

# OPIS ZAŁOŻEŃ PROJEKTU INFORMATYCZNEGO

<b>Tytuł projektu</b>	e-Puszcza 2.0 – rozbudowa funkcjonalna środowiska Open Forest Data jako repozytorium przyrodniczych danych naukowych i powiązanych komponentów wraz z pracami digitalizacyjnym		
<b>Wnioskodawca</b>	Minister Nauki i Szkolnictwa Wyższego		
<b>Beneficjent</b>	Instytut Biologii Ssaków Polskiej Akademii Nauk		
<b>Partnerzy</b>	Uniwersytet Warszawski		
<b>Źródło finansowania</b>	Europejski Fundusz Rozwoju Regionalnego Fundusze Europejskie na Rozwój Cyfrowy 2021-2027 FERC.02 Zaawansowane usługi cyfrowe FERC.02.03 Cyfrowa dostępność i ponowne wykorzystanie informacji Środki zaplanowane w budżecie części dysponenta: część budżetowa nr 27 - informatyzacja		
<b>Całkowity koszt projektu</b>	5 435 292,00 zł		
<b>Planowany okres realizacji projektu</b>	12-2026 do 12-2029		
<b>Osoba kontaktowa</b>	Joanna Łapińska	jlapinska@ibs.bialowieza.pl	692179156

## 1. POWODY PODJĘCIA PROJEKTU

### 1.1. Identyfikacja problemu i potrzeb

W pkt 1.1 proponuje się zastąpić pierwszy akapit następującym brzmieniem:

Projekt odpowiada na problemy wynikające z ograniczeń istniejącego portalu/repozytorium Open Forest Data, planowanego do rozwinięcia w ramach projektu e-Puszcza 2.0. Portal obejmuje trzy zasadnicze komponenty: stronę wejściową www Open Forest Data, repozytorium danych Open Forest Data oparte o oprogramowanie Dataverse oraz portal GIS.OpenForestData.pl oparty o GeoNode. Przez repozytorium danych rozumie się główny system repozytoryjny tj. Open Forest Data(Dataverse) służący do deponowania, opisu metadanymi, wersjonowania, publikowania, cytowania i pobierania datasetów. Przez komponent mapowy rozumie się portal GIS Open Forest Data oparty o GeoNode, służący do publikowania, przeglądania i udostępniania map, warstw przestrzennych i usług geoinformacyjnych.

Obecnie brakuje pełnej integracji repozytorium danych Open Forest Data z portalem GIS Open Forest Data, automatycznego importu i walidacji danych, spójnego procesu opisu metadanymi oraz wygodnych mechanizmów identyfikacji autorów i użytkowników. Część cennych zasobów przyrodniczych IBS PAN i Uniwersytetu Warszawskiego pozostaje w bazach lokalnych, arkuszach, dokumentacji analogowej albo w postaci wymagającej dalszej standaryzacji. Projekt obejmuje modyfikację trzech systemów: repozytorium danych Open Forest Data (Dataverse), portalu GIS Open Forest Data (GeoNode).

<b>Interesariusz</b>	<b>Zidentyfikowany problem</b>	<b>Szacowana wielkość grupy</b>
Pracownicy jednostek	Brak w pełni zautomatyzowanego procesu	80

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
konsorcjum odpowiedzialni za dane, kolekcje, zielniki, dokumentację archiwalną badań naukowych i utrzymanie platformy	importu, walidacji, opisu i publikacji danych z uporządkowanych plików źródłowych i roboczych baz danych, w szczególności arkuszy Excel/CSV. Ograniczona integracja repozytorium Dataverse z portalem GIS. Nadmierna pracochłonność ręcznego dodawania datasetów i aktualizacji metadanych. Ryzyko niespójności opisów, duplikacji rekordów oraz utraty wartości danych pozostających w lokalnych plikach, bazach roboczych, archiwach i kolekcjach. Potrzeba dalszej digitalizacji i zabezpieczenia zasobów, w tym zasobów zielnikowych, entomologicznych, genetycznych, 3D i mikrotomograficznych.	
Naukowcy i doktoranci prowadzący badania z zakresu biologii, ekologii, leśnictwa, geografii, ochrony przyrody i nauk o środowisku	Ograniczony dostęp do kompletnych, dobrze opisanych, cytowalnych i wersjonowanych danych przyrodniczych. Brak wygodnej ścieżki od datasetu do mapy, analizy i ponownego użycia w narzędziach R/Python. Niewystarczająca dostępność danych archiwalnych, długoterminowych, surowych, kolekcyjnych i zielnikowych. Brak pełnego powiązania lokalnych zasobów z międzynarodowymi bazami bioróżnorodności, w tym GBIF. Brak logowania przez ORCID i dostępu instytucjonalnego.	3000
Studenci kierunków przyrodniczych, leśnych, środowiskowych, geograficznych i informatycznych	Ograniczony dostęp do uporządkowanych danych edukacyjnych i badawczych, które można wykorzystać w pracach dyplomowych, ćwiczeniach GIS, analizach statystycznych i projektach data science. Brak prostych przykładów użycia datasetów, map i analiz online. Zbyt wysoka bariera wejścia w pracę z repozytorium naukowym.	5000
Nauczyciele, edukatorzy przyrodniczy, nauczyciele akademicki i popularyzatorzy nauki	Brak łatwo dostępnych, wiarygodnych i atrakcyjnych materiałów opartych na rzeczywistych danych naukowych. Ograniczona liczba prostych wizualizacji, map tematycznych, opisów i scenariuszy użycia danych dla osób niebędących specjalistami od repozytoriów lub GIS.	2000
Jednostki administracji publicznej wykonujące zadania w obszarze ochrony przyrody, środowiska, leśnictwa, planowania	Ograniczony dostęp do aktualnych i historycznych danych przyrodniczych oraz przestrzennych potrzebnych do planowania ochrony przyrody, ocen oddziaływania na środowisko, opracowań strategicznych, dokumentów planistycznych i decyzji środowiskowych. Brak jednego punktu	500

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
przestrzennego i zarządzania zasobami przyrodniczymi, w tym parki narodowe, parki krajobrazowe, RDOŚ/ GDOŚ, Lasy Państwowe i jednostki samorządu terytorialnego	dostępu łączącego dane źródłowe, metadane, mapy, warstwy GIS i informacje o ograniczeniach wykorzystania danych wrażliwych przyrodniczo.	
Organizacje pozarządowe działające w obszarze ochrony przyrody, edukacji ekologicznej i kontroli społecznej	Ograniczony dostęp do wiarygodnych, zrozumiałych i możliwych do ponownego wykorzystania danych o środowisku. Brak prostych narzędzi umożliwiających analizę zmian przyrodniczych, wykorzystanie map i udział w konsultacjach społecznych na podstawie danych naukowych.	30
Przedsiębiorcy wykorzystujący dane środowiskowe, w tym firmy wykonujące ekspertyzy, analizy GIS, usługi turystyczne, wydawnicze i edukacyjne	Brak łatwego dostępu do danych przyrodniczych o jasnym statusie licencyjnym, możliwych do pobrania, cytowania i wykorzystania w usługach, opracowaniach, ekspertyzach oraz produktach edukacyjnych lub turystycznych. Ograniczona dostępność danych przestrzennych w formatach gotowych do dalszego przetwarzania.	400
Obywatele zainteresowani przyrodą, mieszkańcy regionu, przewodnicy terenowi i uczestnicy nauki obywatelskiej	Wysoka bariera wejścia do korzystania z naukowych danych przyrodniczych. Brak prostych map, dashboardów, opisów i wizualizacji pokazujących zasoby przyrodnicze oraz zmiany środowiska w sposób zrozumiały dla użytkowników nieprofesjonalnych. Ograniczone możliwości kontrolowanego włączania obserwacji społecznych i wolontariatu naukowego.	10000
Instytucje krajowe i międzynarodowe agregujące dane o bioróżnorodności oraz otwartej nauce, w tym GBIF, EOSC i repozytoria dziedzinowe	Brak pełnego, spójnego i wystandaryzowanego przepływu danych między lokalnymi bazami kolekcyjnymi a międzynarodowymi systemami danych. Ryzyko duplikacji rekordów, niespójnych identyfikatorów, niepełnych metadanych i niewystarczającej widoczności polskich zasobów przyrodniczych w obiegu międzynarodowym.	20
Jednostki biorące udział w projekcie: Instytut Biologii Ssaków PAN oraz Uniwersytet Warszawski, w tym	Brak jednolitego procesu digitalizacji, importu, walidacji, publikacji i aktualizacji rozproszonych zasobów przyrodniczych. Niewystarczająca automatyzacja tworzenia datasetów, kontroli jakości metadanych, obsługi danych wrażliwych, integracji z GIS i	70

Interesariusz	Zidentyfikowany problem	Szacowana wielkość grupy
Stacja Geobotaniczna Wydziału Biologii UW	przygotowania danych do standardów Darwin Core oraz GBIF/IPT.	

