

Interoperabilita výskumných informácií a európska prax



Keď hovoríme o výskumných informáciách, každému zainteresovanému ako prvé napadnú vedecké databázy, publikačná činnosť a citácie. Do tejto kategórie však zaradíme opis celého životného cyklu výskumu, nielen jeho finálne výsledky. Dôležité je, kto výskum vykonáva (organizácie, výskumníci) a aké výskumné aktivity realizuje (grantové schémy, výzvy na podávanie projektov, projekty). Následne nás zaujímajú výsledky výskumných aktivít, ktorými sú okrem publikácií aj patenty, produkty, inovácie, ale v neposlednom rade aj rozsiahle kolekcie výskumných dát, spracovanie ktorých privedlo výskumné kolektívy k formulovaniu záverov.

Identifikácia, zozbieranie a predovšetkým logické a navzájom previazané usporiadanie informácií, dát, metadát a digitálnych objektov, charakterizujúcich vedu a výskum, je výzva, ktorú riešia tímy odborníkov z celého sveta. Cieľom je zabezpečiť, aby údaje boli voľne dostupné, vyhľadateľné, opakovane použiteľné a tiež uložené v systémoch, ktoré umožnia ich vzájomnú výmenu a spoločné využívanie. Tieto požiadavky sú zosumarizované v podobe princípov riadenia údajov FAIR (findable, accessible, interoperable, reussable).[3]

Jeden z prieskumov zameraných na interoperabilitu[4] výskumných informácií spoločne realizovali v rokoch 2015 – 2016 medzinárodné organizácie EUNIS (European University Information Systems Organization, www.eunis.org) a euroCRIS (European Organization for International Research Information, www.eurocris.org).

Ako dva základné komponenty oblasti manažmentu výskumných informácií pre tento prieskum odborníci označili **systémy CRIS** (Current Research Information System – informačný systém o vede a výskume) a **inštitucionálne repozitáre** (IR): „Zvyšujúci sa strategický význam systémov CRIS a IR pre vysokoškolské vzdelávanie a výskumné organizácie sa týka potreby podporovať výskum a inovácie a zabezpečiť rýchlejší a širší prenos technológií do priemyslu a spoločnosti. Ďalšími dôležitými prvkami, ktoré majú silný vplyv na takýto strategický vývoj, sú nové politiky týkajúce sa otvoreného prístupu, hodnotenia výskumu na celoštátnej úrovni a financovania výskumu.“[1]

Pre potreby prieskumu bola použitá definícia systému CRIS podľa organizácie euroCRIS: **CRIS** je definovaný ako informačný nástroj určený na sprístupňovanie a šírenie výskumných informácií. Systém CRIS je tvorený dátovým modelom opisujúcim predmety záujmu výskumu a vývoja. Obsahuje tiež nástroje (sadu nástrojov) na manažovanie dát. Systémy CRIS používané v európskych organizáciách často používajú dátový formát CERIF (Common European Research Information Format). Znamená to, že architektúra ich dátového modelu je štandardná a interoperabilná. CERIF je štandard, odporúčaný Európskou komisiou pre všetky členské štáty EÚ. Jeho podporu, rozvoj a šírenie zabezpečuje euroCRIS.

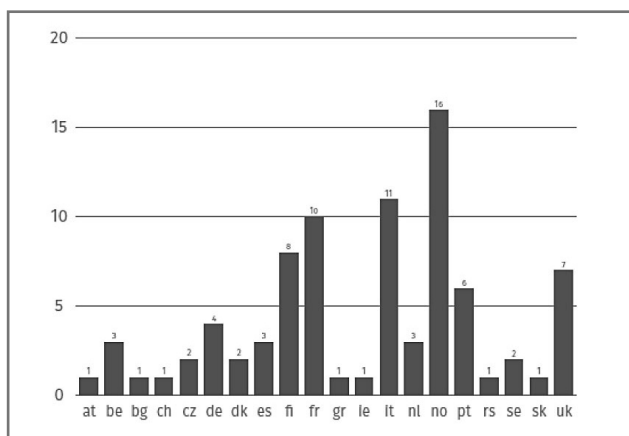
Inštitucionálny repozitár bol definovaný ako digitálna zbierka výstupov výskumu (hlavne publikácií a sád dát – dataset), určená na zhromažďovanie, uchovávanie a šírenie intelektuálnej produkcie univerzity alebo výskumnej inštitúcie.

