



전자기록 장기보존 정책 및 보존포맷 재설계 방향

2019. 3. 27(수)

목차

0. 추진 배경 및 취지

1. 전자기록 장기보존 정책

- 현황 및 문제점
- 주요 해외사례
- 장기보존정책 주요내용

2. 전자기록 보존포맷 재설계

- 현황 및 문제점
- 주요 해외사례
- 보존포맷 및 장기보존패키지 다양화

3. 국가기록원 전자기록 보존 관련 동향

- 전자기록 기술정보 수집·활용 체계
- 전자기록 손상파일 복구 연구
- 향후 계획

추진 배경 및 취지

➤ 추진 배경

- 법 개정('07) 및 기록관리시스템 혁신 ISP 사업('06) 이후 큰 변화 없이 그대로 유지
 - 전자기록의 장기보존을 위한 제도적 기반과 표준 문서 위주의 전자기록 장기보존체계(기록관리시스템, 중앙영구기록관리시스템, 보존포맷변환, 장기검증 등)를 마련하였으나,
 - 전자기록 생산 및 관리 환경의 변화에 따른 적절한 보존 정책 필요
- 국가기록관리 혁신방안('18)의 '전자기록 장기보존 정책 및 보존포맷 재설계 수립 필요성' 제기
 - 장기보존 기본 정책 없이 일부 기록 유형 및 정보기술별 절차만 있어 포괄적 관리와 보존에 어려움

➤ 취지

- 전자기록 장기보존 및 보존포맷 정책에 따라 기록유형별 일관된 개선
 - 보존분야 제도(법, 표준 등), 프로세스 개선 및 전자기록 생산·보존 환경에 적합한 연구개발 활성화



1

전자기록 장기보존정책

01 현황 및 문제점

- **(관련 근거) 『공공기록물 관리에 관한 법률』 제20조, 동법 시행령 제36조, 제46조**
 - 기록관(특수기록관) 및 영구기록물관리기관의 전자기록 보존
- **일관된 전자기록 관리에 필요한 기본 장기보존 정책 부재**
 - 현재는 문서유형에 대한 보존포맷 변환 정책만 존재
 - 데이터세트, 시청각, 웹기록 등 다양한 기록 유형의 보존정책 부재
 - 전자기록 관리단계에서 기록정보 변경에도 진본성·무결성 검증 가능 체계 마련 필요
- **중앙기록물관리기관에서 정한 보존방법을 일괄적으로 적용**
 - 영구기록물관리기관 주도로 관할 기관의 전자기록물 생산·보존 환경 등 특성을 반영한 다양한 기술·방법 적용을 통한 관련분야 기술발전 유도 필요
- **정보기술 변화에 따른 전자기록 생산·관리 환경에 대한 적절한 대응 필요**
 - 다양한 기술변화에 따른 장기보존 및 활용 서비스 전략 필요
 - 기록관리 업무 담당자의 지속적 전자기록관리 업무 훈련 및 연구개발 활성화

02 주요 아카이브 해외사례

구분	미국 NARA	영국 TNA	캐나다 LAC	호주 NAA	스위스 SFA
보존 전략	•마이그레이션 •선호포맷	•에물레이션 •원 포맷 보관	•마이그레이션 •선호포맷	•마이그레이션 •선호포맷 •원 포맷 보유	•마이그레이션 •원본 보존
보존 절차	•보존 우선순위 설정, 일정수립 •보존업무 수행	•보존위험모형 •취약기록 식별	•기술동향 파악 •보존위험성 평가 •보존처리 우선순위 설정	•독점포맷 장기접근 SW 보급 •보존기술적용시 공개 표준 준수 •오픈포맷으로 변환	•이관시SIP로 제출 -결함기록은 거부가능 •SIP는 변경없이 보존 •AIP는 삭제불가능
상태 검사	•바이러스, 불변성검사 •쓰기방지 매체 사용 •이관후 모든 활동 기록 •매년 무결성 매체 표본검사 •분기별 파일검증을 위해 로그검사	•암호화하여 보존 •진본성 검증 •이관절차 표준화	•손상여부 표본검사	•정기적 불변성(fixity)검사 •저장장치 검사 후 주기적 백업	•보존포맷만 이관 •기록에 UUID 부여
위험 관리	•취약포맷 조치계획 •포맷 변환수행 •포맷/매체 노후화 검사	•DROID 관리 •PRONOM 활용 •보존위험 관리 •보존위험 측정 및 결과 공개	•예방차원 보호조치 적용 •보존처리 조치	•오픈 포맷 변환불가 포맷 대책 마련 •위험관리체계구축 -위험도 모니터링 및 관리	•보존포맷 정기적 재평가 •보존기록 모니터링 및 위험평가
인프라 관리	•디지털보존용 인프라 계획수립	•저장소 비용모델, 데이터용량 전망	•적절한 보존인프라 확보 •신규 솔루션에 필요한 비용 등 분석	•모듈방식으로 디지털 아카이빙 구축	•기록만 보존
참조 모델	•OAIS	•신뢰받는 디지털저장소	•OAIS •ISO 16363 신뢰받는 디지털 저장소	•OAIS	•OAIS

