

발 간 등 록 번 호

11-1371029-000181-01



국립중앙도서관 디지털자원 보존전략 수립 연구

2018. 12.

2018 연구보고

국립중앙도서관 디지털자원 보존전략 수립 연구

2018. 12.



국립중앙도서관
National Library of Korea

제 출 문

국립중앙도서관장 귀하

본 보고서를 “국립중앙도서관 디지털자원 보존전략 수립 연구”의 최종
보고서로 제출합니다.

2018. 12.

-
- 연구수행기관 : 한성대학교 산학협력단
 - 책임 연구원 : 서 은 경 (한성대학교 크리에이티브인문학부 교수)
 - 공동 연구원 : 최 상 희 (대구가톨릭대학교 도서관학과 교수)
박 희 진 (한성대학교 크리에이티브인문학부 교수)
윤 소 영 (서울여자대학교 연구원)
김 지 성 (한성대학교 사회과학부 교수)
 - 보조 연구원 : 감 미 아 (연세대학교 문헌정보학과 박사과정)
서 혜 린 (연세대학교 문헌정보학과 석사과정)
정 의 연 (대구가톨릭대학교 문헌정보학과 석사과정)
신 규 정 (한성대학교 조교)
-

이 연구는 2018년도 국립중앙도서관 연구개발비로 수행되었으며,
본 연구에서 제시된 정책 제언이나 의견 등은 국립중앙도서관의
공식 의견이 아니라 본 연구진들의 개인 견해를 밝히 둡니다.

■ 요약문	13
-------	----

제 1 장 서 론 / 17

1.1 연구의 필요성	19
1.2 연구의 목적	22
1.3 연구의 범위	23
1.4 연구의 방법	24
1.5 연구의 기대효과	26

제 2 장 디지털자원의 특성 및 보존 현황 분석 / 27

2.1 디지털 보존의 개념	29
2.1.1 디지털 보존의 정의	29
2.1.2 디지털 보존의 필요성	35
2.1.3 디지털 보존의 범위	39
2.2 디지털자원의 유형 및 특성	44
2.2.1 자원 유형별 분석	44
2.2.2 제공 유형별 분석	46
2.2.3 미디어 형태별 분석	48
2.2.4 구성요소별 분석	52
2.3 디지털자원의 보존 동향	54
2.3.1 전자납본 사례	54
2.3.2 웹 아카이빙 사례	69

2.3.3 디지털화 사례	82
2.4 종합 요약 및 시사점	97

제 3 장 국립중앙도서관 디지털자원 보존 실무 개선 / 101

3.1 디지털자원 보존용 메타데이터	103
3.1.1 디지털자원 보존 메타데이터 설계원칙 및 국제 표준	104
3.1.2 PREMIS 기반 디지털 보존 메타데이터 연구 분석	111
3.1.3 국립중앙도서관 보존용 메타데이터 분석	118
3.1.4 국립중앙도서관 보존용 메타데이터 개선 방안	130
3.2 디지털자원 관련 업무 분석 및 KOLIS II 보존 기능	134
3.2.1 국립중앙도서관 디지털자료 관련 업무 분석	134
3.2.2 KOLIS II 보존 기능 개선(안)	138
3.3 디지털자원 포맷 레지스트리	145
3.3.1 포맷 레지스트리의 필요성	145
3.3.2 포맷 레지스트리 사례 분석	146
3.3.3 국립중앙도서관 포맷 레지스트리 분석	156
3.3.4 국립중앙도서관 포맷 레지스트리 개선(안)	160
3.4 종합 요약 및 시사점	163

제 4 장 국립중앙도서관 디지털자원 보존 정책 / 169

4.1 디지털자원 보존 정책	171
4.1.1 각 국가별 디지털자원 보존 정책 현황	171
4.1.2 디지털자원 보존 정책 개발을 위한 가이드라인	183
4.1.3 국립중앙도서관 디지털자원 보존 정책(안)	191
4.2 디지털자원 보존 환경 및 시스템 지침	201
4.2.1 디지털 보존을 위한 표준 및 시스템 기능 요건 현황	201
4.2.2 디지털자원 보존 리스크 분석	225
4.2.3 디지털자원 보존 운영 지침	235

4.3 디지털자원 보존 조직	258
4.3.1 조직 구성을 위한 개념적 프레임	258
4.3.2 국립중앙도서관 디지털자원 보존 조직 구성(안)	271
4.4 종합 요약 및 시사점	279

제 5 장 결론 및 후속과제 / 285

5.1 종합 요약	287
5.2 후속과제 영역	292
■ 참고문헌	294

[표 2-1] 디지털 큐레이션 프로세스	34
[표 2-2] 디지털 보존 프로세스 범위	40
[표 2-3] 국립중앙도서관 온라인 자료 수집 현황	45
[표 2-4] 일본 의회도서관 수집 온라인 자료 조건	68
[표 2-5] 기관 유형별에 따른 수집 빈도	81
[표 2-6] WARP에서 파일 단위로 수집된 자료의 유형과 비율	82
[표 2-7] 호주 국가도서관의 이미지 캡처 표준	87
[표 2-8] 보존 및 배포를 위한 오디오 디지털 파일 규격	95
[표 2-9] 보존 및 배포를 위한 영상 디지털 파일 규격	95
[표 3-1] 국립중앙도서관의 PREMIS 보존 메타데이터 - 객체 개체 메타데이터 요소	120
[표 3-2] 국립중앙도서관의 PREMIS 보존 메타데이터 - 이벤트 개체 메타데이터 요소	122
[표 3-3] 국립중앙도서관의 PREMIS 보존 메타데이터 - 행위자 개체 메타데이터 요소	122
[표 3-4] 국립중앙도서관의 PREMIS 보존 메타데이터 - 권한 개체 메타데이터 요소	123
[표 3-5] SPOT, OAIS 분석을 통한 보존 핵심 메타데이터 요소 제시	127
[표 3-6] 디지털자원의 보존 생애주기 메타데이터 비교 분석	129
[표 3-7] 국립중앙도서관의 PREMIS 기반 보존 메타데이터 개선(안)	133
[표 3-8] KOLIS II 개발 이력	139
[표 4-1] 미국의 디지털 컬렉션 구축과 디지털 보존 정책수립을 위한 3단계 계획	174
[표 4-2] 캐나다 국가도서관이 제시한 디지털 보존 정책 가이드라인 목차	185

[표 4-3] 디지털 보존 정책에 수록되어야 할 내용에 대한 가이드라인(영국)	186
[표 4-4] METS를 활용한 보존 메타데이터 구조	207
[표 4-5] OAIS의 기능 및 하위기능 요소	214
[표 4-6] 본 디지털자료 보존을 위한 고려 사항	236
[표 4-7] 파일 포맷 및 저장매체의 전송 및 접근 옵션	238
[표 4-8] 입수 절차 - 저장 및 보존을 위한 데이터 및 문서	240
[표 4-9] 수집 및 평가 - 추천 체크리스트	241
[표 4-10] 디지털 보존을 위한 IT 스토리지 시스템 사용 원칙	245
[표 4-11] NDSA 디지털 보존을 위한 권장 수준	248
[표 4-12] 직무분석의 절차	260
[표 4-13] 직무조사 조사항목	261
[표 4-14] 업무량산정 조사항목	262
[표 4-15] 국립중앙도서관 '부' 및 '과' 소속 인력	272
[표 4-16] Brown 보존 업무 프로세스와 국립중앙도서관 장기보존 핵심영역	273

그림 차례

[그림 2-1] 호주 웹 아카이빙 수집 범위의 변화	75
[그림 2-2] Open UK Web Archive	77
[그림 2-3] JISC 영국 웹 도메인 데이터 세트	77
[그림 2-4] DigiLab을 통한 디지털화	86
[그림 2-5] DigiLab을 통한 디지털화 문서 사례 2	86
[그림 3-1] PREMIS Data Model	106
[그림 3-2] PREMIS Data Model version 3	109
[그림 3-3] BnF의 SPAR 보존 메타데이터	116
[그림 3-4] BnF의 SPAR 보존 데이터 흐름	117
[그림 3-5] 생애주기 관리	118
[그림 3-6] 국립중앙도서관의 PREMIS 기반 보존 메타데이터	119
[그림 3-7] SPOT(Simple Property-Oriented Threat) 분석	126
[그림 3-8] SPOT 기반 보존 위험 평가	126
[그림 3-9] PREMIS 보존 메타데이터	130
[그림 3-10] 온라인 납본 업무 프로세스	134
[그림 3-11] 온라인 납본 시스템 흐름도	135
[그림 3-12] 웹사이트 및 웹자료 처리 업무 프로세스	136
[그림 3-13] 웹자료 처리 시스템 흐름도	136
[그림 3-14] 웹사이트 처리 시스템 흐름도	137
[그림 3-15] 디지털화 업무 프로세스	138
[그림 3-16] 디지털화 시스템 흐름도	138
[그림 3-17] KOLIS II 콘텐츠 관리기능	140
[그림 3-18] KOLIS II 업무관리시스템 관리기능	141
[그림 3-19] KOLIS II 와 장기보존 개념도	142

[그림 3-20] PRONOM 이용 예시화면 - 검색어 “HWP”	147
[그림 3-21] PRONOM 파일 포맷 검색 구성요소	148
[그림 3-22] PRONOM4의 high-level information model	149
[그림 3-23] GDFR의 Data model	152
[그림 3-24] UDFR Class Hierarchy (summary)	154
[그림 4-1] 캐나다 국가도서관의 디지털 보존 프로그램 전략 프레임워크	179
[그림 4-2] 정책수립 절차	184
[그림 4-3] METS 구조 맵	207
[그림 4-4] 주요 메타데이터 유형과 관계	211
[그림 4-5] OAIS 기능 요소	212
[그림 4-6] 기록관리시스템의 기능 모델	216
[그림 4-7] 영구기록관리시스템의 기본 기능	219
[그림 4-8] 영구기록관리시스템의 기능 요소	221
[그림 4-9] 영구기록관리시스템의 기능별 프로세스 흐름도	222
[그림 4-10] 기능구조 예시	263
[그림 4-11] 사업구조 예시	264
[그림 4-12] 국립중앙도서관 조직도와 센터조직	265
[그림 4-13] 국가기록원 조직도와 센터조직	267
[그림 4-14] 한국과학기술정보연구원 조직도와 센터조직	268
[그림 4-15] 미국 의회도서관 부서 조직도	269
[그림 4-16] 일본 의회도서관 조직도	270
[그림 4-17] 디지털자료 장기보존과 관련된 영역과 주요 업무	274

요약문

현대 사회에는 다양한 형태의 수많은 디지털자원이 생산·유통되고 있다. ‘디지털 자원’은 유형 면에서 보면 단순히 디지털 정보기술의 산물이자 매개일 뿐이지만, 내용 면에서 보면 ‘디지털 시대’를 대표하는 인류의 지적·문화적 성과이며, 정치, 경제, 사회 그리고 문화 영역 등 인류의 삶에 관련된 모든 영역에서 소비되는 중요한 사회적 자산이자 유산이다. 이에 따라 무형자산인 디지털자원의 중요성에 대한 인식은 점점 높아지고 있고 디지털자원이 지닌 공유성 및 유용성으로 인하여 그 활용가치를 보장 받고 있다. 하지만, 우리 사회에서 디지털 기술의 급속한 발달로 인하여 생산된 지 얼마 안 되는 디지털자원이 빠르게 사라지는 역설적 현상이 빚어지고 있다.

새로운 정보기술의 지속적인 개발로 인하여 디지털자원은 빠르게 노화되고 있고 활용을 위한 접근이 어려워지고 있다. 이에 따라 기술변화와는 상관없이 디지털자원을 지속적으로 활용할 수 있게 하는 디지털 보존 업무는 디지털자원의 관리 업무 중 에서 핵심적인 프로세스로 간주되기 시작하였다. 특히 온라인 자료의 납본이 시행 3년차를 맞이하여 국립중앙도서관이 점차적으로 방대해진 디지털자원을 보유하게 되자, 디지털자원에 대한 체계적인 관리 및 디지털자원 보존 방향을 모색해야 할 필요성이 높아졌다. 이에 본 연구는 국내에서 발간되는 인쇄자료자원과 동일하게 디지털자원을 후대에 영구적으로 전송하기 위하여 국립중앙도서관 디지털자원 보존전략을 수립하는 데 그 목적을 가진다.

이를 위하여 본 연구는 문헌 고찰 및 국내외 자료 분석을 통해, 디지털 보존의 개념과 범위, 필요성을 도출하고, 국내·외 유통되고 있는 디지털자원의 현황 및 디지털자원 보존 프로젝트, 디지털 보존 메타데이터 표준, 포맷 레지스트리에 대한 최근 동향을 분석하였다. 이러한 문헌적 고찰과 사례분석 결과를 기반으로 하여 국립중앙도서관에 적합한 디지털 보존 정책, 디지털 보존 실무 개선방향, 디지털 보존을 위한 조직 등에 관한 개선방안이 제시되었다. 검토된 내용과 제안된 사항을 종합적으로 살펴보면 다음과 같다.

첫째, 각국의 국가도서관 디지털 보존과 관련된 현황 조사에서 나타난 주목할 만한 사항은 디지털자료의 단순한 보존보다는 디지털자료의 활용성을 강조하고 이에 대한

인식이 보편화되어 가고 있다는 것이다. 디지털 장기보존은 단순히 디지털자료 복본을 안전하게 보존하는 체제가 아니라 후세가 디지털 형태로 보존되어 있는 자료를 활용하도록 보존하는 체제를 구축하는 것이다. 특히 디지털 보존 정책은 국가차원 가치가 있는 디지털자원을 선별 입수하여 소장하여 장기적으로 이용자에게 접근할 수 있는 모든 활동을 포함하는 거시적 정책 및 전략을 중심으로 발전하고 있음을 알 수 있었다. 따라서 디지털 보존 정책은 내부적 업무와 전략, 외부적 협력에 대한 내용이 포함되어야 할 것이다. 디지털 보존 정책에는 기술기반의 단기적 보존전략보다는 국가차원의 비전과 목적달성을 위한 장기적 디지털 보존 개발전략 수립이 개발되어야 할 것이다.

둘째, 국립중앙도서관 디지털자료 관련 업무는 온라인 납본, 웹 아카이빙, 디지털화 등 3개 영역으로 나눌 수 있으며 각각의 프로세스가 독립적으로 운영되고 있으나 기본적인 단계는 각 영역별로 기본 처리 후 디지털장서관리시스템과 KOLIS II 온라인 시스템에 등록하는 과정으로 구성되어 있다. 보다 효율적이고 장기적인 디지털 보존을 수행하기 위해서는 납본 원문을 서비스용으로 원문저장소에 보장하여 서비스를 하고 현재 보존용으로 생성하는 것을 장기보존시스템으로 전송하여 장기보존 포맷으로 전환한 후 장기보존저장소에 저장하여 장기보존 자료를 통합하여 관리하는 체제를 제안하였다. 또한 디지털화 자료 역시 온라인 자료와 동일한 체제로 서비스용은 원문저장소에서 관리하고 장기보존용으로 장기보존 포맷으로 전환한 후 장기보존관리체제에 통합하여 관리 방안을 제안한다. 한편으로 웹사이트와 웹자료는 현행에서 수행하고 있는 2개의 프로세스를 반영하여 파일 단위의 웹자료는 납본자료와 유사하게 처리하고 웹사이트는 자동으로 수집하여 메타데이터가 생성되어 원문저장소에 저장시 완료되면 특정 시점을 설정하여 연단위로 추가된 데이터베이스본을 장기보존저장소에 백업형태로 저장하며 웹사이트 구성요소의 포맷을 포맷 레지스트리에 등록하는 체제로 운영하는 것을 제안한다.

셋째, 2011년 설계된 국립중앙도서관 PREMIS 기반 보존 메타데이터는 KOLIS에 적용되어 국립중앙도서관 디지털자원의 기술메타데이터인 MODS의 기본 정보와 연계가 가능하도록 설계되었으나, 그 구조가 매우 복잡하며 대부분의 하위요소들이 필수 값으로 지정되어 있어 실질적으로 적용이 쉽지 않았다. 이러한 문제점을 해결하고자, SPOT과 OAIS 기반의 보존 위험과 기능 분석을 통해 국립중앙도서관의 보존 메타데이터 핵심요소를 파악하였다. 또한 국립중앙도서관의 디지털자원에 적용되고 있

는 메타데이터를 분석하여 연계를 통해 자동으로 추출할 수 있는 요소들을 확인하고 핵심요소들로만 구성된 PREMIS 개선(안)을 제안한다.

넷째, KOLIS II와 장기보존시스템은 독립적으로 구축하고 서로 연계하도록 설계하여 확실한 관리 주체에 의해 일관성 있게 운영, 관리하는 것을 목표로 하며 콘텐츠 메타데이터인 MODS와 함께 METS를 이용하여 패키징하는 장기보존시스템을 제안한다.

다섯째, 국립중앙도서관의 디지털자원의 장기보존 포맷 관리를 위해 파일 식별 기술 연구에 있어 가장 많이 참조되고 있는 PRONOM의 오픈소스 DROID의 식별 방법과 포맷 시그니처 표현 규칙을 참조하여 포맷 레지스트리를 구현할 것을 제안하며 KOLIS II의 보존 기능, 장기보존시스템과 연계하는 것을 제안한다. 또한 마이그레이션, 인캡슐레이션 등 적절한 보존전략을 수립하여 수행하는 과정에서 새로이 생성되는 기술(technical) 메타데이터를 포맷 레지스트리에 등록, 관리하는 체제의 구축을 제안한다.

여섯째, 디지털 장기보존 조직 구성에 관련 사례 및 현황조사 분석결과 국립중앙도서관 업무 중 디지털 장기보존에 관련된 업무가 연구 및 기획, 장기보존운영, 웹보존운영, 디지털지원 등 여러 개로 분할(divisionalization)되어 있다. 그러나 디지털 장기보존의 업무 특성 상 업무의 한계가 명확하며, 자료유형 및 기능의 특정성과 현 국립중앙도서관 업무 프로세스 상 하나의 독립적 단계라는 업무의 독자성 및 계속성이 존재하므로 여러 개로 분할된 업무를 통합하여 전담하는 조직체계를 구성하는 것을 제안한다. 따라서 디지털 장기보존 조직의 규모는 중앙행정기관의 ‘국’에 준하여 설정하는 것이 타당하다고 판단된다. 이에 따라 국립중앙도서관의 경우 하부조직인 ‘부’ 조직이 중앙행정기관의 ‘국’에 준하는 기능을 수행하며 ‘국’ 수준의 규모로 디지털 장기보존 조직의 규모는 국립중앙도서관 내 ‘부’의 위상에 준하여 설정함을 제안한다.

본 연구는 마지막으로 분석된 내용을 토대로 하여 다음과 같이 3단계 후속 연구과제를 제안한다. 지원환경 구축을 목적으로 하는 1단계에서는 디지털 장기보존 세부정책 개발과 효율적인 디지털 장기보존 조직 개발을 위한 세부 직무분석 및 인력산정과 관련된 과제를 제안한다. 이를 토대로 하여 실무적 절차를 시스템화하는 2단계에서는 디지털 보존 자원의 컴포넌트 유형별 보존 속성 및 렌더링 기술 분석, 디지털자원의 장기보존을 위한 포맷 레지스트리 설계 및 구현, 장기보존시스템과 KOLIS II 온라인 시스템의 메타데이터 자동 연계 시스템 구축, 장기보존시스템 패키징 모듈화 및 장기

보존을 위한 세부 기능 설계, 장기보존을 위한 에플레이팅 체제 구축과 관련된 과제를 제안한다. 마지막으로 디지털 장기보존의 안정화를 목표로 하는 3단계에서는 장기보존 컬렉션에 기반 한 콘텐츠 재생산, 장기보존 컬렉션 서비스 방안 수립, 장기보존을 위한 마이그레이션 영역으로 장기보존을 위한 마이그레이션 로드맵과 장기 전략 수립을 제안한다.

디지털 보존은 뜻있는 개인들의 노력만으로는 해결될 수 없는 광범위한 과제이며, 국가도서관과 지식정보의 생산 및 유통, 보존에 관여하는 모든 이해관계자들과 함께 협력하여 풀어나가야 하는 시대적 과제이자 더 나아가 인류 문명적 차원의 과제이다. 이에 따라 국립중앙도서관의 디지털 보존 과제를 수행하는 중요한 첫걸음인 국립중앙도서관 디지털 보존전략 수립은 인류사회 기억을 보존하고 전승하도록 하는 시대적 과제를 수행한다는 의미를 가진다.

제 1 장

서 론

- 1.1 연구의 필요성
- 1.2 연구의 목적
- 1.3 연구의 범위
- 1.4 연구의 방법
- 1.5 연구의 기대효과

제 1 장

서 론

1.1 연구의 필요성

- 현대 사회에는 다양한 형태의 수많은 디지털자원이 생산·유통되고 있다. ‘디지털 자원’은 유형 면에서 보면 단순히 디지털 정보기술의 산물이자 매개일 뿐이지만, 내용면에서 보면 ‘디지털 시대’를 대표하는 인류의 지적·문화적 성과이며, 정치, 경제, 사회 그리고 문화 영역 등 인류의 삶에 관련된 모든 영역에서 소비되는 중요한 사회적 자산이자 유산인 것이다(오성환 2018).
- 이에 따라 무형자산인 디지털자원의 중요성에 대한 인식은 점점 높아지고 있고 디지털자원이 지닌 공유성 및 유용성으로 인하여 그 활용가치를 보장 받고 있다. 하지만, 우리 사회에서 디지털 기술의 급속한 발달로 인하여 생산된 지 얼마 안 되는 디지털자원이 빠르게 사라지는 역설적 현상이 빚어지고 있다. 즉 새로운 정보기술의 지속적인 개발로 인하여 정보환경과 저장방식 및 매체가 지속적으로 변화하게 되고 이러한 잦은 변화는 디지털자원을 빠르게 노화시키고 활용을 위한 접근을 어렵게 만들고 있다.
- 기술변화와는 상관없이 디지털자원이 지속적으로 활용될 수 있게 하는 노력인 디지털 보존 업무는 디지털자원의 관리 업무 중에서 핵심적인 프로세스로 간주되기 시작하였다. 디지털 보존은 현재의 소프트웨어, 하드웨어, 파일 포맷, 관리시스템 등등이 모든 것들이 무용지물이 될지라도 미래의 이용자가 디지털자원에 접근하여 이용할 수 있도록 하는 모든 활동을 말한다(Digital Preservation Coalition 2015).

-
- 이러한 기능을 지닌 디지털자원의 보존 업무의 필요성은 현재와 같이 새로운 기술이 끊임없이 발표되는 환경에서는 더욱 강조되고 있다. 특히 디지털 보존에 대한 관심과 투자가 시급한 이유는 우선 갈수록 더 많은 지적·문화적 성과들이 디지털 형태로 생산되어 다른 형태로는 존재하지 않을 것인데, 장기적 보존을 염두에 두지 않거나 아무런 조치를 취하지 않으면 아무런 경고 없이, 아무도 알아차리지 못하는 사이에 디지털자원 형태의 지적 성과들이 변경되거나 소멸될 수 있기 때문이다.
 - 최근 많은 연구자들은 디지털자원 구축에 대한 기법 및 활용에 대한 관심 못지않게 장기적 접근의 보장성을 높이는 디지털 보존에도 많은 관심을 가지게 되었다. 일찍이 디지털자원의 장기적 보존은 디지털자료의 생애주기 기반으로 한 표준적인 기술 환경에서 이루어져야 성공할 수 있다는 전제 아래, 2002년에 NASA의 CCSDS (Consultative Committee for Space Data Systems)는 기술적 표준 권고안을 개발하였다. 즉 이들은 디지털정보를 영구적이거나 무기한 장기적으로 보존하는데 있어서 요구되는 기술적 권고안인 OAIS 참조모형(Open Archival Information System Reference Model)을 제시하였다. 일 년 뒤에 이 참조모형은 ISO 14721:2003으로 국제적 표준으로 지정되었고 이를 기반으로 해서 디지털 보존에 관한 기본 개념과 용어에 대한 국제적인 합의를 도출하였다(Nicholson & Dobрева 2009).
 - 이후 미국, 영국, 호주 또는 캐나다의 국가도서관이나 국가기록관 또는 연구기관에서 국가적 프로젝트가 진행되었고 2012년에는 OAIS 참조모형이 개정되어 발표되었지만, 여전히 디지털 보존에 대한 장기적 계획 수립 및 활동은 미비한 편이다. 이는 디지털 보존 업무가 디지털자료의 생애주기에 맞추어 수행됨에 따라 각 단계별 디지털 보존 업무가 복잡하게 연관되어 있고 또 각각의 디지털 보존 방법론과 요구조건이 계속해서 변화됨에 따라 적합한 방법을 선택하는 일조차도 어려운 일이 되었기 때문이라 할 수 있다(서은경 2016).
 - 그러나 이해관계에 따라 폐기되거나 변조될 가능성이 높은 디지털자원을 보존하는 일은 개인이나 기관 차원보다는 공익과 안정성을 보장할 수 있는 국가기관에서 수행해야 한다는 인식이 보편화되면서 각 국가별로 디지털자원 보존에 대한 전략이 수립되고 점차적으로 제도화되기 시작하였다. 즉 장기적 보존을 위한 디지털자료의 수집 범위 및 업무절차, 기술적 문제 등에 직면하면서 보다 실질적이고 효과적인 디지털 보존을 위한 정책 및 전략 수립과 실행이라는 당면과제가 대두된 것이다

(Kastellec 2012).

- 더욱이 우리나라의 경우, 2016년 법 개정을 통해 도서관법 제 19조 및 20조에 의거 납본의 시행범위가 인쇄자료에서 국제표준자료번호를 부여 받은 온라인 자료로 확대되었다. 개정된 법에는 오프라인 디지털자료의 납본 시점을 인쇄자료와 동일하게 발행일 또는 제작일을 납본의무 발생일로 규정하고 있지만 납본된 디지털자료의 이용에 대한 법률적 제한점에 대해서는 논의되고 있지 못한 채, 온라인자료의 납본이 시행 3년차를 맞이하였고 국립중앙도서관이 보유하고 있는 디지털자원은 점차적으로 방대해지고 있는 실정이다.
- 비록 현재에는 국가도서관의 디지털자원에 대한 접근에 대한 디지털자원의 이해당사자와 이용자 간의 의견이 충돌되고 있지만, 이용자가 지속적으로 디지털자료를 활용할 수 있도록 하는 것은 국가도서관의 임무라 할 수 있다. 따라서 디지털자료의 장기보존과 이용 보장을 위한 모든 활동은 이제 국가도서관의 핵심 업무로 간주되고 있다. 특히 디지털자료의 보존 업무는 인쇄자료의 보존 방식과는 완전히 다르고 그 책무조차 광범위하므로 국가차원의 디지털 보존 업무에 대한 실용적인 대안을 모색하는 것이 시급하며 절실히 필요하다.
- 분명한 것은 지속적으로 많은 양의 디지털자원들이 국립중앙도서관에 의하여 관리될 것이며 또 이러한 디지털자원은 점점 노화된다는 점이다. 이러한 사실을 직시하고 디지털 보존을 위한 정책 및 업무절차가 미래지향적으로 수립되어야 할 것이며 디지털 보존을 위한 기술적 역량을 확장시켜야 할 것이다. 이제 디지털자원 중심로의 패러다임 변환을 인지해야 하고 미래를 위한 준비의 하나인 디지털 보존에 대한 관심과 노력은 이제 선택이 아닌 필수인 것이다.
- 디지털 보존은 뜻있는 개인들의 노력만으로는 해결될 수 없는 광범위한 과제이며, 국가도서관과 지식정보의 생산 및 유통, 보존에 관여하는 모든 이해관계자들과 함께 협력하여 풀어나가야 하는 시대적 과제이자 더 나아가 인류 문명적 차원의 과제이다. 이에 따라 국립중앙도서관의 디지털 보존 과제를 수행하는 중요한 첫걸음인 국립중앙도서관 디지털 보존전략 수립은 인류사회 기억을 보존하고 전승하도록 하는 시대적 과제를 수행하는 것이라 할 수 있다.

1.2 연구의 목적

- 본 연구는 국내에서 발간되는 인쇄자료자원과 동일하게 디지털자원을 후대에 영구적으로 전승하기 위하여 디지털자원 보존 정책 수립, 담당 조직 제언, 그리고 보존 실무 위한 개선 방안을 마련하는 것이다. 즉 현 시점에서 이루어지고 있는 각국의 디지털자원의 보존 업무와 정책, 표준에 대한 다각적인 분석을 기반으로 하여 국립중앙도서관 디지털자원의 전반적인 보존전략 즉 장기적인 디지털 보존을 위한 조직, 정책, 실무, 시스템과 관련된 디지털 보존전략을 수립하고자 한다. 따라서 본 연구의 세부 목적은 다음과 같다.
- 첫째, 디지털 보존에 대한 이해와 국가적 동향을 인식하기 위하여 디지털 보존의 개념과 범위 및 디지털자원의 유형을 규명하고 해외 주요 국가대표도서관의 디지털자원 보존 업무의 현황 및 특징을 분석하는 것이다.
- 둘째, 장기적인 보존전략을 수행하기 위하여 국립중앙도서관 디지털자원 보존실무의 개선(안)과 디지털 보존의 물리적 환경 및 시스템의 가이드라인을 제시함으로써 효율적인 디지털자원 보존 업무 수행을 지원하는데 있다.
- 셋째, 국외 주요 국가대표도서관의 디지털 보존 업무 프로세스 및 체제에 대한 분석을 토대로 하여 디지털자원의 영구적인 보존 및 활용을 위한 장기보존 전략을 수립하고 국립중앙도서관에 적합한 디지털 보존 체제를 제안하는 것이다.
- 이에 따른 본 연구의 결과는 국립중앙도서관 디지털자원 보존 정책 수립, 디지털자원 보존 업무 수행을 위한 조직 체계 확보 및 보존 실무를 기초자료로 활용할 수 있을 것이다.

1.3 연구의 범위

□ 디지털자원의 특성 및 보존 현황 분석

- 디지털 보존 개념과 범위, 필요성을 조사 분석하여 현실적으로 국립중앙도서관에서 수용할 수 있는 보존 개념과 범위를 제시한다.
- 국내외에서 생산, 서비스 되는 디지털 정보자원의 동향과 유형, 특성을 조사 분석하여 국립중앙도서관에서 수집, 보존해야 할 대상이 되는 핵심 디지털자원의 생산 현황과 유통 현황 등을 파악한다.
- 디지털자원 보존전략과 실제에 대한 국내외 선행연구 및 사례를 조사 분석하여 국립중앙도서관 디지털자원 보존전략 수립에 대한 시사점을 도출한다.

□ 국립중앙도서관 디지털자원 보존 실무 개선

- 디지털자원 보존에 관한 메타데이터 관련 국제표준 및 동향, 사례를 조사한 후 분석 내용과 문제점을 바탕으로 국립중앙도서관 보존 메타데이터 설계(안)를 제시한다.
- 국립중앙도서관 KOLIS II의 보존 관련 기능 및 디지털자원 보존 메타데이터 수용성을 조사하여 국립중앙도서관 KOLIS II의 디지털자원 보존 기능 개선(안)을 제안한다.
- 국내외의 대표적이며 표준적인 포맷 레지스트리와 국립중앙도서관의 포맷 레지스트리와 비교 분석한 내용을 토대로 하여 국립중앙도서관 디지털자원의 장기보존 포맷 레지스트리 가이드라인을 제시한다.

□ 국립중앙도서관 디지털자원 보존전략 모형 제시

- 국외 국가도서관 보존 대상 디지털자원 체제 및 정책을 분석하여 국립중앙도서관 디지털 보존 정책을 제안한다.

- 디지털자원 보존을 성공적으로 운영하기 위한 기반 환경 구축을 위하여 국제 표준에 기반한 국립중앙도서관 디지털자원 보존 시스템 요건 및 운영지침을 제시한다.
- 국가디지털자원 보존의 허브로서 국립중앙도서관의 역할 분석과 국립중앙도서관 디지털자원 관련 운영 조직체계 구성안을 제안한다.

1.4 연구의 방법

□ 문헌 고찰 및 국내외 자료 분석

- 디지털 보존의 개념과 범위, 필요성을 도출하기 위하여 심층적으로 관련 문헌을 조사한다.
- 각국의 국가대표도서관의 디지털 보존전략, 국가적 디지털자원 보존 프로젝트 동향 및 전망을 도출하기 위하여 기관에서 출판된 보고서, 홈페이지자료, 선행연구를 조사한다.
- 국내·외 유통되고 있는 디지털자원 현황 분석을 위하여 ‘전자납본’, ‘웹 아카이빙’, ‘소장자료의 디지털화’ 등의 범주로 디지털자원의 유형을 세분화하여 국내외 선행연구를 통해 그 특성과 변화 추이를 광범위하게 조사하고 분석한다.

□ 사례 분석

- 국립중앙도서관의 디지털 보존 전반을 분석하기 위하여 디지털자원 보존 업무와 프로세스, 디지털 보존을 위한 조직, 디지털 보존 정책 등을 분석한다.
- 보존 업무의 기능을 강화시키기 위하여 국립중앙도서관 통합자료관리시스템 및 보존시스템의 기능과 디지털자원 포맷, 보존용 메타데이터 등에 대한 현황을 조사한다.
- 해외 국가대표도서관의 디지털자원 보존 정책 및 조직을 심층적으로 조사하고 해외 주요도서관의 디지털자원 보존 업무 및 현황을 분석한다.

□ 국제표준 및 프로젝트 분석

- 디지털 보존을 위한 메타데이터 표준과 디지털 보존을 위한 포맷 레지스트리 최근 동향을 분석한다.
- 전자출판 기관에서 수행하고 있는 보존 프로젝트 사례를 수집하여 검토한다.
- 국가차원에서 수행되고 있는 디지털자원 보존 관련 국제 프로젝트 동향을 분석한다.

□ 포커스 그룹 인터뷰(Focus Group Interview, FGI)

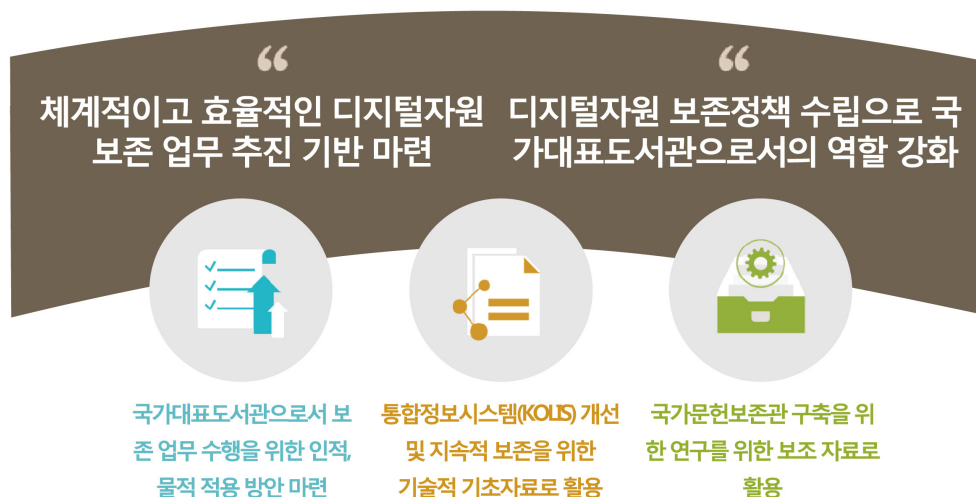
- 디지털자원 보존 실무를 분석하고 개선 방안을 모색하기 위하여, 국립중앙도서관의 디지털자원 보존, 관리 실무 이해관계자를 선별하여, 설문 및 인터뷰를 통해 보존 실무에 대한 요구를 수렴한다.
- 국립중앙도서관 디지털 납본 관련 담당자, 소장자료 디지털화 담당자, OASIS 담당자를 대상으로 한 포커스 그룹 인터뷰를 실시하여 통합정보시스템(KOLIS II 온라인)의 보존 관련 기능, 보존 메타데이터, 포맷 레지스트리와 관련한 문제점을 파악하고 개선 방안을 도출한다.

□ 전문가 자문 회의

- 국립중앙도서관의 보존 방향을 모색하고, 전략적 운영에 대한 개선 방안에 대한 자료를 확보하고, 제안한 보존 정책(보존계획, 보존대상 선별, 평가 원칙 등)을 검증하기 위하여 각 분야의 전문가들의 자문회의를 수행한다.
- 국립중앙도서관 보존 현황에 대한 자문 및 국립중앙도서관이 수행해야 할 기능에 대한 자문을 구하며, 국내 문헌정보학계 교수, 유관기관 전문가, 디지털자원 보존 이해관계자 등을 초빙하여 개최한다.

1.5 연구의 기대효과

- 디지털자원 보존은 한 번에 그리고 단편적으로 이루어지는 업무가 아니다. 즉 여러 부서에서 협동적으로 체계적으로 그리고 장기적으로 이루어지는 업무이다. 따라서 디지털자원 보존전략은 종합적으로 검토되어 각 세부 영역으로 구체화되어야 한다.
- 디지털 보존 업무의 특성에 맞추어 수행된 본 연구의 최대 장점은 성공적으로 디지털 보존이 수행될 수 있도록 정책적인 면, 조직적인 면, 실무적인 면 그리고 기술적인 면을 포괄적으로 다루었다는 점이다.
- 또 하나의 장점은 본 연구에서 제시한 디지털 보존 정책을 비롯한 디지털 보존을 위한 지침, 가이드라인, 리스크 분석 및 체크리스트는 디지털 보존을 수행하는데 있어서 실무적 및 실천적 근거로 활용하는데 유용하다는 점이다.
- 본 연구의 결과는 국립중앙도서관 디지털자원 보존 정책 수립, 디지털자원 보존 업무 수행을 위한 조직 체계 확보 및 보존 실무의 기초자료로 활용할 수 있으며, 이를 통해 체계적이고 효율적인 디지털자원 보존 업무 추진 기반을 마련하고 향후 국가 문헌보존관 구축을 위한 보조 자료로 활용하여 국가대표도서관으로서의 역할을 강화하는 효과가 기대된다.



제 2 장

디지털자원의 특성 및 보존 현황 분석

- 2.1 디지털 보존의 개념
- 2.2 디지털자원의 유형 및 특성
- 2.3 디지털자원의 보존 동향
- 2.4 종합 요약 및 시사점

제 2 장

디지털자원의 특성 및 보존 현황 분석

2.1 디지털 보존의 개념

2.1.1 디지털 보존의 정의

□ 디지털 보존

- 인쇄자료의 보존이란 보존 대상의 형태와 콘텐츠를 변화시키지 않고 있는 그대로 계속 유지하는 것을 말하나, 디지털 보존은 개별적이고 독립적인 기능이 아니라 여러 기능들과 밀접하게 연관되어 있는 종합적인 기능으로 콘텐츠를 장기적으로 유지시켜 지속적으로 이용가능하게 하는 활동을 의미한다.
- 따라서 디지털 보존은 디지털 콘텐츠의 생산 단계에서부터 고려되어야 하며 접근 과도 불가분의 관계에 있다. 보존하지 않은 것에 대한 검색은 불가능하며, 미래의 접근을 보장하지 못하는 보존 또한 의미가 없기 때문이다. 이에 디지털자원을 저장하고 서비스하는 여러 기관들과 이와 관련하여 연구하는 학자들은 다음과 같이 디지털 보존을 정의하고 있다.
- 디지털 보존은 디지털 콘텐츠의 무결성(integrity)을 유지하면서 디지털 매체의 손상 및 하드웨어, 소프트웨어의 문제로 인한 기술적 퇴화의 위협으로부터 디지털 객체를 안전하게 보호하는 일련의 관리활동을 말한다(정혜경 2005).
- 디지털 보존은 특정 데이터 형태가 지속적으로 관리되어 기술적인 변화 속에서도

계속 접근가능하고, 이해될 수 있도록 하는 아카이빙 활동의 일부로 볼 수 있다. 디지털 정보자원의 무결성을 상실하지 않고, 저장매체의 손상 및 퇴화, 기술적 퇴화의 위협으로부터 디지털정보를 안전하게 보호하여 유용한 형태로 존속하게 하는 일련의 관리활동으로 정의할 수 있다(안영희, 박옥화 2009).

- 디지털 보존은 디지털정보자원의 진본성과 무결성을 상실하지 않게 하고, 저장매체의 손상, 퇴화, 기술퇴화로부터 디지털정보를 보호하여 유용한 형태로 존속하게 하는 일련의 활동이며 디지털 보존의 궁극적 목적은 바로 디지털 콘텐츠를 신뢰할 수 있도록 보존하고 지속적으로 이용하도록 하는 활동이다(ICPSR 2012).
- 디지털 보존은 단순한 디지털 데이터의 저장 및 유지가 아니라 보존 가치가 있는 자료의 선정, 복원 및 디지털화, 디지털자원의 소장, 디지털 객체의 진본성 및 무결성 유지, 출처와 상호관련성 유지, 지속적 이용의 활성화 등을 수행하는 일련의 업무이다. 디지털 보존 업무는 다양한 부서와 연관되므로 부서간의 유기적 협력이 이루어지는 것이 중요하다(서은경 2016).
- 디지털 보존은 지속적인 접근이 가능하도록 디지털 콘텐츠를 관리하는 활동이다(Library of Congress 2016).
- 디지털 보존은 디지털 콘텐츠의 보존(preservation)과 보존처리(conservation) 활동을 수행하는 것으로 간주하면 된다. 일상적으로 보존은 지적·문화적 자산을 보호하기 위한 제도적, 물리적, 기술적 측면 모두를 망라한 제반 활동이고, 보존처리는 지적·문화적 자산을 적극적으로 보호하기 위한 처리, 변환과 같은 과학적 방법을 수행하는 작업을 의미하며, 디지털 보존은 이 두 가지 업무를 동시에 수행하는 것으로 보면 된다(The American Institute for Conservation of Historic and Artistic Works 2016).
- 디지털 보존은 필요할 때까지 디지털자원에 지속적으로 접근이 가능하게 하는 일련의 관리 활동이다. 미래에 소프트웨어, 하드웨어 파일 포맷, 관리자 등등 이 모든 것들이 무용지물이 될지라도 이용자가 디지털정보에 접근하여 이용할 수 있도록 하는 모든 활동을 말한다(DPC 2015).
- 디지털 보존은 저장된 디지털자원에 대한 기술적 측면만을 다루는 것도 아니고 단순

히 접근만을 강조하는 것도 아니다. 또한 단 한 번의 작업으로 이루어지는 것이 아니라 지속적이고 체계적인 관리 측면이 강조되는 활동이다(Corrado & Sandy 2017). 따라서 이들은 다음과 같은 세가지면을 강조하였다.

- 디지털 보존은 단순히 백업(backup)과 디지털 복구(recovery) 작업이 아니다.
- 디지털 보존은 오직 접근(access)을 위한 활동도 아니다.
- 디지털 보존은 나중에 생각해서 수행하는 일(afterthought)도 아니다.
- 백업(Backup)은 데이터에 중점을 두어 데이터가 손상되거나 손실될 경우를 대비해 저장하는 데이터의 사본을 말하며, 백업의 목적은 사고가 났을 때 데이터를 복원(Restore)하는 것이다. 즉 누군가 실수로 또는 악의로 파일이나 콘텐츠를 지워버렸다면 또 랜섬웨어에 의하여 모든 파일이 암호화되었을 때, 백업을 이용하여 모든 데이터를 복원할 수 있다는 것이다. 이와 같이 데이터 복원 기능을 위한 백업 업무는 일반적으로 디지털 보존 업무의 하나로 간주되고 있다.
- 디지털 보존은 필요한 만큼 디지털자료에 대한 지속적인 액세스를 보장하기 위해 필요한 일련의 관리 활동을 나타낸다. 따라서 디지털 보존은 그 디지털자료의 보존 기간에 그 목표와 기능을 달리한다.
 - 단기보존: 사용이 예측되는 일정 기간 동안 디지털자료에 접근 할 수 있지만 예측 가능한 미래를 넘어서지 않으며, 기술의 변화로 인해 접근이 불가능해질 때까지 접근할 수 있다.
 - 중기보존: 지정된 기간 동안 기술 변화와 상관없이 디지털자료에 계속 접근할 수 있지만 무기한은 아니다.
 - 장기보존: 디지털자료 또는 디지털자료에 포함된 정보에 계속 접근할 수 있다.
- 결론적으로 디지털 보존이라 함은 미래의 후손들이 용이하게 볼 수 있고 이해할 수 있는 방식으로 디지털자원을 보존하는 것을 말한다. 따라서 디지털 보존은 가독성과 보존성이라는 두 가지 문제 해결을 목표로 한다. 이제 디지털자료의 장기보존은 인류 지식 프로젝트라 칭하며 도서관, 박물관 등을 중심으로 지금까지도 많은 노력이 이루어지고 있다.

□ 디지털 아카이빙

- 아카이빙이라는 용어는 기록이나 문서를 저장하여 보존하기 위하여 수행되는 처리 업무 또는 컴퓨터 파일의 일시적 백업을 의미하는 것으로 사용되어 왔다. 그러나 2000년대 초반에 디지털자원의 확대에 따른 디지털자원의 아카이빙에 대한 논의가 디지털 리포지터리(repository) 구축과 함께 본격적으로 시작되었고 이러한 활동을 디지털 아카이빙이라는 용어로 지칭되었다.
- 일반적으로 디지털 아카이빙은 지속적 가치가 있다고 판단되는 디지털 객체를 장기간 관리하여 이후의 이용을 보장하는 활동을 말하며 디지털 아카이브 운영, 디지털 객체 관리, 그리고 접근을 지원하는 활동으로 정의된다. 즉 가치 있는 디지털자원을 선별하고, 그 내용 및 기능을 보존 관리하여 장기간 접근할 수 있도록 하는 전반적인 활동을 포괄하는 개념이다.
- 장기보존과 지속적 가치가 디지털 아카이빙의 핵심 개념이다. 디지털 보존과의 차이는 보존 행위의 주체가 되는 디지털 아카이브의 구축 및 운영까지의 활동을 포함하는 것이 디지털 아카이빙으로 디지털 보존보다는 더 넓은 의미로 사용되는 경향이 있다(기록학 용어 사전 2008).
- 아카이빙과 큐레이션이 다르듯이 디지털 아카이빙과 디지털 큐레이션은 다르다. 디지털 아카이빙은 디지털 아카이브와 콘텐츠의 관리를 그 중점이 있다면 디지털 큐레이션은 생애주기에 맞추어 선별부터 제공까지의 모든 활동을 뜻한다. 또한 디지털 아카이빙을 백업, 체크섬, 저장과 같은 기술적 관리행위로 정의하는 것도 적절하지 않다. 왜냐하면 정보전문가가 디지털 아카이빙을 위하여 전문적 의사결정을 수행하고 이에 따른 복잡하고 다양한 업무를 수행하는 것을 간과했기 때문이다(Cunningham 2008).
- 디지털 아카이빙은 디지털문서와 문서에 담겨진 콘텐츠를 보존하는 활동으로 다음과 같은 세 가지 조건을 충족시켜야 한다: 1) 디지털문서가 물리적으로 그리고 내용적으로 보존되어야 한다; 2) 디지털문서는 장기적으로(최소 30년 이상) 보존되어야 한다; 그리고 3) 항상 접근이 가능해야 하며 이해 가능한 상태로 디지털문서가 유지되어야 한다(CINES: Centre Informatique National de l'Enseignement Supérieur 2018).

-
- 한편으로 아카이빙을 데이터를 접근할 수 있도록 적절하게 선정·저장하고, 장기간 데이터의 안전성과 진본성을 포함하여 논리적·물리적 무결성이 보장되도록 하는 큐레이션 활동으로 보는 측면도 있다(이재선 2016).

□ 디지털 큐레이션

- 디지털 큐레이션은 디지털자원을 수집, 보존, 아카이빙, 제공하는 것을 지칭한다. 넓게 보면 현재와 미래에 이용될 신뢰할 수 있는 디지털정보를 유지하고, 가치를 부여하는 것이라고 해석할 수 있다. 디지털 큐레이션은 데이터 큐레이션, 디지털 아카이빙, 디지털 보존을 포함하는 디지털 자산 관리 전 과정으로 가장 넓은 의미의 디지털 보존 활동이라 할 수 있다. 다음은 기관 및 연구자들이 디지털 큐레이션에 대하여 정의한 것이다.
- 디지털 큐레이션의 핵심적 개념은 바로 재현가능성(reproducibility)과 재사용(re-use)이다. 즉 디지털 데이터의 지속가능성을 유지하는 보존 활동과 재사용을 위한 데이터의 안정적 관리가 중점적 활동이다(Digital Curation Center 2014).
- 디지털 환경을 전제로 보다 확장된 의미에서 즉, 도서관뿐만 아니라 리포지터리, 박물관, 기록관 등 여러 문화 유산기관들이 다양한 유형과 형식의 디지털정보에 대하여 생애주기 동안 접근 및 재사용을 보장하기 위한 일련의 활동이 디지털 큐레이션이다(김관준 2015).
- 디지털 데이터의 장기보존과 지속적인 이용보장을 위하여 수행되는 모든 활동을 지칭한다(Sabharwal 2015).
- 디지털 큐레이션은 신뢰성 있는 디지털자원의 유지관리와 보존기술 및 전략을 강조하는 ‘디지털 보존’과 가치 있는 디지털자원의 선별 및 장기적 접근을 가능하게 하는 전반적인 활동은 물론이고 보존행위 주체가 되는 기관의 운영까지 포함하는 ‘디지털 아카이빙’을 포괄하는 개념인 것이다(서은경 2017).
- 디지털 큐레이션은 장·단기적인 관점에서 다양한 이점을 갖고 있다. 먼저 단기적으로는 디지털 큐레이션을 통해 디지털자원의 진본성을 보장함으로써 법적으로 유

효한 공식적 자원으로 보존이 가능하고, 재활용(re-use) 할 수 있게 도와준다. 장기적으로는 디지털 큐레이션을 통해 데이터를 보존하고, 손실이나 퇴화와 같은 위험으로부터 디지털자료를 보호할 수 있고, 재정적 변화나 제도적 변화에도 불구하고 신뢰할 수 있는 디지털자원에 대해서 지속적인 접근이 가능하도록 해준다(한나은, 김성희 2014).

- 디지털 큐레이션에 대한 정의가 다양한 것은 디지털 큐레이션의 서비스나 활동에 대한 정의가 처음부터 명확하게 규정된 것이 아니라 여러 기관에서 큐레이션 업무를 수행하면서 또는 이를 기반으로 하는 시스템을 구현하면서 이에 필요한 목표나 서비스 또는 구체적인 업무들을 점점 추가되고, 이를 기반으로 디지털 큐레이션의 정의와 범위가 제시되기 때문이다. 다음의 [표 2-1]은 디지털 큐레이션의 실제업무 프로세스를 정리한 것으로 이러한 업무들에 대한 이해는 실제 디지털 큐레이션 업무의 정의와 범위에 대한 이해를 도와줄 수 있을 것이다.

[표 2-1] 디지털 큐레이션 프로세스

프로세스	개요	세부업무
계획	디지털 큐레이션에 대한 전체적인 업무를 파악하고 이에 대한 정책 수립 업무를 수립한다. 또한 업무수행을 위한 조직 정비, 법적 및 재정적 지원이 이루어지고 있는지를 수행한다.	<ul style="list-style-type: none"> • 장기적 계획과 정책 수립 • 관련 조직 정비 • 관련 법적 문제 해결 • 지속적인 재정 지원 확보 • 타부서(기증자)와의 협력
생산/수집	데이터의 디지털화, 익명처리, 특수문자, 표준코드 변환 등의 수정을 통한 재생산, 생산된 데이터의 수집전략 및 수집기능에 대한 업무를 수행한다.	<ul style="list-style-type: none"> • 디지털화, 컨버전 • 데이터 재생산(클리닝, 익명처리...) • 수집전략의 수립 • 수집대상 범위 • 자동수집 기능
선정/평가	디지털 콘텐츠의 선정 대상을 정하고 선정 및 평가를 수행할 때 사용되는 절차와 검증과 관련된 업무를 말하며, 선정기준을 정하기 위한 이용 및 이용자 분석 업무까지 포함된다.	<ul style="list-style-type: none"> • 선정 및 평가 기준 • 선정/평가 절차 수립 • 재평가 절차 수립 • 검증기능(매체/파일의 품질, 표준...) • 이용자 및 미래이용 분석
입수	기록관리 표준에 준수하는 이관업무와 이관관리에 필요한 서류업무도 포함한다. 또한 보존 및 검색을 위한 메타데이터 기술, 분류, 색인 업무를 수행한다.	<ul style="list-style-type: none"> • 기술적 이관 프로세스 • 관리적 이관(서류, 인수증, 지적소유권) • 메타데이터 기술 • 주제분류 및 정보조직 • 입수를 위한 tools 확보

프로세스	개요	세부업무
보존	영구적/준영구적 디지털 데이터의 보존을 위하여 보존 정책, 보존전략, 위험 관리, 인증절차, 무결성 검사수행 등과 같은 업무를 수행한다.	<ul style="list-style-type: none"> • 보존 및 처분 정책 • 보존전략 및 활동 • 보존 메타데이터 기술 • 파일/미디어 모니터링 • 보안 및 위험관리
소장/접근	이용자에게 디지털 콘텐츠를 제공하기 위한 소장 매커니즘과 관련된 업무와 검색을 위한 다양한 검색지원 및 접근점 제공, 플랫폼 개발 업무를 말한다.	<ul style="list-style-type: none"> • 다양한 소장방법 적용 • 데이터 백업 • 다양한 접근점 지원 • 검색지원 • 적합한 플랫폼(인터페이스, 기능)구축
이용	이용자의 디지털 콘텐츠의 활용이 가능하게 하는 다양한 서비스 업무와 적합한 이용자에게 적합한 콘텐츠를 제공하기 위한 접근권한, 공개여부, 저작권관리 업무를 말한다.	<ul style="list-style-type: none"> • 다양한 서비스 제공 • 특화된 서비스 제공 • 접근권한 및 규제 수립 • 공개기준 및 비밀보장 • 저작권 관리
관리	전반적인 디지털 큐레이션 업무에 대한 관리를 위하여 수행되는 감사 기능과 전문가 확보, 이용자 분석 및 관리 등과 같은 업무를 수행한다.	<ul style="list-style-type: none"> • 데이터 추적 감사 • 전문성 확보 • 이용통계분석, 미래 분석 • 이용약관관, 도큐멘테이션 • 이용자 그룹 관리

(〈출처〉 서은경 2017, 257)

2.1.2 디지털 보존의 필요성

□ 디지털 보존의 필요성

- 디지털 보존이라 함은 미래의 후손들이 용이하게 볼 수 있고 이해할 수 있는 방식으로 디지털자원을 보존하는 것을 말한다. 이에 따라 이러한 디지털 보존 활동이 필요한 이유는 다음과 같다.
- 첫째, 디지털 보존의 필요성은 디지털자원 또는 디지털 데이터의 특성으로 생겨진다고 볼 수 있다. 디지털자원은 인쇄자료에 비해 공간을 초월한 접근성, 배포의 용이성, 그리고 즉시성 등의 강력한 장점을 가지나, 디지털자원이 담겨진 매체와 형식이 기술의 발달에 따라 빠르게 변화하고 또 노후화되어 기존의 디지털자원이 순식

간에 망실되어 질 수 있다는 단점을 가진다.

- 둘째, 디지털 데이터가 비트로 구성됨에 따라 접속과 해독을 위한 도구 즉 하드웨어와 소프트웨어가 필요하기 때문이다. 이러한 도구는 매우 빠르게 노화되거나 쓸모없게 됨에 따라 새로운 것으로 대체하게 되는데 이 경우 콘텐츠는 새로운 소프트웨어로 접근이 불가능하게 되고 이로서 콘텐츠가 접근이 불가능하게 된다. 또한 이러한 기술변화 속에서 무활동은 궁극적으로 데이터의 손실로 귀결된다.
- 셋째, 시스템, 저장매체, 데이터 포맷, 프로그램이 계속해서 변화되고 또 개발됨에 따라 새로운 형태의 디지털자료가 계속해서 생산되고 있다는 점이다. 이로서 하나의 저장소에 다양한 버전의 파일과 콘텐츠들이 소장되어 있어서 신뢰성과 완결성을 지니도록 유지 관리하는 작업이 점점 어려워지고 있다.
- 넷째, 디지털 보존문제를 해결할 수 있는 완성된 단 한가지의 방법과 솔루션이 없다는 점이다. 기술발전과 함께 계속적으로 디지털자원의 유지를 위한 전략과 기술을 개발해야 하는 어려운 점이 있다.
- 다섯째, 디지털자원의 저장과 접근에 관한 법이 기존의 인쇄자료 보존과 관련된 법보다 훨씬 더 엄격하고 개정이 빠른 편이어서 디지털자원과 관련된 법과 규정을 지키는데 어려운 점이 있다.
- 이와 같이 디지털자원의 휘발성 및 기술의존성과 같은 심각한 약점과 디지털환경이 가지는 특성으로 인하여 디지털자원의 유지관리는 인쇄자료의 유지관리보다는 훨씬 어렵고 비용도 많이 든다. 그러나 이러한 문제점과 어려운 점을 적절히 대응하고 해결해야 현 시대에서 생산되고 활용되는 디지털자원을 영구적으로 보존할 수 있다.
- 디지털 보존은 디지털자원의 무결성을 상실하지 않고, 저장매체의 손상 및 퇴화, 기술적 퇴화의 위협으로부터 디지털정보를 안전하게 보호하여 유용한 형태로 존속하게 하는 일련의 관리활동이다. 따라서 미래 세대에 우리의 지적 유산을 계승하기 위해서는 지속적이고 체계적인 디지털 보존이 수행되어야 하고 디지털콘텐츠를 보유하고 있는 모든 기관에서는 디지털보존의 필요성과 중요성을 인지하고 해야 한다.
- 정보자원을 유지 관리하여 서비스하고 후대에까지 보존해서 전달해야 하는 도서관

과 정보서비스 기관의 측면에서 디지털 콘텐츠의 보존 즉 디지털 보존은 매우 큰 도전 과제인 것이다. 각국의 국가도서관은 디지털 보존의 필요성과 그 중요성을 강조하면서 디지털 보존 방법과 전략을 수립되었고 장기적으로 그리고 거시적으로 실행에 옮기고 있다.

- 분명한 것은 지속적으로 많은 양의 디지털자원이 납본되어 관리될 것이며 이러한 디지털자원은 점점 노화된다는 점이다. 납본되어 보존되고 있는 디지털자원에 대한 무활동, 소극적 활동, 비체계적 활동은 소중한 국가의 자산을 손실시키는 행동이다. 따라서 디지털자원을 장기적으로 유지하고 이용자가 항상 그리고 지속적으로 디지털자원을 이용할 수 있게 해주기 위해서는 일련의 디지털 보존 활동이 기본적으로 수행되어야 할 것이다.

□ 디지털 보존전략의 필요성

- 디지털 보존이 언급되기 시작한 2000년 초반에는 디지털 보존을 수행하기 위한 기술적인 면을 중심으로 한 연구와 프로젝트가 많이 진행되었지만, 그로부터 약 15여년이 지난 현재에 이르러서는 디지털 보존을 시스템적으로 수행할 수 있는 신뢰성 있는 디지털 리포지터리가 구축되어 운영되고 있는 실정이다. 특히 리포지터리에 대한 국제적 가이드라인이 제시되고 있고, 기능적 절차나 기술적 처리과정에 대한 인증이 제도적으로 안착될 정도로 기술적인 면에서의 디지털 보존은 시스템화되었고 보편화되었다. 이제 디지털 보존은 단기적이 아닌 장기적인 면에서 또, 기술적인 면이 아닌 관리적 측면에서 바라보아야 한다.
- 뉴질랜드 국가도서관은 장기적인 디지털 보존이 포함해야 하는 활동을 단순한 기술적 절차로 국한시키지 않고 있다. 그들은 디지털 보존을 수행하기 위해서는 업무적 정책 및 절차 수립, 인적/물리적/금전적 자원 확보, 성능평가(효율성과 효과성, 내부적 교육, 디지털자원 생산자 관리, 업무적/기술적 지원, 소통과 같은 업무가 필요하다고 하였다. 특히 그들은 이와 같은 활동이 유기적으로 이루어질 때 장기적인 디지털 보존이 성공할 수 있다는 점을 강조하였다.
- 이와 같이 디지털 보존은 하나의 부서의 업무로, 또한 한 번에 이루어질 수 없다.

또한 디지털 보존은 체계적으로, 지속적으로 그리고 다각적으로 수행되어야 하고 지원되어야 한다는 것이다. 이러한 점들이 디지털 보존을 어렵게 하고 이로 인하여 디지털 보존의 중요성과 디지털 보존 전략 수립의 필요성이 강조되고 있다.

- 성공적으로 디지털 보존전략을 수립하기 위하여 도서관은 먼저 디지털 보존 실무 이외에 지원 조직 및 자원 면을 고려해야 한다. 다음은 디지털 보존전략을 수립하기 전에 도서관이 분석하고 해결방안을 도출해야 하는 이슈들이다.
- 실무처리 이슈(treats issues)
 - 디지털자료의 보존 방법
 - 신뢰성 있는 장기보존 방법
 - 디지털자료의 문맥적 그리고 매체 보존
 - 많은 양의 보존 처리와 보존전략 수행
- 지원 조직 이슈(organizational issues)
 - 지원 조직 체계의 적합성
 - 역할과 책임 할당
 - 내부, 외부, 협력기관간의 협조
 - 규정과 법 관리
- 지원 자원 이슈(resource issues)
 - 예산과 비용
 - 직원과 능력
 - 디지털 리포지터리 시스템
 - 고성능 컴퓨팅
 - 디지털 보존 실험실

2.1.3 디지털 보존의 범위

□ 디지털 보존 업무의 범위

- Corrado와 Sandy(2017)는 디지털 보존을 정보전문가의 업무이며 임무임을 강조하였다. 특히 디지털 보존 업무는 새로운 아이디어나 새로운 기술이라기보다는 새로운 업무절차나 규정이므로 정보전문가가 전적으로 수행해야 하는 활동이라고 하였다.
- 디지털 보존은 장기보존(long term preservation)을 염두에 두고 수행하는 업무로 이를 위해서는 두 가지 주요 업무(activities)가 수행되어야 한다.
 - 비트 보존(Bit preservation): 무결성 유지를 위한 비트 스트림 보존
 - 논리적 보존(Logical preservation): 비트 스트림 해석 능력의 보존
- 이러한 장기 디지털 보존 활동을 수행하기 위해서는 정보기관 즉 대표적으로 도서관은 필요성에 대한 인식 및 정책수립, 관리, 디지털자원 수집, 보존실무, 지속적인 접근환경 제공과 같은 업무를 수행해야 한다.
- ‘필요성에 대한 인식 및 정책수립’은 디지털 보존의 중요도 및 부재 시에 발생하는 문제점에 대한 전사적으로 인식해야 하고 이에 대한 정책수립을 뜻한다. 특히 필요성은 디지털 보존을 원하는 수준으로 이끌어낼 수 있는 중요한 요인이므로 디지털 보존 정책에서 필수적으로 정의되어야 한다.
- ‘관리’는 디지털 보존을 성공적으로 이끌어내는 핵심 업무로 기획, 조직, 자원배정 및 기술적 처리 등과 같은 사안에 대해 의사결정 또는 지원이 이루어지는 것을 말한다.
- ‘디지털자원의 수집’은 디지털 보존 대상을 정하고 그 유형과 특징을 인식하는 작업을 말한다. 디지털자원은 실제 다양한 디지털 개체를 뜻하는 것으로 이미지, 데이터 세트, 오디오 레코딩, 비디오, 3D 파일, 텍스트 기반 콘텐츠 등이 있다. 이와 같이 디지털자원은 상당히 다른 유형으로 구성되므로 디지털 보존과정은 그 유형에 따라 다르게 진행되어야 한다.
- ‘실무활동’은 체계적으로 계획된 절차와 과정아래 보존전략 기법들이 이루어지는 것을 뜻하는 것으로 시스템에서 이루어지는 활동을 말한다.

- 마지막으로 ‘지속적 접근환경 제공’은 디지털 보존의 궁극적 목적이며 핵심 활동이다. 지속적으로 디지털자원에 접근할 수 있도록 한다는 것은 바로 장기적으로 보존된 콘텐츠에 대해 검색과 이용이 가능하다는 것을 의미한다고 할 수 있다.
- 다음은 Brown(2013)이 정리한 디지털 보존을 위하여 수행해야 될 업무의 범위이다.

[표 2-2] 디지털 보존 프로세스 범위

디지털 보존 프로세스		
프로세스		업무활동
1	조직의 실행가능성	<ul style="list-style-type: none"> • 담당 직원 및 업무시간 배정 • 적절한 예산 배정 • 인력개발 요구사항 파악 및 자금지원
2	이해관계자의 참여	<ul style="list-style-type: none"> • 주요 이해관계자 식별 • 식별된 의사소통 목표 및 목적
3	법적 근거	<ul style="list-style-type: none"> • 저작권, 주요 법적 권리와 책임
4	정책 프레임워크	<ul style="list-style-type: none"> • 디지털 보존 정책 성문화
5	획득 및 수집	<ul style="list-style-type: none"> • 수집, 납본되는 디지털콘텐츠 유형을 규명한 정책 수립 • 수집, 납본 절차 및 이와 관련된 업무에 대한 문서화 • 시스템에 의한 자동 입수 도구 개발
6	비트스트림 보존	<ul style="list-style-type: none"> • 네트워크 드라이브, 워크스테이션 또는 이동식 미디어의 전용 저장 공간 • 각 객체에 대해 최소 3개의 copy본 생성 및 유지 • 기본적 무결성 점검 수행 • 바이러스 검사 수행 • 기존 액세스 제어 및 보안 프로세스 적용
7	논리적 보존	<ul style="list-style-type: none"> • 최소한의 포맷을 인지하게 하는 기본적인 특징 확인 • 임기응변용 보존계획 수립 • 임기응변용 보존활동 절차 수립 • 디지털 객체에 대한 다각적 관리능력 확보
8	메타데이터 관리	<ul style="list-style-type: none"> • 문서화된 필수 메타데이터 확립 • 데이터에 대한 메타데이터 및 일관성 있는 접근환경 구축 • 다양한 형태로 메타데이터를 생성하고 사용 • 데이터와 메타데이터 간 지속적으로 연결되도록 관리 • 모든 디지털 객체에 대한 영구 고유 식별자 할당 및 유지 관리
9	유통	<ul style="list-style-type: none"> • 다양한 검색도구 개발 및 이용자 맞춤 검색인터페이스 제공 • 데이터와 메타데이터 보기(view) 기능 및 다운로드 기능 제공
10	기반시설	<ul style="list-style-type: none"> • 충분한 저장 용량 확보 • 향후 스토리지 요건을 충족시키는 계획 수립 • 디지털 보존의 목표에 적합한 IT 시스템 선정

-
- 디지털 보존과 접근과 관련된 블루리본 연구팀(BRTF: Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008)은 디지털 장기보존은 심리적인 면에서부터 예산, 조직, 이해관계자와의 협업까지 모든 면에서 관여가 되는 거대한 프로젝트임을 강조하고 디지털 보존 업무를 수행하는데 도움을 주는 10가지 팁을 다음과 같이 제시하였다.

- ① 디지털자원의 생성에서부터 소멸까지 전 과정에 걸친 세세한 관리와 보존절차에 관한 계획 수립
- ② 하드웨어/소프트웨어, 인프라, 시간, 인적자원 등의 디지털 보존비용을 정확하게 파악
- ③ 데이터 표준화 및 보존용 표준 메타데이터를 구축하고 그 내용과 포맷을 연결해 지속적으로 사용할 수 있도록 정비
- ④ 각기 다른 매체에 중요한 데이터의 복본 저장
- ⑤ 새로운 스토리지 미디어로 디지털 데이터를 이전할 계획을 미리 수립. 새로운 스토리지와 소프트웨어 기술, 파일 포맷 이전, 소요 시간 등에 관한 사항에 대해 미리 인지
- ⑥ 향후 데이터 관리에 대한 사전적 계획 수립
- ⑦ 보존전략 및 아카이빙 방식이 결정되면 이 때 필요한 신뢰 수준도 결정
- ⑧ 데이터 보존과 접근 정도는 이용자의 요구/사용량에 적합하도록 결정
- ⑨ 데이터 무결성 유지를 위한 보안 프로세스 구축
- ⑩ 디지털 보존과 관련된 법, 규제, 표준에 대한 최신 동향 파악

□ 디지털 보존 대상의 범위

1) 디지털자원의 유형

- 디지털자원이란 정보를 디지털로 인코드하고 컴퓨터를 통해 접속, 인지하도록 하는 방식으로 제작하여, 일반 공중의 이용에 제공할 목적으로 배포된 자료를 말한다. 디지털자원이 전통적인 종이기반 인쇄자료를 완전히 대체하지 않을 것은 분명하지만, 도서관자료로서의 중요성이 점점 높아지고 있는 것은 사실이다. 디지털자원의 종류는 매우 다양하지만 대체로 오프라인 자료와 온라인 자료로 나눌 수 있다.

-
- 오프라인 자료는 손으로 잡을 수 있는 물리적 형태가 있는 매체에 고정된 것으로 패키지형 전자출판물이라고도 한다. 오프라인 자료는 복본을 제작하여 배포한다는 점에서 인쇄자료와 유사하다. CD-ROM, 컴퓨터디스켓, DVD 등이 대표적이다.
 - 온라인 자료는 인터넷, PC통신 등 네트워크를 통해 생산되고 이용되는 것이므로 손으로 잡을 수 있는 물리적 형태가 없다. 온라인 자료는 호스트컴퓨터(서버)에 저장되어 있고 이용자가 통신네트워크를 통해서 접속함으로써 배포되므로 네트워크형 전자출판물이라고도 한다. 인터넷 상에서 제공되는 각종 웹정보, 데이터베이스 등을 예로 들 수 있다.
 - 온라인 자료는 다시 정적(static) 온라인 자료와 동적(dynamic) 온라인 자료로 나눌 수 있다. 전자는 파일의 내용이 일정하게 고정된 것인데 비해 후자는 이용자와의 상호작용 속에 지속적으로 변화하는 특징을 갖는다. 디지털자료는 처음부터 디지털형태로만 만들어진 자료와 다른 포맷으로 만들어져있던 자료를 디지털 변환한 자료로 구분할 수도 있다(서혜란 2003).
 - 도서관에서 다루는 디지털자원에는 텍스트 형태의 콘텐츠, 스틸 이미지, 사운드레코딩, 동영상, 디지털자료의 편집물(compilations), 멀티미디어자료(multimedia works), 웹사이트, 데이터세트, 지리정보 등이 있다.

2) 콘텐츠 형태로 본 디지털자원의 영역

- 미국 의회도서관의 NDIIPP(National Digital Information Infrastructure Preservation Program) 프로젝트 연구팀은 다음과 같이 디지털자원을 4가지 영역으로 나누었다.
- 영역 1: 지리 영역(Geospatial data)
 - 국가 및 행정기관 자원(GPS 데이터베이스, 재난대비데이터 등)
 - 과학 자원(지리통계 데이터, 해양정보 데이터, 대기정보 데이터 등)
- 영역 2: 음향 영상 영역(Audiovisual content)
 - 공공 자원(공영 TV 방송, 공영 라디오 방송 등)
 - 개인 자원(상업적 방송, 상업적 음반, 공연 동영상 등)

- 영역 3: 텍스트와 이미지 영역(digital text and Images)

- 개인 및 회사 기록(의료기록, 행정 기록 등)
- 데이터베이스(사회과학연구데이터베이스, 여론데이터베이스 등)
- 정부기록(법원 기록, 주 및 지역 기록 등)
- 전자정보원(전자저널, 신문, 전자책, 지도 등)
- 디지털화된 인쇄기록(디지털화된 지도, 디지털화된 책 등)

- 영역 4: 웹 콘텐츠 영역(Web-based Content)

- 소셜 소프트웨어 사이트(블로그, 비디오 블로그 사이트 등)
- 정부 사이트(정부기관 사이트, 정부정책 사이트 등)
- 최신이슈 사이트(개인 웹사이트, 이페메라(Ephemera) 정보사이트 등)
- 공공데이터베이스(유튜브, 플리커, 위키피디아 등)
- 기관 사이트(회사 및 기관 사이트, 해외투자 및 아웃소싱 사이트 등)

3) 도서관법에서 본 디지털자원 범위

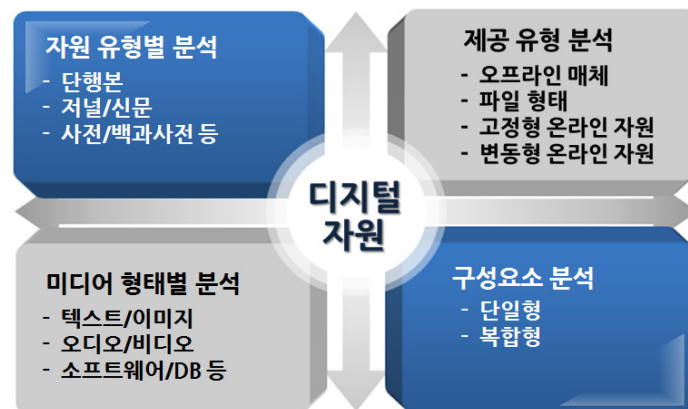
- 디지털자료 납본은 1990년대 이후 전자출판의 다변화와 웹을 기반으로 한 정보서비스가 활성화 되면서 인쇄자료를 중심으로 하는 납본체도로 국가차원에서 문헌수집에 한계가 있음을 인식하고 납본의 대상을 디지털 파일형태의 자료로 확대하는 것으로 시작되었다.
- 국내에서도 2016년 도서관법 납본조항을 아래와 같이 수정하면서 납본의 대상을 디지털 파일형태의 자료로 확장하였다. 온라인 납본과 관련된 조항은 다음과 같다.

제20조(도서관자료의 납본)

- ① 누구든지 도서관자료(온라인 자료를 제외한다. 다만, 온라인 자료 중 제21조에 따라 국제표준자료번호를 부여받은 온라인 자료를 포함한다. 이하 이 조에서 같다)를 발행 또는 제작한 경우 그 발행일 또는 제작일로부터 30일 이내에 그 도서관자료를 국립중앙도서관에 납본하여야 한다. 수정증보판인 경우에도 또한 같다. <개정 2009.3.25., 2016.2.3.>
- ② 국가, 지방자치단체 및 그 밖에 대통령령으로 정하는 공공기관이 제1항에 따라 도서관자료를 국립중앙도서관에 납본하는 경우에는 대통령령으로 정하는 바에 따라 디지털 파일 형태로도 납본하여야 한다. <신설 2016.2.3.>

2.2 디지털자원의 유형 및 특성

- 도서관이 수집, 보존하는 디지털자원은 광범위하다 할 수 있다. 파일 형식으로 보면, 이미지용 TIFF, 텍스트에 대하여 PDF와 XML, 오디오에 대하여 WAV, 웹사이트에 대하여 HTML 등으로 구분할 수 있다. 디지털자원의 자료 유형으로 보면, 전자책, 전자저널, 오디오북, 공공간행물 등, 이 밖에도 제공유형으로 구분하면 데이터베이스 구축기관에서 제공한 파일, 장애인을 위한 데이지 파일 등이 있다. 이처럼 다양한 디지털자원의 유형, 특성을 다각적으로 조사 분석하기 위해 패킷 분석을 적용하고자 한다.
- 패킷 분석이란 복합 주제에 대한 접근 시 하나의 특정 관점에서 시작하여 계층 구조로 세분화하는 것이 아니라 다양한 시각 및 관점 즉 다면적으로 접근하는 방식의 분석법을 말한다. 디지털자원을 다음과 같은 4가지 패킷, 즉 자원 유형, 제공유형, 미디어 형태, 구성요소 측면에서 디지털 보존에서 고려할 사항들을 분석하였다.



