

DDB.201.5.2020

Załącznik Nr 1
do ogłoszenia

Dokumentacja funkcjonalna
serwisu regionalnego agregatora danych
Biblioteki Śląskiej w Katowicach
(wstępny Opis Przedmiotu Zamówienia)

**Opracowanie dla projektu pn.: „Śląskie Digitalium. Digitalizacja i udostępnianie zasobów instytucji kultury województwa śląskiego”,
aplikowanego**

do konkursu ogłoszonego w ramach Regionalnego Programu Operacyjnego Województwa Śląskiego na lata 2014-2020, Osi Priorytetowej II Cyfrowe Śląskie, Działania 2.1. Wsparcie rozwoju cyfrowych usług publicznych.

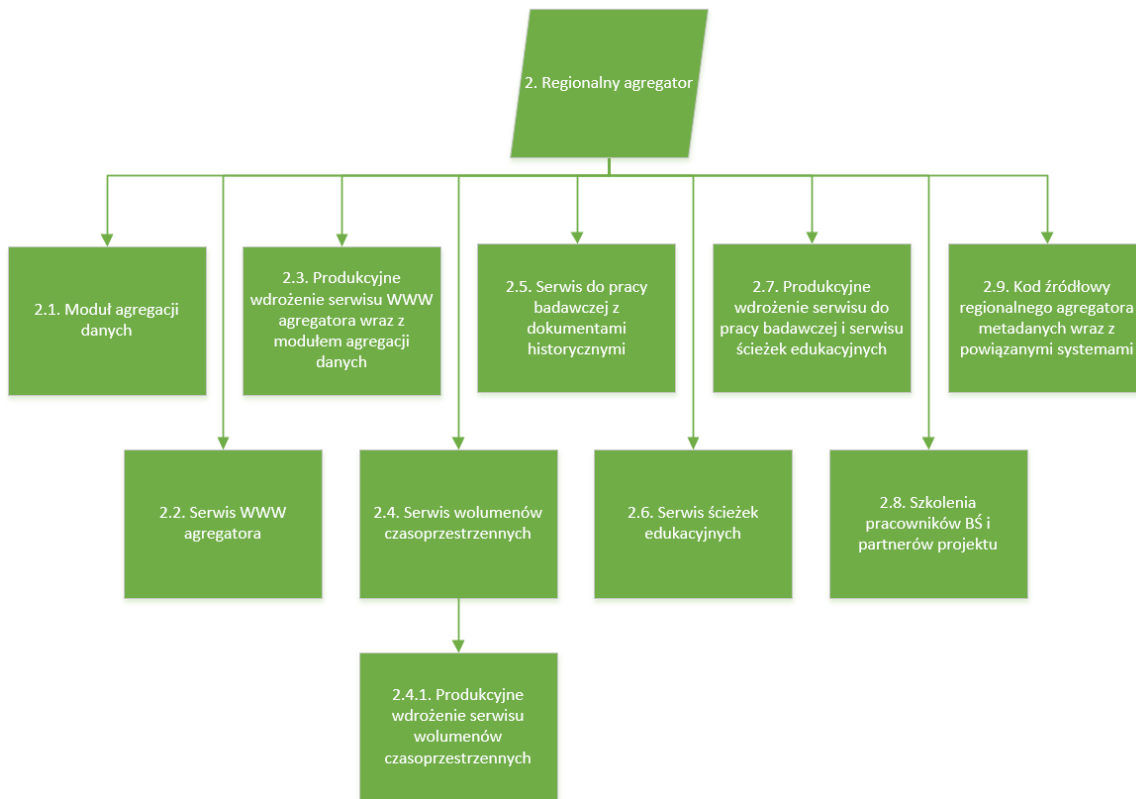
1. Zakres prac i produkty wynikowe

Uruchomienie regionalnego agregatora, który ma stanowić jeden punkt dostępu do dostępnych on-line zasobów instytucji kultury, edukacji i nauki regionu śląskiego, obejmować będzie następujące działania, opisane szczegółowo w dalszych częściach niniejszego opracowania:

- Opracowanie i produkcyjne wdrożenie systemu agregacji danych (zarówno metadanych, jak i treści tekstowych) wraz z dostępowym portalem WWW (produkty 2.1 – 2.3 na Rysunku 1), realizującego podstawowe założenia funkcjonalne agregatora;
- Opracowanie i produkcyjne wdrożenie systemu informacyjnego pozwalającego na zarządzanie wolumenami czasoprzestrzennymi¹, bardzo pomocnego przy pracy badawczej z dokumentami historycznymi (produkty 2.4, 2.4.1 na Rysunku 1);
- Opracowanie serwisu do pracy badawczej z dokumentami historycznymi dostępnymi poprzez agregator (produkt 2.5 na Rysunku 1) i serwisu do budowania ścieżek edukacyjnych na podstawie zasobów agregatora (produkt 2.6 na Rysunku 1) oraz produkcyjne ich wdrożenie w ramach regionalnego agregatora (produkt 2.7 na Rysunku 1).

Po zakończeniu wspomnianych powyżej wdrożeń niezbędne będzie również przeszkolenie pracowników BŚ i partnerów projektu (produkt 2.8 na Rysunku 1). Ponadto, licencja na oprogramowanie regionalnego agregatora wraz ze stowarzyszonymi serwisami, powinna dawać BŚ prawo dostępu do kodu źródłowego i prawo modyfikacji tego kodu na własne potrzeby BŚ (produkt 2.9 na Rysunku 1).

1 Koncepcję wolumenów czasoprzestrzennych opisano szerzej w dalszej części dokumentu.



Rysunek 1. Diagram struktury produktu dla regionalnego agregatora.

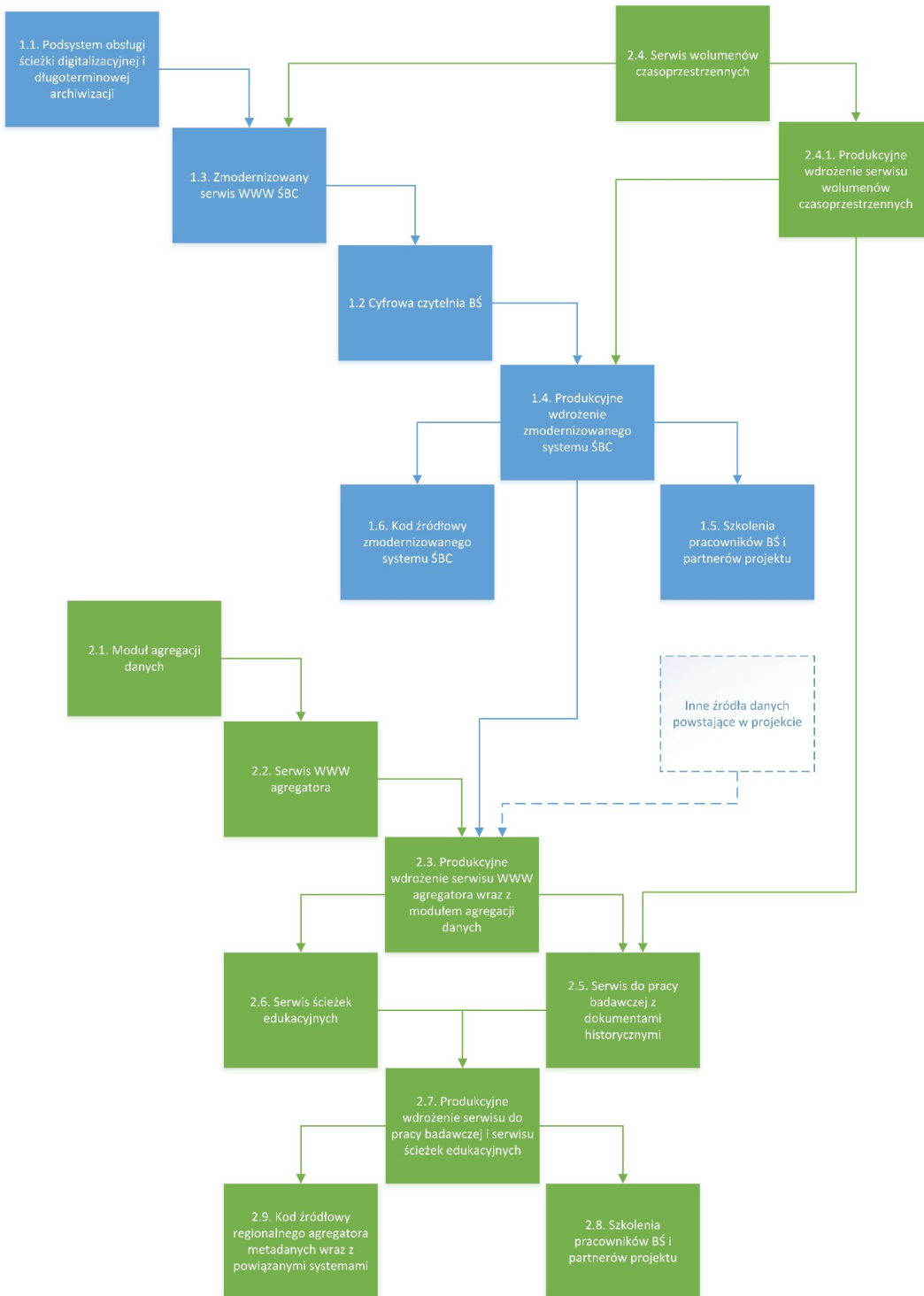
W ramach projektu, w innym zadaniu, zrealizowana zostanie także modernizacja serwisu Śląskiej Biblioteki Cyfrowej (www.scb.org.pl). Ponieważ obydwa produkty (ŚBC i agregator) będą elementami jednego ekosystemu, należy również mieć na uwadze zależności między nimi, kluczowe przy realizacji projektu. Zależności te zostały przedstawione na Rysunku 2, w formie wspólnego diagramu następstwa elementów składowych tychże produktów. Na diagramie ujęto również zależności kolejnościowe pomiędzy elementami składowymi produktów.

Kluczowe zależności pomiędzy produktami są związane z:

- potrzebą wykorzystania odniesień do wolumenów czasoprzestrzennych w ramach opracowywania metadanych obiektów w Śląskiej Bibliotece Cyfrowej — obrazuje zależność pomiędzy produktami 2.4 i 1.3 (narzuca potrzebę wykorzystania API serwisu wolumenów czasoprzestrzennych przy modernizacji Śląskiej Biblioteki Cyfrowej) oraz 2.4.1 i 1.4 (narzuca kolejność produkcyjnego wdrażania zależnych od siebie systemów);
- potrzebą agregacji danych ze Śląskiej Biblioteki Cyfrowej przed publicznym udostępnieniem serwisu WWW regionalnego agregatora – obrazuje to zależność pomiędzy produktami 1.4 i 2.3.



Inne publiczne źródła danych powstające w projekcie (oraz ewentualnie wybrane źródła poza nim) również powinny zostać zagregowane jeszcze przed publicznym udostępnieniem agregatora, przynajmniej na poziomie ustanowienia zautomatyzowanego pobierania danych.



Rysunek 2. Wspólny diagram następstwa produktów.

Projekt współfinansowany przez Unię Europejską z Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Śląskiego na lata 2014 – 2020 „Śląskie Digitalium. Digitalizacja i udostępnianie zasobów instytucji kultury województwa śląskiego”.

2. Specyfikacja funkcjonalna i pozafunkcjonalna regionalnego agregatora metadanych

2.1. Zakres funkcjonalny

Regionalny agregator metadanych, który powstanie w efekcie realizacji projektu będzie rodzajem nowej usługi publicznej – ma stanowić jeden punkt dostępu istniejących, rozproszonych cyfrowych zasobów instytucji kultury, edukacji i nauki regionu śląskiego. Przykładem podobnego polskiego serwisu jest Federacja Bibliotek Cyfrowych (<http://fbc.pionier.net.pl/pro/>), jednak portal FBC agreguje wyłącznie zbiory cyfrowych bibliotek, archiwów i muzeów. Ponadto projektowany agregator regionalny będzie dodatkowo wzbogacony o zintegrowane narzędzia on-line pozwalające na wykorzystanie dostępnych poprzez agregator materiałów w sferze edukacji i nauki. Zakres funkcjonalny poszczególnych elementów składowych agregatora opisano poniżej. Zakłada się, że wszystkie opisane poniżej publiczne serwisy składowe agregatora będą współdzieliły ten sam system uwierzytelniania, dzięki czemu użytkownik będzie mógł się do każdego z nich logować, używając tych samych danych.

Moduł agregacji danych (Produkt 2.1.)

Siła regionalnego agregatora tkwiła będzie w bogactwie dostępnych poprzez ten agregator danych. Stąd też kluczowym elementem tej e-usługi będzie moduł agregacji danych², odpowiedzialny za gromadzenie metadanych i treści tekstowych z rozproszonych źródeł, składowanie ich oraz budowanie na ich podstawie indeksów wyszukiwawczych. Na bardziej szczegółowym Rysunku 5, przedstawiającym architekturę logiczną agregatora, opisywany w niniejszym punkcie moduł agregacji danych obejmuje de facto wszystkie komponenty oznaczone na niebiesko oraz aplikację WWW dla administratorów agregatora.

