

УДК 004.338

І. В. Петров, А. М. Стеценко, Н. В. Солоніна
Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

Обґрунтування вибору програмно-апаратних засобів архівів електронних документів довготермінового збереження

Розглянуто сучасний стан електронних архівів в Україні та програмно-апаратні засоби побудови систем довготермінового збереження цифрової інформації, що пропонуються і використовуються в Україні та світі. Запропоновано підходи до створення архівів електронних документів, які відповідають правовим нормам збереження електронних документів.

Ключові слова: електронні архіви, роботизовані бібліотеки, архівні підрозділи, програмні продукти.

Вступ

Кількість електронних документів, у першу чергу тих, які утворюються за рахунок держбюджету в органах влади та інших державних установах і організаціях, з кожним роком стрімко збільшується, що потребує вирішення проблеми їхнього збереження [1]. Існують види документів, для яких неможливо вказати, коли вони втрачають термін дії або ж стають неактуальними. В багатьох випадках цінність документів може зростати.

З точки зору українського законодавства архівна справа виходить за рамки документообігу і є самостійним процесом, який зберігає електронні документи. Сучасними тенденціями розвитку систем електронного зберігання є їхня інтеграція в корпоративні системи електронного документообігу, системи управління підприємством тощо. У зв'язку з впровадженням корпоративних систем електронного документообігу можна виділити два напрямки в реалізації архівних систем:

— системи, які є частиною системи електронного документообігу та приймають електронні документи на зберігання безпосередньо з неї;

— системи, які наповнюють архівний фонд підприємства за рахунок ретроконверсії, тобто перетворення паперового відомчого архіву на електронний відомчий архів.

У роботі проаналізовано сучасний стан проблеми в Україні, наведено короткий огляд програмних продуктів та апаратних засобів, призначених для побудови систем довготермінового зберігання електронних документів.

Постановка проблеми

Архівне зберігання є достатньо складним і трудомістким процесом. Практично всі органи влади та державні установи здійснюють архівне зберігання своїх електронних документів. Автоматизація архівного зберігання документів виникла та розвивалася практично одночасно з розвитком технічних засобів, які придатні для вирішення задач збереження.

Актуальність вирішення задачі побудови архіву електронних документів зумовлена необхідністю розв'язання проблем постійного, тривалого (понад 10 років) зберігання електронних документів, що стрімко накопичуються при постійному вдосконаленні інформаційних технологій та систем у державних підприємствах і установах, а також необхідністю спрощення процедури доступу до них.

Перспективна платформа для побудови архіву електронних документів

Система зберігання архівів електронних документів повинна включати в себе такі складові: стрічкові або оптичні дискові бібліотеки (сховище довгострокового зберігання електронних документів); спеціалізовану інфраструктуру доступу серверів до пристроїв зберігання даних; програмне забезпечення керування даними, що зберігаються; систему керування якістю сервісу; централізовану систему резервного копіювання та відновлення; підсистему сканування документів.