03 장기보존정책 목적, 대상, 방향

● 목적

- 장기적으로 보존할 가치가 있는 모든 유형의 전자기록물을 장기간에 걸쳐 안전하게 보존하고 접근이 가능할 수 있도록 함

● 적용 대상

- (기관) 기록물을 장기간 보존하는 영구기록물관리기관, 정부산하공공기관 등
- (대상) 전자적으로 생산된 기록물 + 원본이 폐기된 전자화 기록물
- (재량) 기관 특성에 따라 재량으로 보존정책 수립

보존정책의 방향

- ① 환경변화에 관계없이 전자기록물을 장기간 보존할 수 있는 **지속가능성**
- ② 이론과 실무의 토대로서 인정받은 국내외 **표준의 준수**
- ③ 새로운 기술변화에 쉽게 대응할 수 있는 **유연성**
- ④ 정보자원 및 기능의 확대·축소가 용이한 **확장성**

04 전자기록 장기보존 원칙

● 전자기록물의 진본성·무결성 보장 및 유지

- (요건) 프로세스별 메타데이터, 불변성 검사를 통한 검증 및 정보보호
- (재량) 전자기록물 불변성 검사(fixity check) 기술 다양화

요건이 구체적이지 않고 검증기술(전자서명 및 시점확인정보)만 제시



- ① 프로세스별 업무수행
- ② 메타데이터
- ③ 다양한 검증기술로 확대
- ④ 정보보호

● 장기간 접근 및 이용가능성

- 오랜 시간이 지나도 재현이 가능하도록 하여야 하며, 포맷, 저장정보 등 관련 메타데이터 관리 및 이용가능성 확보

● 장기보존을 위한 OAIS 참조모델 준수

- 기록물 유형 및 형태에 상관없이 인수하여 장기보존패키지로 변환

표준전자문서만 장기보존패키징



다양한 기록유형에 장기보존패키지 확대 적용

● 메타데이터

- 기록관리 업무 외에도 장기보존, 재현, 서비스를 위한 메타데이터 관리

05 장기보존 전략

보존전략 확대

- (마이그레이션) 포맷변환, 재패키징, 매체 재수록, 복제

* 마이그레이션 : 전자기록의 필수요소를 보존하면서 하나의 기술을 다른 기술로, 하나의 H/W, S/W를 다른 H/W, S/W로 옮기는 것

- (에물레이션) 기록물 외형과 구성을 그대로 표현해야 하는 경우 등에 적용
- (XML) 공개표준으로 특정 HW나 OS에 관계없이 파일포맷, 보존패키지 등에 사용

마이그레이션, XML 방식



에물레이션 방식 추가

보존포맷 다양화

- 기록유형별 기준에 적합한 보존포맷 기준을 정하고, 보존포맷 다양화

문서유형을 위한 보존포맷 'PDF/A-1'만 규정



기록물 유형별 다양한 보존포맷 확대

장기보존패키지 다양화

- 기록물 유형 및 기관 특성에 따라 적용가능한 장기보존패키지 다양화

장기보존패키지 'NEO'만 규정



기록물 특성 등에 따라
다양한 장기보존패키지 선택

매체 재수록 및 복제

- 매체 특성 및 수명을 고려하여 주기적으로 매체에 재수록하며, 매체 수록시 비트(bit) 변경없이 복제

06 그 밖의 장기보존정책

● 보존 계획 및 인프라

- (보존계획) 전자기록 객체, 매체, 구동SW, 컴퓨팅 플랫폼 등에 대한 보존계획 수립·실행
 - * 기록물 및 보존매체의 불변성 검사 대상, 주기, 결과에 따른 처리, 포맷변환 시기, 스토리지 저장 용량 점검 등
- (인프라) OAIS 참조모형에 따른 분야별 기술에 대한 전략 수립, 포맷변환 도구, 시스템 등 구축
 - * 전자기록의 안전한 보호 조치, 기록/시스템/서비스별 접근권한 설정 등

