

DIGITAL RESOURCES

Webharvesting and e-Born Archiving

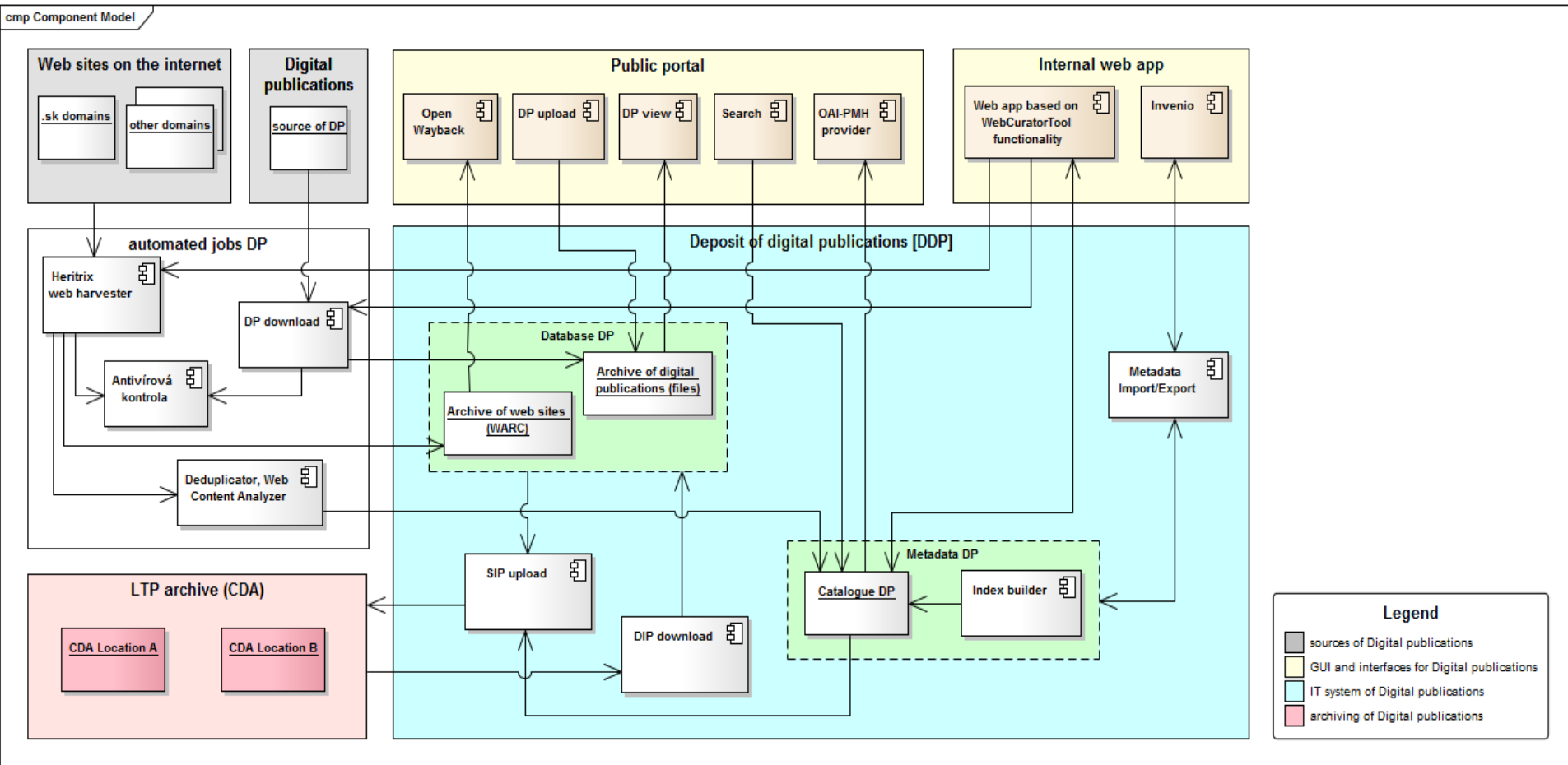
Digital Resources Project

- 2011 Start OPIS – 7 digitisation and 3 Infrastructural Projects
- **ULB:Central Data Archive Project (2012.01 - 2014.12)**
- **ULB:Digital Resources Project (DRP) (2015.05 – 2015.12)**
 - 2015.05 RFP
 - 2015.06 Project Proposal
 - 2015.07 Project Contract
 - 2015.07 – 11 Appl. Development
 - 2015.11 ICT Installation
 - 2015.12 Pilot harvest
- **DRP Mission & Goal**
- **Systematic collection and longtime archivation of slovacical web publications and native e-Born content as of integral part of the Slovak cultural heritage.**
- **Creation of technological and organisational infrastructure for archiving the Web and e-Born resources**

DRP – Functional Requirements

- **Information system** for
 - Survey
 - Registration
 - Harvesting
 - Collection
 - Processing
 - Preservation
 - Access
 - of e-Born Resources
- Architecture (7)
- Survey WWW (5)
- Acquisition (9)
- Scope (8)
- Quality (6)
- Reporting (3)
- Catalogue (7)
- Access (9)
- Preservation (6)
- Control (9)

