

Perzistentné identifikátory digitálnych informačných entít – štandardy, vybrané systémy a aktuálne trendy v kontexte vedeckej komunikácie

Jasná identifikácia informačnej entity je v kontexte rýchlej a efektívnej komunikácie nevyhnutnosťou. V rôznych informačných procesoch sú informačné objekty reprezentované symbolickými znakovými reťazcami – identifikátormi. Globálny charakter komunikácie spojil požiadavku jednoznačnosti a jedinečnosti identifikátora s ďalšími požiadavkami – systémovou nezávislosťou a použiteľnosťou naprieč systémami (interoperabilita), trvalým charakterom (perzistentnosť) a možnosťou ďalšieho rozširovania (extenzibilita). Dôležité je aj jednoduché použitie v čo najširšom prostredí. Aktuálne tieto požiadavky reprezentujú dva platformovo nezávislé systémy, ktoré vznikli na pôde neziskových organizácií a získali globálnu autoritu – identifikátory DOI a ORCID.

Rozmach elektronického komunikačného (a publikačného) priestoru vytvoril tlak na vytvorenie a používanie systému špecifických symbolických reprezentantov informačných objektov/zdrojov. Perzistentné identifikátory digitálnych informačných entít prinášajú do oblasti informačných a komunikačných procesov novú kvalitu. Umožňujú objekt nielen identifikovať, ale zvyšujú jeho vizibilitu a umožňujú integráciu do ďalších informačných procesov. Medzi základné požiadavky na systémy identifikácie digitálnych informačných entít patria tri:

- interoperabilita
- perzistentnosť
- extenzibilita

a k tomu, samozrejme, globálny charakter. Významným momentom v tomto kontexte bol rozvoj digitálnych knižníc. Digitálne knižnice a aktivity spojené s týmto fenoménom stáli pri zrode niekoľkých systémov identifikátorov, ktoré umožnili nastaviť a aplikovať štandardy následne umožňujúce využívať potenciál vytváraného a komunikovaného digitálneho obsahu v prostredí heterogénnej infraštruktúry a systémových riešení. Boli tu však určité limity, ktoré sa stali hranicou ich rozšírenia. Neustále je potrebné mať na pamäti, že každý identifikátor je vlastne dohodnutá a akceptovaná konvencia, kde práve miera a rozsah akceptácie na strane používateľov výrazne determinuje používanie, a tým aj ďalší vývoj (Coyle, 2006).

Pre zabezpečenie interoperability bolo potrebné navrhnuť identifikátory umožňujúce ich používanie mimo kontrolovaného prostredia tvorca. Interoperabilitu však možno interpretovať aj v inom, širšom význame ako „komunikáciu medzi systémami“ a síce ako požiadavku „interoperability s budúcnosťou“. Je tu teda potreba trvalého charakteru (perzistencia). No a v neposlednom rade interoperabilita implikuje požiadavku identifikátora, ktorý možno rozširovať na základe potrieb rôznych autorít, metadátových štruktúr a pod., a to buď na báze konzistentného dátového modelu, alebo pridelovaním identifikátorov novým entitám v kompatibilnom móde (Paskin, 2005).

DOI: Digital Object Identifier

DOI (<http://www.doi.org>) je registrovanou obchodnou známkou (DOI®). Schéma vychádza zo špecifikácie medzinárodného štandardu ISO 26324:2012: Information and documentation – Digital object identifier system, ktorý špecifikuje syntax, popis, rozlíšenie funkčných prvkov a všeobecné princípy pre vytváranie, registráciu a správu mien DOI; nešpecifikuje technológiu. Nie je náhradou ani alternatívou žiadneho identifikátora používaného v rámci iných schém, je to unikátny identifikátor digitálneho informačného objektu v sieti internet. Je priradený k objektu tak, aby poskytoval trvalé prepojenie s objektom/pripojenie k tejto informačnej entity. Unikátnosť DOI je vynútená systémom identifikácie, nesmú byť pridelené dva rovnaké identifikátory tej istej entite, resp. jeden identifikátor nesmie byť pridelený dvom odlišným entitám). Dohodnutý systém udržateľnosti funguje vďaka technickej a sociálnej infraštruktúre, ktorú poskytuje a udržiava nezisková členská organizácia International DOI Foundation (IDF), a to už od roku 1998. IDF je riadiacim a dozorným orgánom federácie registračných agentúr poskytujúcich služby a registráciu identifikátorov digitálnych objektov a registračnou autoritou v zmysle vyššie uvedeného štandardu ISO.