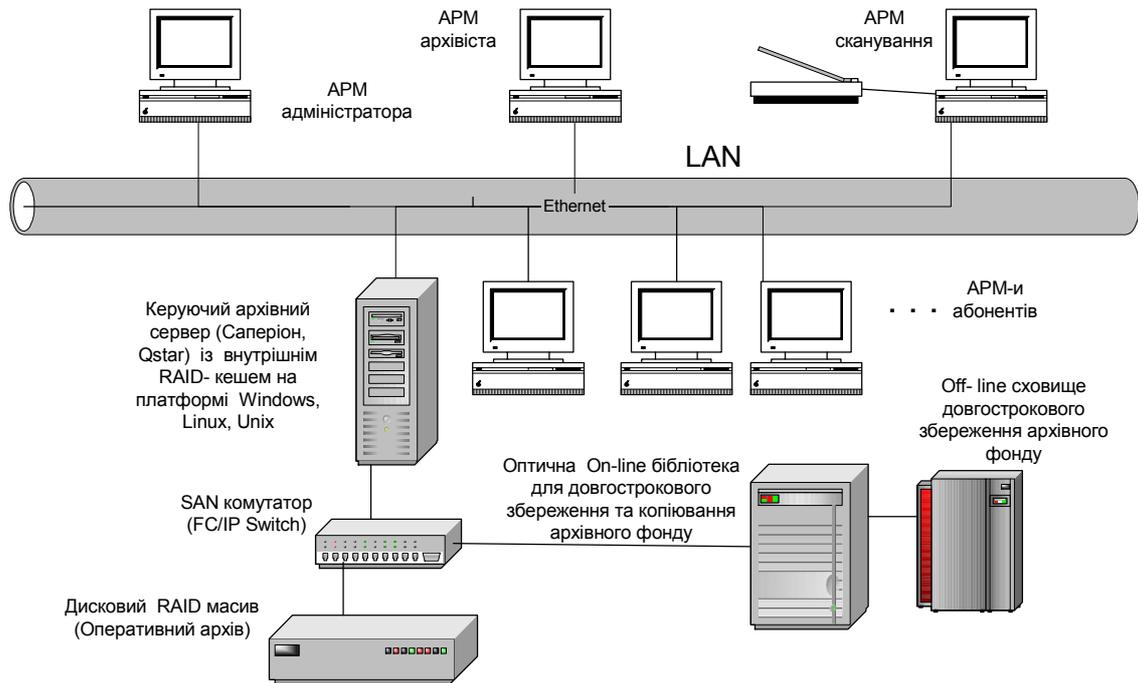
Використані в технічних рішеннях сучасного сховища програмно-апаратні засоби покликані забезпечити наступну функціональність.

Організацію інтелектуального керування сховищем даних і збереження даних на енергонезалежних носіях. Пряме керування архівними накопичувачами CD/DVD/BD, що безпосередньо задіяні в технологічних процесах по збереженню й експлуатації інформаційного ресурсу.

Застосування СКБД для керування індексною базою (MS SQL, Oracle, Database 11g і т. п.). Наявність функції автоматичної конвертації різноманітних форматів файлів додатків (pdf, txt, gif, doc, rtf) в універсальний формат, придатний для доступу до документів на робочих станціях, укомплектованих типовими засобами перегляду.

Подібну типову платформу мережі довготермінового зберігання даних можна представити як трирівневу ІТ-інфраструктуру, що включає рівні комутації, сервісів і сховищ (див. рисунок). Рівень комутації включає в себе адаптери шини хосту та SAN контролери і зв'язує сервер (хости) і сховища по протоколу FC або iSCSI. Цей рівень забезпечує варіантність шляхів доступу до даних у рамках мережі збереження, що дозволяє досягнути високої доступності даних.

Рівень сервісів забезпечує дублювання даних у вигляді створення повної копії робочого тому, snapshot, реплікації, синхронного або асинхронного міroringу і back-up на стрічку і/або диск через FC або iSCSI.



Структурна схема типової платформи (мережі та сховища)
довготермінового зберігання цифрової інформації

Рівень сховищ представляє собою дискові та стрічкові чи оптичні накопичувачі, що поставляються провідними світовими виробниками (IBM, HP, EMC, Overland, Plasmon та ін.). Довготермінове, надійне збереження даних потребує застосування зовнішніх архівних накопичувачів для створення резервних копій електронних документів і архівних ресурсів зі швидким довільним доступом до даних. Найбільш досконалі рішення, наприклад робототехнічні системи, підтримують надщільний оптичний формат (Ultra Density Optical, UDO), а також технології Blue Ray.

Використання перспективних архівних накопичувачів

При виборі технічних рішень збереження архівних ресурсів зі швидким довільним доступом до даних, оптичні технології PDD, UDO, DVD, BD найбільшою мірою забезпечують виконання вимог, що ставляться до сховища, включаючи такі параметри як висока надійність і тривалість збереження, автентичність і незмінюваність даних, швидкий довільний доступ до даних, висока ємність носіїв, можливість розширення.