Agregator regionalny będzie bazował na indeksie wyszukiwawczym zawierającym tzw. dokumenty o ujednocionej strukturze metadanych oraz opcjonalnej zawartości tekstowej. Struktura metadanych będzie oparta o standard Dublin Core (norma ISO 15836:2009), z opcjonalnymi rozszerzeniami (tzw. profil aplikacji Dublin Core), o ile ich wprowadzenie będzie uzasadnione w trakcie prac analitycznych i badań UX zaplanowanych do realizacji na pierwszy etap działań związanych ze stworzeniem agregatora.

Dla każdego zagregowanego zasobu, w procesie normalizacji i przetwarzania danych powstawał będzie dokument o strukturze pasującej do indeksu wyszukiwawczego, który następnie będzie zapisywany w indeksie. Sam indeks będzie zorganizowany w postaci struktury danych optymalnej pod kątem wyszukiwania dokumentów pasujących do określonych (prostych lub złożonych kryteriów). Indeks będzie wspierał też specjalny język wyszukiwawczy pozwalający na stosowanie zaawansowanych operatorów logicznych,

2 Jeżeli w kontekście opisu agregatora mowa o agregowaniu danych, oznacza to zarówno agregowanie metadanych (opisów), jak i powiązanych z nimi treści cyfrowych, o ile są dostępne w postaci tekstu w określonym formacie (por. ostatni akapit tekstu o Produkcie 2.1).

maskowania, wyszukiwania rozmytego, określania ważności poszczególnych elementów zapytania oraz operacji takich jak lematyzacja (lub ewentualnie stemming) tekstu.

W trakcie realizacji projektu zakłada się, że powstanie moduł agregacji danych, który będzie umożliwiał przyłączenie źródeł danych następującego typu:

- Repozytoriów/bibliotek/archiwów/muzeów cyfrowych, które będą udostępniały dane poprzez protokół OAI-PMH (metadane w formacie Dublin Core³ lub pochodnym, treść tekstowa poprzez wskazanie odnośników do plików w metadanych w formacie Dublin Core lub formacie METS⁴, w znacznikach z grupy <filesec/fileGrp/file>). Te standardy są już obecnie implementowane przez ŚBC.
- Serwisów wiedzy (np. encyklopedii) opartych na popularnym otwartym silniku MediaWiki. Ten system jest obecnie wykorzystywany przez Encyklopedię Województwa Śląskiego (<http://ibrbs.pl>).
- Serwisów katalogów bibliotecznych udostępniających metadane poprzez protokół Z39.50 (lub ewentualnie pozwalających na bezpośredni dostęp do bazy danych poprzez zapytania SQL). Agregowanie metadanych poprzez protokół Z39.50 będzie realizowane zgodnie z wytycznymi opracowanymi w ramach projektu TELPlus⁵, który agregował w ten sposób dane z katalogów bibliotecznych bibliotek narodowych krajów UE.
- Jednego rodzaju serwisów typu CMS, opartych na rozwiązaniach open source, udostępniających REST API (np. Wordpress czy Drupal). Wybór wspieranego serwisu CMS zostanie dokonany na podstawie przeglądu technologicznego wybranych serwisów instytucji kultury, edukacji i nauki regionu śląskiego, najbardziej atrakcyjnych treściowo dla planowanego zakresu tematycznego agregatora.
- Poprzez import metadanych ze statycznego pliku CSV oraz XML o określonej strukturze, dostępnego pod zdefiniowanym adresem http – to pozwoli na podstawową realizację agregacji z dowolnym serwisem, który będzie w stanie podstawowe metadane swoich zasobów eksportować do pliku CSV lub XML. Wykorzystanie formatu XML powinno także umożliwić agregację w ten sposób treści tekstowych dokumentów.

3 Np.: http://www.sbc.org.pl/dlibra/oai-pmh-repository.xml?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:www.sbc.org.pl:10794 (metadane XML widoczne w kodzie źródłowym odpowiedzi)

4 Np.: <http://www.sbc.org.pl/dlibra/oai-pmh-repository.xml?verb=GetRecord&metadataPrefix=mets&identifier=oai:www.sbc.org.pl:10794> (metadane XML widoczne w kodzie źródłowym odpowiedzi)

5 Opracowanie „Guidelines for preparing a Z39.50/SRU target to enable metadata harvesting” dostępne online pod [tym adresem](#)

Dodatkowo opracowany zostanie zestaw wytycznych technologicznych pozwalających innym instytucjom na dopasowanie się do metod agregacji wspieranych przez agregator regionalny.

Agregator po skonfigurowaniu połączeń z poszczególnymi źródłami danych funkcjonował będzie w pełni autonomicznie, powiadamiając administratora, gdy zajdzie konieczność ingerencji (np. w przypadku zmiany schematu agregowanych danych w źródle). Dodatkowo administrator będzie miał możliwość wymuszania na module agregacji danych operacji takich jak ponowna agregacja kompletu danych czy ponowna reindeksacja zagregowanych danych (w przypadku zauważenia niespójności). Częstotliwość agregacji będzie dopasowana do możliwości poszczególnych źródeł danych, domyślnie będzie to raz na dobę.

Agregacja treści będzie możliwa, jeżeli źródło danych będzie udostępniało ten tekst w postaci plików tekstowych, plików XML lub będzie on jedną z warstw informacyjnych w pliku PDF, DjVu lub plikach ODT/ODP zgodnych ze standardem ISO/IEC 26300 (Open Document Format).

Serwis WWW agregatora (Produkt 2.2.)

Podstawowe funkcje serwisu WWW agregatora obejmą przeszukiwane zagregowanych i zindeksowanych zasobów, prezentowanie wyników wyszukiwania, prezentowanie szczegółowych informacji o pojedynczym wyniku oraz przekierowywanie użytkownika do adresu URL skojarzonego z danym zasobem (czyli np. do tego zasobu na stronach ŚBC). Funkcje zaawansowane będą związane z filtrowaniem fasetowym i sortowaniem wyników wyszukiwania, personalizacją sposobów prezentacji listy wyników wyszukiwania oraz budową prywatnych kolekcji z interesujących danego użytkownika zasobów wyszukanych w serwisie agregatora. Użytkownicy będą też mogli z poziomu poszczególnych obiektów przejść do serwisu ścieżek edukacyjnych lub do serwisu do pracy badawczej i tam wykorzystać znaleziony obiekt (opcja taka będzie dostępna pod warunkiem zgodności formatu zagregowanego zasoby z formatami akceptowanymi przez określony podserwis). Szczegółowy zakres funkcjonalny zostanie wypracowany w początkowej fazie projektu, z wykorzystaniem metodologii UX opisanej w dalszej części dokumentu.

Dodatkowo serwis agregatora będzie wyposażony w podstawową funkcjonalność CMS umożliwiającą prowadzenie działań informacyjnych i promocyjnych, takich jak tworzenie minikolekcji tematycznych zwracających użytkownikom uwagę na ciekawe zasoby czy informowanie o bieżącej działalności agregatora.

Ponadto agregator będzie posiadał moduł analityczny pozwalający na monitorowanie ruchu użytkowników i ich zachowania w serwisie, a także zliczanie przekierowań użytkowników przechodzących z agregatora do poszczególnych agregowanych serwisów (zarówno na poziomie serwisów, jak i konkretnych zasobów).

Serwis Wolumenów czasoprzestrzennych (Produkt 2.4.)

Serwis wolumenów czasoprzestrzennych będzie pomocniczą usługą pozwalającą w innych serwisach powstających w ramach projektu spójnie opisywać historyczne zasoby cyfrowe w kontekście czasu i przestrzeni. Dane gromadzone w tym serwisie, udostępniane poprzez otwarte API, będą wykorzystywane w Śląskiej Bibliotece Cyfrowej (do wzbogacania metadanych) oraz w opisanym poniżej serwisie do pracy badawczej z dokumentami historycznymi. Dzięki temu możliwe będzie tagowanie dokumentów (w ŚBC) lub ich fragmentów (w serwisie do pracy badawczej) lokalizacjami geograficznymi, ale z uwzględnieniem kontekstu czasowego. Powstające odwołanie będzie dotyczyło zarówno współczesnym koordynat na mapie (w postaci punktu lub obszaru/wielokąta), jak i dawnej, potencjalnie już niestosowanej, nazwy geograficznej czy administracyjnej, możliwe będzie też uwzględnienie dawnych podziałów administracyjnych.

Podstawową funkcjonalnością serwisu będzie możliwość tworzenia wolumenów czasoprzestrzennych i publikowania ich on-line. Zakłada się, że po uruchomieniu serwis zostanie wstępnie zasilony danymi zaimportowanym z istniejących baz takich jak Geonames, ale będzie on też umożliwiał dodawanie kolejnych danych uwierzytelnionym użytkownikom. Użytkownicy ci będą podzieleni na dwie grupy:

- użytkowników profesjonalnych, którzy będą pełnili funkcję administratorów bazy danych i będą odpowiedzialni za jej merytoryczną poprawność – zostaną oni wybrani spośród pracowników BŚ i zainteresowanych współpracujących instytucji;
- użytkowników rejestrujących się samodzielnie, zainteresowanych współtworzeniem danych w serwisie, cechujących się różnym poziomem kompetencji w zakresie wiedzy geograficzno-historycznej.

Dane w serwisie będą podzielone na dwie przestrzenie: dane wiarygodne, wprowadzone lub zaakceptowane przez użytkowników profesjonalnych oraz dane społecznościowe, wprowadzone przez pozostałych użytkowników. Takie podejście pozwoli na rozwój bazy wolumenów czasoprzestrzennych w modelu crowdsourcingu i będzie sprzyjało angażowaniu społeczeństwa w badania naukowe, co jest coraz silniejszym trendem w krajach rozwiniętych Europy Zachodniej (tzn. citizen science⁶) i jest uważane za jeden z istotnych czynników rozwoju społeczeństwa innowacyjnego. Takie podejście jest też uzasadnione badaniami użytkowników bibliotek cyfrowych⁷, które pokazują, że jedną z głównych grup użytkowników zasobów cyfrowych w regionalnych bibliotekach cyfrowych, takich jak ŚBC, są hobbyści zajmujący się badaniami historii regionu lub genealogią.

Osoby wykorzystujące dane zgromadzone w serwisie wolumenów czasoprzestrzennych będą mogły określić, czy chcą korzystać wyłącznie z podzbioru wiarygodnego, czy też są zainteresowane również danymi zebranymi społecznościowo. Będzie to miało duże znaczenie w przypadku serwisu do pracy badawczej, który będzie zrealizowany w analogicznym modelu.

6 <https://ec.europa.eu/digital-single-market/en/citizen-science>

7 Górny Mirosław, Mazurek Jolanta. Key users of Polish digital libraries , The Electronic Library, Vol. 30 Iss: 4, pp. 543-556

Serwis do pracy badawczej z dokumentami historycznymi (Produkt 2.5.)

Kolejnym serwisem składowym agregatora regionalnego będzie narzędzie do pracy badawczej z dokumentami historycznymi. Będzie to uniwersalne narzędzie do transkrypcji, transliteracji czy indeksowania dokumentów historycznych, z możliwością współpracy użytkowników serwisu nad kilkoma warstwami tekstowymi (np. zarówno uzyskanie warstwy tekstowej, jak i opracowanie uwspółcześnionej transkrypcji), z oznaczaniem fragmentów tekstu jako odnoszących się np. miejsc czy osób w tekście, a także budowaniem indeksu haseł oraz słownika na bazie takich dokumentów.

Serwis ten będzie pozwalał na realizację projektów w oparciu o zasoby zgromadzone w agregatorze regionalnym, przy założeniu spełnienia przez nie określonego poziomu interoperacyjności. Możliwa będzie praca na dokumentach zapisanych w formatach graficznych (TIFF, JPG, PNG, PDF, DjVu) oraz dostępnych poprzez protokół IIIF. Oznacza to m.in. większość zasobów Śląskiej Biblioteki Cyfrowej. Funkcjonalność serwisu będzie pozwalała na opracowywanie na podstawie graficznej reprezentacji dokumentu historycznego (np. skan lub zdjęcia) tworzenie dodatkowych tekstowych warstw informacyjnych (transliteracje, transkrypcje, warstwy z adnotacjami itp.). Zakłada się, że będzie możliwość tworzeniu wielu równoległych warstw, a interfejs systemu, przy odpowiednio dużej rozdzielczości ekranu, będzie pozwalał na równoczesną pracę z wizerunkiem graficznym oraz dwiema warstwami tekstowymi (np. transliteracją i uwspółcześnioną transkrypcją). Ponadto na każdej z warstw tekstowym możliwe będzie:

- oznaczanie fragmentów dotyczących osób i tym samym automatyczne budowanie indeksu osób występujących w dokumencie,
- oznaczanie fragmentów dotyczących miejsc (z wykorzystaniem danych z serwisu wolumenów czasoprzestrzennych) i tym samym automatyczne budowanie indeksu miejsc (a właściwie wolumenów czasoprzestrzennych) występujących w dokumencie,
- oznaczanie innych haseł do indeksu (w zależności od potrzeb konkretnego badacza),
- oznaczanie terminów do słownika i tworzenie słownika zawierającego referencje zwrotne do źródła (np. w celu pokazania kontekstu występowania danego słowa w przeszłości).

Schematem danych wykorzystywanym do działania systemu będzie uznany na świecie schemat XML TEI P5⁸, będący dominującym standardem w pracach związanych z zaawansowanymi cyfrowymi edycjami tekstów historycznych. Wykorzystanie tego standardu pozwoli na uzyskanie wysokiej interoperacyjności danych i będzie dodatkową zachętą dla naukowców zainteresowanych wykorzystaniem źródłowych materiałów historycznych w swojej pracy.

