

ЦИФРОВЕ ЗБЕРЕЖЕННЯ ЕЛЕКТРОННИХ ДОКУМЕНТІВ НА ПРИКЛАДІ ДІЯЛЬНОСТІ НАЦІОНАЛЬНОЇ БІБЛІОТЕКИ ФРАНЦІЇ

В статті висвітлена документальна політика електронної бібліотеки Національної бібліотеки Франції, характеризується процес створення та сучасний етап розвитку електронної бібліотеки Gallica, розглядається система довготривалого збереження цифрових ресурсів SPAR, зокрема базові принципи її створення та функціонування, методи та стратегії збереження електронних документів, файли та формати, можливості цифрового співробітництва.

Ключові слова: Національна бібліотека Франції, електронна бібліотека, Gallica, Gallica Labs, оцифровка документів, довготривале збереження цифрових ресурсів, системи архівації OAIS, SPAR.

L.P. Khaletska

DIGITAL PRESERVATION OF E-DOCUMENTS BASED ON AN EXAMPLE OF THE BIBLIOTHÈQUE NATIONALE DE FRANCE

The article highlights the policy of the bibliotheque nationale de France, the establishment and modern development of e-library Gallica is overviewed; the policy of Bibliotheque Nationale de France regarding documents, the long-term saving system of digital resources SPAR is being described – the basic principles of its creation and day-to-day operations; methods and strategies of digital documents' safekeeping, files and their descriptions; the possibilities of digital co-operation.

Keywords: Bibliothèque nationale de France, e-library, Gallica, Gallica Labs, digitalizing of documents, long-time preservation of digital resources, archivation systems OAIS, SPAR.

Створення і користування різними видами електронних документів стало нормою сьогодення. Підприємства, установи та організації, музеї, архіви та бібліотеки накопичують значний обсяг даних, представлених у цифровому форматі. У цих умовах досить високим є ризик втрати даних, на створення чи придбання яких іноді витрачаються значні зусилля і кошти. Тому надзвичайної актуальності набуває питання необхідності довготривалого збереження цифрових ресурсів.

З точки зору впровадження систем та стратегій довготривалого збереження електронних документів корисним для вітчизняних установ буде знайомство з відповідним зарубіжним досвідом, перш за все з досвідом великих національних установ, які протягом

тривалого часу задіяні як у міжнародних, так і в національних проектах з розробки систем довготривалого зберігання цифрових даних. Однією з таких установ є Національна бібліотека Франції.

Аналіз вітчизняної бібліографії з визначеної теми дозволяє зробити висновок про її актуальність з точки зору наукових досліджень та практичного впровадження наукових розробок. Зокрема, інтернет-технологіям бібліотечного сервісу в Україні присвячене дослідження Д.В. Солов'яненка [7], технологіям архівного збереження електронних бібліотечних ресурсів у мережевому комунікаційному просторі присвячене ґрунтовне дисертаційне дослідження Н.О. Стрілець [8]. Остання висвітлює значний європейський та американо-канадський досвід, накопичений у цій сфері. Проблема зберігання електронних документів розглядається у роботах Ю. Ковтанюка [1], Г.В. Лавренюк та Л.В. Жук [2], П. Марченка [4] та інших українських дослідників. Безпосередньо основним напрямом діяльності Національної бібліотеки Франції присвячена стаття Н. Малолетової [3], яка, крім іншого, висвітлює питання автоматизації бібліотечних процесів, створення цифрових документів, відбір документів для європейської цифрової бібліотеки Europeana, створення цифрової бібліотеки Gallica. Проте у вітчизняних дослідженнях не знайшло відображення діяльності Національної бібліотеки Франції з питання створення та використання системи збереження цифрових документів. Отже, нашою *метою* стало висвітлення на прикладі діяльності Національної бібліотеки Франції документальної політики цифрової бібліотеки загалом, а також проблеми створення та функціонування Системи збереження і розподільного архівування SPAR.

