

# 메타데이터 상호운용성을 위한 기록관리 메타데이터 표준 분석

5W1H와 태스크 모델의 관점에서

백 재 은\* 스기모토 시게오\*\*

1. 머리말
2. 선행 연구
  - 1) 기록관리(Record management)를 위한 메타데이터 표준
  - 2) 기록 생애 주기 모델
  - 3) 기록 생애 주기를 바탕으로 메타데이터 표준의 특징 분석
  - 4) 관련 연구
3. 특징 분석을 위한 기초 모델-태스크 모델과 5W1H 모델
  - 1) 태스크 모델과 5W1H 모델의 의의(意義)
  - 2) 기록관리 업무 중심 모델-태스크 모델
  - 3) 메타데이터 기술 요소와 업무 중심 모델의 조합
  - 4) 5W1H 모델
4. 기초 모델을 이용한 메타데이터 표준간의 매핑
  - 1) 5W1H 카테고리 와 업무 식별을 위한 일반적인 어휘
  - 2) 5W1H 카테고리에서의 기술 요소 분류
  - 3) 태스크 그룹(Task Group)에서의 재분류
  - 4) 기초 모델을 이용한 분류와 매핑 정리
5. 맺음말

\* 일본 쓰쿠바 대학 대학원 도서관정보미디어연구과 박사과정 수료(주 저자)

\*\* 일본 쓰쿠바 대학 대학원 도서관정보미디어연구과 교수

## [국문초록]

메타데이터 표준규격(이하 메타데이터 표준으로 기재)은 디지털 자원(Digital resource)의 장기보존 및 디지털 아카이브를 위해 필요한 기본 요소 중 하나로, 이는 현대 정보사회에서 중요한 요소로 잘 알려져 있다.

자원(Resource)의 기록관리와 아카이브, 장기보존을 위한 메타데이터 표준은 다양하며, AGRkMS,<sup>1)</sup> EAD,<sup>2)</sup> ISAD(G),<sup>3)</sup> OAIS,<sup>4)</sup> PREMIS<sup>5)</sup> 등이 이용되고 있다. 우리는 아카이브 시스템의 메타데

- 
- 1) AGRkMs(Australian Government Recordkeeping Metadata Standard)는 기록(records)과 기록물이 가지고 있고, 사용하고 있는 문맥들(contexts)에 대한 정보를 기술한다. 오스트레일리아 국가기록원(The National Archives)이 권장하는 이 정보는 기록을 작성하고 수집하는 오스트레일리아 정부기관에 의해 사용되는 비즈니스 기능 안에서 파일 화 된다. 또한 이 표준은 기록(records)을 위한 메타데이터(ISO 23081)와 기록 관리(ISO 15489)안에서 오스트레일리아의 표준에 준수한다.
  - 2) EAD(Encoded Archival Description)는 기록검색도구에서 넓은 범위 내에 상호관계가 있는 기술 정보를 발견한다. 또한 EAD는 기술의 레벨 사이에서 존재하는 계층에 따른 관계를 나타내고, 계층 정보의 구조 내에서의 이동과 요소의 특정한 색인 작성과 검색을 지지하며 ISAD(G)와의 호환성을 유지하고 있다.
  - 3) ISAD(G)(General International Standard Archival Description)는 ICA(International Council on Archives)가 승인한 보존 기록 자료(Archival records)의 목록 기술의 국제표준이다. 이 표준은 보존기록 기술(Archival descriptions)의 준비를 위한 일반적인 안내를 제공하고 있다. 또한 이는 국가 표준의 개발을 위한 기초로서, 기존의 국가 표준과 함께 사용된다.
  - 4) OAIS(Open Archival Information System)는CCSD(Consultative Committee for Space Data Systems)에 의해 책정된 것으로, OAIS의 참조 모델은 장기간에 걸쳐서 디지털 정보에의 접근(access)을 유지하고 보존하는 데 전담하는 기록보존 시스템을 위한 개념적인 프레임워크이다. 참조 모델은 디지털 오브젝트를 아카이브하기 위해 적절한 개념의 이해와 인식을 높이는 것을 목적으로 하고 있다(Brian Lavoie, 2000).
  - 5) PREMIS(Preservation Metadata: Implementation Strategies)는 OCLC(OCLC and RLG)와 RLG(Research Libraries Group)에 의해 설립된 PREMIS Working group에서 작성된 것으로, 데이터 사전을 가지고 있다. 데이터사전은 데이터베이스 안의 데이터 항목 혹은 의미를 등록, 정의한 사전이다. 데이터사전은 OAIS 참조모델

이더 표준을 디자인하기 위해 목적에 따른 메타데이터 표준을 선택하고 맞춤화(Customization)하지 않으면 안 된다. 한편으로, 다른 시스템의 메타데이터 스키마와의 상호운용성(Interoperability)<sup>6)</sup>에 대한 고찰도 실시하지 않으면 안 된다.

이전 연구에서, 우리는 기록 생애 주기(Records lifecycle)라는 관점으로부터 메타데이터 표준의 특성에 대해 분석을 실시하였다. 이로 인해, 각 메타데이터 표준 요소가 해당하는 기록 생애 주기의 처음 단계를 확인할 수 있었고, 아카이브 혹은 보존을 위해서는 하나의 메타데이터 표준만으로 기록 생애 주기 전체를 포괄할 수 없다는 것을 보여 주었다. 우리는 이 분석을 통해서 기록 생애 주기의 단계와 메타데이터 표준간의 관계, 기록 생애 주기 전체에서의 메타데이터 특성은 볼 수 있었으나, 보다 상세한 분석을 실시하는 것은 앞으로의 과제로 남겨두었다.

지금까지의 연구에 근거하여, 본 논문은 기록 생애 주기의 관점에서 디지털 아카이브와 보존, 기록관리를 위한 메타데이터 표준의 특징 분석을 위해, 기록 생애 주기 안에서 실행되는 업무의 관점으로부터 메타데이터 스키마를 재 파악하고 분석하였다.

지금까지 메타데이터 스키마는 기술대상이 되는 자원을 중심으로 정의되었기 때문에 기록 생애 주기 전체와 생애 주기 안의 각 단계에서 이용되는 메타데이터 표준간의 매핑을 위한 적절한 방법이 없었다. 이에 본 논문에서는 각기 다른 메타데이터 표준의 기술 요소를, 기록 생애 주기에 포함시키는 업무와 연결시키는 것으로 메타데이터 표준간의 매핑 방법을 제안한다.

---

(ISO 14721)을 기초로 하고 있고, 보존 메타데이터(preservation metadata)를, 디지털 보존과정을 지지하기 위해 저장소가 사용하는 정보로 정의되고 있다.

6) 상호운용성(Interoperability)은 메타데이터 실행에서 가장 중요한 원리 중 하나이다. 상호운용성은 하나 혹은 그 이상의 시스템 혹은 구성 요소로부터 각각의 시스템의 특별한 기능 없이 정보를 교환하고, 교환된 정보를 이용하도록 하는 기능이다(CC:DA, 2000).

본 연구에서는 메타데이터 표준 분석을 위한 프레임워크(Framework)로, 기록 생애 주기를 이용하여 작성한 자원의 업무 중심 모델, 즉 태스크 모델(Task Model)을 제안한다. 태스크 모델을 이용함에 있어서 업무를 실행하는 ‘이벤트(Event)’의 관점을 보다 명확하게 할 수 있다.

한편, 업무를 중심으로 기술 요소간의 매핑을 보다 효율적으로 실시하기 위해서는 요소를 카테고리 화하여 매핑의 대상 범위를 좁히는 것이 중요하다. 이를 위해 우리는 범용성을 가진 5W1H 모델(Who, What, Why, When, Where, How)을 이용하여, 기술 요소를 카테고리 화 하는 것을 제안한다. 그리고 태스크 모델과 5W1H 모델을 이용하여 메타데이터 표준 요소에 특징을 부여하고, 요소간의 매핑을 실시하여 표준 간 관계를 확인하였다.

태스크 모델은 기록 생애 주기 전체에 업무의 관점을 반영한 것으로, 이 모델을 이용함으로써 기록 생애 주기와 그 안의 각 업무에 대한 메타데이터 표준의 사용, 그리고 특징 분석을 위해 실시되는 기술 요소간의 매핑이 가능할 수 있었다. 또한, 5W1H 카테고리를 이용하여 업무와 자원에 관련되는 기술 요소간의 관계를 살펴보는 것으로, 관계가 명확해지는 것과 함께 조사대상을 좁히는 것이 가능하게 되었다. 이 프레임워크의 제안과 이용으로 우리는 특징 분석을 위해 실시되는 매핑 혹은 분류가 단순한 일반적인 매핑이 아닌 의미적인 분류를 할 수 있었다.

본 연구에서는 이 모델들을 이용하여, 메타데이터 표준간의 크로스워크를 정의하였다. 그리고 태스크 모델의 각 단계의 문맥 내에서 메타데이터 기술 요소의 특성을 매핑 예로 확인하고, 이를 바탕으로 하여 프레임워크에 대해 고찰하였다.

**주제어:** 기록물, 기록관리, 기록 생애 주기, 디지털 리소스, 디지털 자원, 레코드 키펙, 메타데이터 스키마, 메타데이터

표준, 메타데이터 상호운용성, 보존을 위한 메타데이터,  
아카이브, 아카이브를 위한 메타데이터, 태스크 모델,  
5W1H 카테고리, Task Model, 5W1H Model

## 1. 머리말

인터넷과 웹의 급속한 발전은 대량의 디지털 자원을 생산하고 발신시켜 우리에게 다양한 자원을 이용할 수 있는 환경을 제공하였다. 이 같은 환경은 디지털 자원의 대중화를 가져왔고, 우리는 더 이상 네트워크 없는 생활에 대해 생각할 수 없게 되었다.

디지털 자원은 매일 다양한 형태로 끊임없이 생산되고 있다. 그러나 이 같은 디지털 자원의 대량 생산과 발신은 잇달아 여러 가지 문제를 가져오기 시작했다. 그 중 하나가 바로, 장래를 향한 디지털 자원의 관리와 보존이다. 다시 말하면, 현재의 디지털 자원을 다음 세대에 이용 가능한 상태로 유지하기 위한 관리와 보존이 필요한 것이다.

디지털 자원의 장기보존과 관리는 현대사회에 있어 중요한 과제로 널리 알려져 있다. 특히, 도서관이나 국가기록원처럼 각양각색의 자원을 장래의 이용자를 위해 관리하고 보존하는 역할을 담당하고 있는 기관에 있어 디지털 자원의 증가는 단지 양적인 문제일 뿐만 아니라, 장기보존과 관리를 위한 방법, 방침, 기술(技術), 정책 등, 다양한 문제를 초래하고 있다.

자원, 즉 기록물을 관리하는 모든 기관에 있어 메타데이터는 절대적으로 필요하다. 메타데이터는 디지털 자원의 장기보존을 위해서 필요하고, 디지털 아카이브와 보존에 있어 중요한 구성 요소이다.

주요 메타데이터 표준은 자원의 기록관리, 아카이브, 그리고 보존을

위해서 이용되고 있다. 여기에 조금 더 나열하면, 자료의 검색지원(Finding aids), 자원의 관리에 대한 관리, 접근성(Accessibility) 등의 목적을 가지고, 기록 생애 주기 단계에 따라 이용되어지고 있다. 이들의 예로는 AGRkMS, EAD, ISAD(G), MoReq2,<sup>7)</sup> OAIS, PREMIS 등이 있다.

한편, 특정한 목적을 가지고 만들어지는 시스템을 위한 메타데이터 스키마를 정의하기 위해서는, 시스템의 요구조건에 따라 메타데이터 표준을 적절하게 선택하고 조합시켜야 한다. 다시 말해, 목적에 의해 맞춰진 메타데이터 스키마, 어플리케이션 프로파일(Application profile)<sup>8)</sup>을 정의할 필요가 있는 것이다. 더불어, 시스템간의 데이터 교환을 위해서 메타데이터 스키마간의 매핑을 정의하는 것도 요구 되어 진다.