● 기술과 훈련

- 적절한 기술력을 보유한 직원이 전자기록 보존업무를 수행할 수 있도록 지원
- 조직의 직원들이 전자기록 관련 업무를 수행할 수 있도록 전문 지식 습득 기회 제공

● 연구와 협력

- 장기보존정책 지원을 위해 신기술을 접목한 장기보존 분야 연구개발 활성화

● 정책의 운영 및 유지

- 국가기록관리위원회에 보고되며, 5년마다 혹은 필요시 재검토



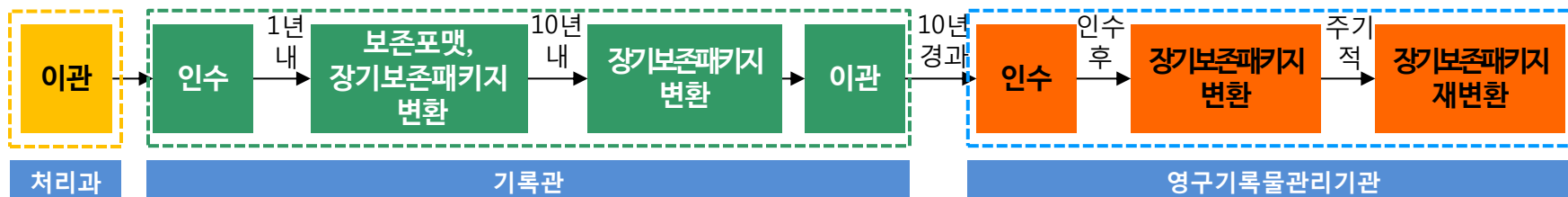
2

전자기록 보존포맷 재설계

01 현황 및 문제점

- (관련 근거) 『공공기록물 관리에 관한 법률』 제20조, 영 제36조, 제46조
 - 중앙기록물관리기관의 장이 보존포맷과 장기보존패키지를 정하도록 함

- 보존포맷 변환 프로세스



- (보존포맷) 오피스계열 문서에 적합한 PDF/A-1으로 규정되어있으나, 기술변화에 따른 다양한 포맷을 적용하지 못하며, 다양한 기록 유형 적용 필요
- (장기보존패키지) 현행 장기보존패키지인 NEO는 Base64 인코딩이 필요하므로 대용량 기록에 적합하지 않아 다른 형태의 장기보존패키지 필요

02 주요 아카이브 해외사례

기관	이관가능포맷	보존전략	기록유형분류	비고
미국 의회도서관 (LOC)	선호, 허용포맷	-	①문서, ②프리젠테이션, ③데이터세트, ④이미지, ⑤오디오, ⑥지리공간, ⑦웹	
미국 국립기록청 (NARA)	선호, 허용포맷	마이그레이션	①문서, ②프리젠테이션, ③데이터세트, ④이미지, ⑤오디오, ⑥동영상, ⑦CAD, ⑧지리공간, ⑨웹, ⑩이메일	
영국 국가기록원 (TNA)	생산당시 포맷	에물레이션 * 보존대안이 없는 경우, 마이그레이션	①문서, ②프리젠테이션, ③데이터세트, ④이미지, ⑤오디오, ⑥동영상, ⑦이메일, ⑧프로젝트관리	
캐나다 국립도서관 (LAC)	선호, 허용포맷	마이그레이션	①문서, ②프리젠테이션, ③데이터세트, ④이미지, ⑤오디오, ⑥동영상, ⑦CAD, ⑧지리공간, ⑨이메일	규정된 포맷 외 이관 불가
호주 국가기록원 (NAA)	선호, 허용, 위험포맷	마이그레이션	①문서 ②프리젠테이션 ③데이터세트, ④이미지, ⑤오디오, ⑥동영상, ⑦CAD, ⑧지리공간, ⑨웹, ⑩이메일, ⑪캡슐화(encapsulation)	생산당시포맷도 보유
스위스 연방기록원 (SFA)	보존포맷	마이그레이션	①문서, ②데이터세트, ③이미지, ④오디오, ⑤동영상, ⑥지리공간, ⑦웹	규정된 포맷 외 이관 불가
중국 당안국	선호, 허용포맷	-	①문서, ②데이터세트, ③이미지, ④오디오, ⑤동영상, ⑥웹, ⑦이메일	
한국 국가기록원	보존포맷	마이그레이션	①문서	생산당시포맷도 보유

03 전자기록 유형별 보존포맷(문서보존포맷) 다양화

보존포맷 선정 방법 개선

- (기존) 중앙기록물관리기관에서 포맷을 정하고 기관에 일괄 적용
- (개선) 포맷선정 기준 수립 후 기준에 맞는 포맷을 영구기록물관리기관 등이 재량으로 선택

