

전자기록관리를 위한 포맷등록시스템 개발 연구

A Study on the Development of Format Registry for ERM

유 영 수(Young-Soo Yu)*

목 차

- | | |
|---------------------|-----------------------|
| 1. 서 론 | 3.1 영국 PRONOM Service |
| 1.1 연구목적 | 3.2 GDFR 프로젝트 |
| 1.2 선행연구 | 3.3 시사점 |
| 1.3 연구방법과 범위 | 4. 포맷등록시스템 개발 방안 |
| 2. 포맷등록시스템의 필요성과 기능 | 4.1 포맷등록시스템 관리체계 |
| 2.1 포맷등록시스템의 필요성 | 4.2 포맷등록시스템 정보요소 |
| 2.2 포맷등록시스템의 기능 | 4.3 포맷등록시스템 기능과 활용 |
| 3. 해외 사례 분석 | 5. 결론 및 제언 |

<초 록>

종이기록과 달리 특정 하드웨어와 소프트웨어의 조합에 의해서만 접근 가능한 전자기록은 그들의 생산에 이용된 저장매체, 하드웨어와 소프트웨어, 포맷의 수명보다 훨씬 더 길뿐만 아니라 시간이 흐르면 퇴화되거나 노후화되기 때문에 이에 대한 대비가 철저하게 이루어져야만 디지털정보의 장기보존과 접근, 미래 특정시점에서의 재현성을 보장할 수 있다. 이에 본 연구에서는 전자기록의 장기보존과 장기 접근성 유지에 근간이 되는 기술정보요소를 지속적으로 수집·관리하여 마이그레이션이나 에뮬레이션과 같은 보존 전략을 효율적으로 지원하는 포맷등록시스템 개발 방안을 제안하였다.

주제어: 포맷등록시스템, 포맷레지스트리, 장기보존, 기술정보

<ABSTRACT>

Unlike paper records, accessing electronic records needs to be supported by specific combination of hardware and software. The electronic records have longer life span than the storage media, hardware, software and the formats that were used for the production of such records. Because the quality of electronic record can become deteriorated as time goes on, we need to have appropriate mechanisms to secure the long-term preservation and access to the digital information at a certain point in the future. Based on this reality, this study proposes the functions and the development method of the format registry, which efficiently supports the preservation strategy such as migration and emulation, collecting and managing technical information elements. These elements consist of the foundation of the long-term preservation of the electronic records.

Keywords: Format Registry, Long-term Preservation, Technical Information

* 명지대학교 기록관리학 대학원(hbgirl@korea.com)

논문접수일자 2007년 6월 7일

게재확정일자 2007년 6월 11일

1. 서론

1.1 연구목적

우리가 살고 있는 오늘날의 사회는 디지털 세상이라 해도 과언이 아니다. 점점 더 많은 정보들이 디지털 형식으로 생산·처리되고 이용 가능해짐에 따라 전자기록에 대한 의존도는 나날이 높아지고, 영향력 또한 증대되고 있다. 이와 동시에 이런 디지털 기록에 대한 장기접근성 유지 문제에 대한 연구 활동의 중요성 역시 증대되고 있다.

일반적으로 디지털정보의 보존과 관리는 기술적(technical), 관리적(administrative), 절차적(procedural), 조직적(organizational), 그리고 정책적(policy) 이슈들과 상호 관련된다. 이 중에서도 가장 기본이 되는 이슈가 기술적 문제이다.

지금 현재 기술노후화 솔루션으로 주목을 받고 있는 마이그레이션이나 에뮬레이션, 인캡슐레이션 기법 모두 디지털 기록의 장기보존을 위한 완전한 해법이라고 할 수는 없다. 그리고 포맷, 소프트웨어와 하드웨어, 문서유형이나 레코드키핑 실무가 변할 때마다 매번 새로운 보존 솔루션을 개발하는 것 또한 현실적으로 불가능하다. 따라서 이 모든 문제를 다방면에서 지원해줄 수 있는 강력한 메커니즘이 필요하다.

이에 본 연구에서는 전자기록의 장기보존과 장기 접근성 유지에 근간이 되는 기술정보(technical information)요소, 즉 포맷정보, 하드웨어와 소프트웨어정보, 운영시스템 정보, 매체정

보를 지속적으로 수집·관리하여 마이그레이션이나 에뮬레이션과 같은 보존 전략을 효율적으로 지원하는 포맷레지스트리¹⁾의 기능과 역할에 대하여 살펴보고, 우리나라 환경에 적합한 포맷등록시스템 개발방안을 제안하였다.

1.2 선행연구

전자기록의 장기보존에 필요한 기술정보(technical information)에 관한 선행연구는 많지 않다. 다만 포맷정보 관리를 위한 몇몇 프로젝트와 기술적 노후화에 대비한 보존전략에 관한 연구를 통해 장기보존의 개념과 세계적 흐름을 살펴볼 수 있는 수준이다.

포맷정보 관리에 관한 가장 대표적인 선행 연구로는 IANA(Internet Assigned Names Authority) MIME type registry가 있다. 이것은 이메일에 첨부된 파일을 설명해주는 표준으로, application, audio, image, message, multipart, text, video 등 총 7가지 유형으로 구분된다. 그러나 MIME포맷유형은 포맷정보도 정교하지 않을 뿐만 아니라 일관성도 없어 장기보존을 지원하기에 적합하지 않다.

1995년 European Commission의 Information Society Technologies(IST) Program은 포맷표준과 규격에 대한 참조정보와 안내정보를 제공하는 Diffuse.org 웹사이트 개발 프로젝트를 지원하였으나, 2003년 1월 서비스가 중단되었다. 이런 정보원들은 보존을 목적으로 설계된 것이 아니므로 장기보존을 지원해 줄 어떤 서비스도 제공하지 않는다는 한계가 있다.

1) 이후 본 논문에서는 포맷레지스트리(format registry)를 포맷등록시스템으로 통일하여 지칭하겠다.

장기보존전략에 관한 연구로는 로텐베르크의 연구를 들 수 있다.²⁾ 로텐베르크는 디지털 보존을 위한 해결책으로 에물레이션과 인캡슐레이션 기법을 강하게 지지하는 동시에 마이그레이션의 필요성도 지적하고 있다.

한편 국내에서는 설문원, 김연정, 천권주가 ISO/TR 18492의 장기보존전략 목표, 요소, 정책적 고려사항을 체계적으로 정리하고, 아울러 마이그레이션, 에물레이션, UVC 등 장기보존 기법에 대한 설명을 덧붙여 제시하였다. 그러나 최신 보존기법에 대한 내용을 다소 소략하여 다루었다는 한계를 지닌다(설문원, 김연정, 천권주 2006).

1.3 연구방법과 범위

본 연구에서는 우선 포맷등록시스템의 개념과 필요성을 살펴본 뒤 그 기능에 대해 제시하였다. 그리고 포맷등록시스템의 핵심적인 구성요소인 기술정보요소를 중심으로 해외사례를 분석하여 시사점을 도출하였다. 마지막으로 이러한 시사점을 근거로 우리나라에 적합한 포맷등록관리시스템의 관리체계와 그 활용방안을 제안하였다. 그리고 PRONOM 정보요소와 GDFR 정보요소를 분석하여 우리 환경에 맞는 기술메타데이터요소를 정리하였다.

본 연구는 전자기록의 장기보존을 위한 포맷등록관리시스템에 대한 개념 정립과 구축 방안을 제안한 국내 최초의 논문이라는데 의의가 있다. 그러나 본 시스템 구성에 필요한 기본적인 정보요소만을 제시하고, 정보공유와 호환을

위한 구체적인 실행방안을 제안하지 못했다는 한계가 있다.

2. 포맷등록시스템의 필요성과 기능

2.1 포맷등록시스템의 필요성

종이기록과 달리 기술 의존적 특성을 갖는 전자기록은 향후 기술노후화라는 심각한 문제를 야기한다. 마이그레이션, 에물레이션, 인캡슐레이션과 같은 보존 기법들이 실행되고는 있으나 하나의 보존전략만을 고집하기 보다는 기록의 특성과 보존 초점에 따라 다양한 보존기법을 선택하여 병행하는 것이 효과적인 장기보존에 도움이 될 것이다. 어떤 보존전략을 선택하든 기록의 생명주기 동안 반복되어 적용될 것이고, 특정 기록 건에 대한 전략이 어느 시점에 바뀔 수도 있다. 기록에 대한 가독성을 최대한 보장하고, 미래의 어느 시점에서든 어떠한 보존전략도 선택이 가능하기 위해, 그리고 다양한 보존전략을 적시에 적절하게 적용하기 위해서는 보다 안전한 방법이 필요하다. 이미 해외 각국에서는 마이그레이션, 에물레이션과 같은 보존 전략 외에 포맷등록시스템에 대한 연구가 활발하게 진행되고 있다.