DRP – Functional Diagram



DRP – Technical Infrastructure

- CISCO Server Cluster

- Public portal Server
- Internal portal Server
- Workers **24**x4 CPU

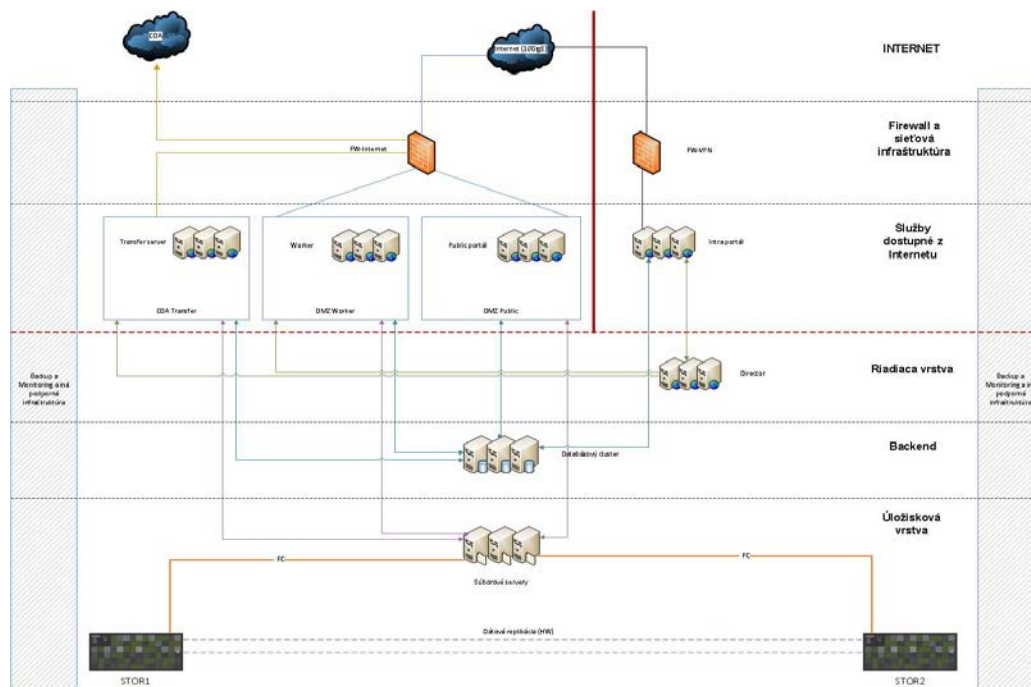
- Hitachi Storage

- 8x8Gbps
- **800** Tbyte

- CISCO

- SAN 48x16Gbps
- Firewall 8x**10**Gbps

- Infrastructure

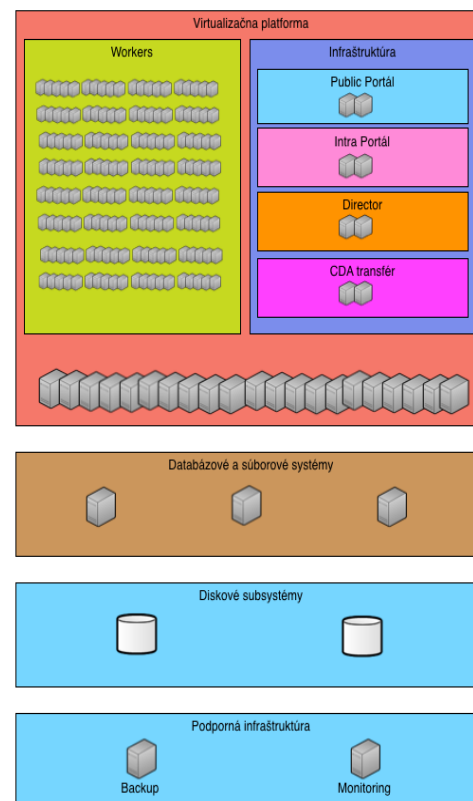


DIP – Software Warehouse

- System SW
 - OS RHEL
 - Apache Tomcat
 - PostgreSQL
 - SOLR
 - ZooKeeper
 - RabbitMQ
 - Java 8u45
 - Python
- Application SW
 - Heritrix
 - OpenWayback
 - Invenio
 - Wordpress
 - CLAMAV
 - Workflow
 - Cluster
 - IBM SmartCloud

DRP – System Structure

- Virtualisation Module
 - System Resources Management
- Database Module
 - Relational DB, Data sharing
- Storage Module
 - SATA(535), SAS(240), SSD(27) TB
- Support Modules
 - Backup, Monitoring, Communication



DRP – Resources I

- **nic .sk** ≈ 370 000 sites
- .eu, .info, .org, .???



DRP – Resources II

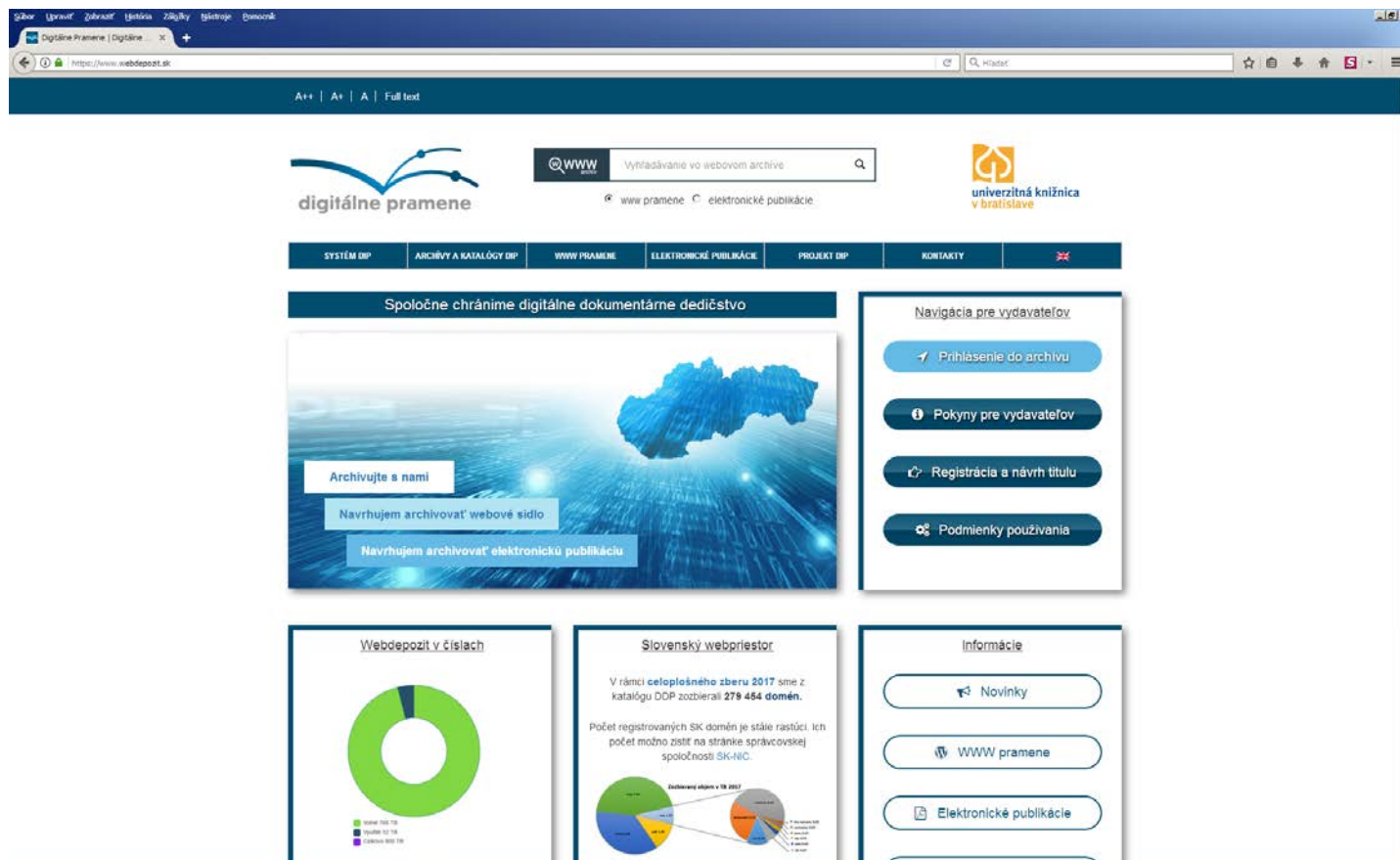
- **e-Serials (issn.issn.sk)**
 - 6500 Titles
 - 1100 on-line **17%**
 - 850 on-line live
 - **250 on-line cancelled 20%**
 - 090 Open Access
- **e-Monographs (www.snk.sk)**
 - 5500 Titles
 - 500 e-Born **11%**



