

УДК 025.4; 004.738.5:02; 004.78:025.4.03

І. В. Лобузін

Національна бібліотека України імені В.І. Вернадського
проспект 40-річчя Жовтня, 03039 Київ, Україна

Цифровий комплекс сучасної наукової бібліотеки

Розглянуто технологічні ланки та окремі модулі життєвого циклу формування бібліотечних цифрових ресурсів. Визначено необхідний цифровий комплекс наукової бібліотеки. Обґрунтовано склад та послідовність технологічних модулів і процесів бібліотечного цифрового проекту.

Ключові слова: *цифрова бібліотека, бібліотечний цифровий ресурс, життєвий цикл цифрового ресурсу, цифровий комплекс наукової бібліотеки.*

Постановка проблеми

Сьогодні колекції рідкісних і цінних видань, рукописів а також краєзнавчі фонди бібліотек України через різні обставини руйнуються і потребують спеціальних заходів щодо їхнього збереження. Головним сучасним напрямом збереження документального національного надбання є організація оцифрування фондів бібліотек. Однак тут необхідно зазначити, що, незважаючи на досить детально розроблені стратегії оцифрування фондів бібліотек, слід констатувати відсутність єдиного системного підходу щодо організації самої технології формування бібліотечних цифрових ресурсів в Україні.

Для вирішення завдань організації цифрових ресурсів бібліотек сучасний інформаційний ринок пропонує різноманітні програмно-технологічні засоби. Кожна бібліотека має визначити пріоритети й обрати з цього розмаїття найбільш раціональні моделі та засоби інтеграції власних інформаційних ресурсів у світове інформаційне середовище. Як показує світова практика, тут є декілька основних підходів: використовувати вже готові рішення провідних світових дистриб'юторів, скористатися безкоштовними програмними платформами, розробленими в рамках ініціативи архівів відкритого доступу, або адаптувати доступні програмні рішення на платформі автоматизованої бібліотечної інформаційної системи. У кожному із цих рішень є свої недоліки і переваги. Розглянемо більш детально головні з них.

1. Готові рішення провідних світових дистриб'юторів. Комплексні рішення управління цифровими колекціями провідних розробників світового ринку інфор-

маційних послуг таких як ABBYY, XEROX, CANON, спеціалізовані платформи (QStar HSM, Saperion) — класу ECM (Enterprise Content Management) та аналогічні професійні програмні середовища для створення електронних бібліотек Rosetta (ExLibris), Dlibra (Poznańskie Centrum Superkomputerowo-Sieciowym), WorldCat, GoogleBook. Ці рішення включають комплексну автоматизацію процесів ретроконверсії довідково-пошукового апарату бібліотеки, виготовлення цифрових копій паперових документів, підготовку електронних версій документів, введення та редагування метаданих, формування електронних колекцій, он-лайнову або локальну публікацію створених цифрових ресурсів, виготовлення поліграфічних аналогів унікальних паперових документів максимально наближених до оригіналу, підтримка пошукового апарату колекцій і системи захисту інформації від розповсюдження та копіювання.

Серед недоліків цієї групи рішень для організації цифрових ресурсів слід зазначити, що такі системи коштують досить дорого, здаються спеціалістами провідних фірм «під ключ», потребують постійної не безкоштовної підтримки і консультацій, є мало доступними сьогодні для державних бібліотек України.

2. Рішення, що засновані на безкоштовних програмних продуктах: DSpace (DSpace Foundation), EPrints (EPrints Free Software), Fedora (Fedora Commons), Greenstone (New Zealand Digital Library Project), у більшості випадків реалізують відкриті міжнародні стандарти та протоколи, які реалізовані в рамках «Ініціативи відкритих архівів» («Open Archives Initiative», OAI, www.openarchives.org/) [28]. Найбільш популярною із вільно розповсюджуваних програмних платформ для організації цифрових ресурсів є програмний продукт Greenstone, який є не тільки засобом пошуку, але й засобом для зберігання та організації цифрових матеріалів, а також їхніх метаданих. Однією з особливостей програми є можливість швидкого пошуку в значних обсягах даних (сотні гігабайт). Спосіб зберігання метаданих сумісний з поширеними форматами, такими як Dublin Core. Поповнення бібліотеки можливе не лише з локального комп'ютера: бібліотечний інтерфейс надає можливість завантажування файлів через Інтернет, а також підтримує поширені протоколи обміну даними між бібліотеками, наприклад Z39-50, OAI-PMH.

Головним недоліком таких рішень, є те що вони, незважаючи на свою безкоштовність, потребують значних зусиль щодо налаштування користувацьких інтерфейсів, мовної адаптації, налагодження стабільної роботи відповідних інформаційно-пошукових можливостей, навчання співробітників бібліотеки роботі з новими інформаційними сервісами. Крім того, ці рішення потребують додаткових розробок та адаптації для організації інтеграції цифрового фонду з іншими бібліотечно-інформаційними ресурсами.