본 시스템이 소장한 기술메타데이터정보는 전자기록의 보존과 활용에 근간이 되는 요소로, 마이그레이션이나 에물레이션과 같은 장기보존전략 지원에 이용될 것이다. 그리고 기술정보의 업데이트는 물론 각각의 보존전략에 관한

2) Jeff Rothenberg, Council on Library and Information Resources, 1998, 『Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation』

상세 정보를 문서화하여 제공하므로, 좀 더 유연한 보존활동을 가능하게 한다. 또한 끊임없는 모니터링과 위험 평가활동을 통해 기술변화에 대해 시의 적절하게 대응할 수 있다.

현재 우리나라 공공기관에서 등록·관리하는 전자기록은 워드, 엑셀, 파워포인트, PDF, ZIP파일, JPEG파일 정도로, 그 종류가 다양하지 않다. 그러나 정부혁신지방분권위원회의 국가기록관리혁신 로드맵 아젠다별 세부과제가 구체적으로 실천되면 보다 다양하고 많은 전자기록들이 생산, 등록, 관리될 것으로 예상된다. 따라서 앞으로 증가할 전자기록을 효율적으로 보존하고 장기접근성을 더 먼 미래로까지 확장하기 위해서는 포맷등록시스템 제도의 도입이 고려되어야 한다.

요컨대 포맷등록시스템은 디지털정보의 장기보존과 접근성을 유지하는데 중요한 정보원을 제공해준다. 그리고 포맷정보의 집중화는 물론, 이에 따른 계획된 보존전략 수립을 가능하게 한다. 또한 본 시스템이 등록·관리하는 재현정보는 보존 리파지토리가 유지하는 디지털객체의 식별, 검증 및 처리에 이용된다. 즉, 이런 제도의 도입은 기록이용서비스의 확대 및 질적 향상을 가능하게 하여 궁극적으로는 기록의 활용성을 높이는 계기가 될 것이다.

2.2 포맷등록시스템의 기능

2.2.1 전자기록의 기술적 특성 관리

(technical characterizations)

디지털 객체의 장기보존을 위해서는 우선 저장된 디지털객체의 특성을 이해해야 한다. 그 객체가 어떤 매체에, 어떤 포맷으로 저장되어

있는지, 그리고 그 기록을 생산하는데 이용된 하드웨어와 소프트웨어는 무엇인지 등을 정확하게 파악해야 한다. 이 기능은 이후에 이루어지는 모든 보존활동의 근원이 되므로, 이에 대한 정확한 이해 없이 디지털객체의 장기보존과 접근가능성을 유지하는 것은 불가능하다.

전자기록의 기술적 특성 관리 과정은 식별-검증-특성추출 과정으로 이루어진다. 식별(identification)은 그 객체의 포맷이 무엇인지를 확인하는 것이고, 검증(validation)은 그 객체의 포맷이 그 포맷 명세에 맞게 형성되었는지, 그리고 그것이 타당한지를 검증하는 단계이다(Brown 2005, p.2). 실질적으로 검증은 가장 확실한 식별을 제공하지만, 기술적으로는 그 과정이 매우 복잡하다. 디지털 객체의 포맷을 식별하고 확인하는 것은 그 객체로의 접근과 관리를 위해 필수불가결한 요소로, 이에 관한 지식 없이 디지털객체를 이용하거나 보존할 수는 없다.

마지막 특성추출(property extraction)은 디지털 객체의 장기보존에 중요한 속성들을 선택하는 과정이다. 이런 과정을 통해 선별된 디지털기록의 특성은 포맷등록시스템을 통해 지속적으로 유지·관리된다.

2.2.2 장기보존계획 지원

모든 보존서비스의 핵심을 이루는 보존계획은 기술 변화를 모니터링하고, 그 기술 변화가 리파지토리에 저장된 전자기록에 미칠 잠재적 영향력을 확인하는 기능을 수행한다. 그리고 이러한 기술 변화의 충격을 완화시켜줄 수 있는 보존전략 개발을 지원하며, 포맷이 노후화되는 시기를 알려줘 보존활동의 시발점으로 작용한다. 이 과정은 위험평가 → 기술감시 → 영향

력평가 → 보존계획 수립 → 실행으로 구성된다.

디지털객체가 리퍼지토리로 입수되는 시점에 실시되는 위험평가는 그 객체로의 지속적 접근에 필요한 기본 요건을 평가기준으로 삼는다. 기본 요건은 일반위험요인과 특별위험요인으로 구분할 수 있으며, 이 결과는 보존활동의 우선순위를 결정하는데 이용된다.

기술 감시는 포맷등록시스템 콘텐츠에 업데이트 될 기술변화를 계속적으로 모니터링 하는 과정으로, 이런 업데이트를 통해 위험기준을 바꿀 수도 있다.

영향력 평가는 리퍼지토리에 저장된 객체에 미치는 실질적인 영향력을 예측하기 위해 위험평가와 기술 감시 결과를 분석하는 과정이다(Brown 2005, p.5).

보존계획의 마지막 단계는 세부적인 보존 활동을 결정하고 실행하는 것이다. 마이그레이션 기법의 경우, 노후화 위험에 처한 소스포맷과 계속 실행 가능한 타겟포맷 간의 직접적인 변환뿐만 아니라 많은 매개포맷을 경유할 경우가 있다(Abrams 2005, p.6). 이때 이 복잡한 변환 순서를 찾는데 포맷등록시스템이 이용된다. 포맷등록시스템은 복잡한 보존행위 과정을 문서화하여 새로운 포맷으로의 마이그레이션 경로 생성을 지원해 준다. 그리고 이런 보존 행위의 문서화는 전자기록의 지속적인 진본성 유지에 있어서 중요한 역할을 수행한다(Brown 2004, p.6).

3. 해외 사례 분석

3.1 영국 PRONOM Service

3.1.1 PRONOM 개발 배경과 목적

2002년 3월에 개발된 PRONOM³⁾은 2004년 2월, 웹에서 이용 가능한 PRONOM3을 거쳐, 지금 현재 TNA 홈페이지에서 이용되고 있는 것은 PRONOM4이다.

PRONOM은 영구저장을 위해 TNA로 이관되는 기록의 장기보존을 지원하는 프로젝트로, 다음과 같은 목적을 갖는다(Brown 2004, p.6).

- 파일포맷, 소프트웨어제품, 운영시스템, 그리고 하드웨어 구성요소 등, 전자기록의 기술의존성에 대한 정보 저장소 기능
- 파일포맷의 식별과 검증, 메타데이터 추출과 마이그레이션을 위한 자동화된 툴 지원
- 기술 노후화나 지원되지 않는 제품에 대한 자동 경고 시스템을 통한 기술감시지원
- 포맷 간 마이그레이션 경로의 자동 생성을 통한 마이그레이션 지원
- 데이터의 교환과 자동화 서비스를 위해 다른 시스템과의 상호운용성 지원
- 월드와이드웹에 접근하기 쉬운 제품을 개발

3.1.2 PRONOM 정보요소

(1) 파일포맷(File Format)

디지털객체 부호화에 사용될 파일포맷은 기술적 구성요소 중 가장 복잡한 객체로, 다양한 기술체계에 따라 분류된 다른 포맷들과 그룹핑될 수 있다. 이 객체를 구성하는 필수요소로는

3) 영국의 TNA가 개발한 PRONOM은 전자기록의 기술의존성에 대한 정보를 관리하기 위한 애플리케이션으로, 데이터 파일포맷과 그들을 지원하는 소프트웨어 제품에 관한 온라인 정보시스템이다.

System ID와 포맷명, 그 포맷의 주요 특징을 기술하는 기술요소가 있으며, 그 외 포맷이 할당된 조직그룹, 포맷 유형(표 11 참조), 포맷이 텍스트 기반인지 바이너리 기반인지를 나타내는 오리엔테이션, 바이트오더, 포맷 식별에 필요한 내·외부 서명, 압축 유형 및 포맷 공개수준, 포맷 발행일 및 포맷 지원 종료일, 포맷 개발자, 포맷 지원자, 지적재산권, 포맷참조파일

등의 선택요소가 있다.

