

전자기록 에뮬레이션 서비스 개발 사례 연구*

A Case Study for the Emulation Service of Electronic Records

임진희 (Jin-Hee Yim)**

최주호 (Joo-Ho Choi)***

이재영 (Jae-Young Lee)****

목차

- | | |
|-------------------------------|------------------------------|
| 1. 머리말 | 4. 에뮬레이션 서비스 모형 연구 |
| 2. 에뮬레이션 프로젝트 사례 분석 | 4.1 에뮬레이션 서비스 개발 방향 |
| 2.1 에뮬레이션 프로젝트의 비교 | 4.2 운영체제 에뮬레이션 프로토타입 개발 결과 |
| 2.2 해외 에뮬레이션 서비스의 비교 | 4.3 애플리케이션 에뮬레이션 프로토타입 개발 결과 |
| 2.3 해외 사례의 시사점 | 5. 맺음말 |
| 3. 공공기록 에뮬레이션 대상 분석 | |
| 3.1 공공기관 디지털 컴포넌트의 포맷 현황 | |
| 3.2 문서보존포맷 변환 시 소실되는 기능성 조사결과 | |

<초록>

우리나라 공공기록의 디지털 컴포넌트를 장기보존하면서 다양한 기능을 충분히 재현해내기 위해서는 에뮬레이션 전략이 필요하다. 이를 위해 이 연구에서는 먼저 대표적인 해외의 에뮬레이션 프로젝트인 CAMiLEON, KB, Planets, KEEP 등에 대한 사례를 통해 디바이스 에뮬레이션, 운영체제 에뮬레이션, 애플리케이션 에뮬레이션 등의 전형적인 에뮬레이션 방안을 비교분석하였다. 다음으로 우리나라 공공기관의 디지털 컴포넌트 포맷 현황을 살펴본 결과 아래아한글문서와 엑셀문서가 전체 디지털 컴포넌트의 90프로에 가까운 비율을 보인다는 점과 파일의 확장자와 실제 내부 포맷이 상이하게 관리되고 있기도 하다는 점을 발견하였다. 아래아한글문서와 엑셀문서 등 주요 포맷들이 문서보존포맷으로 변환되어 보존될 경우 상실하게 되는 기능성과 렌더링 특성을 조사하였으며, 운영체제 에뮬레이터와 애플리케이션 에뮬레이터 프로토타입을 개발하여 이들 문서의 기능성을 손실없이 재현해 보여줄 수 있음을 확인하여 에뮬레이션 전략의 가능성을 보여주고 있다.

주제어: 에뮬레이션 전략, 전자기록 장기보존, 운영체제 에뮬레이션, 애플리케이션 에뮬레이션, 프로토타입 에뮬레이터

<ABSTRACT>

The objective of this research is to show the possibilities of emulation strategy in the long-term preservation for the digital components of Korean public agencies. First, this study analyzed four major emulation projects that were conducted in the world, such as CAMiLEON, KB, Planets and KEEP. Second, it compared the three different emulation approaches, such as device emulation, OS emulation and application emulation. Third, it searched all the digital components that one public agency produced during the last six years. Through the search, this study found out that the HWP and XLS format files comprised around 90% of the files. Fourth, it examined the functionalities of the formats, which may have been lost when the format was converted to PDF/A. Last, this study introduced the two prototype emulators for the National Archives of Korea. One is a type of OS emulator, while the other is a type of application emulator.

Keywords: Emulation strategy, Long-term preservation, OS emulation, Application emulation, Prototype emulator

* 본 연구는 "안전행정부 국가기록원의 기록물 보존기술 연구 개발사업"의 일환인 "2013년도 행정기관 전자 기록물 재현기술 연구 및 프로토타입 개발 연구사업"의 재정지원을 받아 수행되었음.

** 명지대학교 기록정보과학전문대학원 조교수(yimjhkr@mju.ac.kr) (제1저자)

*** (주)이산티에스(iq2chun@esants.co.kr) (공동저자)

**** (주)이산티에스(jaeyoung.2ee@esants.co.kr) (공동저자)

■ 접수일: 2014년 7월 20일 ■ 최종심사일: 2014년 7월 28일 ■ 게재확정일: 2014년 8월 22일

■ 한국기록관리학회지 14(3), 55-82, 2014. <<http://dx.doi.org/10.14404/JKSARM.2014.14.3.055>>

1. 머리말

전자기록을 장기보존한다는 것은 먼 미래 시점에도 해당 기록을 원래의 모습대로 볼 수 있도록 하는 조치를 포함한다. 이를 가능하게 하는 방법으로 마이그레이션, 에물레이션, 인캡슐레이션 등 여러 보존 전략이 논의되고 있다. 이 중에서도 에물레이션은 전자기록이 지닌 본래의 기능성 및 원래의 비트스트림을 유지하며 원래의 모습을 그대로 유지할 수 있는 최적의 장기보존 전략이라 할 수 있다(Digital Preservation Testbed, 2009, p. 25). 에물레이션은 소프트웨어의 원래 기능이 현재의 컴퓨터상에서 재현될 수 있도록 하기 위해 디지털 정보가 생산·활용된 원래의 운영 환경을 재생산하는 것을 의미한다(김명훈, 오명진, 이재홍, 임진희, 2013).

현재 우리나라 공공기록물의 장기보존 전략은 아래아한글문서로 작성된 결재문서를 문서보존 포맷인 PDF/A로 변환하는 마이그레이션 전략과 기록물의 메타데이터를 원문 비트스트림들과 함께 장기보존포맷인 NEO(NAK's Encapsulated Object) 규격의 XML문서로 패키징하는 인캡슐레이션 전략을 채택하고 있다. 이는 인쇄 출력하여 보는 용도의 정적인 결재문서류에는 타당한 장기보존 전략이라 볼 수 있지만 PDF/A로 변환했을 때 기능이 사라지고 마는 요소를 포함한 전자문서류에는 적합하지 않다. 현재까지의 공공기관 결재 관행으로 미루어 볼 때, 아래아한글문서 안에 영상 클립을 삽입한다거나 메모를 붙여 넣거나 하진 않으므로 그러한 기능을 미래에 재현할 필요가 없으며 따라서 PDF/A로 변환하는 과정에서 손실하게 되는 요소는 없다. 하지만, 결재문서 외의 아래아한글문서에는

다양한 OLE(Object Linking and Embedding) 객체가 삽입되거나 책갈피나 스크립트가 포함되어 있을 수 있으며, 이들은 PDF/A로 변환하는 과정에서 모두 사라지게 되어 재현이 불가능하게 된다. PDF/A로 변환하는 것으로 충분한 보존조치가 될 수 있는 문서와 그렇지 않은 문서가 존재한다는 것이다.

다른 예로 엑셀문서를 살펴볼 수 있다. 우리나라 공공기관이 생산하는 디지털 컴포넌트의 유형에서 가장 많은 비율을 차지하는 것은 아래아한글문서(hwp)이며 두 번째로 많은 비율을 차지하는 것이 엑셀문서(xls)로 조사된 바가 있다(김명훈 외, 2013, p. 306). 공공기관에서 결재문서 작성 시 본문은 아래아한글문서로 하지만 첨부자료로 엑셀문서를 첨부하는 경우가 많다. 엑셀시트를 인쇄출력하면 셀의 최종 값만 나타나며 셀에 적용된 함수나 셀 간에 설정된 참조 관계가 사라지고 보이지 않게 된다. 또한, 셀에 부가적인 설명을 위해 추가한 메모도 기본적인 인쇄출력으로는 확인할 수 없다. 결국 엑셀문서가 PDF/A로 변환되면 그 결과는 아래아한글문서의 표를 작성한 것과 동일한 표만 남게 되는데 이것이 우리가 보존해야 할 엑셀문서의 가치로 충분한 것인지가 관건인 것이다. 표만 남겨 보존하는 것으로는 충분하지 않으며 셀 간의 연관관계를 함께 보존하여야 하는 엑셀문서가 있다면 PDF/A로 변환하는 방법 외에 다른 접근이 필요하다.

이처럼 마이그레이션 접근으로는 재현이 불가능한 요소를 지닌 문서들에 대해서는 에물레이션 접근을 취해볼 수 있을 것이다. 물론, 이 경우 마이그레이션 전략과 에물레이션 전략은 상호배타적인 접근이 아니라 상호보완적인 접근으로 이용되는 것이다. 즉, 기록물의 디지털

컴포넌트별로 PDF/A로 변환해도 원래의 기록 내용을 충분히 재현할 수 있는 것과 그렇지 않은 것을 구분하여 전자는 마이그레이션 조치를 취하며 후자는 에뮬레이션이 가능하도록 조치를 취하도록 할 수 있다.

이 연구에서는 앞에서 살펴본 상황을 전제로 하여 국가기록원과 같은 영구기록물관리기관이 에뮬레이션이 요구되는 디지털 컴포넌트들을 재현해 주기 위해 어떤 에뮬레이터를 만들어 장기보존 및 재현기능을 제공할 수 있는지 서비스 모형을 제시하고자 한다. 또한, 특정 포맷의 디지털 컴포넌트에 대한 에뮬레이터 프로토타입을 개발한 사례를 소개함으로써 전자기록 장기보존 전략으로서 에뮬레이션의 가능성을 타진하고자 한다. 이를 위해 이 연구에서는 실제 공공기관에서 생산되고 있는 디지털 컴포넌트의 종류를 조사하고, 각 포맷별로 어떤 기

능성을 가질 수 있으며, 이 중에 PDF/A로 변환 시 상실될 수 있는 것을 조사하고, 이들을 재현하기 위해 한국형 에뮬레이터를 만드는 과정을 보여주고자 한다.

2. 에뮬레이션 프로젝트 사례 분석

2.1 에뮬레이션 프로젝트의 비교

해외에서 진행된 디지털객체에 대한 주요 에뮬레이션 프로젝트는 CAMiLEON, KB, Planets, KEEP 등 네 가지로 요약해 볼 수 있다. 김명훈 외(2013)는 각 프로젝트별로 개요 및 목적, 구체적인 에뮬레이션 기법을 살펴보고, 이 프로젝트들이 제시하는 시사점을 정리하고 있다. 네 개의 프로젝트를 다른 관점에서 정리하면 <표 1>과 같다.

