

대학도서관 서비스의 디지털 큐레이션 전략

Digital Curation Strategy for University Library Service

안 영 희(Young-Hee Ahn)*
박 옥 화(Ok-Wha Park)**

〈 목 차 〉

I. 서 론	2. 국내 사례
1. 연구의 필요성	IV. 대학도서관의 디지털 큐레이션 역할 및 모형설계
2. 연구 방법	1. 디지털 큐레이션 역할
II. 디지털 큐레이션 개념 및 생명주기 모델	2. 기술 적용
1. 디지털 큐레이션 개념	3. 디지털 큐레이션 워크플로우
2. 생명주기 모델	V. 결론 및 제언
III. 디지털 큐레이션 사례	
1. 국외 사례	

초 록

본 연구는 디지털 정보자원의 발전에 따른 보존, 아카이빙, 디지털 큐레이션에 대한 개념을 파악하고, 디지털 큐레이션의 국내·외 사례와 디지털 정보자원의 생명주기 정보처리를 기반으로 대학 도서관에서의 디지털 큐레이션의 역할 및 전략에 대해 제시하고자 하였다. 생명주기를 기반으로 한 생산자, 큐레이터, 이용자의 디지털 큐레이션 관계 모형을 근간으로 하여, 기관 레포지터리인 대학도서관의 디지털 자원에 대한 보존 및 큐레이션 모형을 설계하였다. 그 결과, 대학 간 협력체제와 자원공유에 기초가 되는 디지털 큐레이션 체제를 수립하는 토대를 마련하였다.

키워드: 대학도서관, 디지털 아카이빙, 디지털 큐레이션, 보존, 큐레이션 전략

ABSTRACT

This study is to identify the concept of the digital curation, archiving and preservation resulted in the development of digital information resources, and is to suggest the roles and strategies of the digital curation in university library based on the lifecycle information process of the digital information resources and the examples of the digital curation from domestic and abroad. I have designed a curation model and the preservation of the digital resources in university library by using creators, curators, and users who stand on the basis of the lifecycle. As a result, I have built up a system that is to be a basis to share resources and cooperation among universities.

Keywords: Curation Strategy, Digital Archiving, Digital Curation, Preservation, University Library

* 백석대학교 학술정보관 사서(yha@bu.ac.kr)(제1저자)

** 충남대학교 사회과학대학 문헌정보학과 교수(owpark@cnu.ac.kr)(공동저자)

• 접수일: 2009년 11월 26일 • 최초심사일: 2009년 11월 30일 • 최종심사일: 2009년 12월 26일

I. 서론

1. 연구의 필요성

EMC(Enterprise Content Management)와 IDC(International Data Corporation)가 공동으로 추진하고 있는 보고서에 따르면 정보기술의 발달과 급속한 변화로 국내 디지털 정보자원의 생산량이 2011년에는 2007년 대비 6배 이상 증가한 27,237페타바이트에 이르고, 국민 1인당 평균 디지털 정보량은 571GB에 도달할 것으로 전망했다.¹⁾

국내 대학의 경우 연구소 및 산학협력을 통하여 각 연구기관마다 디지털 데이터를 이용한 결과물의 공동 활용이 가시화되면서 디지털 정보자원의 생산이 증대되고, 산학협력에 따른 기술이전 및 연구가 가파르게 증가하고 있다.²⁾

이와 같은 움직임은 디지털 정보의 생산과 유통의 증가로 이어지고 디지털 보존에 대한 중요성을 재고하는 요인이 된다. 예를 들면 산학협력을 통하여 도출된 연구결과물과 대학 내 연구소에서 생산된 데이터가 지속적인 연구 활동을 위한 학술정보자원으로 활용되기 위해서는 디지털 정보의 생산단계에서부터 데이터의 보존 및 관리가 필수적으로 수반되어야만 가능하다.

그러나 디지털 정보자원은 매체의존도가 높아 새로운 매체의 등장과 기술적 변화 또는 퇴화, 저장 환경, 조작 실수 등으로 쉽게 훼손·파괴될 수 있다. 따라서 디지털 정보자원을 안전하게 보호하여 유용한 형태를 유지하기 위한 장기보존 환경이 마련되어야 한다. 또한 디지털 정보자원이 현재와 미래의 이용자를 위하여 디지털 정보자원의 생명주기를 통해 연구결과 데이터가 효율적으로 재사용될 수 있도록 디지털 보존전략이 필요하다.

이와 같은 보존 및 원활한 재사용을 지속적으로 가능하게 하기 위해서는 대학도서관이 기관 레포지터리로서 디지털 보존과 데이터 큐레이션을 수행할 수 있도록 디지털 큐레이션 역할을 담당해야 한다.

대학도서관에서는 모든 분야를 망라하는 다양한 단행본, 학술논문, 연구보고서 등을 선택, 수집, 조직 및 제공하고 있다. 그러나 엄청난 공공재원으로 수행된 연구과정의 데이터를 체계적으로 수집하고 유통하려는 노력은 상대적으로 미흡하며, 연구에 투자된 재원의 효율성과 관련 분야에서 다양하게 재사용되도록 활용도를 높이려는 노력도 요구되고 있다. 심원식³⁾은 대학 내 연구소를 비롯한

1) EMC와 IDC가 발표한 “2011년까지 전 세계 디지털 정보 성장 전망 보고서(The Diverse and Exploding Digital Universe : An Updated Forecast of Worldwide Information Growth Through 2011).”

2) 매일경제, 2009년 11월 2일,

〈<http://news.mk.co.kr/outside/view.php?year=2009&no=565598>〉 [cited 2009. 11. 2].

3) 심원식, Open Access to Scientific Research Data, 제7회 오픈지식정보포럼 발표자료(서울 : 한국과학기술정보연구원, 2009), p.3, 〈http://icon.ndsl.kr/i_seminar/i_forum/forum_1.jsp〉 [cited 2009. 10. 21].

산학협력으로 도출된 연구데이터는 보존과 재사용을 위해서, 반복실험 및 결과의 검증을 위해서, 반복할 수 없거나 반복하는데 많은 비용이 드는 실험 및 관찰의 경우에, 교육적 목적으로, 여러 실험에서 생성된 데이터를 결합하는 메타연구 또는 모델연구 등을 이유로 반드시 공유되어야만 한다고 보고하였다.