3. Рішення, яке засноване на застосуванні класичних АБІС (автоматизованої бібліотечної інформаційної системи, що призначена для комплексної автоматизації бібліотеки, має насамперед ту перевагу, що в інформаційній системі вже накопичений значний досвід з упорядкування документальних інформаційних ресурсів. Таке рішення має безперечні переваги та є з точки зору автора обґрунтованим в умовах бібліотеки: система має можливості для розробки власних прикладних рішень; сучасні АБІС працюють на основі клієнт-серверної архітектури, що дає змогу організувати мережеву взаємодію користувачів багатьох різних підрозділів; для великої кількості документів уже створено записи в електронному каталозі та

спеціалізованих базах даних; сучасні АБІС підтримують зберігання цифрових об'єктів у бінарному вигляді та гіпертекстові посилення до електронних документів; створений електронний фонд легко інтегрувати з іншими інформаційними сервісами бібліотеки. Адаптація такого рішення для формування цифрового фонду бібліотеки не потребує перенавчання бібліотечних та ІТ-спеціалістів і дає змогу використовувати вже налагоджений пошуковий апарат традиційної бібліотеки.

Серед недоліків рішення, заснованого на АБІС (яке є обґрунтованим з точки зору співвідношення ціни і якості), слід зазначити відсутність стандартних рішень щодо організації всього життєвого циклу формування бібліотечного цифрового ресурсу. Однак, зважаючи на інші значні переваги використання АБІС, як базової платформи, відповідні зусилля були зосереджені на аналізі та обґрунтуванні необхідного бібліотечного цифрового комплексу, його елементів і технологічних ланок.

Провідні спеціалісти з управління цифровими ресурсами (*digital curation*) визначають такі головні ланки організації життєвого циклу повноцінного функціонування цифрового ресурсу: *концептуалізацію* (планування відбору та створення цифрових матеріалів); *опис і відбір* (введення метаданих, оцінка цифрового матеріалу і відбір для довгострокового зберігання і відновлення); *управління* (оцінка цифрового матеріалу і відбір для довгострокового зберігання і відновлення, дотримання правових вимог, видалення матеріалів, які не були відібрані для довгострокового зберігання і відновлення, передача матеріалів до архіву, сховища, центру обробки даних); *збереження* (проведення заходів щодо забезпечення довгострокового зберігання: перевірка метаданих, перевірка цілісності цифрових даних, запис даних із дотриманням відповідних стандартів); *доступ і використання* (надання активного доступу користувачам до цифрового ресурсу: загальнодоступного, авторизованого або платного) [23, 24]. Саме за цими принципами побудовані комплексні рішення управління цифровими архівами (наприклад DocuShare XEROX).

Отже, серед основних модулів (підсистем) інформаційної системи цифрового фонду, що будуть забезпечувати його стабільне та ефективне функціонування, на думку автора доцільно виокремити такі:

- створення цифрових копій;
- опису та обліку цифрових ресурсів;
- обробки та управління цифровими ресурсами;
- зберігання цифрового фонду;
- пошуку та публікації цифрових ресурсів.

Кожна із цих підсистем має свої принципи функціонування і потребує окремого дослідження теоретичних і технологічних вимог у процесі опрацювання та упорядкування цифрових ресурсів. Розглянемо більш детально ці основні структурні блоки інформаційної системи цифрового комплексу наукової бібліотеки.

Підсистема створення цифрових копій

Ця підсистема є аналогом технології комплектування цифрової бібліотеки. Перші питання, які вирішує ця підсистема — це відбір документів для оцифру-

вання. Критеріями такого відбору можуть бути різноманітні фактори, відповідно до рекомендацій Е. Сміт [30] та Г. Ковальчук [6], основними з них є такі:

— *унікальні*, рідкісні або цінні документи, що існують в одному або в обмеженій кількості примірників, у тому числі рукописи та фотографії;

— *пошкоджені*, крихкі або великогабаритні документи, які потребують особливих умов доступу до них. Оцифрування таких документів сприяє їхньому збереженню, за умов надання доступу користувачам до їхніх цифрових копій;

— документи, що відносяться до об'єктів *національного надбання*, введення яких у широкий культурний обіг має велике значення для відновлення національної пам'яті, історії та культури держави.

Для цифрових архівів поповнення інформації відбувається за рахунок переведення паперових документів в електронний вигляд. Для цього використовуються, залежно від типу та формату документів, сканери, цифрове фотообладнання та програмне забезпечення цифрової обробки зображень.

Для підбору скануючого комплексу сучасної наукової бібліотеки, відповідно до практичних рекомендацій Адміністрації національних архівів і документації (*National Archives and Records Administration, NARA, www.archives.gov*), необхідно визначити характеристики та основні параметри оригіналів документів, цифрові копії яких необхідно буде виготовити [31]. До таких основних параметрів оригіналів відносяться: тип паперу та способи виготовлення, локальні недоліки, що потребують попередньої реставрації або ретушування, тональність, різкість (чіткість), габаритні розміри. Необхідно оцінити розміри аркушів паперу для рукописних і друкованих документів, кількість документів кожного формату. Від цієї оцінки буде залежати добір апаратного комплексу для оцифрування конкретного бібліотечного фонду.

Після сканування отримана цифрова інформація вимагає обробки для доведення відсканованого зображення до необхідного користувацького вигляду. Створено багато прикладних програм, за допомогою яких можна коригувати отримані під час сканування зображення. Найбільш розвинуті можливості має популярний програмний пакет Adobe Photoshop, який має величезну кількість різних інструментів для роботи із зображеннями. Серед програм, які дозволяють здійснювати швидко пакетне опрацювання отриманих цифрових зображень, слід зазначити ACDSee, Microsoft PfotoEditor (вбудована в Microsoft Office), Raster Desk, Raster ID, Spotlight. Для перетворення відсканованих паперових документів в електронні формати для подальшого редагування — системи розпізнавання тексту: ABBYY FineReader, OCR CuneiForm, Readiris, Microsoft Office Document Imaging. Для редагування об'ємних зображень використовуються програми StudioMAX, Maya, Rhinoceros, SolidWorks та інші пакети 3D-моделювання [4].