## 1.2. Opis stanu obecnego

Obecnie środowisko repozytorium Open Forest Data obejmuje stronę wejściową Open Forest Data, repozytorium danych Open Forest Data oparte o Dataverse oraz portal GIS Open Forest Data oparty o GeoNode. Strona wejściowa pełni funkcję punktu dostępu do zasobów i przekierowuje użytkowników do repozytorium danych oraz portalu GIS. Repozytorium danych Open Forest Data umożliwia publikację datasetów, opis metadanymi, DOI, wersjonowanie, pobieranie danych i dostęp przez API. Portal GIS Open Forest Data umożliwia publikację, przeglądanie i wizualizację warstw przestrzennych oraz map.

Repozytorium zawiera obecnie ponad 79 tys. datasetów. Procesy importu, walidacji, standaryzacji metadanych oraz powiązania datasetów z mapami są jednak w dużej mierze ręczne albo półautomatyczne. Istniejące systemy wymagają rozbudowy w zakresie automatycznego importu danych, kontroli jakości danych i metadanych, lepszej integracji repozytorium z portalem GIS, rozwoju API, widoków tematycznych, dashboardów oraz dokumentacji dla użytkowników.

Niniejszy projekt wykorzysta istniejące produkty i infrastrukturę środowiska Open Forest Data oraz istniejące zasoby serwerowe i techniczne wykorzystywane obecnie do utrzymania repozytorium, bez migracji zasobów. Zakres digitalizacji i udostępniania projektu obejmie tylko nowe zasoby zielnikowe, zoologiczne, genetyczne, przestrzenne, 3D, mikrotomograficzne oraz historyczne i długoterminowe dane środowiskowe (przykładowo obecnie w repozytorium utworzono 60 tys. datasetów opisujących okazy zoologiczne IBS PAN, a planowane jest opisanie kolejnych 120 tys. okazów, tj. prawie 100% kolekcji zoologicznej IBS PAN, bez ryzyka dublowania prac. W odniesieniu do zasobów Stacji UW są to dane dotychczas w ogóle w żadnej części nie zdigitalizowane i nie udostępnione)

## 2. EFEKTY PROJEKTU

### 2.1. Cele i korzyści wynikające z projektu

<b>Cel - 1</b>	Rozwój repozytorium Open Forest Data w ramach projektu e-Puszcza 2.0 poprzez modyfikację głównego repozytorium danych (Dataverse), modyfikację portalu GIS Open Forest Data (GeoNode), integrację repozytorium z portalem GIS oraz rozwój strony wejściowej Open Forest Data (jako serwisu www informacyjno-nawigacyjnego), w celu udostępniania, wizualizacji, analizy i ponownego wykorzystania przyrodniczych zasobów nauki.
<b>Cel strategiczny</b>	Cel wpisuje się w następujące cele dokumentów strategicznych: - Program Fundusze Europejskie na Rozwój Cyfrowy 2021–2027, Priorytet FERC.02 „Zaawansowane usługi cyfrowe”, Działanie FERC.02.03 „Cyfrowa dostępność i ponowne wykorzystanie informacji”, cel szczegółowy EFRR.CP1.II „Czerpanie korzyści z cyfryzacji dla obywateli, przedsiębiorstw, organizacji badawczych i instytucji publicznych”. - Program otwierania danych na lata 2021–2027, w szczególności: Cel 2 „Poprawa interoperacyjności i jakości danych”, Cel 3 „Wzrost wykorzystywania i wymiany danych” oraz Cel 4 „Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych”.

	<ul style="list-style-type: none"> <li>- Polityka Naukowa Państwa, w zakresie otwartej nauki, otwartego dostępu do danych badawczych oraz zwiększania wykorzystania zasobów nauki.</li> <li>- Polityka Ekologiczna Państwa 2030, w zakresie celów horyzontalnych: „Środowisko i administracja. Poprawa efektywności funkcjonowania instrumentów ochrony środowiska” oraz „Środowisko i edukacja. Rozwijanie kompetencji ekologicznych społeczeństwa”.</li> <li>- Strategia Produktywności 2030, w zakresie rozwoju gospodarki opartej na danych, zwiększania wykorzystania danych i wzmacniania cyfrowych zasobów wiedzy.</li> </ul>
<b>Korzyść:</b>	Zwiększenie wykorzystania przyrodniczych danych naukowych przez naukę, administrację, edukację, ochronę przyrody, organizacje społeczne i gospodarkę.
<b>KPI:</b>	<p>KPI 1: Liczba podmiotów wspartych w zakresie rozwoju usług, produktów i procesów cyfrowych.</p> <p>KPI 2: Liczba podmiotów, które udostępniły informacje sektora publicznego/dane prywatne on-line.</p> <p>KPI 3: Użytkownicy nowych i zmodernizowanych publicznych usług, produktów i procesów cyfrowych.</p>
<b>Wartość aktualna i docelowa KPI:</b>	<p>KPI 1: wartość aktualna: 0</p> <p>KPI 2: wartość aktualna: 0</p> <p>KPI 3: wartość aktualna: 748</p> <p>KPI 1: wartość docelowa: 2</p> <p>KPI 2: wartość docelowa: 2</p> <p>KPI 3: wartość docelowa: 1048</p>
<b>Metoda pomiaru KPI</b>	<p>KPI 1: pomiar ilościowy na podstawie dokumentacji projektowej, umowy partnerskiej, protokołów odbioru produktów oraz raportu końcowego. Źródło danych: dokumentacja projektowa, protokoły odbioru, raport końcowy. Częstotliwość pomiaru: jednorazowo na zakończenie projektu.</p> <p>KPI 2: pomiar ilościowy na podstawie potwierdzenia publicznego udostępnienia zasobów przez IBS PAN oraz Uniwersytet Warszawski. Źródło danych: rejestry publikacji w Dataverse/Open Forest Data, raport z udostępniania zasobów, protokół odbioru końcowego. Częstotliwość pomiaru: kwartalnie oraz na zakończenie projektu.</p> <p>KPI 3: pomiar ilościowy liczby użytkowników nowych i zmodernizowanych produktów cyfrowych. Źródło danych: logi systemowe, statystyki API i raport analityczny zgodny z zasadami ochrony prywatności. Częstotliwość pomiaru: kwartalnie po wdrożeniu oraz w okresie 12 miesięcy od zakończenia projektu.</p>
<b>Cel - 2</b>	Integracja repozytorium Dataverse/Open Forest Data z portalem GIS, umożliwiającą bezpośrednio powiązanie datasetów z mapami, warstwami przestrzennymi i usługami wizualizacji danych.
<b>Cel strategiczny</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027 oraz w cele Programu otwierania danych dotyczące interoperacyjności, jakości danych i wzrostu ich wykorzystywania. Integracja repozytorium z komponentem GIS zwiększa wartość użytkową udostępnianych zasobów nauki i środowiska.
<b>Korzyść:</b>	Usprawnienie wykorzystania danych przestrzennych w badaniach, ochronie przyrody i decyzjach środowiskowych.
<b>KPI:</b>	<p>KPI 1: Liczba datasetów lub warstw przestrzennych powiązanych z funkcją mapową.</p> <p>KPI 2: Rozmiar informacji sektora publicznego lub danych prywatnych</p>

