

О.В. Новицький

ІНТЕГРОВАНЕ СЕРЕДОВИЩЕ АГРЕГАЦІЇ МЕТАДАНИХ ВІДКРИТИХ ДЖЕРЕЛ ДЛЯ ПІДТРИМКИ НАУКОВИХ ДОСЛІДЖЕНЬ

Метадані для електронних бібліотек є невід'ємною частиною організації та доступу до цифрових ресурсів. У цій статті ми досліджуємо типові підходи до інтеграції електронних бібліотек, зосереджуючись на аналізі метаданих, протоколах та програмному забезпеченні. Розглянуті типи метаданих, такі як описові, структурні, адміністративні, метадані збереження, метадані прав та технічні метадані, розкривають різноманітні аспекти цифрових ресурсів. Розглянуто різні протоколи та підходи до інтеграції, зокрема, протокол OAI-PMH, який відіграє ключову роль у забезпеченні доступу та обміну метаданими між репозиторіями. Також надається огляд програмного забезпечення для інтеграції за допомогою протоколу OAI-PMH, включно із порівняльною таблицею характеристик популярних програмних рішень. Ця стаття допомагає розуміти важливість метаданих та ефективність їх використання для забезпечення доступу до цифрових ресурсів у сучасному науковому середовищі. Детально описано процеси екстракції, трансформації, завантаження, які можна застосовувати при семантичній інтеграції даних через OAI-PMH. Також проведено огляд програмного забезпечення для інтеграції електронних бібліотек та журналів відкритого доступу та визначено ключові функції для створення ефективного наукового середовища.

Ключові слова: інтеграція електронних бібліотек, протоколи обміну метаданими, oai-pmh, vufind, Dublin core.

О. Новицький

INTEGRATED OPEN SOURCE METADATA AGGREGATION ENVIRONMENT TO SUPPORT SCIENTIFIC RESEARCH

Metadata for electronic libraries is an integral part of the organization and access to digital resources. In this article, we explore common approaches to e-library integration, focusing on metadata analysis, protocols, and software. The types of metadata considered, such as descriptive, structural, administrative, preservation metadata, rights metadata, and technical metadata, reveal various aspects of digital resources. Various protocols and approaches to integration are considered, in particular the OAI-PMH protocol, which plays a key role in providing access and metadata exchange between repositories. An overview of software for integration using the OAI-PMH protocol is also provided, including a comparative table of characteristics of popular software solutions. This article helps to understand the importance of metadata and the effectiveness of its use to ensure access to digital resources in the modern scientific environment. The processes of extraction, transformation, loading, which can be used in the semantic integration of data through OAI-PMH, are described in detail. Software for the integration of electronic libraries and open access journals was also reviewed and key functions for creating an effective scientific environment were identified.

Keywords: integration of electronic libraries, metadata exchange protocols, oai-pmh, vufind, Dublin core.

Вступ

Метадані для електронних бібліотек є ключовим аспектом організації та надання доступу до цифрових ресурсів. Він передбачає створення, керування та застосування описової інформації про цифрові об'єкти, такі як книги, статті, зображення, аудіофайли та інший цифровий вміст. Проблема представлення та обміну метаданими між бібліотеками завжди актуальна, але

тривалий час вона не досягла значного прогресу. Своєрідний клас систем інтеграції представляють системи, в яких за основу взято технологію Ініціативи відкритих архівів (Open Archive Initiative – OAI) [1]. У більшості відомих систем цієї категорії їхні інформаційні ресурси представляють собою колекції текстових документів, передусім наукових публікацій, які автономно фо-

рмуються у вузлах глобальної мережі, підтримуються та адмініструються їхніми власниками. Агрегація метаданих для репозиторія здійснюється згідно зі спеціально розробленим протоколом Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) [2], що забезпечує глобальні послуги доступу та пошуку. Суть підходу відкритих архівів полягає у тому, щоб дозволити здійснювати веб-доступ до інформаційних ресурсів, розташованих у інтероперабельних репозиторіях, за допомогою організації спільного використання, публікації й архівування метаданих таких ресурсів. Протокол OAI-PMH надає провайдерам даних простий спосіб такого представлення їхніх метаданих, який робить їх доступними для провайдерів сервісів. Водночас як транспортний протокол використовуються HTTP, а як формат обміну даних XML. Проте в протоколі OAI-PMH для забезпечення базового рівня інтероперабельності специфіковано формат Дублінського ядра. Таким чином, метадані з різних неоднорідних джерел поєднуються в єдиній базі даних для того, щоб надати множину сервісів на основі таких агрегованих метаданих.