(2) 소프트웨어 구성요소(Software Component)

이 정보는 특별한 파일포맷을 실행하는데 필요한 운영시스템, 또는 특별한 하드웨어나 저장매체 지원을 위해 필요한 소프트웨어 구성요소이다. 이 객체는 소프트웨어명과 그 소프트

<표 1> 파일포맷 구성요소

요소명	유형	필수여부	내용
System ID	System-dependent	필수	내부시스템에서 생성된 ID
External Identifier	Identifier	선택	포맷 외부 식별자
Name	Name	필수	포맷명
Alias	Name	선택	다른 포맷명
Family	Family	선택	포맷이 속한 조직그룹(Family group)
Format Type	Classification	선택	일반적인 포맷유형 [스킴 제공: 표 11 참조]
Description	String	필수	포맷의 주요특징 기술
Orientation	String	선택	'Text'인지 'binary'기반인지
Byte Order	String	선택	포맷의 바이트오더
Internal Signature	Internal Signature	선택	파일포맷 식별을 위한 내부서명
External Signature	External Signature	선택	파일포맷 식별을 위한 외부서명
Compression Type	Compression Type	선택	파일포맷에 사용된 압축 방법
Character Encoding	Character map	선택	파일포맷에 사용된 문자 부호화
Format Disclosure	String	선택	포맷규정(specification)에 따른 공개 수준 - 지정값: • 전체공개 • 부분공개 • 비공개
Release Date	Date	선택	포맷이 발행된 날짜
Withdrawn Date	Date	선택	포맷 지원일이나 지원 종료일
Developer	Actor	선택	포맷 개발자
Support	Actor	선택	현재 포맷을 지원/유지하는 사람
Documentation	Documentation	선택	포맷에 관한 문서
IPR	IPR	선택	포맷 지적재산권
Note	String	선택	포맷 정보주기
Source	Actor	필수	정보를 제공한 행위자객체
Source Date	Date/Time	필수	그 정보가 PRONOM에 추가된 날짜/시간
Last Updated	Date/Time	필수	그 정보가 마지막으로 업데이트된 날짜/시간
Provenance Note	String	선택	출처정보 주기
Reference File	Reference File	선택	포맷 참조파일

〈표 2〉 소프트웨어 구성요소

요소명	유형	필수여부	내용
System ID	System-dependent	필수	내부시스템에서 생성된 ID
External Identifier	Identifier	선택	소프트웨어 외부 식별자
Name	Name	필수	소프트웨어 명
Alias	Name	선택	다른 소프트웨어 명
Family	Family	선택	S/W가 속한 조직그룹(Family group)
Software Type	Classification	선택	일반적인 S/W 유형
Description	String	필수	S/W 주요특징 기술
Service Pack Level	String	선택	S/W에 적용된 서비스 팩이나 패치 수준
Default File Format	File Format	선택	S/W가 지원하는 기본파일포맷
Software Requirement	Software Requirement	선택	S/W지원에 필요한 운영시스템과 기타S/W 요건
Hardware Requirement	Hardware Requirement	선택	S/W지원에 필요한 H/W 요건
Media Format	Storage Media	선택	S/W 저장매체 유형
Language	String	선택	S/W에 의해 지원된 언어
Release Date	Date	선택	S/W가 발행일
Withdrawn Date	Date	선택	S/W 지원일이나 지원 종료일
Developer	Actor	선택	S/W 개발자
Support	Actor	선택	현재 S/W를 지원/유지하는 사람
Documentation	Documentation	선택	S/W에 관한 문서
IPR	IPR	선택	S/W 지적재산권
Note	String	선택	S/W 정보주기
Source	Actor	필수	정보를 제공한 행위자 객체
Source Date	Date/Time	필수	그 정보가 PRONOM에 추가된 날짜/시간
Last Updated	Date/Time	필수	그 정보가 마지막으로 업데이트된 날짜/시간
Provenance Note	String	선택	출처정보 주기
Image	Image	선택	S/W제품 이미지

웨어의 주요 특징 기술을 필수요소로 하고 있으며, 그 외 소프트웨어의 기타 다른 명칭, 소프트웨어가 할당된 조직그룹, 소프트웨어 유형, 그 소프트웨어에 의해 지원되는 기본파일포맷(default file format), 소프트웨어 지원에 필요한 운영시스템 및 소프트웨어/하드웨어 요건, 저장매체, 언어, 소프트웨어 발행일, 소프트웨어 지원일이나 지원 종료일, 개발자, 지원자, 지적재산권, 제품 이미지 등의 선택요소로 구성되어 있다.

(3) 하드웨어 구성요소(Hardware Component)
특별한 소프트웨어 구성요소나 저장매체 지원에 필요한 요소로, 하드웨어명과 주요 특징 기술, 하드웨어 유형이 필수요소이다. 선택요소로는 외부식별자, 기타 다른 하드웨어 명칭, 하드웨어가 할당된 조직그룹, 발행일, 하드웨어 지원일이나 지원 종료일, 하드웨어 지원에 필요한 소프트웨어와 기타 다른 하드웨어 요건, 개발자와 지원자, 하드웨어에 관한 문서, 지적재산권, 제품이미지 항목이 있다.

〈표 3〉 하드웨어 구성요소

요소명	유형	필수여부	내용
System ID	System-dependent	필수	내부시스템에서 생성된 ID
External Identifier	Identifier	선택	H/W 외부 식별자
Name	Name	필수	하드웨어명
Alias	Name	선택	다른 H/W 명
Family	Family	선택	H/W가 속한 조직그룹(Family group)
Description	String	필수	H/W 주요특징 기술
Hardware Type	Classification	필수	H/W 구성요소 유형 - 지정값: • CPU • Motherboard • RAM • Hard disk drive • CD-ROM drive • DVD drive • Floppy disk drive • Zip drive • Graphics card • Sound card • Dongle • Network card • Modem
Release Date	Date	선택	H/W가 발행된 날짜
Withdrawn Date	Date	선택	H/W 지원일이나 지원 종료일
Software Requirement	Software Requirement	선택	H/W지원에 필요한 S/W
Hardware Requirement	Hardware Requirement	선택	H/W지원에 필요한 다른 H/W 요건
Developer	Actor	선택	H/W 개발자
Support	Actor	선택	현재 H/W를 지원/유지하는 사람
Documentation	Documentation	선택	H/W에 관한 문서
IPR	IPR	선택	H/W 지적재산권
Note	String	선택	H/W 정보주기
Source	Actor	필수	정보를 제공한 행위자 객체
Source Date	Date/Time	필수	그 정보가 PRONOM에 추가된 날짜/시간
Last Updated	Date/Time	필수	그 정보가 마지막으로 업데이트 된 날짜/시간
Provenance Note	String	선택	출처정보 주기
Image	Image	선택	H/W제품 이미지

(4) 저장매체(Storage Media)

이 요소는 디지털객체 저장에 사용될 물리적 저장매체의 특별한 유형을 나타낸다. 필수요소로는 매체명과 매체 특성 기술, 매체 포맷 유형, 쓰기 유형, 비압축 데이터 저장 용량 등이 있다. 그 밖에 쓰기 속도, 쓰기방지, 에러수정, 데이터

전송율, 매체접근 유형, 압축용량, 이용 가능한 데이터 저장 면수와 층수, 매체의 물리적 크기, 매체 접근에 필요한 소프트웨어/하드웨어 요건, 매체의 예상 수명, 매체의 보자력을, 권장할 만한 저장환경조건, 특별저장요건과 처리요건 주기, 매체이미지 등의 선택요소가 있다.

〈표 4〉 저장매체 구성요소

요소명	유형	필수여부	내용
System ID	System-dependent	필수	내부시스템에서 생성된 ID
External Identifier	Identifier	선택	매체에 적용할 외부 식별자
Name	Name	필수	매체명
Alias	Name	선택	다른 매체명
Family	Family	선택	매체가 속한 조직그룹(Family group)
Description	String	필수	매체 주요특징 기술
Media Format Type	Classification	필수	일반적인 매체 유형 - 지정값: • Punched paper tape • Punched card • Flexible magnetic disk • Hard disk drive • Magnetic tape cartridge • Magnetic tape reel • Optical disc • Magneto-optical disk • Solid state
Write Type	String	필수	지원되는 쓰기 유형 - 지정값: • Read only • Write once • Rewritable
Write Speed	String	선택	최소 쓰기 속도
Write Protection	String	선택	쓰기방지메커니즘 기술
Error Correction	String	선택	에러수정메커니즘 기술
Data Transfer Rate	String	선택	최소 데이터 전송율
Media Access Type	String	선택	데이터 접근 방법 - 지정값: • Random • Serial
Uncompressed Capacity	String	필수	비압축 데이터 저장 용량
Compressed Capacity	String	선택	압축 데이터 저장 용량
No. of Sides	Integer	선택	이용 가능한 데이터저장면(sides) 수
No. of Layers	Integer	선택	이용 가능한 데이터저장층(layers)수
Dimensions	String	선택	매체와 관련된 물리적 크기
Software Requirement	Software Requirement	선택	매체 접근에 필요한 소프트웨어 요건
Hardware Requirement	Hardware Requirement	선택	매체 접근에 필요한 하드웨어 요건
Release Date	Date	선택	매체 발행 날짜
Withdrawn Date	Date	선택	매체 지원일이나 지원 종료일
Developer	Actor	선택	매체 개발자
Support	Actor	선택	현재 매체를 지원/유지하는 사람
Documentation	Documentation	선택	매체에 관한 문서
IPR	IPR	선택	매체 지적재산권

