

Пренос метаподатака у национални информациони систем научноистраживачке делатности уз аутоматско придруживање ауторства

УДК 655.411(0.034.2):004.65]:378.4(497.11)

САЖЕТАК: Постојећи и алтернативни CRIS системи (енгл. *Current Research Information System*), који се баве решавањем проблема праћења научноистраживачке продуктивности у Србији, показали су се као непотпуни. Нови национални информациони систем научноистраживачке делатности, под називом еНаука, применом савремених методологија, интеграцијом екстерних сервиса и аутоматизацијом процеса, треба да обезбеди већи квалитет метаподатака и ефикасност у раду свих њених корисника. Како би се уштедело време преузимања и додељивања записа истраживачким профилима и НИО (научноистраживачким организацијама), еНауке користи технике аутоматског придруживања и препознавања ауторства помоћу перзистентних идентификатора и алгоритама сличности.

КЉУЧНЕ РЕЧИ: аутоматско придруживање ауторства, перзистентни идентификатори, валидација метаподатака, дедупликација метаподатака, ОАИ-РМН протокол

РАД ПРИМЉЕН: 21. октобар 2023.

РАД ПРИХВАЋЕН: 14. децембар 2023.

Владимир Оташевић
vladimir.otasevic@rcub.bg.ac.rs
ORCID: 0000-0002-9408-3454

*Универзитет у Београду
Електротехнички факултет
Београд, Србија*

1. Увод

Систем еНаука је јавно доступан портал за праћење научноистраживачких резултата, истраживача и институција у

Републици Србији и намењен је обједињеном приказивању научне продукције, области истраживања и постигнућа научноистраживачке заједнице.¹ Систем је базиран на софтверу отвореног кода – DSpace-CRIS², који омогућава одржавање профила истраживача и научноистраживачких организација, прикупљање различитих научноистраживачких резултата, праћење цитираности и сл.

Главни фокус еНауке јесте на уштеди времена свим крајњим корисницима, независно од њихове улоге у систему. Из угла истраживача уштеда времена се одражава на време уноса свих референци и резултата које је досад публиковао, а са стране администратора система посматра се време које је потребно за проверу и верификацију свих резултата које је истраживач пријавио. За доносиоце одлука, уштеда времена се односи на квалитетније сагледавање стања научне продуктивности. На нивоу података, уштеда времена ће се постићи тако што неће моћи олако да се одбаце доступни извори података у које је већ утрошено истраживачко време или време администратора на верификацију и унос података. Систем еНаука је отворен за преузимање свих метаподатака о научним резултатима који могу да задовоље међународни стандард за размену библиотечких информација, што подразумева имплементацију OAI-PMH протокола (Lagoze et al. 2015). Уштеда времена ће се остварити ако извори метаподатака могу обезбедити пренос ауторстава, што подразумева да се ауторска имена која су разрешена перзистентним идентификатором (PID) на самом извору података могу пренети као таква и у систем еНаука.

Систем еНаука омогућава пренос метаподатака по OAI-PMH протоколу. Сваки информациони систем, база података или интерна инфраструктура унутар било које институције која подржава OAI-PMH протокол, може се потенцијално интегрисати у еНауку. За еНауку, OAI-PMH представља стандард у погледу техничких карактеристика које треба имплементирати. Поред техничке исправности сервера који имплементира OAI-PMH протокол, метаподаци морају задовољити минимум захтева у погледу квалитета и доступности како би се резултати пренели у систем еНаука. Због тога је у склопу развоја еНауке креиран посебан сервис OAIValidator који омогућава транспарентну, јавну и бесплатну семантичку и синтаксну проверу исправности метаподатака који се преузимају кроз OAI-PMH протокол. Синтаксна провера се најчешће примењује на проверу идентификатора

1. Портал еНаука

2. [GitHub - 4Science/DSpace at Dspace-6_x_x-Cris](https://github.com/4Science/DSpace-at-Dspace-6_x_x-Cris)

и нумеричких вредности попут године публикавања. Семантичка провера се користи приликом препознавања категорије научног резултата (нпр. научни чланак, конференцијски рад, монографија, патент, техничко решење, докторска дисертација итд.). Препознате категорије научног резултата усклађене су са важећим правилницима надлежног министарства (РС 2020). OAIValidator сервис подржава препознавање ауторстава и може се користити за проверу тачности преноса перзистентних идентификатора истраживача.³

Новина коју еНаука уводи у односу на претходне системе или алтернативна решења јесте отвореност. Отвореност пружа могућност да су метаподаци о научноистраживачком раду видљивији за све крајње кориснике, независно да ли неко од њих обавља неку од улога коју еНаука препознаје. Отвореност омогућава лакшу проверу коректности и тачности резултата, из разлога што су истраживачи обазривији приликом пријаве научног резултата, док администратори система осећају одговорност према свакој публикацији коју морају да провере. Улоговани корисници система немају никакву привилегију у систему насрам анонимних корисника, осим у погледу функционалности. Транспарентност ка корисницима омогућава да корисници могу бити слободни приликом писања критика и похвала система. Такође, такав приступ се показао врло успешним приликом сакупљања ORCID (енгл. *Open Researcher and Contributor ID*)⁴ броја истраживача. Развојном тиму еНауке, транспарентност према корисницима омогућава да сагледа потребе научноистраживачке заједнице и да реализује додатне функционалности које нису могле да се предвиде у процесу пројектовања система, а које у значајној мери доприносе побољшању квалитета платформе и олакшавају рад свих учесника у систему.

