

오마이뉴스 기사 메타데이터 재설계방안에 관한 연구

정 성 숙*

1. 머리말
2. 오마이뉴스 아카이빙의 필요성
 - 1) 대중매체의 사회적 기능과 오마이뉴스
 - 2) 오마이뉴스 기록의 특성
 - 3) 오마이뉴스 기록의 메타데이터 재설계 필요성
3. 오마이뉴스의 기록 생산과 메타데이터
 - 1) 보도기사 기록생산 프로세스와 메타데이터 현황
 - 2) 국내외 메타데이터 표준 분석
4. 오마이뉴스 기록 메타데이터 재설계 방안
 - 1) 메타데이터 재설계 방향
 - 2) 메타데이터 기술요소 구성
 - 3) 메타데이터의 활용 방안
5. 맺음말

* 한남대학교 일반대학원 기록관리학 전공 석사.

[국문초록]

사회구성원들의 인터넷 사용이 크게 증가하고 있는 만큼 인터넷언론은 우리 시대에 많은 영향을 미치고 있다. 대표적인 국내 인터넷언론인 '오마이뉴스'는 오늘날 국내 독립 인터넷 뉴스를 대표하는 젊은 대안언론으로 평가받고 있으며, 인터넷뉴스의 특징을 보여주는 전형적인 사례이기도 하다.

본 논문은 오마이뉴스 보도기사의 처리과정에서 단계별로 생성되는 기록정보와 그에 해당되는 기술요소를 추출하였고, 메타데이터 설계방안 도출을 위하여 국내외 메타데이터 표준 사례를 분석하였다. 국내 메타데이터 표준은 2010년 10월에 발표된 기록관리 메타데이터 공공표준안을 대상으로 하였으며, 외국 메타데이터 표준 연구는 보존 메타데이터를 위한 PREMIS 데이터 사전으로 하였다. 국내외 사례연구를 바탕으로 이용자 접근의 용이, 연관정보와의 연결, 행위이력의 추적 등에 중점을 두어 메타데이터를 재설계하였다.

오늘날 대중매체의 기록관리 연구에서 인터넷매체를 비롯한 뉴미디어에 대한 연구는 미미한 수준에 있다. 이미 언론정보학, 정보통신학, 커뮤니케이션학 등의 외부 학계에서는 인터넷 매체에 대한 연구가 심도 있게 이루어지고 있다. 인터넷 매체에 대한 다양한 분야의 연구들은 앞으로 이루어질 대중매체 기록관리 연구에 있어서 반드시 고려되어야 할 것이라 예상된다.

주제어: 오마이뉴스, 인터넷언론, 메타데이터

1. 머리말

동시대에 생산된 기록들은 그 시대를 이해하기 위한 필수 요소이다. 따라서 기록을 통한 연구가 다양할수록 그 시대에 대한 설명은 더욱 견고해지며, ‘기록관리’는 역사 연구를 위한 자료를 만들어줌으로써 시대에 대한 설명을 완벽하게 한다고 할 수 있다. 그렇다면 지금 이 시대를 설명하기 위해 연구의 가치가 있는 세부 요소들은 무엇인지 생각해볼 필요가 있으며, 그중에서도 특히 ‘대중매체’는 우리 시대를 설명하는 대표적인 키워드이다. 신문과 TV방송뉴스의 경우 오늘날 역사연구에 가장 중요한 자료이다. 이에 따라 신문 아카이브, 방송뉴스 아카이브에 대한 연구도 자연스럽게 이루어졌다.

오늘날에는 더욱 많은 매체들이 과거 종이신문이 했던 역할들을 분담하고 있다. 그중에서도 인터넷이 대표적이며, 오늘날 대부분의 정보를 인터넷 매체를 통해 얻는다고 해도 과언이 아니다. 많은 영향력을 가짐에도 불구하고 인터넷 기사들은 종이신문이나 방송뉴스에 비해 기록관리의 필요성을 인식하지 못하고 있는 실정이다.

인터넷 기사의 메타데이터 획득을 통해 여러 가지 효과를 실현할 수 있다. 인터넷 뉴스 기사에 기록학적인 생산체계를 적용했을 때 얻어지는 효과를 정리하면, 언론사 내의 효과와 언론사 외부에서의 효과로 나눠 볼 수 있다. 먼저 오마이뉴스 내부에서는 보도기사의 신뢰성을 갖추어 경쟁력을 얻게 된다. 또한 등록기사 비삭제를 원칙으로 하는 오마이뉴스는 방대한 기사 데이터를 좀 더 효율적으로 관리할 수 있을 것이다. 한편 외적인 효과로는 무엇보다 구독자들이 스스로 원하는 기사를 검색하기가 수월해지고, 연관 정보 연결을 통해 더욱 풍부한 정보를 서비스할 수 있다. 나아가 기사 데이터가 연구자들에게 풍부한 정보를 제공하여 그것이 연구 자료로 활용된다면 오마이뉴스는 나름의 역사적

가치를 갖출 수 있을 것이다.

현재 발생하는 기사 검색의 어려움과 인터넷 기사의 신뢰도 문제 해결, 그리고 인터넷 기사 기록관리의 바탕이 되는 출발점을 제시하기 위해 본 논문에서는 <오마이뉴스>를 대상으로 보도기사 메타데이터 재설계 방안을 모색하고자 한다. 오마이뉴스는 분야별 순위 사이트인 랭키닷컴에서 인터넷신문 분야 7위에 위치한 인터넷언론사이다.¹⁾ 미디어다음이나 네이버뉴스와 같은 대형 포털에 소속된 인터넷신문을 제외한 독립형 인터넷신문에서는 그 영향력에 있어서 최상위권에 위치하고 있다. 본 논문에서는 오마이뉴스 보도기사에 대한 메타데이터 기술요소를 재설계함으로써 생산된 정보에 신뢰성을 부여할 뿐만 아니라 이용자의 기사 활용이 용이하도록 하는 방안을 연구하고자 한다.

대중매체를 기록관리의 범주에 포함하여 연구한 논문은 적지 않으나, 인터넷 뉴스에 대한 연구는 찾아보기 힘든 실정이다. 인터넷 언론의 기록학적 연구를 위해서는 먼저 앞서 발표된 대중매체 아카이빙 연구들을 살펴볼 필요가 있다.

먼저 방송뉴스 아카이빙에 대한 연구로 우현민(2007)²⁾의 논문이 있다. 방송뉴스 데이터 관리의 문제, 시스템 상의 문제, 조직상의 문제를 지적하고, 이를 해결하기 위한 분류체계 재설계와 메타데이터 재설정, 조직 개편 방안을 제안했다. 최초로 방송뉴스를 기록학적 관점에서 가치를 두고 연구했다는 점에서 의미가 있다. 박신애³⁾는 TV뉴스의 매체 특성과 기록으로서의 특성을 반영한 메타데이터 표준안을 제시하였다. 아카이브의 국제기술표준인 ISAD(G)뿐만 아니라 국내 방송사의 현행 메타데이터, 해외 방송연맹에서 정하고 있는 방송 콘텐츠 메타데이터를

1) 랭키닷컴(www.rankey.com)의 2012년 3월 통계임.

2) 우현민, 「방송 뉴스의 기록학적 관리 방안」, 명지대학교 기록과학대학원 기록관리학과 석사학위논문, 2007.

3) 박신애, 「TV 뉴스 아카이빙을 위한 표준 메타데이터 설계」, 한국외국어대학교 대학원 정보기록관리학과 석사학위논문, 2010.

연구하여 적절한 메타데이터 표준안을 구상하였다. TV 뉴스의 특성상 기록관리 방안이 주로 생산주체인 방송사의 관점으로 이루어졌으며, 이용자서비스와 관련된 개선방안이 없다는 점이 아쉽다. 양수연(2011)⁴⁾은 디지털시대에 필요한 방송뉴스의 기록관리는 방송기록의 생산단계 및 관리단계에서 기록관리 프로세스 일원화, 생산 시 제어를 위한 디지털 방송제작 환경 재정비, 방송기록의 선별·평가 절차에 대한 구체적인 논의가 필요함을 제언했다. 방송뉴스기록의 관리단계에 있어서 분류 및 활용, 접근에 대한 구체적인 방안이 다소 부족하다는 한계를 가진다.

신동호(2007)⁵⁾는 취재기록이 생산·저장·처리되는 집배신 시스템의 기록생산 프로세스와 아카이빙 단계의 기록관리 프로세스를 점검하고, 그 가운데서 취재기록의 관리에 적용될 수 있는 요소들을 추출하였다. 또한 최근 뉴스 콘텐츠의 표준화를 위하여 도입된 NewsML의 소개 자료와 문헌에서 기록관리 요소를 분석해 각 프로세스에 적용하였다. 이재근(2008)⁶⁾은 매일신문사를 중심으로 신문사 기록관 설립을 위해 규정과 조직, 수집, 관리, 평가, 보존 등 분야별로 기록학적인 적용을 시도했다. 특히 전국언론사 기사자료 표준분류표 등 현존하는 언론 표준에 대한 분석을 통해 신문사 기록관에서 실행할 수 있는 분류, 기술 방안을 제시했다는 점이 주목할 만하다.

본 논문의 방향인 메타데이터 재설계 방안 모색을 위해서는 메타데이터에 대한 이론 연구를 살펴볼 필요가 있다. 김익한(2004)⁷⁾은 기록

4) 양수연, 「디지털시대 방송뉴스의 체계적 기록관리를 위한 연구」, 명지대학교 기록정보과학전문대학원 기록관리전공 석사학위논문, 2011.

5) 신동호, 「통합집배신시스템을 통한 신문사 취재기록의 관리」, 명지대학교 기록과학대학원 석사학위논문, 2007.

6) 이재근, 「신문사기록관 설립 및 운영방안」, 경북대학교 대학원 기록학과 석사학위논문, 2008.

7) 김익한, 「기록의 속성과 메타데이터 표준을 통해 본 한국의 기록·기록기술」, 『기록학연구』 제10호, 한국기록학회, 2004, 3-26쪽.

기술요소는 기록관리기관에서의 관리행위와 이용자들의 기록 이용을 위해 설정되어 있는 반면에, 메타데이터 요소는 기록 그 자체의 속성을 유지하기 위한 수단으로서의 성격을 전제로 생산단계에서부터의 관리와 이용을 위해 설정되어 있음을 언급하며 기록 및 메타데이터 재설계의 방향성을 제시하였다. 또한 메타데이터 요소를 설계할 때 생산자가 기록을 생산하는 시점에 반드시 남겨야 하거나 남기기 쉬운 메타데이터 요소, 프로세스 상에서 변화할 수 있는 요소, 본문의 변경이 이뤄져서는 안 될 시점에 부가해야 할 요소, 기록이 관리되는 과정에서 부가되어야 할 요소를 추출해야 한다고 하였다. 사무관리규정, 기록관리법 등을 참고로 하여 기록서식에 들어있는 요소와 등록시 부가되는 요소를 나열한 뒤, 현행법상 규정되어 있는 메타데이터 요소와 현행 기술요소를 비교분석하였다.

김혜영(2009)⁸⁾은 기록의 한 유형으로서 웹 기록의 특징과 그 구조를 파악하고, 이를 관리하는 방법론과 관련 사례분석 결과를 고려한 웹 기록 보존 메타데이터 요소 세트를 설계하는데 목적을 두었다. 웹 기록관리 프로젝트에 대해 조사하여 웹 기록을 관리하기 위해 어떠한 프로세스와 방법론을 적용해야 하는지를 연구하였는데, IIPC의 웹 기록 메타데이터 연구, OCLC/RLG의 보존 메타데이터에 대한 연구보고서, PREMIS 연구 그룹의 데이터 사전 그리고 NLNZ의 보존 메타데이터 연구보고서를 분석하였다. 이러한 문헌 연구와 사례 분석을 통해 보존 메타데이터로서 포함해야 하는 정보 유형을 구조 정보, 테크니컬 정보, 이벤트 정보, 권한 정보, 변경이력 정보로 밝히고, 이러한 정보 유형을 포함하는 5개의 엔티티, 즉 객체 엔티티, 행위주체 엔티티, 이벤트 엔티티, 법규/권한 엔티티, 변경이력 엔티티를 설계하였다. 웹 기록이라는 넓은 범주의 대상을 연구하였기 때문에 특정 영역에서 실행할 경우 보완이 필요

8) 김혜영, 「웹 기록 보존 메타데이터 요소 설계에 관한 연구」, 명지대학교 기록정보과학전문대학원 기록관리트랙 석사학위논문, 2009.

할 것으로 예상된다. 그러나 웹 기록관리 프로세스와 방법론이 분명치 않은 시점에서 웹 기록의 이론적 개념을 정리하였다는 점에서 의의를 찾을 수 있다.

이경남(2006)⁹⁾은 전자기록의 장기적 보존을 위한 종합적인 보존 메타데이터 요소를 제안하고, 제안된 요소를 기준으로 국내 행정기관의 기록관리 시스템에 반영된 보존 메타데이터 요소 현황을 분석하였으며, 기록관리 과정의 각 단계별로 획득되어야 하는 요소를 정리하였다. OAI 참조모형¹⁰⁾의 정보 패키지 모형을 준용하여 개발된 Cedars 프로젝트의 보존 메타데이터 요소, NEDLIB 프로젝트의 보존 메타데이터 요소, OCLC/RLG에서 제안한 보존 메타데이터 프레임워크, 뉴질랜드 국립도서관의 보존 메타데이터 프레임워크를 소개함으로써 다양한 메타데이터 유형을 접하도록 하였다. 그 후에 호주 빅토리아주 기록보존소의 VERS 보존 메타데이터 스키마, PREMIS 연구 그룹에서 제안한 데이터 사전의 분석을 바탕으로, 두 표준의 공통요소와 기록관리에 필수적으로 반영되어야 하는 요소를 중심으로 보존 메타데이터 요소를 제안하였다. 그러나 제안한 전자기록의 장기적 보존을 위한 메타데이터 요소는 필수 요소, 그리고 상위 요소 중심으로 제안하였기 때문에 종합적이고 상세한 보존 메타데이터 스키마를 제공하지 못하였다는 한계를 지닌다.

