

RiC을 적용한 아카이브 시스템 데이터 모델링 연구

A Study in the Data Modeling for Archive System Applying RiC

신미라(Mira Shin)¹, 김익한(Ikhan Kim)²

E-mail: shinpdnim@gmail.com, ikhan@mju.ac.kr



¹ 제1저자 명지대학교 기록정보과학전문대학원 기록관리전공 석사
² 교신저자 명지대학교 기록정보과학전문대학원 교수

논문접수 2019.1.8
최초심사 2019.1.31
게재확정 2019.2.11

초 록

Records in Contexts(RiC)은 ICA의 네 가지 기술표준을 통합, 정규화하여 개발한 국제적 기술표준이다. RiC은 기록 기술을 다차원적으로 바꾸고 기록의 맥락을 드러내며 이질적 시스템 간 데이터 상호운용성을 확보할 수 있는 장점이 있다. 본 논문에서는 RiC을 아카이브 시스템 설계의 핵심 도구로 설정하고, 데이터베이스 구현을 위한 '논리적 데이터 모델링'을 수행한다. RiC의 개념 모델인 RiC-CM(Conceptual Model)은 데이터 참조 모델로 활용할 수 있기 때문에, 사용자 요구사항에 맞는 데이터 모델로 개발이 가능하다. 이에 본 논문에서는 이를 두 가지 데이터 모델, '관계형 데이터 모델'과 '그래프형 데이터 모델'로 구현하고자 한다. 관계형 데이터 모델은 대부분 레거시 시스템의 데이터베이스가 적용하고 있는 만큼 범용적이다. 한편 그래프형 데이터 모델은 정보 '개체(entity)' 사이의 '관계(relationship)'를 중심으로 개체를 유연하게 확장할 수 있다.

© 한국기록관리학회

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.

ABSTRACT

Records in Contexts (RiC) is an international archival description standard developed by integrating and normalizing four archival description standards of the International Council of Archives (ICA). RiC has the advantage of diversifying archival description, exposing the context of records, and ensuring data interoperability between disparate systems. In this study, RiC is set up as a key tool in the design of archive systems, and logical data modeling is performed to implement the database. Because of RiC's conceptual model, RiC-CM can be used as a data reference model, and which makes it possible to develop a data model that meets user requirements. Therefore, this study intends to implement these two data models: relational data model, which is widely used as the database on legacy systems, and graphical data model, which can flexibly extend objects around the relationship between information entities.

• 본 논문은 신미라의 석사학위논문 「RiC을 적용한 아카이브 시스템 데이터 모델링 연구」(2009)를 수정한 것임.

Keywords: 기록물 기술, RiC(Records in Context), 데이터 모델링, 관계형 데이터 모델, 그래프형 데이터 모델
archival description, RiC (Records in Contexts), data modeling, relational data model, graph data model

1. 머리말

1.1 연구 배경과 목적

Records in Contexts(이하 RiC, '릭')는 ICA(International Council of Archive)에서 2016년에 발표한 이래 기존의 기술표준과 디지털 아카이브 패러다임을 전환할 수 있는 새로운 기술표준(standard of archival description)으로 부상하고 있다. 그렇다면 RiC의 실체는 무엇이고 실무에 어떻게 적용할 수 있는가? 그 답을 찾아가기 위해 본 연구는 RiC을 디지털 아카이브 시스템의 데이터 모델로 다루고자 한다.

국내 기록관리 현장에서 아직도 종이기록을 분류하고 기술하던 기존의 관행을 유지하고 있는 것(현문수, 설문원, 2018)은 기술표준을 시스템 입력 서식에 그대로 반영하는 시스템 설계 관행과도 맞닿아 있다.¹⁾ 시스템 설계를 위해 기술 요소(descriptive element)를 선정하지만, 선정된 기술 요소가 데이터 아키텍처²⁾ 수준에 적용되지 못하고 있는 것이다. 그 결과 시스템 화면(응용 프로그램 수준)과 데이터 구조가 같아지거나, 업무 프로세스 그대로 데이터 구조가 정의되어 데이터 관리 차원에서 무결성을 위반하게 된다. 본 연구는 이 점을 개선할 방법론으로 새로운 기술표준 RiC을 데이터 베이스 설계 단계부터 적용하려 한다.