공공표준으로 '포맷' 선정

표준에서 정한 포맷을 기관에 일괄 적용



(원칙) 공공표준으로 '포맷선정 기준' 규정
(포맷)

- 영구기관 및 정부산하기관은 재량으로 선택
- 국가기록원 관할 기관 등을 위해 '보존포맷' 권고

(적용) 영구기관 및 정부산하기관이 선택한 보존포맷을 관할 기관에 적용

보존포맷 선정 기준 수립

- 공통기준(모든 유형에 동일하게 적용)과 고유기준(유형별 특성) 수립
 - * 공통기준 : 개방성, 안정성, 상호운용성, 자체문서화, 검색기능
- 선정기준에 따른 기록유형별 포맷 평가 후 보존포맷 선정

기록 유형별 보존포맷 다양화

(문서) PDF/A-1



(문서) PDF/A-1, PDF/A-2, TXT 등

* 환경에 따라 ODF, EPUB 등 재량으로 선택

(데이터세트) SIARD, CSV, JSON 등

(이미지) TIFF, JPEG 등

(음성) FLAC, AAC 등

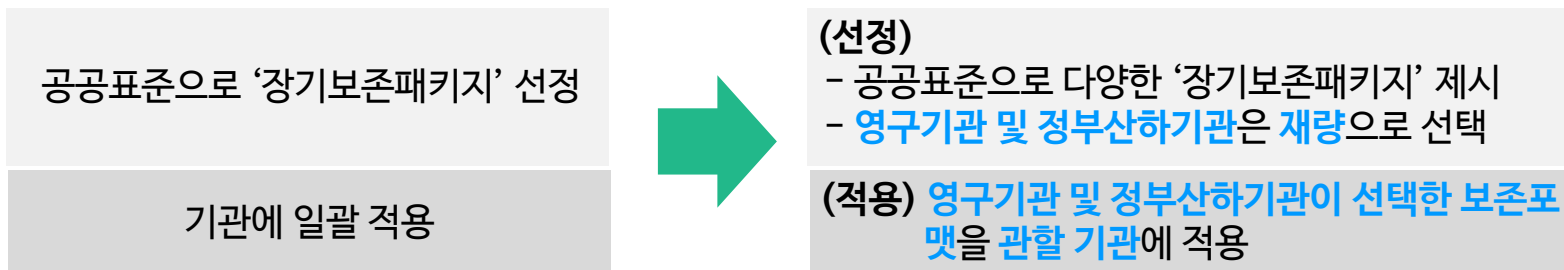
(영상) FFV1, MP4 등

(웹) WARC 등

04 장기보존패키지 (장기보존포맷) 다양화

장기보존패키지 선정 방법 개선

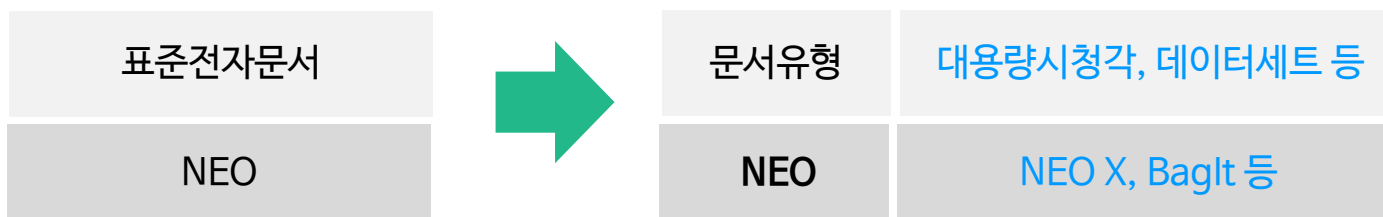
- (기존) 중앙기록물관리기관에서 정하고 기관에 일괄 적용
- (개선) 패키지 선정 기준 수립 후 영구기록물관리기관 등이 재량으로 선택



장기보존패키지 선정 기준 수립

- 전자기록의 진본성과 무결성이 유지될 경우 선정 가능하며, 기록물 특성 등에 따라 선택 가능
 - * (진본성) 기록의 내용, 형상에 대한 것으로 생산자, 생산시점 등이 입증될 수 있는 것
 - * (무결성) 기록이 완전하고 변경되지 않는 것

