

국가R&D 조사분석 검증자료의 장기적 보존 메타데이터 모델에 관한 연구

안정은*, 윤종민**, 고준호***

한국정보통신기술협회(TTA)

e-mail : jean@tta.or.kr*, jmyn@tta.or.kr**, jhko@tta.or.kr***

A Study on Preservation Metadata Model for long-term Preservation of the R&D Information

Jeong-Eun AN*, Jong-Min Yoon**, Jun-Ho ko***

TTA(Telecommunication Technology Association)

요 약

국가R&D사업의 추진결과와 활동현황에 대한 종합적이고 다각적인 분석을 위해 1999년도부터 매년 모든 국가R&D사업에 대해 주요 연구 성과인 논문, 특허, 기술료 등에 대한 국가R&D 조사분석 자료 수집이 실시되고 있다. 국가R&D 표준정보 수집을 위해 2003년도에 '국가연구개발사업 종합관리시스템(CORDI, 현재 SIMS)'이 구축되었고, 매년 국가R&D 조사분석 자료들의 수집·분석이 실시되고 있으나, 전자기록의 속성을 갖는 국가R&D 조사분석 자료, 특히 조사분석의 검증자료에 대한 장기적 보존을 위한 관련 연구는 전무한 상태이다.

따라서 본 연구는 국가R&D 조사분석의 검증자료에 대한 장기적 보존과 나아가서는 평가의 활용 및 정보 시스템간의 상호운용성을 높이기 위한 국가R&D 조사분석 검증자료의 보존 메타데이터 모델을 제안한다.

1. 서론

1999년도부터 시작한 국가R&D사업의 조사분석 및 평가는 국가적 차원에서 R&D 성과의 극대화와 효율화를 위한 프로그램 평가체계의 필요성으로부터 시작되었다. 조사분석의 체계는 정부 주도의 R&D 사업들이 대형화·복잡화됨에 따라 R&D 자원의 전략적 투자 등에 대한 종합조정을 뒷받침하기 위한 기초 통계자료의 요구에 따라 구축되었다. 예산의 분배와 조정에 있어 국가R&D사업 평가결과의 큰 요인으로 작용하고 있음에 따라, 평가의 기초가 되는 국가R&D사업 조사분석에 대한 중요성 또한 커지고 있다. 또한 국가R&D사업을 총괄하는 국가과학기술위원회는 과학기술기본법에 근거하여 매 5년마다 국가기술예측 및 기술수준, 기술지도 등을 작성하여 5년간 국가 과학기술전략을 담은 과학기술기본계획(5개년)을 작성하고 있다. 계획 수립을 위한 기초 통계자료로서 국가R&D사업의 조사분석 자료의 보존에 대한 연구가 시급하다. 조사분석 자료는 국가연구개발사업 종합관리시스템(SIMS)을 통해 정형화된 데이터로 수집·저장되고 있으나, 이를 검증하는 검증자료는 문서, 이미지, 엑셀, PDF 등 다양한 매체로 구성된다. 또한 국가R&D사업 조사분석의 검증자료는 시스템을 통해 수집·저장되고는 있으나 검증자료의 원본이 아닌 사본으로서 추후 복원에 대한 문제가 발생할 수 있다.

따라서 본 연구는 기록관리학의 표준모델인 OAIS 참

조모형(ISO 14721)과 OCLC/RLG¹⁾ PREMIS (Preservation Metadata : Implementation Strategies) 실무그룹에서 개발한 'Data Dictionary'의 객체 엔티티의 보존 메타데이터를 기초로 하여 국가R&D 조사분석 검증자료의 장기적 보존 메타데이터 모델을 제안한다. 또한 사례를 통해 본 연구의 실용성을 검증한다.

2. 관련연구

메타데이터는 통상 '데이터에 대한 데이터'로서 정의되는데, 'metamorphosis'라는 그리스어의 어원으로 'meta'는 '변화'를 의미한다. 따라서 메타데이터는 데이터의 기원, 이력에 관한 정보로, 인터넷의 대중화에 함께 본격적으로 사용되었다. 기록관리학에서는 '특정 대상물의 내용에 관한 조직화된 정보'의 개념으로 기록물과 관련된 메타데이터에 대한 다양한 연구가 진행되고 있다. ISO 15489-1(2001)은 메타데이터의 맥락, 내용, 구조, 일정기간 동안의 기록관리 이력을 기술한 데이터로 정의한다. 메타데이터의 역할은 다음 <표 1>과 같이 5개의 역할로 구분되어 정의된다.