Note	String	선택	매체 정보주기
Longevity Rating	String	선택	매체의 예상 수명
Longevity Documentation	Documentation	선택	수명을 지원 문서
Coercivity Rating	Integer	선택 (자기매체에만 적용)	에르스렛에서의 매체의 보자력율
Storage Conditions	String	선택	권장할만한 저장환경조건
Storage Notes	String	선택	특별 저장요건 기술
Handling Notes	String	선택	특별 처리요건 기술
Storage Documentation	Documentation	선택	저장과 처리를 지원하는 문서
Source	Actor	필수	정보를 제공한 행위자 객체
Source Date	Date/Time	필수	그 정보가 PRONOM에 추가된 날짜/시간
Last Updated	Date/Time	필수	그 정보가 마지막으로 업데이트된 날짜/시간
Provenance Note	String	선택	출처정보 주기
Image	Image	선택	매체이미지

3.2 GDFR 프로젝트

3.2.1 GDFR 개발 배경과 목적

2002년 Harvard University Library와 MIT가 실시한 GDFR프로젝트는 디지털 보존의 분산 프로그램 수립을 가능하게 하여, 개별 보존 프로그램에 드는 시간과 노력을 절약해 준다. 또한 좀 더 많은 보존프로그램이, 좀 더 많은 포맷에 대해, 보다 정교한 정보를 가질 수 있는 방법을 제공하며, 기관 전반을 아우르는 분산 보존서비스 환경 개발에 상당한 기여를 할 것이다.

이 프로젝트의 목적은 보존기능에 필요한 필수적인 포맷 재현 정보를 관리해주는 정보원을 제공하는 것으로, 좀 더 넓은 의미에서 그 기능을 살펴보면,

- 소중한 기술정보를 공동 관리하고 공유하기 위한 공통 메커니즘을 제공하여 기술정보 수집에 드는 중복된 노력을 줄여준다.
- 보존활동과 관련된 모든 이들에게 기술정보의 접근 채널을 제공한다.

- 디지털 포맷에 대한 상세 정보를 필요로 하는 부가서비스의 토대를 제공한다.

GDFR은 NDIIPP⁴⁾와 마찬가지로 ‘디지털보존아키텍처’와 ‘디지털 보존 네트워크’ 측면에서 보존인프라를 지원한다. 보존아키텍처 측면에서는 효과적인 협력보존에 필요한 포맷재현정보를 제공하는 제공자가 되고, 보존네트워크 면에서는 디지털자산의 영구보존에 관심이 있는 모든 기관들이 그들의 업무에 GDFR을 이용하게 될 것이다(Abrams & Flecker 2005, p.1).

3.2.2 GDFR 정보요소

(1) 포맷등록시스템 속성(Registry properties)
이 요소는 시스템 자체에 대한 속성으로, 포맷등록시스템명, 버전, 포맷등록시스템 수립일, 외부 포맷등록시스템으로 구성된다(표 5 참조).

(2) 포맷속성(Format properties)
포맷속성 정보는 식별자, 포맷개발자, 포맷

4) NDIIPP(The National Digital Information Infrastructure Preservation Program)는 2000년 디지털문화유산의 손실을 막기 위해 미국 의회도서관과 OCLC가 시작한 “국가디지털정보기반구축보존프로그램”이다.

상태정보, 정보출처와 최종 업데이트일 등의 필수요소와 포맷에 대한 다른 식별자, 지적소유권자, 유지기관, 분류체계, 관계, 포맷명세서, 포맷에 대한 기술정보의 공개수준, 서명, 특

정포맷을 처리할 수 있는 시스템이나 서비스·애플리케이션 명, 기술적 리뷰 등의 선택요소로 구성된다(표 6 참조).

〈표 5〉 포맷등록시스템 속성 요소

요소명	필수여부	정 의	하위요소
Name	필수	포맷등록시스템 명칭	
Version	필수	포맷등록시스템 버전	
Date	필수	포맷등록시스템 수립일	
External	선택	외부 포맷등록시스템 명	
Note	선택	정보주기	

〈표 6〉 포맷 속성 요소

요소명	필수여부	정 의	하위요소
Identifier	필수	식별자	
Alias	선택	다른 식별자	
Author	필수	개발자	
Owner	선택	지적소유권을 갖고 있는 기관이나 사람	
Maintainer	선택	유지 기관	
Classification	선택	분류체계 [스킴 제공; 표 11 참조]	
Relationship	선택	다른 포맷과의 관계	
Specification	선택	포맷명세 문서	서명
			유형
			저자
			출판사
			출판일
			식별자
			접근가능성
Disclosure	선택	포맷에 대한 기술정보의 공개 수준 - 지정값: • 전체공개 • 부분공개 • 비공개	주기
Signature	선택	포맷 식별을 위한 서명정보	내부서명
			외부서명
Tool	선택	특정포맷을 처리할 수 있는 서비스나 애플리케이션	애플리케이션 명
			버전
			발행일
			벤더
			프로세스

			서비스
			하드웨어 의존성
			소프트웨어 의존성
Status	필수	포맷의 상태 정보	현용(active)
			중료(withdrawn)
			알 수 없음
Provenance	필수	등록, 업데이트, 삭제와 같은 이벤트 출처	출처기관
			이벤트 날짜
			이벤트 유형
			주기
Technical Review	선택	기술적 리뷰 - 지정값: • 전체리뷰 • 부분리뷰 • 기타	
Note	선택	정보주기	
Last Modified	필수	최종 수정일	

3.3 시사점

영국의 PRONOM시스템과 미국의 GDFR 프로젝트를 통해 포맷등록시스템의 기능과 정보요소에 대하여 살펴보았다.

집중형 레지스트리의 대표 유형인 PRONOM 시스템은 일괄적인 정보 수집과 관리, 정책 수행 및 보다 확장된 서비스로의 발전이 용이하며, 더욱이 기록관리 기관인 영국의 TNA가 주관이 되어 진행되었다는 점, 그리고 지금 현재 활발하게 운영되고 있다는 점에서 그 가치를 더한다. 그리고 본 시스템을 구성하는 정보요소, 즉 포맷정보, 소프트웨어 정보, 하드웨어정보, 저장매체 정보들은 서로서로를 상호 연계해 줄 메타데이터 요소를 필요로 하며, 이런 메타데이터요소는 이용자계에 다양한 접근점으로 제공된다.

반면, GDFR은 디지털포맷 재현정보의 저장, 발견, 배달 등에 관한 분산서비스를 제공한

다. 그리고 GDFR 프로토콜을 준수하는 여러 기관과의 상호협력을 지원해 줄 종합 레지스트리 통신망을 형성하여, 이 기관들과의 정보공유와 상호교환, 분산보존 및 분산접근 등의 기능도 제공한다. 각각의 협력기관은 각 기관들이 소장한 정보뿐만 아니라 다른 협력기관들이 소장하고 있는 포맷정보도 원하는 때에 동시에 이용할 수 있다. 무엇보다도 GDFR의 가장 큰 의의는 디지털 보존에 관심이 있는 모든 기관을 아우를 수 있는 이론적 분산시스템 구조를 제안한다는 데 있다. 이는 그동안 각각의 보존기관에서 개별적으로 수행되었던 디지털 자원의 장기보존에 대한 중복된 노력과 비용의 상당부분을 절감해주는 효과가 있다.

이런 시사점을 토대로 포맷등록시스템 개발시 고려할 사항은 다음과 같다.

첫째, 관리적 측면에서 포맷등록시스템은 기관 차원에서 구축하는 집중형과 인터넷을 통해 원거리 접근이 가능한 네트워크형 방식으로 구

분할 수 있다. 따라서 전자기록의 장기보존에 관심이 있는 기관들은 조직의 성격과 보존목적, 그리고 보존환경에 적합한 관리구조를 선택하여 구축하여야 한다.

둘째, 포맷등록시스템 구축을 위한 기본요건으로 기술메타데이터요소와 포맷분류체계가 개발되어야 한다. 이 시스템의 일차적인 목적은 디지털 객체의 기술적 특성을 관리하고 유지하는데 있다. 따라서 포맷정보, 소프트웨어와 하드웨어정보, 매체정보로 구성되는 주요 기술정보메타데이터와 포맷정보의 체계적인 관리와 효과적인 이용을 지원하는 포맷분류체계의 개발은 포맷등록시스템을 형성하는 핵심요건이라 할 수 있다.

