

Станіслава Клочок
Євгенія Коцюба,
мол. наук. співробітники НБУВ

Комплектування, опрацювання та облік цифрових ресурсів Національної бібліотеки України імені В. І. Вернадського

У статті порушуються проблеми формування цифрових ресурсів наукової бібліотеки відповідно до колекційного принципу. Детально розглянуто розроблену в Національній бібліотеці України імені В. І. Вернадського технологічну модель оцифрування, каталогізації та обліку цифрових копій різних документних ресурсів. Акцентується увага на питаннях формування публічного доступу до цифрових ресурсів та збереження страхового фонду цифрових копій.

К л ю ч о в і с л о в а: НБУВ, цифрові бібліотечні ресурси, електронні колекції, оцифрування документів, обробка електронних документів, цифрова бібліотека, страховий фонд цифрових копій.

The problems of the formation of the scientific library of digital resources according to framework collection principle are described. A detailed study is developed the technological model digitizing at The Vernadsky National Library of Ukraine, cataloging and recording of digital copies of various kinds of documentary resources. The formation of public access to digital resources and the preservation of the insurance fund of digital copies are considered.

K e y w o r d s: digital library resources, digital collections, digitization of documents, processing of electronic documents, digital library, digital copies of the insurance fund.

Створення цифрових ресурсів історико-культурних фондів бібліотек є нині надзвичайно важливим питанням у контексті активного входження цих установ у глобальний інформаційний простір. Аналіз публікацій Т. Різа [7], А. Сміт [8], Р. Харвея [6], інших авторів засвідчує, що фахівці розглядають переважно питання стратегії формування бібліотечних цифрових ресурсів і недостатньо приділяють уваги дослідженню таких аспектів формування цифрових ресурсів, як комплектування, опрацювання та облік. Як показала практика, створення масштабного цифрового ресурсу на базі оцифрування полідокументного фонду Національної бібліотеки України імені В. І. Вернадського (НБУВ) є вкрай важливою, потрібною справою. Адже професійні програмні продукти, які дають змогу організувати весь технологічний цикл формування електронних колекцій, за своєю вартістю не завжди доступні для бібліотек України. Тому вирішення цих проблем в НБУВ на базі традиційної автоматизованої бібліотечної інформаційної системи, відповідно до головних

принципів формування бібліотечних фондів, показало свою оптимальність та ефективність.

Нині залучення до фондів електронних ресурсів перетворилося на повсякденну практику бібліотек. При цьому, перед бібліотекарями постають завдання, які мало чим відрізняються від традиційних: визначення джерел комплектування, каталогізація, організація доступу. Тільки тепер всі вони вирішуються з урахуванням специфіки нових інформаційних ресурсів.

Існує кілька джерел отримання електронних ресурсів бібліотеками. Це передплата комерційних повнотекстових баз даних, купівля електронних ресурсів на компакт-дисках і т. п., самостійне створення або збирання електронних ресурсів, інтеграція в бібліотечне обслуговування веб-ресурсів.

Оцифрування належить до завдань самостійного створення бібліотеками електронних ресурсів. Як свідчать реалії, створення національних цифрових ресурсів для політематичних та полідокументних фондів великих бібліотек найбільш перспективним є колекційний принцип, який і було обрано в НБУВ головною стратегією формування цифрового фонду.

Комплектування: відбір матеріалів для оцифрування

У процесі формування електронних колекцій основним критерієм виступає ефективність. А вже оцифрування документів проводиться не для накопичення абстрактних показників у вигляді кількості сторінок і видань, не для виконання грантових зобов'язань, а для вирішення нагальних проблем бібліотеки. Серед останніх головними є підвищена експлуатація деяких видів видань, що змушує активно опікуватися питаннями їх фізичного збереження та створення цифрових об'єктів для документів історико-культурного надбання, що уможлиблює роботу з цифровими копіями широкому загалу користувачів. Якщо більшості читачів надаватиметься цифрова копія видання, а не фізичний примірник, це сприятиме вирішенню як проблем розширення доступу до бібліотечних фондів, так і забезпечення їх фізичного збереження. Відповідно до цих завдань, у НБУВ оцифрування документів здійснюється двома основними напрямками: формування онлайн-оцифрованої бібліотеки та створення страхового фонду цифрових копій [3].

