

# 활동이론을 중심으로 한 연구데이터 큐레이션 개념 모델 제안

## Proposal of a Conceptual Model for Research Data Curation based on Activity Theory

한 나 은 (Na-eun Han)\*

### < 목 차 >

I. 서론	IV. 데이터 큐레이션 활동 현안
II. 이론적 배경	V. 연구데이터 큐레이션 개념 모델 제안
III. 디지털 큐레이션 모델 분석	VI. 결론 및 제안

**요약:** 본 연구는 활동이론을 이론적 틀로 활용하여 연구데이터 큐레이션 모델을 분석한 문헌 연구로, 활동이론에서 사용하는 활동의 구성 요소들을 바탕으로 여러 연구데이터 큐레이션 모델을 분석했을 뿐만 아니라 연구데이터 큐레이션 활동을 수행하는데 있어서 도서관계에서 논의가 필요한 현안을 분석하였다. 그리고 이를 기반으로 새로운 연구데이터 큐레이션 개념 모델을 제안하였다. 본 연구는 기존에 제안된 5개의 디지털 큐레이션 생애주기 모델이 어떻게 구성되어 있는지 분석하고 각각의 모델에서 산발적으로 제시되는 활동요인들을 분석하여 공통 요인을 추출하고 새로운 모델로 통합하여 새로운 연구데이터 큐레이션 개념 모델을 제안하였다. 또한, 도서관 및 리포지토리에서 연구데이터 큐레이션 활동을 수행함에 있어서 고려해야 할 6가지의 현안들을 분석하고 논의하였다. 본 연구에서 제안되는 연구데이터 큐레이션 개념 모델은 총 10단계로 이루어져 있으며, 각 활동의 단계에서 고려해야 할 실천적 문제와 모순을 포함하고 있다.

**주제어:** 연구데이터 큐레이션, 활동이론, 큐레이션 개념 모델, 디지털 큐레이션 활동 현안

**ABSTRACT:** This study is a literature study that analyzed the research data curation models using activity theory as a theoretical framework. Based on the factors of the activity used in the activity theory, this study analyzed various research data curation models, as well as issues that needed discussion in the library field in carrying out research data curation activities. And based on this, a new research data curation conceptual model was proposed. This study analyzed how the five previously proposed digital curation lifecycle models are configured, and analyzed the actions presented sporadically in each model. A new research data curation conceptual model was proposed by analyzing factors, extracting common factors and integrating them into a new model. In addition, six issues to be considered in carrying out research data curation activities in libraries and repositories were analyzed and discussed. The research data curation conceptual model proposed in this study consists of a total of 10 steps, and it contains practical issues and contradictions to consider at each stage of activity.

**KEYWORDS:** Research Data Curation, Activity Theory, Curation Conceptual Model, Issues of Digital Curation Activities

\* 한국과학기술정보연구원(KISTI) 박사 후 연구원(betterhan@kisti.re.kr / ISNI 0000 0004 9212 2493)

- 논문접수: 2023년 2월 20일 • 최초심사: 2023년 2월 27일 • 게재확정: 2023년 3월 16일
- 한국도서관·정보학회지, 54(1), 167-190, 2023. <http://dx.doi.org/10.16981/kliss.54.1.202303.167>

※ Copyright © 2023 Korean Library and Information Science Society  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

## I. 서론

### 1. 연구의 배경 및 목적

데이터 큐레이션과 디지털 큐레이션이라는 용어는 혼용되어 사용되어지는 경우가 많은데 일반적으로 더 많이 사용되는 용어로서 디지털 큐레이션(Digital Curation)은 장기간에 걸쳐 디지털 데이터를 관리하고 보존하는 것을 포함할 뿐만 아니라, 해당 디지털 데이터의 검색 및 향후 재사용을 위해 수행되는 데이터 관리 및 그에 대한 제반 활동을 모두 포함하는 개념이다(Tammaro, Ross, & Casarosa, 2014). 미국의 시장조사기관인 International Data Corporation(IDC)에 따르면 2020년의 전 세계 디지털 정보량은 약 90제타바이트(ZB)에 달한다고 한다. 이와 같은 디지털 데이터의 양적 증가와 더불어 데이터에 대한 체계적인 관리의 중요성도 역시 주목받고 있으며, 이러한 관심은 연구데이터의 체계적 관리, 보존, 활용 및 재활용에도 집중되고 있다(국가과학기술연구회, 2019; Tolle, Tansley, & Hey, 2011). 연구데이터란 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터”를 의미하며(국가연구개발혁신법, 2022), 연구데이터는 연구결과에 대한 핵심 정보를 담고 있기 때문에 공유 및 활용을 통해 연구결과를 증명할 수 있게 해주는 중요한 자료로 사용될 수 있다. 뿐만 아니라 동시에 다른 연구자들의 중복 연구 또는 데이터의 중복 생산을 방지할 수 있도록 하는 가치를 갖는다(국가과학기술연구회, 2019; 김지현, 2012). 연구데이터 관리에 대한 중요성을 바탕으로 보조금 지원 기구 및 출판사 등은 연구데이터에 대한 보다 체계적인 관리를 요구하고 있는데, 미국의 경우에는 미국국립과학재단(National Science Foundation, 이하 NSF) 및 미국국립보건원(National Institutes of Health, 이하 NIH) 등에서 데이터 관리 계획(Data Management Plan, 이하 DMP)을 요구하고 있으며, 국내에서도 과학기술정보통신부 산하 기관에서 공고하는 일부 과제에 한하여 DMP를 작성 및 제출할 수 있도록 요청하고 있다(과학기술기본법, 2020). 효율적인 데이터의 관리를 위해 디지털 데이터의 생애주기를 기반으로 하는 큐레이션 활동을 연구데이터에 적용함으로써 보다 체계적인 연구데이터 관리를 기대할 수 있다.

2001년 ‘큐레이션’이라는 용어가 처음 사용된 이후, 해당 분야에 대한 연구가 활발히 진행되고 있는데, 영어를 주요 언어로 하여 발간된 책, 보고서, 논문 등이 약 800여 편이 넘게 출판되었다(Bailey, 2019). 그러나 디지털 큐레이션 및 연구데이터 큐레이션에 대한 다양한 연구가 진행되어 오는 동안 특정 분야를 다루는 디지털 큐레이션 모델에 대한 연구는 상당히 적게 제안된 바 있고, 특히나 국내의 연구에서는 활동이론의 요소를 기반으로 디지털 큐레이션 모델을 분석한 경우는 존재하지 않는다. 뿐만 아니라 디지털 큐레이션 모델을 제안하는 동시에 현재 도서관계에서 큐레

이선과 관련하여 논의해야 할 현안들을 함께 고려한 연구는 없는 실정이다. 그런 의미에서 본 연구는 디지털 큐레이션 모델에 대한 새로운 문헌연구를 제공함과 동시에 도서관계에서 논의할 필요가 있는 디지털 큐레이션 현안을 포함한 연구데이터 큐레이션 개념 모델을 제안한다는 점에서 의의를 갖는다.

## 2. 연구의 내용 및 방법

본 연구는 활동이론을 이론적 틀로 활용하여 연구데이터 큐레이션 모델을 분석한 문헌 연구이다. 활동이론에서 사용하는 활동의 구성 요소들을 바탕으로 여러 연구데이터 큐레이션 모델을 분석했을 뿐만 아니라 연구데이터 큐레이션 활동을 수행하는데 있어서 도서관계에서 논의가 필요한 현안을 분석하였다. 이에 더해 본 연구는 연구데이터 큐레이션과 관련한 모델 분석 및 현안 분석을 바탕으로 연구데이터 큐레이션 개념 모델을 제안하고, 각 단계별 활동에서 디지털 큐레이터 또는 데이터 관리자가 수행해야 할 역할에 대해 정리하여 제안한다.

본 문헌 연구는 처음 디지털 큐레이션 생애주기 모델이 제안된 2008년의 모델부터 2015년까지 국외에서 제안된 디지털 큐레이션 모델 가운데 5개를 선정하여 분석을 진행하였다. 5개의 모델은 각각 DDI(Data Documentation Initiative) 생애주기 모델, DCC(Data Curation Centre) 모델, ANDS(Australian Research Data Commons) 데이터 공유 모델, DataONE(Data Observation Network for Earth) 모델, 그리고 빅데이터 생애주기 모델이다.