셋째, 기능적 측면에서 포맷등록시스템은 권위 있고 믿을만한 기술정보를 지속적으로 등록·관리하고 유지하여, 시스템 자체에 대한 신뢰성을 높여야 한다. 이를 위해서는 ① 새로이 생성되거나 갱신되는 기술정보를 계속적으로 업데이트해야 한다. ② 소프트웨어 관련 업체와 접촉하여 다양한 포맷정보와 기술정보를 적극적으로 수집해야 한다. ③ 기술변화에 대

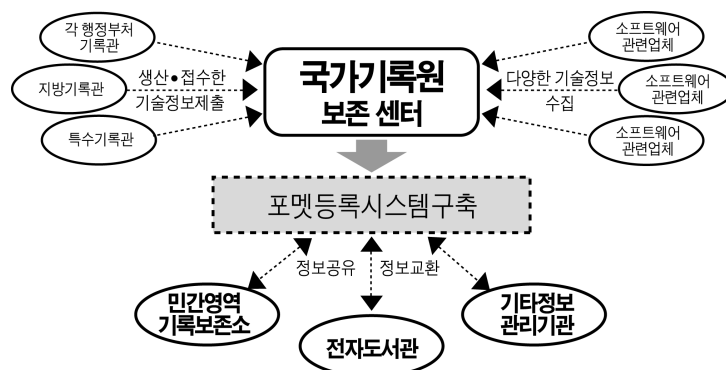
한 지속적인 모니터링과 위험평가서비스를 통해 기술노후화에 시의 적절하게 대응해야 한다.

또 다른 기능면에서 포맷등록시스템은 저장·관리하고 있는 우수한 기술정보를 다양한 보존계획이나 이용서비스에 제공하여야 한다. 기술정보의 유지 및 관리의 기본 목적은 단순한 보관이 아니라 기록의 활용성을 높이는 데 있다. 따라서 다양한 보존전략과 이용자의 요구에 적합한 DIP 생성을 지원할 수 있어야 한다.

4. 포맷등록시스템 개발 방안

4.1 포맷등록시스템 관리체계

포맷등록시스템의 관리방식은 기관 차원에서 하나의 데이터베이스를 구축하여 관리하는 방식(집중형)과 인터넷을 통하여 원거리 접근을 허용하는 네트워크형(분산형) 방식으로 구분할 수 있다. 우리나라는 아직 포맷등록시스템에 대한 인식조차 부재한 상태이다. 따라서 본 연구에서는 정보의 일괄 수집 및 관리가 용이한 집중형



〈그림 1〉 포맷등록시스템 관리체계

관리 방식을 중심으로 설계하였다. 그리고 점점 유사기관과의 정보교환과 정보공유를 지원할 수 있는 분산형 모델로의 확장성을 제시하였다.

구축 초기단계에서는 기록관리 영역을 중심으로 하는 집중형 등록시스템을 설계하였다. 집중형 모델은 무엇보다도 구축 기관의 역할이 중요하다. 본 논문에서는 기록관리 전반에 걸쳐 정책적·실무적 책임 기관이며, 기록관리 영역 내외적으로 중요한 위치를 차지하고 있는 국가기록원을 구축기관으로 선정하였다. 그 역할은 다음과 같다.

첫째, 각 행정부처의 기록관과 처리과, 특수 기록관, 그리고 지방기록관 등을 대상으로 그들이 생산·접수한 기록의 포맷 및 관련 소프트웨어 정보를 수집한다.

둘째, 소프트웨어 관련 업체와 접촉하여 다양한 포맷정보와 관련 기술정보를 적극적으로 수집하여야 한다.

셋째, 이렇게 수집된 기술정보를 지속적으로 유지, 관리하고 기술감시 및 위험평가 서비스를 제공하여 신뢰할 수 있는 포맷등록시스템을 구축하여야 한다.

넷째, 신뢰할 수 있는 포맷등록시스템에서 등록·관리되고 있는 우수한 기술정보를 다양한 보존전략이나 이용서비스에 제공하여야 한다.

이렇게 구축된 집중형 시스템의 안정화를 기반으로, 유사기관과의 정보 공유와 상호교환을 위해 GDFR 분산형 모델로의 확장이 가능하다. 이는 국가기록원과 디지털정보의 장기보존에 관심이 있는 다른 기관들을 아우를 수 있는 제도이다. 그러나 분산형 모델로의 확장에 앞서 다음과 같은 사항에 주의하여야 한다.

- 국가기록원과 이 프로젝트에 참여할 기관

들의 경험과 적극적인 관심이 필요하다.

- 분산모델의 설계와 실행, 테스트 단계에서 해당분야의 전문가나 기관의 참여가 유도되어야 한다.
- PRONOM, GDFR과 같은 유사 프로젝트의 지식과 경험, 결과를 분석, 활용한다.
- 협력기관 간의 효과적인 정보교환과 동시 이용가능성을 제공하기 위해서는 협력 기관들 간에 준수해야 할 네트워크프로토콜이 개발되어야 한다.
- 이런 종합적인 네트워크망을 관리해 줄 관리기관과 운영비용 문제가 해결되어야 한다.

4.2 포맷등록시스템 정보요소

PRONOM정보요소는 구체적이고 상세하여 향후 기술정보의 기술(description)에 있어서 그 확장성이 좋은 반면, GDFR정보요소는 간결하고 포괄적이며 타기관과의 정보교환을 고려하여 개발되었다는 이점이 있다.

본 연구에서는 포맷등록시스템의 시작 단계에 있는 우리나라 상황을 고려하여 다음과 같은 방법으로 기술요소를 제안하였다. 첫째, PRONOM과 GDFR정보요소를 참고하여 시스템 구축에 기본이 되는 기술요소를 중심으로 기술메타데이터 요소를 제안하였다. 둘째, 이 기술메타데이터 요소는 사용하기 간단하도록 구성될 것이다. 셋째, 필요할 경우 하위요소를 추가할 수 있도록 확장가능하게 설계할 것이다.

4.2.1 포맷등록시스템 정보

포맷등록시스템 정보란 포맷등록시스템 자체에 관한 속성으로, 포맷등록시스템 명칭과

〈표 7〉 포맷등록시스템 속성 요소

요소명	필수여부	정 의	하위요소
Name	필수	포맷등록시스템 명	
Version	필수	포맷등록시스템 버전	
Date	필수	수립일	
External	선택	외부 포맷등록시스템 명	
Note	선택	포맷등록시스템 정보주기	

버전, 수립일, 외부관련 포맷등록시스템,⁵⁾ 기타 정보주기로 구성된다(표 7 참조).

4.2.2 포맷정보

(1) 일반 속성

일반속성정보는 포맷에 관한 일반적인 개요를 제공한다(표 8 참조).

System ID는 자원이 시스템 안에 등록됨과 동시에 시스템 내부에서 자동으로 생성되는 일

종의 관리번호이다. 식별자는 포맷등록시스템 내에서 그 포맷을 유일하게 구별할 수 있도록 해준다. 그리고 디지털객체의 상호교환과 관리, 정보공유의 근간이 된다. 포맷명은 포맷을 표현하는 메타데이터로 접근을 위한 주요 접근점이 된다. 따라서 일관되고 공식적인 명칭을 사용하는 것이 좋다. 동일 포맷인 경우 버전 항목을 사용하여 보다 정확한 기술정보를 제공할 수 있다. 기술(Description)은 포맷의 주요 특징에 대

〈표 8〉 포맷 일반 속성 요소

요소명	필수여부	정 의	하위요소
System ID	필수	시스템 내부에서 자동 생성	
Identifier	선택	식별자	
Name	필수	포맷명	
Other Name	선택	기타 다른 포맷명	
Version	선택	포맷 버전	
Description	필수	포맷의 주요 특징 기술	
Classification	필수	포맷유형 분류체계[스킴 제공; 표 12 참고]	
Status	필수	포맷 상태정보[지정값: 현용, 위험상태, 노후화, 기타]	
IPR	선택	지적재산권	유형
			저작권
			특허
			기타
			소유권자
			이름
			주소
전화번호			
이메일			
유효기간			
주기			

5) 외부포맷등록시스템이란 포맷등록시스템 서비스가 분산형 모델로 확장될 경우 종합 네트워크망을 형성하는 다른 보존기관이나 PRONOM과 같은 외부 포맷등록시스템을 가리킨다.