Концепція протоколу OAI-PMH виділяє дві ролі: провайдера даних та провайдера сервісів. Провайдер даних – це служба, що підтримує створення і ведення одного чи більше репозиторіїв (бази документів, архівів, електронних бібліотек), здійснює публікацію своїх ресурсів, а також уможливорює доступ до своїх метаданих для їхнього використання в інших системах. Провайдер сервісів здійснює збір і збереження метаданих, наданих провайдерами даних, для надання кінцевим користувачам різних сервісів. Тривалий час в Україні працював провайдер сервісів для електронних бібліотек на базі РКР Harvester <https://oai.org.ua/> [3]. РКР Open Archives Harvester (РКР OAI Harvester) є хорошим інструментом для збору метаданих з різних архівів через протокол OAI-PMH. Ця система дозволяє збирати метадані з цифрових репозиторіїв в Україні та індексувати 76 репозиторіїв, загальна кількість записів становить понад 630 тисяч. Сторінку списку електронних бібліотек показано на рис. 1.

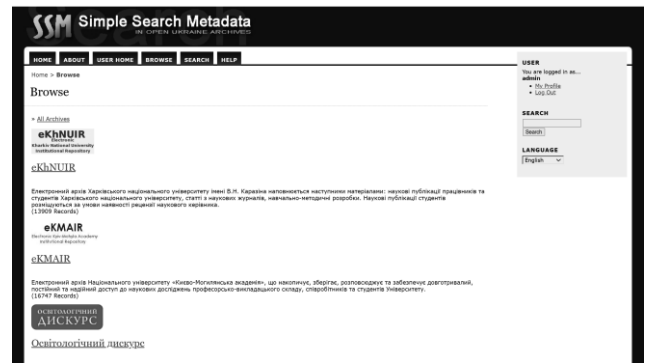


Рис. 1. Сторінка списку електронних бібліотек

РКР Harvester використовує мову програмування PHP 5.6, термін служби якої закінчився. Це створює проблему розвитку та розробки нових сервісів.

У рамках розробки програмних засобів підтримки наукових досліджень необхідно передбачити не тільки пошук метаданих, а й надати розширені можливості їх обробки та інтеграції з іншими системами. Водночас мета – створити систему, орієнтовану на роботу з українськими постачальниками даних. Це спонукає досліджувати сучасні методи та рішення для створення середовищ, які інтегрують цифрові бібліотеки.

Хоча багато проєктів використовують протокол OAI-PMH для інтеграції даних [4], в таблиці 1 наведено список порталів збору метаданих:

Таблиця 1

Перелік відкритих сервіс провайдерів метаданих

Назва та URL	Кількість ресурсів	Список основних сервісів
BASE base-search.net	340 М.	Title and abstract search
OAIster oaiSTER.on.worldcat.org	50 М.	title and abstract search, advanced metadata
CORE core.ac.uk	275М	title and abstract search, advanced metadata

Незважаючи на існування таких проєктів, електронні бібліотеки України не повною мірою представлені в цих агрегато-

рах. Зокрема, це також пов'язано з тим, що мова метаданих є переважно українською. Не всі цифрові бібліотеки належним чином надають багатомовні метадані. Зокрема, в одній з найбільших електронних бібліотек України Науковій електронній бібліотеці періодичних видань НАН України (Рис. 2) такі дані, як опис ресурсу, дублюються, але не вказано, якою мовою. Це створює проблему в обробці даних, джерело цієї проблеми буде розглянуто далі.

```
<?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet type="text/xsl" href="static/style.xml"?><OAI-PNH xmlns="http://www.request-xmlns.org/2003-06-08/190459" metadataPrefix="oai_dc" http://dspace.
<getRecord>
  <record>
    <header>
      <identifier>oai:dspace.nbuv.gov.ua:123456789/190459</identifier>
      <timestamp>2023-06-08T16:00:19Z</timestamp>
      <setSpec>com_123456789_190459</setSpec>
      <setSpec>com_123456789_172</setSpec>
      <setSpec>com_123456789_170</setSpec>
      <setSpec>com_123456789_21</setSpec>
      <setSpec>col_123456789_190333</setSpec>
    </header>
    <metadata>oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://www.lyncode
    <dc:title>Сходимость двухэтапного проксимального алгоритма для задачи о равновесии в пространствах Адамара</dc:title>
    <dc:creator>Вадець, Р.И.</dc:creator>
    <dc:creator>Сандраков, Г.В.</dc:creator>
    <dc:creator>Семенов, В.В.</dc:creator>
    <dc:creator>Найба, Р.И.</dc:creator>
    <dc:subject>Системный анализ</dc:subject>
    <dc:description>Рассмотрен итерационный двухэтапный проксимальный алгоритм для приближенного решения задач о равновесии в
    <dc:description>итерационно-проксимальном пространстве. Представлен алгоритм для приближенного решения задач о равновесии в
    <dc:date>2023-06-08T16:00:19Z</dc:date>
    <dc:date>2023-06-08T16:00:19Z</dc:date>
    <dc:date>2023-06-08T16:00:19Z</dc:date>
    <dc:type>Article</dc:type>
    <dc:identifier>Сходимость двухэтапного проксимального алгоритма для задачи о равновесии в пространствах Адамара / Р.И. Ва
    <dc:identifier>2023-5262</dc:identifier>
    <dc:identifier>http://dspace.nbuv.gov.ua/handle/123456789/190459</dc:identifier>
    <dc:identifier>http://dx.doi.org/10.15687/1728-1728/190459</dc:identifier>
    <dc:language>ru</dc:language>
    <dc:relation>Кибернетика и системный анализ</dc:relation>
    <dc:publisher>Институт кибернетики им. В.И. Глушкова НАН Украины</dc:publisher>
  </oai_dc:dc>
</metadata>
</record>
</getRecord>
</OAI-PNH>
```

