

전자기록물 보존포맷 선정평가를 위한 지원도구 설계방안 연구*

A Study on the Design Framework for a Support Tool in Evaluating Digital Records Preservation Formats

김 지 혜 (Ji-Hye Kim)**

송 채 은 (Chae-Eun Song)***

양 동 민 (Dongmin Yang)****

초 록

국가기록원은 전자기록물의 장기적 접근성과 재현성을 보장하기 위해 보존포맷으로 변환하는 마이그레이션 전략을 채택하고 있다. 보존포맷을 선정하기 위해서는 보존포맷 선정체계에 따른 평가를 지속적으로 수행해야 한다. 그러나 파일포맷에 대한 공신력 있는 정보 접근이 제한적이고, 파일포맷의 기술적 사양과 특성에 대한 전문지식이 요구되는 등 평가 과정에서 여러 어려움이 발생한다. 이러한 문제는 보존포맷 선정체계의 실무 정착을 저해하는 요인으로 작용할 수 있다. 본 연구는 보존포맷 선정평가의 효율성을 높이고자 평가를 지원하는 도구의 기능을 정보관리와 평가지원 영역으로 구분하여 제안하였다. 그리고 해당 도구의 실현 가능성을 평가하기 위해 파일포맷 정보 리소스 3곳에서 제공하는 정보를 보존포맷 선정기준(공통) 평가문항과 비교, 분석하여 정보 연계 가능성을 검토하였다.

ABSTRACT

The National Archives of Korea (NAK) has adopted a migration strategy to convert preservation formats to ensure the long-term accessibility and reproducibility of digital records. To select an appropriate preservation format, it is necessary to conduct evaluations based on the preservation format selection system. Evaluating file formats is challenging due to limited access to reliable information about file formats and the requirement for specialized knowledge of their technical specifications and characteristics. As a result, the evaluation process can act as a barrier to the practical implementation of the preservation format selection system. This study aims to improve the efficiency of preservation format selection evaluations by proposing the functionality of a support tool divided into two domains: information management and evaluation support. Furthermore, to assess the feasibility of this tool, the study analyzed the information provided by three reliable file format information portals and compared it with the evaluation criteria (common) of the preservation format selection system to examine the potential for information integration.

키워드: 보존포맷 선정평가, 장기보존, 전자기록물, 보존포맷, 지원도구

Preservation format selection evaluation, Long-term preservation, Digital records, Preservation format, Support Tool

* 본 연구는 전북대학교 4단계 BK21 대학원혁신지원사업의 지원을 받아 수행된 연구임.

이 논문은 2023년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구지원사업의 지원을 받아 수행된 연구임(NRF-2023S1A5A2A01077759).

** 전북대학교 기록관리학과 박사과정(kimjh129056@gmail.com) (제1저자)

*** 남동발전 기록물관리 전문요원(chaeesung@koenergy.kr) (공동저자)

**** 전북대학교 기록관리학과 교수, 문화융복합아카이빙연구소 공동연구원(dmyang@jbnu.ac.kr) (교신저자)
논문접수일자: 2024년 11월 22일 논문심사일자: 2024년 11월 25일 게재확정일자: 2024년 12월 10일
한국비블리아학회지, 35(4): 163-201, 2024. <http://dx.doi.org/10.14699/kbiblia.2024.35.4.163>

※ Copyright © 2024 Korean Biblia Society for Library and Information Science

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.

1. 서론

1.1 연구의 배경 및 목적

『공공기록물 관리에 관한 법률 시행령』(이하 공공기록물법 시행령) 제36조 1항에 따라, 필요한 경우 보존기간이 10년 이상인 전자기록물을 보존포맷으로 변환하여 관리할 수 있다. 국가기록원은 전자기록물의 장기적 접근성을 보장하기 위해 파일포맷을 변환하는 마이그레이션 전략을 도입한 것이다. 전자기록물은 파일(file) 단위로 저장되며, 파일은 특정 파일포맷(file format)에 따라 구조화된다. 파일포맷은 컴퓨터가 파일을 저장하고 읽는 방식을 정의한 규칙으로, 소프트웨어나 운영체제에서 파일 내 데이터를 처리하는 기준이 된다. 이로 인해 전자기록물의 생성·저장·재현 등 모든 과정은 파일포맷과 이를 지원하는 소프트웨어, 하드웨어, 운영체제 등에 의존하게 된다. 그러나 기술의 변화와 발전으로 인해 기존 파일포맷의 지원이 중단되거나 소프트웨어가 더 이상 파일포맷을 지원하지 않는 상황이 발생하며, 이는 전자기록물의 접근성과 재현성에 부정적 영향을 미친다. 따라서 이러한 문제를 해결하기 위해서는 보존포맷으로의 변환이 필수적이다.

보존포맷은 전자기록물 생산 당시의 내용과 외형 등 주요 특성을 유지한 채 시간의 흐름과 기술의 변화에 상관없이 해당 기록물을 재현할 수 있는 포맷이다(국가기록원, 2022, NAK 37:2022, v1.0). 2022년에 이루어진 『공공기록물법 시행령』 개정을 통해 기존의 ‘문서보존포맷’이 ‘보존포맷’으로 변경되었다. 이는 선택 가능한 보존포맷의 범위를 확대하고 보존정책의

유연성을 강화하려는 목적에서 추진된 것이다. 과거 국가기록원은 PDF/A-1b 포맷만을 보존포맷으로 인정하였다. 그러나 PDF/A-1b 포맷은 기록물의 시각적 외형만 보장할 뿐, 기록물이 가지는 본래의 특성을 모두 보장할 수 없다는 한계가 있어 모든 기록물 유형에 일괄 적용하기 어렵다는 문제가 제기되었다(한능우, 2023). 이러한 문제를 해결하기 위해 국가기록원은 기록물 유형별 보존포맷을 신규 지정하는 게 아닌, 『전자기록물 보존포맷 선정기준(v1.0)』 표준을 제정하여 보존포맷을 자체적으로 선정할 수 있는 체계를 수립하였다.

해당 표준에는 보존포맷 선정기준과 선정평가 방식이 규정되어 있다. 보존포맷 선정기준은 모든 기록물 유형에 적용되는 공통기준과 기록물 유형에 따라 다르게 적용되는 고유기준으로 구성되어 있으며, 각 기준에는 평가지표로 활용되는 문항들이 포함되어 있다. 공통기준은 파일포맷 자체의 기능과 지원 가능성을, 고유기준은 기록물 유형별 고유한 속성과 사양을 평가하는 문항들로 이루어져 있다. 이 두 기준의 평가 목적은 다르지만, 모두 파일포맷의 기술적 특성을 확인하는 데 중점을 둔다. 이는 전자기록물의 장기보존이 정보기술에 의해 좌우되기 때문이다. 따라서 보존포맷 선정평가는 파일포맷의 기술정보(technical information)를 근거로 이루어진다.

그러나 평가자는 파일포맷 관련 정보를 수집하고 분석하는 과정에서 다음과 같은 어려움을 겪게 된다. 첫째, 공신력 있는 파일포맷 정보에 대한 접근 제약이 존재한다. 둘째, 파일포맷의 기술적 사양과 특성을 이해하고 분석하기 위해 평가자에게 전문지식이 요구된다. 이러한 문제

점들은 보존포맷 선정평가 과정을 복잡하게 만들고 하나의 파일포맷을 평가하는 데 많은 시간과 자원을 소모하게 한다. 더 나아가 보존포맷 선정체계의 실무 정착을 저해하는 요인으로 작용할 수 있다.

이에 본 연구는 보존포맷 선정평가를 효율적으로 수행할 수 있도록 평가문항 응답에 필요한 정보를 제공하는 보존포맷 선정평가 지원도구(이하 평가 지원도구)의 설계방안을 제안한다. 이를 위해 먼저 보존포맷 선정체계와 관련된 선행연구와 국내외 사례를 조사한다. 3장에서는 평가 지원도구의 구조와 기능을 설계하고 이에 대한 적용성을 검토한다. 4장에서는 파일포맷 관련 정보 리소스에서 제공하는 정보와 보존포맷 선정기준(공통) 평가문항과의 연계 가능성을 평가하여 정보 유효성을 검토한다. 마지막으로 5장에서는 전반적인 연구 내용을 정리하고, 평가 지원도구 활용을 통해 얻을 수 있는 이점과 시사점, 한계점 등을 제시하며 결론을 맺는다.

1.2 선행연구

현행 기록관리 공공표준 「전자기록물 보존포맷 선정기준(v1.0)」은 전자기록물을 보존포맷으로 마이그레이션하기 위한 체계를 지원한다. 보존포맷으로의 변환은 2007년 「공공기록물법」과 동법 시행령의 개정에서 시작되었으며, 이 개정으로 국가기록원은 보존기간이 10년 이상인 전자기록물을 문서보존포맷과 장기보존포맷으로 변환하여 관리하여야 한다고 규정하였다. 더불어 국가기록원은 법령을 근거로 문서보존포맷과 장기보존포맷의 기술규격에 관한

공공표준을 제정하였고(국가기록원, 2008a, NAK 30:2008, v1.0; 2008b, NAK/TS 3:2008, v1.0), PDF/A-1b 포맷을 공식 보존포맷으로 채택하였다. 그러나 보존포맷으로 변환하는 목적과 다르게 PDF/A-1b 포맷은 다양한 전자기록물 유형의 고유한 특성과 기능을 유지하지 못한다는 기술적 한계가 나타났다. 이러한 문제점을 해결하기 위해 전자기록물 보존포맷 선정체계 수립에 관한 연구가 본격적으로 수행되었다.

소정의 외(2018)는 미국, 캐나다, 영국, 호주, 스위스의 아카이브 기관들이 발표한 장기보존 정책을 비교·분석하여 장기보존을 위한 핵심 정책 요소를 제시하였다. 이 연구에서 제시된 정책 요소 중 하나인 ‘위험 관리’는 보존포맷이나 보존매체에 대한 위험을 식별하고 평가하는 프로세스를 구축하여 시행하는 것을 의미한다. 이는 보존포맷의 위험을 평가하는 것이 전자기록물의 장기보존에 필수적임을 시사한다. 한희정, 오효정 외(2020)는 전자기록물 보존포맷의 다양화를 위한 선정체계를 제안하였다. 보존포맷 선정기준으로 공통기준과 고유기준의 개념을 제시하고, 두 기준 중 공통기준에 대한 평가문항을 도출하였다. 이 연구에서 제시한 보존포맷 선정체계는 모듈형으로 설계되어 새로운 기록물 유형이 등장할 때마다 고유기준만을 수립하여 기존 보존포맷 선정체계에 통합할 수 있는 확장성을 가진다. 이 두 연구는 전자기록물의 장기보존 측면에서 보존포맷의 위험 평가와 체계적 보존포맷 선정기준의 필요성을 강조하였다.

보존포맷 선정체계와 공통기준의 평가문항 설계에 이어 한희정, 윤성호 외(2020)는 데이터세트 유형 전자기록물의 보존포맷 선정을 위

한 고유기준 평가문항을 개발하였다. 국내외에서 데이터세트 보존포맷으로 채택 또는 검토되고 있는 SIARD 2.1을 대상으로 개발한 평가체계를 적용하여 보존포맷으로서의 적합성을 검증하였다. 특히, 고유기준 설계 시 기록물 유형의 본질적 속성인 필수보존속성(Significant Properties)을 우선으로 도출해야 함을 강조하였다. 한편, 윤성호 외(2022)는 문서 유형 전자기록물의 고유한 특성이 문서 생산에 주로 사용되는 소프트웨어의 기능에 기반한다고 보았다. 이에 전자문서의 하위 유형인 텍스트형, 스프레드시트형, 프리젠테이션형의 주요 소프트웨어 기능요소를 도출하여 필수보존속성으로 사용하였다. 이 두 연구는 전자기록물 유형별 필수보존속성을 도출하고, 이를 반영한 고유기준을 설계하는 방향성을 제시하였다.

「전자기록물 보존포맷 선정기준(v1.0)」표준이 제정된 2022년 이후, 보존포맷 선정기준의 고유기준 개발에 관한 연구가 활발히 이루어졌다. 전한역 외(2023)는 국내 기록관리 실정을 반영하여 기존 필수보존속성의 범주별 정의를 재정립하고, 이를 기반으로 시청각기록물 중 오디오형의 필수보존속성을 제시하였다. 전한역, 양동민(2024)은 전한역 외(2023)에서 제시한 필수보존속성을 바탕으로 고유기준 평가문항을 설계하였다. 그리고 범용적으로 사용되는 오디오형 파일포맷을 대상으로 기존 공통기준과 해당 연구에서 수립한 고유기준을 적용한 보존포맷 선정평가 결과를 도출하였다. 이정은, 양동민(2023)은 데이터세트 관리에 대한 사회적 관심이 증대됨에 따라 데이터세트 장기보존을 위한 보존포맷 선정체계의 필요성을 강조하였다. 데이터베이스형과 구조화데이터형으로 구분

하여 데이터세트의 필수보존속성을 도출하였다. 송채은, 양동민(2024)은 시청각기록물 중 이미지형에 대한 고유기준 평가체계를 설계하고, 이를 실제 파일포맷에 적용하여 평가문항의 적합성을 검증하였다. 더불어 공공표준 「전자기록물 보존포맷 선정기준(v1.0)」에 따른 보존포맷 최종 선정 단계에서 공통기준 평가지표 중 하나인 국제화 표준 여부만 반영하여 보존포맷과 수용가능포맷을 구분하는 문제를 지적하며, 고유기준 평가지표도 최종 선정 단계에 반영되도록 보존포맷 선정 방식 일부를 개선해야 함을 주장하였다. 김지혜, 양동민(2024)은 시청각기록물 중 비디오형에 대한 고유기준 평가체계를 제안하였다. 비디오형 전자기록물은 여러 유형의 기록물을 구성요소로 가지며 이를 하나로 묶는 컨테이너(container)와 이형의 데이터를 인코딩/디코딩하는 여러 코덱이 필요하다는 특징에 기반하여 필수보존속성을 컨테이너형과 코덱형으로 구분하여 도출하였다. 그리고 이를 기준으로 고유기준을 설계하였다. 이 연구들은 다양한 기록물 유형별 특성을 반영한 고유기준을 설계하여 보존포맷 선정체계의 확장성과 실효성을 높이는 데 기여하였다.

선행연구를 검토한 결과, 보존포맷 선정체계에 관한 연구는 「전자기록물 보존포맷 선정기준(v1.0)」 표준의 제정 이전과 이후로 나뉜다. 표준 제정 이전에는 단일보존포맷 전략을 넘어서는 새로운 보존포맷 선정체계를 마련하기 위한 기초연구가 주로 이루어졌다. 표준 제정 이후에는 보존포맷 선정체계의 적용 범위를 확장하기 위해 전자기록물 유형별 필수보존속성 도출하고, 이에 따른 고유기준 평가문항을 설계하는 연구가 활발히 진행되었다. 그러나 현행

보존포맷 선정체계 적용 시 발생하는 어려움과 보존포맷 선정평가를 지원하는 방안에 관한 연구는 확인할 수 없었다. 이에 본 연구에서는 기록관리 실무에서 보다 효율적으로 보존포맷 선정평가를 수행할 수 있도록 지원하는 도구의 필요성과 이를 위한 평가 지원도구의 구조와 기능을 설계하고자 한다.

1.3 연구방법 및 대상

본 연구는 전자기록물 보존포맷 선정평가의 효율성을 제고하기 위해 평가에 필요한 정보를 수집하여 평가자에게 제공하는 평가 지원도구의 구조와 기능을 설계하는 것을 목적으로 한다. 이를 위해 먼저 보존포맷 선정체계에 관한 선행연구와 국내외 사례를 조사 및 검토하였다. 둘째, 보존포맷 선정평가 과정에서 평가자가 직면하는 주요 어려움을 분석하여 정리하였다. 또한 보존포맷 선정평가의 효율성을 높이는데 필요한 평가 지원도구의 구조와 기능을 설계하였다. 마지막으로, 앞서 제안한 평가 지원도구 기능의 실현 가능성을 확인하기 위해 적용성을 검토하였다. 특히, 평가 지원도구의 핵심은 정보 제공에 있으므로 정보의 유효성을 확인하여

평가항목별 정보 제공 가능성을 구체적으로 살펴보고자 한다. 이를 위해 파일포맷 관련 정보를 제공하는 정보 리소스를 선정하고 보존포맷 선정기준(공통)의 평가문항을 지표로 삼아 보존포맷 선정평가에 필요한 정보를 분석하였다.

