



Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



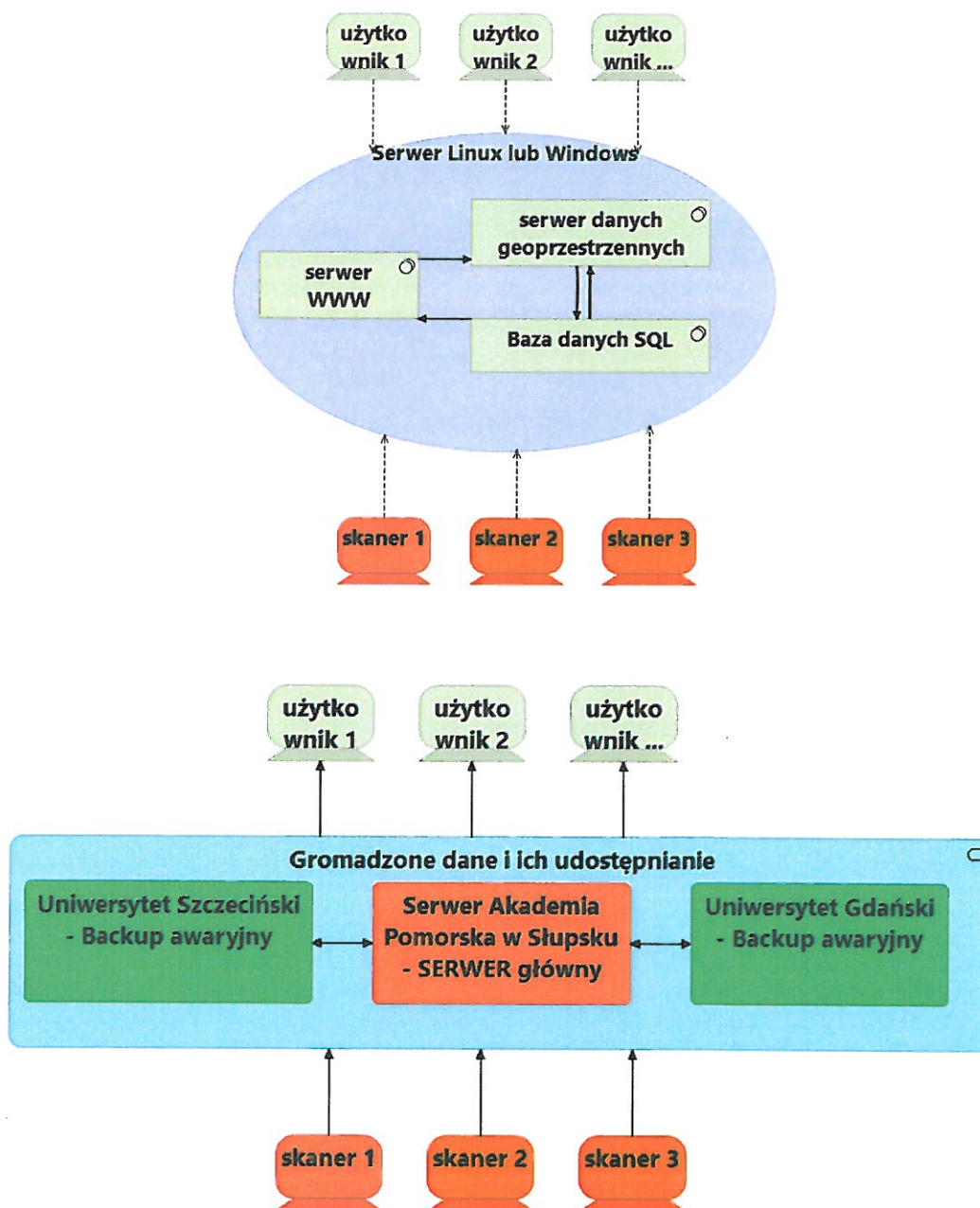
AKADEMIA POMORSKA W SŁUPSKU
ul. Arciszewskiego 22a
76-200 Słupsk

W związku z koniecznością dokonania szacowania wartości zamówienia projektu wykonawczego na potrzeby realizacji projektu pn.: „Zintegrowane wirtualne Herbarium Pomorza Herbarium Pomeranicum – digitalizacja i udostępnienie zbiorów herbariów jednostek akademickich Pomorza poprzez ich połączenie i udostępnienie cyfrowe” współfinansowanego ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa 2014-2020” Akademia Pomorska w Słupsku zwraca się z prośbą o oszacowanie kosztu realizacji zamówienia w odniesieniu do zakresu prac określonego poniżej:

Zakładane standardy i rozwiązania

W ramach projektu przewidziane są trzy „lustrzane” centra danych (Szczecin, Słupsk i Gdańsk) z zainstalowaną niezbędną infrastrukturą sieciowo-serwerową dedykowaną na potrzeby wirtualizacji tworzonych systemów. Każde z centrów wyposażone będzie w macierz dyskową o pojemności ok. 260TB. Przewiduje się po wykonaniu prac przeprowadzić testy określenia klasy dostępności po wykonaniu prac związanych z wdrożeniem rozwiązań hardware i software w ostatnim roku prac (2023). Testy te obejmą również analizy obciążeniowe systemu z zastosowaniem technik „stress” i „load” (np. za pomocą aplikacji JMeter).

Centra danych w Słupsku i Gdańsku przyjmą rolę klastra HA (*high availability*) z replikacją przyrostową danych pomiędzy macierzami. Jeżeli dojdzie do awarii, lub z jakichś powodów któraś ze świadczonych usług składowych budowanego systemu będzie niedostępna to system automatycznie przełączy role w klastrze co z punktu ciągłości pracy użytkownika końcowego systemu będzie nieodczuwalne. W przypadku centra danych w Szczecinie rozwiązanie przewiduje regularne wykonywanie kopii przyrostowej aktualnego stanu systemu z zachowaniem wersjonowania w określonym przedziale czasowym co umożliwi odtworzenie stanu systemu na określony dzień. Ponadto w Szczecinie, podobnie jak w systemie głównym HA uruchomione będą usługi związane z obsługą projektu realizowane poprzez strony WWW. Każde z centrum danych może więc przejść rolę strony wiodącej.