DRP – Resources III

- WWW Collection Policy
 - **.sk** Domain Harvest 1xY
 - Thematic Harvest-events
 - Selective Harvest –
Conspect
 - Agreements
- e-Born Collection Policy
 - Agreements
 - Creative Commons
 - OpenSource

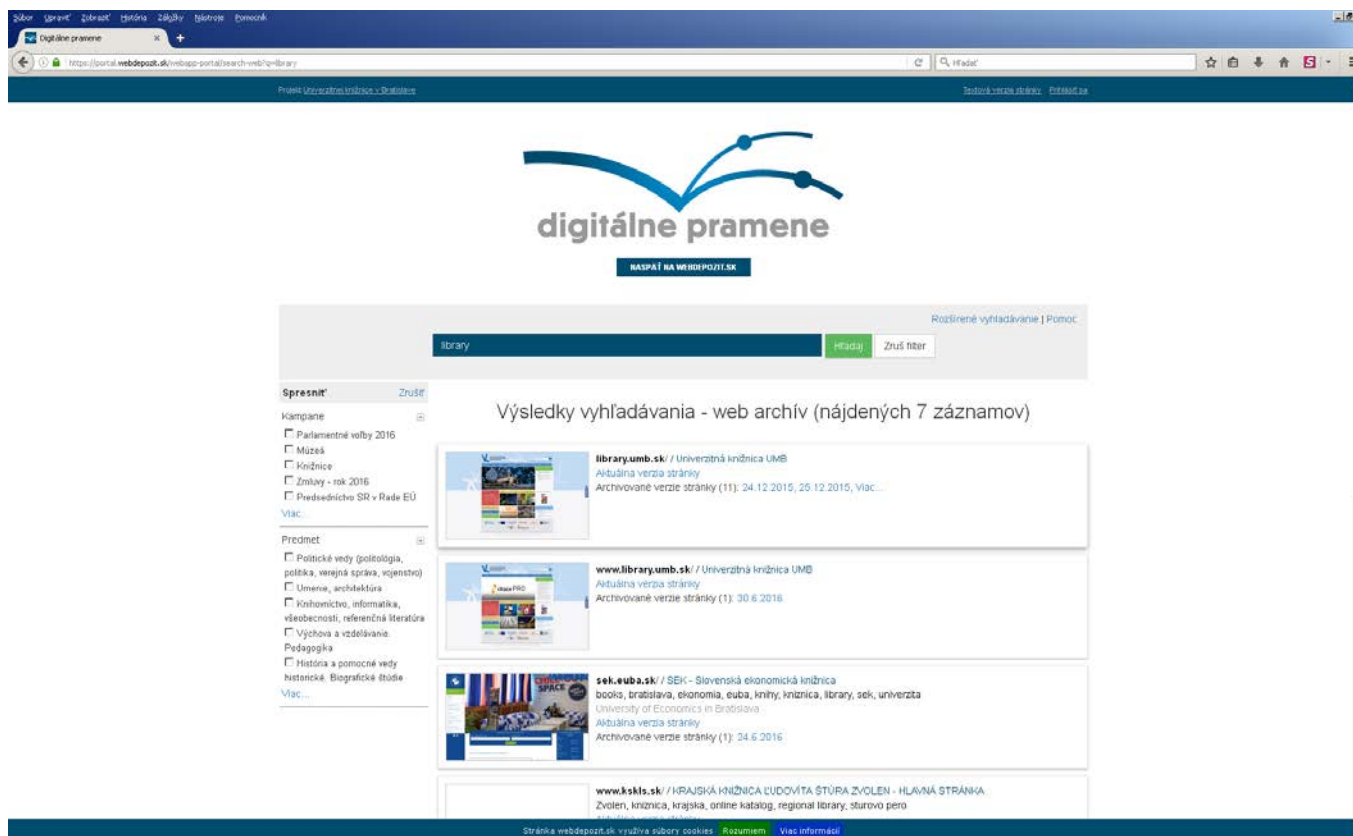
DRP – Public Portal



The screenshot shows the Digital Sources (DRP) Public Portal website. The page layout includes a navigation menu with categories like 'SYSTÉM DRP', 'ARCHÍVY A KATALÓGY DRP', 'WWW PRAMENE', 'ELEKTRONICKÉ PUBLIKÁCIE', 'PROJEKT DRP', and 'KONTAKTY'. The main content area is divided into several sections:

- Spoločne chránime digitálne dokumentárne dedičstvo:** A central banner with a blue abstract graphic and buttons for 'Archivujte s nami', 'Navrhujem archivovať webové sídlo', and 'Navrhujem archivovať elektronickú publikáciu'.
- Navigation for publishers:** A sidebar with buttons for 'Prihlásenie do archívu', 'Pokyny pre vydavateľov', 'Registrácia a návrh titulu', and 'Podmienky používania'.
- Webdepozit v číslach:** A donut chart showing the distribution of web deposits.
- Slovenský webpriestor:** Text stating that in the 2017 survey, 279,454 domains were collected, and a pie chart showing the distribution of registered domains.
- Informácie:** A sidebar with buttons for 'Novinky', 'WWW pramene', and 'Elektronické publikácie'.

