DOI: 10.3745/KIPSTD.2009.16D.5.651

ISO 14721 OAIS 참조모형을 활용한 웹 아카이빙의 메타데이터 구조 및 요소 정의

오 상 훈[†] · 최 영 선[††]

요 약

본 연구에서는 웹 아카이빙에서 가치 있는 웹 자원의 수집, 관리 및 보존을 위해 요구되는 메타데이터의 구조를 설계하고 요소를 정의하였다. 본 연구를 위해 국립중앙도서관'OASIS'등의 웹 아카이빙에서 수집 자원의 장기 보존을 위해 활용되는 메타데이터를 조사하고, 웹 아카이빙의 각 프로세스 단계별 요구사항 및 웹 자원의 특성을 분석하였으며, 특히 장기 보존을 위한 아카이빙의 개념적 틀을 제공하는 ISO 14721 OAIS 참조모형을 기반으로 제안하였다. 또한 웹 아카이빙 간의 자원 공유를 위한 메타데이터의 상호운용성을 고려하였다. 그 결과 본 연구에서는 웹 아카이빙에서 자원의 체계적이고 효율적인 수집, 관리, 운영 및 보존을 위한 설명적, 구조적, 관리적 그리고 보존적 유형의 4개 메타데이터 구조를 설계하고 28개의 필수 메타데이터 요소를 정의하였다.

키워드 : 디지털 자원, 온라인 디지털 자원, 웹아카이빙, 메타데이터, 보존 메타데이터, OAIS 참조모형

A study on Designing Metadata Structure and Element on Web Archiving based on the ISO 14721 OAIS Reference Model

Oh Sang Hoon[†] · Choi Young Sun[††]

ABSTRACT

This study is to develope the structures and the elements of the metadata for harvesting, management and preservation of a valuable web resources in the web archiving. For this study, we investigated the available metadata in the web archiving and surveyed the requirements of web archiving process. And we analyzed the characteristics of web resources. Also, this study was used a based on the ISO 14721 OAIS Reference Model. Finally, to share the metadata elements among the web archiving system, this study considered the interoperability for the exchange of the metadata. Based on the result, this study designed four structures of the metadata and defined the 28 core metadata elements for the web archiving.

Keywords : Digital Resources, Online Digital Resources, Web Arhiving, Metadata, Preservation Metadata, OAIS Reference Model

1. 서 론

1.1 연구배경

오늘날 정보통신기술의 급속한 발달은 전통적인 지식정보자원의 생산·수집·관리·축적 방식에 변화를 가져왔다. 가치 있는 지적 자원들이 다양한 형태의 디지털 자원으로 생산되고 있으며, 기존 자원들도 디지털로 재생산 되고 있다. 그러나 디지털 자원들은 태생적으로 물리적인 형태가 없으며, 가변적인 성질을 갖고 있어 빠르게 소멸될 수도 있다는 문제점이 있다. 이에 2003년 10월 유네스코에서는 이러한 디지털 자원의 수집과 보존을 위해 인터넷 지식자원의 보존 및 이용에 관한 내용을"디지털 유산 보존 헌장"에서 천명하였으며[1], 세계 각국에서는 웹 아카이빙을 운영하며 온라인 자원을 수집하고, 보존하기 위한 노력을 하고 있다. 국립중앙도서관에서도 가치 있는 인터넷 자료(디지털 자원)를 국가적인 차원에서 수집·보존하고자 OASIS(Online Archiving & Searching Internet Sources, 이하 OASIS)프로젝트를 2001년부터 추진해 오고 있다[2].

웹 아카이빙은 쉽게 사라지는 인터넷 상의 진본의 자원을

※ 본 논문은 2007년도 국립중앙도서관 「OASIS 표준화연구」의 지원으로 수행됨.
† 정 회 원 : (사)한국디지털콘텐츠산업협회 사무국장
†† 정 회 원 : (사)한국디지털콘텐츠산업협회 연구원
논문접수 : 2008년 10월 16일
수 정 일 : 1차 2009년 6월 5일, 2차 2009년 7월 29일
심사완료 : 2009년 8월 10일

효과적으로 수집·관리하고 장기간 보존하여 미래 세대에 전달하는 목적을 갖고 있다. 이에 디지털 자원을 기술(Description)하고, 체계적으로 관리하며 이용자들에게 효율적인 검색을 제공해 주기 위해서는 전자정보, 네트워크 정보를 위한 메타데이터가 필요하게 되었다.

특히 웹 자원은 표현하는 수단과 종류가 다양하여 이들을 체계적이고 효율적으로 관리하기 위해서는 각 매체별·유형별 특성화된 관리요소가 있어야 한다. 더불어 웹 아카이빙에서는 자원을 수집·관리 할 뿐만이 아니라 미래의 이용자들에게 현재의 가치 있는 자원을 전달해 주기 위한 보존단계별 기능을 포함한 전체적인 메타데이터 구조와 요소가 제시되어야 한다.

따라서 본 논문에서는 OASIS 웹 아카이빙의 주요 수집 자원인 웹사이트, 웹 문서, 웹 자원(개별 파일)을 대상으로 관리하고 보존하기 위한 단계별·기능별 요구사항을 충족 할 수 있는 메타데이터 구조와 요소를 개발하였다.

1.2 연구목적 및 수행 방법

본 연구는 향후 디지털 자원의 웹 아카이빙을 위해 내·외부에 있는 디지털 자원을 보존하기 위한 주체와 대상 자원의 체계적인 수집, 관리, 저장 및 서비스를 위한 체계적인 관리방안을 제시하고자 하였다.

따라서 OASIS 프로젝트의 현황 분석과 웹 아카이빙 관련 해외 대표사례를 조사 및 분석하였다. 그 결과를 기반으로 OASIS 및 웹 아카이빙에 적용 가능한 메타데이터를 개발하기 위해 ISO 14721 OAIS 참조모형에서 제시하는 웹 아카이빙 업무단계별, 기능별 프로세스에 따른 구조와 요소를 적용하고자 하였다.