<표 1> 해외 에뮬레이션 프로젝트 비교

항목	CAMiLEON	KB	Planets	KEEP
명칭	Creating Creative Archiving at Michigan and LEEDS	Koninklike Bibliotheek	Preservation and Long term Access through NETworked Services	Keeping emulation environments portable
년도	1999~2002	2005~2007	2006~2010	2009~2012
주관 기관	미시건 & 리즈대학	네덜란드국립도서관 & 네덜란드국립아카이브	British Library (대영도서관)	파리국립도서관
내용	에뮬레이션을 중심으로 한 디지털매체의 장기적 보존방법론 개발 연구 프로젝트	에뮬레이션 기반의 새로운 보존전략에 관한 연구 프로젝트	디지털 객체를 관리하기 위한 일련의 도구들을 개발하고 실제 서비스할 수 있는 견고한 프레임워크와 인프라를 제공하려는 지향점을 가짐	렌더링 기반의 에뮬레이션 접근 플랫폼(emulation access platform) 개발
금액			약 206억 원	약 58억 원
적용 사례	BBC Domesday 프로젝트: BBC와 Acom Computer Ltd, Philips, Logica가 공동으로 조사한 영국 국민들의 삶에 대한 기록	하드웨어 아키텍처의 구성 요소를 에뮬레이션한 2007년 Dioscuri란 명칭으로 개발하였는데, 이 에뮬레이터는 Planets 프로젝트에 의해 계속 연구를 수행	Open Planets Foundation을 통해서 그 결과물들이 유지 운영	Open Planets Foundation을 통해서 그 결과물들이 유지 운영

항목	CAMiLEON	KB	Planets	KEEP
URL	http://www2.si.umich.edu/CAMiLEON/index.html	http://www.kb.nl/en/expertise/e-depot-and-digital-preservation	http://www.Planets-project.eu/	http://www.keep-project.eu/
정책	<ul style="list-style-type: none"> • 소프트웨어 장기 지속 에뮬레이션 • 마이그레이션은 리퀘스트 (이용 시점에 객체를 마이그레이션하는 도구와 함께 원래의 바이트스트림을 유지하여 보다 정확하고 비용 효율적이 되도록 함) 	<ul style="list-style-type: none"> • 장기보존 전략으로서의 다양한 에뮬레이션 전략 • 에뮬레이션 기술의 지속성/유연성 	<ul style="list-style-type: none"> • 네트워크 서비스를 통한 장기보존 및 접근 방법 • 구형화된 디지털 자산의 에뮬레이션 툴, 서비스 평가 및 테스트 • 보존 계획을 위한 테스트 베드 제공 	<ul style="list-style-type: none"> • 디지털 컴포넌트 포맷, 플랫폼, 운영체제의 에뮬레이션 • 에뮬레이션 서비스 • 법적 이슈 및 저작권 관리
산출물	<ul style="list-style-type: none"> • 다양한 수준의 에뮬레이션 비교자료 • 도서관용 보존 툴 • 마이그레이션 및 에뮬레이션 전략 가이드라인 • 다양한 디지털 객체의 속성 정의 • 디지털 객체 보존 전략 	<ul style="list-style-type: none"> • 에뮬레이션 방법론 4종 • 모듈 에뮬레이터: Dioscuri • 에뮬레이터 명세 문서 (ESD) 제시 	<ul style="list-style-type: none"> • Planets 상호운용성 프레임워크 • 보존계획 툴: Plato (의사결정지원서비스) • 에뮬레이션 툴: GRATE 	<ul style="list-style-type: none"> • 이관 툴 • KEEP-EF (Emulation Framework) • TOTEM (Trustworthy Online Technical Metadata Database)이라는 이름으로 웹에서 기술 메타데이터를 담은 데이터베이스를 시범적으로 구축

* 출처: 김명훈 외 (2013) 재구성

2.2 해외 에뮬레이션 서비스의 비교

1999년부터 시작한 CAMiLEON 프로젝트부터 2012년까지 수행된 KEEP 프로젝트의 목적과 내용 산출물을 비교 검토한 결과 CAMiLEON과 KB 프로젝트는 방법론 정립 중심의 프로젝트인 반면 Planets과 KEEP 프로젝트를 통해서 다양한 산출물 및 시스템이 구축되었다. CAMiLEON 프로젝트는 장기적 보존 방법론에 대한 연구를 통해 마이그레이션 및 에뮬레이션 방법론을 제시하였고, KB 프로젝트는 에뮬레이션 기반의 새로운 보존 전략을 위해 에뮬레이터 Dioscuri를 개발하였다. Planet 프로젝트를 통해 본격적인 에뮬레이터 3종인 Dioscuri, QEMU, UVC가 개발되어 GRATE 서비스가 제공되었는데 3종의 에뮬레이터는 운영체제에 대한 에뮬레이터로 정의된다. 가장 최근에 수

행된 KEEP 프로젝트에서는 렌더링 기반의 에뮬레이션 프레임워크(KEEP-EF)에 Qemu(x86), Dioscuri (x86), WinUAE(Amiga), Vice(c64), JavaCPC (Amstrad / Schneider), BeebEm (BBCmicro), Thomson(Thomson T07) 등 7종의 에뮬레이터를 제공하는데 모두 운영체제 에뮬레이터임을 확인할 수 있었다.

해외 에뮬레이션 프로젝트를 통해 개발된 시스템과 서비스를 분석한 결과 Planets 프로젝트에서 개발한 Plato는 에뮬레이션 정책 관련 시스템이고, KEEP 프로젝트에서 제공하는 KEEP-TTF, TOTEM은 메타데이터 관리시스템이며 그 외에는 모두 에뮬레이션 프레임워크 및 운영체제 에뮬레이터임을 확인하였다. 네 개의 프로젝트에서 만들어진 에뮬레이션 기법을 정리해 보면 <표 2>와 같다.

해외 에뮬레이션 시스템과 서비스에서 보여

〈표 2〉 해외 에뮬레이션 서비스 비교

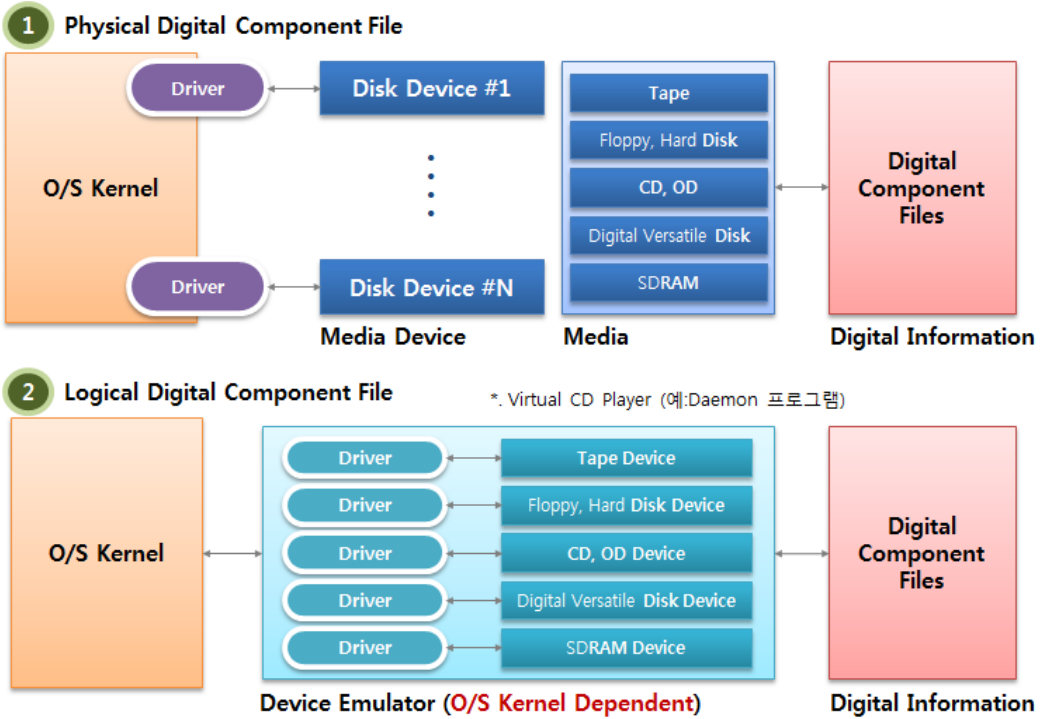
항목	CAMiLEON	KB	Planets	KEEP
정책	×	×	Plato	×
디바이스 에뮬레이션	×	Dioscuri	GRATE • Dioscuri • QEMU • UVC	KEEP-EF • Qemu(x86) • Dioscuri(x86) • WinUAE(Amiga) • Vice(c64) • JavaCPC(Amstrad/Schneider) • BeebEm(BBCmicro) • Thomson(Thomson T07)
운영체제 에뮬레이션	×			
애플리케이션 에뮬레이션	×	×	×	×
메타데이터 관리	×	×	DROID PRONUM	DROID PRONUM KEEP-TTF TOTEM

준 에뮬레이션 방식은 디바이스 에뮬레이션 방안, 운영체제 에뮬레이션 방안, 애플리케이션 에뮬레이션 방안 등 크게 세 가지로 구분된다. 먼저, 디바이스 에뮬레이션이란 테이프, 플로피 디스크 등 미디어 디바이스 및 드라이버에 대한 에뮬레이션 서비스를 제공하는 것이다. 미디어 플레이어에 종속된 디지털 컴포넌트를 조회하기 위해서는 디바이스에 대한 에뮬레이터를 제공하여 미디어에 저장된 디지털 컴포넌트를 조회하는 방식의 에뮬레이션을 취해야 한다. 〈그림 1〉에서 보는 바와 같이 미디어에 종속적인 형태로 저장된 디지털 컴포넌트 조회를 위해서는 Virtual CD Player와 같이 디바이스 에뮬레이터를 이용하여 디지털 객체에 접근하도록 하는 개념이다. 디바이스 에뮬레이션은 운영체제에 종속적이며 신규 운영체제에 대한 지속적인 업그레이드가 필요하므로 보통의 경우

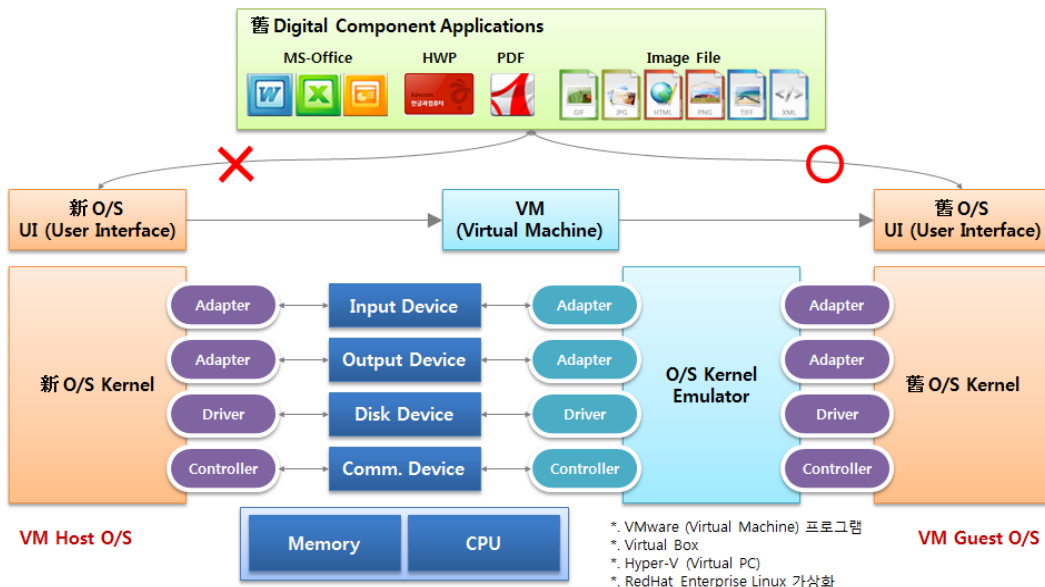
운영체제 에뮬레이션과 통합되어 서비스되어야 한다.

운영체제 에뮬레이션 방안은 신형 하드웨어에 설치된 신형 운영체제에 에뮬레이터(VM)¹⁾ 및 구형 운영체제를 설치하여 기록물이 생산될 당시의 환경을 가상으로 만들고 구형 애플리케이션을 구동하는 방식이다. 과거의 운영체제 자체를 에뮬레이션하므로 해당 운영체제에서 구동되던 모든 애플리케이션이 실행가능하다는 점이 특징이다. 만약 재현해야할 디지털 컴포넌트의 종류가 다양하고 당시의 애플리케이션이 있어야 해당 디지털 컴포넌트가 완전히 재현될 수 있다면 운영체제 에뮬레이션 방안이 효과적인 접근이라 할 수 있다. 하지만, 운영체제 에뮬레이터는 새로운 운영체제가 나올 때마다 그 운영체제에 맞추어 업그레이드되어야 하므로 유지해나가는데 상당한 노력이 동반된다(〈그림 2〉 참조).

1) 운영체제 에뮬레이터는 가상 머신(VM:Virtual Machine)으로 불리는데 현실에 존재하지 않지만 마치 존재하는 것과 같이 만든 가상의 기계 또는 장치를 의미. 현재 많이 사용하고 있는 VM은 VirtualBox와 VMware이다.