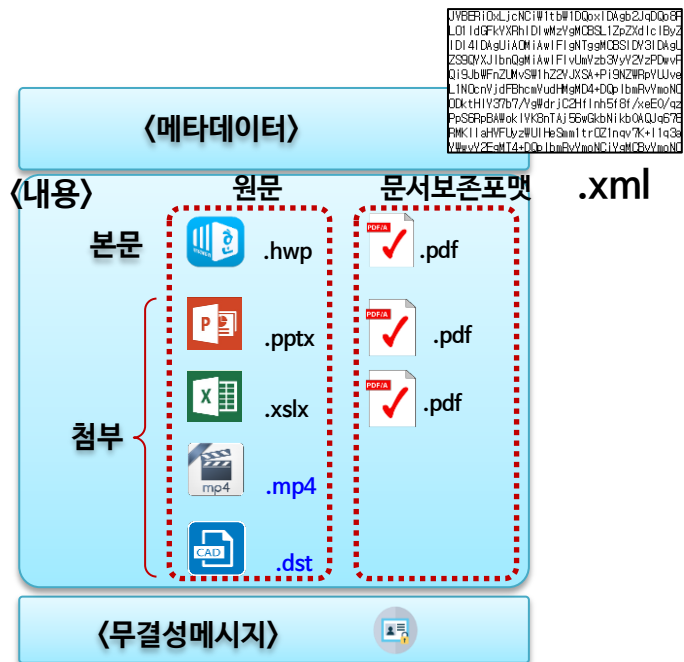
장기보존패키지 다양화



05 장기보존패키지 유형

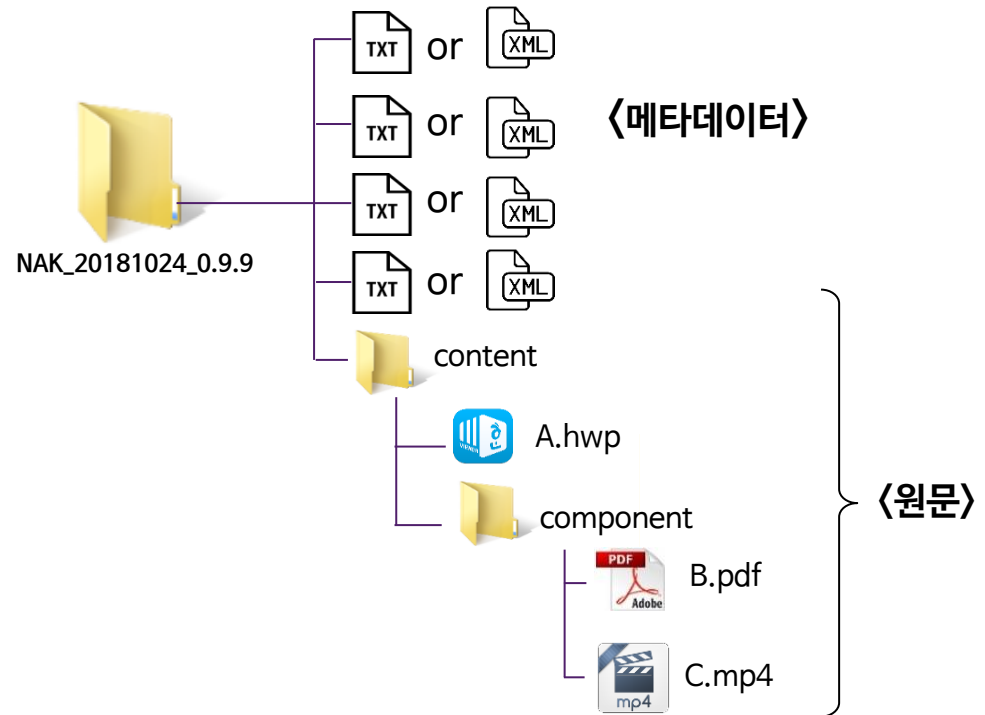
파일형

최종 개체: XML 형태의 텍스트 파일



폴더형

최종개체: 폴더



NEO2@NAK, VEO2@PROC, XENA@NAA

VEO3@NAA, Baglt@LAC, AAP@NARA



3

국가기록원
전자기록 보존관련
동향

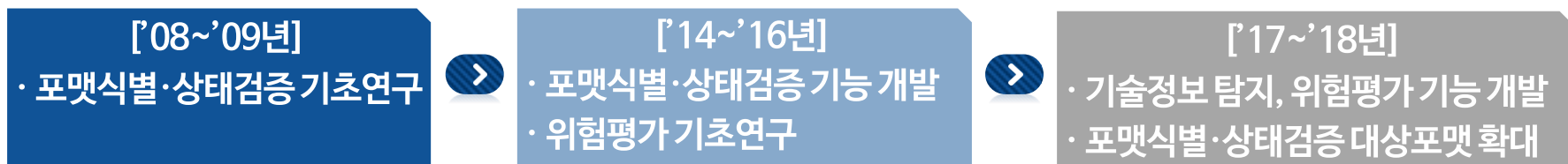
01 추진현황

- (관련 근거) 『공공기록물 관리에 관한 법률 시행령』 제46조

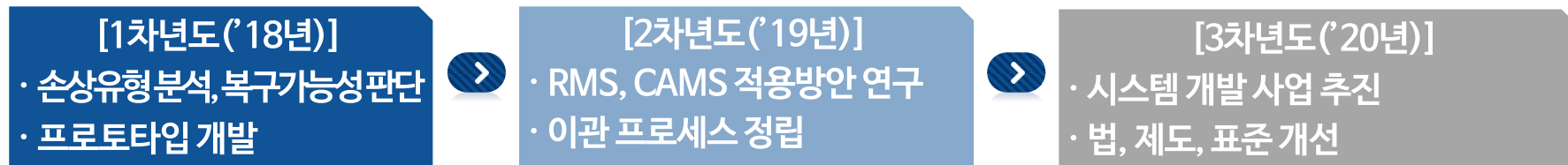
- ④ 영구기록물관리기관은 전자기록 손실방지를 위한 데이터, 매체, 시스템 등 복구체계 수립