Podstawowa funkcjonalność:

System ma zapewnić:

- komunikację poprzez sieć internetową w postaci klient – serwer z zapewnieniem bezpieczeństwa z wykorzystaniem protokołów szyfrujących komunikację np. TLS 1.3, protokół SSH, VPN,
- sprawne przechowywanie danych w postaci bezstratnie kompresowanych danych rastrowych,
- umożliwić utworzenie bazy danych informacji zawartych na arkuszach zielnikowych wraz z plikami zeskanowanych alegatów,
- przeszukiwanie danych w postaci kreowanych przez użytkownika zapytań do bazy danych,
- przeszukiwanie wybranych arkuszy w postaci rastrowej i wyświetlanie ich na ekranie w domyślnie obniżonej rozdzielczości w celu ich sprawnego przeglądania,

- pobieranie przez użytkownika końcowego danych rastrowych w wybranych rozdzielczościach wraz z informacjami o arkuszu w postaci plików tekstowych lub wariantach xml,
- wykonywanie analiz na wybranych z bazy arkuszach polegających np. na wyświetlaniu map rozmieszczenia gatunków na podstawie siatek używanych w botanice oraz współrzędnych geograficznych dla arkuszy, które zostały wcześniej poddane geolokalizacji. Mapy będą udostępniane również w postaci formatów WMS i WFS standaryzowanych przez Open Geospatial Consortium, które podobne jak np. warstwy tematyczne „geoportalu”, można otworzyć w aplikacjach związanych z administrowaniem danymi przestrzennymi (GIS) wykorzystywanymi przez podmioty gospodarcze oraz administrację publiczną
- automatyczne przygotowywanie danych do dalszych analiz naukowych w formatach csv, txt lub innych wykorzystywanych w statystyce przestrzennej wykonywanej przez najbardziej rozpowszechnione oprogramowanie np. STATISTICA, MAXEnt itp.

Cyfryzowane dane zgromadzone w trakcie trwania Projektu udostępniane będą również poprzez API rekomendowane przez gremia standaryzacyjne zgodne m.in. GBIF (The Global Biodiversity Information Facility). Umożliwi to łatwą wymianę danych o rozmieszczeniu gatunków pomiędzy istniejącymi już tego typu projektami na świecie. Dostęp do danych zostanie więc zapewniony zgodnie ze standardem TDWG Access Protocol for Information Retrieval (TAPIR) (<http://www.tdwg.org/activities/tapir/>). Ponadto opis lokalizacji powinien być zgodny ze standardem ATPOL stosowanym w Polskich ośrodkach Akademickich.

System będzie umożliwiać przeglądanie danych na różnych poziomach dostępu:

- Poziom A. Zespół redakcyjny zielnika – osoby mające wpływ na **weryfikację wszystkich danych** tj. mogą dane dodawać, usuwać i dowolnie je zmieniać,
- Poziom B. Edytor bazy danych i osoba skanująca alegaty – osoby mogące dodawać skany do systemu i w pełni edytować **wprowadzane przez siebie dane**,
- Poziom C. Osoby związane czynnie z zielnikiem które będą miały rozszerzone możliwości w stosunku do niektórych danych, **wskazane przez zespół redakcyjny**,
- Poziom D. Użytkownicy zewnątrzni wyszukujący dane i mogący je **przeglądać oraz pobierać**.

Główne dane tj. zdigitalizowane alegaty będą przechowywane w autorskim oprogramowaniu umożliwiającym zbudowanie go według potrzeb i rozbudowę na własnej licencji, bez potrzeby ubiegania się o takie do podmiotów trzecich. W ramach potrzeb wykorzystywane będą również gotowe rozwiązania na licencjach wolnych np. GNU GPL. Przechowywanie danych rastrowych odbywać się będzie dla większości danych w postaci plików graficznych w formatach bezstratnych TIF lub PNG z możliwością przechowywania ich w różnej rozdzielczości tak aby udostępnianie ich było płynne. Dane te będą więc pobierane w postaci plików bezpośrednio z macierzy dyskowej. Ponadto możliwe będzie pobranie informacji z etykiety, która zapisana będzie w bazie danych. Informacje te będą mogły być filtrowane, tak aby użytkownik końcowy miał pełnię możliwości kreowania zapytań dla jak najlepszego wyniku wyszukiwania. Umożliwiać to będzie aplikacje internetowa uruchamiana w przeglądarce a także dedykowana dla systemów mobilnych. Pobrane dane zostaną wyeksportowane do formatów powszechnie używanych do gromadzenia i importu danych tj. txt, csv, xls. Zostanie także zachowana możliwość eksportowania danych do standardów światowych m.in. Access to Biological Collections Data (ABCD) Schema – jest to standard dostępu do danych biologicznych, opracowany w ramach projektu BIOCASE (Biological Collections Access for Europe). Ze względu na objętość, w bazie danych przechowywany

będzie jedynie adres położenia zeskanowanych alegatów w macierzy lub adres url.

Wszystkie dane będą geolokalizowane, co umożliwi wykonywanie map rozmieszczenia zgromadzonych okazów. Mapy te będą wykonywane w zadanej dokładności z wykorzystaniem podkładów mapowych udostępnianych w ramach INSPIRE przez geoportal. Takie rozwiązanie umożliwi lepszą integrację danych z podmiotami wykorzystującymi te informacje dla celów gospodarczych, społecznych i naukowych. Mapy te będą udostępniane w postaci serwisów WFS i WMS oraz plików źródłowych np. gml, kml, shp, umożliwiających otwieranie ich w aplikacjach związanych z gospodarowaniem przestrzenią (GIS). Dane tego typu mogą być następnie prezentowane w postaci atlasów.

Platforma Herbarium Pomeranicum zostanie zaprojektowana w sposób zorientowany na użytkownika w oparciu o następujące normy projektowania:

ISO 9241-210:2010 „Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems”, ISO/TR 16982:2002 „Ergonomics of human-system interaction - Usability methods supporting human-centered design”, ISO TR 18529:2000 „Human-centered lifecycle process descriptions”, PN-EN ISO 9241-210:2011 „ Ergonomia interakcji człowieka i systemu - Część 210: Projektowanie ukierunkowane na człowieka w przypadku systemów interaktywnych”. Dotyczyć to będzie zwłaszcza interfejsu dla którego zostaną zastosowane odpowiednie kryteria, min.

- Interfejs powinien postugiwać się słownictwem i pojęciami interesariuszy, którzy najczęściej będą się nim postugiwać,
- Interfejs powinien przekazywać jasne i jednoznaczne informacje zwrotne w momencie kiedy generowany jest błąd np. w wyniku kreowania zapytania do bazy danych wpisywanego przez użytkownika,
- Powinien kreować i podpowiadać interakcje dostosowane do wyszukiwanych danych,
- Powinien oferować mechanizmy podpowiadające rozwiązanie w przypadku błędnego wpisania danych do systemu (np. autouzupelnianie pól),
- System powinien być przejrzysty a użytkownik nie powinien mieć wrażenia zbyt dużego przeladowania dostępnymi opcjami. Tam gdzie jest to możliwe operacje powinny być wykonywane w podobny sposób
- stosowane oznaczenia w postaci np. ikon powinny być jednoznacznie opisane i tematycznie związane z prezentowaną treścią,
- wszystkie polecenia powinny mieć możliwość wyboru z menu np. w postaci rozwijanych list wielokrotnego wyboru, a nie wpisywane ręcznie. Ponadto powinny być pokazywane jednoznaczne podpowiedzi w stosunku do rodzaju wpisywanych danych np. format daty, format liczb (dziesiętne, całkowite), długość tekstu itp.

