

디지털 정보자원 보존의 위험관리 분석: 대학도서관 전자정보실 중심으로*

Risk Management of Digital Information Resources Preservation in University Libraries

서 은 경(Eun-Gyoung Seo)**

초 록

우리나라 대학도서관에서는 디지털 패러다임에 적합한 디지털 정보자원을 구축하여 서비스를 제공하고 있으며 이를 위하여 디지털 보존에 힘써왔지만, 아직까지 본격적으로 연구되고 논의되지 않고 있다. 디지털 보존은 인쇄매체 보존 방법과 상이하게 다르므로 디지털 정보자원의 안정성과 기능성을 최대한 확보, 유지할 수 있도록 보존전략을 강구해야 할 필요가 있다. 본 연구는 대표적인 디지털 정보자원의 보존기술인 매체재생, 매체변환, 포맷변환, 정보전환, 에뮬레이션 기술의 대학도서관에서의 사용정도와 대학도서관의 디지털 정보관리 담당자들이 생각하는 각각의 보존 기술에 대한 위험가능성과 위험영향력을 조사하여 위험관리 분석을 하였다. 그 결과, 가장 빈번히 사용되고 있는 기법은 포맷변환이고 전혀 사용하지 않고 있는 기법은 에뮬레이션임을 알 수 있었다. 또한 위험이 발생할 확률과 데이터 손실 영향력이 가장 낮은 방법은 매체재생인 것으로 판명되었고 가장 높은 것은 에뮬레이션 기법과 정보전환으로 나타났고 사서들은 디지털 보존활동을 전문적으로 수행할 전담 담당자가 없는 경우 위험발생 확률이 제일 높을 것으로 보았다.

ABSTRACT

University libraries have developed digital information resources for digital services and have tried to preserve valuable digital information. Because digital preservation raises challenges of a fundamentally different nature which are added to the problems to preserving traditional format materials, it is necessary that preserving digital resources must be discussed and researched actively. This study conducts risk management for exploring preserving technologies and assessing tools. It is found that most university libraries have used format conversion frequently and do not use the emulation technology. It is also found that medium refreshing has the lowest risk probability and risk impact, but information migration and emulation technology have the highest in the risk probability and in the risk impact individually and the absence of full-time professional staff causes high risk.

키워드: 디지털 보존, 위험관리, 보존전략, 디지털 보존기술, 매체재생, 매체변환, 포맷변환, 정보전환, 에뮬레이션, 위험가능성, 위험 영향력, digital preservation, risk management, risk probability, risk impact

* 이 논문은 2002년도 한성대학교 교내연구비 지원과제임

** 한성대학교 지식정보학부 부교수(egseo@hansung.ac.kr)

■ 논문 접수일 : 2002. 11. 30

■ 게재 확정일 : 2002. 12. 16

1 서 론

정보기술의 발전은 사회 변화 및 인간의 일상생활에 엄청난 영향을 끼쳐왔고 그 영향력은 앞으로도 더욱 커질 것이다. 특히 인류의 문화유산인 정보를 축적, 유지 계승하고 있는 도서관의 역할과 기능은 디지털 패러다임이 강조되고 있는 환경 속에서 그 어느 때 보다 급속하게 변화되어 가고 있다. 이제 인쇄매체가 아닌 전자매체에 구축된 정보자원은 도서관의 주요 장서 유형의 하나로써 인지되고 있음은 물론이고 그 중요성과 효용성이 강조됨에 따라 도서관에서의 디지털 정보자원의 규모는 점차적으로 확장되고 있다.

디지털 정보자원은 전통적인 인쇄매체에 근간을 둔 정보자원과는 생성, 소장/축적, 접근/검색, 전송/배포, 보존, 이용면에서 크게 다르다고 할 수 있다. 현재 이용되고 있는 디지털 정보자원으로는 전자잡지, 전자책, 웹사이트, 디지털 사운드, 디지털 텔레비전, 디지털 동영상 그리고 도서관에서 구축된 각종 디지털 문서들 등을 들 수 있는데 이러한 디지털 정보는 다양한 포맷 형태로 생성/축적되고 있는 실정이다. 따라서 디지털 정보자원에 접근하기 위해서 이용자는 구축된 포맷 형태 각각에 대한 이해가 필요하며 도서관은 각각의 정보원을 처리할 수 있는 특정 소프트웨어는 물론 다양하며 강력한 하드웨어까지 갖추어야 한다. 또한 디지털 정

보자원 이용에 필요한 소프트웨어와 하드웨어가 어느 정도 시일이 지나 개정되어 그 기능과 사용방식, 플랫폼 등이 바뀔 때마다 새로운 환경에서도 쉽게 이용될 수 있도록 도서관은 정보자원을 이전 또는 변환시켜야 하므로, 디지털 정보의 이용은 더욱더 복잡해질 뿐만 아니라 디지털 정보자원의 보존 역시 어려운 도서관 활동으로 부각되고 있는 실정이다.

디지털 정보자원의 계속적인 이용을 위해서는 디지털 정보 즉 디지털 콘텐츠 자체뿐만 아니라 디지털 정보자원이 생성시킨 모양과 느낌(look and feel)까지도 유지시켜야 하며(Hodge 2000) 동시에 디지털 정보를 사용케 해주는 여러 가지 속성, 즉 액세스, 검색, 처리가 가능하게 하는 메커니즘, 역동적이고 상호작용적인 전달 기능, 이용자 기반의 원문 재생 기능 등이 사라지지 않고 유지되어야 진정한 보존이라는 생각이 보편적이다(Bikson 1998). 그러나 이와 같이 콘텐츠와 여러 기능들을 디지털 정보자원의 생명주기별 보존 관리하여 미래 이용자들이 어려움 없이 정보를 활용할 수 있도록 하는 전반적인 작업을 “디지털 아카이빙(digital archiving)”으로 부르고 “디지털 보존”은 디지털 아카이빙의 하위 기능으로 정보자원이 지닌 콘텐츠와 모양과 느낌을 계속적으로 유지 보존시켜 주는 기술적 활동으로 간주되고 있다.

디지털 아카이빙을 위한 다양한 기능 중 디지털 보존은 다른 기능보다 매우 중

요한 관리 요소로 부각되고 있지만 그 방법은 상당히 난해한 편이다. 그 이유는 여러 측면에서 분석할 수 있지만, 정보기술에 절대적으로 의존하는 보존방식으로 인하여 정보기술이 지닌 속성의 한계점(시간이 지남에 따른 시그널 감쇄, 디지털 정보를 담고 있는 매체의 빠른 노화¹⁾, 대체기술의 잦은 출현현상으로 오는 소프트웨어와 하드웨어의 단명성²⁾)을 고스란히 지닌다는 점과 디지털 정보자원의 구축 포맷이 각각 다르다는 점, 이 두가지가 유지 보존기능을 상당히 어렵게 하고 있다(Friedlander 2002). 그러나 이러한 속성들로 인하여 디지털 데이터에 대한 가독성은 저하되고 또는 부분적으로 기능이 상실되고 있으므로 그 방법이 어렵다 하더라도 주기적인 디지털 보존 활동은 불가피해진 실정이다.

현재 우리나라 대부분의 대학도서관에서는 디지털 패러다임에 적합한 디지털 정보자원의 생산, 축적, 공유, 검색 및 이용을 위한 환경을 제공하는 데에 그 목적을 두고 구축 및 서비스를 제공하는 반면 디지털 보존의 중요성은 간과되고 있는 실정이다. 그러나 디지털 정보자원을 보존하기 위하여 새로운 투자와 정책이 필요하다는 것을 인식하고 있는 외국의 디지털도서관과 아카이브 센터는 디지털 정

보자원 보존에 대한 국가적 관리정책, 보존 프로그램의 모형화 그리고 기금마련에 대한 논의를 본격적으로 진행하고 있는 반면(Brown 2000, Hodge 2000, Thibodeau 2001, Friedlander 2002), 우리나라에서는 디지털 보존에 대한 필요성을 인지하고는 있으나 아직까지 구체적인 논의 및 연구가 진행되고 있지 않고 있다. 어느새 디지털 정보자원을 생성하고 이용한지가 5년 이상 되고 있는 상황에서 보면 대학도서관은 디지털 정보자원의 안정성과 기능성을 최대한 확보, 유지하는 방향으로 보존전략을 강구해야 할 필요가 있다. 비록 보존활동을 수행하고 있더라도 다양한 디지털 정보자원의 출현, 도서관 여건의 변화, 정보기술의 진보 등에 의하여 지금 사용하고 있는 보존방식이 절대적일 수도 없으며 최선의 대안이 아닐 수도 있다. 따라서 현재 사용되고 있는 보존기술을 평가하고 각각의 기술이 지닌 문제점과 대안점을 파악할 필요가 있다. 더 나아가 장기적 디지털 보존을 위한 기술적 해결책이 연구되고 논의되지 않는다면 상당한 많은 양의 정보가 손실될 것이며 엄청난 비용이 투자될 것이므로 디지털 정보자원의 장기적 유지 및 관리에 대한 연구가 본격적으로 수행되어야 할 것이다.

