

РОЗДІЛ 8

**ІНТЕГРАЦІЯ
НАУКОВИХ
ІНФОРМАЦІЙНИХ
РЕСУРСІВ НАН УКРАЇНИ**

Пилип Андон,
Ольга Захарова,
Олександр Новицький,
Галина Проскудіна,
Валерій Резніченко

Харвестер відкритої науки НАН України як систему інтеграції наукових інформаційних ресурсів установ НАН України було розроблено у рамках виконання цільового науково-технічного проекту НАН України «Створення й впровадження інфраструктури відкритої науки в НАН України (OPENS)». Метою цієї розробки є забезпечення ефективнішого оприлюднення отриманих в НАН України наукових результатів завдяки суттєвому полегшенню процесу доступу до наукових інформаційних ресурсів. Ця мета є досяжною з використанням засобів їх інтеграції та наданням єдиної точки пошуку та доступу до них, а також за умови централізованого підключення інтегрованих ресурсів до міжнародних агрегаторів. Це суттєво покращить представлення результатів наукових досліджень НАН України у світі.

У цьому розділі описано обраний для розробки Харвестеру підхід до інтеграції наукових інформаційних ресурсів, запропоноване програмне забезпечення та основні характеристики створеної складової інфраструктури відкритої науки НАН України.

8.1. ШЛЯХИ ВИРІШЕННЯ ЗАВДАННЯ ІНТЕГРАЦІЇ

За останні роки створення систем інтеграції інформаційних ресурсів стало важливим напрямом практичних розробок інформаційних систем різного призначення, зокрема електронних бібліотек (ЕБ) або архівів. Під інтеграцією даних в електронних системах ми розуміємо забезпечення єдиного уніфікованого інтерфейсу для доступу користувачів до сукупності автономних джерел, які зазвичай мають неоднорідність щодо деяких їх властивостей. Існує кілька підходів до вирішення проблеми створення електронних бібліотек з інтегрованими інформаційними ресурсами. Серед них можна виділити два класи систем: з інтегрованим і з розподіленим веденням ресурсів.

Підхід з *інтегрованим веденням ресурсів* передбачає збір, збереження й обробку інформаційних ресурсів у єдиному репозитарії.

Такий підхід видається доцільним у випадку, коли інформаційні ресурси організаційно в одній структурі і безпосередньо належать одному власнику. Прикладом таких організаційних структур можуть бути науково-дослідні інститути НАН України. Будь-який інститут може створити інтегровану наукову електронну бібліотеку (НЕБ), яка б містила колекції її різноманітних наукових ресурсів (наукові статті, звіти, дисертації, навчально-методичні посібники, матеріали конференцій тощо) і вирішувала всі питання ведення таких ресурсів та надання доступу до них. Однак, наприклад, для створення НЕБ НАН України, яка б припускала централізацію всіх ресурсів, такий підхід не зовсім доцільний, оскільки він потребує вирішення багатьох складних організаційно-технічних завдань. Насамперед спрямованих на збір вихідних інформаційних ресурсів з інститутів. Крім того, він потребує створення розвинутої структури ведення такої НЕБ.

Підхід з *розподіленим веденням ресурсів* припускає, що існує багато організацій, які здійснюють самостійне створення і ведення НЕБ і надають можливість доступу до цих ресурсів, включаючи також і організацію пошуку необхідних ресурсів. Крім того, існує «надбудова» над ними, що забезпечує можливість пошуку за цими ресурсами і, за наявності відповідних умов, надавати доступ до самих ресурсів. У цьому випадку інститути НАН України створюють свої НЕБ, а інтегрована служба на рівні НАН України забезпечує пошук відповідних ресурсів і доступ до них. Відомо два концептуальних рішення реалізації цього підходу. Перше – існування механізму *перехресного пошуку* за багатьма архівами, коли всі ресурси, бібліографічні описи та пошуковий сервіс знаходяться в організації. Так працюють системи з використанням протоколу Z39.50¹: клієнт-серверний протокол для пошуку та отримання інформації з віддалених комп'ютерних баз даних. Тоді пошук здійснюється завдяки безпосередньому зверненню до всіх або до вибраних користувачем електронних бібліотек з подальшим зведенням одержаних результатів у єдиний список. Друге рішення пропонує *здійснювати збір (харвестинг, harvesting) метаданих*, що описують інформаційні ресурси «на місцях» для того, щоб можна було надати централізований пошук в одному місці на основі зібраних метаданих. За суттю це є деякий аналог інтегрованого електронного каталогу.

¹ Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. ANSI/NISO Z39.50-2003. URL: <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf> (hereinafter, the date of the last accessed: 30.07.2025).

8.2. ІНІЦІАТИВА ВІДКРИТИХ АРХІВІВ

Ініціатива відкритих архівів (Open Archive Initiative, OAI, <https://www.openarchives.org/>) виникла в зв'язку з тим, що багато організацій, де створюються електронні інформаційні ресурси, насамперед наукові, вирішили надати відкритий доступ до них. Так постала проблема надання інтегрованого доступу до неоднорідних (гетерогенних) репозитаріїв. ОАІ націлена саме на те, щоб розробляти та сприяти розвитку й поширенню середовища і відповідних стандартів, які б допомогли об'єднати зусилля з надання інтегрованого доступу до їх ресурсів. Суть підходу відкритих архівів полягає у тому, щоб надавати вебдоступ до інформаційних ресурсів, розташованих у інтероперабельних репозитаріях, за допомогою організації спільного використання, публікації й архівування метаданих таких ресурсів.

8.2.1. ПРОТОКОЛ ОАІ ДЛЯ ЗБОРУ МЕТАДАНИХ (ОАІ-РМН)

Протокол ОАІ для збору (харвестінгу) метаданих (Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH, <https://www.openarchives.org/pmh/>) визначає механізм збору записів, що містять метадані з репозитаріїв. Протокол надає провайдерам даних простий спосіб такого представлення метаданих, який робить їх доступними для провайдерів сервісів. Для обміну метаданими використовуються технології HTTP (Hypertext Transport Protocol) і XML (Extensible Markup Language). Зібрані в такий спосіб метадані можуть бути представлені в будь-якому форматі, обраному організаціями, які створюють інтегровану федеративну електронну бібліотеку. Проте в протоколі ОАІ-РМН для забезпечення базового рівня інтероперабельності специфіковано формат Дублінського ядра (Dublin Core, DC)². Отже, метадані з різних неоднорідних джерел поєднуються в єдиній базі даних для того, щоб надати множину сервісів на основі таких агрегованих метаданих. Зв'язки між такими об'єднаними метаданими і відповідними інформаційними ресурсами (тобто з контентом інформаційних ресурсів) не визначаються в цьому протоколі, тож він не надає можливості робити повнотекстовий пошук за інформаційними ресурсами, а тільки за їхніми метаданими. Він просто дозволяє об'єднати інформаційні ресурси на рівні метаданих, і саме на цьому рівні виконувати пошук. Хоча концепція протоколу ОАІ-РМН досить проста, однак побудова на її основі відповідного набору сервісів, які б задовольнили потреби користувачів, залиша-

² ANSI/NISO Z39.85-2012. The Dublin Core Metadata Element Set. – National Information Standards Organization. 2013. URL: https://groups.niso.org/higherlogic/ws/public/download/10258/Z39-85-2012_dublin_core.pdf

