

## 장기보존전략으로서의 XML

- 이미화, 현문수 역 (2006). @World Series No. 1  
『전자기록의 유형별 보존기법』, 한국국가기록연구원 -

왕 호 성\*

최근 기록학계에서 가장 활발하게 논의되고 있는 주제의 하나가 전자기록의 장기보존과 관련된 것이다. 개정 기록관리법에서 전자적으로 생산된 기록뿐만 아니라 비전자기록까지 전자적으로 관리해야 함을 명시하고 있는 것을 볼 때, 이미 전자기록과 관련된 실무 차원의 검토가 상당 수준까지 진척되었을 것으로 추측된다. 년초 한국국가기록연구원에서 이슈페이퍼로 발간된 『ISO/TR 18492의 전자기록 장기보존전략』에 이어 나온 @World Series No.1 『전자기록의 유형별 보존기법』은 그 내용적인 연속성과 실무적용에 참조할 만한 구체성을 함께 갖추고 있어 출판을 기획한 연구원의 고민과 노력을 엿볼 수 있다. 또한 기록학적 관점을 견지하면서 컴퓨터공학 또는 정보공학에서 이용하는 기법과 용어들을 비교적 이해하기 편하게 풀어내고 있는 역자들의

---

\* 민주화운동기념사업회 사료관

지식과 노고 또한 돋보인다.

이 책은 2003년 네덜란드 국립기록관의 'Digital Preservation Testbed Project'의 결과물을 번역한 것으로 대표적 전자기록 유형인 텍스트, 전자메일, 데이터베이스, 스프레드시트 기록에 대하여 3가지 장기보존전략인 마이그레이션, 에뮬레이션, XML 기법을 각각 테스트한 후, 적절한 전략을 도출·권고하는 방식으로 구성되어 있다. 테스트베드는 무결성과 검증·인증과 같은 보안기법이 진본성의 핵심개념임을 전제한 후, 전자기록의 5대 특징(맥락, 내용, 구조, 외형, 행위)의 무결함, 전자서명의 적용, 메타데이터의 보존 가능성을 실험하고 적절한 장기보존전략을 제시한다. 첫 번째 전략인 마이그레이션은 소급호환성·상호운영성·비독점 표준으로의 변환을 실험한다. 두 번째 전략인 에뮬레이션은 H/W에뮬레이션, S/W에뮬레이션, 그리고 마이그레이션을 일부 적용한 에뮬레이션 기법인 UVC(Universal Virtual Computer)를 실험한다. 세 번째 전략인 XML에서는 원본파일과 메타데이터, 기타 필요한 파일들에 대한 캡슐화(Encapsulation)<sup>1)</sup>를 실험한다. 테스트베드는 장기보존을 위한 실험결과를 토대로 텍스트기록은 PDF와 결합된 XML 전략을, 전자메일기록은 첨부파일과 전송파일이 프레임워크를 통해 연계되는 XML 전략을, 데이터베이스기록은 어플리케이션이 포함된 다중 스키마 구조의 XML 전략을, 스프레드시트기록은 기본형식의 XML 전략을 각각 권고한다. 각각의 유형

---

1) 캡슐화(Encapsulation)와 함께 래퍼(Wrapper), 컨테이너(Container), 프레임워크(Framework) 등이 혼용되고 있음

별 기록에 따라 XML 보존객체의 구성이 그림으로 제공되는데 전자기록에 따른 XML의 구조와 보존요소를 이렇게 상세히 소개한 자료를 본 적이 없어 업무뿐만 아니라 관련 연구에 두고두고 참조가 될 것 같다. 독자들에게도 이 구조도들에 나온 내용만큼은 놓치지 말 것을 권하고 싶다.

