

연구콘텐츠 라이프 사이클 모델 개발: 한국과학기술정보연구원 콘텐츠 큐레이션 센터를 중심으로*

Developing the Research Contents Life Cycle Model: Based on the Curation Model for KISTI Curation Center

이혜원 (Hyewon Lee)** , 윤소영 (Soyoung Yoon)***
박지영 (Ziyoung Park)**** , 황혜경 (Hyekyong Hwang)*****
김재훈 (Jayhoon Kim)***** , 이혜림 (Hea Lim Rhee)*****

초 록

본 연구에서는 KISTI 콘텐츠 큐레이션 센터의 업무를 중심으로 콘텐츠 라이프 사이클에 맞는 큐레이션 모델인 KISTI CLCM을 개발하였다. 이를 위해 앞서 개발된 모델 DCC, DCC&U, UC3, DCN 모델 등을 살펴보고, 디지털 큐레이션 센터의 업무와 연결되는 모델들의 장점을 도출하였다. KISTI 콘텐츠 큐레이션 센터의 업무담당자들과의 집단 면담을 통해 센터에서 다루는 콘텐츠의 특징을 파악하고 현 업무의 성취정도를 확인하였다. 또한 본 연구에서는 UC3 모델을 기반으로 KISTI 콘텐츠 큐레이션 웨이브를 개발하여 이용자 중심의 서비스를 발굴하고 디지털 자원에 새로운 가치를 추가하여 활용성을 높일 수 있는 방안을 제시하였다.

ABSTRACT

This study developed KISTI CLCM, a curation model suitable for the content life cycle, focusing on the KISTI Content Curation Center. For developing the KISTI Curation Model, the previously developed models DCC, DCC&U, UC3 and DCN models were reviewed and the advantages of the models connected to the mission and tasks of the digital curation center were derived. Interviews with staffs of the KISTI Content Curation Center identified the characteristics of the content and identified the level of achievement of the current task. In addition, this study proposed to enhance usability by developing KISTI content curation waves based on the UC3 model to discover user-centered services and add new values to digital resources.

키워드: 콘텐츠 라이프 사이클, 콘텐츠 큐레이션 모델, 콘텐츠 큐레이션 웨이브, 한국과학기술정보연구원
contents life cycle, contents curation model, contents curation waves, Korea Institute of
Science and Technology Information (KISTI)

* 본 연구는 2018년도 한국과학기술정보연구원(KISTI) 주요사업 과제(KISTI 큐레이션 라이프 사이클 모델 개발; 연구과제번호 K-18-IR-11-01R-1)로 수행한 것임.

** 서울여자대학교 사회과학대학 문헌정보학과 부교수(hwlee@swu.ac.kr) (제1저자)

*** 한성대학교 크리에이티브인문예술대학 디지털인문정보학트랙 강사(corba99@gmail.com) (공동저자)

**** 한성대학교 크리에이티브인문예술대학 도서관정보문화트랙 부교수(zgpark@hansung.ac.kr) (공동저자)

***** 한국과학기술정보연구원 콘텐츠큐레이션센터장 책임연구원(hkhwang@kisti.re.kr) (공동저자)

***** 한국과학기술정보연구원 콘텐츠큐레이션센터 선임연구원(jay.kim@kisti.re.kr) (공동저자)

***** 한국과학기술정보연구원 콘텐츠큐레이션센터 선임연구원(rhee.healim@kisti.re.kr) (공동저자)

■ 논문접수일자: 2019년 8월 22일 ■ 최초심사일자: 2019년 9월 24일 ■ 게재확정일자: 2019년 9월 27일

■ 정보관리학회지, 36(3), 203-228, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.3.203>

1. 서론

정부연구개발예산은 2018년 기준 196,681억 원이며 그 중 과학기술분야는 60,920억원에 달한다. 연구개발을 통해 발생하는 연구보고서 및 학술지 논문에 대한 관리는 체계적으로 진행되는데 반해 이러한 결과물을 잘 활용할 수 있는 정보체계 구축은 한번 더 살펴볼 필요가 있다. 연구개발의 결과물인 '콘텐츠' 그리고 콘텐츠를 구성하는 일부 또는 콘텐츠 자체인 '데이터', 이러한 콘텐츠와 데이터에 대한 기술의 수준이나 표현정보의 품질과 양은 연구결과의 확산에 큰 영향을 준다. 이와 같은 영향력으로 인해 우리나라뿐만 아니라 많은 국가의 연구기관들이 콘텐츠 및 데이터 관리체계의 발전에 많은 시간과 노력을 투자하고 있다. 콘텐츠의 기술 및 표현정보를 확보하기 위해 자연스럽게 콘텐츠의 라이프 사이클에 집중하고 관리체계 재구조화를 위해 큐레이션 모델을 개발하고 활용하고 있다. 이에 반해 우리나라는 콘텐츠의 관리와 서비스 관점에서 도출된 큐레이션 모델에 대한 연구가 미흡하고 특히 특정 주제 영역이나 기관에서 선언된 큐레이션 모델을 확인하는 것이 쉽지는 않다.

본 연구에서는 국내 최대 과학기술 콘텐츠를 보유하고 서비스하는 KISTI의 사명을 반영한 콘텐츠 큐레이션 모델을 개발하여 국가 차원의 고부가가치 정보인프라를 구축하였다. 이를 위해 콘텐츠의 생성, 관리, 보존, 이용 등 전반적인 라이프 사이클을 제안한 모형을 분석하여 시사점을 도출하고 KISTI 과학기술 콘텐츠 인프라 환경을 분석하였다. KISTI 과학기술 콘텐츠 인프라 환경 분석을 위해 KISTI 콘텐츠 큐레이션 센터의 업무를 파악하였는데 첫 번째

로 국내외 학술논문, 국가 R&D 연구보고서 등 주요 콘텐츠 특성을 분석하고, 두 번째로 콘텐츠별 라이프 사이클을 규명하고, 세 번째로 콘텐츠별 라이프 사이클 각 단계에서의 담당 부서와 업무를 파악하고, 부서 간 협업 관계를 조사하였다. 또한 디지털 큐레이션 모형인 영국 디지털 큐레이션 센터(Digital Curation Centre, 이하 DCC라고 함)의 큐레이션 라이프 사이클 모델(Curation Lifecycle Model, 이하 CLM이라고 함)의 체크리스트를 활용하여 순차적인 활동의 성취 정도를 알아보았다.

최종적으로 본 연구는 KISTI 콘텐츠 큐레이션 라이프 사이클 모델 개발과 동시에 모델의 지속성을 보장하고 이용자 서비스에 초점을 둔 KISTI 콘텐츠 큐레이션 웨이브를 개발하였다. 먼저, KISTI 콘텐츠 큐레이션 라이프 사이클 모델은 전 세계적으로 통용되는 큐레이션 모델들을 분석하고 KISTI 콘텐츠 큐레이션 센터의 업무에 맞는 특징을 도출하여 개발하였다. 콘텐츠의 생애 주기에 맞게 핵심 콘텐츠, 전체 라이프 사이클 활동, 순차적인 활동, 비정기적인 활동 등을 선언하였다. KISTI 콘텐츠 큐레이션 웨이브는 큐레이션 모델이 추구하는 이용자 중심의 서비스를 발굴하고 디지털 자원에 새로운 가치를 추가하여 활용성을 높일 수 있는 방안으로 제안하였다.

2. 디지털 큐레이션 모델 분석

2.1 디지털 큐레이션의 필요성 및 적용 범위

디지털 큐레이션은 디지털 정보자원의 진본

성을 확보하고, 미래의 이용자가 재사용할 수 있도록 관리하기 위한 것이다(Pennock, 2007, p. 1). 또한 디지털 큐레이션은 현재의 이용과 원래의 정보 가치를 유지하기 위한 보존 측면과 미래의 재사용을 위한 부가가치의 창출 측면이 모두 포함되어야 하며, 컴퓨터 과학, 기록학, 도서관학, 정보학 등 여러 분야의 협력이 필요한 학제적 영역이다(Constantopoulos, Dallas, Androutsopoulos, Angelis, Deligiannakis, Gavrilis, Kotidis, & Papatheodorou, 2009, p. 37).

디지털 큐레이션이 필요한 현장도 인문학부터 자연과학, 공학에 이르기까지 매우 다양하며, 큐레이션을 통해 제공되는 정보 또한 디지털 리포지터리나 박물관, 도서관, 기록관 등 어느 곳에도 존재할 수 있는 것들이다. 그러므로 디지털 큐레이션 모델은 특정 도메인을 대상으로 중점적으로 개발되거나 전반적인 큐레이션 영역을 대상으로 포괄적으로 개발될 수 있다(Constantopoulos et al., 2009).

본 연구에서는 연구데이터 큐레이션 모델을 살펴보았으며, 전반적인 영역을 대상으로 하는 모형과 특정 도메인인 대학을 중심으로 한 데이터 관리 모형을 분석하였다.

2.2 전반적인 영역을 대상으로 하는 디지털 큐레이션 모델

전반적인 영역을 대상으로 하는 디지털 큐레이션 모델은 라이프 사이클 기반과 컨티뉴엄 기반으로 나눌 수 있다. 먼저, 라이프 사이클 기반 디지털 큐레이션 모델은 디지털 정보가 생산, 배포, 활용, 보존되는 것을 유기체와 같은 생애주기로 보는 프레임으로서, 영국을 중심으

로 확산되었다. 대표적인 라이프 사이클 기반 디지털 큐레이션 모델은 DCC CLM이다. 디지털 큐레이션과 구분되는 디지털 아카이빙 모델에는 대표적으로 OAIS(Open Archival Information System) 참조모형이 있으며, 이 모형 또한 라이프 사이클 기반 접근방식을 채택하고 있다.

2.2.1 DCC CLM을 기반으로 DCU와 DCC&U 모델

DCC CLM은 상위 수준의 개념모형으로 개별 기관이나 협력기관 간에 이루어지는 반복적인 큐레이션 라이프 사이클에 맞춘 단계별 요건을 제시하고 있다. 또한 추상적이고 포괄적인 모델이므로 실제 적용 단계에서 여러 기관이 참여할 경우에는 개별 기관마다 자관이 분담하는 역할에 따라 특정 단계에만 참여할 수도 있다. 즉 특정 기관에 적용할 경우, 필요에 따라 부가적인 단계나 요건을 추가하는 것이 가능하다(DCC Homepage a, n.d).

