디지털 자원의 위험관리 사례연구

Case Study of Risk Management for Digital Resources

이 미 화*
Mi-hwa Lee

차 례

초 록

디지털 아카이빙 실행을 위한 위험관리의 필요성을 인식하고, 위험을 줄이기 위한 방안을 강구하고자 해외의 위험관리 사례를 조사한다. 사례로 선정된 기관은 OCLC와 코넬대학 도서관이다. 양 기관은 실질적인 실험을 바탕으로 위험관리를 수행해 타 기관에 좋은 모범이 되고 있다. 사례를 바탕으로 국내에서 위험관리의 실행을 위한 방안으로 위험관리 요소 규명, 파일 포맷과 마이그레이션 프로그램 등의 실험 테스트, 위험관리 수행 전담기구 설립을 제안한다.

키 워 드

위험관리, 디지털 아카이빙, 디지털 보존, 마이그레이션

* 한성대학교 학술정보관 학술정보팀 주임
(Staff, Dept. of Infomation Services, Hansung University Library, leemh@hansung.ac.kr)
• 논문접수일자 : 2005년 11월 14일
• 게재확정일자 : 2006년 1월 11일

ABSTRACT

This paper is to probe foreign cases about risk management and to develop a method for the risk management of the digital resources. This is under the perception that the risk management is necessary for digital archiving implementation. OCLC and Cornell University Library are the leading institutions in the area of risk management. They are the best examples because they make a decision about the risk management through the testing and experimentation. Based on case studies, identification of element of risk management process in our environment, the foundation of national center for risk management, and testing file format and migration program as practical plan of risk management for digital resources are suggested.

KEYWORDS

Risk Management, Digital Archiving, Digital Preservation, Migration

1. 서 론

1.1 연구목적

디지털 정보의 생산은 급격히 증가하고 있으나 이를 장기적으로 관리하기 위한 실질적 방안이 마련되지 않아 디지털 자원의 소멸가능성이 커지고 있다. 도서관이 구축한 TIFF 등의 이미지 파일은 지속적으로 관리하지 않을 경우 가독이 불가능하며, 구입한 시디롬과 아카이브용으로 소장한 시디 자료도 몇 년 전 것은 가독이 어렵다. 특히 구입한 시디롬의 경우는 운영체제와 애플리케이션이 호환되지 못하며, 일부 아카이브용 파일은 버전이 낮아 내용을 정확히 디스플레이하지 못한다. 도서관이나 아카이브는 인류의 지식자원을 후대까지 보존시켜야 하는 책무를 가지고 있으나, 수집된 디지털 자원의 관리는 인쇄자원보다 어렵다.

디지털 자원의 영구접근을 위한 활동으로 현재 디지털 아카이빙 또는 디지털 보존 활동이 진행되고 있다. 디지털 아카이빙은 단순한 파일의 저장이 아니라 생산부터 보존에 이르기까지 디지털 자원의 생애주기 전반에 걸친 자료보존을 의미한다. 하지만, 현재까지 디지털 아카이빙을 위한 기술 및 보존전략이 현장에 접목될 수 있을 만큼 발전되지 않아 디지털 자원의 손실위험은 지속된다. 이러한 위험에 대처하기 위해 위험관리를 도입해 디지털 아카이빙할 대상을 선정하고, 마이그레이션 후 파일의 변경사항 체크 실험을 통해 보존기법을 결정하고 있다.

본고는 디지털 아카이빙의 안전한 실행을

위해서 위험관리의 필요성을 인식하고, 해외 위험관리 사례를 통해 국내의 위험관리 방안을 제시하고자 한다. 단, 웹 자원의 위험관리가 디지털 자원의 위험관리 절차와 방법에 차이가 있어 웹 자원의 위험관리는 연구범위에서 제외하였다. 해외 사례연구의 대상은 위험관리를 수행한 OCLC와 코넬대학 도서관이다. 두 기관은 PADI에 제시된 위험관리 분야 연구 중에서 일반 디지털 자원의 위험관리 프로젝트를 수행한 기관이다. 국내의 디지털 아카이빙 연구는 시작단계이고, 디지털 자원의 위험관리 필요성에 대한 인식은 낮은 상태이기 때문에 본 연구를 통해 위험관리의 인식을 제고하고, 디지털 자원의 위험을 줄일 수 있는 방안의 첫 단계인 위험요소를 파악하고, 디지털 자원 위험관리를 위한 정책적인 방안을 마련할 수 있을 것이다.

1.2 선행연구

위험관리에 관한 해외 사례는 PADI의 위험관리 범주 내에 수록된 서지 리스트가 전부이고, 국내에서 위험관리를 현장에서 실시한 사례는 거의 없다. 위험관리 분야의 문헌은 위험관리의 일반론, 보존전략에 따른 위험관리, 그리고 웹 자원의 위험관리로 구분할 수 있다.

일반적인 위험관리의 개념을 많은 프로젝트에서 제시하고 있으나, 디지털 보존의 맥락에서 위험관리의 개념과 원칙에 관해 언급한 ERPANET이 대표적이다. ERPANET은 위험 커뮤니케이션 툴(Risk Communication Tool)의 목적과 위험관리의 수행단계에 대한 개괄적인 개념을 설명한다. ERPANET은 디지털 보존에 대한 적극적인 대응을 위한 도구로 ERPAtools(Cost Orientation Tool, Selecting Technologies Tool, Digital Preservation Policy Tool, Risk Communication Tool)을 개발하였다. 이 중 위험 커뮤니케이션 툴은 기관에서 어떤 디지털 자원이 위험에 처했는 지를 파악하고, 위험요인을 찾아, 분류하여 위험의 순위를 매기고, 위험영역이 소통되도록 하고, 위험관리전략개발을 자극하는 도구이다.

