

БОГДАН ТРИФУНОВИЋ, М. А.
Градска библиотека
„Владислав Петковић Дис“, Чачак

УДК 027.54(437.1):004

ПРОЈЕКТИ ДИГИТАЛНЕ ЗАШТИТЕ НАЦИОНАЛНЕ БИБЛИОТЕКЕ ЧЕШКЕ РЕПУБЛИКЕ

Сажетак: Национална библиотека Чешке Републике покренула је почетком 90-тих година прошлог века прве кораке дигиталне заштите писаног научног и културног наслеђа чешког народа од средњег века до савременог доба. Током времена усталила су се три пројекта, који се баве посебним врстама грађе за дигитализацију: Manuscriptorium садржи дигитализоване рукописе и старе књиге, Kramerius монографске и серијске публикације, WebArchiv дигиталну архиву чешког интернет домена. У раду су сажето приказани ови пројекти и основне информације о раду и плановима дигиталне заштите у Националној библиотеци Чешке Републике.¹

Кључне речи: дигитална заштита, дигитализација, Manuscriptorium, Kramerius, WebArchiv, веб архивирање, национална библиотека, Чешка Република

Увод

Средином 80-тих година 20. века лондонски ВВС дошао је на идеју да створи мултимедијално издање „Књиге судњег дана“ Енглеске на крају 20. века, по узору на 900 година стар оригинал, први свеобухватни попис сред-

¹ Аутор је током новембра 2008. године боравио на стручном усавршавању у Националној библиотеци у Прагу, захваљујући стипендији Digital Preservation Europe Exchange Program (DPEX, <http://www.digitalpreservationeurope.eu/exchange/>) и подршци Градске библиотеке „Владислав Петковић Дис“ у Чачку.

њовековне Енглеске, написан 1086. године по наређењу Виљема Освајача. На пројекту који је коштао тадашњих 2,5 милиона фунти, радило је више од 60 сарадника, милион људи доставило је различите податке о савременој Енглеској, развијен је посебан софтвер и хардвер за управљање подацима који су снимљени на двостране ласерске дискове, а за њихово читање коришћени су уређаји под контролом тадашњих микрокомпјутера. Након окончања пројекта, на дисковима је било смештено преко 200.000 фотографија, 24.000 карата, 8.000 разних база података и 60 минута видео записа. Овај, по свим критеријумима изузетан пројекат (прве копије дискова смештене су поред оригиналне „Књиге судњег дана“ у Националном архиву), у великој мери је био испред свог времена, па га је и задесила судбина честа међу визионарима. Почетком 21. века, на свету скоро да није било уређаја који је могао да прочита и користи записане податке на видео дисковима мултимедијалне „Књиге судњег дана“, јер су коришћене технологије у року од једне деценије постале застареле и превазиђене. Петнаест година након што је урађена, „дигитална књига“ постала је бескорисна, док је руком писани оригинал и после 900 година у одличном стању.²

Порука ове поучне приче је да не постоји дигитализација урађена једном за сва времена и да сви субјекти укључени у процесе дигитализације морају узети у обзир трансфер технологија, стандарда и формата у планирању дугорочних програма дигиталне заштите баштине. Иако је крајем 2002. године тим стручњака са Универзитета Мичиген и са Универзитета Лидс успео да створи софтвер за „емулирање“ технологије потребне за читање дискова модерне „Књиге судњег дана“, не постоје гаранције да се за наредних 15-20 година поново не јави потреба за сличним послом. И тако у недоглед. Финансијска и структурална одрживост дигиталне заштите у будућности почиваће на решавању поменутог проблема.³

Термин Digital Preservation („дигитална заштита“, „дигитално очување“) стручна јавност у свету прихватила је за обухватање свих активности, послова, знања и вештина које се примењују у процесу (дугорочног) идентификовања, селектовања, обраде, похрањивања, заштите и приступа информацијама у дигиталном облику. Сведено на терминологију конкретних активности, дигитална заштита подразумева „управљање дигиталним информацијама током времена“, што може да подразумева и дигитализа-

² <http://www.csa.com/discoveryguides/cyber/overview.php>

³ Више о томе: Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. A Report to the Council on Library and Information Resources, Washington 1999 <<http://www.clir.org/pubs/reports/rothenberg/contents.html>>; Kenneth Thibodeau, *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*. CLIR Reports <<http://www.clir.org/pubs/reports/pub107/thibodeau.html>>.

цију извора у папирном облику, дигитализацију аудио и видео материјала и мултимедије, дигитализацију „електронски рођених“ докумената, успостављање система метаподатака за обраду, чување и вођење података о дигиталним колекцијама, стварање поверљивих дигиталних репозиторијума, одређивање нормативних и системских оквира од стране државе, итд. Практично, све активности које се обављају на пољу очувања информација, њихове организације на начин који ће корисницима у будућности пружити јасан и меродаван увид у садржај и форму дигиталних информација. Овај рад приказује решења једне установе на пољу дигиталне заштите националне баштине, кроз анализу конкретних пројеката и активности које спроводи Национална библиотека Чешке Републике у Прагу.

