

# OPIS ZAŁOŻEŃ PROJEKTU INFORMATYCZNEGO

<b>Tytuł projektu</b>	DATA4EOSC - Ekosystem otwartych środowiskowych danych naukowych dla badań, innowacji i zrównoważonego rozwoju		
<b>Wnioskodawca</b>	Minister Nauki i Szkolnictwa Wyższego		
<b>Beneficjent</b>	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, Poznańskie Centrum Superkomputerowo-Sieciowe		
<b>Partnerzy</b>	Instytut Ochrony Roślin - Państwowy Instytut Badawczy Instytut Hodowli i Aklimatyzacji Roślin - Państwowy Instytut Badawczy Instytut Badawczy Leśnictwa Ideas Instytut Badawczy Instytut Włókien Naturalnych i Roślin Zielarskich Państwowy Instytut Badawczy Wielkopolski Park Narodowy Karkonoski Park Narodowy Instytut Technologiczno-Przyrodniczy - Państwowy Instytut Badawczy		
<b>Źródło finansowania</b>	1. Środki UE: Fundusze Europejskie na Rozwój Cyfrowy, Działanie 02.03 - Cyfrowa dostępność i ponowne wykorzystanie informacji. 2. Budżet Państwa – część 27.		
<b>Całkowity koszt projektu</b>	20 800 000,00 zł		
<b>Planowany okres realizacji projektu</b>	11-2026 do 10-2029		
<b>Osoba kontaktowa</b>	Maciej Zacharczuk	mzacharczuk@man.poznan.pl	601943549

## 1. POWODY PODJĘCIA PROJEKTU

### 1.1. Identyfikacja problemu i potrzeb

W Polsce istnieją zasoby danych istotnych dla badań nad środowiskiem, rolnictwem, leśnictwem, klimatem i bioróżnorodnością. Są one rozproszone między instytucjami badawczymi, uczelniami i administracją oraz udostępniane w sposób niejednolity. Portal dane.gov.pl gromadzi ponad 43 tys. zbiorów, lecz w obszarach kluczowych dla projektu dostępnych jest 177 zbiorów w kategorii rolnictwo, 338 w kat. środowisko i 751 w kat. nauka. Sama liczba zbiorów nie zapewnia jednak spójnego, międzysektorowego dostępu do środowiskowych danych naukowych.

Stan obecny odbiega od oczekiwanego. Model funkcjonowania opiera się na niezależnych repozytoriach i systemach dziedzinowych, które nie zapewniają wspólnego modelu interoperacyjności, semantyki ani mechanizmów automatycznego ponownego wykorzystania danych pomiędzy sektorami. Ogranicza to możliwość prowadzenia badań interdyscyplinarnych, analiz przekrojowych oraz wykorzystania danych w rozwiązaniach AI i systemach wspomaganie decyzji. Szczególną barierą jest brak wspólnych modeli informacyjnych i semantycznych umożliwiających spójny opis oraz integrację danych pochodzących z różnych instytucji.

Digitalizacja danych i rozbudowa każdego repozytorium osobno prowadziłyby do powielania

kosztów i utrzymania fragmentacji środowiska danych. Model federacyjny pozwoli zachować autonomię partnerów przy zapewnieniu jednolitego sposobu wyszukiwania, opisu, integracji i ponownego wykorzystania danych zgodnego z FAIR oraz dyrektywą UE 2019/1024 oraz ustawą o otwartych danych.

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
Wnioskodawca – Ministerstwo Nauki i Szkolnictwa Wyższego	Ograniczona dostępność interoperacyjnych danych naukowych utrudnia realizację polityk państwa w zakresie otwartej nauki, cyfryzacji oraz ponownego wykorzystania danych naukowych finansowanych ze środków publicznych.	1
Beneficjent – Instytut Chemii Bioorganicznej PAN / Poznańskie Centrum Superkomputerowo-Sieciowe	Brak wspólnego środowiska interoperacyjności i integracji danych utrudnia koordynację publikacji, standaryzacji, jakości i wymiany danych pomiędzy wieloma dostawcami danych naukowych, w tym realizację projektów i badań interdyscyplinarnych wspieranych przez narzędzia cyfrowe i sztuczną inteligencję.	1
Partnerzy projektu - instytuty badawcze, instytucje posiadające dane naukowe	Dane naukowe partnerów są rozproszone, opisane w sposób niejednolity i często nieprzygotowane do udostępniania zgodnie z wymaganiami interoperacyjności, FAIR i 5 Star Open Data. Utrudnia to prowadzenie badań i uczestnictwo w projektach interdyscyplinarnych bądź obejmujących elementy cyfrowe i sztuczną inteligencję	7
Uczelnie wyższe i instytuty badawcze spoza konsorcjum jako potencjalni wytwórcy i użytkownicy danych naukowych	Brak łatwego dostępu do interoperacyjnego ekosystemu danych ogranicza możliwość współdzielenia zasobów oraz prowadzenia badań interdyscyplinarnych i współpracy międzyinstytucjonalnej.	346
Badacze i zespoły badawcze prowadzące prace w obszarach rolnictwa, środowiska, klimatu, leśnictwa i bioróżnorodności na terenie Polski.	Wysokie nakłady czasu na wyszukiwanie, ocenę jakości i przygotowanie danych ograniczają możliwość prowadzenia analiz międzysektorowych oraz wykorzystania AI i metod data-driven science.	72 000
Administracja publiczna szczebla regionalnego i lokalnego	Utrudniony dostęp do wiarygodnych i interoperacyjnych danych środowiskowych ogranicza możliwość prowadzenia analiz, planowania przestrzennego i zarządzania środowiskiem.	2 875
Podmioty sektora rolnictwa, leśnictwa i	Brak łatwo dostępnych, wiarygodnych danych ogranicza możliwość wykorzystywania	1 300 000

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
ochrony środowiska jako odbiorcy pośredni danych i usług, w tym gospodarstwa rolne, leśnictwa, doradztwo rolnicze, instytucje ochrony środowiska, organizacje pozarządowe, itp.	wyników badań w praktyce gospodarczej, doradczej, środowiskowej i zarządczej, w szczególności przy ocenie ryzyk, planowaniu działań i wdrażaniu rozwiązań wspierających decyzje.	
Operatorzy zewnętrznych repozytoriów, katalogów i platform danych naukowych oraz publicznych, które będą integrowane w projekcie: dane.gov.pl, EOSC Portal, CEADS data space, SAGE data space, KDM4EOSC Services	Brak wspólnego modelu wymiany i opisu danych utrudnia synchronizację metadanych, zwiększenie widoczności zasobów oraz integrację z krajowymi i europejskimi ekosystemami danych.	5
Przedsiębiorstwa i podmioty rozwijające usługi oparte na danych na terenie Polski i za granicą	Utrudniony dostęp do danych gotowych do automatycznego wykorzystania zwiększa koszt budowy produktów i usług cyfrowych, w tym rozwiązań agrotech, środowiskowych, ubezpieczeniowych, predykcyjnych i opartych na AI.	412 800
Instytucje rządowe i podległe odpowiedzialne za polityki publiczne w obszarze nauki, rolnictwa, środowiska, klimatu, cyfryzacji i otwartych danych	Ograniczona dostępność porównywalnych i wiarygodnych danych utrudnia przygotowywanie analiz przekrojowych, monitorowanie efektów polityk publicznych oraz ocenę wpływu działań w obszarach rolnictwa, środowiska, klimatu, leśnictwa i nauki.	554

## 1.2. Opis stanu obecnego

Obecnie procesy związane z gromadzeniem, przetwarzaniem i udostępnianiem środowiskowych danych naukowych realizowane są niezależnie. Partnerzy projektu wykorzystują własne repozytoria danych, lokalne bazy danych, systemy laboratoryjne, arkusze kalkulacyjne, systemy GIS, repozytoria plikowe oraz dziedzinowe systemy monitoringowe i badawcze. Część danych funkcjonuje nadal w postaci dokumentów papierowych, archiwów lub nieustrukturyzowanych skanów PDF wymagających digitalizacji.