미국과학재단(National Science Foundation)⁴⁾은 NSF 지원으로 수행된 연구를 통해 생성·수집된 주요 데이터, 표본, 물리적 자료 및 보조 자료를 최소한의 비용과 적정범위 내에서 다른 연구자에게 제공하도록 요구하고 있다. 또한 영국은 연구데이터에 영구적으로 접근할 수 있도록 디지털 큐레이션 센터(Digital Curation Centre)⁵⁾를 설립하여 디지털 자원에 대한 보존뿐만 아니라 재사용을 위한 다각적인 지원시스템을 구축한 상태다. 한편 국내에서는 학술데이터의 재사용에 대한 필요성과 방법론에 대한 관심이 고조되고 있으나, 중·장기적으로 보존, 아카이빙 및 큐레이션을 통한 데이터의 가치, 재사용 및 보존가치에 대한 기준이 마련되어 있지 않은 실정이다. 또한 디지털 데이터 고유의 특성과 데이터에 대한 학술 커뮤니케이션 시스템의 부재로 데이터의 수집, 배포 및 보존에 장애가 되고 있다.

한국교육학술정보원(KERIS)에서 2008~2009년에 ‘학술자원관리체제전략계획 수립’의 일환으로 대학 내 학술자원 현황을 조사한 대학별 연구소 데이터 현황을 살펴보면,⁶⁾ 185개 대학 1,481개의 연구소에서 약 157,315건의 연구결과물을 디지털화하였다. 그러나 보다 장기적인 데이터 아카이빙과 체계적인 관리를 위해 디지털 큐레이션 서비스를 통한 경쟁력 있는 레포지토리를 구축하여 국내 학술데이터의 자산화 및 공유가 시급한 상황이다.

본 연구는 기관 레포지토리로서 대학도서관의 디지털 큐레이션 체제 구축을 위한 방안으로 생명주기를 기반으로 한 디지털 큐레이션의 개념과 필요성을 구명하고, 해외의 디지털 큐레이션 사례를 분석하여 국내 실정에 맞는 대학도서관의 디지털 큐레이션 모형을 개발하고자 하였다.

2. 연구 방법

문헌연구 및 홈페이지를 통하여 디지털 정보자원의 보존, 아카이빙, 큐레이션에 대한 각각의 개념을 정리하고, 영국의 DCC에 구축된 디지털 정보자원의 생명주기를 중심으로 생산, 수집, 보존, 재사용, 변환 등의 각 단계별 처리과정을 기반으로 국내 실정에 맞는 대학도서관의 디지털 큐레이션 역할과 모형을 설계하였다.

4) NSF Homepage, <<http://www.nsf.gov/>> [cited 2009. 10. 21].

5) DCC Homepage, <<http://www.dcc.ac.uk/>> [cited 2009. 10. 21].

6) 한국교육학술정보원(KERIS) 국가지식정보센터 담당자로부터 받은 자료에서 조사한 내용임.

II. 디지털 큐레이션 개념 및 생명주기 모델

1. 디지털 큐레이션 개념

미국 대학도서관의 '미래에 대한 환경 조사보고서'에 의하면, 디지털 컬렉션, 디지털 아카이브에 대한 관심이 고조되면서 데이터 저장, 검색, 큐레이션, 서비스 방식 등이 개선될 것이라고 예측하고 있다.⁷⁾

아카이빙과 큐레이션은 동의어로 사용되기도 하는데, 디지털 아카이빙은 단순히 디지털 자산(assets)을 관리하는 것을 말한다. 이는 디지털 정보자원이 생산되는 시점부터 디지털 정보자원에 대한 관리가 적절히 고려되지 않으면 보존조치가 미흡할 수 있다는 태생적 한계로 인하여 디지털 큐레이션의 필요성이 제기되었다.

디지털 큐레이션은 디지털 자원의 보존 및 관리뿐만 아니라 생산단계부터 포괄적인 활용까지의 모든 과정을 포함한다는 점에서 디지털 아카이빙보다 넓은 범주의 활동을 지칭한다. 사용되는 분야와 맥락에 따라 디지털 보존, 디지털 아카이빙, 디지털 큐레이션의 개념에는 상당히 다른 이해가 존재하므로, 이들 용어에 대한 개념을 고찰해 보았다.

가. 디지털 보존

디지털 보존은 특정 데이터 형태가 지속적으로 관리되어 기술적인 변화 속에서도 계속 접근가능하고, 이해될 수 있도록 하는 아카이빙 활동의 일부로 볼 수 있다. 디지털 정보자원의 무결성을 상실하지 않고, 저장매체의 손상 및 퇴화, 기술적 퇴화의 위협으로부터 디지털 정보를 안전하게 보호하여 유용한 형태로 존속하게 하는 일련의 관리활동으로 정의할 수 있다.⁸⁾

나. 디지털 아카이빙

아카이빙이란 공적인 기록물의 보존을 처리하는 과정이나 컴퓨터 파일의 일시적 백업을 의미하는 것으로 사용되어 왔다. 그러나 디지털 정보자원의 이용이 가파르게 확대되면서 디지털 정보자원의 원문확보 및 보존의 필요성이 증가하게 되어 디지털 아카이빙이란 용어로 구분하게 되었다.

디지털 아카이빙은 지속적 가치가 있다고 판단되는 디지털 객체를 장기간 관리하는 활동으로 정의되며, 이는 가치 있는 디지털 자원을 선별하고, 그 내용 및 기능을 보존·관리하여 장기간 접

7) The Association of College and Research Libraries Research Committee, *Environmental Scan 2007*(Chicago, Association of College and Research Libraries, 2008),
<http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/Environmental_Scan_2.pdf> [cited 2009. 11. 23].

8) 이수상, 디지털 아카이빙의 워크플로우와 보존처리 기술에 관한 연구, 한국도서관·정보학회지, 제35권, 제3호(2004), pp.119-138.

근할 수 있도록 하는 전반적인 활동을 포괄하는 개념이다.⁹⁾

다. 디지털 큐레이션

큐레이션은 본래 미술관이나 박물관 소장품을 정리하거나 전시 또는 작품에 대한 연구 등을 실시하는 학예직 또는 미술관의 큐레이터처럼 디지털도서관에서 수집된 콘텐츠를 정리하거나 그 이용 및 보존을 보조하고 촉진시키는 사람들이 큐레이터처럼 콘텐츠를 구축하고 보존하는 작업을 말한다.¹⁰⁾

디지털 큐레이션은 2001년 10월 19일 런던에서 개최된 Digital Preservation Coalition and the British National Space Centre 주최로 개최한 ‘Digital Curation : digital archives, libraries and e-science seminar’에서 처음으로 사용된 용어다.

