

Opis predmetu zákazky

Projekt:

Migrácia IS Digitálny archív ŠGÚDŠ do cloudu ako súčasť elektronických služieb štátu

Kód projektu v ITMS2014+	:	311071CFN3
Výzva - kód Výzvy	:	OPII-2021/7/13-DOP
Operačný program:	:	Integrovaná infraštruktúra
Spolufinancovaný fondom:	:	Európsky fond regionálneho rozvoja
Prioritná os:	:	7 Informačná spoločnosť

1 Obsah

1	Obsah	2
2	Zoznam pojmov a skratiek	3
3	Východiskové informácie	6
4	Popis súčasného stavu	7
4.1	Procesné riešenie skenovacej linky ŠGÚDŠ.....	13
4.2	Aplikačná architektúra	15
4.3	Technologická architektúra.....	18
5	Súpis požiadaviek	19
5.1	Legislatívne požiadavky.....	19
5.2	Požiadavky na spracovanie analýzy a návrhu riešenia projektu	19
5.3	Funkčné požiadavky	20
5.3.1	Základné požiadavky	20
5.3.2	Skenovacia linka a workflow spracovania dát	21
5.3.3	Vyhľadávacie služby	23
5.3.4	Sťahovacie služby	25
5.3.5	Ukladacie služby.....	26
5.3.6	Reportovacie služby	27
5.4	Požiadavky na integrácie.....	28
5.5	Požiadavky na výkonnostné parametre, kapacitné požiadavky	30
5.6	Požiadavky na aplikačnú architektúru	31
5.7	Požiadavky na technologickú architektúru	32
5.8	Požiadavky na využitie služieb cloudu	33
5.9	Požiadavky na bezpečnosť	35
5.10	Požiadavky na zdrojový kód	36
5.11	Požiadavky na testovanie.....	37
5.12	Všeobecné požiadavky.....	37

2 Zoznam pojmov a skratiek

Skratka / pojem	Vysvetlenie
AD	Active directory
ALTO	Technické metadáta pre optické rozpoznávanie znakov (OCR)
API	Application programming interface (rozhranie pre programovanie aplikácií) - Ide o zbierku funkcií a tried (ale aj iných programov), ktoré určujú akým spôsobom sa majú funkcie knižníc volať zo zdrojového kódu programu.
CDA	Centrálny dátový archív
DA / IS DA	Digitálny archív geologickej knižnice a geofondu
DDoS	Distributed Denial of Service - Distribuovaný útok zahľtením servera služby sa zameriava na webové lokality a servery a narúša sieťové služby. Útok DDoS sa snaží vyčerpať zdroje aplikácií. Páchatelia týchto útokov zahltia webovú lokalitu pochybnou aktivitou, čím spôsobia nízku funkčnosť lokality alebo jej úplné vyradenie z prevádzky.
DNR	Detailný návrh riešenia
EÚ	Európska únia
GDPR	General Data Protection Regulation - Všeobecné nariadenie o ochrane údajov - plným názvom Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov
GIS	Geografický Informačný Systém
HTTP	HyperText Transfer Protocol - protokol na prenos html dokumentov medzi servermi a klientmi služby
HTTPS	HyperText Transfer Protocol Secure - je zabezpečená verzia komunikačného protokolu HTTP
HW	Hardware
ICM	IBM Content Manager
ICN	IBM Content Navigator
IS	Informačný systém
IS CSRÚ	Informačný systém centrálnej správy referenčných údajov
JPEG2000	Formát súboru JPEG 2000 je štandard pre grafický formát rastrovej grafiky s kompresiou dát. Pri kompresii sa využíva diskretná waveletová transformácia, ktorá ponúka lepšie výsledky oproti staršiemu štandardu JPEG, ktorý používa diskretnú kosínusovú transformáciu.
MetaIS	Centrálny metainformačný systém verejnej správy SR
MARC 21	Reprezentácia a komunikácia popisných metadát o položkách knižnice
METS	Metadata Encoding and Transmission Standard – štandard pre programovanie popisných, administratívnych a štrukturálnych metadát o objektoch
MIX	Schéma XML na kódovanie prvkov technických údajov potrebných na správu zbierok digitálnych obrázkov
MODS	Metadata Object Description Standard (štandardný popis metadátových objektov) - Označenie XML pre metadáta z existujúcich záznamov MARC 21 a pôvodný popis zdroja
MŽP	Ministerstvo životného prostredia SR
NKIVS	Národná koncepcia informatizácie verejnej správy
OAI Cat	Open Source Software, ktorý podporuje štandard OAI-PMH

Skratka / pojem	Vysvetlenie
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting – protokol umožňujúci zber metaúdajových záznamov z digitálnych repozitárov. Protokol využíva bežné štandardy používané na webe - HTTP a XML.
OCR	Optické rozoznávanie znakov alebo OCR (z angl. Optical Character Recognition) je metóda umožňujúca preklad obrazu (grafiky) tlačенých alebo písaných znakov do textovej, editovateľnej formy napr. do ASCII znakov abecedy.
OPIS	Operačný program informatizácia spoločnosti
PDF	Portable Document Format - je súborový formát, ktorý vyvinula v roku 1993 spoločnosť Adobe Systems. Používa sa na ukladanie dokumentov nezávisle od softvéru, hardvéru a operačného systému na ktorom boli vytvorené a taktiež na zariadení, na ktorom sú zobrazované.
PSP balík	Producer Submission Package - balíček dát a metadát, ktorý prichádza od producenta dát
PREMIS	Preservation Metadata - dátový slovník a podporné XML schémy pre základné metaúdaje ochrany potrebné na podporu dlhodobého uchovávaní digitálnych materiálov.
REST	Representational state transfer (REST) je softvérový architektonický štýl, ktorý popisuje jednotné rozhranie medzi fyzicky oddelenými komponentmi, často cez internet v architektúre klient-server.
SaaS	Software as a Service (Softvér ako služba) je model nasadenia softvéru, kedy dochádza k hostingu aplikácie prevádzkovateľom služby. Služba je ďalej ponúkaná zákazníkom cez Internet.
SAŽP	Slovenská agentúra životného prostredia
SIP balík	Submission Information Package - balíček dát a metadát, ktorý je vstupom pre preces archivácie v CDA
S-JTSK	Súradnicový systém
SOAP	Simple Object Access Protocol - je protokolom na výmenu správ založených na XML prostredníctvom siete, hlavne pomocou HTTP. Formát SOAP tvorí základnú vrstvu komunikácie medzi webovými službami a poskytuje prostredie na tvorbu zložitejšej komunikácie.
SOAP	Simple Object Access Protocol - je protokolom na výmenu správ založených na XML prostredníctvom siete, hlavne pomocou HTTP
SR	Slovenská republika
SSL	Secure Sockets Layer - je štandardný protokol, ktorý sa používa na bezpečný prenos dokumentov v sieti. Technológia SSL vytvára bezpečné spojenie medzi webovým serverom a prehliadačom, aby sa zabezpečil súkromný a integrovaný prenos údajov.
ŠGÚDŠ	Štátny geologický ústav Dionýza Štúra
TAR	Tape ARchive - je súborový formát komprimovaného súboru a zároveň názov programu ktorý takýto kompresiu umožňuje.
VM	Virtual Machine – virtuálny stroj
WFS	Web Feature Service – ukladacia služba - Služba pracuje na princípe klient-server a umožňuje zdieľanie priestorových údajov vo forme vektorových údajov v prostredí Internetu. Výsledkom požiadavky klienta na WFS server sú údaje vo vektorovej podobe kódované v jazyku GML (XML).

Skratka / pojem	Vysvetlenie
WGS84	Súradnicový systém
WMS	Web Map Service - Služba pracuje na princípe klient-server a umožňuje zdieľanie priestorových údajov vo forme rastrových máp v prostredí Internetu. Výsledkom požiadavky klienta (tenkého alebo silného klienta) na WMS server sú rastrové údaje (obrázok) najčastejšie vo formáte JPEG, PNG, TIFF, a iné
XML	Extensible Markup Language – jazyk využívajúci na svoj zápis značky, používa hierarchickú štruktúru pre opis dát, umožňuje prehľadne oddeliť obsah a vzhľad

3 Východiskové informácie

Predmet projektu:

ŠGÚDŠ prevádzkuje od roku 2014 informačný systém Digitálneho archívu geofondu a Ústrednej geologickej knižnice (isvs_11180). Projekt implementácie Digitálneho archívu bol realizovaný a financovaný v rámci Operačného programu informatizácie spoločnosti (OPIS) s cieľom skvalitnenia a dobudovania systému digitalizácie kultúrneho, vedeckého a intelektuálneho dedičstva a sprístupnenia digitálneho obsahu Geofondu a Ústrednej geologickej knižnice Slovenskej republiky. Systém je aktuálne prevádzkovaný na hardvérovej a softvérovej virtualizovanej infraštruktúre umiestnenej v prostredí ŠGÚDŠ, ktorá bola súčasťou celkového dodávaného riešenia.

Predmetom tohto projektu je migrácia Digitálneho archívu geofondu a Ústrednej geologickej knižnice do cloud prostredia vrátane všetkých údajov a prevádzka riešenia na obdobie 24 mesiacov.

Identifikácia projektu:

projekt_1683 <https://data.gov.sk/id/egov/project/1683>

4 Popis súčasného stavu

Digitálny archív Štátneho geologického ústavu Dionýza Štúra (ďalej ŠGÚDŠ) isvs_11180 je jedným z kľúčových informačných systémov ŠGÚDŠ. Je to komplexné softvérové riešenie problematiky správy dokumentácie Štátneho geologického archívu SR a Ústrednej geologickej knižnice, jej archivácie a publikácie rôznym skupinám používateľov v rámci organizácie, ako aj širokej odbornej a laickej verejnosti.

Digitálny archív umožňuje vkladanie veľkého spektra formátov dokumentov do digitálneho archívu systémom „drag and drop“. Každý dokument je pri vkladaní do archívu indexovaný pre rýchle vyhľadanie. Indexovanie prebieha nie len nad popisnými atribútmi dokumentu (metaúdajmi), ale aj na základe hĺbkovej analýzy obsahu dokumentu. Digitálny archív tak pomáha rýchlo a jednoducho vyhľadať požadovanú dokumentáciu, a to formou predpripravených vyhľadávacích formulárov, alebo voľným fulltext vyhľadaním slova, resp. frázy. Digitálny archív prispôsobuje výsledky vyhľadávania usporiadávaním a uprednostňovaním zoznamu vyhľadaných reťazcov na základe užívateľských zvyklostí. Rozumie štylistike slovenského jazyka vrátane skloňovania, vďaka čomu sa dosahujú omnoho lepšie výsledky vyhľadávania.

Hlavné funkcie Digitálneho archívu ŠGÚDŠ:

- Automatizované importy skenovaných dokumentov pre hromadné napĺňanie archívu digitálnym obsahom
- Manuálne vkladanie dokumentov formou „drag and drop“, vrátane hromadného vloženia viacerých dokumentov v jednom kroku
- Vyhľadávanie dokumentácie pomocou predpripravených formulárov, ktoré je možné administrátorsky upravovať
- Vyhľadávanie dokumentácie pomocou priestorovej lokalizácie (mapy) v kombinácii s formulárovým vyhľadávaním
- Fulltextové vyhľadávanie pri rozpoznávaní slovenského jazyka a gramatiky
- Automatické indexovanie dokumentov (OCR) v jazykoch: slovenčina, angličtina, nemčina
- Fazetové vyhľadávanie
- Organizácia dokumentov v archíve, možnosť užívateľsky vytvárať hierarchické zložky, priečinky a pridávať obsah do priečinkov
- Presúvanie a zmeny zaradenia dokumentácie v archíve
- Generovanie náhľadov na dokumenty
- Poznámkovanie dokumentov
- Verzionovanie zmien dokumentov a sprístupnenie historických verzií dokumentu
- Uloženie vyhľadaných dokumentov medzi „moje obľúbené“ pre budúce okamžité vyhľadanie bez nutnosti využívania vyhľadávacích formulárov
- Zobrazenie dokumentov pomocou univerzálneho prehliadača
- Riadená online výpožička dokumentov z archívu na základe oprávnení
- Tlač a export dokumentácie z archívu
- Export metaúdajov do tabuľkového formátu
- Podpora bibliografických štandardov pre rešeršné výstupy, generovanie rešeršných prehľadov podľa štandardov
- Podpora tlače katalogizačných štítkov
- Rozsiahla podpora užívateľskej personalizácie

- Konfigurovateľný a znovu použiteľný workflow, ktorý zabezpečí pracovný proces spracovania digitálneho obsahu počas jeho životného cyklu a schvaľovacieho procesu
- Odoslanie hypertextového odkazu na dokumentáciu inému používateľovi
- Podpora rozsiahlych užívateľsky definovaných reportov
- Bezpečnostný auditing
- Otvorené integračné rozhrania pre poskytovanie údajov a dokumentov z archívu, podpora pre harvesting.

Digitálny archív obsahuje v súčasnosti cca 2 600 000 dokumentov. Prírastok novej dokumentácie v archíve predstavuje cca 50 000 dokumentov mesačne. Digitálny archív eviduje v priemere 4000 prístupov používateľov mesačne a vykoná v priemere 7500 online výpožičiek mesačne.