<그림 1> 디바이스 에뮬레이션 개념



<그림 2> 운영체제 에뮬레이션 개념

마지막으로 애플리케이션 에뮬레이션 방안은 구형 하드웨어 및 구형 운영체제 기반의 디지털 컴포넌트에 대한 전자기록을 조회할 수 있도록 신형 운영체제 기반에서 구동되는 뷰어 소프트웨어를 제공하는 방식이다. 만약, 특정 플랫폼에서 소수 애플리케이션으로만 생산된 디지털 컴포넌트를 재현할 필요가 있다면 해당 애플리케이션들만 에뮬레이터로 개발하는게 용이할 것이다. 특히, 디지털 컴포넌트를 생산할 때 필요했던 애플리케이션 기능 중 재현을 위해 필요한 기능이 일부에 국한된다면 더욱이 이 방안은 비용대비 타당성이 있다고 볼 수 있다. 하지만, 구형 운영체제 기반에서 저장된 디지털 컴포넌트 조회를 위해 새로운 운영체제하에서는 구형 운영체제에서 운영되던 뷰어가 작동이 되지 않을 수도 있으므로 이 경우 뷰어를 새로 개발해야 하는 부담이 따를 수 있다.

앞에서 살펴본 세 가지 에뮬레이션 방식을 디바이스&운영체제 에뮬레이션 방안과 애플리케이션 에뮬레이션 방식의 두 가지로 나누어 장단점을 비교 설명하면 <표 3>과 같다. 두 방식에 대한 장단점을 분석하였을 때 운영체제 에뮬레이터를 적용하면 디지털 컴포넌트가 생산될 당시의 환경을 그대로 사용하기 때문에 다양한 응용 소프트웨어에 대한 에뮬레이션 서비스가 가능하지만 중첩된 가상화 방식을 사용할 경우 하드웨어 성능이 현저히 저하되는데 반해 애플리케이션 에뮬레이터는 새로운 운영체제상에서 개별 애플리케이션에 대한 뷰어를 개발하기 때문에 개발 대상이 증가하지만 하드웨어 성능에 영향을 미치지 않는다는 점이 가장 큰 특징으로 볼 수 있다.

시간적으로 앞선 프로젝트는 후행 프로젝트에 영향을 미쳤을 것이다. 선행 프로젝트가 연구적인 성격이 강했다면 최근의 프로젝트들은 실

2.3 해외 사례의 시사점

시간적으로 앞선 프로젝트는 후행 프로젝트에 영향을 미쳤을 것이다. 선행 프로젝트가 연구적인 성격이 강했다면 최근의 프로젝트들은 실

<표 3> 에뮬레이션 방안 비교

항목	디바이스 & 운영체제 에뮬레이션	애플리케이션 에뮬레이션
방식	- 신규 운영체제기반에서 아카이브에 등록된 디지털 컴포넌트를 조회할 수 있도록 구형 운영체제를 구동할 수 있는 가상화(VM, Virtual Machine) 제공	- 신규 운영체제 기반에서 아카이브에 등록된 디지털 컴포넌트를 조회 할 수 있도록 디지털 컴포넌트에 대한 에뮬레이터 제공
장점	- 애플리케이션 Format에 대한 정보 관리 불필요 - 애플리케이션 종류 및 수에 관계없이 에뮬레이션 가능	- 애플리케이션 S/W에 대한 보존 불필요 - 상대적으로 폰트 관리가 용이하여 문서 형식(원본성) 유지가 가능 - 전자기록에 포함된 디지털 컴포넌트 조회 기능 및 구성이 상대적으로 간단함
문제점	- 운영체제에 종속된 애플리케이션 S/W(라이선스, 버전) 관리가 필요하며 폰트 관리가 복잡하여 원본 형식 유지가 어려움 - 중첩(Stacked)방식의 가상화 전략을 사용할 경우 전자기록에 포함된 디지털 컴포넌트 조회기능이 상대적으로 복잡하며 시간이 지날수록 속도 저하	- 신규 운영체제에 대하여 디지털 컴포넌트별 에뮬레이터를 개발하므로 개발 대상 증가 - 애플리케이션 포맷에 대한 정확한 정보와 관리 필요 - 애플리케이션 개발사의 적극적인 협조 필요

제 에뮬레이션 서비스를 제공하는 단계까지 수행이 이루어졌음을 알 수 있다. 그러나, KEEP이 7개의 에뮬레이터와 6개 플랫폼을 제공하는 것으로 알 수 있듯이 에뮬레이션 서비스는 과거 모든 디지털 객체에 대한 재현이 가능한 것은 아니며 재현을 하기 위한 환경도 제한적임을 알 수 있다. 재현 대상의 범위와 종류가 많아 질수록 TOTEM(Trustworthy Online Technical Metadata Database)과 같은 데이터베이스 기반을 구축하는 것이 중요해진다. 이 연구에서도 공공기록물 디지털 컴포넌트로 관리대상이 되는 주요 포맷에 대해 상세한 정보가 DFR(Digital Format Registry)로 구축되어 있다는 전제하에 에뮬레이터의 프로토타입을 개발하였다.

위의 해외 사례 중 이 연구에 직접적으로 연관되는 용어를 살펴보자면 CAMiLEON 프로젝트가 밝히고 있는 핵심목표 세 가지 중 첫 번째 목표에 등장하는 기능성(functionality)이라 할 수 있겠다. CAMiLEON은 '디지털 객체의 원래의 기능성(functionality)과 외형과 느낌(look and feel)의 장기적 유지를 위한 선택사항들을 탐구'하고자 한 프로젝트이다. 아래아한글문서나 엑셀문서를 PDF/A로 변환하는 전략은 인쇄출력 및 디스플레이 렌더링 시의 '외형과 느낌'을 장기적으로 유지한다는 측면에서는 타당성이 있다고 보여진다(National Archives of Australia, 2004). 그러나, '기능성'의 측면에서는 머리말에서 살펴본 사례처럼 문제가 있다고 보여진다. 해외에서 이미 1999년부터 2002년에 수행된 프로젝트에서 제시했던 '기능성'의 장기적 유지를 위한 대안을 우리나라에서는 이제야 검토하고 있는 것이다.

에뮬레이터를 구현할 때 이 연구에 직접적인

영향을 준 서비스 형태는 KB 프로젝트라 할 수 있다. KB에서는 모듈화된 에뮬레이터 Dioscuri를 개발하여 공개하였는데, 이 모듈 에뮬레이터가 동작하기 위해서는 범용가상머신이 필요하다. 모듈 에뮬레이터가 동작하기 위해 어떤 플랫폼별 구성요소가 필요한지를 '에뮬레이터 상세문서(emulator specification document)'를 통해 파악하고 구성요소 라이브러리에서 필요한 모듈을 검색하여 가상머신을 재구성하게 된다. 이러한 동작 방식은 이 세상에 존재했던 모든 하드웨어와 소프트웨어를 재현가능하게 하기 보다는 필요한 것들만 재현가능하도록 구성할 수 있다는 점에서 현실가능한 접근이라 보여진다.

해외 사례를 통해 확인할 수 있는 사실은 에뮬레이션을 위한 기반 구축에는 상당한 비용이 필요하다는 것이다. Planet의 경우 200억 원이 넘는 금액이 투자되었는데 이것이 모든 디지털 객체의 완벽한 에뮬레이션에 소요된 금액이 아니라는 점을 고려한다면 최소한 우리나라 공공기록물에 대한 에뮬레이션 전략의 경우에도 얼마나 많은 비용이 투자되어야 할지 조금은 가늠해 볼 수 있을 것이다. 따라서 이 연구에서는 공공기록물의 디지털 컴포넌트 포맷의 현황 정보와 재현 대상과 범위를 고려한 맞춤형 에뮬레이터 개발을 제안하고 한정적 환경과 대상에서 기능하는 현실가능한 서비스를 구상하고 있다.

3. 공공기록 에뮬레이션 대상 분석

3.1 공공기관 디지털 컴포넌트의 포맷 현황

우리나라 공공기록관리기관에서 에뮬레이션

전략을 고려하고자 한다면 그 필요성을 확인하고 방향을 정하기 위해 먼저 디지털 컴포넌트의 종류와 특성을 확인해보아야 한다. 본 연구에서는 중앙행정기관 중 하나를 선정하여 기록관리시스템에서 관리 중인 디지털 컴포넌트를 전수 조사한 바 있다. 연구의 조사 내용은 첫째, 파일명 중에 확장자 둘째, 파일의 헤더 정보 셋째, 포맷별 파일의 개수 등이었다. 파일의 헤더 정보까지 확인하기 위해서는 연구팀의 조사용 컴퓨터에 디지털 컴포넌트들을 다운로드 받은 후 조사를 위해 개발한 소프트웨어 프로그램을 구동시켜야 했다. 이 과정에서 조사대상 중앙행정기관이 다른 곳으로 변경되게 되었다. 애초에는 A기관을 조사하려 했으나 기록관리시스템에 아카이빙 스토리지 장비를 사용하고 있었고 FTP와 같은 방법으로는 접근이 불가하며 조사용 컴퓨터에 디지털 컴포넌트를 대량으로 복사해주는 API도 마련되어 있지 않아 조사가 불가능하다는 것이었다. 이에 아카이빙 스토리지를 사용하지 않고 일반 스토리지를 사용하고 있는 B기관으로 변경하게 되었다. A기관의 경우 이번 연구 시도 이전에 아카이빙 스토리지에 존재하는 디지털 컴포넌트의 목록을 작성한 적이 없었으며 따라서 파일의 개수나 전체 용량을 계산해 본 바도 없음을 알게 되었다. 만약 A기관의 기록관에서 디지털 컴포넌트의 현황을 알고자 원한다 해도 바로 알 수 없으며 현재와 같은 스토리지 구조에서는 별도의 조치를 취하지 않고는 디지털 컴포넌트의 전모를 알기 어렵게 되어 있다. 기록관리에서 의미 있는 관리행위의 최소 단위는 기록 건(item)이지만 실제적인 관리조치의 최소 단위는 디지털 컴포넌트들이다. 기록 건을 통해서 디지털 컴

포넌트를 접근할 수도 있어야 하지만 디지털 컴포넌트의 통계현황을 직접 조사하고 검토할 수 있어야 한다. 그런데, 이번 연구과정에서 기록관리시스템 내의 기능에는 디지털 컴포넌트 단위의 관리조치 행위가 전혀 반영되어 있지 않다는 점을 확인하게 되었다.

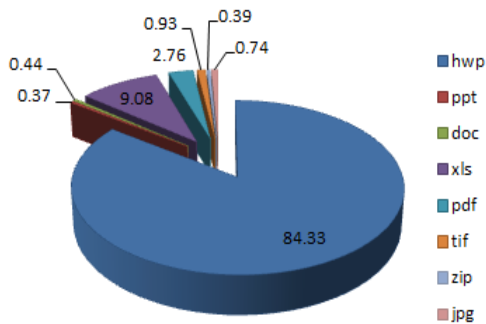
B기관 기록관리시스템 서버에는 2008년부터 2013년에 걸쳐 이관된 기록의 모든 디지털 컴포넌트가 이관된 연도별로 폴더로 나뉘어 저장되어 있었다. 2013년 9월 2일 약 8시간에 걸쳐 60만 여 건, 9월 23일 약 8시간에 걸쳐 200만 건의 디지털 컴포넌트를 조사하였다. 전체 260만 여건의 파일을 조사용 컴퓨터에 다운로드를 받았는데 시간을 절약하기 위해 압축도구의 하나인 TAR를 이용하여 폴더별로 파일들을 묶어서 전송을 하였다. 폴더의 이름이 이관연도 4자리, 월 2자리, 일 2자리, 시간 2자리로 총 10자리로 구성되어 있었는데 전체 파일을 균형적으로 10개로 나누어 전송하기 위해 이 중 끝자리가 같은 것끼리 하나의 TAR 파일로 만드는 방식을 적용했다. 결과적으로 end0.tar, end1.tar, end3.tar ... ent9.tar 등 총 10개의 TAR 파일을 만들게 되었으며 FTP를 이용하여 기록관리시스템 서버에 접속하여 이 파일들을 조사용 컴퓨터로 전송하였다. 전체 TAR 파일의 용량은 400G바이트 정도였고 서버로부터 조사용 컴퓨터로 다운로드 받은 후 압축된 TAR 파일을 풀어내야 했다. DROID6와 Extract Format 도구를 이용하여 디지털 컴포넌트 파일의 헤더 정보를 일일이 읽어내어 파일 포맷정보와 생산년도를 추출하였다. 디지털 컴포넌트를 전수조사한 결과 확장자에 대한 통계, 생산년도별 통계, 포맷종류에 대한 통계 등을 작성할 수 있었다.