Национална библиотека Чешке Републике

Национална библиотека Чешке Републике⁴ (Národní knihovna ČR, НКП) налази се у самом срцу Прага, у згради некадашње високе језуитске школе, данас познате под именом Клементинум. Ово огромно здање из 16. века заузима читав стамбени блок на површини преко два хектара и незаобилазна је туристичка атракција, са изложбама на отвореном и у затвореном простору, организованим посетама барокној библиотеци и читаоници, астрономском торњу или концертној сали. Поред ове зграде НКП, на ободима Прага, у предграђу Хостивар, постоји и климатизовани Централни депо, установљен 1996. године, где је смештено преко пет милиона штампаних публикација. У циљу модернизације и решавања складиштених проблема (јер се показало да ове две зграде нису довољне), 2007. године расписан је конкурс за идејно решење нове зграде Националне библиотеке. Победничко решење, „Око над Прагом“, својим радикалним модерним дизајном, као супротношћу традиционалном архитектонском стилу ширег центра града, покренуло је жучну друштвену расправу између поборника и противника тог решења, па је пројекат на крају остављен.⁵

Централни депо у Хостивару смештен је у неинвентивном индустријском здању, за разлику од зграде Клементинума или предложене нове зграде НКП, и обједињује у себи, између осталог, одељења конзервације и заштите књиге, као и одељење за дигиталну заштиту. Дигитална заштита публикација у Хостивару подразумева свеобухватни процес пријема, обраде, скенирања, креирања метаподатака за дигитализоване публикације из 19. и 20. века и њихово објављивање на порталу Крамериус (Kramerius). У

⁴ <http://nkp.cz/>

⁵ http://www.nkp.cz/novabudova/novabudova_aktuality.htm

оквиру одељења за дигиталну заштиту налази се више професионалних камера и скенер апарата, који у раду дневно створе хиљаде дигиталних копија и истовремено производе копије на микрофилму. НКП се определила да све што се дигитализује буде чувано у дигиталном и микрофилм облику, јер се микрофилм и даље сматра далеко поузданијим медијумом од било ког дигиталног формата. Простор где се налазе камере и скенери у потпуности је замрачен, изузев осветљења које производе сами уређаји током рада, ради што бољег квалитета дигиталних докумената. Одељење за дигиталну заштиту такође води и централни дигитални репозиторијум, док се микрофилмови чувају у посебним кутијама које обезбеђују вишевековно очување квалитета микрофилм записа на одређеној температури и влажности ваздуха. Поред тога, НКП је купила и плотер уређај за израду кутија у које се одлажу старе и ретке публикације, након процеса дигитализације и конзервације. Кутије се могу израдити за најразличитије формате (од књиге до плаката и карата), коришћењем папира од чисте плуте, што спречава појаву влаге или пропадање папира услед ефекта ацидације.⁶



⁶ Уопште је читава атмосфера организације и обима рада који се спроводи у Хостивару на дигиталној заштити код аутора створила осећај да се налази у „фабрици за дигитализацију“ (прим. аут.).

Пројекти дигиталне заштите у НКП**Манускрипториум (*Manuscriptorium*) – дигитална библиотека рукописа**

Манускрипториум⁷ представља систем за прикупљање, опис и приказ информација о старим рукописима и штампаним књигама, повезан са виртуелном библиотеком дигитализованих докумената за детаљан преглед заступљених рукописа и књига. Пројекат финансира Национална библиотека Чешке Републике, док њиме управља прашка компанија AiP Beroun, која је и аутор техничких решења система. То је условило да заступљене колекције и документа делимично буду слободно доступни корисницима на веб сајту пројекта, док се за преглед свих колекција наплаћује годишња лиценца. Са друге стране, сви метаподаци су слободно доступни за преглед и преузимање. Успостављена сарадња Националне библиотеке и AiP Beroun представља пример добре координације послова јавног и приватног сектора на заједничком пројекту, где је Национална библиотека примарно допринела у домену метаподатака.⁸ Партнери у пројекту (преко 10 страних и пар десетина чешких институција) имају пун приступ систему и могу самостално да одлуче у којој мери ће својим корисницима омогућити слободан приступ колекцијама (прегледању докумената). Пројекат је започет још почетком 90-тих година 20. века, сарадњом са Унеском на програму „Памћење света“. Као део тог програма, 1995. године покренут је *Memoriae Mundi Series Bohemica*, пројекат који улази у редовно финансирање Националне библиотеке.⁹ Због богатства збирке рукописа НКП, ова збирка добила је приоритет у процесу дигитализације и она је у највећој мери заслужна за формирање центра за дигитализацију у НКП у првој половини 90-тих година 20. века.¹⁰

⁷ http://www.manuscriptorium.com/Site/ENG/default_eng.asp

⁸ Tamara Butigan-Vučaj, *Studio za digitalizaciju Nacionalne biblioteke Češke Republike*, Glasnik Narodne biblioteke Srbije 1/2005, Beograd, 2005, str. 295.

