



НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
«КИЇВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Політика оцифрування та збереження цифрових даних

Версія 01.03 від 01 серпня 2017

ЗМІСТ

Зміст

| | |
|----------------------------------------------|---|
| 1. Вступ | 2 |
| 2. Призначення та область застосування | 2 |
| 3. Нормативна база..... | 3 |
| 4. Принципи та завдання | 3 |
| 5. Підходи до оцифрування..... | 4 |

Автор: Чорна Тетяна, директор наукової бібліотеки НаУКМА

*Схвалено рішенням Робочої групи по створенню політики оцифрування та збереження
цифрових даних в НаУКМА*

1. Вступ

Академічні бібліотеки з унікальними цінними та рідкісними бібліотечними фондами відіграють все більш важливу роль у створенні наукового контенту в Інтернет. На сьогоднішній день постає необхідність для створення цифрових колекцій, що сприятимуть навчально-освітньому процесу та науковим дослідженням студентів, викладачів та співробітників університету. Рух Відкритого доступу до інформації, активне створення програм для оцифрування та зростання електронних публікацій у мережі Інтернет вимагають активізації процесу оцифрування бібліотечних, музейних та архівних фондів. Оцифрування цих фондів створює багато переваг та можливостей для користувачів НаУКМА. Оцифрований контент використовується для підтримки викладання, дослідження та залучення громадськості. Існує величезний потенціал для подальшого використання оцифрованих ресурсів у рекламі, виставках, залученні громадськості та фандрайзингу НаУКМА. Інвестиції в оцифрування дають прибуток з точки зору відкритості, доступності та впливу цих ресурсів на навчальний та дослідницький процес в університеті і можуть використовуватися одночасно більшою кількістю користувачів без обмежень у часі і просторі. Світові академічні бібліотеки швидко рухаються в напрямку оцифрування та представлення своїх фондів у відкритому доступі, і дуже важлива вимога часу - щоби НаУКМА розвивав та підтримував свій високий рівень у цій галузі.

Наукова бібліотека Національно університету «Києво-Могилянська академія» протягом останніх 10 років працює над оцифрування матеріалів, які необхідні для навчального процесу та створенням цифрової колекції рідкісних та цінних видань з бібліотечних фондів. Політика оцифрування та збереження цифрових даних намагається формалізувати підхід НаУКМА до цього напрямку та встановити основу для її діяльності щодо оцифрування протягом наступних п'яти років.

2. Призначення та область застосування

2.1. Дана Політика сформована спираючись на власний досвід та вивчаючи досвід світових і вітчизняних проєктів з оцифрування.

2.2. Цією Політикою визначено принципи, які будуть формувати стратегію розвитку Наукової бібліотеки НаУКМА в галузі інноваційних бібліотечних послуг та інтегруватися зі стратегічним планом розвитку НаУКМА в частині загально-університетського проєкту "Цифровий університет".

2.3. Цей документ визначає політику та регламентує діяльність в процесі оцифрування та збереження цифрових даних для наукових досліджень та громадськості. Вони спрямовані на вирішення всього життєвого циклу оцифрованого контенту, від відбору для оцифрування до довгострокового збереження оцифрованого вмісту.

2.4. Політика оцифрування та збереження цифрових даних базується на вимогах законодавчих, регуляторних та нормативних документів з бібліотечної, архівної та музейної справ.

3. Нормативна база

- 3.1. Політика розроблена у відповідності до вимог чинного законодавства України, а саме:
- Законів України «Про культуру», «Про інформацію», «Про авторське право і суміжні права», «Про бібліотеки і бібліотечну справу», «Про Національний архівний фонд та архівні установи», «Про музеї та музейну справу», «Про охорону культурної спадщини»;
 - нормативно-правових актів з бібліотечної, архівної та музейної справ;
 - інших нормативних документів, що регламентують вимоги до оцифрування та збереження цифрових даних.

