

Заїка Віктор Миколайович,

ORCID <https://orcid.org/0009-0003-6582-6524>,

кандидат фізико-математичних наук,

завідувач відділу,

відділ інформаційно-комунікаційних технологій,

Інститут інформаційних технологій,

Національна бібліотека України імені В. І. Вернадського,

Київ, Україна

e-mail: zayika@nbuv.gov.ua

Технологічні підходи до створення цифрових репозитаріїв

Розглянуто підходи до створення цифрових репозитаріїв та вибору їхньої технологічної основи. Проаналізовано можливості використання спеціалізованих платформ і розширення діючих інформаційних систем, зокрема автоматизованих бібліотечно-інформаційних систем. Актуалізовано питання довготривалого цифрового збереження, інтеграції даних, використання метаданих, протоколу OAI-PMH та застосування принципів FAIR. Узагальнено, вибір технологічного рішення має визначатися потребами установи, наявною інфраструктурою, складністю інтеграції та можливостями довготривалої підтримки репозитарію.

Ключові слова: платформи цифрового репозитарію, Islandora, цифрові об'єкти, довготривале збереження, метадані, OAI-PMH, FAIR, відкрита наука.

Ідея доповіді виникла після участі у програмі підвищення кваліфікації з метаданих, стандартів оцифрування та управління цифровими колекціями, організованій Представництвом ЮНЕСКО в Україні та ВГО «Українська бібліотечна асоціація» в межах виконання проєкту «Цифрова колекція українсько-єврейської документальної спадщини». Офлайн-частина програми, серед іншого, передбачала практичну роботу з платформою цифрового репозитарію

Islandora, формування та контроль якості метаданих. Участь у навчальній програмі спонукала авторів замислитися над тим, чим саме є платформа цифрового репозитарію, які функції вона має забезпечувати та чим платформа відрізняється від репозитарію, створеного установою самостійно.

Цифрові репозитарії є важливою складовою інформаційної інфраструктури бібліотек, архівів та інших установ, що створюють, зберігають і надають доступ до цифрових ресурсів. Їхнє призначення не обмежується зберіганням окремих файлів. До функцій репозитарію належить забезпечення ідентифікації та опису цифрових об'єктів, робота з метаданими, пошук, керування доступом, встановлення зв'язків між об'єктами та надання доступу до цифрових ресурсів користувачам.

Існують спеціалізовані програмні платформи цифрових репозитаріїв, які мають готові компоненти для зберігання цифрових об'єктів, роботи з метаданими, пошуку, керування доступом та обміну даними. Їх використання дає змогу зменшити обсяг власної розробки. Одним із прикладів є Islandora.

Зауважимо, поняття платформи та цифрового репозитарію не є тотожними. Спеціалізована платформа є програмною основою, на якій створюється цифровий репозитарій, тоді як власне репозитарій являє собою сукупність програмних засобів, що забезпечують необхідну функціональність, а також, даних і правил, які визначають порядок зберігання цифрових об'єктів та їхніх метаданих, опису їх та взаємозв'язків, пошуку та надання доступу до них. Використання спеціалізованої платформи це один із способів створення цифрового репозитарію. Чи завжди цифровий репозитарій створюється на базі спеціалізованої платформи, або ж, за певних обставин, його функції можуть бути реалізовані на основі вже наявної інформаційної системи?

Цифровий репозитарій насамперед виконує функцію довготривалого збереження цифрових об'єктів. Цифровим об'єктом може бути окремий файл або сукупність файлів, доповнених метаданими та іншою інформацією, необхідною для їх ідентифікації, опису та використання. Тому збереження цифрового

об'єкта передбачає збереження не лише його вмісту, а й інформації, необхідної для його подальшого розуміння та використання.

Довготривале цифрове збереження (digital preservation) охоплює організаційні та технічні заходи, спрямовані на те, щоб цифрові об'єкти залишалися доступними, цілісними, автентичними та придатними для використання в умовах технологічних змін. Це передбачає захист даних від втрати та пошкодження, контроль їхньої цілісності, збереження інформації про походження та зміни об'єктів, а також управління ризиками, пов'язаними зі зміною форматів, програмного та апаратного забезпечення.

Цифровий репозитарій є одним із технологічних засобів реалізації цих завдань. Він має забезпечувати зберігання цифрових об'єктів і метаданих, їхню ідентифікацію, контроль цілісності, фіксацію походження та історії операцій. Важливою є також можливість експорту об'єктів разом із необхідними метаданими та іншою пов'язаною інформацією. Це створює умови для перенесення даних до іншої системи у разі зміни технологічної інфраструктури.

Для належної реалізації довготривалого цифрового збереження функцій лише програмного забезпечення не достатньо: необхідні також визначені установою правила збереження, процедури виконання відповідних операцій, технічна інфраструктура, призначена для надійного зберігання та контролю даних, а також визначені фахівці, відповідальні за виконання цих завдань. Тому під час вибору технологічної основи репозитарію необхідно враховувати не лише його поточну функціональність, а й можливість тривалої підтримки відповідних процесів.