8 <https://tei-c.org/Guidelines/P5/>

Użytkownicy serwisu będą mogli realizować projekty badacze indywidualnie lub zespołowo. Bieżące i finalne wyniki realizowanych projektów będą mogły być widoczne tylko dla osób realizujących dany projekt lub też będą mogły być udostępnione w trybie tylko do odczytu, z możliwością poproszenia przez użytkownika o włączenie do danego zespołu badawczego lub możliwością sklonowania publicznego projektu do własnej przestrzeni badawczej w celu jego samodzielnego rozwoju. Dzięki temu będzie powstawał ogólnodostępny katalog opracowanych dokumentów historycznych o warstwie informacyjnej znacznie bogatszej niż same dokumenty źródłowe z podstawowymi metadanymi opisowymi dostępne w bibliotece cyfrowej. Każdy z opracowanych i upublicznionych dokumentów będzie dostępny w postaci interaktywnej, widocznej w przeglądarce internetowej oraz w postaci pliku XML TEI z powiązaniem do danych źródłowych. Powyższy opis funkcjonalności oparty jest na rozmowach przeprowadzonych z naukowcami – humanistami realizującymi różnego rodzaju prace badawcze w oparciu o dane źródłowe z bibliotek cyfrowych. W pierwszej fazie realizacji projektu zakłada się uszczegółowienie tego zakresu, zgodnie z metodyką prac UX opisaną w dalszej części dokumentu.

Podobnie jak w przypadku serwisu wolumenów czasoprzestrzennych, również w serwisie do pracy badawczej zakłada się dwa poziomy wiarygodności użytkowników. Użytkownicy zweryfikowani jako pracownicy biblioteki lub np. osoby zajmujące się zawodowo tego typu pracą naukową będą mogły dodawać swoje projekty do katalogu opracowań profesjonalnych podczas gdy wszyscy inni użytkownicy będą mogli współtworzyć katalog opracowań społecznościowych. Analogicznie jak w przypadku serwisu wolumenów czasoprzestrzennych zakłada się tutaj, że będzie to doskonały serwis do promocji nauk humanistycznych, ich metod i narzędzi wśród członków innowacyjnego społeczeństwa informacyjnego.

Serwis ścieżek edukacyjnych (Produkt 2.6)

Ostatnim z elementów składowych agregatora regionalnego będzie serwis ścieżek edukacyjnych. Serwis ten będzie umożliwiał tworzenie multimedialnych ścieżek edukacyjnych, w postaci edytowanego elektronicznego dokumentu złożonego z opisów tekstowych (z bogatym formatowaniem), adnotacji/komentarzy, odnośników do zasobów internetowych oraz osadzonych dokumentów cyfrowych. Zawartość poszczególnych ścieżek edukacyjnych będzie mogła być budowana w oparciu o materiały wgrywane samodzielnie przez użytkownika oraz przez materiały wybrane z bazy agregatora. Materiały wstawione z agregatora będą prezentowane w postaci odnośników z metadanymi, miniatur lub pełnych osadzonych obiektów, w zależności od możliwości technicznych związanych z formatem konkretnego zasobu.

Każda ze ścieżek edukacyjnych będzie opisana zestawem podstawowych metadanych wg schematu Dublin Core oraz pól specyficznych dla treści edukacyjnych, pozwalających określić wiek odbiorcy (poprzez podanie lat lub klasy czy etapu edukacji) oraz zakres przedmiotowy ścieżki, w odwołaniu do obowiązującej obecnie podstawy programowej. Dzięki temu ze ścieżek edukacyjnych będzie można zbudować katalog uzupełniających

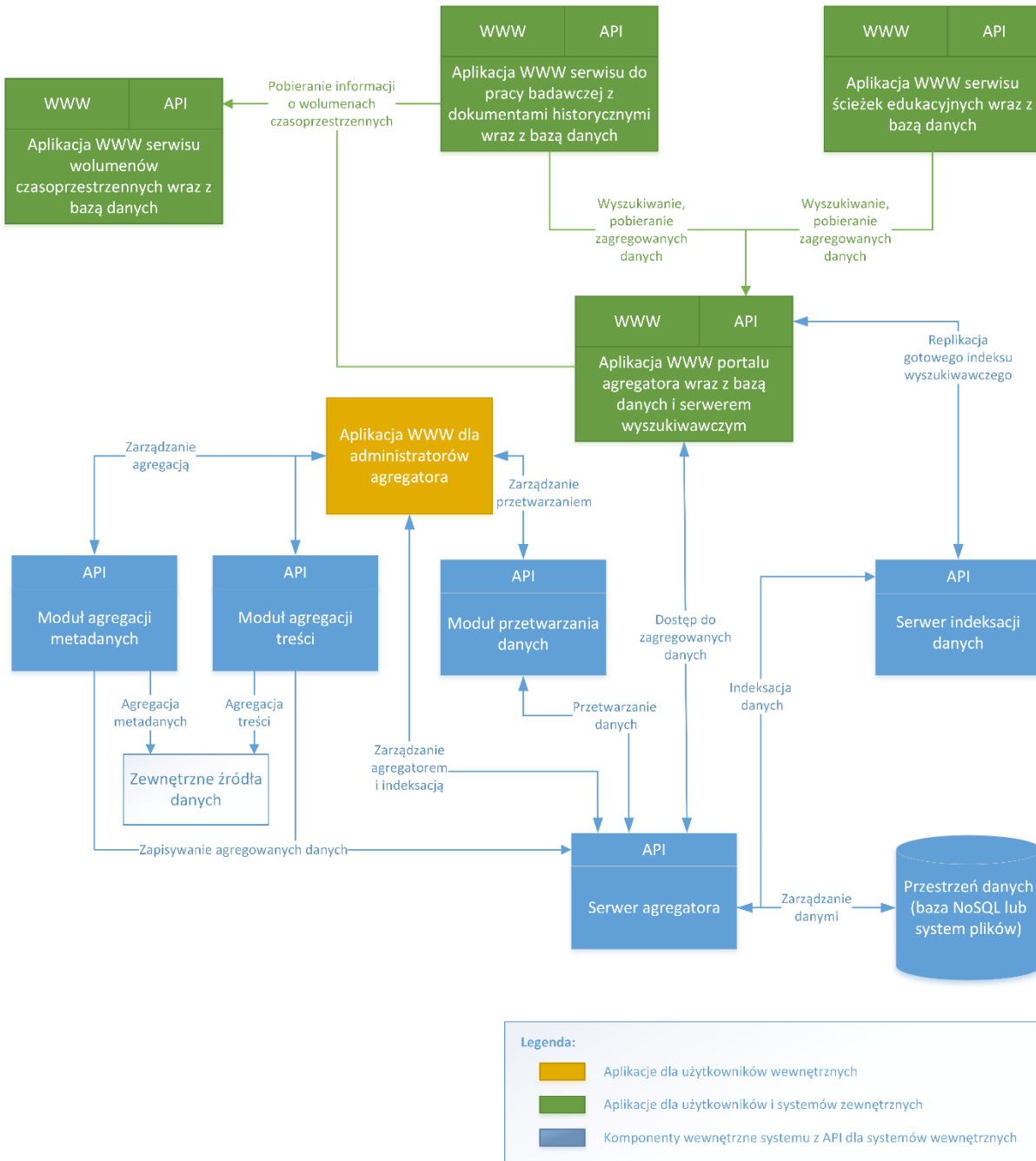
materiałów dydaktycznych, łatwy do przeszukania poprzez zastosowanie filtrów związanych poziomem edukacji czy zakresem przedmiotowym.

Analogicznie jak w przypadku poprzednich serwisów, zakłada się dwupoziomowy model użytkowników systemu. W ramach projektu określona zostanie pewna grupa użytkowników profesjonalnych, którzy będą mogli publikować treści w części zaufanej katalogu oraz moderować treści pozostałych użytkowników (określać jako wiarygodne lub przeciwnie — usuwać). Wszyscy inni użytkownicy posiadający konto w portalu regionalnym również będą mogli tworzyć ścieżki edukacyjne i publikować je, jednak będą one oznaczone jako stworzone przez społeczność, bez gwarancji merytorycznej poprawności.

2.2. Zakres pozafunkcyjny

Architektura Logiczna

Planowana architektura logiczna regionalnego agregatora wraz z powiązаныmi serwisami została przedstawiona na Rysunek 3 poniżej. Zakłada się, że system będzie wyposażony w dwa skalowalne moduły agregacji danych. Jeden z nich będzie odpowiedzialny za agregowanie metadanych, a drugi za agregowanie treści tekstowych. Moduły te będą kontaktowały się okresowo z zewnętrznymi źródłami danych i pobierały informacje związane z nowymi bądź zmienionymi zasobami. Zakłada się opracowywanie specjalizowanych modułów dla różnego typu źródeł danych, w zależności od stopnia ich technologicznej interoperacyjności i wykorzystywanych protokołów. Agregowane dane będą trafiały do głównego serwera agregatora, który będzie odpowiedzialny za zarządzanie nimi. Zarządzanie to w podstawowym zakresie będzie się sprowadzało do: składowania zagregowanych danych, przetwarzania ich do wspólnego schematu, normalizacji i czyszczenia (np. uspoźniania zapisu dat) oraz indeksacji. Zarządzanie pracą agregatora i listą agregowanych źródeł danych odbywało będzie się poprzez dedykowaną aplikację WWW (oznaczoną kolorem pomarańczowym na rysunku).



Rysunek 3. Architektura logiczna regionalnego agregatora wraz z powiązаныmi serwisami.

Projektowana architektura agregatora regionalnego zakłada oparcie indeksacji o dedykowany serwer (wstępnie zakłada się rozwiązania takie jak Solr Cloud lub Elastic Search), będący w stanie budować i replikować indeks pomiędzy różnymi maszynami. Takie rozwiązanie znacząco podnosi niezawodność całego środowiska, a także pozwala na wykonywanie intensywnych obliczeniowo operacji (np. pełne przebudowanie indeksu

wyszukiwawczego po zmienia zasad normalizacji danych) bez zaburzania obsługi żądań użytkowników końcowych, którzy w tym czasie korzystają z działającej równolegle repliki indeksu. W systemie wstępnie uruchomione będą dwa serwery indeksująco-wyszukiwawcze. Jeden z nich będzie odpowiedzialny przede wszystkim za budowanie i aktualizację indeksu (opisany na rysunku jako serwer indeksujący), a drugi będzie odpowiedzialny za obsługę żądań wyszukiwawczych (opisany na rysunku jako serwer wyszukiwawczy). W razie potrzeby serwery będzie można zamienić rolami lub jeden z nich potraktować jako serwer indeksująco-wyszukiwawczy.

Serwer wyszukiwawczy będzie podstawowym źródłem danych dla portalu agregatora, którego kluczową funkcjonalnością będzie przeszukiwanie zagregowanych danych. Metadane będzie można przeglądać na poziomie portalu agregatora, natomiast sięgnięcie do treści będzie powodowało przekierowanie użytkownika do odpowiedniej podstrony serwisu źródłowego (np. Śląskiej Biblioteki Cyfrowej). Portal agregatora, poprzez otwarte API będzie serwował również dane niezbędne do realizacji funkcjonalności serwisu do pracy badawczej z dokumentami historycznymi oraz serwisu ścieżek edukacyjnych. Ponadto serwis do pracy badawczej oraz portal agregatora będą korzystać, przez otwarte API, z danych opisujących wolumeny czasoprzestrzenne, zebranych w serwisie WWW odpowiedzialnym za zarządzanie tymi danymi.

Ponieważ projektowany serwis agregatora regionalnego będzie w praktyce posiadał aż 4 różne składowe serwisy WWW, kluczowe będzie też zapewnianie użytkownikom wrażenia pracy w jednym spójnym środowisku. Częściowo będzie to zapewnione poprzez widoczny na rysunku przepływ danych pomiędzy serwisami, ale niezbędne będzie również zapewnienie jednolitego mechanizmu logowania, w którym nadrzędnym punktem uwierzytelniania będzie główny portal agregatora.

Interoperacyjność i dostępność zasobów

Jak wspomniano wcześniej, systemy tworzone lub rozwijane w ramach projektu nie będą wykorzystywane na potrzeby funkcjonowania rejestrów publicznych (w rozumieniu ustawy o informatyzacji podmiotów realizujących zadania publiczne), a realizacja projektu nie wynika bezpośrednio z konieczności dostosowania lub dołączenia krajowych systemów informacyjnych do systemów informacyjnych w innych krajach Europy. Mimo to agregator regionalny i powiązane z nim serwisy tematyczne będą w pełni przygotowane do udostępniania informacji z zastosowaniem różnorodnych technik interoperacyjności. Każdy uruchomiony serwis WWW posiadać będzie interfejs komunikacyjny (API), który pozwoli na wymianę informacji z zewnętrznymi systemami informatycznymi. W miarę możliwości wykorzystywane będą ustandaryzowane otwarte protokoły komunikacyjne. W sytuacji gdzie nie będzie to możliwe (ze względu na brak szerzej przyjętych standardów komunikacyjnych dopasowanych do specyficznych funkcjonalności projektowanych serwisów), opracowane zostaną dedykowane protokoły komunikacyjne, zgodne z modelem komunikacyjnym REST, a dokumentacja tych protokołów będzie udostępniona publicznie w ramach stron informacyjnych portalu agregatora. Podsumowanie poziomu interoperacyjności i dostępności zasobów zawiera tabela poniżej.