본 연구는 디지털 정보자원의 보존기술

-
- 1) Rothenberg(1995)은 정보매체별 노화기간 및 물리적 수명을 짧게는 1년에서 길게는 30년 정도로 추정하였다.
 - 2) Hodge(2000)는 하드웨어와 소프트웨어의 전환시기가 빠르면 2년에서 길면 10년 정도 되는 것으로 조사하였다.

로서 빈번히 사용되고 있는 매체재생(medium refreshing), 매체변환(medium conversion), 포맷변환(format conversion), 정보전환(information migration), 에뮬레이션 기술(emulation technology) 중, 현재 대학도서관에서는 어떤 기술들이 디지털보존을 위하여 사용되고 있으며 대학도서관의 디지털 정보관리 담당자들이 생각하는 각각의 보존 기술에 대한 위험가능성과 위험영향력을 조사하였다. 즉 디지털 정보자원 보존기술에 대한 위험평가를 측정하여 디지털 정보자원에 대한 위험관리(risk management)의 제반 문제점을 파악하고 더 나아가 대학도서관에 적합한 보존정책 및 기술들을 제안하고자 한다. 본 연구의 궁극적인 목적은 도서관이 사용해야 할 보존기술에 대한 제안과 더 나아가 장기적인 디지털 정보자원 보존정책을 수립하는데 있어서 기초적 기반을 제공하는데 있다.

2 디지털 정보자원 보존

2.1 선행연구

디지털 정보자원의 보존관리에 대한 국내 연구는 매우 적다. 기록물 보존에 관한 관심도가 최근 높아지긴 하였으나 아직까지 디지털 정보자원 보존에 대해서는 많은 연구가 없는 실정이다. 최원태(2001)는 기술적, 경제적, 국가적인 관점에서 디지털 아카이브의 현황을 조사하였

으며 디지털 정보자원의 장기보존과 관련된 배경과 이슈들에 관하여 분석하였다. 특히 디지털 아카이브의 표준화 모형, 디지털 정보의 선택 및 컬렉션 개발, 디지털 정보자원의 구분체계, 디지털 정보의 기술방법에 관한 현황과 문제점을 분석하였다. 그는 디지털 정보자원의 효과적인 보존을 위해서는 디지털 정보의 생산, 수집 등의 초기 과정에서의 단계별 관리와 미래의 계속 접근을 보장하기 위한 지속적인 관리가 필요하다고 하였다. 그리고 디지털 아카이브 구축을 위한 표준화 모델 수립, 기술진보에 따른 기술적 노후현상 문제, 디지털 정보의 지적소유권 문제, 장기 보존을 위한 비용예측 등이 효율적인 디지털 정보자원의 보존을 위하여 먼저 해결되어야 한다고 하였다.

윤희윤(2001)은 국가도서관의 전자출판물 수집 및 보존방안을 제시하는 연구를 수행하였다. 특히 그는 디지털 정보자원 보존의 난해성과 장애요소를 분석하였고 이에 따른 디지털 보존의 전략적 방안과 디지털 보존의 경제성을 분석하였다. 모든 자료의 이용성과 보존성을 동시에 지향해야 하는 국가도서관은 국가적 보존정책 수립, 표준 포맷과 같은 보존과 관련된 표준 모델 제안, 정기적 매체변환 계획 수립, 협동 보존 계획 및 디지털 전략 수립 등을 추진해야 하고 이를 위해서 전자출판물과 관련된 제도적 장치(관계법령의 제정, 저작권 문제 해결 등)를 보완하고, 정책적 과제(디지털도서관 건립, 장서

정책의 성문화, 보존방식의 결정)를 추진 하며, 기술적 문제(원형보존, 매체변화이 나 기술전화, 축적포맷, 데이터베이스의 구조 등)을 해결하여야 한다고 주장하였다. 신은자(2001)는 전자저널의 아카이빙에 대한 연구를 수행하였다. 검색의 다양성과 이용의 편리성 면에서 전자저널의 아카이빙을 선호하고 있으나 장기적인 보존 문제 때문에 안정된 인쇄저널을 선호하고 있는 문제점을 지적하면서 전자저널 아카이빙 구축 및 보존을 위한 모형 개발 및 국가적 정책을 제안하였다.

외국에 경우는 우리나라보다 약 5-7년 정도 앞서 본격적인 디지털 보존에 관한 연구를 진행하였다. 먼저 디지털 아카이빙에 관한 대단위 국가적 프로젝트를 살펴보면 다음과 같다. 미국의 보존위원회(Commission on Preservation and Access)와 미국 연구도서관그룹(Research Library Groups)이 1996년에 수주한 *Task Force on Archiving of Digital Information*은 미국뿐만 아니라 세계적으로 디지털 아카이빙의 문제점을 규명한 최초의 사업이고 다른 프로젝트 착수에 도화선이 되었다. 이 연구는 디지털 아카이빙에 관한 폭넓은 정책을 수립하였고 교육 및 연구와 관련된 디지털 정보라면 현재 물리적으로 분산되어 있더라도 공동으로 관리하고 장기적으로 보존할 수 있도록 다양한 보존 기술과 보존모형을 제안하였다(Garret & Waters 1996). Consultative Committee for Space Data Systems(CCSDS)는 1999년

에 디지털 보존에 대한 방대한 내용의 보고서를 제출하였다. 보고서는 디지털 정보의 장기적 보존과 접근, 즉 디지털 아카이빙 개념에 대한 이론적 기초를 제시하였고 현재 운영되고 있는 아카이브와 미래의 아카이브를 구조와 운영면에서 비교하면서 앞으로의 디지털 아카이브의 역할과 기능을 제안하였다.

영국의 JISC(Joint Information Systems Committee)는 "Digital Preservation Focus" 부서를 세워 디지털 정보자원 보존에 대한 정책과 전략을 수립하고 실제 실행하고 있으며, 1998년부터는 CEDARS(CURL Exemplars in Digital Archives)라는 디지털 아카이빙 프로젝트를 수행하고 있다(Stone & Day 1999, Russell 2000). 이 프로젝트는 elib 프로그램의 일환으로 상세한 디지털 자원 보존정책을 수립하고 구체적인 디지털자원 메카니즘 구축을 목표로 하고 있으며 3개년 계획으로 추진되었다. 특히 디지털 정보자원의 보존, 저작권 관리, 메타데이터 추적, 에뮬레이션 등에 관하여 상세한 분석과 지침을 제공하고 있으며 아울러 각각의 문제점을 파악하고 해결책을 제시하고 있다. 또한 영국의 문화유산 센터(English Heritage Center for Archaeology: CfA)는 디지털 정보자원 보존 전략은 디지털 정보의 생명주기에 따라 세워져야 효율적이라는 전제아래 각각의 생명주기 단계별로 보존관리 전략을 상세하게 제시하였다. 특히 생명주기 단계 중 생성 및 획득의 단계를 보존 전

단계로 구분하여 전략을 제시하였고 이어서 그 다음 생명주기 단계인 소장, 접근, 활용에 적합한 보존전략을 각각 제시하였다(Brown 2000).

캐나다 국립도서관은 1994년부터 캐나다에서 발행된 디지털 정보자원을 아카이빙하는 목적으로 EPPP(Electronic Publication Pilot Project)를 착수하였다. 캐나다의 전자출판물을 수집, 정리, 보존, 그리고 접근이 가능하도록 하기 위하여 이와 관련된 이슈와 문제점들을 파악, 해결하는 것을 목표로 하고 있다. 특히 온라인 장서를 취급시에 도서관에서 부딪치는 이슈를 파악하고 온라인 장서의 처리 및 보존에 대한 중장기 정책을 수립하고 이와 관련된 기술을 개발하는 것이 EPPP의 주요 목적이라 할 수 있다(NLC 1996).

호주에서는 호주국립도서관을 중심으로 표준화된 국가적 디지털 아카이브 모델을 개발하고 보존정책을 제안하고 있다. 호주국립도서관이 실시하고 있는 PANDORA(Preserving and Accessing Networked Documentary Resources of Australia) 프로젝트는 국가적인 보존 차원에서 호주와 관련된 디지털 정보의 장기적인 접근을 위한 절차를 개발하고 구현하는데 그 목적을 두고 있다. 이 프로젝트는 보존시스템의 기능 구분, 보존 과정과 관련된 방법론 및 표준화의 평가 보존비용 장비 인력의 추정, 디지털 보존을 위한 전략 개발, 디지털 보존을 위한 국가적인 모형 개발 등과 같은 사항을 연구하고 있다

(NLA 2001). 특히 PANDORA의 하부 프로젝트인 PADI(Preserving Access to Digital Information)는 디지털 형태의 정보들이 잘 보존되어 쉽게 접근할 수 있도록 하는 여러 매커니즘을 제공하는 것이 주 목적으로 디지털 보존 기술, 정책, 저작권, 표준, 포맷과 매체 등에 관한 이슈와 그 방법에 대하여 연구하고 있다(Berthon et al. 2002).

이와 같이 각국은 장기적 디지털 정보 지원 관리를 위하여 국가적인 디지털 아카이빙 프로젝트를 수행하고 있다. 또 한편으로 디지털 정보자원의 장기적 유지를 위한 적합한 보존기술을 제안하기 위하여 몇몇 연구는 도서관이 실제 사용하고 있는 보존기술을 심층적으로 평가하였다. Bennett(1997)는 디지털 정보의 종류, 매체, 플랫폼에 따른 보존시 요구사항을 분석한 후 초보적인 위험평가를 실시하였다. 그 결과 디지털 매체의 복잡성과 상호의존성이 높을수록 요구사항도 복잡하고 다양한 것으로 나타났다.