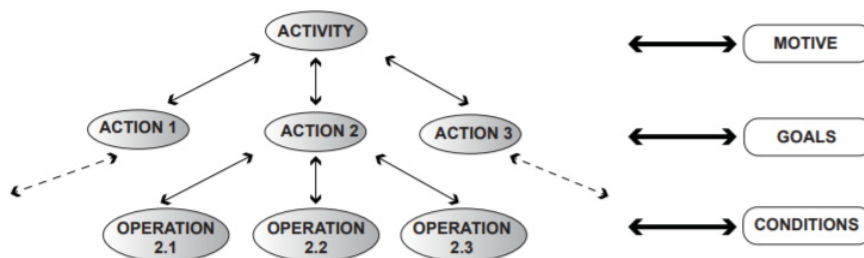
DDI 생애주기 모델은 초기에 제안된 큐레이션 모델로 디지털 데이터 자체뿐만 아니라 메타데이터 유형의 수를 증가시켜 데이터 품질 평가를 증진시키기 위해 제안되었다는 점에서 분석의 의의를 갖는다. DCC 모델은 영국의 데이터 큐레이션 센터에서 제안한 모델로 데이터별로 세밀한 수준에 맞게 큐레이션 및 장기적 보존 활동을 적합하게 수행할 수 있도록 도움을 제공할 수 있는 관련 표준이라고 할 수 있다. ANDS 데이터 공유 모델은 데이터의 생산자와 이용자 모두가 사용할 수 있는 개념으로 특히나 기타 여러 데이터를 포함한다기 보다 연구데이터와 연구 분야에 보다 더 집중하여 제안되었기 때문에 분석의 가치를 갖는다. 다음으로 DataONE 모델은 미국국립과학재단(National Science Foundation, 이하 NSF)에서 지원하는 프로젝트 중 하나로 다른 모델과는 다르게 커뮤니티 기반의 모델을 제안했다는 점에서 분석에 의의가 있으며, 마지막으로 빅데이터 생애주기 모델은 빅데이터를 대상 데이터로 하여 연구데이터를 관리하고 큐레이션 활동의 워크플로우를 계획하는데 활용되어질 수 있도록 일반적인 모델로 개발되었기 때문에 선정되었다. 해당 모델들은 모두 연구데이터를 대상 데이터로 하여 큐레이션 생애주기 모델을 제안했을 뿐만 아니라 각각 미국 또는 영국 내의 규모 있는 데이터 리포지토리에서 제안된 모델이라는 점에서 의의를 갖는다.

## II. 이론적 배경

### 1. 활동이론(Activity Theory)의 개념 및 구성요인

본 연구는 활동이론(Activity Theory)을 이론적 틀로 활용하여 연구데이터 큐레이션 모델과 연구데이터 큐레이션 활동에 관련한 현안을 분석하였다. 활동이론의 기본적 개념은 ‘주체(subject)’와 ‘객체(object)’ 간의 목적이 존재하고, 인간의 변형적이며 발전적인 상호 작용을 통해 특정한 활동을 이해할 수 있다는 것이다(Kaptelinin & Nardi, 2012). 활동이론의 초창기에 Vygotsky(1962)는 ‘중재(mediation)’의 개념을 포함한 활동의 틀을 주장하였고, 이 개념은 자극(stimulus)과 반응(response)의 사이에서 직접적인 연관성을 ‘매개물(mediating artifact)’이 포함된 삼각형으로 표현하는 것이다. 초기의 활동이론은 주체, 객체, 매개물을 각 꼭지점으로 하여 서로가 영향을 미치는 삼각형 형태의 모델이었으며, 이후 확장된 형태의 모델이 Leont’ev에 의해 개발되고(Leont’ev, 1978), Engeström(1987)을 통해 확장되었다.

활동이론에서 모든 활동은 그 형태와 관계없이 주체와 세계 또는 주변 환경과의 의도적인 상호작용으로 인식하며, 이는 주체와 객체의 상호작용이 이루어지는 과정으로 이해한다. 활동이론의 주체와 객체는 상호작용을 통하여 서로에게 영향을 미치고 변화시키며, 이와 같은 중심 과정을 ‘내면화(internalization)’라고 표현한다. 특정한 활동이 어떻게 구성되고 진행되는지를 분석함으로써 활동의 주체와 객체를 모두 이해할 수 있는데(Kaptelinin, 2005), 활동 구조에서 객체는 활동의 계층적 프로세스를 통해 결과로 변환된다. 계층 구조는 Leont’ev에 의하여 개발되었으며 ‘활동(activity)’, ‘행위(action)’, ‘동작(operation)’으로 구성되어 있다. 그리고 각각의 요인들은 ‘동기(motive)’, ‘목표(goals)’, 활동이 수행될 수 있도록 하는 ‘조건(conditions)’과 연관되어 있으며(Leont’ev, 1978; Lompscher, 2006), 이는 <그림 1>에서 보여지는 것과 같다. 본 연구에서는 연구데이터 큐레이션 자체를 활동 구조 가운데 최상위 계층에 존재하는 ‘활동(activity)’이라고 이해할 수 있다.



<그림 1> 활동의 계층적 구조(Kaptelinin & Nardi, 2012)

이후 Engeström(1987)은 기본 계층 구조에 다양한 추가 개념을 포함하여 활동 이론을 집단 활동의 모델로 확장하였는데, 이는 집단 활동뿐만 아니라 개인의 활동 역시 반영한다. 확장된 모델에서는 기존의 '주체(subject)', '객체(object)', '매개물(mediating artifact)'의 개념뿐만 아니라 '공동체(community)', '중재자(mediator)', '도구(tool)', '규칙(rule)', '분업(division of labor)'의 개념을 포함한다(Engeström, 2001). 도구는 주체와 객체 사이의 매개물로 작용하는 것들을 의미하며, 분업은 객체와 공동체 사이에서 관계를 중재한다. 특히 분업은 권력과 지위의 수직적인 분업을 의미할 뿐만 아니라 공동체 내에서의 협력적 구성원들 사이에 발생하는 수평적 업무 분업을 포함하는 개념이다(Engeström, 1987). 그리고 규칙은 명시적이거나 암묵적인 규범, 행동, 규정, 규약, 그리고 운영 및 상호작용을 구현 또는 제한하는 표준을 말한다(Kaptelinin & Nardi, 2012). 본 연구에서는 연구데이터 큐레이션 모델의 구성요소들에 대해 활동이론을 바탕으로 분석하였으며, 세부적으로는 행위, 활동 주체, 활동 객체, 도구, 규범 및 정책을 기준으로 각각을 분석하였다.

## 2. 디지털 큐레이션 모델의 의의

디지털 데이터는 본질적으로 기술적인(technical) 변화에 취약하기 때문에 데이터의 식별 및 보존에서 폐기에 이르기까지 모든 단계에서 관리에 필수적인 단계를 식별 및 계획하고 필요한 조치를 올바른 순서로 구성하고 구현하는 것이 중요한데, 생애주기를 기반으로 한 접근 방식을 통해 이를 수행 가능하다. Pennock(2007)은 디지털 데이터의 연속성을 보장하기 위해 생애주기 기반의 접근 방식이 필요하다고 주장한바 있고, 영국의 디지털 큐레이션 센터(Digital Curation Centre, 이하 DCC)는 디지털 데이터 관리에 대한 생애주기 기반의 접근 방식이 디지털 데이터의 성공적인 큐레이션 및 장기적 보존을 가능케 하는데 중요하다고 주장하였다(Higgins, 2008). 이와 같은 주장을 바탕으로 이미 다양한 분야 및 연구에서 디지털 데이터에 관리에 대한 생애주기 접근 방식을 연구하고 계획하기 위한 도구로서 디지털 생애주기 모델을 제안한 바 있다(Burton & Treloar, 2009; Constantopoulos et al., 2009; Higgins, 2008; Michener & Jones, 2012).

데이터 생애주기는 특정 데이터가 초기 생성되는 순간에서부터 유효 수명이 다하는 시점까지 활용되고 보존 및/또는 폐기되는 일련의 단계를 의미한다. 다양한 연구자들 및 기관에서 제안한 여러 데이터 관리 생애주기 모델이 존재하는데, 생애주기 모델의 중요성은 해당 생애주기 동안에 데이터 기록에서 수행해야 하는 여러 작업을 고려할 수 있도록 단계적 구조를 제공한다는 것이다(Ball, 2012). 디지털 자료들은 생성, 관리, 보존, 활용, 폐기되는 일련의 과정을 거치는데, 해당 과정들에 대한 정확한 이해가 선행되어야 효과적인 큐레이션 및 생애주기 프로세스를 기반으로 한 각 단계별 큐레이션 서비스를 제공할 수 있다. 이와 같은 효과적인 큐레이션 및 단계별 큐레이션 서비스를 제공하는데 있어서 디지털 큐레이션 모델은 개념과 방법을 제안한다는 점에서 의의를 갖는다.

### Ⅲ. 디지털 큐레이션 모델 분석

본 연구는 활동이론을 이론적 틀로 활용하여 여러 디지털 큐레이션 모델을 분석하였다. 본 장에서는 기존에 제안된 5가지 디지털 큐레이션 생애주기 모델을 대상으로 선정하여 활동이론의 구성요소들을 적용하여 분석하였다. 선정된 5가지의 디지털 큐레이션 생애주기 모델은 DDI 생애주기 모델, DCC 모델, ANDS 데이터 공유 모델, DataONE 모델, 그리고 빅데이터 생애주기 모델로, 해당 모델들은 모두 연구데이터를 대상 데이터로 한 큐레이션 생애주기 모델일 뿐만 아니라 각각 미국 또는 영국 내의 데이터 리포지토리에서 제안된 모델이다.