Syntax DOI (štandardizácia ISO 26324:2012) – kombinácia unikátnej predpony/prefixu (kódové označenie registra a numerický reťazec priradený k jednému konkrétnemu registrujúcemu subjektu) a unikátnej prípony/sufixu (priradené jednému konkrétnemu digitálnemu objektu) – umožňuje decentralizovanú alokáciu názvom/mien DOI. Kým tvar prefixu je záväzný, tvar sufixu je ponechaný na registrátorovi a konvencii prijatej interne, musí však spĺňať určité všeobecne dané podmienky. Prefix a sufix sú oddelené lomkou, tvar môže byť v podobe numerického, alebo alfanumerického reťazca.

Príklad:

10.1000/1247.21522

prefix/sufix

10.1000/aate-no14578

prefix/sufix

kde: 10 je číslo registra, 1000 je identifikátor registrátora (napr. vydavateľ'a) a čísla za lomkou konvenciou dohodnuté meno informačného objektu.

Podľa aktuálneho odporúčania pre zápis DOI sa používa plnohodnotný aktívny odkaz v tvare <https://doi.org/10.xxxxx/xxxxxxx>.

Príklad: <https://doi.org/10.15414/afz.2019.22.01.21-25>

Systém DOI funguje na princípe súčinnosti (federácie) registračných agentúr (podobne fungujú aj iné registračné systémy – napr. ISBN a ISSN). Register DOI je súbor jedinečných identifikátorov s informáciami o každej obsahovej položke/digitálnom objekte (vrátane metadát) registrovanej prostredníctvom registrujúceho subjektu (registrátora), ktorým môže byť osoba alebo organizácia. Registrátor má následne povinnosť uchovávať/udržiavať digitálny objekt a v prípade akejkoľvek zmeny na strane objektu zabezpečiť funkčnosť prepojenia. Každá agentúra sa zameriava na určitú konkrétnu oblasť obsahu, takže výsledný register pokrýva informačné entity rovnakého/spoločného typu obsahu. Aktuálne v rámci systému DOI poskytuje služby 10 registračných agentúr:

- *Airiti, Inc.* (DOI pre materiály v tradičnej čínštine).
- *China National Knowledge Infrastructure (CNKI)* (informačné zdroje z/o Číne vrátane politiky, ekonómie, humanitných vied, sociálnych vied, vedy a techniky).
- *Crossref* (akademický/vedecký a odborný obsah, okrem registrácie aj doplnkové či už platené, alebo bezplatné služby).
- *DataCite* (infraštruktúra a obsah spojený s vedeckými dátami/datasetmi, zameriava sa na spoluprácu s dátovými centrami a organizáciami, ktoré uchovávajú dáta).
- *Entertainment Identifier Registry (EIDR)* (register filmov, televíznych relácií a iných komerčných audio/video prostriedkov).
- *The Institute of Scientific and Technical Information of China (ISTIC)* (podpora odkazovacích (linking) služieb pre čínske časopisy, dizertačné práce, knihy, zborníky z konferencií a iné zdroje, multimedálne zdroje a správa vedeckých dát v Číne).
- *Japan Link Center (JaLC)* (register a verejné informačné služby na podporu vedy a techniky v Japonsku, spolupráca s Crossref pri registrácii DOI pre japonské časopisy vychádzajúce v angličtine).
- *Korea Institute of Science and Technology Information (KISTI)* (riadenie projektov na ukladanie, prístup a zdieľanie vedeckých údajov v Kórei).
- *Multilingual European DOI Registration Agency (mEDRA)* (perzistentný referenčný systém pre internetové dokumenty).
- *Publications Office of the European Union (OP)* (v pozícii oficiálneho vydavateľa inštitúcií, orgánov, úradov a agentúr Európskej únie, zodpovedá za pridelenie DOI).