Національна бібліотека Франції (НБФ) – одна з найстаріших у Європі, центр національної мережі бібліотек Франції, найбагатше зібрання літератури французькою мовою в світі. Колекції НБФ налічують понад 14 млн. книг та друкованих видань, а також рукописи, гравюри, фотографії, плани тощо [3, 20].

У 1980-х рр. почалася автоматизація бібліотечних процесів. Протягом 1987 – 2002 рр. для поточної каталогізації використовувалися три бібліографічні бази: BN–OPALE, BN–OPALINE, GEAC–Libre accès [3, 23].

Серед понад 50 європейських проектів, у яких бере участь НБФ, найбільш відомим є Europeana – створення електронної європейської бібліотеки, започатковане ЄС у 2008 р. [3, 24-25]. НБФ бере участь у діяльності міжнародних організацій, які, зокрема, проводять роботи в галузі створення та збереження цифрових ресурсів. Серед них Міжнародна федерація бібліотечних асоціацій та організацій (IFLA). До її програми PAC (Preservation and Conservation), метою якої є довготривале збереження документів на всіх видах носіїв, НБФ приєдналася у 1992 р. [3, 25]. У 2006 р. з ініціативи національних бібліотек Бельгії, Канади, Франції, Люксембурга, Квебека та Швейцарії, зібрання представників яких відбулося в

Парижі з ініціативи НБФ, була створена Франкомовна інформатизована мережа (RFN – Réseau francophone numérique), яка об'єднала 25 найбільших документознавчих установ франкофонії. Її метою є забезпечення збереження та доступності франкомовних оцифрованих документів для наступних поколінь [14].

З 1997 р. НБФ розпочала створення власної цифрової бібліотеки Gallica. Ідея повністю віртуальної бібліотеки була запропонована Жаком Атталі (економіст, письменник, радник Ф. Міттерана у 1981 – 1991 рр.). до неї знову повернулися в рамках проекту створення нової публічної бібліотеки Франції у 1989 р. була поставлена мета модернізації пошуку на основі застосування нових інформаційних електронних засобів, а також демократизації доступу до писемної спадщини [13, 93]. Документальна політика електронної бібліотеки НБФ базується на наступних принципах [13, 94]:

- цифрове відтворення найрізноманітніших видань для забезпечення праці дослідників порівняльним способом;
- першочергова оцифровка текстів найчастіше використовуваних чи з ускладненим доступом;
- представлення в електронному вигляді творів, які презентують ту чи іншу соціально-культурну сферу;
- забезпечення доступу на відстані до документів електронної бібліотеки;
- поряд із створенням енциклопедичного електронного фонду представити тематичні зібрання текстів і зображень.

Для створення електронної бібліотеки були організовані дві проектні групи: перша займалася придбанням документів, друга – інформаційною концепцією і розробкою етапів виробництва електронної бібліотеки. Сьогодні електронна бібліотека інтегрована в діяльність одного структурного підрозділу – Департаменту електронної бібліотеки, який здійснює управління та координацію діяльності, пов'язаної з оцифровкою [13, 157]. До 2001 р. оцифровка документів здійснювалася способом зображення (факсиміле) розрішенням від 300 до 400 dpi, однак зміст книг і періодичних видань оцифровувався текстовим способом [13, 157-158]. Більшість документів була оцифрована не на місці, а надана у тимчасове користування в результаті звернення до європейських пропозицій. Служба оцифровки спостерігає за якістю документальних джерел та за якістю оцифрованих документів. Електронні дані записуються на касети DAT, CD-Worm, зміст яких передається на сервери інформаційної системи (SI). SI контролює якість файлів, які передаються, перед їх використанням, бере участь у розробці та реалізації концепції програмного забезпечення передачі даних через Internet та їх оцінки. Департамент електронної бібліотеки проводить також попередній відбір методів збереження електронних даних [13, 158].

