

공개포맷에 기반한 전자기록 보존 포맷 재설계 방향 연구

Redesigning Electronic Records Preservation Formats Based on Open Formats

오 세 라 (Seh-La Oh)*

정 미 리 (Mi Ri Jung)**

임 진 희 (Jin Hee Yim)***

목 차

- | | |
|-----------------------|-------------------|
| 1. 서 론 | 4. 장기보존포맷의 재설계 방향 |
| 2. 공공 전자기록 장기보존 전략 현황 | 5. 결 론 |
| 3. 문서보존포맷의 재설계 방향 | |

<초 록>

변화하는 컴퓨팅 환경에 따라 공문서 생산 환경이 G-클라우드 및 클라우드 온-나라 시스템으로 전환하고 있다. 더불어 공문서 포맷도 독자 포맷인 HWP을 대신하여 개방형 규격인 ODF가 정착할 것으로 전망된다. 이에 본 연구는 현행 공문서 장기보존 전략을 재검토 하고, ODF 생산 문서 보존에 적합한 문서보존포맷과 장기보존포맷 설계 방향을 제시하고자 한다. 생산 포맷이 ODF인 기록은 문서보존포맷으로의 변환 없이 기록관리 메타데이터를 추가한 확장 ODF 포맷을 제안하고자 하며 독자포맷의 보존에도 ODF를 검토하였다. 장기보존 포맷, 즉 NEO 포맷 변환 현황을 알아보고 BagIt 규격 패키지를 ISO21320으로 제정된 ZIP을 적용하여 장기보존포맷을 설계하였다.

주제어: 문서보존포맷, 장기보존포맷, 전자기록 장기보존, 마이그레이션, 인캡슐레이션, ISO21320

<ABSTRACT>

Along with the changing computing environment, the production of public documents is moving to G-Cloud and On-nara System, which are re-engineered on G-Cloud. In this process, it is predicted that the Open Document Format (ODF) will replace the proprietary document format, HWP, and will become widely used in the near future. To preserve digital records produced in ODF, this study reviewed the current long-term preservation strategies and identified considerations for new formats. For ODF source files, it is suggested to eliminate the conversion to PDF/A-1 and preserve the documents in the extended ODF with added metadata of records management. Moreover, ODF is presented as an alternative format for preserving proprietary formats. Finally, this study reviewed the current long-term preservation format, i.e., NEO, and suggests a zipped BagIt package, which has an ISO 21320 specification.

Keywords: long-term preservation format, migration, encapsulation, ISO 21320

* 네모아이씨티(주) 대표(sehlao@naver.com) (제1저자)

** (사)한국국가기록연구원 선임연구원(fancyball@rikar.org) (공동저자)

*** 명지대학교 기록정보과학전문대학원 부교수·인간과기억아카이브(hmarchives.org) 아키비스트
(yimjhkr@mju.ac.kr) (교신저자)

■ 접수일: 2016년 11월 1일 ■ 최초심사일: 2016년 11월 1일 ■ 게재확정일: 2016년 11월 24일

■ 한국기록관리학회지 16(4), 79-120, 2016. <http://dx.doi.org/10.14404/JKSARM.2016.16.4.079>

1. 서론

1.1 연구배경 및 목적

최근 서울기록원의 정보화전략(ISP)사업에 서는 문서보존포맷인 PDF/A-1과 장기보존포맷인 NEO(NAK(National Archives of Korea) Encapsulated Object)로 변환하기 위한 서버와 스토리지 및 소프트웨어 도입 예산의 부담, 관리 등의 문제들이 발생함에 따라 새로운 문서보존포맷의 채택이 필요하다고 보았고 이에 대한 중점 추진과제로 영구기록물 문서보존포맷의 재설계를 제안하였다(서울시, 2016). 장기보존에 유리한 ODF(Open Document Format)를 문서 생산 포맷으로 규정하여 별도의 보존포맷 변환 없이 보존하고, 장기보존포맷으로 영국의 회도서관 등 기관에서 도입한 Bagit 등의 아카이브 포맷을 검토하여 장기보존포맷에 대한 재설계 방향을 제안하였다(서울시, 2016). 이와 같은 연구 결과는 포맷 변환과 이관에 대한 비효율 문제가 내부적으로 발생하는 상황과 온-나라 클라우드에서 생산 포맷을 ODF로 도입하는 생산 환경의 변화가 맞물리면서 현행 보존포맷에 대한 실효성에 대한 의문이 직접적으로 제기되었음을 알 수 있다.

보존포맷 변환과 이관과 관련된 상황을 살펴 보자면, 『공공기록물 관리에 관한 법률』 제19조(기록물의 관리)에 따라 보존기간 기산일로부터 10년이 경과한 2003~4년도 생산 기록물 중 보존기간 30년 이상 기록물은 문서보존포맷과 장기보존포맷으로 변환되어 국가기록원으로 이관되고 있다. 전국 공공기관 관할 기록관에서는 국가기록원으로서의 이관을 위해 보존기간이 10년

이상인 기록을 문서보존포맷인 PDF/A-1로 변환하고 30년 이상의 기록을 장기보존포맷인 NEO로 변환하였다. 그러나 정식이관에 앞서 2011년과 2013년 두 차례 시범이관을 실시하며 준비했음에도 불구하고(국가기록원, 2016) 이관 시점에 임박하여 다량의 기록을 보존포맷으로 변환하면서 여러 문제가 발생하고 있다. 첫째, 보존포맷 변환 과정 중 시스템 오류로 인한 변환 실패가 발생하고 있다. 오류코드 중 문서보존포맷과 관련된 오류는 문서에 쓰기 권한이나 암호가 설정되어 있는 등 문서에 접근이 되지 않아 변환이 불가하거나 한/글(HWP 포맷을 실행하는 어플리케이션, 이하 한/글)에서 인쇄기능으로 구현되는 PDF/A-1 변환기 실행 중 발생하는 오류가 있으며, 장기보존포맷과 관련된 오류는 메타데이터가 불일치하거나 인증서와 관련된 오류가 있다(국가기록원 RMS유지보수사업단 제공 RMS오류코드, 담당자로부터 이메일로 직접 수신). 둘째, 문서보존포맷으로 변환하는데 많은 시간이 필요하다. 문서보존포맷으로 변환에 드는 시간은 국가기록원에서 제시하는 컴퓨터 사양을 기준으로 최저사양(1건당 9초)과 최고사양(1건당 1초)의 처리속도의 편차가 있으며(국가기록원, 2015a), 이를 기준으로 계산해 보면 작년 전체 이관 대상인 238만 건의 변환에만 소요되는 시간이 대략 661~5,950시간이 걸렸다고 볼 수 있다. 그러나 이는 변환에서만 그치는 것이 아닌 위에 언급한 시스템적 오류들 외에도 메타데이터, 불임파일 누락 등 다양한 오류들을 함께 처리해야 하는 것으로 보존포맷으로 변환하기 위해 소모되는 시간이 많다고 볼 수 있다. 마지막으로 보존포맷을 저장하기 위한 대용량 스토리지가 부족하다. 보존포맷 변환 과

정에서 용량이 원본에 비해 보통 1.3배 증가하고 중앙연구기록물관리기관으로의 이관 전에는 이러한 기록들이 누적되고 있어 기록관에서는 이를 저장하기 위한 스토리지의 확보가 필요하다. 하지만 노후화되거나 증설이 필요한 장비를 즉각적으로 업그레이드하기 위한 비용을 확보하기 어려운 실정이다.

생산 포맷의 변화와 관련된 상황으로는 2016년 6월 행정자치부에서 발표한 “정부지식 공유활용기반 고도화” 2차 사업의 제안요청서에서 살펴볼 수 있다. 그동안 기록을 생산하는 신전자 문서시스템에 한/글이 편집기로 탑재되어 있었고, MS-Office에서 생산되는 다양한 포맷의 문서들이 전자문서의 대부분을 차지하고 있었다. 그러나 ‘정부 3.0 발전계획’의 핵심과제 중 하나인 ‘클라우드 기반 지식 정부 구현’에서 온-나라 시스템 및 기록관리시스템의 고도화를 추진 중이며, 이 사업에는 개방·공유의 정부 3.0 가치를 구현하기 위해 공문서에 개방형 문서표준포맷인 ODT와 PDF 등을 적용할 것을 명시하고 있다. ODT는 XML을 기반으로 설계된 개방형 문서표준포맷인 ODF의 텍스트 문서 포맷으로, 포맷 구조와 개발 내용이 공개된 오픈 포맷이다. 텍스트인 XML을 기반으로 작성되고 압축파일인 ZIP의 형태로 생성되어 바이너리코드로 작성된 포맷보다 장기보존에 유리하다(성환혁, 2007). 이처럼 공공기록의 생산 환경이 클라우드 컴퓨팅 환경으로 변화하면서 ODF의 채택이 가정사실화되고 있음을 알 수 있다.

이와 같은 공문서 생산 환경과 관련된 내부의 직접적인 변화와 서울기록원의 연구에서 제기된 바를 토대로 ODF도입을 고려하면서 포맷 변환과 관련된 이슈를 해결하기 위한 방안을 생

각해볼 필요가 있다. 문서보존포맷, 장기보존포맷 등의 보존포맷들이 지정된 후 10년이 지난 현재, 현장에서 보존포맷의 적용과 활용 과정에서 어떠한 문제점이 있었는지 살펴보고, 기록관에서 포맷으로 변환하는 기술 및 비용 부담과 국가기록원에서 이관 받고 보존하는 과정에서의 비용 및 기술 효율 등을 세밀하게 검토할 필요가 있다.

본 연구는 현재까지 확인된 문서보존포맷, 장기보존포맷과 관련된 기록관과 국가기록원이 경험한 난제들의 원인을 검토해보고, 변화하는 환경에서 전자기록의 안전한 장기보존이라는 본래의 목적에 부합하는 보존포맷을 새롭게 설계할 때의 방향을 제시하는 것이 목적이다. 또한 공문서의 생산 환경이 변화하고 기존 보존포맷에 대한 재검토가 요구됨에 따라 전자기록의 효과적이고 효율적인 보존을 위한 문서보존포맷과 장기보존포맷의 설계 방향을 제안하고자 한다.

1.2 선행연구

전자적으로 생산한 기록을 안전하게 보존하고 장기적인 이용가능성을 보장하기 위한 대표적인 연구는 네덜란드의 디지털 보존 테스트베드(Digital Preservation Testbed), InterPARES 1 Project의 진본성 연구, 2 Project의 장기보존포맷의 검토가 있었으며, 특히 3 Project에서는 한국 TEAM이 장기보존에 적절한 파일 포맷의 기본적인 요건인 자율성(Autonomy), 상호운용성(Interoperability), 진본성(Authenticity), 기능성(Functionality) 등을 기준으로 오픈 표준 파일 포맷 TIFF와 PDF, PDF/A, XML에 대

해 검토하였다(InterPARES 3 Project, 2016). 국내의 전자기록 보존에 관한 연구는 일반, 정책, 기술 및 디지털 자원 유형별 연구의 4가지 주제영역으로 범주화 되고 있으며, 그 중 기술관련 분야는 프로세스, 이관, 시스템, 패키지, 보존 메타데이터, 파일 포맷 등의 세부주제로 범주화 될 수 있다. 그 중 파일 포맷과 관련된 연구는 주로 파일포맷 레지스트리와 같은 포맷 관리에 관한 연구가 진행되고 있었다(이소연, 2013).

보존포맷과 관련된 연구 중 국가기록원의 보존포맷에 대한 연구개발이 진행된 2005~6년 이후 문서보존포맷에 대한 기술적 고찰을 나타낸 연구는 성환혁(2007)의 연구가 있다. 그는 국가 R&D에서 제시하고 있는 문서보존포맷 PDF/A-1에 대한 한계를 다양한 기록유형에 적용할 수 없다는 점을 언급하고 다차원 문서와 다매체 문서에 대해 ODF를 우선 검토할 것을 제안하였다. 문서보존포맷이 다양한 기록을 보존할 수 없다는 점은 임진희, 최주호, 이재영(2014)의 연구에서도 지적된 바 있다. 현재 적용되고 있는 문서보존포맷인 PDF/A-1은 인쇄 출력용의 결재문서에는 타당하지만 다양한 OLE(Object Linking and Embedding) 객체가 삽입된 한/글 문서나 여러 수식을 담고 있는 엑셀 문서를 변환할 경우 기능들이 소실되므로 여러 맥락이 담긴 문서의 가치가 상실되므로 PDF/A-1로의 변환 외에 다른 접근이 필요함을 말하면서 PDF/A-1의 마이그레이션 전략과 상호보완이 될 수 있는 한국형 에뮬레이션 전략을 제시하였다.

최근에는 보존포맷의 정책과 기술적 구조에 대한 연구가 진행되었다. 이승익(2015)은 국가

기록원 계간지 『기록인』의 칼럼을 통해 보존포맷의 정책 재검토를 이야기하였다. 현행 기록 보존 지침은 생산 원천 파일과 원천 파일을 변환한 PDF 파일을 관리하므로, 원천 파일이 따로 존재하기 때문에 PDF를 보존을 위한 포맷보다 생산 포맷으로 규정하는 것이 더 효율적이라고 말하였다. 또한 “PDF를 양질의 파일포맷으로 인정하는 국가들은 대부분 생산포맷으로 보는 경우가 일반적이며, 만약 생산포맷으로 규정할 경우 그 자체로 생산기록이 되어 원천파일과 관계가 달라지므로 병행 이관할 필요가 없어진다”고 보았다. 또한 장기보존포맷(NEO)에 대해서는 구성적 속성, 생성 시점 및 주체를 논의하였다. 또한 ‘국가기록원은 NEO의 보존관리 프로세스나 인프라가 기존 비전자 기록에 비해 매우 미비한 실정임’을 말하면서, NEO의 고안은 규격화된 결재문서에서 시작되었으므로 새로운 유형을 포괄할 수 있는 구조적 유연화가 필요하다고 보았다. 이로써 현행 보존포맷 규정은 생산과 보존 분야로 구분하여 별도로 규정할 필요가 있으며, 양질의 생산 포맷을 국가기록원이 표준화하여 생산기관이 이를 준수하도록 하는 방향이 타당하다고 보았다. 따라서 현재 RMS에서 실행되고 있는 장기보존포맷 변환을 국가기록원과 같은 아카이브 관리 영역으로 재조정하는 것이 오류를 줄일 수 있는 실무적인 차원의 개선방안이고, ISO14721과 같은 디지털기록관리 규범에 부합하는 개선을 의미한다고 주장하였다. 이에 따라 “국가기록원은 생산에서 보존까지의 전 주기를 관리 시야에 넣을 수 있는 통합적 관리체계의 구축이 필요하다”고 말하였다.

곽정(2016)은 구전자-신전자-업무관리시스

템으로 이어지는 전자문서 생산 시스템의 변화와 관련하여 전자기록의 진본성 확보를 위한 전반적인 기록관리 개선방안에 대해 연구하면서, 장기보존포맷의 구조 개선에 대해 언급하였다. 현행 NEO는 “전자기록의 유형에 관계없이 ‘기록물건’이라는 표준화된 메타데이터 구조로 변경하고 생산-관리-보존 과정에서 관리 행위가 발생할 때 해당 요소에 이력정보를 누적하여 축적하는 구조”라 설명하면서, 다양한 유형의 전자기록이 기록관리 단계에서 기록물건으로 일원화되면서 여러 문제점이 발생하였고, 특히 생산 당시 전자기록에 포함되었던 요소들이 유실되는 문제가 발생하였다고 한다. 표준화된 요소에 맵핑되면서 업무 맥락에 따라 다양한 의미를 함축하고 있던 정보들의 본질이 왜곡되는 문제와 원천정보와 기록관리 과정에서 추가된 이력정보가 혼재되어 관리되면서 원천정보가 희석되는 문제가 발생하고 있어 장기보존포맷의 본질적인 개선을 주장하였다. 이처럼 10여 년 간 현장에서 보존포맷을 실제 운영하면서 나타나는 문제들에 대해 인식하고 제도적으로나 기능 및 구조상 변경의 필요성을 주장하는 연구들이 나타나고 있음을 알 수 있다.

최근 공공영역에서 특정 벤더의 어플리케이션에 종속되지 않는 공개포맷을 도입하여 적용하기 시작함에 따라 기록의 생산, 유통, 관리, 활용 및 보존의 측면에서 예상되는 영향에 대한 연구도 시작되었다. 정미리, 오세라, 임진희(2016)는 생산 및 유통과정에서 발생하는 비용 부담의 절감과 RDF를 통한 이용가능성의 확장을 말하였으며 보존 측면으로는 XML의 구조 속에 기록관리를 위한 메타데이터 요소를 추가함으로써 장기보존을 위한 포맷 변환 단계를 생략 또

는 간소화 할 수 있음을 언급하였다. 그러나 세부적인 연구단계는 현재까지 이뤄지지 않은 상태이며, 올해부터 ODF 도입 사업이 시작됨에 따라 이에 대한 세밀한 연구가 시급한 것으로 여겨진다.