2. 웹 아카이빙 메타데이터 사례연구

2.1 메타데이터 개요

메타데이터는 데이터에 대한 데이터로서, 실제 콘텐츠는 아니면서 그에 대한 각종 정보를 갖고 있는 데이터를 의미하며, 일반적으로 다음의 2가지 기능을 갖는다. 첫째는 정보검색을 지원하는 기능으로 정제된 정보를 제공하여 자원의 식별과 이용에 도움을 줄 뿐만 아니라 자원에 가치를 부가하여 자원의 검색에 유용하게 한다. 둘째는 자원을 체계적으로 관리하고 효과적인 이용을 지원하기 위해 자원을 기술하고, 범주화하며 자원의 연혁(history) 정보를 기록하는 것이다[3].

웹 아카이빙을 위한 메타데이터는 추가적으로 보존 정보를 지원하기 위한 기능이 요구된다. 보존정보는 자원의 생명주기 전 과정에 대한 모든 정보를 포함하고 있어야 한다. 이는 데이터 객체에 관한 정보 뿐 아니라 그 자원을 관리하기 위해 요구되는 소프트웨어와 하드웨어 등의 정보 그리고 자원 관리에 관련된 관리자에 관한 정보 등 자원 보존 활동에 관련된 모든 정보를 구체적으로 표현해야한다. 이러한 역할을 하는 메타데이터를 보존 메타데이터라고 한다[4]. 웹 아카이빙에서 보존메타데이터란 "장기적으로 전자 기록들이 환경이 변하여도 계속 활용될 수 있게 하는 생존 능력(Viability), 이용자 또는 관리자의 요구에 따라 처리 및 표기될 수 있는 능력(Renderability), 관리자에 의하여 기록된 정보들을 식별할 수 있게 해주는 능력(Understandability)을 유지하기 위해 필요한 지식 정보"로 정의된다[5].

2.2 웹 아카이빙 메타데이터 사례분석

2.2.1 Dublin Core [6]

더블린 코어(Dublin Core) 메타데이터는 데이터의 호환성을 유지하고 네트워크 자원의 기술에 필요한 일련의 데이터 요소를 규정하여 이들 자원의 신속한 검색을 목적으로 1995년 OCLC와 NCSA(National Center for Supercomputer Application)가 더블린에서 개최된 워크숍에서 합의되었다. Dublin Core Metadata Element Set, Version 1.1이 발표되었으며, ISO Standard 15836-2003 (February 2003)과 ISO Standard Z39.85-2007 (May 2007)로 채택되었다. 더블린 코어 메타데이터는 메타데이터 표준 개발에 폭넓게 이용되는 국제 표준 메타데이터로 보편적이고 단순한 메타데이터 요소를 구성하여 전문가와 비전문가 모두에게 이용 된다. 또한 고유성, 확장성, 구문독립성, 선택성, 반복성, 수정가능성 원칙을 근거로 15개 메타데이터 요소를 제안하였다. 특히 전자도서관 및 초기 웹 아카이빙에서 자원을 수집 및 관리하기 위해 메타데이터 개발 시 기준이 되었다.

2.2.2 OASIS [2]

OASIS는 2001년 11월 '온라인저작물 수집·보존 시스템 구축 ISP 및 시범시스템 개발'을 시작으로 2005년 웹기반 OASIS 시스템 확장 개선 사업을 완료 하였으며, 2006년 OASIS 홈페이지를 통해 대국민 서비스를 실시하고 있는 우리나라를 대표하는 웹 아카이빙이다. OASIS는 수집된 자원의 메타데이터를 기술하기 위하여 디지털 정보자원에 대한 국제 데이터 기술 표준화 형식인 더블린 코어 메타데이터 요소와 OASIS 수집 자원 관리를 위한 요소를 추가적으로 선정하여 사용하고 있다. 특히 OASIS 프로젝트의 정책, 시스템 변화 및 업무 변동 상황에 따라 메타데이터의 확장 및 변경·수정이 가능하도록 구성하였다. OASIS 메타데이터는 더블린 코어의 15개 요소(Elements)를 토대로 내부 관리 메타데이터 요소를 추가하여 총 32개의 메타데이터 요소(하위요소 포함)로 구성되어 있다.

2.2.3 OCLC Digital Archive Metadata [7]

디지털 아카이브 관련 연구를 지속적으로 진행하고 있는 OCLC(Online Computer Library Center)는 디지털 아카이브 시스템 가이드에서 디지털 아카이브에서의 메타데이터 요소를 정의하였다. 여기에는 더블린 코어에 기반을 둔 요소와 자원의 관리 및 기술(技術)을 위한 메타데이터 요소를 포함하고 있으며 웹 아카이빙 운영을 위해 필요한 관리 요소들이 자세히 표현되었고, 특히 이용자에게 서비스하기 위한

부분까지 고려하였다. 2002년 초판 이후 2003년 4월 자원의 생성과 입수 전 과정에 대한 요소가 삭제되고 자원의 생명주기에 관한 이벤트와 자원의 관계정보가 추가되었으며, 2004년 5월 아카이브 운영에 관한 요소를 반영하여 34개의 메타데이터 요소를 제안하고 있다.

2.2.4 WARP [8]

WARP(Web Archiving Project)은 2002년 11월 시작한 온라인 전자 출판물의 수집을 위한 웹 아카이빙 프로젝트로 NDL(National Diet Library)에서 지원한 3년간의 시범적인 프로젝트이다. WARP은 일본 도메인 상의 인터넷 정보를 수집·보존하여 미래세대에게 전달하는데 그 목적이 있다. WARP는 2001년 3월 더블린 코어를 기반으로 메타데이터 생성을 위한 NDL 표준안인 "NDL Metadata ELement Set"을 공표하였다. WARP의 메타데이터는 자원의 설명적 메타데이터, 구조적 메타데이터, 관리적 메타데이터로 구성되어 있으며, 보존적 메타데이터 요소는 포함되어 있지 않다.

2.2.5 NLA (National Library of Australia) [9]

호주국립도서관은 웹 아카이빙을 위해 15개의 기본적인 요소를 개발하였으나 아카이브 관리 시스템이 다양한 디지털 정보 자원의 보존을 지원하는 메타데이터 요소를 도출하지 못하자 독자적으로 보존 메타데이터를 개발하였다.