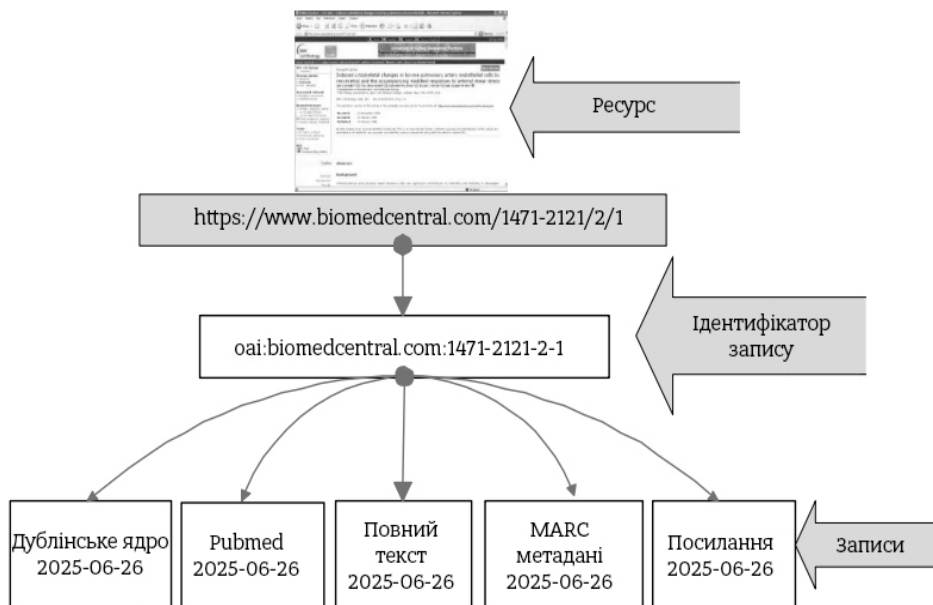


Рис. 8.1. Інформаційна модель OAI-PMH

ється досить складним завданням. Це завдання цілком лежить на «плечах» провайдера сервісів, своєрідної пошукової системи, що допомагає користувачам знаходити інформацію та досліджувати декілька репозитаріїв одночасно.

8.2.2. ІНФОРМАЦІЙНА МОДЕЛЬ OAI-PMH

Виходячи з того факту, що джерелом виникнення протоколу була електронна публікація, модель даних OAI-PMH загалом інтерпретується в термінах бібліографічних даних, що описують академічний ресурс, хоча можливі й інші інтерпретації. OAI-PMH має просту й гнучку інформаційну модель (рис. 8.1).

У верхній частині моделі – описуваний ресурс: традиційний бібліотечний об'єкт (наприклад, книга, стаття), або інші сутності (наприклад, зображення, поняття). Нижче – елемент, складова репозитарію, за допомогою якої поширюються метадані про ресурс. Елемент концептуально є контейнером, що зберігає або динамічно генерує метадані про ресурс у декількох форматах, кожен з яких може бути зібраний у вигляді запису через OAI-PMH. Кожний елемент має ідентифікатор запису (OAI identifier) або унікальний ідентифікатор, який однозначно визначає елемент у репозитарії та використовується у запитах OAI-PMH для складання метаданих з елементів. Елемент

може містити метадані у декількох форматах. Унікальний ідентифікатор вказує на елемент, і всі можливі записи, що є в одному елементі, разом використовують один і той самий унікальний ідентифікатор. Записи описують ресурс у довільному форматі метаданих, що може бути виражений у XML Schema. В ідею протоколу OAI-PMH закладена підтримка будь-яких схем опису метаданих, але він потребує обов'язкового включення в опис ресурсу набору метаданих Дублінського ядра. Необхідно підкреслити, що ідентифікатор запису не є ідентифікатором документа (об'єкта). Очевидно, що багато користувачів захочуть отримати доступ до повного тексту ресурсу, описаного записом метаданих. Протокол рекомендує, щоб архіви використовували елемент запису метаданих для зв'язування запису з ідентифікатором (URL, URN, DOI тощо) асоційованого документа (об'єкта). Для цієї мети обов'язковий формат надає елемент «ідентифікатор» (DC.Identifier).

8.2.3. ПРОВАЙДЕРИ ДАНИХ І СЕРВІСІВ

Концепція протоколу OAI-PMH виділяє дві ролі: провайдера даних і провайдера сервісів³. Провайдер даних — це служба, що підтримує створення і ведення одного чи більше репозитаріїв (бази документів, архівів, електронних бібліотек), здійснює публікацію своїх ресурсів, а також надає можливість доступу до метаданих для використання в інших системах. Як правило, провайдер даних надає вільний доступ до своїх метаданих і, можливо, але не обов'язково, вільний доступ до повних текстів документів. Провайдер даних може мати самостійний вебінтерфейс для організації пошуку, перегляду і доступу до своїх ресурсів, а також інші сервіси, що надаються кінцевим користувачам, а також самостійно вирішує питання про відкритість своїх інформаційних ресурсів і доступність до них. Зокрема, провайдер даних може вирішити інтегрувати усі або частину своїх інформаційних ресурсів на рівні метаданих у провайдера сервісів і для цього організувати експорт відповідних метаданих у форматі протоколу OAI-PMH. Провайдер сервісів здійснює збір і збереження метаданих, наданих провайдерами даних, для надання кінцевим користувачам різних сервісів. Для організації пошуку необхідної інформації не здійснюється звертання до провайдерів даних, тому що він відбувається на основі збережених метаданих у провайдера сервісів. Провайдер сервісів може збирати не всі метадані, надані про-

³ Резніченко В.А., Новицький О.В., Проскудіна Г.Ю. Інтеграція наукових електронних бібліотек на основі протоколу OAI-PMH. *Проблеми програмування*. 2007. № 2. С. 97–112. URL: <https://nasplib.isoftware.kiev.ua/handle/123456789/291>

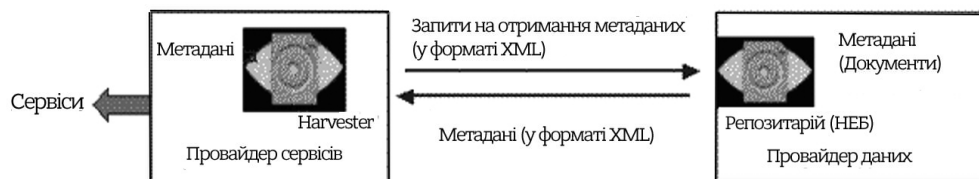


Рис. 8.2. Загальна схема взаємодії провайдера даних і провайдера сервісів

вайдером даних, а тільки ті, які доцільно збирати на основі тих чи інших критеріїв. Схема такої взаємодії наведена на рис. 8.2.

Зазначимо, що в цій схемі той самий провайдер може виконувати обидві ролі: забезпечувати як створення і ведення репозитаріїв інформаційних ресурсів, так і здійснювати збір метаданих від інших провайдерів і надавати на їхній основі необхідні сервіси.

Запропоновані рішення є гнучкими. На рис. 8.3 показані різні варіанти розвитку цього підходу, зокрема з використанням багатьох провайдерів сервісів. Таке рішення дає змогу створювати проблемно чи функціонально орієнтовані провайдери сервісів, а також, за необхідності, знижувати навантаження на провайдера сервісів. На рис. 8.3, б показано розширення з включенням агрегатора, який виконує обидві функції – провайдера даних і сервісів. Як провайдер сервісів він збирає метадані від інших провайдерів даних і потім, виконуючи функцію провайдера даних, робить зібрані метадані доступними для збору іншими провайдерами сервісів. Завданням агрегатора є обробка зібраних метаданих, наприклад, їх нормалізація або перетворення в інший формат метаданих. Також можна використовувати множину провайдерів сервісів та агрегаторів з одночасним комбінуванням з пошуком (рис. 8.3, в).

