

Денис СОЛОВ'ЯНЕНКО,

наук. співробітник НБУВ, канд. іст. наук

Цифровий ідентифікатор об'єкта (DOI): «ISBN суспільства знань»

У статті представлено концептуальні засади технологічної моделі цифрового ідентифікатора об'єкта (DOI). Представлено та охарактеризовано складові системи DOI: стандартизований синтаксис (Z39.84), службу диспетчеризації (Handle System), модель даних DOI та механізми реалізації моделі на рівні інформаційної політики. Висвітлено досвід впровадження технологічної моделі DOI співтовариством видавців наукової літератури CrossRef та Національною науково-технічною бібліотекою Німеччини.

Ключові слова: DOI, CrossRef, ідентифікація змісту, диспетчеризація, технології видавничої справи, семантичні технології.

The paper presents an overview of the Digital Object Identifier (DOI) technological model. It is characterized the DOI System main components: a standardized syntax Z39.84, a resolution service based on the Handle System, a DOI data model and an implementation mechanism through specific information policy. The paper also outlines the experience of implementation DOI model by the society of scientific publishers CrossRef and the German National Library of Science and Technology.

Keywords: DOI, CrossRef, identification of content, resolution service, publishing technologies, semantic technologies.

Технологічна модель DOI, про яку йтиметься нижче, виникла на перетині проблематики видавничої справи та кібернетики, вона стосується питань індексації та обліку інформаційних ресурсів. Стаття має на меті проведення огляду системи DOI для аналізу потенційних можливостей українських бібліотек, інформаційних центрів та окремих видавничих установ щодо запровадження даної технологічної моделі у практиці своєї діяльності.

Постановка проблеми

Необхідність запровадження системи єдиних міжнародних стандартних номерів (ISBN) для документів була усвідомлена ще у середині 1960-х рр. Сьогодні індекси ISBN (а також ISSN, ISMN, ISAN тощо) мають фундаментальне значення для видавничої справи та інформаційної галузі взагалі. Проте, епоха онлайн-комунікацій кинула нові виклики стандартизації видавничої справи, зокрема, виникла потреба у розробленні системи реєстрації універсальних стандартних ідентифікаторів для онлайн-інформаційних ресурсів. Стандартизація електронного документообігу за допомогою ISBN-подібних ідентифікаторів пов'язана з низкою проблем, адже електронні ресурси суттєво відрізняються від традиційних як за формальними ознаками, так і за механізмом їх видання. Отже, індексація електронних ресурсів потребує запровадження зовсім іншої технології обліку.

Наявність єдиних міжнародних ідентифікаторів ресурсів вкрай важлива також для сталого розвит-

ку глобальних обчислювальних мереж. Взагалі, поява Інтернет-мережі у її сучасному вигляді пов'язана з винайденням у 1990 р. технології Uniform Resource Locator (URL). Це технологія адресації інтернет-ресурсів, фундаментальна за своїм значенням: рядок URL є унікальним для Мережі загалом набором символів, який дає змогу здійснювати елементарну навігацію у Всесвітньому павутинні. Проте, рядок URL ідентифікує місце збереження ресурсу, але не сам ресурс. З одного боку, ресурси з однаковим змістом можуть знаходитись одночасно за кількома URL-адресами (публікуватися паралельно на декількох веб-сайтах), з іншого – адреси веб-сайтів та їх структура (адреси у межах веб-сайтів) часто змінюються, а отже ідентифікація інформаційного змісту за місцем його збереження (тобто за URL-посиланнями) є дуже ненадійною.

У жовтні 1997 р., в результаті досягнення домовленостей і компромісів між співтовариствами видавців та інтернет-розробниками, з'явилася ініціатива розроблення універсального цифрового ідентифікатора об'єктів Digital Object Identifier (DOI). Процес розроблення технології DOI переслідував конкретну мету – потужним видавцям наукової періодики необхідний був інструмент обліку для роздрібної торгівлі своєю продукцією через Інтернет. Для того, щоб дозволити кінцевому користувачеві в обхід усіх посередників купити окрему статтю, розміщену на сайті видавця, або навіть її частину, потрібно запровадити одиницю обліку та розробити механізм ціноутворення для

кожної конкретної одиниці інформації. Більше того, оскільки така роздрібна торгівля передбачає виставлення на продаж великої кількості одиниць обліку, цінність кожної одиниці повинна визначатись у автоматизованому режимі. Якщо мова йде про продаж науково-інформаційних ресурсів, для визначення цінності кожної одиниці найбільше підходить індекс цитування. Отже, необхідна технологія для автоматизованого визначення індексів цитування ресурсів. Саме такий зміст було закладено початково у технологічну модель DOI. Сьогодні вона розглядається більш широко, оскільки DOI можуть бути зареєстровані не лише для ресурсів, а й навіть для абстрактних наборів знань, тому дана технологічна модель претендує на статус одиниці обліку «економіки знань» у «суспільстві знань».

Де-факто, система DOI є практичним упровадженням специфікацій URI та URN [18], хоча можливості цієї технології виходять за межі означених специфікацій і вирішують низку самостійних важливих проблем інформаційної галузі.

Digital Object Identifier

Система DOI створена для ідентифікації будь-якого змісту у цифровому середовищі (причому, сам зміст не обов'язково має бути цифровим: DOI розшифровується як «цифровий ідентифікатор об'єкта», а не як «ідентифікатор цифрового об'єкта» [12]). Об'єкти реєструються за допомогою унікальних ідентифікаторів, які складаються з цифр та літер, зв'язані з посиланнями на актуальні джерела в Інтернеті, де ці об'єкти (або інформація про них) можуть бути знайдені. Таким чином, інформація про об'єкти може змінюватися з часом, але зареєстровані унікальні ідентифікатори DOI завжди залишаються незмінними [18]. Цю систему часто порівнюють з системою присвоєння універсальних штрих-кодів EAN/UPC товарам [22], якщо штрих-коди ідентифікують об'єкти фізичного світу, система DOI – інформаційні об'єкти. Ідентифікатори DOI можуть бути зареєстровані для будь-яких об'єктів, які можуть бути асоційовані з певною адресою URL [6].

Система DOI як підсистема Handle. Синтаксис DOI

Система DOI є впровадженням глобальної ієрархічної системи просторів мережеских імен Handle System [10]. Розробником системи Handle є Роберт Кан – один з «батьків Інтернету», винахідник сімейства протоколів TCP/IP, лауреат Премії Т'юрінга (аналог Нобелівської премії у

галузі кібернетики). Серед інших упроваджень Handle System можна назвати, зокрема, проект DSpace (пакет програмного забезпечення для створення цифрових репозитаріїв), програму National Digital Library Program (проект Бібліотеки Конгресу США з представлення у цифровому вигляді фондів історичних матеріалів), Defense Technical Information Center (репозитарій науково-технічної документації Міністерства оборони США), сховище даних NASA, бази даних Національної сільськогосподарської бібліотеки США, австралійський урядовий проект PILDIN, низку потужних проектів європейських науково-дослідних установ тощо. Щоб уникнути плутанини, одразу зазначимо: система DOI – це не певний програмний пакет, готовий до впровадження і не абстрактний стандарт іменування об'єктів у базах даних. Система DOI є технологічною моделлю, сукупністю стандартизованих інформаційних технологій, упровадження яких дає змогу побудувати конкретну онлайн-інформаційну систему. Спрощено кажучи, мова йде про єдину загальносвітову глобальну базу даних інформаційних об'єктів, виданих різними установами в різному форматі та доступних у різних місцях.