2.2.6 EVA Project [10]

EVA 프로젝트는 핀란드의 국립도서관인 헬싱키 대학도서관의 사서와 출판사업자 그리고 전문가그룹 등과 함께 진행한 핀란드 교육부의 정보사회전략 프로그램(Information Society Strategy Program)에 포함된다. EVA는 핀란드의 인터넷에서 공인 또는 자유롭게 출판된 온라인 문서를 캡쳐, 등록 및 보존하고 접근을 지원하기 위한 방법을 실험하는 프로젝트이다. EVA는 수집된 자원을 기술하기 위해 더블린 코어 기반의 Nrodic 더블린 코어 15개 요소를 사용하였다.

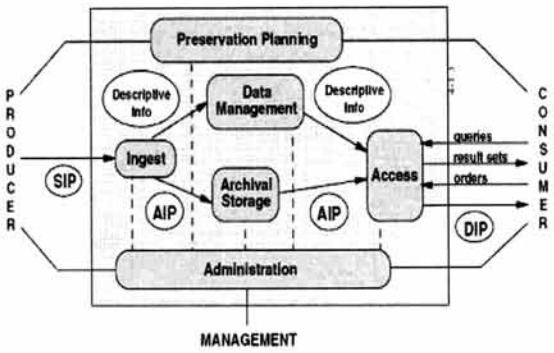
2.3 웹 아카이빙을 위한 표준화 및 사례 연구

2.3.1 ISO 14721 OAIS 참조 모형

2002년 ISO 14721로 제정된 국제표준으로서 장기간에 걸쳐 디지털 정보를 보존하여 이에 대한 지속적 접근을 제공하고자 하는 디지털객체보존 시스템, 즉 아카이브를 위한 개념적 구조 틀이다. 현재는 진행되고 있는 거의 모든 디지털 정보 보존기관과 프로젝트의 기반이 되는 디지털 보존 시스템을 위한 개념틀 이기도 하다. OAIS에서 정의하고 있는 디지털 아카이브의 기능들은 입수, 디지털객체물 저장, 데이터 관리, 행정관리, 보존계획, 이용 이렇게 6가지이며 OAIS 참조모형은 (그림 1)과 같다[11].

2.3.2 NEDLIB 보존 메타데이터

NEDLIB(Networked European Deposit Library) 프로젝트는 1998년부터 2001년까지 유럽 7개국의 국립도서관 (네덜란드, 핀란드, 프랑스, 이탈리아, 노르웨이, 포르투갈, 스위스)과 3곳의 온라인 전자출판물 출판사(Elsevier, Kluwers, Springer Verlag) 그리고 네덜란드 국립기록보존소가 참여한 전자출판물의 장기 보존을 위한 프로젝트이다. NEDLIB 보존 메타데이터 또한 OAIS 참조 모형에 기초하여 개발되었으며, 그 결과는 다시 OAIS 참조모형이 완성되는데 영향을 주었다. CEDARS, NLA, OCLC/RLG의 보존 메타데이터가 디지털 정보의 장기 보존과 접근이라는 두 가지 목적을 두고 개발 된 것과 달리, NEDLIB는 디지털 정보자원의 보존 메타데이터는 엄격하게 보존에만 초점을 맞춰서 개발이 진행되었다[12].



(그림 1) OAIS Reference Model

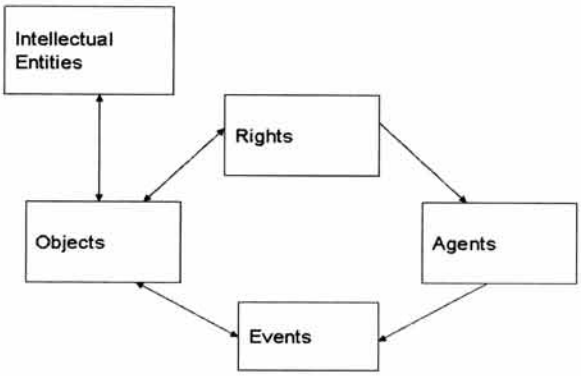
2.3.3 OCLC/RLG 보존 메타데이터

2002년 6월 OCLC/RLG에서는 디지털 객체(object)를 보존하기 위한 메타데이터 프레임 워크를 개발하였다. OAIS 참조 모형을 기초로 하는 보존 메타데이터 프레임워크는 보존 메타데이터를 구성하는 정보 유형의 포괄적이고, 상위 수준의 설명을 제시하였다. 또한 보존 메타데이터 요소의 " 프로토타입(Prototype) "을 구성하여, 공식적인 보존 메타데이터 명세서를 개발위한 합의에 기초하였다[13].

2.3.4 PREMIS Data Dictionary [14]

2003년 6월 OCLC/RLG는 PREMIS 보존메타데이터 구현 전략을 위해 6개국의 대학도서관, 국가도서관, 박물관 아카이브, 정부기관 기업에서 참여한 30명의 전문가로 구성된 PREMIS 실무반을 구성하였다. REMIS 실무반의 목적은 첫째, 보존 메타데이터 입력, 저장, 관리 및 교환을 위한 대안적인 전략들은 식별하고 평가하는 것이고, 둘째, 구현 가능한 필수 보존 메타데이터의 관리 및 이용을 위한 가이드라인과 권고안을 정의 하는 것이다. 그 결과로 PREMIS 데이터 모형과 의미단위(Semantic unit)를 정리한 PREMIS Data Dictionary 를 완성하였다.

PREMIS 데이터 모델에는 엔티티 (entity), 엔티티 속성 (property), 엔티티 간의 관계들이 묘사된다. 5개의 엔티티 (Intellectual Entities, Objects, Events, Rights 그리고 Agents)는 보존 메타데이터로 설명되는 디지털 보존에 적절



(그림 2) PREMIS 데이터 모델

한 개체들이고, 엔티티의 속성은 의미단위로 표현된다. PREMIS 메타데이터 요소들의 논리적인 조직을 용이하게 하기 위해서 (그림 2)와 5가지 유형의 속성들로 구성된 단순한 모델을 개발하였다.

