

# 웹 아카이빙의 법·제도적 문제에 대한 고찰

- 웹 정보자원의 특성을 중심으로 -

## A Study of Legal Issues for Web Archiving

김 유 승(You-Seung Kim)\*

### 목 차

- |                              |                               |
|------------------------------|-------------------------------|
| 1. 문제제기                      | 3.1 저작권                       |
| 2. 웹 정보자원의 특성 분석: 아카이빙의 측면에서 | 3.2 납본과 수집                    |
| 2.1 분산성                      | 3.3 진본성                       |
| 2.2 휘발성                      | 4. 『공공기록물 관리에 관한 법률』과 웹 아카이빙  |
| 2.3 역동성                      | 5. 웹 아카이빙 기관의 과제: OASIS를 중심으로 |
| 2.4 심층 웹                     | 6. 나가는 글                      |
| 2.5 진화하는 웹                   |                               |
| 3. 웹 아카이빙의 법·제도적 문제          |                               |

### 초 록

정보자원의 기록과 보존이라는 측면에서 웹이 차지하는 가치는 날로 증대되고 있다. 웹의 정보적, 문화적 그리고 증거적 가치를 보존하고자 하는 노력들은 웹의 본격적인 대중화와 함께 전개되기 시작했다. 웹을 기반으로 하는 정보자원의 보존은 웹의 고유한 매체 특성으로 인해 기존의 인쇄물을 중심으로 하는 유형 기록물의 보존에서 나타나는 것과는 현저하게 다른 문제점들을 노정하고 있다. 웹의 기술적 아키텍처의 특성에 기반 한 기술적 측면의 문제들이 그 중 하나이고, 웹 정보자원의 저작권, 진본성 등에 관한 법·제도적 문제가 웹 아카이빙에서 제기되는 또 하나의 중요한 문제이다. 이러한 맥락에서 웹 정보자원이 가지는 기술적 법·제도적 특성을 파악하는 것은 필수적이라 할 수 있다. 이 논문은 웹 정보자원 특성에 기반하여 웹 아카이빙과 관련된 기술적 문제와 법·제도적 측면을 살펴보고, 이를 통해 웹 아카이빙의 발전을 위한 우리의 과제를 모색해보고자 한다.

### ABSTRACT

In terms of archiving, an importance of the Web is more increasing. Since the Internet has been popularized, many archivists have made efforts to preserve informational, cultural, and evidential values of the Web. Information resources on the Web raise archival issues that are significantly different from issues of preservation for tangible materials, because they are based on unique characteristics of the Web. One of issues is about a technical aspect which is based on technical architecture of Web. Another important issue is related to legal problems including copyright, authenticity and so on. In this context, it is essential to understand technical and legal characteristics of the Web. A purpose of this article is to review technical and legal aspects of Web archiving and to find an agenda for development of Web archiving.

키워드: 웹 아카이빙, 웹 정보자원, 저작권, 납본, 공공기록물 관리  
Web Archiving, Information Resource on the Web, Copyright, Legal Deposit,  
Public Records Management

\* 중앙대학교 문헌정보학과 강사(yskim@icogito.net)  
논문접수일자 2007년 8월 8일  
게재확정일자 2007년 9월 27일

## 1. 문제제기

아톰(Atom)의 시대는 가고 비트(Bit)의 시대가 도래하였다(Negroponete 1996). 아카이빙의 영역에도 예외는 아니다. 아카이브 패러다임의 변화를 분석한 압들아지즈(2005)의 주장을 빌리지 않더라도 오늘날 생산·소통되는 개인적인 혹은 공적인 정보자원의 상당 부분이 다양한 디지털 양식에 기반하고 있음은 부정할 수 없는 사실이다. 특히 지난 10여 년 간의 웹의 성장은 우리가 일상에서 정보자원을 생산하고 이용하고 나누고 보존하는 양식에 큰 변화를 일으켜왔다.

이제껏 인류가 만들어낸 어떠한 매체보다 더 빠른 속도로 성장하고 있는 웹은, 그 짧은 연혁에도 불구하고, 오늘날 정보 이용자들이 가장 빠르고 손쉽게 접근할 수 있는 정보 매체의 하나가 되었다. 이제 사적 커뮤니케이션의 영역에서부터 공적 커뮤니케이션의 영역까지, 그리고 학술연구 분야에서부터 상거래의 영역까지 국경을 초월하는 전지구적 뉴미디어로 성장한 웹이 우리의 삶에 가지는 중요성은 새삼 강조할 필요조차 없을 것이다. 이러한 맥락에서, 레비(Levy 2000)는 디지털 정보자원의 구조화와 정리를 가능케 하는 아카이브의 구축에 지식정보사회의 성패가 달렸다고까지 주장하였고, 정보자원의 기록과 보존이라는 측면에서 웹이 차지하는 가치는 날로 증대되고 있다.

웹의 정보적, 문화적 그리고 증거적 가치를

보존하고자 하는 노력들은 웹의 본격적인 대중화와 함께 전개되기 시작했다. 최초의 웹 아카이빙 프로젝트로 알려진 1994년 캐나다 국립도서관의 전자간행물 파일럿 프로젝트(Electronic Publication Pilot Project, 이하 EPPP)<sup>1)</sup>를 비롯하여, 미국의 인터넷 아카이브(Internet Archive)<sup>2)</sup>, 호주의 판도라(PANDORA)<sup>3)</sup> 등 각국의 웹 아카이빙 노력들이 확대되고 있으며, 2000년 유네스코(UNESCO)는 온라인 디지털 정보를 납본 대상에 포함시킬 것을 권고하는 지침을 발표한 바 있다(Larivière 2000).<sup>4)</sup>

아톰 시대의 역사적 기록과 유물의 보존은 결코 손쉬운 일이 아니다. 아키비스트들을 비롯한 많은 전문가들의 노력에도 불구하고 무수히 많은 문화적, 역사적 유물들이 사라져갔다. 자연적 부식에 의해, 때론 부주의 혹은 예기치 못한 사고로, 심지어는 고의에 의한 훼손으로 인해 후대에 전해지지 못했다. 이에 비하면, 네트워크에 연결된 컴퓨터들의 저장장치 속에 비트의 형태로 저장되어, 복제, 배포될 수 있는 웹 정보자원들의 보존은 상대적으로 쉽다고 말할 수도 있을 것이다.

하지만 디지털 유산, 비트의 형태로 된 디지털 기록이라고 해서 세월 속에 훼손되지 않는 것은 아니다. 아날로그 방식으로 기록된 기록물의 경우, 육안이나 간단한 장비를 통해 그 내용을 읽을 수 있고, 보존 시 별도의 가독 장치를 필요로 하지 않는다(윤대현, 남성운 2003, 4). 반면 비트를 담고 있는 매체의 수명이 상대적

1) EPPP는 1996년 6월 최종 보고서를 제출하였다. 이 보고서는 캐나다 정부의 기록보존서비스프로그램 웹사이트를 통해 열람할 수 있다(Electronic Publication Pilot Project team and Electronic Collections Committee, 1996).

2) <http://www.archive.org>

3) <http://pandora.nla.gov.au>

4) 룬 박사(Dr. Jean Lunn)의 1981년 유네스코(UNESCO) 납본 지침 개정판.

으로 짧고, 하루가 다르게 변화하는 컴퓨팅 환경 속에서 현재의 디지털 기록이 수십 년 후 여전히 이용 가능한가의 여부는 매우 불투명하다 (Stanescu 2004). 오늘날의 최신 소프트웨어로 생성된 파일이라도 10여 년 후의 이용가능성에 대해서는 누구도 장담할 수 없는 것이 현실이다.<sup>5)</sup>

웹을 기반으로 하는 정보자원의 보존은 웹의 고유한 매체 특성으로 인해 기존의 인쇄물을 중심으로 하는 유형 기록물의 보존에서 나타는 것과는 현저하게 다른 문제점들을 나타내고 있다. 웹 아키텍처의 특성에 기반 한 기술적 측면의 문제들이 그 중 하나이고, 웹 정보자원의 저작권, 진본성 등에 관한 법·제도적 문제가 웹 아카이빙에서 제기되는 또 하나의 중요한 문제이다. 그러나 웹 아카이빙과 관련된 대부분의 현실적 문제들은 기술적 측면과 법적 측면의 문제가 명확히 분리되어 나타나지 않는다. 하부구조로서의 기술적 측면이 법·제도에 직접적인 영향을 미치고 있고, 또한 기술의 발전 방향에 법·제도가 큰 영향을 주고 있기 때문이다. 따라서 웹 아카이빙에서의 문제들은 기술적 측면과 법·제도적 측면을 통합적으로 고려하는 해법을 요구하고 있다.

이러한 맥락에서 웹 정보자원이 가지는 기술적, 법·제도적 특성을 파악하는 것은 필수적이라 할 수 있다. 이 논문은 웹 정보자원 특성에 기반 하여 웹 아카이빙과 관련된 기술적 문제

와 법·제도적 측면을 포괄적으로 살펴보고, 이를 통해 웹 아카이빙의 발전을 위한 우리의 과제를 모색해보고자 한다.