Преузимање метаподатака у систем еНаука не зависи од платформе на којој су ти метаподаци креирани. Неке институције су годинама градиле своју локалну инфраструктуру базирану на сопственом знању, не ослањајући се на међународне стандарде, док са друге стране постоје институције које су се опредељивале за имплементацију готових (енгл. *out of the box*) решења. На овај начин свака публикација, без обзира на институционалну политику депоновања записа, кандидат је за преузимање. Овим се постиже да институција може наставити независан развој своје инфраструктуре и њено прилагођавање својим потребама

3. Последња верзија OAIValidator сервиса доступна на адреси <https://proref.rcub.bg.ac.rs/OAIValidator/>

4. ORCID

без радикалнијег мењања начина депоновања података. Институције које су препознале значај информационих система и имају развијену инфраструктуру зарад чувања научноистраживачких резултата, могу наставити рад на свом систему као и досад, а еНаука ће редовним преузимањем наставити да прикупља резултате из институционалних информационих система. Одређене институције су показале висок степен заинтересованости да унапређују своју инфраструктуру и да подрже већи пренос метаподатака, поготово кад је реч о преносу ауторстава кроз ORCID или други перзистентни идентификатор (Naak et al. 2012; Otašević and Kosanović 2022).

Пројектовање система засновано је на CRIS моделу података који је дуго година присутан и примењен у различитим системима (Jeffery and Asserson 2009). Портал еНаука чине три ентитета: Научноистраживачке организације, Истраживачи и Резултати. Додатно, портал је проширен модулom за статистику, са идејом да се обезбеди континуирано праћење напретка научноистраживачке продуктивности и осталих параметара од значаја за науку. Модел података је креиран тако да се сваки ентитет може независно и самостално одржавати, али уједно се могу успоставити и релације између вишеструких инстанци истог ентитета или између различитих типова ентитета. За тачност релација, значајну улогу има заступљеност перзистентних идентификатора истраживача. Модел података је динамички проширив, што омогућава лакшу надоградњу система са новим ентитетима ако се за тим укаже потреба.

Једна од главних идеја за креирање система еНауке јесте да се на једном месту пронађу све информације везане за целокупну научну продуктивност истраживача и институција у Републици Србији. То подразумева, да еНаука прикупља резултате периодично из других система као и из екстерних сервиса попут ORCID платформе чувајући информацију одакле су резултати пристигли. Систем еНаука дедуплицира и обједињује научноистраживачке резултате на једном месту. Поред тога, обједињени подаци се аутоматски проверавају да ли су синтаксно и семантички исправни. Такође, омогућено је обogaћивањерезултата додатним метаподацима, што пружа крајњим корисницима могућност за боље сагледавање научне продуктивности.

Систем је пројектован тако да подржи два процеса, која су врло значајна за сваког истраживача и НИО. То су научна акредитација институцијеи избор у научноистраживачко звање. Систем тежи да постигне штовећи степен транспарентности у поменутиm процесима јер један од главних принципа овог системјесте отвореност.

Како би се обезбедила одрживост, еНаукасе у техничком смислу ослања на међународне стандарде и добре праксе, али уједно је усклађена са правном регулативом која ближе дефинише научноистраживачку делатност. Ослушкујући потребе научне заједнице еНаука разматра нова решења са циљем да корисници што лакше овладају радом у систему. Такође, еНаука настоји да, кроз своје функционалности и имплементацију нових сервиса, образује и информише крајње кориснике пружајући им могућност стицања нових знања и вештина (нпр. значају ORCID профила и међународна видљивост (Arunachalam and Madhan 2016)).

2. Постојећа решења

До сада је било неколико покушаја и идеја како треба имплементирати и технички подржати систем који би сакупљао научноистраживачке резултате и који би могао да подржи различите процесе. Свако досадашње решење није могло у потпуности да задовољи потребе научноистраживачке заједнице. Остављен је простор за унапређење процеса, поготово ако је фокус на уштеди времена. Свако од наведених решења послужило је као пример из ког су могле да се извуку добре и лоше праксе. Уочене праксе утицале су на еНауку тако дане занемарује познате и већ виђене проблеме, већ да активно ради на њиховом решавању.

Сагледавање одрживости, имплементације и развоја других софтвера, показује да постоји неколико различитих решења. Једна од техничких имплементација система могу бити и готова решења као што је Pure⁵. Са друге стране, постоје решења која су у потпуности имплементирана самостално или се у значајној мери ослањају на самостални развој. Такав систем је РИС – Регистар истраживача Србије⁶. Још један сличан пример јесте и новосадски CRIS-UNS⁷ који се користи на Природно-математичком факултету у Новом Саду (Ivanović et al. 2017). Још једно могуће решење је еCRIS систем заснован на подацима из COBISS базе, комерцијално решење које одржава и развија IZUM (Институт информацијских знаности у Марибору) и који се активно користи у више научноистраживачких библиотека, као и

5. Pure | The World's Leading RIMS or CRIS | Elsevier

6. RIS – Registar Istraživača Srbije

7. CRIS-UNS

универзитетским библиотекама, попут Универзитетске библиотеке „Светозар Марковић“ (Tomic and Ljubišić 2020).

Систем еНауке јесте једна врста комбинације различитих имплементација, с тим што је основа базирана на софтверу отвореног кода (енгл. *Open-source software*). Ипак, еНаука у значајној мери одступа од оригиналног почетног отвореног кода и целокупна надоградња система представља самостални развој. За еНауку је јако важно да се досадашњи резултати, претходни развој и подаци који су прикупљени кроз друге системе не одбацују и да се избегне поновни унос нечега што је већ негде унето.

РИС систем, по својој структури узначајној мери личина CRIS модел. Систем представља самостални развој, односно у мањој мери се ослања на имплементацију и интеграцију већ постојећих софтверских решења. Током активног коришћења и примене, прикупљен је значајан број различитих референци о научноистраживачким резултатима. Ако се посматра ниво доступности информација и отворених података, РИС систем је у значајној мери имплементиран као затворени систем. Затвореност се одразила и на податке, па су истраживачи идентификовани на основу *risID*, интерног идентификатора у систему. Мања транспарентност и видљивост овог система смањује одговорност различитих корисника система према тачности прикупљених података. РИС представља један од главних извора података који су иницијално преузети у еНауку, из следећих разлога:

- Садржи референце о научноистраживачким резултатима који су само у РИС систему били доступни;
- У РИС систему истраживач, најчешће, сам уноси референце. На овај начин резултати су могли само ручним уносом бити придружени појединцу;
- Поред научноистраживачких референци, РИС систем садржи и друге административне податке који су од значаја за успостављање и идентификовање осталих ентитета.