본 연구에서 분석한 5개의 디지털 큐레이션 모델은 데이터 생애주기 전반에 걸쳐 각 큐레이션 프로세스를 제시하고 있으며, 위와 같은 큐레이션 모델들은 디지털 데이터를 보다 적극적으로 관리하여 장기적 보존에 있어서 발생할 수 있는 위협을 줄이고 연구데이터에 대한 장기적인 접근을 가능하게 하도록 하는 동일한 목표를 갖는다. 그러나 연구데이터 큐레이션 활동의 구조, 행위 및 목표 등은 큐레이션을 수행하는 리포지토리 그리고 데이터 자체의 특성에 따라 차이가 존재한다. 이를 바탕으로 큐레이션이 수행되는 맥락에 따라 강조되는 큐레이션 활동, 큐레이터의 역할 및 문화, 작업, 도구 및 소프트웨어 등에는 차이가 존재한다.

각 연구데이터 큐레이션 모델들의 모델 구조와 목표를 살펴보면 다음과 같다. 먼저 DDI(Data Documentation Initiative) 3.0 버전 개념 모델은 연구데이터, 특히 사회과학 데이터에 대한 생애주기 모델을 다루고 있다(Vardigan, Heus, & Thomas, 2008). 해당 모델은 연구의 개념, 데이터 수집, 데이터 처리, 데이터 보존, 데이터 배포, 데이터 검색, 데이터 분석 및 용도 변경의 총 8가지 요소로 구성되어 있으며, DDI는 데이터의 용도를 변경하고 활용을 증진시키기 위해 데이터와 함께 활용되는 메타데이터 유형의 수를 증가시킴으로써 데이터 품질 평가에 기여하는 것을 목표로 제안되었다(Vardigan, Heus, & Thomas, 2008).

다음으로 DCC 생애주기 모델은 큐레이션 서비스를 수행할 수 있는 기준에 따라 데이터 항목별로 큐레이션 및 보존을 적합하게 수행하는 것을 목적으로 개발되었다(Higgins, 2008). 해당 모델이 대상으로 하는 데이터는 특정 도메인 분야의 연구데이터에 국한되지 않으며 디지털 형태의 모든 정보를 포함하는 데이터를 그 대상으로 한다. DCC 큐레이션 생애주기 모델은 데이터의 개념화, 생성, 평가 및 선택, 수집, 보존, 저장, 접근, 사용 및 재사용, 변환의 순서로 진행되며, 해당 모델은 시각화된 모델을 제공함으로써 초기의 개념화에서부터 데이터를 성공적으로 큐레이션하고 보존하는데 필요한 단계를 제안한다(Higgins, 2008).

호주국립데이터서비스(Australian National Data Service, 이하 ANDS)는 여러 이용자들이 데이터를 사용할 수 있도록 하기 위해 수행해야 하는 활동 모델을 제안하기 위해서 'Data Sharing Verbs'를 개발하였다(Burton & Treloar, 2009). 해당 모델은 운영 계획 및 활용 도구를 위한 구조화된 기술

(description)로서 데이터 생산자 및 데이터 이용자 모두가 사용할 수 있는 개념으로 다음 총 8개의 단계를 포함한다: 생성, 저장, 설명, 식별, 등록, 검색, 접근 및 이용(Burton & Treloar, 2009). ANDS는 연구자들로 하여금 보다 효율적인 연구가 가능하도록 디지털 큐레이션 프로세스를 통해 연구데이터를 공유하고 재사용하는 것을 목표로 해당 모델을 제안하였으며, 다른 모델들이 연구 데이터 뿐만 아니라 다양한 분야에서 활용할 수 있는 여러 형태의 데이터를 대상으로 하고 있는데 반해, ANDS가 제안한 생애주기 모델은 연구데이터와 연구 분야에 보다 더 집중되어 있는 모습을 보인다.

DataONE 데이터 생애주기 모델은 NSF에서 지원하는 프로젝트 중 하나인 DataONE에서 수집한 지구과학 데이터를 큐레이션하고 보존하기 위한 모범 사례를 구성하도록 설계된 커뮤니티 기반 모델이다(Michener & Jones, 2012). 이 데이터 생애주기 모델은 계획, 수집, 보장, 설명, 보존, 발견, 통합 및 분석의 8단계로 구성되어 있다(Michener & Jones, 2012).

마지막 분석 대상 모델은 빅데이터 생애주기 모델인데, 이는 Pouchard(2015)가 연구를 수행하는 관점과 디지털 큐레이션을 실시하는 관점을 결합하여 데이터 관리 작업을 식별하고 빅데이터를 처리하는데 필요한 체계를 포함하여 제안한 모델이다. 이 모델은 설명, 보장, 계획, 수집, 준비, 분석, 보존 및 검색의 단계를 포함하는데, 제안한 Pouchard(2015)는 빅데이터 생애주기 모델을 연구자, 데이터 관리자, 도서관 사서 등이 프로젝트 또는 조직에서 연구데이터를 관리하고 큐레이션 활동의 워크플로우를 계획하는데 사용할 수 있도록 하는 일반적인 모델로 개발하였음을 분명히 하였다.

다음 <표 1>은 각 디지털 큐레이션 모델의 구성 요소들을 간략하게 정리하여 보여준다.

<표 1> 디지털 큐레이션 모델 구성요소

모델	행위(모델 구조)	주체	객체	도구	규범 및 정책
DDI	○	○	○	○	○
DCC	○	○	○	○	○
ANDS	○	○	○	X	X
DataONE	○	○	○	○	○
Big Data Life Cycle	○	○	○	X	X

### 1. 행위 (Actions)

이와 같은 5가지의 디지털 큐레이션 모델의 공통적인 목표는 디지털 데이터를 적극적으로 관리하여 장기적인 보존을 수행함에 있어서 발생할 수 있는 위협을 줄이고 연구데이터에 대한 보다 지속적인 접근을 가능케 하는 것이다. 그러나 DDI, DCC 생애주기 모델은 연구데이터에

국한되지 않은 모든 일반 디지털 데이터를 큐레이션하는 것을 목적으로 설계되었으며, ANDS, DataONE, 빅데이터 생애주기 모델은 연구데이터를 그 대상으로 하여 큐레이션할 수 있도록 설계되었다.

본 장에서 분석한 디지털 큐레이션 모델의 구조는 기본적으로 생애주기를 기반으로 한 연속적인 구조를 보이고 있으나 각 행위(action)를 표현하는 명칭에는 차이가 존재한다. 동일한 의미를 갖는 큐레이션 단계 및 요인들을 통합 및 분석하여 표로 제시하면 다음 <표 2>와 같다.

<표 2> 디지털 큐레이션 모델을 구성하는 행위 요인

모델	행위 요인									
	계획	생성	수집	분류 및 색인	표준화	보존	활용 및 배포	평가	재사용	폐기
DDI	○	○	○	○	-	○	○	○	-	-
DCC	○	○	○	-	○	○	○	○	○	○
ANDS	-	○	○	○	-	○	○	-	-	-
DataONE	○	○	○	○	-	○	-	○	-	-
Big Data Life Cycle	○	○	○	○	-	○	-	○	-	-

<표 2>에서 볼 수 있듯이 5개의 디지털 큐레이션 모델에서 모두 공통적으로 나타나는 행위 요인은 생성, 수집 및 보존이며, 이는 큐레이션의 대상 데이터가 무엇이든 디지털 큐레이션에서 장기적 보존이 굉장히 중요한 요인임을 보여준다. 다음으로 5개 모델 가운데 4개의 디지털 큐레이션 모델의 공통적인 행위 요인은 계획, 분류 및 색인, 평가로 나타났다. 이를 통해 대부분의 디지털 큐레이션 모델에서 계획, 수집, 분류 및 색인, 보존, 평가가 기본적인 생애주기 프로세스의 주요 요인으로 활용되고 있음을 추론할 수 있다.

반면에 일부 특정 디지털 큐레이션 모델에서만 나타나는 행위 요인도 존재하는데, 특히 재사용 및 폐기의 구성 요인은 DCC 모델에서만 나타나며, 해당 모델은 접근, 활용 및 재사용을 강조하는 동시에 데이터에 대한 지속적인 접근 및 효율적인 큐레이션 서비스와 디지털 데이터의 장기적 보존을 강조한다.

## 2. 활동 주체(Subjects)

디지털 큐레이션 활동의 주체는 생산자(producer), 관리자(manager), 소비자(consumer)이다. 본 장에서는 5개의 디지털 큐레이션 생애주기 모델의 주체를 분석하며, 아래의 <표 3>은 간략한 분석 결과를 나타낸다.



〈표 3〉 디지털 큐레이션 모델의 주체

모델	주체		
	생산자	관리자	소비자
DDI	연구자	사서, 관리자, 리포지토리 개발자	연구자
DCC	데이터 생산자	데이터 과학자, 사서, 관리자	-
ANDS	연구자	리포지토리	연구자
DataONE	연구자	사서, 개발자	연구자
Big Data Life Cycle	연구자	사서, 관리자	연구자

가. 생산자(Producer)

생산자는 데이터를 생산하고 제공하는 주체로서, 정확한 디지털 큐레이션 활동을 수행하기 위한 데이터 수집의 첫 번째 단계의 주체이다. 5개의 디지털 큐레이션 모델 가운데 4개의 모델(DDI, ANDS, DataONE, 빅데이터 생애주기)은 연구자를 데이터의 생산자로 제시함과 동시에 연구자를 소비자로 정의함으로써 생산자이자 소비자로서 연구자를 정의하였다(Burton & Treloar, 2009; Michener & Jones, 2012; Pouchard, 2015; Vardigan, Heus, & Thomas, 2008). 반면에 DCC 모델에서는 데이터 생산자를 데이터 생산자로 정의함으로써 데이터를 생산하는 주체를 연구자로 제한하지 않는다(Higgins, 2008).