## 2. 웹 정보자원의 기술적 특성 분석: 아카이빙의 측면에서

### 2.1 분산성

웹 아카이빙과 관련하여 논의되는 분산성이란 웹 정보자원 특성으로서의 그것과 웹 아카이빙으로의 접근방식을 일컫는 이중적 의미를 갖는다. 주지하는 바, 웹은 어느 단일 조직의 지배를 받지 않는 탈중심적 매체이자 정보자원이다 (Poster 1997, 204). 아카이브의 측면에서도 웹의 보존에 관련된 정책이나 기준을 제시하고 통제할 단일 기구는 존재하지 않는다. 웹은 스스로의 기술적 근간인 분산 네트워크(Distributed Network)<sup>6)</sup> 아키텍처의 특성을 그대로 발현하는 탈중심적인 정보자원으로 발전하여 왔다. 신문, 방송 등 전통적 매체에서의 정보 생산·유통의 방식과는 달리, 웹에서의 많은 정보들은 사용자 스스로에 의해 분산적으로 생산되며 결정되어진다. 이러한 맥락에서 웹 보존 기관들은 웹의 특정 부분에 집중하려는 경향을 띤다. 고고학자, 미술가, 기록관리자들은 제각기 다른 전문적 시각으로 웹 아카이빙에 접근하고

5) 예를 들어, 1086년 윌리엄 1세에 의해 만들어진 토지 대장의 사본들은 천 년 가까운 세월이 지난 오늘까지 남아있지만, 영국 국영방송 BBC가 원본 토지대장의 900주년을 기념하기 위해 제작한 인터랙티브 비디오디스크 '1980년 돔스데이 프로젝트(Domesday Project)'는 오늘날 사실상 이용이 불가능하다(Charleworth, Andrew 2003a).

6) 1960년대초, 미국정책연구기관 랜드(RAND)는 미 공군의 지원을 받아 중앙통제기구를 가지지 않는 패킷-스위칭(packet-switching) 네트워크를 연구한다. 이 연구는 폴 바란(Paul Baran)을 주저자로 1964년 출간된 '분산 커뮤니케이션에서(On Distributed Communications)'라는 제목의 보고서들로 결실을 맺는다(Baran, 1964).

있다. 아키비스트가 웹 사이트들의 기록관리적 측면에 주 관심을 두고 있다면, 또 다른 학자들은 장기적 관점에서의 웹 사이트들의 사회 정치적 중요성에 보다 많은 관심을 쏟을 것이다. 아카이빙에서 중복을 피하고 사용자의 접근을 보장하기 위해서는 각 조직 간의 긴밀한 상호 협력이 필수적이라는 난점에도 불구하고, 웹 아카이빙에서의 이러한 분산적 접근방식은 매우 유용한 것으로 평가되고 있다(Day, 2003, 8). 이러한 맥락에서, 정보의 다원성과 다양성을 그 특징으로 하는 웹 정보의 보존과 관리를 단일 국가 기관이 모두 책임지기는 어렵기 때문에, 다양한 관리 기관들의 협력 체제 구성이 필요하다는 주장이 설득력을 얻고 있다(서혜란, 2004, 17).

## 2.2 휘발성

웹은 나타났다가는 사라진다. 수시로 업데이트되고 재구성되기도 한다. 내용과 구조의 추가, 수정, 삭제 간에 지속적이고 일상적인 변화가 일어나는 것이 웹의 특성이다. URL은 인터넷 도메인 이름에 기반하고 있기 때문에, 때론 사라지고 심지어는 소유자가 바뀌기도 한다. 이런 이유들로 인해 웹에서는 슬한 브로큰 링크들(broken links)이 생겨나고, 우리는 웹 브라우저에서 해당 정보 혹은 문서를 찾을 수 없음을 뜻하는 'HTTP Error 404 -- Not Found'라는 메시지와 일상적으로 마주치게 된다. 더욱이, 웹을 학술적으로 인용하고자 하는 사람들은 웹 정보의 짧은 수명과 휘발성에 당혹해 한다. 알렉스 인터넷(Alex Internet, <http://www.alex.com>)은 웹 페이지들의 평

균 수명을 약 75일로 추정한 바 있다(Lawrence, et al, 2001, 30). 영국 브리스톨 대학(University of Bristol)의 앤드류 찰스워즈(Andrew Charlesworth)는 웹 정보에 대한 법학적 인용의 문제점을 지적하면서, 단행본 책의 경우 한 장(章)에서 인용된 웹문서의 약 65%는 1년 안에 사라지거나 URL이 변경되며, 한 학술지 논문에 인용된 웹 문서의 50%는 논문이 발행되기도 전에 원래의 위치에 변동이 생긴다고 보고했다(Charlesworth 2003b).

웹 환경에서 대량의 정보자원이 생성되고 또 사라지는 일상적 현상은 웹이 가지는 생태적 진화의 속성이라는 측면에서 어쩔 수 없는 손실이며(민경배, 이한메 2004), 웹의 성공 원인 중 하나라고도 일컬어진다. 하지만, 이러한 웹의 특성은 웹 정보자원의 신뢰성이라는 측면에서는 매우 부정적인 요인으로 작용한다. 어느 한 시점에서 우리가 확인한 대부분의 웹 링크들이 언젠가는 작동하지 않거나 잘못된 웹사이트로 연결될 수 있음을 의미하기 때문이다. 웹을 비롯한 디지털 형식의 기록물들은 태생적으로 내용 - 구조 - 맥락의 요소가 분리되어 있는 논리적 객체이기 때문에 시스템에 대한 사전적 분석을 통해 이를 파악해야 한다(임진희 2006a). 이런 의미에서 웹 인용은 충분한 맥락 정보를 포함할 것을 요구받으며, 이를 통해 사용자가 검색엔진을 이용해서라도 찾을 수 있도록 만들어야 한다.

## 2.3 역동성

웹은 매우 다양한 형태로 존재한다. 모든 방문자에게 동일한 정보를 보여주는 정적(static)

인 웹 사이트에서부터 특정 질의에 대해 역동적인 정보를 제공하는 사이트까지 있다. 사용자 분석을 기반으로, 사용자의 선호를 파악하고 개별화된 정보를 제공하기도 한다. 이러한 동적(dynamic) 웹 사이트들은 디지털 형태의 간행물이라기보다는 '상호작용적인 응용 소프트웨어(Interactive software application)'에 가깝다고 할 수 있다(National Archives of Australia 2001). 반면 정적인 웹 사이트는 하이퍼링크라는 연결고리를 가지고 서버 컴퓨터에 저장된 정적인 문서의 모음에 다름 아니다. 이 경우 웹 사이트는 항상 동일한 URL을 가지며, 한 문서에서 다른 문서로 이동할 수 있는 링크가 상호작용성의 전부이다. 따라서 정적인 웹의 아카이빙은 상대적으로 손쉽다 할 수 있다.

다수의 사이트들은 사용자들로부터의 코멘트나 요구와 같은 정보를 수집하기 위한 폼(form)을 사용한다. 이들 사이트들은 많은 부분을 정적 간행물 메커니즘에 의존하고 있음에도 불구하고, 폼을 작동하게 하는 스크립트 혹은 프로그램과 방문자가 폼을 작성할 때 제공되는 정보 등 역동성을 수반한다.

정적인 웹과는 달리, 또 다른 일단의 웹 사이트들은 특정 데이터베이스로의 접근을 위한 사용자 인터페이스(user interface) 또는 프런트엔드(front-end)로 사용된다. 사용자들은 사이트가 제공하는 목록을 검색하거나 검색어를 입력한다. 질의에 응답한 정보는 사용자들에게 각각의 고유한 식별자를 가진 HTML 문서로 표시된다. 이렇듯 역동적 데이터를 제공하는 사이트들의 아카이빙에는 정적인 웹의 경우와는 다른 문제가 제기된다. 우선, 다수의 사용자가 동일한 정보를 얻고자 할 때, 모두가 동일한

웹 정보를 얻는 것이 아니라는 것이다. 사용자의 브라우저에 표시되는 페이지는 사용자의 개별적 질의에 기초한다. 따라서 사용자의 질의는 결과물을 실질적으로 생성하는 과정의 하나로 기록되어야 할지도 모른다. 데이터베이스의 사용 결과는 사용자의 기호, 능력, 접근 권한 등에 따라 변화할 수 있으므로, 기록하기 손쉬운 단일한 형태로 존재하지 않는다. 또한, 정적인 웹의 속성을 가진 메인 페이지 혹은 최상위 페이지에 변화가 없더라도 사이트 뒤에 감추어진 데이터베이스에 저장된 정보는 지속적으로 변화할 수 있고, 겉으로 드러나지도 않는다. 전체 웹에서 이러한 역동적 웹의 비중은 날로 증가하고 있어, 웹을 다루는 아키비스트들은 역동적으로 생성되는 웹 정보자원을 완벽히 작동하는 상태로 보존해야 하는 기술적 문제들에 직면해 있다.