Rothenberg(1999)은 미국의 보존위원회와 도서관정보자원 협의회(CLIR)의 디지털 도서관 프로그램의 지원을 받아 디지털 보존을 위한 실행 가능한 보존기술을 제안하였다. 그는 먼저 디지털 장기보존의 문제점과 디지털시대에서 보존의 중요성을 제시한 다음, 현재 사용되고 있는 네 가지 보존방식(인쇄형태로 변환, 가독성을 유지시키기 위한 표준방식으로의 전환, 박물관에 있는 기술이용, 정보전환)에

대한 타당성을 조사 분석하였다. 그는 현재의 기술들이 노동집약적 방식이며 진본 형태로 디지털자료를 보존할 수 없다는 큰 문제점을 지적하고 이러한 문제점을 극복할 수 있는 방식으로 디지털 문헌이 지닌 원기능, 모양과 느낌을 유지시킬 수 있는 에뮬레이션방식을 이상적인 해결책으로 제안하였다.

CAMiLEON(Creative Archiving at Michigan and Leeds: Emulating the Old on the New)은 미국의 국립과학재단(NSF)와 영국의 JISC의 재정 지원을 받아 미국 미시간 대학 문헌정보학 대학원과 영국의 Leeds대학과 협동연구로 수행되는 프로젝트로 디지털 보존전략으로 제시되고 있는 에뮬레이션 기법을 평가하는 것을 목표로 하고 있다(Granger 2000). 그동안 에뮬레이션 기법에 대하여 너무 이론적이고 가설 중심으로 논쟁이 있어 왔음에 따라 실제 에뮬레이션이 얼마나 효과적이며 구현 가능한지에 대한 검증을 수행하기 위하여 실시된 프로젝트이다. 따라서 현재 사용 중인 에뮬레이터 평가, 에뮬레이터 개발에 대한 조사, 실제 에뮬레이션을 사용 중인 기관에 대한 사례연구, 다른 기법과의 이용자만족도, 비용 대 효과 분석을 수행하고 있다. Lawrence et al.(2000)은 CLIR의 지원을 받아 도서관에서 많이 사용하고 있는 정보 전환시에 발생하게 될 위험을 평가하였다. 특히 정보전환 기법을 사용할 때 디지털 정보원 내용면에서 나타날 위험을 전반적으로 파

악하였고, 다음 파일포맷 변환에 대한 위험평가를 소프트웨어 전환시 발생될 위험, 방대하고 서로 다른 포맷으로의 변환시 발생될 위험, 파일전환 과정에서 발생될 위험, 메타데이터 작성과 관련된 위험 등 4가지 면에서 실시하였다. 그 외 두 가지 포맷(Lotus1-2-3과 wk1)의 변환시 발생되는 위험요인을 비교하는 사례연구와 이미지 파일포맷 변화시 발생되는 위험을 분석하는 사례연구도 실시하였다. 그 결과 정보전환 기법은 무시할 수 없는 위험을 가지고 있음이 판명되었고 위험의 종류도 다양하고 어떤 것은 심각한 정도로 나타났음을 밝혔다.

Kenney et al.(2002)는 코넬 대학의 PRISM(Preservation, Reliability, Interoperability, Security and Metadata)프로젝트의 한 부분으로 웹자원 보존의 위험관리 연구를 수행하였다. 특히 위험관리가 4단계(위험요인 수집 및 카테고리화; 단순한 위험요인 파악 및 탐지; 전체적 위험요인 파악 및 탐지; 보존정책 자동적 강화)로 수행될 수 있도록 하였고 이에 따른 전략을 제시하였다. 이들은 계속적으로 웹자원을 이용하기 위해서는 독립적인 웹페이지, 하이퍼링크로 연결된 웹페이지, 웹사이트들을 각각 다른 방식으로 모니터하고 관리해야 하는 이유와 그 방법도 제시하였다.

2.2 디지털 보존전략

디지털도서관으로의 전환을 시도하고

있는 많은 대학도서관은 디지털 정보자원을 많이 활용되도록 노력하고 있으나 디지털 정보자원의 장기적 보존에 대한 기술적 해결책을 확정시키지 못한 채 있다. 다만 몇몇 연구자들이 디지털 보존방법을 제안하고 있다. Rothenberg(1999)은 인쇄자료로의 변환, 표준 이용, 컴퓨터 박물관 활용, 정보전환, 그리고 에뮬레이션기술 이용을 제안하였고 윤희운(2001)은 매체 재생, 매체변환, 포맷변환, 기술전환과 같은 네 가지 방안을 제시하였다. 또한 Arms(2000)는 디지털보존 전략으로 더 낫은 매체로의 변환(better media), 비트 재생(refreshing bits), 내용전환(migration of contents), 기술환경 에뮬레이션(emulation of technology environment), 디지털 유물화(digital archeology)와 같은 다섯 가지 방법을 제시하였다.

이와 같은 환경 속에서 몇몇의 가치 있는 데이터들이 매체나 소프트웨어가 퇴화됨에 따라 손실될 가능성에 놓이게 되자 도서관은 100% 만족은 하지는 않으나 데이터 손실을 방지하기 위하여 여러 가지 보존방식을 채택 사용하고 있다. 첫째 가장 손쉬운 방법으로 디지털 정보를 축적한 매체가 한계수명에 도달하면 동일한 기술을 사용하여 동종의 새로운 매체에 복제하는 “매체재생”을 들 수 있다. 사용된 매체가 계속해서 이용된다면 매체노화로 인한 데이터 손실을 방지할 수 있으나, 매체 그 자체가 정보기술에 진보에 따라 진화되어 성능이 떨어지는 그전의

매체를 사용하지 않은 경우에는 “매체변환”이 필요하게 된다. 따라서 두 번째 방식은 매체의 안정성이 떨어지거나 비표준적인 매체를 현재의 안정적이고 표준매체로 변경시켜주는 방법을 말한다. 그 예로 디스켓에 수록된 정보를 마이크로필름, CD-ROM, 호스트 컴퓨터 서버에 축적하는 방식을 들 수 있다. 세 번째 방식으로 원본의 디지털 문헌에 수록된 지적 내용을 그대로 유지하면서 당초의 포맷을 다른 포맷으로 바꾸는 방식인 “포맷변환”을 들 수 있다. 디지털 문헌에 적용되고 있는 축적포맷(ASCII, word Processor Format, Print Format, Database Format, PDF, XML 등)을 다른 포맷으로 바꾸는 프랜스포매팅(transforming), 리포매팅(reformatting)을 지칭한다(윤희운 2001). 포맷변환을 실시할 경우 원본의 파일 구조가 변경될 수도 있고 또는 정보의 문맥이나 데이터간의 내재적 관계를 잃어버릴 수는 있으나 원 데이터 즉 콘텐츠는 계속 유지될 수 있는 기법이다. 따라서 스프레드시트의 데이터가 ASCII로 전환될 때 셀 안에 있는 수치정보는 살아 있으나 셀 안에 내재된 공식과 다른 관계는 잃어버리게 되는 위험이 따르게 된다.

네 번째 방식으로 “정보전환”을 들 수 있다. 정보전환 기법은 기존의 디지털 정보를 새로운 하드웨어 및 소프트웨어에 수용하도록 하여 디지털 정보의 작동 및 검색환경을 바꾸는 방식을 말한다. “*Preserving Digital Information*” 프로젝트

(1996)에서는 정보전환을 한 하드웨어와 소프트웨어 환경에 담겨진 디지털 자료를 다른 것으로 또는 현 컴퓨터기술 세대의 것을 다음 세대의 것으로 정기적으로 옮겨주는 것으로 정의내리고 있다. 즉 기존의 이미지를 위한 디지털 포맷(예, TIFF)를 보다 기능이 강화된 현 표준포맷으로의 변환시키거나, 또는 SGML 문헌을 새로운 표준포맷 문헌으로 변환시키는 것이 이에 해당된다(Green et al. 1999). 강화된 포맷으로의 정보전환은 디지털 재생산의 기능성을 강화시킨 것이라 할 수 있다. 이러한 전환기술에 대하여 Rothenberg (1999)는 노동집약적이며 시간과 비용이 많이 드는 방식일 뿐만 아니라 대규모 전환이 불가능하고 에러가 잘 발생할 수 있는 기법이라 하여 추천하지 않고 있다. 그러나 현재 도서관에서 가장 많이 사용되고 있는 기법으로 정보전환시 데이터 손실 위험을 줄이기 위하여 메타데이터를 적극적으로 활용하고 있다.