Популярність технології UDO (широкий спектр пропозицій виробників і постачальників продуктів, можливість записати на один картридж 30 Гбайт інформації, термін збереження інформації 50 років) надавала підстави стверджувати, що автозавантажувачі оптичних дисків UDO з часом зможуть стати стандартом у галузі оптичного архівування даних. Так модель SER JB4U/22-X має обсяг збере-

ження 660 Гбайт, а ціна при 22 слотах — від 6700 євро і вище. В області рішень старшого класу (з обсягом пам'яті до 150 Тбайт) серія продуктів ASM 1000 може розмістити до 5000 носіїв даних.

Одним із головних конкурентів UDO аналітики вважають формат PDD (Professional Disc for Data), запропонований Sony, який представляє собою модифіковану версію технології Blu-ray Disc. Одна із версій PDD орієнтована на ринок систем збереження даних, інша адресована замовникам радіотелемовних компаній.

Наприкінці 2006 року консорціум INSIC (Information Storage Industry Consortium), що об'єднує головних виробників і дослідницькі центри з технологій збереження даних, представив огляд розвитку архівних накопичувачів на найближчі 5–10 років [2]. У даному огляді основну роль було відведено новому поколінню оптичних бібліотек, що використовують носії Blu-ray.

Широкі перспективи технології Blu-ray полягають в наступному. Насамперед — це універсальність: Blu-ray прийнята в якості кращого архівного носія в основних секторах IT-індустрії, а також у таких специфічних областях, як радіо- і телемовлення, медицина, охорона правопорядку, видавнича справа і багатьох інших. Довговічність BD (більш 50 років), їхня ємність (50 ГБ, у перспективі до 200 ГБ) і швидкісні характеристики (від 9 МБ/с) дозволяють цим дискам забезпечити створення унікальних за конкурентоздатністю систем архівування. Дискководи Blu-ray сумісні по читанню/запису з носіями CD/DVD, підтримуються файловою системою UDF. Технологія Blu-ray підтримується всіма основними виробниками дискководів і носіїв. Завдяки цьому диски BD мають всі передумови стати стандартом архівних носіїв, оскільки інші типи (жорсткі диски, магнітні стрічки) не відповідають повною мірою вимогам нормативних документів зі створення електронних архівів.

Так, наприклад, оптичні бібліотеки серії HCM BD є сімейством модульних продуктів, у яких використана технологія Blu-ray з носіями ємністю 25 або 50 ГБ кожний. Ємність оптичних накопичувачів серії HCM BD складає від 1 ТБ до 35 ТБ, забезпечується до 14 паралельних потоків передачі даних зі швидкістю 9 МБ/с, підтримуються носії одноразового запису (BD-R) і перезапису (BD-Re). Керування даними в режимі оффлайн підтримується інтелектуальною технологією пакетування (SmartPack). Дискководи Blu-ray сумісні по читанню/запису з носіями CD/DVD. Стандарти технології підтримуються всіма основними виробниками дискководів і носіїв, а також файловою системою UDF [3].

Збільшення тривалості часу використання та щільності запису інформації на оптичних носіях є одними з найбільш досліджуваних напрямків у розвитку оптичних носіїв і бібліотек. В ІПРІ НАН України досліджуються і проводяться розробки технологій довготермінового зберігання цифрової інформації та надщільного запису інформації на перспективних реєструвальних середовищах. Співробітниками інституту було досліджено параметри інформаційного рельєфу для оптичних носіїв, де як підкладки оптичних носіїв використано високостабільні оптичні прозорі матеріали — силікатне та кварцове скло, ситал, сапфір, і виготовлено носії для довготермінового зберігання даних у стандартному форматі компакт-дисків [4].

Проведені дослідження показали, що створені в інституті оптичні носії характеризуються значно кращими оптичними та механічними властивостями порів-

няно із стандартними полікарбонатними носіями і забезпечують термін служби носія протягом сотень років.