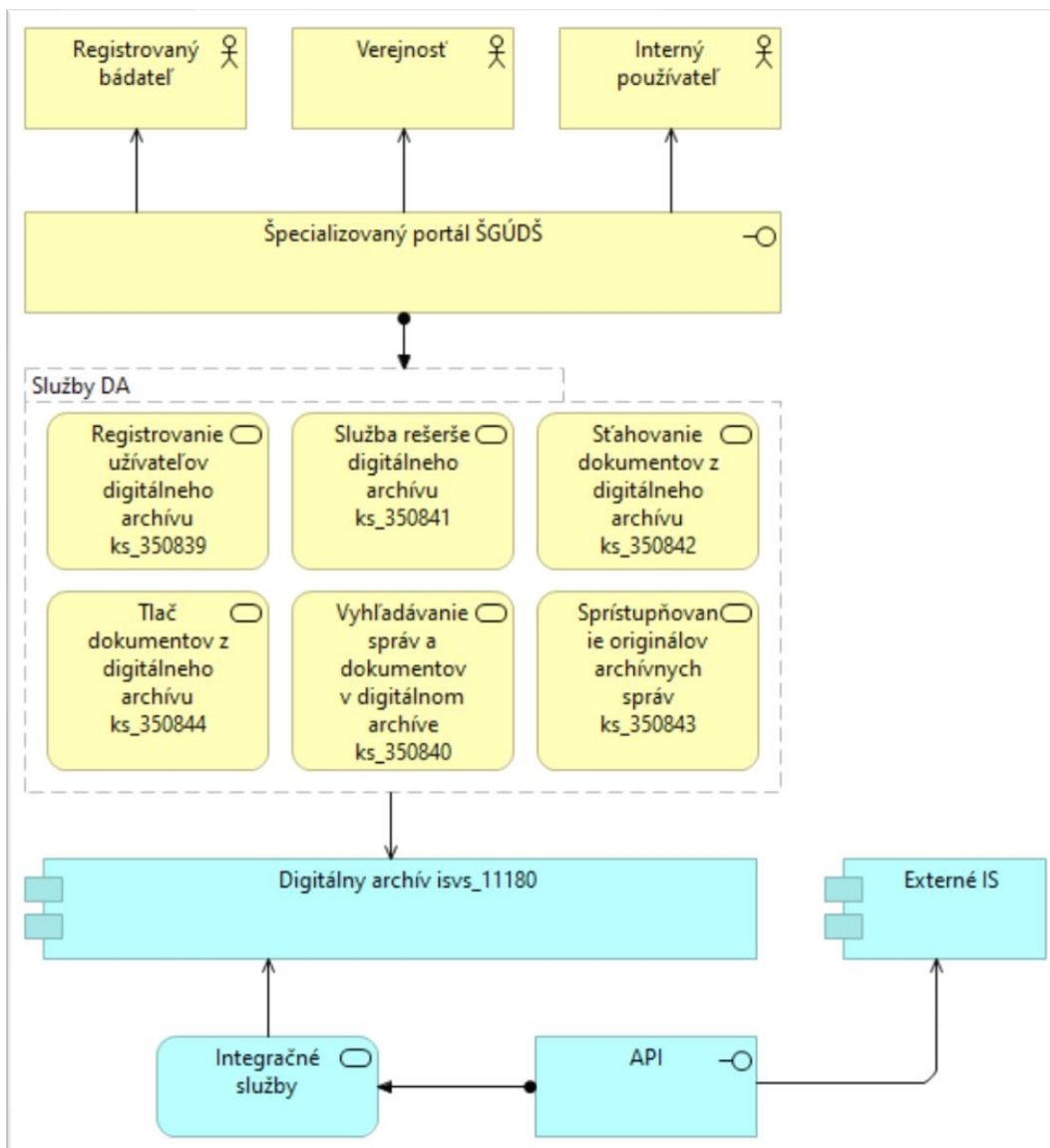
Digitálny archív je dostupný v prostredí internetu a v súčasnosti obsahuje tri používateľské rozhrania:

- Anonym – rozhranie je určené pre neprihlásených používateľov. Verejnosť má dostupné funkcie vyhľadávania dokumentov, tvorby rešeršných prehľadov, bez možnosti ich zobrazenia alebo stiahnutia.
- Registrovaný bádateľ - rozhranie je určené pre registrovaných používateľov, ktorí majú dostupné funkcie vyhľadávania dokumentov s možnosťou zobrazenia a stiahnutia dokumentu.
- Interný používateľ – rozhranie je určené pre interných zamestnancov, ktorí majú sprístupnené funkcie pre správu obsahu digitálneho archívu.

Digitálny archív poskytuje nasledujúce koncové služby:

- Tlač dokumentov z digitálneho archívu ks_350844, úroveň elektronizácie služby: úroveň 4
- Sprístupňovanie originálov archívnych správ ks_350843, úroveň elektronizácie služby: úroveň 3
- Sťahovanie dokumentov z digitálneho archívu ks_350842, úroveň elektronizácie služby: úroveň 4
- Služba rešerše digitálneho archívu ks_350841, úroveň elektronizácie služby: úroveň 4
- Vyhľadávanie správ a dokumentov v digitálnom archíve ks_350840, úroveň elektronizácie služby: úroveň 4
- Registrovanie užívateľov digitálneho archívu ks_350839, úroveň elektronizácie služby: úroveň 2.

Biznis architektúra Digitálneho archívu je znázornená na nasledujúcom obrázku:



Obrázok 1 Biznis architektúra Digitálneho archívu AS IS

Vyhľadávacie služby

Dokumenty je možné v archíve vyhľadávať pomocou elektronických formulárov, ktoré obsahujú aj mapu pre výber lokality. V systéme je v súčasnosti možné vyhľadávať objekty typu:

- Záverečné geologické správy
- Časopisy

- Články časopisov
- Monografie
- Články monografií
- Geologické mapy.

Každý vyhľadávací formulár je osobitne upravovaný pre danú úroveň prístupov (iné vyhľadávacie kritéria môže využívať verejnosť, iné registrovaní bádatelia a iné interní pracovníci). Rozsah sprístupnených vyhľadávacích kritérií a ich parametre je možné administrátorsky konfigurovať. Určené kritéria využívajú služby správy číselníkov, ktoré môžu byť aj rozsiahle (napr. register autorov, register lokalít, mapové listy WGS84 a S-JTSK a pod.).

Dokumenty je možné vyhľadávať kombináciou vyhľadávacích kritérií a výberom záujmovej lokality v mape. V takom prípade archív vyhľadáva všetky dokumenty, ktoré spĺňajú zadané kritéria a sú priestorovo prislúchajúce ku zvolenej lokalite. Pre vyhľadávanie prostredníctvom mapy sú v rámci Digitálneho archívu implementované mapové a aplikačné služby platformy ESRI ArcGIS for Server. Podkladové mapy, mapové listy a súvisiace mapové služby využíva Digitálny archív z celopodnikového Mapového servera GIS nasadeného a prevádzkovaného v rámci ŠGÚDŠ.

Vyhľadané dokumenty sú zoskupované do kategórií a prehľadne sprístupnené spolu s ďalšími popisnými údajmi. Široká paleta doplnkových nástrojov výrazne rozširuje možnosti práce s archivovaným dokumentom. Ich dostupnosť pre používateľa je riadená na základe systému oprávnení. Archív umožňuje konfiguračne nastavovať výsledky vyhľadávania a rozsah zobrazovaných metaúdajov dokumentov.

Súčasťou služieb vyhľadávania je fulltextové vyhľadávanie slov alebo presných fráz. Vyhľadávanie využíva technológiu IBM Watson, ktorá je určená pre extrémne presné a rýchle vyhľadanie slov a fráz vo veľkom množstve dokumentácie. Výsledky vyhľadania sú automaticky kategorizované pre pohodlnú orientáciu a ďalšie bližšie spresnenie hľadanej dokumentácie. V rámci Digitálneho archívu ŠGÚDŠ sú konfigurované tri komplexné vyhľadávače (crawlers) pre geologické správy a ich obsah, kapitoly monografií a ich obsah, články časopisov a ich obsah.

Prezeracie a sťahovacie služby

Na prezeranie dokumentov sú implementované špecifické služby, ktoré podľa „mime type“ formátov súborov sprístupňujú prezeracie aplikácie. Predvolene je využívaný vstavaný prehliadač platformy IBM Content Navigator. Súčasťou digitálneho archívu je aj univerzálny prehliadač dokumentov postavený na softvérovej platforme IIP Image. Univerzálny prehliadač umožňuje zobrazovať archivované dokumenty viac ako 200 formátov súborov bez potreby dodatočnej inštalácie asociovaných programov. Používateľ tak môže prezeráť rôzne výkresy, videá, multispektrálne snímky, súbory špecializovaných aplikácií bez toho, aby tieto aplikácie musel mať nainštalované na vlastnom počítači. Univerzálny prehliadač je súčasne prispôsobený pre rýchle zobrazovanie veľkoobjemových dokumentov, rádovo stoviek megabajtov až gigabajtov (v prípade historických geologických máp skenovaných vo vysokom rozlíšení).

Sťahovanie súborov z archívu je riadené funkcionalitou online výpožičiek. Vypožičanie dokumentu z archívu sa riadi definovanými prístupovými oprávneniami používateľa. Dokumenty

je možné online vypožičiavať vo formáte PDF alebo v originálnom formáte dokumentu. V prípade voľby formátu PDF archív zabezpečuje konverziu dokumentu do tohto formátu.

Digitálny archív má implementované sťahovacie a prezeracie služby aj formou štandardného otvoreného rozhrania (Open API), ktoré umožňuje autorizovaných externým informačným systémom vyhľadávať dokumentácie, harvestovať obsah archívu a taktiež sťahovať dokumenty z archívu.

Ukladacie služby

Digitálny archív umožňuje oprávneným používateľom manuálne vkladanie dokumentácie do archívu. Služba podporuje ukladanie dokumentov formou potiahni a pusť („drag and drop“), a to jednotlivého dokumentu ako aj sady viacerých dokumentov. Každý dokument, podľa jeho zaradenia do príslušnej objektovej triedy, má v archíve pripojené metaúdaje. V rámci objektovej štruktúry je možné vytvárať hierarchickú štruktúru. Digitálny archív má implementované dátové pumpy, ktoré umožňujú automatické zakladanie dokumentov ako aj celých zväzkov dokumentácie geologických správ, časopisov, monografií a geologických máp, ktoré sú výstupom skenovacej linky.

Digitálny archív má súčasne implementované integračné služby formou štandardného otvoreného rozhrania (Open API), ktoré umožňuje automatické zakladanie dokumentácie v archíve z externých informačných systémov. Ukladané dokumenty, v závislosti od formátu súboru a príslušnej konfigurácie indexovania, sú automaticky indexované vrátane indexovania obsahu dokumentov (OCR) pre následné vyhľadávanie v metadátach a fulltextovom vyhľadávaní v obsahu dokumentov. Indexovanie je zabezpečené nástrojmi IBM ECM a IBM Watson.

Reportovacie služby

Reportovacie služby zabezpečujú riadený prístup k výstupným zostavám a štatistickým prehľadom údajov Digitálneho archívu verejnosťou a internými pracovníkmi ŠGÚDŠ. Služby sú postavené na technológii Jasper Reports. V súčasnosti sú implementované nasledovné reportovacie služby:

- Rešeršné prehľady – umožňujú generovanie rešerše podľa bibliografických pravidiel do formátu MS Word podľa zadaných kritérií vo vyhľadávacích formulároch.
- Katalogizačné štítky – umožňuje vytlačenie katalogizačných štítkov podľa knižničných pravidiel ŠGÚDŠ pre používateľom definované dokumentácie.
- Odtajnené správy – umožňuje generovať prehľady geologických správ, ktoré boli v určitom období odtajnené a sprístupnené pre sťahovacie služby.
- Polročný výkaz a Ročné prírastky – komplexná zostava, ktorá je generovaná pre potreby nadriadeného štátneho orgánu.
- Štatistické výkazy sťahovaných správ a početností objektov – sada výstupných zostáv, ktoré sprístupňujú analytické údaje o používanosti Digitálneho archívu.
- Report auditového žurnálu – analytická zostava, ktorá sprístupňuje údaje o prístupoch, online vypožičkách a všetkých iných aktivitách, ktoré používatelia vykonávajú v Digitálnom archíve.

Integračné služby

Digitálny archív je pripravený na integráciu s inými informačnými systémami prostredníctvom štandardných rozhraní a integračných štandardov. Integračné služby sú implementované využitím integračnej platformy Apache Camel. Apache Camel slúži na prenos a smerovanie správ pomocou pravidiel, ktorý implementuje jednotlivé integračné služby. Úlohou tohto integračného komponentu je prepojenie Digitálneho archívu s externými aplikáciami ako aj poskytovanie služieb dovnútra celého systému Digitálneho archívu. Komponent umožňuje prijať dáta z ľubovoľného zdroja a na základe definovaných pravidiel tieto údaje transformovať a dopraviť do cieľa. Archív poskytuje v súčasnosti integračné služby v nasledovnom rozsahu:

Dátová pumpa CM Import

Dátová pumpa slúži k importu bibliografických dát do IBM Content Manager. Importované sú nasledovné typy vstupných dát:

- celé Monografie, kapitoly monografií a metaúdaje
- celé Časopisy, kapitoly časopisov a ich metaúdaje
- celé Správy, jednotlivé strany geologickej správy a ich metaúdaje
- Geologické mapy a ich metaúdaje

Údaje a dokumenty sú importované zo štandardizovaných PSP balíkov, ktoré sú výstupom skenovacej linky. Dokumenty sú prevažne vo formátoch PDF - skenované písomnosti, JPEG2000 - veľké prílohy, mapy a schémy, XML metadáta v bibliografickom štandarde METS.