Dane udostępniane są poprzez strony internetowe, lokalne repozytoria lub indywidualne eksporty danych. W części systemów wykorzystywane są interfejsy API oraz standardowe identyfikatory,

np. kody EPPO, jednak ich stosowanie nie jest jednolite pomiędzy instytucjami, wewnątrz tych instytucji i poszczególnymi zbiorami danych. Poszczególne systemy wykorzystują odmienne słowniki (nawet kilka w jednej instytucji), struktury metadanych i sposoby identyfikacji obiektów, w tym nawet odmienne nazwy roślin, szkodników i parametrów środowiskowych. Proces wymiany danych pomiędzy instytucjami realizowany jest obecnie punktowo i wymaga manualnego przygotowania danych oraz mapowania struktur.

Zakres danych objętych digitalizacją i standaryzacją obejmuje m.in. dane monitoringowe i laboratoryjne dotyczące ochrony roślin, zasoby banków patogenów, dane środowiskowe, GIS i ortofotomapy, dokumentację siedliskową i leśną, wyniki badań genotypowych i hodowlanych, publikacje naukowe oraz zasoby papierowe przeznaczone do digitalizacji.

Projekt zmieni obecny model poprzez budowę federacyjnej platformy DATA4EOSC. Powstanie wspólna warstwa interoperacyjności obejmująca katalog danych, repozytorium metadanych i ontologii, model semantyczny (bazujący na istniejącym AIM), standaryzowane API oraz narzędzia harmonizacji i jakości danych. Platforma umożliwi automatyczne wyszukiwanie, integrację i ponowne wykorzystanie danych zgodnie z FAIR i 5 Star Open Data, w tym publikację w dane.gov.pl, EOSC i przestrzeniach danych.

## 2. EFEKTY PROJEKTU

### 2.1. Cele i korzyści wynikające z projektu

<b>Cel - 1</b>	Zwiększenie dostępności i możliwości ponownego wykorzystania środowiskowych danych naukowych poprzez ich cyfrowe udostępnienie i integrację
<b>Cel strategiczny</b>	<p>„Program Otwierania Danych na lata 2021-2027” - cel szczegółowy 5.4 „Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych” - podpunkt 5.4.2 Zwiększenie dostępności zasobów nauki.</p> <p>Strategia na rzecz Odpowiedzialnego Rozwoju - Obszar: E-państwo w zakresie zapewnienia organizacyjnej, prawnej, semantycznej oraz technicznej interoperacyjności systemów teleinformatycznych.</p>
<b>Korzyść:</b>	<p>Trwałe zwiększenie dostępności i wykorzystania cyfrowych zasobów naukowych z obszarów środowiska, rolnictwa, klimatu i bioróżnorodności przez administrację publiczną, jednostki naukowe oraz przedsiębiorstwa, skrócenie czasu pozyskania danych, obniżenie kosztów dostępu do informacji oraz zwiększenie efektywności prowadzenia analiz, badań i procesów decyzyjnych.</p> <p>Przykładowe wykorzystanie obejmuje: wsparcie administracji publicznej w planowaniu przestrzennym i zarządzaniu ryzykiem środowiskowym na podstawie zintegrowanych danych; prowadzenie badań interdyscyplinarnych poprzez łączenie danych genotypowych, środowiskowych i meteorologicznych; wykorzystanie danych w systemach wspomaganie decyzji (DSS) dla rolnictwa oraz analizie danych monitoringowych (np. fotopułapki, lidar) w ochronie przyrody.</p>
<b>KPI:</b>	<p>KPI 1 Liczba podmiotów, które udostępniły informacje sektora publicznego online</p> <p>KPI 2 Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne</p>

	<p>KPI 3 Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne</p> <p>KPI 4 Rozmiar zdigitalizowanych informacji sektora publicznego/danych prywatnych</p> <p>KPI 5 Rozmiar udostępnionych on-line informacji sektora publicznego/danych prywatnych</p>
<p><b>Wartość aktualna i docelowa KPI:</b></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: 7 podmiotów</p> <p>KPI 2: 190 000 szt</p> <p>KPI 3: 5 800 000 szt</p> <p>KPI 4: 270 TB</p> <p>KPI 5: 170 TB</p>
<p><b>Metoda pomiaru KPI</b></p>	<p>KPI 1: Badanie ewaluacyjne ilościowe na podstawie liczby partnerów publikujących dane online; źródło danych: dokumentacja projektowa, protokoły odbioru, katalog danych; częstotliwość: wartość docelowa mierzona jednorazowo (08-2029).</p> <p>KPI 2: Raporty digitalizacji; pomiar liczby rekordów/plików zdigitalizowanych; źródło danych: repozytoria danych partnerów, raporty z digitalizacji; częstotliwość: wartość docelowa mierzona jednorazowo (04-2029).</p> <p>KPI 3: Badanie ewaluacyjne ilościowe liczby dokumentów opublikowanych online; źródło danych: system publikacji, katalog danych, raport wdrożeniowy; częstotliwość: wartość docelowa mierzona jednorazowo (08-2029).</p> <p>KPI 4: Rozmiar danych w TB na podstawie monitoringu zasobów; źródło danych: repozytoria danych, raporty techniczne; częstotliwość: wartość docelowa mierzona jednorazowo (04-2029).</p> <p>KPI 5: Rozmiar danych w TB opublikowanych online; źródło danych: monitoring platformy, katalog danych, logi systemowe; częstotliwość: wartość docelowa mierzona jednorazowo (08-2029).</p>
<p><b>Cel - 2</b></p>	<p>Zwiększenie poziomu ponownego wykorzystania danych naukowych z sektorów związanych ze środowiskiem poprzez zapewnienie ich poziomu otwartości, jakości, standaryzacji i zgodności z 5 Star Open Data i FAIR</p>
<p><b>Cel strategiczny</b></p>	<p>„Program Otwierania Danych na lata 2021-2027” - cel szczegółowy 5.4 „Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych” - podpunkt 5.4.2 Zwiększenie dostępności zasobów nauki.</p> <p>Strategia na rzecz Odpowiedzialnego Rozwoju - Obszar: E-państwo w zakresie zapewnienia organizacyjnej, prawnej, semantycznej oraz technicznej interoperacyjności systemów teleinformatycznych.</p>
<p><b>Korzyść:</b></p>	<p>Trwały wzrost poziomu ponownego wykorzystania danych naukowych przez sektor publiczny, naukę i przedsiębiorstwa, prowadzący do rozwoju nowych analiz, usług cyfrowych i innowacji opartych na danych oraz zwiększenia kompetencji organizacji naukowych w zakresie zarządzania, standaryzacji i udostępniania danych wysokiej jakości.</p>

