

Организовање дигиталног репозиторијума у Институту за српски језик САНУ

УДК 061.6: 811.163.41]:004.738.5

Владимир Живановић
vladimirludwig@yahoo.com
Институт
за српски језик САНУ
Београд, Србија

САЖЕТАК: Овим радом желели бисмо да представимо случај прикључивања Института за српски језик САНУ (ИСЈ САНУ) свету дигиталних репозиторијума. Године 2018. Министарство просвете, науке и технолошког развоја Републике Србије усвојило је Платформу којом обавезује све јавно финансиране институције да своје научне радове депонују у неки репозиторијум. Намера нам је да представимо основне фазе успешно спроведене Платформе Министарства у ИСЈ САНУ. У периоду од једне године у ИСЈ САНУ припремљено је преко 100.000 страница дигиталног текста које чине 2.500 записа пуног текста. Овим подухватом реализованим у оквиру концепта отворене науке омогућено је да српска језичка хуманистика постане видљива у свету.

КЉУЧНЕ РЕЧИ: дигитални репозиторијум, институционални репозиторијум, ДАИС, отворена наука, самоархивирање, дигитална хуманистика.

РАД ПРИМЉЕН: 10. мај 2019.

РАД ПРИХВАЋЕН: 18. јуни 2019.

1. Увод у процес дигитализације

Године 2018. Министарство просвете, науке и технолошког развоја усваја Платформу за отворену науку (**Платформа, 2018**). Ова Платформа, која се заснива на принципима отворене науке и

смерницама Европске комисије у тој области, уводи обавезу за појединце и институције да део своје научне продукције који је јавно финансиран учини доступним широј јавности депоновањем у неки дигитални репозиторијум. Како би се поштовала „транспарентност научне комуникације и методологије“ и обезбедила доступност научне продукције, било је потребно приступити развоју одговарајуће дигиталне инфраструктуре, која у Србији није постојала.

Отворени приступ у овом документу дефинише се као „право сваког корисника интернета да без финансијских издатака чита, преузима, чува, штампа и користи дигитални садржај публикација“. Поред ових права, корисници ће уз поштовање ауторских права, тј. лиценце која је придружена депонованом раду и уз правилно навођење извора, суделовати у ширењу идеја отворене науке, користећи неки од „институционалних/тематских/националних репозиторијума“. Тип репозиторијума за депоновање научне продукције није строго одређен па је тако остављено институцијама да нађу адекватно решење у складу са својим могућностима. На основу ове Платформе, обавеза је да свака од обухваћених институција усвоји „локалну“ политику, тј. одговарајући документ о спровођењу Платформе којим ће регулисати њено спровођење на институционалном нивоу.

У исто време, током припреме ове Платформе, и Српска академија наука и уметности (САНУ), у сарадњи са Рачунарским центром Универзитета у Београду (РЦУБ) ради на развијању сопственог репозиторијума који би био намењен Академији и припадајућим институтима. Репозиторијум је назван Дигитални архив издања САНУ и Института – ДАИС.¹ Према дефиницији Џозефа Бранина, репозиторијуми представљају „моделе система и услуга који су осмишљени да прикупе, организују, ускладиште, деле и чувају дигиталне информације и знања институције“ (Branin, 2004-2005, 237). Сврха дигиталног репозиторијума ДАИС је да омогући Академији, као и Институтима којима је Академија оснивач, да у њему трајно похрањују своју научну продукцију те тако јавно представљају резултате свога научног рада.

Развој комуникационих процеса омогућава већу доступност научних радова у пуном тексту. Општа доступност је и даље идеал према коме се тежи, али научне институције подстакнуте бројним иницијативама из домена отворене науке, као и праксом великих сервиса технолошких

¹ ДАИС (на вебу)

компанија (Google books и др.) почеле су да заснивају своје дигиталне репозиторијуме. Један од резултата успешно вођеног репозиторијума је повећање видљивости научних радова, посебно из области српске хуманистике. Због комерцијалног потенцијала техничких и природних наука, њихова видљивост у свету је много већа у односу на хуманистички домен. Оваква несразмера између природних и техничких наука није оправдана, с обзиром да оно што је локално представља једну културу на глобалном нивоу.