На сьогодні Gallica надає в користування 3 480 461 документ, серед яких 605 942 книг, 88 614 карт, 56 877 рукописів, 1 166 736 зображень, 1 499 434 примірника періодичних видань, 28 472 партитури, 34 386 звукозаписів [12]. Протягом 2015 р. має бути реалізована нова версія Gallica Labs, яка спрощує способи користування електронними документами. За її допомогою сьогодні можна переглядати унікальний прилад перегляду документів, який дозволяє обрати різні способи виведення документа на екран дисплея (посторінково, по дві сторінки, вертикальне прокручування, мозаїчний вигляд). Серед цих способів новим є одночасний перегляд електронного документа і у вигляді зображення, і у текстовому вигляді. Крім того, спрощуються способи користування електронними документами завдяки покращенню навігації змісту та завдяки можливості гортати документ за допомогою клавіатури. Інтерфейс Gallica Labs прагне бути доступним для осіб з обмеженими можливостями [16]. Крім цього, весь оцифрований матеріал Gallica доступний для користувачів в аудіо форматі: з'явилася можливість посторінкового «прослуховування» документів, що особливо важливо для користувачів зі слабким зором [6].

Розвиток електронної колекції, а також поява публікацій виключно в електронній формі (за прогнозом дослідження, замовленого Британською бібліотекою у 2004 р., на 2016 р. половина академічної періодики існуватиме тільки в електронному форматі [5]) зумовлюють необхідність пошуку методів тривалого зберігання електронних даних. З цією метою НБФ обрала ком пакт-диски, виготовлені з міцного матеріалу тривалого використання – Century Disc. Однак зберігання оцифрованих даних вимагає не лише збереження носіїв інформації, а також постійності доступу до інформаційних кодів, які утворюють зміст (дані, системи використання та індексації тощо). У міжнародному аспекті НБФ застосовує два методи: міграцію даних і розвиток програмного забезпечення, що дозволяє використовувати їх зміст, та емуляцію технічного та програмного середовищ документів. Необхідність розробки спільних норм зумовлює участь НБФ в програмах європейських досліджень, прикладом яких є NEDLIB (Networked Deposit Libraries) [13, 146].

Таким чином, створення та збереження оцифрованих документальних ресурсів є одним з пріоритетних напрямів діяльності НБФ. Результатом багаторічних досліджень засобів тривалого збереження електронних документів стала SPAR (Système de préservation et d'archivage réparti – Система збереження та розподільчого архівування). SPAR є способом реалізації еталонної моделі OAIS (Open Archival Information System – Відкритий проект архівної інформаційної системи), розробленої CCSDS (Consultative Committee for Space Data System – Консультативний комітет систем космічних даних) у 1999 р. і прийнятої ISO (International Organisation for Standardization) як норма 14721 у 2002 р.

Еталонна модель OAIS «пропонує базову модель довготривалого зберігання і доступу до електронних даних, надає концепцію, що дозволяє усім не архівним організаціям бути

активними учасниками в процесі зберігання документів; описує структуру архіву (термінологія, архітектура засобів, експлуатація) для існуючих і майбутніх архівів; слугує підставою для порівняльного аналізу та дискусій з питань функціональної та інформаційної моделей довготривалого зберігання електронних даних; пропонує специфікації функцій прийому, доступу, адміністрування та зберігання електронних даних» [4].

SPAR почала діяти у 2010 р. відповідно до принципів, спрямованих на забезпечення принципів довготривалого зберігання: 1) програмне забезпечення SPAR розроблене на основі використання відкритого програмного забезпечення; 2) розробка має інтегративний характер, який покликаний забезпечити пристосування до нових потреб та змін; 3) представлені документи постійно аналізуються, підтверджуються, збагачуються інформацією, необхідною для їх зберігання; 4) інфраструктура зберігання розташована на двох сайтах; технології регулярно оновлюються з метою запобігання ризику застарівання [15].

Використання SPAR дозволяє вирішити наступні завдання:

- гарантувати зберігання електронних документів і забезпечити читабельність, зрозумілість, придатність для використання на довготривалий термін документальної інформації навіть в умовах зміни технічного і людського оточення, в якому вона виготовлялася;

- здійснювати операції електронного архівування;

- виконувати роль «поширювача» документів: забезпечувати управління багатьма копіями документів на багатьох сайтах з метою запобігання їх втраті та знищенню [15].