## 2.4 심층 웹

1994년 질 엘스워즈(Jill Ellsworth)가 “보이지 않는 웹(Invisible Web)”이라는 용어를 최초로 사용하면서 본격적인 학술적 논의의 대상이 된(Bergman 2001) 심층 웹(Deep Web)은 아무런 제약 없이 접근이 개방되어 있는 웹 사이트들을 일컫는 표면 웹(Surface Web)의 반대 개념이다. 웹 아카이빙에서 심층 웹이 중요한 문제로 제기되는 것은 정확히 알려지지 않는 심층 웹의 크기와 제한된 접근성 때문이다. 바-일란(Bar-Ilan 2001, 9)은 알려진 웹의 예측 가능한 크기는 검색엔진과 사용자들이 자유로이 접근 가능한 정적인 웹 페이지들만을 포괄한다는 점을 지적하였는데, 데이터베이스

에 기반한 사이트, 패스워드 보호 등과 같은 접근 장벽을 가지고 있는 사이트, 마이크로소프트(Microsoft)사의 ASP(Active Server Page)를 통해 제공되는 사이트, 그리고 호환성이 낮은 특정한 포맷을 사용하는 사이트 등이 심층 웹을 구성하는 대표적인 예라고 할 수 있다. 한 보고서는 심층 웹의 크기가 표면 웹보다 400~500배 크다고 추정한 바 있다(Bergman 2001). 2003년의 또 다른 연구는 표면 웹의 크기를 167테라바이트(Terabyte)로, 심층 웹의 크기를 표면 웹의 550배에 달하는 91,850테라바이트(Terabyte)로 추정하였다(Lyman & Hal 2003). 심층 웹은 웹 아카이빙의 어려움을 단적으로 보여준다. 심층 웹은 숨어있는 웹(hidden Web)이다. 그 이유는 상업적인 것일 수도, 조직적 이익을 위한 것일 수도, 혹은 단순한 개인의 선택일 수도 있다. 하지만 그 이유를 불문하고, 심층 웹은 로봇의 자동 수집을 거부하며, 수집을 위해 웹 콘텐츠의 저작권자와의 협의 절차를 필요로 한다는 공통점이 있다. 이는 저작권 문제와 직결됨을 의미한다. 심층 웹의 크기가 표면 웹의 몇백 배에 달함을 감안할 때, 심층 웹에 대한 선별적 아카이빙은 매우 노동집약적, 고비용 작업일 수밖에 없다.

## 2.5 진화하는 웹

인터넷은 기술적인 측면과 사회 경제적인 측면 모두에서 아직도 끊임없이 진화하고 있다.

인터넷 접속 속도는 끝없이 향상되고, 기술은 진보되고 있으며, 그 속에 사용자들의 행태도 빠르게 변하고 있다.

우리나라의 경우, 1990년대 중반 전화모뎀의 28.8Kbps에 불과했던 인터넷 접속 속도는 90년대 후반부터 시작된 디지털가입자회선 또는 케이블 모뎀 등의 보급으로 10Mbps를 상회하고 있으며, 이미 100Mbps 케이블 모뎀 서비스는 물론, 동일 속도의 초고속디지털가입자회선(VDSL)도 상용화되었다(최경섭 2006). 이 빠르고 항상 켜있는 인터넷 접속으로 인해, 인터넷 콘텐츠의 형태와 양은 급속하게 변하고 있다.<sup>7)</sup>

가장 기본적인 웹 표준과 프로토콜은 1990년대 이래로 비교적 안정적으로 남아있는 반면, 어떤 웹 사이트의 관리 기술과 방법에는 엄청난 변화가 일어났다. 많은 웹 콘텐츠가 역동적 데이터베이스로부터 제공되기 시작했다. 소프트웨어와 데이터베이스 구조에 대한 메타데이터와 같은 상세 정보 없이는 보존소에 사본을 둘 수도 없는 상황이다. 플래시(Flash) 등 플러그인(plug-in)과 같은 특정 소프트웨어를 사용하거나, 일부 브라우저에서는 작동하지 않는 비표준적 특성을 적용한 사이트들은 광범위한 이용이 불가능할 수도 있다.

웹의 창안자 팀 버너스 리(Tim Berners-Lee)는 1998년도부터 개방과 공유라는 웹 본연의 의미와 잠재력을 찾고자 하는 자동화 지능화된 차세대 웹인 시맨틱 웹(Semantic Web)을 제안하고 개발을 주도하고 있으며, 플랫폼이 기

7) 사실상 대부분의 인터넷 가입자가 광대역 인터넷을 사용하는 우리나라의 경우, 대역 회선의 보급이 저조한 나라들에 비해 인터넷의 방송 오락 대체적 성격이 확연히 강화되고 있음을 볼 수 있다. 한국인터넷진흥원이 발표한 통계에 의하면 2007년 3월 기준 초고속인터넷 가입자 수는 1400여 만 명을 돌파하여, 2006년 3월 대비 약 14%의 고성장을 기록하였다(한국인터넷진흥원 2007, 111).

반 환경이 되는 웹을 의미하는 ‘웹2.0’에 대한 논의도 뜨겁다. 웹을 둘러싼 환경의 이러한 변화들은 웹 사이트를 수집하고 보존하고자 하는 이들에게 당면한 기술적 과제를 제기하고 있다.

### 3. 웹 아카이빙의 법·제도적 문제

#### 3.1 저작권

이론상으로, 그리고 원칙적으로 웹을 비롯한 디지털 유산의 가장 큰 장점은 빠르고 저렴하고 완벽하게 복제될 수 있다는 것이다. 하지만, 실제 디지털 유산의 복제를 고려할 때 이 장점들은 저작권의 문제에 봉착한다. 경제적 이익을 목적으로 생산되거나 유통되는 상업적 디지털 저작물의 경우, 저자나 저작권자의 허가 없는 디지털 자료의 광범위한 복제는 경제적 이익의 현저한 감소로 이어질 가능성이 높다. 디지털 음악 정보인 MP3 파일의 광범위한 불법적 유통이 음반 산업에 끼친 영향이 그 좋은 예다. 디지털 음악 정보의 순수한 복제로 인해 생산자나 저작권자는 자원의 재생산에 대한 통제력을 잃었고 이는 경제적 손실로 직결되고 있다.

물론 이러한 문제가 디지털 자원에만 국한되는 것은 아니다. 지적 재산과 관련된 일련의 법률들, 특히 저작권에 관련된 법률은 다양한 저

작들에 대한 재정적·지적 투자를 위한 일정한 보호를 제공한다. 따라서 대부분 국가의 법률 시스템은 저작권법이라는 일정한 법적 보호막을 제공한다.<sup>8)</sup> 이를 통해 디지털 자원의 사용자가 복제할 수 있는 범위를 통제할 수 있는 권한을 저작권자에게 제공한다.<sup>9)</sup> 하지만 디지털 정보자원의 불법적 복제물을 만드는 데 요구되는 투자, 기술, 노력은 기존의 유형 기록물의 복제에 요구되는 것들보다 현저하게 낮은 것이 일반적이다. 아날로그 방식의 복제는 음원의 질 저하가 불가피한 반면, 디지털 음악 포맷의 등장과 인터넷의 연계, P2P 기술의 발전은 고품질의 음악 정보를 손쉽게 인터넷으로부터 다운로드받는 것을 일상으로 만들었다.

이런 이유들로 인해, 디지털 자료의 저작권자들은 기존 전통적인 대체에 적용되어오던 법·제도 위에 추가적인 법적 통제를 통해, 자신들 디지털 자원의 통제권을 확보하려 하기 시작했다. 여기서 추가적인 법적 통제는 일반적인 저작권 법률의 개정과 디지털 자원의 복제와 유통을 제한하는 계약조항의 광범위한 사용이라 할 수 있다.

미국의 1998년 디지털 밀레니엄 저작권법(Digital Millennium Copyright Act, 이하 DMCA)<sup>10)</sup>과 유럽연합의 2001년 저작권 지침(Copyright Directive 2001/29/EC)은 저작권자의 허가 없이 디지털 저작물을 복제하는 것을 불법으로

8) 우리나라에서는 1957년 “학술적 또는 예술적 저작물의 저작자를 보호하여 민족문화의 향상발전을 도모함을 목적으로” 『저작권법』이 최초로 제정된 이래, 열네 차례의 일부 혹은 전문 개정을 거쳐 오늘에 이르고 있다. 2006년 12월 28일 전부개정을 통해 법률 제8101호로 공포된 『저작권법』은 “저작자의 권리와 이에 인접하는 권리를 보호하고 저작물의 공정한 이용을 도모함으로써 문화의 향상발전”을 법의 목적으로 규정하고 있다.

9) 『저작권법』 제30조는 사적 이용을 위한 복제를 허용하고 있다. 이 조항에 따르면 “공표된 저작물을 영리를 목적으로 하지 아니하고 개인적으로 이용하거나 가정 및 이에 준하는 한정된 범위 안에서 이용하는 경우에는 그 이용자는 이를 복제할 수 있다”고 규정하고 있다.

