

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



**Author identity,
ISNI – <http://www.isni.org/>,
ORCID – <http://www.orcid.org>**

PV070: ESEJ

Juraj Ondruška

Brno, 9.1.2011

Úvod

Identifikácia autorov patrí medzi problémy v identifikácii. Je dôležitá pri určovaní kto aký článok napísal, vyhľadávaní a pridelovaní mu za to ocenení. Najjednoduchšie by bolo k identifikácii používať meno autora, ale keďže existuje viac spôsobov ako zapísať meno, najmä u ľudí s dlhšími viacslovnými menami, nie je táto cesta optimálna. Problémom je aj fakt, že jedno meno môže náležať viacerým osobám, prípadne sa meno môže meniť najmä u žien ktoré sa vydajú, alebo môže byť meno uvedené v rôznych tvaroch. Kvôli tomu sa začali vyvíjať rôzne iné identifikátory, ako napríklad International Standard Name Identifier (ISNI), Open Researcher and Contributor ID (ORCID) a mnoho ďalších.

ISNI

POPIS PROJEKTU

Projekt je momentálne ešte len v stave návrhu medzinárodnej normy (Draft International Standard – DIS) 27729. Je vyvíjaný pod záštitou Medzinárodnej organizácie pre normalizáciu (International Organization for Standardization – ISO). ISNI umožňuje identifikáciu ľudí, skupín, alebo fiktívnych osôb (napríklad pseudonym, alebo vymyslená postava), pričom je určený k priamemu prístupu k informáciám o verejnej identite, ale dokáže sa prepájať aj s inými systémami, v ktorých sú takéto informácie držané. Mnoho databáz si pri používaní ISNI môže vymieňať informácie bez používania takých chaotických metód ako je porovnávanie reťazcov. ISNI jednoznačne identifikuje verejné identity naprieč rôznymi odbormi tvorivej činnosti. Jeho uvedenie malo byť v roku 2010.

ISNI bude riadené Medzinárodnou agentúrou (International Agency), obvyčajne známou ako ISNI-IA. Táto nezisková organizácia registrovaná vo Veľkej Británii bola založená konzorciom organizácií skladajúcich sa z Medzinárodnej konfederácie spoločností autorov a skladateľov (Confédération Internationale des Sociétés d'auteurs et Compositeurs – CISAC), Konferencie európskych národných knižníc (the Conference of European National Librarians – CENL), Medzinárodnej federácie reprodukčných práv organizácií (the International Federation of Reproduction Rights Organisations – IFRRO), Medzinárodnej interpretačnej databázovej asociácie (the International Performers Database Association - IPDA), Online počítačového knižničného centra (the Online Computer Library Center – OCLC) a ProQuest-u. Z Konferencie európskych knižníc je to Francúzska národná knižnica (Bibliothèque nationale de France) a Britská knižnica (British Library).

ISNI bolo navrhnuté ako premostovací identifikátor, ako otvorená vrstva nad vlastným identifikačným systémom. Zachováva iba minimálne metadáta potrebné na rozlíšenie dvoch identít, všetky ostatné informácie zostávajú v databázach zabezpečených podmieneným prístupom.

STAV

ISNI bude v prvý deň obsahovať 3 milióny záznamov. Akonáhle bude databáza vytvorená, bude pracovať hybridným centralizovaným – decentralizovaným spôsobom. Predpokladaná veľkosť ISNI je okolo 20 miliónov a predpokladaný ročný prírastok je 600 000 ročne. Tieto odhady zahŕňajú už aj spomínané 3 milióny identít, ktoré sú autori článkov, ktorí zatiaľ neboli identifikovaní ako tvorcovia diel katalogizovaných knižnicami. Medzinárodná asociácia vedeckých, medicínskych a technických publikovateľov (The International Association of Scientific, Medical and Technical Publishers – RTM) oznámila, že tu je 7 miliónov tvorcov článkov a ďalších 210 000 pribudne každý rok. ISNI zatiaľ nie je štandardom, je to iba Draft, ale má byť formálne publikované čoskoro. ISNI je zadarmo, ale Registračné agentúry si účtujú poplatky za vedenie databázy.

