

OASIS의 선정지침 개선(안)에 관한 연구

A Study on Improving the OASIS Selection Guidelines

노 영 희(Younghee Noh)*

고 영 선(Youngsun Go)**

초 록

웹상에서 유통되는 정보자원의 역사적, 사회적, 문화적 가치는 전 세계적으로 인정받고 있으며, 인류 유산으로 후세에게 물려주기 위한 웹 아카이빙 사업이 수행되고 있다. 이 사업에서 가장 기본이 되는 것은 수많은 웹 자원 중 수집 및 보존의 가치가 있는 웹 자원을 선정하기 위한 선정지침이라 할 수 있다. 본 연구에서는 선행연구 및 국내외 웹 아카이빙 선정지침을 집중적으로 분석하여 OASIS를 위한 선정지침의 개선안을 제안하였다. 첫째, 웹 자료에 대한 정의, 용어정의, 수집기본원칙, 수집방법, 수집주기, 구체적인 아카이빙 대상자원 선정에 대해서 개선 내용을 제안하였다. 둘째, 수집대상자료에 대해서 대폭적으로 개선 내용을 제안하였으며, 웹 아카이빙 선정제외자료에 대한 제안도 하였다. 마지막으로, 수집방법과 온라인 자료 납본과의 관련성, 선정대상자료에 대한 목록데이터베이스 구축의 필요성, 협력형 아카이빙 정책의 필요성에 대해서 논의하였다.

ABSTRACT

The historical, social, and cultural value of Internet resources is indisputable, therefore many national institutions have created Web archiving projects to hand down this heritage to future generations. The selection guidelines are the most crucial aspect of these projects because they aid in differentiating which resources are worth collecting and preserving from the large number of web resources available. The purpose of this study was to suggest improvements for the OASIS Selection Guidelines by analyzing the selection guidelines of other domestic and international Web archiving projects. First, based on the results of Web archiving projects abroad, we proposed improvements for the definition of Web data and other terms, the basic principles of the collection, collection methods, and collection cycle. Second, we proposed substantial improvements in the target resources for archiving, and also stated what kind of web resources must be excluded from Web archiving. Finally, we discussed the relationship between data collection methods and the legal deposit of online resources, the necessity of constructing a database for selected target materials, and the necessity of cooperative archiving policies.

키워드: 웹 아카이빙, 온라인자료, 웹 자원, 선정지침, 오아시스

Web Archiving, Online Materials, Web Resources, Selection Guideline, OASIS

* 건국대학교 문헌정보학과 부교수(irs4u@kku.ac.kr)

** 건국대학교 지식콘텐츠연구소 선임연구원(forgolib@hanmail.net)

논문접수일자 : 2012년 8월 17일 논문심사일자 : 2012년 9월 5일 게재확정일자 : 2012년 9월 19일

1. 서론

1.1 연구의 필요성 및 목적

웹상에서 유통되는 정보자원의 보존가치에 대해서는 많은 연구자 및 이용자가 인정하고 있는 바이다. 웹상에서 유통되는 정보자원의 양이 급증하고 있고, 오로지 웹으로만 출판되는 정보자원도 많으며, 이러한 정보자원에 대한 의존도가 매우 높은 것이 현실이다. 이들 자원 중에는 일시적인 가치만 있는 것뿐만 아니라 역사적 가치, 문화적 가치, 학술적 가치 등이 높아서 수집, 보존하여 인류 유산으로 후세에게 물려주어야 할 자원들이 늘어나고 있다. 이에 국가적인 차원에서 웹 아카이빙 프로젝트를 세계 여러 나라가 수행하고 있으며, 국내의 경우 국립중앙도서관을 포함한 국가기록원, 한국과학기술연구원 등이 이러한 업무를 수행하고 있다.

이 중 국립중앙도서관은 가치 있는 온라인 디지털자원의 체계적인 수집·보존 정책을 연구하고 실행함으로써 국가 대표도서관으로서의 역할을 강화하고자 하였다. 이에 국립중앙도서관에서는 국가적으로 수집해야 할 디지털 자원의 범위를 제안하고 정부기관이나 교육기관에서 생산되는 자료와 전문가 집단에게 추천 받은 웹 사이트를 포함하는 보다 정확하면서도 폭넓은 수집 정책을 설계하였다. 그 결과 개발된 OASIS(Online Archiving & Searching Internet Sources)는 저작권 소유기관 및 개인으로부터 이용허락을 받아 아카이빙을 진행하며, 캡처가 아닌 원본 웹 사이트의 서버 구조를 파악하고 완전히 동일하게 복제하는 방법인 사

이트 미러링(mirroring)을 통해 아카이빙을 진행한다.

그러나 역사적, 문화적, 학술적, 사회적 가치가 있는 웹 자원을 선정하여 아카이빙하기 위해서는 가치가 있는 웹 자원을 잘 선택할 수 있도록 하기 위한 성문화된 웹 자원 장서개발정책과 선정지침의 마련이 우선되어야 한다. 선정지침(Selection Guidelines)이 필요한 이유에 대해 Masanes(2006)는 다음과 같이 다섯 가지로 제시하고 있다. 첫째, 선정지침은 웹 아카이빙을 담당하는 개개인의 선입견을 줄여 주어 기관이 원래 설정한 목표에 맞는 장서개발을 할 수 있도록 해 준다. 둘째, 계획과 실제 컬렉션에서의 개별성 차이를 줄여주며 선별 과정과 리뷰 및 필터링 과정에서 계속성과 지속성을 보증한다. 셋째, 우선순위를 결정하고, 목적과 범위를 명확히 하는데 도움을 준다. 넷째, 평가된 웹 사이트에 대한 선별을 결정하며, 범위에 포함되지 않는 기록물의 취득에 대해 적절하게 대처할 수 있다. 다섯째, 다른 기관과 자원 공유를 위한 틀을 제공한다.

이에 웹 아카이빙을 위한 선정지침 개발과 관련된 연구가 다수 수행되었으며, 지금까지 수행된 이와 관련된 연구로는 2007년에 수행된 이재선(2008)의 연구, Fritcher와 Cromwell(2001)의 연구 등이 있다. 이들은 평가기준 요소로 저작성, 내용의 정확성 및 객관성, 포괄성, 희귀성 및 유일성, 최신성, 저작자의 평판 등을 제시하고 있으며, 평가기준 요소에 있어서는 연구자마다 차이를 보이고 있다. 이와 같이 웹 자원을 평가하고 아카이빙을 위한 평가기준을 제시한 연구는 있으나 Masanes가 그 필요성을 주장했던 선정지침 개발 및 개선과 관련된 연구는 거

의 없었다.

이에 본 연구에서는 아카이빙 대상 자원을 선정할 수 있는 현장 적용성이 높은 선정지침을 제안하고자 하며, 온라인자료 선정지침(안)을 제안함으로써 이를 실제로 OASIS 자원을 개발하는데 활용될 수 있도록 하고자 한다.

1.2 연구방법

OASIS는 수집, 보존, 서비스하는 웹 아카이빙의 대상을 개별자원과 웹 사이트로 보고 있으며, 개별자원은 하나 또는 그 이상의 파일단위로 정보 가치를 가지는 자원이고, 웹 사이트는 하나의 단위로 해당 사이트의 모든 자원을 미러링하게 하는 자원이다(윤정옥 2010).

그러나 웹상에서 유통되는 정보자원의 양은 엄청날 뿐만 아니라 갱신되는 속도도 점점 빨라지고 있기 때문에 웹상의 모든 정보자원을 아카이빙 하는 것은 사실상 불가능할 뿐만 아니라 현재의 기술로는 가능하지도 않을 것이다. 따라서 아주 엄선된 정보자원을 수집할 필요가 있으며, 엄격하면서도 실효성 있는 웹 아카이빙을 위한 선정지침은 다른 모든 작업들 보다 우선한다고 할 수 있다.

그럼에도 불구하고 OASIS를 위한 선정지침이 유명무실하거나 실제로 적용하고자 할 때 혼선이 있으며, 결과적으로 자주 바뀌는 오직 한 명의 수집담당자의 그때그때의 판단, 임의적 전문가 집단의 자문 등에 의존하게 된다. 시스템의 개발도 중요하고 웹 로봇의 성능도 중요하며, 하드웨어의 성능도 중요하지만 지금이 시점에서 가장 중요한 것은 수집대상 자원을 정확하고 빠르게 선정할 수 있는 선정지침

의 개선이라 생각한다.

따라서 본 연구에서는 OASIS 시스템의 웹 아카이빙을 위한 선정지침을 제안함으로써 수집과 보존을 통해 후세에게 유산으로 물려줄 만한 가치가 있는 웹 자원을 아카이빙 할 수 있도록 하고자 한다. 이러한 목적을 달성하기 위해 다음과 같은 연구절차를 거치고자 하였다.

첫째, 웹 아카이빙 선정지침과 관련된 선행 연구를 집중적으로 분석하였으며, 웹 사이트를 평가하는 기준에 관한 연구는 제외하였다.

둘째, 현재의 OASIS 선정지침을 기반으로 자원을 선정할 때 문제가 될 수 있는 부분을 구체적으로 발견하고자 하였다.

셋째, 해외 웹 아카이빙 사이트에서 사용하고 있는 선정지침을 분석하였다.

넷째, OASIS에 적용할 수 있는 구체적인 웹 아카이빙 선정지침의 개선방향을 제안하였으며, 'OASIS 선정지침(안)'을 개발하였다.

다섯째, 이렇게 개발된 선정지침(안)에 대해 현재 OASIS 업무를 담당하고 있는 국립디지털도서관 자료기획과의 검증작업을 거쳤다.

선정지침의 제안은 현재의 선정지침에서 따르고 있는 순서, 즉 추진배경, 웹 자료의 정의, 수집 기본원칙, 선정대상자원 등의 순으로 하였다.

2. 선행 연구

국내의 웹 아카이빙에 대한 연구는 2000년 초반부터 본격적으로 시작된 것을 알 수 있다. 웹 아카이빙에 대한 초창기 연구는 웹 아카이빙의 필요성을 주장하고 선진국의 웹 아카이빙 사례 및 현황을 소개하였고, 이후 웹 아카이빙

의 법적 및 제도적 문제, 디지털 납본의 문제, 국가차원에서 접근 문제 등이 구체적으로 다루어지고 있는 것을 알 수 있다.

먼저, 웹 아카이빙의 필요성을 주장하고 선진 사례 및 현황을 소개한 연구로, 서혜란(2004)은 웹 정보자원의 성격과 웹 아카이빙의 필요성을 언급하고, 웹 아카이빙을 위한 기술적 접근방법에는 일정한 기준에 의해 수집대상 웹 사이트를 선정한 후 수집하는 선택적 접근방법과 로봇 프로그램을 활용해서 자동적으로 수집하는 포괄적 접근방법이 있다고 하였다. 선택적 접근방법을 취한 PANDORA(호주)와 포괄적 접근방법을 취한 Kulturw3(스웨덴), Internet Archive(미국)의 사례를 배경, 수집범위와 방법, 저장과 보존, 이용에 걸쳐 구체적으로 설명하고 있다. 윤정옥(2011)은 OASIS의 자원 수집 지침 현황을 알아보았고, 수집 자료의 현황을 살핀 후, 단순한 스냅샷과 자료의 공개 기준에 대해 의문점을 제기하고 있다. 각 주제별 웹 사이트의 아카이빙 현황 중 '최신 자료'와 '많이 본 자료' 리스트를 분석하여 문제점을 기술하였다. OASIS는 수집 콘텐츠의 주제 분포의 편중, 웹 사이트 저작자/발행자의 권위 및 학술적 가치의 근거 미약, 아카이빙 대상의 정의 부적합, 수집 및 이용 통계 처리의 모호성 등 문제점을 갖고 있다고 기술하였다.

웹 아카이빙의 법·제도적 문제에 대해서 논의한 연구로 서혜란(2003)은 당시 국립디지털도서관 건립 계획을 확정된 국립중앙도서관으로서는 디지털자료의 수집과 보존을 위한 법률적·기술적 기반을 확립하는 일이 무엇보다 시급한 과제 중 하나라고 주장하였다. 한국, 일본, 영국, 호주, 미국 등 14개국의 디지털자료

납본제도 현황을 조사하였고, 웹 아카이빙 프로젝트 현황도 함께 기술하였다. 김유승(2007)은 웹 정보자원의 특성을 분석한 후 저작권, 납본과 수집, 진본성에 대해 웹 아카이빙의 법·제도적 문제와 공적 가치를 가지는 웹 기반 기록물관리에 대한 법·제도 마련이 시급하다고 하였다. 또한 선택적 수집방법을 하고 있는 OASIS는 대상과 방법에서 좀 더 포괄적이고 유연한 방안의 적용 가능성을 모색하여 연구 및 서비스를 강화해야 한다고 언급하였다. 이를 위해, 디지털 납본제도와 같은 법·제도적 개선 또한 적극적으로 고려되어야 한다고 말하였다. 김지훈(2009)은 국가도서관의 포지셔닝에 의거하여 하이브리드 장서개발체계와 함께 온라인 디지털 자료의 아카이빙 방법을 고찰하였다. 국가도서관은 대표도서관의 입장에서 장서개발을 고려해야 하고, 전자잡지와 웹DB, 웹 자원 두 가지로 나누어서 기술하고 있다. 웹 자료의 대표적인 유형은 무료로 제공되는 전자잡지, 회색문헌, 통계데이터, 정책과 기준, 법령과 제도, 전자게시판 등이다. 이를 수집하는 방법은 선택적 아카이빙, 보편적 아카이빙(포괄적 아카이빙), 그리고 이들의 단점을 보완하기 위해 중요하고 권위 있는 주요기관의 웹 사이트에 대해서는 선택적 방법을 적용하고 그 외에는 로봇을 이용하여 정기적으로 체크하여 수집하는 하이브리드 아카이빙 방법을 기술하였고, 이 중 하이브리드 아카이빙 방법이 바람직하다고 언급하였다. 최재황, 곽승진, 김정택(2009)은 국회에 온라인 디지털자료 납본 및 이용에 관한 법률(안)이 발의된 상황에서 세계 주요국의 납본 관련 법률 및 지침을 참조하여 우리나라 온라인 디지털자료 납본요소를 추출하고 온

라인 납본 체계 및 이용에 대한 가이드라인을 제시하였다. 발의한 법안이 덴마크, 뉴질랜드 등의 납본법을 비교하여 납본대상의 종류와 우선순위, 납본 제외대상의 바람직하다고 판단되는 방향을 제시하였다. 오상훈, 최영선(2008)은 웹 아카이빙을 위한 디지털 자원 납본 프로세스를 개발하기 위해 디지털 자원 납본 주체와 대상을 정의하였고, 납본 프로세스를 위한 단계별 정의와 기능을 기술하였다. 디지털납본 시스템 구성을 위한 업무 흐름도와 단위 업무에 따른 기능 정의와 정보 흐름을 위한 구조를 제시하기도 하였다.