전자기록시대에서 기술 요소의 탑재와 데이

터의 상호운용은 디지털 아카이브의 중요한 과제로 인식된다. 예를 들어, AtoM(Access to Memory)은 시스템 관리자 입력 화면에 ISAD(G)를 반영하고 있으며, 그 내부 데이터 모델은 ISAD(G), ISAAR(CPF) 등을 기반으로 한다. 또한 기술 요소를 기반으로 데이터 상호운용성(interoperability)을 확보하기 위해 데이터를 정의하고 참조모형을 구조화하는 개념 모델을 개발하기 시작했다. 호주의 AGRkMS, 스페인의 CNEDA, 핀란드의 FCMAD 등, 각 국가에서 ISO 23081을 참조하고 자국의 기술표준을 토대로 개발한 개념 모델이 바로 그것이다. 이 과정에서 데이터의 상호운용성 확보 범위가 국가와 분야(도서관·박물관 등)를 넘게 되면서 국제적 기술표준의 필요성이 대두하였고, 그 결과 RiC이 탄생하게 된 것이다(Llanes-Padron & Pastor-Sánchez, 2017).

RiC은 국제적 기술표준으로서 기록이 생산·관리·배포되는 맥락(context)의 복잡한 관계를 효과적으로 표현할 수 있도록 설계되었다(Artifactual Systems Inc., 2017). 이를 위해 RiC은 개념 모델(conceptual model)인 RiC-CM과 온톨로지(ontology)인 RiC-O를 제시한다. RiC-CM은 기록 맥락의 다차원적인(multi-dimensional) 특성을 기록과 연계하여 관리할 수 있도록 기존 ICA 4개의 표준³⁾에 분산된 기술 요소를 정규화하고 통합한 것이며, RiC-O는 기록정보 배포의 범위와 방법적 차원에서 기관 간, 시스템 간

1) 시스템을 전체 아키텍처 관점에서 설계하지 않기 때문인데, 촉박한 개발 일정, 부족한 인력, 개발의 수준 등 다양한 원인이 있다. 그리고 이는 데이터(기록) 구조가 응용 프로그램에 종속되는 결과를 낳을 수 있다.
2) 데이터 아키텍처는 데이터 참조 모델을 기반으로 조직의 업무 수행을 위해 필요한 데이터를 어떻게 정의하고 유지할 것인가를 정의하는 것이다(DBGuide.net, 데이터 전문가 지식포털). [cited 2018.12.31].
<http://www.dbguide.net/db.db?boardUid=12721&boardConfigUid=9&boardIdx=25>
3) ISAD(G), ISAAR(CPF), ISDF, ISDIAH.

상호운용성을 확보하기 위한 것이다. RiC은 데이터 개체와 관계를 기반으로 하고 상호운용성을 중시하고 있어, 도입할 경우 각 시스템에 분산된 데이터를 연계하기 쉽고 개발된 이후에는 적은 비용으로 시스템을 확장할 수 있을 것으로 기대된다. 또, 그 데이터를 이용해 새로운 서비스를 개발할 수 있으며, 세계적인 표준으로서 RiC을 도입한 모든 기관은 상호 데이터 협력을 할 수 있게 될 것이다.

이러한 주장을 뒷받침하기 위해, 현재 ICA에서는 RiC을 개발한 EGAD(Expert Group on Archival Description)의 하위 그룹을 결성해 RiC을 구현하기 위한 파일럿 프로젝트를 진행하고 있으며, 이에 대한 명세를 공개하고 있다(ICA EGAD, 2018). 이 프로젝트는 RiC-CM의 요소를 기록 정보에 매핑(mapping)해 데이터 모델링하는 것으로 RiC-O를 구현한다. 이때 데이터를 RDF(Resource Description Framework)로 표현하며, RDF 데이터로 변환된 기록 정보를 저장하고 시각화하기 위해 트리플 저장소(triplestore)를 채택해 그래프형 데이터 모델을 구현하였다. 이러한 RiC 파일럿 프로젝트를 통해 우리는 아직 개념적으로만 존재하는 RiC이 어떻게 데이터 상호운용성을 확보하고 연계할 수 있는지 알 수 있다. 그리고 실현 가능성 측면에서 앞으로 국내 RiC 연구에 시사하는 바가 크다.