Developer	필수	포맷 개발자	이름
			주소
			전화번호
			이메일
			주기
Support	선택	현재 포맷을 지원/유지하는 사람/기관	이름
			주소
			전화번호
			이메일
			유효기간
Release Date	선택	포맷 발행일	
Withdraw Date	선택	포맷 지원 종료일	
Note	선택	정보주기	

한 서술식 설명으로 간략하고 명확하게 기술하는 것이 좋다. Classification은 포맷에 대한 분류 체계로 포맷등록시스템에 등록되는 모든 포맷들은 <표 12>에서 제시한 포맷분류체계에 따라 하나의 포맷 유형에 할당되게 된다. 상태요소는 그 포맷이 여전히 이용 가능한 상태인지, 노후화 혹은 노후화의 위험에 놓인 상태인지를 제공하는 정보로, 현용(active), 위험상태(at risk), 노후화(obsolete), 기타로 구분된다.

포맷을 개발한 특정 사람이나 조직에 대한 개발자정보와 현재 포맷을 지원하거나 유지하는 사람이나 기관에 대한 지원자요소는 포맷자원 검색 시 하나의 접근점으로 활용되는 벤더 정보의 주요 정보원이 된다. 저작권, 특허권, 기타로 구분되는 지적재산권 요소는 포맷정보의 공유와 활용에 있어서 중요한 정보이다. 날짜 요소를 이용하여 포맷이 발행일과 포맷 지원이 종료되는 날짜를 표현할 수 있다.

(2) 기술 속성

기술속성정보란 포맷이 갖고 있는 구조론적·의미론적 특성을 나타내는 정보이다(표 9 참조).

공개수준은 포맷에 대한 기술정보의 공개 정도를 나타내는 요소로, 전체공개와 부분공개, 비공개로 나누어진다. 포맷간의 상호 관련성을 보여주는 관계요소는 동등관계, 이전·이후버전, 상·하위관계, 기타 등으로 그 관계를 표시한다. 이 요소를 통해 현재 기술되는 포맷정보와 관련 있는 다른 포맷정보를 서로 연계하여 종합적으로 이용할 수 있다.

문서요소는 포맷과 관련된 문서정보로 서명, 문서유형, 저자, 출판사, 출판일, 식별자, 이용가능성 등의 하위요소를 갖는다. 문서유형은 다시 기사, 서신, 매뉴얼, 보고서, 표준, 학위논문, 웹으로 구분된다. 포맷 관련 문서의 이용가능성은 접근성을 나타내는 정보로, 이용가능·제한적 이용·이용불가로 나누어진다.

Orientation 정보는 그 포맷이 텍스트 기반인지 바이너리 기반인지를 기술하는 요소이고, 포맷 식별에 이용되는 서명정보는 외부서명과 내부서명으로 구분된다. 외부서명은 객체의 비트스트림 밖에 있는 포맷 식별요소이다. 대표적인 유형으로는 파일확장자가 있으며, 이용자는 파일 확장자를 사용하여 포맷정보를 검색할

〈표 9〉 포맷 기술 속성 요소

요소명	필수여부	정 의	하위요소
Disclosure level	선택	포맷의 기술정보에 대한 공개 수준 [지정값: 전체공개, 부분공개, 비공개]	
Related Format	선택	관련 포맷	동등 관계
			이전/이후 버전
			상위/하위 유형
			기타
Documentation	선택	포맷 관련 문서	서명
			유형
			지자
			출판사
			출판일
			식별자
			이용가능성
Orientation	선택	text기반/binary기반 표시	text
			binary
External Signature	선택	포맷 식별을 위한 외부서명	
Internal Signature	선택	포맷 식별을 위한 내부서명	이름
			기술(description)
			바이트시퀀스
Note	선택	정보주기	

수 있다. 내부서명은 객체의 비트스트림 내에 존재하는 포맷 식별요소로, 포맷마다 갖고 있는 고유한 내적 구조 특성을 이용하여 포맷을 식별하는 것이다.

(3) 관리 속성

포맷등록시스템에 등록된 포맷을 관리하는데 이용되는 관리 속성은 현재 기술되는 자원의 출처를 제공하는 정보원 요소와 최초 등록일, 최종 업데이트 일, 포맷관리 상태, 주기 정보로 구성된다. 포맷관리 상태를 나타내는 Type 정보는 최초등록, 업데이트, 노후화선언, 삭제 등 그 포맷에 일어난 이벤트 이력을 보여주는 요소이다 (표 10 참조).

(4) 포맷분류체계

〈표 11〉과 같이 PRONOM 분류체계는 분류 계층을 '대-중-소'로 나누었을 때 '대'에 해당하는 상위 개념만을 제시하고 있는 반면, GDFR 분류체계는 '대-중-소'로 이어지는 포맷의 분류계층을 상세히 제공하고 있다. 이에 본 연구에서는 PRONOM과 GDFR 분류체계를 참고하여 〈표 12〉와 같이, '대' 그룹은 Text, Image, Video, Audio, Application 등 총 5가지로 구분하였다. 그리고 각각의 포맷유형은 하위요소를 두어 새로운 포맷유형이 개발될 때마다 적절한 그룹에 추가할 수 있도록 설계하였다.

〈표 10〉 포맷 관리 속성 요소

요소명	필수여부	정 의	하위요소
Source	필수	정보 제공자	이름
			주소
			전화번호
			이메일
			주소
Source Date	필수	포맷등록시스템 등록일	
Last Updated	필수	최종 업데이트 일	
Type	필수	포맷 관리 상태 표시 [지정값: 최초등록, 업데이트, 노후화, 삭제, 기타]	
Note	선택	정보주기	

〈표 11〉 PRONOM과 GDFR 포맷분류체계 비교

PRONOM	Global Digital Format Registry(GDFR)		
	대	중	소
Text(Unstructured)	Text	Structured	CSV
Text(Structured)		Mark-up Language	HTML, SGML, XML...
Text(Mark-up)		Programming Language	C++, C#...
Text(Word-processed)		Message	SMTP, NSENET
Image(Raster)	Image	Still	SVG, VRML, GIF, TIFF, PDF, PostScript....
Image(Vector)		Motion	AVI, MPEG, QuickTime
Audio	Audio		AIFF, MP3, Real, WAV
Video	Application	CAD	DXF, IGES
Database		Communication	MIME
Spreadsheet		Database	SQL DDL, MySQL, Oracle...
Presentation		Executable	ELF, EXE, Java byte code
GIS		GIS	GeoTIFF, STDS
Page Description		Presentation	Powerpoint
Email		Spreadsheet	Excel
		Word processing	Word
		Transformation	Compression
			Container
		Transfer	Base64, BinHex

〈표 12〉 포맷분류체계

Format Classification		
대	중	소
Text	Unstructured	
	Structured	CSV
	Mark-up Language	HTML, SGML, XML...
	Programming Language	C++, C#...
Image	Vector	SVG, VRML
	Raster	GIF, JFIF, PCD, TIFF, JPEG
	Page Description	PDF, PostScript, Quark Xpress
Video		AVI, MPEG, QuickTime
Audio		AUFF, MP3, Real, WAV
Application	CAD	DXF, IGES
	Database	SQL DDL, MySQL, Oracle...
	GIS	GeoTIFF, STDS
	Presentation	Powerpoint
	Spreadsheet	Excel
	Word processing	한글, Word, WordPerfect

4.2.3 소프트웨어 정보

소프트웨어 정보는 특정 파일 포맷을 처리하거나 지원하는데 필요한 요소이다. 그 구성요소로는 System ID, 식별자, 소프트웨어명과 버전, 소프트웨어 지원에 필요한 운영시스템과 소프트웨어·하드웨어 요건, 소프트웨어에 의해 지원되는 프로세스유형, 언어, 소장여부와 위치, 발행일 및 지원 종료일, 개발자, 지원자, 판매자, 지적재산권, 관련문서, 이미지, 그리고 기타 정보주기가 있다.

소프트웨어명 요소는 포맷명과 마찬가지로 접근을 위한 주요 접근점이 된다. 따라서 일관되고 공식적인 명칭을 사용하는 것이 좋다. 기술(Description)요소 역시 소프트웨어의 주요 특징에 대한 서술식 설명으로, 간략하고 명확하게 기술하는 것이 좋다. 소프트웨어 정보가

가지는 소프트웨어요건과 하드웨어요건은 특정 소프트웨어를 지원하는데 필요한 정보이다. 프로세스 포맷(Process format)요소는 특정 소프트웨어에 의해 처리 가능한 포맷유형을 나타내는 정보로 다시 create, render, validate, 기타로 구분된다.