분석 대상은 <표 1>과 같이 공신력 있다고 평가되는 미국 국립문서기록관리청(National Archives and Records Administration, 이하 NARA)에서 제공하는 위험 매트릭스(Risk Matrix), 미국 의회도서관(Library of Congress, 이하 LOC)의 디지털 포맷의 지속가능성(Sustainability of Digital Formats)에서 제공하는 포맷 설명 문서(Formats Description Documents, 이하 FDDs), FileInfo.com으로 선정하였다.

NARA의 위험 매트릭스는 파일포맷의 위험 수준을 측정하기 위한 도구로, 파일포맷의 안정성과 위험성을 평가하는 8개 범주와 27개 문항으로 구성되어 있다. 각 문항은 긍정(Yes), 부정(No), 알 수 없음(Unknown), 해당 없음(N/A) 등 4가지 선택지로 응답할 수 있으며, 응답에 따라 음수 또는 양수 점수가 부여된다(NARA, 2024). NARA의 파일포맷 위험 수준 평가 방식은 국가기록원의 보존포맷 선정평

<표 1> 보존포맷 선정평가 지원도구 적용성 검토를 위한 정보 리소스

구분	웹사이트명 (웹사이트 URL)	운영주체	게시연도
국외 국립기록원	NARA Digital Preservation Risk Matrix (https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Risk_Matrix)	정부/공공기관 (NARA)	2019
국외 국립도서관	Sustainability of Digital Formats (https://www.loc.gov/preservation/digital/formats/intro/intro.shtml)	정부/공공기관 (LOC)	2004
기타	FileInfo.com (https://fileinfo.com/)	전문가협의체	2005

가 방식과 유사하다. 위험 매트릭스의 문항에 관한 결과가 어떤 정보에 근거하는지 알 수 없다는 한계가 있지만, 본 연구에서는 보존포맷 선정기준의 평가문항과 유사한 위험 매트릭스 문항의 점수를 환산하여 보존포맷 선정평가에 활용할 수 있다고 판단하였다.

LOC의 FDDs는 디지털 콘텐츠의 장기보존을 지원하기 위해 다양한 파일포맷에 대한 심층적이고 검증된 기술정보를 체계적으로 제공하는 자료이다. FDDs에서는 파일포맷의 개요와 다른 파일포맷 간의 관계를 서술하는 식별 및 설명(Identification and description), 파일포맷의 적합성을 평가하는 지속가능성 요인(Sustainability factors), 기록물 유형별 품질 및 기능 요인(Quality and functionality factors) 등 8개 범주와 각 범주의 하위 요소에 따른 정보를 제공한다(Library of Congress, 2004). 본 연구에서는 FDDs가 검증된 데이터를 기반으로 구성된 신뢰성 높은 정보 리소스라 간주하였다. 더불어 각 범주의 요소에는 보존포맷 선정기준의 평가문항과 연계될 수 있는 정보가 포함되어 있어 보존포맷 선정평가에 활용할 수 있다고 판단하였다.

FileInfo.com은 다양한 파일 확장자를 기준으로 파일포맷 개요와 이를 지원하는 소프트웨어에 대한 정보를 제공하는 온라인 데이터베이스이다(FileInfo.com, 2005). 비록 정부나 공공기관에서 운영하는 포털은 아니지만, 파일포맷과 정보를 제공하는 권위 있는 포털로 평가된다. 특히 상세한 지원 소프트웨어 정보를 포함하고 있어 이에 대한 정보가 상대적으로 미흡한 NARA의 위험 매트릭스와 LOC의 FDDs를 보완할 수 있다고 판단하여 분석 대상으로 선정하였다.

2. 이론적 배경

2.1 국가기록원의 전자기록물 보존포맷 선정 체계

국가기록원은 「공공기록물법」 제20조의 1항에 따라 전자기록물 관리체계를 구축해야 하며, 동법 시행령 36조에 따라 보존기간이 10년 이상인 전자기록물의 경우 필요에 따라 보존포맷으로 변환하여 관리할 수 있도록 기준을 마련해야 한다. 이에 따라 국가기록원은 보존포맷 선정절차와 방식을 체계화한 공공표준 「전자기록물 보존포맷 선정기준(v1.0)」을 제정하여 국내 보존포맷 선정체계를 확립하였다.

보존포맷 선정평가는 사전에 수립된 보존포맷 선정기준에 기반하여 이루어진다. 보존포맷 선정기준은 특정 파일포맷이 보존포맷으로 사용되기 적합한지를 평가하는 기준이다. 이 기준은 공통기준과 고유기준으로 구분된다. 공통기준은 보존포맷 선정평가 대상이 되는 모든 파일포맷에 적용되고, 고유기준은 기록물 유형별 개별적으로 수립되어 기록물 유형에 따라 다르게 적용된다. 공통기준은 기록물 유형과 관계없이 파일포맷이 전자기록물의 장기보존과 이용가능성을 보장할 수 있는지 확인하는 문항들로 이루어져 있다. 고유기준은 기록물 유형에 따라 필수보존속성(Significant Properties)을 유지 및 재현할 수 있는 사양이나 기능을 제공하는지 확인하는 문항들로 이루어져 있다.

보존포맷 선정평가는 평가 결과점수를 산출한 뒤, 이를 사전에 설정한 기준 점수와 비교하여 보존포맷 적합성을 판별하는 방식으로 이루어진다. 각 평가문항에는 점수가 부여되어 있으

며, 문항에 대해 긍정(Yes) 또는 부정(No)으로 응답하는 방식으로 평가가 진행된다. 긍정으로 응답한 평가문항의 점수를 합산하여 평가 결과 점수를 산출하고, 이 점수를 기준 점수와 비교한다. 평가 결과점수가 기준 점수보다 높은 파일포맷만을 대상으로 다음 단계의 평가를 진행한다.

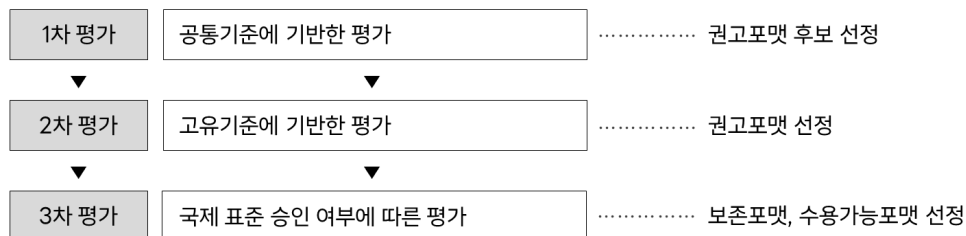
보존포맷 선정평가 절차는 <그림 1>에 제시된 순서로 진행된다. 1차 평가는 보존포맷 선정 기준(공통)을 기반으로 파일포맷을 평가한다. 공통기준 평가 결과점수가 75점 이상인 파일포맷을 권고포맷(보존에 적합한 파일포맷) 후보로 선정하고, 그렇지 않은 포맷은 부적합포맷으로 분류한다. 2차 평가는 보존포맷 선정기준(고유)을 기반으로 권고포맷 후보인 파일포맷을 평가한다. 고유기준 평가 결과점수가 1레벨(75점 이상), 2레벨(50~75점 미만)인 파일포맷을 권고포맷으로 선정하고, 그렇지 않은 파일포맷은 비권고포맷으로 분류한다. 마지막으로 3차 평가는 권고포맷을 대상으로 국제 표준 승인 여부를 평가한다. 국제 표준으로 승인된 권고포맷을 보존포맷으로, 국제 표준화 과정을 거치지 않은 권고포맷은 수용가능포맷으로 분류한다.

국내 보존포맷 선정체계는 기록물 유형별로 고유기준을 수립하도록 설계되어, 새로운 기록

물 유형이 등장해도 기존 체계에서 고유기준만 추가하면 되는 확장성을 가진다. 이는 기술 및 생산 환경 변화에도 유연하게 대응할 수 있는 장점이 있다. 현행 「전자기록물 보존포맷 선정기준(v1.0)」에는 텍스트형, 프리젠테이션형, 스프레드시트형 등의 고유기준이 포함되어 있다. 국가기록원은 고유기준 적용 범위를 확대하기 위해 지속적으로 R&D 연구를 진행하고 있다. 한편, 국가기록원은 디지털 객체의 장기보존을 지원하기 위해 전자기록 기술정보관리시스템(Digital Format Registry, 이하 DFR)을 개발하여 파일포맷의 기술정보와 관련 지원 소프트웨어 정보를 지속해서 수집·관리하고 있다. DFR은 지속적으로 고도화되고 있으며, 최근에는 보존포맷 선정체계를 활용하여 파일포맷의 위험평가와 연계할 수 있는 위험관리 도구 모듈을 개발하고 있다(국가기록원, 2023).

2.2 미국 국립문서기록관리청의 전자기록물 보존포맷 선정체계

미국 국립문서기록관리청(National Archives & Records Administration, 이하 NARA)은 2017년에 처음으로 '디지털 기록의 장기보존을 위한 전략(Strategic for Preserving Digital Archival



<그림 1> 국가기록원의 전자기록물 보존포맷 선정과정

Materials)’을 발표하였다. 현재는 ‘디지털 보존 전략 2022-2026(Digital Preservation Strategy 2022-2026)’이 수립되어 있다. 디지털 보존 전략은 총 8개의 전략으로 구성되어 있으며, 대상 기록에는 디지털 생산기록(born-digital records)과 디지털 대체사본(digital surrogate)이 포함된다. 디지털 보존 전략 중 디지털 보존 프로그램(Digital Preservation program)은 NARA의 보유기록물에 대한 보존 위험을 평가하고, 기록물의 접근성을 유지하기 위한 프로그램이다. NARA는 디지털 보존 프로그램에 따라 디지털 보존 프레임워크(Digital Preservation Framework)를 개발하였다(NARA, 2022).

디지털 보존 프레임워크는 위험 매트릭스(Risk Matrix)와 파일포맷 보존 조치 계획(File Format Preservation Action Plans)으로 구성된다. 위험 매트릭스는 파일포맷의 장기보존 위험성을 평가하기 위한 도구로, 2024년 9월 기준 8개 범주와 27개의 질문으로 구성되어 있다. 각 질문에는 특정 점수가 할당된 Yes, Unknown, N/A, No 선택 옵션이 있으며, 모든 질문의 선택 값을 합산하여 파일포맷의 위험 수준을 식별한다. 이를 통해 향후 변환할 필요가 있는 파일포맷을 판별할 수 있다. NARA는 문서, 프레젠테이션, 스프레드시트, 데이터세트, 이미지, 오디오, 동영상, CAD, 지리공간, 웹, 이메일 등 16개 범주의 기록물 유형에 대해 파일포맷 보존 계획을 수립하였다. 여기에는 기록물 유형별로 파일포맷 마이그레이션 시 반드시 유지해야 할 필수보존속성(Significant Properties)이 정의되어 있으며, 선호(preferred)포맷과 허용가능(acceptable)포맷이 함께 명시되어 있다.

NARA는 Github에 위험 매트릭스와 파일

포맷 보존 조치 계획을 공개하고 있다(NARA, 2019). 특히, 위험 매트릭스의 경우 NARA에서 자체적으로 평가한 파일포맷 위험 평가결과를 주기적으로 갱신하여 공개하고 있다.

2.3 그 외 전자기록물 보존포맷 선정체계

보존포맷 선정체계를 명시적으로 공개하고 있지는 않지만, 전자기록물의 파일포맷 관리를 위한 장기보존 전략을 수립하고 이를 실행하는 기관들이 세계 각국에 존재한다. 이러한 기관들은 각자의 환경과 필요에 맞는 장기보존 정책과 기술을 개발하여 디지털 기록물의 보존과 접근성을 보장하기 위해 노력하고 있다. 이 중 영국 국립기록보존소와 미국 의회도서관은 파일포맷에 대한 기술정보를 제공하는 정보 리소스를 운영하고 있다.

영국의 국립기록보존소(The National Archives, 이하 TNA)는 2011년과 2017년 두 번에 걸쳐 디지털 전략을 발표하여 시간이 지나도 디지털 기록을 사용 가능한 형태로 유지하기 위한 보존 정책을 제시하였다. TNA는 이관 시 디지털 기록의 생성 당시의 파일포맷을 유지하는 것을 원칙으로 삼는다. 에뮬레이션을 장기보존 전략으로 채택한 것인데, 대안이 없는 경우에는 디지털 기록을 다른 파일포맷으로 마이그레이션하는 것을 허용한다. 즉, TNA는 파일포맷 변환을 최소화하며 생산 시 사용된 파일포맷을 그대로 보존하는 것을 기본 정책으로 삼고 있다. TNA는 홈페이지에 이관을 위한 파일포맷(File Format for digital transfers) 가이드를 공개하고 있으며, 워드 프로세싱과 텍스트, 스프레드시트, 프레젠테이션, 그래픽, 오디오, 비

디오, 이메일, 기타 등 8가지 유형별로 파일포맷 명칭, 버전, 확장자, 식별자(PUID) 정보를 제시하고 있다(TNA, n.d). 또한 위험 평가(Risk Assessment)를 통해 사용하는 파일포맷 및 소프트웨어 제품의 지원이 중단될 위험을 언급하였는데(TNA, 2017), 이는 TNA 내부적으로 파일포맷에 대한 평가를 진행하였음을 간접적으로 드러낸다. TNA는 COPTR, InSPECT 등 주요 디지털 보존 프로젝트를 주관하였으며, 신뢰할 수 있는 보존도구를 개발하여 공개하였다. TNA가 개발한 보존도구로 DROID(Digital Record Object Identification)와 PRONOM이 있다. DROID는 파일포맷 식별 도구로, 파일 확장자가 잘못되거나 누락되더라도 내부 시그니처를 검색하여 정확한 파일포맷과 버전을 식별할 수 있다. DROID를 활용해 정보자산을 이해하는 데 필요한 정보를 수집하고 디지털 연속성에 대한 위험을 관리할 수 있다. PRONOM은 2002년에 개발된 파일포맷 레지스트리 데이터베이스로, 디지털 아카이브에 저장된 파일포맷의 특성과 기술정보를 제공하며, DROID와 연계되어 디지털 보존에 활용된다(소정의 외, 2018).

미국 의회도서관(Library of Congress, 이하 LOC)은 디지털 보존을 위해 두 가지 주요 리소스인 '디지털 파일포맷의 지속가능성(Sustainability of Digital Formats)'과 '권장 파일포맷 지침(Recommended Formats Statement, 이하 RFS)'을 제공한다. 디지털 파일포맷의 지속가능성은 파일포맷의 장기보존 적합성을 평가하기 위한 기술정보를 정리한 여러 파일포맷 설명 문서(Format Description Documents, 이하 FDDs)로 구성된다. FDDs는 사전에 설계한 기술요소에 따라 파일포맷의 개요와 관련 정보를 XML 기

반으로 작성한 웹 문서이다. 모든 파일포맷의 FDDs에는 공표(Disclosure), 채택(Adoption), 투명성(Transparency), 자체문서화(Self-documentation), 외부 종속성(External dependencies), 특허의 영향(Impact of patents), 기술보호 메커니즘(Technical protection mechanisms)이 포함된 지속가능성 요소와 파일포맷의 유형에 따라 상이한 품질 및 기능(Quality and functionality)을 추가로 설명한다. 이는 파일포맷의 필수보존속성을 반영한 기능적 요소이다.