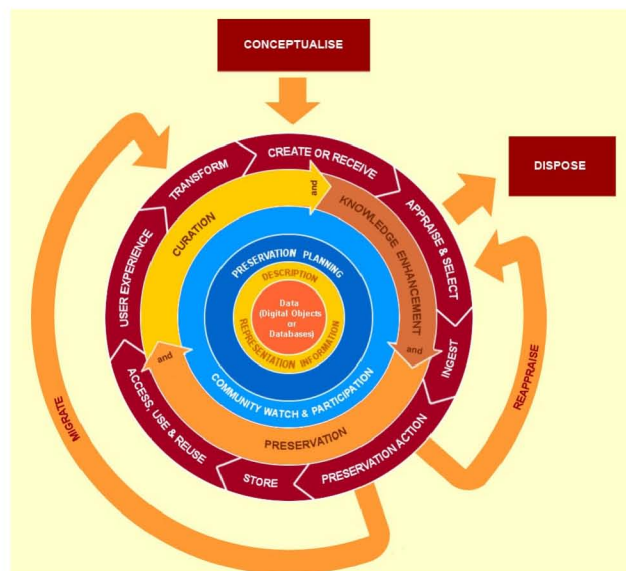
DCC에서는 CLM의 순차적 활동에 대한 체크리스트를 제시하였는데, 이는 큐레이션 계획과 평가에 활용할 수 있다. DCC 모델은 큐레이션에 필요한 계층적 활동과 순차적 활동을 연계한 전반적인 개념 틀을 제공함으로써, 생산자와 큐레이터, 데이터를 재사용하려는 이용자들이 큰 그림 속에서 자신의 역할과 책임을 식별할 수 있도록 지원한다. DCC 모델을 실제로 적용함으로써 특정 단계를 추가하거나 이해당사자의 역할을 연계하는데 결여된 부분을 발견하고 보완할 수도 있는 강점을 가지고 있다(DCC Homepage b, n.d).

DCC CLM을 기반으로 한 DCC&U 모델은 본 연구의 콘텐츠 표현을 강화할 수 있는 방안

을 제시하였다. DCC&U는 영국 디지털 큐레이션 센터의 DCC 모델과 아테네의 연구센터(Athena Research Centre)에서 제안한 DCU(Digital Curation Unit) 모델의 퓨전 버전이다(Constantopoulos et al., 2009). 아테네의 연구센터에서 DCU를 발전시켜 DCC&U를 제안한 것이다. DCU는 주로 문화유산 분야를 중심으로 디지털 휴머니티 측면에서 개발되어 맥락 관리나 문화유산 도메인 내의 특징을 강조한 모형이며, 아테네 연구센터에서는 도메인 중심의 DCU를 보편적인 관점에서 확장시키기 위해 DCC의 모델과 연계하고자 하였다. 연구팀은 DCU 모델을 절차 중심의 접근방식(process-oriented approach)으로 규정하고, 2가지 모델을 통합(harmonization)하는 과정에서 DCC 모델의 핵심인 라이프 사이클에 DCU 모형을 분석하여 적용하였다. 통합 과정을 거친 후 DCU 모

델은 DCC 모델에는 포함되어 있지 않았던 주요 큐레이션 활동을 도출하고 기존 모델을 향상시켰다는 장점이 부각되었지만 반면 향상된 기능을 적용하기 위해 모델이 더 복잡해지고 추가적인 구현 조건을 추가해야 한다는 약점도 드러났다.

연구팀에서 제안한 DCC&U 모델은 <그림 1>과 같으며, DCC 모델에 이용자 경험과 전거, 시맨틱웹 기술 등을 부가한 특징을 지니고 있다. 수정된 부분으로는 기존의 큐레이션과 보존 활동에 '지식 향상(knowledge enhancement)'을 부가하고, 정보 기술과 표현 활동에 '전거'를 추가하였다. 또한 순차적인 활동에는 '이용자 경험(user experience)'을 '접근 및 사용/재사용(access, use & reuse)'과 '변환(transform)' 사이에 추가하였다.

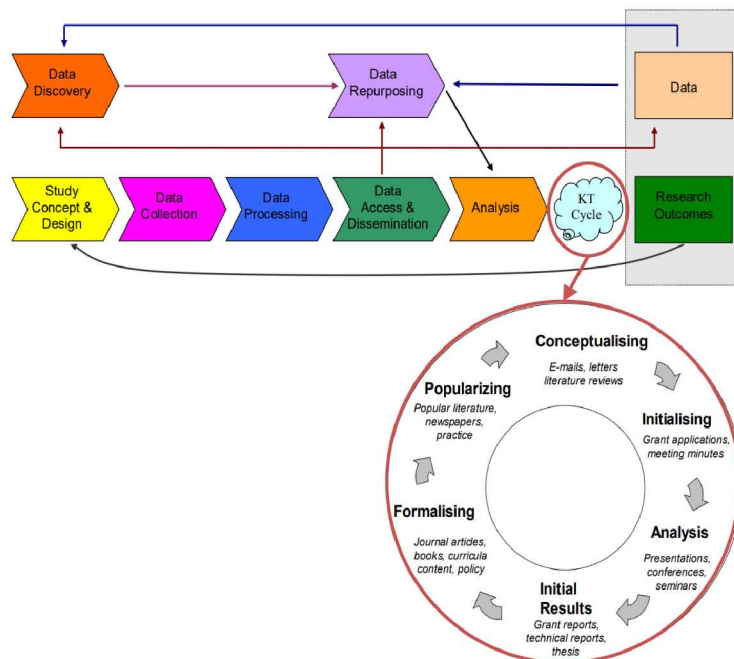


<그림 1> DCC&U 모델 개요
(Constantopoulos et al., 2009)

2.2.2 Lifecycle of Research Knowledge Creation 모델

Lifecycle of Research Knowledge Creation 모델은 라이프 사이클을 바탕으로 정보 처리 과정을 도식화한 모형으로서 데이터 생성 과정을 순차적으로 이해하고 각 과정 간의 연계성을 파악할 수 있도록 지원한 것이다(Humphrey & Hamilton, 2004; Humphrey, 2006; Oliver & Harvey, 2016). <그림 2>는 연구 지식 생산에 대한 라이프 사이클 모델을 나타낸 것으로, 상단의 왼쪽에 위치한 연구결과와 관련 데이터 박스는 문서화되며 이는 데이터와 연구결과 박스를 함께 둘러싼 회색 박스를 통해 표현된다. <그림 2>에서 화살표 모양으로 표현되는 연구과정의 각 단계는 연계되지 않고 단절될 수 있

다. 예를 들어, 데이터 수집에서 데이터 처리가 다른 담당자에 의해 이루어질 경우 각 단계의 결과물은 해당 담당자의 하드 드라이브에 개별적으로 저장될 수 있는 것이다. 그러므로 각 단계를 연결할 수 있는 큐레이션이 필요하다. 또한 상단에 제시된 모델에는 지식 전송 사이클(knowledge transfer cycle)이 포함되어 있는데, 이는 연구자들의 연구결과를 공유하는 다양한 방식을 나타낸 것이다. 초반에는 이메일이나 문헌 리뷰 등의 방법을 통해 공유하며, 점차 연구지원 신청이나 회의록에서 발표, 세미나로 발전하게 된다. 1차 결과가 도출되면 연구보고서, 논문 등으로 공유되며, 이후 학술논문이나 단행본으로 구체화되고, 신문이나 실무 적용 등을 통해 확산되는 과정을 보여주고 있다.



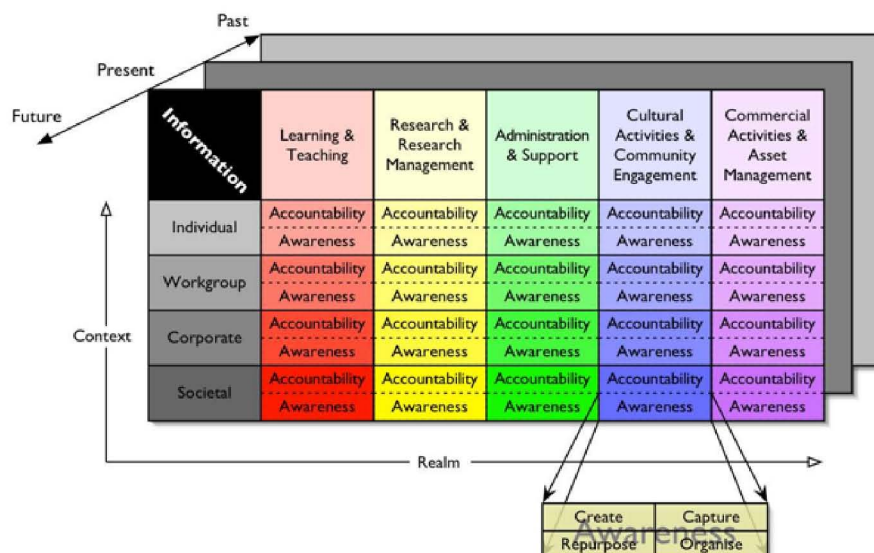
<그림 2> 연구지식 생산 라이프 사이클 모델(위) 및 지식 전송 사이클 내에서의 학술 커뮤니케이션(아래) (Humphrey, 2006)

컨티뉴엄 기반 디지털 큐레이션 모델은 디지털 객체의 특징을 중점적으로 분석하여 호주에서 제안된 접근방식으로 라이프 사이클 모델에 비해 덜 확산되었으나 향후 라이프 사이클 기반 방식에 수용되어 상호 발전될 수 있을 것이다.

2.2.3 Data Curation Continuum

Data Curation Continuum은 호주의 기록관리 분야의 레코드 컨티뉴엄 개념이 디지털 큐레이션 분야에 적용된 것이다(Oliver & Harvey, 2016, pp. 36-37). 레코드 컨티뉴엄 개념은 사회학자인 기든스(A. Giddens)의 구조화이론(Structuration Theory)을 바탕으로 1996년에 제안되었다(Upward, 2005). 컨티뉴엄 이론은 라이프 사이클 이론에 비해 덜 알려져 있는데, 초기에는 호주 Monash 대학에서 학생들의 다양한 정보요구를 관리하고 학습에 활용한 도구로 사용되었기 때문이다(Oliver, 2010). 컨티뉴엄

이론은 라이프 사이클의 단선적인 구조에 대한 대안으로 간주되고 있다(Treloar, Groenewegen, & Harboe-Ree, 2007). 그러나 라이프 사이클 모형은 이론상으로는 명확하지만, 다양한 유형의 정보와 생산주체 등이 존재하는 환경에서는 실제 구현하는데 어려움이 많았기 때문에 컨티뉴엄 이론을 기반으로 이를 극복할 수 있는지를 고민하였다. 컨티뉴엄 이론은 Monash 대학과 정부 지원 프로젝트가 결합하여 제시된 결과물이다. 프로젝트의 시작은 연구, 교육, 행정을 아우르는 기관 리포지터리를 호주 내외의 많은 대학과 연합하여 개발하는 과정에서 단일 구조의 리포지터리에 다양한 데이터를 저장하고 활용하기 어렵다는 것이었다. 다축 분석(multiple-axis analysis)과 같이 다양한 기준으로 데이터를 분석하는 정보 이론에서 정보를 <그림 3>과 같이 다차원 관점(multi-dimensional view)으로 접근하는 정보 컨티뉴엄을 도출하였다.