Ze względu na to, że praca z platformą Herbarium Pomeranicum odbywać się będzie poprzez dedykowane aplikacje działające w przeglądarce internetowej, podczas projektowania aplikacji prezentujących cyfryzowane zbiory (portale oraz aplikacje mobilne dla poszczególnych grup użytkowników) zastosowane zostaną wytyczne ujęte w międzynarodowym standardzie dostępności treści internetowych Web Content Accessibility Guidelines (WCAG), którego wersja 2.0 jest międzynarodową normą ISO/IEC 40500:2012 oraz techniki związane z Responsive Web Design, dostosowujące czytelność aplikacji do rozdzielczości urządzenia na którym została uruchomiona. Przede wszystkim nacisk położony zostanie na punkty zawarte w rozporządzeniu Rady Ministrów z dnia 16 maja 2012r. w sprawie

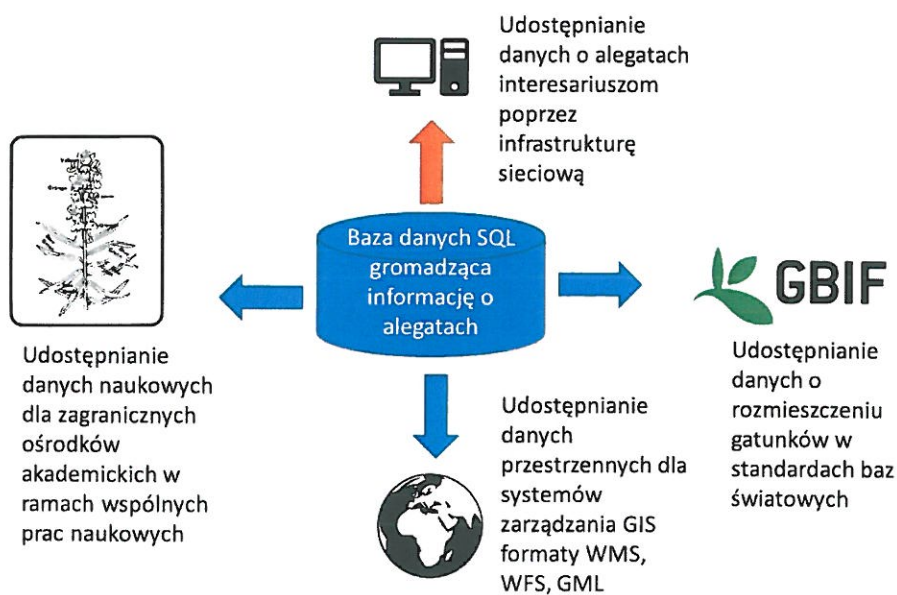
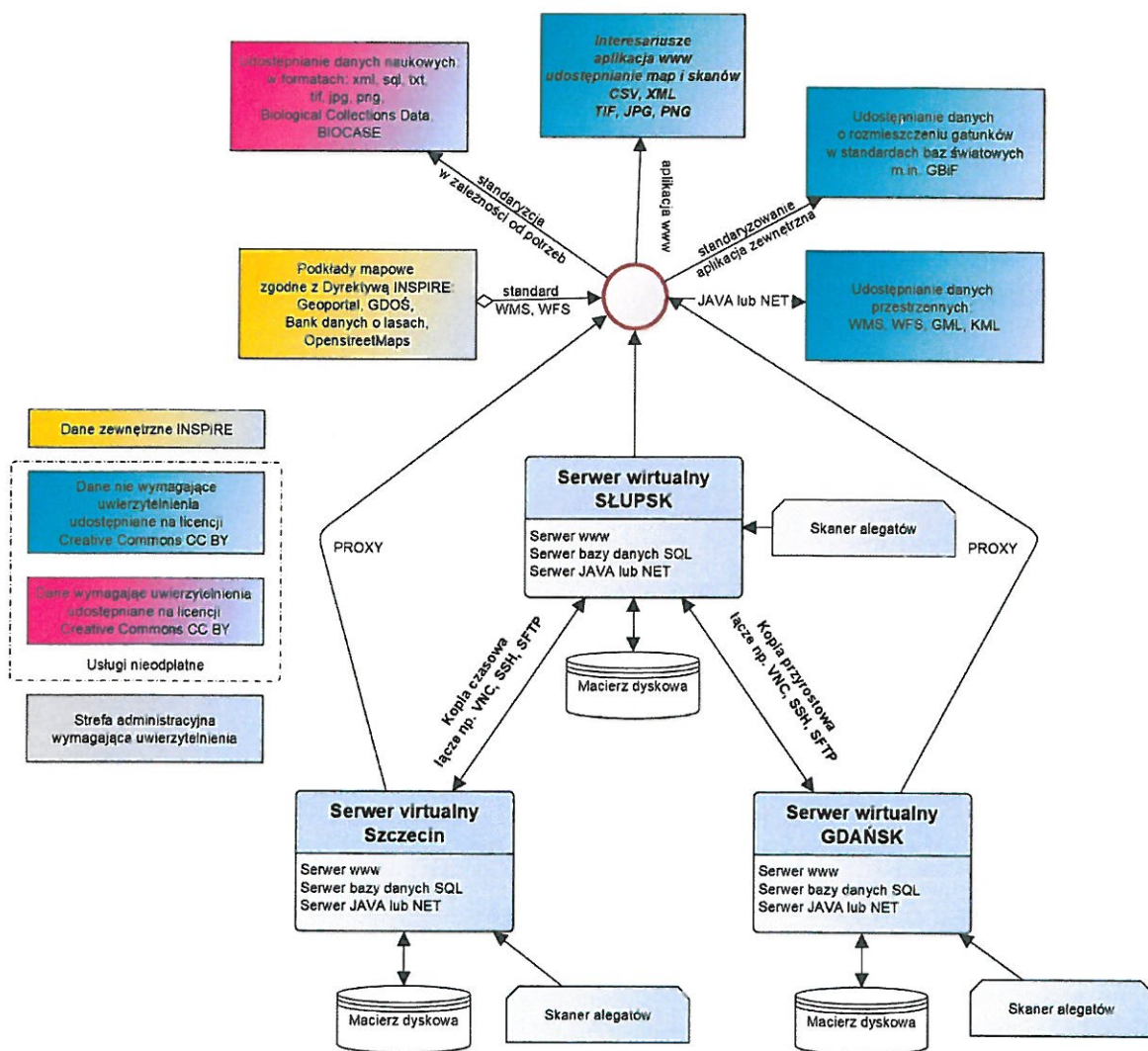
Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych. Biorąc pod uwagę charakter gromadzonych i udostępnianych treści (przede wszystkim treści wizualne: tekstowe i graficzne), najistotniejszym elementem będzie wsparcie dla osób niedowidzących i niewidomych.