디지털 아카이브의 관계자들이 RiC을 이해하기 위해서는 소프트웨어로서의 RiC에 대한 연구가 RiC의 필요성 연구만큼이나 중요하다. 본 연구는 지금까지 국내 RiC 연구 성과에서 한 발짝 더 나아가 RiC을 구현하기 위한 기술적(technical) 영역에서 접근하는 것을 목표로 한다. 동시에 RiC의 개념과 내용을 좀 더 명확

하게 파악하고 그 활용방안을 샘플링 방법을 통한 데이터 모델링으로 제시하고자 하였다.

1.2 선행 연구 검토

RiC에 대한 국내 연구는 주로 소개와 시사점 중심의 경향성을 띤다. 이는 아직까지 RiC이 현재 초안단계로 ICA에서 계속 수정과 보완을 하고 있다는 점과 국내에서 RiC 도입에 대한 공감대가 아직 확산되지 않은 점이 작용한 것으로 보인다. 그리고 무엇보다 연구자 개인이 RiC을 적용한 실험환경(테스트베드)을 구축하고 실효성을 검증하기가 쉽지 않다는 점이 더 큰 원인이라 할 수 있다.

지금까지 RiC 초안의 발표부터 수정과 검토, 각국 기록공동체의 의견 수렴 등 그 과정을 놓치지 않고 국내에 소개한 박지영의 연구(박지영, 2016, 2017a, 2017b)는 RiC에 대한 국내 기록학계의 관심을 불러일으키는 핵심적인 역할을 했다. 박지영은 RiC에 대한 연구 이전에도 이와 비슷한 FRBRoo, CIDOC-CRM과 같은 기술표준을 꾸준히 연구했는데(박지영, 2008, 2017c), 그 연장선에서 차세대 기록물 기술표준으로서 RiC을 분석하고, RiC을 적용하기 위한 테스트를 수행했다.

이유빈과 이해영(2017)은 시맨틱웹 환경에서 기록을 검색하는 인터페이스를 연구하며 그 기반이 되는 정보구조 모형으로서 RiC을 소개한 바 있다. 이 연구는 기록의 온톨로지 검색 인터페이스를 제안하며 RiC이 그 기반이 될 수 있다는 개념(concept)을 제공하였다는 데 그 의미가 있다.

현문수와 설문원(2018)은 차세대 공공 전자

기록관리체계 수립을 위한 연구를 위해 공공 전자기록의 조직화 원칙과 모형 설계의 방향을 제시하였는데, 이때 RiC이 공공 전자기록 조직 모형에 적합한 지에 대한 검토로써 RiC을 분석하였다. 이들은 RiC의 혁신성을 인정하지만 여전히 보완되고 정련될 필요가 있다는 국제사회의 의견을 참조하여, RiC을 무비판적으로 수용하지 말고 RiC의 원칙을 존중하되 후속연구를 통해 국내 디지털 환경에 맞는 실증적 전자기록 조직 모형을 개발해야 함을 시사했다.

이렇듯 RiC을 소개하고 시사점을 전달하는 연구들은 어느 정도 이루어졌으나, 본 논문의 연구주체인 RiC의 데이터 모델링에 관한 직접적인 연구는 아직까지 이뤄진 바가 없다. 따라서 RiC과 비슷한 기술표준이며 RiC의 설계에 바탕이 되기도 한 FRBR/FRBRoo와 CIDOC-CRM을 활용한 데이터 모델링 연구를 참고할 필요가 있다.

현문수(2014)는 로컬리티 정보자원 구조화 연구를 위해 FRBRoo/CIDOC-CRM을 기반으로 하는 개념 모델을 제시하였다. 이 연구에서 데이터베이스 논리 모델링 단계까지 기술되지 않았으나, 로컬리티 정보자원의 사례로부터 물리적 개체(physical entity)를 추출하고, FRBRoo/CIDOC-CRM의 개념적 개체(conceptual entity)와 연결하였다. 이렇게 구축된 개념 모델은 데

이터베이스나 온톨로지 구현을 위한 참조 모델로 활용할 수 있기 때문에 현문수(2014)의 연구는 데이터 모델링을 위한 FRBRoo/CIDOC-CRM의 유용성을 증명한 것이라 볼 수 있다.