Hlavným cieľom prieskumu bolo zistiť, akým spôsobom využívajú organizácie (predovšetkým univerzity) v európskych krajinách prevádzkované systémy CRIS a IR. Prieskum sa zamerával na technické riešenia, ich vzťahy s inými systémami, používanými na univerzitách, na interoperabilitu, dostupnosť dát a metadát a na ich použitie.

Prieskum hľadal odpoveď na dve základné otázky:

1. Sú inštitucionálne repozitáre postupne nahrádzané systémami CRIS?
2. Do akej miery sa prekrýva funkcionality týchto dvoch druhov systémov?

Prieskumu sa zúčastnilo 86 respondentov z 20-tich krajín, podrobnejšie údaje sú v tabuľke 1. Výber respondentov bol plne v režii dvoch organizátorov prieskumu. Otázky boli delené do dvoch častí, týkajúcich sa CRIS a repozitárov.



Obrázok 1 Účastníci prieskumu podľa krajín

Z analýzy získaných odpovedí vyplývalo, že 62 % dopytovaných inštitúcií prevádzkuje obe aplikácie, teda CRIS aj IR. 18 % z nich prevádzkuje CRIS aj IR ako dve funkcionality jednej a tej istej softvérovej aplikácie.

Prvá časť dotazníka bola zameraná na **systémy CRIS**.

Softvér

Najčastejšie používaná platforma, podporujúca CRIS a IR súčasne, bol systém PURE (Elsevier)[7] a IRIS (CINECA)[8]. Pre túto dvojitzú funkcionality sú prispôsobené aj viaceré systémy, vytvorené na mieru. Napríklad CRISin – celoštátny CRIS v Nórsku.

Funkcionality

Systémy CRIS obsahujú množstvo rôznych dátových modulov, najčastejšie sa v prieskume spomínali metadáta výskumných publikácií (81 %). Zaujímavé je, že vo viac ako 50 % dopytovaných systémov CRIS sa nachádzajú aj úplné texty výskumných publikácií, rovnako aj výskumné dáta.

Údaje o projektoch sú dostupné v 76 % systémov CRIS, podobne to platí aj v prípade funkcionality bližšie nešpecifikovaných výstupných zostáv (reporting) – 75 %. V 53 percentách systémov sa nachádza funkcionality podpory vkladania životopisov a správ o výskumnej činnosti (49 %), využitie ktorých na hodnotenie výskumníkov (31 %) sa zatiaľ zdá byť menej početné. Bibliometrické dáta sa nachádzajú v 46 % systémov, avšak ich analytické spracovanie je zriedkavé. Rovnako použitie CRIS na správu údajov o poplatkoch za spracovanie článkov (article processing charges – APC) je zatiaľ ojedinelé.

Interoperabilita

Otázky o interoperabilite boli rozdelené na interoperabilitu so systémami v rámci organizácie a s externými systémami. Z interných systémov je CRIS najčastejšie prepojený s personalistikou (HRM, 68 %) a IR (63 %). Viac údajov sa nachádza v tabuľke 2.

Spolupráca s externými systémami je rozvinutá v menšej miere. Najčastejšie sa týka systémov grantových agentúr. Viac údajov sa nachádza v tabuľke č. 3.

Protokoly, štandardy a slovníky

Ako vidno na obrázku č. 4, 50 % účastníkov prieskumu deklaruje, že ich CRIS podporuje protokol OAI-PMH, 41 % uvádza podporu CERIF a 32 % podporu systému identifikácie ORCID. Čo sa týka slovníkov, najpoužívanejšie z oblasti vedy sú CORDIS (5 %) a FOS (Field of Science and Technology Classification – 3 %), ako aj CASRAI (5 %).