TVAR ISNI

ISNI tvorí 16 číslic v desiatkovej sústave, posledná číslica je kontrolná.
Príklad ISNI: 1422 4586 3573 0476

PRÍKLAD KOMPLETNÉHO ZÁZNAMU

ISNI: 3456 7890 3456 666X

Meno a Priezvisko: Pavol Országh Hviezdoslav

Dátum narodenia: 2.2.1849

Dátum úmrtia: 8.11.1921

Trieda: Literárna

Úloha: Autor

Tvorba: <http://www.osobnosti.sk/index.php?os=zivotopis&ID=1726>

Informácie: <http://www.osobnosti.sk/index.php?os=zivotopis&ID=1726>

PRIRAĐOVANIE ISNI

ISNI-IA bude používať priradovací systém zahrňujúci rozhranie, data-schémy, ne-
dvojznačné algoritmy a databázy, ktoré spĺňajú požiadavky normy ISO a zároveň budú
používať existujúce technológie tam, kde je to možné. Systém je primárne založený
na Virtuálnom medzinárodnom súbore autorít (Virtual International Authority File –
VIAF). VIAF je projekt niektorých národných knižníc a prevádzkuje ho OCLC, jeho
vývoj začala Nemecká národná knižnica a Kongresová knižnica. VIAF bol vyvinutý
kvôli hromadeniu sa knižničných katalógov v OCLC.

Prístup k priradovaciemu systému, k databáze a k číslam, ktoré sú generované ako vý-
sledok procesu bude kontrolovaný nezávislými inštitúciami nazývanými Registračné
agentúry (Registration Agencies). Registračné agentúry budú priamo komunikovať
so zákazníkmi, zabezpečovať, že údaje sú v príslušných formátoch a rekompensovať
ISNI-IA náklady na údržbu priradovacieho systému. Registračné agentúry bude me-
novať ISNI-IA, ale riadené a financované budú nezávisle.

ZÁKLADNÉ PRAVIDLÁ PRE PRIRADENIE ISNI

1. Každá verejná identita bude mať svoje vlastné špecifické ISNI číslo, napríklad Pavol Országh Hviezdoslav bude mať jedno ISNI a Božena Němcová zase iné.
2. Günter Graß, Guenter Grass a Guenter Graß sú spôsoby pre zapísanie toho istého mena (v nemčine „ss“ je takmer to isté ako „ß“, ďalej „ü“ sa prepíše na „ue“ pri preložení do angličtiny) a pripadá im rovnaké ISNI, pretože je to jedna a tá istá osoba.
3. Ciaikovskij Piotr Iljč a jeho meno v azbuke sú transliterály, prvá varianta je prepis mena z azbuky – druhej varianty, takže je to tá istá verejná identita a teda jej pripadá jedno ISNI.
4. Pyotr Tchaikovsky and Peter Tchaikovsky sú rôzne jazykové varianty (vznikli pri prepise z azbuky) rovnakej verejnej identity a teda jej pripadá jedno ISNI.

MINIMÁLNE INFORMÁCIE POTREBNÉ PRE PRIRADENIE ISNI

1. Názov verejnej identity.
2. Dátum a miesto narodenia a alebo úmrtia (alebo registrácie a zániku u právnických osôb).
3. Triedy a úlohy ako sú definované v RA. Triedy definuje repertoár (napríklad hudobné, audio-vizuálne, literárne, ...) a úloha môže byť autor, publikovateľ, ...
4. Názov, alebo odkaz na tvorbu.
5. URI (alebo URL) odkazujúce na detailnejšie informácie o verejnej identite.