보존전략에 따른 위험분석에 대한 연구로는 보존전략 중 일반적으로 사용되는 마이그레이션의 위험분석을 시도한 코넬대학 도서관이 대표적이다. 국내에서는 서은경(2003)이 사서를 대상으로 보존기법 전반에 걸친 설문조사를 실시하였다. 디지털 자원의 보존기술을 매체재생, 매체변환, 포맷 변환, 정보전환, 에뮬레이션으로 나누고, 이 보존기법 중 현재 대학도서관에서 어떤 기술이 디지털 보존을 위해 사용되며, 디지털 정보관리 담당자들이 생각하는 보존기술에 대한 위험가능성과 위험영향력을 설문조사하였다. 설문항목에는 디지털 정보자원 보존활동이 수행되는 부서, 보존활동의 주기, 사용한 기술, 가장 많이 사용된 기술, 변환 후 이전 파일 보존여부, 위험관리 계획의 수립여부 등이며, 이를 통해 현재 우리나라 대학도서관의 위험관리의 현황을 파악하였다. 30개의 응답 도서관 중 8개 도서관만이 정기적으로

보존업무를 수행하였으며, 8개는 비정기적이며, 4개는 실시하지 않았다. 보존활동을 하는 16개 기관 중 포맷 변환을 가장 많이 사용했고, 매체재생, 매체변환, 정보전환의 순으로 에뮬레이션 기법을 제외한 기법이 사용되었다. 에뮬레이션은 최신기법이나 우리나라에서는 사용되지 않았다. 또한, 기법별로 위험가능성과 영향력에 대한 설문조사 결과 매체재생이 가장 안전한 방법으로 선정되었고, 매체변환, 포맷변환, 정보전화, 에뮬레이션 순이었다. 사서들은 에뮬레이션을 가장 위험이 발생할 확률이 높은 보존기법으로 인식하였다. 위험의 영향력을 조사한 결과, 정보전환이 위험영향력이 높고, 매체재생의 영향력은 낮았다.

웹 자원의 위험관리로 호주국립도서관의 『Archiving Web Resources』에는 위험평가 관련 내용을 포함하고 있으나, 웹 자원 자체의 내용에 대한 위험이 아니라 조직 내에서 위험을 감소시키기 위해 레코드 보존측면에서 웹 자원의 계속적인 추적의 필요성을 제시하였다. 실질적인 테스트 사례는 코넬대학의 프리즘(PRISM, Preservation, Reliability, Interoperability, Security and Metadata) 프로젝트가 대표적이다. PRISM에서는 기존의 위험관리 모델을 웹 자원에 맞도록 수정해 4단계로 구분한다. 1단계인 위험요인수집 및 범주화 단계에서는 웹 자원의 위험요인을 파악하기 위해 Mercator 웹크롤러라는 수집도구를 사용하며, 2단계인 단순한 위험요인 파악 및 탐지에서는 웹 자원의 주요 속성을 분석하고, 웹 자원의 생애주기 동안 나타나는 사건을 감지한다. 3단계인 전체적인 위험요인 파악 및 탐지에서는 다양한 기술을 활용해 위험분석 엔진을 개발하고, 4단계는 자동 보존정책 강화로 위험을 방지하기 위한 활동을 수행한다. 웹 자원을 아카이빙하기 위해서는 위험요소를 분석하고 모니터링해야 한다. 이를 위해 대상 웹 자원을 선정하고, 웹 자원의 모니터링 유형(독립형 웹 페이지, 하이퍼링크된 웹 페이지, 웹 페이지 모니터링, 기술과 조직맥락에서 웹 페이지 모니터링)을 4가지로 나누고, 각각의 웹 자원을 모니터링하는 방식의 차이를 설명한다(Kenney 2002; McGovern 2004). PRISM 프로젝트의 일환으로 도서관에서 관리하고 있는 동남 아시아의 웹 자원에 대한 위험관리 연구가 수행되었다. 웹크롤러가 동남 아시아 7개국의 웹사이트의 변경사항을 지속적으로 추적하였다(Botticelli 2003).

2. 위험관리의 개념과 수행과정

2.1 배경 및 정의

위험에 대해 전문가들은 자신의 특수한 배경에서 정의하고 이를 측정한다. 환경학자는 위험은 원치 않는 영향의 발생가능성으로 정의한다. 경영분야에서 위험은 경영자나 회계사가 불확실한 환경에서 일어날 사건의 영향에 관한 자신의 걱정을 표현하기 위해 사용하는 개념이다(Lawrence 2000).

위험관리는 보험, 건강, 비영리조직, 환경모니터, 재정적인 곳에서 주로 이용되었고, 디지털 보존에는 큰 영향을 일으키지는 않았다. 경영학에서 경영의 연속성을 위한 재난계획에서 위험관리로 변환이 이루어지고 있다. 디지털 보존에는 위험관리가 적용되지 않다가 Lawrence의 연구 「Risk management of digital information」에서 마이그레이션에 대한 위험관리 방법론을 개발하였다. 마이그레이션 과정을 구분하고, 위험범주와 특정 위험요소를 정의하였다.

위험관리 정의로 많이 인용되는 것은 Kenney이다. 그는 위험관리는 불확실한 환경에서 생길 수 있는 다양한 결과 및 걱정반경, 이에 따른 인지된 위험과 허용 손해를 확률적으로 파악하고 그 대체 방안을 세우기 위해 사용하는 방법이다(Kenney 2002, 서은경 2003 재인용).

따라서 디지털 자원의 영구적인 접근을 위한 디지털 아카이빙이 수행되는 상황에서 현재의 기술로 아카이빙된 디지털 자원이 미래에 가독될 수 있을 것인 지의 우려 속에서 이에 따른 위험을 최소화하기 위해 전반적인 위험요소를 규명하고, 위험요소별로 위험의 수준과 영향을 평가 분석하여, 대안을 모색하는 것이 위험관리이다.

2.2 디지털 아카이빙과 위험관리

디지털 자원은 태생적으로 위험에 노출되어 있으며, 기술의 변화에 따라 지속적인 관리가 이루어지지 않을 경우 손실의 위험이 매우 높다. 디지털 자원의 손실이나 손상으로 부터 이를 적절히 관리하기 위한 방법이 디지털 아카이빙 또는 디지털 보존이다. 디지털 아카이빙이 지속적 가치를 가졌다고 판단되는 디지털 개체를 장기간 관리하는 활동으로 정의되며, 이는 가치 있는 디지털 자원을 선별하여 그 내용 및 기능을 보존, 관리하여 장기간 접근할 수 있도록 하는 전반적인 활동을 포괄하는 개념이다(설문원 2005).