Характерною рисою системи Handle є унікальність кожного імені в межах системи загалом. Всі ідентифікатори Handle, префікси яких починаються з періоду «10» є ідентифікаторами DOI, що відрізняє їх від ідентифікаторів Handle, зареєстрованих для інформаційних об'єктів у рамках інших проектів (DSpace, NDLP тощо). У свою чергу, всі ідентифікатори DOI також є унікальними в межах системи DOI. Таким чином, кожен ідентифікатор DOI має глобальну унікальність: на рівні системи DOI та, загалом, на рівні системи Handle.

Система DOI має декілька компонентів [3]:

- стандартизований синтаксис побудови ідентифікаторів (Z39.84);
- служба-диспетчеризатор, що базується на Handle System;
- модель словника даних, інтероперабельна з моделями даних е-комерції (база метаданих інформаційних об'єктів);
- механізми впровадження DOI на рівні інформаційної політики.

По суті, DOI-об'єкт – це триплет даних, що складається з унікального ідентифікатора, адреси (або численних адрес) URL та певного набору метаданих. Метадані та URL-адреси об'єкта з часом можуть змінюватись (доповнюватись, коригуватись) за запитом інтелектуального власника, але іден-

тифікатор DOI реєструється для об'єкта одноразово і залишається незмінним протягом усього його «періоду життя». Це дає змогу, зокрема, ідентифікувати власне зміст, а не режим доступу до нього, його адресу або його характеристики (контекст його опису певною схемою метаданих). Звертаючись до DOI-об'єктів, суб'єкти інформаційних відносин взаємодіють через службу-диспетчеризатор (resolution service), яка є джерелом актуальних даних про об'єкт звернення. Таким чином, наприклад, якщо посилання на документ оформлено у вигляді посилання на його ідентифікатор DOI, служба-диспетчеризатор має спрацювати як проксі-сервер і перенаправити кінцевого користувача до актуального місця збереження документа, незалежно від того, хто наразі є його власником і як змінювалися його адреси з моменту оформлення посилання.

За своєю структурою ідентифікатор DOI є рядком даних з цифр та літер, що складається з двох компонентів – *префікса* («Publisher ID») та *суфікса* («Item ID»), – які розподіляються знаком косої риски («/»). Практично не існує обмежень на розмір ідентифікатора DOI (технічний ліміт – приблизно 4 мільярди символів), хоча для надійності та простоти читання людиною пропонується обмеження розміру ідентифікатора 128 символами [22]. На практиці, ідентифікатори DOI – це короткі рядки даних, що містять лише цифри, латинські літери та деякі символи пунктуації.

Префікси для видавців (суб'єктів інтелектуальної власності) реєструються Міжнародною фундацією DOI або однією з її реєстраційних агенцій. Префікси не несуть інформації про поточні права власності на об'єкт, лише вказують на реєстранта, який вніс об'єкт до системи DOI. Деякі суб'єкти інтелектуальної власності мають декілька префіксів для різних типів матеріалів.

Приклади типових префіксів:

10.1016 – Видавничий концерн Elsevier Science;

10.1038 – Видавництво Nature Publishing Group;

10.1535 – Корпорація Intel;

10.2205 – Геофізичний центр Російської академії наук;

10.2298 – Національна бібліотека Сербії.

Суфікси DOI реєструються для об'єктів інтелектуальної власності (книжок та їх розділів, назв журналів, випусків періодичних видань та статей з них, витворів музичного та образотворчого мистецтва, дисертацій, патентів, програмних кодів тощо) реєстрантами, які мають власні префікси (тобто, зареєстрованими постачальниками інфор-

маційного змісту). Стандарт Z39.84-2005 «Синтаксис для DOI» [3] визначає рекомендації та деякі обмеження до структури суфікса DOI, але залишає видавцям свободу вибору оптимальної для них структури суфіксів. Часто у якості суфіксів імплементуються інші ідентифікатори об'єкта (ISBN, ISSN, PII, SICI, «внутрішній» код журналу у видавництві, бібліотечний шифр зберігання тощо).

Приклади ідентифікаторів DOI різних видавців:

doi:10.1038/4541032a – стаття з журналу Nature;

doi:10.1126/science.321.5893.1225b – стенограма подкасту (звукового файлу), опублікованого журналом Science;

doi:10.1016/B978-155860829-0/50015-2 – розділ наукової монографії «Bioinformatics» видавництва Elsevier;

doi:10.1371/journal.pone.0000188.g001 – малюнок зі статті в електронній бібліотеці PLoS One;

doi:10.1002/9780470050118.ecse279 – стаття «Object-Oriented Databases» з енциклопедії «Wiley Encyclopedia of Computer Science and Engineering».

Видавці можуть використовувати різний синтаксис для різних інформаційних об'єктів або переходити з часом до використання більш зручного для них синтаксису. Але важливо, що існуючі ідентифікатори DOI не можуть зазнавати ніяких змін.

Служба-диспетчеризатор DOI

До базової функціональності системи DOI входить функція диспетчеризації інформаційних об'єктів. При реєстрації, з кожним ідентифікатором DOI асоціюється принаймні одна адреса URL, яка вказує на місце знаходження об'єкта. Служба-диспетчеризатор DOI (DOI Resolution Service) – це сервер Handle System, який надає кінцевому користувачеві сервіс, асоційований з певним ідентифікатором DOI; частіше всього, перенаправляє користувача до URL, асоційованого з ідентифікатором DOI. Ця функціональність є повністю відкритою та безкоштовною для кінцевих користувачів Інтернету.

Дуже близькою аналогією є система доменних імен (DNS), на якій базується Всесвітнє павутиння: DNS-сервер отримує від користувача DNS-адресу ресурсу (наприклад, www.nbuv.gov.ua) та перенаправляє його до IP-адреси фізичного сервера, на якому розміщений ресурс (у даному прикладі, 194.44.242.244). Фактично, DNS-адреса – це та ж сама IP-адреса, лише легша для сприйняття людиною. Але крім того, якщо IP-адреса ресурсу змінюється (через зміну телекомунікаційної інфраструктури), наявність рівня абстракції DNS дозволяє зробити запит постачальнику Інтернет на пере-

асоціацію існуючої DNS-адреси з новою IP-адресою і для кінцевого користувача у цьому випадку доменна адреса ресурсу у Всесвітньому павутинні залишиться незмінною. У загальному вигляді диспетчеризація Handle (та DOI, зокрема) виглядає так само, але не на рівні серверів, а на рівні окремих об'єктів інформаційного змісту (сторінок сайтів, статей, наборів даних у таблицях тощо).

Система DOI є впровадженням системи Handle, однією з локальних служб Handle. Наразі ідентифікатори DOI, у більшості впроваджень, диспетчеризуються HTTP-проксі-сервером DOI – <http://dx.doi.org/>. Це DNS-адреса системи DOI. Публічним інтерфейсом системи є проста пошукова форма центральної бази даних DOI-об'єктів (Central DOI Directory) – інструмент для визначення об'єктів змісту за їх ідентифікаторами DOI. HTTP-адреси, які складаються з адреси проксі-сервера DOI та конкретного ідентифікатора, диспетчеризуються автоматично. Умовно кажучи, ідентифікатори DOI – це постійні адреси інформаційних об'єктів на сайті <http://dx.doi.org/>, незалежні від фактичного місця знаходження об'єктів (хоча технічно таке визначення не зовсім правильне, воно дане для спрощення сприйняття моделі). Таким чином, у веб-середовищі для забезпечення можливості оперативної диспетчеризації інформаційних об'єктів системою DOI, гіперпосилання оформлюються у вигляді <http://dx.doi.org/10.1234/5678>, де 10.1234/5678 – ідентифікатор конкретного об'єкта. У найпростіших упровадженнях таке посилання має перенаправити користувача на веб-сайт, де розміщений об'єкт; деякі системи пропонують більшу функціональність (наприклад, перенаправлення до сторінки з цілим меню у вигляді списків посилань до різних копій об'єкта, споріднених об'єктів тощо). Звичайно, диспетчеризація DOI не передбачає, що кінцевий користувач обов'язково отримає доступ до самого ресурсу (повного тексту, аудіофайла, відеофрагмента тощо), але вона гарантує, що користувач обов'язково отримає вторинну інформацію про інформаційний об'єкт, якщо він був зареєстрований у системі DOI. Тобто, якщо доступ до об'єкта є вільним та безкоштовним, користувач, за ідентифікатором DOI з будь-якого комп'ютера в Інтернеті отримає перенаправлення до самого об'єкта (наприклад, до повного тексту), якщо ж об'єкт доступний лише на платній основі і користувач не є його передплатником, він буде перенаправлений на сторінку придбання товару, до анотації об'єкта або до іншої сторінки, яку вказав у якості URL

інтелектуальний власник при реєстрації об'єкта у системі DOI.