Tabela 1. Analiza interoperacyjności i dostępności poszczególnych publicznych serwisów WWW agregatora regionalnego.

Serwis WWW	Interoperacyjność	Dostępność zasobów (w skali 5 Star Open Data)
Główny serwis agregatora	Protokół OpenSearch umożliwiający przeszukiwanie zagregowanych danych.	4*: Eksport metadanych zasobów w formacie XML/RDF oraz formatach bibliograficznych (np. RIS, BiBTeX)
Serwis wolumenów czasoprzestrzennych	Protokół zapytań semantycznych SPARQL lub protokół OpenSearch (do określenia w fazie projektowania i implementacji systemu), umożliwiający przeszukiwanie zgromadzonych danych.	4*: Eksport danych opisujących wolumeny czasoprzestrzenne w formacie XML/RDF, XML lub JSON.
Serwis do pracy badawczej z dokumentami historycznymi	Dedykowane API REST lub protokół OAI-PMH, umożliwiające pobieranie publicznie dostępnych projektów badawczych użytkowników oraz protokół OpenSearch umożliwiający przeszukiwanie publicznie dostępnych projektów badawczych.	3*: Eksport danych badawczych w formie plików XML / TEI P5 (standard powszechnie stosowany w humanistyce cyfrowej).
Serwis ścieżek edukacyjnych	Dedykowane API REST lub protokół OAI-PMH, umożliwiające pobieranie publicznie dostępnych ścieżek edukacyjnych oraz protokół OpenSearch umożliwiający przeszukiwanie publicznie dostępnych ścieżek edukacyjnych.	3*: Eksport ścieżek edukacyjnych w formacie HTML oraz PDF.

3. Dodatkowe wymogi związane z oprogramowaniem wdrażanym w ramach projektu

3.1. Metodyka rozwoju oprogramowania

Prace techniczne w projekcie będą prowadzone w zgodzie ze zwinnymi metodykami rozwoju oprogramowania, w szczególności wykorzystane zostaną dobre praktyki metodyki Scrum w zakresie rozwoju oprogramowania. Takie wymagania zostaną nałożone na dostawcę oprogramowania poszczególnych e-usług.

W przypadku rozwoju oprogramowania kluczowe będzie zastosowanie krótkich (około 3-tygodniowych) przyrostów funkcjonalnych w rozwijanym systemie. W efekcie opracowywana funkcjonalność będzie na bieżąco pokazywana przedstawicielom BŚ do weryfikacji. Uzyskana zostanie dzięki temu możliwość szybkiej reakcji na zmiany rozwiązania w kontekście identyfikowanych potrzeb użytkownika. Krótkie przyrosty funkcjonalności umożliwiają również dostosowanie procesu rozwoju oprogramowania do bieżących uwarunkowań wynikających z przeprowadzonych inspekcji samego procesu oraz wyników weryfikacji prac przez docelowe grupy użytkowników.

3.2. Produkcyjne wdrożenia, opieka serwisowa i gwarancja na produkty 2.3, 2.4.1 oraz 2.7

Produkcyjne wdrożenia poszczególnych komponentów e-usług uruchamianych w projekcie będą prowadzone przez dostawcę oprogramowania zdalnie (poprzez Internet), bądź lokalnie, przy czym finalny odbiór oprogramowanie będzie realizowany już wyłącznie lokalnie w Bibliotece Śląskiej. Będą to również połączone z opisanymi poniżej szkoleniami i udzieleniem dostępu do kodu źródłowego. Do każdego wdrożenia przez dostawcę zostanie przygotowana dokumentacja powykonawcza, zawierająca opis architektury wdrożenia, oraz informacje niezbędne do uruchamiania, zatrzymywania, monitorowania usług, a także tworzenia kopii bezpieczeństwa i odtwarzania systemu po awarii. Dokumentacja będzie zawierała także wytyczne dotyczące monitorowania stopnia obciążenia i dostępności poszczególnych e-usług oraz wytyczne związane z zapewnianiem bezpieczeństwa poszczególnych e-usług i danych przetwarzanych przez te usługi. W obszarze dokumentacji technicznej przygotowane będą m.in. informacje związane z:

- Dokumentacją architektury i projektu rozwiązania informatycznego wraz z opisem poszczególnych komponentów.
- Dokumentacją wymagań funkcjonalnych oraz poza funkcjonalnych dla funkcji/modułów implementowanych w ramach Projektu.
- Dokumentacją API udostępnianych przez rozwiązania informatyczne w ramach projektu uzupełnioną przykładami kodu wykorzystującymi opisane API.
- Dokumentacją powdrożeniową, która specyfikuje charakterystykę wdrożenia oraz parametry konfiguracyjne działającego systemu.

Rodzaj sytuacji

Problem drobny – nie powoduje żadnych ograniczeń w funkcjonalności systemu, oznacza nieznaczne błędy (np. literówka w etykiecie interfejsu użytkownika)

Zapytanie dot. funkcjonalności systemu

Aktualizacja systemu do nowszej wersji

Dodatkowa asysta techniczna nieujęta powyżej

Opis wsparcia

- Zgłoszenie e-mailowe z automatycznym potwierdzeniem odbioru i reakcja wsparcia w ciągu trzech dni roboczych.
- Zgłoszenie e-mailowe z automatycznym potwierdzeniem odbioru i reakcja wsparcia w ciągu trzech dni roboczych.
- Realizowane zdalnie (przez Internet) przez dostawcę oprogramowania, nie rzadziej niż raz na kwartał o ile w takim okresie została udostępniona przez producenta nowsza wersja oprogramowania.
- Rozliczana na zasadzie ustalonej maksymalnej liczby godzin serwisowych w roku, ujętych w gwarancji, obejmujących również prace programistyczne.
- Zakładane minimum to jakie powinien zaoferować dostawca to:
 - 40 h dla agregatora regionalnego i powiązanych z nim e-usług
- Zgłoszenia w ramach dodatkowej asysty technicznej będą realizowane najpóźniej w ciągu miesiąca od momentu ich zgłoszenia poprzez e-mail, przy czym podstawową jednostką określania czasu obsługi zgłoszenia będzie 30 minut.

Zakłada się, że niezależnie od czasu reakcji serwisu opisanego powyżej, maksymalny czas usunięcia błędów znalezionych w oprogramowaniu (z uwzględnieniem aktualizacji wdrożenia produkcyjnego oprogramowania) nie powinien przekroczyć trzech tygodni od momentu zgłoszenia błędu.

Podmiot świadczący gwarancję i powiązaną z nią opiekę serwisową będzie zobowiązany do używania automatycznego systemu śledzenia obsługi zgłoszeń serwisowych, pozwalającego na utrzymanie ciągłości obsługi zgłoszeń i monitorowanie faktycznego czasu reakcji i rozwiązania zgłaszanych problemów.

3.3. Szkolenia (Produkt 2.8)

W ramach realizacji projektu, dla każdej z uruchamianych e-usług, w ramach wdrażania oprogramowania przeprowadzone zostaną również szkolenia. Szkolenia te będą obejmowały trzy poziomy umiejętności:

- A. obsługę administracyjną każdego z serwisów na poziomie systemowym (aspekty takie jak uruchamianie, zatrzymywanie, monitorowanie działania usługi, architektura wdrożenia, wykonywanie kopii bezpieczeństwa i przywracanie systemu z kopii, lokalizacja i znaczenie poszczególnych plików logów, skalowanie systemu itp.);
- B. obsługę administracyjną każdego z serwisów na poziomie merytorycznym (aspekty takie jak tworzenie kont użytkowników i nadawanie uprawnień, zarządzanie strukturami danych, słownikami, kolekcjami, źródłami danych itp.);
- C. obsługę merytoryczną serwisu:
 - o w zakresie specyficznym dla użytkowników wewnętrznych;
 - o w zakresie specyficznym dla użytkowników zewnętrznych (korzystanie z publicznie dostępnych funkcji poszczególnych serwisów; szkolenie w celu możliwości świadczenia wsparcia użytkownikom końcowym zwracającym się z pytaniami do pracowników BŚ oraz weryfikowania tego, jak wprowadzane zasoby prezentują się od strony publicznej).

W ramach każdego ze szkoleń dostawca oprogramowania prześle również BŚ komplet materiałów szkoleniowych w postaci elektronicznej, na licencji umożliwiającej prowadzenie dalszych wewnętrznych szkoleń samodzielnie przez pracowników BŚ.

Zakłada się, że szkolenia będą realizowane w grupach maksymalnie 15 osobowych dla szkoleń na poziomach (B) i (C) oraz 5 osobowych dla szkoleń na poziomie (A). Szkolenia będą odbywały się na miejscu w budynkach BŚ lub innych wskazanych przez BŚ dogodnych lokalizacjach.

Zakres czasowy szkoleń w zakresie obsługi oprogramowania podsumowuje Tabela 3 poniżej.

Tabela 3. Zestawienie planowanej liczby godzin szkoleniowych dla poszczególnych e-usług realizowanych w ramach projektu.

E-usługa	Liczba godzin szkoleniowych na poziomie A	Liczba godzin szkoleniowych na poziomie B	Liczba godzin szkoleniowych na poziomie C
Produkt 2.1. Moduł agregacji danych	12 h	6 h	2 h
Produkt 2.2. Serwis WWW agregatora			
Produkt 2.4. Serwis wolumenów czasoprzestrzennych	6 h	4 h	6 h
Produkt 2.5. Serwis do pracy badawczej z dokumentami historycznymi	6 h	2 h	6 h
Produkt 2.6. Serwis ścieżek edukacyjnych	6 h	2 h	6 h
Łącznie	30 h	14 h	20 h

3.4. Nabycie prawa do korzystania z oprogramowania oraz kodu źródłowego (Produkt 2.9)

Mając na uwadze charakter instytucji jaką jest Biblioteka Śląska, zaleca się, aby oprogramowanie nabywane / rozszerzane w ramach projektu było ostatecznie udostępniane BŚ na licencji otwartej lub na licencji umożliwiającej Bibliotece modyfikowanie kodu prowadzone niezależnie od producenta (w drodze prac własnych lub zleconych podmiotom trzecim).

W związku z powyższym ważne jest, aby w umowie z wykonawcą oprogramowania zawarte zostały następujące wymagania co do kodu źródłowego:

- Kod źródłowy musi być udokumentowany na poziomie technicznym (dokumentacja w plikach z kodem), w sposób zgodny z powszechnie stosowanymi standardami w danych języku programowania. W przypadku wykorzystania języków programowania obiektowego dokumentacja taka powinna być zapewniona co najmniej na poziomie elementów (klas, metod, zmiennych itd.) kodu deklarowanych jako publiczne. Dokumentacja taka może być sporządzona w języku polskim lub angielskim.
- Kod źródłowy powinien być czytelny. Dostawca powinien określić jednoznacznie stosowaną konwencję kodowania (może to być zrobione poprzez podanie

konfiguracji dla automatycznego formatera kodu). Ponadto kod źródłowy nie może być poddany procesom zaciemniania. Jeżeli zaciemnianie kodu jest wymagane ze względów wydajnościowych (np. optymalizacja plików JavaScript), producent powinien również udostępnić pliki niezaciemnione.

- Dokumentacja powinna zawierać instrukcję „krok po kroku” umożliwiającą zbudowanie i uruchomienie poszczególnych komponentów systemu na podstawie kodu źródłowego.
- W przypadku wykorzystania przez producenta komponentów opracowanych przez podmioty trzecie, producent powinien dostarczyć wykaz takich komponentów wraz z określeniem ich producenta oraz licencji, na jakiej komponent jest udostępniany. W takiej sytuacji stosowanie komponentów o zamkniętym kodzie źródłowym dopuszcza się tylko, jeżeli będą to popularne komponenty ogólnego zastosowania (np. komercyjny silnik relacyjnej bazy danych lub komercyjna biblioteka do konwersji pomiędzy formatami dokumentów). W takiej sytuacji koszt dostarczenia licencji pozwalającej na uruchomienie systemu powinien leżeć po stronie dostawcy.

3.5. Użyteczność i doświadczenia użytkownika (UX)

Projektowanie oprogramowania w projekcie będzie prowadzone zgodnie z zasadami projektowania zorientowanego na użytkownika, w szczególności wykorzystana zostanie w tym celu norma PN-EN ISO 9241-210:2011 „Ergonomia interakcji człowieka i systemu – Część 210: Projektowanie ukierunkowane na człowieka w przypadku systemów interaktywnych”. Wymagania takie, zgodne z przykładowym procesem projektowania opisanym poniżej, zostaną nałożone na dostawcę oprogramowania.

Dzięki temu, w toku prac kluczowe będzie pełne zaangażowanie przedstawicieli z poszczególnych grup użytkowników w proces projektowania, w szczególności dotyczyć to będzie udziału w testach, uczestnictwa w grupach roboczych oraz wywiadach pogłębionych, udział w ankietach oraz prowadzonych analizach, a także poddanie się badaniu w zakresie zachowania użytkowników w interakcji z istniejącymi lub projektowanymi rozwiązaniami.

W początkowej fazie projektowania wykonana zostanie ocena podsumowująca⁹ (ang. summarative report) dla projektu interfejsu użytkownika agregatora. Dane z tej ewaluacji posłużą do wstępnej identyfikacji funkcji, zadań i informacji potrzebnych do opracowania projektów interfejsu użytkownika dla nowoprojektowanego serwisu agregatora regionalnego.