2.2.1 자원 유형별 분석

- 디지털자원의 자원 유형별 분류는 디지털자원의 출판목적과 이용형태에 따라 전자책, 전자저널, 오디오북, 공공간행물, 학위논문, 동영상, 웹자료, 학술지, 음원 등으

로 구분할 수 있다

- 국립중앙도서관의 2017년도 연보에 따르면, 납본을 통해 전자책 16,785책, 오디오북 3,013건과 전자저널 5,190건 등을 수집하는 한편, 공공간행물 디지털파일 11,529건 및 학위논문 디지털파일 34,679건을 납본 수집하였다. 또한 오디오북 16,684건뿐만 아니라 동영상 5,246건 등의 온라인 자료를 선정하여 수집하였다. 자체수집을 통해 보존가치는 높으나 휘발성이 높아 소멸하기 쉬운 공개 웹자료, 공개된 학술연구자료 등 90,238건을 수집하였다(국립중앙도서관 2017).

[표 2-3] 국립중앙도서관 온라인 자료 수집 현황

(2017. 12. 31 기준/단위: 건)

자료유형		파일건수
납본	전자책	16,785
	전자저널	5,190
	오디오북	3,013
	공공간행물	11,529
	학위논문	34,679
구입	전자책	7
	오디오북	16,684
	동영상	5,246
자체수집	웹자료	90,238
기증	학술지	1,285,787
	음원	19,035
	공공간행물	43,694
계		1,531,887

- 전자책은 한국전자출판협회에 따르면 “도서로 간행되거나 도서로 간행될 수 있는 저작물의 내용을 디지털 데이터로 CD-ROM, DVD 등의 전자도서 기록매체, 또는 저장장치에 수록하고, 유무선 정보통신망을 경유하여 컴퓨터 또는 휴대용 단말기 등을 이용해 그 내용을 읽고, 보고, 들을 수 있는 것을 말한다”라고 정의하고 있다. 일반적으로 전자책은 단행본 형태로 제작된 도서를 전자형태로 변환하거나, 전자형태로 제작된 단행본 형태의 자료를 의미한다. 형식에 따라 PDF, EPUB, XML 등으로 나눌 수 있다. 전자책은 휴대폰, PMP, PDA 등의 휴대기기나 컴퓨터로 볼 수 있는

특수한 포맷의 파일이다. 일반적으로 전자책이라고 할 때는 텍스트 파일과 같은 범용 파일 포맷이 아니라 저작권 보호를 위해 DRM(Digital Rights Management) 기능을 탑재할 수 있는 특수한 포맷을 가진 파일을 말한다.

- 전자저널의 경우 중요한 디지털 출판물의 형태로 과학, 학술 연구 집단에 높은 커뮤니케이션 도구로 장기보존의 주요 대상이다. 학술적 목적을 위해 제작된 저널과 논문이 전자적으로 제공되는 것을 의미한다. 대부분의 전자저널은 HTML이나 PDF, DOC 등의 포맷으로 제공되며, MP3가 포함되는 경우도 있다. 초기의 저널은 ASCII로 제작되었고, 현재도 이 형태로 발행되는 것이 일부 있으며, 다음과 같이 제공 형태에 따라 전자저널을 다시 구분할 수 있다.

- 온라인으로만 제공되는 경우
- 인쇄된 저널의 온라인 버전
- 인쇄된 형태의 저널을 온라인화 하고, 온라인에만 비디오나 화상 등의 추가 정보를 제공하는 경우

- 장기적인 보존 관점에서 전자저널은 디지털 객체로서의 기술적 표준화와 함께 연구 데이터 이전, 저작권 문제 등이 보존의 주요 고려사항이라 할 수 있다. 학술지의 경우 학술정보원의 주요 자원이며 연구 활동과 관련하여 존재하는 연구데이터의 상황, 원 데이터와의 연계, 데이터 마이그레이션 문제 등이 디지털 보존 시 고려될 수 있을 것이다.

2.2.2 제공 유형별 분석

- 디지털자원은 기증, 납본, 수집 등 자원의 제공 유형에 따라 크게 오프라인(저장)매체, 파일 형태, 고정형 및 변동형 온라인 자료로 구분할 수 있다.
- 미국 의회도서관에서는 디지털자원을 “컴퓨터를 사용하여 입력된 접근 가능한 자료”로 정의하고 그 자원을 보유(acquire)자원, 수집(collect)자원, 링크(link) 자원의 3가지 형태로 구별하여 정의하고 있다. 보유 자원은 도서관이 실제로 소장하거나 라이선스(license) 혹은 동의를 통해 접근을 제공하는 자원으로, 접근만 허용할 뿐 도서관 소유의 자원이 아닐 수도 있다. 수집 자원은 도서관이 만들어낸 자원을

포함해서 도서관이 소유권을 지속적으로 가지고 있는 자원이다. 물론 도서관이 소유권을 가지고 있지만 도서관이 아닌 다른 곳에 소장되어 있을 수도 있다. 링크 자원이란 소장여부나 소유권 여부와 관계없이 각종 자원에 대한 포인터(도서관의 인터넷자원, 도서관 서지 레코드, 비도서관 전자자원)를 가지고 있는 것을 의미한다.

- 국립중앙도서관의 디지털자원은 납본, 기증, 자체수집의 방법으로 구축된다. 온라인 자료의 납본의 경우, 해당 자료(보존용 및 열람용 파일)와 서지정보의 디지털 파일을 국립중앙도서관 서지정보유통지원시스템(<http://seoji.nl.go.kr>)으로 전송하거나, 해당 파일과 서지정보의 디지털 파일을 저장매체에 저장하여 국립중앙도서관으로 송부하는 방법, 국립중앙도서관에 해당 자료와 서지정보의 인터넷상 위치를 통지하고, 국립중앙도서관이 이에 접근하여 수집할 수 있도록 조치하는 방법이 있다.
- 장기보존의 관점에서 오프라인 매체에 저장된 자원의 경우, 수집, 관리되는 디지털 자원은 저장매체마다 서로 다른 저장 조건과 보존 요구사항들을 지니고 있고 컴퓨터 또는 전자장치를 매개체로 해서 직접, 간접적으로 접근이 가능한 자원이므로 관련 소프트웨어와 하드웨어를 포함한 종합적 보존이 필요하다. 저장매체들은 저장 능력과 장단점이 다르며 보존과 접근에 있어 다양한 수준의 안전성을 지닌다. 물리적인 저장매체에 대한 변환 작업은 매우 단순하고 쉽지만 대부분의 전자적인 매체들이 접근을 제공하는 하드웨어와 소프트웨어의 빠른 기술적 퇴화에 따라 매우 빨리 사용불가능하게 된다. 따라서 저장매체에 대한 적절한 선택은 모든 보존전략을 위한 아주 기초적이면서도 필수적이다(정영미, 윤화목, 김정택 2010).
- 파일 형태로 수집, 관리되는 디지털자원은 장기적 보존을 위해서는 애플리케이션의 보관과 함께 생성되는 파일 형식에 대한 보존이 필요하다. 파일 포맷의 선택 범위는 매우 다양하며 그 기준 또한 다양할 수 있다. 파일 포맷 선택은 수집 단계에서부터 결정되며, 수집한 자료가 포맷이 맞지 않는 경우에는 정책에서 선택된 파일 포맷으로 자료들을 변환시켜 저장할 필요가 있다. 디지털 콘텐츠의 휘발성이나 기술 의존성 등과 같은 장기적인 보존 관점에서의 주요 문제점들이 파일 포맷과 관련성이 있으므로, 보다 안정적이고 효율적인 보존을 위해서는 파일 포맷의 선택이 매우 중요하다(정영미, 윤화목, 김정택 2010).
- Digital Preservation Coalition은 파일 포맷의 표준을 선택함과 동시에 디지털 콘텐

츠 제작이나 수집 시 장기적인 관점에서 파일 포맷을 선택할 때 공개되고 소유권이 없는 파일 포맷을 사용하고 자원에 대한 문서화와 표준과 일치하는 메타데이터 제공, 온라인 디지털자원에 대한 영구적인 식별자 할당과 같은 요소들을 포함해야 한다고 제시하였다(DPC 2008).

- 링크를 통해 수집, 관리되는 디지털자원은 주로 웹 정보자원들로 고정형과 변동형으로 구분할 수 있다. 특히 변동형 디지털자원은 언제든지 소멸되거나 잠재적인 손실 가능성을 가지고 있으므로 장기보존의 주요 이슈인 자원 무결성과 진본성 확보가 가장 중요하다.

2.2.3 미디어 형태별 분석

□ 콘텐츠 유형

- 미디어 형태의 구분은 정보기술의 발전에 따라 기존의 인쇄 매체 뿐 아니라, 새로운 유형의 디지털 매체가 계속적으로 개발되어 유통됨에 따라 물리적으로 표현되는 자원의 특성을 기준으로 구분할 수 있다. 디지털자원은 인코딩 포맷 형식에 따라 텍스트, 동영상, 사운드, 이미지, 소프트웨어, 데이터세트, 웹사이트, 컬렉션 등으로 콘텐츠 유형을 구분할 수 있다.
- MODS(Metadata Object Description Scheme)은 디지털자원의 콘텐츠 유형을 인코딩 포맷 형식에 따라 텍스트, 동영상, 사운드, 이미지, 소프트웨어, 데이터세트, 웹사이트, 컬렉션으로 구분한다. 각 유형별 콘텐츠에 대한 정의 및 파일유형(인코딩 포맷)은 다음과 같다.
 - 텍스트(text)
 - 지도(cartographic)
 - 악보(notated music)
 - 녹음자료(sound recording)
 - 이미지(still image)
 - 영상이미지(moving image)
 - 입체자료(three dimensional object)

-
- 소프트웨어/멀티미디어(software/multimedia)
 - 혼합자료(mixed material)

□ 콘텐츠 유형별 특성

- 디지털 보존 콘텐츠의 주요 파일유형(인코딩 포맷)에 대한 특성을 정리하면 다음과 같다.
- 텍스트 및 다른 문서 파일:
 - 도서관 디지털자원의 가장 보편적 형태로, 전자책, 전자저널 등의 자료 유형의 가장 기본적인 형태임
 - PDF, XML, EPUB, HWP, DOC, PPT 등 텍스트 편집기를 통해 생성된 파일 유형을 따름
 - 각 파일 별 다양한 DRM 유형을 하나로 통일하여 통합적으로 관리
 - PDF로 변경하여 보존하도록 하되, 포맷 변경으로 데이터 유실이 우려되는 것은 원본으로 보존 수집된 문서의 전문(Full-text)을 대상으로 검색할 수 있는 시스템을 구축하고 장시간이 지난 이후에도 텍스트 문서를 확인할 수 있도록 리더(Reader)등을 함께 보존해야 함
- 스틸이미지:
 - 디지털 사진, 컴퓨터 그래픽, 애니메이션, 서체 등의 디지털 콘텐츠로 이미지 해상도와 압축 방식에 따라 JPEG, TIFF, GIF 등 다양한 포맷이 존재
 - 일반 이미지에 가장 많이 사용되는 포맷은 래스터 방식의 BMP, JPEG, GIF, PNG, TIFF이며, 벡터 방식의 이미지는 그래프, 네트워크, 도형 등을 표현하는 경우가 대부분임
 - Raw 포맷은 이미지 센서(sensor)에 포착된 그대로의 상태를 의미함
- 오디오 파일:
 - 오디오 북, 음악 자료, 음원 등의 디지털 콘텐츠에 해당되며 인코딩 방식에 따라 WAV, MP3 등 다양한 포맷이 존재

-
- 가장 다양하게 사용되는 포맷은 MP3로 고음질 이외의 대부분의 환경에서 사용됨
 - 음성 데이터는 파일 크기를 줄이기 위해서 일반적으로 압축 알고리즘을 적용하는데, 손실 압축과 비교했을 때 무손실 압축은 상대적으로 음질이 뛰어나고 파일 크기가 크다는 특징을 가짐
 - 음악 CD 원본은 무손실, 무압축에 해당하고 음악 MP3는 손실, 압축에 해당됨. 범용성, 호환성, 재가공성을 고려하여 적합한 보존 파일 포맷을 선택하는 것이 중요함
- 동영상(무빙이미지):
 - 영화, 애니메이션, 비디오 및 공연물 등의 디지털 콘텐츠에 해당됨
 - 일반적으로 재생 플레이어(미디어플레이어 등)에서 지원하는 포맷에 따라 AVI, MPEG 등 다양한 포맷이 존재
 - 동영상 파일은 기본적으로 컨테이너 파일 포맷으로, 영상, 음성, 자막 등을 정해진 알고리즘으로 묶어 파일 생성하며 영상, 음성 등은 각기 다른 코덱으로 압축되는 형태임
 - 영상 파일은 파일교환, 웹 브라우저 내 플레이, 재생플레이어를 통한 스트리밍 서비스로 구분됨
 - 각 상황에 따라 사용되는 파일 포맷은 서로 상이하며 파일 전송을 통한 배포 방식에는 AVI, MKV, MOV, MPEG 등이 주로 사용됨
 - 웹 페이지에 포함되어 재생되는 경우는 플래시(Flash: .flv) 형식이 주로 사용됨
 - 대화용 멀티미디어 응용프로그램
 - 주로 온라인 게임, 모바일 게임, 이러닝 자료 등 디지털 콘텐츠로 제작사가 CD-ROM 또는 DVD-ROM으로 제공하는 경우가 대부분임
 - 데이터가 응용 프로그램과 통합된 것으로, 특정 응용프로그램이 있어야만 가독이 가능함
 - 데이터베이스
 - 데이터베이스 시스템 스토리지(DBS)는 주로 세 가지 구성요소로 구성됨
 - 첫째, Database: 실제 콘텐츠 데이터베이스 테이블의 튜플 또는 테이블

-
- 둘째, DBMS: 데이터베이스 관리 시스템은 데이터베이스를 관리하고 사용자가 요청 한 데이터에 대한 작업을 수행하도록 설계된 컴퓨터 프로그램
 - 셋째, Database 응용프로그램: 특정 응용 프로그램과 문제 세트(Problem set)의 데이터를 관리하기 위해 작성된 컴퓨터 소프트웨어
 - 이러한 구성요소의 배치는 다양한 데이터베이스 시스템에서 다를 수 있으나 크게 데스크톱(단일사용자)의 데이터베이스 시스템과 복합(다중 사용자)의 데이터베이스 시스템으로 구분됨
 - 데이터베이스를 직접 수집해 운영하는 경우, 데이터 백업 및 업데이트 등 데이터베이스 운영과 관련된 전반적 사항에 대한 대비가 필요함
 - 데이터베이스 간 상호운영성이 확보될 수 있도록 데이터베이스의 튜닝과 검색 시스템 구성에 유의하여야 하며, 각 데이터베이스 별로 네임스페이스를 지정해 데이터베이스 간 충돌 및 오류를 방지해야 함
- 장기보존 관점에서 디지털자원은 기술의 발달과 함께 저장 포맷 및 유형이 지속적으로 변화되고 있다. 디지털자원을 생성하는 애플리케이션 중에서 더 이상 사용하지 않는 애플리케이션이 다수 존재한다. 예를 들면, 보석글, 훈민정음, 하나워드부터 CAD, MC, MD, Lotus123 등 스프레드시트 유형의 프로그램들 그리고, 웹에디터 등 다양한 애플리케이션 프로그램들이 사용되다가 지금은 사용이 중단된 상태이다. 현재 가장 많이 사용하는 한글, 엑셀, 파워포인트 등의 애플리케이션도 다양한 버전이 존재하며, 버전 간 파일 포맷 및 속성의 호환에 문제가 발생되고 있다(국가기록원 2013).
 - 컴퓨팅 환경의 변화로 O/S 및 애플리케이션 또한 변화하고 있다. 디지털자원 원본 형식을 지원하는 애플리케이션을 보관하고 있지만, 운영환경이 변화하거나 컴퓨팅 환경이 변화할 경우 보관된 애플리케이션을 더 이상 수행할 수 없는 상황이 발생한다. 현재까지 개발된 거의 대부분의 애플리케이션은 MS의 윈도우 운영환경에서 동작하도록 개발되어 있으나, 운영 환경이 클라우드 서비스 환경으로 바뀔 경우 서버 환경에서 애플리케이션의 구동이 이루어져야 하지만 이미 사양화된 애플리케이션을 클라우드 환경에서 동작하도록 변환하는 기술적, 경제적 측면에서 어려운 경우에는 보관한 애플리케이션의 동작이 불가능하게 되어 원문을 인식 하지 못하는 상태가 발생할 것으로 예측되고 있다(국가기록원 2013).

-
- 디지털자원의 장기적 보존을 위해서는 애플리케이션의 보관과 함께 생성되는 파일 형식에 대한 보존이 필요하다. 현 시점에서 가장 중요한 것은 사라져 가는 애플리케이션의 파일 형식을 정확하게 파악하여 포맷 레지스트리를 구축하는 것이 필요하다. 디지털 보존 자원의 재현을 위한 기술 정보의 수집 및 관리 정보를 위한 요소들을 정의하고 기술정보를 이용하여 장기보존을 위한 기능을 제공할 수 있는 포맷 레지스트리가 요구된다.

2.2.4 구성요소별 분석

- 디지털자원은 구성요소별로 단일형 자원과 복합형 자원으로 구분할 수 있다. 복합형 자원은 “여러 이질적인 디지털 리소스로 구성된 객체”로 정의하고 있었으며, 이질적인 디지털 리소스에는 이미지, 비디오, 오디오, 텍스트, 하이퍼링크, 소프트웨어 프로그램 등이 해당한다(황윤영, 이규철 2013, 2).
- 복합형 자원은 복수의 디지털 객체와 메타데이터로 구성되며 메타데이터는 다양한 형태의 자원의 특성을 기술할 수 있도록 확장되어야 하며, 연속적 관리를 생애주기별 메타데이터가 연계되어야 한다. 복합형 디지털자원의 보존 메타데이터는 여러 종류의 디지털 객체로 구성된 복합형 자원의 구성을 잘 표현해 줄 수 있어야 하며, 객체마다 다른 표준으로 작성된 메타데이터를 포함시켜 줄 수도 있어야 한다. 이와 관련하여 가장 적합한 표준은 METS(Metadata Encoding and Transmission Standard)표이다. METS는 디지털 저작물에 관한 구조, 기술, 관리 메타데이터를 입력하기 위한 XML 언어 스키마로, 리포지터리 간의 디지털 저작물 교환을 표준화하기 위해 사용된다.
- 또한 디지털 컴포넌트와 관련한 파일 포맷 종류가 지속적으로 증가되며, 복합형 자원의 동적 특성이 변화하고 있다. 장기보존의 관점에서 기존의 마이그레이션 및 XML 패키지 변환 등의 방법으로는 오디오, 비디오, 자바스크립트 등 동적인 요소들의 사용이 제한되며, 본문에 첨부된 동영상이나 움직임 효과 등이 렌더링 되지 않아 생산자의 의도나 내용의 맥락을 완벽하게 보존하지 못하는 한계가 발생할 수 있다. 동적인 디지털자원의 보존을 위해서 대표적인 해외의 에뮬레이션 프로젝트인 KEEP, Planets, CAMiLEON, KB 등은 전자문서뿐만 아니라 다양한 OS와 미디

어, 응용 프로그램, 게임, 웹페이지 등을 재현 대상으로 선정하여 에뮬레이터를 개발하고 있다. 국립중앙도서관에서 수집, 관리하고 있는 디지털 보존 자원 중 동적 요소를 포함하고 있는 대상을 파악하고 기술적 구현이 가능한 수준에서 에뮬레이터를 개발하는 것이 요구된다.

2.3 디지털자원의 보존 동향

2.3.1 전자납본 사례

□ 미국

1) 전자납본 의무사항

- Digital Legal Deposit -An IPA Special Report, 24th July 2014
- 의무사항: 전자납본 자체가 의무사항은 아니며 인쇄물 제출 시 전자버전을 제출한다.
- 임시규정 채택: 2010년에 의무적 기탁을 관할하는 임시규정을 채택하였다.

2) eDeposit 프로젝트

- 디지털로 생산된 콘텐츠를 관리하는 목적: 디지털 콘텐츠 관리계획은 처음부터 디지털로 생산된 콘텐츠 관리를 목적으로 한다.
- 전자연속간행물 중심으로 시작: 미국 저작권 사무소(U.S. Copyright Office)로부터 이관된 전자연속간행물을 중심으로 시작되었다.
- eSerials 프로젝트: 2010년 5월 전자로 출판되는 저널의 출판업자들을 모아 전자저널의 판권납본 관련 절차에 대한 토론회를 개최하여 eSerials 프로젝트를 시작하였다. eDeposit 프로젝트의 첫 과제로 2011년 완료되었으나, 현재 eDeposit 프로젝트의 진행사항이 불명확하고 성공적이지 못하다는 평가로 향후 전략적 수정이 필요하다고 판단되는 상황이다.

3) Library of Congress Strategy Plan 2016-2020

- 전략계획의 실행: 21세기 정보 환경에서 효과적이고 반응적인(responsive) 도서관

의 지위를 점하기 위한 전체적(holistic) 접근으로 구성될 것이다.

- 세 번째 전략(Strategy 3): 저작권법과 공적인 목표들에 부합하고 현대화된 저작권 체계를 발전시키기 위해 U.S. Copyright Office와 협업하는 것이다.
- 실행(Actions): 도서관과 U.S. Copyright Office 간 효과적인 데이터 공유를 보장하며, 현재와 미래의 이용을 위하여 도서관의 소장 자료에 가장 중요한 copyright deposits의 원활한 이전을 보장한다.

4) Annual Report 2016

- 자료 수집: U.S. Copyright Office는 3,560만 달러의 순 가치를 가진 저작물의 사본 63만 장 이상을 도서관에 보냈다. 해당 사본의 38만 개 이상은 법의 의무납본 조항에 따라 출판사로부터 받았다.
- 전자납본: 1,674권의 도서, 32,947건의 연속간행물, 그리고 총 3,112,698건의 디지털 파일이 포함된다.

□ 캐나다

1) 전자납본 의무사항

- Digital Legal Deposit - An IPA Special Report, 24th July 2014
- 2007년에 법적인 납본이 온라인 및 디지털 출판물로 확대되었고, 도서관 업로드 플랫폼, FTP, 서버, 전자메일, 게시를 통해 등록이 가능하다. 전자납본 자료의 법적 접근성 허락에 대한 두 옵션은 다음과 같다.
- 개방형 액세스: 누구든지 웹을 통해 발행물을 보고 다운로드 할 수 있으며, 가능한 한 발행인에게 이 옵션을 선택하도록 요청한다.
- 제한된 액세스(주로 판매용 출판물과 관련됨): 오타와의 LAC 본관에 있는 일부 터미널에서 발행물을 볼 수 있으며, 이 단말기에서 파일을 인쇄, 다운로드 또는 전송할 수 없다.

2) 법적 전자납본 개요

- Digital Legal Deposit in Selected Jurisdictions, 2018
- Library and Archives Canada(LAC)의 법적 납본
 - 캐나다 도서관, 기록보관소법 및 출판물의 법적 납본에 의거하여 LAC는 캐나다의 문서를 수집하고 보존하는 연방 기관임
 - 2007년 1월 1일에 발효된 규정에 따라 법적 납본은 매체 또는 형식에 관계없이 캐나다에서 제작된 출판물에 적용되며 온라인 및 디지털 출판물이 포함됨
 - LAC Act, S.C. 2004, c. 11은 캐나다의 연방 법령이며, 합법적으로 LAC에 출판물의 법적 기탁을 하도록 함
 - 이 법은 또한 LAC에 납본되는 출판물의 등급과 요구되지 않는 (캐나다의 사서 및 아키비스트가 서면으로 요청하지 않는) 출판물의 등급을 정의함
 - 모든 캐나다 출판사는 도서관 및 캐나다 기록 보관소에 온라인 출판물의 사본을 납본해야 함
- 디지털 법적납본 요구 사항
 - 매체 또는 형식에 상관없이 캐나다에서 제작된 출판물에 적용되며 온라인 또는 디지털 출판물을 포함함
 - ‘출판물’: 여러 가지 형태로 제공되는 도서관자료. 일반 대중에게 구독 또는 기타 방법으로 제공 가능하며, 출판물은 모든 매체를 통해 입수할 수 있고 인쇄물, 온라인 상품 또는 음반을 포함하여 어떤 형태로도 사용될 수 있음
 - ‘출판인’: 캐나다에서 출판물을 제공하여 복제 할 수 있는 권한을 부여 받았거나 내용을 관리하는 사람을 의미함
- 온라인 또는 디지털 간행물의 납본을 위한 절차
 - Regulations의 제2조: 인쇄자료 이외의 매체를 사용하는 출판물과 그 내용을 만들기 위해서는 출판업자 파악
 - 사서와 아키비스트에게 출판물의 사본을 제공하기 전: 출판물의 암호화 된 데이터를 해독하고, 출판물에 대한 접근을 제한하거나, 이를 위해 고안된 보안 시스템 및 장치를 제거 또는 사용 불가능하도록 해야 함

-
- 사서와 아키비스트에게 출판물 사본 제공 시: 특별히 작성한 소프트웨어 복사본을 제공해야하며, 출판물과 함께 제공되는 매뉴얼 사본, 출판물에 액세스하는 데 필요한 기술 또는 기타 정보의 사본을 제공하고, 출판물의 제목, 작성자, 언어, 발행일, 형식, 제목 및 저작권 정보를 포함하여 발행물에 대한 설명 자료를 제공해야 함

- 지시 사항

- 디지털 출판물을 인쇄하여 전송해서는 안 되며, 디지털 형식으로 기탁해야 함
- Adobe PDF 및 EPUB 형식을 선호함
- 출판물이 디지털 형식과 물리적 형식으로 모두 존재할 경우 두 매체를 모두 배치함
- 웹(HTML, XML, PHP, ASP 등) 형식의 출판물은 같은 출판물(Adobe PDF 또는 EPUB)의 전환 가능한 형식이 있는 경우, 해당 형식을 보관해야 함
- 웹 형식으로만 출판물이 존재하는 경우, 디지털 법적 납본 부서와 계약에 대해 의논해야 함

- 디지털 발행물 제출 시 제공 옵션

- 온라인으로 제출 가능: 디지털 컬렉션의 업로드 플랫폼(한 번에 200MB 이하의 파일)
- 대용량 파일을 위한 FTP
- Email(주로 구독을 위함)
- CD-ROM 또는 기타 저장 장치(일반 우편을 통해 받으며, 해당 자료는 발행자에게 반환되지 않음)

3) Annual Report 2017-2018

- 기록 및 작성, 처분: 매년 연방 정부 기관에서는 모든 유형의 정보 기록을 작성 및 관리하며, LAC는 더 이상 사용하지 않는 기록의 처분을 위해 정부기관에 처분 권한 부여 혹은 DAs를 발급한다.
- DAs의 조건에 따라 기관은 기록을 폐기하거나 보관 가치가 있는 경우 LAC로 전송할 수 있다. LAC는 2017-2018 회계 연도가 끝날 때까지 모든 캐나다 정부기관에

서의 완전 처분 보장을 위한 3개년 계획을 개발하였다.

□ 호주

1) 전자납본 의무사항

- Digital Legal Deposit -An IPA Special Report, 24th July 2014
- 법무부의 자문서 발행: 저작권법 확대에 따라 2012년 3월 법무부에서 자문서를 발행하였다.
- 2016년 이전은 자발적 납본 기준: 2016년 이전에는 전자출판물이 전자형식으로 제공될 때 법적으로 의무사항이 적용되지 않았으며, 자발적인 납본을 기준으로 하고 있었다.
- 2016년 초 저작권법 개정: 전자출판 등에 법적납본이 확대되었다.

2) 법적 전자납본 개요

- Digital Legal Deposit in Selected Jurisdictions, 2018
- 오프라인으로 출판된 자료: 원래 출판된 물리적 매체를 통해 호주 국가 도서관에 전달되어야 한다.
- 온라인 자료: 웹 수집 소프트웨어를 통한 요청을 포함하여 요청이 있을 경우 도서관에 제공되어야 한다.
- eDeposit 서비스: 자료가 공개 웹사이트에서 제공되지 않는 경우 게시자는 도서관의 eDeposit 서비스를 사용하여 자료를 납본할 수 있다.
- PANDORA 아카이브 이용: 주립 도서관 및 유적지 기관과 협력하여 PANDORA 아카이브에 호주 웹사이트를 선택하고 수집한다.
 - 웹사이트 발행인은 웹 수집 과정의 일부로 '요청'을 받고 요청에 대한 법정 권한 통지 관련 링크가 제공됨

-
- 수확된 웹사이트는 PANDORA 플랫폼에서 공개적으로 사용할 수 있음
 - 호주 국가도서관은 모든 호주 정부 웹사이트를 수집함
 - 도서관에 기탁된 기타 전자 출판물은 상업성에 따라 대중에게 공개됨
 - 영리 출판물의 경우 도서관의 현장 사용자만 접근 가능하지만 출판물을 보거나 다운로드 할 수는 없으며, 오프라인에서 활용할 수 있지만 읽기 전용 형식으로만 제공됨
 - 비영리 출판물은 도서관의 카탈로그를 통해 자유롭게 사용할 수 있으며, 도서관이 주최하는 전국 온라인 데이터베이스 수집기인 Trove가 있음
- 국가도서관과 지역도서관의 협력: 2019년에 시작될 국가 eDeposit 플랫폼 개발을 위해 모든 주 및 지역도서관과 협력하고 있다. 국가 eDeposit 플랫폼을 통해 출판사는 전자 출판물을 한 번만 납본하고, 국가 및 주 또는 지역의 법정납본 요구 사항을 충족 할 수 있다.
 - 호주 국가도서관에 적용되는 법적납본: 1968년 Copyright Act 1968 (Cth), pt X div 3에 제시되어 있다. 2016년 이전에는 인쇄본에 한정되어 있어(NLA에 기탁될 필요가 있는 도서관자료 중 호주에서 출판된 자료) 전자자료는 해당되지 않았으며, 자유로운 이용이 가능한 온라인 출판 자료 일부만을 수집하였다.
 - 해당 관할권마다 정해진 각각의 방침 및 법에 따라 지역 도서관에 기탁되어야 하며, 현재 전체 관할구역 중 세 관할구역에서만 생산된 디지털자료의 납본을 요구하고 있다.

3) 연방차원 전자납본

- 1968년 저작권법에서는 문학작품, 극작품, 음악작품, 예술작품 또는 이러한 작품의 저작이나 판을 포함하며, 해당 법에 의거하여 저작이나 판에 저작권이 존속된다.
- 저작이나 판: 웹사이트, 웹페이지, 웹파일, 도서, 정기 간행물, 신문, 팜플렛, 음악, 지도, 계획, 차트 또는 테이블 또는 중앙 도서관장이 정하는 바에 따른다. 판: 본문이나 기타 독서 사항, 삽화, 그림, 판화, 사진 또는 시청각 요소의 추가 또는 변경 사항이 포함된다.

-
- 전자 도서관자료는 “온라인으로 사용 가능” 또는 “온라인으로 사용할 수 없음” 중 하나로 제공한다.
 - 온라인으로 사용 가능한 자료는 인터넷상에서 전달되며, 납본 요구사항이 달라진다. 오프라인에서 사용 가능한 전자자료는 호주에 있는 게시자가 물리적 형식 캐리어에 배포하고 판매할 것인지 또는 무료로 공개할지 결정한다.
 - 온라인에서 구할 수 있는 전자자료: 인터넷을 통해 또는 다른 플랫폼을 통해 판매되거나 무료로 제공되는, 호주 내 대중에게 제공되는 자료(.au 도메인 이름으로 게시된 자료 또는 도메인 이름이 호주 거주자에 의해 소유되거나 허가 된 웹사이트에 게시 된 자료, 오스트레일리아 내 웹사이트에서 온라인으로 액세스 할 수 있는 자료), 국가적 컬렉션인 경우, 호주나 호주 거주자에 의해 인터넷에 출판된 자료 등을 포함한다.
 - 오프라인 전자자료: 인쇄물과 하드 드라이브, USB 플래시 드라이브, CD, DVD 또는 기타 물리적 전자 매체에 출판된 전자 출판물을 포함하며, 발행 후 1개월 이내에 NLA에 전달한다(기존에 공개적으로 사용 가능했던 경우).
 - 온라인 전자자료: NLA 도서관장에게 권한을 부여하여 국가 컬렉션에 포함될 자료의 경우 온라인 제공 자료를 납본하도록 요청 할 수 있고, 요청 받은 자료는 1개월 이내에 NLA에 기탁되어야한다. 온라인 출판물의 납본은 웹 수집 로봇을 사용하여 요청하고, 웹사이트 내에서 자유롭게 사용 가능한 자료는 자동으로 수집 된다. 반면, 자료가 웹사이트에서 사용 가능하지만 자유롭게 액세스 할 수 없는 경우, 게시자가 eDeposit 서비스 또는 도서관과 합의한 다른 수단을 통해 자료를 전달한다. 웹사이트에서 제공되지 않는 온라인 자료인 경우 게시자는 요청한지 한 달 이내의 자료를 eDeposit 서비스 또는 도서관과 합의한 다른 수단을 통해 전달해야 한다.
 - 온라인 자료의 종류 및 형식: 전자 서적, 전자 저널, 잡지 및 뉴스 레터와 같은 온라인 자료. 형식으로는 EPUB, PDF 또는 책, 저널, 잡지, 뉴스 레터 및 MOBI 파일 등은 500MB 미만이어야 한다(이들 중 둘 이상을 게시 할 경우 선호하는 것은 EPUB). PDF, GeoPDF, TIFF 또는 GeoTIFF 파일을 허용하며, Word 문서를 허용하지 않는다.

4) State 및 territory 수준의 전자납본

- 노던 테리토리: 출판법에 의해 출판물의 URI(Uniform Resource Identifier)에 대해서 인쇄된 버전이 없는 인터넷 출판물을 요구한다. 인터넷 출판물이 기술적으로 복사되지 않도록 보호되고 있다면 요청에 따라 출판물의 전자사본을 제공하거나 인터넷에서 동일한 사본을 얻을 수 있는 수단을 제공해야 한다.
- 퀸즈랜드: Libraries Act 1988(Qld)은 발행인에게 주립 도서관과 의회 도서관에 발행된 자료의 사본을 제공하도록 하며, 출판물 생산에 있어 전자 형식의 경우 PDF 또는 Word(.doc) 형식으로 제공되어야 한다. 또한 도서관의 디지털 저장소에 전자 출판물을 저장 및 보존 할 수 있으며, 디스크나 USB로의 우편 전송, Dropbox 폴더 공유 및 다른 파일 공유 서비스를 통해서도 자료 납본이 가능하다.
- 태즈메니아: 출판물은 디지털 및 하드 카피 형식으로 출판되며, 인쇄물의 레이아웃이나 내용이 다른 경우를 제외하고는 디지털 사본만 보관해야하고, 전자 출판물은 LINC Tasmania의 “웹 기반 문서 보관소”를 사용하여 제출할 수 있다(STORS, Stable Tasmanian Online Repository System이라고 불림).

5) National eDeposit Network

- Ned.gov.au 웹사이트: 2019년에 시행될 National eDeposit 시스템 호스팅을 위해 설립되었다.
- 서비스: 현재 국가 납본, 보관, 관리, 검색 및 전송을 위한 혁신적인 온라인 서비스이며, 호주 전역에 출판된 전자 자료를 제공한다.
- 납본: 게시자가 시스템을 사용하여 간행물을 한 번만 납본하면 주 혹은 지역 도서관과 NLA의 법적납본이 완수되는 것으로 설정되어 있어 주별 납본을 중앙정부에서 통합 운영할 수 있다.
- 자료의 접근, 보존과 저장 및 제작: 납본된 자료는 주 혹은 지역 도서관 카탈로그 및 NLA 검색 서비스에서 접근 가능하며 NED(National eDeposit Network)를 통해 사용 가능한 디지털 출판물을 보존, 저장 및 제작이 가능하도록 지원하고 있다.

□ 영국

1) 전자납본 의무사항

- Digital Legal Deposit-An IPA Special Report, 24th July 2014
- Legal Deposit Libraries Act 2003: 납본 요건을 인쇄물을 비인쇄물로 확대하기 위한 규정이 도입 될 수 있는 기본 틀이 포함되어 있다.
- 2013년 4월 6일 법적납본 관련 규정(Legal Deposit (Non-Print Works) Regulations): 법적납본도서관이 전자저널 전국 컬렉션을 구축하고 보존 할 수 있도록 비인쇄물 자료를 기탁해야 할 의무를 확대하였다.
- 콘텐츠: 전자 서적, 디지털로 출판 된 뉴스, 잡지 및 기타 유형의 콘텐츠가 포함된다.
- 2005년 법적납본 자문위원회(Legal Deposit Advisory Panel)라는 독립적인 비영리 공공기관을 설립하였다.
- 2010년 협의 중에 강조되었던 문제는 자료의 범위에 대한 명확한 설명이 없거나 비인쇄물이 영국에서 발행 된 것으로 간주되는 것에 대한 불확실성, 콘텐츠에 대한 액세스 제한 문제, 콘텐츠를 제공하고 유지 보수하는 비용, 기탁 자료의 데이터 보안 및 무결성이다.
- 2012년에 정부는 법안 초안을 개정하여, 웹 수집을 통해 얻을 수 있는 온라인 콘텐츠를 포함하여 오프라인 및 온라인 콘텐츠를 다루는 규정의 범위를 명확히 하였다.
- 법적납본(비인쇄자료)에 대한 2013 규정이 제정되었다.

2) 법적납본 규정 적용 대상 자료

- CD-ROM, DVD 또는 마이크로폼에 포함된 자료와 같은 오프라인 출판물과 전자저널과 같은 온라인 출판물이 모두 포함된다.
- 구체적으로는 인쇄물 이외의 매체(인터넷, 전자 서적 또는 전자 저널의 콘텐츠) 또는 오프라인으로 출판 된 것(인쇄물 이외의 물리적 형태로 출판된 것) 중 출판된 저작물 CD-ROM, DVD-ROM 또는 마이크로필름)을 포함한다.

- 구체적으로 제외되는 자료: 개인 데이터, Facebook 또는 트윗 등 제한된 그룹에서만 사용할 수 있어 타인의 액세스가 제한된 소셜미디어 사이트의 작업, 필름 또는 녹음된 사운드로 주로 구성된 작업 또는 이에 부수되는 작업, 복사가 아닌 요청을 받아야하는 전자자료의 경우에는 시행 규칙이 발행되어야 한다.
- 시행 규정 허가 조건: 영국에서 출판하거나 영국과 관련된 도메인 이름이 있는 웹사이트 또는 영국 내의 장소나 사람에 의해 대중에게 공개된 것이다. 해당 저작물의 창작 또는 출판과 관련된 사람의 행위 중 하나가 영국 내에서 이루어져야 하며, 영국 이외의 사람들만 접근이 가능한 온라인 출판 저작물의 경우 영국에서 발행된 자료로 취급하지 않는다.

3) 비인쇄물 납본 관련 사항

- 영국에서 출판되는 모든 오프라인 저작물을 무료로 제공한다.
- 다른 기탁 도서관은 사본을 요청할 수 있다.
- 계약(동의)에 따른 자료 배포: 출판사가 하나의 출판물에 대해 전자 버전 혹은 인쇄본 두 가지 버전을 발행하는 경우, 발행인과 도서관은 인쇄물이 아닌 전자 형식으로 납본될 수 있다는 것에 대해 상호 동의 할지라도 납본의 기본 양식이 계속 유지된다.
- 웹 수집을 통한 콘텐츠 확보: 온라인 자료의 납본에 대한 동의가 없는 경우, 규정에 따라 납본도서는 웹 수집이라는 자동화 된 프로세스를 통해 규정의 기준에 해당하는 온라인 자료의 사본을 입수할 수 있다(수집이 시작되기 한 달 전에 게시자에게 통지해야 한다).

4) 자료 사용, 제한 및 면제 사항

- 자료 전송 및 보기: 납본된 비인쇄물은 다른 납본도서관으로 이동 및 대여할 수 있으며, 납본도서관은 자체 연구를 위해 자료를 사용할 수 있고, 자료를 받은 후 7일 후에 대중은 납본도서관이 관리하는 도서관 구내의 워크스테이션에서만 볼 수 있다.

- 자료 복사: 납본도서관은 (1) 보존을 목적으로, (2) 연구원 및 시각장애인에게 액세스 할 수 있도록 저작권 침해 없이 비인쇄물의 사본을 만들 수 있으며, (3) 사본은 상업적으로 이용 불가하다. 또한 자료를 파기하기 전, 보존을 위한 가장 적합한 버전으로 모든 관련 자료의 사본 한 부를 보관해야 한다.
- 자료에 대한 접근을 제한하는 저작권 보유자: 규정에 따라 저작권 보유자는 비인쇄물에 대한 액세스가 허용될 시 합법적인 이익을 부당하게 침해받을 수 있음을 입증할 경우, 납본도서관이 자료에 대한 액세스를 제공하는 것을 금지시킬 수 있다. 3년까지 최초 금지 명령을 요구할 수 있으며 매회 최대 3년 단위로 재요청을 할 수 있다.

□ 네덜란드

1) The Digital Heritage Network

- 2018년 3월부터 네덜란드는 국가적 차원에서 디지털자원을 장기보존하는 프로젝트 ‘The Digital Heritage Network’를 교육문화과학부 주도하에 구축운영하고 있으며 네덜란드 국가도서관은 2015년부터 ‘Digital Heritage Network’에 참여하고 있다.
- 네덜란드 국가도서관은 National Archives, Netherlands Institute for Sound, 그리고 Vision and the Cultural Heritage Agency와 협력하여 출판, 기록, 동영상자료, 박물관 등의 영역에서 네덜란드의 디지털 문화유산을 보존하는 책무를 담당하고 있다.
- 네덜란드어를 사용하는 벨기에 지역과 협력하여 네덜란드어협회에 지원을 받는 네덜란드문학 디지털도서관과 함께 디지털 문화유산을 공유하고 있다.
- 네덜란드 국가도서관 디지털자원 보존전략의 특성이 드러나는 디지털자원 보존 서비스는 e-Depot으로 2002년 전자저널 보존에 초점을 맞추어 전자저널 출판사와 협력 체제를 구축하여 디지털 보존을 시작한 것으로 현재는 연구데이터 영역까지 확장하여 네덜란드 내 주요 대학과 협력하여 연구데이터를 국가가 주도하여 보존하고 있다.

2) e-Depot

- 네덜란드 국가도서관의 디지털 보존은 크게 상업적 출판사와 기관과 협력하여 디지털자원을 보존하는 e-Depot과 웹사이트를 보존하는 웹 아카이빙으로 나뉜다.
- e-Depot은 2002년 DIAS 시스템을 구축하여 디지털자원의 장기보존 시스템으로 운영되는 체제였다.
- DIAS 시스템은 OAIS 참조모델에 기반하여 네덜란드 전자출판물을 장기보존하는 목적으로 구축되었으며 장기보존 시스템이 구축되면서 e-Depot을 운영하는 부서가 신설되었다.
- 네덜란드 e-Depot은 세계적인 출판사와 협력을 하면서 디지털 학술 출판물을 수집, 보존하는 체제를 구축하면서 점차 네덜란드 디지털자원을 보존하는 체제에서 국가적인 범위가 모호해지게 되었다. 그 결과 15년이 지난 현재는 전 세계의 디지털 학술 정보자원을 수집하고 프로세싱하는 것이 도서관의 핵심 업무가 되었다.
- e-Depot 협력 전자저널 출판사
 - Elsevier
 - BioMed Central
 - Nederlands Tijdschrift voor Geneeskunde
 - Blackwell Publishing
 - Oxford University Press
 - Taylor & Francis
 - Sage Publications
 - Springer
 - Atlantis Press
 - Koninklijke Brill
 - Directory of Open Access Journals
 - International Union of Crystallography
 - IOS Press
 - Mary Ann Liebert

-
- e-Depot과 기관 리포지터리 차원에서 협력하는 기관
 - KNAW
 - NWO
 - Erasmus University Rotterdam
 - Radboud University Nijmegen
 - University of Groningen
 - Delft University of Technology
 - Eindhoven University of Technology
 - Leiden University
 - Maastricht University
 - Twente University
 - Utrecht University
 - University of Amsterdam
 - Tilburg University
 - VU University Amsterdam
 - Wageningen University and Research Centre
 - Nereus
 - 2012년부터 네덜란드 국가도서관은 DIAS 시스템을 자체적으로 소프트웨어를 개발하여 e-Depot을 운영하고 있다.
 - 자체적으로 개발한 e-Depot 시스템은 향후 국제협력을 하는데 장애가 없도록 하기 위하여 METS, MODS, PREMIS, OAIS 표준을 따르고 있다.

□ 프랑스

1) 전자납본 의무사항

- 프랑스 국가도서관은 1992년 디지털도서관 Gallica를 시점으로 디지털화를 시작하면서 디지털 정보자원을 수집, 보존하는 체제를 구축하기 시작했다. 현재는 디지털 자료도 납본체제에 수용하여 운영하고 있는데 첫 번째 납본대상이 된 디지털자료

는 소리 및 영상 자료로 1938년과 1975년부터 납본을 받아 디지털 형태(CD 또는 DVD) 형태로 변환하여 배포하였고 납본영역을 데이터베이스와 소프트웨어 영역으로 확장하였다.

- 2006년에는 프랑스 도메인에서 출판되는 모든 온라인 자료를 납본대상으로 규정하였다. 이중 온라인 출판물을 중심으로 납본체제가 운영되어 오다 2012년부터 전자책을 대상으로 전자납본 체제가 확장되었다. 프랑스 국가도서관은 기본적으로 프랑스 국가 출판물 납본 시나리오를 인쇄 출판물, 디지털화된 출판물, 온라인 출판물 등 3가지로 운영하고 있다.

2) 전자책 납본

- 2012년부터 프랑스 국가도서관은 프랑스 출판협회와 협력하면서 납본 대상 온라인 자료 중 전자책의 경우 메타데이터 관리를 보존 차원, 서비스 차원, 자료관리 차원으로 구성하여 관리하고 있으며 출판사의 메타데이터 시스템인 ONIX와 연계하는 체제를 구축하려는 시도를 진행 중이다.
- 전자책으로 납본 받는 체제는 기존의 웹 하베스팅을 하는 체제와 별개로 독립적으로 출판사들이 전자책을 제출할 수 있는 플랫폼 체제로 진행 중이며 프랑스 국가도서관에 납본된 전자책은 디지털 보존 리포지터리인 SPAR(Scalable Preservation and Access Repository)로 연계된다.
- SPAR에 전자책이 저장될 때는 아래와 같이 장기보존적인 측면까지 고려해서 관리되도록 설계하고 있다.
- 전자책의 포맷은 포맷 점검 후 첫 단계에서 입수(ingest)되도록 할 것이다.
- 두 번째 단계에서는 장기보존의 측면에서 파일 단위로 입수할 것이다. 예를 들어 Flash나 Javascript가 포함된 EPUB 파일은 입수 대상으로 고려하지 않을 것인데 이는 장기보존시 보존한 콘텐츠를 재생할 가능성이 확실하지 않기 때문이다.

□ 일본

- 일본 의회도서관에서는 인터넷 등에서 출판 (공개) 전자 정보에서 도서 또는 연속 간행물에 해당하는 것(전자 책, 전자 잡지 등)을 온라인 자료로 규정하고 2013년 개정된 의회도서관 법에 따라 온라인 자료를 수집·보존하기 시작하였다. 그러나 보상을 제공하는 납본의 체제로 수집 보존하는 것이 아니라 무료이며 DRM이 없는 것으로 한정하여 수집하여 보존하고 있다. 유상 자료의 경우는 제공기관과 협력하여 의회도서관에서 열람할 수 있는 프로젝트를 시범적으로 운영하고 있다.
- 일본 의회도서관 수집 온라인 자료 수집 조건은 [표 2-4]와 같다.

[표 2-4] 일본 의회도서관 수집 온라인 자료 조건

수집대상	무료/DRM 프리		유상/ DRM 존재
	ISBN, ISSN, DOI가 부여 된 것	PDF, EPUB, DAISY에서 만든 것	
연보, 연감, 편람, 잡지, 사보, 게시판, 논문 집, 학술지 논문, 조사·연구 보고서, 학술지, 뉴스 레터, 학회 요지집, 사업 보고서, 기술 보 고서, CSR 보고서, 회사 연혁, 통계 간행물, 기 타 도서와 연속 간행물에 해당하는 자료	○ (수집 대상)	○ (수집 대상)	X (수집 대상)

- 기본적으로 무료이며 DRM이 없는 자료만을 대상으로 하여 ISBN, ISSN, DOI가 부여되고 있으며 PDF, EPUB, DAISY로 제작된 자료로 제한하고 있다.
- 온라인 자료 파일 및 메타데이터는 세 가지 방법 중 택 1을 하여 시행할 수 있도록 허용하고 있는데 자동 수집, 업로드(전송), 그리고 DVD 형태로 우편으로 제출할 수 있다.
- 자동 수집은 제목, 작성자, 발행인, 발행일, URL 등을 양식에 기재하여 제출하면 의회도서관 크롤러가 수집하는 체제이다. 업로드는 의회도서관의 전송 시스템을 사용하여 파일 및 메타데이터를 업로드하는 방법이며 우편 제출은 DVD-R에 파일 및 메타데이터를 저장하여 의회도서관 앞으로 우편 발송하는 방식이다.

□ 독일

- 독일 국가도서관의 전자납본 시스템은 수집부터 목록, 아카이빙까지 자동화된 체제로 전자책, 학위논문, 전자저널, 전자신문 등을 수집하여 보존하고 있다. 전자납본 시스템은 납본자에 따라 차별화된 납본 플랫폼을 제공하고 있다.
- 독일 국가도서관 납본 플랫폼의 특징은 다음과 같다.
- 웹: 단행본, 악보, 2006년 이전 학위논문 등을 자발적으로 납본하는 체제, 소량출판에 적합한 체제로 제공하며, 개별 업로드 파일 제한 500MB 이다.
- OAI-PMH(Open Archive Initiative, Protocol for Metadata Harvesting): HTTP 기반 프로토콜로 제작된 납본 애플리케이션으로 납본자나 도서관 양 측에서 클라이언트나 하베스터가 수집 요청을 보내면 납본자의 서버에서 자동으로 메타데이터와 원 파일을 수집해오는 체제를 말한다.
- Hotfolder(Ftp/WebDAV): 대량으로 납본하는 체제로 메타데이터와 함께 납본 파일이 패키지로 되어 있으며 FTP 형식으로 전송되는 방식으로 상업적 출판사가 선호하는 방식이다.

2.3.2 웹 아카이빙 사례

□ 미국

1) 웹 아카이빙 전반

- 2000년 파일럿 프로젝트 수립: 웹사이트를 수집하고 보존하기 위한 프로젝트를 수행하였다.
- 도서관 직원은 차세대 연구원을 위해 자료를 평가, 선택, 수집하여 목록을 작성하고, 액세스 권한을 부여하여 보존하는 방법을 연구하였다.
- 이후 ‘미국 내 선거’, ‘이라크 전쟁’ 및 ‘9월 11일’과 같은 주제별 웹 및 이벤트 기반 아카이브를 개발하였다.

-
- 웹 아카이브 방식: 주제, 이벤트 또는 주제 영역별로 그룹화하여 아카이빙된 웹사이트의 모음이며 보관된 사이트는 원본 사이트의 특정 시점에 대한 스냅 샷이다. 웹사이트의 변경 히스토리를 문서화하며, 대부분의 사이트는 두 차례 이상 아카이빙 되었다. 아카이브는 텍스트, 이미지, 오디오, 비디오 및 PDF를 포함하여 가능한 원본 사이트에 해당 자료들을 포함시키고 있다.
 - 웹 아카이빙 팀(Web Archiving Team): 디지털 콘텐츠를 관리하고 유지하는 역할을 수행하며, 웹 콘텐츠를 캡처하기 위한 라이브러리 전체에 대한 이해와 기술 인프라를 구축하였다. 또한 다양한 도서관 직원 및 국내외 파트너와 협력하여 정책 문제를 파악하고 웹 콘텐츠를 수집하고 보존하는 도구를 구축하였다.

2) 웹 아카이빙 도구의 특징

- 주요 기능 영역: 선택, 사용 권한, 획득, 저장 및 액세스 등
- 오픈 소스 및 사용자 정의 개발 소프트웨어를 사용한 전반적인 워크플로우
- 개발, 선정 및 허가: 큐레이터 직원이 보관할 웹사이트를 선택할 수 있도록 도와주는 DigiBoard(PDF)를 개발 및 구현 완료하였으며, 웹사이트 캡처 및 오프사이트 액세스, 웹 아카이브 품질 검토 프로세스에 대한 법적 권한 관리를 용이하도록 하는 플랫폼을 개발 중에 있다.
- 수집: 웹 아카이브는 Heritrix라는 오픈소스 웹 크롤러를 사용하여 생성되며, 이를 사용하여 웹사이트의 복사본을 만든다.
- 저장소: 웹 아카이브는 BagIt 라이브러리를 사용하여 BagIt 준수 패키지로 패키징된다.
- 액세스: 웹 아카이브 재생은 Wayback Machine의 로컬 설치로 활성화한다.

3) 기술적 요구 사항

- 웹사이트를 최대한 완벽하게 재현하는데 필요한 모든 코드, 이미지, 문서, 미디어 및 기타 파일을 검색한다.

-
- 웹 서버(예: HTTP 헤더)와 크롤러(예: 캡처 컨텍스트, 날짜 및 시간 스탬프, 크롤링 조건)의 기술 메타데이터를 캡처하고 보존한다.
 - 날짜/시간 정보는 동일한 자원의 연속 캡처를 구별하는 데 특히 중요하다.
 - 전달된 내용과 정확히 동일한 형식으로 내용을 저장한다.
 - HTML 및 기타 코드는 항상 그대로 유지되며, 동적 수정은 웹 아카이브 재생 중에 즉석에서 이루어진다.
 - 플랫폼 및 파일 시스템 독립성을 유지한다.
 - 기술 메타데이터는 파일 시스템 고유의 메커니즘을 통해 기록되지 않는다.

4) 적용 표준 사항

- 웹 아카이브는 Internet Archive ARC 및 Web ARChive(WARC) 컨테이너 파일 형식으로 저장된다.
- MODS를 사용하여 아카이브 된 웹사이트를 카탈로그화 한다.
- 예비 키워드, 제목 및 주제 메타데이터는 cURL을 사용하여 보관된 웹사이트에서 추출한 다음 제어된 이름과 제목 표제를 추가로 할당한 목록 작성자가 이를 검토하고 향상시킨다.

5) 웹 아카이빙 사이트 범위(선택적 웹 아카이빙)

- 미국 정부(연방, 주, 지방), 외국 정부, 정치적 후보자, 정치적 논평, 정당, 언론, 종교 단체, 지원 단체 및 공무원이 포함된다.
- 기록관리 단체, 옹호 단체, 교육 및 연구 기관, 창조적인 표현물(만화, 시 등) 및 블로그가 포함된다.
- 도서관은 웹사이트를 포함한 전자자료의 선택을 안내하는 수집 정책 진술 및 기타 내부 문서를 보관한다.

-
- 그 외에 URL, 상표, 저작권 선언, 소유권, 게시일 등 모든 식별 사이트 문서를 완전히 복제하려고 시도하고 있다.

6) 웹 아카이브 종류

- U.S. Election Web Archives: 2000년 이래로 이 도서관은 미국 대통령 선거, 의회 선거, 지사 선거와 관련된 웹사이트를 보존했다.
- U.S. Congressional/Legislative Web Archives: 도서관은 의회 구성원뿐만 아니라 의회위원회의 웹사이트를 정기적으로 보존한다.
- Law Library Web Archives: 의회 도서관은 대법원에 변경 사항을 문서화한 웹 콘텐츠를 보존한다. 그들은 또한 미국 변호사 협회가 승인한 법률 학교, 연구 기관, 싱크 탱크 및 기타 전문 지식 기반 조직과 관련된 권위 있는 법률 블로그의 선택적 수집을 개발했다.
- Single Sites Web Archive: 도서관의 전통적인 아날로그자료를 강화하고 확장하기 위해 디지털 형태로 제작 및 생성된 콘텐츠를 선택한다.
- Other Events and Themed Web Archives: 도서관은 다양한 다른 이벤트 및 주제 컬렉션의 콘텐츠를 선택하고 보존한다.
- International Web Archives: 도서관의 웹 아카이브는 전 세계적으로 선거, 이벤트 및 주제를 기록하는 범위에서 국제적이다.
- 전 세계 사건 아카이빙: 미국 및 전 세계에서 전개되는 사건을 아카이빙하기 위해 다른 기관들과 협력하고 있으며 액세스는 파트너 웹사이트를 통해 제공된다.

7) 웹 아카이브 수집 기간

- 콘텐츠가 변경되는 빈도에 따라 일주일에 한 번 또는 매달 한 번 웹사이트를 크롤링되고 있다.
- 일부 사이트는 한 번 또는 두 번 일 년에 한 번도 크롤링되지 않을 수 있다.

-
- 일부 보존 활동은 국가 선거 전후 또는 직후와 같이 시간에 민감한 사안과 관련이 있다.
 - 기타 보존 활동은 지정된 종료일 없이 진행될 수 있다.

□ 캐나다

1) LAC의 웹 아카이빙

- 2005년부터 LAC는 연방 및 비 연방 웹자원을 수집하고 있으며, 캐나다 웹 아카이브(GCWA) 웹사이트는 보관된 연방 웹사이트에 대한 액세스를 제공한다.
- Treasury Board of Canada Secretariat's Web Renewal Initiative 2016-17의 일환으로 대부분의 연방 웹 정보가 캐나다 도메인으로 이전하도록 추진하고 있으며 진행 중인 보존 활동의 일환으로 LAC는 이전의 모든 연방 자료를 수집하여 GCWA에서 이를 제공한다.
- 캐나다 웹 아카이브(GCWA) 포함 사이트: 기존 웹 게시 연방 정부 정보, 더 이상 공개되지 않는 연방 웹페이지를 말한다.
- 비 연방 컬렉션에 대한 향후 계획: 캐나다 사회와 관련된 특정 주제에 중점을 둔 비 연방 웹 컬렉션에 대한 액세스를 제공하기 시작할 것이다.

2) 현재까지 수집된 주제별 웹 아카이브 컬렉션

- 수집된 컬렉션의 종류는 다음과 같다.
 - 연방 선거
 - 1차 세계 대전 100주년(2014-2018) 자료
 - 진실과 화해위원회
 - 2014 소치올림픽 등이 포함

□ 호주

1) 디지털 보존 종류

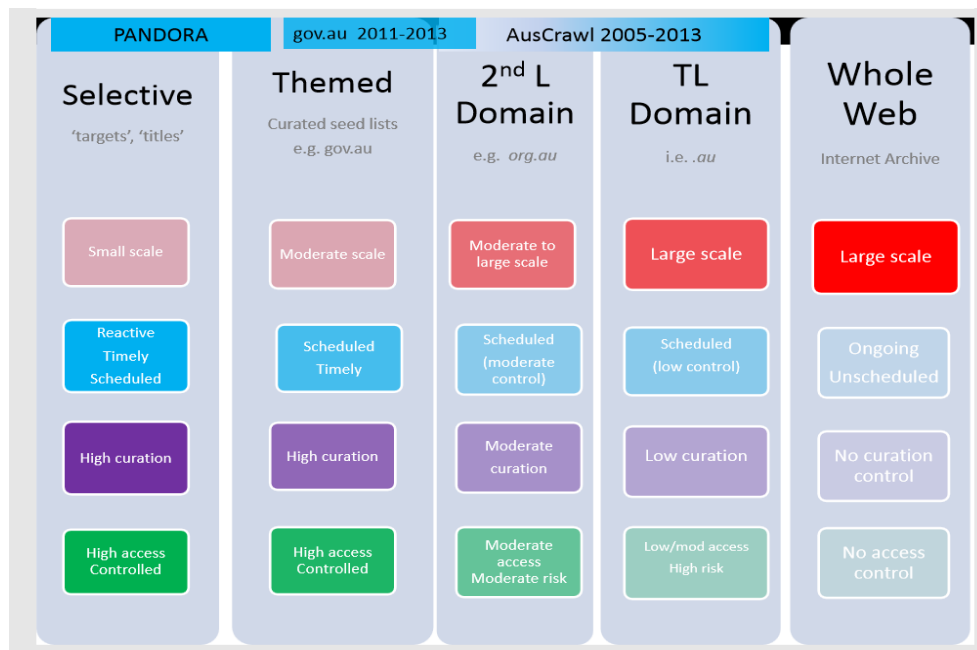
- PANDORA(Preserving and Accessing Networked Documentary Resources of Australia) 프로젝트: 시드니 올림픽 보존 프로그램에서 시작되었으며 가장 성공적인 웹 아카이빙 프로젝트로 평가받고 있다. 국가 차원에서 중요한 역사적 가치를 가지거나 장기적 보존이 필요하다고 판단되는 디지털자원에 대한 수집 및 이에 대한 목록 생성을 목표로 하였다.
- PANDORA 프로젝트에서 아카이빙된 자료는 호주 국가도서관의 목록에서도 접근 가능하다.
- AGWA: 2014년 PANDORA 이후로 발전되어, 2011년 6월부터 연방 정부 웹사이트에 대한 보존을 시작하였다.
- 또한 호주 국가도서관은 Archive-It 서비스를 통해 특정 아시아 태평양 웹사이트의 사본을 선택, 수집 및 보관하기 위해 2007년부터 Internet Archive와 파트너 관계를 맺고 있다. AGWA는 연간 2회에 걸쳐 하베스팅하는 것을 목표로 하고 있다.

2) 웹 아카이빙의 3가지 컬렉션 유형

- PANDORA 프로젝트에서 선택적으로 수집한 디지털자원
- 호주 국가 도메인에서 주기적으로 스냅샷 방식으로 하베스팅한 웹페이지
- 호주 정부 도메인(.gov.au)에서 대량으로 수집한 디지털자원

3) 웹 아카이빙 범위의 변화

- 주제 선택적 수집에서 도메인 선택적 수집으로 변화
- TL Domain으로의 수집 작업은 진행 중



[그림 2-1] 호주 웹 아카이빙 수집 범위의 변화

4) 웹 아카이빙 통계

- PANDORA Web Archive
 - 'Selective', 1996-Aug.2014 (88,500 instances)
 - 2억 4천 1백만 개 파일(11TB)
- Australian Domain(.au) Web Archive
 - 'Country TL domain', 2005-2013 (8 crawls)
 - 53억 8천만 개 파일(204TB)
- Australian Government Web Archive
 - 'Seed-list', 2011-2013 (4 crawls)
 - 3,450만 개 파일(3TB)

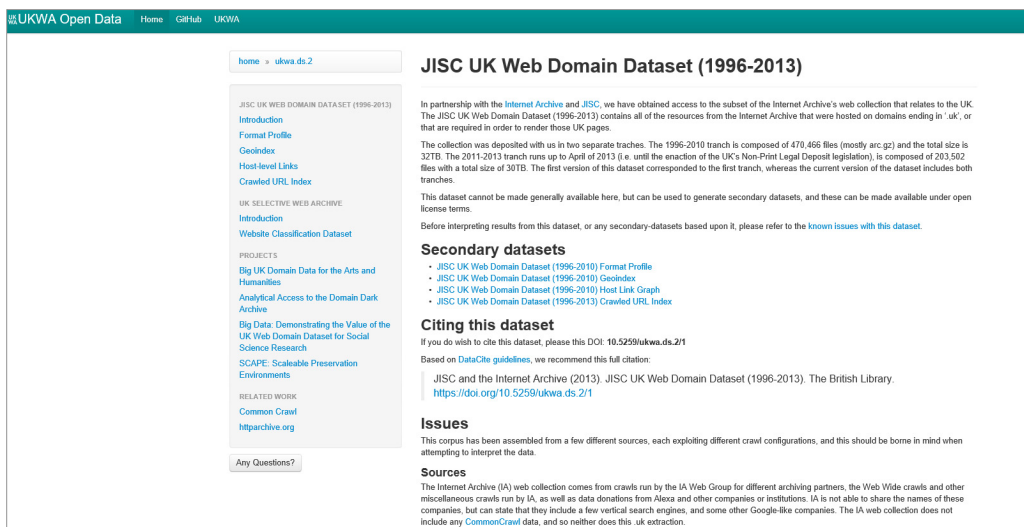
□ 영국

1) 3가지 주요 컬렉션

- Open UK Web Archive
 - 2004년부터 허가를 받아 웹사이트를 수집하였고 세 컬렉션 중 가장 규모가 작음
 - 보관용 사본은 수작업으로 품질을 검사하고 엄격하게 주석을 달고 있음
 - ‘영국의 삶과 문화’, ‘총선’, ‘2012년 올림픽’, ‘제1차 세계 대전 100주년 기념행사’ 등과 같은 특정 행사를 다룸
 - 지속적으로 추가되고 있으며, 누구나 영국 웹사이트를 추천하여 추가를 요청할 수 있음
- JISC 영국 웹 도메인 데이터 세트(1996-2013)
 - ‘.uk’로 끝나는 도메인에서 호스팅 되었거나 영국 페이지를 렌더링 하는데 필요한 인터넷 아카이브의 모든 리소스를 수집하여 보존한 세트
- Non-Print Legal Deposit 영국 웹 아카이브
 - 영국 전역의 연간 아카이브를 통해 확보한 수백만 개의 웹사이트를 포함함
 - 2013년 4월 정부가 발표한 비인쇄물의 법적 납본 규정에 의해 가능해짐



[그림 2-2] Open UK Web Archive



[그림 2-3] JISC 영국 웹 도메인 데이터 세트

2) 접근 가능한 웹 아카이브 자료

- 온라인 접근이 가능한 웹 아카이브 자료

- 특정 웹사이트 및 정보를 찾기 위해 Open UK Web Archive를 검색하거나, 주제 또는 특정 컬렉션별로 찾아보기가 가능함
- 다운로드할 수 있는 많은 데이터 세트가 있음
- 특별히 제작된 프로토타입 검색엔진인 Shine을 사용하여 JISC UK 웹 도메인 데이터 집합(1996-2013)을 검색할 수 있음
- 추세 분석 기능을 통해 17년 동안 웹에서의 변화를 추적 할 수도 있음
- 검색 인터페이스를 통해 postcode, 파일 유형, 연도 또는 도메인별로 전체 텍스트 결과를 필터링할 수 있음

- 관내 접근 가능 자료

- 비인쇄 법적 납본 UK Web Archive는 열람실의 컴퓨터에서만 액세스할 수 있도록 하였음.

- 기타 기관에서 접근 가능한 자료

- 비인쇄 법적납본 UK Web Archive는 해당 도서관의 열람실에서 열람 가능: 웨일즈 국립 도서관, 스코틀랜드 국립 도서관, 옥스퍼드 대학의 보드 랭 라이브러리, 케임브리지 대학 도서관, 트리니티 컬리지 도서관
- 영국 정부 웹 아카이브는 국립 문서 보관소에서 관리함.
- 영국 의회 웹 아카이브(Parliament Web Archive): 의회 웹사이트의 이전 버전 및 관련 외부 웹사이트, 기존 웹사이트에 대한 액세스를 제공함
- 인터넷 아카이브: 전 세계의 웹게시물을 포함하고 있음
- International Internet Preservation Consortium: 전 세계 웹 아카이브를 관리하는 조직 및 국가 도서관 목록을 제공하고 있음

□ 네덜란드

1) 웹 아카이빙 개요 및 선정기준

- 네덜란드 국가도서관에서 2007년부터 수집하면서 구축하고 있는 웹 아카이브는 선택적 아카이빙으로 수집된 자료의 이용성을 기준으로 판단하였을 때 전체 도메인 아카이빙보다 효과적이기 때문이라고 설명하고 있다. 그러나 저작권 문제로 아직은 관내이용만을 허용하고 있어 관외 이용 확장에 대한 방안을 모색하고 있는 단계이다.
- 네덜란드 국가도서관에서 웹 아카이브를 위해 대상을 선정하는 기준과 방식은 두 가지 이다.
 - 첫째, 대량으로 네덜란드 도메인에서 자동으로 웹사이트를 하베스팅 하는 것이며 둘째, 도서관의 정책에 따라 선별하여 웹사이트를 수집하는 것이다. 선별적 수집을 하는 경우는 네덜란드 국가도서관의 장서개발 정책에 기반한 것이며 기본적으로는 문화적, 학술적 가치를 지닌 것을 수집한다.
 - 또한 네덜란드 도메인의 웹사이트 중 최근 동향을 대표하는 혁신적인 성격의 웹사이트도 포함하고 있다.

2) 웹 컬렉션 종류

- 네덜란드 국가도서관은 다음과 같이 국제적 웹 컬렉션 구축에 기여하고 있다.
 - 2012 하계 올림픽 컬렉션
 - 2014 동계 올림픽 컬렉션
 - 1차 세계대전 기념 컬렉션

3) 웹 아카이빙과 관련된 법률적 이슈

- 납본제도가 없는 네덜란드에서 저작권을 침해하지 않고 웹 아카이빙을 하기 위해 서 네덜란드 국가도서관은 웹 아카이빙과 관련된 이슈에 대하여 Leiden 대학과 연구 프로젝트를 수행한 결과, 수집한 웹사이트를 장기보존하기 위하여 마이그레이

선 카피를 생산하였을 경우 웹사이트의 소유자에게 재허가를 받아야 한다는 문제점이 발생하였다.

- 이에 네덜란드 국가도서관은 저작권 제한을 하는 접근방식으로 문제를 해결하고자 네덜란드 국가도서관에서 웹 아카이빙을 하기 위해 웹 수집 로봇이 접근하였을 경우 웹사이트 운영자가 거부할 수 있는 기회를 제공하는 방식을 택하였다.
- 일정 기간 동안 웹사이트 운영자가 네덜란드 국가도서관의 웹 아카이빙에 대하여 거부를 행사하지 않으면 웹 아카이빙을 허용하는 것으로 처리한다.

4) 웹 아카이빙과 관련된 기술적 이슈

- 네덜란드 국가도서관은 International Internet Preservation Consortium (IIPC)에서 설계한 오픈 소스를 활용하여 웹 아카이빙을 하고 있다.
- 네덜란드 국가도서관에서 활용하고 있는 웹 크롤러는 Heritrix crawler 이다. 네덜란드 국가도서관은 단일 홈페이지를 구성하는 모든 파일을 수집하는 것을 목표로 하고 있다. 각각의 파일 단위로 메타데이터가 생성되면 기본적으로 파일 포맷과 수집된 날짜, 파일 크기 등이 생성된다.

5) 장기보존에 관련된 이슈

- 웹사이트는 브라우저와 플러그인 프로그램에 따라 가용성이 결정되기 때문에 장기보존시 보존된 웹사이트의 콘텐츠에 접근할 수 없는 리스크가 발생한다. 네덜란드 국가도서관은 IIPC의 방향과 맞추어 장기보존에 관련된 연구를 수행하고 있다.

□ 일본

1) 웹 아카이빙 구축과정

- 일본 의회도서관은 2002년도부터 일본 도메인의 웹사이트와 파일을 대상으로 사이트 저작권자로부터 허락을 얻어 선택적으로 수집·보존·제공하는 WARP을 구축

하기 시작하였다. 2009년 국가 기관 및 공적 기관의 웹사이트를 의회도서관이 수집하고 저장하는 것을 가능하게 하는 법률조항이 개정되면서 2011년부터 공공 기관을 중심으로 포괄적 크롤링을 시작하였다.

2) 수집범위

- 의회도서관 법률에 의해 수집대상으로 규정된 웹사이트 범위

의회도서관 법 제 24 조에 규정된 기관
 국가 기관 (입법, 행정, 사법: 지방 지분 부국 시설 등 기관 등을 포함한다)
 독립 행정법인
 국립대학 법인 (대학 공동 이용 기관 법인을 포함한다)
 특수 법인 등 (오키나와 진흥 개발 금융 공고 유한 공사 국제 협력 은행, 주식회사 일본 정책 금융 공고, 원자력 손해 배상 페로 등 지원기구 일본 은행 일본 사법 지원 센터 일본 사립학교 진흥·공제 사업단, 일본 중앙 경마 회 일본 연금기구 농수산업 협동조합 저금 보험기구, 예금 보험기구)

의회도서관 법 제 24 조의 2에 정하는 기관
 지방 공공 단체 (법정 합병 협의회를 포함한다)
 지방 공사 등 (港務局, 주택 공급 공사, 도로 공사, 토지 개발 공사, 지방 독립 행정 법인 地方競馬全國協會 지방 공공 단체 금융기구, 일본 하수도 사업단)

- 위와 같이 웹 아카이빙 수집 범위를 국가 및 공공기관으로 제한함으로써 수집 간기를 월간, 분기별로 할 수 있어 업데이트를 단기간에 주기적으로 할 수 있으며 기관 유형별로 구분하여 수집주기를 차별화하여 운영할 수 있다.
- 기관 유형별에 따른 수집 빈도는 [표 2-5]와 같다.

[표 2-5] 기관 유형별에 따른 수집 빈도

기관 유형	빈도
국가 기관	매월
지방 자치 단체, 지방 공사 등	분기 별
독립 행정 법인, 특수 법인 등	분기 별
국립 대학	분기 별

- 웹사이트에서 수집된 파일 유형과 비율은 [표 2-6] WARP에서 파일 단위로 수집된 자료의 유형과 비율과 같다. 이미지와 HTML이 주요 유형이지만 파일 형태를 규정하기 어려운 경우도 20%를 넘어 향후 장기보존시 재생할 수 있는 가능성을 보장하기 어렵다.

[표 2-6] WARP에서 파일 단위로 수집된 자료의 유형과 비율

파일 종류	파일 수	비율
image (jpg, png, tif, etc.)	2,346,090,681	37.70%
html, htm	1,387,641,568	22.30%
pdf	833,069,133	13.39%
php	194,237,127	3.12%
css	89,595,325	1.44%
xls, xlsx	70,841,992	1.14%
doc, docx	48,403,709	0.78%
others	1,252,501,280	20.13%

- 기본적으로 수집된 웹사이트는 의회도서관 관내에서 이용이 가능하며 수집대상 기관의 허가를 획득한 사이트는 의회도서관 홈페이지에서도 이용가능하다.
- 국가기관 및 공공기관을 중심으로 정보를 수집하기 때문에 허가를 획득하는 과정이 용이하다.

2.3.3 디지털화 사례

□ 미국

1) Labs

- 2017년에 웹사이트(<https://labs.loc.gov>)를 개설하였으며, 도서관 디지털 컬렉션에 대한 이용을 장려하도록 고안된 실험, 프로젝트, 이벤트와 자원을 주기적으로 선정하고 업데이트하여 제공함으로써 디지털화된 컬렉션의 활용성에 중점을 두고 있다.

-
- ‘데이터 챌린지’의 우승자, ‘상주 혁신가’의 ‘프로젝트 갤러리’와 도서관 관련 분야 지도자의 블로그 게시물 및 발표영상을 특별히 제공하고 있으며 사이트 이용자들은 실험적인 애플리케이션과 클라우드 소싱 프로그램을 체험해볼 수 있다.