디지털 큐레이션은 디지털 자원을 수집, 유지, 영구보존, 제공하는 아카이빙 활동을 의미하며, 데이터가 학술적인 관심 대상이 되는 생명주기동안 이러한 데이터를 적극적으로 관리하고 평가하는 것으로, 이는 데이터의 재현과 재사용을 위해 반드시 필요한 영역이다. 또한 큐레이션은 미래를 위해 데이터를 보호하고 위험을 관리한다.

이러한 디지털 큐레이션은 연구자, 과학자, 역사학자 등이 현재와 미래에 참고할 수 있도록 디지털 자원의 장기보존소를 설립하고 개발하는 과정이다. 이 정의는 디지털 아카이빙과 보존의 과정뿐만 아니라 양질의 데이터 생산과 관리에 필요한 모든 과정 그리고 새로운 정보와 지식자원을 생산하기 위해 데이터에 새로운 가치를 추가하는 기능까지 포함한다.¹¹⁾

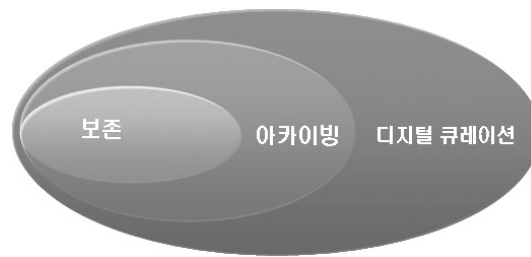
디지털 큐레이션은 데이터 큐레이션, 디지털 아카이빙, 디지털 보존을 포함하는 디지털 자산 관리로서, 세부기능은 <표 1>과 같다.¹²⁾

<표 1> 디지털 큐레이션 세부기능

기능	내용
데이터 큐레이션	<ul style="list-style-type: none"> • 데이터의 사용을 장려하고 관리하는 활동 • 데이터 검색이 편리하도록 지원
디지털 아카이빙	<ul style="list-style-type: none"> • 논리적으로 체계를 유지하여 정보의 신뢰성을 높이는 활동
디지털 보존	<ul style="list-style-type: none"> • 데이터를 지속적으로 이용할 수 있도록 함 • 오랜 기간 동안 데이터가 저장될 수 있도록 함 • 기술의 변화에 대처하여 데이터의 손실을 막도록 함

9) 설문원 외, 국가 디지털 아카이빙 체제 구축에 관한 연구, 한국과학기술정보연구원 최종연구보고서 K-05-ID-01R-5 (2005), p.3.
 10) 국립중앙도서관, 미국 대학도서관의 장래에 관한 환경 조사 보고, 도서관연구소 웹진, 제10권(2008), <bellpower.tistory.com/attachment/ek180000000000.pdf> [cited 2009, 10, 2].
 11) What is digital curation? <http://www.dcc.ac.uk/what.html> [cited 2009, 10, 2].
 12) 중앙대학교 산학협력단, 문화콘텐츠제작센터 운영 프로그램 개발 및 시범제작 사업 단위사업 영역보고서 4, 창작 환경-창작지식생태계 결과보고서(서울 : 중앙대학교, 2008), p.79.

디지털 보존, 디지털 아카이빙, 디지털 큐레이션의 개념 범주의 차이는 <그림 1>과 같이 표현할 수 있다. 이를테면 보존은 아카이빙의 한 측면이며, 아카이빙은 큐레이션을 위해 필요한 활동이라고 볼 수 있다.¹³⁾



<그림 1> 보존, 아카이빙, 디지털 큐레이션 관계도¹⁴⁾

그러므로 실제 디지털 보존과 아카이빙, 큐레이션 활동은 밀접한 연계를 가질 수밖에 없다. 장기적인 디지털 보존 대책이 효과적으로 실행되기 위해서는 적절한 아카이빙 및 큐레이션 전략이 뒷받침되어야 하고, 디지털 보존이나 아카이빙 계획을 수립할 때에는 큐레이션을 포함하는 넓은 범주를 고려하는 것이 바람직하다.

(1) 데이터 큐레이션

데이터 사용을 관리하는 활동으로 생산시점에서부터 데이터의 정보자원으로서의 가치를 발견하고 미래에도 재사용할 수 있도록 하는 활동이다. 데이터 큐레이션은 데이터 셋을 읽을 수 있도록 업데이트하고, 디지털 자산을 사용하기 위해 방대한 데이터 셋을 관리한다.

데이터 큐레이션의 중요성은 전문화된 양질의 자료를 수집·제공하고, 활발한 정보검색을 위하여 지속적인 업데이트 및 동시목적에 부합하도록 세부적으로 관리하여 재사용 및 활용이 원활하도록 데이터 생산자, 큐레이터 및 정보이용자 간의 소통을 바탕으로 한 서비스 개발에 있다.¹⁵⁾

(2) 디지털 큐레이터

디지털 큐레이터는 가장 최근의 정보전문가 유형으로 디지털 정보를 수집·보존·관리하는 역

13) Lord, P. and A. Macdonald, *e-Science Curation Report-Data Curation for e-Science in the UK : An Audit to Establish Requirements for Future Curation and Provision*, Prepared for the JISC Committee for the Support of Research(JCSR), 2003, p.12.

http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf [cited 2009. 10. 2].

14) 설문원 외. 전게서, p.4.

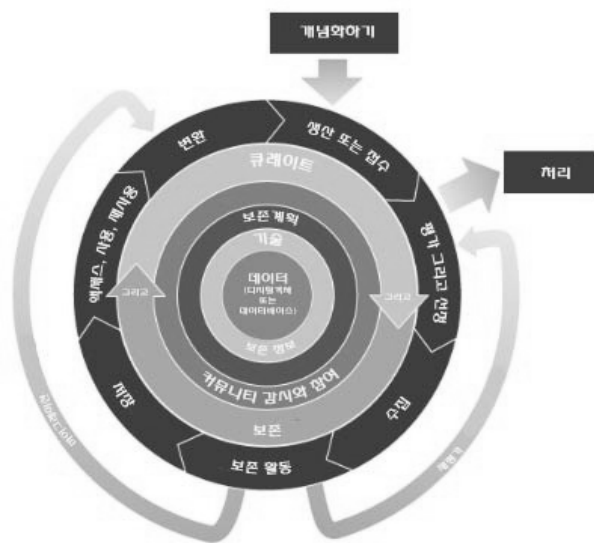
15) Lord, P. and A. Macdonald, *op. cit.*, 2003.

