

서지정보 DB 구축을 위한 표준화(안)의 활용 사례

1998. 11. 4

발표자 : 조현양

연구개발정보센터



목 차



- 연구개발정보센터는
- 표준화의 시발점
- 표준화를 위한 시도
- 서지DB 입력을 위한 표준안
- 표준안 기반 데이터베이스 구축
- 표준 포맷의 활용
- 기대 효과



연구개발정보센터는

KORDIC
연구개발
정보센터

- 연구개발정보센터는
 - 연구개발에 필요한 정보
 - 그리고 연구개발의 성과물로서 생산되는 정보를
 - 정보통신기술을 활용하여
 - 언제 어디서나
 - 필요로 하는 누구든지
 - 신속하고 쉽게 사용할 수 있도록 공급하기 위한 목적으로
 - 과학기술부 산하의 정부출연 연구소
 - 1,250만건의 데이터 보유 (자체구축 : 4백만, 도입 : 8.5백만)

- <http://www.krodic.re.kr>



표준화의 시발점

KORDIC
연구개발
정보센터

- 작업 현장에서의 요구
- 데이터베이스 구축 시장으로의 뒤통은 출발 ('91년)
 - 해외 데이터베이스를 활용하기 위한 방법 도모
- 여러 기관이 하나의 데이터베이스를 구축
 - 과기부 산하 11개 연구소에서 분야별로 데이터베이스를 구축하되 이를 통합하여 서비스
- 국내외 기관간 데이터의 상호공유 및 교환 기대
 - 데이터의 양적, 질적 차원에서 수준 도달
- 서지정보 DB 구축과 관련한 국가적 차원의 표준화 유도
 - 국가적 차원에서의 재정보다는 현장에서의 요구에 따라



표준화를 위한 시도 (1)

- 전문위원회 구성
 - 연구단지 내 기술정보실 중심으로 구성
- 표준화와 관련된 원칙 설정 - DB 구축과 정보제공 분리
 - 구축 실무자를 위한 표준화
 - 검색시 제공되는 정보는 이용자 중심
- 외국의 사례 분석
 - 기계가독형(MARC: Machine Readable Cataloging)
 - LC MARC, OCLC MARC, UNIMARC, KORMARC 등
 - 실제 구축된 해외 데이터베이스의 데이터 분석
 - 공통 항목 (필수항목)과 조건항목을 구분



표준화를 위한 시도 (2)

- DB 구축을 위한 표준화 시안 마련
 - LC MARC, OCLC MARC, UNIMARC, KORMARC 등
 - OCLC MARC를 기준으로 채택
 - Books, Serials, Visual Materials, Mixed Materials, Maps, Scores, Sound Recordings, Computers Format (8가지)
 - 단행본, 보고서, 학위논문 등은 Books Format, 학술지와 Proceedings의 논문 단위의 서지정보 DB 구축은 Serials Format을 기반으로 채택
 - 170개의 항목(Field)과 각 항목하의 식별자를 포함하여 약 300 (레코드당 주어진 모든 사항을 입력한다고 가정)
 - 항목의 단순화 추진 (필수 항목을 선택하여 사용)
 - 기본항목 및 선택항목으로 구분하여 선정하되 선택항목의 채용은 입력기관에 일임



표준화를 위한 시도 (3)

- 정보 및 컴퓨터 전문가들의 의견 수렴
- 확장 개방형 표준안 마련
 - 새로운 형태의 등장에 대비하여 개방형 Format
 - 기관별 용도에 적합하게 데이터를 입력할 수 있도록
- 표준안의 수정, 보완 (실제 작업을 통해서)
- 과학기술분야 전문정보 데이터베이스 및 단행본 Union DB의 구축에 적용



서지DB 입력을 위한 표준안

- 데이터 입력을 위한 표준안 (별첨 자료 참조)
 - 필수항목 : Access Number(기관코드 포함), 언어, 입력일자, 자료형태, 소장기관, 저자사항, 논문제목, 학술지명(Source), 권호사항 등
 - 선택항목 : 초록(필수), 주제어, 주제분야, 기관 청구기호, 기타 데이터의 구분을 위하여 기관의 필요에 따라 표준안에 포함된 모든 사항



표준안 기반 데이터베이스 구축 (1)



- 과학기술전문정보 데이터베이스 (1991년 ~ 현재)
 - 학술지, Proceedings에 수록된 논문 단위의 서지정보 데이터베이스
 - 학위논문, 기술보고서 등 단행본 성격의 책에 대한 서지정보
 - 16개 참여기관에서 구축 ('98년 10월 현재 1,400,000건 구축)
 - KIST, KAIST, KIMM, 원자력연구소 및 포항공대 등
 - 원문복사 서비스 시행 (데이터베이스의 검색 후 원문 입수 목적)

- 과학기술 단행본 Union 데이터베이스 (1991년 ~ 현재)
 - 과학기술계 단행본, 연구보고서 등
 - 49개 기관의 약 800,000 건 데이터 구축
 - 연구단지 출연 연구소 및 충청, 전라, 서울 지역 대학교



표준안 기반 데이터베이스 구축 (2)

- 석박사 학위 논문 데이터베이스 (원문 입력)
- 인력정보 데이터베이스 - 과학재단, KORDIC
- 기자재 데이터베이스
- 연구보고서 데이터베이스 (원문 입력)
- 학회 학술지 데이터베이스 (원문 입력)
- 과학기술 Yellow Page - (과학기술 관련 인터넷 홈 페이지 모음)
 - 분류 및 기술(Description)에 관한 표준안을 준비(DPC)
 - 박사학위 논문 주제로 등장