Офіційні реєстраційні агенції DOI (та деякі окремі постачальники первинного змісту) підтримують для локальних реєстрів DOI власні локальні проксі-сервери Handle/DOI: <http://doi.wiley.com/>, <http://dx.crossref.org/>, <http://doi.ieeeecs.org/>, <http://doi.tib.uni-hannover.de:8000/>, <http://dx.medra.org/> тощо.

Система DOI прив'язана до Інтернет-протоколу TCP/IP, але вона не прив'язана до протоколу HTTP, отже диспетчеризація DOI можлива поза Всесвітнім павутинням, без посередництва веб-проксі-серверів. Передбачається, що кореневий протокол диспетчеризації DOI буде прийнятий найближчим часом у якості одного зі стандартних комунікаційних протоколів Інтернету. Це означає, зокрема, що додавання імені проксі-сервера до ідентифікатора DOI стане не потрібним; кінцевий користувач зможе отримати доступ до об'єкта безпосередньо за протоколом doi, ввівши адресу вигляду doi:10.1234/5678 до адресного рядка Інтернет-браузера, так само, як сьогодні він вводить веб-адресу <http://www.nbuv.gov.ua> або адресу файл-сервера <ftp://example.org>. Але наразі DOI ще не затверджений у якості Інтернет-стандарту* і виробники програмного забезпечення не включають протокол doi до базової комплектації своїх продуктів. Проте, підтримка корневих протоколів диспетчеризації DOI та Handle (без посередництва проксі-серверів) можлива шляхом встановлення додаткових компонентів (plug-in) до Інтернет-браузерів та інших програмних пакетів. Розроблені та безкоштовно доступні програмні модулі для основних веб-браузерів операційних систем Windows та MacOS. Крім того, доступний модуль для програми Adobe Acrobat Reader, який дає змогу проводити розширену Handle/DOI диспетчеризацію pdf-файлів: перевіряти наявність нових версій змісту, встановлювати поточного правовласника тощо [13]. Синтаксис DOI напряму також розуміє низка потужних Інтернет-систем (наприклад, Вікіпедія, CiteULike, Connotea тощо).

Відображення та використання DOI у електронних та друкованих виданнях

Найширшого запровадження система DOI набула у практиці науково-видавничої діяльності. Відповідно, саме у видавничому середовищі почали формуватися традиції щодо форми представ-

* Детальніше про це див. нижче, у частині «DOI як міжнародний стандарт».

лення DOI та його використання кінцевими користувачами інформації.

Система DOI має великі можливості застосування на всіх етапах інформаційної діяльності, зокрема, при запровадженні автоматизованих технологій обліку, обробки, пошуку та поширення інформації. Але для кінцевого користувача DOI – це, насамперед, постійне надійне посилання на інформаційне першоджерело, короткий ідентифікатор, що дає змогу у будь-який час із будь-якого комп'ютера, підключеного до Інтернету, отримати доступ до повного тексту документа або, принаймні, до основних даних про нього. Тому форма представлення DOI у документах має велике значення для користувачів.

Рекомендацією Міжнародної фундації DOI є подання ідентифікатора DOI, разом з іншою бібліографічною інформацією, у заголовку будь-якого фрагмента змісту (статті, розділу книги, зображення тощо) як у електронному, так і у друкованому вигляді. Закріплений також формат представлення DOI: єдиний рядок без пропусків, що складається з позначення «doi:» і власне ідентифікатора (наприклад, doi:10.1038/453979b). Написання позначення «doi:» малими літерами та відсутність пропусків вимагається специфікацією URI (так само, як http:// або ftp://). У електронному вигляді, якщо це можливо, всі згадування DOI оформляються як активні гіперпосилання вигляду <http://dx.doi.org/10.1234/5678>. Оскільки у системі DOI реєструється інформаційний зміст, а не видання у конкретному форматі, друковані та електронні версії видань мають ідентичні ідентифікатори DOI. Отже, якщо DOI зареєстрований для електронної версії видання, у заголовку його друкованої версії він також може (повинен) зазначатися. Формат представлення єдиний для електронних та друкованих версій, але у друкованих виданнях деякі видавці роблять додаткову примітку «Доступно через <http://dx.doi.org/>» для того, щоб кінцеві користувачі розуміли, де саме вони можуть ввести ідентифікатор ресурсу та отримати доступ до його електронної версії. Поки doi: не є стандартним інтернет-протоколом таке уточнення є корисним.

Така форма представлення є рекомендацією, а не вимогою. Більшість видавців наукової періодики, особливо видавничі корпорації, такі як Elsevier та Springer, публікують DOI як у електронному, так і у друкованому вигляді або у заголовку, разом з назвою та іменем автора, або у колонтитулі сторінки, разом з ISBN/ISSN. Часто DOI також вказуються на сторінках змісту видань. Майже всі потужні ре-

феративні та повнотекстові бази даних наукової періодики подають DOI разом з іншою бібліографічною інформацією статей, а потужні пошукові платформи, наприклад SCOPUS, SpringerLink, Wiley InterScience, надають також можливості пошуку ресурсів за їх ідентифікаторами DOI. Проте, дуже багато видавців, які використовують технологічну модель DOI, в явному вигляді не публікують ідентифікаторів DOI у друкованих, а іноді й в електронних, версіях видань.

Звичайно, поки DOI не отримав статусу міжнародного стандарту, ніхто не може примусити науковців використовувати DOI у своїх працях. Хоча для видавців наукових видань модель DOI є дієвим інструментом запровадження автоматизованого підрахунку індексів цитованості ресурсів. Тому, якщо вчені не включають ідентифікатори DOI до бібліографічних записів у пристатейній бібліографії власних праць, то цю роботу повинні виконувати самі редактори. Деякі видавці, зокрема Nature Publishing Group, Oxford, Springer та ін., включають інформацію про використання DOI у власні «Інструкції для авторів», вимагаючи подання ідентифікаторів DOI при оформленні посилань на так звані OnlineFirst видання – джерела, які вже доступні в електронному вигляді, але ще не вийшли друком (а отже, не мають номерів сторінок). Рекомендують (але не вимагають) використовувати DOI також деякі російські видавці академічних (особливо технічних) часописів.

Формат представлення DOI у бібліографічних описах наразі не регулюється відповідними стандартами. Тому видавці в цьому питанні керуються рекомендаціями Міжнародної фундації DOI та власним досвідом. Спільної традиції подання DOI у пристатейній бібліографії ще не склалося: в авторських рукописах ідентифікатори в основному подаються останнім полем бібліографічних записів, після коми/крапки або у дужках.

Модель даних DOI. Онтологія IDD

Як зазначалося вище, система DOI є одним із впроваджень системи Handle. Але технологічна модель DOI не обмежена можливостями ідентифікації та диспетчеризації, закладеними у Handle. Компонентами цієї моделі є також спільна глобальна модель даних та спільна інформаційна політика.

