

GUIDELINES

AN

CHECK LISTS



KISTI



FOR



이혜림 저

DIGITAL

CURATION



## 디지털 큐레이션

가이드라인과

체크리스트

### 저자

-

이혜림

### 약력

-

저자는 이화여자대학교에서  
문헌정보학 전공으로 학사와 석사  
학위를 받았다. 미국에서는 미시건  
대학교에서 정보학 석사학위를  
받고, 피츠버그 대학교에서 정보학  
전공으로 박사학위를 받았다.

저자의 박사학위논문계획서는  
미국도서관협회(ALA)에서  
수여하는 LRRT Jesse H.  
Shera Award for the Support  
of Dissertation Research를  
수상하였다.

현재 저자는  
한국과학기술정보연구원의  
책임연구원으로 재직하고  
있으며, 연구원의 기관지이자  
국제학술저널인 Journal of  
Information Science Theory and  
Practice의 managing editor를  
5년간 역임했다.

GUIDE LINES

AN

CHECK LISTS



KISTI



FOR



이혜림 저

DIGITAL

CURATION

## 서문

데이터의 규모와 다양성이 증가함에 따라 효율적이고 체계적인 디지털 큐레이션의 필요성도 커지고 있다. 디지털 큐레이션은 데이터를 수집, 가공, 관리, 보존하여 이용자들에게 제공하고, 더 나아가 데이터에 가치를 추가하는 프로세스이다. 2000년대부터 미국, 캐나다, 오스트레일리아와 다수 유럽 국가의 정보전문가들이 디지털 큐레이션에 대한 연구를 수행해 왔고, 도서관이나 아카이브, 정보센터 등 정보기관들에서 디지털 큐레이션을 하고 있다. 이에 비해 우리나라에서는 아직까지 디지털 큐레이션에 대한 인식이 부족하다. 또한 이 주제에 대한 연구가 부족하고 디지털 큐레이션 실무 현황에 관해서는 알려진 바가 거의 없다. 그러나, 디지털 정보자원을 다루는 정보기관(도서관, 아카이브, 데이터센터 등)이라면 부분적으로나마 디지털 큐레이션 활동을 하고 있다고 말할 수 있다. 왜냐하면, 그러한 정보기관이라면 최소한 디지털 정보자원의 수집, 관리, 서비스를 수행하기 때문이다. 정보기관은 이러한 기존의 활동에 새로운 디지털 큐레이션 활동을 추가하고, 그러한 활동을 체계적으로 수행함으로써 효율적이고 성공적인 디지털 큐레이션 실무를 수행할 수 있다. 하지만, 아쉽게도, 우리나라에는 디지털 큐레이션 실무를 수행하는 데 참고할 만한 자료가 거의 없다.

이 책은 미국 유학 시절 습득한 디지털 큐레이션에 대한 저자의 지식과 프로젝트 수행 경험, 그리고 KISTI에서 디지털 큐레이션 관련 실무를 수행한 경험을 바탕으로 저술되었다. 이 책의 목적은 국내 디지털 큐레이션 실무자들이 참고 자료로 사용하는 것이다.

이 책은 크게 세 개의 장으로 구성되어 있다. 제1장에서는 ‘한국형 디지털 큐레이션 라이프 사이클 모델’을 제시하고, 제2장과 제3장은 각각 ‘디지털 큐레이션 가이드라인’과 ‘디지털 큐레이션 체크리스트’를 수록하고 있다. 제1장이 디지털 큐레이션에 대한 개념적 지식을 제공한다면, 제2장과 제3장은 실무차원의 지식을 제공한다.

이 책의 세 장은 유기적으로 연결되어 있는 데, 이 책을 보다 유용하게 이용하기 위해서는 ‘책의 구성과 사용법’ 부분을 참조하기 바란다. 저자는 이 책이 디지털 큐레이션 실무를 수행하는 데 조금이나마 도움이 되고, 국내에서 본격적으로 디지털 큐레이션을 수행하는 시발점이 되기를 바란다.

마지막으로, 이 책이 나오도록 여러모로 도움을 주신 KISTI 콘텐츠큐레이션센터 여러분께 감사드리며, 특히 이 책을 리뷰해 주신 황혜경 센터장님께 감사의 마음을 전한다. 또한 저자가 개발한 한국형 디지털 큐레이션 라이프 사이클 모델에 관하여 의견을 주신 국내외 많은 전문가들에게도 감사드린다.

## 이혜림

책임연구원  
콘텐츠큐레이션센터  
국가과학기술데이터본부  
한국과학기술정보연구원

## 책의 구성과 사용법

이 책은 크게 세 개의 장으로 구성되어 있다. 제1장에서는 ‘한국형 디지털 큐레이션 라이프 사이클 모델’을 제시하고, 제2장과 제3장은 이 모델을 기반으로 하여 작성한 ‘디지털 큐레이션 가이드라인’과 ‘디지털 큐레이션 체크리스트’를 수록하고 있다. 이외에도 부록으로 ‘주요 용어 해설’과 ‘디지털 큐레이션 관련 유용한 사이트’를 수록하고 있다. 독자들의 용어에 대한 이해를 돕기 위하여 필요한 경우 한글 용어 옆에 영어 스펠링을 부기하였다.

제1장에서는 저자가 약 2년 동안 국내외 라이프 사이클 모델에 관한 자료들, 디지털 큐레이션 관련 자료들, KISTI 콘텐츠큐레이션센터 직원들의 의견, 국내외 디지털 큐레이션 전문가들의 의견을 참고하여 개발한 ‘한국형 디지털 큐레이션 라이프 사이클 모델’을 소개한다. 이 모델은 체계적이고 효율적인 디지털 큐레이션을 하기 위해 필요한 기본적인 활동들을 도식화하여 그러한 활동들 간의 순서와 관계를 보여준다. 제1장은 한국형 디지털 큐레이션 라이프 사이클 모델을 전체적으로 이해하도록 돕고, 제2장과 제3장의 가이드라인과 체크리스트를 사용하는 데 길잡이 역할을 할 것이다.

제2장의 ‘디지털 큐레이션 가이드라인’은 실제로 디지털 큐레이션 업무를 수행할 때 참고할 수 있도록 하기 위해 작성한 것이다. 이 가이드라인은 독자들이 체계적이고 효율적인 디지털 큐레이션을 위해 수행해야 할 활동들에 대한 실무 수준의 지식을 제공한다. 또한 이것은 디지털 큐레이터들이 그들의 기관에서 큐레이션 기능을 더 잘 실현하거나 모범 실무(best practice)와 비슷한 수준의 실무를 수행할 수 있도록

지원하기 위하여 배경 개념과 방법을 제공한다. 이 가이드라인은 ‘한국형 디지털 큐레이션 라이프 사이클 모델’을 구성하고 있는 각각의 디지털 큐레이션 활동에 따라 서술하였다.

제3장의 ‘디지털 큐레이션 체크리스트’는 정보기관이 현재 자관의 디지털 큐레이션 역량, 인프라, 상황을 진단하는 데 도움을 주고 향후 디지털 큐레이션 계획을 세우는 데 도움을 주기 위한 것이다. 체크리스트도 제2장의 ‘디지털 큐레이션 가이드라인’과 마찬가지로 제1장의 ‘한국형 디지털 큐레이션 라이프 사이클 모델’을 기반으로 하여 모델의 각 요소인 디지털 큐레이션 활동에 따라 구성하였다. 각 체크리스트에 해당하는 디지털 큐레이션 활동을 향상시키기 위해 자세한 정보를 얻고 싶다면, 제2장 ‘디지털 큐레이션 가이드라인’에서 해당 활동에 관한 부분을 참고할 수 있다.

이 책은 제1장부터 제3장까지 순차적으로 읽을 수 있다. 또는 제1장을 읽고, 제3장의 체크리스트를 가지고 기관의 현황을 파악한 후, 제2장의 가이드라인을 보면서 기관의 디지털 큐레이션 실무를 향상시킬 수도 있다. 이 책의 세 장은 유기적으로 연결되어 있기 때문에 독자는 필요에 따라 세 장을 오가면서 이 책을 유용하게 사용할 수 있다. 예를 들어, 개별 활동(e.g., ‘입수’, ‘보존’)에 대한 내용을 세 개의 장에서 찾아 상호참조할 수 있다.

**저 자** 이해림

**펴 낸 곳** 한국과학기술정보연구원

**발 행 인** 최희운

**발 행** 2020년 12월 20일

**주 소** (본원) 대전광역시 유성구 대학로 245  
(분원) 서울특별시 동대문구 회기로 66

**전 화** 042-869-1615

**팩 스** 042-869-1694

**홈페이지** <https://www.kisti.re.kr>

**디 자 인** (주)차이 [www.chai.re.kr](http://www.chai.re.kr)

ISBN: 978-89-294-1167-1-93020

- 
- 이 책은 2020년도 한국과학기술정보연구원(Korea Institute of Science and Technology Information: KISTI) 주요사업과제(과학기술 콘텐츠 큐레이션 체제 구축: 연구과제번호 K-20-L01-C01-S01)로 수행된 연구 결과물입니다.
  - 이 책은 저작권법에 따라 보호받는 저작물이므로 무단전재와 무단복제를 금지하며, 이 책 내용의 전부 또는 일부를 이용하려면 반드시 저작권자와 한국과학기술정보연구원의 서면동의를 받아야 합니다.
  - 이 글의 내용은 저자의 견해이며, 한국과학기술정보연구원의 공식적인 의견이 아님을 밝힙니다.





## 디지털 큐레이션

가이드라인과

체크리스트

# CONTENTS

## 12

제1장  
한국형 디지털 큐레이션  
라이프 사이클 모델

- 전주기 활동
- 순차적인 활동
- 비정기적 활동
- 요약

## 22

제2장  
디지털 큐레이션  
가이드라인

- KOREA 모델의 전주기 활동
- KOREA 모델의 순차적인 활동
- KOREA 모델의 비정기적 활동

## 66

제3장  
디지털 큐레이션  
체크리스트

## 84

참고문헌

## 88

부록

- 1. 주요 용어 해설
- 2. 디지털 큐레이션 관련 유용한 사이트

**01**

디지털 큐레이션  
가이드라인과  
체크리스트



## 제1장

한국형

# 디지털 큐레이션 라이프 사이클 모델

전주기 활동

순차적인 활동

비정기적 활동

## 01

## 제1장.

## 한국형 디지털 큐레이션 라이프 사이클 모델

디지털 큐레이션은 데이터를 적극적으로 관리, 보존하고, 더 나아가 가치를 부가하여 현재와 미래에 이용, 재이용 될 수 있도록 만드는 작업이다. 그러한 디지털 큐레이션은 데이터의 수집, 관리, 보존, 서비스 활동; 데이터에 가치를 부가하는 활동; 데이터의 무결성, 신뢰성, 진본성, 유효성, 접근가능성, 이용가능성을 보장하는 활동 등을 포함해야 한다. 디지털 큐레이션은 데이터 큐레이션으로 불리우기도 한다.

디지털 큐레이션이라는 용어는 2001년 런던에서 개최된 ‘Digital Curation: Digital Archives, Libraries and E-science Seminar’에서 처음으로 사용되었다. 그 후 미국, 캐나다, 오스트레일리아와 다수 유럽 국가의 정보전문가들이 디지털 큐레이션에 대한 연구와 실무를 수행하고 있다. 반면 한국에서는 디지털 큐레이션이 아직 생소한 분야로, 이 분야의 발전을 위해 가장 시급하게 필요한 것은 프레임워크(framework)로 사용할 수 있는 디지털 큐레이션 라이프 사이클 모델(digital curation life cycle model)이라고 저자는 생각했다.

한국에서는 라이프 사이클(life cycle)을 ‘생명주기, 생애주기, 또는 수명주기’라고도 부르는데, 원래 라이프 사이클은 유기체의 탄생에서 죽음에 이르기까지 거치는 일련의 연속적인 단계를 말한다. 디지털 큐레이션 분야에서는 큐레이션 대상이 그 일생 동안 통과하는 일련의 단계를 말하는 은유적인 표현으로 사용된다. 그러한 단계를 그림으로 표현한 것이 바로 라이프 사이클 모델이다. 이미 많은 학자들과 실무자들이 라이프 사이클 모델이 디지털 큐레이션 분야에서 유용하다고 주장해 왔는데, 그 중 몇 가지 주장을 살펴보면 다음과 같다:

- 라이프 사이클 모델은 성공적인 큐레이션과 장기 보존에 필요한 과정, 활동, 관계를 시각화하여 보여준다 (Higgins, 2007).
- 조직의 데이터 서비스 계획/유지 관리/설명/지원을 용이하게 한다 (Cox & Tam, 2018; Humphrey, 2006; Pennock, 2007).
- 큐레이션 계획을 수립하고 실행할 때 필요한 모든 단계들과 그러한 단계들의 가장 적절한 순서를 보여준다(Higgins, 2007;Humphrey, 2006; Oliver & Harvey, 2016).

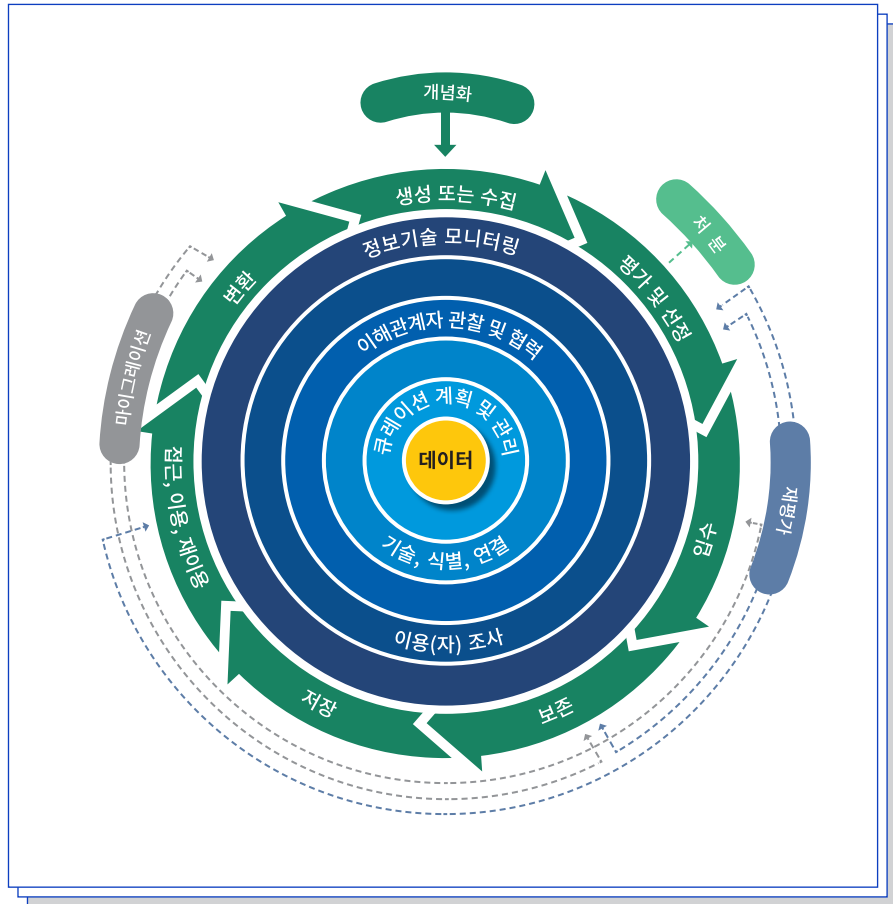
제1장에서는 저자가 약 2년 동안 개발한 ‘한국형 디지털 큐레이션 라이프 사이클 모델(KOREA Digital Curation Life Cycle Model: 이후 본 책자에서는 KOREA 모델이라 칭함)’을 소개한다. 이 모델은 체계적이고 효율적인 디지털 큐레이션을 위해 필요한 기본적인 활동들을 도식화하여 그러한 활동들 간의 순서와 관계를 보여준다. KOREA 모델의 개발 목적은 도서관이나 데이터센터 등 정보를 전문으로 다루는 정보기관에서 디지털 큐레이션 실무를 수행하는 데 프레임워크로 활용할 수 있도록 하는 것이다. 정보기관은 KOREA 모델을 프레임워크로 활용하여 다음과 같은 일을 할 수 있다:

- 기관의 디지털 큐레이션 계획과 전략을 세울 수 있다.
- 기관의 현행 디지털 큐레이션 실무의 강점과 약점을 파악할 수 있다.
- 성공적인 디지털 큐레이션 실무를 수행하기 위해 기관에서 필요한 큐레이션 활동과 단계를 파악할 수 있다.
- 디지털 큐레이션의 각 단계에서 수행해야 할 작업, 정책, 서비스를 파악할 수 있다.
- 기관의 상황에 맞는 고유한 디지털 큐레이션 라이프 사이클 모델을 개발할 수 있다.

저자는 국제적으로 유명한 영국 디지털 큐레이션 센터(Digital Curation Centre: DCC)의 ‘DCC 큐레이션 라이프 사이클 모델(이후 본 책자에서는 DCC 모델이라 칭함)’을 기반으로 하여 KOREA 모델을 만들었다. DCC 모델을 기반으로 한 가장 큰 이유는 이 모델이 세계적으로 많은 라이프 사이클 모델들에게 영향을 미쳤으며, 이러한 사실은 DCC 모델의 타당성과 신뢰성을 시사하는 것이라고 생각했기 때문이다.

더 나아가 저자는 KOREA 모델 개발을 위하여 국내외 디지털 큐레이션 관련 자료들, 연구데이터 관리에 관한 자료들, 라이프 사이클 모델에 관한 자료들을 참고하였다. 또한 회의, 전화 통화, 이메일 교환을 통해 많은 국내외 디지털 큐레이션 전문가들(e.g., DCC 직원, 대학 교수, 디지털 큐레이터)의 의견을 참고하였다. 저자는 2년여 동안 여러 버전의 모델을 만들었고, 국내외 학술대회(e.g., International Digital Curation Conference), 세미나, 포럼 등에서 당시에 개발된 버전의 모델을 발표하고 그에 대한 청중들의 의견을 들었다. 이러한 과정을 거쳐 확정된 모델이 아래의 KOREA 모델이다(그림 1 참조).

그림 1  
KOREA 디지털  
큐레이션 라이프  
사이클 모델  
(이 모델은 DCC  
Curation Lifecycle  
Model을 응용한  
것임)



KOREA 모델의 가장 안쪽에 있는 원은 디지털 큐레이션을 하는 대상을 나타낸다. 모델의 나머지 구성 요소들은 효율적인 디지털 큐레이션 실무 수행을 위해 필요한 디지털 큐레이션 활동들을 보여준다.

KOREA 모델의 큐레이션 대상은 광범위한 개념의 ‘데이터’이다. ‘데이터’는 디지털 형식의 정보로서 ‘디지털 객체’와 ‘데이터베이스’를 포함한다. 디지털 객체는 1) 텍스트, 이미지 또는 사운드 파일—그것들과 관련된 식별자와 메타데이터 포함—처럼 개별 디지털 객체; 2) 웹사이트처럼 여러 디지털 객체가 연관되어 만들어진 복합 디지털 객체를 포함한다. 데이터베이스는 컴퓨터 시스템에 저장된 레코드 또는 데이터의 구조화된 집합체를 말한다(Higgins, 2008, p. 137).