10) DMCA는 1996년 채택된 ‘WIPO저작권조약’과 ‘WIPO실연 및 음반조약’의 이행을 위해 미국 상원의 만장일치와

규정하고 있는 것뿐만 아니라, 저작권자가 복제 방지를 위해 저작물에 설치한 어떠한 기술적 통제를 제거하거나 우회하는 것조차도 범죄행위로 규정하고 있다. 이는 저작권에 대한 실질적인 침해가 없는 경우에도 적용된다. 2002년 제정된 우리나라의 『온라인디지털콘텐츠산업발전법』<sup>11)</sup> 제18조 또한 “누구든지 정당한 권한없이 타인이 상당한 노력으로 제작하여 표시한 온라인콘텐츠의 전부 또는 상당한 부분을 복제 또는 전송하는 방법으로 경쟁사업자의 영업에 관한 이익을 침해하여서는 아니된다”고 명시하면서, “디지털콘텐츠에 적용한 기술적보호조치의 회피·제거 또는 변경[...]을 주된 목적으로 하는 기술·서비스·장치 또는 그 주요부품을 제공·수입·제조·양도·대여 또는 전송하거나 양도·대여를 위하여 전시하는 행위”를 금지하고 있다.<sup>12)</sup>

디지털 정보자원을 다수의 일반 대중이 사용할 수 있는 형태로 보존하기 위해서 도서관과 아키비스트들은 원본의 백업이든 대체물이든 간에 어떠한 형태의 복제물을 만들어야 한다. 경우에 따라서는 하나의 매체에서 또 다른 매체로, 하나의 포맷에서 또 다른 포맷으로의 마이그레이션(migration)이 필요할 수도 있다. 이러한 디지털 저작물에 대한 복제의 필요성은 필연적으로 저작권의 문제를 야기한다.

많은 나라의 저작권법이 현존하는 유형물의 복제를 허용하거나 영구 컬렉션으로 대체하는 것을 허용하고 있지만, 디지털 기록물의 경우 좀 더 복잡한 양상을 띤다. 우리나라는 『저작권법』<sup>13)</sup>을 통해 도서관에 온라인 디지털 정보자료의 복제를 기본적으로 허용하고 있지만, 상당한 예외 규정들도 함께 제시되고 있다. 『저작권법』 제31조는 『도서관법』<sup>14)</sup>에 따른 도서관과 도서·문서·기록 그 밖의 자료를 공중의 이용에 제공하는 시설이 보관된 도서 등을 사용하여 저작물을 복제할 수 있도록 허용하고 있다. 하지만, 디지털 형태의 복제는 도서 등의 자체 보존을 위하여 필요한 경우에만 허용되고, “조사·연구를 목적으로 하는 이용자의 요구에 따라 공표된 도서 등의 일부분의 복제물을 1인1부에 한하여 제공하는 경우”와 “다른 도서관 등의 요구에 따라 절판 그 밖에 이에 준하는 사유로 구하기 어려운 도서 등의 복제물을 보존용으로 제공하는 경우”는 디지털 형태로 복제할 수 없도록 규정하고 있다. 또한 컴퓨터 등을 이용하여 이용자가 도서관 안에서 열람할 수 있도록 보관된 도서들을 복제하거나 전송할 수 있도록 허용하면서도, 이미 디지털 형태로 판매되고 있는 도서 등의 복제를 금지한다. 디지털 형태로 판매되지 않은 경우에도, 열람할 수 있는 이용자의 수를 해당 도서관에

클린턴 대통령의 서명으로 1998년 10월 28일 제정되었다. 이 법은 온라인 저작권을 강화하고 이를 침해하는 기술 개발을 불법화하는 것을 주내용으로 하고 있다. DMAC은 WIPO저작권 조약 및 실연·음반조약의 이행, 온라인서비스제공업자의 책임제한, 컴퓨터 유지보수 목적의 복제에 대한 면책 등에 관한 규정들을 주 내용으로 하고 있다. 이 법의 가장 큰 의의는 온라인상의 저작권에 대한 보호를 강화하고 이를 침해하는 기술의 개발을 불법을 규정한 데 있다.

11) 일부개정 2006.9.27 법률 제7988호.

12) 동법 제22조는 이를 위반한 자를 1년 이하의 징역, 2천만 원 이하의 벌금에 처하도록 규정하고 있다.

13) 전부개정 2006.12.28 법률 제8101호.

14) 일부개정 2006.12.20 법률 제8069호.



서 보관하고 있거나 저작권자로부터 이용허가를 받은 부수를 초과하지 못하도록 제한하고 있다. 이는 미국의 DMCA가 도서관과 아카이브들이 보존 목적으로 특정 저작물을 3부까지 복제할 수 있도록 규정한 것과 같은 맥락이라고 할 수 있다(Hirtle 2003).

하지만 디지털 정보자원 생산자를 보호하고자 하는 이러한 일련의 제재들은 불가피하게 저작권자의 권리와 일반 사용자들의 권리 사이의 균형에 영향을 끼친다. 특히 저작권에 대한 보호가 강화될수록 공정사용(fair use) 혹은 공정거래(fair dealing)와 같은 사용자의 권리, 도서관의 특권은 축소될 수밖에 없다.

### 3.2 납본과 수집

인터넷의 등장 이전, 더 정확히는 웹의 대중화 이전, 책, 저널, 혹은 팜플렛 등과 같은 다양한 매체들로부터 정보를 함께 모으고, 이것들의 복제물을 광범위하게 유통하는 능력은 국제적 출판사, 정부 조직, 국제기구와 같은 소수 조직의 고유한 영역이었다(Charlesworth 2003a). 이와 대조적으로 전통적 매체가 가지고 있었던 정보 제공자로서의 진입장벽이 허물어진 웹 환경에 기인한 정보량과 정보제공자 수의 폭발적 증가는 웹 보존의 문제를 더욱 어렵게 만들고 있다. 만약 정보가 단지 양적으로만 늘어난 것이라면 질적 통제의 방법으로 게이트키퍼의 역할을 하는 기존 매체들에 의존할 수 있을 것이다. 하지만 정보량과 정보제공자의 수가 무한증식하는 웹 환경은 웹 정보자원의 선별과 수집 활동의 효율성과 실효성 담보에 큰 어려움을 주고 있다. 이런 맥락에서 웹 정보자원 보존의 효과적인 방

안의 하나로 대두되는 것이 납본제도이다.

#### 3.2.1 납본의 개념

2006년 12월 20일 일부 개정된 『도서관법』 제2조 8항은 납본의 개념을 “자료를 발행하거나 제작한 자가 일부 부수를 법령에서 정한 기관에 의무적으로 제출하는 것을 말한다.”고 규정하고 있다. 납본의 목적은 자국의 출판 저작물을 완벽하게 수집·보존하고 국가서지를 작성함으로써 당대의 정보유통을 보장하고 후대에 지적 문화적 유산을 전달 계승코자 하는 데 있다(서혜란 2003, 374). 다시 말해, 납본제도는 “국가의 지적 문화유산의 수집 및 보존, 도서관 및 정보서비스를 위한 인프라 구축, 국가 차원의 망라적 정보서비스 제공, 최종 이용자를 위한 직접봉사”를 위한 역할을 담당해왔다(윤희운 2003, 30). 이런 의미에서 웹 정보자원의 중요성을 고려해볼 때, 웹 정보자원의 납본에 대한 논의들은 아키비스트들의 뜨거운 관심의 대상이 되어왔다. 기술적 측면으로 보면, 웹 정보자원은 특정 서버 컴퓨터에 유일본으로 존재하기 때문에 전통적 납본 개념인 복본 출판의 조건에 부합하지 않다고도 할 수 있으나, 웹 정보자원은 패킷(Packet) 형태로 나뉘어 TCP/IP를 통해 전송된 뒤, 클라이언트 컴퓨터에 정보를 임시적으로 복제·재생성하는 과정을 통해 다수의 대중이 이용할 수 있도록 제공된다는 점에서 납본의 대상이라 할 수 있을 것이다(Larivière 2000, 41).

『도서관법』 제20조는 『공공기록물관리예관한법률』<sup>15)</sup> 제3조 제1호에 따른 공공기관의 자료의 경우, 발행 또는 제작일로부터 30일 이내에 국립중앙도서관에 납본하여야 한다고 명시

하고 있으며, 『도서관법 시행령』 제13조에 명시된 국립중앙도서관에 납본하여야 할 자료 중에는 “『출판및인쇄진흥법』<sup>16)</sup> 제2조 제6호에 따른 전자출판물<sup>17)</sup> 중 콤팩트디스크 디지털비디오 등 유형물”과 함께 “출판환경의 변화에 따라 새로운 형태로 발간되는 기록물로서 문화관광부장관이 인정하는 자료”를 포함하고 있다. 문제는 여기에서 “새로운 형태로 발간되는 기록물”로서 웹을 포괄할 수 있는지의 문제다.

### 3.2.2 납본 및 수집 대상의 범위

이렇듯 웹 정보자원의 납본과 관련하여서는 납본 대상의 범위가 첫 번째 문제로 제기되고 있다. 2000년 유네스코 납본지침은 온라인 간행물의 범위를 매우 폭넓게 설정하여, 접속번호, 패스워드 등으로 접근이 통제되는 데이터베이스 안에 저장된 독립 단위에서부터 사용자의 자유로운 정보 접근과 생산을 허용하는 웹사이트까지를 포괄하고 있다. 이는 온라인에서 쉽게 접근할 수 있는 전자저널, 멀티미디어 자원 등도 포함하고 있다(Larivière 2000). 반면 캐나다 국립도서관의 EPPP는<sup>18)</sup> e-간행물의 범주를 상대적으로 좁게 설정하여, 온라인 데이터베이스, FTP, BBS를 수집, 보존되어야 하는 e-간행물의 범주에서 제외하고, 심지어는

웹과 고퍼(Gopher) 사이트에까지 e-간행물로서의 조건을 제시하고 있다. 웹의 경우, 멀티미디어 혹은 하이퍼텍스트 요소와 상관없이, 전통적 간행물로서의 특징을 가져야 하며, 고퍼의 경우도 웹에서와 동일한 조건을 충족하는 것을 전제로 e-간행물로 간주하고 있다(Electronic Publication Pilot Project team and Electronic Collections Committee 1996, 8-9).