일반적으로 메타데이터 스키마는 데이터 모델을 가지고 있고, 모델은 기술 대상이 되는 엔티티(Entity)와 그들이 가지고 있는 속성(Property)을 정의한다. 그리고 메타데이터 스키마에 포함되는 기술 요소는 속성을 나타낸다. 예를 들어, 보존을 위한 메타데이터 표준 PREMIS에서 제공하는 데이터 모델은 5가지 타입의 엔티티(Intellectual entity, Digital Object, Event, Right, Agent)를 가지고 있다. PREMIS에서는 속성을 의미 단위(Semantic units)라고 부르고, PREMIS의 목적에 따라 이들 각 엔티티의 의미 단위를 정의하는 데이터 사전을 준비, 제공하고 있다. PREMIS를 이용하기 위해서는 데이터 사전으로부터 필요한 의미 단위를 골라 기술 항목으로 정하고, 거기에 적절한 기술상의 제약을 부여하여, 목적에 따른 메타데이터 스키마를 만드는 것이 된다.

---

7) MoReq(Model Requirements for the Management of Electronic Records)는 유럽에서 주로 사용되는 것으로 2001년 출판되었다. MoReq2는 MoReq의 2판으로 관련 정보와 문서(documentation)의 테스트와 함께 포괄적인 전자관리기록시스템을 위한 형식적인 요구사항 명세서로 구성되어 있다.

8) 어플리케이션 프로파일(Application profile)은 여러 개의 어휘정의를 가진 요소의 세트로부터, 실제로 응용할 때마다 필요한 것을 선택하여 구조 제약(構造制約)을 정의하는 것이다. 어휘 정의와 구조 제약을 나누는 것은 상호운용성을 높이는 데 도움을 준다(日本總務省, 2012).

다양한 메타데이터 스키마는 기록 생애 주기 내에서 이용되고, 아카이브와 보존을 위한 업무와 관련된다. 그러나 일반적으로 기록 생애 주기는 메타데이터 스키마 안에서 명확하게 나타나 있지 않다. 다시 말하면, 기록 생애 주기의 어느 단계에서, 어느 메타데이터 표준을 사용해야 하는 가, 또한 메타데이터 표준이 정하는 기술 요소가 각 단계에서 어떻게 이용되어야 하는 가 등에 대해 정의되어 있지 않는 것이다.

예를 들어 PREMIS의 경우, 'title'이나 'creator'와 같은 지적 엔티티(Intellectual entity)에는 더블린 코어(DC)<sup>9)</sup>와 같은 표준 요소의 이용을 가정하고 있다. 기록 생애 주기의 처음 단계에서 자원이 작성될 때, 이 같은 엔티티의 기술 요소에 값(Value)이 정해졌다고 해서 작성 단계가 끝난 후에 이 값이 변하는 일은 없다. 그러나 디지털 오브젝트(Digital object) 외의 엔티티의 경우, 기록 생애 주기의 단계가 뒤로 갈수록 값이 변경되는 경우는 있다. 이처럼, 메타데이터의 데이터 모델에 있어서 엔티티와 기록 생애 주기 간의 관계는 명확하게 나타나 있지 않다.

디지털 자원의 아카이브와 보존을 위한 메타데이터 스키마에 대한 고찰에 근거하여, 본 논문에서는 기록 생애 주기의 전체에서 사용되는 메타데이터 스키마의 선택과 조합을 지원하기 위한 메타데이터 스키마의 분석 방법론을 제안한다.

이전의 논문에서 우리는 디지털 자원의 보존과 아카이브를 위한 메타데이터 스키마의 특징 분석을 제안하였다(Baek Jaeun, 2010).<sup>10)</sup> 논문에서는 아카이브와 보존을 위한 메타데이터 스키마로, Digital Preservation

---

9) DC(Dublin Core) 메타데이터는 국제 표준으로 자원에 대한 기술의 이용을 위한 15개의 속성 어휘로 구성되어 있다. 15개의 기술 요소는 Dublin Core Metadata Initiative (DCMI)에 의해 기술적인 설명서(Technical specifications)와 메타데이터 어휘 세트에 대해 기술하고 있다.

10) Baek Jaeun, Shigeo Sugimoto, "Feature Analysis of Metadata Schemas for Records Management and Archives from the Viewpoint of Records Lifecycle", *Journal of Korean Society of Archives and Records Management*, 2010, 10(2), pp. 75~100.

Coalition(DPC)<sup>11)</sup>에 의해 정의된 보존 프로세스인 의사결정나무(Decision tree)로부터 추출한 속성 세트와 AGLS,<sup>12)</sup> EAD, ISAD(G), OAIS, PREMIS가 가지고 있는 속성 세트를 이용하였다. AGLS와 DPC의 속성 세트는 아카이브 혹은 보존을 위한 메타데이터 스키마를 목적으로 디자인된 것은 아니지만, 우리는 아카이브와 보존을 위한 메타데이터의 특징을 보이기 위해 이들을 비교 대상으로 삼고 연구대상에 포함하였다. 이들 메타데이터 요소의 특징 분석 결과로부터 메타데이터 표준과 기록 생애 주기 단계 간의 관계를 볼 수 있었다. 이 결과에 근거하여, 우리는 기록 생애 주기에서 행해지는 업무의 관점으로부터, 메타데이터 표준 요소와 기록 생애 주기와의 관계에 대해 상세한 분석을 실시하였다.

이전 연구의 결과를 통해, 하나의 메타데이터는 아카이브와 보존을 위해 전체의 기록 생애 주기를 포괄하지 못하다는 것을 명백하게 보여주었다. 그리고 기록 생애 주기의 단계 혹은 단계 내의 업무(Task)에 따라서 그들을 조합하거나 메타데이터 스키마로부터 기술적인 요소를 적당하게 선택해서 요구해야 한다는 것도 보여주었다. 이 결과를 기초로, 본 연구는 그들의 기술 요소 간의 관계를 분류하고, 스키마를 분석하는 프레임워크를 제안한다.

메타데이터 표준은 일반적으로 자원의 관점을 중심으로 정의되어 있고 이용목적에 맞춰 설계되었기 때문에, 기록 생애 주기 전체를 보면서 생애 주기 안의 각 단계에서 이용되는 메타데이터 표준간의 매핑을 실시하기 위한 적절한 방법이 없었다. 본 논문에서는 각기 다른 메타데이

---

11) Digital Preservation Coalition (DPC)는 디지털 자원의 안전한 보존을 위한 목표를, 공동 작업을 통해 촉진시키기 위해 2001년 영국에서 설립되었다. DPC는 일반적인 어드바이스 서비스와 가이드라인(예를 들면, 보존에 관한 핸드북, 기술 보고서, DPC의 연차 보고 등)을 제공한다.

12) AGLS(Australian Government Locator Service) 메타데이터 표준은 오스트레일리아의 표준, 5044-2010으로 온라인 정보와 서비스의 상호운용성, 처리 능력 등을 개선할 수 있는 관련 이용 가이드라인과 메타데이터 속성(Metadata Properties) 세트를 제공한다.

터 표준 요소의 사이를 기록 생애 주기에 포함하는 업무와 연결시키는 것으로, 메타데이터 표준간의 매핑 방법을 제안한다.

메타데이터 표준의 분석을 위한 프레임워크(Framework)로, 우리는 기록 생애 주기를 이용하여 자원의 업무를 기초로 해서 작성한 자원의 업무 지향 모델(Task-oriented model), 즉 태스크 모델(Task Model)을 제안한다. 태스크 모델의 이용으로 업무를 실행하는 '이벤트(Event)'의 관점을 보다 명확하게 할 수 있다. 또한 업무를 중심으로 기술 요소 간 매핑을 효율적으로 실시하기 위해서는, 요소를 카테고리화 하여 매핑 대상 범위를 좁히는 것이 중요하다.

우리는 각 메타데이터 스키마의 기술 요소를 균등하게 카테고리화하기 위해서 5W1H 속성세트(what, why, where, who, when, how)를 이용하여 메타데이터 표준에 포함되는 각 기술 요소를 분석하였다.

5W1H 속성은 신문기사나 작문과 같은 이벤트를 설명하는 기초 속성으로 널리 알려져 있다. 본 연구에서는 '자원에 대한 업무를 통해서, 기술대상의 실체에 발생하는 이벤트에 메타데이터 표준의 기술 요소가 이용될 수 있다'라고 하는 관점을 가지고, 메타데이터 기술 요소의 특징을 나타내기 위해 5W1H 속성을 적용하였다. 또한, 5W1H 속성은 메타데이터 표준의 기술 요소를 통일적인 관점에서 분석하는 데 있어서도 유용하다.

본 논문에서는 선택한 메타데이터 표준AGLS, AGRkMS, EAD, ISAD(G), OAIS, PREMIS, DPC의 기술 요소를 가지고, 태스크 모델과 5W1H 모델 그리고 양쪽 모두 분석을 실시하고, 이들 표준의 기술 요소간의 매핑을 예로 제시하고 있다.

본 논문은 다음과 같이 구성하고 있다. 2장은 본 연구를 위한 선행 연구로, 기록관리를 위한 메타데이터 표준과 기록 생애 주기 모델, 기록 생애 주기 모델을 기초로 한 아카이브와 보존 메타데이터 표준의 특징을 설명한 이전 연구, 그리고 관련 연구 순으로 나열, 서술하고 있다.

3장에서는 이 연구에서 제안하고 있는 기초 모델, 태스크 모델과 5W1H 모델에 대해 설명한다. 그리고 4장에서는 이 두 모델을 이용하여 실시할 각 메타데이터 표준 간 매핑 예와 정리를 기술한다. 한편, 이 논문의 범위에서는 매핑 전체 표를 포함시키는 것이 어렵기 때문에, 일부의 예만을 이용해서 설명하고, 마지막으로 결론을 논한다.

## 2. 선행 연구

### 1) 기록관리(Record management)를 위한 메타데이터표준

다음은 이 연구에서 참조하고 있는 메타데이터 표준에 대해서 간단히 소개한다.

우선, 아카이브를 위한 EAD(Encoded Archival Description)와 ISAD(G)(General International Standard Archival Description)에서, ① EAD는 디지털 자원을 위한 메타데이터 스키마로, 디지털 자원의 내용 기술뿐만 아니라, 구조적인 기술을 위한 요소도 포함하고 있고, ISAD(G)와 호환성을 유지하고 있다(Library of Congress, 2002). ISAD(G)는 국가기록원과 같은 기관에서 문서(Documents)를 아카이브 하기 위해서 제안된 것으로, 전통적인 아카이브(Traditional archive)를 위해 디자인 된 표준이다(International Council on Archives, 2000).

② 기록관리를 위한 메타데이터 AGRkMs(Australian Government Record keeping Metadata Standard)는 AGLS 메타데이터를 기초로, 오스트레일리아의 국가기록원(national archives)을 위해서 발표된 표준으로(National Archives of Australian, 2010), 여기서 권장하는 메타데이터에 대해 설명하고 있다(National Archives of Australian, 2011).

이어, 디지털 자원의 보존을 위한 OAIS(Open Archival Information

System)와 PREMIS(Preservation Metadata and Implementation Standard)에서, ③ OAIS는 디지털 자원의 보존을 위한 국제적인 표준으로, 정보에 접근(Access)을 보증하는 아카이브 시스템을 위한 참조모델을 제공하고, 보존을 위해서 필요한 메타데이터의 구조에 대해서 결정하고 있다(Harvard University Library, 2008). ④ PREMIS(Preservation Metadata and Implementation Standard)는 디지털 자원의 보존을 위해서 정의된 메타데이터 표준으로, 5개 엔티티의 속성(즉, 메타데이터의 기술 요소)을 정한 데이터 사전을 준비, 제공하고 있다(Online Computer Library Center, 2008).

이 같은 기록관리를 위한 메타데이터에 좀 더 추가하여, 우리는 웹 자원의 검색(Finding aid)을 위해 디자인된 AGLS, DPC가 제안하고 있는 보존을 위한 결정 프로세스(Decision Process)의 의사결정나무로부터 추출한 속성 세트를 포함하였다.

