

디지털도서관에서 분산자원 검색을 위한 장서 정보에 관한 연구

A Study on Collection Informations for Discovery of Distributed Resources in Digital Libraries

이 성 숙(Sung-Sook Lee)*

목 차

1. 서론	3. 1 RSLP Collection Description Schema
2. Collection Level Description의 개념과 특징	3. 2 Dublin Core Collection Description Application Profile
2. 1 CD의 데이터 모델	3. 3 Encoded Archival Description
2. 2 CLD의 개념	3. 4 RSLP CD Schema와 DC CD AP, EAD의 특징 비교
2. 3 CLD의 역할과 이점	4. Collection Level Description 관련 사례
2. 4 CLD의 구축 단계	5. 결론 및 제언
3. Collection Level Description 관련 표준	

초 록

장서 수준에서 자원을 기술하는 것은 분산된 자원에 효과적으로 접근하는 정보 서비스의 중요한 구성요소로 인식되고 있다. 이 연구는 분산 자원을 효율적으로 탐색하고 관리하기 위한 CLD의 개념과 특징, 효과, 관련 표준을 살펴봄으로써 장서정보 연구에 대한 기초를 제공하기 위한 것이다. 또한 외국에서 수행된 CLD 관련 프로젝트를 검토함으로써 주제 접근을 강화하기 위한 우리나라 디지털도서관의 새로운 방향을 제시하기 위한 것이다.

ABSTRACT

The description of resources at collection-level is being recognised as an important component of information services that seek to provide integrated access to distributed resources. This research investigated the concept, necessities, and standards of collection level description which manages heterogeneous and distributed resources effectively. Also this research reviewed collection level description projects in other countries to show a new direction of subject access in digital libraries.

키워드: 장서수준기술, 장서수준메타데이터, 상호운용성, 정보검색, 메타데이터

Collection Level Description, Collection Level Metadata, Collection Management, Information Retrieval, Metadata, Interoperability

* 충남대학교 사회과학대학 문헌정보학과 강사(inflee@hanmail.net)
논문접수일자 2005년 5월 15일
게재확정일자 2005년 6월 10일

1. 서론

정보사회는 유비쿼터스라는 화두속에서 새로운 세상을 꿈꾸고 있다. 유비쿼터스 도서관이란 언제 어디서나 이용자가 원하는 정보를 정보화 기기를 통해 쉽게 접근하고 획득할 수 있으며, 유무선 통합 환경과 광대역 네트워크 인프라의 지원으로 이용자의 상황에 적절한 정보를 제공하는 디지털 도서관이다(이응봉 2003).

디지털 도서관이 제공하는 서비스는 전통적인 도서관보다 다루는 대상과 형식, 이용자 측면에서 매우 다양하다. 전통적인 도서관에서는 서비스의 범위가 제한적이었지만, 디지털 도서관은 전통적인 도서관의 고정관념과 한계를 뛰어넘고 있다. 디지털 도서관으로의 발전은 장서의 개념을 크게 변화시키고 있다. 전통적인 도서관에서 장서는 도서관의 소장 자료와 밀접하게 관련된 개념이다. 그러나 디지털도서관에서의 장서는 물리적 혹은 전자적 객체의 그룹으로 정의되며, 객체는 도서관에 의해 물리적으로 반드시 소유될 필요는 없기 때문에, 장서는 특정한 도서관에 소장된 장서에 한정되지 않고 분산자원으로 존재한다.

디지털도서관의 구축목적은 사회가 생산해내는 디지털 정보자원을 공간을 초월하여 실시간으로 유통시킴으로써 사회 구성원들의 생산성을 높이고자 하는 것이다. 디지털도서관은 전통적인 도서관과는 달리 가상공간에서 거대한 디지털 네트워크의 노드로서 존재한다(이제환 2004).

이러한 노드에는 다양한 장서가 분산되어 축적되어 있다. 예를 들어 이용자들은 도서관뿐만 아니라 박물관, 기록보존소, 미술관, 정부기관

에서 구축한 복수 장서를 이용할 수 있다. 그러나 이용자의 장서 이용 형태를 조사한 결과에 의하면(Stumm, Moir, and Lawn 2004), 많은 이용자들은 디지털도서관 장서의 깊이와 다양성을 모르고, 그들이 원하는 것을 효과적으로 검색하지 못하는 것으로 나타났다. 또한 대부분의 이용자들은 도서관에서 제공되는 목록서비스 등 다양한 종류의 서비스의 차이를 구분하지 못하고 있다. 이용자들은 개별자료에 대한 정보는 알 수 있지만, 디지털도서관이 제공하는 장서 전체에 대한 이해가 부족한 것으로 나타났다. 이것은 많은 이용자가 분산된 복수 장서를 검색하는데 어려움을 겪는다는 것을 의미한다.

이러한 상황에서 장서수준(collection level)에서 자원을 기술하는 것은 분산된 자원에 효과적으로 접근하는 정보 서비스의 중요한 구성요소로 인식되고 있다. 디지털 도서관의 확산으로 인한 상호운용성(interoperability)의 필요성이 제기되면서, 개별자료 수준으로 자원을 기술하는 것과 같이 장서수준에서도 표준에 의하여 기계가독형식(machine-readable format)으로 자원을 기술하는 장서정보(collection information)에 관한 관심이 높아지고 있다. 장서정보와 유사한 개념으로는 장서수준기술(Collection Level Description, CLD), 장서수준 메타데이터(Collection Level Metadata) 등이 있으며, 이런 개념들은 디지털도서관의 확장으로 새롭게 조명된 주제이다. 이들은 궁극적으로 웹의 정보공간을 효율적으로 검색하기 위한 방안이라고 할 수 있다.

네트워크 환경에서 이질적인 분산자원을 검색하기 위한 방안 중 하나는 간단한 검색 인터페이스 혹은 메타검색을 구현하는 것이다. 이용

자는 메타검색 인터페이스를 이용해서 목록 형식에 대한 이해 없이도 도서관의 모든 데이터베이스와 자원을 검색할 수 있다.

분산자원 검색의 문제를 해결할 수 있는 또 다른 방안은 주제를 통한 접근을 강화하기 위하여 CLD에 기반한 모델을 개발하는 것이다. 영국을 비롯한 여러 나라에서는 디지털도서관의 자원 관리와 검색을 강화하기 위한 수단으로 CLD의 중요성을 인식하고, CLD의 데이터 모델을 개발하였다. 또 그와 관련된 스키마도 제안되고 있다. 이들 나라에서는 데이터 모델과 스키마를 토대로 이미 디지털도서관의 장서를 효과적으로 이용하기 위한 다수의 프로젝트를 진행하고 있다.

이 연구는 분산 자원을 효율적으로 검색하고 관리하기 위한 CLD의 개념과 특징, 효과, 관련 표준을 살펴봄으로써 CLD 연구에 대한 기초를 제공하기 위한 것이다. 또한 외국에서 수행된 CLD 관련 프로젝트를 검토함으로써 주제 접근을 강화하기 위한 우리나라 디지털도서관의 새로운 방향을 제시하기 위한 것이다.

이 연구의 방법은 장서정보에 관한 이론적 배경을 문헌을 통하여 고찰하고, 지금까지 개발되어 온 CLD 관련 표준과 사례들을 문헌 조사와 비교분석을 통하여 수행하였다.

2. Collection Level Description의 개념과 특징

2.1 CD의 데이터 모델

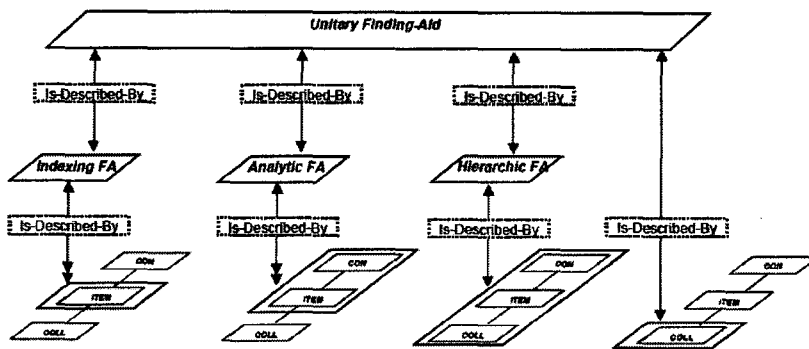
장서기술(Collection Description)의 필요성

이 제기된 이후, Heaney(2000)는 RSLP(Research Support Libraries Program) 장서 모델을 개발하였다. 이 모델은 IFLA에서 제안한 FRBR(Functional Requirements for Bibliographic Records) 접근법에 영향을 받아, 개체-관계 모델에 기반하고 있다.

Heaney의 장서 모델은 도서관과 박물관 자료를 포함해서 모든 종류의 물리적 혹은 디지털 장서에 적용될 수 있도록 설계되었다. 이 모델은 원래 RSLP 프로그램을 위한 장서기술 데이터 모델이었으나, 장서기술에 관심을 가진 모든 기관에서도 유용하게 사용할 수 있다.

이 모델에서는 장서에 대해서는 개체 'Collection'을, 장서의 정보에 대해서는 개체 'Collection-Description'을 정의하였고, 두 개체는 기술되어지는(Is-Described-By) 관계를 가진다. 'Collection-Description'은 아래와 같이 네 가지 유형으로 나눌 수 있고, 이들 간의 관계는 <그림 1>과 같다(Heaney 2000).