Druhá časť dotazníka sa venovala **rezpozitárom**.

Softvér

Najviac respondentov (56 % – vrátane IRIS od CINECA) používa Dspace. Nasledujú rezpozitáre vyvinuté na mieru (16 %) a E-print (12 %). Niektorí prevádzkovatelia IR používajú tiež platformu ARL, knižnično-informačný systém Brocade, CRIS-tin, Fedora, Invenio a HAL, a tiež Pure, ktorý je však primárne používaný ako CRIS.

Na otázku, do akých iných systémov okrem rezpozitára ukladajú výsledky výskumu, odpovedalo 33 respondentov. Najčastejšie uvádzali systémy CRIS, špeciálne systémy určené pre evidenciu patentov, systémy pre záverečné práce, ako aj rezpozitáre na celoštátnej úrovni.

Obsah

Čo sa týka obsahu, IR evidujú hlavne publikácie (96 %) a záverečné práce (86 %). Učebné objekty (22 %) a sady dát (18%) sú zriedkavejšie. Odpovede sú uvedené v tabuľke č. 5.

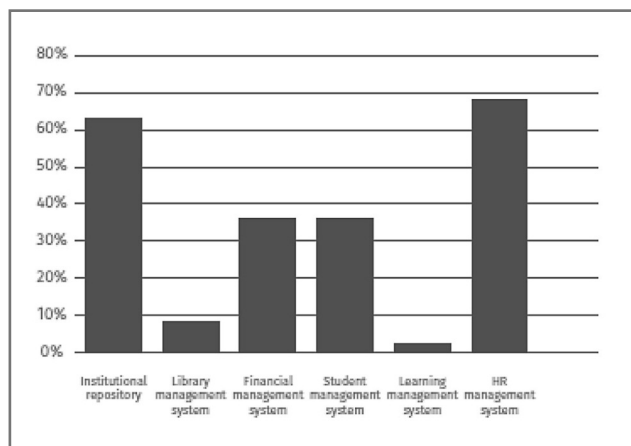
Samostatná otázka bola venovaná politikám otvoreného prístupu, existenciu ktorých deklarovalo 83 % respondentov. Počet mandátov je nižší (26 %). Niektoré organizácie deklarovali používanie oboch princípov.

Interoperabilita

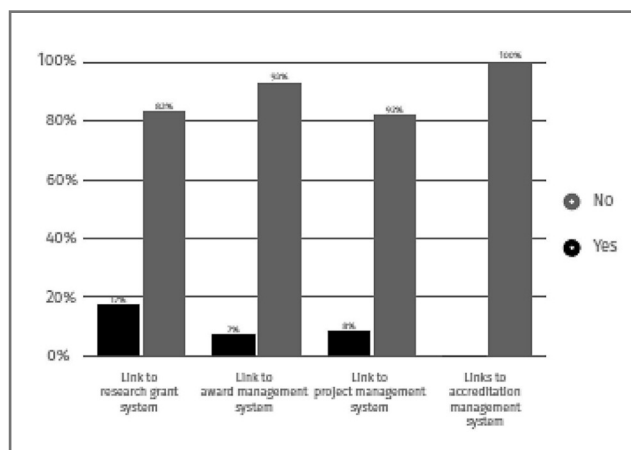
V prípade IR bola zisťované interoperabilita s ostatnými **inštitucionálnymi** službami. Najčastejšie sa vyskytovalo prepojenie so systémom manažmentu knižnice (47 %), na integrované vyhľadávacie systémy (38 %) a internetové stránky výskumníkov (36 %). Prepojenie na finančné systémy je minimálne (4 %).

Interoperabilita IR s externými systémami spočívala predovšetkým v kompatibilite OpenAIRE (69 %) Používanie trvalých identifikátorov ORCID je menej bežné (23 %).