- ⑤ 중앙기록물관리기관의 장이 전자기록 기술정보 수집·활용 방안 수립·시행

- 전자기록 기술정보 수집·활용체계 구축 현황



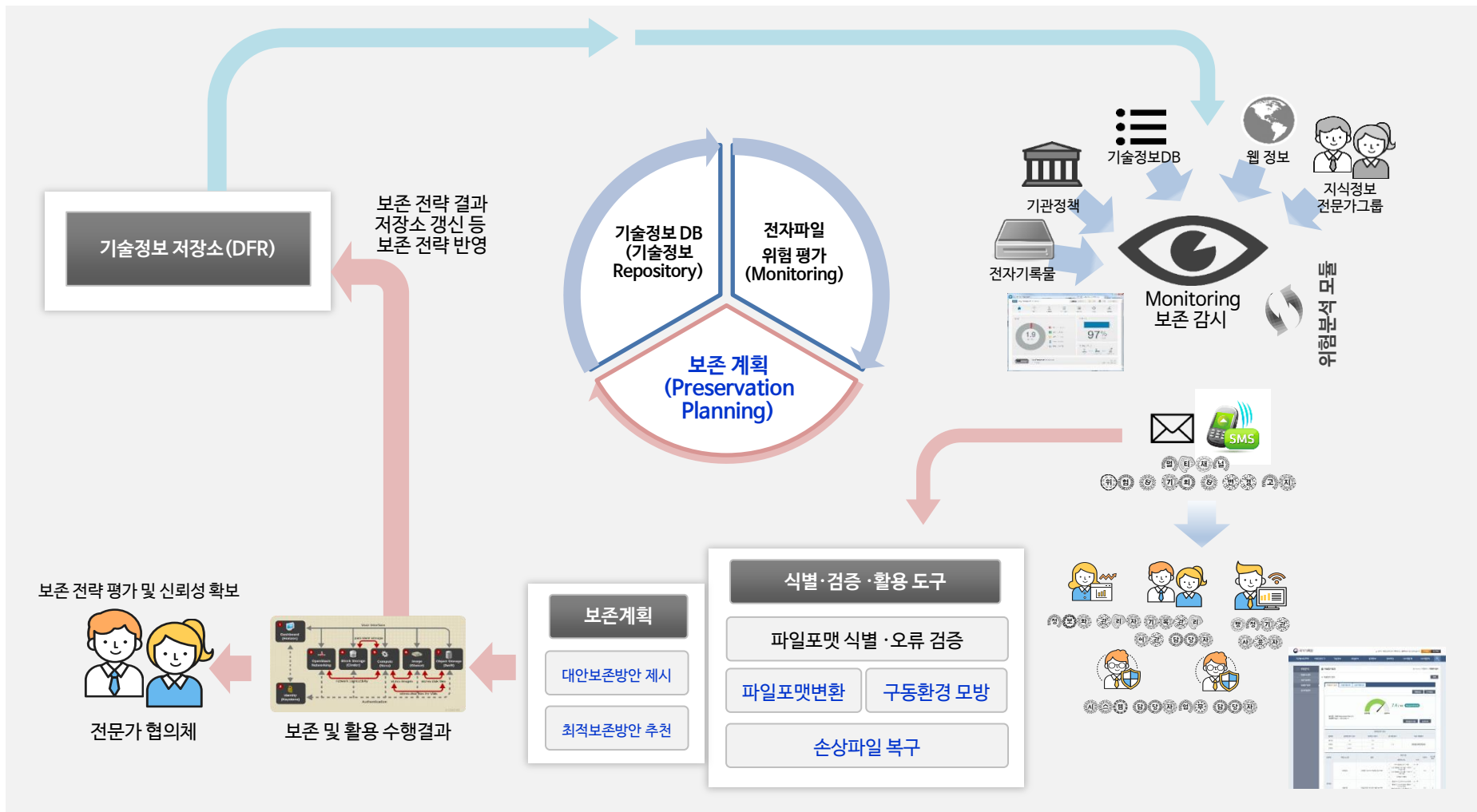
- 전자기록 손실파일 복구 연구



02 전자기록 기술정보 수집·활용 체계 개념도



03 향후 계획



04 기술정보 구축

● 파일포맷, 구동S/W, 제조사, 코덱에 관한 기술정보 구축

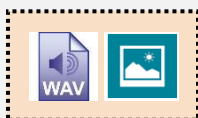
단계

추진
내용



파일포맷(총 31개 항목)

- 파일포맷 및 다수 버전 관련 기술항목 설계
- 내·외부서명, (국제/국가/단체)표준, 규격 문서 등의 정보 기술



- | | |
|----------|----------|
| ① 포맷명/버전 | ⑦ 표준화 |
| ② 관련식별자 | ⑧ 규격문서 |
| ③ 유형 | ⑨ 관련SW목록 |
| ④ 개요 | ⑩ 변환 경로 |
| ⑤ 내/외부서명 | ⑪ 최초발행일 |
| ⑥ 버전목록 | ⑫ 개발/유지사 |



구동S/W(총 16개 항목)

- 파일포맷구동(열기, 조회) S/W기술항목 설계
- 해당 S/W 지원 운영체제(Win, Mac 등) 및 H/W 환경(CPU, MEM, HDD 등) 기술



- | | |
|-----------|----------|
| ① SW명/버전 | ⑦ HW요구사항 |
| ② 관련식별자 | ⑧ 지원언어 |
| ③ 개요 | ⑨ 저작권 |
| ④ 지원포맷 | ⑩ 발행일 |
| ⑤ 관련문서 | ⑪ 종료일 |