KOREA 모델의 디지털 큐레이션 활동들은 크게 ‘전주기 활동’, ‘순차적인 활동’, ‘비정기적 활동’의 세 카테고리로 나누어진다. 노란색 원과 초록색 원 사이의 파란색 계통의 원들은 전주기 활동들을 나타낸다. 전주기 활동들은 데이터의 라이프 사이클 전반에 걸쳐 지속적으로 행해지는 활동들을 말한다. 가장 바깥쪽 초록색 화살표로 표현된 순차적인 활동들은 라이프 사이클에서 일정한 순서대로 행해지는 활동들이다. 비정기적인 활동들은 어떤 조건하에서 이따금씩 일어나는 활동들로 큰 원 밖의 타원형으로 표현되었다.

## 전주기 활동

전주기 활동들은 데이터의 라이프 사이클 전반에 걸쳐 수행되어야 하는 지속적인 활동들로 아래와 같다.

- 1) **큐레이션 계획 및 관리:** 큐레이션 계획 및 관리는 전반적인 큐레이션 계획을 세우고, 계획의 진행과정을 모니터링하면서, 계획 달성을 위해 필요한 관리와 행정 업무를 수행하는 것이다.
- 2) **기술, 식별, 연결:** 기술은 데이터를 이해하고, 관리하고, 검색하여 접근할 수 있도록 메타데이터를 작성하는 것이다. 식별은 데이터에 고유 식별자를 부여하는 것을 말한다. 식별자를 이용하여 관련 있는 데이터를 연결할 수 있으며, 이를 통해 새로운 데이터셋이 생성될 수도 있다.
- 3) **이해관계자 관찰 및 협력:** 정보기관은 자관의 이해관계자들(e.g., 상위 기관, 출판사, 학회, 후원 기관)을 조사하고, 자관의 큐레이션 활동과 이해관계자들이 각각 어떤 관계를 가지고 있는 지 파악한다. 이해관계자들의 요구를 파악하고, 디지털 큐레이션을 위해 그들과 협력한다.
- 4) **이용(자) 조사:** 이용(자) 조사는 기관의 자원(e.g., 데이터, 디지털 큐레이터, 정보시스템)과 이용자들과의 상호작용을 조사하고, 기관의 잠재적 이용자들을 파악하는 것을 포함한다. 이러한 조사는 단순한 이용 통계 분석에서부터 과학적인 방법을 사용하는 이용자 연구까지 다양하게 이루어질 수 있다.
- 5) **정보기술 모니터링:** 정보기술은 디지털 큐레이션에 직접적으로 영향을 미치므로, 발전하는 정보기술을 지속적으로 모니터링하고 새로운 기술을 디지털 큐레이션 실무에 적용한다.

## 순차적인 활동

순차적인 활동들은 일련의 순서대로 수행되는 활동들로, 그 활동들이 일어나는 '단계'라고 생각할 수 있다. KOREA 모델의 순차적인 활동들은 아래와 같다.

- 1) **개념화**: 개념화는 기관의 디지털 큐레이션 과정과 결과물을 고려하여, 데이터 캡처 방법과 저장 필수 사항 등을 포함하는, 데이터의 생성과 수집에 관하여 구상하고 계획하는 것을 말한다.
- 2) **생성 또는 수집**: 생성은 기관에서 데이터를 만들고 그 데이터에 대한 식별자와 메타데이터를 부여하는 것이다. 수집은 법적 규정과 기관의 수집 정책에 따라 다른 조직들(e.g., 도서관, 아카이브, 학회, 데이터 센터)이나 개인 데이터 생성자들로부터 데이터를 획득하고, 필요에 따라 식별자를 부여하고 적합한 메타데이터를 작성하는 것이다.
- 3) **평가 및 선정**: 평가 및 선정은 문서화된 정책, 기준, 법적 요구 사항 등을 준수하면서 생성 또는 수집한 데이터를 평가하여 기관에서 장기간 큐레이션 할 대상을 선정하는 것이다.
- 4) **입수**: 입수는 데이터를 기관의 데이터베이스, 정보시스템 또는 디지털 아카이브에 추가하기 위해 준비하고 실제로 추가하는 행위들을 포함한다.
- 5) **보존**: 보존은 데이터의 무결성, 진본성, 신뢰성, 이용가능성을 유지하면서 데이터의 장기 보존을 위해 실행하는 행위들을 포함한다. 예를 들어, 유효성 검사, 보존 메타데이터 작성이나 변경 등이 이에 속한다.
- 6) **저장**: 저장은 관련 표준에 따라 안전한 방법으로 보관하는 것이다.
- 7) **접근, 이용, 재이용**: 접근, 이용, 재이용은 일상적으로 데이터가 이용, 재이용 될 수 있도록 데이터를 접근가능하게 유지하는 것을 말한다.
- 8) **변환**: 변환은 원본 데이터(original data)를 가지고 새로운 데이터를 만드는 것이다. 데이터를 마이그레이션 하여 새로운 포맷으로 만들거나, 특정 데이터셋을 가지고 새로운 서브데이터셋을 만드는 등의 행위를 포함한다.



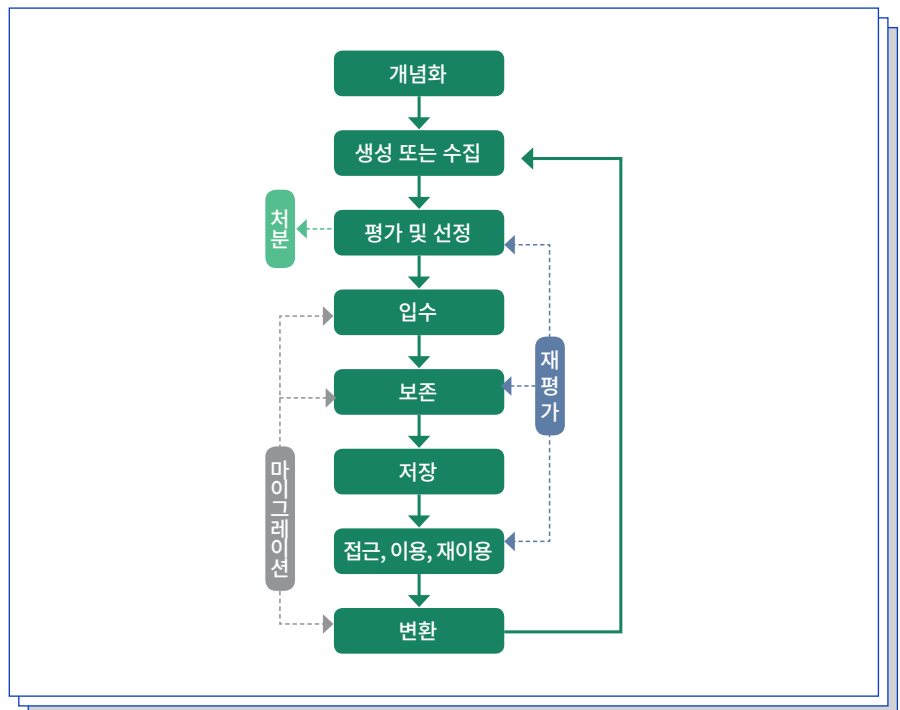
## 비정기적 활동

비정기적 활동들은 이따금씩 일어나는 활동들로 아래와 같다.

- 1) **마이그레이션**: 마이그레이션은 하나의 저장 시스템에서 다른 저장 시스템으로, 하나의 하드웨어/소프트웨어 구성에서 다른 하드웨어/소프트웨어 구성으로, 하나의 포맷에서 다른 포맷으로 데이터를 전송하는 행위들을 포함한다.
- 2) **처분**: 처분은 문서화된 정책, 지침, 법적 요구 사항에 따라 장기 큐레이션에 부적합한 데이터를 다른 정보기관이나 저장소로 이전하거나 폐기하는 것을 말한다.
- 3) **재평가**: 재평가는 시간이 지나 장기 큐레이션에 부적합해진 데이터를 처분하기 위해 보존 또는 이용되고 있던 데이터를 다시 평가하는 것이다.

<그림 1>에서 알 수 있듯이, KOREA 모델이 여러 겹으로 이루어져 있어 세 카테고리 활동들 간의 관계를 파악하기 힘들 수 있다. <그림 2>는 KOREA 모델에 대한 독자들의 이해를 돕기 위하여 KOREA 모델의 순차적 활동들과 비정기적 활동들만 추출하여 그 관계를 표현한 것이다.

**그림 2**  
KOREA 모델의  
순차적 활동들과  
비정기적 활동들  
간의 관계도



## 요약

KOREA 모델의 디지털 큐레이션 활동들 사이에는 서로 관련이 있다. 같은 카테고리에 있는 활동들뿐만 아니라, 서로 다른 카테고리에 있는 활동들 사이에도 관련이 있다. 예를 들어, 전주기 활동 중 하나인 '큐레이션 계획 및 관리'는 순차적인 활동들에 영향을 미친다. 다른 예로, '이용자 조사'와 '접근, 이용, 재이용'사이에도 상호관련이 있다. 이러한 사실이 보여주듯이, 특정 디지털 큐레이션 활동만 잘 수행한다고 해서 성공적인 디지털 큐레이션을 할 수 있는 게 아니라, KOREA 모델의 모든 디지털 큐레이션 활동들을 조화롭게 잘 수행해야 한다.

모든 모델이 그렇듯이 KOREA 모델 또한 완벽한 모델이 아니다. KOREA 모델이 많은 정보기관들에서 사용될 수 있도록 보편성을 추구하였으나, 각 기관의 자원(e.g., 인력, 예산)과 상황이 다르기 때문에 모든 정보기관들이 KOREA 모델을 있는 그대로 적용하기는 어려울 수 있다. 그러한 정보기관은 자관의 고유한 필요와 상황에 따라 KOREA 모델을 기반으로 하여 자관의 고유한 모델을 개발할 수 있다. KOREA 모델 또한 시간과 상황이 변하면서 점점 더 진화할 것이다.



**02**

디지털 큐레이션  
가이드라인과  
체크리스트



## 제2장

# 디지털 큐레이션 가이드라인

### KOREA 모델의 전주기 활동

큐레이션 계획 및 관리  
기술, 식별, 연결  
이해관계자 관찰 및 협력  
이용(자) 조사  
정보기술 모니터링

### KOREA 모델의 순차적인 활동

개념화  
생성 또는 수집  
평가 및 선정  
입수  
보존  
저장  
접근, 이용, 재이용  
변환

### KOREA 모델의 비정기적 활동

마이그레이션  
재평가  
처분

# 02

## 제2장. 디지털 큐레이션 가이드라인

‘모델’은 현실 세계에서 직접 관찰할 수 없는 것을 간단하게 도형으로 시각화 한 것이다. 제1장에서 다룬 한국형 디지털 큐레이션 라이프 사이클 모델(KOREA Digital Curation Life Cycle Model: 이후 본 책에서는 KOREA 모델이라 칭함)은 디지털 큐레이션 활동들에 대한 개념과 관계를 도형화한 것이기 때문에 독자들이 이러한 모델을 이해하고 실무에서 참고하기는 어려울 수 있다. 따라서 제2장에서는 KOREA 모델을 구성하고 있는 각각의 디지털 큐레이션 활동을 좀 더 상세히 서술하여 디지털 큐레이터들이 실무를 수행하는 데 도움을 주고자 하였다.

KOREA 모델에 있는 활동을 시작하기 전에, 정보기관이 해야 할 기관 차원의 필수 선행 활동이 있다. 그것은 정보기관이 디지털 큐레이션 실무를 수행할 수 있는 기본적인 역량과 환경을 갖추고 있는지를 검토하고 준비하는 ‘기관 운영 관리’이다. 본 가이드라인(guideline)은 이러한 활동을 포함하여 KOREA 모델을 구성하는 각각의 디지털 큐레이션 활동에 따라 구성되었기 때문에, 독자들은 가이드라인 전체를 참고할 수도 있고 특정 활동 부분만 선택하여 참고할 수도 있다. 본 가이드라인을 처음부터 끝까지 읽지 않더라도, 최소한 각 활동의 ‘개요’와 ‘주요 행위’ 부분은 읽어보기를 권한다.

가이드라인의 원래 정의대로, 본 가이드라인은 디지털 큐레이션 활동에 대하여 개략적으로 설명하고 있다. 아래에서 서술하고 있는 내용에 대해 좀 더 자세히 알고 싶다면, 이 책의 ‘참고문헌’과 ‘부록: 디지털 큐레이션 관련 유용한 사이트’에 수록되어 있는 자료들을 참고할 수 있다.

- 기관 운영 관리
- KOREA 모델의 전주기 활동
- KOREA 모델의 순차적인 활동
- KOREA 모델의 비정기적 활동

## 기관 운영 관리

### → 개요

‘기관 운영 관리’는 디지털 큐레이션에 관심이 있는 기관들이 KOREA 모델에 따라 디지털 큐레이션을 계획하고 본격적인 활동을 시작하기 전에 수행되어야 하는 선행 활동이다. 이것은 해당 기관이 디지털 큐레이션을 수행할 수 있는 여건을 갖추고 있는지를 확인하기 위해 기관의 역량과 환경을 점검하고 준비하는 활동이다. 다시 말해, 종합적인 발전 전략 수립부터 예산, 조직, 인력, 시스템, 시설설비 관리 등을 포함하여 기관의 전반적인 운영 관리를 파악한다. 특히 기관의 정책 프레임워크 개발; 기관 운영의 모든 영역을 포괄하는 표준의 개발; 적당한 자원과 인프라의 지속적인 공급; 기관의 운영을 모니터링하고 보고하는 관리 프로세스 운영에 관한 현 상황을 점검한 후, 부족한 부분을 파악하여 개선한다.

### → 주요 행위

- 기관 직원들에게 디지털 큐레이션의 필요성에 대한 인식과 합의를 이끌어낸다.
- 기관은 디지털 큐레이션을 수행하는 데 필요한 자원(e.g., 예산, 인력, 시설설비, 정보기술 인프라)을 파악하고, 부족한 자원을 추가로 마련할 수 있는지 검토한다.
- 기관의 종합적인 발전 전략에 디지털 큐레이션을 포함할 수 있는지 검토한다.
- 디지털 큐레이션 실무를 효율적으로 수행할 수 있도록 조직 개편을 고려한다.
- 기관의 전반적인 운영을 모니터링하고 보고하는 관리 프로세스를 운영한다.
- 기관의 자원을 효율적으로 사용하기 위한 자원 할당 정책과 매커니즘을 갖춘다.

### → 비용

데이터를 디지털 큐레이션 하려는 정보기관은 다양한 비용을 부담해야 한다. 그러한 비용에는 인건비, 소프트웨어, 하드웨어, 전력 공급, 네트워크 이용, 마케팅, 저장 매체, 텔레커뮤니케이션 비용 등이 있다. 기관이 디지털 큐레이션 비용을 일찍부터 고려하면 할수록 비용을 효과적으로 지출할 수 있다.

- 기관의 재정적 이슈를 파악한다.
- 안정적 재정보호를 위해 장단기 계획을 수립한다.
- 예산 확보 상황을 주기적으로 모니터링하여 상황 변화에 대처한다.

- 데이터 소유권 구매를 위한 비용은 소유권의 복잡성에 따라 많이 달라질 수 있다.
- 데이터 수집 또는 이전 비용은 단위 당 낮을 수 있으나, 큰 정보기관에서는 전반적으로 이전 비용이 높을 수 있다. 이것은 많은 양의 데이터를 검색하고 다운로드하는 자동 수집에서 일어나는 전송 비용을 포함할지도 모른다. 품질 관리 확인을 사람이 하는 경우에는 많은 비용이 들 것이다.
- 기록되는 정보의 양과 그러한 정보를 찾는 데 어려운 경우들이 있어 데이터를 기술하고 메타데이터를 추가하는 비용은 높을 것이다.
- 데이터를 기관이 사용하기로 정해진 표준 포맷으로 변환하는 일은 변환이 쉬운 경우 비용이 적게 들겠지만 사람이 수작업하거나 수정한다면 비용이 많이 들 것이다. 그러나 기관에서 사용할 포맷의 범위를 제한하는 일은 장기적으로 볼 때 상당한 비용을 절약해 줄 수 있다.
- 기술의 발전과 장기간의 시간프레임 등을 포함한 여러 불확실성 때문에 데이터 보존비용을 예측하기는 어렵다.
- 보존을 위한 데이터를 선정하는 비용은 단위당 낮을 수 있으나, 선정을 위한 인간의 판단은 인건비에서 많은 부분을 차지한다.
- 데이터 리포지토리 운영비용은 제공되는 서비스 수준에 따라 크게 다를 수 있지만, 모든 경우에 기술(하드웨어, 소프트웨어 및 스토리지) 및 인력 비용이 발생하므로, 이에 대한 예산을 계획하고 마련한다.
- 저장 관련 비용은 이론적으로는 낮고 줄어들고 있으나, 그러한 비용은 저장되는 데이터의 양을 반영하므로 결과적으로 높을 수 있다. 저장과 가공 비용을 예측할 때에는 백업과 데이터의 다양한 버전을 고려해야 한다.
- 접근 방식을 제공하는 비용은 접근을 제공하는 프로세스의 난이도에 따라 높을 수도 있고 낮을 수도 있다.

→ 인력

디지털 큐레이션을 위해 필요한 인력의 수는 기관의 규모와 유형에 따라 다르다. 서로 매우 비슷한 데이터가 잘 조직되어 있고, 보존 업무가 간단하다면, 대부분의 프로세스를 자동화하고 필요로 하는 인력을 최소화할 수 있다. 그러나 인간의 개입이 반복적으로 필요하다면, 그 기관은 그에 적당한 인력이 필요하다.

- 기관은 디지털 큐레이션을 수행하는 데 필요한 적절한 수의 직원을 확보하고 있어야 한다.
- 기관은 디지털 큐레이션을 수행하는 데 필요한 능력과 자질을 갖춘 직원을 확보하고 있어야 한다.



## - 디지털 큐레이터에게 필요한 자질

- 문제 해결 능력
- 단기·중기·장기적인 이슈를 고려하여 사전 예방적인 접근
- 관련된 기술, 보존, 기업, 법적 및 정치적 측면에 대한 적합한 인지
- 새로운 아이디어를 수용하고 변화에 적응하는 동시에 비판적으로 생각할 수 있는 능력
- 다른 사람들과 소통하고 협력할 수 있는 능력

## → 정보 기술 인프라

- 기관은 정보시스템을 최적의 상태로 유지하기 위하여 관리지침을 마련하고 문서화한다.
- 기관은 정보시스템의 노후화로 인한 성능 저하 방지를 위해 정보시스템의 성능을 지속적으로 모니터링하고 관리지침에 따라 그 정보시스템을 주기적으로 업데이트한다.

## KOREA 모델의 전주기 활동

전주기 활동은 데이터의 라이프 사이클 전반에 걸쳐 지속적으로 수행되어야 하는 활동이다.

### 큐레이션 계획 및 관리

#### → 개요

‘큐레이션 계획 및 관리’는 체계적이고 효율적인 큐레이션을 수행하기 위해 필요한 기초 활동이다. 전반적인 디지털 큐레이션 활동에 관한 구체적이고 실질적인 장·단기 큐레이션 계획을 수립하고 시행한다. 기관은 자관의 큐레이션 계획을 지속적으로 모니터링하고 검토하여 필요에 따라 계획을 변경한다. 기관은 수립한 계획을 모니터링하고 검토할 책임자와 시기를 결정한다.

디지털 큐레이션을 위한 계획 수립과 관리는 현실에 맞게 수행되어야 한다. 기관의 실제 상황을 고려하여 충분한 수준과 이상적인 수준 사이에서 균형을 유지하면서 계획을 수립하고 관리해야 한다.