2) 클라우드 소싱: 비욘드 워드(Beyond Words) 서비스

- Labs의 첫 번째 특집 프로그램: 대중들이 역사적인 신문에 수록된 만화와 사진을 식별하고 이에 관한 설명을 작성하여 해당 이미지를 검색이 가능한 데이터로 만드는 것이다.
- 시각화, 텍스트 분석 및 기타 디지털 인문학 방법론을 사용하여, 미국의 연대기와 관련한 새 지식을 발견하려는 연구원들이 활용할 수 있도록 텍스트 기반의 데이터셋을 확장하였다.
- 도서관 데이터 연구자들이 관심 분야에 대해 더 많은 것을 배우도록 돕고, 도서관의 클라우드 소싱 역량 상승을 돕기 위한 파일럿 프로젝트이다.

3) Library of Congress Strategy Plan 2016–2020

- 독특하고 희귀하고 이질적인, 연구가치가 있음에도 잘 알려지지 않은 자료들을 중심으로 디지털화가 이루어지고 있다.
- 소장 자료를 위한 공간 문제: 소장 자료들의 품질저하를 줄이고, 더 나은 접근을 제공하고자 하는 측면에서 디지털화가 활발히 논의되고 있다.

4) Annual Report 2016

- 새로운 디지털 컬렉션 관리 및 서비스 부서(DCMS: Digital Collections Management and Services Division): 도서관 내외로 디지털화 및 디지털 컬렉션 관리 지원을 수행하기 위하여 신설되었다.
- 해당 부서는 신문의 디지털화에 관한 기술 사양을 완료하고, Federal Agencies Digital Guidelines Initiative를 완성하는 업무를 수행하고 있다.

□ 캐나다

1) 디지털화 개요

- LAC(Library and Archives of Canada)의 소장 자료에 대한 디지털화는 연구 발전, 발견 촉진, 캐나다의 다큐멘터리 유산에 대한 광범위한 접근을 보장함으로써 가능한 많은 컬렉션을 디지털화하여 온라인에서 자료를 기술하고 검색할 수 있도록 하고 있다.
- 디지털화 우선순위의 효과적인 계획을 보장하기 위해 기관 전체의 전문성을 끌어내는 디지털화 위원회(Digitalization Committee)를 설치하였다.

2) 디지털화 대상 선정 단계

- 1단계의 콘텐츠 평가는 다음과 같이 공익수준, 정부지원 중요도, 유일성, 보존상태를 평가한다.
 - 공익수준: 공익 수준은 특정 컬렉션의 인기를 조사하여 결정되는데, 이 조사에는 캐나다인, 클라이언트 및 파트너 조직과의 협의가 포함됨
 - 캐나다 정부 지원도: 기념행사, 특별행사 및 공개정부계획(open government initiatives)에서 캐나다 정부를 지원하기 위한 디지털화 프로젝트의 잠재력을 평가함
 - 유일성: 입수 가능 사본 및 유사한 자료가 없다면, 컬렉션을 디지털화하는 것이 더 나은 접근을 제공하고 장기적인 디지털 보존을 보장하기 위한 이상적인 방법임
 - 보존상태: 훼손되기 쉬운 자료일수록 장기보존을 위한 디지털화될 가능성이 높음
- 다음 2단계는 디지털 보존의 실현가능성을 평가한다.
 - 권리 문제(Rights Issues): 최종결정을 내리기 전에 저작권, 기증자 조건, 개인 정보보호 문제를 고려하고 다루어야 함
 - 메타데이터 요구(Metadata Needs): 기존 메타데이터를 활용할 수 있는지, 또는 메타데이터를 쉽게 만들 수 있는지를 고려해야 함

-
- 기술적 고려사항: 자료를 얼마나 쉽게 디지털화할 수 있는지의 문제와 관련되며, 여기에는 형식 및 요구되는 장비, 외부 파트너와의 작업공유 가능성 등의 요인을 포함

3) 디지털화 및 파트너십: 참여방법

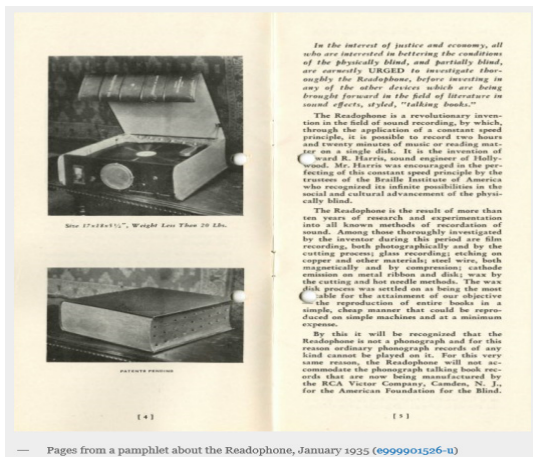
- 파트너는 선정된 문서뿐만 아니라 전체 시리즈 또는 레코드 세그먼트를 디지털화해야 한다.
- LAC는 파트너십 계약을 존중하면서 연구원과 직원에게 디지털 서비스를 제공하기 위해 디지털 사본에 대한 권리를 보유한다.
- 파트너는 동의한 시간 이후에는 LAC가 서비스 위임을 이행할 수 있도록 웹사이트, 열람실 또는 기타 수단에서의 제한 없는 사용을 위하여 LAC에 디지털 사본 및 관련 메타데이터를 전송해야 한다.
- LAC는 파트너 모델에 이상적인 기회를 제공하는 고가치의 많은 관심을 끌 만한 컬렉션 목록을 작성한다. 주제로는 이민, 군대, 범죄 및 교도소, 법원 및 법, 시민 기록, 연방정부 수립 이전의 기록, 기차, 해상 및 선박, 교육 등이 포함된다.

4) Crowded sourced 프로젝트

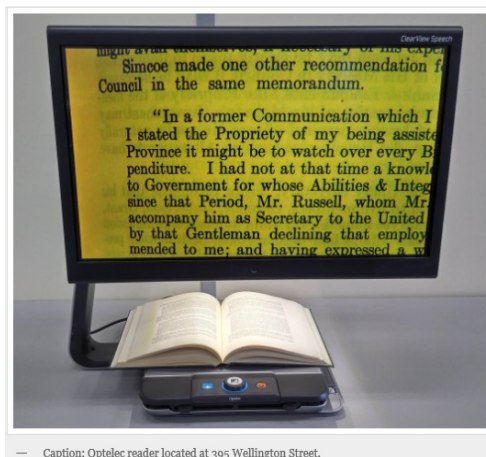
- LAC의 온라인 클라우드 소싱 툴을 사용하여 대중이 문서 작성, 태그 추가 및 검색과 이용을 가능하도록 한다.
- 최근 콜트맨 보고서(Coltman Report, 1818 - Seven Oaks 전투에서 벌어진 범죄에 대한 조사)를 주제로 프로젝트가 시작되었고 한 달 안에 이용자가 534쪽의 모든 문서를 기록한 성과를 보였다.
- 이 과정에서 외부 도서관 및 아카이브 디지털 프로젝트와의 공유를 추진하였으며, LAC의 컬렉션을 사용하는 데이터베이스에 링크하거나 소장 자료에 관련된 연구 정보를 리트윗할 수 있도록 지원하고 있다.

5) LAC의 DigiLab

- 오타와의 395 Wellington Street에 소재한 새로운 실습용 디지털화 시설로서, 사용자가 연구, 직장 및 지역사회에 LAC 컬렉션을 디지털화하고 기술할 수 있는 설비와 도구를 제공한다.
- DigiLab을 통해 디지털화된 모든 자료는 일반 대중이 이용할 수 있도록 온라인으로 제공되며, 디지털 컬렉션을 다양하게 활용할 수 있게 한다.
- DigiLab을 통한 디지털화 문서 사례는 [그림 2-4], [그림 2-5]와 같으며 최신 사례는 블로그에 업데이트되고 있다.



[그림 2-4] DigiLab을 통한 디지털화 문서 사례 1



[그림 2-5] DigiLab을 통한 디지털화 문서 사례 2

6) 2017-2018 연간보고서(Annual Report 2017-2018)

- 국가 유산 디지털화 전략(National Heritage Digitization Strategy)에 의하면 2017년 9월에 54개의 조직이 디지털화 프로젝트를 위해 LAC와 파트너 관계를 맺었다고 약속했다.
- 초기 자금지원으로 인해 3개의 신문을 디지털화하는 파일럿 프로젝트가 가능해졌다.

□ 호주

1) 도서관 컬렉션의 디지털화

- 도서관의 컬렉션을 이용자들이 활발히 이용할 수 있도록 전념하고 있다.
- 디지털화는 모든 호주인의 광범하고도 동등한 접근을 달성할 수 있는 가장 효과적인 방법 중 하나라고 보는 입장이다.
- 디지털화는 귀중하면서도 훼손 위험이 있는 자료를 보존하기 위한 효과적인 전략이다.
- 2000년대 초부터 디지털화 시작: 디지털 캡처 분야 전문지식 축적 및 디지털 프로젝트와 디지털 객체를 장기적으로 관리하고 있다.
- [표 2-7]은 호주 국가도서관의 이미지 캡처 표준이다.

[표 2-7] 호주 국가도서관의 이미지 캡처 표준

Material type	Tonal resolution (pixel depth)	Spatial resolution*
Colour reflective formats, including: • coloured maps • pencil sketches with wash – sepia or coloured photographic prints • printed music • manuscripts • objects	RGB 8 bit (24 bits per pixel)	Larger than A4: 300ppi; Maps: 600ppi A5 to A4: 600ppi A7 to A6: 1200ppi Under A7: 2000ppi
Colour transparencies, including 35mm	RGB 16 bit (48 bits per pixel)	2000ppi
Colour negatives, including 35mm	RGB 16 bit (48 bits per pixel)	2000ppi Note: two digital masters are created: colour negative and colour positive; derivatives are created from the colour positive.

Material type	Tonal resolution (pixel depth)	Spatial resolution*
B&W reflective formats, including: • photographic prints • black and white line art • black and white map	RGB 8 bit (24 bits per pixel)	Larger than A4: 300ppi; Maps: 600ppi A5 to A4: 600ppi A7 to A6: 1200ppi Under A7: 2000ppi
B&W negatives 35mm	Greyscale 16 bits per pixel	3000ppi (TIFF master and derivatives positive)
B&W negatives larger than 35mm	Greyscale 16 bits per pixel	2000ppi (TIFF master and derivatives positive)
B&W microfilm masters of newspapers (for the Australian Newspapers Digitisation Program)	Greyscale and bi-tonal (image optimised for OCR)	400ppi
Print publications	RGB 8 bit (24 bits per pixel)	400ppi at 100%

*: 주어진 해상도는 스캐닝 장치의 광학 기능을 기준으로 함

- 파일 저장: 모든 디지털 마스터 이미지는 압축되지 않은 TIFF 파일로 저장하는데 이에 대한 표준 기술요건은 다음 표의 '호주 국가도서관의 이미지 캡처 표준'과 같다.

2) 컬렉션 디지털화 정책 주요 사항

- 권한: 1960년 호주 국가도서관법(National Library of Australia Act 1960)은 호주 및 호주인들과 관련된 포괄적인 컬렉션을 비롯하여 도서관자료의 국가 컬렉션을 구축하기 위한 권한을 호주 국가도서관에 제공한다. 호주 국가도서관의 컬렉션을 디지털화하면 누구나 도서관을 직접 방문하지 않고도 언제 어디서나 이용할 수 있다. 또한 연구 및 평생학습을 지원하고, 사람들이 혁신적인 방식으로 컬렉션에 참여할 수 있게 하며, 새로운 콘텐츠를 만들기 위해 자료를 다른 용도에 맞게 재사용할 수 있게 한다.
- 컬렉션 종류로 서적, 저널, 신문, 지도, 악보, 단명자료, 웹 간행물 및 웹사이트, 그림, 구술 기록 및 필사본, 미간행 개인 아카이브와 같은 물리적 형식 아날로그 및 '디지털 형식으로 생성된' 자료가 모두 포함된다.

-
- 디지털화 대상 자료로는 정지 이미지, 객체, 텍스트, 소리 및 시청각 자료 등이 있다. 일부 경우에는(특히 신문) 자료는 마이크로필름에서 디지털화되며, 다른 것들은 자료로부터 직접 디지털화된다. 또한 아날로그 사운드 및 시청각 레코딩은 디지털 형식으로 변환되고, 디지털 레코딩은 대량 디지털 저장소로 전송된다. 다음은 디지털 대상자료의 특징을 설명하고 있다.

- 정지 이미지, 객체, 텍스트, 소리 및 시청각 자료 등이 있다.
- 정지 이미지: 실물자료에서 직접 디지털 방식으로 캡처한 그림, 사진, 지도, 포스터, 필사본, 악보 등
- 객체: 광범한 인공물을 포함
- 텍스트: 신문, 서적, 저널, 단명자료 등
- 일부 경우에는(특히 신문) 자료는 마이크로필름에서 디지털화되며, 다른 것들은 자료로부터 직접 디지털화된다.
- 아날로그 사운드 및 시청각 레코딩은 디지털 형식으로 변환되고, 디지털 레코딩은 대량 디지털 저장소로 전송된다.

3) 디지털화 대상의 선정법

- 선택적 디지털화(Selective digitisation): 디지털화 프로그램은 컬렉션의 크기, 권리 조건 및 자원 제약을 가진다. 주요 고려 요소로는 도서관의 장서 개발 우선순위, 문화적 또는 역사적 의의 등이 있다.
- 주문형 디지털화(Digitisation on demand): 디지털 저작물 또는 저작물의 일부를 주문하고자 하는 개인 또는 단체는 저작권 조항에 따라 호주 국가도서관의 Copies Direct 서비스를 통해 서비스 요금을 기준으로 주문을 진행할 수 있다. 이 프로젝트의 수수료 및 처리기한은 개개의 사례에 따라 협의되며, 주문에 의해 디지털화된 자료는 후속 사용자가 자유롭게 접근할 수 있다.

4) 디지털화의 방법

- 정지 이미지 또는 텍스트 디지털화를 진행하는 자료는 물리적 상태로 평가되고, 필

요한 경우 디지털 캡처가 시작되기 전에 보존 처리된다.

- 디지털화된 인쇄물의 전체 텍스트는 일반적으로 광학문자인식(OCR)을 통해 검색할 수 있다.
- 아날로그 소리 및 시청각 레코딩의 경우, 최대 신호량이 추출되어 보존 표준 형식으로 인코딩된다.
- 포맷 특정적 매체의 디지털 소리 및 시청각 레코딩: 콘텐츠가 추출되어 표준 형식으로 변환된다(예: CD 또는 디지털 오디오 테이프).
- 디지털화 작업은 조직 내에서 또는 호주 국가도서관의 사양에 아웃소싱하여 수행할 수 있는데, 대규모 디지털화 프로그램의 경우에 특히 그러하다. 아웃소싱을 할 경우에는 다음과 같은 관리업무를 수행한다.
 - 디지털 객체는 호주 국가도서관의 디지털 보존 정책에 설명된 대로 시스템 아키텍처 및 절차를 사용하여 저장되고 관리된다.

5) 파트너와의 협력 방식

- 기부자 프로그램: 도서관, 조직 또는 개인은 도서관이 지명한 디지털화 프로젝트 또는 도서관의 기부 프로그램을 통해 자체적으로 기금을 지원할 수 있다.
- 상업적 조직 파트너: 상업 조직과의 협력을 통해 디지털화 콘텐츠의 양을 크게 늘릴 기회를 얻게 되며, 이를 위해서는 파트너가 자료를 사용하는 방법에 대한 합의가 필요하다.
- 현물 기부 파트너: 자료 선택, 저작권 허가, 문서 및 설명과 같은 비재무적 기부금으로 도서관의 디지털화 활동을 지원할 수 있다.

6) 디지털화된 콘텐츠로의 접근 방법

- 카탈로그 및 Trove의 기록은 관련 디지털 파일, 검색 도구 또는 색인에 대한 직접 링크를 포함한다.

- Trove는 OCR 프로세스가 적용되는 디지털화된 항목의 전체 텍스트를 색인화하고 검색 접근을 제공한다.
- 디지털화된 저작물은 무료로 내려 받을 수 있다.
- 고해상도 사본은 Copies Direct 서비스를 통해 주문할 수 있다.
- Trove는 도서관의 컬렉션과 다른 참여자의 디지털화된 콘텐츠를 집계한다.
- Trove를 통해 직접 또는 Trove API(Application Programming Interface)를 통해 사용자가 디지털 콘텐츠에 참여할 것을 적극 권장한다.
- 사용자는 Trove를 통해 텍스트, 태그, 댓글을 바로잡거나 디지털화된 콘텐츠를 개인 또는 공개 목록과 대조할 수 있다.
- 검색엔진 및 기타 서비스의 메타데이터를 수집하고 색인할 수 있는 개방형 표준을 사용하여, 디지털화된 콘텐츠에 대한 검색 및 접근을 용이토록 하였다.

7) 디지털화 프로그램 관리 방법

- 내부위원회 감독: 디지털화 정책 및 표준을 감독하고 외부 당사자와의 파트너십을 포함한 주요 디지털화 프로젝트를 승인하는 내부위원회에 의해 감독된다.
- 보존 정책(Preservation Policy) 및 디지털 보존 정책(Digital Preservation Policy)과 관련된다.

□ 영국

1) 영국 국가도서관 전략적 우선순위

- 컬렉션 콘텐츠 및 메타데이터의 장기보존을 위한 새로운 디지털 저장소 및 보존 도구 세트를 구축한다.
- 장기보존을 위해 모든 유형의 콘텐츠와 메타데이터를 새로운 디지털 저장소에서 소장하도록 한다.

-
- 치밀한 프로세스와 리포팅 메커니즘을 구현하여 콘텐츠를 유지관리할 수 있도록 한다.
 - 특정 지역의 디지털 컬렉션 콘텐츠에 대해 적시에 신뢰할 수 있는 액세스가 가능하도록 보장한다.

2) 디지털화 비전

- 신뢰할 수 있는 장기 디지털 저장소에서 디지털 컬렉션을 제공하고 보존하기 위해 2020년까지 워크플로우를 마련하는 것

3) 장기 디지털 컬렉션 제공 및 보존을 위한 계약 기관

- 디지털 보존 연합(Digital Preservation Coalition, DPC)
 - 디지털 메모리를 보존할 수 있도록 2001년에 설립된 비영리 회원조직.
 - 회원은 디지털 콘텐츠 및 서비스에 대한 장기간의 액세스 제공
 - 디지털 컬렉션의 지속적인 가치를 이끌어내고 전략적, 문화적 및 기술적 도전에 대한 인식을 높이는 것을 목적으로 함
 - OCLC(Online Computer Library Center) 및 Wellcome Trust Library와 같은 다른 주요 조직과 함께 DPC의 창립 멤버임
 - 도서관은 정회원으로 집행위원회에 자리를 잡고 있으며 급변하는 환경에서 성장, 지속 가능성 및 관련성이 계속 유지 될 수 있도록 DPC의 방향을 조종하고 관리함
 - 80명이 넘는 회원들이 우수 사례를 발굴하고 개발하여 디지털 보존 문제에 대한 인식을 높이는 등 회원들에게 다양한 혜택을 제공하고 있음
- Open Preservation Foundation(OPF)
 - 2010년에 설립된 회원 조직으로, 디지털 보존 문제와 관련하여 신뢰할 수 있는 솔루션을 제공함
 - 영국국가도서관은 창립 멤버이며 현재 집행 이사회의 부의장을 맡고 있음
 - OPF의 성장을 보장하는 전략적 역할을 수행 중임
 - JHOVE 및 veraPDF를 포함하여 개발 도구 및 서비스를 활용할 수 있도록 하고

있음

- BitCurator 컨소시엄

- BitCurator 사용자가 디지털 컬렉션에 포렌식 접근법과 솔루션을 적용하고, 이에 필요한 네트워크를 제공하는 독립적 멤버십 협회
- 영국국가도서관은 집행위원회이며 창립 멤버임

- UNESCO PERSiSt

- UNESCO PERSiSt는 유네스코 세계 기록 유산(UNESCO Memory of the World) 하에 진행되는 프로젝트
- 목적은 효과적인 정책, 지속 가능한 기술 접근법, 최상의 보존 사례의 개발을 촉진함으로써 디지털 유산에 대한 장기간의 접근을 보장하는 것임
- 영국국가도서관은 정책 및 기술 실무그룹의 구성원으로서의 역할을 맡고 있음

□ 일본

1) 디지털화 및 매체 마이그레이션

- 일본 국가도서관에서 수행하고 있는 디지털 보존 프로젝트의 특성은 디지털자원의 장기보존 프로젝트에서 나타나고 있다.
- 일본 국가도서관은 디지털 장기보존 프로젝트를 성공적으로 수행하기 위하여 이를 위한 연구를 단계별로 시행하였다.
- 2002년부터 2004년까지 수행한 연구를 기반으로 일본의 디지털자원을 보존하는 기본 전략을 수립하였다.
- 기본 전략을 바탕으로 2006년부터 2010년에는 디지털 보존 매체 및 정보유형별로 디지털자원을 장기보존하는 연구를 세분화하여 각 매체와 정보유형별로 포맷과 마이그레이션에 대한 전략을 수립하였다.
- 일본 국가도서관에서 수행하는 장기적 디지털자원 보존은 CD-ROM, 플로피 디스크 등 이전 저장매체에 수록된 디지털 파일을 장기적으로 보존하는 구체적인 방안

을 제시하여 매체에 초점이 맞추어져있는 전략이라고 평가할 수 있으며 두 번째 단계에서는 영상, 음향 자료 등 정보 유형별로 전략을 수립하는 것으로 장기적 보존 방향을 설정하고 있다.

2) 매체 마이그레이션 및 메타데이터

- 2006년부터 시작된 연구 내용은 시청각 자료의 규격 및 재생 장치에 관한 연구에서 시작하여 오래된 시청각 정보매체의 장기보존을 목적으로 시청각 자료의 아날로그에서 디지털로의 변환에 대한 연구로 아날로그 자료를 디지털화하는 과정에서 고려해야할 기술적 이슈에 초점을 맞추었다.
- 특히, A/D 변환 방법을 설명하는 메타데이터에 대한 중요성을 강조하면서 아날로그 자료를 디지털화하고 장기보존에 적합한 파일 형식으로 변환하는 과정을 나타내는 메타데이터 항목을 고려할 때 다음과 같이 크게 세 그룹으로 제시하였다.
 - 원재료에 관한 요소(매체, 기록 방법 등)
 - 디지털화된 자료에 대한 요소(파일 형식, 압축 방법, 압축 후 파일 크기 등)
 - 디지털화에 대한 요소(디지털화 된 날짜, 장비, 응용, A/D 변환 조건 등).

3) 장기보존 대책 관련 연구 수행

- 장기보존시 데이터가 손상되었을 때 복원하는 대책에 대해서도 연구를 수행하였으며 물리적 매체에서 디지털 파일로 변환하는 마이그레이션 프로세스와 변환된 파일을 재생하는 에뮬레이터에 대한 연구를 지속적으로 수행하였다.
- 또한 각 데이터 유형별로 디지털화하는 포맷 조건에 대하여 연구를 하여 디지털화하는 파일의 조건을 각 형식별로 도출하였다. 다음 [표 2-8]과 [표 2-9]는 각각 오디오 및 영상자료의 디지털 파일 규격을 설명해주고 있다.

[표 2-8] 보존 및 배포를 위한 오디오 디지털 파일 규격

용법	파일 형식	샘플링 속도	비트 깊이	비트 전송률
보존 파일	WAV (.wav)	48kHz	16 비트	-
		96kHz	24 비트	-
		192kHz	24 비트	-
	AIFF (.aif)	48kHz	16 비트	-
		96kHz	24 비트	-
		192kHz	24 비트	-
배포 용 파일	MP3 (.mp3)	-	-	128kbps
		-	-	320kbps
	WMA (.wma)	-	-	128kbps

[표 2-9] 보존 및 배포를 위한 영상 디지털 파일 규격

용법	파일 형식	샘플링 속도	비트 깊이	비트 전송률
보존 파일	WAV (.wav)	48kHz	16 비트	-
		96kHz	24 비트	-
		192kHz	24 비트	-
	AIFF (.aif)	48kHz	16 비트	-
		96kHz	24 비트	-
		192kHz	24 비트	-
배포 용 파일	MP3 (.mp3)	-	-	128kbps
		-	-	320kbps
	WMA (.wma)	-	-	128kbps

4) 디지털 장기보존 시스템의 기능

- 일본 의회도서관에서 디지털화 프로젝트를 위해 수행한 디지털정보의 장기보존 및 접근성을 보장하려면 ‘디지털정보의 파일 형식에 대한 기술 정보를 관리하는 기능’, ‘보관 시스템에 수용된 디지털 파일의 형식을 식별하는 기능’, ‘아카이브 시스템에서 더 이상 사용되지 않거나 쓸모없게 될 위험이 있는 파일 형식을 감지하고 이를 방지하기 위해 시스템 관리자에게 경고하는 기능’이 필요하다고 하였다.
- 디지털로 전환된 디지털자료의 장기적인 이용성을 보장하기 위하여 디지털정보에

대한 장기 액세스를 지원하는데 필요한 파일 형식, 소프트웨어 제품 및 기타 기술 구성요소에 대한 기술 정보를 관리하는 형식으로 레지스트리 프로토타입을 구축하고 있다.

2.4 종합 요약 및 시사점

□ 온라인 자료납본 보존 사례

- 전자납본 자체가 의무사항이 아닌 미국이나 네덜란드에서는 상업적 기관과 협력체제를 구축하는 것에 중점을 두고 있으며 영국의 경우도 출판사, 법적납본 도서관 및 정부 단체 간의 협약을 촉진하고 비인쇄물의 납본에 대한 규제 및 비 규제 옵션을 자문하기 위한 자문단이 구성되었다. 일본의 경우 유료와 무료 전자자료에 대하여 차별화하여 수집하는 방안을 모색하고 있다.
- 상업적 전자자료와 저작권관련 이슈로 인하여 국가도서관에서는 디지털 법적 납본 자료의 접근성 허락에 대한 다양한 옵션을 모색하고 있는 추세이다. 일본이나 캐나다 사례에서 나타났듯이 상업적 전자자료에 대하여 저작권 침해 없이 접근을 허용할 수 있도록 협력체제를 구축하고자 하는 시도가 있다.
- 전자책 납본 시 출판사와 연동하여 자동으로 납본시스템으로 연계하여 국가도서관의 전자자료 납본에 대한 업무를 적절 수준까지 자동화하는 시스템이 구축되어 있는 추세이다. 프랑스의 경우 납본시 기초 메타데이터를 입수한 후 PREMIS와 같은 보존 메타데이터까지 한 프로세스로 연동하는 메타데이터 종합관리 체계를 구축하려는 시도를 하고 있으며 독일의 경우 납본자에 따라 차별화된 납본 플랫폼을 제공하는 시스템을 구축하였다.
- 최근 들어 보존된 디지털 파일의 이용성을 고려하여 장기보존에 관련된 이슈를 논의하고 있는데 보존된 디지털자료가 가독될 수 있는지를 고민하며 디지털자료의 포맷 변환에 대하여 주목하고 있다. 프랑스 경우 전자책 보존 시 장기보존을 고려하여 포맷을 관리하고 각 전자책에 포함되어 있는 파일 단위까지 장기보존을 고려하고 있으며 독일 경우 다양한 기관과 포맷뿐만 아니라 리더 애플리케이션 등에 대하여 연구를 수행하고 있다.

□ 웹 아카이빙 보존 사례

- 기본적으로 국가기관 및 공공자료를 대상으로 아카이빙하는 것이 가장 일반적이며 미국의 경우 국제적인 범위로도 인터넷 콘텐츠를 수집하고자 하는 시도를 하고 있어 주제별로 아카이빙을 할 경우에는 국가내외로 범위를 확장하는 추세이다.
- 웹 아카이빙 시 웹페이지에 포함되어 있는 컴포넌트를 처리하는 이슈가 다양하게 대두되고 있다. 기본적으로 웹 아카이빙 시 텍스트, 이미지, 오디오, 비디오 및 PDF 등의 컴포넌트를 가능한 한 원본 사이트에 포함시켜 관리하는 것을 목적으로 시작하였으나 다양한 컴포넌트를 장기보존하였을 때 활용성을 확보하기 어려운 문제점에 대하여 주목하고 있다.
- 기본적으로 온라인 접근과 관내 접근, 기타 기관에서의 접근 가능한 자료가 구분되어 있으나 수집된 웹자료의 활용을 극대화하기 위하여 도서관 통합검색서비스에서 통합검색을 지원하는 시도도 이루어지고 있다.
- 웹 아카이브의 장기보존을 위한 시스템을 자체 개발하는 경우 PREMIS, OAIS 등 기록관리 장기보존 시스템의 표준을 도입하여 장기보존 시스템을 표준화하는 추세이다.

□ 디지털화 자원 보존 사례

- 디지털화는 기본적으로 장기보존을 목적으로 하기 때문에 지속가능한 디지털 파일 형식과 표준화가 주요 이슈이다. 일본의 경우 아날로그 매체를 디지털 파일로 전환하는 작업과 병렬하여 장기보존을 위한 포맷에 대한 연구를 전략적으로 수행하고 있는데 아날로그 매체에서 디지털 파일로 전환하기 전에 디지털 파일의 포맷과 포맷 레지스트리, 디지털 파일 저장매체 등 전반적인 환경 전체에 대하여 중장기적으로 단계별로 연구를 수행하고 있다.
- 최근에는 디지털화에 대한 이슈가 디지털화를 하는 프로세스 중심에서 디지털화한 컬렉션을 활용하는 것으로 바뀌고 있는 추세이다. 디지털화를 하는 궁극적인 목적은 자료를 장기적으로 잘 활용하자는 것이므로 기존에 디지털화 한 컬렉션 이용을

활성화하는 서비스 방안이 다양하게 시도되고 있는 추세이다. 미국의 경우 2017년에 디지털 컬렉션 이용을 위한 새로운 웹사이트 Labs를 개설하여 데이터 연구자들이 다양한 데이터를 활용할 수 있도록 실험적으로 서비스를 제공하고 있으며, 사이트 이용자들에게 실험적인 애플리케이션과 클라우드 소싱 프로그램을 체험해볼 수 있도록 지원하고 있다. 캐나다의 경우 시내에 DigLab 같은 다양한 시설을 설치하여 이를 통해 이용자들이 디지털화된 컬렉션을 바탕으로 새로운 콘텐츠를 생산하는 방안을 모색하였다.

- 디지털화를 하는 대상 선별에 대해서도 다양한 시도가 이루어지고 있다. 캐나다의 경우 콘텐츠 평가와 실현 가능성 평가 등의 2단계 평가 프로세스를 구축하여, 어떤 디지털화 프로젝트가 캐나다인에게 가장 적합할지를 결정하는 데에 활용하고 있으며 호주는 선택적 디지털화, 주문형 디지털 방식을 채택하였다. 영국 경우는 디지털 문화유산 공동체와 협력하면서 디지털화 대상을 선정하고 있다.
- 디지털 장기보존의 경우 파일 포맷 활용성에 대한 이슈와 마이그레이션 이슈가 대두되고 있다. 일본의 경우 디지털 장기보존 시스템에서 디지털 파일의 활용성에 문제가 있을 시기에 시스템에서 자동으로 알람을 하는 기능을 구현하는 시도를 하고 있는 것으로 나타났다.

제 3 장

국립중앙도서관 디지털자원 보존 실무 개선

- 3.1 디지털자원 보존용 메타데이터
- 3.2 디지털자원 관련 업무 분석 및 KOLIS II 보존 기능
- 3.3 디지털자원 포맷 레지스트리
- 3.4 종합 요약 및 시사점

제 3 장

국립중앙도서관 디지털자원 보존 실무 개선

3.1 디지털자원 보존용 메타데이터

- 국립중앙도서관에서는 수집되는 온라인 자료에 대해 디지털장서관리시스템 개발 당시 MODS, METS, PREMIS를 적용하였으며, 2010년부터 본격적으로 데이터를 작성하였다. 외부기관이 작성한 디지털 파일의 연계에는 더블링크어 요약(simple) 버전을 사용하고 있다(이재선 2012).
- 국립중앙도서관의 디지털자원의 장기보존을 위한 메타데이터 개선(안)을 제시하기 위해서는 현재 국립중앙도서관의 보존용 메타데이터의 사용 현황을 조사하고 디지털자원 보존에 따른 메타데이터의 국제적 상호운용성을 확보할 수 있는 표준을 준수하는 방향으로 제시되어야 할 것이다.
- 이를 위해서 국립중앙도서관의 디지털자원 보존 체제에 관련하여 PREMIS 기반의 보존 메타데이터 관련 국제 연구 동향을 조사, 검토하고 현재 국립중앙도서관의 보존용 메타데이터 스킴의 요소와 적용 사례를 조사하고자 한다. 국립중앙도서관의 보존 메타데이터 개선(안)을 제안하고 추후 이를 반영한 시스템 구현에 참고자료가 될 수 있도록 보존 메타데이터 요구사항 및 설계안을 제시하고자 한다.

3.1.1 디지털자원 보존 메타데이터 설계원칙 및 국제 표준

□ 디지털자원 보존을 위한 메타데이터

- 정보자원의 보존은 정보자원 자체뿐만 아니라 해당 정보자원의 보존과 관련된 메타데이터까지도 함께 보존하는 것을 의미한다. 메타데이터는 디지털 객체의 보존과 관리, 검색을 지원할 수 있는 기본으로 디지털 보존의 가장 큰 핵심 부분이라 할 수 있다
- National Information Standards Organization(2004)은 메타데이터의 보존 기능을 다음과 같이 정의하였다.
 - 디지털정보는 깨지기 쉽고, 손상될 수 있음
 - 저장기술의 변경으로 인해 사용불가가 될 여지가 있음
 - 포맷전환과 에뮬레이션기법은 이러한 위기를 극복하는 전략이 될 수 있음
 - 메타데이터는 미래에도 자원들의 영속성과 접근성을 보장하는 핵심적인 역할을 수행함
 - 디지털 객체의 계통을 추적하고 물리적인 상세한 특성을 기술하고, 미래의 기술로 에뮬레이션하기 위한 행동들을 문서화 하는 등과 같이 아카이빙과 보존을 위한 특별한 요소들이 필요함
- 보존 메타데이터는 원 데이터에 대한 접근과 검색을 제공하고, 효율적으로 정보자원을 관리할 수 있도록 지원하는 것을 목적으로 한다(서은경 2005).
- 보존 메타데이터는 디지털 보존 과정을 지원하는 정보로서 디지털자원의 장기간 유지가능성(viability), 재연가능성(renderability), 이해가능성(understandability), 진본성, 무결성을 유지하는 기능을 하며, 이를 위한 핵심 요소를 제시할 수 있어야 한다.
- 보존 메타데이터는 디지털정보를 장기간 보존하기 위한 과정을 기술하는 구조적 정보로 디지털정보의 보존과 접근을 지원하며, 이러한 기능을 수행하기 위해서 다양하고 세부적인 정보를 포함하고 있다. 변환과 허가를 포함하는 관리 메타데이터, 기술 메타데이터, 구조 메타데이터 등의 다양한 범주가 보존 메타데이터에 포함되어 있다. 보존 메타데이터는 장기보존과 접근을 보장하는 모든 정보 인프라를 제공하며, 다른 객체와의 관계, 이력 정보를 제공한다. 데이터 객체 자체와 기술적 환경

정보를 함께 기술하여 급속한 기술적 변화에 대응할 수 있도록 한다. 보존 메타데이터는 디지털정보의 진본성과 무결성을 확보하고, 서비스 단계에서 저작권 관계에 효과적으로 대처할 수 있도록 저작권과 관련된 정보가 반드시 기술되어야 한다.

- 디지털자원의 보존을 위한 메타데이터는 데이터 교환과 이용을 위해 국제적인 상호운용성을 지니도록 국제 표준 OAIS 참조모형을 준용하여 개발되어야 한다. 초기의 국제적인 보존 메타데이터 스키마들은 OAIS 참조모형을 준용하여 개발되었다.

□ 디지털자원 보존 메타데이터 국제 표준: OAIS와 PREMIS

1) OAIS 참조모델

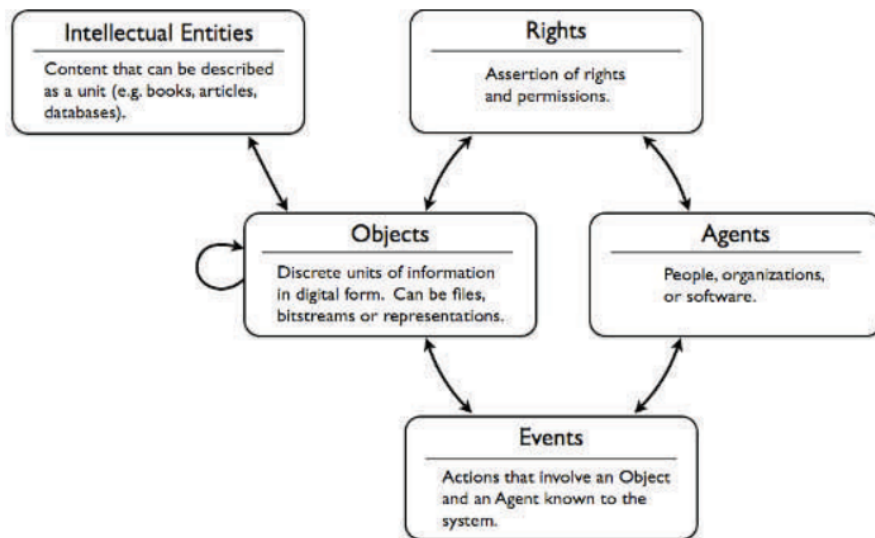
- OAIS 참조모델은 장기간에 걸쳐 디지털 객체를 보존하고 이에 대한 지속적인 접근을 제공해 주는데 있어 기반이 되는 개념적 프레임워크이다(Lavoie 2000, 26; CCSDS 2000: 이소연 2002, 48-49).
- OAIS 참조모델에서는 디지털 아카이브의 장기적인 유지 및 관리를 위해서 입수(Ingest), 아카이브 저장(Archival Storage), 데이터 관리(Data Management), 운영(Administration), 보존계획(Preservation Planning) 및 접근(Access)의 6가지 기능 영역을 제시하고 있으며, 이들 각각의 영역은 상호 유기적으로 작용함으로써 디지털 객체의 보존을 지원하고 있다.
- OAIS 참조모델은 이러한 6가지 영역을 통해 디지털 아카이빙을 구성하는 핵심적인 기능을 정의해 주며, 각각의 영역에서는 Submission Information Package(SIP), Archival Information Package(AIP), Dissemination Information Package(DIP) 등 다양한 정보 패키지를 구성하고 있다. 이들 정보 패키지는 디지털 객체의 장기보존을 위한 상위 개념의 구조와 요건을 제공해 주며, 이외에도 Reference Information, Context Information, Provenance Information, Fixity Information 등 디지털 객체의 보존과 관련된 다양한 측면을 각각의 정보 패키지로 구현하고 있다.

2) PREMIS(Preservation Metadata: Implementation Strategies)

- OAIS는 디지털 보존 시스템의 개념적 프레임워크를 제시한 것이므로 실제 보존

메타데이터를 설계하거나 실질적인 보존 업무에 적용하기에는 한계가 있다. 이에 OCLC/RLG는 각국의 전문가로 구성된 실무그룹을 조직하여, 보존 메타데이터 설계에 필요한 공통된 요소를 찾아 핵심 요소를 개발하고자 하였다. 디지털 보존을 지원하기 위해 보존 리포지터리가 알아야만 하는 요소를 밝히기 위해서 개발된 것이 PREMIS이다.

- PREMIS는 디지털 객체의 장기보존에 필요한 메타데이터 요소를 정리하고, 디지털 보존 시스템이 보존 메타데이터를 구현하기 위한 포괄적이고 실용적인 자원을 제공하는 것을 목적으로 한다.
- PREMIS는 데이터 모델과 데이터 사전으로 구성된다. 데이터 모델은 다음의 [그림 3-1]과 같이 지적개체(Intellectual Entity), 객체(Object Entity), 이벤트(Events), 행위주체자(Agents), 권한(Rights)의 개체 유형으로 구분된다. PREMIS는 실행 가능한 메타데이터 스키마를 개발하려면 각각의 요소를 가능한 한 상세하고 명확하게 정의해야 하는 동시에 이를 기술하고자 하는 개체(Entity) 유형에 연계해야 한다는 원칙 아래 메타데이터 요소의 논리적 조직화를 쉽게 실현하기 위해 지적, 객체, 이벤트, 행위주체, 권한의 5 가지로 개체 유형으로 나누어 메타데이터 요소를 조직화한 데이터 모델을 개발하였다.



[그림 3-1] PREMIS Data Model

-
- 지적 개체(Intellectual Entity)
 - 개념적인 서지 체계를 의미함
 - 기술 메타데이터 표준이 많기 때문에 실제로 지적 개체와 관련된 메타데이터를 정의하지 않음
 - 보존 시스템의 객체가 객체의 메타데이터에 지적 개체의 식별자를 포함시켜 표현된 지적 개체와 연관하여야 함
 - 객체 개체(Object Entity)
 - 보존 리포지터리에 실제로 저장되고 관리되는 객체(고유 식별자, 크기, 형식, 이름, 속성 정보 등)를 의미함
 - 객체를 렌더링하는데 필요한 하드웨어, 소프트웨어 등 다른 객체에 어떤 관계성이 있는지 정의함
 - 이벤트 개체(Events Entity)
 - 리포지터리에 있는 객체에 영향을 주는 행동(action)에 대한 정보 수집
 - 이벤트 개체는 디지털 객체의 출처를 알 수 있으며 진본성 입증 시 중요함
 - 행위주체 개체(Agents Entity)
 - 이벤트와 권한 개체에서 사람, 조직, 소프트웨어, 응용 프로그램 역할을 하는 에이전트를 의미함
 - 행위주체를 식별하는데 필요한 최소 단위의 의미 정의(에이전트에 대한 고유 식별자, 이름, 유형(사람, 조직, 소프트웨어))
 - 권한 개체(Rights Entity)
 - 대부분의 보존전략은 디지털 객체의 사본, 파생 버전 생성, 권리 보유자를 포함한 저작권법에 의해 제한되는 활동을 포함
 - 권한 개체는 리포지터리의 객체 보존과 직접적으로 관련 있는 권한 및 사용 권한에 대한 정보 수집
 - PREMIS의 데이터 사전은 보존활동에 필요하다고 생각되는 다양한 정보를 정리하고 체계화하여, 용어의 의미가 갖는 범위와 용법을 규정한 것이다. 즉 메타데이터

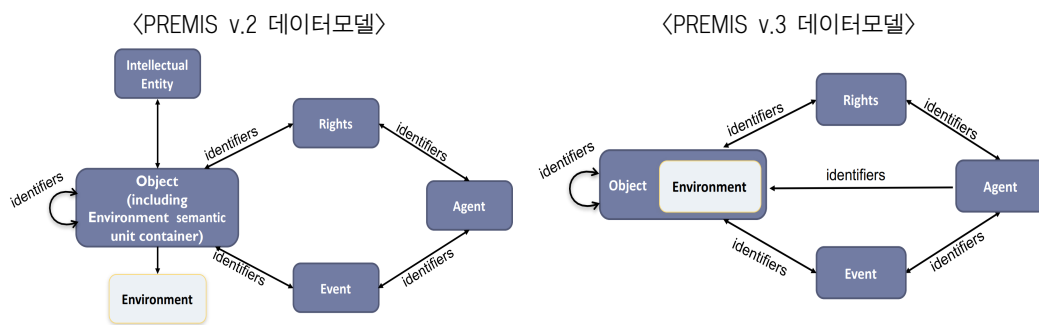
요소의 사전이라 할 수 있으나, 메타데이터 설계 시 환경에 따라 유연하게 응용이 가능하도록 메타데이터 요소(element)라는 표현 대신 의미단위(semantic unit)라는 말을 사용한다.

- 온라인 납본은 개체 유형별로 메타데이터 요소를 제안하고 각각의 요소를 여러 항목으로 구분하여 상세히 설명하고 있다. 각각의 요소에 대한 상세한 정의와 함께 보존 메타데이터 요소 스키마를 개발할 수 있는 기준을 제공한다. 각 요소에 대한 요소의 객체, 이벤트, 행위자, 권한에 대한 이름, 하위요소, 정의, 데이터 제약사항, 객체 범주, 적용가능성(예: 반복가능성, 강제성), 생성/유지 안내, 사용안내를 정의하고 있다. 데이터 사전 요소의 특징을 보면 요소들이 디지털 보존 과정을 지원하기 위해 다양한 범주를 포함하도록 설계되었으며, 특히 객체의 이력과 관련 있는 디지털 출처의 문서화와 다른 객체와의 관계를 문서화 할 것을 강조하고 있다.
- PREMIS가 규정하는 것은 메타데이터의 의미단위이며 그것을 기록하기 위한 방식은 다양한 보존 환경에 따라 자유롭게 결정하도록 허용한다. PREMIS는 데이터 사전에 대응할 수 있는 XML스키마를 작성하여 공개하고 있다.
- PREMIS는 디지털자원의 장기보존을 위한 가장 종합적이고 포괄적인 메타데이터 요소 스키마로 받아들여지고 있으며, 장기보존에 필수적인 메타데이터 요소를 제안하는데 있어 다음과 같은 특징이 있다.
 - 디지털 객체와 관련하여 파일 뿐 아니라 표현물과 비트스트림 수준까지 기술 가능
 - 보존 환경에서 디지털 객체의 장기보존에 있어서 가장 핵심적인 실행 가능성, 제공 가능성, 이해 가능성, 신뢰성을 지원함
 - 보존 리포지터리에서 디지털자료의 장기간 보존을 위해 알아야할 핵심적인 정보를 제공함
 - 엄격하게 정의된 가이드라인에 따른 메타데이터 설계를 지원하며 자동화된 작업 흐름을 지향하는 구현 가능한 메타데이터임
 - 보존 기술, 전략, 메타데이터 저장 및 관리에 대한 가정이 없어 기술 중립적임
- 이에 반해, PREMIS가 가지는 한계점은 다음과 같다.
 - 지적 개체와 표현 개념의 적용이 명확하지 않음

- 데이터 모델의 구현을 위해 준수해야 할 구축상의 지침 및 최소한의 규정이 없음
- ‘의미단위’, ‘의미구성요소’는 다양한 디지털 보존 자원이 내포하고 있는 정보의 특성과 서지정보를 표현하는데 한계가 있음

3) PREMIS version 3.0 (PREMIS Editorial Committee, 2015)

- 2015년 6월 10일, PREMIS 편집위원회는 “Data Dictionary for Preservation Metadata” (보존 메타데이터를 위한 데이터 사전)의 버전 3.0을 공개하였다. 버전 3.0에서는 다음의 [그림 3-2]와 같이 데이터 모델을 수정하였으며, 하드웨어 및 소프트웨어 환경이나 지적 개체 등의 정보를 표현하기 위한 기능을 확장하였다. 이와 함께 XML (eXtensible Markup Language) schema과 OWL(Web Ontology Language) ontology도 함께 개정되었다.



[그림 3-2] PREMIS Data Model version 3
(PREMIS Editorial Committee, 2015)

- PREMIS 3.0은 PREMIS 2.2의 발표 이후 계속해서 실행성과 상호운용성에 초점을 맞추어 연구가 진행되었고 사용자들의 피드백을 받아 문제점을 보완하고 개선시킨 것이다. PREMIS 3.0의 데이터 모델은 객체 개체, 환경 개체, 행위주체자 개체, 권한 개체, 이벤트 개체로 구성되었으며 개체들 간의 관계는 다음과 같이 표현된다. PREMIS 3.0의 주요 개정 내용은 크게 세 가지로 정리할 수 있다.
- 첫째, 지적 개체(Intellectual Entity)를 객체 개체(Object Entity) 범주로 수정: PREMIS

3.0의 데이터 모델 개정에서 가장 주목할 것은 지적 개체가 복잡한 디지털 객체의 속성을 표현할 수 있도록 객체의 일부로 수정된 것이다. 이로써 PREMIS의 지적 개체는 다른 지적 개체를 포함하는 경우, 예를 들어 지적 개체인 웹페이지에 또 다른 지적 개체인 사진이 포함되어 있음을 표현할 수 있게 되었다. 또한 PREMIS 데이터 모델은 많은 보존 리포지터리에서 하나의 지적 개체의 여러 버전인 표현물을 유지하고 있는 것을 파악하고 이를 표현할 수 있도록 수정하였다. 예를 들어 하나의 디지털 이미지가 보존과 사용 환경에 따라 JPEG 이나 TIFF 등 다양한 포맷으로 저장, 관리될 수 있으며, 이러한 경우 지적 개체에 따른 표현물들을 서로 연계할 수 있도록 수정하였다.

- 둘째, 환경 개체(Environment Entity)를 객체 개체(Object Entity)와 독립적으로 수정: PREMIS 2.2의 환경 개체를 객체 개체에서 독립시켜 새로운 데이터 모델로 추가한 것이다. 디지털 객체를 재현하거나 이용할 때 필요한 운영시스템, 응용소프트웨어, 컴퓨팅 자원 등에 대한 환경 정보를 기존의 객체 개체 데이터 모델과 분리하여 별도의 데이터 모델로 추가시킴으로써, 기술 변화에 따른 파일과 소프트웨어의 구체적 정보를 확보할 수 있도록 하였다. PREMIS 3.0의 환경 개체는 모듈화 되어 객체 개체와 별도로 다른 객체들 간, 또는 리포지터리들 간에 재사용하고 공유하는데 필요한 주요 기술적인 정보를 제공한다. 또한 디지털 객체를 재사용하는데 있어서 관련 행위자, 개체와의 관계 표현이 매우 중요함을 인지하고, 환경 개체에 대한 데이터 모델에 `related EnvironmentPurpose`, `relatedEnvironmentCharacteristic` 등을 추가하였다.
- 셋째, 물리적 객체(Physical Object)를 추가: 물리적인 객체는 필사본, 인쇄 문서 등의 내용 정보 뿐 아니라, 디지털정보를 구현하기에 필요한 물리적인 하드웨어 기기 등의 환경 객체, 디지털 객체의 인쇄형 표현물 등을 말한다. 이러한 물리적인 객체는 실제로 많은 디지털 객체와 관련되어 있으며 PREMIS 3.0에는 객체가 디지털 이든, 물리적인 형태이든 관계없이 기술할 수 있도록 데이터모델에 물리적인 객체를 추가하였다. 물리적인 객체는 관련된 디지털 객체와 '관계'로 연계되어야 한다.
- 2016년 Dappert는 PREMIS 버전 3.0 튜토리얼을 통해 PREMIS 버전 3.0의 가장 큰 변화는 자원의 보존 생애주기를 반영하고 있다는 점을 들었다. Dublin Core나 다른 메타데이터 스킴에서 요소세트를 표준으로 제시하고 있는 반면, PREMIS는 다른

메타데이터 표준과 함께 보존 환경에 따라 사용목적에 맞게 설계할 수 있는 확장성이 특징이다. PREMIS의 의미단위는 실제 보존 메타데이터 설계 시, 1:1, 1:多, 혹은 多:1의 관계로 보존환경에 따라 다양하게 구현이 가능하다. PREMIS의 데이터 사전은 디지털 리포지터리의 보존 정보패키지를 교환하는데 있어서 포괄적인 표준으로 활용될 수 있으며, 버전 3.0은 여러 디지털 리포지터리의 복잡한 보존 프로세스들을 염두에 두고 OAIS의 SIP(제출정보패키지), AIP(보존정보패키지), DIP(배포정보패키지)에 적용하여 상호운용성을 확보하는 방안으로 제시되고 있다.

3.1.2 PREMIS 기반 디지털 보존 메타데이터 연구 분석

- 디지털 보존 메타데이터 사례 분석은 국립중앙도서관의 디지털자원 보존용 메타데이터의 개선 방안을 모색하기 위해서 2012년 PREMIS 2.0, 2015년 PREMIS 3.0 이후 실무 현장에서 메타데이터를 적용한 사례 연구들을 중심으로 조사하여 핵심 내용과 시사점을 분석하였다.

□ DPC 보존 메타데이터 보고서 2013

- DPC(Digital Preservation Coalition)는 디지털 보존과 관련된 다양한 활동을 수행하면서 여러 보존 프로젝트들의 주요 성과를 분석하여 다양한 유형의 디지털자원의 장기보존 문제에 적용할 방안을 모색하고 있다. 2013년 DPC는 PREMIS 2.0 적용 사례 분석을 통해서 보존 메타데이터를 구현하는 것과 관련된 주요 이슈 들을 종합, 정리한 보존메타데이터(Preservation Metadata)의 제2차 개정판을 발표하였다(Lavoie & Gartner 2013).
- 보고서는 OCLC의 전문연구원 Brian Lavoie와 King's College London의 Richard Gartner에 의해 공동 저술됐다. 보고서의 주요 내용은 사실상 보존 메타데이터의 국제표준인 PREMIS를 기반으로 구현된 보존 메타데이터 사례들을 중심으로 (특히 PREMIS 데이터 사전의 맥락에서), 구현과 관련된 주요 이슈, 디지털 보존 메타데이터에 관한 전반적인 주제를 다루고 있다.

-
- 보존 메타데이터는 PREMIS가 등장한 이후, 디지털자원의 장기간 관리를 위한 필수적 실무로 발달하고 있다. 보고서의 제1판은 2005년 PREMIS 데이터 사전의 발간에 맞추어 보존 메타데이터의 개념에서부터 표준 설계에 대한 것이라면, 제2판은 사실상의 국제표준인 PREMIS의 등장 이후 적용된 보존 메타데이터의 새로운 개발사항들에 중점을 두고 있다.
 - 이 보고서는 보존 메타데이터 사례들을 중심으로 지난 8년간의 발전사항들을 중심으로 구성되었다. 보고서 서문에 이 보고서는 특히 PREMIS 데이터 사전과 같은 보존 메타데이터의 주요 개발사항에 대해서 알고자 하는 디지털 보존 전문가들을 위해 작성한 것이라고 밝히고 있다.
 - 보존 메타데이터에 대한 관심이 이론에서 실무로 확장됨에 따라, 이 보고서는 데이터 사전 개정 내용, 지역 아웃리치, METS 틀에 중점을 둔 패키징, 디지털 보존 시스템에서 PREMIS 구현 등을 포함하여, PREMIS 데이터 사전의 발간 이후 생겨난 실제 구현에 관한 주요한 이슈들에 초점을 두고 있다.
 - 보고서의 주요 핵심사항을 정리하면 다음과 같다.
 - 디지털자원 관리를 위한 최상의 실무는 메타데이터와 PREMIS 데이터 사전 구축이다.
 - PREMIS의 등장과 적용은 많은 보존 메타데이터 개발이 가능하게 하였다
 - 개정된 PREMIS는 저작권 정보와 확장가능성을 확보하는 것을 강조하고 있다.
 - METS와 함께 PREMIS를 근간으로 하여 보존 메타데이터가 개발되었으며 몇 가지 구현 틀이 등장했다.
 - 보존 메타데이터 구현에 대한 모범사례 분석이 필요하다.
 - 보존 메타데이터에 대한 비용대비 효과를 평가하는 작업이 필요하다.
 - PREMIS는 OAIS의 개념적 프레임워크의 한계를 넘어 실제 다양한 실무에 적용할 수 있도록 개발되었지만, 각 요소들을 실제 데이터와 어떻게 연결시킬 것인지에 대한 사례와 지침이 명확하지 않아 실무에 적용하기에는 어려움이 있었다. 이에 따라 이번 PREMIS 2.0은 디지털자원 관리 및 보존 실무에 실제 적용에 가능하도록 PREMIS 데이터 사전을 개정하고, 그 외 여러 보존 메타데이터를 둘러싼 중점과제들의 해결책에 집중하였다. 그 결과 디지털 보존 전문가뿐만 아니라 보존에 관심을 가지고

있는 일반인들에게도 활용할 수 있다는 강점을 가진다.

□ 실무자를 위한 디지털 보존 메타데이터 보고서

- PREMIS를 기반으로 한 보존 메타데이터 실무 적용 구현 사례들을 수집, 분석한 보고서를 종합하여 OCLC는 2016년 “Digital Preservation Metadata for Practitioners: Implementing PREMIS”(Dappert, et al. 2016)를 출판하였다. PREMIS는 디지털 보존 콘텐츠 유형, 기술적인 인프라구조, 조직의 요구에 따라 구현될 수 있다. 이 책은 특히 보존 메타데이터 실무자가 PREMIS를 기반으로 보존 메타데이터를 설계나 구현할 때 고려해야 할 사항들을 실무적인 관점에서 제시하고 있다.
- 이 보고서는 PREMIS의 특정 버전을 적용하는 방법보다는 현재 작업하고 있는 보존 실무에서 다루고 있는 자료의 특성과 기관의 요구에 따라 응용하는데, 즉 PREMIS 기반 보존 메타데이터 응용프로파일을 설계하는데 초점을 두고 있다.
- 특히, PREMIS의 데이터 사전을 활용하여 보존 메타데이터를 구현 시 제기되는 여러 문제들과 보존 메타데이터 응용프로파일 설계 방법론들을 중심으로 다루고 있다. 보존 메타데이터 구현 실무에서 해결해야 될 영역들, 즉 보존 메타데이터의 핵심 요구사항이 무엇이며 이를 구조화하는 방법, 디지털 보존 대상 객체나 자료로부터 메타데이터를 수동 또는 자동으로 추출할 수 있는지, 다양한 출처에서 입수되는 다양한 포맷을 포괄적으로 수용, 적합한 메타데이터 표준을 선택, 마지막으로 메타데이터 표준을 특정한 보존 환경 및 포맷에 맞추어 적용하는 것을 제안하고 있다.
- 보존 메타데이터는 보존을 수행하는 기관의 목적이나 상황에 따라 그 범위가 다양하게 나타나기 때문에 이를 일관성 있는 방식으로 제안하는 것은 현실적으로 무리가 있다. PREMIS 데이터 모델을 기반으로 보존 메타데이터의 프레임워크로 특정한 보존 환경 시나리오에 따라 다양하게 구현이 가능하다.
- 보존 대상과 범위가 결정되면 PREMIS 기반의 보존 메타데이터 응용프로파일 설계를 위한 구체적인 전략이 수립될 수 있는데 그 핵심 내용과 방법론을 정리하면 다음과 같다.

- 보존 메타데이터 응용프로파일 설계를 위해서는 체계적인 요구 분석이 필요하다. 메타데이터 설계에 있어서 보존 대상 디지털 객체의 유형과 속성, 구조적 관계 등을 조사하기 위해서는 보존 위험과 기능 분석을 바탕으로 한 구체적인 보존 시나리오 분석이 수반되어야 한다.
- 위험 분석은 조직의 정책, 재정 및 법적, 기술적인 위험 측정 등을 포함하여, 실질적으로 디지털자원의 장기보존에 적절한 메타데이터와 대상기관의 포맷 변환이나 관련된 보존 활동의 유형이나 수준을 평가하는 것이다.
- 요구 분석을 위한 프레임워크는 디지털 보존의 포괄적인 국제 수준의 OAIS 표준, 위험 분석을 위해서는 SPOT, DRAMBORA 등을 활용할 수 있다.
- PREMIS 기반의 보존 메타데이터 구현사례들을 살펴보면, 도서관, 기록관 등 보존기관의 특성 뿐 아니라, 웹 아카이브, 동영상, 전자책 등 디지털 객체의 유형에 따라 객체, 이벤트, 권한 등이 다양하게 적용됨을 알 수 있었다. 이용자 커뮤니티의 요구와 함께, 기록관, 도서관, 박물관 등 다양한 커뮤니티와 컬렉션의 목적, 검색, 브라우징, 접근, 데이터 교환, 관리 및 보존 활동 등을 종합적으로 고려하여 메타데이터 기능과 수준을 결정하고 적절한 PREMIS 데이터 사전을 선정할 것을 제안하고 있다.

□ PREMIS 적용 메타데이터 실무 경험 사례 보고서

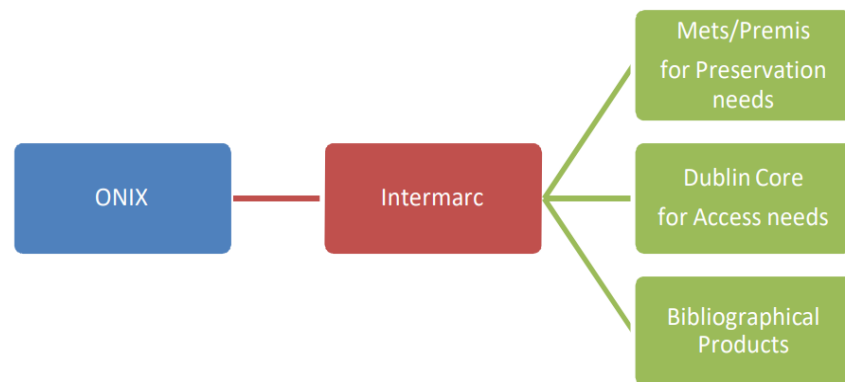
- 미국 도서관협회(American Library Association) 산하의 Association for Library Collections and Technical Services(ALCTS)는 모든 유형의 기관에서 도서관자료의 보존 및 재포맷, 새로운 기술을 적용하여 도서관 소장자원에 대한 지속적인 접근을 지원하는 것을 목적으로 구성되었다. 2016년 ALCTS의 보존 메타데이터 워킹 그룹은 PREMIS 적용 메타데이터 실무에서 경험한 사례를 통해서 보존 메타데이터 구현에 대한 중요한 문제점을 논의하였다(Krewer & Mullins 2017).
- ALA 산하 ALCTS(Association for Library Collections & Technical Services)의 2016년 PARS Preservation Metadata Interest Group Meeting 보고서의 핵심 내용은 다음과 같다.

- PREMIS를 기반으로 하는 보존 메타데이터 구현은 간단하지 않으며, 몇 가지 구현 툴과 함께 상황에 따라 다양한 보존 솔루션을 제공하는 벤더들이 등장하게 되었다. CONTENTdm, Archivematica, Rosetta 등 다양한 디지털자산관리시스템(DAMS: Digital Asset Management System)에 PREMIS를 연계할 수 있는 SIMP 툴이 개발되었다. 입수 과정에서 최소한의 PREMIS 메타데이터가 생성되는데 추후 이벤트 개체 관련한 핵심 요소 등을 확장하여 보완할 예정이다.
- 또한 메타데이터의 수준 결정(granularity)이 보존 메타데이터 설계에 있어 중요하다. 사례 연구들을 통해 기술(descriptive) 메타데이터 레코드를 통해 보존기관으로 입수 당시의 내용, 맥락 정보가 충분히 표현되는데 반해, ‘장기적인 이용가능성’ 측면에서는 객체의 중요 속성, 관계, 이벤트, 환경 정보 기술 등의 정보가 필수적으로 제공되어야 하며, 반드시 보존 메타데이터의 핵심 요소들로 구현될 것을 제시하고 있다.
- 새로운 BagIt 파일 포맷과 관련하여 디지털자원의 무결성 체크를 위해 체크섬 확인을 위한 별도의 태크 체크와 기입이 필요하며 이러한 요구사항이 PREMIS 개선방안에 반영될 수 있도록 권고할 예정이다.

□ BnF, SPAR 보존 메타데이터 연구(2014, 2017)

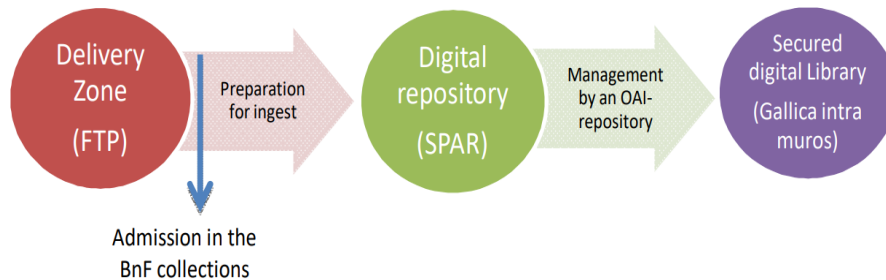
- 프랑스 국가도서관인 BnF는 프랑스 내의 모든 출판물을 수집, 보존하고 대중에게 제공하는 것을 목표로, 납본을 1537년 인쇄물에서, 1992년 전자문서, 2006년 인터넷에 이르기까지 다양한 형태의 출판물로 확대하였다. 디지털 문화유산을 보존하기 위해서 BnF는 확장 가능한 보존 및 아카이빙 시스템(SPAR)을 설계했다. 이 시스템은 OAIS 표준을 준수하고, 오픈 소스 기술 적용을 기반으로 모듈화할 수 있도록 설계되었다.
- SPAR의 보존 메타데이터와 관련하여 2014년 온라인 납본과 관련한 메타데이터 이슈들을 중심으로 한 보고서(Derrot & Oury 2014)와 2017년 보존 생애주기 기반 메타데이터 연구(Caron et. al. 2017)를 발표하였다. 2014년 BnF의 SPAR의 PREMIS 기반 보존 메타데이터에 관한 핵심 내용은 다음과 같다(Derrot & Oury 2014).

- SPAR에서 가장 중요한 부분은 디지털 콘텐츠의 보존으로, 자원의 무결성과 장기적 이용가능성을 지원하는 것이 매우 중요하다. 이를 위해서 FTP 플랫폼으로 납본 제출 시, 자동으로 자원의 무결성과 포맷을 체크하도록 한다.
- SPAR를 통해 온라인 납본으로 제출되는 디지털 콘텐츠에 대한 보존 메타데이터는 자동으로 추출, 생성되는 것을 원칙으로 한다. 출판사로부터 납본 제출 시 ONIX 메타데이터에서 데이터 무결성과 포맷 관련 정보를 추출한 InterMarc는 보존 목적으로 METS와 PREMIS 기반의 메타데이터를 부여한다([그림 3-3] 참조). 정보 검색 서비스를 위해서는 Dublin Core, 효율적인 관리를 위한 서지 데이터를 부여한다.
- 보존정보패키지(AIP)는 보존될 디지털 파일에 따라 데이터 객체, 데이터 객체 이해에 필요한 정보를 제공할 수 있는 메타데이터(특히, 표현정보와 보존기술정보(PDI)), 미디어에 패키지로 저장된 다양한 컴포넌트 간의 실질적, 논리적 링크를 설명하는 패키징 정보를 포함한다([그림 3-4] 참조).



[그림 3-3] BnF의 SPAR 보존 메타데이터

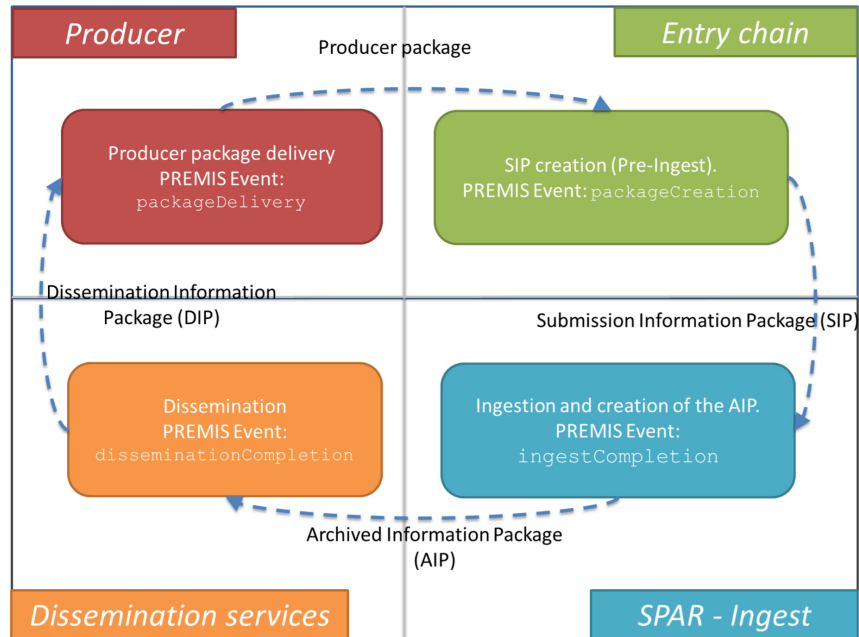
(〈출처〉 Derrot, S. & Oury, C., 2014, p.7)



[그림 3-4] BnF의 SPAR 보존 데이터 흐름

(〈출처〉Derrot & Oury 2014, p.8)

- 2017년 Caron 연구팀이 제안한 BnF의 SPAR의 보존 메타데이터에 관한 핵심 내용은 다음과 같다.
- 거시적인 관점에서 BnF의 디지털 콘텐츠의 장기적인 보존 관점에서 입수부터 관리, 제공까지의 디지털 콘텐츠의 생애주기 전반에 걸쳐 효율적인 메타데이터의 실행과 결정 과정에서 필요한 세부 내용들을 제시하였다.
- 현재 SPAR를 통한 보존 과정에서 다양한 유형의 정보 패키지를 적용하고 있는 것을 지적하고, 패키지는 고정된 형태로 존재하는 것이 아니며 보존 활동의 어떤 단계에 적용되는지에 따라 그 구성이 다르게 나타남을 강조하였다. 따라서, 메타데이터는 SPAR의 전체적인 보존 기능을 지원해 주어야 하며, 디지털 콘텐츠의 생애주기 전반에 걸쳐 각각의 단계에 맞게 PREMIS 기반의 메타데이터를 통해 관련 기술 사항을 제공해야 한다고 하였다.
- 또한 이들은 다음의 [그림 3-5]와 같이 PREMIS 기반의 보존 메타데이터는 생산자로부터 패키지가 생성되는 단계에서부터 선입수 시 SIP와 입수 후 AIP, DIP에 이벤트 개체를 통해 구현할 것을 제안하고 있다.
- 이처럼, SPAR의 보존 메타데이터는 장기적 보존관점에서 디지털 콘텐츠의 생애주기 전반에 적용되는 PREMIS의 이벤트 개체를 활용하여 보존 활동을 기술하는데 필요한 핵심적인 요소들을 제안하고 정보패키지 간의 상호운용성을 지원해주는 환경 구축을 제안하고 있다.



[그림 3-5] 생애주기 관리
(〈출처〉 Caron, et al. 2017)

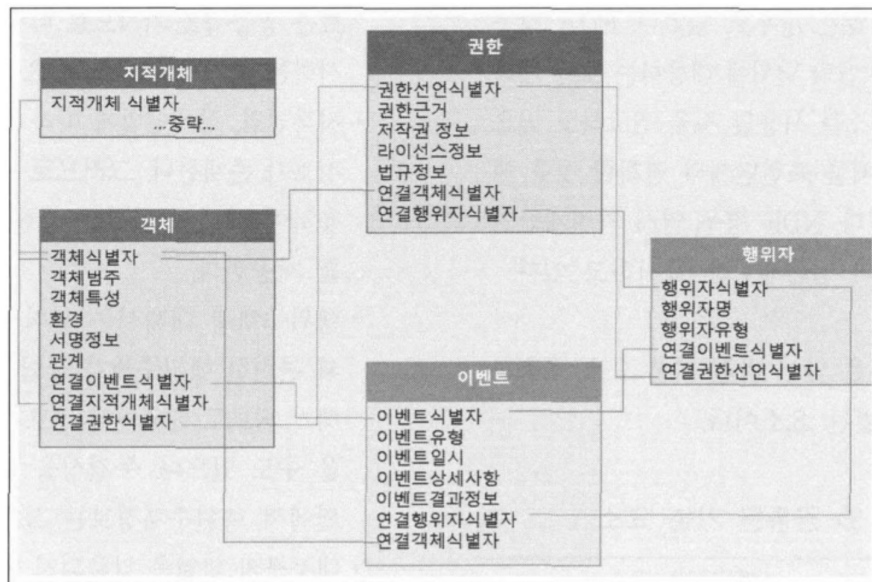
3.1.3 국립중앙도서관 보존용 메타데이터 분석

□ 국립중앙도서관 PREMIS 기반 보존 메타데이터

- 2011년 설계된 국립중앙도서관의 PREMIS 기반 보존 메타데이터는 국립중앙도서관 디지털자원의 기술 메타데이터와의 연계, 객체 개체와의 유기적 연계가 가능하도록 설계되었다. 이와 함께 PREMIS 기반 보존 연구를 통해서 국립중앙도서관의 서지정보 표준 메타데이터 체계인 MODS(Metadata Object Description Schema)와 METS(Metadata Encoding and Transmission Standard)를 기반으로 한 국립중앙도서관의 기존 목록 레코드에 보존 메타데이터를 활용할 수 있는 가이드라인을 제시하였다(박옥남 2012).
- 국립중앙도서관의 디지털자원의 보존과 메타데이터의 사용 현황에 대한 종합적인

분석 결과, PREMIS 기본요소를 도출하였다. 또한 보존 담당자와의 면담을 통해 국립중앙도서관에 필요 없는 요소는 제외하고 기본요소에서는 도출되지 않았으나 보존 정책을 위하여 요구되는 요소를 확장하였다. 메타데이터 설계 당시, 국립중앙도서관의 구체적인 보존 정책이 마련되어 있지 않아, 향후 마이그레이션 등 장기보존 정책이 구현될 것을 고려하여 생성 애플리케이션, 환경정보, 불변성 등의 요소를 포함하도록 하였다.

- 국립중앙도서관의 PREMIS 기반 보존 메타데이터는 객체(object), 이벤트(event), 행위자(agent), 권한(rights)의 총 4개 분야로 나누어졌다.



[그림 3-6] 국립중앙도서관의 PREMIS 기반 보존 메타데이터

(〈출처〉 박옥남, 2012, 96)

- 국립중앙도서관의 PREMIS 기반 보존 메타데이터의 객체(object), 이벤트(event), 행위자(agent), 권한(rights) 개체와 관련한 메타데이터 요소와 사용법은 다음 [표 3-1]에서 [표 3-4]와 같다(박옥남 2012, 97-101).