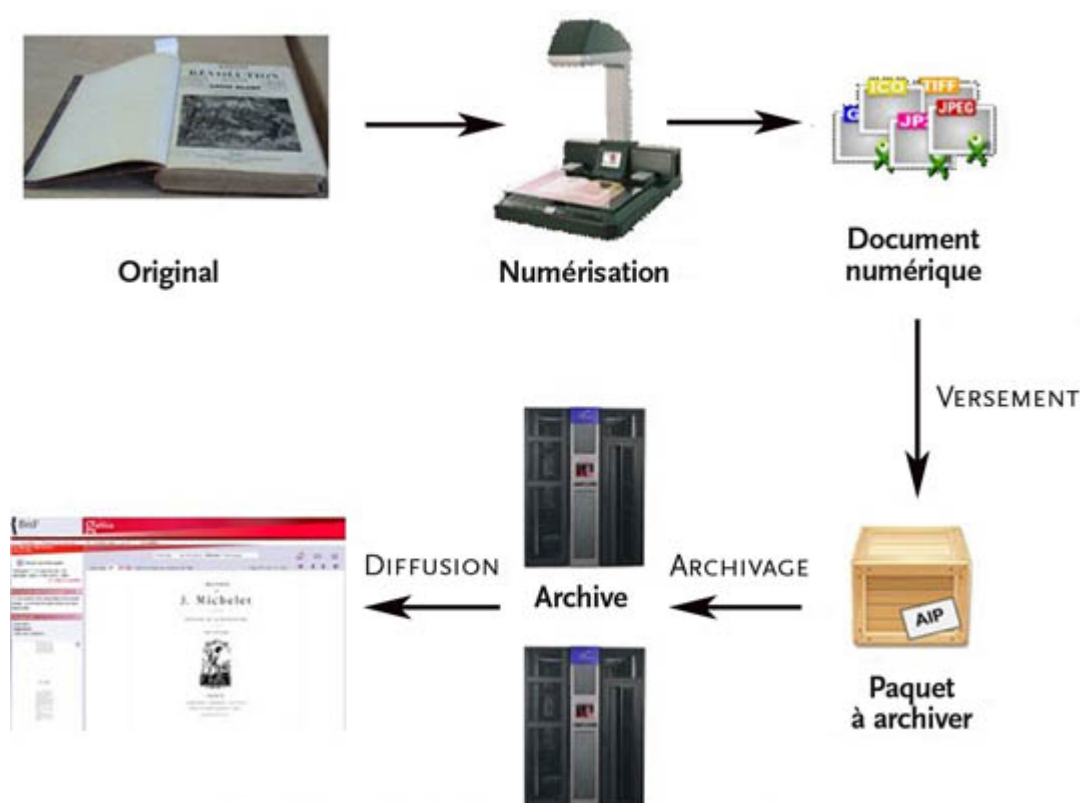
SPAR покликана реалізувати триєдине завдання: збір, зберігання і поширення даних. Висвітлюючи функціонування SPAR, спиратимемося на узагальнюючий огляд, здійснений французькими фахівцями-співробітниками НБФ Е. Берме (архівіст-палеограф, керівник служби «Перспективне прогнозування та документальні послуги» Департаменту бібліографічної та електронної інформації Дирекції служб та мереж НБФ), М.-Е. Фреон (керівник служби оцифровки Департаменту зберігання НБФ, керівник проекту зі зберігання електронних колекцій) та Ф. Мартеном (заступник начальника служби Асоційованих центрів/Gallica НБФ, відповідальний за документальні методи електронних колекцій) [10].