다섯 번째 방식은 “에뮬레이션 기술”이다. 이 방식은 소프트웨어와 하드웨어 에뮬레이터를 채택하여 퇴화된 시스템을 흉내 내게 함으로써 디지털 자료는 원래의 포맷을 유지한 채 보존하면서 이용자가 그전의 방법으로 접근할 수 있게 하는 것이다. 에뮬레이션 기술은 원래의 소프트웨어 환경을 보존하는 방법으로 데이터 파일과 함께 소프트웨어를 운영시켜주는 하드웨어까지 전환시켜줌으로 새로운 환경에서도 콘텐츠는 물론이고 이전의 모양

과 느낌을 지니게 해준다. 그 대표적인 예가 MS-DOS에서 윈도우 95/98/NT로 전환시킨 것을 들 수 있다. 가장 강력하고 최신의 기법으로 Rothenberg가 적극적으로 추천한 기법이지만, 비용 대 효과면에서 실용적인지 또 모든 디지털 데이터를 원본의 모양과 느낌으로 보존해야 하는가에 대한 의문을 제기하면서 적합한 디지털 포맷과 정보전환 계획을 수립 선택하여 필요한 경우에만 에뮬레이션 기술을 활용해야 한다는 주장이 현재로서는 보편적이다.

3 디지털보존에 대한 위험평가

위험관리는 불확실한 환경에서 생겨질 수 있는 다양한 결과 및 “걱정반경(worry radius)”, 이에 따른 인지된 위험과 허용 손해를 확률적으로 파악하고 그 대체 방안을 세우기 위하여 사용되는 방법이다(Kenney 2002). 따라서 효과적인 위험관리를 수행하기 위해서는 분석 대상이 지니고 있는 위험 범위와 카테고리, 구체적인 위험사례 등을 파악해야 한다. 본 연구는 다음과 같은 위험관리 프로그램의 전형적인 진행단계를 따라 연구를 진행시켰다.

1. 위험요인 수집
2. 수집된 위험요인 분류
3. 각 위험요인에 따른 위험평가 수행

4. 위험 분석

5. 위험관리 수행

따라서 대학도서관에서 수행하고 있는 디지털 보전방법에 대한 위험평가를 실시하기 위하여 먼저 각 방법이 지닌 위험요인과 위험가능성 분야를 문헌조사를 통하여 수집하였고 그 결과 다음과 같이 세 가지 위험범주 안에 17개의 위험요인을 선정하였다.

A. 콘텐츠에 직접적 손실을 입히는 위험요인

1. 변환 소프트웨어의 불완전성에 의하여 내용손실이 발생하는 위험
2. bit configuration 변경됨으로써 내용손실이 발생하는 위험
3. 압축이 변경됨으로써 내용손실이 발생하는 위험
4. 암호화 정보가 변경됨으로써 내용손실이 발생하는 위험
5. 연결정보 또는 부수적 정보의 손실로 생겨지는 발생하는 위험

B. 시스템 환경변경으로 발생하는 위험요인

6. 파일헤더 정보가 부분적으로 또는 전혀 변환되지 않음으로써 발생하는 위험
7. 디렉토리 또는 파일 경로변경으로 인한 액세스 제한으로 발생하는 위험
8. 하드웨어 환경 변경으로 발생하는 위험

9. OS 변경이 필요함에 따라 발생하는 위험

10. 인터페이스가 수정됨에 따라 발생하는 위험

C. 행정적 지원 부족으로 발생하는 위험요인

11. 담당 전임직원이 없으므로 발생하는 위험
12. 담당 직원의 기술적 전문지식이 부족함으로 발생하는 위험
13. 장기적 디지털 정보자원 보존 정책이 없음으로 발생하는 위험
14. 디지털 보존 활동에 대한 예산 예측의 어려움에 따라 발생하는 위험
15. 디지털 보존 활동에 드는 충분한 예산이 확보되지 않아 발생하는 위험
16. 이용자가 새로 교육을 받아야 이용할 수 있으므로 발생하는 위험
17. 변환된 자료에 대한 저작권 규제로 인하여 이용제한이 발생하는 위험

다음 단계로 다섯 가지 보존기법 즉 매체재생, 매체변환, 포맷변환, 정보전환, 에뮬레이션기술 기법 각각에 대하여 17가지 위험요인에 관한 위험가능성 수준(risk probability scale)과 위험영향력 수준(risk impact scale)을 측정하였다. 위험가능성은 위험이 일어날 확률을, 위험영향력은 위험이 발생하여 데이터 손실에 미치는

영향력을 측정하는 것으로 다음과 같이 각각 5단계 척도로 가능성 및 영향력을 파악하도록 하였다.

- 위험가능성 척도

1. 매우 낮다 : 5% 미만의 발생할 확률
2. 낮다 : 5 - 15% 사이에 발생할 확률
3. 보통이다 : 16 - 25% 사이에 발생할 확률
4. 높다 : 26 - 40% 사이에 발생할 확률
5. 매우 높다: 41 - 99% 사이에 발생할 확률

- 위험영향력 척도

- A. 약간 : 아주 미소한 부분만 손실 (전체 복구 가능, 원본을 가짐)
- B. 어느 정도 : 부분적 손실 (다른 정보원에 의하여 복구 가능)
- C. 심각한 : 전체 손실 (다른 정보원에 의하여 복구 가능)
- D. 매우 심각한 : 부분적 영구 손실 (복구 가능하지 않음)
- E. 재난적 : 전체 영구 손실 (복구 가능하지 않음)

실제 대학도서관의 디지털 정보자원 관리 담당자가 생각하는 위험정도를 평가하

기 위하여 각 보존기술 방법이 지닌 17가지 위험요인에 대한 위험가능성과 위험영향력 수준을 문의하는 질문지를 작성하였다(참조 <부록 1>). 질문지를 발송할 도서관을 선정하기 위하여 먼저 야후! 코리아 검색엔진을 이용하여 대학도서관으로 카테고리 검색을 하였다. 그 결과 122개 대학도서관 리스트를 작성할 수 있었고 이를 바탕으로 체계적 표본추출 방법을 통하여 40개의 대학을 선택하였다. 다음 40개 대학도서관의 디지털정보 관리자에게 전화로 설문내용을 설명한 후, 응답을 허락한 30개 대학의 담당자에게 전자메일로 질문지를 발송하였다. 이중 20곳³⁾에서만 응답하여 회수율 70%을 나타냈다.

4 위험평가 분석

4.1 디지털 정보자원 보존활동에 관한 분석

디지털 보존기법에 대한 위험평가를 실시하기 전 디지털 정보관리 담당자에게 디지털 보존활동 및 그와 관련 사항에 대하여 질문하였다. <표 1>에서 나타났듯이 대부분의 담당자는 전자정보실 속해 있으며 현 부서 근무연수는 대다수가 보통 1년에서 5년 사이인 것으로 나타났다. 그러나 5년 이상 근무한 사서도 5명이나 있

3) 경기대, 고려대, 광운대, 대구대, 동국대, 서울신학대, 성균관대, 숙명여대, 순천향대, 아주대, 영남대, 울산대, 이화여대, 중앙대, 충남대, 포항공대, 한국기술교육대, 한양대, 한성대, 홍익대

〈표 1〉 대학도서관의 디지털 정보자원 보존활동

번호	질문	항목	빈도수
1	소속부서	전자정보실	13
		학술정보팀	4
		지원팀	3
2	현 근무 연수	1년 미만	2
		1-3년 미만	7
		3-5년 미만	6
		5-7년 미만	2
		7년 이상	3
3	디지털 보존 활동여부	정기적으로 실시	8
		비정기적으로 실시	8
		실시한 적 없다	4
4	디지털 보존시 사용한 기술	매체재생	7
		매체변환	6
		포맷변환	7
		정보전환	7
		에م레이션	0
		무응답	4
5	보존시 가장 많이 사용한 기술	매체재생	5
		매체변환	3
		포맷변환	6
		정보전환	2
		에م레이션	0
		무응답	4
6	변환 후 이전파일 보존 여부	모두 보존	10
		이존 파일은 부분적 보존	5
		새 파일만 보존	1
		무응답	4
7	디지털 보존 관련 정책 여부	있 다	3
		없 다	17
8	위험관리 계획수립 여부	있 다	5
		없 다	15
9	디지털 보존만을 위한 예산확보 여부	있 다	5
		없 다	15
10	보존활동 수행 기관	본 도서관 직원이 수행	14
		외부전문 업체 수행	1
		기 타	1
		무응답	4

는 것으로 보면 디지털 정보자원을 전문성을 지닌 사서가 계속해서 관리하도록 하는 경향을 알 수 있었다.

디지털 보존활동을 살펴보면, 응답한 20개의 도서관 중 8개의 도서관은 정기적으로, 또 다른 8개의 도서관은 비정기적으로 디지털 정보자원의 장기적 보존업무를 수행하고 있는 것으로 나타났으나 나머지 4개 도서관은 보존활동으로서의 업무를 실시한 적이 없는 것으로 응답하였다. 또한 보존활동을 수행하는 16개 대학도서관에서 가장 빈번히 사용한 보존기술은 이전의 것을 다른 포맷으로 변환시켜 사용하도록 하는 “포맷변환”임을 알 수 있었고 그 다음 “매체재생”, “매체변환”, “정보전환” 순이었지만, 한 도서관에서 하나의 기법만을 사용하는 것은 아니므로 “에뮬레이션” 기법을 제외하고는 모든 기법들이 사용되고 있는 것을 알 수 있다(<표 1>의 항목 4 참조). 최신의 기법인 “에뮬레이션” 기술은 아직까지 우리나라에서는 에뮬레이터가 본격적으로 상품화되지 않아 도서관에서는 사용하지 않고 있는 것으로 보인다.