[표 3-1] 국립중앙도서관의 PREMIS 보존 메타데이터 - 객체 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
객체식별자 (PREMIS:object Identifier)	객체식별자유형 (PREMIS:objectIdentifier Type)			1:1	DLC
	객체 식별자 값 (PREMIS:objectIdentifier Value)			1:1	http://nrs.harvard.edu/urn3:FHCL.Loeb:sa1
객체범주 (PREMIS:object Category)				1:1	file
객체특성 (PREMIS:object Characteristics)	구성단계 (PREMIS:compositionLevel)			1:1	0
	불변성 (PREMIS:fixity)	메시지요약 알고리즘 (PREMIS:messageDigest Algorithm)		1:1	SHA-1 MD5
		메시지요약 (PREMIS:messageDigest)		1:1	2e2107b67a1924419563
	사이즈 (PREMIS:size)			0:1	2038937
	포맷 (PREMIS:format)	포맷지시자 (PREMIS:format Designation)	포맷명 (PREMIS:formatName)	1:1	PDF JPEG GIF HTML
			포맷버전 (PREMIS:formatVersion)	0:1	6
	생성 애플리케이션 (PREMIS:creating Application)	생성 애플리케이션명 (PREMIS:creating ApplicationName)		0:1	MSWord
		생성 애플리케이션버전 (PREMIS:creating ApplicationVersion)		0:1	2003
		생성일 (PREMIS:dateCreated ByApplication)		0:1	2003-03-19T12:25:14-5:00
시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)	소프트웨어 버전 (PREMIS:swVersion)		0:M	6
		소프트웨어 유형 (PREMIS:swType)		1:1	
		소프트웨어 기타 정보 (PREMIS:swOther Information)		0:M	Install acroread(PREMIS: Adobe Acrobat)
		소프트웨어 종속성 (PREMIS:swDependencecy)		0:M	GNU gcc >= 2.7.2
	하드웨어 (PREMIS:hardware)	하드웨어명 (PREMIS:hwName)		1:1	Intel Pentium III 1 GB DRAM Windows XP-compatible joystick
		하드웨어 유형 (PREMIS:hwType)		1:1	
		하드웨어 기타 정보 (PREMIS:hwOther Information)		1:M	32MB minimum Required RAM for Apache is unknown

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
관계 (PREMIS: relationship)	관계유형 (PREMIS: relationshipType)			1:1	structural
	관계하위유형 (PREMIS:relationshipSubType)			1:1	is part of
	연관객체식별 (PREMIS: relatedObjectIdentification)	연관객체 식별자유형 (PREMIS:related ObjectIdentifier Type)		1:1	객체 식별자 예시 참조
		연관객체 식별자값 (PREMIS:related Object IdentifierValue)		1:1	객체 식별자 예시 참조
		연관객체 순서 (PREMIS:related ObjectSequence)		0:1	객체 식별자 예시 참조
	연관이벤트식별 (PREMIS: relatedEvent Identification)	연관이벤트 식별자 유형 (PREMIS: relatedEvent IdentifierType)		1:1	이벤트 예시 참조
		연관이벤트 식별자값 (PREMIS: relatedEvent IdentifierValue)		1:1	이벤트 예시 참조
		연관이벤트 순서 (PREMIS: relatedEvent Sequence)		0:1	이벤트 예시 참조
연결이벤트 식별자 (PREMIS: inkingEventIdenti fier)	연결이벤트 식별자유형 (PREMIS:linking Event IdentifierType)			1:1	이벤트 식별자 유형 예시 참조
	연결이벤트 식별자값 (PREMIS:linking Event IdentifierValue)			1:1	이벤트 식별자 값 예시 참조
연결지적개체 식별자 (PREMIS: inkingIntellectual Entity Identifier)	연결지적개체 식별자 유형 (PREMIS:linking Intellectual IdentifierType)			1:1	지적개체 식별자 유형 예시 참조
	연결지적개체 식별자 값 (PREMIS:linking Intellectual IdentifierValue)			1:1	지적개체 식별자 값 예시 참조
연결권한 선언 식별자 (PREMIS: inkingRightsStat ement Identifier)	연결권한선언 식별자 유형 (PREMIS:linking RightsStatement IdentifierType)			1:1	권한식별자 유형 예시와 동일
	연결권한선언 식별자 값 (PREMIS:linkingRights StatementIdentifierValue)			1:1	권한 식별자 값 예시 참조

[표 3-2] 국립중앙도서관의 PREMIS 보존 메타데이터 - 이벤트 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
이벤트 식별자 (PREMIS:event Identifier)	이벤트식별자유형 (PREMIS:eventIdentifier Type)			1:1	Repository Event
	이벤트식별자값 (PREMIS:eventIdentifier Value)			1:1	e-yyyyymmdd-0001
이벤트 유형 (PREMIS:event Type)				1:1	통제어휘(불변성 체크, 메시지 요약, 파일변환, 검증, 바이러스 체크 등)
이벤트일시 (PREMIS:event DateTime)				1:1	20050704T071530-05 0
이벤트상세사항 (PREMIS:event Detail)				0:1	ISO 8601(2003-03-19T12 :25:14-05:00)
이벤트 결과 정보 (PREMIS:event Outcome Information)	이벤트 결과 (PREMIS:eventOutcome)			0:1	통제어휘 (00, 01, 02 등)
	이벤트결과 상세사항 (PREMIS:eventOutcome Detail)	이벤트결과상세정보 (eventOutcome DetailNote)		0:M (상세 사항) 0:NR (상세 정보)	헤더에 비표준 태그 발견
연결 행위자 식별자 (PREMIS:linking AgentIdentifier)	연결 행위자 식별자 유형 (PREMIS:linkingAgent IdentifierType)			1:1	행위자 식별자 유형 예시 참조
	연결 행위자 식별자 값 (PREMIS:linkingAgent IdentifierValue)			1:1	행위자 식별자 값 예시 참조
연결 객체 식별자 (PREMIS:linking ObjectIdentifier)	연결 객체 식별자 유형 (PREMIS:linkingObject IdentifierType)			1:1	객체 식별자 유형 예시 참조
	연결 객체 식별자 값 (PREMIS:linkingObject IdentifierValue)			1:1	객체 식별자 값 예시 참조

[표 3-3] 국립중앙도서관의 PREMIS 보존 메타데이터 - 행위자 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
행위자식별자 (PREMIS:agent Identifier)	행위자식별자 유형 (PREMIS:agent Identifier Type)			1:1	LCNAF SAN MARC 기관 코드 URI
	행위자식별자값 (PREMIS:agent Identifier Value)			1:1	92-79971 Owens, Erik C. MH-CS /a000001 info:lccn/n78890351
행위자명 (PREMIS:agent Name)				0:M	박성환 국립중앙도서관 FDA Virus Check Service

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
연결이벤트식별자 (PREMIS: inkingEvent Identifier)	연결이벤트식별자유형 (PREMIS:linking Event IdentifierType)			1:1	이벤트 식별자 유형 예시 참조
	연결이벤트식별자값 (PREMIS:linking Event IdentifierValue)			1:1	이벤트 식별자 값 예시 참조
연결권한선언식별자 (PREMIS: inkingRights Statement Identifier)	연결권한선언식별자유형 (PREMIS:linking Rights Statement IdentifierType)			1:1	권한선언 식별자 유형 예시 참조
	연결권한선언식별자값 (PREMIS: linkingRights Statement IdentifierValue)			1:1	권한선언 식별자 값 예시 참조

[표 3-4] 국립중앙도서관의 PREMIS 보존 메타데이터 - 권한 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
권한선언식별자 (PREMIS:rights StatementIdentifier)	권한문식별자 유형 (PREMIS: rightstatement IdentifierType)			1:1	
	권한문식별자 값 (PREMIS: rightsStatement IdentifierValue)			1:1	URI 로컬식별자
권한근거 (PREMIS:rightsB asic)				1:1	통제어휘(저작권, 라이선스, 법령)
저작권정보 (PREMIS: copyright Information)	저작권상태 (PREMIS:copyrightStatus)			1:1	통제어휘(미상, 저작권, 저작권만료, 공개)
	저작권적용영역 (PREMIS:copyright Jurisdiction)			1:1	ISO 3166 (410 또는 KOR 또는 KR)
	저작권발효일 (PREMIS:copyrightStatus DeterminationDate)			0:1	20111011
	저작권설명 (PREMIS:copyrightNote)			0:M	저작원은 2010년에 만료될 예정 저작권에 대한 선언문은 파일 헤더에 포함되어 있음
라이선스정보 (PREMIS: license Information)	라이선스식별자 (PREMIS:license Identifier)	라이선스식별자유형 (PREMIS:license IdentifierType)		1:1	URI 로컬식별자
		라이선스식별자값 (PREMIS:license IdentifierValue)		1:1	
	라이선스약관 (PREMIS:license Terms)			0:1	
	라이선스설명 (PREMIS:license Note)			0:M	라이선스는 파일헤더의 XMP 블록에 포함되어 있음

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
법규정보 (PREMIS: statue Information)	법규관할구역 (PREMIS:statue Jurisdiction)			1:1	us de
	법규인용 (PREMIS:statue Citation)			1:1	도서관법 장애인법 National Library of New Zealand
	법규정보결정일 (statueInformation DeterminationDate)			0:1	2007-12-01 ISO 8601(2003-03-19T12: 25:14-05:00)
	법규설명 (statuteNote)			0:M	
연결객체식별자 (PREMIS: inkingObject Identifier)	연결객체식별자유형 (PREMIS:linking Object IdentifierType)			1:1	객체 식별자 유형 예시 참조
	연결객체식별자값 (PREMIS:linking Object IdentifierValue)			1:1	객체 식별자 값 예시 참조
연결행위자 식별자 (PREMIS: linkingAgent Identifier)	연결행위자식별자 유형 (PREMIS:linking Agent IdentifierType)			1:1	객체 행위자 유형 예시 참조
	연결행위자식별자 값 (PREMIS:linking Agent IdentifierValue)			1:1	객체 행위자 값 예시 참조

□ 국립중앙도서관 보존 메타데이터 요구사항 분석

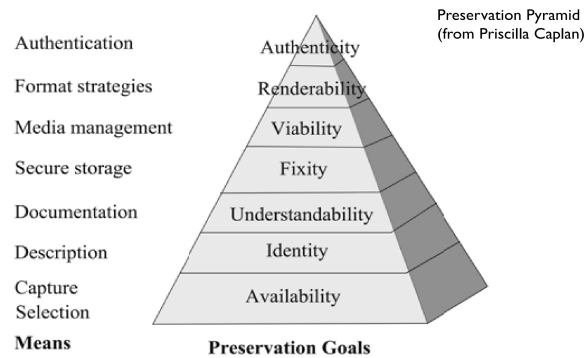
- 2011년 설계된 국립중앙도서관의 PREMIS 기반 보존 메타데이터는 KOLIS에 적용되어 국립중앙도서관의 디지털자원의 기술 메타데이터 표준인 MODS의 기본 정보와 연계가 가능하도록 설계되었다. 현재 국립중앙도서관 보존용 메타데이터의 사용 현황을 조사하고 문제점을 파악하고자 실무담당자와 면담을 수행하였다. 면담의 주요 내용을 정리하면 다음과 같다.
- KOLIS II에서 연계된 PREMIS 기반 보존 메타데이터는 객체(object), 이벤트(event), 행위자(agent), 권한(rights) 범주로 나뉘어 각 개체와 관련한 의미단위와 데이터 모델이 상위, 하위요소들로 구성되어 있어 메타데이터의 구조가 매우 복잡하다. 각 개체의 범주에 따른 보존 속성 정보를 충분하고 완전하게 기술할 수 있도록 하위요소를 확장하여 세부 사항들을 입력하도록 되어 있는데 대부분의 요소가 필수 값으로 지정되어 있어 실질적으로 적용이 쉽지 않다. 핵심 요소들로 이루어진 보다 심플

한 버전이 필요하다.

- 2016년 10월 통합자료관리시스템(KOLIS II)은 웹 기반으로 자료관리시스템의 기능 확대와 함께 디지털장서관리시스템(Digital Contents Management System: DCMS)의 세부 기능을 통합하여 디지털자원 입수방법에 따른 수집 및 관리·보존하는 온라인 납본 기능을 구축하였다. 온라인 납본으로 수집된 자료는 MODS를 기반으로 메타데이터가 입력된 후, KOLIS로 자동이관 돼 PREMIS 보존 메타데이터와 연계되어 있다. 납본 제출 시, 장기적 보존에 있어서 디지털 객체의 무결성과 관련한 PREMIS 메타데이터 정보를 자동으로 추출하여 입력되는 방안을 모색할 필요가 있다.
- 현재 국립중앙도서관의 PREMIS 기반 보존 메타데이터 요소명이 직관적으로 이해하기 쉽지 않으며 업무에 참고할 수 있는 구체적인 메타데이터 사용 지침 및 매뉴얼 등이 요구된다.

□ OAIS & SPOT 분석을 통한 보존 메타데이터 핵심 요소 검토

- 디지털 보존 메타데이터는 디지털 객체의 무결성(integrity)을 유지해야 하고, 수록된 내용이나 형식이 임의적으로 변경 되는 것을 방지할 수 있어야 한다. 이러한 측면에서 보면, 디지털 객체의 보존은 정보환경의 변화에 상관없이 해당 객체를 지속적으로 유지, 관리하고 이를 활용할 수 있도록 지원하는 것으로, 이를 효율적으로 수행하기 위해서는 보존 메타데이터가 매우 중요하다.
- 현재 국립중앙도서관의 PREMIS 메타데이터의 복잡한 구조를 개선하기 위해서 Dappert 등(2016)이 제시한 보존 메타데이터 응용프로파일 설계 방법론에 따라 요구 분석을 할 필요가 있다.
- SPOT(Simple Property Oriented Threat)은 뉴질랜드 통계청, OCLC, 그리고 플로리다 대학에서 디지털자료의 보존을 위해 개발된 것으로, 위협평가를 위하여 사용성, 정체성, 지속성, 렌더링 가능성, 이해가능성, 진본성이라는 6가지 필수 속성을 설정하고 속성별로 구체적인 위협요소를 제시하고 있다([그림 3-7], [그림 3-8] 참조).



[그림 3-7] SPOT(Simple Property-Oriented Threat) 분석
(〈출처〉 Dappert, et al. 2016, 15)

필수 속성	위험 요소
사용성 (Availability)	시간흐름에 따른 디지털객체의 가치와 사용성 약화
	디지털객체의 소멸 가능성
	소유자의 접근 거부
	디지털객체의 부분적인 보존
정체성 (Identity)	불충분한 메타데이터
	메타데이터와 실제 객체와의 연관성 부족
	메타데이터의 사용성 부족
지속성 (Persistence)	부정확한 제어
	저장소 수명 종료
	매체에 대한 독해 도구 사용불가
	매체와 비트시퀀스에 대한 악의적 손상
	매체와 비트시퀀스의 부주의에 의한 손상
렌더링 가능성 (Renderability)	하드웨어와 소프트웨어의 부적절한 조합
	하드웨어와 소프트웨어의 조합이 운영되거나 관리될 수 없음.
	정확한 렌더링 환경이 파악되지 않음
이해가능성 (Understandability)	렌더링 대상의 정확한 특성이 파악되지 않음
	사용자 관심 미 고려
	사용자에 대한 정보제공 부족
	정보의 전체 표현성 미확보에 의한 향후 이해불가능성
진본성 (Authenticity)	추가 정보의 표현성 연계 손상
	객체의 메타데이터와 문서화 부족
	메타데이터의 기술오류
	변화한 디지털객체에 대한 기록 부재

[그림 3-8] SPOT 기반 보존 위험 평가
(〈출처〉 국가기록원 2016, 74)

- OAIS 표준과 SPOT 표준에 기반한 보존 위험 분석과 요구분석을 통해 도출된 국립중앙도서관의 보존 메타데이터의 핵심요소는 [표 3-5]와 같다. 객체식별자, 객체 특성, 관계, 행위자식별자, 이벤트식별자, 이벤트유형, 이벤트일시, 권한선언은 보존 메타데이터와 관련된 대부분의 연구에서 공통적으로 제안하고 있는 요소들로 전반적인 보존 단계에서 반드시 필수적으로 적용되어야 하는 요소로 볼 수 있다.

[표 3-5] SPOT, OAIS 분석을 통한 보존 핵심 메타데이터 요소 제시

SPOT 속성	OAIS 정보	PREMIS 개체/ 의미단위	국립중앙도서관 보존 메타데이터 요소(안)
사용성	참조 정보 맥락 정보 접근권한 정보	권한 개체, 권한 관련 행위자 객체 개체: ObjectIdentifier, storage, preservationLevel	객체식별자 권한선언 행위자식별자
정체성	참조 정보	객체 개체: ObjectIdentifier, originalName	객체식별자
지속성	무결성 정보	객체 형태를 변화시키거나 무결성을 체크하는 이벤트 (예: 무결성 체크, 포맷과 매체 마이그레이션, 위 이벤트 관련 행위자)	이벤트 식별자 이벤트유형 이벤트 일시
	참조 정보 출처 정보	객체 개체: preservationLevel, storage, fixity, size, format 환경 정보	객체특성
렌더링 가능성	구조 또는 표현정보 출처 정보	객체 개체: preservationLevel, format, environment, creatingApplication, structural relationships, objectCharacteristics 객체의 포맷을 변화시키는 이벤트 보존 행위를 수행하는 권한	객체특성 이벤트 식별자 이벤트유형 이벤트 일시
이해 가능성	의미 표현정보 변환 정보 속성	객체: 관련 지적 개체에 연결하는 relatedObjectIdentifier, significantProperties 지적 개체의 외부 기술메타데이터*	관계 객체특성 *MODS와 연계
진본성	무결성 정보	객체 개체: objectCharacteristics (예: size, fixity, signatureInformation)	객체특성

- 사용성과 정체성 측면에서 시간경과에 따른 디지털 객체의 가치를 판단하고 지속적으로 사용가능할 수 있도록 디지털 객체의 생성에서부터 물리적 형태의 변화에 이

르기까지 보존의 대상이 되는 디지털 객체에 대한 포괄적인 기술이 필요하다. 또한 진본성 측면에서 보존된 디지털 객체의 무결성을 확인할 수 있도록 이와 관련된 정보 또한 충분히 기술해 주어야 한다. 이를 위해 객체 개체 메타데이터 요소와 관련하여 객체식별자와 무결성, 포맷 등의 객체 특성이 필수적으로 제공되어야 한다.

- 사용성 측면에서 보존된 디지털 객체를 현재 혹은 향후에 활용하는 것과 관련된 사항 및 해당 객체의 저작권 관련 사항을 포괄적으로 기술하여야 한다. 따라서 권한 개체 메타데이터 요소와 관련하여 연결행위자 식별자, 연결객체 식별자와 같은 권한선언이 필요하다.
- 지속성과 렌더링 가능성 측면에서 디지털 객체의 보존과 관련된 제반 활동을 포괄적으로 기술해 주며 보존활동의 결과로 발생한 사항에 대해서도 상세하게 기술해 주어야 한다. 따라서 이벤트 개체 메타데이터 요소와 관련하여 객체 형태를 변화시키거나 무결성을 체크하는 무결성정보 등 보존행위에 대한 이벤트식별자, 이벤트 유형, 이벤트일시 정보가 필수적으로 제공되어야 한다.
- 이해가능성 측면에서 보존의 대상이 되는 디지털 객체와 관련된 다른 객체에 대한 관계를 기술할 수 있어야 하며, 이를 위해 관계가 필수적으로 제공되어야 한다.

□ PREMIS 3.0 기반 메타데이터 상호운용성 분석

- 국립중앙도서관에서 납본, 수집, 구독, 연계 등 다양한 방법으로 입수되는 디지털자료의 장기적 보존을 위해서 입수, 보존 및 접근 등 전반적인 단계에 걸쳐 적용할 수 있는 메타데이터 요소가 필요하다.
- PREMIS 3.0은 PREMIS의 실행성과 상호운용성에 초점을 맞추어 개정되었으며, 보존 생애주기를 반영하고 있다는 것이 가장 큰 특징이다. PREMIS 보존 메타데이터는 디지털 객체에 관한 포괄적인 기술사항을 기반으로 하고 있기 때문에, 기존의 국립중앙도서관의 다양한 디지털자료의 서지레코드에 활용되는 메타데이터와의 상호운용성을 유지함으로써 보존 메타데이터의 생성 활용에 있어서의 효율성을 확보할 수 있다.

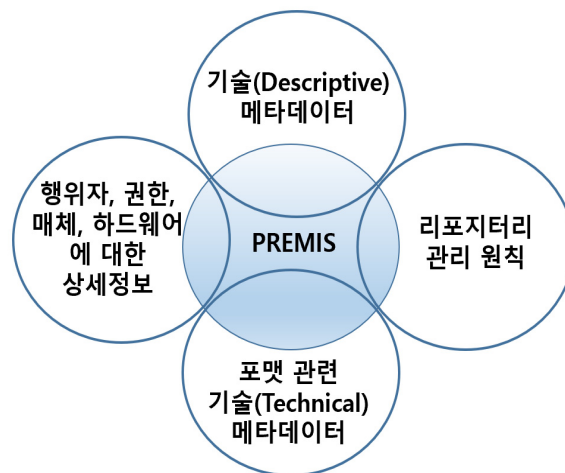
- PREMIS 3.0과 생애주기 전반에 걸쳐 적용되고 있는 서지 메타데이터 MODS와 함께, OASIS 웹 아카이빙에 적용되고 있는 웹자원 저장 포맷의 표준인 WARC의 기술요소를 비교, 분석하여 도출된 국립중앙도서관의 보존 메타데이터의 요소는 [표 3-6]과 같다.
- 제안하는 보존 메타데이터 개선(안)에서는 객체식별자유형, 객체식별자값, 객체범주, 객체특성, 연결객체식별, 이벤트일시, 행위자식별자유형, 행위식별자값, 행위자명, 권한식별자값, 권한근거 등의 요소들이 자원의 생애주기 동안 적용되는 메타데이터들과 연계하여 자동으로 추출하도록 하였다.
- 보존 메타데이터 개선(안)에서 위의 요소들은 자동으로 추출, 필수 항목으로 입력됨으로써, 국립중앙도서관의 보존 자원 생애주기 전반에 걸쳐 서지업무에서 있어서 불필요한 중복 업무를 줄이고 다른 메타데이터와의 상호운용성을 확보할 수 있도록 하였다.

[표 3-6] 디지털자원의 보존 생애주기 메타데이터 비교 분석

PREMIS 3.0		MODS 요소	OASIS 저장포맷 WARC 기술 요소	국립중앙도서관 보존 메타데이터 요소(안)
개체	의미단위			
객체 Object	1.1 objectIdentifier	식별기호		객체식별자유형 객체식별자값
	1.2 objectCategory	식별기호		객체범주
	1.5 objectCharacteristics	콘텐츠유형 형태정보	Format	객체특성
	1.14 linkingObjectIdentifier	연관정보		연결객체식별
이벤트 Event	2.1 eventIdentifier 2.2 eventType 2.3 eventDateTime		Date	이벤트일시
행위자 Agent	3.1 agentIdentifier		Operator	행위자식별자유형 행위자식별자 값 행위자명
권한 Rights	4.1.1 RightsStatement Identifier	접근제한정보		권한문식별자값 권한근거

3.1.4 국립중앙도서관 보존용 메타데이터 개선 방안

- 현재 국립중앙도서관에서 디지털자원에 적용하고 있는 메타데이터 MODS, METS, PREMIS, 더블링크어 등을 검토하여 중복 요소가 있는지 분석하고 이들 보존 생애 주기의 메타데이터들을 연계함으로써 추출 가능한 요소를 확인하고, 핵심요소들로만 구성된 PREMIS 요약 버전을 제안하고자 하였다.



[그림 3-9] PREMIS 보존 메타데이터

□ 디지털자원 생애주기 기반에 따른 보존 메타데이터 적용범위

1) 디지털자원 입수시 보존 메타데이터 자동 추출 및 입력

- 국립중앙도서관에 OASIS 웹 아카이빙, 온라인 납본, 디지털화 등을 통하여 디지털 자원이 입수되면 보존 메타데이터는 자동적으로 추출되어 입력되어야 한다.
- 디지털자원이 국립중앙도서관에 최초 입수되는 시점에서 디지털자원의 데이터 무결성과 포맷 관련 정보는 기술 메타데이터 MODS에 기술되어 있는 형태정보와 웹 자원 저장 포맷의 표준인 WARC에 기술되어 있는 기술 요소들을 바탕으로 연계하여 자동으로 입력되도록 한다.

- 제안하는 보존 메타데이터 개선(안)에서 OASIS 웹 아카이빙 자원의 경우, WARC 파일 포맷에는 정의된 필드(Named-field)와 코드 형식(Record Types)으로 저장되어 있는 정보가 보존 메타데이터의 객체 특성정보로 연계되어 자동, 추출되도록 하였다. 정의된 필드의 콘텐츠 형식(Content-Type)은 MIME 타입으로 정의된 콘텐츠 형식으로 WARC는 기본적으로 표준화된 Application/warc-fields가 저장된다. 콘텐츠 크기(Content-Length)에서는 해당 WARC에 저장된 콘텐츠의 크기가 저장된다.
- 또한 온라인 납본과 디지털화를 통해 입수된 디지털자원은 기술 메타데이터 MODS로 부터 보존 메타데이터의 객체 정보로 연계되어 식별기호, 콘텐츠유형, 형태정보 등이 메타데이터 요소들이 자동, 추출되도록 하였다.
- 자동추출, 입력되는 메타데이터 요소: 객체식별자유형, 객체식별자값, 객체범주, 객체특성, 연결객체식별, 이벤트일시, 행위자식별자유형, 행위식별자값, 행위자명, 권한식별자값, 권한근거

2) 수집 이후 디지털자원의 장기보존을 위한 변화 이력정보 기록

- 디지털자원이 국립중앙도서관에 입수된 이후 디지털자원의 무결성, 이용가능성 보장을 위해서 자원 상태에 영향을 미치는 모든 관리, 보존처리에 따른 이력 정보가 메타데이터에 기록되어야 한다.
- 장기보존 관점에서 디지털자원의 생애주기 동안, 매체변환, 접근 및 이용권한의 변경 등과 같은 보존 관련 행위가 발생할 때 무슨 일이 일어났는지, 언제 발생했는지, 왜 발생했는지, 그리고 누가 수행했는지 등 발생한 모든 행위 이력에 대한 정보가 보존 메타데이터에 저장된다.
- 제안하는 보존 메타데이터 개선(안)에서는 디지털자원의 입수 후, 보존관리 이력과 관련하여 이벤트 개체(이벤트 식별자, 이벤트유형)와 함께 객체(관계), 행위자(연결 이벤트 식별자, 연결권한선언 식별자), 권한(권한문식별자 유형)요소를 필수값으로 제안함으로써, 보존행위를 수행한 행위주체와 행위유형, 행위일자가 명확하게 결합할 수 있도록 하였다.

- 수동으로 추가, 입력되는 메타데이터 요소: 관계유형, 관계하위유형, 연결이벤트식별자, 연결권한식별자, 이벤트식별자, 이벤트유형, 이벤트상세사항, 이벤트결과정보, 연결행위자식별자, 연결객체식별자, 연결이벤트식별자, 연결권한선언식별자, 권한확장

□ 국립중앙도서관 보존 메타데이터 개선(안)의 요소 구성

- 제안하는 국립중앙도서관의 보존 메타데이터 개선(안)은 기존의 보존 메타데이터와 같이 PREMIS의 객체, 이벤트, 행위자, 권한 개체로 구성되며, PREMIS 3.0의 개정된 데이터 사전의 상위 요소들을 중심으로 재구성되었다.
- 그 중 일부는 필수적인 요소로 구분되어 있다. 또한 일부 요소는 연계된 다른 기술 메타데이터로부터 자동으로 추출, 저장되지만 다른 요소는 관리자에 의해 추가되어야 하는 요소이다.
- 앞서 제시한 바와 같이, 데이터 무결성과 포맷 관련 정보는 기술 메타데이터 MODS에 기술되어 있는 형태정보와 웹기록물 저장 포맷의 표준인 WARC에 기술되어 있는 기술 요소들을 기준으로 자동 추출된다.
- 아래의 [표 3-7]에서 제시한 국립중앙도서관의 PREMIS 기반 보존 메타데이터 개선(안)은 보존 단계에서 신뢰할 수 있는 디지털자원의 지속적 접근을 위한 메타데이터 패키지를 구성하는 기준으로 사용될 수 있으며, 세부적인 측면과 관련된 메타데이터 요소들을 결합해 주는 상위요소(wrapper element)로서의 기능을 수행할 수 있다.

[표 3-7] 국립중앙도서관의 PREMIS 기반 보존 메타데이터 개선(안)

1st level	2nd level	3rd level	필수: 반복	자동 여부	연계 외부 메타데이터 요소	입력 예시
객체	객체식별자	객체식별자유형	1:NR	자동	MODS: 식별기호	
		객체식별자 값	1:NR	자동	MODS: 식별기호	
	객체범주		1:NR	자동	MODS: 콘텐츠유형	
	객체특성	포맷	1:R	자동	WarcInfo: Format MODS: 형태정보	
	환경		0:R			
	관계	관계유형	1:NR	수동		structural
		관계하위유형	1:NR	수동		is part of
		연관객체식별	1:R	자동	MODS: 연관정보	
	연결이벤트 식별자		0:R	수동		
	연결권한 식별자		0:R	수동		
이벤트	이벤트 식별자	이벤트식별자유형	1:NR	수동		Repository Event
		이벤트식별자 값	1:NR	수동		e-yyyyymmdd-0001
	이벤트유형		1:NR	수동		무결성체크, 파일변환, 검증, 바이러스 체크 등
	이벤트일시		1:NR	자동	WARCInfo: Date	
	이벤트상세사항		0:R	수동		
	이벤트 결과 정보		0:R	수동		
	연결 행위자 식별자		0:R	수동		
	연결 객체 식별자		0:R	수동		
행위자	행위자식별자	행위자식별자 유형	1:NR	자동	WARCInfo: Operator	
		행위식별자 값	1:NR	자동	WARCInfo: Operator	
	행위자명		0:R	자동	WARCInfo: Operator	
	연결 이벤트 식별자		0:R	수동		
	연결권한선언 식별자		0:R	수동		
권한	권한선언	권한문식별자 유형	1:NR	수동		
		권한문식별자 값	1:NR	자동	MODS: 접근제한정보	
		권한근거	1:NR	자동	MODS: 접근제한정보	
	권한 확장		0:R	수동		

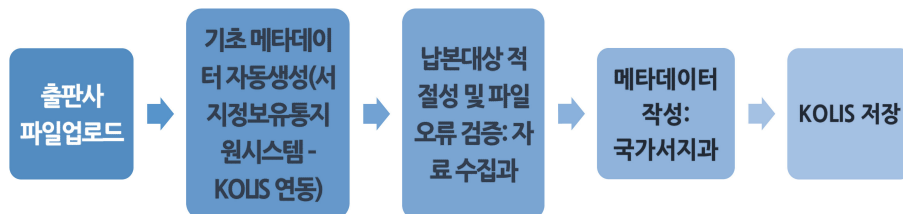
3.2 디지털자원 관련 업무 분석 및 KOLIS II 보존 기능

3.2.1 국립중앙도서관 디지털자료 관련 업무 분석

- 국립중앙도서관 실무 담당부서와의 인터뷰를 통하여 디지털 관련 업무를 수행한 결과 다음과 같은 시나리오가 분석되었다.

□ 온라인 납본 관련 업무

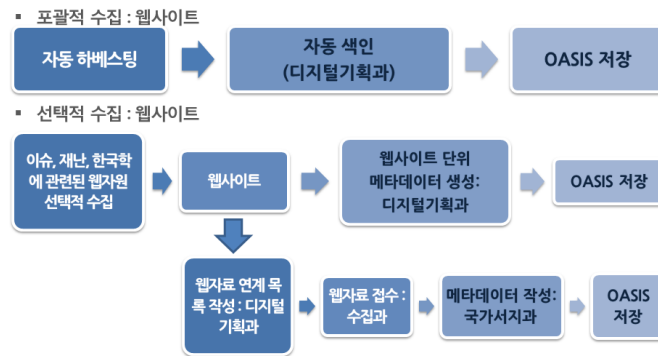
- 온라인 자료 납본 관리 주관 부서는 자료수집과이며, [그림 3-10]과 같이 각 출판사가 서지정보유통지원시스템에 납본대상 파일을 업로드하고 기초 사항을 입력하면 KOLIS II 온라인시스템으로 자동 연동된다.
- 자료수집과에서 자동 연계된 파일 및 기초 메타데이터 오류를 검증한 후 국가서지과로 인계한다.



[그림 3-10] 온라인 납본 업무 프로세스

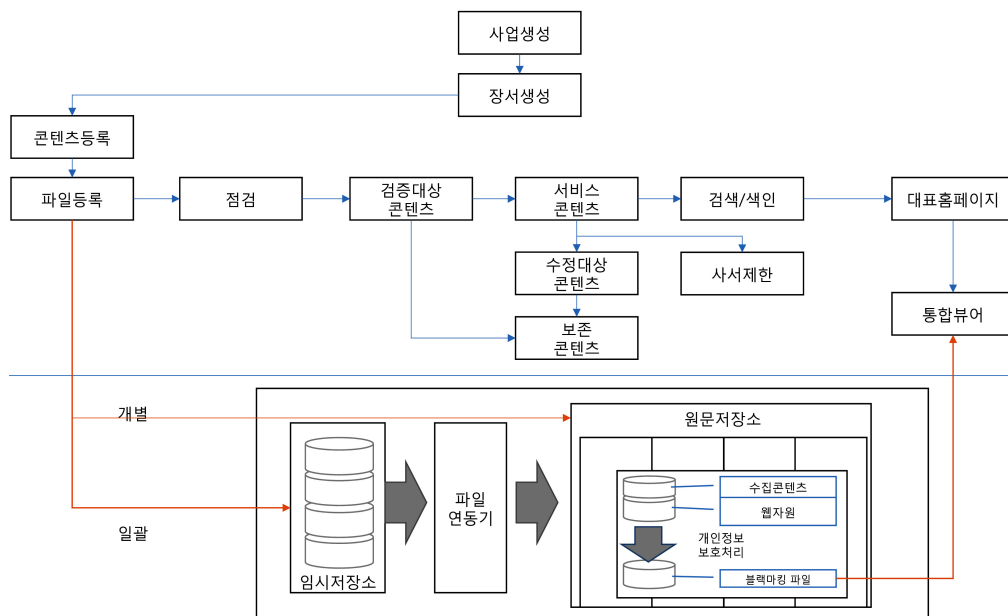
- 시스템 구성상 데이터가 처리되는 과정을 살펴보면, [그림 3-11]과 같이 서지유통시스템에 납본파일이 등록됨과 동시에 서지사항이 서지저장소에 축적되고 원본 파일은 서비스용과 보존용으로 이중으로 저장되는 체제이다.

수집한 웹자료(연속)은 자료수집과에서 자료를 수집하고 파일 오류를 검증한 후 국가서지과에 인계하면 국가서지과에서 세부 메타데이터를 생성하여 등록하는 형태이다.



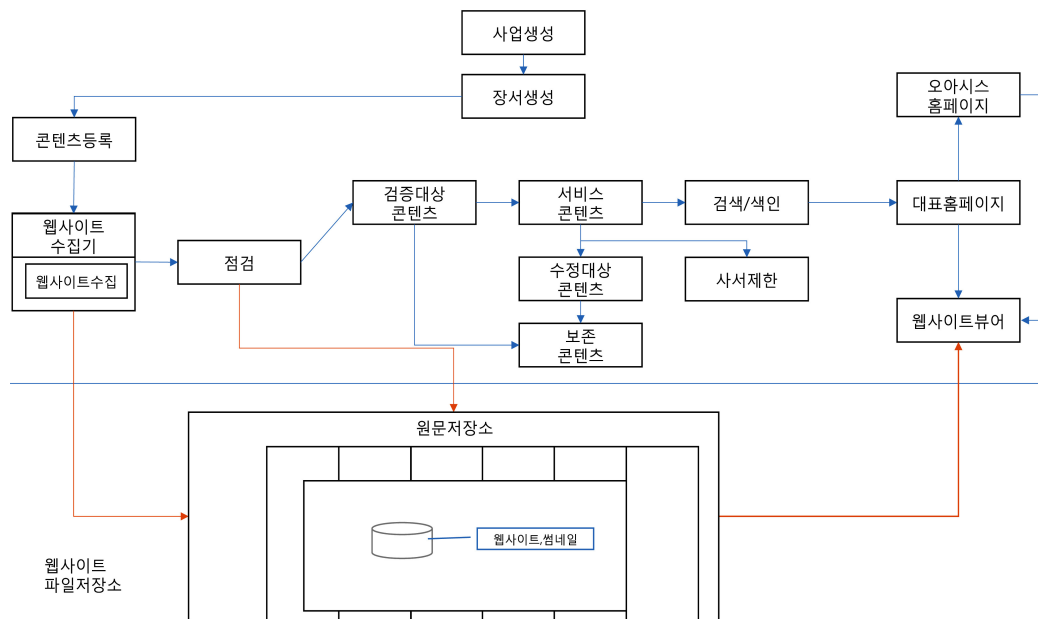
[그림 3-12] 웹사이트 및 웹자료 처리 업무 프로세스

- 시스템 구성상 웹 정보자료가 처리되는 과정을 살펴보면, [그림 3-13]과 같이 파일 단위의 웹자료는 온라인 납본과 동일하게 보존용과 서비스용으로 처리되며 서비스는 오아시스 홈페이지와 국립중앙도서관 홈페이지에서 제공하고 있다.



[그림 3-13] 웹자료 처리 시스템 흐름도

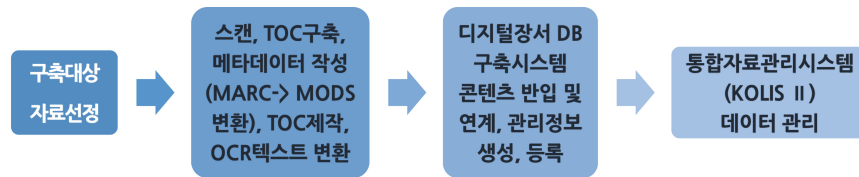
- 웹사이트는 [그림 3-14]와 같이 하베스팅 수집기를 통해 수집되어 KOLIS II 온라인시스템에 저장, 관리된다. 서비스는 오아시스 홈페이지를 통해 제공되고 있다.



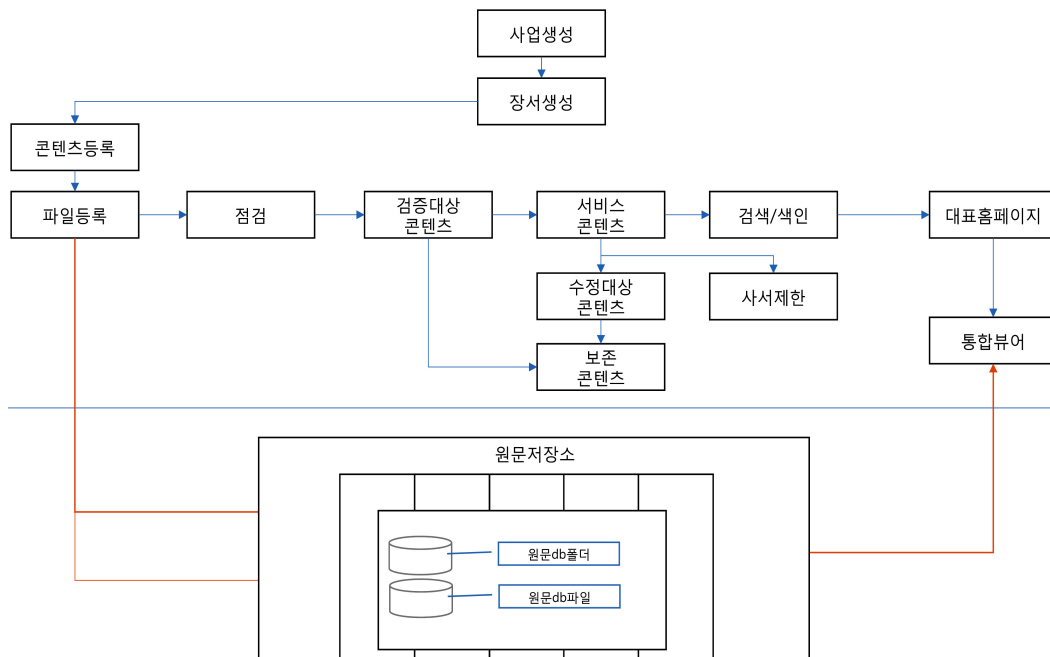
[그림 3-14] 웹사이트 처리 시스템 흐름도

□ 디지털화 관련 업무

- 디지털화 관련 업무 주 담당 부서는 디지털기획과이며 [그림 3-15]와 같은 프로세스로 이루어지는데 자료 선정은 기획과에서 담당하고 변환과 메타데이터는 외주 업체와 협력하여 진행하고 있다.
- 시스템 구성상 디지털화된 정보자료가 처리되는 과정을 살펴보면 파일 단위의 웹 자료와 유사하다. [그림 3-16]과 같이 디지털화된 정보자료는 디지털장서DB구축 시스템에 자료가 개별적으로 업로드되어 등록이 확정되면 KOLIS II 온라인시스템에 등록된다. 디지털화된 정보자료는 서비스용과 구분하여 보존용을 별도로 생성하고 있지는 않다.



[그림 3-15] 디지털화 업무 프로세스



[그림 3-16] 디지털화 시스템 흐름도

3.2.2 KOLIS II 보존 기능 개선(안)

□ KOLIS II 보존 기능 현황 분석

1) 통합자료관리시스템(KOLIS II) 온라인

- 오프라인: 도서/비도서 자료를 관리하는 업무로서 수서, 등록, 정리, 배가, 열람 등 5가지 업무로 구성되어 있는 국립중앙도서관의 자료관리시스템으로 메타데이터로

는 통합KORMARC(KORean Machine Readable Cataloging)을 사용한다.

- 온라인: 국립중앙도서관의 온라인 자료 업무관리 시스템으로 모든 유형의 디지털 정보 자원을 입수, 관리, 영구보존하기 위한 시스템. 메타데이터는 MODS(XML 형식)를 사용한다.
- 1998년 오프라인 자료 관리를 위한 통합정보시스템(KOLIS)을 개발 후 2008년에는 온라인 자료 관리를 위해 디지털장서관리시스템을 개발하였다. 그 후, 시스템 고도화 및 확장 개선을 거쳐 2016년에 온/오프라인을 통합한 현재의 통합자료관리시스템으로 모습을 갖추게 되었다. 자세한 개발 이력은 [표 3-8]과 같다.

[표 3-8] KOLIS II 개발 이력

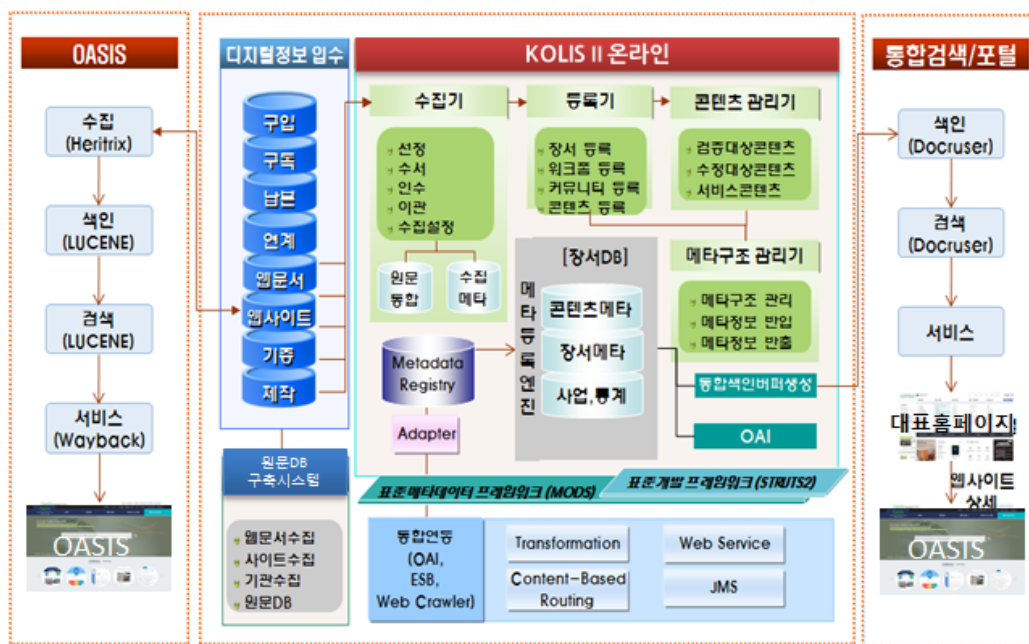
일시	경과내용
2016.12	• 온/오프라인 통합관리를 위해 통합자료관리시스템(KOLIS II)으로 통합
2013.12.	• 온라인 메타데이터 DB구축 시스템 확장개선 • MODS v.3.6 적용 • 보존메타데이터 표준 PREMIS 관리 기능 추가 • 종·권호 관리기능 추가
2010.12.	• 아시아 디지털도서관 프로젝트 추진을 위한 정보시스템 개선으로 시스템 고도화
2008. 6.	• 온라인 자료 관리를 위한 디지털장서관리시스템 개발
1998.12.	• 오프라인 자료 관리를 위한 통합정보시스템(KOLIS) 개발

2) MODS(Metadata Object Description Schema)

- 미국 의회도서관(Library of Congress)의 ‘네트워크 개발 및 MARC 표준 사무국’에서 도서관의 다양한 목적에 사용되도록 표준을 개발하고 유지, 관리한다.
- 디지털자원의 서지정보 표준 메타데이터로 다양한 형태의 자원에 대한 서지레코드를 생성한다.
- MARC와 DC(Dublin Core)를 절충하여 기존의 MARC 레코드를 변환할 수 있도록 설계하여 MARC에서 추출된 데이터 표현이 가능하고 데이터 요소가 간단하나 DC보다 풍부하다. 각 태그가 숫자 형태인 MARC와 달리 누구나 이해 가능한 문자기반 태그를 사용하여 사용이 편리하다.

3) 사업·장서·콘텐츠 관리기능

- KOLIS II의 콘텐츠 관리기능은 [그림 3-17]과 같이, 온라인 콘텐츠 구축을 위한 사업정보를 관리한다. 콘텐츠를 여러 기준으로 구분하여 관리할 수 있도록 장서관리 기능을 제공하고 관리된 콘텐츠를 국립중앙도서관 대표홈페이지로 서비스할 수 있게 한다.



[그림 3-17] KOLIS II 콘텐츠 관리기능

4) 업무관리시스템관리 기능

- KOLIS II의 업무관리시스템은 [그림 3-18]과 같이, 전체 사업장에서의 콘텐츠 관리, 연속간행물 중·권호 관리, 그리고 오아시스 홈페이지를 관리하는 업무관리 기능을 제공한다. 또한 시스템의 MODS XML 항목과 요소가 등록된 메타사전 관리, 시스템 공통코드, 그리고 저장소를 관리하는 시스템관리 기능을 제공한다.



[그림 3-18] KOLIS II 업무관리시스템 관리기능

□ KOLIS II 보존 기능 개선 설계(안)

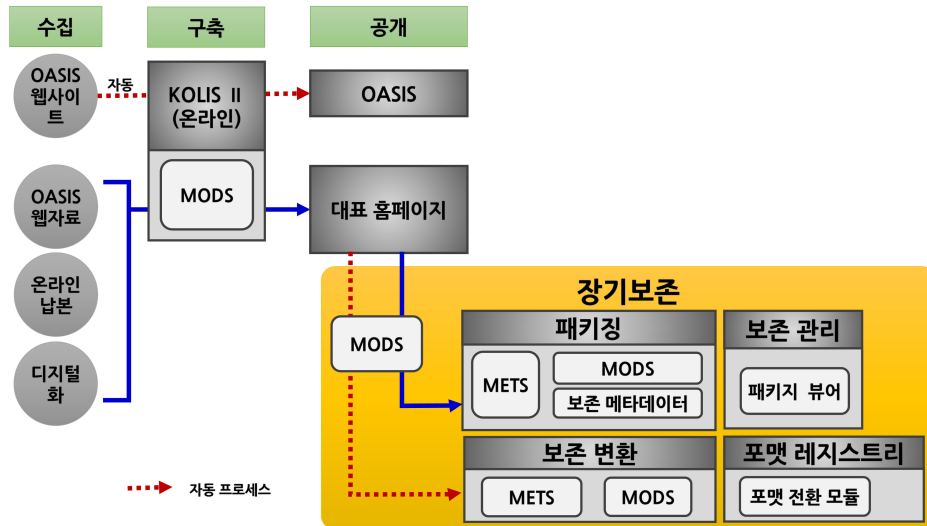
1) KOLIS II 보존 기능 방향성

- 통합자료관리시스템(KOLIS II)은 기존의 오프라인, 온라인 DB 구축시스템을 2016년 통합한 자료관리시스템으로, 2013년 보존 기능을 위해 보존 메타데이터인 표준 PREMIS 관리 기능을 추가하였으나 실제로는 별개로 운영되고 있다.
- 현재 통합자료관리시스템에 보존을 위한 PREMIS 관리 기능을 비롯하여 영구보존을 위한 기능이 일부 구현되어 있으나 장기보존 기능을 기존 통합자료관리시스템에 추가하는 것은 운영 및 관리의 복잡성을 증가시키게 되어 보존 기능의 효율적인 관리를 기대하기 어려울 것으로 예상된다.
- 따라서 KOLIS II와 장기보존시스템은 독립적으로 구축하고 서로 연계하도록 설계하여 확실한 관리 주체에 의해 일관성 있게 운영 및 관리하도록 하는 방향이 바람직한 장기보존 기능 구현을 위한 현실적인 대안이 될 것이다.

2) 장기보존 기능 설계(안)

- 국립중앙도서관의 전체적인 장기보존 기능 개념도는 [그림 3-19]와 같이 구성된다.

기본적으로 자료 입수 경로와 업무를 고려하여 온라인 납본, 웹 아카이빙, 디지털화 등 3가지로 구분하고, 웹 아카이빙은 다시 OASIS 웹사이트 수집과 웹자료 수집으로 구분한다.



[그림 3-19] KOLIS II와 장기보존 개념도

- 장기보존시스템에서 KOLIS II DB를 미러링 형식으로 복제하여 현행 서비스의 오류 복구를 위한 백업 기능 뿐 아니라 장기보존 대상 DB로 활용한다.
- KOLIS II의 메타데이터 DB인 MODS를 추출, 변환하여 METS로 MODS와 보존 메타데이터를 패키징한다.
- 패키지를 확인하기 위한 패키지 뷰어와 보존 관리 기능이 있고 포맷 관리를 위한 포맷 레지스트리로 구성된다.
- 패키징
 - 업무 및 보존 프로세스를 기준으로 웹사이트는 자동으로 수집부터 서비스, 그리고 보존까지 원스톱 워크플로우로 진행할 수 있다. 따라서 보존 메타데이터가 KOLIS II의 MODS DB에서 자동으로 변환되어 구축되지 않는 현재 상황에서는 장기보존시스템에서도 MODS를 추출, 변환하여 METS를 구축한다. OASIS 웹사이트

수집은 스냅샷 단위 웹사이트 보존을 목적으로 하므로 장기보존시스템에서도 이 부분을 반영하도록 한다.

- 온라인 납본, 웹자료, 디지털화 자료는 디지털자원 보존을 위한 보존 메타데이터를 구축하고 콘텐츠 메타데이터인 MODS와 함께 METS를 이용하여 패키징을 한다.
- 보관 저장소는 보존 활동을 지원하는 장기 저장 구성요소이다. 모든 콘텐츠 파일은 여기에 저장된다. 모든 아카이브 메타데이터는 콘텐츠와 연결되며 아카이브 저장소에도 배치된다. 아카이브 저장소의 메타데이터가 운영 액세스에 사용되지는 않지만 아카이브 저장소 자체가 완전함을 보장하기 위해 동일한 시스템 내에서 콘텐츠와 메타데이터를 보관하는 것이 좋다.
- 이 아카이브 메타데이터는 모든 콘텐츠 파일(이미지, 전체 텍스트 파일 등)을 참조하는 PREMIS 및 MODS 구성요소가 있는 METS 파일의 계층 구조로 표시된다. METS 및 콘텐츠 파일 묶음(bundle)은 AIP(Archival Information Package)로 구성된다.
- METS는 다음 3가지 문법을 제공한다.
 - ① 전송 구문(Transfer syntax): 디지털 객체의 전송/교환을 위한 표준으로 OAIS 참조모델의 SIP(Submission Information Package)에 해당한다.
 - ② 기능 구문(Functional syntax): 디지털 콘텐츠 및 관련 메타데이터에 대한 보기 및 내비게이션을 최종 사용자에게 제공하기 위한 기반으로 OAIS 참조모델의 DIP(Dissemination Information Package)에 해당한다.
 - ③ 저장 구문(Archiving syntax): 디지털 객체들을 아카이빙하기 위한 표준으로 OAIS 참조모델의 AIP(Archival Information Package)에 해당한다.
- METS 스키마에는 보존 메타데이터 전용 섹션이 없는 대신 보존 메타데이터를 기술, 디지털 출처, 소스 및 권한 메타데이터로 분할된다. 그러므로 PREMIS에 표현된 보존 메타데이터는 분할되어 이러한 섹션에 분산되어야 한다.

3) 보존 기능 업무별 흐름

- 온라인 납본

- 온라인 납본 자료는 각 출판사가 서지정보유통지원시스템에 납본대상 파일을 업

로드하고 기초 사항을 입력하면 KOLIS II 온라인시스템으로 자동 연동된다.

- 보존용으로 생성된 콘텐츠는 장기보존시스템으로 전송하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.

- 웹 아카이빙: 파일 단위 웹자료

- 파일 단위의 웹자료는 온라인 납본과 동일하게 보존용과 서비스용으로 처리되며 서비스는 오아시스 홈페이지와 국립중앙도서관 홈페이지에서 제공하고 있다.
- 온라인 납본 자료와 마찬가지로 보존용으로 생성된 콘텐츠는 장기보존시스템으로 전송하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.

- 웹 아카이빙: 웹사이트

- 웹사이트는 수집기를 통해 수집된 자료가 자동으로 디지털장서구축시스템을 거쳐 KOLIS II 온라인시스템에 등록된다. 다른 자료와 다르게 OASIS 홈페이지를 통해 서비스되고 있다.
- 웹사이트는 자동으로 수집하여 메타데이터가 생성되어 원문저장소에 저장이 완료되면 특정 시점을 설정하여 연단위로 추가된 자료를 장기보존저장소에 백업 형태로 저장하고 웹사이트 구성요소의 포맷을 포맷 레지스트리에 등록한다.

- 디지털화

- 디지털화된 정보자료는 파일 단위의 웹자료와 유사하게 디지털장서구축시스템에 자료가 개별적으로 업로드되어 등록이 확정되면 KOLIS II 온라인시스템에 등록된다. 파일 단위의 웹자료처럼 보존용이 생성되고 있지 않다.
- 디지털화된 정보자료는 서비스용과 보존용으로 이원화하고 있지 않으나 장기보존을 위해서는 파일 단위의 웹자료 및 납본자료와 유사한 프로세스로 변경하는 것이 바람직하다. 서비스용으로 원문저장소에 저장하여 서비스를 하고 추가로 보존용으로 복제하여 장기보존시스템으로 이관하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.

3.3 디지털자원 포맷 레지스트리

3.3.1 포맷 레지스트리의 필요성

- 일반적으로 종이기록과 달리 기술 의존적 특성을 갖는 전자기록은 하드웨어나 소프트웨어 기술의 노후화에 직접적인 영향을 받기 때문에 디지털자원의 장기적 유지 및 보존을 위해 지속적이고 체계적인 장기보존전략을 통해 기술이 노후화에 대한 계획을 마련해야 한다.
- 마이그레이션, 에뮬레이션, 인캡슐레이션, 변환(conversion), 아날로그 보존, 기술 보존, 리프레쉬먼트(refreshment)와 같은 보존 기법들이 있으나 하나의 보존전략만을 선택하기보다는 자원의 특성과 보존 초점에 따라 적어도 두 개 이상의 다양한 보존 기법을 선택하여 병행하는 것이 효과적이다.
- 어떤 보존전략을 선택하든 자원의 생명주기 동안 반복되어 적용될 것이고, 특정 자원 건에 대한 전략이 어느 시점에 바뀔 수도 있으나 자원에 대한 가독성을 최대한 보장하고, 미래의 어느 시점에서든 어떠한 보존전략도 선택이 가능하도록, 그리고 다양한 보존전략을 적시에 적절하게 적용하기 위해서는 보다 안전한 방법이 필요하다.
- 포맷 레지스트리는 기술보존의 직접적인 방법이자 기술정보요소를 지속적으로 수집 및 관리하여 마이그레이션, 에뮬레이션과 같은 보존전략을 효율적으로 수행할 수 있도록 지원하는 강력한 토대가 되는 개념이라 할 수 있다.
 - 유영수(2006): 전자기록의 장기보존과 장기접근성 유지에 필요한 기술(technical) 정보, 즉 포맷 정보, 그와 관련된 하드웨어 및 소프트웨어 정보, 매체정보를 소장한 일종의 데이터베이스시스템
 - 국가기록원(2008): 특정 디지털정보 파일의 포맷 구문 정보와 의미 정보를 저장하는 일종의 데이터베이스
 - 박지화, 정영미(2015): 특정 전자기록에 사용되는 기술요소 표준을 등록하고 관리하는 포맷의 기술정보요소를 위한 데이터베이스

- 일반적인 보존전략 외에도 포맷 등록 시스템에 대한 연구가 북미와 유럽을 중심으로 국가별 또는 국가 간 협력 프로젝트 형태로 매우 활발하게 전개되고 있는 반면, 국내의 경우는 해외 사례를 소개하는 등의 한정적인 연구가 일부 존재한다.
- 보존 생애주기 전반에 걸쳐 수집, 관리, 보존, 이용 등 단계에서 정보자원의 무결성을 확보하고 지속적인 접근을 위해서는 각 단계의 작업이 유기적으로 이루어질 수 있도록 메타데이터의 통합적 연계가 필요하다. 납본 제출 단계에서는 디지털자원의 하드웨어, 소프트웨어 요소 등 보존에 있어서 중요한 정보가 자동으로 추출, 연계하여 추후 보존 단계에서 정보패키징 변환을 지원하고 관련 기술 메타데이터를 추출하여 레지스트리에 등록 관리할 수 있다

3.3.2 포맷 레지스트리 사례 분석

□ 영국 국가기록원의 PRONOM

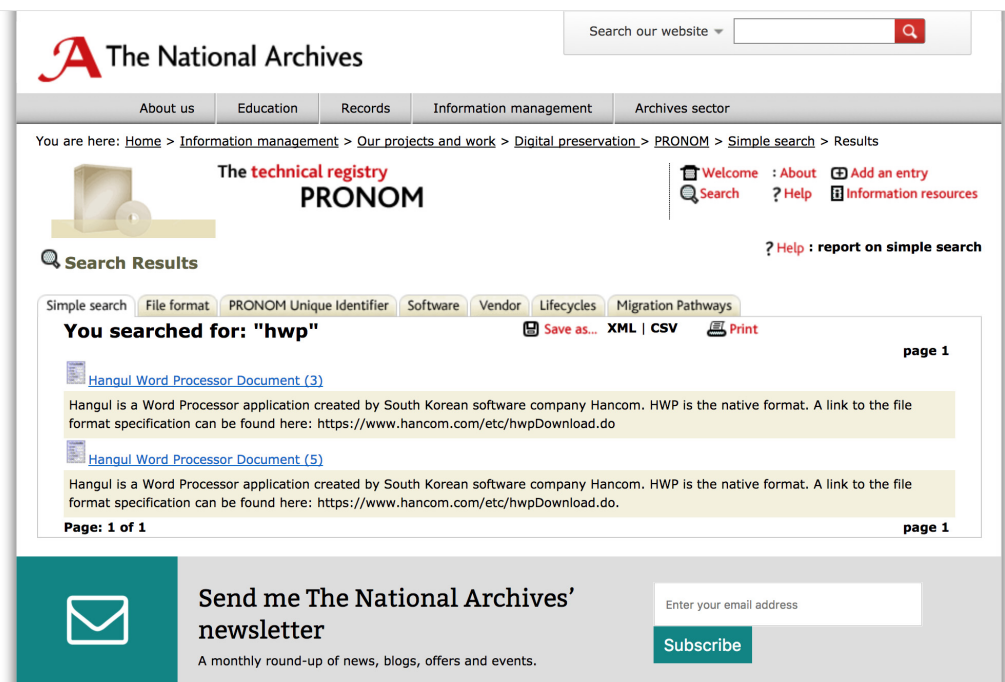
1) 개요

- 영국의 국가기록원(TNA)에서 다양한 포맷의 전자기록을 장기보존하기 위해 개발되었다.
- PRONOM은 데이터 형식 및 지원 소프트웨어 제품에 대한 정보를 제공하는 온라인 정보 시스템으로 디지털 아카이브에 저장되는 전자기록에 대해 신뢰할 수 있는 기술정보에 대한 접근을 제공하기 위해 개발되었고, PRONOM 3부터 웹 버전을 공개하고 있다.
- 현재 제공되는 PRONOM 6.2는 지속적으로 콘텐츠를 수집하여 약 800여개 이상의 포맷 엔트리를 포함하고 있으며, 단순히 기술 데이터베이스가 아니라 보존위험평가, 마이그레이션 경로 계획, 개체 식별 및 검증, 메타데이터 추출 등의 디지털 보존 기능을 지원하는 서비스까지 제공한다.
- PRONOM의 주된 목적은 다음과 같다.
 - 파일 포맷, 소프트웨어 제품, 운영시스템, 하드웨어 구성요소 등 디지털자원의 기

술의존성에 대한 정보 리포지토리를 포함하는 기능

- 파일 포맷의 식별 및 검증
- 메타데이터의 추출과 마이그레이션을 위한 자동화된 툴의 지원
- 기술 노후화나 지원되지 않는 제품에 대한 자동 보고 시스템을 통한 기술 감시지원
- 포맷 간 마이그레이션 경로의 자동생성을 통한 마이그레이션의 지원
- 데이터의 교환과 자동화 서비스 대비를 위한 목적을 위해 다른 시스템과의 상호 작용을 지원

- 최근에는 PRONOM의 링크드 데이터(Linked Data) 버전이 개발되어 PRONOM 데이터의 공유 및 연결을 가능하게 하고 있다.



[그림 3-20] PRONOM 이용 예시화면 - 검색어 “HWP”
(〈출처〉 <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)

Simple search

File format

PRONOM Unique Identifier

Software

Vendor

Lifecycles

Migration Pathways

Details for: Hangul Word Processor Document 3

Save as...

XML

CSV

Print

Go to: Summary > | Documentation > | Signatures | Compression > | Character encoding > | Rights > | Reference files > Properties >

Signatures

External signatures

File extension: hwp

Internal signatures

Name

HWP 3

Description

BOF: "HWP Document File V3.00 " 0x1A0102030405
https://www.hancom.com/etc/hwpDownload.do

Byte sequences

Position type

Absolute from BOF

Offset

0

Maximum Offset

0

Byte order

Value

48575020446F63756D656E742046696C652056332E3030201A0102030405

Top of page

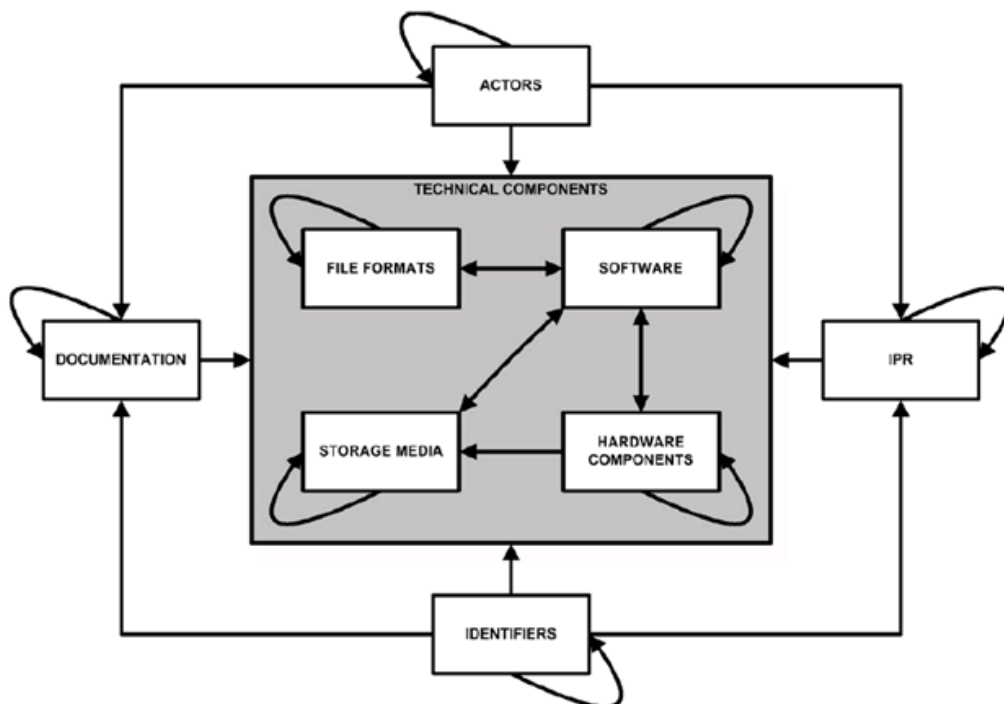
[그림 3-21] PRONOM 파일 포맷 검색 구성요소

2) 구성요소

- PRONOM에서 관리하는 핵심적인 기술적인 구성요소는 파일 포맷(File Format), 소프트웨어 구성요소(Software Component), 하드웨어 구성요소(Hardware Component), 저장매체(Storage Media)를 포함한다.
 - 파일 포맷 구성요소: 내부시스템에서 생성된 ID, 포맷 외부 식별자, 이름, 다른 포맷명, 'Text' 인지 'binary' 기반인지, 포맷의 바이트오더, 파일 포맷 식별을 위한 내부서명/외부서명, 압축 방법, 문자 부호화 등 포맷 자체에 대한 기술적인 요소들이 광범위하게 포함
 - 소프트웨어 구성요소: 내부시스템에서 생성된 ID, 소프트웨어 외부 식별자, 이름, S/W에 적용된 서비스 팩이나 패치 수준, 필요한 운영시스템과 기타S/W 요건, 필요한 H/W 요건, 저장매체 유형, 지원된 언어, 지원일이나 지원 종료일 등이 포함
 - 하드웨어 구성요소: 내부시스템에서 생성된 ID, 하드웨어명, 구성요소 유형 (CPU, Motherboard, RAM, Hard disk drive 등), 필요한 S/W, 발행된 날짜, 주요특징 기술, 개발자 등이 포함
 - 저장매체: 내부시스템에서 생성된 ID, 매체에 적용할 외부 식별자, 매체명, 주요

특징 기술, 매체 유형(Punched paper tape, Punched card, Flexible magnetic disk, Hard disk drive, Magnetic tape cartridge, Magnetic tape reel, Optical disc, Magneto-optical 등), 지원되는 쓰기 유형(Read, Write, R/W), 쓰기 속도, 데이터 접근 방법(Random, Serial), 데이터 저장 용량 등 물리적으로 디지털 객체를 저장하는 데에 사용되는 유형을 나타냄

- 현재 TNA(영국 국가기록원) 홈페이지에서 서비스되는 PRONOM 시스템에서는 파일 포맷과 소프트웨어 구성요소 위주의 정보를 제공하고 있다.
- 서비스되고 있는 시스템과 위에서 소개한 기술적인 요소(technical components)에서는 자원 관리와 장기보존을 직접적으로 지원하는 데에 관련된 특별한 요소들은 드러나진 않지만, PRONOM 4의 정보모델을 보면 앞서 소개한 기술적인 요소(technical components) 4개를 중심으로 하여, ACTOR, DOCUMENTATION, IPR, IDENTIFIER 등의 핵심 엔티티들을 중심으로 구성되어 있다.



[그림 3-22] PRONOM4의 high-level information model

-
- Actor: 다른 각 엔티티에 역할을 수행하는 개인이나 조직을 나타내며, 내부시스템에서 생성된 ID, 개인이름, 조직이름, 직책, 주소, 국가, 연락처, 노트, Source Actor 등이 포함
 - Documentation: 다른 엔티티와 연관되는 문서를 나타내며, 내부시스템에서 생성된 ID, 유형(authoritative, informative or speculative), 서지정보, 작성자, 제목, 출판자, 제목, 가용성, Source Actor, 출처 정보 등을 포함
 - IPR: 다른 엔티티에 적용할 수 있는 지적 재산을 나타내며, 내부시스템에서 생성된 ID, 형태, 소유자, 관할권(Jurisdiction), 라이선스, 날짜, Source Actor 등이 포함
 - Identifier: 다른 엔티티에 적용할 수 있는 외부 식별자이며, 내부시스템에서 생성된 ID, 식별자값, 형태, 정보 노트, Source Actor, 국가, 연락처, 노트 등이 포함

3) PRONOM의 정보 모델

- PRONOM 정보 모델에서는 두 가지 종류의 관계가 허용된다. 동일한 종류의 엔티티 사이의 관계는 source 엔티티, target 엔티티, 관계의 유형, 즉 Previous version of, Subsequent version of, Can contain, Can be contained by, Equivalent to, Subtype of, Supertype of, Other 등 주로 계층을 나타내기 위한 것과 관계에 대한 정보성 노트를 포함한다.
- 서로 다른 종류의 엔티티 사이의 관계는 예를 들면, documentation과 actor 엔티티 사이의 “author” 관계는 actor가 documentation 엔티티의 하나의 속성값으로 (actor 형식의) author를 포함하는 것에 의해 정의되며, 그 외에 복잡하게 정의된 관계로는 파일 포맷 엔티티와 소프트웨어 엔티티 사이의 관계인 “Software Process”, 기술적인 요소(technical components)와 요구되는 소프트웨어 엔티티 사이의 관계인 “Software Requirement”, 그리고 기술적인 요소(technical components)와 요구되는 하드웨어 엔티티 사이의 관계인 “Hardware Requirement” 등이 있다.
- 위와 같은 레지스트리 정보 모델을 이용한다면 기술적인 요소 내의 엔티티들과 함께 그 이외의 핵심 엔티티들을 이용하여 포맷의 기술적인 정보뿐만 아니라 디지털 자원의 장기보존 관점에서 필요한 정보들도 포함하여 관리할 수 있다.

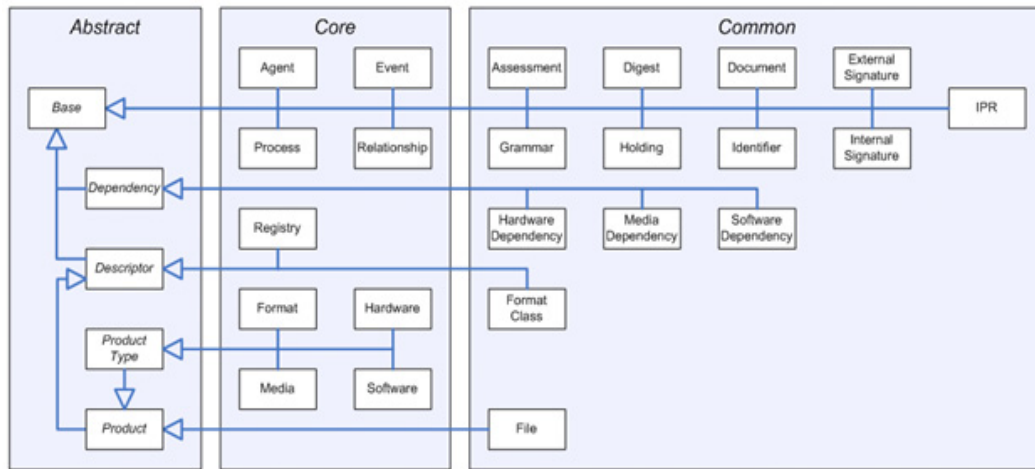
□ 미국 GDFR (Global Digital Format Registry)

1) 개요

- 하버드 대학의 도서관(Harvard University Library)과 MIT가 실시한 GDFR 프로젝트는 디지털 보존 리포지토리를 구현할 때, 개별 보존 프로그램에 드는 시간과 노력을 절약해 주고 디지털자원의 지속적 사용을 위해 포맷 정보를 제공하는데 활용되고 있다(유영수 2006).
- 인력 및 통신 문제와 사용가능한 소프트웨어 및 레지스트리 미비로 2008년에 프로젝트가 종료되었으며 하버드 대학교 도서관에서 GDFR 프로젝트를 오픈 액세스로 제공하고 있다.
- GDFR 프로젝트의 목적은 다음과 같다.
 - 디지털자원의 기술 정보를 공동으로 관리하고 공유하기 위한 공통 매커니즘을 제공하여 기술 정보 기술에 드는 중복된 노력을 줄여주기 위함
 - 보존 활동과 관련된 모든 이들에게 기술정보에 대한 접근 채널 제공
 - 디지털 포맷에 대한 상세정보를 필요로 하는 부가서비스의 토대 마련(국가기록원 2009)

2) 구성요소

- Use case model에서 제시된 기술요소는 실제 레지스트리를 서비스할 때에 필요하다. 14개의 요소들 중 Add Collection Record와 같은 것은 디지털자원 생산 및 편집에 활용될 수 있을 것으로 보인다. 또한 이러한 use case는 사용자가 컬렉션의 포맷을 알고 있는 상황에서 사용할 수 있음을 명시한다.
- GDFR 데이터 모델(Data Model, Version 5.0.6)은 PRONOM 4 정보 모델(version 1 of 4 January 2005)에 강한 영향을 받았으며, 그 외에 DCC Representation Information Model, ISO/IEC 11179, 그리고 OASIS/ebXML Registry Information Model에도 영향을 받았다 할 수 있다.



[그림 3-23] GDFR의 Data model

3) GDFR의 데이터 모델

- 데이터 모델 내의 Format[Core] 엔티티 자체 내에 정의된 속성의 수는 적으나, 그 외에도 관련된 다양한 엔티티들을 정의할 수 있어서, 이를 통해 포맷에 대한 기술적인 정보를 풍부하게 설명할 수 있다.
- Process 등의 엔티티를 통하여 하나의 포맷 상에서 작동하는 소프트웨어의 처리 또는 포맷의 결과를 정의하는데, 이를 활용하여 디지털자원의 상호작용과 동적인 요소를 정의할 수 있다.
- Registry 등의 엔티티를 통해 자원의 모델 체계를 서술하는 역할을 하는 레지스트리를 별도 정의할 수 있다. 만약 구성이 서로 다른 이질적인 디지털자원 세트를 동일한 GDFR 시스템으로 관리해야 할 때에 해당되는 레지스트리를 별도로 정의하여 구분할 수 있다.
- Relationship 엔티티는 두 포맷 사이의 관계를 정의하는 것으로, Affinity, Containment, Equivalence, Extension, Implements, Subtype, Version, Other 등을 포함한다.
- Assessment [Common] 엔티티에는 FDD(LC Format Description Document)를 활용하여 포맷 평가에 대한 상세 정보를 기술할 수 있다.

- 그 안에 정의된 identification And DescriptionSection의 relationships 요소의 하위에 typeOfRelationship(관계 유형)을 정의할 수 있는데, relation 엔티티 보다는 더 다양한 21개의 종류의 type이 존재한다.
- 그 중에 다수의 복합적인 포맷으로 구성된 기록물의 파일 간의 관계를 기술하는 데에 사용될 수 있는 것의 예를 들면, “Must have component”는 하나의 디렉토리 내에서 관련된 이름을 가지며, 공존하고, 함께 기능을 행하는 파일 그룹을 다루기 위해서 사용될 수 있다.
- 위의 Assessment 엔티티에서 외부의 표기법 형태를 차용한 것처럼 FormatClass 엔티티의 유형 또한 FDD, GDFR, 그리고 PRONOM 중에서 선택하여 format classification 값을 지정할 수 있으며, Identifier 엔티티에서는 40여개의 다양한 Namespace 들을 지원한다.
- Document [Common] 엔티티를 통하여 기획자등 생산자의 요구사항을 Specification 등의 형태로 저장할 수 있다.

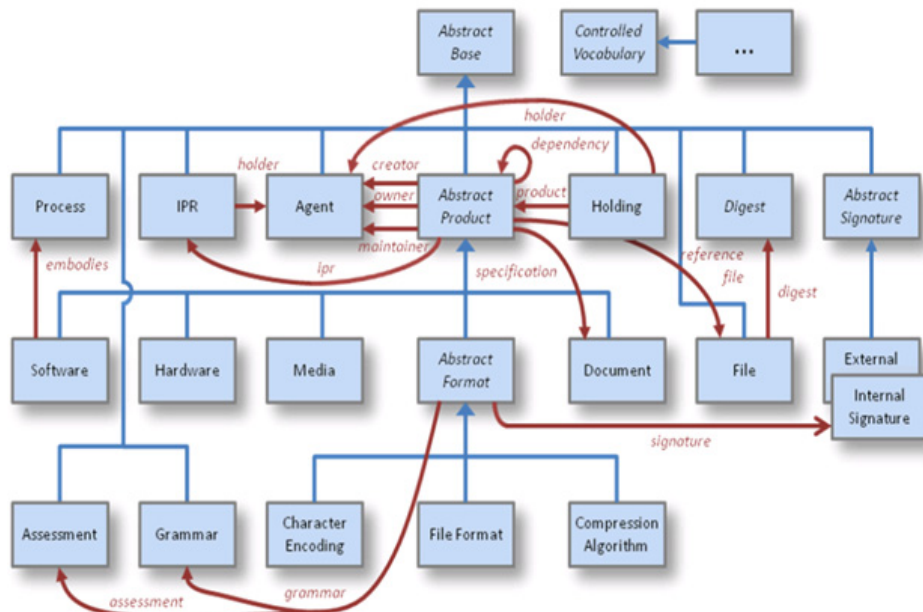
□ 미국 UDFR (Unified Digital Format Registry)

1) 개요

- UDFR은 미국의 디지털정보 인프라 보존 프로그램인 NDIIPP의 일환으로 미국 의 회도서관에서 투자하고 캘리포니아 디지털도서관의 큐레이션 센터에 의해 개발되었다.
- UDFR의 초기에는 지역 사회 참여가 매우 높았지만 배포 후에 지속되지 않았다.