#### → 주요 행위

- 큐레이션에 필요한 자원(e.g., 인력, 예산, 시설)이 준비되었는지 확인한다.
- 큐레이션을 위한 장·단기 전략을 수립하고 문서화한다.
- KOREA 모델의 모든 큐레이션 활동 각각에 대한 계획 수립의 필요성을 인식한다.
- KOREA 모델의 모든 큐레이션 활동 각각에 대한 계획을 세우는 데, 행정과 관리 측면도 고려하여 그에 대한 내용도 포함한다.
- 수립한 계획을 누가, 언제, 어떻게, 얼마나 자주 검토하고 업데이트할 것인지 결정하여 이를 기록해둔다. 결정된 대로 수립한 계획을 정기적으로 검토하고 업데이트한다.
- 기관은 정보기술과 표준의 개발 상황, 자관의 자원(e.g., 데이터, 직원, 정보시스템, 서비스)과 이용자 사이의 상호작용을 모니터링한 결과에 따라 큐레이션 계획을 변경할 수 있는 매커니즘을 갖춘다.
- 큐레이션을 위해 필요한 정책을 수립하고 문서화한다.
- 기관이 성장하고 정보기술이 발전함에 따라 기관의 디지털 큐레이션 절차와 정책을 검토, 업데이트, 개발할 수 있는 매커니즘을 갖춘다.

- KOREA 모델의 모든 활동을 대상으로 하여 각 활동에서 어떤 내용을 문서로 작성하여 보관할지 결정한다.
- 기관의 데이터를 사용할만한 이용자/커뮤니티를 정한다.
- 큐레이션 업무를 수행할 수 있는 지식, 경험, 스킬이 있는 직원을 파악한다.
- 큐레이션을 수행할 직원/부서의 역할과 책임을 규정하고 문서화한다.
- 큐레이션의 업무절차와 워크플로우를 결정하고 문서화한다.
- 직원들의 디지털 큐레이션 전문성을 개발할 수 있도록 관련 교육이나 훈련 프로그램을 계획하고 제공한다. 기관 내에서 독자적으로 이러한 프로그램을 제공할 수 없다면, 직원들이 외부기관의 관련 프로그램을 수강할 수 있도록 한다.
- 큐레이션 수행을 위한 예산이 안정적으로 충당되도록 지속적인 노력을 한다.
- 데이터 품질 측정 지표와 기준을 정립한다.
- 표준은 신뢰할 수 있는 디지털 큐레이션을 수행하기 위하여 필수적으로 필요하다. 대부분의 디지털 큐레이션 행위는 표준(e.g., 파일 포맷 표준, 메타데이터 표준, 파일 식별 표준)을 필요로 하는 데, 그러한 표준을 파악한다.
- 저장 매체, 데이터 포맷, 파일 포맷을 결정한다.
- 장기간 사용으로 인한 저장 매체의 성능저하와 노후화에 대비하여 저장 매체의 변경계획(e.g., 마이그레이션, 에뮬레이션)을 세운다.
- 사고나 재해 발생을 대비하여 매뉴얼과 데이터 복구 계획을 세운다.

### → 위기관리(risk management)를 위한 계획

데이터를 장기 보존하여 미래에도 이용할 수 있도록 하기 위해서는 위기를 관리할 필요가 있다. '위기관리'는 위험을 파악, 평가, 대응하는 프로세스를 말한다. 즉, 위험을 최소화하기 위하여 조치를 취하고 복구 프로그램을 실행하여 손실 또는 피해에 대하여 체계적으로 통제하는 것을 말한다. '위기관리'는 위험한 사건이 발생할 가능성을 줄이고, 사건이 일어났을 때에는 그 영향력을 줄이는 것을 목표로 한다. 데이터 관리와 관련하여 미리 대처할 수 있는 손실을 방지하려면 위기관리 계획이 중요하므로, 기관은 위기(재난 포함)에 대처하는 절차, 방법, 대책에 관한 계획을 수립해야 한다. 효과적인 위기관리 계획을 수립하기 위해서는 국제 위기관리 표준 AS/NZS ISO 3100:2009를 준수해야 한다.

### → 큐레이션을 위한 정책

효율적인 디지털 큐레이션을 위해서는 디지털 큐레이션의 모든 측면을 고려한 정책 개발과 시행이 중요하다. 그러한 정책은 명확하고 장기적인 방향을 제시해야 한다. 수립된 정책은 정기적으로 검토되고 업데이트 되어야 한다.

#### - 정책의 유용성

- 정책은 장기적인 방향과 안내를 제공한다.
- 원칙, 가치, 의도를 명확하게 서술하고 있어, 그러한 서술을 바탕으로 하여 일관성 있는 결정을 내릴 수 있다.
- 기관의 디지털 큐레이션 정책 수립은 그 기관이 디지털 큐레이션 전략과 일관된 디지털 큐레이션 프로그램을 개발하는 데 도움을 준다.
- 정책은 디지털 큐레이터들이 불법행위를 하여 고소당하는 것을 방지해 줄 수 있다.
- 기관이 정책을 가지고 있다는 것은 그 기관이 큐레이션 책임을 진지하게 받아들이고 수행하고 있음을 외부에 보여줄 수 있다.

#### - 좋은 정책의 일반적인 구성 내용

- 장기적인 방향
- 무엇이 허락되고, 무엇이 허락되지 않는지
- 해당 정책을 누가 어떻게 모니터링 할 것인지
- 그 정책을 얼마나 자주 검토할 것인지
- 다른 관련 정책으로의 링크

## 기술, 식별, 연결

### → 개요

‘기술, 식별, 연결’은 데이터의 라이프 사이클 내내 진행되는 활동이다. 기술은 데이터를 이해하고, 관리하고, 검색하여 접근할 수 있도록 메타데이터를 작성하는 것이다. 식별은 데이터에 1) 기관의 고유 식별자를 부여하거나; 2) 디지털 객체 식별자(Digital Object Identifier: DOI), 국제표준도서번호(International Standard Book Number: ISBN), 국제표준연속간행물번호(International Standard Serial Number: ISSN)와 같이 많은 기관에서 통용되는 식별자를 부여하는 것을 의미한다. 식별자를 이용하여 관련이 있는 데이터끼리 연결할 수 있으며, 그러한 연결의 결과로 새로운 데이터셋을 생성할 수 있다.

### → 주요 행위

- 직원들이 기술, 식별, 연결에 대한 필요성과 중요성을 인식해야 한다.
- 기술, 식별, 연결을 위한 정책을 수립하여 문서화한 후 시행한다.
- 기술, 식별, 연결과 관련하여 표준을 사용한다.
- 미래의 이용자들이 기관의 소장 데이터를 이해할 수 있도록 하기 위해 그 데이터에 관하여 어떤 정보를 제공해야 하는지 인지한다.
- 메타데이터를 어떻게 캡처 또는 생성할 것인지 결정한다.
- 표준을 사용하여 기술·관리·구조·보존 메타데이터를 작성한다.
- 기관에서 사용할 식별자를 결정하고 사용한다.
- 데이터와 관련된 정보자원들(e.g., 보고서, 학술논문, 데이터셋)을 어떻게 연결할 것인지 결정한다.
- 메타데이터는 해당 데이터와 연결되어 있는 관련 정보자원들(e.g., 보고서, 학술논문, 데이터셋, 소스 데이터)에 대한 정보도 충분히 제공해야 한다.
- 가능한 한 작업 과정을 자동화한다.

→ 데이터 문서화(Data documentation)

데이터를 이용자-친화적으로 장기간 이용될 수 있도록 만들기 위해서는 이용자들이 데이터를 이해하고 해석할 수 있도록 만드는 일이 중요하다. 이를 위해서는 상세하고 명확한 데이터 기술(data description), 주석(annotation), 맥락 정보(contextual information)가 필요하다.

데이터 문서화는 데이터가 어떻게 생성되었는지 또는 디지털화되었는지, 데이터가 무엇을 의미하는지, 데이터 내용과 구조가 어떠한지 등에 대해 설명한다. 좋은 데이터 문서화는 기본적으로 다음과 같은 정보를 포함한다:

- 데이터를 생성하거나 데이터에 기여한 사람
- 데이터의 이름
- 데이터의 생성 날짜
- 데이터 수집 맥락 및 방법
- 데이터 포맷과 파일 유형
- 데이터 파일의 데이터셋 구조, 파일 간의 관계
- 데이터 가공, 클리닝, 유효성 확인, 품질 보장 절차
- 데이터의 생성부터 시간에 따른 변화
- 데이터 파일의 여러 버전에 대한 식별데이터 접근 조건과 이용 조건에 관한 정보
- 데이터의 기밀성

데이터에 행해진 어떠한 활동이라도 문서화한다. 데이터의 문서화는 데이터가 생성될 때 시작하는 것이 가장 좋고, 큐레이션 과정에서 지속적으로 행해져야 한다. 이것은 이차 이용자들(secondary users)이 해당 데이터를 이해하고 재이용할 수 있도록 돕는다.

## → 메타데이터

- 메타데이터의 정의와 기능: 메타데이터는 ‘데이터에 대한 데이터’로서, 표준화되고 구조화된 데이터 문서화의 서브셋이다. 메타데이터는 “형식화된 방법, 수집시기 및 수집한 사람 등을 포함하는, 특정 데이터셋, 객체 또는 자원에 대한 기술 정보(descriptive information)”이다(American Library Association, [n.,d.]). 메타데이터는 데이터 컬렉션의 기원, 목적, 시간 참조, 지리적 위치, 생성한 사람 또는 조직, 접근 조건과 이용 약관 등을 설명한다. 메타데이터는 이용자들이 원하는 데이터를 쉽게 검색하여 발견할 수 있도록 하고, 인용을 위한 서지 레코드로도 이용된다. 디지털 객체와 그 객체의 메타데이터를 서로 연결하고, 그 메타데이터를 보존해야 한다.

- 메타데이터에 포함될 수 있는 정보

- 디지털 객체의 내용을 분류하거나 색인하기 위해 사용하는 통제 어휘에 대한 정보
- 특정 디지털 객체와 관련 있는 다른 디지털 객체에 대한 정보
- 디지털 객체 생성을 위해 사용된 하드웨어, 소프트웨어, 프로세스에 관한 정보
- 시간이 경과하면서 데이터셋에 어떤 일들이 발생했는지에 대한 과거 이력 정보
- 이용을 위해 요구되는 절차에 관한 정보
- 디지털 객체와 관련된 저작권에 관한 정보
- 디지털 객체를 획득한 장소와 방법에 관한 정보
- 이용자에 관한 정보
- 현재 디지털 객체의 버전에 관한 정보
- 디지털 객체의 무결성 확인과 관련된 정보
- 디지털 객체에 행해진 보존 작업에 관한 정보
- 메타데이터 자체에 대한 정보(e.g., 어떻게 작성 되었는지, 언제 변경되었는지)

- 보편적인 메타데이터 표준의 예

- Preservation Metadata: Implementation Strategies(PREMIS)
- Metadata Encoding and Transmission Standard(METS)
- Metadata Object Description Schema(MODS)
- Metadata Authority Description Schema(MADS)

- 메타데이터 관련 권고 사항

- 메타데이터 기술 시 기관이 정해둔 기술 표준이나 원칙 준수
- 표준 스키마(e.g., 더블린 코어)에 따라 메타데이터를 만들어 다른 기관들과의 상호호환성 향상
- 데이터의 발견가능성을 최대한 높이기 위해 풍부하고 정확하며 완전한 메타데이터 작성
- 메타데이터를 기계가독형으로 제공
- '입수'와 '변환'의 각 단계에서 누가 어떤 메타데이터(기술·구조·관리·보존 메타데이터)를 어떻게 만들 것인지 문서화
- 품질이 좋은 메타데이터 생성을 위해 메타데이터 입력이 간단하도록 고안된 도구 사용
- 메타데이터가 해당 데이터의 재이용을 가능하게 하는지 확인
- 기관은 입력된 메타데이터를 검증하고 관리
- 메타데이터 레코드와 그것이 기술하고 있는 디지털 객체의 링크 유지

- '큐레이션 과정 자동화'를 위한 메타데이터의 중요성: 데이터의 양은 많으나, 큐레이션 인력은 제한되어 있기 때문에 큐레이션 과정 중 가능한 한 많은 부분에서 자동화가 필요하다. 메타데이터는 이러한 자동화를 높은 수준으로 만드는 데 매우 중요하다. 데이터가 생성될 때 그것의 메타데이터가 자동으로 만들어지고, 해당 데이터가 정보기관에 입수될 때 그 메타데이터가 사용가능한 것이 가장 이상적인 경우이다.



### → 발견가능성(discoverability)을 위한 데이터 구조화

디지털 큐레이션은 시간이 지나도 데이터가 발견될 수 있도록 해야 한다. 이를 위해 정보자원들을 기술하고 식별하기 위한 표준화된 방법을 적용한다. 정보자원들의 발견가능성은 메타데이터—특히 기술 메타데이터(descriptive metadata)—와 밀접하게 관련이 있다.

신뢰할 수 있는 장기적인 접근은 데이터, 데이터베이스, 디지털 객체의 지속적인 식별에 기초를 둔다. 만약 이들을 안정적으로 식별하지 못 하거나 찾을 수 없다면, 이들을 잃게 될 것이고 이들에 적용한 어떠한 큐레이션 행위도 소용이 없을 것이다.

### → 영구 식별자(persistent identifiers)

영구 식별자는 디지털 객체의 위치에 상관없이 동일하게 유지되는 그 객체의 라벨(label)이다. 디지털 객체의 위치가 변하더라도(e.g., 다른 서버, 리포지토리, 아카이브) 해당 객체의 영구 식별자는 변하지 않으며 해당 객체가 발견될 수 있도록 한다. 디지털 객체의 신뢰할 수 있는 식별은 해당 디지털 객체에 장기적으로 접근할 수 있도록 하고 그 객체의 신뢰성과 진본성을 보장해준다. 영구 식별자는 디지털 객체가 시간이 흘러도 재이용될 수 있도록 한다. 하나의 디지털 객체에 하나의 유일무이한 영구 식별자가 부여되므로, 그 객체의 버전이 여러 개 존재하더라도, 이용자들의 혼돈을 줄일 수 있다. 영구 식별자를 사용하는 것은 웹자료뿐만 아니라 데이터셋을 가지고 있는 1차 연구에 연결하고 인용하는 데에도 중요하다(Oliver & Harvey, 2016, p. 72).

현재 사용되고 있는 주요 영구 식별자 스키마로는 다음과 같은 것들이 있다:

- Digital Object Identifier(DOI)
- Handle
- Archival Resource Key(ARK)
- Persistent Uniform Resource Locator(PURL)
- Universal Resource Name(URN)

(각각의 스키마에 대한 설명은 <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>에서 참고할 수 있음)

→ 연결

- 데이터와 관련하여 어떤 정보(e.g., 참고문헌, 보고서, 연구 논문, 글꼴)를 소장하고, 그러한 정보 중 어떤 정보를 그 데이터에 연결할 것인지 결정한다.
- 데이터가 충분히 연결(e.g., 관련 보고서/논문으로의 연결, 관련 데이터셋으로의 연결, 소스 데이터로의 연결) 될 수 있도록 작업한다.
- 새로운 데이터셋과 소장 데이터 사이에 관계가 있는 지 조사하고, 관계가 있다면 그것들을 연결한다.
- 시간 경과에 따른 데이터의 맥락정보를 충분히 제공하여 원래 지정된 이용자 집단뿐만 아니라 새로운 이용자 집단도 데이터를 찾아가 이용할 수 있도록 해야 한다. 이를 위해 메타데이터가 데이터와 관련된 맥락정보를 충분히 제공해야 한다. 또한 그 데이터와 연결되어 있는 관련 데이터에 대한 정보도 충분히 제공해야 한다.
- 메타데이터를 사용하여 데이터셋을 관련 출판물, 논문, 데이터 생성 방법 및 이유에 대한 추가 맥락정보를 제공하는 프로젝트에 연결한다.

→ 기술, 식별, 연결을 위한 정책이 갖추어야 할 요소

- 누가 어떤 조건에서 기술, 식별, 연결을 할 것인지에 대한 결정
- 기술할 정보
- 기술할 정보를 가져올 소스
- 기술 정보의 재이용: 우선 허락이 필요한가? 상업적인 목적으로 사용될 수 있는가? 다른 기관들에서 사용할 수 있게 할 것인가?
- 적용될 메타데이터 표준과 스키마
- 사용할 식별자: 자관에서 고유 식별자를 부여할 것인가? 보편적으로 쓰이는 영구 식별자를 사용할 것인가?
- 연결 대상

## 이해관계자 관찰 및 협력

### → 개요

**이해관계자 관찰 및 협력**은 라이프 사이클 전반에 걸쳐 지속적으로 수행되어야 하는 활동이다. 정보기관은 이해관계자들의 요구나 기관의 필요에 따라 이해관계자들과 협력해야 한다. 정보기관은 자관의 이해관계자들(e.g., 상위 기관, 출판사, 학회, 후원 기관, 연구자)을 조사하고, 자관의 큐레이션 활동과 이해관계자들이 각각 어떤 관계를 가지고 있는 지 파악하고, 그에 대해 기록해 두어야 한다. 특히 디지털 큐레이션 커뮤니티와 소통하고 협력하여 디지털 큐레이션을 위한 표준, 도구, 소프트웨어 등을 개발해야 한다.

### → 주요 행위

- 기관은 자관의 이해관계자들을 파악하고, 이에 대해 기록해 둔다.
- 큐레이션 라이프 사이클 모델의 모든 활동과 관련하여 기관이 어떤 이해 관계자들과 어떤 관계를 가지고 있는지, 그들과 어떻게 협력할 수 있는지 파악한다.
- ‘위기 평가’는 어떤 이해관계자들이 특히 중요한지와 그들과 어떤 관계를 맺어야 하는지를 보여주기도 한다.
- 기관은 다른 관련 기관들의 디지털 큐레이션 활동에 관한 정보를 수집하고, 디지털 큐레이션에 관한 지식과 경험을 공유한다.
- 디지털 큐레이션과 관련이 있는 커뮤니티를 관찰하고 협력하여 자관에서 디지털 큐레이션 실무를 수행할 때 최신성을 유지한다.
- 이해관계자의 요구나 기관의 필요에 따라 이해관계자와 협력한다.
- 이해관계자들과 협업하는 경우 각 이해관계자의 책임과 역할을 명확하게 정해둔다.
- 디지털 큐레이션 관련 정책을 세우는 데, 이해관계자들(e.g., 데이터 생성자, 데이터 주제 분야의 전문가)과 협력한다.
- 기관은 자관의 직원들과 이해관계자들의 협업을 위하여 적합한 협업 환경 (e.g., 정보시스템, 네트워크)을 제공한다.
- 현실적이고 실행 가능한 워크플로우를 파악하기 위해 이해관계자들과 협력한다.
- 큐레이션을 위한 표준, 도구, 소프트웨어 등을 개발하기 위하여 디지털 큐레이션 커뮤니티와 협력한다.