Історико-культурні фонди НБУВ дуже різні (стародруки, рідкісні видання, рукописи, образотворчі видання та альбоми, фотоматеріали, ноты, архівні документи, газети, карти тощо), тому вже на першому етапі комплектування електронних колекцій важливо організувати масиви цих документів так, щоб полегшити їх використання у майбутньому.

На першому етапі технологічного циклу формування цифрового фонду підрозділи-фондоутримувачі НБУВ подають пропозиції до плану оцифрування документів на рік, відповідно до визначених адміністрацією бібліотеки річних завдань створення електронних колекцій. Збиранню замовлень сприяє база даних «Цифрові ресурси НБУВ», розроблена і введена в експлуатацію наприкінці 2012 р. Вона призначена для формування та редагування річного плану оцифрування [5]. До бази даних мають доступ як співробітники відділів-фондоутримувачів, так і співробітники групи оцифрування документів. Структура замовлень бази даних «Цифрові ресурси НБУВ» включає: контактну інформацію про підрозділ, назву електронної колекції або електронної виставки, обґрунтування доцільності формування цифрових ресурсів, список документів зі стислим описом та інформацією про шифри, місце зберігання, кількість документів та сторінок. Для електронних виставок вказується дата, до якої мають бути проведені роботи з оцифрування для своєчасного розміщення електронних

ресурсів на порталі НБУВ. Для кожного замовлення наводиться інформація про призначення цифрової копії: онлайн-публікація, страхова копія, службове використання.

Другим етапом формування плану є визначення адміністрацією бібліотеки найбільш пріоритетних замовлень на поточний рік електронних колекцій та формування проекту річного плану оцифрування в межах пропускну здатності цифрового обладнання НБУВ.

Третім етапом є розподіл запланованих на рік для оцифрування документів згідно з наявним обладнанням та його продуктивністю. Для правильного розподілу документів за обладнанням проводиться експертна оцінка – співробітник групи оцифрування особисто попередньо оглядає документи і визначає, на якому обладнанні його можна оцифрувати, і ця інформація вноситься для кожного документа в базу даних «Цифрові ресурси НБУВ».

На четвертому етапі, відповідно до визначених у річному плані завдань та норм використання на відповідні процеси робочого часу, робота розподіляється між безпосередніми виконавцями у групі з оцифрування документів.

Опис цифрових ресурсів: формування метаданих

Перед початком робіт з безпосереднього оцифрування документів для кожного з них оформляється відповідний запис у службовій базі даних «Оцифрування документів», до якої обов'язково вносяться метадані майбутнього цифрового ресурсу як описового характеру, так і технічні та адміністративні. Описові метадані цифрових об'єктів сумісні з бібліографічними описами та описами архівних документів, що дає змогу легко експортувати у систему вже створені відповідними бібліотечними спеціалістами описи документів в UNIMARC-форматі. Технічні метадані містять інформацію про обладнання, проведені роботи, обсяг робіт (назва папки / файлу, кількість файлів, обсяг у байтах, формат файлів, роздільна здатність, колір, глибина кольору). Адміністративні метадані зазначають: № замовлення, замовника (підрозділ НБУВ або установу / організацію), тип замовлення, підставу для виконання, призначення копії, найменування проекту або колекції, стан виконання робіт, дати (початок – кінець), виконавця [2].

Оформлення цієї метадані дає змогу надалі однозначно ідентифікувати цифровий ресурс, запобігти дублюванню робіт з оцифрування однакових документів, забезпечити поступове допов-

нення створеного ідентифікаційного запису інформацією про наступні технологічні етапи його опрацювання.