Następnie, w celu uzyskania adekwatnych rozwiązań technologicznych proces projektowania interfejsów użytkownika oprogramowania będzie złożony z czterech

⁹ Na przykład wg normy ISO/IEC 25062 Software engineering — Software Product Quality Requirements and Evaluation (SQuaRE) - Common Industry Format (CIF) for usability test reports.

zasadniczych etapów, pogłębiających i doprecyzowujących wymagania funkcjonalne co do systemów, opisane w niniejszym dokumencie:

1. Zrozumienie i specyfikacja kontekstu użytkowania projektowanego systemu lub jego części.
2. Specyfikacja wymagań użytkownika.
3. Opracowanie projektów spełniających wymagania i oczekiwania użytkownika.
4. Przeprowadzenie oceny opracowanego projektu w kontekście wymagań użytkownika.

Aby uzyskać możliwie najlepszy produkt etapy te będą iteracyjnie powtarzane w celu uzyskania pełnej akceptacji wszystkich osób biorących udział w procesie projektowania, w szczególności przedstawicieli poszczególnych grup użytkowników. Rozpoczęcie kolejnej iteracji polega na powrocie z etapu 4. do dowolnego innego etapu procesu projektowania rozwiązań – w zależności od potrzeb. Każdy z przewidywanych do realizacji etapów skutkować będzie przygotowaniem odpowiednich rezultatów, np. dokumentacji wymagań użytkownika lub prototypami w zakresie interfejsu użytkownika projektowanego rozwiązania, lub funkcji. Poszczególne rezultaty etapów będą oparte o adekwatne międzynarodowe normy i standardy lub przyjęte powszechnie dobre praktyki. W szczególności rezultaty poszczególnych etapów obejmować będą:

1. Dla etapu dotyczącego zrozumienia i specyfikacji kontekstu użytkowania projektowanego systemu lub jego części:
 - a. Opis zidentyfikowanych grup użytkowników oraz interesariuszy wraz z powiązaniem w kontekście kluczowych celów oraz ograniczeń.
 - b. Charakterystyka zidentyfikowanych grup użytkowników, np. poprzez przygotowanie profili użytkowników oraz definicję tzw. person - reprezentatywnych użytkowników systemu.
 - c. Opis kluczowych celów i zadań, np. poprzez przygotowanie scenariuszy dla zadań, które będą realizowane przez użytkownika w kontekście korzystania z projektowanego systemu/funkcji lub przygotowanie modeli koncepcyjnych funkcjonowania systemu i jego komponentów z perspektywy poszczególnych grup użytkowników. Opisy będą uwzględniać również aspekty niezwiązane z funkcjami czy właściwościami systemu, np. częstotliwość wykonywania zadania, czynności do wykonania równoległe by móc w pełni zrozumieć kontekst funkcjonowania użytkownika.
 - d. Opis środowiska, w którym funkcjonuje system, włączając w to środowisko sprzętowo-programowe, materiały lub urządzenia, charakterystykę społecznych lub kulturowych aspektów, a także fizycznego otoczenia użytkownika, który wykorzystuje oprogramowanie.

2. Dla etapu dotyczącego przygotowania specyfikacji wymagań użytkowników wykorzystana zostanie norma „NISTIR 7432 Common Industry Specification for Usability-Requirements”, która w ogólności składa się z trzech części, mianowicie:
 - a. Kontekst użytkowania systemu, zawierający informacje m.in. o użytkownikach systemu i ich celach wraz z przykładowymi scenariuszami, środowisku, w którym system będzie działał oraz wyposażeniu związanym z funkcjonowaniem systemu.
 - b. Kryteria wydajności i satysfakcji użytkownika, które przedstawiają sposoby na mierzenie użyteczności planowanego produktu w zakresie scenariuszy korzystania z systemu.
 - c. Metody testowania, opisujące sposoby testowania systemu, które w efekcie pokażą, czy system spełnia zdefiniowane kryteria wydajności i satysfakcji użytkownika.

Punkty b. oraz c. posiadają 3 poziomy szczegółowości, co oznacza, że wraz z kolejnymi iteracjami procesu projektowania uszczegóławiane będą informacje na temat kryteriów wydajności, kryteriów satysfakcji użytkownika oraz adekwatnych metod testowania.
3. Dla etapu dotyczącego opracowania projektów — przygotowania rysunków i szkiców odzwierciedlających pomysły w zakresie interfejsu użytkownika. Projektowanie rozpocznie się od tzw. projektów niskiej szczegółowości (ang. low-fidelity), pobudzających dyskusje z docelowymi użytkownikami w zakresie kluczowych charakterystyk projektowanych rozwiązań (np. szkiców zawierających jedynie idee związane z zawartością poszczególnych ekranów, bez precyzji w zakresie docelowego wyglądu). Z biegiem czasu, w efekcie konsultacji z docelowymi użytkownikami i na podstawie ich uwag, projekty nabierać będą coraz większej precyzji, idąc w kierunku projektów wysokiej szczegółowości (ang. high-fidelity). Pozwoli to w dalszej kolejności na implementację przygotowanych projektów (po pozytywnej ich ocenie przez użytkowników). Szkice będą przygotowywane z wykorzystaniem oprogramowania do tworzenia szkiców/atrap interfejsów (ang. wireframes/mock-ups), np. Axure. W razie konieczności przygotowane zostaną makiety projektowanego systemu wraz z zamarkowanymi interakcjami pomiędzy poszczególnymi ekranami.
4. Przeprowadzanie oceny (ewaluacji) opracowanego projektu będzie się odbywać iteracyjne w formie tzw. oceny kształtującej (ang. formative evaluation), której celem jest weryfikacja zgodności projektowanego rozwiązania z wymaganiami użytkowników przygotowanymi w ramach 2-go etapu procesu projektowania. W efekcie każdej takiej sesji oceniającej wytworzony zostanie formularz oceny danego

prototypu, co będzie równocześnie stanowić dane wejściowe dla kolejnej iteracji w procesie projektowania. W proces oceny zaangażowani będą przedstawiciele poszczególnych grup użytkowników. W końcowej fazie projektowania interfejsów dla planowanego systemu przeprowadzona zostanie ewaluacja podsumowująca (summarative evaluation), w efekcie czego uzyskany zostanie raport podsumowujący.











Wyżej opisane działania, prowadzone przez dostawcę oprogramowania we współpracy z BS i partnerami projektu, mające doprowadzić do użytecznego, ergonomicznego i efektywnego rozwiązania, będą prowadzone przy założeniach:

- Pełnego zaangażowania użytkowników w całym procesie projektowania i rozwoju rozwiązań technicznych.
- Zaangażowania w prace rozwojowe multidyscyplinarnego zespołu, który będzie w stanie w pełni zrozumieć użytkowników, ich zadania oraz środowisko, w którym funkcjonować będzie system.
- Wykonywania ocen przez użytkowników systemu, które pozwolą na udoskonalenie projektowanego rozwiązania.

3.6. Wielojęzyczność interfejsu użytkownika

Interfejs użytkownika publicznych serwisów WWW, które zostaną utworzone lub zmodernizowane w ramach projektu, będzie wielojęzyczny. Na podstawie danych dotyczących użytkowników Śląskiej Biblioteki Cyfrowej (patrz Tabela 5, dane za rok 2016), można stwierdzić, że interfejs użytkownika powinien być dostępny w trzech wersjach językowych: polskiej, niemieckiej i angielskiej. Ponadto oprogramowanie będzie przygotowane w taki sposób, aby w przyszłości można je było rozszerzyć o kolejne wersje językowe interfejsu, bez konieczności interwencji programistycznych. Opisana tutaj wielojęzyczność dotyczy elementów interfejsu użytkownika, a nie danych gromadzonych w systemie.

Tabela 4. Statystyki odwiedzin serwisu ŚBC w 2016 r. z podziałem na kraje odwiedzających i ich ustawienia językowe (dane wg Google Analytics).

Kraj ?	Język ?	Sesje ? ↓
		556 929 % całości: 100,00% (556 929)
1.  Poland	pl	355 882 (63,90%)
2.  Poland	pl-pl	64 933 (11,66%)
3.  Germany	de	32 052 (5,76%)
4.  Germany	de-de	18 092 (3,25%)
5.  United States	en-us	7 468 (1,34%)
6.  Poland	en-us	6 695 (1,20%)
7.  Czech Republic	cs	4 755 (0,85%)
8.  Ukraine	ru	3 038 (0,55%)
9.  Ukraine	uk	2 665 (0,48%)
10.  Germany	pl	2 660 (0,48%)

3.7. Mobilność

Opracowane w ramach projektu publiczne portale internetowe przygotowane zostaną zgodnie z techniką określaną Responsive Web Design (RWD). Oznacza to, że treść stron internetowych jest przekazywana do przeglądarki internetowej w jednej formie, natomiast aspektem wizualnym sterują techniczne pliki stylu (CSS), które definiują różnego rodzaju wygląd strony w zależności od parametrów ekranu urządzenia użytkownika. Dzięki temu rozwiązaniu portale będą dostępne dla użytkowników w dowolnym miejscu i na dowolnym urządzeniu (np. mobilnym, stacjonarnym), które ma dostęp do Internetu i obsługuje wyświetlanie stron w standardzie HTML5.

Tabela 5. Udział procentowy różnego rodzaju urządzeń i rozdzielczości ekranów (najpopularniejsze 20 kombinacji) wśród użytkowników Śląskiej Biblioteki Cyfrowej (dane zebrane przez Google Analytics, za okres roku 2016).

Kategoria urządzeń ?	Rozdzielczość ekranu ?	Sesje ? ↓
		556 725 % całości: 99,96% (556 929)
1. desktop	1366x768	150 638 (27,06%)
2. desktop	1920x1080	64 410 (11,57%)
3. desktop	1280x1024	47 096 (8,46%)
4. desktop	1280x800	33 564 (6,03%)
5. desktop	1600x900	33 055 (5,94%)
6. desktop	1440x900	27 361 (4,91%)
7. mobile	360x640	21 685 (3,90%)
8. desktop	1680x1050	20 127 (3,62%)
9. desktop	1024x768	19 197 (3,45%)
10. desktop	1536x864	10 463 (1,88%)
11. desktop	1280x720	9 314 (1,67%)
12. desktop	1920x1200	8 646 (1,55%)
13. tablet	768x1024	7 172 (1,29%)
14. desktop	1360x768	5 259 (0,94%)
15. desktop	1280x768	5 232 (0,94%)
16. mobile	320x534	3 955 (0,71%)
17. tablet	1280x800	3 765 (0,68%)
18. desktop	1344x840	3 384 (0,61%)
19. desktop	1024x819	3 380 (0,61%)
20. desktop	1024x600	3 155 (0,57%)

Tabela 5 powyżej pokazuje udział procentowy różnego rodzaju urządzeń i rozdzielczości ekranów (najpopularniejsze 20 kombinacji) wśród użytkowników Śląskiej Biblioteki Cyfrowej (dane zebrane przez Google Analytics, za okres roku 2016). Sumarycznie około 86,92% sesji odbyło się w tym okresie z wykorzystaniem urządzeń typu PC/notebook

(oznaczenie desktop w tabeli), 9,69% sesji z telefonów komórkowych (oznaczenie mobile w tabeli) i 3,39% sesji z tabletów (oznaczenie tablet w tabeli). To wyraźnie potwierdza potrzebą zastosowania podejścia RWD i pokazuje również potrzebę projektowania portali także pod duże rozdzielczości (w szczególności dedykowany widok Full HD).

3.8. Dostępność dla osób niepełnosprawnych

Portale udostępniające informacje w projekcie będą dostosowane do wytycznych WCAG 2.0 w stopniu przewyższającym wymagania zdefiniowane w KRI (czyli ponad poziom AA). Funkcje te będą zrealizowane tak, aby nowy portal osiągnął dostępność dla osób niepełnosprawnych na poziomie opisanym poniżej. To samo założenie dotyczy wszystkich innych publicznych portali, które będą tworzone od podstaw w niniejszym projekcie.

Zgodność portali tworzonych lub modyfikowanych w projekcie z wytycznymi WCAG 2.0 będzie przy wdrażaniu oprogramowania weryfikowana z wykorzystaniem istniejących w tym celu narzędzi, np. stron internetowych z adekwatnymi walidatorami. poniżej przedstawia wymagania WCAG 2.0, które zostaną spełnione przez wszystkie portale tworzone i modyfikowane w projekcie, poprzez określenie ich jako minimum, które musi spełnić dostawca oprogramowania. Każde wymaganie posiada adekwatne uzasadnienie. Pozycje w tabeli przedstawiają wszystkie wymagania wyspecyfikowane w ramach KRI wraz z uzasadnieniem. Pozycje 10 oraz 23 przedstawiają elementy przewyższające wymagania KRI i są oznaczone pogrubieniem. Oznacza to, że rozwiązania techniczne, planowane do realizacji w ramach Projektu przewyższają minimalne wymagania wyspecyfikowane w KRI.

Zgodność portali tworzonych lub modyfikowanych w projekcie z wytycznymi WCAG 2.0 będzie przy wdrażaniu oprogramowania weryfikowana z wykorzystaniem istniejących w tym celu narzędzi, np. stron internetowych z adekwatnymi walidatorami.

Tabela 6. Wymagania WCAG 2.0 zaplanowane do spełnienia przez wszystkie portale tworzone i modyfikowane w projekcie.