정재윤·김태수(2006)¹¹⁾의 논문은 유일한 인터넷뉴스 메타데이터 관련 연구이며, 컴퓨터가 이해할 수 있는 웹 언어인 RDF¹²⁾로 웹상에 흩어져 있는 인터넷 뉴스 정보의 메타데이터를 만드는 것을 목적으로 연구

9) 이경남, 「전자기록의 장기적 보존을 위한 보존 메타데이터 요소 분석」, 한국의국어대학교 정보기록관리학과 석사학위논문, 2006.

10) Reference Model for an Archival Information System. 디지털 정보를 장기간 보존하는 데 필요한 개념적 구조틀을 제시하는 ISO 표준(ISO 14721)이다. 미국 항공우주국(NASA)의 우주데이터시스템기획위원회가 주체가 되어 개발했다.

11) 정재윤·김태수, 「인터넷뉴스의 메타데이터 구축에 관한 연구」, 『지식처리연구』 제7권 1/2호, 문헌정보처리연구회, 2006, 125-148쪽.

12) Resource Description Frame. 웹 자원에 관한 정보를 표현하기 위한 언어이다.

하였다. 디지털 자원의 메타데이터 포맷을 비교 분석하여 인터넷 뉴스 정보의 기술에 적합한 메타데이터를 선정하고, 실제로 그것을 활용하여 인터넷 뉴스 자원의 메타데이터를 구현하였는데, Dublin Core의 요소와 MARC 필드를 비교하여 웹 자원을 기술하기 위한 핵심요소를 추출하였다. 실제 인터넷 뉴스 자원의 정보를 RDF 메타데이터를 사용하여 기술 함으로써 실제 활용에 도움이 되도록 하였다. 인터넷 뉴스기사 전용 메타데이터의 부재로 기본적인 메타데이터를 설계하였기 때문에 세부요소의 손실을 고려하지 않은 한계점이 있다.

앞선 연구들을 통해 대중매체의 기록관리는 어떤 의미를 가지는지 또 무엇을 얻을 수 있는지 알아볼 수 있으며, 메타데이터 관련 연구를 통해 메타데이터에 대한 이론적 이해와 최근 동향을 살펴볼 수 있다. 이러한 선행연구를 바탕으로 본 연구에서는 오마이뉴스를 대표사례로 하여 인터넷매체에 적용할 수 있는 메타데이터 설계 방안을 탐색하고자 한다.

오늘날 대중매체의 기록관리 연구는 아직도 TV방송에 머물고 있는 상태이며, 인터넷매체로 시작되는 그 이후의 뉴미디어에 대한 연구는 미미한 수준에 있다. 따라서 최근 동향에 맞는 대중매체의 기록관리 방안을 모색한다는 점에서 본 연구의 의의를 찾을 수 있을 것이다.

2. 오마이뉴스 아카이빙의 필요성

1) 대중매체의 사회적 기능과 오마이뉴스

대중매체는 똑같은 메시지를 불특정 다수의 수신자에게 대개 일시에 전달하는 대량 전달 수단이다. 대중매체라는 말은 본래는 메시지의 대

량 전달 수단을 지칭하는 말에서 오늘날에는 그것들로 구성된 하나의 사회 제도라는 뜻으로 또는 대중매체를 생산하는 기관이라는 뜻으로 전화되어 쓰이기도 한다.¹³⁾ 대중매체에 관한 모든 정의들은 대중매체가 정보 송신자와 대중 사이에서 메시지를 전달하는 중간자의 역할을 한다는 의미를 반드시 포함하고 있다. 대중매체를 형태상으로 분류하면 크게 인쇄 매체와 전자 매체로 나눌 수 있다.¹⁴⁾ 대중매체의 형태는 점점 더 다양해지고 있으며, 수용자에게 정보를 전달하는 데에 있어서 더욱 효율적인 방향으로 변화하고 있다. 정보 송신자가 어떠한 형태의 매체를 이용하는가, 또한 매체가 정보를 어떻게 다루는가에 따라 대중이 받는 메시지는 그 영향이 아주 달라질 수 있다.

오늘날 대중매체는 없어서는 안 될 중요한 사회적 제도가 되었다. 그 중요성은 그것이 수행하는 여러 사회적 역할 또는 기능에서 뚜렷이 드러난다. 언론학에서는 대중매체의 사회적 기능¹⁵⁾에 대하여 크게 여섯 가지를 이야기하고 있으며, 정보제공, 사회조정, 문화전수, 오락, 동원 기능이 그것이다.

대중매체의 영향이 매순간 인간의 사고(思考)와 행동으로 발현되고 있는 지금 시점에서 대중매체를 역사의 일부분으로 인정하는 것은 피할 수 없다. 따라서 대중매체에서 발생하는 기록물에 대해 수동적인 자세를 탈피하여, 그것을 가치 있게 이용할 수 있도록 대중매체 아카이빙에 관한 꾸준한 논의가 필요하다.

인터넷뉴스의 특성은 ① 하이퍼텍스트성, ② 저장성, ③ 멀티미디어성, ④ 상호작용성, ⑤ 즉시성과 속보성으로 정리할 수 있다. 인터넷뉴스가 갖고 있는 이러한 특성은 공통적으로 시공간적 제한이 없어진다는 맥락에서 나타날 수 있는 것들이며, 이는 초기 대중매체라고 할 수

13) 강상현·채백, 『대중 매체의 이해와 활용』, 한나래, 2002, 29쪽.

14) 강상현·채백, 위의 책, 2002, 29쪽.

15) 강상현·채백, 위의 책, 2002, 33-35쪽 참조.

있는 종이신문의 한계점을 극복한 형태이다. 인터넷뉴스는 무한한 웹 공간에서 실시간 업데이트가 이루어지기 때문에 종이신문이 발휘했던 파급효과에 비해 훨씬 빠르고 광범위한 결과를 불러온다. 오늘날 구독자가 증가함에 따라 인터넷뉴스의 수적인 증가가 이루어지고 있는 추세이다. 이용자들은 선택적으로 뉴스를 구독할 수 있다는 점을 활용하여 보다 정확하고 신뢰할 만한 인터넷매체를 구독할 필요가 있다.

본 논문에서는 다양한 인터넷뉴스 매체들 가운데 <오마이뉴스>에 주목하였다. 오마이뉴스는 1999년 창간호 발행을 시작으로 하여 2000년에는 주식회사로 설립되었으며, 오늘날에는 국내 독립 인터넷뉴스를 대표하는 젊은 대안언론으로 평가받고 있다. ‘모든 시민은 기자다.’라는 모토를 가지고 창간하였으며, ‘열린 진보’를 편집 철학으로 한다. 그리고 ‘또 하나의 NGO’라는 말로써 운영방식을 설명한다.¹⁶⁾ 특히 시민기자제와 자발적 유료화는 오마이뉴스가 가진 특징적인 운영방식이며, 2003년 <시사저널>의 매체영향력 조사에서 6위를 차지하는 등 매체 내 상위권에 해당되는 언론이기도 하다. 또한 오마이뉴스가 어떤 매체보다도 인터넷뉴스의 특성이 뚜렷하게 나타나는 언론이기에 연구의 대상으로 선정하였다.

2) 오마이뉴스 기록의 특성

아카이빙을 위해 선행되어야 할 것은 무엇보다도 기록주체에서 생산되는 기록의 특성을 파악하는 것이라 할 수 있다. 오마이뉴스의 기록은 다른 매체와 구별되는 여러 가지 특성을 가지는데, 그 내용을 다섯 가지로 정리하면 다음과 같다. 첫째, 오마이뉴스의 기록은 그 생산주체의 유형이 다양하다. 즉, 상근기자를 비롯하여 다양한 직업군의 시민기자가 보도기사를 생산한다. 오마이뉴스의 보도기사는 ‘시민기자제’의 운영

16) 오연호, 『대한민국 특산품 오마이뉴스』, 휴머니스트, 2004, 297쪽.

으로 인해 절반 이상의 보도기사가 직업기자가 아닌 시민에 의해 생산된다. 이는 오마이뉴스의 가장 핵심적인 특징이며, 내부의 직업기자들만이 기사를 생산하는 타 언론사들과 구별되는 부분이라 할 수 있다.

〈표 1〉 오마이뉴스의 기록 생산주체

보도 기사 생산주체	세부분류
상근기자	정치팀
	경제팀
	사회팀
	방송팀
	사진팀
	지역팀
시민기자	다양한 직업군의 개인

〈표 1〉은 오마이뉴스의 보도기사 생산주체의 구성을 설명한다. 상근기자는 오마이뉴스 소속의 전문기자이며, 분야별로 정치, 경제, 사회, 방송, 사진, 지역의 팀으로 분류된다. 한편 시민기자들은 ‘뉴스케털라’라는 이름으로 활동하는데, 이용자들은 무료 회원가입을 하면 누구나 기사를 쓸 수 있다. 생산자 유형이 다양하다는 것은 기사 작성자의 직업군이 다양하다는 의미로 달리 해석해 볼 수 있는데, 이것은 오마이뉴스의 보도기사가 각계각층에서 나타나는 현상들을 드러내어 이용자로 하여금 사회전반에 대한 파악을 가능하도록 한다. 또한 시민기자들은 아이템 발굴, 취재, 편집과 웹페이지 업로드에 이르기까지 철저한 ‘1인 체제’로 운영된다. 이것은 기존의 매스미디어 편집국체제에서 한 가지 기사를 취급하는데 있어 취재, 편집, 사진기자로 그 역할이 분화된 것과 달리 통합적인 형태를 띠는 것이다.¹⁷⁾

17) 이은주, 「인터넷 시민미디어와 시민기자제도에 관한 연구 - <오마이뉴스>의 분석을 중심으로」, 연세대학교 영상대학원 멀티미디어저널리즘전공 석사학위논문

두 번째 특성은 초 단위의 생산시점이다. 오마이뉴스의 회원은 홈페이지에서 기사쓰기 페이지를 이용하여 시시때때로 기사를 작성할 수 있다. 따라서 기사 기록의 생산시점을 최초로 서버에 등록되는 시점을 기준으로 판단한다고 할 때, 오마이뉴스 보도기사의 생산시점은 초단위로 달라진다. 잉겔기사로 분류되지 않아 메인에서 볼 수 없는 생나무기사의 경우에는 오후 8시에서 오전 8시 사이를 제외하고는 웹상에서 그 내용을 볼 수 있으며, 편집부의 검토 현황에 따라 검토전·검토중·검토완료로 표시된다.

종이신문은 뉴스 생산과정에서 기사를 일정한 시간까지 일정한 분량으로 작성하여 전달되어야 하기 때문에 이러한 시공간의 제약을 극복하기 위한 전문 직업기자와 기사작성 표준이 반드시 필요한 구조였다. 반면 오마이뉴스는 기록 생산에 있어 시간적 제약으로부터 자유롭다. 따라서 전문적인 직업기자가 아닌 일반시민도 자유롭게 기사를 송고할 수 있는 환경을 갖출 수 있었던 것이다. 이처럼 기록마다 달라지는 등록시점은 기록들과의 구별을 돕기 때문에 기록의 식별에 유용하게 작용한다. 따라서 오마이뉴스 보도기사의 생산시점은 메타데이터 설계과정에서 중요한 기술요소가 될 수 있다.

세 번째 특성은 오마이뉴스는 최소한의 게이트키퍼¹⁸⁾를 수행한다는 것이다. 게이트키퍼를 최소화할수록 구독자가 기사를 선택할 수 있는 폭이 넓어지기 때문에 이용자는 더욱 다양한 정보를 얻을 기회를 얻게 된다. 앞에서 언급했듯이 오마이뉴스의 기사는 생나무 기사와 잉겔 기사로 분류되는데 2003년에는 오마이뉴스에 의해 ‘생나무’와 ‘잉겔’이라는 용어가 특허상품으로 지정되기도 하였다. 잉겔 뉴스는 편집부의 검토

문, 2002, 44쪽.

18) 게이트키퍼(Gatekeeping)이란 ‘문지기 효과’라는 의미로, 정보의 유통 흐름에 있어서 특정 단계에서 그 정보의 통과 여부를 결정하는 것을 뜻한다(구본권, 「인터넷언론의 게이트키퍼 구조와 특성」, 한양대학교 언론정보대학원 석사학위논문, 2005, 6쪽).

를 거쳐 공식기사로 채택된 것으로, 이때 비로소 원고료가 발생한다.¹⁹⁾ 즉, 생나무는 뉴스가치가 없는 삭제 대상의 의미가 아니라, 잉결의 가능성을 가진 후보기사라고 할 수 있다.

오마이뉴스는 기본적으로 작성기사 비삭제를 원칙으로 하기 때문에 편집부에 의해서 검토된 기사들은 모두 생나무 혹은 잉결로서 서버에 존재하게 된다. 과거 종이신문은 기자가 작성한 원고를 편집부에서 내부 규정에 따라 걸러낸 뒤에 구독자에게 전달되었는데, 이러한 게이트키퍼 과정은 편집주체의 주관적인 성향이 개입될 여지가 많아 비판을 받기도 한다. 그러나 오마이뉴스에서 이루어지는 게이트키퍼는 걸러진 기사까지도 구독자에게 전달됨으로써 구독자 스스로가 게이트키퍼가 되도록 하는 것이다. 즉 게이트키퍼가 최소한으로 수행됨으로써 구독자로 하여금 구독기사의 선택 기회를 확장시킨 것이라고 할 수 있다.

네 번째 특성은 기사 형식의 파괴이다. 오마이뉴스는 기사의 형식을 일정한 규격으로 한정하지 않기 때문에 기사체 뿐만 아니라 대화체, 편지체 등으로 다양하게 보도된다. 내용의 분류를 10개로 나누고 있지만 이것은 편집진 내부의 검토를 통한 분류일 뿐이고, 기사의 작성자는 이를 염두에 두지 않고 자유롭게 기사를 작성할 수 있다.

이와 같은 내용 형식의 파괴뿐 아니라 기술(技術)적 형식도 다양하게 나타난다. 오늘날 인터넷신문에서는 기사에 대한 흥미와 이해를 높이기 위해서 멀티미디어 자료를 첨부한다. 종이신문에서는 글과 그림·사진만으로 작성되었지만, 시공간의 제약이 사라진 인터넷뉴스에서는 글과 이미지뿐 아니라 오디오와 동영상을 첨부하여 작성하고 있다. 오마이뉴스에서는 시민기자도 멀티미디어 자료를 첨부하여 기사를 작성할 수 있도록 하였으며, 오마이포토, 오마이TV 섹션에서는 글을 쓰지 않고도 사진이나 동영상만을 등록하여 공유할 수 있도록 하였다.