	Dane będą ponownie wykorzystywane m.in. do tworzenia modeli predykcyjnych i analiz, rozwoju usług cyfrowych przez przedsiębiorstwa oraz przygotowania zbiorów danych dla sztucznej inteligencji (AI-ready datasets), umożliwiającymi trenowanie modeli analitycznych i automatyzację przetwarzania danych.
<b>KPI:</b>	KPI 1 Liczba instytucji publicznych otrzymujących wsparcie na opracowywanie usług, produktów i procesów cyfrowych KPI 2 Liczba podmiotów wspartych w zakresie rozwoju usług, produktów i procesów cyfrowych KPI 3 Użytkownicy nowych i zmodernizowanych publicznych usług, produktów i procesów cyfrowych KPI 4 Liczba pracowników IT objętych wsparciem szkoleniowym KPI 5 Liczba pracowników nie będących pracownikami IT, objętych wsparciem szkoleniowym
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: 0 KPI 2: 0 KPI 3: 0 KPI 4: 0 KPI 5: 0 KPI 1: 2 szt KPI 2: 7 szt KPI 3: 8000 użytkowników/rok KPI 4: 14 osób KPI 5: 140 osób
<b>Metoda pomiaru KPI</b>	KPI 1: Badanie ilościowe liczby instytucji objętych wsparciem; źródło danych: umowa dofinansowania, raport realizacji projektu; częstotliwość: półroczna, pomiar końcowy w (04-2029).  KPI 2: Badanie ilościowe liczby podmiotów korzystających ze wsparcia; źródło danych: dokumentacja projektu, raporty wdrożeniowe; częstotliwość: pomiar końcowy w (10-2029).  KPI 3: Badanie ilościowe użytkowników usług cyfrowych na podstawie logów i kont użytkowników; źródło danych: baza danych systemu, logi systemowe; częstotliwość: kwartalna, wartość docelowa do 12 miesięcy po zakończeniu projektu.  KPI 4: Badanie ilościowe liczby pracowników IT objętych szkoleniami; źródło danych: listy obecności, raporty szkoleniowe; częstotliwość: po każdym cyklu szkoleniowym, pomiar końcowy (10-2029).  KPI 5: Badanie ilościowe liczby pracowników nie-IT objętych szkoleniami; źródło danych: listy obecności, raporty szkoleniowe, ankiety ewaluacyjne; częstotliwość: po każdym szkoleniu, pomiar końcowy (10-2029).
<b>Cel - 3</b>	Osiągnięcie interoperacyjnego i międzysektorowego dostępu do środowiskowych danych naukowych w modelu federacyjnym, umożliwiającego ich automatyczne wyszukiwanie, integrację i ponowne wykorzystanie
<b>Cel strategiczny</b>	“Program Otwierania Danych na lata 2021-2027” - cel szczegółowy 5.4 „Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych

	<p>naukowych” - podpunkt 5.4.2 Zwiększenie dostępności zasobów nauki.</p> <p>Strategia na rzecz Odpowiedzialnego Rozwoju - Obszar: E-państwo w zakresie zapewnienia organizacyjnej, prawnej, semantycznej oraz technicznej interoperacyjności systemów teleinformatycznych.</p> <p>Polityka rozwoju sztucznej inteligencji w Polsce od 2020 roku, obszar: AI i dane jako fundament rozwoju innowacji i cyfryzacji</p>
<b>Korzyść:</b>	<p>Trwale zwiększenie efektywności współpracy i wymiany danych pomiędzy instytucjami naukowymi, administracją publiczną i podmiotami gospodarczymi dzięki wdrożeniu wspólnej interoperacyjnej infrastruktury danych, obniżenie kosztów integracji danych, zwiększenie gotowości instytucji i firm do wykorzystania danych w zaawansowanych usługach, analizach i rozwiązaniach sztucznej inteligencji, skrócenie czasu wyszukiwania i pozyskania danych przez użytkowników dzięki centralnemu katalogowi i API oraz zwiększenie liczby ponownego wykorzystania danych przez administrację, naukę i przedsiębiorstwa.</p> <p>Środowisko interoperacyjne umożliwi m.in. automatyczną integrację danych z różnych domen i ich wykorzystanie w analizach międzysektorowych, dostęp do danych poprzez katalog i API, integrację z krajowymi i europejskimi ekosystemami (np. dane.gov.pl, EOSC) oraz rozwój zaawansowanych zastosowań, takich jak analizy środowiskowe i systemy oparte na sztucznej inteligencji.</p>
<b>KPI:</b>	<p>KPI 1 Liczba uruchomionych systemów teleinformatycznych</p> <p>KPI 2 Liczba platform udostępniania dokumentów zawierających informacje sektora publicznego/dane prywatne</p> <p>KPI 3 Liczba baz danych udostępnionych on-line poprzez API</p> <p>KPI 4 Liczba utworzonych API</p> <p>KPI 5 Liczba wygenerowanych kluczy API</p> <p>KPI 6 Liczba rozwiązań wykorzystujących informacje sektora publicznego</p>
<b>Wartość aktualna i docelowa KPI:</b>	<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 6: 0</p> <p>KPI 1: 1 szt</p> <p>KPI 2: 6 szt</p> <p>KPI 3: 35 szt</p> <p>KPI 4: 11 szt</p> <p>KPI 5: 80 szt</p> <p>KPI 6: 30 szt</p>
<b>Metoda pomiaru KPI</b>	<p>KPI 1: Badanie ilościowe uruchomionych systemów; źródło danych: protokół odbioru systemu, dokumentacja wdrożeniowa; częstotliwość: jednorazowo na zakończenie wdrożenia (10-2028).</p> <p>KPI 2: Badanie ilościowe liczby platform; źródło danych: dokumentacja projektowa, protokół odbioru platformy; częstotliwość: jednorazowo w dniu uruchomienia produkcyjnego (10-2028).</p> <p>KPI 3: Badanie ilościowe liczby baz danych dostępnych przez API; źródło</p>

	<p>danych: katalog API, repozytorium danych, dokumentacja techniczna; częstotliwość: pomiar końcowy (10-2029).</p> <p>KPI 4: Badanie ilościowe liczby utworzonych API; źródło danych: dokumentacja techniczna, repozytorium kodu, katalog API; częstotliwość: kwartalna, pomiar końcowy (10-2029).</p> <p>KPI 5: Badanie ilościowe liczby wygenerowanych kluczy API; źródło danych: system zarządzania API, logi systemowe; częstotliwość: kwartalna, wartość docelowa do 12 miesięcy po zakończeniu projektu.</p> <p>KPI 6: Badanie ewaluacyjne ilościowe liczby rozwiązań wykorzystujących dane; źródło danych: ankiety użytkowników, raporty wdrożeń, deklaracje partnerów; częstotliwość: roczna, wartość docelowa do 12 miesięcy po zakończeniu 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)
Zdigitalizowane i udostępnione zasoby Instytutu Ochrony Roślin obejmujące wieloletnie dane badawcze i monitoringowe dotyczące stosowania środków ochrony roślin w Polsce. Zakres obejmuje: (1) bazę pozostałości środków ochrony roślin (dane laboratoryjne, wyniki analiz, lokalizacja prób, parametry chemiczne), (2) dane o sprzedaży i zużyciu środków ochrony roślin (raporty roczne, zestawienia tabelaryczne, analizy sektorowe), (3) zalecenia stosowania	31-08-2028	45 000 rekordów i dokumentów

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
<p>środków ochrony roślin powiązane z bazą substancji czynnych, w tym środków ekologicznych, (4) archiwalne komunikaty o występowaniu i nasileniu agrofagów (1960–2000) w układzie miesięcznym i regionalnym, (5) zasoby banku patogenów obejmujące kolekcję ok. 2000 izolatów grzybów i ponad 200 szczepów bakterii wraz z dokumentacją identyfikacyjną, pochodzeniem i metodami przechowywania. Zakres prac obejmuje digitalizację dokumentów archiwalnych, OCR, ekstrakcję danych tabelarycznych, standaryzację struktur danych, tworzenie metadanych, anonimizację danych wrażliwych oraz przygotowanie do udostępnienia poprzez API i repozytorium danych.</p>		
<p>Zdigitalizowane i udostępnione zasoby IHAR - banku genów roślin uprawnych obejmujące kolekcję zasobów genetycznych przechowywanych długoterminowo w formie nasion oraz danych molekularnych. Zakres obejmuje: (1) ok. 60 000 obiektów biologicznych (prób nasion) wraz z dokumentacją paszportową, pochodzeniem, cechami morfologicznymi i agronomicznymi, (2) digitalizację materiału poprzez skanowanie nasion oraz opracowanie morfometryczne, (3) ok. 30 000 genotypów wymagających opracowania cyfrowego zapisu informacji</p>	31-08-2029	90 000 obiektów