Wytyczne standardu WCAG zostaną spełnione między innymi poprzez m.in.:

1. Unikanie elementów animowanych i zmieniających się w sposób agresywny, które rozpraszają użytkowników a w szczególnym przypadku mogą stanowić zagrożenie dla osób cierpiących na padaczkę fotogenną,
2. używanie tekstu strukturalizowanego poprzez użycie nagłówków, w celu zapewnienia łatwiejszej nawigacji po tekście ciągłym dla osób niewidomych,
3. zastosowanie kontrastu treści do tła na poziomie przynajmniej 4,5 do 1,
4. adekwatny opis wszystkich elementów portalu: stron, ramek, tabel, formularzy,
5. umożliwienie powiększenia strony do poziomu przynajmniej 200% bez utraty jej struktury i treści oraz konieczności użycia poziomego paska przewijania
6. zapewnienie pełnej funkcjonalności przy wykorzystaniu samej klawiatury,
7. w przypadku formularzy przyciski wyszukiwarki posiadać będą tekst alternatywny z etykietą opisującą funkcję,
8. w przypadkach, gdzie konieczne będzie użycie mechanizmów CAPTCHA zostaną udostępnione co najmniej dwa alternatywne sposoby rozwiązania, które będzie można wybrać,
9. automatyczne przetaczanie języka strony internetowej po zdeklarowaniu przez użytkownika.

Wszystkie aplikacje wykonane w ramach projektu, będą poddane audytowi pod kątem bezpieczeństwa w specjalizowanym oprogramowaniu np. narzędziem w3af (Web Application Attack and Audit Framework), zgodnie z zaleceniami organizacji OWASP zawartej w specyfikacji OWASP Top 10 2017 The Ten Most Critical Web Application Security Risks. Poziom dostępności zostanie przeanalizowany przez narzędzia zewnętrzne np. A3Web, Achecker, google accessibility, narzędzia W3C itp. Moduły tworzone w ramach projektu poddane zostaną procesowi modelowania zagrożeń opartemu na metodykach STRIDE oraz DREAD. Kod źródłowy aplikacji utworzonych w ramach projektu zostanie poddany audytowi bezpieczeństwa, z szczególnym uwzględnieniem ich kompletności i wzajemnej kompatybilności. Wszystkie osoby związane z projektem zostaną przeszkolone w zakresie bezpieczeństwa, zgodnie z poziomem posiadanych uprawnień.

Zaimplementowane elementy środowiska zostaną skonfigurowane w sposób zgodny z zaleceniami bezpieczeństwa, ale umożliwiającymi spełnienie wymagań funkcjonalnych i późniejszej ich rozbudowy w ramach potrzeb. Komunikacja pomiędzy wszystkimi elementami projektu realizowana będzie w z wykorzystaniem szyfrujących protokołów kryptograficznych np. TLS 1.3, protokół SSH, VPN. O bezpieczeństwo danych dbać będą również administratorzy poszczególnych serwerów



Oprogramowanie do publikacji danych

Założenia:

Ze względu na współczesne standardy stosowane na świecie w podobnych projektach, pełna obsługa udostępniania danych odbywać się będzie za pomocą aplikacji uruchamianych w najpopularniejszych przeglądarkach internetowych tj. Chrome, Firefox, Edge i Opera w najnowszych wersjach. Tym samym zostaną zastosowane najnowsze standardy programistyczne związane z językami Javascript, HTML5, XML oraz PHP i SQL itp.

Z punktu widzenia otwartości danych w ramach Projektu wszyscy Partnerzy zobowiązani są do bezpłatnego cyfrowego udostępniania zbiorów danych wykonanych w ramach projektu bez ograniczeń dla ich ponownego wykorzystania i zgodnie z zasadami tzw. wolnych licencji np. CC BY-SA 3.0 (Creative Commons) z tzw. zachowaniem autorstwa. Wszystkie dane zostaną również „otagowane” co pozwoli na przeszukiwanie metadanych poprzez tagi Highwire Press dla systemów zewnętrznych typu Google Scholar oraz eksport informacji do systemów zarządzania bibliografią naukową (np. formaty RIS, CSV, RDF).

Same zasoby nauki będą udostępniane z wykorzystaniem typów i formatów dokumentów zgodnych z Krajowymi Ramami Interoperacyjności (KRI) m.in.: PDF, XML, TXT, TIF, PNG, JPG itp.

Przed pobraniem danych interesariusz będzie musiał zaakceptować regulamin oraz licencję obejmującą zakres pobieranych danych.

Wszystkie dane zostaną oznaczone odpowiednim znakiem licencji, który będzie jednoznaczny i nie będzie pozostawiał wątpliwości co do zasad jego wykorzystania.

100% zdigitalizowanych w ramach projektu zasobów nauki zostanie udostępniona w formie cyfrowej bezpłatnie, bez ograniczeń dla ich ponownego wykorzystania i zgodnie z zasadami tzw. wolnych licencji np. CC BY-SA 3.0 z tzw. zachowaniem autorstwa. Wszystkie dane zostaną również „otagowane” co pozwoli na przeszukiwanie metadanych poprzez tagi Highwire Press dla systemów zewnętrznych typu Google Scholar oraz eksport informacji do systemów zarządzania bibliografią naukową (np. formaty RIS, CSV, RDF). Digitalizowane w ramach projektu alegaty stanowią zasoby nauki, w związku z tym dostęp do nich będzie otwarty. Wszelkie metadane będą w domenie publicznej z wykorzystaniem licencji CC BY-SA 3.0 (Creative Commons) z tzw. zachowaniem autorstwa.

Każdy interesariusz będzie mógł wykorzystywać dane w celach również komercyjnych z wskazaniem źródła i udostępnieniem na takiej samej licencji. Tym samym zachowana zostanie:

- **Dostępność:** dane będą dostępne w 100% bez opłat w postaci łatwo dostępnej strony internetowej. Ponowne wykorzystanie i redystrybucja: dozwolone będzie ich ponowne wykorzystanie oraz dalsza redystrybucja na zasadach zgodnych z licencją CC BY.
- **Powszechne uczestnictwo:** dostęp do danych będzie miał każdy, bez względu na rodzaj działalności gospodarczej lub jej brak. Każdy będzie miał więc wgląd do zeskanowanych danych oraz powiązanych metadanych, które będzie mógł przeszukiwać.

Zdigitalizowane w ramach projektu alegaty będą posiadały charakter źródłowy (surowy), który następnie będzie mógł być przetwarzany przez interesariusza w specjalizowanych programach do obróbki grafiki. Metadane oraz dane przestrzenne gromadzone w bazach danych będą mogły być pobierane w standaryzowanych formatach umożliwiającymi ich dowolne obrabianie (WFS, XML, txt, csv, GML) i prezentowanie w innych formach w postaci np. map, tabel lub danych surowych dla innych użytkowników.

Same zasoby nauki będą udostępniane z wykorzystaniem typów i formatów dokumentów zgodnych z Krajowymi Ramami Interoperacyjności (KRI) m.in.: PDF, XML, TXT, TIF, PNG, JPG itp.