1) OCLC(Online Computer Library Center) 전 세계 109개 나라의 54,000여개 도서관 회원을 대상으로 서지정보, 연구 지원 서비스를 제공하는 비영리기관 /RLG(Research Libraries Group) 150여개의 미국 연구중심대학 도서관, 박물관 및 각종 연구소 회원을 중심으로 연구와 학습을 지원하는 비영리기관

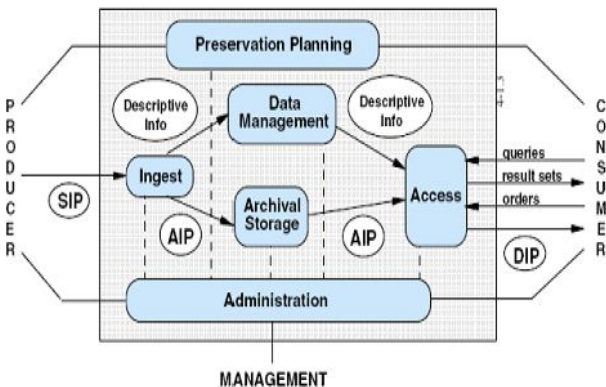
<표 1> 메타데이터의 역할

역할	내용
사용가능성	원하는 지역에 관한 데이터 셋이 존재하는 지에 관한 정보
사용적절성	데이터 셋이 특정한 목적에 적합한 지에 관한 정보
접근가능성	원하는 데이터 셋을 획득하는데 필요한 정보
교환가능성	데이터 셋을 처리하고 사용하는데 필요한 정보
관리정보	현재 존재하는 자료의 상태(자료모델, 품질, 완전성, 시간적 유용성 등)를 문서화하는데 필요한 정보로, 갱신 작업에 관한 내부 정책을 결정하는데 사용

메타데이터는 관련 데이터를 탐색하는데 소요되는 시간, 노력, 경비를 절감할 수 있다. 또한 특정 데이터를 관리하는 담당자가 교체되더라도 메타데이터의 정확한 정보 및 체계성으로 인해 업무의 연속성 확보가 용이하다. 일반적으로 메타데이터는 관리 메타데이터, 기록관리 메타데이터, 권한 메타데이터, 구조기술적 메타데이터, 보존 메타데이터로 구분한다.

이중 보존 메타데이터는 21세기 전자시대의 도래와 함께 문서, 이미지, 동영상 등 다양한 매체에 대한 정보 저장의 필요성에 따라 그 중요성이 부각되고 있다. 보존 메타데이터는 디지털 보존 과정을 지원하는 데이터로 디지털 정보의 장기간 유지가능성, 재연가능성, 이해가능성, 진본성, 무결성을 유지하는 기능을 한다. 문서, 이미지, 동영상 등 디지털 정보를 장기간 보존하기 위한 시스템, 즉 아카이브를 위한 개념적 구조는 2002년도에 표준으로 제정된 ISO 14721 OAIS(Open Archival Information System) 참조모형에서 정의하고 있고, 본 표준에서는 정보 모형, 정보 패키지 모형, 아카이브 기능모형 등의 디지털 아카이빙(Digital Archiving)과 관련된 기본 개념을 정의한다. (그림1)은 OAIS 참조모형을 나타낸 것으로, 디지털 객체물 저장, 데이터 관리, 행정관리, 보존계획, 이용 등 6가지로 디지털 아카이브를 정의한다.

생산자를 통해 장기 보존되어야 할 정보가 제공되면, OAIS를 통해 장기보존에 필요한 절차와 요소들이 체계화 되고, 이후 미래 이용자에게 장기 보존된 정보가 제공된다.



(그림 1) OAIS 참조모형

2) 파일 전송을 위해 파일 원본을 백업용으로 한 곳에 모아두는 것으로, 원본을 보호하지 않는 백업과는 차이가 있음

보존 메타데이터 요소에 대한 연구와 개발은 도서관 등 기록보존소를 중심으로 진행되고 있고, 연구와 개발의 큰 틀은 OAIS 참조모형을 준용 또는 참조하느냐에 따라 구분된다. 본 연구에서는 OAIS 참조모형을 준용하는 OCLC/RLG PREMIS 실무그룹에서 최근 발표한 'Data Dictionary' 메타데이터 요소를 상세히 분석하고, 본 연구에 기초 프레임워크로 활용한다. 'Data Dictionary'는 보존된 객체에 적용할 수 있는 핵심 메타데이터 요소로서, 지적 엔티티, 객체 엔티티, 이벤트 엔티티, 행위주체 엔티티, 권한 엔티티 등 5개의 엔티티 유형으로 구분되고 각 유형의 메타데이터 요소들을 조직화한 데이터 모델이다.