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
<p>genetycznej poprzez sekwencjonowanie DNA oraz przygotowanie danych w formatach interoperacyjnych. Dane zostaną opisane metadanymi zgodnymi ze standardami FAIR, powiązane z ontologiami oraz przygotowane do integracji z europejskimi infrastrukturami danych.</p>		
<p>Udostępnione zbiory danych pomiarowych pochodzące z Instytutu Badawczego Leśnictwa (IBL) obejmujące dane teledetekcyjne i inwentaryzacyjne lasów. Zakres obejmuje: (1) zaanotowane chmury punktów LiDAR (ALS) dla różnych typów obszarów, (2) dane z systemów ALS i TLS dla tych samych powierzchni, (3) dane strukturalne drzew (pnie, korony, biomasa), (4) anotacje dla klasyfikacji gatunkowej i analiz fuzji danych (ALS + CIR), (5) dane referencyjne do trenowania modeli AI.</p>	31-08-2029	5 500 000 rekordów pomiarowych
<p>Zdigitalizowane i udostępnione zasoby pochodzące z Instytutu Włókien Naturalnych i Roślin Zielarskich – Państwowego Instytutu Badawczego (IWNiRZ-PIB) obejmujące dane naukowe, środowiskowe i produkcyjne dotyczące roślin włóknistych i zielarskich. Zakres obejmuje: (1) wyniki badań jakości materiału siewnego konopi włóknistych, (2) zasoby banku genów (Linum, Cannabis), (3) dane meteorologiczne, (4) dokumentację produkcji i obrotu materiałem siewnym,</p>	31-08-2029	27000 rekordów pomiarowych + 1000 zdigitalizowanych dokumentów/obiektów

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
(5) bazę infrastruktury sektora, (6) materiały naukowe i wdrożeniowe. Dane będą udostępniane w modelu zróżnicowanego dostępu.		
Zdigitalizowane i udostępnione zasoby pochodzące z Wielkopolskiego Parku Narodowego (WPN) obejmujące dane archiwalne, naukowe i monitoringowe dotyczące środowiska przyrodniczego. Zakres obejmuje: (1) dokumentację archiwalną i kartograficzną, (2) dokumentację planów ochrony i badań, (3) dane z monitoringu przyrodniczego, w tym zdjęcia i nagrania z fotopułapek, (4) dane przestrzenne o siedliskach i gatunkach.	31-08-2029	80 000 digitalizowanych dokumentów i stron oraz skany wielkoformatowe ok. 150 arkuszy
Zdigitalizowane i udostępnione zasoby pochodzące z Karkonoskiego Parku Narodowego (KPN) obejmujące dane archiwalne, środowiskowe i przestrzenne. Zakres obejmuje: (1) archiwalne mapy i operaty planów ochrony, (2) dokumentację siedliskową i leśną, (3) kroniki i materiały historyczne, (4) wyniki badań naukowych, (5) dane meteorologiczne (1961–2002), (6) ortofotomapy i numeryczny model rzeźby terenu (NMRT), (7) digitalizację zdjęć lotniczych.	31-08-2029	65000 obiektów - digitalizacja dokumentów, skany wielkoformatowe ok. 120 arkuszy
Digitalizacja publikacji naukowych obejmujących szeroki zakres tematyczny Instytutu Technologiczno-Przyrodniczego: rolnictwo, nauki o Ziemi i środowisku,	31-08-2029	25 000 stron publikacji – 220 publikacji

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
inżynieria środowiska, w tym budowlę wodne, budowa obwałowań przeciwpowodziowych, zarządzanie zasobami wodnymi, melioracje wodne, jakość wody, ekosystemy łąkowe i pastwiskowe, gospodarowanie na użytkach zielonych, gospodarowanie nawozami, ekosystemy wodne, hydrologia, rybactwo, wykorzystanie odnawialnych źródeł energii, produkcja zwierzęca, chów zwierząt, zrównoważone rolnictwo, rolnictwo ekologiczne, agrometeorologia, agronomia, ochrona gleb, ogrodnictwo, planowanie przestrzenne, zanieczyszczenia biosfery, zmiany klimatu, monitoring środowiska z wykorzystaniem innowacyjnych technologii (m.in. GIS, teledetekcja) oraz analizy na poziomie komórkowym.		

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	02-2027
Repozytorium metadanych i słowników (rejestr metadanych i ontologii)	04-2028
Modyfikacja systemów teleinformatycznych partnerów w zakresie integracji i udostępniania danych (repozytoria danych)	08-2028
Interfejs API do udostępniania danych i metadanych	10-2028
Narzędzia zarządzania danymi (Data Governance System)	04-2029
Narzędzia monitorowania jakości danych (Data Quality System)	04-2029
Narzędzia przygotowania danych dla AI (AI-ready data tools)	04-2029
Pozytywny raport z testów bezpieczeństwa	08-2029

Nazwa produktu	Planowana data wdrożenia
Pozytywny raport z testów wydajności	08-2029
Raport z testów UX	08-2029
System teleinformatyczny DATA4EOSC Open Data System (platforma integracji i udostępniania danych)	08-2029
Materiały informacyjno-promocyjne projektu	08-2029
Materiały szkoleniowe dotyczące przygotowania i udostępniania danych naukowych	08-2029
Przewodnik przygotowania i udostępniania danych naukowych (standardy, procedury, model prawny)	10-2029

### 3. KAMIENIE MIŁOWE

Kamienie milowe	Planowany termin osiągnięcia
KM1 Zakończona dogłębna inwentaryzacja i opracowany szczegółowy projekt systemu, w tym architektura, model semantyczny oraz wykonany test prywatności	2027-04-30
KM2 Przygotowana infrastruktura i wdrożone podstawowe komponenty systemu, w tym API	2027-10-31
KM3 Uruchomiona wersja beta platformy integracji danych	2028-04-30
KM4 Zmodyfikowane systemy teleinformatyczne partnerów i udostępnione pierwsze zbiory danych obejmujące co najmniej $\geq 30\%$ planowanego zakresu	2028-08-31
KM5 Uruchomiony system integracji danych w wersji beta, w tym API do udostępniania danych i metadanych	2028-10-31
KM6 Uruchomione narzędzia do zarządzania, jakości danych, AI oraz udostępnione co najmniej $\geq 90\%$ planowanych danych	2029-04-30
KM7 Osiągnięta pełna operacyjność narzędzi i integracji danych (100% danych), uruchomiony system produkcyjny po pozytywnych testach bezpieczeństwa, wydajności i UX oraz osiągnięta gotowość do promocji i szkoleń	2029-08-31
KM8 Zakończona rzeczowa realizacja projektu i wdrożony model trwałości	2029-10-31

### 4. KOSZTY

#### 4.1. Koszty ogólne projektu wraz ze sposobem finansowania

<b>Całkowity koszt projektu (netto oraz brutto), w tym</b>	Netto 19 512 724,59 zł Brutto 20 800 000,00 zł	
<b>Procent dofinansowania ze środków UE (brutto)</b>	79,71%	
<b>Procent środków z budżetu państwa (brutto)</b>	20,29%	
<b>Podział całkowitego kosztu projektu na poszczególne lata (netto oraz brutto)</b>	2026	Netto 827 233,57 zł Brutto 837 319,45 zł
	2027	Netto 5 525 596,54 zł Brutto 5 826 583,33 zł
	2028	Netto 6 491 548,93 zł Brutto 7 244 869,05 zł
	2029	Netto 6 668 345,55 zł Brutto 6 891 228,17 zł

## 4.2. Wykaz poszczególnych pozycji kosztowych

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
Oprogramowanie	Przygotowanie dokumentacji analitycznej i projektowej, opracowanie architektury rozwiązania, wytworzenie i rozwój oprogramowania platformy DATA4EOSC, repozytorium metadanych i ontologii, systemów integracji danych, API, narzędzi Data Governance, Data Quality i AI-ready data tools, integracja systemów partnerów, digitalizacja danych oraz	15 662 250,00 zł	Koszty obejmują wytworzenie kluczowych produktów projektu umożliwiających integrację, przetwarzanie i udostępnianie danych zgodnie z zasadami FAIR i 5 Star Open Data. Obejmują również digitalizację oraz przygotowanie danych i systemów do interoperacyjnego udostępniania i wykorzystania w analizach oraz AI. Są niezbędne do osiągnięcia celów projektu w zakresie udostępniania informacji sektora publicznego. Koszty obejmują również prace programistyczne, analityczne i eksperckie, tworzenie prototypów, testy wewnętrzne (deweloperskie, jednostkowe, integracyjne i funkcjonalne), OCR, strukturyzację i anotację danych.