| ⑥ S/W요구사항 | ⑫ 개발/유지사 |



업체 및 코덱(총 19개 항목)

- S/W업체 및 시청각 코덱 기술항목 기술
- S/W업체 연락처 및 코덱 개요, 표준, 규격문서 등의 정보 기술



- | | |
|-----------|-----------|
| ① 업체명 | ⑦ 코덱명 |
| ② 개요 | ⑧ 유형 |
| ③ 주소 | ⑨ 개요 |
| ④ 국가 | ⑩ 표준화 |
| ⑤ 전화번호 | ⑪ 규격문서 |
| ⑥ 사이트/이메일 | ⑫ 발행일/개발사 |

05 포맷별 위험평가

- 파일위험요소마다 중요도에 따라 가중치 부여 등을 통해 점수화

단계

추진
내용



가중치 부여



PDF/A 저위험(보존우수) 파일포맷 적용 사례

No.	분류명	요소명	점수	가중치	보정점수
1	공개성 (3)	규격존재	10	0.5	5
2		식별수준	10	0.3	3
3		검증수준	10	0.2	2
4	대중성 (4)	활용빈도	10	0.5 (0.5)	5
5		⑤브라우저지원	10	(0.1)	0
6		⑥구동 수준	10	0.5 (0.4)	5
7		의존도 상호 운용도			
8		보존포맷 채택	10	0.1 (0.1)	1

No.	분류명	요소명	점수	가중치	보정점수
9		표준화	10	0.4 (0.3)	4
10		⑤코덱표준화	-	(0.1)	-
10		제작정보 존재	10	0.2 (0.2)	2
11	안정성 (8)	⑥보존 지원 수준	10	0.3 (0.3)	3
12		변환경로 이용가능			
13		휴폐이지 수			
14		업체수			
15		버전변화 도			