디지털 아카이빙에서 모든 디지털 자원을 보존할 수 없기 때문에 아카이빙할 대상자원을 선정하고, 적합한 파일 포맷을 결정하고, 보존 시스템을 설계하고, 파일의 손상이 가장 낮은 보존전략을 결정해야 하는데, 이는 위험관리와 연계 속에서 이루어진다. 예를 들어, 파일 포맷의 마이그레이션 전후를 비교한 결과 파일의 내용과 구조 등이 적절히 변경되지 못했을 경우 마이그레이션 프로그램을 수정하거나, 또는 여러 프로그램의 마이그레이션 결과를 조사해 확률적으로 가장 위험이 낮은 프로그램으로 변경한다.

따라서 위험관리는 디지털 아카이빙의 일부이며, 위험요인분석과 위험평가 등을 통해 아카이빙의 전과정을 지원한다.

2.3 보존전략과 위험관리

디지털 아카이빙은 생산 당시의 디지털 자원의 내용과 외형을 미래의 어느 시점에서 완벽

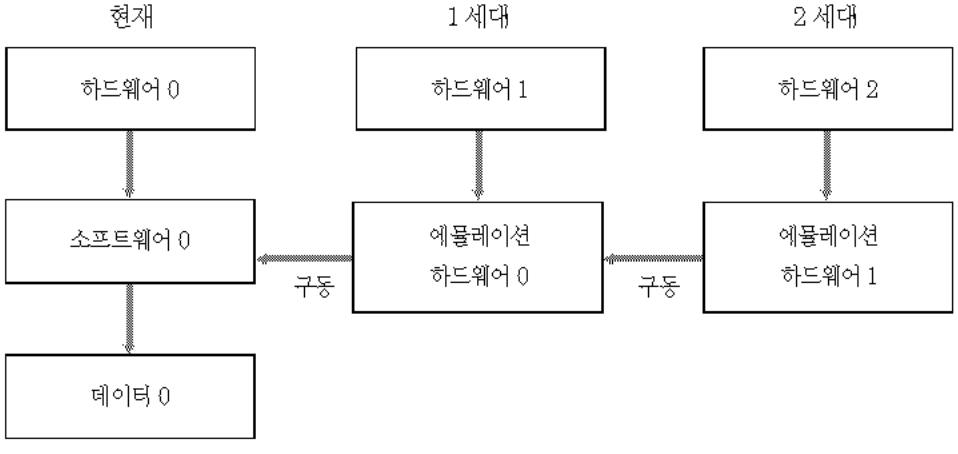
하게 재현하기 위한 것이므로 보존전략이 가장 핵심적 내용이며, 보존전략은 파일 포맷의 가독과 연관되어 있어 위험관리에서도 핵심 분야이다. 따라서 위험관리를 위한 보존전략으로 마이그레이션과 에뮬레이션의 특징과 이에 따른 파일의 위험요인을 파악하는 것이 중요하다.

마이그레이션은 하나의 하드웨어/소프트웨어 환경에서 다른 하드웨어/소프트웨어 환경으로 디지털 정보자원을 주기적으로 이전하는 것으로 정의된다. 즉, 이전 버전으로 만들어진 디지털 정보를 최신 컴퓨터 플랫폼에서 구동되는 신규 포맷으로 변환하는 것이다.

마이그레이션 실행 시에는 다음의 사항을 고려한다. 가장 중요한 점은 마이그레이션의 목적은 단순히 개별적인 아이템이나 포맷을 보존하기 위한 것이 아니라 미래의 어느 시점에서 그 자원에 접근할 수 있도록 보장하기 위한 것이다. 둘째, 마이그레이션 실행한 후의 디지털 정보자원을 삭제 및 폐기하기 전에 반드시 마이그레이션 과정이 성공했음을 점검해야 한다. 또 마이그레이션 시점에서 내용 데이터 뿐 아니라 모든 연계 및 메타데이터를 동시에 획득하도록 보장해야 한다. 마이그레이션 기법이 디지털 자원의 아카이빙에서 가장 위험부분이 큰 것은 정보손실의 위험을 가지고 있기 때문이다. 따라서 마이그레이션의 결과를 반드시 점검해야 한다(설문원 외 2005).

마이그레이션에 따른 위험요인은 파일을 변환하는 변환 소프트웨어이며, 변환 소프트웨어에 의해 이전 파일의 값, 기능, 내용, 외형이 이후 파일에서 유지되는 지를 파악해야 한다. 실험결과 변환이 안정적이지 못할 경우 마이그레이션의 중지 및 소프트웨어의 재설계가 요구된다.

에뮬레이션은 디지털 정보자원을 생산하고



〈그림 1〉 에뮬레이션 전략

출처 : DPT. 2003. *Preserving text documents.*

디스플레이하는 데 사용된 소프트웨어의 원래 기능을 이용시점에서 보편적으로 사용하는 컴퓨터 환경에서 그대로 재현될 수 있게 하기 위해 생산시점과 동일한 운영환경을 재생산하는 것이다. 에뮬레이션에서는 하나의 컴퓨터에서 가동되면서 이를 통해 다른 컴퓨터를 실제적으로 재현할 수 있는 프로그램인 에뮬레이터란 소프트웨어가 필요하다. 에뮬레이션은 원래 데이터 포맷의 원본 모습을 그대로 재생산할 수 있다는 장점을 지니지만, 에뮬레이션의 직접적인 도구가 되는 에뮬레이터 소프트웨어 생산에 고도의 기술력과 고비용을 요구하며, 공급자가 그 권리를 독점하는 소프트웨어를 에뮬레이션하는 경우 지적 재산권과 관련된 문제가 발생하여 장기적으로 디지털 자원의 안정성을 손상시킬 수 있다(설문원 외 2005).

에뮬레이션 전략을 통한 아카이빙에서도 중요한 것은 에뮬레이션 소프트웨어의 생산이다. 하지만, 에뮬레이션에 대한 것은 아직까지 실험 중이며, 실질적인 데이터를 통한 위험관리 연구는 진행되지 않았다.