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	przygotowanie danych do udostępniania i ponownego wykorzystania.		
Infrastruktura	Zakup i konfiguracja niezbędnych elementów infrastruktury teleinformatycznej obejmujących głównie sprzęt do digitalizacji oraz dodatkowo niezbędną rozbudowę o macierze dyskowe, urządzenia sieciowe, wyposażenie serwerowe oraz oprogramowanie systemowe i narzędziowe związane z utrzymaniem infrastruktury. W ramach projektu planowane jest przede wszystkim wykorzystanie istniejących usług infrastrukturalnych i zasobów środowisk teleinformatycznych PCSS i Partnerów.	896 000,00 zł	Koszty obejmują środki trwałe oraz oprogramowanie niezbędne do przechowywania, przetwarzania, zabezpieczenia i udostępniania danych. Koszty ograniczono do minimum poprzez wykorzystanie istniejącej infrastruktury teleinformatycznej i usług współdzielonych PCSS, bez zakupu wyspecjalizowanej infrastruktury obliczeniowej.
Koszty UX i grafiki	Projektowanie doświadczenia użytkownika (UX), interfejsów użytkownika (UI), dostępności cyfrowej (WCAG), testy z użytkownikami oraz opracowanie warstwy graficznej	145 000,00 zł	Koszty zapewniają użyteczność i dostępność systemu dla różnych grup użytkowników, zgodnie z wymaganiami projektowania zorientowanego na użytkownika. Obejmują badania potrzeb, testy i iteracyjne doskonalenie interfejsu. Są wymagane dla zapewnienia efektywnego wykorzystania udostępnionych danych.

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	platformy i katalogu danych.		
Bezpieczeństwo	Audyty bezpieczeństwa, analiza statyczna kodu, testy podatności, testy prywatności, analiza zgodności z RODO i ustawą o otwartych danych, ocena skutków dla ochrony danych (DPIA) oraz wdrożenie mechanizmów ochrony danych i zabezpieczeń systemowych.	150 000,00 zł	Koszty zapewniają ochronę danych, zgodność z regulacjami prawnymi oraz bezpieczeństwo systemów teleinformatycznych. Obejmują niezależne testy i audyty wymagane dla systemów publicznych. Są kluczowe dla bezpiecznego udostępniania danych, w tym danych wrażliwych.
Wydajność rozwiązań	Testy wydajnościowe systemu, optymalizacja przetwarzania danych, testy obciążeniowe API i usług danych oraz dostosowanie architektury do obsługi dużych wolumenów danych i użytkowników.	250 000,00 zł	Koszty zapewniają stabilność i wydajność systemu przy przetwarzaniu i udostępnianiu dużych zbiorów danych. Obejmują niezależne testy i optymalizację działania platformy. Są niezbędne dla zapewnienia jakości usług cyfrowych i dostępności danych.
Szkolenia	Szkolenia dla partnerów projektu w zakresie zarządzania danymi (Data Governance), standardów FAIR, przygotowania danych do AI, wykorzystania API i narzędzi projektowych.	340 000,00 zł	Koszty zapewniają rozwój kompetencji partnerów w zakresie przygotowania i udostępniania danych. Obejmują szkolenia techniczne i organizacyjne niezbędne do efektywnego wykorzystania produktów projektu. Wspierają osiągnięcie wskaźników dotyczących użytkowników i kompetencji.
Działania informacyjno-promocyjne	Działania informacyjne i promocyjne obejmujące	315 000,00 zł	Koszty wspierają upowszechnienie rezultatów projektu i zwiększenie wykorzystania udostępnionych

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	komunikację projektu, upowszechnianie wyników, organizację wydarzeń, przygotowanie materiałów informacyjnych oraz promocję zbiorów danych i ich wykorzystania.		danych. Obejmują działania skierowane do nauki, administracji i biznesu. Są wymagane dla osiągnięcia efektów w zakresie ponownego wykorzystania danych.
Koszty zarządzania i wsparcia (w tym wynagrodzenia personelu wspomagającego)	Zarządzanie projektem (kierownik projektu, koordynatorzy), wsparcie administracyjne, prawne i finansowe, koordynacja partnerów oraz koszty pośrednie związane z realizacją projektu.	3 041 750,00 zł	Koszty obejmują zarządzanie złożonym projektem wielopodmiotowym oraz zapewnienie jego prawidłowej realizacji zgodnie z wymaganiami UE. Obejmują wsparcie organizacyjne, prawne i finansowe oraz koszty pośrednie. Są niezbędne dla zapewnienia skutecznej realizacji i kontroli projektu.

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

Całkowity koszt utrzymania trwałości projektu (brutto)	1 080 260,00 zł		Źródło finansowania
Podział całkowitego kosztu utrzymania trwałości projektu na poszczególne lata (netto oraz brutto)	2029	20 000,00 zł (brutto) (16 260,00 zł netto)	krajowe środki publiczne - budżet państwa
	2030	246 000,00 zł (brutto) (200 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2031	258 300,00 zł (brutto) (210 000,00 zł netto)	krajowe środki publiczne - budżet państwa
	2032	271 215,00 zł (brutto) (220 500,00 zł netto)	krajowe środki publiczne - budżet państwa
	2033	284 745,00 zł (brutto)	krajowe środki

		(231 500,00 zł netto)	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
Brak dostępności wykwalifikowanej kadry w terminach wymaganych harmonogramem	Duża	Znikome	Zaangażowanie szerokiego zespołu projektowego od samego początku projektu, zaangażowanie dodatkowych ekspertów w jednostkach uczestniczących lub rekrutacja
Niedoszacowanie pracochłonności digitalizacji i przygotowania zasobów do publikacji	Średnia	Niskie	Metodyczne planowanie i regularny przegląd prac, zaangażowanie dodatkowych osób do procesu digitalizacji
Opóźnienie realizacji wynikające z przedłużającej się procedury związanej z PZP	Średnia	Niskie	Rozpoczęcie procedur PZP w jak najwcześniejszym etapie projektu
Nieosiągnięcie zakładanych wskaźników dotyczących digitalizacji i udostępnienia zasobów na dany kamień milowy	Średnia	Niskie	Cykliczny monitoring osiągnięcia wskaźników, w celu wczesnego wykrycia potencjalnych odstępstw. Harmonogram szczegółowy uwzględniający ścieżki krytyczne.
Rozbieżność modeli danych, metadanych i	Średnia	Średnie	Użycie narzędzi do konwersji danych pomiędzy formatami, użycie narzędzi AI do augmentacji danych

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
słowników pomiędzy partnerami projektu			