Рис. 2. Дуплікація описових метаданих без ідентифікатора мови

У статті розглянуто типові підходи до інтеграції електронних бібліотек. Зокрема, проведено аналіз типів метаданих, наявних в електронних бібліотеках, розглянуто протоколи та методи інтеграції, а також проведено порівняльний аналіз програмних продуктів, призначених для інтеграції даних. У статтях [5], [6], [7], [8] було описано типові компоненти електронних бібліотек та основні проблеми, пов'язані з типовими підходами до побудови архітектури електронних бібліотек.

Як було зазначено, в основі інтеграції лежить обмін метаданими. Проте ці метадані бувають різними. Ось деякі з основних типів метаданих, які використовуються в цифрових бібліотеках [9], [10]:

– Описові метадані: описові метадані надають основну інформацію про ресурс, таку як його назва, автор, тема, ключові слова, анотація та дата публікації. Ці метадані відіграють більшу роль у розширеній архітектурі для обміну даними.

– Структурні метадані: структурні метадані описують внутрішню організацію та зв'язки між компонентами в цифровому

ресурсі. Він визначає послідовність, ієрархію та логічну структуру ресурсу, як-от розділів у книзі, розділів у статті, тому для журналу.

– Адміністративні метадані: адміністративні метадані включають інформацію про управління та адміністрування цифрових ресурсів. Вони містять відомості про права, дозволи, обмеження доступу, формати файлів, розміри файлів, технічні характеристики та інформацію про збереження.

– Метадані збереження важливі для збереження та архівування цифрових ресурсів. Вони містять інформацію, наприклад, звідки надійшов ресурс, у якому форматі файлу він знаходиться, контрольні суми для забезпечення його цілісності, історію міграції та інші технічні метадані. Ці метадані необхідні для гарантування того, що цифрові об'єкти залишатимуться автентичними та доступними з часом.

– Метадані прав: метадані прав визначають права інтелектуальної власності та дозволи на використання, пов'язані з цифровими ресурсами. Він містить інформацію про авторські права, умови ліцензування, обмеження використання та вимоги до посилання.

– Технічні метадані: Технічні метадані надають інформацію про технічні характеристики цифрових ресурсів. Він містить відомості про формати файлів, роздільну здатність, методи стиснення, колірні простори та інші технічні характеристики, необхідні для візуалізації, відтворення чи обробки цифрового вмісту.

– Метадані використання: метадані використання відстежують використання та взаємодію користувача з цифровими ресурсами. Вони містять таку інформацію, як кількість завантажень, переглядів, оцінки, коментарі та створений користувачами вміст, пов'язаний із певним ресурсом.

Усі ці типи метаданих можуть бути задіяні в інтеграції електронних бібліотек. Для інтеграції використовуються різноманітні протоколи та підходи, які дозволяють через єдину точку доступу отримувати доступ до метаданих. Метадані працюють ра-

зом, щоб забезпечити повний опис цифрових ресурсів у цифровій бібліотеці, забезпечуючи ефективний пошук, виявлення, доступ і керування ресурсами.

Протоколи обміну даними для електронних бібліотек

Однією з важливих проблем для електронних бібліотек є інтеграція даних [11]. Обмін даними в цифровій бібліотеці означає передачу інформації або ресурсів між різними системами, платформами або сховищами в екосистемі бібліотеки. Процес включає обмін, імпорт, експорт або синхронізацію даних для гарантування того, що колекція цифрової бібліотеки залишається актуальною, доступною та однорідною на різних платформах. Ось деякі загальні компоненти обміну даними в електронній бібліотеці:

- Обмін метаданими: метадані описують характеристики та атрибути цифрових ресурсів у стандартизованому форматі. Механізми обміну даними дозволяють обмінюватися метаданими між цифровими бібліотеками, дозволяючи їм знаходити, отримувати доступ до ресурсів з різних джерел. Такі формати метаданих, як Dublin Core, MARC (Machine-Readable Cataloging) або MODS (Metadata Object Description Schema), сприяють взаємодії та обміну даними.