2.4 메타데이터 사례 분석 및 이슈

앞서 살펴본 대표적인 웹 아카이빙 메타데이터 해외 사례들의 경우 웹 크롤러, 로봇이 자원을 수집하는 방법을 채택하고 있다. 또한 자동적으로 수집된 메타데이터 또는 전자자원의 관리와 검색을 위한 더블린코어를 근거로 메타데이터 요소를 개발해 왔다. <표 1>은 프로젝트 사례별 메타데이터에서 분야별 ·요소별 메타데이터의 콘텐츠정보, 구조정보, 권리정보, 관리정보 및 기술정보로 구분하여 각 사례들에 대한 구조와 요소를 비교·분석하였다. WARP은 서명정보와 식별자 정보를 상세하게 구분하였고, OCLC는 웹 아카이빙 시스템에서 자원을 관리하기 위한 정보를 세분화하고, 자원 관리를 위해 필수적인 기술 정보를 추가하였다.

이처럼 초기의 웹 아카이빙 메타데이터가 기존의 다양하고 방대한 자원 수집에 초점을 맞췄었다면, 최근에는 수집, 관리 및 디지털 자원의 장기적인 보존을 위한 요소까지 메타데이터 항목에 추가하여 개발을 진행하고 있다.

또한 해외 보존메타데이터 개발 프로젝트의 결과인 ISO 14721 OAIS 참조모형은 아카이빙이 갖춰야 할 개념적 틀에서 출발하여 이상적인 모형을 제공하기는 하였으나 실제 활용하기에는 현실적인 상황과 차이가 있다. 최상위 범위, 모든 것을 포괄하는 구조 또는 간단하거나 최소의 필수적인 요소만을 제공하다 보니 상세하고, 구체적인 자원의 정보관

〈표 1〉 웹 아카이빙 메타데이터 비교

Catetory	DC	OAIS	OCLC DAM	WARP	NLA	Nordic DC
Content Info	Title Subject Dsecription Source Language Relation Coverage	Title Subject Dsecription Source Language Relation Coverage Pages Annotation	Title Original Functionality Content Description Resource language Relation	Bibliographic Title Other Title Web Page Title Domain, NDL Description Source Language Relation Coverage Volume Number Annotation	Title of web page Subject fo Keyword Description Source Language Relation Coverage	Title Subject Dsecription Source Language Relation Coverage
Sturcture Info	Type Format Identifer	Object Type Media Type Format Format Expansion Original URL Havest URL	Object Type Object Composition OCLC Identifier Standard Identifier	Type Collection NDL Resource Type Original URL Past URL Web Page URL ISSN Other ISSN ISBN Publishing Frequency	Format Resource ID (URL)	Type Format Identifier
Right Info	Creator Publisher Contributor Right	Creator Creator Contact Publisher Contributor Right COI Copyright Aggrement NO	Creator Publisher	Editor Publisher	Author or Creator Publisher Other Contributor Right Management	Creator Publisher Contributor Right
Management Info	Date	Archiving Resource NO Archiving Group NO Creation Date Havested on Date Ingested on Date Modification Date Modification Manager Resigtration Magner Service Decision	Other Metadata Identifier Administrative elements Object Locator Digital Archive Record ID Digital Archive Record Language Digital Archive Save File NO Created (date) Havested on (date) Ingested on (date) Expires in (date) Date Modified (date) Status Event Local Note Functionality chagnge Functionality in Archive	Metadata ID Collection Date	Creation Date Modification Date	Date
Technical Info		Acquisition File	Application Encording Standard Logical object Size Operating System Peripherals	File		

리가 요구되는 웹 아카이빙에서는 계속해서 새로운 메타데이터, 실제 사용하는데 적합한 메타데이터가 요구되어야 한다는 것이다.

OASIS는 웹 아카이빙을 위한 단계별·기능별 요구사항이 반영되지 않았고, 멀티미디어 자원에 대한 기술적 표현 방법에 대한 제한이 있다. 또한 장기보존을 위한 웹 아카이빙임에도 보존보다는 우선 자원 수집을 중점적으로 하고 있어 디지털 자원 보존 기능을 위한 요소 개발에 대한 요구가 있다.

따라서 향후 OASIS 등의 웹 아카이빙에서는 웹 사이트, 웹 자원의 장기적인 보존을 위한 방향과 필요한 업무 프로세스에 따른 관리 방안에 대한 메타데이터 요소가 개발되어 활용되어야하는 이슈를 제시 한다.

3. 웹 아카이빙을 위한 메타데이터 구조 및 요소 개발

3.1 웹 아카이빙 자원 정의 및 특징

3.1.1 웹 아카이빙 자원 정의

디지털 자원은 컴퓨터에서 생산되어진 자원(Born Digital) 또는 재생산 된 자원(Digitizing)을 말하며, 인터넷을 통해 접근 가능한 자원(Online)과 디지털 매체에 저장된 자원(Offline)을 포함한다. 웹 크롤러에 의한 자원 수집 방법을 택하고 있는 OASIS는 「온라인 디지털 자원 선정지침」에 따라 다음과 같이 정의하고 있다[2].

- 인터넷을 통해 전자화된 디지털 자원의 내용을 내려 받아 개인용 및 휴대용 컴퓨터 등의 정보통신 단말기에서 볼 수 있는 디지털 버전의 출판물
- 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
- 웹 사이트, 웹 페이지, 문서 파일 (pdf, hwp, doc, txt 등), 이미지, 동영상, 음악, 압축 파일

3.1.2 웹 아카이빙 자원 특성

본 논문에서는 웹 자원 특성에 관한 3가지 관점을 통해 웹 아카이빙을 위한 메타데이터 요구사항을 기술하고자 하였다.

첫째, 자원의 내용적 측면 특성이다. 디지털 자원, 특히 처음부터 디지털 자원으로 생산된 경우는 누구나 손쉽고, 자유롭게 저작물을 창작하고 바로 온라인상에 유통시킬 수 있어 공식적인 검증 과정이 생략되는 경우가 많아 자원의 신뢰도나 공신력에 제한이 있다.