RFS는 LOC가 디지털 기록의 장기보존을 위해 권장하는 파일포맷을 정리한 지침서이다. 이 지침은 기록물 유형별로 안정적이고 호환성이 높으며 기술적 요구사항을 충족하는 파일포맷을 선호(preferred)포맷과 수용가능(acceptable)포맷으로 구분하여 제시한다. RFS는 장기적 보존과 접근성을 보장하기 위해 글로벌·커뮤니티(Global/community) 기준과 로컬·기관(Local/institutional) 기준을 고려한다. 전자는 상기 디지털 파일포맷 지속가능성의 7가지 지속가능성 요소를 바탕으로 평가하는 기준이며, 후자는 소장 자료를 장기보존하는데 필요한 인적, 물적 자원의 수준을 평가하는 기준이다. LOC는 이러한 기준을 바탕으로 파일포맷을 평가하여 위험을 관리한다. 최근 RFS가 개정되면서 파일포맷 평가를 위해 사용하는 평가 매트릭스 템플릿(Evaluation matrix template)이 공개되었다. 해당 매트릭스는 글로벌·커뮤니티 기준에 따른 8가지 문항과 로컬·기관 기준에 따른 6가지 문항으로 구성되어 있으며, 파일포맷에 대한 장기보존 적합성을 평가하여 선호 또는 허용 여부를 결정하는 평가 방식이다(Library of Congress, 2024).

3. 전자기록물 보존포맷 선정평가 지원도구의 설계방안

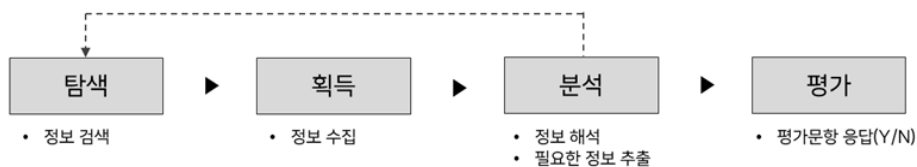
전자기록물의 파일포맷을 필요에 따라 보존포맷으로 변환하는 업무를 수행하려면, 기록관리 공공표준 『전자기록물 보존포맷 선정기준(v1.0)』에 수립된 선정체계를 따라야 한다. 본 장은 해당 표준에 기반한 보존포맷 선정평가를 수행하는 과정에서 평가자가 마주하는 어려움을 분석하였다. 또한 이에 대한 해결방안으로 보존포맷 선정평가 지원도구를 제안하고, 평가 지원도구에 필요한 기능요건을 함께 제시하였다.

3.1 전자기록물 보존포맷 선정평가 과정의 복잡성

전자기록물 보존포맷 선정평가는 미리 수립된 보존포맷 선정기준(공통, 고유)에 따라 특정 파일포맷을 평가하여 보존포맷으로서의 적합성을 판단하는 절차이다. 선정기준은 파일포맷을 평가하기 위한 문항들로 구성되어 있고, 각 문항에 긍정 또는 부정으로 응답하는 방식으로 평가가 이루어진다. 평가자는 파일포맷이 각 평가문항에 명시된 조건을 충족하는지 판단하기 위해 <그림 2>와 같이 필요한 정보를 수집하고 분석하는 단계를 거쳐야 한다.

보존포맷 선정평가의 핵심은 정확하고 신뢰할 만한 파일포맷 관련 정보를 확보하는 데 있다. 일반적으로 정보 탐색은 구글이나 네이버와 같은 검색엔진을 통한 검색으로 이루어진다. 이 과정에서 어떤 경로를 통해 정보를 찾느냐가 중요하다. 개인 블로그나 비공식적인 웹사이트에서 제공된 정보는 사실 여부를 검증하기 어렵다. 따라서 공신력 있는 기관에서 제공하는 문서에서 정보를 확보하는 것이 바람직하다. 여기서 공신력 있는 기관은 파일포맷을 개발한 기업 또는 조직, 파일포맷을 표준화한 표준 기구이다. 이들이 제공하는 파일포맷 문서는 사양서(specification)라 불리며, 파일포맷의 사양과 구조, 동작 방식 등이 기술되어 있다.

그러나 공신력 있는 기관에서 제공하는 문서에 접근하여 정보를 확보하는 데에는 다음과 같은 제약이 따른다. 첫째는 파일포맷 문서 자체에 대한 접근 제약이고, 둘째는 파일포맷의 기술 사양에 대한 접근 제약이다. 이는 독점 파일포맷(proprietary file format)에서 주로 나타난다. 독점 파일포맷은 상업적 목적으로 개발되며 특정 기업이나 조직이 소유권을 가진다. 이러한 파일포맷은 경쟁력을 유지하기 위해 자사 소프트웨어에서만 지원되거나 특허와 라이선스 조건이 붙는다. 이로 인해 파일포맷 관련 문서 자체가 공개되지 않거나 극히 일부 기술 사양만 문서로 공개되는 경우가 많다. 독점 파



<그림 2> 전자기록물 보존포맷 선정평가 절차

일포맷 외 파일포맷에서도 위에서 언급한 제약이 나타난다. 표준 파일포맷은 표준화 기구에 의해 검토되고 승인된 파일포맷으로, 개방성과 상호운용성을 목적으로 기술 사양이 표준 문서로 공개된다. 그러나 일부 표준 파일포맷은 기술 사양이 공개되더라도 열람과 다운로드를 위해 유료 구매가 요구된다. 공신력 있는 문서에 대한 접근이 제한되면 다른 경로를 통해 대체 자료를 찾아야 한다. 이를 위해 여러 출처에서 동일한 정보를 교차 확인하거나, 전문가 자문 등을 통해 정보의 정확성과 신뢰성을 검토하는 과정이 필요하다. 그러나 이러한 과정은 상당한 시간과 노력이 요구되며, 정보의 완전성을 보장하기 어렵다.

공신력 있는 파일포맷 문서를 확보하였음에도 평가자는 또 다른 어려움에 직면한다. 보존포맷 선정기준의 평가문항에 응답하는 근거로 사용할 정보를 분석할 수 있어야 하기 때문이다. 평가자는 수집한 파일포맷 문서 내에서 필요한 정보를 찾아내어 파일포맷이 평가문항을 충족하는지 판단할 수 있어야 한다. 그러나 이러한 분석 과정은 파일포맷에 관한 전문기술 지식이 없으면 수행하기 어렵다. 파일포맷 문서는 대부분 기술적 용어와 개념을 기반으로 작성되어 관련 지식이 부족한 경우, 문서에 명

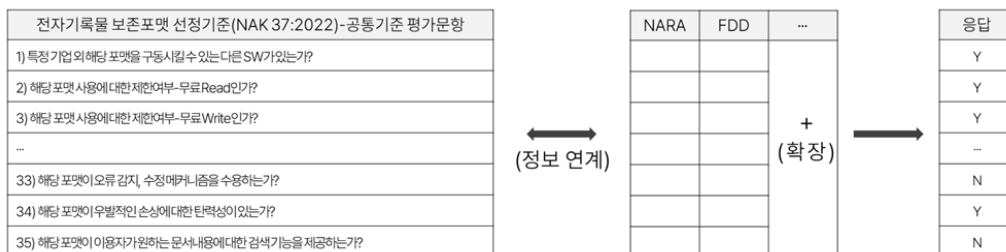
시된 파일포맷의 동작 원리나 구조 등을 이해하는 데 한계가 있다.

정리하면, 보존포맷 선정평가를 수행하기 위해 공신력 있는 정보를 수집하고 이를 해석하고 분석할 수 있는 전문기술 지식이 필수적이라는 점이 보존포맷 선정평가 과정을 어렵고 복잡하게 만든다. 보존포맷 선정평가 과정에서 발생하는 비효율성을 줄이기 위해서는 보존포맷 선정평가에 필요한 정보를 체계적으로 수집 및 분석하고, 평가자가 이해하기 쉬운 형태로 제공하는 지원도구가 필요하다.

3.2 전자기록물 보존포맷 선정평가 지원도구의 구조 및 기능 설계

보존포맷 선정평가 지원도구의 목적은 보존포맷 선정기준의 평가문항별 관련 정보를 제공함으로써 보존포맷 선정평가 과정의 효율성을 높이는 데 있다. 다음 <그림 3>은 본 연구에서 제안하는 보존포맷 선정평가 지원도구의 개념 구조를 나타낸다. 평가 지원도구는 파일포맷 관련 정보를 체계적으로 수집하고 이를 각 평가문항과 연계하여 정보를 직관적으로 제공하는 구조로 설계되어야 한다.

따라서 본 연구는 <표 2>와 같이 평가 지원



<그림 3> 보존포맷 선정평가 지원도구 개념 구조

〈표 2〉 전자기록물 보존포맷 선정평가 지원도구의 기능 요건

구분	기능	내용
정보 관리	정보 수집	파일포맷 관련 정보를 자동 또는 수동으로 수집하여, 이를 체계적으로 저장하는 기능
	정보 분석	수집된 파일포맷 정보를 분석하여 평가문항과 연계되는 정보를 추출·가공 하는 기능
평가 지원	보존포맷 선정평가표 생성	평가문항별 필요한 정보를 연계하여 보여주는 보존포맷 선정평가표를 생성하는 기능
	파일포맷 카탈로그 조회	파일포맷 관련 정보와 참고 URL 등을 카탈로그화하여 열람 및 검색을 지원하는 기능

도구의 기능을 크게 정보관리 영역과 평가지원 영역으로 구분하여 제시하였다. 정보관리 영역은 보존포맷 선정평가에 필요한 정보를 구축하는 내부적 기능을, 평가지원 영역은 보존포맷 선정평가 과정에서 평가자가 효율적으로 작업할 수 있도록 돕는 직접적인 서비스 기능을 포함한다.

3.2.1 정보관리 영역

정보관리 영역에는 보존포맷 선정평가에 필요한 정보를 수집하고, 이를 분석하여 필요한 정보만 추출 및 가공하는 기능이 포함된다. 가

장 먼저 평가 지원도구는 보존포맷 선정평가에 필요한 파일포맷 관련 정보를 확보하여 저장할 수 있어야 한다. 〈그림 4〉는 본 연구에서 제안하는 수집 기능의 구조와 흐름을 정리한 것이다. 이는 파일포맷 관련 정보 리소스에서 자동화 또는 반자동화된 방식으로 정보를 가져와 데이터베이스에 저장하는 과정을 나타낸다.

웹상에서 정보를 수집하는 대표적인 방법으로는 웹 크롤링(web crawling)과 API(Application Programming Interface)가 있다. 웹 크롤링은 크롤러(crawler)라는 도구를 사용하여 정보를 탐색 및 수집하는 기술로, 시작점으로 설정한



〈그림 4〉 보존포맷 선정평가 지원도구 정보 수집 기능의 구조

특정 웹페이지에 포함된 링크를 따라가며 필요한 정보를 체계적으로 수집한다. 이는 파일포맷 정보가 공개된 웹페이지에서 필요한 정보를 하나씩 가져오는 데 효과적으로 활용될 수 있다. API는 특정 요청에 대해 미리 구조화된 정보를 바로 제공하는 기술이다. 예를 들어, 파일포맷 정보를 제공하는 서비스의 API를 통해 특정 파일포맷에 대한 정보를 요청하면, 해당 정보가 구조화된 형태로 즉시 반환된다. 두 기술 모두 파일포맷 정보 수집에 활용될 수 있어 목적과 상황에 따라 적절한 방식을 선택할 수 있다. 다만, 정보를 수집하는 과정은 자동화할 수 있지만 수집 대상을 선정하거나 데이터베이스 스키마를 설계하는 과정에서는 여전히 사람의 판단과 개입이 필요하다. 그럼에도 불구하고 수작업으로 정보를 수집하는 과정을 자동화함으로써 보존포맷 선정평가에 소요되는 시간을 단축할 수 있고, 반복되는 정보 탐색 및 확보 과정을 생략하여 보존포맷 선정평가의 효율성을 높일 수 있다.

평가 지원도구는 수집된 정보에서 실제 보존포맷 선정평가에 필요한 내용을 자동 또는 수동으로 분석 및 추출하는 기능을 갖추어야 한다. 보존포맷 선정기준(공통, 고유)의 평가문항을 충족하는지 판단하기 위해서는 수집된 정보를 분석하고 해석하는 능력이 요구된다. 평가 지원도구에서 평가문항에 연관된 정보를 선별하여 제공하면, 평가자는 필요한 정보만을 직관적으로 확인하며 효율적으로 보존포맷 선정평가를 수행할 수 있을 것이다. 정보 분석과 추출의 대표적인 기술로는 내용 요약과 자동 색인 등이 있다. 이러한 기술은 자연어 처리(Natural Language processing, NLP)의 일종으로, 인

공지능 분야에서 주로 활용된다. 예를 들어 내용 요약 기술은 대량의 정보를 간결하게 정리하여 핵심 내용을 빠르게 파악하게 도울 수 있다. 자동 색인 기술은 수집된 정보에서 주요 키워드를 추출하고, 이를 기반으로 색인을 생성하는 데 활용될 수 있다. 특히, 추출된 키워드를 평가문항의 키워드와 연계하여 필요한 정보를 효율적으로 추출하고 가공할 수 있을 것이다.

정보 수집 기능과 정보 분석 기능은 평가 지원도구가 정보 연계 제공 서비스를 구현하는데 필요한 핵심 기반 작업이라 할 수 있다. 그러나 구체적인 기술 적용 방식은 실제 시스템 구현 단계에서 정해지는 사항으로, 본 연구에서는 이를 제안하기 어려운 한계가 있다. 또한 특정 기술이 적용될 가능성은 시스템 설계 과정에서의 다양한 요구와 상황에 따라 달라질 수 있다. 따라서 본 연구는 이러한 기능의 필요성을 제시하는 데 중점을 두었다.

3.2.2 평가지원 영역

평가 지원도구는 보존포맷 선정평가 과정을 처음부터 끝까지 포괄적으로 지원할 수 있어야 한다. 본 연구에서는 이를 위한 기능으로 보존포맷 선정평가표 생성과 파일포맷 카탈로그 조회 기능을 제안한다. 보존포맷 선정평가표는 평가를 진행하는 데 필요한 도구로, 평가문항에 대한 이원적 응답값(Yes/No)을 기록하고, 평가결과 점수를 계산할 수 있도록 설계되어야 한다. <그림 5>는 파일포맷별 맞춤형 보존포맷 선정평가표를 생성하고 등록하는 기능의 예시 모형이다.

평가를 시작하기에 앞서 기록물 유형, 파일포맷명, 선정기준 유형을 선택하여 조건에 맞



〈그림 5〉 보존포맷 선정평가표 생성 기능 예시 서비스 모형

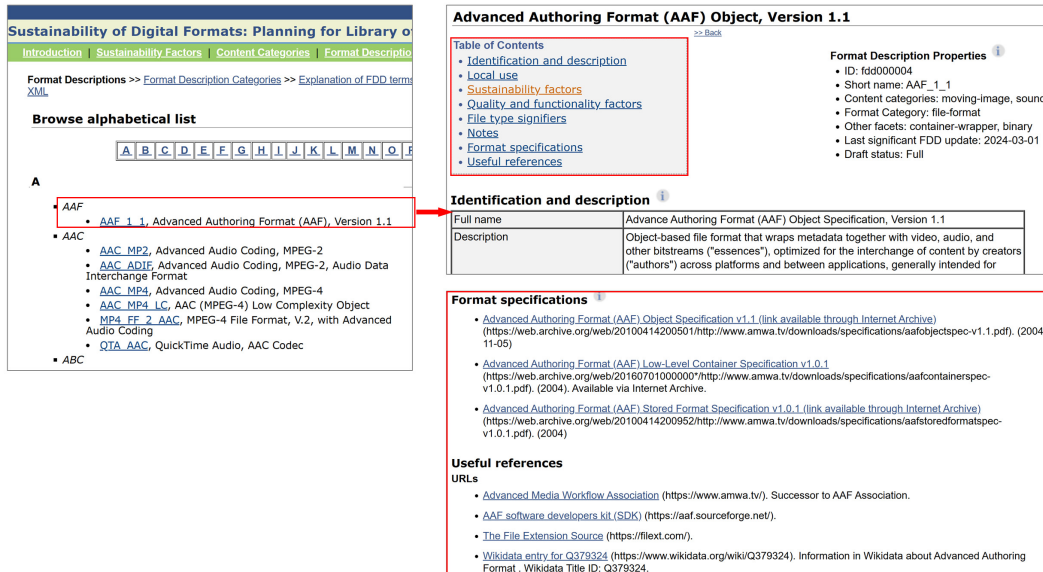
는 평가표가 생성되도록 설계하였다. 이는 파일포맷별로 평가문항에 응답하는데 필요한 정보가 다르기에 특정 파일포맷에 적합한 정보를 불러올 수 있도록 한 것이다. 그리고 각 평가문항에 대해서 평가자가 필요한 정보를 즉시 확인하여 Yes 또는 No를 선택해 응답할 수 있어야 한다. 평가가 완료되면 평가 지원도구가 자동으로 점수를 계산하여 기준 점수를 충족 여부를 확인할 수 있도록 제공해야 한다. 최종적으로, 보존포맷 선정평가표를 저장하여 이후에도 평가결과를 조회할 수 있어야 한다. 〈그림 6〉은 파일포맷 선정평가표 결과목록 조회 예시 모형으로, 이전에 수행하였던 보존포맷 선정평가 결과를 확인할 수 있다. 이 화면에서는 파일포맷별 공통기준 및 고유기준에 따른 평가결과표, 최종 보존포맷 선정 여부 등을 한눈에 확인할 수 있도록 통합적으로 구성하였다. 만약 여러 기관이 하나의 평가 지원도구 플랫폼을 공동으로 사용한다면, 보존포맷 선정평가

완료된 파일포맷의 결과를 공유할 수 있는 화면이 될 것이다. 이를 통해 다른 평가자는 같은 평가를 반복하지 않고도 기존 평가결과를 참고하여 업무 효율성을 높이는 환경을 조성할 수 있다.

