

도서관을 위한 디지털 포맷 선정 전략 연구

A Study on Digital Format Selection Strategies for a Library

서 태 설* · 장 덕 현**
Tae-Sul Seo · Durk-Hyun Chang

차 례

- | | |
|-----------------------|----------------------|
| 1. 서 론 | 5. 결 론 |
| 2. 디지털 포맷 유형 | • 참고문헌 |
| 3. 디지털 포맷 선정 기준 사례 고찰 | • 〈부록〉 전자자원 개발 정책 사례 |
| 4. 디지털 포맷 선정 전략 개발 | |

초 록

이 논문에서는 디지털 포맷의 기본 유형을 정의하였고, 미국의 NISO와 LC, 그리고 덴마크의 국가 및 대학 도서관/왕립 도서관의 디지털 포맷 선정 기준의 사례를 소개하였다. 이를 분석 및 종합하여 5 가지 유지관리 측면의 선정 요인(설명성, 독립성, 일관성, 범용성, 품질)과 품질 및 기능성 선정 요인을 얻었다. 이를 기반으로 도서관을 위한 디지털 정보자원에 대한 포맷 선정 전략을 제시하였다.

키 워 드

디지털도서관, 디지털 포맷, 장서 개발, 디지털 장서

* 한국과학기술정보연구원 지식기반팀 책임연구원
(Principal Researcher, Knowledge Resources Team, KISTI, tsseo@kisti.re.kr)

** 부산대학교 문헌정보학과 부교수
(Associate Professor, Dept. of Library, Archive and Information Studies, Pusan National University, dchang@pusan.ac.kr)

• 논문접수일자 : 2008년 7월 29일
• 게재확정일자 : 2008년 9월 4일

ABSTRACT

In this paper, typical types of digital format were defined and digital format selection guidelines of US's NISO(National Information Standards Organization) and LC(Library of Congress) and Denmark's The State and University Library and the Royal Library were introduced. Five sustainability factors and quality and functionality factors were raised through analysing the three sample guidelines. The sustainability factors include descriptive, independency, consistency, widely adopted, and quality. Thus, digital format selection strategies for a library were developed.

KEYWORDS

Digital Library, Digital Format, Collection Development, Digital Collection

1. 서론

한 도서관에서 정보 자원의 입수 대상 자료의 범위를 검토하기 위해서는 여러 가지 기준을 설정할 수 있다. 예컨대, 자료의 유형, 주제 분야, 발간 매체 형식, 사용 언어, 발행 국가 등의 여러 가지 기준으로 검토할 수 있을 것이며(서태설 외 2005), 선택에 있어서는 항상 최근의 변화를 반영하여야 한다. 예를 들어 발간 매체의 경우만 보아도 과거에는 도서관이 종이매체를 중심으로 자원을 개발하였지만, 최근 컴퓨터 및 인터넷 기술의 발달로 발간 매체가 급속히 전자화 되고 있음은 주지의 사실이다. 또한 구독 및 입수를 위한 도입 가격 체계도 매우 복잡한 양상을 띠어서 전략적인 접근이 필요한 시점이다.

약 30년 전에 음성 카세트테이프가 등장한

이래 비디오카세트, CD-ROM, DVD, 플레이스테이션 게임 등의 연이은 등장은 도서관의 장서 개발의 문제를 복잡하게 만들어 왔다. 한편, 디지털도서관의 확산과 온라인 데이터베이스 및 인터넷의 등장으로 예전에 존재하지 않던 다양한 유형의 정보 매체들이 속속 등장하게 되어 정보의 수집과 축적, 검색 등 모든 부문에서 있어서 다양한 변화를 야기하고 있는 것 또한 사실이다.

이제는 도서관이 단지 친숙한 이용자에게 정보를 효과적으로 서비스하는 것만으로는 불충분하게 되었다. 디지털 환경 자체가 성숙한 결과 절대 다수의 국민들에게 정보를 보급하기 위한 제일의 수단이 됨에 따라 정보의 통합(integration)과 신뢰성이 디지털 장서 개발의 중요한 기준으로 떠오르게 되었다. 이는 웹 표준과 기술이 이중 정보의 광범위한 통합을 지

원하게 되었고, 이용자들의 윈스톱 정보 이용 요구가 급증하고 있기 때문이며, 동시에 수많은 가용 정보량은 이용자로 하여금 신뢰할 수 있는 정보를 찾기 어렵게 만들기 때문이다 (NISO 2007).

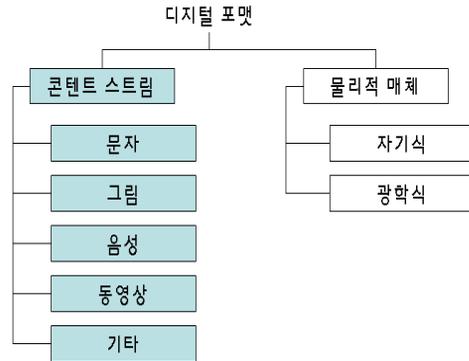
이에 따라 전자자원의 수집과 아카이빙을 담당하는 디지털도서관은 각 정보의 유형에 따라 적절한 디지털 포맷(digital format) 유형을 선택하는 기준을 마련해야만 한다. 그런데, 디지털 매체는 유형과 제작 회사에 따라서 수많은 종류가 존재하여 일관적인 선택 기준을 만들기가 쉽지 않다. 게다가 정보 기술의 빠른 변천도 반영해야 하기 때문에 문제는 더욱 복잡적일 수밖에 없다.

본 연구에서는 디지털도서관의 디지털 장서 개발과 관련하여 장서를 구성하게 될 요소인 디지털 포맷의 유형을 살펴보고 수집 범주에 대해서 검토한 후, 효과적인 디지털 포맷 선정 전략을 제시하였다.

2. 디지털 포맷 유형

2.1 디지털 포맷 기본 유형

디지털 포맷은 크게 콘텐츠 스트림(content stream)과 물리적 매체(physical media)로 구분할 수 있다(Abrams 2005). 콘텐츠 스트림은 별도의 재생 장치(physical medium)와 독립적으로 재생할 수 있는 포맷인 반면, 물리



〈그림 1〉 디지털 포맷의 유형 구분

적 매체는 특정한 물리적 저장 구조상에서 감지되는 형태로 나타나는 콘텐츠 스트림을 위한 것이다. 콘텐츠 스트림은 매체 유형에 따라서 문자(논리, 수치), 그림, 동영상, 음성, 기타 등으로 나누어지고, 물리적 매체는 저장 기술에 따라서 자기(magnetic) 방식과 광학(optical) 방식으로 나누어진다(〈그림 1〉 참조).