Наступним етапом створення еталонної цифрової копії документів є оцінка повноти та якості отриманих цифрових зображень. До повноти відноситься перевірка наявності всіх аркушів або зображень усіх боків оригіналу. До поняття якості, у першу чергу, відносяться чіткість отриманих зображень і кольорова відповідність оригіналу. Для знімання страхових копій рідкісних і старовинних документів з кольоровими зображеннями, оцифрування документів відбувається разом з еталонною шкалою кольорів, яка часто додається до комплекту супровідної документації професійних сканерів.

Підсистема опису та обліку цифрових ресурсів

Облік та інвентаризація цифрових масивів, у тому числі в сфері освіти, науки і культури, є предметом ряду національних і міжнародних проєктів: створення Національної електронної бібліотеки (Мінкультури Росії), <http://www.rusneb.ru/>; Проєкт Мінерва (Європейський союз), <http://www.minervaeurope.org/>; Цифрові бібліотеки для освіти, науки і культури (ЮНЕСКО), www.wdl.org.

Система обліку цифрових ресурсів забезпечує у цифровому фонді ті ж самі функції, що і система обліку традиційної бібліотеки — збереженість фонду. Саме ці питання збереженості цифрових ресурсів, обсяги яких лавиноподібно зростають у всьому світі, викликали дослідження питань опису (створення метаданих) цифрових ресурсів для забезпечення повноцінного керування створеними цифровими ресурсами та збереження їх у майбутньому. Дуже важливим компонентом стратегій збереження цифрової спадщини є метадані, що містять інформацію, необхідну для документування процесу зберігання [3]. Метадані збереження представляють собою інформацію про формат, структуру і використання цифрового ресурсу, історію всіх операцій, проведених над об'єктом збереження, в тому числі будь-які зміни, автентичність, технічну історію, історію зберігання, відповідальність, права, пов'язані зі збереженням тощо [29].

Отже, система обліку цифрових ресурсів тісно пов'язана з вибором метаданих для опису об'єктів створюваного цифрового фонду. Серед метаданих виділяють три основних типи: системні (службові), структурні та семантичні [16]. Основними вимогами до метаописів об'єктів є: повнота представлення знань, закріплених в онтології; можливість використання метаописів об'єктів у інших системах; використання загальноприйнятого стандарту, який сумісний з іншими системами.

Додавання метаданих до електронних ресурсів системи створює можливість більш точного позиціонування інформації про об'єкти, покращує механізм фільтрації та відбору знань, спрощує і прискорює процеси доступу до необхідних програм, серверів, ресурсів.

Однак таких систем метаданих, як універсальних, так і орієнтованих на окремі види цифрових масивів, налічується зараз вже більше 50 [14]. Найбільш перспективним і розповсюдженим є стандарт Дублінського ядра метаданих (*Dublin Core Metadata Initiative, DCMI*) [20]. У середовищі бібліотечного співтовариства більш популярний MARC-формат, в освітньому середовищі — LOM, у музейному середовищі — CIMI, в архівній справі — EAD і т.д. Крім того, навіть у рамках Дублінського ядра можливі різні версії, обумовлені різним набором кваліфікаторів, різними методичними принципами та ін. Однозначного вирішення цієї проблеми не існує, хоча найбільш прийнятним виходом з положення є застосування ієрархії систем метаданих, в якій вершину утворює Дублінське ядро як найбільш загальна система. Інші системи метаданих можуть бути ієрархічним розвитком Дублінського ядра, що забезпечить однозначне конвертування метаданих.

Однак для цифрових бібліотечних ресурсів все ж таки певні переваги мають рішення, засновані на родині MARC-форматів, бо це дає змогу використовувати вже готові описи бібліотечних документів з електронних каталогів і бібліографічних баз даних [9].

Однією із суттєвих проблем інтеграції полівидових ресурсів фонду наукової бібліотеки є забезпечення сумісності опису архівних і бібліотечних документів. Так, основними стандартами, що прийняті для опису архівних документів, є ISAD (G) (Загальний міжнародний стандарт архівного описування) [25]. Для забезпечення стандартизованого обміну записами між архівними каталогами призначений формат метаданих для архівного опису EAD (Encoded Archival Description), що забезпечує функції аналогічні MARC-формату для бібліографічних записів [21], формат передбачає у кодованих даних забезпечення сумісності з MARC 21 [27].

Провідні фахівці зі створення електронних архівів, такі як Адміністрація національних архівів і документації (*National Archives and Records Administration, NARA, www.archives.gov*) [31], рекомендують під час формування фонду цифрових копій вводити не тільки описові метадані документів, а обов'язково передбачити технічні та правові метадані. Технічні метадані мають відобразити основні характеристики зображення (колір, роздільну здатність, розмір, обсяг), обладнання та програмне забезпечення, засобами яких створено зображення та електронна версія документа. Правові метадані особливо важливі для документів архівного фонду, бо за правилами на копії розповсюджуються ті самі обмеження, що стосуються їхніх паперових аналогів. Однією із суттєвих проблем, що пов'язана з регламентацією доступу до інформації (поза службовими взаємовідносинами), з якою зустрічаються бібліотеки під час обслуговування електронними інформаційними ресурсами, є дотримання авторських прав і правил архівного обслуговування [2].

