

전자기록물의 장기보존을 위한 보존포맷 선정 방안에 관한 연구

A Study on the Selection of Preservation Format for Long-Term Preservation of Electronic Records

한희정(Hui-Jeong Han)¹, 오효정(Hyo-Jung Oh)², 양동민(Dongmin Yang)³

E-mail: freebirdhj@naver.com, ohj@jbnu.ac.kr, dmyang@jbnu.ac.kr



¹ 제 1저자 전북대학교 문화융복합아카이빙연구소 전임연구원
² 전북대학교 문헌정보학과 부교수, 문화융복합아카이빙연구소 공동연구원
³ 교신저자 전북대학교 기록관리학과 부교수, 문화융복합아카이빙연구소 공동연구원

논문접수 2020.1.21
최초심사 2020.1.28
게재확정 2020.2.11

ORCID

Hui-Jeong Han
https://orcid.org/0000-0003-2153-5963
Hyo-Jung Oh
https://orcid.org/0000-0001-8067-2832
Dongmin Yang
https://orcid.org/0000-0002-4029-9372

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

- 본 연구는 "2019년 행정안전부 국가기록원 기록관리 연구개발사업"의 연구비를 지원받아 수행되었음.
- 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5B8099507).
- 이 논문은 2017년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.

초 록

문서 유형의 전자기록물 장기보존을 위해 국가기록원에서는 PDF/A-1을 보존포맷으로 선정하였고 문서보존포맷으로 명명하여 공공표준으로 제정하였다. 문서 중심의 하나의 보존포맷인 PDF/A-1을 선택할 수 밖에 없기 때문에, IT 발전과 업무 변화에 따라 실제 업무에 적용할 수 있거나 적용해야만 하는 다양한 전자파일 포맷들을 활용하는데 제약을 주고 있으며, 문서 이외에 다른 유형의 전자기록물(행정정보데이터세트, 시청각기록물, 웹기록물 등)에는 적용되기 어렵다. 그러므로 전자기록물의 보존포맷을 다양화할 수 있는 방안이 필요하다. 본 논문에서는, 우선 전자기록물의 보존포맷을 다양화할 수 있는 보존포맷 선정체계를 제시한다. 또한, 보존포맷을 선정할 때, 모든 전자기록물에 공통으로 적용되는 기준인 공통기준 및 평가방식, 그리고 전자기록물 유형별로 적용되는 고유기준을 도출하기 위한 방법론을 제안하고자 한다.

ABSTRACT

For the long-term preservation of document-type electronic records, the National Archives of Korea has chosen PDF/A-1 as the preservation format named as the document file format, and established it as a public standard. The only option of selecting PDF/A-1 restricts the use of various electronic file formats that can or must be applied to actual works as IT advances and tasks change. Moreover, it is difficult to apply PDF/A-1 to other types of electronic records (administrative information datasets, audiovisual records, web records, etc.). Therefore, it is necessary to diversify the preservation formats of electronic records. We suggest a framework for selecting various preservation formats. Furthermore, we propose common criteria and evaluation methods frequently applied to all electronic records when selecting a preservation format, and introduce a methodology for deriving intrinsic criteria applied to each type of electronic records.

Keywords: 전자기록물, 장기보존, 보존포맷, 문서보존포맷, 선정기준, 공통기준, 고유기준
Electronic records, long-term preservation, preservation format, document file format, selection criteria, common criteria, intrinsic criteria

<https://jksarm.koar.kr>

www.kci.go.kr

1. 서론

1.1 연구 필요성 및 목적

공공기록물법 시행령 제36조 1항에 따르면 보존기간이 10년 이상으로 책정된 공공기관의 전자기록물은 문서보존포맷 및 장기보존포맷으로 변환하여 관리하여야 한다. 문서보존포맷은 전자기록물의 장기보존을 위해 필수적인 요소들을 규정하여 문서의 내용(Content)과 구조(Structure)를 보존할 수 있는 포맷이어야 하며, 국가기록원에서는 2008년에 제정한 PDF/A-1를 공공표준으로 규정하고 있다. 문서보존포맷은 OAIS 참조모형의 내용정보(CI: Content Information)에 해당된다. 장기보존포맷은 OAIS 참조모형의 보존기술정보(PDI: Preservation Description Information)에 해당하는 맥락(Context) 정보를 문서보존포맷과 함께 패키징하는 것으로, 국가기록원에서는 2008년에 제정하고 2013년과 2017년에 개정된 NEO version 2.1를 공공표준으로 규정하고 있다. 장기보존포맷은 OAIS 참조모형의 정보패키지(IP: Information Package)에 해당된다. 본 논문에서는 두 포맷 중에서 문서보존포맷에 초점을 맞추고 있다.

전자기록물 생산환경과 기술의 발전으로 다양한 전자기록물 유형이 업무에 활용되고 있으나 문서보존포맷 표준은 문서 유형에 특화되어 있어 행정정보데이터세트, 시청각기록물, 웹기록물 등 다른 유형에는 적용하기 어렵다. 또한, 포맷의 개수도 하나로 극히 제한적이기 때문에 모든 종류의 문서 유형에도 활용되기 어려운 실정이다. 그러므로 보존포맷)유형의 다양화 및 각 유형의 포맷들을 확대에 대한 연구가 필요하다.

보존포맷과 관련된 파일포맷의 기술 변화는 계속해서 이루어지고 있다. 보존포맷 확대를 위해 PDF/A-1 이외에 몇 개의 파일포맷들을 보존포맷으로 추가만으로는 기술 변화에 유연하게 대응할 수 없다. 전자기록물의 안전한 장기보존을 위한 대책을 수립·시행하는 영구기록물관리기관은 관련 공공기관들의 현황 및 특성을 고려하여 정책과 전략을 세워야 한다. 그런데 영구기록물관리기관들이 모두 동일한 보존포맷들을 공공표준으로 활용하면 몇몇 공공기관들의 현황과 특성을 반영할 수 없는 경우도 발생할 수 있으며, 해당 공공기관의 기록물들을 장기보존하기 어려울 수 있다. 그래서 보존포맷을 유연하게 확대하고, 전자기록물 유형별로 다양한 보존포맷을 사용할 수 있는 방안을 모색해야 한다. 또한, 공공기관들이 다른 보존포맷을 선정할 수 있게 된다면 혼란이 발생할 수 있으므로 보존포맷을 선정하기 위한 공통적인 체계를 만들어질 필요가 있다

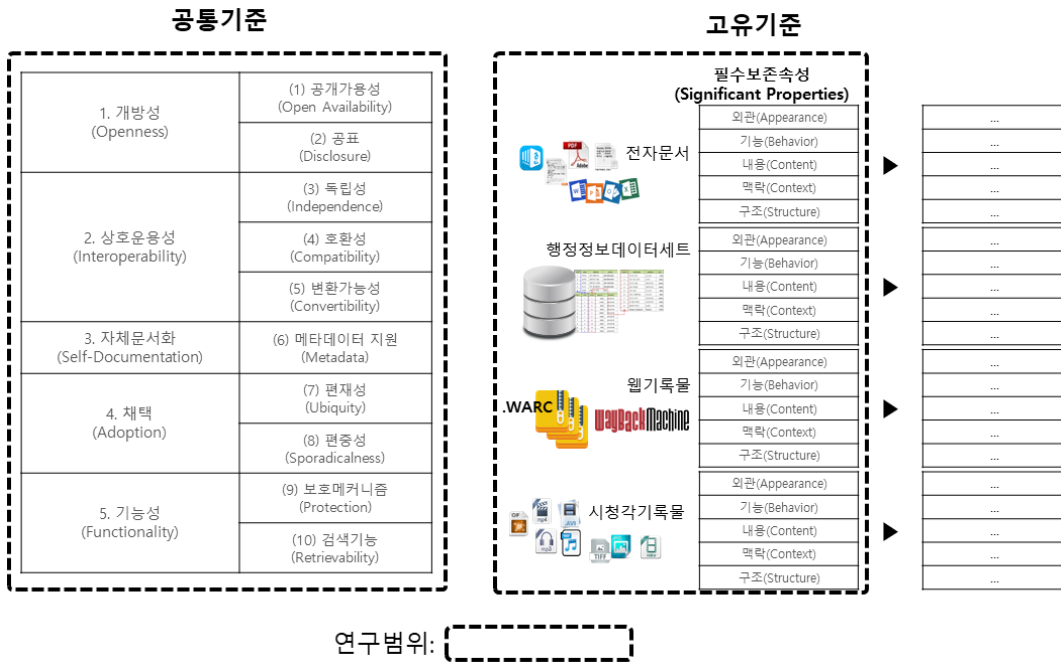
이에 본 논문은 전자기록물 유형별로 보존포맷을 다양화할 수 있는 선정체계를 제시하고, 모든 전자기록물에 공통으로 적용될 수 있는 기준인 공통기준 및 평가방식, 그리고 전자기록물 유형별로 적용될 수 있는 고유기준을 도출하는 방법론을 제안하는 것을 목적으로 한다.

1.2 연구범위

본 연구에서 가정하고 있는 보존포맷 선정 방안은 세 가지 내용으로 이루어져 있다. 첫 번째는 보존포맷 선정체계로, 파일포맷을 보존포맷으로 선정하기 위한 방법론이다. 보존포맷 선정체계에서는 공통기준 및 고유기준 중심의 선정 절차를 정의한다. 두 번째는 공통기준이다. 공통기준은 모든 파일포맷에 대해 공통으로 적용되어 보존포맷으로서의 적합성을 판단하는 기준이다. 이를 기반으로 공통기준에 따라 적합성을 판단하는 평가방식을 제안한다. 세 번째는 고유기준이다. 고유기준은 전자기록물 유형별로 보존될 특성에 따라 보존포맷으로서의 적합성을 평가하는 기준이며, 평가방식은 고유기준에서와 동일하다.