### → 디지털 큐레이션을 위한 협력의 장점

- 비용을 효율적으로 지출
- 경험과 정보 공유
- 학습 기회 발생
- 상호 보완
- 큐레이션이 되는 데이터의 범위 확장
- 노력의 낭비를 줄일 수 있는 계획 수립
- 필요한 전문성 확보
- 다른 기관의 시설, 도구, 시스템 이용
- 소프트웨어와 시스템 개발 비용 절약
- 영향력 있는 이해관계자들이 디지털 큐레이션을 진지하게 받아들이도록 분위기 조성
- 데이터 생산자들과 시스템 개발자들에게 미치는 큐레이터들의 영향력 향상
- 표준과 실무를 위한 공동 연구 및 개발
- 효율적인 큐레이션 프로그램을 위한 자원과 기타 후원 증진 (National Library of Australia, 2003)

### → 성공적인 협력을 위한 제안

- 적합한 협력 기관을 선택한다.
- 협력하고자 하는 기관이 협력을 통해 무엇을 얻고자 하는지 이해한다.
- 협의를 통해 각각의 협력 기관이 할 일을 명확하게 정하고 이를 기록하여 공유한다.
- 협력을 위해 필요한 자원을 할당한다.
- 협력 기관과 빈번하게 소통한다.
- 현실적인 목표를 세우고 협약한 내용의 수행을 정기적으로 평가한다.

## 이용(자) 조사

### → 개요

**이용(자) 조사는 라이프 사이클 전반에 걸쳐 지속적으로 행해져야 하는 활동**이다. 정보기관의 자원(e.g., 데이터, 디지털 큐레이터, 정보시스템)과 그 기관 이용자와의 상호작용을 조사하고, 기관의 잠재적 이용자를 파악하는 것을 포함한다. 이러한 조사는 단순한 이용 통계 분석에서부터 과학적인 방법을 사용하는 이용자 연구까지 다양하게 수행될 수 있다. 기관은 조사한 결과를 바탕으로 이용자 요구에 부합하고 이용자 만족도를 높일 수 있도록 데이터를 큐레이션 해야 한다. 기관은 문서화된 방법과 절차에 따라 기관 이용자와 이용에 대한 조사를 주기적으로 수행하고, 그 결과에 따라 아래와 같은 필요한 조치를 취해야 한다.

- 기관의 큐레이션 관련 정책에 반영
- 큐레이션을 위해 사용하는 정보시스템에 새로운 기능 추가 및 기존 기능 향상
- 기존 큐레이션 서비스 향상 및 새로운 큐레이션 서비스 제공
- 이용자가 요구하는 데이터 생성 및 수집

### → 주요 행위

- 기관의 이용자와 이용을 조사할 수 있는 방법, 도구, 채널을 마련한다.
- 자관의 현재 이용자와 잠재 이용자를 파악한다.
- 기관은 자관의 데이터 이용통계를 조사하고, 분석하고, 문서화한다.
- 정보기술의 발전에 따라 변화하는 이용자의 요구를 파악한다.
- 데이터, 정보시스템, 서비스 등에 관한 이용자 만족도를 조사한다.
- 기관 이용자의 데이터 추구 행태를 조사한다.
- 기관의 이용(자)에 대한 조사 결과를 기관의 디지털 큐레이션 실무에 활용한다.

### → 이용(자) 조사를 위해 많이 쓰이는 조사 방법

- 설문 조사
- 인터뷰
- 관찰
- 포커스 그룹
- 참고질문 분석
- 인용 분석

**정보기술 모니터링**

→ **개요**

‘정보기술 모니터링’은 라이프 사이클 전반에 걸쳐 지속적으로 행해져야 하는 활동이다. 정보기술은 디지털 큐레이션 실무에 직접적으로 영향을 주기 때문에 발전하는 정보기술을 지속적으로 모니터링 해야 한다. 더 나아가 새로운 정보기술을 적극적으로 수용하여 실무에 적용해야 한다. 이를 위해 기관의 정보기술 담당자들, 정보기술 커뮤니티와 소통하고 협력하는 것이 바람직하다.

→ **주요 행위**

- 정보기술의 발전을 지속적으로 모니터링한다.
- 새로운 정보기술을 도입하여 기관의 디지털 큐레이션 실무를 향상시킬 수 있는 방안을 마련한다.
- 발전하는 정보기술을 적극적으로 수용하여 큐레이션 활동과 서비스를 향상시킨다.
- 기관의 정보시스템 성능을 지속적으로 모니터링하고, 새로운 기술을 도입하여 성능을 업그레이드한다.
- 데이터의 보존과 접근이 지속 가능하도록 최신 정보기술을 적용한다.
- 기관의 디지털 큐레이터들은 자관의 IT 담당자들과 소통하고 협력한다.
- 새로운 정보기술을 신속하게 받아들여 다른 기관과 협업할 수 있는 환경을 조성한다.
- 디지털 큐레이션 커뮤니티와 협력하여 새로운 정보기술을 디지털 큐레이션 발전을 위해 활용한다(e.g., 큐레이션 도구 개발, 큐레이션 프로세스 자동화).

→ **정보기술 모니터링의 필요성**

- 정보기술의 발전에 따라 이용자의 요구도 변화한다.
- 새로운 정보기술은 새로운 큐레이션 도구 개발을 가능하게 한다.
- 발전하는 정보기술은 큐레이션 업무의 자동화에 기여할 것이다.
- 새로운 정보기술은 보다 나은 정보시스템의 개발과 향상으로 이끌 것이고, 이러한 시스템은 디지털 큐레이션 실무 수행에 도움이 될 것이다.
- 정보기술의 발전은 기관에서 ‘재평가’를 해야 할 시점을 알려주기도 한다: 예를 들어, 포맷, 매체 또는 렌더링 응용 프로그램의 노후화; 마이그레이션이나 정규화가 불가능하거나 비용이 너무 많이 들 때.

## KOREA 모델의 순차적인 활동

순차적인 활동들은 일련의 순서대로 수행되는 활동들로, 그 활동들이 일어나는 ‘단계’라고 생각할 수 있다.

### 개념화

#### → 개요

‘개념화’는 KOREA 모델의 순차적 활동들 중 첫 번째 활동이자 단계이다. 개념화’는 데이터의 생성과 수집을 구상하고 계획하는 것을 말한다. 부연하자면, 디지털 큐레이션 과정과 결과를 생각하면서 데이터 생성과 수집 절차를 개발하고 계획한다. 철저한 계획은 KOREA 모델에서 뒤따라오는 순차적인 활동들을 보다 용이하게 수행할 수 있도록 도움을 줄 것이다.

#### → 주요 행위

- 기관에서 어떤 유형, 포맷, 수량의 데이터를 생성할 것인지 정한다.
- 기관에서 어떤 유형, 포맷, 수량의 데이터를 수집할 것인지 정한다.
- 기관에서 사용할 데이터 포맷을 결정한다.
- 새로운 데이터의 생성을 계획할 때 생성 비용을 예측한다.
- 새로운 데이터의 수집을 계획할 때 수집 비용을 예측한다.
- 기관에서 데이터를 생성하거나 수집하기 위한 표준과 방법을 정한다.
- 데이터 수집, 분석, 관리, 저장에 위한 표준을 결정해야 한다. 특히 ‘개념화’와 관련된 표준으로는 파일 포맷 표준, 메타데이터 표준, 파일 식별 표준 등이 있다.
- 효율적인 디지털 큐레이션을 위하여 데이터를 생성, 수집, 관리, 보존하는데 사용할 정보시스템을 디자인한다.
- 생성 또는 수집하는 데이터의 저장 매체를 결정하고, 데이터를 저장할 안전한 장소를 찾아 확인한다.
- 저작권에 관하여 고민하고 어떻게 할 것인지를 결정한 다음 문서화한다.
- 데이터 공개 시 발생할 수 있는 요구사항(e.g., 엠바고, 여러 곳에서의 출판 금지)을 파악하고, 파악한 요구사항에 대하여 적절하고 신속하게 대응한다.

**생성 또는 수집**

→ **개요**

‘생성 또는 수집’은 KOREA 모델의 순차적 활동들 중 두 번째 활동이자 단계이다. ‘생성’은 기관에서 데이터를 만들고 그 데이터에 대한 식별자와 메타데이터를 부여하는 활동이다. 관리·기술·구조 메타데이터를 작성하고, 경우에 따라 보존 메타데이터도 생성 단계에서 작성할 수 있다. ‘수집’은 법적 규정과 기관의 수집 정책에 따라 다른 조직들(e.g., 데이터 센터, 도서관, 아카이브, 학회, 출판사)이나 개인 데이터 생산자들(e.g., 연구자)로부터 데이터를 획득하고, 필요에 따라 식별자를 부여하고 적합한 메타데이터를 작성하는 것이다. 새로운 데이터를 생성 또는 수집할 때 그 데이터가 왜 필요한지 설명할 수 있어야 한다.

→ **주요 행위**

- 데이터 생성과 수집에 관한 정책을 만들고, 문서화하고, 시행한다.
- 데이터 생성자들(e.g., 연구자, 다른 기관의 직원)에게 큐레이션을 할 수 있게 준비된 데이터(curation-ready data)를 생성하도록 그들을 지도한다.
- 데이터와 디지털 객체를 표준화된 데이터 포맷과 파일 타입으로 생성한다.
- 데이터의 생성 단계에서 데이터의 콘텐츠(content), 구문(syntax), 구조(structure)에 대한 표준을 정한다.
- 새로운 데이터를 생성 또는 수집할 때 그 데이터의 저작권을 누가 소유하는지 명확하게 정하고 이를 문서화한다.
- 새로운 데이터를 생성 또는 수집할 때 그 데이터의 이용과 재이용을 위한 라이선스를 명확히 하고 이를 문서화한다.
- 데이터 계약 또는 기탁 협약을 문서화한다. 계약상의 의무 또는 법률 위반에 대비한 법률적 대처방안을 문서로 소장한다.
- 데이터, 포맷, 소프트웨어, 이용 동의사항, 그리고 출처에 대한 문서를 수집하고 보관한다.
- 데이터를 기관으로 받아들이는 절차를 개발하고 실행한다.
- 데이터 생성자들, 데이터 기탁자들과 협력한다.
- 데이터 품질 보장 프로세스를 정한다.
- 데이터 품질 평가 기준을 세우고, 기관에서 생성했거나 수집한 데이터가 최소한의 데이터 품질 기준을 만족하는지 확인한다.



### → 데이터 생성과 수집에 관한 정책

데이터 생성과 수집에 관한 정책은 데이터 생성자들과 큐레이터들의 책임과 의무사항을 분명하게 정의해야 한다. 생성과 수집에 관한 정책이 포함해야 할 내용으로는 다음과 같은 것들이 있다:

- 데이터를 기관에 기탁할 수 있는 자격 요건
- 수집할 데이터의 종류
- 데이터 품질 요구사항
- 데이터 생성자가 제출해야 하는 메타데이터
- 기밀성(confidentiality)과 정보공개
- 데이터의 접근 상태(e.g., 엠바고 유무)
- 권리와 소유권
- 데이터 파일 포맷
- 기관이 수집하는 데이터의 수량과 부피의 제한

효과적인 큐레이션을 위해서는 데이터 생성과 수집에 관한 정책이 만들어지고, 문서화되고, 시행되어야 한다. 수립된 정책은 필요에 따라 수정되어야 한다.

### → 데이터 생성

기관은 자관에서 새로운 데이터를 생성시 데이터를 어떻게 생성할 것인가를 미리 결정해 두어야 한다. 예를 들어, 콘텐츠 선택, 계측, 정보기술과 접근 방식, 명명법, 버전 관리 등에 대해 미리 결정해 두어야 한다.

기관에서 새로운 데이터의 생성 시 기관의 큐레이션 범위에 드는 데이터인지를 확인하고 이를 기록해 두어야 한다. 기관에서 데이터를 생성 시 누구를 위해 그 데이터를 생성하는지, 그 데이터들이 어떻게 쓰일 수 있는지를 알고 있어야 한다. 또한 그러한 정보는 회의나 정기적인 문서 보고 등을 통해 관련 업무 직원들 간에 공유되어야 한다.

데이터 생성 과정에서 데이터 보호를 위해 필요한 사항들을 파악하고, 그것들에 대해 다른 직원들과 이야기한다. 데이터 생성의 이른 단계에서 콘텐츠(content), 구문(syntax), 구조(structure)를 위해 사용할 표준에 대해 다른 직원들과 이야기를 나누고 동의를 받는다.

### → 데이터 수집

기관은 데이터 생성자들로부터 그들의 데이터를 가져올 수도 있고, 데이터를 생성하지는 않았으나 수집하여 소장하고 있던 데이터 기탁자들로부터 그들의 데이터를 가져올 수도 있다. 어떤 소스에서 데이터를 가져오든 기관의 문서화된 수집정책에 따라 어떤 데이터를 어떻게 수집할 것인지를 결정한다. 필요에 따라, 큐레이터는 수집한 데이터를 성공적으로 유지하고 접근할 수 있도록 하기 위해 적합한 메타데이터를 부여한다.

다른 기관으로부터 데이터를 받아들일 때 큐레이터는 기본적으로 다음과 같은 일을 해야 한다:

- 이전할 데이터를 정렬, 식별, 열거한다. 파일 포맷, 매체, 운영 시스템과 버전, 데이터를 생성하는 데 사용된 프로그램과 버전, 데이터 생성자명, 데이터 생성 이유, 데이터 생성 날짜, 데이터의 수량, 현재 메타데이터에 관한 정보가 이용 가능해야 한다. 보충자료의 하드 카피가 포함될 수도 있다.
- 기탁자 동의서가 완전한 지 확인한다.
- 데이터를 받을 방법과 절차를 확실하게 한다(e.g., FTP, 물리적 매체).
- 받은 데이터가 기관에 적합한 지, 파일 포맷이 기관에서 받아들일 수 있는 것인지, 스팸은 제외되었는지를 검토한다.
- 제출 과정에서 데이터의 무결성이 유지되었는가를 확인한다.
- 제출받은 메타데이터와 메타데이터 레코드가 정확한지 검토한다.
- 지속적인 식별자를 부여한다.
- 데이터의 품질을 평가한다.

### → 오픈 포맷과 오픈 소스

디지털 객체를 큐레이션 할 수 있게 준비하기 위해서는 문서화가 잘 된 소프트웨어 프로그램으로 가공될 수 있는 표준 데이터 포맷과 파일 타입으로 생성해야 한다. 가능하면, 파일 포맷과 소프트웨어는 오픈된—즉, 암호화되어 있지 않고 공공 도메인에서 사용가능한—것이 좋다. 오픈 파일 포맷과 오픈 소스 소프트웨어는 큐레이터들이 소프트웨어 소스 코드를 통제할 수 있고 파일 포맷의 구조를 파악할 수 있도록 하여 큐레이션을 돕는다.

### → 데이터 품질

데이터 품질 보장은 중요하며 큐레이션 과정에서 빨리 시작되어야 한다. 데이터 품질은 데이터를 생성하는 행위와 관련하여 중요하다. 디지털 큐레이션 과정에서 데이터 품질과 관련하여 다음과 같은 세 번의 중요한 시점이 있다:

- 데이터가 수집되는 시점: 사용되는 데이터 수집 방법의 품질이 데이터 품질에 중요한 영향을 미친다.
- 데이터가 분석되는 시점: 이 단계에서 데이터는 글로 씌어지고, 코드화가 된다. 표준화된 절차를 철저히 지키는 것은 데이터 품질을 획득하는 데 도움이 된다.
- 데이터가 확인되는 시점: 데이터 클리닝과 데이터 체킹을 한다. 데이터 클리닝—에러를 발견하고 없앴—은 데이터 품질 보장에 있어 필수적이다.

데이터 품질은 데이터 생성자가 제공하는 메타데이터에 따라서도 영향을 받는다. 불충분한 기술·구조·관리 메타데이터는 해당 데이터를 큐레이션 하는 기관의 능률을 떨어뜨린다.

가능한 한 빨리 데이터 품질 측정 지표를 설정하고, 데이터 품질 측정 전반에 걸쳐 모니터링한다. 생성하거나 수집한 데이터가 최소한의 데이터 품질 기준을 만족하는지 확인한다.

**평가 및 선정**

→ **개요**

‘평가 및 선정’은 KOREA 모델의 순차적 활동들 중 세 번째 활동이자 단계이다. 이 단계에서는 기관이 어떤 데이터를 장기간 큐레이션 할 것인지 어떤 데이터를 처분할 것인지를 결정하는 기준을 마련하고, 그러한 기준을 적용한다.

→ **주요 행위**

- 기관은 누구를 위해 어떤 데이터를 기관에서 소장해야 하는지 파악한다.
- 데이터 평가와 선정을 위한 정책을 수립하고, 문서화하고, 시행한다.
- 평가 기준을 만들고, 문서화하고, 적용한다.
- 기관의 정책과 기준에 따라 큐레이션을 할 데이터를 선택한다.
- 데이터 평가와 선정 시 현재의 이용뿐만 아니라 미래의 이용도 고려한다.
- 선정된 데이터의 보존 기간을 결정한다.

→ **평가와 선정이 필요한 이유**

- 생성되는 모든 데이터를 영원히 큐레이션 할 수 없다.
- 감소하는 데이터의 저장 비용보다 데이터의 양이 훨씬 빠르게 증가하기 때문에, 실제 기관의 데이터 저장 비용은 지금과 같거나 더 높을 수 있다.
- 데이터의 백업은 저장 비용을 최소 2배 증가시킨다.
- 데이터가 많으면 필요한 데이터를 발견하기가 더 어려워진다.
- 보존 메타데이터의 작성 및 관리 비용과 보존조치 비용이 많이 소요된다.
- 데이터를 큐레이션 할 수 있는 숙련된 인력이 부족하다.
- 데이터의 장기 보존과 큐레이션을 위한 각 기관의 자원은 한정되어 있다.

→ **평가와 선정 정책 및 기준**

디지털 큐레이터들은 이해관계자들의 의견을 들으면서, 평가와 선정을 위한 정책과 기준을 만들어야 한다. 특히 데이터를 생성했거나 이용할 것으로 추정되는 커뮤니티의 자문을 받을 필요가 있다.

정책은 일관되고 투명하며 책임 있는 의사 결정을 보장해야 한다. 정책은 개인 정보 보호와 저작권 같은 관련된 법적 요건에 부합해야 한다. 정책은 데이터의 가치를 평가하기 위한 기준을 설정하고 그에 따라 수행해야 하는 작업—즉, 보관해야 하는 기간 또는 폐기할 수 있는 시기—을 정한다. 또한 평가 정책은 ‘재평가’ 원칙과 ‘재평가’ 스케줄에 관한 내용도 포함해야 한다.

각 기관은 자관의 상황과 소장하는 데이터의 종류와 특성에 따라 자관의 평가 기준을 세울 수 있다. 평가 기준은 다음과 같은 내용을 포함할 수 있다:

- 기관의 사명: 기관의 사명에 부합하는 데이터인가?
- 고유성: 다른 데이터가 가지고 있지 않은 고유한 정보를 가지고 있는가? 복사본이 존재하는가?
- 대체가능성: 데이터의 재생산과 연결 지어 고려해본다. 예를 들어 다음과 같은 질문들을 생각해 본다: 재실험 시 같은 데이터가 생산될 수 있는가? 데이터 재생산 시 얼마나 그 작업이 어려운가? 합리적인 비용으로 데이터를 재생산할 수 있는가? 데이터의 가치와 보존비용을 비교해본다.
- 신뢰성: 데이터가 정확하거나 권위가 있는가?
- 중요성: 데이터의 내용이 해당 주제분야에서 얼마나 중요한가? 현재와 미래의 연구를 위해 중요한 내용을 가지고 있는가?
- 진본성: 원본인가? 내용은 손실되거나 변경되지 않았는가?
- 출처: 수집 출처의 적합성(e.g., 정부 부처, 기증자)
- 동료 리뷰: 같은 분야 학자들에 의해 데이터가 공식적으로 확인, 증명, 인정되었나? 데이터의 무결성과 완전성을 확인하기 위해 공식적인 동료 리뷰를 거쳤나? 데이터셋이 동료-리뷰를 받는 저널에 실렸다는 근거가 문서화되어 있나?
- 이용: 해당 데이터의 이용자 그룹, 이용자의 정보 요구, 연구 경향, 그리고 이용 빈도 등을 고려한다.
- 수집 정책: 기관의 현 수집 정책에 부합하는 데이터인가?
- 법적 고려사항: 예를 들어, 프라이버시와 보존을 금지하는 데이터 보호 법률
- 권리와 관련된 이슈들: 예를 들어, 저작권 관련 이슈들
- 데이터의 크기: 기술적 (technical) 측면에서 데이터의 크기 고려
- 다른 데이터와의 관계: 해당 데이터가 다른 데이터의 정보를 확인시켜 주는가? 다른 데이터의 의미와 가치를 보완해 줄 수 있나? 기관에서 소장하고 있는 다른 데이터와 관계가 있는가? 소장하고 있는 다른 데이터의 정보를 보충해 주는가? 소장하고 있는 다른 관련 데이터를 찾는 데 도움이 되는가?