19) 오연호, 앞의 책, 2004, 50-51쪽.

〈표 2〉 오마이뉴스의 기사 규정²⁰⁾

형식	규정 내용
글	200자 원고지 20매 내외 권장
이미지	파일형식: jpg, gif 이미지 크기: 가로 500픽셀, 세로 사진은 가로 300픽셀 한 기사에 이미지 10개로 제한 사진 설명과 태그는 필수입력 한 기사에서 동일 사진 중복 불가 이미지 슬라이드 기능
오디오	배경음악 기능
동영상	파일형식: avi, wmv, asf, mpeg, mov, dv, mp4, 3gp 파일 크기: 최대 100mb 한 기사에 동일 영상 중복 불가

<표 2>는 오마이뉴스의 기사작성 규정이다. 표에서도 볼 수 있듯이 글의 분량에 관한 규정은 구독에 알맞은 양을 권장하는 수준이고, 내용적 형식에 대한 규정이 정해져 있지 않아 자유로운 형식으로 기사를 작성하도록 하였다. 또한 이미지, 오디오, 동영상과 같은 자료를 필요한 위치에 첨부하도록 하고 있으며, 멀티미디어에 대한 파일형식이나 용량 등에 대한 규정이 있을 뿐이다.

다섯 번째 특성은 독자의견란에서 찾아볼 수 있다. 오마이뉴스는 모든 기사마다 익명과 실명 모두를 사용해 독자의견을 쓸 수 있도록 한다. 네티즌의 독자의견은 종이신문의 그것과는 다른 몇 가지 특징²¹⁾을 갖고 있다. 첫째, 즉각성이 있다. 의견을 올리고 싶을 때 바로 올릴 수 있다. 둘째, 쌍방향성이 있다. 독자의견에 제2의 독자가 또 다른 의견을 달거나 독자와 기자와의 대화의 광장이 될 수도 있다. 셋째, 집단성이

20) 오마이뉴스 홈페이지 이용안내(http://www.ohmynews.com/NWS_Web/help/srv/h_sub01.aspx?GB=101&REF=13&PgNo=2) 참고.

21) 오연호, 앞의 책, 2004, 147-148쪽.

있다. 독자의견이 많을수록 시너지 효과는 더 강해진다. 여론이 여론을 낳는다는 말은 여기에서도 가능하다. 넷째, 솔직성이 있다. 익명으로도 독자의견을 쓰는 것이 가능하기 때문에 자신의 솔직한 입장을 드러낼 수 있다. 다섯째, 때로는 전문성이 있다. 독자의견을 남겨놓은 사람들 가운데는 그 분야에 전문적인 이해를 갖고 있는 이들도 적지 않다. 나열한 특징에서 볼 수 있듯이 네티즌의 독자의견은 기사의 본문내용 만큼이나 많은 정보를 담고 있으며, 현대인의 다양한 양상을 살펴볼 수 있는 중요한 자료가 된다. 따라서 독자의견 혹은 댓글에 등록되는 글 역시 기사에 대한 기록관리 과정에서 빠뜨릴 수 없는 요소이다.

3) 오마이뉴스 기록의 메타데이터 재설계 필요성

오마이뉴스가 기사의 내용과 출처, 그리고 편집내역에 대한 메타데이터를 재설계함으로써 여러 가지 효과를 얻을 수 있는데, 언론사 내부에서 발생하는 내적 효과와 언론사 외부에서 발생하는 외적 효과로 나눠보면 다음과 같다.

오마이뉴스 내부에서 얻을 수 효과는 첫째로, 보도기사가 신뢰성을 갖추으로써 경쟁력을 얻게 된다는 것이다. 인터넷언론에 대하여 평가할 때, 취약점으로 빠지지 않고 등장하는 것이 바로 인터넷언론의 ‘낮은 신뢰도’이다. 구본권(2005)은 인터넷언론의 신뢰도가 낮게 나타나는 원인에 대해 ① 게이트키퍼 시스템이 생성중인 신생언론, ② 전문 인력의 부족, 비전문 인력, ③ 수익모델의 부재, ④ 시민기자에 의한 기사 작성의 문제, ⑤ 사실과 의견의 결합, ⑥ 수정 가능성, ⑦ 독자 영합적 편집 관행, ⑧ 속보에 밀려난 정확성이라고 분석하였는데,²²⁾ ④에서 ⑧까지의 문제, 즉 시민기자에 의한 기사 작성의 문제, 사실과 의견의 결합, 수정 가능성, 독자 영합적 편집관행, 속보에 밀려난 정확성은 뉴스 기

22) 구본권, 앞의 논문, 2005, 29-41쪽.

사 자체의 문제이며, 기사의 내용과 정보원, 수정내역에 대한 데이터를 통제함으로써 문제를 완화시킬 수 있다. 이와 같은 정보를 통제할 수 있도록 메타데이터를 설계한다면 기사 내용의 정확성이 보완될 수 있으며 언론에 대한 신뢰를 높일 수 있을 것이다.

두 번째 내적 효과는 방대한 양의 등록 기사를 효율적으로 관리할 수 있다는 것이다. 오마이뉴스에는 하루에 약 200개의 기사가 오른다. 그 가운데 약 150개는 시민기자가, 나머지는 상근기자가 올린다. 모든 기사는 예외 없이 생나무 뉴스 상태에서 편집부에 의해 검토된다.²³⁾ 이렇게 축적된 기사의 양은 상근기자로만 운영되는 타 언론사에 비해 훨씬 많은 양일뿐더러, 오마이뉴스는 등록된 기사를 삭제하지 않는 것을 원칙으로 하기 때문에 축적된 기사 데이터를 보존하는데 있어서 보다 체계적인 관리가 필요하다. 적절한 기술 요소를 갖춘 메타데이터를 구축하는 것은 기사 데이터에 이용가능성을 부가할 것이다. 효과적인 메타데이터와 함께 보존된 이전 기사들은 현재 송고되는 기사를 검토하는 경우나 이용자가 요청한 이전 기사를 서비스할 경우 등 언론사 내부에서 발생하는 상황에서 유용한 참고 자료가 될 것이라 예상된다.

다음으로 언론사 외부에서 얻을 수 있는 효과를 살펴보면 다음과 같다. 첫째, 이용자가 원하는 기사에 정확히 접근할 수 있다. 메타데이터를 통해 나타나는 이용가능성의 증가는 언론사 내부뿐만 아니라 외부의 이용자들에게 유용하게 작용할 것이다. 검색은 주로 기사 입력시 등록되는 사항이 접근점이 되어 이루어진다. 그러나 입력되는 사항 이외에 기사 데이터에 대한 다각적인 접근점이 존재한다면 더욱 정확한 기사 검색이 가능할 것이다.

두 번째 외적 효과는 이용자에게 풍부한 정보를 제공할 수 있다는 것이다. 메타데이터는 기록에 대한 각종 데이터를 포착함으로써 기록의 이용가능성을 최대화하는 기능을 수행한다.²⁴⁾ 관련 기록을 연결시킨다

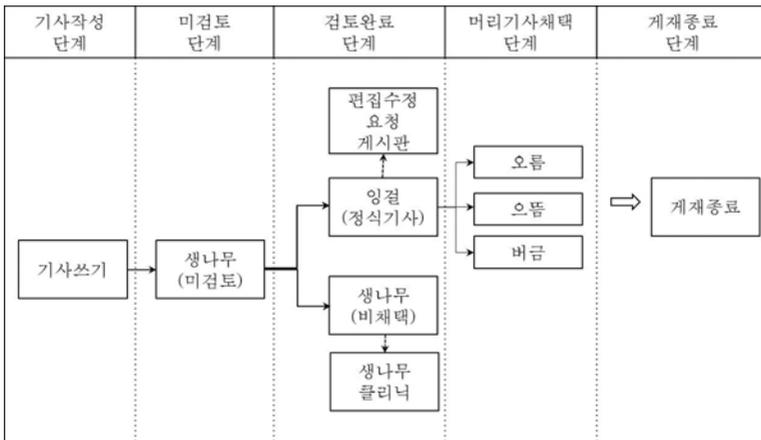
23) 오연호, 앞의 책, 2004, 51쪽.

는 점에서 메타데이터는 수용자가 온라인상에서 뉴스 기사를 검색할 때에 유용하게 작용할 수 있다. 즉, 수용자는 검색하고자 하는 기사의 메타데이터를 이용하여 기사의 원자료뿐만 아니라 관련되는 기사를 획득함으로써 보다 풍부한 정보를 얻을 수 있다는 것이다.

3. 오마이뉴스의 기록 생산과 메타데이터

1) 보도기사 기록생산 프로세스와 메타데이터 현황

<그림 1> 오마이뉴스의 기사처리과정²⁵⁾



<그림 1>은 이용안내에서 공개하고 있는 기사처리과정이다. 각 단계

24) 김익한, 앞의 논문, 2004, 9쪽.

25) 오마이뉴스 홈페이지 이용안내(http://www.ohmynews.com/NWS_Web/Help/srv/h_articleprocess.aspx) 참고.

마다 기록에 대한 정보가 생성되는데, 단계별로 생성되는 기록정보와 그에 해당되는 기술요소를 살펴봄으로써 메타데이터의 구체적인 구성을 그려볼 수 있을 것이다. 해당 메타데이터 요소는 현재 운영되고 있는 우리나라 공공표준 메타데이터의 상위요소를 바탕으로 하고, 기사작성 단계를 시작으로 그 내용을 살펴보도록 하겠다.

〈표 3〉 기사작성 단계의 기록정보와 해당 기술요소

단계	기록정보	입력	해당 메타데이터 요소 ²⁶⁾	필수여부
기사작성 단계	제목	필수	기록명(Title)	필수
	부제	선택		필수
	내용	필수	기술(Description)	선택
	태그	필수		선택
	덧붙이는 글	선택		선택
	파일첨부	선택	관계(Relation)	필수
	취재경위	필수	생산이력(Business History)	필수
	기본분류	필수	분류(classification)	필수
	지역분류	선택		필수
	연재기획	선택	-	

<표 3>은 기사작성 단계에서 입력하는 사항들과 필수입력 여부에 대한 내용이다. 입력사항 중에서도 제목, 내용, 태그, 취재경위, 보도매체 선택, 기본분류는 필수 입력 사항임을 알 수 있다. 기사 내용은 작성란 상단에 있는 도구를 이용하여 글의 모양과 링크를 편집할 수 있으며, 기사 중간에 사진이나 동영상을 첨부하여 편집이 가능하다. 이때 사진이나 동영상 등의 멀티미디어 콘텐츠에 대한 파일설명, 저작년월일, 저작권자, 태그의 필수 입력이 요구된다. 태그는 기사의 키워드를 선택하여 입력하면 그 단어가 접근점이 되어 기사 검색 시 이용된다. 또한 필

26) 해당 메타데이터 요소는 국가 공공표준의 상위요소로 한다.

수입력 사항 중 취재경위에는 직접취재, 경험담, 보도자료 참고 등의 취재 방법을 기재한다. 이 내용은 편집부만 볼 수 있으며, 기사 검토의 참고 사항이 된다. 그리고 기사분류 중 기본분류만이 필수입력이 요구되는데, 기본분류는 사는 이야기, 사회, 경제, 정치, 문화, 민속·국제, 교육, 스포츠, 책동네, 여행, 영화, 미디어, 여성 13개 분류항목으로 구성되어 있다. 이러한 필수사항을 입력함으로써 기사 등록이 가능해지며, 등록된 기사는 편집부로 송고된다.

기사쓰기 단계의 입력항목을 메타데이터 요소와 비교했을 때, 기사의 내용에 해당하는 항목이 메타데이터 요소 중에서 기술(Description)에 해당하는 것을 볼 수 있다. 주목할 점은 기사의 내용항목은 필수입력사항인데 반해 메타데이터의 기술(Description)은 선택사항으로 규정하였다는 것이다. 보도기사는 기록 자체의 내용이 기록정보 중에서도 큰 중요성을 갖기 때문에 현행 공공표준상의 기술(Description) 영역과는 달리 필수사항이 되어야 한다는 점을 고려하여 메타데이터 재설계시 적용해야 할 것이다.

〈표 4〉 미검토 단계의 기록정보와 해당 기술요소

단계	기록정보	해당 메타데이터 요소	필수여부
미검토 단계	송고일시	일시(Date), 생산이력(Creation History)	필수
	기자명	생산자(Agent), 생산이력(Creation History)	필수
	기자분류	생산자(Agent), 생산이력(Creation History)	필수
	첨부파일형식	포맷(Format)	필수
	첨부파일용량	크기(Size)	필수

미검토 단계는 기사 원고가 편집부에 송고된 상태로, 이 단계의 원고

는 생나무 상태의 기사이다. 생나무는 메인화면 하단 ‘생나무 목록’에서 볼 수 있다. 편집부가 검토하지 않은 기사는 ‘검토전’으로 표시되며 사내에서 기사를 검토하는 중에는 ‘검토중’으로, 검토가 끝났을 때는 ‘검토완료’로 표시된다. 생나무 검토 시간은 오전 8시부터 오후 8시까지(연중무휴)이다. 미검토 단계에서는 송고된 날짜와 시간, 기자명, 그리고 첨부파일의 형식과 용량 등의 기록정보가 발생할 수 있다. 이 단계는 기사의 데이터가 단지 작성자에게서 편집부로 이동되어있는 상태이므로 기사 처리 과정 중에서 발생할 수 있는 기록정보 요소가 많지 않다.

미검토 단계는 기사 데이터가 최초로 서버에 등록된 상태라고도 할 수 있다. 따라서 메타데이터 요소 중에서 생산과 관련된 요소들, 즉 생산이력(Creation History)이나 생산자(Agent)에 관련된 기록정보가 다수 생성된다.