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
1.	1.1.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby elementy nietekstowe (tj. pliki graficzne) miały adekwatną reprezentację tekstową, np. miniaturka obiektu cyfrowego będzie opatrzona odpowiednim opisem pochodzącym z metadanych obiektu. Ponadto, w przypadku udostępnianych zasobów cyfrowych każdy obiekt cyfrowy opatrzony będzie metadanymi, które

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		opisują jego zawartość.
2.	1.3.1, poziom A	<p>Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by poszczególne strony zawierały informacje o logicznych powiązaniach pomiędzy poszczególnymi elementami strony, a same elementy były odpowiednio semantycznie zidentyfikowane. Przykładowo, strona główna będzie miała zaznaczone obszary związane z nagłówkiem, z formularzem wyszukiwawczym, a także z prezentacją przykładowych/interesujących obiektów z biblioteki cyfrowej. Dotyczy to również wykorzystywania odpowiednich znaczników dla tabel lub list elementów (jeśli takie są prezentowane), a także grupowania powiązanych ze sobą elementów (np. opcji wyboru lub pól formularza).</p> <p>Niezależnie od tego portale internetowe będą tworzone zgodnie z praktykami Responsive Web Design¹⁰, co oznacza, że treść strony zapisana będzie w ustrukturyzowanym formacie zgodnym z HTML, natomiast wygląd (ułożenie) tych treści będzie determinowany odpowiednimi plikami stylu (CSS). Gwarantuje to niezależność struktury przekazywanych informacji od ich prezentacji na ekranie urządzenia wyświetlającego informacje.</p>
3.	1.3.2, poziom A	<p>Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by poszczególne ich elementy były umieszczone w odpowiedniej kolejności, zgodnej ze znaczeniem prezentowanych informacji. Przykładowo, strona internetowa zawierająca informacje ogólne na temat obiektu cyfrowego będzie posiadała w pierwszej kolejności link do treści obiektu, a następnie metadane (lub odwrotnie - zależnie od wyników badań w zakresie potrzeb użytkowników). Będzie to niezależne od mechanizmu prezentowania informacji na stronie, gdzie część elementów metadanych może być wyświetlana przed linkiem do obiektu (np. tytuł i autor).</p>

10 https://en.wikipedia.org/wiki/Responsive_web_design

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
4.	1.3.3, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w ten sposób, by elementy, które mają rolę inną niż ich techniczna reprezentacja, były opatrzone odpowiednim opisem tekstowym. W szczególności dotyczyć to może opcji zwiększania czcionki na stronie internetowej, które zwykle reprezentowana jest przez serię liter (grafik) „A” o różnym rozmiarze. W przypadku występowania na stronie tego rodzaju elementów, będą one opatrzone stosownym opisem, który precyzuje ich znaczenie.
5.	1.4.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by wykorzystywane na nich kolory nie były jedynym sposobem na przekazanie użytkownikowi informacji. Przykładem może być przycisk przypisania obiektu cyfrowego do ulubionych elementów użytkownika. Jego kolor może reprezentować stan obiektu w tym zakresie, tzn. czy obiekt jest przypisany do ulubionych elementów użytkownika, czy też nie. W tego rodzaju przypadkach stosowane będą dodatkowe formy opisu tekstowego danego elementu (w opisanym przypadku przycisku).
6.	1.4.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by materiały audio udostępnianie w ramach projektu nie uruchamiały się automatycznie. Niezależnie od tego w przypadku odtwarzania materiałów audio dostępna będzie dla użytkownika możliwość wstrzymania odtwarzania oraz regulacji głośności niezależnie od głośności ustawionej na poziomie systemu operacyjnego urządzenia.
7.	1.4.3, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by wszelkie teksty oraz obrazy zawierające teksty zachowywały kontrast 4.5:1 w stosunku do ich tła.
8.	1.4.4,	Portale publiczne tworzone lub modyfikowane w ramach

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
	poziom AA	Projektu będą przygotowane w taki sposób, aby na każdej stronie było możliwe zwiększenie czcionki do 200% bez utraty prezentowanych treści czy dostępnej funkcjonalności.
9.	1.4.5, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby wszelkie informacje, które mogą być przekazane w postaci tekstu, były w ten sposób reprezentowane.
10.	1.4.6, poziom AAA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą spełniały normy dla osób niedowidzących. W szczególności przygotowany zostanie dedykowany widok wspierający osoby z problemami w rozpoznawaniu kolorów. Dotyczy to technicznych wymagań WCAG 2.0 w zakresie ogólnym¹¹, a w szczególności uwzględnienia zachowania kontrastu pomiędzy tekstem i jego tłem na poziomie 7:1. Wymaganie to wykracza poza minimalne wymagania KRI.
11.	2.1.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby umożliwiły operowanie stronami z wykorzystaniem samej klawiatury, bez konieczności korzystania z innych urządzeń sterujących (np. myszki).
12.	2.1.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby przechodzenie pomiędzy komponentami stron odbywało się z wykorzystaniem klawiszy strzałek lub klawisza tabulatora, bez konieczności wykorzystywania klawiszy modyfikujących (np. Ctrl). Stąd nie będzie konieczności wskazywania użytkownikowi innych niż standardowe sposoby na przejście do kolejnych komponentów strony. W szczególności chodzi o to, by użytkownik nie utknął na żadnym elemencie/komponencie strony ze względu na wykorzystywane urządzenia do sterowania

11 <http://www.w3.org/WAI/GL/WCAG20-TECHS/general.html>

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		stroną/komputerem.
13.	2.2.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, by sytuacje związane z ograniczeniem czasowym wykonania określonych czynności były adekwatnie obsługiwane. Ponadto, funkcje dostępne bez logowania w ramach rozwiązań technicznych będą przygotowane w taki sposób, by miały charakter bezstanowy, tzn. nie jest konieczne przechowywanie (np. w sesji użytkownika) informacji o jego bieżących działaniach na stronie, np. w przypadku wyników wyszukiwania linki do następnej lub poprzedniej strony wyników wyszukiwania nie będą bazować na sesji utrzymywanej po stronie serwera - przeciwnie - będą zawierać wszystkie niezbędne informacje do wygenerowania żądanej strony.
14.	2.2.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu nie będą zawierały mechanizmów automatycznego uruchamiania treści ruchomych, migoczących, przewijających się automatycznie lub aktualizujących się automatycznie.
15.	2.3.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu nie będą zawierały migoczących elementów.
16.	2.4.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby bloki treści, które pojawiają się na poszczególnych stronach w takiej samej formie (np. nagłówek, menu lub stopka) były tak osadzone na stronie, aby możliwe było ich pominięcie, np. poprzez zastosowanie hiperłącza na górze strony prowadzącego bezpośrednio do głównej treści strony (pomijając nagłówek i menu).
17.	2.4.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą tak przygotowane, aby każda strona internetowa posiadała opisujący ją tytuł. W przypadku stron WWW tytuł

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		będzie zamieszczony w odpowiednich znacznikach HTML (<title></title>).
18.	2.4.3, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą tak przygotowane, aby trawersowanie klawiaturą po poszczególnych elementach projektowanych stron internetowych było zgodne ze znaczeniem i powiązaniem logicznymi poszczególnych elementów. Przykładowo w przypadku obszaru logowania na stronie - najpierw uzyska aktywność pole tekstowe do wprowadzenia loginu użytkownika, a następnie do wprowadzenia hasła, a dalej przycisk "Zaloguj" (niezależnie od możliwości wybrania przycisku Enter w celu wykonania logowania np. na polu hasła).
19.	2.4.4, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą tak przygotowane, aby wszystkie hiperłącza ze strony internetowej były opatrzone odpowiednim tekstem - tak by z tekstowej reprezentacji hiperłącza użytkownik mógł zrozumieć cel zamieszczonego hiperłącza.
20.	2.4.5, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą umożliwiały co najmniej trzy sposoby na dotarcie do interesującej użytkownika informacji: <ol style="list-style-type: none"> 1. Wyszukiwanie ogólne oraz wyszukiwanie zaawansowane pozwolą na wprowadzenie frazy lub logicznie powiązanych ze sobą fraz w celu odnalezienia interesującej informacji. 2. Przeglądanie zawartości serwisu z wykorzystaniem mapy strony. 3. Przeglądanie indeksów wartości związanych z konkretnymi polami metadanych obiektów cyfrowych. Przykładowo dostępna będzie lista wszystkich autorów, którzy występują w metadanych obiektów cyfrowych. Wybór pozycji (osoby) z listy kierować będzie do listy

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		wszystkich obiektów, których autorem jest wskazana osoba. Podobne listy dostępne będą do innych elementów metadanych (np. słowa kluczowe, tytuł).
21.	2.4.6, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby wszystkie nagłówki oraz etykiety opisywały temat lub cel danego elementu, np. każda wiadomość na portalu będzie posiadała nagłówek opisujący skrótowo zawartość wiadomości, z kolei poszczególne pola formularzy (np. formularz wyszukiwawczy) będą opatrzone odpowiednimi etykietami.
22.	2.4.7, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby w przypadku trawersowania po stronie internetowej z wykorzystaniem klawiatury elementy aktywne były odpowiednio zaznaczone, np. w przypadku pola tekstowego pojawiający się w nim kursor oraz jego zmiana wizualna, a w przypadku przycisku stosowna zmiana wizualna tego przycisku.
23.	2.4.8, poziom AAA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby użytkownicy przeglądający zasoby cyfrowe mieli do dyspozycji informację o lokalizacji aktualnej strony w kontekście całego portalu. Będzie to zrealizowane z wykorzystaniem struktury portalu — użytkownik będzie miał do dyspozycji nawigację okruszkową (ang. breadcrumb trails lub breadcrumbms) w tym zakresie oraz hiperłącze do strony głównej z poziomu każdej strony portalu. Wymaganie to wykracza poza minimalne wymagania KRI.
24.	3.1.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby informacje o języku strony internetowej były przesyłane do użytkownika wraz z zawartością strony, np. poprzez wykorzystanie stosownego atrybutu znacznika HTML (np. atrybut lang dla

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		dokumentu HTML).
25.	3.1.2, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby w przypadku zmiany wykorzystywanego języka w obrębie strony internetowej stosowna informacja była umieszczona w ramach elementu ze zmienionym językiem, np. w atrybucie lang znacznika, w ramach którego występuje inny język.
26.	3.2.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby elementy na stronach WWW, które uzyskują aktywność (ang. focus), nie wprowadzały zmian w kontekście prezentowanej strony, np. w przypadku przycisku efekt będzie miało dopiero przyciśnięcie go (klawiszem Enter lub myszką), a nie samo przekazanie aktywności (ang. focus).
27.	3.2.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby zmiany właściwości poszczególnych elementów na stronach nie wprowadzały zmian w kontekście prezentowanej strony, np. wprowadzanie wartości do pola tekstowego nie będzie skutkowało zmianą zawartości strony, podobnie jak zaznaczenie pola wyboru czy wybranie jednej z dostępnych opcji na formularzu (np. wyszukiwania).
28.	3.2.3, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby elementy stron były ułożone względem siebie zawsze w takiej samej kolejności w zakresie nawigacji. W przypadku braku któregoś z elementów na specyficznej stronie, kolejność pozostałych zostanie zachowana.
29.	3.2.4, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby elementy stron, które mają tę samą funkcję, były

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
		nazywane/identyfikowane spójnie/jednolicie. Przykładowo linki do kolejnej/poprzedniej strony wyników wyszukiwania będą jednolicie nazywane na każdej stronie wyników niezależnie od tego, czy wyszukiwanie było uruchomione z poziomu prostego formularza, czy zaawansowanego.
30.	3.3.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby w przypadku wykrycia błędu w danych wprowadzanych przez użytkownika przedstawiony został tekstowy komunikat z wyjaśnieniem zaistniałego błędu, np. w przypadku niewypełnienia obowiązkowych pól formularza kontaktu.
31.	3.3.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby w przypadku wprowadzania przez użytkownika danych, każde pole było opatrzone stosowną etykietą (wyjaśniającą cel pola) oraz jeśli to potrzebne (np. w przypadku wprowadzania dat) instrukcją związaną ze sposobem wprowadzania określonego typu informacji. Przykładem takiej informacji jest data lub hiperłącze do listy dopuszczalnych wartości w ramach określonego elementu formularza wyszukiwawczego.
32.	3.3.3, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby w przypadku błędnie wprowadzonych informacji przez użytkownika, o ile to możliwe, przygotowane były sugestie w zakresie poprawnego wprowadzenia tych informacji. Przykładowo w przypadku błędnie wprowadzonej wartości w komponencie daty przekazany zostanie komunikat o możliwych wartościach, jakie mogą zostać umieszczone w tym polu.
33.	3.3.4, poziom AA	Portale publiczne tworzone lub modyfikowane w ramach projektu nie będą użytkownikom końcowym umożliwiały wykonywania operacji o nieodwracalnych skutkach.

Lp.	Pozycja WCAG 2.0 i poziom	Planowany sposób spełnienia wymagania w projekcie
34.	4.1.1, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą zgodne z odpowiednimi standardami. W szczególności w celu zapewnienia tej zgodności przeprowadzane będą testy walidacji poszczególnych stron portali tworzonych lub modyfikowanych w ramach Projektu w odniesieniu do zastosowanych na danym portalu technik. Przykładowo strony HTML będą walidowane z wykorzystaniem usługi W3C ¹² .
35.	4.1.2, poziom A	Portale publiczne tworzone lub modyfikowane w ramach projektu będą przygotowane w taki sposób, aby wszystkie dedykowane komponenty portali, wykraczające poza standardowe komponenty/kontrolki HTML, były opracowane w taki sposób, aby technologie asystujące osobom niepełnosprawnym mogły z nich w pełni korzystać, np. jest w takim przypadku możliwość sprawdzenia nazwy czy roli/typu komponentu z poziomu oprogramowania, jak również jest możliwość ustawienia właściwości/wartości komponentu przez zewnętrzne oprogramowanie wraz z możliwością podłączenia się oprogramowania pod generowane przez komponent zdarzenia.