- Збирання та агрегація: цифрові бібліотеки часто збирають вміст із різних джерел, таких як видавці, сховища чи інші бібліотеки. Збирання – це процес систематичного збору даних із цих джерел та імпортування їх до колекції бібліотеки. Такі технології, як OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), полегшують збір та обмін метаданими, забезпечуючи ефективну синхронізацію даних. Окрім Дублінського ядра чи Metadata Object Description Schema (MODS) для обміну метаданими використовуються схеми метаданих такі як Journal Article Tag Suite (JATS) [12] - це стандартизований формат розмітки для наукових статей у веб-публікаціях. JATS використовується для структурування та представлення метаданих, тексту, посилань та інших елементів, що сто-

суються наукових статей. Він забезпечує єдність формату, який полегшує обмін, інтеграцію та аналіз наукових даних між різними платформами та системами. JATS заснований на стандарті XML, що дозволяє легко обробляти та адаптувати дані для різних потреб в обробці та візуалізації. Фактично OAI-PMH є транспортним протоколом, який дозволяє обмінюватися метаданими між різними системами.

- Об'єднаний пошук: об'єднаний пошук дозволяє користувачам одночасно здійснювати пошук у кількох цифрових бібліотеках або сховищах і отримувати відповідні результати з кожного джерела. Обмін даними має вирішальне значення в сценаріях об'єднаного пошуку, оскільки він передбачає надсилання пошукових запитів до різних систем, отримання результатів пошуку та їх агрегування в уніфікований інтерфейс для користувачів. Це інший метод забезпечення єдиної точки доступу, за якого пошук здійснюється розподіленим чином. Типовим представником даного протоколу є Z39.50 або SRW (Search/Retrieve Web Service) [13]

- Збереження та архівування: цифрові бібліотеки часто співпрацюють із установами, що займаються збереженням, або архівними сховищами для забезпечення довгострокової доступності і цілісності своїх колекцій. Обмін даними в цьому контексті передбачає передачу ресурсів, метаданих або інформації, пов'язаної зі збереженням, до архівних систем, що забезпечує належне збереження та майбутній доступ. Існує набір стандартів, які забезпечують реалізацію довготривалого зберігання.

Розглянемо дещо детальніше протокол OAI-PMH, який є найпоширенішим протоколом для обміну метаданими в електронних бібліотеках відкритого доступу. Найпоширенішою схемою метаданих, що підтримується в OAI-PMH, є Dublin Core, яка надає базовий набір елементів для опису ресурсів.

Базуючись на тому факті, що протокол походить від електронної публікації, модель даних OAI-PMH зазвичай інтерпретується в термінах бібліографічних даних, що описують академічні ресурси, хоча також можливі інші інтерпретації [14]. OAI-

РМН має просту та гнучку модель даних
Рис. 3.

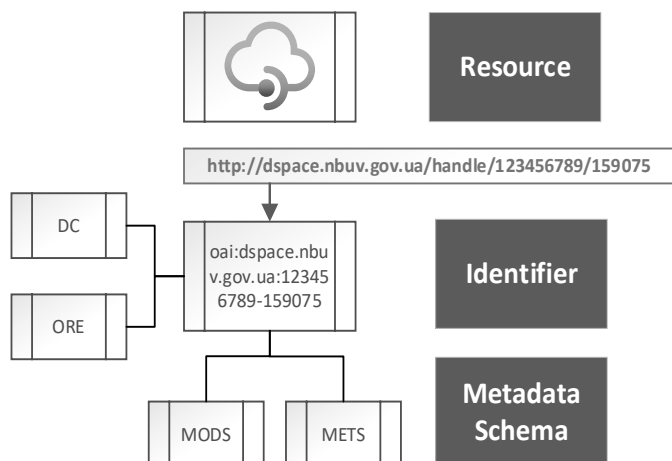


Рис. 3. Модель даних OAI-PMH опису ресурсу

Це може бути традиційний бібліотечний об'єкт (наприклад, книга, стаття), а також інші сутності (наприклад, зображення, концепції). Далі - ідентифікатор запису (ідентифікатор OAI) або шлюз до всіх метаданих, що описують ресурс. Нарешті у нижній частині моделі даних є записи. Записи описують ресурс у будь-якому форматі метаданих, який може бути виражений у схемі XML. Протокол OAI-PMH розроблено для підтримки будь-якої схеми опису метаданих, але основною, включеною в OAI-PMH, є набір метаданих Dublin Core (DC), що описує ресурс. Також бажано включити в опис більші набори метаданих (наприклад, MARC).

На відміну від OAI-PMH, протокол OAI-ORE фокусується виключно на створенні моделі даних, а не на визначенні протоколу обміну. Він пропонує потенційні формати обміну, наприклад XML/RDF. У моделі OAI-ORE існує чотири різні типи ресурсів: агрегація розроблена як ресурси для групування інших ресурсів, які називаються агрегованими ресурсами. Агрегований ресурс — це ресурс, який є частиною агрегації, що представляє інформаційний об'єкт у складеному об'єкті за стандартами ORE; карта ресурсів — це серіалізоване зображення агрегації, містить перераховані агреговані ресурси та властивості щодо агрегації та її агрегованих ресурсів,

включаючи зв'язки із зовнішніми ресурсами [15], [16].