이중에서 콘텐츠 스트림 유형의 디지털 매체는 〈표 1〉처럼 다양하게 구분될 수 있으나, 문자, 음성, 그림, 동영상, 복합 매체, 기타 유형(software, database, CAD file 등) 등으로 구분하는 것이 일반적이다. 각 유형들은 부어 또는 재생기에 따라서 다양한 파일 포맷으로 세분화 되고 최신화 된다. 이러한 세분화된 포맷은 기술의 발달과 함께 새로운 포맷이 등장하고 있어서 지속적으로 관찰하고 그에 따른 수집 정책을 수립하여야 할 것이다.

본 연구에서는 모든 디지털 매체 중 문서, 그림, 음성, 동영상 등의 콘텐츠 스트림 유형을 중심으로 살펴보았다.

〈표 1〉 디지털 매체 유형 구분 사례

분야	DCMI ¹⁾ Type Vocabulary	MIME ²⁾ Media Types	GDFR ³⁾ format category	국가지식 포털
매체 구분	Collection			
	Dataset			
	Event	Application		원문(full-text)
	Image	Audio	Logical	메타데이터(text)
	Interactive Resource	Example	Numerical	이미지(image)
	Moving Image	Image	Text	녹음자료(audio)
	Physical Object	Message	Image(still)	동영상(video)
	Service	Model	Image(motion)	멀티미디어(multipart)
	Software	Multipart	Audio	소프트웨어(application)
	Sound	Text		기타
	Still Image	Video		
	Text			

2.2 문자 유형

문자 유형은 도서관을 위한 가장 보편적인 디지털 매체로서 PDF(Portable Document Format)가 가장 대표적이라고 할 수 있다. 그 외에 마크업 문서인 XML(Extensible Markup Language)과 전자책을 위한 OEBPS(Open eBook Forum Publication Structure)가 있고, 기타 DTB, NCBI Arch, NITF 등이 있다.

PDF는 원래 Postscript에 기초해서 그것을 다양한 컴퓨터와 프린터로 프린트할 수 있게 하기 위해 만들어진 것으로서 종이 문서와 같은 느낌과 외관을 갖는다. 이중 PDF/A는 장기 보존을 위해서 ISO에서 표준화 한 유형이다. XML은 각 항목의 의미의 마크업과 정보

의 표현을 모두 수용할 수 있는 태그를 포함하는 포맷으로 상호운용성 면에서 가장 유리하다.

2.3 그림 유형

그림 유형의 파일 포맷에는 픽셀에 기반한 것으로 JPEG(Joint Photographic Experts Group), TIFF(Tagged Image File Format), GIF(Graphics Interchange Format), PNG(Portable Network Graphics) 등이 대표적이며, 벡터 방식으로는 SVG(Scalable Vector Graphics)가 있다.

JPEG은 모든 디지털 카메라에서 지원하는 포맷으로 손실(lossy) 압축 알고리즘을 채택하기 때문에 비교적 작은 파일 크기를 갖는다

1) DCMI : Dublin Core Metadata Initiative

2) MIME : Multipurpose Internet Mail Extensions

3) GDFR : Global Digital Format Registry

(http://en.wikipedia.org/wiki/Image_file_formats). 그래서 PDF에서는 JPEG가 이미 지 압축 알고리즘으로 사용되고 있다. TIFF는 LZW라는 손실 없는(lossless) 압축 알고리즘을 채택하기 때문에 비교적 큰 파일 크기를 갖는다. 그래서 웹브라우저에서는 TIFF를 지원하지 않는다. GIF는 256색으로 한정되어서 적은 색의 그림 파일에 적합하다. GIF는 애니메이션을 지원하기 때문에 이미지를 이용한 애니메이션 효과를 내는 데 널리 활용된다. PNG는 GIF의 후속으로 나온 최신의 오픈 소스 파일 포맷으로 GIF와는 달리 트루 컬러(true color)를 지원하며, Internet Explorer 7부터 지원이 된다. 벡터 방식인 SVG는 스크립팅이라는 장점을 이용해서 상호작용 웹(interactive web)에 중요한 요소로 사용된다.

2.4 음성 유형

음성 저장 매체는 LP에서 카세트테이프, DAT(Digital Audio Tape), 음성 CD, 미니 디스크, 플로피 디스크, 하드 디스크 등으로 변천해 왔다. 이 중에서 디지털 매체에 저장되는 음성 포맷으로는 MPEG 계열의 MP3 (MPEG-1 Level 3) 파일이 현재 가장 보편적으로 활용되고 있고, 플러그인에 따라서 Real Audio와 MIDI, WAVE, QTA(QuickTime Audio), WMA(Windows Media Audio) 등이 대표적이다.

2.5 동영상 유형

동영상 저장 매체는 16밀리 필름에서 카세트형 비디오테이프를 거쳐 음성 유형과 마찬가지로 CD, 미니 디스크, 하드 디스크 등으로 변천해 왔다. 이 중에서 디지털 매체에 저장되는 동영상 포맷으로는 MPEG 계열과, 플러그인에 따른 RealVideo, AVI, DivX, QT(QuickTime), WMV(Windows Media Video) 등이 대표적이다.

2.6 디지털 매체의 식별

디지털 매체는 여러 가지 유형이 있을 뿐만 아니라 같은 유형 내에서도 버전(version)이 다를 수 있고, 특정한 목적을 위해서 일부 변형한 하부 유형이 있을 수 있으며, 여러 개의 포맷을 포괄(wrapper or bundle)하는 유형도 있다. 그래서 단지 매체 유형 이름만으로는 정확하게 구분하기 어려운 경우도 있다. 그래서 미국의 NDIIPP(National Digital Information Infrastructure and Preservation Program)의 디지털 포맷 홈페이지에서는 아래와 같이 관련 포맷들 간의 관계를 표시하는 기술 형식을 마련하여 각 포맷을 기술하고 있다(<http://www.digitalpreservation.gov/formats/>).