나. 관리자(Manager)

관리자는 지속적인 큐레이션 활동을 통해 데이터를 유지하고 공유할 책임이 있는 주체를 의미하며, 데이터의 수집, 보존, 접근 등 다양한 활동을 수행할 수 있도록 하기 위해서는 관리자에 대한 전문적인 교육을 실시하고 관련된 정책을 성문화하는 것이 필수적이다. 디지털 큐레이션 생애주기 모델에서는 관리자에 사서, 개발자, 데이터 과학자, 데이터 관리자 등이 포함됨을 보여주며, 개인이 관리자인 대부분의 모델과 달리 DDI 모델과 ANDS 모델은 개인뿐만 아니라 리포지토리 역시 관리의 주체로 활동할 수 있음을 명시하고 있다(Burton & Treloar, 2009; Higgins, 2008; Michener & Jones, 2012; Pouchard, 2015; Vardigan, Heus, & Thomas, 2008).

다. 소비자(Consumer)

소비자는 좁은 개념으로는 수집된 데이터의 주요 사용자라고 할 수 있다. 디지털 큐레이션 활동에서의 소비자는 데이터 사용의 역할뿐만 아니라 어떠한 데이터가 의미 있는 데이터이며, 유지 및 보존이 필요한지에 대한 의견 및 방향을 제시하는 역할을 할 수 있다. 본 연구에서 분석한 5가지 디지털 큐레이션 생애주기 모델 가운데 소비자 역할에 대해 명시하지 않은 하나의 모델을 제외하고는

모두 소비의 주체를 연구자로 지칭하고 있음을 알 수 있다(Burton & Treloar, 2009; Higgins, 2008; Michener & Jones, 2012; Pouchard, 2015; Vardigan, Heus, & Thomas, 2008).

### 3. 활동 객체(Object)

디지털 큐레이션은 디지털 데이터의 생애주기 전반에 걸쳐 데이터를 체계적으로 관리함으로써 해당 데이터에 가치를 부여하고 디지털 데이터의 지속적인 접근 및 재사용을 보장하는 활동이다. 따라서 연구데이터 큐레이션의 목적은 연구데이터에 대한 성공적인 보존 및 접근, 재사용을 가능하게 하는 것이다.

이와 같은 디지털 큐레이션의 대상은 기본적으로 디지털 데이터이며, 디지털 데이터는 단순한 디지털 개체(관련 식별자 및 메타데이터가 포함된 텍스트 형식의 파일, 이미지 파일 또는 음성 파일과 같은 개별 디지털 개체)에서부터 복잡한 디지털 개체(여러 디지털 개체를 결합하여 만든 개별 디지털 개체 혹은 웹사이트 등)까지를 포함한다(Constantopoulos et al., 2009). 본 연구에서 분석한 디지털 큐레이션 모델에서도 각각 기본적으로 디지털 큐레이션은 디지털 데이터를 대상으로 수행되지만, 연구데이터, 과학연구데이터, 생물학적, 환경적, 사회경제적 데이터 등 리포지토리의 특성에 따라 다양하게 수집 및 관리되고 있음을 확인하였다.

### 4. 도구(Tool)

본 연구에서 분석하는 5개의 디지털 큐레이션 생애주기 모델은 각각의 다른 리포지토리 또는 프로젝트에서 제안되었는데, 이 가운데 3개의 리포지토리에서는 디지털 큐레이션 활동에 사용할 수 있는 도구에 대한 내용을 제공함으로써 큐레이터, 데이터 관리자 및 사용자에게 편의를 제공한다.

먼저, DDI는 사용하는 용도에 따라서 총 31개의 도구를 제공하는데, 이와 같은 도구는 제어 어휘 적용, 프로필 생성, 편집, 색인 생성, 설문지 작성, 데이터의 저장 및 편집 등에 사용된다.

DCC 생애주기 모델을 제안한 DCC는 사용자가 누구인지에 따라 분류 가능한 도구를 제공하는데, 이는 크게 큐레이터를 위해 제공되는 도구와 연구자를 위해 제공되는 도구로 분류할 수 있다. 큐레이터를 위한 도구는 디지털 객체의 배치 및 수집, 데이터 패키지 보관 및 보존, 리포지토리 관리 및 운영을 목표로 하는 것인 반면에 연구자를 위해 제공되는 도구는 활성 연구데이터 관리, 연구결과 공유 및 활용 내용 추적을 목표로 하여 사용할 수 있는 도구들을 제공한다.

다음으로 DataONE은 데이터를 검색, 사용, 제공하는데 활용될 수 있는 소프트웨어 도구 모음인 'Investigator Toolkit'을 제공한다. 해당 소프트웨어 도구 모음은 DMP 도구, 통계 도구, 인용 도구, 메타데이터 편집기 및 클라우드 보관 시스템 등을 포함한다.

아래의 <표 4>는 디지털 큐레이션 생애주기 모델의 순차적인 작업 순서에 따라 각 리포지토리에서 제공하는 도구가 적용되는 것을 보여준다.

<표 4> 디지털 큐레이션 단계별 활용 도구

모델	디지털 큐레이션 단계					
	계획	생성	분류 및 색인	보존	활용 및 배포	평가
DDI		<b>데이터 편집</b> • SledgeHammer • Colectica Designer • DDI to DDL • DDI-XSLT • Colectica for excel • XCONVERT  <b>설문 작성</b> • Archivist • Colectica Questionnaire • Colectica Designer • QDDT	<b>DDI 프로필 생성</b> • DDIProfileXSLT  <b>색인/검색</b> • CED2AR • Colectica Designer • Colectica Portal • DBKfree • DDI Index  <b>분류</b> • NADA	<b>보호</b> • DDI Data Description Statistics Protection Software  <b>저장</b> • CED2AR • Colectica Repository • Colectica Portal • DBKfree	<b>통계</b> • Mapper • Colectica Resolver	
DCC	• DMPonline			• Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)		• Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) • Data Asset Framework (DAF) • Curation Costs Exchange
DataONE	• DMP Tool	• Morpho	• Zotero • Mendeley		• DataONE Search • Dash • ONE R • Kepler • VisTrails	

<표 4>는 각 리포지토리에서 제공하고 있는 도구가 디지털 큐레이션 생애주기 단계에 따라 구성 및 활용되고 있음을 나타내며, 이를 바탕으로 10개의 순차적인 작업 가운데 계획, 생성, 분류 및 색인, 보존, 사용 및 보급, 평가의 6가지 활동에 사용할 수 있는 도구가 제공됨을 알 수 있다.

## 5. 규범 및 정책

디지털 큐레이션을 지속적이고 효과적으로 시행하기 위해서는 활동을 위한 정책을 성문화하는 것이 바람직한데, 본 논문에서 분석한 5개의 모델을 제안한 리포지토리 가운데서는 DCC에서 연구데이터 관리에 관한 정책을 서면으로 제공하고 있다. DCC에서 제공하는 정책은 연구데이터를 대상으로 하는 내용 뿐만 아니라 자금 제공자 및 기관의 데이터에 관한 내용도 포함하고 있다. 또한 이와 같은 정책을 쉽게 검색하고 활용할 수 있는 지침 및 도구도 함께 제공한다. 아래의 <표 5>는 DCC가 제공하고 있는 정책을 각 영역별로 정리하여 보여준다.

<표 5> DCC 디지털 큐레이션 정책

	데이터 관리 및 공유 계획	리포지토리 및 데이터 센터 정책	부서 및 기관 정책
정책 분야	개념화	평가	데이터 생성
	데이터 생성	수집	메타데이터
	메타데이터	메타데이터	평가
	저장	저장	저장
	보존	보존 및 변환	보존
	접근 및 재사용	접근 및 재사용	접근 및 재사용

뿐만 아니라 여러 리포지토리는 그들이 제안하고 있는 디지털 큐레이션 모델을 최대한 효율적으로 활용할 수 있도록 표준 또는 모범사례집을 제공하는 경향이 있다. DDI는 모범사례에 대한 문서와 여러 표준을 간략하게 정리하고 제공하는 코드북을 제공한다. DCC의 경우에는 특히 메타데이터에 중점을 둔 표준을 제공하고 있으며, DataONE은 데이터베이스를 통한 디지털 데이터 생애주기의 모든 단계에서 데이터를 효과적으로 사용하는 방법에 대한 권장 사항을 개인에게 제공하고 있는데, 이는 DDI에서 제공하는 모범사례와 유사한 형태라고 할 수 있다.