또한 평가 지원도구는 평가문항별 정보 제공을 넘어, 평가자가 다양한 방식으로 정보를 열람할 수 있도록 파일포맷 카탈로그를 구축하여 제공할 수 있어야 한다. 〈그림 7〉은 미국 의회도서관이 제공하는 ‘디지털 파일포맷의 지속가능성(Sustainability of Digital Formats)’의 포맷 설명 문서(The format description documents, 이하 FDDs) 서비스 화면이다. FDDs는 정지 이미지, 오디오, 텍스트, 비디오 등 다양한 유형의 파일포맷에 대한 정보를 체계적으로 정리한 카탈로그를 공개하고 있다. 특히 하단에 파일포맷 사양서와 관련 참고자료의 URL을 제공하여 이용자들이 간략하게 가공된 정보 외에도 본 출처에 직접 접근할 수 있다. 이처럼 평가



〈그림 6〉 보존포맷 선정평가 결과 조회 예시 서비스 화면



〈그림 7〉 미국 의회도서관의 FDDs 서비스 화면

지원도구에서도 파일포맷 카탈로그를 구축하여 서비스하면 사용자가 파일포맷의 개요, 기술적 특성, 사양 등 종합적인 정보를 한곳에서 파악할 수 있을 것이다. 이를 위해 카탈로그에

포함될 정보 요소, 즉 메타데이터를 명확히 정의하는 것이 필수적이다. 예를 들어 공통기준과 고유기준의 각 평가문항에서 추출한 키워드를 메타데이터로 활용할 수 있다. 메타데이터

는 파일포맷 카탈로그의 검색 색인으로 사용되며, 키워드 검색기능을 통해 평가자가 특정 파일포맷 정보를 신속하게 찾을 수 있도록 한다. 또한 통합검색, 상세검색 등 다양한 검색 방식을 제공하여 사용자 편의를 높일 수 있다. 여기에 필터링, 정렬과 같은 부가기능을 추가하면 검색 효율성이 더욱 향상될 것이다.

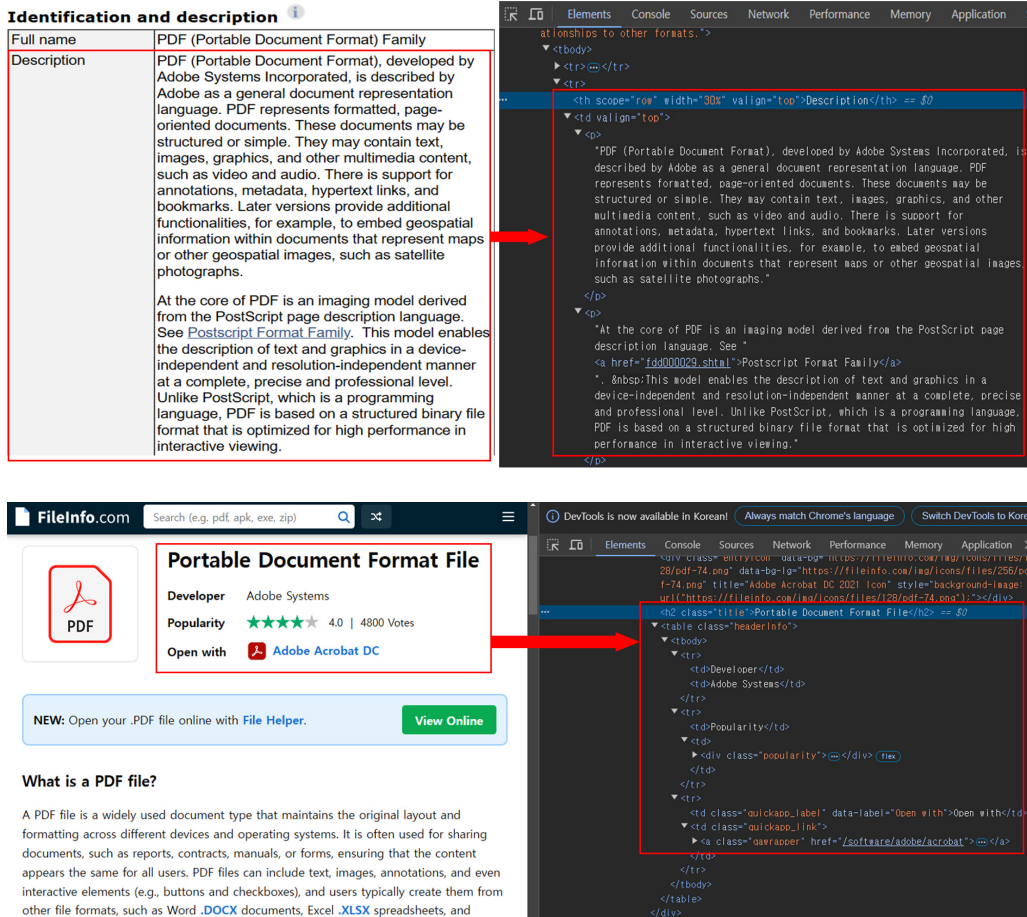
3.2.3 전자기록물 보존포맷 선정평가 지원도구 기능의 적용 가능성 검토

정보 수집과 정보 분석 기능은 보존포맷 선정평가의 효율성을 높이는 데 핵심적인 역할을 한다. 이는 보존포맷 선정평가를 지원하는 기능이 정보 제공을 목적으로 설계되었기 때문이다. 보존포맷 선정평가표 생성과 파일포맷 카탈로그 조회 기능은 평가자가 필요한 정보에 쉽고 효율적으로 접근할 수 있도록 지원하는 역할을 한다. 본 절에서는 이러한 기능들이 실제 환경에서 작동 가능하며 구현으로 이어질 수 있는지를 분석하였다.

정보 수집 기능은 파일포맷 관련 정보를 확보하고 저장하는 과정으로, 정보 리소스의 정보 제공 방식에 따라 적합한 수집 방식을 선택해야 한다. 본 연구에서는 정보 수집 기능의 실현 가능성을 검토하기 위해 여러 정보 리소스에서 데이터를 직접 수집하는 과정을 수행하였다. NARA의 위험 매트릭스(Risk Matrix)는 파일포맷의 위험 수준을 평가하는 질문과 그 결과를 엑셀 파일 형태로 공개하여 손쉽게 정보를 수집할 수 있었다. LOC의 FDDs와 FileInfo.com은 웹페이지에서 정보를 제공하고 있었으며 HTML 소스코드를 함께 공개하고 있었다. HTML 소스코드는 크롤러를 구축하

여 파일포맷별 정보를 추출하고 저장하는 데 활용된다. 예를 들어 <그림 8>과 같이 LOC의 FDDs Description(기술) 항목에 있는 내용을 수집하기 위해 크롤러에 HTML 태그 <th [...]> Description </th>와 그 하위에 있는 태그의 위치 경로를 입력하면 정보를 자동으로 추출하여 저장할 수 있다. FileInfo.com에서도 유사한 방식으로 파일포맷 관련 정보를 수집하였다. 이처럼 웹 기반 정보 리소스를 활용하여 실제 파일포맷 관련 정보를 직접 수집해 봄으로써 정보 수집 기능의 구현 가능성을 확인하였다.

정보 분석 기능은 수집한 정보 중 보존포맷 선정평가에 활용할 수 있는 유용한 정보를 추출하는 과정을 의미한다. 이는 단순히 정보 리소스에 파일포맷 정보가 존재하는지 확인하는 것을 넘어, 해당 정보가 실제 평가문항과 연계될 수 있는지를 검토하는 과정이다. 본 연구에서는 정보 분석 기능의 적용성을 검토하기 위해 NARA의 위험 매트릭스, LOC의 FDDs, FileInfo.com에서 수집한 정보를 대상으로 보존포맷 선정기준(공통) 평가문항과의 연계 가능성을 분석하였다. 각 평가문항에 필요한 정보를 대응시켜 정보 유효성을 확인한 것이다. 분석결과, 일부 문항에서 필요한 정보를 식별하거나 연계하는 데 한계가 있음을 확인하였다. 정보 분석의 적용성 검토 과정은 4장에서 구체적으로 다루었으며, 이를 통해 보존포맷 선정평가를 지원하기 위해 어떤 정보를 추출하여 연계하여야 하는지를 확인하였다. 정보 분석 기능은 평가 지원도구가 실질적인 평가 환경에서 작동할 수 있는 핵심 자원을 제공하는 데 필요한 기능이며, 보존포맷 선정평가의 효율성을 높이는 데 기여할 것으로 판단된다.



〈그림 8〉 LOC의 FDDs(위)와 FileInfo.com(아래)의 HTML 소스코드 제공 화면

보존포맷 선정평가표 생성 기능은 전자정부 표준프레임워크를 활용하여 구현할 수 있다. 전자정부 표준프레임워크는 대한민국 공공부문 정보화 사업에서 플랫폼 간 호환성과 표준화를 지원하기 위한 오픈소스 기반의 개발 프레임워크이다. 이 프레임워크는 자바 기반의 스프링(Spring) 프레임워크를 활용하며, 웹 애플리케이션 개발에 필요한 기능 및 아키텍처를 사전에 만들어 제공함으로써 효율적인 애플리케이션 구축을 지원한다. 표준프레임워크는 개

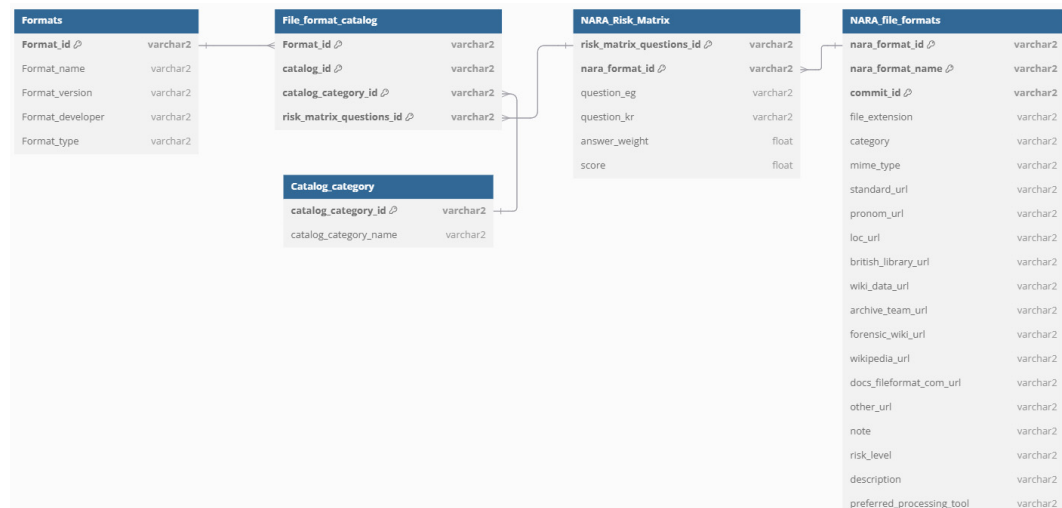
발환경, 실행환경, 관리환경, 운영환경 등 다양한 서비스 영역을 포함하며 사용자 화면 구성, 데이터베이스 연결 및 트랜잭션 관리 기능, 일관 데이터 처리를 위한 설정 기능 등 핵심 요소를 제공한다. 또한 개발자의 코드 작성을 돕는 도구와 테스트 도구 등을 지원한다. 이러한 이유로 우리나라 중앙정부 홈페이지를 포함한 대부분의 공공 웹 시스템이 이 프레임워크를 기반으로 구축되어 있다. 본 연구에서 제안한 〈그림 5〉, 〈그림 6〉과 같은 평가표 생성 기능을 구

현하기 위해서는 평가문항별 연계 정보를 데이터베이스에 저장하고, 해당 정보를 기반으로 선정평가표를 동적으로 생성하는 로직 등이 필요하다. 이는 표준프레임워크의 트랜잭션 관리 기능과 데이터 처리 기능 등을 통해 설계할 수 있다.

파일포맷 카탈로그 조회 기능은 여러 정보 리소스에서 수집한 정보를 파일포맷별로 통합 제공하는 서비스를 의미한다. 이 기능을 구현하기 위해서는 데이터베이스 구축이 필수적이다. 데이터베이스는 수집된 정보를 체계적으로 저장하고 이를 기반으로 정보 조회와 검색 등을 가능하게 하는 핵심 기반이다. 본 연구에서는 파일포맷 카탈로그 조회 기능의 적용 가능성을 검토하기 위해 NARA의 위험 매트릭스를 대상으로 데이터베이스 설계 및 구현 과정을 수행하였다. 먼저 <그림 9>와 같이 ERD(개체-관계 다이어그램)를 설계하고, 이에 맞게 정보를 저장하는 과정을 수행하였다. 또한 데

이터베이스의 조회 가능성을 평가하기 위해 Query를 통한 검색 가능성을 평가하여 필요한 정보를 효과적으로 추출하여 제공할 수 있음을 확인하였다. 이와 같은 결과는 파일포맷 카탈로그 조회 기능의 구현 가능성을 뒷받침하며, 다양한 정보 리소스에서 수집한 정보를 통합적으로 관리하고 제공할 수 있는 데이터베이스 기반 설계의 중요성을 보여준다. 나아가 이 기능은 보존포맷 선정평가 과정에서 정보 검색의 복잡성을 줄이고, 평가자가 효율적으로 작업할 수 있는 환경을 제공할 수 있을 것이다.

본 절에서는 정보 수집, 정보 분석, 보존포맷 선정평가표 생성, 파일포맷 카탈로그 조회 등 네 가지 기능을 중심으로 적용성을 검토하였다. 그 결과, 각 기능이 구현 가능함을 확인하였으며, 평가 지원도구가 보존포맷 선정평가 과정에서 효율성을 높이고 실질적으로 활용될 수 있음을 파악하였다.



<그림 9> NARA의 Risk Matrix의 ERD

4. 전자기록물 보존포맷 선정평가를 위한 정보 유효성 검토

본 장에서는 3장에서 제시한 평가 지원도구의 정보 분석 기능에 대한 적용성을 확인하기 위해 정보 유효성을 검토하였다. 먼저 보존포맷 선정기준(공통)의 평가항목을 해석하였다. 이후 NARA의 위험 매트릭스(Risk Matrix), LOC의 FDDs, FileInfo.com 등 세 개의 정보 리소스를 분석 대상으로 선정하여, 각 평가문항에 필요한 정보를 해당 정보 리소스에서 확인할 수 있는지 비교·분석하였다. 이 과정에서 평가문항의 요구사항을 충족하며 평가문항과 연계 가능한 정보 존재 여부를 평가하여 평가 지원도구의 실효성을 확인하고자 하였다.