	udostępnionych w komponentach mapowych.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 2: wartość aktualna: 0 MB KPI 1: wartość docelowa: 90 KPI 2: wartość docelowa: 250 MB
<b>Metoda pomiaru KPI</b>	KPI 1: pomiar ilościowy na podstawie liczby datasetów, map lub warstw GIS, dla których wdrożono powiązanie dataset–mapa, funkcję „Zobacz na mapie” albo link zwrotny mapa–dataset. Źródło danych: raport z Dataverse/Open Forest Data, raport z GeoNode, protokół odbioru integracji Dataverse–GIS. Częstotliwość pomiaru: kwartalnie po wdrożeniu integracji oraz jednorazowo na zakończenie projektu. KPI 2: pomiar ilościowy na podstawie rozmiaru danych przestrzennych udostępnionych on-line w powiązaniu z komponentem GIS. Źródło danych: raport z GeoNode, raport z repozytorium, protokół odbioru integracji. Częstotliwość pomiaru: jednorazowo na zakończenie projektu.
<b>Cel - 3</b>	Zdigitalizowanie i udostępnienie nowych zasobów naukowych partnerów projektu, w tym danych kolekcyjnych, zielnikowych, entomologicznych, genetycznych, archiwalnych, długoterminowych, przestrzennych, 3D i mikrotomograficznych.
<b>Cel strategiczny</b>	Cel wpisuje się w typ projektu „Cyfrowe udostępnienie zasobów nauki” oraz w cele Programu otwierania danych dotyczące zwiększenia dostępności danych, poprawy ich jakości i stymulowania ponownego wykorzystywania. Digitalizacja obejmuje zasoby dotychczas niedostępne lub trudno dostępne cyfrowo.
<b>Korzyść:</b>	Zwiększenie dostępności unikalnych zasobów naukowych dotychczas niedostępnych lub trudno dostępnych cyfrowo.
<b>KPI:</b>	KPI 1: Liczba zdigitalizowanych dokumentów zawierających informacje sektora publicznego/dane prywatne. KPI 2: Liczba udostępnionych on-line dokumentów zawierających informacje sektora publicznego/dane prywatne. KPI 3: Rozmiar zdigitalizowanych informacji sektora publicznego/danych prywatnych. KPI 4: Rozmiar udostępnionych on-line informacji sektora publicznego/danych prywatnych. KPI 5: Odsetek skontrolowanych próbek digitalizacji spełniających wymagania jakości technicznej.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 2: wartość aktualna: 0 KPI 3: wartość aktualna: 0 GB KPI 4: wartość aktualna: 0 GB KPI 5: wartość aktualna: 0% KPI 1: wartość docelowa: 140 000 KPI 2: wartość docelowa: 140 000 KPI 3: wartość docelowa: 1 200 GB KPI 4: wartość docelowa: 1 200 GB KPI 5: wartość docelowa: co najmniej 95% próbek bez błędów krytycznych
<b>Metoda pomiaru KPI</b>	KPI 1–2: pomiar ilościowy na podstawie ewidencji digitalizacji, raportów z digitalizacji, rejestru partii digitalizacyjnych, systemu repozytorium i protokołów odbioru partii danych. Częstotliwość pomiaru: kwartalnie oraz jednorazowo na zakończenie projektu.

	<p>KPI 3–4: pomiar ilościowy na podstawie wielkości plików powstałych w procesie digitalizacji i wielkości danych udostępnionych on-line. Źródło danych: raport z repozytorium, raport z zasobów dyskowych, ewidencja partii digitalizacyjnych, raport końcowy. Częstotliwość pomiaru: kwartalnie oraz jednorazowo na zakończenie projektu.</p> <p>KPI 5: badanie jakościowe i ilościowe prowadzone na próbkach partii digitalizacyjnych, obejmujące czytelność, kompletność, poprawność przypisania metadanych, zgodność z formatami i brak błędów krytycznych uniemożliwiających ponowne wykorzystanie. Źródło danych: protokoły kontroli jakości digitalizacji, raporty walidacji metadanych, rejestr błędów i korekt. Częstotliwość pomiaru: dla każdej partii digitalizacyjnej oraz zbiorczo kwartalnie; pomiar docelowy na zakończenie projektu.</p>
<b>Cel - 4</b>	Uporządkowanie i automatyzacja procesu publikowania datasetów (do modułu repozytorium dataverse.openforestdata.pl) z uporządkowanych plików źródłowych i roboczych zasobów danych kolekcyjnych uczestniczących jednostek.
<b>Cel strategiczny</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027, ponieważ automatyzacja publikacji danych zwiększa efektywność procesów cyfrowych w organizacjach badawczych. Cel jest zgodny z Programem otwierania danych w zakresie poprawy interoperacyjności i jakości danych.
<b>Korzyść:</b>	Podniesienie efektywności publikowania i aktualizacji danych naukowych przez jednostki uczestniczące w projekcie.
<b>KPI:</b>	KPI 1: Liczba platform udostępniania dokumentów zawierających informacje sektora publicznego/dane prywatne.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 1: wartość docelowa: 1
<b>Metoda pomiaru KPI</b>	KPI 1: pomiar ilościowy na podstawie protokołu odbioru modyfikacji modułu repozytorium Dataverse. Źródło danych: protokół odbioru systemu, dokumentacja techniczna, raport końcowy. Częstotliwość pomiaru: jednorazowo po wdrożeniu platformy.
<b>Cel - 5</b>	Przygotowanie danych okazowych i zielnikowych do integracji z GBIF/IPT oraz standardami opisu bioróżnorodności i zasobów nauki.
<b>Cel strategiczny</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027 oraz w kierunku dotyczące otwartej nauki i interoperacyjności danych o bioróżnorodności. Przygotowanie danych do standardu Darwin Core i GBIF/IPT zwiększa ich widoczność oraz ponowne wykorzystanie.
<b>Korzyść:</b>	Zwiększenie krajowej i międzynarodowej widoczności oraz ponownego wykorzystania polskich danych o bioróżnorodności i zasobów nauki.
<b>KPI:</b>	KPI 1: Liczba rekordów lub obiektów przygotowanych do eksportu lub powiązania z GBIF/IPT w standardzie Darwin Core.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 1: wartość docelowa: 140 000
<b>Metoda pomiaru KPI</b>	KPI 1: pomiar ilościowy na podstawie raportu z modułu jakości danych i metadanych, eksportów Darwin Core Archive, dokumentacji mapowania pól

	oraz raportu końcowego. Źródło danych: raport z systemu, eksporty Darwin Core, dokumentacja techniczna. Częstotliwość pomiaru: kwartalnie w okresie publikowania zasobów oraz jednorazowo na zakończenie projektu.
<b>Cel - 6</b>	Podniesienie jakości danych przez walidację, uzupełnienie metadanych, standaryzację taksonomii, lokalizacji, licencji, identyfikatorów i zasad wersjonowania.
<b>Cel strategiczny</b>	Cel wpisuje się w Program otwierania danych w zakresie poprawy jakości, interoperacyjności i ponownego wykorzystywania danych oraz w zasady FAIR właściwe dla danych naukowych.
<b>Korzyść:</b>	Poprawa wiarygodności i użyteczności danych naukowych udostępnianych przez platformę.
<b>KPI:</b>	KPI 1: Pozytywny wynik walidacji reguł jakości danych i metadanych dla zasobów objętych publikacją.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 1: wartość docelowa: 1 pozytywny raport walidacji jakości danych i metadanych
<b>Metoda pomiaru KPI</b>	KPI 1: badanie jakościowe i ilościowe na podstawie raportów walidacji metadanych, rejestru błędów, statusów jakości, dokumentacji mapowania pól, potwierdzenia stosowania Darwin Core oraz procedur licencjonowania i wersjonowania. Źródło danych: raport walidacji, dokumentacja techniczna, raport końcowy. Częstotliwość pomiaru: kwartalnie oraz jednorazowo na zakończenie projektu.
<b>Cel - 7</b>	Rozwój funkcji wyszukiwania, filtrowania, podglądu i mapowej eksploracji danych w platformie Dataverse.
<b>Cel strategiczny</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027, ponieważ zwiększa dostępność i użyteczność danych dla zróżnicowanych grup użytkowników. Cel wspiera wzrost wykorzystania danych naukowych przez lepsze wyszukiwanie, mapową eksplorację i poprawę UX.
<b>Korzyść:</b>	Poprawa dostępności i łatwości korzystania z przyrodniczych danych naukowych.
<b>KPI:</b>	KPI 1: Liczba zmodyfikowanych systemów teleinformatycznych KPI 2: Pozytywny wynik badań UX lub potwierdzenie wdrożenia korekt wynikających z testów z użytkownikami.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 2: wartość aktualna: 0 KPI 1: wartość docelowa: 1 KPI 2: wartość docelowa: 1
<b>Metoda pomiaru KPI</b>	dla KPI 1: pomiar ilościowy na podstawie odbioru modyfikacji platformy Dataverse. Źródło danych: protokół odbioru modyfikacji systemu Dataverse. Częstotliwość: jednorazowa po odbiorze prac dla systemu Dataverse. Dla KPI 2: pomiar ilościowy na podstawie odbioru protokołu z przeprowadzonych działań UX. Częstotliwość: jednorazowa po odbiorze raportu UX.
<b>Cel - 8</b>	Wdrożenie narzędzi analitycznych umożliwiających pracę z danymi środowiskowymi bezpośrednio w środowisku platformy.
<b>Cel</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027,

