

Licenčná zmluva na počítačový program

uzatvorená podľa ustanovenia § 65 a nasl. zákona č. 185/2015 Z. z. Autorský zákon (ďalej len „Autorský zákon“)

Objednávateľ

Obchodné meno: **Univerzitná knižnica v Bratislave**
Sídlo: Michalská 1, 814 17 Bratislava
IČO: 00 164 631
IČ DPH: SK202082990
Rozpočtová organizácia v zriaďovateľskej pôsobnosti Ministerstva kultúry Slovenskej republiky
Konajúci: Ing. Silvia Stasselová, generálna riaditeľka
(ďalej len „Nadobúdateľ“)

a

Poskytovateľ

Obchodné meno: **TEMPEST a.s.**
Sídlo: Krasovského 14, 851 01 Bratislava
IČO: 31 326 650
IČ DPH: SK2020327716
Zápis: Obchodný register Okresného súdu Bratislava I, odd. Sa, vl. č. 3771/B
Konajúci: Ing. Andrej Bališ, člen predstavenstva
(ďalej len „Poskytovateľ“)

I. Úvodné ustanovenia

1. Dňa 25.11.2013 bola uzatvorená Rámcová dohoda ev. č. MK-82/2013/M časť 2: Servery, dátové úložiská, sieťové komponenty, ochranné komponenty sietí s príslušenstvom, perifériami, súvisiaci SW a súvisiace služby medzi Ministerstvom kultúry Slovenskej republiky, so sídlom Námestie SNP 33, 813 31 Bratislava, IČO: 00 165 182 ako objednávateľom a centrálnym obstarávateľom pre organizácie v jeho zriaďovateľskej pôsobnosti a spoločnosťou SOITRON, s.r.o., so sídlom Plynárenská 5, 829 75 Bratislava, IČO: 35 955 678 ako dodávateľom (ďalej len „Dodávateľ“ a „Rámcová dohoda“).
2. Dodávateľ na základe osobitného zmluvného vzťahu, objednávky zo dňa 25.09.2015 (ďalej len „Objednávka“), objednal u Poskytovateľa realizáciu informačného systému Digitálne pramene, pričom špecifikácia informačného systému sa nachádza v Prilohe č. 1 tejto Zmluvy (ďalej len „Vytvorený IS“). Vytvorený IS bol vytvorený ako zamestnanecké dielo v zmysle ustanovenia §90 Autorského zákona, ku ktorému vykonáva majetkové práva autora Poskytovateľ. Vytvorený IS Poskytovateľ dňa 04.12.2015 odovzdal Dodávateľovi, o čom strany spísali akceptačný protokol. Dodávateľ následne Nadobúdateľovi odovzdal Vytvorený IS dňa 4.12.2015, o čom strany spísali akceptačný protokol.

3. Zmluvné strany uzatvárajú túto zmluvu za účelom udelenia práva užívania Vytvoreného IS Poskytovateľom na Nadobúdateľa. Zmluvné strany vyhlasujú, že užívanie Vytvoreného IS Nadobúdateľom v čase od jeho dodania (odovzdania Dodávateľom Nadobúdateľovi) do účinnosti tejto zmluvy bolo vykonávané so súhlasom Poskytovateľa a teda bolo oprávneným užívaním Vytvoreného IS Nadobúdateľom bez nároku Poskytovateľa na úhradu akejkoľvek odplaty voči Nadobúdateľovi za toto obdobie užívania.

II. Predmet zmluvy

1. Predmetom tejto zmluvy je udelenie súhlasu Poskytovateľom Nadobúdateľovi na použitie Vytvoreného IS, a to v rozsahu a za podmienok uvedených v tejto zmluve (ďalej len „Licencia“).
2. Predmetom tejto zmluvy je aj odovzdanie komentovanej verzie zdrojového kódu k Vytvorenému IS Nadobúdateľovi.

III. Licencia

1. Poskytovateľ poskytuje na základe tejto zmluvy Nadobúdateľovi nevýhradnú licenciu na používanie Vytvoreného IS a to na účely na ktoré bol vytvorený.
2. Poskytovateľ udeľuje Nadobúdateľovi územne a vecne neobmedzenú Licenciu k Vytvorenému IS a to na dobu trvania majetkových práv Poskytovateľa ako autora Vytvoreného IS.
3. Poskytovateľ poskytuje Nadobúdateľovi Licenciu k Vytvorenému IS bezodplatne.
4. Nadobúdateľ je oprávnený s Vytvoreným IS nakladať akýmkoľvek spôsobom, a to už či sám alebo prostredníctvom tretej osoby, a to najmä:
 - a. dielo upraviť alebo dať upraviť;
 - b. rozmnožovať dielo alebo dať rozmnožovať technickými prostriedkami zvolenými podľa jeho uváženia;
 - c. rozširovať dielo akýmkoľvek spôsobom (napr. verejným rozširovaním diela, jeho časti alebo rozmnoženiny, nájmom, vypožičiavaním alebo akoukoľvek formou prevodu vlastníckych práv alebo práva držby, verejným vystavením diela alebo jeho rozmnoženiny, verejným vykonaním diela, vysielaním, prípadne jeho káblovou retransmisiou alebo iným spôsobom prenosu);
 - d. dielo preložiť alebo dať preložiť;
 - e. adaptovať dielo, usporiadať alebo inak spracovať alebo dať adaptovať, usporiadať alebo inak spracovať;
 - f. použiť a šíriť dielo na akýchkoľvek nosičoch;
 - g. spojiť dielo s iným dielom;
 - h. prihlásiť dielo ako ochrannú známku.
5. Licencia prechádza na právnych nástupcov Nadobúdateľa.
6. Nadobúdateľ je oprávnený Licenciu postúpiť na akúkoľvek tretiu osobu alebo subjekt za predpokladu, že takáto tretia osoba alebo subjekt spĺňa definíčné znaky orgánu riadenia podľa zákona č. 95/2019 Z. z. o informačných technológiách vo verejnej správe a o zmene a doplnení

niektorých zákonov, pričom definícia orgánu riadenia z uvedeného právneho predpisu sa použije aj v prípade, ak by bol tento právny predpis nahradený, zmenený alebo zrušený.

7. Nadobúdateľ je oprávnený v rozsahu udelenej Licencie poskytnúť k Vytvorenému IS sublicenciu osobám, ktoré spĺňajú definičné znaky orgánu riadenia podľa bodu 6 tohto článku tejto zmluvy.

IV. Platnosť a účinnosť zmluvy

1. Táto zmluva je platná dňom jej podpísania oprávnenými zástupcami zmluvných strán.
2. Táto zmluva je účinná dňom nasledujúcim po jej povinnom zverejnení v zmysle ustanovenia §5a zákona č. 211/2000 Z. z. o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov (zákon o slobode informácií) v spojení s ustanovením §47a zákona č. 40/1964 Zb. Občiansky zákonník, ak je povinne zverejňovanou zmluvou, inak táto zmluva nadobúda účinnosť dňom jej platnosti.

V. Záverečné ustanovenia

1. V prípade, že ktorékoľvek z ustanovení tejto zmluvy je alebo sa stane neplatným, resp. neúčinným, táto skutočnosť nemá vplyv na platnosť a/alebo účinnosť ostatných ustanovení tejto zmluvy.
2. Táto zmluva a práva a povinnosti z nej vyplývajúce alebo s ňou súvisiace sa riadia právnym poriadkom Slovenskej republiky.
3. Zmluvné strany sa dohodli, že túto zmluvu je možné meniť a doplniť výlučne na základe písomnej dohody medzi jej stranami, ktorá bude v podobe číslovaného dodatku, ktorý podpíšu obidve zmluvné strany.
4. Táto zmluva sa vyhotovuje v dvoch rovnopisoch v slovenskom jazyku, pričom každá zo zmluvných strán obdrží po jednom vyhotovení.
5. Zmluvné strany vyhlasujú, že ich spôsobilosť na právne úkony nie je obmedzená, že túto zmluvu uzatvorili na základe svojej slobodnej vôle, vážne a že táto zmluva nebola uzatvorená pod nátlakom alebo za nápadne nevýhodných podmienok. Zmluvné strany spoločne vyhlasujú, že si túto zmluvu prečítali a že jej obsahu porozumeli a na znak súhlasu ju vlastnoručne podpísali.