Dátová pumpa využíva webové služby IBM Content Manager a je realizovaná využitím nástrojov platformy Apache Camel.

OAI-PMH

Digitálny archív poskytuje otvorené rozhranie OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) pre web harvesting obsahu archívu. Rozhranie je implementované ako data provider na poskytovanie metadát v štandarde DublinCore z repozitára harvestovaných iným systémom prostredníctvom protokolu OAI-PMH. Na implementáciu sú využívané nástroje OAI-Cat. Súčasťou implementácie je konektor na úložisko, ktorým je Content Management digitálneho archívu. Rozhranie OAI-PMH na ŠGÚDŠ podporuje volania:

- služba na získanie zoznamu podporovaných typov objektov Digitálneho archívu
- služba na zoznam identifikátorov pre daný typ objektu (SPRAVA)
- služba na získanie údajov konkrétnej položky na základe identifikátora objektu
- služba na vyhľadávanie v metadátach pomocou atribútu from

Webservice Codelist

Integračná webová služba sprístupňuje číselníky Digitálneho archívu pre externý informačný systém. Služba používa Basic HTTP Authentication a na autentifikáciu sa používajú platné účty IBM CM. Služba je implementovaná využitím SOAP.

Webservice Search

Integračná webová služba poskytuje rozhranie pre prehľadávanie metaúdajov dokumentov Digitálneho archívu z externého informačného systému. Služba používa Basic HTTP Authentication a na autentifikáciu sa používajú platné účty IBM CM. Služba je implementovaná využitím SOAP. Pri vyhľadávaní sa validuje: štruktúra požiadavky (správnosť názvov položiek a atribútov), správnosť typov atribútov, správnosť hodnôt atribútov. Rozhranie podporuje volania:

- vyhľadanie jednoznačnej položky objektu podľa identifikátora. Ako identifikátor slúži hodnota jednoduchého atribútu, ktorá je jednoznačná v rámci
- všetkých záznamov danej položky objektu
- vyhľadanie záznamov danej položky objektu podľa zadaných atribútov
- vyhľadanie záznamov danej položky objektu pomocou fulltextového vyhľadávania (v prílohách)
- stiahnutie dokumentu vyhľadaného objektu prostredníctvom servlet download.

Sprístupnenie vydavateľskej činnosti na webovom sídle ŠGÚDŠ

Na webovom sídle ŠGÚDŠ, sekcii Vydavateľská činnosť je integrovaný špecificky pripravený komponent Digitálneho archívu, ktorý priamo v prostredí webového sídla sprístupňuje všetky vydania časopisov a všetky články týchto časopisov zoskupených podľa rokov vydania.

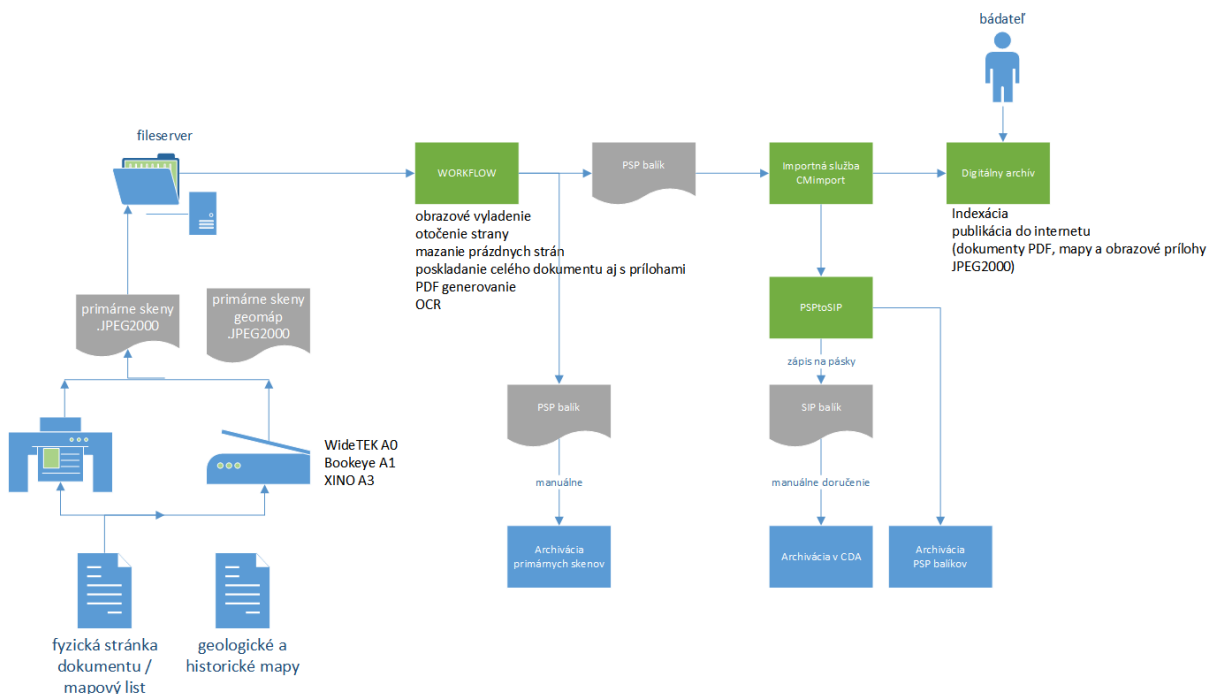
DA má priame napojenie na externý IS (REST, SOAP služby) v rámci MŽP:

- Informačný systém environmentálne záťaže, kód v MetaIS MŽP - isvs_300, prevádzkovateľ SAŽP.
- Priame napojenie na externý IS (REST, WMS a WFS služby) pre súkromný sektor
- IS Albert, IS na efektívne vyhodnocovanie rizík v konkrétnom území
- Priame napojenie na interné aplikácie Geologického IS v rámci ŠGÚDŠ, isvs_314:
- Samostatné aplikácie – Geofyzikálne mapy; Registre geofondu (8 aplikácií), 2D/3D Mapový portál.

4.1 Procesné riešenie skenovacej linky ŠGÚDŠ

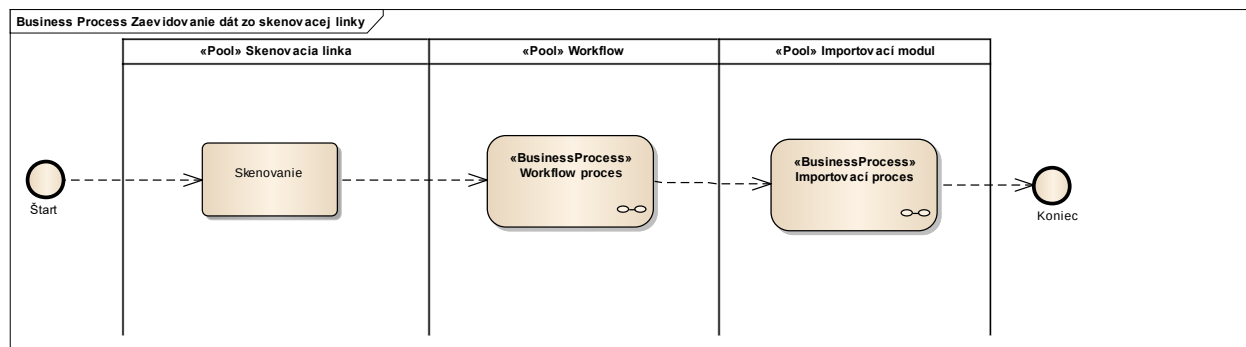
Fyzické dokumenty, veľkoformátové výkresy a mapy sú skenované veľkokapacitnými a vysokovýkonnými skenermi na skenovacom pracovisku Ústrednej geologickej knižnice. Dokumenty sú skenované vo vysokom rozlíšení vo formáte JPEG2000. Tie sú následne spracované softvérovým riešením workflow, ktoré zabezpečuje konverziu dokumentov, obrazové vyladenie, otočenie strán, odstránenie šumu a prázdnych strán, pospájanie strán do celkového dokumentu, OCR indexovanie pre štandard METS. Výstupom workflow sú vygenerované štandardné PSP balíky dokumentov. Tie sú predmetom automatizovaného importu do Digitálneho archívu, následného indexovania pre vyhľadávanie prostredníctvom vyhľadávacích služieb archívu a ich sprístupnenie pre koncových používateľov. Po publikácii dokumentov Digitálny archív automaticky archivuje PSP balíky na archivačnú páskovú mechaniku. PSP balík

je následne konvertovaný pomocou osobitnej aplikácie do štandardu SIP balíkov dokumentácie a TAR súborového archívu. SIP balíky sú ukladané na určené archivačné pásky páskovej mechaniky. Pásky sú fyzicky doručované do Univerzitnej knižnice, kde sa vykonáva ich dlhodobá archivácia v rámci historického uchovávanía kultúrneho dedičstva SR.



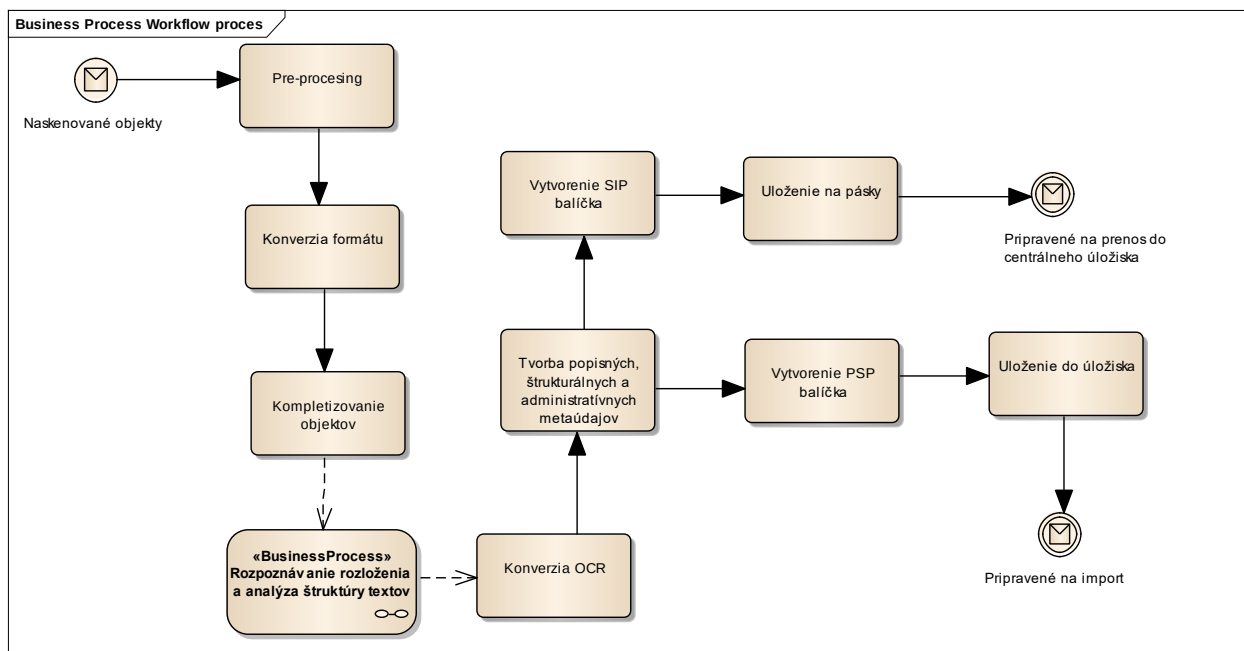
Obrázok 2 Proces skenovacej linky Digitálneho archívu

Proces zaevidovania dát zo skenovacej linky:



Obrázok 3 Hlavný proces – zaevidovanie dát zo skenovacej linky

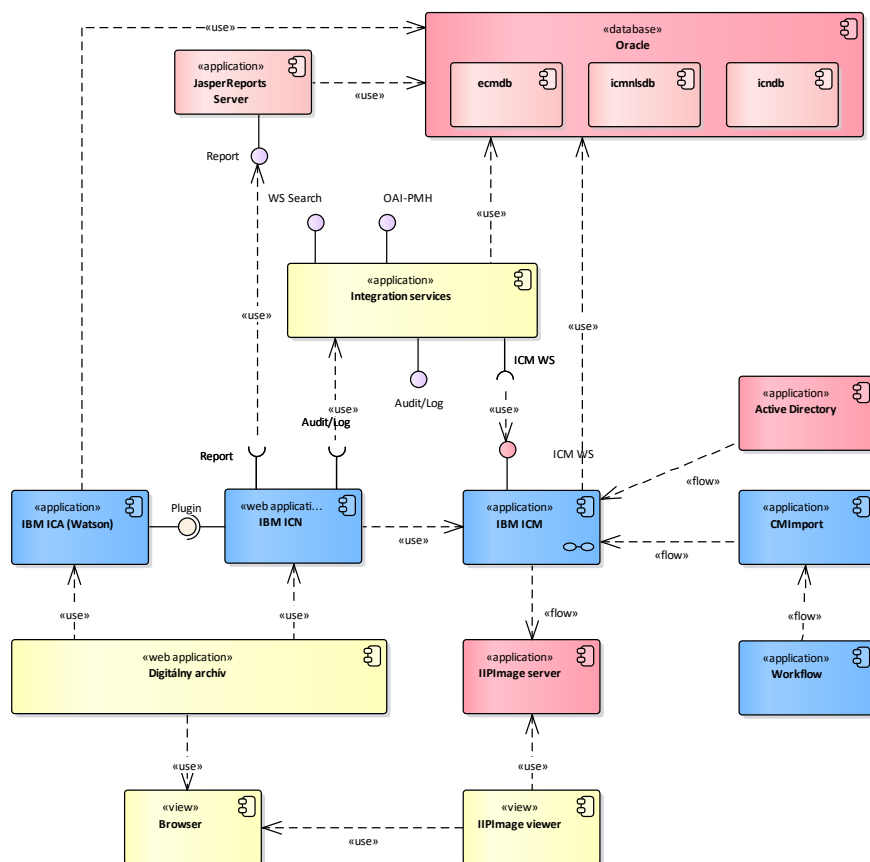
Na nasledujúcom obrázku je znázornený workflow proces dokumentov



Obrázok 4 Workflow proces skenovaných dokumentov

4.2 Aplikačná architektúra

Digitálny archív je postavený na technologickej softvérovej platforme IBM Content Manager, konfigurovaný a rozšírený o pluginy a ďalšie technológie pre zabezpečenie potrieb ŠGÚDŠ. Je optimalizovaný pre vysokú kapacitnú a výkonovú záťaž. Pracuje pri objeme rádovo miliónov dokumentov, pri súčasnom prístupe stoviek používateľov.