1) 공공기록물법 시행령에서 문서보존포맷은 전자기록물 전체에 해당되는 용어이다. 그러나 문서, 행정정보데이터세트, 시청각기록물, 웹기록물 등 여러 유형의 전자기록물에 대해 논의하는 과정에서, 문서보존포맷이 문서 유형에만 해당되는 것으로 오해할 수 있으므로, 본 논문에서는 ‘보존포맷’이란 용어를 사용한다.

보존포맷 선정 방안의 세 가지 내용 중에서 보존포맷 선정체계 그리고 공통기준을 본 연구의 주요 연구범위로 설정하였다. 이는 고유기준 수립을 위해 전자기록물 유형별로 보존해야 할 특성을 심도있게 조사 및 분석하고, 그 특성을 반영하는 기준과 이를 기반으로 평가하는 방법을 도출해야 하기 때문이다. 이에 본 연구는 <그림 1>처럼 선정기준에 관련하여, 공통기준은 정의, 근거, 평가방식까지 제안하고, 고유기준에 대해서는 전자기록물 유형별로 고유기준을 도출하는 방법론까지만 제시하고자 한다.



<그림 1> 선정기준 관련 연구범위

1.3 선행연구

국가기록원(2004)에서는 문서 유형의 전자기록물에 대해 XML, TEXT, 이미지파일, PDF, CSD, PDF/A-1을 대상으로 보존포맷으로서의 적합성 여부를 평가하였다. 보존포맷은 공개용 표준(Open Standard), 편재성(Ubiquity), 안정성(Stability), 메타데이터지원(Metadata Support), 상호운용성(Interoperability), 진본성(Authenticity), 처리능력(Processability), 표현력(Presentation), 검색 기능(Retrieval) 의 9가지 선정기준에 대한 평가를 통해서 가장 좋은 평가를 받은 PDF/A-1이 문서보존포맷으로 선정되었다. 성환혁(2007), 국가기록원(2013)에서는 다양한 유형의 전자기록물을 PDF/A-1로 변환했을 때 많은 기능과 내용들이 소실된다는 문제점들을 발견하였고 이에 대한 대책 마련이 필요하다고 언급하고 있다.

다양한 연구들이 PDF/A-1의 문제점을 해결하기 위해 진행이 되었다. 성환혁(2007)에서는 텍스트 문서, 다차원 문서, 다매체 문서의 3가지 유형으로 구분하고, PDF/A-1를 보존포맷으로 선정했을 때와 동일한 9개의 선정기준으로 평가하였다. PDF/A-2, ODF²⁾, Open XML³⁾ 3개의 파일포맷을 대상으로 보존포맷으로서의 가능성을 확인하였

2) ODF(Open Document Format for Office Applications)는 오피스 문서를 위한 XML 기반 개방형 파일 형식으로 OASIS가 추진하여 ISO/IEC 26300:2006 표준으로 승인되었다. 확장자는 .odt(문서), .odp(프리젠테이션), .ods(스프레드시트) 등과 같다.
 3) OpenXML(Office Open XML 또는 OOXML)도 오피스 문서를 위한 XML 기반 개방형 파일 형식으로 MS가 추진하여 유럽 ECMA-376 및 ISO/IEC 29500 표준으로 승인되었다. 확장자는 .docx(문서), .pptx(프레젠테이션), .xlsx(스프레드시트) 등과 같다.

다. 오세라, 정미리, 임진희(2016)에서는 오피스 유형에 대한 파일포맷으로 XML 기반 개방형 표준인 ODF가 PDF/A-1을 보존포맷의 역할을 대체 가능성을 높게 보고 있으며, 정미리, 오세라, 임진희(2016)에서는 ODF가 공문서의 생산·유통·관리·활용과 보존까지 문서의 전 생애주기에 걸쳐 활용될 수 있음을 제안하고 있다. 강현민(2016), 임나영, 남영준(2019), 박준영, 이명규(2019)에서는 전자기록물 유형 중 시청각기록물을 이미지를 보존하기 위한 파일포맷 및 디지털화 기준에 대해 구체적으로 논의하였다.

현재의 보존포맷은 파일포맷 기술의 변화, 보존포맷 위험도 증가에 따라 추가되거나 변경될 수 있다. 보존포맷 뿐만 아니라 다양한 파일포맷에 대한 위험평가가 지속적으로 이루어져야 하며, 이를 위해 가장 먼저 파일포맷 기술정보요소가 체계적으로 수집·관리되어야 한다. 국가기록원에서는 디지털 정보 파일포맷에 대한 구문정보와 의미정보 등을 저장하는 데이터베이스인 디지털 포맷 레지스트리(DFR: Digital Format Registry)를 2008년 최초로 구축하고 지금까지 지속적으로 고도화하고 있다(국가기록원, 2008; 국가기록원, 2009; 손원성 외, 2009).

송치호, 차현철(2017), 차현철, 최주호(2019)에서는 전자기록이 담겨져 있는 파일포맷의 위험도를 평가하였다. 송치호, 차현철(2017)에서는 위험도 평가에 대해 간략하게 개요를 제시하였다. 차현철, 최주호(2019)에서는 앞서 언급한 파일포맷에 대한 데이터베이스가 구축되어 있다고 가정하고, 파일포맷에 대한 위험도를 평가하기 위해 필요한 위험요소 항목을 도출한 후 각 항목에 대한 평가 기준을 제시하고, 이 기준을 기반으로 위험평가를 수행하는 방식을 제안하였다.

이렇듯 보존포맷 관련하여 다양한 연구들이 수행되었지만 다양한 유형의 전자기록물에 대해 보존포맷을 제시하기에는 한계를 가지고 있다. 즉, PDF/A-1의 단일 보존포맷 전략으로 발생하는 문제점들을 해결하기 위한 다양한 연구가 진행되어 왔지만 다양한 유형의 전자기록물들이 생산되고 이를 관리해야 하는 현재의 상황을 근본적으로는 해결하지는 못한다. PDF/A-1의 단점을 보완하기 위해 몇 개의 파일포맷을 보존포맷에 추가한다고 해도 PDF/A-1의 단일 보존포맷 전략에서와 동일한 상황을 직면하게 된다. 그러므로 파일포맷에 대한 데이터베이스를 구축·관리해야 하고, 보존포맷 선정기준을 도출하고, 평가해야 할 수 있는 체계가 마련되어야 한다. PDF/A-1을 보존포맷으로 선정했을 때 사용된 9개의 선정기준이 적용되었지만 해당 선정기준을 채택한 근거와 정의, 평가방식 등이 구체적으로 드러나 있지 않다.

보존포맷 선정기준과 평가방식에 대한 연구는 송치호, 차현철(2017)과 차현철, 최주호(2019) 이외에 많이 이루어지지 않았다. 차현철, 최주호(2019)는 선정기준과 평가방식을 포함한 보존포맷 선정방안을 체계적으로 제시하였다는 점에서 의의를 갖는다. 그러나 선정기준의 항목들을 선택하게 된 근거가 구체적으로 제시되지 않았으며, 다양한 전자기록물 유형별 특성을 반영할 수 있는 구조로 설계되지 않았다.

기존 연구들은 특정 유형에 대해 부분적으로 이루어졌거나, 기록관리 또는 전자·컴퓨터공학 관점 중 한쪽에 치우쳐 있어서 전자기록물 관련 전문가들이 모두 동의하지 못하는 경우가 있었다. 이는 기록관리 전문가가 전자·컴퓨터공학을 고려하기 어렵고, 반대로 전자·컴퓨터공학 전문가가 기록관리까지 고려하는 것에 한계가 있기 때문이다. 이에 본 연구는 모든 유형의 전자기록물을 대상으로 하였고, 기록관리 및 전자·컴퓨터공학 전문가들이 각각의 관점에서 집중하여 보존포맷을 선정하는 체계를 수립하였으며, 기준항목에 대한 근거부터 정의, 평가방식까지 제시하였다는 점에서 기존연구들과 차별성을 가진다.

2. 보존포맷 및 선정기준 현황

국의 아카이브 기관들은 보존포맷을 선정하기 위한 기준을 가지고 있으며 그 선정기준에 따라 다양한 보존포맷들을 채택하고 있다. 이 장에서는 국외 보존포맷 및 선정기준 현황을 살펴봄으로써 보존포맷을 어떠한 방식으로 다양화할지, 그리고 보존포맷을 선정시 어떠한 특성 및 항목들을 고려해야 할지에 대한 방향성을 확립하는데

도움을 얻을 수 있다.

2.1 보존포맷 현황

본 연구에서는 4개국(미국, 스위스, 캐나다, 호주)의 국립 아카이브 기관을 중심으로 보존포맷 현황에 대해서 조사하였다. 조사된 4개국은 보존포맷 선정 작업이 선행되어야 하는 마이그레이션을 장기보존의 주전략으로 채택한 국가이다. <표 1>과 <표 2>는 각각 4개국의 국립 아카이브 기관의 보존포맷 유형과 현황을 LAC(2010), LAC(2015), NAA(2020b), NARA(2014), NARA(2019), SFA(2018), 소정의, 한희정, 양동민(2018), 국가기록원(2018)로부터 요약, 정리하여 보여주고 있다.