DOREMUS(Doing Reusable Musical Data) 프로젝트는 FRBR/CIDOC-CRM의 요소를 적용해 클래식 음악 정보 온톨로지를 구현하고자 했다.⁴⁾ 데이터 구조 설계를 위해 FRBR/CIDOC-CRM의 핵심 정보 개체를 event, work, expression으로 정하고 개체와 개체간 맥락을 표현하는 하위 개체 및 관계를 세밀하게 조직화하였다. 이렇게 도출된 개념 수준 데이터 모델을 RDF로 변환하고 통제어휘집(controlled vocabularies)을 발행하였으며 RDF의 저장, 질의와 검색을 위해 그래프형 데이터 모델을 구현하였다. DOREMUS 프로젝트의 전체 개발 프로세스는 방법론적 측면에서 본 연구에 시사하는 바가 크다. 이미 검증된 방법이라 할 수 있으며 RiC-O를 구현하기 위해 ICA가 수행한 파일럿 프로젝트와 유사한 프로세스를 갖기 때문이다.

한편, RiC은 ISO 15489(기록관리 표준), 23081과 호환될 수 있도록 설계되었고(EGAD, 2016), ISO 23081과 개발 목적 및 구성 측면에서 상당히 유사하다.⁵⁾ 또 RiC의 적용단계에서 '레거시 시스템(legacy system)'의 데이터를 재활용할 필요도 있기 때문에 메타데이터 표준을 데이터

4) 2014년 말 시작된 DOREMUS는 시멘틱 웹 프로젝트로 인터넷 상의 클래식 음악 카탈로그를 기술하고 배포하고 맥락을 해석(contextualize)하는 도구와 수단을 개발하는 것을 목적으로 하고 있다. DOREMUS는 BnF, Radio France, Philharmonie de Paris, Eurecom 등 프랑스의 다양한 기관과 업체가 컨소시엄을 통해 개발하였으며 그 내용을 <http://www.doremus.org>(cited 2018.12.31) 사이트를 통해 공개하고 있다.

5) ISO 23081 2부에 해당하는 'ISO/TS 23081-2: 개념상·실행상 쟁점'에서 정의한 목적은 기록과 기록의 맥락을 표준화된 방식으로 기술할 수 있게 해주고, 기록의 생애 주기 전반에서 어떠한 공간이나 응용 소프트웨어 환경에서도 메타데이터를 재사용하고 표준화하기 위한 것이다(기록학용어사전, 2008). 이는 데이터의 재사용과 표준화, 시스템간 기관간 데이터 상호운용성, 기록을 둘러싼 다차원적 맥락의 확보 등 RiC을 개발한 목적과 의도에 정확히 부합한다.

모델링에 적용한 국내 사례를 살펴보고자 한다. ISO 23081을 참조하여 이를 데이터 모델링에 직접 적용한 연구 중 황진현과 임진희(2012)의 시각예술기록정보 관리를 위한 데이터 모델 설계 연구는 주목할 만하다. 이 연구는 시각예술정보를 데이터 모델링하기 위해 KS X ISO 23081 다중개체 모형을 적용하여 주요 개체(e.g. 기록 정보, 기록 하위정보 등)를 정리하고, 개체-관계를 모델링하는 절차와 방법을 제시하였다. 관계 모델링을 연관된 두 개체에 대해 각각 제시하면서 메타데이터 표준 요소를 적용한 개체와 개체로서의 관계를 정의하는 과정을 상세히 설명했다는 점에서 의미가 있다.