Sumárny prehľad technológií:

- **Jasper Reports** – pre tvorbu a generovanie reportov a štatistických prehľadov
- **Apache Camel** – pre automatizované dátové pumpy
- **IIP Image** – pre prehliadanie veľkoobjemových dokumentov, schém, príloh a máp
- **ESRI ArcGIS server API** – pre podporu vyhľadávania dokumentácie prostredníctvom priestorovej lokalizácie
- **Špecifické pluginy pripojené do IBM Content Manager** pre rozšírené funkcionality Digitálneho archívu, vytvorené v prostredí Java, Spring
- **Javascript framework Dojo** – pre vytvorenie používateľského rozhrania.
- **Oracle databáza 11.2.0.4** – údaje sú organizované v troch Oracle databázach:
 - o ecmdb - obsahuje centrálny registre a číselníky používané v celom Digitálnom archíve
 - o icmnsdb – centrálna databáza IBM CM obsahujúca všetky údaje archívu
 - o icndb - centrálna databáza IBM ICN obsahujúca konfigurácie archívu
- **Údaje sú v jednotlivých databázach** ukladané v objektovo orientovanej štruktúre. Prístup k údajom je riešený prostredníctvom API nástrojov IBM Content manager.
- **Konfigurácie IBM Watson** – štandardné konfiguračné súbory IBM Watson pre indexovanie dokumentov pre fulltextové a fazetové vyhľadávanie.
- **IS Sirius v1.0.88.0** – komplexné riešenie pre prácu s dokumentami. Sirius je modulárny systém navrhnutý na vytváranie komplexných aplikácií v oblastiach, ako je digitalizácia a správu dokumentov, systémy pre archiváciu, spracovanie a skladovanie textových a obrazových informácií, optické rozpoznávanie a spracovanie rukopisu formy. Používa sa aj v priemyselných aplikáciách, ktoré pracujú s opticky získanými informáciami, napríklad čítanie, čiarové kódy atď. Vďaka svojej orientácii na databázu architektúra klient/server a rozsiahlej sade výkonných nástrojov, umožňuje rýchly vývoj robustných aplikácií. Dávkové spracovanie, spracovanie obrazu, rozpoznávanie formulárov, OCR/ICR/OMR/čiarové kódy, vyťažovanie dát, indexovanie dokumentov, organizácia databázy, zobrazovacie nástroje, podpora archivačných médií.

Prístup používateľov do archívu je riadený systémom užívateľských účtov a prístupových zoznamov IBM ICM. Možnosti používateľa pri práci s archívom riadi konfigurácia pracovnej plochy IBM ICN. Pre každú skupinu používateľov je nadefinovaná samostatná pracovná plocha, ktorá určuje, ku ktorým funkciám archívu má používateľ prístup. Správa používateľských účtov je realizovaná napojením na MS Active directory. Zaradenie používateľa do príslušnej skupiny pri synchronizácii s IBM ICM zabezpečuje správne priradenie užívateľských práv (prístupových zoznamov). Synchronizácia s AD prebieha pomocou nástrojov IBM ICM.

Prístup k používateľským rozhraniám archívu prebieha cez zabezpečený HTTPS protokol s využitím bezpečnostného SSL certifikátu vydaným certifikačnou autoritou.

Digitálny archív loguje všetky udalosti systému a všetky aktivity používateľov zaznamenáva prostredníctvom auditového žurnálu. Digitálny archív má implementované špecifické aplikačné služby, ktoré zabezpečujú súlad celého archívu s nariadením o ochrane osobných údajov (GDPR). Riadi retenciu uchovávaní údajov a umožňuje uplatnenie práva na zabudnutie.

Digitálny archív nevyužíva nadrezortné centrálny bloky a podporné spoločné bloky (SaaS) a projektom sa ani nepredpokladá rozšírenie v tejto oblasti.

Digitálny archív nevyužíva integrácie na nadrezortné centrálné bloky (IS CSRÚ) a projektom sa ani nepredpokladá rozšírenie v tejto oblasti.

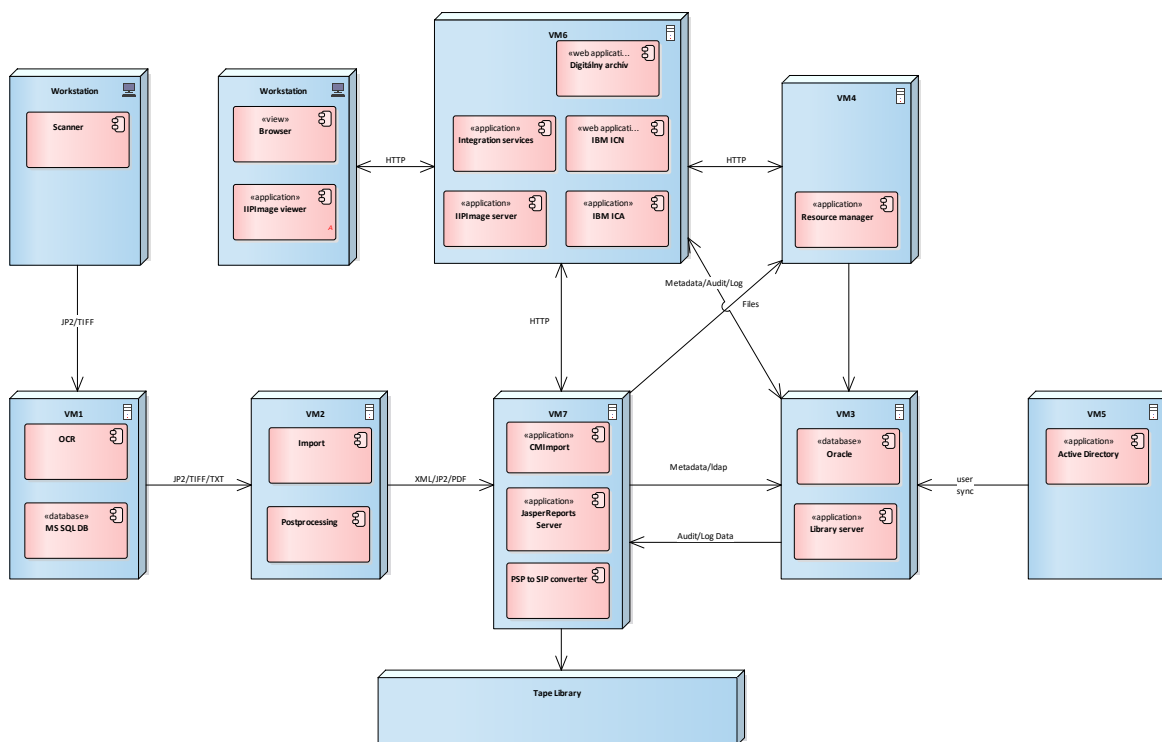
Digitálny archív neposkytuje údaje do IS CSRÚ, ani nekonzumuje údaje z IS CSRÚ a projektom sa ani nepredpokladá rozšírenie v tejto oblasti.

4.3 Technologická architektúra

Digitálny archív je prevádzkovaný na dedikovanej HW infraštruktúre umiestnenej v sídle ŠGÚDŠ. Pozostáva z nasledovných komponentov:

- 4x Fyzický server SUN
- Operačný systém Oracle Solaris (2x) cluster a MS Windows server 2012 (2x) v prevedení failover cluster
- Pásková knižnica Oracle StorageTek SL150 (zásobník 60 pásov) a archivačným SW Oracle SAM-QFS
- Virtualizácia MS Hyper-V.

Súčasná technologická architektúra:



Obrázok 6 Model nasadenia Digitálneho archívu vo virtualizovanej infraštruktúre

5 Súpis požiadaviek

5.1 Legislatívne požiadavky

V rámci projektu migrácie sa požaduje dodržať legislatívne požiadavky vyplývajúce z nasledovných predpisov:

- Zákon č. 95/2019 Z. z., o informačných technológiách vo verejnej správe a o zmene a doplnení niektorých zákonov v znení neskorších predpisov
- Zákon č. 69/2018 Z. z. o kybernetickej bezpečnosti a o zmene a doplnení niektorých zákonov v znení neskorších predpisov
- Zákon č. 45/2011 Z.z. o kritickej infraštruktúre o zmene a doplnení niektorých zákonov v znení neskorších predpisov
- Zákon č. 18/2018 Z.z. o ochrane osobných údajov (GDPR) a o zmene a doplnení niektorých zákonov v znení neskorších predpisov
- Zákon č. 305/2013 Z.z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov v znení neskorších predpisov (zákon o e-Governmente)
- Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné technológie verejnej správy
- Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č.85/2020 o riadení projektov
- Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 179/2020 Z. z., ktorou sa ustanovuje spôsob kategorizácie a obsah bezpečnostných opatrení informačných technológií verejnej správy
- Vyhláška Úradu na ochranu osobných údajov Slovenskej republiky č. 158/2018 Z. z. o postupe pri posudzovaní vplyvu na ochranu osobných údajov
- Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov)

5.2 Požiadavky na spracovanie analýzy a návrhu riešenia projektu

Verejný obstarávateľ požaduje, aby úspešný uchádzač spracoval v úvodnej etape projektu podrobnú analýzu a návrh riešenia migrácie Digitálneho archívu a zabezpečenie implementácie požadovaných funkčných a nefunkčných požiadaviek v nasledovnom rozsahu:

ID požiadavky	CBA ID	Požiadavka
RQ_ANL_1	1	Realizovať analýzu dátového modelu Digitálneho archívu a jeho obmedzení na spôsob migrácie
RQ_ANL_2	2	Realizovať analýzy obmedzení aktualizácie verzií operačných systémov s ohľadom na technológie využívané v celom riešení Digitálneho archívu a jeho aplikačných komponentoch

ID požiadavky	CBA ID	Požiadavka
RQ_ANL_3	3	Realizovať analýzy obmedzení aktualizácie verzií databázových systémov s ohľadom na technológie využívané v celom riešení Digitálneho archívu a jeho aplikačných komponentoch
RQ_ANL_4	4	Spracovať návrh bezpečného prepojenia cloudu a prostredia ŠGÚDŠ
RQ_ANL_5	5	Spracovať návrh databázového riešenia Digitálneho archívu v prostredí cloudu
RQ_ANL_6	6	Realizovať analýzu bezpečnostných riešení súčasného riešenia a navrhnuť opatrenia pre plnenie bezpečnostných požiadaviek vyplývajúcich z legislatívy a bezpečnostných štandardov
RQ_ANL_7	7	Spracovať návrh bezpečnostných riešení implementácie v cloude
RQ_ANL_8	8	Vytvoriť varianty nasadenia súčasných technológií riešenia a spracovať návrh optimalizácie spôsobu prístupu užívateľa k službám IS DA
RQ_ANL_9	9	Spracovať návrh spôsobu novej inštalácie IS DA v prostredí cloudu
RQ_ANL_10	10	Spracovať Proof of Concept vzorovej inštalácie IS DA - overenie spôsobu inštalácie IS DA v prostredí cloudu
RQ_ANL_11	11	Spracovať návrh sieťového nasadenia a sieťových prestupov
RQ_ANL_12	12	Spracovať a odovzdať Detailný návrh riešenia (DNR)
RQ_ANL_13	13	Zabezpečiť dokumentáciu potrebnú pre vytvorenie projektu v cloude
RQ_ANL_14	-	Spracovať návrh aplikačných úprav systému smerom k prispôsobeniu systému princípom „Cloud-native computing“

5.3 Funkčné požiadavky

5.3.1 Základné požiadavky

ID požiadavky	CBA ID	Požiadavka
RQ_FP_ZP_1	15	Zabezpečiť všetky potrebné aplikačné a infraštruktúrne konfigurácie potrebné pre vykonanie migrácie.
RQ_FP_ZP_2	16	Koncový používateľ musí mať prístup ku všetkým existujúcim údajom z existujúcich dátových polí aj v prostredí cloudu.
RQ_FP_ZP_3	21	Zabezpečiť optimálnu decentralizovanú architektúru pre komfortné využívanie používateľom. Existujúce riešenie je postavené na lokálnej

		sieti, diskové pole prepojené optimalizovanými aktívnymi sieťovými prvkami. Pri presune do cloudu bude riešenie čiastočne decentralizované tak, aby užívateľovi nevznikali kroky navyše pri spracovávaní skenov (dátové sklady – privátny cloud, aplikačné servery - cloud, pracovné stanice zabezpečujúce skenovanie umiestnené v ŠGÚDŠ).
RQ_FP_ZP_4	48	Správa obsahu užívateľmi Požaduje sa umožnenie užívateľom vytvárať, spravovať a publikovať obsah a zároveň si zachovať možnosť ho spravovať aj ďalej. Požaduje sa kompletne riešenie uchovávaní a získavania dokumentov v takmer všetkých formátoch.
RQ_FP_ZP_5	51	Správa používateľských účtov Systém musí umožniť pre každú skupinu používateľov sprístupniť samostatnú pracovnú plochu, ktorá určuje, ku ktorým funkciám archívu má používateľ prístup. Prístup k používateľským rozhraniám archívu musí prebiehať cez zabezpečený HTTPS protokol s využitím bezpečnostného SSL certifikátu vydaným certifikačnou autoritou.
RQ_FP_ZP_6	74	Umožniť používateľskú personalizáciu vizualizovaného obsahu.