<표 1> 4개국 보존포맷 유형

국가	보존포맷 전자기록물 유형
미국 (NARA)	• 텍스트(Textual Data), 프레젠테이션(Presentation Formats), 디지털 오디오(Digital Audio), 디지털 동영상(Digital Moving Images), 디지털 정지 이미지(Digital Still Images), 지리데이터(Geospatial Formats), 구조화된 데이터(Structured Data Formats), 이메일(Email), 웹기록(Web Records), 컴퓨터 지원 디자인(Computer Aided Design) 등
스위스 (SFA)	• 텍스트(Text), 이미지(Image), 오디오(Audio), 비디오(Video), 테이블·데이터베이스(Tables, Databases), 지리 데이터(Geodata), 기타(XML/XSD, Web archiving) 등
캐나다 (LAC)	• 텍스트(Text), 프레젠테이션(Presentation), 이메일(Email), 정지 이미지(Still Images), 디지털 오디오(Digital Audio), 디지털 동영상(Digital Moving Images), 특정 지역과 연관된(Geospatial), 컴퓨터 지원 디자인(Computer Aided Design), 데이터세트(Datasets) 등
호주 (NAA)	• 컴퓨터 지원 디자인(Computer Aided Design), 데이터세트(Datasets), 디지털 오디오(Digital Audio), 디지털 동영상(Digital Video), 이메일(Email-Aggregates, Email-Individual Messages), 인캡슐레이션(Encapsulation), 지리 데이터(Geospatial Formats), 이미지(Images-Raster, Images-Vector), 프레젠테이션(Presentation), 프로젝트 관리(Project Management), 원시 이미지(Raw images), 텍스트(Text), 웹사이트(Web formats) 등

먼저, <표 1>에서 볼 수 있듯이, 4개의 국립 아카이브 기관에서는 보존포맷을 다양한 유형으로 구분하고 있다. 각 기관에서 분류한 보존포맷 유형들의 개수와 명칭들은 다를 수 있지만, 국내 공공기록물법 시행령에서 언급되고 있는 전자문서, 웹기록물, 행정정보데이터세트, 시청각기록물 등의 전자기록물 유형으로 대응될 수 있다. 그리고, <표 2>는 4개국 아카이브 기관의 보존포맷 현황을 보여주고 있다. 기관마다 보존포맷 유형이 다르기 때문에 국내 공공기록물법 시행령에 명시된 전자기록물 유형에 각 기관의 보존포맷들을 대응하여 분류하였고, 지면의 한계로 각 기관의 모든 보존포맷들을 열거하지 않았다. 캐나다(LAC)의 웹기록물처럼 아직 채택되지 않은 보존포맷 유형도 있지만, 4개국 모든 아카이브 기관은 유형별로 다수의 보존포맷들을 선정하여 소속 또는 연관 기관들의 업무의 효율성과 유연성을 높이고 있다. 미국(NARA), 호주(NAA), 캐나다(LAC)의 경우에는 같은 보존포맷 유형에서 선호(Preferred), 허용(Acceptable) 등으로 보존포맷의 위험 수준을 구분하기도 한다.

4개국 국립 아카이브 기관의 보존포맷 현황을 살펴본 결과, 보존포맷은 다양한 유형으로 분류되어 있었다. 또한, 유형별로 다수의 보존포맷들을 선정하여 전자기록물을 생산, 관리, 보존할 때 활용 도구 및 매체의 범위를 확대하여 업무의 효율성을 고려함을 알 수 있었다.

<표 2> 4개국 보존포맷 현황

구분	미국(NARA) ⁴⁾	스위스(SFA) ⁵⁾	캐나다(LAC) ⁶⁾	호주(NAA) ⁷⁾
전자문서	• TXT, DOC(X), PDF/A-1/-2, ODT, PPT(X), XLS(X) 등	• TXT, PDF/A-1/-2, DOC, XML 등	• TXT, EPUB, ODF, PDF/A-1, PDF/A-2, ODP(X), PDF DOC(X) 등	• ODT, TXT, PDF/A-1/-2, DOC(X), ODP, PPT(X), XSL(X) 등
시청각 기록물	• TIFF, PNG, GIF, FLAC, MP3, WAVE, AVI, WMV, JP2, MPEG-2/-4 등	• TIFF, JPEG-2000, WAVE, MPEG-4, FFV1 등	• TIFF, PNG, GIF, BWF, WAV, AAC, JPEG, AVI, WMV, MOV, MPEG-2/-4, 등	• PDF/A-1/-2, PNG, GIF, JPEG, TIFF, BWF, FLAC, MPEG-1/-2/-4, WAV 등
행정정보 데이터세트	• JSON, XML, CSV, XSL(X) 등	• CSV, SIARD	• CSV, DBF, XML, ODS 등	• CSV, ODS, XML, XLS(X), SIARD, JSON 등
웹기록물	• (W)ARC	• PDF/A, SIARD, XML	-	• (W)ARC, (X)HTML, ASP(X)

2.2 보존포맷 선정기준 현황

보존포맷 선정기준(Selection Criteria)은 위험요소(Risk Matrix), 지속가능성 요소(Sustainability Factor or Criteria), 연관 요구사항(Binding Requirement) 등 다양한 형태로 논의되고 있다. 앞서 보존포맷 현황을 조사한 4개국 국립 아카이브 기관 중에서는 미국과 캐나다에서만 선정기준 관련 내용을 찾을 수 있어서 국내·외 아카이브, 도서관, 프로젝트, 연구결과 등에서 제시한 내용을 토대로 전자기록 보존포맷 기준에 대한 현황을 조사하였다. 기본적으로 보존포맷 선정기준은 기준 항목들과 각 항목들에 대한 정의로 이루어져 있다. 다수의 기준 항목들이 존재하고, 각 기준 항목들에 대한 용어와 정의들이 다양하게 존재하며 같은 용어라도 다르게 정의되어 있거나, 다른 용어이지만 같은 의미를 지닌 경우도 존재한다.

본 연구에서는 기관 및 연구자에 따라 다르게 사용하고 있는 보존포맷 기준에 대한 용어, 정의 및 내용들을 조사하고 정리하였다. 결과적으로 <표 3>에서처럼 7개의 대분류 수준의 선정기준 항목으로 구분할 수 있으며, 각 기준항목을 채택하고 있는 기관들과 해당 항목들을 정의 및 설명하고 있는 다양한 출처들을 정리하였다. 선정기준 항목들은 훨씬 더 다양하지만 비슷하거나 유사한 항목들은 제외하였다. 각 선정기준 항목들은 유사하지만 조금씩 다르게 정의되어 있었다. 여러 기관들은 이러한 선정기준 항목들을 채택하고 각 항목들을 보존포맷 선정할 때 적용하고 있는 것으로 조사되었다. 기관의 특성 및 정책에 따라 각 선정기준 항목 하위에 다수의 세부기준들을 두거나 여러 개의 선정기준을 하나로 병합하기도 하였으며, 새로운 항목을 추가하거나 제외하는 경우도 있다.

4) NARA(National Archives And Records)는 미국 국립문서기록관리청

5) SFA(Swiss Federal Archives)는 스위스 연방기록보관소

6) LAC(Library and Archives Canada)는 캐나다 국립도서관기록관

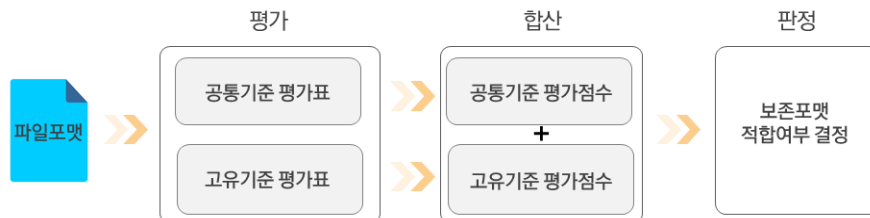
7) NAA(Nation Archives Australia)는 호주 국립기록보관소

<표 3> 보존포맷 선정기준 현황

항목(대분류)	적용 기관	출처
진본성 (Authenticity)	NAK, MSA	Abrams et al., 2005; Arms & Fleischhauer, 2006; Barnes, 2006; CENDI, 2007; Clausen, 2004; Park & Oh, 2012; Folk & Barkstrom, 2003; Frey, 2000; Hodge & Anderson, 2007; InterPARES 2; Lesk, 1995; Todd, 2009; Rog & Wijk, 2008; Sullivan, 2006; Wijk & Rog, 2007
개방성 (Openness)	LAC, TNA, NARA, NAK, LOC ⁸⁾ , WHS ⁹⁾	Brown, 2003; ECMA, 2008; Park & Oh, 2012; Folk & Barkstrom, 2003; Frey, 2000; Hamm & Becker, 2011; Todd, 2009; Puglia et al., 2004
상호운용성 (Interoperability)	LAC, TNA, NAK, LOC, WHS	Abrams et al., 2005; Arms & Fleischhauer, 2006; Brown, 2003; CENDI, 2007; Clausen, 2004; Frey, 2000; Park & Oh, 2012; Hodge & Anderson, 2007; Müller et al., 2003; InterPARES 2; Todd, 2009; Puglia et al., 2004; Rog & Wijk, 2008; Sullivan, 2006; Rog & Wijk, 2007; Moreq2010, 2011
독립성 (Independence)	LAC, TNA, NAK, MSA ¹⁰⁾	Abrams et al., 2005; Arms & Fleischhauer, 2006; Barnes, 2006; Brown, 2003; CENDI, 2007; Park & Oh, 2012; Folk & Barkstrom, 2003; Hodge & Anderson, 2007; Johnson, 1999; Lesk, 1995; Todd, 2009; Müller et al., 2003; Puglia et al., 2004; Rog & Wijk, 2008; Sullivan, 2006; Wijk & Rog, 2007
자체문서화 (Self-Documentation)	TNA, NARA, NAK, LOC, WHS	Brown, 2003; Park & Oh, 2012; Folk & Barkstrom, 2003
채택 (Adoption)	LAC, TNA, NARA, NAK, LOC, WHS	Abrams et al., 2005; Arms & Fleischhauer, 2006; Brown, 2003; CENDI, 2007; Park & Oh, 2012; Folk & Barkstrom, 2003; Hodge & Anderson, 2007; InterPARES 2; Todd, 2009; Puglia et al., 2004; Rog & Wijk, 2008; Sullivan, 2006; Wijk & Rog, 2007
기능성 (Functionality)	NARA, NAK, LOC, MSA, WHS	Abrams et al. 2005; Anderson, 2007; Arms & Fleischhauer, 2006; Brown, 2003; CENDI, 2007; Park & Oh, 2012; Folk & Barkstrom, 2003; Frey, 2000; Hodge & Puglia et al., 2004; InterPARES 2; Puglia et al., 2004; Rog & Wijk, 2008; Sullivan, 2006; Wijk & Rog, 2007