- ① 단일(Unitary) 탐색보조도구: 개별자료에 대한 정보는 제공하지 않고, 전체로서 장서에 관한 정보만을 제공한다.
- ② 계층적(Hierarchical) 탐색보조도구: 개별자료에 대한 정보와 장서에 관한 정보로 구성되며, 주로 기록물의 기술이 해당된다.
- ③ 분석적(Analytic) 탐색보조도구: 개별자료에 대한 정보로 구성되며, 주로 도서관 목록이 해당된다.
- ④ 색인(Indexing) 탐색보조도구: 로봇 검색엔진처럼 개별자료에서 발생된 정보로 구성된다.



〈그림 1〉 장서기술의 네가지 유형간의 관계
 (출처: <http://www.ukoln.ac.uk/metadata/rsip/model/>)

개체 'Collection-Description'은 지적으로 생산되는 자원과 로봇 검색엔진에 의해 모아진 자원을 대상으로 한다. 이 모델의 주요 목적은 위에서 정의된 장서기술의 네 가지 유형 중 장서에 대한 단일 탐색보조도와 모든 유형을 포괄할 수 있는 장서기술을 개발하는 것이다. 이 모델에서는 장서를 어떻게 구성할 것인가, 구체적으로 어떤 요소까지 사용할 것인가의 판단은 장서기술을 적용하는 기관의 몫으로 남겨두고 있다.

이 모델이 갖는 의의는, 장서기술 관련 연구와 프로젝트의 이론적 기반이 되었으며 특히 장서기술을 위한 표준으로 간주되는 RSLP CD

스키마와 DC-장서연구에 직접적인 영향을 미쳤다는 것이다.

2. 2 CLD의 개념

이 절에서의 CLD(Collection Level Description)는 장서 기술의 종류 중 단일 탐색보조도를 의미한다. CLD에서 장서(Collection)란 '물리적 혹은 전자적 개별자료(item)의 집합체'(aggregation)이다. 일반적으로 장서에 포함되는 대상은 〈표 1〉과 같다(Proffitt 2003).

CLD에서 장서의 개념은 장서의 크기에 제한을 두지 않는다. 장서는 여러 개의 개별자료

〈표 1〉 장서의 대상이 되는 객체*

구분	예
자연물	화석, 광물, 표본...
창작물	예술품, 문서, 레코드...
디지털자원	문서, 이미지, 멀티미디어 객체, 데이터, 소프트웨어...
물리적 객체의 디지털 대응물	문서, 이미지...
메타데이터	목록레코드, 개별자료 기술, 장서수준 기술...

로 구성될 수도 있고, 단지 하나의 개별자료만을 포함할 수도 있다. 또한 장서는 영구성의 정도도 다양하여, 특별한 목적을 위해 제한된 기간에만 유지될 수도 있다. 또 장서는 분산되어 존재할 수도 있다(Johnston and Robinson 2002).

CLD의 기술 대상은 장서에 포함된 개별자료가 아닌 개별자료의 전체, 즉 '장서'에 대한 정보를 기술하는 것이다. CLD는 구조화되어야 하고, 개방적이어야 하며, 기계가독의 형식을 갖추어야 하고, 표준화되어야 한다.

CLD에서 제공하는 정보는 크게 고유정보와 상황정보로 나눌 수 있다. 고유정보는 어떤 장서의 내용을 컴퓨터 분석을 통해 도출할 수 있는 정보로서, 예를 들어 수록범위, 개별자료의 유형과 수, 개별자료의 형식 등이 이에 해당한다. 상황정보는 장서제공자 혹은 장서관리자에 의해 제공되는 정보로서, 제목, 책임사항부분, 범위와 목적, 장서의 유형, 생성 혹은 갱신날짜, 갱신횟수, 장서 이용을 위한 조건, 연락처 등이 이에 해당된다(Hill 1999).

CLD의 유래는 DIALOG Bluesheets와 같은 온라인 DB에서 찾아볼 수 있다. 이러한 사용자 가이드는 사용자 도큐멘테이션(user documentation)을 위해 설계되어, 파일에 대한 정보, 생산자에 대한 정보, 포괄 범위와 갱신 주기, 검색과 인쇄 등에 관한 정보를 제공한다(Hill 1999).

CLD의 개념은 실제적으로 새로운 것은 아니다. 기록보존소에서는 자원 기술에 CLD의 관점을 오랫동안 적용해왔다. 그러나 기록보존소에서는 그들이 기술하는 자료의 집합체에 대하여 '장서'라는 용어를 사용하지 않고, 기록물군(fonds)이라는 용어를 사용하여, 개별적인 단

체, 개인에 의해 생성되고 축적된 개별기록물의 집합으로 이해하였다. 기록학의 관점에서는 오히려 기록물군의 개념이 강조되면서, 개별기록물은 간단하게 기술하는 경향이 있다(Johnston and Robinson 2002).

도서관은 장서의 범위를 정의하기 위해 전체 소장자료, 특정 주제 장서, 자료유형에 따른 장서 등 다양한 기준을 사용해왔다. 비록 도서관이 장서를 정의하고 관리하는 단위로서 인식했지만, 자원 기술의 초점은 개별자료(item)에 두었다. 도서관에서 장서 기술은 비공식적이고 비구조적이며, 도서관의 관례에 의해 이뤄진다. 한편 박물관에서도 장서의 개념이 사용되며, 도서관에서와 같이 개별적인 객체에 초점을 두고 있다. 그러나 도서관의 경우와 같이 체계적인 장서수준기술을 위한 노력은 부족했다(Macgregor 2003).

최근 위의 세 영역의 정보 관리자는 디지털 개별자료로 구성되는 장서를 어떻게 관리할 것인가의 문제에 직면해있다. 여기에는 원래 물리적인 자원을 디지털화한 결과물로서의 디지털 자원 혹은 원래 디지털 형태로 만들어진 자원이 포함된다. 웹 자원 관리자는 디지털 자원의 집합체를 기술할 필요가 있다. 이에 웹 자원의 집합을 기술하기 위한 다양한 시도가 있었지만, 아직 일반적으로 받아들여지고 있는 표준은 없다.

디지털 도서관의 확산에 따라 복수 자원을 검색하고 관리하기 위한 가이드 역할을 하는 '정보지도'(information map or landscape)와 같은 도구가 필요한 시점에서, 산발적으로 존재하던 장서기술을 포괄하는 CLD의 개념이 정립된 것은 그 의미가 있다고 할 수 있다.

2. 3 CLD의 역할과 이점

Heaney(2000)는 CLD의 역할을 정보지도와 같다고 하였다. 이용자는 정보지도를 통해 어디에 산맥, 계곡, 평지, 고원이 있는지를 확인할 수 있다. 정보 지도를 본 이용자는 원하는 정보를 얻을 가능성이 높은 지점을 찾을 것이다. 이용자는 초기 조사단계에서 먼저 나무보다는 숲을 식별하는 것에 관심을 가지므로, CLD는 정보검색 과정의 초기에 중요한 역할을 할 수 있다. 또한 CLD는 상이한 개별자료의 기술 표준을 사용하는 기관의 통합에도 유용하고, 다양한 장서 수준 스키마를 사용한 유사한 스키마간의 상호작용에도 도움이 된다.

많은 연구자들이 CLD의 잠재력에 대해 언급하였다. Hill(1999)은 ADL(Alexandria Digital Library) 프로젝트에서 CLD의 역할을 ① 검색과 접근이 가능한 클라이언트 소프트웨어에 의한 장서 식별, ② 장서가 포함하고 있는 것이 무엇인가에 대한 네트워크 검색엔진에 관한 정보 제공, ③ 장서와 디지털 도서관 인터페이스에 관한 정보를 제공하는 이용자 도큐멘테이션, ④ 장서의 저장 위치를 제공하는 것이라고 하였다.

Johnston과 Robinson(2002)은 CLD가 장기적으로 여러 영역의 자원탐색에 중요한 역할을 한다고 하였다. 자원탐색이란 지리적으로 분산되어 있는 각각의 디지털도서관들을 통합환경에서 검색하고 해당 자원에 대해 편리하게 접근하는 것을 말한다. 즉, 자원탐색은 통합검색과 자원접근이라는 두가지의 주요한 정보활동으로 구성된다(이수상 2003). 예를 들어 도서관은 특정 화가의 책을 소장하고 있고, 기록보

존소는 기록물들을 소장하고 있으며, 박물관은 그 화가의 작품을 소장하고 있을 수 있다. 이용자는 이들 기관의 장서를 모두 이용하여 자원에 접근하기를 원하고, 그것을 지원하는 기술적인 기반이 마련되기를 원한다. DC와 같은 시도는 소규모의 세트를 정의함으로써 분야마다 존재하는 기술 관습과 용어의 차이에서 오는 문제를 극복하는데 도움이 되었다. 그러나 이용자는 질의에 대한 결과로서 개별자료 수준의 많은 검색 결과를 통제하는 문제에 직면해 있다. 이런 점에서 장서수준기술은 복수 장서에 접근하는 통제된 검색에도 사용할 수 있고, 질의 초기 검색 결과에서 개별적인 검색결과 수를 줄이기 위해 사용될 수도 있다.

또한 Macgregor(2003)는 목록되지 않은 개별자료의 개관을 위해 CLD를 이용할 수 있다고 하였다. CLD의 접근법은 흔히 대규모로 분산된 이질적인 자원을 대상으로 한 항해를 가능하게 한다. CLD가 RSLP 스키마와 같이 동일한 스키마에 의해 구현될 때, 디지털도서관의 상호운용성을 증대할 수 있고, 분산 네트워크 서비스를 개선할 수 있다.

DC CD WG(2003)는 장서 관리자의 입장에서 CLD는 개별자료 메타데이터가 존재하지 않거나 혹은 이용이 가능하지 않을 때 이용자에게 장서의 내용과 이용가능성에 관한 정보를 알려줄 수 있다. 이용자는 ① 관심있는 장서를 검색하고, 접근할 수 있으며, ② 요약된 정보를 토대로 검색할 장서를 선택할 수 있고, ③ 접근과 이용의 조건을 이해하며, ④ 장서와 장서에 포함된 개별자료를 알 수 있다. 또한 ⑤ 장서의 탐색, 선정, 비교 등의 작업이 소프트웨어에 의해 수행될 수 있다.