<b>strategiczny</b>	ponieważ umożliwi zaawansowane ponowne wykorzystanie cyfrowo udostępnianych zasobów nauki. Cel wspiera analizę danych środowiskowych w nauce i politykach publicznych.
<b>Korzyść:</b>	Zwiększenie efektywności prowadzenia analiz naukowych i środowiskowych na podstawie otwartych danych.
<b>KPI:</b>	KPI 1: Liczba rozwiązań wykorzystujących informacje sektora publicznego/dane prywatne. KPI 2: Pozytywny wynik testów wydajności systemu.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 2: wartość aktualna: 0 KPI 1: wartość docelowa: 1 KPI 2: wartość docelowa: 1 pozytywny raport z testów wydajności
<b>Metoda pomiaru KPI</b>	KPI 1: pomiar ilościowy na podstawie wdrożonego środowiska wykorzystania danych obejmującego API, przykładowe notebooki Jupyter/R/Python, scenariusze analityczne i dokumentację techniczną. Źródło danych: protokół odbioru środowiska analitycznego, dokumentacja techniczna, repozytorium przykładów użycia. Częstotliwość pomiaru: jednorazowo po wdrożeniu rozwiązania oraz weryfikacja końcowa na zakończenie projektu. KPI 2: test wydajnościowy systemu po wdrożeniu docelowych komponentów, obejmujący wyszukiwanie, przeglądanie metadanych, pobieranie plików, działanie API oraz obsługę komponentu mapowego. Źródło danych: raport z testów wydajności, scenariusze testowe, protokół odbioru. Częstotliwość pomiaru: jednorazowo przed odbiorem końcowym systemu; retest w przypadku wyniku negatywnego.
<b>Cel - 9</b>	Rozwój otwartego API, eksportów maszynowych i dokumentacji technicznej w celu zwiększenia ponownego wykorzystania danych.
<b>Cel strategiczny</b>	Cel wpisuje się w cel szczegółowy EFRR.CP1.II Programu FERC 2021-2027 oraz Działanie FERC.02.03, ponieważ rozwój API, eksportów maszynowych i dokumentacji technicznej zwiększa ponowne wykorzystanie zasobów nauki przez użytkowników i systemy zewnętrzne.
<b>Korzyść:</b>	Usprawnienie maszynowego dostępu do danych i ich ponownego wykorzystania w systemach zewnętrznych.
<b>KPI:</b>	KPI 1: Liczba utworzonych API. KPI 2: Liczba baz danych udostępnionych on-line poprzez API. KPI 3: Pozytywny wynik oceny gotowości procedur repozytorium do certyfikacji jakości, wiarygodności i trwałości zasobów.
<b>Wartość aktualna i docelowa KPI:</b>	KPI 1: wartość aktualna: 0 KPI 2: wartość aktualna: 0 KPI 3: wartość aktualna: 0 KPI 1: wartość docelowa: 2 KPI 2: wartość docelowa: 6 KPI 3: wartość docelowa: 1 raport samooceny i gotowości certyfikacyjnej
<b>Metoda pomiaru KPI</b>	KPI 1: pomiar ilościowy na podstawie wdrożonego i udokumentowanego publicznego API platformy; źródło: dokumentacja API, protokół odbioru API, repozytorium dokumentacji technicznej; częstotliwość: jednorazowo po odbiorze API. KPI 2: pomiar ilościowy na podstawie liczby logicznych zasobów danych udostępnionych przez API po imporcie z plików źródłowych i roboczych baz danych po przejściu procesu importu, walidacji i publikacji; źródło:

	<p>dokumentacja API, rejestr plików źródłowych i zasobów danych, raport z systemu, protokół odbioru integracji; częstotliwość: jednorazowo po odbiorze API.</p> <p>KPI 3: badanie jakościowe obejmujące ocenę kompletności procedur publikacji danych, zarządzania metadanymi, wersjonowania, trwałości identyfikatorów, kopii bezpieczeństwa, zasad dostępu, jakości danych i dokumentacji użytkowej. Źródło danych: raport samooceny, procedury repozytorium, dokumentacja techniczna i organizacyjna, raport końcowy. Częstotliwość pomiaru: jednorazowo przed zakończeniem 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 zielnikowe Stacji Geobotanicznej UW oraz zasoby zielnikowe jednostek konsorcjum	30-11-2029	40000
Zdigitalizowane i udostępnione zasoby kolekcji zoologicznej IBS PAN, w tym okazy ssaków, czaszki, skany kart katalogowych i dokumentacja fotograficzna	30-11-2029	50000
Zdigitalizowane i udostępnione kolekcyjne bazy prób i izolatów genetycznych oraz powiązane metadane	30-11-2029	10000
Zdigitalizowane i udostępnione zasoby 3D oraz dane z mikrotomografu	30-11-2029	2000
Zdigitalizowane i udostępnione dane archiwalne, długoterminowe i	30-11-2029	80

Rodzaj informacji/zasobów	Planowana data udostępnienia	Szacowana liczba obiektów objętych digitalizacją (udostępnianiem informacji)
surowe dane środowiskowe		

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
Modyfikacja systemu teleinformatycznego Open Forest Data w zakresie modułów: Dataverse, GeoNode, integracji Dataverse-GeoNode, importu plikowego, automatycznego generowania metadanych, kontroli jakości danych i metadanych, API publicznego, eksportów do GBIF/IPT, możliwości eksportu metadanych do KRONIK@, portalu WWW/dashboardu oraz środowiska analitycznego Jupyter/R/Python	09-2029
Interfejsy API publicznego dostępu do danych i mehttps://www.gov.pl/web/gov/uslugi-dla-przedsiębiorcydanych wraz z dokumentacją techniczną oraz przykładami użycia w systemach Dataverse i Geonode	09-2029
Materiały i instrukcje dla digitalizatorów oraz opiekunów danych, obejmujące standardy digitalizacji, opis metadanymi, zasady walidacji i procedury publikacji	06-2027
Raporty z weryfikacyjnego testu prywatności systemu Dataverse i Geonode	06-2029
Raport z badań UX wraz z pozytywnym wynikiem testów użyteczności albo wykazem wdrożonych korekt po testach dla strony www, modułu Dataverse i modułu GIS	08-2029
Raport z testów bezpieczeństwa wraz z pozytywnym wynikiem testów albo potwierdzeniem usunięcia podatności krytycznych i wysokich modułu Dataverse i modułu GIS systemu Open Forest Data	08-2029
Raport z testów wydajności wraz z pozytywnym wynikiem testów albo potwierdzeniem wdrożenia korekt wydajnościowych dla modułu Dataverse i modułu GIS platformy Open Forest Data	08-2029
Materiały szkoleniowe dla użytkowników API, użytkowników danych i zespołów wykorzystujących zasoby platformy Open Forest data	10-2029
Materiały informacyjno-promocyjne dotyczące platformy Open Forest Data, nowych zasobów i zasad korzystania z danych	10-2029
Infrastruktura w zakresie sprzętu komputerowego i do digitalizacji ( w tym skaner bezdotykowy do zielnika A2+)	06-2027

## 3. KAMIENIE MIŁOWE

Kamienie milowe	Planowany termin osiągnięcia
Powołano zespół zarządczy projektu i uruchomiono projekt	2027-01-30
Przeprowadzono inicjalny test prywatności oraz analizę progową DPIA dla zakresu projektu	2027-02-28
Zakończono rekrutację kluczowego personelu projektu	2027-03-30
Zatwierdzono plan digitalizacji i udostępniania zasobów IBS PAN oraz Uniwersytetu Warszawskiego	2027-06-30
Rozstrzygnięto postępowania zakupowe na usługi IT, usługi digitalizacyjne, sprzęt, testy i szkolenia	2027-06-30
Uruchomiono stanowiska i narzędzia do digitalizacji zasobów objętych projektem	2027-06-30
Zatwierdzono model danych, metadanych, identyfikatorów i procedur jakości danych	2027-09-30
Odebrano pierwszą wersję modyfikacji repozytorium z poprawionym UX i integracją identyfikacji ORCID	2028-03-31
Odebrano podstawową integrację Dataverse-GIS z funkcją powiązania dataset-mapa	2028-06-30
Zdigitalizowano i udostępniono 60% zaplanowanych zasobów kolekcyjnych i zielnikowych	2028-10-30
Uruchomiono pipeline tworzenia datasetów z uporządkowanych plików źródłowych i roboczych zasobów danych IBS PAN oraz UW	2028-12-31
Odebrano moduł jakości danych i metadanych oraz mechanizmy eksportu do GBIF/IPT	2029-03-31
Odebrano publiczne API modułu Dataverse i Geonode wraz z dokumentacją techniczną	2029-06-30
Przeprowadzono weryfikacyjny test prywatności platformy	2029-06-30
Uzyskano pozytywny wynik testów badań UX dla strony www, modułu Dataverse, modułu Geonode	2029-08-31
Uzyskano pozytywny wynik testów bezpieczeństwa dla strony www, modułu Dataverse, modułu Geonode	2029-08-31
Uzyskano pozytywny wynik testów wydajności dla strony www, modułu Dataverse, modułu Geonode	2029-08-31
Wdrożono docelową wersję platformy repozytoryjnej Open Forest Data w tym zmodyfikowaną stronę wejściową www, moduły systemu Dataverse oraz systemu Geonode	2029-09-30
Przygotowano i udostępniono materiały szkoleniowe dla użytkowników API i użytkowników danych	2029-10-31
Zakończono digitalizację, opracowanie metadanych i publiczne udostępnianie zasobów objętych projektem	2029-11-30
Zakończono rzeczową realizację projektu i odebrano wszystkie produkty projektu	2029-11-30