Створення цифрових об'єктів: оцифрування документів та опрацювання цифрових зображень

За сформованим планом документи замовляються у відповідних підрозділах. Враховуючи цінність документів, переважно більшість оригіналів на початку робочого дня замовляють для оцифрування, а наприкінці дня повертають фондоутримувачам.

Якщо в процесі роботи за будь-якими обставинами (відсутність спеціаліста, вихід з ладу обладнання, форс-мажор) оцифрування документа не може бути здійснено у вказаний термін, до бази даних вноситься відмітка «відкладено». Так замовник має змогу бачити не лише стан виконання, а і перенесення термінів оцифрування.

У процесі роботи важливу роль відіграє призначення копії (онлайнova публікація, страхова копія, службове використання). Відповідно до цих вказівок, планується завантаження існуючого цифрового обладнання та роботи з формування цифрової бібліотеки та страхового фонду.

З урахуванням цінності документа, його формату, фізичного стану і вимог до якості цифрового ресурсу, визначається обладнання, на якому його можна оцифрувати та який спосіб подальшого опрацювання цифрових зображень використовувати. Специфіка обробки зображень різними комплексами програмного забезпечення і підвищення якості зображень обумовлена особливостями кожного типу цифрового обладнання. Швидке та ефективне відцифрування забезпечується відповідною підготовкою документів. Необхідно оцінити документ, його формат, товщину, наявність ілюстрацій, згорнутих вкладок, формат яких іноді значно перевищує формат основного документа, зручність розгортання книги, повноту комплексу сторінок. Часто виникає питання про доцільність оцифрування документа до реставрації. Так, після реставрації часто частина інформації може бути недоступна для оцифрування: після перепрошивки неможливо якісно оцифрувати сторінки біля зрізу книги, а також у тих місцях, де сторінки підклеєні, адже на цифровій копії з'являються білі плями. Всі питання, які виникають у процесі такої попередньої оцінки, необхідно з'ясувати з фондоутримувачами ще до початку оцифрування.

Відповідно до попередньої експертної оцінки, документи за характером їх цифрового опрацювання можна умовно поділити на такі основні групи:

- великоформатні видання (карти, плакати, газети);
- рідкісні та цінні видання (архівні документи, стародруки, рукописи);
- сучасні видання;
- видання кінця XIX – початку XX ст. (підвищеної складності);
- матеріали електронних виставок, фотоматеріали, печатки, тиснення.

На першому етапі оцифрування важливо підготувати комп'ютер та цифрове обладнання, за допомогою пробних знімків налаштувати сканер (цифрову камеру) для досягнення максимальної якості знімків. Після оцифрування обов'язково перевіряється якість і кількість усіх отриманих зображень, невдалі вилучаються та повторно оцифровуються. Це важливо, бо з неякісного цифрового зображення неможливо, навіть за допомогою корекції, отримати якісну цифрову копію. Крім того, згідно з вимогами до страхового фонду цифрових копій, до переважної більшості документів корекція майже не застосовується для забезпечення максимальної відповідності оригіналу. До цінних та унікальних документів неможливо застосовувати режим автоматичного сканування. Щоб зберегти усі особливості оригіналу, добирається спосіб і режим найбільш оптимального оцифрування.

Під час сканування великоформатних видань спочатку оцифровуються окремі блоки: права та ліва сторона документа, а потім перевіряються всі сторінки. Це досить трудомістка операція, особливо якщо оцифровуються стародруки, рукописи, документи іноземною мовою. У цьому випадку візуально неможливо перевірити послідовність тексту, доводиться уважно слідкувати за знаками, специфічними для кожного виду документа (для деяких стародруків характерно, що в кінці попередньої сторінки розміщені декілька літер першого слова наступної сторінки). Якщо ж формат видання невеликий, оцифровуються одразу дві сторінки книги.

У процесі роботи фахівці неухильно дотримуються правил страхового копіювання, зокрема, в процесі сканування особливо цінних видань, архівних документів, стародруків, рукописів під кожен сторінку підкладається аркуш цупкого чорного паперу, на фоні якого можна детально розглянути край сторінки.