5.3.2 Skenovacia linka a workflow spracovania dát

Pre zaistenie migrácie Digitálneho archívu do cloudu bude nevyhnutné vykonať úpravy, rozšírenie a migráciu existujúceho Workflow, ktoré zaisťuje kompletný proces spracovania fyzických dokumentov od skenovania až po vygenerovanie štandardizovaných PSP balíkov, ktoré sú následne automaticky importované do Digitálneho archívu pre sprístupnenie koncovým užívateľom.

ID požiadavky	CBA ID	Požiadavka
RQ_FP_WF_1	-	Skenovacia linka ostáva v ŠGÚDŠ Z IS DA ostáva export metadát pre potreby workflow Primárne skeny vstupujú do workflow, ktorý ostáva v prostredí ŠGÚDŠ Výstupné PSP balíky sa automatizovane bezpečne migrujú do cloud
RQ_FP_WF_2	-	Požaduje sa zachovanie existujúcich funkcionalít workflowu: - Centrálna integrovaná správa digitalizovaných dokumentov a procesov ich spracovania. - Modulárna klient-server architektúra s centrálnou správou pracovných staníc, aplikačnou vrstvou (obsahujúca moduly pre konverzie dát, pre-procesing a post-procesing digitalizovaných súborov, OCR modul, analytický modul na detekciu štruktúry strán, moduly na tvorbu metadát a ich export do štandardizovaných PSP balíkov, modul pre správu užívateľov, vyhľadávací engine a pod.), centrálnou databázovou vrstvou a interným zabezpečeným úložiskom pracovných aj finálnych dokumentov. - Validačný modul zaisťujúci priebežnú validáciu automaticky aj ručne vytváraných metadát k jednotlivým dokumentom podľa

ID požiadavky	CBA ID	Požiadavka
		nastavených pravidiel vrátane finálnej validácie popisných, štrukturálnych a administratívnych metadát PSP balíkov importovaných do Digitálneho archívu. - Reportovací modul zaisťujúci detailný monitoring a evidenciu priebehu digitalizačného procesu.
RQ_FP_WF_3	-	Požaduje sa migrácia existujúceho aplikačného, procesného a databázového servera na nové virtuálne servery. Súčasťou procesu bude analýza existujúceho stavu databázy, verifikácia dát, vytvorenie migračných skriptov a následná migrácia dát do nového pracovného úložiska, optimalizácia databáz, indexácia, novo nastavené zálohovanie. V rámci migrácia sa požaduje vykonať prechod na databázový server SQL Server 2019 a taktiež konsolidácia databáz do jednej spoločnej a vytvorenie archívnej databázy.
RQ_FP_WF_4	-	Požaduje sa rozšírenie exportného modulu o vytvorenie finálneho PSP balíka v štruktúre vhodnej pre import do cloudu vrátane aktualizácie použitých medzinárodných štandardov pre monografia a periodiká, METS, MODS, Dublin Core, ALTO, MIX, PREMIS na najnovšie verzie (podľa popisu na stránkach Kongresovej knižnice https://www.loc.gov/librarians/standards)
RQ_FP_WF_5	-	Požaduje sa aktualizácia existujúceho OCR modulu na novú verziu ABBYY FineReader Engine 12 s ohľadom na tvorbu ALTO sekcií v PSP balíku a zabezpečenie možnosti využitia aj iných externých OCR nástrojov.
RQ_FP_WF_6	-	Požaduje sa rozšírenie validačného modulu s ohľadom na použitie aktuálnych medzinárodných štandardov a jeho začlenenie do WorkFlow na všetkých úrovniach spracovania.
RQ_FP_WF_7	-	Požaduje sa modernizácia a rozšírenie užívateľského rozhrania WorkFlow s ohľadom na nové funkcie a systémové nastavenia po prevode Digitálneho archívu do cloudu.
RQ_FP_WF_8	-	Požaduje sa rozšírenie modulu pre Reporting pre zaistenie prehľadu o stave interného úložiska a evidencii prípadných chýb.
RQ_FP_WF_9	22	Používateľ pri práci s naskenovanými dokumentami bude mať určený čas na zasielanie balíkov údajov tak, aby to neohrozilo kontinuitu práce a zabezpečil sa plynulý presun údajov do cloudu. Je potrebné dostatočne nadimenzovať šírku pásma medzi jednotlivými uzlami topológie riešenia. Riziko je však možné limitovať správnou architektúrou riešenia a dávkovými presunmi údajov mimo najviac vyťažенých časových úsekov.
RQ_FP_WF_10	45	Workflow musí zabezpečiť konverziu dokumentov, obrazové vyladenie, otočenie strán, odstránenie šumu a prázdnych strán, pospájanie strán do celkového dokumentu, OCR indexovanie pre štandard METS. Výstupom workflow sú vygenerované štandardné PSP balíky dokumentov. Tie sú predmetom automatizovaného importu do Digitálneho archívu, následného indexovania pre vyhľadávanie prostredníctvom vyhľadávacích služieb archívu a ich sprístupnenie pre koncových používateľov. • Konfigurovateľný a znovu použiteľný workflow, ktorý zabezpečí pracovný proces

ID požiadavky	CBA ID	Požiadavka
		spracovania digitálneho obsahu počas jeho životného cyklu a schvaľovacieho procesu

5.3.3 Vyhľadávacie služby

ID požiadavky	CBA ID	Požiadavka
RQ_FP_VS_1	56	Customizované formuláre pre každú úroveň prístupu. Každý vyhľadávací formulár musí byť osobitne upravovaný pre danú úroveň prístupov (iné vyhľadávacie kritéria môže využívať verejnosť, iné registrovaní bádatelia a iné interní pracovníci).
RQ_FP_VS_2	60	Správa vyhľadávacích formulárov Požaduje sa, aby bolo možné administrátorsky upravovať vyhľadávacie formuláre.
RQ_FP_VS_3	70	Riadená online výpožička Zabezpečiť riadenú online výpožičku dokumentov z archívu na základe oprávnení
RQ_FP_VS_4	25	Umožniť využívanie vyhľadávacích služieb anonymnému používateľovi bez nutnosti prihlásenia. Služby pre verejnosť sú určené pre neprihlásených používateľov. Verejnosť má dostupné funkcie vyhľadávania dokumentov a tvorby rešeršných prehľadov, bez možnosti zobrazenia alebo výpožičky samotného dokumentu. Vyhľadávací formulár musí byť samostatný od ostatných vyhľadávacích formulárov a musí obsahovať položky vyhľadávania definované v DNR. Súčasťou vyhľadávacieho formulára pre anonymných používateľov musí byť aj mapa, ktorá využíva mapové služby poskytované mapovým serverom ŠGÚDŠ.
RQ_FP_VS_5	26	Umožniť využívanie rozšírených vyhľadávacích služieb pre registrovaného bádateľa (prihláseného používateľa). Registrovaný bádateľ má dostupné služby a funkcie DA vyhľadávania dokumentov s možnosťou ich zobrazenia a online výpožičky. Vyhľadávací formulár musí byť samostatný od ostatných vyhľadávacích formulárov a musí obsahovať položky vyhľadávania definované v DNR. Súčasťou vyhľadávacieho formulára pre registrovaných bádateľov musí byť aj mapa, ktorá využíva mapové služby poskytované mapovým serverom ŠGÚDŠ.
RQ_FP_VS_6	27	Umožniť využívanie rozšírených vyhľadávacích služieb pre interných používateľov. Interný používateľ má dostupné osobitné služby a funkcie DA vyhľadávania dokumentov s možnosťou ich zobrazenia a online výpožičky, vrátane funkcionalít pre správu obsahu digitálneho archívu. Vyhľadávací formulár musí byť samostatný od ostatných vyhľadávacích formulárov a musí obsahovať položky vyhľadávania definované v DNR.

ID požiadavky	CBA ID	Požiadavka
		Súčasťou vyhľadávacieho formulára pre interných používateľov musí byť aj mapa, ktorá využíva mapové služby poskytované mapovým serverom ŠGÚDŠ.
RQ_FP_VS_7	28	Umožniť vyhľadávať objekty v archíve typu: - Záverečné geologické správy - Časopisy - Články časopisov - Monografie - Články monografií - Geologické mapy
RQ_FP_VS_8	61	Vyhľadávanie spojené s priestorovou lokalizáciou Požaduje sa zabezpečiť vyhľadávanie dokumentácie pomocou priestorovej lokalizácie (mapy) v kombinácii s formulárovým vyhľadávaním. Mapový komponent je osobitným rozšírením CM, ktorý umožňuje využiť priestorové informácie pre vyhľadávacie služby. Požaduje sa zabezpečiť inštaláciu a konfiguráciu komponentu v prostredí cloud, pričom je potrebné zabezpečiť: - Konfiguráciu podkladovej mapy Open street maps v súradnicovom systéme WGS84 - Sprístupnenie vyhľadania lokality v mape pomocou textového vyhľadávača s ponúkanými možnosťami - Označenie lokality priamo v mape - Zabezpečenie identifikácie mapových listov WGS84 - Poskytnutie označených lokalít pre vyhľadávacie služby CM Zabezpečiť dostupnosť základnej obsluhy mapy – posun v mape, zmena mierky zobrazenia mapy.
RQ_FP_VS_9	50	Analýza obsahu archívu Požaduje sa implementovať analytický nástroj na analýzu obsahu archívu. Komponent musí zabezpečiť automatickú indexáciu vkladáných dokumentov, vrátane správy vyhľadávacích indexov potrebných pre rýchle a presné vyhľadávanie a analýzy. Komponent taktiež bude zabezpečovať služby pre fazetové vyhľadávania.
RQ_FP_VS_10	55 57	Indexácia a fulltext vyhľadávanie Každý dokument musí byť pri vkladaní do archívu indexovaný pre rýchle vyhľadanie. Požaduje sa, aby indexovanie prebiehalo nie len nad popisnými atribútmi dokumentu (metaúdajmi), ale aj na základe hĺbkovej analýzy obsahu dokumentu. Digitálny archív musí umožniť koncovým používateľom rýchlo a jednoducho vyhľadať požadovanú dokumentáciu, a to formou predpripravených vyhľadávacích formulárov, ale aj voľným fulltext vyhľadaním slova, resp. frázy.
RQ_FP_VS_11	81 82	Výsledok fulltextového vyhľadávania sa požaduje realizovať aj s podporou vyznačenia (highlighting) a ukážky odstavce s hľadanými a nájdenými výrazmi.

ID požiadavky	CBA ID	Požiadavka
RQ_FP_VS_12	58	Usporiadanie výsledkov hľadania Digitálny archív musí prispôbiť výsledky vyhľadávania usporiadaním a uprednostňovaním zoznamu vyhladaných reťazcov na základe užívateľských zvyklostí. Musí rozumieť štylistike slovenského jazyka vrátane skloňovania, vďaka čomu sa dosiahnu omnoho lepšie výsledky vyhľadávania.
RQ_FP_VS_13	63	Fazetové vyhľadávanie Požaduje sa sprístupniť výsledky vyhľadávania s podporou pre fazetové vyhľadávanie podľa minimálne podľa rokov, autorov, monografií, geologických časopisov a typov dokumentov.
RQ_FP_VS_14	79	Fuzzy vyhľadávanie Požaduje sa sprístupniť výsledky vyhľadávania s podporou pre fuzzy vyhľadávanie.
RQ_FP_VS_15	80	Vyhľadávanie podobných výrazov Požaduje sa sprístupniť výsledky vyhľadávania s podporou pre vyhľadávanie podobných výrazov.
RQ_FP_VS_16	66	Náhľady na dokumenty Požaduje sa implementovať generovanie náhľadov na dokumenty
RQ_FP_VS_17	78	Prezeracie služby Na prezeranie dokumentov sú implementované špecifické služby, ktoré podľa „mime type“ formátov súborov sprístupňujú prezeracie aplikácie. Súčasťou digitálneho archívu je aj univerzálny prehliadač dokumentov postavený na softvérovej platforme IIP Image. Univerzálny prehliadač umožňuje zobrazovať archivované dokumenty viac ako 200 formátov súborov bez potreby dodatočnej inštalácie asociovaných programov. Používateľ tak môže prezerat rôzne výkresy, videá, multispektrálne snímky, súbory špecializovaných aplikácií bez toho, aby tieto aplikácie musel mať nainštalované na vlastnom počítači. Univerzálny prehliadač je súčasne prispôbený pre rýchle zobrazovanie veľkoobjemových dokumentov, rádovo stoviek megabajtov až gigabajtov (v prípade historických geologických máp skenovaných vo vysokom rozlíšení). Požaduje sa zabezpečiť univerzálny prehliadač aj v zmigrovanom riešení.
RQ_FP_VS_18	69	Uloženie dokumentov do "Moje obľúbené" Požaduje sa umožniť uloženie vyhladaných dokumentov medzi „Moje obľúbené“ pre budúce okamžité vyhľadanie bez nutnosti využívania vyhľadávacích formulárov
RQ_FP_VS_19	75	Umožniť odoslanie hypertextového odkazu na vyhladanú dokumentáciu inému používateľovi.