8.2.4. АРХІТЕКТУРНІ РІШЕННЯ

Організації, виходячи зі своїх можливостей і потреб, створюють, збирають і підтримують в електронному вигляді свої власні інформаційні ресурси як електронні колекції. У загальному випадку склад інформаційних ресурсів може бути довільним, як і перелік користувацьких сервісів. Мінімальними сервісами є перегляд і пошук необхідної інформації та надання результатів пошуку кінцевим користувачам. Крім того, можуть включатися інші сервіси, які є характерними для НЕБ⁴, зокрема і механізми обмеження прав доступу до тих чи інших ресурсів.

⁴ Резніченко В.А., Захарова О.В., Захарова Е.Г. Електронні бібліотеки: інформаційні ресурси та сервіси. *Проблеми програмування*. 2005. № 4. С. 60–72. URL: <https://nasplib.isofts.kiev.ua/handle/123456789/1374>

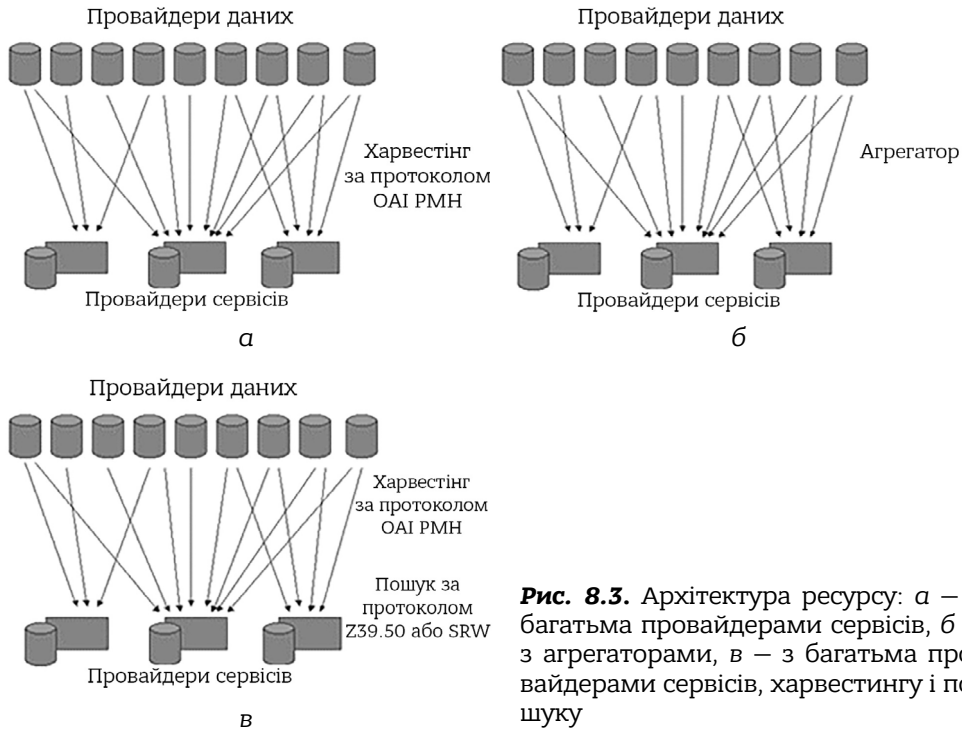


Рис. 8.3. Архітектура ресурсу: а – з багатьма провайдерами сервісів, б – з агрегаторами, в – з багатьма провайдерами сервісів, харвестингу і пошуку

З огляду на вищеописану концепцію OAI-PMH ми запропонували таку архітектуру інтеграції розподілених НЕБ НАН України (рис. 8.4).

Виходячи з розроблених концепцій функціонування НЕБ, організації вибирають програмне забезпечення, здатне задовольнити необхідні (передбачувані) потреби. У розглянутій архітектурі не накладається принципових обмежень щодо обраного програмного забезпечення за винятком одного – воно повинно підтримувати механізм експорту метаданих у форматі OAI-PMH. Існує багато інструментальних засобів створення НЕБ, що підтримують протокол OAI-PMH, зокрема безкоштовні програмні продукти, які виконують у розглянутій архітектурі функції провайдерів даних. Для інтеграції НЕБ створюється служба, яка буде виконувати функції провайдера сервісів. В її обов'язки входить:

- збір метаданих від провайдерів даних, які надають свої інформаційні ресурси в інтегровану пошукову службу провайдера сервісів;
- індексування метаданих з метою надання сервісу з організації пошуку інформаційних ресурсів;
- надання сервісів з перегляду і пошуку інформаційних ресурсів провайдерів даних з можливим наданням посилань до повних текстів документів та інших сервісів.

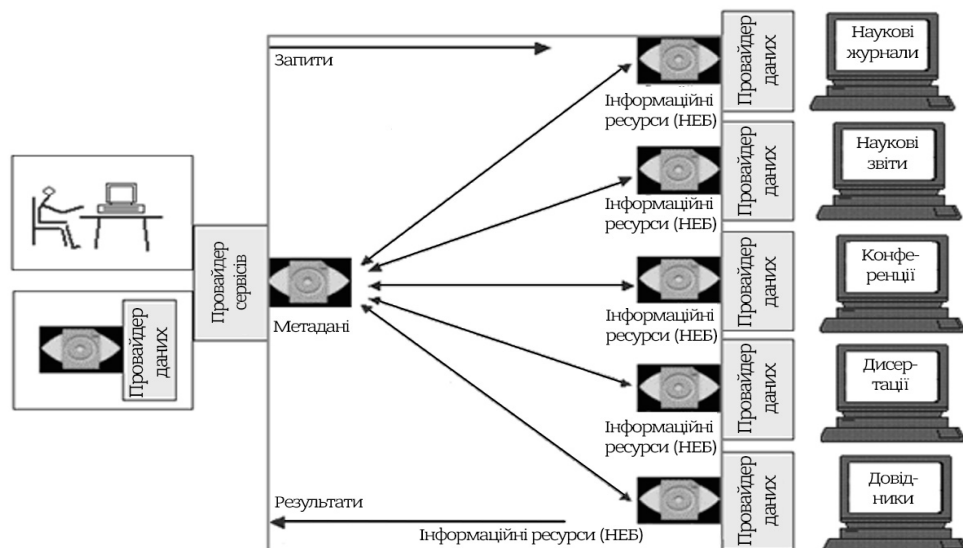


Рис. 8.4. Архітектура інтеграції НЕБ установ НАН України

Суттєвою перевагою такої архітектури є те, що пошук стає незалежним від організації репозитарію інформаційних ресурсів. Пошук в одній великій базі даних або в тисячі дрібних для користувача є одним, оскільки цим займаються харвестери, які здійснюють пошук у відомих їм базах даних.

8.3. ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ПІДТРИМКИ OAI-PMH

З метою обрання для Харвестера НАН України оптимального програмного забезпечення, що підтримує протокол OAI-PMH, ми виконали порівняльний аналіз доступних і поширених продуктів, результати якого показані у табл. 8.1 і стисло викладені нижче.