3. 보존포맷 선정체계

본 연구에서 제안하는 보존포맷 선정체계는 선정기준과 평가방식으로 구성되어 있다. 이때, 선정기준은 모든 파일포맷에 공통적으로 적용되는 공통기준과 전자기록물 유형에 따라 다르게 적용되는 고유기준이 있다. 보존포맷 선정체계에 보존포맷 선정 여부를 알기 위한 파일포맷을 입력하면, 공통기준과 고유기준을 기반으로 각각의 평가표로 평가를 실시한다. 공통기준과 고유기준의 점수를 합산한 최종 평가점수에 따라 보존포맷 선정 여부를 알 수 있다(<그림 2> 참고).



<그림 2> 보존포맷 선정체계를 통한 보존포맷 선정 과정

8) LOC(Library of Congress)는 미국 의회도서관
 9) WHS(Washington Headquarters Services)는 워싱턴 본부관리국
 10) MSA(Minnesota State Archives) 미네소타 주립기록원

전자기록물은 보존하고 표현하고자 하는 내용과 형태, 즉 기록물 유형별로 적합한 파일포맷에 저장된다. 예를 들어, 문서 유형은 아래아 한글 또는 MS 오피스 파일에 저장하고, 시청각기록물은 TIFF(이미지), WAV(음성), MP4(영상) 파일에 저장된다. 한편, 전자기록물의 명칭에서 알 수 있듯이 기록물은 전자 및 컴퓨터 관련 기술로 설계된 파일에 저장되는데, 해당 파일은 파일포맷과 상관없이 동일한 원리로 컴퓨터에 의해서 생성·저장·구동된다.

정리하면, 기록물의 유형에 따라 보존 및 표현하고자 하는 콘텐츠는 다르지만, 기록물이 담겨진 파일이 SW와 HW에 의해서 생성·저장·구동되는 기본 원리는 동일하다. 그래서 보존포맷 선정기준에는, 기록물의 유형에 따라 다른 필수 속성을 보존할 수 있는가의 여부를 판단할 수 있는 고유한 선정기준도 필요하지만, 파일포맷이 장기보존에 적합한 기술로 설계되어 있는지의 여부를 기록 유형에 상관없이 공통적으로 판단할 수 있는 선정기준도 필요하다.

본 연구에서 제시하는 보존포맷 선정체계의 선정기준은 기록물 유형에 상관없이 적용되는 공통기준과 기록물 유형마다 다른 고유기준으로 이루어져 있다. 공통기준은 기록물 관점이 아닌 파일포맷 자체에 대한 관점으로 바라볼 수 있어야 한다. 예를 들어, 파일포맷의 표준이 있는지, 그 표준에 쉽게 접근이 가능한지, 파일포맷 구동시 특정 HW나 SW에 영향을 받는지, 파일포맷을 구동할 수 있는 SW를 쉽게 구할 수 있는지 등의 여부를 반영할 수 있는 기준이어야 한다. 고유기준에서 가정하는 기록물 유형은 국가 또는 기관마다 다를 수 있다. 국내 공공기관의 경우는 공공기록물법 시행령에 명시되어 있는 기록물 유형으로 구분될 수 있다. 그래서 전자문서, 행정정보데이터세트, 시청각기록물, 웹기록물의 유형마다 고유기준의 도출이 필요하다.

<표 3>의 7개 선정기준 항목 중에서 진본성(Authenticity)은 기록물이 지닌 고유한 속성을 훼손하지 않고 보존할 수 있는지 여부를 판단하는 기준으로 기록물 보존에 있어서 가장 중요한 특성이자 기준이라고 할 수 있다. 진본성을 보장하기 위해서는 기록물이 담고 있는 내용(Content), 맥락(Context), 구조(Structure)에 해당되는 특성들이 보존되어야 한다. 내용, 맥락, 구조는 기록물 유형마다 다르다. 예를 들어, 전자문서 유형의 경우는 문자(내용), 제목·생산자·생산일자·메타데이터(맥락), 문자인코딩·스키마(구조) 등이 진본성을 유지하기 위한 특성이며, 행정정보데이터세트 유형의 경우는 테이블데이터(내용), 제목·생산자·생산일자·메타데이터(맥락), 테이블구조·관계(구조) 등이 있다. 전자기록물의 진본성 보장에 관한 연구는 Gareth Knight에 의해서 체계가 만들어졌으며, 보존되어야 할 특성을 전자기록물 유형별로 도출하는 연구가 지금까지 진행되고 있다. 그러므로 전자기록물 유형별로 고려되어야 하는 진본성은 고유기준의 영역에 속한다고 할 수 있다. 그래서 보존포맷 선정기준에서 고유기준을 진본성으로 한다.

나머지 6개 기준들은 기록물 유형별로 보존해야 할 속성과 직접적으로 관련 없는 파일포맷 자체에 대해서 고려해야 할 항목이다. 파일포맷이 얼마나 체계적으로 만들어 졌는지, 얼마나 다양한 환경에서 확인이 가능한지, 얼마나 많이 사용되고 있는지를 판단한다. 이러한 항목들은 전자기록물의 유형에 관계없이 모든 파일포맷에 공통적으로 고려되어야 할 공통기준이다. 공통기준은 4장에서 구체적으로 다루기로 한다.

Gareth Knight(2008)에서는 전자기록물 보존을 위해 해결되어야 할 근본적인 과제는 보존기간과 변화하는 기술환경 전반에 걸쳐 디지털 객체의 접근 가능성과 신뢰성 및 기록의 4대 요건의 보장이며, 이를 위해서는 기록에서 필수적으로 보존되어야 할 속성과 디지털 객체의 물리적 환경과의 분리가 필요하다고 하였다. 디지털 기록은 진본성이 보장된 상태로 보존해야 하는 기록이 담고 있는 고유의 특성을 필수보존속성(Significant Property)의 항목에 따라 특징을 도출하여 고유기준을 만들 수 있다. <표 4>에서 볼 수 있듯이 필수보존속성은 총 5개의 항목으로 이루어져 있으며, 기록물 유형별로 5개 항목에 대한 특징들을 도출할 수 있다. 예를 들어, 전자문서 유형의 경우 Rendering 속성으로는 글꼴, 색상 및 크기, 비트심도, Behavior는 하이퍼링크, Content는 기간, 문자수, Context는 작성자 이름, 작성 날짜, Structure는 첨부파일 등을 필수보존속성으로 들 수 있다. 고유기준은 전자기록물 유형에 따라 제안되어야 하므로, 향후 유형별로 보존되어야 할 필수적인 속성을 심도있게 조사하고 필수보존속성을 도출하는 연구가 지속적으로 이어져야 한다.

〈표 4〉 필수보존속성(Significant Property) 항목 및 정의

항목	정의
Rendering	• 기록물의 모습, 기능, 성능을 표현 또는 재생성하기 위한 정보 예) 폰트, 색상, 크기 등
Behavior	• 객체가 외부 자극과 상호 작용하는 방식을 나타내는 정보 예) 하이퍼링크 등
Content	• 기록에 담긴 내용으로 정보객체(Information Object)에 포함된 정보 그 자체 예) 텍스트, 이미지, 음성, 영상 등
Context	• Content가 생성된 환경 또는 Content에 포함되어 전달하고자 하는 의미와 관련된 주변 환경을 설명하는 전자기록에 포함된 모든 정보 예) 저자, 생성날짜 등
Structure	• 기능을 재구성하는데 필요한 두 가지 이상의 유형 Content 간의 외인 또는 내인적 관계를 설명하는 정보 예) 이메일과 첨부파일

출처: (소정의, 2019) 참고 및 재구성

본 연구에서 제안하는 보존포맷 선정여부 평가방식은 Yes/No로 문항으로 이루어진 평가표로 수행되며 절대평가로 진행된다. 공통기준 평가표를 통해 계산된 평가점수와 고유기준 평가표를 나온 평가점수를 합산된 최종 평가점수에 의해서 보존포맷 여부가 결정된다.

4. 보존포맷 공통기준 및 평가방식

4.1 보존포맷 공통기준

앞서 정리했던 <표 3>의 7개 선정기준을 <표 5>와 같이 5개의 보존포맷 공통기준으로 정리하였다. 진본성은 고유기준에서 다루고 독립성은 상호운용성의 세부기준으로 포함하였다. 상호운용성은 하나의 시스템에서 생성된 기록물을 다른 시스템에서 확인할 수 있는 특성이다. 이를 위해서는 컴퓨터를 구성하고 있는 SW 및 HW의 독립성이 선행되어야 한다. 독립성은 상호운용성의 필요조건이기 때문에 상호운용성의 세부기준으로 두었다.