За разлику од класичне библиотеке, формирање дигиталног репозиторијума захтева техничку припрему. ДАИС је настао прилагођавањем софтвера отвореног кода DSpace.² Прилагођавање овог софтвера подразумевало је његову локализацију, адаптацију корисничког интерфејса, усклађивање са смерницама конзорцијума OpenAIRE за дигиталне репозиторијуме, интеграцију са сервисима ORCID и Altmetric.com и развој додатних екстерних апликација које омогућавају нормирање имена, као и преузимање записа и масован унос и корекције метаподатака.³

Разлог за стварање оваквих софтвера почива на развоју библиотечког пословања. Бодвајн и Бранскофска, са МИТ-а, тврде да универзитетске заједнице зависе од својих библиотека које им омогућавају стални приступ истраживању и научном раду, и то их ставља у позицију да траже решења за проблем складиштења и преузимања интелектуалног рада на дуже стазе (Baudoin and Branschovsky, 2004, 32). Имплементирање и развој софтвера, као и његово прилагођавање специфичним потребама српске научне продукције извршио је РЦУБ. У рачунарском центру се стално ради на одржавању репозиторијума и његовом континуираном усавршавању.⁴

Потребно је уочити разлику између дигиталне колекције и дигиталног репозиторијума. Док колекција остаје локална, какве су нпр. колекције дигиталних објеката које на своје веб-сајтове

² Dspace софтвер отвореног кода је развио Масачусетски институт за технологију (МИТ) заједно с фирмом Hewlett-Packard 2002. године. Његов квалитет види се у лакој и отвореној приступу различитим типовима дигиталних објеката: тексту, сликама, аудио- или видео-материјалу.

³ За више о DSpace софтверу као ослонцу за развој дигиталних репозиторијуми видети (Rajović and Ševkušić, 2018)

⁴ Прилагођавање софтвера отвореног кода локалним потребама је највећи изазов у почетним фазама организовања репозиторијума и захтева квалитетну техничку подршку. На коду софтвера се континуирано ради.

постављају институције или појединци, она не задовољава стандарде интероперабилности (преко DOI броја, других перзистентних идентификатора, метаподатака и атрибута, структуре и стандарда за описивање документа, протокола за размену метаподатака и др. (Van de Sompel and Nelson, 2015)). Дигитални репозиторијум пак представља дигиталну колекцију која је организована тако да се његов садржај може трајно архивирати, чувати и дисеминирати кроз оријентисане циљеве и протоколе компатибилне с другим базама података. Његова сврха је „представљање интелектуалног учинка“ и промовисање доступности (Winter and Bowen-Chang, 2010, 320).

Искуство ИСЈ САНУ у организовању и вођењу дигиталног репозиторијума може послужити другим институцијама које се укључују у организовање дигиталних колекција с пуним текстом у отвореном приступу.

2. Структура

Како би институционални репозиторијум заживео, његово попуњавање је прва највећа препрека. Зато смо у ИСЈ САНУ приступили двома фазама попуњавања. Правилно организован репозиторијум повећава видљивост научне продукције одређене установе, те тиме омогућава и бољу цитираност истраживача (Piwowar and Haustein, 2018).

Прва фаза подразумева прикупљање, дигитализовање и постављање дигиталног садржаја часописа и монографија ИСЈ САНУ. Историја периодичних публикација Института веома је дуга. Два од четири научна часописа које Институт објављује излазе дуже од сто година.⁵ Објављени томови ових часописа броје неколико стотина свезака. Поред серијских публикација у првој фази у институционални репозиторијум уносили смо научне радове истакнутих научника Института. Ту су убројане колеге које су у пензији, а неки од њих нису више с нама. Уношењем ових дигитализованих дела успешно се испуњава неколико критеријума за функционисање репозиторијума: успоставља се добра пракса будућег рада, учествује се у формирању навика будућих

⁵ Први број часописа *Српски дијалектолошки зборник* изашао је 1904. год., *Јужнословенског филолога* 1913., часописа *Наш језик* 1933. године, и *Лингвистичких актуелности* 2000.