물론 이 통계정보의 대상이 된 디지털 컴포넌트는 기록관리시스템에서 관리하고 있는 디지털 컴포넌트의 목록 정보를 이용하여 추출한 것이 아니라 파일 시스템에 있는 디지털 컴포넌트를 직접 조사를 한 것이기 때문에 결과적으로는 기록관리시스템의 기록 건 정보와 연결되지 않는 파일들도 포함되어 있을 수 있다.

조사 대상 기관의 기록관리시스템(RMS)에 등록된 기록물철/건에 첨부된 전체 디지털 컴포넌트 포맷의 비율은 <표 4>와 같으며 이를 <그림 3>의 그래프로 나타내었다.

<표 4> 디지털 컴포넌트 유형 비율

포맷	파일 개수	백분율
hwp	2,272,837	84.33%
ppt	10,030	0.37%
doc	11,849	0.44%
xls	244,782	9.08%
pdf	74,417	2.76%
tif	24,958	0.93%
jpg	20,045	0.74%
zip	10,395	0.39%
(8종) 합계	2,669,313	99.04%
기타	1,617	0.06%
합계	2,670,930	100.00%



<그림 3> 디지털 컴포넌트 유형 분포

앞의 <표 4>에서 보는 바와 같이 장기보존을 위해 공공기관의 기록관리시스템(RMS)에 등록된 디지털 컴포넌트 유형 중 아래아한글(hwp) 파일 포맷이 84.33%를 차지하여 거의 대다수 파일이 아래아한글로 작성됨을 알 수 있고, 다음으로 마이크로소프트사의 엑셀(xls) 파일 포맷이 9.08%를 차지하여 전체 디지털 컴포넌트 중 2가지 포맷이 차지하는 비중(93.41%)이 거의 대부분임을 알 수 있었다. 그러나, 예상과 달리 파워포인트(ppt) 파일 포맷으로 작성된 디지털 컴포넌트의 비중은 0.37%로 다른 포맷 유형보다 적은 반면 pdf 파일 포맷은 2.76%를 유지하여 비교적 많이 사용됨을 알 수 있었다.

또한 아래아한글(hwp) 파일 포맷, 마이크로소프트의 오피스(ppt, doc, xls) 파일 포맷, 보존문서 표준 포맷(pdf), 이미지 파일 포맷(tif, jpg), 압축 파일 포맷(zip) 인 8종의 파일 포맷이 차지하는 비율이 전체 파일 중에서 99.04%를 차지하여 거의 대다수임을 확인할 수 있었다. 기타로 분류한 파일 포맷은 <표 5>에서 보듯이 약 47종의 파일 포맷 유형이 검색되었는데 파일 포맷이 존재하는 경우도 있지만 포맷 확장자 정보가 잘못 입력된 경우도 발행한 것으로 확인하였다.

기타 포맷 유형에서 가장 많이 차지한 것은 bmp, gif, png, swf, mp3, wav, wma, wmv, psd 등의 이미지/오디오/비디오 파일이 존재하였고 html, xml 등 마크업랭귀지로 작성된 파일도 보존 중이었다. 또한 dwg, exe, bat, sql 등의 파일은 시스템 구동 또는 도면 이미지로서 기관에 따라 다양한 파일이 존재할 가능성이 있음을 시사하고 있다.

〈표 5〉 디지털 컴포넌트 기타 포맷 유형

확장자	포맷 유형	확장자	포맷 유형
7z	7Zip format	k3g	
ai	Acrobat- Portable Document Format	mht	
alz		mp3	MPEG 1/2 Audio Layer 3
bat	Batch file (executable)	nxl	Microsoft Excel 97 Workbook (xls)
bmp	Windows Bitmap	opt	
body		ozd	
csv	Comma Separated Values	png	Portable Network Graphics
dwg	AutoCAD Drawing	psd	Adobe Photoshop
egg		rqr	Extensible Markup Language
eml	Hypertext Markup Language	rst	
enc		rtf	Rich Text Format
erl		sql	Structured Query Language Data
exe	Windows Portable Executable	swf	Adobe Flash
flw		tmp	OLE2 Compound Document Format
gif	Graphics Interchange Format	txt	Plain Text File
gux		wav	Waveform Audio(PCMWAVEFORMAT)
hml	Extensible Markup Language	wma	Windows Media Audio
hox	Extensible Markup Language	wmv	Windows Media Video
hpt		xlk	Microsoft Excel 97 Workbook (xls)
htm	Hypertext Markup Language	xlsm	Microsoft Excel Macro-Enabled
html	Hypertext Markup Language	xlsm	Microsoft Excel Macro-Enabled
hwt	OLE2 Compound Document Format	xlt	Microsoft Excel 97 Workbook (xls)
hwx		xml	Extensible Markup Language
jp001			

2008년부터 2013년까지 조사 대상 기관의 연도별 파일 포맷(7종)의 비율 변화를 〈표 6〉에 나타내었는데 2008년도에는 tif 포맷의 비율이 21.34%를 차지하였으나 그 이후로는 거의 파일이 존재하지 않는 것으로 보아 특정 사업을 수행하면서 생성된 tif 파일임을 알 수 있으며, 연도별 조사에서도 아래아한글(hwp)이 가장 많은 비율을 차지하였고 마이크로소프트사의 엑셀(xls) 포맷이 뒤를 이었다.

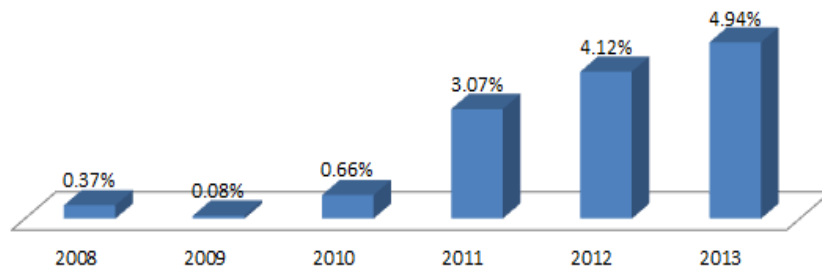
특이 사항은 문서보존포맷(pdf) 비율이 지속적으로 증가하고 있는 점이다. 2008년도에는

0.37%였으나 지속적으로 증가하여 2012년도에는 4.12%, 2013년도에는 4.94%를 차지하고 있는데, 이처럼 문서보존포맷(pdf)으로 변환된 첨부 파일이 증가하는 현상이 타 기관에서도 발생하는지에 대한 검토가 필요하다(〈그림 4〉 참조).

기록물철/건에 포함된 디지털 컴포넌트의 포맷 유형을 조사하여 분석한 결과 아래아한글(hwp)과 마이크로소프트의 엑셀(xls) 포맷이 93.41%로 가장 많으며, 8종(hwp, xls, ppt, doc, pdf, jpg, tif, zip) 포맷이 99.04%로 거의 대부분

〈표 6〉 연도별 주요 포맷 비율

연도	포맷	건수	비율	연도	포맷	건수	비율
2008	hwp	29,290	67.76%	2011	hwp	115,383	83.80%
	xls	3,915	9.06%		xls	15,107	10.97%
	doc	139	0.32%		doc	598	0.43%
	pdf	162	0.37%		pdf	4,225	3.07%
	tif	9,222	21.34%		tif	145	0.11%
	jpg	171	0.40%		jpg	1,684	1.22%
	ppt	324	0.75%		ppt	545	0.40%
합계	43,223	100.00%	합계	137,687	100.00%		
2009	hwp	171,370	98.43%	2012	hwp	89,029	82.25%
	xls	2,197	1.26%		xls	12,919	11.94%
	doc	132	0.08%		doc	309	0.29%
	pdf	147	0.08%		pdf	4,460	4.12%
	tif	6	0.00%		tif	75	0.07%
	jpg	111	0.06%		jpg	1,105	1.02%
	ppt	148	0.09%		ppt	346	0.32%
합계	174,111	100.00%	합계	108,243	100.00%		
2010	hwp	30,477	91.71%	2013	hwp	184,793	84.57%
	xls	2,189	6.59%		xls	18,869	8.64%
	doc	26	0.08%		doc	679	0.31%
	pdf	220	0.66%		pdf	10,791	4.94%
	tif	14	0.04%		tif	154	0.07%
	jpg	242	0.73%		jpg	2,652	1.21%
	ppt	63	0.19%		ppt	577	0.26%
합계	33,231	100.00%	합계	218,515	100.00%		



〈그림 4〉 보존문서 중 PDF 포맷이 차지하는 비율

분을 차지함에 따라 향후 애플리케이션 전략을 수립함에 있어 가장 먼저 아래아한글 포맷과

마이크로소프트의 오피스 포맷에 대한 연구가 시급함을 알 수 있었다. 또한, 보존문서 표준 포

맷(pdf), 이미지 파일 포맷(tif, jpg)은 포맷 정보가 표준으로 공개되어 있을 뿐만 아니라 정적 이미지 정보로 구성되어 있으므로 애플리케이션 전략 수립 대상에서 제외할 수 있다고 판단하였다.

따라서, 본 연구에서는 향후 동적 요소가 포함되어 있는 디지털 컴포넌트를 재현하는 대상으로 아래아한글과 마이크로소프트사의 엑셀 포맷을 대상으로 선정하여 애플리케이션을 수행할 때 필요한 고려 사항들을 검토함으로써 공공기관에 필요한 애플리케이션 전략을 수립하였다.

또한, B기관에서 생산한 확장자가 hwp, xls, ppt인 문서에 대한 포맷 정보를 확인한 결과 확장자와 실제 내부 포맷이 다른 디지털 컴포넌트도 다수 발견되었다. <표 7>과 같이 확장자가 hwp로 되어 있으나 pdf문서, 암호화된 문서, 이미지 문서, xml문서, 압축문서 등 확장자가 hwp로 부여되어서는 안되는 문서들이 다수 존재하고 있었다. 또한 <표 8>과 같이 확장자는

xls로 되어 있으나 문서의 내용을 확인해본 결과 파워포인트, 이미지, pdf, HTML 포맷의 문서도 상당수 존재하고 있었으며 <표 9>와 같이 확장자는 ppt로 되어 있으나 워드나 엑셀 포맷의 문서도 존재하고 있었다.

확장자는 디지털 컴포넌트 파일명에서 확인한 것이며 내부 포맷은 비트스트림을 읽어 내어 확인한 정보이다. 확장자가 실제 포맷과 다르면 해당 비트스트림을 렌더링하고자 할 때 적절한 애플리케이션이 연결되지 않아 오류가 발행하는 등 관리과정에서 처리 상 오류가 발생할 가능성이 높다. 따라서 확장자와 내부 포맷이 일관되도록 보정하는 작업이 필요하다고 판단된다. 이 조사에서는 왜 내부 포맷과 불일치하는 확장자가 부여된 것인지 원인을 추적하지는 못했다. 다만, 이관시점 등 기록관리 단계로 넘어오는 디지털 컴포넌트에 대해 포맷검증 등의 품질검사와 더불어 확장자의 일치성을 확인하는 작업이 필요하다는 점을 확인할 수 있었다.