Підсистема обробки та управління цифровими ресурсами

Після створення оцифрованих копій бібліотечних документів отриманий цифровий фонд потребує подальшого упорядкування. Основні процеси, які необхідно здійснити на цьому етапі, — це створення електронних версій документів для повноцінного їхнього представлення користувачам і формування цифрових колекцій з відповідним пошуковим апаратом. Розглянемо більш детально ці процеси з точки зору опрацювання цифрових ресурсів наукової бібліотеки.

Вимоги до електронних версій документів мають відповідати у першу чергу критерію достовірності, а саме: документ в електронному вигляді має максимально відтворювати оригінал. Крім того, інформація в електронному вигляді надає додаткові можливості з боку зручності маніпулювання інформацією: зручна навігація у матеріалі та пошукові індекси.

Для представлення документів у електронних бібліотеках можуть використовуватися різні формати, в тому числі найбільш розповсюдженими є такі: PDF, DjVu, DOC, TXT, TIFF, JPEG, HTML, XML, FB (FictionBook), CHM, SWF (FLASH-документи).

Можуть бути також передбачені особливі формати для представлення аудіо-відеоінформації, геоданих, тривимірних або анімаційних об'єктів, експериментальних даних та інших особливих видів інформації.

Формати зображень для зберігання страхових копій документів мають відповідати таким основним технічним вимогам: підтримка бітової глибини більше 24 для однотонних документів; схеми стиснення даних без утрат; підтримка різних

колірних моделей і можливості управління кольором; можливість підтримувати високу роздільну здатність майстер-копій; відкритість і добра документованість, широка підтримка та крос-платформна сумісність [7].

Проблему захисту авторських прав у бібліотеках успішно вирішують системи віддаленого захищеного перегляду документів такі як «DefView», «Vivaldi», «SDViewer» тощо. Реалізація засобів захищеного перегляду надає можливість бібліотекам організувати обслуговування віддалених користувачів документами захищеними авторським правом у віртуальному читальному залі, допомагає залучити до фондів бібліотеки сучасні видання авторів, зацікавлених у популяризації та розповсюдженні їхніх інтелектуальних творів.

Ще одним важливим питанням є розпізнавання документів і зберігання їх у символічних форматах. Такі рішення пропонуються практичними рекомендаціями ЮНЕСКО щодо створення електронних бібліотек освітніх ресурсів [17]. Розпізнавання тексту (*Optical character recognition, OCR*) може бути повним цілого документа, можливі також проміжні рішення — розпізнавання деяких частин документів (зміст, анотація, титульна сторінка) або так зване «чорне» розпізнавання, тобто розпізнавання без коректури. Прийняття рішення в даному випадку є досить непростим. Розпізнавання надає можливість посимвольної обробки тексту, отже, різноманітних способів роботи з текстами (пошук, редагування, експорт, імпорт тощо). У той же час оцифрування документів з перекладом в символічні формати значно дорожче [18]. Програми OCR працюють не ідеально, допускаючи певну кількість помилок, після їхньої роботи для якісної підготовки тексту потрібна ще й ручна коректура. У процесі перекладу якісного чіткого тексту кількість помилок складає 1 % і менше. У той же час у процесі оцифрування старовинних книг кількість помилок доходить до 50 % [12].

Для текстів «неоптимальних» для розпізнавання, наприклад рукописних текстів і текстів з математичною нотацією, нетрадиційних алфавітів тощо, проблема часто не може бути задовільно вирішена у близькому майбутньому. Альтернативою технології «сканування – розпізнавання – коректура» є технологія ручного введення (передруку) текстів, яка у багатьох випадках виявляється дешевшою [1].

У процесі підготовки цифрових копій унікальних і цінних друкованих і рукописних об'єктів відповідно до сучасних вимог професійного кураторства цифровим ресурсом та забезпечення його збереженості у майбутньому необхідно дотримуватися технологічної політики, що забезпечує можливість збереження електронних копій протягом тривалого часу і заснованої на наступних принципах: створення базової (депозитарної) цифрової копії об'єкта; формування на основі депозитарної копії необхідних для користувача представлень цифрового об'єкта в різних зручних для роботи форматах; можливість ітеративного повернення з метою модифікації і розвитку цифрової копії об'єкта в процесі її експлуатації.

Після вирішення завдань визначення необхідних форматів виготовлення цифрових об'єктів, наступним організаційним рівнем упорядкування цифрових ресурсів є визначення моделі формування цифрових ресурсів. За даними узагальнюючої публікації Т.В. Майстрович можлива реалізація кількох методичних підходів (моделей): 1) суцільна оцифровка фондів бібліотеки; 2) переведення в електронну форму частини документів, що активно використовуються; 3) формування електронних колекцій за окремими критеріями; 4) створення комплексної куль-

турно-освітньої програми і бази знань [11]. Вибір рішення визначається завданнями кожної бібліотеки.

На думку автора для універсальних наукових бібліотек найбільш прийнятними є моделі 3) та 4). На початкових етапах формування цифрового фонду в межах однієї наукової бібліотеки прийнятною є модель формування окремих електронних колекцій за тематичною ознакою та видами документів. Надалі такий ресурс може стати частиною національного проекту реалізації державної історико-культурної програми.