5.3.4 Sťahovacie služby

ID požiadavky	CBA ID	Požiadavka
RQ_FP_SS_1	29	Sťahovanie súborov z archívu je riadené funkcionalitou online výpožičiek. Vypožičanie dokumentu z archívu sa riadi definovanými prístupovými oprávneniami používateľa. Dokumenty musí byť umožnené online vypožičiavať vo formáte PDF alebo v originálnom formáte dokumentu. O každom vypožičaní dokumentu musí byť zaevidovaný príslušný auditný záznam kto, kedy a aký dokument vypožičal.
RQ_FP_SS_2	30	Konverzia dokumentu do formátu PDF pri online výpožičke V prípade voľby formátu PDF sa požaduje zabezpečiť automatickú konverziu dokumentu do tohto formátu, pokiaľ je dokument registrovaný v digitálnom archíve v inom formáte.
RQ_FP_SS_3	33	Sťahovanie - Open Api Digitálny archív má implementované sťahovacie a prezeracie služby aj formou štandardného otvoreného rozhrania (Open API), ktoré umožňuje autorizovaných externým informačným systémom vyhľadávať dokumentácie, harvestovať obsah archívu a taktiež sťahovať dokumenty z archívu. Požaduje sa túto funkcionalitu zabezpečiť aj po migrácii systému.
RQ_FP_SS_4	71	Tlač a export dokumentácie z archívu Zabezpečiť tlač dokumentov a export zoznamu vyhladaných dokumentov z archívu.
RQ_FP_SS_5	72	Zabezpečiť export meta údajov vyhladaných dokumentov do tabuľkového formátu.

5.3.5 Ukladacie služby

ID požiadavky	CBA ID	Požiadavka
RQ_FP_US_1	31	Digitálny archív umožňuje oprávneným používateľom manuálne vkladanie dokumentácie do archívu. Požaduje sa, aby služba podporovala ukladanie dokumentov formou potiahni a pusť („drag and drop“), a to jednotlivého dokumentu ako aj sady viacerých dokumentov. Každý dokument, podľa jeho zaradenia do príslušnej objektovej triedy, musí mať v archíve pripojené metaúdaje. V rámci objektovej štruktúry musí byť umožnené vytvárať hierarchickú štruktúru. Digitálny archív má implementované dátové pumpy, ktoré umožňujú automatické zakladanie dokumentov ako aj celých zväzkov dokumentácie geologických správ, časopisov, monografií a geologických máp, ktoré sú výstupom skenovacej linky. Požaduje sa zabezpečiť hromadné importovanie celých adresárov s dokumentami (napr. v prípade dodaného DVD média s dokumentov geologickej správy a jej prílohami organizovanými v podadresároch).
RQ_FP_US_2	59	Požaduje sa zabezpečiť automatizované importy skenovaných dokumentov pre hromadné napĺňanie archívu digitálnym obsahom, predovšetkým zo vstupných dátových sád v štandarde PSP balíkov.

ID požiadavky	CBA ID	Požiadavka
RQ_FP_US_3	54	Digitálny archív musí umožniť vkladanie veľkého spektra formátov dokumentov do digitálneho archívu systémom „drag and drop“.
RQ_FP_US_4	32	Automatické zakladanie dokumentácie v DA z externých IS Digitálny archív má súčasne implementované služby formou štandardného otvoreného rozhrania (Open API), ktoré umožňuje automatické zakladanie dokumentácie v archíve z autorizovaných externých informačných systémov. Ukladané dokumenty, v závislosti od formátu súboru a príslušnej konfigurácie indexovania, sú automaticky indexované vrátane indexovania obsahu dokumentov (OCR) pre následné vyhľadávanie v metadátach a fulltextovom vyhľadávaní v obsahu dokumentov. Požaduje sa umožniť hromadné zakladanie dokumentov z externých systémov vrátane následného indexovania.
RQ_FP_US_5	62	OCR v jazykoch Požaduje sa zabezpečiť automatické indexovanie dokumentov (OCR) v jazykoch: slovenčina, angličtina, nemčina
RQ_FP_US_6	-	Ukladanie geologických máp Požaduje sa zabezpečiť ukladanie digitálnych verzií geologických máp do Digitálneho archívu. Pri ukladaní sa požaduje umožniť označenie lokality umiestnenia geologickej mapy na mapovom podklade. Mapový podkladovú musí obsahovať podkladovú mapu sveta. Pri uložení sa požaduje zabezpečiť indexácia priestorového umiestnenia a spolu s popisnými údajmi poskytnúť výsledky vo formáte GeoJSON pre externý systém.
RQ_FP_US_7	64	Organizácia dokumentov Požaduje sa umožniť organizáciu dokumentov v archíve, možnosť užívateľsky vytvárať hierarchické zložky, priečinky a pridávať obsah do priečinkov. Požaduje sa zabezpečiť napojenie na číselníky systému, pričom musí byť zabezpečená aj podpora pre viacnásobný výber z číselníkových hodnôt.
RQ_FP_US_8	65	Presúvanie dokumentov Požaduje sa umožniť presúvanie a zmeny zaradenia dokumentácie v archíve
RQ_FP_US_9	68	Verziónovanie dokumentov a sprístupnenie historických dokumentov Implementovať verziónovanie zmien dokumentov a sprístupnenie historických verzií dokumentu

5.3.6 Reportovacie služby

ID požiadavky	CBA ID	Požiadavka
RQ_FP_RS_1	34	Reportovacie služby zabezpečujú riadený prístup k výstupným zostavám a štatistickým prehľadom údajov Digitálneho archívu verejnosťou a internými pracovníkmi ŠGÚDŠ. Služby sú postavené na technológii Jasper Reports.

ID požiadavky	CBA ID	Požiadavka
		Požaduje sa zabezpečiť všetky existujúce reportovacie služby v migrovanom riešení.
RQ_FP_RS_2	76	Súčasne sa požaduje umožniť tvorbu nových reportov administrátormi objednávateľa.
RQ_FP_RS_3	73	Rešeršné prehľady – zabezpečiť generovanie rešerše podľa bibliografických pravidiel do formátu MS Word podľa zadanych kritérií vo vyhľadávacích formulároch.
RQ_FP_RS_4	35	Katalogizačné štítky – zabezpečiť vytlačenie katalogizačných štítkov podľa knižničných pravidiel ŠGÚDS pre používateľom definované dokumentácie.
RQ_FP_RS_5	36	Odtajnené správy – zabezpečiť generovanie prehľadov geologických správ, ktoré boli v určitom období odtajnené a sprístupnené pre sťahovacie služby.
RQ_FP_RS_6	37	Polročný výkaz a Ročné prírastky – zabezpečiť generovanie komplexnej zostavy, ktorá je generovaná pre potreby nadriadeného štátneho orgánu
RQ_FP_RS_7	38	Štatistické výkazy sťahovaných správ a početností objektov – zabezpečiť generovanie výstupných zostáv, ktoré sprístupňujú analytické údaje o používanosti Digitálneho archívu.
RQ_FP_RS_8	39 77	Report auditového žurnálu – zabezpečiť generovanie prehľadu ktorá sprístupňuje údaje o prístupoch, online výpožičkách a všetkých iných aktivitách, ktoré používatelia vykonávajú v Digitálnom archíve.
RQ_FP_RS_9	-	Business Intelligence – požaduje sa implementovať komponent pre analýzu a operatívny reporting nad údajmi digitálneho archívu formou business intelligence nástroja. Súčasne sa požaduje zabezpečiť vytvorenie dátových sád a analýz dohodnutých v rámci DNR (predpokladaný rozsah do 20 dátových sád a následných analytík).

5.4 Požiadavky na integrácie

ID požiadavky	CBA ID	Požiadavka
RQ_PnI_1	40	Požaduje sa v migrovanom riešení implementovať integračnú platformu pre zabezpečenie požadovaných integrácií. Služby prepojenia s inými IS musia umožniť prenos a smerovanie správ pomocou pravidiel, ktoré implementujú jednotlivé služby. Úlohou tohto komponentu je prepojenie Digitálneho archívu s externými aplikáciami ako aj poskytovanie služieb dovnútra celého systému Digitálneho archívu. Komponent musí umožniť prijať dáta z ľubovoľného zdroja a na základe definovaných pravidiel tieto údaje transformovať a dopraviť do cieľa pre užívateľa.
RQ_PnI_2	41	Požaduje sa zabezpečiť zachovanie automatizovanej dátovej pumpy pre import bibliografických údajov do IS DA zo vstupných importných dávok štandardu PSP balíkov pre: - celé Monografie, kapitoly monografií a metaúdaje

ID požiadavky	CBA ID	Požiadavka
		- celé Časopisy, kapitoly časopisov a ich metaúdaje - celé Správy, jednotlivé strany geologickej správy a ich metaúdaje - geologické mapy a ich metaúdaje.
RQ_PnI_3	-	Požaduje sa sprístupniť otvorené rozhranie OAI-PMH pre web harvesting obsahu archívu, Rozhranie musí byť implementované v štandarde DublinCore v rozsahu: - služba na získanie zoznamu podporovaných typov objektov Digitálneho archívu - služba na zoznam identifikátorov pre daný typ objektu (SPRAVA) - služba na získanie údajov konkrétnej položky na základe identifikátora objektu - služba na vyhľadávanie v metadátach pomocou atribútu from.
RQ_PnI_4	42	Požaduje sa sprístupniť webové služby číselníkov IS DA pre externý informačný systém.
RQ_PnI_5	43	Požaduje sa sprístupniť webové služby pre prehľadávanie metaúdajov dokumentov Digitálneho archívu z externého informačného systému v rozsahu: - vyhľadanie jednoznačnej položky objektu podľa identifikátora. Ako identifikátor slúži hodnota jednoduchého atribútu, ktorá je jednoznačná v rámci - všetkých záznamov danej položky objektu - vyhľadanie záznamov danej položky objektu podľa zadaných atribútov - vyhľadanie záznamov danej položky objektu pomocou fulltextového vyhľadávania (v prílohách) - stiahnutie dokumentu vyhladaného objektu prostredníctvom servlet download.
RQ_PnI_6	44	Požaduje sa zabezpečiť sprístupnenie vydavateľskej činnosti na webovom sídle ŠGÚDŠ. Vydavateľská činnosť bude komponent Digitálneho archívu, ktorý priamo v prostredí webového sídla sprístupní všetky vydania časopisov a všetky články týchto časopisov zoskupených podľa rokov vydania.
RQ_PnI_7	20	Požaduje sa zabezpečiť priame napojenie IS DA na externé informačné systémy v rozsahu: - Informačný systém environmentálne záťaže, kód v MetaIS MŽP - isvs_300, prevádzkovateľ SAŽP. - Priame napojenie na externý IS (REST, WMS a WFS služby) pre súkromný sektor - IS Albert, IS na efektívne vyhodnocovanie rizík v konkrétnom území - Priame napojenie na interné aplikácie Geologického IS v rámci ŠGÚDŠ, isvs_314: - Samostatné aplikácie – Geofyzikálne mapy; Registre geofondu (8 aplikácií), 2D/3D Mapový portál.
RQ_PnI_8	46	Generovanie SIP balíkov a archivácia PSP balíkov Požaduje sa zabezpečiť automatizované generovanie SIP balíkov na základe PSP balíkov generovaným workflow komponentom digitálneho archívu.

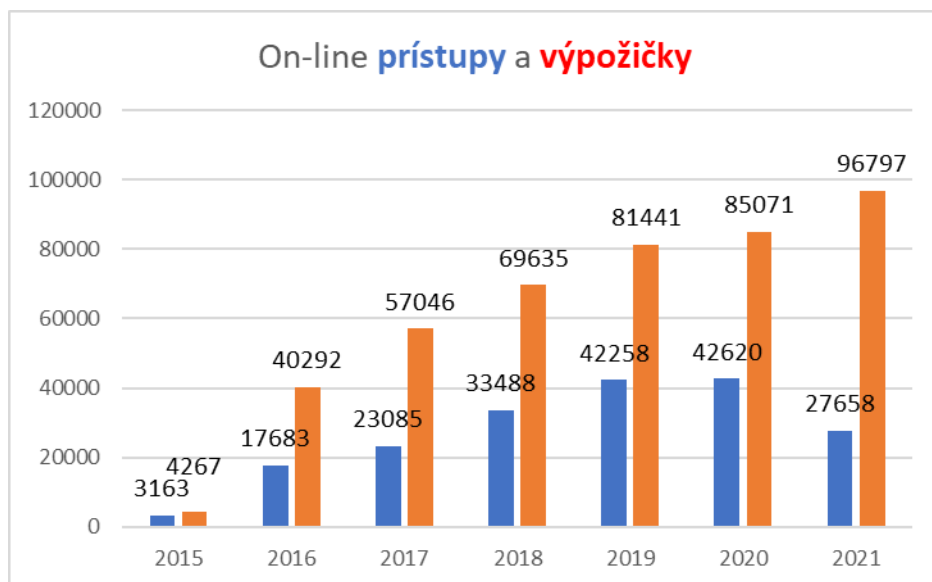
ID požiadavky	CBA ID	Požiadavka
		SIP balíky musia byť generované v štandarde TAR súborového archívu kompatibilného so systémom archivácie kultúrneho dedičstva Univerzitnej knižnice.
RQ_PnI_9	-	Sprístupnenie historických máp a metadát s multijazykovou podporou pre externý systém vo formáte GeoJSON. Požaduje sa zabezpečiť automatické indexovanie metadát a priestorovej lokalizácie historických máp a sprístupnenie vyhľadávacej služby na báze GeoJSON výstupov pre externý systém zverejňovania historických máp.