Після попереднього оцифрування і перевірки повноти копії відбувається опрацювання цифрових зображень. Кількість операцій залежить від віку, типу документа. Для стародруків та архівних документів застосовуються лише дві операції: вирівнювання і обрізання, але їх не можна робити автома-

тично, тільки вручну. Кожна сторона обрізається окремо. А якщо застосовується для обрізання шаблон, то лише за сторінкою найбільшого розміру, і цим частково автоматизується процес.

Для сучасних документів зі сторінками одного формату, якісним папером і чітким друком застосовується пакетне опрацювання програмами, аналогічними «Book Restorer».

Для цифрового опрацювання зображень використовуються два основних комплекси програмного забезпечення: перший – програма ASDSee; другий – дві програми, які доповнюють одна одну – «Book Restorer» та «Scan Kromsator». Перший спосіб більш зручний для оброблення складних документів, де потрібне нестандартне ручне опрацювання, другий спосіб – автоматизований і краще підходить для сучасних видань з якісним друком, що дає змогу пришвидшити роботу.

Далі настає етап остаточної зв'язки цифрової копії документа з оригіналом, перевірка послідовності сторінок правої та лівої частин документа.

Оцифровані і відкореговані цифрові копії документів розміщуються на файловому сервері у папках, позначених номерами відповідних замовлень для їх подальшої однозначної ідентифікації. Цифрові зображення у папках відповідають вимогам страхової копії і підготовлені для подальшого використання.

Формування страхового фонду цифрових копій

Для довготривалого зберігання цифрових копій документів у НБУВ використовуються оптичні диски (UDO). Вони невибагливі у зберіганні, і виробники гарантують їх надійність упродовж 50 років. Ємність таких дисків 60 Гб, кожна сторона – 30 Гб. UDO комплектується з урахуванням інформації, яка міститься в замовленнях на оцифрування. Особливе значення має колекція, до якої належить даний документ, підрозділ, де зберігається фізичний оригінал документа, і розмір його цифрової копії.

Для оптимізації комплектування UDO у службовій базі даних «Оцифрування документів» створюється окремий вид записів для страхового фонду. У цьому виді записів зазначається інформація про дату запису, номер диска, стан виконання, до яких проектів або колекцій належать документи, записані на диск, з яких вони підрозділів, хто записав диск, відомості про пошкодження та дефекти, якщо такі мали місце. Співробітник групи оцифрування вводить також повну інформацію про зміст UDO.

Зміст поділяється, відповідно до фізичного устрою носіїв інформації, на дві сторони: «Сторона А» та «Сторона В». Кожен рядок змісту містить відомості про цифрову копію документа: стислий бібліографічний опис, місце зберігання оригіналу, шифр документа, кількість файлів, розмір документа, якість зображень (формат файлів, роздільну здатність, глибину кольору), порядковий номер замовлення у базі даних. Вся ця інформація автоматично вводиться із відповідних замовлень на оцифрування документів. У форматі виведення даних автоматично підраховується сумарний обсяг цифрових зображень, за цими цифрами оператор бази даних може слідкувати за комплектацією відповідної сторони UDO за обсягом. Якщо обсяг наближається до 30 Гб, то необхідно комплектувати іншу сторону диска або перейти до комплектування нового диска [1; 3; 4].

Розподілені для UDO цифрові копії документів розміщуються у папках з назвами «Side А» та «Side В», які, у свою чергу, розміщені в папці, позначеній номером UDO. Перед записом UDO необхідно підготувати. Під час цих дій диск форматується, записується його номер і пароль. Після процесу підготовки на диск копіюються відібрані файли. Запис однієї сторони UDO займає від 2 до 5 годин, залежно від кількості файлів.

Записані UDO розміщуються на зберігання на полицях сейфа у спеціалізованому сховищі. Дані, внесені у записи змісту UDO, надалі використовуються для оформлення офіційних описів страхового фонду цифрових копій.