Takmer všetky agentúry rozvíjajú svoju činnosť na princípe (niekoľkých typov) členstva a finančnej participácii formou ročných členských príspevkov a poplatkov za služby. Okrem jasne deklarovaných povinností majú členovia právo spolupodieľať sa na rozhodovacích procesoch a dozornej činnosti.

V akademickom prostredí (a vo vydavateľskom prostredí) majú najsilnejšiu pozíciu registračné agentúry Crossref a DataCite.

Crossref aktuálne ponúka registráciu obsahu typu:

- Časopisy a články z časopisov, vrátane suplementárnych materiálov.
- Knihy, kapitoly z kníh a referenčné práce.
- Konferenčné zborníky (celok/príspevky).
- Správy/pracovné materiály.
- Štandardy/normy.
- Datasetsy/databázy.
- Dizertácie.
- Preprinty.
- Oponentské posudky, prehľadové správy.
- Ilustračné/doplnkové komponenty (obrázky, tabuľky...).
- Granty.

Obsah možno registrovať manuálne prostredníctvom webového formulára, alebo využitím dostupných online mechanizmov (napr. publikačné systémy, ktoré majú zabudovanú definovanú schému XML). Suplementárne materiály a špecifické metadáta sa vkladajú do systému formou importov súborov predpísaných/podporovaných formátov (napr. XML, CSV, TXT...). V prípade potreby opravy (napr. zmena URL zdroja, formálna chyba a pod.) sú na stránkach podpory Crossref k dispozícii opäť šablóny pre spracovanie dátovej dávky pre opravu, komunikuje sa prostredníctvom kontaktu na stránke Crossref.

ORCID

Potreba unikátneho identifikátora vedca/autora/pôvodcu obsahu v kontexte jeho jednoznačnej identifikácie v digitálnom komunikačnom prostredí vyústila do celého radu aktivít, ktoré sústredili svoje úsilie na vytvorenie „unikátnej DNA“ v globálnej sieti. Vlastné systémy zaviedli a používali/používajú napr. databázy vytváraním autoritných záznamov personálnych a korporátnych autorít. Asi najznámejším príkladom môžu byť databázy Web of Science alebo Scopus (ResearcherID a Author ID). Stále to však boli lokálne identifikátory, založené síce na určitých štandardoch, ale... Riešenie, s ktorým prišla nezisková organizácia ORCID, aktuálne integrovali do svojich rozhraní aj vyššie spomínané databázové informačné produkty práve z dôvodu funkcionalít, ktoré umožňuje. ORCID je nezisková organizácia členských organizácií, vrátane výskumných inštitúcií/subjektov, vydavateľov, donorov, profesionálnych združení a iných subjektov z prostredia vedeckého výskumu. Buduje register vedcov a ďalších prispievateľov vedeckého obsahu založený na identifikátore rovnakého mena: **ORCID** (Open Researcher and Contributor ID) a je súčasťou globálnej digitálnej infraštruktúry pre výskumných pracovníkov na zdieľanie informácií. Finančne svoju činnosť pokrýva z príspevkov členov. Participácia je možná dvomi spôsobmi:

Korporátne členstvo (s členským poplatkom, inštalácie na úrovni korporátnych informačných systémov a integrácia s ďalšími korporátnymi informačnými systémami).

Individuálna registrácia (bezplatne) používateľa, aktívna správa účtu a informácií, vrátane možnej integrácie informácií z iných systémov (napr. publikácie z databáz WoS, Scopus, Medline...).