웹 아카이빙 정책에 관한 연구로 김유승(2008)은 웹 아카이빙의 특성별 모델을 비교하고, 유형별로 사례를 분석하였다. 유형을 선택적 아카이빙, 복합적 아카이빙, 전체 도메인 아카이빙으로 나누었고, 그 중 복합적 아카이빙인 프랑스 국립도서관의 BnF 사례를 구체적으로 기술하였다. 현재 우리나라의 웹 아카이빙인 OASIS 문제점을 지적하고, 아카이빙 정책의 발전을 위해 복합적 아카이빙을 검토해야 하는 것과 타 기관과의 협력적 웹 아카이빙 전략을 수립해야 한다고 주장하였다.

그 외 웹 아카이빙의 기술적인 측면에서 접근한 연구로 김광영, 이원구, 윤화목, 신성호, 이민호(2011)는 웹 자원 수집 모델을 조사하고, 기존에 사용하고 있는 웹 크롤러의 문제점을 분석하여 가장 효율적인 웹 자원 수집 도구를 위한 대안으로 웹 페이지 사진을 찍는 스냅샷 로봇을 활용한 모델을 제안하였다. 이를 활용하여 실제 웹 사이트들을 대상으로 수집을 한 결과 심층 웹, 플래시, 자바 등의 문제점들을 해결할 수가 있다고 기술하였다. 김희정(2010)

은 미국, 영국, 캐나다, 호주, 싱가포르를 대상으로 국가 웹 아카이빙 사례의 웹 인터페이스를 영역에 따라 평가하였다. 영역은 시각 디자인, 웹 접근성, 보안, 검색, 지원기능 인터페이스로 나누었고, 국가마다 강점을 보이는 영역 인터페이스에는 차이가 있었다. 이를 통해 국내 OASIS에 적용하여 개선할 수 있는 방안을 제안하였다.

웹 아카이빙의 선정 가이드라인 개발과 관련된 연구로 국내에는 이재선(2008)이 있으며, 그는 국가기록원이 웹 사이트 아카이빙을 구축하기 위한 선별 방안을 제안하는 것을 목적으로 국내외 사례분석을 통해 수집 대상과 선별 기준을 살펴보았다. 국내는 OASIS와 민간에서 진행된 정보트러스트 어워드를 살펴보았고, 외국은 영국의 TNA, 미국의 MINERVA, 호주의 PANDORA, 덴마크의 Netarchive.dk를 분석하였다. 이 연구에서는 국립중앙도서관에서 OASIS 프로젝트를 진행하고 있지만, 기록학적 가치로 접근한 것은 아니며 사회적, 역사적인 이슈를 다룬 웹 사이트의 보존은 미미한 실정이라고 기술하였다. 또한 공적인 가치를 가진 공공기관의 웹 사이트는 수집 대상에 포함되지 않아 이롭고, 국가기록원 주체의 웹 사이트 아카이빙 구축이 필요하다고 하였다.

웹 아카이빙과 관련된 해외 연구는 국내의 많은 연구자에 의해 이미 언급되어 있으며, 그 중에서 본 연구와 관련이 있는 웹 아카이빙 선정 기준과 관련된 연구를 살펴보면, 먼저 Masanes(2006)는 웹 아카이빙을 위한 선별 단계가 선정, 캡처, 아카이빙, 접근, 질적 리뷰의 5가지 단계로 구분되어 있다고 언급하였다. 이 중 아카이빙을 한 선정은 준비, 발견, 필터링의 단계를

포함하고, 그 과정은 대상, 캡처 정책, 사용 도구 등을 포함하는 예비조사 단계와 내부 노드를 타고 접근하는 방식, 외부의 링크를 타고 횡적으로 접근하는 방식, 그리고 질, 주제, 유형, 생산자로 구분할 수 있는 필터링 단계로 나뉜다고 하였다. Brown(2006)은 선정 과정을 정책 정의, 선정, 유지의 3단계로 구성되어 있다고 하였으며, 선정을 잘 하기 위해서는 선정과정의 지속성과 수집의 사이클 운영 및 피드백이 중요하다고 주장하였다. 웹 자원의 선정을 위해서는 웹 자원에 대해 평가하기 위한 기준이 필요한데, 웹 자원의 평가요소로 Fritch와 Cromwell(2001)은 저자식별요소, 저자소속기관, 정보원의 권위를 들고 있고, Alexander와 Tate(1998)는 웹 자원의 평가요소로 정확성, 신뢰성, 객관성, 최신성, 주제의 포괄성, 하이퍼링크사항, 안전성, 소프트웨어의 요구사항 등을 들고 있다.

이와 같이 선행연구 및 국내의 현황을 조사해 보았을 때 가이드라인의 방향성을 제시한 연구(이재선 2008)는 있으나 웹 아카이빙을 위한 구체적인 가이드라인 개발 연구는 거의 없는 것으로 조사되었으며, 따라서 본 연구는 웹 아카이빙 선정지침 제안에 중점을 두고 있다. 선행연구를 종합적으로 분석하여 도출된 시사점을 정리하면 다음과 같다.

첫째, 웹 아카이빙 방법은 크게 선택적 수집방법과 포괄적 수집방법이 있는데(서혜란 2004; 윤정옥 2011), 각각의 장단점을 보완하여 하이브리드 아카이빙 방법이 바람직하다고 주장하는 연구(김지훈 2009)도 있다. 그러나 법적인 문제를 해결하는 것이 쉽지 않기 때문에 선택적 수집방법으로 가는 것이 바람직하다. 실제로 세

계의 많은 아카이빙 사이트들이 선택적 수집방법을 채택하고 있는 실정이다(서혜란 2004).

둘째, 안정적 웹 아카이빙을 위해 온라인자료에 대한 디지털 납본법이 수립되어야 한다는 연구가 다수 있음을 알 수 있다(서혜란 2003; 김유승 2007). 즉 웹 아카이빙을 하고 있는 나라들이 디지털자료의 수집과 납본을 위해 납본법에 웹 자원을 납본 대상으로 포함시키도록 법을 개정해 가고 있다는 것이다.

셋째, 웹 아카이빙을 위한 수집대상은 표면웹과 심층웹으로 구분할 수 있다. 지금 현재 OASIS는 표면웹과 심층웹 모두를 대상으로 하고 있다. 단 심층웹의 경우 사이트의 저작권 문제를 해결한 사이트에 한하고 있다. 또한 OASIS는 수집, 보존, 서비스하는 웹 아카이빙의 대상을 개별자원과 웹 사이트로 하고 있으며, 개별자원은 하나 또는 그 이상의 파일단위로 가치를 가지는 정보 자원이고, 웹 사이트는 하나의 단위로 해당 사이트의 모든 자원을 미러링하게 하는 자원이다(윤정옥 2010).

넷째, 웹 아카이빙을 위한 선정지침 개발과 관련된 연구는 그다지 많지 않다. 대부분의 연구가 선정지침 개발의 방향성을 제시하거나, 아카이빙을 위한 평가기준을 제시하는 데 중점을 두고 있음을 알 수 있다.

다섯째, 웹 아카이빙과 관련된 문헌들을 살펴보면, 디지털자원 아카이빙과 웹 자원 아카이빙, 온라인디지털자료 아카이빙, 웹 아카이빙, 온라인자료아카이빙, 온라인아카이빙 등과 같이 매우 다양하게 사용되고 있는 것을 알 수 있다. 웹 아카이빙은 웹상에서 유통되는 자료의 아카이빙이고, 온라인디지털자료 아카이빙은 온라인상에서 유통되는 디지털자료의 아카

이빙을 의미한다고 할 수 있다. 본 연구에서는 웹상에서 유통되는 디지털자료를 아카이빙 한다는 의미에서 디지털자원 아카이빙과 웹 자원 아카이빙, 온라인디지털자료 아카이빙, 웹 아카이빙, 온라인자료아카이빙, 온라인아카이빙을 모두 포괄할 수 있는 웹 아카이빙이라는 용어를 선택하고자 한다.

3. 연구질문

본 연구의 목적은 웹 아카이빙을 위한 구체적인 선정지침을 제시하는 것이다. 국내의 연구 및 OASIS의 홈페이지 등에 게재된 선정과 관련된 내용들을 집중적으로 분석하여 보았으나 실무자들이 즉시 적용할 수 선정지침으로는 많은 한계가 있는 것으로 생각되었다. 국립디지털도서관의 아카이빙 대상 웹 자원 선정과정을 보면, 먼저 관리자가 수집할 대상을 선정하고 저작권 문제를 해결하기 위해 각각의 사이트와 접촉하여 저작권 동의를 받은 사이트에 한하여 웹로봇으로 하여금 사이트의 정보를 정기적으로 수집하도록 하고 있다. 또한 수집로봇이 지속적으로 등록된 사이트의 자원을 수집하여 변화 상태를 감시하고 이전에 저장한 상태와 비교하여 변화 상태를 수치로 표현하면, 관리자가 변화 정도를 판단하여 현재 수집한 상태를 저장하여 보존하는 방식을 취하고 있다.

문제는 현재의 선정지침이 관리자가 어떤 자원을 웹 아카이빙 대상으로 선정할지에 대한 어떤 명확한 답을 줄 수 있는가 이다. 매년 상당한 예산을 들여 웹 아카이빙 사업을 하고 있지만 외주업체가 가장 어려운 점으로 제시하고

있는 것은 수집대상 자원의 확보인 것으로 조사되었다. 가치 있는 수집대상 자원의 빠른 확보를 위해 온라인자료선정위원회도 운영하고 있지만 선정된 자원의 적절성을 평가하기 보다는 수집대상 자원을 제안하는 데 초점을 두고 있어 선정위원이 기존의 선정지침을 명확히 알고 있고, 기존 자료를 명확히 이해하지 않고서는 적절한 자원을 제안하는 것이 쉽지 않다. 게다가 1년에 한두 번 회의를 갖는 것으로 충분하지 않을 것으로 사료된다.

본 연구에서는 현 OASIS 선정지침에 대한 개선방향을 포괄적으로 모색하고 있다. 따라서 선정지침의 구체적인 방향을 제안하는 근거로 선행연구자들의 연구결과와 해외 웹 아카이빙 선정지침이다. 분석대상이 된 해외 사례는 지금까지 선행연구자들이 우수사례로 꼽아 왔던 사례 중에서 선정지침이 홈페이지에 공지되어 있는 것으로서, 대표적으로 호주의 PANDORA, 미국의 MINERVA, 영국의 National Archives, 덴마크의 Netarchive.dk 등이 있다. 본 연구를 진행하는 과정에서 제기된 연구질문은 다음과 같다. 연구질문은 선정지침을 분석하고 개발함에 핵심이 되는 내용만을 뽑은 것이며, 이를 기반으로 분석된 내용을 선정지침의 세부적인 내용 개발에 적용하고자 하였다. 또한 논의부분에서 별도로 논의하고자 한다.

RQ1: 해외 웹 아카이빙 사이트들은 OASIS와 비교하여, 명확한 웹 아카이빙 정책을 가지고 있는가, 그리고 수집대상자원과 수집제외자원에 차이가 있는가?

RQ2: 기타 OASIS에 나와 있지 않는 매우 중요한 부분이 해외 웹 아카이빙 사이트들의 선정지침에 나와 있는가?

RQ3: 디지털 납본법의 수립 유무가 아카이빙 정책에 영향을 미치는가?

4. 해외 아카이빙 사이트의 수집대상 자원분석

4.1 영국 National Archives

National Archives는 주로 영국 중앙 정부 웹 사이트를 아카이빙하며, 현재 아카이빙 되어 테마별로 서비스되고 있는 주제는 다음과 같다.