Архів повинен гарантувати цілісність та автентичність переданих до нього документів, а також збирати всю необхідну інформацію, що забезпечить сприйняття документальних даних користувачем. З цією метою архів використовує такі стратегії зберігання, як міграція, яка передбачає трансформацію об'єктів у більш довговічні формати чи/та формати, краще керовані архівом, та емуляція, яка дозволяє відновлювати електронні об'єкти, імітуючи їх початкове інформаційне оточення. Використання подібних стратегій вимагає здійснення постійного контролю за документами та накопичення чисельної додаткової інформації:

технічних, структурних, правових метаданих, інформації про формати, опис базисних платформ, на яких можуть функціонувати об'єкти, засоби контролю цілісності файлів тощо. У SPAR всі ці дані завантажені у картотеки (файли) метаданих (цей процес має назву «упаковка») у форматі METS, який дозволяє завантажувати у XML всю інформацію про об'єкт і гарантує довговічність. Потім копії цих метаданих передаються до модуля «управління даними» – склад метаданих, закодованих у RDF (Resource Description Framework), куди здійснюються комплексні запити, які дозволяють проводити операції зі зберігання.

Життєвий цикл документів у SPAR забезпечується координованою діяльністю учасників проекту, серед яких: управляючі електронними колекціями, які відповідають за зміст об'єктів, що зберігаються в архіві; адміністратори; експерти зі зберігання, які відповідають за технічний бік справи; керівники сектора (особи, уповноважені зобов'язувати бібліотеку до довготривалого зберігання окремих електронних колекцій).

Приклад ланцюжка «Оцифровка – збереження» [10].



На 2008 р. в електронному архіві було представлено 560 тераоктет даних, на кінець десятиліття прогнозувалося їх зростання до 1,4 петаоктет. SPAR, враховуючи різноманітність цифрових об'єктів, які зберігаються у НБФ, об'єднала їх у різні сектори, відповідно до вимог зберігання. Зокрема, перший опис дозволив виділити наступні сектори: 1) оцифровка збереження (відтворення високої якості і надійності з оригіналу, яке створює допоміжний екземпляр [11]); 2) оцифровка тиражування; 3) автоматизоване надання

обов'язкового зразка друкованої продукції (*dépôt légal automatique* – вже створені в електронному вигляді документи збираються автоматизованим способом, подібні колекції розміщуються у веб-архіві [11]); 4) договірне розміщення (*dépôt légal négocié* – документи передаються на зберігання після переговорів стосовно форми і умов з видавцями чи розповсюджувачами [11]); 5) розміщення в архіві адміністративних і технічних документів, придбань і подарунків. Крім усього іншого, модулі SPAR можуть використовуватися третіми особами в рамках політики національного співробітництва НБФ.

Успішне функціонування проектів забезпечується чіткою взаємодією з суміжниками, яка базується на обміні необхідною інформацією з виробництва різних типів даних: файлів-зображень (цифрові факсиміле оригіналу); метаданих; текстових файлів за допомогою OCR (*Optical character recognition* – оптичне розпізнавання символів, яке широко використовується для конвертації документів в електронний вигляд); файлів змісту та індексу.

На початку роботи з об'єктами НБФ створює два елементи. Перший – це реєстр обробки у форматі XML, що забезпечує розпізнавання документів і визначає обробку запитів. Цей реєстр містить також бібліографічну інформацію, яка дозволяє ідентифікувати документ та засвідчити його фізичний стан. Другий елемент – це ордер вилучення в паперовій формі, призначений для транспортування та страховки.

Кожний електронний документ, який повертається після оцифровки. Супроводжується паперовим ордером повернення та файлом метаданих, який містить:

- повтор (копію) бібліографічної інформації, зокрема ідентифікатора оригіналу (штрих-код) та ідентифікатора бібліографічного опису, які дозволяють співвіднести оригінал з електронним екземпляром;
- ідентифікатор електронного документа, наданий провайдером зі списку даних НБФ;
- класифікаційні рамки Dewey (*Dewey Decima, DDC* – десяткова класифікація Дьюї) та, в разі необхідності, нумерацію періодики, томів тощо;
- таблицю відповідності між зображеннями та фізичними сторінками оригіналу, дані виробництва (взаємовідповідні файли зображень та змісту чи текстові файли, створені OCR), управлінська інформація за відділами виробництва, матеріалів, програмного забезпечення;
- опис повторної обробки об'єктів у форматі XML;
- опис доставки, який містить список доставлених електронних документів.

Сукупність файлів документа формує реєстр доставки – пакет передачі, до якого провайдер додає файл-копію для контролю цілісності доставлених даних.