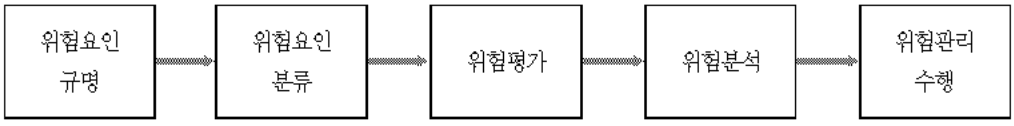
두 보존전략 중 코넬대학 도서관에서는 마이그레이션을 선택하였다. 마이그레이션을 보존전략으로 선정한 이유는 첫째 디지털 파일을 정기적으로 리프레시할 수 있고, 둘째 애플리케이션이 변경될 때 디지털 포맷이 변경되고, 셋째 독점 포맷에서 아스키 파일로 변경하는 디지털 포맷의 변환이 가능하며, 넷째 TIFF에서 PDF로의 변환같이 접근을 보완할 수 있기 때문이다. 코넬은 디지털 정보의 마이그레이션 전략의 안전성을 평가하기 위한 개발도구에 집중하였다(Lawrence 2000).

2.4 위험관리 수행과정

ERPANET의 위험 커뮤니케이션 툴은 조직 내에서 특정 디지털 자원의 위험여부를 밝히고, 둘째 디지털 자원의 위험요소를 규명하고, 셋째 디지털 자원에 위험이 될 조직 내의 영향요인을 파악하고, 넷째 디지털 자원을 관리하기 위해 위험을 범주화하고 대처를 위한 우선순위를 선정하며, 다섯째 위험영역에 관한 조직 내의 커뮤니케이션을 강화하고, 여섯째 위험관리 전략 발전을 강화하기 위한 도구이다. 위험관리의 3단계로 위험규명, 위험분석, 위험관리 실행을 〈표 1〉과 같이 제시한다(ERPANET 2003).

〈표 1〉 위험평가와 관리의 3단계

단 계	내 용
위험규명	위험에 처한 자원, 위험의 유형, 자원의 가치, 조직 내의 위험요소 등
위험분석	수용할 위험의 수준, 위험가시화, 직간접 비용, 위험가시화의 결과
위험관리 실행	위험경감을 위한 선택과 대처, 위험 우선순위, 관리전략, 위험감소 등



〈그림 2〉 위험관리 모델(5단계)

위험규명은 시스템의 이해를 통해 가능한 것으로, 위험에 처한 자원, 유형, 가치, 조직 내의 위험요소를 밝히는 과정이다. 이러한 요인이 규명되면, 각 요인별 위험의 수준을 측정하고, 위험 시나리오가 적용되었을 때 비용을 계산하는 것이 위험분석이며, 최종단계인 위험관리 실행은 위험경감을 위한 방안을 마련하고, 이를 감소하기 위한 대책을 마련하는 것이다.

Kenney는 세분화된 위험관리 모델로 위험요인 규명(risk identification), 위험요인 분류(risk classification), 위험평가(risk assessment), 위험분석(risk analysis), 위험관리 수행(risk management implementation)의 5단계를 제시한다.

위험요인 규명은 디지털 자원 전체의 잠재적 위험이나 재앙을 감지하는 과정이다. 디지털 자원의 범위, 수행도구와 기법이 포함된다. 웹 자원일 경우, 웹의 잠재적 위험을 파악하기 위해 데이터를 수집하는 웹크롤러같은 수동과 자동 기법을 사용한다.

위험분류는 위험요인을 범주화하기 위한 구조적 모델을 개발해 관찰된 위험의 속성과 사건을 개발된 범주 속에 적용시키는 과정이다.

위험평가는 위험이나 손실을 일으킬 수 있는 요인(사건)의 위험 시나리오와 결과, 사건의 발생 가능성을 규명하는 과정으로 최근 위험관리 분야의 많은 문헌은 위험평가를 주로 다루고 있다. 위험평가 시 고려해야할 변수는 자산의 가치, 위협, 영향요인, 손실 등이다. 웹 자원 위험관리 프로젝트인 프리즘(PRISM)에서는 웹 자원에 대한 주요 속성, 웹 자원의 생애주기 동안 관찰된 사건, 정보자원 환경에 관한 정보를 고려한다.

위험분석을 통해 위험 패턴이나 시나리오의 잠재적 영향, 손실의 정도, 회복을 위한 직간접 비용이 결정된다. 이 단계에서 취약사항을 규명하고, 잠재적 결과의 위험을 받아들일 기관의 의지를 고려하고, 손실을 완화시킬 방안을 마련한다. 인공지능방법, 의사지원 시스템, 조직의 프로파일 모두가 위험분석을 지원한다.

위험관리 수행에서는 규명된 위험을 관리할 정책, 절차, 메커니즘을 규정한다. 실행된 프로그램은 자산의 가치와 손상을 방지하거나 회복하는 직간접 비용 간의 균형을 유지해야 한다. 조직과 관련 당사자 모두는 프로그램을 알고 이해해야 한다. 효과적인 프로그램이 되기 위해 범위가 포괄적이고, 정기적으로 체크를 하고, 검증된 전략을 사용해야 한다(Kenny 2002).

3. 해외 사례

3.1 OCLC

3.1.1 개 요

디지털 매체의 장기보존은 2가지 문제를 갖는다. 첫째는 매체의 노후이며, 둘째는 매체가 읽히더라도 저장된 정보를 최신 프로그램으로 가독할 수 없다는 것이다. 현재 1,000여개의 디지털 포맷과 버전이 있고, 대부분은 10년 이상 사용되지 못한다. 이러한 문제를 해결하기 위해 도서관과 아키비스트는 시간을 초월할 수 있는 포맷을 연구한다. 일부 아카이브는 포맷 변경을 시도하며, 새로운 전문 포맷과 변환 과정을 개발하고 있다. OCLC는 특히 파일 포맷의 중요성을 인식하고, 디지털 포맷을 분석하고, 보존활동을 평가하기 위해 INFORM(INvestigation of FOrmats based on Risk Management)을 개발하였다. INFORM은 디지털 포맷의 위험요소를 조사·측정하여 보존실행을 위한 가이드라인을 제공한다. INFORM은 파일 포맷과 보존전략을 비교하고, 보존결정을 위한 객관적 접근을 제공하기 위해 파일의 변화를 측정한다. 파일의 변경사항에 대한 지속적인 추적을 통해 잠재된 손실과 위험을 파악한다.

3.1.2 위험범주

INFORM에서 제시한 위험평가를 위한 6가지 위험 유형은 <표 2>와 같다. 디지털 포맷과 관련된 위험 범주는 디지털 포맷자체, 포맷이 구동되는 하드웨어와 소프트웨어, 관련 조직, 디지털 아카이브, 보존전략(마이그레이션)이다.