Формування цифрової бібліотеки

Велике значення для отримання якісних електронних версій документів, призначених для онлайн-публікації, мають різні способи опрацювання зображень. У НБУВ електронні версії документів виготовляються у двох основних форматах: PDF та флеш. Формат PDF було обрано для представлення сучасних видань – він легко і зручно переглядається браузерами та індексується пошуковими системами, такими як Google. Формат флеш (flip-book) – книги з гортанням сторінок – обрано для перегляду унікальних документів, наприклад, з рукописного фонду та фонду стародруків. Цей формат дає змогу з достатньою якістю відобразити зміст оригіналу, але не дає змоги копіювати та розповсюджувати його цифрову копію. Виготовлення флеш-версії документа також потребує попереднього створення документа у PDF-форматі. Всі виготовлені PDF-файли зберігаються в архіві та записуються на UDO.

Виготовлена електронна версія документа та зображення обкладинки розміщуються на веб-сервері НБУВ. У базі даних «Оцифрування документів» робиться відмітка «опубліковано» та інтернет-посилання до повного тексту документа, розміщеного на порталі НБУВ.

Для надання користувачам бібліотеки доступу до створених електронних документів інформація про нові надходження цифрових об'єктів розміщується у цифровій бібліотеці НБУВ. До останньої експортуються записи з бази даних «Оцифрування документів». Нові надходження розподіляються за відповідними електронними колекціями. Оскільки різні документи мають свої особливості опису, співробітники відділів-фондоутримувачів переглядають описові метадані, редагують і доповнюють їх у разі необхідності. Записи поповнюються анотацією, інформацією про зміст, предметними рубриками та класифікаційними індексами, відомостями про розміщення цифрових копій документів із фондів НБУВ у міжнародних цифрових проєктах. У базі даних цифрової бібліотеки також уможливлено представлення метаданих цифрового ресурсу трьома мовами (українською, російською та англійською), відповідно до стандартів Світової цифрової бібліотеки (WDL – World Digital Library) [3; 4].

Облік цифрових ресурсів та звіти

База даних «Оцифрування документів» завдяки введенню в процесі технологічного опрацювання цифрового ресурсу всіх необхідних метаданих уможливорює автоматичний облік робіт та формування широкого спектру звітів.

Головний інформаційний блок становлять звіти за датою виконання робіт: місячні, кварталні, піврічні та річні. Відібрані за обраним діапазоном дат або датою записи можна згрупувати за будь-яким необхідним додатковим показником: за видом документів, колекцією, відділом-фондоутримувачем, виконавцем, цифровим обладнанням. Аналогічні звіти можна отримати за датою онлайн-публікації ресурсів або запису на UDO страхової копії документа. Легко можна виокремити документи, які пройшли опрацювання, але ще неопубліковані на порталі НБУВ, що дає змогу формувати найближчі плани онлайн-публікацій [2].

За базою даних можна також отримати річні довідки про оцифрування архівних документів або книжкових пам'яток.

Традиційний річний звіт групи з оцифрування документів містить інформацію, розподілену за видом

документів про кількість оцифрованих документів та отриманих цифрових зображень, про кількість підготовлених до публікації документів, кількість опублікованих на порталі НБУВ документів.

Ретельно введений зміст UDO дає змогу отримувати в автоматизованому режимі описи надходжень до страхового фонду цифрових копій документів із фондів НБУВ. Опис одиниці зберігання страхового фонду «Диск UDO № _____» складається з двох розділів «Сторона А» та «Сторона В». Кожен розділ містить згрупований за назвою підрозділу НБУВ список записаних документів: шифр, місце зберігання, стислий опис, назва папки / файлу, формат файлів, кількість цифрових зображень, обсяг у байтах. В кінці кожного розділу наводяться підсумкові дані про кількість документів, кількість зображень та загальний обсяг у байтах. Опис завершується відомостями про його загальний обсяг та місце зберігання диску, про осіб, відповідальних за формування та зберігання страхового фонду.

Висновки

Розроблена в НБУВ технологічна модель (рисунок) опрацювання цифрових ресурсів передбачає упорядкування та опис усіх етапів робіт, що значно полегшує супровід всього технологічного циклу бібліотечних цифрових ресурсів від комплектування до онлайн-публікації.