전자기록물의 내용을 확인하기 위해서는 파일과 해당 파일을 구동할 수 있는 컴퓨팅 환경이 필요하다. 그래서 첫 번째와 두 번째 보존포맷 공통기준은 각각 파일포맷과 이를 구동하는 컴퓨팅 환경과 관계가 깊다. 파일에 원하는 내용을 담고, 공유하여 확인하기 위해서는 내용을 구성하는 요소들의 비트스트림 위치, 순서, 크기 등의 파일포맷이 정해져야 한다. 이러한 파일포맷들은 개인에서부터 기업, 표준화 기관 등에 의해서 표준화된다. 표준화 과정이 체계적으로 이루어지는 파일포맷은 대부분 오픈포맷 정책을 사용하며 최근에는 오픈포맷을 구동하는 오픈소스 프로젝트가 병행하여 진행되고 있다. 표준을 만드는 표준화 과정이 투명하고 개방적으로 이루어질수록, 해당 파일포맷을 구동할 수 있는 오픈소스 프로젝트가 추진될수록 보존포맷에 적합하다고 할 수 있다. 파일포맷과 관련된 표준화 과정과 오픈포맷 정책 그리고 오픈소스 제공은 파일포맷의 기술적인 측면이 아니라 정책적인 측면과 관련이 있으며, 파일포맷을 중심으로 정책적으로 개방성을 지향할수록 보존포맷에 적합하다고 판단할 수 있다. 그래서 파일포맷의 정책적인 측면에 초점을 맞춰 보존포맷 여부를 판단하는 것이 가장 중요하다고 생각하여 개방성(Openness)를 첫 번째 선정기준으로 채택하였다.

〈표 5〉 보존포맷 공통기준 및 정의

기준	정의·설명·세부기준	
개방성 (Openness)	정의	• 특정 기관에 독점적이지 않고 공개적으로 사용 및 접근할 수 있는지 여부에 관한 기준
	설명	• 파일포맷의 '표준(Standard)'과 '오픈소스(Open Source)'의 존재와 파일포맷과 구동SW 소유권의 독점 여부로 판단
	세부기준	(1) 공표(Disclosure), (2) 공개가용성(Open Availability)
상호운용성 (Interoperability)	정의	• 하나의 시스템에서 생산된 기록물이 독점적 또는 상업용 HW와 SW와 상관없이 다른 시스템에서 확인·동작할 수 있는지 여부에 관한 기준
	설명	• 파일포맷과 외부의 다양한 요인(HW, OS, 구동SW 등)과의 독립성, 호환성 및 변환가능성으로 판단
	세부기준	(3) 독립성(Independence), (4) 호환성(Compatibility), (5) 변환가능성(Convertibility)
자체문서화 (Self-Documentation)	정의	• 다양하고 풍부한 설명 및 기능 메타데이터의 파일포맷 내재화할 수 있는지 여부에 관한 기준
	설명	• 파일포맷에서 자체 메타데이터 기능 지원 여부로 판단
	세부기준	(6) 메타데이터(Metadata)
채택 (Adoption)	정의	• 정보 작성자, 유포자 또는 사용자 등에 의해 해당 파일포맷을 사용하고 있는지 여부에 관한 기준
	설명	• 파일포맷의 사용 범위로 판단
	세부기준	(7) 편재성(Ubiquity), (8) 편중성(Sporadicness)
기능성 (Functionality)	정의	• 암호화, 오류감지, 보호메커니즘, 검색기능 등 기능의 제공하는지 여부에 관한 기준
	설명	• 파일포맷이 지원하는 다양한 기능으로 판단
	세부기준	(9) 보호메커니즘(Protection)

파일에 담겨 있는 내용을 확인하기 위해서는 HW와 OS가 설치된 컴퓨터에서 해당 파일포맷을 구동할 수 있는 SW를 설치하고 실행해야 한다. 파일은 그 자체만으로 확인할 수 없고, 컴퓨터와 구동SW를 포함한 컴퓨팅 환경에 절대적인 의존성을 갖고 있다. 구동SW가 다양한 OS에서 동작할수록, HW의 의존성이 없을수록, 구동 SW 제작·배포 기관이 해당 파일포맷에 대해 수립한 장기적인 마스터플랜(Master Plan)을 보유할수록 해당 파일포맷은 보존포맷에 적합하다고 할 수 있다. 파일포맷을 구동하는 컴퓨팅 환경의 관계를 통해서 의존성을 조사하는 것은 기술적인 측면과 관련이 있으며, 컴퓨팅 환경에 의존성이 없을수록 보존포맷에 적합하다고 판단할 수 있다. 그러므로 두 번째 보존포맷 공통기준으로 파일을 실행하기 위한 컴퓨팅 환경과의 의존성 및 연관성의 기술적인 측면에 초점을 맞춰서 보존포맷 적합성을 판단하기 위해 상호운용성(Interoperability)을 두 번째 선정기준으로 채택하였다.

메타데이터는 기록의 4대 속성과 아주 밀접한 관련이 있다. 이러한 메타데이터를 별도의 파일로도 관리할 수 있지만 물리적으로 하나의 개체로 존재하도록 파일포맷 자체에 포함하는 기능이 지원된다면 장기보존에 보다 효과적이다. 파일포맷에서 지원가능한 메타데이터의 내용이 많을수록, 파일포맷을 구동할 수 있는 기술적인 내용이 많이 포함될수록, 자동으로 생성되는 부분이 많을수록, 사용자 정의를 통해 추가할 수 있는 정보가 많을수록 보존포맷에 적합하다고 판단된다. 그래서 메타데이터(Metadata)를 공통기준에 포함하였다.

개방성의 보존포맷 선정기준에 아주 적합하게 파일포맷의 표준화 과정이 잘 이루어졌어도 해당 파일을 사용하는 사람들이 거의 없다면 해당 파일포맷은 도태되어 사람들의 기억 속에서 없어지게 된다. 지금은 사용되지 않은 수많은 표준들이 존재하는 것이 이를 증명한다. 예를 들어, 지금은 LTE 서비스로 전환되었지만 2018년 9월 서비스가 종료된 WiBro(IEEE 802.16e) 기반의 에그 서비스 그리고 마이크로소프트에서 표준화했지만 지금은 많이 사용되지 않고 있는 멀티미디어 파일포맷인 asf 등이 있다. 그러므로 보존포맷의 적합성 여부는 해당 파일포맷을 얼마나 많이 사용하고 있는지와 직결되어 있다. 사용하는 사람들이 많을수록, 사람들의 범위가 넓은 지역에 걸쳐 있을수록, 시장 점유율이 높을수록, 구동SW를 별도로 설치할 필요없이 파일포맷을 확인¹¹⁾할 수 있을수록 보존포맷에 적합하다고 할 수 있다.

그래서 채택(Adoption)을 공통기준으로 선정하였다. 채택 항목과 관련하여 예외적인 상황이 존재할 수 있다. 만약, 특정 국가 또는 특정 기관에서 특수한 목적으로 만들어져서 사용되는 파일포맷인 경우는 극히 일부에서만 사용되지만 장기보존해야 할 가치가 있는 기록일 수 있다. 이러한 경우로 판단되면 다른 4가지 공통기준에 의한 평가와 상관없이 보존포맷으로 선정되어야 한다. 다만 예외적인 보존포맷으로 별도로 관리하는 것이 바람직하다.

마지막 공통기준은 기능성(Functionality)으로, 파일포맷이 장기보존을 위한 유용한 기능들을 제공하는지를 판단하는 항목이다. 복사방지, 디지털서명, 인쇄방지, 콘텐츠 추출 보호 등과 같은 보호메카니즘(Protection)이 적용되었는지, 문서 내용에 대한 검색기능(Retrievability)을 지원하는 파일포맷인지를 판단할 수 있는 항목을 세부기준으로 두었다. 이와 같이 5개 상위 공통기준을 채택하였고, 각 항목별로 구체화하여 총 9개의 하위 세부기준을 마련하였다.

공표(Disclosure)

공표는 파일포맷에 대한 ‘표준(Standard)’이 존재하는지 여부와 ‘오픈소스(Open Source)’가 존재하는지에 대한 여부를 통해 보존포맷 적합성을 판단하는 기준이다. 아래 4개의 항목에 대해 조사함으로써 확인할 수 있다.

- (1) 표준을 다운로드나 열람을 통해서 공개적으로 쉽게 참조하거나 이용하는 것이 가능한지 여부
- (2) 체계적이고 권위있는 기관에 의해 표준화 과정이 이루어졌는지 여부
- (3) 파일포맷을 구동할 수 있는 오픈소스가 존재하는지 여부
- (4) 오픈소스 라이선스 형태

공개가용성(Open Availability)

공개가용성은 해당 파일포맷 및 구동 SW가 공개되어 있는지 또는 특정 기관 또는 기업에 ‘독점’되어 있는지 여부를 통해 보존포맷 적합성을 판단하는 기준이다. 아래 4개의 항목을 조사하여 공개가용성을 확인할 수 있다.

- (1) 해당 포맷 및 구동 SW가 특정 기업에 ‘독점’ 되어 있는지 여부
- (2) 파일포맷을 구동시킬 수 있는 다른 SW 존재 여부
- (3) 파일포맷 사용에 대한 제한여부(라이선스, 구독, 특허료 등)
- (4) OS에 내재되어 있는 기본 도구(메모장, 그림판 등) 로 분석이 가능한지 여부

독립성(Independence)

독립성은 파일포맷을 구동할 때 OS, HW, 구동SW에 의해서 영향받는 정도를 통해 보존포맷 적합성을 판단하는 기준이다. 아래 3개의 항목에 대해 조사함으로써 독립성을 확인할 수 있다.