보존포맷 선정기준은 공통기준과 고유기준으로 이루어져 있다. 고유기준은 전자기록물의 유형별 고유속성을 평가하는 데 초점을 맞추므로, 본 연구에서는 모든 기록물에 일관되게 적용해야 하는 공통기준의 평가문항을 비교 지표로 삼아 정보 연계성을 분석하였다. 보존포맷 선정기준(공통)은 <표 3>과 같이 개방성, 상호

운용성, 채택, 기능성 등 4가지 상위기준으로 구성되며, 각 상위기준에는 세부기준과 관련 평가문항이 포함되어 있다.

분석결과, 평가문항에 대한 응답 근거로 활용할 수 있는 정보는 '직접적 정보(○)', '간접적 정보(△)', '정보 없음(X)'으로 분류할 수 있었다. 직접적 정보는 평가문항에 긍정으로 응답할 수 있는 정보를, 간접적 정보는 평가문항에 응답하는데 간접적으로 참고할 수 있는 정보를 의미한다. 보존포맷 선정기준(공통)의 평가항목별 정보 연계표는 [부록]으로 제시하였다.

4.1 개방성

4.1.1 공개가용성

공개가용성 세부기준은 파일포맷의 접근 용이성을 평가하는 7개 문항으로 구성되어 있으며, <표 4>는 평가항목별 정보 연계 현황을 정리한 것이다.

1번 문항은 파일포맷을 개발한 기업이나 조직에서 운영하는 소프트웨어 외에도 해당 파일 포맷을 지원하는 다른 소프트웨어가 존재하는

<표 3> 전자기록물 보존포맷 선정기준(공통) 구성

공통기준	내용	평가문항 수
개방성	<ul style="list-style-type: none"> 파일포맷이 누구나 접근할 수 있고 투명하게 공개된 상태로 유지되는지 평가하는 기준 세부기준: 공개가용성, 공표 	11개
상호 운용성	<ul style="list-style-type: none"> 파일포맷이 다양한 소프트웨어, 운영체제, 하드웨어 환경에서 일관되게 사용될 수 있는지 평가하는 기준 세부기준: 독립성, 호환성, 변환가능성 	11개
채택	<ul style="list-style-type: none"> 파일포맷이 보편적으로 사용되고 특정 분야에서 표준으로 권장되는지 평가하는 기준 세부기준: 편재성, 편중성(예외적 적용) 	7개
기능성	<ul style="list-style-type: none"> 파일포맷이 전자기록물의 이용가능성에 연관된 기능적 특성을 갖추었는지 평가하는 기준 세부기준: 자체문서화, 기계가독성, 보호메커니즘, 검색기능 	8개

〈표 4〉 보존포맷 선정기준(공통)의 개방성-공개가용성 평가문항별 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
1	특정 기업 외 해당 포맷을 구동시킬 수 있는 다른 SW가 있는가?	△	△	○
2	해당 포맷 사용에 대한 제한 여부 (라이선스, 구독, 특허료 등)	무료 Read인가?	△	○
3		무료 Write인가?	△	○
4	기본 도구를 통해 해당 포맷을 구성하는 콘텐츠 전체를 해석할 수 있는가?	X	X	○
5	기본 도구(메모장, 그림판 등) 사용을 통한 분석가능 여부	텍스트 콘텐츠가 표준 문자 인코딩(UTF-8, 유니코드, 아스키 코드 등)으로 되어 있는가?	△	△
6		압축되어 있는 경우 표준 압축(zip, gzip, lzw 등)으로 되어 있는가?	X	○
7		멀티미디어 콘텐츠가 공개 포맷(jpeg, gif, mpeg 등)으로 되어 있는가?	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

지를 평가한다. 특정 기업의 의미가 명확하지 않으나 파일포맷을 개발한 기업이나 조직으로 해석하였다. 예를 들어 PDF는 어도비(Adobe)에서 개발하였으며 어도비는 PDF를 지원하는 소프트웨어인 아크로벳(Acrobat)을 배포하고 있다. 이때 아크로벳 말고도 PDF를 처리할 수 있는 소프트웨어가 있는지 확인하는 것이다. 2번, 3번 문항은 파일을 열어 내용을 확인하거나(read) 생성 및 편집(write)할 수 있는 소프트웨어가 무료로 배포되고 있는지를 평가한다.

4번에서 7번 문항은 운영체제에 기본으로 설치된 소프트웨어(일명 기본도구)를 통해 파일을 처리할 수 있는지를 평가한다. 이는 별도의 기술 지원 없이도 전자기록물의 내용에 접근하고 재현할 수 있는지 판단하기 위함이다. 4번 문항은 파일 내 저장된 데이터를 운영체제의 기본도구로 열 수 있는가를 평가한다. 이는 Windows, macOS 등에 기본으로 설치된 소프트웨어에서 파일포맷을 지원하는지를 보는 것이다. 5번 문항은 파일포맷의 텍스트 데이터 저장 방식이 표준 문자 인코딩을 따르는지 평가

한다. 문자 인코딩(character encoding)이란 기록물 생산자가 입력한 문자나 기호를 컴퓨터가 이용할 수 있는 이진 데이터로 변환하는 방식을 의미한다. 표준 문자 인코딩 사용 여부를 묻는 이유는 표준화된 방식을 사용할 경우 서로 다른 환경에서 파일을 열어도 깨짐 없이 텍스트를 일관되게 표시할 가능성이 크기 때문이다. 6번 문항은 파일포맷이 ZIP과 같은 표준 압축 포맷(아카이브 포맷)을 사용하여 여러 유형의 데이터(텍스트, 이미지, 오디오, 비디오 등)를 하나의 파일로 패키징하여 저장하도록 설계되었는지 평가한다. 대표적으로 HWPX, DOCX, PPTX, XLSX 포맷 등이 있다. 이러한 파일포맷의 특징은 파일에 담긴 여러 유형의 데이터를 개별 파일로 확인할 수 있다는 것이다. 예를 들어 *.hwpX 파일의 확장자를 *.zip으로 변경한 뒤 압축을 해제하면 본문의 텍스트 데이터는 *.xml 파일로, 이미지 데이터는 *.jpg 등 이미지 파일로 분리되어 확인할 수 있는지를 보는 것이다. 다만, 해당 문항은 파일포맷의 저장 규격에 압축 포맷이 사용되는지가 아닌 이미

생성된 파일을 압축할 때 표준 압축 방식을 사용할 수 있는지로 해석될 모호성을 가진다. 이에 평가문항 수정 및 보완이 필요하다. 7번 문항은 6번과 같이 표준 압축포맷을 사용하여 패키징하는 파일포맷이 시청각 데이터를 공개포맷(open format)으로 저장하도록 설계되었는지 평가한다. 그러나 문서 내 시청각 데이터는 생산자가 첨부한 파일의 형태 그대로 저장되기에 공개포맷으로만 저장된다고 단정하기 어렵다. 따라서 해당 문항에 대한 수정 및 보완이 필요하다.

공개가용성 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 1번에서 3번 문항에 응답하기 위해서는 파일포맷을 지원하는 소프트웨어의 다양성과 사용 비용 여부에 관한 정보가 필요하다. NARA의 위험 매트릭스에서는 파일포맷에 따른 파일을 렌더링할 수 있는 소프트웨어가 있는지, 있다면 그중 하나 이상이 오픈소스인지에 대한 질문과 결과 점수를 확인할 수 있었다. 그러나 파일포맷을 지원하는 소프트웨어가 여러 개인지 알 수 없었으며, 오픈소스 소프트웨어는 대체로 무료로 배포되지만, 항상 무료라고 단정할 수 없기에 '△'로 표시하였다. LOC의 FDDs에서는 일부 파일포맷에 한해서만 Description(기술), Adoption(채택) 항목에서 지원하는 소프트웨어 종류와 사용 비용을 확인할 수 있어 '△'로 표시하였다. 반면 FileInfo는 Program List 섹션에서 파일포맷을 지원하는 여러 종류의 소프트웨어와 각 소프트웨어를 무료로 사용할 수 있는지에 대한 정보가 명시되어 있었다.

4번 문항에 응답하기 위해서는 각 운영체제에 기본으로 탑재되는 소프트웨어에서 해당 파일포

맷을 지원하는지에 대한 정보가 필요하다. 이에 대해 FileInfo의 Program List 섹션에서 운영체제에 기본으로 설치된 소프트웨어를 Included with OS(운영체제에 포함됨)로 표시하여 제공하고 있음을 확인하였다. 5번 문항의 경우 파일포맷이 어떤 문자 인코딩을 사용하며, 그 인코딩 방식이 표준인지에 대한 정보가 필요하다. NARA의 위험 매트릭스에서 관련 정보를 확인할 수 있었으나 단순히 문자 기반 파일포맷인지에 대한 질문만 존재하여 표준 문자 인코딩 사용 여부는 알 수 없어 '△'로 표시하였다. LOC의 FDDs는 텍스트 유형 파일포맷에 한정하여 Quality and functionality factors(품질 및 기능 요소)의 Normal rendering(일반 렌더링) 항목에서 문자 인코딩 정보를 확인할 수 있었다. 6번 문항은 특정 압축포맷으로 패키징하는 파일포맷인지에 대한 정보가 필요하며, 이는 LOC의 FDDs에서 해당 조건을 충족하는 파일포맷에 한정하여 관련 정보를 확인할 수 있었다. 7번 문항에 응답하기 위해서는 파일포맷에 따른 파일을 생성하는 소프트웨어에서 지원하는 파일포맷의 종류를 확인해야 한다. 그러나 분석 대상으로 선정한 세 정보 리소스에서 해당 정보를 확인할 수 없었다.

4.1.2 공표

공표 세부기준은 파일포맷의 사양과 구조가 공개되었는지를 평가하는 4개 문항으로 구성되어 있으며, <표 5>는 평가항목별 정보 연계 현황을 정리한 것이다.

8번에서 10번 문항은 파일포맷의 기술 사양과 설계 구조가 문서화되어 공개되고 있는지, 파일포맷의 사양서(specification)에 접근 가능

〈표 5〉 보존포맷 선정기준(공통)의 개방성-공표 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용		NARA	LOC	FileInfo
8	해당 포맷의 '표준' 존재 여부	해당 포맷의 표준을 인터넷 등을 통해 공개적으로 참조 및 이용이 가능한가?	○	○	X
9		해당 포맷의 표준을 인터넷 등을 통해 공개적으로 참조 및 이용할 때 무료인가?	X	△	X
10		체계적이고 권위 있는 기관에 의해 표준화 과정을 거쳤는가?	○	△	X
11	해당 포맷의 '공개코드' 존재 여부	해당 포맷이 오픈 소스 라이선스인가?	○	○	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

하며 이를 이용할 수 있는지, 파일포맷이 국제적으로 인정받는 표준 기구에 의해 표준으로 승인되었는지를 평가한다. 다만, 8번과 9번 문항에서 사용된 '표준'이란 용어는 표준 문서를 의미하기 때문에, 표준화 과정을 거치지 않은 파일포맷의 경우 공개된 사양서가 있어도 두 문항에 긍정으로 응답할 수 없게 된다. 10번 문항은 어떤 기관이 권위 있는 기관으로 인정되는지에 대한 조건이 명시되어 있지 않다는 모호성이 있다. 따라서 명확한 보존포맷 선정평가를 위해 8번에서 10번 문항은 수정 및 보완될 필요가 있다. 11번 문항은 파일포맷이 오픈소스 라이선스를 따르는지 평가한다. 해당 문항에서는 파일포맷에 오픈소스 라이선스가 적용되어 있는지, 파일포맷을 지원하는 소프트웨어의 소스 코드가 오픈소스로 배포되는지 등을 확인한다.

공표 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 파일포맷의 사양서가 공개되어 있으면 8번 문항은 긍정으로, 공개된 사양서를 열람하기 위해 비용을 지불해야 한다면 9번 문항은 부정으로, ISO와 같은 표준 기구에 의해 포맷이 표준화되었다면 10번 문항은 긍

정으로 응답하면 된다. NARA의 위험 매트릭스에는 파일포맷의 사양서가 있는지, 있다면 그 사양서는 국제적으로 인정받는 표준 기구에 의해 승인 및 공표되었는지에 대한 질문을 확인할 수 있었으나 사양서의 비용 여부에 대한 정보는 확인할 수 없었다. LOC의 FDDs에서는 파일포맷의 사양서를 확인할 수 있는 웹페이지의 URL을 제공하고 있으나 사양서의 열람 비용에 대한 정보는 제공하지 않았다. 그러나 FDDs에서 제공하는 사양서의 URL로 이동하여 사양서가 유/무료로 제공되는지 간접적으로 확인할 수 있었다. 또한, 파일포맷이 권위 있는 표준 기구에 의해 표준화 과정을 거쳤는지에 대한 정보는 없었으나 파일포맷의 표준명을 통해 어느 표준 기구에서 표준화한 것인지 간접적으로 확인할 수 있었다. 11번 문항에 대한 정보는 NARA의 위험 매트릭스와 LOC의 FDDs에서 확인할 수 있었다. NARA의 위험 매트릭스에는 파일포맷이 오픈소스 라이선스 조건을 가졌는지에 대한 질문이 있었으며, LOC의 FDDs에서는 Licensing and patents(라이선스 및 특허권) 항목에 라이선스에 관한 정보가 존재하였다.

4.2 상호운용성

4.2.1 독립성

독립성 세부기준은 파일을 처리(생성·저장·편집·열기)하기 위해 특정 운영체제, 소프트웨어, 하드웨어 등에 기술적으로 의존해야 하는지를 평가하는 4개의 문항으로 구성되어 있다. <표 6>은 평가항목별 정보 연계 현황을 정리한 것이다.

12번 문항은 Windows, macOS 등 여러 운영체제에서 파일을 처리할 수 있는지 평가한다. 이는 운영체제에 파일포맷의 지원 소프트웨어를 설치할 수 있는지를 통해 파악할 수 있다. 13번과 14번 문항은 파일이 별도의 전문 장비나 특수한 하드웨어 없이 일반적인 컴퓨터 환경에서 처리될 수 있는지를 평가하여 하드웨어 의존성을 확인한다. 15번 문항은 파일포맷이나 지원 소프트웨어를 실행하는데 코덱, 플러그인 같은 특정 기술이나 DVD 플레이어와 같은 특정 장치가 필요한지를 평가한다. 그러나 하드웨어 의존성은 이미 13번 문항에서 다루고 있어 일부 내용이 중복되므로 문항 개선이 필요하다.

독립성 세부기준의 평가항목별 정보 연계 결

과는 다음과 같다. NARA의 위험 매트릭스는 하드웨어 측면에서 그래픽 카드, 칩셋, 메모리 등과 같은 특정 하드웨어 환경이 필요한지를, 소프트웨어 측면에서 파일포맷이 여러 운영체제에서 렌더링될 수 있는지, 렌더링을 위해 플러그인이나 스크립트 등이 추가로 필요한지를 평가하는 질문이 존재함을 확인하였다. LOC의 FDDs에는 External dependencies(외부 의존성) 항목이 있으나 대부분 'None(없음)'으로 작성되어 있었다. 그리고 파일포맷을 지원하는 운영체제에 대한 정보는 Description(기술) 또는 Adoption(채택) 항목에서 발견할 수 있었다. 하지만 극히 일부의 포맷에서만 관련 정보를 확인할 수 있어 '△'로 표시하였다. FileInfo의 경우, Program List 섹션에서 운영체제별 파일포맷의 지원 소프트웨어 목록을 제공하고 있어 파일포맷이 다수의 운영체제에서 처리될 수 있는지 확인할 수 있었다.