Загальна характеристика спеціалізованого програмного забезпечення підтримки оптичних роботизованих бібліотек

Для роботи будь-яких оптичних бібліотек різних виробників потрібне спеціалізоване програмне забезпечення. Вибір програмного забезпечення, в першу чергу, залежить від специфіки задачі, для якої застосовується довгострокове збереження інформації.

Фактично всі задачі використання роботизованих систем умовно можна розділити на такі дві групи.

Перша група задач — резервне копіювання на бібліотеку (Backup). Ця група задач у свою чергу поділяється на два різних засоби копіювання:

- 1) Backup з одного Серверу, до якого безпосередньо залучена бібліотека;
- 2) Backup із декількох Серверів і/або Робочих станцій, тобто мережний Backup.

У першому випадку вартість ПЗ порівняно незначна і, як правило, нижча вартості ПЗ для монтування томів (див. другу групу задач). У другому випадку (мережний Backup) вартість ПЗ дуже сильно може варіюватися і залежить від наступних факторів.

- кількості користувачів;
- кількості серверів;
- кількості баз даних, що знаходяться на цих серверах, і необхідності робити Backup «відкритих баз» і «відкритих файлів», тобто миттєвий знімок стану бази даних без необхідності блокування доступу до неї на тривалий час;
- різномірності та кількості встановлених операційних систем. Спостерігається наступна закономірність — чим більша кількість і різномірність встановлених ОС, БД, серверів, тим більш кошторисною буде створена система збереження. Далеко не кожний виробник ПЗ для Backup підтримує всіх виробників операційних систем, баз даних, поштових серверів, що передбачається використовувати (можливо використовуються) у даній мережі;
- необхідності в інших додаткових опціях (необхідності агрегації потоків і т.п.).

Друга група задач — так зване монтування томів, коли всі диски, що знаходяться в бібліотеці повинні бути представлені як один мережний диск, або групи дисків повинні бути представлені як кілька логічних томів (тобто, як мережні диски, наприклад F:, H:, M: тощо). Для можливості мережного доступу до дисків або каталогів на читання і запис на основі політик доступу системним адміністратором мережі налаштовуються відповідні права доступу для різного класу користувачів (наприклад: лише читати, читати і писати, заборона перегляду вмісту окремих дисків або всієї бібліотеки).

Вартість такого ПЗ складає істотну частину вартості всього рішення. Як правило — це біля 30 % від вартості бібліотеки. У першу чергу на вартість впливає максимальна кількість підтримуваних дискових слотів.

Принцип роботи ПЗ усіх розробників схожий. Це процеси запису/зчитування через два види дискового кешу — кешу запису і кешу зчитування, у якості яких можуть виступати як два різних SCSI-диски, так і RAID-масив. Кеш запису використовується для миттєвого запису даних з наступним переносом у фоновому режимі на оптичний диск (ПЗ проаналізує, що робот і привід вільні, і перепише дані з кешу в бібліотеку).

Кеш зчитування використовується для прискорення зчитування даних користувачем. ПЗ, проаналізувавши, що користувач послідовно зчитує дані, починає впереджене зчитування в кеш і при зверненні користувача за черговою порцією даних, він уже не чекає зчитування з диска, а миттєво одержує ці дані мережею з кеша. Таким же чином з кеша одержують дані й інші користувачі мережі, у випадку якщо їм знадобилися дані, розташовані на тих же оптичних дисках, що вже використовувались. Термін збереження даних в обох видах дискового кеша задається мережним адміністратором.

Можливості сучасних програмних продуктів для реалізації електронного архіву з використанням оптичних бібліотек

На сьогодні все більшої актуальності набуває використання архівних рішень у різноманітних системних середовищах. Інтеграція з ERP-системою об'єднує розрізнені додатки і процеси в єдину корпоративну інформаційну систему: документообіг, керування ресурсами підприємства (ERP), опрацювання копій паперових документів, оцифровка, збереження документів і даних. Таким чином, реалізується можливість організувати ефективну взаємодію між структурними підрозділами (відділами і філіями) підприємства та зберігання архівних документів.