корисника, представља се научна продукција у најбољем светлу.⁶ Тиме се покрива и дијахронија научне продукције Института.

Друга фаза, која представља синхронијски ниво, подразумева објављивање текуће научне продукције, најкасније 18 месеци од датума објављивања научног чланка или монографије, и дефинисана је Правилником који је усвојен на нивоу Института. Тиме се испуњавају услови које прописује Платформа Министарства.

Институције које приступају формирању репозиторијума морале би да дефинишу своје циљеве те да јасно идентификују скуп особина добро обученог особља које ће учествовати у процесу одржавања репозиторијума (*Winter and Bowen-Chang, 2010, 323*). Обука кадрова за рад у репозиторијуму може да представља изазов. Недостатак иницијативе и ентузијазма код прихватања другачијег приступа научним публикацијама, те неразвијена свест о његовом значају, може да се одрази у првим фазама организовања репозиторијума. Веома је важно да руководећи кадар, као и сарадници научне установе, увиде важност успостављања репозиторијума.

Сваки научни сарадник у Институту дужан је да уноси своје научне радове кроз поступак самоархивирања. За истраживаче је организована обука, у виду радионица које садрже и практични део (самосталан унос метаподатака и депоновање објеката), а припремљено је и детаљно упутство у виду PowerPoint презентације које могу да користе у самосталном раду. Након што истраживач депонује рад и опише га свим припадајућим метаподацима, библиотекар као администратор контролише тачност уноса и одобрава запис, и једини има могућност да накнадно мења метаподатке или уклони дигитални објекат.⁷

Скуп метаподатака подразумева уношење основних података о документу, текст сажетка и навођење кључних речи, те одређивање степена доступности документа и типа лиценце. Такође се обавезно уносе и подаци о пројекту у оквиру ког је рад финансиран, па се, поред тога, подаци о типу и верзији документа, доступности садржаја и лиценци уносе у складу са актуелним смерницама за дигиталне репозиторијуме конзорцијума OpenAIRE. Домен који задобија све већи

⁶ За разлику од техничких наука у лингвистици велики део старих научних радова не губе на значају зато што су за истраживања и даље релевантни.

⁷ ДАИС репозиторијум претпоставља више нивоа приступа и дозвола за управљање садржајем тако да администратор има могућност да ограничи овлашћења корисницима када самоархивирају, и тиме обезбеди контролу квалитета, тачност података и интегритет базе.

значај је надгледање поштовања ауторских права и ступање у контакт са часописима и издавачима ради регулисања ауторских права.

Такође, библиотекар користи додатне сервисе ДАИС-а, као што су сервис за масовно едитовање метаподатака, нормативну базу, и сервис за преузимање постојећих дигиталних садржаја из других репозиторијума и сл. Тим за развој репозиторијума РЦУБ-а је самостално развио апликацију Ellena2 у којој су интегрисани нормативна датотека, сервис за масовно кориговање метаподатака и сервис за масовни увоз метаподатака. У раној фази развоја, постојале су две одвојене апликације: Ellena (нормативна датотека и масовно кориговање метаподатака) и MultiLoad (преузимање записа из других DSpace репозиторијума и масовни увоз метаподатака у XML или RIS формату). Почетком 2019. године, ове две апликације су интегрисане у апликацију Ellena2. Разлог зашто се приступило изради самосталне апликације јесу недостаци платформе DSpace, која сама по себи не садржи модул са нормативном датотеком. Поред тога, иницијални тестови су показали да систем за преузимање метаподатака који је већ био уграђен у DSpace не даје задовољавајуће резултате.