Media요소는 소프트웨어가 저장된 물리적 매체정보로, 저장기술에 기반하여 magnetic, optical로 구분된다. 한국어나 영어 등 그 소프트웨어에 의해 지원되는 언어정보는 언어요소에 기술되고, 이런 소프트웨어의 소장여부 및 소장위치는 위치 정보에 기술된다. 소프트웨어 간의 관계성을 보여주는 관계정보와 소프트웨어 관련 문서정보, 지적재산권, 소프트웨어 개발자와 지원자정보 역시 포맷의 동일 요소와 같은 기능을 수행한다.

〈표 13〉 소프트웨어 정보요소

요소명	필수여부	정 의	하위요소
System ID	필수	시스템 ID	
Identification	선택	식별자	
Name	필수	소프트웨어명	
Other Name	선택	다른 소프트웨어명	
Version	선택	소프트웨어 버전	
Description	필수	소프트웨어 주요특징 기술	
Software Requirement	선택	운영시스템을 포함한 S/W지원에 필요한 기타 다른 소프트웨어 요건	
Hardware Requirement	선택	S/W지원에 필요한 하드웨어 요건	
Process format ⁶⁾	선택	소프트웨어에 의해 처리 가능한 포맷	create render validate 기타
Media	선택	소프트웨어 저장매체 유형	
Language	선택	소프트웨어에 의해 지원되는 언어	
Location	선택	소장여부 및 소장위치	
Related Software	선택	관련 소프트웨어	동등관계 이전/이후 버전 상위/하위 유형 기타
Documentation	선택	소프트웨어 관련 문서	서명 저자 출판사 출판일 식별자 이용가능성 주기
IPR	선택	지적재산권	유형: 저작권, 특허, 기타 소유권자: 이름, 주소, 전화번호, 이메일 유효기간 주기
Developer	선택	소프트웨어 개발자/개발기관	이름 주소 전화번호 이메일 주기

6) 프로세스 포맷정보의 create는 본 소프트웨어를 이용하여 생산 가능한 포맷유형을 나타내고, render란 본 소프트웨어를 이용하여 불러올 수 있는 포맷유형을 가리킨다.

Support	선택	현재 소프트웨어를 지원/유지하는 사람/기관	이름
			주소
			전화번호
			이메일
			유효기간
주기			
Release Date	선택	소프트웨어 발행일	
Withdraw Date	선택	소프트웨어 지원 종료일	
Image	선택	소프트웨어 제품 이미지	
Source	필수	정보제공자	이름
			주소
			전화번호
			이메일
			주기
Source Date	필수	포맷등록시스템 등록일	
Last Updated	필수	최종 업데이트일	
Type	필수	소프트웨어 관리 상태 표시 [지정값: 최초등록, 업데이트, 노후화, 삭제, 기타]	
Note	선택	소프트웨어 정보주기	

4.2.4 기술정보 입력사례

Microsoft Word 2002에 대한 포맷정보와 소프트웨어정보 입력 사례이다.

<표 14> 포맷정보 입력사례

	요소명	정의	예시
포맷 일반 정보	System ID		
	Identifiers	식별자	PUID: fmt/40 MIME: application/msword
	Name	포맷명	Microsoft Word for Windows Document
	Other name	다른 포맷명	Microsoft Word for Windows Document(97-XP)
	Version	포맷버전	97-2002
	Description	특징 설명	Word97의 발행과 함께, 마이크로소프트사는 바이너리 워드프로세싱 포맷을 개정하였다. 이 포맷은 독점포맷으로, 마이크로소프트는 이 포맷의 상세구조를 공개하지 않는다..... Word2000, 2002에서도 이 포맷은 변하지 않고 그대로 유지된다.
	Classification	포맷유형 분류	Application(Word processing)
	Status	포맷 상태 정보	현용(active) 포맷
	Developer	개발자	Microsoft Corporation
포맷 기술 속성	Disclosure	포맷공개	None(비공개)
	Related file formats	관련 파일포맷	Has priority over OLE2 Compound Document Format Is subsequent version of Microsoft Word for Windows Document(6,0/95) Is subtype of OLE2 Compound Document Format

Documentation	관련 문서	Title	OpenOffice.org's documentation of the Microsoft Compound Document file format		
		Author	OpenOffice.org		
		Publication date	15 Sep 2004		
		Publisher	OpenOffice.org		
		Identifiers	URL : sc.openoffice.org/compdocfileformta.pdf		
		Availability	Public(공개)		
Orientation	text/binary	Binary			
External signature	외부형식	File extension: doc			
Internal signature	내부형식	Name	Microsoft Word for Windows 97-2003		
		Description	Compound document header + WordDocument user stream + Version number		
		Byte sequences	Position type	Absolute from BOF	
			Offset	0	
			Value	DOCF11E0A1B11AE1 20} FEFF	
			Position type	Variable	
			Value	57006F007200640044006F00630075006D006 5006E007400{42}02(00 01)	
			Position type	Variable	
Value	4D6963726F736F667420576F7264 (20382E30 20392E30 2031302E30 2D44 6F6B756D656E74)				
Source	정보 제공자	국가기록원			
Source date	최초 입력일	11 Mar 2005			
Last updated	최종 업데이트	02 Aug 2005			
Type	포맷관리 상태	Updated			

〈표 15〉 소프트웨어정보 입력사례

	요소명	정의	예시
소프트웨어 정보	System ID		
	Name	소프트웨어명	Microsoft Word
	Version	소프트웨어 버전	2002/XP(10.0)
	Software requirements	소프트웨어 요건	operating systems: Windows98SE, ME, NT4 SP6a, 2000, XP
	Hardware requirements	하드웨어 요건	Pentium133, Windows 98 with 32mb RAN 또는 Windows NT4.0 with 40mb RAM, 150mb for typical install
	Processing	프로세스 정보	create

			- Microsoft Word for Windows Document1.0, 2.0, 6.0/95, 97-2002(*.doc) - Rich Text Format1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6/2000, 1.7/XP/2002(*.rtf)
		render	- Microsoft Word for Windows Document2.0, 6.0/95, 97-2002(*.doc) - Rich Text Format1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6/2000, 1.7/XP/2002(*.rtf)
Media format	저장매체	CD-ROM	
Language	언어	English	
Released date	발행일	31 May 2001	
Developer	개발자	Microsoft Corporation	
Support	지원/유지기관	Microsoft Corporation	
소프트웨어 관리 속성	Source	정보 제공자	국가기록원
	Source date	최초 입력일	11 Mar 2005
	Last updated	최종 업데이트	02 Aug 2005
	Type	소프트웨어 관리 상태	Updated

4.3 포맷등록시스템의 기능과 활용

4.3.1 생산단계에서의 기능과 활용

일반적으로 디지털객체의 생산단계는 보존 활동과는 무관한 외부활동으로 볼 수도 있다. 그러나 이 단계는 향후의 보존활동에 커다란 영향을 미치는 파일포맷, 하드웨어와 소프트웨어 요소들이 결정되는 중요한 시점이다. 따라서 업무에 이용되는 문서의 파일포맷은 특정기관에 의해 일정한 파일포맷으로 규정되기보다는 기록생산프로세스의 기능적 요구조건, 즉 업무상

의 프로세스와 장기보존을 위한 요구조건에 따라 선택하는 것이 바람직하다. 이렇게 기록 생산자가 데이터의 생산에 앞서 기록의 장기보존 지속가능성을 고려하여 파일포맷을 선택한다면 디지털 보존에 드는 비용의 상당부분을 절약할 수 있다.

그러나 이를 실현하기 위해서는 신뢰할 만한 포맷정보를 소장한 포맷등록시스템이 전제되어야 하며, 생산자들은 다음 요건을 고려하여 포맷을 선택함으로써 기록의 생산시점부터 장기보존을 염두에 둘 수 있다.