Wymagane elementy rozwiązań dedykowanych:

1. Zaprojektowanie aplikacji do zarządzania i gromadzenia danych

Projekt polega na skanowaniu i udostępnieniu publicznym „alegatów” (zbiorów zielnikowych) w przestrzeni internetowej. Skanowane obiekty mają być zapisane w postaci plików graficznych w standardach bezstratnych (docelowo tif LZW) oraz stratnym (jpg) mającym na celu szybki bezpośredni dostęp do danych. Wymaga to jednak budowy infrastruktury informatycznej w postaci sprzętu oraz oprogramowania. Aplikacja ma za zadanie gromadzenie informacji o zeskanowanych danych w postaci bazy danych zawierającej opis zbioru w postaci metadanych np. gatunku, autorze zbioru, daty wykonania np. Baza ma umożliwiać również bezpośredni dostęp do zeskanowanych plików i umożliwiać ich pobranie. Aplikacja musi zawierać webowy interfejs użytkownika umożliwiający jednoczesną pracę wielu użytkowników nad wprowadzaniem metadanych oraz zarządzanie zeskanowanymi obrazami. Musi też zapewniać określoną strukturę którą zaproponuje wykonawca. Wymagać to będzie budowy bazy z wieloma tabelami zawierającymi określone relacje, które określi zleceniodawca.

Ponadto możliwe będzie pobranie wprowadzonych metadanych przez anonimowych użytkowników. Informacje te będą mogły być filtrowane, tak aby użytkownik miał pełnię możliwości kreowania zapytań dla jak najlepszego wyniku wyszukiwania. Niezbędne jest też wykonanie modułów raportujących w zadanych konfiguracjach oraz możliwość wydruków w postaci pdf np. etykiet poszczególnych alegatów zawierających informacje wprowadzone do baz danych. Etap ten obejmuje:

- Zaprojektowanie struktury bazy danych,
- Projekt graficzny i wykonanie API dostępnego poprzez przeglądarkę lub w formie „desktop” do zarządzania i wprowadzania danych,
- Projekt graficzny i wykonanie API „web” udostępniającego dane użytkownikowi końcowemu

- i/lub zaproponowanie gotowych rozwiązań płatnych lub bezpłatnych.

2. API do wymiany danych z bazą GBIF

Cyfryzowane dane zgromadzone w trakcie trwania Projektu udostępniane będą również poprzez API rekomendowane przez gremia standaryzacyjne zgodne np. GBIF (The Global Biodiversity Information Facility). Umożliwi to łatwą wymianę danych o rozmieszczeniu gatunków pomiędzy istniejącymi już tego typu projektami na świecie. Dostęp do danych zostanie więc zapewniony zgodnie ze standardem TDWG Access Protocol for Information Retrieval (TAPIR) (<http://www.tdwg.org/activities/tapir/>). Ponadto opis lokalizacji powinien być zgodny ze standardem ATPOL stosowanym w Polskich ośrodkach Akademickich.

Ten element wymaga zaprojektowania API w standardzie umożliwiającym bezpośredni (np. json) lub pośredni (np. xml) udostępnienie danych dla organizacji GbIF.

Aplikacja musi zawierać interfejs użytkownika (GUI) oraz odpowiednią strukturę konwersji danych z bazy do standardu GbIF.

3. API do udostępniania danych geoprzestrzennych (WMS, WFS)

Podmioty związane z wykonywaniem dokumentacji z zakresu ochrony środowiska i przyrody oraz podmioty administracji rządowej, muszą mieć dostęp do map zawierających prezentację przestrzennego rozmieszczenia gatunków prezentowanych przez alegaty zielnikowe. Jest to podstawą do analiz mających na celu zachowanie bioróżnorodności, co jest zgodne z ekologiczną strategią Państwa i Unii Europejskiej. Niezbędne jest więc wykonanie platformy informatycznej udostępniającej tego typu dane w formatach standaryzowanych globalnie przez np. [Open Geospatial Consortium](#) (OGC), umożliwiających bezpośrednią wymianę danych pomiędzy systemami geoinformatycznymi np. WMS, WFS i innymi związanymi z XML tj. Gml, kml. Wszystkie dane będą więc geolokalizowane, co umożliwi wykonywanie map rozmieszczenia zgromadzonych okazów. Mapy te będą wykonywane w zadanej dokładności z wykorzystaniem podkładów mapowych udostępnianych w ramach INSPIRE przez geoportal.

Zapewnienie udostępniania danych w standardach GIS wymienionych powyżej generuje konieczność zastosowania serwerów informacji przestrzennej np. ArcGIS (komercyjny) lub/i Geoserver, Mapserver (Opensource). Opierając się na ideologii budowy systemu na oprogramowaniu OpenSource sugeruje się sprawdzone w innych projektach rozwiązania. Do dyspozycji są dwa serwery danych GIS tj. GeoServer (wykorzystywany np. przez GbIF Global Biodiversity Information Facility – największa baza bioróżnorodności na świecie; bazy ZARRiG i Flora Pomeranica Orientalis) i MapServer (wykorzystywany np. przez geoportale administracji publicznej).

Etap ten wymaga wskazania dedykowanego sposobu wykonania eksportu danych do plików np. geojson, GML, KML itp. oraz udostępnienia danych w formatach OGC np. WMS, WFS, WCS. Zaproponowana powinna zostać zarówno struktura danych jak i format jej wyświetlania wraz z stroną dostępu oraz projektem wykorzystania podkładów mapowych z serwisów internetowych np. OpenStreetMaps, GeoPortal itp.

4. API udostępnienia danych naukowych w formatach oprogramowania do statystyki (np. maxent, statistica, canoko, np.) – oraz formatów xml, txt, csv

Liczba osób korzystających z projektowanego serwisu będzie wyższa, gdy gromadzone w ramach prac będą łatwo dostępne w standardach umożliwiającym integrację ich z oprogramowaniem naukowym. W tym celu przewiduje się wykonanie aplikacji automatycznie przygotowującej dane do dalszych analiz naukowych w formatach csv, txt lub innych wykorzystywanych w statystyce przestrzennej wykonywanej przez najbardziej rozpowszechnione oprogramowanie naukowe np. STATISTICA, MAXEnt, Weka np.

Etap ten wymaga zaproponowania i wykonania odpowiednich interfejsów do udostępnienia eksportu danych do wymaganych szablonów struktury danych, proponowanych przez zleceniodawcę.

Poszczególne API mogą zostać zbudowane od podstaw lub zaproponowane rozwiązanie alternatywne np. w postaci gotowego oprogramowania. Projekt wykonawczy powinien więc zawierać analizę wariantów zakupu gotowego oprogramowania i/lub budowy aplikacji od podstaw.

Rozwiązania software podmiotów zewnętrznych oraz system operacyjny

Wszystkie rozwiązania informatyczne w postaci systemu i aplikacji, będą wirtualizowane z wykorzystaniem platform takich jak np. vmware, virtualbox, hyper-v lub kvm . Projekt wykonawczy powinien zawierać propozycję rozwiązania wirtualizacji wraz z uzasadnieniem wyboru.