위의 언급된 내용을 종합하면 CLD의 이점을 자원탐색과 장서관리 측면 두 가지로 나누어 생각할 수 있다. 우선 자원탐색 면에서 보면, ① 장서에 관한 정보를 표현할 수 있다. ② 목록이 작성되지 않은 개별자료의 개관을 제공한다. ③ 이용자가 요약된 기술에 기반하여 검색하기 위한 장서를 선정할 수 있도록 한다. ④ 복수장서에 관한 통제된 탐색을 지원한다. ⑤ 장서에 대한 폭넓은 접근이 가능하다. ⑥ 기계가독형 장서 기술로 소프트웨어를 활용할 수 있다.

장서관리 면에서 보면, ① 장서 제공자에게 장서관리를 가능하게 한다. ② 내부적으로 소장자료의 통제/심사/검토를 가능하게 한다. ③ 장서의 취약점을 식별할 수 있다. ④ 개별자료 수준 목록작성을 위한 우선권을 부여할 수 있다. ⑤ 다른 제공기관과의 협력을 가능하게 한다. ⑥ 장서의 장점과 단점에 관한 정보를 확인하고 기록하고 공유할 수 있다. ⑦ 협력이 가능한 부분을 제시할 수 있다.

2. 4 CLD의 구축 단계

CLD 구축 단계는 크게 장서수준기술 프로그램을 계획하는 단계와 개별적인 장서수준기술을 작성하는 단계로 나눌 수 있다. 첫 번째 장서수준기술 프로그램을 계획하는 단계에서는 표준과 이용자 요구(user requirements)에 부합되도록 포함시킬 장서를 식별하는 것과 스키마를 결정하는 것, DB 플랫폼을 선택하는 것이 포함된다. 이용자 요구에는 최종 이용자와 장서관리자의 요구를 포함해야 한다.

장서수준기술 프로그램 단계에서는 구체적으로 ① 기술할 장서를 식별해야 하고 ② 포함

할 데이터를 식별해야 하고 ③ 다양한 이용자들을 고려한 디스플레이도 고려해야 하고 ④ 사용할 스키마를 선택해야 한다. 또한 ⑤ 표준을 고려해야 하고 ⑥ 재사용 문제도 고려해야 하고 ⑦ 테스트 단계를 포함해야 한다(Keane 2003). 어떤 객체 그룹을 장서로 취급할 것인가의 결정은 도서관의 목표와 환경, 도서관 정책에 따른다(Hill 1999).

두 번째 단계는 개별적인 장서수준기술을 생성하는 것이다. 장서수준기술은 대개 장서에 관한 다양한 세목을 다루기 위한 구조화된 포맷을 사용한다. CLD의 대표적인 스키마로는 RSLP Collection Description Schema, DC CD AP (Dublin Core Collection Description Application Profile), EAD(Encoded Archival Description) 등이 있다.

3. Collection Level Description 관련 표준

3. 1 RSLP Collection Description Schema

CLD의 대표적인 스키마로는 RSLP CD 스키마를 들 수 있다. RSLP CD 스키마의 개발 동기는 다음과 같다. ① 이용자는 관심있는 장서를 검색하고 접근할 수 있도록 한다. ② 이용자는 일련의 통제된 방법으로 복수 장서를 한꺼번에 검색할 수 있도록 한다. ③ 소프트웨어가 이용자를 위해 그런 업무를 수행하도록 한다(Powell, Heaney, and Dempsey, 2000).

RSLP CD 스키마 개발은 UKOLN(UK Office for Library Networking)과 Michael

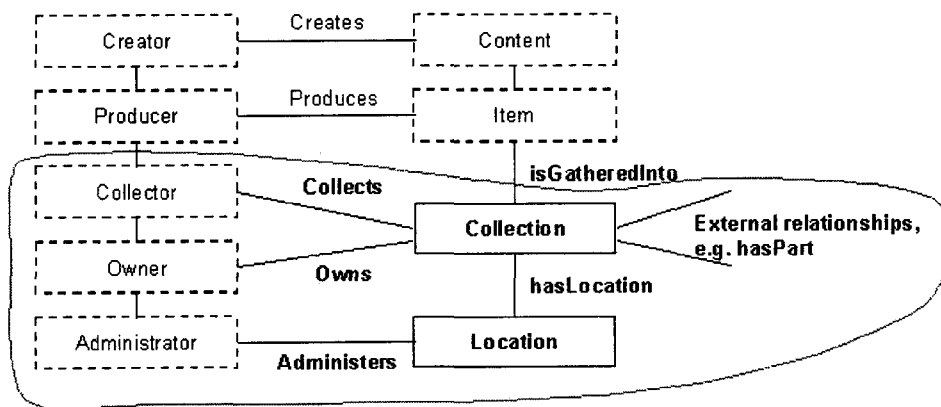
Heaney의 협력과 RSLP의 재정적 후원에 의해 수행되었다. RSLP 스키마는 대체로 장서에 대한 도서관과 기록보존소의 견해에 기초를 두고 있으며, Heaney의 모델에 기반하여 메타데이터 스키마를 개발하였다. RSLP 스키마는 디지털 자원뿐만 아니라 모든 유형의 자료를 다루기 위한 것이다(Powell, Heaney, and Dempsey, 2000).

RSLP 스키마가 가지는 의의는, 첫째 장서 데이터모델에 기반한 스키마라는 것이다. 둘째 특정 장서만을 대상으로 하는 것이 아니라 범용적으로 적용될 수 있는 모델로서 개발된 것이다. 셋째 RSLP CD 스키마가 RDF를 사용해서 구현되었다는 것이다(Xin 2004). RDF는 메타데이터를 교환하기 위한 하부구조로서, 데이터의 의미와 구문, 구조의 통일을 통해 메타데이터의 상호운용성을 확보하기 위한 것이다(김태수 1999). RSLP CD 스키마는 간단한 장서수준기술의 생성에 유용하며, 다양한 장서의 유형을 포괄할 수

있고, 다양한 상황에서 사용가능하다는 장점이 있다(Johnston and Robinson 2002). RSLP CD 스키마의 개발은 DC CD AP의 개발에 촉진제 역할을 하였다.

RSLP 스키마는 <그림 2>와 같이 개체 'Collection', 'Location'과 3개의 관련 에이전트들이 관계를 통해 연결되며, 각 개체는 속성을 갖는다. 개체 'Collection'은 물리적 혹은 전자적 개별자료의 집합이며, 개체 'Location'은 장서가 소장된 물리적 혹은 온라인 디지털 장소이다.

RSLP CD 스키마의 속성은 <표 2>와 같고, RSLP CD 스키마에는 장서에 관한 속성, 위치에 관한 속성, 에이전트에 관한 속성, 다른 개체 간의 관계, 복수 장서 간의 외부 관계가 포함된다(Powell, Heaney, and Dempsey 2000).



<그림 2> RSLP CD 스키마 데이터 모델

(출처: <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/september00/powell/09powell.html>)

〈표 2〉 RSLP CD 스키마의 속성

구분	속성	정의
Collection	Title	장서에 부여한 제목
	Description	장서의 내용에 대한 개요
	Type	장서의 유형
	Strength	장서의 특징
	Accumulation date range	장서가 축적된 날짜 범위
	Contents date range	장서내 개별자료의 날짜 범위
	Custodial history	장서의 소유권과 관리자 변경 내력
	Note	기타 정보
	Physical characteristics	장서의 물리적 표현형식 및 디지털 표현형식
	Language(s)	장서내 개별자료의 내용을 기술하고 있는 언어
	Concept(s)	장서와 관련된 키워드
	Object(s)	장서내 개별자료와 관련된 객체
	Name(s)	장서와 관련있는 개인/단체
	Place(s)	장서의 공간적 범위
	Time(s)	장서의 시간적 범위
	Sub-collection(s)	현 장서내에 포함되어 있는 하위장서
	Super-collection(s)	현 장서를 포함하고 있는 슈퍼장서
	Catalogue or description	현 장서를 기술하는 다른 장서
	Described collection	현 장서에 의해 기술된 다른 장서
	Associated collection	현 장서의 출처가 되는 다른 장서의 식별기호/제목
	Associated publication	장서의 이용, 연구, 분석에 관한 출판물의 식별기호/제목
	Access control	이용자 제한, 요금과 같은 장서에 대한 접근제한
	Location	장서의 물리적 및 온라인 위치
	Identifier	장서를 식별하기 위한 식별기호
	Accrual status	보관 내력
	Legal status	장서의 법적 상태
	Creator	장서내 개별자료를 모은 에이전트
Owner	장서의 법적 소유권을 가진 에이전트	
Location	Name	위치의 이름
	Identifier	위치의 식별기호
	Access conditions	이용가능시간, 이용자 등에 관한 접근조건
	Held collection	물리적 혹은 온라인 위치에 포함된 장서의 식별기호
	See also	장서 위치의 홈페이지(URL)와 같이 위치에 관한 보다 많은 정보를 제공하는 자료의 식별기호
	Administrator	장서 위치에 관한 물리적인 혹은 전자적인 환경에 책임을 지고 있는 에이전트의 식별기호
	Postal address	장서가 물리적으로 위치해 있는 곳의 우편주소
	Post/zip code	장서가 물리적으로 위치해 있는 곳의 우편번호
	Country	장서가 물리적으로 위치해 있는 국가
	Locator	온라인 디지털 장서의 온라인 위치(URL)
Agent	Name	에이전트의 이름
	Identifier	에이전트를 위한 공식적인 식별기호
	Organisation name	에이전트와 관련있는 조직 이름
	Role	에이전트가 수행하는 역할
	Telephone number	에이전트의 전화 번호
	Fax number	에이전트의 팩스 번호
	Email address	에이전트의 전자 메일 주소
	Agent history	에이전트에 관한 관리 내력

3. 2 Dublin Core Collection Description Application Profile

2003년 DCMI(Dublin Core Metadata Initiative)는 장서정보를 위해 사용할 수 있는 DC CD AP를 제안하였다. DC CD AP는 그 동안 장서수준기술에 대한 선도적인 역할을 했던 RSLP, CD Focus(UKOLN), JISC IESR 등의 경험을 바탕으로 한 것이다.