⁹ Манускрипториум је део ширег програма *Memoria*, а тренутно стање тог програма дигитализације старих историјских извора представља управо Манускрипториум. Bohdana Stoklasova, *Czech Digital Library*, <<http://www.svkbb.sk/colloquium/zbornik/data/stoklas.pdf>>; Adolf Knoll, Digitized Cultural Heritage, Pregled Nacionalnog centra za digitalizaciju 5 (2004), 63-66, <<http://elib.mi.sanu.ac.rs/files/journals/ncd/5/d009download.pdf>>.

¹⁰ Tamara Butigan-Vučaj, *нав. дело*, стр. 295-296.



Дигитална библиотека састоји се од каталога (ОСНВР - Open Catalogue of Historical Book Resources) и похрањених докумената у репозиторијуму за чување података (data storage system). Каталог информације о физичким документима (рукописима) приказује у облику каталогског записа, користећи XML фајл формат, повезујући метаподатке из централног каталога са дигиталним документима у репозиторијуму. Комуникација унутар читавог система успостављена је према међународним стандардима за размену података: Z39.50, OAI-MPH (The Open Archives Initiatives Protocol for Metadata Harvesting).¹¹ Манускрипториум укључује и виртуелно окружење за истраживачки рад из области историјских извора. Ту се подразумевају како алати за прегледање дигиталних слика, односно претраживање и читање текста, тако и линкови ка стручним радовима који се баве анализом заступљених рукописа, другим електронским изворима и метаподацима који ближе описују рукописе. Размена података са другим системима дигиталних библиотека омогућена је захваљујући коришћењу већ поменутих стандарда. Поред тога, пројекат подразумева укључивање у колекције и збирки из установа-партнера (библиотека, архива, музеја), али према стандардима и смерницама усвојеним на самом почетку рада Манускрипториума. Захваљујући свему описаном, доброј организацији и резултатима који углавном надмашују сличне програме у свету, овај пројекат прерастао је

¹¹ <http://www.loc.gov/z3950/agency/>; <http://www.openarchives.org/OAI/openarchivesprotocol.html>

националне оквири и данас представља програм дигитализације европске рукописне баштине.

Да би се обезбедила дугорочна дигитална заштита заступљених докумената, сама информација је независна од алата који су коришћени за њено стварање, односно који се користе за управљење подацима и њихов приказ. На тај начин се обезбеђује приступ информацијама у условима коришћења других технологија, односно новог софтвера и хардвера. Све информације чувају се у XML формату, са прихваћеном METS структуром фајлова Конгресне библиотеке. Описни метаподаци за рукописе користе MASTER стандард за приказивање у METS окружењу. MASTER стандард посебно је развијен од стране међународног пројектног тима у сарадњи са Оксфорд универзитетом, за опис рукописа, старих штампаних књига и других историјских извора.

Демо верзија¹² Манускрипториума предвиђена је за тестирање од стране нелиценцираних корисника, како би се стекао пун увид у карактеристике система. Ту је заступљен мањи број докумената, доступних за прегледање, са свим метаподацима и библиографским описом. Прегледање докумената врши се у посебној апликацији за прегледање, MS Viewer, која се отвара у новом прозору веб читача, док је за потребе стварања метаподатака развијена апликација M-Tool. M-Tool је предвиђен за рад библиотекара, архивиста и других стручњака који припремају описе писаних историјских докумената. Уз његово коришћење упрошћава се поступак повезивања описних метаподатака са дигитализованим сликама, сачуваних у већем броју формата. Подаци о описаним документима чувају се у XML формату, уз поштовање свих техничких критеријума за будуће укључивање у Manuscriptorium. M-Tool је бесплатан и може се преузети са сајта пројекта.

Пројекат Крамериус (Kramerius)

Пројекат Крамериус¹³ започет је 2003. године за потребе дигитализације монографских и серијских публикација. Крамериус је резултат сарадње НКП са Библиотеком Академије наука Чешке Републике и компанијом Qbizm Technologies, која је заслужна за развој Крамериус система за управљање документима (Document Management System). Крамериус систем представља основу истоименог пројекта дигитализације и дигиталне би-

¹² http://www.manuscriptorium.com/Site/ENG/free_documents.asp

¹³ <http://kramerius.nkp.cz/kramerius/Welcome.do;jsessionid=ECDD86375315A016447717588D603426?lang=en>.