## IV. 데이터 큐레이션 활동 현안

앞의 3장에서는 활동이론을 바탕으로 여러 연구데이터 큐레이션 모델을 분석하였다. 본 장에서는 실제 도서관 현장에서 연구데이터 큐레이션 활동을 수행하는데 논의가 필요한 현안을 분석한다. 본 연구는 디지털 큐레이션 분야에 있어서 특히나 IJDC(The International Journal of Digital Curation), JASIST(Journal of the Association for Information Science and Technology), Elsevier 등을 활용하여 ‘연구데이터 큐레이션’, ‘학술도서관에서의 연구데이터 관리’, ‘디지털 큐레이션 서비스’ 등을 주제로 하여 연구데이터 큐레이션 서비스 현안과 관련된 약 20여 개의 논문을 살펴보았으며

이를 바탕으로 최근 10년간 논의되었던 내용을 정리하여 제시한다. 본 장에서는 연구데이터 큐레이션과 관련한 여러 문헌 분석을 바탕으로 총 6가지의 현안에 대한 논의를 진행하였다.

## 1. 디지털 큐레이션 정책의 부재

활동이론에 따르면 정책과 절차는 규칙 및 규범을 통해 활동 주체의 행동을 증재한다. 조직 및 커뮤니티의 정책 및 절차도 이와 같은 범주에 속할 수 있으며 연구데이터 큐레이션 활동에 있어서 주체를 증재하는 메커니즘으로 정책 및 절차가 활용될 수 있다. 연구데이터를 체계적으로 관리하고 서비스하기 위해서는 각 리포지토리의 특성에 맞는 성문화된 정책이 필요한데, 실제적으로는 디지털 큐레이션에 대한 성문화된 정책을 구축하고 있는 리포지토리가 많지 않은 것이 문제라고 할 수 있다. 디지털 큐레이션 활동에 대한 명확하고 적절한 정책이 존재하지 않거나 디지털 큐레이션 정책 대신 보존 정책을 큐레이션 정책에도 적용하여 사용하는 경우에는 효과적인 디지털 큐레이션 서비스를 제공하기가 어렵다. 특히나 데이터 관리 및 서비스에 대한 정책을 따로 두어야 함에도 불구하고 데이터 보존에 더욱 중점을 두어 정책을 구축하는 경우가 많으며, 큐레이션과 보존 정책을 혼합하여 모호하게 적용하는 경우도 많은 실정이다(Faniel & Zimmerman, 2011). 디지털 큐레이션은 보존을 포함하는 광범위한 활동으로, 엄밀히 따지면 보존과는 다른 개념이기 때문에 디지털 큐레이션 활동 자체에 대한 정책 마련이 필요하다.

## 2. 연구데이터 큐레이션을 위한 표준의 부재

연구데이터 큐레이션을 수행함에 있어서 표준(standard)과 관련한 문제는 파일 형식 또는 메타데이터의 생성과 관련이 있다. 정책 및 절차 이외에도 연구데이터 큐레이션 활동에서 활동 주체와 리포지토리 사이의 상호작용은 메타데이터 및 파일 형식 표준을 매개로 이루어진다. 그럼에도 불구하고 연구데이터 큐레이션과 관련하여 널리 공유되고 있는 표준이 부족한 실정이다. 메타데이터가 불분명하고 검증할 수 없는 경우에는 이를 통해 데이터를 검색하기 어렵고 접근을 방해할 수 있기 때문에 명확한 메타데이터의 입력은 디지털 보존 및 큐레이션에서 매우 중요한 부분이라고 할 수 있다. 그렇기 때문에 가능한 경우에 데이터를 생성 및 수집하는 시점에서 동시에 메타데이터를 최대한 함께 수집하는 것이 중요하다(Burton & Treloar, 2009). 디지털 큐레이션의 정책 이슈와 마찬가지로 효과적인 디지털 큐레이션 서비스를 수행하고 제공하기 위해서는 여러 연구자들 및 현장의 실무자들 사이에서 공유되는 규범과 기준이 필요하다. TIFF 또는 PDF와 같은 일부 파일 형식은 기본 제공 메타데이터를 지원하고 있지만, 종종 매우 제한된 설명 요인만 지원하는 경우가 존재하고 종종 잘못된 메타데이터를 제공하기도 한다(Burton & Treloar, 2009). 메타데이터는

데이터 파일의 신뢰성 뿐만 아니라 접근 가능성을 보장할 수 있도록 도와주기 때문에 모든 종류의 데이터 변경 사항을 메타데이터에 정확히 입력할 수 있도록 할 필요가 있다.

이와 같은 중요성을 바탕으로 디지털 큐레이션 활동에 있어서의 표준을 수립하려는 다양한 노력들이 존재하는데, 그 중 하나로 자발적인 국제 표준을 개발하는 독립적인 비정부 회원 기구인 국제표준화기구(International Organization for Standardization, 이하 ISO)는 OAIS(Open Archival Information System) 참조모델을 발표하였다((Lee, 2010). 또한 DCC는 2002년에 처음 활동을 시작하여 디지털 큐레이션에 대한 표준 및 관행과 필요한 기술적인(technical) 도구 활용에 대한 지침을 제공하고 있다(Higgins, 2008). 그러나 국내에서는 디지털 큐레이션을 위한 단일 중점 국가 센터 혹은 협회가 존재하지 않으며 개별 연구자들, 혹은 프로젝트 형식으로 디지털 큐레이션을 위한 표준을 개발하고 모범사례를 정의하기 위해 노력하고 있다.

메타데이터 이외에도 파일 형식은 디지털 보존에서 계속적으로 논의되고 있는 문제이며 디지털 데이터의 장기적 보존을 위협하는 가장 큰 위험으로 간주된다. 그러나 이는 개방형 파일 형식의 높은 가용성으로 인하여 대부분의 파일 형식이 갈수록 더 많은 소프트웨어의 응용 프로그램으로부터 지원받을 수 있기 때문에 초기에는 큰 위험으로 인지되지 않았다(Houghton, 2016). 반면에, 독점 파일 형식은 해당 형식에 접근할 수 있는 소프트웨어의 사양을 공개적으로 제공받을 수 있는 가능성이 적기 때문에 지속적인 문제가 될 수 있다. 장기적으로 접근이 가능하도록 하는 호환 소프트웨어를 만들기 위해서는 특허권자의 허가가 필요하며, 이와 같은 이유로 비호환성을 갖는 파일 형식의 경우에는 파일에 대한 접근성이 급격하게 감소하게 되며 장기적 보존에 대한 복잡성이 증가하게 된다(Houghton, 2016). 이미지 파일에 사용되는 TIFF와 같은 일부 유형의 디지털 미디어에는 일반적으로 표준이라고 합의된 파일 형식이 존재한다. 그러나 비디오를 포함한 모든 미디어 유형의 데이터 보존을 위해 표준으로 합의된 파일 형식이 존재하지 않는다는 것이 논의사항이라고 할 수 있겠다(Godby & Denenberg, 2015).

### 3. 지속가능한 경영 모델의 부재

많은 형태의 학술도서관 또는 리포지토리가 연구데이터를 관리 및 보존하는 업무와 관련한 그들의 역할과 책임을 평가받기 시작하면서 동시에 디지털 큐레이션과 데이터 보존에 투입되는 비용을 고려하지 않을 수 없게 되었다. 연구데이터 큐레이션은 연구데이터의 수집, 관리, 보존, 폐기에 이르는 전 과정을 아우르는 활동으로, 성공적인 연구데이터 큐레이션 서비스를 제공하기 위해서 학술도서관 또는 리포지토리는 추가 예산이 필요할 수 있는 각 단계의 작업에 대한 전략 뿐만 아니라 관리자 배치에 대한 내용을 고려해야 한다. 학술도서관 또는 리포지토리에서 연구데이터 큐레이션과 같은 추가적인 활동에 대해 예산을 설계하고 재정을 계획하는 것은 쉽지 않을 일이

될 수 있으나 안정적인 큐레이션 서비스를 제공하기 위해서는 필수적으로 고려해야 할 사안이다. 그러나 대부분의 학술도서관 또는 리포지토리에서는 디지털 큐레이션 서비스에 대한 재정 계획을 거의 갖고 있지 않은 상황이다. 특히 국외에서는 일부 데이터 센터를 중심으로 장기적인 데이터 관리 작업 및 디지털 큐레이션 활동을 유지할 수 있도록 하기 위한 실용적인 경영 모델을 개발하기도 하였지만 학술도서관 및 리포지토리의 연구데이터 큐레이션을 위한 지속가능한 경영 모델은 일반적으로 많이 부족한 실정이다(Steinhart et al., 2008). 연구데이터의 성공적인 중장기적인 접근성 제고와 데이터의 재사용을 증대시키기 위해서는 재정 계획을 포함한 실질적인 경영 모델이 필요하지만, 현재 도서관 현장에서는 이와 같은 경영 모델이 부족한 것이 문제점이라고 할 수 있다.