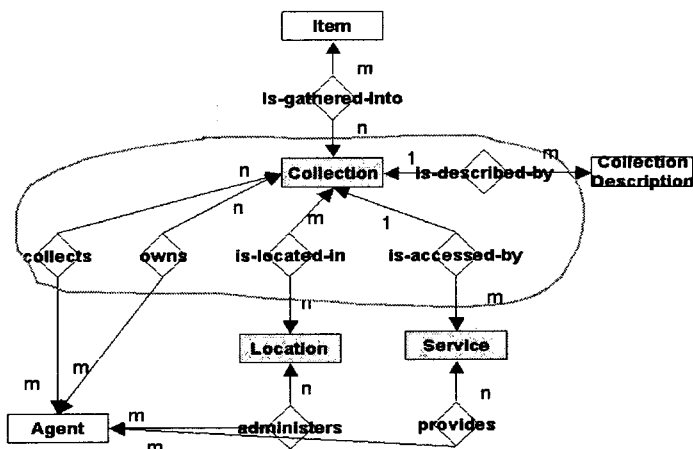
DC CD AP는 RSLP 장서기술 스키마에 기반을 두고 있으며, DC CD AP의 데이터 모델은 Heaney(2000)의 장서 데이터모델을 토대로 하고 <그림 3>과 같이 개체 'Service'를 추가하였다(DC CD WG 2003).

DC CD AP는 DC 메타데이터 요소가 장서 수준기술에서 어떻게 사용될 수 있는지에 대한 가이드라인을 제시하며, 장서의 간단한 CLD 생성 방법을 제공하기 위한 것이다. DC CD AP는 공통된 접근점을 이용한 장서의 탐색, 이

미 알고 있는 장서의 식별, 검색된 다수의 장서에서 적절한 장서의 선택, 장서의 위치 확인, 장서에 접근할 수 있는 서비스 확인 등의 기능요건을 충족시킨다. DC CD AP의 속성은 <표 3>과 같다(DC CD WG 2003).

3. 3 Encoded Archival Description

영국, 미국, 호주의 몇몇 프로젝트와 기관은 '기록물' 장서의 기술을 위한 표준으로서 EAD(Encoded Archival Description)를 사용해왔다(Stumm, Moir, and Lawn 2004). EAD는 1993년 탐색보조도구 전산화의 표준을 마련하기 위하여 미교육부의 제안에 의해 버클리대학에서 수행한 Berkeley Finding Aids Project에서 시작된 것이다. EAD 실무그룹은 1998년 EAD DTD 1.0 버전을 발표하였고, 그 후 SGML/XML 스펙을 만족시키며 다양한 소프트웨어를 이용하여 테스트된 EAD DTD 2002



<그림 3> DC CD의 데이터 모델

(출처: <http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>)

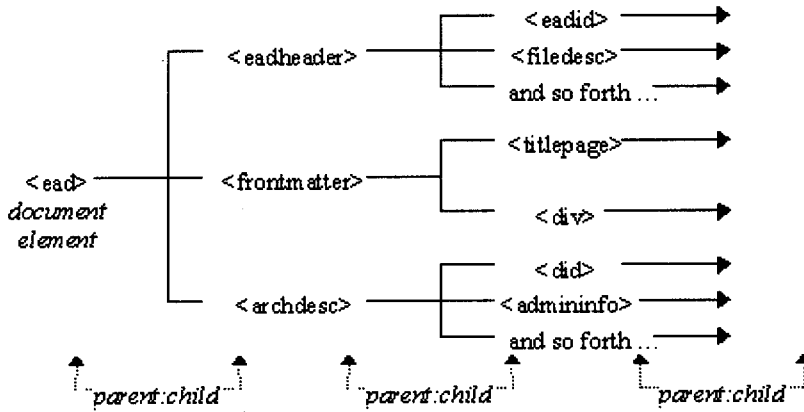
〈표 3〉 DC CD AP의 속성

속성	정의
Collection Identifier	장서를 식별하기 위한 식별기호
Title	장서에 부여한 제목
Alternative Title	공식적인 제목이외의 제목
Description	장서의 내용에 대한 개요
Physical Characteristics	장서의 물리적 표현형식 및 디지털 표현형식
Size	장서의 크기
Language	장서내 개별자료의 내용을 기술하고 있는 언어
Type	장서의 유형
Rights	장서에 대한 권리정보
Access Rights	장서 접근제한
Accrual Method	장서에 개별자료가 누적되는 방법
Accrual Periodicity	장서에 개별자료가 누적되는 주기
Accrual Policy	장서에 개별자료가 누적되는 정책
Custodial History	장서의 소유권과 관리자 변경 내력
Audience	장서에 대한 이용자 범주
Logo	장서의 로고
Subject	장서내 개별자료와 관련된 주제
Spatial Coverage	장서내 개별자료의 내용에서 다루어진 공간적 범위
Temporal Coverage	장서내 개별자료의 내용에서 다루어진 시간적 범위
Accumulation Date Range	장서가 축적된 날짜의 범위
Contents Date Range	장서내 개별자료의 날짜 범위
Collector	장서내 개별자료를 모은 에이전트
Owner	장서의 법적 소유권을 가진 에이전트
Is Located At	장서의 위치
Is Accessed Via	장서에 대한 접근을 증개하는 서비스
Associated Publication	장서의 이용, 연구, 분석에 관한 출판물
Super-Collection	현 장서를 포함하고 있는 수퍼장서
Sub-Collection	현 장서내에 포함되어 있는 하위장서
Catalogue or Description	현 장서를 기술하는 다른 장서
Associated Collection	현 장서와 관련있는 다른 장서

버전을 발표하였다. EAD DTD는 전자검색도구에서 포함해야 하는 데이터와 데이터간의 상관관계를 집약한 정의집이다. EAD의 목표는 기록보존서, 도서관, 대학, 연구소 및 역사적 자료를 보유하고 있는 단체에 의해 생산되고 있는 검색 자료를 전자기록형태로 부호화하는 것이며, 특히 고문서기술을 위해 고도로 구조화된 기술요소를 제공한다.

이 연구에서는 EAD DTD 2002 버전에 관

한 실무지침서(Application Guideline)가 공개되지 않아 확인할 수 없었기 때문에 EAD DTD 1.0 버전을 중심으로 검토하였다. 〈그림 4〉와 같이 <ead>는 <eadheader>, <frontmatter>, <archdesc>의 세가지 영역으로 구분되며, 총 145개의 태그로 구성된다. <eadheader>와 <frontmatter>는 검색도구에 대한 정보를 기술하는 영역이며, <archdesc>는 기록물 자체에 대한 서지, 행정 및 접근 정보 등을 기술하도록 하고



〈그림 4〉 EAD DTD 요소의 전개

(출처: <http://www.loc.gov/ead/ag/agconc.html>)

있다(김현희 2001).

EAD DTD의 요소 중 주요 요소만을 중심으로 살펴보면 〈표 4〉와 같다. 같은 수준의 요소들은 들여쓰기로 구분하여 상하위 관계를 표시하며, 들여쓰기의 값이 같은 경우 해당 요소들은 동일한 수준이라는 것을 의미한다. 〈eadheader〉는 필수영역이고, 〈frontmatter〉는 선택 영역이다. 〈eadheader〉에는 고유번호를 할당하는 〈eadid〉, 검색도구의 저자, 표제 등을 기술하는 〈filedesc〉, 검색 도구의 언어와 코드(coder)에 대한 정보를 기술하는 〈profiledesc〉, 검색도구에 대한 개정 정보를 수록하는 〈revisiondesc〉로 구성된다. 〈frontmatter〉는 〈eadheader〉에 기술하지 못한 정보를 기술하기 위한 것이며, 표제에 관한 부차적인 정보를 기술하는 〈titlepage〉와 기타 정보를 기술하는 다용도 태그인 〈div〉로 구성된다.

〈archdesc〉의 하위요소 중 〈did〉 태그는 기록물의 기본 정보를 기술하는 태그로 구성된다. 〈did〉의 하위요소인 〈physdesc〉는 형태사항을

기술하며, 〈physloc〉는 기록물이 소장된 구체적인 장소를 기술한다. 〈dao〉와 〈daogrp〉는 기록물의 디지털 자료를 기술하기 위해서 만든 태그로 생산자의 사진, 기록물 텍스트의 디지털 및 이미지 자료 등을 기술하는데 사용된다. 〈admininfo〉는 행정 정보를, 〈bioghist〉 태그는 전기 또는 역사에 대한 정보를 기술한다. 〈controlaccess〉는 통제어휘로 된 색인어를 기술하는 하위태그들로 구성된다. 〈scopecontent〉와 〈organization〉 태그는 범위와 내용, 기록물의 조직을 각각 기술한다. 〈arrangement〉는 기록물의 배열 방법을 기술하는 하위태그들로 구성된다. 〈add〉 태그의 하위 태그인 〈bibliography〉는 기록물을 기반으로 하여 만든 단행본, 논문 및 TV 프로그램 등에 대한 정보를 기술하는 태그이며, 〈relatedmaterial〉은 관련 자료에 대한 정보를 기술하는 태그이다. 마지막으로 〈dsc〉는 계층적으로 구조화된 기록물을 기술할 수 있는 태그이다(김현희 2001, 63-66; 강소연 2003).