4. Принципи та завдання

- 4.1. Політики оцифрування та збереження цифрових даних стосуються всієї оцифрованої колекції документів та окремих оцифрованих елементів, які становлять історико культурне надбання, забезпечують навчальний процес та вимагають довготривалого зберігання.
- 4.2. Оцифрування має на меті наступне:
- підвищення якості обслуговування користувачів;
 - покращення зберігання оригіналів та можливість відновлення їх після втрати;
 - розповсюдження інформації про рідкісні та цінні колекції НаУКМА серед міжнародної наукової спільноти.
- 4.3. Створення цифрових копій документів та об'єктів має наступні переваги:
- зменшується зношеність оригіналів, оскільки оцифрування відбувається один раз, і в подальшому користувачам надаються у користування цифрові копії. Створюється можливість доступу до тих об'єктів, оригінали яких знаходяться в незадовільному фізичному стані і на запит користувачів не видаються;
 - цифрова копія має необмежений термін зберігання;
 - при втраті оригіналу є можливість його відновити поліграфічним засобом;
 - можливість розпізнавання образів (у т.ч. тексту) в цифрових копіях дозволяє застосовувати пошук по тексту, що перетворює послуги, які надаються користувачам на якісно новий рівень.
- 4.4. Оцифрування ставить наступні завдання:
- визначення параметрів відбору документів та об'єктів для оцифрування;
 - аналіз та експертиза документів та об'єктів з метою визначення матеріально-технічної бази, необхідної для оцифрування документів та вимог, щодо електронного контенту;
 - забезпечення матеріально-технічної бази (придбання обладнання та його налаштування);
 - програмне забезпечення та його налаштування для представлення цифрового контенту;
 - навчання, підвищення кваліфікації, стажування співробітників, що займаються оцифруванням;
 - планування та впровадження системи звітності;
 - управління цифровим контентом.

5. Відбір контенту на оцифрування та експертиза

5.1. Наукова бібліотека активно співпрацює з академічною спільнотою НаУКМА, для того щоб визначити пріоритети для оцифрування унікальних колекцій. Головним напрямком для програми оцифрування повинно бути залучення зовнішнього фінансування через грантову діяльність, що дозволяє зробити доступними великі частини колекцій.

5.2. Пріоритетом у відборі контенту для оцифрування становлять історичні джерела, що репрезентують історію Києво-Могилянської академії на всіх етапах її існування: Києво-Могилянська академія, Київська духовна академія, Національний університет «Києво-Могилянська академія». Контент для оцифрування буде відбиратися з різних джерел: власних академічних колекцій, з інших українських та іноземних організацій, установ, приватних колекцій. До цифрової колекції буде долучено рідкісні та цінні книжкові, архівні, археологічні, історичні, мистецькі та меморіальні колекції.

5.3. Фізичні формати включатимуть, але не обмежуватимуть, друковані книги, рукописи, карти, фотографії, творів мистецтва, аудіо- та відеозаписи. Для кожного формату створено спеціальні технічні вимоги. Додаток 1.

5.4. При відборі для оцифрування фізичний стан рідкісних і цінних документів та об'єктів буде враховуватися як наступний рівень пріоритетності.

6. Підходи до оцифрування

6.1. **Обов'язковий:** отримання копій документів та об'єктів у вигляді графічних зображень, які створюються шляхом сканування чи фотографування з наступною їх обробкою та зберігання в одному з форматів графічних файлів. У цьому випадку повністю зберігається оригінальна верстка книги або оригінальне відтворення об'єкту.

6.2. **Пошуковий (факультативний):** оптичне розпізнавання тексту з послідовним збереженням розпізнаного тексту як графічного набору символів. У цьому випадку з'являється можливість повнотекстового у документі та індексація великих масивів електронних документів. Для повної ідентичності змісту потребує ручної перевірки.

7. Типи електронних копій

Для об'єктів зберігання в межах планового оцифрування можливе виготовлення як мінімум 3 типів цифрових копій, які відрізняються за об'ємом інформації в залежності від призначення та особливостей використання.