5.5 Požiadavky na výkonnostné parametre, kapacitné požiadavky

V nasledujúcej tabuľke sa nachádzajú požiadavky na výkonnostné parametre a kapacitné požiadavky:

Parameter	Jednotky	Predpokladaná hodnota	Poznámka
Počet interných používateľov	Počet	14	
Počet súčasne pracujúcich interných používateľov v špičkovom zaťažení	Počet	8	častejšie 4 používatelia súčasne, výnimočne 7-8 používateľov súčasne
Počet externých používateľov (internet)	Počet	350/tisíce	Registrovaný/anonymný
Počet externých používateľov používajúcich systém v špičkovom zaťažení	Počet	Cca 20 registrovaných naraz	
Počet transakcií (podaní, požiadaviek) za obdobie	Počet/rok	Vid' graf pod tabuľkou	Marec 2022 už vyše 12000 výpožičiek
Objem údajov na transakciu	Objem/transakcia	Cca 30/70 TB	30 TB údaje na aktívny prístup, 70TB úložisko
Diskové pole	Objem	Cca 115 TB	Z toho transakčné dáta: Workflow: 10 TB (IS Sirius) Digitálny archív 10 TB Statické dáta: Dátové zálohy primárnych skenov a zálohy Workflow a SIP balíky : 60 TB Aplikačné zálohy: 14TB ŠGÚDŠ (prípravné skeny na nové dávky a interné zálohy, opravy ...): 20 TB
Požadovaný nárast dát - ŠGÚDŠ (prípravné skeny na nové dávky a interné zálohy, opravy ...)	%	50	Statické dáta: ŠGÚDŠ (prípravné skeny na nové dávky a interné zálohy, opravy ...): aktuálne 20 TB – navýšenie o 50%
Ročný nárast objemu dát	%	10	Celkových dát

Tabuľka 1 - Prehľad vybraných kapacitných a výkonových požiadaviek – budúci stav



Obrázok 7 - Počet transakcií (podaní, požiadaviek) za rok

ID požiadavky	CBA ID	Požiadavka
RQ_PnVP_1	47	Požiadavka optimalizovať systém pre vysokú kapacitnú a výkonovú záťaž – práca pri objeme rádovo miliónov dokumentov, pri súčasnom prístupe desiatok používateľov.

5.6 Požiadavky na aplikačnú architektúru

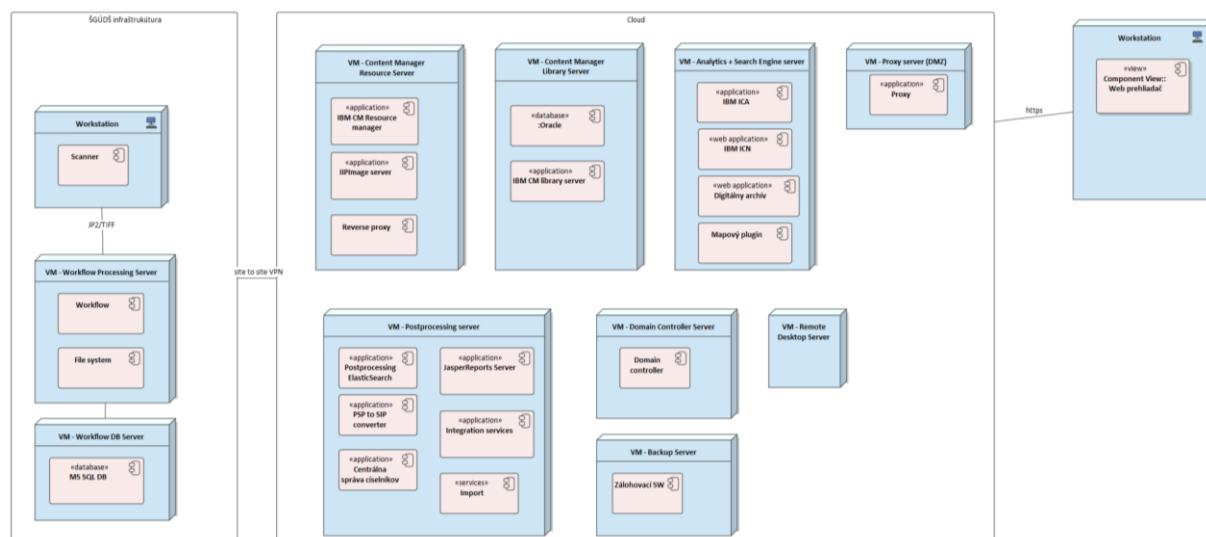
ID požiadavky	ID	Požiadavka
RQ_PnAA_1	-	Verejný obstarávateľ upozorňuje, že v predloženom projekte sa nejedná o presun existujúcich virtuálnych serverov do prostredia cloudu. Všetky technológie bude potrebné nanovo inštalovať na novej virtualizovanej infraštruktúre podľa inštalčných a konfiguračných postupov výrobcu IBM CM riešenia.
RQ_PnAA_2	-	Požaduje sa prispôbiť systém princípom „Cloud-native computing“ – návrh úprav bude analyzovaný v rámci spracovania analýzy a návrhu riešenia projektu tak, aby v optimálnej miere prispôbil systém presunu do cloud prostredia.
RQ_PnAA_3	-	Požaduje sa zachovanie implementovaných verzií jednotlivých aplikácií, kvôli zabezpečeniu plnenia licenčných podmienok a kompatibility aplikácií v rámci celého systému: - IBM Content Manager v8.5.0.0 - IBM Content Navigator v2.0.3.6 - IBM Content Analytics v3.5.0.3 - Modulárny digitalizačný systém Sirius v1.0.88.0

ID požiadavky	ID	Požiadavka
RQ_PnAA_4	-	V rámci migrácie databázového servera do cloudu, sa požaduje zachovanie databázového systému Oracle vo verzii 11.2.0.4., prípadne dodávateľ môže využiť iné dostupné databázové riešenia so zabezpečením kompletnej funkčnosti. Výsledné rozhodnutie bude výstupom analytickej etapy projektu.

5.7 Požiadavky na technologickú architektúru

Od uchádzača sa požaduje v rámci verejného obstarávania vypracovať a odovzdať návrh technologickej architektúry riešenia, ktorý bude vychádzať z nasledovných princípov, prípadne ich optimalizuje v rámci prvej analytickej etapy.

Návrh budúceho riešenia technologickej architektúry sa nachádza na nasledujúcom obrázku.



Obrázok 8 - Budúce riešenie v cloude

ID požiadavky	CBA ID	Požiadavka
RQ_PnTA_1	49	Požiadavka na prácu s obsahom z viacerých serverov Požaduje sa umožniť užívateľom vytvárať vlastné pohľady na obsah webového klienta vytvorením teamspaces, ktoré poskytujú cieľový pohľad na príslušné dokumenty, zložky a vyhľadávanie. Poskytnúť formou API sadu nástrojov, ktoré sa použijú na rozšírenie webového klienta a vytvorenie všetkých špecifických funkcionalít IS DA. Systém zároveň umožní pripájanie sa na viacero údajových zdrojov umiestnených vo viacerých databázach digitálneho archívu
RQ_PnTA_2	-	Z pohľadu aplikačnej vrstvy sa požaduje:

ID požiadavky	CBA ID	Požiadavka
		- scanovanie a následné spracovanie primárnych skenov (workflow proces) prebehne v rámci lokálnej infraštruktúry ŠGÚDŠ (teda sa nebude migrovať do cloudu, vykoná sa inštalácia na novú infraštruktúru v rámci ŠGÚDŠ) - zachovávajú sa súčasné prevádzkované veľkokapacitné skenery, vrátane spôsobu ich súčasnej prevádzky a obsluhy - PSP balíky dokumentov sa budú tvoriť lokálne a následne budú už v prostredí cloud importované do DA a automatizovane zverejňované pre koncových používateľov. - Softvérové riešenie DA bude zmigrované do cloudu a modernizované. - Do cloudu budú zmigrované aj súčasné databázy DA a fyzické dokumenty.

5.8 Požiadavky na využitie služieb cloudu

V súlade s NKIVS by technologická architektúra ISVS mala byť založená na cloudových službách. Projekt predpokladá migráciu do vládneho cloudu, prípadne optimálnejšiu alternatívnu lokalitu spĺňajúcu jednotlivé požiadavky.

ID požiadavky	Požiadavka										
RQ_PnVSC_1	Požaduje sa vytvorenie len produkčného prostredia, nakoľko systém je dostatočne stabilný a nie je tak potrebné permanentné udržiavanie testovacieho a vývojového prostredia.										
RQ_PnVSC_2	V rámci služieb IaaS služieb privátneho cloudu sa požaduje využitie služby 1.1. Virtuálny server s nasledovnými parametrami: <table> <tr> <th>VM</th><th>Popis parametrov služby</th></tr> <tr> <td>Proxy Server</td><td> Popis: Proxy – riešenie bezpečného prístupu verejnej časti riešenia z internet siete Požadované parametre: Linux Ubuntu, 2 vCPU, 8GB RAM, 40GB DISK </td></tr> <tr> <td>Content Manager Library Server</td><td> Popis: DB Oracle – centrálna úložisko dát IBM CM library server – prístup k databáze systému Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 300GB DISK </td></tr> <tr> <td>Content Manager Resource Server</td><td> Aplikácie: IBM CM resource manager – prístup k dátam systému uloženými na file systeme Reverse proxy – riešenie bezpečnosti IIPimage – multiforátový prehliadač formátov dokumentov Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 6000GB DISK </td></tr> <tr> <td>Postprocessing Server</td><td> Popis: PSP to SIP – aplikácia pre generovanie SIP balíkov, Centrálna správa číselníkov – aplikácia pre centrálnu správu číselníkov ElasticSearch – aplikácia pre rýchle vyhľadávanie dát Jasper reports server – nástroj pre tvorbu a sprístupnenie reportov Integration services – sprístupňuje integračné rozhrania systému </td></tr> </table>	VM	Popis parametrov služby	Proxy Server	Popis: Proxy – riešenie bezpečného prístupu verejnej časti riešenia z internet siete Požadované parametre: Linux Ubuntu, 2 vCPU, 8GB RAM, 40GB DISK	Content Manager Library Server	Popis: DB Oracle – centrálna úložisko dát IBM CM library server – prístup k databáze systému Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 300GB DISK	Content Manager Resource Server	Aplikácie: IBM CM resource manager – prístup k dátam systému uloženými na file systeme Reverse proxy – riešenie bezpečnosti IIPimage – multiforátový prehliadač formátov dokumentov Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 6000GB DISK	Postprocessing Server	Popis: PSP to SIP – aplikácia pre generovanie SIP balíkov, Centrálna správa číselníkov – aplikácia pre centrálnu správu číselníkov ElasticSearch – aplikácia pre rýchle vyhľadávanie dát Jasper reports server – nástroj pre tvorbu a sprístupnenie reportov Integration services – sprístupňuje integračné rozhrania systému
VM	Popis parametrov služby										
Proxy Server	Popis: Proxy – riešenie bezpečného prístupu verejnej časti riešenia z internet siete Požadované parametre: Linux Ubuntu, 2 vCPU, 8GB RAM, 40GB DISK										
Content Manager Library Server	Popis: DB Oracle – centrálna úložisko dát IBM CM library server – prístup k databáze systému Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 300GB DISK										
Content Manager Resource Server	Aplikácie: IBM CM resource manager – prístup k dátam systému uloženými na file systeme Reverse proxy – riešenie bezpečnosti IIPimage – multiforátový prehliadač formátov dokumentov Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 6000GB DISK										
Postprocessing Server	Popis: PSP to SIP – aplikácia pre generovanie SIP balíkov, Centrálna správa číselníkov – aplikácia pre centrálnu správu číselníkov ElasticSearch – aplikácia pre rýchle vyhľadávanie dát Jasper reports server – nástroj pre tvorbu a sprístupnenie reportov Integration services – sprístupňuje integračné rozhrania systému										