둘째, 자원의 관리적 측면 특성이다. 디지털 자원은 동일 자원에 대한 복수의 포맷으로 생산이 가능하고, 다양한 유형의 포맷으로 구성된 디지털 자원의 구성도 존재한다. 또한, 디지털 자원은 인쇄자료에 비해 자원의 저장 및 다운로드가 용이하여 복제본의 생성이 쉬우며, 손쉽게 이동이 가능하다. 그리고 기존 인쇄 자원들은 물리적인 형태의 변경이나 소멸되는 경우가 적지만, 디지털 자원은 응용 기기의 순간적인 오작동이나 오류발생 또는 버전 업그레이드 등을 통해서 자원이 유실되거나 훼손될 가능성이 있어 관리 시 주의가 필요하다[15].

셋째 자원의 이용적 측면 특성이다. 디지털 자원은 온라인으로 연결된 곳이라면 장소와 시간에 구애받지 않고 접근이 가능하다. 또한 디지털 자원을 이용하기 위해서는 반드시 정보가 수록된 매체의 응용프로그램과 이를 활성화하기 위한 하드웨어가 요구된다. 뿐만 아니라, 네트워크를 통해 손쉽게 복제와 이동이 가능한 까닭에 저작물에 대한 불법적인 복제와 유통 등 저작권자의 권리를 침해하는 사례가 비일비재하게 발생하고 있다.

3.2 웹 아카이빙 메타데이터 개발

본 논문에서 디지털 자원 웹 아카이빙을 위한 메타데이터를 개발하기 위해서 지속적으로 연구되어 온 더블린코어 메타데이터, OCLC/RLG 보존메타데이터, NEDLIB 보존메타데이터, NLA 보존 메타데이터, OCLC 웹 아카이빙 메타데이터 그리고 최근 완료된 보존 메타데이터의 결정판이라 할 수 있는 PREMIS Data Dictionary (1.1)를 비롯하여 그동안의 웹 아카이빙 관련 메타데이터 사례들을 분석하여 메타데이터 개발 목표와 원칙을 제시하였다.

3.2.1 메타데이터 개발 목표

첫째, 다양한 디지털 자원에 적용 가능하도록 디지털 자원의 매체별·유형별 메타데이터 요소를 구조화 및 상세화

둘째, 웹 자원의 특성을 고려한 연계정보를 구성

셋째, 웹 아카이빙에서 보존 전략 및 계획을 추진 할 때 지원이 가능한 보존 요소 개발

넷째, 향후 아카이빙 기관들 간의 메타데이터 호환과 자원 공유를 위해 상호운영성 확보가 가능한 메타데이터 구조를 제시

3.2.2 메타데이터 개발 원칙 및 검증

첫째, ISO 14721 OAIS 참조모형을 준수한다. OAIS 참조모형은 웹 아카이빙에서 발생하는 모든 과정에서 최상위의 개념적 프레임워크를 제공한다. 현재 운영 또는 준비 중인 세계의 웹 아카이빙 프로젝트나 기관 또한 OAIS 참조모형의 기준을 적용하고 준수하기 위해 노력하고 있다[3]. 따라서 ISO 14721 OAIS 참조모형에서 제시하고 있는 아카이빙의 절차에 따른 단계별 요구정보를 통한 메타데이터의 구조와 요소를 개발하고자 한다.

둘째, 표준 또는 권위 있는 메타데이터를 준용한다. 이러한 메타데이터는 미국과 유럽을 중심으로 각국의 전문가와 앞선 사례들이 중심이 되어 오랜 기간 논의와 조사를 통해 개발되었다. 이에 현재 웹 아카이빙을 준비·계획 중이거나, 운영하고 있는 많은 프로젝트와 기관에서는 이들이 제시한 메타데이터 구조와 요소를 참고하여 메타데이터 요소를 추

출하고 있다[16]. 따라서 본 논문에서는 국제적인 아카이빙 프로젝트 및 각 국가별 아카이빙에서 활용되는 메타데이터의 다양한 요구사항과 특성들을 고려한 구조와 요소에 대한 확보방안을 마련한다.

3.2.3 웹 아카이빙 메타데이터 구조 설계

웹 아카이빙에서 비트스트림과 파일 등의 형태로 존재하는 자원을 식별, 관리 및 보존을 하기 위해서는 자원에 대한 다양한 정보를 기록한 메타데이터가 필요하다. 이때 어떤 정보를 갖고 있느냐에 따라 설명적 메타데이터, 구조적 메타데이터, 관리적 메타데이터 등으로 구분 지을 수 있다[17]. 이를 위해 본 논문에서는 OAIS 참조모형에서 제시한 아카이빙의 기능 단계별(입수 전-입수, 저장, 데이터관리, 보존, 접근/서비스, 아카이빙 운영)로 요구되는 정보를 활용할 수 있는 구조적인 특징을 반영한 메타데이터 구조를 제시하고자 한다.

첫째, 입수 전-입수단계 : 아카이빙에서 수집 할 자원을 선별하여 입수하는 과정으로 무엇보다도 디지털 자원 객체 자체에 관한 정보와 자원의 진본성과 무결성 확인을 위한 정보들이 제공된다. 자원에 대한 정보로는 제목, 주제, 표현언어 등 자원의 서지 사항을 설명하는 정보가 있다. 또한 자원의 진본성과 무결성 검증을 위해 저작자, 권리 관계 등의 권리 정보 그리고 수집 자원의 파일 정보를 다루는 기술정보 뿐만 아니라 고유 식별자(URL,ISBN 등) 및 자원의 구조와 구성에 대한 정보 등이 요구된다.

둘째, 저장단계 : 저장은 수집 된 자원에 대한 장기 보존을 위해 저장소(서버)에 추가하는 과정으로 데이터가 실제 아카이빙에 제공되는 단계다. 이때는 자원에 대한 마지막 확인 과정(에러 점검)과 저장된 공간의 위치 정보, 저장소에 대한 관리정보 등이 요구된다.