блиотеке монографских и серијских публикација, старих часописа, листова, књига и других публикација штампаних на папиру који садржи киселину. Због тога се овај пројекат (подржан од стране норвешких државних фондова¹⁴) може сврстати у један од ретких националних пројеката у Европи који има за циљ да дигитализацијом публикација реши питање дугорочне заштите осетљивог материјала културне и научне баштине, који је подложен пропадању током времена. Крамериус систем за управљање документима (или дигитална библиотека за штампу периодике и монографске публикације, ради лакшег разумевања) може се слободно преузети са матичног сајта.¹⁵ Систем подржава OAI-PMH v2.0 протокол за електронске репозиторијуме, чиме је обезбеђена дугорочна компатибилност и интероперабилност са другим репозиторијумима заснованим на том протоколу. Док се процес стварања дигиталних копија одвија у више установа, остатак процеса дигитализовања одвија се централно у НКП.¹⁶

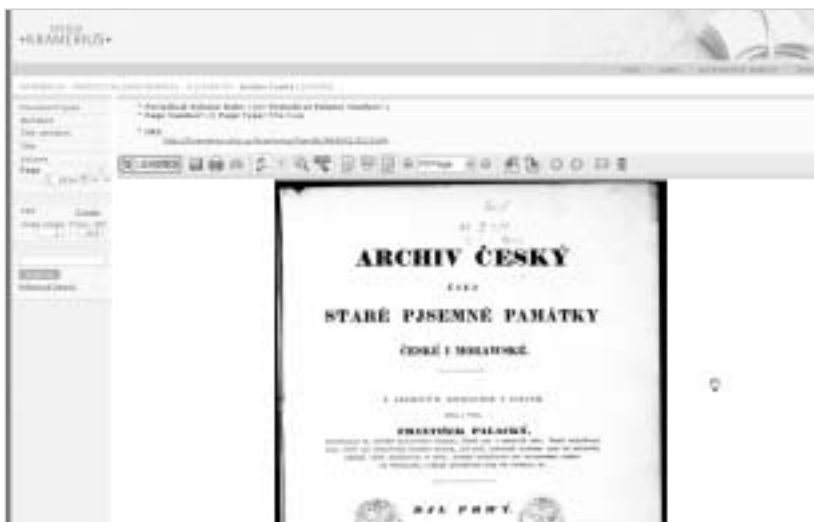
Пун приступ садржајима дигиталне библиотеке резервисан је за кориснике који користе приступне тачке у НКП (који могу прегледати и претраживати метаподатке и дигитализоване слике публикација), због присуства материјала који су заштићени ауторским правима. Сви остали имају пун приступ метаподацима (библиографским и другим подацима о публикацијама), као и пун приступ оним дигитализованим публикацијама које спадају у јавно добро. За приступ дигиталној библиотеци потребан је неки од савремених интернет (веб) читача, са инсталираним бесплатним додатком за прегледање слика у DjVu формату. Овај формат докумената изабран је због односа компресије (смањења меморијске запремине) и очувања квалитета приказа, који је у случају DjVu формата један од најповољнијих. DjVu је бесплатан формат и примарно је намењен за чување скенираних слика које садрже текст или графику. Карактерише га слојевити приказ садржаја скениране слике, где се слој са текстом одваја од слоја где се налази позадинска слика или графика. Квалитет приказа при максималној компресији података бољи је од конкурентских формата, а уз то је излазна величина документа мања у односу на PDF или JPEG документе. Оно што је недостатак тог формата управо се односи на неопходност додатне инсталације софтвера који ће омогућити приказ слика на рачунару корисника. DjVu формат није шире прихваћен као стандард за графички приказ слика на интернету (где доминирају JPEG и GIF формати), па веб читачи у највећем броју случајева неће ни приказати тражену слику, уколико се софтверски додаток не инсталира. С обзиром да постоје корисници који не знају да инсталирају додатне

14 <<http://www.eeagrants.org/projects>>.

15 <<http://kramerius.qbizm.cz/menu/Download/index.html>>.

16 Adolf Knoll, *nav. delo*, str. 64.

алате на рачунар, односно који то не могу да ураде (као што је случај у многим великим организацијама, где радници нису слободни да самовољно манипулишу рачунарима) или то једноставно не желе, употребљивост Крамериус система дигиталне библиотеке је тиме умањена.



Документи у оквиру Крамериуса могу се претраживати, при чему претраживач користи метаподатке о документима за резултате (претражују се ISSN односно ISBN број, наслов, поднаслов, име и презиме аутора, издавач, штампарија, година издавања и штампања...), а омогућена је и претрага пуног текста публикација. Подржана је претрага коришћењем Булових оператора, као и напредна претрага, коришћењем одређених метаподатака за лимитирање резултата претраге. Приликом једноставног претраживања корисницима се резултати приказују на основу релевантности резултата у оквиру пуног текста публикација, сортирани према меродавности. Резултат претраге отвара приказ дигиталне слике оне странице на којој се термин из упита налази. Приказ странице врши се кроз DjVu приказивач слика. Корисници кроз приказивач слика могу снимити приказани документ (слику), одштампати га, зумирати приказ, одабрати величину и начин приказа, ротирати слику, итд. Такође је подржан PDF формат, кроз опцију да корисници једним кликом дођу до PDF верзије документа. Међутим, пре тога се мора дефинисати највише 20 страница које се желе конвертовати у PDF формат, јер није могуће добити читаву публикацију у том формату (мада корисници могу из више делова то учинити).