파일 포맷은 하드웨어와 소프트웨어에 의존적이기 때문에 보존을 위해 포맷의 명세 분석이

<표 2> 위험의 6가지 범주

6 위험유형	내용
디지털 개체의 포맷	포맷 명세에 따른 위험으로, 압축 알고리듬, 독점 포맷 대 개방 포맷, 저작권 관리, 암호화, 디지털 서명
소프트웨어	운영 시스템, 애플리케이션, 아카이브 실행, 마이그레이션 프로그램, 압축 알고리듬 수행, 암호화와 디지털 서명과 같은 소프트웨어 요소에 의한 위험
하드웨어	매체유형(CD, DVD, 마그네틱 디스크, 테이프, WORM), CPU, I/O 카드, 주변장치를 포함한 하드웨어에 의한 위험
관련 조직	조직과 관련된 위험으로 아카이브, 콘텐츠 소유자, 벤더, 오픈 소스 커뮤니티 등
디지털 아카이브	디지털 아카이브 자체에 의한 위험(아키텍처, 프로세스, 조직구조 등)
마이그레이션과 보존계획	마이그레이션 과정에 의한 위험

중요하다. 파일은 하드웨어, 소프트웨어를 통해 구현되기 때문에 의존성이 크고, 시간이 지나면서 개발자들로부터 외면되면 가독성이 떨어져 장기보존에 큰 위협요인이 된다.

INFORM 방법은 6개 위험유형별로 위험요소를 규정하고 있으며, 위험요소가 위험 발생가능성과 위험이 발생할 경우 저장된 개체에 미칠 영향력의 측면에서 측정된다. 발생가능성과 영향력은 5점 척도로 측정되며, 두 가지의 결합을 통해 위험에 노출 정도를 표시한다.

3.1.3 조사 및 실험결과

위험요소의 위험 정도를 측정하기 위해 컴퓨터 학자, 포맷 전문가, 사서, 하드웨어 전문가, 법률가, 오픈 소스 개발자, 아키비스트, 저자 등으로 구성한 리뷰자에게 위험요소별로 위험가능성과 영향력을 표시하는 설문지를 보낸다. 리뷰자는 많은 경험과 기술이 있어야 하며, 그룹이 크고 다양해야 조사의 편견이 적고, 신뢰성 있는 데이터가 도출된다. 리뷰자가 설문지를 완성하면 각 요소는 3가지 범주로 나뉜다. 즉, 즉각적인 조치가 필요하지 않은 위험, 가까운 미래에 계획과 조치가 필요한 위험, 즉각적인 조사가 수행되어야할 위험이다. 이를 통해 가장 위험한 요소는 무엇이며, 즉각적인 관심이 필요한 것은 무엇인 지를 규명한다.

둘째 TIFF JEPG의 두 가지 포맷을 비교하고, 서로 다른 시점에서 파일을 평가한다. 전체 JPEG를 TIFF로 마이그레이션하는 것이 최선인 지를 연구한다. JPEG와 TIFF를 비교한 결과 포맷 간의 차이가 없고, JPEG도 TIFF만큼 적절한 포맷이다. 다만, JPEG 개체에 사용된 압축 알고리듬은 보존에 문제를 일으킬 수 있어 TIFF가 더 적합하다. 위의 비교결과는 보존을 위한 파일 포맷의 결정에 사용될 수 있으나, JPEG가 대부분인 컬렉션의 경우 TIFF로 변환하는 것이 그림의 질을 향상시키지 않는다. 따라서 단순히 포맷 간의 특징만으로 마이그레이션할 파일 포맷을 결정하는 것은 옳지 않다(Stanescu 2004).

3.2 코넬대학 도서관(CUL)

3.2.1 위험범주

디지털 보존의 전략으로 마이그레이션을 선정하고, 마이그레이션의 위험 정도를 파악하기 위해 마이그레이션 전후의 파일 비교를 통해

〈표 3〉 위험범주 및 평가도구

위험범주	위험내용	평가도구
일반 컬렉션	기관지원, 재정, 시스템 H/W, S/W	위험평가 워크북
데이터 파일 포맷	변환으로 인한 파일의 내부 구조	리더 소프트웨어 이용
파일 포맷 변환과정	변환 소프트웨어에 의한 파일 변환의 적절성	테스트 파일 비교

위험관리를 수행한다. 보존전략으로 마이그레이션을 고려할 때 〈표 3〉과 같이 3가지 위험범주가 측정되어야 한다.

첫째, 일반 컬렉션(General collection)과 관련된 위험에는 기관의 지원, 재정적 요인, 시스템 하드웨어, 소프트웨어, 직원의 존재 여부가 해당한다. 이 항목은 디지털 아카이브의 기본 인프라(deep infrastructure)이며, 본질적 요소이다. 디지털 정보와 관련된 법적, 정책적 문제도 부차적 위험으로 이에 해당한다.

둘째, 데이터 파일 포맷과 관련된 위험으로 변환에 의해 영향 받은 파일 내부 구조적 요소이다.

셋째, 파일 포맷 변환과정과 관련된 위험으로 변환 소프트웨어가 의도한 결과를 생산했는지, 변환 에러가 어느 정도인 지를 파악한다.