Poszczególne aplikacje i uruchomione usługi będą dzielić się na:

- Administracyjne dla systemu np. Openstack czy inne aplikacje dedykowane dla administratorów serwera i przechowywania danych w chmurze,
- Aplikacje obsługiwane przez zatrudnionych przy cyfryzacji danych pracowników, służące do wprowadzania i zapisywania danych na serwerze. Aplikacje będą umożliwiać automatyzację niektórych zadań np. Geolokalizację, automatyczne rozpoznawanie tekstu na etykietach zielnikowych,
- Aplikacje umożliwiające przeglądanie i udostępnianie danych w sieci przez interesariuszy.

Dane w postaci zeskanowanych alegatów, przechowywane będą w specjalizowanych macierzach dyskowych. Zakłada się, że komunikacja z aplikacjami obsługującymi wymianę danych, odbywać się będzie za pomocą krajowej sieci światłowodowej pionier. Jednak ze

względu na dużą ilość danych w postaci plików graficznych, które będą miały znaczną objętość, część z nich będzie przesyłana w partiach w okresie najmniejszego obciążenia sieci. Dotyczyć to będzie zwłaszcza wykonywanych kopii bezpieczeństwa. Na gruncie krajowym sieć pionier posiada umowy peeringowe o wymianie ruchu z wszystkimi większymi operatorami. Umożliwia to spełnienie kolejnego kryterium o wymianie danych pomiędzy chmurą danych naukową oraz chmurą administracji publicznej, jak również gwarantuje szerokopasmowy dostęp obywatela do usług, które opracowane zostaną w projekcie.

L.P.	System źródłowy	System docelowy	Zakres wymianianych danych	Sposób wymiany danych	Typ modyfikacji	Typ Interfejsu
1	Platforma edycji danych	Magazyn danych	Informacje zapisywane w bazie danych SQL oraz pliki rastrowe w formatach bezstratnych. Uzgodnione z Partnerami na etapie projektowania i rozwoju	Kopiowanie i uzupełnianie w bazie danych	-	autorski
2	Platforma udostępniania danych	Dokumenty i pliki	Pliki rastrowe w formatach bezstratnych, pliki pdf, pliki txt, xml, formaty danych GIS WMS i WFS. Uzgodnione z Partnerami na etapie projektowania i rozwoju	Kopiowanie, odwrotania bezpośrednie do pobieranych danych	-	autorski
3	Platforma udostępniania danych w strukturze GBiF	Format baz danych GBiF	Pliki XML, txt, pliki rastrowe.	Kopiowanie i pobieranie danych	-	Interfejs GBiF

Pobrane dane zostaną wyeksportowane do formatów powszechnie używanych do gromadzenia i importu danych tj. txt, csv, xls. Zostanie także zachowana możliwość eksportowania danych do standardów światowych m.in. Access to Biological Collections

Data (ABCD) Schema – jest to standard dostępu do danych biologicznych, opracowany w ramach projektu BIOCASE (Biological Collections Access for Europe). Ze względu na objętość, w bazie danych przechowywany będzie jedynie adres położenia zeskanowanych alegatów w macierzy lub adres url.

Do wykonania zadania niezbędny jest zakup oprogramowania, które zostanie zaprojektowane i zaprogramowane indywidualnie do potrzeb poszczególnych ośrodków. Zawierać ono będzie co najmniej kilka elementów wymagających odrębnych konfiguracji lub rekonfiguracji niektórych elementów dotyczących zwłaszcza wymiany danych między ośrodkami realizowanych za pomocą zabezpieczonych poprzez szyfrowanie VPN. Prawa autorskie do poszczególnych elementów zaprogramowanych samodzielnie przez wykonawcę zostaną przekazane na zamawiającego. W przypadku wykorzystania innych licencji zostaną one uwzględnione odrębnie. Dotyczy to zwłaszcza oprogramowania licencjonowanego przez inne podmioty np. Microsoft lub oprogramowania OpenSource.

Do realizacji zadania potrzebne jest zainstalowanie oraz konfiguracja m.in.:

- Systemu operacyjnego uruchomionego w przestrzeni wirtualnej Linux i/lub Windows Server posiadający wsparcie producenta,
- Baza danych PostgreSQL (opensource),
- Serwery informacji przestrzennej: GeoServer lub Mapserver (opensource)
- Poszczególnych usług oraz samych serwerów i systemów operacyjnych dla potrzeb oprogramowania,
- Odrębnie wykonanego oprogramowania służącego do wymiany informacji i składowania ich w bazie danych.

Projekt wykonawczy powinien zawierać analizę wyboru wariantu:

- **podstawy rozwiązań wirtualizacji,**
- **wykorzystanego systemu operacyjnego,**
- **bazy danych gromadzącej dane wraz ze strukturą oraz dodatków w postaci rozszerzeń, triggerów itp.**
- **rozwiązań webowych w aspekcie wykorzystanego CMS lub projektu własnych rozwiązań lub/i rozwiązań komercyjnych,**
- **oprogramowania do zarządzania usługami serwerowymi,**
- **oprogramowaniem do obsługi macierzy danych,**
- **oprogramowania do wykonywania kopii zapasowych itp.**

Infrastruktura serwerowo-sieciowa wraz z oprogramowaniem do wirtualizacji

W ramach zadania przewidziano zakup, dostawę, instalację i uruchomienie urządzeń infrastruktury sieciowo-serwerowej niezbędnych do uruchomienia zwirtualizowanych serwerów projektowanej platformy "Herbarium Pomeranicum".

Zakupione i uruchomione zostaną trzy centra danych:

- -główne centrum danych umieszczone w Akademii Pomorskiej w Słupsku, który będzie miało za zadanie gromadzenie wszystkich danych oraz obsługę edycji danych jak i ich przeglądanie przez interesariuszy,
- centrum danych umieszczone na Uniwersytecie Gdańskim, które w razie wystąpienia awarii przejmie na siebie wszystkie funkcje głównego centrum,
- centrum danych na Uniwersytecie Szczecińskim przechowujące wersjonowane w określonym zakresie czasowym dane kopii bezpieczeństwa.
- Każde z centr będzie miało zainstalowane na serwerach pełne oprogramowanie umożliwiające wprowadzanie zeskanowanych danych.

Zakupione urządzenia zostaną zainstalowane i podłączone do infrastruktury technicznej w już funkcjonujących serwerowniach:

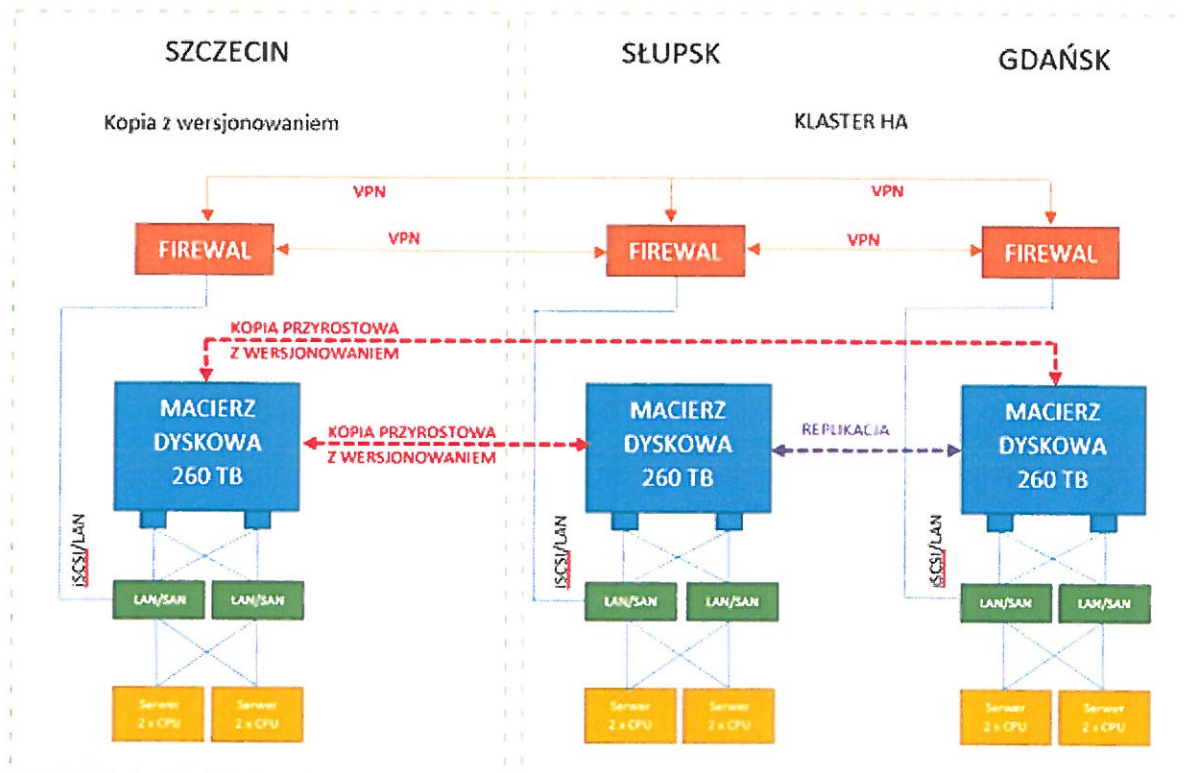
- Akademii Pomorskiej w Słupsku
- Uniwersytetu Szczecińskiego
- Uniwersytetu Gdańskiego

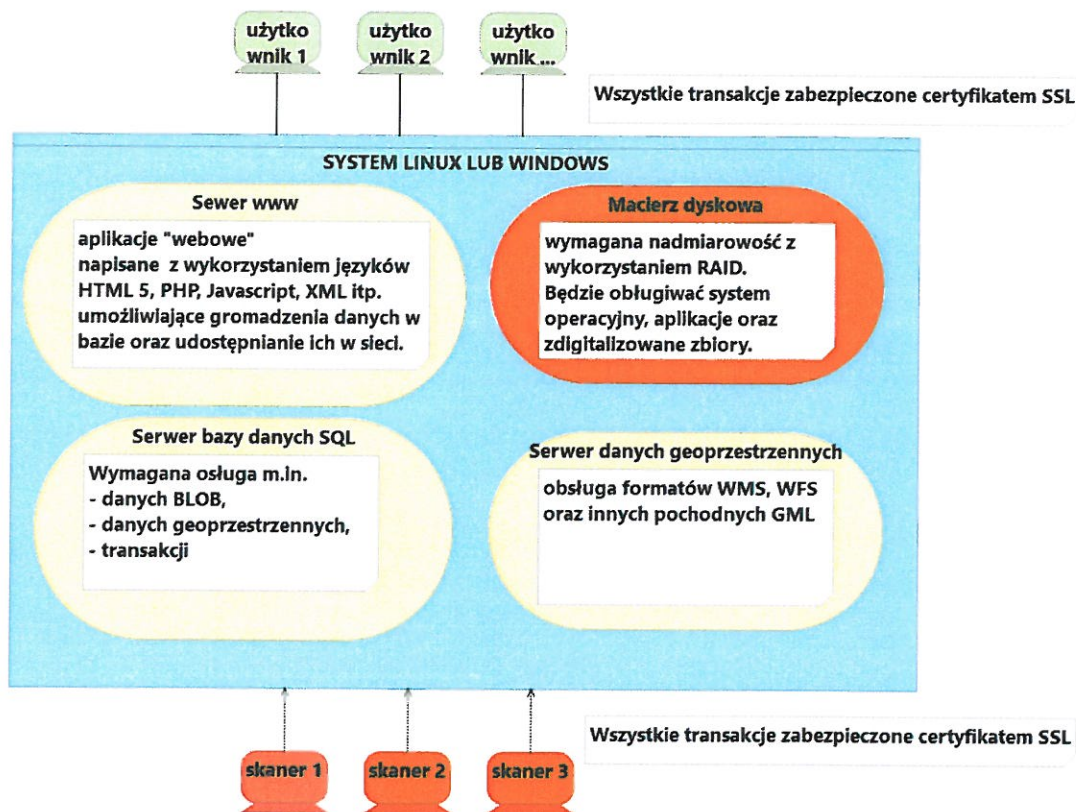
W ramach zadania projektu zaplanowano instalację następującego sprzętu teleinformatycznego w każdej serwerowni:

1. Szafa Rack 19" wraz z niezbędnymi akcesoriami.
2. Zasilacz bezprzerwowy z modułami baterijnymi umożliwiający utrzymanie zasilania urządzeń elektronicznych zainstalowanych w szafie w przypadku wystąpienia krótkich przerw lub nieprawidłowych parametrów zasilania sieciowego. Pojemność baterii musi umożliwić bezpieczne wyłączenie pracujących urządzeń zgodnie ze zdefiniowaną procedurą.
3. Listwy dystrybucji zasilania - dwie sztuki na szafę, zarządzane, z monitorowaniem parametrów środowiskowych wewnątrz szafy (wilgotność/temperatura).
4. Macierz dyskowa z podwójnym kontrolerem obsługująca deduplikację, kompresję, kompaktowanie i thin provisioning zapewniająca przestrzeń użytkową netto (przestrzeń użytkową z uwzględnieniem oferowanego mechanizmu zabezpieczenia RAID bez zastosowania mechanizmu kompresji, deduplikacji i thin provisioning'u, która będzie zaalokowana w 100% - widziana przez HOSTA) w wielkości 260TB. Macierz musi umożliwiać wykonywanie replikacji międzymacierzowej (synchroniczna, asynchroniczna) wolumenów logicznych.
5. Dwa, 24 portowe modułarne (SFP+) przetączniki DataCenter umożliwiające połączenie macierzy, serwerów i firewall'a. Urządzenia dostarczone wraz z kompletem niezbędnych wkładek i kabli umożliwiających połączenie macierzy, serwerów, firewall'a, listw PDU i UPS'a.
6. Dwa serwery w obudowie rack 19, o minimalnej konfiguracji: 2 x Intel(r) Xeon(r) Silver - 4110, 128GB RAM, 2 x zasilacz, 2 x 10G SFP+, dysk/karta pamięci umożliwiająca zainstalowanie oprogramowanie wirtualizatora.
7. Firewall sprzętowy umożliwiający podłączenie do infrastruktury sieciowej w danej lokalizacji, zapewniający minimalną przepustowość komunikacji szyfrowanej VPN 400Mb/s.