〈표 5〉 검토완료 단계의 기록정보 요소

단계	기록정보		해당 메타데이터 요소	필수여부
검토완료 단계	검토완료일시		관리이력 (Management History)	필수
	검토자			
	검토결과			
	결과사유			
	원고료지급여부			
	기사배치일시			
	기사배치위치			
	편집수정 요청	편집수정요청일시	관리이력 (Management History), 관계(Relation)	필수
		편집수정담당자		
		편집수정요청내용		
		편집수정첨부파일		
		수정일시		
수정결과				
생나무	생나무클리닉의뢰일시	관리이력	필수	

	클리닉	클리닉의뢰공개여부	(Management History), 관계(Relation)	
		클리닉의뢰내용		
		클리닉의뢰상태		
		클리닉소견작성자		
		클리닉소견작성일시		
		클리닉소견내용		

검토 후의 기사는 정식기사와 비채택기사로 분류되며, 정식기사로 채택된 기사가 바로 잉결기사이다. 잉결 기사는 2천원의 원고료가 지급되며 메인면 하단 잉결 뉴스에서 볼 수 있다. 검토 후의 생나무기사는 편집부가 정식 기사로 채택하지 않은 글이다. 정해진 기준이 존재하는 것이 아니라 기사의 전반적인 사항을 검토하기 때문에 비채택의 원인은 다양하게 나타난다. 문장이 부적절하거나 기사 가치가 낮을 때, 논리가 맞지 않을 때, 중복기사이거나 단순 외신 번역일 경우, 주장성 글이 지나치게 감정적이거나 주장이 분명하지 않을 때, 취재의 사실 관계가 명확하지 않을 때, 기자가 일방의 편을 들고 있을 때는 정식 기사로 채택하지 않는다. 오마이뉴스는 생나무 클리닉 페이지를 두어 상근기사와 다른 시민기사로부터 정식기사로 채택되지 않은 비채택 생나무에 대한 조언을 받을 수 있도록 하였다. 검토완료 단계에서는 검토완료 날짜와 시간, 검토자, 검토결과, 결과사유, 원고료 지급 여부, 그리고 기사의 배치시점과 화면상에서 배치된 위치가 기록에 대한 정보로 포함될 수 있다. 검토가 완료된 기사는 작성자가 직접 수정할 수 없으며, 편집 수정요청 게시판을 통해 수정을 요청해야 한다. 또한 비채택된 기사의 작성자가 생나무 클리닉을 통해 본인의 기사에 대한 조언을 의뢰하는 경우 의뢰 공개·비공개, 요청내용, 의뢰상태(의뢰·검토중·완료·보류·의뢰취소), 소견내용의 기록정보가 발생하게 된다.

검토완료 단계는 기록물의 관리과정에 해당되기 때문에 주로 관리이력(Management History)과 관련된 정보가 생성된다. 그리고 이 단계에서

놓치지 말아야 할 것은 편집수정요청 기록과 생나무클리닉 기록이다. 메타데이터 설계시 관계(Relation)요소들을 기술함으로써 해당 보도기사에 관련된 편집수정요청 기록, 생나무클리닉 기록을 연결시켜야 한다.

〈표 6〉 머리기사채택 단계의 기록정보 요소

단계	기록정보	해당 메타데이터 요소	필수여부
머리기사채택 단계	채택일시	관리이력 (Management History)	필수
	채택 담당자		
	채택등급		
	채택사유		
	원고료 지급여부	—	

잉겔로 채택된 기사는 그 뉴스가치에 따라 오름, 으뜸, 버금으로 분류되어 메인화면에 게재된다. 오름기사는 메인면 머리기사 1부터 3까지 배치된 기사로, 머리기사 1은 5만원, 머리기사 2, 3은 3만원의 원고료가 지급된다. 다음으로 으뜸기사는 오름 아래에 배치되는 기사이며, 으뜸은 2만4천원의 원고료가 지급된다. 마지막 버금기사는 메인면 으뜸 아래에 배치된 기사와 이미지 뉴스, 지역 맛보기, 사는이야기 맛보기에 배치된 기사이며, 1만2000원의 원고료가 지급된다.²⁷⁾ 머리기사로 채택된 기사는 채택일시, 담당자, 오름·으뜸·버금의 채택등급, 채택사유, 원고료 지급 여부 등이 기록정보 요소로 발생된다.

머리기사채택 단계는 생나무기사에서 잉겔기사로 채택된 원고에만 해당되는 단계이기 때문에 모든 기사원고에 적용되지 않는다는 사실을 고려해야 하며, 이와 같은 부분은 메타데이터 설계에 있어서 ‘해당시 필수’사항으로 규정할 수 있을 것이다.

27) 오마이뉴스 홈페이지 이용안내(http://www.ohmynews.com/NWS_Web/help/srv/h_sub01.aspx?GB=101&REF=14) 참고.

〈표 7〉 게재종료 단계의 기록정보 요소

단계	기록정보	해당 메타데이터 요소	필수여부
게재종료 단계	보도종료일시	일시(Date)	필수
	담당자	관리이력(Management History)	필수
	저장위치	위치(Location)	필수
	원고료지급내역	-	
	연관기사	관계(Relation)	필수
	조회수	이용이력(Use History)	필수
	댓글		
	기사점수		
	기사공유내역		
	검색내역		

게재종료 단계는 해당 기사가 더 이상 메인화면에 게재되지 않는 단계이다. 게재가 끝난 기사는 메인화면에서는 볼 수 없지만 통합검색을 이용하여 찾아볼 수 있다. 게재가 종료되더라도 검색이 가능하기 때문에 기사에 대한 피드백은 여전히 진행중에 있다고 할 수 있으며, 갱신되는 조회수, 댓글, 공유, 검색내역이 이에 해당한다. 이러한 사항은 모두 기사 기록정보에 포함된다.

이 단계에서는 데이터의 위치(Location), 이용이력(Use History)에 해당되는 정보를 얻을 수 있다. 보도기사는 웹서버에 존재하는 웹기록물에 해당하기 때문에 위치(Location)요소는 기본식별자로 URL 주소 등을 표기하여 기술할 수 있을 것이다. 또한 이용이력(Use History)요소의 경우 검색을 통해 끊임없이 갱신되는 조회수, 댓글, 기사점수, 공유내역을 갖는 인터넷뉴스의 특성을 고려하여 기술되어야 하며, 지속적인 관리가 필요한 부분이다.

이처럼 오마이뉴스의 기사처리과정에서는 각 단계에 따라 기사에 대한 다양한 기록정보가 발생하며, 이렇게 발생하는 정보들은 메타데이터

재설계를 위한 기술요소가 된다. 그렇다면 현재 오마이뉴스는 기사원고 혹은 보도된 기사의 데이터를 어떻게 관리하고 있으며, 어떠한 정보를 추출하여 이용하고 있는지 알아볼 필요가 있다.

오마이뉴스는 등록 기사 비삭제를 원칙으로 하기 때문에 게재된 기사를 그대로 웹서버에 보존한다. 편집부로 송고되는 기사는 상근기자의 원고를 포함하여 하루 200개 정도이다. 이렇듯 많은 양의 기사 기록이 발생하지만 보존을 위한 메타데이터 생산은 제대로 이루어지지 않고 있으며, DB 저장 시스템에 의존하고 있는 실정이다.

현재 오마이뉴스가 기사 데이터를 통제하기 위해 관리하고 있는 기록정보 항목을 파악하기 위하여 언론사내에서 기사 검색시 이용하는 검색항목을 살펴보도록 하겠다.

〈그림 2〉 사내 기사검색화면²⁸⁾

〈그림 2〉은 오마이뉴스 내에서 편집진이 기사를 검색할 때에 볼 수 있는 화면이다. 이를 통해 어떠한 기록정보를 이용하여 기사 데이터를 통제하고 있는지 짐작해볼 수 있다. 검색항목은 기본분류, 지역분류, 형식분류, 검토상태, 승인상태, 기사등급, 기자유형, 전송일, 전송일시, 채택일시 그리고 단어 검색창으로 이루어져 있다. 분류별 항목을 자세히 살펴보면 다음과 같다.

28) 2012년 3월 12일 오마이뉴스 기획팀 김○○ 서면 인터뷰 답변.

〈표 8〉 분류별 항목

분류	세부항목
기본분류	사은이아기/사회/경제/정치/문화/민족/국제/교육/스포츠/책동네/여행/영화/미디어/여성
지역분류	광주전라/대구경북/부산경남/대전충남/인천경기/강원제주
형식분류	정보/주장/사건/고발/일반/예고/르포/인물인터뷰/동정/기타
검토상태	검토전/검토중/검토완료
승인상태	승인전/승인완료
기사등급	옴/오름/버금/잉걸/생나무
기자유형	상근기자/시민기자
전송일	

오마이뉴스는 이와 같은 분류별 항목에 따라 기사 데이터를 분류하여 기록정보를 통제하고 있다. 분류별 항목 외에 이용되는 단어 검색창은 기사작성 단계에서 입력되는 태그를 포함하여 기사를 검색한다. 종합해보면, 기사 데이터의 통제는 주로 기사 작성시 입력되는 정보에 의존하고 있다. 사실상 이러한 기사 검색조차 오마이뉴스 언론사내에서만 가능하고, 외부 이용자들은 정확한 기본분류, 지역분류, 태그 정보를 알지 못할 경우 이전 기사를 검색하기가 쉽지 않다. 이 외에 오마이뉴스가 가진 구조적인 문제점은 콘텐츠로서의 관리부서, 데이터로서의 관리부서가 각각 관리하고 있기 때문에 보도기사의 통합적인 메타데이터를 생산하기가 어렵다는 것이다. 기사의 내용적 측면과 형식적 측면에서의 정보를 통합적으로 관리할 수 있는 구조적 개선이 필요하다.

지나간 기사 데이터도 관리수준에 따라 그 가치를 증가시킬 수 있음에도 불구하고 현실은 단순히 보존상태를 유지하는 수준에 그치고 있다. 먼저 보도기사에 대한 체계적인 관리 필요성을 인식하는 것이 중요할 것이다.

2) 메타데이터 표준 분석

오마이뉴스의 보도기사 관리를 위한 메타데이터 구성을 위하여 현재 운영되고 있는 국내외 메타데이터 표준을 파악하는 과정이 필요하다. 먼저 국제표준에 입각하여 설계된 국내 공공기관의 메타데이터 구성을 알아봄으로써 메타데이터가 갖추어야 할 기본적인 틀을 구상해볼 수 있다. 그리고 오마이뉴스의 기사 콘텐츠와 같은 웹 자원을 관리하는 메타데이터 표준을 분석하여 메타데이터 재설계시 세부 기술요소 추출에 참고하려 한다. 국내 메타데이터 표준으로는 2010년 10월에 발표된 기록관리 메타데이터 공공표준안을 대상으로 하며, 외국 메타데이터 표준은 보존 메타데이터를 위한 PREMIS 데이터 사전을 대상으로 한다.

① 국내 기록관리 메타데이터 공공표준

2010년 10월 국가기록원에서 『기록관리 메타데이터 국가표준화 연구용역 최종보고서²⁹⁾』가 발간되었다. 근래의 공공표준이 적용된 “표준 기록관리시스템” 보급이 2~3년이 경과함에 따라 메타데이터 요소들의 적절성 평가 시점의 도래하였으며, 기존 공공표준이 전통적 문서의 등가물에 해당하는 문서 형태의 전자기록에 초점을 두어 개발되어 웹 기록이나 데이터 세트 기록, 기타 다양한 유형의 시청각기록의 특성을 충분히 기술할 수 있도록 확장의 필요성 인식이 있었다. 또한 기타 공공기관의 기록관리시스템 개발, 기업 영역에서의 전자기록관리 표준화가 진전됨에 따라 기록관리의 핵심 메커니즘인 메타데이터의 국가표준화 방향 정립과 준비가 필요하였다.

이러한 배경에서 제시된 메타데이터 표준안은 그 적용 범위를 특별한 형태나 매체와 관련된 표준이 별도로 제정되지 않는 한 기록의 형태

29) 국가기록원, 『기록관리 메타데이터 국가표준화 연구용역 최종보고서』, 2010.

나 매체와 상관없이 모든 기록을 대상으로 한다. 그리고 현용·준현용 단계의 기록에 적용할 수 있는 메타데이터이므로 전자기록생산시스템, 기록관리시스템을 구축·운영하고자 할 때 적용해야 하는 표준이지만, 장기보존을 위한 요소도 고려한다.

표준안은 근래의 공공표준에서 제기되는 문제점을 보완하면서 실행 적합성과 상호운용성을 고려하였으며, 국제표준과의 격차를 줄여 호환이 용이하도록 하는 방안을 연구하였다. 특히 메타데이터 요소에 있어서 국제 표준 및 해외표준에서와 같이 각 요소에 대한 의무수준, 적용계층, 할당값 등에 대한 설명체계를 개선하였다. 의무수준에 대한 ‘해당시 필수’ 요소의 경우 작성방법에서 어떤 경우 반드시 사용하고, 필요에 따라 생산시스템에서 이미 작성한 것을 어떤 방식으로 가져올 수 있는지에 대해 제시하며 다양한 선택값과 사용례를 추가하였다. 또한 폐기된 기록에 대한 메타데이터 요소 중 관리이력에서 폐기 이외의 관리이력정보는 남기지 않고, 폐기와 관련된 관리이력정보만을 남겨 효율적인 메타데이터를 관리할 수 있도록 제시하였다.

메타데이터 국가표준화 사업의 공공표준안에서 제시된 메타데이터 요소의 구성은 다음에 나오는 <표 9>와 같다.