⑤ AGLS(Australian Government Locator Service Metadata)는 온라인 환경에서 디지털 자원과 비디지털 자원(Non-digital resource)에 관계없이, 검색을 개선시키기 위해 만들어진 오스트레일리아의 정부에서 제공하는 표준이다(National Archives of Australian, 2006).

마지막으로 메타데이터 스키마로 디자인 된 것은 아니지만, 본 연구에서 사용한 ⑥ 의사결정나무(Decision Tree, DPCset)는 DPC에서 제공하고 있는 가이드라인 DPH(Digital Preservation Handbook)에 의해 정의되고 있는 디지털 자원의 보존을 위한 결정 프로세스(decision process)로, 내용에 의해 Selection, Rights & Responsibility, Technology & Metadata, Documents & Costs의 순으로 질문과 응답이 하나의 조로 구성, 반복 된다(Digital Preservation Coalition, 2006).

의사결정나무는 보존을 위한 평가와 선택을 목적으로 하고 있으나, 기술 대상의 속성을 아카이브의 관점으로부터 기술하는 메타데이터 표준과는 다르다. 또한 의사결정나무, 그 자체는 메타데이터 스키마와 같

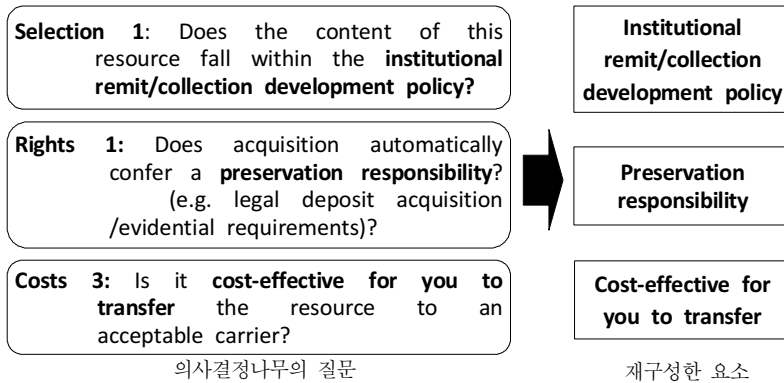
이 메타데이터가 가지는 구조나 모델, 기술 요소 등을 가지고 있지 않다. 즉, 의사결정나무는 메타데이터 스키마가 아닌, 다른 메타데이터의 기술 요소와 동일한 관점에서 비교하는 것이 어렵다. 그러나 다른 한편으로 의사결정나무는 메타데이터 표준과 동일하게 평가 대상으로 디지털 자원을 나타내고 있고, 평가를 위한 질문 항목은 자원이 가지는 속성을 나타내고 있다고 생각 할 수 있다.

평가를 위한 질문에는 보존을 위한 방법 혹은 적절한 기술을 선택하기 위한 결정적인 의미적 속성(시멘틱 속성, Semantic property)을 포함하고 있다. 이 같은 질문 내 의미적인 속성은 메타데이터 속성으로 변경시킬 수 있다. 또한 질문의 대답은 속성을 위한 값의 등급(Class of values) 혹은 속성 값이라고 할 수 있다. 이 같은 관점으로 본 연구에서는 DPC의 의사결정나무를 앞서 설명한 다른 표준과 동일하게 취급하여, 결정 프로세스의 질문 항목을 메타데이터 기술 요소로 변환하기로 하였다.

메타데이터 스키마로 의사결정나무를 전환하기 위해서 우리는 우선, 1. 의사결정나무의 각 질문에서 의미적인 특징을 찾고, 2. 질문 내용을 나타내는 중요한 구절 혹은 키워드를 추출하여, 3. 추출한 키워드와 구절을 메타데이터 스키마의 기술 요소로 재편성하였다. 예를 들어 “Do you need to acquire for other purpose?”라고 하는 Selection 3의 질문으로부터, 이 문장을 설명하고 있는 주요 단어인 ‘Acquire for other purpose’를 선택하여 기술 요소로 정하였다.

<그림 1>은 결정 프로세스의 질문을 메타데이터의 기술 요소로 변환한 몇 가지의 예를 제시하고 있다. 본 연구에서는 의사결정나무의 4개의 부분으로부터 질문을 추출하여 전부 27개 속성으로 정의하였다.

<그림 1> 질문에서 속성으로의 변환



## 2) 기록 생애 주기 모델

각 조직이나 기관은 그들이 가지고 있는 기록물의 관리 방침이나 가이드라인에 따라 자원을 작성하고, 이용, 폐기 그리고 보존한다. 일반적으로 이 같은 과정을 기록 생애 주기(Resource lifecycle 혹은 Records lifecycle)라고 부른다. 이 논문에서는 기록 생애 주기의 기초 모델로, 미국의 NARA(National Archives and Records Administration)<sup>13)</sup>의 기록 생애 주기 모델(Records Lifecycle)을 참조하였다. NARA의 기록 생애 주기는 자원의 타입에 관계없이 7단계, (1) Creation(작성), (2) Maintenance and use(유지와 이용), (3) Disposition(처분), (4) Arrangement and description(정리와 기술), (5) Preservation(보존), (6) Reference(참조), (7) Continuing use(연속적인 이용)로 구성되어 있다.

우리는 NARA의 기록 생애 주기의 7단계를, 이 연구에 맞춰 6단계로 재구성하였다. 우선, 각 단계의 명칭을 단순하게 하였고, NARA의 마지

13) National Archives and Records Administration의 웹 페이지 참조.  
<http://www.archives.gov/about/info/whats-a-record.html>.

막 2단계 Reference와 Continuing use를 하나의 단계로 정리하였다.

Reference와 Continuing use는 메타데이터의 관점에서 보면 주로 자원에의 접근(Access)과 검색을 나타내는 단계로, 이 두 단계의 차이는 그다지 크지 않다. 그러므로 본 연구에서는 이 두 단계를 하나의 단계인 Reference & Re-use로 재구성하였다. 그리고 우리는 명칭을 단순화하기 위해서, NARA의 Continuing use의 명칭을 Re-use로 변경, 정의하였다. NARA의 Continuing use는 자원의 지속적인 이용을 정의하는 단계로, 우리는 이 단계를 보존이 마친 자원이 재이용되는 단계로 보고 명칭을 변경하였다. 이렇게 재구성한 기록 생애 주기 모델은 본 연구에서 작성하고 제안하는 태스크 모델에 참조되었다. <그림 2>는 본 연구에서 이용하는 기록 생애 주기 모델을 보여 주고 있다.

<그림 2> 이 연구에서 이용되는 기록 생애 주기 모델



### 3) 기록 생애 주기를 바탕으로 메타데이터 표준 특징 분석

메타데이터는 기록 생애 주기 내의 단계에서 작성된다. 기록 생애 주기에서 메타데이터 작성 일정, 즉 메타데이터의 작성 타이밍은 메타데이터 관리 방침을 바탕으로 결정된다. 우리는 이전 연구에서 기록 생애 주기의 단계와 요소 간 관계에 대해 간단하게 특징 분석을 실시하였다 (Baek Jaeun, 2010).<sup>14)</sup>

14) Baek Jaeun, Shigeo Sugimoto, 앞의 글.

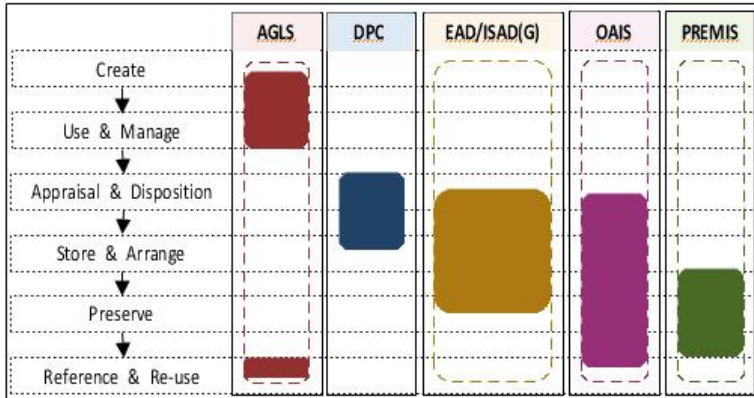
각 메타데이터 요소를 위한 값(Value)은 기록 생애 주기 내에서 처음 단계를 정할 수 있다. 예를 들어 ‘title’과 ‘creator’와 같은 요소는 자원의 작성 단계(Creation)에서 결정할 수 있으나, ‘provenance’와 ‘fixity’와 같은 요소는 평가 단계(Appraisal stage)의 다음에 결정되어진다. 우리는 이 연구에서 메타데이터 표준으로부터 가져온 메타데이터 기술 요소를 기록 생애 주기의 각 단계에 적용시켰다. 그리고 메타데이터 표준으로부터 가져온 모든 요소의 값이 제일 처음으로 지정되는 혹은 수정되는 최초의 단계를 결정하였다.

<표 1> 수치로 살펴본 기록 생애 주기 내의 메타데이터 표준 (%)

Lifecycle Metadata	AGLS	DPC decision tree	EAD / ISAD(G)	OAIS	PREMIS
Create (작성)	16		11	1	5
Use & Manage (이용과 관리)	28		13	2	22
Appraisal & Disposition (평가와 폐기)	5	61	14	13	
Store & Arrange (저장과 정리)	18	39	33	30	21
Preserve (보존)	15		20	39	45
Reference & Re-use (참조와 재이용)	18		9	15	7

<표 1>과 <그림 3>에서는 기록 생애 주기 단계에서의 메타데이터 요소 매핑을 간단하게 보여주고 있다. <표 1>에서는 각 표준 요소의 백분율을 나타내고 있고, <그림 3>에서는 그들 값이 할당하는 또는 재 할당되는 각 표준의 요소가 가장 많은 숫자를 나타내는 최초 단계를 어두운 부분으로 나타내고 있다. 이 연구에서 우리는 각 표준이 그들의 성질, 특징, 목표에 따라서 작성, 이용되고 있는 것을 고찰하였다.

<그림 3> 기록 생애 주기의 단계에 의해 보여지는 메타데이터 요소



#### 4) 관련 연구

이 장에서는 넓은 범위에서 우리의 연구와 관계가 있는 연구로, 디지털 자원의 기록관리, 디지털 아카이브를 위한 메타데이터의 문제점, 메타데이터의 상호운용성과 매핑, 메타데이터의 선택 등에 대한 연구에 대해서 간단히 소개한다.

① Baca(2003)<sup>15)</sup>는 문화유산자원을 위한 적절한 메타데이터 스키마의 선택에 주목하고 있다. 이를 위해 CDWA,<sup>16)</sup> EAD, MARC,<sup>17)</sup> VRA Core<sup>18)</sup>

15) Baca, Murtha, "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information", *Cataloging & Classification Quarterly*, 2003, 36(3-4), pp. 47~55. ([http://polaris.gseis.ucla.edu/gleazer/260\\_readings/Baca.pdf](http://polaris.gseis.ucla.edu/gleazer/260_readings/Baca.pdf))

16) CDMA(Categories for the Description of Works of Art)는 예술품, 건축물 (architecture), 그 외의 물질문화, 예술품의 수집품과 관련된 사진 등의 정보에 접근하고 기술하기 위한 개념적인 프레임워크를 이용하고, 프레임워크를 표현할 수 있는 예술 데이터베이스의 콘텐츠에 대해 기술하고 있다.

와 같이 각각 다른 목적을 가지고 있는 다양한 메타데이터를 이용하여, 그들의 요소 간의 크로스워킹(Crosswalk)과 메타데이터의 매핑을 진행하고, 이에 대해 서술하고 있다. Baca의 연구는 메타데이터의 적절한 선택을 위한 매핑과 분류를 다루고 있는 점에서 본 연구와 관련성을 가지고 있다.

② Chan와 Zeng (2006)<sup>19)</sup>은 다양한 메타데이터 스키마와 함께 상호운용성에 대한 문제에 관해서 연구를 하고 있다. 그리고 동일한 상호운용성의 관점으로부터 Schema level, Record level, Repository level의 3가지 레벨에 대해 설명하고 있다. 우리는 이 연구를 통해서 메타데이터의 선택과 조합의 필요성이 상호운용성에서부터 비롯된다는 것을 다시 확인하였고, 이 연구배경(상호운용성의 필요성 등)을 본 연구에 참조하였다.