Pure платформа је такође једно од могућих техничких решења која подржавају CRIS модел података и релације између ентитета. Као комерцијално решење које развија и одржава Elsevier, то је савремен систем заснован на модерним информационим технологијама који користи податке доступне у Scopus бази.⁸ Један од главних изазова

8. Детаљи о Scopus бази на <https://www.elsevier.com/products/scopus>

за овакав систем био би проширење модела података због тога што се научноистраживачка продуктивност, а пре свега њено вредновање, темељи првенствено на подацима доступним у Web of Science⁹ (Kosanović 2004), а мање на Scopus подацима. Други изазов био би преузимање домаће продуктивности. Научноистраживачки резултати из друштвених и хуманистичких наука мање су видљиви и доступни у међународним базама, а много присутнији у домаћим индексним базама.

Решење које јесте комерцијално, али које боље разуме како настаје научноистраживачка продуктивност публикована у домаћим часописима или код домаћих издавача јесте COBISS. Систем препознаје неколико различитих корисничких улога у зависности од степена одговорности и могућности које дата улога носи. Подаци у систему су јавно доступни, што у значајној мери уноси још један ниво одговорности према подацима. За еНауку, COBISS је значајан извор метаподатака, претежно због тога што представља главни извор метаподатака о публикованим и објављеним резултатима код домаћих издавача. У комбинацији са eCRIS системом, могуће је успоставити пренос метаподатака заједно са информацијом о ауторству.

Добар пример да је могуће увести CRIS систем као поуздано решење у рад једне научноистраживачке организације (НИО) јесте CRIS-UNS који је развијен као самостално решење. Ипак, ово решење је првенствено осмишљено и примењено за рад на нивоу једне институције. Њихова примарна улога јесте да подрже рад појединачних НИО и омогући да се на једноставан начин уведу процедуре уноса референци. Препознато је као пример добре праксе које је примењено и код још неколико НИО. Систем у техничком смислу испуњава потребне захтеве како би био интегрисан у еНауку.

3. Имплементација модела података и инфраструктуре система еНаука

Систем еНаука заснован је на софтверском решењу отвореног кода Dspace-CRIS, који је развио 4Science, али у значајној мери зависи од заједнице и појединаца који континуирано доприносе унапређењу софтвера (Mornati and Bollini 2013). Такође, Dspace-CRIS базиран је на једном другом софтверском решењу отвореног кода, односно Dspace платформи (2017). У тренутку писања овог рада, постоји

9. Web of Science

206 акредитованих НИО. Од тог броја 111 има локално развијену инфраструктуру или користи заједничку која испуњава минималне техничке захтеве за интеграцију са еНауком. Од 111 пријављених извора које се харвестују преко OAI-PMH протокола, значајан број заснована је на Dspace или Dspace-CRIS платформи (64%) која је лако проширива да подржи и пренос ауторстава (Kosanović et al. 2019; Smederevac et al. 2020).

Као већ добро познат модел података због значајне заступљености, еНаука се темељина Dspace-CRIS 6.3 моделу података као и већи део програмског кода. Разлог за одабир верзије 6.3 је тај што у тренутку кад је започет пројекат еНаука није постојала међународна стабилна продукциона верзија Dspace 7.X, или Dspace-CRIS 7.X. Већина продукционих инстанци су на верзијама 6.X или 5.X које су међусобно врло сличне. Оне се драстично разликују, како технолошки тако и по моделу података од верзије 7.X.

Dspace-CRIS 6.3 представља надоградњу Dspace 6.X која је развијена помоћу двеју различитих технологија. Једна је базирана на XML UI технологији (Sarang 2006) која је флексибилна, проширива и лако примењива, али постоји велики проблем са одрживошћу јер се технологија већ дуго сматра застарелом и превазиђеном. Са друге стране, Dspace нуди и имплементацију развијену у JSPUI технологији (JCP 2000). Решење на Java програмском језику је врло популарно међу апликацијама које су већ дуже време присутне и активне. Dspace-CRIS 6.3. развијен је као надоградња Dspace 6.X и то у JSPUI технологији. Зато и еНаука своју главну компоненту (енгл. *core*) заснивана JSPUI технологијама. Систем еНаука тежи да свој развој који се пре свега односи на проширење постојећих функционалности или увођењу нових компонената, ствара у савременијим технологијама. Додатни алати, скрипте и додаци у зависности од потреба, примењују развојне оквири Angular (Google 2024) и Spring (Deinum et al. 2012), REST API структуру (Fielding 2000) и остале технологије које се најчешће могу наћи код савремених софтверских решења. Посебна пажња посвећена је писању скрипти и пројектовању алата који се периодично извршавају, међу којима су скрипте које генеришу варијанте имена (нпр. пермутације, пресловљавање и сл.) и алати за аутоматско придруживање рада истраживачу.

Систем еНаука препознаје неколико различитих улога корисника. Анонимни корисници свакако представљају највећу групу. Систем тежи да буде отворен и транспарентан у сваком погледу, што омогућава

крајњим корисницима да буду главни контролори квалитета рада система и приказаних података. Према групи истраживача, еНаука тежи да поједностави процесе рада у систему. Поред могућности да уређују додатне податке о себи и да „ручно“ унесу своје референце искључиво кроз екстерне сервисе, истраживачи се постепено информишу и едукују кроз систем. За истраживаче највећа новина јесте ORCID профил, чија је заступљеност пре увођења еНауке била на врло ниском нивоу (<15%), како би после четири месеца од пуштања продукције достигла 83,24%. Овај проценат ће бити већи када еНаука почне да се промовише кроз научноистраживачку заједницу. Трећа врло значајна улога у систему еНаука јесте НИО уредник. Ту улогу добијају појединци које ангажује НИО и који представљају главне чуваре квалитета метаподатака о научноистраживачким резултатима. Већина функционалности, које се односе на проверу, одобравање и уређивање записа, прилагођено је раду НИО уредника, тако да се штеди време које му је потребно за прегледања записа и доделу ауторстава.