V Bratislave, dň

V Bratislave

Ing. Andrej Bali
člen predstavenstva
TEMPEST a.s.

Ing. Silvia Stasselová
generálna riaditeľka
Univerzitná knižnica

1 VŠEOBECNÝ POPIS SYSTÉMU A JEHO FUNKCIE

IS Digitálne pramene je komplexný informačný systém. Medzi jeho hlavné funkcie patria záznam, zbieranie, sprístupňovanie a ochrana pôvodného digitálneho obsahu, webharvesting a webarchiving online digitálnych prameňov. Systém zbiera, ukladá, sprístupňuje a spravuje pôvodné (e-Born) elektronické pramene ako sú elektronické monografie a periodiká, webové stránky a ďalšie druhy elektronických publikácií a artefaktov, vrátane elektronických tvorivých a publikačných modelov a predlôh.

Sprístupňovanie pôvodných elektronických a WWW prameňov je realizované formou otvorenej digitálnej knižnice a webových služieb s konsolidovaným rozhraním (online repozitár, dostupný verejnosti), ktorá je prepojená na systém dlhodobej ochrany kultúrneho dedičstva (CDA).

Systém je koncipovaný ako integrovaný IS navzájom automaticky spolupracujúcich a komunikujúcich autonómnych funkčných modulov (serverov).

Moduly sú navzájom harmonizované a komunikujú štandardnými rozhraniami a protokolmi (webservises, OAI-PMH). Moduly, ktoré vyžadujú interakciu používateľa majú implementované GUI rozhrania na ich obsluhu. Tieto GUI rozhrania predstavujú verejne prístupný portál pre vyhľadávanie archivovaného digitálneho obsahu návštevníkmi, dedikovanú sekciu verejného portálu určenú len registrovaným vydavateľom a webovú intranetovú aplikáciu pre kurátorov digitálneho obsahu.

Ako fyzické úložisko je použitý zdieľaný súborový systém, na ktorý sa ukladajú výsledky WWW zberov v tvare WARC súborov. Ostatné dáta (metadáta) sú uložené v PostgreSQL databáze.

Zastrešujúce GUI je napísané v jazyku Java vo forme webovej aplikácie. Aplikácia beží na aplikačnom serveri a volá externé utility tretích strán - Heritrix, OpenWayBack.

Súčasťou celého riešenia sú aj headless moduly na integráciu s CDA (generovanie SIP balíkov) a dve služby OAI-PMH.

OBRAZOK 2 – KOMPONENT DIAGRAM – OBRAZOVKY, ULOŽISKÁ A SYSTÉMOVÉ MODULY IS DIGITÁLNE PRAMENE

2.2 MODULY

Zoznam logických funkčných modulov:

Logický modul	Implementácia	Popis
Databáza domén	PostgreSQL	Obsahuje zoznam všetkých sledovaných domén (nie len z .sk priestoru) a základné atribúty o nich. Je východiskom pre plánovanie periodického a tematického zberu. Databáza domén je chápaná ako headless komponent, s ktorým používatelia pracujú pomocou modulu Kurátor . Táto databáza je samozrejme využívaná technickými procesmi.
Úložisko WARC a newebového obsahu	Súborový systém	Zdieľaný súborový systém, na ktorý sa budú ukladať súbory WARC vytvorené Harvesterom .
Harvester	Heritrix + Deduplicator	Slúži na zber a archiváciu vybraných webových stránok v podobe súborov WARC. Súčasťou je aj riešenie (add-on) na deduplikáciu ukladaného obsahu. Za logickú súčasť Harvestera považujeme aj ukladanie newebového obsahu vkladanej cez Portál Pôvodcom newebového obsahu.
Prehliadač WARC	OpenWayBack	Slúži na zobrazenie obsahu uloženého vo WARC súboroch.
Metadátový index	PostgreSQL	V procese harvestingu sa vytvára aj metadátový index pre potreby následného vyhľadania. Metadát sa extrahuje hlavne z html / head sekcia a podľa potreby sa kombinujú s metadátami z Doménovej databázy . Tento modul je chápaný ako headless komponent, s ktorým používatelia pracujú pomocou modulu Portál .
Portál	Wordpress, Java	Portál má verejnú a privátnu časť. Vo verejnej časti vytvárané pomocou CMS sú publikované a udržiavané relevantné informácie o projekte. Verejná časť obsahuje aj GUI na vyhľadanie metódou full-text alebo kombinácie metadát. Privátna časť slúži okrem iného aj autorizovaným Pôvodcom newebového obsahu na upload newebového obsahu.
Integrácia s CDA	Java	Tento funkčný celok zabezpečuje zabalenie WARC a metadát do SIP a ich vklad do CDA. Predpokladá sa online vklad s prípadnou explicitnou žiadosťou o vklad.
Integrácia s SK-NIC	PostgreSQL + bash	Tento modul zabezpečuje periodické načítanie zoznamu slovenských domén z www.sknic.sk, zistenie ich reálneho funkčného stavu a aktualizáciu ich záznamu v Databáze domén .
Modul Kurátor	Java	Tento modul slúži na prácu nad Databázou domén a na plánovanie zberu webového obsahu. Prístup je pomocou GUI, vykonané zmeny sa zapisujú do Databázy domén a riadiacich súborov Harvestera .
Modul Invenio	Invenio	Tento modul obsahuje alternatívnu evidenciu záznamov o doménach a uchovávanom newebovom obsahu a to vo forme opisných metadátami v štandarde MARC 21. V module je možné vykonávať štandardné knižničné operácie nad uloženými záznamami.
Modul OAI-PMH	Java	Tento modul vystavuje metadáta z Databázy domén pre potreby ich zberu tretími stranami.
Modul Auditing	Java	Tento modul centralizuje spracovanie audit záznamov.
Infraštruktúrne aplikácie	Rôzne podporné aplikácie	Tento logický komponent tvoria rôzne podporné infraštruktúrne aplikácie typu Monitoring, Antivirus a pod.

3 APLIKAČNÝ DIZAJN

Operačným systémom na všetkých serveroch IS Digitálne pramene je Linux OS.

3.1 RELAČNÁ DATABÁZA POSTGRESQL

PostgreSQL je voľne šíriteľný objektovo-relačný databázový systém, šírený pod flexibilnou licenciou BSD. Ponúka alternatívu k ostatným proprietárnym ako aj voľne šíriteľným databázovým systémom.

V IS Digitálne pramene slúži na perzistenciu opisných údajov o digitálnych objektoch ako aj na uloženie riadiacich a stavových dát.

3.2 APACHE

Apache HTTP Server je webový server s open-source licenciou pre Linux, BSD, Microsoft Windows a iné platformy. V dnešnej dobe je najrozšírenejším webovým serverom na celom svete a pokladá sa aj za najstabilnejšiu alternatívu.

3.3 APACHE TOMCAT

Apache Tomcat je open-source webový server a kontajner serverov vyvinutý Apache Software Foundation (ASF). Tomcat implementuje viacero Java EE špecifikácií vrátane Java Servlet, JavaServer Pages (JSP), Java EL a WebSocket, and poskytuje prostredie "pure Java" HTTP pre spustiteľný Java kód.