③ Day (2001)<sup>20)</sup>는 디지털 보존을 위한 메타데이터에 관한 최신 연구와 개발에 대해 나열하고 있다. 또한 보존 계획을 위한 메타데이터의 중요성, 디지털 보존 문제 등에 대해서도 소개하고 있다. 특히 이 논문에서는 기록관리를 위한 메타데이터(Metadata formats of record keeping), 기록관리 프로젝트, OAIS 참조모델 등에 대해 서술하고 있다. 본 연구에서는 기록관리를 위한 메타데이터와 기록관리에 대한 여러 정의와 서술 등을 참조하였다.

---

17) MARC(Machine Readable Cataloging)는 미국 의회도서관과 MARC표준의 사무소에 의해 개발된 것으로, 온라인 장서 목록 내에서 목록 작성 데이터를 부호화하기 위한 표준 세트이다.

18) VRA Core(Visual Resources Association Core)는 시각적 문화 작품의 기술을 위한 데이터 표준이다. 게다가 이들 요소가 조직적으로 구조화하는 방법으로 최초의 청사진부터, 타이틀, 지역, 요일등과 같은 정보의 단위로 메타데이터 요소 세트로 구성되어 있다.

19) Chan, Lois Mai, Marcia Lei Zeng, "Metadata Interoperability and Standardization - A Study of Methodology Part 1", D-Lib Magazine, 12(6), 2006.  
(<http://www.dlib.org/dlib/june06/chan/06chan.html>)

20) Day, Michael, "Metadata for Digital Preservation: A Review of Recent Developments", 2001, *ECDL 2001*, LNCS 2163, pp. 161~172.

④ Evans et al (2005)<sup>21)</sup>는 미래의 아카이브 시스템 Clever Record keeping Metadata Project(CRKMP)의 적합성과 다양한 아카이브를 목적으로 하는 메타데이터 개발에 대해 설명하고 있다. CRKMP는 비즈니스 시스템과 레코드 키팅의 현 시스템, 아카이브 시스템 사이에서 자동적으로 메타데이터를 작성하고 공유하는 것에 대한 과제를 검토하고 있다. 우리는 이 연구를 통해서 레코드 키팅에 대한 현황과 관련 연구 그리고 CRKMP에서 제공하는 개념적인 프레임워크인 레코드 연속이론과 모델(Records continuum model)에 대한 개념 설명으로 현 기록관리 시스템과 메타데이터의 필요성, 관련 연구에 대해 조사할 수 있었다.

⑤ Haslhofer와 Klas(2010)<sup>22)</sup>는 현 정보 시스템에서 이용되고 있는 메타데이터 시스템, 개념, 메타데이터의 상호운용성 그리고 그들이 가지고 있는 문제점에 대해 기술하고 있다. 또한 메타데이터 상호운용성을 위한 기술(Technique)에 대한 설명과 상호운용성의 기술의 특성 등에 대해 나열하고 있다. 이 논문에서 우리는 메타데이터의 상호운용성에 대한 기본개념과 상호운용성의 필요성, 기술(Technique), 의의 등을 본 연구의 기초적인 개념과 특징 분석을 위한 상호운용성의 필요성에 대한 설명을 참조하였다.

⑥ McKemish 외 (2009)<sup>23)</sup>는 기록 연속모델(Records continuum model)에 근거하여, 기록관리를 위한 개념적인 프레임워크를 기술하고 있다. 이 연구에서는 기록관리에 대한 전반적인 설명을 나타내고 있어 우리는 이 설명들과 함께 본 연구에서 사용한 AGLS 메타데이터, AGRkMS 그

---

21) Evans ed., "Create Once, Use Many Times: The Clever Use of Record keeping Metadata for Multiple Archival Purposes", *Archival Science*, 2005(5), pp. 17~42.

22) Haslhofer ed., "A Survey of Techniques for Achieving Metadata Interoperability", *ACM Computing Surveys* 42(2), 2010, Article 7, pp. 1~37.

23) McKemish ed., 2009, Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema. (<http://infotech.monash.edu/research/groups/rcrg/publications/archiv01.html>)

리고 기록관리를 위한 메타데이터에 대한 설명을 참조하였다.

⑦ Zeng (1999)<sup>24)</sup>는 역사적 가치가 있는 패션 수집품(Collection)에 대한 기존의 메타데이터 포맷의 어플리케이션에 대해 설명하고 있다. 그리고 디지털화한 역사적 가치가 있는 패션 수집자원을 위한 목록을 개발하고 있다. 이를 위해 3개의 스키마(AACR, Dublin Core, Visual Resources Associations (VRA) core)가 이 연구에서 이용되고 있다. 이 연구의 기술 대상은 본 연구의 기술 대상과는 다르나, 디지털 자원이라는 공통점을 가지고 있고, 메타데이터 표준 어플리케이션의 필요성에 대해 서술하고 있다. 기술대상의 목록작성을 위해 스키마의 요소를 선택하여 비교하고 고찰하는 Zeng의 연구 내용은 기록관리를 위해 메타데이터 표준의 매핑을 실시하는 본 연구와 관련성을 가지고 있다.

마지막으로, ⑧ 광정(2007)은 기록관리 메타데이터의 설계와 행정 기관 사례에 대한 연구를 기술하고 있다. 그리고 기록 생애 주기에서의 현용·준 현용 단계에 주목하고, 기록관리 메타데이터와 업무와의 관련성, 필요성에 대해 자세하게 설명하고 있다. 기록관리 메타데이터에의 접근, 메타데이터와 업무의 관련성, 기록 생애 주기에서 요구되는 메타데이터의 정립 등에 대한 이 연구의 서술은 본 연구에서 설명하고 있는 기록관리 메타데이터의 중요성과 기록 생애 주기, 그리고 기록관리 메타데이터의 분석과 관련성을 보여주고 있다.

### 3. 특징 분석을 위한 기초 모델 - 태스크 모델과 5W1H모델

---

24) Zeng, Marcia Lei, "Metadata Elements for Object Description and Representation: A Case Report from a Digitized Historical Fashion Collection Project", *Journal of the American Society for Information Science*, 50(13), 1999, pp. 1193~1208.

이 연구에서 우리는 기록관리를 위한 메타데이터 표준 요소를 분류하기 위해 2개의 모델—태스크 모델(Task Model)과 5W1H 모델—을 작성, 제안한다. 이 모델들에 대한 설명과 두 모델의 조합, 5W1H 모델의 개념에 대한 정의를 다음에서 설명한다. 우리는 이 모델을 이용하여, 메타데이터 표준의 특징을 분석하였다.

### 1) 태스크 모델과 5W1H 모델의 의의(意義)

우리는 이전 연구를 바탕으로 상세한 메타데이터 표준의 특징 분석을 위해서, 아카이브와 보존을 위한 메타데이터를 포함한 기록관리를 위한 메타데이터 표준을 선택하여 매핑과 분류를 실시하였다.

메타데이터 표준은 일반적으로 자원의 관점을 중심으로 이용 목적에 맞춰 설계되었기 때문에, 기록 생애 주기 전체를 보는 업무의 관점은 반영되지 않는다. 그러나 기록 생애 주기 전체를 보는 업무의 관점을 반영하는 것으로, 기록 생애 주기와 그 안의 업무에 따른 메타데이터 표준의 사용, 그리고 메타데이터 요소간의 매핑을 실시하는 것이 보다 쉬워진다.

요소 간의 매핑은 지금까지 널리 행해지고 있지만, 원래부터 자원중심의 관점에서의 매핑이었을 뿐, 업무의 관점은 포함되어 있지 않았다. 다시 말해, 업무의 관점을 포함하는 것으로 인해 요소의 의미를 보다 확실히 할 수 있는 것이다.

본 연구에서는 자원을 위한 메타데이터를 업무 중심으로 보기 위해서, 기록 생애 주기를 업무의 관점으로 정리한 태스크 모델을 제안하고 작성하였다. 그리고 태스크 모델과 함께 5W1H 모델을 이용하였다.

태스크 모델은 기록 생애 주기 안에서 메타데이터 표준이 사용되는 단계를 명확하게 할 수 있다. 단, 기록 생애 주기 전체를 볼 경우에는

복수의 메타데이터 표준만을 다루는 것이 되기 때문에, 우리는 상세한 분석을 위해 기술 요소간의 의미적인 매핑을 실시하지 않으면 안 된다. 본 연구에서는 기술 요소가 무엇을 표현하고 있는지를 알기 위한 의미적인 관점(Semantic viewpoint)에서 매핑을 실시하기 위해, 5W1H를 이용하여 카테고리화 하였다.

태스크 모델은 업무를 실행하려고 하는 ‘이벤트’의 관점이 명확하기 때문에 업무에 관한 5W1H로 나눠 분석이 가능하다. 즉, 관련하는 자원과, 업무와 자원을 연결하는 요소를 5W1H 카테고리로 나눠서 분석하는 것이 가능한 것이다. 또한, 업무 내에서 사용되는 요소의 의미를 5W1H를 이용하여 카테고리화 하는 것으로, 요소의 의미가 보다 확실하게 되고 한정시킬 수 있으며, 대응관계가 명확해짐과 동시에, 조사대상을 좁힐 수 있다.

이 프레임워크의 제안과 이용에 의해서, 메타데이터의 특징 분석을 위해 실시된 매핑 혹은 분류는 단순히 일반적인 매핑이 아니라 의미적인 분류가 가능하게 된다. 또한 우리는 매핑이 단순한 개념 이용이 아닌 의미적인 분류가 될 수 있도록 하기 위해, 키워드를 정하였다. 다음에서는 태스크 모델과 5W1H모델에 대해 설명하고, 정의를 나열한다.

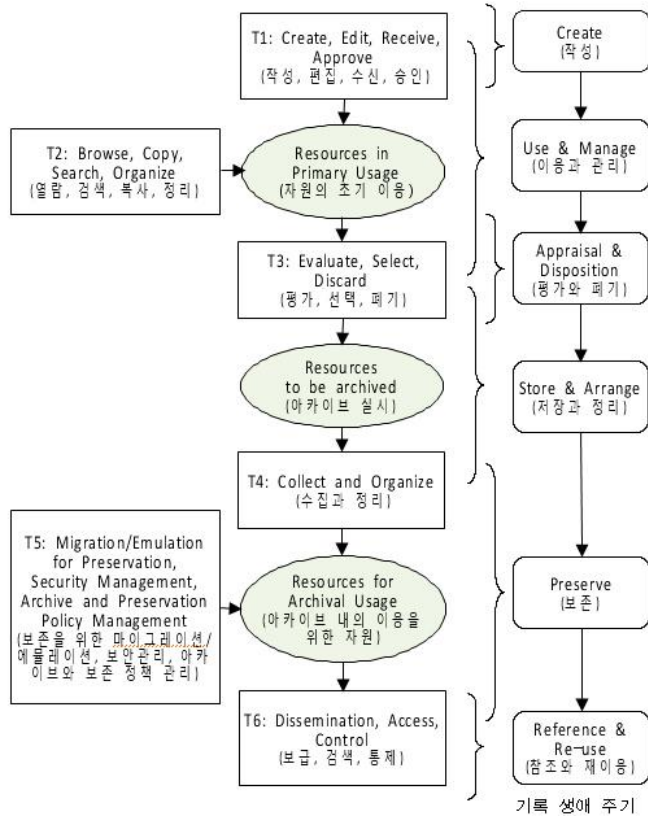
## 2) 기록관리 업무 중심 모델-태스크 모델

기록 생애 주기는 기관 혹은 조직 내에서 자원 작성부터 보존에 이르기까지의 과정을 각 단계별로 나눠서 정의하고 있다. 본 연구에서는 기록 생애 주기를 이용하여 태스크 모델을 작성하였다.