No.	분류명	요소명	점수	가중치	보정 점수	No.	분류명	요소명	점수	가중치	보정 점수		
1	공개성 (3)	규격존재	9	0.5	4.5	9	안정성 (8)	표준화	5	0.4	2		
2		식별수준	10	0.3	3	10		㉠코덱표준화	-	-	-		
3		검증수준	10	0.2	2	11		제작정보 존재	10	0.2	2		
4	대중성 (4)	활용빈도	10	0.5	5	12		㉡보존 지원 수준	변환경로 이용가능	7.66	0.3	2.3	
5		㉢브라우저지원	-	-	-	13			휴폐이지 수				
6		㉣구동 지원 수준	의존도	8.417	0.5	4.209			14				업체수
7			상호 운용도						15				버전변화 도
8		보존포맷 채택	10	0.1	1								

공개성 : $4.5 + 2 + 2 = 9.5 \times 0.3 = 2.85$
 대중성 : $5 + 4.209 = 9.209 \times 0.4 = 3.68$
 안정성 : $1 + 2 + 2 + 2.3 = 7.3 \times 0.3 = 2.19$

$2.85 + 3.68 + 2.19 = 8.72(\text{보존우수})$

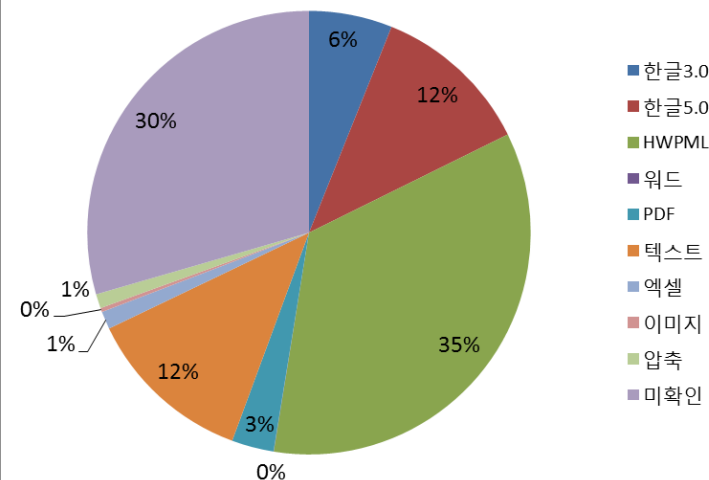
최종점수	0≤점수<2	2≤점수<4	4≤점수<6	6≤점수<8	8≤점수<10
표시색상					
위험등급	E등급 (보존위험)	D등급 (보존취약)	C등급 (보존보통)	B등급 (보존양호)	A등급 (보존우수)

06 전자기록 손상파일 복구 연구

- 국가기록원 이관 전자기록물 중 손상현황('17년까지 이관된 기록물 대상)
 - 손상파일 59,398개 중, 실제 분석이 필요한 파일 개수는 4,724개

포맷별 손상현황

한글3.0	287개
한글5.0	549개
HWPML	1,646개
워드	3개
PDF	145개
텍스트	579개
엑셀	61개
이미지	39개
ZIP	49개
미확인	1,366개



복구가능성 분석 결과

복구 가능

- 확장자 오기재
- 서식 수정
- 인코딩 파일 등

복구 불가능

- 구조는 정상이나 열리지 않는 파일
- 포맷구조의 큰 손상으로 포맷식별 불가

복구 대상 제외

대체파일, 빈파일, DRM, 링크파일 등

추가분석 필요

미확인, PDF, OLE파일(엑셀, 한글, etc) 등

손상으로 분류된 파일의 손상현황

- 59,398개의 손상으로 분류된 파일 중, 대부분이 HTML(링크파일) 29,044개, 원본 손실 파일 25,583개임 / 실제 분석대상 파일은 4,724개
- 4,724개의 파일에 대한 분석으로 복구 가능 케이스 분석(문서 서식정보만 깨진 경우, Base64 인코딩 파일 등)
- 약 **1,822개(39%)의 파일은 복구 가능** / 약 2,902개(61%)의 파일은 현재 복구 불가능하며 추가 연구 필요

향후 일정

▶ 전자기록 장기보존 정책 및 보존포맷 재설계 수립 및 실행

- 영구기록물관리기관, 각급 기록관 등 정책 의견조회(4월 중순)
- 국가기록관리위원회 보고(5월 예정) 및 확정(6월 예정)
- 전자기록 장기보존정책의 상세 이행 계획 수립(~12월)

▶ 보존포맷 및 장기보존패키지 공공표준 제·개정 추진

- 전자데이터세트 유형 보존포맷 및 폴더형 장기보존패키지 연구 추진(~12월)
 - * R&D 사업 ‘데이터베이스 유형 전자기록 장기보존기술 연구’ 수행(~11월)
- 보존포맷 및 장기보존패키지 공공표준 제·개정 추진(~’20년)

Q & A