셋째, 데이터 관리 단계 : 데이터 관리는 최초 입수된 자원이 아카이빙에서 관리되고 있는 단계에 관한 정보가 제공된다. 즉 데이터가 현재 어떤 상태(예: 포맷 변경, 저장 매체 변경 등)에 있는지에 대한 정보가 요구된다.

넷째, 보존 단계 : 보존을 위해서는 우선 아카이빙에 보존 전략이 있어야 하며, 데이터 그 자체 뿐 아니라 그 보다 더 많은 정보를 수집해야 한다. 그 자원의 출처가 어디였는지, 자원의 진본성 확인은 언제, 어떻게 받았는지, 다른 아카이빙 자원과도 연결이 되는지, 생성부터 보존까지 어떤 과정을 거쳤는지 등에 관한 자세한 정보가 필요하다. 따라서 이 단계에서는 보존 전략 및 보존 계획에 관한 정보, 기술정보 변화에 따라 환경 정보를 지속적으로 관리해야 한다.

다섯째, 접근/서비스 단계 : 접근은 관리자가 접근점을 제공해 주는 경우와 OAIS 참조모형의 DIP(Dessemination Information Package:분배정보패키지)를 제공하는 경우처럼 이용자가 원하는 정보를 아카이빙에 요청하는 경우가 있다. 전자는 현재 검색 정보를 제공하는 것과 마찬가지로 해당 자원의 고유 식별 정보나 관리 정보 중 몇 개의 요소를 선정하면 된다. 후자의 경우는 아카이빙이 제공 가능한 정보에 대해서 이용자에게 공지를 하면 된다.

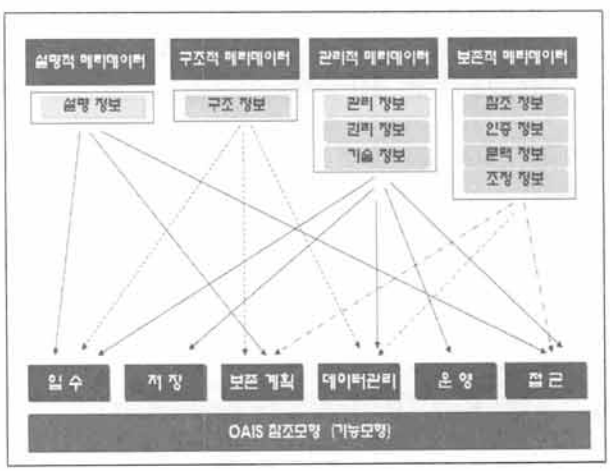
〈표 2〉 기능 단계별 요구 정보

기능 단계	요구 정보
입수전-입수	설명정보, 구조정보, 기술정보, 관리 정보, 권리 정보
저장	기술정보, 인증정보
데이터관리	관리 정보, 이력정보
보존	출처 정보, 인증정보, 고정정보, 참조 정보
접근	식별정보, 관리정보
운영	관리자 정보, 아카이빙 고유정보, 정책 정보

여섯째, 아카이빙 운영 단계 : 운영은 아카이빙 전반에 걸쳐 처리해야 하는 정보들이 포함 된다. 디지털 자원을 체계적으로 관리하기 위한 정보, 아카이빙 정책과 관련된 정보(예를 들면 자원의 서비스 여부, 내부 정책 수립, 이용자 및 정보 제공자 관련 정책 및 지원 정보) 그리고 아카이빙 및 자원을 관리하는 업무를 하는 담당자에 관한 정보, 시스템의 하드웨어와 소프트웨어 등의 환경 설정 정보 등이 해당된다.

본 논문에서 메타데이터를 구조화하기 위해 아카이빙의 각 기능 단계별(6단계)에서 요구되는 정보로 구분하여 정리하면 <표 2>와 같다. 각각의 메타데이터 정보들은 특정 단계(예: 운영의 관리자 정보, 데이터관리의 이력정보 등)에서만 나타나기도 하고, 아카이빙의 전 분야에 걸쳐 공통적으로 요구(예: 입수 전, 입수, 저장 단계의 기술정보)될 수 있다.

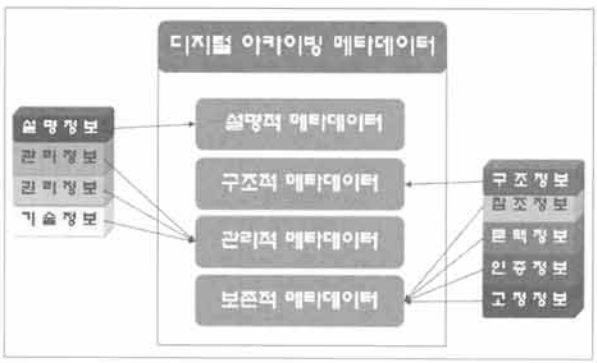
(그림 3)은 <표 2>에 나타난 각 기능단계별 주요 메타데이터 정보가 표현하는 내용에 따라 메타데이터를 구조화한 것이다. 일반적으로 관리적 메타데이터로 분류되던 보존 정보가 웹 아카이빙에서는 출처정보, 인증정보, 고정정보, 참조정보를 포함하며 구조화 및 상세화 되고 있다. 따라서 본 논문에서는 아카이빙 단계별 요구되는 정보에 따라 설명적 메타데이터(설명정보, 식별정보), 구조적 메타데이터(구조정