본 연구는 문서, 이미지, 동영상 등 전자자원의 저장, 즉 디지털 아카이빙을 목적으로 하기 때문에 위의 5개 엔티티 중 객체 엔티티만을 분석하고 그 상세한 데이터 모델은 아래 <표 2>에 나타내었다.

<표 2> 객체 엔티티의 보존 메타데이터 모델

main element	sub element		
object(M)	object type(M)		
	object value(M)		
storage stage(M)			
object category(M)			
object character(M)	composition stage(M)		
	fix(O)	message digest algorithm(M)	
		message digest(M)	
		message digest manufacturer(O)	
	size(O)		
	format	format direction(O)	format name(M)
			format version(O)
		format Registry(O)	format name(M) Registry
			format Registry key(O)
			format Registry role(O)
object's main attribute(O)			
inhibitorsn(O)	limit type(M)		
	limit scope(O)		
	limit clear key(O)		
manufacture application	manufacture application name(O)		
	manufacture application version(O)		
	manufacture date(O)		
manufacturer giving name(O)			
storage(M)	content site(O)	content site type(M)	
	content site(M)		
storage media(M)			
environment(O)	environment character(O)		

* (M) : Mandatory, (O) : Optional

<표 2>와 같이, 'Data Dictionary'의 객체 엔티티의 보존 메타데이터는 5개의 필수요소(M)과 3개의 선택요소(O)로 구성되고, 보존 데이터의 정보가 어떻게 저장되는지에 관계없이 디지털 아카이빙 시에 필요한 정보만을 핵심

요소로 정의한다. 객체의 관리를 위해 디지털 객체에 대한 정보만의 특성을 기술한 것이 특징으로, 본 연구의 목표인 국가R&D 조사분석 검증자료 저장에 적합한 데이터 셋이다.

3. 국가R&D 조사분석 검증자료의 장기적 보존 메타데이터 모델

본 연구의 국가R&D 조사분석 검증자료의 보존 메타데이터 모델은 9개의 메타데이터로 구성되는데, 참조한 OAIS 참조모형과 OCLC/RLG PREMIS 실무그룹에서 개발한 ‘Data Dictionary’의 데이터 모델과의 차별성은 국가 R&D사업이 갖는 고유 데이터 요소인 program과 performance를 갖는다는 것이다. program은 국가R&D사업별로 고유로 부여되는 식별번호이고 performance 또한 사업의 성과별로 부여되는 고유번호이다. 본 연구의 보존 메타데이터 모델은 범주(Category), 요소(Element), 하위 요소(Sub Element), 정의(Definition)로 구분되며, 9개의 필수요소와 1개의 선택요소로 구성된다. 특히 object(객체) 체, 즉 논문, 특허, 기술료의 검증자료 저장 장소 또는 저장매체의 이관, 폐기 시에도 변동사항을 확인할 수 있는 이력정보를 갖는다.

<표 3> 국가R&D 조사분석 검증자료의 보존 메타데이터 모델

category	main element	sub element	definition
origin	object Identifier(M)	classification(M)	객체의 종류로 논문, 특허, 기술료 중 택 1
		program(M)	객체에 부여된 고유 식별자로 과제고유번호 기재
		performance(M)	객체에 부여된 고유 식별자로 성과관리번호 기재
object Description(M)	object Description(M)	subject(M)	객체의 내용이 기술된 주제를 기술
		language(M)	객체의 내용 언어
		auxiliary description(M)	객체를 생산한 내용의 목적 등
		media format(M)	기록을 구성하는 정보의 포맷
format(M)	format(M)	data format(M)	기록을 구성하는 데이터의 논리적 포맷
		media(M)	저장된 매체에 대한 정보
		size(M)	기록물의 물리적 크기
right(M)	right(M)	title(M)	기록에 부여된 이름
		manufacturer(M)	기록 생산자
		creation date(M)	기록물이 생산된 날짜와 시간
		inhibitors(O)	기록물의 접근, 이용, 보존 활동에 제약이 있는 자
manage	storage(M)	content location(M)	기록물이 저장된 위치
		medium(M)	기록물이 저장된 물리적 매체
		environment(M)	소프트웨어(M)
	disposal(M)	hardware(O)	객체를 이용을 지원하는 하드웨어 환경
		status(M)	신규, 이관, 폐기 등 상태 기록
	disposal(M)	sentence(M)	기록에 지정된 보유기간
		action due(M)	신규, 이관, 폐기 등 처리 종류에 대한 날짜