Keyword „library“

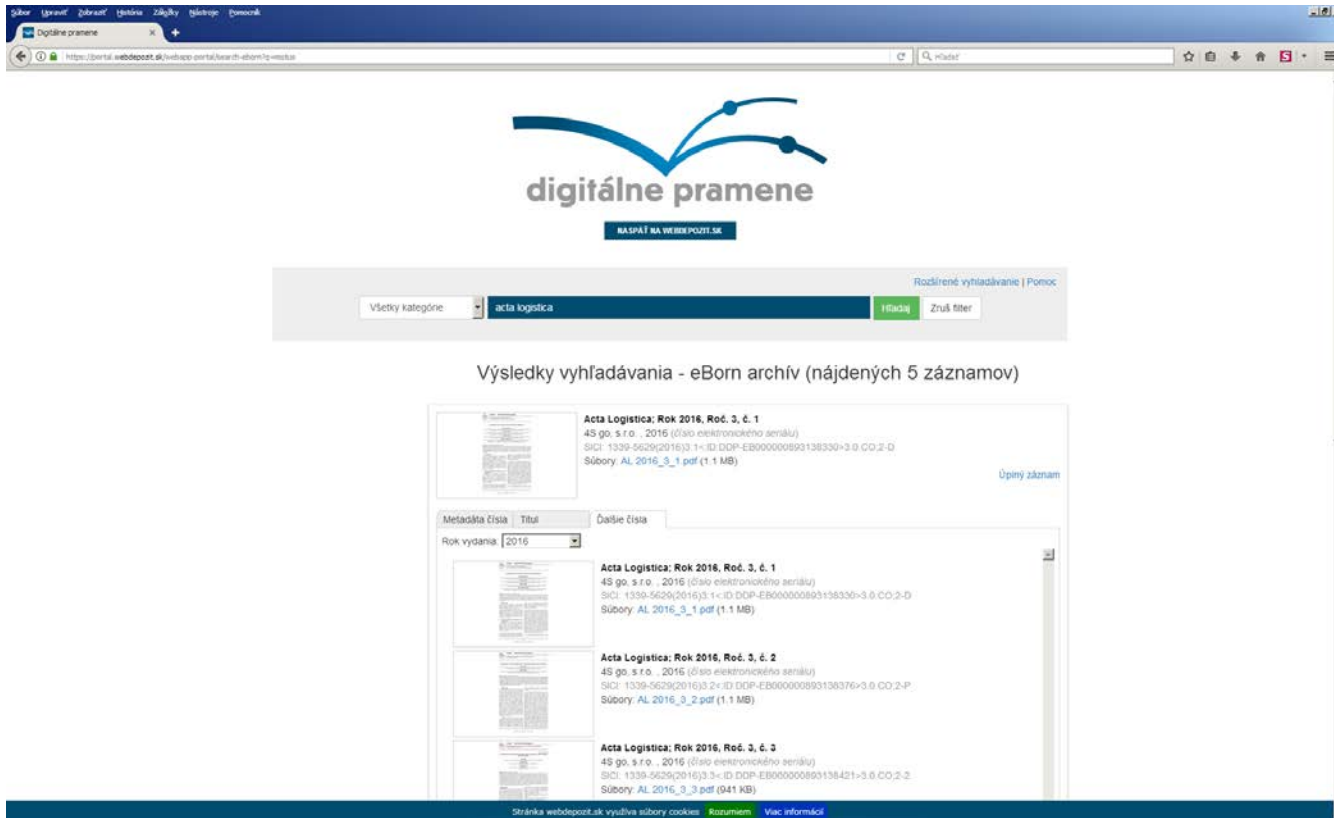


The screenshot shows a web browser window displaying the 'digitálne pramene' portal. The search bar contains the keyword 'library'. The search results are displayed as follows:

- librarijumb.sk / Univerzitná knižnica UMB**
Aktuálna verzia stránky
Archivované verzie stránky (11): 24.12.2015, 26.12.2015, Viac...
- www.librarijumb.sk / Univerzitná knižnica UMB**
Aktuálna verzia stránky
Archivované verzie stránky (1): 30.6.2016
- sek.euba.sk / SEK - Slovenská ekonomická knižnica**
books, bratislava, ekonomia, euba, knihy, knižnica, library, sek, univerzita
University of Economics in Bratislava
Aktuálna verzia stránky
Archivované verzie stránky (1): 24.6.2016
- www.kkskls.sk / KRAJSKÁ KNÍŽNICA ĽUDOVÍTA ŠTÚPA ZVOLEN - HLAVNÁ STRÁNKA**
Zvolen, knižnica, krajska, online katalog, regional library, starovo pero

At the bottom of the page, there is a footer: [Stránka webdepozičiek](#) [vyráža súbory cookies](#) [Recenzium](#) [Viac informácií](#)

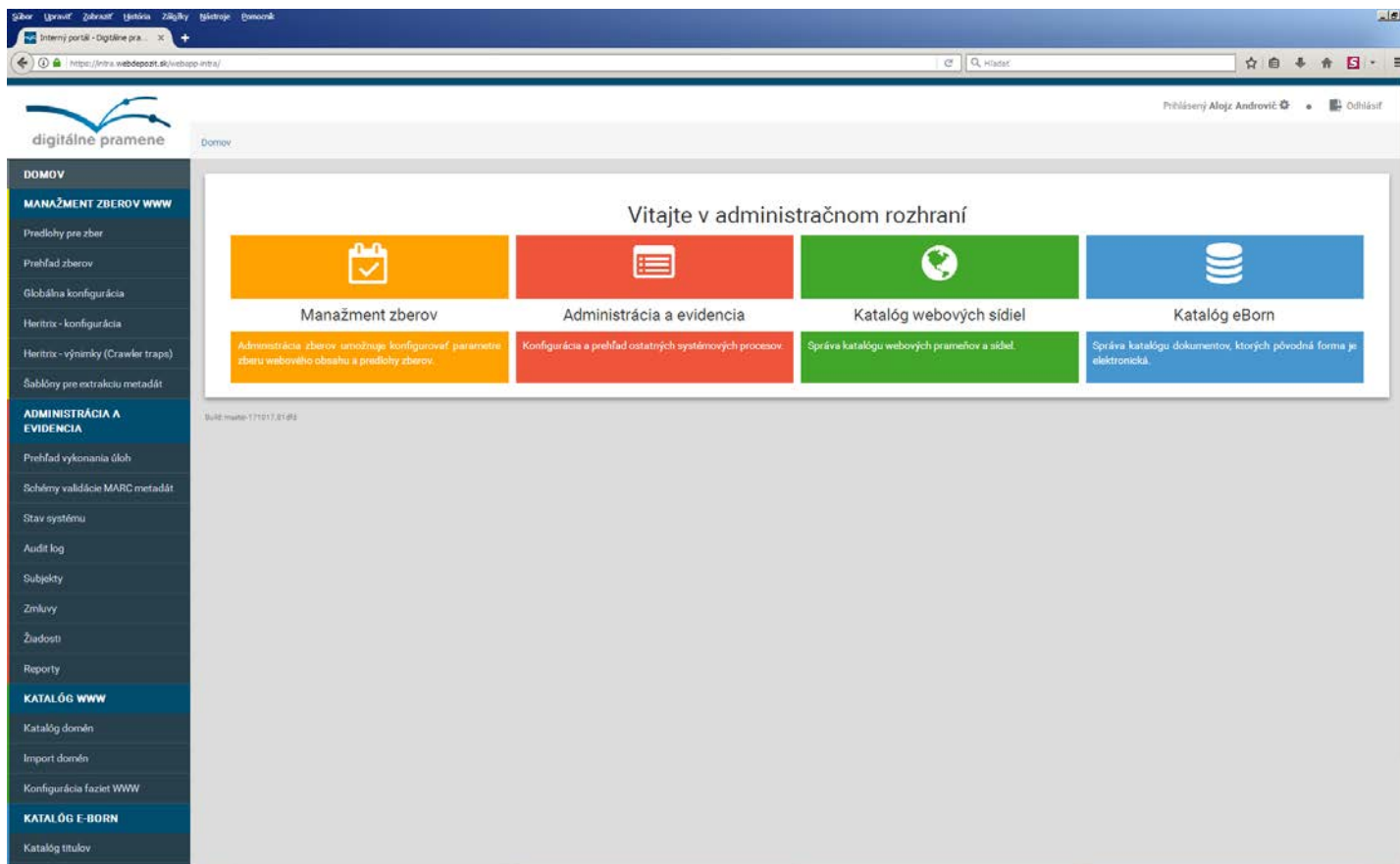
Keyword „acta logistica“



The screenshot shows a web browser window displaying the search results for the keyword "acta logistica" on the "digitálne pramene" website. The search results are displayed in a table format, showing the title, year of publication, and file size for each record. The search results are as follows:

Metadáta	Číslo	Titul	Ďalšie číslo
4S go, s.r.o., 2016 (ďalšie elektronického seriálu)	1339-5629(2016)3 1-ID DDP-EB000000893138330-3.0.CO.2-D	Acta Logistica: Rok 2016, Roč. 3, č. 1	Subory: AL 2016_3_1.pdf (1.1 MB)
4S go, s.r.o., 2016 (ďalšie elektronického seriálu)	1339-5629(2016)3 1-ID DDP-EB000000893138330-3.0.CO.2-D	Acta Logistica: Rok 2016, Roč. 3, č. 1	Subory: AL 2016_3_1.pdf (1.1 MB)
4S go, s.r.o., 2016 (ďalšie elektronického seriálu)	1339-5629(2016)3 2-ID DDP-EB000000893138376-3.0.CO.2-P	Acta Logistica: Rok 2016, Roč. 3, č. 2	Subory: AL 2016_3_2.pdf (1.1 MB)
4S go, s.r.o., 2016 (ďalšie elektronického seriálu)	1339-5629(2016)3 3-ID DDP-EB000000893138421-3.0.CO.2-Z	Acta Logistica: Rok 2016, Roč. 3, č. 3	Subory: AL 2016_3_3.pdf (941 KB)

DRP – Curator Interface







digitálne pramene

Domov

Přihlášený Alojz Androvit

Vítajte v administračnom rozhraní

 Manažment zberov Administrácia zberov umožňuje konfigurovať parametre zberu webových obsahů a predlohy zberov.	 Administrácia a evidencia Konfigurácia a prehľad ostatných systémových procesov.	 Katalóg webových sídiel Správa katalógu webových prameňov a sídiel.	 Katalóg eBom Správa katalógu dokumentov, ktorých pôvodná forma je elektronická.
---	---	--	--

Built using 17101.81@#

DOMOV

- MANAŽMENT ZBEROV WWW
 - Predlohy pre zber
 - Prehľad zberov
 - Globálna konfigurácia
 - Heritrix - konfigurácia
 - Heritrix - výnimky (Crawler traps)
 - Šablóny pre extrakciu metadát
- ADMINISTRÁCIA A EVIDENCIA
 - Prehľad vykonania úloh
 - Schémy validácie MARC metadát
 - Stav systému
 - Audít log
 - Subjekty
 - Zmluvy
 - Žiadosti
 - Reporty
- KATALÓG WWW
 - Katalóg domén
 - Import domén
 - Konfigurácia faziet WWW
- KATALÓG E-BORN
 - Katalóg titulov

DRP – Harvest Control

Predlohy pre zber

Nová predloha výberového zberu
 Nová predloha tematického zberu
 Nová predloha celoplošného zberu

Názov predlohy:

Typ predlohy:

Aktívna predloha:

Plánované spustenie od: do:

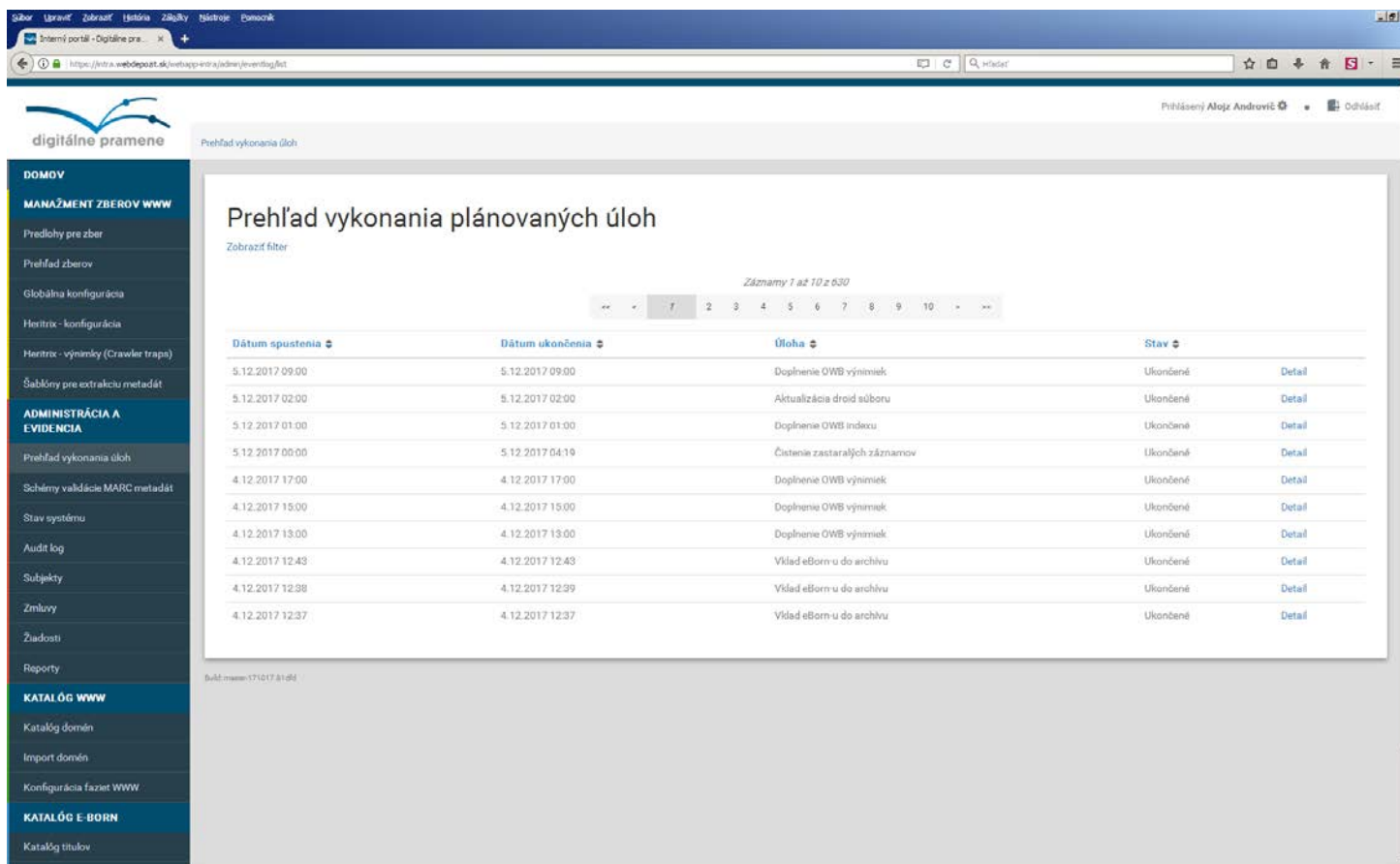
Plánované ukončenie od: do:

Najbližšie spustenie od: do:

Záznamy 1 až 10 z 17

Názov predlohy	Typ predlohy	Plánované spustenie	Plánované ukončenie	Plánovaný počet	Najbližšie spustenie	Počet spustení	Aktívna predloha		
Celoplošný zber 2016	Celoplošný zber	7.10.2016	7.10.2016	-	-	2	✓	Detail	História zberov
Celoplošný zber 2017	Celoplošný zber	24.2.2017	24.2.2017	-	-	2	✓	Detail	História zberov
E-Born WWW	Výberový zber	28.3.2017	28.3.2017	-	-	3	✓	Detail	História zberov
Kultúrny profil Slovenska	Výberový zber	21.4.2017	21.4.2017	-	-	5	✓	Detail	História zberov
Kultúrny profil Slovenska	Výberový zber	21.5.2017	21.5.2017	-	-	4	✓	Detail	História zberov
Kultúrny profil Slovenska	Výberový zber	19.5.2017	19.5.2017	-	-	5	✓	Detail	História zberov
Medzinárodný festival súčasného umenia Biela noc	Výberový zber	9.10.2017	9.10.2017	-	-	6	✓	Detail	História zberov
Mesačný zber ULIB	Výberový zber	30.6.2017	-	200	27.12.2017	9	✓	Detail	História zberov
Mesačný zber Webdepozit	Výberový zber	13.4.2017	-	200	9.12.2017	10	✓	Detail	História zberov
Časopis na žeb. športovcov	Výberový zber	19.7.2017	19.7.2017	-	-	1	✓	Detail	História zberov

DRP – Administration



The screenshot shows the 'Prehľad vykonania plánovaných úloh' (Overview of planned task execution) page in the DRP Administration system. The page includes a sidebar with navigation options and a main content area with a table of task execution records.

Navigation Sidebar:

- DOMOV
- MANAŽMENT ZBEROV WWW
 - Predlohy pre zber
 - Prehľad zberov
 - Globálna konfigurácia
 - Heritrix - konfigurácia
 - Heritrix - výnimky (Crawler traps)
 - Šablóny pre extrakciu metadát
- ADMINISTRÁCIA A EVIDENCIA
 - Prehľad vykonania úloh
 - Schémy validácie MARC metadát
 - Stav systému
 - Audit log
 - Subjekty
 - Zmluvy
 - Žiadosti
 - Reporty
- KATALÓG WWW
 - Katalóg domén
 - Import domén
 - Konfigurácia faziet WWW
- KATALÓG E-BORN
 - Katalóg titulov

Main Content Area:

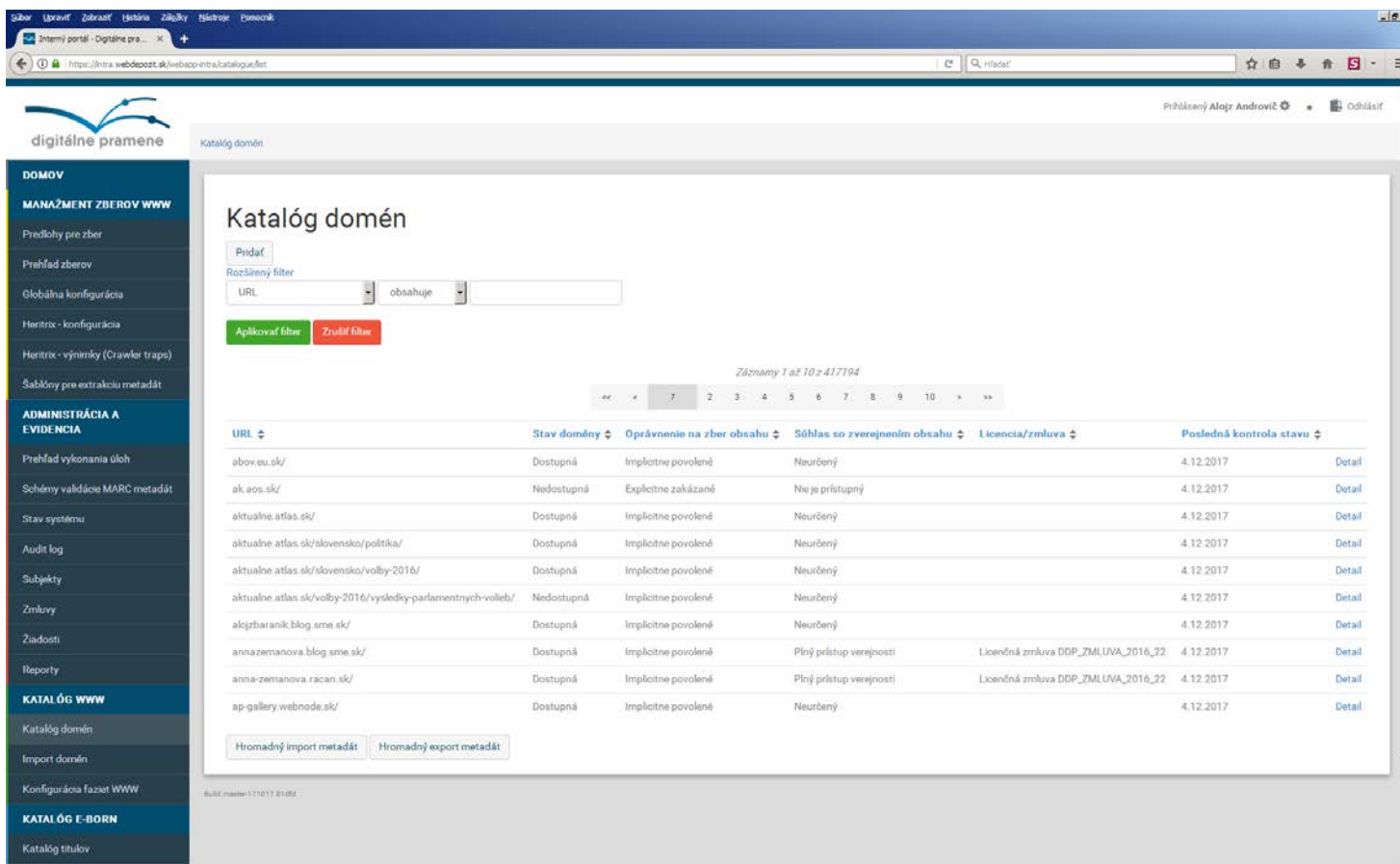
Prehľad vykonania plánovaných úloh

Zobraziť filter

Záznamy 1 až 10 z 630

Dátum spustenia	Dátum ukončenia	Úloha	Stav
5.12.2017 09:00	5.12.2017 09:00	Doplnenie OWB výnimiek	Ukončené
5.12.2017 02:00	5.12.2017 02:00	Aktualizácia droid súboru	Ukončené
5.12.2017 01:00	5.12.2017 01:00	Doplnenie OWB indexu	Ukončené
5.12.2017 00:00	5.12.2017 04:19	Čistenie zastaralých záznamov	Ukončené
4.12.2017 17:00	4.12.2017 17:00	Doplnenie OWB výnimiek	Ukončené
4.12.2017 15:00	4.12.2017 15:00	Doplnenie OWB výnimiek	Ukončené
4.12.2017 13:00	4.12.2017 13:00	Doplnenie OWB výnimiek	Ukončené
4.12.2017 12:43	4.12.2017 12:43	Vklad eBORN-u do archívu	Ukončené
4.12.2017 12:38	4.12.2017 12:39	Vklad eBORN-u do archívu	Ukončené
4.12.2017 12:37	4.12.2017 12:37	Vklad eBORN-u do archívu	Ukončené