Підхід до інтеграції даних

З огляду на вищеописану концепцію архітектури OAI-PMH, у цій статті пропонується семантичний підхід для інтеграції даних у ЕБ [17]. Як було зазначено, сам протокол не передбачає ніяких механізмів для семантичної інтеграції даних. Пропонується застосувати наступні відомі підходи до інтеграції семантичних даних у поєднанні з OAI-PMH:

1. Зіставлення метаданих: визначте зіставлення між схемами метаданих, що використовуються в різних сховищах. Цей процес відображення вирівнює семантику відповідних елементів у різних схемах, забезпечуючи значущу інтеграцію. Наприклад, ви можете відобразити елементи Dublin Core у більш специфічну онтологію, що використовується в цільовому сховищі.

2. Відображення онтології: використовуються існуючі онтології, щоб забезпечити загальний словник для інтеграції даних. Інструменти та методи відображення онтологій можна використовувати для виявлення відповідностей і співставлення онтологій.

3. Зв'язані дані. Використання принципу зв'язаних даних для з'єднання та зв'язування наборів даних між сховищами. Зв'язані дані дозволяють встановлювати явні зв'язки між ресурсами за допомогою стандартних протоколів, таких як RDF.

Важливо відзначити, що семантична інтеграція даних є складним завданням і вимагає ретельного розгляду знань предметної області, моделей даних і стратегій інтеграції. Хоча OAI-PMH не може безпосередньо вирішити проблему семантичної інтеграції, він може бути будівельним блоком для полегшення обміну метаданими. Його можна покращити за допомогою додаткових методів для досягнення глибшої інтеграції. Процеси програмного забезпечення, які полегшують інтеграцію метаданих, широко відомі як процеси вилучення-перетворення-завантаження (ETL). ETL (Extract, Transform, Load) означає процес

Таблиця 2

Перелік сучасного ПЗ для інтеграції ЕБ та журналів

ПЗ	VuFind	DSpace	Omeka S
Опис	VuFind — це платформа підтримує збір та інтеграцію OAI-PMH сумісних репозиторіїв, що дозволяє збирати метадані з багатьох джерел і забезпечувати уніфікований пошук.	DSpace — це платформа цифрового сховища, яка підтримує протокол OAI-PMH.	Omeka S — це веб-платформа для створення цифрових колекцій. Включає підтримку OAI-PMH, що дозволяє обмінюватися метаданими з іншими OAI-PMH-сумісними системами.
Технології	PHP, MySQL, SORL	JAVA, MySQL	PHP, MySQL
ETL	Так	Ні	Ні
Підтримка інтегрованих бібліотечних систем	Aleph, Alma, DAIA, Evergreen, Folio, GeniePlus, Horizon, Innovative, Koha, KohaLSDI, KohaRest, NewGenLib, NoILS, PAIA, Polaris, SierraRest, Symphony, Unicorn, Virtua, Voyager,	Ні	Ні
Фасетна навігація	Так	Так	Так
Фільтрація отриманих записів/фільтрація пошуку	Так	Так	Так
Система рекомендацій в інтерфейсі користувача	Так	Ні	Так
Наявність механізму зв'язування з повнотекстовим екст-рактором	Так	Ні	Ні

вилучення даних із різних джерел, перетворення їх у потрібний формат і завантаження в цільову систему чи сховище даних. Модель процесу ETL виглядає наступним чином:

Екстракція (E): можна визначити як комплексний пошук усіх даних, що містяться у вихідній системі, не залишаючи жодних даних неврахованими чи пропущеними під час процесу екстракції.

Трансформація (T) означає зміну структури даних на нову: очищення даних, що включає видалення або виправлення помилок, невідповідностей, дублікатів і відсутніх значень у даних; перевірка даних на цілісність, якість та узгодженість даних із попередньо визначеними правилами або бізнес-логікою; збагачення даних, що включає покращення даних шляхом додавання додаткової інформації, похідних атрибутів або обчислених значень на основі бізнес-правил або зовнішніх джерел даних; фільтрування даних: вибір або виключення певних даних на основі попередньо визначених критеріїв.

Завантаження (L): цей процес включає в себе завантаження перетворених даних безпосередньо в цільову систему або базу даних за допомогою власних механізмів завантаження або API.

Дані процеси доречно застосувати у вирішенні проблеми семантичної інтеграції в рамках протоколу OAI-PMH.

Огляд програмного забезпечення для інтеграції даних за протоколом OAI-PMH

Кілька програмних рішень доступні для інтеграції OAI-PMH, що дозволяє організаціям збирати та показувати метадані з різних сховищ. Ми зробили короткий аналіз популярного програмного забезпечення для створення комбайна OAI-PMH і визначили ключові функції для створення ефективного наукового середовища. Під час формування списку було висунуто наступні вимоги: відкритий код; регулярні оновлення системи; час існування ПЗ на ринку понад 10 років. Порівняльна таблиця характеристик популярних ПЗ для інтеграції електронних бібліотек OAI-PMH.