4.2.2 호환성

호환성 세부기준은 파일포맷이 여러 소프트웨어에서 일관되게 지원되는지 평가하는 4개의 문항으로 구성되어 있으며, <표 7>은 평가항목

<표 6> 보존포맷 선정기준(공통)의 상호운용성-독립성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용		NARA	LOC	FileInfo
12	OS 관점	해당 포맷을 구동할 수 있는 OS의 개수가 다수인가?	○	△	○
13	HW 관점	해당 포맷을 특별한 HW없이 구동할 수 있는가?	○	X	X
14		해당 포맷을 개인용 컴퓨터 수준의 HW에서 구동할 수 있는가?	X	X	X
15	특정 기술, 표준, 부가SW	해당 포맷 또는 구동 SW에 특수 코덱 및 특수 플레이어와 같은 특정 기술이나 부가 SW 등의 영향이 없는가?	○	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

〈표 7〉 보존포맷 선정기준(공통)의 상호운용성-호환성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
16	해당 포맷이 현재 구동 SW에서 지원하는가? (동일한 SW(같은 제조사, 계열사, 인수회사 등)에 한함)	X	△	○
17	해당 포맷이 이전/이후 구동 SW 버전과 호환이 가능한가? (동일한 SW(같은 제조사, 계열사, 인수회사 등)에 한함)	X	X	X
18	해당 포맷은 구동하는 SW의 배포(Release) 주기에 따라 형식이나 사양이 자주 업데이트되는가? (현재 가장 대표성 있는 구동 SW)	X	X	X
19	해당 포맷의 버전 업데이트 개발 로드맵 또는 계획이 존재하는가?	X	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

별 해당하는 정보 연계 현황을 정리한 것이다.

16번 문항은 파일포맷이 개발사(또는 관련 계열사)가 배포한 여러 소프트웨어에서 일관되게 지원되는지를 평가한다. 예를 들어 어도비가 개발한 파일포맷이 어도비가 배포하는 소프트웨어(Photoshop, Acrobat 등)에서 지원되는지 보는 것이다. 17번 문항은 파일포맷이 파일포맷 개발사(또는 관련 계열사)가 운영하는 지원 소프트웨어의 버전 차이와 관계없이 상호 호환성을 유지하며 일관되게 지원되는지를 평가한다. 가령 HWP 1.0 버전과 HWP 1.2 버전의 두 *.hwp 파일이 한컴오피스 2024 버전과 2022 버전에서 문제없이 열리는지를 확인할 수 있다. 다만, 평가 시점에서 가장 최신 버전을 기준으로 바로 이전 버전과 비교하는 것이 적절하기에, '이전/이후 버전'이라는 표현이 평가 기준을 모호하게 만들 수 있다. 따라서 이 표현에 대한 수정과 명확한 기준 설정이 필요하다. 18번 문항은 파일포맷을 지원하는 대표적인 소프트웨어의 버전이 업데이트될 때마다 파일포맷의 구조나 사양이 이에 맞춰 변경되는지를 평가한다. 그러나 이 문항은 논리적으로 잘못된 문항이라 볼 수 있다. 파일포맷은 소프트웨어와 무관하게 독립적으로 관리되는 것이 일반

적이며, 파일포맷의 구조와 사양을 기준으로 소프트웨어가 개발된다. 따라서 소프트웨어의 버전 변화에 영향을 받아 파일포맷의 구조나 사양이 변한다고 보기 어렵기에 문항 수정 및 보완이 필요하다. 19번 문항은 파일포맷의 개발 로드맵의 유무를 평가한다. 로드맵은 파일포맷의 업데이트 계획과 개발 방향성을 포함한 유지보수 계획이다. 여기에는 파일포맷의 호환성 관련 내용이 담겨 있을 가능성이 크다. 따라서 로드맵의 존재는 해당 파일포맷의 개발 기업이나 조직이 호환성을 고려하고 있음을 보여주는 근거가 될 수 있다.

세부기준의 평가항목별 정보 연계 결과는 다음과 같다. FileInfo에서는 파일포맷의 개발사를 확인하고, Program List 섹션의 지원 소프트웨어 목록에 파일포맷 개발사 소유의 소프트웨어가 존재하는지 확인하여 16번 문항에 응답할 수 있었다. LOC의 FDDs에서도 같은 방법으로 지원 여부를 확인할 수 있었으나, 모든 포맷에 대해 일관된 수준의 정보를 제공하지 않아 일부 파일포맷에 대해 개발사 소유의 지원 소프트웨어 여부를 알 수 없는 경우도 있었다. 17번에서 19번 문항의 경우, 분석 대상으로 선정된 정보 포털에서 관련 정보를 확인할 수 없었다.

4.2.3 변환가능성

변환가능성 세부기준은 파일포맷의 변환 가능 여부를 평가하는 3개의 문항으로 구성되어 있으며, <표 8>은 평가항목별 정보 연계 현황을 정리한 것이다.

20번 문항은 파일의 파일포맷을 다른 파일포맷으로 변환할 때 정보손실이 발생하는지를 평가한다. 파일포맷 변환 과정에는 정보손실의 위험이 상존한다. 파일포맷은 각기 다른 목적과 기술적 요구에 따라 설계되어 데이터 저장 구조, 지원 사양, 메타데이터 처리 방식, 인코딩 방식 등에 차이가 있다. 이로 인해 변환 대상 파일포맷이 원래의 파일포맷 사양을 동일하게 지원하지 못하거나 인코딩 방식의 차이로 데이터가 다르게 처리되어 저장될 수 있다. 따라서 정보손실 없이 파일포맷을 변환하는 건 불가능에 가까우며, 정보손실로 인정되는 범위가 필요하다. 21번 문항은 파일포맷이 다양한 종류의 파일포맷으로 변환될 수 있는지 평가한다. 22번 문항은 목적에 따라 기존 파일의 파일포맷을 다양한 도구나 소프트웨어에서 쉽게 활용될 수 있는 파일포맷으로 변환할 수 있는지를 평가한다. 다만, 문항에 사용된 표현이 모호하여 활용하기 쉬운 파일포맷으로의 변환이 아닌 원래의 파일포맷이 소프트웨어나 특정 도구들

통해 편집되고 활용될 수 있는지로 해석된다. 따라서 문항이 의도를 명확하게 표현할 수 있도록 보완이 필요하다.

변환가능성 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 20번 문항의 경우, 변환 과정에서 무엇을 정보손실로 간주해야 하는 기준이 없어 평가 근거로 활용할 정보를 찾지 못하였다. 21번 문항에 대해 FileInfo의 How to convert 섹션에서 변환 가능한 파일포맷 종류를 확인할 수 있었다. 하지만 다양함의 기준은 평가자가 판단해야 한다. 22번 문항은 어떤 목적을 위해 파일포맷이 다른 파일포맷으로 변환되어야 하는지, 어떤 파일포맷으로 변환해야 하는지, 어떤 도구나 소프트웨어에서 조작되고 재사용될 수 있는지에 대한 조건이 명시되어 있지 않다. 이로 인해 파일포맷의 변환 및 재사용성을 평가하기 어렵다고 판단하였으며, 정보 리소스에서도 관련 정보를 확인할 수 없었다.

4.3 채택

4.3.1 편재성

편재성 세부기준은 파일포맷이 널리 사용되어 접근이 용이한지 평가하는 5개의 문항으로 구성되어 있다. <표 9>는 평가항목별 정보 연계

<표 8> 보존포맷 선정기준(공통)의 상호운용성-변환가능성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용		NARA	LOC	FileInfo
20	보존, 추후 안정적인 마이그레이션 보장 가능성	해당 포맷이 정보의 손실없이 다른 포맷으로 변환 가능한가?	X	X	X
21		변환 가능한 포맷이 다양한가?	X	X	○
22	해당 포맷을 활용하기 쉬운 포맷으로 변환가능 여부 (AIP → DIP)	해당 포맷이 SW, 서비스 및 틀과 상호운용되어 새로운 목적으로 콘텐츠를 조작하고 재사용할 수 있는가?	X	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

〈표 9〉 보존포맷 선정기준(공통)의 채택-편재성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
23	OS에서 별도의 응용 SW 설치 없이 해당 포맷을 인식하고 내용을 확인할 수 있는가?	X	X	○
24	브라우저(Microsoft Edge, Internet Explorer, Chrome, Firefox 등)에서 별도의 확장 응용 SW 설치 없이 해당 포맷을 인식하고 내용을 확인할 수 있는가?	X	△	△
25	해당 포맷이 표준화 단체에 의해 표준화 과정을 거쳐 저명한 컨소시엄과 그룹에 의해 채택되어 전 세계에서 사용하는가?	X	△	X
26	해당 포맷이 시장을 선도하는가?	X	△	X
27	해당 포맷을 제작/조작/렌더링하는 많은 경쟁 제품이 존재하는가?	X	△	○

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

현황을 정리한 것이다.

23번 문항은 파일포맷이 별도의 소프트웨어 설치 없이 운영체제의 기본 소프트웨어에서 인식되고 처리될 수 있는지를 평가한다. 운영체제의 기본 소프트웨어에서 처리할 수 있다는 것은 파일포맷이 범용적으로 사용되어 파일의 내용에 쉽게 접근할 수 있음을 의미한다. 그러나 개방성-공개가용성 세부기준의 4번 문항과 내용이 중복된다. 24번 문항은 포맷이 특정 확장 프로그램을 설치할 필요 없이 웹 브라우저에서 바로 인식되어 내용을 확인할 수 있는지를 평가한다. 가령 이미지 포맷인 JPEG는 여러 운영체제에 기본으로 설치된 뷰어 소프트웨어에서 지원되고 있으며, Chrome, Edge, Safari 등 다양한 웹 브라우저에서 일반적으로 지원된다. 따라서 JPEG는 이미지 저장 및 공유를 위해 다양한 환경에서 보편적으로 사용되는 포맷이며, 높은 편재성을 가진다고 볼 수 있다. 25번 문항은 표준 단체에 의해 표준화 과정을 거친 파일포맷이 저명한 산업 연합에서 권장 사양으로 채택되어 국제적으로 널리 사용되는지를 평가한다. 이 문항은 개방성-공표 세부기준의 10번 문항에서 다루는 표준화 여부를 반복 평가하여 독점 포맷, 사실상 표준을 배제하게 만든

다. 그러나 저명한 산업 연합에서 채택된다는 것은 표준과 사실상 표준(De facto)을 모두 포함할 수 있어 표준화 여부와의 의미가 충돌된다. 또한 전 세계적으로 사용된다고 평가할 수 있는 구체적 기준이 없어 정성적 평가가 요구된다. 따라서 문항에 대한 수정과 명확한 기준 설정이 필요하다. 26번 문항은 파일포맷이 시장 내 우위를 점하고 있는지 평가한다. 이는 파일포맷이 특정 업계에서 기술적 영향력을 행사하거나 사용률이 높은지를 평가하는 문항이다. 다만, 이 역시 구체적 조건이 명시되지 않아 객관적으로 평가하기 어렵다는 문제가 있어 보완이 필요하다. 27번 문항은 파일포맷을 지원하는 다수의 경쟁 소프트웨어가 존재하는지를 평가한다. 파일을 생성, 편집하고 내용을 읽을 수 있도록 지원하는 소프트웨어가 다수 존재한다는 것은 이 포맷이 널리 사용되고 그만큼의 수요가 있음을 의미한다. 그러나 이 문항은 개방성-공개가용성 세부기준의 1번 문항에 '많은'이라는 표현만 추가될 뿐, 평가 의도는 동일하다고 볼 수 있다.

편재성 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 23번 문항은 개방성-공개가용성 세부기준의 4번 문항과 같이 FileInfo의

Program List 섹션에서 Included with OS(운영체제에 포함된)인 소프트웨어가 있는지 확인하면 된다. 24번 문항에 대한 웹 브라우저에서의 지원 여부는 LOC의 FDDs Adoption(채택) 항목에서, FileInfo의 How to Open 섹션에서 확인할 수 있었다. 그러나 두 곳 모두 확장 프로그램 필요 여부를 다루고 있지 않으며, LOC의 FDDs는 일부 포맷에 한정하여 관련 정보를 제공하고 있어 ‘△’로 표시하였다. 25번에서 27번 문항에 필요한 정보는 표준 여부, 파일포맷의 보편적 사용 여부, 특정 업계에서의 채택 여부 등 다양하다. 대부분 LOC FDDs의 Description(기술), Adoption(채택) 항목에서 관련 정보를 확인할 수 있었으나 이외의 항목에도 정보가 흩어져 있었다. 그러나 25번에서 27번 문항에 부정으로 응답되는 파일포맷에 대해서는 관련 정보를 제공하지 않았다.

4.3.2 편중성

편중성 세부기준은 파일포맷이 보편적으로 사용되지 않더라도 특수한 목적으로 사용되거나 특정 지역이나 분야에서만 전문적으로 사용될 경우 예외적으로 적용하는 세부기준이다. 필요에 따라 점수 비중을 조절하여 평가 결과에 영향을 줄 수 있으며, 편중성 있다고 판단되는 파일포맷은 평가와 상관없이 보존포맷으로

채택할 수 있다. 편중성 세부기준은 2개의 문항으로 구성되어 있으며, <표 10>은 평가문항별 정보 연계 현황을 정리한 것이다.

예외 1번 문항은 파일포맷이 기록 보존을 목적으로 채택되어 사용되는지를 평가한다. 이 문항에서 언급된 기관(국립도서관, 기록원, 기타 기록유산 기관)에서 채택한 파일포맷은 장기보존에 적합한 파일포맷일 가능성이 크므로, 보존포맷으로서 신뢰성을 갖춘 파일포맷이라 판단할 수 있다. 예외 2번 문항은 파일포맷이 특정 분야에서 업계 표준으로 사용되고 있는지를 평가한다. 이는 파일포맷이 특정 분야에서 반드시 사용되어야 하는 불가피성을 확인하기 위함이다. 파일포맷이 보편적으로 사용되지 않더라도 특정 목적에 있어서 필수적이거나 특정 분야에서만 독보적인 기능을 제공하는 경우, 그 파일포맷을 보존포맷으로 고려할 수 있는 근거가 될 수 있다.

편중성 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 예외 1번 문항에 대해 NARA의 위험 매트릭스에서는 파일포맷이 연방정부 내 영구기록물을 생성 또는 유지하는 데 사용된 적이 있는지 확인하는 질문이 있었다. 여기에는 채택이라는 표현은 없지만 파일포맷이 보존포맷으로 사용될 수 있는가로 해석할 수 있다고 판단하여 ‘△’로 표시하였다. LOC의 FDDs

<표 10> 보존포맷 선정기준(공통)의 채택-편중성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
예외 1	해당 포맷이 국립도서관, 기록원 및 기타 기록유산 기관이 공식적으로 채택한 보존포맷인가?	△	○	X
예외 2	해당 포맷이 특정 전문 분야의 커뮤니티에서 채택되어 사용되고 있는가?	X	△	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

Adoption(채택) 항목에는 도서관과 아카이브에서 수용 혹은 선호하는 파일포맷으로 채택하고 있는지 확인할 수 있는 정보가 존재하였다. 예외 2번 문항에 필요한 특정 업계에서 채택되어 사용되는가에 대한 정보는 LOC의 FDDs Adoption(채택) 항목에서 확인할 수 있었다. 그러나 이는 관련 분야에서 주로 사용되는 일부 파일포맷에서만 확인되었다.

4.4 기능성

4.4.1 자체문서화

자체문서화 세부기준은 파일을 재현하거나 내용을 이해하는데 필요한 메타데이터를 포함하는지 평가하는 3개의 문항으로 구성되어 있으며, <표 11>은 평가항목별 정보 연계 현황을 정리한 것이다.