또한 대학도서관이 축적하고 있는 디지털 정보자원의 양이 비교적 적어서인지 변환파일은 물론이고 이전파일도 보존하고 있는 도서관이 10개나 되며, 선별적으로 보존하는 도서관은 5개이고 새 변환파일만 보존하는 도서관은 하나인 것으로 나타났으며 이러한 보존활동을 대다수 도서관(70%)에서는 해당 도서관 사서들이

직접 수행하는 것으로 보인다.

디지털 보존에 대한 위험관리 계획을 수립하고 이에 따른 예산 확보를 하고 있는 도서관은 5개 도서관으로 겨우 25%에 해당하며 디지털 보존 관련 정책을 세운 도서관은 겨우 세 개로 장기적 정책안에 체계적이고 일상적이며 하나의 독립적인 (독자적 예산을 지닌) 도서관 활동으로 보존활동이 이루어지지 않고 있음을 알 수 있다.

4.2 보존기법에 따른 위험평가 분석

4.2.1 위험가능성 분석

위험을 측정하는 것은 위험요인을 파악하는 것만큼 복잡하고 어려운 일이다. 특히 위험가능성과 같이 실제 일어난 빈도가 아닌 확률을 문의하는 질문에 대한 답은 매우 정황적이고 주관적인 면을 가지게 되며 응답자 또한 어려움을 느끼게 된다. 이러한 어려움을 줄이는 시도로 척도 수준을 세 단계(높다, 보통, 낮다)로 사용하는 것이 일반적이기는 하나, 본 연구에서는 Lawrence et al.(2000) 연구에서 사용한 것처럼 위험발생 확률 분포를 5단계 비대칭 척도(“매우 낮다” 의미는 위험 발생확률 5% 미만을 가리키며, “매우 높다”의 의미는 위험발생 확률이 41% 이상임을 뜻한다)로 만들어 위험가능성 정도를 측정하였다. 20개 대학도서관 사서들이 각 위험요인별 위험가능성 수준을 어떻게 평가하였는지를 쉽게 파악하기 위

하여 각 척도에 따라 1, 2, 3, 4, 5값(위험 가능성이 높아질수록 큰 값을 가지도록)을 배점하였다.

〈표 2〉는 5개의 보존기술에 대한 각 위험요인 별 위험가능성 척도의 평균값을 보여주는 것으로 평균값을 살펴보면, 위험이 발생할 확률에 가장 낮은 방법은 “매체재생”(2.17)인 반면, 가장 높은 것은 “에물레이션”(2.65)으로 나타났다. 그 사이 “매체변환”(2.35), “포맷변환”(2.41), “정보전환”(2.58) 순으로 위험 발생 확률이 높은 것으로 나타났으나 그 차이는 미비하다고 볼 수 있다.

“매체재생” 기법을 사용할 때 위험가능성이 가장 높은 요인은 매체변환 담당자의 전문지식의 부족과 전임담당자가 부재요인을 들었는데 이는 사서들이 “매체재생” 시 담당자의 실수 또는 무지에서 생겨진 위험이 가장 크게 보고 있음을 알 수 있다. 그 반면 이용자 재교육 시에 생겨질 위험을 가장 낮게 보았는데 이는 같은 매체로의 변환이므로 이용자 재교육은 실제로 필요 없으므로 이로 인한 위험 또한 낮다고 보는 것이 당연하다. 또 재미있는 현상으로 다른 어떤 기법보다도 “매체재생” 기법을 통하여 디지털 정보자원을 재생산할 경우 저작권 문제가 가장 심각할 것이라는 생각을 하고 있는 것으로 나타났다. 또한 사서들은 “매체재생” 기법은 콘텐츠에 직접적 손실을 입히는 위험(1.94)이 행정적 지원 부족에서 오는 위험(2.57)보다는 훨씬 적을 것이라 예상

하고 있다(〈표 4〉 참조).

“매체변환” 기법인 경우, 위험가능성이 가장 높은 요인을 예산확보의 어려움(2.94)과 직원의 전문지식 부족(2.77)으로 보았고 비트 변경이나 이용자 재교육시에 발생하는 확률을 가장 낮게 보는 것으로 나타났다. 여전히 행정적 지원 부족에서 오는 위험확률이 콘텐츠 변경이나 시스템 환경변경에서 오는 확률보다 높게 나타났으나 “매체재생”과는 다르게 시스템 환경변경에서 오는 위험확률(2.55)이 콘텐츠 변경에서 기인하는 것(2.27)보다 높게 표시하였다. 현재 사용 중인 포맷이 아닌 다른 포맷으로 변환시켜주는 “포맷변환” 기법인 경우 모든 요인이 2.00에서 3.00 사이의 값을 가지므로 5-15%의 위험발생 확률을 지닌다고 할 수 있다. 역시 가장 위험가능성이 높은 요인은 담당직원의 부재를 들었고 이와 동시에 예산확보의 어려움도 지적하였다. 한편 가장 낮은 요인은 암호화변경에 따른 내용손실 요인으로 나타났다. “매체변환” 기법과 마찬가지로 행정적 지원부족으로 오는 위험가능성(2.66)이 가장 높았고 다음 콘텐츠 변경(2.34), 시스템 환경변경(2.25)으로 나타났다.

“정보전환”인 경우도 역시 담당직원 부재와 담당직원 전문성 부족에서 오는 위험가능성이 가장 높은 것으로 보았고 또한 “정보전환” 시 하드웨어 변경과 암호변경으로 인한 내용손실 및 부분적 파괴 위험가능성도 역시 높을 것으로 보고 있

음을 알 수 있다. 여전히 이용자 재교육으로 인한 위험과 인터페이스 수정으로 인한 위험을 가장 낮게 보았다. 다른 기법과 마찬가지로 행정적 지원 부족에서 오는 위험확률이 가장 높게 나타났으나 다른 범주와의 차이가 매우 근소하게 나타났다. 전반적으로 다 위험발생 확률이 이전 세 가지 기법보다는 높게 나타났다(<표 4> 참조).

실제 도서관에서 사용하지 않고 있는 “에물레이션” 기법인 경우에 가장 위험가능성이 높은 요인은 보존정책이 없음으로 나타났다. 이는 현재 “에물레이션”이 비

용과 기술적 문제로 인하여 모든 데이터에 적용시키는 것이 아니라 선별적으로 이용가치가 높은 데이터에만 실시하고 있는 환경에서 “에물레이션” 활용에는 장기적 계획 및 정책이 필요하다는 사서들의 생각을 추정할 수 있다. 그 외 여전히 담당직원 부재 및 전문지식 부족에서 오는 위험가능성을 높이 보았고 다른 기법보다 변환소프트웨어 불완전성에서 오는 내용 손실 위험가능성도 높을 것으로 보았다. 역시 “에물레이션” 기법을 이용하여 데이터가 변환되기는 하나 포맷이나 느낌이 변화되지 않았으므로 이용자 재교육시에

<표 2> 5개 보존기술에 대한 위험요인별 위험가능성 비교

범 주	번호	위험 내 용	매체 재생	매체 변환	포맷 변환	정보 전환	에물레 이션	평 균
컨텐츠 변경	1	변환 소프트웨어의 불완전성	2.00	2.47	2.29	2.71	2.88	2.47
	2	bit configuration 변경	1.88	1.88	2.35	2.26	2.75	2.22
	3	압축방식의 변경	1.94	2.59	2.35	2.41	2.63	2.38
	4	암호화 정보의 변경	2.06	2.06	2.06	2.82	2.63	2.33
	5	연결정보/부수적 정보 손실	1.84	2.35	2.65	2.53	2.31	2.34
시스템 환경 변경	6	파일헤더 정보손실	1.94	2.18	2.24	2.65	2.63	2.33
	7	디렉토리/파일경로 변경	1.82	2.06	2.41	2.77	2.38	2.29
	8	하드웨어 환경 변경	2.23	2.29	2.29	2.82	2.63	2.45
	9	OS 변경	2.17	2.47	2.12	2.35	2.81	2.38
	10	인터페이스 수정	1.77	2.12	2.18	2.18	2.50	2.15
행정적 지원 부족	11	담당 전임직원 부재	2.71	2.59	2.94	3.00	2.94	2.84
	12	담당 직원 전문지식 부족	2.94	2.77	2.88	2.82	3.00	2.88
	13	장기적 보존 정책 부재	2.59	2.47	2.53	2.77	3.12	2.70
	14	보존활동 예산 예측의 어려움	2.47	2.53	2.47	2.53	2.56	2.51
	15	충분한 예산 확보의 어려움	2.65	2.94	2.94	2.59	2.81	2.79
	16	이용자 재교육 필요	1.77	1.88	2.12	2.12	2.06	1.99
	17	변환된 자료에 대한 저작권 규제	2.88	2.65	2.71	2.65	2.56	2.69
평 균			2.17	2.35	2.41	2.58	2.65	2.43

발생되는 위험가능성을 가장 낮게 보았다. 특히 이 값(2.06)은 “포맷변환”(2.12), “정보전환”(2.12) 기법에서 나타난 값보다 낮았다. 또한 행정적 지원 부족에서 오는 위험가능성(2.72)이 가장 높았으나 내용에 손실을 입힐 위험(2.64) 역시 높을 것이라 예상한 것으로 보인다.