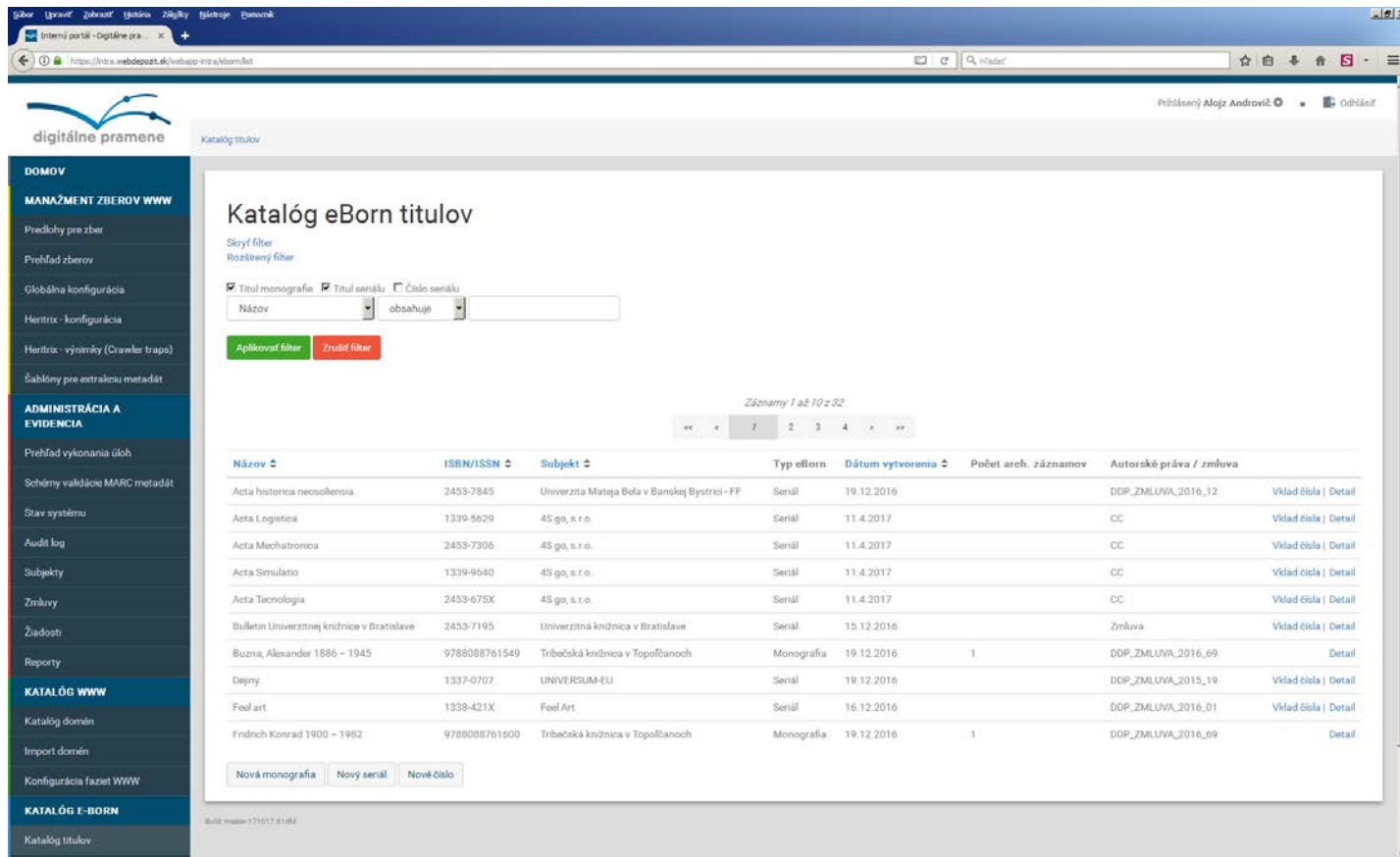
DRP – WWW Catalogue



The screenshot shows the 'Katalóg domén' (Domain Catalogue) interface. It features a sidebar with navigation options like 'DOMOV', 'MANAŽMENT ZBEROV WWW', 'ADMINISTRÁCIA A EVIDENCIA', and 'KATALÓG WWW'. The main content area displays a table of domain records with columns for URL, status, permissions, consent, license, and last control date. A search filter is visible at the top of the table.

URL	Stav domény	Oprávnenie na zber obsahu	Súhlas so zverejnením obsahu	Licencia/zmluva	Posledná kontrola stavu
abov.eu.sk/	Dostupná	Implicitne povolené	Neurčený		4.12.2017
ak.aos.sk/	Nedostupná	Explicitne zakázané	Ne je prístupný		4.12.2017
aktualne.atlas.sk/	Dostupná	Implicitne povolené	Neurčený		4.12.2017
aktualne.atlas.sk/slovensko/politika/	Dostupná	Implicitne povolené	Neurčený		4.12.2017
aktualne.atlas.sk/slovensko/volby-2016/	Dostupná	Implicitne povolené	Neurčený		4.12.2017
aktualne.atlas.sk/volby-2016/vysledky-parlamentnych-volieb/	Nedostupná	Implicitne povolené	Neurčený		4.12.2017
algzbaranik.blog.sme.sk/	Dostupná	Implicitne povolené	Neurčený		4.12.2017
annazemanova.blog.sme.sk/	Dostupná	Implicitne povolené	Plný prístup verejnosti	Licenčná zmluva DDP_ZMLUVA_2016_22	4.12.2017
anna-zemanova.racan.sk/	Dostupná	Implicitne povolené	Plný prístup verejnosti	Licenčná zmluva DDP_ZMLUVA_2016_22	4.12.2017
ap-gallery.webnode.sk/	Dostupná	Implicitne povolené	Neurčený		4.12.2017