할을 담당한다. 어떤 정보를 어떻게 보존할 것인지, 보존하기 위한 최선의 방법은 무엇인지, 정보의 생명주기부터 보존 및 재사용을 위한 변환작업 등의 문제를 컨트롤할 수 있는 사람이 바로 디지털 큐레이터다. 그러므로 디지털 큐레이터는 아카이빙 된 아이টে에 대한 감시와 더불어 매체 이전, 전략 개발, 보존상의 기술적 문제 등을 전문가와의 협력을 통해 처리해야 한다.

2. 생명주기 모델

디지털 정보자원의 큐레이션에 필요한 활동은 디지털 객체의 생명주기를 기반으로 파악할 수 있으며, 생명주기 단계를 통하여 큐레이션에 필요한 각종 활동들을 규정할 수 있다.¹⁶⁾

영국의 디지털 큐레이션 센터(DCC)가 개념화 한 생산, 평가, 수집, 보존, 저장, 액세스, 대체 변환까지 데이터를 보존하고 관리해나가는 디지털 큐레이션 생명주기 모델은 <그림 2>와 같다.



<그림 2> 디지털 큐레이션 생명주기 모델

디지털 큐레이션 생명주기 모델은 성공적인 큐레이션과 데이터의 보존을 위한 우수한 그래픽 모형으로, 각 기관마다 필요로 하는 단계별 활동계획을 수립하는데 사용할 수 있다.

또한 디지털 큐레이션 생명주기 모델은 크게 데이터, 전체 생명주기 활동, 연계 활동, 비정기 활동으로 구분할 수 있으며, 각각의 분야별 활동은 <표 2>와 같다.

16) Digital Curation Centre Curation Lifecycle Model,
 <<http://www.dcc.ac.uk/lifecycle-model/#11>> [cited 2009, 10. 2].

〈표 2〉 디지털 큐레이션 생명주기

구 분		내 용
데이 터	디지털 객체	텍스트, 이미지 또는 사운드 파일과 같은 디지털 아이템의 단순한 디지털 객체 또는 웹사이트와 같은 복잡한 디지털 객체
	데이터베이스	컴퓨터 시스템에 저장된 레코드 또는 데이터의 구조화된 수집품
전 체 생 명 주 기 활 동	기술 및 표현정보	적절한 표준을 사용하여 장기적으로 제어하기 위해 관리, 설명, 기술, 구조 및 보존 메타데이터를 할당
	보존계획	디지털 자료의 큐레이션 생명주기를 통해 보존을 위한 계획모든 큐레이션 생명주기의 경영 및 관리를 위한 계획 포함
	커뮤니티 감시와 참여	커뮤니티 활동 감시 표준, 도구, 적합한 소프트웨어 개발 참여
	큐레이트 및 보존	큐레이션 생명주기를 통하여 큐레이션과 보존을 원활히 하기 위한 계획 관리 및 행정조치를 담당
연 계 활 동	개념화하기	캡처 방법과 저장 옵션을 포함하는 데이터의 생성, 수집, 저장에 관한 표현과 계획
	생산 또는 접수	관리, 설명, 기술, 구조 및 보존 메타데이터를 포함하는 데이터 생성 접수되는 데이터의 메타데이터 수집을 위한 정책 필요
	평가 및 선정	데이터 평가 또는 장기 큐레이션과 보존을 위해 적합 여부 선정
	수집	기록관, 레포지터리, 데이터 센터 또는 다른 관리자에게 데이터를 전송시킴
	보존 활동	본질을 유지하고 장기 보존을 하기 위해 활동 보존 활동은 데이터의 무결성을 유지하고, 진실성, 신뢰성이 유용해야 함 이 활동은 데이터의 클리닝, 유효성 검증, 보존 메타데이터의 지정, 표현정보 지정과 구조데이터 또는 파일 형식 유지 포함
	저장	데이터를 안전한 방식으로 저장하기 위해 관련 표준을 준수
	접근, 사용, 재사용	데이터 유지는 사용자와 재사용자 모두 접근 가능 강한 접근 제어와 인증 절차를 적용
	변환	기존 데이터로부터 새로운 데이터를 생성(마이그레이션, 부분발췌, 발행 등)
비 정 기 적 활 동	처리	데이터처리, 문서화 정책, 지침 또는 법적 요구사항에 따라 장기 큐레이션 보존을 위해 부적합한 데이터의 처리 일반적으로 데이터는 다른 아카이브, 레포지터리, 데이터 센터 또는 다른 관리자에게 양도 및 파기 데이터의 성질, 법적 이유로 보안 및 파기
	재평가	보다 나은 평가를 위해 검증 절차가 실패로 돌아가면 데이터를 재평가
	마이그레이션	데이터를 다른 포맷으로 마이그레이션 하드웨어 또는 소프트웨어의 의존성 탈피를 포함

Ⅲ. 디지털 큐레이션 사례

1. 국외 사례

가. 디지털 큐레이션 센터(Digital Curation Center)¹⁷⁾

DCC는 디지털 데이터의 가치를 향상시키고 영속적으로 이용되도록 영국의 각 기관들이 데이터

17) DCC Home page, <http://www.dcc.ac.uk> [cited 2009. 10. 2].

를 저장·관리·보존하고 체계적으로 활용할 수 있도록 전문기술과 경험을 제공·지원하는 기관으로 2004년 11월 에든버러 National e-Science Centre에서 공식 발족되었다.

JISC(Joint Information Systems Committee)가 2002년 ‘지속적 접근과 디지털 보존전략 2002-2005’에서 “디지털 보존을 위해 분산형 개발 활동을 수행하고, 연구자 및 서비스 기관들이 네트워크를 연결하여 모일 수 있는 중심을 마련하여, 분산된 디지털 데이터센터 및 보존기관들을 위한 통합서비스, 표준, 도구를 개발하기 위한 조직이 마련되어야 한다”는 것이 계기였다.