<표 9> 2010년 메타데이터 공공표준안의 기술요소구성

상위요소	하위요소	필수여부	비고
1. 생산자 (Creator)	1.1 기관명	필수	
	1.2 기관코드	해당시	행정표준관리시스템의 기관코드
	1.3 개인명	해당시	
	1.4 개인코드	해당시	생산자 ID
	1.5 부서명	해당시	
	1.6 부서코드	해당시	행정표준관리시스템의 부서코드
	1.7 직위명	선택	
	1.8 직위코드	선택	로그인정보에 의해 자동부여

2. 기록계층 (Aggregation Level)		필수	기록물철/기록물건/첨부
3. 고유식별자 (Identifier)	3.1 기본식별자	필수	RMS에서 자동생성된 ID
	3.2 참조식별자	해당시	기록물철분류번호/기록물건등록번호/ 첨부일련번호
	3.3 보조식별자	해당시	URL주소/ISBN번호/ISSN번호 등
4. 기록명(Title)	4.1 공식제목명	필수	기록물철과제명/기록물건표제명
	4.2 부제목명	해당시	
	4.3 기타 제목	해당시	총서명/부가제목/정정제목/영문제목 등
5. 기술 (Description)		해당시	-초록이나 요약문으로 자유롭게 기술 -기록물철일 경우 단위과제카드의 내용 및 취지, 기록물건일 경우 문서관리카드의 문서취지, 메모보고의 보고내용 등
6. 주제 (Subject)	6.1 주제유형	필수	일반주제명/인명/단체명/지명
	6.2 주제명	필수	
7. 유형 (Type)	7.1 기록유형	필수	일반문서/간행물/카드/도면/시청각기 록/웹기록/데이터세트/전자우편/행정 박물
	7.2 사본유형	해당시	대체사본/보존사본/열람사본/발췌사본
8. 포맷 (Format)	8.1 포맷명	해당시	HWP/TXT/XML/PDF/ZIP 등
	8.2 포맷버전	해당시	
	8.3 생성어플리케이션명	해당시	Adobe Acrobat 등
	8.4 생성어플리케이션버전	해당시	
9. 저장매체 (Medium)		필수	종이류/마이크로필름/CD/하드디스크 등
10. 크기 (Extent)	10.1 용량	필수	
	10.2 단위	필수	KB/MB/GB/장/쪽/건/분 등
11. 분류 (Classification)	11.1 분류체계 유형	필수	BRM/LBRM/기록물분류기준/공문서분 류체계/문서분류체계/기타
	11.2 분류값	필수	분류ID/분류명/분류계층

12. 일시(Date)	12.1 생산일시	필수	
	12.2 철시작일시	필수	
	12.3 철종료일시	해당시	
13. 생산이력 (Business History)	13.1 전자기록생산시스템	해당시	업무관리시스템/전자문서시스템/행정정보시스템/기록관리시스템/자료관리시스템/기타
	13.2 생산근거	해당시	
	13.3 비치기록	해당시	
14. 보존기간 (Retention)	14.1 보존기간	필수	1년/3년/5년/10년/30년/준영구/영구
	14.2 보존기간책정사유	선택	
15. 권한 (Rights)	15.1 보안	해당시	비밀기록에만 해당. 비밀분류/근거/기간
	15.2 접근권한	필수	기관/부서/개인단위로 접근권한부여, 목록열람/내용열람/다운로드/복사/접근제한없음
	15.3 공개	필수	공개/비공개/부분공개, 비공개근거, 공개제한부분, 공개예정일시, 공개관련근거
16. 위치 (Location)	16.1 소장처	필수	소장기관명
	16.2 소장위치	필수	서고번호, 서가번호, 관리번호, 서가배치번호, 보존상자번호 전자기록일 경우 저장된 스토리지
17. 관리이력 (Management History)	17.1 관리유형	필수	인수인계/인수/이관/공개재분류/보존기간재책정/보류/폐기/재분류/기술/비밀재분류/접근권한재책정/발체사본생성/수집기록등록/기타
	17.2 관리내용	해당시	
	17.3 관리발생일시	필수	
	17.4 관리행위주체	필수	기관명/기관코드/개인명/개인코드/부서명/부서코드/직위명/직위코드
	17.5 관리 근거	필수	해당 법규/지시사항/정책/표준/지침 등
	17.6 변경요소	해당시	변경요소/변경이전값

18. 이용이력 (Use History)	18.1 이용유형	필수	전체열람/다운로드/복사/허가받지않 은접근
	18.2 이용내용설명	해당시	
	18.3 이용일시	필수	
	18.4 이용자	필수	기관명/기관코드/개인명/개인코드/부 서명/부서코드/직위명/직위코드
19. 보존이력 (Preservation History)	19.1 보존처리유형	필수	매체수록/디지털변환/포맷변환/소독/ 상태평가/정수점검/제본/탈산/복원/기타
	19.2 보존처리설명	해당시	
	19.3 보존처리일시	필수	
	19.4 보존행위주체	해당시	기관명/기관코드/개인명/개인코드/부 서명/부서코드/직위명/직위코드
20. 관계 (Relation)	20.1 관계유형	필수	포함/선행/후행/사본/원본/포맷변환/ 참조/기타
	20.2 관계대상식별자	필수	RMS에서 자동생산된 식별자
	20.3 관계설명	선택	
21. 언어 (Language)	21.1 언어스키마	필수	ISO639-2/RFC3066
	21.2 언어코드	필수	kor
22. 무결성체크 (Integrity Check)	22.1 무결성체크기법	필수	기관 내 무결성체크기법 예) 해쉬함수MD5
	22.2 무결성체크값	필수	시스템 자동부여

출전 : 국가기록원, 『기록관리 메타데이터 국가표준화 연구용역 최종보고서』, 2010, 참고하여 재구성함.

제시된 공공표준안의 몇몇 특징은 오마이뉴스 메타데이터의 구조 설계에 적용할 수 있으며, 그 내용은 세 가지로 정리해 볼 수 있다. 첫째, 현행 표준에서 행위주체(Agent)로 기술되던 부분이 표준안에서는 생산자(Creator)로 변경되었다. 생산자(Creator)는 오직 기록의 생산단계에 관련되는 정보이며, 이후 기록에 행해지는 관리·이용·보존 등의 행위주체는 각각 관리이력(Management History), 이용이력(Use History), 보존이

력(Preservation)의 하위요소로 행위주체를 두어 그에 대한 정보를 기술하도록 하였다. 오마이뉴스의 보도기사 데이터도 역시 생산단계 이후에 기록에 연관되는 행위주체가 다양하게 나타나기 때문에, 이렇게 생산자 요소를 따로 두고 관리·이용·보존이력 요소에 행위주체를 하위요소로 기술하는 방법을 적용할 필요가 있다.

둘째, 주제(Subject)요소에 현행 표준에는 없었던 주제유형(Subject Type)을 추가하였다. 현행 표준이 주제명만을 기술하여 나열하였다면 표준안은 주제유형을 구분하여 검색의 효율성과 접근점을 확대하려고 한 것이다. 주제유형을 일반주제명, 인명, 단체명, 지명으로 분류하기 때문에 이용자로 하여금 원하는 기록물을 빠르고 정확하게 검색하도록 할 수 있다. 주제요소와 그 하위요소들은 보도기사 기술에 있어서 빠질 수 없는 요소라고 할 수 있는데, 기록물의 주제와 주제유형 요소는 보도기사의 주제와 기본분류와 대응시킬 수 있기 때문이다. 오마이뉴스 보도기사의 일반분류는 이미 나름의 체계가 갖추어져 있으므로 메타데이터 설계에 수월하게 적용할 수 있을 것이라 예상된다.

셋째, 관리이력(Management History)에 변경요소가 신설되었다는 점이다. 현행 공공표준에서는 관리행위가 발생할 경우 관리에 의해 변화 이전의 값을 기술할 항목이 없어 그 변경사항에 대한 처리과정을 추적할 수 없었다. 그러나 제시된 표준안은 기록의 안정적 관리를 위해 반드시 관리행위 발생시 요소값의 변화가 생길 때 이전의 실제값을 축적시켜 또 다른 메타데이터로 관리함으로써 기록의 무결성을 획득할 수 있도록 하였다.³⁰⁾ 송고·채택단계를 거칠 뿐만 아니라, 편집수정 요청을 통해 그 내용이 수정될 수 있는 보도기사의 경우에도 관리이력을 축적하여 기술할 필요가 있다. 관리이력에서 변경과정을 기술하여 출처와 책임자를 명확히 밝힌다면 기사의 신뢰성을 증가시킬 수 있다.

오마이뉴스 보도기사의 특성과 관련하여 표준안 메타데이터 요소의

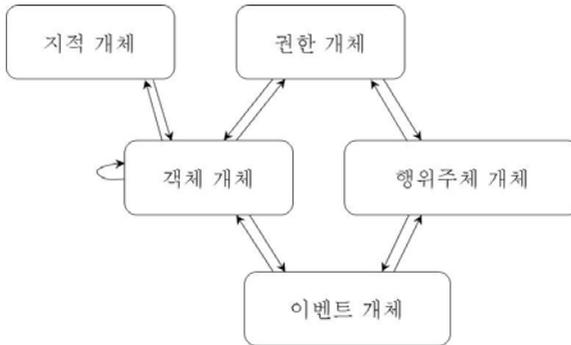
30) 국가기록원, 앞의 책, 2010, 145쪽.

세 가지 주요특성을 살펴보았다. 이 외에도 표준안의 메타데이터 요소는 현행 표준을 보완하여 무결성체크(Integrity Check)요소와 포맷(Format)의 생성어플리케이션요소 신설 등의 특이사항을 가지고 있다.

② PREMIS 데이터 사전

PREMIS는 2011년에 데이터 사전 2.1버전을 발표하여 최신의 보존 메타데이터 모델을 보여주었으며, 이는 모든 형태의 디지털 객체를 적용 대상으로 한다. 2003년 7월 OCLC와 RLG는 PREMIS(보존 메타데이터: 구성 전략) 워킹그룹의 조직을 지원하여, 전자적 보존 활동을 지원하는 메타데이터 전문가들을 구성하였다. 2005년 보존 메타데이터를 위한 데이터 사전: PREMIS 워킹그룹의 최종보고서가 발표되었으며, 2011년 2월에 발표된 2.1 버전에서는 오류목록, 일부 의미단위에 대한 부가해설, 지속성 유지를 위한 보완사항 그리고 몇몇 의미단위가 추가되었다.³¹⁾

〈그림 3〉 PREMIS의 데이터 모형



출처: PREMIS Editorial Committee, 「Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata version 2.1」, 2011, p.5

31) PREMIS Editorial Committee, 「Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata version 2.1」, 2011, pp.1-2 <<http://www.loc.gov/standards/premis/v2/premis-report-2-1.pdf>> [인용. '12.03.30.]

보존 메타데이터를 위한 PREMIS 데이터 사전은 기본적으로 지적 개체, 객체 개체, 이벤트 개체, 행위주체 개체, 권한 개체의 5개 개체로 이루어진 데이터 모형을 바탕으로 한다.³²⁾

객체 개체는 ‘디지털 형태로 된 정보의 개별 단위(a discrete unit of information in digital form)’로 정의되며 재현물과 파일, 비트스트림의 3개의 서브 타입 유형을 가진다. 행위주체 개체는 데이터 객체의 생애주기에서 보존과 관련된 활동과 권한에 관계된 행위를 하는 주체에 관한 정보로 개인을 비롯하여 조직, 컴퓨터 소프트웨어 등도 행위주체에 포함된다. 이벤트 개체는 하나 이상의 객체나 행위주체를 포함하는 활동에 관한 정보이다. 이벤트에 관한 메타데이터는 디지털 객체와 분리되어 기록되고 저장되며, 객체를 수정하는 추가 활동 등도 이벤트로 기록된다. 즉 이관, 입수와 같은 기록관리 행위가 이에 해당한다. 이벤트 개체의 요소는 이벤트가 발생하는 시기마다 그 활동에 관한 정보인 메타데이터로 획득되어야 한다. 권한 개체는 저작권이나 지적 재산권법 또는 협약에 의해 행위주체에 부여된 권한에 대한 정보이다. 권한 정보는 아카이브에서 소장하고 있는 개체와 관련하여 아카이브가 수행할 수 있는 권한과 허가를 받은 것이 무엇인가에 대한 정보라 할 수 있다. 보존 목적으로 PDF 파일의 복사본을 생산할 수 있는 권한 등이 해당된다.

5개 개체는 상호연관성을 가지며, 식별자(Identifier)를 이용하여 개체 간의 관계를 나타낸다. 이러한 데이터모형을 바탕으로 하는 PREMIS 데이터 사전은 기존의 메타데이터 표준에서 사용했던 ‘메타데이터 요소’ 대신에 ‘의미단위(semantic unit)’라는 용어를 사용한다. PREMIS 데이터 사전의 의미단위 구성은 다음 <표 10>와 같다.

32) 장보성, 「디지털 연구성과물 장기보존을 위한 메타데이터에 관한 연구」, 중앙대학교 기록관리학과 기록물관리학 박사학위논문, 2011, 83-85쪽.