3.4 SOLR

Solr je open-source platforma pre vyhľadávanie vytvorená v programovacom jazyku Java v rámci projektu Apache Lucene. Jeho najvýznamnejšími vlastnosťami sú full-text vyhľadávanie, hit highlighting, faceted search, real-time indexing, dynamic clustering, database integration, NoSQL a spracovanie rich-text dokumentov (napr. Word, PDF). Solr podporuje distribuované vyhľadávanie a replikáciu indexov, Solr je veľmi dobre škálovateľné a odolné voči poruchám (fault tolerant).

3.5 HERITRIX

Heritrix je aplikácia na prieskum a zber webového obsahu (web crawler) využívaná na jeho archivovanie. Bola vytvorená skupinou Internet Archive v jazyku Java. Je voľne šírená bez licenčných poplatkov. Hlavné rozhranie je aplikácie je dostupné cez GUI, pričom sú poskytované nástroje na riadenie cez príkazový riadok (CLI). Riadenie úloh a konfigurácie sa uskutočňuje zmenou XML konfiguračných súborov.

IS Digitálne pramene obsahujú štandardnú inštaláciu (vždy najaktuálnejšiu verziu) a na mieru konfiguráciu základných parametrov vzhľadom na vysoký (ale škálovateľný) výkon systému. Je implementovaná aj priama integrácia s modulom plánovania www zberov, správy šablón www zberov a modulom konfigurácie vlastných zberových politík. Tieto funkcie sú vzhľadom na svoju vysokú komplexnosť a odbornosť prístupné len správcovi systému a www kurátorovi.

3.6 OPENWAYBACK

OpenWayback je open-source projekt a aplikácia určená na zobrazovanie archivovaného webového obsahu pomocou bežného prehliadača. OpenWayback je podporovaný členmi konzorcia International Internet Preservation Consortium (IIPC) a stala sa de-facto štandardom pre zobrazovanie uložených WARC súborov vytvorených napr. aplikáciou Heritrix. OpenWayback má jednoduché webové GUI na vyhľadanie archivovanej stránky podľa URL a dátumu.

IS Digitálne pramene obsahujú štandardnú inštaláciu, aplikácia je však integrovaná a úzko prepojená na verejný portál (vlastné Java komponenty) a preto jej konfigurácia je dostupná len správcovi systému.

3.7 INVENIO

Invenio je voľne šírená aplikácia umožňujúca vytváranie vlastných katalogizačných záznamov o publikáciách a dokumentoch. Aplikácia podporuje mnohé aspekty vedenia knižničných záznamov vrátane klasifikácie, indexovania a práce kurátora. Invenio podporuje protokol OAI-PMH a používa štandard MARC 21 ako formát na uchovávanie opisných metadát.

Aplikácia funguje nezávisle na ostatných komponentoch systému. Na jej katalóg je však integrovaná intranetová aplikácia pre e-Born kurátorov, ktorá automaticky preberá z Invenia záznamy. Ku konfigurácii tohto pripojenia má prístup len administrátor systému. Integrácia je realizovaná v programovacou jazyku Java.

3.8 VLASTNÉ APLIKAČNÉ PROGRAMOVÉ VYBAVENIE

Špecificky celoplošný harvesting všetkých slovenských domén na základe ich zoznamu z SK-NIC si vyžaduje špecifickú funkcionality pre import zoznamu domén, verifikáciu ich stavu, extrakciu prvotných metadát, atď. Túto funkcionality zabezpečuje na mieru vyvinutá aplikácia, ktorá je schopná automatickej interakcie so zvyšnými komponentami na zber, spracovanie a sprístupňovanie digitálneho obsahu (Heritrix, OpenWayback).

Ďalším komponentom vyvinutým na mieru je automatizovaný proces vytvárania a odosielania SIP balíčkov do Centrálného dátového archívu (CDA) za účelom dlhodobej ochrany kultúrneho dedičstva SR. V procese je využitý pokročilý spôsob referencovania jednotlivých SIP balíčkov navzájom, vzťahujúcich sa na rovnaký kultúrny objekt (KO) a to prostredníctvom vytvárania ich verzií, edícií a derivácií.

Komponenty vyvinuté na mieru boli implementované v programovacou jazyku Java a bežia na aplikačnom serveri Apache Tomcat. Využívajú databázu PostgreSQL na ukladanie riadiacich a stavových dát, ale aj na ukladanie popisných údajov o digitálnych objektoch, s ktorými pracujú.

Ovládanie a správa týchto komponentov je pomocou webového GUI, ktoré je ale pre vysokú komplexnosť ich funkcionality a nutnú odbornú znalosť významu jednotlivých riadiacich parametrov prístupné len správcovi systému a vybraným kurátorom.

4 KATALÓG OBJEKTOV

Katalóg objektov IS Digitálne pramene je meta-model založený na **MARC21** štruktúre tagov a poli. Tento katalóg tvoria dokopy 3 zdroje metadát:

- Archív WWW (katalóg WARC/zberov)
- Katalóg WWW (katalóg "živých" .sk domén)
- Katalóg e-Born (katalóg uložených dokumentov)

Každý z týchto metadátových zdrojov má vlastný typ základnej entity, na ktorú sa viaže sada metadát a každá entita má jedinečný identifikátor.

4.1 ARCHÍV WWW

- základnou entitou, na ktorú sa viaže metadátová sada je **1 zber 1 webového sídla**
- pre zber je kľúčovým parametrom dátum zberu
- pre webové sídlo je kľúčová URL
- jedinečným identifikátorom entity je URN NBN - URN:NBN sa generuje pri zbere

4.2 KATALÓG WWW

- základnou entitou, na ktorú sa viaže metadátová sada je **webové sídlo**
- pre webové sídlo je kľúčová URL
- jedinečným identifikátorom entity je URN NBN

4.3 KATALÓG E-BORN

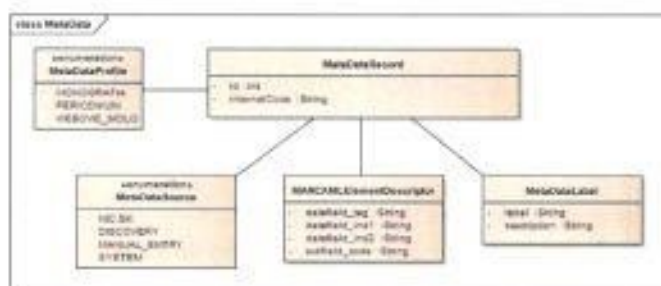
- základnou entitou, na ktorú sa viaže metadátová sada je **e-Born prameň**
- pod e-Born prameňom rozumieme:
 - e-Born monografiu
 - e-Born periodikum - titul
 - e-Born periodikum - ročník
 - e-Born periodikum - číslo
 - e-Born vo forme aktualizovanej www stránky
- jedinečným identifikátorom entity je URN NBN

4.4 DOMÉNOVÝ MODEL METADÁT

Metadáta sú zoskupené do setov (MetaDataSet), kde ku každému profilu je priradený súbor metadát. Metadáta sú identifikované buď cez jedinečný internalCode alebo cez kombináciu päťice metadataprofil+datafield_tag+datafield_ind1+datafield_ind2+subfield_code.

MetaDataRecord.label and MetaDataRecord.description slúži pre human readable zobrazenie názvu a popisu meta tagu (napr. label: Kód predmetovej kategórie, description: Klasifikačný znak konspekt). Na GUI bude názov metadátovej položky zobrazený ako MetaDataRecord.label

Jednotlivé MARCXML podpolia môžu mať tiež priradené label a description, s ktorým sa "prekryje" label a description z entity MetadataRecord. Ak MARCXML podpole má uvedený label, alebo description, tak sa na GUI zobrazí jeho hodnota.