У техничком смислу, еНаука кроз своју доградњу треба да подржи аутоматизацију процеса у сваком сегменту. Основна компонента CRIS система јесу ентитети. Главни ентитети у еНауци су научноистраживачки резултати, истраживачи и НИО. Циљ је да се уз примену информационих технологија аутоматизује сваки процес који би уштедео време истраживачу и НИО уреднику. Главни изазов је аутоматско препознавање PID-ова (енгл. *persistent identifiers*) који се примењују на свим ентитетима, односно на публикацијама, истраживачима и организацијама. Њихова примена обезвређује једнозначну идентификацију сваког појединачног ентитета, као и њихову сталну интеграцију са спољним сервисима. Како главни податак у еНауци јесте научноистраживачки резултат, односно његова припадност појединачном истраживачу или НИО, онда главни изазов јесте постизање аутоматизације у процесу успостављања релације између ентитета (Слика1).

Сваки ентитет има особину јаког ентитета што омогућава да постоје самостално и независно. То подразумева да је могуће сваки од њих уређивати, додавати или уклањати без обзира на друге ентитете. Међутим, поред ентитета, CRIS систем чине и релације које постоје између њих. Зато без обзира на њихову самосталност и независност сваку измену над подацима треба чинити пажљиво и са опрезом, да се не би нарушили остали ентитети. Поред научноистраживачких података, у еНауци постоје и други подаци који ближе описују или



Слика 1. Приказ ентитета у еНауци и података који одређују директне и индиректне релације

једнозначно идентификују сваког истраживача или НИО. Такви подаци представљају административне податке и препознати су као засебна целина у систему. Због тога податке у еНауци треба поделити у две групе:

1. Научноистраживачки подаци – метаподаци о референцама. Јавно видљиви и свима доступни након одобравања НИО уредника или преузимања из поузданих извора.
2. Административни подаци – уређују НИО референти. Посебан затворен модул са овлашћеним правом приступа. Садржи већи број података о истраживачима и НИО (нпр. ORCID и запослење). Од значаја су приликом покретања процедура за избор у звање или акредитацију.

4. Извори метаподатака

Истраживачи и НИО као ентитети у еНауци дефинисани су административним подацима. Ако у еНауци није могуће пронаћи одређеног истраживача или институцију, или ако постоји грешка у административним подацима попут имена, презимена, запослења, звања, ORCID и сл., онда НИО референт у оквиру своје институције кроз административни модул треба да унесе, измени или обрише појединог истраживача или одређени податак. Подаци између јавног

портала еНауке и модула за уређивање административних података преузимају се на дневном нивоу. Овакав модел синхронизације одабран је зато што измене у административним подацима нису масовне и учестале сад кад су успостављање релације између ентитета.

За јавни део еНауке, поред административних података, значајну улогу имају и метаподаци о научноистраживачким резултатима. Такође, сваки запис који је дошао редовним преузимањем из постојећих извора носи у себи и метаподатак (`dc.identifier.uri`), који представља трајну везу према самом извору где је запис преузет. Такав приступ обезбеђује проверљивост доступних информација и проверу ауторстава. Треба узети у обзир два важна фактора која су утицала наодабир, сакојим ће се приоритетом извори метаподатака интегрисати, и на који начин:

1. Поузданост извора метаподатака – ако су подаци стизали из извора код којих је већ неко одговоран и квалификован прегледао метаподатке, онда не постоји потреба да се записи верификују. Ипак, ако је те податке унео само истраживач, онда је неопходно да их НИО уредник провери, па тек онда придружи аутору.
2. Значај извора метаподатака – ако је извор метаподатака већ коришћен као информациони систем за праћење научне продуктивности или садржи велики број значајних публикација, у научном смислу, онда такви извори имају приоритет приликом интеграције јер доступност таквих метаподатака у значајној мери штеди време истраживачу и НИО уреднику.

Извори метаподатака могу се поделити према типу резултата. Прва група представља информационе системе и базе које прикупљају резултате који имају заједничку особину публиковања. Такви извори могу прикупити метаподатке записа који су публиковани код домаћих издавача или који су само публиковани у Web of Science и сл. Са друге стране, постоје инфраструктуре које су развијене у оквиру појединачне НИО које прикупљају све метаподатке о научноистраживачким резултатима својих запослених. Препознат је проблем да одређене НИО немају капацитет да развију сопствену инфраструктуру, тада се опредељују за заједничке репозиторијуме. Према моделу преузимања, еНаука се обogaђује записима тако што се редовно харвестују поуздани извори који се не верификују или истраживачи иницирају преузимање публикација из спољних извора које НИО уредник верификује.

Постоје различита софтверска решења која омогућавају имплементацију OAI-PMH протокола. Како би се у техничком смислу

подржао овај протокол, мора постојати софтверско решење (у даљем тексту ОАИ-РМН сервер) које може слати податке у очекиваном формату и мора постојати софтверско решење са друге стране које ће моћи те податке да преузме. Свака НИО има могућност да одабере софтверско решење које обезбеђује слање метаподатака. Исправну имплементацију ОАИ-РМН протокола успоставило је 111 НИО и овај број наставља да расте. Очекује се да ће доћи до пораста сервера који имплементирају ОАИ-РМН кад систем буде промовисан међу научноистраживачком заједницом. Препознате имплементације сервера за ОАИ-РМН протоколдате су у Табели 1.

Табела 1. Типови софтверских решења које се користе као извори из којих се преузимају метаподаци

Софтверско решење	Опис
Институционални репозиторијум	НИО су препознале значај репозиторијума, чија је примена много шира од интеграције у еНауку.
Заједнички репозиторијум	Најчешће инфраструктура универзитета или библиотека
Међународни репозиторијум	Међународна инфраструктура
Самосталан развој	Поједине институције овладале су ОАИ-РМН стандардом и успеле су у томе да самостално развију ОАИ-РМН сервер.
Екстерни сервис	Истраживачи само из екстерних сервиса могу повући метаподатке. НИО не подржава ниједну инфраструктуру.