3.9. Monitoring dostępności

W celu zapewnienia wysokiej dostępności serwisów, które zostaną uruchomione lub zmodernizowane w ramach projektu, planuje się objąć je automatycznym mechanizmem monitorowania i restartu. Oprogramowanie tego typu (np. system Nagios,) zostanie wdrożone w infrastrukturze informatycznej Biblioteki Śląskiej przez pracowników Działu Informatycznego BŚ, a następnie, we współpracy z dostawcą sprzętu i oprogramowania zostaną opracowane odpowiednie zautomatyzowane testy i procedury awaryjne. W szczególności dostawca oprogramowania będzie odpowiedzialny za przygotowanie specyfikacji testów i (tam, gdzie to będzie możliwe) skryptów automatycznego restartu poszczególnych usług po awarii. Dzięki tym mechanizmom oraz zvirtualizowanej infrastrukturze sprzętowej zakłada się osiągnięcie poziomu dostępności nie gorszego niż 99% w skali roku (z wyłączeniem planowanych przerw serwisowych).

12 <http://validator.w3.org/>

3.10. Monitoring wykorzystania

Wdrożenie monitoringu wykorzystania udostępnianych w ramach projektu zasobów oraz serwisów będzie oparte o działania dwupoziomowe. Z jednej strony każdy z serwisów zostanie objęty monitoringiem narzędzia do analityki webowej (np. Google Analytics), które pozwoli na szczegółowe analizy zachowań użytkowników serwisu, na podstawie pewnej próbki danych. Z drugiej strony wykorzystywane będą narzędzia do analizy logów serwerów WWW (takie jak np. Webalizer czy AWStats) oraz narzędzia statystyczne wbudowane w zamawiane oprogramowanie. Dzięki temu możliwe będzie dokładne określanie faktycznego ruchu jaki jest w poszczególnych serwisach¹³ i rzeczywistej liczby pobrań/odtworzeń udostępnianych obiektów, co jest jednym z podstawowych wskaźników rezultatów projektu.

3.11. Skalowalność

Na potrzeby nowopowstającego agregatora regionalnego, ze względu na jego potencjalnie większą popularność, zakłada się docelowe obciążenie serwisu głównego na poziomie do 30 mln żądań miesięcznie w sytuacji zwiększonego ruchu. Jeżeli chodzi o serwisy powiązane, przyjmuje się, że ze względu na ich wyspecjalizowany charakter, przewidywany ruch maksymalny będzie na poziomie:

- Dla serwisu ścieżek edukacyjnych: 15% obciążenia serwisu głównego – do 4,5 mln żądań miesięcznie.
- Dla serwisu wolumenów czasoprzestrzennych i serwisu do pracy badawczej z dokumentami historycznymi: 5% obciążenia serwisu głównego – do 1,5 mln żądań miesięcznie dla każdego z tych dwóch serwisów.

Średni czas odpowiedzi serwera na żądania użytkowników powinien pozostać nie gorszy niż 2 s.

Opisane powyżej wymagania wydajnościowe zostaną postawione przed dostawcą oprogramowania dla w/w serwisów.

3.12. Bezpieczeństwo

Wszystkie systemy teleinformatyczne wdrożone lub wykorzystywane w projekcie będą zapewniały należyte bezpieczeństwo przetwarzania danych. W dalszej części opisano poszczególne aspekty bezpieczeństwa oprogramowania w projekcie.

Klasyfikacja danych przetwarzanych w projekcie

13 Ze względu na różne metody pomiaru narzędzia do analityki webowej i do analizy logów dają zazwyczaj mocno odmienne wartości. Od kilku do kilkudziesięciu razy mniejsze mogą być wartości pokazywane przez narzędzia do analityki webowej (<https://www.devonwebdesigners.com/2567/awstats-vs-webalizer-vs-google-analytics/>).

Projekt operuje na danych przeznaczonych do udostępnienia szerokiemu gronu użytkowników. Będzie to informacja o charakterze publicznym (zasoby regionalnego agregatora będą udostępnione bez zalogowania, w większości jako zbiory z domeny publicznej, nieobjęte ochroną prawnoautorską).

Opracowany system musi jednak przechowywać informacje o pewnych grupach użytkowników, w tym dane osobowe (zwykłe), takie jak np. imię i nazwisko czy miejsce zatrudnienia. Przechowanie tych danych jest niezbędne na potrzeby zarządzania systemem, np. dla potrzeb uwierzytelnienia w interfejsie zarządzającym. System nie będzie przetwarzać wrażliwych danych osobowych ani informacji niejawnych w rozumieniu Ustawy o ochronie informacji niejawnych z 5 sierpnia 2010 r. Zapewnienie bezpieczeństwa danych w oczywisty sposób nie może się opierać jedynie na klasyfikacji ich wrażliwości oraz zastosowaniu wymaganych stosownymi przepisami środków ochrony komponentów systemu, które dane te przechowują oraz przetwarzają. Należy w odpowiedni sposób zabezpieczyć wszystkie moduły docelowego systemu, aby w wyniku przejęcia źle chronionego, ale potencjalnie zaufanego dla systemu, elementu nie nastąpiła eskalacja ataku prowadząca do wycieku danych czy przejęcia uprzywilejowanego konta w systemie.

Mechanizmy uwierzytelniania

Zasoby projektu będą dostępne bez potrzeby uwierzytelnienia w systemie, ale zarządzanie systemem i zamieszczanie zasobów wymagać będzie utworzenia mechanizmów zarządzania użytkownikami. W ramach projektu planuje się wykorzystanie uwierzytelniania przy pomocy pary: nazwa użytkownika oraz hasło. Na etapie rozwoju projektu zakładanym przez Projekt korzystanie z wielu różnych mechanizmów uwierzytelniania prowadziłyby do zbyt wysokiego stopnia komplikacji – czego w przypadku systemów zabezpieczeń należy unikać.

Mechanizmy uwierzytelniania zostaną zaprojektowane i zaimplementowane w sposób maksymalnie utrudniający atak na nie. Należy zwrócić uwagę na fakt, że liczba kont użytkowników wewnętrznych będzie względnie niewielka – ale za to konta te będą w systemie uprzywilejowane. Należy więc je szczególnie chronić, za to można pozwolić sobie na względnie zauważalny koszt obliczeniowy na tę ochronę poświęcony.

Zostaną zastosowane rekomendowane mechanizmy zabezpieczenia informacji uwierzytelniającej (por. np.), przede wszystkim:

- Baza danych nie będzie przechowywać haseł otwartym tekstem, ale ich skróty kryptograficzne.
- Skróty kryptograficzne będą generowane przy pomocy rekomendowanego algorytmu SHA-256.
- Zostanie wykorzystany mechanizm tzw. soli, dodatkowo utrudniające ataki siłowe w przypadku kradzieży skrótów haseł.
- Użyty będzie mechanizm tzw. *key stretching*, wymuszający wielokrotne liczenie skrótów haseł – spowalniający ataki siłowe.

- Funkcjonalność systemu uwzględni:
 - Rozsądne polityki tworzenia haseł użytkowników, niepozwalające na stosowanie słabych, łatwych do odgadnięcia lub złamania przy pomocy ataku siłowego, haseł.
 - Odpowiednie mechanizmy przypomnienia hasła (bez stosowania tzw. pytań pomocniczych, oparte na wysłaniu użytkownikowi na zarejestrowaną w systemie skrzynkę poczty elektronicznej ważnego przez krótki czas linku prowadzącego na odpowiednią podstronę systemu i umożliwiającego jednorazową zmianę hasła).

Zapewnienie ochrony danych osobowych

Dane osobowe są szczególnym przypadkiem danych, które chronić należy w sposób wyjątkowo drobiazgowy. Konieczność ich ochrony wynika między innymi z wymogów zapewnienia zgodności z polskimi, unijnymi oraz międzynarodowymi przepisami nakazującymi każdemu administratorowi danych osobowych konieczność ich szczególnego zabezpieczenia.

Dla każdego ze zbiorów danych osobowych przechowywanych w systemach realizowanych w projekcie, we współpracy z dostawcą oprogramowania utworzony zostanie zestaw procedur wpisujących się w ogólną politykę bezpieczeństwa ICT, który wraz z instrukcjami zarządzania oraz obsługi systemu stanowić będzie zabezpieczenie dla danych na poziomie organizacyjnym. Podczas tworzenia polityki bezpieczeństwa pod uwagę zostaną wzięte również wszelkie zabezpieczenia na poziomie technicznym, które mogą znaleźć zastosowanie w ochronie zbiorów danych osobowych.

W każdym przypadku podejście takie zgodne jest z zapisami zawartymi w Ustawie z dnia 29 sierpnia 1997 r. o ochronie danych osobowych, Rozporządzeniem Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony danych osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE nazywanym ogólnym rozporządzeniem o ochronie danych osobowych oraz wytycznymi GIODO.

Bezpieczeństwo w fazie utrzymania

Bezpieczeństwo nie jest raz osiągniętym stanem, lecz procesem¹⁴. Ocena bezpieczeństwa przygotowana na podstawie zrealizowanych i zaraportowanych prac jest ważna jedynie na moment jej wydania, przede wszystkim z dwóch następujących przyczyn:

- Zmian w stanie wiedzy w zakresie bezpieczeństwa IT (np. odkrywania nowych podatności i błędów bezpieczeństwa, opracowania nowych technik ataków).
- Zmian w analizowanym środowisku (np. dodawania funkcjonalności do aplikacji – w okresie utrzymania projektu mało prawdopodobnej, ale również migracji na nowsze

14 <https://www.schneier.com/crypto-gram/archives/2000/0515.html>

wersje oprogramowania, rekonfiguracji środowiska pod kątem zmieniających się uwarunkowań, dodawania zasobów albo użytkowników zarządzających).

Stąd dla zapewnienia stosownego poziomu zabezpieczeń utworzonego systemu konieczne będzie wdrożenie dodatkowych działań, wyszczególnionych poniżej, w szczególności w okresie utrzymania wyników projektu. W kontekście oprogramowania będącego przedmiotem tego opracowania zakłada się zobowiązanie dostawcy oprogramowania, aby w okresie trwałości projektu (5 lat) zapewniał gwarancję na oprogramowanie, obejmującą również usuwanie powstałych faktycznych i potencjalnych luk bezpieczeństwa w dostarczonym oprogramowaniu. Inne podejmowane działania mające na celu utrzymanie bezpieczeństwa systemu będą wynikały bezpośrednio z polityk bezpieczeństwa ICT funkcjonujących w danym okresie w Bibliotece Śląskiej.

Audytowalność

Poza opisanym wcześniej monitorowaniem wykorzystania udostępnianych zasobów, planowane w ramach projektu oprogramowanie będzie wyposażone również w mechanizmy logowania wystąpienia czynności dodawania, modyfikacji i usuwania danych, wykonywane przez użytkowników wewnętrznych. Dzięki temu, w razie wystąpienia niepożądanego zdarzenia zmiany lub utraty danych, bądź też np. opublikowania on-line danych objętych prawem autorskim, możliwe będzie określenie, kto jest odpowiedzialny za powstałą niepożądaną sytuację i zapobieżenie powtórzeniu jej wystąpienia w przeszłości.

3.13. Kopie zapasowe i zdolność do odtworzenia po awarii

W ramach wdrażania oprogramowania opracowane zostaną również procedury wykonywania kopii zapasowych i odtwarzania danych i serwisu z kopii po wystąpieniu awarii. W tym zakresie na dostawcę oprogramowania nałożone zostanie wymaganie, aby system zrealizowany był w sposób umożliwiający wykonywanie kopii bezpieczeństwa danych bez blokowania dostępu do usług publicznych (w szczególności bez konieczności wyłączania ich). Ponadto dostawca będzie zobowiązany do przygotowania zakresu danych i elementów oprogramowania, które powinny podlegać wykonywaniu kopii bezpieczeństwa oraz procedury odtwarzania systemu w oparciu o tak zabezpieczone dane i elementy. Samo zaimplementowanie i monitorowanie funkcjonowania procedur zautomatyzowanego wykonywania kopii bezpieczeństwa będzie zrealizowane przez pracowników Działu Informatycznego Biblioteki Śląskiej.

4. Opis niezbędnej infrastruktury sprzętowej

4.1. Wstęp

Na potrzeby określenia niezbędnej infrastruktury sprzętowej potrzebnej do uruchomienia e-usług realizowanych w ramach projektu niezbędne będzie przygotowanie odpowiedniej infrastruktury sprzętowej. Zakłada się, że środowisko to powinno być wysoce niezawodne i realizowane w oparciu o mechanizmy wirtualizacji. Szczegółowy opis planowanej infrastruktury przedstawiony jest w osobnym opracowaniu, poniżej natomiast określono

minimalne wymagania z punktu widzenia opisywanego w niniejszym dokumencie planowanego oprogramowania.