각 기관은 누가 평가와 선정 기준을 개발하고, 누가 개발한 기준을 적용할 지에 대한 역할과 책임을 분명하게 정하고, 이를 문서화해야 한다. 이상적으로는 기관의 디지털 큐레이터들이 데이터 평가와 선정 작업을 할 때, 해당 데이터의 생성자들, 해당 데이터 주제 분야의 전문가들과 협력하는 것이 좋다. 특히 평가와 선정을 위한 정책과 기준을 만들기 위해 기관 이해관계자들의 견해를 들을 필요가 있다. 해당 데이터 생성자들의 견해, 해당 데이터 주제 분야의 전문가들 견해, 해당 데이터를 이용할 것으로 추정되는 커뮤니티들의 견해는 매우 중요하다.

→ 평가와 선정 정책 및 기준

평가 및 선정 프로세스에 해당 정책과 기준을 적용한다. 이 프로세스는 현재와 미래의 이용자들에게 데이터의 선택 결정이 정당하다는 것을 보여줄 수 있도록 명확하고 객관적인 기준을 따라 진행되어야 한다.

미래의 이용자들이 특정 데이터셋이 보존 또는 폐기된 이유를 이해할 수 있도록 모든 평가와 선정 결정을 기록으로 남겨야 한다. 의사 결정 기록은 데이터를 관리하고 제어하는 데 사용되는 정보시스템에 보관되는 메타데이터이며, 이러한 메타데이터는 영구적으로 유지되어야 한다.

기관은 평가와 재평가를 통해 자관의 자원을 어떤 데이터의 큐레이션을 위해서는 사용하지 않겠다고 결정할 수 있다. 그러한 결정이 내려진 데이터는 다른 정보기관이나 저장소로 옮겨지거나 폐기된다.

입수

→ 개요

‘입수’는 KOREA 모델의 네 번째 활동이자 단계이다. ‘입수’는 데이터를 기관의 데이터베이스, 디지털 아카이브 또는 정보시스템에 추가하기 위해 준비하고 실제로 추가하는 행위를 포함한다. 다른 말로, ‘입수’는 이관된 데이터의 처리 과정이라고 할 수 있다. Open Archival Information System(OAIS) 참조 모델에서 입수는 Submission Information Package(SIP)에서 Archival Information Package(AIP)로 전환되는 과정이다. 입수는 노동집약적인 활동이므로 가능한 한 입수 프로세스를 자동화해야 한다.

### → 주요 행위

- 입수에 관한 정책을 수립하고, 문서화하고, 시행한다.
- 기관은 입수 프로세스를 가능한 한 간단하게 만들고 기관에서 할 수 있는 지원과 참고자료(e.g., 지침서)를 제공한다.
- 기관의 입수 프로세스를 가능한 한 자동화한다.
- 입수를 위해 사용할 수 있는 도구를 파악하고 사용한다.
- 입수 프로세스 중 미리 정해진 시점에 입수하고 있는 데이터의 생성자 또는 기탁자에게 그 진행 상황을 알린다.
- 입수된 데이터를 식별이 가능하고 관리될 수 있도록 등록한다.
- 등록된 데이터에 고유 식별자를 부여한다.
- 누가 어떤 메타데이터를 어떻게 작성할 것인지 결정하고 그 결정을 문서화한다.
- 데이터 품질을 검증한다.
- 기관에서 데이터를 입수하는 시점에서 데이터 품질 보장을 마지막으로 누가(e.g., 데이터 생성자, 디지털 큐레이터) 책임져야 하는지 결정한다.
- 데이터 품질 보장의 결과가 모든 이해관계자들에게 전달되었는지 확인한다.
- 기관은 입수 시점에서 데이터와 연관되어야 하는 정보를 명확하게 지정해 둔다.
- 기관은 모든 데이터의 출처를 인증하는 매커니즘을 마련한다.
- 데이터의 장기 보존을 위한 준비를 한다. 예를 들어, 바이러스가 있는지 확인하고 제거; 관련 있는 메타데이터를 추출, 작성, 추가; 데이터의 무결성 체크를 돕기 위해 체크섬(checksum), 해시값(hash value)을 생성; 기술적 상세정보(technical details)를 확인; 필요하다면, 데이터의 압축을 풀거나 암호화를 해제; 기관에서 사용하는 파일 포맷인지 확인하고, 필요하다면 포맷 변환; 데이터와 그 데이터의 메타데이터를 아카이벌 정보 패키지(Archival Information Package)와 연결한다.

### → 입수를 위한 정책

좋은 정책과 안내서가 가장 효율적으로 사용될 수 있는 활동이 바로 '입수'다. 입수를 위한 정책은 법적 규정의 변화와 실제 입수 활동의 현실성을 반영하면서 최신 상태로 유지되고 문서화되어야 한다. 문서화된 정책은 다른 활동에서와 마찬가지로 입수 절차에 있어서 유용한 데, 그 이유는 다음과 같다.

- 책임과 소통 라인을 명확히 한다.
- 표준화를 촉진한다.
- 위험 요소를 관리할 수 있도록 한다.
- 규정 준수 이슈를 해결한다.

기관은 선호하는 포맷이나 정규화 프로세스 등 장기 보존에 영향을 미칠 수 있는 기관의 정책에 대해서도 알고 있어야 한다.

### → 입수를 위한 전제 조건

입수되는 데이터는 '평가 및 선정' 단계에서 큐레이션 대상으로 적합하다고 판단되어 선택된 것이다. 입수 전에 보안, 기밀유지, 저작권 등에 관하여 미리 합의를 해두어야 한다.

### → 법률적 대비

외부에서 생산된 데이터를 입수하고자 할 때 해당 데이터에 대한 계약 또는 기타 협약을 체결하고 문서화하여 보관해야 한다. 기관의 데이터에 대한 계약서 또는 기타 협약문은 계약자/기탁자, 이전된 권리 등을 포함하여 데이터를 큐레이션 할 때 발생할 수 있는 법적 권리 및 책임 등에 관한 내용을 구체적으로 명시해야 한다. 기관은 계약 또는 기타 협약과 관련하여 법률적 분쟁에 대비한 법률적 대처 방안을 문서화하여 가지고 있어야 한다.

### → 데이터 품질 보장

데이터 품질과 클리닝의 수준은 기관의 목적에 맞추어 결정한다. 데이터 품질 보장/관리를 위한 기준을 가지고 사용한다. 기관은 데이터를 기탁 받는 시점에서 데이터 품질 보장을 누가(e.g., 연구자, 아카이브, 큐레이터) 최종적으로 책임져야 하는 지 결정한다. 모든 이해관계자들에게 데이터 품질 보장의 결과를 전달한다.



## 보존

## → 개요

‘보존’은 KOREA 모델의 순차적 활동들 중 다섯 번째 활동이자 단계이다. ‘보존’은 데이터의 무결성, 진본성, 신뢰성, 이용가능성을 유지하면서 데이터의 장기 보존을 위해 실행하는 행위를 포함한다. 시간이 지나도 데이터의 접근과 재이용이 가능하도록 하려면 사전 예방적이고 확장 가능하며 지속 가능한 디지털 보존 전략의 실행이 필요하다. 성공적인 디지털 보존 전략은 시간이 지남에 따라 데이터의 접근성과 유용성을 위협하는 다양한 요인들을 예상하고 그러한 요인들의 영향을 완화시킨다. 보존을 위해 수행한 활동을 기록해 두는 것은 데이터를 장기간 이용할 수 있게 만드는 데 매우 중요하다.

## → 주요 행위

- 보존 정책을 수립하고, 문서화하고, 시행한다.
- 기관의 보존 정책을 기반으로 하여, 사전 예방적이고 확장 가능하며 지속가능한 디지털 장·단기 보존 전략을 수립하고, 문서화하고, 실행한다.
- 보존 절차를 수립한다.
- 큐레이션 할 수 있는 데이터의 양을 늘리기 위해서 보존 작업의 자동화를 위해 노력한다.
- 보존 메타데이터를 작성, 변경, 또는 추가한다.
- 데이터 보존을 담당하는 직원들은 메타데이터나 다른 방법을 통하여 보존할 데이터의 속성을 파악한다.
- 기관은 현재의 정보기술에 적합한 특정 보존 포맷을 채택한다.
- 본래의 비트열과 미래에 사용될 것 같은 ‘보존 버전’을 유지한다.
- 데이터 클리닝을 한다.
- 유효성 검사를 한다.
- 데이터에 행한 모든 보존 조치를 문서화한다.
- 디지털 보존은 위기 평가와 관리까지 포함한다.
- 기관이 책임지고 있는 데이터를 보존하기 위하여 기관의 법적 권리를 명확하게 해둔다.
- 디지털 보존을 위해 디지털 큐레이션 커뮤니티, 디지털 보존 커뮤니티, 정보기술 커뮤니티 등과 협력한다.

→ 디지털 보존 정책과 기준

- 보존 정책을 수립하여 문서화하고 이를 준수한다.
- 수립한 보존 정책을 계속 모니터링하고 업데이트한다.
- 보존 정책은 보존 전략을 뒷받침해야 한다.
- 데이터 종류에 따라 의무적인 보존 수준을 결정한다.
- 보존 방법(e.g., 마이그레이션, 에뮬레이션)을 결정한다.
- 데이터의 보존 기간을 결정하는 데 도움을 줄 지침이나 기준을 마련한다.

→ 디지털 보존 방법의 주요 선택 기준

- 보존하는 데이터의 성격
- 필요한 기술적 인프라(technical infrastructure)
- 비용
- 해당 기관의 요구사항

기관은 하나의 디지털 보존 방법을 사용할 수도 있고, 여러 디지털 보존 방법을 혼용할 수도 있다.

→ 주요 디지털 보존 방법

- 마이그레이션과 에뮬레이션이 가장 많이 사용되는 보존 방법이다.

- 마이그레이션(migration): 디지털 객체의 지적 콘텐츠를 보존하고 이용자들이 그것을 발견하고 이용할 수 있도록, 하나의 기술(하드웨어나 소프트웨어 생성)에서 다른 기술로 디지털 객체를 이전하는 과정이다.
- 에뮬레이션(emulation): 응용 프로그램, 운영 체제 및/또는 하드웨어 아키텍처를 포함하여 현재와 미래 세대의 컴퓨터에서 시대에 뒤떨어진 시스템을 모방하는 소프트웨어를 개발하는 과정이다.
- 디지털 포렌식(digital forensics) 또는 디지털 고고학(digital archaeology): 손상되었거나 접근할 수 없는 저장 매체로부터 데이터를 복원하기 위해 데이터 복원 테크닉을 사용하는 것이다.
- 기술 보존(technology preservation): 시대에 뒤떨어진 하드웨어와 소프트웨어를 작동 가능한 상태로 유지하여 그러한 하드웨어와 소프트웨어를 필요로 하는 디지털 자료에 접근할 수 있도록 한다.

- 데이터 포맷의 표준화(standardizing data formats) 또는 정규화(normalization): (1) 오랫동안 사용될 것 같은 파일 포맷으로 변환 (2) 시간이 지남에 따라 이러한 포맷을 지원하기 위한 노력을 말한다.
- 캡슐화(Encapsulation): 디지털 객체와 그 객체로의 접근을 제공하는 수단을 함께 패키징하는 것이다.

### → 보존 활동의 문서화

- 보존할 데이터를 어떻게 선택하는 지 문서화 한다.
- 데이터의 보존 주기를 문서화 한다.
- 데이터의 보존 조치를 문서화하여 시간이 지나면서 데이터에 어떤 작업이 수행되었는지 알 수 있도록 한다.

### → 보존 활동의 자동화

현재 대부분의 디지털 보존 방법이 노동집약적이기 때문에 큐레이션을 할 수 있는 데이터의 양은 제한적이다. 큐레이션 할 수 있는 데이터의 양을 늘리기 위해서는 큐레이션 워크플로우와 보존 활동의 자동화가 필수적이다. 큐레이션 절차를 자동화하기 위해서는 소프트웨어 도구들이 필요하다. 이러한 소프트웨어 도구들은 영국과 미국의 디지털 보존 프로젝트(e.g., CASPAR, PLANETS)에서 개발되었고, 대부분 오픈 소스로 제공되고 있다. 여러 큐레이션 도구들—특히 보존 활동을 위한 도구들—이 이미 사용가능하거나 현재 개발 중에 있다. 디지털 큐레이터는 목적과 필요에 따라 큐레이션 도구를 선택하여 사용할 수 있다.

### → 보존 책임에 대한 실용적인 원칙

- 누군가는 책임을 져야 한다.
- 모든 큐레이터가 모든 것에 대해 책임질 필요는 없다. 책임을 나누어질 수 있다.
- 책임을 나누어지더라도 누군가는 지도자 역할을 해야 한다.
- 책임을 영원히 질 필요는 없으며, 정해진 시간 동안만 책임을 진다.
- 포괄적인 대규모 보존 프로그램을 개발하는 데에는 시간이 걸린다.

(National Library of Australia, 2003, p. 48)

→ 디지털 보존의 위협 요인

- 소프트웨어, 하드웨어 및 미디어 포맷 노후화
- 하드웨어 오류
- 자연 재해(e.g., 화재, 홍수, 지진)

→ 위기관리 단계의 예

위기관리 단계는 각 기관에 따라 약간씩 다를 수 있는 데, 오스트레일리아 국립 도서관에서는 다음과 같은 위기관리 단계를 제시했다:

- 1 단계: 보호해야 할 기관의 자산 파악
- 2 단계: 기관의 목표에 위협을 가져올 위협 요소 파악
- 3 단계: 각각의 위협 요소가 발생할 가능성 예측
- 4 단계: 그러한 위협 요소가 발생했을 때 그 결과의 영향력 예측
- 5 단계: 위협 요소의 발생 가능성과 실제로 발생했을 때의 영향력을 같이 려하여 위기 단계를 계산
- 6 단계: 위협 요소의 발생 가능성과 영향력을 줄일 수 있는 대책 제안
- 7 단계: 위협을 줄일 수 있는 대책이 필요한 위기 수준 결정
- 8 단계: 위기 대처 책임자 결정
- 9 단계: 파악한 위협 요소의 위험 정도를 비교, 어떠한 위협 요소에 우선순위를 둘 것인지 결정
- 10 단계: 위기 평가와 우선순위 평가가 합리적인지 확인
- 11 단계: 당장 필요한 조치가 있는지 확인

(National Library of Australia, 2003, p. 53)

## 저장

### → 개요

‘저장’은 KOREA 모델의 순차적 활동들 중 여섯 번째 활동이자 단계이다. ‘저장’은 관련 표준에 따라 데이터와 관련 기술(description)을 안전하게 보관하는 활동이다. 이를 위해 시간이 지나도 데이터의 무결성을 지킬 수 있는 저장 시설이 필요하다. 저장 시설은 관련 표준을 따르고, 신뢰할 수 있고, 데이터의 이용과 재이용을 위해 데이터의 접근이 가능해야 한다.

### → 주요 행위

- 안전한 데이터 저장을 위한 정책을 개발하여 문서화하고 적용한다.
- 데이터가 저장되는 매체와 장소를 문서화한다.
- 기관은 입수한 데이터를 서로 다른 지역의 안전한 장소에 보관한다.
- 동일한 데이터를 여러 곳에 분산하여 보관할 때에는 그 데이터의 버전을 관리한다.
- 지리적으로 떨어져 있는 백업 시스템과 같은 신뢰할 수 있는 저장 매체를 이용한다.
- 백업에 대한 기관의 지침과 절차를 마련하고 정해진 스케줄에 따라 정기적으로 백업한다.
- 기관은 충분한 저장소(storage)를 제공한다.
- 데이터를 어떻게 저장소로 전송할 것인지(필요하다면 암호화)를 문서화한다.
- 데이터와 함께 충분한 디스크립션(description)이 저장되도록 한다.
- 저장되어 있는 데이터와 관련 디스크립션(description)의 무결성을 정기적으로 체크한다.
- 저장에 대한 책임자를 지정하고, 이를 문서화한다.
- 저장에 대한 책임이 바뀌는 경우 (가능하면) 공식 수령증을 받거나 폐쇄 및 이전에 대한 비공식적 승인을 받는다.
- 데이터 보안 정책과 규정을 수립하고 그것을 문서화하고 시행한다.
- 보존 활동을 촉발하는 요인(e.g., 파일 포맷 마이그레이션, 파일 손상)이 있는지 모니터링한다.
- 시간의 경과에 따른 저장 매체의 성능저하와 노후화에 대비한 저장매체의 변경 계획을 수립한다.
- 기술 인프라(technical infrastructure)를 유지관리하고 필요에 따라 교체 한다.
- 화재, 홍수, 지진 등의 재난상황에 대비하여 데이터 복구 절차와 매뉴얼을 개발하고, 필요에 따라 그 절차를 수행한다.

→ 디지털 보존의 위협 요인

- 데이터 저장을 위해 보존 친화적인 포맷(보통 비독점적이거나 오픈 표준 포맷)을 사용한다.
- 가능한 한 빨리 모든 데이터의 백업 복사본을 두 개 이상 만들어 원본과 떨어진 두 곳 이상의 장소에 저장한다.
- 정기적(2년이나 5년 주기 추천)으로 데이터를 새로운 매체에 복사하거나 마이그레이션 한다.
- 체크섬이나 다른 방법을 사용하여 저장되어 있는 모든 데이터의 무결성을 정기적으로 체크한다.
- 저장 매체를 모니터링한다.
- 파일 포맷을 모니터링하다가 필요하면 새로운 포맷으로 마이그레이션한다.
- 저장된 데이터를 즉각적으로 찾을 수 있도록 조직한다.
- 데이터 저장 지역이 불이나 홍수의 피해를 최소화할 수 있도록 구조적으로 좋은 지역인지 확인한다.