(1) DCC의 목적

데이터나 정보자원의 저장소 역할이 아닌 데이터복구(Curation) 및 장기적 보존과 관련한 이슈들을 다룬 연구프로그램을 기반으로 데이터를 큐레이션하는 기관이나 개인을 지원하는 확장서비스와 현장서비스 제공이 목적이다.

(2) DCC의 활동

DCC는 디지털 큐레이션에 관한 개요설명이나 매뉴얼, 관련기술이나 표준 동향을 소개하는 자료, 용어집, 사례, 도구의 소개, 전문가와의 인터뷰 등의 정보를 웹사이트로 제공하는 것 외에 전자 메일이나 전화를 통한 헬프데스크를 운영하며, 관련기술이나 과제 등에 관한 워크숍, 이벤트 개최, 국제협력 등 다양하게 활동하고 있다. 또한 디지털 보존을 위하여 여러 곳으로 분산되어 있는 개발 활동, 연구자들, 서비스의 광범위한 네트워크에 연결하여 큐레이터 교육을 이수한 직원과 연구에 초점을 맞춰 디지털자료센터와 보존서비스를 위한 핵심적 서비스, 표준, 도구들을 개발하고 있다.

나. 캐나다국가연구회의-캐나다과학기술정보연구소((NRC-CISTI)¹⁸⁾

NRC(National Research Council) 국립과학도서관에서는 연구자가 용이하게 캐나다의 과학·기술·의학(STM)분야의 데이터 셋과 데이터 레포지터리에 액세스하여 데이터를 이용할 수 있는 사이트 ‘Gateway to Scientific Data(GSD)’를 공개하였다.

GSD는 캐나다 연구자들이 연구한 방대한 데이터를 보다 쉽게 검색·이용 가능하게 하여 다른 연구자들을 돕고, 다양한 방법으로 자료에 접근함으로써 연구자들은 보다 혁신적이고 종합적인 연구 성과를 양산하고 있다. 또한 데이터 관리 및 큐레이션에 관한 우수사례와 정책, 관련저널, 참고 문헌, 세미나 및 관련회의 자료에 대한 정보도 제공한다.

GSD의 데이터 관리와 큐레이션을 위한 관련도구로는 식별(Identification), 데이터 기술(Data Description), 관리 소프트웨어(Management Software), 보존(Preservation), 데이터 저장 도구

18) NRC-CISTI Home page, <<http://cisti-icist.nrc-cnrc.gc.ca/eng/services/cisti/gateway-scientific-data.html>> [cited 2009. 10. 2].

(Data Storage Tools), 디지털 레포지터리 및 도서관 모델(Digital Repository and Library Models), 문서 및 표준(Documentation and Standards), 전략 도구 및 지침(Strategy Tools and Guidelines), 웹 아카이빙 도구(Web Archiving Tools) 및 관련 프로젝트(Relevant Projects) 등이 있다.

NRC 국립과학도서관 사이트에는 데이터 셋에 링크되는 'Scientific Data Sets'와 데이터 관리 및 큐레이션에 관한 'Data Management and Curation'도 제공하고 있으며, NRC-CISTI(Canada Institute for Scientific and Technical Information)의 Open Access 성과는 전 세계 R&D분야의 발전에 지대한 영향을 미치고 있다.

다. UCL(University College London) Library¹⁹⁾

UCL 도서관에서는 디지털 큐레이션에 대하여 “생명주기를 통하여 선택된 디지털 파일이 영구히 남을 수 있도록 지적 콘텐츠가 시간이 지나고 변화하는 환경에서도 누구라도 오랫동안 필요한 자료를 이용할 수 있도록 보장하는 것”이라고 정의하였다. 이를 위하여 소장하고 있는 디지털 자산에 디지털 큐레이션의 원칙을 적용하여 정보기술과 커뮤니케이션을 위한 디지털 컬렉션 개발, 메타데이터, 포맷, 정보공개, 저작권, 보존, 유지, 관계, 비즈니스 프로세스 등 9가지 디지털 큐레이션 전략을 지원한다.

예를 들면 UCL 도서관이 소장하는 디지털 자원과 각 학과 또는 다른 기관에서 소장하는 라이선스가 확보된 자료들을 관리하고, 디지털 큐레이션 활동을 위해 다양한 기술적 환경을 제공하며, IT 서비스 그룹을 별도로 운영한다. 이를테면 디지털 큐레이션 팀이 메타데이터 표준과 보존에 관한 표준포맷을 채택하여 디지털 큐레이트 콘텐츠의 텍스트 마이닝과 데이터 마이닝을 지원한다. 디지털 콘텐츠는 자유롭게 이용되고, 이용제한자료는 도서관 관리자에 의해서만 접근 할 수 있다.

라. MSU(Michigan State University) Digital Curation Planning²⁰⁾

MSU 디지털 큐레이션 프로젝트는 디지털 자산의 보존과 관리를 'Digital Preservation Management : Implementing Short-term Strategies for Long-term Problems' 웹사이트를 통하여 소개하고, 디지털 큐레이션을 “생애주기 전반에 걸쳐 신뢰할 수 있는 디지털 자원의 강화와 관리활동이며, 컴퓨터 파일을 캡처, 저장, 보존과 장기적으로 접근할 수 있도록 디지털 객체를 안전하게 하는 특정한 큐레이션”이라고 정의하였다. 또한 디지털 자원이 기술의 발달과 비례하여 변화의 위험에 노출되어 있으므로 여러 기관들이 협력하여 디지털 큐레이션 계획과 가이드라인을 발전시키고, 신뢰할 수 있는 보존, 관리 및 지적 고유의 디지털 자산을 안전하게 유지하는 가이드라인을 실행하고 있다.

19) UCL Home page, <<http://www.ucl.ac.uk/Library/dig-cur-strategy.shtml>> [cited 2009. 10. 2].

20) MSU Home page, <<http://msudcp.archives.msu.edu/>> [cited 2009. 10. 2].

2. 국내 사례

한국과학기술정보연구원(KISTI)에서 e-Science Grid나 국가과학기술종합정보시스템(NTIS)의 발전방향에 대한 연구보고서²¹⁾에서 데이터베이스 아카이빙 활동을 수용할 때 디지털 큐레이션 센터를 국가적 차원에서 설립할 필요가 있다고 제안하였다.