- (1) 파일포맷을 구동할 때 특정 OS 또는 제한적인 OS에서만 동작하는지 여부
- (2) 파일포맷을 구동할 때 특수한 HW에서만 구동되거나 별도의 HW가 필요한지 여부
- (3) 파일포맷 또는 구동SW가 특정기술, 표준, 별도의 SW 등이 존재해야만 구동이 가능한지 여부

호환성(Compatibility)

호환성은 파일포맷에 대해 구동SW 제조사가 가지고 있는 호환성 계획 여부 및 구체성을 통해 보존포맷 적합성

11) 보편화된 파일포맷인 경우 OS 또는 웹브라우저에서 해당 파일을 확인할 수 있는 기능을 내재화한다. jpg, gif 등 이미지 파일포맷과 pdf 등이 대표적이다. 구동SW 설치없이 확인가능한 파일포맷은 편재성 관점에서 보존포맷에 적합하다고 할 수 있다.

을 판단하는 기준이다. 여기서 제조사는 현재 가장 대표성 있는 구동SW를 제작하는 기업 또는 기관이다. 아래 2개 항목을 조사하면 호환성을 확인할 수 있다.

- (1) 구동SW의 Release 주기(공개주기)에 따른 형식 및 사양의 업데이트 여부
- (2) 파일포맷의 버전 관련 개발 로드맵 또는 계획의 존재 여부

변환가능성(Convertibility)

변환가능성은 장기보존을 위해 새로운 보존포맷으로 안정적으로 변환 가능 여부 및 활용을 목적으로 다른 파일포맷으로의 변환 가능 여부를 통해 보존포맷 적합성을 판단하는 기준이다. 아래의 2개 항목을 조사하여 변환가능성을 확인할 수 있다.

- (1) 정보 손실없이 다른 포맷으로 변환 가능 여부 및 변환 가능한 포맷의 다양성
- (2) 정보 손실이 발생할 수도 있는 다양한 활용 목적으로 SW, 서비스, 도구 등에 의해서 다른 포맷으로 변환되고 재사용 가능 여부

메타데이터(Metadata)

메타데이터는 파일포맷 자체에서 파일 렌더링을 위한 정보나 기록관리 측면에서 부가적인 정보의 내재화가 가능한지 여부를 조사함으로써 보존포맷의 적합성을 판단할 수 있는 기준이다. 아래의 4개 항목을 조사하여 메타데이터 관련 속성을 알 수 있다.

- (1) 파일포맷을 구동하기 위한 렌더링 정보(버전, 표준 등)를 제공하는 메타데이터 제공 여부
- (2) 사용자 정의 메타데이터 지원이 가능한지 여부
- (3) 외부 표준 메타데이터(예: JPG, TIFF의 EXIF: Exchangable Image File) 지원이 가능한지 여부
- (4) 파일포맷으로부터 쉽게 메타데이터 추출이 가능한지 여부

편재성(Ubiquity)

편재성은 해당 포맷에 대한 수요와 공급의 법칙이 잘 확립되어 있고 광범위하게 사용되는 포맷인지의 여부에 대한 조사를 통해 보존포맷의 적합성을 판단할 수 있는 기준이다. 아래의 5개 항목을 조사하여 편재성을 알 수 있다.

- (1) OS에서 별도의 응용 SW의 설치 없이 해당 포맷을 인식하고 내용 확인이 가능한 지 여부
- (2) 웹브라우저(Microsoft Edge, Internet Explorer, Chrome, Firefox, Safari, Swing, Whale 등)에서 별도 SW 설치 없이 해당 파일포맷을 인식하고 내용을 확인할 수 있는지 여부
- (3) 파일포맷이 표준화 단체에 의해 표준화 과정을 거쳐 저명한 컨소시엄과 그룹(ISO/IEC, ITU, W3C, OASIS 등)에 의해 채택되어 전 세계에서 사용하고 있는 지 여부
- (4) 시장을 선도하는 파일포맷인지 여부
- (5) 파일포맷을 제작·조작·렌더링 할 수 있는 많은 경쟁 제품의 존재 여부

편중성(Sporadicalness)

편중성은 특정 국가 또는 기관에서 특수한 목적으로 예외적으로 만들어진 파일포맷에 대해 보존포맷으로의 적합성을 판단하는 기준이다. 아래의 2개 항목을 조사하여 편중성을 알 수 있다. 만약 편중성이 있는 파일포맷으

로 평가되는 경우 다른 공통기준의 평가와 상관없이 보존포맷으로 채택될 수 있다.

- (1) 특정 전문 분야의 커뮤니티에서 채택되어 사용되고 있는 파일포맷인지 여부
- (2) 국립도서관, 기록원 및 기타 기록유산기관이 공식적으로 채택한 보존포맷인지의 여부

보호메커니즘(Protection)

일반적으로 OS에서 쓰기 권한이 부여된 사용자는 대부분 파일포맷을 사용할 수 있지만 최근 파일포맷 자체에서 파일을 수정·삭제할 수 있는 권한을 제어하거나 훼손된 데이터를 복구할 수 있는 보호메커니즘을 제공한다. 아래 3개 항목을 조사하면 보호메커니즘 기능에 대한 제공 수준을 파악할 수 있다.

- (1) 복사 방지, 디지털 서명, 인쇄 방지 및 콘텐츠 추출 보호와 같은 보호메커니즘이 적용되는 포맷인지의 여부
- (2) 오류 감지, 수정 메커니즘 및 암호화 옵션의 수용 여부
- (3) 우발적인 데이터 손상에 대한 탄력성 여부

4.2 평가방식

보존포맷 선정을 위한 평가 방식은 절대평가 방식이며 다음과 같이 4단계로 진행된다.

- (1단계) 공통기준 평가 : 보존포맷 선정을 위한 <표 6>의 공통기준 평가표를 적용하여 전자기록 보존포맷으로서 적합성 평가
- (2단계) 고유기준 평가 : 특정 유형의 보존포맷 선정을 위한 고유기준 평가표를 적용하여 특정유형 전자기록 보존포맷으로서 적합성 평가
- (3단계) 1·2단계 평가점수를 합산한 후 평점으로 환산(100%)하여 등급을 부여한 다음 최종 보존포맷으로서 적합성 평가
- (4단계) 부적합으로 평가된 보존포맷의 경우 예외 규정(편중성)의 적용여부 판단해 최종 결정

<표 6> 공통기준 평가표

공통기준	세부기준	평가항목	Y/N (1점)	
개방성 (Openness)	1. 공개가용성 (Open Availability)	1.1 특정 기업 외 파일포맷을 구동시킬 수 있는 다른 SW가 있는가?	Y/N (1점)	
		1.2 파일포맷 사용에 대한 제한여부(라이선스, 구독, 특허료 등)	1.2.1 무료 Read인가?	Y/N (1점)
			1.2.2 무료 Write인가?	Y/N (1점)
		1.3 기본 도구(메모장, 그림판 등) 사용을 통한 분석 가능 여부	1.3.1 기본 도구를 통해 해당 포맷을 구성하는 콘텐츠 전체를 해석할 수 있는가?	Y/N (1점)
			1.3.2 텍스트 콘텐츠가 표준 문자 인코딩(UTF-8, 유니코드, 아스키 코드 등)으로 되어 있는가?	Y/N (1점)
			1.3.3 압축되어 있는 경우 신뢰성 있는 압축(zip, gzip, lzw 등)으로 되어 있는가?	Y/N (1점)
			1.3.4 멀티미디어 콘텐츠가 공개 포맷(jpeg, gif, mpeg 등)으로 되어 있는가?	Y/N (1점)