OBRAZOK 3 – DOMÉNOVÝ MODEL METADÁT (KATALÓG)

4.4.1 PRAVIDLÁ PRE EXPORT METADÁT

- exportuje sa len plný model (všetko)

4.4.2 PRAVIDLÁ PRE IMPORT ZÁZNAMOV/URL DO KATALÓGU WWW

- import je formou INSERT záznamov
- ak sa nájde zhoda v URL, tak sa záznam neimportuje
- niektoré vybrané polia sú zakázané, tzn. import ich prečíta, ale nevloží

4.4.3 PRAVIDLÁ PRE AKTUALIZÁCIU METADÁT IMPORTOM ALEBO HARVESTOM OAI-PMH INVENIO

- import je formou REPLACE záznamov
- ak sa nájde zhoda v identifikátore, tak sa záznam prepíše
- ak sa nenájde identifikátor, tak sa záznam preskočí (neimportuje)
- niektoré vybrané polia sú zakázané, tzn. import ich prečíta, ale nedovolí zmenu

4.4.4 PRAVIDLÁ PRE EDITÁCIU METADÁT NA INTRA PORTÁLI

- je možné editovať metadáta 1 domény v Katalógu WWW
- je možné editovať metadáta 1 zberu 1 domény v Archive WWW
- editácia konkrétnych MARC polí (číselný kód)
- niektoré vybrané polia sú read-only bez možnosti editácie/zmeny
- ak je pole len editovateľné môže sa hodnota vyskytovať iba raz
- ak je pole možné pridať alebo odstrániť môže sa hodnota vyskytovať viackrát
- niektoré vybrané polia sa vyplňajú cez combobox/číselník (napr. Konspekt)

5 VEREJNÝ PORTÁL

5.1 LOKALITA PORTÁLU

Portál pre verejnosť je prístupný z Internetu na URL adrese: <http://webdepozit.sk>.

5.2 HLAVNÁ STRÁNKA PORTÁLU

Stránka pozostáva z nasledujúcich hlavných prvkov:

- Menu - Poskytuje používateľovi štruktúrovanú a prehľadnú navigáciu v obsahu portálu.
- Linky na návrhové formuláre - Otvorenie formulára pre zaevidovanie návrhu na zber webového sídla, resp. vklad elektronickej publikácie (e-Born).
- Vyhľadávacie pole - Rýchle a jednoduché prehľadávanie obsahu archívu webových prameňov, resp. elektronických publikácií.
- Informačná časť - Textové a grafické informácie.



OBRÁZOK 4 – HLAVNÁ STRÁNKA VEREJNÉHO PORTÁLU IS DIGITÁLNE PRAMENE

5.3 VYHĽADÁVANIE NA VEREJNOM PORTÁLI

S cieľom dosiahnuť používateľskú prívetivosť a intuitívne použitie vyhľadávania v archívoch, na hlavnej stránke má používateľ k dispozícii jediné textové pole pre zadanie vyhľadávaného výrazu.

Pod poľom je umiestnený prepínač, ktorým používateľ určí, v ktorom z archívov sa má vyhľadávať:

- archív webového obsahu,
- archív elektronických publikácií.

Do poľa môže používateľ zadať niekoľko slov, ktorých výskyt sa vyhľadáva v metadátach archivovaných entít. Vzťah medzi jednotlivými výrazmi je 'alebo', pričom akceptované sú iba slová s dĺžkou 3 a viac znakov.

Výsledky vyhľadávania sú zobrazené na stránke otvorenej v novej záložke webového prehliadača, v stránkovanom zozname.

5.3.1 VÝSLEDOK VYHĽADÁVANIA VO WEB ARCHÍVE

Položky výsledkov obsahujú náhľad stránky v čase zberu a základné informácie (metadáta) o nájdennej entite. V prípade vyhľadania viacerých zberov jednej domény sa zobrazia všetky dostupné verzie s príslušným časom.

Používateľ kliknutím na položku v zozname výsledkov otvorí, vo webovej aplikácii OpenWayback, archivovanú verziu sídla. Pomocou kalendára alebo po kliknutí na počet záznamov v ovládacom paneli sa dostaneme ku všetkým zozbieraným verziám danej domény v čase.

5.3.2 VÝSLEDOK VYHĽADÁVANIA V E-BORN ARCHÍVE

Ako výsledok vyhľadávania v e-Born archíve sa zobrazí zoznam e-Born prameňov so základnými informáciami (metadátami) o nájdennej entite.

6 INTRANET

6.1 PRIHLÁSENIE POUŽÍVATEĽA

Prístup do používateľského rozhrania interného portálu podlieha štandardnému mechanizmu prihlasovania. Po úspešnom prihlásení sú používateľovi pridelené prístupy k funkciám, na ktoré má oprávnenie.

6.2 ÚVODNÁ OBRAZOVKA

Po úspešnom prihlásení sa používateľovi zobrazí úvodná obrazovka umožňujúca navigáciu k jednotlivým funkciám portálu, ku ktorým má používateľ povolený prístup.



OBRÁZOK 5 – ÚVODNÁ OBRAZOVKA INTRANETOVÉHO PORTÁLU IS DIGITÁLNE PRAMENE

6.3 ADMINISTRÁCIA ZBEROV WWW

6.3.1 PREDLOHY PRE ZBER

Táto funkcia umožňuje prezerať predlohy zberov. Predlohy definujú, akým spôsobom sa má uskutočniť alebo periodicky uskutočňovať zber vybranej skupiny domén z *Katalógu*.

Zobrazený zoznam je možné filtrovať podľa:

- názvu predlohy,
- typu predlohy,
- plánovaného spustenia,
- plánovaného ukončenia,
- najbližšieho spustenia.

Pri predlohe je evidovaný aj počet spustení, teda koľkokrát bol vykonaný zber podľa tejto predlohy.

6.3.2 PREHĽAD ZBEROV PREDLOHY

Na tejto stránke sa zobrazuje história zberov vykonaných na základe danej predlohy. Pomocou tlačidla "Report zberov" je možné stiahnuť report zberov za posledných 30 dní.

6.3.3 DETAIL ZBERU DOMÉN

Stránka zobrazuje sumárne informácie ohľadom zberu spolu s tabuľkou jednotlivých zberov domén, ktoré boli v danom zbere zbierané.

6.3.4 ZBER DOMÉNY

Kliknutím na link *Detail* v poslednom stĺpci tabuľky sa zobrazia detailné informácie týkajúce sa príslušného zberu domény.

6.3.5 PREHĽAD ZBEROV

Prehľad zberov umožňuje sledovať všetky uskutočnené zbery domén.

Zobrazený zoznam je možné filtrovať podľa:

- typu predlohy,
- stavu zberu,
- spustenia zberu,
- ukončenia zberu.

Doplňujúcou informáciou pre kurátora je objem zozbieraných dát, ako aj počet úspešne/neúspešne zozbieraných domén.

6.3.6 NASTAVENIE GLOBÁLNEJ KONFIGURÁCIE

Pomocou tejto funkcie môže administrátor (prípadne kurátor) definovať základné parametre ovplyvňujúce zber domén.

6.3.7 HERITRIX - KONFIGURÁCIA

Funkcia zobrazuje zoznam pripravených konfigurácií pre utilitu Heritrix, ktorá uskutočňuje samotný zber. Pridanie novej konfigurácie do zoznamu je dostupné cez tlačidlo *Pridať*. Po kliknutí naň sa otvorí formulár, ktorý umožňuje zaradenie nového konfiguračného súboru do zoznamu.