Повнотекстовий пошук	Так	Так	Так
Нечіткий пошук	Так (Sorl)	Так (Sorl)	Так (Elasticsearch)
Ролі користувачів	Так	Так	Так
Авторизація LDAP	Так	Так	Так
Створення титульної сторінки	Так	Ні	Ні
DOI	Так	Так	Так
EZргоху	Так	Так	Так
Правопис для пошуку	Так	Так	Ні
Експорт	RefWorks, EndNote, EndNoteWeb, MARC, MARCXML, RDF, BibTeX, RIS, OAI	Linked Open Data, SWORD, OAI	Json, XML, Spreadsheet, Bibtex, Csl, Ris
Інтерфейс на основі конфігурації	Так	Обмежена	Ні
Багатомовна підтримка для метаданих	Обмежена	Так	Ні
Матомо аналітика	Так (з коробки)	Так	Так (з коробки)
API підтримка	REST API	REST API	REST API
Підтримувані схеми метаданих (Імпорт та перегляд)	Dublin Core, METS, Dublin Core Terms, MARC, XML, CSV	Dublin Core, Dublin Core Terms	Dublin Core, METS
Редагування метаданих	Ні	Так	Так
Веб-інтерфейс для управління ресурсами	Ні	Так	Так
Автодоповнення при пошуку	Так	Ні	Ні

Оцінюючи характеристики, слід розуміти, що порівнювати ряд параметрів

дуже складно. Наприклад, у Vufind [18], [19], архітектура системи виконана таким чином, управління відображенням метаданих ресурсу повністю контролюється за допомогою змін у темі. VuFind — це програмне забезпечення для створення порталу бібліотечних ресурсів, основною метою якого є покращення взаємодії з користувачем шляхом перетворення традиційного онлайн-каталогу публічного доступу (OPAC) [20]. Ця платформа є бібліотечною пошуковою системою з відкритим вихідним кодом і розроблена бібліотекою університету Вілланова, яка вперше була стабільно опублікована для громадськості в 2010 році. Програмна архітектура цього програмного продукту реалізована дуже вдало завдяки орієнтованому на розробника набору інструментів, фреймворку Laminas і великій кількості системних налаштувань. Це дозволяє змінити структуру метаданих, які потрібно показати користувачеві без необхідності змінювати програмний код системи. Правила форматування об'єкта контролюються з коду теми, таким чином ми встановлюємо правила, для яких метаданих слід використовувати та які методи системи будуть відповідати за отримання даних. Це надсилається на серверну частину і після обробки результат повертається до інтерфейсу, де відображається для користувача. Тобто в архітектурі системи дані та правила форматування цих даних розділені. Це дуже зручно для кастомізації. Розглянемо більш детально, як реалізований метод ETL у Vufind. Фактично в даному програмному продукті процес перетворення даних розділений на два етапи. У процесі отримання даних відбувається початкова трансформація метаданих для зміни ідентифікатора запису з архіву. Це пов'язано з тим, що ідентифікатори мають бути унікальними і, з іншого боку, структура ідентифікатора не повинна містити похилих рисок. Оскільки кожен ресурс має URL-адресу, яка відповідає його ідентифікатору в основній електронній бібліотеці, яка є джерелом метаданих. Власне на основі таких переваг даних програмний продукт був обраний за основу для побудови інтегрованого середовища для підтримки наукових досліджень.

Результат розгортання та індексації наукової електронної бібліотеки періодичних видань НАН України представлено на Рис 4.

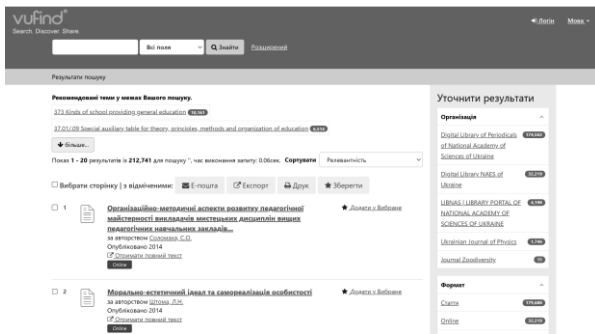


Рис. 4. Інтерфейс списку ресурсів з факетним фільтром

Інтеграція семантичних даних не передбачена в VuFind, але її можна досягти за допомогою функцій користувача, які можуть виконувати відображення семантичних даних. Однією з переваг VuFind є можливість використовувати такі виклики. Процес інтеграції та організації доступу до інформації в VuFind складається з таких етапів: 1) збір метаданих за допомогою протоколу OAI-PMH; 2) перетворення даних відповідно до моделі ETL. На етапі екстракції VuFind дозволяє здійснювати частково операції трансформації. Цей процес дозволяє VuFind створювати уніфікований і повний індекс ресурсів із багатьох джерел, надаючи користувачам можливість централізованого пошуку; 3) надання користувацького пошуку на агрегованих даних за допомогою зручного інтерфейсу з глибокою доступністю для конфігурації; 4) доступ до ресурсу. Кожен ресурс має доступ безпосередньо через надані посилання, включаючи необхідні ідентифікатори (наприклад, URL-адреси або DOI) до повного вмісту, розміщеного оригінальними постачальниками даних; 5) відображення метаданих представляє зібрані метадані у стандартизованому ETL процесі та зручному для користувача вигляді. Це може збагатити метадані додатковою інформацією або аспектами для покращення пошуку та допомоги у виявленні ресурсів.