1.3 연구 방법 및 범위

전자기록의 장기보존전략의 종류와 특징 그리고 우리나라에서 문서보존포맷과 장기보존포맷을 채택하게 된 경위 등을 살펴보기 위해 국가기록원에서 실시한 2차례의 R&D 보고서 및 보존포맷과 관련된 공공표준을 바탕으로 문헌 연구를 실시하였다. 또한 기록관에서 수행된 보존포맷 변환 과정에서 발생하는 문제에 대해서는 2015년 국가기록원 백서의 이관 사업 결과와 관련 연구에 담긴 실무자와의 인터뷰 내용, 해당 백서의 사례에 제시된 기관의 담당자와의 인터뷰를 통해 문헌만으로는 느낄 수 없는 현장 실무자의 고충을 엿볼 수 있었다. 문헌 및 현장 조사를 통해 생산 포맷의 변화가 시스템 및 현업의 문제 개선의 출발점이 될 수 있음을 깨달았다. 이에 따라 실제 공공기관에서 생산되고 있는 공문서 파일(*.hwp)과 새로 변경될 ODF(*.odt)의 구조를 비교해보고, ODF 표준인 ISO/IEC 26300를 바탕으로 ODF 포맷 내에 기록관리 요소 메타데이터 적용 방안을 고려해보았다. 이로써 문서보존포맷 변환 단계를 생략하여도 기록의 신뢰성, 무결성, 진본성, 이용가능성을 확보하고 비용적, 시간적 효율성을 보장할 것으로 여겨진다. 또한 원본 파일(ODF)과 메타데이터 및 전자서명 등을 패키징하기 위해 해외 선진사례인 Bagit을 살펴보고, ISO21320

으로 제정된 ZIP을 이용한 장기보존포맷을 제안하였다. 본 연구의 기록유형은 전자문서류에 한정하였으나, 행정정보데이터세트와 웹 기록, SNS 기록 등 다양한 포맷의 기록도 포괄할 수 있는 보존포맷을 고려하였음을 밝혀둔다.

2. 공공 전자기록 장기보존 전략 현황

2.1 공공 전자기록의 장기보존 전략

전자적으로 생산된 기록은 종이기록과는 달리 시스템에 의존해야 하는 특성 상 기술력이 노후되거나 포맷이 변화되는 경우 접근 및 이용가능성을 보장하기 어렵다. 전자기록의 장기보존을 위한 대표적인 세 가지 전략으로 마이그레이션(Migration), 에뮬레이션(Emulation), 인캡슐레이션(Encapsulation)을 꼽을 수 있다. 국가기록원에서는 2005년 「전자기록물 영구보존 기반기술 용역」과 2006년 「기록관리시스템혁신 ISP」 두 차례의 R&D를 통해 공문서의 보존 포맷 표준을 확정·고시하였다. 두 연구에서는 전자적으로 생산된 기록과 디지털화 한 기록을 전자적으로 이관·보존·활용하면서 각 단계별 과정에서 불법적인 위·변조를 방지하고 전자기록 관리의 핵심원리인 원본성·진본성의 유지, H/W와 S/W에 의존하지 않는 연속성 확보, 디지털 데이터의 영구보존에 필요한 전산 기술적·법적 요소들을 분석하여 문서 보존포맷과 장기보존포맷에 대한 기술규격을 제안하였으며, 2008년 국가기록원 표준으로 제정하여 현재에 이르고 있다(국가기록원, 2008). 표준을 각각 살펴보면 각 시스템에서 한/글로

작성된 결재문서를 문서보존포맷인 PDF/A-1로 변환하는 일종의 마이그레이션 전략과 기록의 메타데이터를 원문 비트스트림들과 함께 XML으로 패키징하는 NEO규격의 인캡슐레이션 전략을 혼합하여 전자기록의 보존포맷을 결정하였음을 알 수 있다(임진희, 2013).

마이그레이션은 한 세대의 컴퓨터 기술로부터 다음 세대로, 또는 한 가지 조합의 하드웨어 및 소프트웨어로부터 다른 것으로 정기적으로 디지털 자료를 옮기는 방법이다(임진희, 2013). 즉 기록이 원래 생산된 하드웨어와 소프트웨어 환경에서 새로운 환경으로 변환하는 것으로, 옛날 포맷으로 작성된 기록을 최신 컴퓨터에서 실행될 신규 포맷으로 변환하는 전자기록 장기보존 전략이다. 이를 InterPARES(International Research on Permanent Authentic Records in Electronic Systems)의 용어 데이터베이스에서는 '시간이 흐름에 따라 시스템과 기록 매체가 유실되거나 노후 될 수 있는 상황에서 기록물의 지속적인 접근가능성을 보장하기 위해 기록물을 하나의 시스템 또는 저장 매체로부터 다른 것으로 이전하는 프로세스'라고 정의하고 있다(권도윤 외, 2009). 이는 전자기록의 기능(functionality)을 보존하고, 이용자들이 기록에 접근할 수 있도록 함으로써 물리적 매체의 노후화 문제를 해결할 수 있게 한다. 그러나 반복적인 마이그레이션 작업은 원본의 논리적 구조 또는 처리(rendering) 방법을 다르게 하여 새로 작성하므로 비트스트림의 손실이 불가피하다고 볼 수 있다. 또한 마이그레이션은 디지털 환경이 변경되는 시점마다 지속적으로 실행해야 하며, 이로 인한 비용의 부담은 단점으로 볼 수 있다(국가기록원, 2004, 2008).

인캡슐레이션은 하나의 전자 기록 객체에 데이터와 기능 등 필요한 모든 자원을 포함하여 물리적인 컨테이너(container)나 랩퍼(wrapper)로 고정하는 것으로, 전자기록의 내용과 구조에 관한 메타데이터, 데이터 객체, 진본임을 확인할 수 있는 전자서명 등을 하나의 객체로 결합하는 것을 말한다(임진희, 2013). 메타데이터는 기록이 지적으로 이해될 수 있게 하고 미래에 기술적으로 접근될 수 있도록 하는 역할을 한다. 이러한 기록의 맥락정보를 패키지와함으로써 기록의 무결성과 진본성의 보증이 가능할 수 있지만, 인캡슐레이션이 진행되는 동안 중요 메타데이터가 누락될 위험이 있으므로 마이그레이션이나 에물레이션 등과 같은 기록의 접근가능성을 보장할 수 있는 전략들과 함께 결합되어 사용되어야만 효과를 발휘할 수 있다(국가기록원, 2004).

문서보존포맷은 독자포맷으로 생산된 기안문과 첨부파일에 대한 프로그램별 상호호환성, 데이터 레거시, 라이선스 등의 문제를 탈피하면서 문서에 대한 접근과 이용가능성을 장기적으로 보장하기 위해 문서가 생산된 당시의 애플리케이션(예를 들어, 한/글 또는 MS-Word)이 없어도 해당 문서의 내용과 외형을 그대로 재현하고 시스템에 관계없이 시각적 모양을 지속적으로 보장하기 위해 도입되었다. 이를 위해 호주 NAA의 XML데이터 포맷, 호주 VERS의 VEO, 영국 EROS에서 권고하고 있는 문서보존포맷 등과 같은 선진 사례를 조사하고 문서보존용으로 활용할 수 있는 XML, Text, 이미지, PDF, CSD와 같은 포맷에 대해 영국 TNA에서 가이드 하고 있는 전자문서 장기보존포맷의 요건을 참고하여 검토하였다. 8가지 검토 기준은

다음과 같다. 첫째, 공개용 표준(Open Standard)으로, 특정 벤더에 속하는 제품이거나 사용자에게 경제적으로 영향을 주지 않아야 한다. 둘째, 편재성(Ubiquity)으로, 오랜 기간이 지나도 언제 어디서든 사용 가능할 수 있어야 한다. 셋째, 안정성(Stability)으로, 포맷 변환시 원본 문서의 내용이나 구조, 맥락정보 등을 훼손시키지 않아야 한다. 넷째, 메타데이터 지원으로, 보존하기 위해 필요한 메타데이터를 지원할 수 있어야 한다. 다섯째, 상호운영성(Interoperability)으로, 한 기관에서 생성된 포맷은 다른 기관이나 다른 사용자도 사용할 수 있어야 한다. 여섯째, 진본성(Authenticity)으로, 문서가 훼손되거나 위조 또는 변조되지 않도록 기록물의 진본성을 유지할 수 있어야 한다. 일곱째, 표현력(Presentation)으로, 원래 생산된 문서의 모양과 구성이 같도록(Look and Feel) 복원할 수 있어야 한다. 여덟째, 검색 기능으로, 사용자가 문서 내부의 정보를 검색하고 적합한 문서를 찾을 수 있어야 한다. 이러한 기준을 바탕으로 앞서 언급한 문서보존포맷들에 대하여 종합적으로 고려하여 적합성을 분석하였으며, 검토 결과 PDF/A-1(ISO 19005-1:2005)이 문서보존포맷으로 채택되었다(NAK/TS 3:2008). 2004년 국가기록원의 R&D 당시 StarOffice의 ODF에 대한 구조를 검토하였고 2006년의 연구에서는 XML을 문서보존포맷의 비교대상으로 언급한 바 있으나, PDF/A-1과는 표현력에서 부족함으로 평가되었다. ODF는 2004년 당시 공개용 표준 검토 단계였고 2006년에서야 ISO/IEC 인증이 이뤄지게 되었다. 2008년 7월 기록관리 표준전문위원회 초안 검토 보고 내용에 의하면, PDF/A-1 규격으로 보존할 수 없는 전자문서

에 대한 보존방법이 필요하고, ODF와 OOXML 등에 대한 연구가 필요하다는 전문가의 의견이 있었으나(행정자치부, 2008), 이에 대한 후속 조치가 없는 것으로 미루어 보아 2008년 당시에는 PDF/A-1보다 ODF에 대한 사례와 검증이 부족한 단계에서 기인한 것으로 추측해볼 수 있다. 현재의 관점에서 ODF를 장기보존포맷 요건에 맞춰 평가해본다면 표현력을 포함한 모든 요소를 충족한다고 볼 수 있을 것이다.

문서보존포맷이 한/글 등에서 생산된 전자 결재문서의 내용을 생산 당시의 모습(인쇄형태) 그대로 지속적으로 확인하는 것을 보장하

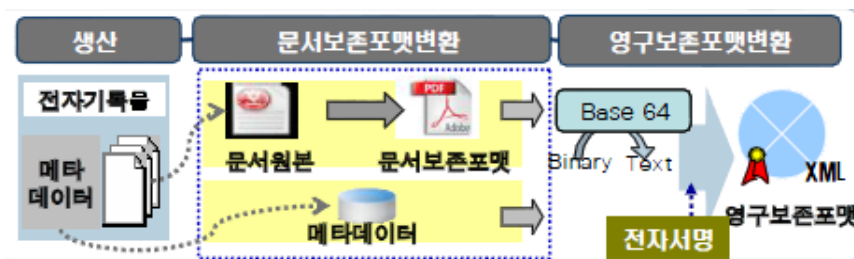
기 위함이라면, 장기보존포맷은 업무활동에 대한 증거와 업무에 대한 책임소재를 분명히 할 수 있는 법적증거를 장기간 안전하게 보존하기 위한 포맷으로 볼 수 있다. 이를 위해서는 처음 생산된 파일을 비트스트림의 변형 없이 그대로 보존하면서 내용에 대한 접근을 보장해야 하며, 관리에 대한 이력과 진본임을 보장할 수 있는 검증파일을 하나의 객체로 묶어 진본성과 무결성을 유지해야 한다(NAK/TS 4).

이에 따라 시스템에 의존하지 않는 자체충족성과 맥락정보를 포함하는 자체 문서화, 육안으로 정보를 확인할 수 있는 구조화된 텍스트

〈표 1〉 문서보존포맷 비교

	XML	Text	이미지 (TIFF/BitMap)	PDF	CSD	PDF/A-1
공개용 표준	○	○	○	X	X	○
편재	○	○	○	△	△	○
안정성	○	X	○	○	○	○
메타데이터 지원	○	X	X	○	○	○
상호운영성	○	○	○	○	○	○
진본성	○	○	○	○	○	○
처리능력	○	X	△	○	○	○
표현력	△	X	○	○	○	○
검색기능	○	○	X	○	○	○

출처: 국가기록원(2006b).



〈그림 1〉 장기보존포맷(영구보존포맷) 변환 프로세스

출처: 국가기록원(2006b), p. 198.

인코딩, 기록이 변조·위조되지 않음을 보장하는 무결성 등의 요소를 고려하여 호주의 빅토리아 주에서 구현한 VERS@DOI 프로젝트에서 양과모델을 적용한 인캡슐레이션 전략인 VEO와 다양한 포맷의 전자기록물을 XML 기반의 임의 구조화된 패키지로 보존하는 미국의 사례를 도입하여 장기보존포맷 규격인 NEO를 마련하였다. NEO는 전자기록철의 계층적 구조를 반영할 수 있는 텍스트 기반의 표기 언어인 XML로 표현되며, 기록물의 메타데이터와 콘텐츠, 인증정보를 텍스트로 구조화한 것이다. NEO는 진본성을 보장하기 위해 생산자가 처음 생산한 기록의 원본 파일과 시간과 기술의 변화에 상관없이 내용을 열람할 수 있도록 변환한 문서보존파일(PDF/A-1), 해당 기록에 대한 관리 이력을 담은 메타데이터 및 진본성 및 무결성을 보장하기 위한 전자서명을 모두 Base64로 인코딩하여 XML로 랩핑하여 보존하는 인캡슐레이션 전략이다(NAK/TS 4). Base64 인코딩은 바이너리데이터를 아스키(ASCII)텍스트로 변환하거나 그의 반대로 변환하는 인코딩 방법으로, 2진 데이터의 각 3바이트씩을 4개의 6비트 단위로 나누어 하나의 64진수 문자로 변환하게 되어 인코딩을 거치게 되면 파일 크기가 원래보다 약 1/3정도 증가한다(서울시, 2016). 이처럼 바이너리 형식의 원본파일을 Base64로 인코딩을 하면 원래의 크기보다 1.88~2.25배 용량이 증가하게 되며, 문서보존포맷으로 변환한 파일과 메타데이터, 인증서를 모두 포함하기 때문에 최초 생산한 문서 파일보다 2~3배 정도 파일 용량이 증가한다(국가기록원, 2006a).

문서보존포맷과 장기보존포맷으로 변환하기

위해서는 포맷변환 모듈이 기록관리시스템 상에 탑재되어 기능을 구현해야 한다. 국가기록원에서는 문서보존포맷 변환 모듈에 대해 파일구조, 그래픽, 폰트, 투명도, 주석, 메타데이터, 액션, 폼 등에 대한 규격을 마련하였다. 현재 한국정보통신기술협회(TTA)에 문서보존포맷 변환 모듈 규격을 인증 받은 곳은 (주)유니닥스의 'ezPDFWorkflow'라는 1개의 제품이며(정보통신시험인증연구소 홈페이지), 이 제품을 탑재한 표준RMS는 중앙부처를 비롯한 군기관, 광역시도, 시군구 및 교육청 등에 보급되었다(보급률 99.5%, 2015년 말 기준). 표준RMS에는 보존포맷 변환을 위한 문서보존포맷변환 기능, 장기보존포맷변환 기능이 구현되어 있으며, 중앙연구기록물관리시스템(CAMS)에는 포맷변환 기능을 비롯하여 이관된 파일을 확인할 수 있는 NEO 전용 뷰어가 탑재되어 있다. 이처럼 장기보존전략을 구현하기 위한 체제와 시스템, 규격 등이 마련되어 있지만, 실제 장기보존전략 구현 과정에 대해 검토해볼 필요가 있다.

2.2 장기보존전략 구현 시 검토사항

서울기록원 ISP사업에서는 PDF/A-1로의 변환 과정에서 용량문제, 예산문제, 관리상의 어려움과 같은 문제점이 발생하고 있으며, 업무담당자가 10년 이상으로 책정된 기록물을 PDF/A로 변환하는 과정에서 많은 예산과 인력이 소요되고 있다고 지적하였다(서울시, 2016). 이에 따라 보존포맷과 관련된 법적 요건, 변환 프로세스, 시스템 환경, 비용의 측면에서 장기보존전략 구현에 대해 검토해보고자 한다.

첫째, 법적 요건에 대해 살펴볼 필요가 있다.