다음 <표 4>는 국가R&D 조사분석 자료 중 논문 성과의 검증자료인 ‘매체가 한글인 연구논문’을 저장하기 위해 본 연구의 국가R&D 조사분석 검증자료의 보존 메타데이터 셋에 적용한 예이다. 과제고유번호 1575000013과 성과관리번호 2012 paper-1가 objet Identifier 데이터화되어 검증자료의 고유성을 확인할 수 있고, format을 통해 검증자료를 활성화시킬 수 있는 매체와 크기 등의 정보를 확인할 수 있다. 또한 disposal을 통해 검증자료의 신규, 이관, 폐기 정보를 확인할 수 있다.

<표 4> ‘매체가 한글인 연구논문’을 저장하기 위한 보존 메타데이터 셋

category	main element	sub element	data set	
origin	object Identifier(M)	classification(M)	논문	
		program(M)	1575000013	
		performance(M)	2012 paper-1	
object Description(M)	object Description(M)	subject(M)	국가R&D 조사분석 메타데이터	
		language(M)	한글	
		auxiliary description(M)	연구 성과	
	format(M)	media format(M)	문자, 이미지	
		data format(M)	한글	
		media(M)	한글 2012	
manage	disposal(M)	size(M)	50M	
		action due(M)	2012/1/1	
right(M)	right(M)	title(M)	국가R&D 조사분석 메타데이터에 관한 연구	
		manufacturer(M)	한국정보통신기술협회	
		creation date(M)	국가R&D 조사분석 메타데이터 논문	
		inhibitors(O)	없음	
manage	storage(M)	content location(M)	KISTEP 서버실	
		medium(M)	1번 서버	
		software(M)	한글	
	disposal(M)	hardware(O)	-	
		status(M)	신규	
	disposal(M)	disposal(M)	sentence(M)	2022/12/31
			action due(M)	2012/1/1

4. 결론 및 시사점

본 연구는 기록관리학의 표준모델인 OAIS 참조모형 (ISO 14721)과 OCLC/RLG PREMIS (Preservation Metadata: Implementation Strategies) 실무그룹에서 개발한 ‘Data Dictionary’의 객체 엔티티의 보존 메타데이터를 기초로 하여, 8개의 필수요소와 1개의 선택요소로 구성된 국가 R&D 조사분석 검증자료의 장기적 보존 메타데이터 모델을 제안하고, 실 예를 통해 유효성을 확인하였다. 본 연구는 예로서 유효성을 검증하였으나, 본 연구결과의 논리적 실효성을 확보하기 위해서는 수학적 분석을 통한 추가 연구가 필요하다. 또한 본 연구는 대상은 국가R&D 조사분석의 검증자료로 제한하였으나, 국가R&D 예산조정 및 배분에 직접적인 요인인 평가결과에 대한 응용 또한 가능할

것으로 기대한다.

Acknowledge

본 연구는 지식경제부 기술표준원의 지원을 받는 ‘정보통신표준화 및 인증지원사업’과 방송통신위원회의 ‘방송통신기술력향상사업’의 연구결과로 수행되었음

참고문헌

- [1] 박옥남(2012), PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구, 한국문헌정보학회지
- [2] 전방진(2011), 활용도 제고를 위한 지적정보의 메타데이터 표준화 적용방안, 한국지적정보학회
- [3] 김관준(2010), 연구정보를 위한 보존 메타데이터 요소 개발에 관한 연구, 정보관리학회지
- [4] 서은경(2005), 디지털자원의 보존 메타데이터 요소세트 개발에 관한 연구, 정보관리학회지
- [5] 전방진(2011), 활용도 제고를 위한 지적정보의 메타데이터 표준화 적용방안, 한국지적정보학회
- [6] 이경남(2007), 전자기록의 장기적 보존을 위한 보존 메타데이터 요소 분석, 한국외국어대학교
- [7] 이소연(2002), "디지털 아카이빙의 표준화와 OAIS 참조모형", 정보관리연구
- [8] 김희정(2003), "전자저널 아카이빙을 위한 OAIS 참조모형의 적용방안에 관한 연구", 연세대학교
- [9] 이소연(2002), "디지털 아카이빙의 표준화와 OAIS 참조모형", 정보관리연구
- [10] OCLC. "Digital Archiving Metadata Elements", Dublin, Ohio, 2004.
- [11] OASIS homepage : <http://www.OASIS.go.kr>