PDF/A: ISO 19005-1. Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF

- 설명 : PDF/A-1 is a constrained form of Adobe PDF version 1.4 intended to be suitable for long-term preservation of page-oriented documents for which PDF is already being used in practice. The ISO standard [ISO 19005-1:2005] was developed by a working group with representatives from government, industry, and academia and active support from Adobe Systems Incorporated. Part 2 of ISO 19005 will be a version of PDF/A based on a more recent version of PDF.
- 상태 : A final-state format for delivery to end users and long-term preservation of the document as disseminated to users.
- 관계 :
 - * Subtype of PDF
 - * Subtype of PDF_1_4
 - * Has subtype PDF/A-1a
 - * Has subtype PDF/A-1b

3. 디지털 포맷 선정 기준 사례 고찰

3.1 미국의 NISO: 디지털 객체 선정 지침

미국의 NISO에서는 디지털 컬렉션 구축 지침 프레임워크 보고서를 발간하였다(NISO 2007). 이 보고서에서 디지털 객체(Digital Object)란 하나의 파일(file) 또는 여러 개의 연결된 파일로 구성되거나, 여러 개의 파일과 그것들을 묶는데 필요한 구조적 메타데이터로 구성된다. 이런 의미에서 객체는 도서관의 소장 품목(item)과 개념적으로 유사하다. 또한 디지털 객체는 컬렉션(collection)의 구성 요소가 된다.

디지털 객체는 소장을 위한 원본(master copy)과 이용을 위한 임시적 복사본으로 구분하여 관리하기도 한다. 예컨대, 그림 파일 원본은 압축이 없는 TIFF 포맷으로, 이용을 위

한 것은 압축 방식의 JPEG 포맷으로 함으로써 웹 이용자들이 빠른 시간에 다운로드할 수 있도록 할 수 있다.

- 원칙 1 : 디지털 객체는 의도하고 있는 현재와 미래의 이용을 지원하는 포맷 안에 존재한다.
- 원칙 2 : 디지털 객체는 장기간 저장 가능해야(preservable) 한다.
- 원칙 3 : 디지털 객체는 국지적 상황 밖에서도 사용 가능하고 의미가 있어야 한다.
- 원칙 4 : 디지털 객체는 지속성이 있고 전지구적으로 유일한 식별자로 명명되어야 한다.
- 원칙 5 : 디지털 객체는 인증될 수 있어야 한다.
- 원칙 6 : 디지털 객체는 관련 메타데이터가 있어야 한다.

3.2 미국 의회 도서관 : 아카이빙을 위한 디지털 포맷 선정 요인

미국의 의회 도서관(LC: Library of Congress)에서는 디지털 자원의 장기적 아카이빙을 위해서 디지털 포맷 관련 프레임워크와 전략 수립을 하였다. 여기서는 디지털 포맷을 평가할 때 고려해야 할 요인을 ‘유지관리 요인’과 ‘품질 및 기능성 요인’으로 구분하여 제시하고 있다(Arms and Fleischhauer 2003).

3.2.1 유지관리 요인

- 공개 정도(disclosure) : 이것은 기술적 완결성(integrity)을 검증할 수 있는 명세와 도구가 완비되었고 디지털 콘텐츠로의 접근 가능한 정도를 말한다. 즉, 정보가 어떻게 비트(bit)와 바이트(byte)로 부호화 되었느냐를 이해할 수 있어야 한다는 것이다. 독점 포맷보다는 비독점의 공개 표준이 더 문서화가 잘 되어 있고 검증 도구가 더 잘 지원되는 것이 보통이다.

- 채택 정도(adoption) : 이것은 그 포맷이 정보원의 일차 생산자, 보급자 또는 이용자에 의해서 이미 이용되고 있는 정도를 말한다. 어떤 포맷이 널리 이용된다는 것은 그만큼 폐기될 가능성이 낮고 이전과 에뮬레이션 도구를 쉽게 얻을 수 있다는 것을 의미한다.

- 투명성(transparency) : 이것은 디지털 표현이 문자 에디터를 이용하여 사람이 읽을 수 있는 것을 포함해서 기본도구로 직접 분석할 수 있도록 공개된 정도를 말한다. 중요한 정보가 쉬

우면서도 직접적으로 표현되어 있는 디지털 포맷이 아무래도 새 포맷으로의 변환이 쉬운 것이다. 음성, 동영상, 그림 등과 같은 많은 포맷들이 압축되어 있거나 암호화 되어 있는데, 이 경우는 널리 사용되고 공개되어 있으며 손실 없는(lossless) 형식의 경우는 무리가 없을 것이다.

- 기록의 자체 보유(self-documentation) : 해당 디지털 포맷을 설명하고 그것의 생성과 초기 단계와 관련된 기술적 및 관리적 사항이 포함된 메타데이터가 매체 자체에 포함되어 있으면 그렇지 않은 경우보다 더 유리할 것이다. OAIS 참조 모델은 메타데이터에 포함되어야 할 몇 가지 사항을 제시하고 있다. 1) 표현: 데이터가 정보로서 사용될 수 있게 하는 표현 정보, 2) 참조: 해당 콘텐츠를 식별하고 설명할 수 있는 참조 정보, 3) 상황: 예를 들어 콘텐츠 생성과 관련된 기록, 4) 불변성(fixity): 콘텐츠 데이터의 완결성의 검사를 허락하는 불변성 정보, 5) 기원(provenance): 콘텐츠가 생성된 이후의 변화 및 보관 이력 기록.

- 외부 의존성(external dependencies) : 이것은 특정 포맷이 특정 하드웨어, 소프트웨어, 운영 시스템 등과 관련된 정도와 미래의 기술적 환경에서 그러한 의존성을 다루기 위해 예측되는 복잡성을 말한다.

- 기술적 보호 장치(technical protection mechanisms) : 콘텐츠에 암호화 되었거나, 패스워드를 걸었거나, 특허권 보호를 위한 이용 제한 등과 같은 기술적인 보호 장치가 없어야 미래의 기술 환경에서도 쉽게 이용하게 될 것이다.