- 2012 Olympic and Paralympic Games and Cultural Olympiad
- Budget 21 March 2012
- Regional Development Agencies
- Budget 23 March 2011
- Japan Earthquake
- Pakistan Flood - Summer 2010
- Transparency Agenda
- Budget 22 June 2010
- General Election - May 2010
- Volcanic Ash Cloud
- Budget 24 March 2010
- Financial Crisis
- Avian Influenza
- Swine Influenza
- National Security 2006

National Archives는 UKWAC(Web Archiving Consortium)이라는 컨소시움을 구성하여 2004년부터 2009년까지 운영하였으며, British

Library, TNA, The National Libraries of Wales and Scotland, JISC, The Wellcome Trust의 6개 파트너들로 구성되는데 각 파트너들은 웹 아카이빙의 주제를 나눠서 담당했었다. British Library는 문화·역사·정치 분야, TNA는 주요 정부기관 웹 사이트, The National Libraries of Wales는 웨일즈의 문화·역사 분야, The National Libraries of Scotland는 스코틀랜드의 문화·역사 분야, JISC는 정보통신 분야, 첨단 혁신 사업, Wellcome Trust는 의학 분야를 담당했었으며, 현재 National Archives 사이트에서 모두 서비스되고 있다.

4.2 미국 MINERVA

MINERVA는 107대 의회, 2002년 선거, 2001년 9/11 테러, 2000년 선거, 2002년 동계 올림픽, 9/11 추모, 이라크 전쟁, 108대 의회, 2004년 선거 등 현재 이슈가 되고 있는 주제에 대해 각각 컬렉션을 구성하고 웹 아카이빙을 하고 있으며, 현재 다음과 같은 주제가 홈페이지에서 서비스 되고 있다.

- Crisis in Darfur, Sudan, Web Archive, 2006
- Indian General Elections 2009 Web Archive
- Indonesian General Elections 2009 Web Archive
- Iraq War 2003 Web Archive
- Law Library Legal Blawgs Web Archive
- Library of Congress Manuscript Division Archive of Organizational Web Sites
- Papal Transition 2005 Web Archive

- September 11, 2001 Web Archive
- Single Sites Web Archive
- United States 107th Congress Web Archive
- United States 108th Congress Web Archive
- United States Election 2000 Web Archive
- United States Election 2002 Web Archive
- United States Election 2004 Web Archive
- United States Election 2006 Web Archive
- United States Election 2008 Web Archive
- Visual Image Web Sites Archive

또한 MINERVA는 미 의회와 파트너 관계에 있는 기관들에 의해 개발된 모든 웹 아카이브도 제공하고 있으며, 다음의 주제로 구분하여 제공하고 있다.

- U.S. Elections Web Archives: 도서관은 2000년대 이래의 미국 대통령선거, 의원선거, 지방의원 선거와 관련된 웹 사이트를 아카이빙해서 최근 자료는 홈페이지에 공개한다.
- U.S. Congressional/Legislative Web Archives: 도서관은 정기적으로 의원 및 의원위원회의 웹 사이트를 아카이빙 한다.
- Law Library Web Archives: 의회법률 도서관은 대법원 내 변화를 문서화하는 웹 콘텐츠를 아카이빙 한다. 또한 미국변호사 협회와 관련된 권위 있는 법률 블로그를 선택적으로 수집한다.
- Single Sites Web Archive: 특정 주제나 이벤트와 관련 없는 웹 사이트도 아카이빙하는 방법을 시도하였는데, Library's Rec-

ommending Officers는 도서관의 전통적인 아날로그 장서의 확장을 위해 디지털로만 생산된 콘텐츠를 선정해서 서비스 하고 있다.

- Other Events and Themed Web Archives: 그 외에 다양한 이벤트별 및 테마별 컬렉션을 개발하여 보존하고 있으며, Small Press Expo and Comic Art Collection, Performing Arts Web Archive, Civil War 150th Web Archive, Public Policy Topics, LC Manuscript Division Archive of Organizational Web Sites 등이 있다.
- International Web Archives: 도서관은 카이로, 자카르타, 뉴델리, 리오 등에 도서관 해외지사(Library's Overseas Office)를 두고 해당 지역의 선거나 이벤트를 다루는 웹 사이트를 아카이빙 하고 있다. Burma/Myanmar Election 2012, Timor Leste Election 2012, Egyptian Elections 2012 등의 주제가 서비스되고 있다.
- Collaborations: 도서관은 미국에 알려지지 않은 세계의 다른 지역에서 일어나는 이벤트를 아카이빙 하기 위해 다른 기관과 협력하고 있으며, 협력기관을 통해서 제공되도록 한다. 버지니아텍과 협력해서 구축된 Japanese Earthquake, National Library of France와 협력해서 구축한 North Africa & the Middle East 2011 등이 있다.

4.3 호주 PANDORA

호주 PANDORA는 파트너 기관의 특성을 살펴 자료 수집을 하고 있는 것이 특징적이다.

즉, 파트너 기관들이 자관의 목적에 맞게 수집 정책을 펼치고 있으면서도, 수집이 중복적이지 않고 상호보완적이라는 것이다(이재선 2007). PANDORA에 수집되는 정보원은 국가적으로 중요한 정보(NLA), 각 주에 관련된 중요한 정보(주립도서관들), 호주 원주민과 관련된 자료(Australian Institute of Aboriginal and Torres Strait Islander Studies), 전쟁에 관련된 정보(호주 전쟁기념관), 영화나 음향과 관련된 정보(Screen Sound Australia: The National Screen and Sound Archive) 등을 수집대상으로 하고 있다.

특히 NLA는 우선수집자원과 기타로 구분하여 수집대상 자원을 구분하고 있으며, 우선 수집 자원은 Commonwealth and ACT government publications, Publications of tertiary education institutions, Conference proceedings, E-journals, Titles referred by indexing and abstracting agencies, Topical sites이며, 기타 직원의 시간이 허용한다면 추가적으로 수집하는 자원은 Annual reports, Digitised materials, Educational material, Exhibitions, Ethnic community sites, Juvenile publications, Literary works, Maps, Music, Newspapers, Organisational and personal sites, Other protocols(e.g.), FTP, Gopher, Email, Pacific publications, Religious sites, State government publications이다.

4.4 덴마크의 netarchive.dk

덴마크의 웹 아카이빙을 위한 수집전략은 크게 3가지 유형으로 구분되며, 주로 덴마크에 관련된 인터넷 웹 사이트를 아카이빙 한다. 첫째,

'dk' 도메인 전체에 대한 '스냅샷' 수집방법을 1년에 4회 수행한다. 둘째, 역동적 사이트에 대해서는 선택적 수집을 수행한다. 셋째, 특별한 이벤트와 관련된 웹 사이트를 수집한다.

① 스냅샷 수집(벌크 하베스팅)

대량수집방법은 최상위 도메인인 'dk'에 대한 완전한 그림을 수집하기 위한 것이다. 수집은 수집될 도메인의 목록으로부터 시작되며, 이 도메인 목록은 최상위도메인 'dk' 관리자에 의해 제공된다.

② 선택적 수집

선택적 수집은 자주 갱신되는 웹 사이트와 대량수집방법에 의해 수집되지 않을 수 있는 웹 사이트를 수집하기 위해 수행된다. 즉, 뉴스 사이트(국가 미디어 또는 지역 미디어), 시민 사회, 상업영역, 공적기관들을 대표해서 매우 많이 이용되는 역동적 특성이 있는 웹 사이트, 그리고 인터넷을 이용해서 새로운 방법을 시도하는(예를 들어 net art) 실험적이나 독특한 사이트 등이 수집된다.

③ 이벤트 사이트 수집

이벤트 사이트 수집은 매우 특별한 사건에 대해서 기술하는 새로운 사이트이지만, 그 사건이 종료되었을 때 사라져 버릴 수 있는 사이트를 대상으로 한다. 즉, 수집될만한 가치가 있는 사건으로 첫째, 덴마크 역사에 중요할 것으로 기대되는 국민들 간의 토론이나 덴마크 사회발전에 영향을 줄 수 있는 토론내용이 있는 이벤트 사이트, 둘째, 새로운 이벤트 웹 사이트가 나오게 만든 사건의 웹 사이트, 셋째, 기존의

〈표 1〉 해외 아카이빙 사이트의 수집 특징

국가	사업명	사이트 URL	수집 특징
영국	National Archives	http://www.nationalarchives.gov.uk/	<ul style="list-style-type: none"> • 컨소시움(UKWAC) 구성 • 6개 파트너들로 구성되어 각 파트너들은 웹 아카이빙의 주제를 나누어 담당
미국	MINERVA	http://www.loc.gov/	<ul style="list-style-type: none"> • 현재 이슈가 되는 주제에 대해 각 컬렉션을 구성 • 미 의회와 파트너 관계에 있는 기관들에 의해 개발된 모든 웹 아카이브 제공
호주	PANDORA	http://pandora.nla.gov.au/	<ul style="list-style-type: none"> • 파트너 기관의 특성과 목적에 맞게 수집 정책 실현 • 중복적이지 않고 상호보완적으로 수집
덴마크	netarchive.dk	http://netarkivet.dk/	<ul style="list-style-type: none"> • '.dk' 도메인 전체에 대한 스냅샷 수집 • 역동적 사이트는 선택적 수집 • 특별한 이벤트 관련 웹 사이트 수집

웹 사이트들에 대해서 광범위하게 다루고 있는 이벤트 웹 사이트 등이다.

위의 사례들을 분석해 보았을 때 각 나라마다 선정대상이 매우 다르다는 것을 알 수 있다. 호주의 경우 수집원칙이 우리나라와 비슷하지만 구체적인 선정대상 자원이 매우 다르다. 또 어떤 나라가 선정대상을 웹 아카이빙 대상으로 가장 잘 하고 있다고 평가하기도 어렵다. 이는 각 나라의 웹 아카이빙 기관의 목적과 방향에 따라 달라질 것이기 때문이다. 해외 아카이빙 사이트의 수집 특징을 간단히 정리하면 다음 〈표 1〉과 같다.

5. 웹 아카이빙 선정지침 개선방향

본 연구에서는 웹 아카이빙 선정지침을 제안하는 것을 목적으로 하고 있으며, 특히 우리나라 국가 대표 아카이브인 OASIS를 위한 선정지침을 제안하고자 하였다. 선정지침을 분석함에 있어서 분석 포인트를 다양하게 잡을 수 있으며, 김유승(2008)은 수집대상 정보자원 기준

으로 정보자체의 특성과 한국과의 관련성 측면으로 나누어 OASIS의 선정지침을 분석한 바 있다.

현재 OASIS의 아카이빙 수준을 보면 보통 3단계 수준까지 저장하지만 경우에 따라서는 초기화면만 저장되는 경우도 있다. 수집 대상 웹 사이트는 외부와 연결된 내용들은 수집되지 않고 내부의 페이지만 수집하고 있다. OASIS는 선별적 수집을 택하고 있으며, 수집 대상에 포함된 웹 사이트는 지속적으로 관찰된다. 수집 대상 웹 사이트는 정적인 사이트이며 자바, 플래쉬 등으로 제작된 동적인 사이트와 DB로 제작된 심층웹은 기술적인 이유로 수집이 되지 않는다. OASIS의 웹 로봇은 최근 수집된 사이트를 기준으로 변동여부를 주기적으로 체크하여 최근 저장본과의 변동비율을 알려준다. 수집 팀에서는 이러한 변동비율을 기준으로 해당 사이트의 재수집 여부를 판단하며, 전체적인 디자인이 바뀌었거나, 특별한 사회적 이슈가 되는 사이트일 경우 매일 수집하기도 한다(이재선 2007).

이와 같이 OASIS의 전반적인 수집 정책은

정해져 있으나 그 가이드라인의 구체성이 매우 미비한 관계로 실제로 선정할 정보자원을 지정할 때 많은 애로사항이 있는 것으로 알려져 있다. 현재 OASIS 선정지침의 문제점은 앞에서도 부분적으로 언급하였지만, 개선안의 필요성을 부각시키기 위해 다음과 같이 종합적으로 정리하여 제시하고자 한다.

첫째, 선행연구 분석결과에서 보듯이 웹 아카이빙을 위한 선정지침 개발과 관련된 연구는 많지 않으며, 현재의 OASIS 선정지침은 2006년 개발된 이후 전혀 개정이 이루어지지 않아 시대적 흐름을 반영하지 못하고 있는 문제점이 있다.

둘째, 디지털납본정책의 미비로 선택적 수집을 할 수 밖에 없으며, 효과적 선택적 수집을 위해 수집대상 목록이 개발되어 있어야 함에도 불구하고 이러한 목록은 없는 것으로 조사되었다.

셋째, 웹 아카이빙과 관련된 문헌들에 나타나는 용어들은 디지털자원 아카이빙, 웹 자원 아카이빙, 온라인디지털자료 아카이빙, 웹 아카이빙, 온라인자료아카이빙, 온라인아카이빙 등과 같이 매우 다양하게 사용되고 있는 것을 알 수 있는데, 선정지침에서는 용어들에 대한 각각의 정의를 포함할 필요가 있으며, 대표적으로 사용할 용어에 대한 지침도 포함하여야 할 것이다.

넷째, 수집대상인 웹자료에 대한 정의가 명확하지 못하기 때문에 선정자에게 혼선을 줄 수 있다.

다섯째, 수집기본원칙에 수집 대상 자원에 대한 명확한 지시 내용이 포함되어 있지 않아서, 자료선정을 위한 가치판단에 한계가 있을

것으로 보인다.

여섯째, 수집방법과 평가기준, 수집주기에 대한 내용이 명확히 기술되어 있지 않다.

일곱째, 선정대상자원이나 선정제외자원에 대한 근거가 명확하지 않고 포괄적이어서 혼선이 생길 수 있다. 예를 들어 '포털사이트 제외'라고만 하지 말고 '인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트는 제외'라고 하는 것이 아카이빙 담당자를 보다 효과적으로 지원해 줄 수 있을 것이다.