Відповідно до правил упорядкування знань та укладання електронних колекцій необхідно, щоб вони були детально описані, бо вони також є об'єктом зберігання у цифровому фонді бібліотеки. Створення метаописів колекцій має дві мети: надає користувачам експертних знань щодо підготовленого інформаційного ресурсу та значно покращує навігацію у цифровому середовищі. Представлення метаданих колекцій є обов'язковим елементом інтеграції електронних колекцій бібліотек до світових цифрових ресурсів (The European Library, Europeana, World Digital Library).

Підсистема зберігання цифрового фонду

З точки зору забезпечення збереження великих обсягів цифрових даних, в останні роки широкого поширення набула концепція управління життєвим циклом інформації, в основі якої лежить принцип поділу загального масиву даних на класи в залежності від змісту, частоти звернень та терміну зберігання. Відповідно до цього підходу виділяються три ключові завдання зберігання цифрових даних: оперативний доступ до інформації, резервне копіювання та архівне зберігання. Для вирішення кожної з них застосовується різне устаткування — згідно зі специфічними вимогами до зберігання і доступу [7].

Оперативний архів реалізують розміщенням на сервері у спеціальному розділі жорсткого диска (для невеликих обсягів інформації), на окремому жорсткому диску або RAID-масиві. Для підвищення надійності функціонування подібних систем використовують засоби резервного копіювання, дублювання інформації та «дзеркальних дисків». RAID-масиви також використовують для зберігання метаданих та облікової бази даних архіву. В практиці американських електронних архівів ці дані мають назву Master References File (еталонний файл) [20]. Найбільш перспективним рішенням, що одночасно задовольняє вимогам швидкості та надійності, є RAID-5 або -6.

Подальшим розвитком рішень щодо зберігання великих обсягів даних є спеціалізовані системи зберігання даних (СЗД). Сучасна система зберігання даних має включати такі основні компоненти: пристрої зберігання даних (жорсткі диски, компакт-диски тощо); інфраструктуру доступу серверів до пристроїв зберігання даних; систему архівування та резервного копіювання даних; програмне забезпечення керування зберіганням даних. Основними завданнями, які виконує СЗД, є надійне зберігання даних, а також відмовостійкий та високопродуктивний доступ серверів до засобів зберігання даних. Основні методи, які використовуються сьогодні для реалізації цих завдань, можна охарактеризувати одним словом — дублювання (дублювання даних і каналів доступу до них) [7]. Більшість сучасних сис-

тем зберігання даних здійснюють процеси реплікації та резервного копіювання в автоматизованому режимі.

Для організації області довгострокового (архівного) зберігання використовуються змінні носії інформації CD/DVD/UDO та роботизовані бібліотеки на їхній основі. Найбільш досконалі рішення підтримують надщільний оптичний формат (Ultra Density Optical, UDO) та технології Blue-Ray. Популярність технології UDO ґрунтується на можливості записати на один носій 30–60 Гб інформації з терміном зберігання інформації 50 років. Роботизована бібліотека є наступним етапом створення електронного архіву довгострокового зберігання, вона представляє собою масив змінних дисків, розміщений у спеціальному корпусі. Рішення на основі роботизованої бібліотеки зі змінними носіями інформації також дозволяє досить легко вирішити питання технічного захисту інформації і забезпечити її надійне зберігання [13].

З точки зору зберігання самих файлів електронних документів є також два основних підходи: за умов використання першого з них електронні документи зберігаються у бінарному вигляді у відповідних полях самої бази даних архіву, другий підхід базується на тому, що файли можуть зберігатися окремо, а в базі даних є тільки електронні адреси файлів (вказується шлях доступу до файлу). Перший із способів швидко призводить до цілої низки проблем: ускладнення резервування даних, перехід до іншої інформаційної платформи, зниження швидкодії системи. Він може бути рекомендований лише для архівів з невеликим фізичним обсягом інформації.

Крім надійності та продуктивності електронного архіву або сховища даних важливою також є логічна структура сховища [15]. За базову структуру рекомендується прийняти ієрархію папок з відповідними документами, яка буде відображати логічний зв'язок основних вузлів системи.

До кола питань збереження цифрової інформації відносяться також питання інформаційної безпеки. В процесі створення цифрових копій доступ користувачів до цифрового архіву необхідно регулювати залежно від виду, призначення і ступеня важливості інформації, способу обробки даних тощо. З точки зору інформаційної безпеки інформацію можна розділити на чотири класи: критичну (цифрові копії унікальних та особливо цінних документів), модифікація або пошкодження якої призведе до невиправної втрати (внутрішньо-системне програмне забезпечення цифрової техніки); важливу, доступну невеликій групі користувачів (налаштування прикладного програмного забезпечення, місце зберігання цифрових копій і розподіл доступу до ресурсів системи); інформацію, несанкціонований доступ до якої може привести до витоку більш цінних даних (ідентифікатори, паролі користувачів, мережеві налаштування); інформацію, яка не представляє конкретного інтересу для зловмисників, але вимагає захисту від випадкових порушень через безвідповідальність виконавців (службові бази даних обліку робіт, реєстрації адміністративних та технічних метаданих, структура та наповнення цифрової бібліотеки на веб-порталі).