본 연구에서 제안하는 태스크 모델은 기록 생애 주기와 매우 비슷하나, 기록 생애 주기의 각 단계에서 행하는 업무를 정의하고, 자원과 자원에 행해지는 업무와의 관계를 상세하게 나타내고 있다. <그림 4>에서는 태스크 모델과 기록 생애 주기 모델을 나열하여 보여주고 있다. 태

스크 모델은 6개의 태스크 그룹(Task group) (T1-T6)으로 구성하고 있고, 각 업무를 다음과 같이 정의한다.

<그림 4> 태스크 모델 (Task Model)



- Task 1 (Creation task): 자원의 초기 작성 과정으로, 자원의 작성이 완성될 때까지 행해지는 여러 가지 편집, 승인, 수속 등의 업무
- Task 2 (Primary Usage task): 주로 이용자를 위해서 작성한 자원을 이용 및 유통하는 과정으로, 이 과정에서의 검색이나 접근에 대한 관리 등의 업무

Task 3 (Appraisal and Discard task): 자원의 목적이나 기간에 의해서 이용이 끝난 자원을 보존하기 위해 선택하는 과정으로, 보존을 위해 필요한 자원을 선별하고, 선택을 위한 평가, 파기, 보존 정책의 적용 등의 업무

Task 4 (Management and Archival Transformation task): 아카이브가 시작되는 단계로, 보존 대상의 자원을 모아서 변환하는 과정이다. 보존 대상 자원의 수집, 축적, 이전(移轉), 정리, 보관, 관리, 관련 자원 수집 등의 업무

Task 5 (Preservation task): 자원의 장기보존을 위해서, 자원의 유지관리를 위해 필요한 관리, 변환, 시스템, 보존정책의 적용 등의 업무

Task 6 (Re-use task): 보존된 자원을 재이용하는 과정으로, 보존된 자원을 유지, 관리하면서 재이용을 위해서 자원의 복사, 검색, 관리 등의 업무

### 3) 메타데이터 기술 요소와 업무 중심 모델의 조합

#### (1) 업무 중심의 관점에서의 메타데이터

업무는 기록 생애 주기의 각 단계에 따라서 실행된다. 이 같은 업무로는 예를 들면, 작성, 검색이나 수정, 평가, 수집이나 관리, 보존 등이 있다. 자원은 기록 생애 주기를 바탕으로 작성한 업무 중심의 태스크 모델에서 실행되는 업무와 업무 내용, 목적에 의해 변한다.

자원은 메타데이터 표준 요소에 의해 기술(記述)되고, 자원을 위한 메타데이터는 각 업무를 행하는데 이용된다. 메타데이터는 태스크 모델의 각 단계에 의해서 내용이 갱신되는 자원에 대해 기술해야 한다. 다시 말해, 태스크 모델의 단계에 의해 자원 수집 혹은 폐기, 보존을 위

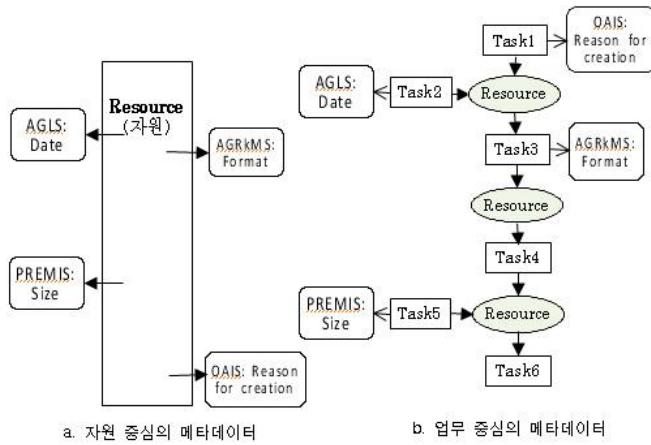
한 포맷의 변환 등과 같은, 업무 실행에 의해 내용이 갱신되는 자원의 메타데이터를 기술해야 하는 것이다.

본 연구에서는 자원을 기술하는 메타데이터 표준 요소가 자원에 실행되는 업무를 기술할 수 있는 지에 대해서 조사하였다. <그림 5>에서는 이 고찰의 결과를 예로 보여주고 있다.

이전 연구에서 우리는 아카이브와 보존을 위한 메타데이터가 기록 생성 주기와 관련이 있다는 것을 설명하였다. 이를 바탕으로 한 메타데이터 요소와 업무의 관계를 <그림 5a, b>를 통해서 설명한다.

<그림 5a>는 자원을 기술하기 위해 이용되고 있는 여러 가지 메타데이터 표준 요소를 보여 주고 있고, 이 기술 요소들을 태스크 모델 위에 적절하게 적용한 것을 <그림 5b>에서 보여 주고 있다.

<그림 5> 자원과 업무중심에서의 메타데이터



예를 들어 설명하면, 자원을 기술하는 메타데이터 중 OAIS의 reason for creation은 자원 작성의 이유를 기술하는 요소이다. 이 요소를 태스크 모델에 적용해 보면, reason for creation은 태스크 모델의 1단계

(Creation Task)에서 자원이 작성된 이유를 나타내기 위해서 사용된다.

우리는 이전의 연구와 자원을 기술하는 메타데이터 표준의 요소를 태스크 모델에 적용함으로써, 기술 요소가 자원의 업무와 태스크 모델을 표현할 수 있다고 하는 것을 확인할 수 있었다. 업무에 의해 실행되는 자원, 자원을 기술하는 메타데이터, 즉 메타데이터는 업무를 기술하는 것이 가능한 것이다.

## (2) 업무 지향 모델과 메타데이터 요소의 조합

태스크 모델에서는 자원과 관련이 있는 여러 가지 업무가 행해진다. 우리는 자원을 기술하는 메타데이터 요소를 업무 관점에서 의미적인 분류를 하기 위해 5W1H 모델을 이용하여 카테고리화를 실시하였다. <그림 6>에서는 하나의 업무 내에서 기술하고 실행되는 관련 업무들을 분류하여 나타내고 있다.

<그림 6a>에서 보여주고 있듯이, 「A Task(하나의 업무)」는 권리, 시간, 목적, 사람 등과 같은 엔티티와 관련이 되어 있고, 이 엔티티는 업무를 행하는 이유나 가이드라인, 업무를 실시하는 장소나 기관 등, 업무 내에서 역할을 가지고 나타내는 실체이다. 일반적으로 하나의 업무와 그들과 관련이 있는 엔티티와의 관계는 매 업무마다 결정되어진다.

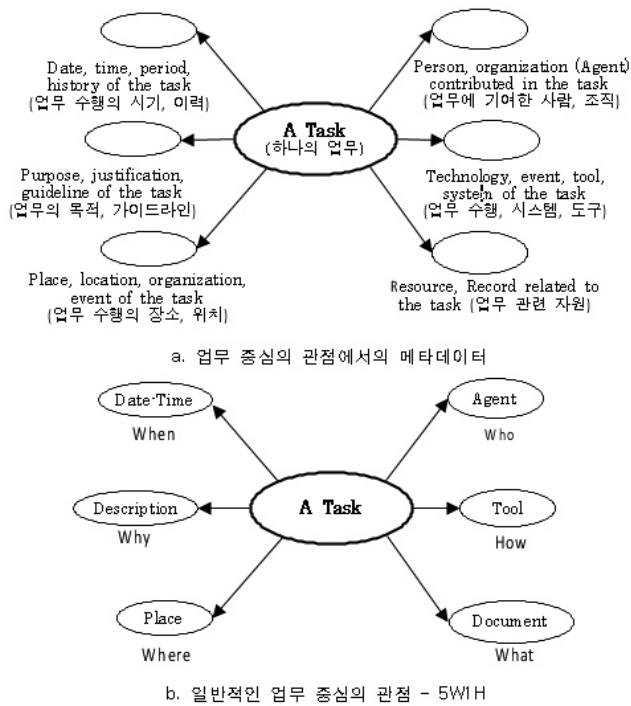
이 연구에서 우리는 <그림 6b>에서 보여지고 있는 것과 같이 업무와 관련 엔티티의 관계를 나타내는 일반적인 카테고리로 5W1H 속성을 이용하는 것을 제안한다. 5W1H 카테고리를 이용함으로써 업무에 관련되고 있는 엔티티의 타입에 따라 업무와의 관계를 나타낼 수 있었다.

업무와 관련이 있는 엔티티는 각 목적마다 결정된 특정한 스키마에 의해 메타데이터의 값(Value)으로 기록된다. 그러나 일반적으로 메타데이터 표준의 데이터 모델은 업무가 아니라, 데이터의 엔티티에 근거하

여 정의되고 있다. 이는 우리의 이전 연구에서 보여준 기록 생애 주기와 메타데이터 요소 간의 관련성에도 불구하고, 메타데이터의 기술 요소가 업무와의 관련을 명확하게 하고 있지 않다는 것을 의미한다.

메타데이터 표준의 데이터 모델과 본 모델의 차이는 자원을 기술하는 요소에 참조되는 업무를 명확하게 하고, 업무와의 관계가 5W1H 모델을 이용하여 분류된다는 점이다. 또한, 이 연구의 근본적인 아이디어는 전통적인 매핑에서 실시하는 데이터의 엔티티 중심의 관점을 대신 하여, 일반적인 업무 중심 관점에서 메타데이터를 이용한 것이다. 이 연구에서 우리는 앞서 서술한 업무의 관점에서 기술 요소의 특성을 이

<그림 6> 업무중심의 관점에서의 메타데이터와 5W1H



용하여, 제시한 메타데이터 표준의 특성을 확인하는 동시에, 그들의 기술 요소간의 매핑을 실시하였다.

#### 4) 5W1H 모델

5W1H 모델 속성은 메타데이터 표준 요소의 카테고리를 확인하기 위해 이용한다. 5W1H 속성에 의해 나타내고 있는 카테고리를 이 논문에서는 5W1H 카테고리라고 부른다. 다음에서는 5W1H 카테고리 정의에 대해 설명하고 있다.

- 1) Who: 자원에 관련된 에이전트(Agent), 업무에 관련하는 사람 혹은 조직, 기관(작성자, 이용·관리자 혹은 기관, 평가자 혹은 부서, 기록관리 전문가(Archivist) 등)
- 2) When: 자원에 대한 작업이 행해진 시간, 시기, 기간, 일정, 시대(작성날짜, (재)이용 혹은 관리기간, 수집 날짜, 평가 날짜, 보존 기간 등)
- 3) Where: 자원에 대한 작업을 실시하는 장소, 위치, 지역, 기관이나 조직(작성부서, 이용 혹은 관리조직, 수집 혹은 정리기관, 보존 기관 등)
- 4) What: 자원에 대한 기술, 자원에 행하는 작업 내용의 사례에 대한 기술(보존업무와 프로세스에 대한 정보, 자원의 서지사항, 이용·평가·수집·보존에 관한 기준, 법률, 규칙, 관련 자원 등)
- 5) How: 자원에 관한 기술(技術)적인 특징, 자원과 관련 자원에 실시하는 작업/운동(파일 포맷, 소프트웨어와 하드웨어, 검색과 환경 시스템, 법률, 변환 도구 등)
- 6) Why: 자원에 관한 작업의 이유 혹은 목적(작성 목적, 보존을 위한 이유 등)

## 4. 기초 모델을 이용한 메타데이터 표준간의 매핑

이 장에서는 앞서 선택한 메타데이터 표준의 기술 요소 간 매핑 예를 제시한다. 우리는 선택한 메타데이터 표준 요소(AGLS, EAD, AGRkMS, OAIS, PREMIS 그리고 DPC 의사결정나무의 속성)를 가지고, 6.1장에서 보여줄 키워드를 이용하여 2개의 타입으로 매핑을 실시하였다.

우선 각 메타데이터 표준 요소를 해당하는 업무 내에서 의미적으로 분류하기 위해서 5W1H 카테고리를 이용하여 매핑을 실시하였다. 그리고 5W1H 카테고리를 통해 실시한 매핑을 자원의 업무로 구성하고 있는 태스크 그룹에서 다시 재분류하였다. 단, 실시한 매핑의 전체 표는 사이즈가 너무 크기 때문에 본 논문에는 포함하지 않고, 표의 윤곽을 보여주기 위해 일부분을 <부록 1>에 수록하였다. 전체 매핑 표는 저자의 웹 사이트에 기재하여 제공하고 있다(<http://www.slis.tsukuba.ac.jp/digitalarchive/DC2011/> for the mapping tables).