〈표 16〉 생산자가 파일포맷 선택 시 고려사항

요건	정의
공개표준(Open standards)	어느 특정 벤더에 속하지 않은 제품
편재성(Ubiquity)	오랜 기간이 지난 후에도 언제, 어디서든 사용가능한 포맷
안정성(Stability)	기술적 변화에도 안정적이며, 소급호환성이 있는 포맷
메타데이터 지원(Metadata support)	파일포맷 안에 메타데이터 필드가 포함된 포맷
특성(Feature set)	업무프로세스에 필요한 모든 특성과 기능을 지원 가능한 포맷
상호운용성(Interoperability)	다른 이용자나 기관들도 사용가능한 포맷
실행가능성(Viability)	에러검출 기능을 제공하는 포맷

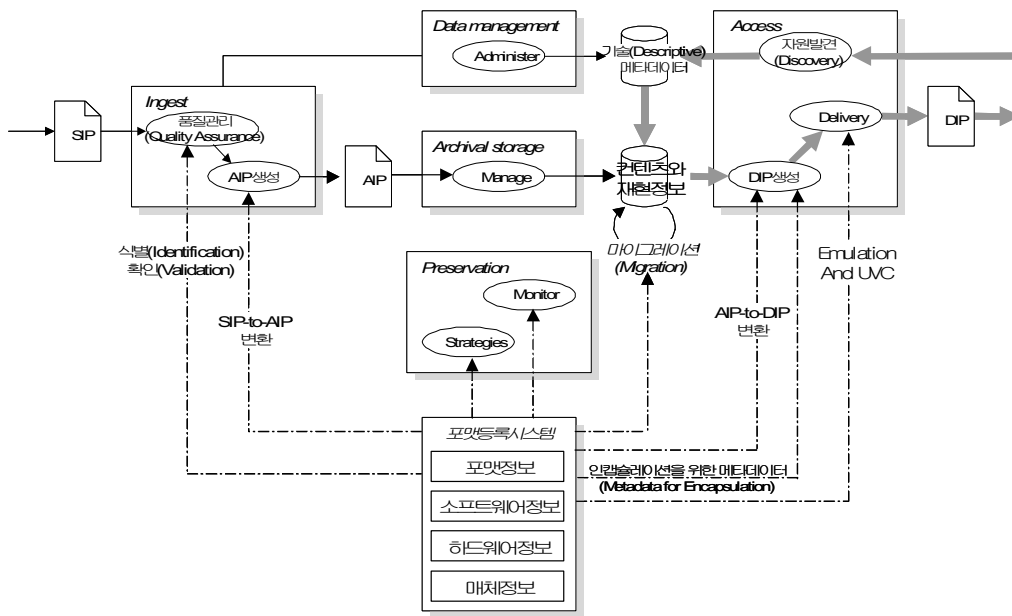
4.3.2 보존 및 접근단계에서의 기능과 활용
 본 장에서는 Abrams가 제안한 OAIS참조모델에 기반한 포맷등록시스템 데이터모델을 근거로, 입수-저장-접근에 이르는 기본적인 보존 프로세스와 포맷등록시스템과의 연관관계, 그리고 그 기능 및 활용방안에 대하여 살펴보았다.

〈그림 2〉를 살펴보면, 우선 생산자가 생산한 디지털정보 중 장기보존의 가치가 있는 기록은 제출정보패키지(SIP) 형태로 변환하여 OAIS로 보낸다. 이때 포맷등록시스템은 품질관리(Quality Assurance) 프로그램을 통해 SIP 형태를 식별하여 그 타당성을 검증한다. 그리고 보존시스템으로 보낼 적절한 저장정보패키지(AIP)로의 변환을 지원하고 관련 기술메타데이터를 추출하여 등록시스템에 기술한다. 이렇게 변환된 AIP는 OAIS저장시스템으로 옮겨지고, AIP

를 구성하는 내용정보와 이에 대응하는 보존기술정보에서 적절한 기술정보(technical information)를 추출하여 포맷등록시스템에 추가한다.

둘째, 저장된 AIP는 이용자의 요구에 따라 다양한 형태의 배부정보패키지(DIP)로 변환하여 이용자에게 제공된다. AIP에서 DIP로의 변환 타당성 또한 본 시스템을 통해 확인된다.

셋째, 포맷등록시스템은 OAIS 보존시스템과 연계하여 기술변화와 노후화를 모니터링하고 그런 변화가 저장된 기록에 미칠 영향력을 평가한다. 이런 과정을 기반으로 마이그레이션나 에뮬레이션, 인캡슐레이션 등 그에 맞는 적절한 보존전략을 수립하여 수행함과 동시에 이 과정에서 새로이 생성되는 기술메타데이터를 포맷등록시스템에 기술해준다.



〈그림 2〉 OAIS참조모델과 포맷등록시스템

5. 결론 및 제언

지금까지 전자기록의 장기보존 및 장기접근성 유지에 효과적인 포맷등록시스템에 대하여 살펴보았다. 포맷등록시스템에 대한 해외 각국의 관심과 활발한 연구 활동에 비해 우리나라 기록학계에서는 아직 이렇다 할 연구 성과가 이루어지지 않았다. 이에 마지막으로, 전자기록의 안전하고 효과적인 장기보존을 위해 몇 가지 제안점을 제시하고자 한다.

우선 신뢰할 만한 전자기록보존환경 수립에 앞서 국가기록원은 전 공공기관, 넓게는 민간 영역에 이르기까지, 모든 분야에서 생산·활용되는 전자기록의 유형과 기술정보에 대해 조사하고 파악해야 한다. 그리고 이런 정보들이 어떤 위험에, 얼마나 노출되어 있으며 최종적으로 어떻게 처리되어 관리되고 있는지에 대한 사전 평가가 수행되어야 한다. 그리고 이런 선행조사 결과를 바탕으로 다음과 같은 추가적인 조치들이 이루어져야 한다.

첫째, 전자기록의 장기보존을 위한 국가 차원에서의 규제와 제도적 장치 마련이 요구된다. 급속도로 증가하는 디지털기록과 하루가 다르게 변화하는 IT기술을 조율하기 위해서는 성문화된 정책과 절차, 지침이 필요하며, 이런 제도적 장치 하에서 포맷 및 관련 기술정보에 대한 체계적인 관리가 이루어져야 한다.

둘째, 포맷 및 관련 기술정보를 집중적으로 관리해 줄 포맷등록시스템이 개발되어야 한다. 보존시스템을 구성하는 보존메타데이터 안에서 기술정보를 관리하는 것은 더 먼 미래를 대비할 때 적합하지 않다. 보존시스템 안에서는

기술메타데이터를 위한 기본필드조차 충분하게 제공되지 못하기 때문에 디지털객체에 관한 구체적인 기술정보를 기록으로 남기기 어렵다. 이런 문제는 향후 기록의 이용 및 활용영역에서 심각한 문제를 야기할 것이다.

셋째, 포맷등록시스템에 저장된 포맷 및 관련 기술정보를 계속적으로 관리하고 유지하여야 한다. 기술정보에 대한 지속적인 업데이트가 이루어지지 않으면 본 시스템에 대한 신뢰성과 지속가능성이 위협받게 된다. 따라서 여러 소프트웨어 관련 업체들과 협력하여 급변하는 기술정보를 끊임없이 수집하고 갱신하여 포맷시스템에 저장된 기술정보에 대한 신뢰성을 높여주어야 한다.

넷째, 우리나라에서 개발한 소프트웨어 정보는 반드시 수집하고 관리하여야 한다. 외국에서 생산한 소프트웨어정보는 다른 해외 사이트를 통해 향후 정보 수집이 가능하지만, 우리나라에서 생산한 소프트웨어인 경우 국가차원에서 수집·관리하지 않으면 미래의 정보 활용을 보장할 수 없다.

다섯째, 이렇게 개발된 시스템과 이용자를 상호 연결시켜 주는 편리한 인터페이스가 개발되어야 한다. 인터페이스란 사용자인 인간과 컴퓨터 또는 시스템을 연결하는 매개 장치로, 인간 중심의 이상적인 인터페이스 설계를 통해 기록의 검색 및 활용성을 높일 수 있다.

여섯째, 여러 관련기관들의 적극적인 참여와 실질적이고 구체적인 연구 활동이 필요하다. 아직은 미흡한 상황이지만 많은 기관들의 연구 노력과 동기부여가 한층 발전된 보존환경 수립의 기틀을 마련할 것이다.

참 고 문 헌

- 설문원, 김연정, 천권주., 2006. 「ISO/TR18492의 전자기록 장기 보존 전략」. 서울: 한국국가기록연구원.
- Abrams, Stephen. L. 2005. "Establishing a Global Digital Format Registry." *Library Trends*, 54(1): 125-143.
- Abrams, Stephen., Flecker, Dale. 2005. *A Proposal for a Global Format Registry*. Cambridge: Harvard University Library.
- Brown, Adrian. 2005. "Automating preservation: new developments in the PRONOM service." *RLG DigiNews*, 9(2). [cited 2006.10.2].
<http://www.rlg.org/en/page.php?Page_ID=20571#article1>
- Brown, Adrian. 2005. *PRONOM 4 Information Model*. Richmond: The National Archives.
- Brown, Adrian. 2004. *PRONOM 4 User Requirements*. Richmond: The National Archives.
- Brown, Adrian. *The PRONOM Service: a technical registry to support long-term preservation*. Richmond: The National Archives.
- Rothenberg, Jeff. 1998. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. A Report to the CLIR. [cited 2006.10.6].
<<http://www.clir.org/pubs/reports/rothenberg/contents.html>>