Найбільшу загрозу інформаційній безпеці становлять випадкові загрози, які призводять до найбільших утрат інформації, у цьому випадку може відбуватися знищення, порушення цілісності та доступності інформації. Основними засобами захисту від випадкових загроз є регулярне дублювання (резервування даних) та

програмне блокування помилкових операцій (формально-логічний контроль за цілісністю даних).

Навмисні загрози відповідно до їхньої фізичної сутності та механізмів реалізації можуть бути розподілені на такі основні групи: несанкціонований доступ до інформації, модифікація структури інформаційної системи та даних, шкідливе небезпечне програмне забезпечення.

Для протистояння як випадковим, так і навмисним загрозам, систему інформаційної безпеки доцільно розглядати у вигляді єдності трьох компонентів: інформація; технічні та програмні засоби; обслуговуючий персонал і користувачі. На практиці застосовують системи захисту у вигляді трьох складових частин: основного контуру безпеки, засобів протидії випадковому несанкціонованому доступу та засобів управління системою захисту. Основний контур виконує функції захисту носіїв даних (як змінних, так і незмінних), захисту технічних засобів від несанкціонованого проникнення, контролю введення апаратури в режим виконання регламентних робіт. Випадковому несанкціонованому доступу можна запобігти постійним контролем за якістю та цілісністю інформації (контрольною сумою файлів), застосуванням мережевого програмного забезпечення із засобами аутентифікації. Центральною ланкою системи захисту є адміністрування системи безпеки, що включає реєстрацію користувачів, регламентацію їхнього доступу, реєстрацію паролів, перевірку журнальних записів, за якими можна прослідити за зверненням до інформації, її зміною та передаванням у межах локальної мережі [5].

Основний контур інформаційної безпеки забезпечується обмеженням доступу у приміщення з серверною та цінною цифровою апаратурою, допуском до роботи з професійним цифровим обладнанням підготовлених спеціалістів (системних адміністраторів і спеціалістів з цифрової обробки зображень), регулярним проведенням навчання та заходів з підвищення кваліфікації. Цілісність і збереження цифрового ресурсу забезпечується регулярною перевіркою якості цифрових матеріалів і резервуванням даних (оперативним на серверах через захищені протоколи передавання даних) та довгостроковим (на носіях довготривалого зберігання даних, аналогічних UDO-дискам). Носії довготривалого зберігання даних страхового фонду цифрових копій захищаються паролем та зберігаються у спеціально облаштованому сховищі (сейфі).

Захист інформації від навмисного несанкціонованого доступу забезпечує система розмежування доступу до інформації, заснована на керуванні доступом. Найбільш складним є розмежування доступу користувачів до баз даних, де він регламентується правилами, визначеними для окремих частин бази даних, включає такі основні ланки як «Ідентифікація → Аутентифікація → Авторизація». Основним засобом на сьогодні залишається парольна аутентифікація, яка потребує певного свідомого відношення персоналу до своїх приватних даних.

Стосовно шкідливого програмного забезпечення найнадійнішим методом захисту є використання ліцензійних професійних програмних антивірусних засобів, а також виконання користувачами правил безпечної роботи з інформацією: дублювання (резервування) особливо цінної інформації, обережність у разі використання змінних носіїв інформації та нових файлів (унеможливлення використання змінних носіїв інформації та доступу до мережі Інтернету в зоні доступу до цифрових копій унікальних та особливо цінних документів).

Як показав практичний досвід, для повноцінної реалізації центральної ланки системи захисту інформації найбільш дієвим засобом адміністрування є застосування доменних політик і забезпечення серверного централізованого керування системою антивірусного захисту (аналогічного Kaspersky Security Center), автоматичне протоколювання сеансів авторизованого доступу до FTP-серверів та веб-серверів.

Підсистема пошуку та публікації цифрових ресурсів

Остаточною метою будь-якого проекту з оцифрування історико-культурних та наукових бібліотечних інформаційних ресурсів є реалізація доступу до створеного інформаційного масиву для широких верств населення. Очевидно, що цифрові бібліотеки вирішують низку важливих питань у сучасних соціальних комунікаціях: ліквідація інформаційної нерівності (доступ до інформації надається незалежно від часу, місця у просторі, соціальної приналежності); сприяння організації освітніх процесів і безперервного навчання; активізація перебігу інформаційно-знаннєвих процесів у наукових комунікаціях суспільства; сприяння збереженню Пам'яті Світу; створення умов для формування високоякісних ресурсів наукового та історико-культурного надбання; сприяння процесам віртуальної реконструкції важливих суспільних подій і реституції визначних національних колекцій. Інтерактивні властивості цифрового середовища надають унікальну можливість візуалізації інформації про культурну спадщину суспільства. Он-лайніві інтернет-сервіси, забезпечені когнітивними засобами подання інформації дають нові можливості бібліотекам та архівам для розробки інформаційних сервісів, орієнтованих на співтовариство кваліфікованих користувачів.

Будь-яка цифрова бібліотека призначена для певного кола користувачів і має надавати підтримку конкретних видів діяльності. Таким чином, для розробки якісного користувацького інтерфейсу, функціонування цифрові бібліотеки мають бути оцінені в контексті їхньої цільової аудиторії та галузі діяльності. [19]. Відповідно до досліджень російських учених НТЦ «Інформрегістр» розвинуті функціональні та пошукові можливості системи, з точки зору ергономічного інтерфейсу користувача наукової бібліотеки, мають включати [1]: навігацію у всьому доступному інформаційному просторі, лексичний пошук, символний пошук, атрибутивний пошук (автор, назва, місце видання, дата видання тощо); перегляд змісту цифрового об'єкта і його структури; багатовіконний режим роботи; підтримка апарату гіпертекстових і гіпермедійних зв'язків; протоколювання сеансу роботи користувача; експорт інформації із системи з вказівкою на джерело інформації.