2) UDFR의 데이터 모델

- UDFR은 GDFR과 PRONOM의 기능과 자산을 통합하고자 하였으며, UDFR의 클래스 계층관계도 GDFR의 데이터 모델과 거의 유사하다.
- 데이터 모델은 다음의 [그림 3-24]와 같이 구성되며, 주요한 엔티티의 특징은 다음과 같다.



- Agent 엔티티를 사용하여 생산자, 기획자 등을 표현함
- Document 객체를 통하여 요구사항을 저장할 수 있음
- Process를 통하여 기록물의 기능성을 나타낼 수 있음
- 추상 엔티티인 Product를 통하여 대상 자원을 표현함
- Assessment 및 Grammer 등을 통하여 포맷의 특징을 기술할 수 있음

□ 국가기록원 기술정보은행

1) 개요

- 국내의 대표적인 사례인 국가기록원의 위탁으로 2009년 ‘한글과컴퓨터’에서 프로토타입으로 구축한 디지털 보존을 위한 기술정보은행의 개발 목적은 디지털 콘텐츠 공유의 책임을 갖고 해결책을 마련하고자 하는 것이었으나 현재는 서비스 되고 있지 않다.
- 2008년까지 151개의 포맷 종류가 구축되었으며, 2009년 19개의 포맷이 추가되어 총

170개의 포맷 정보를 포함하고 있다.

- 1차 연구 이후 보다 다양한 문서 형식을 지원하기 위한 고도화 사업을 수행하 업무 환경에 가장 빈번히 사용되는 문서 형식으로 2006년, 2008년에 각각 ISO 26300 과 ISO 29500 국제 표준으로 승인된 ODF(Open Document Format) - (추가 15종의 문서 형식)과 OOXML(Office Open XML) - (추가 3 종의 문서 형식) 개방형 오피스 문서 형식을 추가 지원하게 되었으며, ISO 28500 WARC라는 웹 기록물 저장 국제 표준 형식에 대한 유효성 검증 기능을 개발하여 지원하고 있다.
- 또한 국내 실정을 반영하여 MS오피스, 한컴오피스, 하나워드, 아리랑, 훈민정음과 같은 레저시 포맷과 국가기록원의 장기보존포맷인 NEO 패키지 파일에 대한 유효성 검증 기능까지 추가하고 있다.

2) 기술정보은행의 구성요소

- 기술정보은행의 정보는 크게 파일 포맷 정보와 소프트웨어 정보의 두 가지로 구성되어 있다.
- 포맷 구성요소: 파일 포맷을 확인할 수 있을 정도의 정보. General(포맷의 일반적인 요소들), Technical(포맷의 기술적인 요소들), Management(포맷의 관리에 관한 요소들) 하위에 분류되어 있으며, XML 스키마로 표현될 수 있다.
- 소프트웨어 구성요소: 소프트웨어의 이름, 버전, 제작사, 연락처 등과 같은 기본적인 정보. 해당 포맷이 의존적인 소프트웨어적인 환경에 대해서 정의하는데, 여기에는 파일 포맷 구성요소와 유사하게, General(소프트웨어의 일반적인 요소들), Technical(소프트웨어의 기술적인 요소들), Management(소프트웨어의 관리에 관한 요소들) 하위에 분류되어 있으며, XML 스키마로 표현될 수 있으며 그 외에 포맷 식별 및 검증도구, 텍스트 추출 기능 등 부가기능 등이 프로토타입에 포함된다.
- 포맷 및 소프트웨어 구성요소 중 Management 요소들은 포맷의 일반적인 그리고 기술적인 정보들을 관리하는 제공자에 대한 항목이 주를 이루나, 이것을 활용하여 디지털자원의 장기보존을 위한 생산자의 맥락 정보를 관리해보는 것도 시도할 수 있다.

- 포맷 및 소프트웨어 구성요소 중 Technical의 TechnicalEnvironment 요소는 포맷의 기술적인 환경인데, 이 요소를 통하여 포맷의 기능성을 설명하는 데에 활용해볼 수 있을 것으로 판단된다.
- 또한 소프트웨어 구성요소의 Feature(S/W의 일반적인 특징), Note(정보주기) 등을 활용하여 기능성과 상호작용에 대한 요소를 표현해 볼 수 있다.

3.3.3 국립중앙도서관 포맷 레지스트리 분석

□ 국립중앙도서관 OASIS 디지털자원정보

- 국가 지식 자원의 수집과 보존에 있어 책임을 지고 있는 국립중앙도서관은 국가의 대표도서관으로서 가치 있는 인터넷 자료를 국가적인 차원에서 수집 및 축적하여 미래 세대에 연구 자료로 제공하고, 온라인 디지털자원 수집 및 보존의 표준모델 연구 등에 목적을 두고 온라인 디지털자원 수집 및 보존사업인 OASIS 프로젝트를 추진하였다.
- OASIS는 디지털자원이 장기간 수집되고 보존되는 경우 해당 자원을 구동하기 위한 정보를 검색하고 상세 내용 조회가 가능하도록 기술정보를 관리하기 위해 보존되어 있거나 향후 보존될 가능성이 있는 디지털자원에 대한 파일 포맷, 해당 파일 포맷을 지원하는 소프트웨어 정보 등을 체계적으로 관리하기 시작하였다.
- 디지털자원정보는 다양한 디지털자원정보에 대한 파일 포맷, 소프트웨어정보, 해당 제조사 정보를 제공하고 있다.
- 제조사 정보는 2006년 11월 16일 최종 수집 되었고, 검색 결과 130건이 제공되고 있다.
- 파일 포맷 검색은 디지털자원을 생성하는 각 소프트웨어의 기술지원기한과 배포된 날짜로 검색하는 기능을 제공하며 2013년 12월 마지막으로 갱신되어 1,063건 엔트리를 포함하였다.
- 디지털자원정보는 세 가지 엔티티를 정의하는 단순한 구조의 레지스트리로, 주로

파일 포맷과 소프트웨어에 대한 기술정보를 저장하는 데에 치중하고 있다.

- 기술자원 및 소프트웨어의 ‘메모’ 속성, 그리고 포맷의 ‘구분’, ‘파일 포맷 분류’ 속성을 통하여 어떤 종류의 포맷인지를 서술하고, 해당 포맷의 기능성 및 상호작용성 등에 대한 정보 또한 제공 가능할 것으로 보인다.
- 그러나 2013년 이후, 업데이트가 전혀 이루어지지 않고 있으며 국립중앙도서관의 디지털자원의 장기보존을 위해서 전혀 OASIS의 디지털자원정보의 기능을 활용하고 있지 않은 실정이다.
- 국립중앙도서관 OASIS의 디지털자원정보를 포맷 레지스트리로 활용하고자 한다면, 현재 디지털자원정보가 실제 포맷 정보가 아니라 포맷을 설명하기 위한 메타정보로만 구성되어 있고 소프트웨어 정보 역시 소프트웨어에 대한 기본적인 정보만을 포함하고 있으므로 동적인 렌더링 기능성 요소에 대한 속성을 포함하도록 디지털자원정보 구조 개선에 대한 연구가 필요할 것이다.

□ 국립중앙도서관의 포맷 관리 및 포맷 레지스트리에 대한 요구분석

- 디지털자원의 파일 포맷 노후화는 디지털 보존의 잠재적으로 가장 큰 본질적인 문제라고 할 수 있다. 앞서 국립중앙도서관에서 수집, 보존하고 있는 디지털자원의 유형 및 특성을 조사, 분석한 결과 파일 포맷이 매우 광범위함을 확인하였다.
- 국립중앙도서관에서 보존하고 있는 디지털자원은 기증, 납본, 수집 등 자원의 제공 유형에 따라 다양하며 디지털자원을 구성하고 있는 디지털 컴포넌트와 관련한 파일 포맷 종류가 지속적으로 증가하고 있다. 특히 컴퓨터 기술 발전에 따른 디지털 자원의 동적 특성이 변화하고 있으며, 파일 포맷이 퇴화되거나 더 이상 사용하지 않게 되는 경우가 발생할 수 있다.
- 현재 국립중앙도서관에서 온라인 자료의 납본의 경우, 해당 자료(보존용 및 열람용 파일)와 서지정보의 디지털 파일을 국립중앙도서관 서지정보유통지원시스템에 등록하거나 해당 파일과 서지정보의 디지털 파일을 저장매체에 저장하여 국립중앙도서관으로 송부, 국립중앙도서관에서 해당 자료와 서지정보의 인터넷상 위치를 확

인하고 접근하여 수집하고 있다.

- 장기보존의 관점에서 오프라인 매체에 저장된 자원의 경우, 수집, 관리되는 디지털 자원은 저장매체마다 서로 다른 저장 조건과 보존 요구사항들을 가지고 있으며 관련 소프트웨어와 하드웨어를 포함한 종합적인 보존이 필요하다.
- 물리적인 저장매체에 대한 변환 작업을 통한 보존은 비교적 단순하고 쉽지만, 하드웨어, 소프트웨어의 빠른 기술적 퇴화에 따라 사용 불가능하게 되는 경우가 발생한다. 따라서 예측 불가능한 매체의 변화에 대해서 특정 콘텐츠의 장기보존 필요성이 높아지므로 운영 환경에 대한 애플리케이션을 통해 데이터 조회 가능성 유지가 필요하다.
- 파일 형태로 수집, 관리되는 디지털자원은 장기적 보존을 위해서 애플리케이션의 보관과 함께 생성되는 파일 형식에 대한 보존이 필요하다.
- 최근에는 3D 프린팅, 3D 애니메이션, 증강현실 등 더욱 복잡해진 시청각 콘텐츠 제작이 활발해지면서 과거와 달리 데이터 파일 포맷 유형과 제작 형태가 다양화되고 있다. 또한 하나의 파일 포맷이 2개 이상 유형의 자료를 저장할 수 있도록 파일 포맷 구조가 다변화하면서 복합 구조의 파일 저장이 가능하도록 파일 포맷의 변화가 이루어지고 있다.
- 또한 복잡한 멀티미디어 환경에서 작성되어진 디지털자원을 안전하게 보존하기 위해서는 표준 파일 포맷의 사용이 절실하나 기술 개발을 주도하는 기업의 입장에서 표준에 얽매어 새로운 기술 구현이 지연되는 단점을 피하고 기술 개발의 주도권을 잡기 위해 비공개된 독점적인 파일 포맷 구조를 선호하고 있는 것이 현실이다.
- 따라서 디지털자원을 안전하게 장기보존하기 위해서는 신기술 분야에서 새롭게 정의되어 사용되는 파일 포맷을 주기적으로 확인하여 파일 포맷 정보를 확인할 필요가 있다.
- 현 시점에서 가장 중요한 것은 사라져 가는 애플리케이션의 파일 형식을 정확하게 파악하여 포맷 레지스트리를 구축하는 것이 필요하다. 디지털 보존 자원의 재현을 위한 기술 정보의 수집 및 관리 정보를 위한 요소들을 정의하고 기술정보를 이용하여 장기보존을 위한 기능을 제공할 수 있는 포맷 레지스트리가 요구된다.

-
- 영국 TNA의 PRONOM과 같이 포맷 레지스트리를 구축하여 장기적인 차원에서 기술노후화에 포맷의 기술 변화를 모니터링하고 기술 변화가 저장된 디지털자원에 미칠 변화에 대비하여야 한다.
 - 포맷 레지스트리는 OAIS 참조모형에서 제시한 안전하고 신뢰할 수 있는 보존전략으로 활용될 수 있다. 포맷 레지스트리의 기능을 통해 입수, 보존, 접근 등의 주요 보존 생애주기에서 파일식별, 파일검증, 파일특성파악, 파일 포맷 위험평가를 통해 디지털자원의 적절한 보존전략으로 활용할 수 있다.
 - 디지털자원의 보존에 있어서 입수(수집)-저장-접근에 이르는 기본적인 프로세스에 파일 포맷 레지스트리를 연계함으로써, SIP, AIP 정보패키징 변환을 지원하고 관련 기술메타데이터를 추출하여 포맷 레지스트리에 등록, 관리하도록 한다.
 - 포맷 레지스트리는 OAIS 모델과 연계하여 기술변화와 노후화를 모니터링하고 그런 변화가 보존 자원에 미칠 위험을 평가하는 역할을 하게 된다. 이런 위험평가를 통해 마이그레이션, 에물레이션 등 디지털자원에 대한 적절한 보존전략을 수립하여 수행함과 동시에 이 과정에서 새로 생성되는 기술메타데이터를 레지스트리에서 등록, 관리함으로써 디지털자원의 장기보존을 지원할 수 있다.
 - 현재 국립중앙도서관의 포맷 레지스트리라고 할 수 있는 OASIS의 디지털자원정보는 오랫동안 유지, 관리되지 않고 있으며 국립중앙도서관의 디지털자원의 장기보존을 위해서 전혀 그 기능을 활용하고 있지 않은 실정이다.
 - 우선 디지털자원의 장기보존을 위한 포맷에 대한 가이드를 제시하기 위해서는 현재 국립중앙도서관의 디지털자원 포맷에 대한 전반적인 현황을 조사하고 이를 수용하면서 장기적인 안정성을 확보하는 방안 마련이 시급하다.
 - 이를 위하여 주요 국가도서관에서 규정한 디지털자원 포맷 규격과 주요 국가기록원에서 제안하는 디지털자원의 표준 포맷과 규격에 대하여 조사한 후 최신 정보 포맷의 수용성을 평가한 후 디지털자원의 장기보존에 적절한 포맷에 대한 가이드라인을 제시하고 최종 제안한 디지털자원 포맷으로 변환하는 전략을 마련할 필요가 있다.

3.3.4 국립중앙도서관 포맷 레지스트리 개선(안)

□ PRONOM의 오픈소스 도구 활용

- 파일 포맷 레지스트리는 디지털자원의 보존과 시스템 간 상호호환성을 확보하기 위하여 각종 파일 포맷에 대한 명료한 정의와 식별 정보를 관리하고 제공하기 위하여 개선되어 왔으며 특히 영국 국가기록원의 PRONOM은 기구축 운영 중인 디지털 아카이브 및 관련 연구에 가장 많이 활용되고 있음을 알 수 있었다.
- 국내외 주요 오픈 소스 식별 도구들은 DROID의 식별 방법을 채택하고 있으며 PRONOM에서 얻은 시그니처 파일을 이용하고 있다. 국립중앙도서관의 포맷 레지스트리도 파일 식별 기술 연구에 있어 DROID의 식별 방법과 포맷 시그니처 표현 규칙을 참조할 필요가 있다.
 - DROID(Digital Record Object Identification): PRONOM에서 제공되는 포맷 정보를 이용하여 파일의 포맷을 식별하는 별도의 소프트웨어
 - PUID(Persistent Unique Identifier): PRONOM 레지스트리의 기록을 지속적이고 유일하며 명확히 식별하기 위한 식별 체계
- DROID는 오픈소스로서 누구나 다운로드하여 사용하고 개발에 참가할 수 있다. PRONOM 데이터베이스에 저장되어 있는 파일 포맷 정보를 XML 파일 포맷의 시그니처 파일로 다운로드 받아 이 정보를 이용하여 파일 포맷을 식별한다. PRONOM과 별도로 추진되고 제공되는 소프트웨어지만 PRONOM의 파일 포맷 서명 파일을 기반으로 어느 정도 의존적이라 할 수 있다.
 - 실행: Java 애플리케이션으로서 GUI 또는 명령어로 실행하며 공개 API가 제공되며, 명령어 라인에서 식별할 파일 위치, 출력 옵션 등을 직접 입력하여 실행하거나 또는 그래픽인터페이스 도구를 이용하여 실행할 수 있다. 폴더를 지정하면 순환적으로 하위 폴더 및 파일을 찾아 실행한다.
 - 실행 결과: 파일 포맷 이름, 버전, PUID, MIME 타입, 식별 방법 등이 출력된다. 그래픽인터페이스에서 실행했을 때 식별 결과 역시 도구 내에 출력되며 CSV 파일 포맷으로 추출할 수 있다.

-
- 파일 포맷 식별: 파일 포맷 시그니처를 확장자와 비교한 뒤 시그니처 일치 여부, 확장자 일치 여부에 따라 결과가 식별 성공, 잠정적, 식별 불가로 분류됨. 성공적으로 식별된 파일 포맷은 일치한 시그니처가 범용인지 유일하게 사용되는지를 구분하며 파일 확장자를 비교하여 실제 확장자와 파일 포맷 확장자가 일치하는지를 검사하며 일치하지 않을 경우 경고를 결과에 추가한다.
 - 신뢰도 표시: 식별한 파일 포맷에 대하여 신뢰도를 함께 표시해준다. 파일 포맷과 시그니처의 관계가 1:1인 유일 파일 포맷의 시그니처와 일치하는 경우가 가장 신뢰할 수 있는 결과이며 확장자만 일치하는 파일 포맷이 발견된 경우의 신뢰도는 낮은 편이다.
 - PRONOM의 파일 포맷 식별도구인 DROID에서는 포맷 특성 정보가 비트스트림 외부에 존재하느냐 내부에 존재하느냐에 따라 외장 시그니처와 내장 시그니처로 구분된다.
 - 파일 외부에 존재하는 외장 시그니처란 디지털 객체의 비트스트림 바깥에 존재하는 포맷 특성 데이터로서 Windows의 경우 파일 확장자, Mac OS의 경우 데이터 폭스(data forks)가 외장 시그니처에 해당된다.
 - 내장 포맷 시그니처란 디지털 객체의 비트스트림 속의 특정 위치에 존재하는 포맷에 대한 특성 값을 말한다.
 - 포맷과 시그니처와의 관계를 살펴보면, 각 포맷은 다수의 내장 시그니처 또는 파일 확장자를 가지며 각 내장 시그니처는 1개 이상의 바이트 시퀀스로 구성된다. 포맷에는 식별자가 부여되고 포맷과 포맷 사이에는 수많은 관계를 가질 수 있다.
 - 파일 확장자는 포맷 식별에 불충분하며 식별 도구는 포맷 시그니처를 이용해야 한다. 따라서 식별 도구가 포맷 시그니처에 대한 정보를 신속히 획득하고 최신 정보로 유지하기 위한 수단을 갖추고 있느냐 역시 식별 도구 개발에서 고려해야 할 것이다.
 - 국내외 포맷 레지스트리 및 식별 도구 사례에서 보듯이 각 포맷별 포맷 시그니처에 대한 정보는 포맷 레지스트리에 저장하여 관리하며 식별 도구는 레지스트리로부터 이 정보를 얻어 식별에 사용하는 체계를 갖추고 있다.

□ KOLIS 보존 기능과 연계

- 국립중앙도서관의 포맷 레지스트리를 구축하여 KOLIS 보존 기능과 연계함으로써, 디지털자원의 장기적인 보존 차원에서 기술노후화에 따른 포맷의 기술 변화를 모니터링하고 기술 변화가 아카이빙에 저장된 디지털자원에 미칠 변화의 충격을 완화시킬 수 있다.
- 포맷 레지스트리는 디지털자원의 입수, 보존 및 접근 등 전반적인 단계에 걸쳐 안전하고 신뢰할 수 있는 보존전략으로 활용될 수 있다.
- 수집, 등록단계에서는 파일 식별, 파일 검증 등의 기능을 통해 디지털자원의 장기보존 지속가능성을 고려하여 파일 포맷을 선택하도록 한다. DROID의 파일 식별을 통해, 디지털자원의 파일 포맷이름, 버전, PUID, MINE 타입 등 파일 포맷의 중요한 정보를 자동으로 획득, 추출하여 보존 메타데이터로 자동 입력할 수 있다. 파일 포맷 레지스트리를 통해 다양한 유형의 콘텐츠가 수집될 것과 장기보존 해야 하는 것을 감안하여 콘텐츠의 파일 형식을 분석하고, 향후 보존활동에 영향을 미치는 파일 포맷 정보를 등록하여 체계적으로 관리할 수 있다.
- 보존 및 접근 단계에서는 입수-저장-접근에 이르는 기본적인 프로세스와 파일 포맷 레지스트리를 연계하여 SIP, AIP 정보패키징 변환을 지원하고 관련 기술메타데이터를 추출하여 레지스트리에 등록, 관리함으로써 디지털자원의 장기보존에 있어서 필요한 정보를 관리할 수 있다. 파일 포맷을 식별하여 포맷 레지스트리를 통해 그 타당성 검증하고 보존 패키지로의 변환을 지원하고 내용정보와 보존기술정보에서 적절한 기술(technical)정보를 추출하여 포맷 레지스트리에 추가한다.
- 이처럼 포맷 레지스트리는 OAIS 모델과 연계하여 기술변화와 노후화를 모니터링하고 저장된 디지털자원에 미칠 영향력을 평가한다. 파일 포맷의 위험평가를 기반으로 마이그레이션, 인캡슐레이션 등 적절한 보존전략을 수립하여 수행하도록 하며, 이 과정에서 새로이 생성되는 기술(technical) 메타데이터를 포맷 레지스트리에 등록, 관리함으로써 효율적인 보존전략으로 활용될 수 있을 것이다.

3.4 종합 요약 및 시사점

□ 디지털자원 보존용 메타데이터

- 보존 메타데이터는 원 데이터에 대한 접근과 검색을 제공하고, 효율적으로 정보자원을 관리할 수 있도록 지원하는 것을 목적으로 한다. 이를 위해 보존 메타데이터는 디지털자원의 장기간 유지가능성, 재연가능성, 이해가능성, 진본성, 무결성을 제공할 수 있어야 한다.
- PREMIS는 디지털자원의 장기보존을 위한 가장 종합적이고 포괄적인 메타데이터 요소 스키마로 받아들여지고 있으며, 2015년 보존 생애주기를 반영할 수 있는 PREMIS 3.0이 발표되었다. PREMIS 3.0은 다른 메타데이터 표준과 함께 보존 환경에 따라 사용목적에 맞게 설계할 수 있는 확장성을 갖는다.
- PREMIS를 기반으로 하는 보존 메타데이터는 디지털자원의 입수 과정에서 기술 메타데이터 레코드와 연계를 통해 최소한의 메타데이터가 자동으로 생성될 수 있는데, 장기적인 이용가능성 측면에서는 객체의 중요한 속성, 관계, 이벤트, 환경 정보 기술이 반드시 보존 메타데이터의 핵심요소들로 구현되어야 한다.
- 2011년 설계된 국립중앙도서관 PREMIS 기반 보존 메타데이터는 KOLIS에 적용되어 국립중앙도서관 디지털자원의 기술메타데이터인 MODS의 기본 정보와 연계가 가능하도록 설계되었으나, 그 구조가 매우 복잡하며 대부분의 하위요소들이 필수 값으로 지정되어 있어 실질적으로 적용이 쉽지 않다.
- 이러한 문제점을 해결하고자, SPOT과 OAIS 기반의 보존 위험과 기능 분석을 통해 국립중앙도서관의 보존 메타데이터 핵심요소를 파악하였다. 또한 국립중앙도서관의 디지털자원에 적용되고 있는 메타데이터를 분석하여 연계를 통해 자동으로 추출할 수 있는 요소들을 확인하고 핵심요소들로만 구성된 PREMIS 개선(안)을 제안하였다.
- 디지털자원 보존 생애주기 전반에 걸쳐 수집, 관리, 보존, 이용 등 단계에서 정보자원의 무결성을 확보하고 지속적인 접근을 위해서는 각 단계의 작업이 유기적으로

이루어질 수 있도록 메타데이터의 통합적 연계가 필요하다.

- 납본 제출 단계에서는 디지털자원의 하드웨어, 소프트웨어 요소 등 보존에 있어서 중요한 정보가 자동으로 추출, 연계하여 추후 보존 단계에서 정보패키징 변환을 지원하고 관련 기술메타데이터를 추출하여 레지스트리에 등록 관리할 수 있다.
- BnF의 SPAR의 보존 메타데이터(2014, 2017)의 사례와 같이, 현재 국립중앙도서관의 디지털자원 기술을 위한 METS와 PREMIS 기반의 연계형 메타데이터 구현 방안을 모색할 필요가 있다.
- 제안하는 국립중앙도서관의 보존 메타데이터의 요소는 PREMIS 기반의 디지털자원의 보존 활동을 기술하는데 필요한 핵심적인 요소들만을 대상으로 하고 있으며, 이외의 세부적인 기술요소들은 기존의 메타데이터 표준과의 상호운용성 확보를 통해 추가할 수 있다.

□ 장기보존을 위한 국립중앙도서관 디지털자료 프로세스 확장 방향성

- 국립중앙도서관 디지털자료 관련 업무는 온라인 납본, 웹 아카이빙, 디지털화 등 3개 영역으로 나눌 수 있으며 각각의 프로세스가 독립적으로 운영되고 있으나 기본적인 단계는 각 영역별로 기본 처리 후 디지털장서관리시스템과 KOLIS II 온라인 시스템에 등록하는 과정으로 구성되어 있었다.
- 장기보존을 위해 기본 업무를 바탕으로 장기보존 프로세스를 결합하여 기존 프로세스를 확장하는 방안을 다음과 같이 세가지 영역별로 제시하였다.
- 첫째, 온라인 자료 납본 관련 업무
 - 납본 원문을 서비스용으로 원문저장소에 보장하여 서비스를 하고 현재 보존용으로 생성하는 것을 장기보존시스템으로 전송하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.
 - 통합정보시스템 온라인(KOLIS II) 등록완료 후에 등록본의 메타데이터를 납본 원문과 함께 장기보존시스템으로 전송한다.
- 둘째, 웹 아카이빙 관련 업무

-
- 수집된 웹 정보자료를 웹자료와 웹사이트로 구분하여 장기보존 프로세스를 분리하는 것이 업무 프로세스 및 자료 특성상 바람직하다.
 - 파일 단위의 웹자료는 납본자료와 유사하게 서비스용으로 원문저장소에 저장하여 서비스를 하고 추가로 보존용으로 복사본을 장기보존시스템으로 전송하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.
 - 파일 단위의 웹자료는 통합정보시스템 온라인(KOLIS II) 등록완료 후에 등록본의 메타데이터를 납본원문과 함께 장기보존시스템으로 전송한다.
 - 웹사이트는 자동으로 수집하여 메타데이터가 생성되어 원문저장소에 저장이 완료되면 특정 시점을 설정하여 연단위로 추가된 데이터베이스본을 장기보존저장소에 백업형태로 저장하며 웹사이트 구성요소의 포맷을 포맷 레지스트리에 등록하는 체제로 운영한다.
- 셋째, 디지털화 관련 업무
 - 디지털화된 자료는 서비스용으로 원문저장소에 저장하여 서비스를 하고 추가로 보존용으로 복사본을 메타데이터와 함께 장기보존시스템으로 전송하여 장기보존포맷으로 전환한 후 장기보존저장소에 저장한다.

□ KOLIS II 보존 기능 개선(안)

- KOLIS II와 장기보존시스템은 독립적으로 구축하고 서로 연계하도록 설계하여야 확실한 관리 주체에 의해 일관성 있게 운영, 관리할 수 있다.
- 온라인 납본, 웹자료, 디지털화 자료는 디지털자원 보존을 위한 보존 메타데이터를 구축하고 콘텐츠 메타데이터인 MODS와 함께 METS를 이용하여 패키징하도록 한다.
- 웹사이트는 자동으로 수집부터 서비스, 그리고 보존까지 원스톱 워크플로우로 진행하도록 한다. 보존 메타데이터는 MODS DB에서 자동으로 추출, 변환하여 METS를 구축하도록 한다.

□ 디지털자원 포맷 레지스트리 개선(안)

- 포맷 레지스트리는 각 파일 포맷에 대한 기술 정보는 물론 그 파일 포맷을 생산한 소프트웨어 및 변환에 필요한 정보 및 소멸 위험에 놓인 파일 포맷에 대한 정보를 파일 포맷별로 포괄적이며 전문적으로 관리하는 데이터베이스를 말한다.
- 국내외 포맷 레지스트리의 사례를 분석한 결과, 디지털자원의 보존과 시스템 간 상호호환성을 확보하기 위하여 각종 파일 포맷에 대한 명료한 정의와 식별 정보를 관리하고 제공하기 위하여 개선되어 왔으며 특히 영국 국가기록원의 PRONOM은 디지털 아카이브 및 관련 연구에 가장 많이 활용되고 있음을 확인하였다.
- 국립중앙도서관은 2006년 OASIS 사업을 통해 수집, 보존하는 다양한 디지털자원 정보에 대한 파일 포맷, 소프트웨어정보, 해당 제조사 정보를 제공하기 위해 OASIS 디지털자원정보를 구축하였다. 그러나 OASIS의 디지털자원정보는 오랫동안 유지, 관리되지 않고 있으며 국립중앙도서관의 디지털자원의 장기보존을 위해서 전혀 그 기능을 활용하고 있지 않은 실정이다.
- 국립중앙도서관의 디지털자원의 장기보존 포맷 관리를 위해 파일 식별 기술 연구에 있어 가장 많이 참조되고 있는 PRONOM의 오픈소스 DROID의 식별 방법과 포맷 시그니처 표현 규칙을 참조하여 포맷 레지스트리를 구현할 것을 제안하였다.
- 포맷 레지스트리는 KOLIS의 보존 기능, 장기보존시스템과 연계하여 파일 포맷의 위험평가를 기반으로 마이그레이션, 인캡슐레이션 등 적절한 보존전략을 수립하여 수행하도록 하며, 이 과정에서 새로이 생성되는 기술(technical) 메타데이터를 포맷 레지스트리에 등록, 관리함으로써 효율적인 보존전략으로 활용될 수 있을 것이다.
- 국립중앙도서관의 디지털 보존 자원의 양적 증가와 다양화는 가속화되고 있으나, 디지털 보존 자원의 안정적이고 지속적인 사용성을 보장하는 장기보존에 대한 전략과 실무는 미비한 실정이다.
- 포맷 레지스트리는 디지털자원의 수집부터 보존까지 생애주기에 있어서 장기보존 전략 수립에 있어서 필수적이다. 포맷 레지스트리를 통해 디지털 보존 자원의 재현을 위한 기술 정보의 수집 및 관리 정보를 위한 요소들을 정의하고 기술정보를 이

용하여 장기보존을 위한 성공적인 전략을 수립할 수 있을 것이다.

- 포맷 레지스트리 구축을 위해서는 국립중앙도서관의 디지털 보존 자원 포맷에 대한 전수조사를 하여 이를 수용하면서 장기적인 안정성과 사용성을 확보할 수 있도록 자원 유형에 관한 분석과 보존을 위한 디지털 객체의 중요 속성 표현에 관한 연구가 필요하다.

제 4 장

국립중앙도서관 디지털자원 보존 정책

- 4.1 디지털자원 보존 정책
- 4.2 디지털자원 보존 환경 및 시스템 지침
- 4.3 디지털자원 보존 조직
- 4.4 종합 요약 및 시사점

제 4 장

국립중앙도서관 디지털자원 보존 정책

4.1 디지털자원 보존 정책

4.1.1 각 국가별 디지털자원 보존 정책 현황

□ 미국 의회도서관

- 미국 의회도서관은 21세기에 들어서서 디지털 보존 대상을 미국 의회도서관에 납본되는 디지털자료 이외에 국가유산으로 지정될 수 있는 모든 자료까지 확대하였다. 이를 위하여 미국 의회도서관은 미국 전역에 위치한 주정부 및 국가기관과 파트너십을 맺고 디지털 보존을 위한 수집 전략, 디지털화 전략, 그리고 디지털 장기 보존 전략을 수립하였다.
- 2000년 미국 의회도서관은 디지털자원의 보존 문제를 해결하기 위하여 국가 디지털정보인프라 및 보존 프로그램(National Digital Information Infra structure and Preservation Program, NDIIPP)이라는 10년간의 장기 프로젝트를 시작하였다. 프로젝트 초기의 활동은 국가차원의 디지털자료 수집활동에 동참하는 기관과의 파트너십을 맺는 것과 디지털자원의 장기보존 및 관리에 대한 다각적인 연구의 수행이었고, 다음 단계에는 디지털 환경과 디지털 유형에 적합한 디지털 보존 시나리오를 수립하는 것이었고, 최종적인 단계는 국가차원의 디지털 보존 인프라를 구축하는 것이었다.

- 특히 미국 의회도서관은 디지털 텍스트뿐만 아니라 사진, 만화, 영화, 사운드레코딩, 비디오게임 등과 같은 다양한 형태의 콘텐츠 수집전략과 보존방법을 개발하는데 많은 노력을 기울였다. 2011년에는 그동안 수행된 다각적인 연구와 조사를 토대로 하여 「우리의 디지털 유산을 보존하는 방법: 국가디지털정보 인프라 및 보존 프로그램 2010 보고서」를 발표하였다(Library of Congress 2010).
- 이 보고서는 디지털 보존이 성공적으로 수행되기 위해서는 ‘디지털정보의 니즈(needs)에 대한 이해’, ‘국가적 디지털 보존 네트워크 구축’, ‘미래를 위한 지식 확보(securing)’가 필요하다고 보았고, 디지털 보존의 세부적 전략을 다음과 같이 4가지 측면에서 다루었다: 1) 디지털 보존에 관한 책임 네트워크 구축, 2) 국가 디지털 컬렉션 구축, 3) 디지털 보존을 위한 기술적 인프라 개발, 4) 디지털 보존에 적합한 공공정책(법, 규제, 저작권 등) 개발’.
- 이후, 국가적인 사업으로 확장된 디지털 보존 및 인프라 구축 프로젝트는 미국 의회도서관에서 관리하는 국가 디지털 책임의무 연합 기관(National Digital Stewardship Alliance, NDSA)에서 진행되다가 2016년부터는 디지털도서관연맹(Digital Library Federation, DLF)에서 운영되고 있다.
- 국가차원에서 디지털 보존을 중점적으로 다루는 기관이 별도로 수립됨에 따라 미국 의회도서관은 디지털 보존 정책을 별도로 수립하지 않고 있고 국가 디지털유산 보존 정책을 따르고 있다. 그러나 미국 의회도서관의 수집정책이 유형별 또는 주제 영역별 등과 같이 각각 별도의 목적에 따라 수립되어 도서관업무에 사용되고 있듯이, 디지털 보존에 관련된 가이드라인이나 정책 및 규정은 기능별 또는 이슈별로 수립되어 있다.
- 최근에는 디지털자원의 유형별(스틸이미지, 사운드, 텍스트, 동영상, 웹 아카이브, 데이터세트, 지리정보 데이터, 유전정보 데이터 등)로 디지털화 및 디지털포맷, 메타데이터 등에 대한 가이드라인이 수립되었다.
- FADGI(Federal Agencies Digital Guidelines Initiative)가 제시한 최신 가이드라인을 연도 별로 정리하면 다음과 같다.

-
- 2017
 - ① 오디오 아날로그-디지털 변환기 지침(저비용)
 - ② MXF 적용 명세서
 - ③ DPX 파일의 메타데이터 연결 가이드라인
 - 2016
 - ① 영상필름 디지털화 가이드라인
 - ② 문화유산 객체의 이미지 디지털화 가이드라인
 - ③ 오디오 아날로그-디지털 변환기 지침(고품질)
 - 2014
 - ① 파일 포맷 비교 지침
 - 2012
 - ① 스틸이미지를 위한 최소단위 임베드 메타데이터 기술요소
 - ② 브로드캐스팅 WAVE 파일을 위한 임베드 메타데이터 기술요소
 - 2009
 - ① 디지털 이미지 프레임워크
 - ② 내용 카테고리화 디지털화된 객체
 - ③ 디지털화 활동: 프로젝트 기획
 - ④ TIFF 이미지 메타데이터
- 다음 [표 4-1]은 미국 의회도서관의 NDIIPP가 고려한 4가지 측면 중 컬렉션 구축 및 정책수립과 관련한 진행현황을 보여준다(LC 2010).

[표 4-1] 미국의 디지털 컬렉션 구축과 디지털 보존 정책수립을 위한 3단계 계획

활동 영역	A Decade of Action 2000-2009	Work in Progress 2010-2013	The Decade Ahead 2013-2020
컬렉션 구축	<ul style="list-style-type: none"> • 공공이용자나 연구자를 위하여 1400 개 이상의 디지털 컬렉션에서 위험 상태의 디지털 콘텐츠를 선택하고 보존함 • 체계적인 방식으로 특히 정책과 연계하여 디지털 콘텐츠의 수집, 보존 및 액세스 기능을 제공하는 조직/기관을 육성 • 문화유산 컬렉션 이용의 활성화를 위하여 디지털 콘텐츠를 도메인별, 수준별로 나누어 관리함 • 포맷, 업무프로세스, 프로토콜 등 디지털 보존관련 주요 표준 및 모범 사례 제시 	<ul style="list-style-type: none"> • 현재 및 미래가치 있는 콘텐츠에 대한 협력적 수집을 위한 국가 디지털 공동수집 프레임워크 구축 • 디지털 콘텐츠 보존과 액세스에 영향을 미치는 신규 디지털자료의 특징과 포맷에 대해 조언을 해줄 수 있는 전문가그룹 구성 • 정부, 정치/법, 지도/지리, 뉴스, 미디어/저널리즘에 관한 디지털자료의 집중적인 수집을 위한 단계별 전략 수립 • 디지털 컬렉션에 대한 향상된 통합검색 기능을 갖춘 웹 포털 또는 검색도구 구현 	<ul style="list-style-type: none"> • 세계 역사와 문화, 미국 역사, 예술과 문화, 종교와 철학, 스포츠, 레크리에이션, 여행 등과 같은 주제의 광범위하고 국가 단위의 디지털 컬렉션 구축 • 신뢰할 수 있는 여러 기관과의 협약을 통해 지속적인 국가차원의 디지털정보 수집 • 전 세계에서 접근이 가능하도록 국제 표준 메타데이터 구축 및 접근기능 제공 • 시각화, 매핑, 데이터 마이닝 등을 이용하여 다채롭게 컬렉션을 제공하고 이와 관련된 새로운 연구를 지원
정책	<ul style="list-style-type: none"> • 디지털 콘텐츠 보존을 지원하고 장려하기 위해 미국 저작권법 개정 지원 • 저작권법과 디지털 보존과의 관련성 및 영향에 대한 국제적 연구 수립 • 디지털 보존 및 접근의 지속 가능성에 관한 경제적 분석 연구 수립 	<ul style="list-style-type: none"> • 디지털 시대의 프라이버시 문제와 문화 유산 보존과의 관계 연구 후원 • 공공 이익을 위해 사적 자산 보존에 관한 인센티브 생성에 대한 권고 • 지속적으로 소장된 방대한 콘텐츠의 접근에 영향을 미치는 정책 문제 연구 	<ul style="list-style-type: none"> • 개정된 미국 저작권법에 따른 보존 활동 지원 • 디지털 보존을 위한 인센티브 장려 • 공공 이익을 위해 사적 자산 아카이빙을 위한 정책 지원

□ 영국 국가도서관

- 영국 국가도서관은 디지털 컬렉션을 보다 신뢰성 있게 보존하고 또 용이하게 이용할 수 있게 하는 인프라와 업무프로세스 구축을 위한 ‘디지털 보존전략 2017-2020’을 2017년에 발표하였다.

- 이번 2017년에 수립된 디지털 보존전략은 디지털 보존 활동 중 우선적으로 실행해야 하는 4가지 업무와 구현하기 위한 절차를 명확하게 제시하였다. 새로 규명된 업무는 ‘새로운 인프라 구축’, ‘디지털 데이터 전환’, ‘레포팅 매커니즘’, ‘이용자의 접근 기능 강화구’이다.
- 영국 국가도서관이 발표한 디지털 보존전략 보고서인 “Sustaining The Value: The British Library Digital Preservation Strategy 2017-2020”에서 다루고 있는 내용은 다음과 같다(British Library 2017).

1. 목적

본 보고서의 목적은 2017-2020에 이루어질 디지털 보존 활동을 위한 프레임워크를 제공하는 데 있다. 기존 2013-2016 활동에 대한 분석결과를 토대로 앞으로의 전략적 우선순위와 전략 이행과제, 이에 대한 직원의 역할과 책임을 제시한다.

2. 비전

디지털 보존 정책의 비전은 디지털 컬렉션을 미래 사용자가 지속적으로 이용할 수 있도록 신뢰할 수 있는 디지털 리포지터리 구축 및 디지털 보존에 대한 전체적 워크플로우를 제공하는 데 있다. 즉 디지털 지적 유산에 대한 미래 세대의 접근 보장이 디지털 보존의 비전이다.

3. 정의

- 디지털 컬렉션:

eBooks, eJournals, UK 웹 아카이브, 디지털 신문 및 원고, 디지털 지도, 디지털 음악, 선거 등록부, 특허, 개인 디지털 아카이브 등

- 우선적으로 수행하는 과제:

- ① 디지털 리포지터리 인프라를 새로 구축하거나 기존의 인프라를 강화하는 것
- ② 장기보존을 위한 기존의 디지털 컬렉션을 새로운 인프라로 전환시키는 것
- ③ 신뢰하는 보존처리 업무를 수행하고 이에 대한 증빙을 제공할 수 있는 레포팅 매커니즘을 개발하는 것
- ④ 콘텐츠에 도서관 이용자의 접근을 확고히 하는 것

- 도전과제:

- ① 디지털자원 생애주기 기반의 사전 예방적 관리
- ② 무결성 유지 및 검증
- ③ 저장매체 취약성 보완

4. 전략적 요소

- 리포지터리 인프라 대체 및 기술적 요소 강화: 새로운 저장 인프라 구축, 위험관리, 복제
- 장기보존을 위한 디지털 컬렉션 수집: 자동입수시스템 구축, 자동 메타데이터임베드
- 관리와 레포팅: 관리정보 생성, 담당자의 전문기술 강화, 문서화
- 이용자에게 적합한 접근: 사용성 및 신뢰성 제공

5. 기대되는 성과

4개의 핵심과제의 성공적인 구현

6. 전략 수행계획

- 정부의 지원
- 2017/18 수행계획:
 - ① 디지털저장소 시스템 기획,
 - ② 새 시스템으로의 마이그레이션
 - ③ 디지털 보존 정책 조정
- 2018/19 수행계획:
 - ① 디지털 저장소 시스템 구현,
 - ② 마이그레이션된 데이터 검증
 - ③ 새로운 시스템에 적합한 프로세스 개발
 - ④ 직원 및 이용자 교육
- 2019/20 수행계획:
 - ① 시스템 사용성 평가
 - ② 데이터 감사 및 인증
- 국제적 협조: 국가적인 메모리 기관, 학술기관 및 디지털 보존 협회, 오픈엑세스 재단, BitCurator 컨소시엄
- 평가

□ 캐나다 국가도서관과 기록관(LAC)

- 캐나다 국가도서관과 기록관(LAC)의 디지털 아카이브에서는 영구적으로 소장할 LAC의 디지털 컬렉션을 보존하고 다양한 소프트웨어와 하드웨어 그리고 운영시스템을 관리하고 있다. 2017년에 “디지털 보존 프로그램을 위한 전략(Strategy for a Digital Preservation Program)”을 수립하여 미래지향적인 전략적 대책을 마련하

였다(Library and Archives Canada 2017a).

- 디지털 보존 프로그램을 위한 전략에 포함된 내용은 다음과 같다.

1. 서론

LAC는 국가 디지털 지적유산을 장기적으로 보존하는 기관이며, 보존 센터인 LAC 디지털 아카이브는 LAC이 보유하고는 디지털 컬렉션의 저장소이다. LAC 디지털 아카이브의 임무는 텍스트, 시각자료, 오디오비디오 자료, 웹 콘텐츠 등과 같이 다양한 유형의 디지털자료를 장기적으로 보존하는 것이며 이와 동시에 디지털 장기보존을 지원해주는 소프트웨어, 하드웨어 및 운영 체제를 유지 관리하는 것이다.

2. 정의와 범위

디지털 보존은 “지속적인 액세스를 보장하기 위해 디지털 콘텐츠를 능동적이고 체계적으로 관리”하는 것으로 정의될 수 있다. 지속적인 액세스를 보장하기 위해 디지털 보존 담당직원은 기술적 노후로 인하여 생겨지는 문제점을 사전에 모니터링하고 이에 대한 대책 및 전략을 수립해야 한다.

기술적 절차 및 업무는 다음과 같은 표준에 준하여 수행된다.

- ISO 14721: 2012 Reference Model for an OAIS
- Digital Curation Center(DCC) Lifecycle Model
- Digital Preservation Solution(DPS) at LAC

3. 디지털 보존을 위한 비전

2024년까지 LAC는 ‘ISO 16363: 신뢰할 수 있는 디지털 리포지터리의 감사 및 인증’을 준수하는 지속가능한 디지털 보존 프로그램 수립과 이를 수행할 수 있는 신뢰할 수 있는 디지털 리포지터리를 구축할 것이다.

이 비전을 실현하기 위해 LAC는 다음과 같은 활동을 수행한다:

- 지속 가능한 기술적 테크니컬 솔루션 구현
- 디지털 컬렉션이 체계적으로 보존되도록 컬렉션 관리 프레임워크 수립
- ISO 16363 인증의 핵심 요구사항 인 프로그램 개발에 대한 증거데이터 제시
- 계획/실무/운영 프로세스의 체계화 및 프로그램의 조직 프레임워크 설계
- 충분한 인적 자원과 재정적 자원 확보

신뢰성 있는 디지털 장기보존을 위하여 LAC이 수행해야하는 단계는 다음과 같다.

- 1단계: 정보수집

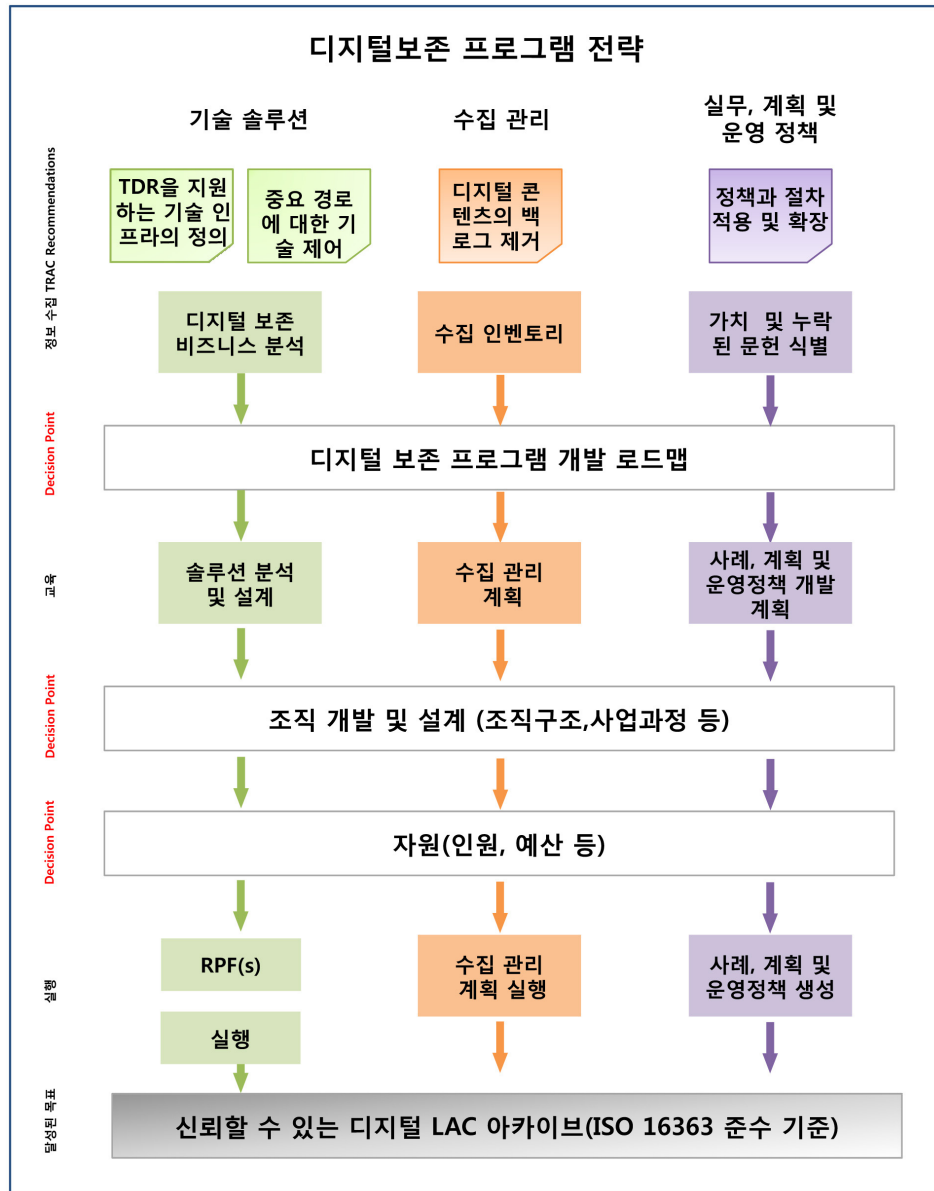
- ① 디지털 보존 니즈와 이슈, 실행시 발생하는 문제점, 관련 프로그램에 관한 정보
- ② 운영계획과 실제 실무와 절차간의 갭 분석에 관한 정보
- 2단계: 프로그램 개발
 - ① 포괄적으로 문제를 파악하고 이를 해결할 수 있는 시스템, 솔루션 개발
 - ② 디지털 보존 역량을 확보
- 3단계: 프로그램 수행
 - ① 지속가능한 디지털 보존 프로그램을 위한 목표 달성
 - ② 공유, 지원, 역량 파악

LAC은 디지털 보존을 위하여 다음과 작업을 정기적으로 수행해야 한다.

- 보존할 컬렉션의 백로그를 포함하여 디지털 보유량 확인
- 보존할 디지털 컬렉션의 책임자 확인
- 디지털 컬렉션의 소장 및 관리 장소 확인
- 보존되고 있는 컬렉션의 파일 형식 및 특성 확인
- 장기보존 여부 결정
- 영구 손실의 위험 평가

4. 결론

- 다음 [그림 4-1]은 LAC가 수립한 디지털 보존 프로그램 전략 프레임워크를 정리한 것이다.



[그림 4-1] 캐나다 국가도서관의 디지털 보존 프로그램 전략 프레임워크

□ 호주 국가도서관

- 호주 국가도서관은 2013년에 디지털 정보자원의 장기적인 보존을 위하여 디지털 컬렉션의 보존과 디지털 보존을 위한 다른 기관간의 협조를 위하여 디지털 보존 정책 4판을 수립하였다(National Library of Australia 2013).
- 디지털 보존 정책에 담겨진 내용은 다음과 같다.

1. 목적

보고서는 호주의 국가도서관에서 수집한 디지털자료의 보존과 다른기관에서 생산되고 있거나 호주 국가도서관에서 관리하는 디지털 정보자원의 보존을 위하여 작성된 정책으로 중요 디지털 보존 프로세스를 명확하게 하는데 그 목적이 있다. 또한 디지털 보존 정책은 도서관 일반 보존 정책 및 디지털자료의 생산, 선별, 획득, 기술 및 접근하는 방법과 관련된 기타 정책 문서 및 지침들과 상호 관련성이 있음을 밝힌다.

2. 도서관의 디지털 보존 업무의 목적

디지털 보존 역할은 현재는 물론 미래에도 지속적으로 사용할 수 있도록 디지털 컬렉션을 유지하는 것으로 도서관은 디지털정보자원의 수집 및 보존에 대한 막중하고 법적인 책임을 갖는다. 또한 도서관의 디지털 보존 업무는 디지털 컬렉션을 수집, 관리, 보존 및 유지할 수 있는 인프라를 개발하고 제공하는 데 그 목적이 있다.

이에 따라 도서관이 수행해야 하는 업무는 다음과 같다.

- 수집, 관리, 보존, 그리고 접근가능성 유지
- 비트(bit) 수준의 보존
- 진본성과 출처 유지
- 적합한 보존정보 유지
- 지속적 접근에 영향을 미치는 리스크에 대한 이해
- 적합한 보존전략 수행
- 포맷, 메타데이터 표준 리뷰

3. 도서관의 디지털 컬렉션 속성

호주 국가도서관은 1980년대 중반부터 지속적인 가치를 지닌 디지털정보자원을 획득하고 생산하기 시작하였고 다른 기관과 협약 하에 보존 관리하고 있는 디지털정보자원은 다음과 같다:

- PANDORA를 포함한 호주의 웹자원
- 디지털화된 자료의 master copies의 보존: 디지털화된 신문, 디지털화된 자료
- 호주의 디지털 출판물

- 오리지널 미출판된 디지털 파일
- 도서관의 보존자료(필사본, 사진, 지도 등)
- 이 정책은 다른 기관의 정책에서 관리되고 있는 구술역사 자료를 포함하지 않는다.

4. 디지털정보자원의 접근가능성에 영향을 미치는 요소

- 컬렉션 접근에 영향을 미치는 미디어 및 기술의 노후화
- 보존된 파일 형식의 다양성
- 디지털 컬렉션의 확장 및 증가
- 영구적으로 보존해야 하는 자료와 특정기간 또는 어느 시점에 폐기되는 자료와 함께 관리되는 환경
- 객체 구조, 상호관계, 매체 의존성 등 디지털자료의 복잡성
- 지적인 통제의 복잡성
- 보존 계획 및 처리 수준의 복잡성
- 보존 조치를 위한 일정표(timetable)의 수시 변경
- 다양한 보존 접근방식의 필요성
- 반복적인 변환 및 전환 필요
- 필요한 보존전략 조치 및 관련 기술의 제약성
- 도서관의 디지털 보존 프로그램은 이러한 문제를 해결하거나 최소한 고려해야한다.

5. 도서관의 디지털 컬렉션을 보존하기 위한 광범위한 지침

- 보존 접근성의 목표
 - ① 비트스트림 수준에서 신뢰할 수 있는 데이터에 대한 액세스 유지
 - ② 비트 스트림으로 인코딩된 콘텐츠에 대한 액세스 유지
 - ③ 콘텐츠의 맥락과 지적 의미에 대한 접근성 유지
 - ④ 디지털자원에 대해 명시된 장기보존의 목표를 지속적으로 유지
- 참고모델과 표준
 - ① OAIS(Open Archival Information Systems) 참조모델
 - ② 국제 표준 및 모범 사례에서 정의된 원칙(e.g. PREMIS, Open Planets Foundation)
 - ③ 유용한 새로운 접근법 또는 프로세스 모델 개발에 대해 지속적인 지원

6. 정책 이행

정책의 이행은 도서관이 제공하는 인프라와 비용 효율적인 솔루션의 가용성에 달려 있다.

- 도서관의 임무
 - ① 디지털 수집 자료에 대한 접근성 평가, 획득, 관리 및 유지와 관련된 비용 이해
 - ② 보존능력과 이를 위한 자원의 가용성을 고려하여 수집 프로그램 맞춤화 필요
 - ③ 디지털 컬렉션 보존에 관한 교육을 충분하게 받은 직원 확보
- 담당자의 인식

- ① 디지털 보존 업무가 획득단계에서 디지털자료의 분석 및 평가까지임을 인식
 - ② 컬렉션에서 보유하고 있는 것에 대한 이해
 - ③ 디지털 객체에 대한 보존 의도와 실제 업무에 대한 이해
 - ④ 접근제한 및 접근요건에 대한 도서관 지원수준에 대한 이해
 - ⑤ 디지털 보존의 위험정도 평가 및 모니터링에 대한 인식
 - ⑥ 지원 변경의 영향 예상
 - ⑦ 계획 트리거에 대한 인식과 변경 조치 계획
- 디지털 보존을 위한 노력
- ① 디지털 콘텐츠에 대한 구성 및 특성, 메타데이터 특성, 관리리스크 정리
 - ② 디지털 보존 위험도 평가 및 보존계획 및 조치의 우선순위 수립
 - ③ 적절한 인프라 구축을 위한 사전적 계획 수립
 - ④ 보존전략 기술에 대한 문서화: 파일 마이그레이션, 포맷 마이그레이션
 - ⑤ 다양한 수준의 소프트웨어 및 하드웨어 환경 시뮬레이션
 - ⑥ 적절한 소프트웨어 또는 하드웨어의 유지 보수
 - ⑦ 디지털 보존 활동에서 일어나는 문제점을 문서화

7. 연구와 표준 개발

도서관은 디지털 컬렉션을 효과적으로 관리하기 위해 이에 관련된 연구를 지속적으로 장려하고 수행 할 것이다. 도서관은 이해관계자들의 요구에 잠재적으로 이익이 될 수 있도록 호주 및 국제 표준 개발을 계속 지원할 것이다.

8. 국가의 디지털정보 자원을 보존하기 위해 다른 사람들과 협력

도서관은 국가 지식유산이 보존되고 접근 가능하도록 보장하는 데 노력을 기울려 왔다. 이를 위하여 호주의 디지털정보자원의 유산을 보존해야 할 책임이 있는 여러 기관들과 협력을 해야 한다.

□ 덴마크 왕립도서관

- 덴마크 왕립도서관은 납본된 디지털정보자원을 안정된 시스템에 저장하고 정기적으로 백업기능을 수행하면서 데이터의 손실이 되지 않도록 보존을 해왔다. 다만 백업은 단기적 디지털 보존전략 솔루션임을 인지하고, 2012년에는 디지털자원의 장기적인 보존 정책을 수립하였다(The Royal Library, Demark 2012).
- 특히 덴마크 왕립도서관은 새로운 형태의 디지털자료가 계속해서 생산되고 있고,

시스템/저장매체/데이터 포맷/프로그램이 계속해서 업그레이드 되거나 새로 개발되고 있으며, 디지털 보존문제를 해결할 수 있는 완성된 단 한가지의 방법과 솔루션이 없고, 디지털 보존 비용은 지속적으로 증가한다는 점을 인식하고 이러한 어려운 점을 해결하기 위해서는 장기보존을 위한 디지털자원 정책수립과 장기 디지털 보존을 위한 국제적 협력이 필요하다고 보았다.

- 디지털 보존 정책은 덴마크 왕립도서관에서 이미 수립한 ‘디지털자원 선별정책’과 ‘디지털자원 보존전략’을 따르고 있다.

□ 디지털 보존 정책에 포함되어야 할 내용

- 대표적인 국가도서관의 디지털 보존 정책을 살펴본 결과, 디지털 보존 정책에는 포함될 수 있는 세부적인 내용은 다음과 같다.
- 서론에는 ① 범위, ② 내용, ③ 과제/전략, ④ 필수사항, ⑤ 준수해야 할 표준 등이 포함될 수 있다.
- 실행부분에는 ① 보존 업무 계획, ② 선별과 평가, ③ 입수, ④ 메타데이터, ⑤ 접근과 이용, ⑥ 소장, 복제, 백업, ⑦ 검증, ⑧ 도큐멘테이션 등이 포함될 수 있다.
- 운영부분에는 ① 권리와 제한 관리, ② 역할과 책임, ③ 지속가능성 계획, ④ 보안관리 등이 포함될 수 있다.
- 지원부분에는 ① 협력, ② 직원교육 등이 포함될 수 있다.

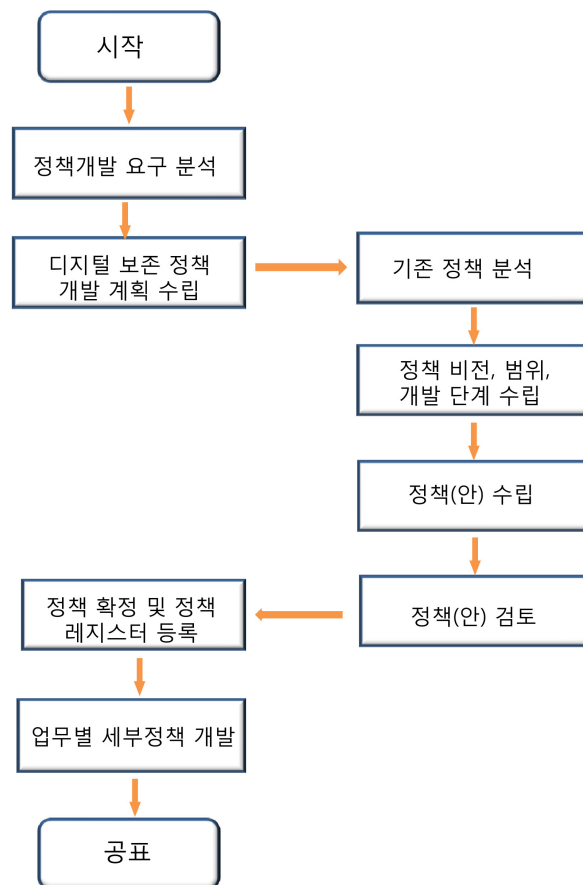
4.1.2 디지털자원 보존 정책 개발을 위한 가이드라인

□ 디지털자원 보존 정책 가이드라인 사례

- 디지털 보존을 체계적으로 또는 원하는 수준으로 이끌어가기 위해서 우선적으로 수행되어야 하는 작업이 디지털 보존 정책을 수립하는 것이다.
- 정책에는 디지털 보존을 수행하는데 실무적 솔루션의 방향이 될 수 있는 일관성이

있으며 빈틈없고 자세한 요건이 기술되어야 할 뿐만 아니라, 조직과 정책의 비전과 목적을 반영하는 원칙과 가이드라인이 기술되어야 한다. 또한 정책은 반드시 지속성(longevity), 효과성(effectiveness), 명료성(clarity), 실용성(practicality)을 가져야 한다.

- Adrian Brown(2013)의 “Practical Digital Preservation”은 디지털 보존 정책에는 다음과 같은 내용 즉 1) 목적, 2) 배경, 3) 범위, 4) 정책 원칙, 5) 디지털 보존을 위한 요건(생성/선별/수집/보존전략/접근과 재사용/인프라), 6) 표준, 7) 역할과 책임, 8) 협력과 소통, 9) 모니터링, 10) 개정, 11) 용어정의가 담겨져야 한다고 권고하였다. 다음의 [그림 4-2]는 정책수립 절차를 보여준다.



[그림 4-2] 정책수립 절차

- 캐나다의 역사와 유산을 담당하는 정부기관에서는 디지털 리포지터리 또는 디지털 보존 프로그램을 운영하는 기관에서 디지털 보존 정책을 수립하는데 도움을 줄 수 있는 디지털 보존 정책 개발 가이드라인(Digital Preservation Policy Framework: Development Guideline ver. 2.1)을 제공하였다(Library and Archives Canada 2017b). [표 4-2]는 디지털 보존 정책 가이드라인에 수록된 내용의 목차이다.

[표 4-2] 캐나다 국가도서관이 제시한 디지털 보존 정책 가이드라인 목차

항목	세부항목
1. 비전과 배경	- 궁극적 목적 - OAIIS 참조모델 준수
2. 관리 책임성	- 목적 - 위임사항: 법, 규정, 정책, 사명 - 실질적 목적: 전략적 계획, 예산, 보존계획, 기술적 계획
3. 실행가능성	- 범위 - 운영원칙 - 역할과 책임 - 선별과 입수 - 접근과 이용 - 도전과 리스크
4. 재정 지속가능성	- 담당기관 지원 - 협력기관 지원
5. 기술적/절차적 적합성	- 기술요건 - 절차적 요건
6. 시스템 보안	- 보안 - 접근권한
7. 절차적 책임	- 투명성 - 정책관리

- 2013년에 영국에서도 도서관의 보존전략을 수립하기 위한 가이드라인을 제시하였다. 영국 국가도서관 소속의 보존자문센터(Preservation Advisory Centre)는 디지털 보존전략을 수립하기 위해서는 기존에 수립된 전략은 물론이고 다른 기관이 수립한 전략 표준 등을 참조하여 수립해야 한다고 하였다. [표 4-3]과 같이 보존전략 가이드라인에는 보존전략에 담겨야 할 내용과 체크리스트에 담겨질 내용이 상세히

제시되고 있다(Building a Preservation Policy 2013).

[표 4-3] 디지털 보존 정책에 수록되어야 할 내용에 대한 가이드라인(영국)

구분	항목	내용
보존 정책	1. 목적	- 기관이 추구해야 할 철학과 비전
	2. 컬렉션의 목적	- 자료의 범위 - 자료의 형태 - 자료의 규모
	3. 기관의 기능	- 기관이 수행하는 모든 업무에 대한 기술
	4. 입수	- 입수시 보존과 관련된 결정해야 사항 제시 - 입수시 자료의 특징 문서화
	5. 보유	- 보유기간 결정 - 보존 매체 유형 및 포맷 결정
	6. 접근	- 접근도구 및 인터페이스 - 액세스 환경
	7. 이용	- 이용자의 요구 - 이용현황
	8. 협력	- 이해관계자 정립 - 보존협력 기관
	9. 컬렉션 평가	- 보존 가치평가 - 매체 평가
	10. 리스크 및 보존요구 평가	- 위험관리 - 보존에 대한 니즈 평가
체크 리스트	1. 기관의 목적과 기능	- 기관의 목적 - 보존 정책의 필요성 - 보존 정책의 배경
	2. 정의	- 보존 전 정책에서 사용되는 용어에 대한 정의
	3. 범위	- 보존 정책이 포괄해야 하는 대상: 컬렉션, 매체 - 보존 정책의 유효기간
	4. 보존철학, 표준, 윤리	- 보존 정책의 근간이 되는 철학 및 윤리 - 보존 정책이 준수해야 하는 표준
	5. 접근	- 접근 대상 - 접근 정책
	6. 보안	- 물리적 보안 - 시스템적 보안
	7. 매체전환 및 리포맷	- 매체전환 시기 및 방법 - 리포맷, 이전 시기 및 방법

구분	항목	내용
	8. 전시 및 대여	- 전시 및 대여에 대한 정책
	9. 위험평가	- 이용 시 위험평가 - 보존전략의 기술 리스크 평가
	10. 기관서비스	- 저작권문제 - 전송/다운로드 문제
	11. 저장	- 저장 기술 - 디지털 리포지터리
	12. 환경	- 물리적 환경 - 기술적 환경
	13. 교육	- 담당자 교육 - 이해관계자 교육
	14. 예산	- 예산에 대한 이해와 고려
	15. 책임 및 모니터링	- 정책 관리 책임 - 재조정

□ 디지털자원 보존 정책 수립시 고려해야 할 리스크

- NDIIPP보고서는 국가 차원의 디지털 보존을 수행할 때 참작해야 할 주요 리스크를 규명하였고 리스크 관리는 장기적 보존 시 이루어지는 모든 활동, 모든 단계에서 고려되어야 함을 강조하고 있다. 이러한 리스크에 대한 정확한 인식과 솔루션에 대한 활동은 디지털 보존 정책에서도 포함되어야 한다고 본다. 따라서 디지털 보존 정책을 수립할 때 고려되어야 할 리스크는 다음과 같다.
- 기술차원의 리스크(Technology Risks)
 - 하드웨어와 소프트웨어: 사용 소프트웨어나 오픈소스 모든 경우 하드웨어와 소프트웨어 현 시점의 구동성을 확보하는 문제점
 - 포맷: 내용 포맷을 설명하는 정보의 부재로 인한 보존자료 활용성 저하
 - 생애주기 기반의 기록관리 문제: 데이터 마이그레이션과 파일 디그레이션 등 자료변환시 문제
- 법률 및 정책적 차원의 리스크(Legal and Policy Risks)

-
- 저작권법: 도서관이 보존을 목적으로 복사본을 만드는 과정에서 발생할 수 있는 저작권 침해사항
 - 개인정보보호: 보존 콘텐츠 내용에 포함된 개인정보와 관련하여 발생할 수 있는 법률적, 윤리적 문제점
 - 회계법률 개정으로 인한 영향: 미국의 회계 개혁에 관한 연방법률(사베인스 옥슬리법)의 영향으로 인해 회사나 개인기관에서 회사나 개인의 이익을 위해 역사적으로 중요한 기록물을 폐기하는 현상이 발생
 - 법률 미비: 현행 법률상 디지털 기록보존의 사회적 의의와 가치에 대한 인식과 반영이 부족함
- 콘텐츠 차원의 리스크(Contents Risks)
 - 콘텐츠 무결성: 디지털 보존 시 콘텐츠의 양과 복잡한 구조로 인하여 누락되는 내용이 발생하는 문제점
 - 메타데이터 보존: 보존되는 디지털 콘텐츠에 대한 메타데이터 부족으로 검색 및 활용 시 발생하는 문제점
 - 조직 차원의 리스크(Organizational Risks)
 - 자원 부족: 디지털 보존을 위한 인적, 물적 자원 부족
 - 책임 권한 체계 부족: 디지털 보존을 위한 조직 내 체계가 설정되지 못하였을 때 발생할 수 있는 문제점
 - 전문 인력 부족: 디지털 보존 업무를 위한 숙련된 인력부족
 - 이에 국립중앙도서관에서 제도적 차원에서 고려해야할 리스크는 법률 및 정책과 조직 차원에서 발생하는 문제점이 해당되며, 콘텐츠 차원의 리스크는 보존 기록 수집 범위 설정 및 보존을 위한 메타데이터 개발 측면에서 고려되어야 할 요소이다. 또한 기술 차원의 리스크는 디지털자원의 장기보존 실효성을 보장하는 중요한 요소로서 국가차원에서 중장기 전략을 수립해야할 시 분석, 평가하여야 할 사항이다.

□ 디지털 보존 정책수립을 위한 체크리스트

1) 운영 면에서

- 선별/수집:

- ① 보존할 가치가 있는 자료인가?
- ② 디지털 보존 인프라에 적합한 디지털자료인가?

- 역할과 책임

- ① 디지털 객체에 대한 정보가 있는가?
- ② 디지털 객체 접근에 대한 책임과 제한이 명시적인가?

- 보유

- ① 디지털 객체 유형에 따른 관리체계가 수립 되었는가?
- ② 보유 매체 및 관리에 대한 전문가가 있는가?
- ③ 기관의 목적에 디지털 객체가 적합한가?

2) 매체 면에서

- 매체 유형

- ① 디지털 객체의 미디어 및 하드웨어의 벤더가 다양한가?
- ② 기술적 환경의 변화에도 지속적으로 매체유지가 가능한가?
- ③ 매체가 견고한가?
- ④ 매체유지 조건은?

- 매체관리

- ① 매체의 무결성과 가독성 체크 기간은?
- ② 외부 복본에 대한 관리는?
- ③ 매체에 대한 이력관리는?

3) 포맷 면에서

- 보존포맷 선택

- ① 포맷이 일반적인 뷰어와 에디터를 지원하는가?
- ② 오픈 포맷이거나 포맷에 대한 명세서가 있는가?
- ③ 원본이 가지는 특성을 포맷이 지원하는가?
- ④ 포맷이 압축과 암호화를 지원하는가?

- 파일 관리

- ① 파일이 생성 시에 사용된 소프트웨어가 있는가?
- ② 파일 에뮬레이션과 마이그레이션을 얼마나 자주하는가?
- ③ 포맷의 속성의 손실이 데이터의 손실인가?

4) 메타데이터 면에서

- 일반

- ① 국제표준 메타데이터를 사용하고 있는가?
- ② 국가 수준 표준을 사용하는 이유는?

- 자원기술

- ① 디지털 객체에 대한 기술이 표준이나 국내요건에 따르는가?
- ② 상호호환성을 따르는가?
- ③ 선택한 표준이 데이터 애그리게이션 시에도 적합한가?

- 관리/보존

- ① 진본성 지표가 될 수 있는 것을 기술하고 있는가?
- ② 이력정보가 기술되고 있는가?

- 지속성 식별자

- ① 유일성을 보장하는 식별자가 있는가?
- ② 식별시스템을 활용하고 있는가?

5) 저작권 면에서

- ① 저작권 정보가 있는가?
- ② 저작권 보호와 관련된 정보가 명시되고 있는가?
- ③ 이용제한에 대한 정책이 있는가?
- ④ 영구적 보존에 대한 권리를 가지고 있는가?
- ⑤ 디지털 장기보존 및 접근을 위한 권리확보에 대한 비용이 적합한가?

4.1.3 국립중앙도서관 디지털자원 보존 정책(안)

- 국립중앙도서관이 효율적이며 체계적으로 디지털자원을 보존하기 위해서는 디지털 보존에 대한 비전과 업무내용이 포괄적으로 담겨진 정책이 필요하다. 이에 따라 본 연구는 국립중앙도서관 디지털자원 보존 정책을 다음과 같이 제안하였다.
- 디지털 보존 정책에 담겨진 내용은 다음과 같다.

1. 서문

1.1 디지털 보존을 위한 비전

- 디지털 보존은 디지털 지적 유산에 대한 미래 세대의 접근 보장의 근간이 되는 작업으로 미래 사용자가 아무런 어려움이 없이 현재 구축된 디지털자원을 이용할 수 있도록 하는 도서관의 핵심 활동이다. 이러한 디지털 보존은 디지털자료 형태의 인류사회 기억을 수집·보존·전승하여 「헌법」 제21조에 규정된 ‘국민의 알 권리’를 실현하고 지적 자유를 수호하며 지식정보에 대한 접근 이용의 평등권을 확립하는데 기여한다.

1.2 디지털 보존 정책의 목적

- 디지털 보존 정책의 기본적 목적은 국립중앙도서관(이하 도서관)에서 수집한 디지털 자원 및 디지털화된 자료를 장기적으로 보존하기 위한 전체적 업무활동과 중요 디지털 보존 프로세스를 명확하게 제시함으로써 디지털 보존의 일관성과 지속성을 확보하는 것이다.
- 또한 디지털 보존 정책은 디지털자원의 보존관리 및 지속적 활용에 대한 도서관의 의지의 표명, 기본지침에 근거한 체계적인 관리, 그리고 성문화를 통한 법적 책임과 역할을 대내외 알리는 데 그 목적이 있다.

1.3 디지털 보존 정책의 범위

- 디지털 보존 정책

디지털 보존은 “지속적인 접근과 활용을 보장하기 위해 디지털자원을 신뢰성 있게 보유하고 체계적으로 관리”하는 것이다. 따라서 디지털 보존 정책은 디지털자원의 체계적인 보존관리를 위한 논리적 토대를 제공하여야 하며, 이와 동시에 현실적/실무적 지침으로 활용될 수 있어야 한다. 이에 따라 본 정책은 디지털 보존 정책의 비전과 목적, 디지털 보존 업무 지침, 정책이행을 위한 노력, 책임사항과 협력, 디지털 보존 정책의 관리로 구성되어 있다.

- 디지털 보존의 대상이 되는 디지털자원

- 도서관에 납본되어 저장된 온라인 자료(대한민국의 디지털 출판물)
- 도서관에서 디지털화된 자료
- 수집된 웹자원

2. 디지털 보존 업무 지침

- 디지털 보존 업무의 목적은 현재는 물론 미래에도 지속적으로 사용할 수 있도록 디지털자원을 유지 관리하고 장기적으로 보존하는 것이다. 이에 도서관은 디지털자원의 수집 및 보존에 대한 막중한 법적 책임을 갖고 디지털자원을 수집, 관리, 보존 및 유지하는 일련의 디지털 보존 업무를 수행하여야 한다.