6.3.8 HERITRIX - VÝNIMKY

Táto funkcia umožňuje definíciu výnimiek (crawler traps) pre utilitu Heritrix. Pridanie novej konfigurácie do zoznamu je dostupné cez tlačidlo *Vytvoriť*.

6.3.9 ZOZNAM SÚBOROV PRE EXTRAKCIU METADÁT

Táto funkcia obsahuje zoznam aktuálne používaných súborov, ktoré definujú pravidlá pre extrakciu opisných metadát z webových stránok. Pridanie novej konfigurácie do zoznamu je dostupné cez tlačidlo *Pridať*.

6.4 ADMINISTRÁCIA OSTATNÝCH ÚLOH

6.4.1 PREHĽAD ÚLOH

Funkcia Prehľad úloh zobrazuje zoznam evidovaných systémových úloh. V tabelárnej podobe je k jednotlivým úlohám evidované:

- typ úlohy,
- dátum spustenia,
- stav úlohy.

6.4.2 POTVRDENIE ARCHIVÁCIE SIP BALÍKOV

Zobrazený formulár obsahuje zoznam naplánovaných SIP balíkov (v stave *Čaká na spracovanie*), ktoré ešte neboli schválené na odoslanie do CDA. Filter umožňuje zobrazenie všetkých balíkov resp. iba SIP-ov pre e-Born vklady alebo WWW zbery.

Používateľ (s rolou *Administrátor*) označí balíky pre export do CDA pomocou výberového tlačidla na začiatku riadku. Formulár umožňuje označiť viacero SIP balíkov a hromadné označenie pomocou výberového tlačidla v riadku hlavičiek stĺpcov. Po použití tlačidla "*Potvrdiť archiváciu*" budú predmetné balíky presunuté na súborovom systéme do priečinka, z ktorého budú automaticky odoslané do CDA a vytvorí sa objednávka na vklad do CDA (viď používateľská príručka CDA pre PFI). Potvrdenie archivácie je nevratný proces. Archiváciu je možné potvrdiť aj hromadne, použitím tlačidla "*Potvrdiť archiváciu všetkých vyfiltrovaných*".

Pred odoslaním balíka do CDA je vhodné vykonať zbežnú kontrolu SIP-u použitím tlačidla *detail* na konci riadku predmetného záznamu. Z detailu SIP-u je možné zobraziť detail archívneho záznamu.

Informácia o úspešnom spracovaní SIP-u na strane CDA je okrem PFI modulu na stránkach CDA dostupná aj v detaile Zberu domény resp. na obrazovke Archívny záznam pre e-Born.

Detail zberu domény, ako aj detail archívneho záznamu e-Born takisto obsahujú údaj o stave archivácie do CDA daného objektu - položka *Archivácia do CDA* (zobrazená iba ak je objekt už priradený do SIP balíka). Okrem stavu archivácie údaj obsahuje aj SIP a AIP identifikátory (v závislosti od stavu archivácie).

6.5 KATALÓG WWW

6.5.1 KATALÓG DOMÉN

Katalóg domén obsahuje zoznam všetkých slovenských domén druhej úrovne (.sk), ako aj iných domén vybraných kurátorom. Je podporované filtrovanie, triedenie a stránkovanie. Pod zoznamom domén sa nachádzajú tlačidlá umožňujúce hromadný import a export metadát.

V prípade potreby môže používateľ využiť rozšírené filtrovanie dostupné cez link *Rozšírený filter*. Pomocou neho môže vytvoriť kritérium ako kombináciu logických podmienok. Počet podmienok spájaných cez logické AND a OR je možné dynamicky meniť.

6.5.2 HROMADNÝ IMPORT METADÁT

Tlačidlo *Hromadný import metadát* odkazuje na stránku Import metadát, kde je možné importovaním súboru vo formáte XML obohatiť metadáta domén.

6.5.3 HROMADNÝ EXPORT METADÁT

Tlačidlo *Hromadný export metadát* slúži na získanie metadát vybraných domén (pomocou filtra). Exportovaný súbor je vo formáte XML.

6.5.4 MANUÁLNE ZADANIE DOMÉNY

Pomocou tlačidla *Pridať* je možné otvoriť formulár na zadanie nového záznamu domény (napr. stránky slovenskej inštitúcie v .eu, .com alebo .tv priestore) cez GUI. S pridaným záznamom sa dá ďalej pracovať ako s ostatnými (automaticky importovanými) záznamami domén. Táto akcia podlieha auditingu.

6.5.5 IMPORT DOMÉN

Funkcia slúži na hromadné pridanie zoznamu domén v tvare CSV. Táto akcia podlieha auditingu.

6.5.6 DETAIL DOMÉNY

6.5.6.1 ZÁKLADNÉ ÚDAJE

Záložka obsahuje základné údaje o doméne, ktoré je možné editovať (URL je možné editovať len pokiaľ nebol vykonaný zber domény).

6.5.6.2 METADÁTA

Záložka Metadáta zobrazuje zoznam metadát prislúchajúcich k danej doméne. Pod zoznamom sú tlačidlá pre import a export metadát domény.

6.5.6.3 HISTÓRIA ZBEROV

Záložka obsahuje históriu zberov danej domény.

6.5.6.4 HISTÓRIA PRIESKUMOV

Na záložke sú zobrazené vykonané prieskumy danej domény.

6.5.6.5 NASTAVENIE ZBERU

Na tejto záložke je možné nastaviť špecifickú konfiguráciu pre zber danej domény. Zmena nastavení je možná vložením špecifického konfiguračného súboru pre Heritrix.

6.6 KATALÓG E-BORN PRAMEŇOV

Katalóg e-Born prameňov obsahuje zoznam všetkých e-Born objektov (monografie, periodiká, www stránky) pridaných do katalógu producentom / vkladateľom, alebo e-Born kurátorom.

6.6.1 PREHĽAD E-BORN PRAMEŇOV

Táto funkcia zobrazuje zoznam entít vložených do *Katalógu e-Born prameňov*. Zoznam je možné filtrovať podľa týchto kritérií:

- dátum vytvorenia (interval),
- producent / vkladateľ,
- status,
- metadátový profil.

Pomocou tlačidla *Pridať* je možné otvoriť formulár na zadanie žiadosti o prevzatie e-Born prameňa.

Kliknutím na tlačidlo *Detail* v príslušnom riadku s e-Born prameňom je možné otvoriť formulár na úpravu záznamu.

6.6.2 PRIDANIE E-BORN PRAMEŇA

Táto funkcia slúži na zadanie žiadosti o prevzatie e-Born prameňa. Používateľ zadá:

- producenta / vkladateľa,
- profil pre vklad,
- metadátový profil,
- základné opisné metadáta

a môže pridať jeden alebo viacero súborov.

6.6.3 ZOZNAM PROFILOV

Funkcia zobrazuje zoznam profilov využívaných pri zadávaní žiadosti o vklad e-Born prameňov.

Kliknutím na tlačidlo *Detail* v príslušnom riadku s profilom je možné otvoriť formulár na úpravu profilu.

Pomocou tlačidla *Pridať* vo funkcii *Zoznam profilov* je možné otvoriť formulár na vytvorenie nového profilu.

7 WEBOVÉ SLUŽBY A ROZHRANIA

7.1 EXTERNÉ SLUŽBY

7.1.1 VYHĽADÁVANIE

Databázy archívu WWW a archívu e-Born sú prístupné na verejnom portáli www.webdepozit.sk. Relevantné domény pre WWW pramene sa vyhľadávajú podľa zadáných kľúčových slov. Vyhľadávať je možné tiež pomocou rozšíreného vyhľadávania. Dostupné je aj filtrovanie na základe faziet. Na správu nearchívneho obsahu tohto webu sa využíva redakčný systém (CMS) Wordpress.