국립중앙도서관의 OASIS(Online Archiving & Searching Internet Sources) 프로젝트가<sup>19)</sup> 2006년 발표한 온라인 디지털자원 선정지침은 ‘온라인 디지털자원’을 “인터넷을 통해 전자화된 디지털자원의 내용을 내려 받아 개인용 및 휴대용 컴퓨터 등의 정보통신 단말기에서 볼 수 있는 디지털 버전의 출판물”과 “인터넷 등을 통해 내려 받아 개인용 컴퓨터, 노트북 컴퓨터, PDA 등의 정보통신 단말기에서 읽을 수 있는 전자적으로 유통되는 텍스트, 소리, 동영상정보 등을 담고 있는 컴퓨터 파일”로 규정하고, 구체적인 범위를 웹사이트, 웹페이지, pdf, hwp, doc, txt 등의 파일, 이미지, 동영상, 음악, 압축파일 등으로 제시하였다(국립중앙도서관 2006a). 또한 일반 선정지침으로는, 웹 정보자원의 국적판별의 기준으로 서,<sup>20)</sup> 한국에 관한 것, 한국 도메인(.kr)상에 있는 것, 한국과 관련된 사회, 정치, 문화, 종교, 과학

15) 일부개정 2007.4.27 법률 제8395호

16) 『출판및인쇄진흥법』은 2007년 7월 19일 공포된 일부 개정안을 통해 법명을 『출판문화산업진흥법』(법률 제8533호)으로 변경하였다.

17) 『출판문화산업진흥법』는 전자출판물의 개념을 동법에 의하며 신고한 출판사가 “저작물 등의 내용을 전자적 매체에 실어 이용자가 컴퓨터 등 정보처리장치를 이용하여 읽거나 보고 들을 수 있도록 발행한 전자책 등의 간행물”로 규정하고 있다.

18) 각주 1) 참조.

19) <http://www.oasis.go.kr>

20) 1996년 유럽위원회는 디지털 정보자원의 국적 판별 기준으로 ‘생산기관의 위치’, ‘제1저자의 주소’, ‘저자의 국적’, ‘저작물의 네트워크 주요 위치’를 제시하였다(European Commission, 1996).

또는 경제적으로 중요한 주제이고 한국 저자에 의해 쓰인 것, 권위 있는 한국 사람에 의해 저술되거나 국내·외적으로 해당 학문분야에 기여를 한 것, 해외 온라인 디지털자료의 경우에는 한국 사람에 의해서 쓰이거나 한국을 주제로 다룬 것, 다섯 가지를 제시하고 있다.

이와 함께, 수집 우선 자료로 중앙정부가 생산한 온라인 디지털자원, 대학 간행물, 회의자료,<sup>21)</sup> 전자저널, 도서관 온라인 디지털자원심의위원회규정이 추천한 온라인 디지털자원, 정치 사회적으로 이슈가 되는 온라인 디지털자원, 정부기관, 도서관, 대학교, 연구기관 등의 웹사이트 등을 들면서, 수집의 대상에서 제외되는 것으로 블로그 사이트, CAMS,<sup>22)</sup> 토론리스트, 채팅 사이트, 게시판과 뉴스 그룹, 뉴스 사이트, 인쇄물이 존재하는 온라인 일간 신문, 포털 사이트, 광고 사이트 등을 규정하였다. 한편, 지방행정기관과 기타 정부 투자기관의 온라인 디지털자원, 교육 목적의 우수자료, 연보, 아동용 간행물, 문학작품, 온라인상에서만 서비스되는 신문, 특수한 연구 가치가 있는 정보를 제공하는 개인 혹은 조직의 사이트, 종교 사이트, 만화, 게임, 비디오, 애니메이션 등을 제한적 수집 대상으로 선정하였다.

### 3.2.3 납본 및 수집의 방법과 절차

납본과 관련된 또 하나의 문제는 납본의 방법

과 절차에 관한 것이다. 신문·잡지와 같은 종이매체 혹은 CD-Rom과 같은 오프라인 디지털 매체처럼 유형의 매체 형식에 고정되어 있는 정보자원과는 달리, 웹 정보자원은 무형의 비트 형태로 존재하고 일상적으로 변화하는 콘텐츠를 담고 있다는 점에서 전통적인 납본의 방식과는 다른 수집방법과 절차가 요구된다. 따라서 각국의 웹 아카이빙 기관들은 별도의 수집 서버를 두고 웹 로봇과 자동분류기 등의 기술적 해법을 사용하고 있다. 국립중앙도서관의 OASIS의 경우, 관리자와 일반사용자의 요청을 처리하는 웹서버, 데이터 관리와 검색 기능을 수행하고 수집된 파일을 체계적으로 관리하는 데이터베이스서버와 함께, 웹로봇 에이전트, 자동분류기, 자동요약기 등을 탑재한 수집서버로 시스템을 구성하고 있다(최원태 2006, 50).

웹 정보자원의 납본에 대한 법·제도가 아직 마련되지 않은 우리나라는 저작권자의 동의하에 선별적 정보에 대한 수집 및 보존이 이루어지고 있다. 웹 정보자원의 납본을 법제화한 대표적인 예로는 덴마크와 프랑스를 들 수 있다. 덴마크는 1997년 납본법을 개정하여 동적(dynamic) 웹을 제외한 자국의 모든 정적(static) 웹 정보자원을 납본 대상에 포함시켰고(Henriksen, 2001), 프랑스는 2006년 제정된 DADVSI법<sup>23)</sup>을 통해, 기호, 이미지, 사운드, 메시지를 포함한, 공적 인터넷 공간에서 소통되는 모든 유형의 온라인 전

21) 선정지침은 국내외적으로 권위있고 규모가 큰 회의로부터 발간되는 자료와 정부기관, 전문 협회나 기관, 대학교에서 개최되는 회의자료를 우선적으로 수집한다고 밝히고 있다. 다만 대학학과와 같이 소규모 집단에서 개최되는 회의자료는 수집대상에서 제외되며, 국제회의의 기준은 전체참가인원 300명 이상, 40% 이상이 외국인 참가자, 5개국 이상의 참가, 3일 이상의 회의기간을 기준으로 하였다.

22) LAN 전용선이 카메라에 직접 연결되어 실시간 이미지를 인터넷에 전송하여 세계 어느 곳에서든지 볼 수 있도록 한 사이트.

23) Droit d'auteur et droits voisins dans la société de l'information- loi 2006-961.

자 간행물을 납본의 대상으로 하는 가장 포괄적인 납본제도를 실시하고 있다(BnF 2006).

웹 정보자원의 보존 방법으로서 선택적 아카이빙과 포괄적 아카이빙은 어느 한 쪽이 월등히 우월하다고 할 수 없는 각각의 장단점을 가지고 있다. 로봇의 자동 수집(automatic harvesting)에 기반 한 포괄적 아카이빙은 상대적으로 수집 비용이 저렴하다는 점이 있을 뿐만 아니라, 가치 유동적인 정보자원의 속성에 능동적으로 대처할 수 있다는 장점이 있다. 반면, 역동적 데이터베이스, 플러그-인, 스크립트(script) 기술을 적용하는 웹사이트를 다루는 데 적절치 못하다는 단점이 있다. 다시 말해, 자동 수집 접근법은 공개적으로 이용가능한 표면 웹을 다루는 데 장점이 있음에도 불구하고 심층 웹을 다루는 데는 선택적 접근법이 좀 더 현실적일 수 있다는 것이다. 이런 맥락에서 포괄적 아카이빙은 상대적으로 도메인의 수가 적고 구분이 편한 경우, 또 정적이고 표준적인 HTML을 사용한 웹을 다룰 경우에 큰 장점이 있다는 평가를 받고 있다. 한편, 선택적 아카이빙은 정보가치의 변화를 빠르게 반영하지 못하여 중요한 정

보자원을 놓칠 수 있다는 점과 노동집약적으로 운영비용이 높다는 점이 단점으로 지적되고 있다. 따라서 웹 아카이빙에서 대상의 범위, 방법 및 절차의 선택은 어느 한 방법의 기술적 우위나 편이성에 근거한다기보다는 웹을 둘러싼 각국의 기술적, 사회적 환경과 정책적 판단에 좌우된다고 할 수 있다.<sup>24)</sup>

### 3.3 진본성

기록물관리의 세계표준인 ISO 15489<sup>25)</sup>는 기록물이 가치를 가지기 위해 진본성(authenticity), 신뢰성(reliability), 무결성(integrity), 가용성(usability)이라는 특성을 충족시켜야 한다고 명시하였다(ISO 2001). 진본성은 기록물 형상의 유래가 적법함을 의미하고 무결성은 그 형상이 변조되지 않았음을 뜻한다. 또한 가용성은 기록물의 정보를 실제 사용할 수 있어야 함을, 신뢰성은 기록물의 정보가 사용하기에 충분한 내용을 담고 있어야 함을 지적하고 있다(송병호 2005, 45). 이중 진본성은 기록물의 증거능력을 대표하는 것으로 기록관리의 핵심적

24) 각국 웹 아카이빙 기관과 수집방법의 예는 다음과 같다.