НИО које су препознале значај сопствене инфраструктуре најчешће су се опредељивале за успостављање институционалних репозиторијума. Институционални репозиторијуми постоје и користе се независно од еНауке. Као другу могућност, неке НИО су се определиле да се удруже, најчешће на нивоу Универзитета и да користе заједничку инфраструктуру. Универзитетске репозиторијуме имају Универзитет у Крагујевцу и Криминалистичко-полицијски Универзитет. Такође,

библиотеке попут Универзитетске библиотеке „Светозар Марковић“ пружају своју инфраструктуру и подршку НИО да би се прикључиле заједничком репозиторијуму.¹⁰ Као трећа могућност, НИО које теже успостављању репозиторијума могу се одредити и за међународну инфраструктуру. До сада Zenodo¹¹ је препознат као међународни репозиторијум. Ако се посматра из угла метаподатака, Zenodo прикупља свеврсте резултата (енгл. *catch all repository*). Одређене НИО су и пре еНауке поседовале локалне базе и информационе системе који су успешно унапређени тако да подрже OAI-PMH сервер. Природно-математички факултет у Новом Саду је кроз CRIS-UNS интегрисан у еНауку и помогао је другим НИО да се по сличном моделу интегришу. Електротехнички факултет у Београду је такође унапредио свој информациони систем и успео да задовољи техничке захтеве за интеграцију у еНауку и помогао осталим НИО да се интегришу. Институције су почеле да препознају значај ORCID и проширују базе новим подацима ако знају да ће њихов пренос у еНауку уштедети време на уређивању записа.

Метаподаци који су доступни у COBISS бази такође се преузимају у еНауку користећи OAI-PMH протокол. За записе који се преузимају из COBISS базе важно је да имају одређену типологију. Систем еНаука приликом преузимања записа из COBISS-а врши аутоматско препознавање категорије научног резултата. OAI-PMH који је имплементиран за потребе дисеминације записа из COBISS базе проширен је тако да се преузимање врши по COMARC спецификацији (IZUM 2023a, 2023b). Видљивост и транспарентност записа који су стигли из COBISS базе омогућили су да се лакше уоче и исправе грешке на нивоу метаподатака на самом извору.

5. Инфраструктура система

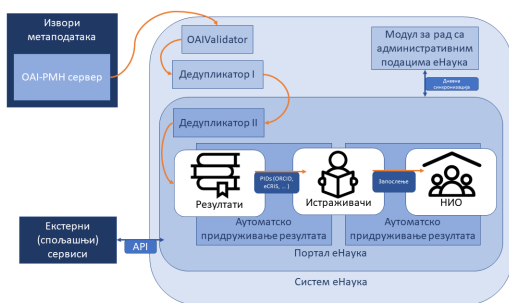
Јавни портал система еНауке развијен је у окружењу које је чешће код некомерцијалних информационих система. Као оперативни систем коришћен је Ubuntu оперативни систем. Као веб-сервер коришћен је Apache Tomcat® 9.x (The Apache Software Foundation 2017). Основу јавног портала еНауке чини Dspace-CRIS 6.3. платформа базирана на Java програмском језику и имплементирана као Maven пројекат. Apache

10. PHAIDRA, Дигитални репозиторијум Универзитета у Београду.

11. Zenodo

Tomcat® 9.x сервер јесте погодан и компатибилан са Dspace-CRIS верзијом 6.3. Као систем за управљање релационим базама коришћен је PostgreSQL (2023), а као систем за индексирање и претрагу користи Apache Solr (The Apache Software Foundation 2014).

Поред техничких карактеристика окружења у којем је развијена еНаука, сам систем чини јавни портал и неколико сервиса. Због сложености имплементираних процеса, као и вишеструких извора метаподатака и административних података, систем еНаука подељен је на неколико засебних модула, при чему је сваки развијен као самостална апликација која може постојати и радити независно од других модула. Ипак, да би систем еНауке могао да функционише као сервис, неопходно је да неки од модула буду у честој комуникацији. На Слици2 приказан је упрошћен модел инфраструктуре који представља систем еНауке. Наранџаста стрелица показује путању кретања метаподатака приликом преузимања. Плава стрелица приказује синхронизације са интерним и екстерним сервисима/модулима.



Слика 2. Приказ инфраструктуре еНауке.

На Слици2 приказан је портал еНауке. То је главни сегмент система еНауке јер је јавно доступан и чини главну компоненту која интегрише остале модуле. Портал еНауке обједињује административне податке о истраживачима и НИО, као и метаподатке о научноистраживачким резултатима. Такође, портал еНауке представља главно чвориште ка осталим екстерним сервисима који су у њему интегрисани. У Табели2

дат је приказ екстерних сервиса који су значајни за придруживање ауторстава у еНауци.

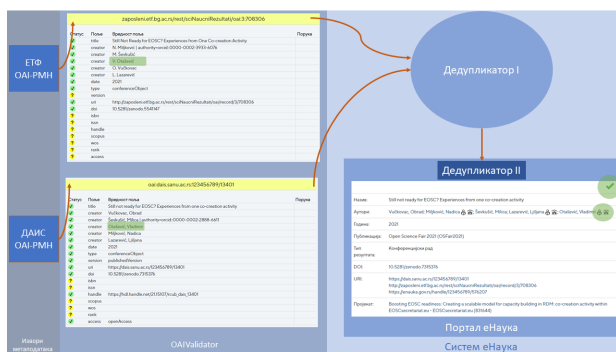
Табела 2. Списак екстерних сервиса који су интегрисани у еНауку, а од значају су за пренос и придруживање ауторстава.