MyCoRe – це програмний фреймворк з відкритим кодом, який в основному використовують для створення та управління цифровими репозитаріями, архівами та бібліотеками. Він надає набір основних функцій для цих застосувань, таких як управління метаданими, підтримка пошукових запитів та інтерфейс OAI-харвестінгу. Має понад 70 застосувань. MyCoRe розроблено спільнотю бібліотек і комп'ютерних центрів німецьких університетів.

Europeana OAI-PMH server – OAI-PMH сервер електронної бібліотеки Europeana, яка містить відскановані інформаційні ресурси, що відображають різні аспекти європейської культури.

Таблиця 8.1. Програмне забезпечення підтримки OAI-PMH

Назва продукту	Посилання	Технологія	Дата останнього релізу	Наявність інтерфейсу користувача	підтримка функцій провайдера даних	Підтримка функцій провайдера сервісів	URL
MyCoRe	https://www.mycore.de/	Java, JavaScript, XSLT	Грудень, 22, 2023	+	+	+	https://github.com/MyCoRe-Org
Europeana OAI-PMH server	https://github.com/europeana/OAI-PMH2	Java, HTML	Липень 21, 2022	?	+	Обмежена	https://github.com/europeana/OAI-PMH2
OAI-PMH Harvester (плагін до Omeka)	https://omeka.org/classic/plugins/OaipmhHarvester/	PHP, HTML	Червень, 06, 2025	+	+	+	https://github.com/Daniel-KM/Omeka-S-module-OaipmhHarvester
Yoda	https://utrechtuniversity.github.io/yoda/	Python, HTML, Jinja	Лютий, 2022	?	+	+	https://github.com/UtrechtUniversity/yoda
OAI-PMH harvester for CKAN	https://extensions.ckan.org/extension/oaipmh/	Python	?	?	-	+	https://github.com/kata-csc/ckanext-oaipmh
jOAI Software	https://europeana-oai.clarin.eu/	Java, XSLT	Лютий 19, 2022	-	+	+	https://github.com/NCAR/joai-project
Apache Camel	https://camel.apache.org/ , https://camel.apache.org/components/4.10.x/oaipmh-component.html	Java	Червень 27, 2025	-	+	+	https://camel.apache.org/download/ , https://github.com/apache/camel
Dataverse OAI-PMH	https://dataverse.org/	Java	Грудень 12, 2024	+	+	+	https://github.com/IQSS/dataverse
VuFind	https://vufind.org/vufind/	PHP, Java	Червень 23, 2025	+	+	+	https://github.com/vufind-org/vufind

Плагін *OAI-PMH Harvester* для *Omeka* дає користувачам змогу імпортувати записи від зовнішніх постачальників даних, які використовують протокол *OAI-PMH*. Плагін ефективно витягує метадані з цих репозитаріїв та відображає їх у моделі даних *Omeka*, що допомагає користувачам включати дані з різних онлайнових архівів до своїх колекцій *Omeka*.

Yoda – це платформа управління дослідницькими даними, яка використовує протокол *OAI-PMH*, що дає змогу харвестерам отримувати метадані у форматах *Dublin Core* та *DataCite*, щоб поширювати метадані з наборів даних *Yoda* в інші системи та портали.

OAI-PMH harvester for SKAN. Забезпечує харвестинг метаданих по протоколу *OAI-PMH* у *SKAN* – систему керування даними з відкритим кодом для підтримки центрів і порталів даних, яка забезпечує роботу сотень порталів даних по всьому світу. Підтримувані схеми метаданих: *DC* та *CMDI* (*Component MetaData Infrastructure*).

jOAI – вебзастосунок на *Java*, який реалізує протокол *OAI-PMH*. Ця програма виконує функції як провайдера даних, що робить *XML*-файли метаданих доступними для інших, так и харвестера даних, який може отримувати *XML*-файли метаданих від інших.

Apache Camel OAI-PMH – компонента, яка забезпечує зв'язок *Apache Camel* з репозитаріями за допомогою протоколу *OAI-PMH* і дає йому змогу бути як споживачем, так і виробником метаданих *OAI-PMH*.

Dataverse OAI-PMH для обміну метаданими з іншими репозитаріями. Інсталяції *Dataverse* можуть бути як клієнтами, так і серверами харвестинга метаданих, тобто збирати метадані з інших інсталяцій *Dataverse* або будь-якого іншого репозитарію, сумісного з *OAI-PMH*, а також надавати власні метадані іншим клієнтам.

Детальніше програмне забезпечення *VuFind*, *DSpace* і *Omeka S* проаналізовано у спеціальній публікації⁵. На основі аналізування сучасних програмних продуктів для Харвестеру була обрана система *VuFind*, розроблена в університеті Вілланова, США⁶. Це відкрите програмне забезпечення поширюється за стандартною публічною ліцензією *GNU* (https://en.wikipedia.org/wiki/GNU_General_Public_License) – система інтеграції та пошуку бібліотечних інформаційних ресурсів.

⁵ Novytskyi O. Metadata harvesting for digital library integration in Ukraine: a comparative study of the *OAI-PMH* protocol and *VuFind*'s efficacy. *Digital Library Perspectives*. 2025. Vol. 41. No. 1. P. 74–90. URL: <https://www.emerald.com/insight/content/doi/10.1108/dlp-03-2024-0034/full/html>

⁶ Houser J. The *VuFind* implementation at Villanova University. *Library Hi Tech*. 2009. Vol. 27. No. 1. P. 93–105. <https://doi.org/10.1108/07378830910942955>

8.4. VuFind: ВІДКРИТЕ РІШЕННЯ ДЛЯ ІНТЕГРАЦІЇ БІБЛІОТЕЧНИХ КОЛЕКЦІЙ

У VuFind механізм пошуку реалізований на пошуковій системі Apache Solr, яка оснований на дослідницькій бібліотеці Lucene (www.lucene.apache.org). Останню на сьогодні використовують більшість електронних бібліотечних систем. Вигляд і функції системи VuFind детально проілюстровані (<http://vufind.org/demo/>), до того ж вона використовує технологію Wiki (<https://vufind.org/wiki/>), де користувачі VuFind розміщують практичні поради, сценарії та іншу корисну інформацію. Однією з ключових рис VuFind є здатність об'єднувати результати пошуку з різних джерел даних, забезпечуючи користувачам єдиний і зручний інтерфейс. Пошук у VuFind дає змогу ефективно шукати одразу в декількох каталогах бібліотек, цифрових архівах, базах даних тощо, надаючи комплексний огляд доступних матеріалів. Грунтуючись на відкритому вихідному коді, VuFind забезпечує бібліотеки гнучкістю в налаштуванні та адаптації під свої унікальні потреби. Розширені можливості конфігурації дають змогу впроваджувати індивідуальні налаштування відповідно до вимог кожної організації. Основою концепції системи є пошук, виявлення та спільне використання.

Пошук – основна функція, що надається користувачам, щоб знаходити інформацію в бібліотечних колекціях і ресурсах. Охоплює пошук книг, журналів, статей, електронних ресурсів, цифрових архівів і багатьох інших видів даних. Користувачі можуть виконувати пошук за ключовими словами, авторами, темами та іншими критеріями, що допомагає їм знайти потрібні ресурси швидко і точно. VuFind індексує всі наявні записи всіх бібліотек, підключених до системи, що дає змогу реалізувати простий і розширений пошук.

Виявлення означає, що система надає не тільки результати пошуку, а й активно допомагає користувачам відкривати нові та цікаві ресурси: передбачені рекомендаційні системи, засновані на аналізі користувацьких запитів, пов'язані ресурси, аналоги тощо.

