

Шиненко М. А.,  
зав. сектору мережних технологій і баз даних  
Інституту цифровізації освіти НАПН України  
Кільченко А. В.,  
науковий співробітник сектору мережних технологій і баз даних  
Інституту цифровізації освіти НАПН України

## СЕРВІС DOI CROSSREF ЯК ДЖЕРЕЛО МЕТАДАНИХ АКАДЕМІЧНИХ ВИДАВЦІВ ТА НАУКОВИХ ЖУРНАЛІВ

Цифрові ідентифікатори об'єктів – *Digital Object Identifiers (DOI)* відіграють важливу роль у сучасній науковій комунікації всього світу. Цифрові ідентифікатори публікацій, що зареєстровані міжнародною базою CrossRef з коректними метаданими, дозволяють вирішувати важливе завдання встановлення зв'язків між автором з його публікаціями, між бібліографічними посиланнями на публікації та адресами їхнього місцезнаходження в електронному середовищі.

Міжнародна база описів наукових публікацій CrossRef, що реєструє DOI з 2000 р., нині налічує понад 100 млн DOI. Її сервіси аналізують документальні потоки, відстежують традиційне цитування та нові типи ідентифікації публікацій в мережі Інтернет [1]. На її основі розвиваються нові функції DOI CrossRef.

Тому *основне завдання цієї публікації* – привернути увагу вітчизняної науково-інформаційної спільноти до тих функцій, які, на думку авторів, не дуже широко відомі. Як наслідок, не усвідомлюється ряд проблем, що виникають через недостатню поширеність DOI CrossRef в українських публікаціях, бідності метаданих, що реєструються у цій базі та відсутності формалізованого оформлення DOI CrossRef на сторінках вітчизняних електронних видань. Опанування новими функціями DOI CrossRef, розуміння механізму їх роботи необхідні для ефективного використання можливостей, які вони надають авторам і видавцям наукових публікацій [1]. Вирішення цих проблем є важливими для поглибленої інтеграції українських публікацій у міжнародну інфраструктуру наукових комунікацій та підвищення комфортності роботи користувачів з науковими базами публікацій.

Наукові матеріали, що завантажують метадані та реєструють DOI, поділено за трьома *типами публікацій*: *журнали* (<https://www.CrossRef.org/06members/51depositor.html>); *матеріали конференцій* (<https://data.CrossRef.org/reports/depositorCP.html>); *книги, монографії* (<https://data.CrossRef.org/reports/depositorB.html>).

*Мета роботи* – проаналізувати застосування сервісу DOI CrossRef для отримання метаданих академічних видавців та наукових журналів.

На початку своєї діяльності база даних CrossRef обмежувалася мінімальним набором метаданих для реєстрації DOI. До обов'язкових елементів не включалася навіть назва публікації [2]. З роками повнота та точність метаданих набувала все більшого значення. Конференція CrossRef "How good is your metadata" (2018 р.) була присвячена якості метаданих, що завантажуються в базу CrossRef. Нині в універсальних метаданих Crossref (застосованих до всіх типів контенту) виділяються *базові* (назва, автор, дата публікації, назва джерела, том, випуск, сторінки, електронна адреса) і *розширені*, до яких відносяться реферати/анотації, списки використаної літератури, відомості щодо ліцензій, фондів, афіліації та ORCID авторів, зв'язків з іншими публікаціями тощо. Причому кількість записів, що містять розширений набір метаданих, постійно зростає.

Щоб продемонструвати редакторам наукових журналів, а також дослідникам ключові показники метаданих, фахівцями CrossRef розроблено спеціальний безкоштовний і простий у використанні інструмент, доступний для всіх користувачів – сервіс *Participation Reports* (Звіти про участь). Для того, щоб скористатися сервісом, потрібно перейти на його офіційну сторінку [3] та додати назву організації-депозитора метаданих. Participation Reports дозволяє

візуалізувати у відсотковому співвідношенні повноту представлених метаданих окремої організації (загальний префікс DOI для всіх наукових журналів видавця), а також всередині видавця для окремого наукового журналу за десятьма [1] ключовими показниками (рис. 1).