이러한 문제점들을 해결하기 위해 본 연구에서는 OASIS 선정지침의 개선방향을 모색하고자 하였다. ① 추진배경, ② 웹자료의 정의, ③ 용어정의, ④ 수집기본원칙, ⑤ 수집방법, ⑥ 수집대상자원, ⑦ 수집제외자원으로 구분하고, 각각에 대한 문제점을 분석하고 개선방향을 제시하고 있으며, 이를 기반으로 한 개선안에 대한 내용은 <부록 1>로 제시하였다.

5.1 사업추진배경

정보통신 환경의 급속한 발달로 많은 지적 창작물이 디지털 형태로 생산되어 인터넷상으로 유통되고 있으나 웹 자원의 짧은 생명주기, 소멸성 등으로 가치 있는 자원을 보존하고 이용시키는데 많은 문제가 있는 것으로 지적되어 왔다. 이에 세계 여러 나라에서는 90년대 중반부터 인터넷 지적 자원을 수집하고 보존하는 프로젝트를 국가적인 차원에서 추진하여 PANDORA 등을 비롯한 많은 시스템들이 개발되어 전 세계적으로 활용되고 있다.

국내에서 이러한 시대의 흐름에 발맞추어 2000년 초반에 OASIS 사업을 추진하였는데, 그 추

진 배경은 국가 지식 자원의 수집과 보존 책임 기관인 국립중앙도서관이 국가대표도서관으로서 가치 있는 인터넷 자료를 국가적인 차원에서 수집·축적하여 미래 세대에 연구 자료로 제공하고 온라인 디지털자원 수집·보존의 표준모델 연구 등에 기여하기 위한 것이다.

5.2 웹자료의 정의

OASIS의 '온라인자료 선정지침'에 따르면 온라인자료의 정의 부분에 다음과 같이 나와 있다.

- ① 인터넷을 통해 전자화된 온라인 자료의 내용을 내려 받아 개인용 및 휴대용 컴퓨터 등의 정보통신 단말기에서 볼 수 있는 디지털 버전의 출판물
- ② 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
- ③ 범위: 웹 사이트, 웹페이지, 문서파일(pdf, hwp, doc, txt,...), 이미지, 동영상, 음악, 압축파일

위의 내용을 종합적으로 정리하면, 온라인 자료는 온라인상에서 유통되는 모든 형태의 정보자원을 의미하며 온라인상에서 유통되고 활용될 수 있는 자료라고 할 수 있다. 그러나 ①과 ②의 내용이 중복되어 보이고 ③이 분리되어 나타나는 느낌이 든다. 따라서 '온라인 자료에 대한 정의'는 다음과 같이 수정되는 것을 권

장한다.

“온라인상에서 유통되고 활용될 수 있는 모든 형태의 정보자원으로서, 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 파일을 포함한다.”

5.3 용어정의

수집대상을 명확히 하기 위해 해외의 대부분의 가이드라인은 용어정의(terminology)를 포함하고 있다. PANDORA의 경우 'Online', 'Home Page', 'Site', 'Title', 'Archive', 'Preserve'에 대한 용어에 대해 간단히 소개하고 있다. 온라인, 홈페이지, 사이트, 타이틀 등은 명확히 정의되지 않으면 혼선이 야기될 수 있으며, 아카이빙의 맥락에서 명확히 정의될 필요가 있다. 따라서 OASIS는 다음과 같은 용어정의를 포함할 것을 제안하며, 이 정의는 문헌정보학용어사전¹⁾과 네이버 국어사전을 기본으로 하되, 학자들의 정의 및 해외 선정지침의 자료 번역으로 보충하였다.

- 온라인(Online)은 일반 PC에서 외부의 다른 시스템과 통신 회선이 전화에 연결된 상태를 뜻하는 것으로 시스템이 온라인 상태일 때 외부의 정보를 이용할 수 있다.
- 홈페이지(Home Page)는 URL을 사용해 특정 호스트에 접속했을 때 가장 먼저 나타나는 웹 페이지이다.
- 사이트(Site)는 인터넷 상에 있는 서버를 가리키는 것으로, 정보가 필요할 때 언제

1) <<http://203.241.185.12>>.

든지 그것을 볼 수 있도록 웹 서버에 저장된 집합체이다.

- 타이틀(Title)은 아카이빙을 위해 고려되고 있는 실체이다. 1개의 문서가 될 수도 있고, 인터넷 주소로 다수의 문서로 연결되기도 한다.
- 아카이브(Archive)는 인터넷으로부터 다운로드하고, 국가도서관 서버에 저장되는 현상의 의미로 사용된다.
- 보존(Preserve)은 원래의 모습 그대로 보호하고 간수하여 남아 있게 하는 방법이다.

또한 한국기록학회(2008)에서 엮은 기록학 용어 사전을 보면 온라인, 홈페이지, 사이트, 타이틀의 뜻은 나와 있지 않고, Archive(아카이브)도 보존 기록, 혹은 보존 기록관으로 두 가지 의미로 사용하고 있어서 정확한 뜻을 가져 오기는 어렵다. 단, Preserve(보존)은 다음과 같이 명시되어 있다. 열화(劣火)나 손상으로부터 자료를 보호하는 과정. 정보를 담고 있는 기록이나 도서류를 보관하기 위한 시설·인원·기술 등 관리적인 측면과 재정적인 측면을 모두 포함하는 포괄적인 개념이다.

5.4 수집기본원칙

온라인 자료의 수집원칙 및 수집 범위에 대한 것은 이미 많은 연구자들이 언급하고 있으며, 따라서 본 연구에서는 이들 연구자들의 연구결과와 해외사례를 집중적으로 분석함으로써 수집기본원칙의 방향성을 제안하고자 한다. 온라인 자료의 수집범위에 대한 것은 2008년에

수행된 김유승의 연구에서 자세히 잘 나와 있고, 2010년의 윤정옥의 연구에서 다시 한 번 정리된 바 있으며, 그 외의 연구에서도 부분적으로 언급되고 있다(서혜란 2004; 김지훈 2009). 즉, 온라인자료의 수집 범위는 ① 개별자원과 웹 사이트 중심으로 하는 방법, ② 주제 중심으로 하는 방법(OASIS, MINERVA)과 도메인 중심으로 하는 방법, ③ 지리적 경계를 중심으로 하는 방법(BnF)과 지리와 관계없이 한 국가를 대상으로 하는 방법 등으로 구분할 수 있다.

영국의 국립아카이브²⁾는 주제 중심이며, 현재 The Record: A Collections Strategy for the 2012 Olympic and Paralympic Games and Cultural Olympiad, Business Archives, Black, Asian and Minority Ethnic Performing Arts, Web Archives, Religious Archives 등을 중심으로 수집하고 있다.

미국 MINERVA도 주제 중심이며, 2000년 선거, 9/11, 2002년 선거, 107대 의회 등의 사건과 같이 사회, 문화적으로 영향을 끼친 사건에 맞는 내용으로 웹 아카이빙을 진행하고 있다. MINERVA는 선정된 주제를 잘 표현할 수 있는 사이트를 수집하기 위해 해당 주제의 전문가들의 의견을 반영하고 있다(Groke & Jones 2003). OASIS는 최근 이슈가 되고 있는 자원을 선정함에 있어 국회의원 선거, 신행정수도 등 최근 이슈가 되는 자원 중 최신성, 희소성, 유용성을 판단하여 수집·보존 가치가 높은 디지털자원, 국제적인 명성과 권위 있는 기관으로부터 평가 받은 저널 등을 수집대상에 포함한다고 하고 있다. 그러나 최근 이슈 중 MINERVA와 같이 사회적 및 문화적 가치를 기준으로 한

2) <<http://www.nationalarchives.gov.uk/archives-sector/collections-strategies.htm>>.

다는 내용이 포함되어야 할 것이다.

호주 PANDORA는 우리나라와 유사하게 선정 대상을 다음과 같이 지정하고 있다. ① 호주인에 관한 것이어야 한다. ② 호주와 관련하여 사회, 정치, 문화, 종교, 과학 또는 경제적인 중요성이 있는 주제이거나 호주인에 의해 쓰인 것이어야 한다. ③ 권위 있는 호주인에 의해 쓰인 연구가치가 있는 자료여야 한다.

덴마크의 Netarchive.dk는 도메인 중심이자 지리적 경계를 기준으로 선정대상 자원을 수집하는 원칙으로 최상위도메인을 중심으로 광범위한 영역을 포괄하는 방식이다. 그 구체적인 선정대상자원은 ① 덴마크 도메인 상에 존재하는 것, ② 업데이트가 잦은 중요한 내용을 담고 있는 것, ③ 예측 가능한 혹은 예측 불가능한 국가적인 행사 등이다.

현재 우리나라의 OASIS는 지리와 관계없이 한국과 관련된 중요한 정보원을 수집하고 있으며, 따라서 다른 나라에서 생산된 한국과 관련된 것이면 모두 수집하고 있다. 따라서 선택적 수집방법을 통해 한국과 관련된 정보원을 제공하는 기관과 협약을 통해 이러한 자료를 수집해야 한다. 또한 OASIS는 주제 중심의 수집방식을 채택하고 있다고 할 수 있으며, 따라서 한국과 관련된 사회, 정치, 문화, 종교, 과학 또는 경제적으로 중요한 주제를 수집 대상 자원으로 하고 있다. 단순히 수집대상 도메인을 지정해주는 것이 아니라 수집 대상 주제를 분석하여 하나하나 선택해야 한다. OASIS는 정보자원의 특성도 고려하고 있다고 할 수 있는데(윤정옥 2010; 김유승 2008), 즉 현재 또는 미래의 정보요구에 봉사하는 유용성, 저작자의 평판, 제공된 정보의 유일성, 학술적 내용, 정보의 최신성,

업그레이드 빈도, 접근의 용이성 등을 고려하여 수집 대상 자원을 결정하고 있다(OASIS).

그러나 OASIS의 수집기본원칙은 홈페이지에 나와 있는 것과 2007년에 개발된 것의 내용이 다르며, 이에 대한 수정 보완이 요구된다. 따라서 위의 분석 내용을 기준으로 OASIS의 수집기본원칙에는 다음의 내용이 들어가도록 제안하고자 한다.

첫째, OASIS의 수집 대상 자원이 지리적 위치와 관계없이 주제 중심임을 명확히 해주어야 한다. 즉, 수집기본원칙에는 다음의 내용이 들어가도록 해야 한다. “수집대상 자원의 주제는 한국과 관련된 국가적 보존가치가 있는 것으로, 한국과 관련된 사회, 정치, 문화, 종교, 과학 또는 경제적으로 중요한 주제이고, 따라서 지리적 위치에 관계없이 한국과 관련된 것으로 보존가치가 있는 것은 해외에서 생산된 것도 수집하도록 한다. 또한 한국의 저자에 의해 쓰인 것이어야 한다. 또한 국내의 저명한 대학 교수나 대학 연구원 등 권위 있는 한국 사람에 의해 쓰이거나 국내외적으로 해당 학문분야에 기여를 한 것이어야 한다.”

둘째, OASIS는 국가적으로 보존가치가 있는 수집대상 자원을 선정함에 있어 수집대상 자원의 특성을 고려하고 있음을 명확히 해야 한다. 즉, 수집기본원칙에는 ‘수집대상 자원의 특성으로 현재 또는 미래의 정보요구에 봉사하는 유용성, 저작자의 평판, 제공된 정보의 유일성, 학술적 내용, 정보의 최신성, 업그레이드 빈도, 접근의 용이성 등을 고려한다.’는 내용이 들어가야 한다.

셋째, OASIS는 수집대상 자원을 수집함에 있어 최근 이슈와 관련된 자원을 수집대상으로

하고 있음을 명시해야 한다. 즉, 수집기본 원칙에는 정치적 이슈, 사회적 이슈, 문화적 이슈, 그리고 국제적 이슈가 있을 경우 그 역사적 가치를 고려하여 수집하여야 함을 명확히 제시해야 한다.

넷째, 오프라인 자료가 있는 온라인 자료와 처음부터 온라인 자료만으로 발간되는 자료에 대한 명확한 구분제시가 필요하다. 즉, 오리지널 온라인 디지털 자원은 망라적으로 수집하며, 오프라인 자료가 있는 온라인 자료는 수집을 지양한다는 내용이 명확히 제시되어야 한다.

다섯째, 소위 가변성이 높은 동적 출판물은 업데이트 전후의 내용변화와 구성체제의 상당한 변화 등을 감안하여 선택하며, 온라인 디지털자료의 포맷이 여러 종류일 경우에는 표준포맷으로 구성된 자료를 우선적으로 선택한다는 내용이 구체적으로 명시되어야 한다.

5.5 수집의 방법

온라인 자료의 수집방법에 대한 것도 국내외 많은 저지들에 의해 언급되고 있다(서혜란 2004; 윤정옥 2011; 김지훈 2009). 수집의 방법에는 망라적 수집과 선택적 수집, 그리고 두 방법을 혼합한 하이브리드 방법이 있다.