## 4. KOSZTY

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

Całkowity koszt projektu (netto oraz brutto), w tym	Netto 5 435 292,00 zł Brutto 5 435 292,00 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)	2026	Netto 40 000,00 zł Brutto 40 000,00 zł
	2027	Netto 2 029 238,27 zł Brutto 2 029 238,27 zł
	2028	Netto 1 956 244,09 zł Brutto 1 956 244,09 zł
	2029	Netto 1 409 809,64 zł Brutto 1 409 809,64 zł

### 4.2. Wykaz poszczególnych pozycji kosztowych

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
Oprogramowanie	Pozycja obejmuje bezpośrednio koszty wytworzenia, modyfikacji i wdrożenia produktów projektu (w tym prace deweloperskie IT zlecone w postępowaniach przetargowych oraz wykonane in-house przez zatrudniony do celów projektowych personel) oraz przygotowanie udostępnienia zasobów. Wyjaśnienie: prace	4 922 492,00 zł	Pozycja obejmuje bezpośrednio koszty wytworzenia, modyfikacji i wdrożenia produktów projektu oraz przygotowania i udostępnienia zasobów. Ujęto tu koszty personelu merytorycznego bezpośrednio realizującego digitalizację, opracowanie metadanych, przygotowanie datasetów, prace Data/GIS i prace deweloperskie, niezależnie od tego, czy są ponoszone jako koszty własne czy usługi zewnętrzne. Pozycja nie obejmuje niezależnych kosztów UX, bezpieczeństwa i wydajności, które wykazano odrębnie.

Nazwa pozycji kosztowej	Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	<p>deweloperskie zlecane i realizowane zatrudnionym personelem obejmują wszystkie systemy, tj. stronę www, dataverse, geonode oraz wdrożenie wszelkich powiązanych integracji zewnętrznych pomiędzy tymi systemami i systemami zewnętrznymi. Prace deweloperskie, digitalizacja i udostępnianie danych: aktualizacja oprogramowania do skanów i obróbki 3D, modyfikacja Dataverse/Open Forest Data, integracja Dataverse-GIS, mechanizmy importu plikowego z uporządkowanych arkuszy Excel/CSV i innych plików źródłowych, automatyczne generowanie opisów metadanych, kontrola jakości metadanych, API, eksporty do GBIF/IPT oraz przygotowanie możliwości eksportu metadanych do</p>	

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	KRONIK@, opłaty DOI, obsługa techniczna serwerowa, konsultacje bazodanowe, genetyczne, botaniczne i mapowe, przetwarzanie plików oraz prace personelu merytorycznego bezpośrednio realizującego digitalizację, opracowanie metadanych, przygotowanie datasetów i publikację zasobów.		
Infrastruktura	Komputery robocze dla zespołów digitalizujących i opracowujących dane, dyski serwerowe, dyski przenośne SSD, nośniki przechowywania i wymiany danych, skaner A2+ do digitalizacji zielników oraz aktualizacja oprogramowanie do skanera 3D.	142 400,00 zł	Infrastruktura jest niezbędna do prowadzenia digitalizacji, przetwarzania dużych plików, bezpiecznego przechowywania zasobów oraz pracy zespołów odpowiedzialnych za opracowanie i publikację danych. Partner, UW, nie kupował dotychczas infrastruktury do digitalizacji, potrzebuje więc skanera do celów zielnikowych. Wnioskodawca nie kupuje sprzętu do digitalizacji, będzie korzystał ze sprzętu zakupionego w poprzednim projekcie. Zakupiona zostanie przez Wnioskodawcę jedynie aktualizacja oprogramowania do skanera 3D.
Koszty UX i grafiki	Badanie użytkowników, analiza użyteczności, projekt UX, projekt graficzny elementów portalu/ dashboardu, testy z użytkownikami,	9000,00 zł	Koszty są niezbędne do obniżenia bariery korzystania z platformy, poprawy ścieżek dostępu do danych, czytelności map i dashboardów oraz zapewnienia użyteczności dla użytkowników profesjonalnych i nieprofesjonalnych.

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	opracowanie rekomendacji UX oraz weryfikacja wdrożenia poprawek.		
Bezpieczeństwo	Inicjalny i weryfikacyjny test prywatności, analiza wymagań ochrony danych i prywatności, analiza danych wrażliwych przyrodniczo, testy podatności, audyt bezpieczeństwa, weryfikacja konfiguracji systemu, zgodność z KRI i RODO, retesty oraz raport z testów bezpieczeństwa.	25 000,00 zł	Pozycja jest konieczna dla potwierdzenia bezpieczeństwa systemu przed produkcyjnym udostępnieniem danych, ograniczenia ryzyk cyberbezpieczeństwa, zapewnienia ochrony dostępu, integralności danych, prywatności użytkowników i właściwego postępowania z danymi wrażliwymi przyrodniczo.
Wydajność rozwiązań	Testy wydajnościowe systemu, API, repozytorium i komponentu GIS, przygotowanie scenariuszy testowych, wykonanie testów obciążeniowych, analiza wyników, rekomendacje optymalizacyjne, retesty po wdrożeniu korekt oraz raport z testów wydajności.	15 000,00 zł	Pozycja jest niezbędna ze względu na skalę digitalizacji i udostępniania danych, planowane wykorzystanie API, obsługę plików przestrzennych i dużych zasobów cyfrowych oraz konieczność zapewnienia stabilności i efektywności działania platformy.
Szkolenia	Materiały i instrukcje dla digitalizatorów, szkolenia opiekunów danych, szkolenia z jakości metadanych i walidacji danych, szkolenia użytkowników API	12 000,00 zł	Szkolenia są niezbędne do zapewnienia prawidłowego procesu digitalizacji, spójnej jakości metadanych, poprawnego publikowania zasobów oraz efektywnego wykorzystania platformy przez użytkowników danych i API.

Nazwa pozycji kosztowej		Przewidywany koszt brutto	Uzasadnienie pozycji kosztowej (przeznaczenie)
	oraz szkolenia z korzystania z danych, map i przykładów analitycznych.		
Działania informacyjno-promocyjne	Materiały promocyjne i informacyjne, udział w wydarzeniach naukowych i branżowych, prezentacja nowych zasobów i funkcjonalności, przygotowanie posterów, banerów, komunikatów i materiałów popularyzujących korzystanie z platformy.	90 400,00 zł	Działania umożliwią poinformowanie użytkowników o nowych zasobach, zasadach dostępu, API, mapach i możliwościach ponownego wykorzystania danych, co jest istotne dla osiągnięcia efektów projektu i wskaźników wykorzystania.
Koszty zarządzania i wsparcia (w tym wynagrodzenia personelu wspomagającego)	Kierownik projektu, koordynacja rzeczowa i finansowa, obsługa administracyjna, rozliczenia, sprawozdawczość, zamówienia, wsparcie prawne i finansowe, koszty pośrednie oraz inne koszty wsparcia niezwiązane bezpośrednio z wytwarzaniem produktów merytorycznych.	219 000,00 zł	Pozycja obejmuje wyłącznie koszty niezbędne do prawidłowego zarządzania projektem, obsługi administracyjnej, finansowej, prawnej i sprawozdawczej - zarządzanie projektem, koordynacja prac Beneficjenta i Partnera, obsługa administracyjna, finansowa i sprawozdawcza, obsługa zamówień, wsparcie prawne i finansowe oraz koszty pośrednie rozliczane zgodnie z zasadami projektu współfinansowanego ze środków UE

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

Całkowity koszt utrzymania trwałości projektu (brutto)	80 000,00 zł	Źródło finansowania
--	--------------	---------------------