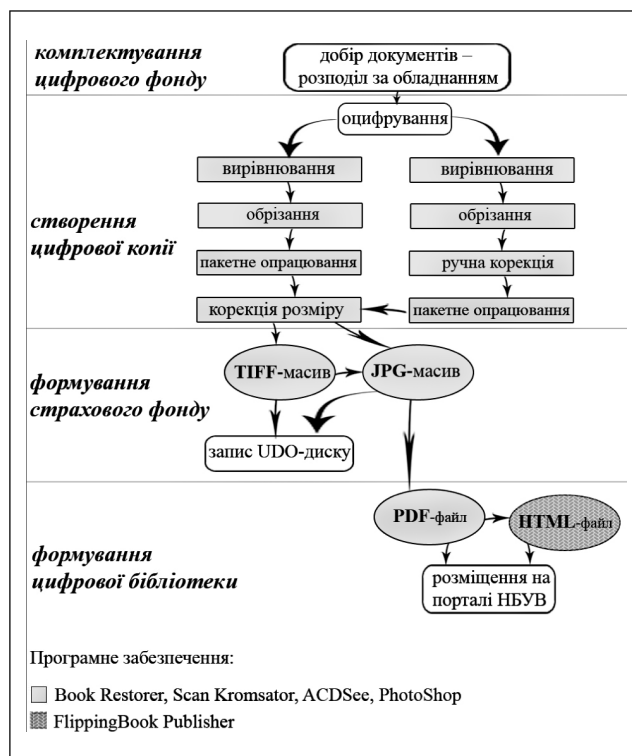


Схема технологічного циклу формування цифрового фонду НБУВ

Запропонована модель організації роботи відповідає світовим тенденціям у підходах до формування цифрових ресурсів бібліотек, де велика увага приділяється процесам цифрового кураторства та збереження цифрових ресурсів для майбутніх поколінь. Інвентаризація та введення метаданих, зазначення всієї необхідної інформації про технологію створення та формати цифрових ресурсів створює передумови для постійної підтримки електронних колекцій наукової бібліотеки в актуальному стані.

Список використаних джерел

1. *Лобузин І. В.* Створення, представлення та перспективи використання електронного фонду цифрових копій документів бібліотеки / І. В. Лобузин // Реєстрація, зберігання і обробка даних. – К., 2011. – Вип. 2. – С. 60–73.
2. *Лобузін К. В.* Інформаційні технології в організації роботи спеціалізованих підрозділів Національної бібліотеки України імені В. І. Вернадського / К. В. Лобузін, І. В. Лобузин // Наукові праці Національної бібліотеки України імені В. І. Вернадського. – К., 2011. – Вип. 29. – С. 173–184.
3. *Лобузін К. В.* Оцифровка библиотечных фондов как средство предоставления доступа к историко-культурному и научному наследию / Е. В. Лобузін // Библиотеки национальных академий наук: пробл. функционирования, тенденции развития: науч.-практ. и теор. сб. – К., 2012. – Вип. 10. – С. 222–231.
4. *Лобузін К. В.* Технології організації знаннєвих ресурсів у бібліотечно-інформаційній діяльності: монографія / К. В. Лобузін; відп. ред. О. С. Онищенко; НАН України, Нац. б-ка України імені В. І. Вернадського. – К., 2012. – 252 с.
5. *Лобузін К. В.* Формування електронних інформаційних ресурсів рукописної та архівної спадщини / К. В. Лобузін // Рукописна та книжкова спадщина України. – К., 2012. – Вип. 16. – С. 234–245.
6. *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; no. 170).
7. *Reese T.J.* Collection Management / T. J. Reese, K. Banerjee. – New York : Neal-Schuman Publishers, 2008. – 277 p.
8. *Smith A.* Strategies for Building Digitized Collections / Digital Library Federation; Council on Library and Information Resources. – Washington (DC), 2011. – 35 p.