Стосовно файлів та форматів, які використовує SPAR, кожний електронний документ містить відтворення своїх сторінок у вигляді зображень, до яких додаються XML-файли для

конверсії змісту, сторінок тексту за допомогою OCR та для метаданих, які здійснюють загальне управління.

Зображення еталонного електронного документа. створення якого передбачене статутом оцифровки НБФ, виготовляється у форматі Tiff V.6 нестислому, за виключенням чорних і білих текстів, стислих у IUT groupe IV (повністю реверсивне стиснення дозволяє реституцію всіх пік селів на їх початкове місце). До 2006 р. для стиснень сірих і кольорових зображень використовувався формат JPEУ, від якого згодом відмовились. Кожен Tiff-файл містить тему, яка у закодованій формі відповідно до стандарту Adobe включає технічну інформацію, інформацію про виробництво та право власності, необхідні для управління зображеннями. Стандартне розрешення – 300 dpi, але можна підвищити до 600. Якщо цього потребує оригінал.

Тексти, які конвертуються за допомогою OCR, відповідають схемі XML Alto (Analyzed Layout and Text Object), що дозволяє одночасно зберігати презентацію і зміст інформації. Кожна конвертована сторінка створює файл «Alto», який являє собою координати кожного ідентифікованого OCR елемента (текстові блоки, в тому числі координати кожного рядка і кожного слова, а також ілюстративні та графічні блоки). Ці координати дозволяють встановлювати відповідність тексту й оригінального зображення, коли їх накладають, щоб виділити терміни, знайдені в результаті запиту.

Файл змісту надає змогу отримати прямий доступ до деяких частин документа. Він відповідає схемі tdmNum. Цей формат дозволяє кодувати ієрархічні рівні змісту чи списки індексів в один файл, керуючись дуже спрощеним формальним підходом TEI (Text Encoding Initiative) та використовувати заголовки для назв частин: «Т» для змісту, «І» для індексу. Всередині частини кожний запис пов'язаний із зображенням і номером зазначеної сторінки.

Метадані у форматі XML відповідають правилам схеми refNum. Інформація згрупована в три базових елементи: бібліографія (опис та ідентифікатори), продукція (число сторінок та зображень тощо) та структура (таблиця відповідності для кожної сторінки коментарів чи підписів під зображеннями).

Е. Берме, М.-Е. Фреон, Ф. Мартен підкреслюють важливість бібліотечного співробітництва в галузі створення, зберігання та поширення електронних документів, оскільки його головним наслідком є зростання пропозиції документів користувачам, а також ефективніше використання фінансових і людських ресурсів. Крім цього, бібліотеки зацікавлені у спільному складанні списків оцифрованих документів, що дозволяє якнайширше представити обрану тематику. З'являється можливість використання електронного екземпляра з найкращими характеристиками. Саме тому з ініціативи НБФ та університетської бібліотеки Cujas було покладено початок програмі з оцифрування юридичних фондів та об'єднання їх у глобальну національну мережу.

Отже, створена НБФ система SPAR покликана забезпечити довготривале зберігання електронних документів великого обсягу і різних типів. У ній зберігаються 2 млн. оцифрованих документів, 16 млрд. веб-сторінок, - загальною кількістю близько петабайту (1000 терабайт) даних. З лютого 2013 р. НБФ пропонує послуги архівації третій стороні за встановленими розцінками, гарантуючи захист та збереженість електронних документів протягом 50 – 100 років [9].

Використані джерела

1. Ковтанюк Ю. Проект нормативно-правового акта України «Порядок роботи з електронними документами та їх підготовка до передачі на архівне зберігання» [Електронний ресурс] / Ю. Ковтанюк // Студії з архівної справи та документознавства. – 2012. – Т. 20. – С. 269-278. – Режим доступу: http://nbuv.gov.ua/j-pdf/sasd_2012_20_43.pdf. – Заголовок з екрана.