<그림 3> 정보관리의 다양한 차원

출처: <http://www.dlib.org/dlib/september07/treloar/treloar-fig1.png>

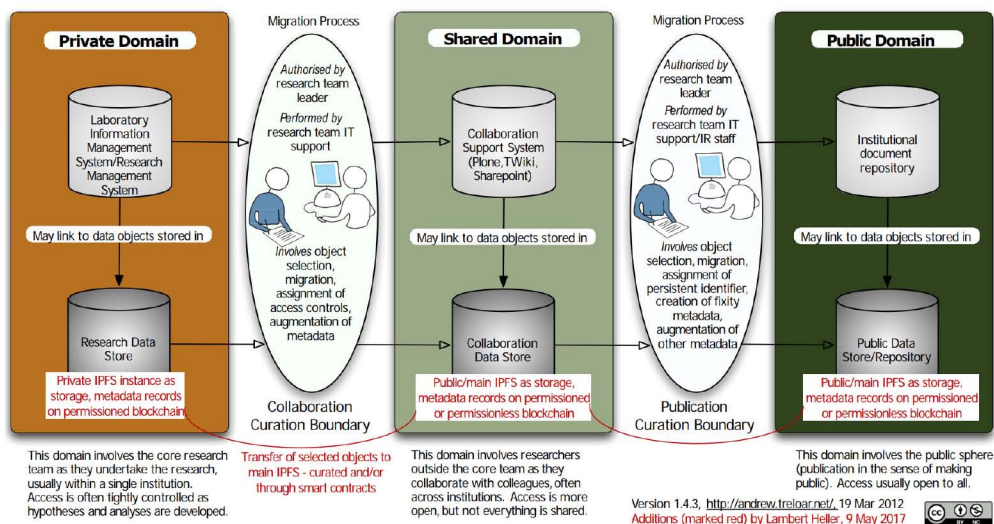
〈그림 3〉은 데이터의 타입을 x축(Realm)으로, 생산단위를 y축(Context), 생산시기를 z축(Past-Present-Future)으로 구분하였으며, 각 셀의 정보를 공개용(Awareness)과 설명책임용(Accountability)으로 다시 나누었다. 그리고 각 셀에 속한 정보는 생산-획득-조직-가치 재설정 단계를 거치게 된다. 큐레이션 컨티뉴엄은 라이프 사이클의 단위를 다차원으로 분할하여 구조화한 것으로 볼 수 있겠다. 〈그림 4〉는 Australian National Data Service(2016)에서 제시한 데이터 큐레이션 컨티뉴엄의 3가지의 활용 도메인인 ‘개인 연구’, ‘공동 연구’, 그리고 ‘출판물’을 중심으로 도메인 간의 사이클 ‘경계 전환 과정’으로 설명한 것이다(Groenewegen & Hahnel, 2014).

〈그림 4〉의 개인 도메인에서는 정보가 연구 관리시스템에 저장되며 연구데이터 저장소와 연계되고, 특정 프로젝트의 ‘핵심 연구팀’도 여

기에 속한다. 그러나 외부에서는 접근이 통제되는 경우가 많다. 개인 도메인의 데이터는 ‘협업’을 통해 공유 도메인으로 이전되는데 복수의 기관에 소속된 연구자들 간의 데이터가 통합될 수 있으며 위키 서비스와 같이 공유 가능한 저장소에 데이터가 저장된다. 공유 도메인의 데이터는 ‘출판’을 통해 공공 도메인으로 이전되는데 공공 도메인에 저장된 데이터는 연구팀뿐 아니라 모든 공공 영역에서 접근할 수 있도록 개방된다.

2.3 대학에서의 데이터 관리를 위한 디지털 큐레이션 모델

대학 중심의 디지털 큐레이션 모델은 캘리포니아 디지털 도서관(California Digital Library; 이하 CDL이라고 함)의 University of California Curation Center(UC3)와 University of Minnesota



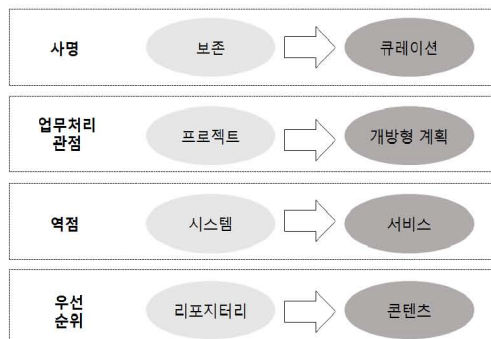
〈그림 4〉 연구데이터의 도메인과 공유 정도의 전환 과정
(Australian National Data Service, 2016; Heller, 2017)

가 주축이 되어 구축한 The Data Curation Network(DCN)를 살펴보았다.

2.3.1 University of California Curation Center(UC3) 마이크로 서비스 모델

University of California는 교수·학습 과정에서 생산되는 다양한 디지털 자산을 대학의 지적 자원이라고 간주하고 그 가치를 보존하여 미래에도 사용이 가능한 시스템을 구현하고자 하였다. UC3는 디지털 큐레이션의 관점에서 대학의 디지털 자산을 장기적으로 (재)사용할 수 있도록 목표를 설정하고, 정책 및 실무지침을 개발하였다(이혜원, 2016, p. 240; CDL, 2010).

〈그림 5〉는 UC3 모델을 설명한 것이다(CDL, 2010).



〈그림 5〉 UC3 모델의 특징
(CDL, 2010, pp. 1-2 재구성)

〈그림 5〉에 나타난 UC3의 사명은 보존보다는 큐레이션으로 무게중심을 옮겼다. UC의 장기적인 이용가능성을 보장하기 위해 보존보다는 더 큰 개념인 큐레이션이라는 용어로 표현하여 디지털 자산의 보존, 접근, 관리 활동을 통합적으로 관리하겠다는 의지를 담고 있다. 디

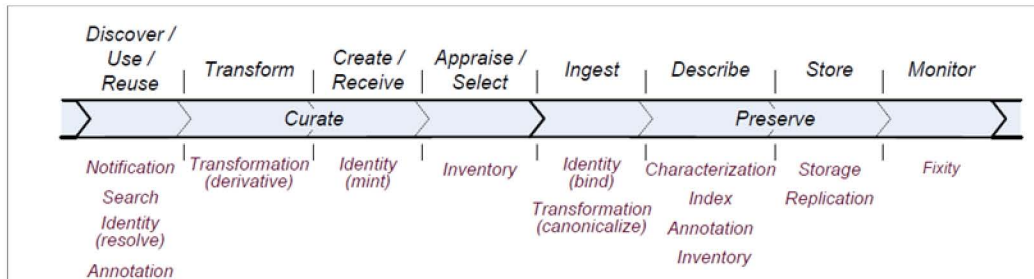
지탈 큐레이션은 시간이 지남에 따라 신뢰할 수 있는 디지털 콘텐츠의 가치를 유지하고 추가하는 것을 목표로 하는 정책 및 관행의 집합인 동시에, 관리 자산을 보존하고 액세스하는 조정된 활동의 필요성을 강조한다. UC3의 모든 활동은 전체 디지털 라이프 사이클과 관련이 있으며 이상적으로는 자산이 처음 생성되어 최종 처분될 때까지 지속된다.

〈그림 5〉에 나타난 UC3의 업무처리 관점은 보존프로젝트보다는 계획에 중심을 두고 있다. 즉 시간에 집중하는 프로젝트보다는 프로그램 자체에 집중하는 개방형 계획 중심의 업무처리 관점(programmatic approach)으로 전환하였다.

〈그림 5〉에 나타난 UC3의 업무처리 역점은 시스템보다는 서비스로 전환되었다. 기술 발전에 따른 정보시스템의 짧은 수명에 집중하기보다는 필수 기능을 지속적으로 제공하고 필요에 따라 기능을 쉽게 추가하고 구현할 수 있는 서비스에 집중하고자 하였다.

〈그림 5〉에 나타난 UC3 업무처리의 우선순위는 장소의 개념인 큐레이션 리포지터리 중심성을 벗어나는 것부터 시작하였다. 큐레이션된 결과물은 개념적으로 단일 시스템에 위치하기보다는, 결과물로 생성되기 위해 중앙장치로 옮겨져야 한다는 전제 없이 콘텐츠가 저장된 원위치에서 소규모로, 자급형으로, 그리고 느슨하게 결합된 방식의 분산형 서비스의 산물이어야 한다.

UC3에서는 독립적이지만 상호운영적인 마이크로 서비스(micro-service)가 가능하고 동시에 복잡한 큐레이션 기능이 전략적으로 세분화된 서비스 모형을 제안하였다. 〈그림 6〉에서



〈그림 6〉 UC3 마이크로 서비스 모델(CDL, 2010)

큐레이션과 보존 위에 있는 활동들은 큐레이션 생명주기에 맞는 서비스를 설명한 것이며, 아래는 UC3에서 제공하고자 하는 서비스들을 나열한 것이다(CDL, 2010, p. 17).

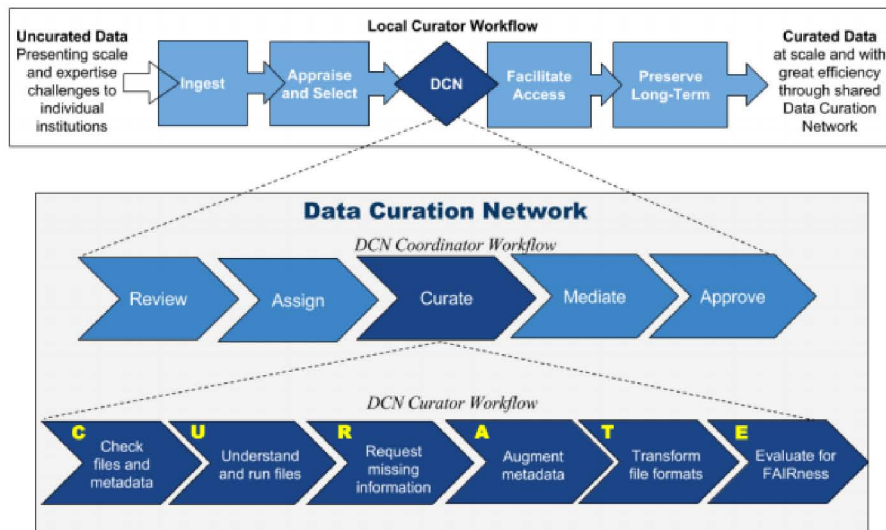
UC3 마이크로 서비스는 디지털 큐레이션은 장기적인 관점에서 신뢰할 수 있는 디지털 콘텐츠에 대한 관리 및 가치 부가를 목표로 식별(identity), 저장(storage), 고정성(fixity), 복제(replication), 목록(catalog), 특성화(characterization), 입수(ingest), 색인(index), 검색 서비스(search), 변환(transformation), 출판(publication), 주석(annotation) 등 12개의 서비스를 제공자 하였다(CDL, 2010, pp. 16-21).