Для кожного видавця, який співпрацює з Crossref, існує окремий *Звіт про участь (Participation Reports)*, що показує, який відсоток їх депонованих даних зареєстровано для кожного з десяти ключових елементів метаданих [4].

На прикладі міжнародного електронного наукового фахового видання «Інформаційні технології і засоби навчання» (Information Technologies and Learning Tools) розглянемо *Звіт про участь*, який наочно показує, де є прогалини і що можна поліпшити щодо повноти метаданих. Одним із засновників журналу є Інститут інформаційних технологій і засобів навчання НАІН України (Institute of Information Technologies and Learning Tools of NAES of Ukraine).

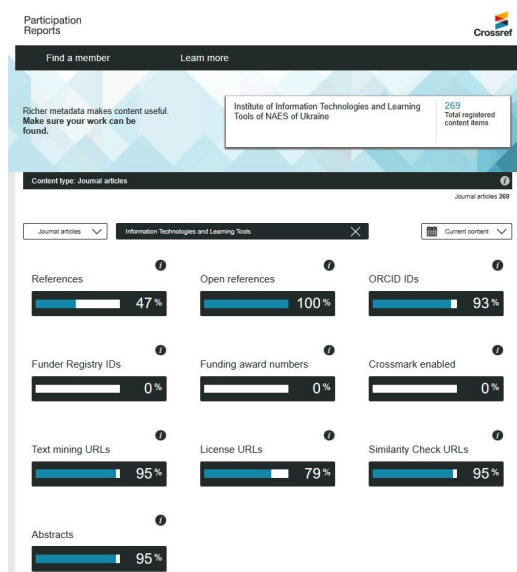


Рис. 1. Статистичні показники корисності контенту обраної наукової установи

На рис. 1 представлений заголовок *Звіту про участь* видавця Institute of Information Technologies and Learning Tools of NAES of Ukraine – загальна кількість елементів контенту 269. Центральне поле заголовка *Звіту про участь* дозволяє ввести назву журналу, збірника або навіть назву публікації та проаналізувати повноту відповідно до завантажених метаданих. Проаналізуємо отримані метадані:

- для 47% публікацій завантажено список літератури (*References*);
- всі (100%) посилання відкриті (*Open References*) – доступні для всіх користувачів сервісів Crossref;
- для 93% авторів вказано *ORCID*;
- для 0% публікацій вказано ім'я та ідентифікатор (*Funder Registry IDs*) спонсора – хоча б однієї з організацій, які фінансували дослідження;
- для 0% публікацій вказано номер гранту фінансування (*Funding award numbers*);
- частка контенту (у нашому випадку 0%), що використовує службу Crossmark (*Crossmark-enabled*), яка надає читачам швидкий та легкий доступ до поточного статусу елемента контенту (у рамках політики видавця щодо виправлень, спростування, відкликання та інших оновлень);
- 95% зареєстрованого контенту містять URL-адресу для інтелектуального аналізу тексту та даних (*Text-mining URLs*) наукової публікації – автоматичного аналізу та вилучення інформації з великої кількості документів. На цей час більшість наукових організацій світу не зацікавлені в створенні спеціального набору інструкцій, за допомогою яких хтось буде досліджувати їхні наукові матеріали;
- 79% метаданих публікацій містять URL-адреси, що вказують на ліцензію (*License URLs*), яка визначає умови, за якими читачі можуть отримати доступ до контенту;

- 95 % метаданих публікацій включають URL-адреси для перевірки схожості (*Similarity Check URLs*), для видань, що співпрацюють з CrossCheck та iThenticate;
- 95 % метаданих публікацій включають інструкції (*Abstracts*), що надають більш глибоке розуміння змісту роботи.

Не потрібно прагнути до 100% показників, але при цьому треба розуміти, що більш повне і акуратне заповнення метаданих публікації тією чи іншою мірою впливає [5] на рейтинги видань, авторів та організацій. А вказівка грантів та фондів підтримки наукової діяльності позитивно впливає на взаємини із цими фондами.