#### 4. 연구데이터 큐레이션을 위한 연구자의 새로운 역할에 대한 이해 부족

디지털 큐레이션을 수행하는 주체로서 연구자들에게는 새로운 역할이 요구된다. 연구데이터 관리에 대한 중요성이 증대하면서 국내외 모두에서 국가 지원을 받는 연구에서는 연구 결과를 오픈 액세스 형태로 제공할 것을 요구하거나, 연구데이터에 대한 체계적인 관리를 요구하고 있다(과학기술기본법, 2020). 이와 같은 과정에서 이전과는 달리 연구자들은 연구데이터를 공유하고 관리하기 위해 다양한 규범, 표준 및 도구 등을 이해하고 사용할 수 있어야 한다. 효과적인 디지털 큐레이션 프로세스를 구현하기 위해서는 활동의 주체인 연구자들의 적극적인 참여가 요구되지만 대부분의 연구자가 디지털 큐레이션 프로세스 활동 자체에 참여하는 경우는 드문 편이고 디지털 큐레이션 자체에 대한 이해도가 낮은 경우도 존재한다.

모든 연구데이터에 대한 디지털 큐레이션 서비스를 제공하기 위해서는 연구 자체에 대한 면밀한 이해가 선행되어야 하며(Constantopoulos et al., 2009; Higgins, 2008), 데이터를 수집하는 과정에서 연구자들은 디지털 큐레이터와 수집 대상 데이터에 대한 평가 및 선별, 수집 범위 및 방법에 대해 논의할 필요가 있다. 뿐만 아니라 연구자들은 본인의 연구데이터를 기관 또는 리포지토리에 제출한 이후에도 그들의 데이터가 지속적으로 업데이트될 수 있도록 시스템을 모니터링 해야 한다.

또한 이용자가 연구데이터에 접근하고 활용할 수 있도록 하기 위해서는 데이터 공유가 필수적이다. 그러나 데이터 공유에 대한 결정권은 개별 연구자가 갖고 있으며 다양한 요인들이 데이터 기탁에 관련한 연구자들의 결정에 영향을 미칠 수 있다(Davis & Connolly, 2007). 뿐만 아니라 도서관 및 리포지토리에서 연구데이터를 제공하기 위해 데이터 공유 작업을 지속적으로 수행하고 지원하더라도, 연구자는 스스로 데이터를 보관하고 메타데이터를 입력하는데 요구되는 시간, 지식 또는 기술이 본인에게 없다고 느낄 수 있다는 문제점이 존재한다(Wallils, Rolando, & Borgman, 2013). 또한 데이터를 공유하는 것은 여러 이해 관계자들의 동기(motivation)와 보상체계에 따라 달라질 수 있음을 고려해야 한다(Borgman, 2012).

끝으로 디지털 큐레이션을 수행하기 위한 디지털 큐레이터와 연구자들의 역할 및 역량이 모호하다는 점도 문제점으로 지목되고 있다. 미국에서는 노스캐롤라이나 대학, 존스홉킨스 대학과 같은 일부 대학에서 디지털 큐레이션 과정과 함께 수료증을 제공하고 있으나, 학교별로 요구되는 과정에는 차이가 존재하며 공통적으로 요구되는 필수 과정 등에 대한 기준이 성립되어 있지 않아 현장에 배출되는 디지털 큐레이터들의 역할 및 역량에 차이가 존재하는 문제가 발생한다(Fulton, Botticelli, & Bradley, 2011).

## 5. 원본데이터의 보존과 기밀유지의 상충성

디지털 데이터는 데이터의 특성상 자연적으로 손실의 위험을 갖고 있다. 따라서 데이터를 보존하는 방법에 있어서 일반적으로 디지털 데이터를 복제하여 여러 위치에 저장하고 백업하여 장기간 보존하도록 노력한다(Sheldon, 2013). 연구데이터 큐레이션은 발표되어진 연구의 결과물뿐만 아니라 연구 과정에서 생성된 원시데이터(raw data) 및 부산물을 그 대상으로 한다. 이와 같은 이유로 연구데이터 큐레이션은 그 대상을 연구 중에 생산된 설문 조사 자료, 인터뷰 자료 등의 원시데이터를 포함하게 된다. 연구자는 해당 연구에 참여하는 참여자의 비밀을 보호하기 위하여 설문 응답, 인터뷰 등의 원시데이터를 일정 기간 보관한 후 파기하는 것이 보통이다. 이 과정에서 데이터 보존과 기밀 유지 사이에 모순이 발생하게 되는데, 원시데이터 및 부산물을 포함한 연구데이터의 장기 보존은 기밀 유지라는 측면에서 위험을 증가시키게 된다. 반면에 기밀성을 우선 순위로 고려할 경우에는 많은 연구의 원시데이터를 폐기해야 하기 때문에 장기적 보존에 있어서 어려움을 겪게 된다. 데이터의 장기적 보존을 위해서 디지털 큐레이터는 다양한 저장소의 유형을 파악하고 그 특성에 따라 데이터 유형에 맞는 적합한 방식으로 저장할 수 있어야 한다. 그와 동시에 디지털 데이터를 보존하고 이용자가 데이터를 지속적으로 사용할 수 있도록 하는 과정에 있어서 고려해야 할 중요한 문제 중 하나는 기밀성과 저작권의 문제이다(Abowd & Lane, 2004). 특히나 개인을 대상으로 하여 수집된 사회 및 의료 데이터의 수집 및 사용에 있어서 응답자의 사생활을 보호하는 문제는 연구자들 사이에서 지속적으로 논의되고 있다. 또한 디지털 데이터는 복제 및 유통이 용이하기 때문에 저작권 문제에 대해 특히 예민하게 반응할 필요가 있다(Abowd & Lane, 2004).

## 6. 보존과 비용의 상충성

가능한 한 완전한 데이터의 형태로 장기간 보존한다는 디지털 큐레이션의 목표와 해당 활동의 총 비용을 줄이는 것에는 상충이 발생하게 된다. 디지털 데이터의 양이 증가함에 따라 보존을 위해서 수행되는 데이터를 대상으로 한 선택과 평가 그리고 보존 및 폐기는 저장소 관리의 중요한



활동 중 하나로 간주된다. 이 과정에서 수집된 모든 디지털 데이터를 원본 상태 그대로 보존해야 한다는 전략과 디지털 큐레이션 활동의 총 비용을 합리적으로 유지하기 위해 평가와 폐기 혹은 데이터를 편집하여 용량을 줄이는 활동을 해야 한다는 전략 사이에 모순이 발생하게 된다.

기존의 연구에 따르면 데이터를 선택하고 평가하는데 있어서 두 가지의 주요한 관점이 존재한다(Harvey, 2007). 첫째는 물리적인 데이터를 보존할 때와 비교할 때 컴퓨터 저장소를 활용함으로써 비용 절감을 이루게 된 것을 바탕으로 데이터 보존에 있어서 선택과 평가를 최소화하고 모든 디지털 데이터를 최대한 원본 상태로 유지 및 보존하는 것이 실용적이라는 관점이다(Harvey, 2007). 그러나 이 관점에서는 메타데이터 생성 및 마이그레이션과 같은 활동에 있어서 발생할 수 있는 추가적인 비용을 고려하지 않는다는 한계가 있다. 때문에 또 다른 관점에서는 디지털 데이터는 데이터의 성격 및 생명주기에 따라 선택, 평가, 보존 및 폐기되어야 한다고 주장한다(Harvey, 2007).

연구데이터는 그 형태와 형식이 매우 다양하며, 접근 권한이나 용도는 연구자의 의도나 연구의 특성에 따라 달라질 수 있다. 따라서 연구데이터를 선택하고 평가하여 보존 또는 폐기를 결정하기 위해서는 보존 기관의 사명 및 목표, 법적 의무를 고려한 성문화된 표준 및 정책이 필요할 것이다. 뿐만 아니라 이에 대한 논의도 지속적으로 이루어져야 할 필요성이 있다.