또한 황혜경 등²²⁾은 디지털 큐레이션 프로세스를 중심으로 디지털 정보자원의 보존관리의 필요성을 보고하였고, 심원식²³⁾은 과학데이터의 공유와 활용을 위한 전략적인 방안으로 데이터의 보존 및 큐레이션을 담당할 전문기관과 인적기반이 필요하다고 하였으며, 박진호²⁴⁾는 국립중앙도서관을 중심으로 디지털 장서관리시스템과 관련하여 디지털 생명주기기반 정보자원처리에 대하여 발표하였다. 그러나 해외에서의 활발한 움직임과는 달리 디지털 큐레이션과 관련된 문제의 중요성을 제안하는 초기단계에 머물러 있는 실정이며, 디지털 큐레이션 모형설계에 이르지 못하는 실정이다.

IV. 대학도서관의 디지털 큐레이션 역할 및 모형설계

국내 디지털 큐레이션의 현황은 아직 초기단계에 머물러 있기 때문에 검토하고 발전시켜야 할 과제들이 많다. 우선, 소속기관 및 기관장의 디지털 큐레이션에 대한 인식을 확장하여 미래의 이용자에게 디지털 학술자원을 유용하게 제공하고자 하는 인식의 전환이 필요하다. 기관 레포지터리로서 대학도서관은 디지털 큐레이션의 역할을 담당하고 디지털 정보자원에 대한 보존, 관리 및 재사용을 위한 정책과 계획이 수립되어야 한다.

또한 각 학문 분야에 따라 연구 특성에 맞는 큐레이션 전략을 세우고, 학술 정보자원 유형 중 우선순위에 따라 디지털 자원의 특성, 생산 및 활용 등 기관의 요구에 맞게 큐레이션의 메커니즘을 구축하는 표준 및 도구도 개발되어야 한다.

1. 디지털 큐레이션 역할

디지털 큐레이션은 대학 내에서 생산되는 연구결과물 등의 디지털 데이터에 대한 학술적, 보존적, 문화적 가치를 선별한 후 보존, 관리 및 재사용을 위해 디지털 정보자원을 변환하여 새로운

21) 설문원 외, 전계논문.

22) 황혜경 외, 과학기술정보자원의 장기보존 : 디지털 큐레이션 프로세스를 중심으로(서울 : 한국과학기술정보연구원, 2009).

23) 심원식, 전계논문.

24) 박진호, 디지털 생명주기기반 정보자원처리, 한국정보관리학회 학술발표회, 2009. 11.

가치를 부여하는 것이다. 이를 위해서는 디지털 정보자원, 소프트웨어 및 하드웨어 등의 시스템을 중심으로 데이터 확인, 메타데이터 보호, 데이터의 탐색과 검색, 연계데이터와 출판 및 학문 분야의 통합적 접근을 지향하는 큐레이션 기술을 적용하여 데이터를 관리하여야 한다.

가. 디지털 큐레이션 전략

대학 도서관에서 단일 기관으로 디지털 큐레이션을 수행하기에는 불가능하므로 여러 대학이 협력하여 자원을 공유하기 위한 연구, 이용 및 파트너십을 위한 협동 체제를 구축해야 한다.

디지털 큐레이션을 위한 각 학문 분야별 특성에 맞는 가이드라인을 작성하고 디지털 정보자원의 생명주기에 따른 적합한 표준을 채택하여 공유된 디지털 데이터를 변환하여 연구에 재사용할 수 있도록 활동한다.

나. 디지털 큐레이션 미션

디지털 정보자원의 지속적인 큐레이션을 위해 대학 간 연구결과물의 재사용, 비용절감, 교육목적 등으로 협력 체제를 구축한다. 또한 디지털 정보자원의 장기적인 보존을 위해 프로그램 및 계획을 개발하고, 각 연구기관에서 진행되고 있는 연구에 대한 정보 등을 배포하는 등 전문지식 및 공유를 위한 시스템을 구축하여야 한다. 또한 디지털 정보자원을 관리하는 디지털 큐레이터도 양성하여야 한다.

디지털 큐레이터는 디지털 큐레이션을 위한 생산자, 큐레이터, 이용자 등의 이해 당사자 간의 협력 및 커뮤니티가 원활하도록 활동한다. 각 기관에서 관리되었던 디지털 정보자원을 통합하여 기관 데이터 센터에서 관리될 수 있도록 하며, 각 연구자들에게 인센티브제를 도입하여 연구데이터의 공유 및 재사용을 활발하게 하고, 기관과 국가차원에서 기금조성을 위한 지원 대책도 수립한다.

큐레이션을 지원하기 위한 분산형 기본시설로 실험실 레포지터리, 기관 레포지터리, 주제 또는 학문분야 레포지터리, 데이터베이스와 데이터 센터를 설치한다.

2. 기술 적용

디지털 정보는 생성시점에서 그 내용을 저장매체로부터 다른 곳으로 옮기는 순간 물리적 매체와 논리적 개체가 분리되면서 최초의 내용이 의미하는 바가 달라질 수 있다. 디지털 정보는 환경이 변화되고 기술이 노화하더라도 정보내용을 읽을 수 있도록 물리적 매체를 구동하는 환경을 변경시켜야 한다. 생산 당시의 내용과 외관을 재현하지 못한다면 디지털 정보를 장기적으로 보존할 의미가 없으므로 장기적 보존을 위한 요소기술, 즉 보존 전략을 위한 기술 적용이 필요하다.

이를 위해서는 생명주기 관리를 통하여 영속적으로 디지털 객체를 활용할 수 있도록 하는 마이그레이션, 에뮬레이션, UVC 등의 기술적 보존전략을 계획하고 수행해야 한다. 특히 각 대학들은

디지털 큐레이션을 하기 위해 상호 협력을 위한 학문별 데이터의 보존, 제공 및 기술 변화에 적합한 저장매체를 마이그레이션하여 사용자들이 접근 및 이용할 수 있도록 해야 한다.

가. 마이그레이션(Migration)

영국의 국립 보존 사무국(National Preservation Office)의 보고서²⁵⁾에서는 마이그레이션 전략이 모든 유형의 디지털 정보자원을 보존하는데 적합한 방식이라고 소개하였다.

마이그레이션은 하나의 하드웨어, 소프트웨어 환경에서 다른 하드웨어, 소프트웨어 환경으로 디지털 정보자원을 주기적으로 이전하는 것을 말한다. 이를테면 시스템 혹은 매체가 시간이 경과하여 노후화되거나 가치가 저하되어도 지속적으로 정보접근을 보장하기 위하여 또 다른 정보시스템 또는 저장매체로 데이터를 변환하는 과정이며, 물리적 매체의 노후화 문제를 해결할 수 있다.