우리나라는 전자기록에 대한 장기보존에 대한 조치를 기록관 단계에서 실시하도록 법령에서 규정하고 있다. 공공기관의 기록관 또는 특수 기록관은 처리과에서 이관된 전자기록물 중 보존기간이 10년 이상인 전자기록물에 대해 품질 검사와 검수 등 인수처리 결과를 통보한 후 1년 이내에 문서보존포맷 및 장기보존포맷으로의 변환을 수행해야 한다(공공기록물법 시행령 제 36조). 문서보존포맷 변환 대상은 전자기록물 건에 포함된 본문파일과 첨부파일로, 문서형식이 아닌 시청각 유형(동영상, 음성파일 등)이나 압축파일, PDF 등은 변환 대상에서 제외된다(국가기록원, 2016). 법령에 따르면 영구기록물관리기관은 전자기록물의 진본성·무결성·신뢰성 및 이용가능성이 보장되도록 관리정보 메타데이터와 행정전자서명 및 시점확인 정보 등에 대한 검증을 실시하고, 주기적으로 장기보존포맷으로 변환해야 한다(시행령 제46조 2항). 이에 따라 기록물 인수 시 행정전자서명의 확인 등 진본 확인 절차를 수행하고, 메타데이터 오류, 바이러스 검사 등 품질검사를 실시하여야 한다(NAK/S 9:2015). 이를 위해 국가기록원은 법령에 근거하여 영구기록물을 전자적으로 관리하기 위한 중앙기록관리시스템(CAMS)을 구축하였다. 2007년 시스템 구축 이래 매년 고도화를 실시하였으며 2011년에는 전자기록물의 메타데이터, 바이러스, 전자서명을 검사하기 위한 인수기능을 탑재하였고 2014년에는 인수 기능 중 '육안검수' 프로세스를, 2015년에는 DFR (Digital Format Registry) 기능을 활용한 기계 검수 기능을 적용하였다(국가기록원, 2016). 국가기록원에서는 이관 받은 기록에 대한 품질검사를 마치고 이관 확정된 기록에 대해 이관 이

력을 메타데이터에 포함하여 NEO로 재포맷해야 한다. 보존포맷 변환에 대한 법령 개정은 2016년 4월에 이뤄졌으며, 이전까지는 보존포맷 변환 시점에 대한 언급이 없었으나, 본 개정 에 따라 처리과에서 이관된 기록은 1년 이내에 변환을 수행해야 한다.

둘째, 변환 프로세스 측면에서 검토해볼 필요가 있다. 기록관 표준운영절차에 따르면 기록관은 기록관리시스템을 통해 문서보존포맷 변환대상 기록물을 검색·조회하여 목록을 작성하고 그 대상 기록물을 문서보존포맷으로 변환해야 한다. 기록관은 문서보존포맷으로 변환된 기록물의 이상 유무를 확인한 후 변환이 완료된 기록물의 목록을 작성하고 기록관리시스템에 등록한다. 또한 기록관은 다음의 경우에 기록의 장기보존포맷 변환 작업을 수행하여야 한다. 하나, 보존기간 10년 이상 전자기록물의 문서보존포맷 변환 완료시, 둘, 장기보존포맷 변환 후 메타데이터 수정이 발생한 경우, 셋, 영구기록관리기관으로 이관 시 재 포맷 변환 요구가 발생한 경우 장기보존포맷인 NEO로 변환해야 한다. 기록관은 장기보존포맷 변환이 완료된 기록물을 기록관리시스템에 등록하고 장기보존포맷 변환 관련 메타데이터 정보를 갱신한다(NAK/S 10:2012). 장기보존포맷 변환 프로세스에는 Base64로 인코딩하는 시간이 소요되며, 검수나 내용 확인을 위해서는 원본 파일에 대한 디코딩이 필요하다. 국가기록원의 전자기록물 235만 건의 검수사업에는 보존기간 30년 이상의 장기보존포맷도 포함되어 있으며, 이에 대한 육안 검수를 위해 NEO파일을 일일이 확인하는 과정이 포함되어 있었다. 원천문서에 대한 파일 손상 확인, 암호 입력 확인, 내용 손상 확인

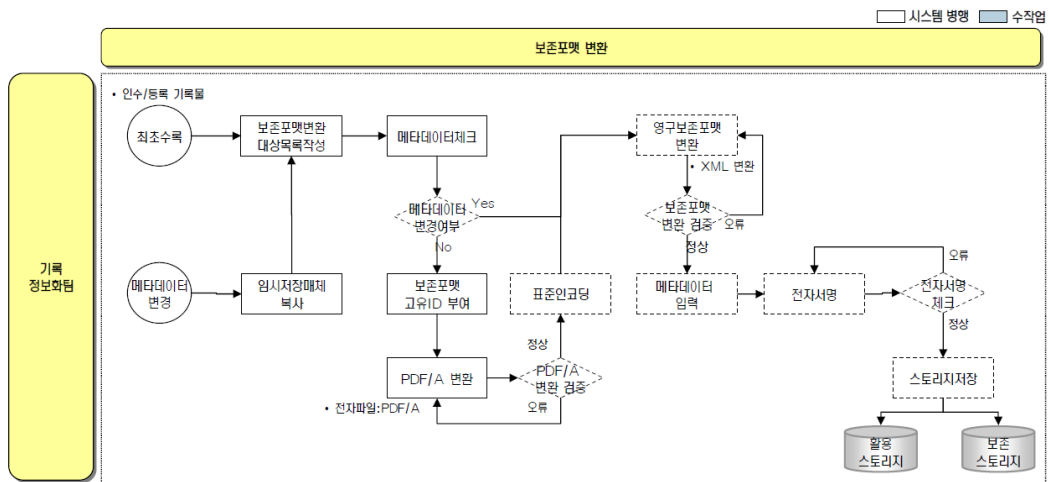
등의 육안검수를 시행하였는데 인코딩 된 데이터를 NEO뷰어를 통해 디코딩하여 원본 파일에 대한 확인을 위한 시간이 상당부분 포함되었을 것으로 예상된다. 또한 법령에 따라 처리과에서 이관 받은 기록물을 1차로 장기보존포맷으로 변환하고 나서 이관 등 기록관리 요소 메타데이터의 변경이 생기거나 전자서명 인증서의 갱신 등 재포맷 사유가 생길 때마다 2차, 3차 재포맷을 실시하여야 하므로 이를 실행하는 과정에는 인적·물적 자원이 많이 소요됨을 예상할 수 있다(〈그림 2〉 참조).

셋째, 보존포맷 변환 시스템 측면에서 검토해볼 필요가 있다. 표준RMS에서의 문서보존포맷 변환 프로세스는 포맷관리>문서보존포맷변환 기능에서 “변환요청 → 변환요청 작업공유(공유 데이터베이스) → 변환작업과 대상파일 획득 → 파일 변환 → 변환파일 저장 → 변환결과 전송”의 절차를 가지며, 장기보존포맷 변환 프로세스는 포맷관리>장기보존포맷변환 기능에서 “변환

요청 → 변환요청 작업공유·메타데이터 저장(공유 데이터베이스) → 변환작업과 대상파일 획득 → 파일변환 → 변환파일 저장 → 변환결과 전송”의 절차를 갖는다(국가기록원, 2015a).

문서보존포맷의 변환 시간은 시스템 환경에 따라 속도의 차이가 있으며, 특히 통합형 또는 공동형 표준RMS를 사용하는 경우 서버를 공유하므로 변환시간이 장시간 소요될 가능성이 높다(〈표 2〉 참조).

넷째, 비용 측면에서 검토해볼 필요가 있다. 우선, 하드웨어 요구사항으로서 서버를 도입해야 한다. 기록관리시스템 구축사업의 입찰공고에 나타난 변환서버의 사양은 CPU(6-Core 이상, 2.4GHz 이상), 메모리(32G 이상), 하드디스크 300GB 2개 이상이고, 소프트웨어 요구사항으로서 변환모듈의 구매가 포함된다(나라장터 홈페이지). 현재 표준RMS에 탑재된 보존포맷변환 모듈은 앞 절에서 언급한 바와 같이 1개의 모듈만 인증되어 있으며, 표준RMS를 도



〈그림 2〉 장기보존포맷 변환 프로세스

출처: 국가기록원(2006b)

〈표 2〉 시스템 환경별 문서보존포맷 변환소요시간

변환 시스템 테스트 환경			파일 건당 평균변환시간 *전체 파일수: 1,382개 (기록물건수: 870건)
시스템사양	대수	서버당 변환 프로세스 수	
저사양 CPU(Intel Xeon 4core 2Ghz) MEM(2G) OS(Win2003 enterprise)	1대	1	9.06초
		4	2.10초
고사양 CPU(Intel Xeon 32core 2.4Ghz) MEM(32G) OS(Window2012 enterprise)	1대	1	4.37초
		8	0.92초

출처: 국가기록원(2016)

입한 기관에서는 모두 해당 업체의 모듈을 사용하고 있다. 표준RMS 구축 시 필요한 모듈 1식의 비용은 약 560만 원이다(경기도교육정보기록원, 2016). 표준 RMS를 구축하였을지라도 이관 건수가 많을수록 처리 시간이 오래 걸리므로 분산 처리를 하기 위해 시스템의 증설이 필요하다. 따라서 이미 표준RMS 도입 시 지불하였던 모듈 비용에 추가로 증설해야 하는 서버 및 스토리지, 모듈 증설을 위해 예산을 편성해야 하는 부담이 생기는 것이다.

실제 국가기록원의 2015년 백서에는 문서보존포맷 변환 장애 등 각종 기술 오류, 폐쇄망 기관의 장기보존포맷 변환 환경 미비, 표준기록관리시스템 서버 및 스토리지 노후화 등에 대한 문제점을 인식하고 있었으며, 폐쇄망 기관인 국세청의 저장 공간 부족 사례와 대량이관 기관인 고용노동부의 문서보존포맷 변환을 위한 서버 메모리 부족 등의 사례를 제시하였다(국가기록원, 2016). 이처럼 현재의 여건으로 봐선 보존포맷 변환에는 물리적인 시간과 장비 요건이 절대적으로 요구된다. 국가기록원 2015년 백서에 제시된 국세청의 사례를 보면, 이관 대상 기록물은 75만 건으로 고성능 서버를 운영하

였을지라도 순수하게 포맷 변환에 필요한 시간은 약 208시간으로 계산된다. 그러나 국세청 기록연구사와의 인터뷰를 실시한 결과 실제 변환 기간은 대략 6개월 정도 소요되었다고 한다. 기존 시스템 장비를 업그레이드하면서 서버의 성능은 갖추어졌지만, 워낙 양이 많았고 순차적으로 오류를 해결하면서 변환을 진행해야 했기에 많은 시간이 소요되었다고 한다. 또한 스토리지 부족에 대한 문제는 국가기록원에서 임시로 이관용 외장하드를 대여해주어 해결했다고 한다(2016.10.25. 전화인터뷰 실시). 또한 서울시 강남구는 1,602,813건을 2014년 8월부터 시작하여 2016년 4월 완료하였다고 보고하였다(강남구, 2016). 기록관담당자는 RMS의 기능을 직접 실행해야 하는데 타 업무와 병행하여 실행해야 하고, 변환 완료시까지 소요 시간을 고려하여도 업무 부담의 가중을 예측할 수 있다.

기록관의 보존포맷 변환 과정에서 나타나는 실무자들의 고충은 현문수(2013)의 연구를 통해서도 간략하게나마 다음과 같이 전해들을 수 있다. 첫째, 문서보존포맷 변환모듈이 기본적으로 표준RMS에 탑재해 있더라도 수차례의 시스템 수정을 통해야 기능이 정상 작동한다. 둘째,

오류 발생이 잦아 수시로 변환 과정을 모니터링해야 하고, 셋째, 변환이 완료되었어도 건 단위 정보를 수작업으로 확인하여 변환되지 않은 첨부파일과 같은 컴포넌트를 확인해야 한다. 넷째, 보존포맷변환 기능을 사용하기 위해 별도의 아카이브 스토리지를 구입해야 하지만 영구기록물관리기관으로의 이관이 시급하지 않아 제한된 예산을 이곳에 할당할 필요성을 느끼지 못하는 상황이다(현문수, 2013). 이처럼 시스템의 보존포맷 변환 기능은 이관을 위해 반드시 필요한 과정이지만 빈번하게 나타나는 오류들로 변환에 드는 시간과 인력에 부담이 된다는 점, 기술적 오류가 발생할 때마다 직접 해결해야 하거나 IT 관련 부서 또는 국가기록원에 요청을 해야 한다는 점에서 이러한 문제들이 지속적으로 발생하거나 이관 대상 기록이 많은 기관인 경우 포맷을 변환하는 업무 자체가 기록연구사에게 스트레스와 업무 과중으로 이어질 가능성이 높다.

현행 문서보존포맷과 장기보존포맷은 독자 포맷으로 생산된 기록을 고려하여 설계되어있다. 살펴본 바와 같이 장기보존전략 구현 과정에서 여러 가지 난제가 발생하고 있고 생산 포맷이 오픈 포맷으로 변경될 것으로 예상되므로 보존 포맷 또한 개선되어야 할 필요가 있다. 본 절에서 시사한 바를 바탕으로 보존포맷의 재설계 목표 수립 시 다음과 같은 사항을 고려해볼 수 있다. 첫째, 기록물을 안전하게 장기보존할 수 있어야 한다. 둘째, 현재의 기술 수준을 고려하여 구현하기에 까다로움이 없어야 한다. 셋째, 실행이 불편하지 않아야 한다. 넷째, 비용의 부담이 적어야 한다.

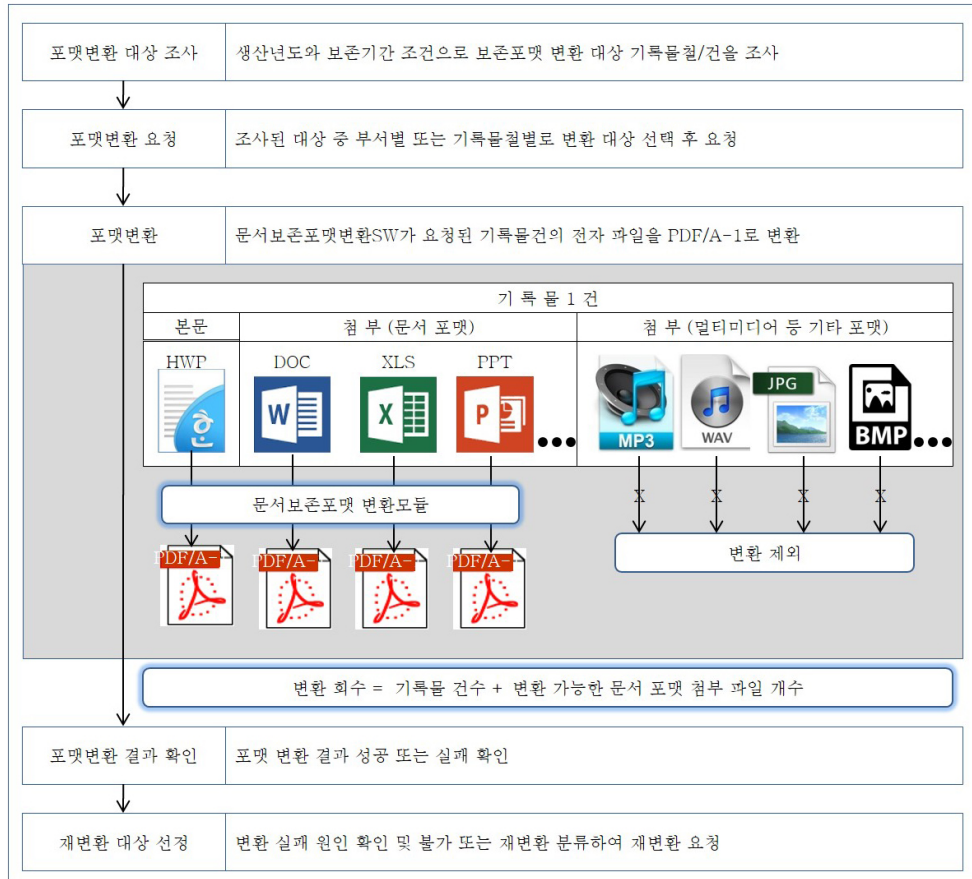
3. 문서보존포맷의 재설계 방향

3.1 현행 문서보존포맷 변환과정과 시사점

기록관리시스템에서 본문과 첨부물을 문서보존포맷으로 변환하기 위하여 보존기간으로 대상을 조사한 뒤 부서별 또는 기록물칠별로 변환을 요청하면 포맷변환 서버에 설치된 포맷변환 모듈이 전자기록물건에 포함된 전자 파일을 문서보존포맷(PDF/A-1)으로 변환한다. 기록관리시스템 도입 시 윈도우가 탑재된 변환서버와 문서보존포맷 변환모듈이 구축되며, 변환모듈이 대상 파일을 PDF/A-1로 변환하여 저장한다. 문서형식이 아닌 시청각 유형(동영상, 음성 파일 등)이나, PDF, EXE 등은 변환 대상에서 제외되며 사유가 자동 입력된다. 그 과정은 <그림 3>과 같으며, “포맷 변환” 단계에서 개별 파일을 하나씩 변환하게 된다. 보통 공문서 전자기록물 1건은 본문 파일 1개와 0개 이상의 첨부 파일들로 구성되므로 변환 대상 파일의 개수는 “기록물 건수 + 변환이 가능한 문서 포맷의 첨부 파일 개수”로서 기록물 건수를 훨씬 상회하게 된다.