〈표 4〉 EAD의 주요 요소

〈eadheader〉	〈frontmatter〉	〈archdesc〉
〈eadid〉	〈titlepage〉	〈did〉
〈filedesc〉	〈titleproper〉	〈head〉
〈titlestmt〉	〈author〉	〈abstract〉
〈titleproper〉	〈div〉	〈physdesc〉
〈author〉	〈head〉	〈repository〉
〈sponsor〉	〈p〉	〈origination〉
〈publicationstmt〉		〈unidata〉
〈publisher〉		〈unitid〉
〈date〉		〈unittitle〉
〈profiledesc〉		〈container〉
〈creation〉		〈physloc〉
〈language〉		〈dao〉
〈revisiondesc〉		〈daogrp〉
		〈admininfo〉
		〈accessrestrict〉
		〈acqinfo〉
		〈bioghist〉
		〈controlaccess〉
		〈persname〉
		〈corpname〉
		〈famname〉
		〈geogname〉
		〈occupation〉
		〈function〉
		〈genreform〉
		〈subject〉
		〈title〉
		〈scopecontent〉
		〈organization〉
		〈arrangement〉
		〈add〉
		〈bibliography〉
		〈relatedmaterial〉
		〈dsc〉
		〈c〉
		〈c01〉
		〈did〉
		〈unittitle〉
		〈container〉
		〈add〉...
		〈c02〉....〈c12〉

3. 4 RSLP CD Schema와 DC CD AP, EAD의 특징 비교

이 절은 RSLP CD 스키마와 다른 장서수준 기술 표준이 어느 정도 부합되지를 중심으로 그 특징을 살펴보기 위한 것이다. RSLP CD 스키마와 DC CD AP, EAD간의 주요 요소 매핑

결과는 <표 5>와 같다. <표 5>는 UKOLN(2004)에서 공개한 RSLP CD 스키마와 DC CD AP의 매핑 테이블과 Johnston(2001)의 EAD 1.0과 RSLP CD 스키마 매핑 테이블을 참고로 주요 항목을 중심으로 작성되었다. <표 5>는 RSLP CD 스키마를 중심으로 왼편에는 DC CD AP를, 오른편에는 EAD를 비교하였다.

<표 5> RSLP CD 스키마와 DC CD AP, EAD 비교

DC CD AP		RSLP CD Schema		EAD
Property label	Property name	Property label	Property name	Tag
Collection				
Identifier	dc:identifier	Identifier	dc:identifier	archdesc/did/unitid/@countrycode archdesc/did/unitid/@repositorycode archdesc/did/unitid eadheader/eadid
Title	dc:title	Title	dc:title	archdesc/did/unittitle eadheader/filedesc/titlestmt/titleproper
Description	dcterms:abstract	Description	dc:description	archdesc/scopecontent
Physical Characteristics	dc:format	Physical Characteristics	dc:format	archdesc/did/physdesc/extent archdesc/did/physdesc/physfacet
Language	dc:language	Language	dc:language	archdesc/@langmaterial eadheader/profiledesc/language/language
Access Rights	dcterms:accessRights	Access Control	rsip:accessControl	archdesc/admininfo/accessrestrict archdesc/admininfo/userrestrict
Accrual Status	cid:accrualStatus	Accrual Status	rsip:accrualStatus	archdesc/admininfo/accruals
Custodial History	gen:provenance	Custodial History	rsip:custodialHistory	archdesc/admininfo/custodhist archdesc/admininfo/acqinfo archdesc/admininfo/appraisal
Subject	dc:subject	Concept	dc:subject	archdesc/controlaccess/subject
Spatial Coverage	dcterms:spatial	Place	dcterms:spatial	archdesc/controlaccess/geogname
Temporal Coverage	dcterms:temporal	Time	dcterms:temporal	archdesc/did/unitdate
Accumulation Date Range	cid:accumulationDateRange	Accumulation Date Range	rsip:accumulationDateRange	archdesc/admininfo/custodhist/date[@type="accumulation"]
Contents Date Range	cid:contentsDateRange	Contents Date Range	rsip:contentsDateRange	archdesc/did/unitdate eadheader/revisiiondesc/list/item/date
Type	dc:type	Type	dc:type	"collection.archive.text" "collection.archive.image" etc "finding-aid.archive"
Catalogue or Description	dc:description	Catalogue or Description	rsip:hasDescription	archdesc/add/otherfindaid
Associated Collection	dc:description	Associated Collection	rsip:hasAssociation	archdesc/add/relatedmaterial archdesc/add/separatedmaterial
Associated Publication	dcterms:isReferencedBy	Associated Publication	rsip:hasPublication	archdesc/add/bibliography

Collector	dc:creator	Collector	dc:creator	
Owner	marcrel:own	Owner	rsp:owner	
Is Available At	gen:isAvailableAt	Location	rsp:hasLocation	
Sub-collection	dcterms:hasPart	Sub-collection	dcterms:hasPart	
Super-collection	dcterms:isPartOf	Super-collection	dcterms:isPartOf	
		Legal Status	rsp:legalStatus	archdesc/@legalstatus archdesc/@otherlegalstatus
		Note	rsp:note	archdesc/odd eadheader/filedesc/notestmt/note
		Name	rsp:agentName	archdesc/did/origination archdesc/controlaccess/persname archdesc/controlaccess/corpname
		Strength	rsp:strength	
		Described Collection	rsp:isDescriptionOf	
		Object	rsp:objectName	
Alternative Title	dcterms:alternative			
Audience	dcterms:audience			
Size	dcterms:extent			
Rights	dc:rights			
Logo	cid:logo			
				archdesc/@level archdesc/@otherlevel
				archdesc/arrangement
				archdesc/add[@encodinganalog="isadg351"] archdesc/odd[@encodinganalog="isadg351"]
				archdesc/admininfo/altformavail admininfo/altformavail
Location				
		Postal Address	rsp:address	archdesc/did/repository/address did/repository/address
		Identifier	dc:identifier	
		Name	dc:title	
		Access Conditions	rsp:accessConditions	
		See also	rsp:seeAlso	
		Post/zip Code	rsp:postcode	
		Country	rsp:country	
		Locator	rsp:locator	
		Administrator	rsp:administrator	
		Held Collection	rsp:isLocationOf	
Agent				
		Name	vcard:fn	archdesc/did/origination eadheader/filedesc/titlestmt/author eadheader/filedesc/publicationstmt/publisher
		Organisation Name	vcard:org	archdesc/did/origination archdesc/did/repository
		Agent History	rsp:agentHistory	archdesc/bioghist
		Identifier	dc:identifier	
		Role	vcard:role	
		Telephone Number	vcard:voice	
		Fax Number	vcard:fax	
		Email Address	vcard:email	

세 개의 장서기술 표준이 비교적 정확히 부합되는 요소는 'Collection'에서는 RSLP CD 스키마의 Identifier, Title, Description, Physical Characteristics, Language, Access Control, Accrual Status, Custodial History, Concept, Place, Time, Accumulation Date Range, Contents Date Range, Type, Catalogue or Description, Associated Collection, Associated Publication 등이다.

RSLP와 DC CD AP만이 매핑되는 속성으로는 Collector, Owner, Location, Sub-collection, Super-collection 등이며, RSLP와 EAD만이 매핑되는 속성으로는 개체 'Collection'의 Legal Status, Note, Name, 개체 'Location'의 Postal Address, 개체 'Agent'의 Name, Organisation Name, Agent History 등이다.

다른 2개의 표준 요소와 매핑되지 않은 속성 중 일부를 살펴보면, RSLP DC 스키마는 'Collection'의 Strength, Described Collection, Object 등이고, DC CD AP는 Alternative Title, Audience, Size, Rights, Logo 등이며, EAD는 archdesc/@level, archdesc/arrangement, archdesc/admininfo/altformavail, admininfo/altformavail 등이다.

세 스키마 간에 매핑되는 요소에 차이가 발생하는 것은 세 스키마 모두 장서를 기술한다는 공통점은 있지만, 그 목적과 대상 유형이 다르기 때문이다. 우선 제공하는 속성의 수에서도 EAD가 가장 광범위하고, 그 다음으로 RSLP CD 스키마, DC CD AP 순이다. 그래서 RSLP CD 스키마의 한 속성에 EAD의 여러 속성이 동시에 매핑되는 사례를 쉽게 발견할 수 있다.

예를 들어 RSLP CD 스키마의 Custodial History에는 EAD의 archdesc/admininfo/custodhist, archdesc/admininfo/acqinfo, archdesc/admininfo/appraisal이 매핑된다.

또한 이들 세 표준간의 기술 대상의 차이를 들 수 있다. RSLP와 DC CD AP의 기술 대상은 장서 수준이지만, EAD는 장서 수준과 개별 자료 수준 모두를 포괄하기 위해 탐색가능한 모든 부분을 마크업하며 상세수준으로 기술하기 때문에 상대적으로 각 요소가 망라적이며 일반적이다. 더욱이 기술하는 대상의 유형에서도 차이가 난다. RSLP는 일차적으로 기관에서 소장한 인쇄형태의 장서를 우선적으로 고려하면서 출발한 후 디지털 장서로 확대하였고, EAD 역시 인쇄형태의 탐색보조도구를 전자형태로 제공하기 위한 것이다. 반면 DC CD AP는 DC가 장서 기술에 어떻게 적용되는가를 검토한 것으로 네트워크 장서를 중심으로 한다. 데이터모델링에 기반한 스키마의 차이를 보면, RSLP CD 스키마와 DC CD AP는 Heaney의 장서 모델에 기반한 것이고, EAD는 필요에 의해 요소들을 추가한 것으로 어떠한 모델에도 기반하지 않는다.