선택적 수집을 주장하고 있는 연구자들의 논리는 선택적 과정으로 인한 웹자료의 품질 보장이 가능하고, 정보제공기관과의 저작권협의를 거친 경우 포괄적 수집으로 수집할 수 없는 사이트들도 수집 가능하며, 이러한 사이트들에 대해서는 영구적 보존 및 공공이용이 가능하다는 것이다. 수많은 대상 자료 중 일정한 기준을 적용해서 일부를 선정한 후 수집하는 선택적 수집

을 하고 있는 경우는 미국의 MINERVA, 호주의 PANDORA, 일본의 WARP이다. 현재 우리나라의 OASIS도 선택적 수집을 하고 있다. 그러나 어떤 웹 자원이 수집 가치가 있는지를 결정하는 전문 인력이 필요한 노동집약적 방법이라는 한계를 극복하기 위한 전략이 필요하다고 본다.

망라적 수집을 주장하고 있는 연구자들의 논리는 일단 수집 가치가 있는 웹 자원을 미리 선정할 필요 없이 무조건 수집하고 미래 세대에 필요할 모든 정보를 수집할 수 있다는 것이다. 웹 로봇 프로그램을 사용해서 망라적 자동수집을 하고 있는 경우는 스웨덴의 Kulturarw3, 핀란드 국립도서관, 스웨덴 왕립도서관이다. 그러나 저장 공간의 한계, 데이터 저장 비용, 웹 자료의 품질, 심층웹에 대한 접근의 어려움 등의 문제는 해결되어야 할 것이다.

한편, 각각의 장단점을 보완하여 하이브리드 아카이빙 방법이 바람직하다고 주장하는 연구(김지훈 2009)도 있다. 최재황, 광승진, 김정택(2009)은 우리나라의 경우, ac, re, go, or 등의 도메인을 대상으로 선택적 수집을 먼저하고 그 범위를 점진적으로 확대해 나가는 하이브리드 방법을 권장하고 있다.

현재 우리나라의 경우 온라인자료에 대한 디지털 납본법이 통과되지 않고 있어서 수집사이트에 대해 정보제공기관들로부터 하나씩 수집 허락을 받아야 하는 상황이다. 따라서 우리나라의 웹 아카이빙은 선택적 수집방법을 채택할 수밖에 없다. 또한 모든 사이트를 아카이빙 하는 것이 실제로 가능하지도 않고 그럴 필요도 없기 때문에 철저한 선정지침을 기반으로 선택적 수집을 하는 것이 바람직하며, 따라서 타당성 있는

선정지침의 개발은 매우 중요한 일이라 할 수 있다. 이러한 이유로 MINERVA, PANDORA, WARP 등 세계의 많은 아카이빙 사이트들이 선택적 수집방법을 채택하고 있는 실정이다(서혜란 2004). 따라서 OASIS의 온라인자료 선정지침에는 OASIS가 선택적 수집방법을 채택하고 있음을 명확히 명시하여야 한다.

5.6 수집주기

선정지침에 수집주기에 대한 부분이 명확히 제시될 필요가 있다. 현재 OASIS의 경우 수집로봇이 지속적으로 등록된 사이트의 자원을 수집하여 변화 상태를 감시하고, 이전에 저장한 상태와 비교하여 변화 상태를 수치로 표현하면, 관리자가 변화 정도를 판단하여 현재 수집한 상태를 저장하여 보존하는 방식을 취하고 있다. 관리자가 이를 매번 판단하여 변화 정도를 판단하고 수집 주기를 정하는 것은 판단과 시간의 한계가 있을 것으로 보인다. 따라서 수집로봇으로 하여금 웹 자원의 변화 주기를 자동적으로 체크하고(Arvidson, Persson, & Mannerheim 2001), 특정 기준을 정해 그 기준 이상의 변화를 보이는 사이트를 재수집하는 방법이 권장된다. 가장 적절한 방법은 각 사이트 당 평균변화 시간을 측정하고 이를 기준으로 각 사이트의 수집주기를 정하는 지침이 들어가야 할 것이다.

결과적으로 OASIS의 온라인자료 선정지침에는 OASIS가 웹 사이트 수집을 위해 수집로봇을 사용하고 있으며, 수집로봇이 자원의 변화 정도를 판단하여 수집주기를 정하고 있음을 명시할 필요가 있다.

5.7 평가기준

웹 자원의 평가기준과 관련된 연구가 다수 연구자에 의해 수행되었으며, 제안되는 평가기준은 각 연구자마다 다르다는 것을 알 수 있다. 이재선(2008)은 국가기록원의 웹 아카이빙을 위한 선별 기준으로 웹 사이트 내용과 성격에 대한 평가를 위한 저작성, 내용의 정확성 및 객관성, 포괄성, 희귀성 및 유일성, 최신성, 저작자의 평판, 가변성, 접근의 용이성으로 총 8가지 항목을 제안하였다. Fritcher와 Cromwell(2001)은 저자 식별요소(identity), 저자의 소속기관(affiliation), 그리고 인식측면의 권위성(cognitive authority)을 들고 있고, Alexander와 Tate(1998)는 웹 자원의 평가요소로 정확성, 신뢰성, 객관성, 최신성, 주제범위의 포괄성, 하이퍼링크 사항, 안전성, 소프트웨어 요구사항 등을 제시하고 있다. 해외의 선정지침을 조사해 보면 이러한 평가기준을 포함하고 있는 경우는 발견하지 못했으나 평가기준은 선정대상자원의 범위를 정하고, 그 범위 내에서 구체적인 자료를 선정할 때 각 자료를 평가하는 판단의 근거가 되므로 다음과 같은 평가기준 요소가 선정지침에 포함될 것을 제안한다.

즉, 웹 아카이빙을 위한 평가기준 요소에는 정확성 및 객관성, 포괄성, 최신성, 저작성 및 저작자의 평판, 안정성, 접근성, 희귀성 및 유일성 등이 포함되도록 한다.

5.8 선정대상자원

선정지침은 웹 아카이빙 대상 자원을 선정하는 담당자로 하여금 자원 선정 판단을 매우 용

이하게 하는 것을 지원해야 한다. 따라서 선정 대상자원에 대한 지침이 매우 구체적이고 명확해야 하며, 구체적인 사례가 포함되면 더 좋을 것이다.

현재 OASIS가 선정대상으로 삼고 있는 정보자원은 중앙정부가 생산한 온라인 자료, 대학 간행물, 회의자료, 전자저널, 최근 이슈가 되는 온라인 자료, 국내 웹 사이트 등이다. 이러한 선정대상 자료는 2006년에 수립된 것으로 그 이후의 많은 변화를 반영하고 있지 못하다. 해외 웹 아카이빙 사이트들의 경우 수집대상 자료에 대한 부분을 지속적으로 수정, 보완하고 있음을 알 수 있으며, 그 개정 내용이 홈페이지에 즉시 공시되고 있다.

또한 OASIS의 경우 국내 웹 사이트라는 제목으로 정부기관, 도서관, 대학교, 연구기관 등의 웹 사이트를 주기적으로 수집한다고 하고, ① 정부기관, 도서관, 대학교, 연구기관 등의 웹 사이트와 그 외의 한국과 관련된 보존가치가 있는 사이트를 주기적으로 수집하고, ② 정부기관, 도서관, 대학교, 연구기관 및 국내적으로 이슈가 되는 블로그나 트위터 등의 SNS(Social Network Service)도 선별적으로 수집한다고 하고 있다. 그러나 이 항목은 앞서 나오는 다섯 개의 항목과 중복되는 부분이 있다. 예를 들어 정부기관과 중앙부처의 구분, 대학교와 대학간행물 등의 구분이 어려울 수 있다. 따라서 '기타 국내 웹 사이트'라는 제목으로 수정하는 것이 바람직하며, 그 내용도 수정되어야 할 것이다.

이에 본 연구진은 국내외 연구결과 및 사례의 분석결과를 기반으로 선정대상자원을 제안함에 있어, 정보제공기관의 성격에 따라 크게 구분하여 선정대상자원을 포함시키고 기타 이

범주에 포함되지 않는 예외 대상을 추가적으로 기술하고자 하였다. 정보제공기관의 성격에 따라 웹 아카이빙 대상을 교육기관, 영리목적의 기관, 비영리기관, 정부기관으로 구분한 예도 있다(이지연, 김성연 2002). 즉 정부기관, 교육 및 연구기관, 도서관, 상업기관 등으로 크게 구분하고, 정부기관 자료는 중앙정부기관 및 그 산하부처가 포함되도록 하고, 교육 및 연구기관에는 대학이나 각종 연구소 등이 포함되도록 하되 상업기관에 소속된 연구소를 포함하도록 한다. 상업기관은 영리기관 중 한국의 역사, 사회, 문화에 지대한 영향을 미친 기관의 웹 사이트를 포함하도록 하였다. 도서관은 정보유통에 가장 큰 목적을 가지고 있는 기관으로서 모든 유형의 도서관을 아카이빙 함으로써 정보유통 동향을 후세에 전달해 줄 수 있기 때문에 포함할 것을 제안한다.

5.8.1 정부기관이 생산한 자료 및 웹 사이트

정부기관이 생산한 온라인 자료 및 웹 사이트를 수집해야 한다. 현재 OASIS도 중앙행정기관과 그 산하기관의 온라인 자료를 수집한다고 하고 있다. 영국 National Archives의 경우 중앙정부 웹 사이트를 아카이빙하고, 미국 MINERVA는 의원 및 의원위원회의 웹 사이트를 아카이브하고 있으며, 호주 PANDORA의 경우 국가적으로 중요한 또는 각 주에 관련된 중요한 정보를 수집하고 있다. 따라서 OASIS는 정부기관이 생산한 자료 및 웹 사이트를 수집함에 있어 다음과 같은 자원을 수집대상으로 포함할 것을 고려해야 한다.

첫째, 중앙정부 및 지방정부, 그리고 그 산하기관의 웹 사이트들을 수집해야 한다. 또한 그

웹 사이트에 게재되어 있는 온라인자료들도 수집해야 한다.

둘째, 중앙정부 및 지방정부, 그리고 그 산하 기관에서 생산하여 온라인으로 유통시키고 있는 자료들을 수집해야 한다. 이러한 자료는 해당 기관의 웹 사이트에 게재될 수도 있고, 새로운 웹 사이트를 개발하여 유통시키고 있을 수도 있다.

셋째, 중앙정부 및 지방정부, 그리고 그 산하 기관과 관련된 국가적으로 중요한 이슈가 되는 웹 정보자원을 수집해야 한다.

5.8.2 교육기관 및 연구기관의 웹 사이트 및 자료

OASIS는 교육기관 및 연구기관이 생산한 자료 및 이 기관들의 웹 사이트를 아카이빙 대상으로 포함시켜야 한다. 현재 OASIS는 대학 간행물에 대해 아카이빙을 하고 있으며, 국내의 저명한 대학 교수나 대학 연구원에 의해서 쓰여진 온라인자료를 우선적으로 수집한다고 하고 있다. 그러나 대학간행물의 범위를 어디까지로 보아야 할지에 대한 고민이 필요하다. 해외 사례를 보았을 때는 대학 및 연구기관의 웹 사이트를 아카이빙 한다고 명시한 경우는 발견하지 못했으나 포괄적 수집을 하는 나라들은 모두 교육기관 및 연구기관을 대상으로 웹 아카이빙을 한다고 보아야 한다.

교육기관 및 연구기관이 생산한 자료 및 이 기관들의 웹 사이트는 학술적 가치가 있는 사이트로 연구 및 교육의 역사를 보여줄 뿐만 아니라 가치 있는 많은 연구 결과를 웹 사이트에만 출판하는 경우가 많기 때문에 아카이빙 대상으로 포함시키는 것이 권장된다. 따라서 교

육기관 및 연구기관이 생산한 자료 및 웹 사이트를 수집함에 있어 다음과 같은 자원을 수집 대상으로 포함할 것을 고려해야 한다.

첫째, 교육기관 및 연구기관의 웹 사이트를 웹 아카이빙 대상으로 포함하되, 이 기관에서 웹으로 유통시키는 자료들을 집중적으로 아카이빙 한다.

둘째, 국내의 저명한 교수나 연구자에 의해서 쓰인 온라인자료를 수집한다.

셋째, 교내 행정적인 사무에 대한 온라인자료는 수집대상에서 제외하도록 한다.

5.8.3 도서관 웹 사이트

국내의 사례를 조사해 보았을 때 도서관 사이트를 아카이빙 대상으로 포함시키는 경우는 없었다. 도서관을 웹 아카이빙 대상 기관으로 포함시키느냐에 대해서는 논란의 여지가 있을 수 있다. 그러나 도서관은 정보유통에 가장 큰 목적을 가지고 있는 기관으로서 모든 유형의 도서관을 아카이빙 함으로써 정보유통 트렌드를 후세에 전달해 줄 수 있을 것으로 보인다. 다만, 어느 정도까지를 대상으로 해야 하느냐는 또 다른 논의가 필요하다. 따라서 도서관이 생산한 자료 및 웹 사이트를 수집함에 있어 다음과 같은 자원을 수집대상으로 포함할 것을 고려해야 한다.

첫째, 아카이빙 대상 도서관의 유형은 대학 도서관, 공공도서관, 전문도서관으로부터 출발하며, 이후 민간이 운영하는 도서관을 포함하여 모든 유형의 도서관 웹 사이트로 확장시키도록 한다.

둘째, 도서관이 소장하고 있는 자원은 주로 출판된 자료로, 이 자료에 대한 아카이빙은

지양하고, 주로 스냅샷을 이용한 홈페이지 수집방식을 적용하는 것이 바람직하다고 본다.