4.2.2 위험영향력 분석

위험영향력은 위험이 발생되어 데이터 손실에 미치는 영향력을 측정하는 것으로 데이터 손실정도와 복구가능성에 따라 위험영향력 척도 값이 틀려지도록 하였다.

따라서 “재난적 손실”은 디지털 정보자원이 전체적으로 손실된 상태에서 어떠한 정보원도 남겨있지 않아 복구도 가능하지 않은 경우를 말하고 “심각한”의 의미는 변환시 전체 손실이 되었으나 다른 정보원에 의하여 복구가 가능한 것을 말한다. 또한 “약간”의 의미는 미소한 부분만 손실되었으며 복구 또한 쉽게 가능한 것을 말한다.

〈표 3〉은 5개의 보존기술에 대한 각 위험요인 별 위험영향력 정도를 수치로 환산하여 평균값을 산출한 것이다. 다섯 가지 방법을 비교한 결과 정보전환의 위험

〈표 3〉 5개 보존기술에 대한 위험요인별 위험영향력 비교

범 주	번호	위험내용	매체 재생	매체 변환	포맷 변환	정보 전환	에물레 이션	평균
컨텐츠 변경	1	변환 소프트웨어의 불완전성	1.57	2.06	2.00	2.47	2.19	2.06
	2	bit configuration 변경	1.77	1.84	2.17	2.29	2.25	2.06
	3	압축방식의 변경	1.71	2.12	1.71	2.53	2.31	2.08
	4	암호화 정보의 변경	1.94	1.82	1.82	2.24	2.38	2.04
	5	연결정보/부수적 정보 손실	1.71	1.88	2.00	2.35	2.25	2.04
시스템 환경 변경	6	파일헤더 정보손실	2.24	2.24	1.82	2.53	2.25	2.22
	7	디렉토리/파일경로 변경	1.82	2.18	2.00	2.24	2.13	2.07
	8	하드웨어 환경 변경	1.82	2.12	1.94	2.53	2.50	2.18
	9	OS 변경	2.12	2.00	2.06	2.29	2.56	2.21
	10	인터페이스 수정	1.59	1.82	1.82	2.17	2.13	1.91
행정적 지원 부족	11	담당 전임직원 부재	2.29	2.12	2.24	2.65	2.50	2.36
	12	담당 직원 전문지식 부족	2.35	2.24	2.12	2.71	3.06	2.50
	13	장기적 보존 정책 부재	2.24	2.06	1.94	2.18	2.06	2.10
	14	보존활동 예산 예측의 어려움	2.06	2.06	1.94	2.06	1.69	1.96
	15	충분한 예산 확보의 어려움	2.18	2.41	2.18	2.41	2.19	2.27
	16	이용자 재교육 필요	1.65	1.47	1.77	1.71	1.81	1.68
	17	변환된 자료에 대한 저작권 규제	1.88	2.18	2.12	2.24	2.06	2.12
평 균			1.92	2.03	1.97	2.33	2.26	2.10

영향력(2.33)이 가장 높은 것으로 나타난 반면 “매체재생”의 위협영향력(1.92)이 가장 낮은 것으로 나타났다. 이로써 “매체재생” 기법이 가장 단순하며 위협 발생 확률과 손실정도가 가장 적을 것으로 보인다. 또한 위험가능성이 가장 높은 것으로 “에물레이션” 기법을 지정하였으나, 데이터 손실 영향력에서는 “정보전환” 기법이 가장 높은 것으로 예측하고 있었다.

위험영향력 값이 가장 높은 요인은 역시 위험가능성과 마찬가지로 담당직원의 전문적 지식 부족과 담당직원 부재로 나타났다. 가장 낮은 요인은 이용자 재교육 요인인 것으로 판명되었다. 이는 어떤 보존기술 방법을 사용하더라도 이용자는 편리하게 그리고 용이하게 디지털 정보를 사용할 수 있을 것이라 생각하고 있음을 보여준다. 또한, “매체재생” 기법을 사용할 경우 변환소프트웨어 불완전성에서 기인하는 위험영향력이 가장 낮게 나타난 것으로 보아 “매체재생”을 위하여 사용되는 소프트웨어는 매우 안정적임을 알 수

있었다. 또한 위험영향력 값이 가장 높은 것은, 즉 데이터 손실영향력이 높을 것으로 예상되는 요인으로 “에물레이션” 기법 사용시 담당직원의 전문지식 부족 요인을 꼽았는데 이는 사서들이 “에물레이션” 기법을 사용할 때에는 전문지식이 절대적으로 필요하다고 생각하고 있음을 알 수 있었다.

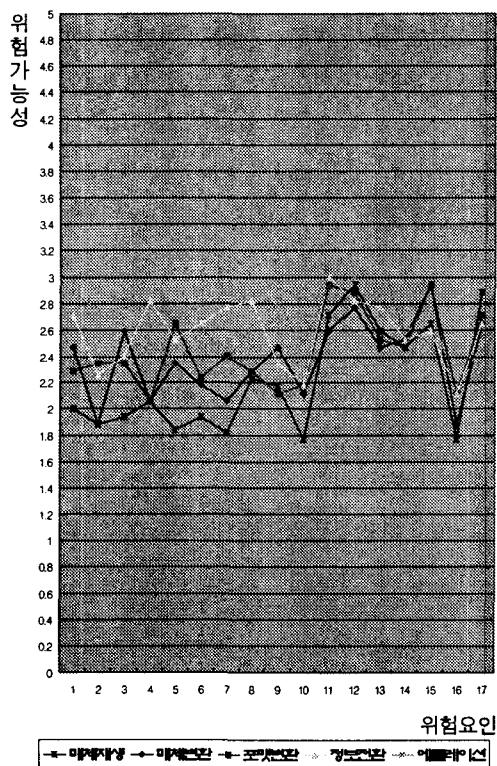
〈표 4〉를 살펴보면 “매체재생”, “매체변환”, “포맷변환” 기법인 경우 행정적 지원 부족 요인들이 데이터 손실을 일으키는 가장 높은 위험영향력을 가진 반면 “정보전환”과 “에물레이션” 기법인 경우에는 행정적 지원 부족이 가장 낮게 나타났다. 특히 “에물레이션” 기법인 경우 시스템 환경변경 요인이 가장 위험영향력이 높은 것으로 낮는데 이는 “정보전환”과 “에물레이션”이 변환시 하드웨어/소프트웨어가 전환되기 때문에 이로 인한 데이터손실과 복구 어려움이 있을 것이라는 예상을 하는 것 같다. 전체적으로 위험영향력 값이 위험가능성 값보다 낮게 나타

〈표 4〉 세가지 위험범주별 위험가능성과 위험영향력 비교

범 주	위 험 내 용	매체 재생	매체 변환	포맷 변환	정보 전환	에물 레이션	평 균
컨텐츠 변경	위험가능성	1.94	2.27	2.34	2.55	2.64	2.35
	위험영향력	1.74	1.94	1.94	2.37	2.28	2.06
시스템 환경 변경	위험가능성	1.99	2.22	2.25	2.55	2.59	2.32
	위험영향력	1.92	2.07	1.93	2.35	2.31	2.12
행정적 지원 부족	위험가능성	2.57	2.55	2.66	2.64	2.72	2.63
	위험영향력	2.09	2.8	2.04	2.28	2.20	2.14
평 균	위험가능성	2.17	2.35	2.41	2.58	2.65	2.43
	위험영향력	1.92	2.03	1.97	2.34	2.26	2.10

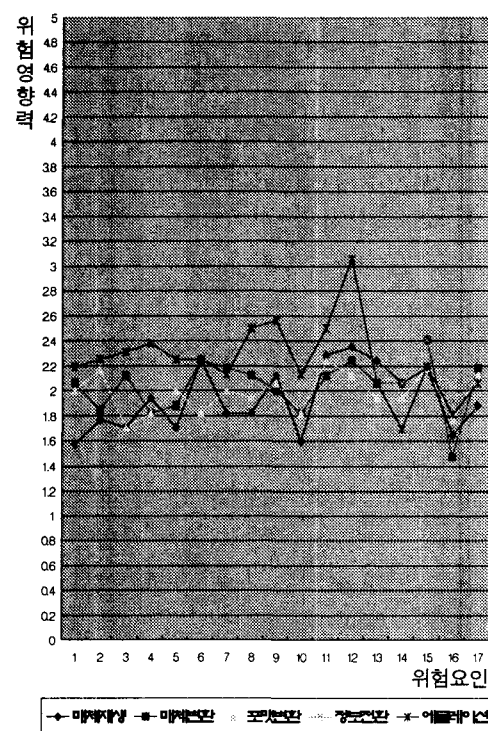
났다.

〈그림 1〉과 〈그림 2〉는 위험요인 별 위험가능성 값과 위험영향력 값을 비교한 그래프로 “에물레이션” 기법이 대체로 높은 위험발생 가능성을 보여준 반면 높은 위험영향력을 보여준 기법은 “정보전환”임을 알 수 있다. 따라서 “매체재생”, “매체변환”, “포맷변환”을 수행할 경우보다는



〈그림 1〉 위험요인에 대한 보존기술별 위험가능성

“정보전환”과 “에물레이션” 기법을 사용할 때 위험발생 탐지에 노력해야 할 것이며 전문적인 지식을 가진 전문가가 보다 체계적으로 진행하도록 하여 위험발생 확률을 낮추어야 할 것이다.