Стосовно подальшого розвитку когнітивних можливостей інтерфейсу цифрової бібліотеки слід розглянути коло питань, які відносяться до організації інтелектуального пошуку інформаційних ресурсів бібліотеки. Розробку такого інтелектуального інструментарію забезпечують бібліотечні технології семантичної (наукової) обробки інформації, коли під час опису ресурсів враховуються не лише формальні описові ознаки документів, а, у першу чергу, розкривається їхній зміст. Додатковими засобами розкриття змісту цифрових ресурсів, особливо для неструктурованої графічної інформації оцифрованих сторінок, є введення змісту документів та анотацій. Текстова інформація цих додаткових точок доступу автоматично

включається у пошукові індекси, що значно розширює пошуковий потенціал створеного інформаційного масиву.

Використання усталених і добре розроблених семантичних бібліотечних технологій значно оптимізує зусилля на удосконалення пошукового апарату цифрової бібліотеки. Під час формування цифрового масиву є можливість скористатися вже готовими засобами тематичного пошуку: рубрикаторами, класифікаторами, тезаурусами, авторитетними файлами; авторитетними файлами осіб та організацій; науковими описами рукописних та архівних фондів наукової бібліотеки. Всі ці засоби значно покращують пошуковий апарат цифрового фонду, надають користувачу у зручному візуалізованому гіпертекстовому вигляді додаткову довідково-бібліографічну інформацію щодо його інформаційного запиту, дають змогу реалізовувати наукоємні засоби інформаційного пошуку та отримати повну й достовірну інформацію [10].

Висновки

Запропонований модульний цифровий комплекс було сформовано та апробовано протягом чотирьох років у рамках цифрового проекту Національної бібліотеки України імені В.І. Вернадського. Як показав набутий досвід, кожна з підсистем (модулів): створення, опису та обліку, обробки та управління, зберігання, пошуку та публікації цифрових ресурсів відповідає певному етапу життєвого циклу формування цифрового фонду наукової бібліотеки. Розробка окремих модулів з певною мірою автономними завданнями створило передумови для детального дослідження особливостей кожної технологічної ланки, більш раціональної організації і розподілу професійних обов'язків спеціалістів бібліотеки. Остаточо, створений цифровий комплекс і послідовність модулів із визначеним колом технологічних процесів забезпечили:

- вирішення двох головних завдань цифрового проекту наукової бібліотеки: 1) створення цифрової бібліотеки відкритого доступу та 2) організацію страхового фонду цифрових копій документів;
- логічну послідовність виконання операцій технологічного циклу;
- необхідну технологічну інфраструктуру для виконання всіх завдань, які передбачає цифровий проект наукової бібліотеки;
- раціональне використання метаданих цифрових об'єктів на всіх технологічних етапах: створення – збереження – облік – публікація цифрових ресурсів;
- обґрунтований розподіл функціональних обов'язків і відповідальності співробітників, які приймають участь у формуванні цифрового ресурсу;
- розвинуту систему пошуку інформації у цифровій бібліотеці, орієнтовану на проведення наукових досліджень;
- сумісність із загальноприйнятими стандартами та можливість інтеграції створених цифрових ресурсів до міжнародних цифрових проектів.

1. Антопольский А.Б. Правовые и технологические проблемы создания и функционирования электронных библиотек / А.Б. Антопольский, Е.А. Данилина, Т.С. Маркова. — М.: ПАТЕНТ, 2008. — 207 с.