База података пројекта Крамериус у јануару 2009. године имала је 540 наслова периодике из 19. и 20. века (са преко 3,8 милиона дигитализованих страница) и 6.083 монографских публикација (са 1,5 милион страница). Периодика је разврстана према наслову, док су монографске публикације разврстане по наслову и ауторима. Приликом избора одређеног наслова, кориснику се отвара прозор са приказом записа из базе података, који сажето приказује основне податке о публикацији (наслов, тип документа, физички опис, језик публикације, локацијске податке, јединствени идентификатор извора (Uniform Resource Identifier - URI) изабране публикације, као и линк ка дигиталној копији публикације. Овај линк отвара пун приказ метаподатака, где су, поред поменутих, излистани и упоредни наслов, издавач, година издавања, место издавања, физичко стање оригинала, УДК број. Исти прозор садржи и листу свих страница публикације у дигиталном облику, које се могу отворити и прегледати (уколико је књига или часопис отворен за приступ свим корисницима).

Сви метаподаци публикација заступљених у Крамериус пројекту дефинисани су кроз METS стандард. METS¹⁷ стандард (Metadata Encoding & Transmission Standard) развијен је у Конгресној библиотеци за кодирање описних, административних и структуралних метаподатака који описују објекте унутар дигиталних библиотека. Један METS фајл описује читаву публикацију и исписан је кроз XML језичку шему, а корисницима је омогућен увид и слободно преузимање фајла за изабрану публикацију.

Структура дела METS документа:

```
<mets:mets OBJID="ABA001/11547373" TYPE="MonographUnit" LABEL="Abbildungen und kurzgefaßte Geschichten verklaerter Diener und Freunde Gottes. - Jahrgang 2" PROFILE="Mets profile - mu" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd">
```

```
-
<mets:metsHdr CREATEDATE="2009-01-21T13:39:10">
```

```
-
<mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
```

```
<mets:name>ABA001</mets:name>
```

```
</mets:agent>
```

```
</mets:metsHdr>
```

```
-
<mets:dmdSec ID="DMD_MARC">
```

```
-
<mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="MARC" MIMETYPE="text/xml" LABEL="MARC XML">
```

```
-
```

¹⁷ <http://www.loc.gov/standards/mets/>

```

<mets:xmlData>
-
<marc:collection>
-
<marc:record>
<marc:leader>—nam-a22—uu-4500</marc:leader>
<marc:controlfield tag="001">dtd20080009786</marc:controlfield>
<marc:controlfield tag="003">CZ-PrNK</marc:controlfield>
<marc:controlfield tag="005">090121133910</marc:controlfield>
<marc:controlfield tag="008">090121b—xx—ger—</marc:controlfield>
-
<marc:datafield tag="100" ind1="1" ind2=" " >
<marc:subfield code="a">***Donator NF***</marc:subfield>
<marc:subfield code="e">Other</marc:subfield>
</marc:datafield>
-
<marc:datafield tag="245" ind1="0" ind2="0">
<marc:subfield code="n">Jahrgang 2</marc:subfield>
</marc:datafield>
-
[...]
```

Вебархив (WebArchiv) пројекат – дигитална архива националног домена

Пројекат Вебархив¹⁸ представља дигиталну архиву за дугорочну заштиту чешких интернет извора. Пројектом руководи Национална библиотека Чешке Републике од 2000. године, у сарадњи са Моравском библиотеком (задуженом за техничка решења) и Институтом за информатику Масариковог универзитета у Брну. Тренутно се (у 2008) Вебархив води као пројекат у НКП, мада је суштински постао организациони део НКП и члан је Међународног конзорцијума за заштиту интернета (International Internet Preservation Consortium, ИПС¹⁹) од 2007. године. Финансирање пројекта током свих година вршено је скоро искључиво кроз различите донације.

Основни циљ чешког пројекта веб архивирања је заштита и очување дигиталних докумената културне, уметничке и историјске вредности на интернету за будуће генерације корисника. Услед наглог развоја интернета и ефемерног карактера многих докумената и електронских извора на мрежи, који брзо настају, али исто тако брзо и нестају, НКП је у сарадњи са

¹⁸ <http://www.webarchiv.cz/>

¹⁹ <http://netpreserve.org/about/index.php>

другим институцијама развила и прилагодила модел и алате за дугорочну заштиту тих докумената и извора (као што су генератор метаподатака према Dublin Core стандарду,²⁰ Uniform Resource Name-URN генератор,²¹ калкулатор за MD5 криптографску функцију²²). Међу сачуваним изворима могу се наћи појединачни документи, публикације, текстуални, видео и аудио фајлови и читави веб сајтови, уколико испуњавају критеријуме за архивирање. Основни критеријум да извор буде укључен у процес архивирања је припадност чешком националном интернет домену .cz. Посебни критеријуми се могу дефинисати према расположивом простору за складиштење података, типу докумената, величини (документи до 100 мегабајта), формату, протоколу за пренос података (за сада само http/ftp протоколи), броју докумената (до 5.000 по једној веб адреси), итд. Селективно (тематско) архивирање узима у обзир додатне критеријуме, као што су садржај који одговара „националном аспекту“, седиште издавача садржаја, језик, слободно доступан оригинални садржај значајне културне или научне вредности.