결국 RSLP CD 스키마와 DC CD AP, EAD는 장서수준으로 기술한다는 것은 동일하지만, 각각의 목적, 대상 유형, 포괄하는 기술 대상이 조금씩 다르다. RSLP CD 스키마는 EAD와 같은 기존의 장서 기술 스키마를 대체하기 위한 것은 아니다. 오히려 DC(Dublin Core) 메타데이터와 같이, 장서 수준에서의 간단한 속성 세트를 제공하기 위한 것이다(Powell, Heaney, and Dempsey, 2000; Johnston 2003). 특히 RSLP는 다른 메타데이터 스키마 특히 DC 메

타데이터 요소 세트와 호환된다.

RSLP CD 스키마가 장서의 관리와 검색을 위해 설계되었다면, DC CD AP는 주로 장서의 탐색과 선정에 초점을 두고 설계되었기 때문에 주로 RSLP의 개체 장서의 속성과만 매핑된다. 물론 DC CD AP는 장서 관리와 같은 다른 기능도 지원할 수 있지만, 중점을 두는 부분은 장서의 탐색과 선정에 관한 부분이다. 그리고 DC CD AP는 장서의 모든 유형의 특징을 전부 기술할 의도를 가진 것은 아니다. RSLP와 비교해서 DC CD AP는 RSLP보다 더 간단한 스키마라고 할 수 있다. 이러한 각 스키마간의 차이는 향후 장서수준기술을 채택하는데 기초자료로 활용될 수 있을 것이다.

4. Collection Level Description 관련 사례

1) eLib clumps projects

1998년 영국은 영국디지털도서관 프로젝트인 eLib의 2단계를 마무리하면서, eLib 3단계 사업의 일환으로 clumps 프로젝트를 진행하였다. 이 프로젝트에서 'clumps'는 목록의 집합체로 정의된다. clumps의 목적은 Z39.50의 효율적인 이용과 목록 집합을 지역적으로, 주제적으로 관리하고 검색할 수 있는 방안을 제시하기 위한 것이다.

clumps가 진행한 하위 프로젝트 중 CAIRNS (<http:// Cairns.lib.strath.ac.uk>), M25 (<http:// www.m25lib.ac.uk>), RIDING (<http:// www.riding.ac.uk>)는 지역적으로 종합목록을 어떻게 활용할 것인가에 대한 것이고, 영국 음악

도서관들이 컨소시엄을 구성하고 진행한 프로젝트인 MLO(Music Libraries Online)는 주제적으로 종합목록을 어떻게 활용할 것인가에 대한 것이다.

clumps가 진행한 이들 프로젝트들은 여러 도서관의 목록을 통합 검색할 수 있는 단일 인터페이스를 제공하는 시스템을 구현하였다. 구현된 인터페이스의 특징은 검색할 수 있는 목록을 선택할 수 있고, 목록에 의해 기술되는 장서에 관한 정보를 제공한다는 것이다(Dunsire and Macgregor 2003).

이 프로젝트가 갖는 의의는 Z39.50에 관한 중요한 기술적 진보와 함께 장서수준기술과 접근 정책과 같은 개념들에 대한 논의가 이루어졌다는 것이다. 또한 목록이 주제적으로 혹은 지역적으로 통합 관리될 수 있는 전형을 보여줬다는 것에 있다.

2) RSLP projects/ UKOLN

2000년부터 2002년까지 RSLP는 진행하는 다수의 프로젝트 중 CLD에 관한 몇몇 프로젝트를 지원하였다. 특히 장서정보 검색서비스를 위한 세가지 프로젝트가 진행되었다. Wales 관련 연구 장서를 위한 Wales 프로젝트(<http:// www.mappingwales.ac.uk>), 인문과학과 사회과학 연구자를 위한 연구와 특수 장서 프로젝트 (<http:// www.rascal.ac.uk>), 스코틀랜드 도서관과 박물관, 기록물이 소장하고 있는 장서에 대한 통합검색을 지원하는 SCONE(<http:// www.scone.strath.ac.uk>)이다. 이 중에서 SCONE은 clump와 연결되어 진행되었다. SCONE은 스코틀랜드 지역에 분산된 약 2,500개의 장서를 대상으로 국가 자원을 통합 구축하고 유지하

기 위한 효율적인 모델을 제시하기 위한 것이다. 장서는 제목, 주제, 위치, 관련 개인과 단체 등에 의해 검색할 수 있다(Dunsire and Macgregor 2003).

또한 RSLP는 자신들의 최상위 목표인 자원 탐색과 접근을 강화하기 위해 장서에 관한 일관된 접근이 중요하다고 판단하고, 장서의 데이터 모델과 그런 이론에 기반한 메타데이터 스키마를 개발하는 프로젝트를 UKOLN의 협력하에 진행하였다. 프로젝트의 결과물로 앞 절에서 살펴본 Heaney의 장서 데이터모델과 RSLP CD 스키마가 발표되었다.

RSLP 스키마가 공개되면서, 이에 대한 지속적인 관리와 배포를 위해 UKOLN 내에 CD

Focus(Collection Description Focus)를 설치하였다. CD Focus는 JISC, RSLP, British Library 등의 협력 지원하에 국가적인 사업으로 진행되었다. CD Focus의 목표는 간단하고 일관된 접근을 보장하기 위하여, 장서의 기술 방법, 스키마, 도구에 대한 상호협력을 증진하기 위한 것이다. CD Focus의 성과는 장서 수준 관련 세미나의 정기적 개최, CLD online tutorial (<http://www.ukoln.ac.uk/cd-focus/cdfocus-tutorial/>)의 개발, RSLP CD 스키마의 배포 사업 등으로 요약할 수 있다(Chapman and Robinson 2004).

RSLP CD 스키마를 사용한 주요 프로젝트의 사례는 <표 6>과 같다(Keane 2003).

<표 6> RSLP CD 스키마 적용 사례

서비스명	특징	URL
Backstage	- 영국의 공연예술장서 - 장서수준 기술은 RSLP CD 스키마 기반, 개별자료수준 기술은 ISAD(G)에 기반	http://www.backstage.ac.uk/backstage/
Cecilia	- 음악 장서	http://www.cecilia-uk.org/
Collections Wales	- Wales의 연구 장서 - 광범위한 주제, 다양한 포맷, 인쇄 자료부터 사진, 문화유물, 전자적 데이터까지 포함	http://www.mappingwales.ac.uk/
Cornucopia	- 6,000개 이상의 전 분야 장서	http://www.cornucopia.org.uk/
Crossroads	- North staffordshire의 도자기 산업 관련 장서	http://www.crossroads-wm.org.uk/html/
Enrich UK	- New Opportunities Fund NOF-digi programme에 의해 지원된 150개 장서 - 내용 범위는 평생학습을 위한 문화, 역사, 사회와 경제발전, 과학 예술 등	http://www.enrichuk.net/
MASC25	- M25 컨소시엄내 도서관이 소장한 인쇄형태의 희귀본 장서	http://www.ucl.ac.uk/ls/masc25/
RASCAL	- 북아일랜드 연구 자원	http://www.rascal.ac.uk/
Revealweb	- 시각장애인을 위한 서비스와 자원에 관한 장서	http://www.revealweb.org.uk/
SCONE	- 스코틀랜드 지역의 분산된 약 2,500개의 장서	http://scone.strath.ac.uk/
Tap into Bath	- Bath 지역의 문화 유산과 학술 장서	http://www.bath.ac.uk/library/tapintobath/

위의 사례에서 보면, CLD는 동일한 기관들의 장서를 통합하기 위해 또는 상이한 기관들의 이질적인 장서를 통합하기 위해 사용되기도 하였다. 구축된 장서가 학술활동을 위한 전 분야를 대상으로 하기도 하고, 음악, 평생학습, 도자기와 같이 특정 주제를 대상으로 하는 경우도 있다. 또한 시각장애자와 같은 특정이용자를 대상으로 하는 경우도 있다. 참여기관에 있어서도 도서관, 박물관, 기록보존소, 장서의 주제와 관련된 단체와 조직 등 다양했다. 제공되는 서비스면에서도 RSLP CD 스키마에서 제공하는 접근점을 통한 장서의 검색과 브라우징이 가능하며, 적절한 장서를 검색한 후 개별자료 수준의 검색을 위한 링크를 제공한다. RSLP CD 스키마는 초기에는 RSLP 프로그램내 도서관, 기록보존소, 박물관의 물리적인 장서를 위해 사용되었고, 최근 Enrich UK 내 New Opportunities Fund-Digitise programmes에서 생성된 분산 디지털 자원을 위해서도 사용되었다(Johnston 2003).

3) Archives hub

Archives hub는 영국내 모든 대학이 소장하고 있는 기록물에 접근하기 위한 게이트웨이이다. Archives hub는 JISC(Joint Information Systems Committee)의 후원 하에 영국 Liverpool 대학팀에 의해 개발된 것으로 CURL(Manchester Consortium of University Research Libraries)의 MIMAS(Manchester Information & Associated Services)에서 운영하고 있다.

Archives Hub는 그동안 보존에만 관심을 가졌던 기록물에 대한 검색을 강화하기 위한 것으로, 2005년을 기준으로 90개 이상의 기관이

참여하고 있으며 약 20,000개의 장서를 포함하고 있다. 영국에서는 Archives hub 이전에도 기록물의 공동 활용을 위한 다양한 프로젝트가 진행되었지만, Archives hub만큼 대규모로 진행되지는 않았다. Archives hub는 장서기술 표준으로는 EAD를 기반으로 변환하여 사용하였다(Hill 2002).