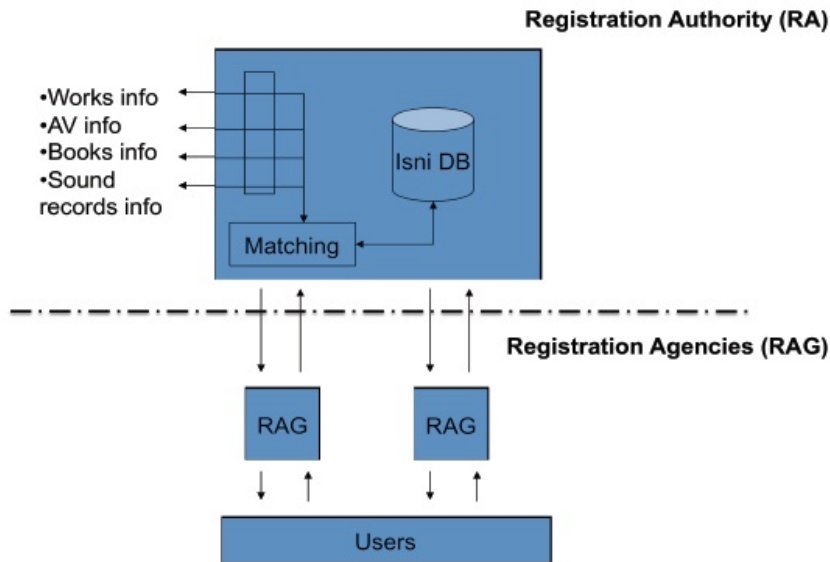
ISNI ORGANIZÁCIA

1. **Registračný úrad (Registration Authority)**
Za registračný úrad bolo menované ISNI-IA. Má na starosti vytváranie a udržiavanie referenčnej databázy. Nebude poskytovať priame služby pre zákazníkov, to majú na programe registračné agentúry. Ďalej bude zodpovedné za celkovú administráciu a správu ISNI Štandardu. Pôvodná databáza bude vytvorená zo súčasných zakladajúcich členov
2. **Registračné agentúry (Registration Agencies)**
Registračné agentúry sú menované ISNI-IA. Budú pôsobiť ako rozhranie medzi užívateľmi a registračným úradom. Registračnou agentúrou sa môže stať akýkoľvek podnikateľský subjekt. Registračné agentúry budú mať neobmedzený prístup do celej databázy avšak iba na čítanie (read-only). Agentúry sú len subjekty oprávnené vydávať ISNI. Všetky žiadosti užívateľov na získanie ISNI budú vybavované prostredníctvom agentúr.

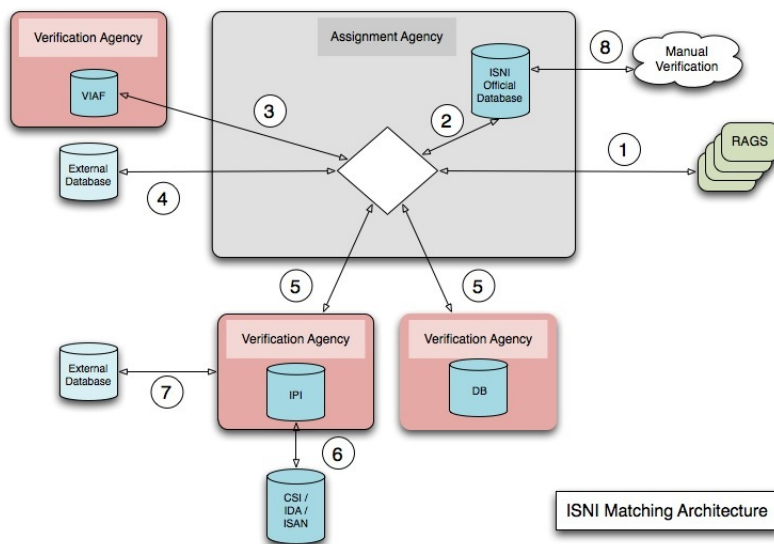
3. Overovacie agentúry (Verifikation Agencies)

Sú menované registračným úradom. Budú pôsobiť ako špecializované zodpovedajúce služby Registračnému úradu. Porovnávajú dáta s externými databázami, teda riadia proces overovania.

UKÁŽKA



Registračný úrad bude stavať a udržiavať databázu spájaním všetkých prichádzajúcich metadát od Registračných agentúr a priradovaním ich k verejným identitám. Registračné agentúry budú prevádzkovať všetky nástroje a služby zákazníkom.



Verifikačné agentúry overujú dáta pomocou externej databázy

DATABÁZA

ISNI	Meno	Priezvisko	Trieda	Rola
0000 1000 0000 1774	John	Smith	Literárna	Autor
0000 1000 0000 115 x	Sam	Davis	Hudobné práce	Skladateľ
0000 1000 0000 1264	West Side	Inc	Literárna	Vydavateľ

ORCID

POPIS PROJEKTU

ORCID má byť štandardom de facto. Tento projekt rieši problém mien spôsobom vytvorenia centrálného registru unikátnych identifikátorov. Ďalej prináša možnosť prepojenia s inými súčasnými identifikačnými systémami na zvýšenie efektivity vyhľadávania. ORCID identifikuje vedeckých a iných akademických autorov. Podobne ako ISNI aj ORCID rieši problém nejednoznačnosti mien. Je určený na ponúknutie otvoreného a nezávislého registra. Prototyp má byť sprevádzkovaný na softvéri, ktorý používa Thomson Reuters pre ich Researcher ID systém. Výsledný systém má bežať v nezávislej organizácii, ktorá bude pridelať ORCID, ten má byť voľne použiteľný každou takou skupinou a spolupracovať s ich identifikačnými systémami. ORCID má sprostredkovať lepšiu komunikáciu medzi vedeckými organizáciami.