Реєстр Handle – це база даних пар «ключ – значення», де ключами завжди є унікальні для системи загалом ідентифікатори, а значеннями – асоційовані з ними URL (або e-mail, паролі, сертифікати тощо).

Але справжню потужність має впровадження, в якому разом з унікальними ідентифікаторами зберігаються описові дані про об'єкти, що ідентифікуються (комплекти метаданих), завдяки яким стає можливим контекстне звернення до інформаційних об'єктів. У системі DOI використовується структура записів «ключ – значення – метадані» (doi – URL – Kernel). Оскільки система DOI не обмежується певним інформаційним сегментом, зрозуміло, що йдеться про прийняття певної спільної глобальної моделі даних.

Різні інформаційні об'єкти характеризуються абсолютно по-різному, вони мають різні аспекти змісту, важливі у різних контекстах. Тому, для забезпечення подальшого багатоаспектного використання, важливо зберегти максимальний обсяг даних про об'єкти (забезпечити максимально можливу кількість контекстів звернення). Але прийняття спільного формату метаданих у межах інформаційного середовища загалом є ускладненим. Семантичні технології пропонують новаторський підхід до проблематики опису інформаційних об'єктів: не стандартизація певних схем метаданих, а навпаки, заохочення до створення власних схем, максимально зручних для опису і пошуку конкретних інформаційних об'єктів у конкретному контексті. При цьому жорсткою вимогою є забезпечення семантичної інтероперабельності локальних схем, тобто прийняття спільної онтології – певного загального словника даних та логічної моделі відстеження семантичних зв'язків між даними у цьому словнику. Така онтологія має забезпечити «семантичну інтероперабельність галузевих знань», тобто можливість перенесення знань з однієї галузі до іншої (з однієї інформаційної системи до іншої) без втрати змісту. При цьому вкрай важливою є наявність певних «глобальних точок прив'язки» (елементів, які на глобальному рівні однаково інтерпретуються всіма учасниками інформаційної взаємодії).

Саме такий підхід застосовується у системі DOI. Реєстранти (постачальники первинного інформаційного змісту) застосовують для опису ресурсів формати метаданих, зручні саме у їх діяльності. Але при цьому, на рівні реєстраційних агенцій DOI, застосовуються моделі даних реєстраційних агенцій, нейтральні до різних інформаційних контекстів у межах профілю впровадження кожної агенції. На рівні системи загалом застосовується єдина модель даних DOI, нейтральна до всіх аспектів інформації. Наприклад, видавництво Nature Publishing Group та Національна бібліотека Сербії

реєструють власні науково-інформаційні ресурси через реєстраційну агенцію CrossRef. Цілоком природно, що національна бібліотека і комерційне видавництво використовують різні формати опису ресурсів. Але до бази даних CrossRef вони переносять з локальних систем не всі метадані, а лише визначені відповідною схемою метаданих CrossRef. Схеми CrossRef розроблені так, щоб гарантувати відсутність колізій у значеннях термінів у межах всього науково-видавничого співтовариства. У свою чергу, агенція CrossRef реєструє у Central DOI Directory не всі метадані, отримані від видавців, а лише мінімальний (стрижневий) набір полів, який нейтрально розуміється не лише науковим співтовариством, але й видавцями офіційних матеріалів, освітніми установами тощо.

Такий мінімальний набір називається DOI Kernel (повна назва: The DOI Resource Metadata Kernel Declaration). Набір метаданих Kernel з'явився в рамках ініціативи Dublin Core. Він призначений для максимального спрощення процедури опису ресурсів, пропонуючи «стрижень» даних про ресурс у вигляді відповідей на чотири базові питання: хто, що, де і коли. Набір DOI Kernel є одним із впроваджень Kernel, він містить 4 базові поля та 7 підполів: реєстраційна агенція DOI, DOI-об'єкт (з подальшою деталізацією), дата випуску DOI та номер випуску DOI (перший випуск, перевипуск).

Таким чином, у межах системи DOI будується глобальний «стрижень» надпотужної розподіленої бази метаданих (по суті, бази знань), зв'язаної єдиним синтаксично та семантично інтероперабельним «жорстким» ключем. Повна синтаксична інтероперабельність гарантується на рівні ідентифікатора DOI, що є певною стійкою послідовністю цифр та літер, яка гарантовано однаково інтерпретується всіма учасниками інформаційної взаємодії. Семантична інтероперабельність забезпечується жорсткою прив'язкою ідентифікатора до асоціації. Тобто на глобальному рівні існує гарантія, що користуючись ключем, всі учасники взаємодії вкладають у нього саме той зміст, який презентується URL, асоційованим з ним. При цьому сам ключ позбавлений жодних контекстів, будь-який учасник взаємодії може додавати до нього власний контекст (власні метадані) або використовувати інші вже асоційовані з ключем контекстні дані (метадані).

Все це справедливо не лише для системи DOI, але для Handle загалом. Такий саме «жорсткий» ключ має, наприклад, мережа архівів відкритого

доступу DSpace. Але, як зазначалося вище, для орієнтації у таких глобальних розподілених репозитаріях потрібна онтологія – «карта» взаємозв'язків інформаційних контекстів.

Онтологія системи DOI називається IDF's Data Dictionary (IDD), вона розробляється спільно членами Міжнародної фундації DOI (див. далі, у розділі «Інфраструктура DOI. Фундація IDF»). Реєстраційні агенції розробляють свої онтології для власних цілей, при цьому IDD є своєрідною «есенцією» розробок різних агенцій і забезпечує глобальну інтероперабельність різних впроваджень системи DOI. Принцип зрозумілий: коли велика кількість потужних інформаційних установ займається розробленням власних онтологій, стає важко підтримувати глобальну інтероперабельність на горизонтальному рівні, контролюючи двосторонню інтероперабельність з кожною онтологією; виникає необхідність у кореневій онтології. Підтримуючи інтероперабельність з однією онтологією (IDD), реєстраційні агенції DOI, таким чином, підтримують двосторонню горизонтальну інтероперабельність з онтологіями усіх інших реєстраційних агенцій DOI.

Процедури розроблення та використання IDD і обміну метаданими між реєстраційними агенціями регулюються політикою метаданих системи DOI. Принципи цієї політики такі:

- реєстраційні агенції вільні у виборі схем для збору, збереження, використання та поширення метаданих;
- кожен об'єкт DOI має супроводжуватися мінімальним набором метаданих DOI Kernel;
- реєстраційні агенції мають обмінюватися метаданими лише у форматі DOI Resource Metadata Declaration (RMD), що є частиною IDD;
- всі терміни (назви елементів метаданих та їх значення), які використовуються реєстраційними агенціями у DOI Kernel та RMD, повинні бути зареєстрованими у IDD. Таким чином, якщо реєстраційні агенції хочуть обмінюватися метаданими, неузгодженими з IDD, вони повинні розробити таке RMD-узгодження;
- додаючи власні терміни до IDD, реєстраційні агенції можуть визначати права доступу до цих термінів, але публічні терміни, зокрема DOI Kernel, RMD, MPEG-21 RDD тощо, повинні бути доступними всім реєстраційним агенціям.

Інфраструктура DOI. Фундація IDF

Адміністрування системи DOI за своєю інфраструктурою є доволі схожим на адміністрування сис-

теми ISBN-номерів, але має також низку істотно інших ознак.

Централізоване адміністрування системи DOI здійснюється Міжнародною фундацією DOI (The International DOI Foundation, IDF), яка була заснована у 1998 р. трьома провідними світовими асоціаціями видавців (Association of American Publishers, International Publishers Association та International Association of Scientific Technical and Medical Publishers), як неприбуткова міжнародна організація [22]. Керує IDF виконавча колегія, яка обирається членами Фундації. Вона опікується питаннями розгортання єдиної світової політики функціонування моделі DOI, формулювання стандартів та рекомендацій для видавців, реєстраційних агенцій по всьому світу, забезпечення технологічної сумісності та інтеграції зусиль з організаціями, котрі займаються розробленням паралельних інформаційних технологій.