Екстерни сервис	Примена
ORCID	Аутентикација; Ауторизација; Повлачење радова; Аутоматско придруживање ауторства; *Слање радова
CrossRef	Обогаћивање записа; провера квалитета
КоBSоN	Аутоматска категоризација резултата; Преузимање записа
KNR	Преузимање записа; Аутоматско придруживање ауторства; веза ка профилу истраживача унутар КНР платформи
eCRIS	Аутоматско придруживање ауторства; веза ка профилу истраживача/НИО унутар eCRIS систему
RIS	Аутоматско придруживање ауторства

У Табели 2 приказани су најзначајнији екстерни сервиси са којим је еНаука интегрисана. Наведени сервиси доприносе еНауци тако што се ентитети истраживача и резултата редовно ажурирају новим подацима. Из угла аутоматског придруживања ауторства, наведени сервиси пружају неопходне податке за успостављање механизма у еНауци који аутоматски придружују резултате истраживачима, проверавају тачност пренетих ауторстава и проверавају постојање дупликата. Постоје и други екстерни сервиси интегрисани у еНауку, али је њихова примена специфична и није од значаја за крајњекориснике. Записи који се могу пронаћи на порталу еНауке редовно се аутоматски обогаћују подацима добијеним из наведених екстерних сервиса. У колони „Примена“ одређене вредности исписане су са знаком „*“ што подразумева да се наведена примена очекује и да није доступна у тренутку писања овог рада.

6. Фазе провере и обраде метаподатака

Метаподаци који се преузимају редовним харвестом пре него што стигну до еНауке, морају проћи кроз неколико фаза провере и обраде. Први корак јесте валидација OAI-PMH сервера. За потребе провере OAI-PMH сервера и метаподатака развијен је OAIValidator. Провера OAI-PMH сервера подразумева проверу да ли је сервер доступан, има ли имплементиране функције које враћају листе скупова по којима су груписани записи, да ли постоји *Identify* страница са основним информацијама о извору метаподатака и да ли је подржан очекивани формат за пренос метаподатака. Поред основне функционалности провере валидности OAI-PMH сервера, OAIValidator помаже администраторима репозиторијума да проверена једном месту семантичку и синтаксну исправност метаподатака. Такође, OAIValidator пружа основне статистичке показатеље о записима као и детаљне поруке о неисправним метаподацима. У контексту еНауке, сви извори метаподатака који се редовно харвестују пролазе аутоматски кроз процес валидације који ће машински препознати записе/метаподатке, који могу ући, у систем еНауке. На Слици 3 приказан је пример валидације једног извора метаподатака и како еНаука види један од преузетих записа са ауторствима. Приказан пример је додатно проверио НИО уредник како би се разрешио захтев за ауторство.



Слика 3. Пример валидације, дедупликације и обједињавања записа који је аутоматски пренет из два извора метаподатака.

Други корак представља дедупликација метаподатака. Како је тај процес сложен и временски захтеван, одвија се у засебном модулу. Дедупликацију чине два шаблона препознавања:

1. Идентификатори (DOI, WoS UT, Scopus ID) – наведени идентификатори морају бити једнозначни, у супротном реч је о дупликату. Постоје изузеци али тад се примењује други шаблон;
2. Наслов, година, тип публикације – Немају све публикације PID-ове. Као публикације које имају сличан нормиран наслов, годину објављивања са грешком од \pm једне године и исти тип научноистраживачког резултата сматрају се дупликатима.

Препознати дупликат се не одбацује већ се покушава аутоматско обједињавање записа. Најзначајнији метаподатак на запису јесте ауторство. Дедупликатор ће покушати на записима да препозна ауторе који немају додељено ауторство (нпр. ORCID, eCRIS-ID и сл.), ако дупли запис садржи додатни идентификатор. Поред ауторства, дедупликатор ће за препознате дупликате објединити све PID-ове на записима као и URI вредности, како би трајно остао траг одакле су записи преузети.

Портал еНауке је стално активан и њени корисници непрекидно уређују или повлаче записе у еНауку. Како је процес дедупликације временски скуп, може се догодити да подаци у дедупликатору не буду синхронизовани са подацима у еНауци. Због тога ће приликом преноса дедуплицираних записа у еНауку радити још једна аутоматска дедупликација у реалном времену, ако се препозна да запис већ постоји или да је запис на верификацији код НИО уредника. Додатно, за записе који постоје у еНауци радиће се обједињавање и обогаћивање метаподатака. Изузетно је важно ако из других извора метаподатака стигне дупли запис који носи нове информације о ауторствима.

7. Модел података

Модел података у еНауци заснован је потпуно на моделу података који постоји у Dspace-CRIS 6.3. те није било потребе да се уводе додатне табеле и колоне. Овакав приступ омогућава да еНаука остане стандардизована и да не одступа у техничком смислу од изворног кода кад је реч о моделу података. Ипак, начињене су одређене модификације како се примењују вредности из базе. Подразумевано, Dspace-CRIS 6.3. као вредности ауторстава препознаје ORCID. Са друге стране,

интегрисани системи чувају ауторства кроз друге PID-ове попут eCris, RISiD, Картон Војводине ID и сл. Стога је еНаука пројектована тако да омогућава чување и придруживање осталих идентификатора.

Како еНаука преузима метаподатке из спољних извора редовним харвестом преко OAI-PMH сервера, неопходно је било дефинисати који ће се стандарди и спецификације метаподатака примењивати. Омогућено је преузимање метаподатака као записима који следе једну од следећих спецификација:

1. Даблинско језгро (енгл. *Dublin Core*, dc) – стандардизована и међународно прихваћена спецификација за размену метаподатака (DCMI 2012; Kunze and Baker 2015; ISO 2017; Хокнер and Будрони 2011);
2. Dspace Intermediate Format (dim) – нестандардизована спецификација метаподатака осмишљена за размену података унутар саме Dspace платформе.