## 5.2. Ryzyka wpływające na utrzymanie efektów

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Utrata lub uszkodzenie danych związane z długoterminowym przechowywaniem	Duża	Niskie	Wdrożenie polityki zarządzania danymi z replikacją danych
Niedostępność danych, poprzez udostępnioną platformę, wskutek błędów programistycznych lub wystąpienia awarii infrastruktury IT	Duża	Średnie	Włączenie platformy do systemu monitorowania usług NOC(24/7) PCSS w celu wczesnego wykrycia awarii, korzystanie z systemów o wysokiej dostępności, utrzymanie zespołu reagowania i programistycznego, podłączenie do istniejącego systemu ticketowego dla raportowania błędów, przestrzeganie procedur ISO (9001, 27001)
Zagrożenia cyberbezpieczeństwa: atak na repozytorium cyfrowe, naruszenie integralności danych,	Duża	Wysokie	Wykorzystanie zaawansowanych mechanizmów ochrony systemów i danych, cykliczne audyty bezpieczeństwa, backup z kopią offline, cykliczne testy przywracania danych z kopii, okresowa weryfikacja spójności danych
Brak środków na utrzymanie efektów projektu	Duża	Niskie	Przewidziano różnorodne źródła finansowania utrzymania efektów projektu.
Nieciągłość finansowania infrastruktury (np. wzrost kosztów energii, usług IT)	Duża	Niskie	Przewidziano różnorodne źródła finansowania utrzymania efektów projektu.
Brak możliwości	Średnia	Średnie	Budowa kompetencji wewnętrznych

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
zatrudnienia osób o odpowiednich kompetencjach niezbędnych do utrzymania efektów projektu			partnerów, szkolenia, dokumentacja techniczna oraz wykorzystanie istniejących zespołów IT i danych.
Brak wystarczających zasobów kadrowych do utrzymania efektów projektu	Średnia	Niskie	Określenie modelu utrzymania systemu i odpowiedzialności partnerów już na etapie realizacji projektu.
Zmiany przepisów dotyczących otwartych danych i ochrony danych	Średnia	Niskie	Stały monitoring zmian prawnych oraz utrzymanie wsparcia prawnego i organizacyjnego.
Ryzyko ograniczonej interoperacyjności z zewnętrznymi systemami i przestrzeniami danych	Średnia	Średnie	Stosowanie otwartych standardów, API oraz architektury modularnej i federacyjnej.
Nieosiągnięcie wszystkich zaplanowanych korzyści projektu	Średnia	Średnie	Zaplanowano działania promocyjne, szkoleniowe oraz rozwój interoperacyjnych API i mechanizmów ponownego wykorzystania danych.

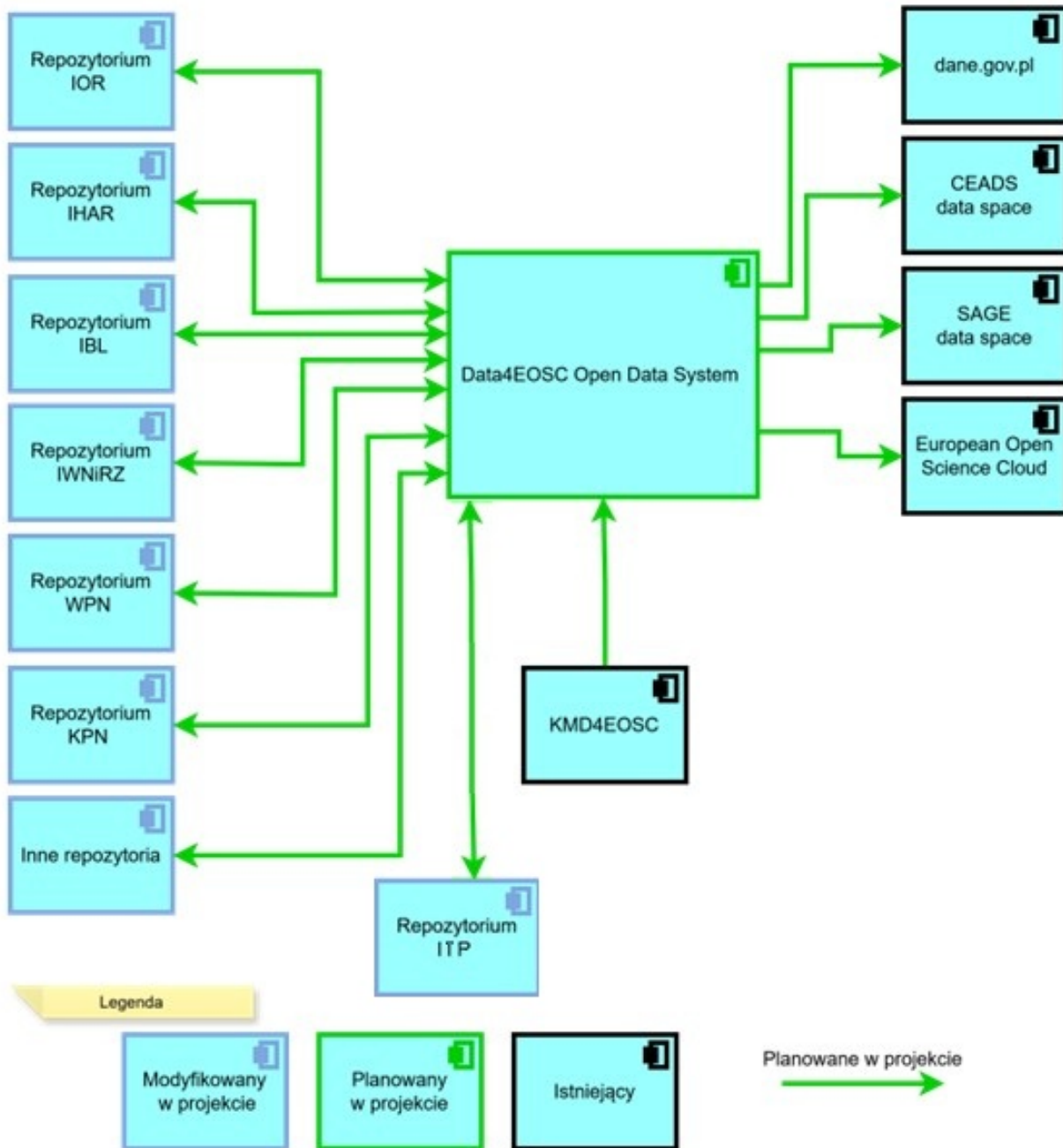
## 6. OTOCZENIE PRAWNE

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	TAK/NIE		
2	Rozporządzenie Rady Ministrów z dnia 21 maja 2024 r. w sprawie Krajowych Ram Interoperacyjności	TAK/NIE		
3	Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym	TAK/NIE		

Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
	wykorzystywaniu informacji sektora publicznego			
4	Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych	TAK/NIE		
5	Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych	TAK/NIE		
6	Ustawa z dnia 4 kwietnia 2019 r. o dostępności cyfrowej stron internetowych i aplikacji mobilnych podmiotów publicznych	TAK/NIE		
7	Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych wraz z Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. (RODO/GDPR)	TAK/NIE		
8	Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego (Open Data Directive)	TAK/NIE		
9	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/868 z dnia 30 maja 2022 r. w sprawie europejskiego zarządzania danymi (Data Governance Act)	TAK/NIE		
10	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 (Data Act)	TAK/NIE		
11	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 (AI Act),	TAK/NIE		
12	Dyrektywa Parlamentu Europejskiego i Rady (UE) 2022/2555 z dnia 14 grudnia 2022 r. w sprawie środków na rzecz wysokiego wspólnego poziomu cyberbezpieczeństwa na terytorium Unii (NIS2)	TAK/NIE		
13	Ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa wraz ze zmianami z dnia 23 stycznia 2026 r.	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	Data4EOSC Open Data System	PCSS	Federacyjny system służący integracji, katalogowaniu, udostępnianiu i wyszukiwaniu otwartych danych naukowych partnerów projektu. Zapewnia interfejs	Planowany	Budowa i wdrożenie platformy integrującej.