<표 7> 아래아한글문서 확장자를 가진 다른 포맷 문서의 사례

확장자	포맷	버전	건수
hwp	pdf	0	322
hwp	FasooSecureContai	0	822
hwp	JFIF	0	242
hwp	xml	0	3,896
hwp	SESSION	0	8
hwp	BM	0	11
hwp	Handysoft	0	333
hwp	rtf1	0	3
hwp	GIF89	0	4
hwp	Exif	0	20
hwp	Photoshop	0	3
hwp	ALZ	0	6

〈표 8〉 엑셀문서 확장자를 가진 다른 포맷 문서의 사례

확장자	포맷	버전	건수
xls	JPEG File Interchange Format	1.01	2
xls	Exchangeable Image File Format (Compressed)	2.2	3
xls	Graphics Interchange Format	89a	1
xls	Acrobat PDF 1.4 - Portable Document Format	1.4	3
xls	Acrobat PDF 1.1 - Portable Document Format	1.1	2
xls	Microsoft Powerpoint Presentation	97-2002	4
xls	Acrobat PDF 1.5 - Portable Document Format	1.5	1
xls	Windows Bitmap	3	1
xls	Acrobat PDF 1.3 - Portable Document Format	1.3	1
xls	Acrobat PDF 1.6 - Portable Document Format	1.6	1
xls	Hypertext Markup Language	unknown	200
xls	Hypertext Markup Language	4.01	1
xls	Extensible Hypertext Markup Language	1	1
xls	Hypertext Markup Language	4	1

〈표 9〉 파워포인트 문서 확장자를 가진 다른 포맷 문서의 사례

확장자	포맷	버전	건수
ppt	Microsoft Word Document	97-2003	1
ppt	Microsoft Excel 97 Workbook (xls)	8	5

3.2 문서보존포맷 변환 시 소실되는 기능성 조사결과

앞에서 살펴본 바와 같이 우리나라 공공기관에서 최근 몇 년간 주로 사용하는 디지털 컴포넌트 포맷은 몇 가지 종류로 한정됨을 알 수 있다. 만약 이 디지털 컴포넌트들을 장기보존하면서 재현하기 위해 에뮬레이션 전략을 취한다면 다음으로는 2장에서 살펴본 에뮬레이션 방법 중 어떤 방법을 채택하는 것이 비용효과적이면서도 현실가능한 방안일까를 판단할 차례이다. 이 판단을 위해서는 공공기관에서 생산하는 문서 포맷들이 보유한 기능성과 렌더링 방식을 먼저 파악하고 문서보존포맷인 PDF/A로 변환시

소실되는 기능성이 무엇인지, 렌더링이 불가능한 객체가 무엇인지 파악해야 할 것이다. 에뮬레이션 대상이 되는 디지털 컴포넌트별 기능성은 디지털 컴포넌트를 생산하는 애플리케이션의 버전 및 파일 포맷 버전별로 다르다. 〈표 10〉은 보존포맷 변환시 소실되는 렌더링 객체 속성 분석을 통하여 아래아한글문서, 엑셀문서, 파워포인트문서를 문서보존포맷인 PDF/A로 변환했을 때 소실되는 문서의 기능성들을 분석한 결과이다. 이러한 기능성은 파일의 비트스트림에 코드화된 내용을 애플리케이션이 인식하여 동작함으로써 구현된다.

예를 들어, 아래아한글문서에는 OLE개체를 삽입하여 실행하는 기능이 있다. 아래아한글문

〈표 10〉 문서보존포맷 변환 시 소실되는 기능

애플리케이션	기능성 명칭	설명
한글	OLE 개체	문서 내에 Windows에 설치된 프로그램들이 자신의 기능을 서로 사용하는 기능
	메모장	단어나 문장을 선택하고 설명을 추가하는 기능
	교정부호	단어나 문장이 틀렸을 때 수정할 수 있게 하는 기능
	양식 개체	명령단추, 선택상자, 라디오단추, 입력상자, 목록상자 등을 사용하는 기능
	책갈피	문서를 편집하는 도중에 표시를 해 두었다가 바로 이동할 수 있게 하는 기능
	스크립트	양식 개체와 연결하여 여러 가지 선택을 할 수 있게 하는 기능
파워포인트	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 플레이 하는 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
엑셀	스파크라인	단일 워크시트 셀 안의 작은 차트로서 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 접혀있는 경우, PDF/A로 변환할 때 접힌 상태로 변화되기 때문에 문제 발생

서의 비트스트림에는 OLE개체가 삽입된 부분에 해당 개체에 대한 설명 정보와 함께 삽입된 개체를 코드화한 결과가 포함되어 있다. 이 문서를 열어서 OLE개체를 실행할 때는 한글 애플리케이션뿐만 아니라 비트스트림에 포함된 개체를 실행시키기 위한 별도의 애플리케이션의 기능이 필요하다. 즉, OLE개체가 삽입된 아래아한글문서는 비트스트림만 보관해서는 해당 기능성을 온전히 보존할 수 없으며 OLE개체와 관련된 애플리케이션 기능까지 보관해야 하는 것이다. 그런데 OLE 개체가 삽입된 아래아한글문서를 PDF/A로 변환하게 된다면 어떻게 될까? 당연히 PDF/A로 변환된 문서에서는 OLE개체를 실행할 수가 없게 된다. 따라서 만약 어떤 아래아한글문서가 OLE개체를 포함하고 있고 먼 미래에

도 OLE개체가 실행될 필요가 있다면 OLE개체의 실행가능성을 장기보존하기 위한 전략을 수립해야 하며, PDF/A로 마이그레이션하기 보다는 에플레이션 전략에 따라 아래아한글문서를 보존해야할 것이다. 혹은 어떤 아래아한글문서가 OLE개체를 포함하고 있다 하더라도 실행가능성까지 장기보존 할 필요는 없다고 평가한다면 PDF/A와 같은 문서보존포맷으로 변환하는 것으로 충분할 수 있다. 이러한 판단은 개별 디지털 컴포넌트를 생산하는 사람이 어떤 의도와 목적으로 각 기능성을 문서에 포함했느냐, 그리고 그 기능성이 문서의 기록으로서의 가치에 얼마나 중요한 기여를 하느냐 등에 달려있다고 본다. 누가 언제 이러한 평가를 할 것인지 절차와 기준을 정해야 할 것이

며, 평가 결과에 따라 마이그레이션을 할 대상과 에뮬레이션을 할 대상을 나누어 보존처리하는 프로세스를 마련하는 등 향후 에뮬레이션 전략을 채택하는 경우 보다 세밀한 업무 절차가 정의되어야 할 것이다.

4. 에뮬레이션 서비스 모형 연구

4.1 에뮬레이션 서비스 개발 방향

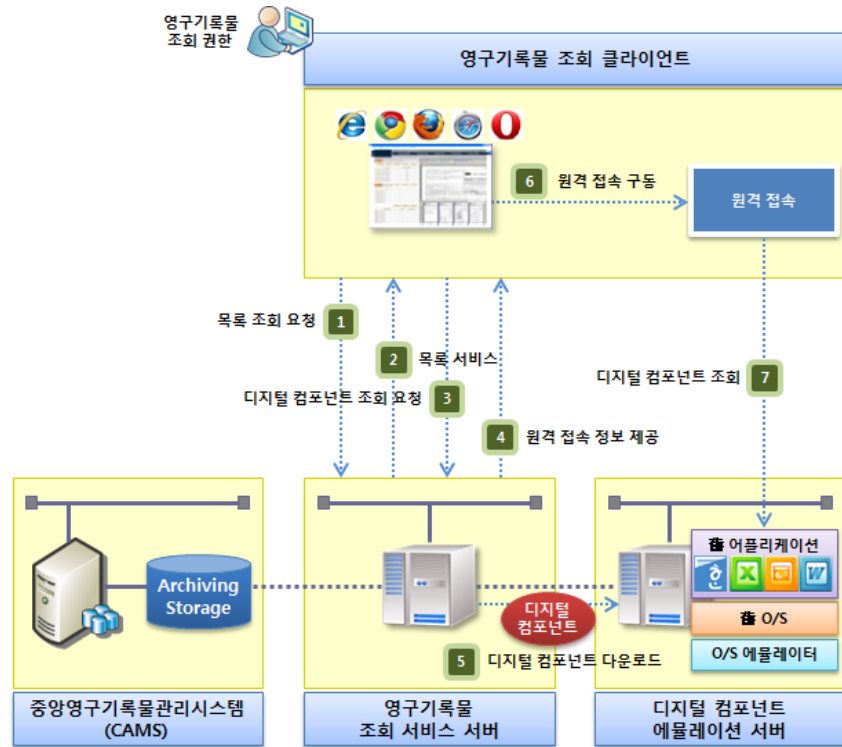
국가기록원은 2015년부터 본격적으로 공공기관의 기록관리시스템(RMS)으로부터 대량의 전자기록을 매년 이관받아 관리해야 한다. 국가기록원의 중앙영구기록물관리시스템(이하 CAMS)에는 장기보존 대상의 기록들 즉 30년, 준영구, 영구 보존기간으로 설정된 전자기록을 보존하면서 이용자들이 언제든지 렌더링해 볼 수 있도록 보장해야 한다. 3장에서 살펴본 바와 같이 이관될 전자기록의 디지털 컴포넌트들은 마이그레이션만으로는 충분한 재현이 불가능하며 이중 일부는 반드시 에뮬레이션이 필요하다. 이에 이 연구에서는 한국의 특수한 상황을 감안하여 국가기록원이 CAMS에 보관하게 될 디지털 컴포넌트를 대상으로 에뮬레이션하는 서비스를 제안하고 에뮬레이터의 프로토타입을 개발하였다. 여기서 말하는 한국의 특수한 상황이란 대부분의 공문서가 아래아한글문서 포맷을 사용한다는 점, 그리고 최상위 비중을 차지하는 디지털 컴포넌트 포맷이 한글과 엑셀이라는 점 등을 고려한 것이다.

2장에서 살펴본 해외 사례 및 기술 정보 동향을 기반으로 판단할 때 국가기록원과 같이 모든

기록물을 아카이빙 스토리지에 집적하여 모아두는 방식에는 첫째, 서버 기반의 운영체제 에뮬레이션 방안과 둘째, 애플리케이션 에뮬레이션 방안 등 두 가지의 모형이 가능하다고 정리할 수 있다.

먼저 서버기반의 운영체제 에뮬레이션 방안은 <그림 5>와 같이 CAMS의 아카이빙 스토리지에 등록된 전자기록을 에뮬레이션 서버와 같은 공간에 두고 이용자가 원격 접속으로 서버에 접근하여 기록물을 보는 방식이다. <그림 5>에서 제시하는 서버 기반의 운영체제 에뮬레이션 서비스의 흐름을 살펴보면 다음과 같다.

- (1) 이용자가 영구기록물을 조회하기 위한 클라이언트 프로그램을 이용하여 '영구 기록물조회 서비스 서버'에 원하는 기록물의 목록을 조회요청 한다.
- (2) '영구기록물 조회 서비스 서버'는 이용자가 제시한 조건을 만족한 기록물의 목록을 제공한다.
- (3) 이용자가 특정 기록물의 디지털 컴포넌트를 열어서 보고자 조회 요청을 한다.
- (4) '영구기록물 조회 서비스 서버'는 이용자에게 디지털 컴포넌트를 에뮬레이션해서 보여줄 서버에 접속하기 위해 필요한 관련 정보를 제공해 준다.
- (5) '영구기록물조회 서비스 서버'는 이용자가 열어서 보고자 요청한 디지털 컴포넌트를 '디지털 컴포넌트 에뮬레이션 서버'로 다운로드한다.
- (6) 이용자는 원격 접속 정보를 이용하여 '디지털 컴포넌트 에뮬레이션 서버'에 원격 접속할 준비를 한다.