Спецификацију Даблинског језгра било је лакше применити и интегрисати јер је реч о широко распрострањеној шеми за опис метаподатака која се може пропагирати кроз OAI-PMH сервер (Jackson et al. 2008). Ипак, такав приступ носи и одређена ограничења. По dc спецификацији није могуће пренети ауторство (нпр. ORCID). Како би се овај проблем превазишао, еНаука је пратећи смернице међународних организација и добрих пракси применила решење које проширује етикету са ауторствима за још један атрибут („id“) у који OAI-PMH сервер може сместити вредност ауторства (2018). Постоје и друга примењива алтернативна решења за пренос ауторстава, разматрана приликом проширења OAI-PMH преноса метаподатака која имплементирају dc спецификацију (BASE 2023). Други значајан недостатак ове спецификације јесте ограниченост у препознавању етикета (енгл. *tag*). Свака етикета се састоји од два дела, „schema“ и „element“, који служе за препознавање. Овај проблем се испољава приликом преноса PID вредности, што је велики проблем за еНауку која већину операција заснива на тим вредностима. Као пример можемо узети запис који има DOI, WoS UT, Scopus iD, PMID, COBISS iD и интерни нумерички идентификатор. Све наведене PID вредности стићиће кроз „dc.identifier“ поља, што би из угла еНауке значило да постоје препознати идентификатори, али се незна за одређену вредност који конкретно идентификатор представља. Наравно, еНаука има механизме семантичког препознавања и синтаксне провере PID

вредности и покушаће да разврста вредности по тачно одређеним идентификаторима, али постоји могућност да ће НИО уредник морати накнадно да исправља такав запис.

Са друге стране, еНаука нуди могућност преузимања података по `dim` спецификацији. Како је ова спецификација нестандардизована, могуће је њоме радити све врсте модификација. Овде не постоји проблем са идентификаторима и њиховим препознавањем јер свака етикета може бити дефинисана са максимално три сегмента, односно поред „`dc`“ и „`element`“ постоји „`qualifier`“ атрибут. Такође, `dim` спецификација без додатних модификација подржава пренос вредности ауторства на свим етикетама. Једина разлика је та што се атрибут назива „`authority`“. Применом XSLT (енгл. *Extensible Stylesheet Language Transformations*) трансформације (W3C 1999), могуће је једнозначно превести `dim` у `dc` спецификацију. XSLT представља језик за трансформисање XML докумената. Може се закључити да `dc` заправо представљаподскуп `dim` спецификације. Могуће је XSLT трансформацијом превести `dc` у `dim` спецификацију, али таква трансформација није једнозначна нити поуздана, поготово ако је запис богат идентификаторима.

Портал еНаука је надограђен додатним механизмима који олакшавају рад НИО уредника и истраживача приликом провере квалитета метаподатака. У реалном времену, портал еНауке ће пријавити потенцијалне дупликате по наслову које није могао аутоматски да се дедуплицира јер није постигнут степен поузданости да је посредни стварни дупликат. То су ситуације код којих је неопходно да човек прегледа запис. Поред препознавања дуплих наслова, еНаука ће приликом уређивања записа, ако се дописују додатни идентификатори, радити проверу да ли већ постоји запис са таквим идентификатором. Уочен је проблем са записима који имају неисправне и непостојеће DOI бројеве. Систем еНауке аутоматски ради проверу сваког DOI броја и обележава оне који нису валидни.

8. Закључак

Систем еНаука преставља искорак у односу на постојећа алтернативна решења. Прилагођен је раду НИО уредника и истраживача. Аутоматско придруживање ауторстава на раду знатно штеди време свих корисника. Ако извори метаподатака испуњавају техничке услове, односно имају OAI-PMH сервер и њихови метаподаци задовољавају критеријуме синтаксне и семантичке исправности, онда

ће еНаука подржати интеграцију таквих извора. Прикупљање података на једном месту, уз чување веза ка оригиналним изворима, омогућава преглед појединачних записа у еНауци, као и проверљивостистих записа код оригиналних извора из којих су записи преузети. Извори података су децентрализовани и омогућавају свакој НИО да настави коришћење својих информационих система. Са друге стране, систем еНауке ће редовним преузимањем података аутоматски радити на валидацији, дедупликацији, провери и обogaћивању записа, што ће крајњим корисницима пружити бољи увид у записе.

Даље, анонимни корисници имају посебну улогу у овом систему и њихов допринос кроз непрекидну проверу квалитета доступних информација унапређује сам систем. Транспарентност у информацијама уводи један нови степен обавезности и смањује могућност да се поткраду (не)намерне грешке.

Што се тиче инфраструктуре, пројектована је тако да се постиже што већа модуларност. Такав приступ обезбеђује да се сваки модул може развијати и одржавати независно од другог. Такође, модули могу радити независно један од другог, али да би цео систем функционисао, неопходно је да буду доступни и међусобно усклађени. Својом иновативношћу и искорак наспрам алтернативних решења, систем еНауке је постао извор научних информација. Крајњи корисници имају прилику да овладају новим вештинама које су неопходне сваком ко жели да буде део савремене међународне научноистраживачке заједнице.

Систем еНауке са ентитетима истраживача, НИО и научноистраживачких резултата представља CRIS систем. Следећи искорак за еНауку био би додавање пројеката као нов ентитет, поред већ постојећих. Пројекти би допринели отварању једног новог погледа на праћења научноистраживачке продуктивности.

Литература

- Arunachalam, SSubbiah, and Muthu Madhan. 2016. "Adopting ORCID as a Unique Identifier Will Benefit All Involved in Scholarly Communication." *The National Medical Journal of India* 29 (4): 227–234.
- BASE. 2023. *Bielefeld Academic Search Engine. n.d. Golden Rules for Repository Managers*. Преузето 24.10.2023, https://www.base-search.net/about/en/faq_oai.php#dc-creator.