3.2.2 품질 및 기능성 요인

- 음성 녹음 자료 ;
 - 충실도(fidelity): 고음 해상도를 지원
 - 음장(sound field): 다채널 음 지원
 - 정상음 표현 범위를 넘는 기능성 : 주해에 의한 표현(예, MIDI)
- 정지 영상 자료 ;
 - 청명도(clarity): 화질의 고해상도를 지원
 - 색상 유지(color maintenance)
 - 정상 그림 표현을 넘는 기능성: 벡터 그래픽스, 삼차원 모델 등
- 문자 자료 ;
 - 문서 구조와 내비게이션의 완결성
 - 배지, 폰트, 디자인 특성 등의 완결성
 - 수학, 화학식, 도식 등을 표현 완결성
- 동영상 자료 ;
 - 청명도(clarity): 화질의 고해상도를 지원
 - 충실도(fidelity) : 고음 해상도를 지원
 - 음장(sound field) : 다채널 음 지원
 - 정상음 표현 범위를 넘는 기능성 : 부호화 애니메이션, 프레임 정밀 편집, 추가 음 궤적

3.3 덴마크 국립/왕립 도서관 : 파일 포맷 평가 기준

덴마크에서는 국립 도서관(The State Library)과 왕립 도서관(The Royal Library)이 공동으로 인터넷 저장 프로젝트를 수행했다. 그 일환으로 파일 포맷 관리 전략을 수립하였

다(Clausen 2004). 이 자료에서는 개방성, 이식성(portability), 품질 등의 세 부류의 기준을 제안하고 있다.

- 개방성 ;
 - 제품 정보가 공개되어 있을 것
 - 특허나 저작권에 얽매이지 않을 것
 - 뷰어 소스가 무료일 것
 - 뷰어 소스가 GPL 라이선스를 가질 것
 - 암호화 되지 않을 것
- 이식성 ;
 - 하드웨어에 독립적일 것
 - 운영 시스템에 독립적일 것
 - 타 소프트웨어와 독립적일 것
 - 특정 기관, 단체 또는 사건과 독립적일 것
 - 널리 이용되고 있을 것
 - 내장 기능이 최소일 것
 - 버전이 하나이거나 잘 정의되어 있을 것
- 품질 ;
 - 공간 비용이 적을 것
 - 추가적 관리가 필요한 유관 포맷이 적을 것
 - 견실성
 - 단순성
 - 검증이 많이 되어 있을 것
 - 손실이 없을 것
 - 메타데이터가 지원될 것

4. 디지털 포맷 선정 전략 개발

본 연구에서는 앞 절에서 살펴본 사례를 중

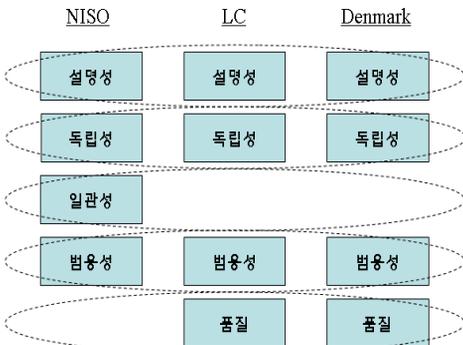
합하고 이를 참고해서 디지털 포맷 선정 전략을 마련하였다.

4.1 디지털 포맷 선정 기준 종합

앞의 세 기관의 사례를 분석·종합하여 <그림 2>와 같은 5가지 요인(설명성, 독립성, 일관성, 범용성, 품질)을 얻었다. 5가지 선정 기준 가운데 마지막의 품질의 경우는 공통적 품질 기준과 매체별 품질 기준으로 나눌 수 있다. 품질 기준은 미국 의회도서관의 품질 기능 요인을 그대로 활용할 수 있을 것이다.

- 요인 1 : 설명성(descriptive)
 - 상호 호환과 재사용을 할 수 있도록 포맷의 생성, 이력, 용도 등과 관련된 문서가 마련되어 있어야 한다.
 - 포맷의 기술적인 특성을 쉽게 이해할 수 있는 메타데이터를 가지고 있어야 한다.
- 요인 2 : 독립성(independency)

- 비독점적이어야 한다.
- 암호화 되지 말아야 한다.
- 압축되어 있지 말아야 한다.
- 지적재산권 제한이 없어야 한다.
- 특정 하드웨어나 운영 시스템에 독립적이어야 한다.
- 특정 소프트웨어에 독립적이어야 한다.
- 특정 기관 또는 사건과 무관해야 한다.
- 요인 3 : 일관성(consistency)
 - 기술이 변해도 계속해서 사용할 수 있어야 한다.
 - 버전 관리가 잘 되어야 한다.
 - 포맷 이름이 변동되지 않고 유일 식별되어야 한다.
- 요인 4 : 범용성(widely adopted)
 - 표준적인 매체여야 한다.
 - 널리 사용되고 있는 것이어야 한다.
 - 좋은 이용 사례가 있어야 한다.
- 요인 5 : 품질(quality)
 - 공간 비용이 적어야 한다.
 - 추가적 관리가 필요한 유관 포맷이 적어야 한다.
 - 견실성
 - 단순성
 - 변환 시 손실이 없어야 한다.
 - 매체별 품질 요인(별도 설명)



<그림 2> 디지털 포맷 선정 요인 도출

매체별 품질선정 요인은 미국의 LC와 동일하다.

4.2 디지털 포맷 선정 모형

디지털 매체의 선정은 <그림 3>과 같이 우선적으로 품질과 기능성 요인을 만족시킨 후에 유지관리 요인을 반영하는 순서로 하여야 할 것이다. 왜냐하면 유지관리 요인에 대한 평가가 아무리 좋다고 하여도 요구되는 품질이나 기능을 구비하지 못한다면 쓸모가 없는 것이기 때문이다.

다음에서는 미국의 NDIIPP 프로젝트에서 디지털 매체별 선정 특성에 따라서 평가한 결과를 소개한다(NDIIPP 2008). 각 매체 내에서는 또 다시 문서 유형별로 특성이 달라지기 때문에 매우 복잡한 양상을 보이고 있다.

4.2.1 문서

문서 유형의 디지털 매체에서 고려하여야 할 품질 및 기능성 요인으로는 1) 문서 구조와 내비

<표 2> 문서 유형별 선정 특성 분석

순번	유형 구분		문서구조	배치	수식표현
	유형 설명	예			
1	문서 구조와 내비게이션을 중요시 하는 문서	사전, 단행본, 보고서, 극본 등	◎	△	×
2	단순 문서구조를 가지는 짧은 문서	기사, 수필, 연설문 등	○	△	×
3	배치와 디자인을 중요시 하는 문서	브로슈어, 포스터, 광고물, 어린이 도서 등	△	◎	×
4	비문자형 시각요소(수식, 도형 등) 내의 정보 내용을 중요시 하는 문서	학술논문, 연구보고서 등	○	○	◎

(범례) ◎: 매우중요, ○: 중요, △: 보통, ×: 중요하지 않음.