Висновки

Побудова інтегрованого середовища для агрегації наукових ресурсів потребує вирішення низки проблем. У статті розглянуто підходи до інтеграції електронних архівів, а також описано досвід практичного застосування інтеграції електронних архівів України за протоколом OAI-PMH.

Побудова інтегрованого середовища для агрегації наукових ресурсів потребує вирішення низки проблем. У статті розглянуто підходи до інтеграції електронних архівів України за протоколом OAI-PMH. Розглянуто основні протоколи інтеграції електронних бібліотек. Як показав аналіз за останні 10 років, жодного значного протоколу обміну, альтернативного OAI-PMH, не з'явилося. Проаналізовано підходи до структурної інтеграції електронних бібліотек та здійснено порівняльний аналіз функціональних можливостей кожного програмного забезпечення. Було показано, що VuFind є найефективнішим інструментом для інтеграції цифрових бібліотек.

VuFind — це платформа з відкритим вихідним кодом, призначена для надання доступу до різноманітних бібліотечних ресурсів і ресурсів культурної спадщини, включаючи книги, журнали, статті, цифрові колекції тощо. Однією з його ключових особливостей є підтримка протоколу OAI-PMH, що означає Open Archives Initiative Protocol for Metadata Harvesting.

Література

1. H. Van de Sompel, M. Nelson, C. Lagoze и S. Warner, «Resource harvesting within the OAI-PMH framework,» D-lib magazine, № 10, 2004.
2. "The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14," [Online]. Available: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
3. V. O. Kopanieva, L. I. Kostenko, O. V. Novytskyi та V. A. Reznichenko, «The task of digital transformation of the scientific information environment,» Problems in programming, т. 1, pp. 3-10, 2023.

4. S. Roy, B. Sutradhar та P. Das, «Large-scale Metadata Harvesting—Tools, Techniques and Challenges: A Case Study of National Digital Library (NDL),» *World Digital Libraries: An International Journal.*, т. 10, 2017.
5. R. Gartner, *Metadata for digital libraries: state of the art and future directions*, JISC, 2008.
6. A. Getaneh, B. Stevens та P. Ross, «Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach,» *New Library World*, т. 113, pp. 38-54, 2012.
7. К. Лобузiна, «Сучаснi пiдходи до iнтеграцiї електронних iнформацiйних ресурсiв бiблiотек,» *Вiсник Книжкової палати*, т. 12, pp. 24-28, 2012.
8. О. М. Спiрiн, С. М. Иванова, О. В. Новицький, З. Савченко, В. А. Резниченко, А. В. Яцишин, Н. М. Андрiйчук та В. Ткаченко, *Електроннi бiблiотечнi iнформацiйнi системи наукових i навчальних закладiв.*, Педагогiчна преса, 2012.
9. J. Pomerantz, *Metadata*, MIT Press Essential Knowledge series, 2015.
10. W. M. Beyene, «Metadata and universal access in digital library environments,» *Library Hi Tech*, т. 35, № 2, pp. 210-221, 2017.
11. M. Agosti, N. Ferro та G. Silvello, «Digital library interoperability at high level of abstraction,» *Future Generation Computer Systems*, т. 55, pp. 129-146, 2016.
12. National Center for Biotechnology Information, U.S. National Library of Medicine, «Journal Article Tag Suite,» 2024. [Онлайн-вий]. Available: <https://jats.nlm.nih.gov/>. [Дата звернення: 10 2024].
13. . A. S. Lingam, «Federated search and discovery solutions,» *IP Indian J. Libr. Sci. Inf. Technol.*, Томи 1-2 January-June 5, № 1, pp. 39-42, 2020.
14. C. Lagoze та H. Van de Sompel, «The Open Archives Initiative Protocol for Metadata Harvesting,» 2015. [Онлайн-вий]. Available: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
15. C. Lagoze та H. Van de Sompel, «ORE User Guide - HTTP Implementation,» [Онлайн-вий]. Available: <https://www.openarchives.org/ore/1.0/http>. [Дата звернення: 2023].
16. C. Lagoze та H. Van de Sompel, «ORE User Guide - Resource Map Implementation in RDF/XML,» [Онлайн-вий]. Available: <https://www.openarchives.org/ore/1.0/rdfxml>. [Дата звернення: 2023].
17. В. А. Резниченко, О. В. Новицкий та Г. Ю. Проскудина, «Интеграция научных электронных библиотек на основе протокола OAI-PMH,» *Проблемы програмування*, № 2, pp. 97-112, 2007.
18. Villanova University's Falvey Library., «VuFind® - Search. Discover. Share.,» [Онлайн-вий]. Available: <https://vufind.org/>. [Дата звернення: 2023].
19. D. Katz, R. LeVan та Y. Ziso, «Using authority data in VuFind,» *Code4Lib Journal*, т. 14, 2011.
20. H. Yu та M. Young, «The impact of web search engines on subject searching in OPAC,» *Information technology and libraries*, т. 4, № 23, pp. 168-180, 2004.