예를 들면 네덜란드의 DPT(Digital Preservation Testbed)는 보존 전략과 관련하여 텍스트, 데이터베이스, 스프레드시트 등 다양한 디지털 정보자원을 대상으로 소급호환성유지, 상호운용성유지, 표준으로 변환 등의 마이그레이션 전략으로 디지털 정보자원을 큐레이션하여 영구보존하고 있다.

나. 에뮬레이션(Emulation)

디지털 정보자원이 생산당시의 본래기능이 그대로 재현되도록 생산시점과 동일한 운영환경으로 재생산하는 것이 에뮬레이션인데, 하나의 컴퓨터(호스트 시스템)에서 가동되면서 이를 통해 다른 컴퓨터(타깃 시스템)에서 실제적으로 재현할 수 있는 에뮬레이터란 소프트웨어 및 하드웨어가 필요하다.²⁶⁾

예를 들면 디지털 정보자원의 생산시점에서 원본형식으로 비스트림을 저장하고, 정보를 재현할 시점에 생산당시에 사용된 소프트웨어와 이를 가동할 수 있는 본래의 컴퓨터 플랫폼 환경을 재생산하여 비스트림을 재현하는 것이다.

다. UVC(Universal Virtual Computer)

IBM의 Raymond Lorie에 의해 제안된 UVC는 가상으로 존재하는 단순 아키텍처 컴퓨터의 형태로, 미래의 소프트웨어 개발자가 UVC를 위한 에뮬레이터를 작성할 수 있도록 UVC 데이터 포맷 디코더가 가동되는 기반이다. 이러한 변환 프로그램은 생산시점에 작성되어야 하며, 데이터 자체의 보존과 프로그램 행위의 보존을 서로 분리시켜 미래에 보다 이해하기 쉬운 포맷으로 데이터 원본을 디코딩 하는 것이다.²⁷⁾ 이는 생산시점에 작성한 UVC 에뮬레이터가 미래에 사용하는 컴퓨

25) Digital Preservation Testbed, Migration : Context and Current Status, 2001,

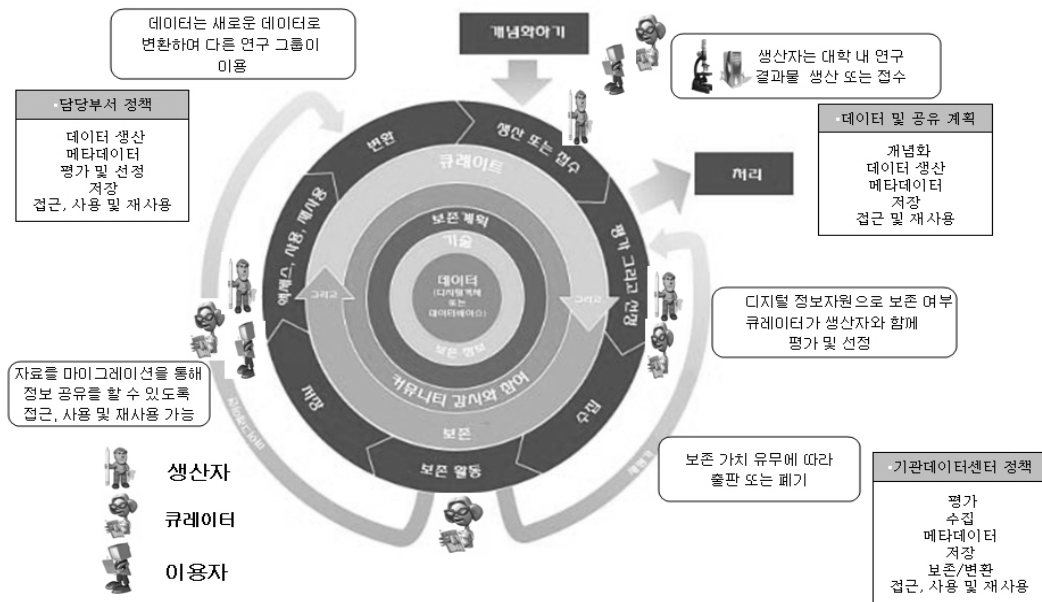
〈<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf>〉 [cited 2009. 11. 2].

26) National Archive of Australia, Digital Recordkeeping : Guidelines for Creating, Managing and Preserving Digital Records[Exposure Draft], 2004.

터에서 구동되는 UVC 에뮬레이터에 의존하며, 마이그레이션 전략요소도 포함한다.

3. 디지털 큐레이션 워크플로우

디지털 큐레이션 생명주기 모형에는 생산자, 큐레이터, 이용자 사이에 긴밀한 협조가 요구되는데, 각각의 활동에 따른 역할관계는 <그림 3>과 같다.



<그림 3> 생산자, 큐레이터, 이용자와 디지털 큐레이션 정책과의 관계 모형

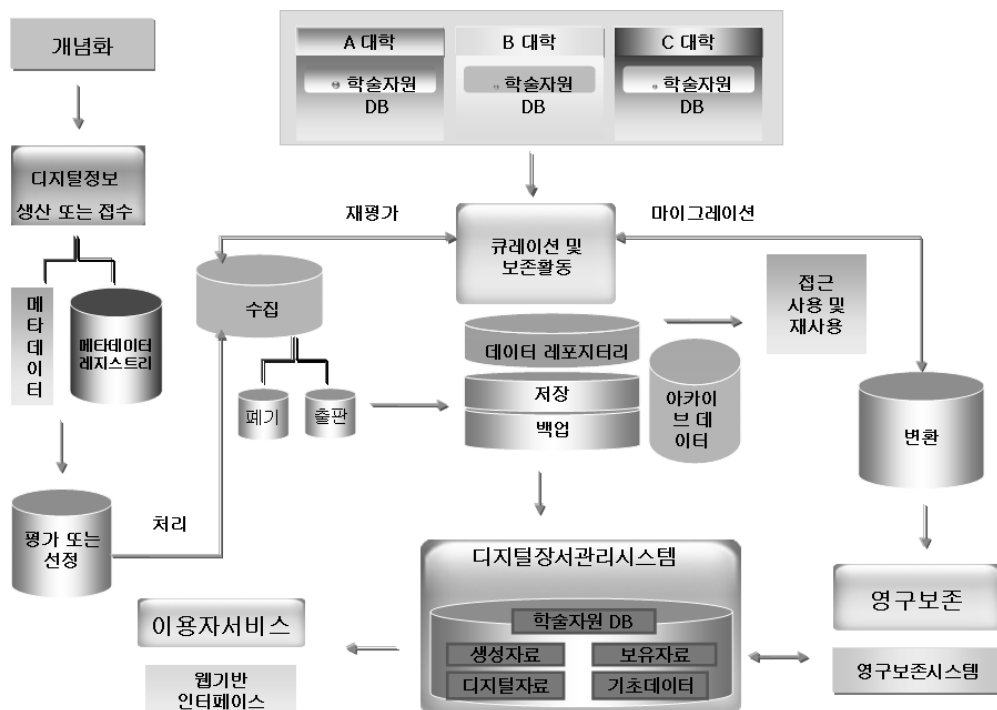
디지털 큐레이션 정책은 데이터관리 및 공유계획, 담당부서 정책, 기관데이터센터 정책에 따라 생명주기 워크플로우가 달라진다. 또한 각 생명주기별로 생산자, 큐레이터, 이용자가 개입하는 상황도 달라진다.