<그림 3> 디지털 매체 선정 모형

게이션의 완결성, 2) 배치, 폰트, 디자인 특성 등의 완결성, 3) 수학, 화학식, 도식 등을 표현 완결성 등이 대표적으로 거론될 수 있다. <표

2>와 <표 3>에 이러한 기준을 따라서 각 유형에 맞는 선정 특성 분석과 적정 포맷을 제시하였다.

<표 3> 문서 유형별 적정 포맷

유형 구분		선호 매체 유형	허용 매체 유형
순번	유형 설명		
1	문서 구조와 내비게이션을 중요시 하는 문서	· XML+DTD · SGML+DTD	· HTML · PDF
2	단순 문서 구조를 가지는 짧은 문서	· XML+DTD · SGML+DTD · PDF	· HTML
3	배치와 디자인을 중요시 하는 문서	· HTML · PDF	· XML+DTD+XSL
4	비문자형 시각요소(수식, 도형 등) 내의 정보 내용을 중요시 하는 문서	· PDF	· XML+DTD+수식표현 기능
5	맹인을 위한 말하는 책	· XML+ANSI/NISO Z39.86	-
6	이메일 내용	· RFC 2822	-
7	일러스트레이션을 중요시 하는 문서	· 그림 유형 선정 기준 활용	

<표 4> 그림 유형별 선정 특성 분석

유형 구분			해상도	색상유지	정상이상 의 기능성	보관 요구	이용 요구	보호 효과
순번	유형 설명	예						
1	고도의 예술적 그림	전문사진, 광고, 예술 작품 등	◎	○	미세 표현	단순 가공	-	해상도에 영향주지 말아야 함
2	공간 해상도를 중요 시 하는 그림	지도, 그래프, 도형 등	◎	-	미세 표현	단순 가공	-	해상도에 영향주지 말아야 함
3	작품성이 낮은 그림	-	○~△	-	-	-	-	해상도에 영향주지 말아야 함
4	부수적 그림 또는 애 니메이션	웹용 이미지	△	-	-	-	-	×
5	벡터 그래픽스	CAD 도면	원본의 정밀도 유지	-	특정 SW 활용	원본의 기능성 유지	원본의 기능성 유지	기능에 영향주지 말아야 함
6	정적 삼차원 객체용 가상현실 묘사	-	-	-	-	-	-	-

(범례) ◎: 매우중요, ○: 중요, △: 보통, ×: 중요하지 않음.

4.2.2 그림

그림 유형의 디지털 매체에서 고려하여야 할 품질 및 기능성 요인으로는 1) 해상도(화질의 고 해상도를 지원), 2) 색상 유지, 3) 정상 그림 표

현을 넘는 기능성(벡터 그래픽스, 삼차원 모델 등) 등이 대표적으로 거론될 수 있다. <표 4>와 <표 5>에 이러한 기준을 따라서 각 유형에 맞는 선정 특성 분석과 적정 포맷을 제시하였다.

<표 5> 그림 유형별 적정 포맷

유형 구분		선호 매체 유형		허용 매체 유형	
순번	유형 설명				
1	고도의 예술적 그림	비트맵, 리치컬러, 비압축	• TIFF UNC (RGB, 24bit 이상)	• TIFF LZW, • JP2_J2K_C_LL in JPX J2K • JP2_J2K_C_LL in JP2 J2K	
		디지털 카메라	• TIFF/EP • TIFF_UNC_EXIF 또는 TIFF UNC(RGB, 24bit 이상)		
		인쇄 원본	• TIFF/IT • PDF/X • TIFF_UNC		
		다광 데이터	-		
2	공간 해상도를 중요시 하는 그림	비트맵, 컬러, 비압축	• TIFF_UNC	고품질 또는 인덱스된 컬러	• JP2_J2K_C_LSY in JPX_FF or JP2_FF • GIF • BMP
		벡터 delivery	• PDF/A • PDF/X	-	-
3	작품성이 낮은 그림	• JP2_J2K_C_LSY in JPX_FF or JP2_FF • JPEG_DCT in JFIF 또는 JPEG_EXIF • GIF		• GIF	
4	부수적 그림 또는 애니메이션	-		-	-
5	벡터그래픽스	-		-	-
6	정적 삼차원 객체용 가상현실 묘사	-		-	-

〈표 6〉 음성 유형별 선정 특성 분석

순번	유형 구분		행상도	음장	정상이상 의 기능성	보관 요구	이용 요구	보호 효과
	유형 설명	예						
1	서라운드 음향	-	◎	최소변화 유지	복수 스피커	단순가공	-	해상도에 영향주지 말아야 함
2	모노 또는 스테레 오 음향	-	◎~○	-	-	단순가공	-	해상도에 영향주지 말아야 함
3	보통의 음향	스트림, 웹캐스트 등	△	-	-	-	-	정상표현에 영향주지 말아야 함
4	부수적인 음향	배경 음악	×	-	-	-	-	×
5	악보기반 표현	MIDI	원본의 정밀도 유지	n/a	특정성능 SW 활용	원본 성능 유지	원본 성능 유지	기능에 영향주지 말아야 함
6	음향 효과	-	-	-	-	-	-	-
7	오디오복	-	-	-	-	-	다양함	-

(범례) ◎: 매우중요, ○: 중요, △: 보통, ×: 중요하지 않음.

〈표 7〉 음성 유형별 적정 포맷

순번	유형 구분		선호 매체 유형	허용 매체 유형
	유형 설명			
1	서라운드 음향		• AAC_ADIF	• AAC_M4A • QTA_AAC • WMA_WMA9_PRO
2	모노 또는 스테레오 음향		• WAVE_LPCM-BWF • WAVE_LPCM • AIFF-LPCM	• MP3+ID3 • AAC_ADIF • AAC_M4A • QTA_AAC • WMA_WMA9_PRO • WMA_WMA9
3	보통의 음향		• MP3 • AAC_ADIF • AAC_M4A • QTA_AAC • WMA_WMA9	-
4	부수적인 음향		모든 형식 가능	-
5	악보기반 표현		• SMF • XMF • RMID	• MODS
6	음향 효과		-	-
7	오디오복		• DTB	-

4.2.3 음성

음성 유형의 디지털 매체에서 고려하여야 할 품질 및 기능성 요인으로는 1) 해상도(고음 해상도를 지원), 2) 음장(다채널 음 지원), 3) 정상음 표현 범위를 넘는 기능성(악보에 의한 표현) 등이 대표적으로 거론될 수 있다. <표 6>과 <표 7>에 이러한 기준을 따라서 각 유형에 맞는 선정 특성 분석과 적정 포맷을 제시하였다.