У системах, які наповнюють архівний фонд підприємства за рахунок ретроконверсії, застосування програмного продукту Saperion та Qstar HSM надасть можливість архівним підрозділам підприємств створити сховище, що реалізує функції архіву організації, а також функції служб діловодства в частині формування та оформлення справ.

Система Saperion входить до переліку провідних світових платформ для Enterprise Content Management (ECM) — комплексного, повнофункціонального управління всім контентом підприємства. Система Saperion, що була створена як система електронного архіву документів, за рахунок подальшого розвитку функціональних можливостей забезпечує комплексну побудову вертикальних рішень класу HSM (Hierarchical Storage Management — Управління ієрархічним сховищем), зберігаючи всі свої ключові переваги в частині надійності та безпеки архівації та зберігання.

Компанія Saperion AG (Берлін/Цюріх) поставляє нову Версію 5.7 системи керування даними (СКД). Кожна версія такої системи має більше 20 компонентів програмного забезпечення. Кожний компонент має своє призначення та виконує функцію, належну тільки їй [5]. При виборі програмного продукту СКД ціна буде залежати від переліку вирішуваних задач підрозділами архівного зберігання державних підприємств та установ. Зі збільшенням функцій системи збільшується і кількість компонентів системи, яка приймає участь у життєвому циклі архівних

електронних документів, при цьому відповідно збільшується і ціна програмного продукту.

Система може застосовуватись як типове рішення для автоматизації архівної діяльності в державних установах і державних органах влади з великими об'ємами архівної інформації та поточного документообігу [6].

Архітектура системи передбачає наступні функціональні пристрої:

- сервер баз даних;
- документ-сервер (може імпортувати до двох мільйонів документів за годину);
- брокер-сервер ;
- мережний файл-сервер;
- документ-сервер роботизованої MO/CD/DVD/UDO бібліотеки;
- документ-сервер резервної копії;
- Web-Internet Gateway-сервер;
- DocFlow-сервер;
- спеціалізоване робоче місце сканування.

Saregion відповідає наступним вимогам по створенню електронних інформаційних ресурсів, ефективному керуванню та зберіганню документів:

- завантаженню різних форматів і типів електронних документів з різних носіїв;
- надійному зберіганню документів. Для зберігання бази даних використовується власний формат (мета-файл), який не можна прочитати або змінити зовні;
- неможливості вилучення (документ виключається тільки із доступу);
- неможливості зміни (завжди створюється нова версія документа);
- прямому керуванню обладнанням зберігання (HDD, RAID і т.п.);
- підтримки більшості моделей промислових та офісних сканерів;
- регламентації прав доступу;
- моніторингу системи.

Saregion має багато додаткових модулів, розроблених спеціально для архівації та управління документами. Система працює на різних платформах: Windows'9x/Me/NT/2000/XP, Linux (Intel®), Solaris® (SPARC), підтримується робота клієнтських станцій в Intranet/Internet через Microsoft Explorer і Netscape.

У світі Saregion працює більше ніж в 7000 організаціях. За кордоном система успішно впроваджена в державних структурах, відомих корпораціях, у банках і на підприємствах. У Росії система Saregion впроваджена більше як в 50 державних установах. В Україні, в ДНВП Геоінформ України, ведеться робота по створенню комп'ютерного архіву геологічних матеріалів на базі Saregion. Наповнюють архівний фонд підприємства за рахунок ретроконверсії, тобто перетворення паперового відомчого архіву в електронний відомчий архів.

Програмне забезпечення Qstar HSM є продуктом компанії Qstar Technologies (США) і відноситься до класу HSM. Вирішує завдання віддзеркалювання, міграції даних, забезпечення безпеки і безліч інших завдань, пов'язаних з архівним зберіганням.