Спільне використання передбачає, що користувачі можуть легко ділитись знайденими ресурсами з іншими через соціальні медіа, надсилання посилань електронною поштою, створення закладок і багато іншого.

Важливими є й інші характеристики. *Безліч джерел даних*. VuFind підтримує різноманітні джерела даних, включно з різноманітними системами управління бібліотеками, електронними каталогами, репозитаріями, базами даних і багатьма іншими. Це дає змогу інтегрувати та керувати різними видами ресурсів в одній системі.

Персоналізація та профілі користувачів. Система надає можливість налаштування профілів користувачів, щоб адаптувати інформа-

цію та ресурси під конкретні потреби. Можливе налаштування рекомендаційних систем, оформлення інтерфейсу та інші функції, що сприяють кращій персоналізації.

Інтеграція з відкритими інтернет-сервісами, такими як Google Books, OpenStreetMap тощо.

Багато локацій і багатокористувацький доступ. VuFind підтримує роботу з великою кількістю бібліотечних систем, а також забезпечує багатокористувацький доступ. Це допомагає організаціям із розподіленою структурою ефективно керувати своїми ресурсами та надавати широкий доступ користувачам.

Відкрите програмне забезпечення і співтовариство. Оскільки VuFind розробляється як відкрите програмне забезпечення, це дає змогу адаптувати і розширювати систему відповідно до своїх потреб. Існує також активне товариство розробників і користувачів, які співпрацюють для поліпшення системи.

Харвестінг у VuFind. Система може бути використана як харвестер для збору інформації з різних джерел та інтеграції в єдину систему. Нижче описано декілька прикладів використання VuFind як харвестеру даних.

Імпорт каталогів бібліотечних систем. VuFind можна використати для імпорту даних з наявних бібліотечних систем або каталогів. Бібліотека може використовувати його для збору даних з її інтегрованих бібліотечних систем і надавати своїм користувачам зручніший та сучасніший інтерфейс для пошуку та навігації.

Інтеграція з цифровими архівами та репозитаріями допомагає надавати доступ до цифрових колекцій, а також ресурсів, як-от електронні тезауруси та архіви з відкритим доступом.

Інтеграція із зовнішніми базами даних та онлайнними ресурсами є одним із сильних аспектів системи VuFind, яка може бути налаштована для збору даних із різноманітних зовнішніх баз даних та онлайнних ресурсів, як от наукові журнали, бази даних, електронні книги. Це дає змогу користувачам шукати інформацію в різних ресурсах і різних типів через єдиний інтерфейс і отримувати результати пошуку, а також посилання на повні тексти статей і журналів.

Імпорт і аналіз даних відкритого доступу, таких як відкриті архіви, бібліотеки з відкритими даними та онлайнні репозитарії. Це допомагає бібліотекам та установам збирати та інтегрувати безкоштовні ресурси.

Гарні приклади застосування VuFind надають бібліотечні системи: на його основі працюють понад 200 систем (<https://vufind.org/wiki/community:installations>). Її потужні можливості яскраво демонструє система BASE (Bielefeld Academic Search Engine, <https://www.base-search.net/about/en/index.php>), одна з найоб'ємніших пошуко-

вих систем у світі, особливо для академічних вебресурсів, як от наукові статті, дисертації, препринти тощо.⁷

Опис системи VuFind. Архітектура системи охоплює такі компоненти:

- інтерфейс користувача (UI);
- ядро системи (Application Core);
- індексація та пошуковий движок (Indexing and Search Engine, Apache SOLR);
- інтеграція з джерелами даних (Data Source Integration);
- управління автентифікацією та авторизацією (Authentication and Authorization);
- пошукові фільтри та розширення (Search Filters and Extensions);
- база даних (Database, DB);
- кешування та керування ресурсами (Caching and Resource Management);
- інтерфейс для адміністрування (Admin Interface).

VuFind дає змогу створювати та керувати різними типами облікових записів, які мають різні права та можливості в системі, і передбачає три основні групи користувачів: анонімні користувачі, зареєстровані користувачі та адміністратори. Опис їхніх можливостей наведено у підрозділі 8.5.1.

Набір пошукових полів або характеристик для документів VuFind відповідає стандарту Дублінського ядра і підтримує пошук за метаданими з використанням стандарту SPARQL, що є мовою запитів до даних, представлених у форматі RDF (основному у семантичному вебі). Також можна використовувати інші формати метаданих: MARC 21, MODS, METS, EAD тощо, які використовують для опису різних типів бібліотечних ресурсів.

Поле тег у системі VuFind є елементом форми, який дає користувачеві змогу вводити чи обирати теги (ключові слова чи фрази, які описують зміст або характеристики ресурсу). Функція додавання власних тегів допомагає краще організувати власну колекцію.

Перегляд за фасетами є потужним інструментом для уточнення і навігації за результатами пошуку. Фасети (або фасетні поля) – це категорії, які групують результати пошуку за певними атрибутами або метаданими, наприклад, автор, організація, рік публікації, тема, тип ресурсу. Цей підхід дає змогу користувачам точніше визначити параметри, що їх цікавлять, і звужувати результати пошуку.

Авторитетний пошук широко застосовують в бібліотечних системах і наукових ресурсах, де стандартизація та ідентифікація авто-

⁷ Pieper D., Summann F. Bielefeld Academic Search Engine (BASE) An end-user oriented institutional repository search service. *Library Hi Tech*. 2006. 24. 4. <https://doi.org/10.1108/07378830610715473>

рів, термінів і тем є критично важливою для точного дослідження. Наприклад, використання авторитетних файлів, нормалізація імен і термінів для дотримання стандартів і запобігання неоднозначності, уніфікація результатів запиту, тобто надання стандартизованих посилань на відповідні авторитетні файли, використання унікальних ідентифікаторів, наприклад, ORCID дослідників.

8.5. ХАРВЕСТЕР ВІДКРИТОЇ НАУКИ НАН УКРАЇНИ

На основі виконаних раніше⁸ і поточних досліджень зі створення системи інтеграції ресурсів з різних відкритих академічних джерел були сформульовані основні вимоги до майбутнього Харвестеру.

1. Можливість імпортувати наукові інформаційні ресурси установ НАН України як на рівні метаданих, так і самі ресурси. Інформаційні ресурси описуються схемою метаданих Дублінського ядра. Передання ресурсів здійснюється з використанням стандартного протоколу обміну метаданими OAI PMH. Надається можливість передавати метадані усіх основних типів наукових інформаційних ресурсів: текстові, набори даних, програмне забезпечення, аудіо, відео, мапи, музичні твори тощо.

2. Пошукові можливості на рівні Apache Lucene. Можливості зруження простору пошуку. Зручний спосіб перегляду результатів.

3. Розвинуті засоби адміністрування (резервне копіювання, відновлення, реєстрація користувачів, права доступу тощо).