2. Лавренюк Г.В., Жук Л.В. Архівне зберігання електронних документів та інформаційних ресурсів в Республіці Білорусь [Електронний ресурс] / Г.В. Лавренюк, Л.В. Жук. – Режим доступу: irbis-nbuv.gov.ua/www.archives.gov.ua/Publicat/AU_1_2010/15.pdf. – Заголовок з екрана.

3. Малолетова Н. Національна бібліотека Франції: основні напрями діяльності [Електронний ресурс] / Н. Малолетова // Бібліотечний вісник. – 2013. – № 5. – С. 20-27. – Режим доступу: http://nbuv.gov.ua/j-pdf/bv_2013_5_6.pdf. – Заголовок з екрана.

4. Марченко П. Шляхи вирішення проблеми зберігання електронних документів в архіві (аналіз інтернет-ресурсів) [Електронний ресурс] / П. Марченко. – Режим доступу: www.archives.gov.ua/Publicat/Studii_2004.12_05.php. – Заголовок з екрана.

5. Масевич А.Ц., Ходоровский Л.А. Раздел «Электронные (цифровые) библиотеки» в учебном курсе «Проектирование библиотечно-информационных систем» в Санкт-Петербургском государственном университете культуры и искусств [Электронный ресурс] / А.Ц. Масевич, Л.А. Ходоровский. – Режим доступа: cyberleninka.ru/.../razdel-elektronnye-tsifrovye-biblioteki-v-uchebnom-... – Заголовок с экрана.

6. Национальная Библиотека Франции предоставляет контент в аудио формате [Электронный ресурс]. – Режим доступа: <http://www.rsl.ru/ru/s6/s373/s3734405/s37344054408>. – Заголовок с экрана.

7. Солов'яненко Д.В. Інтернет-технології бібліотечного сервісу в Україні: становлення і розвиток (1990-ті рр. – початок XXI ст.) [Текст] : дис. ... канд. іст. наук: 07.00.08 / Солов'яненко Денис Володимирович. — К., 2008. — 208 с.

8. Стрілець Н.О. Технології архівного збереження електронних бібліотечних ресурсів у мережевому комунікаційному просторі [Текст]: дис. ... канд. наук з соціальних комунікацій: 27.00.03 / Наталя Олександрівна Стрілець . – Харків, 2013 . – 243 с.

9. Франция: Национальная библиотека предлагает организациям услуги по электронной архивации [Электронный ресурс]. – Режим доступа: <http://tsdea.archives.gov.ua/pdf/news/12032013.pdf>. – Заголовок с экрана.

10. Bermes E., Fréon M.-E., Martin F. Tous les chemins mènent au numérique [Ressource électronique] / E. Bermes, M.-E. Fréon, F. Martin. – Régime d'accès : <http://bbf.enssib.fr/consulter/bbf-2008-06-0034-006>. – Écran titre.

11. Collections numériques préservées dans SPAR : filières et modèles [Ressource électronique]. – Régime d'accès : http://www.bnf.fr/fr/professionnels/spar_systeme_preservation_numerique/a.spar_filieres.html. – Écran titre.

12. Gallica, bibliothèque numérique [Ressource électronique]. – Régime d'accès : <http://gallica.bnf.fr/?&lang=FR>. – Écran titre.

13. La Bibliothèque nationale de France: collections, services, publics / dir. Daniel Renoult, Jacqueline Melet-Sanson. – Paris: Electre-Éditions du Cercle de la Librairie, 2001. – 238 p.

14. Le Réseau francophone numérique | RFN [Ressource électronique]. – Régime d'accès : http://www.bnf.fr/fr/collections_et_services/bibliotheques_num_internationales/a.portail_reseau_francophone_numerique.html. – Écran titre.

15. SPAR, le système de préservation numérique de la BnF [Ressource électronique]. – Régime d'accès : http://www.bnf.fr/fr/professionnels/spar_systeme_preservation_numerique.html. – Écran titre.

16. Sur Gallica Labs, un nouveau visualiseur de documents [Ressource électronique]. – Régime d'accès : <http://blog.bnf.fr/gallica/index.php/2015/01/26/sur-gallica-labs-un-nouveau-visualiseur-de-documents/>. – Écran titre.