Аутори пројекта покушали су двојним приступом да реше проблем проналажења, издвајања и архивирања извора на интернету, кроз комплетно архивирање чешког интернет домена у одређеном периоду сваке године (започето септембра 2001), односно кроз селективно претраживање (четири пута годишње), издвајање и архивирање интернет извора на одређену тему (приликом чега се формирају тематске колекције, попут колекције председничких избора у Чешкој 2008. године, што је популаран избор у многим пројектима интернет архивирања широм света). Као илустрација количине података прикупљених кроз комплетно архивирање чешког интернет домена, крајем јануара 2009. године укупно је архивирано 206,8 милиона докумената, који запремају 8,8 терабајта података.

С обзиром да је материјал на интернету такође заштићен ауторским и власничким правима (аутора, издавача, власника интернет домена), држава је правно регулисала то питање на тај начин што је НКП законом добила слободу да се бави активностима веб архивирања и да ствара своје колекције, али да приступ колекцијама корисницима на интернету буде ограничен само на оне делове за које су носиоци ауторских права дали писмену дозволу. Вебархив у највећем броју случајева склапа уговор са издавачима веб сајтова, којим се регулише будуће укључивање тог интернет извора у јавно доступну колекцију на сајту пројекта. У тренутку писања овог текста (јануар 2009) корисницима је путем интернета било могуће да прегледају колекцију од 800 архивираних веб сајтова, који припадају култур-

20 <http://www.webarchiv.cz/generator/dc.php?lang=en>

21 <http://en.webarchiv.cz/urn-generator/>

22 <http://en.webarchiv.cz/md5-generator/>

но-научној баштини чешког народа у матици и расејању. Уколико издавачи траже, Вебархив доставља метаподатке за њихов архивирани веб сајт или интернет садржај у Dublin Core стандарду.²³



Веб архивирање²⁴ (Web Archiving) се дефинише као складиштење у дигиталном репозиторијуму целине или одређеног дела интернет садржаја, зарад обезбеђивања трајног приступа тим дигиталним садржајима у будућности. Услед несагледивог раста садржаја на интернету, који се експоненцијално шири, веб архивирање постаје актуелна тема у многим земљама света, које посматрају своје националне интернет домене као саставни део културне баштине (што ће највероватније прихватити већина држава током времена, како се сразмерно буде повећавало коришћење и присуство на интернету). Овде је у питању веома посебна врста културне баштине, која не постоји у физичком (аналогном) облику, па се и њено дугорочно чување (архивирање) мора решити на другачији начин. Нарочито ако се узме у обзир променљивост садржаја на интернету, са многим изворима и сајтовима који постоје у врло кратком временском периоду (свега неколико месеци, као у случају одређених манифестација, где се након окончања програма често угаси/обрише и интернет запис о томе). Истраживања од пре неко-

23 Petr Žabička, *WebArchiv – Czech Web Archive revisited*. <http://www.iwaw.net/06/PDF/iwaw_06-zabicka.pdf>.

24 http://en.wikipedia.org/wiki/Web_archiving

лико година утврдила су да просечни (активни) животни век нових интернет сајтова не превазилази дужину живота једне муве, тачније 44 дана.²⁵ Постоје случајеви масовног одумирања интернет простора, као у случају Србије гашење .yu домена, када ће 30. септембра 2009. неминовно нестати део националне „дигиталне културе“.²⁶

Као и већина новина, с којима се установе културе хватају у коштац, и овде постоји много експериментисања и вишегодишњих припремних радова, тако да је врло мало земаља изградило националну стратегију веб архивирања целог или дела свог националног домена. Занимљиво је да ниво технолошке развијености или културе коришћења интернета не одређује и преимућство, јер у суштини процес „хватања“ садржаја са интернета није претерано компликован (постоје и бесплатни програми који се могу инсталирати на кућном рачунару и који пренесу на локални диск читав садржај неког сајта или интернет ресурса). Многобројне фазе морају се проћи пре саме „жетве“ интернета (дословни превод енглеског термина harvest): стварање правног оквира који би неким установама омогућио да приступају садржајима под заштитом ауторских и власничких права и преузимају их у целости (у суштини најдужа фаза); стварање и организовање националног репозиторијума, што није само питање куповине неког система великог капацитета за складиштење података, већ начина како ће се омогућити дугорочни и сигурни приступ подацима у будућности, уз гарантовање аутентичности тих података; критеријуми за селекцију садржаја који ће бити архивиран (ни овде не треба све архивирати, већ само оно што се сматра корисним за будуће кориснике); решавање проблема застаревања софтвера и хардвера, тј. миграција података на нове платформе и системе, како генерације после нас не би имале само музејске примерке хард дискова, са нечитљивим подацима на њима; итд. Након што се реше ова и друга питања, може се приступити планском раду на дигиталној заштити (Digital Preservation), односно управљању дигиталних информација током времена (или на дуги рок).