〈그림 5〉 서버 기반의 운영체제 에뮬레이션 서비스

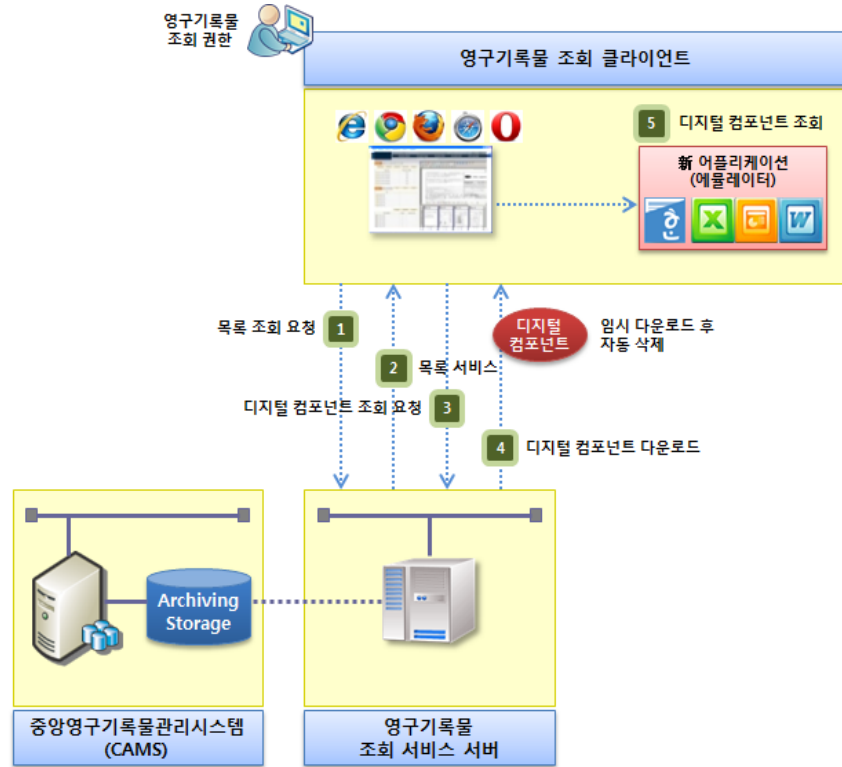
(7) 이용자는 '원격 접속' 프로그램을 통해 '디지털 컴포넌트 에뮬레이션 서버'에 접속하여 '영구기록물 조회 서비스 서버'에서 다운로드된 디지털 컴포넌트를 생산 당시의 컴퓨팅 환경에서 열어보게 된다.

디지털 컴포넌트를 조회하기 위해서는 전자기록에 대한 조회 권한을 부여받은 사용자에 한하여 접근이 가능하도록 하여야 한다. 디지털 컴포넌트 에뮬레이션 서버는 구형 애플리케이션의 실행을 지원하기 위해 운영체제 에뮬레이터 및 구형 운영체제가 설치되어야 하고, 서버에 구형 운영체제를 설치하기 위해서는 1:1로 에뮬레이터를 설치하여야

하기 때문에 에뮬레이션 대상 운영체제의 수가 늘어나면 설치하는 에뮬레이터도 비례하여 증가한다.

둘째, 애플리케이션 에뮬레이션 방안은 〈그림 6〉과 같이 에뮬레이터에 해당하는 통합 뷰어 프로그램을 이용자 컴퓨터에 설치하여 디지털 컴포넌트를 렌더링해 보도록 하는 것이다. 〈그림 6〉에서 제시하는 애플리케이션 에뮬레이션 서비스의 흐름을 살펴보면 다음과 같다.

(1) 이용자가 영구기록물을 조회하기 위한 클라이언트 프로그램을 이용하여 '영구기록물조회 서비스 서버'에 원하는 기록물의 목록을 조회요청 한다.



〈그림 6〉 애플리케이션 에뮬레이션 서비스

- (2) '영구기록물 조회 서비스 서버'는 이용자가 제시한 조건을 만족한 기록물의 목록을 제공한다.
- (3) 이용자가 특정 기록물의 디지털 컴포넌트를 열어서 보고자 조회 요청을 한다.
- (4) '영구기록물 조회 서비스 서버'는 이용자가 원하는 디지털 컴포넌트를 이용자의 '영구기록물 조회 클라이언트' 쪽으로 다운로드한다.
- (5) 이용자의 클라이언트 쪽에 설치되어 있는 '에뮬레이터' 프로그램을 이용하여 현재의 컴퓨팅 환경에서 원하는 디지털 컴포넌트를 열어보게 된다.

현재까지 개발된 대부분의 애플리케이션은 클라이언트에서 실행되는 방식으로 클라이언트에 에뮬레이터가 설치되어 있어야 한다. 디지털 컴포넌트에 대한 뷰어는 현재 개발되어 서비스되는 뷰어와 같은 방식으로 하나로 통합되어서 서비스할 수 있도록 개발되어야 한다. 신형 운영체제가 출시되면 구형 운영체제에서 생산된 디지털 컴포넌트를 조회할 수 있도록 에뮬레이터 개발이 필요하므로 구형 디지털 컴포넌트에 대한 포맷 정보에 대한 보존 관리가 필요하다. 그리고 디지털 컴포넌트를 조회하기 위해서는 PC나 디바이스마다 애플리케이션이 설치되어야 하고, 각 클라이언트에 애플리케이션을 설

치할 때 발생하는 저작권에 대한 정책 수립도 병행하여야 한다.

4.2 운영체제 에뮬레이션 프로토타입 개발 결과

이 연구에서 개발한 한국형 에뮬레이터 프로토타입은 운영체제 에뮬레이션 방식과 애플리케이션 에뮬레이션 방식에 대한 서비스를 제공하는 것으로 각각 설계되었다. <표 11>에서 보는 바와 같이 운영체제 에뮬레이션은 윈도우7(Host O/S)에 오라클(Oracle)사의 VirtualBox를 설치한 후 윈도우 XP(Guest O/S)에서 디지털 컴포넌트에 대한 조희가 가능하도록 구현하였고, 애플리케이션 에뮬레이션은 엑셀 포맷에 대한 뷰어를 개발하여 서비스를 구현하였으며 디지털 컴포넌트 구성기(Configurator)는 MySQL, Tomcat 기반의 웹 서비스 형태로 구성한 후 구형 운영체제와의 통신을 위해 VMServer를 별도로 개발하였다.

<그림 7>과 같이 운영체제 에뮬레이션을 수행하기 위해 프로토타입 시스템에서는 오라클

사의 VirtualBox를 도입하여 에뮬레이션 환경을 구축한 후 기록물 철/건 목록에서 디지털 컴포넌트를 선택하였을 때 구형 애플리케이션을 자동으로 구동하기 위해 디지털 컴포넌트 구성기와 VMServer를 개발하여 환경을 구축하였다.

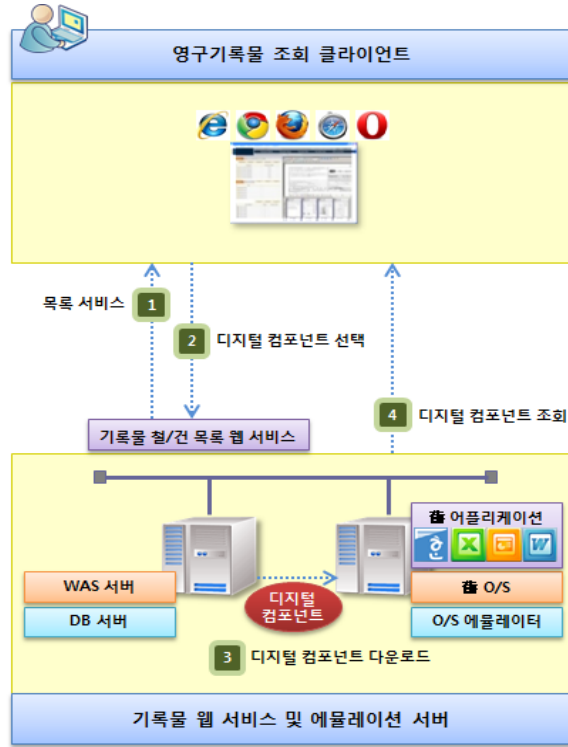
운영체제 에뮬레이션 서비스 결과 CAMS와 동일한 형식의 전자기록 철/건에 포함된 디지털 컴포넌트 목록에서 특정 디지털 컴포넌트를 선택하면 운영체제 에뮬레이터를 통해 구형 운영체제에서 해당 디지털 컴포넌트를 조회할 수 있도록 하였다. <그림 8>과 같이 “에뮬레이터 보기”를 선택하면 구성기에서 운영체제(Guest O/S)를 선택할 수 있게 되어 있고, 운영체제에서 디지털 컴포넌트를 볼 수 있는 애플리케이션을 구동시켜 보여준다.

프로토타입으로 개발된 운영체제 에뮬레이터를 실행하기 위해서는 첫째 에뮬레이션 서버 실행, 둘째 웹 서비스 서버를 실행하여 에뮬레이션 서버 환경을 구성하여야 한다.

에뮬레이션 서버 실행은 <그림 8>과 같이 VirtualBox를 실행한 후 VirtualBox 기반에서

<표 11> 에뮬레이션 프로토타입을 위한 시스템 구성

도입 또는 개발 항목		프로토타입	방식
운영체제 에뮬레이션	신형 운영체제	윈도우 7	도입
	구형 운영체제	윈도우 XP	도입
	구형 애플리케이션	한/글 파워포인트 엑셀 워드 훈민정음	뷰어 도입(무상)
	에뮬레이터	VirtualBox(오라클)	도입
	디지털 컴포넌트 구성기		개발
애플리케이션 에뮬레이션	에뮬레이터(뷰어)	엑셀	개발
	디지털 컴포넌트 구성기		개발



〈그림 7〉 운영체제 에뮬레이션 서비스 구성



〈그림 8〉 VirtualBox 실행 화면

구형 운영체제를 구동시킨 후 웹 서비스 프로그램과의 통신을 위해 <그림 9>와 같이 자체 제작한 VMServer를 실행시킨다.

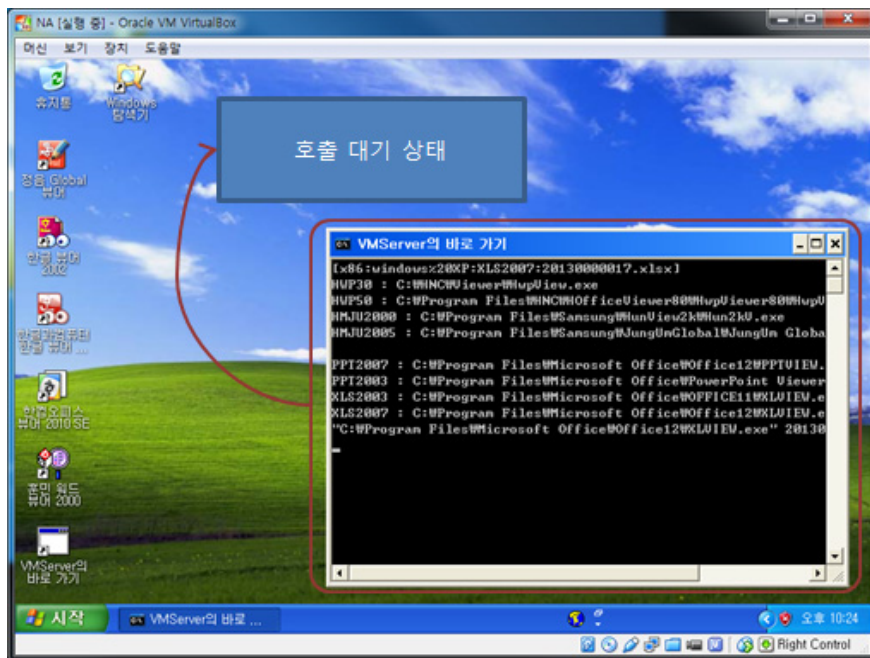
웹 서비스 서버 실행은 기록물 목록에 대한 웹 서비스를 위해 <그림 10>과 같이 DB 서버를 실행 한 후 <그림 11>과 같이 WAS 서버를 실행시킨다.