Згідно зі статутом, членство у IDF можуть мати будь-які прибуткові чи неприбуткові організації інформаційного профілю (видавництва, звукозаписувальні компанії, бібліотеки та бібліотечні асоціації, архіви, музеї, агенції з авторського права, передплатні консорціуми тощо). Фундація працює за неприбутковим принципом, але при цьому членство та активне використання системи DOI є платним. Існує чотири основні класи членства у IDF:

- Базове членство дає змогу організації брати безпосередню участь у розробленні засад функціонування системи DOI. Згідно з поточною політикою IDF, цей клас членства є перехідним у процесі надання організації статусу офіційної реєстраційної агенції. Станом на 2009 р. IDF налічує 8 базових членів:
 - CENL (Conference of European National Librarians);
 - Copyright Clearance Center;
 - EDP Sciences;
 - Joint Information Systems Committee (JISC);
 - NamesforLife LLC;
 - New England Journal of Medicine;
 - Publishers Licensing Society;
 - Wolters Kluwer International Health & Science.
- Привілейоване членство спочатку мали організації-засновники IDF. Наразі привілейованими членами IDF є три провідні світові видавничі корпорації: Elsevier, John Wiley & Sons та Springer SBM.
- Асоційоване членство доступне для асоціацій та професійних об'єднань, до яких входять організації-члени IDF. Станом на 2009 р. у IDF –

два асоційованих члени: International Federation of Reproduction Rights Organisations (IFRRO) та STM International Association.

- Членство у якості реєстраційної агенції, як і базове членство, передбачає особисту участь організації або об'єднання у розробленні політики IDF, водночас воно додатково надає члену низку особливих прав та накладає на нього додаткові обов'язки. Головна функція реєстраційної агенції – координація впровадження та використання системи DOI постачальниками первинного інформаційного змісту: створення, підтримка необхідної інфраструктури, розподіл префіксів DOI між видавцями та іншими постачальниками інформації, забезпечення можливостей реєстрації об'єктів DOI, підтримка Central DOI Directory, а також моніторинг баз даних первинної інформації своїх реєстрантів, збір та синхронізація наборів метаданих, з якими асоційовані DOI. Кожна реєстраційна агенція має власну «спеціалізацію» (профіль застосування DOI), працюючи у власній інформаційній ніші (за типами інформаційної продукції, географічними та іншими ознаками) [22]. Хоча ця «спеціалізація» часто є доволі умовною. Кожна агенція вільна обирати оптимальну для неї бізнес-модель роботи зі своїми реєстрантами. Станом на 2009 р. IDF має 8 офіційних реєстраційних агенцій:
- австралійська національна агенція Copyright Agency Limited;
- спілтовариство видавців наукової періодики CrossRef;
- mEDRA – багатомовна європейська агенція (Італія);
- британська компанія Nielsen BookData;
- OROCE (Офіс офіційних публікацій Європейських Спілтовариств);
- американська компанія R. R. Bowker;
- Національна науково-технічна бібліотека Німеччини (ТІВ);
- китайська компанія Wanfang Data Co., Ltd.

Розглянемо окремо діяльність двох агенцій, що спеціалізуються, відповідно, на реєстрації DOI для вторинних та первинних наукових даних.

Реєстраційна агенція CrossRef

CrossRef – офіційна реєстраційна агенція DOI, заснована у 2000 р. провідними світовими видавцями наукової періодики для створення єдиної потужної служби взаємних цитувань в академічних науко-

вих часописах [7]. За формою організації CrossRef є спілтовариством видавців. Спеціалізація агенції – науково-дослідні матеріали (журнальні статті, монографії, дисертації тощо), а основні зусилля направлені на створення глобальної бази даних вторинної наукової інформації та взаємних цитувань. Обмежень за географічною ознакою CrossRef не має, її членами є організації з усіх континентів. Станом на березень 2009 р. членами CrossRef є 2730 видавців наукової літератури, які реєструють через агенцію DOI для 20 тис. назв наукових видань [6]. Загальна кількість DOI, зареєстрованих через CrossRef, сягає 35,6 млн. На сьогодні це найпотужніше впровадження технологічної моделі DOI. У раді директорів CrossRef представлені такі видавці, як AAAS (Science), AIP, ACM, APA, Elsevier, IEEE, IOP, Informa UK, Wolters Kluwer Health, Nature, Sage, Springer, Thieme, University of California Press, University of Chicago Press та Wiley-Blackwell. Потужні постачальники наукової інформації мають членство у CrossRef додатково до прямого членства у IDF. Це дає їм змогу, з одного боку, справляти вирішальний вплив на стратегію діяльності Фундації DOI, а з другого – мати безпосередню користь від спеціалізованих рішень, виключно для сегмента наукової інформації, зокрема, від служби взаємних цитувань. До складу консультативної ради CrossRef, яка робить незалежні оцінки щодо перспективних напрямів подальшого розвитку агенції, входять представники найпотужніших інформаційно-технологічних установ світу: CNI, Drexel University, Harvard University, Los Alamos National Laboratory, Microsoft, MIT, OCLC, Protein Data Bank, Research Information Network UK та SirsiDynix.

Незважаючи на те, що найбільшої популярності використання системи DOI набуло в сегменті наукової періодики, останніми роками CrossRef докладно зусиль для популяризації використання цієї технологічної моделі під час публікації монографій, довідково-енциклопедичних матеріалів та інших неперіодичних наукових видань. Станом на вересень 2008 р. понад 40 членів CrossRef використовували DOI для книжок та їх складових частин, а загальна кількість DOI, зареєстрованих для неперіодичних наукових видань становить 1,4 млн, у т. ч. Oxford Dictionary of National Biography має окремі DOI для кожної з 66 тис. словникових статей [6]. Іншою потужною ініціативою CrossRef є використання системи DOI для повернення в науковий обіг архівів провідних наукових видань. Так, на початку 2007 р. Лондонське ко-

ролівське товариство розпочало проект реєстрації у системі DOI історичних архівів власних видань, починаючи з 1665 р. Ініціатива була підтримана видавництвами Elsevier, Springer, SAGE, Wolters Kluwer, Blackwell, John Wiley & Sons та AAAS, які також розпочали проекти ретро-оцифрування та реєстрації власних архівів з використанням DOI [19]. Хоча «спеціалізацією» агенції є наукові матеріали, вона також реалізує проекти у деяких суміжних галузях. Так, у жовтні 2007 р. CrossRef розпочала спільний проект з британською національною організацією зі стандартизації. У рамках цього проекту було створено робочу групу з проблематики реєстрації стандартів у системі DOI [5]. Оскільки стандарти традиційно є високоцитованими документами, то передбачається, що ініціатива їх реєстрації за допомогою DOI та подальшого відслідковування їх цитованості матиме великі перспективи.

Хоча агенція працює за безприбутковим принципом, для видавців членство у CrossRef платне. Річний членський внесок становить від \$275 (для організацій з річним доходом менше \$1 млн) до \$50.000 (для організацій з доходом понад \$500 млн). При цьому CrossRef дозволяє своїм членам реєструвати DOI для таких типів інформаційних об'єктів:

- журнали та доповіді: назва, том, випуск, стаття;
- книжки: книжкова серія, назва, розділ/частина/стаття;
- матеріали конференцій: серійне видання, назва, праця;
- компоненти: включення журнальної статті, додаток до розділу книжки/статті/довповіді;
- малюнки, таблиці, графіки та супровідні дані;
- дисертації;
- стандарти;
- набори даних та елементи даних.