5.8.4 상업기관의 웹 사이트

공공영역뿐만 아니라 영리목적을 가진 상업기관의 웹 사이트도 선별적으로 수집하도록 해야 한다. 상업기관 중 세계적인 기업의 웹 사이트, 역사적 및 사회적, 그리고 문화적 영향력이 있는 웹 사이트를 아카이빙 대상으로 포함시킴으로써 공공영역 뿐만 사적 영역에서도 선별적으로 아카이빙을 해야 한다.

5.8.5 회의자료

현재 OASIS는 국내외적으로 권위 있고 규모가 큰 회의로부터 발간되는 자료를 망라적으로 수집하고 있으며, 정부기관, 전문 협회나 기관 그리고 대학교에서 개최되는 회의자료를 우선적으로 수집하고 있다. OASIS는 현재 아카이빙 대상 국제회의는 국제협회연합(Union of International Associations)의 기준인 전체 참가자 수 300명 이상, 참가자중 40% 이상 외국인, 참가국수 5개국 이상, 회의기간 3일 이상을 기준으로 선정하고 있다. 다만 국내 회의의 경우 그 기준에 대해 명확히 제시하지 못하고 있다. 수집된 회의자료의 신뢰성 및 가치를 높이기 위해서는 다음과 같은 기준을 명확히 적용하여 아카이빙을 하도록 해야 한다.

첫째, 정부기관, 전문 협회나 기관 그리고 대학교에서 개최되는 회의로부터 발간되는 자료를 망라적으로 수집하도록 하며, 회의 개최를 위해 개발된 웹 사이트도 수집대상으로 포함시킨다.

둘째, 국제회의는 국내에서 개최되는 경우로

제한하며, 국제협회연합(Union of International Associations)의 기준인 전체 참가자 수 300명 이상, 참가자 중 40% 이상 외국인, 참가국 수 5개국 이상, 회의기간 3일 이상을 기준으로 아카이빙 대상을 포함한다.

셋째, 국내 회의의 경우 학회 및 기관 차원에서 수행되는 회의의 자료를 대상으로 한다.

넷째, 아카이빙을 위한 회의의 목록을 개발하고, 이를 기준으로 아카이빙 하도록 한다.

5.8.6 전자저널

웹 아카이빙 대상 자료에 전자저널을 포함시켜야 한다고 주장하는 연구는 다수이며, 해외의 많은 웹 아카이빙 사이트가 수집대상으로 전자저널을 포함하고 있다. 전자저널을 아카이빙해야 하는 이유는 전자저널은 특정 이용 조건 하에서 특정 이용자에게 특정 기간 동안만 이용하게 함으로써 영구적 보존 및 접근의 한계가 있기 때문이다(최호남, 이응봉 2005). 문제는 수많은 전자저널 중 어떤 저널들을 수집 대상으로 하느냐이다. 최호남과 이응봉 연구자는 해외 전자저널의 디지털 아카이브 구축전략에 관한 연구에서 구축 모델링 및 콘텐츠 전략과 기술전략으로 구분하여 상세히 제안하고 있다. 특히 아카이빙 우선순위 도출을 위한 저널 평가 기준과 우선협상 대상 출판사 선정을 위한 판단기준을 수립해야 한다고 주장하고, 자료의 소유자인 출판사와 협상전략을 개발할 것을 제안했다. 아카이빙 대상으로 전자저널을 선정함에 있어 다음과 같은 구체적인 기준을 적용하도록 한다.

첫째, 오프라인으로 출판되지 않고 온라인으로만 출판되는 전자저널로 제한한다.

둘째, 같은 분야의 연구자 심사 평가를 거침으로써 논문의 높은 수준과 질을 보장받는 온라인 저널을 수집한다.

셋째, 국제성과 지역성에 역점을 두어 편집인, 심사위원, 저자들이 국제적인 명성과 다양성이 있는지를 권위 있는 기관으로부터 평가받은 저널을 수집하되, 다른 연구자에 의해서 논문 인용도가 높은 저널을 수집한다.

5.8.7 최근 이슈가 되는 온라인 자료

최근 이슈가 되고 있는 온라인 자료를 아카이빙 대상으로 포함시키는 해외 사례를 많이 발견할 수 있다. 최근 이슈 중 어떤 주제를 그 대상으로 할지는 아카이빙 기관의 정책에 따라 달라질 것이다. 미국 MINERVA는 선거와 관련된 웹 사이트, 올림픽, 이라크전쟁, 9/11추모 등 최근 이슈가 되는 것들을 아카이빙하고 있고, 덴마크 Netarchive.dk는 특별한 이벤트와 관련된 웹 사이트로서 덴마크 역사에 영향을 줄 수 있는 사건의 웹 사이트를 아카이빙 한다. OASIS도 최근 이슈가 되는 온라인 자료를 수집하고 있으며, 다음과 같이 구체적으로 선정 지침을 보완할 것을 제안한다.

첫째, 대통령 선거, 국회의원 선거를 포함한 지방자치선거, 신행정수도, 올림픽, 월드컵 등 국내적으로 이슈가 되는 온라인자료 및 웹 사이트를 수집하도록 한다.

둘째, 한국의 역사, 문화, 사회에 지대한 영향을 미칠 수 있는 사건에 관련된 온라인 자료 및 웹 사이트를 수집한다.

셋째 국내적으로 이슈가 되는 블로그나 트위터 등의 SNS(Social Network Service)도 선별적 수집을 고려한다.

넷째, 출처가 불분명하거나 포함하고 있는 내용이 불충분한 디지털자원은 수집대상에서 제외한다.

5.9 선정제외자료

웹 아카이빙 사이트마다 아카이빙 제외 대상 자료에 있어서 차이를 보이고 있다. PANDORA의 경우 ① 홍보 및 광고사이트, ② 인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트(디렉토리 사이트나 포털 사이트), ③ 공공에게 이용될 수 없는 자료들(인트라넷의 자료, 사내자료), ④ 진행 중인 작업 문서의 초안 등의 자료는 수집대상에서 제외하고 있다.

PANDORA의 파트너들이 제시하는 제외 대상은 약간씩 다르지만, 호주의 국립도서관인 NLA의 제외 대상에 모두 포함된다. NLA에서 지정한 제외 대상 목록은 총 13개 범주이다.

- ① 방송을 위해 디지털 이미지를 업로드 한 웹 카메라의 내용
- ② 데이터세트
- ③ 토론 리스트, 채팅방, 게시판, 뉴스 그룹
- ④ 초안이나 진행 중인 작업
- ⑤ 게임
- ⑥ 개인 문서나 아티클
- ⑦ 뉴스 사이트
- ⑧ 프린트 버전이 존재하는 온라인 뉴스
- ⑨ 조직기록
- ⑩ 포털과 인터넷 정보를 조직화하는 목적으로만 제공되는 사이트
- ⑪ 광고 사이트
- ⑫ 다른 정보원의 편집 사이트나 원본 콘텐츠가 아닌 것

㉓ 논문

정보트러스트 어워드는 3가지의 경우는 수 상대상에서 제외하였다. ① 저작권 침해 소지가 있는 것으로 해당 사이트에 저작권이 없는 경우, ② 자료가 허술하거나 사전에 제시된 심사 기준에 비추어 상대적으로 가치가 떨어진다고 판단되는 경우, ③ 실질적인 자료를 담고 있지 않고 하이퍼링크 목록만 제공하는 경우이다.

OASIS는 ① 블로그 사이트, ② CAMS(LAN 전용선이 카메라에 직접 연결되어 실시간 이미지를 인터넷에 전송하여 세계 어느 곳에서든지 볼 수 있도록 한 사이트), ③ 토론리스트, 채팅 사이트, ④ 게시판과 뉴스 그룹, ⑤ 뉴스 사이트, ⑥ 인쇄물이 존재하는 온라인 일간 신문, ⑦ 포털 사이트 등의 디지털자원은 수집대상에서 제외한다.

위에서 웹 아카이빙 제외대상 자료를 검토한 결과 제외 대상 가이드라인이 비교적 포괄적이어서 실제로 현장에서 적용할 때 어려움이 있을 수 있다고 생각한다. 예를 들어 단지 '포털 사이트'라고 하지 않고 '인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트'라고 하고 예시로 '디렉토리 사이트 또는 포털 사이트 등'으로 제시함으로써 아카이빙 담당자의 판단을 지원하고 있다. 이와 같은 맥락에서 본 연구진은 OASIS가 현재의 가이드라인을 다음과 같이 보완할 것을 제안한다.

첫째, 인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트로서, 디렉토리 사이트나 포털 사이트 등은 제외하도록 한다. 이는 사이트 자체가 정보 자체를 제공하지 않고 정보원이 있는 사이트로의 링크정보만을 담고 있기 때문이다.

둘째, 다른 정보원의 편집 사이트나 원본 콘텐츠가 아닌 사이트는 제외하도록 한다. 이는 포털 사이트와는 약간 다른 개념으로 단순히 링크를 재조직하는 것에서 좀 더 나아가 특정 주제에 대해 편집된 내용 등을 수록한 사이트이다.

셋째, 공공에게 이용될 수 없는 자료들이 있는 사이트로서, 인트라넷의 자료, 사내자료 등은 제외한다. 이와 같은 자료들에는 웹 로봇이 접근할 수 없기 때문이다.

넷째, 웹에 자신의 생각을 기사처럼 쓰고 게재하는 웹 사이트로서, 블로그 사이트 등은 제외하도록 한다. 이는 일부 블로그 사이트의 경우 연구가치가 있는 풍부한 자료를 제공할 수 있다고 할 수 있으나, 대부분의 경우 개인의 생각이나 경험을 쓰는 경우가 많기 때문에 대부분의 아카이빙 사이트에서 제외시키고 있다.

다섯째, 상품이나 서비스를 알리는데 목적이 있는 사이트로서, 홍보 및 광고 사이트는 제외하도록 한다.

여섯째, 계속 중이거나 완성되지 않은 사이트로서, 초안이나 진행 중인 작업결과를 제공하는 사이트는 제외하도록 한다. 이들 자료들은 완성이 되면 다른 사이트에 출판될 확률이 높기 때문이다.

일곱째, 게시판이나 채팅사이트로서, 토론리스트나 뉴스그룹 사이트는 제외한다. 이러한 사이트들은 대부분 정보적 가치가 없는 엔터테인먼트적 성격을 띠는 경우가 많기 때문이다. 게임 사이트는 오락적 기능이 강한 사이트로서 제외하도록 한다.

여덟째, 인쇄물이 존재하는 온라인 자료는 제외하도록 한다. 즉 인쇄물이 존재하는 온라인

인 일간 신문이나 뉴스 사이트 등은 선정대상에서 제외함으로써 중복적 아카이빙이 되지 않도록 한다.

아홉째, 방송을 위해 디지털 이미지를 업로드한 웹 카메라의 내용이 있는 사이트와 LAN 전용선이 카메라에 직접 연결되어 실시간 이미지를 인터넷에 전송하여 세계 어느 곳에서든지 볼 수 있도록 한 사이트(생방송 사이트 등)는 제외하도록 한다.

열째, PANDORA의 경우 논문이나 데이터 세트, 개인 문서나 아티클은 제외하도록 하고 있으며, OASIS도 이러한 기준을 적용하도록 해야 한다. 이는 개인논문이나 데이터 세트 등은 일반적으로 오프라인으로 출판되어 유통되고 있을 가능성이 높거나, 아니면 평가절차를 거치지 않아 신뢰성에 문제가 있을 수 있기 때문이다.

6. 논의 및 향후연구

본 연구는 웹 아카이빙을 위한 선정지침에 대한 개선안을 제안하는 것이다. 많은 연구자들이 웹 아카이빙의 방향성을 제시하는 연구를 수행하여 왔으며, 이러한 연구결과를 기반으로 우리나라의 아카이빙 정책이나 아카이빙 사업은 지속적으로 발전하여 왔다고 할 수 있다. 그러나 역사적, 사회적, 문화적 가치가 있는 아카이빙 대상 자원을 선별하기 위한 구체적인 선정지침의 개발에 대한 연구는 거의 없는 것으로 조사되었다. 이에 본 연구에서는 웹 아카이빙 선정지침을 개선할 수 있는 방안을 모색했다. 또한 온라인 자료 납본법에 대한 것, 온라인 자료 납

본이 수집방식에 미치는 영향, 그리고 선정대상 자료 목록 개발에 대해서도 논의하였다.

첫째, 해외 웹 아카이빙 사이트들은 OASIS와 비교하여 명확한 웹 아카이빙 정책을 가지고 있는지, 그리고 수집 대상 자원을 명확히 제시하고 있는지를 분석하였을 때, 국내외 모든 웹 아카이빙 사이트들은 홈페이지를 통해 웹 아카이빙 정책과 수집 대상 자원을 제시하고 있는 것으로 분석되었다. 그러나 해외 웹 아카이빙 사이트들은 OASIS와 비교하여 수집 제외 자원은 상당히 다른 것으로 나타났으며, 이에 본 연구에서는 어떠한 수집 대상 자원이 OASIS의 선정지침에 포함되어야 하는지와 어떤 자원이 수집 제외 자원이 되어야 하는지를 앞 절에서 제안하였다.