국 가	기 관	수집방법
오스트레일리아	PANDORA(NLA)	선택적 수집
오스트리아	AOLA(ONB/TU Wien)	자동 하베스팅
핀란드	Helsinki University Library	자동 하베스팅
프랑스	Bibliothèque nationale de France	자동 하베스팅 & 선택적 수집
스웨덴	Kulturarw(KB)	자동 하베스팅
영국	Britain on the Web(BL)	선택적 수집
미국	Internet Archive	자동 하베스팅
미국	MINERVA(LoC)	선택적 수집

출처: Day, Michael(2003). Collecting and preserving the World Wide Web. UKOLN, University of Bath. 18쪽.

25) 2001년 공표된 ISO 15489는 자료관리 및 자료관리 시스템 개발을 위한 표준으로서, 자료 관리정책과 절차를 표준화하고 모든 자료를 효율적으로 보호하고 검색할 수 있도록 함을 목적으로 한다. ISO 15489는 1부에서 기록관리의 일반원칙과 요건을 제시하고, 2부에서는 1부에서 기술한 일반적인 기록관리 원칙과 요건을 실행하기 위한 세부사항을 제안하고 있다(ISO, 2001).

개념으로 강조되어 왔다.

하지만 웹 정보자원은 진본성이라는 측면에서 태생적으로 매우 취약한 한계를 가지고 있다. 변조와 복제가 상대적으로 용이하고, 기술 의존도가 높은 웹 아카이빙의 영역에서 진본성을 검증하고 확보하는 데는 많은 어려움이 따를 뿐더러, 누구나 손쉽게 접근할 수 있는 오픈 플랫폼(open platform)에 기반하고 있는 웹 정보자원은 그 속성상 익명성을 지닌 불특정 다수에 의해 노출되고, 훼손될 여지가 높다. 또한 웹과 같은 디지털 정보자원을 담고 있는 저장 매체의 불안정성과 짧은 수명으로 인해 마이그레이션 혹은 에뮬레이션(emulation) 등을 통한 보존 전략이 필수적이라는 점에서 진본성의 유지는 더욱 어려운 과제로 남는다. 그러나 이러한 진본성 확보의 취약성은 진본성 확보의 어려움을 의미할 뿐 진본성 가치의 상실을 뜻하는 것은 결코 아니다.

유형의 형태를 가진 기록물의 경우, 유일본의 개념이 중요시되었지만, 무수한 복제물이 존재할 수 있는 웹 정보자원에서 진본성의 개념은 오히려 더욱 강조되고 중요시되고 있다(김명훈 2005). 기록의 내용과 내용을 담는 매체가 일체화되어 있는 종이 기록물을 비롯한 아날로그 기록물과는 달리, 디지털 기록에서는 기록을 담아내는 구조와 내용정보가 분리되어 존재한다. 비록 내용정보는 그대로 유지된다

하더라도 기록을 담아내는 구조는 지속적인 변화의 대상이 된다. 디지털 정보자원의 이러한 변화는 정보자원의 끊임없는 재생산을 의미하며, 변화, 재생산에 의해 생성된 사본이 변화, 재생산 이전의 디지털 정보자원의 내용과 동일하도록 유지하는 것, 다시 말해 원본과 동일한 진본 사본(authentic copies)을 유지하는 것이 웹을 포함한 디지털 기록 관리의 핵심과제라 할 수 있다(임진희 2006a; 2006b). 기록의 형태와는 무관하게, 진본성을 가진 기록물만이 영구보존할 당위성을 가지며, 연구 자료로서의 가치를 갖는다는 점에서 기록물의 진본성을 확보하는 것은 아키비스트에게 매우 중요한 의의를 가지기 때문이다.

우리나라에서 디지털 정보자원의 보존에 적극 나서고 있는 대표적 기관으로는 국가기록원과 국립중앙도서관을 들 수 있다. 국가기록원은 정부가 보관하고 있는 아날로그 기록물들을 디지털화하고 장기보존하기 위해 ISO 15489,<sup>26)</sup> ISO 23081,<sup>27)</sup> OAIS 참조모형(Reference Model for an Open Archival Information System)에 따라 보관 시스템을 구축하고 있으며, 국립도서관은 웹 아카이빙 프로젝트 OASIS를 운영 중에 있다.

하지만 웹 정보자원의 진본성을 보장하고 유지하도록 하는 방안에 대한 법·제도는 매우 미비한 실정이다. 현행법상 디지털 정보자원의

26) 2001년 공표된 ISO 15489는 자료관리 및 자료관리 시스템 개발을 위한 표준으로서, 자료 관리정책과 절차를 표준화하고 모든 자료를 효율적으로 보호하고 검색할 수 있도록 함을 목적으로 한다. ISO 15489는 1부에서 기록관리의 일반원칙과 요건을 제시하고, 2부에서는 1부에서 기술한 일반적인 기록관리 원칙과 요건을 실행하기 위한 세부사항을 제안하고 있다(ISO, 2001).

27) ISO 23081은 기록관리 메타데이터를 생산, 관리, 활용하는 틀을 제시하고, 메타데이터를 통제하는 원칙을 설명하고 있다. 메타데이터 관리에 대한 지침을 직접 제시하고 있지 않고 있는 국제표준인 ISO 15489를 보완하는 목적으로 가지고 있다. 현재 제1부만 출판되었고 제2부와 제3부는 2008년경 출판될 예정이다(최삼미, 이상용, 2007, 41-42).

진본성과 무결성을 보장하는 장치는 공인전자 문서보관서 제도로서, 2005년 3월에 개정된 『전자거래기본법』<sup>28)</sup>에 그 법적 근거를 두고 있다. 공인전자문서보관소의 지정과 기술 표준은 한국전자거래진흥원이 담당하고 있다.

#### 4. 『공공기록물 관리에 관한 법률』과 웹 아카이빙

2007년 4월부터 시행된 『공공기록물 관리에 관한 법률』은 종이문서 기록물에서 전자기록물로 중심이 이동하고 있는 기록물관리의 패러다임을 반영하고 있다는 점에서 큰 의의가 있다. 이 법의 제6조는 “공공기관 및 기록물관리기관의 장은 기록물이 전자적으로 생산·관리되도록 하여야 하며, 전자적으로 생산되지 않은 기록물에 대하여도 전자적으로 관리되도록 필요한 조치를 강구”하여야 한다고 명시하고 있으며, 제5조는 전자기록물의 구체적인 관리지침으로써 기록물의 생산에서부터 활용까지의 전 과정에 걸친 “진본성, 무결성, 신뢰성 및 이용가능성”의 보장을 강조하고 있다. 또한 모법과 함께 시행령과 시행규칙을 통해 전자기록물의 원칙과 관리방식(법 제5조, 6조, 시행령 제5조, 제39조, 제50조), 등록(시행령 제21조, 제24조, 제28조), 이관(시행령 제26조, 제37조, 제47조), 보존(시행령 제38조)에 이르는 전 과정에 대한 법적 근거를 제공하고 있다.

하지만 안타깝게도 웹 아카이빙에 대한 구체적인 언급은 찾아볼 수가 없다. 물론 전자기록

물이라는 큰 범주 안에 웹 기반 기록물이 자리하고 있다. 그러나 앞서 살펴보았듯이 오프라인의 전자기록물과는 현저하게 다른 특성을 가지고 있는 웹 기반 기록물관리를 고려한 법 조항의 미비는 크게 아쉬운 점이라고 할 수 있다. 특히 우리 사회에서 웹이 차지하는 정치사회적, 문화적 영향력을 고려할 때 공적 가치를 가지는 웹 기반 기록물관리에 대한 법제도 마련은 시급하다 할 수 있다. 빠르게 변화하는 웹 환경으로 인해, 모법에 구체적인 내용의 규정이 어렵다 한다면, 적어도 시행령 혹은 규칙에서는 웹 아카이빙 활동의 법적 근거와 원칙 그리고 책임 있는 주체의 명시를 고려해보아야 할 것이다.

#### 5. 웹 아카이빙 기관의 과제: OASIS를 중심으로

이상에서 살펴보았듯이, 웹 정보자원의 보존은 아직도 숙한 미완의 과제를 안고 있다. 웹의 태생적 특성에 따른 정보자원의 취약성으로부터, 수집의 기술적 문제들, 그리고 저작권의 문제에 이르기까지 방대한 영역의 문제들을 모두 포괄하는 해법은 현존하지 않는다.

호주 국립 아카이브는 효율적 웹 아카이빙을 위해 고려해야 할 네 가지 요소로 사이트의 목적, 사이트의 복합성, 보존기관의 공적 투명성, 정보자원의 변화빈도와 일정성을 들었다(National Archives of Australia 2001)(그림 1 참조).