Крім того, CrossRef збагачує метадані, що подані організацією під час реєстрації DOI, доповнюючи їх новими виявленими зв'язками (наприклад, між оригіналом та перекладом, статтею та набором даних), інформацією про фонди, класифікаційними кодами журналів Scopus та ін. Усвідомлюючи важливість розширеного складу метаданих для пошуку публікацій та аналізу публікаційних процесів, видавці не тільки включають їх у поточні реєстрації, а й активно додають у раніше сформовані метадані.

DOI присвоюються не тільки поточним публікаціям, але й сотням матеріалів, що вийшли багато років тому. Кожному DOI зіставлений набір метаданих, що включає базові і все частіше розширені метадані. Це закономірно призводить до розвитку нових опцій DOI, крім функцій, що спочатку планувалися, однозначній ідентифікації об'єктів та зв'язуванню елементів списків використаної літератури з документами на видавничих платформах.

Можна виділити принаймні **нові функції DOI CrossRef**:

- база для аналізу публікацій та публікаційної активності;
- основа роботи сервісів відстеження посилання публікацій у широкому контексті мережних публікацій: власний сервіс CrossRef Event Data, сервіс Altmetric та подібні;
- основа роботи сервісу – визначення статусу вільного доступу до публікацій Unpaywall.

**Аналіз публікаційних потоків та публікаційної активності.** База метаданих CrossRef з обсягом понад 100 млн записів є цінним джерелом даних для бібліометричного та наукометричного аналізу, включаючи аналіз цитування, але не обмежуючись ним. У роботі [2] наведено приклади матеріалів, у яких виконано багатоаспектний аналіз публікацій на базі метаданих CrossRef. Дані з бази можна збирати за допомогою API (<https://www.crossref.org/education/retrieve-metadata/rest-api>) без обмежень за контентом: метадані не охороняються ліцензією, вони поширюються вільно відповідно до законодавства про авторське право.

**CrossRef Event Data та сервісу альтметрик.** Альтернативні метрики або альтметрики належать до посилань публікацій у контекстах, відмінних від цитування, яке в рамках цього протиставлення розглядається як об'єкт традиційних метрик. Посилання (mentions) відстежуються в соціальних мережах загального призначення, таких як Twitter та reddit, наукових соціальних мережах (Mendeley), новинних агрегаторах, блогах та ін. Ці посилання стають все більш важливими показниками впливу публікацій, випереджаючи цитування за оперативністю та доповнюючи показники цитування характеристиками нових аспектів публікацій у мережі [6].

Першим сервісом, що відстежує і підраховує посилання, у 2011 р. став **Altmetric.com** (<https://www.altmetric.com>). Він орієнтований на видавничий світ; цей сервіс інтегрований у безліч видавничих платформ, для яких постачаються відомості щодо посилань публікацій на комерційній основі. Крім цього, Altmetric.com пропонує безкоштовний букмарклет (браузерна закладка, що виконує невеликий JavaScript-код, вказаний в її URL-адресі) Altmetric it!, який може вивести дані альтметрик на вебсторінці публікації. Основною умовою роботи Altmetric it! є наявність DOI. Якщо на сторінці публікації немає DOI або сервіс неспроможний його розпізнати як ідентифікатор цієї публікації, з'являється повідомлення: "Sorry! We couldn't find a DOI on this page" – «Ми не можемо знайти DOI на цій сторінці». Існує також список зареєстрованих у сервісі журналів, для яких можна отримати відомості щодо посилань їх публікацій навіть за відсутності DOI.