Руковођење садржајем омогућава сарадницима да репозиторијум користе у ширем распону него што их Платформа отворене науке обавезује јер ДАИС подржава унос и обраду више типова докумената и формата. Поред научног чланка, поглавља у монографији, монографије и сл., сарадници су у могућности да депонују и препринт верзију научног рада, докторске или магистарске тезе, извештаје и све друге садржаје који се називају „сивом“ литературом.⁸ У DSpace-у индексира се сав читљив текст из депонованих докумената, па се самим тим може и претраживати. Рангирање индексираних докумената врши се преко формалних параметара, тј. по релевантности, наслову и датуму. DSpace подржава похрањивање практично свих формата. У DSpace-у не постоји систем за прегледање докумената на онај начин на који он функционише у Omeki и другим платформама које су првенствено намењене приказивању културне баштине и дигитализоване грађе. DSpace је намењен презентацији научних садржаја, а похрањене датотеке се могу снимити на рачунар корисника и оданде отворити помоћу одговарајуће апликације. PDF, JPG и слични формати отварају се у новом прозору интернет претраживача. DSpace се може надоградити тако да у оквиру

⁸ За више о научној „сивој“ литератури видети (Ferrerias-Fernández and Merlo-Vega, 2015) и (Ćirković, 2018)

саме платформе приказује садржаје (нпр. листање књига), али нама, а ни корисницима, за сада то није потребно.

3. Пракса

Пред крај 2017. године у ИСЈ САНУ приступило се раду на попуњавању ДАИС-а. С обзиром на релевантност часописа ИСЈ САНУ у целом славистичком свету, један део годишњака већ је био скениран у оквиру сервиса Google Books, али није био јавно доступан.⁹ Више америчких универзитета је у оквиру својих колекција у сарадњи са Гуглом дигитализовало наше часописе. Те дигиталне примерке смо прикупили и они су чинили отприлике 50% од укупног броја свезака часописа. Остатак је требало дигитализовати. Приступило се послу који је подразумевао организовање, скенирање, редиговање дигиталних копија, рад на оптичком препознавању карактера садржаја и његово постављање у базу. На тај начин је за годину дана припремљено преко 50.000 страница текста. Недостатак инфраструктуре у одређеним сегментима посла надоместила је сарадња са сервисом Google Books: захваљујући томе што смо скениране публикације учинили доступним и на овом сервису, успели смо да скенирани материјал, који је био веома обиман у меморијском смислу, смањимо на оптималне величине,¹⁰ као и да обезбедимо оптичко препознавање текста.¹¹

У првој фази досадашњу продукцију Института учинили смо јавно доступном депоновањем целих годишта периодичних публикација (као

⁹ Једна од основних мисија сервиса Google Books је организација информација на светском нивоу. Дигитализацијом светске библиотечке баштине компанија Google постаје универзална библиотека светског знања. За више о партнерском програму са сервисом Google Books видети (Шевкушић, 2013)

¹⁰ Како би учинили наша документа читљивим и смањили величину фајлова (јер су неки имали и по 600 Мб) а то све оптерећује како репозиторијум и његове ресурсе, тако и крајњег корисника који жели да добије документ оптималне величине за кратко време. На пример, једно годиште часописа запремало је у просеку 200 и 300 Мб, али је након Гугловог оптичког препознавања карактера документ смањен на отприлике 30 Мб.

¹¹ До сада није урађена евалуација успешности оптичког препознавања текста. Када буду биле доступне и приступачне услуге домаће иницијативе која се бави оптичким препознавањем текста, посебно ћириличног, радо ћемо поново приступити OCR-у, те у репозиторијум депоновати документа са прецизније препознатим текстом.

једног документа), док је за касније остало да се та годишта поделе на чланке, који би се потом појединачно депоновали и тако учинили видљивијим. Поред часописа и монографска продукција је унета у ДАИС и обухвата преко 60 свезака. Овим је део прве фазе био готов и Институт је у ДАИС-у имао представљену ретроспективу своје издавачке делатности (250 томова часописа и монографија).

Други сегмент у првој фази успостављања ДАИС-а представљао је сакупљање и припремање те објављивање чланака истакнутих научника ИСЈ САНУ. У првих годину дана је у репозиторијум унесено више од 1.200 научних чланака.¹² Мали број ових чланака је већ постојао у дигиталној форми те је цео посао дигитализације требало да прође процес физичке припреме и рада на скенирању, сређивању скенираних докумената и креирању металодатака. Према Фостеру и Гибонсу институционални репозиторијум „без садржине, је исто што и низ празних полица“ (Gibbons and Foster, 2005, пар. 4).¹³ Зато је за формирање корисничких навика и успостављање доброг примера од суштинске важности оснаживање почетне динамике једног репозиторијума.