3.2.2 일반 컬렉션 위험평가

위험평가 도구로 워크북은 디지털 정보 마이그레이션과 관련된 잠재적 위험을 규명한다. 워크북은 위험과 문제를 평가하는 체계적인 접근

〈표 4〉 일반 컬렉션 위험평가 내용

평가요소		평가내용
소스/대상 포맷 평가	변화 소프트웨어	자체개발 여부, 상업 소프트웨어일 경우 적합성, 변환 S/W의 기능 등
	포맷	대상 파일의 표준 여부, 대상 파일의 가독 여부, 소스 포맷의 기능 유지 여부, 조직과 개발자의 지원 등
시스템 평가	하드웨어	컴퓨터 하드웨어의 일반상태, CPU의 상태, 메모리 상태, CPU 업그레이드나 대체할 계획의 유무, 저장매체 상태, 저장매체 대체나 업그레이드 계획 유무, 주변장치의 상태, 주변장치 대체나 업그레이드 계획 유무 등
	소프트웨어	OS 변경 계획여부, 데이터 조직 변경여부(저장장치의 정보밀도, 파일의 계층조직 여부, 독점적 파일 관리 시스템 여부) 등
	압축	압축여부, 압축률, 압축 스펙의 기록과 동일한 압축방법사용 여부 등
	보안	읽고 쓰기의 권한 소지, 암호화 여부 등
메타데이터 평가 (아카이브에 저장된 디지털 정보를 정확히 표현하기 위해 필요한 정보)		다큐멘테이션의 유지 여부, 다큐멘테이션의 접근가능 여부, 다큐멘테이션의 디지털화 계획, 메타데이터의 우선목적, 메타데이터 수정 계획, 소스와 대상 메타데이터 모두를 위한 내용 표준 여부, 다큐멘테이션 포맷의 독점성 여부 등
조직평가	계획	디지털 보존계획 여부, 계획의 인지 등
	재정	현 아카이브의 가치평가, 보존을 위한 예산, 마이그레이션 프로젝트 예산 등
	인력	직원수, 직원 기술(skill) 등
	이용자	이용자의 보존결정에 관여 여부 등

을 제공할 수 있다. 개발된 위험평가 워크북을 위한 위험평가 척도는 위험이 발생할 가능성과 위험의 영향을 측정한다. 평가 스케일에 따라 위험이 낮으면 마이그레이션을 수행하고, 위험이 높을 경우 마이그레이션 수행을 지연한다.

워크북의 내용은 〈표 4〉와 같은 내용으로 구성되며, 이는 마이그레이션 결정을 위한 기관의 현상태에 대한 평가를 위한 것이다.

3.2.3 파일 포맷 위험평가

보존전략으로 마이그레이션이 결정되면 위험평가를 통해 마이그레이션 프로그램을 선정한다. 파일 마이그레이션은 파일 포맷의 구조와 데이터 요소를 다른 포맷으로 변환하는 과정이다.

1) 변환 포맷 위험평가

변환 프로그램의 위험평가는 마이그레이션 전후의 파일을 조사함으로 이루어진다. 테스트 파일이 소스 파일을 대상 파일로 마이그레이션하는 컨버전 소프트웨어를 거친다. 원본의 필드와 필드 값이 대상 파일로 적절히 생성된다면, 마이그레이션으로 인한 위험은 감소된다. 반면, 필드나 값이 적절하게 변환되지 않았다면 마이그레이션의 위험은 증가한다.

변환 소프트웨어 평가 시 수치 값이 적절한 지, 특성과 텍스트는 적절한 지, 워크 시트 특성(칼럼 길이 등)은 보존되었는가, 셀 포맷(텍스트, 수치, 데이터, 시간)은 유지되는가, 셀 기능은 그대로 전달되는가를 조사하기 위해 테스트 파일을 만든다. 즉, 스프레드 시트 파일이 다른 파일 포맷으로 변환 · 마이그레이션될 때 파일의 어떤 특성이 그대로 보존되는 지를 파악하기 위한 것이다.

실험에서 테스트 파일을 이용해 Lotus의 기능을 평가했다. 소스 파일(원본 파일)은 wk 1 포맷으로 Lotus 1 2 3 버전 2.2이다. 대상 포맷은 소스 포맷이 마이그레이션을 거친 후 생성되는 파일이다. 수작업으로 테스트 파일의 마이그레이션 전후 내용을 비교하는 데 3시간이 소요되며, 정확한 결과를 얻을 수 있다. 마이그레이션 전후 파일 포맷 간의 구조적 요소와 데이터 요소의 변환이 잘 이루어졌으며, 완벽하지는 않았지만 〈표 6〉과 같이 위험이 낮았다.

Lotus 1 2 3을 상위 버전으로 업그레이드, 엑셀로 변환, 아스키 파일로 변환할 때 위험도를 평가함에 있어 상위 버전 마이그레이션은 전

〈표 5〉 변환 S/W 위험평가

평가요소	평가내용
변환 소프트웨어의 위험평가	• 수치 값 • 특성과 텍스트 변환 • 워크 시트 특성(칼럼 길이 등) • 셀 포맷(텍스트, 수치, 데이터, 시간) • 셀 기능

〈표 6〉 마이그레이션 대상 파일별 위험도

마이그레이션 유형	파일의 전후 위험분석
Lotus 1-2-3 → 최신 Lotus로 마이그레이션	위험 낮음
Lotus 1-2-3 → 엑셀로 마이그레이션	위험 낮음
Lotus 1-2-3 → 아스키로 마이그레이션	위험 낮음

〈표 7〉 Examiner의 에러 보고 예

Examiner가 검토한 파일이 존재하는 디렉터리	/USDA/ftp/usda/data-sets/crops/94018.budget.wk 1:
위험수준	risk level 5
구조적 요소번호 : 구조적 요소명 : 구조적 요소의 간략기술 : 특정 파일에서 발견되는 총 포인트 수	Tag 14 : number : floating point number -Qty : 584 ----

혀 문제없이 이루어졌으며, 엑셀로 마이그레이션하는 것은 위험이 낮았으나, 엑셀은 배치 처리를 못해 변환 소프트웨어의 기능 측면에서 미흡했다. 아스키로 변환할 경우 스프레드 시트의 값은 변환되지만, 스프레드 시트의 기능, 등호, 포인트 등은 손실된다. 이 실험에 따라 상위 버전 마이그레이션과 엑셀 마이그레이션을 우선으로 하고, 두 가지가 불가능할 경우 아스키를 사용할 것을 결정했다.

파일 구조 위험평가를 위해 위에 기술된 수작업 체크는 대규모 파일에는 적합하지 않다. 따라서 위험을 가진 파일을 측정하기 위해 각 파일을 검토하는 Examiner라는 파일 리더 소프트웨어를 준비하였다. 이 프로그램은 파일 내 2개 이상의 위험이 발견되면 파일, 위치, 유형, 위험요소를 보고한다. Examiner는 파일을 읽고, 특정 파일 포맷 요소의 존재와 간기(frequency)를 감지한다. 〈표 7〉은 Examiner가 보고한 보고내용이다. 파일이 존재하는 디렉터리, 위험수준, 구조적 요소번호, 구조적 요소명, 요소의 간략기술, 총 포인트 수가 제시된다. 시간당 1만 개의 로터스 파일을 스캐닝해 대규모 파일의 마이그레이션에서 나타나는 문제를 파악할 수 있다.