7.1.2 SPRÍSTUPŇOVANIE OBSAHU

Sprístupnenie štruktúrovaného elektronického obsahu zabezpečuje aplikácia OpenWayback, ktorá stránky zobrazuje a umožňuje ich vyhľadávať na základe URL a prezerať historicky v čase. V prípade e-Born obsahu je dokument zobrazený priamo z archívu. Obsah databáz je sprístupňovaný v troch režimoch:

- a) lokálny prístup - prístup v bádateľni digitálnych prameňov
(<https://www.webdepozit.sk/archivy-a-katalogy-dip/sprístupnovanie-archivu/>)
- b) voľný prístup – na verejnom portáli
- c) bez prístupu – prístup zakázaný

7.1.3 SPRÍSTUPŇOVANIE METADÁT

Sprístupňovanie metadát sa realizuje prostredníctvom služby OAI-PMH: <https://www.webdepozit.sk/archivy-a-katalogy-dip/katalogy-digitalnych-pramenov/#sluzby-OAIPMH>

Zdrojmi metadát, ktoré sa publikujú cez OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) sú dva katalógy: Archív WWW a Katalóg e-Born.

Zoznam publikovaných OAI-PMH služieb v IS Digitálne pramene:

#	Názov služby	URL	Typ prístupu
1	OAI-PMH provider, WWW zbery	https://oai.webdepozit.sk/www?verb=Identify	public access
2	OAI-PMH provider, e-Born pramene	https://oai.webdepozit.sk/eborn?verb=Identify	public access

1. Služba OAI-PMH pre WWW

Základnými entitami, na ktoré v tejto skupine viažeme metadáta, sú zbery jednej konkrétnej domény, resp. webového sídla. Každý zber konkrétneho webového sídla má vlastný (jedinečný) URN:NBN identifikátor (URN = Uniform Resource Name, NBN = National Bibliography Number). Služba je dostupná na nasledovnej URL: <https://oai.webdepozit.sk/www>

2. Služba OAI-PMH pre e-Born pramene

Základnými entitami, na ktoré v tejto skupine viažeme metadáta, sú e-Born pramene: elektronické monografie a periodiká. Jedinečným identifikátorom je URN:NBN identifikátor. Služba je dostupná na nasledovnej URL: <https://oai.webdepozit.sk/eborn>

7.1.4 DLHODOBÉ UCHOVÁVANIE V LTP ÚLOŽISKU

Domény a informácie o ich zberoch sú uchované vo WWW archíve vo formáte WARC. Súbor WARC a e-Born dokumenty sa odosielaajú v podobe SIP balíkov na dlhodobú archiváciu do Centrálného dátového archívu (CDA). Služba je interná aj externá.

Interná – vykonávajú sa interné úkony v IS Digitálne pramene (DIP), aby sa zabezpečilo uloženie webových stránok a elektronických publikácií v archíve, a tiež sa dokumenty vo forme SIP balíkov odosielaajú do CDA.

Externá – CDA služby pre verejnosť za asistencie DIP, žiadosti na sprístupnenie dokumentov z archívu budú používatelia adresovať DIP a DIP ich vyžiada od CDA.

8 REPORTING

8.1 REPORTY ZBEROV DOMÉN

Reporty sa nachádzajú v internom portáli po ukončení každého zberu, pri zobrazení detailu tohto zberu - Prehľad zberov/Detail zberu domén. Nachádzajú sa tam 3 reporty zberu.

8.1.1 REPORT ZBERU DOMÉN

Report obsahuje základné údaje o ukončenom zbere. Zarátané sú iba štatistiky pre zbory domén, ku ktorým bol Heritrixom vygenerovaný crawl-report. Stĺpce reportu:

- Doména - URL adresa
- Začiatok zberu – dátum a čas začiatku archivovania webového obsahu
- Koniec zberu - dátum a čas ukončenia archivovania webového obsahu
- Stav procesu – zobrazuje, že proces zberu domény je hotový (DONE). Po ukončení zberu domény obsahuje úspešný, neúspešný, zrušený, neznámy.
- Dôvod ukončenia zberu – úspešne ukončený (Finished) alebo že zber domény skončil na limite času (Timelimit hit), objemu (Maximum amount of data limit hit) alebo počte objektov (Object hit)
- Počet zozbieraných objektov – počet archivovaných objektov z danej domény
- Zozbieraný objem [byte] – presný objem z danej domény v bajtoch, pred komprimáciou
- Zozbieraný objem – približný zaokrúhlený objem v KB, MB, GB, TB

Doména	Začiatok zberu	Koniec zberu	Stav procesu	Dôvod ukončenia zberu	Počet zozbieraných objektov	Zozbieraný objem [byte]	Zozbieraný objem
www.tien.sk/	2.3.2021 8:47	2.3.2021 17:33	DONE	Finished	13298	898304685	857 MB

8.1.2 REPORT OBJEKTŮ ZBERU DOMÉN

Report obsahuje údaje zo zberu o všetkých typoch objektov, ich počet a objem. Nachádza sa tu reálny počet všetkého čo sa zbieralo (úspešné aj neúspešné, napríklad aj záznamy o dopytoch na chybové stránky...). Stĺpce reportu:

- Typ – typ objektu, súboru archivovaného v danom zbere (MIME type)
- Počet zbieraných objektov – počty daných typov objektov z daného zberu
- Zbieraný objem [byte] – objem daných typov objektov z daného zberu v bajtoch
- Zbieraný objem – zaokrúhlené vyjadrenie objemu v KB/MB/GB
- Spolu – posledný riadok, súčet údajov v danom stĺpci

Typ	Počet zbieraných objektov	Zbieraný objem [byte]	Zbieraný objem
application/json	6561	44717519	43 MB
image/jpeg	3146	660514480	630 MB
Spolu	13298	898304685	857 MB

8.1.3 REPORT STATUS KÓDOV ZBERU DOMÉN

Report vyjadruje http odpovede zo všetkých zbieraných objektov zberu a počet ich výskytu.

Nachádza sa tu reálny počet všetkého, na čo sa náš robot dopytoval (úspešne, neúspešne, ...). Stĺpce reportu:

- Status kód – číselné vyjadrenie odpovede
- Status text – textové vyjadrenie odpovede
- Počet – počty výskytov konkrétnych odpovedí zbieraných objektov

Status kód	Status text	Počet
1		228
200	OK	8009
301	Moved Permanently	779
400	Bad Request	26
403	Forbidden	12
404	Not Found	3841
500	Internal Server Error	5

8.2 REPORTY ZBERU DOMÉNY V DETAILE DOMÉNY

Reporty sa nachádzajú v internom portáli po ukončení každého zberu, pri zobrazení detailu tohto zberu a kliknutím na detail konkrétnej domény v zbere - Prehľad zberov/Detail zberu domén/Zber domény. Nachádza sa tu Log súbor na stiahnutie, možnosť zobraziť pozbieraný obsah v archíve po skončení indexácie pre OpenWayback, report objektov a report status kódov pre jednu konkrétnu doménu.

8.2.1 LOG SÚBOR DOMÉNY

Súbor slúži na hlbšiu analýzu detailu konkrétnej domény. Obsahuje všetky údaje o zbere, nastavenia zberu, heritrix logy, nastavenia pre metadáta (extrakčná šablóna). Súbor je na stiahnutie v ZIP formáte.