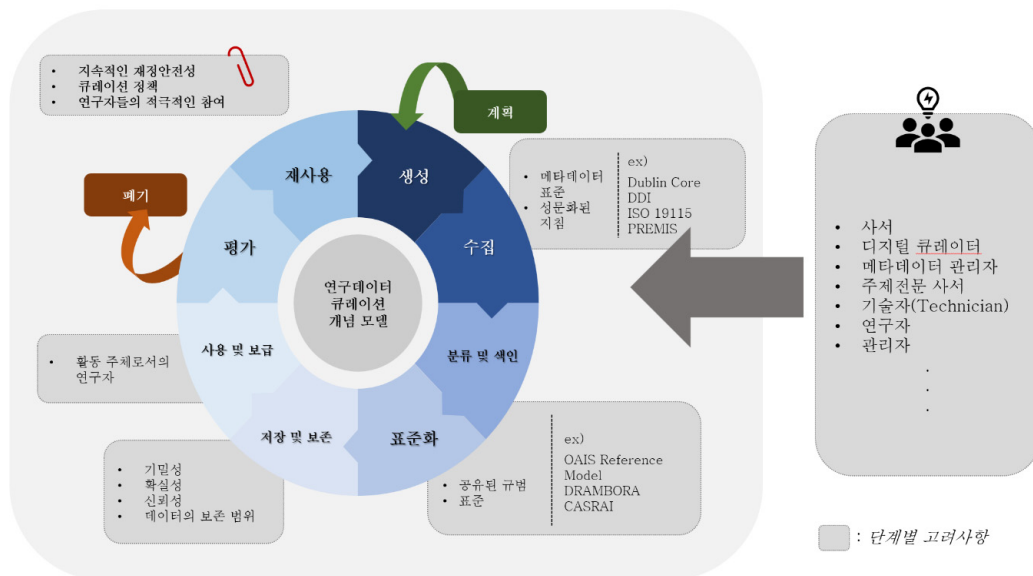
## V. 연구데이터 큐레이션 개념 모델 제안

앞 장에서는 활동이론을 이론적 틀로 활용하여 여러 리포지토리에서 제안한 디지털 큐레이션 생애주기 모델을 분석하였으며, 문헌 분석을 통해 도서관 현장에서 연구데이터 큐레이션 활동을 수행할 때 고려해야 할 현안들에 대해 논의하였다. 본 장에서는 디지털 큐레이션 모델을 분석한 내용과 현장에서 고려해야 할 사안들을 종합하여 연구데이터 큐레이션의 새로운 개념 모델을 제안한다. 앞서 분석한 5개의 디지털 큐레이션 생애주기 모델은 다양한 리포지토리에서 각각 제안되었으며, 해당 모델들은 리포지토리의 특성에 따라 큐레이션의 대상 데이터나 서비스 활동에 있어서 차이가 존재하였다. 일부 리포지토리는 빅데이터 또는 과학연구데이터와 같은 데이터의 관리에 초점을 맞춘 큐레이션 모델을 제안한 반면 또 다른 리포지토리에서는 메타데이터의 생성 또는 데이터의 장기적 보존에 중점을 둔 모델을 제안한 바 있다.

리포지토리 자체의 목표를 기반으로 하여 제안된 일부 디지털 큐레이션 생애주기 모델은 제공해야 하는 큐레이션 활동보다 데이터 자체의 생애주기에 더 초점을 맞추는 경우가 존재하였는데, 본 연구에서 제안하는 연구데이터 큐레이션 개념 모델은 비교적 최근까지 제안된 디지털 큐레이션 모델들로부터 분석하여 추출된 공통 구성 요인을 기반으로 구축되었다. 또한 앞서 분석한 각 디지털 큐레이션 모델의 특성에 따라 산발적으로 제안되었던 행위요인을 하나의 모델로 통합하여 종합

모델을 제시하는 것을 목표로 한다. 본 연구에서 제안하는 연구데이터 큐레이션 개념 모델은 데이터 생애주기를 기반으로 구축되었으나 동시에 각 활동의 체계에서 고려해야 할 실천적 문제와 모순을 포함한 개념 모델로서 대상 데이터를 연구데이터로 한정하고 있으나 특정 연구 분야에만 초점을 맞추고 있지 않기 때문에 비교적 넓은 범위의 큐레이션 서비스 활동에 적용할 수 있을 것으로 기대한다.

〈그림 2〉는 연구데이터 큐레이션 생애주기 모델과 각 단계에서 발생할 수 있는 실천적 문제 또는 모순을 결합한 개념 모델을 나타낸다.



〈그림 2〉 연구데이터 큐레이션 개념 모델

제안하는 연구데이터 큐레이션 개념 모델은 계획, 생성, 수집, 분류 및 색인, 표준화, 보존, 사용 및 보급, 평가, 재사용, 폐기의 10단계로 구성되는데, 해당 단계는 본 연구의 3장에서 실시한 디지털 큐레이션 모델 분석을 통해 추출되었다. 이는 각 모델에서 나타나는 행위 요소들을 정리하여 제시한 것이다(〈표 2〉 참고). 생애주기의 외부에 존재하는 요인들은 활동이론을 기반으로 하여 도서관 현장 또는 리포지토리에서 연구데이터 큐레이션 활동을 수행함에 있어서 발생할 수 있는 잠재적인 문제를 나타낸다.

먼저 연구데이터 큐레이션 활동의 전반적인 과정에 있어서 데이터 큐레이션 및 보존을 지속적으로 수행하기 위해서는 재정적 안정성을 고려해야 한다. 특히 보존 활동에 있어서는 컴퓨터를 활용하여 저장 비용을 절감하게 되면서 최대한 모든 디지털 데이터를 유지 및 보존하려는 관점이

존재하는데, 동시에 저장소는 컴퓨터를 활용한 저장 비용뿐만 아니라 메타데이터 생성 및 마이그레이션과 같은 추가 비용도 고려해야 한다는 점을 염두해야 한다. 또한 연구데이터 큐레이션 과정의 모든 활동에는 각 리포지토리의 특성에 맞는 성문화된 정책이 필요한데, 디지털 큐레이션은 보존을 포함하는, 보존과는 다른 개념의 광범위한 활동이기 때문에 성문화된 정책은 디지털 큐레이션 자체에 대한 독립적인 정책이 구축되어야 한다.

연구데이터를 생성하고 수집하는 단계에서는 데이터를 행정적(administrative), 서술적(descriptive), 구조적(structural), 기술적(technical) 메타데이터를 입력하여 생성해야 한다. 대개 학술도서관 및 리포지토리는 Dublin Core, PREMIS(PREservation Metadata: Implementation Strategies)와 같은 다양한 메타데이터 표준을 활용하는 경향이 존재한다.

생성 및 수집된 데이터는 분류 및 색인을 통해 접근성을 고취시킬 수 있도록 하고, 표준화를 통해 각 데이터의 효용성을 증대할 수 있도록 하는데, 이 과정에서 공유된 규범이 존재하는 것이 바람직하다.

저장 및 보존의 단계에서 디지털 데이터는 그 특성상 데이터 손실의 위험이 크기 때문에 장기적 보존은 연구데이터 큐레이션 활동의 핵심 부분 중 하나가 된다. 동시에 디지털 형태로 된 연구데이터를 보존함에 있어서 보존의 문제뿐만 아니라 원시데이터가 담고 있을 수 있는 연구 참여자들의 프라이버시와 연구자들의 지적 재산을 보호하기 위해서 기밀성과 저작권 문제를 고려해야 한다.

또한 데이터의 활용 단계에서는 연구자가 연구데이터의 생산자이자 사용자로서 연구데이터 큐레이션 활동 과정 자체에 참여하는 것이 매우 중요하다. 기존의 전통적인 도서관의 이용자와 달리 연구자는 본인의 연구데이터를 소속 기관 또는 리포지토리에 제출한 후 해당 연구데이터의 지속적인 업데이트를 위해 시스템을 모니터링하고 데이터를 관리할 필요가 있다. 이는 평가의 단계에서 이루어지며 평가를 통해 데이터를 폐기시킬 것인지를 결정할 수 있고, 장기적 보존을 유지함으로써 재사용을 장려할 수도 있다.

뿐만 아니라 연구데이터 큐레이션 활동을 수행함에 있어서 디지털 큐레이터의 역할과 역량에 대해 논의해야 할 필요가 있다. 디지털 큐레이션이라는 용어와 디지털 큐레이터라는 이름이 등장한 지는 꽤 시간이 흘렀지만 여전히 그 용어와 역할에 대해서는 모호함이 존재한다. 또한 큐레이션 활동의 규범과 표준에 있어서도 유사한 문제가 남아있다. 그렇기 때문에 효과적인 연구데이터 큐레이션 활동을 위해서는 큐레이션과 큐레이터의 기준을 정립하기 위한 많은 노력이 여전히 필요하다.

본 연구에서는 연구데이터의 생애주기와 각 단계에서 고려해야 할 논의사항을 포함한 연구데이터 큐레이션 개념 모델을 제공함과 동시에 각 활동의 단계에서 디지털 큐레이터 혹은 데이터 관리자가 수행해야 할 역할을 문헌 분석을 바탕으로 정리하여 제안한다.

앞선 논의에서 살펴본 것과 같이 디지털 큐레이션 활동에는 다양한 단계가 존재하며, 각 단계마다 디지털 큐레이터 또는 데이터 관리자가 수행해야 하는 역할이 존재한다. Johnston(2017)은 DCC

생애주기 모델에 제시된 순차적 활동을 기반으로 디지털 큐레이터의 역할과 성과를 제안한 바 있다. 디지털 큐레이터는 데이터의 생애주기에 맞게 각 단계별로 역할을 수행할 필요가 있으며, 해당 연구에서는 총 10단계의 연구데이터 큐레이션 프로세스를 제안하였다. 아래의 <표 6>은 본 연구에서 제안하는 연구데이터 큐레이션 개념 모델의 단계별 활동에 적용하여 디지털 큐레이터의 역할을 요약 및 정리한 내용을 보여준다.