〈표 10〉 PREMIS의 의미단위

	의미단위(semantic unit)	의미구성요소(semantic component)	필수여부
객체 요소 Object entity	1. 객체식별자 (object identifier)	1.1 객체식별자유형	필수
		1.2 객체식별자값	필수
	2. 객체유형 (object category)		필수
	3. 보존수준 (preservation level)	3.1 보존수준값	필수
		3.2 보존수준기능	선택
		3.3 보존수준사유	선택
		3.4 보존수준승인일시	선택
	4. 주요속성 (significant properties)	4.1 주요속성유형	선택
		4.2 주요속성값	선택
		4.3 주요속성확장정보	선택
	5. 객체특성 (object characteristics)	5.1 구성등급	필수
		5.2 고정값	선택
		5.3 크기	선택
		5.4 포맷	필수
		5.5 제작프로그램	선택
		5.6 접근제한	선택
		5.7 객체특성확장정보	선택
	6. 원객체명 (original name)		선택
	7. 저장(storage)	7.1 콘텐츠위치	필수
		7.2 저장매체	선택
	8. 환경(environment)	8.1 환경특성	선택
8.2 환경목적		선택	
8.3 환경주기		선택	
8.4 부분정보		선택	
8.5 소프트웨어		선택	
8.6 하드웨어		선택	
8.7 환경확장정보		선택	

	9. 서명정보 (signature information)	9.1 서명	필수
		9.2 서명정보확장	선택
	10. 관계(relationship)	10.1 관계유형	필수
		10.2 관계보조유형	필수
		10.3 관계된객체식별	필수
		10.4 관계된행위식별	선택
	11. 연결행위식별자 (liking event identifier)	11.1 연결행위식별자유형	필수
		11.2 연결행위식별자값	필수
	12. 연결된지적개체식별자 (linking intellectual entity identifier)	12.1 연결된지적개체식별자유형	필수
		12.2 연결된지적개체식별자값	필수
	13. 연결된권한사항식별자 (linking rights statement identifier)	13.1 연결된권한사항식별자유형	필수
		13.2 연결된권한사항식별자값	필수
	이벤트 요소 Event entity	14. 이벤트식별자 (event identifier)	14.1 이벤트식별자유형
14.2 이벤트식별자값			필수
15. 이벤트유형 (event Type)			필수
16. 이벤트일시 (event date time)			필수
17. 이벤트상세내용 (event dtail information)			선택
18. 이벤트결과정보(event outcome information)		18.1 이벤트결과	선택
		18.2 이벤트결과상세정보	선택
19. 연결된주체식별자 (linking agent identifier)		19.1 연결된주체식별자유형	필수
		19.2 연결된주체식별자값	필수
		19.3 연결된주체역할	선택
20. 연결된객체식별자 (linking object identifier)		20.1 연결된객체식별자유형	필수
		20.2 연결된객체식별자값	필수
		20.3 연결된객체역할	선택

행위주체 요소 Agent entity	21. 주체식별자 (agent identifier)	21.1 주체식별자유형	필수
		21.2 주체식별자값	필수
	22. 주체명(agent name)		선택
	23. 주체 유형 (agent type)		선택
	24. 주체 주기 (agent notes)		선택
	25. 주체확장정보 (agent extension)		선택
	26. 연결된이벤트식별자 (linking event identifier)	26.1 연결된이벤트식별자유형	필수
		26.2 연결된이벤트식별값	필수
27. 연결된권한사항식별자 (linking rights statement identifier)	27.1 연결된권한사항식별유형	필수	
	27.2 연결된권한사항식별값	필수	
권한 요소 Rights entity	28. 권한사항 (rights statement)	28.1 권한사항식별자	필수
		28.2 권한배경	필수
		28.3 저작권정보	선택
		28.4 인가정보	선택
		28.5 법령정보	선택
		28.6 인정된 권한	선택
		28.7 연결된객체식별자	선택
		28.8 연결된주체식별자	선택
	29. 권한확장정보 (right statement extension)		선택

출전 : PREMIS Editorial Committee, 「Data Dictionary section from PREMIS Data Dictionary for Preservation Metadata」, 2011. <<http://www.loc.gov/standards/premis/v2/premis-dd-2-1.pdf>> [인용. '12.03.30.]

PREMIS 데이터 사전은 모든 유형의 디지털 객체를 대상으로 하므로, 오마이뉴스 보도기사와 관리할 객체의 성격이 여러 가지 면에서 중복

된다고 할 수 있다. 보도기사 콘텐츠 관리를 위해 필요한 특성을 데이터 사전에서 찾아볼 수 있는데 그 내용은 다음과 같다.

첫째, 주요속성(significant properties)을 기술한다. 하위 구성요소로 주요속성 유형(significant properties type)을 기술함으로써 어떤 유형의 특성을 기술하는지 알도록 한다. 주요속성 유형에는 내용(content), 구조(structure), 행위(behavior), 쪽수(page count), 쪽 너비(page width), 서체(typeface), 하이퍼링크, 그림 개수(image count), 색상(color space) 등이 제시되어 있다. 그리고 주요속성유형과 함께 주요속성값(significant properties value)을 짝을 이뤄 함께 기술함으로써 객체의 특성을 상세하게 나타내도록 한다.³³⁾ 이러한 주요속성 의미단위를 오마이뉴스 메타데이터에 응용한다면 보도기사 콘텐츠의 다양한 형식을 설명할 수 있다. 여기에는 오마이뉴스 보도기사에서 나타나는 특성, 즉 문체 유형, 첨부파일 유무, 멀티미디어 자료 여부 등이 해당된다.

둘째, 연결객체식별자(Linking object identifier)를 적용할 수 있다. 연결객체식별자는 어떤 이벤트와 관련된 객체에 대한 정보이다. 디지털 상의 출처는 객체와 문서화된 이벤트 간의 관계를 요구하기 때문에 기록물과 연결된 객체를 식별할 필요성이 있다. 연결된 의미단위는 저장소가 그 정보를 파악해야 한다는 점에서 필수로 기술되어야 마땅하지만 PREMIS가 연계되어야 하는 방향을 일일이 열거하지 않기 때문에 선택사항으로 한정된다.³⁴⁾ 이 의미단위를 적용한다면 하나의 보도기사를 연관기사와 연결할 수 있을 것이다. 데이터 사전의 연결객체식별자가 객체와 이벤트 기록 간 관계에 관한 것이라면 오마이뉴스 메타데이터

33) PREMIS Editorial Committee, 『Data Dictionary section from PREMIS Data Dictionary for Preservation Metadata』, 2011, p.41 <<http://www.loc.gov/standards/premis/v2/premis-dd-2-1.pdf>> [인용. '12.03.30.]

34) PREMIS Editorial Committee, 『Data Dictionary section from PREMIS Data Dictionary for Preservation Metadata』, 2011, p.147<<http://www.loc.gov/standards/premis/v2/premis-dd-2-1.pdf>> [인용. '12.03.30.]

는 객체와 객체 간의 관계에 중점을 두어야 한다. 그 후에 이를 URL 등의 식별자를 이용하여 연결할 수 있다.

셋째, 데이터 사전의 비트스트림(bitstream)의 개념에 주목할 필요가 있다. 비트스트림은 파일 내에 존재하는 연속적·비연속적 데이터를 의미하며, 파일 내에 들어가 있는 비트의 집합이라고도 할 수 있다. 비트스트림은 파일구조의 추가 혹은 특정 파일 포맷의 형식에 따른 재포맷 과정이 없다면 단독적인 파일로 전환될 수 없다. 이러한 비트스트림에는 TIFF 6.0 파일에 들어가 있는 이미지, WAVE 파일의 오디오 데이터, Microsoft Word 파일에 들어가 있는 이미지가 해당된다. 보도기사 콘텐츠와 비교해 보면 비트스트림은 보도기사에 포함된 이미지, 동영상 파일과 같은 맥락이라고 할 수 있다. 보도기사에 들어가 있는 이미지, 영상, 오디오 등의 콘텐츠는 그 자체로 참고자료로서의 가치를 가지기 때문에 따로 관리해야 함에도 불구하고 텍스트를 포함한 하나의 보도기사 콘텐츠로 묶여서 관리되기 쉽다. 그러나 메타데이터 설계시 비트스트림 구조를 적용하여 첨부된 멀티미디어 콘텐츠를 따로 기술한다면 활용도가 더욱 높을 것이라 예상된다.

이와 같이 PREMIS 데이터 사전은 디지털 객체에 중점을 두어 설계되었기 때문에, 인터넷신문인 오마이뉴스 기록물에 대한 메타데이터 설계에 있어 적합한 참고모델임을 알 수 있다. 그러나 PREMIS 데이터모델은 보존에 중점을 둔 메타데이터이므로 지속적인 활용도를 필요로 하는 인터넷뉴스 콘텐츠에 맞도록 보완이 필요하다.

4. 오마이뉴스 기록 메타데이터 재설계 방안

1) 메타데이터 재설계 방향

오마이뉴스 기록 메타데이터 재설계시 중점을 두어야 할 사항은 다음과 같다.

첫째, 이용자 접근이 용이하도록 한다. 이용자가 원하는 정보에 정확하게 접근하기 위해서는 무엇보다 접근점이 다양하게 나타나야 하므로, 다각적인 측면에서 기술요소를 구성한다.

둘째, 연관 정보와의 연결에 중점을 둔다. 연관 정보에는 다른 보도 기사가 될 수도 있고 오디오나 동영상과 같은 멀티미디어 콘텐츠가 될 수도 있다. PREMIS 데이터 사전에서 각 개체를 식별자로 연결하는 방식을 응용하여 연관 정보를 연결할 수 있다.

셋째, 행위이력, 즉 생산·관리·이용·보관에 대한 이력 축적을 고려한다. 행위이력의 축적은 기록을 구성하는 내용·구조·맥락 가운데 맥락을 밝힘으로써, 기록의 설명책임성과 신뢰성을 보장한다. 보도기사의 메타데이터 상에 생산이력(Creation history), 관리이력(Management history), 이용이력(Use history), 보관이력(Preservation history)을 구성하여 데이터에 행해진 행위이력을 밝힐 수 있다. 또한 각 요소에는 하위요소를 두어 행위주체에 관한 정보를 기술하도록 한다.

넷째, 콘텐츠의 활용을 높이는 방향으로 설계한다. 보도기사 뿐만 아니라 보도기사에 포함된 멀티미디어 파일을 하나의 콘텐츠로 분리하여 활용도를 높이도록 한다. 그러기 위해서는 앞서 살펴본 PREMIS 데이터 사전의 비트스트림과 같은 맥락에서 보도기사 내의 멀티미디어 콘텐츠의 메타데이터를 작성할 필요가 있다.

다음에 나오는 2절에서는 이러한 사항을 고려하여 오마이뉴스 기록

의 메타데이터를 재설계하도록 할 것이다. 오마이뉴스 메타데이터 기술요소 구성을 위해서 선행되어야 할 것은 보도기사 기록에 해당되는 필수요소를 선정하는 것이다. 그 후 오마이뉴스 보도기사 기술에 적합한 방향으로 부가적인 수정과 보완이 이루어져야 한다. 먼저 필수요소 선정에 있어 국내외 메타데이터 표준의 비교를 통해 해당되는 요소를 선정해볼 수 있다. 각 표준의 기술요소를 비교분석하는 과정에서 중복요소, 그리고 비중복요소가 발생하게 되는데 오마이뉴스 보도기사의 특성을 고려하여 적합한 방향으로 필수요소를 선정해야 할 것이다.

사실상 국내사례와 해외사례는 중복되는 요소가 많지 않다. 이는 국내사례가 현용·비현용 기록물을 대상으로 하고 과거 종이기록물까지 포함하는 데 반해, 해외사례는 보존을 위한 메타데이터이며 디지털 객체만을 대상으로 하기 때문일 것이다. 또한 국내사례의 상위요소에 해당되는 부분이 해외사례의 하위요소로 포함되는 등 요소간 상하단위 차이가 있을 수 있다. 하지만 이러한 차이점을 포괄하며, 두 사례의 상위요소만을 대상으로 분석한다.

필수요소를 선정한 뒤에는 필요한 요소를 추가하거나 기존의 요소를 수정함으로써 오마이뉴스 보도기사에 맞게 보완한다. 이 과정에서는 기사처리과정에서 발생하는 기록정보를 모두 포함할 수 있도록 하는 것에 중점을 두어야 할 것이며, 앞서 국내외 사례 분석에서 서술했던 적용 가능한 사례별 특성을 고려해야 한다.

2) 메타데이터 기술요소 구성

본 절에서는 국내외 메타데이터 표준 사례의 기술요소 구성을 비교하였다. 그리고 비교 내용을 오마이뉴스의 기록에 적용하여 메타데이터를 구성할 기술요소명을 설정하였다.

〈표 11〉 국내외 사례 요소 비교 및 오마이뉴스 기록에의 적용

해당 영역	공공표준안 메타데이터 기술요소	PREMIS 데이터 사전 의미단위	오마이뉴스 메타데이터 기술요소명
생산영역	1. 생산자 2. 기록계층 3. 고유식별자 4. 기록명 12. 일시 13. 생산이력	1. 객체식별자 6. 원객체명 21. 주체식별자 22. 주체명 23. 주체 유형 24. 주체 주기 25. 주체 확장정보	고유식별자 생산유형 제목 생산일시 생산자명 생산자유형 생산자확장정보 취재경위
형태영역	8. 포맷 9. 저장매체 10. 크기	2. 객체유형 8. 환경	객체유형 파일형식 용량 크기
행위이력 영역	12. 일시 14. 보존기간 15. 권한 16. 위치 17. 관리이력 18. 이용이력 19. 보존이력 22. 무결성체크	3. 보존수준 7. 저장 9. 서명정보 14. 이벤트식별자 15. 이벤트유형 16. 이벤트일시 17. 이벤트상세내용 18. 이벤트결과정보 21. 주체식별자 22. 주체명 23. 주체 유형 24. 주체 주기 25. 주체 확장정보 28. 권한사항 29. 권한확장정보	관리유형 관리일시 관리주체 관리사유 관리결과 이용유형 이용일시 이용주체 이용결과 보존유형 보존일시 보존주체 저장위치

연결영역	20. 관계	10. 관계 11. 연결이벤트식별자 12. 연결된지적객체식별자 13. 연결권한사항식별자 19. 연결된주체식별자 20. 연결된객체식별자 26. 연결된이벤트식별자 27. 연결된권한사항 식별자	연결객체명 연결객체유형 연결객체식별유형 연결객체식별값 연결객체생산자명 연결객체생산자유형 연결객체생산자확장정보 연결주체유형 연결주체명 연결주체확장정보 연결행위유형 연결행위결과 연결행위주체명 연결행위주체유형 연결행위주체확장정보
내용기술 영역	4. 기록명 5. 기술 6. 주제 7. 유형 11. 분류 21. 언어	4. 주요속성 5. 객체특성 6. 원객체명	기본분류 지역분류 형식분류 기사등급 태그 내용주기 연재기획 특이사항

먼저 오마이뉴스 기록을 기술하기 위한 5개의 영역을 설정하고, 영역 별로 각 사례의 해당되는 기술요소를 비교하였다.

생산영역에 해당되는 국내 표준안의 기술요소는 생산자, 기록계층, 고유식별자, 기록명, 일시, 생산이력이다. 국내 행정기록물의 경우 행정 기능업무에 따른 증거 혹은 결과로서 기록물이 생산되기 때문에 기능 분류에 의한 기록계층 요소가 존재한다. 반면 모든 디지털 객체를 대상

으로 하는 PREMIS 데이터 사전과 오마이뉴스의 보도기사에는 해당되지 않는다. 생산자, 고유식별자, 기록명, 일시, 생산이력은 생산단계에서 이미 결정되어지는 사항이기 때문에 생산영역에서 취득될 수 있는 요소에 포함한다.