4.2.4 동영상

동영상 유형의 디지털 매체에서 고려하여야 할 품질 및 기능성 요인으로는 1) 해상도(화질

의 고해상도를 지원), 2) 충실도(고음 해상도를 지원), 3) 음장(다채널 음 지원), 4) 정상 동영상 표현 범위를 넘는 기능성(부호화 애니메이션, 프레임 정밀 편집, 추가 음 궤적) 등이 대표적으로 거론될 수 있다. <표 8>과 <표 9>에 이러한 기준을 따라서 각 유형에 맞는 선정 특성 분석과 적정 포맷을 제시하였다.

4.3 디지털 포맷 선정 지침 적용 사례

<부록>에는 미국의 LC, 영국의 BL, 호주의 NLA 세 기관의 전자자원 개발 정책을 소개하였다. 이를 디지털 포맷에 따라 구분하면 <표

<표 8> 동영상 유형별 선정 특성 분석

순번	유형 구분		해상도	음장	정상이상 의 기능성	보관 요구	이용 요구	보호 효과
	유형 설명	예						
1	극장 배급용 동영상	-	◎	최소변 화유지	복수 스피커	단순가 공	단순가 공	해상도에 영향주지 말아야 함
2	방송용 동영상	지상파 방송, 위성 방 송, 케이블 방송 등	◎	최소변 화유지	복수 스피커	단순가 공	단순가 공	해상도에 영향주지 말아야 함
3	비디오용 동영상	비 디 오 테 이 프, DVD, 인터넷 등	◎	서라운드 스테레오에 정규화	-	단순가 공	-	정상표현에 영향주지 말아야 함
4	자료용 동영상	뉴스, 토크쇼 등	△	-	-	-	-	-
5	인터넷 방송용 동영상	-	△	-	-	-	-	x
6	부가적 동영상	웹캐스트, 단순 애 니메이션 등	x	-	-	-	-	-
7	동적 애니메이션	대화용 프로그램	원본의 정밀도 유지	n.a.	특정 SW활용	원본 성능 유지	원본 성능 유지	성능에 영향주지 말아야 함

(범례) ◎: 매우중요, ○: 중요, △: 보통, ×: 중요하지 않음.

10)과 같이 정리할 수 있다. 이것을 살펴보면, 문자 유형에는 전자저널, 전자책, 악보, 신문, 간행물 등이 있고, 그림 유형으로는 지도와 사진이 있다. 음성 유형의 경우는 언급만 있고 동영상 유형의 경우는 세 곳 모두 언급하

지 않고 있다.

본 연구에서 개발한 디지털 포맷 선정 지침을 위의 사례에 적용하여 보면 다음과 같이 정리할 수 있다.

- 문자 유형의 전자자원 : <표 10>에서는

<표 9> 동영상 유형별 적정 포맷

유형 구분		선호 매체 유형		허용 매체 유형	
순번	유형 설명				
1	극장 배급용 동영상	<ul style="list-style-type: none"> • DPX_2 • DCDM_1_0 • MXF(MXF_GC_UNC) 		<ul style="list-style-type: none"> • MPEG-2 또는 MXF(MXF_GC_MPEG-2) 	
2	방송용 동영상	비압축, 비손실 압축	<ul style="list-style-type: none"> • MXF(MXF_GC_UNC) • MJ2_FF_LL 	고품질 손실 압축(독점)	-
		고품질 손실 압축(공개)	<ul style="list-style-type: none"> • MPEG-2 또는 MXF(MXF_GC_MPEG-2) 	저품질 손실 압축	<ul style="list-style-type: none"> • MPEG-2 • MPEG-4_AVC • MPEG-4_V
3	비디오용 동영상	고품질 손실 압축(공개)	<ul style="list-style-type: none"> • MPEG-2 또는 MXF(MXF_GC_MPEG-2) • MPEG-4_AVC • MPEG-4_V 	고품질 손실 압축(독점)	-
		-	-	저품질 손실 압축	<ul style="list-style-type: none"> • MPEG-2 • MPEG-4_AVC • MPEG-4_V
4	자료용 동영상	<ul style="list-style-type: none"> • MPEG-2 • MPEG-4_AVC • MPEG-4_V 		-	
5	인터넷 방송용 동영상	고품질 손실 압축(공개)	<ul style="list-style-type: none"> • MPEG-2 • MPEG-4_AVC • MPEG-4_V 	고품질 손실 압축(독점)	-
		-	-	저품질 손실 압축	<ul style="list-style-type: none"> • MPEG-2 • MPEG-4_AVC • MPEG-4_V
6	부가적 동영상	모든 형식 가능		-	
7	동적 애니메이션	<ul style="list-style-type: none"> • FLA • SVG_1_2 • SVG_1_1 		<ul style="list-style-type: none"> • SWF 	

〈표 10〉 세계 주요 국가 도서관의 전자자원 범위

구분		LC	BL	NLA
콘텐츠 스트림	문자	전자 저널, 전자 책	전자 저널	약보, 책, 신문, 간행물
	그림	-	지도	사진, 지도
	음성	-	음성 자원	구전 역사
	동영상	-	-	-
물리적 매체		CD-ROM, DVD-ROM	음성 CD	-
기타		웹 사이트, 온라인 DB	웹 아카이빙	-

문자 유형의 전자자원으로 전자 저널, 전자 책, 약보, 신문, 간행물 등을 확인할 수 있다. 〈표 2〉와 〈표 3〉에 따르면, 전자책에 선호되는 디지털 포맷으로는 XML과 SGML이라 할 수 있고, 전자 저널의 경우는 PDF라고 할 수 있다.

- 그림 유형의 전자자원 : 〈표 10〉에서는 그림 유형의 전자자원으로 지도와 사진이 있다. 〈표 4〉와 〈표 5〉에 따르면, 사진에 선호되는 디지털 포맷은 TIFF 계열의 파일 포맷이 대표적이라 할 수 있고, 지도의 경우는 TIFF_UNC와 PDF/A, PDF/X 등이 선호된다.

- 음성 유형의 전자자원 : 〈표 10〉에서는 그림 유형의 전자자원으로 특별한 지칭이 없이 음성 자원과 구전 역사 등이 제시되고 있다. 따라서 음성 유형의 경우는 원하는 음질의 수준에 따라서 〈표 6〉과 〈표 7〉에 따라 다양한 디지털 포맷을 선택하여 활용할 수 있을 것이다.