<표 6> 각 활동단계별 디지털 큐레이터의 역할

순차적 활동	각 활동단계별 디지털 큐레이터의 역할
계획	데이터 큐레이션 서비스가 적절한 인력 및 경영 모델을 통해 유지될 수 있도록 함.
생성	데이터 생산자가 데이터 저장소(기관 또는 리포지토리)에 보관할 데이터를 제출할 수 있도록 지원함.
수집	적절한 리포지토리 환경을 활용하여 안전한 위치에서 데이터를 수집 및 저장할 수 있도록 함.
분류 및 색인	적절한 구조화를 통해 색인어 간의 계층/연관 관계를 표현하여 데이터를 배열할 수 있도록 함.
표준화	적절한 설명 메타데이터를 작성 및 제공함으로써 데이터 생산자가 제출한 메타데이터를 개선하여 검색을 용이하게 함.
저장 및 보존	데이터 변환을 최적화하기 위해 제출된 모든 데이터에 대한 처리 작업을 포함하여 성문화된 절차에 따라 데이터를 저장할 수 있도록 함.
사용 및 보급	검색, 배포, 다운로드 등의 기능을 통해 데이터에 대한 접근성을 강화할 수 있도록 함.
평가	수집 정책의 기준에 충족하지 않는 데이터를 수정 또는 폐기하고, 디지털 데이터가 내재하고 있는 위험성을 완화하기 위해 제출된 데이터를 재검토하도록 함.
재사용	검색, 배포, 다운로드 등의 기능을 통해 데이터에 대한 접근성을 강화할 수 있도록 함.
폐기	데이터의 영향 또는 가치를 평가하여 해당 데이터의 보관 또는 폐기 여부를 결정함.

## VI. 결론 및 제언

본 연구는 활동이론을 이론적 틀로 활용하여 연구데이터 큐레이션 모델을 분석한 문헌 연구로, 활동이론에서 사용하는 활동의 구성 요소들을 바탕으로 여러 연구데이터 큐레이션 모델을 분석했을 뿐만 아니라 연구데이터 큐레이션 활동을 수행하는데 있어서 도서관계에서 논의가 필요한 현안을 분석하였다. 그리고 이를 기반으로 새로운 연구데이터 큐레이션 개념 모델을 제안하였다.

본 연구는 기존에 제안된 5개의 디지털 큐레이션 생애주기 모델(DDI 모델, DCC 모델, ANDS 모델, DataONE 모델, 빅데이터 생애주기 모델)이 어떻게 구성되어 있는지 분석하였다. 각각의 큐레이션 모델은 데이터 생애주기를 기반으로 구축되었으며, 공통적으로 저장 및 보존 활동을 가장 중요한 단계로 인식하고 있음을 알 수 있다. 뿐만 아니라 디지털 큐레이션 모델에서 활동의 주체인 연구자의 역할을 강조하고 있는데, 이는 연구데이터의 생산자이자 소비자로서의 연구자 역할을 의미한다. 본 연구는 기존에 제안된 디지털 큐레이션 모델에서 산발적으로 제시되는 활동요인들을

분석하여 공통 요인을 추출하고 새로운 모델로 통합하여 새로운 연구데이터 큐레이션 개념 모델을 제안하였다. 해당 모델은 계획, 생성, 수집, 분류 및 색인, 표준화, 저장 및 보존, 사용 및 보급, 평가, 재사용, 폐기의 총 10단계로 이루어져 있다.

또한, 도서관 및 리포지토리에서 연구데이터 큐레이션 활동을 수행함에 있어서 고려해야 할 현안들을 분석하였다. 본 연구에서는 보존 정책과 병용되지 않는 큐레이션 자체를 대상으로 하는 정책의 부재, 연구데이터 큐레이션을 위한 표준의 부재, 재정적 안정성에 대한 고민의 필요성, 디지털 큐레이터의 역할과 역량에 대한 명확한 정의, 원본데이터를 보존하는 문제와 참여자 기밀 유지의 상충성 등에 대해 논의하였다. 해당 논의는 미국의 학술도서관을 주요 대상으로 한 문헌을 분석하였기 때문에 국내 실정을 정확히 반영하지 못한다는 한계점을 지니고 있으나, 상대적으로 국내와 비교할 때 미국을 포함한 국외의 학술도서관에서 디지털 큐레이션 서비스가 보다 활발하게 진행되고 있음을 고려할 때 우리나라에서 역시 고려해볼만 한 논의 사항일 수 있다. 해당 논의는 연구데이터 큐레이션에 서비스에 관하여 현장 중심의 관점을 제공하고 있다는 점에서 의의를 갖는다.

본 연구에서 제안하는 연구데이터 큐레이션 개념 모델은 연구데이터를 대상 데이터로 하여 연속적 단계를 포함하는 생애주기 모델임과 동시에, 해당 모델에 각 활동의 단계에서 고려해야 할 실천적 문제와 모순을 포함함으로써 실제 현장에 적용함에 있어 제고된 활용성을 가질 수 있을 것으로 기대한다.

## 참 고 문 헌

- 과학기술기본법. 법률 제18727호.  
국가과학기술연구회 (2019). 연구데이터 관리 가이드라인 (2019-07).  
국가연구개발혁신법. 법률 제18645호.  
김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455. <https://doi.org/10.16981/kliss.43.3.201209.433>  
Abowd, J. M. & Lane, J. (2004, June). New approaches to confidentiality protection: Synthetic data, remote access and research data centers. In International Workshop on Privacy in Statistical Databases. Springer, Berlin, Heidelberg, 282-289.  
Bailey, C. W. (2013). Research data curation bibliography. Digital Scholarship.  
Ball, A. (2012). Review of data management lifecycle models. University of Bath, IDMRC.  
Borgman, C. L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6), 1059-1078.

- Burton, A. & Treloar, A. (2009). Designing for discovery and re-use: the 'ANDS data sharing verbs' approach to service decomposition. *International Journal of Digital Curation*, 4(3), 44-56.
- Constantopoulos, P., Dallas, C., Androutopoulos, I., Angelis, S., Deligiannakis, A., Gavriliis, D., ... & Papatheodorou, C. (2009). DCC&U: An extended digital curation lifecycle model. *International Journal of Digital Curation*, 4(1), 34-45.
- Davis, P. M. & Connolly, M. J. (2007). Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace.
- Engeström, Y. (1987). *Learning by Expanding*. Cambridge University Press.
- Engeström, Y. (2001). Expansive learning at work: toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14(1), 133-156.
- Faniel, I. M. & Zimmerman, A. (2011). Beyond the data deluge: a research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, 6(1), 58-69.
- Fulton, B., Botticelli, P., & Bradley, J. (2011). DigIn: a hands-on approach to a digital curation curriculum for professional development. *Journal of Education for Library and Information Science*, 95-109.
- Godby, C. J. & Denenberg, R. (2015). *Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC*. OCLC Online Computer Library Center, Dublin.
- Harvey, R. (2007). Appraisal and selection. *Curation Reference Manual*. Digital Curation Centre.
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134-140.
- Houghton, B. (2016). Preservation challenges in the digital age. *D-lib Magazine*, 22(7/8), 1-6.
- Johnston, L. (2017). *Curating Research Data: A Handbook of Current Practice*. Chicago: Association of College and Research Libraries.
- Kaptelinin, V. (2005). The object of activity: making sense of the sense-maker. *Mind, Culture, and Activity*, 12(1), 4-18.
- Kaptelinin, V. & Nardi, B. (2012). Activity theory in HCI: fundamentals and reflections. *Synthesis Lectures Human-Centered Informatics*, 5(1), 1-105.

- Lee, C. A. (2010). Open archival information system (OAIS) reference model. *Encyclopedia of Library and Information Sciences*, 4020-4030.
- Leont'ev, A. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, New Jersey.
- Lompscher, J. (2006). The cultural-historical activity theory: some aspects of development. *Critical Perspectives on Activity: Explorations across Education, Work, and Everyday Life*, 35-51.
- Michener, W. K. & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, 27(2), 85-93.
- Pennock, M. (2007). Digital curation: a life-cycle approach to managing and preserving usable digital information. *Library & Archives*, 1, 34-45.
- Pouchard, L. (2015). Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation*, 10(2), 176-192.
- Sheldon, M. (2013). Analysis of current digital Preservation Policies. *Archives, Libraries and Museums*. [pdf] Available:  
<http://www.digitalpreservation.gov/documents/Analysis%20of%20Current,20>
- Steinhart, G., Saylor, J., Albert, P., Alpi, K., Baxter, P., Brown, E., ... & Lowe, B. (2008). Digital research data curation: Overview of issues, current activities, and opportunities for the Cornell University Library. A report of the CUL Data Working Group.
- Tamaro, A. M., Ross, S., & Casarosa, V. (2014). Research data curator: the competencies gap. *BOBCATSSS 2014 Proceedings*, 1(1), 95-100.
- Tolle, K. M., Tansley, D. S. W., & Hey, A. J. (2011). The fourth paradigm: data-intensive scientific discovery [point of view]. *Proceedings of the IEEE*, 99(8), 1334-1337.  
<https://doi.org/10.1109/JPROC.2011.2155130>
- Vardigan, M., Heus, P., & Thomas, W. (2008). Data documentation initiative: toward a standard for the social sciences. *International Journal of Digital Curation*, 3(1), 107-113.
- Vygotsky, L. S. (1962). *Thought and Word*. Cambridge, MA: MIT Press.

• 국한문 참고문헌의 영문 표기

(English translation / Romanization of references originally written in Korean)

Framework Act on Science and Technology. No. 18727.

Kim, Ji Hyun (2012). A study on university researchers' data management practices.

한국도서관·정보학회지(제54권 제1호)

Journal of Korean Library and Information Science Society, 43(3), 433-455.

<https://doi.org/10.16981/kliss.43.3.201209.433>

National Research and Development Innovation Act. No. 18645.

National Research Council of Science and Technology (2019). Research Data Management Guidelines (2019-07).