2.1 디지털 보존 담당조직

- 업무내용

- 업무활동별 담당 직원 및 업무시간 배정
- 적절한 예산 배정
- 인력개발 요구사항 파악
- 지원부서 및 지원 비용 파악

- 요건수준

- 기본수준:
 - 직원 업무가 적합하게 할당되고 이에 대한 책임이 부여되었다.
 - 활동에 적합한 예산이 부여되었다.
 - 직원개발에 관한 사항이 명시되고 예산이 부여되었다.
- 관리수준:
 - 예산 및 직원역할 개발에 대한 니즈분석이 정기적으로 또 환경의 변화에 따라 평가되고 있다.
- 최적수준:
 - 예산 및 직원역할 개발에 대한 니즈분석이 미래 변화를 예측하여 사전 대비적으

- 로 평가되고 있다.
- 직원개발의 효율성이 정기적으로 모니터링되고 있다.

2.2 이해관계자 관리

- 업무내용
 - 주요 이해관계자 파악과 참여 지원
 - 이해관계자와의 의사소통을 하는 목적에 대한 인식
- 요건수준
 - 기본수준:
 - 담당자는 주요 이해관계자를 파악하고 있다.
 - 담당자는 이해관계자와의 소통을 하는 목적과 방법을 알고 있다.
 - 관리수준:
 - 담당자는 이해관계자와 정기적으로 소통하고 있다.
 - 최적수준:
 - 이해관계자가 사전 대비적으로 관련된 정보를 제공하고 있다
 - 이해관계자가 디지털 아카이빙에 적극적으로 참여하고 있다.

2.3 법적 이슈 관리

- 업무내용
 - 저작권 파악
 - 디지털 보존과 관련된 주요 법적 권리와 책임 인식
- 요건수준
 - 기본수준:
 - 담당자는 주요 법적 권리 및 책임을 파악하고 있다.
 - 관리수준:
 - 담당자는 법적 이슈를 정기적으로 점검하고 있다.
 - 담당자는 발생된 새로운 이슈에 대한 처리를 즉시 수행하고 있다.
 - 최적수준:
 - 법적 이슈가 사전 대비적으로 모니터링되고 조치되고 있다.

2.4 정책 프레임워크 관리

- 업무내용
 - 디지털 보존 정책의 성문화
 - 디지털 보존 정책의 모니터링과 갱신
- 요건수준

- 기본수준:
 - 디지털 보존에 관한 인증된 성문화 정책이 있다.
- 관리수준:
 - 도서관에서 수립한 다른 정책과 같이 디지털 보존 정책이 정기적으로 점검 및 리뷰되고 있다.
- 최적수준:
 - 디지털 보존 정책은 조직 내의 변화 또는 외부 환경의 변화를 영하여 모니터링되고 있고 사전 대비적으로 개정되고 있다.

2.5 수집

- 업무내용
 - 수집 정책 수립 및 문서화
 - 수집·납본 절차 및 납본 대상 확립
 - 시스템에 의한 자동 입수도구 개발
- 요건수준
 - 기본수준:
 - 수집 및 납본대상을 정확히 정의하는 세부정책이 있다.
 - 문서화된 수집·납본 절차가 수립되어 있다.
 - 납본절차를 도와주는 지원도구가 개발되어 있다.
 - 관리수준:
 - 수집정책이 정기적으로 점검되고 개정되고 있다.
 - 납본을 위한 세부지침이 제공되고 있다.
 - 입수와 관련된 프로세스가 자동화되어 있다.
 - 최적수준:
 - 디지털자원 납본 프로세스와 리포지터리 저장과정이 연계되어 있다.
 - 입수와 관련된 프로세스가 하나의 시스템으로 이루어지고 있다.

2.6 비트스트림 보존

- 업무내용
 - 디지털 매체별 전용 저장공간 확보
 - 디지털자원에 대한 복제 및 복제본 관리
 - 데이터 무결성 및 바이러스에 대한 주기적인 점검
 - 액세스 제어 및 보안 프로세스 관리
- 요건수준
 - 기본수준:

- 디지털자원의 매체별 특성에 따라 저장 공간이 확보되어 있다.
- 디지털자원은 복제되고 이동성 매체로 백업되고 있다.
- 무결성 체크가 주기적으로 수행되고 있다.
- 바이러스 체크가 주기적으로 수행되고 있다.
- 접근통제와 보안 프로세스가 적용되고 있다.
- 관리수준:
 - 자동적으로 복제 및 백업이 진행되고 있다.
 - 디지털자원의 내용을 관리하기 위하여 리포지터리 플랫폼이 사용되고 있다.
 - 자동 무결성 검증 및 바이러스 체크 시스템이 사용되고 있다.
 - 리포지터리 보안 프로세스가 운영되고 있다.
- 최적수준:
 - 계층적 저장관리 시스템이 운영되고 있다.
 - 사전 대비적으로 보안과 접근통제 프로세스가 모니터링 되고 있다.
 - 사전 대비적으로 비트스트림 보존을 위협하는 요인을 확인하고 조치를 취하고 있다.

2.7 장기보존 기술 및 전략 수립

- 업무내용
 - 보존 기술 및 전략 계획 수립
 - 보존전략 절차 수립
 - 포맷 레지스트리 관리
 - 디지털 객체의 유형에 따른 다각적 관리방식 개발
- 요건수준
 - 기본수준:
 - 각각의 보존 기술 및 전략의 특징과 방법을 인식하고 있다.
 - 사안별 보존계획이 수립되어 있다.
 - 필수적으로 수행되어야 하는 보존활동이 사안별로 명시되어 있다.
 - 관리수준:
 - 장기보존 기술이 자동적으로 그리고 적합하게 수행되고 있다.
 - 디지털 객체의 사용성을 위협하는 위험요인을 인지하고 적합한 방법으로 조치를 취하고 있다.
 - 최적수준:
 - 보존 기술 적용 시에 발생하는 위험정도를 인식하고 모니터링하고 있다.
 - 디지털 장기보존의 위험요인과 문제점을 예상하고 이에 대한 대비를 사전에 준비하고 있다.

2.8 메타데이터 관리

- 업무내용
 - 보존 관련 필수 메타데이터 요소 확립 및 문서화
 - 보존 메타데이터 관리
 - 디지털 객체 식별자 관리
- 요건수준
 - 기본수준:
 - 문서화된 보존 메타데이터 요소 리스트가 있다.
 - 일관성 있게 디지털 객체와 메타데이터에 접근할 수 있다.
 - 메타데이터가 별도로 저장되어 있다.
 - 관리수준:
 - 메타데이터의 내용이 일관성 있게 기술되어 있다.
 - 메타데이터가 체계적으로 관리되고 있다.
 - 최적수준:
 - 메타데이터 관리가 디지털 보존 시스템의 하나의 기능으로 수행되고 있다.

2.9 활용을 위한 접근기능 관리

- 업무내용
 - 이용자 맞춤 검색을 위한 다양한 검색도구 개발
 - 디지털자원 내용 및 메타데이터 보기(view) 기능과 다운로드 기능을 제공하기 위한 제반 업무
- 요건수준
 - 기본수준:
 - 검색도구가 제공되고 있다.
 - 이용자는 온라인으로 디지털자원을 다운로드하거나 보기기능을 통하여 볼 수 있다.
 - 관리수준:
 - 다양한 접근 옵션이 제공되고 있다.
 - 접근방식 및 기능이 이용자 피드백 기반의 분석을 통하여 향상되고 있다.
 - 최적수준:
 - 최신의 검색환경이 제공되고 있다.
 - 담당자는 이용자의 니즈와 기대를 맞추기 위하여 이용자 커뮤니티와 주기적으로 소통한다.

2.10 기반시설 관리

- 업무내용
 - 디지털 리포지터리 관리
 - 디지털 보존에 적합한 IT 시스템 선정 및 관리
 - 디지털자원 증가에 따른 향후 리포지터리 확충 및 관리 계획 수립
- 요건수준
 - 기본수준:
 - 리포지터리의 소장 용량이 충분하다.
 - IT 시스템이 현재의 디지털 보존 활동에 적합하다.
 - 관리수준:
 - 담당자는 향후 필요한 리포지터리 용량과 기능을 예상하고 있다.
 - IT 시스템이 정기적으로 보완되거나 업데이트되고 있다.
 - 최적수준:
 - 향후 IT 시스템의 개발을 위한 상세한 로드맵이 수립되어 있다.
 - 새로운 도구나 시스템이 사전 대비적으로 파악되고 준비되고 있다.

3. 정책 이행을 위한 노력

3.1 담당 부서의 임무

- 행정적 지원
 - 수집된 디지털자원 자료에 대한 평가·획득·관리 및 유지와 관련된 비용 확보
 - 기관의 보존능력과 이를 위한 자원의 가용성을 고려하여 디지털자원 자동 수집 및 관리 프로그램 개발
 - 디지털 보존 업무의 일관성과 장기적 관리 확보를 위하여 디지털 보존과 관련 문서화된 정책 수립
 - 디지털자원을 보존하는 데 필요한 프로세스를 관리할 수 있도록 디지털 보존에 관한 교육을 충분하게 받은 직원 확보 및 지속적인 교육 실시
- 담당자 인식
 - 디지털 보존 업무가 획득 단계에서 디지털자료의 분석 및 평가까지 임을 인식
 - 보유하고 있는 컬렉션에 대한 이해
 - 디지털 객체에 대한 보존 의도와 실제 업무에 대한 이해
 - 접근통제 및 접근요건에 대한 도서관 지원수준에 대한 이해
 - 디지털 보존의 위험정도 평가 및 모니터링에 대한 인식
 - 지원 변경의 영향 예상
 - 계획 트리거에 대한 인식과 변경 조치에 대한 이해
- 디지털 보존 업무 시에 고려해야 하는 사항

- 사용하고 있는 디지털 매체와 기술은 점점 노후화되고 있다.
- 보존되는 파일의 형식이 점점 다양해지고 있다.
- 디지털 컬렉션은 지속적으로 증가되고 있다.
- 영구보존 자료와 어느 시점에 폐기되는 자료가 함께 관리되고 있다.
- 디지털자원의 객체 구조, 구현 매체, 구현 HW/SW 등이 복잡하다.
- 반복적으로 보존 기술과 전략이 사용되며 그 처리 수준이 복잡하다.
- 보존 기술 적용을 위한 계획 및 일정이 수시로 변경될 수 있다.
- 필요한 보존전략 조치방식 및 관련 기술이 여전히 완벽하지 않다.

3.2 연구와 표준개발

- 도서관은 디지털자원을 효과적으로 관리하기 위해 이에 관련된 연구를 지속적으로 장려하고 수행해야 한다. 또한 도서관은 이해관계자들의 요구를 반영하고 디지털 보존의 장기적인 효과를 가질 수 있는 국내표준을 개발해야 하며 이와 동시에 디지털 보존 관련 국제표준 개발에도 지원해야 한다.

3.3 저작권

- 디지털 보존 및 디지털화는 모두 콘텐츠 복제와 관련이 있으므로 보존 및 활용을 위한 복제와 관련된 모든 것은 저작권법을 따른다. 이에 따라 도서관은 지적 재산권 및 콘텐츠의 복사, 저장, 수정 및 디지털 기록 사용과 관련된 기타 법적 및 윤리적 권리를 준수해야 한다.

4. 책임사항과 지원

4.1 역할과 책임

- 디지털 보존을 위한 비전을 실현하기 위해 도서관의 디지털 보존 부서는 다음과 같은 책임과 업무를 수행해야 한다:
 - 디지털자원의 체계적 보존을 위한 디지털 컬렉션 관리 프레임워크 수립
 - 지속 가능한 보존전략 기술솔루션 구현
 - 디지털 보존 업무의 효율적 수행을 위한 디지털 보존 조직 프레임워크 설계
 - 디지털 보존 위험관리 및 문제점 해결을 위한 매커니즘 개발
 - 신뢰성 있는 디지털 장기보존을 위한 활동/이력 레포팅 매커니즘 개발
 - 충분한 인적 자원과 재정적 자원 확보
- 이에 디지털 보존 부서와 담당자는 'ISO 14721: 2012 Reference Model for an OAIS'와 'ISO 16363: 신뢰할 수 있는 디지털 리포지터리의 감사 및 인증' 등과 같은 국제표준을 준수하기 위하여 조직의 지원체제 수립, 지속가능한 디지털 보존 프로그램 수립, 신뢰할 수 있는 디지털 리포지터리 구축에 대한 임무를 가진다.

4.2 교육

- 도서관은 디지털 보존과 관련된 지식과 기술을 갖춘 직원이 디지털 보존 활동을 수행하도록 보장해야 한다. 이를 위하여 도서관은 담당자들이 들이 디지털 보존 전문 지식을 익히고 향상시킬 수 있는 교육 기회를 제공해야 한다. 이러한 기회에는 자기 주도적 학습, 강좌 참가, 국내외 세미나 참석, 워크숍 및 컨퍼런스, 해외 기관 방문, 전문 단체와의 업무 교류 등이 포함된다.

4.3 협력

- 도서관은 국가의 모든 디지털 지식유산이 장기적으로 그리고 신뢰성 있게 보존되도록 노력해야 한다. 이를 위하여 도서관은 디지털자원 생산기관, 디지털자원 출판업체, 디지털자원의 유산을 보존해야 할 책임이 있는 여러 기관들과 협력을 강화한다.
- 디지털 보존을 효율적으로 수행하기 위해서는 보존 기술 및 전략과 관련된 전문 기술 개발과 공유기능을 강화하기 위하여 관련 전문 업체와 협력한다.
- 디지털 보존 시 수행되는 모든 업무는 디지털 보존 기술과 환경의 상호호환성을 높이기 위해서는 국제적 표준을 준수하여야 한다. 이와 동시에 도서관은 디지털 보존과 관련된 국제표준기구 및 보존과 관련된 국제기구들과의 국제적인 교류 및 협력을 강화한다.

5. 디지털 보존 정책의 개정

- 디지털 보존 정책은 디지털 보존 업무와 이에 관련된 보존 활동 수준을 규정한 문서이므로 실무수행의 일관성, 합리성, 현실성, 실용성을 담보해야 한다. 또한 디지털 보존은 디지털 환경 및 기술의 변화에 직접적인 영향을 받는다. 따라서 도서관은 디지털 보존 정책을 3년 주기로 디지털 환경, 디지털자원의 생산·유통 패러다임, 디지털자원의 이용자 요구 및 이용행태 등의 변화 등에 대한 분석결과 내용을 반영하여 수정 보완해야 한다.

- 이와 같이 국립중앙도서관에서 사용할 수 있는 디지털 보존 정책(안)을 제시하였다. 다만, 최근에 발표한 국가도서관의 디지털 보존 정책을 살펴본 결과, 국가도서관에서 업무에서 활용되는 디지털자원의 장기보존 정책이라기보다는 국가차원 가치가 있는 디지털자원을 선별 입수하여 소장하여 장기적으로 이용자에게 접근할 수 있는 모든 활동을 포함하는 거시적 국가 정책 및 전략내용이 주를 이룬다는 점을 파악할 수 있었다. 특히 2010년대에 들어와서는 단기적 보다는 장기적 보존전략에 그 비중을 더 주고 있고 이러한 디지털 보존이 한 기관에서 이루어질 수 없다는

점을 인식하고 있는 것 같다.

- 따라서 국립중앙도서관은 디지털 보존 정책을 지속적으로 발전시켜야 할 것이다. 특히 1단계의 정책에서 제시된 내부적 업무와 기술기반의 전략이 어느 정도 확고히 자리를 잡힌 다음에는 2단계의 정책 즉 디지털 국가유산의 영구적 보존에 대한 국가차원의 비전과 목적달성을 위한 장기적 디지털 보존전략이 수립되어야 하고 이에 따른 각국의 국내 협력기관은 물론 다른 나라와의 협력도 시도하고 이에 대한 정책이 포함되어 제시되어야 할 것이다.

4.2 디지털자원 보존 환경 및 시스템 지침

4.2.1 디지털 보존을 위한 표준 및 시스템 기능 요건 현황

□ 디지털 보존 관련 표준

1) 운영 표준

- 디지털 보존을 위한 운영 모델의 개발을 도울 수 있는 많은 표준이 있다. 디지털 자료를 관리하려면 디지털 보존 환경으로의 이전을 관리하는 일련의 절차가 필요하다. 여기에는 이관될 자료를 식별하고 정량화하고, 보존하는 비용을 평가하고, 미래의 인증 및 기밀성을 위한 요구 사항을 파악하는 것을 포함한다.
- ISO 20652: 2006 공간 데이터 및 정보 전송 시스템 - 생산자 아카이브 인터페이스 - 방법론 요약 표준(ISO, 2006)은 작성자가 디지털 보존 환경으로 디지털자료를 공식적으로 전송하는 절차를 개발하기 위한 방법론적 프레임워크를 제공하는 국제 표준이다. 목표, 행동 및 예상 결과는 4단계로 식별되는데, 생산자의 초기 협상(사전 단계), 요구 사항 정의(공식 정의 단계), 디지털자료의 디지털 보존 환경으로의 이전(전송 단계) 및 디지털자료 확보 및 관련 메타데이터는 동의한 내용(검증 단계)을 준수하도록 보장한다.
- ISO/TR 18492: 2005 전자 문서 기반 정보의 장기보존(ISO/TR, 2005)은 매체 갱신에 관한 기술 중립적 지침을 포함하는 진정한 전자 문서 기반 정보의 지속적인 보존 및 검색을 위한 실용적인 방법론을 제공한다(마이그레이션, 품질, 보안 및 환경 제어 등). 이 지침은 원본 정보 보관 시스템의 수명을 넘어서는 기록의 신뢰성을 보장하기 위해 개발되었다.
- ISO 14721: 2012 공간 데이터 및 정보 전송 시스템 - 개방형 아카이브 정보 시스템 - 참조모델 (OAIS)(ISO, 2012)은 장기 디지털정보 보존 및 액세스에 필요한 아카이브 개념을 이해하고 구현하기 위한 체계적인 프레임워크를 제공한다. 기존 및 미래 아카이브의 아키텍처와 운영을 설명하고 비교하며, 장기보존을 위한 역할, 과정

및 방법을 설명한다.

- ISO 15489: 2016 정보 및 문서화 - 기록 관리 (ISO, 2016)는 장기보존 기록을 중심으로 디지털 보존 구현을 위한 역할, 프로세스 및 방법을 정의할 때 유용한 표준이 될 수 있다. 이 표준은 비즈니스 레코드가 신뢰할 수 있고 액세스 가능한 상태로 유지되는 동안 비즈니스 레코드가 큐레이팅되고 문서화되도록 비즈니스 레코드를 관리하기 위한 모범 사례의 프레임워크를 간략하게 설명한다.
- ISO 16175: 2011 전자 사무 환경의 레코드에 대한 원칙 및 기능 요구 사항 (ISO, 2011)은 전자 문서 및 레코드 관리 시스템은 물론 콘텐츠 관리 시스템과 관련된다. 디지털 보존에 대한 특정 요구 사항은 포함되어 있지 않지만 시간이 지남에 따라 기록을 유지할 필요가 있음을 인정하고 이러한 전자 시스템 사양에서 포맷 노후화 문제를 고려하도록 한다.

2) 기술 표준

- 주로 메타데이터 및 파일 형식과 관련하여 디지털 보존 프로그램의 기술적 측면에서 표준을 적용하면 특정한 이점이 있다.
- 관련 설명 메타데이터 표준과 관련하여 PREMIS 및 METS는 사실상 디지털 보존 프로그램을 향상시키는 표준이다. PREMIS는 미국 의회도서관에서 주도하여 2005년에 처음 발표된 표준으로 디지털자료의 보존을 지원하기 위해 데이터 사전 및 지원 도구를 특별히 개발하였다. METS는 디지털자료를 보관 정보와 함께 패키징할 수 있는 XML 인코딩 표준이다.

3) ISO 14721:2012 OAIS 참조모델

- ISO 14721: 2012 Space Data and Information Transfer System - Open Archival Information System- Reference Model (OAIS)
- NASA의 CCSDS(Consultative Committee for Space Data Systems)는 일반적인 데이터 시스템에 대한 문제들에 대해 논의하고 실질적인 기술적 해결책을 찾기 위해 노력하며, 장기간에 걸쳐 디지털자료를 보존하고 사용자의 지속적인 접근을 제

공하기 위한 개념적인 참조모델 개발의 필요성에 따라 OAIS 참조모델을 개발하였고, 2003년 3월에 ISO 14721로 확정되었다.

- OAIS 참조모델은 디지털정보를 장기적으로 보존하는데 필요한 개념적 프레임워크로 디지털정보를 보존하는데 기반이 되는 표준적인 정보 모델링으로 널리 사용되고 있다. OAIS를 준수하는 것은 신뢰할 수 있는 디지털 아카이브 구축의 기본조건이다.
- 전자정보 영구보존을 위하여 개발된 보존시스템에 대한 광범위한 합의사항을 중심으로 개발된 권고사항으로 다음 사항을 포함한다.
 - 정보에 대한 정의
 - 정보패키지 정의
 - 기능적 모델 제시
 - 보존전략 제시 등
- ISO 14721에서 명시하고 있는 구체적인 목적은 다음과 같다.
 - 첫째, 아카이브가 디지털정보를 영구/장기간 보존하는데 있어서 광범위한 의견 일치에 도달하기 위해 개발된 기술적 권고안이다.
 - 둘째, 디지털정보를 영구/장기보존하기 위한 사항을 개발하는데 바탕이 되는 기술적 권고안을 마련한다.
 - 셋째, 아카이브가 정보를 보존하고 적합하게 이용할 수 있도록 하는 사항을 기술하기 위해서이다.
 - 넷째, 장기보존에 필요한 접근 및 관리에 관한 개념을 이해시키는 기본적인 틀 제공한다.
 - 다섯째, 아카이브 구조와 운영에 대해 기술하고 장기보존 전략 기법 비교를 위한 분석의 틀을 제시한다.
 - 마지막으로 OAIS에 관련된 표준을 생산하도록 유도, 디지털정보의 장기보존과 접근에 폭넓고 구체적인 합의점을 만드는데 도움이 되고자 한다.
- OAIS는 장기적인 정보보존과 접근기능을 수행하는 시스템이나 기관에서 활용할 수 있다. ISO 14721의 적용 가능한 범위는 정부기관, 아카이브즈, 기업체나 대학, 디지털정보를 보존, 이용할 수 있게 하는 모든 기관, 그리고 아카이빙 기능을 수행

하고 있다고 인지하지 못하는 기관까지 자료의 보존이 필요한 전체 기관을 대상으로 하고 있다.

- ISO 14721 OAIS 참조모델 표준은 OAIS 개념, OAIS 책임, 세부모델, 보존 전망, 아카이브 간 상호운용성을 포함하여 다음과 같이 6개의 장과 6개의 부록으로 구성되어있다.

1. Introduction
 - 1.1 목적과 범위
 - 1.2 적용가능성
 - 1.3 근거
 - 1.4 순응
 - 1.5 관련 표준 개발을 위한 로드맵
 - 1.6 설명서 구조
 - 1.7 정의
2. OAIS 개념
 - 2.1 OAIS 환경
 - 2.2 OAIS 정보
 - 2.3 OAIS 상위수준의 외부 상호작용들
3. OAIS 책임
 - 3.1 의무적인 책임
 - 3.2 책임수행을 위한 절차의 예
4. 세부모델
 - 4.1 기능모델
 - 4.2 정보모델
 - 4.3 정보패키지 변환
5. 보존 전망
 - 5.1 정보 보존
 - 5.2 접근 서비스 보존
6. 아카이브 간 상호운용성
 - 6.1 OAIS 아카이브간 상호작용의 기술수준
 - 6.2 연합아카이브의 관리문제

- 부록 A 복합 기능 뷰
부록 B 다른 표준 또는 노력과의 관계
부록 C 통합된 모델링 언어 (UML)에 대한 간단한 지침
부록 D 참고 자료
부록 E 표현 정보에 사용된 소프트웨어 모델
부록 F 보안 고려 사항

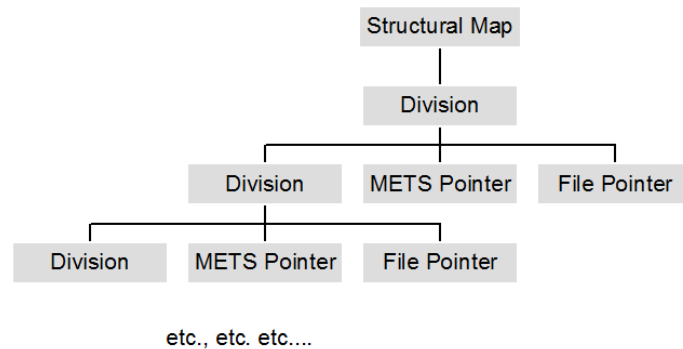
- 지금까지 OAIS 참조모델의 가장 중요한 성과는 디지털 보존의 공통어로 거의 보편적으로 인정되고 있다는 것이며, 참조모델에 명시된 개념과 용어는 디지털 보존 실무자에게 유용한 도구가 되고 있다. OAIS 참조모델은 디지털 보존에서의 협업에 대한 이해, 토론 및 지원을 위한 공유된 어휘를 제공할 뿐만 아니라 저장소가 관리하는 정보 객체뿐만 아니라 보존 저장소의 기본 구성요소를 설명하는 개념적 청사진과 유사한 역할을 한다.
- 이러한 OAIS 참조모델은 디지털 보존 분야로 제한되고 있는데 이는 모델 자체의 문제가 아니라 참조모델이 적용되는 방식 및 관련 업무의 성격 때문이라고 할 수 있다. 또한 정보패키지와 관련하여 실제 적용에 있어서 이 개념을 직접 구현하는 경우는 거의 없으며 PAIMAS와 PAIS가 표준화 프로세스에 이 개념을 정의한 가장 가까운 사례에 해당한다. 표준 모델은 표준화를 촉진하는 것이 그 핵심 개발 이유이고 표준화라는 지속적인 방향성을 제시하는 것이다.

4) METS

- METS 메타데이터 입력 및 전송 표준으로 리포지터리 간의 디지털 저작물 교환을 표준화하기 위해 사용하는, 디지털 저작물에 대한 구조, 기술, 그리고 관리 메타데이터를 입력하기 위한 XML 스키마 기반의 표준이다.
- 디지털도서관의 객체가 어떻게 조직되어 있는지 이해하기 위해서 구조 메타데이터가 필요하고, 디지털 형태의 자원이 원 자원을 얼마나 정확하게 표현하는지를 이해하기 위한 기술 메타데이터가 필요하다는 요구에 대한 해결책으로 제시되었다.
- 오디오 또는 비디오 형태와 같이 시간에 기반을 둔 디지털자원의 구조에 대한 기술(description)과 외부 자원에 대한 링크 기능, 기술 및 관리 메타데이터에 더 많은 확장성을 제공한다.
- METS 스키마의 주된 역할은 다음과 같다.
 - 디지털 객체의 내용을 구성하는 파일들을 구별하고, 그 내용의 구조를 표현
 - 디지털 콘텐츠와 서술적 메타데이터의 링크
 - 디지털 콘텐츠와 관리적 메타데이터의 링크

-
- 디지털 콘텐츠와 동작 정의나 프로그램 코드와의 링크 및 관련된 서술적, 관리적 메타데이터와의 링크
 - 디지털 콘텐츠와 관련 서술적, 관리 메타데이터를 이진데이터로 패키징
- METS는 다음 3가지 문법을 제공한다.
 - 전송 구문(Transfer syntax): 디지털 객체의 전송/교환을 위한 표준으로 OAIS 참조모델에서 SIP에 해당한다.
 - 기능 구문(Functional syntax): 디지털 콘텐츠 및 관련 메타데이터에 대한 보기 및 내비게이션을 최종 사용자에게 제공하기 위한 기반으로 OAIS 참조모델에서 DIP에 해당한다.
 - 저장 구문(Archiving syntax): 디지털 객체들을 아카이빙하기 위한 표준으로 OAIS 참조모델에서 AIP에 해당한다.
 - METS 문서는 7개의 주요 섹션으로 구성된다.
 - ① **METS 헤더(Header) [필수]**: METS 헤더는 제작자, 편집자 등의 정보를 포함하여 METS 문서 자체를 설명하는 메타데이터를 포함한다.
 - ② **기술 메타데이터(Descriptive Metadata Section)**: 설명 메타데이터 섹션은 METS 문서 외부의 설명 메타데이터(예: OPAC의 MARC 레코드 또는 WWW 서버에서 유지 관리되는 EAD 검색도구)를 가리킬 수 있으며 내부적으로 포함된 설명 메타데이터 또는 둘 다를 포함할 수 있다. 외부 및 내부 설명 메타데이터의 여러 인스턴스가 설명 메타데이터 섹션에 포함될 수 있다.
 - ③ **관리 메타데이터(Administrative Metadata Section)**: 관리 메타데이터 섹션은 파일 작성 및 저장 방법, 지적 재산권, 디지털 라이브러리 객체가 파생된 원래 소스 객체에 관한 메타데이터 및 디지털 라이브러리 객체를 구성하는 파일의 출처에 관한 정보를 제공한다(즉, 마스터/파생 파일 관계 및 마이그레이션 / 변환 정보). 설명 메타데이터와 마찬가지로 관리 메타데이터는 METS 문서의 외부에 있거나 내부적으로 인코딩될 수 있다.
 - ④ **파일 섹션(File Section)**: 파일 섹션에는 디지털 버전의 전자 버전을 구성하는 내용이 포함된 모든 파일이 나열된다. <file> 요소는 <fileGrp> 요소 내에 그룹화되어 개체 버전별로 파일을 세분화할 수 있다.

- ⑤ **구조 맵(Structure Map Sections) [필수]**: 구조 맵은 METS 문서의 핵심이다. 디지털 라이브러리 객체에 대한 계층적 구조를 설명하고 해당 구조의 요소를 각 요소와 관련된 내용 파일 및 메타데이터에 연결한다.
- ⑥ **구조 링크(Structure Link Sections)**: METS의 구조 링크 섹션에서는 구조 작성자가 구조 맵에 설명된 계층 구조의 노드 사이에 하이퍼링크가 있는지 기록할 수 있다. METS를 사용하여 웹사이트를 보관할 때 특히 유용하다.
- ⑦ **행위(Behavior Section)**: 행위 섹션을 사용하여 실행 가능 행위를 METS 객체의 내용과 연결할 수 있다. 행위 섹션 내의 각 행위에는 특정 행위 섹션이 나타내는 행위 집합의 추상 정의를 나타내는 인터페이스 정의 요소가 있다. 각 행동에는 인터페이스 정의에 의해 추상적으로 정의된 행동을 구현하고 실행하는 실행 가능 코드 모듈을 식별하는 메커니즘 요소도 있다.



[그림 4-3] METS 구조 맵

[표 4-4] METS를 활용한 보존 메타데이터 구조

METS 섹션	보존 메타데이터 내용
metsHdr(헤더 섹션)	METS 문서 생성기관
dmdSec(서술 메타데이터 섹션)	논문/보고서에 대한 서술 메타데이터
amdSec(관리 메타데이터 섹션)	관리를 위한 메타데이터
fileSec(파일 섹션)	article, abstract, journal, volume
structMap(구조 맵 섹션)	article, abstract, journal, volume 간의 관계성
structLink(구조 링크 섹션)	같은 journal, volume에 포함된 논문간의 관계
behaviorSec(행위 섹션)	보존/폐기 행위

5) ISO 15489: 2016 문헌정보-기록관리

- ISO 15489: 2016 - Information and documentation - Records management
- ISO 15489는 기록관리 및 기록관리 시스템 개발을 위한 표준으로 공표되었는데, 기록 관리정책과 절차를 표준화하고 모든 기록을 적절히 보호하여 기록 정보를 효율적으로 검색할 수 있도록 함을 목적으로 한다. ISO 15489의 개념과 원칙은 다양한 업무 및 기술 환경에 적용되지만, 이러한 환경은 기록 통제, 프로세스 및 시스템의 구현에 대해 서로 다른 접근 방식을 요구할 수 있다.
- ISO 15489에서는 기록을 생산, 획득, 관리하는데 있어서 특정 환경에 대한 자세한 구현을 제안하기 위한 것이 아니다. 오히려 주요 개념을 정의하고, 모든 환경에서 기록을 관리하기 위한 기록 통제, 프로세스, 시스템을 개발할 수 있는 상위 수준의 원칙을 제시하고 있다.
- ISO 15489의 part 1인 '개념과 원칙'은 기록관리에 대한 공통된 요소를 종합하여 얻은 하나의 보편적인 관점과 기록관리의 이론적, 실질적 측면에서 일관되게 지켜야 하는 기본적인 규칙이나 법칙을 이루겠다는 것을 의미한다. 조직으로 하여금 기록의 개념을 자산으로 인식시키고, 기록관리의 중요성과 이점을 부각시킴으로써 ISO 15489의 활용을 극대화하고자 하였다.
- 제시된 기록관리의 이점 중에는 구 버전에는 없었던 '업무 위험의 관리', '소송에서의 보호 및 지원', '기업의 설명책임을 입증하는 능력의 향상', '지적 재산권 보호' 등이 새로운 이점으로 제시되었다. 특히, '업무 효율성 향상을 위한 비용절감'은 기하급수적으로 증가하는 데이터세트 또는 대용량으로 생산되는 시청각기록 등 새로운 기록 유형의 장기보존을 위해 수반되는 비용 문제가 최근 기록관리에서 배제할 수 없는 중요한 요소가 되었기 때문이다(이정은, 윤은하 2018). ISO 15489:2016 표준의 목차는 다음과 같다.

-
1. 범위 Scope
 2. 인용표준 Normative references
 3. 용어 정의 Terms and definitions
 4. 기록관리의 원칙 Principles for managing records
 5. 기록 및 기록시스템 Records and records systems
 - 5.1 일반사항 General
 - 5.2 기록 Records
 - 5.3 기록시스템 Records systems
 6. 정책과 책임 Policies and responsibilities
 - 6.1 일반사항 General
 - 6.2 정책 Policies
 - 6.3 책임 Responsibilities
 - 6.4 모니터링 및 감사 Monitoring and evaluation
 - 6.5 적격성과 훈련 Competence and training
 7. 평가 Appraisal
 - 7.1 일반사항 General
 - 7.2 평가범위 Scope of appraisal
 - 7.3 업무이해 Understanding the business
 - 7.4 기록요구사항 결정 Determining records requirements
 - 7.5 기록요구사항 실행 Implementing records requirements
 8. 기록통제 Records controls
 - 8.1 일반사항 General
 - 8.2 기록 메타데이터 스키마 Metadata schemas for records
 - 8.3 업무분류체계 Business classification schemes
 - 8.4 접근과 허용규칙 Access and permissions rules
 - 8.5 처분 승인도구 Disposition authorities
 9. 기록생산 획득 관리프로세스 Processes for creating, capturing and managing records
 - 9.1 일반사항 General
 - 9.2 기록생산 Creating records
 - 9.3 기록획득 Capturing records
 - 9.4 분류와 색인작업 Records classification and indexing
 - 9.5 접근통제 Access control
 - 9.6 기록저장 Storing records
 - 9.7 이용과 재이용 Use and reuse
 - 9.8 마이그레이션과 변환 Migrating and converting records
 - 9.9 처분 Disposition

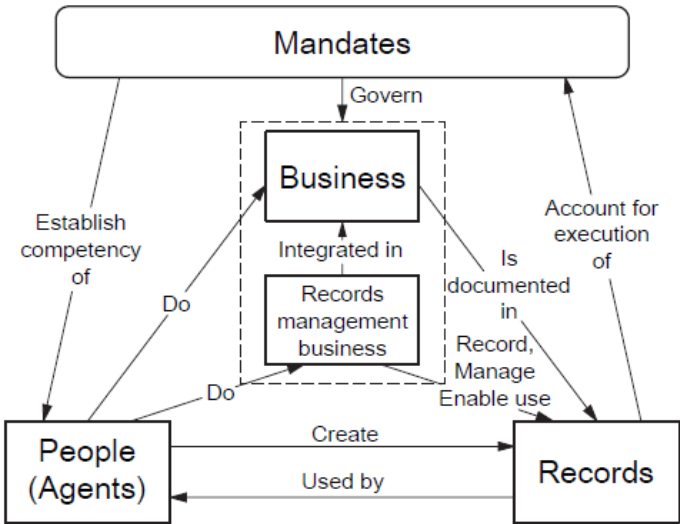
6) ISO 23081: 2017 문헌정보-기록관리과정-기록메타데이터

- ISO 23081: 2017 Information and documentation - Records management processes - Metadata for records - Part 1: Principles
- ISO 23081은 기록관리 메타데이터를 생성, 관리 및 사용하기 위한 프레임워크를 설정하고 이를 관리하는 원칙을 설명한다. ISO 15489가 메타데이터를 설명하고 있지만 메타데이터 관리 지침을 직접 제시하고 있지는 않으므로 이 표준이 ISO 15489 프레임워크 내에서 기록관리를 위한 메타데이터를 이해하고 구현하며, 그리고 사용하기 위한 메타데이터 표준으로 제시되었다.
- 업무 프로세스 내에서 기록관리 메타데이터의 관련성과 업무 및 기록관리 프로세스를 지원하는 메타데이터 역할 및 유형에 관하여 다루고 있으며 이러한 메타데이터를 관리하기 위한 프레임워크를 제공한다. ISO 23081 표준의 목차는 다음과 같다.

1. 범위
2. 참고 문헌
3. 용어 정의
4. 기록관리 메타데이터
5. 기록관리 메타데이터의 관점과 목적
 - 5.1 기록관리 메타데이터의 목적과 이점
 - 5.2 조직에 적용되어야하는 기록관리 메타데이터
6. 역할과 책임
7. 다른 메타데이터 영역과 관련한 기록관리 메타데이터
 - 7.1 일반 사항
 - 7.2 e- 비즈니스를 위한 메타데이터
 - 7.3 보존을 위한 메타데이터
 - 7.4 자원 기술을 위한 메타데이터
 - 7.5 자원 발견을 위한 메타데이터
 - 7.6 권한 관리를 위한 메타데이터
 - 7.7 시스템 경계를 넘나드는 지속 가능성을 위한 메타데이터
8. 메타데이터 관리
 - 8.1 일반
 - 8.2 메타데이터 적용 수준
 - 8.3 메타데이터를 작성하고 적용 할 때 기록이 존재하는 지점
 - 8.4 메타데이터 관리 프로세스

8.5 메타데이터 구조
8.6 시스템의 역할
9. ISO 15489-1을 지원하는데 필요한 메타데이터 유형
9.1 메타데이터 유형 소개
9.2 기록에 관한 메타데이터
9.3 업무 규칙, 정책 및 위임 사항에 관한 메타데이터
9.4 에이전트 메타데이터
9.5 업무프로세스 메타데이터
9.6 기록관리 프로세스에 대한 메타데이터

- ISO 15489-1을 지원하는데 필요한 메타데이터 유형은 다음 구성 요소로 나눌 수 있다.
 - ① 기록 자체에 관한 메타데이터
 - ② 업무 규칙 또는 정책 및 위임에 관한 메타데이터
 - ③ 에이전트에 대한 메타데이터
 - ④ 업무 활동 또는 프로세스에 대한 메타데이터
 - ⑤ 기록관리 프로세스에 대한 메타데이터



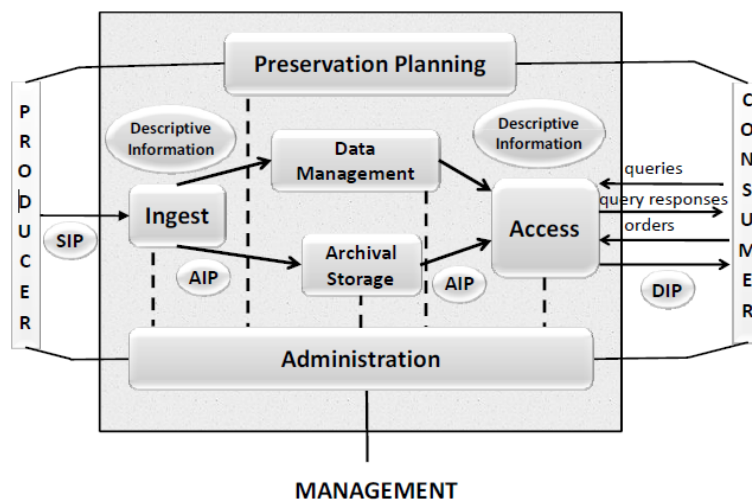
[그림 4-4] 주요 메타데이터 유형과 관계

- ISO 23081에서 제시하고 있는 기록관리의 개념적 데이터 모델은 기록(record), 생산하고 활용하는 주체(agent), 처리행위로서의 업무(business), 관할하는 규정(mandate)이라는 개체(entity)와 그 개체간의 관계로 구성된다. 메타데이터를 활용하면 업무 시스템(business system)이 존재하는 업무환경으로부터 개체를 재현해낼 수 있다. 개체는 업무 환경을 이해할 수 있도록 기록에 대한 개념을 제공하지만 개체 자체가 항상 만질 수 있는 객체(object)는 아니다.

□ 디지털자원 보존 시스템의 기능 요건

1) OAIS의 기능 요소

- OAIS의 장기보존은 크게 「비트열 보존(bit preservation)」과 「논리적인 보존(logical preservation)」이라는 2개의 활동으로 구별할 수 있다. 전자는 0과 1이 되는 문자열을 결손없이 완전히 보존하기 위한 활동이며, 후자는 전자정보를 사람들이 이해할 수 있는 형태로 재현하기 위한 활동이다.
- OAIS에서는 특히 논리적인 보존을 고려하여 장기보존을 담당하는 시스템에서 요구되는 기능요소를 6개로 설정하고 거기에서 취급하는 정보의 구성을 다음과 같이 정의하고 있다.



[그림 4-5] OAIS 기능 요소

• OAIS의 기능 요소는 다음과 같다.

- ① 입수(Ingest): 생산자로 부터 SIP를 받기 위한 서비스와 기능을 제공하고, 아카이브 내에서 저장하고 관리하기 위한 내용을 준비
- ② 아카이브 저장(Archival Storage): AIP를 저장하고 유지하며 검색하기 위한 서비스와 기능을 제공
- ③ 데이터 관리(Data Management): 아카이브 소장물을 식별하고 문서화하기 위한 기술 정보과 아카이브 내부의 관리 데이터를 생성하고, 유지하고, 접근하기 위한 서비스와 기능을 제공
- ④ 관리(Administration): 아카이브 시스템의 전반적인 운영을 통제
- ⑤ 보존 계획(Preservation Planning): OAIS의 환경을 모니터하고 초기의 컴퓨팅 환경이 노후화되어 사라져도 지정된 이용자 집단이 OAIS에 저장된 정보를 장기간에 걸쳐 접근할 수 있도록 보증
- ⑥ 접근(Access): 이용자(Consumer)가 OAIS에 저장된 정보에 대해 존재여부, 기술내용, 위치정보, 가용 여부 등을 확인하고, 정보를 요구하여 생산물을 받을 수 있도록 함

[표 4-5] OAIS의 기능 및 하위기능 요소

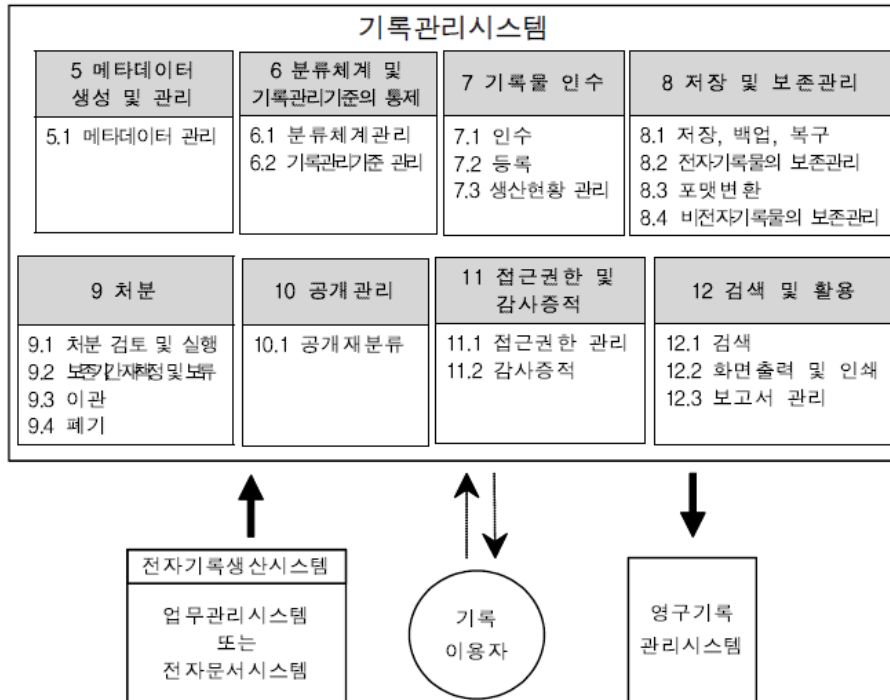
기능	하위 기능
입수	제출정보접수 품질확인 AIP 생성 기술정보생성 갱신조정
저장	데이터접수 저장체계관리 매체수록 에러점검 재난복구 데이터제공
데이터관리	데이터베이스관리 질의어처리 보고서생성 데이터베이스 갱신
운영	제출동의협상 시스템구성관리 보존정보갱신 물리적접근통제 표준 및 정채 굿립 감사 수요활성화 이용자서비스
보존계획	유관공동체 모니터링 기술모니터링 보존전략 및 표준 개발 패키징설계와 마이그레이션계획
접근	접근활동조정 DIP생성 응답전달

-
- 단계별 프로세스는 다음과 같다.

- ① 생산자가 SIP를 입수(Ingest)에 제공한다.
- ② 입수는 AIP를 생성하여 아카이브 저장(Archival Storage)으로 전달한다.
- ③ 입수는 AIP에서 관련된 기술정보(Descriptive Information)를 추출하여 데이터 관리(Data Management)로 넘긴다.
- ④ 이용자는 적절한 기술정보와 접근도구(Access Aids)를 활용하여 정보를 탐색하고 요청한다.
- ⑤ 접근(Access)은 아카이브 저장에서 이용자가 탐색하여 요청한 적절한 AIP를 적절한 DIP로 변형한다.
- ⑥ 이러한 모든 활동은 관리(Administration) 지침에 따라 수행된다.
- ⑦ 보존 계획(Preservation Planning)이 보존전략과 기법을 안내하여 제자리에 위치시킨다.

2) 기록물관리시스템 기능 요건(NAK 6 2015)

- 기록관리시스템은 일반적으로 업무활동의 증거를 획득하고, 적절한 통제 하에서 이를 활용할 수 있도록 다음과 같은 기능이 구현되어야 한다.
- 기록물의 내용과 맥락을 모두 포함하여 관리할 수 있어야 하며 이를 위하여 메타데이터를 획득하고 지속적으로 추가·관리할 수 있어야 한다. 특히 기록관리시스템은 진본성, 신뢰성, 무결성 및 이용가능성을 확보할 수 있는 방식으로 기록물을 저장하고 관리할 수 있어야 하며, 신뢰할 수 있고 감사추적이 가능한 상태에서 기록물을 처분할 수 있어야 한다.
- 이용자에게 편리한 방식으로 최대한 기록물을 검색하고 열람을 지원할 수 있어야 하며, 인수, 이관 등을 효율적으로 수행하기 위해 유관 시스템과 원활하게 상호연계할 수 있어야 한다.
- 마지막으로 기록관리시스템은 진본성과 무결성을 확보하기 위해 전자서명, 암호화 등과 같은 기술적인 보호조치를 취할 수 있어야 한다.



[그림 4-6] 기록관리시스템의 기능 모델

- 각각의 기능에 대한 핵심 요건은 다음과 같다.
- **메타데이터 생성 및 관리:** 기록물의 진본성과 무결성을 입증하고, 이용가능성을 보장할 수 있도록 기록 메타데이터 요소의 필수여부, 선택값, 작성방법 등에 따라 적절한 메타데이터가 생성, 관리될 수 있도록 필요한 기능을 제공하여야 한다.
- **분류체계 및 기록관리기준의 통제:** 기록물이 생산된 업무 맥락을 제공할 수 있도록 생산당시의 조직, 기능, 업무 처리행위에 대한 분류체계를 유지하고 해당 기록물에 대한 기준정보를 관리할 수 있어야 한다. 또한, 조직의 업무목표에 따라 필요한 분류체계를 반입하여 적용할 수 있어야 하며, 분류체계에 의한 다양한 검색을 지원할 수 있어야 한다.
- **기록물 인수:** 다양한 유형과 포맷으로 생산, 등록된 기록물을 그 수와 양에 관계없이 인수할 수 있어야 한다. 또한, 기록물의 진본성과 무결성이 유지될 수 있도록 기록물의 내용, 구조, 맥락을 함께 인수하고, 인수 과정을 감사하고 추적할 수 있어야

한다. 적법한 절차에 따라 기록물을 등록하고 편철할 수 있는 기능을 제공하여야 하며, 기록물의 생산현황을 들여오기 및 내보내기할 수 있어야 한다.

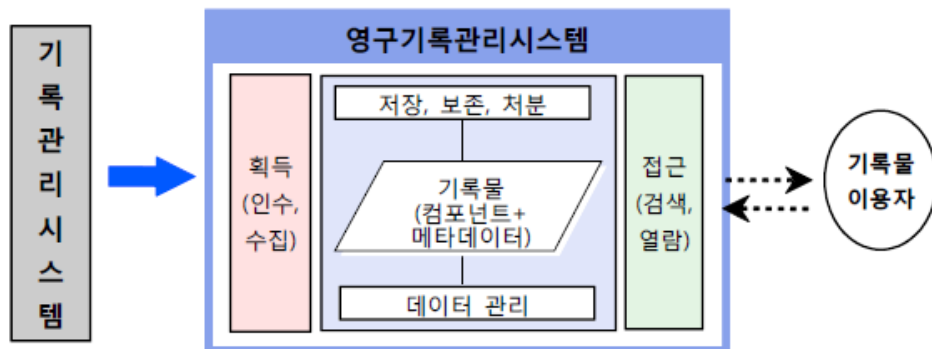
- **저장 및 보존관리:** 인수한 기록물의 진본성, 무결성 및 지속적인 접근 가능성을 보장하기 위해 디지털 저장소를 적절히 관리하고 시스템 오류나 예기치 못한 사고에 대비한 백업 및 복구 기능을 구비하여야 한다. 또한 전자기록물의 장기보존을 위해 보존포맷 변환, 마이그레이션 등의 기능을 수행하여야 하며, 비전자기록물의 보존을 위한 서가관리, 기록물 위치 추적 기능 등을 제공하여야 한다.
- **처분:** 적법한 절차와 기준에 따라 이관 또는 폐기하거나 보존기간을 재책정 하는 등 기록물의 처분을 실행할 수 있는 기능을 제공하여야 한다. 또한 적법한 절차에 따른 처분의 검토와 실행이 증명될 수 있도록 처분에 의해 변경된 기록물의 정보와 관련된 행위에 대한 이력을 메타데이터로 관리할 수 있어야 한다.
- **공개관리:** 비공개 기록물에 대해 5년마다 공개재분류를 수행하고, 영구기록물관리 기관으로 이관할 시기가 도래한 기록물에 대해 공개재분류를 수행할 수 있는 기능을 제공하여야 한다. 또한 수행된 공개재분류 행위에 대한 이력을 메타데이터로 관리할 수 있어야 한다.
- **접근권한 및 감사증적:** 사용자, 기록물, 시스템 기능에 대한 접근권한과 접근범위를 설정하고 접근권한 관리에 따라 기록물과 시스템에 대한 접근을 통제할 수 있는 기능을 제공하여야 한다. 또한 중요 기록관리 업무 수행 이력을 감사증적으로 남길 수 있어야 하며, 감사 증적의 내용을 임의로 변경하지 못하도록 하여야 한다.
- **검색 및 활용:** 기록물과 그 메타데이터에 대한 정보가 효율적으로 활용될 수 있도록 다양한 검색서비스를 제공하여야 한다. 또한 검색된 기록물과 메타데이터를 이용자 요구에 따라 화면에 출력하거나 인쇄할 수 있어야 하며, 사용자 권한별로 요청된 사항을 보고서로 생성할 수 있어야 한다.

3) 영구기록물관리시스템 기능 요건(NAK 7 2015)

- 영구기록관리시스템은 영구적 보존가치가 있는 기록물을 장기간 보존하고 관리하는 시스템으로서 다음과 같은 사항을 고려하여 구축되어야 한다.

-
- 해당 영구기록물관리기관의 기록물관리 업무를 효율적으로 수행하기 위한 절차와 기능이 구현되어야 한다.
 - 기록물관리 원칙과 장기적인 기록물 보존 및 서비스에 필요한 사항을 고려하여 구축되어야 한다.
 - 해당 기관이 수집 및 관리하는 기록물의 특성, 업무환경, 기록물관리 정책, 사회적, 기술적 변화에 대한 대응 및 위험관리 등을 반영할 수 있어야 한다.
 - 기록물의 형태나 유형에 관계없이 모든 기록물을 관리할 수 있어야 한다.
- 영구기록관리시스템이 갖추어야 할 기본적인 기능은 다음과 같다([그림 4-7] 참조).
 - 시스템은 인수, 수집, 관리, 보존, 활용 및 기타 지원기능에 이르기까지 전체 기록물 관리 프로세스 수행에 필요한 기능을 제공해야 하며, 기록물의 전자·비전자여부, 기록물의 유형과 관계없이 모든 종류의 기록물을 관리할 수 있어야 한다.
 - 시스템은 기록물의 진본성과 무결성 증명에 필요한 절차와 기술적 조치를 포함하여야 한다. 기록물을 획득하는 시점부터 기록물관리의 전 단계에 걸쳐 기록물의 진본성과 무결성을 보장하며, 이에 대한 검증 결과를 신뢰성 있게 제시할 수 있어야 한다.
 - 시스템은 장기보존에 필요한 기록관리 메타데이터를 구성하고 관리하여야 한다. 기록물의 구조와 맥락 뿐만 아니라 장기간에 걸친 기록물 관리 이력을 메타데이터로 기록함으로써 기록물의 진본성, 무결성 유지 및 지속적인 기록물 이용에 필요한 도구로 활용되도록 하여야 한다.
 - 시스템은 기록물관리 원칙을 준수하고 기록물을 적극적으로 서비스할 수 있도록 정리, 기술, 공개관리, 열람, 검색 등 필요한 업무 기능을 지원해야 한다. 또한 효율적인 기록물 검색과 활용성 증대를 위해 다양한 검색도구를 구축하고 이를 제공할 수 있도록 지원해야 한다.
 - 시스템은 기술의 변화에 관계없이 장기간에 걸쳐 기록물이 안전하게 보호되고 서비스될 수 있도록 기관이 정한 기록물 장기보존 전략을 이행하고 이를 위해 필요한 업무 기능을 지원할 수 있어야 한다.
 - 시스템은 승인받지 않은 접근에 의한 기록물의 이용, 변경, 훼손 및 삭제가 발생하지 않도록 접근권한을 포함한 권한관리를 수행하고, 시스템과 기록물에 대한 임의의 접근시도 이력을 감사증적으로 남겨 관리할 수 있어야 한다.

- 시스템은 기록물 관리자나 이용자의 요구를 수용하고, 해당되는 법령과 규정을 준수하며, 사회적, 기술적 변화를 신축성 있게 수용할 수 있도록 장기적으로 점검 및 유지관리 되어야 한다. 시스템의 개선이나 유지관리와 관련된 모든 행위는 기록으로 관리되어야 한다.



[그림 4-7] 영구기록관리시스템의 기본 기능

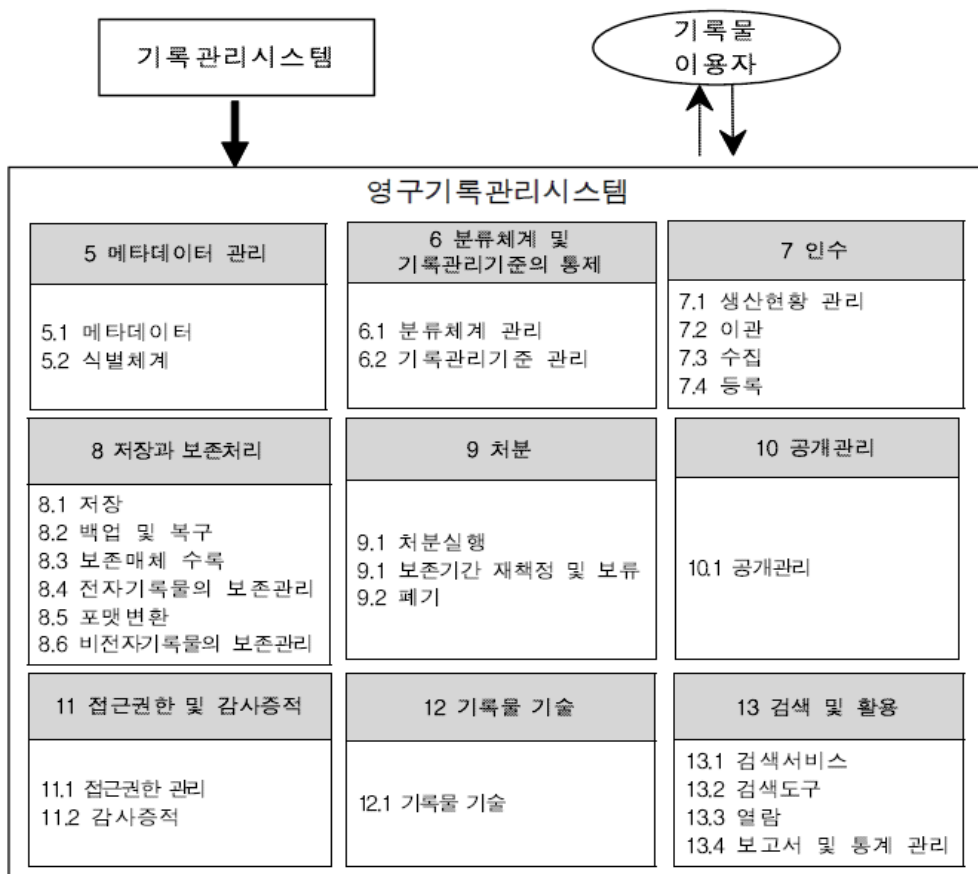
- 영구기록관리시스템 각각의 기능에 대한 핵심 요건은 다음과 같다([그림 4-8]과 [그림 4-9] 참조).
- **메타데이터 관리:** 메타데이터는 기록물을 구성하는 필수요소로서 기록물의 내용과 구조를 기록하고 시간의 경과에 따른 관리사항을 기술함으로써 기록의 진본성, 무결성을 보장하고, 기록물 관리의 신뢰성을 확보할 수 있도록 관리되어야 한다. 또한 기록물 서비스에 재 활용되어 기록물에 대한 이용가능성을 높일 수 있도록 구조가 설계되고 효율적으로 운용되어야 한다. 메타데이터는 기관의 승인된 정책에 따라 추가, 변경이 가능해야 하며 허가받지 않은 접근으로부터 보호되어야 한다.
- **분류체계 및 기록관리기준의 통제:** 기록물의 분류체계를 유지하고 해당 기록물에 대한 기준 정보를 관리할 수 있어야 한다. 또한 영구기록물관리기관의 특성에 따라 기록물에 다중 분류체계를 적용할 수 있어야 하며, 분류체계를 통해 다양한 기록물 검색이 가능하여야 한다.
- **인수:** 법적인 절차에 따라 시스템으로부터 보존대상 기록물을 파악하기 위한 생산현황 관리, 실제 기록물을 들여오는 인수, 기증, 위탁, 회수 등의 방법으로 개별 기

기록물을 획득하는 수집, 그리고, 시스템으로부터 고유식별자를 부여하여 공식적인 기록물로 획득하는 절차인 등록으로 구성된다. 다양한 유형의 기록물을 일괄적 혹은 개별적으로 인수받을 수 있어야 하며, 기록물의 내용, 구조, 맥락을 함께 인수받아야 한다.

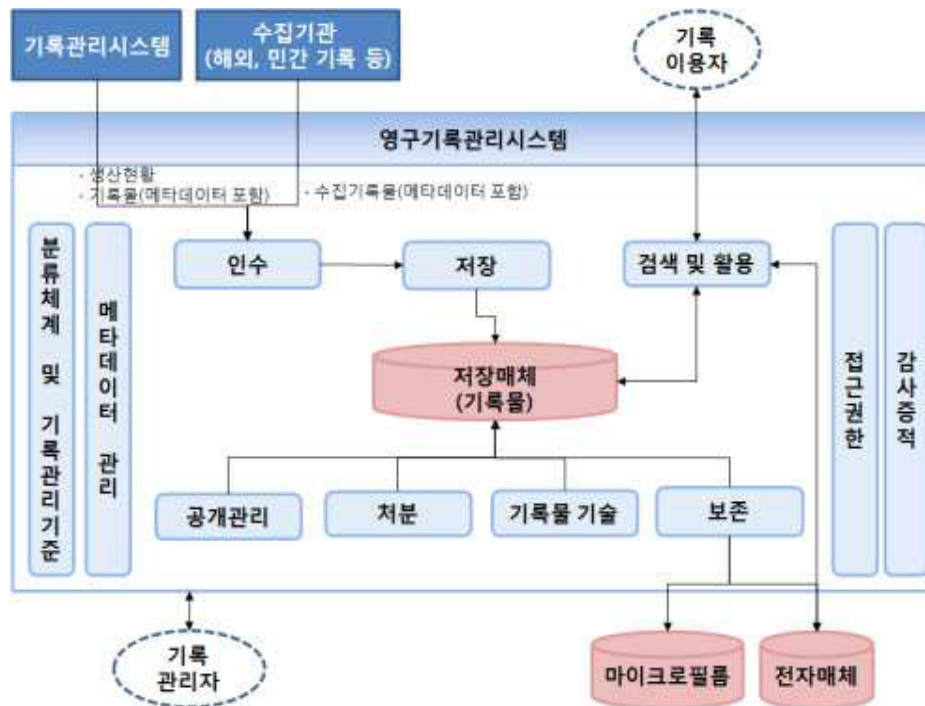
- **저장과 보존처리:** 저장 및 보존처리는 시스템에 등록 완료된 기록물을 저장장치에 저장하고, 저장된 기록물을 장기간 보존, 관리하기 위해 필요한 백업 및 복구, 매체수록, 포맷변환으로 구성된다. 특히, 전자, 비전자기록물은 각각의 특성에 따라 구분되어 보존관리될 필요가 있다. 시간의 흐름에 따른 기술의 변화, 예상치 못한 재난 등으로부터 기록물을 안전하게 보호하기 위한 정책적, 기술적 조치와 업무절차를 기능으로 구현할 수 있어야 한다.
- **처분:** 처분은 법령에서 정한 적법한 절차와 기준에 의해 기록물의 보존가치를 평가하고 기록물을 처분하는 행위로서, 승인된 결정에 따라 기록물의 처분을 실행할 수 있어야 한다. 기록물의 처분은 보존기간 재책정 및 보류, 폐기로 구성된다. 처분의 결정과 실행에 대한 이력정보는 메타데이터로 관리되고 처분의 적법성을 증명할 수 있도록 감사증적으로 남길 수 있어야 한다.
- **공개관리:** 공개관리는 기록물건에 대한 공개여부를 관리하고, 공개재분류 등의 업무를 수행하기 위한 기능으로 공개관리를 위한 이력정보가 메타데이터로 누적되고 기록물 서비스를 위한 정보로서 활용될 수 있도록 관리되어야 한다. 접근권한 및 감사증적: 기록물의 추적, 접근 제어, 기록물 변경에 대한 통제를 통해 기록물의 진본성과 무결성을 보장하고, 기록물 관리 행위에 대한 신뢰성을 제시하기 위한 기능으로 접근권한 관리 및 감사증적 기능이 요구된다. 시스템은 기록에 대한 허가받지 않은 접근, 변경, 은폐 및 삭제가 발생하지 않도록 사용자의 접근권한을 관리하고 지속적으로 모니터링 하여야 한다. 기록물에 적용되는 통제조치와 접근이력은 메타데이터로 관리되고 감사증적으로 남겨야 한다.
- **기록물 기술:** 기록물 기술은 영구기록물관리기관이 수집한 기록물을 물리적으로 또는 논리적으로 분류하고 해당 분류계층의 기록물을 표준화된 형식으로 기술하기 위한 기능이다. 기록물 기술을 수행하기 위해서 시스템은 기관에서 정한 고유한 분류체계를 시스템에 반영하고 분류와 기술을 효율적으로 수행할 수 있도록 지원하

여야 한다. 기술 분류체계와 요소별 기술내용은 단계별 기록관리 업무 프로세스와 타 검색도구의 구축, 기록물 서비스에 연계되어 활용될 수 있어야 하며 표준화된 기술 요소는 메타데이터로 관리되어야 한다.

- **검색 및 활용:** 검색 및 활용은 기관이 수집한 기록물과 그 메타데이터를 내외부적으로 서비스하기 위한 기능으로 검색서비스, 검색도구, 열람, 보고서 및 통계관리로 구성된다. 기록물의 효율적인 검색과 활용을 위해 메타데이터를 재활용하고 이용자의 요구사항을 반영하여 검색서비스와 검색도구를 구현할 수 있어야 하며, 기관이 정한 진본성 보장을 위한 통제조치를 통해 기록물을 원본 열람을 지원할 수 있어야 한다. 또한 요청에 따른 보고서 생성과 통계관리가 가능하도록 지원하여야 한다.



[그림 4-8] 영구기록관리시스템의 기능 요소



[그림 4-9] 영구기록관리시스템의 기능별 프로세스 흐름도

4) UNHCR 디지털 보존 프로젝트의 디지털 보존 시스템 기능 요건

- 유엔 난민 고등 판무관실(United Nations High Commissioner for Refugees, UNHCR)은 1950년 유엔 총회에서 난민과 실향민에 대한 보호와 지원을 제공하기 위해 설립되었다. UNHCR의 기록보존소는 1950년 조직 설립 당시의 정보를 포함하여 전 세계적으로 그리고 역사적으로 범위와 내용면에서 유일한 컬렉션을 보유하고 있으며, 더 많은 자료를 현장에서 가져온 디지털자료를 보존하기 위한 최첨단 시스템을 구현하기 위해 노력하고 있다. 다음은 UNHCR의 디지털 보존 프로젝트에서 정의하고 있는 디지털 보존 시스템 기능 요건이다(UNHCR 2017).

3 기능 요건 FUNCTIONAL REQUIREMENTS

- 3.1 사전 입수 PRE-INGEST
- 3.2 입수 기능 INGEST FUNCTIONAL FACILITY
- 3.3 자동 입수 AUTOMATED INGEST
- 3.4 그래픽 이용자 인터페이스 GRAPHICAL USER INTERFACE
- 3.5 제출 정보패키지 유효성 확인 VALIDATION OF THE SUBMISSION INFORMATION PACKAGE
- 3.6 기술 메타데이터 TECHNICAL METADATA
- 3.7 AIP와 보존 메타데이터 저장을 위한 관계 메타데이터 RELATIONSHIP METADATA FOR AIP AND PRESERVATION METADATA STORE
- 3.8 AIP 생성 GENERATION OF ARCHIVAL INFORMATION PACKAGE
- 3.9 아카이브 저장 기능 ARCHIVAL STORAGE FUNCTION
- 3.10 파일 시스템 정보 보존 PRESERVATION OF FILE SYSTEM INFORMATION
- 3.11 저장 ARCHIVAL STORAGE
- 3.12 API의 무결성 자동체크 AUTOMATED INTEGRITY CHECKING OF AIPs
- 3.13 저장은 완전해야 한다 ARCHIVE STORE SHOULD BE ATOMICALLY COMPLETE
- 3.14 보존 저장 수준 LEVELS OF PRESERVATION STORAGE
- 3.15 데이터 관리 DATA MANAGEMENT
- 3.16 보존 메타데이터 저장 PRESERVATION METADATA STORE
- 3.17 관계성(버전, 원본과 복제) RELATIONSHIPS
- 3.18 관리 기능 개체 ADMINISTRATIVE FUNCTIONAL ENTITY
- 3.19 유지 정책 RETENTION POLICY
- 3.20 보존 계획 기능 개체 PRESERVATION PLANNING FUNCTIONAL ENTITY
- 3.21 접근 기능 개체 ACCESS FUNCTIONAL ENTITY
- 3.22 지원과 교육(SLA 모델) SUPPORT AND TRAINING

5) NLM 디지털 리포지터리의 기능 요건

- 미국 국립의학도서관(National Library of Medicine, NLM)은 포맷에 관계없이 생물 의학 및 보건 전문 분야의 출판된 기록을 수집, 정리, 보존 및 접근 할 수 있도록 도서관이 다양한 디지털 형식으로 많은 양의 자료를 관리하고 보존하는 데 필요한 견고한 인프라를 개발하였다. NLM 디지털 저장소는 모든 유형의 디지털 객체에 대한 저장, 보존 및 액세스 기능을 제공하며 다음과 같은 기능 요건을 제시하고 있다(NLM 2007).

7. 기능 요건 Functional Requirements

7.1 입수 Ingest

7.1.1 접수 Receive Submission

7.1.2 품질 보증 Quality Assurance

7.1.3 아카이브 정보패키지 생성 Generate Archival Information Package (AIP)

7.1.4 기술 정보/메타데이터 생성 Generate Descriptive Information/Metadata

7.1.5 위치정보 업데이트 Coordinate Updates

7.2 저장 Archival Storage

7.2.1 접수일 Receive Data

7.2.2 저장 계층 관리 Manage Storage Hierarchy

7.2.3 매체 교체 Replace Media

7.2.4 오류 체크 및 재난 복구 Error Checking and Disaster Recovery

7.2.5 데이터 제공 Provide Data

7.3 데이터 관리 Data Management

7.3.1 관리지 DB Administer Database

7.3.2 질의 수행 Perform Queries

7.3.3 보고서 생성 Generate Report

7.3.4 DB 업데이트 Receive Database Updates

7.4 관리 Administration

7.4.1 제출 합의 협상 Negotiate Submission Agreement

7.4.2 시스템 구성 관리 Manage System Configuration

7.4.3 아카이브 정보 업데이트 Archival Information Update

7.4.4 표준과 정책/절차 수립 Establish Standards and Policies/Procedures

7.4.5 감사 제출 Audit Submission

7.4.6 요청 활성화 Activate Requests

7.5 저장 계획 Preservation Planning

7.5.1 (지정된) 지역사회 모니터링 Monitor Designated Community

7.5.2 기술 모니터링 Monitor Technology

7.5.3 보존전략과 표준 개발 Develop Preservation Strategies and Standards

7.5.4 패키징 설계와 마이그레이션 계획 개발 Develop Packaging Designs and Migration Plans

7.6 접근 Access

7.6.1 위치 접근 활동 Coordinate Access Activities

7.6.2 DIP 생성 Generate DIP

7.6.3 배달 Deliver Response

4.2.2 디지털자원 보존 리스크 분석

- 디지털자원 보존을 위한 리스크는 DPC의 디지털 보존 핸드북에 따르면, 디지털자료 관련, 조직 관련, 그리고 자원 관련 리스크로 구분할 수 있다(DPC 2015).

□ 디지털자료 관련 리스크

1) 데이터 유지

- 모든 디지털 파일은 일련의 0과 1 또는 비트로 구성되는데 비트 스트림은 디지털자료의 생존을 보장하기 위해 손실이나 손상없이 시간이 지남에 따라 획득되고 유지되어야 한다. 이러한 비트를 보존하려는 시도에는 여러 가지 위협이 있다. 저장매체는 시간이 지남에 따라 부식되어 파일이 손상될 수 있다. 저장매체는 쓸모없어지게 되고 최신 컴퓨터와 해당 컴퓨터를 이해하고 액세스할 수 있는 소프트웨어에 의해 지원되지 않을 수 있다. 비트는 무시 혹은 포기되거나 실수로 삭제 혹은 악의적으로 파괴될 수 있다.
- 비트 보존을 위한 체계적인 프로세스를 유지하는 것은 장기적인 디지털 보존을 보장하는데 있어 기본적인 요구 사항이다. 저장소 미디어를 모니터링하고 새로 갱신해야 한다. 다양한 종속 기술을 도입하고 단일 지리적 위치에서의 치명적인 재난 방지를 위해 파일 복제 또는 백업을 이중화해야 한다. 손실을 식별하고 비트의 무결성을 효율적이고 자동화된 방식으로 확인할 수 있도록 체크섬을 생성하고 자주 점검해야 한다. 디지털자료가 저장되는 장소는 신중하게 기록되어야 하며 보존 책임자를 지정해야 한다.

2) 데이터의 의미 유지

- 비트스트림 내에서 인코딩된 정보를 재구성하려면 일반적으로 데이터의 특정 인코딩이나 형식을 렌더링, 조작, 분석 또는 다른 방식으로 상호 작용하도록 설계된 컴퓨터 소프트웨어가 필요하다. 시간이 지남에 따라 인코딩 (또는 파일 형식)이 변경될 수 있으며 해당 인코딩과 상호 작용하는 소프트웨어 응용 프로그램이 작동하지

않을 수도 있다. 잘 알려진 파일 형식에서는 드물지만 자주 사용되지 않는 파일 형식은 렌더링 소프트웨어가 더 이상 지원되지 않아 시간이 지나면 쓸모없게 될 수 있다.

- 특정 디지털자료가 종속된 기술을 이해하면 자료의 보존을 위해 적절한 조치를 취할 수 있다. 고려된 보존 계획 프로세스에 따라 디지털 파일 포맷에서 포맷으로 마이그레이션하거나, 오래된 소프트웨어의 에뮬레이션 또는 대체 소프트웨어 응용 프로그램을 사용하여 데이터를 렌더링할 수 있다. 각 옵션마다 장점과 단점이 있으며 사례별로 세심하게 평가해야 한다.
- 파일 형식 노후화가 이전에는 압도적인 위험으로 떠오르지 않았지만 도전적인 요소가 되었다. 오래된 파일 형식을 렌더링하는 방법(예: 오래된 소프트웨어를 에뮬레이트하는 방법)을 찾는 것이 가능할 수도 있지만 렌더링이 얼마나 정확하며 소프트웨어를 실행하는 것이 합법적인가, 이 복잡한 노력의 보존비용과 유지비용이 타당한가에 대한 문제가 있다.

3) 데이터에 대한 신뢰 유지

- 디지털자료는 시간이 지남에 따라 유동적이거나 쉽게 편집되거나 변경 혹은 미디어 오류로 인해 손상될 수 있다. 또한 소프트웨어 렌더링을 신뢰할 수 없거나 사용자에 의해 부정확하게 읽혀질 수 있다. 디지털 보존 작업의 결과에 대해 최종 사용자가 신뢰를 가지기 위해서는 디지털자료의 전체 수명주기와 시간이 지남에 따라 누구와, 무엇과 상호 작용했는지에 대해 신중히 고려해야 한다.
- 디지털자료의 진본성과 무결성이 중요하다. 예를 들어, 학자들은 인용문이 시간이 지남에 따라 동일하게 유지될 것이라는 확신을 가질 필요가 있다. 법원은 법적인 증거 요구 사항을 잘 유지하고 있다고 확신해야 하며 정부 부처는 진위성과 관련하여 법적으로 시행 가능한 요건을 갖추어야 한다. 이 문제는 법률 및 조직 문제와 중복되며 일반적인 절차 대신 개별 분야에서 해결하는 것이 가장 좋다.
- 데이터 무결성 기술을 적용하고 감사 추적을 유지하면 아카이브에 보관된 이후 디지털 객체가 변경되지 않았다고 확신할 수 있다. 궁극적으로 사용자에게 대한 진정성

은 보존 조직 전체의 폭넓은 신뢰성에 훨씬 더 의존할 수 있다. 적절한 감사 및 인증으로 검증된 고품질 보존 프로세스를 유지하는 것이 중요하다.