Ponadto zakupione zostanie oprogramowanie do wirtualizacji i zarządzania dla sprzętu teleinformatycznego, który zostanie zlokalizowany w serwerowni Akademii Pomorskiej w Słupsku.

Projekt wykonawczy powinien zawierać proponowane rozwiązania sprzętowe oraz software wraz z uzasadnieniem wyboru.





Propozycje rozwiązań z zakresu bezpieczeństwa danych

Na serwerach w ramach projektu nie będą przechowywane informacje podlegające rozporządzeniu o ochronie danych osobowych (RODO). System może być jednak celem ataków komputerowych. Tym samym niezbędne jest zawarcie w projekcie wykonawczym rozwiązań zabezpieczeń oraz testów umożliwiających ocenę bezpieczeństwa gromadzonych danych.

▪ Systemowe

Założono, że infrastruktura związana z realizacją projektu będzie zabezpieczona m.in. poprzez specjalizowane sprzętowe „firewall” w klasie NGFW lub analogiczne rozwiązanie softwarowe, analizujące na bieżąco ruch sieciowy w komunikacji pomiędzy poszczególnymi grupami użytkowników oraz oprogramowanie zapewniające ochronę przed szpiegowaniem i atakami np. typu DDos. Pełny dostęp do danych związanych z ich edycją posiadać będą wyznaczone komputery powiązane funkcjonalnie z stanowiskami do skanowania i edycji danych, które będą uwierzytelnione poprzez przypisany unikalny adres MAC, IP lub „token kryptograficzny”. Administrowanie serwerami odbywać się będzie w z wykorzystaniem najnowszych implementacji protokołu SSH i/lub VPN oraz plikiem\tokenem kryptograficznym.

Struktura i konfiguracja baz danych SQL i noSQL, zostanie opracowana zgodnie z zasadą minimalnych przywilejów, z zachowaniem bezpieczeństwa dla odpowiednich grup użytkowników oraz transakcji. Bazy danych nie będą bezpośrednio udostępniane na

zewnętrzny i stanowić będą jedynie zaplecze (backend) dla wyeksponowanych aplikacji dostępowych uruchamianych w przeglądarkach internetowych.

▪ **Bezpieczeństwo danych**

Głównym zadaniem projektu będzie operowanie na danych przeznaczonych do udostępnienia szerokiemu gronu użytkowników tj. administracji publicznej, podmiotom gospodarczym oraz gremium naukowemu. Informacja ta będzie więc miała charakter publiczny zgodnie z licencją CC BY-SA 3.0 z tzw. zachowaniem autorstwa. W pewnych przypadkach w zależności od poziomu uprawnień użytkownika, może być wymagane zalogowanie, tym samym będą zbierane informacje podlegające ochronie danych osobowych (Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych - RODO). Zalogowanie będzie dotyczyć osób, które będą uczestniczyć w procesie wprowadzania danych bezpośrednio do środowiska bazodanowego. Taka sytuacja może wystąpić w przypadku konieczności weryfikacji alegatów przez eksperta z danej dziedziny botaniki, gdzie bezpośredni dostęp z uprawnieniami edycji może przyspieszyć proces udostępnienia danych interesariuszom. Ponadto dla wygody użytkownika np. zapamiętania wykonanych poprzednio działań, czy listy pobieranych plików możliwe będzie również utworzenie konta, które jednak nie będzie wymagane a dobrowolne. Użytkownicy często chcą personalizować swoje dane w związku z czym taka opcja pojawi się w systemie.

Dla tego typu informacji zostanie zapewniony odpowiedni poziom ich ochrony z podaniem podmiotu, który będzie nimi administrował oraz informacją umożliwiającą użytkownikowi podjęcie działań w celu ich administracji.

Zebrane dane będą przechowywane w bazie danych (informacje o alegatach, użytkownikach, logach systemowych itp.) oraz plikach graficznych (zeskanowane alegaty). Ich bezpieczeństwo będzie zapewnione poprzez redundancję na serwerze macierzystym oraz kopię zapasową również z redundancją na serwerze zewnętrznym, który będzie wykonywał przyrostową (incremental) lub różnicową (differential) kopię zapasową, wykonywaną w określonych interwałach czasowych.

System musi zapewniać przechowywanie logów dotyczących jego użytkowania, w szczególności osób będących posiadaczą podwyższonych uprawnień a także numerów ip i/lub mac adres komputerów łączących się z aplikacjami internetowymi.

▪ **Bezpieczeństwo aplikacji i serwerów**

Wszystkie aplikacje udostępnione w internecie poprzez wybraną domenę, będą poddane audytowi pod kątem bezpieczeństwa w specjalizowanym oprogramowaniu np. narzędziem w3af (Web Application Attack and Audit Framework), zgodnie z zaleceniami organizacji OWASP zawartej w specyfikacji OWASP Top 10 2017 The Ten Most Critical Web Application Security Risks.

Moduły tworzone w ramach projektu poddane zostaną procesowi modelowania zagrożeń opartemu na metodach STRIDE oraz DREAD. Kod źródłowy aplikacji utworzonych w

ramach projektu zostanie poddany audytowi bezpieczeństwa, z szczególnym uwzględnieniem ich kompletności i wzajemnej kompatybilności. Wszystkie osoby związane z projektem zostaną przeszkolone w zakresie bezpieczeństwa, zgodnie z poziomem posiadanych uprawnień. Zaimplementowane elementy środowiska zostaną skonfigurowane w sposób zgodny z zaleceniami bezpieczeństwa, ale umożliwiające spełnienie wymagań funkcjonalnych i późniejszej ich rozbudowy w ramach potrzeb.

Komunikacja pomiędzy wszystkimi elementami projektu realizowana będzie w z wykorzystaniem szyfrujących protokołów kryptograficznych np. TLS 1.3, protokół SSH, VPN. O bezpieczeństwo danych dbać będą również administratorzy poszczególnych serwerów.

Projekt wykonawczy powinien zawierać opis proponowanych rozwiązań ochrony danych i serwerów. Powinna także zostać opracowana procedura testowa.

Wycenę prosimy przesłać za pomocą aktywnego formularza na platformie zakupowej https://platformazakupowa.pl/pn/apsl_edu w terminie do dnia **06.02.2020 r. do godz. 15.00**

UWAGA:

Niniejsze Rozeznanie nie stanowi oferty w rozumieniu przepisów ustawy z dnia 23 kwietnia 1964 r. – Kodeks cywilny (t.j. Dz. U. z 2019 r. poz. 1145), jak również nie jest ogłoszeniem w rozumieniu przepisów ustawy z dnia 29 stycznia 2004 r. – Prawo zamówień publicznych (Dz. U. z 2019 r. poz. 1843 ze zm.).

REKTOR

dr hab. inż. Zbigniew Osadowski, prof. AP