7.1. **Майстер-копія.** Ця копія містить максимально можливу кількість інформації. Вона використовується для відновлення об'єкта у випадку його втрати, для іншої поліграфічної мети, для деяких видів досліджень та як основа для виготовлення інших типів цифрових копій. До файлів, які містять архівні копії, не застосовується стиснення з втратами. Дані копії розміщуються на носіях з тривалим терміном існування. З метою збереження авторських прав доступ до копій даного типу обмежений, а користувачам вони видаються в особливих випадках.

7.2. **Користувацька копія високого або середнього розрешення.** Ця копія створюється з майстер-копії для цінних та рідкісних видань. Вона призначена для представлення

користувачам в базах даних та електронних бібліотеках. Допускається застосування стиснення з втратами, але при цьому повинна зберегтися читабельність тексту і можливість розрізнати деталі графічних елементів. Об'єм файлу повинен відповідати пропускній можливості локальної комп'ютерної мережі (зображення повинно завантажуватися не більше 3-5 секунд).

7.3. Службова копія низького розрішення. Ця копія створюється з майстер-копії або користувацької копії. Вона призначена для представлення її на веб-сайті та для попереднього перегляду при прошенні її в локальній мережі, в тому числі і для представлення користувачам в базах даних та електронних бібліотеках. Допускається застосування стиснення з втратами, зменшення розрішення графічного зображення, впровадження в документ додаткових знаків (водяних, тощо).

8. Обладнання

8.1. Оцифрування поділяється на основні типи: контактне і безконтактне. Застосування контактного сканування дозволяється у тому випадку, якщо:

- поверхня оцифрування плоска або може бути випрямлена без шкоди об'єкту;
- поверхня оцифрування не може бути пошкоджена контактуючи зі склом;
- світло пристрою не шкодить об'єкту;
- для спеціальних видів об'єктів, які потребують максимального розрішення (більше 1000 dpi, наприклад, слайдів).

8.2. В решті випадках повинно застосовуватись безконтактне сканування.

8.3. Освітлення, яке застосовується для безконтактного сканування, повинно бути підібране таким чином, щоб забезпечити якісне освітлення об'єкта, але при цьому світло не повинно шкодити об'єкту.

9. Обробка цифрових копій, розміщення на носіях, зберігання та контроль за зберіганням носіїв

9.3. Подальша обробка цифрових майстер-копій проводиться з метою покращання якості відтворення копії на екрані, в той же час не повинна втрачатись інформація, яка міститься в цифровій копії. Діапазон можливих способів обробки копій низького та середнього розрішення досить широкий, допускається деяка втрата інформації.

9.4. Типовими діями програмної обробки майстер-копій є:

- вирівнювання зображення по полю, якщо необхідно;
- зменшення розмірів зображення за рахунок неінформаційних полів;
- видалення зайвих частин зображення, якщо такі існують (наприклад, частина сусідньої сторінки, зайві предмети, що потрапили в кадр і т.п.), але не допускається втрата інформації;
- розширення діапазону яскравості зображення до максимально можливого;
- збереження зображення у файлі, які визначено для формату майстер-копії (не допускається стиснення з втратами)

9.5. Після обробки та запису майстер-копії, на її основі створюються копії середнього та низького розрішення. Цей процес передбачає наступне:

- зменшення розмірів зображення шляхом зниження розрішення та/або видалення полів, які вважаються неважливими для цієї копії, при цьому слід слідкувати, щоб текст на екрані залишався читабельним, а деталі графіки були розбірливими;
- вирівнювання країв зображення;
- зведення до єдиного розміру зображень, які стосуються одного видання. Для видань, які мають великий корінець або нерівно розміщені сторінки, розмір графічного зображення визначається за найбільшим з зображень. Тому для таких видань дозволяється збільшення чорних полів для деяких зображень;

- збірка зображень для форматів, які перевищують формат поля сканера;
- зменшення розрешення;
- зменшення розмірів зображення в пікселях або сантиметрах відносно оригінального розміру (наприклад, 50% від оригінального розміру);
- зміна типу та ступеня стиснення графічного файлу;
- створення в документів додаткових даних (наприклад, водяні знаки, метадані і т.п.)