DRP – eBorn Catalogue

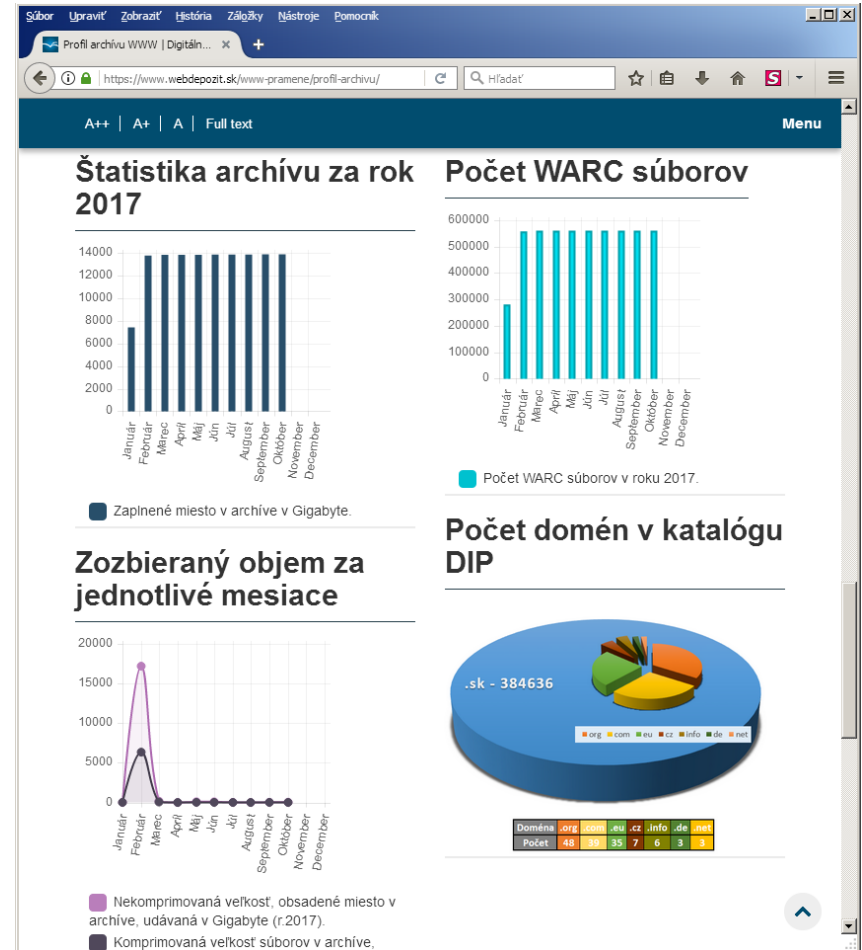
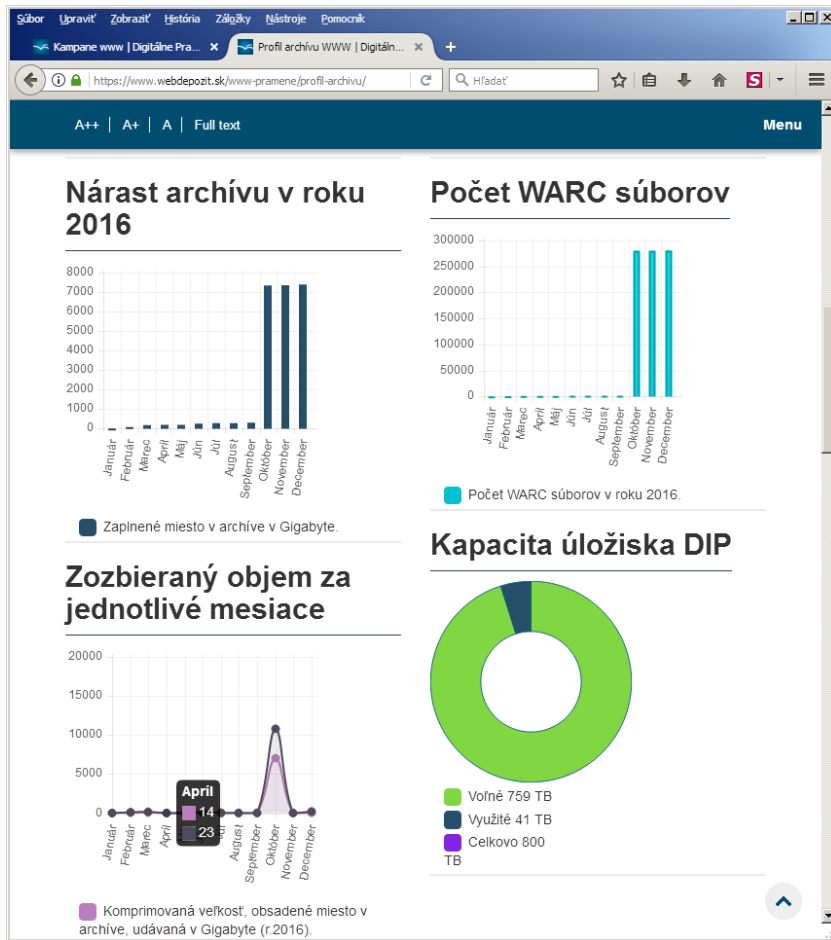


The screenshot shows the 'Katalóg eBorn titulov' (eBorn Catalogue) web application. The interface includes a navigation menu on the left with sections like 'DOMOV', 'MANAŽMENT ZBEROV WWW', 'ADMINISTRÁCIA A EVIDENCIA', 'KATALÓG WWW', and 'KATALÓG E-BORN'. The main content area features a search filter section with options for 'Titul monografia' and 'Titul seriálu', and a search input field. Below the search is a table of records with columns for 'Název', 'ISBN/ISSN', 'Subjekt', 'Typ eBorn', 'Dátum vytvorenia', 'Počet arch. záznamov', and 'Autorské práva / zmluva'. The table lists various titles such as 'Acta historica neosolvensia', 'Acta Logistica', 'Acta Mechatronica', 'Acta Simulatio', 'Acta Technologia', 'Bulletin Univerzitetnej knižnice v Bratislave', 'Buzna, Alexander 1886 – 1945', 'Dějiny', 'Feel art', and 'Fridrich Konrad 1900 – 1982'. A footer note indicates the build version: 'Build: master-1.1017.8188'.

DRP – Organisational Support

- **ULB - Elektronisation and Integration section**
 - **National ISSN Agency and DRP**
 - Department: **Digital resources Deposit** (1+3)
 - External capacities
 - Support (1 FTE)
 - Experts (0.5 FTE/2)
- **DRP Staff**
 - DRP Manager
 - Curator I
 - DRP Information system
 - Curator II
 - WWW Resources
 - Curator III
 - E-Born Resources

DRP - Figures



DRP – Sustainability

- 2015 (6 800 000)
- 2016–2020 (3 400 000)

Project Budget €	
HW	5 800 000
SW sys.+apl.	800 000
Personal	160 000
Production	40 000
Total	6 800 000

Running Costs €	
SLA Support	500 000
Staff	60 000
Production	120 000
Total	680 000

Saving and keeping!

- Digital technologies are enriching the information, publication and creation processes. Digital Resources – webpages and objects and native electronic publications are due to their origin endangered. They evolve quickly and massive, and they change and disappear even faster.
- Our effort is the comprehensive collection and long-term preservation of the testimonies of the present “forgetting epoque“. We believe that thank to close cooperation with the politicians, authors, editors, providers and creators of the digital content we could master this!