2.3.2 The Data Curation Network (DCN) 모델

최신 데이터 큐레이터 커뮤니티를 위한 전문가 서비스, 협업 인센티브, 표준화된 큐레이터 사례 및 전문 개발 교육을 제공하는 로컬 데이터 리포지터리 단계에서 'human layer'를 강조(Johnson, Carlson, Hudson-Vitale, Imker, Kozlowski, Olendorf, & Stewart, 2017, p. 3)하는데, 이를 대표하는 것이 DCN 모델이다. DCN 참여 기관은 University of Minnesota을 중심으로 University of Michigan, Washington

University in St. Louis, University of Illinois at Urbana Champaign, 그리고 Cornell University, Pennsylvania State University 등이다. Data Curation Network 모델은 다수의 학문적 데이터세트를 찾기 쉽고, 접근·상호운용·재사용이 가능하도록 하며, 데이터 큐레이션 서비스를 집합적으로 제공하는 기관의 전문성을 잘 활용할 수 있도록 설계되었다. 이를 기반으로 한 DCN의 큐레이션 워크플로우는 〈그림 7〉과 같으며 DCN의 업무 흐름을 보여주고 있다(Johnson et al., 2017, p. 29).

DCN의 장점은 기존 데이터 큐레이션 서비스가 있는 대학도서관, 데이터 서비스를 위한 인력이 충분하지 않은 대학도서관, 특정 전공 또는 주제 리포지터리에서도 활용될 수 있다는 것이다(Johnson et al., 2017, p. 29-30). 먼저 기존 데이터 큐레이션 서비스가 있는 대학도서관에서는, 첫 번째로 로컬에서 사용할 수 있는 것보다 더 많은 학문분야 및 형식에 대한 데이터 큐레이션의 전문 지식에 액세스할 수 있다는 것이고, 두 번째로 데이터 큐레이션 실무에 있어 규모가 더 큰 에코시스템에 참여할 수 있으며, 세 번째로 공동 표준 개발에 참여할 수 있고, 네 번째로 데이터 큐레이터 교육 및 전문 데이터 큐레이션 관행을 확립하기 위한 파이프



〈그림 7〉 DCN 큐레이션 워크플로우(Johnson et al., 2017)

라인을 구축할 수 있다는 것이다. 다섯 번째로는 로컬 큐레이션 서비스의 개발 및 발전에 기여할 수 있고, 마지막으로는 직원의 이동 및 부족 시 서비스를 원활하고 안정적으로 유지할 수 있다는 것이다. 데이터 서비스를 위한 인력이 충분하지 않은 대학도서관에서는, 첫 번째로 로컬 자원에 제한되어 있을 때 -고용 요구가 없을 경우- 새로운 데이터 큐레이션 서비스가 중요함을 제안할 수 있고, 두 번째로 로컬 데이터 큐레이션의 전문가가 규모가 크고 강력한 네트워크에 참여할 기회를 가지며, 세 번째로 DCN 파트너가 제공하는 명확한 로드맵을 통해 데이터 큐레이션 서비스 성숙도 및 확장의 이점을 누릴 수 있다. 마지막으로 도서관 리포지터리에서 실행되는 데이터 수집, 기탁, 아카이빙 등과 관련된 업무에 대한 표준화를 제공할 수 있다. 특정 전공 또는 주제 리포지터리에서는 첫 번째로 DCN 파트너 기관으로부터 더 우수하고 가치 있는 데이터 제출을 받을

수 있으며, 두 번째로 DCN의 파트너들을 통해 학술 레파지터리에 최신의 데이터를 추가하거나 자주 접하지 못한 데이터 유형을 확장할 수도 있으며 세 번째로 외부 기관에 상주하는 큐레이션 직원을 활용하여 인력에 대한 간접비용을 최소화시킬 수 있다. 네 번째로는 참여 기관들의 네트워크 덕분에 더 많은 연구자들이 학술 리포지터리에 접속하게 되며, 마지막으로는 학술 리포지터리가 파트너로 참여하고자 하는 경우, 소비 규모가 증가함에 따라 잠재적인 새로운 수입원을 확보하게 된다.

DCN에서는 다양한 인력 및 참여자를 나누어서 그 역할을 설명하는데, 그 중 DCN 큐레이터의 업무가 중요하다(〈그림 7〉 참고). DCN 큐레이터는 전문 지식과 도메인별 지식을 사용하여 재사용에 적합한 데이터를 검토하면서 표준화된 파일 형식의 특정 작업을 수행하는데, 'CURATE' 단계에서 데이터 세트를 관리하고 운영하게 된다(Johnson et al., 2017).

- C - 데이터 파일을 확인하고 문서를 읽는다.
- U - 데이터를 이해한다 또는 실행한다.
- R - 누락된 정보 또는 변경을 요청한다.
- A - 검색 가능성을 높이기 위해 메타데이터로 제출 내용을 보강한다.
- T - 재사용 및 장기 보존을 위한 파일 형식을 변환한다.
- E - FAIRness에 대해 평가한다.

3. KISTI 콘텐츠 큐레이션 센터의 업무 분석 및 특징 파악

KISTI 콘텐츠 큐레이션 센터는 큐레이션팀, 콘텐츠구축팀, 국가 R&D 성과물관리팀, 한국 DOI센터운영팀, 자료실운영팀 등 총 다섯 개의 팀으로 이루어져 있다. 본 연구에서는 센터의 업무 및 콘텐츠를 분석하기 위해 콘텐츠 관리에 초점을 맞춘 큐레이션팀, 콘텐츠구축팀, 국가 R&D 성과물관리팀, 한국DOI센터운영팀 대표 담당자들과 집단 면담을 실시하였다. 팀별 집단 면담은 2018년 5월 24일부터 6월 21일까지 약 한달간 진행되었으며, 각 면담은 평균 2시간 정도였다. 큐레이션팀과 한국DOI센터운영팀은 1회 집단 면담을 실시하였고 국가R&D 성과물관리팀과 콘텐츠구축팀은 2회 집단 면담을 진행하였다.

3.1 콘텐츠 큐레이션 센터 업무 분석 및 현황

본 연구에서 파악한 콘텐츠 큐레이션 센터의 팀별 주 업무는 <표 1>과 같다.

팀별 면담은 <표 2>의 질문지가 바탕이 되었다. 팀별 업무에 차이가 있기 때문에 <표 2>의 질문지를 큰 틀로 세부적인 내용은 조정하였다.

집단 면담의 내용을 통해 확인한 콘텐츠의 범위를 정리하면 <표 3>과 같다. 국내 학술지는 학회 및 협회에서 원문 데이터를 받아서 메타데이터, 식별자 등을 부여 및 관리한다. 해외 학술지는 메타데이터 및 식별자를 부여 및 관리하고 KESLI(Korean Electronic Site License Initiative) 컨소시엄을 통해 데이터를 추가한다. R&D 보고서는 원문과 함께 메타데이터 및 식별자를 관리하며, 국립중앙도서관 및 국회도서관과도 데이터를 공유한다. R&D 연구의 결과물에 포함된 논문은 메타데이터를 관리하고, 한국학술지인용색인(Korea Citation Index, KCI), Web of Science(WoS), SCOPUS와 연결되어 있다. 그 외 R&D 성과물인 도서와 CD 등은 메타데이터와 식별자를 부여하고 관리한다.

또한 본 연구에서는 센터 내 자체 평가를 통해 DCC에서 제안한 큐레이션 활동이 얼마나 잘 이루어지고 있는지를 분석하였다. 업무 담당자들은 리커드 척도, 0부터 5사이의 숫자로 센터의 큐레이션 활동의 성취도를 표시하였다. 질문의 큰 범주를 '개념화', '생산 혹은 접수', '평가와 선별', '입수', '보존 활동', '저장', '접근, 이용 그리고 재사용' 등으로 나누고, 하위 50개의 질문을 구체화하였다. 각 활동에 대한 체크리스트의 주요 내용은 다음과 같으며, 세부항목의 개수도 함께 표시하였다.

- 개념화(Conceptualise): 14개 세부항목
- 연구업무와 콘텐츠 큐레이션을 연속적인 일련의 활동으로 인식한다.(8)

〈표 1〉 콘텐츠 관리에 초점을 둔 팀별 주요 업무

팀 구분	업무내용
큐레이션팀	• 콘텐츠 큐레이션 모델 개발 및 평가 • 콘텐츠 큐레이션 정책 기획 • 메타데이터 자동추출 기술 연구 개발 • 학술논문 관리체계 개선을 위한 수준진단 및 개선계획 수립 • 참고문헌 자동식별 기술 연구 개발 • 개체 자동식별 기술 고도화 • 주제어 자동분류 기술 연구 개발 • 해외학술논문 데이터 운영방안 연구
국가R&D성과물관리팀	• 국가연구 성과물 활용 체제 구축·운영 - 국가R&D 보고서원문/ 논문 성과물 관리·유통 전담기관 역할 • 국가R&D 보고서원문 수집 및 고부가가치 DB 구축·관리(보고서원문 DOI 등록 및 관리 체제 구축) • 국가R&D 논문 성과물 수집 및 원문 연계 DB 구축·관리 (NTIS 논문성과데이터 입수·'18년부터, API를 통해 상시 입수) 및 점검, 주요등재DB 입수·구축·관리(SCIE, SCOPUS, KCI, KSCI), NDSL 미검색 논문 성과 메타데이터 입력 및 관리) • 국가R&D 성과물 활용·확산 및 홍보(성과물 정보의 NDSL/NTIS 연계) • 성과물 등록·관리시스템 • 고도화 및 운영 - NRMS, NARS, PAPER, 보고서원문프레임워크
콘텐츠구축팀	• 국내학술논문DB(OCEAN DB) 구축 • 해외학술논문DB(eGate DB) 구축
한국DOI센터운영팀	• 한국DOI센터 운영: 표준식별체계 운영기관과 협력체제 구축, 국제 DOI 재단(IDF) • 학술단체의 국제식별 및 연결 체계 지원: DOI, DLI, ISNI, 과제정보, 참고문헌 연결, Cited by 등 • DOI 등록관리 시스템 고도화: 특허, 단행본(ISBN-A), 학위논문에 대한 DOI 등록 체계 구축 • DOI 등록관리 시스템 운영

〈표 2〉 질문지 주요 내용 및 예시

주요 내용	예시
콘텐츠 큐레이션 대상 콘텐츠 및 업무내용	• 학술논문과 R&D 보고서를 대상 콘텐츠로 설정하였는데 각 콘텐츠 별로 주의를 기울여야 하거나 상세 범위에서 선별이 필요한 부분이 있으면 말씀해 주십시오. • 국가 R&D 과제관리기관에 대해 설명해 주십시오.
2018년 중점 추진 사항 및 본 연구(콘텐츠 큐레이션 모델 개발)와의 관계	• 2018년도 중점 추진 업무 중 큐레이션 모델 개발에 대한 내용을 설명해 주십시오. • DOI RA 체제에서 KISTI 콘텐츠 큐레이션 라이프 사이클 모델 개발에 참고할 부분이 있으면 설명해 주십시오.
이용자 요구 분석	• 콘텐츠의 선별과 배치라는 큐레이션 측면에서 성과물 이용자의 구분과 기준에 변경이 필요하다고 생각하신다면 그 이유에 대해 설명해 주십시오. • 분석형 맞춤형 제공 측면에서 각 이용자의 요구사항에 대해 설명해 주십시오.
데이터베이스 고부가가치화, 활용 강화	• 원문 구조화 및 메타데이터 가공 프로세스에 대해 설명해 주십시오, 그리고 추가 메타정보의 출처에 대해 알려 주십시오. • DOI 데이터 표준화 업무에 대해 설명해 주십시오.