- DCMI. 2012. *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description*. Преузето 25. 10. 2023, <https://www.dublincore.org/specifications/dublin-core/dces/>.
- Deinum, Marten, Koen Serneels, Colin Yates, Seth Ladd, and Christophe Vanfleteren. 2012. “Pro Spring MVC: With Web Flow.” Chap. *Spring Framework Fundamentals*, edited by Marten Deinum, Koen Serneels, Colin Yates, Seth Ladd, and Christophe Vanfleteren, 25–50. Berkeley: Apress. https://doi.org/10.1007/978-1-4302-4156-0_2.
- DSpace 6.x Documentation*. 2017. Преузето 28. 10. 2023, <https://dspace.umkt.ac.id/handle/463.2017/19>.
- Fielding, Roy Thomas. 2000. “Architectural Styles and the Design of Network-Based Software Architectures.” Преузето 25. 10. 2023, https://ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf. PhD diss., University of California, Irvine.
- Google. 2024. *Angular - Introduction to Angular Concepts*. Преузето 6. 1. 2024, <https://angular.io/guide/architecture>.
- Haak, Laurel L., Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. 2012. “ORCID: A System to Uniquely Identify Researchers.” *Learned Publishing* 25 (4): 256–264. <https://doi.org/10.1087/20120404>.
- ISO. 2017. *ISO 15836-1:2017: Information and documentation The Dublin Core metadata element set*. Преузето 25. 10. 2023, <https://www.iso.org/standard/71339.html>.
- Ivanović, Dragan, Dušan Surla, Miroslav Trajanović, Dragan Misić, and Zora Konjović. 2017. “Towards the Information System for Research Programmes of the Ministry of Education, Science and Technological Development of the Republic of Serbia.” *Procedia Computer Science*, no. 106, 122–129. <https://doi.org/10.1016/j.procs.2017.03.044>.
- IZUM. 2023a. *COMARC/A. Format za normativne podatke: priručnik za korisnike*. Преузето 25. 10. 2023, https://home.izum.si/izum/e-prirucnici/2_COMARC_A/Ceo_2_COMARC_A.pdf.
- IZUM. 2023b. *COMARC/B. Format za bibliografske podatke: priručnik za korisnike*. Преузето 25. 10. 2023, https://home.izum.si/izum/e-prirucnici/1_COMARC_B/Ceo_1_COMARC_B.pdf.

- Jackson, Amy S., Myung-Ja Han, Kurt Groetsch, Megan Mustafoff, and Timothy W. Cole. 2008. "Dublin Core Metadata Harvested Through OAI-PMH." *Journal of Library Metadata* 8 (1): 5–21. https://doi.org/10.1300/J517v08n01_02.
- JCP. 2000. *Java™ Servlet Specification Version 2.3*. Preuzeto 27. 10. 2023, https://jcp.org/aboutJava/communityprocess/first/jsr053/servlet23_PFD.pdf.
- Jeffery, Keith, and Anne Asserson. 2009. "Institutional Repositories and Current Research Information Systems." *New Review of Information Networking* 14 (2): 71–83. <https://doi.org/10.1080/13614570903359357>.
- Kosanović, Biljana. 2004. "The Availability of Scientific Information in Serbia: Trends and Perspectives." *Hemijska industrija* 58 (4): 158–160. <https://doi.org/10.2298/HEMIND0404158K>.
- Kosanović, Biljana, Milica Ševkušić, Vasilije Rajović, and Nenad Popović. 2019. *Setting the Scene for a Sustainable National Repository Network in Serbia*. Преузето 24. 10. 2023, <https://zenodo.org/records/3509971>.
- Kunze, John A., and Thomas Baker. 2015. *The Dublin Core Metadata Element Set. RFC 5013*. Preuzeto 27. 10. 2023, <https://datatracker.ietf.org/doc/rfc5013/>.
- Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, and Simeon Warner, eds. 2015. *Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0*. Preuzeto 27. 10. 2023, <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
- Mornati, Susanna, and Andrea Bollini. 2013. "DSpace-CRIS: An Open Source Solution." EuroCRIS Membership Meeting, https://dspacecris.eurocris.org/bitstream/11366/73/1/CINECA_DSpace_CRIS_An_Open_Source_Solution_v2.pdf.
- Otašević, Vladimir, and Biljana Kosanović. 2022. "Resolving authorship using the ORCID identifier." In *Zbornik radova 28. IKT konferencije "YU INFO 2022"*, 48:167–172. Informaciono društvo Srbije.
- PostgreSQL 12.18 Documentation*. 2023. Preuzeto 27. 10. 2023, <https://www.postgresql.org/docs/12/index.html>.

- Sarang, Poornachandra. 2006. "Pro Apache XML." Chap. The Apache Cocoon Framework, 279–325. Berkeley: Apress. https://doi.org/10.1007/978-1-4302-0166-3_7.
- Smederevac, Snežana, Dejan Pajić, Sanja Radovanović, Silvia Gilezan, Petar Čolović, and Branko Milosavljević. 2020. *Otvorena nauka: praksa i perspektive*. Novi Sad: Univerzitet u Novom Sadu.
- Tomic, Emina Cano, and Ljubica Ljubišić. 2020. "General Overview of the Growth and Development of the Services Provided by Virtual Library of Serbia Network and Their Impact on Collection of Statistical Library Data in Serbia." *Qualitative and Quantitative Methods in Libraries* 9 (1): 77–98.
- Using Persistent Identifiers with Literals in Dublin Core-Based Metadata in XML*. 2018. Preuzeto 28. 10. 2023, https://github.com/dcmi/pids_in_dc/blob/master/proposal/The_Association_of_Persistent_Identifiers_with_Literals_in_XML-formatted_Metadata_using_Dublin.md.
- W3C. 1999. *XSL Transformations (XSLT)*. Преузето 29. 10. 2023, <https://www.w3.org/TR/xslt-10/>.
- РС. 2020. *Правилник о стицању истраживачких и научних звања ("Службени гласник РС", бр. 159 од 30. децембра 2020, 14 од 20. фебруара 2023.)* Преузето 25. 10. 2023, <https://www.pravno-informacioni-sistem.rs/SIGlasnikPortal/eli/rep/sgrs/ministarstva/pravilnik/2020/159/18/reg>.
- Хокнер, Маркус, and Паоло Будрони. 2011. "Пројекат Репозиторијума Универзитета у Бечу." Преузето 25. 10. 2023, http://infoteka.bg.ac.rs/pdf/Srp/2011-1/INFOTHECA_XII_1_August_23-33.pdf, *Инфотека* 12 (1): 23–33.