표준(?) 포맷의 활용

- 과학기술전문정보 데이터베이스의 구축을 위하여 정해진 Format은 여러 기관에 전수되어 활용되고 있음
 - 국가 전자도서관 사업 (서지정보)
 - 공동 구축기관 (연구단지 내 15개 출연 연구소 및 KAIST 및 Union DB 참여 49개 기관)
 - 현대중공업 등 기업체 자료실
 - 건설기술연구원 등 타부처 연구소
 - 정보과학회를 비롯한 60개 학회
 - 포항공대

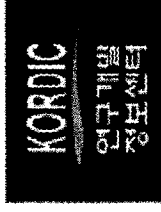


향후 계획

- 표준(안)에 대한 평가
 - 전문가 그룹
 - 실제 구축기관에서의 활용을 통한 평가
- 표준(안)의 보급
 - 지속적 홍보 및 교육 실시
 - Contents의 확보
 - Know-how의 전수
- DB구축을 위한 다양한 형태의 정보에 대한 표준화 유도
 - Full-text, Multimedia 등



기대효과



- 세계적 표준안에 근거하여 작성
 - 해외 데이터베이스의 데이터 활용이 가능
 - Cambridge Scientific Abstracts 데이터의 실제 활용
- 데이터의 상호교환 및 공동활용이 가능 (국내 기관간)
- 데이터베이스의 구축과 관련하여 축적된 Know-how의 전수가 가능
 - 자료 관리 담당자에게 친숙한 형식
 - 실제 데이터베이스 구축기관의 의견을 충분히 반영하여 작성
- 국가적 차원의 통일된 데이터 형태 유지 가능

태그별 정의

태그	태그명	항목수	크기	서브 필드	입력 예
001	AN	1	11	제어 번호	KDIC9712345
008	KD	1	6	기사 입력일자(0 - 5 col)	970102
	DT	1	1	자료 유형(6 col)	j, p, x, t, s
	DTC	1	16	자료 유형(Display Only)	
	LA	1	3	언어 코드(23 col)	KOR, ENG, JPN
010	LN		64	▼a LCCN	
020	IB		64	▼a ISBN(국제표준도서번호) ▼c price(가격)	
027	SR	2	64	▼a 표준기술레포트번호	
037	SI	2	128	▼a stock no.(서적상 번호) ▼b source(기관명) ▼c price(가격) ▼f form of issue(자료형태)	
	(주문정보)				
<u>049</u>	LO	5	64	▼a 원문 소장기관 ▼l 소장기관 등록번호	
088	RN	3	64	▼a 보고서번호	
090	CN	2	64	▼a 분류 기호 ▼b 저자 기호 ▼f 별칭 기호	
	(청구기호)				
099	SC	3	64	▼a 주제 코드	
100	PNH	1	192	▼a 개인 저자 ▼u 저자 소속 기관	
700	PNA	10	128	▼a 공동 저자 ▼u 저자 소속 기관	
110	COH	1	192	▼a 기관 저자	
710	COA	5	128	▼a 공동 기관 저자	
245	TI	1	256	▼a 기사 제목 ▼b 부제 ▼n 권책번호 ▼p 권책명	
250	ED	1	64	▼a 판차 사항	

태그	태그명	항목수	크기	서브 필드	입력 예
260	PB (출판사항)	1	128	▼a 출판지 ▼b 출판사(학위논문은 대학명) ▼c 출판일	
265	AI	1	128	▼a 구입처(출판사나 서적상의 이름, 주소)	
300	PH (형태사항)	1	128	▼a 페이지 ▼b 삽도 표시 ▼c 자료 크기	
440	SN	3	128	▼n 규격 번호	
500	GN	2	256	▼a 주기 사항	
502	DN	2	128	▼a 학위 정보(학위, 대학, 학위수여일)	
503	SH	2	128	▼a 규격 갱신 기록	
506	RA	2	64	▼a 접근 제한 정보	
510	CI	2	128	▼a 인용 자료명(DB인 경우 DB명) ▼c location within source(수록된 권호, 페이지, 연도. DB인 경우 DB access no.)	
513	RT	2	64	▼a 보고서 종류(연구, 경영, 예술, 기술, 기타) ▼b 수록 기간	
520	AB	1	2046	▼a 초록	
533	PS (복제사항)	2	128	▼a 복제 형태(microfilm, microfiche, microcard 등) ▼b 복제 장소 ▼c 복제 기관 ▼d 복제일 ▼e 형태 상세 기술 ▼f 복제 일련 번호	
536	FI (연구후원정보)	3	128	▼a 연구 후원 기관 ▼b 연구 계약 번호 ▼c 연구 기금 번호 ▼d 연구 번호	
590	SA (규격승인사항)	5	64	▼a 규격 승인 국가 ▼b 규격 승인 기관 ▼c 규격 승인 일자 ▼d 규격 종류	
599	AD	5	64	▼a 논문 지도 교수	
650	SU	20	64	▼a 주제명(자연어, 저자 keywords)	

태그	태그명	항목수	크기	서브 필드	입력 예
653	ID	20	64	▼a 식별자(identifier, 통제어)	
773	HI (원문수록처)	5	512	▼t 서명(title, 기사제목, 보고서제목, 학위논문제목) ▼a 주표목(main entry) ▼b 판차 사항(edition) ▼d 발행일자(출판지, 출판사, 출판년) ▼g 권호 또는 통권(년도, 페이지-페이지) ▼k 시리즈 사항 기술 ▼p 축약 서명(abbreviated title) ▼r 보고서 번호(report no.) ▼S 통일 서명(uniform title) ▼u 표준 보고서 번호(standard rn) ▼x ISSN ▼y 잡지 약어 코드(coden) ▼z ISBN	
810				작성자명 또는 작성자 고유번호	
820				▼a 참고 문헌수	
830				▼a 원 제목(번역했을시)	

*** ▼ : ALT 31 (ASCII code 값)