

Національні проекти з оцифрування

Metamorfoze (metamorfoze digitalization project) це загальнонаціональна програма Нідерландів, що сприяє збереженню культурного надбання. Програма існує з 1997 р., головний офіс розміщується в Королівській бібліотеці (Національна бібліотека Нідерландів – Koninklijke Bibliotheek). Будь-яка установа пам'яті Нідерландів може внести у список пропозиції щодо створення нових проектів і збереження фондів в офісі Metamorfoze. Після цього комітет з кількох незалежних експертів розглядає та оцінює ці пропозиції і їх релевантність для програми Metamorfoze. Основне завдання комітету це розробка рекомендацій з фінансування пропозицій, які подаються у Metamorfoze. Програма віддає перевагу матеріалам голландського походження, які зберігаються в установах культурної спадщини: наукових бібліотеках, великих регіональних або місцевих бібліотеках, архівних установах, регіональних історичних центрах, історичних центрах документації, музеїв та науково-дослідних інститутів.

Національна інфраструктура цифрової інформації та програма збереження [The National Digital Information Infrastructure and Preservation Program (NDIIPP)] – це програма, яку веде Бібліотека Конгресу США з архівування і надання доступу до цифрових ресурсів. Конгрес США організував програму у 2000 р. Бібліотека Конгресу як одна з провідних постачальників високоякісного контенту в Інтернеті була обрана головним офісом проекту. У липні 2010 р. бібліотека організувала Національний електронний опікунський Альянс (National Digital Stewardship Alliance (NDSA)), щоб залучити до проекту більше установ. Основними цілями NDSA є розвиток існуючих стандартів і методів збереження, для цього організована співпраця з експертами для визначення форматів цифрової інформації, які є найбільш придатними для збереження. Крім того NDSA бере участь в об'єднанні цифрового контенту у загальнонаціональній колекції. Серед членів NDSA є університети, професійні асоціації, комерційні підприємства, консорціуми, і державні установи.

Digidaily проект з оцифрування шведських газет, тривав з 2010 р. по 2013 р. Digidaily розвивався в Media Conversion Center (МКС), Національному архіві (Riksarkivet in Swedish) і Національній бібліотеці Швеції (Kungl. biblioteket in Swedish), де і було розроблено раціональні методи й процеси оцифрування та інтерпретації тексту газет. Оцифровувалися відібрані видання шведських газет, як оригіналів так і копій з мікрофільмів. SAP – проект з оцифрування шведсько-американської преси, розвиток якого підтримується в Центрі конверсії медіа ресурсів (Media Conversion Center (МКС)), Національному архіві (Riksarkivet in Swedish) і Національній бібліотеці Швеції (Kungl. biblioteket in Swedish). Національна бібліотека Швеції співпрацює з Europeana у проекті EUscreenXL.

У проектах SAP і Digidaily, для опису цифрових об'єктів використовується декілька стандартів метаданих – METS, PREMIS, MODS, MIX, ALTO. [35]

ELuxemburgensia – це другий проект з оцифрування в Національній бібліотеці Люксембургу (Bibliothèque nationale de Luxembourg (BnL)) що був відкритий у грудні 2009 р. Перший – luxemburgensia був менш досконалим, тепер є можливість шукати не тільки за змістом, а й повнотекстовим пошуком через новий інтерфейс на www.eluxemburgensia.lu. Крім того, оцифровані ресурси були інтегровані в систему цифрового архівування. Для реалізації повнотекстового пошуку оцифрований контент було повністю переведено в текстовий формат за допомогою програмного забезпечення оптичного розпізнавання символів. Цілями проекту є фізичне збереження друкованої спадщини, забезпечення доступу в Інтернеті (у межах регулювання авторських прав), забезпечення сучасних методів пошуку цифрового змісту.

Національна бібліотека Люксембургу використовує METS і ALTO у своїх проектах з оцифрування. Формат METS дає змогу структурувати й шукати в логічній структурі документів (сторінок, статей тощо). Він також відстежує технічні метадані (права, цифрове обладнання тощо). Формат метаданих ALTO описує структуру окремих сторінок і зберігає результат обробки програмами OCR (Optical Character Recognition). Поєднання METS і ALTO дає можливість здійснити пошук за специфічними термінами, які містяться в оригінальній статті на зображеннях сторінок.

У *The Australian Newspaper Digitisation Program* (Австралійська програма з оцифрування газет) розроблено сервіс, який забезпечує інтернет- доступ до газет. Державні й територіальні бібліотеки та інші власники мікрофільмів майстер-якості з копіями газет забезпечили мікрофільми необхідних примірників для оцифрування. Цифрові зображення перетворили в повнотекстовий контент за допомогою OCR. Контент аналізується людиною оператором і розбивається на статті, сегменти, сторінки (зони) і створює метадані для статей.

Газети описуються таким чином. XML файл для кожної сторінки, відповідає схемі ALTO і містить результати OCR обробки. Файл XML для кожного блоку газет, відповідає схемі METS і містить велику частину створених людиною метаданих для кожного випуску (*Підсумки другого науково-практичного семінару «Оцифроване надбання: збереження, доступ, репрезентація» // Національна бібліотека України ім. В. І. Вернадського (<http://www.nbuv.gov.ua/node/1969>)*).