첫째, 웹 사이트의 목적 파악이다. 웹 사이트

28) 일부개정 2007.5.17. 법률 제8466호

의 목적은 비상업적 정보의 소통에서부터, 상품의 광고·판매·서비스를 비롯하여 정치·사회·문화적 사건에 대한 공적 기록의 제공에 이르기까지 그야말로 광범위하다. 이러한 웹 사이트의 목적은 정보자원을 생산한 주체와 내용의 공적 성격에 크게 좌우된다. 사이트의 목적을 파악하여 공공성과 보존가치, 그리고 예상되는 법적 문제를 판단하는 일은 보존기관 본연의 가장 핵심적인 역할이라 할 수 있다.

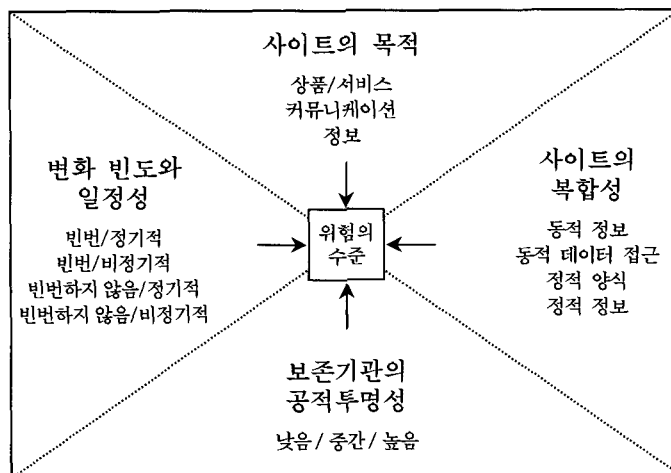
둘째, 사이트의 기술적 분석이다. 사이트가 동적인지 정적인지, 상호작용성의 수준은 어느 정도인지를 분석하여 효과적인 방법론을 찾는 노력을 기울여야 한다. 특히 동적 페이지 혹은 데이터를 포함하고 있는 사이트의 높은 상호작용성을 효과적으로 보존하는 방안에 대한 지속적인 연구가 요구된다.

셋째, 정보자원의 변화에 대한 효율적 대처다. 변화는 정기적으로 혹은 부정기적으로 일어날 수 있다. 매일, 매주, 혹은 매월 단위로 자료

가 업데이트될 수 있고, 모든 정보가 새것으로 교체될 수도 있다. 심지어, 데이터베이스 기반의 사이트인 경우, 변화는 사용자 요구에 따라 실시간으로 일어난다. 따라서 내용 변화가 얼마나 빈번한지 정기적으로 업데이트가 되고 있는지의 여부를 파악하여 수집의 방법과 주기를 설정하는 것은 매우 중요하다. 이는 웹 정보자원 아카이브의 신뢰성과 유지성에 직결되어 있다.

넷째, 보존기관의 공적 투명성 확보다. 공적 가치를 가지는 정보자원을 보존하며, 일반 대중들을 대상으로 공적 서비스를 제공하는 임무를 담당하고 있는 웹 아카이빙 기관에서 실질적 수집·보존 기능과 관련된 기술적, 제도적 방법에 대한 고민만큼이나 중요한 것이 웹 아카이브 기관의 공정성, 투명성을 확보하고자 하는 노력이다.

이러한 맥락에서 웹 아카이브를 담당하는 기관들은 본연의 수집·보존 기능뿐만 아니라 미래를 위한 연구 기능 및 일반 이용자를 위한 서



〈그림 1〉 웹 아카이브에서 고려해야 할 요소  
(National Archives of Australia, 2001, 23)

비스 기능을 함께 수행할 것을 요구받고 있다. 여기서 연구 기능은 크게 웹 아카이빙 기술의 개선과 개발을 중심으로 한 기술 분야의 연구와 웹 아카이빙과 관련된 법·제도에 대한 연구로 구분된다. 한편 서비스 기능이라 함은 일반 이용자가 웹 아카이브를 교육·연구 혹은 기타의 목적으로 활용할 수 있도록 기관이 적극 지원하는 것을 의미한다. 디지털 아카이브의 목적 중 하나가 저장된 모든 정보자원에 대한 일반 사용자들의 접근성을 최대한 보장함으로써 디지털 격차(digital divide)를 해소하고 이를 통해 전자민주주의를 실현하는 데 있다(민경배, 이한배 2004)는 측면에서 웹 아카이브 서비스의 중요성은 거듭 강조되고 있다. 이러한 서비스 기능 강화를 위해서는 일반 사용자들의 웹 아카이브로의 접근성 개선과 사용자 리터러시(literacy)를 강화하는 제도적 장치의 마련이 필요하다.

국가적 차원의 온라인 디지털 유산 수집·보존과 교육·연구 자료로의 활용을 목적으로 하는 우리나라의 대표적 웹 아카이브는 국립중앙도서관에 의해 주도되고 있다.<sup>29)</sup> 국립중앙도서관에 의해 2004년 시작된 OASIS 프로젝트는 2005년과 2006년 각각 5억6백만 원, 4억6천만 원의 예산이 투입되었고, 2006년 2월 홈페이지를 연 이래 대국민 서비스를 실시하고 있다(국립중앙도서관, 2006b). 선택적 수집방법을 채택하고 있는 OASIS는 수집 대상을 개별자원으로서의 “웹문서”와 하나의 완결된 사이트

단위로 해당 사이트의 모든 자원을 수집하는 “웹사이트”로 구분하고 있다. 수집된 정보자원은 KDC 분류체계에 따라 분류되고, 더블린코어(Dublin Core) 메타데이터의 15개 요소를 기준으로 기술된다. 이러한 수집 보존활동은 사전협의를 통한 저작권자의 동의를 전제로 하고 있다.

OASIS는 세계 최고의 인터넷 인프라와 이용률을 자랑하는 우리나라가 첫발을 내딛는 국가적 웹 아카이빙이라는 점에서 큰 의의를 지닌다. 하지만 OASIS에 대한 비판의 여지가 없는 것은 아니다. 2008년 국립디지털도서관 사업의 완성이라는 전향적 발전의 계기를 목전에 두고 있는 OASIS가 당면한 과제로는 크게 두 가지를 들 수 있다.

첫째, 수집방법에 대한 고민이다. 현재 OASIS가 채택하고 있는 선택적 수집방법은 웹 아카이빙에 따른 법적 문제 등과 관련된 위험을 현저하게 감소시킨다는 장점이 있다. 그러나 웹의 활용이 일상 그 자체가 되어 있고 웹 환경의 변화 속도가 매우 빠른 우리의 현실을 감안할 때, 이러한 수집 방법의 타당성에 대한 연구와 논의가 필요할 것으로 판단된다. 프랑스와 같은 전면적이고 포괄적인 수집은 어렵다 하더라도 현재의 수동적인 선택적 수집보다는 대상과 방법에서 좀 더 포괄적이고 유연한 방안의 적용가능성이 모색되어야 한다. 이를 위해, 디지털 납본 제도와 같은 법·제도 개선 또한 적극적으로 고려되어야 할 것이다.

29) 『도서관법』 제2조는 도서관이 다루고 제공하는 “도서관자료”를 “도서관이 수집·정리·보존하는 자료로서 인쇄자료, 필사자료, 시청각자료, 마이크로형태자료, 전자자료 그 밖에 장애인을 위한 특수자료 등 지식정보자원 전달을 목적으로 정보가 축적된 모든 매체”로 규정함으로써 국립중앙도서관이 국가적 웹 아카이빙의 주체로 나설 수 있는 법적 근거를 제공하고 있다.



둘째, 연구 및 서비스 기능의 강화 문제다. 일차적 목적인 수집·보존에 비해 연구 활동의 영역과 다양한 서비스 및 활용방안에 대한 계획과 비전이 상대적으로 미흡한 점은 시급히 시정되어야 할 것이다. 예를 들어, 2007년 7월 현재, OASIS는 국가자료공동목록시스템 KOLIS (Korean Library Information System Network)와 연계되어 있지 않은 상태로, 향후 연계 서비스에 대한 구체적인 계획도 수립되어 있지 않은 상태이다. 연구 및 서비스 기능의 강화를 위해서는 웹 아카이빙 사업 예산의 대폭 확대와 함께 정부 및 민간 조직을 망라한 타 기관과의 적극적인 협력 네트워크 구축을 강화해야 할 것으로 보인다.

이 두 가지는 당연한 과제이면서도, 결코 단기간에 해결될 성질의 것이 아니라, 장기적 관점에서 지속적으로 살펴야 할 웹 아카이빙 기관의 과제라 할 수 있다. 이러한 맥락에서 OASIS의 발전을 위한 대안들은 단기적 처방보다는 근본적인 인식과 제도의 개선을 통한 항구적인 웹 아카이빙 환경의 개선에 주안점을 두어야 할 것으로 판단된다.

## 6. 나가는 글

웹은 지금 이 순간도 매우 빠른 속도로 성장하고 있는 거대한 정보자원이다. 기록 보존이라는 측면에서 웹은 “움직이는 표적(moving target)”이라 해도 과언이 아닐 것이다. 살아있는 생물처럼 끝없는 변화와 발전을 거듭하는 웹 정보자원의 속성은 기록의 보존이라는 측면에서 무척 곤혹스러운 것이다.