4) 데이터의 맥락과 종속관계 유지

- 디지털정보의 의미는 원래 작성되었거나 사용된 맥락 안에서 내재된 추가 정보에 의존하지만 나중에 다시 방문할 때는 명확하지 않을 수 있다. 관련 문맥 정보를 확인, 이해 및 획득하는 것이 성공적인 보존 노력에 관건이 될 수 있다.
- 디지털정보가 계속해서 더 복잡하고 상호 연결된 방식으로 생성되기 때문에 관련 정보 자원의 광범위한 맥락에서 특정 디지털자료의 위치를 유지해야 할 수도 있다. 겉으로는 단순하고 독립적인 문서가 실제로 관련 파일과 참조된 글꼴에 의존할 수 있고 웹에서 관련 정보에 대한 포인터를 가지고 있을 수도 있다. 간단한 웹 페이지로 볼 수 있는 것도 인터넷의 다른 위치에서 출처가 지정된 실제 데이터로부터 즉석에서 생성되었을 수 있다.
- 데이터의 사용 방법을 이해하면 종속관계와 맥락은 적절한 방식으로 보존을 위해 획득되고 향후 지적 내용을 보존하고 이해할 수 있도록 충분히 명시적으로 문서화될 수 있다.

5) 적시에 보존 활동을 수행

- 적시에 디지털 보존 활동의 우선순위를 정하고 적용하는 것은 손실을 회피하는 것뿐만 아니라 한정된 자원을 최대한 활용하는 데 중요할 수 있다. 수명주기 초반에 개입할 수 있다면 디지털자료는 앞으로 더 잘 살아남을 수 있을 것이다. 파일 형식의 선택, 중요한 문서의 획득 또는 메타데이터의 주요 관계에 대한 설명은 적은 투자를 필요로 하지만 투자 대비 비용절감 효과가 훨씬 높다.
- 이것이 불가능하고 데이터에 대한 위협이 확인되면 보존 조치를 위한 최상의 시기가 명확하지 않을 수 있다. 기술 노후화를 피할 수 있는 조기 개입은 장기적인 지속가능성에 대한 더 큰 확신을 제공하지만 개입이 궁극적으로 필요하지 않고 자원이 낭비될 위험이 있다. 적절한 시기에 조치를 취하면 불필요한 활동을 최소화할 수

있지만 더 이상 사용되지 않는 특수한 지식을 필요로 경우에는 구식 기술을 조사하는데 필요한 노력이 증가한다. 적절한 조치가 경우에 따라 취해져야한다.

6) 데이터 홍수에 대처

- David Rosenthal(2014)의 연구에 따르면 데이터 생성 속도는 연 60%씩 증가하고 있고 데이터 스토리지는 연간 약 25%씩 확대되고 있으며 그리고 데이터 센터 예산은 연 2%씩 증가하고 있다.
- 이는 선택 정책 및 기타 조직의 의사 결정에 대한 도전적인 압력을 가하는 동시에 기술적인 문제를 야기한다. 한 수준에서 효과적으로 기능하는 단순한 보존 프로세스가 대용량의 데이터 또는 매우 큰 개별 파일에 쉽게 적용, 확장할 수 있는 것은 아니다. 대규모 오디오 비주얼 컬렉션, 연구 데이터 및 웹 기반 아카이브 처리에 대한 전문성이 커짐에 따라 대규모로 작업할 수 있는 기술 및 이해가 빠르게 진행되고 있다.
- 그러나 일부 저장소는 늘어나는 데이터 처리를 위한 확장 가능한 아키텍처와 절차를 개발하고 유지, 관리하는데 있어 여전히 중대한 문제에 직면하고 있다. 디지털자료에 대한 접근, 관리 및 접근 기술 및 관리상의 어려움을 과소평가해서는 안 된다. 선택, 평가 및 처분이 디지털 관리 활동의 중요한 구성요소임을 잊지 않는 것이 중요하다.

□ 조직 관련 리스크

- 기술 문제도 어려울 수 있지만 조직 문제와 관련된 수많은 문제가 있다. 여기에는 디지털 보존이 어떻게 조직되고 전달되는지, 그러한 책임이 디지털자료의 시간과 라이프 사이클에 따라 어떻게 변하는지 등이 포함된다. 조직 전반에 공통적인 디지털 보존 문제가 있지만 조직이 처한 모든 상황이 다를 수 있다. 조직의 동인을 확인하고 이러한 요구를 충족시키는 실제 솔루션을 개발하는 것이 중요하다. 디지털 보존을 위한 모든 접근법을 충족하는 솔루션은 없다.

-
- 디지털자료의 생성, 보존 및 액세스는 널리 배포된다. 결과적으로 기술의 이점을 극대화하고 공통된 문제를 해결하며 비용 효과적인 문제를 극복하기 위해 개별 조직 또는 국가의 경계를 뛰어넘을 필요성이 커지고 있다.

1) 사내 또는 아웃소싱

- 외주(제3자) 또는 사내 또는 아마도 이 두 가지를 결합하여 디지털 보존 전체 또는 일부를 수행할지 결정하는 것은 복잡한 문제이다. 인력 충원 및 인프라가 충분하면 디지털 보존이 사내에서 수행될 수 있으나 일부 활동이나 지원을 아웃소싱하는 것이 비용 효과적이며 내부 역량과 능력을 활용할 수도 있다.
- 저장소에서 특정 작업이나 서비스를 아웃소싱하는 것이 결코 새로운 현상은 아니다. 효과적인 사양을 준비하고 성능을 모니터링 할 수 있는 충분한 지식을 보유하고 유지하는 것이 가장 중요하다. 아웃소싱 작업을 쉽게 확인하고 품질을 확인해야하며, 사양을 세심하게 설계하고 외주사가 제공하는 보고를 통해 가장 잘 수행할 수 있어야한다. 디지털 보존 계약 여부를 결정할 때 비용은 분명히 중요한 고려 사항이지만 법적 문제와 같이 고려해야 할 다른 요소도 있다. 예를 들어 개인 정보 보호 또는 기밀 유지로 인한 법적 조항은 아웃소싱의 적합성 여부에 영향을 줄 수 있다.

2) 공동 작업(협력)

- 디지털 보존 문제는 모든 조직과 모든 부문에서 직면하게 될 문제이므로 전문 지식과 경험을 쌓는 것이 중요하다. 디지털 보존 과제에 효과적으로 대처하고 이를 극복하기 위해 조직 내에서 그리고 조직 간에 더 큰 협력을 하도록 하는 강력한 이유와 경우에 따라 정치적 압력이 있다.
- 대부분의 조직은 협업 증가의 이점을 쉽게 인식하지만 상이한 의제, 시간 척도 또는 자금 조달 메커니즘에서 발생할 수 있는 잠재적 어려움도 나타내고 있다. 그렇다하더라도, 잠재적인 어려움을 완화하는 특정 영역이나 강도에 따라 협업이 가능하다. 디지털 보존 분야에서 가장 중요하고 성공적인 최근 이니셔티브의 일부는 사실상 협력 작업으로 이루어졌다.

3) 조직 변화

- 현대의 디지털 세계는 기술적으로나 조직상의 변화가 급격한 분야이다. 디지털 보존은 장기적인 활동이며 시간이 지남에 따라 조직 변화에 의해 영향을 받을 가능성이 있다.
- 이것은 상위 조직의 변화뿐만 아니라 주요 제공자 및 사용자, 공급 업체 또는 공동 협력자의 변화를 통해 저장소에 영향을 미칠 수 있다. 따라서 조직 변화는 관리해야 할 주요 위험이다.

4) 조직 구조

- 기술의 성격과 디지털자료의 보존에 대한 의존성은 조직 구조에 영향을 미칠 수 있다. 많은 활동이 수렴되는데 예를 들어, 수집과 보존에 대한 결정이 동시에 현명하게 이루어져야 한다. 조직 구조는 디지털자료에 필요한 모든 기술과 전문 지식을 끌어내기 위해 경계를 넘어야 한다.
- 조직에서 획득 및/또는 생성 한 디지털자료의 보존 책임을 할당하는 것은 필연적으로 함께 작업하는 조직의 여러 부서의 직원들과 관련되어 있다. 직원에게 전달할 수 있는 강력한 조직의 비전을 뒷받침하지 않는 한 잠재적으로 어려움을 겪을 수 있다.

5) 역할과 책임

- 특정 주제 영역 또는 특정 형식에 대한 책임을 수행하는 기존의 저장소가 몇 가지 있다. 예를 들어, 영국에서는 영국 데이터 서비스가 선정된 사회 과학 연구 데이터에 책임을 지고 있으며 영국 국가도서관의 국립 사운드 아카이브는 음원 수집에 대한 책임을 맡고 있다. 각 저장소는 자체 수집 정책과 수집 기관의 소관 범위를 넓게 고려해야 한다.
- 디지털 환경은 많은 이해 관계자 그룹과의 관계를 요구한다. 디지털 보존에 대한 생명주기 접근 방식은 장기보존을 담당하는 조직이 생산자, 게시자 및 기타 중개자, 그리고 상호간에 상호 작용하고 협력해야 하는 방식에 중대한 함의를 지니고 있다.

-
- 디지털자료 생산자는 자신이 만든 디지털자료의 중장기 생존 가능성 측면에서 자신의 행동이 갖는 의미를 이해할 수 있어야한다. 부서의 일상 업무에서 작성된 기록이든 아날로그 수집 자료의 디지털 사본이든, 지침 및 지원은 물론 적절한 기술 및 조직 인프라가 효율적인 관리 및 보존을 위한 크게 개선된 전망을 촉진하는데 도움이 될 것이다.

6) 선택

- 디지털 방식으로 생산되는 엄청난 양의 정보, 가변적인 품질 및 장기 접근을 유지하는 책임을 지는 사람들에게 대한 자원 제약으로 인해 진행 중인 접근을 보존하는 것이 목적이려면 선택이 불가피하다.
- 디지털 환경에서 보존을 위한 비선택(non-selection)은 가치있는 것으로 여겨지더라도 거의 확실하게 항목의 손실을 의미할 수 있다.
- 여러 버전이 있는 경우 보존을 위해 가장 적합한 버전을 선택할지 둘 이상의 버전을 선택해야 할지 결정해야한다. 각 변경 사항을 저장하려고 시도하는 것과 달리 동적 리소스를 샘플링하는 것이 실용적인 유일한 옵션일 수 있다. 그러나 잘 정의된 프레임워크 내에서 샘플링이 수행되지 않고 사용자의 현재 및 미래의 필요성을 적절하게 고려하면 심각한 반향을 일으킬 수 있다.
- 선택시 디지털 보존을 보장하는 데 필요한 중복성 수준까지 고려해야한다. 누가 그 책임을 맡을 것인지, 어느 기간 동안인지 명확하게 이해해야한다. 그렇지 않으면 여러 사본이 다양한 저장소에 저장되더라도 모든 저장소가 여러 가지 이유로 어느 시점에서 디지털 개체의 유지 관리를 중단할 수 있다.

7) 보안 및 접근 균형 조정

- 보존과 접근 사이에는 항상 강한 연관성이 있다. 리포지터리는 대부분의 경우 디지털자료에 다양한 사용자가 접근하면서도 안전함을 보장해야할 필요가 있다. 실제 사용자에게 의한 접근은 보존 시설 설계에 귀중한 도움을 줄 수 있어 불필요한 조치를 피할 수 있을 뿐 아니라 피드백주기를 확인하고 도입하는데도 도움을 줄 수 있다.

-
- 장기보존을 위해 선택된 많은 종류의 디지털자료에는 비인가 사용자가 액세스하지 못하도록 보호해야하는 기밀 정보 및 기밀 정보가 포함될 수 있다. 다른 경우에는 저장소에 대한 접근에 영향을 주는 법적 또는 규제 의무가 있을 수 있다. 이 두 가지 역할 사이에는 긴장이 있을 수 있으며 보안과 접근의 용이성 사이에 균형을 유지할 필요가 있다.

8) 법률 준수

- 디지털 보존의 법적 문제는 단순하지 않다. 디지털자료에는 여러 사본 및 파생 버전이 존재하며 서로 다른 출처의 소프트웨어 및 메타데이터와 관련될 수 있다. 디지털 콘텐츠는 보다 광범위한 생산자 그룹에 의해 생성되며 아날로그 세계에서 적용되는 것보다 다양한 형식 및 지적 재산권(IPR)과 관련되어 있다. 또한 법률은 종종 기술 변화와 디지털 보존 요구보다 뒤떨어져 있다. 디지털자료의 수집, 보존 및 접근 제공에 있어 저장소에 영향을 미치는 주요 법적 문제는 다음과 같다.
 - 계약과 합의 또는 정부의 입법을 통해 기증자 및 기금 제공자가 저장소 및 그 모기관에 취하는 관리, 보존 및 접근에 관한 법적 요구 사항 기록, 법정 보증금 등 (예: 접근성, 가용성, 정보 보안, 유지, 감사 및 준수, 공개, 납본 등)
 - 저장소가 보유한 디지털자료의 제3자 권리 관련 법적 의무(예: 저작권, 데이터 보호)
 - 저장소와 제 3자 제공자 혹은 제공자들 간의 관계에 대한 법적 요소(예: 서비스 계약 및 서비스 수준 합의 조건)

□ 자원 관련 리스크

1) 예산 및 비용

- 디지털 보존비용은 다른 조직의 지출과 쉽게 분리 될 수 없으며 그렇게 해서는 안 된다. 디지털 보존은 본질적으로 시간의 흐름에 따라 액세스를 보존하는 것이므로 디지털 라이프 사이클의 모든 부분에 대한 비용을 고려하는 것이 적절하다. 이러한 맥락에서 디지털자료를 만드는 비용조차도 궁극적으로 장기보존을 용이하게 하는

비용 요소로 포함할 필요가 있다.

- 적절한 기술을 갖춘 직원을 고용하고 개발할 수 있는 능력은 기술 변화의 속도와 필요한 기술의 범위에 따라 더 어려워질 수 있다. 또한 리소스가 부족한 전통적인 컬렉션 및 디지털 컬렉션을 관리해야 할 필요가 있는 조직의 리소스 제약에 따라 제한된다.
- 디지털 보존에 필요한 인건비가 가장 중요한데 전담 전문가뿐만 아니라 행정, 관리, IT 지원, 법률 고문 등 다양한 직원이 포함된다.
- 비용에 영향을 미치는 다른 주요 쟁점은 조직의 사명과 목표, 수집의 유형과 크기, 보존 수준, 필요한 접근의 양과 수준, 조치를 위한 시간 프레임 등을 포함한다.

2) 인력 충원 및 기술

- 디지털 보존은 다양한 기술과 조직의 역할을 포함한다. 일반적으로 디지털 보존은 일반적이지 않은 다양한 기술을 사용한다. 즉 규모가 큰 조직에서는 여러 분야의 팀을 구성해야하며 소규모 조직에서는 분산된 팀 또는 지원 부서를 활용해야한다.
- 인력 및 기술과 관련하여 고려해야 할 세 가지 주요 쟁점은 다음과 같다.
 - 첫째, 최근 몇 년간 상당한 발전이 있었지만 디지털 보존 교육은 종종 현재의 모범 사례보다 뒤떨어져 있거나 신입 사원을 위한 관련 정보 관리 프로그램은 대체로 이론 수준에 머무른다. 따라서 실용적인 기술과 경험을 가진 인력은 공급보다 수요가 많아 직원 채용에 어려움을 겪을 수 있다.
 - 둘째, 직무 기술서를 작성하기 어려울 수 있다. 특히 기관이 처음부터 새로운 역할부터 효과적으로 시작하고자 할 때, 이를 위해 많은 연구 프로젝트가 디지털 보존에 필요한 일반적인 기술을 설명하려고 시도해 왔는데 조직의 여러 수준에서 서로 다른 기술이 필요하다는 가정을 기본으로 한다.
 - 마지막으로, 디지털 보존 업무를 담당하는 직원들은 적극적으로 경력 개발에 참여해야한다. 기술과 사용자의 요구가 시간이 지남에 따라 발전할 것이라는 기대를 감안할 때 이러한 변화하는 요구 사항을 충족시키는 데 참여하는 직원은 전문가 브리핑 및 전문 네트워킹 등 기술을 지속적으로 향상시킬 수 있는 방법을 찾

아야한다.

3) 시설

- 효과적인 디지털 보존을 위해서는 운영 워크플로우 및 디지털자료 처리가 기반이 될 수 있는 몇 가지 기본 시설 또는 인프라가 필요하다. 조직이 디지털 보존을 시작할 때는 초보적이거나 최소 규모일 수 있지만 대용량 데이터를 처리하는데 필요한 시설에는 상당한 투자가 필요하다.

4) 저장

- 손실을 방지하기 위해 보존 데이터를 복제해야한다는 요구 사항으로 인해 스토리지 하드웨어는 가장 중요한 디지털 보존 시설 중 하나이다. 스토리지 기술은 최근 수십 년 동안 급속히 변했다. 장기간 보관을 위해 CD 또는 DVD와 같이 널리 사용되는 매체가 이용되나 자기 미디어의 급속한 발전으로 인해 빠르고 안정적인 스토리지가 등장하였다. 엔터프라이즈 스토리지 시스템은 합리적인 비용으로 대량의 스토리지 볼륨을 제공한다. 일반적으로 약 4-8년의 유한한 수명을 지니고 있지만 쉽게 모니터링하고 교체할 수 있다.
- 조직은 보존 인프라로 클라우드 서비스 “임대”를 원할 수도 있다. 클라우드의 유연성으로 비교적 신속하고 저렴한 테스트 및 파일럿이 가능하다. 클라우드 서비스는 여러 위치에 간편하고 자동화된 복제를 제공하고 전문적으로 관리되는 디지털 저장소 및 무결성 검사를 수행할 수 있다. 저장소는 전문 도구, 절차, 워크플로우 및 서비스 계약에 대한 액세스를 추가하고 전문 공급 업체를 통해 디지털 보존 요구 사항에 맞게 조정된 디지털 저장소 시스템을 제공할 수 있다.

5) 디지털 저장소 시스템

- 디지털자료를 보존하기 위한 핵심 요구 사항 중 많은 부분이 전용 디지털 보존 시스템 또는 신뢰할 수 있는 디지털 저장소에 의해 자동화된 방식으로 제공된다. 저장소 응용 프로그램은 내부에 있는 각 디지털 객체를 고유하게 식별한다. 객체의 저장을

관리하고 객체의 특성을 식별하며 저장소 관리자의 보존 계획을 도울 수 있다.

- 또한 객체에 대한 접근을 용이하게 한다. 시간이 지남에 따라 효과적으로 디지털 자료를 관리하는데 있어 전용 저장소 애플리케이션이 필수적이다. OAIS 모델은 저장소에서 요구하는 기능에 대해 높은 수준의 모델을 제공한다.

6) 고성능 컴퓨팅

- 대규모 데이터는 더 많은 스토리지뿐만 아니라 더 큰 계산 능력을 필요로 한다. 데이터의 기술적 특성의 설정 및 평가, 검색 및 액세스를 위한 데이터 인덱싱, 무결성 검사 및 기타 여러 작업을 수행하려면 상당한 계산 성능이 필요하다. 연구 데이터 또는 웹 아카이브는 일반적으로 고성능 컴퓨팅과 이러한 요구를 충족시키기 위해 상용 하드웨어 클러스터에서 실행되는 아파치 하둡(Apache Hadoop)과 같은 기술을 필요로 한다.

4.2.3 디지털자원 보존 운영 지침

- 디지털자원 보존을 위한 조직차원의 활동은 디지털자원의 생산부터 수집, 저장, 보존, 그리고 접근에 이르는 각 단계별 활동으로 구성된다. 각 단계별 활동 시에 참고할 수 있는 가이드라인은 다음과 같다(DPC 2015).

□ 생산

- 생산자는 디지털자원에 장기간 접근할 수 있도록 지원하는데 있어서 중요한 역할을 한다. 대부분의 개인 생산자는 디지털 콘텐츠를 보존하는 장기적인 노력을 기울일 수는 없다. 모든 디지털자원에는 수명주기가 있으며 관련된 여러 이해 관계자와 이해관계가 있다.
- 여기에서는 전자기록, 연구 데이터 또는 개인 디지털정보와 같은 디지털 대용물 또는 본 디지털 파일의 생성 과정에서 디지털 보존을 위해 고려해야 할 주요 문제와

활동에 대한 가이드를 제공한다. 특히 가이드라인은 다음과 내용을 담고 있다.

- 디지털 보존은 필요할 때까지 디지털자료에 대한 지속적인 액세스를 보장하는 데 필요한 일련의 관리 활동을 의미한다. 여기에는 생산자가 디지털자료를 생산할 때 디지털 보존을 위한 요구 사항을 충족시키기 위해 필요한 활동을 포함한다.
- 실제로 많은 조직과 개인이 수십 년 동안 사용하고 관리해야하는 디지털자료를 생산하는데 디지털 보존을 처음부터 고려하지는 않는다. 박물관, 도서관 또는 보존 임무를 가진 기록보존소가 아닌 경우 이러한 행동이 디지털 보존에 어떻게 영향을 미치는지 의사소통하는데 “디지털 연속성”과 같은 용어를 사용한다.
- 생산자가 연속성을 보장하기 위해 적극적으로 노력하지 않으면 디지털자료를 쉽게 사용할 수 없게 될 수 있음을 깨닫는 것이 중요하다. 정보가 완전하고 가용성이 뛰어나고 업무 요구에 맞게 사용할 수 있는지 확인하는 것이다.
- 다음 [표 4-6]은 장기간 보존 잠재력과 장기보존 가능성을 보장하기 위해 디지털자료를 생성할 때 고려해야 할 주요 문제 및 조치에 대한 지침이다.

[표 4-6] 본 디지털자료 보존을 위한 고려 사항

본 디지털자료 보존	
1	<ul style="list-style-type: none"> • 소프트웨어 및 형식(Software and formats) <ul style="list-style-type: none"> - 잘 지원되는 소프트웨어를 선택하고 다양한 프로그램에서 읽을 수 있는 파일을 만든다.
2	<ul style="list-style-type: none"> • 파일 명칭(File names) <ul style="list-style-type: none"> - 컨텍스트를 제공하고 현재와 미래의 사람과 컴퓨터가 쉽게 이해할 수 있는 내용과 날짜로 구성된 짧은 파일 명을 사용한다. - 공백이나 특수 문자 (- 또는 _ 이외의 문자)를 사용하면 컴퓨터 하드웨어 나 소프트웨어의 오역을 방지할 수 있다. - 날짜 정보는 ISO 8601: 2004 표준 형식 (YYYY-MM-DD)을 준수한다. 이는 버전 추적을 위한 일관된 방법을 제공한다. 시스템에서 생성된 개별 파일 날짜 메타데이터는 이후 작업으로 자동 변경될 수 있다. - 일관된 방법으로 파일 버전을 표시한다. 날짜를 기본으로 하며 필요에 따라 버전 번호 v1, v2, v_final 등으로 보완할 수 있다.
3	<ul style="list-style-type: none"> • 저장 및 백업(Storage and Backup)
4	<ul style="list-style-type: none"> • 전환 계획(Plan for transactions) <ul style="list-style-type: none"> - 일부 전환은 예측할 수 있고 일부는 예측할 수 없지만 훌륭한 계획과 절차를 통해 완화될 수 있다.

□ 수집 및 평가

- 디지털 환경에서 생산과 선택에 관한 결정은 보존에 중요한 영향을 미친다. 접근과 보존 사이의 연결은 처음부터 유지(즉 보존)하지 않으면 디지털자원에 대한 액세스가 비교적 짧은 기간 내에 손실 될 수 있으므로 종이 및 기타 전통적인 자료보다 훨씬 명확하다.
- 기술의 다양한 변화를 통해 디지털자원을 보존하는 것이 실현 가능하지 않고 바람직하지도 않은 경우 획득을 재평가해야 할 수도 있다. 전통적인 보존 환경과 동일한 원칙이 유용하게 적용될 수 있지만 디지털 환경에 맞게 정책과 절차를 조정해야 한다.
- 정확한 문서화는 디지털 환경에서도 매우 중요하다. 이는 시간이 지남에 따라 자원을 관리하기 위한 필수 세부 사항뿐만 아니라 디지털 객체 자체를 유지하는 기술적, 물리적 상황에 대한 정보를 제공한다.
- 네트워크에 연결된 디지털자원의 경우 자원에 대한 액세스를 제공한다고 해서 반드시 자원을 물리적으로 소장할 필요가 없는 경우 수집 개념은 기존 소장과 완전히 다르다. 액세스를 제공하거나 '가상 컬렉션'을 구축하는데 사용할 수 있는 다양한 옵션이 있다. 예를 들어 액세스용 사본/미러 만들기, 리소스에 대한 하이퍼링크 제공, 온라인 카탈로그 및 지원도구 찾기.
- 특히 수집 및 평가 과정을 위해 필요한 표준 절차문서와 함께 협상 프로세스를 위한 표준 제출 계약 및 라이선스 세트가 수립되어야 한다. 이는 보존에 필요한 최소한의 허가 및 지적 재산권이 부여되도록 하는데 유용하기 때문이다. 충분한 라이선스 계약이 없으면 조직은 디지털 컬렉션을 소유하고 있어도 적극적으로 보존하거나 제공할 권리가 없다.
- 다음은 수집 워크플로우에서 다루어야 할 내용이다.
- 협상
 - 저작권 관련 사항
 - 보존 목적으로 레코드를 변경하는 조직의 권한(예: 장기간 보존하거나 액세스할

수 있는 형식으로 마이그레이션).

- 전송 내용이 레코드의 이해에 필수적이지만 레코드 자체를 구성하지 않는 내용을 포함하는 경우, 조직의 콘텐츠가 다른 곳에서 획득될 경우 해당 파일을 삭제할 수 있다는 데 동의(예: 레코드의 메타데이터).

• 이전(transfer)

- 본 디지털자료를 조직의 IT 환경으로 전송할 때 전송의 보안 및 완전성을 보장하기 위해 이것이 어떻게 수행되어야하는지에 대한 고려가 있어야한다. 소규모 조직의 경우 관련 디지털 파일을 드라이브 또는 유사한 하드웨어에 제공하고 내용을 대조하는 것으로 충분할 수 있다.
- [표 4-7]은 디지털자료의 전송 및 접근 옵션을 설명한다.

[표 4-7] 파일 포맷 및 저장매체의 전송 및 접근 옵션

파일 포맷 및 저장매체의 전송 및 접근 옵션		
옵션	발행물	요구 사항
• 모든 옵션		<ul style="list-style-type: none"> • 파일 형식 및 표준 에 관한 정책 • 스토리지 및 레거시 미디어 형식의 개발에 대한 보존 계획 및 기술 감시
<ul style="list-style-type: none"> • 수신 된 파일 형식의 범위 제한 • 수신 된 미디어의 범위 제한(가장 경제적인 장기 옵션) 	<ul style="list-style-type: none"> • 관리를 간소화하고 전반적인 비용을 절감한다. • 예금주는 자원이나 전문 지식이 부족할 수 있다. • 사용되는 다양한 파일 형식 및 공개 표준에 대한 독점 확장. • 전송에 사용되는 물리적 저장매체는 임시 운송 업체 일 수 있으며 콘텐츠는 사용 된 장기 저장 장치로 전송된다. 	<ul style="list-style-type: none"> • 선호하는 파일 형식에 대한 지침. • 보증금에 대한 영향력 정도. • 옹호 및 협력 전략은 바람직한 결과를 달성한다. • 선호 전송 매체 및 전송 절차에 관한 지침.
<ul style="list-style-type: none"> • 받은 파일 형식을 수락하지만 표준 파일 형식으로 변환 • 받은 저장매체를 수락하지만 사용 된 표준 저장 장치로 내용을 전송 	<ul style="list-style-type: none"> • 관리를 간소화하고 장기간 비용을 절감한다. • 기술적으로 표준 파일 형식으로 변환할 수 없다. • 우발적인 데이터 손실이 발생하지 않았는지 확인해야한다. 	<ul style="list-style-type: none"> • 법률 준수, 저작권 허가 또는 법정 보존 권리. • 호스트 기관의 자원 및 전문 기술. • 선호하는 형식의 선거. • 변환이 가능한 네이티브 형식의 문서화. • 변환 프로세스의 무결성 검사.

파일 포맷 및 저장매체의 전송 및 접근 옵션		
옵션	발행물	요구 사항
<ul style="list-style-type: none"> • 수령 한 대로 보관 (초기 비용은 낮지만 장기적으로 비용 효과가 가장 적은 옵션). 	<ul style="list-style-type: none"> • 복잡한 관리 및 시간이 지남에 따라 리소스 관리 비용이 증가한다. • 높은 위험 옵션, 특히 다수의 디지털자료가 수집되는 경우. • 파일 형식을 선택할 수 있다. 보관 된 것은 보존에 가장 적합하지 않을 수 있다. • 저장매체는 품질이 미숙하고 장기보존에 적합할 수 있다. • 형식은 기관에서 쓸모없거나 지원되지 않을 수 있다. 	<ul style="list-style-type: none"> • 단기 및 장기보존을 위한 우선 순위를 명확하게 정의. • 수신 된 항목에 암호화, 독점 소프트웨어 등의 문제를 해결할 수 있는 기능. • 항목에 포함 된 정보에 대한 미래의 액세스를 보장하는 능력.

• 검증(validation)

- 디지털자료는 (매니페스트에 따라) 완전성을 확인하고, 무결성을 유지하며, 바이러스 및 악성 코드가 없어야 한다. 디지털 보존 시스템에 입수 준비가 된 것이다.
- 파일 및 폴더 이름 목록과 각 파일의 체크섬/고정 값으로 구성된 검증 가능한 목록
- 파일의 크기(전체 볼륨 포함)
- 파일 포맷 목록
- 기록과 관련된 모든 IPR을 기술한 진술서

• 입수(Ingest) 절차

- 항목을 검사하고, 기술 메타데이터를 수집하고, 설명 메타데이터를 향상시키고, 모든 객체에 적용할 수 있는 조직의 일반 접근 프로세스를 시작한다.
- 검사 시에 조직의 수집 또는 수집 프로필, 지침 문서에 제시된 파일 형식 사양을 충족시키지 못하거나 조직의 정보 수집을 위해 부적절하다고 판단되는 경우 다시 고려할 수 있다. 장기간의 평가 기간 동안 또는 다른 이유, 즉 개인 정보나 상업적으로 민감한 정보, 예를 들어 명예 훼손법을 위반하는 항목과 같은 특정한 민감한 자료에는 유예기간(moratorium)이 부과될 수 있다. 이러한 경우 파일의 종료 기간을 명확하게 지정하는 것이 중요하다.

- 다음 [표 4-8]은 입수 절차에서 저장 및 보존을 위한 데이터 및 문서에 대한 내용이다.

[표 4-8] 입수 절차 - 저장 및 보존을 위한 데이터 및 문서

입수 절차-저장 및 보존을 위한 데이터 및 문서
<ul style="list-style-type: none"> • 고유 번호 매기기 <ul style="list-style-type: none"> - 기관이 사용하는 각 디지털자원에는 고유한 식별자가 할당되어야한다. 이 번호는 기관의 카탈로그에 있는 리소스를 식별하고 실제 미디어 및 설명서를 찾거나 식별하는 데 사용된다. 자원이 어떤 이유로든 고유 번호는 다시 할당되지 않아야한다. 영구 식별자 체계를 사용하는 경우 영구 식별자에서 조연을 참조한다. • 취급 및 저장 이전 지침 <ul style="list-style-type: none"> - 직원을 고용하기 위한 취급 및 이전 지침은 IT 및 보존 직원이 다양한 저장매체 및 장기 보관 시스템으로의 파일 전송에 대한 모범 사례에 대한 조언을 반영하여 개발해야한다. • 파일 형식 다시 포맷 <ul style="list-style-type: none"> - 자원을 전송하는 데 사용된 파일 형식이 장기보존에 적합하지 않은 경우, 기관은 자원을 선호하는 파일 형식으로 재형식화 할 수 있다. 아카이브 형식 외에도 사용자에게 제공하기에 적합한 다른 형식의 버전도 원본 파일에서 생성될 수 있다. • 복제 <ul style="list-style-type: none"> - 기관의 보관 및 보존 정책의 일환으로 가입하는 동안 항목의 다중 백업 사본이 생성될 수 있으며 재해 복구 절차가 가능하다. • 보안 <ul style="list-style-type: none"> - 시스템 및 물리적 보안 정책 및 절차는 가입하는 동안 항목의 보전 및 무결성을 보장하기 위해 마련되어야한다. 이것들은 보안에 관한 제도적 정책과 절차에서 개발되고 반영되어야한다. • 버전 및 버전 제어 <ul style="list-style-type: none"> - (예: 정기적으로 업데이트되는 데이터베이스의 연례 스냅 샷) 또는 적절한 경우 접근 항목의 버전 제어 (예: 다른 형식으로 액세스하거나 보존 및 액세스를 위해 다른 형식을 사용하는 항목)를 업데이트 및 편집하기 위한 절차가 생성된다. • 카탈로그 및 문서화 표준 <ul style="list-style-type: none"> - 플랫폼, 개인 간에 정보와 문서를 효과적으로 교환하기 위해서는 전송, 검증 및 수집 중에 수신되거나 생성된 그림 및 문서가 필수적이다. 최소한 항목의 출처 및 관리 내역 (작성된 이후의 모든 데이터 처리 포함), 내용, 구조 및 IPR 권한과 기간을 포함한 후속 관리 및 사용에 첨부된 이용 약관에 관한 정보를 제공해야한다. 다음 사항을 충분히 자세히 설명해야한다. • 처리 시간 <ul style="list-style-type: none"> - 이상적으로 목표는 가입 과정 중에 발생하는 미처리 및 위험 가능성이 있는 자료의 백 로그를 방지하기 위해 수집과 목록 작성 간의 최대 시간 설정 및 모니터링이 필수적이다.

- 한편으로, 조직은 디지털 컬렉션을 확보하기 위한 충분한 기술 및 인력 자원이 있는지 여부를 고려해야한다. 그러나 이 정보는 특정 디지털 컬렉션의 다양한 큐레이

팅 문제가 시간이 지남에 따라 밝혀질 수 있기 때문에 인수 초기에는 분명하지 않을 수 있다. 따라서 조직은 지식, 기술 및 인력 부족을 계획하고 가능한 경우 전문가용 디지털 큐레이팅 서비스의 교육, 채용 또는 참여를 통해 이를 해결해야 한다. 자금 조달 자원이 이를 충족시키지 못하는 경우에는 종종 사내 지식 개발을 위한 전용 드라이브만으로도 충분할 수 있다.

- 다음 [표 4-9]는 수집 및 평가를 위한 권장사항을 요약한 추천 체크리스트이다.

[표 4-9] 수집 및 평가 - 추천 체크리스트

수집 및 평가 - 추천 체크리스트	
제출자를 계약과 안내	
<input type="checkbox"/>	표준 제출 계약서 및 라이선스 모음 만들기
<input type="checkbox"/>	제출자를 위한 적절한 지침 작성
이전 절차	
<input type="checkbox"/>	공급 업체의 디지털자료 전송 지침 및 지원을 위한 설명서 제공
<input type="checkbox"/>	저장 및 보존 정책을 지원하기 위해 이전 절차 개발을 결정
검증 절차	
<input type="checkbox"/>	미디어, 콘텐츠 및 구조 확인
저장 및 보존을 위한 데이터 및 문서 준비 절차	
<input type="checkbox"/>	각 항목별 고유 번호 매기기
<input type="checkbox"/>	다른 매체에 대한 취급 및 보관 전송 지침
<input type="checkbox"/>	동일한 지침에 따라 필요한 경우 파일 형식 다시 포맷
<input type="checkbox"/>	기관의 보관 및 보존 정책의 일환으로 여러 사본 생성
<input type="checkbox"/>	접근 동안 아이템에 대한 시스템 및 물리적 보안 정책 및 절차
목록 작성 및 문서화 절차	
<input type="checkbox"/>	지적 재산권 정보를 포함한 목록 작성에 필요한 최소한의 정보 표준
<input type="checkbox"/>	회고 문서화 또는 카탈로그 강화 지침.
<input type="checkbox"/>	항목의 버전 또는 에디션을 업데이트하고 관리하는 절차.
<input type="checkbox"/>	컬렉션 관리 데이터베이스를 업데이트하는 절차
<input type="checkbox"/>	목록 작성 및 문서화 표준의 선택
<input type="checkbox"/>	가입 작업 및 완료를 위한 시간 척도 목표
절차 검토	
<input type="checkbox"/>	가이드라인 및 일정은 매년 또는 조직의 개발 요구 사항 및 컬렉션 개발 정책에 보조를 맞추기 위해 가능한 한 자주 검토해야 한다.
직원 교육	
<input type="checkbox"/>	지식, 기술 및 인력 부족을 계획하고 가능하면 전문 타사 서비스의 교육, 채용 또는 참여를 통해 이를 해결한다.
소송 비용	
<input type="checkbox"/>	수집의 라이프 사이클 비용 평가 및 계획

□ 유지 및 검토 지침 개요

- 시간이 지나면서 변화하는 요구와 상황을 반영하기 위해 수집 및 수집 정책을 검토할 필요성이 제기될 수도 있다. 디지털 환경에서 보존을 위한 선택에 대한 조기 결정(아날로그 환경에서 흔히 볼 수 있는 사후 인식 기간 없음)의 필요성은 디지털자료의 보존 수명주기에서 향후 검토가 필요할 수 있음을 의미한다.
- 수집 수준이 적용된 디지털 도서관 환경에서 모든 수집 수준 범주의 디지털자료는 변화하는 요구와 상황을 충족시키기 위해 필요에 따라 정기적인 검토를 거쳐 한 수준에서 다른 수준으로 재지정 되거나 철회 혹은 취소될 수 있다. 그러나 영구 보존을 위해 선택된 자료의 경우 드물지만 엄격하게 통제된 상황에서 검토 및 탈퇴가 일어날 것으로 예상된다.
- 미러된 또는 라이선스 획득 자원과 같은 다른 컬렉션 레벨에는 다음 사항이 포함될 수 있다.
 - 허용 수준 이하로 사용이 지속적으로 감소한다.
 - 더 높은 수준의 품질 또는 상당히 저렴한 비용으로 다른 곳에서 콘텐츠를 사용할 수 있다.
- 콘텐츠는 대체되거나 실행 품으로 유지관리를 정당화하기에 더 이상 충분히 정확하지 않은 경우, 그 콘텐츠는 후속 판 또는 철회 정보와 함께 보존될 수 있다.
 - 라이선스 또는 데이터 교환 계약의 만료 또는 종료 및 디지털자료의 데이터 공급업체로의 철회 혹은 반환
 - 데이터 리소스를 유지하는 데 드는 비용은 받은 가치나 이익보다 큼
 - 공급 업체가 제공하는 품질 서비스의 저하 또는 데이터 리소스 내부의 인덱싱, 이미징 또는 기타 특성의 업데이트 불량으로 인한 콘텐츠의 액세스 가능성 저하
- 기록보관소 및 기록관리 전문직 내에서 보존 기간 및 일정의 사용은 잘 정립되어 있다. 기록은 보관 기간이 끝나면 폐기되거나 더 오래 보관되거나 장기보존을 위해 기관으로 이전될 수 있다.
- 어떤 수집 환경에서든 지명된 개인이나 거버넌스 및 수집 개발을 담당하는 조직의

섹션에 할당된 명확한 책임으로 보존 및 검토 프로세스에 필요한 서면 절차가 중요하다. 검토를 위한 기간, 상황 및 허가 절차는 명확하게 기술되어야 한다. 유지 및 검토 일정 문서 자체는 새롭게 등장하는 조직의 요구 사항에 보조를 맞추기 위해 주기적으로 검토되어야 한다.

□ 저장

1) 개요

- 디지털 보존을 위한 스토리지 기술의 사용은 지난 20년 동안 극적으로 변화했다. 이전에는 개별 CD, 테이프 등의 개별 미디어 항목을 사용하여 디지털 자료를 저장한 다음 주기적으로 마이그레이션하여 저하 및 노후화를 해결하였다.
- 오늘날은 온라인 액세스의 문화에서 쉽고 빠르게 검색할 수 있어야 하며 보존되어야 할 더 중요한 디지털 자료에 탄력적인 IT 스토리지 시스템을 사용하는 것이 일반적이다. 이러한 방식으로 디지털 자료는 기본 저장 메커니즘과 분리하여 다양한 보존 활동을 독립적으로 처리할 수 있는 이점을 가진다.

2) 탄력적인 스토리지 시스템

- 탄력적인 IT 스토리지 시스템은 내장된 중복성 및 복구 기능을 사용하여 다양한 장애 모드에 대한 복원력을 제공하는 서버 내에 포함된 스토리지 미디어로 구성된다. 예를 들어, 스토리지 시스템은 RAID(Redundant Array of Independent Disks)의 하드 디스크 드라이브, 테이프 라이브러리의 데이터 테이프 또는 HSM(Hierarchical Storage Management System) 스토리지 유형의 조합일 수 있다. 온사이트 스토리지 또는 원격 클라우드 스토리지와 여러 사이트 및 시스템에서 디지털 자료의 자동 복제를 포함할 수 있다.
- 이러한 시스템은 시간이 지남에 따라 쓸모없게 되고 디지털 자료는 스토리지 시스템 간에 정기적으로 마이그레이션되어 폐기되어야 한다. 스토리지 시스템 간 마이그레이션은 파일 포맷 간의 마이그레이션과는 별도로 이루어지며 보존 요구 사항

을 충족시킬 수 있도록 적절한 감독이 이루어져야하지만 대부분 IT 문제로 처리할 수 있다. 장점은 데이터 스토리지를 위한 IT 시스템의 사용이 훨씬 더 빠른 액세스, 보다 확장 가능한 솔루션, 더 쉬운 관리 및 궁극적으로는 특히 규모의 감소를 제공할 수 있다는 것이다.

- 표준 IT 스토리지 솔루션과 장기보존의 추가 요구 사항 간의 차이점을 이해하는 것이 중요하다. 이러한 차이점을 IT 부서 또는 스토리지 서비스 공급자에게 설명하고 시스템 또는 서비스를 조달할 때 이러한 요구 사항을 지정할 수 있어야 한다. 표준 스토리지 시스템은 현재 사용 중인 디지털 객체용으로 설계되었다. 일반적으로 백업 절차가 포함되지만 일반적으로 디지털자료의 장기보존을 보장하기 위한 보다 엄격한 요구 사항을 충족하지 못한다. 백업 및 디지털 보존은 동일한 것이 아니며 많은 IT 부서 또는 전문가가 이를 인식하지 못할 수도 있다.
- 보존 저장소에 대한 요구 사항을 충족하는 방법은 다양하며 조직의 상황에 따라 규모와 복잡성이 다양하다. 사내 자원을 평가하고 아웃소싱 및 클라우드 스토리지 옵션을 고려해야 한다. 취해진 접근법은 종종 수집 크기 및 수집 복잡성 및 조직 내에서 사용 가능한 리소스에 따라 달라진다. 기본 설정으로 보존 저장 요구 사항을 충족할 수는 있지만 컬렉션이 증가함에 따라 확장성 및 자동화와 같은 문제를 해결해야 한다.
- 다음 [표 4-10]은 디지털 보존을 위한 IT 스토리지 시스템 사용 원칙에 대한 내용이다.

[표 4-10] 디지털 보존을 위한 IT 스토리지 시스템 사용 원칙

디지털 보존을 위한 IT 스토리지 시스템 사용 원칙	
다음은 보존을 위한 스토리지 시스템을 설계하거나 선택할 때 채택되어야하는 원칙을 나타낸다.	
1	<p>중복성 및 다양성</p> <p>디지털자료의 여러 독립 사본을 만들어 이를 서로 다른 지리적 위치에 저장한다.</p> <p>온라인 스토리지 시스템과 오프라인 미디어를 함께 사용한다.</p> <p>다양한 유형의 스토리지 기술을 사용하여 위험을 분산시키고 데이터 안전, 쉬운 액세스 및 관리 가능한 비용의 균형을 이룬다.</p>
2	<p>수리, 모니터링,</p> <p>체크섬과 같은 고정 값 측정을 사용하여 디지털자료의 각 복사본의 무결성을 기록하고 정기적으로 모니터링한다.</p> <p>손상이나 분실이 발견되면 다른 사본 중 하나를 사용하여 교체품을 만든다.</p> <p>디지털자료뿐만 아니라 별도의 데이터베이스 또는 시스템에 고정 정보를 저장한다.</p>
3	<p>기술 및 공급 업체 감시, 위험 평가 및 사전 예방적 마이그레이션</p> <p>스토리지 기술, 제품 및 서비스의 수명이 짧음을 이해한다.</p> <p>기술 감시를 사용하여 마이그레이션이 필요한시기를 판단한다.</p> <p>스토리지 공급 업체 또는 스토리지 솔루션의 생존 가능성에 주목한다.</p> <p>디지털자료가 위험에 처하기 전에 스토리지 마이그레이션에 적극적으로 참여한다.</p>
4	<p>통합, 단순성, 문서화, 출처 및 감사 추적</p> <p>보존에 사용되는 기존 미디어 유형 및 스토리지 시스템의 확산을 최소화한다.</p> <p>최소한의 저장 시스템에 디지털자료를 통합한다(위의 중복 요구 사항 적용).</p> <p>스토리지 시스템을 설치하고 운영하는 방법뿐만 아니라 스토리지 시스템에 디지털자료를 수집하여 전송하는 방법을 문서화한다.</p> <p>이를 사용하여 데이터 신뢰성에 대한 감사 정보를 제공한다.</p>

3) 스토리지 안정성

- 온사이트 또는 클라우드의 스토리지 솔루션을 살펴볼 때 얼마나 신뢰할 수 있는지, 문제가 발생하면 데이터 손실 측면에서 무엇이 의미가 있는지 등의 문제가 발생한다.
 - 디지털 보존의 관점을 비롯하여 스토리지 안정성을 설명하고 측정하고 예측하는 방법에 대한 몇 가지 중요한 고려 사항은 다음과 같다.
 - IT 스토리지 기술은 일반적으로 신뢰할 수 있는 기능을 제공한다. 실패는 상대적으로 드물게 발생하지만 일어날 수 있다. 유감스럽게도 개인 수준에서 특정 유형의 장애를 경험할 수 없기 때문에 스토리지 기술이 일반적으로 실제보다 신뢰성

이 높다고 가정해야한다.

- 실패 및 오류는 비교적 드문 경우이기 때문에 공급 업체의 안정성 통계는 일반적으로 모델 및 시뮬레이션을 기반으로 하며 실제로 발생하는 현상을 장기간 관측하지는 않는다.
- 상관관계를 고려할 때 데이터 손실의 가능성이 극적으로 증가한다. 상관관계는 시스템의 일부 또는 디지털자료의 다른 사본을 독립적으로 간주할 수없는 부분이다. 시스템의 한 부분이나 디지털자료의 사본에 문제가 있는 경우 다른 부분이나 사본에 문제가 있을 수 있다. 이러한 상관관계 요인은 신뢰성 통계가 다루는 특정 장애 모드보다 훨씬 더 중요할 수 있다.
- 이 결과 다음과 같은 권고 사항을 제시할 수 있다.
 - 제조업체가 얼마나 독창적으로 설계했는지에 관계없이 IT 스토리지 솔루션에서 발생하는 오류를 계획한다. 실제로 실패율은 제조업자 통계치보다 높을 수 있다.
 - 데이터 손실은 스토리지 기술 자체뿐만 아니라 IT 스토리지 사용과 관련된 적절한 프로세스 및 절차를 마련하지 않아서 발생할 수 있다. 적절한 위험 평가는 이러한 문제를 식별하고 관리하는 방법이다.
 - 최상의 전략은 여러 위치에 디지털자료의 여러 독립 사본을 작성하고 가능하면 다른 기술을 사용하여 디지털자료를 저장하는 것이다. 여기에는 모든 사본의 데이터 무결성을 능동적으로 정기적으로 검사하여 문제가 발생하는 장소와 문제가 무엇이든 문제를 감지할 수 있는 프로세스가 포함되어야한다.

4) 다중 사본 저장 전략

- 디지털 저장 기술은 디지털 객체의 장기보존에 여러 가지 위험을 안겨준다. 이러한 위험은 하나 이상의 스토리지 시스템과 적어도 두 개의 데이터 복사본이 포함된 디지털 스토리지 전략을 사용하면 줄일 수 있다.
- 스토리지 전략의 특징은 다음과 같다.
 - 디지털자료의 여러 독립 사본이 존재한다.
 - 이들 사본은 지리적으로 다른 위치로 분리되어있다.

-
- 사본은 다른 저장 기술을 사용한다.
 - 사본은 온라인 및 오프라인 저장 기술의 조합을 사용한다.
 - 스토리지를 적극적으로 모니터링하여 문제를 신속하게 감지하고 수정한다.
- 디지털 스토리지 전략은 기본 수준의 보호 및 디지털 콘텐츠에 대한 액세스부터 시작하여 더 높은 수준의 데이터 안전성과 보안을 제공하는 보다 자동화되고 확장 가능한 접근 방식을 단계적으로 구현할 수 있다.
 - 디지털 콘텐츠에 대한 위험은 다양한 출처에서 비롯되며 디지털 스토리지 전략은 디지털 스토리지 비용과 이러한 위험을 줄이는데 도움이 된다. 고려해야 할 위험의 예로는 화재, 홍수, 적절한 프로세스 또는 절차의 시작 또는 실패, 악성 공격, 미디어 성능 저하 및 스토리지 시스템 및 기술의 노후화 등이 있다.
 - 콘텐츠 손실의 많은 예가 반드시 스토리지 기술의 기술적 결함으로 발생하는 것은 아니지만 사람의 실수, 예산 부족 또는 스토리지 마이그레이션 계획 등으로 인해 발생할 수 있음을 인식하는 것이 중요하다.
 - 디지털자료의 사본을 두 개 이상 만드는 것은 기본 데이터 안전 수준을 달성하는데 있어 기본적인 것이다. 각 복사본에 대해 서로 다른 유형의 저장소를 사용하면 위험을 분산시키고 한 기술의 문제가 다른 기술에 영향을 미치지 않도록 할 수 있다. 각 사본이 저장되는 방식은 허용 가능한 전반적인 수준의 비용, 위험 및 복잡성을 달성하도록 조정될 수 있다.
 - NDSA(National Digital Stewardship Alliance) 보존 수준에 따라 저장 및 지리적 중복을 통해 디지털 보존을 지원할 수 있는 4가지 권장 수준은 다음과 [표 4-11]과 같다(NDSA 2013).

[표 4-11] NDSA 디지털 보존을 위한 권장 수준

수평	접근	해결 된 위험 및 이점
1	<ul style="list-style-type: none"> • 함께 배치되지 않은 디지털자료의 완전한 사본 2부. 하나의 사본은 오프라인이어야 한다. • 이기종 미디어 (광디스크, 하드 드라이브 등) 디지털자료의 경우 매체 및 스토리지 시스템에 콘텐츠를 저장한다. 	<ul style="list-style-type: none"> • 스토리지 시스템 장애를 포함한 다양한 문제로부터 복구할 수 있는 기본 기능. 한 사본의 손실 또는 손상은 다른 사본을 사용하여 복구할 수 있다. • 단일 스토리지 시스템에서 디지털자료를 보다 쉽게 관리할 수 있다.
2	<ul style="list-style-type: none"> • 적어도 3개의 완전한 사본. • 다른 지리적 위치에 적어도 하나의 사본. • 스토리지 시스템 및 스토리지 미디어와 이를 사용하기 위해 필요한 것을 문서화한다. 	<ul style="list-style-type: none"> • 위와 마찬가지로 자연 재해 및 기타 주요 사건으로부터의 보호. • 우수한 수준의 액세스 및 디지털자료 안전성. • 직원은 따라야 할 분명한 정책과 절차를 갖고 있으므로 보다 효율적이고 비용이 절감되며 직원의 변경 사항을 관리할 수 있다.
3	<ul style="list-style-type: none"> • 다른 재난 위험이 있는 지리적 위치에 적어도 하나의 사본. • 스토리지 시스템 및 미디어에 대한 모니터링 및 마이그레이션 프로세스. 	<ul style="list-style-type: none"> • 위와 마찬가지로 기술 노후화와 관련된 장기적인 위험으로부터 보호한다. • 마이그레이션 및 재해 발생시에도 콘텐츠에 대한 지속적인 액세스가 가능하다.
4	<ul style="list-style-type: none"> • 다른 재난 위험이 있는 지리적 위치에 최소 3부. • 현재 액세스할 수 있는 미디어 또는 시스템에 파일과 그림을 보관할 종합 계획을 마련한다. 	<ul style="list-style-type: none"> • 위와 같이 우발적인 손실 및 악의적인 공격, 공급 업체 고정 및 예산 불안정과 같은 모든 위험 요소를 해결한다. • 콘텐츠 가용성이 높고, 비용을 예측하고 관리할 수 있으며 신뢰할 수 있는 저장소 인증을 획득할 수 있다.

5) 스토리지 시스템 노후화 및 위험 관리

- 스토리지 기술 및 솔루션을 사용하려면 디지털 보존을 지원하기 위한 효과적인 접근 방식으로 신중한 계획 및 관리가 필요하다. 제대로 완료되면 결과적으로 매우 우수한 수준의 데이터 안전성이 필요하다.
- IT 스토리지 기술이 실패하거나 데이터 손상을 야기할 수 있으며 미디어 및 시스템의 수명은 일반적으로 짧다(예: 3-5년). 솔루션이 빨리 폐기되고 디지털자료가 위험에 처하지 않도록 하려면 마이그레이션이 필요하다. 이 컨텍스트에서 마이그레이션이란 기존 스토리지 시스템에서 새 스토리지 시스템으로 데이터를 이전하는 것을 의미한다. 디지털자료 자체는 변하지 않지만 저장소 솔루션은 변하지 않는다.

IT 부서 또는 스토리지 서비스 제공 업체는 스토리지 수준에서 마이그레이션을 고려할 것이다. 이는 파일 형식이 변경되는 파일 형식 마이그레이션과 대조적이지만 파일 저장 방법은 변경되지 않는다.

□ 보존 계획

1) 개요

- 보존 계획은 저장소가 보유하고 있는 디지털자료의 지속 가능성 또는 액세스에 영향을 줄 수 있는 변경 사항을 모니터링하기 위한 디지털 저장소 내의 기능이다. 변경 사항은 저장소 내에서, 저장소가 상주하는 조직 내에서 또는 저장소 외부와 조직 외부에서 발생할 수 있다. 다음 영역에서 변경 사항을 모니터링할 수 있다.
 - 기술 감시(Technology watch): 패키징, 저장, 포맷, 도구, 환경, 액세스 메커니즘
 - 지정된 지역사회(Designated communities): 사용자의 필요와 기대, 제공자의 필요와 기대, 기계 대 기대 접근을 위한 도구, 이용자와 제공자의 공식 피드백.
- 보존 계획의 개념은 OAIS 표준의 기능 모델 내에서 정의된다. 여기에서는 주로 OAIS 정의내의 모니터링 구성 요소에 중점을 둔다. OAIS의 ‘모니터 기술’과 ‘지정된 지역사회 모니터링’ 기능은 보존 계획 활동을 알리는 조사를 제공한다. 이는 스토리지 및 액세스 기술의 혁신이나 지정된 커뮤니티의 범위 또는 기대치 변경과 같이 보관된 정보에 대한 액세스를 보존하고 유지하는 능력에 영향을 줄 수 있는 외부 환경 및 위협의 변화에 대해 리포지터리에 경고한다. 그런 다음 보존 계획에서는 이러한 변경 사항을 수용하기 위해 저장소의 정책과 절차를 업데이트하기 위한 권장 사항을 개발한다.
- 보존 계획 기능은 끊임없이 진화하는 사용자 및 기술 환경에 대한 OAIS의 안전장치를 나타낸다. 리포지터리의 책임을 수행하는 능력에 영향을 미치는 변경 또는 위협을 감지하고 이를 해결하기 위한 전략을 수립하며 아카이브 시스템 내에서 이러한 전략을 구현하는데 도움을 준다.

2) 보존 계획의 목적

- 디지털자료 보존을 위한 조치를 취하기 위한 트리거 식별
 - 변경이 확인된 경우 위험 평가 프로세스를 사용하여 저장소의 디지털자료에 대한 중대한 위험을 나타내는 변경 사항을 분석하고 식별할 수 있다.
 - 적절한 보존 조치를 결정하기 위한 보존 계획 수립에 따라 위험을 처리하고 완화할 수 있다. 이 경우 모니터링 또는 기술 감시 프로세스는 추후 분석, 보존 계획 및 해당되는 경우 디지털자료 보존을 위한 트리거 포인트를 식별하는 것이다.
- 보존 활동을 알리기 위한 지식 기반 구축
 - 보존 계획 활동의 일환으로 내부 및 외부 요소를 모니터링하는 프로세스는 조직의 지식 기반을 알릴 수 있으며 이를 통해 디지털 보존 활동을 효과적으로 수행할 수 있는 능력을 향상시킬 수 있다. 예를 들어 조직의 “지식 기반”은 새로운 소프트웨어 도구의 기능 또는 기존 도구의 폐기 및 사용할 수 없는 기능에 대한 정보로 보강될 수 있다.
 - 경우에 따라 이 프로세스는 개별적으로 또는 조직 내에서 수행하는 것이 가장 좋을 수도 있지만 협업 방식으로 보다 효율적으로 수행할 수도 있다. 디지털 보존에 필요한 광범위한 지식과 폭 넓은 지식은 자연스럽게 협업 접근 방식을 선호한다.
- 보존 계획 서비스의 구현
 - 기술 감시가 필요한 정도는 기관이 수행할 수 있는 포맷 및 매체에 대한 통제 또는 제어의 정도에 따라 달라질 수 있다. 대부분의 기관에서 모든 사람이 동일한 소프트웨어 및 버전을 사용하고 제품이 개발되면 새로운 버전으로 마이그레이션 되도록 기관 표준을 개발해야 한다.
 - 이를 포함하는 효과적인 기술 감시 또는 IS 전략을 구현하지 않으면 디지털 자산에 대한 액세스가 손실되고 비용이 높아질 수 있다. 예를 들어 디지털 포렌식을 통해 액세스를 다시 설정하는 것이 가능하다.
 - 디지털 보유 및 위험 평가 및 실행 계획에 대한 소급 조사가 많은 기관에서 기술 감시 실행에 앞서 필요한 첫 번째 단계일 수 있다.
 - 통합 보존 시스템 및 개별 도구 및 레지스트리도 이 기능을 지원할 수 있다.

□ 보존 활동

1) 포맷의 기술적 노후화

- 기술 노후화는 오랜 기간 디지털 보존의 중요한 과제로 고려되어 왔으나 최근 몇 년간의 연구 결과에 따르면 포맷 노후화가 이전에 두려워했던 것만 큼 널리 퍼진 것은 아니라고 한다(Rosenthal 2015a, Jackson 2012). 디지털자료가 여러 세대의 기술 변화를 극복하기 위해서는 반드시 인식되고 해결되어야 할 하나의 문제이지만 분명히 유일한 도전은 아니다. 많은 기존 파일 형식이 여전히 지원되며 사용 가능하다. 대다수의 파일 형식은 쓸모없는 것이 아니라 일반적으로 이해되고 잘 지원될 것이다.
- 노후화에 대한 간단한 정의는 시대에 뒤떨어지거나 더 이상 사용되지 않는 프로세스이다. 노후화가 발생하는 정확한 순간을 정확히 찾아내는 것은 어려울 수 있다. 특히 최근에는 쓸모없게 된 자료의 경우 특히 더하다. 유사하게 한 기관은 다른 사람들이 새 버전으로 업그레이드한 후에도 한참 동안 기존 소프트웨어를 계속 사용하고 유지 관리할 수 있다. 따라서 '제도적 노후화'에 대해 이야기하는 것이 더 유용할 수 있다.
- 모든 파일이 자체 하드웨어 및 소프트웨어 종속성을 가지고 있기 때문에 더 이상 지원되지 않는 기능이 문제가 된다. 변경은 사용자가 콘텐츠의 의미 또는 해석의 의미를 손상시킬 때 문제가 된다. 디지털 보존 활동의 핵심 목표는 컴퓨팅 기술의 세대 간 변화에도 불구하고 보존되는 자료의 무결성과 신뢰성을 유지하는 것이다.

2) 보존전략

- 디지털정보를 보존하기 위해 사용할 수 있는 기술적 전략으로 여기에서 다룰 기술은 다음과 같다.
 - 포맷 마이그레이션
 - 에뮬레이션
 - 컴퓨터 박물관

- 포맷 마이그레이션(Format migration)

- 포맷 마이그레이션은 가장 널리 사용되는 보존전략 중 하나이며 대부분의 디지털 보존 시스템에는 마이그레이션 솔루션을 가정하는 기능 또는 시스템 데이터가 포함되어 있다.
- 포맷 마이그레이션은 데이터의 구조와 내용을 변경할 가능성이 있는 모든 개입과 마찬가지로 오류 및 정보 손실을 초래할 수 있다. 따라서 가능한 정보 손실을 측정하기 위한 매트릭을 정의하고 이를 사용하여 포맷 마이그레이션의 정확성과 품질에 대한 테스트를 수행하는 것이 중요하다.
- 이 프로세스와 관련된 실질적인 문제에는 마이그레이션 시기가 포함된다. 대대로 마이그레이션하는 것이 더 나은지 또는 일부 세대를 건너 뛸 수 있어야 하는지에 대한 것이다.

- 에뮬레이션(Emulation)

- 에뮬레이션은 아카이브가 원본 파일에서 직접 사용자에게 액세스 권한을 보존하고 전달할 수 있도록 하는 대체 솔루션을 제공한다. 이 기술은 정보의 내용뿐만 아니라 응용 프로그램의 원래 동작과 모양과 느낌을 보존하려고 시도한다. 원본 프로그램만 형식에 대한 권한이라는 관점에 기반하며 이는 게임이나 대화형 응용 프로그램과 같이 여러 가지 상호의존성이 있는 복잡한 개체에 특히 유용하다.
- 에뮬레이터는 현재 컴퓨터 아키텍처에서 실행되지만 이전 컴퓨터와 동일한 기능 및 동작을 제공하는 프로그램이다. 이 접근법은 많은 기술 기관과 공동으로 기술 기관의 지원을 받아 왔으며 최근에는 문화 유산을 위한 에뮬레이션 솔루션을 구현하는데 있어 주목할 만한 성공을 거두었다.
- 에뮬레이션의 특별한 이점은 모든 개체가 동일한 운영체제 또는 하드웨어 스택에서 배달해야 하는 경우 많은 수의 개체에 대한 액세스를 제공하기 위해 단일 솔루션을 배포할 수 있다는 것이다. 그러나 레거시 컴퓨팅 장비를 사용에 있어 사용자는 어려움을 겪을 수 있다.

- 컴퓨터 박물관(Computer museums)

- 이 방법론은 컴퓨터 및 시스템 소프트웨어(운영 체제, 드라이버 등)는 물론 데이터 및 응용 프로그램의 유지를 제안한다. 모든 플랫폼을 잘 유지하고 기계와 프로

그램을 유지 관리하고 사용하는 데 필요한 모든 지식을 보유하려면 노력을 기울여야 한다. 따라서 이 전략은 장기적인 해결책이 아닌 임시적인 방법이 되는 경향이 있다. 캘리포니아의 컴퓨터 역사박물관(Computer History Museum)과 캠브리지(Cambridge)의 컴퓨팅 역사센터(Centre for Computing History)와 같은 박물관이 존재한다. 이것들은 보존 서비스를 제공하지 않지만 일반적으로 기계를 작동 순서대로 유지한다.

□ 접근

1) 개요

- 보존과 접근 사이에는 항상 강한 연관성이 있다. 전통적인 자원의 정보 내용을 보존하는 주요 목적은 현재와 미래 세대 모두에게 접근할 수 있도록 하는 것이다. 디지털 객체에 대한 액세스를 유지하는 것은 디지털 보존 프로그램의 핵심 목표이지만 리소스를 확보하기 전에 리소스 수명주기 동안 더 적극적인 관리가 필요하다.
- 그러므로 보존 단계 초기부터 이상적으로는 획득 단계 초기에 접근이라는 중요한 문제를 고려하는 것이 필수적이다. 여기에서는 고려해야 할 주요 쟁점, 접근 준비를 계획할 때 결정해야 할 사항, 그리고 이것이 보편적으로 보존에 미칠 수 있는 영향을 파악한다.

2) 사용자 이해하기

- 잠재적인 사용자를 이해하는 것은 디지털 객체에 대한 액세스 제공을 계획할 때 필수적이고 보다 광범위한 디지털 보존 활동의 주요 고려 사항이기도 하다. 잠재적인 사용자에게 대해 얻은 지식은 라이프 사이클 전반에 걸친 의사 결정에 영향을 미치지만 적절한 액세스 전달 솔루션을 선택할 때 리소스 및 기술적 고려 사항과 균형을 이룰 때 가장 큰 비중을 차지할 것이다. 사용자 커뮤니티와 그들의 요구 사항을 체계적이고 객관적으로 파악하는 것이 중요하다.
- 이 정보를 수집하는 데 사용되는 방법론은 조직의 상황에 따라 달라지나 가능한 용

선 및 도구에는 다음이 포함될 수 있다.

- 현재 사용 현황 분석(물리적 및 디지털 객체에 대한 액세스 요청, 웹사이트 통계 등)
 - 설문 조사
 - 포커스 그룹
 - 인터뷰
 - 사용 사례
 - 작업 분석
- 사용자 분석을 수행할 때 기존 사용자와 비사용자를 모두 고려하는 것이 중요하다. 비사용자와의 상호 작용이 본질적으로 더 어렵지만 잠재적인 새로운 시장을 식별할 뿐만 아니라 현재의 장벽을 이해하는 데 유용한 프로세스가 될 수 있다. 사용자 커뮤니티 개발을 지속적으로 모니터링하는 것이 중요하며 이는 조직 내의 표준 보존 계획 활동에 통합되어야 한다.

3) 접근 형식

- 접근 계획 시 주요 고려 사항은 디지털 객체가 사용자에게 전달되는 형식이다. 디지털 보존 프로그램의 최우선 목표라는 관점에서 보존과 접근 사이에는 강력한 연결이 존재하지만, 그것들을 명확히 구분할 필요가 있다. 보존 사본과 접근 사본을 분리하는 기술적, 합법적 및 실용주의적인 이유가 복합적으로 존재할 수 있으므로 디지털 객체의 접근 사본을 다른 형식으로 사용자에게 전달하는 것이 바람직하거나 필요하다.
- 액세스 형식을 선택할 때 조직에서 고려해야 할 몇 가지 내용이 다음과 같다.
 - 객체 유형에 대해 가장 일반적으로 사용되고 널리 지원되는 형식은 무엇인가?
 - 사용자가 제안된 파일 형식을 지원하는 무료 뷰어나 소프트웨어에 접근할 수 있는가?
 - 어떤 파일 크기로 생성되며 사용자에게 전달할 때 어떤 영향을 미치는가?
 - 포맷을 쉽게 사용할 수 있는가?

-
- 사용자가 물체에 접근/사용할 수 있도록 안내 또는 지원 문서가 필요한가?
 - 조직은 액세스 요구 사항이 다른 별도의 사용자 커뮤니티가 있는가?

4) 접근에 대한 법적 문제

- 배포되는 기술 솔루션과 자료에 액세스할 수 있는 사람 모두에게 영향을 끼치게 되는 디지털 객체에 대한 액세스를 제공할 때 해결해야 할 다양한 법적 문제가 있다. 이는 앞서 언급한 바와 같이 수집과 중복되는 주요 액세스 고려 사항 중 하나이다. 이 정보가 없으면 액세스를 적절하게 관리하지 못할 수 있으며 잠재적인 법적 위험에 노출 될 수 있다.
- 고려해야 할 법적 문제는 다음과 같다.
 - 민감도 및 데이터 보호와 관련된 사용 제한
 - 조기 접속이 계약 위반을 나타낼 수 있는 콘텐츠에 대한 합의된 엠바고
 - 저작권과 같은 지적 재산권 관리
- 특히, 지적 재산권의 관리는 양도 절차와 관련이 있으며, 양도 계약과 소유권 계약 및 저작권 라이선스가 신중히 고려되어야 한다. 라이선스는 제 3자 라이선스를 포함하여 허용된 접근 및 재사용 권한을 명확하게 명시해야 한다. 권한 관리 시스템을 통하든 다른 방법을 통해 관리하든 관계없이 정책 및 절차에 명확하게 표현되어야 한다.

5) 접근 제공 형식

- 조직에서 반드시 거쳐야 할 최종 결정 사항은 다음과 같다.
 - 정책
 - 절차
 - 무료 또는 유료 서비스
 - 온라인/오프라인 액세스 및 제공되는 액세스 환경
 - 장애인을 위한 액세스

- 스토리지 및 보안

- 액세스 사본이 디지털자원의 유일한 사본일 경우 도난 또는 손상으로 인한 손실의 위험은 분명히 매우 높으므로 이에 대비해야 한다.

□ 정보 보안

- 정보 보안 문제는 시스템 보안(예: 디지털 보존 및 네트워크 시스템, 서비스를 외 내부 위협에 노출되지 않도록 보호), 수집 보안(예: 손실 또는 변경으로부터 콘텐츠 보호, 저장소 프로세스의 권한 부여 및 감사), 그리고 법률 및 규제 측면(예: 디지털자료의 개인 정보 또는 기밀 정보, 안전한 액세스, 수정)과 관련된다.
- 정보 보안은 일반적으로 정보 시스템에 대한 복잡하고 중요한 주제로 조직 내의 관련 전문 지식에 의존하고 일반 정보 보안 절차 및 조언을 위해 정부 및 기타 네트워크를 통하는 것이 중요하다. 특정 디지털 보존 절차 및 요구 사항에 대한 적절한 옹호가 필요할 수도 있다.
- 엄격한 보안 절차는 다음과 같다.
 - 법적 및 규제 요구 사항 준수 보장
 - 부주의한 변경이나 고의적인 변경으로부터 디지털자료를 보호
 - 책임 요구 사항을 충족시키는 감사 추적을 제공
 - 잠재적인 내부 보안 침해에 대해 역지력을 행사
 - 디지털자료의 진본성을 보호
 - 도난 또는 분실로부터 보호
- 장기보존을 위해 선택된 많은 종류의 디지털자료에는 비인가 사용자가 액세스하지 못하도록 보호해야하는 기밀 정보가 포함될 수 있다. 많은 경우 조직의 법적 의무 또는 규제 의무 사항에 해당될 수 있다. 이러한 자료는 보안 침해로부터 시스템을 보호하기 위해 조직의 정보 보안 정책에 따라 관리해야한다. 경우에 따라 조직의 정보 보안 정책이 디지털 보존 활동에 영향을 줄 수 있으며 프로세스를 용이하게 하기 위해 정보 거버넌스 및 ICT 팀의 지원을 필요로 할 수도 있다.

-
- 암호화와 같은 정보 보안 방법은 보존 프로세스의 복잡성을 증가시키므로 가능한 경우 보관용 사본에 적용은 피해야한다. 따라서 다른 보안 접근법이 민감한 암호화되지 않은 파일에 보다 엄격하게 적용될 필요가 있다. 여기에는 제어된 위치(보안실)의 잠긴 단말기에 대한 액세스 제한 또는 원격 액세스에 대한 강력한 사용자 인증 요구 사항이 포함될 수 있다.
 - 그러나 이러한 대체 접근법은 항상 충분하거나 실현 가능하지 않을 수도 있다. 또한 수집한 파일이 암호화되어 있을 수 있으므로 암호화, 암호화 키 관리 및 디지털 보존에 대한 영향과 같은 정보 보안 옵션을 알고 있어야 한다.

4.3 디지털자원 보존 조직

4.3.1 조직 구성을 위한 개념적 프레임

□ 조직단위

- 국립중앙도서관의 디지털 장기보존 조직은 보조(계선)기관의 성격을 지니므로 중앙행정기관의 보조(계선)기관 조직을 검토한다.
- 계선조직(line organization)이란 조직의 최고관리자를 중심으로 수직적이며 위계적으로 연결되어 있는 운영조직이며 조직의 목표를 수행하며 이에 대한 책임을 진다. 정부조직의 장관-차관-실장-국장-과장-계장으로 연결되는 조직이 대표적이다(윤재풍 2014). 이 중 보조기관은 계선기관장의 직무를 보조하는 직위의 계선기관으로서 중앙행정기관의 보조기관은 차관, 차장, 실장, 국장 등으로 법에 구체적으로 명시되어 있다(정부조직법 제2조 제3항).
- 정부조직법에 따른 보조(계선)기관 체계는 실, 국, 과 등으로 위계적이며 구체적인 내용은 다음과 같다. 아래의 내용은 정부조직법 등 관련 법령 및 조석준·임도빈(2016)을 참조하여 재구성하였다.

□ 한국 중앙행정기관의 보조(계선)기관

1) 실

- 중앙행정기관의 보조기관의 핵심 조직은 국이다. 실의 경우 ‘본부’라는 명칭을 사용하기도 하며 ‘국’으로서는 그 목적달성이 곤란하다고 인정되는 특별한 사유가 있는 경우’에 설치 가능하나(행정기관의 조직과 정원에 관한 통칙 제14조 1항) 대개 모든 부처에 실이 설치되어 있다.
- 이는 상위직 인사적체 등의 실질적 이유로 국의 상위 조직으로 실이 설치되는 경우가 많기 때문이다.

2) 국

- 중앙행정기관의 상급 보조기관으로서 장관, 차관 아래 가장 핵심이 되는 조직 단위이다. 국장은 일반적으로 고위공무원단에 속하는 나등급 일반직공무원이 보임된다.
- 국은 첫째, 중앙행정기관의 소관업무를 업무의 성질이나 양에 따라 수 개로 분담하여 수행할 필요가 있고, 둘째, 업무의 한계가 분명하고 업무의 독자성과 계속성이 있는 경우에 둘 수 있다(행정기관의 조직과 정원에 관한 통칙 제14조 제1항).
- 내용상 기준은 없으며 부의 기능 중 대분류에 해당하는 업무를 담당하며 국민 전체 혹은 정책집단에 대해 정책프로그램을 기획 및 집행할 근거 법률과 사업성 예산을 가지고 있다.

3) 과

- 중앙행정기관의 전형적 하급 보조기관이다. 기본적으로 실 또는 국의 하위 조직으로 과, 반, 팀 등을 설치할 수 있다(정부조직법 제2조 제2항).
- 실이나 국의 소관업무를 업무량 혹은 성질에 따라 수개로 분담 수행의 필요가 인정되는 경우에 과, 반, 팀 등을 둘 수 있으며 그 기준은 부처마다 다양하며 모호하다. 소관 정책의 근거가 되는 법을 하나 이상 가지고 있다.
- 과장·반장·팀장은 3급 내지 5급 공무원으로 보하며 과장은 고위공무원단 나등급, 혹은 3급 혹은 4급으로 보임한다(행정기관의 조직과 정원에 관한 통칙 제14조 제3항).

4) 종합

- 중앙행정기관의 보조(계선)기관의 핵심 조직은 ‘국’이며 국은 내용상 특별한 기준은 없으나 중앙행정기관의 기능 중 대분류에 해당하는 업무를 담당한다. ① 중앙행정기관의 소관업무를 업무의 특성과 범위에 따라 수 개로 분담하여 수행할 필요가 인정되고 ② 업무의 명확한 한계, 독자성, 지속성이 있는 경우에 설치할 수 있다. 국장은 고위공무원단 나등급의 일반직공무원으로 보임한다.

- 국의 소관업무를 업무량이나 특성에 따라 분할할 필요가 인정되는 경우 과를 둘 수 있다. 과장은 고위공무원단 나급이나 3급 혹은 4급으로 보임한다.