### 1) 5W1H 카테고리와 업무 식별을 위한 일반적인 어휘

우리는 메타데이터 표준 요소간의 의미적인 관계를 조사하기 위해 5W1H 카테고리와 태스크 모델을 이용하여 매핑과 분류를 실시하였다. 이 매핑은 의미적인 관계를 정하는 것이기 때문에, 저자가 제시하는 통일적인 매핑의 규칙에 근거하여 수작업으로 진행하였다.

우리는 매핑을 위해 <표 2>와 <표 3>에서 나타내고 있는 키워드를 이용하여 분류하였다. 이 표에서 나타내고 있는 키워드는 각 카테고리 혹은 업무 내에서 이용되는 일반적인 용어로, 이 키워드를 중심으로 같은 의미, 넓은 의미 혹은 좁은 의미 등의 관계에서 매핑을 실시하였다.

<표 2> 5W1H 모델에 해당하는 분류 어휘 (예)

5W1H model	Key words
Who	Agent, Institution, Name, Organization, People, Person etc
When	Date(s), Period, Time etc
Where	Agent, Country, Institution, Location, Organization, Place etc
What	Bibliography, Description, History, Relationship, Right etc
How	Action, Event, Technique, Tool, Transference etc
Why	Purpose, Reason etc

<표 3> 태스크 모델에 해당하는 분류 어휘 (예)

Task model	Keywords
T1: Create, Receive, Approve	Create, Make, Produce etc
T2: Browse, Copy, Search, Organize	Access, Manage, Use etc
T3: Evaluate, Select, Discard	Accept, Appraise, Destruct, Select etc
T4: Collect and Organize	Archive, Collect, Manage, Store etc
T5: Migration/Emulation for Preservation, Archive/ Preservation Policy Management	Archive, Manage, Store, Preserve etc
T6: Dissemination, Access, Control, Search	Access, Search, Use etc

## 2) 5W1H 카테고리에서의 기술 요소 분류

매핑을 위해서, 우리는 각 요소가 나타내고 있는 정의, 가이드라인, 어플리케이션을 위한 서비스, 다양한 정보 등을 참조하여, 앞서 서술한 키워드를 중심으로 분류를 실시하였다.

다음에서는 메타데이터 표준에서 선택된 요소와 5W1H 카테고리를 이용하여 실시한 매핑과 분류에 대해 예를 통해서 설명한다.

(1) AGLS Metadata의 *Publisher*

AGLS에서 예로 추출한 요소, *Publisher*는 자원을 이용할 수 있도록 하는 사람이나 조직의 엔티티를 의미한다. 이 요소는 정보 서비스에 접근을 허락하고 제공하는 조직의 세부사항을 제공하는 데 이용된다.

5W1H 카테고리의 키워드에 근거하면, 조직이나 기관과 같은 에이전트는 장소로써 이용되고 있고, 이로 인하여 *Publisher*는 Who와 Where 양쪽 모두에 분류되어진다.

*Publisher*에 해당하는 요소는 EAD와 OAIS의 요소로, 그 외의 표준에는 해당하는 기술 요소가 없고, 이는 다른 메타데이터에는 일치하는 요소가 없다는 것을 의미 한다.<sup>25)</sup> *Publisher*와 일치하는 요소를 가지고 있지 않은 AGRkMS와 PREMIS, DPC에 대해 간단히 서술하면, AGRkMS는 기록관리를 위한 최소한의 메타데이터 표준으로, 자원을 기술하는 일반적인 메타데이터 요소는 AGLS로부터 이용한다. PREMIS의 경우, 지적 엔티티에 관련이 있는 요소는 다른 메타데이터 표준을 사용하는 것을 전제로 하고 있다. 그리고 DPC의 경우는 보존을 위한 평가 프로세스에서 추출한 속성이기 때문에, AGLS가 다루는 지적 내용에 관한 기술 요소는 포함되어 있지 않다.

<표 4> AGLS: *Publisher*

5W1H model	Metadata Standards					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
Who	Publisher			Publication Statement Publisher	Name of publisher	
Where	Publisher			Publication Statement Publisher	Place of Publication Name of publisher	

25) 표 안에서 비어 있는 곳은 해당하는 요소가 없음을 의미한다.

(2) AGRkMs의 *Date Range*

AGRkMs의 요소, *Date Range*는 엔티티에 관련되는 날짜, 시간을 의미하고, 하위요소로 *Start Date*와 *End Date*를 가지고 있다.

*Date Range*의 정의에 의하면, 이 요소는 5W1H 모델의 키워드에서 제공하는 Dates와 Times를 가지고 있다. 이들 요소는 명확히 When에 해당된다. <표 5>에서는 AGRkMs의 *Date Range*에 해당하는 AGLS, EAD, OAIS, PREMIS의 각 요소를 보여주고 있다.

<표 5> AGRkMS: *Date Range, Start Date and End Date*

5W1H model	Metadata Standards					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
When	Date	Date Range		Date	Date of Publication	dateCreatedByApplication
		Start Date			Change History Before Archiving	
						PreservationLevelDateAssigned
		End Date				

(3) DPC의 의사 결정 나무(Decision Tree)속성의 *Multiple media formats*

DPC의 속성 세트에서 추출한 *Multiple media formats*는 자원의 종류, 디지털 혹은 비디지털 자원에 관계없이, 자원이 가지고 있는 하나 이상의 미디어 포맷을 의미한다.

*Format*은 두 가지의 의미로 설명할 수 있다. 하나는 자원의 미디어 타입을 의미하고, 다른 하나는 자원을 만드는 데 필요로 하는 기술 (Technology)을 의미한다. 이 요소는 이 두 가지 의미를 모두 가지고 있어, 미디어 타입을 나타내는 처음의 의미 What에, 두 번째의 의미로

5W1H 카테고리의 How에 분류된다.

<표 6> DPC Attribute Set: *Multiple media formats*

5W1H model	Metadata Standards					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
What	Format	Format	Multiple media formats			
How	Format	Format	Multiple media formats			Format

(4) EAD의 *Title of the Unit*

EAD의 요소, Title of the Unit는 기술하고 있는 자원의 이름, 명칭을 의미한다. Title of the Unit는 업무에서 다루는 자원의 이름을 나타내고 있기 때문에, 5W1H 카테고리에서는 What에 분류된다. 그리고 EAD의 Title of the Unit에 해당하는 메타데이터 표준으로는 AGLS, AGRkMs, OAIS가 있다.

<표 7> EAD: *Title of the Unit*

5W1H model	Metadata Schemas for Archive					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
What	Title	Name		Title of the Unit	Resource description	

(5) OAIS의 *Reason for Creation*

OAIS의 요소인 Reason for Creation element는 자원의 작성 이유를 명시하는 데 이용된다. 이 요소는 이유나 목적의 의미를 가지고 있는 Why에 카테고리 되고, 이 요소에 해당하는 AGLS의 요소, Description은 넓은 범위의 요소로 이 요소와 매핑 된다.

<표 8> OAIS: Reason for Creation

5W1H model	Metadata Schemas for Archive					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
Why	Description				Reason for Creation	

### (6) PREMIS의 Size

PREMIS의 요소, *Size*는 자원의 크기, 파일의 크기와 같이 기술적인 값 (Technical value)을 나타내고 있다. 기술적인 값을 나타내는 이 요소를 키워드에 근거하면, 대부분 How에 분류된다. *Size*에는 비슷한 의미를 가지고 있는 AGRkMS의 *Format*, EAD의 *Extent*와 매핑 하고 있고, 넓은 의미를 가지는 AGLS의 *Description*에도 대응하고 있다.

<표 9> PREMIS: Size

5W1H model	Metadata Schemas for Archive					
	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
How	Description	Format		Extent		Size

### 3) 태스크 그룹(Task Group)에서의 재분류

우리는 6.2장에서 매핑한 메타데이터 표준 요소(5W1H 카테고리에서의 요소 분류)를 5W1H 모델과 조합시켜 태스크 그룹에서 재분류를 실시하였다. 매핑 표는 5.1장에서 설명한 태스크 그룹(Task groups)에 따라 정리하였다.

<표 10> 기록관리를 위한 메타데이터 스키마의 매핑의 예  
(T3: Evaluate, Select, Discard)

TASK Model	SWIH model	AGLS	AGRKMS	Decision Tree of DPC	OAIS	EAD	PREMIS	
T3	Who		Position*			Sponsor*		
						Publisher*		
						Publication Statement*		
	When	Date*	Date Range* Start Date*				Date*	
			End Date*				Publication Statement*	
			Identifier Scheme*				Date of the Unit*	
	Where		Position*				Sponsor*	
							Author*	
	What		permissions*				Publisher*	
		Mandate*		Security Covenant*				
		Rights*		Rights*				
					Long term value justify preservation			
					Acquire for other purposes			
					Commit adequate staff			
						Ingest Process History	Processing Information	
							Appraisal Information	
			Identifier*	Identifier Scheme*				
			Format*	Format*				
	How		permissions*					
		Mandate*		Negotiate for the source to supply				
Rights*			Rights*					
				Digital version be selected for preservation		Table Column Specification*		
Format*			Format*	Manageable file format Carrier that is acceptable for transfer and/or storage				
				Document Form				
				Technically feasible for you to transfer the material Available to you online or on a physical carrier				
			Change History*	Ingest Process History*				
Why						Revision Description*		
				Long term value justify preservation*				
			Acquire for other purposes*					
			Document Form*					

앞서 언급했듯이 매핑의 전체 표는 논문에 포함할 수 없어 이 장에서는 일부분, 태스크 모델의 T3: Evaluate, Select, Discard, 내에서의 매핑을 예로서 설명한다. 이는 <표 10>에서 보여주고 있다.

태스크 그룹에서 T3: Evaluate, Select, Discard 은 아카이브를 위해 보존해야 할 자원을 선택하고 평가하는 등의 업무를 실시하는 단계이다. 태스크 모델에 해당하는 분류 키워드(<표 3> 참조)에 의하면, T3은 평가, 선택, 파괴(폐기), 승인 등의 키워드를 가지고 있다. 이 키워드를 바탕으로 분류를 실시한 결과, T3에서는 PREMIS를 제외한 모든 메타데이터 표준 요소가 이 매핑에 포함되었다.

T3의 일부분을 예로 설명하면, T3에서 자원을 실행하기 위한 포맷/환경(기술적인 의미를 가진 포맷/Technology)을 나타내는 How의 Format에는 AGLS의 Format\*, AGRkMs의 Format\*, DPC의 다수 요소, EAD의Table Column Specification\* 등과 관련을 가지고 매핑되고 있다. 이는 일부분의 예로, Format에는 더 많은 메타데이터 요소가 대응되고 있다.

이 매핑을 통해서, 우리는 기록관리를 위한 메타데이터 표준 요소가 나타내는 태스크 모델의 단계와, 각 태스크 그룹 내에서 5W1H 카테고리 분류된 요소와 요소 간 관계를 확인할 수 있었다.

#### 4) 기초 모델을 이용한 분류와 매핑 정리

우리는 선택한 기존의 메타데이터 표준의 기술 요소를 가지고, 5W1H 카테고리화 태스크 그룹으로 매핑과 분류를 실시하였다. 이 장에서는 6.2장과 6.3장에서 실시한 매핑과 분류의 결과를 정리하고 이에 대해 간단히 설명한다.

6.2장에 실시한 5W1H 카테고리에 의한 메타데이터 요소의 매핑을 통해서, 우리는 메타데이터 표준 요소가 나타내는 5W1H 카테고리화 5W1H 카테고리 내에서의 요소 관계를 살펴 볼 수 있었다. 앞서 설명한 매핑의 예에 이어, 이번에는 기술 요소의 5W1H 카테고리화의 결과를 숫자로 정리하였다. 이는 <표 11>에서 나타나고 있다.