→ 데이터 보안

- 데이터 보안은 승인되지 않은 접근, 이용, 변경, 공개, 폐기 등으로부터 데이터를 보호하는 것이다. 또한 데이터에 원하지 않는 변화가 일어나지 않도록 보호하는 것이다.
- 개인 정보를 가지고 있는 데이터는 그렇지 않은 데이터보다 높은 수준의 보안 조치를 취해야 한다.
- 데이터를 효과적으로 보호하기 위하여 보안 정책과 규정을 수립하여 시행한다. 보안 정책은 보안 평가 기준의 국제 표준을 준수한다.
- 기관의 보안 정책과 규정에 따라 데이터로의 접근을 통제하는 보안시스템을 구축하여 운영한다.
- 보안시스템은 보안등급을 분류하여 적용한다.
- 보안업무와 관련하여 직원들의 임무, 책임, 역할에 따른 권한 등을 분명하게 기록해 둔다.
- 모든 보안시스템의 보안 침해 사례를 기록해 두고 분석하여 대처한다.
- 보안과 관련된 요인들(e.g., 시스템, 데이터, 인력, 시설설비, 보안 요구)을 주기적으로 점검한다.
- 데이터 보안에서 유의해야 할 세 가지 측면

- 물리적 보안: 데이터가 저장되고 컴퓨터가 있는 물리적 장소가 안전하고 승인 받은 사람들만 접근할 수 있도록 한다.
- 네트워크 보안: 인터넷과 같은 외부 네트워크와 연결된 컴퓨터나 서버에 소장된 데이터를 안전하게 보관한다.
- 파일 보안: 파일을 암호화하여 승인을 못 받은 이용자들이 접근할 수 없도록 만든다.

## 접근, 이용, 재이용

### → 개요

‘접근, 이용, 재이용’은 KOREA 모델의 순차적 활동들 중 일곱 번째 활동이자 단계이다. ‘접근, 이용, 재이용’은 데이터가 일상적으로 이용, 재이용 될 수 있도록 데이터를 접근가능하게 유지하는 것을 말한다. 각 기관은 필요에 따라 소장 데이터로의 접근을 강력하게 통제할 수 있으며, 이를 위해 인증 절차를 갖출 수 있다.

### → 주요 행위

- 기관의 데이터를 이용 또는 재이용할 수 있도록 하기 위해 그 기관에게 요구되는 법적 권한이 있는지 확인한다.
- 계약 및 기탁 협약을 준수하는 접근 정책을 수립하고 문서화하여 시행한다.
- 데이터가 웹에서 검색 될 수 있도록 적합한 메타데이터를 보장하는 표준을 사용한다.
- 데이터를 웹 검색 엔진에서 검색 될 수 있도록 만든다.
- 기관의 정보기술 인프라가 데이터의 접근 제한을 기록하고 지원할 수 있도록 만든다(e.g., 암호, 윤리, 접근 동의 및 계약).
- 이용자들이 데이터로의 접근 권한을 얻을 수 있는 프로세스를 마련한다.
- 데이터에 대한 접근 권한 및 접근 제한을 명확하게 서술하여 공지한다.
- 오직 승인받은 이용자들만 데이터에 접근할 수 있도록 접근 통제와 인증절차를 사용한다.
- 데이터는 명확한 데이터 이용 조건(e.g., Creative Commons License)과 함께 공개한다.
- 시간 경과에 따른 데이터의 맥락정보를 충분히 제공하여 원래 지정된 이용자 집단이든 새로운 이용자 집단이든 데이터를 찾아 이용할 수 있도록 한다. 이를 위해 메타데이터가 해당 데이터와 관련된 맥락 정보를 충분히 제공해야 한다. 또한 해당 데이터와 연결되어 있는 관련 데이터에 대한 정보도 충분히 제공한다.
- 가치가 높은 데이터의 경우 데이터가 누구에 의해 어떻게 재이용될 수 있는 지 생각해 보고, 해당 데이터에 새로운 가치를 부가한다.
- 데이터의 재이용을 위해 데이터 생성자, 데이터 소유자, 데이터 보관자를 명확하게 표시한다.
- 데이터의 이용과 재이용을 위한 협업 도구를 제공한다.
- 기관 이용자들의 데이터 접근과 이용을 신속하게 도울 수 있도록 기관의 연락 정보(e.g., 전화번호, 팩스 번호, 이메일 주소)를 공지한다.

→ 법적 이슈

- 데이터 접근, 이용, 재이용은 법적 요구사항을 충실히 따라야 한다.
- 법적 요구사항은 다음과 같은 세 개의 카테고리로 나누어진다.

- 연구 자금 제공자의 요구사항
  - 기밀과 프라이버시에 관하여 법령으로 정해진 요구사항
  - 저작권과 디지털 권한 관리
- (Oliver & Harvey, 2016, p. 204)

→ 접근 통제와 인증 절차

특정 환경에서 기관의 민감한 데이터나 기밀의 데이터를 보호하기 위해 그러한 데이터로의 접근과 이용을 제한할 수 있다. 일반적으로 정보기관이 보유하고 있는 데이터는 누구나 이용할 수 있는 공적 도메인(public domain)에 속하지 않는다. 그러한 데이터의 경우 이용자 등록 후에 특정 목적을 가지고 사용하도록 접근을 제한할 수 있다.

- 접근 정책을 실행할 수 있는 정보시스템(인증 기능 포함)을 구축하여 운영 한다.
- 이용자들이 데이터로의 접근 권한을 얻을 수 있는 프로세스를 마련하고 데이터에 대한 접근 권한 및 접근 제한을 명확하게 문서화하고 이를 공지 한다.
- 데이터에 접근할 수 있는 접근 권한이 변경된 경우 변경 내용을 공지해야 한다.
- 접근 권한이 있는 이용자가 이용 대상인 데이터에 접근하지 못한 경우를 모니터링하고, 이를 기록하여 관리하고 해결방법을 찾는다.

→ 데이터의 공유와 재이용 촉진을 위해 필요한 기본 요건

- 오픈 데이터: 저작권이나 다른 제한 없이 데이터를 무료로 이용할 수 있어야 한다.
- 상호운용성: 데이터의 재이용을 위해서는 널리 쓰이는 표준을 사용하여 해당 데이터가 공유될 수 있어야 한다.
- 이용자의 신뢰: 이용자가 데이터를 신뢰할 수 있도록 데이터가 진본성을 가지고 있고, 데이터의 출처가 분명하고, 데이터 검색·접근·이용이 가능해야 한다.



## 변환

### → 개요

‘변환’은 KOREA 모델의 순차적 활동들 중 마지막 활동이자 단계이다. ‘변환’은 원본 데이터(original data)를 가지고 새로운 데이터를 만드는 것이다. 데이터를 마이그레이션 하여 새로운 포맷으로 만들거나, 특정 데이터셋을 가지고 새로운 서브셋을 만드는 등의 행위를 포함한다. ‘변환’을 통해 새로 생성된 데이터도 큐레이션 해야 한다.

### → 주요 행위

- 파일 포맷 변환에 필요한 소프트웨어를 마련한다.
- 데이터 포맷을 변환 할 때는 데이터, 주석 또는 내부 메타데이터가 손실되거나 변경되지 않았는지 확인한다.
- 최신 정보기술을 반영하는 데이터 포맷에 관심을 가지고 그에 대한 지식을 지속적으로 업데이트 한다.
- 새로 파생된 결과를 만들기 위해, 선택이나 질의(query)를 하여 데이터셋의 서브셋을 생성한다.

### → 포맷 변환

- 시간이 지남에 따라 기술이 발전하기 때문에 포맷 변환은 언젠가는 일어나는 일이다.
- 사용되지 않는 포맷으로 보관되고 있는 데이터를 제거하기 위해 데이터 포맷을 반복적으로 변환할 수도 있다.
- 기관은 최신의 정보기술을 반영하는 데이터 포맷에 대한 지식을 지속적으로 업데이트해야 하는데, 누가 이러한 책임을 질 것인지를 결정한다.

### → 새로운 데이터셋이나 서브셋의 생성

기존에 소장하고 있는 데이터셋의 서브셋을 만들거나 이러한 서브셋을 변환 시켜 새로운 데이터를 생성할 수 있다. 이러한 서브셋은 데이터를 선택하거나 질의(query)를 하여 생성할 수 있다. 예를 들어, '한국 도서관 현황'이라는 데이터셋에서 '국립도서관'이라는 질의어를 통해 '한국 국립도서관 현황'이라는 새로운 서브셋을 만들 수 있다.

이미 생성되어 있는 데이터셋을 재처리하거나 수정하거나 부수적인 데이터를 추가하여 새로운 버전의 데이터셋이 만들어지기도 한다. 이러한 경우 데이터셋의 버전을 관리하고 기록해 두어야 한다. 이러한 버전 관리는 연구를 위해 사용한 데이터셋을 정확하게 인용해야 하는 연구자들이 그들 연구의 재생산성과 신뢰성을 보여주는 데 매우 중요하다. 기관은 데이터 버전에 번호를 일관성 있게 매겨서 데이터 이용자들이 해당 데이터의 컬렉션이 어떻게 변화했는지, 사용할 수 있는 최신 버전은 어떤 것인지를 알 수 있도록 해야 한다.

## KOREA 모델의 비정기적 활동

비정기적 활동들은 이따금씩 일어나는 활동들이다.

### 마이그레이션

#### → 개요

‘마이그레이션’은 비정기적 활동들 중 하나로 주로 순차적 활동인 ‘보존’이나 ‘접근, 이용, 재이용’ 단계에서 어떤 계기에 의해 촉발된다. **마이그레이션은 디지털 객체의 지적 콘텐츠를 보존하고, 이용자들이 그것을 발견하고 이용할 수 있도록 어떤 기술(하드웨어나 소프트웨어에서의 생성)에서 다른 기술로, 어떤 포맷에서 다른 포맷으로, 디지털 객체를 이전하는 과정이다.** 또한 마이그레이션은 파일 포맷의 변화 또는 저장 환경이나 기술 플랫폼의 변화에 반응하는 프로세스이다. 마이그레이션은 가장 널리 사용되는 보존 방법이다..

#### → 주요 행위

- 얼마나 자주 마이그레이션을 해야 하는지, 어떤 포맷으로 마이그레이션을 해야 하는지, 어떤 데이터를 마이그레이션 해야 하는지, 누가 마이그레이션에 대한 책임을 질 것인지 등에 대한 정책을 세운다. 그 정책을 문서화하고 시행한다.
- 마이그레이션의 절차와 방법을 문서화한다.
- 파일 포맷의 노후화를 모니터링하고, 필요에 따라 노후화된 파일 포맷을 표준에 따라 새로운 포맷으로 변환한다.
- 마이그레이션하여 생성할 파일 포맷과 데이터 구조를 선정한다. 미래에 회복될 확률을 증가시키기 위하여 비독점적이고 문서화가 잘 된 데이터 포맷 표준을 사용한다.
- 필요에 따라 마이그레이션 한다.
- 메타데이터에 마이그레이션 절차를 충분히 기록한다.
- 원본의 정확성과 진본성을 유지하기 위해 엄격한 품질 확인 절차를 만들고 철저히 시행한다.

### → 마이그레이션을 위한 파일 포맷의 선택 기준

- 공개 표준(open standards)
- 보편성(Ubiquity)
- 안정성(Stability)
- 메타데이터 지원(Metadata Support)
- 기능 셋(Feature Set)
- 상호운용성(Interoperability)
- 생존 능력(Viability)
- 진본성(authenticity)
- 가공 능력(processability)
- 표현력(presentation)

(각 기준에 대한 설명은 [https://webarchive.nationalarchives.gov.uk/+https://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](https://webarchive.nationalarchives.gov.uk/+https://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)에서 참고할 수 있음)

### → 마이그레이션의 단점

마이그레이션이 많이 사용되는 보존 방법이지만 완벽하지는 않다.

- 단순한 디지털 객체나 단일한 포맷의 자료 컬렉션에는 마이그레이션이 적합하지만, 복잡한 자료의 경우에는 마이그레이션이 적합하지 않다.
- 자료의 일생동안 일정한 주기로 마이그레이션을 계속 해주어야 하기 때문에 지속적인 비용이 든다.
- 마이그레이션은 그 절차를 충분히 문서화해야 하기 때문에 노동집약적인 방법이다.
- 마이그레이션이 데이터를 조금씩 변화시킬 수 있기 때문에, 이러한 변화들이 시간에 따라 계속되면서 본래의 데이터를 크게 변화시킬 수 있다. 이것은 데이터의 무결성과 진본성 문제를 일으킬 수 있다.

## 재평가

### → 개요

‘재평가’는 KOREA 모델의 비정기적 활동 중에 하나로, 어떤 계기에 의해 데이터를 다시 평가하는 활동이다. ‘재평가’는 주로 ‘보존’이나 ‘접근, 이용, 재이용’ 단계에서 일어나는 데, 이러한 단계에서 기존의 데이터가 평가 기준에 부합하는지 다시 평가한다. 만약 그 데이터가 더 이상 큐레이션 될 필요가 없다고 결정되면 그 데이터를 ‘처분’한다—즉 그 데이터를 폐기하거나 다른 정보기관이나 저장소로 보낸다.

### → 주요 행위

- 재평가에 대한 문서화된 정책을 마련하고 준수한다.
- 재평가를 할 조건을 상세히 규정한다.
- 재평가 절차와 방법을 문서화 한다.
- 재평가해야 할 데이터셋을 평가 기준에 맞추어 다시 평가한다.
- 재평가를 책임질 직원을 결정하고 문서화 한다.

### → 재평가를 해야 하는 상황

재평가를 수행하는 데에는 자원이 많이 들어가는데, 전통적으로 데이터가 생성된 지 몇 년이나 몇 십 년 뒤에 행해진다. 데이터의 재평가를 촉발하는 상황으로는 다음과 같은 것들이 있다.

- 기관의 운영 상황에 큰 변화
- 기관 지시에 큰 변화
- 예상하지 못 했던 심각한 상황 발생
- 정보기술의 발전: 예를 들어, 마이그레이션이나 정규화가 불가능하거나 비용이 너무 많이 들 때; 포맷, 매체 또는 렌더링 응용 프로그램의 노후화
- 지정된 커뮤니티(designated community)에 변화: 예를 들어, 대학이 특정 학과를 없앴을 때
- 법적 변화: 예를 들어, 특정 데이터에 대한 법적 기탁 규정이 개정되었을 때

### → 평가와 재평가

평가 정책은 재평가 원칙과 재평가 스케줄에 대한 내용을 포함해야 한다. 평가와 재평가 결과는 특정 데이터의 큐레이션에 기관의 자원을 사용하지 않겠다는 결정을 가져올 수 있다. 그러한 결정이 내려진 데이터는 다른 정보기관이나 저장소로 옮겨지거나 폐기된다.

**처분**

→ **개요**

‘처분’은 비정기적으로 일어나는 활동으로, ‘평가 및 선정’ 또는 ‘재평가’의 결과로 수행된다. 기관에서 장기간 보존과 큐레이션을 할 필요가 없다고 결정한 데이터를 다른 정보기관이나 저장소로 이관하거나 안전하게 폐기한다. 저작권 등의 법률 위반, 국가 안보에의 위험, 허위 연구 결과, 데이터의 기밀성과 관련된 사항 때문에 처분 결정을 내릴 수 있다.

→ **주요 행위**

- 데이터 처분(폐기 포함)을 위한 기관의 정책과 절차를 수립하고 문서화하여 실행한다.
- 데이터 처분 기준이나 지침을 마련하고 그것을 준수한다.
- 처분 절차와 프로세스를 정한다.
- 처분을 위하여 정기적인 검토를 한다.
- 처분을 수행할 책임자를 정한다.
- 데이터 처분을 촉발하는 요인을 기관 차원에서 결정해 둔다.
- 기관의 정보시스템이 처분작업의 자동화를 얼마나 보조할 수 있는지 확인한다.
- 데이터를 영구 폐기하는 경우 법적으로 요구되는 사항이 있는지 확인하고, 만약 요구되는 사항이 있다면 그에 따라 영구 폐기한다.

→ **평가, 재평가, 처분**

평가와 재평가 결과는 특정 데이터의 큐레이션에 기관의 자원을 사용하지 않겠다는 결정을 가져올 수 있다. 그러한 결정이 내려진 데이터는 다른 정보기관이나 저장소로 옮겨지거나 완전히 폐기된다.

### → 처분 정책

다음과 같은 세 가지 면을 고려하여 처분 정책을 수립한다.

- 데이터 처분 조건
- 처분된 데이터의 처리
- 처분된 데이터의 메타데이터 처리

### → 처분 과정

데이터를 처분하기 전에 자관이 어떤 데이터를 가지고 있으며 어떤 데이터를 가지고 있을 필요가 있는지 검토한다. 기관은 처분 절차와 결과를 문서화한다.

### → 데이터 이전

기관과 관련이 없는 데이터는 다른 정보기관이나 저장소로 보내는 것을 고려한다. 보내는 경우 해당 데이터에 대한 메타데이터와 관련 문서도 함께 보낸다.

### → 데이터 폐기

“폐기”는 완전히 데이터를 없애는 행위이다. 어떤 데이터는 법적 규정에 의해 안전하게 폐기되어야 한다. 어떤 법령은 데이터 폐기 결정에 관한 문서를 남기도록 규정하기도 한다. 기관은 특정 데이터를 영구 폐기 시 법적으로 요구되는 사항이 있는지 확인하고 있을 경우 그에 따라 조치해야 한다.

폐기되는 모든 데이터는 다시 이용되거나 구축될 수 없도록 안전하게 폐기되어야 한다. 가장 믿을 만한 폐기 방법은 데이터를 가지고 있는 저장 매체(e.g., 컴퓨터, 외부 하드 드라이브)를 물리적으로 파괴하는 방법이다. 단순히 파일을 삭제하거나 하드드라이브를 리포맷팅하는 것은 복원이 가능하기 때문에 안전한 방법이 아니다. 폐기를 위해서는 기관에서 허락하는 안전한 파괴 시설을 이용하여 저장 매체를 물리적으로 파괴해야 한다.

03



디지털 큐레이션  
가이드라인과  
체크리스트



## 제3장

# 디지털 큐레이션 체크리스트

01. 기관 운영 관리
02. 큐레이션 계획 및 관리
03. 기술, 식별, 연결
04. 이해관계자 관찰 및 협력
05. 이용(자) 조사
06. 정보기술 모니터링
07. 개념화
08. 생성 또는 수집
09. 평가 및 선정
10. 입수
11. 보존
12. 저장
13. 접근, 이용, 재이용
14. 변환
15. 마이그레이션
16. 처분
17. 재평가

## 03

## 제3장. 디지털 큐레이션 체크리스트

디지털 큐레이션을 수행할 의향이나 계획이 있는 정보기관은 가장 먼저, 현재 자관의 디지털 큐레이션 역량, 인프라, 상황을 진단해 볼 필요가 있다. 정보기관은 그러한 결과를 디지털 큐레이션 계획을 세우는 데 기초 자료로 사용할 수 있다.

정보기관은 제3장에서 제공하는 체크리스트를 다음과 같은 목적으로 사용할 수 있다.

**첫째**, 디지털 큐레이션과 관련한 기관의 현재 상황을 확인: 디지털 큐레이션과 관련하여 어떤 활동들을 어떻게 하고 있는 지, 디지털 큐레이션과 관련하여 어떤 자원을 어떻게 사용하고 있는 지 등

**둘째**, 체계적이고 효율적인 디지털 큐레이션을 위해 기관이 개선해 나가야 할 부분 파악: 어떤 활동들을 추가 또는 보강해야 하는 지, 어떠한 요소(e.g., 표준, 정책)를 보충해야 하는 지 등

**셋째**, 디지털 큐레이션의 본격적인 수행을 위해 준비해야 할 부분 파악: 예산, 인력, 정책, 시스템 등

**넷째**, 기관의 기존 작업과 디지털 큐레이션 활동과의 관련성 인지

**다섯째**, 직원들에게 디지털 큐레이션에 대한 관심 유발

제3장에서는 KOREA 모델의 각 활동에 따라 체크리스트를 제공하는 데, 제2장의 ‘디지털 큐레이션 가이드라인’에서와 마찬가지로, ‘기관의 운영과 관리’라는 체크리스트를 추가하였다. 독자는 제3장에 있는 모든 체크리스트들을 참고할 수도 있고, 독자가 점검하고 싶은 특정 활동에 해당하는 체크리스트만 참고할 수도 있다. 제3장의 결과를 보고, 기관이 부족하거나 개선해야 할 활동이 있다면, 제2장에서 해당 활동 부분을 찾아 참고할 수 있다. 이렇게 제2장과 제3장을 상호 참조하여 이용할 수 있다.