2) 변환 프로그램 위험 평가

여러 변환 소프트웨어의 특징과 관련된 위험을 조사하기 위한 실험을 수행했다. 코넬에서는 직접 간단한 기능의 평가 프로그램을 독자 개발하였다. 〈표 8〉은 마이그레이션 프로그램의 위험도를 나타낸 것으로 Lotus 1 2 3을 엑셀로 변환할 때 변환 프로그램별로 위험을 측정한 결과이다. 엑셀 프로그램 이용 시 위험이 낮았으며, 상업적인 프로그램도 위험이 낮았으나, Conversion Plus라는 프로그램에서는 마이그레이션이 적절히 이루어지지 못하였다.

〈표 8〉 마이그레이션 소프트웨어별 위험도

소스 파일	마이그레이션 S/W	대상 파일	위험도
.wk 1	엑셀	xls	낮음
.wk 1	DataJunction	xls	낮음
.wk 1	Conversion Plus	xls	높음

〈표 9〉 위험관리 사례 비교

비교항목	OCLC	CUL
목적	위험요소 및 파일 포맷의 특성 파악	마이그레이션의 여부 결정 변환 파일 포맷의 결정 변환 프로그램 평가
실험대상 파일 포맷	이미지 파일 JPEG, TIFF	수치 파일 Lotus 1-2-3
실험결과	• 설문조사를 통해 위험이 큰 요소 조사 • 파일 포맷의 지속적 추적을 통해 적합한 파일 선정	• 마이그레이션의 수행 결정 • Lotus를 엑셀, 상위 로터스로 변환은 적합, 아스키는 부적합 • 변환 프로그램의 적합성 테스트를 통해 변환 프로그램 결정

3.3 비교분석

위험관리의 사례로 OCLC와 코넬대학의 사례를 조사한 결과 〈표 9〉와 같이 특징을 정리할 수 있다.

OCLC의 INFORM 연구에서 설문결과나 조사된 파일의 특징을 세부적으로 기술하고 있지는 않지만, 위험관리에 대한 방향을 제시해 주고 있다. 실험에서 이미지 파일 중 JPEG, TIFF 파일의 특징을 조사하였다는 점과 지속적인 추적을 통해 마이그레이션할 파일을 결정한 것은 위험관리에서 필요한 방법이다.

코넬대학의 연구는 마이그레이션이 디지털 자원의 위험에 큰 영향요인임을 인식하고 마이그레이션 전후의 파일 포맷에 대한 실험을 수행하였다. 파일의 구조적, 내용적 요소의 적절한 변경이 이루어졌는 지를 조사하고, 마이그레이션 소프트웨어 중 위험이 낮은 소프트웨어를 실험을 통해 증명하고, 파일의 적절한 변환을 검토하기 위한 프로그램을 작성하였다.

OCLC의 경우 주로 이미지 파일을 조사하였으며, 파일의 지속적인 추적을 통해 위험이 낮

은 파일 포맷을 파악하며, 코넬대학은 파일 포맷 중 수치 파일을 대상으로 마이그레이션 전후를 조사하고, 마이그레이션 프로그램의 적합성 여부를 실험으로 증명하였다. 이 연구들은 미래에 나타날 수 있는 파일 포맷의 위험을 증명하고, 위험을 줄이기 위한 방안과 실천을 모색한다.

4. 국내에서 위험관리 수행방안

4.1 위험요소 규명

디지털 자원의 위험요소를 규명하고, 위험요소의 평가와 분석을 통해 실질적인 실행계획이 마련될 수 있다. 국내에서 위험관리에 대한 중요성이 인식되고 있지는 않으나, 해외의 연구를 기반으로 국내에서 수행전략이 신속히 마련될 수 있을 것이다. 위험관리의 단계를 도입하고, 위험요소를 규명하고, 분석 평가한 후 실질적인 위험관리 실행의 단계를 거쳐야 할 것이다.

위험관리의 기초는 디지털 자원과 아카이빙에 관련된 위험요소를 정확히 규명하는 것이다. 이미 연구된 해외의 연구에서는 디지털 포맷, 소프트웨어, 하드웨어, 관련조직, 디지털 아카이브(아키텍처, 프로세스, 조직), 보존전략으로 위험요인을 분류하고 있으며, 해당 요인별로 평가가 수행되어 이를 위험관리 결정과 아카이빙에 사용할 수 있다.

해외 사례에서 밝혀진 위험요인과 우리나라의 환경을 조사하여 국내의 디지털 자원의 위험요소를 밝혀야 한다. 또한, 위험요인 규명 및 위험요인에 따른 마이그레이션 등의 보존전략 실행여부 등을 결정하기 위한 평가척도가 개발되어야 할 것이다.

4.2 위험요소별 실험 테스트 실시

해외의 경우 아카이빙에 사용할 파일을 결정하기 위해 위험관리 중 파일 포맷의 위험평가가 실시되고 있다. 코넬의 실험에서는 로터스 파일을 로터스 상위 버전이나 엑셀, 아스키로 마이그레이션되었을 때 위험의 정도를 파악하여 로터스와 엑셀을 권고 포맷으로 결정하고, 두 가지 포맷의 사용이 불가능할 때 아스키로 할 것을 결정하였다. Digital Preservation Testbed(DPT)에서는 가능한 원본 파일을 유지하면서 PDF와 함께 아카이빙을 결정하였다. 해외에서는 파일 포맷을 실험 테스트하고 있으나 국내의 토종 파일에 대한 실험 테스트는 전무하다.

예를 들면, <표 10>과 같이 과거 훈민정음과 보석글, 한글의 낮은 버전을 마이그레이션 프로그램을 이용해 변환시킨 후, 내용과 구조 형태 기능 등이 적절한 지를 수치화해야 한다. 전담기구에서 국내의 파일 포맷을 이용한 위험 테스트 및 마이그레이션 소프트웨어의 위험 테스트를 수행하고 적절한 파일 포맷을 권고해야 한다.