□ 직무분석

1) 직무분석의 개념

- 직무분석은 타 직무와 차별화되는 직무의 고유한 특성과 정보를 체계적으로 유형화 및 기술하는 과정이며 이를 통해 직무 종류 및 유형을 결정한다(이창길 2016).
- 직무분석은 직무기술서 작성, 인적자원의 모집 및 선발, 직무평가 및 보수등급 결정, 직무설계·재설계 등을 위한 기초자료로 활용된다(이창길 2016).

2) 직무분석 및 인력산정 방법

- 직무분석은 디지털 장기보존 조직의 업무 특성을 파악하고 적정인력 규모를 산출하기 위해 ① 사전조사 ② 직무조사 및 설계 ③ 직무분석의 단계를 거쳐 추진한다.

[표 4-12] 직무분석의 절차

[1단계] 사전조사	(1) 기초현황자료 수집 (2) 직무분류체계 작성
[2단계] 직무조사	(3) 직무조사표 설계 (4) 직무조사 실시
[3단계] 직무분석	(5) 직무 성격 및 업무량 산정 (6) 적정업무량 추계

• 사전조사

- 구성원 및 직무와 관련한 기본 현황에 대한 자료를 수집하여 직무조사 설계를 위한 기초자료로 활용한다.
- 주요 조사 항목은 구성원의 소속 부서, 이전 근무부서 직위, 직렬, 직급, 연령, 성별, 입직 시기, 현 보직 발령 시기, 최근 1년 간 월 평균 근무 일수, 휴가 일수,

초과 근무시간 등이 포함된다.

- 구성원 개인별 업무분장표 및 세부 업무 현황을 파악하고 기초자료를 토대로 직무 중심의 분류체계를 작성한다. 이때 분류 기준은 업무 프로세스에 기반하여 업무 내용 및 구성원 역할을 설정한다.
- 직무분류체계는 대분류 - 중분류 - 소분류 - 단위 업무로 구성된 수평적 분류기준 및 실무 관련 직무로 구성된 수직적 분류 기준을 적용한다.
- 업무 목적 및 범위, 자격 조건 등이 명확히 구분되는 직무군의 유형화는 유사한 기능을 수행하는 단위 업무로부터 추출한다.

• 직무조사 설계

- 직무조사표는 실제 업무의 분장에 근거한 단위 업무 중심으로 업무량 및 직무 특성 등을 측정할 수 있도록 설계한다.
- 조사 대상은 관리자 및 실무 담당을 포함한 전 구성원이며, 구체적 지표로서 업무 비중, 업무주기, 업무중요도, 전문성 요구정도, 업무시간 추가투입 필요 정도, 고객지원 업무 유형, 직무교육 필요 수준 등이 포함된다.

[표 4-13] 직무조사 조사항목

구분	주요 질문
업무비중(업무량)	담당 업무 전체 가운데 해당 업무가 차지하는 시간 비중
업무주기	수시, 일, 주, 월, 분기, 반기, 년
중요도	결재권자 혹은 보고 우선순위 등에 따른 업무의 중요도
전문성 요구정도	원활한 업무수행을 위한 숙련 요구기간에 따른 전문성 요구수준
업무시간 추가투입 필요정도	해당 업무를 보다 충실히 수행하기 위한 업무시간의 증가·혹은 감소의 필요 정도
고객 지원업무 유형	해당 업무가 내부고객(타 팀·부서) 혹은 외부고객(외부기관·개인)을 위한 지원업무의 성격을 가졌는지 여부
직무교육 필요정도	기본적인 직무교육 외 추가로 집중적 보수교육이 필요한 정도

• 직무분석

- 유형별 업무량 산정 및 분석: 업무량은 표준 근로업무량, 실제 근로업무량, 추가

투입 업무량 등으로 유형화하고 단위업무 및 부서별 업무 성격에 따라 업무량을 분석한다.

- 업무 성격에 따른 업무량은 업무주기, 중요도, 전문성 요구정도, 직무교육 필요 정도를 기준으로 분석한다([표 4-14] 업무량산정 조사항목 참조).
- 단위 업무(직무) 우선순위 파악을 위해 업무 중요도 및 난이도를 기준으로 한 매트릭스 분석을 수행한다.
- 적정 업무량 추계: 부서의 업무 특성, 사업 유형별 업무 특성에 따라 적정 인력을 추계하기 위해 각 단위 업무별로 산출된 업무량을 종속변수로, 직급·직군, 업무 체계·난이도·중요도 등을 독립변수로 하는 위계선형모형을 구성하여 추정한다.

[표 4-14] 업무량산정 조사항목

구분	정의 및 산식	
표준근로 업무량	정의	1년 동안 법정 공휴일을 제외한 전일제 총 근로일수에서 초과근무를 하지 않는다고 가정할 때의 총 근무시간
	산식	$(52\text{주} \times 7\text{일} \times 8\text{시간}) - (\text{법정 공휴일수} \times 8\text{시간}) - (\text{평균 연차일수} \times 8\text{시간})$
실제근로 업무량	정의	1년 동안 근로자가 업무를 수행하는 데 실제로 투입한 총 근무시간
	산식	$(\text{실제 월평균 근무일수} \times 12\text{개월} \times 8\text{시간}) - (\text{법정 공휴일수} \times 8\text{시간}) - (\text{월평균 휴가일수} \times 8\text{시간}) + (\text{월평균 초과근무시간} \times 8\text{시간})$
추가투입 업무량	정의	담당업무를 보다 충실히 수행하기 위해 투입시간의 증가가 필요한 정도를 반영한 총 근무시간
	산식	$(\text{실제근로업무량}) \times (\text{현재 대비 업무시간 추가 필요 정도 비율})$

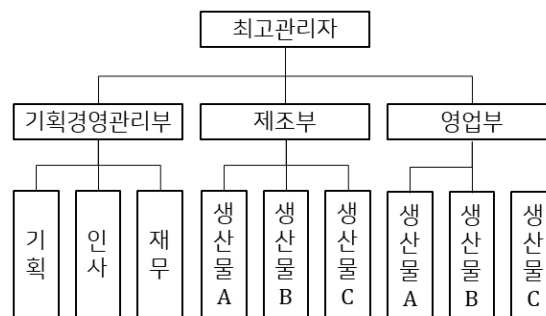
• 적정인력의 산정

- 적정 센터 및 센터 하위 조직단위별 적정 인력의 산정은 온라인 설문조사를 통한 직무분석 결과, 내·외부 구성원 대상 심층면접 및 전문가 자문 결과 등을 활용하여 단위 업무 당 투입량을 산출하고, 1단위 업무량 대비 연간 투입 및 업무 실적 정보를 활용하여 인력을 산출한다. 미래 기능에 해당하는 직무의 경우 현재 시점에서 업무량을 조사할 수 없으므로 전문가 자문 및 사례분석 결과를 활용하여 도출한다.

□ 기능구조와 사업구조

1) 기능구조(Functional Structure)

- 전통적 부서화(departmentalization) 원리에 따라 형성된 계서제적 구조로서 본질적으로 수직적이고 조직구성원들을 전문화하고 조직의 기능과 연계된 직무분야별로 집단화 및 협업하도록 하는 구조이다(오석홍 2018).
- 장점
 - 전문분야별로 집단화된 인적 자원을 다양한 사업의 집행에 유연하게 활용 가능하다.
 - 여러 사업의 수행에 있어 기능상 기준, 절차, 규칙 등의 일관성을 유지할 수 있다.
- 단점
 - 특정 사업보다 기능적 단위를 더 중요시하여 기능 간 갈등이 조장되기 쉬우므로 사업의 통합적 추진력이 약해질 수 있다.
 - 하나의 사업에 대한 전체적 책임 소재가 불명확하다.
 - 기능별로 구성원들의 성향이 달라 조정의 어려움이 있다.
 - 고객에 대한 신속한 대응이 어렵고, 강력한 수직적 계서제가 변동에 저항하도록 하여 창의성을 발휘하도록 할 유인체계가 부족하다.



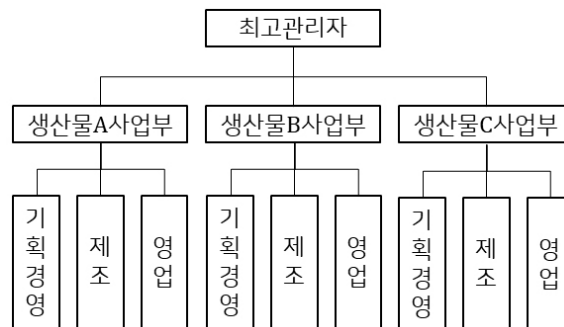
[그림 4-10] 기능구조 예시

2) 사업구조(Project Structure)

- 특정 사업의 수행에 요구되는 권한을 중관관리계층인 사업관리자에 위임하는 구조

로서 본질적으로 수평적이고 사업수행 참여자들은 사업관리자에게 직접 보고하므로 평평한(flat) 형태의 구조를 형성한다.

- 맡은 사업에 대한 개선적 권한 행사로 사업을 효율적으로 수행 가능하다는 장점을 지닌다.
- 복수의 사업을 수행하는 조직이 사업구조에만 의존하게 되면 막대한 비용이 소요(사업별로 전문 직원 및 시설이 중복 배치)되며 업무중복성의 문제가 발생하고, 전문가들의 기능별 집단화가 이루어지지 않아 조직의 기능별 역량을 발전시키기 어렵다는 단점이 있다.



[그림 4-11] 사업구조 예시

□ 조직 사례

- 일반적으로 ‘센터’ 조직 단위의 설치 기준이 모호하며 조직에 따라 활용의 예가 다양하게 나타난다. 이중 지식정보를 공공서비스 자원으로 제공하는 조직의 몇 가지 사례를 조직구조상 위치, 주요업무, 인력을 기준으로 제시하면 다음과 같다. 사례는 각 기관의 공식 홈페이지에 제시된 정보를 근거로 정리하였다.

1) 국립중앙도서관 자료보존연구센터

- 국립중앙도서관의 자료보존연구센터는 3부(기획연수부, 자료관리부, 디지털자료운영부), 3관(국립어린이청소년도서관, 국립장애인도서관, 국립세종도서관)과 더불어 독립

된 조직단위로 설치되어 있다(국립중앙도서관은 3부 3관 1센터 18과 1팀으로 조직).

- 이 중 소속기관인 자료보존연구센터의 주요 업무는 도서관자료의 과학적 보존 및 복원과 보존기술연구개발 등이며 구체적으로 도서관 발전을 위한 조사 및 연구, 도서관 관련 국내외 각종 실태조사, 도서관 관련 조사 및 연구의 협력 체계 구축 및 운영, 도서관 관련 조사 및 연구 자료의 발간과 배포, 도서관자료(기록매체 포함)의 보존·복원 연구, 도서관자료의 보존·복원처리 및 관련 시설 운영, 도서관 소장 원본자료 보존 및 디지털도서관 보존서고 관리, 도서관자료 보존 및 복원 지원, 도서관자료의 보존·복원 관련 표준화 업무 등이 포함된다(문화체육관광부와 그 소속기관 직제 시행규칙 제32조 2항).
- 자료보존연구센터에는 센터장 1명을 두도록 되어있으며 센터장은 4급 서기관 또는 학예연구관으로 보한다(문화체육관광부와 그 소속기관 직제 제6장 제48조: 문화체육관광부와 그 소속기관 직제 시행규칙 제32조 1항).
- 자료보존연구센터의 인력은 센터장을 포함하여 12명으로 구성되어 있다.



[그림 4-12] 국립중앙도서관 조직도와 센터조직

2) 국가기록원 기록보존복원센터 및 기록관리교육센터

- 국가기록원의 경우 상위 조직단위인 부 내에 과 및 센터를 설치하고 있으며 기록보존복원센터 및 기록관리교육센터 등 2개의 센터가 존재한다.
- 기록보존복원센터의 주요업무는 총 16가지로 분류되며 열거하면 다음과 같다. ① 기록물의 보존 정책 기획 및 보존 업무의 총괄·조정 ② 기록물 보존시설의 기반구축 계획 수립·조정 및 지원 ③ 원내 보존시설의 구축 및 보존분담 총괄·조정 ④ 기록물관리기관의 보안 및 재난대비 기준 마련 및 점검 관리 ⑤ 기록물관리기관의 보존시설·장비·환경 기준의 수립 및 점검 관리 ⑥ 기록물 서고배치 및 기록물 반·출입 관리의 총괄·조정 ⑦ 기록물 정수 점검 및 통계 관리의 총괄·조정 ⑧ 기록물 상태검사·점검 및 보존환경 기준 수립과 관리 ⑨ 기록매체 보존규격의 개발·관리 및 보존용품 인증제도의 운영 ⑩ 기록물 보존처리·매체수록의 기준 수립·관리 및 시청각실·복원실 운영 ⑪ 복원·복제 및 탈산·소독의 기준 관리 ⑫ 기록물의 유형별 보존복원기술·장비 연구 개발, 보존시설 대여 및 기술 지원의 총괄·조정 ⑬ 수복용 한지 등 기록물 복원·복제 재료 및 기술 연구 ⑭ 연구 개발된 기술의 시험·적용 및 교육·보급 ⑮ 재난피해 기록물의 응급복구 및 복원 지원 ⑯ 국내외 보존 복원 관련 교류 협력 및 지원
- 기록보존복원센터의 인력은 센터장을 중심으로 19명으로 구성된다(사무관, 연구관, 전문임기제 다급 및 라급 및 주무관 포함).
- 기록관리교육센터의 주요업무는 총 6가지로 분류되며 다음과 같다. ① 기록물관리 교육훈련 기본계획의 수립·추진 ② 기록물관리 맞춤형 교육과정의 개설·운영 ③ 기록물관리 교과목·교육기법의 개발 및 교육교재의 발간 ④ 사이버 기록물관리 교육 콘텐츠의 개발 및 교육운영 ⑤ 국내외 기록물관리 관련 교육훈련기관 간 협력·지원 ⑥ 그 밖에 기록물관리 교육에 관한 사항
- 기록관리교육센터의 인력은 센터장을 중심으로 10명으로 구성된다(사무관, 연구관 및 주무관 포함).



[그림 4-13] 국가기록원 조직도와 센터조직

3) 한국과학기술정보연구원

- 한국과학기술정보연구원은 국가 과학기술정보 분야의 전문연구기관으로 ① 과학·기술 및 이와 관련된 산업정보를 종합적으로 수집·분석·서비스하고 ② 정보 관리 및 유통 관련 기술·정책·표준화 등에 대한 전문적 조사와 연구를 수행하며 ③ 첨단 정보 및 연구개발 인프라를 체계적으로 구축하고 운영한다.

- 한국과학기술정보연구원의 경우 상위조직 단위인 '본부' 내 센터를 다수 설치(총 21센터)하고 있다.
- 센터별 인력 수는 각기 다르게 나타나는데 중소기업지원센터가 총 50명으로 가장 많고 연구전략센터가 5명으로 가장 적게 나타나며, 센터 당 평균 인력 수는 19.33명이다.



[그림 4-14] 한국과학기술정보연구원 조직도와 센터조직

4) 국립의과학지식센터

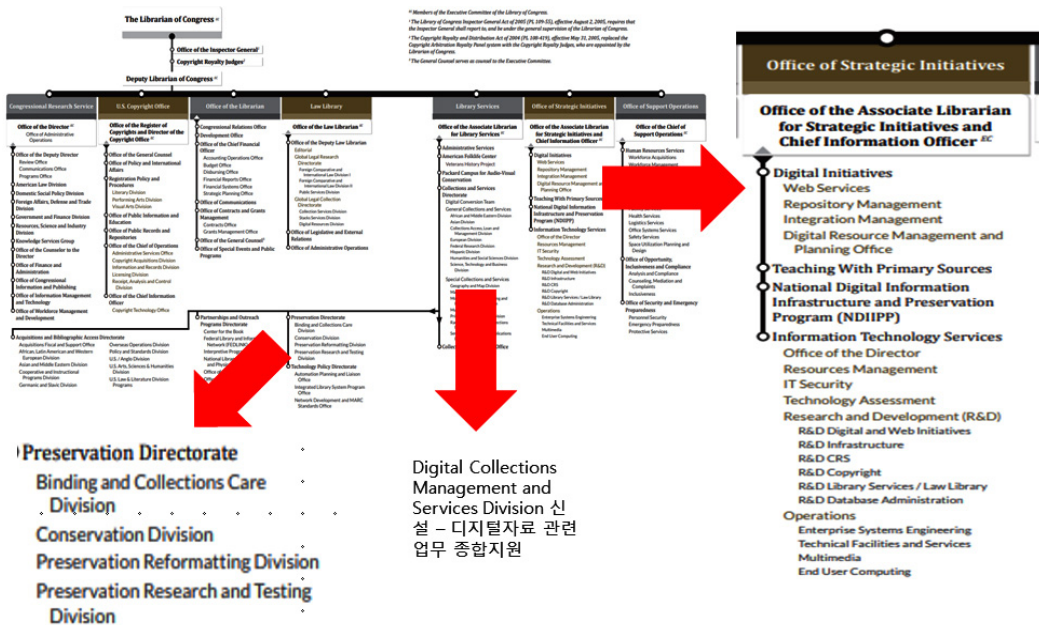
- 국립의과학지식센터는 질병관리본부 산하의 독립된 센터로서 ① 국내·외 의과학 정보자원의 망라적 수집·가공·보존 및 제공 ② 의과학 연구성과물 통합관리 및 제공 ③ 의과학 연구 데이터의 수집·보존·공유 인프라 구축 ④ 의과학 최신정보 및 교육 프로그램 제공 등의 주요 기능을 수행한다.

- 의과학지식센터 업무를 총괄하고 센터 내 핵심기능을 수행하는 의과학지식관리과 내에 12명의 인력이 배치되어 있으며 그 외에도 종합자료실 및 멀티미디어자료실에 각각 3명 및 1명이 배치되어 있다.

□ 도서관 장기보존관련 조직 사례

1) 기능별로 구성된 사례

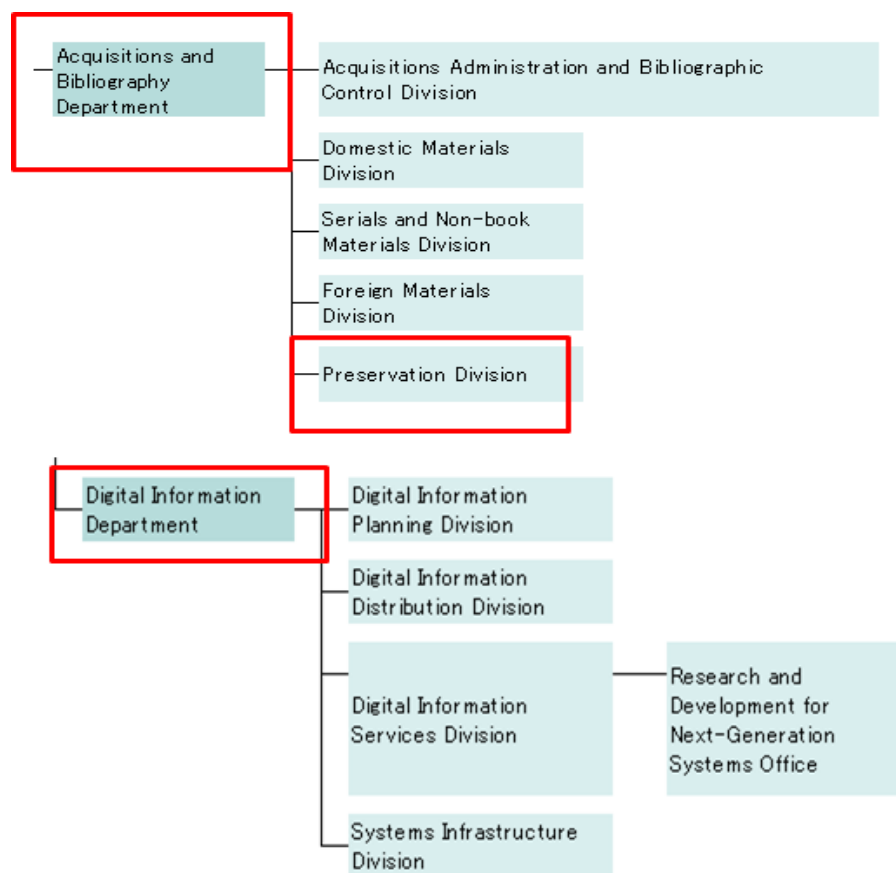
- 미국 의회도서관 사례를 살펴보면 기획, 컬렉션관리 및 서비스, 보존 등 기능별로 부서가 구성되어 있으며 각 기능별로 구성된 부서하에 디지털 담당 업무와 조직이 배치되어 있다.
- 기획부서는 Office of the Associate Librarians for Strategic Initiative and Chief Information Officer이며 디지털자료는 Digital Collection Management and Services Division이 담당하고 있고 보존 전체는 Preservation Directorate가 담당하고 있다.



[그림 4-15] 미국 의회도서관 부서 조직도

2) 디지털자료 유형별 구성된 사례

- 일본 의회도서관 사례를 살펴보면 디지털자료를 중심으로 기획, 연구, 수집, 서비스, 시스템까지 한 부서로 구성되어 자원유형별로 조직을 구성 한 것으로 조사되었다.
- 보존부서는 수집, 서지 부서의 하위 부서로 배치되어 있어 인쇄자료, 아날로그 자료의 보존에 초점이 맞춰져 있다.



[그림 4-16] 일본 의회도서관 조직도

3) 센터조직 사례

- 프랑스 국가도서관은 보존 업무를 담당하는 조직을 센터형태로 운영하고 있다. 현

재 보존관련 설치된 센터는 Bussy-Saint-Georges technical center와 Joël-le-Theule Technical Preservation Center가 있다.

- Bussy-Saint-Georges technical center는 1995년에 복원 관련 워크숍 운영과 보존서고 기능을 목적으로 독립된 부서로서 설립되었고 Joël-le-Theule Technical Preservation Center는 탈산화를 통해 인쇄자료를 보존하는 목적으로 2016년에 설립되었으나 최근 디지털자료 관련 워크숍과 교육활동도 하고 있는 보존관련 센터이다.
- Bussy-Saint-Georges technical center 주요 기능
 - 인쇄자료 보존, 디지털자료 보존 관련 연구, 교육, 디지털 컬렉션 등 담당
 - 인쇄자료 보존관련 워크숍
 - 디지털화 워크숍
 - 시청각자료 워크숍
 - 방재처리
 - 연구실 운영
 - 보존관련 교육
 - 국가도서관(미테랑, 리슐리외) 보존서고
 - 디지털 보존 컬렉션 프로젝트 진행 및 담당
- Joël-le-Theule Technical Preservation Center 주요 기능
 - 산성지로 인한 보존위험성에 처한 인쇄물 재생산 및 보존 중심
 - 디지털화
 - 마이크로필름화
 - 포스터 및 기타 자료 보존 등

4.3.2 국립중앙도서관 디지털자원 보존 조직 구성(안)

- 디지털 장기보존 업무의 특성상 수 개(연구, 기획, 운영, 선별 및 디지털화, 디지털 자원환경보존 등)로 분할되어야 한다는 점에서 그에 준하는 조직구성을 고려할 필요가 있다.

- 디지털 장기보존 업무는 자료유형 및 기능상으로 업무의 한계와 구분이 분명하고, 장기보존은 현 국립중앙도서관의 업무 프로세스 중 단일하고 독립적인 단계로서 업무의 독자성과 계속성이 있다.
- 디지털자료의 장기보존은 장기적 연구와 기획이 수반되며 마이그레이션 등 업무의 독자성과 계속성이 가장 핵심적인 사항이므로 ‘국’에 준하는 단위의 조직편성이 바람직하다.
- 다만 국립중앙도서관의 조직체계를 중심으로 볼 때 하부조직인 ‘부’가 중앙행정기관의 ‘국’에 상응하는 위상을 가지고 있다. 따라서 디지털 장기보존 조직의 규모는 국립중앙도서관 내 ‘부’의 규모에 준하여 설정한다.
- 현재 국립중앙도서관 본관의 ‘부’ 및 ‘과’의 인력 현황은 다음 표와 같다.

[표 4-15] 국립중앙도서관 ‘부’ 및 ‘과’ 소속 인력

부서	과	인원(명)
기획연수부	총무과	44
	기획총괄과	11
	사서교육문화과	10
	국제교류홍보팀	10
	소계	75
자료관리부	자료수집과	27
	국가서지과	37
	자료운영과	31
	고문헌과	9
	소계	104
디지털자료운영부	디지털기획과	18
	디지털정보이용과	13
	정보시스템운영과	18
	소계	49

- 부의 평균 인력은 76명이며 과의 평균 인력 수는 20명이다. 도서관 전문 영역이 아닌 일반 운영업무 중심인 총무과가 가장 많은 인력 수를 가지고 있으나 이는 도서관 기능 단위 조직이라고 해석하기 어려우므로 수집, 조직, 운영, 디지털 등 주요

기능 및 자료유형으로 구성된 조직 부서를 중심으로 판단해야 할 것이다.

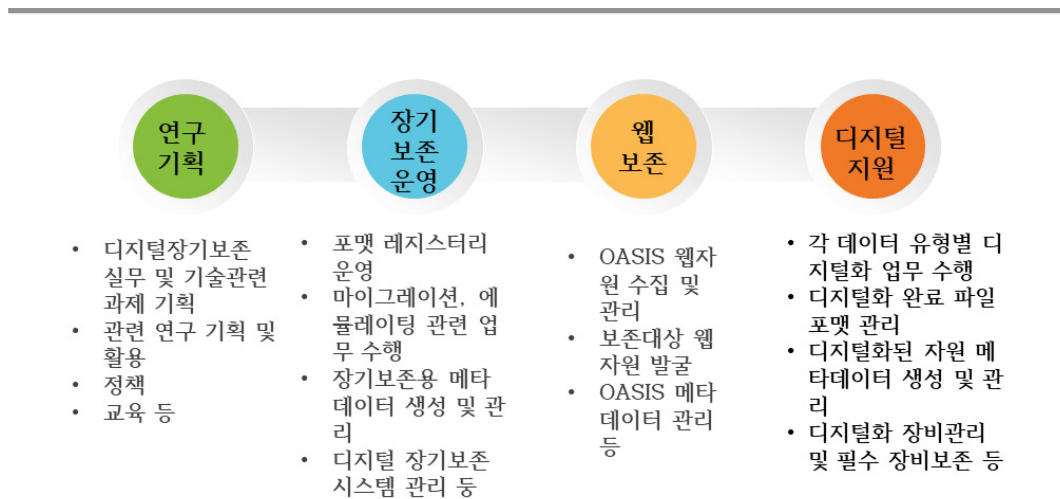
- 따라서 수집, 조직, 운영 등 주요 기능단위로 구성된 ‘과’에 해당하는 인력을 중심으로 분석한 결과 기능단위 ‘과’에 해당하는 인력은 평균 32명에 해당하였고 디지털로 자료유형이 조직구성의 상위분류로 적용된 후에 하위 조직단위로 세부 기능이 구성된 디지털자료운영부의 사례에서는 각 ‘과’의 인력은 평균 16명에 해당하였다.
- 디지털 장기보존의 경우 디지털로 자료 유형 범위가 설정된 후 하위 기능단위로 구성되는 사례에 해당하므로 디지털자료운영부의 사례에 준하여 조직을 구성하는 것이 타당할 것이다.

□ 기록보존 업무 영역과 프로세스

- 디지털자료 장기보존과 관련된 핵심 영역은 연구기획, 장기보존운영, 웹보존, 디지털 지원 등 총 4개영역으로 도출되었다.
- 이 영역들은 현재 국립중앙도서관에서 수행하고 있는 디지털자료관련 업무와 2장에서 설명한 Brown(2013)의 기록보존 업무 프로세스 10개에 해당하는 영역과 비교 분석하여 도출한 결과이다([표 4-16], [그림 4-17] 참조).

[표 4-16] Brown 보존 업무 프로세스와 국립중앙도서관 장기보존 핵심영역

국립중앙도서관 장기보존 핵심영역	Brown 보존 업무 프로세스	국립중앙도서관 장기보존 핵심영역	Brown 보존 업무 프로세스
연구기획	이해관계자의 참여	웹보존	획득 및 수집
	법적 근거 확보		비트스트림 보존
	정책 프레임워크		논리적 보존
장기보존운영	획득 및 수집		메타데이터 관리
	비트스트림 보존	디지털지원	유통
	논리적 보존		획득 및 수집
	메타데이터 관리		비트스트림 보존
			논리적 보존
			메타데이터 관리
			기반시설 구축 및 관리



[그림 4-17] 디지털자료 장기보존과 관련된 영역과 주요 업무

□ 디지털 장기보존 핵심영역과 주요 업무

1) 연구기획

- 인력규모: 약 10-15명
- 기존 디지털기획과 18명, 사서교육문화과 10명 규모와 비교하여 디지털 장기보존 전략기획과 교육을 전담하는 조직규모를 국립중앙도서관 기존 인력 수준을 기준으로 하되 기획과 교육을 2:1의 비율로 구성
- 디지털자원 관련 기획 영역
 - 디지털자원 수집, 큐레이션, 활성화 등에 관련된 기획업무
 - 국립중앙도서관의 소장자료를 포함하여 사회의 변화와 주요 이슈에 따라 국가도서관으로서 수집해야하는 민간 디지털자료를 조사 분석하여 국립중앙도서관의 컬렉션으로 개발할 수 있는 기획 등
 - 국립중앙도서관에 소장되어 있는 디지털자원을 활용하여 지식재생산으로 환류될 수 있는 디지털자원을 큐레이션하여 서비스 하는 등 디지털자원 활성화에 관련된 동향 분석 및 기획 등
 - 국가적 가치가 있는 디지털자원 발굴 및 생산(디지털화) 기획
 - 국가도서관으로서 보존해야할 민간 자원을 발굴하여 디지털화하여 새로운 디지

털 콘텐츠로 재생산함으로써 국가지식정보자원으로 개발하는 단위과제 기획 및 수행 등

- 디지털 장기보존 협력체제 구축

- 디지털자원 장기보존 체제가 구축되어 있지 않은 공공도서관을 위해 디지털 장기보존 협력 체제를 구축하여 전국 공공도서관에 소장되어 있는 국가적 가치가 있는 자원을 발굴하여 국립중앙도서관의 디지털자원으로 확보하고 장기보존 대상으로 포함

- 디지털자원 장기보존 정책수립

- 디지털자원 환경변화에 따른 장기보존 동향을 파악하고 디지털 장기보존 정책을 수립 후에도 환경변화에 신속하게 대응할 수 있도록 정책 수정 및 보완 등을 포함

- 온라인 자료 납본체제, 웹 아카이빙 관련 등 장기보존 관련 연구 과제 기획 및 관리

- 국가대표도서관으로서 디지털자료 장기보존 환경변화에 대응하기 위한 동향 파악 및 기초 자료 제공 등을 위해 장기보존 관련 연구 과제를 기획하여 연구 결과를 도서관계에 공유

- 디지털자료 장기보존 방식 및 체제 관련 연구(마이그레이션, 에뮬레이터, 포맷 레지스트리 등) 및 연구결과 보급

- 디지털자료를 장기보존하기 위하여 마이그레이션 및 에뮬레이터, 포맷 레지스트리 등에 대하여 개념 및 이론은 소개되어 왔으나 국가도서관과 같이 장기보존의 책무를 가지고 있는 기관에서 대용량 데이터 환경에 적용하여 수행한 사례는 많지 않다.
- 따라서 국립중앙도서관은 디지털 장기보존에 적절한 장기보존 체제 이론을 실제 디지털자료 환경에 적용하여 결과 활용 과정까지 실제 적용해보는 실험적 연구를 수행하여 국내 도서관계에서 활용할 수 있는 디지털자료 장기보존 방식을 도출해내는 연구를 기획하여 수행한 후 그 연구결과를 도서관계에 보급하는 역할을 수행

- 국립중앙도서관 디지털 장기보존 관련 중장기 기획

- 디지털 장기보존 관련 도서관 교육

-
- 디지털자료 장기보존은 보존을 위해 아날로그 자료를 디지털화한 후속 단계이자 디지털자료를 수집하여 저장, 축적 이후의 단계로 고려되었던 사항임.
 - 도서관자료의 디지털화와 디지털자료 수집이 20여년이 지난 현 시점에서는 도서관에서 해결해야 할 현행 과제로 부각됨
 - 도서관 실무현장에서는 아날로그 자료를 수장고에 보존하는 개념은 수용되고 있으나 디지털 장기보존에 대한 이해도는 낮은 편임
 - 이에 국립중앙도서관 디지털 장기보존 부서에서 기획하여 수행한 연구결과를 바탕으로 도서관에서 수행해야할 디지털 장기보존에 대하여 사서재교육 및 세미나 등 다양한 교육프로그램을 통해 확산시켜야 함
 - 기관 내 디지털 장기보존에 대한 이해도를 높이는 목적을 가지고 교육의 대상에 국립중앙도서관의 기존 실무부서 인력도 포함시켜야 함

2) 장기보존운영

- 인력규모: 약 10-15명
- 기존 디지털자료운영부 하위 조직인 ‘과’의 평균 인력 수준인 16명에 해당하는 인력 규모로 대략적으로 산정
- 장기보존 자료 수집 및 관리
 - KOLIS II 온라인 시스템에 등록된 디지털 파일 연계 수집
 - 온라인납본, 디지털화, 구입, 기증 등으로 수집된 파일 단위의 디지털자료 수집
- 장기보존용 메타데이터 생성 및 관리
 - KOLIS II 에서 연계된 메타데이터에 장기보존을 위한 관리용 메타데이터 추가 및 관리
- 장기보존용 디지털자원 포맷 레지스트리 시스템 구축 및 운영
 - KOLIS II 온라인 시스템을 통하여 저장된 디지털자원의 포맷 레지스트리 시스템 구축 및 운영
- 디지털 포맷 마이그레이션 수행 및 관리

-
- 장기보존을 위한 표준 포맷의 변화에 따른 마이그레이션 전략 수립 및 수행
 - 디지털자원 활용을 위한 에뮬레이터 관리
 - 장기보존을 위한 emulating 체제 구축 및 관리
 - 디지털 장기보존 시스템 관리
 - Ingest, packaging 모듈 구축 및 관리
 - 장기보존용 포맷 전환 및 관리

3) 웹보존

- 인력규모: 약 3-5명
- 기존 디지털자료운영부 하위 조직인 ‘과’의 평균 인력인 16명 수준에는 크게 못 미치나 웹 아카이브 업데이트가 연간단위로 이루어지며 자동화업무 의존도가 높은 직무의 특성을 반영하여, 디지털자료운영부 ‘과’ 평균 인력의 1/3 수준으로 산정
- 웹 아카이브(OASIS) 운영
 - 웹 보존을 위한 웹 아카이브 운영
 - OASIS 홈페이지 운영
 - 포괄적 수집, 선택적 수집 방식으로 수집되어 처리된 웹자원 관리
 - KOLIS II 온라인 시스템과 연계된 OASIS 모듈 관리
 - 장기보존시스템 내 OASIS 자료 관리
- 웹 아카이브 수집 대상 선정 및 웹 보존대상 발굴
 - 포괄적 수집, 선택적 수집 대상 선정
- 웹 아카이브 메타데이터 관리
 - 자동생성된 웹 아카이브 메타데이터 관리
 - 장기보존용 메타데이터 자동변환, 생산 및 관리
- OASIS 파일 뷰어 관리
 - OASIS 파일 뷰어 관리 및 서비스 관리

4) 디지털지원

- 인력규모: 약 10-15명
- 기존 디지털자료운영부 하위 '과'의 평균 인력 수준인 16명에 해당하는 규모로 대략적으로 산정
- 디지털화를 위한 컬렉션 선정
 - 디지털화를 위한 계획 수립
 - 디지털화 대상 선정
- 디지털화 프로세스 관리 및 수행
 - 단위과제별, 데이터 유형별 디지털화 자료 수집, 디지털화 프로세스 수행
 - 디지털화 외주 업무 관리
- 디지털화 원본 및 디지털 파일 관리
 - 디지털화 대상 자료 및 디지털화한 파일 관리
 - 디지털화 자원 검수 및 보정 수행
- 디지털화 파일 포맷 관리
 - 디지털화 포맷 선정 및 장기보존용 포맷 레지스트리 등록 및 관리
- 디지털자료 메타데이터 생성 및 관리
 - 디지털화한 파일 메타데이터 생성 및 관리용 메타데이터 생성 및 관리
- 디지털화 장비관리 및 필수 디지털화 장비 보존(아날로그 자료를 재생할 수 있는 최소장비 보존)
 - 핵심 자료 디지털화 환경 구축 및 장비 관리
 - 아날로그 자료를 재생할 수 있는 최소장비 보존

4.4 종합 요약 및 시사점

□ 디지털 보존 정책

- 디지털 보존 정책의 사례를 조사, 분석한 후 도출한 시사점은 아래와 같다.
- 최근에 발표한 국가대표도서관의 디지털 보존 정책의 특징은 납본 받은 디지털자료에 대한 장기보존을 위한 정책이라기보다는 국가차원 가치가 있는 디지털자원을 선별 입수하여 소장하여 장기적으로 이용자에게 접근할 수 있는 모든 활동을 포함하는 거시적 정책 및 전략내용이 주를 이룬다는 점이다.
- 이러한 거시적이며 국가차원의 전략을 수립한다는 것은 이미 기관에 납본된 디지털자료에 대한 보존전략 방법과 시스템은 어느 정도 완성되었다는 것을 의미한다. 물론 이러한 보존 정책은 단기적 디지털 보존전략을 근간으로 하고 있다. 최근 2010년대에 들어와서는 단기적 보다는 장기적 보존전략에 그 비중을 더 두고 있고 이러한 디지털 보존이 한 기관에서 이루어질 수 없다는 점을 인식하고 있는 것 같다. 이에 따라 각국의 국내 협력기관은 물론 다른 나라와의 협력도 시도하고 이에 대한 정책이 포함되어 제시되고 있다.
- 따라서 디지털 보존 정책은 내부적 업무와 전략, 외부적 협력에 대한 내용이 포함되어야 할 것이며 기술기반의 단기적 보존전략보다는 국가차원의 비전과 목적달성을 위한 장기적 디지털 보존 개발전략이 수립되어야 할 것이다.
- 또한 디지털 보존 정책에 포함될 수 있는 세부적인 내용은 다음과 같다. 즉 1) 서론에는 ① 범위, ② 내용, ③ 과제/전략, ④ 필수사항, ⑤ 준수해야 할 표준; 2) 실행부분에는 ① 보존 업무 계획, ② 선별과 평가, ③ 입수, ④ 메타데이터, ⑤ 접근과 이용, ⑥ 소장, 복제, 백업, ⑦ 검증, ⑧ 도큐멘테이션; 3) 운영부분에는 ① 권리와 제한 관리, ② 역할과 책임, ③ 지속가능성 계획, ④ 보안관리; 마지막으로 4) 지원부분에는 ① 협력, ② 직원교육 등이 포함될 수 있다.

□ 디지털자원 보존 환경 및 위험

- 디지털 보존을 위한 표준 및 시스템 기능 요건을 조사, 분석한 결과 디지털자원의 보존을 위한 기관 활동은 보존 초기 단계인 디지털자료 생산부터 수집 및 평가, 보유 및 검토, 저장, 보존 계획, 보존 활동, 그리고 접근까지 각 단계별로 연계되어 이루어지는 것으로 나타났다. 주요 시사점은 아래와 같다.
 - 디지털자원 보존은 진행 중인 프로세스이므로 디지털자료의 수명주기 전체에 걸쳐 보존 계획을 세워야 한다. 보존 활동은 디지털자료의 권위 있는 성격이 장기간 보호되도록 보장하기 위해 계획되어야 한다.
 - 이러한 작업에는 유효성 검사, 보존 메타데이터 할당, 표현 정보 할당 및 수용 가능한 데이터 구조 또는 파일 형식 확보가 포함된다. 이러한 보존 활동을 통해 디지털 객체가 항상 신뢰성 있고 가용성이 유지되도록 보장한다.
 - 기술의 다양한 변화를 통해 디지털자원을 보존하는 것이 실현 가능하지 않고 바람직하지도 않은 경우, 획득을 재평가해야 할 수도 있다. 전통적인 보존 환경과 동일한 원칙이 유용하게 적용될 수 있지만 디지털 환경에 맞게 정책과 절차를 조정해야 한다.
- 디지털 보존은 단순히 위험에 관한 것이 아니라 기회를 창출하고 디지털자료를 보호함으로써 새롭거나 확장된 가치가 파생 될 수 있음을 의미한다. 조직이나 기술의 변경을 포함하여 디지털자원과 관련한 많은 위험은 디지털 보존 활동 및 광범위한 업무 기능 수행 혹은 법규를 준수하는 능력에 중대한 영향을 미치므로 위험관리가 필요하다. 디지털자원이 직면하는 위험에는 다음과 같은 내용이 포함된다.
 - 조직 간의 기능 합병, 폐쇄 또는 이전
 - 전략 방향 또는 자금 지원 및 조직에서 지원하는 기능의 변화
 - 디지털 보존 전문가 및 실무자의 주요 변화
 - 미래 보존 필요성에 대해 고려하지 않은 아웃소싱
 - 파일 포맷 노후화(obsolescence)로 인한 데이터 변환비용 증가
 - 미디어 노후화로 인한 복구의 불가능
 - 데이터 손상 혹은 변경으로 인한 미디어 성능저하
 - 문맥 정보 손실에 따른 의미 인식의 어려움

-
- 데이터 검색의 어려움
 - 저작권 또는 기타 법적 정보의 지속적 변경에 따른 권리 및 의무에 대한 불확실성 초래
 - 출처 정보나 문서 인증정보의 손실에 따른 진본성 훼손
 - 버전 제어 실패로 인한 문서의 신뢰성 훼손
 - 의도하지 않은 삭제로 인한 데이터 변경 또는 손실
 - 사용 정도(수준)에 따른 손실
 - 자연 재해
- 디지털 콘텐츠에 대한 위험은 다음과 같은 실무적 문제점을 야기한다.
 - 평판의 훼손
 - 중요한 업무에 부적절한 자원을 활용하게 됨
 - 이용자에게 적합한 콘텐츠를 지원할 수 없음
 - 법적 이슈 또는 규제에 대한 적절한 대응의 어려움
 - 데이터를 활용하거나 재사용할 수 없음
 - 정체성 및 이력정보(memory)의 손실
 - 재생산 및 복구를 위한 고비용

□ 디지털 장기보존 조직

- 디지털 장기보존의 업무 특성 상 분담하여 수행할 필요성이 있고, 디지털 장기보존이라는 업무의 한계가 명확하며, 자료유형 및 기능의 특정성과 현 국립중앙도서관 업무 프로세스 상 하나의 독립적 단계라는 업무의 독자성 및 계속성이 존재한다. 특히 디지털자료의 보존 업무는 장기적 연구와 기획이 동반되며 마이그레이션 등 업무 독자성과 지속성이 가장 핵심적이므로 디지털 장기보존 조직의 규모는 중앙행정기관의 ‘국’에 준하여 설정하는 것이 타당하다고 판단된다.
- 국립중앙도서관 업무 중 디지털 장기보존에 관련 업무가 연구 및 기획, 장기보존운영, 웹보존운영, 디지털지원 등 여러 개로 분할(divisionalization)되어야 하므로 국에 준하는 전담 조직의 위계 하에 과에 준하는 보조(계선)조직을 설치할 필요가 있다.

- 조직의 장은 업무 특성에 적합한 직위 및 권한을 가져야 한다. 행정기관의 조직과 정원에 관한 통칙에 근거할 때, 국에 준하는 디지털 장기보존 조직이 설치될 경우 그 장은 고위공무원단에 속하는 ‘나’ 등급 일반직공무원을 보임하고, 과에 준하는 하위조직의 장은 3급 혹은 4급 공무원으로 보임하는 것이 적절하다고 판단된다.
- 다만 국립중앙도서관의 경우 하부조직인 기획연수부, 자료관리부, 디지털자료운영부 등 ‘부’ 조직이 중앙행정기관의 ‘국’에 준하는 기능을 수행하며 부의 장을 고위공무원단에 속하는 일반공무원으로 보하고 있으므로(문화체육관광부와 그 소속기관 직제 제6장) 디지털 장기보존 조직의 규모는 국립중앙도서관 내 ‘부’의 위상에 준하여 설정해야 할 것이다.
- 문화체육관광부와 그 소속기관 직제 및 시행규칙에 따르면 디지털 장기보존 조직이 설립될 국립중앙도서관의 기 설립된 센터인 자료보존연구센터에는 센터장 1명을 두도록 되어있으며 센터장은 4급 서기관 또는 학예연구관으로 보하도록 되어 있다. 한편 국립중앙도서관의 하부조직 부 단위 조직(자료관리부, 디지털자료운영부 등)의 장인 부장은 고위공무원단에 속하는 ‘나’ 등급 일반직공무원으로 보하도록 되어있다.
- 새로 설립될 디지털 장기보존 조직의 장은 업무 특수성 및 중요성을 고려하고 추진력을 강화할 수 있도록 국립중앙도서관의 부장급에 준하는 고위공무원단 소속 일반직공무원으로 보하고, 하위부서(과)의 장을 중앙행정기관의 과장에 준하는 3급 혹은 4급 공무원으로 보하도록 하여 권한과 조직에서의 위상을 확립할 필요성이 있다.
- 디지털 장기보존 조직의 구성은 도서관 업무 전체를 프로세스 단계별로 구분하여 기능구조화 한 뒤 이를 아날로그와 디지털이라는 자료 유형별로 분할하는 방법과, 자원 유형별로 사업부화하여 조직을 분할하고 하위 조직구조를 도서관 업무 프로세스 기능을 반영하여 구성하는 방법이 모두 가능하다.
- 디지털자료 유형의 특수성 및 장기보존 업무의 특수성에 초점을 둔다면 디지털 장기보존 업무 수행의 계선적 권한 및 책임을 디지털 장기보존 조직의 장에게 두고 해당 사업 수행의 효율성을 증가시키는 방안이 효과적일 것으로 판단된다.
- 다만 기존 부서 업무와 인력 및 시설 등의 중복성 문제가 최소화되도록 조직 설계

단위에서 조직 차원의 논의 및 합의가 필요하며, 이에 대한 추가적인 직무분석이 요구된다. 또한 디지털 장기보존 조직의 주요 영역을 기능별 업무전문성을 중심으로 적절히 분할하여 조직의 기능별 역량을 보완할 필요가 있다.

- 센터 조직의 활용 사례를 조직구조상 위치, 주요업무, 인력을 기준으로 검토한 결과, 기능의 유사성을 가진 조직임에도 불구하고 조직에 따라 센터 조직의 설치 기준이 모호하고 및 활용예가 다양하다. 인력의 경우 센터에서 수행하는 주요 업무의 범위에 따라 다르지만 대략 20명 내외의 인원으로 구성되는 것으로 나타났다.
- 센터 이름을 사용하고 있는 조직의 사례는 다양하나 주요 업무를 고려하여 20명 내외의 수로 인원을 구성할 필요가 있다. 정확한 구성원 규모를 제시하기 위해서는 추가로 구체적인 직무분석 및 업무량 산정이 필요하다.
- 디지털 장기보존 조직을 구체적으로 구성하기 위해서는 디지털 장기보존 조직 직무를 체계적으로 세분화하여 조사하고 분석하는 과정이 반드시 선행되어야 되므로 이에 대하여 향후 전략과제로 수행해야 한다.

제 5 장

결론 및 후속과제

5.1 종합 요약

5.2 후속과제 영역

제 5 장

결론 및 후속과제

5.1 종합 요약

□ 디지털 보존의 필요성

- 디지털자원의 중요성에 대한 인식은 점점 높아지고 있지만 디지털 기술의 급속한 발달로 인하여 디지털자원 이용환경이 급격히 변화하고 있고 이에 디지털자원을 빠르게 노화되고 활용이 어려워지는 문제점이 발생하고 있다.
- 이에 디지털자원의 환경변화와는 상관없이 디지털자원이 장기적으로 활용될 수 있게 하는 노력인 디지털 보존 업무는 보존 업무의 핵심적인 프로세스로 간주되기 시작하였다. 디지털 보존은 현재의 소프트웨어, 하드웨어, 파일 포맷, 관리시스템 등이 변화하거나 소멸되더라도 미래의 이용자가 디지털자원을 활용하는데 문제가 없도록 보존하는 모든 활동을 말한다(DPC 2015).
- 이러한 디지털 보존에 대한 관심과 투자가 시급한 이유는 지적·문화적 성과들이 디지털 형태로만 생산되어 저장되고 있는데 현 시대에서 활용하는 것은 보장될 수 있지만 장기적 시각에서 보존을 염두에 두어 처리하지 않거나 아무런 조치를 취하지 않고 단순히 백업하는 수준으로 처리할 경우 결국 디지털자원이 접근불가하거나 내용을 열어볼 수 없어 소멸될 수 있기 때문이다.
- 국가차원에서 디지털자원 보존에 대한 이슈가 핵심 문제로 대두되면서 국가의 문화유산과 지적자산을 보존하는 차원으로 공익과 안정성을 보장할 수 있는 국가수

준의 정보기관에서 수행해야 한다는 인식이 보편화되고 있다. 이에 각 국가도서관 별로 디지털자원 보존에 대한 전략이 수립되고 점차적으로 제도화되기 시작하였다.

- 본 연구는 국내에서 발간되는 디지털자원을 후속 세대에 영구적으로 전송하기 위한 기본 틀을 구축하기 위하여 국립중앙도서관이 수용 가능한 디지털자원 보존 정책을 수립하고 담당 조직구성을 위한 시사점을 제시하고 보존 실무를 위한 개선(안)을 제안하였다. 특히 본 연구에서 조사된 디지털 장기보존 정책, 수집, 조직 사례 등은 국립중앙도서관의 디지털 장기보존 정책 및 업무 정의, 조직 구성 시 참조할 수 있는 사례가 될 수 있으며 장기적인 디지털 장기보존 방향 설정에 활용될 수 있다.

□ 각국 국가도서관의 디지털 보존과 관련된 현황 조사

- 국가도서관의 디지털 보존과 관련된 현황 조사에서 나타난 주목할 만한 사항은 디지털 보존 자료의 활용성에 대한 중요도가 높아지고 있다는 것이다. 디지털 장기보존은 단순히 디지털자료 복본을 안전하게 보존하는 체제가 아니라 후세가 디지털 형태로 보존되어 있는 자료를 활용하도록 보존하는 체제를 구축하는 것이다.
- 보존되어 있는 디지털자료를 활용하는 방안에 대하여 각국 국가도서관에서 장기적 계획을 수립하여 포맷 관리와 장기보존 체제를 구축하는 현실적 방안에 대하여 다양한 시도가 시작되는 단계로 파악되었다.

□ 디지털 보존 정책

- 초기의 디지털 보존 정책은 납본 받은 디지털자료에 대한 장기보존을 위한 정책에 국한하여 개발·수립한 반면, 최근에는 국가차원 가치가 있는 디지털자원을 선별 입수하여 소장하여 장기적으로 이용자에게 접근할 수 있는 모든 활동을 포함하는 거시적 정책 및 전략을 중심으로 발전하고 있다.
- 따라서 디지털 보존 정책은 내부적 업무와 전략, 외부적 협력에 대한 내용이 포함되어야 할 것이다. 디지털 보존 정책에는 기술기반의 단기적 보존전략보다는 국가차원의 비전과 목적달성을 위한 장기적 디지털 보존 개발전략 수립이 개발되어야

할 것이다.

- 본 연구에서 수립한 국립중앙도서관 디지털 보존 정책의 세부적인 내용은 다음과 같다. 1) 서론에는 ① 디지털 보존을 위한 비전, ② 디지털 보존 정책의 목적, ③ 디지털 보존 정책의 범위; 2) 실행부분에는 ① 디지털 보존 담당조직, ② 이해관계자 관리, ③ 법적 이슈 관리, ④ 정책 프레임워크 관리, ⑤ 수집, ⑥ 비트스트림 보존, ⑦ 장기 보존기술 및 전략, ⑧ 메타데이터 관리, ⑨ 활용을 위한 접근기능 관리, ⑩ 기반시설 관리; 3) 운영부분에는 ① 담당 부서의 임무, ② 연구과 표준개발, ③ 저작권; 4) 지원부분에는 ① 역할과 책임, ② 교육, ③ 협력; 5) 마지막으로 디지털 보존 정책의 개정이 포함되어 있다.

□ 디지털자료 프로세스와 장기보존 프로세스의 연계방안

- 국립중앙도서관 온라인 자료 업무와 관련하여 장기보존과 관련된 제안 사항은 납본 원문을 서비스용으로 원문저장소에 보장하여 서비스를 하고 현재 보존용으로 생성하는 것을 장기보존시스템으로 전송하여 장기보존 포맷으로 전환한 후 장기보존저장소에 저장하여 장기보존 자료를 통합하여 관리하는 체제이다.
- 디지털화하는 자료 역시 온라인 자료와 동일한 체제로 서비스용은 원문저장소에서 관리하고 장기보존용으로 장기보존 포맷으로 전환한 후 장기보존관리체제에 통합하여 관리한다.
- 웹사이트와 웹자료는 현행에서 수행하고 있는 2개의 프로세스를 반영하여 파일 단위의 웹자료는 납본자료와 유사하게 처리하고 웹사이트는 자동으로 수집하여 메타데이터가 생성되어 원문저장소에 저장이 완료되면 특정 시점을 설정하여 연단위로 추가된 데이터베이스본을 장기보존저장소에 백업형태로 저장하며 웹사이트 구성요소의 포맷을 포맷 레지스트리에 등록하는 체제로 운영하도록 한다.

□ KOLIS II 보존 기능 개선사항

- KOLIS II와 장기보존시스템은 독립적으로 구축하고 서로 연계하도록 설계하여 확

실한 관리 주체에 의해 일관성 있게 운영, 관리하는 것을 목표로 하며 콘텐츠 메타데이터인 MODS와 함께 METS를 이용하여 패키징을 하는 체제를 기반으로 하는 것을 제안한다.

- 웹사이트는 자동으로 수집부터 서비스, 그리고 보존까지 원스톱 워크플로우로 진행하도록 하며 이때, 보존 메타데이터는 MODS DB에서 자동으로 추출, 변환하여 METS를 구축하는 것을 제안한다.

□ 디지털 장기보존을 위한 디지털자원 환경 관리 방안

- 국립중앙도서관의 디지털자원의 장기보존 포맷 관리를 위해 파일 식별 기술 연구에 있어 가장 많이 참조되고 있는 PRONOM의 오픈소스 DROID의 식별 방법과 포맷 시그니처 표현 규칙을 참조하여 포맷 레지스트리를 구현할 것을 제안하며 KOLIS II의 보존 기능, 장기보존시스템과 연계하는 것을 제안한다.
- 또한 마이그레이션, 인캡슐레이션 등 적절한 보존전략을 수립하여 수행하는 과정에서 새로이 생성되는 기술(technical) 메타데이터를 포맷 레지스트리에 등록, 관리하는 체제를 구축해야 한다.
- 디지털 보존을 위한 표준 및 시스템 기능 요건을 조사, 분석한 결과 유효성 검사, 보존 메타데이터 할당, 표현 정보 할당 및 수용 가능한 데이터 구조 또는 파일 형식 확보를 포함하여 개발되어야 하는 것으로 나타났다. 이러한 보존 활동을 통해 디지털 객체가 항상 신뢰성 있고 가용성이 유지되도록 보장되어야 하며 기술의 다양한 변화를 통해 디지털자원을 보존하는 것이 실현 가능하지 않고 바람직하지도 않은 경우, 획득을 재평가해야 할 수도 있다.

□ 디지털 장기보존 조직 구성

- 디지털 장기보존 조직 구성에 관련 사례 및 현황조사 분석결과 국립중앙도서관 업무 중 디지털 장기보존에 관련된 업무가 연구 및 기획, 장기보존운영, 웹보존운영, 디지털지원 등 여러 개로 분할되어 있으므로 이를 통합하여 전담하는 조직체계를

구성해야 하는 것으로 나타났다.

- 디지털 장기보존의 업무 특성 상 업무의 한계가 명확하며, 자료유형 및 기능의 특성과 현 국립중앙도서관 업무 프로세스 상 하나의 독립적 단계라는 업무의 독자성 및 계속성이 존재한다. 따라서 디지털 장기보존 조직의 규모는 중앙행정기관의 ‘국’에 준하여 설정하는 것이 타당하다고 판단된다.
- 국립중앙도서관의 경우 하부조직인 ‘부’ 조직이 중앙행정기관의 ‘국’에 준하는 기능을 수행하며 ‘국’ 수준의 규모로 디지털 장기보존 조직의 규모는 국립중앙도서관 내 ‘부’의 위상에 준하여 설정해야 할 것이다.

□ 기대효과

- 디지털자원의 경우 상업적 이해당사자의 저작권을 보호하고 이용자의 접근을 최대한 활용하는 두 가지 이슈가 충돌하고 있는 상황이지만 국가차원에서 국가에서 생산되고 있는 디지털자료를 이용자가 지속적으로 활용할 수 있도록 하는 것은 국가도서관의 임무라 할 수 있다. 따라서 디지털자원의 장기보존과 이용보장을 위한 모든 활동은 이제 국가도서관의 핵심 업무로 간주되고 있다. 특히 디지털자원의 보존 업무는 인쇄자료의 보존 방식과는 완전히 다르고 그 책무조차 광범위하므로 국가차원의 디지털 보존 업무에 대한 실용적인 대안을 모색하는 것이 시급하며 절실히 필요하다.
- 본 연구의 결과는 국립중앙도서관 디지털자원 보존 정책 수립, 디지털자원 보존 업무 수행을 위한 조직 체계 확보 및 보존 실무의 기초자료로 활용할 수 있으며, 이를 통해 체계적이고 효율적인 디지털자원 보존 업무 추진 기반을 마련하고 향후 국가문헌보존센터 구축을 위한 보조 자료로 활용하여 국가대표도서관으로서의 역할을 강화하는 효과가 기대된다.

5.2 후속과제 영역

- 디지털자원 보존은 한 번에 그리고 단편적으로 이루어지는 업무가 아니다. 즉 여러 부서에서 협동적으로 체계적으로 그리고 장기적으로 이루어지는 업무인 것이다. 따라서 디지털자원 보존에 필요한 종합적인 사항을 제안한 본 연구를 토대로 하여 다음과 같은 우선순위로 후속연구과제가 수행되는 것을 제안한다.

□ 1단계: 지원환경 구축

- 디지털 장기보존 세부정책 영역
 - 디지털 장기보존을 위한 국가유산 자료 선별 및 세부 보존 정책
 - 디지털 장기보존을 위한 로드맵 구축
- 디지털 장기보존 조직 영역
 - 디지털 장기보존을 위한 조직 구성에 필요한 세부 직무분석
 - 디지털 장기보존을 위한 조직 구성에 필요한 세부 인력산정

□ 2단계: 실무적 절차 시스템화

- 디지털 장기보존 메타데이터 영역
 - 디지털 보존 자료의 컴포넌트 유형별 보존 속성 및 렌더링 기술 분석
 - 디지털자원의 장기보존을 위한 포맷 레지스트리 설계 및 구현
 - 장기보존시스템과 KOLIS II 온라인 시스템의 메타데이터 자동 연계 구현
- 디지털 장기보존시스템 영역
 - 장기보존시스템 패키징 모듈화 및 장기보존을 위한 세부 기능 설계
 - 장기보존을 위한 애플레이팅 체제 구축

□ 3단계: 장기보존 안정화

- 디지털 장기보존 컬렉션 서비스 영역
 - 장기보존 컬렉션에 기반한 콘텐츠 재생산
 - 장기보존 컬렉션 서비스 방안
- 장기보존을 위한 마이그레이션 영역
 - 장기보존을 위한 마이그레이션 로드맵과 장기 전략 수립

참고문헌

- 국가기록원 (2008). 디지털 포맷 및 애플리케이션 기술정보은행(DFR) 프로토타입 개발.
- 국가기록원 (2009). 기술정보은행(digital format registry) 기능고도화 및 기술정보데이터 구축.
- 국가기록원 (2013). 행정기관 전자기록물 재현기술 연구 및 프로토타입 개발.
- 국가기록원 (2016). 전자기록 기술정보 기반 장기보존 의사결정 지원체계에 관한 연구.
- 국립중앙도서관 (2012). 주요국의 도서관정보정책에 관한 연구. 국립중앙도서관 최종 보고서.
- 국립중앙도서관 (2017). 2017년도 국립중앙도서관 연보.
- 김명옥, 이상용 (2010). 전자기록물의 장기보존을 위한 기능요소 연구. 한국기록관리학회지, 10(2), 101-126.
- 김명훈, 오명진, 이재홍, 임진희 (2013). 전자기록 장기보존 전략으로서의 에물레이션 사례 분석. 기록학연구, 38, 265-309.
- 김관준 (2015). 디지털 큐레이션 연구동향 분석과 과제. 정보관리학회지, 32(1), 265-295.
- 박옥남 (2012). PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구: 국립중앙도서관 디지털 자료를 중심으로. 한국문헌정보학회지, 46(2), 83-113.
- 박지화, 정영미 (2015). 포맷 레지스트리 운영 사례 분석. 한국도서관정보학회 하계 학술발표회, 135-148.
- 서은경 (2005). 디지털 자원의 보존 메타데이터 요소세트 개발에 관한 연구. 정보관리학회지, 22(3), 233-260.
- 서은경 (2016). 대학도서관에서의 디지털보존 실태분석 연구. 정보관리학회지, 33(1), 269-292.
- 서은경 (2017). 전문도서관의 디지털 큐레이션 프로세스 분석. 정보관리학회지, 34(4), 247-271.
- 서혜란 (2003). 디지털자료의 납본과 보존을 위한 각 국가의 노력. 정보관리학회지, 20(1), 373-399.
- 안영희, 박옥화 (2009). 대학도서관 서비스의 디지털 큐레이션 전략. 한국도서관·정

-
- 보학회지, 40(4), 311-326.
- 오석홍 (2018). 조직이론. 서울: 박영사.
- 오성환 (2018). 문화유산 정보의 디지털 보존 정책연구. 박사학위논문. 건국대학교 대학원 문화콘텐츠학과.
- 유영수 (2006). 전자기록의 장기보존을 위한 포맷등록관리 시스템 개발 방안 연구: 영국PRONOM Service를 중심으로. 석사학위논문. 명지대학교 대학원 기록관리학과.
- 윤재풍 (2014). 조직론. 서울: 대영문화사.
- 이소연 (2002). 디지털 아카이빙의 표준화와 OAIS 참조 모형. 정보관리연구, 33(3), 45-68.
- 이승민 (2015). 디지털 아카이빙을 위한 보존 메타데이터 패키지 구축. 정보관리학회지, 32(3), 21-47.
- 이재선 (2012). 디지털 도서관 아카이빙 현황 및 발전과제. 제5차 기록보존기술 공동학술세미나 자료집: 디지털 기록, 아카이빙 동향과 신기술. 국가기록원.
- 이정은, 윤은하 (2018). ISO 15489 개정판의 주요 특징에 관한 연구. 기록학연구, 57, 75-11.
- 이창길 (2016). 인적자원행정론. 서울: 법문사.
- 정영미, 윤화목, 김정택 (2010). 디지털 콘텐츠의 무결성 유지를 위한 장기적인 보존 정책에 관한 연구. 정보관리연구, 41(4), 205-226.
- 정혜경 (2005). 디지털보존의 비용요소에 관한 연구. 정보관리학회지, 22(1), 47-64.
- 조석준, 임도빈 (2016). 한국행정조직론. 법문사.
- 한국기록학회 (2008). 기록학 용어 사전. 역사비평사.
- 한나은, 김성희 (2014). 외국 대학도서관의 디지털 큐레이션 프로세스 비교분석. 한국도서관·정보학회지, 45(2), 93-116.
- 황윤영, 이규철 (2013). 이질적 전자기록물로 구성된 복합전자기록물 장기보존에 관한 연구 조사 및 분석. 한국전자거래학회지, 18(4), 1-24.
- British Library (2017). Sustaining the Value: The British Library Digital Preservation Strategy 2017-2020.
https://www.bl.uk/britshlibrary/~media/b/global/digital%20preservation/b_digitalpreservationstrategy_2017-2020.pdf. (accessed 2018.09.30.)

-
- Brown, Adrian (2013). *Practical Digital Preservation: A how-to guide for organization of any size*. London: Facet Publishing.
- Canadian Heritage Information Network (2015). Checklist for creating a preservation policy.
<http://www.canada.ca/en/heritage-information-network/services/digital-preservation/>.
- Caron, B., Houssaye, J., Ledoux, T., & Reecht, S. (2017). Life and Death of an Information Package: Implementing the Lifecycle in a Multi-Purpose Preservation System. iPRES 2017 14th International Conference on Digital Preservation, Sep 2017, Kyoto, Japan.
- CCSDS (2002). Reference Model for an Open Archival Information System(OAIS), January 2002.
- Corrado, E. M., & Sandy, H. M. (2017). *Digital Preservation for Libraries, Archives, and Museums*. 2nd ed. Lanham: Rowman & Littlefield.
- Cunningham, Adrian (2008). Digital curation/digital archiving: a view from the National Archives of Australia. *The American Archivist*, 71(2), 530-543.
- Dappert, A. et al. (2016). *Digital Preservation Metadata for Practitioners: Implementing PREMIS* (2016). London: Springer.
- Derrot, S., & Oury, C. (2014). Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France.
https://www.oaic.gov.au/images/documents/about-us/working-with-others/icon-meetings/Koerbin_ICON_2014.pdf. (accessed 2018.08.15.)
- Digital Curation Center(DCC) (2014). What is digital curation?
<http://www.dcc.ac.uk/digital-curation/what-digital-curation>. (accessed 2018.09.28.)
- Digital Preservation Coalition (2015). *Digital Preservation Handbook*. DPC.
<https://dpconline.org/handbook>. (accessed 2018.06.30.)
- Gartner, R., & Lavoie, B. (2013). *New Preservation Metadata*, DPC Technology Watch Report 13-3.
- Gömpel, Renate, & Lars G. Svensson (2011). "Managing Legal Deposit for Online

-
- Publications in Germany”, World Library and Information Congress: 77th IFLA General Conference and Assembly, 13-18 August 2011, San Juan, Puerto Rico. <https://www.ifla.org/past-wlic/2011/193-goempel-en.pdf>
- ICPSR (2012). ICPSR Digital Preservation Policy Framework. <http://www.icspr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html>. (accessed 2018.08.15.)
- IFLA Statement on Legal Deposit (2011). <https://www.ifla.org/publications/ifla-statement-on-legal-deposit-2011> (accessed 2018.08.15.)
- International Federation of Library Association (2017). Born Digital Legal Deposit Policies and Practices. <http://library.ifla.org/1905/> (accessed 2018.08.30.)
- International Publishers Association (2014). Digital Legal Deposit-An IPA Special Report, 24th July 2014. International Publishers Association. <https://www.internationalpublishers.org/images/news/2014/digital-legal-deposit-2014.pdf>
- Jackson, Andrew N. (2012). Formats over Time: Exploring UK Web History. <http://arxiv.org/abs/1210.1714>
- Kastellec, Mike (2012). Practical limits to the scope of digital preservation. *Information Technology and Libraries*, 31(2), 63-71.
- Krewer, D., & Mullins, J. (2017). Current Approaches to Implementing Preservation Metadata. A Report of the ALCTS PARS Preservation Metadata Interest Group Meeting. American Library Association Annual Conference, Orlando, June 2016. *Journal Technical Services Quarterly*, 34(2), 194-198.
- Lavoie, B. (2000). Meeting the challenges of digital preservation: the OAIS reference model. *OCLC newsletter (Online)* 243(January/February, 2000). (accessed 2018.08.30.)
- Lavoie, B., & Gartner, R. (2013). Preservation Metadata Second edition: DPC Technology Watch Report 13-03 May 2013. <https://www.dpconline.org/docs/technology-watch-reports/894-dpctw13-03/file> (accessed 2018.08.30.)

-
- Library and Archives Canada (2017a). Strategy for a Digital Preservation Program.
<https://www.bac-lac.gc.ca/eng/about-us/publications/Documents/LAC-Strategy-Digital-Preservation-Program.pdf>
- Library and Archives Canada (2017b). Digital Preservation Policy Framework: Development Guideline Version 2.1.
<https://www.canada.ca/en/heritage-information-network/services/digital-preservation/policy-framework-development-guideline.html>.
- Library and Archives Canada (2018). Annual Report 2017-2018.
<https://www.bac-lac.gc.ca/eng/about-us/annual-reports/annual-report-2017-2018/Documents/Report2018.pdf>
- Library of Congress (2010). The National Digital Information Infrastructure and Preservation Program 2010 Report.
http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf
- McGath, Gary (2013). The Format Registry Problem. Code4Lib Journal, 19, 1940-5758.
- National Digital Stewardship Alliance (2013). STAFFING FOR EFFECTIVE DIGITAL PRESERVATION AN NDSA REPORT.
<https://ndsa.org/documents/NDSA-Staffing-Survey-Report-Final122013.pdf>
- National Information Standards Organization (2004). Understanding Metadata, NISO Press, Bethesda, MD.
- National Library of Australia (2013). Digital Preservation Policy 4th ed.
<https://www.nla.gov.au/policy-and-planning/digital-preservation-policy>.
(accessed 2018.09.30.).
- National Library of Australia (2018). Collection digitisation policy.
<https://www.nla.gov.au/policy-and-planning/collection-digitisation-policy>
(accessed 2018.09.30.)
- National Library of Medicine (2007). Digital Repository Policies and Functional Requirements Specification. NLM Digital Repository Working Group.
<https://www.nlm.nih.gov/digitalrepository/NLM-DigRep-Report-rev03200>

7.pdf

Nicholson, D., & Dobрева, M. (2009). Beyond OAIS: Towards a reliable and consistent digital preservation implementation framework. In: 16th International Conference on Digital Signal Processing. IEEE.

National Library of Medicine (2007). Requirements for an NLM Digital Repository: Report and Recommendations.

<https://www.nlm.nih.gov/digitalrepository/NLM-DigRep-Report-rev032007.pdf>

Nicholson, D., & Dobрева, M. (2009). Beyond OAIS: Towards a reliable and consistent digital preservation implementation framework. In: 16th International Conference on Digital Signal Processing. IEEE.

PREMIS Editorial Committee (2008). Data Dictionary for Preservation Metadata Version 2.0. Retrieved from

<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>

PREMIS Editorial Committee (2015). PREMIS Data Dictionary for Preservation Metadata Version 3.0. Retrieved from

<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

PREMIS Working Group (2005). Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group. Retrieved from

<http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf>

PREMIS Working Group (2015). PREMIS Data Dictionary for Preservation Metadata Version 3.0. Retrieved from

<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

Preservation Advisory Centre (2013). Building a Preservation Policy.

https://www.bl.uk/abouts/stratpolprog/collectioncare/publication/bookleaves/building_a_preservation_pol.

Rosenthal, D. (2014). Talk "Costs: Why Do We Care?", DSHR's Blog.

<http://blog.dshr.org/2014/11/talk-costs-why-do-we-care.html>

Rosenthal, D. (2015a). "The Prostate Cancer of Preservation" Re-examined.

<http://blog.dshr.org/2015/09/the-prostate-cancer-of-preservation-re.html>
Rosenthal, D. (2015b). Emulation & Virtualization as Preservation Strategies.
https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf.
Sabharwal, Arjun (2015). Digital Curation in the Digital Humanities, Chandos Publishing.
The Law Library of Congress, Global Legal Research Center (2018). Digital Legal Deposit in Selected Jurisdictions. The Law Library of Congress, Global Legal Research Center
The Royal Library, Denmark (2012). Policy for long term preservation of digital materials at the Royal Library.
http://www.kb.dk/export/sites/kb_dk/da/kb/downloadfiler/PreservationPolicyDigitalMaterials_21092012.pdf
United Nations High Commissioner for Refugees (2017). DER - Digital Preservation Project (DPP).
<https://www.unhcr.org/UNUser/Documents/DownloadPublicDocument?docId=553356>
World Library (2017). 미국의회도서관, 디지털 컬렉션 이용을 위한 새로운 웹사이트 개설.
http://wl.nl.go.kr/usr/com/prm/BBSDetail.do?bbsId=BBSMSTR_000000000456&menuNo=12002&upperMenuId=12&nttId=5660&boardTab=null&boardGubun=null

[웹사이트]

국가기록원. <http://www.archives.go.kr/> (accessed 2018.11.30.)
국립의과학지식센터.
<http://library.nih.go.kr/ncmiklib/cmmu/intro/> (accessed 2018.11.30.)
국립중앙도서관. <http://www.nl.go.kr/> (accessed 2018.12.10.)
한국과학기술정보연구원. <https://kisti.re.kr/> (accessed 2018.10.30.)
American Institute for Conservation of Historic and Artistic Works
<http://www.conservation-us.org/> (accessed 2018.08.15.)

Bibliothèque nationale de France. <http://www.bnf.fr/en/> (accessed 2018.10.30.)

British Library. <https://www.bl.uk/> (accessed 2018.11.30.)

Centre Informatique National de l'Enseignement Supérieur (Le C.I.N.E.S.)
<https://www.cines.fr/en/long-term-preservation/a-concept-problems-2/long-term-preservation>. (accessed 2018.09.28.)

German National Library.
http://www.dnb.de/EN/Erwerbung/Pflichtablieferung/pflichtablieferung_node.html. (accessed 2018.09.30.)

JISC 영국 웹 도메인 데이터 세트.
<http://data.webarchive.org.uk/opendata/ukwa.ds.2/> (accessed 2018.08.30.)

Library and Archives Canada.
<https://www.bac-lac.gc.ca/eng/> (accessed 2018.09.30.)

Library of Congress. <https://www.loc.gov/> (accessed 2018.10.30.)

National Diet Library, Japan.
<http://www.ndl.go.jp/en/preservation/dlib/index.html>. (accessed 2018.08.15.)

National Library of Australia. <https://www.nla.gov.au/> (accessed 2018.11.30.)

National Library of the Netherlands. <https://www.kb.nl/en/> (accessed 2018.08.30.)

Open UK Web Archive.
<https://www.webarchive.org.uk/ukwa/> (accessed 2018.09.30.)

The National Archives.
<http://www.nationalarchives.gov.uk/> (accessed 2018.10.30.)

연구진

- 연구수행기관 : 한성대학교 산학협력단
- 책임 연구원 : 서 은 경 (한성대학교 크리에이티브인문학부 교수)
- 공동 연구원 : 최 상 희 (대구가톨릭대학교 도서관학과 교수)
박 희 진 (한성대학교 크리에이티브인문학부 교수)
윤 소 영 (서울여자대학교 연구원)
김 지 성 (한성대학교 사회과학부 교수)
- 보조 연구원 : 감 미 아 (연세대학교 문헌정보학과 박사과정)
서 해 린 (연세대학교 문헌정보학과 석사과정)
정 의 연 (대구가톨릭대학교 문헌정보학과 석사과정)
신 규 정 (한성대학교 조교)

2018 연구보고

국립중앙도서관 디지털자원 보존전략 수립 연구

인 쇄 : 2018년 12월 21일

발 행 : 2018년 12월 23일

발행처 : 국립중앙도서관

서울특별시 서초구 반포대로 201(반포동)

Tel. 02-590-0653 Fax. 02-590-0546

<http://www.nl.go.kr>

연구기관 : 한성대학교 산학협력단

ISBN 979-11-89585-30-3 93020 [비매품/무료]

인쇄 : 조은클터(02-2269-1481)



국립중앙도서관

National Library of Korea

비매품/무료

93020



9 791189 585303

ISBN 979-11-89585-30-3