4) Online Archive of California

California Digital Library의 일환으로 진행된 OAC(Online Archive of California)는 California내 도서관과 박물관, 기록보존소가 소장한 기록물들을 통합 검색하기 위한 것으로, 미국에서 진행된 대표적인 장서수준기술 사례에 해당된다. OAC에서는 장서 기술을 위한 표준으로는 EAD를 채택하고, 구체적으로 각 기관에서 사용해오던 메타데이터를 EAD 표준으로 변환하는 작업과 새로운 EAD기반 레코드를 생성하는 작업을 진행하였다. 또한 이러한 가상 장서 구축을 위한 기술적인 인프라 기반을 마련하였다(Online Archive of California 2005). EAD가 디지털 정보 생성 및 보존, 정보의 상호 운용성과 일관성 문제를 해결하리라는 판단 하에 수행된 이 프로젝트는 박물관, 도서관, 기록보존소간을 연결하는 대규모 합동 프로젝트로서의 의의를 갖는다(조현주 1998).

5) Jean Garling Performing Arts Project

호주의 New South Wales 주립도서관에서는 2002년 도서관 장서에 대한 접근을 개선하기 위해 Jean Garling Performing Arts Project를 진행하였다. 이 프로젝트에서는 예비 조사를 통해 장서수준기술의 필요성을 검토하고,

기존의 장서수준기술 표준을 활용하기로 결정하였다. 이 프로젝트에서는 미국과 영국, 호주 의 사례를 검토한 후 EAD가 갖는 안정성, 융통성, 공개 표준이라는 장점 때문에 EAD를 선택하였다. 스키마를 결정한 후 기존 개별수준 레코드에서 예술공연 프로토타입에 포함시킬 레코드를 조사하고 EAD DTD의 2002 버전을 약간 수정하여 프로토타입을 생성하였다. 프로젝트의 결과는 웹에 공개되었다(<http://www.sl.nsw.gov.au/collguides/dance/>).

프로젝트의 결과물로 공개된 '공연예술 가이드'(guide)는 목록을 통해 MARC 레코드로부터 링크할 수도 있고, 구글과 같은 검색엔진을 통해 이용할 수도 있고, 주제에 관한 주요 사이트를 안내하는 웹사이트에서 링크할 수도 있다. 이 가이드의 이용자 평가 결과, 이용자들은 공연예술에 관한 장서 가이드가 오랫동안 진행되어 온 공연예술 관련 지식을 파악할 수 있는 매우 훌륭한 메커니즘이라는 것에 만족을 나타냈다. 장서수준에 기반해서 구축된 이 프로젝트는 미래 세대를 위한 지식을 보존하는데 지대한 역할을 할 것이라고 예측하고 있다(Stumm, Moir, and Lawn 2003).

6) Chinese Biodiversity Information System

중국에서는 분산 환경하에서 이중의 디지털 장서를 효과적으로 다루기 위한 몇몇 프로젝트가 진행되었다. 그중 하나가 CBIS(Chinese Biodiversity Information System)이다. CBIS는 Chinese Academy of Sciences의 Biodiversity Committee에 의해 지원되었다. CBIS는 40여 개의 DB로 구성되며, 생물학 관련 자원을 통합

하는 분산 시스템이다. 이 시스템에서는 자체 개발한 장서수준기술 스키마를 사용하였다.

또 다른 사례로는 CSDL(Chinese National Science Digital Library)을 들 수 있다. 디지털 도서관을 통합하기 위한 이 사업에서는 장서수준 기술 표준으로서 DC를 사용하였다. 또 다른 사례인 CDLS(Chinese Digital Library Standardization Project)에서는 무수한 이중 디지털 자원을 탐색하기 위해 공개된 표준을 사용하는 것이 적합하다고 판단하고, RSLP CD 스키마를 사용하였다. 이러한 일련의 프로젝트를 계기로 중국에서는 국가 차원에서 장서수준 기술을 위한 표준을 마련하기 위한 준비를 하고 있다(Xin 2004).

7) Alexandria Digital Library

Santa Barbara 소재 California 대학교에서 주도한 ADL(Alexandria Digital Library)은 지도를 포함한 지리적 자료를 공간적으로 색인 하는데 중점을 둔 프로젝트로서, 디지털 도서관에서 다양한 장서를 통합하기 위한 핵심적인 요소로 장서 메타데이터를 인식하였다. 이런 인식 하에 ADL에서는 장서 메타데이터인 CLaM(Collection-Level Metadata)를 자체 개발하였다. CLaM에 포함되는 정보는 ① 장서의 범위, 목적, 조건, 책임사항 등의 정보, ② 장서내의 개별자료, 유형과 형식, 개별자료의 지리적 공간적 분포 등의 정보, ③ 장서 메타데이터에서 사용하는 어휘와 같은 구조적인 정보 등이다.

ADL에서는 이러한 CLaM에 기반한 검색을 위해 STARTS(Stanford protocol for internet retrieval and search)를 사용하였다. STARTS는 Stanford 대학에서 개발한 것으

로, 검색엔진들 간의 상호 작용을 수행할 수 있는 프로토콜이다. 이 프로토콜은 장서를 탐색하기 위한 질의어, 질의에 대한 검색 결과 및 장서의 메타데이터 구조를 기술한다(Hill 1999).

8) ITAM: MIMAS Collection Description

JISC에서 수행되고 있는 프로젝트 중 하나인 ITAM(Implementing the DNER Technical Architecture at MIMAS)은 MIMAS(Manchester Information & Associated Services)에 의해 운영되는 데이터세트에 관련된 메타데이터에 접근하기 위해 Cheshire 온라인 정보검색시스템을 사용한 Z39.50 기반구조를 개발하고 있다. ITAM이 추진하고 있는 과제 중 하나는 Z39.50을 통해 어떻게 MIMAS내의 장서를 검색할 것인가에 관한 것이다.

이를 위해 Z39.50에서 사용되는 속성 집합과 RSLP CD 스키마를 매핑하는 연구를 진행하였다. 예를 들어 Z39.50에서 사용되는 속성 집합에서 대표적인 Bib-1의 4, 34, 1016, 1017, 1035, 1097은 RSLP CD 스키마의 'Name of Collection'에 매핑되고, Bib-1의 59, 1016, 1017, 1018, 1035, 1101은 'Name of Location'에 매핑된다(MIMAS 2004).

5. 결론 및 제언

이 연구에서는 분산 자원의 효율적인 검색과 관리를 위한 CLD의 개념과 특징, 효과, 관련 표준을 검토한 후, 외국에서 수행된 CLD 관련 프로젝트를 살펴보았다.

CLD에서 장서는 물리적인 혹은 전자적인 객체의 집합이며, CLD란 장서에 대한 구조적이며, 안정적이고, 표준화된 기계가독형 기술을 의미한다. CLD는 네트워크 분산 환경에서 이용자들에게 이용할 수 있는 장서의 전체를 조망할 수 있는 정보지도(information map)의 역할을 하며, 궁극적으로는 검색성능을 높여 효율적으로 장서를 관리할 수 있다.

이전에 산발적이던 장서수준기술에 대한 논의가 RSLP CD 스키마와 같은 CLD 표준을 통해 하나로 묶일 수 있는 것이다. CLD 관련 표준 중 대표적인 것은 장서수준기술을 위해 개발된 RSLP CD 스키마, DC를 장서수준기술에 적용한 DC CD AP, 기록물 기술을 위해 사용되어 온 EAD가 있다. 이들 각각은 그 출발, 대상 유형, 포괄하는 기술 대상이 조금씩 다르다. RSLP는 EAD의 간단한 속성 세트라고 할 수 있으며, DC CD AP는 RSLP 보다 더 간단하다고 할 수 있다.

앞에서 살펴본 CLD 관련 사례들은 분산 자원의 주제적 접근을 용이하게 하기 위한 방안으로 시도된 것이다. CLD에 대한 관심은 영국뿐만 아니라 미국, 호주, 중국 등에서 점차적으로 확산되고 있다. CLD는 동일한 영역에 속하는 기관들의 장서를 통합하기 위해 또는 상이한 영역에 속하는 기관들의 이질적인 장서를 통합하기 위해 사용할 수 있다. 전 주제의 장서를 포괄하기도 하고, 특정 주제만을 대상으로 하기도 한다. 특히 영국에서는 전자도서관 프로젝트인 eLib의 3, 4 단계에서 CLD의 개념에 대한 논의와 함께 장서의 데이터 모델, RSLP CD 스키마 등을 개발함으로 주제접근을 강화하고 구축된 디지털 자원의 효과적인 이용을 꾀하였다.

또한 ADL에서는 장서수준기술의 메타데이터를 정의하고, 장서의 검색과 접근을 위한 프로토타입인 STAIRS를 사용하였고, 최근 MIMAS에서는 Z39.50과 CLD와의 통합을 위한 노력을 하고 있다.

국내의 경우, CLD에 대한 개념의 논의조차 미미하다. 다만 몇몇 기관에서 장서수준기술의 필요성을 인식하고 있음을 엿볼 수 있는 사례가 있다. 예를 들어 연세대학교 도서관에서는 2001년 디지털 도서관 구축시 YDL 메타데이터를 만들면서, 장서수준기술에 해당되는 '학술 데이터베이스 메타데이터'의 항목을 포함시켰다(김은수, 문숙희 2001). 그러나 이 사례는 학술데이터베이스만을 대상으로 한 것으로 연세대학교 중앙도서관의 전체 장서에 대한 것은 아니다.