(그림 3) 아카이빙 단계별 메타데이터 유형

〈표 3〉 메타데이터의 내용과 사례

구분		내용	예시
설명적 메타데이터	설명 정보	디지털 자원에 대한 설명으로 주로 서지정보 일부를 포함하는 자원을 유일하게 식별할 수 있는 기본 정보를 제공	디지털 자원의 제목, 주제, 표현 언어
구조적 메타데이터	구조 정보	디지털 자원을 표현하기 위한 수단 및 방법에 따라 자원의 구조와 구성에 대한 정보를 제공	문서, 이미지, 텍스트, 비디오/ 웹사이트, 연속 간행물, 전자 도서
관리적 메타데이터	관리 정보	웹 아카이빙에서 입수된 자원을 관리하고 아카이빙을 운영 하기 위해 요구되는 정보를 제공	아카이빙 등록번호, 그룹 번호, 날짜, 관리자, 사건 정보
	권리 정보	디지털 자원의 지적 재산권에 관련된 개인, 단체, 기관에 대한 정보를 제공	저작자, 출판사, 지적재산권정보,COI정보
	기술 정보	디지털 자원을 사용하기 위해 요구되는 운영시스템, 소프트웨어, 하드웨어 정보 등 기술적인 환경에 대한 정보를 제공	Operating System 정보, 파일 크기, 파일포맷, (Window XP 등), 각종 소프트웨어 정보
보존 메타데이터	출처 정보	자원의 출처에 대한 정보를 제공	서지정보, 메타데이터 정보
	참조 정보	디지털 자원의 내용 정보에 대한 히스토리, 즉 이력에 대한 정보를 제공	입수전-입수 정보, 자원의 기원 정보
	문맥 정보	디지털 자원 내용 정보의 생산 이유와 다른 내용 정보와의 관련에 대한 정보를 제공	자원의 관계정보
	인증 정보	해당 디지털 자원이 진본의 자원인지 분명한가를 확인하고 보장하는 정보를 제공	진본성 확인



(그림 4) OASIS 메타데이터 구조

보), 관리적 메타데이터(관리정보, 관리자 정보, 기술정보, 정책정보, 권리정보), 그리고 보존적 메타데이터 (참조정보, 인증정보, 문맥정보, 출처 정보)로 구분하였다.

보존적 메타데이터를 관리적 메타데이터와 구분한 것은 현재 OASIS와 같이 보존 정책과 계획이 수립되지 않은 웹 아카이빙 상황에서 보존과 관련된 정보들 - 문맥정보, 참조정보, 출처정보, 인증 정보 - 등에 대해 별도로 관리해야 할 필요가 있음을 고려하였기 때문이다. 결과적으로 현재 OASIS 메타데이터와 앞서 사례 분석한 연구내용을 토대로 4개의 섹션(Section)과 9가지 메타정보로 구성된 구조를 제안하고 그에 따른 구성을 (그림 4)와 같이 표현하였다.

<표 3>은 웹 아카이빙 메타데이터 구조에서 제시된 4개의 섹션의 9가지 메타데이터 정보에 대한 내용과 그에 대한 설명 예시를 기술하였다.

3.2.4 웹 아카이빙 메타데이터 요소정의

본 논문에서는 앞서 제안한 OASIS 메타데이터 구조와 개발 원칙을 기준으로 총 29개의 메타데이터 요소를 추출하고 구조화, 상세화를 통해 하위 요소를 선정하였다. <표 4>부터 <표 7>까지는 각 단위 섹션별 메타데이터 요소에 대

〈표 4〉 설명적 메타데이터 요소

범주	요소	하위요소	기원	정의	의무
설명 정보	Title		DC	디지털 자원을 대표하는 이름	필수 반복
	Subject	Collection	DC	디지털 자원 내용의 주제를 아카이빙 컬렉션과 분류번호로 표현	필수
		Classification number			
	Description	-	DC	디지털 자원을 설명할 수 있는 요약정보 및 부연 설명 정보	선택 반복
	Source	-	DC	디지털 자원의 원래 출처	선택 반복
	Language	-	DC	디지털 자원을 기술하고 있는 대표 언어	선택 반복
	Coverage	-	DC	디지털 자원이 적용되는 기간과 장소	선택 반복

〈표 5〉 구조적 메타데이터 요소

범주	요소	하위요소	기원	정의	의무
구조정보	Object Type	Text - Format & version - compression	NLA	디지털 자원을 표현하는 수단의 유형 - 텍스트 - 이미지 - 비디오 - 오디오 - 멀티미디어	필수 반복
		Image - Format & version - image resolution - image dementions - image color - image orientation - compression			
		Video - Format & version - frame dimensions - duration - frame rate - compression - video encoding structure - video sound			
		Audio - format & version - audio resolution - duration - bitrate - compression - track & type			
		Multimedia			
	Object Genre	Web Site	-	디지털 자원의 유형별 종류에 따라 웹사이트와 웹자원에 대한 정보	필수 반복
		Web Resource - Individual - Group			
	Object Identifier	Original URL	OCLC	디지털 자원의 유일한 식별자	필수 반복
		Source URL			
		Harvest URL			

〈표 6〉 관리적 메타데이터 요소

범주	요소	하위요소	기원	정의	의무
권리정보	Creator	Name	DC	디지털 자원의 창작한 자 또는 기관의 명과 연락처	필수 반복
		Contact - Telephone Number - Address - e-mail	OASIS		선택 반복
	Publisher	Name	DC	디지털 자원을 발행한 사람 또는 기관, 발행한 장소	필수
		Locator	-		선택
	Right	Agent - Personal - Agencies	DC	디지털 자원에 대한 권리정보를 갖는 사람 또는 기관, 권리의 기간, 저작권 동의서에 관한 정보	필수 반복
		Right Period - Start date - End date	-		필수 반복
		Copyright Agreement Number	OASIS		필수
	COI	COI Identifier number	OASIS	해당 디지털 자원과 연계된 COI 식별자 번호	필수
관리정보	Date	Created	DC	디지털 자원의 생애 주기에서 일어나는 사건에 대한 날짜	필수
		Issued			필수
		Harvested			필수
		Ingested			필수
		Modified			필수 반복
	Digital Archiving save file number	Object Number	OCLC	디지털 자원이 아카이빙에 입수된 번호 - 디지털 객체의 입수번호와 그 객체가 포함된 자원의 번호	필수
		Group Number			필수
	Service level	-	OCLC	디지털 자원의 내부/외부 서비스 여부	필수
	Management person	Registration person	OASIS	아카이빙 내부에서 디지털 자원의 관리와 목록작업에 관련된 사람	필수 반복
		Modification person			필수 반복
	Event	Harvest	OCLC	디지털 자원의 아카이빙 되는 단계별 상황	필수 반복
		Ingest			
		Archive			
		Modification			
기술정보	File	Name	OASIS OCLC	디지털 자원이 저장된 파일의 정보	필수
		Size			
	Object Format	Name	NEDLIB	디지털 자원의 파일 포맷 정보	필수
		Value			
	Operating System	Name	NEDLIB	디지털 자원을 이용하기 위해 요구되는 권장되는 운영체계 정보	필수 반복
		Value			
	Application	Name	NEDLIB	디지털 자원 이용하기 위해 요구되는 응용 프로그램	필수 반복
		Value			

하여 메타데이터의 범주(Category), 요소(Element), 하위요소(Sub Element), 기원(Origin), 정의(Definition) 및 의무사항(Cardinality)에 관한 설명이다.