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			użytkownika dostawcy danych, katalog danych, API, mechanizmy publikacji metadanych, model informacyjny (AIM) i warstwę interoperacyjności semantycznej, mapowanie schematów i standaryzację danych, komponenty do zarządzania jakością danych, pochodzeniem i cyklem życia danych oraz komponenty wspierające anotację, enrichment danych i przygotowanie zbiorów treningowych dla AI.		
2	CEADS + SAGE	PCSS	Komponenty europejskich przestrzeni danych dla rolnictwa i gran deal wspierające interoperacyjność, katalogowanie i wymianę danych zgodnie z architekturą data spaces.	Istniejący	Integracja z europejskimi przestrzeniami danych. Reużycie technologii EDC.
3	KDM4EOSC Services	PCSS	Usługi repozytoryjne i zarządzania danymi naukowymi, obejmujące przechowywanie, identyfikatory trwałe i publikację danych.	Istniejący	Wykorzystanie istniejącej infrastruktury jako warstwy przechowywania i publikacji wybranych zbiorów danych.
4	Repozytorium IOR	IOR	Repozytorium danych naukowych i monitoringowych z zakresu ochrony roślin, doświadczeń i modeli prognostycznych	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
5	Repozytorium IHAR	IHAR	Repozytorium danych hodowlanych, genotypowych, fenotypowych i zasobów	Modyfikowany	Integracja danych i metadanych z platformą

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			genowych.		centralną oraz dostosowanie do standardów.
6	Repozytorium IBL	IBL	Instytucjonalne repozytorium danych otwartych Instytutu Badawczego Leśnictwa	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
7	Repozytorium WPN	WPN	Repozytorium danych środowiskowych, monitoringowych i bioróżnorodności.	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
8	Repozytorium IWNiRZ	IWNiRZ	Repozytorium danych dotyczących roślin zielarskich, włóknistych i analiz laboratoryjnych.	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
9	Repozytorium KPN	KPN	Repozytorium danych środowiskowych, monitoringowych i bioróżnorodności.	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
10	Repozytorium ITP	ITP	Instytucjonalne repozytorium danych otwartych Instytutu Technologiczno-Przyrodniczego	Modyfikowany	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.
11	dane.gov.pl	Ministerstwo Cyfryzacji	Centralny portal danych publicznych służący publikacji i udostępnianiu	Istniejący	Integracja platformy centralnej w

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			metadanych zbiorów informacji sektora publicznego.		zakresie automatycznej publikacji i synchronizacji metadanych zbiorów danych projektu.
12	EOSC Portal	European Open Science Cloud	Europejski katalog usług i zasobów naukowych umożliwiający publikację i wyszukiwanie danych badawczych.	Istniejący	Integracja w zakresie publikacji metadanych i zwiększenia widoczności danych projektu na poziomie europejskim
13	Inne repozytoria	Uczelnie wyższe i Instytuty naukowe	Repozytorium danych środowiskowych	Istniejący	Integracja danych i metadanych z platformą centralną oraz dostosowanie do standardów.

## Lista przepływów

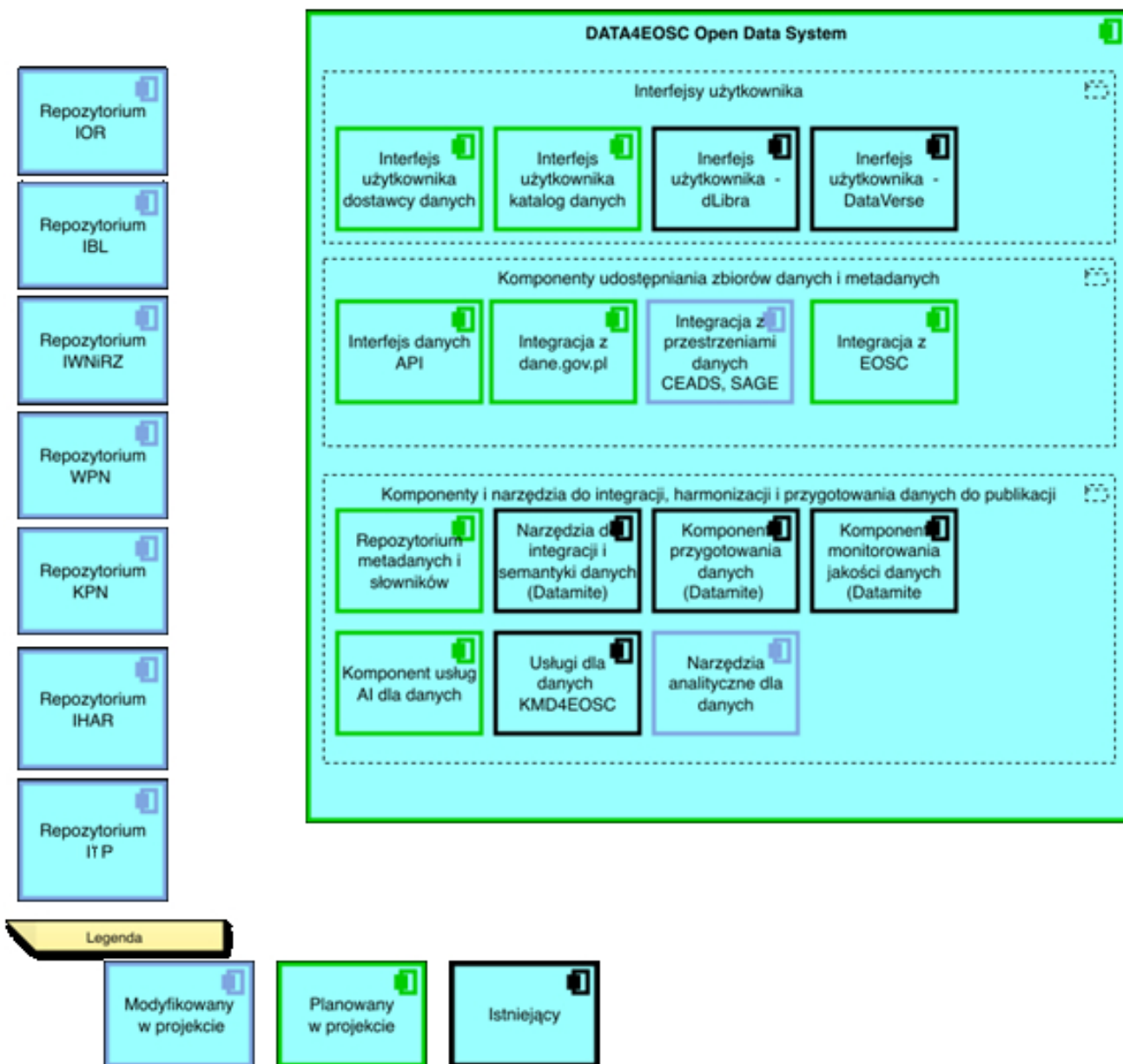
Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
1	Data4EOS C	dane.gov.pl	Publikacja metadanych i bezpośredniego interfejsu dostępowego	kopiowanie danych	realizowalny inną metodą	REST API
2	Data4EOS C	EOSC (European Open Science Cloud)	Publikacja metadanych i bezpośredniego interfejsu dostępowego	kopiowanie danych	realizowalny inną metodą	REST API
3	Data4EOS C	CEADS data space	Publikacja metadanych i bezpośredniego interfejsu	kopiowanie danych	realizowalny inną metodą	REST API