또한 대상 장서가 포괄적인 '국가지식정보통합검색시스템'은 분산 자원 검색을 위한 대표적인 사례라고 볼 수 있다. 국가지식정보통합검색시스템은 '국가지식정보DB 구축사업 5개년 계획'(2000년~2004년)에 따른 '지식정보자원관리사업'의 핵심적인 결과물로서, 과학기술, 교육학술, 문화유산·예술, 역사, 정보통신, 건설기술, 산업경제, 해양수산 분야 등 전국 700여 개 기관이 협력해 구축한 지식정보 DB를 원스톱으로 검색할 수 있다(국가지식정보통합검색시스템 2005). 그러나 국가지식정보통합검색시스템의 품질을 평가한 결과에 의하면(한국데이터베이스진흥센터 2005; 전자신문 2004), 데이터의 양과 범위, 신뢰성은 우수한 것으로 나타났지만, 최신성, 검색 결과 오류와 검색 속도, 전혀 상관없는 주제가 지나치게 많이 검색되는 검색 성능의 문제, 검색결과의 제공형식의 통일 등은 개선할 점으로 지적되었다.

지금까지 이 연구에서 논의한 장서수준기술의 관점에서 보면, 국가지식정보통합검색시스템은 각 기관에서 구축한 장서를 모아 개별자료 수준의 검색만을 제공하며, 장서에 대해서는 구조화되지 않은 형식으로 간단한 설명만이 제공될 뿐이다.

이 연구에서는 장서수준기술이 분산자원의 검색과 관리에 효과적이라는 점과 외국에서 진행된 사례를 바탕으로 국가지식정보통합검색시스템의 메타 DB 구축시 장서수준기술을 포함할 것을 제안하고자 한다. 이를 통해 앞에서 지적된 전혀 상관없는 주제가 지나치게 많이 검색되는 검색 성능의 문제가 해결될 수 있을 것으로 예상된다. 국가지식정보통합검색시스템은 참여 기관이 이미 개별자료에 대한 별도의 스키마를 가지고 있고, 전자적인 개별자료와 인쇄형태의 개별자료의 장서도 다루어야 하므로, EAD와 DC CD AP보다는 RSLP CD 스키마의 수용을 검토하는 것이 타당할 것으로 예상된다. 이에 대한 구체적인 후속 연구가 필요하다.

마지막으로 우리나라에서 CLD 관련 연구와 사업이 활성화되기 위한 몇가지 제언을 하고자 한다.

(1) CLD의 필요성에 대한 인식의 확대가 필요하다. CLD의 필요성에 대한 인식과 함께, 구체적으로 영국에서 진행된 사례를 참고하여, 장서 기술을 위한 스키마 연구와 관련 프로젝트를 진행해야 하고, 이를 지속적으로 관리할 수 있는 조직이 마련되어야 한다. 또한 국가적 차원에서 디지털자원에 대한 마스터플랜을 세워, 각 기관마다 중복투자를 방지하고 특색있는 장서 개발이 필요하다.

(2) CLD 사업의 추진 방향은 디지털화된

자원을 중심으로 단기적으로는 주제별 거점을 선정하여 관련 사업을 진행하고, 장기적으로는 우리나라에서 생산되는 모든 장서에 대해 통합하는 방향으로 진행되어야 할 것이다. 영국의 사례를 보면, 한 분야의 통합검색을 먼저 시도하고 그 후 좀 더 범위를 확대해 나가는 것을 볼 수 있다.

(3) 각 기관마다 별도의 CLD를 유지하는 것이 아니라 하나의 CLD 표준에 따르는 것이 타당하다. 우리나라에서 표준 CLD 스키마를 만들려고 할 때 상호운용성과 재사용의 문제를

고려해야 한다. CLD 정보에 대한 표준화가 이루어지면, 개별 기관마다 장서 기술을 따로 만들 필요가 없다. 어느 기관에서 표준화에 따라 장서수준 기술을 하면 다른 기관에서는 이미 작성된 장서수준 기술을 재사용할 수 있을 것이다. 국내 디지털도서관들도 기관별로 장서정보를 작성하는 것이 아니라, 장서수준기술의 형식과 내용을 표준화해서 함께 사용하는 것이 바람직하다.

참 고 문 헌

- 강소연. 2003. 『기록물 검색도구 전산화를 위한 EAD DTD에 관한 연구』. 석사학위논문. 명지대학교 기록과학대학원.
- “국가지식정보통합검색시스템.” 2005. [cited 2005. 4.15].
 <<http://www.knowledge.go.kr/intro/SupportService.jsp>>.
- 김은수. 문숙희. 2001. “연세대학교 디지털 도서관의 메타데이터 구축 사례.” 『도서관』, 56(4): 52-76.
- 김태수. 1999. “더블린코어(Dublin Core) 메타데이터.” [cited 2005.4.5].
 <<http://dewey.yonsei.ac.kr/metadatas/DC.htm>>.
- 김현희. 2001. 기록물정보 관리를 위한 메타데이터 설계와 구현에 관한 연구. 『정보관리학회지』, 18(4): 57-81.
- 이수상. 2003. 메타데이터의 상호운용성 보장을 위한 요소기술. 『한국도서관·정보학회지』, 34(1): 91-109.
- 이응봉. 2003. Ubiquitous computing & Digital Library. 『제6회 디지털도서관컨퍼런스』, 2003년 12월 9일 서울: 한국과학기술회관
- 이제한. 2004. 디지털도서관품질평가에 관한 연구. 『한국문헌정보학회』, 38(1): 143-172.
- 『전자신문』. 2004. 국가지식정보 통합검색시스템 ‘헛바퀴’ 돈다. 2월 12일.
- 조현주. 1998. 문화유산 메타데이터 표준의 국내 적용 연구. 『국회도서관보』, 35(6): 76-89.
- 한국데이터베이스진흥센터. 2005. 데이터베이스 리뷰. [cited 2005.3.25].
 <<http://www.dpc.or.kr/02dbinfo/dbreview.php?field=&keyword=&page=1&dbnum=16&mode=detail>>.

- "Archives Hub." [cited 2005.2.24].
<<http://www.archiveshub.ac.uk/>>.
- Chapman, Ann and Bridget Robinson. 2004. "Collection-level Description: Thinking globally before acting locally." *ARI-ADNE*, 40. [cited 2005.4.22].
<<http://www.ariadne.ac.uk/issue40/robinson/>>.
- Dublin Core Collection Description Working Group(DC CD WG). 2003. Dublin Core collection description application profile, working draft. [cited 2005.3.22].
<<http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>>.
- Dunsire, G. and G. Macgregor. 2003. "Clumps and collection description in the information environment in the UK with particular reference to Scotland." *Electronic library and information systems*, 37(4): 218-225.
- "Encoded Archival Description(EAD) official EAD version 2002 web site." [cited 2005.2.24].
<<http://lcweb.loc.gov/ead/>>.
- Heaney, Michael. 2000. RSLP collection description: An analytical model of collections and their catalogues. [cited 2005.1.5].
<<http://www.ukoln.ac.uk/metadata/rslp/model/>>.
- Hill, L. L. et al. 1999. "Collection Metadata Solutions for Digital Library Applications." *JASIS*, 50(13): 1169-1181.
- Hill, Amanda. 2002. "The Archives Hub: a new resource for historical research." *SCONUL Newsletter* 26: 37-39. Information & Associated Services(MIMAS). 2004. ITAM: MIMAS Collection Level Description Z39.50 Mapping. [cited 2005.4.29].
<<http://metadata.mimas.ac.uk/ITAM/zmap.html>>.
- Johnston, P. 2001. ISAD(G)/EAD to RSLP CLD Schema/Application Profile Mapping. [cited 2005.2.1].
<<http://www.ukoln.ac.uk/metadata/rslp/isadg/crosswtk.html>>.
- Johnston, P. and B. Robinson. 2002. "Collections and Collection Description." *Collection Description Focus, Briefing Paper*, 1. 2002. [cited 2005.1.5].
<<http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf>>.
- Johnston, P. 2003. "The RSLP Collection Description Schema." *Collection Description Focus, Briefing paper*, 2. 2003. [cited 2005.2.3].
<<http://www.ukoln.ac.uk/cd-focus/briefings/bp2/bp2.pdf>>.
- Keane, Shirley. 2003. CLD Online Tutorial. [cited 2005.1.15].
<<http://www.ukoln.ac.uk/cd-focus/cdfocus-tutorial/>>.
- Macgregor, G. 2003. "Collection-level descriptions: metadata of the future?" *Library Review*, 52(6): 247-250.

- "Online Archive of California." [cited 2005.1.26].
<<http://www.oac.cdlib.org/>>.
- Powell, A., M. Heaney, and L. Dempsey. 2000. "RSLP Collection Description." *D-Lib Magazine*, 6(9). [cited 2005.2.15].
<<http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/september00/powell/09powell.html>>.
- Proffitt, Merrilee. 2003. Library and Archives Data Structures: EAD, MODS, RSLP Collection Description. [cited 2005.3.10].
<<http://www.rlg.org/en/downloads/2003metadata/proffitt.ppt>>.
- Stumm, D., S. Moir. and M. Lawn. 2004. Collection-level description: separating the woods from the trees. [cited 2005.3.1].
<<http://www.vala.org.au/vala2004/2004pdfs/16StMoLa.PDF>>.
- UKOLN. 2004. Comparison of the RSLP CD Schema and the Dublin Core Collection Description Application Profile. [cited 2005.1.22].
<<http://www.ukoln.ac.uk/events/cdfocus-schema-forum/breakout/breakout-2-3/rslpcd-dccdap.doc>>.
- Xin, Wang. 2004. "Research and usage of collection level metadata in Chinese digital libraries." *The International Information & Library Review*. 36: 291-295.