한편 PREMIS 데이터 사전의 생산영역 의미단위는 국내 표준안의 기술요소와 많은 차이를 보인다. 생산영역에 해당되는 의미단위는 객체 식별자, 원객체명, 주체에 관한 정보에 불과하다. 이러한 차이를 보이는 이유는 데이터 사전이 보존을 위한 메타데이터이기 때문일 것이다. 표에서 볼 수 있듯이 데이터 사전은 생산자 요소를 따로 두고 있지 않다. 그러나 여러 가지 주체 관련 요소를 상위요소로 드러냄으로써 주체에 대한 정보를 더욱 상세하게 취득하도록 한다.

앞서 오마이뉴스의 기사 처리과정에서 볼 수 있듯이 기사 작성작성 단계에서 시민기자가 참여하기 때문에 오마이뉴스 기록은 생산단계가 갖는 비중이 크다. 국내 표준안과 같이 생산단계에서 취득할 수 있는 정보를 고려하여, 고유식별자, 생산유형, 제목, 생산일시 요소를 구성하도록 한다. 그리고 오마이뉴스의 기록은 생산자에 관한 정보 획득이 중요한 의미를 갖는다. 그러므로 생산자에 관한 요소, 즉 생산자명, 생산자유형, 생산자확장정보를 구성하여 생산자에 관한 정보를 다양하게 취득할 수 있도록 한다.

형태영역을 살펴보면 국내 표준안은 포맷, 저장매체, 크기로 기록물의 형태정보를 획득한다. 국내 표준안이 기술하는 기록물의 범주는 종이기록물부터 마이크로필름, 행정박물, 전자기록물 등 모든 형태를 아우른다. 따라서 정보를 담고 있는 기록물의 물리적 형태에 대한 기술요소가 포맷, 저장매체, 크기 요소로 분명하게 나타난다.

반면 PREMIS 데이터 사전은 이미 그 형태를 디지털 객체로 한정하고 있기 때문에 물리적 형태에 관한 의미단위를 분명히 규정하고 있지 않고 있다. 다만 객체의 형태를 재현물(representation), 파일, 비트스트림

(bitstream)으로 나누는 객체유형 요소가 형태영역에 포함된다고 할 수 있다. 그리고 하드웨어나 소프트웨어적인 환경을 기술하는 환경 의미 단위를 형태영역에 포함시킬 수 있다.

오마이뉴스의 기록은 주로 PREMIS 데이터 사전과 같이 디지털 객체로 생산된다. 따라서 형태영역의 구성에 있어 주로 데이터 사전의 의미 단위를 참고해야 할 것이다. 무엇보다 객체유형을 설정하는 구조는 보도기사와 멀티미디어 자료를 기술하는 데에 적합하기 때문에 데이터 사전과 같이 오마이뉴스 메타데이터에도 객체유형 요소를 구성하였다. 더불어 객체유형에 다른 파일형식, 용량, 크기 요소가 기술되어야 한다.

행위이력에 관한 국내표준안의 기술요소는 일시, 보존기간, 권한, 위치, 관리이력, 이용이력, 보존이력, 무결성체크가 해당된다. 일시 요소의 경우 생산·관리·이용·보존이력에 모두 필요한 사항이기 때문에 생산영역과 행위이력영역에 반복하여 기입하였다.

데이터 사전에서는 관리·이용·보존이력에 대한 의미단위를 따로 분류하지 않고 이벤트로 묶어 구성하고 있다. 이벤트이력에 관한 의미단위는 이벤트식별자, 이벤트유형, 이벤트일시, 이벤트상세내용, 이벤트결과정보이며, 이벤트주체에 관한 의미단위는 구성되어 있지 않다. 따라서 주체 관련 의미단위의 적용이 필요한 부분이다. 이벤트 관련 의미단위 외에도 보존수준, 저장, 서명정보, 권한사항 등의 의미단위 또한 행위이력 영역에 포함시킬 수 있다.

행위이력영역의 구성의 큰 틀은 국내 표준안의 관리·이용·보존이력의 분류를 참고해볼 수 있다. 그 이후에 각 행위이력의 세부 내용, 즉 행위유형, 행위일시, 행위주체, 행위사유, 행위결과를 설계함으로써 행위이력영역을 구성하도록 한다.

국내 표준안의 기술요소 가운데 연결영역에 해당하는 것은 관계 요소이다. 관계 요소의 하위요소를 통해 연관정보를 연결하고 있지만 데이터 사전이 8개의 의미단위를 통해 연관정보를 기술하는 것에 비한다

면 그 상세도에 있어서 한계가 있다고 할 수 있다.

PREMIS 데이터 사전에서 연결객체, 연결주체, 연결이벤트를 분류하여 구성하고 있다. 이를 적용하여 오마이뉴스 메타데이터는 상위요소로 연결객체 요소, 연결주체 요소, 연결이벤트 요소를 두고, 하위요소로 세부정보가 취득되도록 설계할 것이다.

마지막 내용기술 영역은 기록 자체의 내용에 관한 기술요소가 해당되는데, 국내 표준안은 기록명, 기술, 주체, 분류, 언어 요소가 포함된다. 그리고 데이터 사전은 주요속성, 객체특성, 원객체명이 해당된다. 내용기술 영역에서 데이터 사전이 비교적 빈약한 의미단위 구성을 보이고 있는데, 이는 PREMIS 데이터 사전이 보존을 위한 메타데이터 모델로서 주로 객체의 식별과 연관정보의 연결에 중점을 두고 있기 때문이다.

따라서 기록 자체에 대한 내용기술 영역은 주로 국내 표준안을 참고하여 오마이뉴스 기록에 적합하도록 보완하는 방향으로 구성해야 한다. 특히 오마이뉴스에서 현재 운영되고 있는 다양한 기사분류를 내용기술 영역에서 다룸으로써 기록에 대한 접근점을 만들 수 있다.

지금까지의 국내외 사례의 비교분석을 바탕으로 오마이뉴스 메타데이터 기술요소를 구성한 결과는 다음 <표 12>와 같다.

<표 12> 오마이뉴스 메타데이터 기술요소 구성

영역	상위요소	하위요소	기술값
생산 영역	생산유형		기사/편집수정요청계시판/생나무클리닉/덧글
	생산자명	ID	
		실명	
		기타이름	닉네임
생산자유형		상근기자/시민기자/비회원	

	생산자확장정보	나이	회원가입 시 입력하는 회원정보에서 자동취득
		지역	
		소속	
		기타	
	생산일시		송고일시
	제목	제목	
		부제목	
기타제목			
취재경위			
형태 영역	객체유형		텍스트포함콘텐츠/단일멀티미디어 콘텐츠
	파일형식		jpg, bmp, wav, avi, ...
	용량	용량단위	mb
		용량값	
	크기	크기단위	inch, cm
크기값		가로×세로	
행위이력 영역	관리유형		기사수정/인결채택/머리기사채택/ 계재종료/원고료지급
	관리일시		
	관리주체	관리주체명	
		관리주체부서	
		관리주체소속	
	관리사유		
	관리결과		인결채택여부/채택등급/원고료
	이용유형		검색/조회/댓글/공유/기사접수
	이용일시		
	이용주체	이용주체유형	
이용주체명			
이용주체 확장정보			

	이용결과		검색수/조회수/댓글수/공유수/기사 접수통계	
	보존유형			
	보존일시			
	보존주체	주체명		
		부서		
		소속		
저장위치	위치식별자유형			
	위치식별자값			
연결 영역	연결객체	연결객체명		
		연결객체유형	연관기사/멀티미디어콘텐츠/편집수 정요청게시판/생나무클리닉/댓글	
		연결객체식별유형	URL	
		연결객체식별값		
		연결객체생산자명		
		연결객체생산자유형		
	연결주체	객체생산자확장정보		
		연결주체유형	국가/기관/단체/개인	
		연결주체명		
		연결주체확장정보		
	연결행위	연결주체식별자		
		연결행위유형	원고수정/게재종료	
		연결행위식별자		
		연결행위주체명		
		연결행위주체유형		
행위주체 확장정보				
내용기술 영역	기본분류		사는이야기/사회/경제/정치/문화/ 민족/국제/교육/스포츠/책동네/여 행/영화/미디어/여성	
	지역분류		광주전라/대구경북/부산경남/대전 충남/인천경기/강원제주	

	형식분류		정보/주장/사건/고발/일반/예고/리포/인물인터뷰/동정/기타
	기사등급		오피/옴/버금/일경/생나무
	태그		
	내용주기		덧붙이는 글(자유 서술)
	연재기획	연재기획 연결기사명	
연재기획 연결기사식별			

<표 12>와 같이 5개의 영역과 30개의 상위요소를 구성함으로써 오마이뉴스 메타데이터를 설계하였다. 먼저 큰 범주의 영역, 즉 생산영역, 형태영역, 행위이력영역, 연결영역, 내용기술영역으로 구성하였으며, 각 영역에는 그에 해당되는 상위요소가 포함된다.

생산영역은 생산유형, 생산자명, 생산자유형, 생산자확장정보, 생산일시, 제목, 취재경위를 포함한다. 생산유형은 기사, 편집수정요청 게시글, 생나무클리닉 게시글, 댓글 중에서 메타데이터로 기술되는 기록이 어떤 유형에 속하는지 밝힌다. 오마이뉴스의 주요 기록은 보도기사지만 보도기사를 중심으로 수정요청,클리닉, 댓글 등에 대한 기록을 함께 연결해야 하기 때문에 보도기사 외의 기록에도 생산유형을 밝히고 메타데이터를 작성할 필요가 있다.

그리고 생산자명, 생산자유형, 생산자 확장정보를 포함하는 생산자 정보가 생산영역에 해당된다. 생산자유형은 상근기자과 시민기자로 분류되는데, 확장정보에는 회원가입시 입력하는 기본정보가 해당된다. 생산주체에 관한 요소는 시민의 개인정보가 다루어지는 부분이므로 가입자 본인이 공개를 동의하는 범위 안에서 기술하는 방향으로 한다.

생산일시와 제목, 취재경위는 생산과정에서만 발생하는 정보이기 때문에 생산영역에서 다루도록 한다. 제목의 하위요소인 기타 제목에는 수정요청의 결과로 제목의 변경이 있을 시 변경 이전의 제목이 기술되거나, 필요에 따라 외국어 제목을 기술할 수 있다. 그리고 취재경위는

기사쓰기 페이지에서 취재경위란에 작성자가 자유롭게 서술한 내용이 기술된다.

형태영역은 기록의 형태적 특성을 밝히는 콘텐츠 유형, 파일형식, 용량, 크기요소로 구성하였다. 먼저 기술되는 객체가 텍스트와 그 이상을 포함한 하나의 콘텐츠인지, 콘텐츠에 포함된 하나의 멀티미디어 파일인지 구분해야 한다. 이는 PREMIS 데이터 사전의 비트스트림 구조를 적용한 것으로, 보도기사 콘텐츠를 구성하는 멀티미디어 파일을 하나의 독립된 단일 콘텐츠로 기술함으로써 멀티미디어 자료의 활용도를 높일 수 있다. 단일 멀티미디어 파일일 경우 파일형식을 필수로 기술하도록 하고, 이미지 파일은 픽셀 단위의 이미지 크기를, 동영상 파일은 용량을 기술한다. 오마이뉴스의 기사 규정에서 권장하는 규격이 있지만 추후의 상황에 융통성 있게 적용하기 위해서는 메타데이터 상에 파일 규격에 대한 기술요소를 구성할 필요가 있다.

행위이력영역은 크게 관리이력에 관한 요소, 이용이력에 관한 요소, 보존이력에 관한 요소로 나눌 수 있으며, 각 행위이력에는 행위주체에 대한 정보가 포함된다. 관리이력의 유형은 오마이뉴스의 기사처리과정에 따라 잉걸채택, 머리기사채택, 게재종료, 원고료지급 그리고 기사수정으로 선택하여 기술한다. 관리주체는 대부분이 오마이뉴스 내의 구성원이므로 소속부서를 함께 기술하도록 하고, 기사 채택이나 게재종료 등의 처리과정에 따른 관리사유와 관리결과를 기술한다.

이용이력에 관한 요소에서는 먼저 검색, 조회, 댓글, 공유, 기사접수로 이용유형을 기술한다. 더불어 이용일시와 이용주체를 기술하는데, 이용주체의 경우 주체유형에서 비회원, 회원, 사내 구성원 중 선택하여 밝히도록 한다. 인터넷뉴스의 특성상 비회원 이용률이 높기 때문에 이용이력에 관한 상세한 정보를 기술하는 것에 한계가 있다. 따라서 검색, 조회, 댓글 등의 비회원 이용은 통계적인 정보만을 기술할 수 있을 것이고, 그 내용은 이용이력결과에 기술한다. 또한 공개되어 있는 데이

터는 조회수, 댓글수 등의 통계가 지속적으로 갱신되기 때문에 주기적으로 그 이용이력을 남기는 방법이 실행될 수 있을 것이다.

보존이력에 관한 요소는 향후 오마이뉴스가 보도기사 콘텐츠를 특정 저장공간에 보존하려고 할 때 필요한 부분이다. 오마이뉴스의 모든 기사가 웹상에 공개되어 있기 때문에 오늘날에는 사실상 큰 의미를 갖지 못한다. 그러나 방대한 양의 보도기사 데이터를 웹 서버에 삭제 없이 축적하는 것은 재정적 측면에서도 분명 한계가 있을 수 있다. 뉴욕타임스와 워싱턴포스트지 등 미국의 인터넷신문은 일정 시기 이전의 보도 기사를 데이터베이스로 저장하여 이용자에게 유료로 제공함으로써 수익을 창출한다.³⁵⁾ 오마이뉴스는 NGO의 성격을 가지기 때문에 뉴욕타임스나 워싱턴포스트지와 같이 수익창출을 목적으로 하지는 못하더라도, 축적되는 기사의 양이 방대하기 때문에 보존이력에 대한 메타데이터 기술이 반드시 필요하다고 할 수 있다. 본 논문에서 구성한 메타데이터에서는 보존이력에 대한 기본적인 요소만을 제시하고 있지만 앞으로의 지속적인 운영을 위해서 더욱 높은 수준으로 기술할 필요가 있는 부분이다.