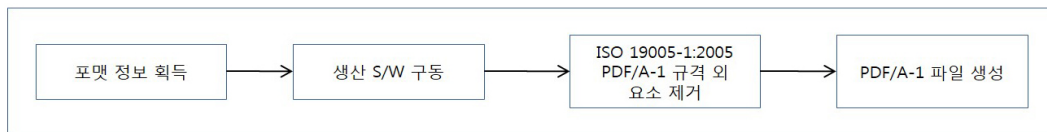
포맷 변환 단계의 상세 구현 방안은 변환모듈에 따라 차이가 있으나 기본적으로 <그림 4>와 같은 단계를 거쳐 진행된다.

순차적으로 1개씩 파일을 취한 뒤 우선 이 파일의 포맷정보를 획득한다. 포맷정보를 획득하는 방법은 이를 구현한 변환모듈마다 다를 수 있는데 파일 확장자를 이용하는 방법 또는 변환을 요청할 때 파일 정보로 별도로 제공하는 방법 등이 있다. 파일 확장자를 이용하는 경우 확장자가 잘못 주어져 있으면 포맷을 잘못 인식하



〈그림 3〉 문서포맷 변환 과정

출처: 표준RMS에서 구현되는 문서보존포맷 변환과정을 저자 구성



〈그림 4〉 PDF/A-1 변환 단계

출처: 표준RMS에서 구현되는 문서보존포맷변환 기능을 저자 구성

여 변환에 실패하게 된다. 확장자 오류는 현재 까지 알려진 변환 오류에서 가장 큰 비중을 차지한다(국가기록원 RMS유지보수사업단 제공 RMS오류코드). 알아낸 포맷에 따라 이를 생산

한 응용S/W 또는 해당 포맷을 열 수 있는 호환 S/W를 API를 통해 구동하여 열거나 내용에 접근해야 하는데, 이 때 반드시 변환을 수행하는 컴퓨터에 관련 응용S/W나 호환S/W가 미리

설치되어 있어야 한다. 포맷변환 단계는 개별 파일을 해당 파일을 생산했던 응용S/W(예, 한/글, 엑셀, 파워포인트 등)로 열어 PDF/A-1로 저장 또는 파일로 인쇄할 때와 동일하다. 그리고 무엇보다 아래와 같이 PDF/A-1 규격에 맞추어 일부 요소(암호화, 내장파일, LZW압축, 투명성, 멀티미디어, 자바스크립트)를 제거하는 것이 표준 포맷 변환 과정에서 필수적이다.

- 폰트(Embedding): 시스템 폰트 사용 금지 (Embedding 되지 않은 폰트 사용 금지)
- 암호화: 암호화 금지
- XMP 메타데이터(Document Information): 문서정보와 XMP 메타데이터 삽입
- 압축 필터: 금지된 압축필터 사용 금지
- 첨부파일: 첨부파일 금지
- Optional Content: 선택적인 보기모드 금지
- 투명도(이미지,XObject): 키워드 상의 투명도 금지
- Annotation(첨부파일, 동영상, 사운드): 첨부파일, 동영상, 사운드 주석 금지
- Action(자바스크립트, Launch): 자바스크립트 삽입 금지
- Interactive Form: 폼 금지

위 규격에서 보듯이 PDF/A-1의 “A”는 기록물(Archive)를 나타내며, “1”은 인쇄가 가능한 모든 매체를 대상으로 하고 있다는 의미이다(NAK/TS 2). 즉, 새로운 메타데이터 추가와 일부 요소 제거 후 최종 저장된 PDF/A-1 포맷의 문서는 육안으로는 파일을 인쇄한 결과와 동일하다. 따라서 작성자가 인쇄를 염두에 두고 서식을 지정하지 않은 경우 전혀 기대하

지 않은 변환 결과를 얻을 수 있다. 원형 보존이 어려운 예로 엑셀 포맷을 들 수 있다. 엑셀 포맷은 국가기록원이 2016년, 85개 공공기관이 운영하는 전자결재, 기록관리 및 고유업무 관련 시스템 995개를 대상으로 조사한 포맷 보유 현황에서 HWP, PDF에 이어 68.8%의 시스템이 세 번째로 가장 많이 생산하는 문서 포맷이다(행정자치부, 2016b). 생산자가 인쇄된 결과를 염두에 두고 엑셀 포맷 문서의 용지, 여백, 확대/축소 비율을 미리 설정해두지 않으면 변환된 PDF/A-1 문서에서는 행·열이 여러 페이지로 분리되거나, 컬럼이 좁아 숫자가 #로 표시되어 내용의 이해가 불가능해진다. 다수의 시트가 포함된 경우 최종 저장 시 활성화된 시트만 변환되어 다른 시트는 보존하지 못 하는 결과를 가져온다. 컬럼 너비를 조정하지 않고 변환할 경우 <그림 6>과 같이 여러 시트 중 활성화 되어 있던 첫 번째 시트가 내용을 이해할 수 없는 형태로 PDF/A-1 파일이 생성된다.

예시로 살펴본 엑셀의 간단한 사례 외에도 공문서의 대표적인 3개 포맷인 HWP, XLS, PPT를 대상으로 문서보존포맷으로 변환 시 소실되는 기능을 임진희, 최주호, 이재영(2014) 연구는 <표 3>과 같이 밝혔다.

PDF/A-1 포맷으로 변환하려면 먼저 파일을 열어야 한다. 그러나 문서 내 이미지 유실, 데이터 왜곡, 외부 링크 데이터 연계 오류, 확장자 누락, 포맷 정보 오류, 손상된 파일, 암호 적용으로 인해 문서 열기에 실패하면 변환에 실패하게 된다. 국가기록원에서 제공한 표준RMS 오류코드를 살펴보면 파일을 생산한 S/W의 부재, 생산한 한/글버전과 변환서버에 설치된 한/글버전 차이, 하드웨어 또는 운영체제 오작동으로

〈그림 5〉 변환 전 엑셀 파일 원본

〈그림 6〉 변환 후 PDF/A-1

인한 한글 작동 실패가 보고되었고, 파일 내용에 삽입된 외부 이미지 및 기타 객체, 파일 손상 역시 변환 실패의 원인이 되고 있다(국가기록원 RMS유지보수사업단 제공 RMS오류코드).

지금까지 살펴본 바와 같이 현행 문서보존포맷은 변환 실패와 원형 변형의 문제점을 안고 있어 기록물의 품질을 보장할 수 없다. 같은 이유

로 국가기록원은 2015년 국내 공공기관에 기록관리시스템이 구축된 이래 최초로 영구기록물관리기관에서 기록물을 인수한 후 육안검수가 필요했을 것이다. 국가기록원은 성공적인 전자기록물 인수를 위해 전자기록물 오류 확인 및 상태 점검 등 이관 단계에서부터 신뢰할만한 검수체가 필요하다는 취지로 기록물 164만 건(파일

〈표 3〉 문서보존포맷 변환 시 소실되는 기능

애플리케이션	기능성 명칭	설명
한글 (HWP)	OLE 개체	문서 내에 Windows에 설치된 프로그램들이 자신의 기능을 서로 사용하는 기능
	메모장	교정부호 단어나 문장이 틀렸을 때 수정할 수 있게 하는 기능
	양식 개체	명령단추, 선택상자, 라디오단추, 입력상자, 목록상자 등을 사용하는 기능
	чек갈피	문서를 편집하는 도중에 표시를 해 두었다가 바로 이동할 수 있게 하는 기능
	스크립트	양식 개체와 연결하여 여러 가지 선택을 할 수 있게 하는 기능
엑셀 (XLS)	스파크라인	단일 워크시트 셀 안의 작은 차트로써 데이터의 추세를 시각적으로 표현하는 기능
	메모장	셀 단위로 부가적인 설명을 추가하는 기능
	차트	셀 단위의 정보를 이용하여 시각적으로 표현하는 기능
	수식정보	셀 단위의 정보를 이용하여 셀에 합계, 평균과 같은 수식을 이용하여 계산하는 수학적 기능
	접힌 셀	셀이 겹쳐있는 경우, PDF/A로 변환할 때 접힌 상태로 변화되기 때문에 문제 발생
파워포인트 (PPT)	삽입된 비디오 파일	문서 내에 비디오 파일을 플레이 할 수 있게 하는 기능
	삽입된 오디오 파일	문서에서 오디오 파일을 삽입할 수 있고, 오디오 링크를 클릭하여 오디오를 플레이하는 기능
	OLE 개체	Windows에서 제공하는 외부 응용 프로그램의 결과물을 삽입하는 기능
	애니메이션	프리젠테이션을 하기 위해 화면에 나타내기, 날아오기 등의 기능을 사용하여 부가적인 설명을 하는 기능

출처: 임진희, 최주호, 이재영(2014)

254만 개)에 대한 육안검수를 과업으로 하는 “2015년 전자기록물 육안검수 용역사업”(사업비 약 6억 2천만원)을 진행한 바 있다(국가기록원, 2015b).

3.2 ODF 중심의 문서보존포맷 선택 전략

신전자문서시스템 보급이 시작된 2004년 이래 한/글 편집기의 연계는 강제 사항이 아니었다. 그러나 공공기관에 이미 한/글 편집기가 지배적으로 보급되어 있었기 때문에 새로 도입된 신전자문서시스템에는 이견없이 한/글 편집기를 기안기로 채택되었다. 이후 개발된 정부업무 관리시스템 역시 기안기에 대한 변경 없이 한/글 편집기를 기반으로 문서 포맷을 HWP로 구현하였다.

그러나 2015년 정부 클라우드 업무 환경 기반을 조성한 행정자치부는 최우선 대상으로 온-나라시스템을 클라우드로 전환 개발하였으며, “문

서 생산 시부터 HWP포맷이 아닌 개방형 표준 포맷(ODF, PDF 등)으로 생산 필요”에 의해 “문서생산 및 유통포맷에 개방형 문서포맷 적용”하고자 (1) 클라우드 온-나라를 기반으로 기존 HWP 기반의 생산 및 유통 포맷을 ODF 또는 PDF로 변경 적용, (2) 아래한글 기안기를 클라우드 온-나라에 적용된 웹기안기를 적용하여 대체 구축하고 있다(행정자치부, 2015). 행정자치부는 클라우드 온-나라시스템으로 일원화할 계획이므로 향후 공문서 포맷은 공식적으로 ODF로 정착될 것으로 전망된다.

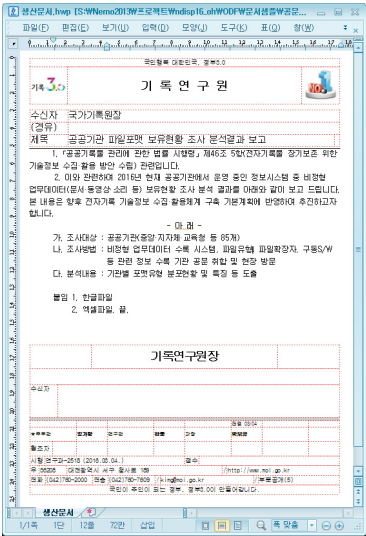
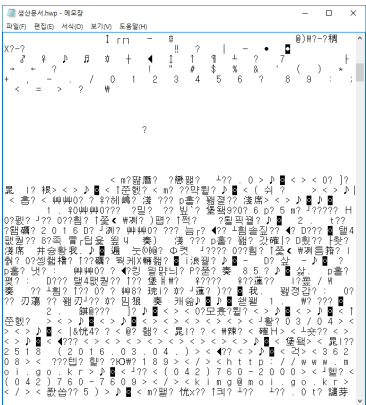
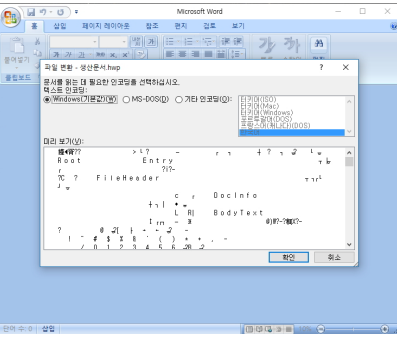
본문 작성 시 웹기안기를 사용하고 ODF로 저장한다면 기관에서 반드시 한/글 편집기를 구매할 필요가 없게 되며 그 결과 한/글 편집기의 대안을 찾게 될 것이다. 정부부처 및 공공기관들의 발 빠른 오픈소스 S/W 활용 동향을 고려할 때, 개인 문서 편집 도구 역시 ODF를 지원하는 오픈소스 편집기 활용을 기대해볼 수 있다.

차세대 공문서 포맷으로 ODF가 유력한 만큼 ODT를 확장자로 갖는 ODF의 워드프로세서 문서의 보존 전략은 달라져야 할 것이다. <표 4>에서 보듯이 HWP 포맷은 한/글에서만 재현 가능하나, ODT 포맷 문서는 다양한 뷰어에서 재현 가능하므로 표준 포맷인 PDF/A-1로 변환하는 마이그레이션 전략 외에 새로운 전략을 검토해볼 수 있다.

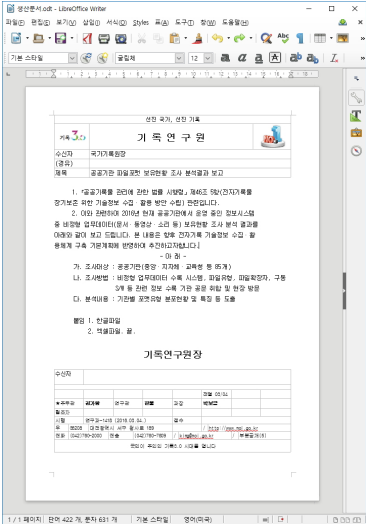
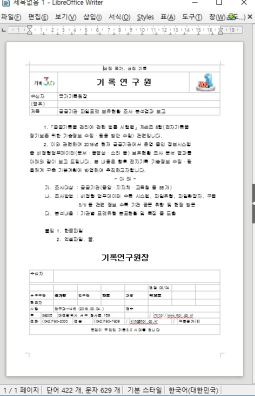
한/글 2000 이상 버전에서 생산한 HWP 포맷은 한/글에서만 열람, 편집이 가능하며, Microsoft

Office Word 2007·2010 등에서 열게 되면 인식이 되지 않아 열리지 않는다. 또한 이를 텍스트 편집기에서 열어보면 한/글만이 인식하는 바이너리 값의 집합으로 표현된다. 반면 OpenOffice에서 편집한 ODT파일은 다른 오픈소스 편집기 LibreOffice에서 동일하게 재현되며, 약간의 서식 변형은 있지만 상용 문서 편집기인 폴라리스 오피스와 심지어 한/글로도 열 수 있다. 메모장에서 열어보면 태그로 표현된 문서 구조 속에 텍스트가 포함되어 있어 문서의 기본적 내용을

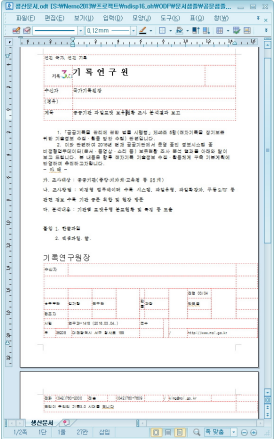
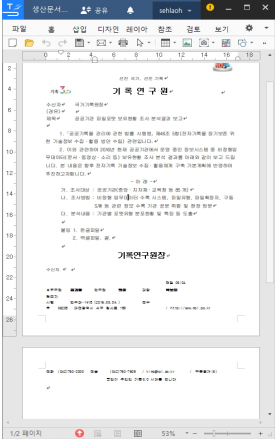
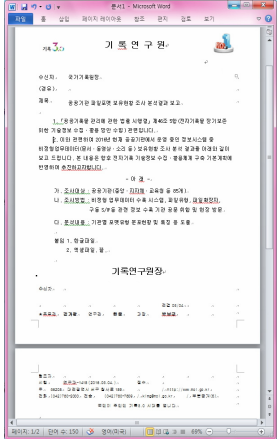
< 표 4 > HWP 포맷 공문서를 다양한 편집기에서 연 결과

HWP 포맷 공문서	
한/글(2010)에서 열었을 때	타 편집기에서 열었을 때
	<div style="text-align: center;"> <p><메모장></p>  </div> <div style="text-align: center; margin-top: 20px;"> <p><Microsoft Office Word(2007)></p>  </div>

〈표 5〉 ODF 포맷 공문서를 다양한 편집기에서 연 결과

ODF 포맷 공문서	
OpenOffice(4.1.2)에서 열었을 때	타 편집기에서 열었을 때
	<p>〈메모장〉</p> <pre> 0 316 </text:p> 317 </table:table-row> 318 </table:table-row> 319 </table:table> 320 <text:p <text:style-name='P157'> 321 <text:p> 322 <text:tab/1. 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 323 </text:p> 324 <text:p <text:style-name='P157'> 325 <text:p/1. 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 326 </text:p> 327 <text:p <text:style-name='P157'> 328 <text:p/1. 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 329 </text:p> 330 </text:p> 331 <text:p <text:style-name='P157'> 332 <text:p/1. 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 333 </text:p> 334 </text:p> 335 <text:p <text:style-name='P157'> 336 <text:p 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 337 </text:p> 338 </text:p> 339 </text:p> 340 <text:p <text:style-name='P157'> 341 <text:p 기록 연구원 제 462 (보) 기록 연구원 장기보존특별 기록연구 342 </text:p> 343 </text:p> 344 <text:p <text:style-name='P157'> 345 </text:p> 346 </text:p> </pre> <p>〈LibreOffice (5.1)〉</p> 

〈표 6〉 ODF 포맷 공문서를 상용 소프트웨어에서 연 결과

ODF 포맷 공문서를 상용 소프트웨어로 열었을 때		
한/글(2010)	Polaris Office(8.1.152)	Microsoft Office Word(2007)
		

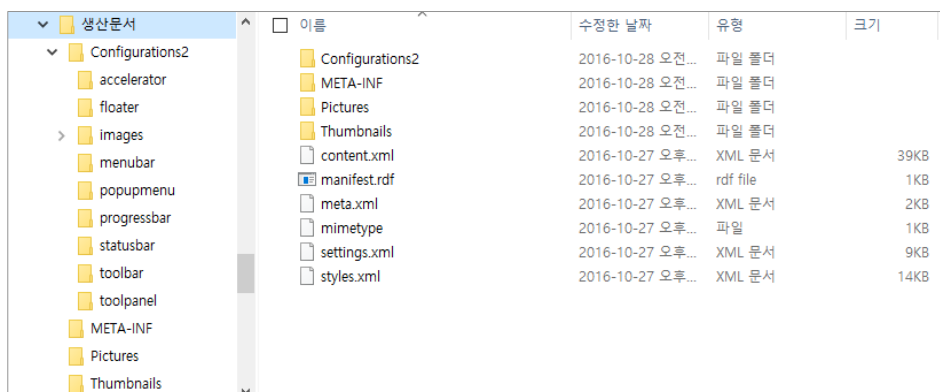
이해할 수 있다. 또한 ODF의 규격은 공개되어 있으므로 다양한 뷰어의 제작이 가능하며 현재 이미 여러 응용S/W가 만들어져 사용 중이다. 그렇다면 표준 문서 포맷인 ODF의 문서를 데이터 및 표현 속성을 잃으면서 다른 표준 포맷으로 변환할 필요는 없다.