# 01

## 기관 운영 관리

**목표:** 기관의 디지털 큐레이션 계획 수립 이전, 기관의 기본적인 역량, 환경, 인프라를 점검한다.

### CHECKLISTS

- 01 기관의 전반적인 운영을 모니터링하고 보고하는 관리 프로세스가 있는가?

---

- 02 기관의 자원을 효율적으로 사용하기 위한 자원 할당 정책과 매커니즘을 갖추고 있는가?

---

- 03 기관의 종합적인 발전 전략에 디지털 큐레이션을 포함할 수 있는가?

---

- 04 기관 직원들 사이에 디지털 큐레이션의 필요성에 대한 인식과 합의가 있는가?

---

- 05 기관에는 디지털 큐레이션을 수행할 수 있는 충분한 인력이 있는가?  
→ 만약 없다면, 인력을 보강할 수 있는가?

---

- 06 디지털 큐레이션 업무를 효율적으로 수행하기 위해 기관의 조직을 변경하거나 새로운 팀이나 부서를 만들 필요가 있는가?

---

- 07 디지털 큐레이션을 수행할 수 있는 충분한 예산을 가지고 있는가?  
→ 만약 없다면, 추가 예산을 마련할 수 있는가?

---

- 08 디지털 큐레이션을 수행할 수 있는 충분한 정보기술 인프라를 가지고 있는가?  
→ 만약 없다면, 정보기술 인프라를 추가로 마련할 수 있는가?

---

- 09 디지털 큐레이션을 수행할 수 있는 충분한 시설설비를 갖추고 있는가?  
→ 만약 없다면, 시설설비를 추가로 마련할 수 있는가?

# 02

## 큐레이션 계획 및 관리

**목표:** 1) 기관에서 디지털 큐레이션 계획을 세우기 전, 어떠한 점을 고려하여 계획을 세워야 하는 지 파악한다; 2) 기관이 디지털 큐레이션 계획을 세운 후, 중요한 내용이 빠지진 않았는지 점검한다.

### CHECKLISTS

- 01 큐레이션에 필요한 자원(e.g., 인력, 예산, 시설)이 준비되었나?
- 02 큐레이션을 위한 장·단기 전략을 가지고 있는가?
- 03 큐레이션 계획을 수립하고 이를 문서화 했나?
- 04 수립된 큐레이션 계획을 실행하기 위한 책임자를 결정하고 이를 문서화 했나?
- 05 수립한 큐레이션 계획을 누가, 언제, 어떻게, 얼마나 자주 검토하고 업데이트할 것인지 결정하고 이를 문서화 했나?
- 06 큐레이션 계획을 변경할 수 있는 매커니즘을 가지고 있는가?
- 07 큐레이션과 관련된 기관 정책(e.g., 수집, 평가, 보존, 접근)이 있는가?
- 08 큐레이션을 수행할 부서/직원의 역할과 책임을 규정하고 이를 문서화 했나?
- 09 큐레이션의 업무절차와 워크플로우를 결정하고 문서화했나?
- 10 큐레이션 활동별로 어떤 내용을 문서로 작성하여 보관할 것인지 결정했나?
- 11 기관의 데이터가 사용될 수 있는 대상 사용자/커뮤니티(target user/community)를 정했나?
- 12 디지털 큐레이션에 필요한 표준(e.g., 파일 포맷 표준, 메타데이터 표준, 파일 식별 표준)을 파악하여 알고 있는가?
- 13 사고나 재해 발생을 대비하여 매뉴얼과 데이터 복구 계획을 세웠나?
- 14 기관 직원들의 디지털 큐레이션 기술과 전문성 개발을 위한 교육이나 훈련 프로그램 제공 계획(다른 기관의 프로그램 수강 포함)을 세웠나?

# 03

## 기술, 식별, 연결

목표: 기관이 데이터를 효율적으로 기술, 식별, 연결하고 있는 지를 평가한다.

### CHECKLISTS

- 01 직원들이 기술, 식별, 연결에 대한 필요성과 중요성을 인식하고 있는가?

---

- 02 기술, 식별, 연결과 관련하여 표준을 사용하고 있는가?

---

- 03 기술, 식별, 연결을 위한 정책을 수립하여 문서화 했나?  
→ 그렇다면, 그러한 정책을 시행하고 있는가?

---

- 04 미래의 이용자들이 기관의 소장 데이터를 이해할 수 있도록 하기 위해,  
해당 데이터에 관하여 어떤 정보를 제공해야 하는지 알고 있는가?

---

- 05 메타데이터의 캡처 또는 생성 방법이 정해져 있는가?

---

- 06 표준을 사용하여 메타데이터를 작성하는가?

---

- 07 메타데이터가 해당 데이터와 연결되어 있는 다른 정보자원들(e.g., 보고서,  
학술논문, 데이터셋, 소스 데이터)에 대한 정보도 충분히 제공하는가?

---

- 08 데이터와 관련된 정보자원들(e.g., 보고서, 학술논문, 데이터셋)을 어떻게  
연결할 것인지 결정해 두었나?

---

- 09 기관에서 사용할 식별자를 정하여 사용하고 있는가?

---

- 10 작업 과정이 가능한 한 자동화되어 있는가?

# 04

## 이해관계자 관찰 및 협력

**목표:** 기관의 이해관계자들(e.g., 상위 기관, 출판사, 학회, 후원 기관, 연구자)을 알고 있고 그들과 협력하고 있는 지 확인한다.

### CHECKLISTS

- 기관이 보유하고 있는 데이터의 생성자들, 기탁자들, 이용자들로부터

**01** 문서화된 절차에 의해 피드백을 지속적으로 제공받는가?

→ 그렇다면, 제공받은 피드백에 따라 필요한 조치를 하는가?

---

- 02** 자관의 이해관계자들을 파악하고, 이에 대해 기록해 두었나?

---

- 디지털 큐레이션과 관련하여 이해관계자들과 협업하는가?

**03** → 그렇다면, 이해관계자들의 책임과 역할을 명확하게 정해 두는가?

---

- 현실적이고 실행 가능한 워크플로우를 파악하기 위해

**04** 이해관계자들과 협력하는가?

---

- 다른 정보기관의 데이터를 재이용 할 수 있는 권한과

**05** 데이터 공유에 필요한 제한 사항을 고려하는가?

→ 그렇다면, 그러한 제한 사항에 대해 해당 정보기관과 협의하는가?

---

- 기관은 이해관계자들과의 협업을 위해 적절한 협업 환경

**06** (e.g., 정보시스템, 네트워크)을 제공하는가?

---

- 07** 다른 관련 기관들의 디지털 큐레이션 활동에 관한 정보를 수집하는가?

# 05

## 이용(자) 조사

**목표:** 기관이 자관의 현재 이용자와 잠재적 이용자를 알고 있는 지, 기관의 자원(e.g., 데이터, 디지털 큐레이터, 정보시스템)과 이용자 간의 상호작용을 조사하는지를 점검한다. 더 나아가 그러한 조사 결과를 기관의 실무에 활용하고 있는지 확인한다.

### CHECKLISTS

- 01 기관의 실제 이용자를 파악하고 있는가?

---

- 02 기관의 잠재적 이용자를 파악하고 있는가?

---

- 03 기관은 자관의 이용자와 이용에 대해 조사할 수 있는 방법, 도구, 채널을 가지고 있는가?

---

- 04 기관은 자관의 데이터 이용통계를 조사하고, 분석하고, 문서화하는가?

---

- 05 이용자의 변화하는 요구를 지속적으로 조사하는가?  
→ 그렇다면, 그 조사 결과를 실무에 반영하는가?

---

- 06 기관 이용자의 데이터 추구 행태를 지속적으로 조사하는가?  
→ 그렇다면, 그 조사 결과를 실무에 반영하는가?

---

- 07 데이터, 정보시스템, 서비스 등에 관한 이용자 만족도를 주기적으로 조사하는가?  
→ 그렇다면, 그 조사 결과를 실무에 반영하는가?

---

- 08 기관의 이용(자)에 대한 조사 결과를 기관의 실제 디지털 큐레이션 실무에 활용하는가?

# 06

## 정보기술 모니터링

**목표:** 발전하는 정보기술을 지속적으로 모니터링하고, 새로운 정보기술을 적극적으로 수용하여 실무에 적용하는 지를 확인한다.

### CHECKLISTS

- 01 발전하는 정보기술을 지속적으로 모니터링 하는가?

---

- 02 새로운 정보기술을 도입하여 기관의 큐레이션 실무를 향상시킬 수 있는 방안을 마련하는가?

---

- 03 발전하는 정보기술을 적극적으로 수용하여 큐레이션 활동과 서비스를 향상시키는가?

---

- 04 기관의 정보시스템 성능을 지속적으로 모니터링하고, 새로운 기술을 도입하여 성능을 업그레이드 하는가?

---

- 05 새로운 정보기술을 신속하게 받아들여 다른 기관과 협업할 수 있는 환경을 만드는가?

---

- 06 디지털 큐레이션 커뮤니티와 협력하여 새로운 정보기술을 디지털 큐레이션 발전을 위해 활용하는가(e.g., 큐레이션 도구 개발, 큐레이션 프로세스 자동화)?



# 07

## 개념화

**목표:** 디지털 큐레이션 과정과 결과를 생각하면서 데이터의 생성과 수집을 구상하고 계획하는 활동을 하고 있는지 검토한다. 기관에서 데이터를 생성하지 않는 경우 아래의 질문들 중에서 데이터 생성과 관련된 질문들은 건너뛴다.

### CHECKLISTS

- 01 기관에서 생성하고자 하는 데이터의 유형, 포맷, 수량을 계획하는가?

---

- 02 기관에서 수집하고자 하는 데이터의 유형, 포맷, 수량을 계획하는가?

---

- 03 기관에서 사용할 데이터 포맷이 결정되어 있는가?

---

- 04 기관에서 데이터를 생성하는가?  
→ 그렇다면, 새로운 데이터의 생성을 계획할 때 생성 비용을 예측하는가?

---

- 05 기관에서 데이터를 수집하는가?  
→ 그렇다면, 새로운 데이터의 수집을 계획할 때 수집 비용을 예측하는가?

---

- 06 기관에서 데이터를 생성하거나 수집하기 위한 표준이 정해져 있는가?

---

- 07 기관에서 데이터를 생성하거나 수집하기 위한 방법과 절차가 정해져 있는가?

---

- 08 생성 또는 수집하는 데이터의 저장 매체와 저장 장소를 결정해 두는가?

---

- 09 저작권에 관하여 고민하고 어떻게 할 것인지를 결정한 다음 문서화하는가?

---

- 10 데이터 공개 시 발생할 수 있는 요구사항(e.g., 엠바고, 여러 곳에서의 출판 금지)을 파악하고 있는가?  
→ 그렇다면, 파악한 요구사항에 대하여 적절하고 신속하게 대응하고 있는가?

# 08

## 생성 또는 수집

**목표:** 기관이 데이터의 생성과 수집을 효율적으로 하고 있는지 검토한다. 기관에서 데이터를 생성하지 않는 경우 아래의 질문들 중에서 데이터 생성과 관련된 질문들은 건너뛴다.

### CHECKLISTS

- 01 데이터 생성과 수집에 관한 정책을 수립하고 문서화 했나?  
→ 그렇다면, 그러한 정책을 실제로 시행하고 있는가?

---

- 02 누구를 위하여 기관에서 데이터를 생성하는지 알고 있는가?

---

- 03 잠재적 이용자들이 기관에서 생성한 데이터를 가지고 무엇을 하고 싶어 할지 예측하고 있는가?

---

- 04 누구를 위하여 기관에서 데이터를 수집하는지 알고 있는가?

---

- 05 잠재적 이용자들이 기관에서 수집한 데이터를 가지고 무엇을 하고 싶어 할지 예측하고 있는가?

---

- 06 새로운 데이터를 생성 또는 수집할 때 그 데이터가 왜 필요한 지 설명할 수 있는가?

---

- 07 새로운 데이터를 생성 또는 수집할 때 그 데이터의 저작권을 누가 소유하는지 명확히 하고 이를 문서화 하는가?

---

- 08 새로운 데이터를 생성 또는 수집할 때 그 데이터의 이용과 재이용을 위한 라이선스(license)를 명확히 하고 이를 문서화 하는가?

---

- 09 데이터 계약 또는 기탁 협약을 문서화 하는가?  
→ 그렇다면, 계약상의 의무 또는 법률 위반에 대비한 법률적 대처방안을 문서로 가지고 있는가?

---

- 10 데이터, 포맷, 소프트웨어, 이용 동의사항, 그리고 출처에 대한 문서를 수집하고 보관하는가?

---

- 11 데이터의 생성 단계에서 데이터의 콘텐츠(content), 구문(syntax), 구조(structure)를 위해 기관에서 사용하는 표준이 있는가?

---

- 12 데이터 생성 과정에서 데이터 보호 요구사항을 파악하고 그것에 대해 다른 직원들과 이야기하는가?

---

- 13 기관에 데이터 품질 평가 기준이 있는가?  
→ 만약 있다면, 기관에서 생성했거나 수집한 데이터가 최소한의 데이터 품질 기준을 만족하는지 확인하는가

# 09

## 평가 및 선정

**목표:** 기관에서 생성 또는 수집한 데이터를 평가하여 장기간 큐레이션 할 데이터를 선정하는 기관의 활동을 검토한다.

### CHECKLISTS

- 01 누구를 위해 어떤 데이터를 소장해야 하는지 알고 있는가?

---

- 02 데이터 평가와 선정을 위한 기관의 정책을 만들어서 문서화 했나?  
→ 그렇다면, 그러한 정책이 실제로 시행되고 있는가?

---

- 03 데이터 평가 기준이 있는가?  
→ 만약 있다면, 그 기준에 따라 장기 큐레이션 할 데이터를 선택하는가?

---

- 04 데이터 평가와 선정 시 현재의 이용뿐만 아니라 미래의 이용도 고려하는가?

---

- 05 선정된 데이터의 보존 기간을 결정하는가?

# 10

## 입수

**목표:** 다른 곳에 있던 데이터를 기관의 데이터베이스, 디지털 아카이브 또는 정보시스템에 추가하기 위해 준비하고 실제로 추가하는 활동에 대해 점검한다.

### CHECKLISTS

- 01 입수에 대한 정책을 수립하고 문서화 했나?   
→ 그렇다면, 그 정책이 실제로 시행되고 있는가?
- 02 입수를 위해 기관에서 해 줄 수 있는 지원과 지침을 제공하고 있는가?
- 03 입수를 위해 사용할 수 있는 도구를 파악하고 사용하고 있는가?
- 04 입수 프로세스를 정립해 두었나?
- 05 입수 프로세스의 자동화를 위해 노력하고 있는가?
- 06 입수 프로세스 중 미리 정해둔 시점에, 입수하고 있는 데이터의 생성자 또는 기탁자에게 그 진행 상황을 알리는가?
- 07 누가 어떤 메타데이터를 어떻게 작성할 것인지를 결정하고 그 결정을 문서화 했나?
- 08 다른 기관이나 개인으로부터 데이터를 전송해 올 때 데이터 품질 평가를 하는가?
- 09 기관에서 데이터를 입수하는 시점에서 데이터 품질 보장을 마지막으로 누가(e.g., 데이터 생성자, 큐레이터) 책임져야 하는 지 결정되어 있나?
- 10 데이터 품질 보장의 결과가 모든 이해관계자들에게 전달되었는지 확인하는가?
- 11 기관에는 모든 데이터의 출처를 인증하는 메커니즘이 있는가?   
데이터의 장기 보존을 위한 준비(e.g., 바이러스가 있는 지 확인하고 제거; 관련 있는 메타데이터를 추출, 작성, 추가; 데이터의 무결성 체크를 돕기 위해 체크섬, 해시값을 생성; 기관에서 사용하는 파일 포맷인지 확인하고, 필요하면 포맷 변환)를 하는가?

# 11

## 보존

**목표:** 데이터의 무결성, 진본성, 신뢰성, 이용가능성을 유지하면서 데이터의 장기 보존을 위해 기관에서 실행하고 있는 활동을 점검한다.

### CHECKLISTS

- 01 기관의 보존 정책을 수립하고 문서화 했나?  
→ 그렇다면, 수립한 보존 정책을 시행하고 있는가?
- 02 장·단기 보존 전략을 수립하고 문서화 했나?  
→ 그렇다면, 수립한 보존 전략을 실행하고 있는가?
- 03 보존 프로세스를 정립해 두었나?
- 04 보존 프로세스의 자동화를 위해 노력하고 있는가?
- 05 특정 보존 포맷을 채택했나?  
→ 그렇다면, 그 포맷은 현재의 정보기술에 적합한가?
- 06 보존 메타데이터를 작성, 변경, 또는 추가하는가?
- 07 데이터 보존을 담당하는 직원들은 메타데이터나 다른 방법을 통하여 보존 할 데이터의 속성을 파악하는가?
- 08 데이터 클리닝을 하는가?
- 09 유효성 검사를 하는가?
- 10 데이터에 행해진 모든 보존 조치를 기록하여 남기는가?
- 11 디지털 보존과 관련하여 위기 평가와 관리까지 고려하는가?
- 12 기관이 책임지고 있는 데이터를 보존하기 위하여 기관의 법적 권리를 명시해 두었나?

# 12

## 저장

**목표:** 관련 표준에 따라 기관의 데이터와 관련 기술(description)을 안전하게 보관하고 있는지를 점검한다.

### CHECKLISTS

- 01 안전한 데이터 저장을 위한 정책을 수립하여 문서화 했나?  
→ 그렇다면, 수립한 저장 정책을 시행하고 있는가?
- 02 데이터의 저장 매체와 저장 장소를 문서화하는가?
- 03 기관은 입수한 데이터를 서로 다른 지역의 안전한 장소에서 보관하는가?
- 04 데이터가 여러 곳에서 보관되고 있다면, 그 데이터의 버전 관리를 하는가?
- 05 백업에 대한 기관의 지침과 절차가 있는가?  
→ 만약 있다면, 그러한 지침과 절차를 실제로 따르고 있는가?
- 06 입수된 파일은 정기적으로 충분히 백업되는가?  
→ 그렇다면, 그 백업된 파일은 안전하게 저장되는가?
- 07 안전하게 저장되어야 하는 개인 또는 민감한 데이터가 기관에 있는가?  
→ 만약 있다면, 그 데이터를 제대로 보호하고 있는가?
- 08 기관은 충분한 저장소(storage)를 가지고 있는가?
- 09 데이터를 어떻게 저장소로 전송할 것인 지를 문서화 하는가?
- 10 데이터와 함께 충분한 기술(description)을 저장하고 있는가?
- 11 저장되어 있는 데이터와 관련 기술(description)의 무결성을 정기적으로 체크하는가?
- 12 저장에 대한 책임자를 정하고 이를 문서화 했나?
- 13 저장에 대한 책임이 바뀌는 경우, 공식 수령증을 받거나 폐쇄 및 이전에 대한 비공식적 승인을 받는가?
- 14 데이터 보안 정책과 규정을 수립하고 그것을 문서화 했나?  
→ 그렇다면, 수립한 보안 정책과 규정을 시행하고 있는가?
- 15 보존 활동을 촉발하는 요인(e.g., 파일 포맷 마이그레이션, 파일 손상)이 있는지 모니터링하는가?
- 16 시간의 경과에 따른 저장 매체의 성능저하와 노후화에 대비한 저장매체의 변경 계획을 가지고 있는가?
- 17 데이터 복구를 위한 문서화된 지침과 절차가 있는가?  
→ 만약 있다면, 데이터를 복구 할 때 그 지침과 절차를 따르는가?