Процеси реєстрації та подальшої підтримки DOI-об'єктів для видавців передбачають вивантаження XML-файлів на сервер CrossRef. Потужні видавці та видавничі корпорації автоматизують для себе ці процеси, використовуючи веб-сервіси та інші складні програмні рішення; для невеликих видавничих структур на сайті CrossRef доступний веб-орієнтований інструментарій для адміністрування DOI. Реєстрація ідентифікаторів DOI для назв журналів безкоштовна для членів CrossRef, за реєстрацію інших типів інформаційних об'єктів вони повинні додатково сплачувати: \$0,06 – за кожний DOI, зареєстрований для набору даних, \$0,15 – за кожний DOI, зареєстрований для архівних об'єктів

будь-якого типу, \$0,17 – за DOI для книжки або її частини двох останніх років видання, \$1,00 – за кожний DOI для інших інформаційних об'єктів двох останніх років видання: випуск журналу, стаття, доповідь, дисертація тощо. Видавці – члени CrossRef зазнають додаткових витрат, якщо вони вступають у конфлікт з політикою реєстраційної агенції. Типовим прикладом такого конфлікту може бути реєстрація двох або більше DOI для об'єктів з фактично тотожним інформаційним змістом [6].

Членство у CrossRef дає змогу використовувати сервіс Cited-by Linking. Даний сервіс дозволяє видавцям обмінюватись інформацією щодо цитованості інформаційних об'єктів, використовуючи метадані з пристатейної бібліографії видань. Наведена послуга добровільна і безкоштовна, вона стає можливою на основі взаємної заінтересованості. При цьому від видавців вимагається подання мінімального обсягу метаданих щодо джерел, які цитуються в їх виданнях. Такий набір метаданих має містити ідентифікатори DOI цитованих джерел, що дає змогу точно встановити кількість посилань на певний DOI. Отже видавці отримують доступ до бази даних взаємної цитованості і можуть упроваджувати ці дані на власних видавничих платформах. Наприклад, Science, IOP Publishing, низка американських наукових товариств, інші видавці у власні онлайн-статті включають посилання «Статті, що цитують дану статтю». Ця функція забезпечується через Cited-by Linking. Станом на 2008 р. ініціативу підтримали 126 видавців, у т. ч. майже всі найбільші видавці наукової періодики (крім Elsevier). Хоча інші продукти, які дозволяють встановлювати глобальні показники цитованості наукових публікацій, також глибоко інтегровані з системою DOI [20], сервіс Cited-by Linking суттєво відрізняється від них за низкою ознак. Так, Cited-by Linking, сам по собі, не є базою даних, до якої може звернутися кінцевий користувач інформації, оскільки цей сервіс не має «публічного» інтерфейса [6]. Тобто, для того, щоб дозволити користувачам отримувати дані щодо цитованості публікації, видавець повинен упровадити на власній платформі API для взаємодії з віддаленою базою даних. Водночас дані про цитованість Cited-by Linking є конфіденційними, тому видавці можуть отримувати дані щодо цитованості лише тих публікацій, які вони самі реєструють у системі DOI. Отже, Cited-by Linking – це наукометричний інструмент, але не наукометрична база даних. Ще однією відмінністю від відомих баз даних ци-

тованості є те, що даний сервіс надає комплексні показники цитованості незалежно від дисципліни та типу ресурсу, в якому джерело було процитовано (враховуються всі цитати, у т. ч. міждисциплінарні: у статтях, монографіях, дисертаціях, презентаціях тощо). Важливим є також те, що сервіс Cited-by Linking пропонує тип взаємодії «видавець – видавець»: CrossRef виступає лише організаційним, юридичним та технічним посередником у цьому процесі, що має як плюси, так і мінуси. З одного боку, засобами цієї ініціативи видавці фактично рекламують ресурси один одного, оскільки дані про цитованість ресурсу «вбудовуються» безпосередньо до повного тексту першоджерела і стають доступними будь-якому кінцевому читачеві процитованого інформаційного ресурсу. З другого, Cited-by Linking надає інформацію про цитованість лише з тих джерел, які постачаються учасниками проекту. Наприклад, видавництво Elsevier, як зазначалося вище, не приєдналося до цієї ініціативи, тому видавці засобами Cited-by Linking не можуть отримати даних щодо цитованості їхніх ресурсів у виданнях Elsevier. Таким чином, дані щодо цитованості ресурсів, отримані через сервіс Cited-by Linking, можуть бути неповними. Даним щодо цитованості, отриманим зі SCOPUS чи ISI Web of Knowledge, також може бракувати повноти, оскільки ці бази даних проводять вибіркове індексування видань на основі попереднього відбору. Cited-by Linking не передбачає попереднього відбору, враховуються всі цитати в усіх джерелах, які надаються учасниками ініціативи.

Іншим цікавим сервісом, доступним членам CrossRef, є система пошуку плагіату в науковій літературі CrossCheck. Вона була запропонована у червні 2008 р. Це спільний продукт агенції та спеціалізованої компанії iParadigms LLC, яка займається розробленням високотехнологічних рішень у сфері інтелектуальної обробки текстової інформації та захисту авторських прав. Система CrossCheck складається з двох компонентів: бази даних проіндексованих першоджерел та пакета програмних засобів, кожен з яких має власні алгоритми знаходження плагіату у наукових текстах. У найпростішому вигляді, система порівнює проіндексовані повні тексти та формує для першоджерел «звіти про оригінальність», у яких фрагменти тексту, запідозрені у наявності плагіату, виділяються червоним кольором з посиланнями на DOI інших документів, де системою було знайдено схожі фрагменти тексту, але на які

відсутні оформлені посилання у фрагменті, що підозрюється. Для проведення подальшої ручної експертизи надається також доступ до повних текстів інших документів, що містять схожі фрагменти. У вересні 2008 р. CrossCheck удостоїлася премії ALPSP у номінації «Найкраща інновація 2008 р. у видавничій справі». Подальший розвиток сервісу спрямований на його інтеграцію з системами електронної доставки авторських рукописів, що дасть змогу видавцям мати «звіти про оригінальність» вже на етапі отримання рукописів від авторів. Станом на листопад 2008 р. 31 видавець приєднався до ініціативи та надав свій повнотекстовий зміст для індексації у базі даних CrossCheck. В основному це потужні видавничі корпорації (Elsevier, Springer, Taylor & Francis, Wiley-Blackwell, Science тощо), для яких даний проект є перш за все іміджевим.

Ще однією додатковою послугою для членів агенції є передплата доступу до репозитарію метаданих CrossRef (CMS). Взагалі, пошук інформації у цьому репозитарії на сайті CrossRef є безкоштовним для всіх бажаючих, і не вимагає навіть ніякої реєстрації, але пошуковий інструментарій сайту вкрай обмежений, зокрема, форма містить мінімальну кількість пошукових полів, а у результатах пошуку пропонується лише основна бібліографічна інформація. Це безкоштовний інструмент для встановлення ідентифікаторів DOI, але не пошукова система. Проте, CrossRef дозволяє своїм членам за додаткову платню використовувати CMS у повному обсязі для створення власних інформаційних систем. CMS є репозитарієм метаданих обсягом понад 35 млн записів, повністю сумісним з форматом обміну метаданими OAI-PMH. Типовими впровадженнями, заснованими на використанні CMS, є пошукова система Scirus компанії Elsevier, цифрова бібліотека NASA Astrophysics Data System, база даних EMBL Європейської лабораторії молекулярної біології, проект MESUR щодо побудови семантичної моделі процесу наукової комунікації, деякі інші проекти.