Identifikátor ORCID má tvar 16 miestneho numerického reťazca – 0000-0001-8165-3369. Štandardný zápis v súlade s používanou konvenciou: <http://orcid.org/0000-0001-8165-3369>

Identifikátor je nezávislý, globálny a jedinečný nástroj na zvýšenie vlastnej viditeľnosti v sieti, získanie širšieho vplyvu, ako aj nástrojom prezentujúcim výsledky práce. Významným benefitom v tomto smere je možnosť priamej komunikácie s prostredím vybraných bibliografických a bibliometrických databáz (Web of Science, Scopus, Medline...) a zdieľania dát z ich prostredia, alebo možnosť jednoduchého importu dát v podporovanom formáte (napr. TXT, BibTeX...) do vlastného profilu.

Benefity DOI a ORCID

Keby sme chceli jednoduchým spôsobom zhrnúť benefity systémov DOI a ORCID v kontexte súčasných potrieb vedeckej komunikácie, môžeme uviesť:

1. Zvýšenie viditeľnosti v digitálnej sieti (vydavateľa, časopisu, vedca, inštitúcie...).
2. Pozitívny posun v oblasti metrík (citovanosť...).
3. Rozšírenie možností spolupráce.
4. Transparentnosť aktivít a ich výsledkov.
5. Aktívny manažment aktivít a ich výsledkov (napr. projekty, publikácie).
6. Globálny charakter a udržateľnosť.

Záver

Univerzálny systém identifikátorov neexistuje a ani nie je možné, objektívne, takýto systém vytvoriť. Je to odraz heterogenity a diverzity informačného prostredia. Je tu však možnosť využiť už existujúce systémy, ktoré sú postavené na akceptovaných štandardoch a poskytnúť integračnú platformu pre novú pridanú hodnotu. Identifikátory DOI a ORCID v súčasnosti dokážu saturovať požiadavky (nielen) akademického prostredia a efektívnej komunikácie. Vytváranie spoločných platforiem na úrovni spomínaných systémov a ďalších subjektov ponúka nový potenciál, napríklad v oblastiach altmetrík, rozšírenia dopadu výsledkov výskumu, dostupnosti výsledkov výskumu, zlepšenia spolupráce vedeckej/akademickej komunity v globálnom rámci a pod.

Použitá literatúra

- COYLE, Karen. 2006. Identifiers: Unique, Persistent, Global. In *The Journal of Academic Librarianship* [online], vol. 32, issue 4, pp. 428-431 [cit. 2018-12-20]. ISSN: 0099-1333. Dostupné na <https://doi.org/10.1016/j.acalib.2006.04.004>
- CROSSREF. 2019. Crossref [online]. [cit. 2019-02-10]. Dostupné na <https://www.crossref.org/>
- CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. 2019. Perzistentné identifikátory digitálnych objektov. In EIZ: Elektronické informačné zdroje [online]. [cit. 2019-02-10]. Dostupné na <http://eiz.cvtisr.sk/doi/perzistentne-identifikatory-digitalnych-objektov/>
- DATA CITE. 2019. *DataCite: Locate, identify, and cite research data with the leading global provider of DOIs for research data* [online]. [cit. 2019-02-10]. Dostupné na <https://datacite.org/>
- DOI FOUNDATION. 2019. *The DOI® System* [online]. [cit. 2019-02-10]. Dostupné na <http://www.doi.org/>
- ORCID. 2019. *ORCID: Connecting research and researchers* [online]. [cit. 2019-02-10]. Dostupné na <https://orcid.org/>
- PASKIN, Norman. 2005. "Digital Object Identifiers for scientific data". In *Data Science Journal*, vol. 4, pp.12-20. ISSN 1683-1470.

PhDr. Ľubica Jedličková, PhD.

lubica.jedlickova@uniag.sk

(Slovenská poľnohospodárska knižnica pri SPU v Nitre)