2. *Бобришева О.* Правові засади формування комплексної системи захисту інформації в бібліотеках / *О. Бобришева* // Вісн. Кн. палати. — 2009. — № 12. — С. 23–26.
3. *Браккер Н.В.* Сохранение цифрового наследия — мировой опыт / *Н.В. Браккер, Л.А. Куйбышев* // Информационный бюллетень РБА. Ч. 1. Доклады и сообщения. — 2009. — № 53. — С. 23–27.
4. *Гонсалес Р.* Цифровая обработка изображений / *Р. Гонсалес, Р. Вудс.* — [3-е изд., испр. и доп.] — М.: Техносфера, 2012. — 1104 с.
5. *Шаньгин В.Ф.* Защита информации в компьютерных системах и сетях / *В.Ф. Шаньгин.* — М.: ДМК Пресс, 2012. — 592 с.
6. *Ковальчук Г.І.* Критерії відбору книжкових пам'яток для першочергового поцифрування / *Г.І. Ковальчук* // Наук. пр. Нац. б-ки України ім. В.І. Вернадського. — К., 2010. — Вип. 28. — С. 141–147.
7. *Корепанов И.* Как сохранить архив на десятилетия? [Электронный ресурс] / *И. Корепанов* // «Журнал сетевых решений (LAN)». — 2008. — № 3. — Режим доступа: <http://www.osp.ru/lan/2008/03/4899898/>
8. *Васяева Н.С.* Анализ архитектур интеллектуальных систем хранения данных / *Н.С. Васяева, М.А. Сибиряков* // Математика. Механика. Информатика. — 2013. — № 2. — С. 32–51.
9. *Лобузін І.В.* Створення бази даних цифрового фонду бібліотеки: основні проблеми та технологічні рішення / *І.В. Лобузін* // Науково-технічна інформація (НТІ). — 2013. — № 1. — С. 67–62.
10. *Лобузін І.В.* Створення, представлення та перспективи використання електронного фонду цифрових копій документів бібліотеки / *І.В. Лобузін* // Реєстрація, зберігання і оброб. даних. — 2011. — Т. 13, № 4. — С. 60–73.
11. *Майстрович Т.В.* Электронный документ как объект библиотечного дела / *Т.В. Майстрович*; Рос. гос. б-ка. — М.: Пашков дом, 2004. — 247 с.
12. *Масевич А.Ц.* К созданию электронных коллекций старопечатных книг в библиотеке Российской академии наук: на примере работы над двумя проектами / *А.Ц. Масевич, Е.А. Савельев, А.К. Багажков* // Новые технологии в информац. обесп. науки: тр. конф. — М.: Биоинформсервис, 2001. — С. 132–140.
13. *Меленець А.В.* Обрання типу сховища для зберігання документів страхового фонду в електронному вигляді // Системи обробки інформації. — Х.: ХУПС. — 2010. — № 6(87). — С. 199–203.
14. *Обзор* форматов метаданных [Электронный ресурс] / сост.: *Ю.Е. Хохлов, С.А. Арнаутов*; Институт развития информационного общества // Российские электронные библиотеки. — Режим доступа: http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/md_rev
15. *Петров І.В.* Обґрунтування вибору програмно-апаратних засобів архівів електронних документів довготермінового збереження / *І.В. Петров, А.М. Стеценко, Н.В. Солоніна* // Реєстрація, зберігання і оброб. даних. — 2010. — Т. 12, № 1. — С. 79–88.
16. *Тузовский А.Ф.* Системы управления знаниями (методы и технологии) / *А.Ф. Тузовский, С.В. Чириков, В.З. Ямпольский* / под общей ред. *В.З. Ямпольского.* — Томск: Изд-во науч.-техн. лит., 2005. — 260 с.
17. *Электронные библиотеки в образовании* : специализированный учебный курс: практическое руководство / Институт ЮНЕСКО по информационным технологиям в образовании. — М.: ЮНЕСКО, 2007. — 247 с.
18. *Best Practice Guidelines for Digital Collections* Best Practice Guidelines for Digital Collections [Electronic resource] / ed. by *Susan Schreibman* // University of Maryland Libraries. — 2006. — Made of access: http://www.lib.umd.edu/dcr/publications/best_practice.pdf

19. Chowdhury S. Usability and Impact of Digital Libraries: A Review / Sudatta Chowdhury, Monica Landoni and Forbes Gibb // Online Information Review. — 2006. — N 6. — P. 656–680.
20. DCMII (Dublin Core Metadata Initiative) [Electronic resource]. — Way of access: <http://dublincore.org>
21. EAD: кодированное архивное описание. Библиотека тегов. Версия 2002 / гл. ред. пер. Е.Д. Жабко. — Санкт-Петербург: ФГБУ «Президентская библиотека имени Б.Н. Ельцина», 2011. — 337 с. — (EAD Технический документ; № 2).
22. Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, Particularly Those Held by Libraries and Archives [Electronic resource] / IFLA; UNESCO // IFLA. — 2003. — Way of access: <http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>
23. Harvey R. Digital Curation: A How-To-Do-It Manual / R. Harvey. — New York; London: Neal-Schuman Publishers, Inc., 2010. — 17 p. — (How-to-do-it manuals; N 170).
24. Higgins S. Digital Curation: The Emergence of a New Discipline / S. Higgins // The International Journal of Digital Curation. — 2011. — Vol. 6, N 2. — P. 78–88.
25. ISAD (G): Загальний міжнародний стандарт архівного описування. Схвалено комітетом з дескриптивних стандартів. — Стокгольм (Швеція). — 19–22 вересня 1999 р. / пер. та укр. версія Г.В. Папакіна; Державний комітет архівів України; Український НДІ архівної справи та документознавства. — [2-е вид.]. — К., 2001. — 48 с.
26. Knight G. Modelling OAIS Compliance for Disaggregated Preservation Services / Gareth Knight, Mark Hedges // The International Journal of Digital Curation. — 2007. — Vol. 2, N 1. — P. 62–72.
27. MARC 21 Format for Bibliographic Data [Electronic resource] // Library of Congress. — Way of access: <http://www.loc.gov/marc/>
28. Open Archives Initiative (OAI) [Electronic resource]. — 2007. — Way of access: www.openarchives.org/
29. PREMIS Data Dictionary for Preservation Metadata [Electronic resource]: Final Report of the PREMIS Working Group. — 2005. — Way of access: www.loc.gov/standards/premis/v2/premis-2-0.pdf
30. Smith A. Strategies for Building Digitized Collections / A. Smith // Digital Library Federation; Council on Library and Information Resources. — Washington (DC), 2011. — 35 p.
31. Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files — Raster Image [Електронний ресурс] / S.T. Puglia, J. Reed, E. Rhodes. — U.S. National Archives and Records Administration (NARA). — 2004. — 87 p. — URL: <http://www.archives.gov/preservation/technical/guidelines.html>

Надійшла до редакції 13.02.2014