# 13

## 접근, 이용, 재이용

**목표:** 기관의 데이터가 일상적으로 이용, 재이용 될 수 있도록 데이터를 접근가능하게 유지하는지 점검한다.

### CHECKLISTS

- 01 데이터의 이용/재이용을 위해 기관에게 요구되는 법적 권한이 있는지 확인하는가?
- 02 계약 및 기탁 협약을 준수하는 접근 정책을 수립하고 문서화 했나?  
→ 그렇다면, 그 접근 정책을 시행하고 있는가?
- 03 이용자들이 웹 검색 엔진에서 기관의 소장 데이터에 접근할 수 있는가?
- 04 시간 경과에 따른 데이터의 맥락정보를 충분히 제공하여 원래 지정된 이용자 집단이든 새로운 이용자 집단이든 데이터를 찾아 이용할 수 있도록 하는가?
- 05 이용자들이 데이터로의 접근 권한을 얻을 수 있는 프로세스가 있는가?
- 06 데이터에 대한 접근 권한 및 접근 제한을 명확하게 서술하여 공지하고 있는가?
- 07 오직 승인받은 이용자들만 데이터에 접근할 수 있도록 접근 통제와 인증 절차를 사용하는가?
- 08 기관의 정보기술 인프라가 소장 데이터로의 접근 제한을 기록하고 지원하는가(e.g., 암호, 윤리, 액세스 동의 및 계약)?
- 09 데이터를 명확한 데이터 사용 조건(e.g., Creative Commons License)과 함께 공개하는가?  
소장 데이터의 재이용을 위해서  
기존 데이터에 새로운 '가치'를 추가하는 것을 고려해 본 적이 있나?  
→ 그렇다면, 실제로 기존 데이터에 새로운 가치를 추가하여 이용시키고 있는가?
- 11 데이터의 재이용을 위하여  
데이터 생성자, 데이터 소유자, 데이터 보관자를 명확하게 표시하는가?
- 12 기관 이용자들의 데이터 접근과 이용을 신속하게 도울 수 있도록 기관의 연락처 정보(e.g., 전화번호, 팩스 번호, 이메일 주소)를 공지하고 있는가?

# 14

## 변환

**목표:** 데이터를 마이그레이션 하여 새로운 포맷으로 만들거나, 특정 데이터셋을 가지고 새로운 서브셋을 만드는 등의 행위를 함으로써 원본 데이터(original data)를 가지고 새로운 데이터를 만드는 활동에 대해 점검한다.

### CHECKLISTS

- 01 파일 포맷 변환에 필요한 소프트웨어를 가지고 있는가?

---

- 02 데이터 포맷을 변환 할 때 데이터, 주석 또는 내부 메타데이터가 손실되거나 변경되지 않았는지 확인하는가?

---

- 03 최신의 정보기술을 반영하는 데이터 포맷에 대한 지식을 지속적으로 업데이트 하는가?

→ 그렇다면, 기관에서 누가 이러한 책임을 질 것인지 정해져 있는가?

---

- 04 선택이나 질의(query)를 하여 데이터셋의 서브셋을 생성하는가?



# 15

## 마이그레이션

**목표:** 디지털 객체의 지적 콘텐츠를 보존하고, 이용자들이 그것을 발견하고 이용할 수 있도록 어떤 기술(하드웨어나 소프트웨어에서의 생성)에서 다른 기술로, 어떤 포맷에서 다른 포맷으로, 디지털 객체를 이전하는 활동에 대해 점검한다.

### CHECKLISTS

- 01 필요에 따라 마이그레이션을 하는가?

---

- 02 얼마나 자주 마이그레이션을 해야 하는지, 어떤 포맷으로 마이그레이션을 해야 하는지, 어떤 데이터를 마이그레이션 해야 하는지, 누가 마이그레이션에 대한 책임을 질 것인지 등에 대한 기관의 정책이 있는가? 
  - 만약 있다면, 그 정책을 문서화 하고 실제로 따르고 있는가?

---

- 03 마이그레이션의 절차와 방법을 문서화 했나? 
  - 그렇다면, 그 절차와 방법을 실제로 따르고 있는가?

---

- 04 마이그레이션 될 파일 포맷과 데이터 구조를 정해 두었나?

---

- 05 메타데이터에 마이그레이션 절차를 충분히 기록하는가?

---

- 06 원본의 진본성을 유지하기 위하여 마이그레이션 후 원본의 품질을 확인하는가?

# 16

## 처분

**목표:** 기관에서 장기간 큐레이션을 할 필요가 없다고 결정한 데이터를 다른 정보기관이나 저장소로 이관하거나 안전하게 폐기하는 지를 점검한다.

### CHECKLISTS

- 01 데이터 처분(폐기 포함)을 위한 기관의 성문화된 정책과 절차가 있는가?   
→ 만약 있다면, 실제로 그것을 따르고 있는가?

---

- 02 데이터 처분 기준이나 지침이 있는가?   
→ 만약 있다면, 실제로 그것을 사용하고 있는가?

---

- 03 처분을 위한 정기적인 검토 프로세스를 정해두었나?

---

- 04 처분을 수행할 책임자가 정해져 있는가?

---

- 05 데이터 처분을 촉발하는 요인을 기관 차원에서 결정해 두었나?

---

- 06 기관의 정보시스템이 처분작업의 자동화를 얼마나 보조할 수 있는지 알고 있는가?   
데이터를 영구 폐기하는 경우,
- 07 법적으로 요구하는 사항이 있는지 확인하는가?   
→ 만약 요구되는 사항이 있다면, 그에 따라 영구 폐기하는가?

# 17

## 재평가

목표: 기관의 데이터를 다시 평가하는 활동에 대해 검토한다.

### CHECKLISTS

- 01 재평가에 대한 문서화된 정책이 있는가?  
→ 만약 있다면, 그러한 정책을 시행하고 있는가?

---

- 02 기관의 평가 정책이 재평가에 대한 내용  
(e.g., 재평가 원칙, 재평가 스케 줄, 재평가 조건)을 포함하고 있는가?

---

- 03 재평가 절차와 방법이 정해져 있는가?  
→ 만약 있다면, 그러한 절차와 방법을 따르는가?

---

- 04 재평가 책임자를 결정하고 이를 문서화 했나?

## 참고문헌

2018 KISTI 콘텐츠 큐레이션 가이드라인 (2018). 대전: 한국과학기술정보연구원

이혜원, 윤소영, 박지영, 황혜경, 김재훈, 이혜림 (2019). 연구콘텐츠 라이프 사이클 모델 개발: 한국과학기술정보연구원 콘텐츠 큐레이션 센터를 중심으로. 정보관리학회지, 36(3), 203-228. <http://dx.doi.org/10.3743/KOSIM.2019.36.3.203>

충남대학교 산학협력단 (2010). 수명주기 기반 디지털 콘텐츠 아카이빙 정책 연구. 대전: 한국과학기술정보연구원.

최정태 외 엮음 (2005). 기록관리학사전. 파주: 한울아카데미.

한국기록학회 엮음 (2008). 기록학용어사전. 경기: 역사비평사.

American Library Association ([n.,d.]). Metadata. <http://www.ala.org/tools/atoz/metadata/metadata>

Association of Recorded Sound Collections, 2009. Preservation of Archival Sound Recordings. Annapolis, MD: ASRC Technical Committee. [www.arcs-audio.org/pdf/ARSCTC\\_preservation.pdf](http://www.arcs-audio.org/pdf/ARSCTC_preservation.pdf)

Brown, Adrian (2003). Selecting File Formats for Long-Term Preservation. U.K.: National Archives. [https://webarchive.nationalarchives.gov.uk/+https://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](https://webarchive.nationalarchives.gov.uk/+https://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)

Center for Research Libraries (2007). Ten Principles. Chicago: Center for Research Libraries. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>

Cox, A. M., & Tam, W. W. T. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, 70(2), 142-157. <https://doi.org/10.1108/AJIM-11-2017-0251>

Craig, Barbara (2004). *Archival Appraisal*. Muchen: K.G. Saur.

Data Curation Network (2018). Checklist of CURATED Steps performed by the Data Curation Network. <http://z.umn.edu/curate> (also found at <https://datacurationnetwork.org>)

Digital Curation Centre (2013). Checklist for a Data Management Plan, v.4.0. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/data-management-plans>

Digital Curation Centre (2014). Five Steps to Decide What Data to Keep: A Checklist for Appraising Research Data, v.1. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides>

Digital Curation Centre ([n.d.]). Checklists. <https://www.dcc.ac.uk/guidance/how-guides>

Digital Preservation Coalition (2020). Digital Preservation Handbook. <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>

Higgins, S. (2007). Draft DCC curation lifecycle model. *International Journal of Digital Curation*, 2(2), 82–87. <https://doi.org/10.2218/ijdc.v2i2.30>

Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>

Humphrey, C. (2006). E-science and the life cycle of research. Retrieved from <https://era.library.ualberta.ca/items/3334684b-fa6a-4c9d-a74b-559fec42f9f/view/79b064d6-7b51-4d18-8e4e-3d42b9faa81f/Lifecycle-science060308.pdf>

ISO/TS 23081-2:2007: Information and documentation — Records management processes — Metadata for records — Part 2: Conceptual and implementation issues. <https://www.iso.org/standard/43390.html>

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2018). Data Curation Network: A Cross-institutional Staffing Model for Curating Research Data. *International Journal of Digital Curation*, 13(1), 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>

## 참고문헌

National Library of Australia (2003). Guidelines for the Preservation of Digital Heritage. [N.Y.]: Information Society Division. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000130071>

OAIS Reference Model (ISO 14721): The Fundamental Standard for Digital Preservation ([n.d.]). <http://www.oais.info/>

Oliver, G., & Harvey, R. (2016). Digital Curation. 2nd ed. Chicago, IL: 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. Retrieved from [http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch\\_curation.pdf](http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf)

UK Data Archive (2011). Managing and Sharing Data: A Best Practice Guide for Researchers. 3rd ed. Colchester: UK Data Archive. <https://ukdataservice.ac.uk/media/622417/managingsharing.pdf>

United Kingdom. The National Archives ([n.d.]). Checklist: Testing for Continuity of a Digital Asset. <https://www.nationalarchives.gov.uk/documents/information-management/testing-for-continuity-checklist.pdf>

United Kingdom. The National Archives ([n.d.]). Disposal Checklist. <https://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/disposal/disposal-checklist/>

University of Essex, University of Manchester and Jisc (2020). Data Management Checklist. <https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx>

Whyte, A. & Wilson, A. (2010). How to Appraise and Select Research Data for Curation. DCC How-to Guides. Edinburgh: Digital Curation Centre. <https://www.dcc.ac.uk/resources/how-guides>



## 부록

-

1. 주요 용어 해설
2. 디지털 큐레이션 관련  
유용한 사이트

## 부록 1. 주요 용어 해설

### ㄷ 디지털 큐레이션 *digital curation*

디지털 데이터/정보를 오랜 기간 관리하고 보존하는 것이며, 많은 생각과 적당한 시간과 자원의 투입을 요구하는 지속적인 프로세스이다. 데이터의 수집, 관리, 보존, 서비스 활동; 데이터에 가치를 추가하는 활동; 데이터의 무결성, 신뢰성, 진본성, 유효성, 접근가능성, 이용가능성을 보장하는 활동 등을 포함한다.

### 디지털 큐레이터 *digital curator*

기관에서 디지털 큐레이션 실무를 수행하는 직원들을 총칭한다.

### 데이터 *data*

디지털 형식의 정보로서 “디지털 객체”와 “데이터베이스”를 포함한다. “디지털 객체”는 1)텍스트, 이미지 또는 사운드 파일—그것들과 관련된 식별자와 메타데이터 포함—처럼 개별 디지털 객체; 2) 웹사이트처럼 여러 디지털 객체가 연관되어 만들어진 복합 디지털 객체를 포함한다. “데이터베이스”는 컴퓨터 시스템에 저장된 레코드 또는 데이터의 구조화된 집합체를 말한다 (Higgins, 2008, p.137).

### 데이터 클리닝 *data cleaning*

데이터의 결함이나 에러를 찾아 고치거나 삭제함으로써 데이터의 품질을 향상시키는 프로세스이다.

### ㄹ 라이프 사이클 *life cycle*

우리말로는 “생명주기”, “생애주기”, “수명주기”라고 불린다. 원래 라이프 사이클은 “유기체의 탄생에서 죽음에 이르기까지 거치는 일련의 연속적인 단계”를 말한다. “디지털 큐레이션” 분야에서는 큐레이션 대상이 그 일생 동안 통과하는 일련의 단계를 말하는 은유적인 표현으로 사용된다.

### ㄴ 맥락 정보 *contextual information*

맥락(전후 사정, 사물 따위가 서로 이어져 있는 관계)과 관련된 정보이다. 데이터와 관련된 맥락 정보로는 데이터의 출처, 생성 목적, 생성자, 수집처 등이 있다.



**메타데이터 metadata**

검색을 위해 기계가독형으로 구조화된 데이터셋에 대한 정보이다. 메타데이터 요소에는 기본 정보(e.g., 제목, 저자명, 생성 날짜) 및/또는 데이터셋의 고유한 특정 요소(e.g., 공간적 범위, 시간적 기간)가 포함될 수 있다.

**모범 실무 best practice**

소정의 목표를 달성하는 데 적절하고도 효과적인 수단이라는 점이 경험과 연구를 통해 입증되어 널리 인정받는 절차나 지침을 말한다 (한국기록학회, 2008, p. 98).

**문서화 documentation**

기록하여 문서(인쇄본 또는 전자본)로 보관하는 것을 말한다.

**버전 관리 versioning**

버전 이력과 각 버전에 대한 변경 사항을 설명하는 메타데이터를 포함하여 시간이 지남에 따라 새 버전의 데이터를 입수하기 위한 메커니즘을 제공한다.

**비트열 bit stream**

한 번에 한 비트씩 0과 1이 직렬 통신선로를 통해 끊임없이 연속적으로 전송되는 데이터의 흐름을 의미한다. 비트열은 8비트 단위로 구성할 경우 바이트열(byte stream) 또는 바이트스트림이라 한다 (기록관리학사전, 2005, p. 242).

**소스 데이터 source data**

의미 있는 사용을 위한 처리가 안 된 원시 데이터(raw data)를 말한다.

**○ 오픈 아카이벌 정보 시스템 Open Archival Information System: OAIS 참조 모델**

미국 National Aeronautics and Space Administration의 Consultative Committee for Space Data Systems에서 2002년에 발표한 참조 모델이다. 디지털 보존과 기록물 관리 영역에서 사용되는 개방형 정보시스템 개발을 위한 표준 참조 모델이다. 이 모델은 디지털 객체를 설명하는 정보 모델과 디지털 객체의 장기적 보존과 검색을 위한 시스템의 여섯 가지 기능 요건(수집, 저장, 데이터 관리, 접근, 보존 계획, 운영)을 정립하고 있다 (최정태 외, 2005, p. 347; OAIS Reference Model (ISO 14721): The Fundamental Standard for Digital Preservation, [n.d.]).

**이해관계자 stakeholder**

기관과 이해관계가 있는 조직들과 개인들을 말한다. 이해관계자들로는 해당 정보기관의 상위 기관들·후원 기관들, 다른 정보기관들, 출판사들, 학회들, 연구자들, 소장 데이터 주제 분야의 전문가들, 디지털 큐레이션 커뮤니티들을 포함한다.

**엠바고 embargo**

일반적으로 일정 기간 동안 데이터셋으로의 액세스를 제한하는 것을 말한다. 어떤 경우에는 접근 금지뿐만 아니라 데이터가 존재한다는 사실을 모르게 하기 위해 엠바고를 사용할 수 있다.

**위기 관리 risk management**

어떠한 위기가 올 것인지를 예상하고, 그러한 위기의 정도를 줄일 수 있도록 계획하는 것을 말한다.

**ㄷ 정규화 normalization**

(1) 오랫동안 사용될 것 같은 파일 포맷으로 변환 (2) 시간이 지남에 따라 이러한 포맷을 지원하기 위한 노력을 말한다.

**정보기관 information institution**

도서관, 아카이브, 데이터 센터 등 정보를 다루는 기관들을 포괄적으로 말한다.

**정보시스템 information system**

기관이 디지털 큐레이션을 수행하는 데 사용하는 컴퓨터 시스템(하드웨어, 소프트웨어, 네트워크 등 포함)을 총괄하여 말한다.

**지정된 커뮤니티 designated community**

아카이브에서 특정 정보 집합을 이해할 수 있어야 하는 잠재적 소비자 그룹이다. 이러한 소비자는 여러 커뮤니티로 구성될 수 있으며, 아카이브에 의해 지정되며, 시간이 지남에 따라 변경 될 수 있다 (OAIS Reference Model (ISO 14721): The Fundamental Standard for Digital Preservation, [n.d.]).

---

**㉔** **체크섬** *checksum*

데이터가 오류 없이 저장 또는 전송되었는지를 확인하는 데 사용되는 값이다.

---

**㉕** **큐레이션-준비가 되어있는 데이터** *curation-ready data*

데이터를 큐레이션하여 미래에도 접근가능하고, 공유되고, 재이용 될 수 있도록 준비하는 것이다.

---

**㉖** **핸들** *handle*

핸들은 URL을 현재 위치로 확인하는 중앙 레지스트리를 가지고 있는 인터넷 리소스에 대한 고유하고 영구적인 식별자이다. 각 핸들은 단일 리소스와 리소스를 생성했거나 현재 유지하는 조직을 식별한다. 핸들 시스템은 또한 특수한 유형의 핸들인 DOI의 기술 인프라를 뒷받침 한다 (Digital Preservation Coalition, 2020).

---

---

## 부록 2. 디지털 큐레이션 관련 유용한 사이트

- 
- 01** Canadian Heritage Information Network, <https://www.canada.ca/en/heritage-information-network.html>
- 
- 02** Curation Costs Exchange, <https://www.curationexchange.org/>
- 
- 03** Data Curation Network, <https://datacurationnetwork.org>
- 
- 04** Digital Curation Centre, <http://www.dcc.ac.uk>
- 
- 05** Digital Preservation Management: Implementing short-term strategies for long-term problems, [http://dpworkshop.org/dpm-eng/eng\\_index.html](http://dpworkshop.org/dpm-eng/eng_index.html)
- 
- 06** Library of Congress, <http://www.digitalpreservation.gov/>
- 
- 07** National Archives, <https://www.archives.gov/global-pages/subject-index.html>
- 
- 08** OAIS Reference Model (ISO 14721): The Fundamental Standard for Digital Preservation, <http://www.oais.info/>
- 
- 09** Open Preservation Foundation, <https://openpreservation.org/>
- 
- 10** UK Data Archive, <https://ukdataservice.ac.uk>
- 
- 11** University of California Curation Center (UC3), <https://cdlib.org/services/uc3/>
-





## 디지털 큐레이션

가이드라인과 체크리스트



9 788929 411671  
ISBN 978-89-294-1167-1

비매품/무료

93020