〈표 3〉 콘텐츠 유형에 따른 및 구축 현황 외부 시스템과의 연결 여부

구축 현황 및 연결 콘텐츠 유형	원문	메타데이터	용어관리	식별자	KCI	WoS	SCOPUS
국내 학술지	○	○		○			
해외 학술지		○		○			
R&D 보고서	○	○		○			
R&D 논문		○			○	○	○
그 외 R&D 성과물		○		○			

- 국내 최대 과학기술지식정보 보유기관으로서 KISTI는 보유 데이터로 무엇을 하고 얼마동안 수행할 것인지 알고 있다. KISTI는 이러한 기대를 충족하기 위해 추가 자금이나 인력이 필요한가?(2)
- 데이터 수집 단계부터 지적 재산을 밝히고 문서로 남겨 보장한다.(2)
- 여러 사이트에 걸쳐 출판함을 금지하거나 제한하는 등의 출판 요구 사항을 확인한다.(2)
- 생산 혹은 접수(Create or Receive): 12개 세부항목
 - KISTI는 누구를 위해 데이터를 생성하고 있고 해당 데이터로 그들이 할 수 있는 (그리고 할 수 없는) 것을 알고 있다. 이를 해당 업무 관련 담당자에게 전달한다.(3)
 - KISTI 데이터 구축 업무 과정에서 해결해야 할 데이터 보호 요구 사항을 파악하고 이를 모든 직원에게 반드시 전달한다.(2)
 - 초기 단계부터 콘텐츠, 구문 및 구조에 활용할 어떤 표준에 합의한다. 일단 합의가 이루어지면, 프로젝트의 다른 담당자 및 함께 일하게 될 데이터/정보 관리자에게 이 사항을 확실하게 전달하도록 한다. 필요한 경우 교육을 제공한다.(3)
- 가능한 한 빨리 데이터 품질 측정 지표를 설정하고 이들 지표가 전달되어 모니터링 (추적 관찰)되는지 확인한다.(3)
- 함께 일한다. 즉 콘텐츠 큐레이션 센터와 관련 부서/팀 그리고 콘텐츠 구축 담당자들이 정기적으로 의사소통해야 한다(고립된 상태로 일을 할 수는 없다).(1)
- 평가와 선별(Appraise and Select): 5개 세부항목
 - 초기부터 선택과 평가를 시작한다.(1)
 - 현재 필요하거나 향후 필요할 것으로 파악되는(생각되는) 콘텐츠 및 메타데이터에 대한 계획을 세운다. 시간 경과에 따라 필요할 최소한(핵심)은 무엇인가?(2)
 - 무엇을 폐기해야 하는지 알고 있다. 폐기는 때로는 법률적 요건 준수를 보장하기 위해 필수적이다.(1)
 - 콘텐츠 큐레이션 센터와 관련 부서/팀 담당자가 협력하여 정책을 개발하고 현실적이고 실행 가능한 워크플로우를 확인한다.(1)
- 입수(Ingest): 2개 세부항목
 - 기술 표준을 준수/참고하여 계층적 데이

- 기술 업무를 수행한다.(1)
- '입수' 프로세스를 가능한 한 간단하게 만들고 상세한 지침을 제공한다. 할 수 있으면 프로세스를 자동화한다.(1)

• 보존 활동(Preservation Action): 5개 세부항목

- '모범 사례'와 권장된 접근법을 검토할 때 비판적인 자세를 두려워하지 않는다. 위의 방법들은 그것들이 만들어진 특정 시나리오를 위해 작동하는 것이지 KISTI를 위한 것이 아니다. '선호되는 (우선)' 포맷과 같은 것들을 평가하는데 이용되는 기준을 알고 있는가?(4)
- 보존 활동을 문서화해서 사람들이 시간이 지남에 따라 데이터가 어떻게 처리되었는지 알 수 있도록 한다.(1)

• 저장(Store): 5개 세부항목

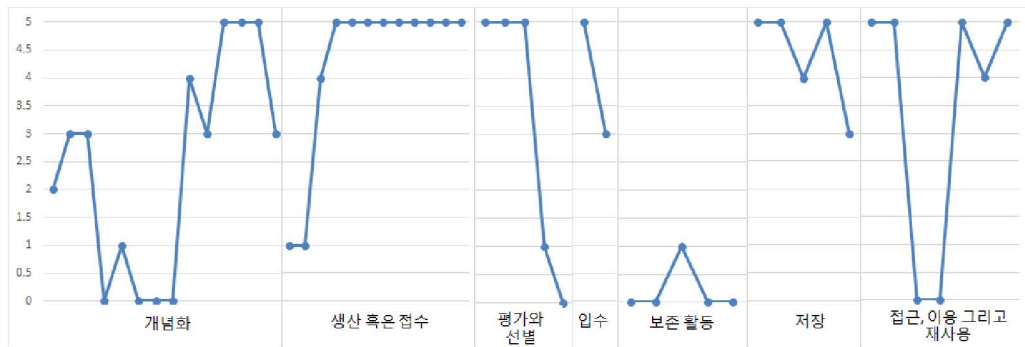
- 보관 단계에서 최종적인 데이터 품질 보증을 책임질 책임자(연구원, 아카이브, 정보 관리자 등)를 결정한다. 이 QA의 최종 상태는 모든 이해 관계자에게 반드시 전달되어야 한다.(3)
- 콘텐츠 구축 완료 후 관련 (부서) 시스템 연계/이전(transfer)을 위해 (가능한 경우) 공식적인 수령증이나 비공식적인 접수 통지를 받는다.(2)

• 접근, 이용 그리고 재사용(Access, Use and Reuse): 7개 세부항목

- KISTI는 이용자가 데이터를 이용하여 수행할 수 있는 작업과 기간을 알고 있다.(4)

- 접근 및 사용에 관한 어떤 제한 사항도 반드시 전달되고 준수되어야 한다.(1)
- 시간 경과에 따른 충분한 맥락정보(context) 제공을 보장하여 원래 지정된 이용자 집단이든 신규 이용자인든 상관없이 반드시 데이터를 찾아 이용할 수 있도록 한다.(2)

〈그림 8〉은 DCC의 큐레이션 활동 체크리스트에 콘텐츠 큐레이션 센터의 업무를 연결한 것으로 센터 내 중점 임무를 확인하고 미진한 부분을 파악할 수 있었다. 좀 더 자세히 살펴보면, 첫 번째로 콘텐츠 큐레이션 센터는 데이터 베이스 구축 및 관련 서비스를 주 업무로 삼고 있으며 '보존 활동'에 대한 부분은 실행하고 있지 않다. 두 번째로 콘텐츠는 주로 '접수' 채널 통해 수집되고 있으며 중복 및 가짜(fake) 저널의 폐기에 대해서는 처분 단계에서만 고려하고 있고 '평가 및 선별' 활동에서의 폐기는 구체적인 지침이나 기준이 마련되어 있지 않다. 세 번째로 이용자 참여는 국내 학회, 국가 R&D 연구과제 수행기관(산학연), 과제 관리 기관 등 기관 단위 중심으로 이루어지고 있다. 개인 연구자에 대해서는 KISTI의 국가과학기술정보센터 NDSL 서비스 운영 부서에서 이용자 통계 로그 및 행태를 파악하고 있다. KISTI 내부 부서간의 협업 체계를 개선할 필요가 있다. 마지막으로 5점에 가까운 완성 단계에 있는 활동과 거의 0점에 가까워서 앞으로 많은 발전을 이루어져야 할 활동이 공존한다는 것이다. 즉 활동별 완성 정도의 격차가 크다는 것을 알 수 있었다. 이러한 결과는 콘텐츠 큐레이션 센터가 디지털 큐레이션 라이프 사이클에 맞춘 업무 방향성을 재설계하는데 매우 큰 영향을 주었다.



〈그림 8〉 DCC 체크리스트를 통한 디지털 큐레이션 센터의 업무 분석

3.2 특징

KISTI는 과학기술 콘텐츠 수집유통을 사명으로 하고 있으며 1990년대 후반부터 국내 학술단체와는 협약을 통해 학술논문DB를 구축하고 있다. KISTI는 국내 학술논문의 영문 서비스 KoreaScience를 통해 국내 논문의 글로벌 활용확산을 촉진하는 점에서 차별화된 우수성이 있다. 또한 전 세계 40여 정보원으로부터 해외 학술논문 메타데이터를 수집, 구축, 품질 정제를 통해 고품질 학술논문 DB를 구축하고 있다. 1억여 건에 달하는 논문 메타데이터의 정보량은 국내에서는 최대량이며 세계적 수준에 달한다. KISTI가 구축한 참고문헌 데이터는 7억여 건에 달하며 이를 활용하면 세계적인 인용색인DB에서만 가능하던 학술연구성과 영향력 분석 등 다양한 정보 분석 및 서비스가 국내에서도 가능할 것으로 보인다.

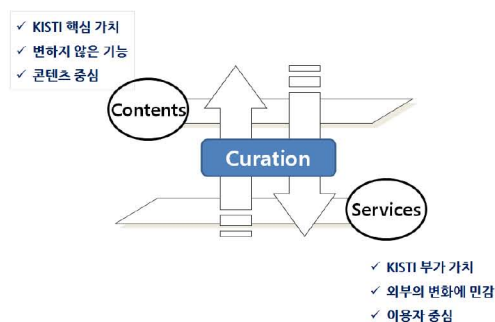
2008년부터는 연구성과 관리 및 유통 전담 기관으로 국가 R&D 보고서원문, 논문 수집, 구축, 정제, 보존 기능을 체계적으로 수행하고 있다. 기술적으로는 인공지능 기술을 적용하여 문헌으로부터 메타데이터 추출, 참고문헌 식별,

그리고 저자식별을 수행하는 등 기술적인 인프라가 우수한 것으로 보인다. 특히 참고문헌 식별은 정확율 97%로 세계적 수준으로 보인다.