왜 문서보존포맷으로 PDF/A-1을 정하여 모든 파일을 변환하여 보존하게 되었는지 이유를 상기해보면, 공문서 파일의 지배적인 포맷인 HWP가 “한글과컴퓨터”가 개발하여 한/글에서만 읽을 수 있는 포맷으로서 미래에도 이 포맷이 유지되고 기록을 읽을 한/글 소프트웨어가 있을지 장담할 수 없기 때문이다. 그러나 <표 6>에서 보듯이 ODT 포맷은 이미 사용이 보편화된 다양한 소프트웨어를 이용해 읽을 수 있으므로 또 다른 보존 포맷으로 변환할 필요가 없다.

개방형 문서 포맷을 채택하는 가장 큰 이유는 특정 개발 업체의 기술과 준폐로부터 독립적이라는 점이다. 앞 장에서 살펴본 바와 같이 국가기록원 “전자기록물 문서보존포맷 기술규격(NAK/TS 2)”은 문서보존포맷 고려요소로서 여덟 가지 기준을 제시하고 있다. 이중 “공개

용 표준”이란 특정 업체가 독점적으로 소유권을 가지고 있지 않아야 하며 누구나 참조하고 이용할 수 있게 공개되어야 하고, 사용자에게 경제적 영향을 주지 않아야 하는 요구이다. ODF는 다른 어떤 포맷보다 “공개용 표준” 요건을 충족시키며, 애플리케이션의 버전 및 소멸, 운영체제 또는 플랫폼 차이에 영향 받지 않고 문서 내용과 구조 및 표현을 재현할 수 있다는 점에서 안정성, 표현력, 상호운용성, 진본성, 편재성 요건도 충족시킨다. ODF는 XML 기반으로서 메타데이터를 담기에 최적의 형식이며, 검색 역시 가능하다. 따라서 모든 요건을 충족시키는 ODF는 생산 포맷으로서만이 아니라 보존 목적에도 적합한 포맷이라 볼 수 있다.

보존 전략 재검토를 위해 ODF 포맷의 텍스트 문서 규격인 ODT 포맷을 좀 더 살펴보겠다. ODT 문서는 단일 1개의 파일일 수도 있으나 보통 패키지로서 폴더와 파일을 포함하여 압축되어 있다. OpenOffice4.1.2에서 생산한 문서 한 건에 해당하는 파일, “생산문서.odt” 패키지를 풀면 <그림 7>과 같은 구조로 Configurations2, META-INF 등의 폴더와 content.xml, styles.xml



<그림 7> ODT 포맷의 내부 구조

며 이해하기 쉬운 형식으로서 문서 파일 자체의 보존과 문서 파일에 대한 기록관리 메타데이터를 문서 파일에 함께 넣어 보존할 수 있는 가능성을 열고 있다. 보통 기록관리 단계에서 일어난 다양한 관리 및 활용 행위들은 데이터베이스에 저장하고 관리하고 장기보존포맷 생성 시 별도의 파일에 내역을 기술하여 같이 패키지로 만든다. 그러나 ODF는 파일 내용 외에도 보다 풍부한 정보를 담을 수 있는 형식이므로 기록관리 내역을 보존 파일에 함께 추가하면 파일 1개가 파일의 내용은 물론 기록관리 메타데이터까지 포함하여 자체적으로 완결된 보존 형식이 될 수 있다.

기록관리 속성 값과 기록관리 이력을 문서 파일에 통합하여 보존 포맷을 생성하는 방안을 3가지로 검토해보았으며, 이에 대한 개념도는 <표 8>과 같다. 1안은 생산 포맷인 ODF 규격을 변형 없이 그대로 이용하는 방안이며, 2안은 규격을 확장하는 방안, 마지막으로 3안은 ODF 외부에 메타데이터 파일을 추가하는 방안이다.

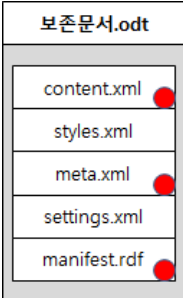

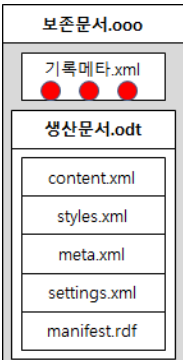
제1안은 ODF 규격에 정의된 메타데이터 규격을 이용, 속성 값을 넣어 다시 ODF 파일을 생성하는 방안이다. ODF 규격에서 허용하는 메타데이터는 ① W3C Resource Description Framework(RDF)에 따라 ODF 내 파일과 위치를 선언한 메타데이터 선언 파일(manifest.rdf)과 ② content.xml이나 styles.xml에 RDF로 기술된 메타데이터와 ③ meta.xml 기술된 18개 메타데이터 항목, ④ 사용자 정의 메타데이터 항목이다. 이 중 RDF로 기술된 메타데이터는 기록관리 속성을 표현하는데 제약적이며, 기 정의된 메타데이터는 제목, 저자, 작성 소프트웨어, 검색어 등 18개에 제한되어 이 역시 충분하지

않다. 마지막으로 사용자 정의 메타데이터 항목은 <meta:user-defined>로 표현되는 1개의 항목으로서 하위에 다른 항목은 포함할 수 없고 텍스트만 기술할 수 있기 때문에 1개의 기록관리 속성 값 표현만 가능하다. 단, 다수의 에트리뷰트를 임의로 정의할 수는 있으나 다양한 데이터 항목을 기술할 수 없기 때문에 ODF 규격 내에서 기록관리 메타데이터를 추가하기에 제한적이다.

제2안은 ODF 규격에서 정의되어 있지 않은 엘리먼트를 추가 정의하고 메타데이터를 저장한 확장 ODF 파일을 생성하는 방안이다. ODF 규격, "Information technology - Open Document Format for Office Applications (OpenDocument) v1.2"의 Part1 OpenDocument Schema와 Part3 Packages는 각각 OpenDocument Extended Document와 OpenDocument Extended Package를 허용하고 있다. 규격에 정의된 XML 엘리먼트의 엘리먼트와 에트리뷰트를 정의하고 이용할 수 있는 여지를 두고 있다. 따라서 OpenDocument Extend Document로서 기록관리 메타데이터를 추가 정의하고 파일을 추가하여 Extended Package로서 패키징한 ODF의 확장 버전을 문서보존포맷으로 정할 수 있다. 확장 ODF에 대해서는 현재 보편적으로 사용되고 있는 OpenOffice나 LibreOffice가 인식하지 못 하므로 재현 시 추가한 기록관리 메타데이터가 조회되지 않기 때문에, 고유 뷰어가 개발되어야 한다. 이 경우 관리자의 입장에서 확장된 포맷임을 표시하기 위해 확장자를 'odt'에서 'odtx'와 같은 방식으로 변경하는 것도 고려해 볼 수 있다.

제3안은 ODF 파일 외에 별도의 메타데이터 파일을 추가하여, ODF와 메타데이터를 다시

〈표 8〉 ODF파일의 보존포맷 생성 3안

	방안	포맷	개념도
1안	ODF 규격에 정의된 메타데이터 규격을 이용, 속성 값을 넣어 다시 ODF 파일 생성	ODF	 <p>● 기록관리 메타데이터</p>
2안	ODF 규격에서 정의되어 있지 않은 엘리먼트를 추가 정의하고 메타데이터를 저장한 확장 ODF 파일 생성	확장 ODF (OpenDocument Extended Document)	 <p>● 기록관리 메타데이터</p>
3안	ODF 파일 외에 별도의 메타데이터 파일을 추가하여, ODF + 메타데이터를 다시 패키지화한 신규 포맷 생성	신규 패키지	 <p>● 기록관리 메타데이터</p>

패키지화한 신규 포맷 생성 방안이다. 기록관리 메타데이터를 별도 파일에 기술한 뒤 생산 ODF 파일과 함께 패키징한 파일을 보존한다. 패키징 방법은 tar, zip, bagIt, gzip 등 다양하며

압축여부, 손실여부, 해제 용이성 여부 등을 고려하여 결정할 수 있다. 이 방안에서 문서의 내용은 OpenOffice, LibreOffice 등 ODF 호환 뷰어로 열고 읽을 수 있다. 그러나 추가한 기록

관리 메타데이터는 별도의 XML 파일이므로 일반 텍스트 편집기로 읽을 수 있으나 사용자가 이해하기 쉬우려면 전용 XML 뷰어가 개발되어야 하며, 새로 정의한 포맷임을 표시하기 위하여 일반 확장자와 중복되지 않는 새 확장자를 부여해야 한다.

위의 3가지 방안은 생산 포맷을 유지하며 메타데이터를 추가하는 방법의 차이로서 생산 문서 포맷을 변환하지 않기 때문에 원문의 포맷 변환 과정이 없어진다. 변환 과정을 제거함으로써 첫째, 원문의 내용과 기능을 모두 최대한 보존할 수 있고, 둘째, 변환 작업에 투입되는 시간과 인력을 절약할 수 있다. 각 방안 별 구체적 구현 방안과 전용 뷰어에 대해서는 향후 문서보존포맷만을 주제로 한 세부 연구 과제로 남기고, 본 연구는 방안 제시로 범위를 제한하겠다.

3.3 독자포맷의 문서보존포맷 선택 전략

이상 검토한 ODF 생산 문서의 보존전략과 설계 방향은 본문 파일의 보존에 적용될 수 있으나, 첨부 파일은 다양한 포맷으로 만들어지며 첨부 파일의 포맷을 ODF와 같은 표준포맷으로 모두 통일하기는 어려운 일이다. 국가기

록원에서 2016년 85개 공공기관을 대상으로 전자결재, 기록관리, 고유업무 관련 시스템 995개에서 파일 포맷을 조사한 결과에 따르면 문서 파일 포맷 수 49개, 동영상 파일 포맷 수 48개, 소리 파일 포맷 수 23개, 이미지 파일 포맷 수 38개로서 다종 및 다수의 파일 포맷이 존재함을 알 수 있다(행정자치부, 2016b). 본문 파일 포맷은 ODF가 보편화된다 할지라도 첨부 파일은 다양한 포맷으로 존재할 것으로 보며 현재의 독자 포맷도 계속 존재하고, 향후 새로운 독자 포맷의 출현 가능성도 배제할 수 없다. 이에 독자 포맷의 장기보존 전략 또한 수립할 필요가 있다. 그 첫 번째 방안은 현행의 문서보존 전략을 그대로 유지하여 PDF/A-1로 변환하는 방안, 두 번째 방안은 PDF/A-1 대신 ODF를 문서보존포맷으로 정하여 변환하는 방안이 있으며, 장단점을 비교하면 <표 9>와 같다.

현행 전략의 유지는 무엇보다 이미 투자한 비용과 시범 운영을 통해 축적한 경험을 활용할 수 있다는 장점이 있고, 이 장점은 반대로 문서보존포맷으로 ODF를 제정하는 방안의 단점으로서, 새로운 전략을 국내 현실에 적용하기 위한 연구와 개발이 선행되어야 한다는 부담이 있다.

<표 9> 독자포맷 보존전략 비교

	현행 문서보존포맷 변환	ODF 포맷 변환
보존 포맷	PDF/A-1	ODF
대상 포맷	문서	문서, 소리, 이미지
장점	- 기구축 하드웨어 및 소프트웨어 자원 유지 - 2015년부터 실제 보존포맷 변환 경험 축적	- 문서 내용을 다양한 기능 포함 최대 보존 - 드로잉, 소리 포맷 변환 가능 - 문서보존포맷을 ODF로 통일
단점	- 문서 내 기능 소실 - 문서 포맷만 변환 가능 - 문서보존포맷의 다원화(PDF/A-1과 ODF)	- 신기술 적용에 따른 초기 연구 및 개발 부담

그러나 기록을 본래 생산 당시 모습으로 유지한다는 보존 목적을 고려할 때 ODF 포맷으로 변환하여 보존하는 방향으로 가야할 것이다. 앞서 살펴 본 PDF/A-1 변환 시 소실되는 기능들이 ODF 변환 시 보존되는지 알아보기 위하여 대표적인 ODF 편집기인 OpenOffice와 LibreOffice에서 해당 요소들이 포함된 HWP, XLS, PPT 파일을 열어보았다. 그 결과 한글 (HWP) 포맷은 인식이 되지 않았으며, 엑셀 포맷은 스파크라인 외의 모든 기능들이 보존되었고, 파워포인트 포맷에서 애니메이션과 일부 OLE 개체는 유지되었다(〈표 10〉 참조).

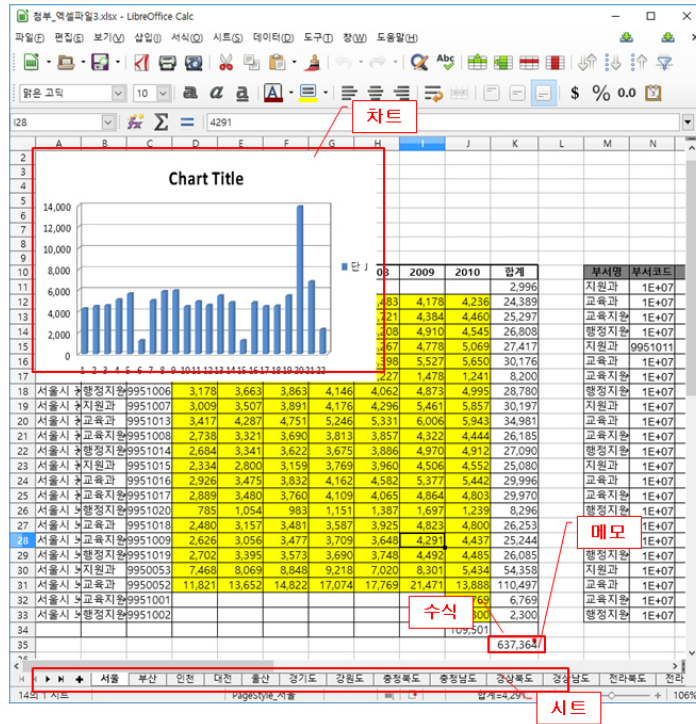
특히 엑셀 포맷은 PDF/A-1에서 소실되는 거의 모든 요소들이 그대로 남아있어 보존에 보다 적합하다고 볼 수 있다. 엑셀 포맷 파일을

LibreOffice를 이용하여 변환한 ODF를 캡처한 〈그림 8〉에서 그대로 남아 있는 차트, 메모, 수식, 시트를 확인할 수 있다.