Крім видавців, членство у CrossRef можуть мати також бібліотеки, хоча це інша категорія членства. Станом на 2009 р. членами CrossRef є півтори тисячі бібліотек з усього світу. Згідно з політикою CrossRef, для більшості бібліотек членство є безкоштовним, платня стягується лише з потужних бібліотек, які дуже активно використовують сервер CrossRef. Завдяки членству бібліотеки мають можливість суттєво покращувати та збагачувати електронні каталоги і картотеки за рахунок

комбінування диспетчеризації DOI та OpenURL. Набуваючи членства у CrossRef, бібліотека має змогу обслуговувати інформаційні запити через сервер DOI та використовувати отримані від сервера метадані у власних інформаційно-пошукових та навігаційних системах. Однак тут слід сказати, що перспективи впровадження DOI для бібліотек суттєво залежать від програмного забезпечення, яке вони використовують. Для західних бібліотек використання функціональності OpenURL вже давно стало звичним, а у країнах СНД робляться лише перші кроки у цьому напрямі [1].

Реєстраційна агенція ТІВ

Національна науково-технічна бібліотека Німеччини (Technische Informationsbibliothek) у 2003 р. за фінансової підтримки Німецького науково-дослідного товариства (DFG) розпочала пілотний проект «Публікація та цитування первинних наукових даних», суть якого полягає у запровадженні моделі DOI для створення баз первинних наукових даних [11]. Цей проект вкрай важливий, адже фактично йдеться про введення первинних даних у науковий обіг та репрезентацію цих наборів даних у бібліотечних каталогах. Тобто дослідники отримують можливість посилатися у своїх роботах безпосередньо на дані обсерваторій, дослідних станцій, космічних супутників тощо, а не на публікації колег науковців. Передбачється також запровадження індексу цитування первинних наукових даних.

Першим етапом проекту стало впровадження моделі DOI для публікації і цитування наборів даних у галузі наук про Землю. Учасниками ініціативи стали Інститут метеорології Макса Планка (світовий центр даних про клімат, Гамбург), Університет Бремена та Інститут Альфреда Вегенера (світовий центр даних про екологію морського середовища), Німецький авіакосмічний центр (світовий центр даних дистанційного зондування атмосфери, Оберфаферховен) та Німецький науково-дослідний центр наук про Землю (Потсдам). Ці організації діють як агенти ТІВ і координують впровадження технологічної моделі DOI у профільних для них наукових напрямках. Для реалізації корпоративної взаємодії учасників проекту було реалізовано програмно-обчислювальний комплекс з розподіленою архітектурою на базі використання технології веб-сервісів [16]. Хоча ТІВ позиціонується як некомерційна агенція, витрати на реєстрацію та підтримку DOI частково перекладаються на виробників первинного змісту, агенти сплачують бібліотеці щорічну платню у розмірі

€250 [8]. Крім того, кошти на виконання проекту закладаються у бюджетне фінансування ТІВ у рамках виконання нею національної місії.

ТІВ реалізує технологічну модель DOI у різних інформаційних сегментах, зокрема підтримуючи партнерські відносини з низкою наукових установ поза межами Німеччини. Так, у рамках проекту eBank спільно з UKOLN агенція реєструє у системі DOI для кристалічних структур; Європейська академія алергології та клінічної імунології, Європейське товариство радіології реєструють через агенцію дані клінічних випадків; у рамках співпраці з Європейською асоціацією комп'ютерної графіки (Eurographics) до системи вводиться «сіра література» (неопубліковані, але важливі матеріали). Реєстрацію у системі DOI проходять дисертації CERN та звіти про результати науково-дослідних робіт, фінансованих німецьким урядом. Крім того, агенція ТІВ бере участь у проекті ELEONET зі створення єдиного загальноєвропейського е-навчального середовища.

Глобальна мета ТІВ – перетворення системи бібліотечних каталогів і картотек TIBORDER з джерела вторинної інформації про ресурси, доступні в бібліотеці, на портал для навігації по всьому спектру джерел достовірної науково-технічної інформації з забезпеченням можливостей оперативного встановлення місця знаходження та режиму доступу до будь-якого об'єкта наукового знання за принципом «єдиного вікна».

На середину 2008 р. ТІВ та її агентами було зареєстровано 550 тис. комплектів первинних наукових даних, 12 тис. наукових відеофрагментів, 6,5 тис. клінічних випадків, сотні звітів, наукових статей, навчальних об'єктів та інших матеріалів [4]. Наразі фахівці ТІВ виступають за створення агенції DOI всесвітнього масштабу, якій бібліотека могла б передати повноваження щодо координації впровадження технологічної моделі DOI для первинних наукових даних. Форматом такої агенції ТІВ бачить всесвітню асоціацію наукових бібліотек та галузевих інформаційних центрів, здатних на локальному рівні координувати функціонування системи DOI та співпрацювати на засадах рівноправного партнерства. Станом на середину 2008 р. в рамках створення такої агенції ТІВ підписані дві угоди: з ETH Zurich Library (найбільша бібліотека та центр науково-технічної інформації Швейцарії) та з L'Institut de l'Information Scientifique et Technique (Інститут науково-технічної інформації Франції) [4]. Розгляда-

ються перспективи та статус членства Міжнародної ради з науково-технічної інформації (ICSTI).

DOI як міжнародний стандарт

Технологічна модель DOI дуже складна і багатогаспектна, тому стандартизація даної моделі потребує прийняття цілої низки стандартів стосовно різних аспектів функціонування DOI.

Процес прийняття синтаксису DOI у якості стандарту розпочався у другій половині 1990-х рр., із упровадженням даної моделі у практику науково-видавничої діяльності. У 2000 р. NISO затвердила синтаксис DOI як національний стандарт США – Z39.84-2000 «Syntax for the Digital Object Identifier». Тоді ж IDF стала членом Американської організації з інформаційних стандартів. У 2005 р. прийнято оновлену версію Z39.84, який є чинним наразі: ANSI/NISO Z39.84-2005 «Syntax for the Digital Object Identifier» [3].

Роботи над стандартом ISO TC46/SC9 для моделі DOI розпочалися у 2004 р. [14] Предметом стандарту було вирішено зробити систему DOI в цілому: її функціонування на рівні принципів синтаксису, опису та диспетчеризації, а також механізмів створення, реєстрації та адміністрування імен (ідентифікаторів) DOI. Після проведення підготовчих робіт, у лютому 2006 р. ISO створила робочу групу ISO TC46/SC9 з розроблення міжнародного стандарту ISO 26324 «Information and documentation – Digital Object Identifier (DOI)». Зараз прийняття стандарту перебуває на стадії 30.99 (прийнятий за основу, як міжнародний стандарт). Під час попереднього голосування 22 країни, у т. ч. і Україна (Держспоживстандарт), висловилися за прийняття проекту; Болгарія, Хорватія і Данія утрималися; Угорщина та Росія не брали участі у голосуванні; проти проекту висловилися Франція [15]. Остаточне прийняття та публікація стандарту заплановані на 2009–2010 рр. [18]

Система DOI перебуває також у процесі затвердження в якості інтернет-стандарту. Хоча у випадках, коли йдеться про інтернет-технології, поняття «стандарт» є доволі розмитим, традиційно «законодавцем Інтернет» прийнято вважати Робочу групу з розвитку інтернет-технологій (IETF), у яку входять найпотужніші інформаційно-технологічні установи світу. Набуття технологією статусу стандарту (RFC STD) – традиційно дуже складний процес, який триває роками. І це цілком закономірно, адже до IETF входять установи з абсолютно протилежними інтересами. Система DOI проходить процес затвердження IETF у якості

однієї з базових схем URI. У 2006 р. простір імен DOI був стандартизований як один з реєстрів схеми info-URI [21]. Схема info-URI включає 28 найважливіших реєстрів імен, які використовуються бібліотечним та видавничим співтовариством. Серед них – системи ідентифікації Бібліотеки Конгресу США, Національних бібліотек Франції та Австралії, OCLC Worldcat, OpenURL, Handle, DOI тощо. Отже, використання системи DOI в межах функціональності бібліотечно-інформаційних та видавничих систем регулюється відповідним Інтернет-стандартом вже сьогодні, процес затвердження даної моделі у якості базової схеми URI наразі триває.