둘째, 웹 아카이빙을 하는데 있어서 가장 문제가 되는 것은 온라인자료에 대한 납본법이 제정되지 않고 있다는 것이다. 따라서 현재 OASIS는 심층웹을 수집함에 정부기관, 학술단체, 연구기관 등을 포함한 정보제공 기관들로부터 일일이 정보제공동의서를 받는 절차를 거쳐야 한다. 해외 많은 나라에서 웹 아카이빙 프로젝트를 위해 자발적 납본을 포함하여 온라인 자료를 납본의 대상으로 포함하고 있으며, 많은 연구자(서혜란 2004; 최재황, 광승진, 김정택 2009; 윤희운 2002; 오상훈, 최영선 2007)가 이미 웹 아카이빙을 위한 온라인 자료 납본의 필요성에 대해서 주장해 왔다. 특히 최재황, 광승진, 김정택(2009)은 온라인 디지털자료의 납본 체계 및 이용에 대한 가이드라인을 ① 출판지, ② 납본 대상 자료의 종류와 우선순위, ③ 납본 제외대상 자료의 유형, ④ 납본 포맷, 납본 시점, 납본 부수, ⑤ 납본 자료의 수집, 증표 교부,

⑥ 납본 주체, ⑦ 납본 기관과 의무, ⑧ 납본 주체의 납본거절 및 납본된 자료의 이용거절, ⑨ 납본된 자료의 갱신, 폐기, ⑩ 납본된 자료의 하드카피 제공, ⑪ 납본된 자료의 다운로드와 저장, ⑫ 납본된 자료의 보존을 위한 복사, ⑬ 납본의 효력 날짜, ⑭ 납본 자료의 이용, ⑮ 납본의 보상, ⑯ 납본 불이행에 대한 제재 등 16개 항목으로 나누어 구체적으로 제시하고 있는 것이 주목할 만하다.

셋째, 웹 아카이빙 정책 중 수집방법으로 선택적 수집방법과 포괄적 수집방법이 있는데, 이러한 수집방법을 선택하게 된 것이 디지털 납본법에 온라인 자료의 납본 포함여부와 관련이 있는지 간단히 분석해 보았다. <표 1>에서 보는 바와 같이 온라인자료 납본이 적용되고

있는 나라는 영국, 덴마크, 프랑스, 스웨덴, 캐나다, 독일이며, 이중 포괄적 수집을 하고 있는 나라는 덴마크, 프랑스, 스웨덴이다. 이 세 나라가 포괄적 수집정책을 펼칠 수 있었던 것은 온라인 자료에 대한 납본법이 제정되어 있기 때문이다.

어떠한 수집방식을 선택하느냐는 웹 아카이빙 기관의 정책에 달려 있다. 그러나 온라인자료의 납본이 법으로 제정되어 있지 않다면, 포괄적 수집방식을 채택하는 것은 불가능하다.

넷째, 본 연구에서는 선정지침을 개발하였으나, 더 나아가서 선정대상자료에 대한 목록데이터베이스를 따로 구축하여 운영하는 것이 바람직하다고 본다. 예를 들어 선정지침의 수집 대상으로 회의자료 부분에서 '국내외적으로 권

<표 2> 온라인자료 납본과 수집방식의 관련성

국가	사업명	수집방법	디지털 납본법명(연도)	온라인자료의 납본적용유무
한국	OASIS	선택	온라인디지털자료 납본 및 이용에 관한 법률(안)(2007)	-
호주	PANDORA	선택	개정안 제출 혹은 검토 중(곽승진 2008)	-
영국	TNA	선택	Legal Deposit Libraries Act(2003)	온라인 출판물을 납본 대상에 포함
미국	MINERVA	선택	개정안 제출 혹은 검토 중(곽승진 2008)	-
덴마크	Netarchve.dk	선택/포괄	Act on Legal Deposit of Published Material(2004)	전자 커뮤니케이션 네트워크 형태의 저작물을 납본 대상에 포함
프랑스	BnF Web archiving initiative	선택/포괄	Droit d'auteur et droits voisins dans la société de l'information(2006)	웹페이지를 납본 대상에 포함
스웨덴	Kulturarw3	포괄	The Legal Deposit Act(1993)	웹 아카이빙을 범제화하는 특별 법령을 선포
일본	WARP	선택	개정안 제출 혹은 검토 중(곽승진 2008)	-
캐나다	EPPP	선택	Library and Archives of Canada Act(2004)	도서, 잡지, 연례보고서, 연구 논문, 학술 저널 등의 온라인 출판물을 납본 대상에 포함
독일	DAHCS	선택	Gesetz über Die Deutsche Nationalbibliothek(2006)	온라인출판물을 납본대상에 포함
노르웨이	PARADIGMA	선택	The Norwegian Legal Deposit Act(1989)	온라인자료를 납본대상에 포함

위 있고 규모가 큰 회의'로부터 발간되는 자료를 수집한다고 하고 있으나 그와 같은 회의가 어떤 회의인지를 그 리스트를 개발하고 운영하는 것을 권장한다. 더불어 OASIS 온라인자료 선정위원회로 하여금 이에 대한 목록을 제시해 주도록 자문을 받는 것을 권장한다. 영국 TNA의 경우 영국 정부 기관의 웹 사이트 수집 대상과 수집 주기에 대한 목록을 관리하고 공개하고 있다. 또한 덴마크는 1년에 4회, 덴마크의 전체 웹 사이트를 대상으로 스냅샷을 진행하는데, 덴마크의 전체 도메인을 수집하기 위해 수집해야 할 도메인의 목록을 확보한다. 이를 위해 법률을 제정하고 국가 도메인인 dk로부터 dk의 TLD(최상위도메인) 아래에 등록된 모든 도메인 목록을 제공 받는다. OASIS는 수집정책이 선택적이기 때문에 이와는 다른 방법으로 수집 대상 자료에 대한 목록 DB를 반드시 구축해야 할 것이며, 이를 홈페이지에 공개하는 것을 권장한다.

향후 연구로서 본 연구진은 협력형 아카이빙 정책에 대해 제안한다. 국가기관일지라도 가치 있는 모든 웹 사이트를 아카이빙하는 것은 불가능하다고 생각한다. 따라서 기관 간 협력을 통해서 아카이빙 정책을 펼치는 것이 바람직하다. 미국 MINERVA는 미국의 버지니아텍과 협력하여 일본 지진에 관한 웹 아카이빙을 구축했고, 프랑스 국립도서관과 협력하여 프랑스의 특별한 이벤트에 대한 웹 사이트를 아카이빙 했다. 호주 PANDORA도 국립도서관이나 기타 여러 기관들과 협력하여 아카이빙을 수행하고 있다. 따라서 OASIS도 자체적으로 웹 아카이빙을 할 뿐만 아니라 국내외 다른 기관들과 적극적인 협력을 통해 웹 아카이빙을 할 필

요가 있으며, 협력 모형에 대한 연구는 추가적으로 수행될 필요가 있다고 본다.

7. 결 론

웹상에서 유통되는 정보자원의 양은 기하급수적으로 증가할 뿐만 아니라, 역사적 및 문화적으로 가치 있는 정보자원이 상당히 많이 있는 것으로 평가되고 있다. 또한 최근에는 오프라인으로는 전혀 출판되지 않고 온라인으로만 출판되는 자료들이 증가하고 있으며, 온라인 정보자원에 대한 의존도도 높아지고 있다. 온라인으로 출판되어 웹상에서 유통되는 정보자원들의 문제는 오프라인 자료처럼 도서관에 소장되는 것이 아니라 웹 사이트의 유지종료, 기술적인 문제, 관심사의 변경 등의 이유로 그 역사적 가치 및 사회적 가치에도 불구하고 일순간에 사라져 버릴 수 있다는 것이다. 이에 세계 많은 나라에서는 국가적인 차원에서 웹 아카이빙을 수행하여 왔으며, 국내의 경우 대표적으로 OASIS가 국가적인 차원에서 프로젝트로 진행되었다.

국립중앙도서관은 OASIS사업을 통해 저작권 소유기관이나 개인으로부터 이용허락을 받아 아카이빙을 진행하고 있으며, 매년 새로운 아카이빙 대상 사이트를 발굴하여 자원을 확충하고 있다. 그러나 선정지침이 추상적이고 지금 개발되어 있는 지침이 2006년에 개발된 이래 새로운 변화나 요구를 반영하지 못하고 있는 상황이다.

본 연구에서는 세계적인 웹 아카이빙 사이트들을 집중적으로 조사하여 웹 아카이빙 정책이

나 웹 아카이빙 대상자원을 분석하였으며, 이를 기반으로 OASIS를 위한 온라인자료 선정 지침을 제안하였으며, 주요 연구성과는 다음과 같다.

첫째, 해외 웹 아카이빙 사이트나 OASIS 모두 명확한 웹 아카이빙 정책이나 수집대상자원에 대한 내용, 수집 제외자원에 대한 내용을 선정지침에 명확히 제시하고 있었으며, 각 기관마다 각각 다른 정책을 가지고 있다는 것을 알 수 있었다.

둘째, 본 연구에서는 OASIS 선정지침에 대한 개선안을 제안하고 있으며, 웹 자료에 대한 정의, 용어정의, 수집기본원칙, 수집방법, 수집주기, 구체적인 아카이빙대상 자원에 대해서 해외사례 분석결과를 기반으로 개선내용을 제안하였다.

셋째, 수집대상자원에 대해서 대폭적인 개선을 제안하였으며, 선정대상으로 포함시켜야 할 자원으로 정부기관, 교육 및 연구기관, 도서관, 상업기관 등에서 생산된 자원 및 웹사이트, 국내외적으로 권위 있고 규모가 큰 회의로부터 발간되는 회의자료, 온라인으로만 출판되는 전자저널, 최근 이슈가 되고 있는 온라인 자료 등을 아카이빙할 것을 제안하였다.

넷째, 웹 아카이빙 제외 대상으로 인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트, 다른 정보원의 편집 사이트나 원본 콘텐츠가 아닌 사이트, 공공에게 이용될 수 없는 자료들이 있는 사이트, 웹에 자신의 생각을 기사처럼 쓰고 게재하는 웹 사이트, 상품이나 서비스를 알리는데 목적이 있는 사이트 등 10개 유형을 제안하였다.

마지막으로 웹 자원 수집방법과 디지털 납본법과의 관련성, 선정대상자료에 대한 목록데이터베이스 구축의 필요성, 협력형 아카이빙 정책의 필요성에 대해서 논의하였다.

선행연구에서 살펴보았듯이 웹 아카이빙과 관련된 연구는 다수 수행되었다. 웹 아카이빙을 필요성과 개념정의, 선진 사례 및 해외 현황 소개, 법적·제도적 문제에 관한 논의, 기술적인 측면에 대한 연구, 디지털 납본법과 관련된 연구 등 다양한 각도에서 웹 아카이빙의 발전을 모색한 연구가 많이 수행되었다. 그러나 선정지침의 개선방향 모색과 관련된 연구는 거의 없고 본 연구가 국내에서 최초로 수행되었다는 점에서 연구의 의의가 있다고 보며, 이를 기반으로 보다 많은 연구자가 다양한 관점을 반영하여 선정지침을 개선시켜 나아가기를 바란다.

참 고 문 헌

기록관리학회. 2008. 『기록학 용어 사전』. 서울: 역사비평사.
김광영, 이원구, 윤화목, 신성호, 이민호. 2011. 웹 자원 아카이빙을 위한 웹 크롤러 연구

개발. 『한국콘텐츠학회논문지』, 11(9): 9-16.
김유승. 2008. 복합적 웹 아카이빙 정책에 관한 고찰 『한국문헌정보학회지』, 42(4): 159-179.

- 김유승. 2007. 웹 아카이빙의 법·제도적 문제에 대한 고찰. 『한국문헌정보학회지』, 41(3): 5-24.
- 김지훈. 2009. 국가도서관의 포지셔닝에 의한 하이브리드 자원 개발. 『한국비블리아학회지』, 20(2): 45-58.
- 김희정. 2010. 웹 아카이빙 인터페이스 유형 및 특성 분석. 『한국기록관리학회지』, 10(2): 147-170.
- 도쟁이&슬미. [online]. [cited 2012.7.13]. <<http://203.241.185.12>>.
- 서혜란. 2004. 웹 아카이빙의 성과와 과제. 『한국비블리아학회지』, 15(1): 5-22.
- 서혜란. 2003. 디지털자료의 납본과 보존을 위한 각 국가의 노력. 『情報管理學會誌』, 20(1): 373-399.
- 오상훈, 최영선. 2008. 디지털 자원의 웹 아카이빙을 위한 납본 프로세스 개발 및 기능 설계. 『情報管理學會誌』, 25(4): 5-23.
- 윤정옥. 2011. 웹 아카이브 OASIS의 현황에 관한 연구. 『정보관리연구』, 42(3): 95-116.
- 윤정옥. 2010. 웹 아카이브 OASIS에 관한 고찰. 『한국문헌정보학회지』, 44(2): 5-27.
- 이재선. 2008. 『국가기록원의 웹 사이트 아카이빙 구축을 위한 선별 방안에 관한 연구』. 석사학위논문. 이화여자대학교 정책과학대학원, 기록관리학전공.
- 이지연, 김성연. 2002. 디지털 도서관의 웹자원 기반 장서개발. 『한국기록관리학회지』, 3(1): 111-127.
- 최재황, 박승진, 김정택. 2009. 온라인 디지털자료의 납본체계 및 이용에 관한 연구. 『한국도서관·정보학회지』, 40(1): 209-232.
- 최호남, 이응봉. 2005. 해외 전자저널의 디지털 아카이브 구축 전략에 관한 연구. 『한국문헌정보학회지』, 39(2): 161-183.
- Alexander, J. and M. A. Tate. 1998. Web Resource Evaluation Techniques [online]. [cited 2012.7.13]. <<http://www.widener.edu/Media/Website%20Resources/docs/library/evalrevised2005.ppt>>.
- Arvidson, Allan, Krister Persson and Johan Mannerheim. 2001. "The Royal Swedish Web Archive: a 'Complete' Collection of Web Pages." *Journal of International Preservation News*, 26: 10-12.
- Brown, Adrian. 2006. *Archiving websites: a practical guide for information management professionals*. London: Facet publishing, 24-41.
- Frithch, J.W. and R. L. Cromwell. "Evaluating Internet Resources: Identity, Affiliation, and Cognitive Authority in a Networked World." *Journal of the American Society for Information Science and Technology*, 52(6): 499-507.
- Groke, Abigail, Jonesand and Gina. 2003. Web preservation projects at Library of Congress. [online]. [cited 2012.7.13]. <<http://www.archimuse.com/mw2003/papers/grotke/grotke.html>>.
- Masanés, J. 2006. *Web Archiving*. Berlin: Springer, 71-92.
- nationalarchives. [online]. [cited 2012.7.13]. <<http://www.nationalarchives.gov.uk>>.