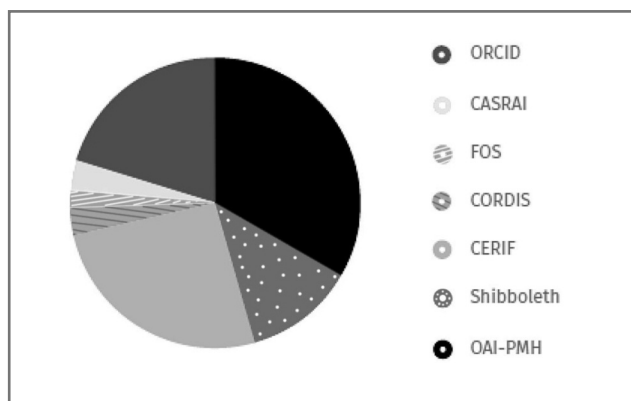
Využitie systémov CRIS a inštitucionálnych rezpozitárov na hodnotenie výskumu resp. na hodnotenie autorov alebo inštitúcie zatiaľ nie je veľmi časté (obrázok č. 6).



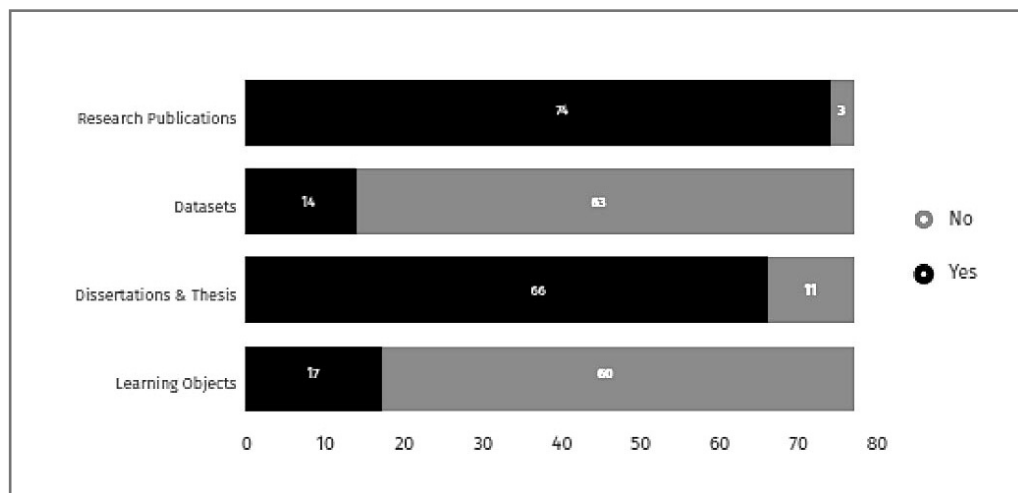
Obrázok 2 CRIS – interoperabilita so systémami organizácie



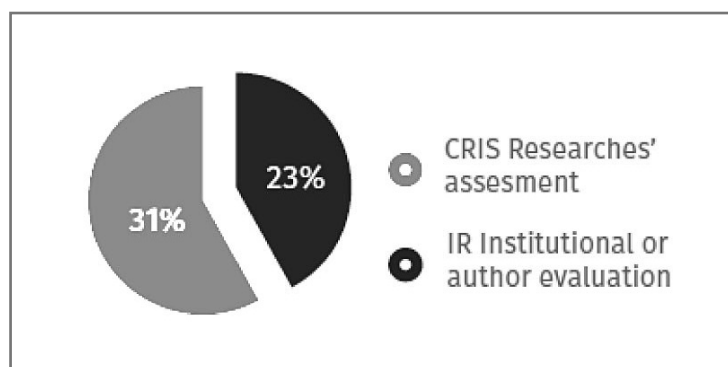
Obrázok 3 CRIS – interoperabilita s externými systémami



Obrázok 4 CRIS - protokoly, štandardy a slovníky



Obrázok 5 IR – obsah



Obrázok 6 Využitie CRIS a IR na hodnotenie – podiel kladných odpovedí

Autori prieskumu konštatujú, že odpoveď na v úvode položené dve otázky je záporná. Dva systémy, CRIS a IR sa vzájomne dopĺňajú. Kým IR je prvotne určený na správu publikácií a záverečných prác, úlohou CRIS je správa inštitucionálnych výskumných informácií, vrátane metadát pre výsledky a výskumu.