Podział całkowitego kosztu utrzymania trwałości projektu na poszczególne lata (netto oraz brutto)	2030	1154,00 zł (brutto) (1154,00 zł netto)	środki własne Wnioskodawcy i Partnera
	2031	14 535,00 zł (brutto) (14 535,00 zł netto)	środki własne Wnioskodawcy i Partnera
	2032	15 262,00 zł (brutto) (15 262,00 zł netto)	środki własne Wnioskodawcy i Partnera
	2033	16 022,00 zł (brutto) (16 022,00 zł netto)	środki własne Wnioskodawcy i Partnera
	2034	16 827,00 zł (brutto) (16 827,00 zł netto)	środki własne Wnioskodawcy i Partnera
	2035	16 200,00 zł (brutto) (16 200,00 zł netto)	środki własne Wnioskodawcy i Partnera

#### 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
Niezrealizowanie zamówień publicznych i zakupów w planowanym terminie	Średnia	Niskie	Wczesne przygotowanie opisów przedmiotu zamówienia, harmonogram postępowań z buforami czasowymi, monitoring postępowań przez kierownika projektu oraz przygotowanie wariantów ograniczających zakres usług bez naruszania celów projektu.
Brak możliwości zatrudnienia osób o odpowiednich kompetencjach	Średnia	Średnie	Publikacja ogłoszeń w kanałach naukowych i branżowych, wykorzystanie kontaktów jednostek konsorcjum, możliwość pracy częściowo zdalnej dla wybranych stanowisk oraz szkolenia

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			wdrożeniowe dla osób z kompetencjami pokrewnymi.
Brak wystarczających zasobów kadrowych do realizacji digitalizacji i metadanych	Duża	Niskie	Organizacja prac w dwóch jednostkach uczestniczących w projekcie: IBS PAN oraz Uniwersytecie Warszawskim. Wdrożenie wspólnych instrukcji digitalizacyjnych i metadanych. Priorytetyzacja zasobów oraz bieżący monitoring wydajności prac. Częściowa automatyzacja tworzenia datasetów i walidacji metadanych.
Niezgodność wersji systemu przy aktualizacji Dataverse lub GeoNode	Duża	Niskie	Beneficjent zweryfikował wstępnie zgodność. Wybrane do zaimplementowania narzędzia pochodzą z oficjalnych integracji do Dataversa. Audyt konfiguracji i modyfikacji przed aktualizacją, środowisko testowe, aktualizacja etapowa, testy regresji oraz unikanie modyfikacji kodu core Dataverse.
Nieosiągnięcie wskaźników produktu lub celu projektu	Duża	Niskie	Bieżący monitoring wskaźników kwartalnych, raportowanie postępu digitalizacji i publikacji, możliwość przesunięcia zasobów kadrowych między pakietami prac oraz priorytetyzacja zasobów o najwyższej gotowości do digitalizacji.
Niedoszacowanie nakładu pracy na digitalizację i opracowanie metadanych	Duża	Niskie	Wykorzystanie doświadczeń z projektu e-Puszcza, między innymi poprzez dywersyfikację prac deweloperskich i data science pomiędzy usługi zlecone zewnętrznie i pracę etatową personelu., etapowa kontrola produktywności zespołów, próbkowanie zasobów przed digitalizacją oraz korekta organizacji prac w trakcie projektu.
Nadmierna złożoność zakresu funkcjonalnego modernizowanej platformy	Duża	Niskie	Wykorzystanie istniejących, utrzymywanych przez społeczność międzynarodową rozwiązań open source Dataverse i GeoNode oraz ich dokumentowanych interfejsów. Realizacja modyfikacji przez konfigurację, API i warstwę integracyjną bez forka kodu podstawowego Dataverse. Priorytetyzacja funkcji krytycznych: publikacja datasetów, metadane,

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			integracja GIS, import danych, jakość danych i API. Etapowe odbiory modułów i testy regresji.

## 5.2. Ryzyka wpływające na utrzymanie efektów

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
Brak wystarczających środków na utrzymanie efektów projektu	Średnia	Średnie	Zaplanowanie finansowania utrzymania ze środków własnych jednostek, ograniczenie kosztów przez wykorzystanie rozwiązań open source oraz utrzymanie wariantu minimum po okresie trwałości.
Brak wystarczających zasobów kadrowych do utrzymania efektów projektu	Duża	Wysokie	Przygotowanie instrukcji administracyjnych i technicznych. Szkolenie pracowników po stronie IBS PAN i Stacji Uniwersytetu Warszawskiego. Dokumentowanie procedur utrzymaniowych i aktualizacyjnych. Zapewnienie jak najdalej idącej "bezobsługowości" systemu - tj. w trakcie projektu zapewnione zostanie jak najszersze przeprowadzenie procesu digitalizacji, żeby zaspokoić określone potrzeby użytkowników. Po projekcie planowane jest zapewnienie możliwości digitalizacji i udostępniania własnymi zasobami jednostek, a kluczowa dla głównych użytkowników naukowych będzie możliwość samodzielnego wykorzystywania repozytorium jako platformy do samodzielnego deponowania, publikowania i współpracy naukowej, co do tej pory przyniosło zadowolenie użytkownika naukowego.
Odejście osób z wypracowanymi kompetencjami digitalizacyjnymi lub technicznymi	Duża	Wysokie	Dokumentowanie procedur digitalizacyjnych i technicznych. Tworzenie materiałów szkoleniowych dla następców i opiekunów danych. Wdrożenie pipeline'ów i instrukcji

Nazwa ryzyka	Siła oddziaływania	Prawdopodobieństwo wystąpienia ryzyka	Sposób zarządzania ryzykiem
			postępowania dla pracowników i zarządzających kolekcjami zoologiczną i botaniczną. Przekazywanie wiedzy w zespołach IBS PAN i UW. Możliwość ponownego szkolenia pracowników w okresie trwałości.
Nieosiągnięcie wszystkich zaplanowanych korzyści w okresie trwałości	Duża	Niskie	Monitorowanie wykorzystania zasobów, działania informacyjne, aktualizacja dokumentacji użytkowej oraz utrzymywanie współpracy z naukowcami, administracją i agregatorami danych.
Sytuacja geopolityczna lub zdarzenia nadzwyczajne wpływające na dostępność zasobów	Duża	Średnie	Zabezpieczenie zasobów cyfrowych, kopie bezpieczeństwa, procedury ciągłości działania i rozproszenie odpowiedzialności między partnerów.

## 6. OTOCZENIE PRAWNE

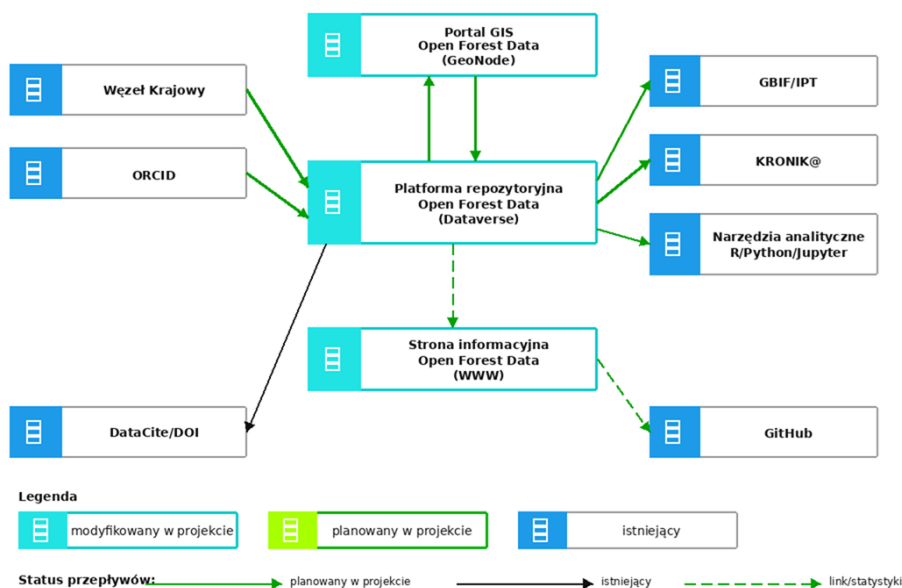
Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
1	Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego	TAK/NIE		
2	Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne	TAK/NIE		
3	Rozporządzenie Rady Ministrów z dnia 21 maja 2024 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych	TAK/NIE		
4	Ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa	TAK/NIE		

Lp.	Tytuł aktu prawnego	Czy wymaga zmian	Opis zmian (jeśli dotyczy)	Etap prac legislacyjnych (jeśli dotyczy)
5	Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych	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 5 września 2016 r. o usługach zaufania oraz identyfikacji elektronicznej	TAK/NIE		
8	Rozporządzenie Ministra Cyfryzacji z dnia 10 marca 2020 r. w sprawie szczegółowych warunków organizacyjnych i technicznych, które powinien spełniać system teleinformatyczny służący do uwierzytelniania użytkowników	TAK/NIE		
9	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. - ogólne rozporządzenie o ochronie danych	TAK/NIE		
10	Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/868 z dnia 30 maja 2022 r. w sprawie europejskiego zarządzania danymi	TAK/NIE		
11	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	TAK/NIE		
12	Rozporządzenie Prezesa Rady Ministrów z dnia 20 lipca 2011 r. w sprawie podstawowych wymagań bezpieczeństwa teleinformatycznego	TAK/NIE		