References

1. H. Van de Sompel, M. Nelson, C. Lagoze и S. Warner, «Resource harvesting within the OAI-PMH framework,» *D-lib magazine*, № 10, 2004.
2. "The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14," [Online]. Available: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
3. V. O. Kopanieva, L. I. Kostenko, O. V. Novytskyi and V. A. Reznichenko, "The task of digital transformation of the scientific information environment," *Problems in programming*, vol. 1, pp. 3-10, 2023.
4. S. Roy, B. Sutradhar and P. Das, "Large-scale Metadata Harvesting—Tools, Techniques and Challenges: A Case Study of National Digital Library (NDL)," *World Digital Libraries: An International Journal.*, vol. 10, 2017.
5. R. Gartner, *Metadata for digital libraries: state of the art and future directions*, JISC, 2008.
6. A. Getaneh, B. Stevens and P. Ross, "Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach," *New Library World*, vol. 113, pp. 38-54, 2012.
7. К. Лобузина, "Suchasni pidkhody do intehratsiyi elektronnykh informatsiynykh resursiv bibliotek," *Visnyk Knyzhkovoyi palaty*, vol. 12, pp. 24-28, 2012.
8. О. М. Спiрiн, С. М. Иванова, О. В. Новицький, З. Савченко, В. А. Резниченко, А. В. Яцишин, Н. М. Андрiйчук та В. Ткаченко, *Електроннi бiблiотечнi iнформацiйнi системи наукових i навчальних закладiв.*, Педагогiчна преса, 2012.

- matsiyni systemy naukovykh i navchalnykh zakladiv., Pedagogichna presa, 2012.
9. J. Pomerantz, Metadata, MIT Press Essential Knowledge series, 2015.
 10. W. M. Beyene, "Metadata and universal access in digital library environments," Library Hi Tech, vol. 35, no. 2, pp. 210-221, 2017.
 11. M. Agosti, N. Ferro and G. Silvello, "Digital library interoperability at high level of abstraction," Future Generation Computer Systems, vol. 55, pp. 129-146, 2016.
 12. National Center for Biotechnology Information, U.S. National Library of Medicine, "Journal Article Tag Suite," 2024. [Online]. Available: <https://jats.nlm.nih.gov/>. [Accessed 10 2024].
 13. . A. S. Lingam, "Federated search and discovery solutions.," IP Indian J. Libr. Sci. Inf. Technol., Vols. January-June 5, no. 1, pp. 39-42, 2020.
 14. C. Lagoze and H. Van de Sompel, "The Open Archives Initiative Protocol for Metadata Harvesting," 2015. [Online]. Available: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
 15. C. Lagoze and H. Van de Sompel, "ORE User Guide - HTTP Implementation," [Online]. Available: <https://www.openarchives.org/ore/1.0/http>. [Accessed 2023].
 16. C. Lagoze and H. Van de Sompel, "ORE User Guide - Resource Map Implementation in RDF/XML," [Online]. Available: <https://www.openarchives.org/ore/1.0/rdfxml>. [Accessed 2023].
 17. V. A. Reznichenko, O. V. Novytskyy and H. Yu. Proskudina, "Intehratsiya naukovykh elektronnykh bibliotek na osnovi protokolu OAI-RMN," Problemy prohramuvannya, no. 2, pp. 97-112, 2007.
 18. Villanova University's Falvey Library., "VuFind® - Search. Discover. Share.," [Online]. Available: <https://vufind.org/>. [Accessed 2023].
 19. D. Katz, R. LeVan and Y. Ziso, "Using authority data in VuFind," Code4Lib Journal, vol. 14, 2011.
 20. H. Yu and M. Young, "The impact of web search engines on subject searching in OPAC," Information technology and libraries, vol. 4, no. 23, pp. 168-180, 2004.
- Одержано: 12.02.2024
Внутрішня рецензія отримана: 19.02.2024
Зовнішня рецензія отримана: 08.03.2024

Про авторів:

¹Новицький Олександр Вадимович,
Кандидат технічних наук,
науковий співробітник.
<https://orcid.org/0000-0002-9955-7882>.

Місце роботи авторів:

¹Інститут програмних систем
НАН України,
тел. +38 (044) 526-33-19
E-mail: alex.google@gmail.com
<https://iss.nas.gov.ua/>