4.2. Serwery usługowe

Na potrzeby ułatwionego wyspecyfikowania skali niezbędnej infrastruktury przyjęto, że powinna ona być oparta o dwa podstawowe rodzaje serwerów, opisane poniżej:

- „Serwer L” o silniejszych parametrach, stosowany tam, gdzie konieczna może być duża liczba równoległych żądań i są to np. żądania od użytkowników końcowych, które nie mogą być kolejgowane do asynchronicznego przetworzenia z opóźnieniem. Proponowane podstawowe parametry minimalne tego serwera to:
 - Liczba rdzeni procesora: 16 z taktowaniem 3.3 GHz
 - RAM: 32 GB
 - Dysk twardy: 512 GB SSD
- „Serwer S” o słabszych parametrach, stosowany tam, gdzie nie jest spodziewany bardzo duży ruch użytkowników końcowych lub gdzie przetwarzanie może być kolejgowane i realizowane asynchronicznie z opóźnieniem. Proponowane podstawowe parametry minimalne tego serwera to:
 - Liczba rdzeni procesora: 8 z taktowaniem 3.3 GHz
 - RAM: 16 GB
 - Dysk twardy: 270 GB SSD

Obydwa powyższe rodzaje serwerów mogą być uruchamiane w ramach infrastruktury zwirtualizowanej.

Tabele poniżej zawierają zestawienie liczby serwerów poszczególnego typu wraz z rozłożeniem komponentów projektowanych e-usług na te serwery:

Tabela 7. Zestawienie serwerów na potrzeby regionalnego agregatora.

Serwer	Komponenty
RA.S.1	Aplikacja WWW serwisu wolumenów czasoprzestrzennych wraz z bazą danych
RA.S.2	Aplikacja WWW serwisu do pracy badawczej z dokumentami historycznymi
RA.S.3	Baza danych na potrzeby aplikacji WWW serwisu do pracy badawczej z dokumentami historycznymi
RA.S.4	Narzędzia przetwarzania treści na potrzeby serwisu do pracy badawczej z dokumentami historycznymi
RA.S.5	Aplikacja WWW serwisu ścieżek edukacyjnych wraz z bazą danych

Serwer	Komponenty
RA.L.1	Aplikacja WWW portalu agregatora wraz z bazą danych i serwerem wyszukiwawczym
RA.L.2	Serwer indeksacji danych
RA.S.6	Moduł agregacji metadanych
RA.S.7	Moduł agregacji treści
RA.S.8	Moduł przetwarzania danych
RA.S.9	Serwer agregatora
RA.S.10	Aplikacja WWW dla administratorów agregatora

Podsumowanie zapotrzebowania na zwirtualizowaną infrastrukturę sprzętową zebrano w tabeli poniżej.

Tabela 8. Zestawienie zapotrzebowania na zwirtualizowaną infrastrukturę sprzętową.

Nazwa serwisu	Liczba sztuk serwerów	Liczba rdzeni procesora (3.3GHz) [szt.]	Pamięć RAM [GB]	Pojemność dysków SSD [GB]
Regionalny agregator		112	224	3 724
<i>Serwery S</i>	10	80	160	2 700
<i>Serwery L</i>	2	32	64	1 024

4.3. Składowanie danych

W przypadku agregatora metadanych bierze się pod uwagę zastosowanie bazy danych typu NoSQL lub systemu plików (w zależności od ostatecznego rozwiązania zaproponowanego przez wykonawcę serwisu agregatora). Na potrzeby składowania danych w serwisie agregatora zakłada się w związku z tym konieczność zapewnienia 6 maszyn fizycznych o następujących podstawowych parametrach minimalnych:

- Liczba rdzeni procesora: 16 z taktowaniem 3.3 GHz
- RAM: 64 GB
- Dyski twarde: HDD 2x 1TB oraz SSD 512 GB

Harmonogram prac

Zamawiający wymaga ponownego ustalenia harmonogramu prac dla zadania, ponieważ planowany bieg projektu uległ zmianie.

Zakłada się, że zadania będą realizowane w trzech równoległych wątkach opisanych poniżej. Pierwszy wątek dotyczy modernizacji ŚBC, a dwa kolejne budowy regionalnego agregatora. Każdy z tych wątków może być realizowany do pewnego stopnia niezależnie, ale występują też pomiędzy nimi punkty synchronizacji oznaczone jako kamienie milowe. Są one również widoczne na Rysunek 2. oraz na diagramie Gantta widocznym na Rysunek 4 poniżej tabeli. Diagram Gantta kończy się na 42 miesiącu projektu, jeden miesiąc po zakończeniu prac opisywanych w niniejszym dokumencie. Cały projekt realizowany będzie przez okres 48 miesięcy.

Analiza zależności pomiędzy zadaniami pokazują, że kluczowym serwisem dla osiągnięcia rezultatów projektu jest serwis wolumenów czasoprzestrzennych. Od niego zależą bowiem pozostałe serwisy: zmodernizowana ŚBC oraz agregator regionalny, a konkretnie jego serwisy składowe – do badań z dokumentami historycznymi. W związku z tym realizacja tej części projektu będzie objęta szczególnym monitorowaniem postępów i ryzyka. Ponadto, w razie wystąpienia opóźnień, przy realizacji tego komponentu nie powinno wystąpić zagrożenie w realizacji wskaźników rezultatu projektu związanych z ilością udostępnianych zasobów informacji publicznej ani ich pobrań czy odtworzeni. Wynika to z faktu iż serwis wolumenów czasoprzestrzennych jest niezbędny do zrealizowania usług dodanych w stosunku do treści cyfrowych. W przypadku ŚBC jest to wzbogacanie metadanych, a w przypadku serwisu do badań jest to adnotowanie treści dokumentów historycznych. Poza tym nawet mało prawdopodobne kilkumiesięczne opóźnienie w osiągnięciu kamienia milowego K2.1 nie spowoduje, że zależne od niego podzadania przesuną się poza harmonogram realizacji całego projektu

Tabela 9. Harmonogram realizacji prac informatycznych związanych z uruchomieniem e-usług w projekcie.

Początek realizacji	Koniec realizacji	Opis działania	Powiązany produkt ¹⁵
---------------------	-------------------	----------------	---------------------------------

T1. Modernizacja Śląskiej Biblioteki Cyfrowej

M10	M14	T1.1. Prace projektowo-implémentacyjne związane z podsystemem obsługi ścieżki digitalizacyjnej i długoterminowej archiwizacji.	Produkt 1.1
M15	M15	T1.2. Pilotażowe wdrożenie i konfiguracja podsystemu obsługi ścieżki digitalizacyjnej i długoterminowej archiwizacji w BŚ.	
M16	M24	T1.3. Prace projektowo-implémentacyjne nad zmodernizowanym portalem Śląskiej Biblioteki Cyfrowej (Produkt 1.3) – Etap 1.	Produkt 1.2 Produkt 1.3
M25	M26	T1.4. Produkcyjne wdrożenie pierwszej wersji zmodernizowanego portalu ŚBC.	Produkt 1.4 Produkt 1.5 Produkt 1.6
M27	M29	T1.5. Prace projektowo-implémentacyjne nad zmodernizowanym portalem Śląskiej Biblioteki Cyfrowej – Etap 2. <u>Zależność:</u> Rozpoczęcie prac na tym etapie wymaga realizacji kamienia milowego K2.1 (patrz zadanie T2.2 poniżej)	Produkt 1.2 Produkt 1.3

15 Por. Rozdział 1. Zakres prac i produkty wynikowe.

Początek realizacji	Koniec realizacji	Opis działania	Powiązany produkt
M30	M30	<p>T1.6. Produkcyjne wdrożenie drugiej wersji zmodernizowanego portalu ŚBC. Produkcyjne uruchomienie całości zmodernizowanego systemu ŚBC.</p> <p><u>Kamień milowy K1.1:</u> Zmodernizowany serwis ŚBC jest dostępny produkcyjnie do wykorzystania przez inne systemy.</p>	<p>Produkt 1.4</p> <p>Produkt 1.5</p> <p>Produkt 1.6</p>
<p>T2. Regionalny agregator - Uruchomienie serwisu wolumenów czasoprzestrzennych i serwisu do pracy badawczej z dokumentami historycznymi</p>			
M10	M24	<p>T2.1. Prace analityczne, projektowe, implementacyjne związane z opracowaniem serwisu wolumenów czasoprzestrzennych</p>	Produkt 2.4
M25	M26	<p>T2.2. Produkcyjne wdrożenie serwisu wolumenów czasoprzestrzennych.</p> <p><u>Kamień milowy K2.1:</u> Serwis wolumenów czasoprzestrzennych jest dostępny produkcyjnie do wykorzystania przez inne systemy.</p>	Produkt 2.4.1
M27	M39	<p>T2.3. Prace analityczne, projektowe, implementacyjne związane z opracowaniem serwisu do pracy badawczej z dokumentami historycznymi</p>	Produkt 2.5
M40	M41	<p>T2.4. Produkcyjne wdrożenie serwisu do pracy badawczej z dokumentami historycznymi, integracja z portalem agregatora regionalnego.</p> <p><u>Zależność:</u> Rozpoczęcie prac na tym etapie wymaga realizacji kamienia milowego K3.1 (patrz zadanie T3.4 poniżej)</p>	Produkt 2.7

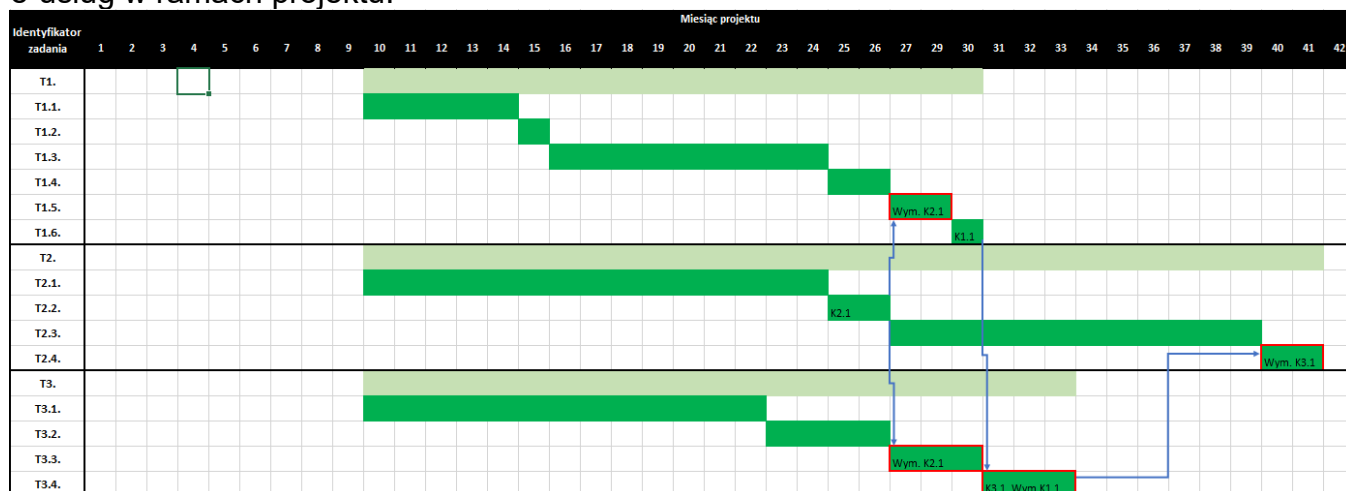
Początek realizacji	Koniec realizacji	Opis działania	Powiązany produkt
---------------------	-------------------	----------------	-------------------

T3. Regionalny agregator - Uruchomienie serwisu WWW agregatora regionalnego wraz z serwisem ścieżek edukacyjnych

M10	M22	T3.1. Prace analityczne, projektowe, implementacyjne związane z opracowaniem modułu do agregacji danych i serwisu agregatora regionalnego – Etap 1 (agregacja metadanych, portal).	Produkt 2.1 Produkt 2.2
M23	M26	T3.2. Produkcyjne wdrożenie agregatora regionalnego, rozpoczęcie napełniania danymi	Produkt 2.3
M27	M30	T3.3. Prace analityczne, projektowe, implementacyjne związane z opracowaniem modułu do agregacji danych i serwisu agregatora regionalnego – Etap 2 (agregacja treści, serwis ścieżek edukacyjnych). <u>Zależność:</u> Rozpoczęcie prac na tym etapie wymaga realizacji kamienia milowego K2.1 (patrz zadanie T2.2 powyżej)	Produkt 2.1 Produkt 2.2 Produkt 2.6
M31	M33	T3.4. Produkcyjne wdrożenie agregatora regionalnego wraz z modułem agregacji treści i serwisem ścieżek edukacyjnych. <u>Kamień milowy K3.1:</u> Serwis agregatora regionalnego jest dostępny produkcyjnie do wykorzystania przez inne systemy. <u>Zależność:</u> Rozpoczęcie prac na tym etapie wymaga realizacji kamienia milowego K1.1 (patrz zadanie T1.6 powyżej)	Produkt 2.7 Produkt 2.8 Produkt 2.9



Rysunek 4. Wykres Gantta dla zadań związanych z wdrażaniem oprogramowania e-usług w ramach projektu.



Na wykresie Gantta niebieskimi strzałkami zaznaczono zależności pomiędzy zadaniami, wpisano również kamienie milowe z tabeli z harmonogramem znajdującej się powyżej. Nazwa kamienia milowego wpisane w zadanie oznacza, że pozytywne zakończenie zadania jest równoznaczne z osiągnięciem kamienia milowego. Nazwa kamienia milowego poprzedzona określeniem „Wym.,” oznacza, że do wykonania zadania jest niezbędne osiągnięcie danego kamienia milowego.