## 7. ARCHITEKTURA

### 7.1. Widok kooperacji aplikacji

## Open Forest Data - widok kooperacji aplikacji



## Lista systemów wykorzystywanych w projekcie

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
1	Dataverse/ Open Forest Data	Instytut Biologii Ssaków Polskiej Akademii Nauk	Dataverse/Open Forest Data to system teleinformatyczny wspierający deponowanie, opis metadanymi, wersjonowanie, publikowanie, cytowanie i udostępnianie przyrodniczych danych naukowych oparty na własnej . System obsługuje kolekcje, datasety, pliki, metadane, wersje, licencje, trwałe identyfikatory, role użytkowników i publiczne API. W projekcie system przyjmuje dane przygotowane w uporządkowanych plikach źródłowych, w szczególności Excel/CSV, i przetwarza je do postaci datasetów oraz metadanych	Modyfikowany	Modyfikacja obejmuje moduły: repozytorium danych, udostępnianie danych, API publiczne, uwierzytelnianie i dostęp, dane wrażliwe przyrodniczo, import plików, automatyczne generowanie metadanych, kontrolę jakości metadanych oraz eksporty maszynowe, w tym GBIF/IPT i możliwość eksportu

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			publikowanych w repozytorium.		metadanych do KRONIK@.
2	GeoNode/ Portal GIS	Instytut Biologii Ssaków Polskiej Akademii Nauk	GeoNode/Portal GIS to system teleinformatyczny wspierający publikowanie, przeglądanie, stylowanie i udostępnianie warstw przestrzennych, map i usług geoinformacyjnych. System obsługuje dane wektorowe i rastrowe, metadane przestrzenne, usługi mapowe oraz podgląd danych przestrzennych.	Modyfikowany	Modyfikacja obejmuje moduł map i warstw GIS, powiązania dataset-mapa, linki zwrotne mapa-dataset, funkcję „Zobacz na mapie”, style mapowe i integrację z Dataverse/ Open Forest Data.
3	Portal Open Forest Data	Instytut Biologii Ssaków Polskiej Akademii Nauk	Portal Open Forest Data to wejściowa strona www wspierająca prezentację zasobów, ścieżki użytkownika, dashboardy, widoki tematyczne, informacje o danych, instrukcje, statystyki i dostęp do usług platformy. Poza stroną www trzonem repozytorium przechowującym dane są obecnie funkcjonujące własne instancje oprogramowania typu open source Dataverse i Geonode (odpowiednio do przechowywania danych ogólnych i GISowych)..	Modyfikowany	Modyfikacja obejmuje dashboard, widoki tematyczne, prezentację zasobów, ścieżki dostępu do danych i map, publikację informacji o API oraz prezentację informacji o możliwości ponownego wykorzystania danych.
4	Jupyter/R/ Python	Instytut Biologii Ssaków Polskiej Akademii Nauk	Jupyter/R/Python to środowisko analityczne wspierające przygotowanie przykładów pracy z danymi, notebooków analitycznych, scenariuszy wykorzystania API, danych przestrzennych i eksportów maszynowych.	Istniejący	

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
			System służy do demonstracji i ułatwienia ponownego wykorzystania danych.		
5	ORCID	ORCID, Inc.	ORCID to międzynarodowy system identyfikacji badaczy i kontrybutorów, wspierający jednoznaczne powiązanie osób z dorobkiem naukowym, datasetami i innymi zasobami cyfrowymi. System jest wykorzystywany do identyfikacji autorów i użytkowników oraz powiązania identyfikatora ORCID z zasobami przez standardowe mechanizmy integracyjne.	Istniejący	
6	DataCite/DOI	DataCite e.V.	DataCite/DOI to międzynarodowy system obsługi trwałych identyfikatorów DOI dla danych badawczych i innych zasobów nauki. System wspiera rejestrację metadanych, trwałe odwołania do datasetów i interoperacyjność danych cytowania.	Istniejący	
7	GBIF/IPT	GBIF Secretariat oraz gestor właściwej instancji IPT	GBIF/IPT to międzynarodowa infrastruktura agregacji i publikacji danych o bioróżnorodności oraz narzędzie IPT służące do publikacji danych zgodnych ze standardem Darwin Core. System jest wykorzystywany jako kanał zwiększenia widoczności danych okazowych i zielnikowych.	Istniejący	
8	GitHub	GitHub, Inc./	GitHub to zewnętrzny system repozytoriów	Istniejący	

Lp.	Nazwa systemu	Gestor systemu	Opis systemu	Status	Krótki opis ewentualnej zmiany
		Microsoft	kodu, dokumentacji technicznej, przykładów użycia API, skryptów pomocniczych i informacji o wersjach komponentów. System wspiera utrzymanie dokumentacji technicznej i przykładów integracyjnych.		
9	Węzeł Krajowy	Ministerstwo Cyfryzacji	Węzeł Krajowy to krajowy system identyfikacji elektronicznej wspierający uwierzytelnianie użytkowników usług online. System może być wykorzystywany do obsługi dostępu instytucjonalnego lub personalnego do platformy Open Forest Data.	Istniejący	
10	KRONIK@	Ministerstwo Cyfryzacji	KRONIK@ - Krajowe Repozytorium Obiektów Nauki i Kultury to istniejący system teleinformatyczny służący do gromadzenia, wyszukiwania i udostępniania cyfrowych obiektów nauki i kultury. W projekcie system jest zewnętrznym, dodatkowym kanałem eksportu metadanych i odwołań do wybranych zasobów udostępnianych w repozytorium Open Forest Data. Eksport do KRONIK@ jest możliwy do wykonania na etapie realizacji projektu, po doprecyzowaniu formatu i trybu technicznego, ale nie warunkuje podstawowego udostępnienia zasobów w repozytorium, API i komponentach publikacyjnych.	Istniejący	

## Lista przepływów

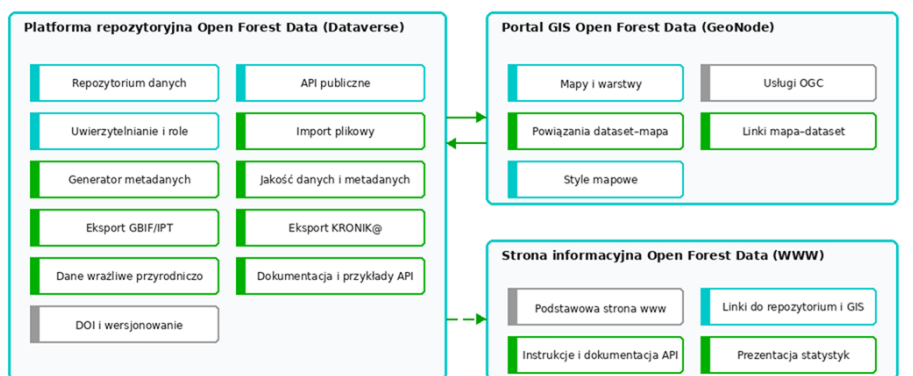
Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
1	Dataverse/ Open Forest Data	GeoNode/ Portal GIS	Metadane datasetów przestrzennych, pliki przestrzenne, informacje o lokalizacji, powiązania dataset-mapa i dane do publikacji warstw GIS.	kopiowanie danych	krytyczny dla sukcesu projektu	API, middleware, GeoJSON, GeoPackage, SHP, WMS, WFS, ISO 19115
2	GeoNode/ Portal GIS	Dataverse/ Open Forest Data	Identyfikatory warstw i map, linki zwrotne do zasobów źródłowych, informacje o warstwach GIS powiązanych z datasetami.	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	API GeoNode, middleware, JSON, REST
3	Dataverse/ Open Forest Data	Portal Open Forest Data	Metadane datasetów, informacje o zasobach, kolekcjach, licencjach, DOI, statusach publikacji i powiązaniach z mapami.	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	REST API, JSON, HTTPS
4	GeoNode/ Portal GIS	Portal Open Forest Data	Mapy, warstwy tematyczne, podglądy przestrzenne i odnośniki do zasobów źródłowych.	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	WMS, WFS, WMTS, GeoJSON, API
5	Dataverse/ Open Forest Data	DataCite/ DOI	Metadane datasetów wymagane do rejestracji i obsługi trwałych identyfikatorów DOI.	tryb odwołań bezpośrednich	krytyczny dla sukcesu projektu	Mechanizm DOI Dataverse, DataCite Metadata Schema, API, HTTPS