28번 문항은 파일을 생성 또는 편집 시 미리 설계된 메타데이터 항목에 값이 자동으로 입력되고 저장되는지를 평가한다. 파일포맷마다 메타데이터 항목 구성이 다르며, 일부 항목은 자동으로 입력되도록 설계되어 있다. 일반적으로 파일포맷은 제목, 작성일, 수정일 등의 메타데이터를 자동으로 기록하는 기능을 지원한다. 29번 문항은 파일포맷이 사용자 정의 메타데이터를 수용할 수 있는지를 평가한다. 이는 파일포

맷이 확장 가능한 메타데이터 구조를 갖추어 사용자가 임의로 메타데이터 항목을 추가하고 값을 입력할 수 있도록 설계되었는지를 확인하는 것이다. 예를 들어, 이미지 포맷 JPEG는 EXIF(Exchangeable Image File Format) 표준을 사용하여 메타데이터 확장성을 지원하기에 *.jpeg 파일은 소프트웨어를 통해 사용자 정의 항목을 추가할 수 있다. 이처럼 메타데이터 확장성은 파일포맷이 특정 메타데이터 표준을 따르는지 또는 소프트웨어에서 메타데이터 추가 기능을 지원하는지를 통해 확인할 수 있다. 30번 문항은 소프트웨어가 파일포맷 내 메타데이터 구조를 이해하고, 그 내용을 추출하여 사용자에게 표시할 수 있는지 평가한다. 이 문항은 파일포맷 자체의 기능이 아닌 파일포맷을 지원하는 소프트웨어의 기능을 평가하므로, 소프트웨어를 통해 파일에 저장된 메타데이터를 확인할 수 있다면 긍정으로 답할 수 있다.

자체문서화 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 파일포맷에는 메타데이터를 저장하기 위한 특정 위치와 항목이 설계되어 있다. 28번 문항의 메타데이터 항목에 값이 자동으로 입력되는지에 대한 여부는 NARA의 위험 매트릭스와 LOC의 FDDs에서 확인할 수 있었다. NARA의 위험 매트릭스에는 파일 생성 시 기술적 환경을 기록하는 메타데이터를

<표 11> 보존포맷 선정기준(공통)의 기능성-자체문서화 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
28	해당 포맷이 자동 생성 메타데이터 기능을 제공하는가?	○	△	X
29	해당 포맷이 사용자 지정 메타데이터 기능을 제공하는가?	X	△	X
30	해당 포맷으로부터 메타데이터를 추출할 수 있는 기능을 지원하는가?	X	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

포함하는지, 수동 또는 자동화된 과정을 통해 설명적, 관리적 메타데이터를 내장하는 걸 지원하는지에 대한 질문이 존재하였다. LOC의 FDDs Self-documentation(자체 문서화) 항목에서 파일포맷의 메타데이터 관련 정보를 제공하고 있지만, 메타데이터 값이 자동으로 입력되는지에 대한 직접적 정보는 확인할 수 없어 '△'로 표시하였다. 29번 문항에 필요한 정보는 LOC의 FDDs Self-documentation(자체 문서화) 항목에서 확인할 수 있었다. 다만, 메타데이터 확장성을 직접 언급하기보다는 확장 가능한 메타데이터 규격을 따른다는 내용을 통해 간접적으로 유추할 수 있었다. 30번 문항 관련 정보는 분석 대상 정보 리소스에서 확인할 수 없었다. 다만, 파일 대부분은 지원 소프트웨어를 통해 메타데이터를 확인할 수 있으므로 '○'라고 봐도 무방하다.

4.4.2 기계가독성

기계가독성 세부기준은 데이터 분석과 인공지능 학습에 활용할 수 있는 파일포맷인가를 평가하는 1개의 문항으로 구성되어 있으며, <표 12>는 평가항목별 정보 연계 현황을 정리한 것이다.

31번 문항은 파일포맷이 데이터 기반 분석 및 학습을 목적으로 컴퓨터가 처리할 수 있는지를 평가한다. 컴퓨터가 처리할 수 있는 파일포

맷인지 평가할 때는 주로 기계가독성을 가진 파일포맷인지 확인한다. XML, CSV와 같이 정형 데이터를 담는 파일포맷은 기계가독성을 가져 31번 문항에 긍정으로 응답할 수 있다. 그러나 기계가독성이 없는 파일포맷이라도 추가 가공을 통해 데이터 분석과 학습에 활용 가능하다면 평가문항에 긍정으로 응답할 수 있다. 이 경우 HWPX, PPTX 등과 같이 파일에 포함된 이형의 데이터를 각각 분리할 수 있는 파일포맷도 해당 문항에 긍정으로 답할 수 있을 것이다. 정리하면, 파일포맷이 기계가 처리하기 쉬운 형태로 되어있는지가 핵심이지만 데이터 분석과 학습에 사용되는 데이터 유형(텍스트, 이미지, 오디오, 비디오 등)을 담는 파일포맷도 해당된다고 볼 수 있다. 기계가독성 세부기준의 평가문항별 연계 결과는 다음과 같다. NARA의 위험 매트릭스와 FileInfo에서는 관련 정보를 확인할 수 없었으나 LOC의 FDDs Transparency(투명성) 항목에서 일부 파일포맷에 대해 기계가 처리하기 쉬운 형태로 되어있다는 정보를 확인할 수 있었다.

4.4.3 보호메커니즘

보호메커니즘 세부기준은 파일포맷이 장기적 접근성과 무결성을 유지하면서 우발적 손상에 대체할 수 있는 능력을 갖추었는지 평가하는 3개의 문항으로 구성되어 있다. <표 13>은

<표 12> 보존포맷 선정기준(공통)의 기능성-기계가독성 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
31	해당 포맷이 데이터 분석, 인공지능 학습 등의 목적으로 컴퓨터가 처리할 수 있는 형태(기계가독성을 위한 요소(태깅 등) 포함, 개방형 파일 포맷, 표준화된 파일 포맷 등)로 설계되었는가?	X	△	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

〈표 13〉 보존포맷 선정기준(공통)의 기능성-보호메커니즘 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
32	해당 포맷이 암호 보호, 복사 방지, 디지털 서명, 인쇄 방지 및 콘텐츠 추출 보호와 같은 기술 보호 메커니즘이 적용되어 있지 않은가?	○	○	X
33	해당 포맷이 오류 감지, 수정 메커니즘을 수용하는가?	X	X	X
34	해당 포맷이 우발적인 손상에 대한 탄력성이 있는가?	X	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

평가항목별 정보 제공 현황을 정리한 것이다.

32번 문항은 파일포맷이 암호화나 복사 방지, 디지털 서명 등 정보 보안 메커니즘을 수용할 수 있는지를 평가한다. 보안 메커니즘은 정보에 대한 접근 통제, 변조 및 훼손 등을 방지하기 위한 목적으로 사용되지만, 장기보존 관점에서는 기록 활용 가능성을 저해하는 요소로 작용할 수 있다. 따라서 이 문항에서는 보안 메커니즘을 수용하는 포맷일 경우 부정으로 응답해야 한다. 33번 문항은 파일포맷이 데이터 손상을 감지할 수 있는 구조로 설계되었는지를 평가한다. 구체적으로, 파일포맷이 체크섬(checksum) 메커니즘이나 오류 감지를 위한 메타데이터 구조를 포함하고 있는지를 확인하는 것이다. 다만, 수정 메커니즘은 오류를 감지하고 자동으로 수정하는 과정을 포함하는데, 이는 파일포맷 자체의 기능이 아니라 파일을 처리하는 소프트웨어나 시스템의 기능이다. 따라서 33번 문항에서 평가하는 요소가 명확해지도록 수정 메커니즘과의 차이를 분명히 할 필요가 있다. 34번 문항은 파일포맷이 데이터 손상에 대처할 수 있는 복구 메커니즘을 가지는지를 평가한다. 그러나 데이터 손상 복구는 파일포맷이 아닌 소프트웨어에서 주로 수행되는 기능이다. 그럼에도 불구하고 데이터 손상 복구는 전자기록물 관리에서 매우 중요한 요소이다. 해당 문항은 파일포

맷의 복구와 관련된 구조적 특성과 이를 지원 하는 소프트웨어의 역할을 모두 고려하도록 수정 및 보완될 필요가 있다.

보호메커니즘 세부기준의 평가항목별 정보 연계 결과는 다음과 같다. 32번 문항에 필요한 정보는 파일포맷의 보안 메커니즘 수용 여부이다. NARA의 위험 매트릭스에서는 암호화를 사용해야 하는 포맷인지, DRM과 같은 기술적 보호 조치를 허용하는지 등을 확인하는 질문과 그 결과를 확인할 수 있었다. LOC의 FDDs는 Technical protection considerations(기술 보호 고려사항) 항목에서 포맷에 암호화 등과 같은 기술적 보호 기능을 허용하는지에 대한 정보를 제공하고 있음을 확인하였다. 33번, 34번 문항과 관련된 정보는 NARA의 위험 매트릭스, LOC의 FDDs, FileInfo에서 확인할 수 없었다.

4.4.4 검색기능

검색기능 세부기준은 파일에 저장된 내용을 검색할 수 있는지 평가하는 1개의 문항으로 구성되며, 〈표 14〉는 평가항목별 정보 연계 현황을 정리한 것이다.

35번 문항은 파일포맷에 검색기능이 있는지를 평가한다. 이는 파일 내용을 검색할 수 있는지 확인하는 것이다. 그러나 실제 검색은 파일

〈표 14〉 보존포맷 선정기준(공통)의 기능성-검색기능 평가문항 관련 정보 연계 현황

no.	평가문항 구성 내용	NARA	LOC	FileInfo
35	해당 포맷이 이용자가 원하는 문서내용에 대한 검색 기능을 제공하는가?	X	X	X

○: 직접적 정보 확인 가능, △: 간접적 정보 확인 가능, X: 정보 없음

포맷이 아닌 소프트웨어나 시스템을 통해 이루어짐을 유의해야 한다. 예를 들어 Windows에서는 파일탐색기를 통해 PPT(X), DOC(X), XSL(X)과 같은 자사 파일포맷이나 TXT, HTML 등 텍스트 파일의 내용을 검색할 수 있다. 따라서 검색기능을 평가하려면 운영체제가 어떤 파일포맷에 대해 색인을 지원하는지를 확인해야 한다. 그러나 NARA의 위험 매트릭스, LOC의 FDDs, FileInfo에서는 해당 문항과 관련된 정보를 확인할 수 없었다.

4.5 시사점

평가 지원도구의 목적은 보존포맷 선정평가에 필요한 정보를 효과적으로 제공하는 것이다. 이를 위해 파일포맷과 관련된 정보가 먼저 확보되어야 한다. 단순히 정보를 수집하는데 그치지 않고 수집된 정보 중에서 보존포맷 선정기준의 평가항목에 응답하는데 필요한 정보를 분석 및 추출하여 제공하는 것이 중요하다. 본 연구는 보존포맷 선정기준(공통) 평가문항별로 필요한 정보를 확인하고, 정보 연계 가능성을 분석함으로써 정보 유효성을 검토하였다. 그 결과, 평가 지원도구가 보존포맷 선정평가 과정의 복잡성과 어려움을 줄이는데 기여할 수 있음을 확인하였다. 이는 시간과 자원이 많이 소요되며 일관성 있는 평가를 보장하기 어렵다는 한계를 해결하고, 나아가 보존포맷 선정체계가 실무에 효과

적으로 정착하는 데 도움을 줄 수 있음을 시사한다. 더불어 정보 유효성을 평가하는 과정과 결과를 통해 다음과 같은 시사점을 도출할 수 있었다.

첫째, 보존포맷 선정기준 평가항목의 개선이 필요하다. 정보 유효성 평가결과, 보존포맷 선정기준(공통)의 37개 평가문항(예외적 적용 문항 포함) 중 18개 문항은 세 정보 리소스에서 제공하는 정보를 종합하여 명확한 응답이 가능하였다. 반면, '△'로 표시된 8개의 평가문항은 연계 가능한 일부 간접적 정보를 평가자가 주관적으로 해석하여 응답값을 결정해야 했다. 나머지 11개 문항은 분석 대상으로 선정한 세 정보 리소스에서 연계 가능한 정보를 확인할 수 없었다. 이러한 연계 가능성의 차이는 주로 선정기준 평가문항에 의해 결정된다고 볼 수 있었다. 평가문항의 의미를 해석하는 과정에서 일부 평가문항에 단어로 인한 모호성, 구체적이지 않은 범위/기준에 의한 모호성, 논리적 오류로 인한 모호성이 있음을 발견하였다. 이러한 문제는 평가문항이 여러 의미로 해석될 여지를 남겨 보존포맷 선정평가의 일관성을 저해함과 동시에 정보 연계를 어렵게 만드는 주요 요인으로 작용하였다. 보존포맷 선정평가에 필요한 정보를 확인하기 어려웠던 평가문항 대부분이 이러한 모호성을 가지고 있다는 문제점이 있었다. 이는 평가 지원도구의 한계라기보다는 보존포맷 선정기준 평가문항의 개선 필요성을

보여주는 결과로 해석된다. 따라서, 평가 지원 도구와는 별개로 평가문항의 명확성과 적합성을 강화하기 위한 개선이 선행되어야 함을 시사한다.

둘째, 평가 지원도구의 정보 수집 대상을 다양하고 폭넓게 설정할 필요가 있다. 본 연구에서는 기록관리 및 보존 관점에서 파일포맷 관련 정보를 제공하는 NARA의 위험 매트릭스(Risk Matrix)와 LOC의 FDDs, 파일의 처리(생성, 편집 등) 관점에서 파일포맷 관련 정보를 제공하는 FileInfo.com을 정보 유효성 검토에 활용하였다. 이 세 정보 리소스는 모두 파일포맷에 관한 정보를 제공한다는 공통점이 있으나 제공하는 정보의 내용이나 수준에서 차이를 보였다. 하나의 파일포맷에 대해 보존포맷 선정평가를 진행할 때, 각 평가문항에 필요한 정보가 모든 리소스에서 균등하게 제공되지 않았다는 점을 확인하였다. 예를 들어 파일포맷을 지원하는 소프트웨어가 있는지를 묻는 문항에 대해 NARA의 위험 매트릭스에서는 파일 렌더링 소프트웨어의 존재 여부만을 확인할 수 있었다. 반면, FileInfo.com은 운영체제별로 파일포맷을 지원하는 소프트웨어 목록을 체계적으로 제공하고 있다. LOC의 FDDs는 지원 소프트웨어를 명시하거나 관련 정보를 포함한 웹사이트를 하이퍼링크로 제공하지만, 일부 파일포맷에 대해서는 이러한 정보가 제공되지 않았다. 이는 정보 리소스별로 정보의 유무와 설명 정도가 상이함을 보여준다. 평가 지원도구의 정보 수집 대상이 특정 정보 리소스에만 의존할 경우, 일부 평가문항에 필요한 정보를 확보할 수 없거나 평가자의 주관적 해석에 의존해야 하는 상황이 발생할 수 있다. 따라서 평가 지원도구는 다양

한 관점과 범주에서 정보를 제공하는 정보 리소스들을 수집 대상으로 포함해야 하며, 이를 통해 평가문항별 요구사항에 부합하는 정보를 보다 풍부하게 확보해야 할 수 있어야 한다. 이러한 접근은 평가문항과 정보 간의 연계를 강화하고 보존포맷 선정평가의 일관성을 높이는 데 기여할 것이다.

셋째, 국내 고유의 파일포맷에 대한 자체적 정보를 구축할 필요가 있다. 조사 과정에서 파일포맷에 관한 정보를 통합적으로 제공하는 국내 웹사이트를 찾을 수 없었다. 국외 정보 리소스는 보존포맷 선정평가에 유용한 정보를 제공하지만, 국내에서만 사용되는 파일포맷에 대한 정보는 거의 포함되어 있지 않다. 예를 들어 한국에서 보편적으로 사용되는 HWP는 해외 정보 리소스에서 관련 정보를 제공하지 않거나, 제공된 정보가 보존포맷 선정평가에 필요한 수준에 미치지 못하는 경우가 대부분이었다. 따라서 국내 고유의 파일포맷에 대한 정보는 해외 정보 리소스에 의존할 수 없으며 평가 지원도구의 운영 및 관리기관이 별도로 수집하고 구축할 필요가 있다.

본 절에서는 평가 지원도구의 실효성을 높이기 위한 방안을 시사점으로 제시하였다. 평가문항의 명확성과 적합성을 개선하여 보존포맷 선정평가의 일관성을 확보하고, 정보 수집 대상을 확대하여 평가문항별 요구사항에 부합하는 정보를 제공해야 한다. 또한, 국내 고유 파일포맷에 대한 정보를 자체적으로 구축하여 평가 지원도구에서 활용할 수 있어야 한다. 이러한 방안은 평가 지원도구의 효율성을 강화하고 보존포맷 선정체계가 실무에 안정적으로 정착하는 데 기여할 것이다.