현재로서는 ODF도 원본의 모든 요소를 완벽하게 보존한다고 할 수는 없지만, 그림에도 문서 내용에 멀티미디어와 타 포맷 개체를 포함한 전자 문서에 대해서는 PDF/A-1보다 높은 보존율을 보이기 때문에 독자포맷의 보존포맷으로서 고려할 가치가 있다. 더욱이 공개포맷이 점차 보편화될 것으로 예상되므로 변환 도구도 다양해지고 정밀해질 것으로 기대된다. 국내 독점 포맷인 HWP는 아직 ODF 편집 도구에서 인식하지 못 하지만 이미 공개되어 있는 HWP 개발자용 API를 이용한 변환 도구 개발도 가능하며, 반대로 HWP를 ODF로 저장하려

〈표 10〉 문서보존포맷과 ODF 변환시 소실되는 기능 비교

애플리케이션	기능성 명칭	설명	OpenOffice 4.1.2	LibreOffice 5.1.0.3
한글 (HWP)	모든 기능	문서 내용 및 포함된 모든 기능	X	X
엑셀 (XLS)	스파크라인	단일 워크시트 셀 안의 작은 차트로서 데이터의 추세를 시각적으로 표현하는 기능	X	X
	메모장	셀 단위로 부가적인 설명을 추가하는 기능	○	○
	차트	셀 단위의 정보를 이용하여 시각적으로 표현하는 기능	○	○
	수식정보	셀 단위의 정보를 이용하여 셀에 합계, 평균과 같은 수식을 이용하여 계산하는 수학적 기능	○	○
	접힌 셀	셀이 접혀있는 경우, PDF/A로 변환할 때 접힌 상태로 변화되기 때문에 문제 발생	○	○
파워포인트 (PPT)	삽입된 비디오 파일	문서 내에 비디오 파일을 플레이 할 수 있게 하는 기능	X	X
	삽입된 오디오 파일	문서에서 오디오 파일을 삽입할 수 있고, 오디오 링크를 클릭하여 오디오를 플레이하는 기능	X	X
	OLE 개체	Windows에서 제공하는 외부 응용 프로그램의 결과물을 삽입하는 기능	△	△
	애니메이션	프리젠테이션을 하기 위해 화면에 나타내기, 날아오기 등의 기능을 사용하여 부가적인 설명을 하는 기능	○	○



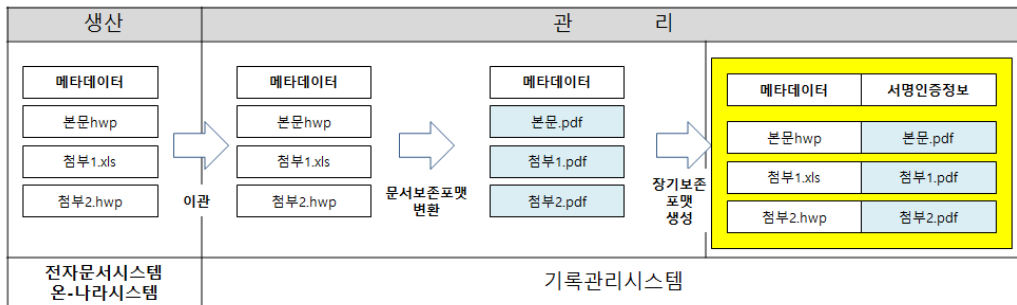
〈그림 8〉 LibreOffice 5.1.0.3에서 연 엑셀 포맷 파일

는 요구사항이 많아져 한/글의 ODF 포맷으로 저장 기능이 사용 가능한 수준으로 향상될 가능성도 있다. 이러한 현실적 조건이 구비됨에 따라 독자포맷의 보존포맷은 선택의 폭이 넓어진다. 컴퓨팅 환경에서 전자적인 재현보다는 인쇄 결과물의 시각적 완성도에 민감한 문서의 보존에는 PDF/A-1가 적합함은 부인할 수 없다. 특히 편재성, 뷰어 호환성, 포맷의 지속가능성에서 PDF/A-1은 이미 검증된 포맷이다. 그러나 PDF/A-1에서 손실이 많거나 변환 불가능 기록까지 획일적으로 PDF/A-1 변환을 고집할 것이 아니라 ODF를 선택할 수 있도록 추가 고려할 것을 제안하는 바이다.

4. 장기보존포맷의 재설계 방향

4.1 현행 NEO 생성 및 관리과정과 시사점

현행 장기보존전략 하에서는 생산 기록이 기록관리시스템으로 이관된 뒤 문서보존포맷으로 변환된 후 장기보존포맷으로 변환되는 과정을 간략히 표현하면 〈그림 9〉와 같다. 본문(hwp)과 첨부1(hwp), 첨부2(xls) 파일로 구성된 기록물건을 예로 들면, 원문 파일과 생산 메타데이터가 기록관리시스템으로 이관된다. HWP 포맷과 엑셀 파일 포맷을 먼저 문서보존포맷으로 변환하고 원문, 문서보존포맷, 기록관리 메타데이터, 전자서명을 넣어 장기보존포맷 패키지(NEO)



〈그림 9〉 장기보존포맷 생성 과정

를 만든다. 장기보존포맷은 기록물철과 기록물건에 대하여 각각 정의되어 있는데, 전자기록 컴포넌트를 포함하는 포맷은 기록물건이므로 본 연구는 기록물건 장기보존포맷만을 대상으로 하였다.

기록물건 장기보존포맷 구성 요소는 크게 메타데이터, 원문 파일, 문서보존포맷 파일, 전자서명이다. 메타데이터는 해당 기록의 생산부터 보존의 전 과정을 기술한 정보로서 생산자, 기록물명, 보존기간, 분류, 주제, 생산이력, 관리이력, 보존이력 등의 정보를 전자기록물 장기보존포맷 기술규격(NAK/TS 3:2013(v2.0))에서 정의한 장기보존포맷 스키마를 따라 XML 형식으로 작성한다. 원문 파일과 문서보존포맷 파일은 디지털 컴포넌트로서 각각 컴포넌트 식별자, 포맷, 포맷버전, 어플리케이션, 크기 등의 메타데이터를 기술하고 원문 파일을 Base64 방식으로 인코딩한 텍스트 스트림을 붙인다. 인증정보는 전자기록물의 진본성 및 무결성 보장을 위해 이용되는 정보로, 기관인증서와 서명정보 및 잠김인증정보로 구성된다. 현행 인증 정보에 대하여 최근 서울기록원의 정보화전략(ISP)사업 결과는 인증서 대신 해쉬함수 이용 방식의 사례 연구 등을 통해 무결성 보장 방안의 재검토 논의

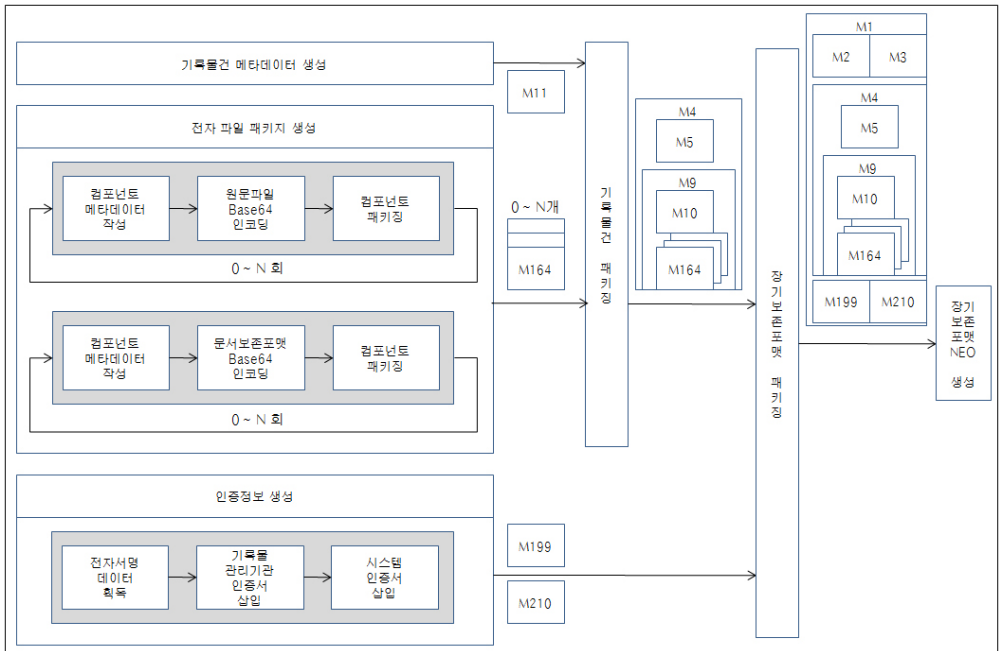
를 이끌어 냈다. 인증 정보 재설계 방안은 또 다른 연구 주제로 연구되어야 할 것으로 보고 본 연구에서는 제외하고자 한다.

상세 과정을 분해하여 보면 〈그림 10〉과 같다. 〈그림 10〉에서 장기보존포맷의 상위 컨테이너를 중심으로 이 컨테이너를 생성하는 프로세스를 표시하고 그 결과 생성되는 컨테이너를 전자기록물 장기보존포맷 기술규격(NAK/TS 3:2013(v2.0))에서 정의한 컨테이너 기호로 표시하였다.

〈그림 10〉에서 프로세스 산출물로 표시된 M1 ~ M210의 의미는 〈표 11〉로 정리하였다.

규격에 따라 생성한 장기보존포맷 샘플을 〈그림 11〉에서 보자. 본문 원문 HWP 포맷 파일과 보존포맷으로 변환된 PDF 포맷 파일이 Base64 인코딩되어 메타데이터와 함께 XML 내 텍스트 값으로 들어가 있다. 앞서 2장에서 지적한 바와 같이 Base64 인코딩에 의한 패키지 방식은 복호화해야 전자 파일을 얻을 수 있고, 1.88~2.25배 용량 증가를 가져오는 문제가 있다.

XML로 패키지화된 장기보존포맷 파일에서 메타데이터와 원문 등을 쉽게 조회하려면 인코딩된 전자 파일을 복호화하여야 하고 XML에



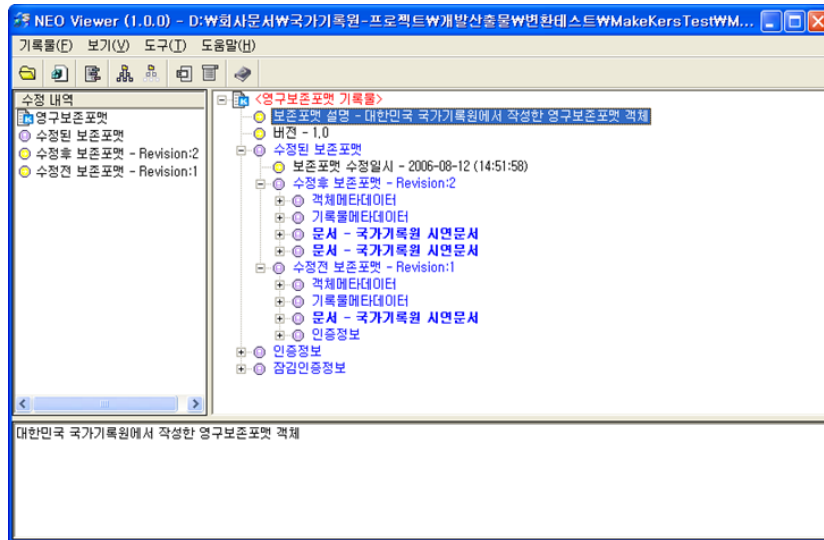
〈그림 10〉 장기보존포맷 변환 과정

출처: 국가기록원 전자기록물 장기보존포맷 기술규격(NAK/TS 3)을 바탕으로 필자 재구성

〈표 11〉 장기보존포맷 구성 컨테이너

기호	엘리먼트	이름	설명
M1	NARSEncapsulatedObject	장기보존포맷객체	장기보존포맷 전체 컨테이너
M2	NEOFormatDescription	장기보존포맷 설명	장기보존포맷 설명
M3	Version	장기보존포맷버전	장기보존포맷 설명
M4	SignedObject	서명된 객체	장기보존포맷 요소 중 전자서명된 부분의 컨테이너
M5	ObjectMetadat	객체메타데이터	장기보존포맷에 대한 기술(技術) 특성에 대한 일반 정보를 포함하는 컨테이너
M9	ObjectContent	객체컨텐츠	기록물건, 기록물철, 수정된 보존포맷을 포함하는 컨테이너
M10	Record	기록물건	정의 기록물건 메타데이터와 컴포넌트를 포함하는 컨테이너
M11	RecordMetadata	기록물건메타데이터	기록물건을 설명하는 메타데이터 요소를 포함하는 컨테이너
M164	Component	컴포넌트	전자기록물 건에 포함된 컴포넌트 메타데이터와 인코딩을 포함하는 컨테이너
M199	SignatureBlock	인증정보	보존포맷의 인증을 위한 전자서명정보를 포함하는 컨테이너
M210	LockSignatureBlock	잠금인증정보	인증정보(M199)의 서명값을 시스템인증서(장기보존포맷 생성시스템에 부여한 인증서)로 전자서명한 서명정보를 포함하는 컨테이너

출처: 전자기록물 장기보존포맷 기술규격(NAK/TS 3:2013(v2.0))



〈그림 12〉 중앙영구기록관리시스템(CAMS)에 구현되어 있는 NEO 뷰어의 예시
출처: 국가기록원(2006a)

대한 지식 없이도 내용물을 볼 수 있는 전용 뷰어를 이용해야 한다. 전용 뷰어에서 장기보존포맷 패키지를 열면 패키지의 구성 요소들이 구조화되어 보이고 분류정보, 생산정보, 이력정보와 원본 파일, 문서보존포맷 파일의 내용이 조회된다. 〈그림 12〉는 NEO 전용 뷰어에서 장기보존포맷 패키지를 조회한 화면의 예이다.

전용 뷰어에서만 장기보존포맷 파일을 열어 볼 수 있기 때문에 처리과에서는 장기보존포맷으로 변환한 뒤 결과를 확인하거나, 오류의 정정을 위해 해체한 뒤 다시 패키지화하고 정상 생성 여부를 확인하기 어렵다.

현행 장기보존포맷파일은 장기보존포맷 파일의 생성만이 아니라 국가기록원에서 인수 후 검수 시에도 어려움이 있다. 메타데이터 데이터 값의 유효성 검증, 문서보존포맷 파일과 원천 파일의 일치성, 파일이 정상적으로 열리는지 검증하려면 전용 뷰어로 하나씩 열어 육안검수를

해야 하는데 장기보존포맷 파일을 암호화 및 복호화 할 때 뷰어가 컴퓨터 자원을 많이 필요로 하여 검수 생산성을 저하시키는 원인이 되고 있다.

4.2 ZIP과 XML 기반의 장기보존포맷 선택 전략

장기보존포맷은 자체 충족성, 자체 문서화, 무결성을 충족시키기 위해 기록건을 구성하는 요소들을 한 개 패키지로 묶어 보존하며, 패키지 하는 방법에 여러 가지 방식이 존재한다. 현행 보존포맷 패키지는 메타데이터를 구조화된 텍스트로 기술하고 바이너리 파일은 Base64 인코딩을 통해 텍스트화하여 XML 구조에 넣는다. Base64 인코딩을 간단히 설명하면 전자 파일의 연속된 bit를 대상으로 6bit씩 잘라 Base64 색인 표에서 그 값에 해당하는 문자로 대체하는 방식으로서 그 결과는 의미 없는 문자의 연속이다.

앞에서 살펴본 <그림 11> 보존포맷 샘플에서 <ComponentData> 컨테이너에 삽입된 Base64 인코딩 결과를 보면 <그림 13>과 같다.

바이너리 파일을 그대로 텍스트 편집기에서 열 수는 없으나 Base64 인코딩을 하면 텍스트 파일과 동일하여 메모장에서 열 수 있다. 하지만 그 텍스트에는 의미가 없기 때문에 복호화하여 본래의 비트스트림으로 복원하여 생산했던 어플리케이션으로 조회해야 한다.

현행 방식의 가장 큰 문제는 Base 64 인코딩 시 원본 3바이트가 인코딩 후 4바이트로 늘어나 파일 크기가 약 33% 증가하므로 보존하는 작업이 거대해지는 문제가 발생하며, 인코딩과 복호화 때 메모리와 CPU 자원 소요, 네트워크 전송량 증가의 문제를 가져온다. 또한 쉽게 패키지를 풀고 XML 구조 속 메타데이터와 전자 파일을 보기 어려워 활용이 어렵고 기록관리 메타데이터 변경 등 수정 시 재패키징 또한 어렵다.