공통기준	세부기준	평가항목		Y/N
개방성 (Openness)	2. 공표 (Disclosure)	2.1 '표준' 존재 여부	2.1.1 파일포맷의 표준을 인터넷 등을 통해 공개적으로 참조 및 이용이 가능한가?	Y/N (1점)
			2.1.2 파일포맷의 표준을 인터넷 등을 통해 공개적으로 참조 및 이용할 때 무료인가?	Y/N (1점)
			2.1.3 체계적이고 권위있는 기관에 의해 표준화 과정을 거쳤는가?	Y/N (1점)
	2.2 '공개코드' 존재 여부	2.2.1 파일포맷이 오픈소스 라이선스인가?	Y/N (1점)	
상호운용성 (Interoperability)	3. 독립성 (Independence)	3.1 OS 관점	3.1.1 파일포맷을 구동할 수 있는 OS의 개수가 다수인가?	Y/N (1점)
		3.2 HW 관점	3.2.1 특별한 HW없이 구동할 수 있는가?	Y/N (1점)
			3.2.2 파일포맷을 개인용 컴퓨터 수준의 HW에서 구동할 수 있는가?	Y/N (1점)
	3.3 특정 기술, 표준, 부가 SW	3.3.1 파일포맷 또는 구동 SW에 특수 코덱 및 특수 플레이어와 같은 특정 기술이나 부가 SW 등의 영향이 없는가?	Y/N (1점)	
	4. 호환성 (Compatibility)	4.1 파일포맷이 현재 구동 SW에서 지원되는가? (동일한 SW(같은 제조사, 계열사, 인수회사 등)에 한함)	Y/N (1점)	
		4.2 파일포맷이 이전/이후 구동SW 버전과 호환이 가능한가? (동일한 SW(같은 제조사, 계열사, 인수회사 등)에 한함)	Y/N (1점)	
4.3 파일포맷은 구동하는 SW의 Release 주기(공개 주기)에 따라 형식이나 사양이 자주 업데이트되는가? (현재 가장 대표성 있는 구동 SW)		Y/N (1점)		
4.4 파일포맷의 버전 업데이트 개발 로드맵 또는 계획이 존재하는가?		Y/N (1점)		
5. 변환가능성 (Convertibility)	5.1 보존, 추후 안정적인 마 이그레이션 보장 가능성	5.1.1 파일포맷이 정보의 손실없이 다른 포맷으로 변환 가능한가? 5.1.2 변환 가능한 포맷이 다양한가?	Y/N (1점)	
	5.2 활용하기 쉬운 포맷으로 변환가능 여부	5.2.1 SW, 서비스 및 톨과 상호운용되어 새로운 목적으로 콘텐츠 조작 및 재사용이 가능한가?	Y/N (1점)	
자체문서화 (Self - Documentation)	6. 메타데이터 (Metadata)	6.1 파일포맷이 자동 생성 메타데이터 기능을 제공하는가?	Y/N (1점)	
		6.2 파일포맷이 사용자 지정 메타데이터 기능을 제공하는가?	Y/N (1점)	
		6.3 파일포맷으로부터 메타데이터를 추출할 수 있는 기능을 지원하는가?	Y/N (1점)	
채택 (Adoption)	7. 편재성 (Ubiquity)	7.1 OS에서 별도의 응용 SW 설치 없이 파일포맷을 인식하고 내용을 확인할 수 있는가?	Y/N (1점)	
		7.2 브라우저(Microsoft Edge, Internet Explorer, Chrome, Firefox 등)에서 별도의 확장 응용 SW 설치 없이 파일포맷을 인식하고 내용을 확인할 수 있는가?	Y/N (1점)	
		7.3 표준화 단체에 의해 표준화를 거쳐 저명한 컨소시엄 또는 그룹에 의해 채택되어 세계적으로 사용하는가?	Y/N (1점)	
		7.4 파일포맷이 시장을 선도하는가?	Y/N (1점)	
		7.5 파일포맷을 제작/조작/렌더링하는 많은 경쟁 제품의 존재하는가?	Y/N (1점)	
	8. 편중성 (Sporadicalmes)	8.1 파일포맷이 국립도서관, 기록원 및 기타 기록유산기관이 공식적으로 채택한 보존포맷인가? 8.2 파일포맷이 특정 전문 분야의 커뮤니티에서 채택되어 사용되고 있는가?	Y/N (1점)	
기능성 (Functionality)	9. 보호메커니즘 (Protection)	9.1 파일포맷이 암호 보호, 복사 방지, 디지털 서명, 인쇄 방지 및 콘텐츠 추출 보호와 같은 기술보호메커니즘이 적용되어 있지 않은가?	Y/N (1점)	
		9.2 파일포맷이 오류 감지, 수정 메커니즘 및 암호화 옵션을 수용하는가?	Y/N (1점)	
		9.3 파일포맷이 우발적인 손상에 대한 탄력성이 있는가?	Y/N (1점)	

1~2단계에서 도출된 보존포맷 평가점수에 따라 점수가 90% 이상인 경우는 'A등급(매우 우수)', 80% 이상인 경우는 'B등급(우수)', 70% 이상인 경우는 'C등급(양호)', 60% 이상인 경우는 'D등급(보통)', 60% 미만인 경우는 'E등급(미흡)'으로 정할 수 있다(<표 7> 참조).

<표 7> 보존포맷 등급 및 평점 기준

등급	평점(환산점수)	수준정의
A (매우 우수)	90 이상	<ul style="list-style-type: none"> 매우 높은 수준의 안정적인 전자기록 보존포맷 보존포맷 적합성 : 적합 - 10년마다 재평가 실시하여 등급 재설정
B (우수)	80 이상 (80이상~90미만)	<ul style="list-style-type: none"> 높은 수준의 보존포맷이지만 정기적인 평가 필요 보존포맷 적합성 : 적합 - 5년마다 재평가 실시하여 등급 재설정
C (양호)	70 이상 (70이상~80미만)	<ul style="list-style-type: none"> 전자기록 보존포맷으로 선정하기에는 다소 미흡한 부분이 있으므로 보존포맷 선정 여부는 상대평가로 결정 보존포맷 적합성 : 부분적합 - B 등급 이상의 보존포맷이 없거나 적은 경우 채택 - 3년마다 재평가 실시하여 등급 재설정
D (보통)	60 이상 (60이상~70미만)	<ul style="list-style-type: none"> 전자기록 보존포맷으로 선정하기에는 상당히 미흡한 부분이 있으므로 보존포맷 선정 여부는 상대평가로 결정 보존포맷 적합성 : 부분적합 - C 등급 이상의 다른 보존포맷이 없는 경우에만 채택 - 3년마다 재평가 실시하여 등급 재설정
E (미흡)	60 미만	<ul style="list-style-type: none"> 전자기록 보존포맷으로서 매우 미흡하므로 선정 불가 보존포맷 적합성 : 부적합

보존포맷 선정평가의 결과는 정량적으로 도출되지만 각각의 기준 항목들을 평가하는 과정은 정성적으로 판단해야 부분이 상당한 비중을 차지한다. 때문에 보존포맷 선정에 객관성을 높이기 위해서는 지속적으로 정량적으로 평가할 수 있는 부분을 확대해야 한다. 그리고 평가의 전문성을 제고하기 위해서 IT 및 기록관리 분야의 전문가들로 구성된 보존포맷 선정평가 위원회를 주기적으로 개최해야 한다. 필요한 경우, 위원회 소속 위원의 재량으로 소집할 수도 있다.

5. 결론

본 연구에서는 선행연구 조사를 통해 단일 보존포맷 전략의 한계를 파악하였고, 보존포맷 확장을 위해서는 단순히 몇 개의 보존포맷을 추가하는 것에 그치지 않고 보존포맷 선정할 수 있는 체계가 필요하다고 것을 알게 되었다. 이에 보존포맷 및 선정기준 현황 조사를 기반으로 전자기록물의 보존포맷을 다양화할 수 있는 보존포맷 선정체계를 제시하였다. 그리고 보존포맷을 선정할 때, 모든 전자기록물에 공통으로 적용될 수 있는 기준인 공통 기준을 근거와 함께 도출하였고 평가방식까지 제안하였다.

전자기록물 생산 환경이나 기술의 발전으로 다양한 전자기록물 유형이 업무에 활용되고 있는 상황에서 이에 대해 유연하게 대처할 수 있는 보존포맷 평가방안에 대한 전체적인 틀구조와 공통기준 및 평가방식을 제안하였다. 그러나 전자문서, 행정정보데이터세트, 시청각기록물, 웹기록물 등 공공기록물법 시행령에 명시되어 있는 다양한 전자기록물 유형별로 도출되어야 하는 고유기준은 앞으로도 계속해서 연구가 이루어져야 한다. 보존포맷 선정평가 과정에는 정성

적으로 판단해야 부분이 상당한 비중을 차지한다. 때문에 평가의 객관성을 높이기 위해서는 지속적으로 정량적으로 평가할 수 있는 부분을 확대해야 한다. 그리고 현재는 공통기준 및 고유기준의 항목들에 대한 가중치가 동일한 비중으로 설정되어 있는데 이 가중치에 대해서도 지속적으로 연구가 되어 합리적인 가중치 도출이 필요하다.

참고문헌

- 국가기록원 (2004). 전자기록물 영구보존 기반기술 용역 완료보고서. 대전: 국가기록원.
- 국가기록원 (2013). 행정기관 전자기록물 재현기술 연구 및 프로토타입 개발 완료보고서. 대전: 국가기록원.
- 국가기록원 (2008). 기술정보은행(Digital Format Registry)기능 고도화 및 기술 정보 데이터 구축. 대전: 국가기록원.
- 국가기록원 (2009). 디지털 포맷 및 애플리케이션 기술정보은행(DFR) 프로토타입 개발. 대전: 국가기록원.
- 국가기록원 (2018). 해외 내셔널 아카이브즈 전자기록관리 전략 자료집. 대전: 국가기록원.
- 강현민 (2016). 중앙기록물관리기관의 종이기록물 영구보존용 마스터 파일로서 JPEG 포맷의 표준화에 대한 연구. 한국도서관·정보학회지, 47(4), 489-510. <https://doi.org/10.16981/kliiss.47.4.201612.489>
- 박준영, 이명규 (2019). 디지털 사진기록물 관리를 위한 Raw 이미지 파일 포맷의 도입에 관한 연구. 한국기록관리학회지, 19(3), 155-178. <https://doi.org/10.14404/JKSARM.2019.19.3.155>
- 성환혁 (2007). 전자기록의 장기적 보존 및 활용을 위한 유형별 문서보존포맷에 관한 연구. 석사학위논문. 한국외국어대학교 대학원.
- 소정의, 한희정, 양동민 (2018). 국외 전자기록물의 장기보존 정책 비교 분석. 한국기록관리학회지, 18(4), 125-148. <http://doi.org/10.14404/JKSARM.2018.18.4.125>
- 소정의 (2019). 데이터세트 보존포맷 선정을 위한 주요 항목 도출에 관한 연구 - 관계형 DB의 데이터세트를 중심으로 석사학위논문. 전북대학교 대학원.
- 손원성, 임순범, 남동선, 김은미 (2009). 디지털 객체 보존을 위한 디지털 포맷 레지스트리에 관한 연구. 멀티미디어학회지, 12(10), 1397-1406.
- 송치호, 차현철 (2017). 장기보존 전자기록의 위험평가에 관한 연구. 한국컴퓨터정보학회 동계학술대회, 25(1), 29-30.
- 오세라, 정미리, 임진희 (2016). 공개포맷에 기반한 전자기록 보존 포맷 재설계 방향 연구. 한국기록관리학회지, 16(4), 79-120. <https://doi.org/10.14404/JKSARM.2016.16.4.079>
- 임나영, 남영준 (2019). 기록의 디지털화 기준에 관한 연구. 한국비블리아학회지, 30(3), 5-30. <https://doi.org/10.14699/kbiblia.2019.30.3.005>
- 정미리, 오세라, 임진희 (2016). 공문서 컴포넌트 오픈포맷 채택이 기록관리에 미치는 영향 분석. 한국기록관리학회지, 16(2), 29-55. <https://doi.org/10.14404/JKSARM.2016.16.2.029>
- 차현철, 최주호 (2019). 전자기록의 장기보존을 위한 위험평가 방법의 제안. 멀티미디어학회지, 22(1), 79-87. <https://doi.org/10.9717/kmms.2019.22.1.079>
- Abrams, Stephen et al. (2005). PDF-A: The Development of a Digital Preservation Standard. Paper presented at the 69th Annual Meeting for the Society of American Archivists, New Orleans, Louisiana, August, 14-21.
- Barnes, Ian (2006). Preservation of Word Processing Documents. Australian Partnership for sustainable repositories.
- CENDI (2007). Formats for Digital Preservation: A Review of Alternatives and Issues.
- Clausen, Lars R. (2004). Handling File Formats. Statsbiblioteket.
- ECMA (2008). Office Open XML File Formats—Part 1. 2nd Edition.
- Folkj, Mike & Barkstrom, Bruce (2003). Attributes of File Formats for Long-Term Preservation of Scientific and Engineering Data in Digital Libraries. Paper presented at the Joint Conference on Digital Libraries, Houston, TX, May 27-31.
- Knight, Gareth (2008). Framework for the definition of significant properties. The National Archives.