K podobným záverom ohľadne interoperability systémov CRIS a repozitárov sa dopracovala aj medzinárodná pracovná skupina Knowledge Exchange (<http://www.knowledge-exchange.info>), pozostávajúca zo zástupcov šiestich kľúčových národných organizácií, vyvíjajúcich infraštruktúru a služby týkajúce sa použitia digitálnych technológií na skvalitnenie vysokoškolského vzdelávania a výskumu a na podporu otvoreného prístupu.

Míting v Kodani sa venoval problematike monitoringu OA publikácií a monitoringu údajov o nákladoch na publikovanie OA publikácií (tzv. APC – Article Publication Charges) a formuloval odporúčania v tejto oblasti. Odporúčania stručne a výstižne reflektujú základné úlohy na zabezpečenie interoperability metadát o publikáciách a možno ich zhrnúť do 3 oblastí podľa určení zainteresovanej strany, ktorej sa týkajú. [2]

1. Current Research Information Systems (CRIS)

Ukázalo sa, že kľúčovým faktorom pre úspešné monitorovanie publikácií OA sú systémy CRIS, ktorých sa týkajú nasledujúce odporúčania:

- Zlepšiť integráciu medzi systémami CRIS a inštitucionálnymi repozitármi
- Integrovať v systémoch CRIS rôzne kategórie (výsledkov výskumu, publikovaného v rámci) otvoreného prístupu
- Zabezpečiť, aby systémy CRIS dodržiavali pravidlá interoperability s OpenAIRE
- Používať systémy CRIS na zber dát o publikáciách zeleného aj zlatého otvoreného prístupu
- Prepojenie účtovných systémov a CRIS

2. Vydavatelia

Vydavatelia sú základnými aktérmi v procese monitorovania, pretože sú nositeľmi potrebných informácií o publikáciách a príslušnosti (k organizácii) autorov a o nákladoch. Vydavatelia boli preto zahrnutí do mnohých odporúčení.

Vydavatelia by na evidenciu príslušnosti autorov mali používať štandardizované dátové formáty. Do prehľadu Crossref¹ by mali zaradiť licenčné vyhlásenie pre každú publikáciu a uviesť, v akej forme otvoreného prístupu je publikácia zverejnená (zelená, zlatá alebo hybridná forma otvoreného prístupu). Knižnice a financujúce organizácie by zasa mali od vydavateľov vyžadovať príslušné metadáta, a zahrnúť tieto požiadavky do zmlúv o kompenzáciách alebo licenčných zmluvách. Spoločnosť Jisc vyvinula šablónu, ktorú je možné použiť alebo ďalej rozvíjať.

Proces, v ktorom vydavatelia odošlú svoju faktúru za publikovanie s otvoreným prístupom (APC), by mal byť štandardizovaný. Vydavatelia by mali byť vyzvaní, aby do svojich faktúr zahrnuli všetky metaúdaje súvisiace s nákladmi a licencovaním: DOI, informácie o zadávateľovi a licencii, mená autorov a príslušnosť, identifikátory poskytovateľa grantu a ORCID.

3. Knižnice

Knižnice hrajú v procese monitorovania publikácií a nákladov dôležitú úlohu. Mali by od vydavateľov v rámci uzavretých kontraktov požadovať všetky potrebné údaje. Mali by konať spoločne a pracovať v kontextoch väčších konzorcií a spolupracovať medzi krajinami, aby predložili jednotné požiadavky vydavateľom (napr. Účinnosť a normy pre poplatky za otvorený prístup/ Efficiency and Standards for Open Access Article Charges, skratka ESAC).