이 책은 여러 가지 측면에서 다양한 장점을 보여주고 있다. 첫째 정보공학적 사고와 용어에 익숙하지 않으면서도 전자기록을 다루어야 하는 기록전문가들의 부담을 다양한 사례제시를 통해 상당부분 해소해주고 있다. MS Word와 Adobe의 PDF를 예로 한 텍스트기록에 대한 설명, 우리가 흔히 사용하는 아웃룩과 HTML 메일 예시를 통한 전자메일기록, 오라클·MS SQL·IBM DB2·MS Access 등을 이용한 데이터셀 기록의 구조설명 등 비교적 익숙한 어플리케이션의 이름과 인터페이스를 곁들인 설명을 제공함으로써 익숙하지 않은 길을 잘 안내하고 있다. 둘째 단지 설명만을 통해 이해하기가 쉽지 않은 UVC전략에 관하여 IBM에 의해 개발되었음을 명시하여, 독자들이 하여금 해당 사이트에서 UVC 어플리케이션을 다운받아 자신의 PC에서 독특한 에뮬레이션을 경험할 기회를 제공하고 있다<sup>2)</sup>. 셋째 모든 유형별 기록의 장기보존 전략으로 권고되는 XML 보존객체들의 구조를 상세하게 설명하여 독자들이 비록 비전문가일지라도 비교·검토하기에 편리하다. 뿐만 아니라 역시 장기보존전략으로 검토되어지는 PDF 파일에 대한 상세 소개는 이 책을 본 기록전문가들과 ICT전문가들 간의

---

2) <http://www.alphaworks.ibm.com/tech/uvc>

의사소통에 도움이 될 수 있을 것이다. XML이 진본성 유지에 필요한 특정 개체를 유형에 관계없이 함께 보존하여 맥락관계를 표현할 수 있고, 스타일시트를 통해 외형을 표현하며, 스키마를 통해 기록의 계층성을 표현할 수 있음을 보여주고 있는 점들은 왜 XML이 전자기록의 장기보존전략으로 권고되고 있는지를 잘 설명해준다. 넷째 마이그레이션을 테스트하는 기준으로 제시된 드쥬르표준(De jure standard)과 드팩트표준(De facto standard)의 구분은 기록관리시스템의 도입이나 행정정보DB구축 사업과 같은 디지털화 사업을 추진하는 사업담당자에게 유용한 아이디어가 될 수 있을 것으로 생각된다. 다섯째 기술적인 권고와 함께 장기보존전략의 성공적 수행을 위한 정책 권고를 각각의 유형별 기록에 따라 제시하고 있다는 점은 이 책의 짜임새 있는 구성을 더욱 돋보이게 함과 동시에 이 책의 혜택을 볼 다양한 사람들에게 유익할 것이다. 아쉬운 점을 굳이 지적한다면 소개되는 각종 S/W들의 명칭과 버전이 2006년 현재의 우리 상황과 다소 다른 부분이 있으므로 이를 독자들이 각자 판단해야 한다는 것이다. 예컨대 Adobe Distiller 4.0을 기준으로 한 설명에서 현재의 상용화 버전이 7.0이고 세부기능에 약간의 차이가 있다거나 우리가 거의 사용하지 않는 Foxpro나 dbase의 예를 드는 정도가 그것일 것이다. 그러나 역시 사소하다.

2006년 9월 국가기록원에서 나온 “전자기록 영구보존기술 적용을 위한 테스트베드 구축”이라는 제목의 최종보고회 발표 자료에 의하면, 이 책에서 제시하는 방향과 거의 유사하게 장기보

존 또는 영구보존 전략이 적용되고 있음을 확인할 수 있다. 자료의 내용을 보면 전자기록의 보존포맷으로 ISO 19005-1:PDF/A-1, 영구보존포맷으로는 원본파일+PDF파일+인증정보+메타데이터가 캡슐화(encapsulation)된 XML이 채택된 테스트베드 프로젝트가 현재 우리나라 중앙기록물관리기관의 주도로 진행되고 있음을 알 수 있다. 제한된 시간 내에 원하는 최종산출물을 향하여 다양한 분야의 전문가들이 끊임없이 의사소통을 해야 하는 것이 프로젝트의 특징임을 감안할 때, PDF의 변환과 XML 캡슐화를 실험하는 프로젝트팀에 본 번역서가 기여할 부분이 상당부분 있을 것으로 생각한다. 기록관리전문가와 ICT전문가들에게 상호간의 전문지식을 교차 이해시킬 수 있는 참고서적을 발견하기가 쉽지 않은 터라 이 책이 전할 수 있는 가치와 혜택은 다양하고 넓을 것으로 생각된다. 마땅히 기록관리 업무종사자와 연구자 및 관계자들에게 일독을 권한다. 직접 번역하는 수고로움을 덜면서 전자기록을 중심으로 알아두어야 할 정보공학영역의 핵심 키워드가 가득 담겨있는 이 책의 가치를 구구절절 말하기가 어찌 쉽겠는가? 참으로 기쁜 마음으로 이 책을 읽었음을 옮기신 두 분 선생님께 전하며, 혹여 미흡한 평이 발견되더라도 너그러운 양해를 바란다.