마지막으로 운영체제 기반의 에뮬레이션 서비스를 위해 웹 브라우저를 통해 에뮬레이션 서비스 사이트에 접속하면 <그림 12>와 같은 초기 화면이 제공되는데 여기에서 에뮬레이션을 하고자 하는 기록물 철/건에 포함된 디지털 컴포넌트를 선택하면 <그림 13>과 같이 에뮬레이션 서비스를 제공받을 수 있다. '에뮬레이터 보기'를 선택하면 해당 디지털 컴포넌트를 이용할 때 작동할 수 있는 운영체제 종류와 예

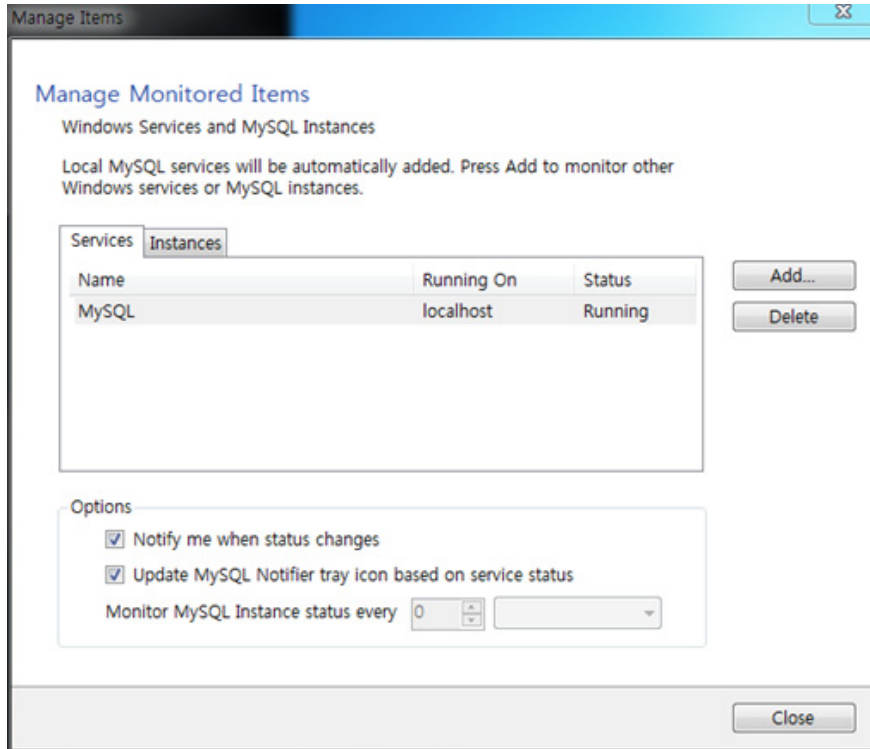
레이터 종류 등 여러 구성이 뜨고, 이 중 원하는 조합으로 '에뮬레이션 선택'을 하면 해당 조건에 맞춰 실행된다.

4.3 애플리케이션 에뮬레이션 프로토타입 개발 결과

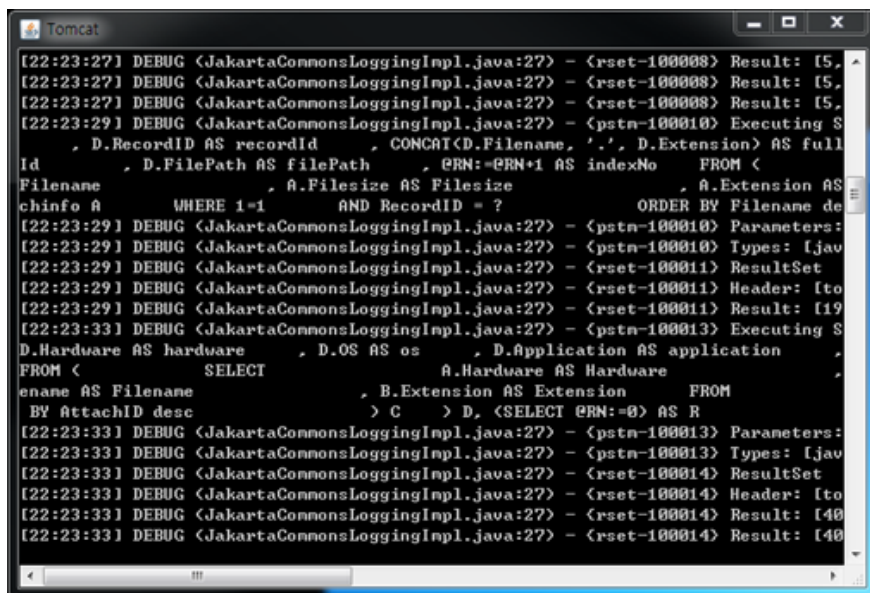
<그림 14>와 같이 애플리케이션 에뮬레이션 실행을 위해 엑셀 포맷에 대한 웹 뷰어를 개발한 후 기록물 철/건 목록에서 디지털 컴포넌트를 선택하였을 때 '웹 뷰어 에뮬레이터'가 자동으로 구동되도록 디지털 컴포넌트 구성기를 개발하여 애플리케이션 에뮬레이션 환경을 구축하였다. 프로토타입에서는 운영체제 에뮬레이션과 애플리케이션 에뮬레이션에 대한 실행이 하나의 서버에서 구동되도록 구축하였으나 향후 웹 서



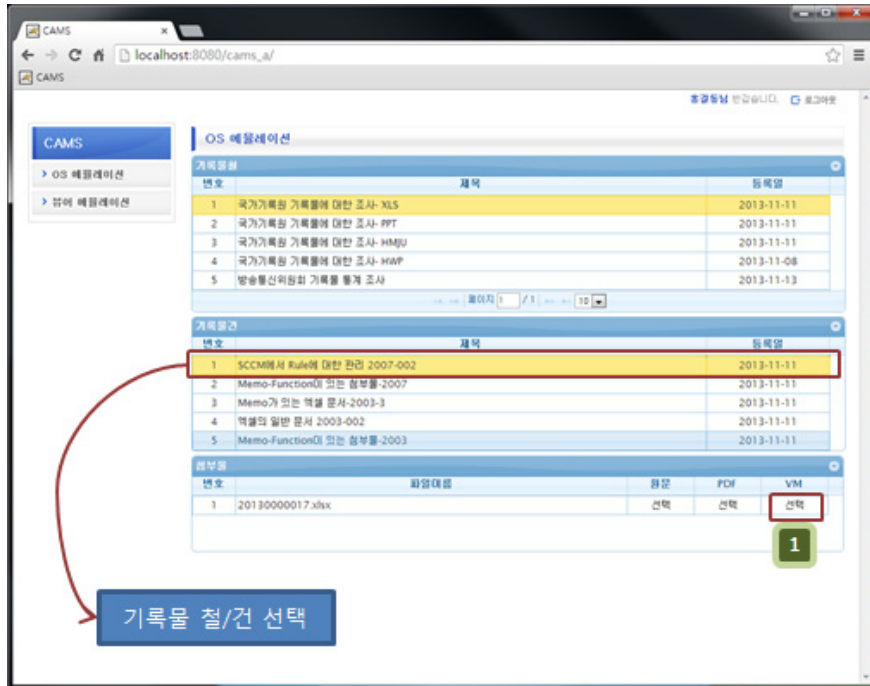
<그림 9> VMServer 실행 화면



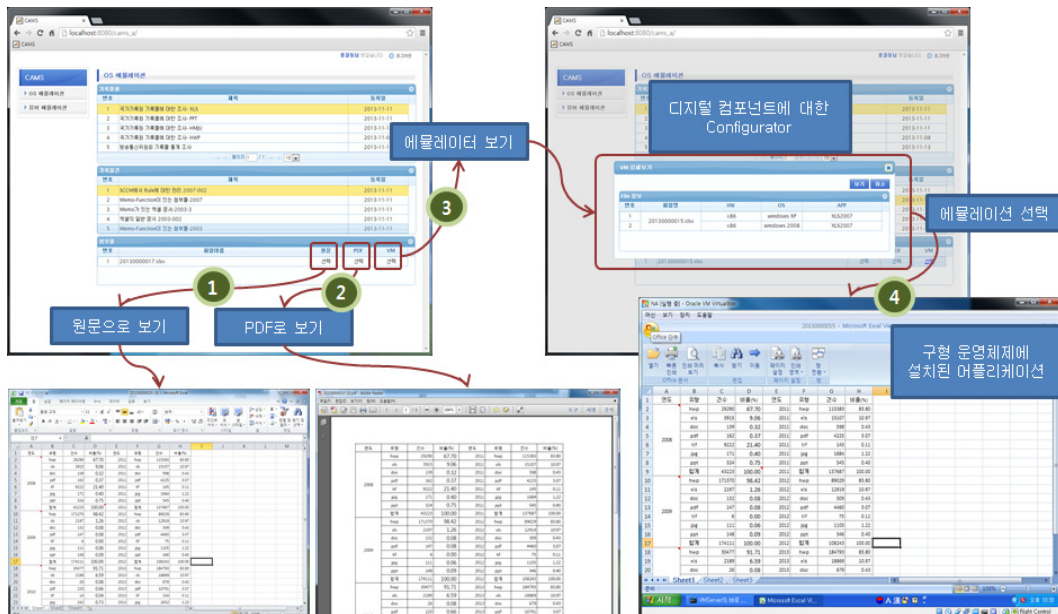
〈그림 10〉 MySQL 실행 화면



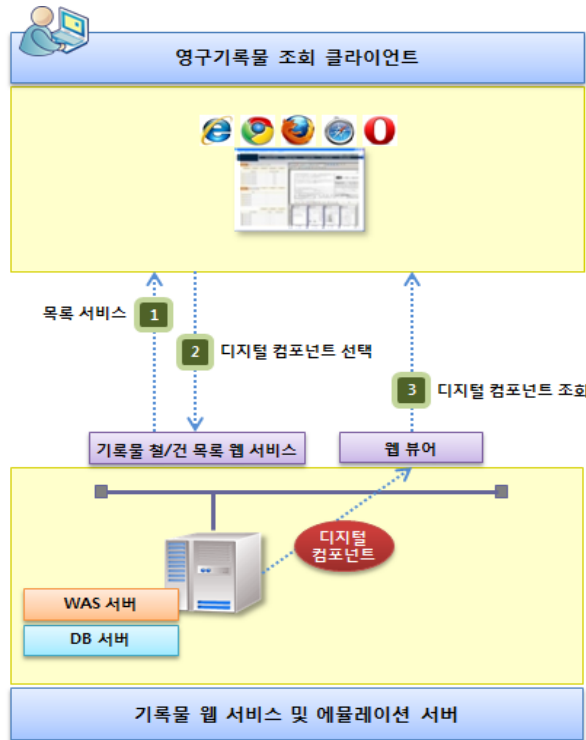
〈그림 11〉 Tomcat 실행 화면



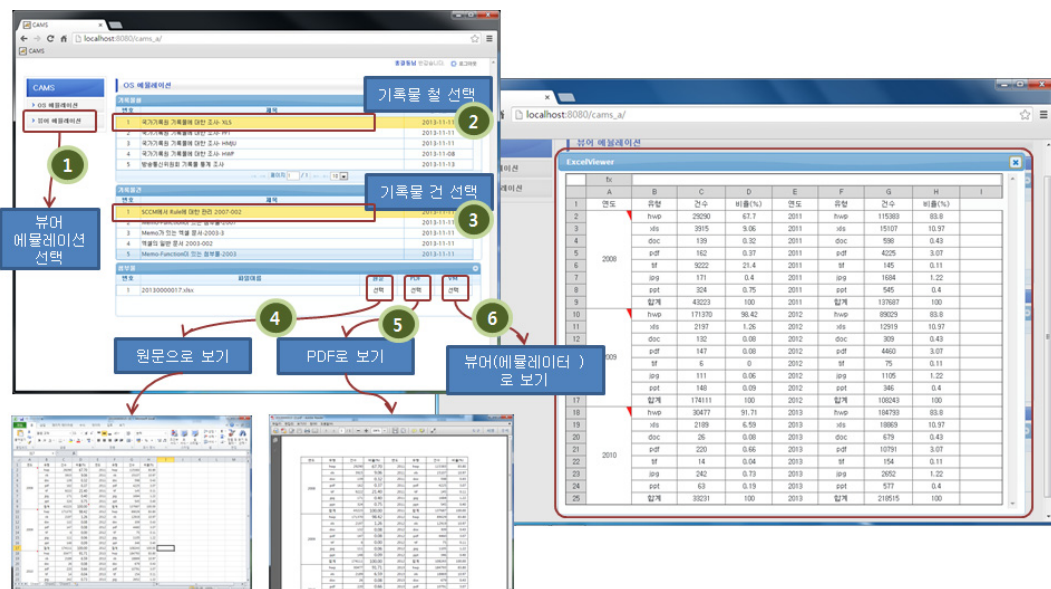
〈그림 12〉 운영체제 에뮬레이션 서비스 초기 화면



〈그림 13〉 운영체제 에뮬레이션 서비스 결과 화면



<그림 14> 애플리케이션 에뮬레이션 서비스 구성



<그림 15> 애플리케이션 에뮬레이션 서비스 결과 화면

스 서버와 에뮬레이션 서버로 분리가 가능한 구조이다.