우리는 분류를 위해 각 메타데이터 표준에서 추출한 기술 요소를 해

당하는 카테고리에 적용하였다. 기술 요소의 적용기준으로 앞서 언급한 키워드를 중심으로 분류를 실시하였으나, 단, 기술 요소를 나타내는 키워드가 5W1H 내에서 2개의 카테고리에 포함될 경우는 양쪽 모두에 분류·적용시켰다.

<표 11> 숫자로 나타낸 5W1H의 카테고리에 해당하는 메타데이터 표준

Metadata 5W1H	AGLS	AGRkMs	Decision tree of DPC	EAD	OAIS	PREMIS
Who	4	1	0	4	18	2
When	1	3	0	2	9	7
Where	4	3	0	8	23	7
What	15	31	16	24	99	21
How	7	28	12	43	47	125
Why	1	1	2	3	0	2

예를 들어, AGLS의 *format*은 자원의 기술(Description)에 대한 정보와 기술(Technology)에 대한 정보를 가지고 있다. 기술 요소가 다양한 의미를 가지고 있고 이에 대한 키워드를 모두 나타내고 있을 경우, 우리는 기술 요소가 해당되는 카테고리 모두에 분류하였다. 즉, AGLS의 *format*을 What과 How 양쪽 모두에 적용되는 것이다. 우리는 이 의미적인 분류로 인해 각 메타데이터 표준 요소가 가지고 있는 ‘이벤트’의 관점을 명확하게 할 수 있었다.

우리는 이 매핑을 통해 각 메타데이터 표준 요소를 의미적으로 분류할 수 있었고, 또한 메타데이터 표준이 가장 많이 적용되는 5W1H 카테고리를 살펴볼 수 있었다. 매핑 결과를 숫자로 정리한 <표 11>에서 AGLS 메타데이터 요소를 예로 설명하면, AGLS는 5W1H 카테고리 중에서 What에 가장 많이 적용되고, PREMIS의 기술 요소는 How에 가장 많이 분류되고 있었다. 이는 AGLS의 기술요소가 ‘자원에 대한 기술’에 관한 의미를 강하게 가지고 있거나 혹은 관련되는 요소가 많은 것을 의미

하고, PREMIS의 기술 요소는 ‘자원에 관한 기술(技術)적인 특징’에 대한 의미를 나타내는 요소가 많다는 것을 의미한다.

5W1H 카테고리에 의한 분류에 이어, 6.3장에서 실시한 매핑을 통해서 우리는 각 메타데이터 표준 요소가 해당하는 태스크 그룹과, 태스크 그룹 내의 5W1H 카테고리와 다른 메타데이터 표준 요소간의 관계를 살펴 볼 수 있었다. 우리는 각 태스크 그룹 내의 5W1H 카테고리에 해당하는 메타데이터 표준 매핑 역시 숫자로 정리하였다. 이는 <표 12>에서 보여주고 있다.

<표 12> 숫자로 나타낸 태스크 그룹과 5W1H에 해당하는 메타데이터 표준

Task model	5W1H	AGLS	AGRkMs	Dicision tree of DPC	EAD	OAIS	PREMIS
Task 1	Who	3 / 4			15 /18		
	When	1 / 1	3 / 3		5 / 9		1 / 7
	Where	3 / 4	1 / 3		16 /23		
	What	6 /15	4 /31		26 /99		
	How	3 / 7	1 /28		11 /47		2 /125
	Why	1 / 1				1 / 3	
Task 2	Who	2 / 4	1 / 1		15 /18	1 / 4	
	When	1 / 1	3 / 3		5 / 9		
	Where	1 / 4	3 / 3		18 /23	1 / 8	
	What	11 /15	21 /31		29 /99	3 /24	
	How	5 / 7	15 /28		9 /47	2 /43	11 /125
	Why		1 / 1				
Task 3	Who		1 / 1		16 /18		
	When	1 / 1	3 / 3		5 / 9		
	Where		2 / 3		15 /23		
	What	10 /15	11 /31	5 /16	19 /99	1 /24	
	How	6 / 7	11 /28	11 /12	10 /47	1 /43	
	Why	1 / 1	1 / 1	2 / 2			
Task 4	Who		1 / 1		17 /18	3 / 4	
	When	1 / 1	3 / 3		9 / 9	2 / 2	1 / 7
	Where		3 / 3		12 /23	7 / 8	1 / 7
	What	3 /15	20 /31		91 / 99	23 /24	2 /21
	How	3 / 7	18 /28		46 /47	42 /43	31 /125
	Why		1 / 1			2 / 3	

Task 5	Who				17 / 18	2 / 4	2 / 2
	When	1 / 1	3 / 3		8 / 9	2 / 2	5 / 7
	Where		1 / 3		21 / 23	7 / 8	5 / 7
	What	3 / 15	1 / 31	11 / 16	61 / 99	21 / 24	18 / 21
	How	3 / 7	1 / 28	1 / 12	32 / 47	42 / 43	119 / 125
	Why					3 / 3	2 / 2
Task 6	Who	2 / 4			16 / 18	1 / 4	
	When	1 / 1	3 / 3		5 / 9		
	Where	1 / 4			17 / 23	1 / 8	
	What	4 / 15			24 / 99	3 / 24	
	How	3 / 7			10 / 47	3 / 43	11 / 125
	Why						

<표 12>는 <표 11>과 다르게 매핑의 결과를 2개의 숫자로 나타내고 있다. 간단히 이 숫자들에 대해 설명하면, 숫자는 크게 두 개의 의미를 가지고 있다

우선, 앞의 숫자는 각 태스크 그룹 내의 5W1H 카테고리에 해당하는 메타데이터 표준의 숫자이고, 뒤의 숫자는 각 메타데이터 표준을 5W1H의 각 카테고리에 분류한 숫자로, <표 11>에서 나타내고 있는 숫자이다. 즉, 5W1H의 각 카테고리에 해당하는 메타데이터 요소의 수 중, 각 태스크 그룹 내의 5W1H에 해당하는 기술 요소의 수를 나타낸 것이다.

예를 들면, AGLS의 Task 1에서 Who는 '3/4'라고 표시하고 있다. '3/4'에서 4는 AGLS의 요소 중 Who에 해당되는 요소의 개수, 3은 AGLS의 요소 중 Task 1의 Who에 해당하는 요소의 개수이다. 우리는 이 같은 방식으로 <표 12>를 작성하였고, 표의 빈 칸은 모두 0/0을 나타내고 있다.

<표 12>에서 어두운 부분은 각 태스크 그룹 내의 5W1H 카테고리 전체를 포괄하는 메타데이터 표준이다. 태스크 모델에서 메타데이터 표준이 포괄하는 각 업무 단계(태스크 그룹)를 살펴보면, 태스크 그룹1(작성 업무)은 리소스의 검색을 위해 작성된 AGLS, 태스크 그룹 2(주요한 이용업무), 3(평가와 폐기업무), 4(관리와 아카이브를 위한 변환업무)는 리소스 관리를 위한 메타데이터 표준 AGRkMs가 해당되고 있다. 태스크

그룹 4와 5(보존업무)는 보존을 위한 메타데이터 OAS, 태스크 그룹 5에 해당하는 메타데이터는 PREMIS이다. 마지막으로, 태스크 그룹6(재이용 업무)에서 5W1H 전체를 포괄하는 메타데이터 표준은 없었다. 이 외의 메타데이터 표준 DPC 속성과 EAD는 태스크 그룹과 5W1H 카테고리에 해당되고 있지만, 태스크 그룹 내의 5W1H 카테고리 전체에 해당되고 있지는 않았다.

또한 우리는 이 분류와 매핑을 통해 각 태스크 그룹에 해당하는 메타데이터 표준과, 각 태스크 그룹 내에서 메타데이터 표준이 해당하는 5W1H 카테고리를 조사할 수 있었다. 예를 들어 아카이브와 보존을 위해 자원을 평가하고 선택 혹은 폐기에 관한 업무를 진행하는 태스크 그룹3(Task 3: Evaluate, Select, Discard)을 살펴보면, PREMIS를 제외한 모든 메타데이터 표준은 5W1H 카테고리의 What, How를 나타내고 있다. What와 How를 제외한 5W1H 카테고리에 해당하는 메타데이터 표준으로, AGLS는 When, AGRkMs는 When, Where, Why, DPC는 Why, 그리고 EAD는 Who, When, Where에 해당되고 있다. OAS는 What, How에 분류되고 있다.

우리는 이 2 개의 매핑을 통해 업무의 관점으로 작성한 프레임워크, 즉, 태스크 모델 내에서 각 메타데이터 표준이 어떤 업무 단계를 중점적으로 나타내고 있으며, 의미적인 분류를 통해 메타데이터 표준이 해당하는 태스크 그룹 내의 5W1H 카테고리가 무엇인지에 대해 살펴보았다. 그리고 다른 메타데이터 표준과의 매핑을 통해서 요소들 간의 관계도 볼 수 있었다.

## 5. 맺음말

디지털 자원의 급증에 따라 다양한 문제점이 발생하기 시작하고, 디지털 자원의 장기 관리와 보존이 점차 어려워지고 있다. 디지털 자원의 아카이브와 보존은 풀려고 노력하지만, 좀처럼 풀기 어렵다고 잘 알려진 이슈이고 이 이슈를 위해 현재 다양한 연구가 진행되고 있다. 여기서 우리는 아카이브와 보존을 포함한 기록관리 전체를 위한 메타데이터 표준의 특징 분석을 목표로 연구를 진행하고 있다.

메타데이터 커뮤니티에서 잘 알려져 있듯이 어플리케이션 프로파일의 개념은 메타데이터의 상호운용성에 있어서 매우 중요하다. 또한 어플리케이션 프로파일의 개념에서는 메타데이터의 상호운용성을 높이기 위해 기존의 메타데이터 표준 요소를 적절하게 선택하고 조합하는 것이 필요하다는 것을 우리에게 알려주고 있다. 다시 말해, 이는 어플리케이션 프로파일을 적절하게 작성하는 것이 메타데이터 표준의 특징을 잘 분석하기 위해 필요하고 중요하다는 것을 말한다. 이를 통해서 우리는 메타데이터 표준의 분석을 위해서 여러 메타데이터 표준 사이를 동일한 카테고리에서 적절하게 매핑 하는 것이 중요하다는 것을 배웠다.

우리는 기록관리를 위한 메타데이터 표준 요소의 특징 분석을 위해 우선, 기록 생애 주기를 바탕으로 자원의 업무 관점에서 작성한 태스크 모델과 5W1H 모델을 제안하였다. 그리고 이 모델들과 함께 선택한 기존의 메타데이터 표준을 이용하여, 표준 간의 매핑과 분류를 실시하였다. 이 연구를 통해서 우리는 기록관리를 위한 메타데이터의 조합이 중요하다는 것을 다시 확인할 수 있었다.

이 연구의 핵심적인 포인트는 기록 생애 주기를 기초로 하여 업무중심의 관점에서 메타데이터 표준을 보는 것이다. 메타데이터 요소의 의미는 주로 그들의 기본적인 데이터 모델에서부터 부여된다. 데이터 모

텔은 도메인에 포함되어 있는 엔티티의 분석과 엔티티 상의 업무, 이 모두를 기초로 하여 정의하고 있다. 그러나 기록 생애 주기 전체를 보면, 하나 이상의 메타데이터 표준을 조합시키는 것이 필요하기 때문에, 단지 표준간의 매핑을 하는 것이 아니라 데이터 모델을 기초로서 고찰해야 한다고 생각한다.

이 연구의 또 다른 포인트는 기록 생애 주기 모델에서 추출한 문맥상의 정보를 이용한 것이다. 우리는 업무와 5W1H 속성과 같은 동일한 문맥상의 정보가 메타데이터의 요소를 의미적으로 연결하는 것이 중요하다고 생각한다.

이 같은 관점에서 우리가 제안한 2개의 모델, 태스크 모델과 5W1H 모델이 유용하다고 생각한다. 이 프레임워크는 기록 생애 주기 내 업무 관점에서부터, 기술 요소의 의미를 확인하는 데 도움이 되는 단순한 의미의 기반을 제공하기 때문이다. 단순한 의미란 각각의 업무를 식별하는 데에 필요로 하는 부분과 업무의 관점에서부터 기술 요소를 동일시하는 데에 도움이 된다는 것을 나타낸다.