Висновки

Технологічна модель DOI початково розроблялася як один з інструментів підвищення функціональності онлайн-науково-видавничих систем. Проте сьогодні сфера застосування цієї моделі є дуже широкою. За 12 останніх років з'явилися тисячі інформаційних проектів різної спрямованості, що базуються на моделі DOI. Враховуючи обмежений обсяг статті, увагу було сфокусовано на діяльності лише двох «наукових» реєстраційних агенцій DOI. Хоча реально існує поле запровадження моделі DOI дуже широке. Серед іншого, функціонує навіть проект щодо застосування DOI-функціональності для створення цифрових бібліотек клинописних табличок з внутрішніми індексами цитування клинописних документів один одним [17]. З 2004 р. обов'язкову реєстрацію у системі DOI проходять всі офіційні документи Євросоюзу [9], модель запроваджена у проектах наукових та освітніх установ, бібліотек та інформаційних центрів, книговидавничих та книготорговельних організацій, урядових установ тощо. У 2008 р. розгорнулася жвава дискусія щодо необхідності запровадження міжнародних «цифрових ідентифікаторів авторів» (т. зв. DAI) [2], які за аналогією з DOI будуть використовуватися у всіх наукометричних інструментах та унікально ідентифікувати персоналії (науковців).

З наведених прикладів запровадження системи DOI може скластися враження, що дана модель має основну цінність як інструмент для відслідковування цитованості. Однак це не зовсім так. Будь-які семантичні системи мають у своєму базисі певну наскрізну схему ідентифікації, і цитованість об'єктів – лише окремий приклад реалізації семантичних взаємозв'язків. Побудова будь-якого інформометричного апарата розпочинається з прийняття у

проекті єдиної системи ідентифікації, адже для проведення будь-яких математичних обчислень у інформаційних базах потрібно спочатку провести нормалізацію даних у цих базах і ввести унікальний та постійний для бази загалом ключ. А отже питання вибору схеми ідентифікації та обліку мають велике значення принаймні в двох аспектах: «стійкості ключа» та інтероперабельності.

На жаль, у питаннях ідентифікації онлайн-ресурсів вітчизняні виробники інформаційного змісту та інформаційні агрегатори помітно відстають від західних колег, що призводить, зокрема, до проявів безсистемності у розвитку українського інтернет-середовища. Дана прогалина в умовах глобалізації може мати серйозні наслідки для інформаційної політики держави в цілому. Сказане свідчить, що питання впровадження технологій ідентифікації цифрових ресурсів (зокрема, DOI) наразі є актуальними і потребують широкого фахового обговорення.

Список використаних джерел

1. *Литвинова, Н. Н.* Организация интегрированной среды пользователя для поиска в сетевых удаленных ресурсах // Научные и технические библиотеки. – 2008. – № 6. – С. 5–13.
2. *Aerts, R.* Digital identifiers work for articles, so why not for authors? // *Nature*. – Vol. 453, N. 7198. – P. 979. doi:10.1038/453979b.
3. ANSI/NISO Z39.84-2005: Syntax for the Digital Object Identifier: an American national standard developed by the National Information Standards Organization; approved September 30, 2005 by the American National Standards Institute. – Bethesda: NISO Press, 2005. – 21 p.
4. *Brase, J.* Access to Non-textual Information: The Big Challenge for Libraries – What the DOI System Can Do to Help [Virtual Resource] // Resource Access for a Digital World: IDF Open Meeting, Brussels, Belgium, 17 June 2008. – Access Mode : www.doi.org/doi_presentations/members_meeting_2008/presentations/Brase.ppt.
5. BSI British Standards joins CrossRef: Press Release [Virtual Resource] // CrossRef: DOIs for Research Content. – Access Mode : URL : www.crossref.org/01company/pr/press100907.htm.
6. CrossRef: DOIs for research content [Virtual Resource]. – Access Mode : URL : www.crossref.org. – Date of Access : March 2009.
7. CrossRef [Virtual Resource] // Wikipedia: The Free Encyclopedia. – Access Mode : URL : en.wikipedia.org/wiki/CrossRef. – Date of Access : September 2008.
8. DOI Registration Agency [Virtual Resource] / German National Library of Science and Technology. – Access Mode : www.tib-hannover.de/en/the-tib/doi-registration-agency/.
9. European Union adopts DOIs [Virtual Resource] // CAL digital news. – December 2004. – Is. 1. – P. 5. – Access Mode : <http://www.doi.copyright.com.au/reports%20&%20papers/DOI/z1.pdf>.
10. Handle System® [Virtual Resource]. – Access Mode : URL : www.handle.net. – Date of Access : March 2009.
11. New DOI Registration Agency for scientific data appointed: 1,500,000 datasets to be registered by end of 2005: News release [Virtual Resource] / The International DOI Foundation. – 2005. – May, 4. – Access Mode : www.doi.org/news/TIBNews-050405.html.
12. *Paskin, N.* Digital Object Identifier (DOI®) [Virtual Resource] // Encyclopedia of Library and Information Sciences: Third Edition / Taylor & Francis. – Access Mode : URL : www.doi.org/overview/080625DOI-ELIS-Paskin.pdf.
13. *Paskin, N.* DOI: A 2003 progress report [Virtual Resource] // D-Lib Magazine. – 2003. – Vol. 9, № 6. – Access Mode : URL : dx.doi.org/10.1045/june2003-paskin.
14. *Paskin, N.* DOI Working Group Report to the ISO TC46/SC9 plenary [Virtual Resource]. – 2007. – 10 May. – Access Mode : www.lac-bac.gc.ca/iso/tc46sc9/docs/sc9n465.pdf.
15. SC9 № 450: Approved New Work Item 26324, Digital Object Identifier (DOI) system: Results of the voting on the New Work Item proposal distributed in SC9 № 447 [Virtual Resource] / ISO/TC 46/SC 9 Secretariat. – 2006. – 3 August. – Access Mode : www.lac-bac.gc.ca/iso/tc46sc9/docs/sc9n450.pdf.
16. *Sens, I., Brase, J.* A system for easy access to scientific information using DOIs [Virtual Resource] // Global Access to Science: 28th IATUL Conference, Stockholm, Sweden, June 11 – 14 2007. – Access Mode : www.iatul.org/doclibrary/public/Conf_Proceedings/2007/Brase_J_full.pdf.
17. *Sowa, J., Palmero, J.* RA Panel Discussion new applications for the DOI/other industry developments // IDF Members Meeting (Bologna, Italy, 14th June 2005). – Access Mode : www.doi.org/doi_presentations/members_meeting_2005/presentations/PanelApplicationsPalmero.ppt.
18. The DOI® System [Virtual Resource]. – Access Mode : URL : www.doi.org. – Date of Access : March 2009.
19. The Royal Society Registers Archive with CrossRef // Information Today. – 2007. – Vol. 24, N. 3. – P. 37.
20. Thomson Increases ISI Web of Knowledge // Computers in Libraries. – 2004. – Vol. 24. – Iss. 8. – P. 54.
21. Van de Sompel, H. et al. The «info» URI Scheme for Information Assets with Identifiers in Public Namespaces: IETF Request for Comments 4452 [Virtual Resource] / The Internet Society. – 2006. – Access Mode : www.ietf.org/rfc/rfc4452.txt.
22. Wang Jue Digital Object Identifiers and Their Use in Libraries // Serials Review. – 2007. – Vol. 33. – Iss. 3. – P. 161–164.