또한 KISTI는 현재 한국 유일의 DOI 등록 기관으로서 국내 산학연 기관이 출판하는 과학 기술 콘텐츠에 항구적 식별자인 DOI 등록을 지원하고 있다. 장기적으로는 과학기술 콘텐츠와 연구데이터를 연계함으로써 국내 과학기술 콘텐츠의 글로벌 활용확산을 위한 중요한 국가적 인프라이다.

이러한 맥락에서 콘텐츠 큐레이션 센터의 업무는 KISTI가 가지는 핵심 가치와 부가가치를 모두 추구하고 있는 것이 특징이다. 〈그림 9〉에서 제시한 것처럼 콘텐츠 큐레이션 센터에서 다루는 콘텐츠는 이용자를 고려한 새로운 서비스로 생성되어야 한다. 콘텐츠를 관리하고 확대하는 것은 센터의 변하지 않는 기능이며, 그 기능이 외부의 변화를 반영할 때 이용자 중심의 서비스로 전환될 수 있다. 이런 점들을 감안할 때, 콘텐츠 영역에는 식별 기능, 실행가능성, 유틸리티, 진본성, 영구성, 검색 및 발견, 이동성, 온톨로지 지원, 평가, 적시성 등과 같은 특성이 고려되어야 하며, 서비스 영역에는 지속

가능성, 이용가능성, 반응 및 응답, 확장성, 신뢰성, 보안, 상호운용성 등이 포함되어야 한다. 그리고 두 영역에 걸쳐있는 특성으로는 완전성과 이용자 중심의 운영이 있다. 두 영역을 함께 고려하기 위해서는 센터 업무의 방향성은 UC3에서 제시한 태스크(프로젝트) 중심에서 계획(프로그램매틱) 중심의 접근방법을 채택해야 하고, 현재의 업무에 자연스럽게 큐레이션 모델을 접목시키는 방안이 필요하다. 즉 콘텐츠 큐레이션 센터에 맞는 디지털 큐레이션 모델을 제공하여 현재의 기능을 잘 수행하고 변화하는 이용자 요구를 충족시킬 수 있는 업무 활동의 기준을 마련하여야 한다.



〈그림 9〉 콘텐츠 큐레이션 센터 업무 특징

4. KISTI Curation Life Cycle Model(CLCM) 및 업무 가이드라인 개발

본 연구에서는 콘텐츠 라이프 사이클과 KISTI 콘텐츠 큐레이션 센터의 업무 및 기능을 연결하여 KISTI 큐레이션 라이프 사이클 모델(KISTI Curation Life Cycle Model: KISTI CLCM)을 개발하였다.

4.1 KISTI Curation Life Cycle Model (CLCM) 개발

본 연구에서 제안한 KISTI CLCM은 〈그림 10〉과 같다. KISTI CLCM은 전반적으로 DCC 모델을 차용하였으며, 콘텐츠 큐레이션 센터의 업무 및 관리 콘텐츠를 분석하여 전체 라이프 사이클 활동과 순차적 활동으로 나누어서 정리하였다.

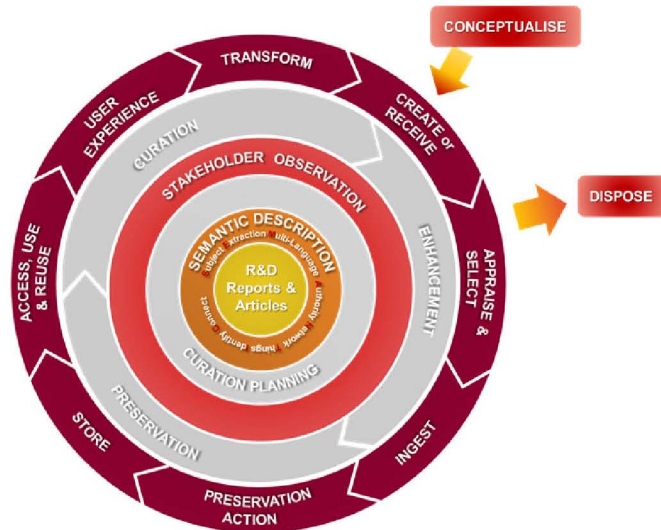
KISTI CLCM의 핵심 콘텐츠는 콘텐츠 큐레이션 센터에서 관리하고 있는 콘텐츠로 디지털로 표현된 정보인 국가 R&D 보고서와 학술 논문이다. 또한 DCC 모델의 가장 중앙에 위치한 디지털 객체와 데이터베이스의 개념도 포함된다.

4.1.1 전체 라이프 사이클 활동(Full Lifecycle Actions)

전체 라이프 사이클 활동은 라이프 사이클 전반에 적용되는 지속적인 활동을 말하며, 가장 안쪽에 위치하는 것은 의미 기술이다.

1) 의미 기술(Semantic Description)

의미 기술은 콘텐츠의 장기적인 제어 장치로 관리(administrative), 기술(descriptive), 구조(structural), 기술(technical), 보존(preservative) 메타데이터를 할당하여 콘텐츠에 대한 설명을 제공한다. 해당 활동은 콘텐츠 큐레이션 센터의 업무를 반영한 것으로 전체 라이프 사이클 활동의 시작점이라고 할 수 있다. 의미 기술 활동에서는 표준화된 메타데이터 셋을 적용하여 데이터를 기술하고 제어해야 한다. 종합적으로 의미 기술 활동에서는 디지털 객체에 대한 식별기호



〈그림 10〉 KISTI Curation Life Cycle Model(CLCM)

체계를 구축 및 관리하여 다음과 같은 기능을 수행해야 한다. 첫째, 특정 데이터 및 콘텐츠를 식별 및 확인하고, 둘째, 디지털 객체에 대한 설명을 추가하며, 마지막으로 외부 데이터 및 콘텐츠와의 연계 환경을 구축해야 한다. 콘텐츠의 속성으로 처리될 수 있는 데이터의 전거 체계, 즉 저자, 기관, 주제 등의 전거레코드, 과학기술 분야 개념어(디스크립터)에 대한 정의, 다국어 표현, 시소러스 구축 등과 같은 용어의 의미 정보를 함께 제공하는 것이다. 본 연구에서는 의미 기술과 관련된 활동을 'SEMANTIC'이라는 용어로 표현하였다. 'SEMANTIC'을 구성하는 알파벳으로 시작되는 단어들을 제시하여 의미 기술의 개념을 표현하였다(〈표 4〉 참고).

〈표 4〉의 내용을 좀 더 이해하기 쉽게 〈그림 11〉로 표현하였다.

'Thing'은 개별 데이터 및 콘텐츠를 표현한 것이고, 개별 데이터 및 콘텐츠의 속성은 식별요소(identity), 주제(subject), 전거(authority),

다양한 이형(multi language) 등으로 추출(extraction)된다. 이러한 속성들은 새로운 조합을 통해 새로운 서비스에 활용되는데, 본 연구에서는 'Network'로 표현하였다. 새로운 서비스는 해당 기관의 시스템뿐만 아니라 통합 시스템이나 타 기관의 서비스와도 연결(connect)된다.

2) 큐레이션 계획(Curation Planning)

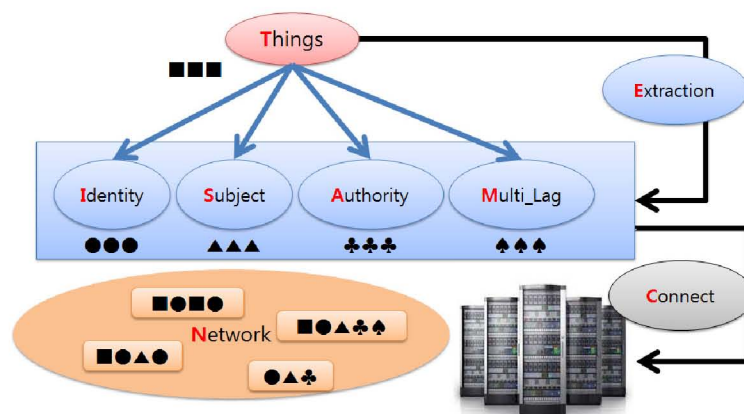
큐레이션 계획은 콘텐츠 큐레이션 라이프 사이클을 위한 계획으로 전체적인 활동들을 관리하는 것을 포함한다.

3) 이해당사자 관찰(Stakeholder Observation)

이해당사자 관찰 활동은 큐레이션의 각 단계에 적합한 이해당사자의 활동을 확인하는 것으로, 콘텐츠 큐레이션 라이프 사이클 활동과 관련된 이해당사자들의 요구를 파악하고 이를 공유할 수 있는 환경을 제공한다. 또한 정보사회

〈표 4〉 의미 기술 활동의 ‘SEMANTIC’ 개념

Key Words	Definitions
Subject	- 주제 전거데이터 구축 - 과학기술 분야 개념어에 대한 정보 제시 - 이형 관리 및 계층 구조 설명 - 시소러스나 온톨로지 모형으로 구체화
Extraction	- 데이터 및 콘텐츠의 중요 속성을 추출 - 이용자 요구와 부합되는 정보를 이끌어내야 함 - 표준을 준수하여 추출하고 표현
Multi-Language	- 다국어 사전 및 시소러스 구축 - 한국, 중국어, 일본어 사전 구축 - 개념에 대한 언어 표기를 확인하고 차별점을 제시
Authority	- 전거데이터 구축 - 인명, 조직명에 대한 정보 수집 및 관리 - 이형 관리 및 계층 구조 설명 - 한국인명의 특징인 동명이인 처리
Network	- 데이터 및 콘텐츠 연결의 구조화 - 콘텐츠 간의 연결(예, R&D보고서와 학술논문, 학술논문과 표), 데이터 간의 연결(예, 저자와 저자, 저자와 주제), 콘텐츠와 데이터 간의 연결(예, 학술논문과 저자, 그림과 다운로드 수) 등을 시도 - 식별기호를 활용하는 것이 중요함
Thing	- 콘텐츠를 기술하고 표현하는데 활용되는 모든 것 - 온톨로지 ‘Thing’의 개념과 유사한 것으로 이 세상에 존재하는 모든 것을 포함함 - 콘텐츠 자체가 될 수도 있으며, 콘텐츠를 설명하는 정보로 활용될 수 있는 모든 것을 뜻함
Identity	- 데이터 및 콘텐츠의 식별기호 - 디지털 객체를 식별하는 특수 기호 및 문자
Connect	- 콘텐츠와 정보시스템 - 콘텐츠는 이용자를 위한 다양한 정보시스템의 자원으로 활용됨. 단, 콘텐츠는 하나의 정보시스템에 종속되지 않으며 독립적인 가치를 가지고 있음



〈그림 11〉 의미 기술 개념도

의 변화 및 정보기술변화 등을 감지하고, 핵심 이용자 도출 및 이용자의 정보 요구도 파악해야 한다. 이해당사자는 합의된 표준, 도구, 소프트웨어 등의 개발에 참여할 수 있는 기회를 제공받게 된다.