8.2.2 REPORT OBJEKTOV DOMÉNY

Report objektov zberu konkrétnej domény, pri jednom zbere vykonanom v konkrétny deň. Report obsahuje po ukončení zberu údaje o všetkých typoch objektov, ich počet a objem. Nachádza sa tu reálny počet všetkého, čo sa zozbieralo (úspešné, neúspešné, ...). Stĺpce reportu:

- Typ - typ objektu, súboru archivovaného v danom zbere (Mime type)
- Počet zozbieraných objektov - počty daných typov objektov z daného zberu
- Zozbieraný objem [byte] - objemy daných typov objektov z daného zberu v bajtoch
- Zozbieraný objem - zaokrúhlené vyjadrenie objemu v KB/MB/GB
- Spolu - posledný riadok, súčet údajov v danom stĺpci

Typ	Počet zozbieraných objektov	Zozbieraný objem [byte]	Zozbieraný objem
application/json	6561	44717519	43 MB
image/jpeg	3146	660514480	630 MB
Spolu	13298	898304685	857 MB

8.2.3 REPORT STATUS KÓDOV ZBERU DOMÉNY

Report status kódov konkrétnej vybranej domény po zbere. Report vyjadruje http odpovede zo všetkých zbieraných objektov zberu a počet ich výskytu. Nachádza sa tu reálny počet všetkého, čo sa zozbieralo (úspešné, neúspešné, ...). Stĺpce reportu:

- Status kód – číselné vyjadrenie odpovede
- Status text – textové vyjadrenie odpovede
- Počet – počty výskytov konkrétnych odpovedí zbieraných objektov

Status kód	Status text	Počet
1		228
200	OK	8009
301	Moved Permanently	779
400	Bad Request	26
403	Forbidden	12

8.3 REPORTY IS DIP

Obrazovka s reportami sa nachádza v Internom portáli v menu Administrácia a evidencia /Reporty. Obsahuje väčšinu reportov dostupných pre kurátorov o archivovaných údajoch v IS DIP.

8.3.1 REPORT KATALÓGU DOMÉN

Aktuálny report domén obsahujúci informácie o jednotlivých doménach z katalógu domén. Vytvára sa vždy aktuálne v daný deň a hodinu. Obsahuje momentálne aktuálne údaje o všetkých doménach. Stĺpce reportu:

- URL – adresa URL, doména
- stav domény – dostupnosť domény z posledného prieskumu Available/Unavailable
- dátum evidencie - dátum vytvorenia domény v katalógu (zo zoznamu sk-nic alebo manuálnym pridaním)
- oprávnenie na zber – možnosti: implicitne povolené, explicitne povolené, explicitne zakázané
- poznámka k oprávneniu na zber – ľubovoľný text kurátora slúžiaci pre internú potrebu
- súhlas so zverejnením obsahu – možnosti: verejné, limitované, neurčené, zakázané zobrazenie cez OW
- poznámka k súhlasu so zverejnením obsahu – poznámka od kurátora – spravidla licenčná zmluva alebo CC

URL	Stav domény	Dátum evidencie	Oprávnenie na zber	Poznámka k oprávneniu na zber	Súhlas so zverejnením obsahu	Poznámka k súhlasu so zverejnením obsahu
100.snd.sk/	Available	18.12.2020 13:53	IMPLICIT_PERMISSION		LIMITED	zml. na www.snd.sk
1989.knihovny.cz/	Available	22.11.2019 11:47	EXPLICIT_DISABLED		UNKNOWN	
ak.aps.sk/	Available	27.11.2015 20:49	EXPLICIT_DISABLED	Akademická knižnica	FORBIDDEN	
annazemanova.blog.sme.sk/	Available	5.2.2016 15:42	IMPLICIT_PERMISSION		PUBLIC	DDP_ZMLUVA

8.3.2 ŠTATISTICKÝ REPORT DOMÉN

Aktuálny report obsahuje počty a percentuálne podiely domén prvej úrovne (tzv. Top Level Domain - TLD) v katalógu domén. Obsahuje aktuálne údaje o všetkých doménach, posledný riadok je súčet všetkých domén. Stĺpce reportu:

- Doména – názov TLD (.sk, .cz, .eu, ...)
- Počet – súčet danej TLD
- Percent – percentuálny podiel danej TLD

8.3.3 REPORT DOMÉN SO ZMLUVOU

Aktuálny report domén s referenciou na zmluvu. V IS DIP ide o typ licencie, ktorý obsahuje hodnoty ako Zmluva (pre tento report hodnota UKB), CC a PV. V reporte sa zobrazujú vždy aktuálne údaje o predmetných doménach. Stĺpce reportu:

- url – konkrétna doména, URL adresa
- typ licencie – UKB (UKB – v IS DIP je to typ licencie Zmluva), CC, PV

- urn-nbn zmluvy – jedinečný identifikátor objektu zmluva, ktorý generuje IS DIP pri zakladaní zmluvy. Generuje sa len pre typ licencie Zmluva (UKB). Nachádza sa v metadátach [www](#), konkrétne tag 542__f
- označenie čísla zmluvy – interné označenie zmluvy, ktoré sa uvádza pri jej zakladaní v IS DIP (DDP_ZMLUVA_rok_poradové číslo)
- súhlas so zverejnením obsahu – možné hodnoty FORBIDDEN, PUBLIC, LIMITED a UNKNOWN. Na doméne uvedené ako Nie je prístupný, Plný prístup verejnosti, Limitovaný prístup a Neurčený
- poznámka k súhlasu so zverejnením obsahu – prenáša sa do metadát [www](#), konkrétne do tagu 542__n

URL	Typ licencie	URN-NBN zmluvy	Označenie	Súhlas so zverejnením obsahu	Poznámka k súhlasu so zverejnením obsahu
cda.kultury.sk	CC			PUBLIC	CC
dejum.sav.sk/	UKB	urn:nbn:sk:ddp-zml0000008510	DDP_ZMLUVA_2015_17	PUBLIC	Licenčná zmluva DDP_ZMLUVA_2015_17
openaccess.cvtisr.sk/	CC			PUBLIC	CC-BY-SA 4.0

8.3.4 REPORT POČTU ZOZBIERANÝCH OBJEKTŮ

Report obsahuje množstvo zozbieraného obsahu za vybrané obdobie (od – do na dni). Nachádzajú sa v ňom počty zozbieraných objektov a ich objem rozdelený podľa typu predlohy. Posledný riadok obsahuje súčet hodnôt pre každý daný stĺpec v reporte. Stĺpce reportu:

- Typ predlohy – podľa predlohy výberový, tematický, celoplošný
- Počet zozbieraných objektov v danom období
- Zozbieraný objem [byte] – presný počet v základnej jednotke
- Zozbieraný objem – zaokrúhlený nahor na KB, MB, GB, TB

Typ predlohy	Počet zozbieraných objektov	Zozbieraný objem [byte]	Zozbieraný objem
Výberový zber	1419787	1,05E+11	99 GB
Tematický zber	3775095	3,12E+11	291 GB
Spolu	5194882	4,17E+11	389 GB

8.4 REPORTY ZBEROV WWW

Nachádzajú sa tu reporty zo zberov za rok (výber konkrétneho roku) alebo prehľady za celé obdobie archívu. Po zadaní potrebného roku sa na pozadí vytvorí report a následne sa môže stiahnuť. Generovanie reportov na pozadí môže trvať niekoľko minút. Sťahujú sa vždy nové reporty za daný rok, alebo aktuálne reporty, t.j. k dnešnému dátumu za celé obdobie.