5. 결 론

기록관리 공공표준 「전자기록물 보존포맷 선정기준(v1.0)」이 제정됨에 따라, 특정 파일포맷이 보존포맷으로서 적합한지 확인하는 보존포맷 선정평가를 수행해야 한다. 보존포맷 선정평가는 보존포맷 선정기준(공통, 고유)의 평가 문항에 대한 이원적 응답(Y/N) 방식으로 이루어지며, 파일포맷 관련 기술적 정보를 근거로 한다. 따라서 보존포맷 선정평가를 수행하려면 필요한 정보를 우선으로 확보해야 한다. 그러나 이 과정에서 직면하게 되는 파일포맷 정보 접근의 제약과 확보된 정보를 해석하기 위한 전문기술 지식의 필요성은 평가를 어렵고 복잡하게 만드는 주요 요인으로 확인되었다.

이에 본 연구는 평가자가 정보를 직접 수집하거나 분석할 필요 없이 보존포맷 선정평가에 필요한 정보에 쉽게 접근할 수 있도록 평가 지원도구의 기능을 정보관리 영역(정보 수집, 정보 분석)과 평가지원 영역(보존포맷 선정평가표 생성, 파일포맷 카탈로그 조회)으로 구분하여 설계하였다. 이후 제안한 평가 지원도구의 네 가지 기능에 대한 적용성을 검토하였다. 특히 평가 지원도구의 핵심 역할이 보존포맷 선정평가에 필요한 정보 제공이므로 정보의 유효성을 검토하여 평가 지원도구의 실효성을 확인하였다. 분석 결과를 바탕으로 평가 과정의 효율성을 높이

고 정보 연계의 정확성을 강화하기 위해 보존포맷 선정기준(공통) 평가문항의 개선 필요성, 평가 지원도구의 정보 수집 대상 확대, 국내 고유의 파일포맷에 대한 정보 구축 등의 개선방안을 제시하였다. 이는 평가 지원도구가 보존포맷 선정평가 과정에서 효율적이고 일관되게 작동할 수 있는 기반을 마련하며, 평가자의 작업 부담을 줄이고 평가결과의 일관성을 높이는 데 기여할 것이다.

보존포맷 선정평가는 보존포맷 선정기준(공통, 고유)을 기반으로 이루어진다. 따라서 각 선정기준의 평가문항은 오류 없이 명확한 의도를 포함해야 평가 과정의 효율성과 일관성을 보장할 수 있다. 특히, 공통기준은 모든 파일포맷에 적용되는 평가의 출발점이 되므로, 각 평가문항이 명확하고 논리적으로 작성되어야만 평가 지원도구가 제 역할을 수행할 수 있다. 현재 보존포맷 선정체계는 기존에 제시되지 않았던 기록물 유형의 고유기준을 수립하고, 이를 통해 선정평가의 적용 대상을 확장하는 데 초점을 두고 있다. 그러나 이런 확장과 별개로 기존에 설계된 보존포맷 선정기준의 평가문항의 개선은 그동안 고려되지 않았다. 따라서 본 연구의 후속 연구로 공통기준의 개선방안에 관한 연구를 수행하고자 한다. 이는 보존포맷 선정평가의 정확성을 더욱 강화할 수 있을 것이다.

참 고 문 헌

- 공공기록물 관리에 관한 법률 시행령. 대통령령 제34487호.
공공기록물 관리에 관한 법률. 법률 제20309호.

- 국가기록원 (2008a). 전자기록물 문서보존포맷 기술규격 (NAK 30:2008, v1.0).
- 국가기록원 (2008b). 전자기록물 장기보존포맷 기술규격 (NAK/TS 3:2008, v1.0).
- 국가기록원 (2022). 전자기록물 보존포맷 선정기준 (NAK 37:2022, v1.0).
- 국가기록원 (2023). 시청각 및 데이터세트 유형 보존포맷 선정기준 개발 및 전자기록물 보존포맷 선정체계 구축 연구.
- 김지혜, 양동민 (2024). 비디오 유형 전자기록물의 보존포맷 선정기준 및 평가체계에 관한 연구. 한국 기록관리학회지, 24(1), 163-186. <https://doi.org/10.14404/JKSARM.2024.24.1.163>
- 소정의, 한희정, 양동민 (2018). 국외 전자기록물의 장기보존 정책 비교 분석: 미국, 캐나다, 영국, 호주, 스위스를 중심으로. 한국기록관리학회지, 18(4), 125-148. <https://doi.org/10.14404/JKSARM.2018.18.4.125>
- 송채은, 양동민 (2024). 이미지 유형 전자기록물의 보존포맷 선정체계 구축방안 연구. 기록학연구, 79, 343-387. <https://doi.org/10.20923/kjas.2024.79.343>
- 윤성호, 김지호, 양동민 (2022). 전자문서 보존포맷 선정기준을 위한 오피스스위트 분석 기반 기능 재분류. 디지털문화아카이브지, 5(1), 171-192. <https://doi.org/10.23089/jdca.2022.5.1.010>
- 이정은, 양동민 (2023). 데이터세트 유형 전자기록의 필수보존속성 연구. 한국비블리아학회지, 34(4), 259-283. <https://doi.org/10.14699/kbiblia.2023.34.4.259>
- 전한역, 김지혜, 김현태, 양동민 (2023). 시청각 유형 보존포맷 선정을 위한 필수보존속성 연구: 디지털 오디오를 중심으로. 디지털문화아카이브지, 6(2), 27-53. <https://doi.org/10.23089/jdca.2023.6.2.002>
- 전한역, 양동민 (2024). 시청각 유형 보존포맷 선정기준 및 권고포맷 연구: 오디오 유형을 중심으로. 한국비블리아학회지, 35(1), 273-300. <https://doi.org/10.14699/kbiblia.2024.35.1.273>
- 한능우 (2023). 전자기록 장기보존정책의 현황과 과제. 기록학연구, 78, 79-151. <https://doi.org/10.20923/kjas.2023.78.079>
- 한희정, 오효정, 양동민 (2020). 전자기록물의 장기보존을 위한 보존포맷 선정 방안에 관한 연구. 한국기록관리학회지, 20(1), 69-87. <https://doi.org/10.14404/JKSARM.2020.20.1.069>
- 한희정, 윤성호, 오효정, 양동민 (2020). 데이터세트 보존포맷 검증방안에 관한 연구: 재난안전정보 데이터세트의 SIARD 적용을 통해. 정보관리학회지, 37(2), 251-284. <https://doi.org/10.3743/KOSIM.2020.37.2.251>
- FileInfo.com (2005). FileInfo.com-The File Format Database. Available: <https://fileinfo.com/>
- Library of Congress (2004). Sustainability of Digital Formats: Planning for Library of Congress Collections Format Descriptions. Available: <https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>
- Library of Congress (2024). Recommended Formats Statement 2024-2025.

- NARA (2019). NARA Digital Preservation Risk Matix. Available:
https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Risk_Matrix
- NARA (2022). Digital Preservation Strategy 2022-2026.
- NARA (2024). 2024 NARA Risk Matrix Updates.
- TNA (2017). Digital Strategy 2017-2019.
- TNA (n.d.). The National Archives homepage-Information management. Available:
<https://www.nationalarchives.gov.uk/information-management/>

• 국문 참고자료의 영어 표기

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

- Enforcement Decree of the Public Records Management Act. Presidential Decree No. 34487
- Han, Hui-Jeong, Oh, Hyo-Jung, & Yang, Dong-min (2020). A study on the selection of preservation format for long-term preservation of electronic records. *Journal of Korean Society of Archives and Records Management*, 20(1), 69-87.
<https://doi.org/10.14404/JKSARM.2020.20.1.069>
- Han, Hui-Jeong, Yoon, Sung-Ho, Oh, Hyo-Jung, & Yang, Dong-min (2020). Empirical verification of conversion and restoration of preservation format for dataset: application of dataset with disaster safety information to SIARD. *Journal of the Korean Society for Information Management*, 37(2), 251-284. <https://doi.org/10.3743/KOSIM.2020.37.2.251>
- Han, Nung-woo (2023). Long-term preservation strategy of digital records. *Journal of Archival Studies*, 78, 79-151. <https://doi.org/10.20923/kjas.2023.78.079>
- Jeon, Han-yeok & Yang, Dong-min (2024). Research on audiovisual type preservation format selection criteria and recommended formats: focusing on audio types. *Journal of the Korean Biblia Society for Library and Information Science*, 35(1), 273-300.
<https://doi.org/10.14699/kbiblia.2024.35.1.273>
- Jeon, Han-yeok, Kim, Ji-Hye, Kim, Hyun-Tae, & Yang, Dong-min (2023). A study on significant properties for selection of audiovisual type preservation format: focused on digital audio. *Journal of D-Culture Archives*, 6(2), 27-53. <https://doi.org/10.23089/jdca.2023.6.2.002>
- Kim, Ji-Hye & Yang, Dong-min (2024). A study on selection criteria and evaluation system for preservation formats of video-type digital records. *Journal of Korean Society of Archives and Records Management*, 24(1), 163-186.

<https://doi.org/10.14404/JKSARM.2024.24.1.163>

- Lee, Jung-Eun & Yang, Dong-min (2023). A study on significant properties for dataset type preservation format. *Journal of the Korean Biblia Society for Library and Information Science*, 34(4), 259-283. <https://doi.org/10.14699/kbiblia.2023.34.4.259>
- National Archives of Korea (2008a). *Digital Document File Format for Long-term Preservation (NAK 30:2008, v1.0)*.
- National Archives of Korea (2008b). *Standard of Archival Information Package (NAK/TS 3:2008, v1.0)*.
- National Archives of Korea (2022). *Selection Criteria for Preservation Format of Digital Records (NAK 37:2022, v1.0)*.
- National Archives of Korea (2023). *Research on development of selection criteria for preservation formats of audiovisual and dataset-type records and establishment of a selection system of preservation formats for electronic records*.
- Public Records Management Act, Act No. 20309.
- So, Jeong-Eui, Han, Hui-Jeong, & Yang, Dong-min (2018). A comparative analysis of long-term preservation policies in foreign electronic records: NARA, LAC, TNA, NAA, and SFA. *Journal of Korean Society of Archives and Records Management*, 18(4), 125-148. <https://doi.org/10.14404/JKSARM.2018.18.4.125>
- Song, Chae-eun & Yang, Dong-min (2024). A study on the development of a selection system for preservation formats of image-type electronic records. *Journal of Archival Studies*, 79, 343-387. <https://doi.org/10.20923/kjas.2024.79.343>
- Yoon, Sung-Ho, Kim, Ji-ho, & Yang, Dong-min (2022). A study on re-classification of functions through OfficeSuite analysis to prepare criteria for selecting preservation format for electronic documents. *Journal of D-Culture Archives*, 5(1), 171-192. <https://doi.org/10.23089/jdca.2022.5.1.010>

[부록] 보존포맷 선정기준(공통)의 평가항별 정보 연계표

NAK-선정기준(공통) 범주	NARA-Risk Matrix*		LOC-FDDs**	FileInfo.com***
	범주	문항		
공개 가용성	1	(△) External Software Dependencies 6.1	(△) Sustainability Factors > Adoption에서 확인 가능	Program List에서 확인 가능
		6.2		
	2	• NARA-6.1 = 2점 ⇨ NAK-01 = Y	(△) Sustainability Factors > Adoption에서 확인 가능	Program List에서 확인 가능
		• NARA-6.2 = 1점 ⇨ NAK-01 = Y		
	3	(△) External Software Dependencies 6.2	(△) Sustainability Factors > Adoption에서 확인 가능	Program List에서 확인 가능
6.2				
4	• NARA-6.2 = 1점 ⇨ NAK-02 = Y	(△) Quality and functionality factors > Normal rendering에서 확인 가능	Program List에서 확인 가능	
개방성	5	(△) Transparency 3.1	Identification and description > Description에서 확인 가능	-
		3.1		
	6	• NARA-3.1 = 1점 ⇨ NAK-05 = Y	Sustainability Factors > Documentation, Adoption에서 확인 가능	-
		-		
	7	-	Note > General, History에서 확인 가능	-
공표	8	Disclosure 1.2	Sustainability Factors > Disclosure, Documentation에서 확인 가능	-
		1.2		
	9	• NARA-1.2 = 2점 ⇨ NAK-08 = Y	(△) Sustainability Factors > Disclosure, Documentation에서 확인 가능	-
		-		
	10	Disclosure 1.3	(△) Sustainability Factors > Disclosure, Documentation에서 확인 가능	-
1.3				
11	• NARA-1.3 = 2점 ⇨ NAK-10 = Y	Identification and description > Description에서 확인 가능	-	
		Impact of Patents 7.2	Sustainability Factors > Disclosure, Documentation, Adoption, Licensing and patent claims에서 확인 가능	-
		• NARA-7.2 = 2점 ⇨ NAK-11 = Y		

NAK-선정기준(공통)	NARA-Risk Matrix*		LOC-FDDs**	FileInfo.com***	
	범주	문항			문항
상호 운용성	독립성	12	External Software Dependencies • NARA-6.4 = 2점 ⇨ NAK-12 = Y	(△) Sustainability Factors > External dependencies에서 확인 가능	Program List에서 확인 가능
		13	External Hardware Dependencies • NARA-5.1 = 2점 ⇨ NAK-13 = Y	-	-
		14	-	-	-
		15	External Hardware Dependencies External Software Dependencies • NARA-5.1 = 2점 ⇨ NAK-15 = Y • NARA-6.3 = 2점 ⇨ NAK-15 = Y	-	-
		16	-	(△) Sustainability Factors > Adoption에서 확인 가능	Developer, Program List에서 확인 가능
채택	변환 가능성	17	-	-	-
		18	-	-	-
		19	-	-	-
		20	-	-	-
		21	-	-	How to Convert에서 확인 가능
		22	-	-	-
기능성	편제성	23	-	-	Program List에서 확인 가능
		24	-	(△) Sustainability Factors > Adoption에서 확인 가능	(△) How to Open에서 확인 가능
		25	-	(△) Sustainability Factors > Adoption에서 확인 가능	-
		26	-	(△) Sustainability Factors > Adoption에서 확인 가능	-
		27	-	(△) Sustainability Factors > Adoption에서 확인 가능	Program List에서 확인 가능
가용성	편증성	예외1	(△) Adoption • NARA-2.1 = 2점 ⇨ NAK-예외1 = Y	(△) Sustainability Factors > Adoption에서 확인 가능	-
		예외2	-	(△) Sustainability Factors > Adoption에서 확인 가능	-
		28	Self-Documentation • NARA-4.1 = 2점 ⇨ NAK-28 = Y • NARA-4.2 = 1점 ⇨ NAK-28 = Y	(△) Sustainability Factors > Self-documentation에서 확인 가능	-

NAK-신장기준(공통)		NARA-Risk Matrix*		문항	LOC-FDDs**	FileInfo.com***
범주	문항	범주	문항			
기능성	자체 문서화	-	-	-	(△) Sustainability Factors > Self-documentation에서 확인 가능	-
	기계 가독성	-	-	-	(△) Sustainability Factors > Transparency에서 확인 가능	-
	보호 메커니즘	32	Technical Protection Mechanisms •NARA-8.1 = 2점 ⇨ NAK-32 = Y •NARA-8.2 = 0점 ⇨ NAK-32 = Y •NARA-8.3 = 2점 ⇨ NAK-32 = Y	8.1	Sustainability Factors > Technical protection considerations에서 확인 가능	-
				8.2		
				8.3		
	33	-	-	-	-	-
	34	-	-	-	-	-
검색기능	35	-	-	-	-	-

* NARA(2019). https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Risk_Matrix

** Library of Congress (2004). <https://www.loc.gov/preservation/digital/formats/intro/intro.shtml>

*** FileInfo.com(2005). <https://fileinfo.com/>