〈그림 2〉 위험요인에 대한 보존기술별 위험영향력

4.2.3 위험판단 분석

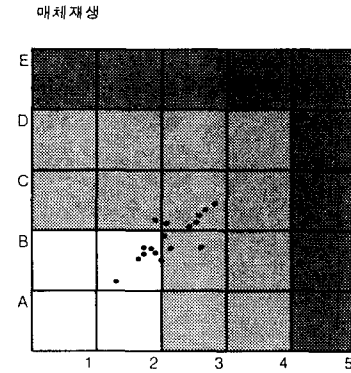
위험가능성과 위험영향력의 조합을 하나의 값⁴⁾으로 산출하여 각 요인에 대한

4) 조합된 값의 예는 다음과 같다.

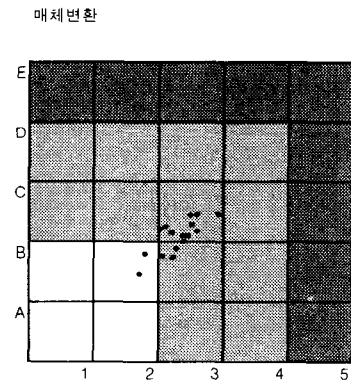
- 5E: 재난적 영향력을 가지며 위험발생 확률이 매우 높음
- 3D: 매우 심각한 영향력을 가지며 위험발생 확률이 높음
- 2C: 심각한 영향력을 가지며 위험발생 확률이 보통임
- 1B: 어느 정도의 영향력을 가지며 위험발생 확률이 매우 낮음
- 1A: 약간의 영향력을 가지며 위험발생 확률이 매우 낮음

위험판단 분석을 하였다. 이러한 위험판단 결정을 보다 용이하게 하기 위하여 세 가지 위험판단 범주를 파악할 수 있게 하는 매트릭스를 작성하였다. 일반적으로 흰색의 격자 셀(1A-B, 2A-B)에 놓이게 되면 낮은 위험 확률을 가진 것으로 조심스럽게 보존기술 기법을 사용하면 위험발생 확률을 낮출 수 있게 된다. 또한 연한 회색의 격자 셀(1C-D, 2C-D, 3A-D, 4A-D)에 놓이게 되면 꽤 높은 위험을 가지게 되므로 그 요인이 지닌 위험가능성이 낮추어 질 때까지 연기하는 것이 좋다. 만약 진한 회색 격자 셀(1E, 2E, 3E, 4E, 5A-E)에 놓인 경우에는 변환작업을 실시하지 말아야 하고 궁극적으로 그 요인을 제거시킨 후에 실행하도록 결정하는 것이 좋다.

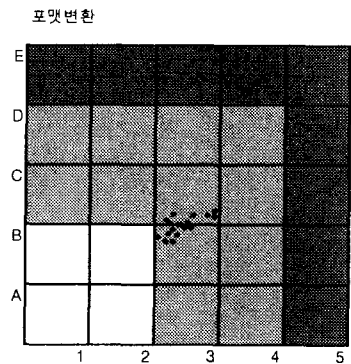
〈그림 3-7〉을 살펴본 결과 대부분의 요인들이 연한 회색 격자 셀에 놓인 것을 알 수 있으므로 디지털 보존활동 전에 각각의 요인들이 지닌 위험요소들을 미리 파악하고 제거하도록 노력하여 위험이 발생되지 않도록 조치를 취해야 할 것이며 실행시에는 전문가가 체계적으로 수행하여야 할 것이다. 특히 모든 요인들이 연한 회색 격자 셀에 놓인 “포맷변환”, “정보전환”과 “에뮬레이션” 기법을 사용할 경우 담당자는 모든 면에서 유의해야 할 것이다. 또한 도서관은 보다 적극적으로 행정적 지원(예, 전문 담당 직원 채용, 담당자 전문교육 훈련, 예산확보, 보존에 관련된 정책수립 등)에 힘써야 할 것이다.



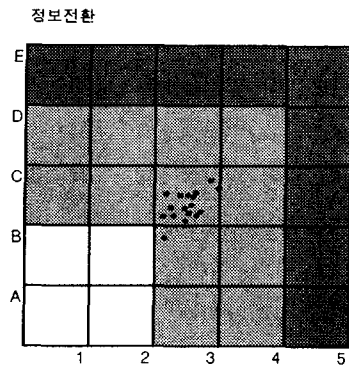
〈그림 3〉 매체재생의 위험판단



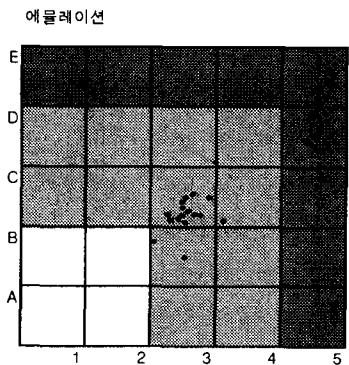
〈그림 4〉 매체변환의 위험판단



〈그림 5〉 포맷변환의 위험판단



〈그림 6〉 정보전환의 위험판단



〈그림 7〉 에물레이션의 위험판단

5 결 론

장기적 디지털 정보자원 보존을 위한 업무는 계속되는 기술적 변화로 인하여 생겨지는 미래의 불확실성에 대한 도전과 같은 것이라 할 수 있다. 따라서 기술적으로 디지털 보존방식 중 어떤 방식이 최고이고, 어떤 방식이 표준이고, 어떤 방식만으로 사용해야 한다는 절대적 결론은 있을 수 없다. 이런 환경 때문에, 즉 궁극

적 솔루션이 없다는 이유만으로 현재 사용 중인 방식을 평가하고 이 안에서의 최선의 방법을 찾는 것을 기피할 필요는 없다. 미래 이용자들을 위한 정보자원의 보존활동이 불가피한 상황에서는 더욱 그렇다.

본 연구는 디지털 정보자원 보존기술로서 도서관에서 빈번히 사용되고 있는 다섯 가지 보존기술(매체재생, 매체변환, 포맷변환, 정보전환, 에물레이션)에 대하여 대학도서관 디지털 정보관리 담당자들이 위험평가를 실시하였다. 위험평가는 위험이 발생할 확률을 물어보는 위험가능성과 위험이 데이터손실에 까지 미치는 위험영향력을 측정함으로써 실시되었다. 조사 결과를 요약하면 다음과 같다.

1. 80%의 대학도서관은 현재 디지털 보존활동을 실시하고 있으며 가장 빈번히 사용하고 있는 기법은 포맷변환이고 전혀 사용하지 않고 있는 기법은 에물레이션임을 알 수 있었다. 또한 25%의 도서관만이 디지털 보존활동을 위한 예산을 확보하고 위험관리 계획을 수립하고 있는 것으로 나타나 아직까지 대부분의 대학도서관에서는 디지털 보존활동이 도서관 일상적 업무로 인식되지 못하고 있음을 알 수 있었다.
2. 위험요인별 5개 보존 기술의 위험가능성을 비교한 결과 위험이 발생할 확률이 가장 낮은 방법은 매체재생인 것으로 판명되었고 가장 높은 것은 에물레이션으로 나타났다. 특히 사서들은 디지털 보존활동을 전문적 수행할 전담

담당자가 없는 경우에 실시되는 애플레이션이나 정보전환 기법이 가장 큰 위험확률을 가져올 것이라 본 반면, 이용자 재교육에 의한 위험 발생확률이 가장 낮을 것으로 예상하였다. 따라서 대부분의 도서관이나 디지털 보존기법을 개발하는 업체들은 기술적 솔루션 개발만을 염두에 두고 이것이 최선책으로 보고 있으나, 이와 동시에 디지털 보존에 대한 전문지식을 갖는 사서나 전문가를 양성해야 한다.

3. 위험요인별 보존기술의 위험영향력을 비교한 결과 정보전환 기법의 위험영향력이 가장 높게 나타난 반면, 매체재생에서 가장 낮게 나타났다. 이로서 매체재생 기법은 위험발생확률, 데이터 손실 영향력 모두 가장 적은 것으로 보인다. 여전히 디지털 정보관리자의 전문지식 부족 및 전문 담당 직원 부재가 높은 위험요인으로 나타났다. 또한 매체재생, 매체변환, 포맷변환 이 세 기법인 경우 행정적 지원 부족요인들이 데이터 손실을 일으키는 가장 높은 영향력을 가진 것으로 나타난 반면, 정보전환과 애플레이션 기법에서는 가장 낮게 나타났다. 이는 두 기법을 수행할 때 사용되는 소프트웨어가 일킬 위험을 가장 크다고 예상하고 있는 것으로 보인다.
4. 위험판단 결정을 보다 용이하게 하기 위하여 위험가능성과 위험영향력의 조합 값을 세 가지 범주로 분류시켰다.

그 결과 대부분의 요인들이 “위험범주”에 속해 현재 도서관은 보존활동시 위험요인을 파악하고 또 제거하도록 노력하여 위험발생 확률을 낮추어야 하여야 할 것이다. 특히 행정적 지원 부재로 인한 위험발생을 낮추도록 노력해야 한다.