- 동영상 유형의 전자자원 : 〈표 10〉에서는 동영상 유형의 전자자원으로 특별히 지칭된 것이 없다. 따라서 동영상 유형도 음성 유

형의 경우처럼 음질 및 화질의 수준에 따라서 〈표 8〉과 〈표 9〉에 따라 다양한 디지털 포맷을 선택하여 활용할 수 있을 것이다.

5. 결론

본 연구에서는 디지털 포맷의 기본 유형 구분을 알아보고, 물리적 매체와 무관한 콘텐츠 스트림에 대해서 특성과 종류를 알아보았으며, 이러한 디지털 매체별 세부 유형 선정 기준, 모형 및 지침을 제시하였다. 이것들을 요약하면 아래와 같다.

- 1) 디지털 포맷의 기본 유형으로 문서, 그림, 음성, 동영상 등이 있다.
- 2) 디지털 포맷 선정 요인으로는 일반적 요인과 품질 및 기능성 요인이 있는데, 품질 및 기능성 요인을 먼저 평가한 후 관리적 요인을 평가하여야 한다.
- 3) 일반적 요인은 설명성, 독립성, 일관성, 범용성, 품질 등으로 구분된다.
- 4) 품질 및 기능성 요인은 포맷에 따라 다

른 기준이 적용된다.

또한 좋은 디지털 컬렉션을 구축하기 위한 원칙 및 우선적으로 검토해야 할 주요 전자자원 유형과 그에 따른 대표적 디지털 포맷 유형을 제시하였다.

본 연구에서 살펴본 바와 같이 디지털 포맷은 매우 다양할 뿐만 아니라 기술 발전에 따라 빠르게 진화하고 있어서 획일적인 기준을 적용하기 어렵다. 따라서 여기서 제시된 내용들은 지속적으로 수정·보완되어야 한다.

참고문헌

- 서태설, 이태웅, 장덕현. 2005. 과학기술 정보자원 입수 효율화를 위한 정책 및 평가모델. 『KOSTI 2005』, 2005년 11월 28-29일. [서울: COEX 인터컨티넨탈호텔].
- NISO. 2007. *A Framework of Guidance for Building Good Digital Collections*. 3rd Ed., Maryland, USA, [cited 2008, 7, 21].
 <<http://www.niso.org/publications/rp/framework3.pdf>>.
- Abrams, Stephen L. 2005. "Establishing a Global Digital Format Registry, Library trends." *Library Trends*, 54 (1): 125-143.
- Arms, C. R. and Fleischhauer, C. 2003. *Digital Formats for Library of Congress Collections: Factors To Consider When Choosing Digital Formats(DRAFT)*. Library of Congress, USA, [cited 2008, 7, 21].
 <http://memory.loc.gov/ammem/techdocs/digform/DigForm_Intro_v04.pdf>.
- Clausen, L. R. 2004. *Handling file formats*. The State and University Library, Denmark. [cited 2008, 7, 21].
 <<http://netarchive.dk/publikationer/FileFormats-2004.pdf>>.
- NDIIPP. *Sustainability of Digital Formats Planning for Library of Congress Collections*. [cited 2008, 7, 21].
 <<http://www.digitalpreservation.gov/formats/>>.
- OCLC. 2004 *Information Format Trends*. OCLC Library & Information Center. [cited 2008, 7, 21].
 <http://www5.oclc.org/downloads/community/2004infotrends_content.pdf>.
- Wikipedia. *Image File Format*. [cited 2008, 7, 21].
 <http://en.wikipedia.org/wiki/Image_file_formats>.

<부록> 전자자원 개발 정책 사례

1. 미국 LC(<http://www.loc.gov/acq/devpol/electronicselectionguidelines.html>)

- 학술적 콘텐츠
- 디지털 형태로만 이용 가능한 콘텐츠
- 위험성(우연적 또는 계획적 자원의 제거)
- 한시적인 자원(예, 회색 혹은 숨겨진 문헌)

1.1 전자자원의 범위

전자자원이란 컴퓨터의 이용을 통해 접근할 수 있고 입력된 저작으로 정의된다. 이는 (1) 원격 접근과 (2) 고정 매체에 의해 이용 가능한 전자 데이터를 포함한다. 원격 접근 전자자원은 컴퓨터 네트워크를 통한 전자자원의 이용을 나타낸다. 고정 매체 전자자원은 컴퓨터 장치나 그 보조 장치로 삽입되도록 만들어진 매체(discs, cassettes, cartridges)를 통한 전자자원의 이용을 나타낸다. 전자자원으로는 웹 사이트, 온라인 데이터베이스, 전자저널, 전자책, 전자적으로 통합된 자원, CD-ROM, DVD-ROM 등이 있다.

1.2 전자자원 수집 일반 지침

다음의 기준에 관해서 높은 순위를 차지한 전자자원을 수집한다.

- 의회와 연구자의 현재나 미래의 정보 요구에의 유용성
- 정보 제공자의 명성
- 유일한 정보

1.3 특정 원칙

- 내용의 적정성
- 부가가치
- 접근성 : 서버의 신뢰성, 멀티미디어의 합법적 이용, 저작권 준수 및 공정한 이용, 프라이버시 보호, 플러그인을 위한 소스 링크 지원 등
- 디자인, 이용자 인터페이스, 내비게이션 특성
- 표준
- 영구 보존(Permanent Retention)

2. 영국 BL(<http://www.bl.uk/aboutus/stratpolprog/digi/dom/index.html>)

2.1 전자자원 범위

- 도입된 디지털화 자원
- 자체 자원의 디지털화
- 전자저널

- 지도 정보
- 음성 CD 및 기타 음성 자원
- 웹 아카이빙

2.2 Legal Deposit Libraries Act 2003

- 전자 출판물과 기타 비인쇄 매체가 소장 될 것이다.
- 이러한 자원은 미래의 연구자 및 학자 세 대에게 매우 중요하게 될 것이다.