4) 큐레이션, 강화 그리고 보존(Curation, Enhancement and Preservation)
큐레이션, 강화, 그리고 보존의 관점에서 이루어져야 할 활동은 다음과 같다.

- 큐레이션: 이용자 중심 서비스를 위한 핵심 콘텐츠 개발(필터링 기능) 및 기존 콘텐츠와 의미 정보(메타데이터, 전거데이터, 온톨로지, 시소러스 등) 연결
- 강화: 디지털 정보자원 자체는 물론이고 자원에 언급된 실세계의 개체나 상황, 사건 등을 연계시킬 수 있도록 시맨틱 웹 기술을 활용한 주석, 규칙, 온톨로지 등을 도입하여 기존 지식을 향상
- 보존: 장기적인 관점에서 디지털 콘텐츠 관리에 대한 계획 및 실행

4.1.2 순차적인 활동(Sequential Actions)
디지털 큐레이션의 선행 모형들을 참고로 KISTI CLCM의 순차적인 활동을 설계하였다.

- 개념화(Conceptualization): 콘텐츠 큐레이션을 위한 콘텐츠 구축 방법 및 저장 옵션을 비롯하여 디지털 객체 생성을 구상하고 계획함
- 생산 또는 접수(Create or Receive): 디지털 객체(학회지/보고서)를 수집하기 위

한 채널을 관리하고 생성 및 접수할 때 기본적인 식별기호를 포함한 관리, 설명, 구조 및 기술 보관 메타데이터를 생성함

- 평가/선별(Appraise and select): 디지털 객체(학회지/보고서)를 평가하여 장기적인 큐레이션이 필요한 대상을 선택함
 - (재)평가를 통해 폐기할 디지털 객체를 선택함: KISTI 업무를 고려하여 가짜 및 약탈적(predatory) 학술지 확인 등 일반적인 기록 평가와는 구별되는 활동이 포함됨
 - 문서화된 지침, 정책 및 법적 요구 사항을 준수함
- 입수(Ingest): 데이터베이스 구축을 위해 디지털 객체를 구축시스템으로 이전함
 - 데이터베이스 구축에 필요한 문서화 된 지침, 정책 및 법적 요구 사항을 준수함: 이 과정에는 (a) 이미지, 사운드, 텍스트 및 데이터의 디지털 레코딩, (b) 다양한 물리적 매체에 대한 아날로그 레코딩의 디지털화, (c) 리포지토리를 포함한 다른 출처의 디지털 리소스 가져 오기가 포함됨
- 보존 활동(Preservation action): 장기간 보존 및 디지털 객체의 권위 있는 속성정보 유지를 위한 활동
 - 물리적 원인 또는 기술적 진화로 인해 발생할 수 있는 위험으로부터 보호하기 위한 프로세스임
 - 영속성에 대한 보장이 필요함
- 저장(Store): 데이터베이스 구축이 완료된 디지털 객체를 관련 표준에 따라 안전한 방법으로 보관함
 - 데이터 품질 검수 및 검증을 위한 유효성 검사를 수행함

- 접근, 사용/재사용(Access, Use & Reuse):
최초 사용이나 재사용을 위해 지정된 사용
자가 데이터에 접근할 수 있도록 보장함
 - 인터페이스 모듈화, 검색 및 필터링, 관련
정보시스템 연계 등을 지원함
- 이용자 경험(User Experience): 사용자와
콘텐츠(또는 데이터) 간의 상호작용을 담
고 있음
 - 이용자가 주석 및 태그를 작성하는 등 콘텐
츠 이용 경험을 공유하는 플랫폼을 제공함
 - 이용자 통계 및 이용 행태 분석을 통한
인사이트를 발견할 수 있는 기반으로 제
공하는데 이는 자동으로 추출할 수 있음
- 변환(Transform): 원본에서 새로운 디지
털 객체를 생성함
 - 다른 형식으로서의 마이그레이션
 - 출판 등과 같은 새로운 결과물이 생성됨

4.1.3 비정기적인 활동(Occasional Actions)
디지털 큐레이션의 선행 모형들을 참고로
KISTI CLCM의 비정기적인 활동을 설계하
였다.

- 처분(Dispose): 문서화된 지침, 정책 및
법적 요구 사항에 따라 장기 큐레이션 보
존에 부적합한 데이터를 처분하는 것

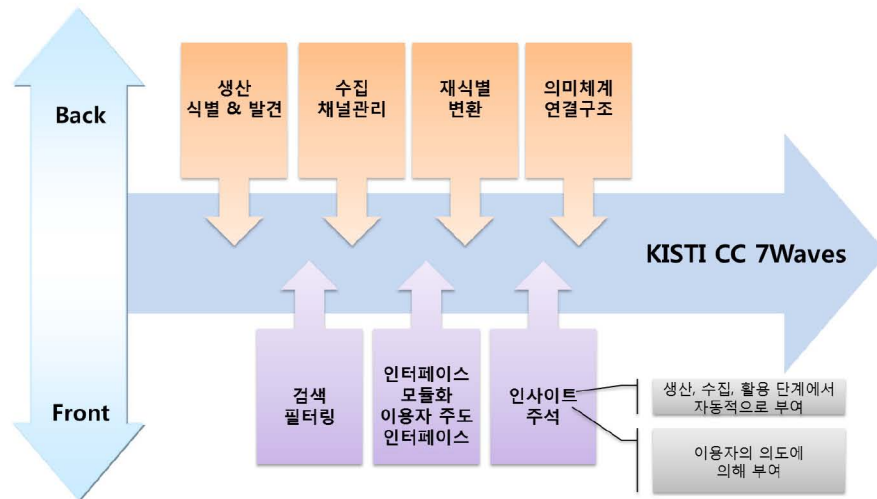
4.2 콘텐츠 큐레이션 웨이브 개발

본 연구에서는 UC3 모델을 기반으로 KISTI 콘
텐츠 큐레이션 웨이브를 개발하여 KISTI CLCM
의 기능을 제안하였다. KISTI 콘텐츠 큐레이
션 웨이브는 큐레이션 모델이 추구하는 이용자

중심의 서비스를 발굴하고 디지털 자원에 새로
운 가치를 추가하여 활용성을 높일 수 있는 방
안을 제시한다. <그림 12>와 같이 본 연구에서
는 7개의 웨이브를 14개 기능으로 세분화하여
각 단계별 완성도를 점검할 수 있게 하고 서비
스를 강조하는 큐레이션 모델을 지원한다.

<그림 12>에서 표현된 생산, 식별 및 발견,
수집, 채널관리, 재식별, 변환, 의미체계, 연결
구조 기능은 콘텐츠 큐레이션 센터의 업무와
직접적인 연결이 이루어지므로 'Back'이라고
표시하였다. 반면 검색, 필터링, 인터페이스 모
듈화, 이용자 주도 인터페이스, 인사이트, 주석
은 콘텐츠 큐레이션 센터가 제공하는 앞선 8개
기능을 활용하여 이용자와 소통하는 장에서 발
현되는 기능이므로 'Front'라고 표시하였다. 14
개의 기능에 대한 설명은 다음과 같다.

- 생산: 이용자 요구에 의해 또는 기관의 비
전에 맞는 업무를 수행하기 위해 콘텐츠를
생산하고 이를 관리하는 것. 현재 콘텐츠
큐레이션 센터에서는 자체 생산하는 콘텐
츠가 많지는 않으나 앞으로는 이용자 분석
자료나 전거데이터 생산 등과 같은 관리하
는 콘텐츠에 대한 설명 자료를 생산하고
배포할 필요가 있음
- 식별 및 발견: 기관 내부 및 외부에서 생산
되거나 관리되는 자료를 수집 정책 및 지
침에 의해 식별하고 찾아내는 작업임
- 수집: 수집 원칙에 맞는 자료들을 적절한
과정에 의해 콘텐츠 큐레이션 센터로 반입
하는데, 콘텐츠 자체뿐만 아니라 메타데이
터와 같은 콘텐츠 표현정보도 함께 수집해
야 함



〈그림 12〉 KISTI 콘텐츠 큐레이션 7 웨이브

- 채널관리: 콘텐츠는 다양한 개인 및 기관에 의해 생산·관리되기 때문에 이해당사자들의 상황을 파악하고 콘텐츠 유통의 과정에 문제가 없는지를 확인해야 함. 또한 유통 채널의 효율화에도 관심을 가져야 하며 이해당사자들의 복잡한 이해관계를 이해하고 업무에서는 그들의 권한을 보장하고 책임 역할을 수행할 수 있는 기반을 제공해야 함
- 재식별: 수집 단계에서 데이터나 콘텐츠의 식별체계를 확인할 수 있지만 재식별 단계에서 다시 콘텐츠 큐레이션 센터에서 관리하는 식별체계를 부여하여 KISTI 콘텐츠임을 확인하고 일관성 있는 콘텐츠 관리를 보장해야 함
- 변환: 정보기술 환경의 변화에 대처하고, 지속적인 서비스를 보장하기 위해 콘텐츠 자체의 포맷 변환, 기기 변환 등을 고려해야 하며, XML, PDF/A로의 변환 등과 같이 콘텐츠 공유 및 활용의 측면에서도 고려되어야 함
- 의미체계: KISTI CLCM의 의미 기술에 해당하는 기능을 포함
- 연결구조: 데이터 및 콘텐츠를 연결하여 내재된 의미가 더 확장되고 새로운 의미를 부여할 수 있는 기능을 고려해야 함
- 검색: 모든 데이터 및 콘텐츠 활용에 초점을 맞추고, 이용자의 검색 활동을 지원해야 함
- 필터링: 데이터 및 콘텐츠의 다양한 패킷을 제공하여 이용자 중심의 필터링 기능을 제공함
- 인터페이스 모듈화: KISTI 정보시스템 중심의 접근이 아닌 이용자가 원하는 콘텐츠들을 함께 활용할 수 있는 콘텐츠 중심의 서비스를 제공해야 함
- 이용자 주도 인터페이스: 이용자가 원하는 검색 및 브라우징 기능만을 모으고 조작할 수 있는 기능을 제공함
- 인사이트: 이용자가 콘텐츠를 생산, 수집,

활용하는 단계에서 자동적으로 부여되는 정보

- 주석: 이용자의 의도를 부여하는 장치로 별점이라든지 문자열을 통한 콘텐츠 설명 등이 추가될 수 있음

4.3 KISTI CLCM 활동과 디지털 큐레이션 센터 업무 연결

KISTI CLCM의 순차적인 활동과 비정기적인 활동을 실제 업무와 연결하면 <표 5>와 같다. KISTI CLCM의 순차적인 활동과 비정기

적인 활동을 프로세스로 하고 각 프로세스별로 연결될 수 있는 업무를 표시하였다. 이는 각 프로세스별 업무의 완성도 확인을 위한 점검기준표로 사용할 수도 있다.