연결영역은 연결된 객체, 주체, 행위로 나누어 구성하였다. 연결객체의 유형에는 연관기사, 멀티미디어 콘텐츠, 편집수정요청 게시글, 생나무 클리닉 게시글, 댓글이 있다. 하나의 기사 기록을 중심으로 연관되

35) 미국의 인터넷신문들은 'archives'라는 서비스 메뉴를 두어 검색 서비스를 제공하고 있는데, 풀텍스트 제공은 대부분 유료화하고 있다. 『워싱턴포스트』는 1977년부터의 기사가 데이터베이스로 구축되어 있으며 검색 서비스를 받기 위해서는 회원 가입을 해야 한다. 검색 결과 기사의 제목과 요약은 보는 것은 무료이나 2주가 경과한 기사의 풀텍스트를 보기 위해서는 기사 1건당 2.95달러의 요금을 지불해야 한다. 『시카고 트리뷴』은 1985년부터의 기사를 DB화했는데, 회원 가입을 해야 하며 풀텍스트를 보려면 기사당 2.55달러의 요금을 지불해야 한다. 『뉴욕타임즈』는 1996년부터의 기사를 DB로 구축했는데, 역시 검색은 무료지만 풀텍스트를 보려면 기사의 시기에 관계없이 기사당 2.5달러의 요금을 지불해야 한다(계수미, 『인터넷신문의 콘텐츠 특성에 관한 연구』, 연세대학교 언론홍보대학원 잡지출판전공 석사학위논문, 2000, 53쪽).

는 객체를 연결시켜줌으로써 기사에 대한 정확한 파악이 가능하게 할 뿐만 아니라 더욱 다양한 정보를 제공할 수 있다. 객체식별은 웹상에서 검색되는 기록은 URL 주소를 사용하여 식별자를 만들고, 객체생산자 관련 요소를 이용하여 연결객체의 생산자에 대한 정보를 간략히 기술하도록 한다.

연결주체는 기록과 관련된 주체에 관한 요소이다. 연관되는 주체는 국가, 기관, 단체, 개인이 될 수 있으며, 이 중에서 주체유형을 선정하여 주체명 요소에서 공식명칭을 밝히는 방식으로 한다.

연결행위 요소에서는 기록과 연관된 행위, 즉 기사수정, 게재 종료, 원고료 지급 등의 관리행위와 이용행위, 보존행위를 기술한다. 행위이력의 축적을 위하여 행위일시, 행위결과, 행위주체 요소를 통해 연관된 기록 처리 행위의 기본적인 정보를 기술하도록 한다.

마지막 내용기술영역은 기록 자체의 내용에 대한 정보를 담으며, 다른 유형의 기록보다는 주로 기사기록에 해당하는 부분이다. 기본분류, 지역분류는 기사작성시 입력하는 기본분류와 지역분류가 동일하게 적용된다. 그리고 형식분류는 기사검토 과정에서 기사의 서술형식에 따라 분류된다. 형식분류의 세부내용은 현재 오마이뉴스에서 운영되는 10개 분류를 그대로 적용할 수 있다. 기본분류, 지역분류, 형식분류와 더불어 기사의 등급 또한 하나의 기사분류로 기술한다. 그리고 태그와 내용주기 요소를 이용하여 기사의 내용 자체에 대하여 기술한다. 태그는 기사작성 시 태그란에 입력하는 내용을 적용하고, 내용주기는 덧붙이는 글란에 작성하는 내용이 들어간다. 기사 내용에 특이사항이 있을 경우 메타데이터 작성시 태그와 내용주기 요소에 임의로 작성할 수 있도록 한다. 연재기획은 오마이뉴스에서 승인한 연재기획에 해당하는 기사일 경우 이를 표시하고, 식별자를 이용하여 연재기획 시리즈 내의 다른 기사들과 연결시켜 준다.

이와 같이 오마이뉴스 기록에 적합한 방향으로 영역과 기술요소를

설계해보았다. 본 논문에서 설계한 메타데이터는 오마이뉴스만의 운영적 특성에 맞추어 구성하였다는 사실을 고려해야 하며, 메타데이터의 적용 시 인터넷매체의 유동적인 성격에 맞게 메타데이터의 활용 또한 융통성 있게 운영될 필요가 있다.

3) 메타데이터의 활용 방안

앞서 구성한 메타데이터를 활용하기 위해서는 각 단계에서 관리자뿐만 아니라 시민기자과 이용자의 역할이 적절하게 수행되어야 한다. 이때 주의해야 할 것은 시민기자과 이용자의 입장에서 수행해야 하는 역할 비중이 과할 경우, 기사작성과 구독이 감소할 위험이 따를 수 있다는 것이다. 따라서 시민기자과 이용자에게는 부담을 느끼지 않을 정도로 최소한의 역할을 기대해야 하며, 가능한 한 현재 운영되는 수준을 유지해야 할 것이다. 오마이뉴스 메타데이터의 운영·활용을 위하여 관리자, 시민기자, 이용자가 수행해야 하는 역할은 무엇이며, 각 주체의 입장에서 메타데이터를 어떠한 용도로 활용할 수 있는지 살펴보고록 하겠다.

먼저, 관리자 측면에서의 메타데이터 활용방안이다. 관리자는 메타데이터를 이용하여 많은 양의 콘텐츠를 효율적으로 관리할 수 있다. 메타데이터는 분류와 검색이 필요한 업무에서 특히 유용하게 활용되는데, 먼저 메타데이터가 갖추어진 각각의 기사 기록은 다양한 분류체계에 따라 유동적으로 분류될 수 있다. 따라서 관리자가 기사 기록을 이용하여 다양한 컬렉션을 기획하여 서비스하는데 용이하며, 이에 따른 결과로 기획된 콘텐츠 서비스를 통해 수익창출까지도 고려해 볼 수 있다. 그리고 메타데이터는 관리자의 기사 검색 업무에 활용할 수 있다. 이용자뿐만 아니라 관리자도 여러 가지 목적에서 지나간 기사를 검색해야 하는 상황이 발생한다. 수정요청이 접수된 기사를 검색하거나, 지나간

자료와의 중복을 확인하거나, 또는 공정한 머리기사 채택 심사에 있어서 앞선 사례를 참고하는 등의 상황이 그러하다. 이렇듯 기사의 검색이 요구되는 상황에서 기사 기록에 대한 메타데이터가 큰 역할을 할 수 있을 것이다.

둘째로, 시민기자 측면에서의 활용이다. 시민기자는 메타데이터를 통한 기사 열람을 통해 기사작성 시 중복기사를 예방할 수 있으며, 본인이 작성한 기사와 다른 시민기자가 작성한 기사를 연결시킬 수 있다. 연관기사의 연결이 지속적으로 진행된다면 한 사건에 대한 전체 맥락이 형성되어 구독자로 하여금 사건에 대한 이해를 도울 수 있다. 뿐만 아니라 연관기사로 묶인 보도기사의 집합은 그 자체로 주제를 갖는 하나의 주제 콘텐츠로 발전한다. 시민기자의 메타데이터 활용은 이렇듯 기사 간의 연계뿐만 아니라 시민기자 간의 연계까지도 가능하게 한다. 현재 오마이뉴스에서 운영되고 있는 연재기획의 릴레이 프로젝트가 이와 같은 방향성을 갖는 것으로 보인다.

셋째로, 이용자 측면에서의 메타데이터 활용이다. 이용자는 무엇보다 다양하고 정확한 정보를 획득하는 것을 중시하는데, 그러한 목적을 위하여 메타데이터가 활용된다. 이용자의 메타데이터 활용은 대부분 검색 차원에서의 활용이라고 할 수 있다. 다만 이용자가 직접 메타데이터에 접근하기에는 이용에 어려움이 따를 수 있기 때문에, 시스템 환경을 통해 이용자가 메타데이터를 활용할 수 있도록 해야 할 것이다. 메타데이터의 활용으로 이루어지는 기사 검색이나 관리자가 제공하는 정보서비스를 통해 이용자는 지위고하를 막론하고 공평하게 정보를 제공받을 수 있어야 할 것이다.

5. 맺음말

오늘날 대중매체는 없어서는 안 될 중요한 사회제도가 되었다. 언론학에서 말하는 대중매체의 사회적 기능들 중에서도 특히 사회조정 기능과 동원 기능의 관점에서 볼 때 오늘날의 사회에서는 ‘영향을 미치는’ 수준을 넘어 사회를 제어하는 단계까지 왔다고 볼 수 있다. 또한 이용률과 영향력 측면에서 대중매체 가운데 인터넷의 비중이 크게 증가했다. 인터넷뉴스는 하이퍼텍스트성, 저장성, 멀티미디어성, 상호작용성, 그리고 즉시성과 속보성의 특성을 가지며, 종이신문이 발휘했던 파급효과에 비해 훨씬 빠르고 광범위한 결과를 불러온다. 인터넷뉴스의 메타데이터 재설계가 불러올 수 있는 내부적 효과는 보도기사가 신뢰성을 갖추으로써 경쟁력을 얻는다는 점, 그리고 방대한 양의 등록 기사를 효율적으로 관리할 수 있다는 점이다. 그리고 외부적 효과는 이용자가 원하는 기사에 정확하게 접근할 수 있다는 점, 이용자에게 더욱 풍부한 정보를 제공할 수 있다는 점이 있다.

국내의 대표적인 인터넷 언론 오마이뉴스의 기록은 보도기사를 중심으로 생산되며, 생산부터 보존까지 일정한 생애주기를 갖는다. 보도기사의 처리과정은 기사작성 단계, 미검토 단계, 검토완료 단계, 머리기사 채택 단계, 게재종료 단계로 나누어 볼 수 있으며, 앞에서 단계별로 생성되는 기록정보와 그에 해당되는 기술요소를 살펴보았다. 또한 본 논문은 메타데이터 설계방안을 구상하기 위하여 국내외 메타데이터 표준 사례를 분석하였다. 국내 메타데이터 표준은 2010년 10월에 발표된 기록관리 메타데이터 공공표준안을 대상으로 하였으며, 외국 표준은 보존 메타데이터를 위한 PREMIS 데이터 사전을 대상으로 하였다.

메타데이터 재설계시 중점을 두어야 할 사항은 다음과 같다. 첫째, 이용자 접근이 용이하도록 한다. 둘째, 연관정보와의 연결을 고려한다.

셋째, 생산·관리·이용·보관 행위이력의 축적을 고려한다. 넷째, 콘텐츠의 활용도를 높이는 방향으로 설계한다. 메타데이터 기술요소의 설계는 국내외 사례의 기술요소를 비교 분석하여 해당되는 필수요소를 선정한 후, 그 밖에 보도기사의 특성에 맞는 기술요소를 추가하고 보완하는 방법으로 수행하였다.

오늘날에는 많은 매체들이 과거 종이신문이 했던 역할들을 분담하고 있으며, 그중에서도 인터넷이 대표적이다. 정치, 사회, 경제, 문화 등 모든 분야에서 영향력을 가지고 있음에도 불구하고 인터넷 기사들은 종이신문이나 방송뉴스에 비해 기록관리의 필요성을 인식하지 못하고 있는 실정이다. 또한 오늘날 대중매체의 기록관리 연구는 아직도 TV방송에 머물고 있는 상태이며, 인터넷매체로 시작되는 그 이후의 뉴미디어에 대한 연구는 부족한 수준이다.

이미 언론정보학, 정보통신학, 커뮤니케이션학 등의 외부 학계에서는 인터넷 매체에 대한 연구가 심도 있게 이루어지고 있다. 그 연구의 범위는 인터넷 신문의 콘텐츠와 보도 프로세스뿐만 아니라 신뢰도 문제, 기사의 주제어 추출을 통한 자동분류 시스템 개발 등으로 다양하다. 그리고 최근에는 웹사이트 방문, 온라인 검색 통계, 위치정보, 소셜미디어 등 기업의 내·외부에 축적된 빅데이터(Big Data)를 관리하는 것이 기업 경쟁력의 핵심요소로 부상하고 있으며, 그에 대한 연구가 등장하기도 하였다.³⁶⁾ 인터넷 매체에 대한 이러한 연구들은 앞으로 이루어질 대중매체 기록관리 연구에 있어서 반드시 고려되어야 할 것이라 예상된다. 이러한 배경에서 본 논문은 최근 동향에 맞는 대중매체의 기록관리 방안을 모색한다는 점에서 의의를 찾을 수 있을 것이다.

그러나 본 논문에서는 오마이뉴스 내부 보안상의 문제로, 웹상에 공

36) 빅데이터란 기존의 관리 및 분석 체계로는 감당하기 어려운 막대한 데이터를 일컫는다(이용수, 「스마트혁명 시대 빅데이터 활용과 프라이버시 사이의 충돌에 관한 연구」, 경원대학교 소프트웨어대학원 소프트웨어학과 전자거래학전공 석사학위논문, 2001).

개된 사항만을 바탕으로 연구했기에 미흡한 점이 있을 것이라 예상된다. 이러한 부분은 언론사에서 실제로 메타데이터를 적용할 경우 보완되어야 하며, 기사 데이터의 관리에 초점을 두어 연구를 진행하였기 때문에 최근 스마트 기기의 개발, 소셜미디어의 활성화 등 최신 환경에 적합한 구체적인 활용방안을 제시하지 못하였다. 이와 관련된 구체적인 방안은 향후 인터넷 매체에 대한 연구에서 보충되어야 할 것이다.

ABSTRACT

A Study on the redesign Oh my News article metadata

Jeong, Seong-Suk

The share of the internet has grown significantly in terms of usage and level of influence. Among the internet media, OhmyNews is considered a alternative media representing Korea's independent internet news, and a typical example displaying characteristics of the internet news. The processing step can be divided as five stage, with recorded information and applicable technical element extractable from each stage.

In this thesis paper, we have analyzed domestic and overseas metadata standard examples to devise metadata design plans. Items to be focused when redesigning metadata based on domestic and foreign case studies are as follows: First, user access should be convenient; second, connection with related information should be considered; third, accumulation of production, management, usage, storage, and action history should be considered; and fourth, the design should allow higher utilization of contents.

In depth researches over internet media are already being carried out in external academic disciplines including the media information studies, information and communication studies, and communications studies. We expect that it will also be necessary to consider such researches over the internet media for mass media record management research going forward.

Key words: Archiving, internet news, OhmyNews, metadata