- /archives-sector/collections-strategies.htm), <<http://www.nationalarchives.gov.uk/webarchive/other-collections.htm>>, <<http://www.nationalarchives.gov.uk/webarchive/information.htm>>.
- MINERVA. [online]. [cited 2012.7.13]. <<http://www.loc.gov/webarchiving/collections.html>>.
- netarchive.dk. [online]. [cited 2012.7.13]. <<http://netarkivet.dk/in-english/faq/#anchor6>>.
- OASIS. [online]. [cited 2012.7.13]. <http://www.oasis.go.kr/intro_new/intro_selguide.jsp>.
- PANDORA. [online]. [cited 2012.7.13]. <<http://pandora.nla.gov.au/guidelines.html>, http://pandora.nla.gov.au/guidelines/slwa_guidelines.doc>.

[부록 1] OASIS 웹 자료 선정지침(안)

1. 사업 추진 배경

국가 지식 자원의 수집과 보존 책임 기관인 국립중앙도서관은 국가대표도서관으로서 가치 있는 인터넷 자료를 국가적인 차원에서 수집·축적하여 미래 세대에 연구 자료로 제공하고 온라인 디지털 자원 수집 보존의 표준모델 연구 등에 기여한다.

2. 웹 자료의 정의

온라인상에서 유통되고 활용될 수 있는 모든 형태의 정보자원으로서, 인터넷 등을 통해 내려 받아 개인용 컴퓨터, 노트북 컴퓨터, PDA 등의 정보통신 단말기에서 읽을 수 있는 전자적으로 유통되는 텍스트, 소리, 동영상 정보 등을 담고 있는 컴퓨터 파일을 포함한다.

3. 용어 정의

- ① 온라인(Online)은 일반 PC에서 외부의 다른 시스템과 통신 회선이 전화에 연결된 상태를 뜻하는 것으로 시스템이 온라인 상태일 때 외부의 정보를 이용할 수 있다.
- ② 홈페이지(Home Page)는 URL을 사용해 특정 호스트에 접속했을 때 가장 먼저 나타나는 웹 페이지이다.
- ③ 사이트(Site)는 인터넷 상에 있는 서버를 가리키는 것으로, 정보가 필요할 때 언제든지 그것을 볼 수 있도록 웹 서버에 저장된 집합체이다.
- ④ 타이틀(Title)은 아카이빙을 위해 고려되고 있는 실체이다. 1개의 문서가 될 수도 있고, 인터넷 주소로 다수의 문서로 연결되기도 한다.
- ⑤ 아카이브(Archive)는 인터넷으로부터 다운로드하고, 국가 도서관 서버에 저장되는 현상의 의미로 사용된다.
- ⑥ 보존(Preserve)은 원래의 모습 그대로 보호하고 간수하여 남아 있게 하는 방법이다.

4. 수집 기본 원칙

- ① 수집대상 자원의 주제는 한국과 관련된 국가적 보존가치가 있는 것으로, 한국과 관련된 사회, 정치, 문화, 종교, 과학 또는 경제적으로 중요한 주제이고, 따라서 지리적 위치에 관계없이 한국

과 관계된 것으로 보존가치가 있는 것은 해외에서 생산된 것도 수집하도록 한다. 또한 한국의 저자에 의해 쓰여진 것이어야 하며, 국내의 저명한 대학 교수나 대학 연구원 등 권위 있는 한국 사람에 의해 쓰이거나 국내외적으로 해당 학문분야에 기여를 한 것이어야 한다.

- ② 수집대상 자료의 특성으로 현재 또는 미래의 정보요구에 봉사하는 유용성, 저작자의 평판, 제공된 정보의 유일성, 학술적 내용, 정보의 최신성, 업그레이드 빈도, 접근의 용이성 등을 고려한다.
- ③ 정치적 이슈, 사회적 이슈, 문화적 이슈, 그리고 국제적 이슈가 있을 경우 그 역사적 가치를 고려하여 수집한다.
- ④ 오리지널 온라인 디지털 자료는 망라적으로 수집하며, 오프라인 자료가 있는 온라인 자료는 수집을 지양한다.
- ⑤ 소위 가변성이 높은 동적 출판물은 업데이트 전후의 내용변화와 구성체제의 상당한 변화 등을 감안하여 선택하며, 온라인 디지털자료의 포맷이 여러 종류일 경우에는 표준포맷으로 구성된 자료를 우선적으로 선택한다.

5. 수집 방법

수집의 방법에는 포괄적(망라적) 수집과 선택적 수집이 있다. OASIS는 선택적 수집을 한다. 선택적 수집은 웹 자료의 품질 보장이 가능하고, 정보제공기관과의 저작권협의를 거친 경우 포괄적 수집으로 수집할 수 없는 사이트들도 수집 가능하며, 이러한 사이트들에 대해서는 영구적 보존 및 공공이용이 가능하다.

6. 수집 주기

OASIS는 평균 변화 시간을 기반으로 각 사이트의 수집 주기를 정한다. 즉, OASIS는 수집로봇이 지속적으로 등록된 사이트의 자료를 수집하여 변화 상태를 감시하며, 이전에 저장한 상태와 비교하여 변화 상태를 수치로 표현하면, 관리자가 변화 정도를 판단하여 현재 수집한 상태를 저장하여 보존하는 방식을 취하고 있다.

7. 평가기준

선정 대상 자원 중 하나 하나의 자료에 대해 평가하는 기준은 정확성 및 객관성, 포괄성, 최신성, 저작성 및 저작자의 평판, 안정성, 접근성, 희귀성 및 유일성 등의 요소를 기반으로 한다.

8. 선정 대상 자원

1) 정부기관이 생산한 자료 및 웹 사이트

- ① 중앙정부 및 지방정부, 그리고 그 산하기관의 웹 사이트들을 수집한다. 또한 그 웹 사이트에 게재되어 있는 온라인자료들도 수집한다.
- ② 중앙정부 및 지방정부, 그리고 그 산하기관에서 생산하여 온라인으로 유통시키고 있는 자료들을 수집한다. 이러한 자료는 해당 기관의 웹 사이트에 게재될 수도 있고, 새로운 웹 사이트를 개발하여 유통시키고 있을 수도 있다.
- ③ 중앙정부 및 지방정부, 그리고 그 산하기관과 관련된 국가적으로 중요한 이슈가 되는 웹 정보자원을 수집한다.

2) 교육기관 및 연구기관의 웹 사이트 및 자료

- ① 교육기관 및 연구기관의 웹 사이트를 웹 아카이빙 대상으로 포함하되, 이 기관에서 웹으로 유통시키는 자료들을 집중적으로 아카이빙 한다.
- ② 국내의 저명한 교수나 연구자에 의해서 쓰인 온라인자료를 수집한다.
- ③ 교내 행정적인 사무에 대한 온라인자료는 수집대상에서 제외한다.

3) 도서관 웹 사이트

- ① 아카이빙 대상 도서관의 유형은 대학도서관, 공공도서관, 전문도서관으로부터 출발하며, 이후 민간이 운영하는 도서관을 포함하여 모든 유형의 도서관 웹 사이트로 확장시키도록 한다.
- ② 도서관이 소장하고 있는 자료는 주로 출판된 자료로, 이 자료에 대한 아카이빙은 지양하고, 주로 스냅샷을 이용한 홈페이지 수집방식을 적용한다.

4) 상업기관의 웹 사이트

- ① 국내 상업기관 중 세계적인 기업의 웹 사이트를 수집한다.
- ② 역사적 및 사회적, 그리고 문화적 영향력이 있는 상업기관의 웹 사이트를 수집한다.

5) 회의자료

- ① 정부기관, 전문 협회나 기관 그리고 대학교에서 개최되는 회의로부터 발간되는 자료를 망라적으로 수집하며, 회의 개최를 위해 개발된 웹 사이트도 수집대상으로 포함시킨다.
- ② 국제회의는 국내에서 개최되는 경우로 제한하며, 국제협회연합(Union of International Associations)의 기준인 전체 참가자 수 300명이상, 참가자중 40%이상 외국인, 참가국수 5개국 이상, 회의기간 3일 이상을 기준으로 아카이빙 대상을 포함한다.

- ③ 국내 회의의 경우 학회 및 기관 차원에서 수행되는 회의의 자료를 대상으로 한다.
- ④ 아카이빙을 위한 회의의 목록을 개발하고, 이를 기준으로 아카이빙 한다.

6) 전자저널

- ① 오프라인으로 출판되지 않은 온라인으로 출판되는 전자저널로 제한한다.
- ② 같은 분야 연구자의 심사 평가를 거침으로써 논문의 높은 수준과 질을 보장받은 온라인 저널을 수집한다.
- ③ 국제성과 지역성에 역점을 두어 편집인, 심사위원, 저자들이 국제적인 명성과 다양성이 있는지를 권위 있는 기관으로부터 평가받은 저널을 수집하되, 다른 연구자에 의해서 논문 인용도가 높은 저널을 수집한다.

7) 최근 이슈가 되는 온라인 자료

- ① 대통령 선거, 국회의원 선거를 포함한 지방자치선거, 신행정수도, 올림픽, 월드컵 등 국내적으로 이슈가 되는 온라인자료 및 웹 사이트를 수집한다.
- ② 한국의 역사, 문화, 사회에 지대한 영향을 미칠 수 있는 사건에 관련된 온라인 자료 및 웹 사이트를 수집한다.
- ③ 국내적으로 이슈가 되는 블로그나 트위터 등의 SNS(Social Network Service) 도 선별적 수집을 고려한다.
- ④ 출처가 불분명하거나 포함하고 있는 내용이 불충분한 디지털자원은 수집대상에서 제외한다.

9. 선정 제외 자료

- ① 인터넷 정보를 조직해서 제공할 목적으로 개발된 사이트로서, 디렉토리 사이트나 포털 사이트 등은 제외하도록 한다. 이는 사이트 자체가 정보 자체를 제공하지 않고 정보원이 있는 사이트로의 링크 정보만을 담고 있기 때문이다.
- ② 다른 정보원의 편집 사이트나 원본 콘텐츠가 아닌 사이트는 제외한다. 이는 포털 사이트와는 약간 다른 개념으로 단순히 링크를 재조직하는 것에서 좀 더 나아가 특정 주제에 대해 편집된 내용 등을 수록한 사이트이다.
- ③ 공공에게 이용될 수 없는 자료들이 있는 사이트로서, 인트라넷의 자료, 사내자료 등은 제외한다. 이는 이와 같은 자료들에는 웹 로봇이 접근할 수 없기 때문이다.
- ④ 웹에 자신의 생각을 기사처럼 쓰고 게재하는 웹 사이트로서, 블로그 사이트 등은 제외한다. 이는 일부 블로그 사이트의 경우 연구가치가 있는 풍부한 자료를 제공할 수 있다고 할 수 있으나, 대부분의 경우 개인의 생각이나 경험을 쓰는 경우가 많기 때문이다.

- ⑤ 상품이나 서비스를 알리는데 목적이 있는 사이트로서, 홍보 및 광고 사이트는 제외한다.
- ⑥ 계속 중이거나 완성되지 않은 사이트로서, 초안이나 진행 중인 작업결과를 제공하는 사이트는 제외한다. 이들 자료들은 완성이 되면 다른 사이트에 출판될 확률이 높기 때문이다.
- ⑦ 게시판이나 채팅사이트로서, 토론리스트나 뉴스그룹 사이트는 제외한다. 이러한 사이트들은 대부분 정보적 가치가 없는 엔터테인먼트적 성격을 띠는 경우가 많기 때문이다. 게임 사이트 등도 오락적 기능이 강한 사이트로서 제외한다.
- ⑧ 인쇄물이 존재하는 온라인 자료는 제외한다. 즉 인쇄물이 존재하는 온라인 일간 신문이나 뉴스 사이트 등은 선정대상에서 제외함으로써 중복적 아카이빙이 되지 않도록 한다.
- ⑨ 방송을 위해 디지털 이미지를 업로드 한 웹 카메라의 내용이 있는 사이트, LAN 전용선이 카메라에 직접 연결되어 실시간 이미지를 인터넷에 전송하여 세계 어느 곳에서든지 볼 수 있도록 한 사이트(생방송 사이트 등) 등은 제외한다.
- ⑩ 논문이나 데이터 세트, 개인 문서나 아티클은 제외한다. 이는 개인논문이나 데이터 세트 등은 일반적으로 오프라인으로 출판되어 유통되고 있을 가능성이 높거나, 아니면 평가절차를 거치지 않아 신뢰성에 문제가 있을 수 있기 때문이다.