<표 5>를 살펴보면 '개념화' 활동에서는 기관의 사명을 확인하고 기관에서 사용하고 있는 표준을 이해하는 것으로 시작하여 장기적인 관점에서 앞으로의 활동을 계획하고, 데이터를 공유하고 활용하는 방안을 모색하고, 변환 전략을 수립하고, 데이터 구조를 재설계 하며, 메타데이터 개발 및 추가를 고려해야 한다. '생산/접수' 활동에서 콘텐츠 큐레이션 센터가 할 수 있는

<표 5> KISTI CLCM 활동과 디지털 큐레이션 센터 업무 연결

활동/프로세스	업무 1	업무 2	업무 3	업무 4	업무 5	업무 6
개념화	KISTI mission 표준 준수	장기적인 관점	데이터 공유 및 활용	변환 전략	데이터 구조	메타데이터 개발 및 추가
생산/접수	식별(minting) - DOI, 저자기호, 기관식별기호 등	메타데이터수집				
평가/선별	인벤토리	DOI 부여 여부	OA 수집가능여부	저작권/ 라이선스 여부	Fake/ Predatory 판단	메타데이터 수준 확인 - 원문포함, 메타 데이터 수준
입수	식별(binding) 서비스 결합 - 등록번호	AIPbinding (데이터와 의미기술 정보 바인딩)				
의미 기술	인벤토리	특성화	색인	주석(관리)		
보존	데이터 정제	데이터 검수	보존메타데이터 추가	마이그레이션		
저장	내적 무결성 - 자원(asset) 관리	외적 무결성 확인 (평가와 연결)	백업(복제)	영구성	오류체크	보안 (기술 인프라)
접근/(재)사용	검색	필터링	식별(resolving) URL/URI: 페이지 접근 가능	접근 통제	인증절차	
이용자 경험	인사이트	주석(입력)				
변환	마이그레이션	포맷 변환 - PDF -> PDF/A - PDF -> XML - TIFF -> JPEG				
처분	처분					

업무로는 데이터 및 콘텐츠의 식별을 가능하게 하는 것인데 DOI, 저자기호, 기관식별기호 등을 연결하는 것이 중요하다. 또한 메타데이터의 수집을 통해 생산 및 접수한 콘텐츠나 데이터의 설명 정보를 확보하는 단계이다. '평가/선별' 활동에서는 인벤토리를 작성하고, DOI 부여 여부를 확인하고, OA(Open Access)을 통해 수집할 수 있는가를 확인하고 저작권과 라이선스 현황을 정확하게 분석하고 가짜 및 약탈적 학술지에 포함되는지를 판단한다. 마지막으로 메타데이터 수준 및 원문 포함여부를 확인한다. '입수' 활동에서는 수집된 데이터 및 콘텐츠가 서비스와 결합할 수 있도록 식별체계를 수립하고 연결하는 것이 중요하다. 또한 AIP(Archival Information Package)에서 수행하는 데이터와 의미 기술을 함께 묶는 활동도 여기에 포함된다. '의미 기술' 활동에서는 의미를 부여한 데이터 및 콘텐츠의 인벤토리를 작성하고, 기관 및 주제에 맞는 특성화를 진행하며, 색인을 부여·관리하고 이용자가 부여한 주석이나 외부에서 가지고 온 설명 정보를 추가한다. '보존' 활동에서는 데이터를 정제하거나 검수하고, 보존 메타데이터를 추가하며 마이그레이션을 실시하는 업무들이 수반된다. '저장' 단계에서는 내적 무결성을 통해 자산을 관리하고, 외적 무결성을 통해 기관의 사명에 맞는 데이터 및 콘텐츠를 수집·관리하고 있음을 확인한다. 또한 이 활동에서는 복제를 통한 백업을 시행하고, 데이터 영구성을 보장을 하며, 오류체크를 통해 시스템의 완전성을 추구하고 지속적인 기술 지원으로 보안체계를 유지한다. '접근/(재)사용' 활동에서는 검색 및 필터링 기능을 지원하고 URL/URI을 통해 데이터 및 콘텐츠에 접근

하며, 인증절차를 진행하고 접근을 통제하는 장치를 제공한다. '이용자 경험' 활동에서는 이용자가 콘텐츠를 생산, 수집, 활용하는 단계에서 자동적으로 부여되는 정보를 통해 이용자의 요구 및 이용행태를 찾아내는 인사이트 추출 기능을 제공하고, 이용자의 의도를 부여하는 장치로 별점이라든지 문자열을 통한 콘텐츠 설명 등을 추가할 수 있는 주석 기능도 제시한다. '변환' 활동에서는 마이그레이션을 통해 PDF에서 PDF/A로, PDF에서 XML으로, 그리고 TIFF에서 JPEG로 포맷 변환이 이루어진다. '처분' 활동에서는 기관의 처분 정책을 따르고, 가짜 및 약탈적 학술지 확인 등을 통해 외적인 무결성을 유지한다.

5. 결 론

본 연구에서는 KISTI 콘텐츠 큐레이션 센터의 업무를 중심으로 콘텐츠 라이프 사이클에 맞는 큐레이션 모델인 KISTI CLCM을 개발하였다. 모형 개발을 위해 본 연구에서는 DCC, DCC&U, UC3, DCN 모델 등과 같은 선행 큐레이션 모델을 살펴보고, 콘텐츠 큐레이션 센터의 업무와 연결되는 모델들의 장점을 도출하였다. 또한 KISTI 콘텐츠 큐레이션 센터의 업무 담당자들과의 집단 면담을 통해 센터에서 다루는 콘텐츠의 특징을 파악하고 현 업무의 성취정도를 확인하였다.

본 연구에서는 UC3 모델을 기반으로 KISTI 콘텐츠 큐레이션 웨이브를 개발하여 KISTI CLCM의 기능을 제안하고 이용자에게 제공해야 할 주요 서비스들을 구체화하였다. KISTI

콘텐츠 큐레이션 웨이브의 14개의 기능은 콘텐츠 큐레이션 센터의 임무를 한 눈에 보여주며, 이를 좀 더 구체화하기 위해 4.3장을 통해 실무와 연결된 부분들을 제시하였다.

본 연구에서 제안한 KISTI CLCM은 다음과 같이 활용될 수 있다. 첫 번째로는 KISTI 콘텐츠 큐레이션 정책 개발에 기초 자료로 활용하여 현재 KISTI 콘텐츠 관리 및 서비스 체계 점검하고 과학기술 지식정보인프라의 연구개발 및 서비스 체계 확립이라는 KISTI 임무를 수행할 수 있는 기반을 제공할 수 있다. 두 번째로는 국가 차원의 고부가가치 과학기술 정보인프라 구축에 활용되어 과학·기술 및 관련 산업에 관한 정보를 종합적으로 수집·분석·관리를 위한 콘텐츠 관리 정책 기초 자료로 활용 가능하며, 국가

과학기술 및 산업발전에 기여할 수 있는 과학기술정보의 관리 및 유통에 관한 기술·정책·표준화를 위한 참조 모델로 제공될 수 있다.

본 연구는 KISTI 콘텐츠 큐레이션 센터에 맞는 큐레이션 모델을 개발하는 것에 목적을 두었다. 특정 연구기관이 라이프 사이클에 맞는 디지털 콘텐츠 큐레이션 모델을 구축했다는 것은 디지털 객체를 관리하는 타 연구기관 및 대학에 긍정적인 영향을 주며 좋은 사례로 남을 수 있다. 그러나 본 연구에서는 개발된 모형이 얼마나 실무에 도움이 되는지 그리고 적용될 때 발생하는 문제는 없는지 등에 대한 검증은 이루어지지 않았다. 후속 연구를 통해 모형의 안정화 과정을 정리하고 모형의 평가 분석도 제시되어야 한다.

참 고 문 헌

- 이혜원 (2016). 역사콘텐츠 활용을 위한 디지털 큐레이션 프레임워크 제안. 한국기록관리학회지, 16(3), 235-256. <https://doi.org/10.14404/JKSARM.2016.16.3.235>
- Australian National Data Service (2016). ANDS Guide - Curation continuum, Retrieved from <http://www.ands.org.au/guides/curation-continuum>
- CDL(California Digital Library) (2010). UC3 Curation Foundations. retrieved from <https://confluence.ucop.edu/download/attachments/13860983/UC3-Foundations-latest.pdf>
- Constantopoulos, P., & Dallas, C. (2008). Aspects of a digital curation agenda for cultural heritage. 2008 IEEE International Conference on Distributed HumanMachine Systems, Athens, Greece: IEEE.
- Constantopoulos, P., Dallas, C., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., Kotidis, Y., & Papatheodorou, C. (2009). DCC&U: An extended digital curation lifecycle model. International Journal of Digital Curation, 4(1), 34-45. <https://doi.org/10.2218/ijdc.v4i1.76>

- DCC Homepage a (n.d.). Curation Lifecycle Model. Retrieved from <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- DCC Homepage b (n.d.). Lifecycle model FAQ. Retrieved from <http://www.dcc.ac.uk/resources/curation-lifecycle-model/lifecycle-model-faqs>
- Groenewegen, D., & Hahnel, M. (2014). Combining cloud management and discoverability with institutional storage. Retrieved from <https://eresearchau.files.wordpress.com/2014/11/thursday-1230-grand-1-2-david-groenewegen.pdf>
- Heller, L. (2017). Private research, shared research, publication, and boundary transitions: Additions (marked red). Retrieved from https://blogs.tib.eu/wp/tib/wp-content/uploads/sites/3/2017/05/data_curation_continuum_heller2017.pdf
- Humphrey, C. (2006). E-Science and the life cycle of research. <https://doi.org/10.7939/R3NR4V>
- Humphrey, C., & Hamilton, E. (2004). "Is it working? assessing the value of the canadian data liberation initiative." *Bottom Line*, 17(4), 137-146. <https://doi.org/10.1108/08880450410567428>
- Johnson, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2017). Data curation network a cross-institutional staffing model for curating research data. Retrieved from <https://conservancy.umn.edu/handle/11299/188654>
- Oliver, G. (2010). Transcending silos, developing synergies: libraries and archives. *Information Research: An International Electronic Journal*, 15(4). Retrieved from <http://www.informationr.net/ir/15-4/colis716.html>
- Oliver, G., & Harvey, R. (2016). *Digital curation*. 2nd edition. Chicago: ALA Neal-Schuman, an imprint of the American Library Association.
- Pennock, M. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. *Library and Archives Journal*, Issue 1. Retrieved from http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf
- Treloar, A., Groenewegen, D., & Harboe-Ree, C. (2007). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine*, 13(9/10). Retrieved from <http://www.dlib.org/dlib/september07/treloar/09treloar.html>

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Lee, Hyewon (2016). Suggestion of a digital curation framework for historical contents. Korean Society of Archives and Records Management, 16(3), 235-256.

<https://doi.org/10.14404/JKSARM.2016.16.3.235>