Vytváranie, upravovanie a udržiavanie ORCID profilu je zadarmo. Všetky profilové dáta sú zadarmo dostupné. Všetok software vyvinutý ORCID-om bude zadarmo dostupný pod licenciou Open Source. ORCID identifikátory a dáta budú k dispozícii kombináciou zdarma a za poplatok a ceny budú nastavené na udržanie ORCID-u ako neziskovej organizácie.

Vedú sa diskusie o centralizovaní verzus združovaní, ale vyzerá to tak, že ORCID bude centralizovaný systém podobný DOI (Digital Object Identifier). Vedci budú mať nad svojím ORCID záznamom veľkú kontrolu a to nie len kvôli ochrane osobného súkromia, ale aj preto, že je to najjednoduchší spôsob ako opraviť chyby, ktoré vytvoria aj tie najsofistikovanejšie algoritmy v procese automatického priradenia autora. Ako som spomínal ORCID bude používať software Researcher ID od spoločnosti Thomson Reuters. Open ID nebude centrálnou časťou ORCIDU, ale ORCID bude určite postavený tak, aby dokázal pracovať s Open ID a ďalšími autentizačnými mechanizmami. ORCID bude spravované jednou nezávislou organizáciou, pravdepodobne to bude Cross-Ref.

STAV

ORCID ešte len bude prijatý, ak sa stane široko používaným, bol predstavený v decembri 2010.

TVAR ORCID

ORCID má alfanumerický tvar.

Príklad: 447766410380

RIADENIE

Niekoľko organizácii prispelo dátovými zdrojmi a technológiou na podporu počiatočného prototypu ORCID, patrí sem napríklad: Researcher ID profilového systému od Thomson Reuters a bibliografické údaje z metadatabázy CrossRef.

ORCID bol oficiálne spustený v auguste 2010, je riadený štrnástymi členmi predstavenstva:

1. **Wellcome Trust**
2. **Harvard University**
3. **Wiley-Blackwell**
4. **Hannover Medical School**
5. **OCLC**
6. **Thomson Reuters**
7. **CERN – Európska organizácia pre jadrový výskum**
8. **CrosdsRef**
9. **Nature Publishing Group**
10. **Association for Computing Machinery, Inc**
11. **Elsevier**
12. **Massachusetts Institute of Technology Libraries**
13. **National Institute of Informatics (NII) (Japonsko)**
14. **Cornell University Library**

Zhodnotenie prínosu

Zlepšenie identifikácie autorov by malo so sebou priniesť zdokonalenie systému rozoznávania. Podobne ako iné dnešné systémy aj identifikácia autorov poskytuje zvýhodnenie a uľahčenie práce pre užívateľa. ORCID sa špecializuje na užšiu skupinu autorov a je zadarmo. ISNI toho zahŕňa viac nesie možnosť registrácie napríklad aj fiktívnych osôb, no registrácia je platená, pretože treba udržiavať databázu a poskytovať služby užívateľom. Zlepšená identifikácia bude mať obrovský prínos pre vyhľadávače.

Použitá literatúra

<http://www.orcid.org>

<http://www.isni.org>

<http://en.wikipedia.org/wiki/ORCID>

<http://en.wikipedia.org/wiki/ISNI>

<http://blogs.nature.com/mfenner/2010/01/03/>

[orcid-or-how-to-build-a-unique-identifier-for-scientists-in-10-easy-steps](http://blogs.nature.com/mfenner/2010/01/03/orcid-or-how-to-build-a-unique-identifier-for-scientists-in-10-easy-steps)

Metadáta v Dublin Core

```
<link rel="schema.DC"href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title"content="Author identity" />
<meta name="DC.Creator"content="Juraj Ondruška" />
<meta name="DC.Subject"content="Author identity" />
<meta name="DC.Description"content="Spôsoby riešenia nejednoznačností v rozoznávaní autorov" />
<meta name="DC.Date"content="28.11.2010" />
<meta name="DC.Type"content="Text" />
<meta name="DC.Type"content="Esej" />
<meta name="DC.Format"content="application/pdf" />
<meta name="DC.Format"content="computerFile" />
<meta name="DC.Source"content="http://www.isni.org" />
<meta name="DC.Source"content="http://en.wikipedia.org/wiki/ISNI" />
<meta name="DC.Source"content="http://www.orcid.org" />
<meta name="DC.Source"content="http://en.wikipedia.org/wiki/ORCID" />
<meta name="DC.Language"content="sk" />
```

Obsah

Úvod, ii

ISNI, ii

Literatúra, viii

Metadáta, viii

ORCID, vi

Zhodnotenie prínosu, vii