이 연구에서 기록관리를 위한 메타데이터 표준의 기술 요소의 특징을 분석하기 위해 실시한 전체 매핑 표는 온라인에서 제공하고 있다. 우리는 매핑 표를 모두 수작업으로 작성하였고, 이 표와 결과에 대한 실용상의 평가가 필요하다고 생각한다. 그리고 이 매핑 작업을 지원하는 도구(Tool), 매핑 표를 이용하기 위한 도구에 관해서도 앞으로의 연구 과제로 생각하고 있다. 우리는 다양한 메타데이터 스키마의 조합과 선택에 대한 연구의 진행과 함께 매핑 표에 대한 평가, 매핑의 이용을 지지하는 소프트웨어의 도구를 앞으로의 연구 과제로서 남겨놓았다.

## [참고문헌]

- 곽정, 「현용·준현용 단계의 기록관리를 위한 메타데이터」, 『기록학연구』 제16호, 2007
- Baca, Murtha, “Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information”, *Cataloging & Classification Quarterly*, 36(3-4), 47-55, 2003.
- Baek, Jaeun, Shigeo Sugimoto, “Feature Analysis of Metadata Schemas for Records Management and Archives from the Viewpoint of Records Lifecycle”, *Journal of Korean Society of Archives and Records Management*, 10(2), 2010.
- Brian F Lavoie, OCLC, “Meeting the challenges of digital preservation: The OAIS reference model”, *OCLC newsletter (Online)*, 243. 2000.  
<http://www.oclc.org/research/publications/archive/2000/lavoie/>  
(2011.03.06 참조)
- CC:DA (ALCTS/CCS/Committee on Cataloging: Description and Access), 2000, Task Force on Metadata: Final report, <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html> (2011.03.06 참조)
- Chan, Lois Mai, and Marcia Lei Zeng, 2006, Metadata Interoperability and Standardization - A Study of Methodology Part 1. *D-Lib Magazine*, 12(6).  
<http://www.dlib.org/dlib/june06/chan/06chan.html> (2011.03.06 참조)
- Consultative Committee for Space Data Systems (CCSDS), 2002, Reference Model for an Open Archival Information System (OAIS), <http://public.ccsds.org/publications/archive/650x0b1.pdf> (2011.03.06 참조)
- Day, Michael, “Metadata for Digital Preservation: A Review of Recent Developments”, *ECDL* 2001, LNCS 2163, 2001.
- Digital Preservation Coalition. 2006. Decision Tree for Selection of Digital

- Materials for Long-term Retention. <http://www.dpconline.org/advice/preservationhandbook/decision-tree> (2011.03.06 참조)
- Dublin Core Metadata Initiative, 2010. Dublin Core Metadata Element Set, Version 1.1. <http://dublincore.org/documents/2010/10/11/dces/> (2011.03.06 참조)
- Evans, Joanne, Sue Mckemmish and KarunaBhoday, “Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes”, *Archival Science*, 2005(5).
- Harvard University Library., 2008, Metadata Standards., <http://hul.harvard.edu/ois/digproj/metadata-standards.html> (2011.03.06 참조)
- Haslhofer, Bernhard, and Wolfgang Klas, “A Survey of Techniques for Achieving Metadata Interoperability”, *ACM Computing Surveys* 42(2), Article 7, 2010.
- International Council on Archives, 2000, ISAD(G): General International Standard Archival Description, Second edition. <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html> (2011.03.06 참조)
- International Organization for Standardization, 2006, ISO 23081-1:2006: Information and documentation-Records management processes-Metadata for records- Part 1: Principles.
- International Organization for Standardization, 2009, ISO 23081-2:2009: Information and documentation-managing metadata for records - Part 2: Conceptual and implementation issues.
- Library of Congress, 1999, MARC (MACHINE-Readable Cataloging) Formats, <http://lcweb.loc.gov/marc/> (2011.03.06 참조)
- Library of Congress, 2002, Encoded Archival Description, <http://www.loc.gov/ead> (2011.03.06 참조)

- Library of Congress, 2007, VRA Core 4.0, [http://www.loc.gov/standards/vracore/VRA\\_Core4\\_Intro.pdf](http://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf) (2011.03.06 참조)
- McKemmish, Sue, Glenda Acland, Nigel Ward and Barbara Reed, 2009, Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema, <http://infotech.monash.edu/research/groups/rcrg/publications/archiv01.html> (2012.03.06 참조)
- National Archives and Records Administration, <http://www.archives.gov/about/info/whats-a-record.html> (2011.03.06 참조)
- National Archives of Australia, 2006, Australian Government Implementation Manual : AGLS Metadata., [http://www.naa.gov.au/Images/AGLS%20Manual%203.0%20-%20most%20recent%20revision%2030%20September%202011\\_tcm16-49605.pdf](http://www.naa.gov.au/Images/AGLS%20Manual%203.0%20-%20most%20recent%20revision%2030%20September%202011_tcm16-49605.pdf) (2011.03.06 참조)
- National Archives of Australia, 2008, Australian Government Recordkeeping Metadata Standard Version 2.0, [http://www.naa.gov.au/Images/AGRkMS\\_Final%20Edit\\_16%2007%2008\\_Revised\\_tcm16-47131.pdf](http://www.naa.gov.au/Images/AGRkMS_Final%20Edit_16%2007%2008_Revised_tcm16-47131.pdf) (2011.03.06 참조)
- National Archives of Australia, 2011, Australian Government Recordkeeping Metadata Standard Implementation Guidelines Version 2.0, [http://www.naa.gov.au/Images/AGRkMS%20Implementation%20Guidelines\\_tcm16-50156.pdf](http://www.naa.gov.au/Images/AGRkMS%20Implementation%20Guidelines_tcm16-50156.pdf) (2012.03.06 참조)
- National Archives of Australia, 2011, AGLS Metadata Australian Government Implementation Manual Version 3.0, [http://www.naa.gov.au/Images/AGLS%20Manual%203.0%20-%20most%20recent%20revision%2030%20September%202011\\_tcm16-49605.pdf](http://www.naa.gov.au/Images/AGLS%20Manual%203.0%20-%20most%20recent%20revision%2030%20September%202011_tcm16-49605.pdf) (2012.03.06 참조)
- Online Computer Library Center, 2008, Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group, <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> (2011.03.06 참조)

- Shimazu, Keiko, atsya Arisawa, Isao Saito, 2006, “Interdisciplinary contents management using 5W1H interface for metadata”, IEEE/WIC/ACM International Conference on Web Intelligence (WI’06), pp. 909 - 912
- The getty, 2009, Categories for the Description of Works of Art, [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/introduction.html](http://www.getty.edu/research/publications/electronic_publications/cdwa/introduction.html) (2011.03.06 참조)
- Zeng, Marcia Lei, “Matadata Elements for Object Description and Representation: A Case Report form a Digitized Historical Fashion Collection Project”, *Journal of the American Society for Information Science* 50(13), 1999.
- 日本總務省, 2012, 「新 I C T 利活用サービス創出支援事業」 메타데이터 정보基盤構築事業, <http://www.meta-proj.jp/metaproj2010.pdf> (2012.03.06 참조)

<부록 1> 태스크 그룹과 5W1H 카테고리에서 기록관리를 위한  
메타데이터 표준 매핑 표 (T1, T2)

TASK Model	5W1H model	AGLS	AGRkMS	Decision Tree	OAIS	EAD	PREMIS	
T1	Who	Creator						
		Contributor						
		Publisher					Publication Statement	
						Publisher		
	When	Dates	DateRange Start Date				DateoftheUnit	
							Publication Statement*	
			End Date				Date	
							Imprint*	
							Item*	
	Where	Creator*						
		Contributor *						
			Location					
		Publisher*						Sponsor*
							Publication Statement*	
							Publisher*	
	What	Title	Name NameWords					Subject*
								Title
								Subtitle
		Language	Language					
							Language	
	Coverage							
	Description					Abstract		
How	Description *							
		Extent*						
	Mandate*							
	Rights*							
Why	Description *				Reason of creation			
T2	Who	Audience				Actors *		
			Position					
	When	Date*	DateRange* Start Date*				Date*	
							Publication Statement*	
			End Date*				Date of the Unit *	
Where	Contributor							

	*						
		Position*					
		Location*					
What	Format	Identifier					
		Format				Scope and Content	
						Dimensions*	
		Extent					
						Physical Description	
	Type	Category				Genre/Physical Characteristic	
	Description	Description				Abstract*	
	Function	Keyword					
	Subject						
	Coverage	Coverage					
	Mandate*						
	Rights*	Rights					
		Caveat text					
		Jurisdiction					
Source							
How		Identifier*					
	Format*	Format*				Table Column Specification*	
		Extent*					
		Document Form					
		permissions *					
	Mandate*						
	Relation*						
		Related Entity*					
	Category*						
				Actions *			
	Description*						
Why	Document Form*						

ABSTRACT

## **Analysis of Metadata Standards of Record Management for Metadata Interoperability**

From the viewpoint of the Task model and 5W1H

Baek, Jae-Eun·Shigeo Sugimoto

Metadata is well recognized as one of the foundational factors in archiving and long-term preservation of digital resources. There are several metadata standards for records management, archives and preservation, e.g. ISAD(G), EAD, AGRkMs, PREMIS, and OAIS. Consideration is important in selecting appropriate metadata standards in order to design metadata schema that meet the requirements of a particular archival system. Interoperability of metadata with other systems should be considered in schema design.

In our previous research, we have presented a feature analysis of metadata standards by identifying the primary resource lifecycle stages where each standard is applied. We have clarified that any single metadata standard cannot cover the whole records lifecycle for archiving and preservation.

Through this feature analysis, we analyzed the features of metadata in the whole records lifecycle, and we clarified the relationships between the metadata standards and the stages of the lifecycle. In the previous study, more detailed analysis was left for future study. This paper proposes to

analyze the metadata schemas from the viewpoint of tasks performed in the lifecycle. Metadata schemas are primarily defined to describe properties of a resource in accordance with the purposes of description, e.g. finding aids, records management, preservation and so forth. In other words, the metadata standards are resource- and purpose-centric, and the resource lifecycle is not explicitly reflected in the standards. There are no systematic methods for mapping between different metadata standards in accordance with the lifecycle.

This paper proposes a method for mapping between metadata standards based on the tasks contained in the resource lifecycle. We first propose a Task Model to clarify tasks applied to resources in each stage of the lifecycle. This model is created as a task-centric model to identify features of metadata standards and to create mappings among elements of those standards. It is important to categorize the elements in order to limit the semantic scope of mapping among elements and decrease the number of combinations of elements for mapping.

This paper proposes to use 5W1H (Who, What, Why, When, Where, How) model to categorize the elements. 5W1H categories are generally used for describing events, e.g. news articles. As performing a task on a resource causes an event and metadata elements are used in the event, we consider that the 5W1H categories are adequate to categorize the elements. By using these categories, we determine the features of every element of metadata standards which are AGLS, AGRkMS, PREMIS, EAD, OAIS and an attribute set extracted from DPC decision flow. Then, we perform the element mapping between the standards, and find the relationships between the standards.

In this study, we defined a set of terms for each of 5W1H categories,

which typically appear in the definition of an element, and used those terms to categorize the elements. For example, if the definition of an element includes the terms such as person and organization that mean a subject which contribute to create, modify a resource the element is categorized into the Who category. A single element can be categorized into one or more 5W1H categories. Thus, we categorized every element of the metadata standards using the 5W1H model, and then, we carried out mapping among the elements in each category.

We conclude that the Task Model provides a new viewpoint for metadata schemas and is useful to help us understand the features of metadata standards for records management and archives. The 5W1H model, which is defined based on the Task Model, provides us a core set of categories to semantically classify metadata elements from the viewpoint of an event caused by a task.

**Key words:** Archive, Archival Metadata, Digital Resource, Metadata Interoperability, Metadata Schema, Metadata Standard, Preservation Metadata, Records, Recordkeeping, Records Lifecycle, Record Management, Task Model, 5W1H Model, 5W1H Categories