⁸ Резніченко В.А. та ін., 2007; Проскудіна Г., Овдій О., Резніченко В., Кудім К. Досвід використання наукової електронної бібліотеки NASPLIB. *Інформація, комунікація, суспільство 2013: матеріали 2-ої Міжнародної наукової конференції ІКС-2013* (16–19 травня 2013 року, Україна, Львів, Славське). Львів, 2013. С. 254–255. URL: <https://ena.lpnu.ua/handle/ntb/20832>; Кудім К.О., Новицький О.В., Проскудіна Г.Ю., Резніченко В.А. Статистика використання наукової електронної бібліотеки періодичних видань НАН України. *Наука України у світовому інформаційному просторі*. 2014. № 10, С. 60–77. URL: <https://nasplib.isofts.kiev.ua/handle/123456789/122992>; Кудим К.А., Проскудіна Г.Ю., Резніченко В.А. Сравнительный анализ функциональных возможностей систем электронных библиотек. *Проблеми програмування*. 2007. № 4. С. 32–49. URL: <https://nasplib.isofts.kiev.ua/handle/123456789/273>; Захарова Е.Г., Захарова О.В., Резніченко В.А. Опис наукових електронних ресурсів метаданими Дублінського ядра. *Проблеми програмування*. 2008. № 2–3. С. 507–514. URL: <https://nasplib.isofts.kiev.ua/handle/123456789/1472>; Резніченко В.А., Проскудіна Г.Ю. Концептуальна модель наукових інформаційних ресурсів НАН України. *Інженерія програмного забезпечення*. 2013. Т. 14. № 2. С. 29–37. URL: <https://jrnl.nau.edu.ua/index.php/IPZ/article/view/6021>; Novytskyi O., Spirin O. Methods and Tools for Building Open Systems of Scientific Research Support. *CEUR Workshop Proceedings*. 2024. Vol. 3806. P. 449–461. URL: https://ceur-ws.org/Vol-3806/S_7_Novytskyi_Spirin.pdf; Кудім К.О., Резніченко В.А., Новиць-

4. Продуктивність (здатність ефективно функціонувати з обсягом 2–3 млн документів).

5. Ведення статистичної інформації.

6. Підтримка багатомовності.

7. Можливість передавати метадані до інших харвестерів, агрегаторів за протоколом OAI-PMH.

Загальна архітектура Харвестеру наведена на рис 8.5. Ядром цієї архітектури є ПЗ VuFind, що підтримує усі основні функції Харвестеру. Постачальником даних для Харвестеру є репозитарії провайдерів даних, які повинні бути реалізовані з використанням ПЗ, що підтримують схему опису метаданих DC та протокол обміну даними OAI-PMH. Прикладами таких ПЗ є DSpace, EPrints, Fedora, OJS.

VuFind надає розвинені функції пошуку та перегляду інформаційних ресурсів на рівні стандарту Apache Lucene та їх маніпулювання. Результати пошуку можуть бути роздруковані, відправлені на пошту, експортовані та збережені для подальшого використання. Харвестер підключений та передає свої інформаційні ресурси до найбільших у світі агрегаторів: OpenDOAR, Open AIRE, BASE, CORE. Для отримання статистичної та аналітичної інформації щодо функціонування Харвестер підключений до Google Analytics, Matomo, Flag Counter.

Харвестинг інформаційних ресурсів відбувається, як зазначено вище, за такими двома умовами: інформаційні ресурси репозитарію установи НАН України описані схемою метаданих простого або розширеного Дублінського ядра; репозитарій підтримує стандартний протокол обміну метаданими OAI-PMH.

кий О.В., Проскудіна Г.Ю., Овдій О.М. Розробка інтегрованої системи періодичних наукових видань на основі OJS. *Проблеми програмування*. 2015. № 3. С. 72–85. URL: <https://pp.isoftware.kiev.ua/index.php/ojs1/article/view/150>; Світунов С.Я., Перконос П.І., Суботін С.В., Твердохліб Є.М., Резніченко В.А. Особливості формування інфраструктури відкритої науки в Україні. *Проблеми програмування*. 2022. № 3–4. С. 335–348. <https://doi.org/10.15407/pp2022.03-04.335>; Копанева В.О., Костенко Л.Й., Новицький О.В., Резніченко В.А. Завдання цифрової трансформації науково-інформаційного середовища. *Проблеми програмування*. 2023. № 1. С. 3–10. <http://doi.org/10.15407/pp2023.01.003>; Резніченко В., Проскудіна Г., Кудім К. Сучасний підхід до створення та інтеграції бібліотечних електронних інформаційних ресурсів. *Вісник Книжкової палати*. 2024. № 6. С. 24–33. [http://doi.org/10.36273/2076-9555.2024.6\(335\).24-33](http://doi.org/10.36273/2076-9555.2024.6(335).24-33); Проскудіна Г.Ю., Кудім К.О. Огляд глобальних служб агрегації ресурсів відкритого доступу та їхніх вимог до постачальників даних. *Проблеми програмування*. 2025. № 1. С. 38–54. <https://doi.org/10.15407/pp2025.01.038>; Новицький О.В. Інтегроване середовище агрегації метаданих відкритих джерел для підтримки наукових досліджень. *Проблеми програмування*. 2024. Вип. 2–3. С. 408–417. <http://doi.org/10.15407/pp2024.02-03.408>; Проскудіна Г.Ю., Кудім К.О., Резніченко В.А. VuFind: відкрите рішення для інтеграції бібліотечних колекцій. *Проблеми програмування*. 2023. № 4. С. 15–26. <http://doi.org/10.15407/pp2023.04.01>

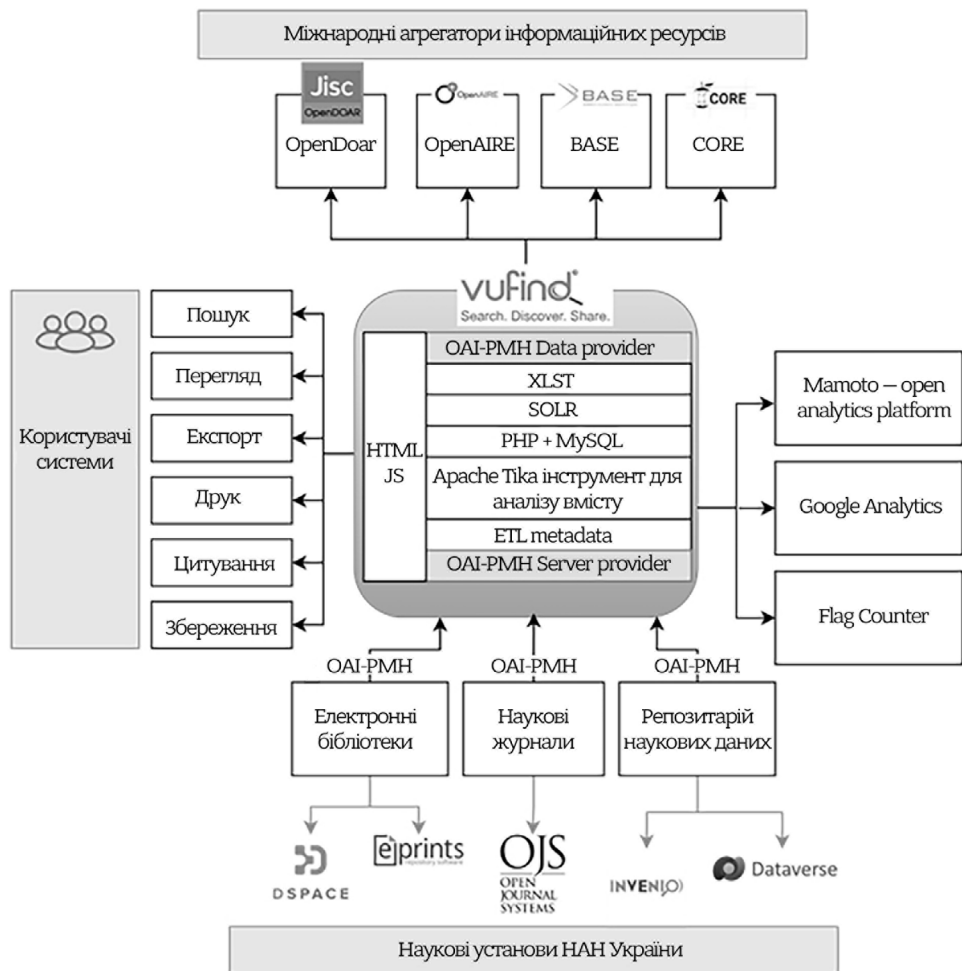


Рис. 8.5. Загальна архітектура Харвестеру

Для установ НАН України можливі два варіанти: автоматично без відома провайдера даних; за бажанням провайдера даних.