“모든 정보가 진흙 탁자에 써졌거나 대리석 위에 새겨졌더라면 그것들의 보존은 훨씬 쉬웠을 것이다”라는 허틀(Hirtle 2003)의 언명은, 21세기 첨단 기술의 집약이라 일컬어지는 디지털 정보자원의 취약성을 역설적으로 이야기하고 있다. 정성스레 만들어지고 보존된 종이는 수백 년의 세월을 견디어낸다. 반면, 오늘날의 디지털 정보의 수명은 십 수 년을 넘지 못한다. 불과 몇 년 전까지 우리의 소중한 정보를 담아 주었던 플로피 디스크를 생각해보자. 지금 책상 서랍 속 어디인가의 플로피 디스크에 담긴 마그네틱 데이터가 아직 유효하다고 해도, 정작 디스크를 읽어낼 수 있는 드라이버는 점점 귀해지고 있다. 무어의 법칙(Moore's Law)은 여전히 유효하며, 네트워크의 환경과 기술 또한 하루가 다르게 발전하고 있다. 우리가 웹을 포함한 디지털 정보자원의 보존에 적극적으로 대처하지 않는다면 ‘디지털 암흑시대’의 경고(Kuny 1998)는 머지않은 현실이 될 수도 있다. 다행스러운 일은 앞서 살펴보았듯이 웹 아카이빙에 대한 많은 전문가들의 노력들이 나름의 결실들로 나타나고 있다는 것이다. 하지만 이것은 결실의 시작일 뿐이다. 웹 아카이빙의 더 큰 결실로 가는 길에는 수집·보존 방법의 개선, 연구 및 서비스 기능의 강화, 아카이빙 기관의 공공성과 투명성 제고 등, 산적한 과제가 놓여있다.

기록이 없으면 역사도 없다. 올바른 기록 없이 민주주의의 발전을 기대할 수 없다(곽건홍 2003). 21세기 가장 대중적인 정보매체로 등장한 웹의 올바른 보존은 우리의 역사와 민주주의 지키는 일에 다름 아니다. 이런 맥락에서, 이제 웹 아카이빙에 관한 논의들은 당면과제에

대한 기술적이고 방법론적인 대안 모색을 위한 연구 성과를 바탕으로 점차 그 외연을 넓혀 나가야 할 것이다. 디지털 정보자원의 보존은 통합과 수렴을 위한 노력을 전제로 한다(이소연 2004, 55). 웹 아카이빙이라는 화두를 중심으로

한 학제간 연구와 관련 주체간의 협력이 절실히 요구되는 시점이다. 이러한 거시적 관점을 기반으로 웹 아카이빙의 제반 문제에 대한 논의들이 더욱 활발히 전개되었으면 하는 바람이다.

### 참 고 문 헌

곽건홍. 2003. 『한국 국가기록 관리의 이론과 실제』. 서울: 역사비평사.

국립중앙도서관(2006a). 온라인 디지털자원 선정 지침 [cited 2007.7.15]  
<http://www.oasis.go.kr>

국립중앙도서관(2006b). 2006년도 OASIS 자체평가 보고서 [cited 2007.7.15]  
<http://www.oasis.go.kr>

김명훈. 2005. 전자기록의 평가: 기록의 속성과 관련하여. 『제3회 한국기록학회 월례발표회』. 한국기록학회.

김익한. 2006. 전자기록의 진본평가 시스템 모형 연구. 『기록학연구』, 14: 91-117.

민경배, 이한메. 2004. 디지털유산보존국내현황 및 협력틀 구축방안. 디지털유산보존에 관한 기초연구. 국립중앙도서관.

서혜란 2003. 디지털자료의 납본과 보존을 위한 각 국가의 노력. 『정보관리학회지』, 20(1): 373-399.

서혜란. 2004. 웹 아카이빙의 성과와 과제. 『한국비블리아학회지』, 15(1): 5-22

송병호 2005. 진본성 확보를 위한 전자기록물 관리 방안. 『한국비블리아학회지』, 16(2): 43-59.

압둘아지즈 아비드. 2005. 디지털 유산의 보존: UNESCO의 관점에서. 『한국기록관리학회지』, 5(2): 183-193.

우학명. 2006. 국회도서관의 전자도서관과 웹 아카이빙: 현재와 미래. 『국회도서관보』, 328: 21-30.

윤대현, 남성운.(2005). 전자기록물 관리체계에 대한 제언: 전자기록물 개념과 국제표준 사례 등을 중심으로. 『기록관리보존』, 10: 1-32.

윤희윤. 2003. 한국 납본제도의 문제점과 개선방안 『한국문헌정보학회 학술발표논문집』, 16: 27-45.

이소연 2004. 디지털유산의 장기적 보존: 국가정책수립을 위한 제언. 『기록학연구』, 10: 27-64.

임진희. 2006a. 『OAIS정보모델과 기록AIP』. 한국국가기록연구원.

\_\_\_\_\_. 2006b. 전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조. 『기록학연구』, 13: 41-90.

최경섭. 2006. 하나로 · KT, 100Mbps VDSL

- 상용화 '눈앞.' 『디지털타임스』
- 최상미, 이상용. 2007. 현용전자기록물의 메타데이터 요소에 관한 연구. 『한국기록관리학회지』, 7(1): 39-60.
- 최원태. 2006. 『OASIS 중장기 정보전략계획 세부과제 도출』. 국립중앙도서관.
- 한국인터넷진흥원(2007). 『2007 한국인터넷 백서』.
- Baran, Paul. 1964. Distributed communications: I. Introduction to distributed communications networks. [cited 2004. 6. 22] <<http://www.rand.org/publications/RM/RM3420/>>
- Bergman, Michael K. 2001. The Deep Web: Surfacing Hidden Value. *The Journal of Electronic Publishing*, 7(1).
- BnF. 2006. Web Archiving at BnF. [cited 2007.7.20] <[http://www.bnf.fr/pages/version\\_anglaise/depotleg/pdf/BnFnews200609.pdf](http://www.bnf.fr/pages/version_anglaise/depotleg/pdf/BnFnews200609.pdf)>
- Charlesworth, Andrew. 2003a. Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia.
- \_\_\_\_\_. 2003b. Law and the Would-be Web Archivist. *Preserving the Web*. University of Kerkira, Kerkira, Greece.
- Day, Michael. 2003. Collecting and preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust. UKOLN, University of Bath.
- Electronic Publication Pilot Project team and Electronic Collections Committee. 1996. Electronic Publication Pilot Project (EPPP) Final Report. [cited 2007.7.20] <<http://dsp-psd.pwgsc.gc.ca/Collection/SN3-331-1997E.pdf>>
- European Commission. 1996. Deposit collections of electronic publications: Libraries in the information society. European Communities.
- Henriksen, Birgit. 2001. Legal Deposit from the Internet in Denmark: Experiences with the Law from 1997 and the Need for Adjustments. *Preserving the Present for the Future: Strategies for the Internet conference*. The Royal Library, Copenhagen 18th -19th of June 2001. [cited 2007.7.25] <[http://www.deflink.dk/upload/doc\\_filer/doc\\_alle/1023\\_BNH.doc](http://www.deflink.dk/upload/doc_filer/doc_alle/1023_BNH.doc)>
- Hirtle, Peter B. 2003. Digital preservation and copyright. Stanford University Libraries. [cited 2007.7.20]. <[http://fairuse.stanford.edu/commentary\\_and\\_analysis/2003\\_11\\_hirtle.html](http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html)>
- ISO. 2001. ISO 15489L 2001(E) Information and documentation - Records management.
- Kuny, Terry. 1997. A Digital Dark Ages? Challenges in the Preservation of Electronic Information. *63<sup>rd</sup> IFLA Council and General Conference*. [cited 2007. 7. 25]

- <<http://www.ifla.org/IV/ifla63/63kunya1.pdf>>
- Larivière, Jules. 2000. Guidelines for Legal Deposit Legislation. Paris: UNESCO.
- Lawrence, S., Pennock, D. M., Flake, G. W., Krovetz, R., Coetzee, F. M., E., Nielsen, F.Å, Kruger, A. & Giles, C.L. 2001. Persistence of Web references in scientific research. *Computer*, 34(2), February, 26-31.
- Levy, Pierre. 2002. 『집단지성: 사이버공간의 인류학을 위하여』. 권수경 역, 서울: 문학과 지성사.
- Lyman, Peter and Hal R. Varian. 2003. How Much Information, 2003. [cited 2007. 7. 20]  
<<http://www.sims.berkeley.edu/how-much-info-2003>>
- National Archives of Australia. 2001. Archiving Web Resource: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government. [cited 2007. 7. 30]  
<[http://www.naa.gov.au/recordkeeping/er/web\\_records/guide\\_contents.html](http://www.naa.gov.au/recordkeeping/er/web_records/guide_contents.html)>
- Negroponte, N. 1996. *Being Digital*. London: Coronet.
- Poster, Mark. 1997. Cyberdemocracy: Internet and the public sphere. In Porter, David (Ed.) *Internet culture*(201-217). New York: Routledge.
- Stanescu, Andreas. 2004. "Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology." *D-Lib Magazine*, 10(11). [cited 2007.7.25]  
<<http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>>