**CrossRef Event Data**, який з'явився у 2017 р., також відстежує посилання в Мережі публікацій, що мають DOI. Але на відміну від Altmetric.com та інших сервісів альтметрик (наприклад, PlumX), Event Data не підраховує значення метрик, а генерує за запитами користувачів через безкоштовний API інформацію щодо всіх подій посилання для певного DOI або сукупності DOI, що мають загальний префікс (в джерелах, що відстежуються цим сервісом). Їхній девіз: – «Ми надаємо необроблені дані – вирішуєте, як їх використовувати» (<https://www.crossref.org/services/event-data>). CrossRef Event Data – сервіс, що розвивається; після періоду бета-тестування розробники продовжують активно його вдосконалювати.

Дані поставляються у файлі *json*, з якого можна отримати інформацію про джерела посилання, час посилання, адресу (URL) сторінки, що згадує публікацію та інші дані.

CrossRef Event Data є цікавим для відстеження посилань як окремих публікацій, так і публікацій певних журналів та видавництва, що мають індивідуальні префікси DOI. За допомогою цієї інформації можна виявити канали просування публікацій та групи зацікавлених користувачів. Але для повноцінного спрацьовування Event Data потрібно, щоб публікації мали DOI, а в соціальних мережах та інших мережних спільнотах утвердилася практика включення DOI до посилання публікацій. Можна констатувати, що така ситуація вже є у країнах «умовного Заходу», але ще далека від цього стану в Україні: посилання на публікації в соціальних мережах не так часто включають DOI. Це закономірно проявляється у результатах Event Data щодо посилань публікацій у вітчизняних журналах.

**Визначення статусу вільного доступу до публікацій за допомогою Unpaywall.** Сервіс **Unpaywall** (<http://unpaywall.org>) став дуже популярним за останні роки. Він дозволяє досить точно визначити для кожної публікації, що має DOI CrossRef, наявність вільнодоступних версій як на сайтах видавців, так і в численних репозитаріях відкритого доступу. Цей сервіс інтегрований в авторитетні бази даних – Web of Science, Scopus, Dimensions та ін., забезпечуючи можливість фільтрації результатів пошуку за ознакою доступності повних текстів. Тисячі бібліотек по всьому світу використовують цей сервіс для інформування користувачів щодо вільнодоступних документів завдяки його інтеграції в популярні link resolvers, такі як Primo, Summon, EBSCO Full Text Finder (через API).

Крім того, безкоштовне розширення Unpaywall, яке можна встановити з адреси <http://unpaywall.org/products/extension> у браузері Chrome і Firefox, визначає доступність документа на відкритій у браузері сторінці, якщо знаходить на ній DOI CrossRef цього документа. Якщо статус доступності визначено, на сторінці з'являється значок «замочка»: сірий, якщо документ закрито, і зелений, якщо він доступний. Можна налаштувати розширення, щоб колір «замочка» для вільного доступу був різним: золотим, зеленим або бронзовим відповідно до типу відкритого доступу, як він визначений у роботі [7].

До бази даних Unpaywall звертаються програми, постійно перевіряють статус доступності документів для Web of Science та Scopus. Однак у поданні результатів опрацювання в цих базах є різниця: Scopus враховує результати без змін, а Web of Science додатково надає статус золотого відкритого доступу всім записам на статті з журналів, що входять до Directory of Open Access Journals (DOAJ), незалежно від результатів роботи Unpaywall, зокрема у разі відсутності DOI у статтях. Внаслідок цього статті з журналу, зареєстрованого в DOAJ і що входить одночасно у Web of Science і Scopus, можуть отримати різний статус доступності в цих базах. Слід мати на увазі, що дані розширення Unpaywall, яке звертається до бази Unpaywall для визначення статусу доступності документа з конкретним DOI, можуть не збігатися з даними, отриманими з бази на запит DOI. Можлива причина – у тому, що результат аналізу безпосередньо на вебсторінці документа ще не занесено до бази, яка оновлюється з певною періодичністю.

Робота Unpaywall спирається на базу даних, що постійно оновлюється, формується індексуванням безлічі джерел, перелік яких наведено на сторінці <http://unpaywall.org/sources>. Основа індексування – база CrossRef; DOI є ключовим елементом записів у базі, до нього додається інформація про знайдені в інших джерелах відкритих версій документа з даними DOI та підставах (evidence) для визначення статусу документа як вільнодоступного.