전자기록 보존을 위한 패키지 생성 대안으로서 서울기록원 정보화전략계획(ISP) 사업은 BagIt 규격을 검토하였다. BagIt은 미국 의회도서관(Library of Congress)과 캘리포니아 디지털 도서관에서 대량의 전자 파일을 손실 없이 보관하기 위한 포맷으로 개발하였고 스탠포드 대학교 도서관에서 80TB 이상의 다양한 포맷의 전

자 기록 활용 및 보존에 사용된 공개 패키지 규격이다(서울시, 2016, p. 209). BagIt의 특징은 패키지에 담은 내용물과 설명 표현에 충분한 구조를 제공하면서 그 내용물 자체에 대해서는 알 필요 없기 때문에 규격이 단순하면서 상이한 디지털 컴포넌트를 보존하기 적합하다. 또한 BagIt은 패키지 구조에만 응용할 수 있는 것이 아니고 이관, 무결성 체크(Fixity Check) 등 다수 모듈의 컴포넌트로 이용될 수 있다(안대진, 임진희, 2016).

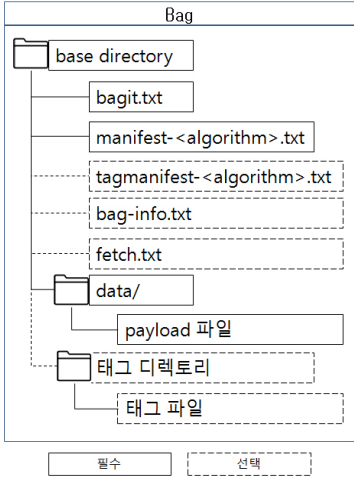
BagIt 포맷은 태그(tag)와 페이로드(payload)로 구성된다. 태그는 BagIt 패키지의 저장과 이동을 위한 메타데이터이고 페이로드는 패키지에 들어가 있는 내용물로서 BagIt 패키지에서는 일련의 비트 값으로만 다를 뿐, 내용물의 패키징을 위해 구조나 의미를 드러낼 필요가 전혀 없다. 이런 이유로 BagIt 포맷의 규격은 상이한 디지털 컴포넌트를 간단하고 제약 없이 패키지로 만들어 보존하거나 옮길 수 있다. 이러한 BagIt에 대한 구조와 구성요소에 대한 설명은 <표 12>와 같다.

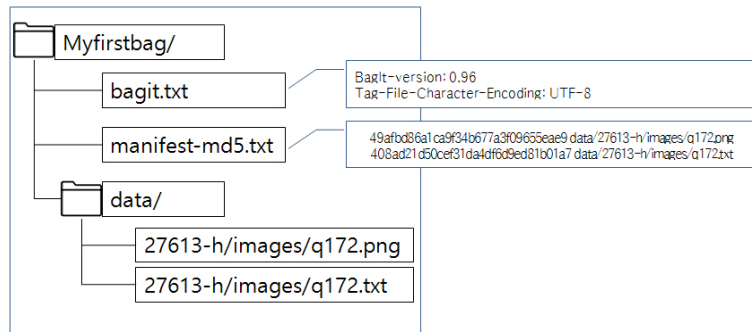
필수 구성 요소만을 가진 BagIt 패키지의 예를 <그림 14>에서 보면 패키지 내용물을 선언하는 manifest 파일의 내용이나 내용물의 파일 구조가 매우 간단하다.

```
<ComponentData>
PD94bWwgdMvyc2lvbj0iMS4wliBibmNvZGluc290iRvVVDUtSlj8+Cjw/eG1sOnN0eWxic2hlZXQg
PSJhZGp1c3QiPiZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNw
OzwvcD48cCBzdHlsZT0iZm9udC1mYW1pbHk6Pz8/OyBmb250LXNpemU6MTRwdDsgbGluc290ZWln
aHQ6Ny45MDJtbTsiGFsaWduPSJhZGp1c3QiPiZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNwOyZuYnNw
OyZuYnNwOyZuYnNwOz8/ID8/PyA/Pz8/ID8/PyA/Pz8gPz8/PyA/Pz8gPz8/Py48L3A+PHAgc3R5
bGU9ImZvbnQtZmFtaWx5Oj8/PzsiGFsaWduPSJhZGp1c3QiPjwvcD48cCBzdHlsZT0iZm9udC1m
</ComponentData>
```

<그림 13> 전자파일의 Base64 인코딩 결과의 예

〈표 12〉 BagIt 구조와 구성 요소

BagIt 구조	구성 요소	필수	설 명
	bagit.txt	○	첫 줄에 BagIt 버전을 명시하고 두 번째 줄에 tag 파일의 문자 세트를 명시한 파일
	manifest- \langle algorithm \rangle .txt	○	- data 디렉토리에 있는 파일과 체크섬 목록 - 파일명에 체크섬 알고리즘 이름 명시 - 알고리즘 선택에 제한 없음
	data/	○	패키지의 내용물에 해당되는 파일이 위치한 디렉토리
	tagmanifest- \langle algorithm \rangle .txt	X	- 추가 태그 파일과 체크섬 목록 - 파일명에 체크섬 알고리즘 이름 명시 - 알고리즘 선택에 제한 없음
	bag-info.txt	X	- bag와 payload를 설명하는 메타데이터 파일 - 인간이 해독 가능하도록 기술
	fetch.txt	X	bag과 함께 전송되어 bag의 유효성 검증 전에 내용물에 추가될 파일 목록
	태그 디렉토리	X	선택적으로 추가할 태그 파일이 위치할 디렉토리
	태그 파일	X	추가적으로 manifest- \langle algorithm \rangle .txt 파일에 선언한 태그 파일



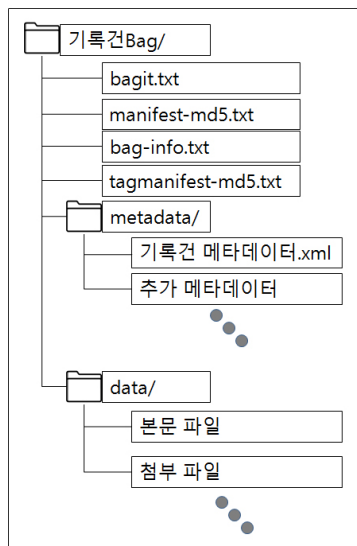
〈그림 14〉 BagIt의 단순한 패키지 예시

출처: The BagIt File Packaging Format (V0.97)

파일 무결성을 검증하기 위한 체크섬 알고리즘은 자유롭게 구현할 수 있으며 manifest 파일명에 붙이는 알고리즘 이름을 영문 소문자와 숫자로 쓰는 규칙만 지키면 된다. 그러나 BagIt을 생성하고 검증 및 활용하는 도구들은 md5와 sha1 알고리즘을 반드시 지원해야 한다.

BagIt 규격은 확장성이 좋아서 인증 정보나 추가되는 정보는 tag 파일로 추가하고 manifest 파일에 선언하면 된다. 따라서 장기보존포맷 패키지도 쉽게 BagIt 규격으로 〈그림 15〉와 같이 설계할 수 있다. 필수 요소인 bagit.txt, manifest- \langle algorithm \rangle .txt, data 디렉토리 및

게 선택 요소인 tagmanifest-〈algorithm〉.txt와 metadata 및 기록건과 파일에 대한 메타데이터를 기술한 태그 파일들을 정의해야 한다. 메타데이터를 1개 파일에 기술할지 아니면 컴포넌트별로 1개 파일에 기술할지 각 전자파일과 1대1 대응하도록 각각의 파일에 기술할지는 더 논의가 필요하며 본 연구에는 최소한의 구조만 설명하였다.



〈그림 15〉 BagIt 규격 장기보존 포맷의 구조

BagIt 장기보존포맷 구성요소의 예시를 〈표 13〉에서 설명하였다.

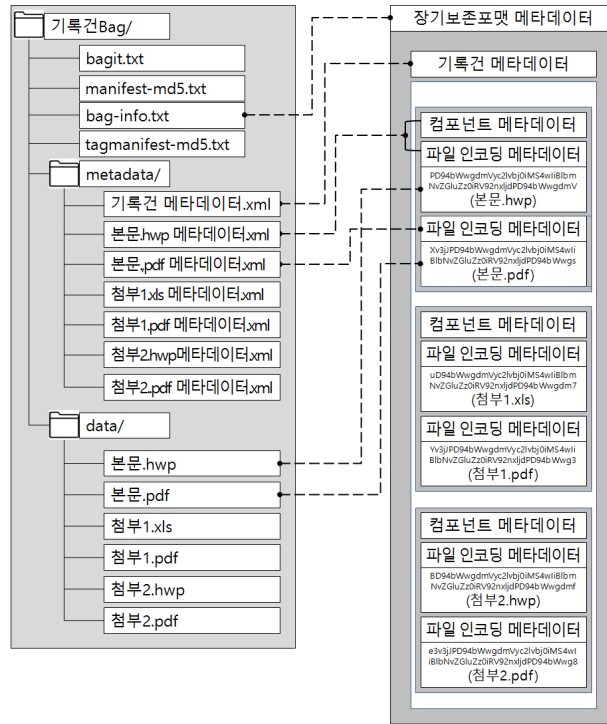
장기보존포맷인 BagIt의 내용물인 전자 파일이 반드시 ODF 포맷일 필요는 없다. BagIt 패키지의 내용물 포맷은 자유로우며 현재와 같이 HWP와 PDF/A-1 파일일지라도 NEO 대신 BagIt 포맷으로 묶을 수 있다. 문서보존포맷은 현재와 같이 PDF/A-1로 변환하고 패키지만 BagIt 규격으로 생성할 경우 NEO와 비교

하면 〈그림 16〉과 같다. 현 NEO 포맷에서 기록건 메타데이터는 그대로 별도의 기록건메타데이터.xml 파일로 만든 뒤 BagIt의 추가 태그 파일로 정의한다. 본문과 첨부 파일에 대한 메타데이터의 경우 현행 NEO 포맷에서는 컴포넌트 제목, 컴포넌트유형, 컴포넌트 버전, 컴포넌트 보존이력, 컴포넌트 관계정보 등의 메타데이터를 컴포넌트 메타데이터라고 불리는 〈ComponentMetadata〉 컨테이너에 기술한다. 각 컴포넌트 객체는 원본 파일 이외에도 문서보존포맷으로 변환되는 파일이 있기 때문에 1개 이상의 객체를 가지므로, 모든 객체에 공통되는 사항은 컴포넌트 메타데이터로 관리하되 각 파일별로 관리되어야 하는 정보는 인코딩 메타데이터라 불리는 〈Encoding〉 엘리먼트의 하위 엘리먼트인 〈EncodingMetadata〉에 기술한다 (NAK/TS 3). BagIt 규격으로 설계한 장기보존 포맷의 예시에서는 각 컴포넌트 메타데이터를 그대로 추출하여 각각 1개 파일로, 각 인코딩 메타데이터도 그대로 추출하여 각각 1개 파일로 분리하여 생성하였고 meta 디렉토리에 모든 메타데이터 파일을 넣었다.

〈그림 16〉으로 표현되는 장기보존포맷은 패키지 구조는 BagIt이지만 패키지 내용은 현행과 동일하게 전자 파일을 문서보존포맷으로 변환하여 원본과 PDF/A-1 파일을 모두 넣는 경우로 변경사항을 최소화하면서 BagIt 규격을 적용하였다. 이 예시를 통해 어떤 패키지이든 쉽게 적용할 수 있는 BagIt 규격의 유연성과 확장성을 확인할 수 있다. 그러나 메타데이터가 개별 파일로 분리되면서 메타데이터 파일의 개수가 증가하고 전자 파일과의 연관성이 드러나지 않아 최적의 구조로 보기는 어렵다. BagIt은

<표 13> BagIt 규격 장기보존포맷 구성요소 예시

구성 요소		예 시
bagit.txt	BagIt 버전과 문자 인코딩 명시	BagIt-version: 0.96 Tag-File-Character-Encoding: UTF-8
manifest-md5.txt	data/에 위치한 전자 파일 목록	203td21dae0cf31rda4df6d9ed81b017 data/본문파일 78adx1d50cef31d3acdffeed81b01a1 data/첨부파일
bag-info.txt	장기보존포맷의 생성일, 생성자, 유형, 크기와 같은 속성 정보	Source-Organization: 국가기록원 Organization-Address: 35208 대전광역시 서구 청사로 189, 2동 Contact-Name: 김기록 Contact-Email: info@archives.go.kr Contact-Phone: 012-481-6300 External-Description: 장기보존포맷 Bagging-Date: 2016-11-30 External-Identifier: ark:/13030/fk4jm2bcp Bag-Size: 10MB Payload-Oxum: 21836794142,831
tagmanifest-md5.txt	기록건과 전자 파일에 대한 메타데이터 파일 목록	b2dd8bbda34a1d363d3f1e7c9853b696 metadata/기록건 메타데이터.xml 496412702417e7a933d3318006e11e3e metadata/추가 메타데이터
metadata/	기록건과 전자 파일에 대한 메타데이터 파일 위치	메타데이터 파일 위치
metadata/기록건 메타데이터.xml	기록건 메타데이터를 기술한 파일	<p>NAK/TS 3:2013(v2.0) 전자기록물 장기보존포맷 기술규격(v2.0)에서 정의한 기록건 메타데이터</p> <p>출처: NAK/TS 3:2013(v2.0) 전자기록물 장기보존포맷 기술규격(v2.0)</p>
data/	전자 파일 위치	전자 파일의 위치
data/본문 파일 등	전자 파일	기록건의 모든 전자 파일들



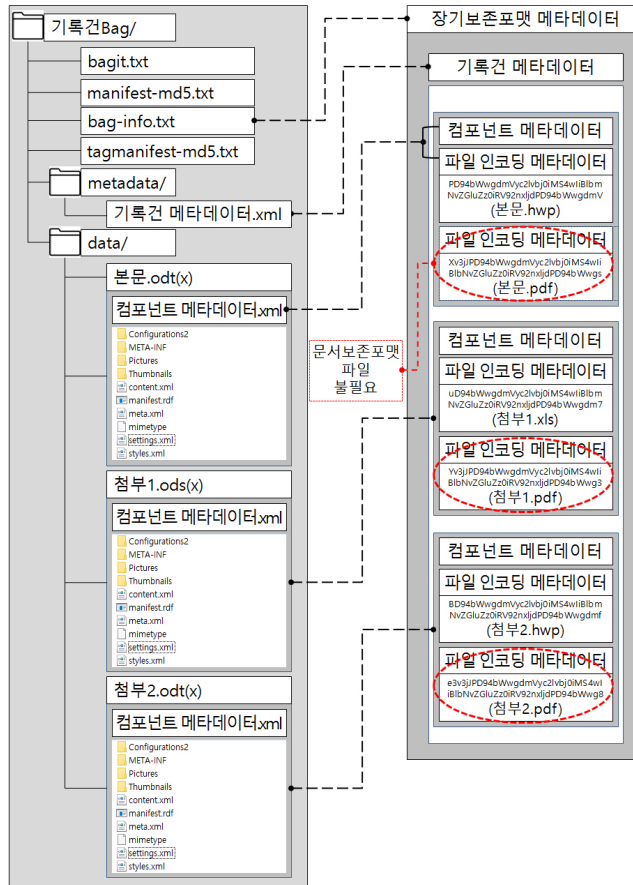
〈그림 16〉 BagIt 규격 장기보존포맷 패키지와 NEO 비교
(원본과 문서보존포맷 파일 포함)

ODF를 문서보존포맷으로 보존할 때 자기충족성을 만족시키는 컴포넌트를 단순한 구조로 패키지화하는 BagIt의 장점을 잘 살릴 수 있다.

앞서 본 BagIt 장기보존포맷 구조처럼 기록건 메타데이터는 그대로 기록건메타데이터.xml 파일로 만든 뒤 BagIt의 추가 태그 파일로 정의한다. 전자 파일에 대한 컴포넌트 메타데이터와 인코딩 메타데이터는 별도 파일로 BagIt 요소로 넣지 않고 ODF의 요소로 넣는다. 앞 장에서 살펴본 ODF 기반 문서보존포맷 중 확장 ODF 포맷의 경우라면 사용자 정의 메타데이터 파일에 컴포넌트 메타데이터와 인코딩 메타데이터를 함께 기술하여 확장 ODF 패키지 안에 넣고 BagIt은 이 ODF 패키지들을 다시 패키징한다.

ODF 포맷 파일을 내용물로 갖는 BagIt과 NEO 포맷의 예시를 〈그림 17〉에서 비교하였다.

지금까지 제시한 BagIt 기반 장기보존포맷은 Base64 인코딩이 필요가 없고 구조의 깊이가 2레벨에 지나지 않기 때문에 이해하기 쉬워 생성이나 검증이 간단하여 시간과 자원 소비를 줄일 수 있다. 이러한 기술적인 장점만이 아니라 오픈 포맷이기 때문에 규격 적용의 비용이 들지 않으며, 활용을 위한 오픈 소스 어플리케이션을 쉽게 구할 수 있으므로 포맷 생성이나 활용 시스템 구축이 쉬울 것이다. 미국 의회도서관을 비롯하여 디지털 아카이브 구축에 BagIt을 이용했던 기관들은 자체 프로젝트에서 개발한 어플리케이션이나 라이브러리 소스를 공개하고



〈그림 17〉 ODF 기반 BagIt 장기보존포맷과 NEO 비교

있으며 이를 <표 14>에 요약하였다.