4. 결론 및 제언

본 논문에서는 ISO 14721 OAIS 참조모형을 적용한 국내 웹 아카이빙 메타데이터의 구조와 활용요소를 개발하였다. 그 결과 첫째, 웹 아카이빙을 위한 디지털 자원으로 표현되는 수단인 문서, 이미지, 동영상, 사운드의 매체별 자원에 대

〈표 7〉 보존적 메타데이터 요소

범주	요소	하위요소	기원	정의	의무
출처 정보	Resource Description		CEDARS	디지털 자원의 출처에 관한 설명	필수 반복
참조 정보	Origin	-Designation -Procedure -Date -Responsible agency -Outcome -Note -Next occurrence	OCLC/RLG	디지털 자원의 생성과 관련된 정보	필수
	Pre-ingest		OCLC/RLG	디지털 자원의 입수 전 정보	필수
	Ingest		OCLC/RLG	디지털 자원의 수집 시 정보	필수
	Archival retention		OCLC/RLG	디지털 자원의 아카이빙 시 정보	필수
문맥 정보	Relation-ships	-	OCLC/RLG	디지털 자원과 연관되는 정보	필수 반복
인증 정보	Object Authenti-cation	Type	OCLC/RLG	디지털 자원의 진본성 확인에 대한 정보	필수 반복
		procedure			
		date			
		result			

한 정보는 자원의 구조정보를 기술하는 구조적 메타데이터 요소로 구조화, 상세화 하였다. 둘째, 웹 자원의 특성을 고려하여 웹사이트, 웹 자원의 연계 정보를 구체화하였다. 셋째, 디지털 자원의 장기 보존을 지원하기 위한 필수 보존 메타데이터 요소를 추가하였다. 그 결과 메타데이터 개발 원칙과 기준에 의해 4개의 메타데이터 섹션 구조 - 설명적, 구조적, 관리적, 보존적 - 와 9개의 범주정보(Category)를 담을 수 있는 메타데이터 구조를 제안하고, 각 섹션구조에 따른 단위 정보를 제공할 수 있는 29개의 메타데이터 요소와 각 요소별 필수하위 요소를 개발하였다.

향후 연구에서 보존 메타데이터 개발은 국립중앙도서관의 OASIS 및 각 웹 아카이빙의 정책적인 방향에 따라 보존 전략과 계획을 준비해야 할 것이다. 현재의 보존 메타데이터 요소는 자원의 장기 보존을 지원하기 위해 요구되는 기본적인 요소만을 추출한 것이다. 향후 보존 정보가 강화된다면 보존 메타데이터 섹션에 구조 정보, 기술 정보, 관리 정보의 일부 정보를 포함하는 광의의 보존 메타데이터로도 변경도 가능하다. 따라서 디지털 자원의 단계별 보존 전략과 계획에 의한 구체적인 메타데이터 요소에 대한 연구가 필요할 것이다.

참 고 문 헌

[1] 서혜란, "디지털 납본제도 방안".디지털 유산 보존을 위한 포럼, 2004.

[2] OASIS homepage : http://www.OASIS.go.kr

[3] 국립중앙도서관,"OASIS 표준화과제연구", 2003.

[4] Deborah Woodvard, "Preservation Metadata" OCLC/SCURL NEW Directions in Metadata, Edinburght, 15-16 August 2002.

[5] 김희정, "전자저널 아카이빙을 위한 OAIS 참조모형의 적용방안에 관한 연구", 박사학위논문, 연세대학교 문헌정보학과, 2003.

[6] DCMI : http://dublincore.org/

[7] OCLC. "Digital Archiving Metadata Elements", Dublin, Ohio, 2004.

[8] WARP : http://warp.ndl.go.jp

[9] NLA : http://www.nla.gov.au/

[10] Kristi Lounamaa, "EVA-The Acquisition and Archiving of Electironic Network Publications In Finland", Tietolinja News. 1999.

[11] 이소연, "디지털 아카이빙의 표준화와 OAIS 참조모형", 정보관리연구 33(3), pp.45-68, 2002.

[12] NEDLIB : http://nedlib.kb.nl/

[13] OCLC.RLG, A Metadata Framework to Support the Preservation of Digital Objects,2002.

[14] PREMIS Working Group, " Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group", 2005.

[15] 김태수, 목록의 이해,서울,한국도서관협회, 2000.

[16] OCLC/RLG, "Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community", 2004.

[17] NISO, "Understanding Metadata", 2004.

오 상 훈

e-mail : oshosh24@gmail.com

1988년 한국외국어대학교 전자계산학과 (학사)

1990년 한국외국어대학교 경영정보대학원 응용전산학과(석사)

2006년 충남대학교 정보통신공학부 정보검색 및 자연어처리(박사)

1994년~2000년 한국과학기술정보연구원 연구원

2000년~2001년 (재)한국데이터베이스진흥센터 팀장

2001년~현 재 (사)한국디지털콘텐츠산업협회 사무국장

관심분야 : 디지털아카이빙, 디지털콘텐츠 유통보호, 메타데이터, 정보검색

최 영 선

e-mail : ming279@gmail.com

2003년 숙명여자대학교 정법학부(학사)

2007년~현 재 (사)한국디지털콘텐츠산업협회 연구원

관심분야 : 디지털아카이빙, 메타데이터