1.3 연구 방법

데이터 모델링(data modeling)이란 주어진 개념으로부터 논리적인 데이터 모델을 구성하는 작업을 말하며, 일반적으로 이를 물리적인 데이터베이스 모델로 환원하여 고객의 요구에 따라 특정 정보시스템의 데이터베이스에 반영하는 작업을 포함한다(이규화, 2011). 본 연구에서는 일반적인 개념으로서의 논리적 데이터 모델을 구성하는 작업에 국한하며, 데이터베이스 모델링이라 불리는 후자의 작업은 연구의 범위에서 제외하였다.

데이터 모델링 연구는 다음과 같이 진행되었다. 이론적 배경에서는 RiC 아키텍처를 살펴보고 RiC의 개체, 속성, 관계가 어떻게 구성되어 있는지를 파악한다. RiC은 개념 모델 RiC-CM과 온톨로지 모델 RiC-O로 구성되어 있는데,

이중 RiC-O의 개발 명세가 아직 발표되지 않았으므로 RiC 아키텍처에서는 RiC-CM의 요소를 중심으로 살펴볼 것이다. 이어서 앞서 검토한 RiC 아키텍처와 특징을 분석하여 모델링 개요를 살펴본다. 이때 모델링하는 각각의 데이터 모델, 관계형과 그래프형의 매커니즘을 소개하고 앞으로 도출될 데이터 모델이 어떤 모습으로 구현될 것인지 살펴볼 것이다.

이상의 내용을 토대로 본 연구에서는 본격적으로 RiC을 이용해 데이터 모델링을 시도해보려고 한다. 먼저 현행 디지털 아카이브 시스템이 대부분 관계형 데이터베이스를 기반으로 하고 있음을 고려하여 관계형 데이터 모델을 구현할 것이다. 관계형 데이터 모델링을 위한 도구로는 MySQL Workbench 8.0 CE를 활용하였다. MySQL Workbench는 MySQL 데이터베이스를 위한 오픈소스 비주얼 데이터베이스 설계와 질의 도구이다. 또, 정보 개체 간의 '관계'를 연결하여 네트워크 데이터 모델을 구축할 수 있는 그래프 데이터 모델의 구현도 함께 수행할 것이다. 이 때 그래프 모델을 구조화하고 질의하는 몇 가지의 방법 중 속성 그래프 모델을 구현하고자 한다. 도구는 현재 가장 대중적인 그래프 데이터베이스인 Neo4j(네오 포 제이) 그래프 데이터베이스를 활용한다(DB-Engines ranking of Graph DBMS, 2018). Neo4j는 GPL3 라이선스⁶⁾의 오픈소스 커뮤니티 에디션으로 개발사 홈페이지(<https://neo4j.com/>)에서 자유롭게 다운로드하여 이용할 수 있다.

각각의 모델링을 위한 샘플 데이터는 음악 저작물로서 임의로 선정한 클래식 음악 CD와 DVD

6) GNU General Public License, GNU GPL 또는 GPL, 자유 소프트웨어 재단에서 만든 자유 소프트웨어 라이선스.

메타정보를 활용하였다. CD와 DVD는 우리가 일상에서 흔히 접할 수 있는 기록물로 분류와 정리 개념을 설명하기 쉬운 대상이다. 하나의 작품이 완성되기 까지 음악 작곡, 레코딩, 촬영, 매체 제작 등 다양한 활동을 포함하기 때문에 여러 인물과 단체, 작품 관련 주제와 역사, 클래식 공연과의 연계 등 다양한 맥락을 표현할 수 있다.