ID požiadavky	Požiadavka										
	<table> <tr> <td></td><td>Import – automatizované služby importu dát do systému Požadované parametre: Windows server 2016, 8 vCPU, 16GB RAM, 1000GB DISK</td></tr> <tr> <td>Analytics + Search Engine Server</td><td> Popis: IBM ICA – implementácia IBM Watson pre fazetové vyhľadávanie IBM CN – IBM Content Navigator pre sprístupnenie obsahu koncovým používateľom Digitálny archív – aplikácia pre koncových používateľov Mapový plugin – sprístupňuje mapové služby a priestorové vyhľadávanie Požadované parametre: Windows server 2016, 16 vCPU, 32GB RAM, 4000GB DISK </td></tr> <tr> <td>Remote Desktop Server</td><td> Popis: remote desktop – slúži pre vzdialenú správu celého riešenia Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 100GB DISK </td></tr> <tr> <td>Domain Controller Server</td><td> Popis: doménový radič Domain controller a DNS server Požadované parametre: Windows server 2016, 1 vCPU, 4GB RAM, 40GB DISK </td></tr> <tr> <td>Backup Server</td><td> Popis: zálohovanie Zálohovací SW Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 15000GB DISK </td></tr> </table>		Import – automatizované služby importu dát do systému Požadované parametre: Windows server 2016, 8 vCPU, 16GB RAM, 1000GB DISK	Analytics + Search Engine Server	Popis: IBM ICA – implementácia IBM Watson pre fazetové vyhľadávanie IBM CN – IBM Content Navigator pre sprístupnenie obsahu koncovým používateľom Digitálny archív – aplikácia pre koncových používateľov Mapový plugin – sprístupňuje mapové služby a priestorové vyhľadávanie Požadované parametre: Windows server 2016, 16 vCPU, 32GB RAM, 4000GB DISK	Remote Desktop Server	Popis: remote desktop – slúži pre vzdialenú správu celého riešenia Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 100GB DISK	Domain Controller Server	Popis: doménový radič Domain controller a DNS server Požadované parametre: Windows server 2016, 1 vCPU, 4GB RAM, 40GB DISK	Backup Server	Popis: zálohovanie Zálohovací SW Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 15000GB DISK
	Import – automatizované služby importu dát do systému Požadované parametre: Windows server 2016, 8 vCPU, 16GB RAM, 1000GB DISK										
Analytics + Search Engine Server	Popis: IBM ICA – implementácia IBM Watson pre fazetové vyhľadávanie IBM CN – IBM Content Navigator pre sprístupnenie obsahu koncovým používateľom Digitálny archív – aplikácia pre koncových používateľov Mapový plugin – sprístupňuje mapové služby a priestorové vyhľadávanie Požadované parametre: Windows server 2016, 16 vCPU, 32GB RAM, 4000GB DISK										
Remote Desktop Server	Popis: remote desktop – slúži pre vzdialenú správu celého riešenia Požadované parametre: Windows server 2016, 4 vCPU, 16GB RAM, 100GB DISK										
Domain Controller Server	Popis: doménový radič Domain controller a DNS server Požadované parametre: Windows server 2016, 1 vCPU, 4GB RAM, 40GB DISK										
Backup Server	Popis: zálohovanie Zálohovací SW Požadované parametre: Windows server 2016, 4 vCPU, 8GB RAM, 15000GB DISK										
RQ_PnVSC_3	V rámci služieb IaaS služieb privátneho cloudu sa požaduje využitie služby 1.2. Diskový priestor s nasledovnými parametrami: <table> <tr> <th>Služba</th><th>Popis parametrov služby</th></tr> <tr> <td>Tier2 or SSD</td><td>Kapacita: 1000GB</td></tr> <tr> <td>Tier3</td><td>Kapacita: 12000GB</td></tr> <tr> <td>Tier3 a menej</td><td>Kapacita: 15000GB</td></tr> </table>	Služba	Popis parametrov služby	Tier2 or SSD	Kapacita: 1000GB	Tier3	Kapacita: 12000GB	Tier3 a menej	Kapacita: 15000GB		
Služba	Popis parametrov služby										
Tier2 or SSD	Kapacita: 1000GB										
Tier3	Kapacita: 12000GB										
Tier3 a menej	Kapacita: 15000GB										
RQ_PnVSC_4	V rámci služieb IaaS služieb privátneho cloudu sa požaduje využitie služby 1.3. Služba pripojenia do špecifickej siete – pripojenie do siete internet										
RQ_PnVSC_5	V rámci služieb IaaS služieb privátneho cloudu sa požaduje využitie služby 1.4. Sieťové služby: - Vytvorenie preddefinovaného sieťového modelu a základných FW pravidiel - Vytvorenie dodatočných FW pravidiel										

5.9 Požiadavky na fyzickú HW infraštruktúru ŠGÚDŠ

ID požiadavky	Požiadavka						
RQ_PnFWH_1	V existujúcej fyzickej infraštruktúre ŠGÚDŠ sa požaduje zabezpečiť VM s nasledovnými parametrami. UPGRADE ani prípadné rozširovanie fyzickej infraštruktúry nie je predmetom projektu.						
	<table><tr><th>VM</th><th>Popis parametrov služby</th></tr><tr><td>Workflow DB Server</td><td>Popis: databázový server workflow a OCR server Požadované parametre VM: Windows server 2016, 24 vCPU, 16GB RAM, 300GB DISK</td></tr><tr><td>Workflow Processing Server</td><td>Popis: workflow Požadované parametre VM: Windows server 2016, 32 vCPU, 24GB RAM, 10000GB</td></tr></table>	VM	Popis parametrov služby	Workflow DB Server	Popis: databázový server workflow a OCR server Požadované parametre VM: Windows server 2016, 24 vCPU, 16GB RAM, 300GB DISK	Workflow Processing Server	Popis: workflow Požadované parametre VM: Windows server 2016, 32 vCPU, 24GB RAM, 10000GB
	VM	Popis parametrov služby					
Workflow DB Server	Popis: databázový server workflow a OCR server Požadované parametre VM: Windows server 2016, 24 vCPU, 16GB RAM, 300GB DISK						
Workflow Processing Server	Popis: workflow Požadované parametre VM: Windows server 2016, 32 vCPU, 24GB RAM, 10000GB						

5.10 Požiadavky na bezpečnosť

ID požiadavky	CBA ID	Požiadavka
RQ_PnB_1	14	Požaduje sa, aby zmigrovaný IS DA spĺňal bezpečnostné požiadavky, ktoré vyplývajú z Vyhlášky Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné systémy verejnej správy.
RQ_PnB_2	-	Požaduje sa zabezpečiť súlad zmigrovaného IS DA s požiadavkami zákona č. 69/2018 Z. z. o kybernetickej bezpečnosti a o zmene a doplnení niektorých zákonov a prijať bezpečnostné opatrenia, ktorých cieľom je zabezpečenie kybernetickej bezpečnosti počas životného cyklu sietí a systému.
RQ_PnB_3	53	V zmigrovanom IS DA sa požaduje implementovať pravidlá ochrany osobných údajov v súlade so zákonom č. 18/2018 Z. z. o ochrane osobných údajov (GDPR), vrátane riadenia retencie uchovávaní údajov a uplatnenie práva na zabudnutie.
RQ_PnB_4	-	Požaduje sa, aby prístup k údajom bol riadený prostredníctvom funkčných rolí a ich oprávnení vedených v IS DA.
RQ_PnB_5	-	Požaduje sa, aby zhotovené dielo na technologickej úrovni v čo najvyššej možnej miere využívalo existujúce bezpečnostné politiky, komponenty a technológie cloudu, predovšetkým v oblastiach: - Bezpečnosť dátových centier - Bezpečné „Multi-tenancy“ Izolované údaje - DDoS Ochrana údajov - Segregácia Údajov.
RQ_PnB_6	14	Požaduje sa vytvoriť a otestovať šifrované spojenie medzi prostredím ŠGÚDŠ a cloudom. Používatelia a automatizované služby budú komunikovať prostredníctvom zabezpečeného spojenia, ktoré bude spĺňať bezpečnostné štandardy.

ID požiadavky	CBA ID	Požiadavka
RQ_PnB_7	-	Prevádzka systému musí byť zabezpečená prostredníctvom HTTPS s bezpečnostným certifikátom vydaným certifikačnou autoritou.
RQ_PnB_8	-	Požaduje sa automatizované zálohovanie prevádzkových údajov a databáz na dennej báze. Virtuálna serverová architektúra musí byť zálohovaná aspoň raz týždenne s trojmesačnou archivačnou dobou záloh. Zálohy budú ukladané na diskové polia cloudu. Pre automatizované zálohovanie sa požaduje v rámci projektu nasadiť štandardnú zálohovaciu technológiu.
RQ_PnB_9	-	Požaduje sa zabezpečiť detailné logovanie udalostí jednotlivých komponentov DA. Záznam v logu musí obsahovať minimálne informáciu o dátume a čase výskytu udalosti, kategóriu udalosti (informáciu, upozornenie, chybová udalosť) a podrobný opis parametrov udalosti
RQ_PnB_10	52	Požaduje sa, aby zmigrovaný IS mal implementované nástroje pre zaznamenávanie aktivít používateľov systému. Každý takýto záznam musí obsahovať informáciu kedy a kým bola daná aktivita vykonaná. Údaje z auditingu budú dostupné iba oprávneným používateľom.
RQ_PnB_11	-	Od zhotoviteľa sa požaduje spracovať a odovzdať bezpečnostný projekt k zmigrovanému riešeniu.
RQ_PnB_12	-	Od zhotoviteľa sa požaduje vypracovať havarijný plán pre zmigrovaný IS DA, ktorého súčasťou bude definovanie procesov pre prípady technických závad veľkého rozsahu a postupy vedúce k obnove funkcionality systému

5.11 Požiadavky na zdrojový kód

ID požiadavky	CBA ID	Požiadavka
RQ_PnZK_1	-	Pri akceptácii projektu musí byť odovzdané funkčné vývojové a produkčné prostredie, ktoré je súčasťou Informačného systému
RQ_PnZK_2	-	Pri akceptácii Informačného systému alebo jeho časti bude potrebné odovzdať vytvorený zdrojový kód v jeho úplnej aktuálnej podobe, zapečatený, na neprepisovateľnom technickom nosiči dát s označením časti a verzie Informačného systému, ktorej sa týka. Za odovzdanie Vytvoreného zdrojového kódu sa rozumie odovzdanie technického nosiča dát Oprávnenej osobe. O odovzdaní a prevzatí technického nosiča dát bude oboma stranami spísaný a podpísaný preberací protokol.
RQ_PnZK_3	-	Vytvorený zdrojový kód Informačného systému vrátane jeho dokumentácie bude prístupný v režime podľa § 31 ods. 4 písm. b) Vyhlášky č. 78/2020 (s obmedzenou dostupnosťou pre orgán vedenia a orgány riadenia v zmysle Zákona o ITVS – vytvorený zdrojový kód je dostupný len pre orgán vedenia a orgány riadenia)

ID požiadavky	CBA ID	Požiadavka
RQ_PnZK_4	-	Vytvorený zdrojový kód musí byť v podobe, ktorá zaručuje možnosť overenia, že je kompletný a v správnej verzii, t. j. v takej, ktorá umožňuje kompiláciu, inštaláciu, spustenie a overenie funkcionality

5.12 Požiadavky na testovanie

ID požiadavky	ID	Požiadavka
RQ_PnT_1	17	Od zhotoviteľa sa požaduje zrealizovať komplexné pretestovanie splnenia funkčných požiadaviek.
RQ_PnT_2	18	Od zhotoviteľa sa požaduje zrealizovať testovanie funkčnosti VM a sieťových prestupov definovaných v DNR.
RQ_PnT_3	19	Od zhotoviteľa sa požaduje zrealizovať integračné testovanie pre potvrdenie funkčnosti všetkých migrovaných a implementovaných integračných rozhraní.
RQ_PnT_4	-	Od zhotoviteľa sa požaduje zrealizovať záťažové a výkonnostné testovanie jednotlivých komponentov DA.
RQ_PnT_5	-	Od zhotoviteľa sa požaduje zrealizovať bezpečnostné testovanie.
RQ_PnT_6	-	Od zhotoviteľa sa požaduje zrealizovať UX testovanie a UAT testovanie.

5.13 Všeobecné požiadavky

ID požiadavky	ID	Požiadavka
RQ_VP_1	24	Riadenie projektu musí byť v súlade s vyhláškou 85/2020.
RQ_VP_2	-	Od zhotoviteľa sa požaduje realizovať predmet plnenia po etapách, podľa nasledovného časového harmonogramu aktivít, zodpovedajúcich realizácii projektu ako celku: - Analýza a dizajn - Implementácia a testovanie - Nasadenie.
RQ_VP_3	-	Od zhotoviteľa sa požaduje ukončiť všetky etapy a aktivity projektu do 31.12.2023. Za ukončenie realizačnej fázy predmetu zákazky sa rozumie situácia, kedy zadávateľ prevezme a akceptuje predmet projektu od zhotoviteľa preberacím a akceptačným protokolom. Pri akceptácii budú vyhotovované vopred definované akceptačné kritériá a požiadavky z katalógu funkčných a nefunkčných požiadaviek vzťahujúce sa k migrácii Digitálneho archívu.
RQ_VP_4	-	Od zhotoviteľa sa požaduje vypracovať a odovzdať dokumentáciu k projektu migrácie DA v rozsahu:

ID požiadavky	ID	Požiadavka
		- dokumentácia projektového riadenia v zmysle štandardu pre riadenie informačno-technologických projektov - detailný návrh riešenia (DNR) - plán testovania a výsledky z testovania: Od zhotoviteľa sa požaduje zdokumentovať výsledky všetkých testov, ktoré zhotoviteľ vykoná, resp. sú požadované. - školiaca dokumentácia - používateľská príručka: požaduje sa, aby zhotoviteľ v rámci projektu vytvoril a dodal používateľskú príručku, ktorá musí obsahovať popis všetkých funkcionality informačného systému. - prevádzková dokumentácia: požaduje sa dodať inštalačnú, konfiguračnú a prevádzkovú dokumentáciu, ktorá musí obsahovať popis inštalácie, konfigurácie, zálohovania, chybových stavov, popis logovania, administrátorských prístupov ako aj podrobný popis integračných rozhraní vrátane ich konfigurácie, ako aj konfigurácie celého informačného systému - bezpečnostný projekt - havarijný plán
RQ_VP_5	-	Od zhotoviteľa sa požaduje realizovať školenia pre kľúčových používateľov DA v počte do 10 osôb, minimálne v rozsahu: - školenie administrátorov systému zamerané na inštaláciu a správu jednotlivých komponentov DA v prostredí cloudu pre zabezpečenie prevádzky informačného systému. Predpokladaný rozsah je 2 dni. - školenie kľúčových používateľov zamerané na používanie zmigrovaného systému. Predpokladaný rozsah je 2 dni.