Підставами можуть бути ліцензія в метаданих CrossRef, ліцензія на сторінці документа, факт власності журналу відкритого доступу або репозитарію відкритого доступу, факт доступності файлу pdf на сторінці публікації та ін.

Інформацію з бази даних CrossRef можна отримати через API (<https://unpaywall.org/products/api>) або через онлайн-форму запиту з переліком DOI (<https://unpaywall.org/products/simple-querytool>).

**Висновки.** DOI CrossRef відіграє важливу роль у сучасній науковій комунікації. За два десятиліття свого існування у CrossRef створено більш ніж стомільйонну базу описів наукових публікацій. На її основі розвиваються різноманітні функції, що дозволяють аналізувати документальні потоки, відстежувати традиційне цитування та нові типи посилання публікацій в Інтернеті. На основі бази даних DOI CrossRef розвивається сервіс Unpaywall, що використовується для визначення доступності повних текстів публікацій, до яких можна перейти від результатів пошуку в авторитетних базах даних та пошукових системах у тисячах бібліотек світу. Знайомство з новими функціями DOI CrossRef, розуміння механізму їх роботи необхідне для ефективного використання можливостей, що вони надають авторам та видавцям наукових публікацій.

Враховуючи ці тенденції, можна сформулювати **рекомендації** для авторів та видавців:

- присвоєння та реєстрація DOI CrossRef стає необхідним елементом запровадження публікації в науковий обіг;
- метадані публікації під час реєстрації DOI у CrossRef повинні бути детальними і формалізованими (у тих випадках, коли є рекомендації з цього приводу, наприклад при введенні даних про ліцензію);
- включення DOI публікацій має стати необхідним елементом не тільки у списках літератури, а й за посиланнями публікацій у соціальних мережах та інших мережних спільнотах;
- журнали, що практикують відкритий доступ до всього контенту, слід орієнтувати на ліцензії Creative Commons; реєстрація таких журналів у Directory of Open Access Journals підвищить оприлюднення їх публікацій як вільно доступних;
- на вебсторінках журнальних публікацій необхідно відображати в метаданих елементи, які допомагають формально визначити статус публікації як вільнодоступної.

### Список використаних джерел

1. Іванова С.М., Кільченко А.В. Використання сервісу Participation Reports бази даних Crossref для отримання метаданих академічних видавців та наукових журналів. Імерсивні технології в освіті: матеріали наук.-практ. конф. з міжнар. участю, м. Київ, 22 верес. 2021 р. К.: ІТЗН НАПН України, 2021. С. 88-92. URL: <https://lib.iitta.gov.ua/727346>.
2. Hendricks G., Tkaczyk D., Lin J., Feeney P. Crossref: The sustainable source of community-owned scholarly metadata. Quantitative Science Studies. 2020. Vol. 1 (1). P. 414-427. URL: [https://doi.org/10.1162/qss\\_a\\_00022](https://doi.org/10.1162/qss_a_00022).
3. Participation Reports. URL: <https://www.CrossRef.org/members/rep>.
4. Tolwinska Anna Participation Reports help Crossref members drive research further/. Sci Ed. 2021. № 8 (2). P. 180-185. Published online August 20, 2021. DOI: <https://doi.org/10.6087/kcse.253>.
5. Piwowar H., Priem J., Orr R. The Future of OA: A largescale analysis projecting Open Access publication and readership. 2019. URL: <https://doi.org/10.1101/795310>.
6. Ortega J.L. Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data. Scientometrics. 2018. Vol. 116 (3). P. 2123-2138. URL: <https://doi.org/10.1007/s11192-018-2838-z>.
7. Открытый доступ сегодня: широкомасштабный анализ распространенности и влияния статей открытого доступа / Х. Пивовар и др. Наука и научная информация. 2019. Т. 2. № 4. С. 228-247. URL: <https://doi.org/10.24108/2658-3143-2019-2-4-228-247>.