8.4.1 REPORT OBJEMU ARCHÍVU V ROKU

Ročný report slúži na mesačný, kvartálny a ročný prehľad počtu spustených zberov a počtu domén v týchto zberoch, pre ktoré sa zobrazuje aj pozbieraný objem. V posledných stĺpcoch sa nachádza

celkový objem archívu v ne/komprimovanej forme, ktorý sa pripočítava za každý mesiac. Január obsahuje údaj z minulého roka a December konečný objem archívu na konci roka. Objem archívu v roku: YYYY. Stĺpce reportu:

- Počet www zberov – sumárny počet www zberov, dátum spustenia zberu je menší alebo rovný end-date (posledný deň) daného obdobia a zároveň dátum ukončenia zberu je väčší ako start-date (prvý deň) daného obdobia
- Počet URL v zberoch - sumárny počet URL v zberoch (seeds), dátum spustenia zberu je menší alebo rovný end-date (posledný deň) daného obdobia a zároveň dátum ukončenia zberu je väčší ako start-date (prvý deň) daného obdobia
- Pozbieraný objem [Byte] - suma objemu (zozbieraný objem v Bajtoch) www zberov (harvest items). Dátum spustenia zberu je menší alebo rovný end-date (posledný deň) daného obdobia a zároveň dátum ukončenia zberu je väčší ako start-date (prvý deň) daného obdobia
- Komprimovaná veľkosť archívu [Byte] – suma objemu (zozbieraný objem v Bajtoch). Komprimovaná veľkosť všetkých uložených dát na Storage 1,2,3. Objem vyjadruje veľkosť WARC súboru v zazipovanej .gzip forme
- Nekomprimovaná veľkosť archívu [Byte] – suma objemu (zozbieraný objem v Bajtoch). Nekomprimovaná veľkosť všetkých uložených dát na Storage 1,2,3. Objem vyjadruje reálnu nekomprimovanú zozbieranú veľkosť
- Suma za celý rok – posledný riadok, súčet údajov pre daný stĺpec

8.4.2 POČET HTTP STATUS ZA ROK

Ročný report zobrazuje počty výskytov konkrétnych http odpovedí zbieraných objektov v zberoch za vybraný rok v daných mesiacoch a kvartáloch. Obsahuje agregované údaje z discovery, počty objektov s konkrétnymi HTTP status odpoveďami v zberoch. Nachádza sa tu reálny počet všetkého, čo sa zozbieralo (úspešné, neúspešné, ...). Počet http status za rok: YYYY. Stĺpce reportu:

- Status kód – číselné vyjadrenie http odpovede
- Status text – textové vyjadrenie http odpovede
- Q1-Q4 – súčet výskytov status kódu v prvom až štvrtom kvartáli vybraného roka
- Suma za celý rok – posledný stĺpec, súčet údajov pre daný status kód za rok

8.4.3 REPORT WARC V ROKU

Ročný report obsahuje postupný pribúdajúci počet súborov WARC a ich objem ku konkrétnemu mesiacu a kvartálu pre vybraný rok. Vo vybraných mesiacoch a kvartáloch sa nachádza počet a objem novo pribudnutých a následne všetkých WARC spolu. Počet WARC za rok: YYYY. Stĺpce reportu:

- 2020 – vybraný rok
- Počet nových WARC – sumárny počet WARC súborov, ktorých dátum poslednej zmeny je v danom období od-do (prvý deň, posledný deň roka) a zároveň dátum vytvorenia je menší alebo rovný end-date (posledný deň) daného obdobia a zároveň majú stav spracovania: Ukončený a zároveň sú typu: WARC
- Objem nových WARC [Byte] – suma objemu (veľkosť súborov v Bajtoch) WARC súborov, ktorých dátum poslednej zmeny je v danom období od-do (prvý deň, posledný deň roka) a zároveň dátum vytvorenia je menší alebo rovný end-date (posledný deň) daného obdobia a zároveň majú stav spracovania: Ukončený a zároveň sú typu: WARC

- Počet všetkých WARC – celkový sumárny počet WARC súborov v archíve
- Objem všetkých WARC [Byte] – celková suma objemu WARC súborov v archíve v bajtoch
- Suma za celý rok – posledný riadok, súčet údajov pre daný stĺpec

8.4.4 POČET TYPOV OBJEKTŮ V ROKU

Ročný report obsahuje počet zozbieraných typov obsahu (html, obrázky, audio/video) ku konkrétnemu kvartálu vo vybranom roku (mimetype, heritrix log). Posledný stĺpec obsahuje súčet všetkých kvartálov v danom roku. Nachádza sa tu reálny počet všetkého, čo sa zozbieralo (úspešné, neúspešné, ...). Tento report započítava všetky zbery vybraného roku. Počet typov obsahu za rok: YYYY. Stĺpce reportu:

- Typ formátu – typ objektu, súboru archivovaného v danom zbere
- Q1 až Q4 Počet zozbieraných objektov – koľko, počet daných typov objektov v danom kvartáli
- Suma za celý rok - stĺpec je súčtový za celý rok, za zozbieraný počet objektov
- Q1 až Q4 Zozbieraný objem [byte] - objem daných typy objektov v danom kvartáli vybraného roka v bajtoch
- Suma za celý rok - posledný stĺpec je súčtový za celý rok za zozbieraný objem v bajtoch

8.4.5 POČET TYPOV OBJEKTŮ CELKOVO

Celkový report, ktorý obsahuje údaje o všetkých typoch objektov, ich počet a objem. Nachádza sa tu reálny počet všetkého čo sa zozbieralo (úspešné, neúspešné, ...). Obsahom sú celkové počty typov obsahu – mimeType a ich veľkosť v bajtoch. Objekty, ktorých mime-type nie je validný, sú sčítané na konci tabuľky pod názvom "Iné". Stĺpce reportu:

- MimeType – typ objektu, všetky vyskytujúce sa formáty súborov v zberoch
- Počet zozbieraných objektov celkovo – počty všetkých objektov daného typu
- Zozbieraný objem celkovo – veľkosť všetkých objektov daného typu v bajtoch

8.4.6 CELKOVÝ PREHĽAD VŠETKÝCH ZBEROV

Celkový report obsahuje všetky zbery až po dnešok. Je vytvorený z obrazovky „Prehľad zberov“. Zobrazenie reportu je podľa názvu predlohy zberu a jej základných údajov. Stĺpce reportu:

- Názov predlohy – predloha vytvorená kurátorom
- Typ predlohy – tematická, výberová, celoplošná
- Dátum spustenia – spustenie zberu, dátum a čas
- Dátum ukončenia – ukončenie zberu, dátum a čas
- Stav zberu – ukončený, ukončený manuálne, prebiehajúci,...
- Úspešné – počet úspešných domén (URL) v zbere
- Neúspešné - počet neúspešných domén (URL) v zbere
- Vynechané - počet vynechaných domén (URL) v zbere
- Objem (GB) – objem daného zberu prepočítaný na GB (nekomprimovaný)

8.4.7 POČET HTTP STATUS CELKOVO

Celkový report zobrazuje http odpovede a počty ich výskytov zo všetkých zbieraných objektov zo všetkých realizovaných zberov. Nachádza sa tu reálny počet všetkého, čo sa zbieralo (úspešné, neúspešné, ...). Report obsahuje aktuálne agregované údaje z discovery, teda počty objektov s konkrétnymi HTTP status odpoveďami v zberoch (1, 100, 201, 300, 403, 404, ... 999). Report sčítava všetky zbery a z týchto zberov sčíta stavy odpovedí z každej zbieranej domény. Stĺpce reportu:

- Status kód – číselné vyjadrenie http odpovede
- Status text – textové vyjadrenie http odpovede
- Počet – počty výskytov konkrétnych odpovedí zbieraných objektov

Status kód	Status text	Suma celkovo
1		56748277
200	OK	904130130
201	CREATED	9869
202	ACCEPTED	6282