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
6	ORCID	Dataverse/ Open Forest Data	Identyfikatory użytkowników, autorów i kontrybutorów oraz dane niezbędne do powiązania autora z datasetem lub kontem użytkownika.	tryb odwołań bezpośrednich	realizowalny inną metodą	OAuth 2.0, ORCID iD, HTTPS; wykorzystanie standardowego mechanizmu integracyjnego ORCID bez odrębnych uzgodnień gestorskich na etapie przygotowania projektu
7	Dataverse/ Open Forest Data	GBIF/IPT	Dane okazowe i zielnikowe przygotowane do publikacji lub powiązania z GBIF, metadane taksonomiczne, lokalizacyjne i źródłowe.	kopiowanie danych	realizowalny inną metodą	Darwin Core Archive, Darwin Core, CSV, EML, IPT, HTTPS
8	Dataverse/ Open Forest Data	Jupyter/R/ Python	Datasety, metadane, pliki źródłowe i przykładowe dane do analiz, notebooków i scenariuszy użycia API.	tryb odwołań bezpośrednich	realizowalny inną metodą	REST API, JSON, CSV, pobieranie plików, notebooki Jupyter
9	Węzeł Krajowy	Dataverse/ Open Forest Data	Dane uwierzytelniające i identyfikacyjne użytkowników niezbędne do obsługi dostępu instytucjonalnego lub personalnego do platformy.	tryb odwołań bezpośrednich	realizowalny inną metodą	OpenID Connect lub SAML; integracja realizowana zgodnie z dokumentacją techniczną i procedurą właściwą dla usługi uwierzytelniania
10	GitHub	Dataverse Open Forest Data	Dokumentacja techniczna, przykłady użycia API, skrypty	tryb odwołań bezpośrednich	realizowalny inną metodą	HTTPS, Git

Lp.	System źródłowy	System docelowy	Zakres wymienianych danych	Sposób wymiany danych	Typ modyfikacji	Typ interfejsu
			pomocnicze, informacje o wersjach komponentów.			
11	Dataverse/ Open Forest Data	KRONIK@	Eksport opisów obiektów, metadanych i odwołań do zasobów cyfrowych z Dataverse/ Open Forest Data do KRONIK@, obejmujący w szczególności identyfikatory, tytuły, opisy, słowa kluczowe, informacje o autorach lub instytucjach, licencjach, datach, typach obiektów oraz odwołania do plików źródłowych lub DOI. Przepływ nie jest warunkiem krytycznym dla podstawowego udostępnienia zasobów w projekcie.	kopiowanie danych	realizowalny inną metodą	Eksport plikowy lub API, zgodnie z możliwościami technicznymi KRONIK@ doprecyzowanymi na etapie realizacji projektu; XML, JSON lub CSV/ZIP

## 7.2. Kluczowe komponenty architektury rozwiązania

### Open Forest Data - widok komponentów



#### Legenda



## 7.3. Przyjęte założenia technologiczne

Lp.	Obszar	Założenie technologiczne
1.	Infrastruktura	Rozwiązanie będzie rozwijane z zachowaniem zgodności z istniejącą infrastrukturą repozytorium Open Forest Data (w tym wykorzystanie zasobów serwerowych Politechniki Białostockiej, oraz własnych do przesyłu i tymczasowego przechowywania plików) oraz komponentów powiązanych. Zakłada się wykorzystanie skalowalnej, istniejącej obecnie w sieci nauki infrastruktury umożliwiającej przechowywanie dużych plików danych naukowych, odwzorowań cyfrowych, danych przestrzennych, danych 3D/microCT oraz metadanych. Infrastruktura musi zapewniać już obecnie możliwość wykonywania kopii zapasowych, odtwarzania danych, aktualizacji bezpieczeństwa oraz dalszego utrzymania repozytorium po zakończeniu projektu.
2.	Sieć i bezpieczeństwo	Komunikacja z systemem będzie realizowana z użyciem szyfrowanych połączeń HTTPS/TLS. System będzie wykorzystywał mechanizmy uwierzytelniania i autoryzacji użytkowników zgodne z architekturą Dataverse, w tym konta lokalne oraz integrację z ORCID, a w zakresie przewidzianym w projekcie także możliwość integracji z Węzłem Krajowym. Dostęp do funkcji administracyjnych, importowych i redakcyjnych będzie ograniczony rolami użytkowników. Dla danych wrażliwych przyrodniczo, w szczególności danych o lokalizacji gatunków i siedlisk wymagających ochrony, zakłada się stosowanie ograniczeń widoczności, zakresu publikacji lub poziomu szczegółowości udostępnianych danych. Zapewnienie bezpieczeństwa i sieci będzie oparte na istniejących zasobach Politechniki Białostockiej.

Lp.	Obszar	Założenie technologiczne
3.	Standardy wymiany danych	Rozwiązanie będzie wykorzystywało otwarte i powszechnie stosowane formaty wymiany danych, w szczególności CSV, XLSX, JSON, XML, ZIP, GeoJSON oraz formaty rastrowe i wektorowe właściwe dla danych przestrzennych. Import danych źródłowych będzie możliwy z uporządkowanych plików roboczych Excel/CSV, które nie stanowią odrębnych systemów teleinformatycznych. Dla danych przyrodniczych zakłada się stosowanie mapowania metadanych do standardów właściwych dla repozytoriów danych naukowych i bioróżnorodności, w szczególności DataCite, DOI, ORCID, Darwin Core, GBIF/IPT oraz standardów OGC dla danych przestrzennych. Eksport do KRONIK@ będzie realizowany w formacie i trybie technicznym doprecyzowanym na etapie realizacji projektu.
4.	Systemy operacyjne serwerowe	Rozwiązanie będzie utrzymywane w środowisku serwerowym Politechniki Białostockiej, z którego na podstawie odrębnych umów korzysta Beneficjent. Są one zgodne z wymaganiami technologicznymi Dataverse, GeoNode i komponentów open source wykorzystywanych w projekcie. Zakłada się wykorzystanie stabilnych, wspieranych wersji systemów operacyjnych klasy Linux lub środowisk równoważnych, zapewniających aktualizacje bezpieczeństwa, zgodność z wymaganiami aplikacji oraz możliwość długoterminowego utrzymania po zakończeniu projektu.
5.	Bazy danych	Warstwa danych będzie zachowywać zgodność z wymaganiami Dataverse i GeoNode. Zakłada się wykorzystanie relacyjnych baz danych oraz komponentów indeksowania i wyszukiwania zgodnych z tymi rozwiązaniami, w szczególności PostgreSQL/PostGIS oraz mechanizmów indeksowania metadanych stosowanych przez Dataverse. Struktury danych i metadanych będą projektowane w sposób umożliwiający import, walidację, publikację, wyszukiwanie, eksport i ponowne wykorzystanie danych naukowych.
6.	Serwery aplikacji	Warstwa aplikacyjna będzie rozwijana z zachowaniem zgodności z architekturą Dataverse, GeoNode oraz powiązanych komponentów open source i nowo utworzonych. Modyfikacje będą realizowane przede wszystkim przez konfigurację, API, komponenty integracyjne i moduły rozszerzające, bez tworzenia ryzykownego forka kodu podstawowego Dataverse. Zakłada się stosowanie udokumentowanych interfejsów aplikacyjnych, w szczególności REST API, oraz mechanizmów umożliwiających import danych, generowanie metadanych, kontrolę jakości, integrację GIS i eksport danych do systemów zewnętrznych.
7.	Portale	Wykorzystany zostanie obecnie istniejący portal openforestdata.pl oraz znajdujące się na podstronach moduły: dataverse.openforestdata.pl oraz gis.openforestdata.pl, gdzie wprowadzone zostaną modyfikacje poprawiające użyteczność i dodane zostaną nowe funkcjonalności. Portal Open Forest Data będzie zapewniał publiczny dostęp do opisów zasobów, datasetów, metadanych, map, warstw przestrzennych, dokumentacji API oraz materiałów

Lp.	Obszar	Założenie technologiczne
		informacyjnych. Portal zostanie zmodyfikowany i unowocześniony z uwzględnieniem dostępności cyfrowej, responsywności, spójnego nazewnictwa i jednoznacznych ścieżek dostępu do danych. Zakłada się rozbudowanie integracji repozytorium Dataverse z komponentem GeoNode / Portal GIS oraz rozwój widoków tematycznych i dashboardów prezentujących zasoby udostępniane w projekcie.
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~~
- ~~-dodatkowe zabezpieczenia powyżej wymogów KRI: należy wskazać uzasadnienie~~