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			dostępowego			
4	Data4EOS C	SAGE data space	Publikacja metadanych i bezpośredniego interfejsu dostępowego	kopiowanie danych	realizowalny inną metodą	REST API
5	KMD4EOS C	Data4EOS C	Wymiana, harmonizacja i przygotowanie metadanych	kopiowanie danych	realizowalny inną metodą	REST API
6	Repozytorium IOR	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
7	Repozytorium IBL	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
8	Repozytorium WPN	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
9	Repozytorium KPN	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
10	Repozytorium ITP	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
11	Repozytorium IHAR	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
12	Repozytorium IWNiRZ	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
13	Inne repozytoria	Data4EOS C	Metadane obiektów i cyfrowe obiekty	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
14	Data4EOS C	Repozytorium IOR	Zharmonizowane metadane, status walidacji,	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			identyfikatory zasobów, raporty jakości danych			
15	Data4EOS C	Repozytorium IHAR	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
16	Data4EOS C	Repozytorium IBL	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
17	Data4EOS C	Repozytorium IWNiRZ	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
18	Data4EOS C	Repozytorium WPN	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
19	Data4EOS C	Repozytorium KPN	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API
20	Data4EOS C	Repozytorium ITP	Zharmonizowane metadane, status	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			walidacji, identyfikatory zasobów, raporty jakości danych			
21	Data4EOS C	Inne repozytoria	Zharmonizowane metadane, status walidacji, identyfikatory zasobów, raporty jakości danych	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API

## 7.2. Kluczowe komponenty architektury rozwiązania



### 7.3. Przyjęte założenia technologiczne

Lp.	Obszar	Założenie technologiczne
1.	Infrastruktura	Wykorzystanie istniejącej infrastruktury PCSS, infrastruktura dostosowana do założeń rozproszonego systemu wielu współpracujących ze sobą serwisów. Adaptacja oraz uruchamianie skonteneryzowanych aplikacji na zwirtualizowanych zasobach udostępnionych w ramach podejść Infrastructure-as-a-Service oraz Platform-as-a-Service. Zastosowane frameworki: Openstack oraz OpenShift. Komponenty systemu zostaną zlokalizowane w centrum danych PCSS, wyposażonym w mechanizmy kontroli dostępu fizycznego, ochrony przeciwpożarowej, zapewniające właściwe warunki pracy sprzętu (temperatura, wilgotność) oraz odpowiednie zasilanie awaryjne, monitorowanym przez zespół NOC (ang.

Lp.	Obszar	Założenie technologiczne
		Network Operations Center) w trybie 24/7/365
2.	Sieć i bezpieczeństwo	Zastosowanie uwierzytelniania, autoryzacji i szyfrowanie komunikacji pomiędzy modułami. Zastosowanie standardu OpenID Connect zarówno do obsługi użytkowników, ale również granularna kontrola wszystkich aplikacji klienckich używanych w ramach systemu
3.	Standardy wymiany danych	<p>Do asynchronicznej wymiany danych między elementami rozproszonego systemu wykorzystana będzie platforma Apache Kafka wraz z repozytorium modeli wykorzystywanych przez system. Rekomendowane formaty to JSON oraz AVRO. W przypadku bezpośredniej interakcji między serwisami zastosowane są interfejsy programistycznych spełniających założenia standardu RESTful (REST API). Kontrola dostępu do serwisów odbywa się za pomocą standardu OIDC (otwarte oprogramowanie Keycloak).</p> <p>Projekt adresuje wszystkie poziomy interoperacyjności zgodnie z Europejskimi Ramami Interoperacyjności, obejmując – interoperacyjność prawną, organizacyjną, semantyczną i techniczną. Szczególny nacisk położony jest na interoperacyjność semantyczną jako kluczowy czynnik umożliwiający integrację danych między domenami.</p> <p>Rdzeniem architektury danych systemu będzie wspólny model semantyczny oparty na Agriculture Information Model (AIM) opracowany pierwotnie w projekcie Demeter i aktualnie wpierany na poziomie EU przez PCSS (między innymi w europejskich przestrzeniach danych takich jak CEADS i SAGE), który zostanie rozszerzony na domeny środowiskowe, leśne i bioróżnorodności. Model ten zapewni spójny opis danych oraz umożliwi mapowanie schematów z heterogenicznych systemów źródłowych.</p> <p>Rozszerzenie AIM będzie realizowane poprzez modularne ontologie dziedzinowe, takie jak SOSA/SSN, SKOS oraz Darwin Core, zapewniające skalowalność i interoperacyjność międzysektorową.</p> <p>Warstwa metadanych systemu będzie oparta na standardzie DCAT-AP oraz jego rozszerzeniach (GeoDCAT-AP dla danych przestrzennych oraz DQV dla reprezentacji jakości danych), zapewniając interoperacyjność z krajowymi i europejskimi katalogami danych, w tym dane.gov.pl oraz EOSC. Standardy te stanowią podstawę warstwy metadanych modelu AIM.</p>
4.	Systemy operacyjne serwerowe	Unix/Linux
5.	Bazy danych	Dzięki zastosowaniu architektury rozproszonej sugerowane jest podejście polyglot persistence, gdzie każdy serwis decyduje o najoptymalniejszym rodzaju przechowywanych danych z uwagi na zastosowanie i przeznaczenie samej aplikacji. W ramach repozytoriów systemu oferowane jest przechowywanie danych w następujących podsystemach: tematach Apache Kafka (o nieskończonej retencji); relacyjnych bazach danych MySQL i PostgreSQL; nierelacyjne bazy danych (MongoDB), bazach typu timeseries InfluxDB; repozytoriach KV (key-value) Redis oraz

Lp.	Obszar	Założenie technologiczne
		storage w standardzie S3 (MinIO).
6.	Serwery aplikacji	Podejście mikroserwisowe pozwala na budowanie, osadzenie oraz uruchamianie skonteneryzowanych aplikacji (np. jako obrazy Docker) na frameworkach Openstack/OpenShift z wykorzystaniem wielorakich rozwiązań serwerów aplikacji, głównie planowane jest wykorzystanie Nginx/Redis.
7.	Portale	Proces projektowania aplikacji klienckich zakłada zastosowanie wytycznych ujętych w międzynarodowym standardzie dostępności treści internetowych Web Content Accessibility Guidelines (WCAG 2.2).
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

PCSS w projekcie jako operator infrastruktury centralnej posiada SZBI zgodny z normą ISO 27001:2017 i świadczy usługi cyberbezpieczeństwa i konsultingu ICT w zgodzie z normą ISO 9001:2015. Zarządzanie bezpieczeństwem informacji w niniejszym projekcie również będzie oparte na ww. normach.

Infrastruktura centralna działać będzie w dedykowanej podsieci logicznej i będzie skonfigurowana zgodnie z zasadą minimalnych przywilejów oraz chroniona za pomocą systemów klasy NGFW (frewall nowej generacji) Palo Alto z aktywną licencją Advanced Threat Prevention (zaawansowanej ochrony przed zagrożeniami). Planowane jest również niezależne wdrożenie i wykorzystanie w projekcie systemów klasy WAF do ochrony interfejsów Web oraz SIEM do zarządzania całokształtem informacji związanej z bezpieczeństwem. Infrastruktura centralna, podłączona do szerokopasmowej sieci dla nauki PIONIER, będzie chroniona przez zespół PIONIER CERT i lokalnych specjalistów SOC, co zapewni monitorowanie bezpieczeństwa sieci i obsługę incydentów.

Infrastruktura centralna projektu chroniona będzie zgodnie z paradygmatami ochrony dogłębnej (ang. Defence-in-depth) i bezpiecznego cyklu życia oprogramowania (ang. SDLC), co przekłada się na szereg proponowanych działań. M.in. dla architektury projektu zrealizowany zostanie proces modelowania zagrożeń, oparty na podejściach STRIDE oraz DREAD. Kod źródłowy aplikacji (usług Web, mobilnych i interfejsów API) utworzonych w ramach projektu, a także jego infrastruktura widziana jako całość, będą poddane wewnętrznym i zewnętrznym audytom bezpieczeństwa. Wybór zewnętrznej infrastruktury chmurowej podyktowany będzie m.in. właściwymi gwarancjami bezpieczeństwa jej operatora (w tym również zgodnością z GDPR/

RODO). Do poszczególnych prac użyte będą właściwe standardy ogólne (w szczególności wytyczne Narodowych Standardów Cyberbezpieczeństwa) i techniczne (np. OWASP Application Security Verification Standard).

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