CAMS와 동일한 형식의 전자기록 철/건에 포함된 디지털 컴포넌트 목록에서 특정 디지털 컴포넌트를 선택하면 <그림 15>와 같이 웹 뷰어 형태의 애플리케이션 에뮬레이터를 통해 조회 서비스를 제공하도록 하였다. 디지털 컴포넌트의 원문을 생산시의 애플리케이션을 통해 열어볼 수도 있고(<그림 15>의 (4)), PDF/A로 변환된 버전을 열어볼 수도 있고(<그림 15>의 (5)), 프로토타입으로 개발한 에뮬레이터를 통해 열어볼 수도 있도록(<그림 15>의 (6)) 구성하였다. 예를 들어, 열기의 대상이 된 디지털 컴포넌트 원문이 엑셀파일이었다면, (4)의 방식으로 마이크로소프트 엑셀을 이용하여 열기를 한 엑셀파일은 수정, 편집, 삭제 등 모든 기능의 구사가 가능한 상태가 된다. (5)의 방식으로 뷰어를 이용하여 열기를 한 PDF/A 파일은 보기와 인쇄만이 가능한 상태가 된다. 만약 (6)의 방식으로 에뮬레이터를 이용하여 열기를 한 엑셀파일은 편집, 삭제 등은 불가능하며 보존하고자 원한 메모 기능만 작동하는 상태가 된다.

5. 맺음말

전자기록 장기보존 전략으로서 에뮬레이션이 국내 기록관리 영역에 본격적으로 인식하기 시작한 것은 ISO 14721에서 마이그레이션 전략과 더불어 대표적인 보존 전략으로서 에뮬레이션을 소개한 것에서 시작된다. 그러나 ISO 14721 표준에서는 에뮬레이션에 대해 개념적인 설명에 머물고 있을 뿐 실행 방법으로서 에

뮬레이션을 어떻게 실제 기록관리 환경에서 적용할 수 있을지에 관한 내용은 다루지지 않았다. 우리나라 공공기록관리 부문은 최근에 전자기록의 장기보존 전략으로서 에뮬레이션 전략을 연구하기 시작하여 현재까지는 초기 조사 연구 단계에 와있는 수준이다.

에뮬레이션 전략은 기술의 복잡성이 높고 구현 비용의 추산이 어려워 효과적인 에뮬레이션이 현실적으로 적용가능한지 판단이 어려우며, 해외의 경우에도 장기간, 대규모 예산을 들인 공동 연구를 통해서 지속적인 검토와 연구를 진행한 바는 있지만 현실적으로 각국에서 이러한 에뮬레이션 전략을 성공적으로 도입한 사례는 아직 파악되고 있지 않다.

그러나 다양한 가능성을 보유한 디지털 컴포넌트들이 계속적으로 생산되고 있고 원래의 기록 그대로를 보려고 하는 인간의 욕구 또한 변함없을 것이며 에뮬레이션 전략이 아니고서는 재현이 불가능한 디지털 컴포넌트들의 기능성들은 여전히 존재할 것이므로 장기보존에 있어서 에뮬레이션이 필요한 영역이 분명하고 각국의 환경에 근거한 에뮬레이션 전략의 필요성 자체는 변함없다(국가기록원, 2009, p. 3).

본 연구에서는 운영체제 에뮬레이션 방안과 애플리케이션 에뮬레이션 방안 각각에 대한 프로토타입을 개발한 사례를 제시하고 있다. 두 가지 방식은 개발 과정과 비용, 적용 디지털 컴포넌트, 서비스 절차와 방식 등에서 차이를 보인다. 연구의 중요한 시사점은 에뮬레이션이라는 장기보존 전략을 취할 때 세부적인 접근법에는 상이한 방식이 존재한다는 것과 그 모두가 구현가능하다는 것을 확인했다는 점이다. 또한, 본 연구의 결과를 응용하여 우리나라의

공공기록물의 장기보존을 위한 최선의 한국형 에물레이션 전략을 도출할 수 있다는 점이다. 즉, 에물레이션을 검토했다고 해서 현재의 장기보존 전략을 통째로 바꾸어야 한다거나 혹은 에물레이션 전략과 현재의 장기보존 전략이 공존할 수 없는 것이 아니라 에물레이션 전략을 이용하여 현재의 장기보존 전략을 보완할 수 있다는 것이다.

현재 우리나라 공공 전자기록의 장기보존 전략으로 마이그레이션과 인캡슐레이션 전략이 함께 적용되고 있다. 즉, 전자기록을 구성하는 디지털 컴포넌트 각각은 PDF/A라는 문서보존포맷으로 변환하면 되는데 이 전략은 마이그레이션 전략이다. 메타데이터와 디지털 컴포넌트 원문, 디지털 컴포넌트 PDF/A 버전, 전자서명 등을 모두 XML 문서화하는데 이 전략은 인캡슐레이션 전략이다. 그런데, 앞에서 살펴본 듯이 디지털 컴포넌트 중에는 PDF/A로 변환할 수 없는 것이나 변환했을 때 기능성을 상실하는 것이 존재하며, 이런 기록물의 경우에는 마이그레이션 전략 대신 에물레이션 전략을 채택할 수 있도록 하는 것이 필요하다. 기록관리 과정에서 OLE개체를 가진 아래아한글문서, 복잡한 수식과 함수를 통해 셀의 값을 계산한 결과값을 가진 엑셀문서 등 마이그레이션으로는 충분히 이용가능성을 보존하기 어려운 디지털 컴포넌트들을 에물레이션해야할 대상으로 지정하고 이후 에물레이션 방식으로 디지털 컴포넌트를 열어 볼 수 있도록 하는 것이다. 이 때, 운영체제 에물레이션의 방식을 채택하느냐 애플리케이션 에물레이션 방식을 채택하느냐에 따라 에물레이션 대상 디지털 컴포넌트에 대해 관리해야할 메타데이터가 달라지게 될 것이다.

운영체제 에물레이션을 하는 경우에는 해당 디지털 컴포넌트를 열어보는데 필요한 애플리케이션에 대해 기술하고 애플리케이션 자체를 잘 보존하는 것이 중요하다면, 애플리케이션 에물레이션을 하는 경우에는 해당 디지털 컴포넌트에 어떤 기능성을 보존해야하는지를 기술해두고 에물레이터를 개발할 때 필요한 포맷정보를 잘 남기는 것이 중요하다. 애플리케이션 에물레이션 방식을 취하는 경우 전자기록을 생산하는 시점에 업무담당자들이 각 디지털 컴포넌트 별로 보존해야할 기능성을 기술하도록 하는 것이 필수적이다. 어떤 기능성을 보존할 필요가 있는지는 업무담당자 만이 제대로 파악할 수 있기 때문이다. 물론, 특성화(Characterization) 도구를 사용하여 디지털 컴포넌트의 비트스트림을 필터링하여 PDF/A와 같은 문서보존포맷으로 변환 시 상실되는 기능성을 찾아내어 업무담당자에게 제시하고 이 중에서 보존해야만 하는 기능성을 선택하도록 하는 방법도 구상할 수 있다. 이러한 특성화 도구를 마련하기 위해서는 현재 공공기관에서 작성되고 있는 포맷별로 에물레이션 대상이 될 수 있는 기능성의 종류와 각 비트 패턴을 데이터베이스화하는 것이 필요하다.

이처럼 우리나라 공공기록물관리 분야에서 에물레이션 전략은 현재의 마이그레이션 전략, 인캡슐레이션 전략과 함께 보완적 관점에서 적용될 수 있다. 해외 에물레이션 프로젝트 사례에서 살펴본 바와 같이 운영체제 에물레이션 전략을 적용하기 위해서는 상당한 재정 투자가 필요함을 알 수 있다. 따라서, 현재의 마이그레이션, 인캡슐레이션 전략을 그대로 두고 반드시 에물레이션이 필요한 디지털 컴포넌트에 대

해서만 에뮬레이션 전략을 복합적으로 적용하는 한국형 전자기록 장기보존 전략을 수립함으로써 비용효과적이면서도 실현가능한 전략 구사가 가능해질 것이다.

본 연구에서 제시한 에뮬레이션 서비스는 국가기록원과 같은 큰 규모의 공공기록물관리 기

관에서 제공할 수 있는 서비스이다. 향후 아래 아한글문서와 같이 우리나라에서 만들어진 독창적인 포맷들의 에뮬레이션 서비스를 제공할 수 있도록 국가기록원에서는 에뮬레이션 관련 투자와 연구를 지속적으로 해나가야 하며, DFR 정보도 지속적으로 갱신해 나가야 할 것이다.

참 고 문 헌

국가기록원 (2009). 전자기록물 장기보존을 위한 에뮬레이션 전략 연구. 대전: 국가기록원

국가기록원 (2013). 행정기관 전자기록물 재현기술 연구 및 프로토타입 개발. 대전: 국가기록원.

김명훈, 오명진, 이재홍, 임진희 (2013). 전자기록 장기보존 전략으로서의 에뮬레이션 사례 분석. 기록학연구, 38, 265-309.

Digital preservation Testbed (2003). Emulation: Context and current status, Digital preservation testbed white paper.

National Archives of Australia (2004). Digital recordkeeping: Guidelines for creating, managing and preserving digital records [Exposure Draft].

[웹사이트]

CAMiLEON. Retrieved 2013. 12. 1, from <http://www2.si.umich.edu/CAMILEON/index.html>

KB. Retrieved 2013. 12. 1, from <http://www.kb.nl/en/expertise/e-depot-and-digital-preservation>

Planets. Retrieved 2013. 9. 1, from <http://www.Planets-project.eu/>

KEEP. Retrieved 2013. 9. 1, from <http://www.keep-project.eu/>

• 국문 참고자료의 영어 표기

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

Kim, Myung-Hun, Oh, Myung-Jin, Lee, Jae-Hong, & Yim, Jin-Hee (2013). An analysis of cases of emulation for long term electronic records preservation strategy. The Korean Journal of Archival Studies, 38, 265-309.

The National Archives of Korea (2009). A study on the emulation strategy for long-term

preservation of electronic records. Daejeon: The National Archives of Korea.
The National Archives of Korea (2013). A study on the reproduction technology and the prototype for the electronic records of administrative agency. Daejeon: The National Archives of Korea.