2. 이론적 배경

2.1 RiC 아키텍처

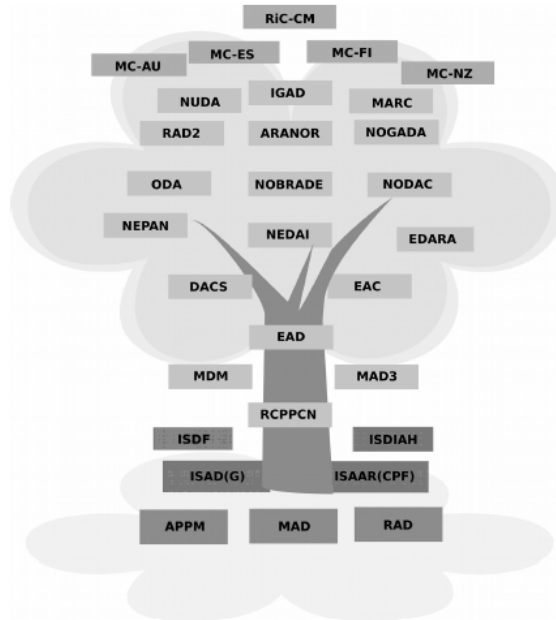
2016년 RiC 초안(RiC-CM v0.1)이 발표된 이후, ICA 산하 ‘기록물 기술 표준을 개발하기 위한 전문가 그룹’(EGAD, Expert Group on Archival Description, 이하 EGAD)은 계속해서 RiC을 보완하고 있으며 RiC을 적용한 참조 모델을 개발하는 등⁷⁾ RiC이 국제 표준으로서 활용될 수 있도록 노력하고 있다. 그 과정에서 ICA EGAD는 2018년 10월, RiC-CM v0.2와 RiC-O의 베타 버전을 준비하고 2019년 초 공개할 예정임을 밝혔다(EGAD, 2018). EGAD의 RiC 연구 과정과 내용은 “RiC-IM HOME: ICA Records in Context - Compendium” 사이트에서 공유되고 있다. RiC 포털과도 같은 이 사이트는 RiC과 관련한 모든 자료와 내용(관련 문서 파일, RiC의 개체, 속성, 관계의 상세한 명세와 자료 등)을 확인할 수 있으며, RiC이 업데이트 되는 최신의 내용을 반영하고 있다. 2018년

12월 31일 현재까지는 2018년 7월 수정 반영된 v0.2의 초안을 제공하고 있으며, ICA가 예고한 대로 2019년 RiC-CM v0.2가 공식 문서로서 발표되면 이 사이트가 정리한 내용이 포함될 것으로 보인다. 따라서 본 장의 RiC 아키텍처에서는 “ICA Records in Context - Compendium” 사이트에서 제공하는 최신 버전의 RiC(RiC-CM v0.2 Draft Version-2018.07)을 중심으로 다루고자 한다.

RiC 이전부터 상호운용성 확보를 위한 통합적 기술(description) 개발이 호주, 뉴질랜드, 스페인, 핀란드 각 국과 세계 도서관과 박물관기구에서 시도되었다. 그 결과 ISAD, ISAAR (CPF), ISDF, ISDIAH 등 다계층 기술 규칙(multi-level description rule), 호주 AGRkMS, 스페인 CNEDA, 핀란드 FCMAD 등 각국의 개념 모델, 도서관의 FRBR, 박물관의 CIDOC-CRM 등이 개발되었고, 이 모든 것을 분석한 그 정점에 지금의 RiC이 위치하고 있는 것이다(<그림 1> 참조).

Llanes-Padron, Pastor-Sánchez(2017)에 따르면 RiC 개념 모델의 목적은 기술적(descriptive) 과정에 관여하는 모든 개체를 상호 연결하여 표현하고, 이를 통해 다차원적 기술을 생성하는 것이다. 따라서 기록 정보 간 상호운용성이 RiC 개념 모델의 기본 원칙이며, 상호운용성 확보대상은 여러 아카이브, 도서관·박물관 등 서로 다른 기관·분야의 정보시스템에 담겨진 기록 정보이다. 한편, RiC-O는 개념적으로 제시되어 있을 뿐 아직까지 발표된 개발 명세가 없다.

7) ICA EGAD 산하 소그룹은 ‘프랑스 아카이브 기관들을 위한 상호운용성 프로젝트’(Pilote d’interopérabilité pour les autorités archivistique française, 이하 PIAAF 프로젝트)를 추진하였는데, 이 프로젝트의 참조 모델은 본 논문의 2.3 시맨틱웹 검색 도구로서의 RiC-O에서 다시 살펴볼 것이다.



〈그림 1〉 아카이브 기술을 위한 규범 및 개념 모델의 진화 과정

출처: Dunia Llanes-Padrón, Juan-Antonio Pastor-Sánchez(2017)

그러나 RiC-O를 구현하기 위해서는 RiC-CM의 요소를 기반으로 하기 때문에 본 논문에서는 RiC 개념 모델을 RiC 아키텍처로 소개하고자 한다.