За допомогою QStar HSM здійснюється переміщення даних через кеш жорстких носіїв на оптичні носії. QStar HSM забезпечує роботу з багаторівневою струк-

турую архіву, що дозволяє надійно зберігати велику частину інформаційних ресурсів підприємства на зовнішніх носіях та отримати швидкий доступ до даних завдяки кешу. QStar HSM — багатофункціональний програмний продукт, тому архівне рішення, засноване на його основі, може бути різної конфігурації відповідно до вимог підприємства чи компанії.

QStar HSM забезпечує високу продуктивність і збереження даних при роботі з інформаційними ресурсами підприємства. Тисячі впроваджень по всьому світу і підтримка широкого спектра архівних платформ — головна конкурентна перевага продукту.

До основних можливостей системи Qstar можна віднести наступне:

- об'єм архіву від 650 МБ до петабайтів;
- робота з різними носіями UDO, BD, DVD і CD в одному архівному рішенні;
- програмне забезпечення Qstar працює на всіх поширених серверних платформах (Windows, UNIX, Linux) [7].

На вибір системи Qstar HSM впливають такі фактори як простота використання, надійність середовища зберігання, швидкість передачі даних у процесі резервного копіювання. Системи керування даними дозволяє уніфікувати роботу з файлами і використовувати в межах підприємства єдиний користувацький інтерфейс.

До недоліків системи слід віднести надлишковість функціональності в системах управління даними, що веде до значного збільшення об'ємів програмного забезпечення, складності управління ними та вартість, що знижує зацікавленість користувачів до придбання таких комплексних програмних продуктів, особливо під час економічної кризи.

Вартість зберігання електронних документів в архівних системах довгострокового зберігання досить значна.

Так, при розрахунку вартості архівного рішення на оптичних дисках, необхідно врахувати наступне:

- вартість оптичного роботизованого накопичувача та технічної підтримки;
- вартість керуючої станції, до якої підключається оптичний накопичувач (включаючи вартість операційної системи з необхідною кількістю клієнтських ліцензій);
- вартість програмного забезпечення керування оптичним накопичувачем із комплектом необхідних ліцензій і технічної підтримки. В основному ліцензується кількість слотів в оптичному накопичувачі, кількість приводів, тип використовуваного носія (з можливістю однократного запису або з можливістю перезапису);
- вартість оптичних носіїв необхідного об'єму;
- вартість інсталяції, інтеграції в систему і часу консультування.

За даними фірми IDC (Японія), за 10 років зберігання однакової кількості інформації в системах збереження даних з носіями на оптичних дисках ціна складає \$6 350 950, а на стрічкових носіях — \$946 405 (Доповідь фірми Fujifilm Recording Media на семінарі Storage Academy фірми «Юстар» у червні 2009 року, м. Київ).

Архівні підрозділи державних підприємств та установ найчастіше зберігають дані на жорстких дисках, використовуючи розповсюджені програмні продукти, тримаючи копії на стрічках, оптичних дисках, виготовлених за технологією

CD/DVD/UDO (архівні підрозділи: УкрНТЕІ, бібліотеки ім. Вернадського, Київської та Кримської лазерної обсерваторії, Гідрометеорологічного центру та інші).

Центральний державний електронний архів (ЦДЕА), який був створений два роки тому, використовує власне програмне забезпечення під платформу Windows. Згодом планується проведення уніфікації й переведення архівного підрозділу ЦДЕА на крос-платформне програмне забезпечення. Така власна розробка ЦДЕА України дозволить вільно вносити зміни та корегувати алгоритм роботи залежно від вимог часу.

На даний час ЦДЕА створюють архівний фонд веб-ресурсів Державного комітету архівів України. На стадії розробки перебуває система для перегляду скопійованих веб-ресурсів і пошукова система [8].

Висновки

1. На даний час у світі до найбільш розповсюджених програмних продуктів, що обслуговують архівне зберігання електронних документів з використанням оптичних носіїв і є найбільш функціональними, належать Qstar HSM компанії Qstar Technologies (США) та Saperion німецької компанії Saperion AG (Берлін/Цюріх). Головна перевага цих програмних продуктів полягає в тому, що вони розроблені з урахуванням вимог архівного зберігання документів. До недоліків системи слід віднести його високу вартість.