Автоматичний харвестинг. Установа НАН України самостійно вирішує питання про відкритість своїх інформаційних ресурсів в інтернеті та доступність до них, зокрема самі ресурси та їхні метадані. У зв'язку з цим Харвестер без будь-якого дозволу самостійно та автоматично проводить сканування відкритих інформаційних ресурсів установ НАН України, і за коректного представлення метаданих виконує їх імпорт, індексацію й надає можливість усім користувачам виконання пошуку та, у разі необхідності, переходу до репозитарію установи НАН України для можливого отримання самого інформаційного ресурсу. Існує можливість перевірити, чи відповідають дані,

що передаються, протоколу OAI-PMH. Підтримка функціонування програмно-технічних засобів Харвестеру та їх оновлення, а також підтримка його цілісності та інформаційної безпеки здійснюється Службою супроводу Харвестеру, яка також здійснює підключення до Харвестеру репозитаріїв провайдерів даних, формуючи перелік провайдерів даних установ НАН України та вебадрес «точок входу» для Харвестеру. Натепер імпортовано понад 320 тис. статей з 36 репозитаріїв.

Харвестінг за бажанням провайдера даних. Установа НАН України може самостійно ухвалити рішення про передання метаданих своїх інформаційних ресурсів у Харвестер і звернутись із відповідною заявкою до Служби супроводу. Процес отримання метаданих підтримує технологію ETL (Extract, Transform, Load), коли після отримання метаданих надається можливість їх редагування ще до завантаження у Харвестер. Для роботи з провайдерами даних створено реєстр, який містить майже 90 репозитаріїв періодичних видань установ НАН України, які підтримують DC-OAI-PMH.

8.5.1. ПОШУК І ПЕРЕГЛЯД ІНФОРМАЦІЙНИХ РЕСУРСІВ

Анонімним (незареєстрованим) користувачам надають усі можливості з пошуку, які передбачені Apache Lucene: звуження простору пошуку (за репозитарієм, автором, мовою, діапазоном дат, типом інформаційного ресурсу); надання умов пошуку на значення метаданих, відсікання слів, використання в словах групових символів, пошук по близькості звучання слова, пошук за фразами, пошук з використанням відстані між словами, пошук за важливістю слів або фраз, обов'язкова наявність слів або фраз, будування логічних пошукових виразів; результати пошуку можна переглянути, зберегти, надрукувати або експортувати в інші бібліографічні формати, помітити тегами для подальшого використання.

Зареєстровані користувачі можуть робити все те саме, що й анонімні користувачі, а також їм надаються додаткові функції. Вони можуть: переглядати історію своїх пошукових запитів і результатів, а також зберігати їх під своїм обліковим записом; створювати і редагувати списки вибраних документів, а також ділитися ними з іншими користувачами; додавати або вибирати свої власні теги для каталогізації чи пошуку ресурсів в Харвестері; залишати відгуки, коментарі та оцінки до документів, а також читати відгуки інших, надсилати електронною поштою, друкувати, цитувати.

Загальна схема пошуку та перегляду є такою (рис. 8.6): спочатку звужується простір пошуку з використанням фільтрів за: тематичними рубриками, репозитарієм, типом ресурсу, автором, мовою та

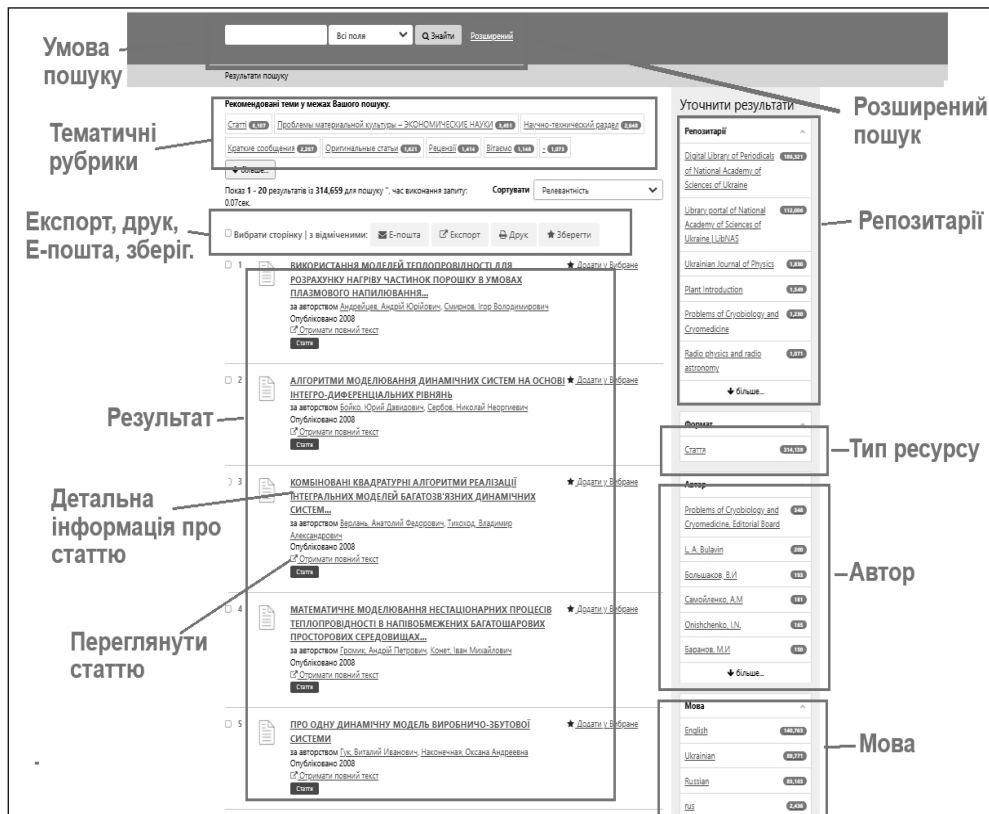


Рис. 8.6. Сторінка пошуку та перегляду інформаційних ресурсів

діапазоном дат. Потім задається умова пошуку та ініціюється пошук. Результат пошуку надається у стислому вигляді, який можна розгорнути. За потреби можна продивитись повний текст отриманого ресурсу, отримані результати можна експортувати, друкувати, надіслати та зберегти. Детальніше формулювання умов пошуку можна виконати на сторінці розширеного пошуку.

8.5.2. ПРЕДСТАВЛЕННЯ ХАРВЕСТЕРУ У МІЖНАРОДНИХ АГРЕГАТОРАХ

VuFind надає можливість передавати свої метадані іншим харвестерам і агрегаторам за протоколом OAI-PMH. Це надало можливість підключити Харвестер НАН України до найбільших у світі міжнародних агрегаторів електронних інформаційних ресурсів.