〈표 10〉 국내 파일 포맷의 위험평가 가상 사례

	변환 포맷		변환 소프트웨어	
	실험조건	실험결과	실험조건	실험결과
보석글	보석글 → 한글	위험 높음	한글 프로그램	위험 높음
	보석글 → 아스키	위험 낮음	별도 작성 프로그램	위험 낮음
훈민정음	훈민정음 → 한글	위험 높음	한글 프로그램	위험 높음
	훈민정음 → 아스키	위험 낮음	별도 작성 프로그램	위험 낮음
한글 3.0	한글 3.0 → 한글 상위 버전	위험 낮음	한글 프로그램	위험 낮음
	한글 3.0 → 아스키	위험 낮음	별도 작성 프로그램	위험 낮음
	한글 3.0 → MS 워드	위험 높음	MS 워드	위험 높음

4.3 위험관리 수행 전담기구 조직

디지털 아카이빙을 위한 테스트베드를 개별 기관에서 수행하기보다 중앙 센터에서 수행해야 하는 것처럼 위험요인의 평가도 중앙기관에서 실험하고, 실험을 기반으로 위험관리 수행 가이드라인을 마련하고 법제를 정비도록 해야 한다.

예를 들어, 디지털 아카이빙 체제에서 정책기구로 디지털 큐레이션 센터(DCC) 등에서 여러 실험을 통해 적절한 가이드라인을 만드는 것과 같이 중앙통제기구로 DCC와 같은 기관에서 위험관리의 실행과 이를 통한 정책을 마련해야 한다. 특히 아카이빙을 위한 테스트베드 실험과 함께 위험요인의 평가분석을 수행하여 국가적인 권고를 작성할 수 있다. 개별 기관은 중앙기관의 실험결과를 기초로 적절한 파일을 선정하고, 마이그레이션 프로그램을 선택할 수 있다.

5. 결 론

지금까지 위험관리의 정의, 위험관리와 디지털 아카이빙 및 보존전략과의 관계, 위험관리 과정, 실험과 테스트 파일을 이용해 위험관리의 일부인 위험평가를 수행한 OCLC와 코넬대학의 사례를 살펴보았다. 이를 바탕으로 국내 환경에 적용할 방안을 모색하였다.

위험관리란 미래 사건의 가능성에 대처하기 위한 다양한 분석과 대처기법이며, 디지털 아카이빙의 일부이다. 즉, 디지털 자원의 미래 접근을 보장하는 것이 디지털 아카이빙이며, 디지털 아카이빙에 내재하는 위험을 최소화하는 것이 위험관리이다. 위험관리는 위험평가로 종료되는 것이 아니라, 디지털 아카이빙의 전체 과정에서 다양한 위험요소를 규명하고 이를 평가해야 한다. 위험평가 수행을 통한 결과는 정책에 반영되어, 보전전략을 수정하고, 보존기법을 변경하고, 프로그램을 수정하는 등의 작

업이 동반되어야 한다.

사례에서와 같이 OCLC, 코넬대학, 사례로 넣지는 않았으나 호주국립도서관의 경우 신속한 연구와 실험 데이터가 작성되고 있으나, 우리나라의 경우 디지털 아카이빙과 위험관리에 관한 연구가 시작되고 있다. 특히, 호주와 코넬은 웹 자원의 위험관리를 위해 지속적인 모니터링 프로그램을 개발 · 연구하고 있으며, OCLC와 코넬은 위험요소를 규명하고, 개별 파일 포맷이 마이그레이션에 얼마나 안전한 지, 마이그레이션 소프트웨어의 적합성에 대한 연구를 수행하였다. 앞으로 위험관리는 기술의 발전과 디지털 아카이빙의 보존전략의 변화에 따라 많은 연구와 실험이 진행되어야 할 것이다.

참고문헌

서은경. 2003. 디지털 정보자원 보존의 위험관리분석. 『정보관리학회지』, 20(1): 5 29.

설문원 외. 2005. 『국가 디지털 아카이빙 체제 구축에 관한 연구』. 서울: 한국과학기술정보연구원.

Botticelli, P. 2003. "Risk management for web resources: a case study on southeast Asian web sites." *RLG DigiNews*, 7(1). [cited 2005. 10. 20]. 〈http://www.rlg.org/preserv/diginews/diginews7 1.html#feature2〉.

DPT. 2003. "Preserving text documents." *From digital volatility to digital permanence series* 3. [cited 2005. 10. 20]. 〈http://www.digitaleduurzaamheid.nl/bibliotheek/ docs/volatility permanence textdocs en.pdf#search=' Preserving%20text%20documents.%20%28From%20digital%20volatility%20to%20digital%20permanence%20series%29〉.

Erpanet. 2003. *Risk communication tool*. [cited 2005. 10. 20]. 〈http://www.erpanet.org/guidance/docs/ERPANETRiskTool.pdf〉.

Kenney, A. R. et al. 2002. "Preservation Risk Management for web resources : VRC in Cornell's project prism." *D Lib magazine*, 8(1). [cited 2005. 10. 20]. 〈http://www.dlib.org/dlib/january02/kenney/01kenney.html〉.

Lawrence, G., et al. 2000. *Risk management of digital information: a file format investigation*. Washington: CLIR. [cited 2005. 10. 20]. 〈http://www.clir.org/pubs/reports/pub93/pub93.pdf#search=' Risk%20management%20of%20digital%20information%3A%20a%20file%20format%20investigation〉.

McGovern, N. Y. 2004. *Virtual Remote Control : building a preservation risk management toolbox for web resources*. [cited 2005. 10. 20]. 〈http://www.dlib.org/dlib/april04/mcgovern/04mcgovern.html〉.

Rieger, Oya Y. and A. R. Kenny. *Risk management of digital information case study for image file format*. [cited 2005. 10. 20]. 〈http://www.library.cornell.edu/imls/CLIRImageStudy.pdf〉.

Stanescu, A. 2004. "Assessing the durability of formats in a digital preservation environment." *D Lib magazine*, 10(11). [cited 2005. 10. 20]. 〈http://www.dlib.org/dlib/november04/stanescu/11stanescu.html〉.

Van Diessen, R. 2002. "Preservation requirements in a deposit system." *IBM/KB long term preservation study series 3*. [cited 2005. 10. 20]. 〈http://www.kb.nl/hrd/dd/dd_onderzoek/reports/3 preservation.pdf〉.