Креирање и организовање дигиталног садржаја у ДАИС-у усклађено је са захтевима Европске комисије у вези са отвореним приступом публикацијама, и омогућава даље ширење научних информација преко европских и међународних портала као што су OpenAire, BASE, CORE, Google Scholar. Једно од поља за унос метаподатака бележи име пројекта на коме је рад настао. То поље, преко линка, повезује све радове једног пројекта у ДАИС-у, али и у бази OpenAire, која на једном месту прикупља податке о резултатима свих пројеката Европске комисије, као и податке о резултатима националних пројекат из десетак земаља, међу којима је и Србија (OpenAire, 2019). Такође, коришћењем идентификатора ORCID (Open Researcher and Contributor ID) на једном месту се окупују и повезују све публикације аутора, и када су објављене под различитим варијантама именима.

¹² Међу њима су чланци академика Ирене Грицкат, академика Митра Пешикана, др Ивана Поповића, др Берислава Николића, научног саветника др Егона Фекета, проф. др Димитрија Е. Стефановића, академика Александра Ломе, научног саветника др Јасне Влајић-Поповић, научног саветника др Стане Ристић, научног саветника др Срета Танасића и др.

¹³ Као пример добре праксе узели бисмо репозиторијум Масачусетског технолошког института (МИТ) који до 2019. године у свом репозиторијуму поседује преко 106.000 дигиталних записа (DSpace, 2019).

При одређивању профила дигиталне колекције ИСЈ САНУ се одлучио за обезбеђивање пуног текста у отвореном приступу за сва своја издања, као и за објављене радове запослених уз ограничења која одређују ауторска права других издавача. Одлучено је да се на издања Института примењују СС лиценце (Creative Commons). Овај једноставан и стандардизован начин одређивања ауторских права има више нивоа. Институт се одлучио за модул СС BY-NC-ND (Ауторство – Некомерцијално – Без прерада), што подразумева да је обавезно навођење имена аутора и да аутор даје дозволу за преузимање дела и његову даљу дистрибуцију. Ова најрестриктивнија од слободних лиценци не дозвољава даљу прераду дела нити комерцијалну употребу.

Софтвер DSрpace у својој архитектури омогућава стварање колекција у дигиталном репозиторијуму. Ове колекције категоришу садржај, те га тематски, формално или на други начин разврставају. ИСЈ САНУ у ДАИС-у тренутно има седам колекција. Поред четири колекције које разврставају периодичку Института, две су намењене монографским серијама, док је једна општа. Општа колекција сабира зборнике Института, друге материјале и радове истраживача објављене изван Института. Проблем класификације увек намеће разна решења. Грађа је могла бити и другачије разврстана, али оптимални приступ води рачуна и о лакој сналажењу крајњих корисника.

4. Закључак

На пољу дигиталне хуманистике у ИСЈ САНУ за релативно кратко време дошло је до великих промена. У пракси, институционални репозиторијум је заживео и служи за више намена. Главна намена је промоција концепта отворене науке и представљање научне продукције установе, чиме се испуњавају услови које је Министарство прописало Платформом. Оно што је јавно финансирано требало би да буде и јавно доступно. Други правац којим се институционални репозиторијум креће јесте и представљање ретроспективне научне продукције. Објављено је готово све што је ИСЈ САНУ публиковао. Тиме се стари научни радови који су и даље релевантни чине лако доступним. Библиотека радом репозиторијума партиципира у научном истраживању. Према Бодвајн и Бранскофској, заснивањем дигиталног репозиторијума мења се начин на који размишљамо о животном циклусу научног истраживања (Vaudoine and Branschofsky, 2004, 43).