Харвестер у OpenDOAR. OpenDOAR – це глобальний каталог понад 2600 репозитаріїв відкритого доступу із гарантованою якістю,

які надають безкоштовний відкритий доступ до академічних результатів і ресурсів. Базується у Великій Британії.

Система OpenDOAR містить інформацію винятково про репозитарії, які повністю підтримують концепцію відкритого доступу до повних текстів документів. Репозитарії, які вимагають реєстрації або мають документи, доступ до яких обмежений чи платний, не заносяться у каталог, так само і ті, що містять тільки метадані.

Харвестер представлений у OpenDOAR за посиланням: <https://opendoar.ac.uk/repository/10991>

Харвестер у OpenAIRE. Це європейська електронна мережа сховищ, архівів і журналів відкритого доступу, які підтримують політики відкритого доступу. Охоплює 194 млн публікацій, 64 млн даних досліджень, 405 тис. програмного забезпечення для наукових досліджень.

OpenAIRE має на меті просувати відкриті наукові дослідження та покращувати доступність, можливість спільного використання, багаторазове використання, відтворюваність і моніторинг результатів досліджень на основі даних у всьому світі. Це європейська електронна інфраструктура, що пропонує різноманітний набір державних послуг для прискорення впровадження відкритої науки, підтримана мережею експертів, які працюють у ключових національних організаціях у європейських країнах. Серед користувачів послуг OpenAIRE є дослідники, дослідницькі спільноти, політики, наукові організації, університети та бібліотеки. OpenAIRE – це ключовий розробник Європейської хмари відкритої науки (EOSC).

Харвестер представлений у OpenAIRE за посиланням: https://explore.openaire.eu/search/dataprovider?datasourceId=openoar_____:35d58f24d7a7701719c983a1ef1979b1

Харвестер у BASE. Це одна з найбільших пошукових систем у світі, особливо для академічних вебресурсів. BASE управляє бібліотека Білефельдського університету (Німеччина).

Система індексує метадані всіх видів академічних ресурсів – журналів, інституційних репозитаріїв, цифрових колекцій тощо, які надають інтерфейс OAI і використовують протокол OAI-PMH для надання свого вмісту.

Харвестер представлений у BASE за посиланням: https://www.base-search.net/about/en/about_sources_date.php?&search_country=ua

Харвестер у CORE. Це агрегатор, що містить найбільшу у світі колекцію повних текстів дослідницьких робіт з відкритим доступом. Забезпечує єдину точку доступу до понад 314 млн статей відкритого доступу від більш ніж 11 000 постачальників контенту. CORE супроводжує Knowledge Media Institute, що базується у Відкритому університеті, Велика Британія.

Метою проекту є агрегування усього вмісту відкритого доступу, розподіленого в різних системах, як-от репозитарії та журнали відкритого доступу, збагачення цього вмісту за допомогою інтелектуального аналізу тексту та інтелектуального аналізу даних і забезпечення вільного доступу до нього. Проєкт CORE також спрямований на сприяння відкритому доступу до наукових результатів і тісно співпрацює з цифровими бібліотеками та інституційними репозитаріями.

Харвестер представлений у CORE за посиланням: <https://core.ac.uk/data-providers/22737/>

Статистика функціонування Харвестеру. Для отримання статистики відвідувань Харвестер підключений до статистичних сервісів Flag Counter, Matomo, Google Analytics.

Flag Counter надає статистику відвідувань по країнах за різні часові періоди. Харвестер представлений у Flag Counter за посиланням: <https://info.flagcounter.com/vb3w>

Matomo – найпоширеніший безкоштовний додаток вебаналітики з відкритим кодом для відстеження онлайн-відвідувань і відображення звітів про ці відвідування для аналізування. Є альтернативою Google Analytics. Використовує понад 460 000 сайтів, підтримує більш ніж 45 мов. Функціональні можливості: відстежування відвідувань сайту за їх типом (прямі переходи, переходи з сайтів, переходи з пошукових систем тощо); відстежування ключових слів, за якими відвідувачі перейшли на сайт; відстежування джерела трафіку; відстежування дій на сайті (скачування файлів, переходи за посиланням, тривалість відвідування та ін.); відстежування географії відвідувачів; автоматична звітність по електронній пошті у форматі PDF або HTML; представлення статистики в числовому та графічному вигляді; експорт даних у форматі CSV, XML, PHP, EXCEL, JSON.

Харвестер представлений у Matomo за посиланням: https://matomo.iitta.gov.ua/index.php?module=CoreHome&action=index&idSite=4&period=day&date=Takterday#?period=day&date=Takterday&category=Dashboard_Dashboard&subcategory=1

Google Analytics – сервіс від Google для аналізу інтернет-сайтів та мобільних додатків. Показує детальну статистику відвідувачів сайту та допомагає оптимізувати видимість вебсторінок. Є одним із найпоширеніших сервісів вебаналітики в інтернеті.

Харвестер представлений у Google Analytics за посиланням: https://analytics.google.com/analytics/web/#/p439363834/reports/intelligenthome?params=_u..nav%3Dmaui

За менше ніж за рік існування Харвестеру понад 40 000 читачів зі 110 країн світу відвідали його майже 154 000 разів.

8.5.3. РЕГЛАМЕНТ РОБОТИ ХАРВЕСТЕРУ

З метою визначення порядку роботи Харвестеру як елемента інфраструктури відкритої науки в НАН України було розроблено його регламент. У ньому визначено основні поняття, призначення, склад і структуру, засади організації та управління Харвестером, правила захисту від розголошення інформації з обмеженим доступом, дотримання прав інтелектуальної власності та прав на інформацію, процедуру застосування ліцензій відкритого доступу, використання персональних даних. Регламент має на меті допомогти установам та менеджерам інституційних репозитаріїв у налаштуванні збору метаданих за допомогою Харвестеру, а також забезпечити максимальну ефективність використання кожного наукового репозитарію, підключеного до нього.

Висновки. Вище розглянуто проблематику інтеграції електронних інформаційних ресурсів, яка є актуальною та важливою для забезпечення побудови автоматизованих інформаційних систем з автономними, розподіленими, неоднорідними джерелами інформації для забезпечення відкритого доступу до них на сучасному рівні.

Розглянуто підходи до вирішення проблеми побудови таких агрегаторів: детально розкрито підхід Ініціативи відкритих архівів, який отримав широке практичне застосування. Суть підходу полягає у забезпеченні єдиного уніфікованого інтерфейсу для вебдоступу до електронних інформаційних ресурсів, розміщених в інтероперабельних репозитаріях, за допомогою організації спільного використання, публікації й архівування (харвестингу) метаданих таких ресурсів.

Викладено загальну інформаційну модель підходу, розкрито ролі провайдерів даних і сервісів, описана схема їх взаємодії з агрегаторами тощо, представлена архітектура системи інтеграції наукових електронних бібліотек установ НАН України.

Виконано аналітичний огляд поширеного відкритого програмного забезпечення підтримки підходу до формування відкритих систем харвестингу.

Виходячи з вимог, сформульованих до інфраструктури відкритої науки в НАН України та поставлених завдань ЦНТП «OPENS», для побудови Харвестеру НАН України було обрано програмне забезпечення VuFind, широкі можливості якого детально розкрито у цьому розділі.

Розроблений і описаний тут Харвестер відкритої науки НАН України навіть за короткий час своєї роботи вже показав результативність, свідченням чого є не тільки входження до системи інших міжнародних агрегаторів наукової інформації, а й статистичні показники.