9.6. Записуючи цифрові копії на носії, надається перевага групуванню файлів, які стосуються одного об'єкта, на одному носіїві або носіях з суміжними номерами. Після закінчення запису обов'язково проводиться перевіряється читаність зображення з носія. Кожному носіїві визначається унікальний номер.

9.7. Кількість екземплярів майстер-копії залежить від збереження та унікальності оцифрованого об'єкта. Копії для об'єктів, які стосуються цінних та рідкісних видань, рукописів, гравюр і т.п., а також для об'єктів, повторне оцифрування яких викликає труднощі або з інших причин не бажане, розміщуються в двох екземплярах на різних носіях (дублюється весь вміст носія). Це дозволяє практично виключити ймовірність втрати копії у випадку пошкодження носія.

9.8. Стан носіїв перевіряється не менше ніж раз на п'ять років. При виявленні ознак пошкодження носія інформація переноситься на інший носій, перевіряється ідентичність файлів на двох носіях та їх кількість.

10. Параметри оцифрування

10.3. Якщо не можливо оцифрувати об'єкт в цілому, допускається оцифрування його окремими частинами (з перетином сторінок). Частина зберігаються або окремо, або накладаються на одне зображення так щоб між ними залишався проміжок у кілька пікселів.

10.4. Книги скануються повністю, включаючи лицьову та тильну сторінки обкладинки, титульну сторінку, кінцевий титульний аркуш (колофон), зворотню сторону переплетеної кришки, вклейки та зворотні сторони вклейок і т.п. Пусті сторінки (вакати) також підлягають скануванню, незалежно від наявності на них номерів чи іншої значної інформації.

10.5. Вшиті чи вклеєні в оригінал аркуші, на яких розміщена логічна суцільна інформація (карта/зображення, кілька зображень під однією назвою, таблиця) повинні бути представлені як єдине зображення. Вшиті чи вклеєні в оригінал аркуші, інформація на яких логічно не пов'язана, повинні бути представлені як два зображення. Зустрічаються в книгах сторінки, які містять ілюстрації, що розміщуються повністю на 2 сторінках, скануються як дві сторінки.

10.6. Текстові сторінки книг скануються з налаштуванням яскравості та контрасту, які виставлені для максимальної ефективності розпізнавання тексту.

10.7. Графічні об'єкти повинні бути чітко сфокусовані (різкість текстових графічних об'єктів повинна фокусуватися по тексту). Відтінки, глибина та насиченість кольорів об'єкту повинні бути максимально наближені до оригіналу.

10.8. Можливі 2 варіанти встановлення параметрів поля оцифрування для зшитих видань: по межі важливої інформації (можливо для сучасних видань), або по межі документа, включаючи каскад сторінок, обкладинку та максимально фон сканування (рекомендовано для всіх рідкісних та цінних видань).

10.9. Для сторінок одного формату графічні зображення повинні мати однаковий розмір в пікселях по висоті та ширині. Єдиний розмір повинен бути обраний з найбільшого зображення сторінки з полями не більше 1-3 мм.

10.10. Зображення великих форматів або малих форматів сторінок, які відрізняються розміром від основної частини сторінок джерела сканування, повинні зберігати оригінальний розмір.

10.11. В результаті сканування книг оригіналів отримують графічні об'єкти у

форматі TIFF (LZW-compressed) в посторінковому представленні з наскрізною нумерацією файлів-зображень сторінок.