Захваљујући успешној сарадњи библиотекара и руководства Института, као и научних сарадника, дигитални репозиторијум је заснован, попуњен научним радовима и спреман да прати будућу научну продукцију. Кроз успешну сарадњу свих фактора у овом сложеном процесу гради се и промовише кредибилитет успостављеног репозиторијума. Важно је напоменути да су сарадници Института показали спремност да се укључе у цели подухват што се одражава у резултатима. Сада је њихова продукција скоро у целости у слободном приступу. Процент садржаја у отвореном приступу у колекцијама ИСЈ САНУ је 100%, што је велики успех, имајући у виду да је отворени приступ у хуманистичким наукама у свету мање заступљен и спорије се развија него у другим дисциплинама. Према извештајима начињеним за потребе Европске комисије, највеће ограничење код отвореног приступа је баш у оквиру друштвених наука и области хуманистике (Archambault and Roberge, 2014, 20).

Литература

- “DSpace”. Duraspace. Приступљено 12.03.2019, <https://duraspace.org/dspace/about/>
- “OpenAire”. European Union/European Commission. Приступљено 08.03.2019, <https://www.openaire.eu>
- “Платформа за отворену науку”. Министарство просвете, науке и технолошког развоја. Приступљено 21.02.2018, <http://www.mpn.gov.rs/wp-content/uploads/2018/07/Platforma-za-otvorenu-nauku.pdf>
- Archambault, É., D. Amyot, P. Deschamps, A. Nicol, F. Provenc-her, L. Rebout, and G. Roberge. “Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels – 1996–2013”. *Science-Metrix* (2014). Приступљено 20.03.2019, <http://science-metrix.com/en/publications/reports/proportion-of-open-access-papers-published-in-peer-reviewed-journals-at-the>
- Baudoin, Patsy and Margret Branschofsky. “Implementing an Institutional Repository: The DSpace Experience at MIT”. *Science & Technology Libraries* Vol. 24, no. 1/2 (2004): 31–45.
- Branin, Joseph. *Encyclopedia of Library and Information Science*. Institutional Repositories. New York, N.Y.: Marcel Dekker, 2004-2005.

- Ferreras-Fernández, Tránsito, Francisco J. García-Peñalvo and José A. Merlo-Vega. “Open Access Repositories as Channel of Publication Scientific Grey Literature”. У *TEEM '15 Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality*, 419–426. New York, NY, USA: ACM, 2015.
- Gibbons, Susan and Nancy Fried Foster. “Understanding Faculty to Improve Content Recruitment for Institutional Repositories”. *D-Lib Magazine* Vol. 11, no. 1 (2005). Приступљено 25.03.2019, <http://www.dlib.org/dlib/january05/foster/01foster.html>
- Piowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West and Stefanie Haustein. “The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles”. *PeerJ* Vol. 6 (2018). Приступљено 01.04.2019, <https://peerj.com/articles/4375/;6:e4375>, DOI:10.7717/peerj.4375
- Rajović, Vasilije, Biljana Kosanović and Milica Ševkušić. “DSpace – institutional repositories – dissemination of research results: A local case study”. У *Primena slobodnog softvera i otvorenog hardvera PSSOH 2018*. Belgrade, Serbia: University of Belgrade – School of Electrical Engineering and Academic Mind, 2018.
- Van de Sompel, Herbert and Michael L. Nelson. “Reminiscing About 15 Years of Interoperability Efforts”. *D-Lib Magazine* Vol. 21, no. 11/12 (2015). Приступљено 15.03.2019, <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html>
- Winter, Marsha and Portia Bowen-Chang. “Dealing with DSpace: The Experience at the University of the West Indies, St. Augustine”. *New Library World* Vol. 111, no. 7/8 (2010): 320–332.
- Ćirković, S. “Grey Literature – The Chameleon of Information Resources”. *Infoteka - Journal For Digital Humanities*, Vol. 18, no. 1 (2018): 75–83. doi:10.18485/infoteka.2018.18.1.5
- Шевкушић, Милица, “Партнерски програм Гугл књиге као платформа отвореног приступа у научним библиотекама”. У *Отворен приступ знању у библиотекама, организатори конференције Библиотекарско друштво Србије [и] Народна библиотека Србије*, 187–207. Београд: Библиотекарско друштво Србије, 2013.