- Hamm, Markus & Becker, Christoph. (2011). Report on decision factors and their influence on planning. Scalable Preservation Environment.
- Hodge, Gail & Anderson, Nikkia (2007). Formats for Digital Preservation: A Review of Alternatives and Issues. *Information Services & Use*, 27, 45-63.
- InterPARES2 (2006). General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation, InterPARES 2 Project.
- Johnson, Amy Helen (1999). XML Etends its Reach: XML Finds Favor in Many IT Shops, but It's Still Not Right for Everyone. *Computer World*, 33(42), 76-81.
- Lesk, Michael E. (1995). Preserving Digital Objects: Recurrent Needs and Challenges. Paper presented at the 2nd NPO Conference on Multimedia Preservation, Brisbane, Australia.
- Moreq2010 (2011). Modular Requirements for Records Systems. Core services & plug-in modules (version 1.1). Volume 1. Europe: DLM Forum
- Müller, Eva et al. (2003). Using XML for Long-Term Preservation: Experiences from the DiVA Project. Paper presented at the Sixth International Symposium on Electronic Theses and Dissertations, Berlin, May, 109-116.
- Park, Eun G. & Oh, Sam (2012). Examining Attributes of Open Standard File Formats for Long-term Preservation and Open Access. *Information Technology and Libraries*, 31(4), 46-67.
- Puglia, Steven, Reed, Jeffrey & Rhodes, Erin (2004). Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files—Raster Images. US National Archives and Records Administration.
- Rog, Judith, & van Wijk, Caroline (2008). Evaluating File Formats for Long-term Preservation. National Library of the Netherlands.
- Sullivan, Susan J. (2006). An Archival/Records Management Perspective on PDF/A. *Records Management Journal*, 16(1), 51-56.
- Todd, Malcolm (2009). File formats for preservation. DPC Technology Watch Series Report, 09-02.
- van Wijk, Caroline & Rog, Judith (2007). Evaluating File Formats for Long-Term Preservation. Paper presented at International Conference on Digital Preservation, Beijing, China, Oct 11-12, 2007.

[웹사이트]

- Arms, Caroline R. & Fleischhauer, Carl (2006). Sustainability of Digital Formats: Planning for Library of Congress Collections. Retrieved January 10, 2020, from <https://www.loc.gov/preservation/digital/formats/index.shtml>
- Frey, Franziska (2005). 5. File Formats for Digital Masters., *Guides to Quality in Visual Resource Imaging*, Research Libraries Group and Digital Library Federation. Retrieved January 10, 2020, from <https://pdfs.semanticscholar.org/d63d/d3c6515fafb5fdf4925a26cac2e799436a0d.pdf>
- LAC (2010). File Format Guidelines for Preservation and Long-term Access(v1.0). Retrieved January 10, 2020, from <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx>
- LAC (2015). Guidelines on File Formats for Transferring Information Resources of Enduring Value. Retrieved January 10, 2020, from <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx>
- LAC (2017). Strategy for a Digital Preservation Program. Retrieved January 10, 2020, from <https://www.bac-lac.gc.ca/eng/about-us/publications/Pages/strategy-digital-preservation-program.aspx>
- NAA (2020a). Digital Preservation Policy. Retrieved January 10, 2020, from <https://www.naa.gov.au/about-us/our-organisation/accountability-and-reporting/archival-policy-and-planning/digital-preser>

vation-policy

- NAA (2020b). Long-term File Formats. Retrieved January 10, 2020, from <https://www.naa.gov.au/information-management/storing-and-preserving-information/preserving-information/preserving-digital-information/long-term-file-formats>
- NARA (2014). Appendix A: Tables of Formats. Retrieved January 10, 2020, from <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>
- NARA (2019). Frequently asked questions about Selecting Sustainable Formats for Electronic Records. Retrieved January 10, 2020, from <https://www.archives.gov/records-mgmt/initiatives/sustainable-faq.html>
- SFA (2015). Federal Archives Strategy 2016 - 2020. Retrieved January 10, 2020, from <https://www.bar.admin.ch/bar/en/home/about-us/the-federal-archives/strategy.html>
- SFA (2018). STANDARDS FOR ARCHIVING DIGITAL DOCUMENTS-Archivable File Formats. Retrieved January 10, 2020, from <https://www.bar.admin.ch/bar/de/home/archivierung/ablieferung/digitale-unterlagen.html>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Cha, Hyun-Chul & Choi, Joo Ho (2019). A Risk Assessment Method for the Long-term Preservation of Electronic Records. *Journal of Korea Multimedia Society*, 22(1), 79-87. <https://doi.org/10.9717/kmms.2019.22.1.079>
- Jung, Mi Ri, Oh, Seh-La, & Yim, Jin Hee (2016). Effects of Adopting the Open Document Format in Public Records Management. *Journal of Korean Society of Archives and Records Management*, 16(2), 29-55. <https://doi.org/10.14404/JKSARM.2016.16.2.029>
- Kang, HyenMin (2016). A Study on the Standardization of JPEG Format as a Long-Term Preservation Master File for Paper Archives in the Central Archives of Korea. *Journal of the Korean Library And Information Science Society*, 47(4), 489-510. <https://doi.org/10.16981/kliss.47.4.201612.489>
- Lim, Nayoung & Nam, Youngjoon (2019). A Study on the Criteria for Digitization of Records. *Journal of the Korean BIBLIA Society for library and Information Science*, 30(3), 5-30. <https://doi.org/10.14699/kbiblia.2019.30.3.005>
- National Archives of Korea (2004). *Electronic Record Permanent Preservation Based Technology Service*. Daejeon: National Archives of Korea.
- National Archives of Korea (2008). National Archives of Korea (2013). *A study on the reproduction technology and the prototype for the electronic records of administrative agency*. Daejeon: National Archives of Korea.
- National Archives of Korea (2009). *DFR Prototype Development for Digital Formats and Applications*. Daejeon: National Archives of Korea.
- National Archives of Korea (2013). *A study on the reproduction technology and the prototype for the electronic records of administrative agency*. Daejeon: National Archives of Korea.
- National Archives of Korea (2018). *Global Trend in Digital Archiving Strategy*. Daejeon: National Archives of Korea.
- Oh, Seh-La, Jung, Mi Ri, & Yim, Jin Hee (2016). Redesigning Electronic Records Preservation Formats Based on Open Formats. *Journal of Korean Society of Archives and Records Management*, 16(4), 79-120. <https://doi.org/10.14404/JKSARM.2016.16.4.079>
- Park, Junyoung & Lee, Myounggyu (2019). A Study on the Introduction of Raw Image File Formats for the Management of Digital Photographic Records. *Journal of Korean Society of Archives and Records Management*, 19(3), 155-178. <https://doi.org/10.14404/JKSARM.2019.19.3.155>
- Seong, Hwan Hyuk (2007). *A study on document preservation format classified by the type for long-term preservation and use of electronic records*. Master's Thesis, Hankuk University of Foreign Studies. Seoul.

- So, Jeong Eu (2019). A Study on Derivation Critical Factor for Selection of Dataset Preservation Format: Focus on Dataset of Relational Database. Master's Thesis, Jeonbuk National University of Graduate School. Jeonju.
- So, Jeong Eui, Han, Hui Jeong, & Yang, Dong Min (2018). A Comparative Analysis of Long-Term Preservation Policies in Foreign Electronic Records. *Journal of Korean Society of Archives and Records Management*, 18(4), 125-148.
<http://doi.org/10.14404/JKSARM.2018.18.4.125>
- Sohn, Won-Sung, Lim, Sun-Bum, Nam, Dong-Sun, & Kim, Eun-Mi (2009). A study on the Digital Format Registry for digital objects preservation in Korea. *Journal of Korea Multimedia Society*, 12(10), 1397-1406.
- Song, Chi-Ho & Cha, Hyun-Chul (2017). A Study on the risk evaluation of electronic records for long-term preservation. *Journal of The Korea Society of Computer and Information Winter Conference*, 25(1), 29-30.