생산자는 디지털 연구 결과물을 생산 또는 접수하고, 평가 및 선정과 접근, 사용 및 재사용하는 역할을 담당한다. 큐레이터는 개념화, 평가, 수집, 보존활동, 접근, 사용 및 재사용하는 역할을 수행하며, 이용자는 개념화, 접근, 사용 및 재사용을 결정할 때 개입한다.

아직은 국내 대학도서관의 디지털 현황이 디지털 정보자원의 큐레이션에 대한 환경은 마련되지

27) Raymond Lorie, *The UVC : A Method for Preserving Digital Documents IBM/KB Long-Term Preservation Study*, 2002, pp.5-6, <http://www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf> [cited 2009. 11. 02].

않았지만 디지털 자원에 대한 보존, 아카이브 및 큐레이션에 대한 중요성은 제기된 상태다.²⁸⁾ 그러므로 앞으로 대학도서관이 나아갈 방향을 제시하기 위하여 영국의 DCC에서 제시한 디지털 큐레이션 모형을 기반으로 각 대학 도서관에서 디지털 큐레이션 대상 자료들을 서로 협력하고, 기관간의 레포지토리를 구축하여 공유할 수 있도록 <그림 4>와 같이 워크플로우를 설계하였다.



<그림 4> 대학 간 협력 디지털 큐레이션 워크플로우

디지털 정보자원을 공유하기 위해서는 각 대학 간 긴밀한 협력 체제를 구축하는 것이 선행되어야 한다. 또한 각 기관별 생명주기 기반 디지털 큐레이션을 위한 정책 및 전략이 우선적으로 협의되어 기관 레포지터리로서 대학도서관이 디지털 큐레이션 역할을 수행해야 한다.

대학도서관에서는 대학 내 연구기관 및 산학협력을 통해 완성된 연구결과물인 디지털 정보자원을 큐레이션하기 위해 저장방법과 캡처방식 등의 계획을 세우고, 연구 데이터를 생산 및 접수하고, 수집된 정보자원들의 메타데이터를 추출하여 미래에 접근 가능한 기술, 관리, 구조적 메타데이터를 통해 메타데이터 레지스트리에서 관리한다. 보존할 가치가 있는 정보자원에 대한 보존 계획을 세우고, 지속적으로 유지될 수 있는 가치가 있는지를 평가 또는 선정하는 처리과정을 거쳐 수집한다.

28) 황혜경 외, 전계논문.

수집된 자료들은 출판하여 데이터 레포지터리에 큐레이션 및 보존활동을 거쳐 디지털 정보자원을 기관의 레포지터리 소프트웨어에 저장, 백업하고 이를 접근, 사용 및 재사용할 수 있도록 마이그레이션 과정을 거쳐 변환 및 통합한다. 이를 다른 연구그룹이 이용할 수 있도록 영구보존시스템에 보존한다. 이렇게 큐레이트 된 자료들은 웹기반 레포지터리 인터페이스를 통하여 검색되어 자유롭게 이용된다. 큐레이트 된 정보수준에 따라 이용범위는 제한될 수 있다.

이상과 같은 생명주기를 기반으로 한 생산자, 큐레이터, 이용자의 디지털 큐레이션 관계 모형 및 워크플로우는 보다 완결성을 갖춘 디지털장서관리시스템으로서 이용자에게 우수한 서비스를 제공하는 수단이 될 것이다.

V. 결론 및 제언

국내에서는 디지털 정보자원의 중요성과 필요성에 대한 인식이 대두되고 있지만 아직까지 보존, 관리 및 재사용을 위한 공동 활용체제가 제대로 구축되어 있지 않아 디지털 큐레이션 전략이 절실히 요구된다.

대학도서관 서비스의 디지털 큐레이션을 위해서는 각 학문분야별 유형에 따른 전략과 대학당국의 이해관계자들의 협력네트워크가 구축되어야 하며, 대학도서관을 중심으로 디지털 큐레이션이 수행되어야 한다. 이를 위하여 디지털 큐레이션 생명주기 모델을 기반으로 대학 상호간 긴밀한 협력 체제를 구축하는 디지털 큐레이션 워크플로우를 구상하였다.

각 생명주기를 기반으로 생산자가 연구데이터의 결과물이 생산되면 평가, 선정 등의 처리과정을 거쳐 수집, 보존 및 큐레이션 절차를 거쳐 변환되어 재사용 및 영구 보존되는 시스템이다. 예를 들면 대학 내에서 생산된 연구 결과물을 중심으로 기관 레포지터리를 구축하여 학술데이터를 공유하고 재사용하여 자산으로 관리하는데 의의가 있다.

앞으로 디지털 큐레이션 기술의 방향은 과거를 대상으로 한 디지털 큐레이션 복원기술뿐만 아니라 미래에도 유지해야할 디지털 큐레이션 보존기술로도 확장되어야 한다. 또한 디지털 큐레이션을 위한 전문 인력을 확보하기 위하여 디지털 보존교육 프로그램을 개발하고, 전문교육 프로그램을 통하여 자격 요건을 갖춘 인력이 배출되는 디지털 큐레이터 양성도 시급하다.

〈참고문헌은 각주로 대신함〉