2.3 디지털 객체 관리(DOM) 프로그램

- 임무 : 영국이 디지털 출력을 영구적으로 저장하고 사용하도록 하는 것.
- 비전 :
 - 모든 디지털 자원의 영구적 저장과 보전
 - 적절한 허용기준에 따른 이용자 접근성 제공
 - 자원 탐색의 용이성 제공
 - 현재 기술로 자원을 볼 수 있도록 지원
 - 원본의 느낌과 외관을 유지토록 지원
- 고려 사항 :
 - 디지털 자원의 생애주기: 생성, 수집, 보존, 접근, 저작권 관리, 제거
 - 외부 기술, 법규, 이용자 요구
 - 디지털 자원의 더 짧은 내구성과 기술적 노후화
 - 다양한 메타데이터
 - 물리적 저장소와 관련된 문제들

3. 호주 NLA(<http://www.nla.gov.au/policy/acqguide.html>)

3.1 전자자원의 범위

- 물리적 포맷 전자자원 : 물리적 매체로 출판되고 아날로그 전자자원(예 : cassette tapes, videos)과 디지털 전자자원(예: floppy disks, CD-ROMs, DVDs, VCDs) 모두를 포함하는 것.
- 온라인 자원 : 웹으로 출판되거나 이메일과 같이 다른 도구에 의해 인터넷을 통해 배포되는 디지털 자원. 기관이나 조직 업무 과정에서 내부적 이용을 위해 생산된 온라인 문서는 '기관의 기록'으로 간주하여 컬렉션의 범위에서 제외.
- 디지털화 자원 : 기존 장서에 대한 접근을 향상시키기 위한 디지털화 프로그램의 결과로 생산 것 ; 호주의 사진, 지도, 악보, 책, 구전 역사, 민간전승 기록물, 신문, 기타 간행물의 디지털 사본.

3.2 전자자원 수집 지침

(1) 물리적 포맷 전자자원

- 전자 출판물을 위한 국가 납본 법 제정 때까지, 도서관은 호주의 출판물을 CD-ROMs, DVDs, 플로피 디스크의 포맷으로 자발적인 납본을 통해 수집하기 위해, 출판사, 생산자와 협상한다. 자발적인 납본으로 수집되어질 수 없거나, 다수의 사본이 요구될 경우, 구매를 고려한다.

- 이러한 포맷의 출판물은 Acquisition Guidelines for Australian Physical Format Electronic Publications(www.nla.gov.au/policy/acqguide.html)에 정해진 선정 기준을 따른다.

- 호주에서 출판되는 것을 수집한다. 내용이 호주에 대한 것이나, 호주 저자에 의한 것이어야만 하는 것은 아니다. 출판물의 인쇄와 물리적 포맷 전자 버전 모두 존재하는 경우, 도서관은 인쇄버전을 수집한다. 출판물의 온라인과 물리적 포맷 전자 버전이 모두 이용 가능한 경우, 도서관은 온라인 버전을 수집한다. 버전 간에 중요한 차이가 있다면, 두 가지 또는 모두를 수집한다.

- 인쇄 출판물을 동반하거나 혼합된 미디어 키트의 부분인 물리적 포맷 전자 출판물을 수집한다. 필름, 음악, 소프트웨어, 클립아트 단독으로 존재하는 아이템은 수집하지 않는다.

- 비디오 포맷은 수집하지 않을 수도 있다. 비디오 포맷 자료를 수집하는 것과 관련하여, National Film과 Sound Archive와 같은 타 기관과 협력한다.

(2) 온라인 자원

- 호주의 웹 아카이브인 PANDORA에서 선정된 호주의 온라인 자원을 수집하고 보존한다. PANDORA는 mainland and territory libraries, National Film and Sound Archive, Australian War Memorial, Australian Institute of Aboriginal and Torres Strait

Islander Studies와 협력한다. 참여하는 각 기관은 자신의 선정 가이드라인에 충족하는 온라인 출판물을 선정하고 보존한다.

- 출판사의 허가를 얻어, 도서관은 출판사의 웹사이트로부터 온라인 출판물을 복사하고 아카이브에 사본을 배치한다. Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia(pandora.nla.gov.au/selectionguidelines.html)에 따라 선정한다.

- 사회, 정치, 문화, 과학, 경제적 중요성에 대한 것이거나 호주와 관련한 것 또한 호주 저자에 의해 생산된 온라인 자원을 수집하고 보존한다. 권위 있는 내용, 연구 가치, 그것의 유형에 대해 혁신적이거나 뛰어난 사례로서 화제나 관심과 같은 요소들은 선정 과정에서 고려한다. 도서관은 일반적으로 정확한 인쇄 동등물이 있는 경우 온라인 자원을 수집하지 않고, 인쇄 동등물을 대신 수집한다.

- 국가 복지와 ACT 정부 온라인 출판물을 수집하는 것에 우선순위를 둔다. 온라인 정부 출판물을 수집하는 것에 대한 책임은 PANDORA Archive에 참여하고 있는 관련 주와 Northern Territory 도서관에 있다.

- 제3의 교육 기관, 회의록의 온라인 출판물과 전자저널에도 우선순위를 둔다.

- 표본 추출은 개인의 홈페이지, 판촉 사이트, 기관의 웹사이트, 무상 출판과 같은 온라인 출판물의 많은 다른 유형을 수집하는데 사

용되어 진다. 진행 중인 초안과 저작은 선정하지 않는다.

- 온라인 출판물에 대한 납본 법 조항의 부재와 저작권의 제한으로 인해, 하비스트를 통한 자원에 대한 접근은 제한될 것이다.

(3) 디지털화 자원

- 호주의 사진, 지도, 악보, 구전 역사, 민간전승 기록물, 자관의 컬렉션으로부터 선정된 원고 아이템과 신문에 대한 디지털 사본을 생산하는 것에 초점을 둔 디지털화 프로그램을 가진다. 사본은 도서관의 목록을 통해 접근이 가능하도록 만들어지고, 목록되고, 저장된다. 디지털화 프로그램의 내용은 정기적으로 검토되며 선정은 도서관의 디지털화 정책(www.nla.gov.au/policy/digitisation.html)

을 따른다.

(4) 전자자원의 보존

- 모든 유형의 전자자원은 손실의 위험이 있다. 전자자원을 위해 호주의 기록 유산의 부분으로 수집하거나 생산하기 위한 결정을 하는 보존 프로그램을 가진다. (www.nla.gov.au/policy/digpres.html).

- 호주의 물리적 포맷 전자자원을 위한 보존 업무는 자료의 연구 가치에 따라 행해진다. 실제 물리적 아이템 보다는 보존되는 정보의 내용에 따른다. 더 많은 정보는 Preserving Australian Physical Format Electronic Publications-Selection Guidelines(www.nla.gov.au/policy/selectgl.html)에서 얻을 수 있다.