- 10.12. Після сканування виконуються процедури автоматичного вирівнювання отриманих зображень горизонтально та обрізання у формат видання (з врахуванням полів).
- 10.13. Не допускається зберігання кількох зображень в один файл, сканування з різним розрішенням в межах одного об'єкта.
- 10.14. При оцифруванні допускається застосування апаратної корекції гама. Кольорова корекція, індивідуальна програмна корекція зображень не застосовується.

11. Режим кольору

- 11.3. Режим Grayscale допускається, якщо на об'єкті, що оцифровується немає ніякої значної кольорової інформації. В усіх решту випадках використовується режим Color RGB.
- 11.4. Оператор сканування має право самостійно прийняти рішення про необхідність кольорового сканування видання у випадку поганого контрасту (наприклад, коричневий колір на жовтому фоні).
- 11.5. Всі видання скануються в режимі 24-bit RGB (кольоровий режим), якщо інше не зазначено.

12. Розрішення майстер-копії

12.3.1. У випадку необхідності окремі зображення фотографій/гравюр/карт можуть бути збережені у файлі.

- фотоальбоми та фотографії: максимально можливе в діапазоні 300-700 dpi;
- текст з малюнками та без: 300-450 dpi;
- гравюри: 300-1200 dpi;
- карти, атласи: 300-700 dpi.

13. Формати файлів

В якості основних форматів представлення цифрових зображень від сканованих документів повинні використовуватись формати TIFF, JPG та PDF.

При скануванні використовується TIFF з LZW-компресією, оптичне розрішення 300 dpi та більше.

Для користувацьких та службових копій використовується JPG з компресією збереження 80% якості та при необхідності, зниженням розрішення.

Для представлення копії у вигляді електронної книги повинен використовуватися формат багатосторінкового PDF. Файл PDF повинен бути оптимізований для Web та створений з копій у форматі jpg з параметрами стиснення та розрішення, що дозволені для даного типу копій (користувацька, службова).

| Назва формату та мета його використання | Програми, які можуть відкривати файли | Рекомендований метод стиснення |
|-----------------------------------------|---------------------------------------|--------------------------------|
|-----------------------------------------|---------------------------------------|--------------------------------|

| | | |
|-------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| <p>TIFF Tagged Image File Format</p> <p>Для майстер-копій</p> | <p>Більшість растрових редакторів та видавничі системи; векторні редактори, які підтримують растрові об'єкти</p> | <p>Без стиснення LZW (стиснення без втрати якості)</p> |
| <p>JPEG Joint Photographi с Experts Group</p> <p>Для користувачьких та службових копій</p> | <p>Більшість графічних програм, Інтернет-браузери</p> | <p>JPEG (можливість вибору ступеня стиснення. Не рекомендується стиснення більше 50% від оригіналу)</p> |
| <p>PDF - Portable Document Format</p> <p>Для користувачьких та службових копій</p> | <p>Міжплатформовий формат файлу, який переноситься на комп'ютери будь-яких систем. Для перегляду файлу необхідна програма Acrobat Reader або аналог, в т.ч. плагіни браузера</p> | <p>LZW, JPEG</p> |

14. Вимоги доступності інформації

- 14.1. Файли повинні відкриватися на перегляд стандартними засобами, які призначені для роботи з ними в середовищі операційної системи, починаючи з MS Windows 2000/XP без попереднього виводу на екран будь-яких попереджень чи повідомлень про помилки.
- 14.2. Не допускається у файлах встановлення опції заборони друку вмісту файлу.
- 14.3. Не допускається встановлення у файлах захисту за паролем на відкриття файлу.
- 14.4. Шрифти, ілюстрації та інші файлові об'єкти повинні бути вбудовані у тіло PDF-файлу.
- 14.5. Не припустимі гіпертекстові посилання з одного PDF-файлу на інші файли.
- 14.6. Копіюючи файли з носіїв інформації, переглядаючи їх на комп'ютері чи роздруковуючи не повинно виникати помилок, які пов'язані з неякісною підготовкою до запису на носій, фізичним пошкодженням чи пошкодженням носієм.