2. В Україні не створена цілісна система нормативно-методичного забезпечення процесів архівного зберігання електронних документів, гармонізована з нормами міжнародного права з питань розвитку інформаційного суспільства. Тому єдиних правил по веденню електронних архівів відомчих державних підприємств та установ немає. Створюються електронні архіви залежно від вирішуваних архівом задач по збереженню даних та економічних можливостей тієї чи іншої установи.

Найбільш поширеніший спосіб збереження даних — використання жорстких дисків із застосуванням розповсюджених програмних продуктів і створення резервних копій документів на магнітних стрічках, оптичних UDO та CD/DVD дисках (Центральний державний електронний архів, архівні підрозділи: УкрНТЕІ, бібліотеки ім. Вернадського, Київської та Кримської лазерної обсерваторії, Гідрометеорологічного центру та інші). ДНВП Геоінформ, створює архівний фонд геологічних документів шляхом ретроконверсії на базі системи Saperion з використанням оптичного DVD-накопичувача HSM 3000.

3. При виборі технічних рішень збереження архівних ресурсів зі швидким довільним доступом до даних, оптичні технології UDO, DVD, BD найбільшою мірою забезпечують виконання вимог, що ставляться до сховища, включаючи такі параметри як висока надійність і тривалість збереження, автентичність і незмінюваність даних, швидкий довільний доступ до даних, висока ємність носіїв, можливість розширення.

Таким вимогам, певною мірою, задовольняють бібліотеки серії HSM BD, де використана технологія Blu-ray, і носії ємністю 50 ГБ, що здатні зберігати інформацію протягом 50 років. Ємність оптичних накопичувачів можна поетапно на-

рощувати від 1 ТБ до 35 ТБ і використовувати до 14 паралельних потоків передачі даних зі швидкістю 9 МБ/с кожен.

Найбільшою мірою відповідають вимогам по створенню архівних рішень з використанням довгострокового збереження великих обсягів інформації програмні продукти Saperion та Qstar HSM, які можуть бути використані як типові рішення для автоматизації архівної діяльності в державних установах і державних органах влади з великими об'ємами архівної інформації та поточного документообігу.

1. *Додонов О.Г.* Методологія створення Національного реєстру електронних інформаційних ресурсів / О.Г. Додонов, О.В. Нестеренко, А.В. Бойченко // Реєстрація, зберігання і оброб. даних. — 2005. — Т. 7, № 3. — С. 88–97.

2. *Будущее хранения информации* — за Blu-ray [Электронный ресурс] // Корпорация Электронный архив. — Режим доступа: http://ncm.ru/news/news_000.shtml

3. *Оптические накопители ЭЛАР HCM BD* [Электронный ресурс] // Корпорация Электронный архив. — Режим доступа: http://ncm.ru/nsm_bd.shtml

4. *Горбов І.В.* Вплив показника заломлення матеріалу підкладки на глибину інформаційного рельєфу оптичних носіїв / І.В. Горбов // Реєстрація, зберігання і оброб. даних. — 2009. — Т. 11, № 1. — С.

5. *Описание компонентов ПО Саперион* [Электронный ресурс]. — Режим доступа: http://saperion.ru/comp_55.htm

6. *Система электронного архива с обеспечением доступа к нему граждан и организаций.* Описание программного обеспечения. — Технический проект. — 2004. — С. 5–10.

7. *Описание программного обеспечения QStar HSM* [Электронный ресурс] // Корпорация Электронный архив. — Режим доступа: http://ncm.ru/po/qstar_hsm.shtml

8. *Лавринюк А.Г.* Електронні архіви. Проблеми та перспективи / А.Г. Лавринюк // Щорічна зустріч керівників державних архівних служб України, Росії та Білорусі. — 10-12.09.2009.

Надійшла до редакції 01.02.2010