BagIt 패키지는 패키지이지만 규격 상 반드시 1개의 파일로 생성될 필요는 없다. 그러나 대량의 기록을 이관하고 활용함에 있어서 1개 파일로 만드는 편이 더 효율적이고 안정적이라고 볼 수 있다. 그렇다면 어떤 알고리즘으로 BagIt 패키지를 1개 파일로 만들지 검토할 필요가 있다. 다수의 디렉토리나 파일을 포함한 구조 자체를 1개 패키지 파일로 만드는 방법은 TAR, ZIP 등 제약이 없다. 패키지 파일 생성 방안으로는 전 세계적으로 널리 사용되고 일반

사용자가 쉽게 사용 가능한 ZIP 포맷을 검토하고자 한다.

ZIP 파일은 PKWARE® Inc.에서 개발한 뒤 20여년 이상 널리 사용되어 왔지만 국제 표준으로 제정되지는 않았었다. 보편적으로 사용됨에도 불구하고 표준이 아닌 결과 오히려 규격 활용 시 저작권 문제, 규격의 신뢰성과 존속성에 대한 의심 등이 ZIP 사용의 장애가 되어오다가 2015년 ISO/IEC 21320으로 공식적인 국제 표준 규격으로 제정되었다. ISO/IEC 21320은 PKWARE® Inc가 개발한 Zip 파일 포맷 규격

〈표 14〉 BagIt 활용을 위한 오픈 소스

어플리케이션	기능	개발 언어	개발 주체
bagger	BagIt 규격에 따라 패키지 생성	Java	미국 의회도서관
bagit-java	BagIt 패키지 생성, 처리와 검증을 위한 라이브러리	Java	미국 의회도서관
chronam	미국 의회도서관의 Chronicling America (http://chroniclingamerica.loc.gov) 구축에 사용한 Django 어플리케이션	Python	미국 의회도서관
bagit-python	BagIt 패키지 활용을 위한 Python 라이브러리와 유틸리티	Python	미국 의회도서관
bagger-js	BagIt 규격의 순수 JavaScript 어플리케이션	JavaScript	미국 의회도서관
MarcMods3.5.xsl	MARC(Achine-Readable Cataloging) 21과 MODS(Metadata Object Description Schema) 사이 매핑용 XSLT	XSLT	미국 의회도서관
django-tabular-export	행과 행제목을 빠르고 안전하게 XLSX나 CSV로 내보내는 기능을 제공하는 모듈	Python	미국 의회도서관
BagIt for ruby	BagIt 패키지 생성을 위한 Ruby 라이브러리	Ruby	Florida Center for Library Automation
BagIt PHP	BagIt 패키지 생성 및 검증을 PHP 라이브러리	PHP	University for Virginia
archive-utils	BagIt 패키지 처리를 위한 각종 Ruby 유틸리티	Ruby	University of Stanford

(버전 6.3.3)을 인용하고 있다. 다만 ISO/IEC 21320에서는 ZIP 규격 중 아래 사항이 허용되지 않는다.

- 압축하지 않거나(0) deflate 매커니즘을 이용한 무손실 압축이어야 함
- 암호화 금지
- 전자서명 금지
- 추가 데이터
- ZIP 파일은 여러 개로 나뉠 수 없음

ZIP 방식은 압축 알고리즘으로 알려져 있어 비트 손실을 가져올지 모른다는 우려에서 국가 기록원 보존포맷에서 제외되었으나 국제 표준으로 규격화 되면서 압축하지 않음(0)이나 deflate 압축만 허용하고 있다. 따라서 손실 문제에 대한 우려는 없을 것으로 보인다.

ZIP 형식의 패키지 생성이나 활용을 위해 별

도의 전용 뷰어를 반드시 이용할 필요는 없다. 생성 및 해제 도구는 오픈 소스부터 상용 제품까지 구하기 쉽다. 따라서 전용 뷰어가 따로 없어도 압축 해제가 가능하므로 ZIP방식의 패키징이 보편성, 효율성, 편의성 면에서 유리하다고 본다.

5. 결론

전자기록의 보존포맷에 대한 규격이 마련된 지 10년이 되었다. 공공기록물법 시행령 제53조(영구기록물관리기관의 기록물 평가 및 폐기) 2항에 따르면 준영구기록물은 보존기간 기산일로부터 70년 뒤에 재평가를 거쳐 보존 여부를 판단해야 한다. 이에 따라 2005년에 생산되고 보존기간이 준영구 이상으로 분류된 기록물은 2076년에 첫 번째 평가를 받게 된다. 2076년이

도래하였을 때 해당 기록물이 과연 원활하게 내용을 확인할 수 있을지, 확인을 위해 드는 시간이 얼마나 걸릴지 의문이 든다. 2장에서 살펴본 다양한 장기보존 전략은 앞서 제기한 기록에 대한 접근과 내용 확인과 같은 기본적인 것을 해결하는 데에서 시작되었다고 볼 수 있다. 구조적 변경이나 바이너리의 손실 없이 원래의 형태를 그대로 보존하는 것이 기록의 내용, 구조, 맥락을 유지하기 위한 최선의 방안이지만 현재와 같이 독자포맷으로 생산된 기록을 그대로 패키징 하는 것은 장기적 측면에서 접근성, 이용 가능성을 보장하기는 어렵다. 역설적으로 오랫동안 보존하기 위한 방법이 오랫동안 보존하다 보니 보기 어려워질 수 있다는 아이러니한 상황이 벌어질 수 있게 되는 것이다. 이러한 우려 속에서 공기록 생산 표준 포맷으로 ODF를 적용하는 것은 참 반가운 일이다. ODF는 2000년대 초반에 개발된 포맷이지만 최근 들어 각광받기 시작하였다. 독점 라이선스와 거대 자본으로 프로그램 시장을 점령한 대형 벤더에 대한 폐해가 날날이 드러나게 되면서 더 이상 그들에게 휘둘리지 않고 독자적으로, 자유롭게 사용할 수 있는 오픈포맷이 주목받게 된 것이다. 독점 라이선스가 주는 편의성은 분명히 있다. 그러나 그 편의성에 젖어있는 동안 생겨난 시장질서의 저해, 비용부담의 증가, 지속가능성에 대한 막연한 불안은 시간이 갈수록 가시화되고 있다. 이를 탈피하고자 유럽연합을 중심으로 오픈포맷을 공공영역에 도입하고 있으며, 우리나라 또한 이를 준비 중이다.

본 연구에서는 G-클라우드 기반 온-나라 업무관리시스템에 XML을 기반으로 한 ODF를 채택함으로써 ODF가 생산포맷으로 정착됨에

따라 보존포맷으로의 역할이 가능함을 확인하고 이 포맷에 대한 효율적이고 효과적인 장기보존 전략에 대해 살펴보았다. 검토 결과 ODF는 문서보존포맷으로의 변환 단계 없이 기록관리 메타데이터 요소를 추가하는 것만으로도 장기보존을 위한 요건을 충족하고 있다고 보았다. 이에 따라 현행 문서보존포맷 외에도 ODF를 더 추가하는 방식으로 확장해 갈 수 있을 것임을 제안하였다. 또한 현행 NEO 포맷을 재검토하고 보완하기 위해 국제 표준 ISO21320으로 제정된 ZIP 알고리즘을 사용한 Bagit 구조를 제안하였다. 본 연구에서 모두 세밀하게 설계하지는 못하였지만 이 연구를 바탕으로 ODF로 생산한 기록뿐만 아니라 행정정보데이터세트, 웹 기록, SNS 기록 등 여러 기록유형을 포괄할 수 있는 상세한 보존포맷의 연구와 이를 기반으로 하는 테스트베드가 필요할 것으로 보인다.

유구한 기록문화 유산을 지닌 우리나라의 가장 대표적인 기록문화유산은 조선왕조실록이다. 수차례 국난을 거쳤어도 지금까지 현존할 수 있었던 까닭은 병충해에 강하고 보존에 유리한 한지를 사용하여 실록을 생산하고, 소설에 대비한 4대 사고(史庫)의 분산보존 체계, 보존하기에 적합한 온도와 습도를 유지할 수 있는 환경 등 보존 환경을 잘 구축함에 있다. 이처럼 기록 내용을 보존하기 위해 보존 매체의 특성을 잘 살리고 환경을 갖추는 것이 중요함을 알 수 있다. 현대에 들어 전자적으로 문서를 생산하고 그 생산량이 기하급수적으로 늘어가고 있는 지금, 공공영역에서 영구·준영구로 책정된 기록은 전체 생산 기록의 극히 일부에 지나지 않는다. 하지만 그들이 갖고 있는 가치는 우리나라의 역사성을 대표할만한 기록이 될 것이며 이들을 꾸준히

안전하게 보존하는 것이 우리 아키비스트의 역할로 볼 수 있다. 앞으로 100년, 200년 앞을 내다 보자면 장기적으로 부족한 면을 채우고 보완해나가야 함이 마땅하다고 여겨진다. 이처럼 문서 생산 환경과 컴퓨팅 환경이 급속도로 변화하는 시점에서 기록관리계와 관련 업계에서 이에 조속히 대응할 수 있는 체제와 안정적으로 오랫동안 쓸 수 있는 보존 전략을 세우는 것이 필요하다.

최근 전자기록의 안정적인 보존포맷 변환을 위해 문서보존포맷과 장기보존포맷의 변환 시점을 인수 종료 후 1년 이내에 실시하도록 「공공기록물 관리에 관한 법률」 시행령이 개정되었다. 2004년 생산한 문서들을 10년이 지난 2015년

국가기록원 이관 시점에 임박하여 포맷 변환을 진행하면서 여러 문제점들이 나타났고, 이를 순차적으로 해결해가며 포맷 변환 및 이관을 실시하게 되면서 이관 업무에 많은 시간이 소요되었다. 법령이 개정됨에 따라 당장 기록관에서는 10년 전 생산 기록과 현재 생산 기록을 동시에 변환해야 하므로 기록연구사의 업무가 더욱 가중될 것으로 예상된다. 본 연구를 시작으로 기록의 보존이라는 목적 아래 뒤에서 감내하고 있는 실무자들의 고충이 조금이나마 덜어질 수 있도록 중앙기록물관리기관과 영구기록물관리기관에서 실무차원의 대책을 검토해줄 것을 기대해본다.

참 고 문 헌

- 곽정 (2016). 전자기록의 진본성 확보를 위한 공공기록관리 개선방안 연구. 박사학위논문. 한국외국어대학교 대학원, 정보기록관리학과.
- 권도윤, 김희섭, 오삼균 (2009). 전자기록물 장기보존을 위한 마이그레이션 전략에 관한 연구. 한국기록관리학회지, 9(2), 19-40.
- 성한혁 (2007). 전자기록의 장기적 보존 및 활용을 위한 유형별 문서보존포맷에 관한 연구. 석사학위논문. 한국외국어대학교 대학원, 정보기록관리학과.
- 안대진, 임진희 (2016). 디지털 아카이브 시스템 구축을 위한 공개 소프트웨어 활용방안 연구. 정보관리학회지, 33(3), 345-370.
- 이소연 (2013). 국내 디지털 보존 연구의 동향 분석. 한국기록관리학회지, 13(2), 247-283.
- 임진희 (2006). 전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조. 기록학연구, 13, 41-90.
- 임진희 (2013). 전자기록관리론. 서울: 선인.
- 임진희, 최주호, 이재영 (2014). 전자기록 에뮬레이션 서비스 개발 사례 연구. 한국기록관리학회지, 14(3), 55-82.
- 정미리, 오세라, 임진희 (2016). 공문서 컴포넌트 오픈포맷 채택이 기록관리에 미치는 영향 분석. 한국기록관리학회지, 16(2), 29-55.

현문수 (2013). 표준 기록관리시스템의 전자기록 보존 기능 평가 연구 - 문서보존포맷변환 기능을 중심으로 -. 한국기록관리학회지, 13(2), 115-147.

[법령]

공공기록물 관리에 관한 법률

공공기록물 관리에 관한 법률 시행령

[관련표준]

국가기록원 기록관리업무 표준 NAK/TS 2:2008(v1.0) 「전자기록물 문서보존포맷 기술규격」〈행정안전부 고시 제2008-43호, 2008.11.4.〉

국가기록원 기록관리업무 표준 NAK/TS 3:2013(v2.0) 「전자기록물 장기보존포맷 기술규격」〈안전행정부 고시 제2013-53호, 2013.12.30.〉

ISO/IEC 21320-1:2015 Information technology - Document Container File - Part 1:Core.

ISO/IEC 26300-1:2015 Information technology - Open Document Format for Office Applications (OpenDocument) v1.2-Part 1:OpenDocument Schema.

Internet Engineering Task Force, The BagIt File Packaging Format (V0.97).

OASIS, Open Document Format for Office Applications (OpenDocument) Version 1.2.

Public Record Office Victoria (2003). Introduction to the Victorian Electronic Records Strategy (VERS) PROS 99/007 (Version 2).

[기사]

이승역 (2015). 전자기록 관리정책 전환을 위한 재검토 - 보존포맷·보존매체·관리체계를 중심으로 -. 기록인, 2015 Autumn, Vol32, 22-29.

[보고서]

강남구 (2016). 전자기록물 문서보존포맷 변환 결과(2016.04.22.).

국가기록원 (2004). 전자기록물 영구보존 기반기술 용역 종합보고서.

국가기록원 (2006a). 전자기록 영구보존기술 적용을 위한 테스트베드 구축 -최종보고회- 발표자료.

국가기록원 (2006b). 기록관리시스템혁신 ISP 사업 보고서.

국가기록원 (2016). 2015 국가기록백서.

서울특별시 (2016). 서울기록원 정보화전략계획(ISP) 수립 용역 완료보고서.

InterPARES 3 Project (2016). General Study 20 - Examining the Criteria for Open Standard File Formats: General Study Report.

[웹페이지]

나라장터. 검색일자: 2016. 9. 10. <http://www.g2b.go.kr/>

한국정보통신기술협회 정보통신시험인증연구소. 검색일자: 2016. 10. 1. <http://test.tta.or.kr>

[기타자료]

경기도교육정보기록원 (2016). 기록관리시스템에서 운영중인 보존포맷변환솔루션 증설 관련 자료 요청 (정보공개청구 3732098).

국가기록원 (2015a). 2016년도 기록물 관리지침.

국가기록원 (2015b). 2015년 전자기록물 육안검수 용역사업 제안요청서.

행정자치부 (2008). 기록관리 표준화과제 초안검토 결과보고(2008.07.09. 정보공개청구 3727072).

행정자치부 (2016a). 2016년도 전자정부지원사업 제안요청서.

행정자치부 (2016b). 공공기관 파일포맷 보유현황 조사 분석결과보고서(2016.10.12. 정보공개청구 3727076).

• 국문 참고자료의 영어 표기

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

An, Dae-Jin & Yim, Jin-Hee (2016). A Study on using open source software for building a digital archive system. *Journal of the Korean Society for Information Management*, 33(3), 345-370.

Hyun, Moonsoo (2013). Evaluation of the preservation functions in standard records management system for public agencies. *Journal of Korean Society of Archives and Records Management*, 13(2), 115-147.

Jung, Mi Ri, Oh, Seh-La, & Yim, Jin Hee (2016). Effects of adopting open document format in public record management. *Journal of Korean Society of Archives and Records Management*, 16(2), 29-55.

Kwag, Jeong (2016). A Study on improvement of public records management for the authenticity of electronic records. doctoral dissertation. Hankuk University of Foreign Studies.

Kwon, Do-Yun, Kim, Heesop, & Oh, Sam-Gyun (2009). A Study on migration strategy for long-term preservation of electronic records. *Journal of Korean Society of Archives and Records Management*, 9(2), 19-40.

Lee, So-Yeon (2013). Trends analysis of digital preservation research in Korea. *Journal of Korean Society of Archives and Records Management*, 13(2), 247-283.

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.

Yim, Jin-Hee (2006). The composition and structure of archival information package (AIP) for a long-term preservation of electronic records. *The Korean Journal of Archival Studies*, 13, 41-90.

Yim, Jin-Hee, Choi, Joo-Ho, & Lee, Jae-Young (2014). A case study for the emulation service of electronic records. *Journal of Korean Society of Archives and Records Management*, 14(3), 55-82.

[Standards]

National Archives of Korea. NAK/TS 2:2008(v1.0). Digital Document File Format for Long-term Preservation.

National Archives of Korea. NAK/TS 3:2013(v2.0). Technical Specification for Long-Term Preservation Format.

[Reports]

National Archives of Korea (2004). A study of foundation technique for Long-term Preservation of Electronic Records.

National Archives of Korea (2006a). Test bed for permanent preservation of electronic records technology.

National Archives of Korea (2006b). Records Management System Innovation ISP Report.

National Archives of Korea (2016). National Archives of Korea White Paper 2015.

Seoul Metropolitan Government (2016). Seoul Metropolitan Archives Information Strategy Planning (ISP) Report.