Спеціалізована платформа дає змогу реалізувати основні функції репозитарію без створення всієї системи «з нуля». Наприклад, Islandora використовує Drupal як основу для вебінтерфейсу та керування контентом, Fedora – для зберігання цифрових об'єктів і пов'язаних із ними даних, а Solr – для їх індексації та пошуку. Інші програмні компоненти забезпечують обробку та представлення цифрових об'єктів. Архітектура платформи є модульною, тому конкретний набір компонентів та їх конфігурація можуть залежати від певної реалізації.

Розгортання платформи передбачає налаштування взаємодії її компонентів, адміністрування, оновлення та відповідну технічну інфраструктуру.

Вибір платформи є водночас вибором певної технологічної архітектури та пов'язаних із нею вимог до ресурсів установи. Чим ширшою є функціональність платформи, тим більше зусиль може потребувати її розгортання, інтеграція та подальший супровід.

Труднощі можуть виникнути під час інтеграції репозитарію з уже наявними інформаційними системами, особливо якщо дані установи вже зберігаються в спадковій автоматизованій бібліотечно-інформаційній системі (АБІС). У такому випадку доцільно використовувати вже створені записи як джерело метаданих для цифрових об'єктів репозитарію. Для цього необхідно забезпечити відповідність між записами АБІС і цифровими об'єктами та визначити спосіб обміну даними. За відсутності відповідних програмних інтерфейсів або стандартних механізмів обміну це може потребувати написання додаткових скриптів.

Альтернативою є повторне створення описів безпосередньо в середовищі репозитарію. Однак якщо відповідні дані вже існують та підтримуються в АБІС, такий підхід означає дублювання виконаної роботи. Фахівці установи витрачають робочий час не на створення нової інформації, а на повторне введення та підтримку ще однієї системи.

Тому під час вибору технологічного рішення варто враховувати не лише можливість технічної інтеграції систем, а й спосіб використання вже наявних даних. Повторне введення інформації може бути виправданим лише тоді, коли її перенесення або обмін між системами є технічно чи організаційно недоцільним. Робочий час фахівців є ресурсом, який, як і технічна інфраструктура, має використовуватися раціонально.

Тому під час вибору архітектури репозитарію доцільно передбачити можливість приймання вже створених метаданих разом із цифровими об'єктами у структурованому форматі обміну. XML, зокрема, може використовуватися як формат експорту даних з бібліотечної системи та їх подальшого імпорту в ре-

позитарій. При такому підході розмежовуються процеси створення та введення даних і процеси їх публікації та поширення через репозитарій. Тобто, оцінюючи перспективи впровадження окремої платформи необхідно враховувати не лише її функціональні можливості, а й те, наскільки ефективно вона може використовувати уже створену інформацію.

Використання певної платформи не є єдиним можливим рішенням. Іншим варіантом може бути розширення вже наявної інформаційної системи функціями, необхідними для роботи з цифровими об'єктами. Такий підхід може зменшити складність інтеграції, оскільки дані та функції роботи з цифровими ресурсами реалізуються в межах одного інформаційного середовища.

Прикладом такого підходу є цифровий репозитарій Національної бібліотеки України імені В. І. Вернадського, створений на основі ISIS-сумісної бібліотечної системи, що доступний за адресою <https://irbis-nbuv.gov.ua/everlib>. Репозитарій реалізовано як вебсайт, призначений для надання відкритого доступу до інтелектуальних продуктів НБУВ засобами інтернет-технологій, поширення цих матеріалів та інтеграції у середовище світового науково-освітнього товариства. Для опису та ідентифікації наукових робіт використовуються бібліографічні записи діючої АБІС. Файли повнотекстових документів зберігаються окремо у файловому сховищі, а їх накопичення здійснюється шляхом використання FTP-сервера та клієнтського програмного забезпечення. Вебсайт забезпечує функціональний пошук та надає доступ до повних текстів.

Розглянутий приклад характеризується обмеженими можливостями і не претендує на безпосереднє порівняння з повнофункціональною платформою цифрового репозитарію. Його значення полягає в іншому: продемонструвати, що за відносно простих вимог функції цифрового репозитарію можуть бути реалізовані шляхом розширення наявної інформаційної системи та використання вже створених бібліографічних даних.

Власна реалізація репозитарію зовсім не означає його ізолюваність від зовнішньої інформаційної інфраструктури. Якщо репозитарій має бути її частиною, на етапі проектування передбачаються стандартизовані способи обміну

даними: формати метаданих, стійкі ідентифікатори, програмні інтерфейси та відповідні протоколи. Одним із поширених протоколів обміну метаданими є OAI-PMH, який дає змогу зовнішнім сервісам автоматично отримувати та індексувати метадані репозитарію. Таким чином репозитарій, створений на основі нестандартної або спадкової інформаційної системи, може взаємодіяти із зовнішніми ресурсами за умови реалізації необхідних механізмів обміну. В якості ілюстрації можна привести репозитарій наукових текстів НАН України та одну з його головних складових – електронний архів академічної періодики (<http://jnas.nbu.gov.ua/>).