У техничком смислу, веб архивирање је аутоматизован процес који спроводи неколико софтверских алата за прикупљање, индексирање и чување података у архиви, према унапред дефинисаним параметрима. Вебархив пројекат се ослања на софтвер отвореног кода Heritrix²⁷ за идентификовање и прикупљање извора са интернета. Поред тога је дефинисан скуп критеријума за селектовање интернет-насталих (online-born) докумената, како би били укључени у чешку националну библиографију. У том контексту правно питање веб архивирања постаје веома битно, јер законодавац у

²⁵ <http://www.loc.gov/acq/devpol/webarchive.html>

²⁶ http://www.rnids.rs/tacka_na_yu

²⁷ <http://crawler.archive.org/>

Чешкој Републици још увек није предвидео обавезни електронски примерак као у случају штампаних публикација.

Прикупљени документи са релевантним метаподацима чувају се у стандардизованом архивском формату ARC, који подржава ИПС конзорцијум. ARC²⁸ је формат за архивирање података компресијом без губитака, који се користи за складиштење и чување интернет извора у многим пројектима веб архивирања. Развијен од стране Internet Archive²⁹ за рад са Xeritrix програмом-трагачем (crawler), ARC се примарно користи за складиштење блокова информација прикупљених од стране робота за претрагу и прикупљање извора. У оквиру сваког ARC фајла налазе се архивирани документи (html, gif, jpeg...) који чине структуру и садржај интернет страница. Сваком од докумената у ARC структури претходи информација у заглављу о формату, величини, спољним линковима у оквиру документа, итд. ARC се показао као одличан формат за масовно архивирање интернет садржаја (као што је национални интернет домен), када контрола процеса не може да подразумева појединачно праћење квалитета и исправности архивираних материјала. У циљу што бољег дугорочног веб архивирања из ARC формата развијен је WARC³⁰ формат, за који се очекује да постане стандард у будућности.

Структура архивираних HTML документа у ARC формату:

```
filedesc://IA-001102.arc 0 19960923142103 text/plain 76
1 0 Alexa Internet
URL IP-address Archive-date Content-type Archive-length
http://www.dryswamp.edu:80/index.html 127.10.100.2 19961104142103 text/html 202
HTTP/1.0 200 Document follows
Date: Mon, 04 Nov 1996 14:21:06 GMT
Server: NCSA/1.4.1
Content-type: text/html Last-modified: Sat,10 Aug 1996 22:33:11 GMT
Content-length: 30
<HTML>
Hello World!!!
</HTML>
filedesc://IA-001102.arc 0.0.0.0 19960923142103 text/plain 200 - - 0
IA-001102.arc 122
2 0 Alexa Internet
URL IP-address Archive-date Content-type Result-code Checksum
Location Offset Filename Archive-length
http://www.dryswamp.edu:80/index.html 127.10.100.2 19961104142103
text/html 200 fac069150613fe55599cc7fa88aa089d - 209 IA-001102.arc 202
```

28 <http://www.digitalpreservation.gov/formats/fdd/fdd000235.shtml>

29 <http://www.archive.org/index.php>

30 <http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>

```

HTTP/1.0 200 Document follows
Date: Mon, 04 Nov 1996 14:21:06 GMT
Server: NCSA/1.4.1
Content-type: text/html Last-modified: Sat,10 Aug 1996 22:33:11 GMT
Content-length: 30
<HTML>
Hello World!!!
</HTML>
    
```

Информациони портал Крамериус система

Крајем 2008. године Национална библиотека Чешке Републике покренула је „Информациони портал Крамериус система“.³¹ Портал је и даље у развојној фази и доступан је, за сада, само на чешком језику. Предвиђено је да се преко овог портала повежу описани пројекти дигиталне заштите (Manuscriptorium, Kramerius, WebArchiv), као и да се на једном месту обједине информације о прошлим и актуелним активностима дигиталне заштите у НКП и Чешкој Републици. Информациони портал ће, такође, најављивати све предстојеће догађаје у земљи из области дигитализације. Поред информација и линкова ка актуелним пројектима, корисници ће имати приступ релевантној литератури и документацији о дигитализацији у Чешкој Републици.



31 <http://kramerius-info.nkp.cz/>

О конзорцијуму DigitalPreservationEurope

DigitalPreservationEurope³² (DPE) је дугорочна иницијатива за унапређење координације, кооперације и конзистентности у постојећим активностима на обезбеђивању ефектне заштите дигитализованих извора. DPE подстиче сарадњу и размену знања, идеја и практичних решења између више националних програма дигиталне заштите³³ (digital preservation) широм Европе. На самом веб сајту конзорцијума налази се страница која дефинише дигиталну заштиту као „скуп активности потребних зарад лоцирања, приказивања, коришћења и разумевања у будућности“.³⁴ DPE је практично наставак ERPANET пројекта³⁵ (Electronic Resource Preservation and Access Network), који је финансирала Комисија ЕУ (The European Commission) за потребе успостављања конзорцијума установа културе, софтверских компанија, истраживачких институција, невладиног сектора, у циљу развоја најбољих теоријских и практичних оквира дигиталне заштите културног и научног наслеђа.