Na nutnosť integrácie systémov poukazuje aj Jyrki Ilva [5] v článku, týkajúcom sa používania systémov CRIS a IR v škan-dinávskych krajinách, keď konštatuje: „CRIS a repozitár ako pravdepodobné koncepcie budú pravdepodobne existovať v budúcnosti, zdá sa však, že integrácia týchto dvoch systémov zostane populárnym cieľom... Zatiaľ čo úložiská vykonávajú dobrú prácu pri poskytovaní prístupu k publikáciám a iným digitálnym materiálom, moderný výskumný informačný systém ponúka sofistikovanejšie formuláre na vkladanie dát, formáty metadát a pracovné postupy prispôbené potrebám výskumnej správy, knižníc a vedcov.“

Daniela Luzi [6] už v roku 2012 upozorňovala na početné novovytvárané repozitáre otvorených dát a analyzovala možnosť ich prepojenia so systémami CRIS.

„Štandardizovaný opis metadát projektov prepojených s novou entitou CERIF Result_dataset poskytuje kontext, v ktorom sa vytvárajú výskumné dáta, a určite zlepšuje ich vyhľadávanie a opakované použitie spájajúc rôzne typy údajov získaných v rovnakom alebo súvisiacom projekte.“ Samozrejme za podmienky, že budované repozitáre výskumných dát obsahujú identifikáciu projektu (ID, názov, akronym a pod.) Poukazuje na skutočnosť, že „...bohatý vzťah modelu CERIF spolu s dobre vymedzenou sémantikou môže vylepšiť archívy výskumných dát aj CRIS, napríklad jasným rozlíšením autorov dát od ich vlastníkov, ich príslušnosti alebo uvedením úlohy inštitúcie ako vlastníka autorských práv k dátam, uvedením spolupracujúcej organizácie alebo financujúcej agentúry.“

Záverom možno konštatovať, že štandardy, odporúčania aj technické predpoklady pre interoperabilitu výskumných informácií v celoeurópskom meradle sú vytvorené. V európskom meradle je braná do úvahy aj spoločenská potreba, dokonca nevyhnutnosť vzájomného prepojenia systémov podpory vedy.

U nás, kde zatiaľ stále dominuje podporná infraštruktúra vedy a výskumu na národnej úrovni na úkor tej inštitucionálnej, by sme sa mali oboznámiť s príkladmi osvedčenej praxe a nájsť v nich inšpiráciu. Je potrebné využiť všetky dostupné zdroje a budovať výskumné databázy aj repozitáre, digitalizovať, uchovávať a hlavne vytvorené zdroje informácií sprístupňovať a navzájom logicky prepájať.

Zdroje

1. Ligia Ribeiro; Pablo de Castro; Michele Mennielli: EUNIS – EUROCRIS joint survey on CRIS and IR. ERAI (Eunis Research and Analysis Initiative), 2016. Online: <http://www.eunis.org/wp-content/uploads/2016/03/cris-report-ED.pdf>
2. Niels Stern: Knowledge Exchange consensus on monitoring Open Access publications and cost data. Report from workshop held in Copenhagen 29-30 November 2016. Online: <https://zenodo.org/record/480852#.WRmjciZF3cs>
3. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0. Online: <https://www.force11.org/node/6062>
4. <http://www.informatizacia.sk/interoperabilita/3481s>
5. Jyrki Ilva :Integrating CRIS and repository – an overview of the situation in Finland and in three other Nordic countries. Online: https://www.doria.fi/bitstream/handle/10024/97606/OR2014_CRIS%2Brepositories_in_Nordic_Countriesfinal.pdf?sequence=3
6. Daniela Luzi, Rosa Di Cesare, Roberta Ruggieri: Toward the integration of datasets in the CRIS environment: A preliminary analysis. Online: http://dspacecris.eurocris.org/bitstream/11366/104/1/CRIS2012_Luzi_datasets.pdf
7. Elsevier. PURE. Online: <https://www.elsevier.com/solutions/pure>
8. CINECA. IRIS – Institutional Research Information System. Online: <https://www.cineca.it/en/content/iris-institutional-research-information-system>

Ing. Danica Zendulková

danica.zendulkova@cvtisr.sk

(Centrum vedecko-technických informácií SR)

¹ Crossref je oficiálna agentúra, registrujúca DOI (Digital Object Identifier). Web: <https://www.crossref.org/>