Під час створення репозитарію слід враховувати не лише внутрішню функціональність, а й можливість подальшого використання отриманих даних. Це важливо в контексті відкритої науки та принципів FAIR, відповідно до яких цифрові об'єкти та їхні метадані мають бути придатними для знаходження, доступу, машинної обробки та повторного використання. Саме для цього потрібні стійкі ідентифікатори, структуровані та стандартизовані метадані, визначені механізми доступу, а також інформація про походження об'єктів, права та умови їх використання. Реалізація згаданих вимог не потребує обов'язкового використання спеціалізованої платформи. Репозитарій, створений на основі діючої бібліотечної системи, також може забезпечувати інтероперабельність і відповідати принципам FAIR, якщо його дані мають узгоджену структуру та можуть передаватися іншим системам у стандартизованому форматі. Наприклад, метадані можуть експортуватися у форматі XML та надаватися зовнішнім сервісам через згаданий вище протокол OAI-PMH. Відповідність репозитарію вимогам ініціативи відкритої науки визначається передусім не технологічною основою, а тим, як організовані дані та яким чином забезпечуються їх опис, доступ, обмін і повторне використання.

Спосіб створення цифрового репозитарію обумовлюється потребами установи та можливостями її актуальної інфраструктури. Слід співставляти не лише функціональність програмного забезпечення, а й те, які дані та технологічні засоби вже доступні, наскільки складними будуть їх перенесення або інте-

грація, скільки додаткової роботи потребуватиме створення репозитарію та хто зможе забезпечувати його подальшу підтримку.

Використання спеціалізованої платформи є доцільним, якщо потрібен широкий набір готових функцій і установа має ресурси для її розгортання, адміністрування та оновлення. Адаптація діючої системи може бути оптимальнішим кроком, якщо в ній уже зберігаються необхідні дані та реалізовано частину потрібних функцій. Власна реалізація виправдана і тоді, коли вимоги до репозитарію є відносно простими або специфічними, а розширення наявної інфраструктури дає змогу уникнути зайвої складності та повторного виконання вже проведеної роботи. Варто оцінювати не лише зусилля на створення репозитарію, але і його підтримку протягом всього життєвого циклу, що включає підготовку та передавання даних, адміністрування, оновлення програмного забезпечення, забезпечення безпеки, резервування, контроль цілісності та можливість подальшого перенесення даних. Важливо мати розуміння, яким чином підтримка згаданих процесів буде відбуватися через кілька років.

Таким чином, вибір технологічного рішення має бути результатом зіставлення конкретної потреби з реальною складністю її реалізації. Якщо окрема платформа вимагає створення нового інформаційного середовища, перенесення вже наявних даних і значних інтеграційних ресурсів, її переваги слід порівнювати з можливістю розширення наявної системи. Якщо ж наявна система не забезпечує необхідної функціональності або її розширення потребує надмірних зусиль, спеціалізована платформа може бути раціональнішим рішенням.

Таким чином, спеціалізовані платформи є важливим і часто оптимальним інструментом створення цифрових репозитаріїв. Водночас їх використання не є обов'язковою умовою. Репозитарій може бути створений також шляхом адаптації або розширення наявної інформаційної системи, якщо такий підхід відповідає потребам установи. Технологічна основа репозитарію не визначає його функціональну цінність. Незалежно від способу реалізації, репозитарій має забезпечувати належне збереження цифрових об'єктів, роботу з метаданими, доступ до них та можливість обміну даними з іншими системами.

Створений самотужки цифровий репозитарій має право на існування тоді, коли його створення має чітке технологічне, економічне та інституційне обґрунтування. У такому випадку він є не альтернативою «правильній» платформі, а іншим способом вирішення того самого завдання.

Viktor Zayika,

ORCID <https://orcid.org/0009-0003-6582-6524>,

Candidate of Physical and Mathematical Sciences,

Head of Department,

Information and Communication Technologies Department,

Institute of Information Technologies,

V. I. Vernadskyi National Library of Ukraine,

Kyiv, Ukraine

e-mail: zayika@nbn.gov.ua

Approaches to the development of digital repositories and the selection of their technological foundation are considered. The possibilities of using specialized platforms and extending existing information systems, particularly integrated library systems, are analyzed. Issues of long-term digital preservation, data integration, metadata management, the OAI-PMH protocol, and the FAIR principles in the context of open science are discussed. It is demonstrated that the choice of a technological solution should be determined by the institution's needs, existing infrastructure, the complexity of integration, and the capacity for long-term maintenance of the repository.

Keywords: digital repository platforms, Islandora, digital objects, long-term preservation, metadata, OAI-PMH, FAIR, open science.