Укупно је 11 институција заступљено као партнери у пројекту, углавном националних установа европских земаља из сектора културе и науке. Још је већи број тзв. „придружених партнера“ (укупно 20 у јануару 2009), јер је комплексност проблема дигиталне заштите и обимност послова условила интердисциплинарну сарадњу међу многобројним институцијама различитих профила и активности. DPE је покренуо и сопствени програм размене професионалаца из области дигиталне заштите (DigitalPreservationEurope Exchange Programme – DPЕХ), како би појединци стекли прилику да у многобројним установама Европе стекну драгоцену искуство, размене идеје и усаврше знање, односно праксу. Програм је био отворен до краја 2008. године и омогућавао је да примљени кандидати бораве у некој установи у периоду између две недеље и три месеца (коју су сами изабрали и контактирали, како би добили званичан позив). Организација DPE стипендираним појединцима дарежљиво је покривала све трошкове пута, смештаја и боравка (до одређеног максимума).

³² <http://www.digitalpreservationeurope.eu/>

³³ „Digital Preservation“ смо слободно превели као „дигитална заштита“, јер сматрамо да одговара смислу коришћења тих термина на енглеском језику. „Библиотекарски терминолошки речник“ Народне библиотеке Србије (2004, интернет издање на <http://btr.nbs.bg.ac.yu/>) преводи Digital Preservation као „дигитализација ради заштите“, што сужава значење оригинала, јер заштита у дигиталном облику не мора нужно да подразумева сам процес дигитализације (као што је случај код материјала који постоји само у дигиталном облику – нпр. „електронски рођени документи“) (прим. аут.).

³⁴ <http://www.digitalpreservationeurope.eu/what-is-digital-preservation/>

³⁵ <http://www.erpanet.org/index.php>

Литература

1. Butigan-Vučaj, Tamara: *Studio za digitalizaciju Nacionalne biblioteke Češke Republike*, Гласник Народне библиотеке Србије 1/2005, Београд, 2005, стр. 295-301.
2. Žabička, Petr: *WebArchiv - Czech Web Archive revisited*, <<http://www.iwaw.net/06/PDF/iwaw06-zabicka.pdf>>.
3. Knoll, Adolf: *Digitized Cultural Heritage*, Pregled Nacionalnog centra za digitalizaciju 5 (2004), str. 63-66, <<http://elib.mi.sanu.ac.rs/files/journals/ncd/5/d009download.pdf>>.
4. Kovačević, Ljiljana, Injac, Vesna, Begenišić, mr Dobrila: *Bibliotekarski terminološki rečnik : englesko-srpski, srpsko-engleski*, Београд, 2004. < <http://btr.nbs.bg.ac.yu/>>.
5. Rothenberg, Jeff: *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. A Report to the Council on Library and Information Resources, Washington, 1999. < <http://www.clir.org/pubs/reports/rothenberg/contents.html>>.
6. Stoklasova, Bohdana: *Czech Digital Library*, <<http://www.svkbb.sk/colloquium/zbornik/data/stoklas.pdf>>.
7. Thibodeau, Kenneth: *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*, CLIR Reports < <http://www.clir.org/pubs/reports/pub107/thibodeau.html>>.
8. http://en.wikipedia.org/wiki/Web_archiving
9. <http://kramerius.nkp.cz/kramerius/Welcome.do;jsessionid=ECDD86375315A016447717588D603426?lang=en>
10. <http://kramerius-info.nkp.cz/>
11. <http://crawler.archive.org/>
12. <http://www.archive.org/index.php>
13. <http://www.digitalpreservation.gov>
14. <http://www.digitalpreservationeurope.eu/>
15. <http://www.eeagrants.org/projects>
16. <http://www.loc.gov/acq/devpol/webarchive.html>
17. <http://www.loc.gov/z3950/agency/>
18. <http://www.loc.gov/standards/mets/>
19. http://www.manuscriptorium.com/Site/ENG/default_eng.asp
20. <http://netpreserve.org/about/index.php>
21. <http://www.nkp.cz/>
22. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
23. http://www.rnids.rs/tacka_na_yu
24. <http://www.csa.com/discoveryguides/cyber/overview.php>
25. <http://www.webarchiv.cz/>

DIGITAL PRESERVATION PROJECTS AT NATIONAL LIBRARY
OF THE CZECH REPUBLIC*Summary*

The National Library of the Czech Republic (NKP) started the first steps in digital preservation of Czech scientific and cultural heritage in the beginning of the 90's. During the time three projects have stabilized among other digital preservation activities at NKP: Manuscriptorium for digitized manuscripts and old printed books, Kramerius for monographic and serial publications and Web-Archiv as digital archive of online-born documents. This paper reviews these projects, but also presents the main information about present and future work on digital preservation at NKP.

Key words: digital preservation, digitization, Manuscriptorium, Kramerius, WebArchiv, web archiving, national library, Czech Republic, DPE

Bogdan Trifunovich, M. A.