이와 같이 도서관에서 사용되고 있는 디지털 보존을 위한 전략적 기법들을 분석하였으나 이 방법만이 보존활동에 필요한 것은 아니다. 다양한 데이터의 보존을 용이하게 하기 위해서는 이질적 객체를 정규화 시키는 기술, 진본성을 입증해주는 알고리즘, 상호운영을 위한 메커니즘, 대규모 백업 메커니즘, 복구 메커니즘 등 여러 기술이 필요하다. 이러한 기술적 문제 이외에도 저작권 문제, 이용제한 등 실제 보존과 관련된 많은 이슈들이 아직 해결되지 못한 채 있어 디지털 보존을 위협하게 하고 있다. 따라서 이러한 문제도 앞으로 해결해야 할 것이다. 또한 본 연구가 디지털 보존을 위한 도서관에서 사용되고 있는 전략적 기법 자체를 분석한 것으로써 각각의 기법들이 지닌 특성과 예상된 위험만이 파악되었다는 한계점을 가진다. 따라서 다음 연구에서는 어떤 경우에 실제 위험이 발생하는 지에 대한 즉 시스템 안에서 실제 사용할 때 발생하는 위험에 대하여 상세 연구를 수행해야 할 것이다. 이를 위해서는 보다 정제된 위험평가 도구가 개발되어야 할 것이다.

모든 자료는 보존되어야 한다는 전체

아래, 한 때는 디지털 정보자원도 모두 보존해야 한다는 생각을 지녔다. 그러나 이것은 이론적으로는 가능하나, 디지털 환경에서는 엄청난 비용이 필요하므로 실제 모든 데이터가 보존되고 있지는 않다. 그렇다면 디지털 정보자원의 보존은 기술적 문제뿐만 아니라 “선별”이라는 또 하나의 어려운 업무가 뒤따르게 된다. 이 또한 미래 우리에게 다가올 또 하나의 문제로 남는다 하겠다. 그러나 분명 가치 있는 정보자원은 장기적으로 또는 영구적으로 유지 보존되어야 함으로 앞으로 디지털 정보자원의 보존에 더 많은 연구가 수행되어야 할 것이다.

참 고 문 헌

- 신은자. 2001. 전자저널의 아카이빙에 관한 연구. 『정보관리학회지』, 18(3): 139-158.
- 윤희윤. 2001. 국가도서관의 전자출판물 수집 및 보존방안. 『도서관』, 56(3): 3-48.
- 최원태. 2001. 디지털 아카이브의 현황 및 구성 요소에 관한 연구. 『한국문헌정보학회지』, 35(2): 23-40.
- Arms, C.R. 2000. "Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress," *RLG DigiNews*, 4(3).
<<http://www.rlg.org/preserv/diginews/diginews4-3.html>>.
- Bennett, John. 1997. *A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material*. British Library Research and Innovation Report, 50. West Yorkshire, UK: BLRIC.
<<http://www.ukoln.ac.uk/services/elib/papers/supporting/#blric>>.
- Berthon, H., S. Thomas, and C. Webb. 2002. "Safekeeping: A Cooperative Approach to Building a Digital Preservation Resources," *D-Lib Magazine*, 8(1).
<<http://www.dlib.org/dlib/january02/berthon/01berthon.html>>.
- Bikson, T.K. 1998. "Managing Digital Documents: Technology Challenges and Institutional Responses." In *Proceedings of the International Conference of the Round Table on Archives*. Stockholm, Sweden: International Council on Archives.
- Brown, Adrian. 2000. *Digital Archiving Strategy*. English Heritage Center for Archeology.
<http://www.ifs.tuwien.ac.at/~aola/links/safekept/e-heritage_strategy.pdf>
- Consultative Committee for Space Data

- System. 1999. *Recommendation Concerning Space Data System Standards: Reference Model for an Open Archival Information System(OAIS)*, CCSDS 650.0-R-1, NASA.
[<http://wwwdev.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf>.](http://wwwdev.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf)
- Friedlander, A. 2002. "The Nation Digital Information Infrastructure Preservation Program: Expectations, Realities, Choices and Progress to Date." *D-Lib Magazine*, 8(4).
[<http://www.dlib.org/dlib/april02/friedlander/04friedlander.html>.](http://www.dlib.org/dlib/april02/friedlander/04friedlander.html)
- Garret, J. and D. Waters. 1996. *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information*. Commission on Preservation and Access and the Research Libraries Group Inc.
[<http://www.rlg.org/ArchTF>.](http://www.rlg.org/ArchTF)
- Granger, S. 2000. "Emulation as a Digital Preservation Strategy." *D-Lib Magazine*, 6(10). [<http://www.dlib.org/dlib/october00/granger/10granger.html>.](http://www.dlib.org/dlib/october00/granger/10granger.html)
- Green, A. et al. 1999. *Preserving the Whole: A Two-Track Approach the Rescuing Social Science Data and Metadata*. Washington, D.C.: Digital Library Federation.
[<http://www.clir.org/pubs/reports/pub83/contents.html>.](http://www.clir.org/pubs/reports/pub83/contents.html)
- Hodge, G.M. 2000. "Best Practices for Digital Archiving: An Information Life Cycle Approach." *D-Lib Magazine*, 6(1). [<http://www.dlib.org/dlib/january00/01hodge.html>.](http://www.dlib.org/dlib/january00/01hodge.html)
- Kenney, A.R. et al. 2002. "Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism." *D-Lib Magazine*, 8(1). [<http://www.dlib.org/dlib/january02/kenney/01kenney.html>.](http://www.dlib.org/dlib/january02/kenney/01kenney.html)
- Lawrence, G.W., et al. 2000. *Risk Management of Digital Information: a File Format Investigation*. Council on Library and Information Resources.
[<http://www.diglib.org/preserve.htm>.](http://www.diglib.org/preserve.htm)
- National Library of Australia. 2001. *PANDORA Archive*.
[<http://pandora.nla.gov.au>.](http://pandora.nla.gov.au)
- National Library of Canada. 1996. *Electronic Publications Pilot Project*. Summary of the Final Report.
[<http://www.collection.nlc-bnc.ca/e-coll-e/ereport.htm>.](http://www.collection.nlc-bnc.ca/e-coll-e/ereport.htm)
- Rothenberg, J. 1995. "Ensuring the Longevity of Digital Documents." *Scientific American*, 272(1): 46-47.

- Rothenberg, J. 1999. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Council on Library and Information Resources.
<http://www.clir.org/pubs/report_s/rothenberg/contents.html>.
- Russell, Kelly. 2000. *CEDARS: Long-term Access and Usability of Digital Resources: the Digital Preservation Conundrum*.
<<http://www.ariadne.ac.uk/issue18/cedars>>.
- Stone, A. and M. Day. 1999. *CEDARS Preservation Metadata Elements: CEDARS Project Document*. AIW02, CEDARS.
- Thibodeau, K. 2001. "Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration," *D-Lib Magazine*, 7(2).
<<http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html>>.

〈부록 1〉 디지털 보존 기술에 대한 위험도 조사:
매체재생, 매체변환, 포맷변환, 정보전환, 에뮬레이션

범 주	번호	위험 내용	위험 가능성 수준	위험 영향력 수준
보존성: 내용면	1	변환 소프트웨어의 불완전성에 의하여 내용손실이 발생하는 위험정도		
	2	bit configuration 변경됨으로써 내용손실이 발생하는 위험정도		
	3	압축이 변경됨으로써 내용손실이 발생하는 위험정도		
	4	암호화 정보가 변경됨으로써 내용손실이 발생하는 위험정도		
	5	연결정보 또는 부수적 정보의 손실로 생겨지는 발생하는 위험정도		
보존성: 시스템면	6	파일헤더 정보가 부분적으로 또는 전혀 변환되지 않음으로써 발생하는 위험정도		
	7	디렉토리 또는 파일 경로변경으로 인한 액세스 제한이 발생하는 위험정도		
	8	하드웨어 환경 변경으로 발생하는 위험정도		
	9	OS 변경이 필요함에 따라 발생하는 위험정도		
	10	인터페이스가 수정됨에 따라 발생하는 위험정도		
지원성	11	담당 전임직원이 없으므로 발생하는 위험정도		
	12	담당 직원의 기술적 전문지식이 부족함으로 발생하는 위험정도		
	13	장기적 디지털 정보자원 보존 정책이 없음으로 발생하는 위험정도		
	14	디지털 보존 활동에 대한 예산 예측이 어려움에 따라 발생하는 위험정도		
	15	디지털 보존 활동에 드는 충분한 예산이 확보되지 않아 발생하는 위험정도		
	16	이용자가 새로 교육을 받아야 이용할 수 있으므로 발생하는 위험정도		
	17	변환된 자료에 대한 저작권 규제에 의하여 이용제한이 발생하는 위험정도		

위험가능성 수준: ① 매우 낮다. ② 낮다. ③ 보통이다. ④ 높다. ⑤ 매우 높다.
위험영향력 수준: ⑥ 약간 ⑦ 어느 정도 ⑧ 심각한 ⑨ 매우 심각한 ⑩ 재난적