach open source.
6.	Serwery aplikacji	System FEGA został wytworzony w języku JAVA i wykorzystuje framework Spring. Uruchamiany jest na serwerze Tomcat. Wykorzystane technologie są szeroko wykorzystywane w systemach klasy biznes zapewniając bezpieczeństwo danych oraz wydajność aplikacji.

Lp.	Obszar	Założenie technologiczne
7.	Portale	
8.	Inne	

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

Beneficjent jest operatorem Węzła EGA. System został opracowany przez Europejski Instytut Bioinformatyki jako rozwiązanie paneuropejskie.

Obecnie nie ma wymogów dotyczących zabezpieczania danych objętych projektem. Mechanizmy zabezpieczeń wynikają z wymogów RODO i zostały dobrane na podstawie analizy ryzyka zarówno na poziomie Węzła jak i Całego EGA. W chwili obecnej trwają prace nad spójną polityką bezpieczeństwa infrastruktury FEAGA jako całości. Kluczowe jest szybkie przywrócenie systemu po awarii/incydencie - zastosowano dwie redundantne geograficznie kopie. Dane wrażliwe przechowywane są w formie szyfrogramu. Komunikacja odbywa się poprzez zaszyfrowane łącza a poszczególne serwisy muszą uwierzytelnić się certyfikatem. Wdrożone zostały mechanizmy bezpieczeństwa wypracowane przez Europejski Instytut Bioinformatyki. Bezpieczeństwo systemu potwierdzone zostało testami penetracyjnymi wykonanymi przez podmiot trzeci.

~~-dodatkowe zabezpieczenia powyżej wymogów KRI: należy wskazać uzasadnienie~~