RiC 개념 모델은 아카이브라는 현실 세계를 개념 세계로 추상화 했을 때 어떤 요소로 이루어져 있는지를 표현한 개념 구조이다. 일종의 데이터 모델로 아카이브 기술과 메타데이터 요소를 통합하고 정규화하여 개체(entity), 속성(attribute), 관계(relationship)로 정의하였다. 현실 세계를 표현하기 위한 주요 데이터를 식별하고 개념 세계로 옮기는 작업을 데이터 모

델링 과정 중에서도 개념적 모델링(conceptual modeling)이라 한다(김연희, 2013). 개념적 모델링의 산출물인 개념 모델은 데이터 모델링을 거쳐 논리적 모델로 구현되어 정보시스템에 반영된다.

RiC 개념 모델의 구성요소는 현재 15개의 주요 개체(primary entity)와 16개의 보조 개체(supporting entity), 개체의 속성(attribute)과 관계(Relationship)로 정의되어 있다(RiC-CM v0.2 Draft Version-2018.07).⁸⁾ 주요 개체는 객체지향적 상속(inheritance)⁹⁾ 개념을 활용하여 부모 개체에 해당하는 ‘최상위 기본 개체’와 자

8) RiC v0.1에서는 주요 개체와 보조 개체의 구분 없이 다음의 14개 개체로 제시되었다: Record, Record Component, Record Set, Agent, Occupation, Position, Function, Function(Abstract), Activity, Mandate, Documentary Form, Date, Place, Concept/Thing. (아직 공식 문서로 발표되지는 않았지만) RiC v0.2로 명세가 업그레이드되면서, 위의 개체는 주로 주요 개체로 이동하고, 나머지는 보조 개체나 속성 등으로 재정의 되는 등 전반적인 조정이 이루어졌다.

식 개체에 해당하는 하위 개체로 세분화하고 있다(〈표 1〉 참조).

주요 개체는 기록을 기술할 때 필수적으로 고려되어야 할 개념적·물리적 개체이다. 주요 개체 안에는 하위개체를 그룹화하는 Record Source(RiC-E00), Agent(RiC-E04), Activity(RiC-E11) 등 세 개의 최상위 기본개체로 구성되어 있다. 주요 개체는 ISAD(G), ISAAR(CPF) 및 ISDF 등의 기존 기술 표준, 호주의 시리즈 시스템, ISO 23081에 부합하는 기술 요소를 제공할 수 있도록 호환성을 갖는다(Records in Contexts-Compendium 홈페이지). 이에 따라 ISAD(G) 기반 기술데이터나 ISO 23081을 적용한 레저시 데이터를 RiC으로 전환하거나 적용하는 데 용이하고, RDF와 같은 표현으로 데이터를 재사용·재활용할 수 있다(Llanes-Padrón & Moro-Cabero, 2017).

보조 개체는 주요 개체를 기술한다는 면에서 속성에 해당하기도 하지만, 독립적인 개체로 정의되어 자체적인 속성을 보유할 수 있다. 이러한 점은 다중 값을 갖는 개체를 표현하는 데 유용하다.

예를 들어 하나의 Record(주요개체:RiC-E01)는 속성으로 여러 개의 Documentary Form(보조개체:RiC-E24)을 가질 수 있고, Documentary Form(보조개체:RiC-E24)도 개체로서 자체적으로 속성을 가질 수 있다(Records in Contexts-Compendium 홈페이지). 한편, 보조 개체는 4개의 범주(Appellation Entities, Classification Entities, Dates and Places, Concept or Thing)로 구분되며, 각 개체와 개체 간 관계에 대한 부여·식별·구체화·구분·분류의 기능을 띠고 있다(〈표 2〉 참조).

속성은 개체가 가지는 고유속성(properties) 또는 특성(characteristics)을 나타내는 요소로, 데이터 모델링의 유연성을 보장하는 수단으로 활용된다. 예를 들어, 개체-속성을 매핑 할 때 해당 개체를 도메인으로 하는 고유한 속성을 가질 수도, 다른 개체와 공유하는 속성을 가질 수도 있는 것이다. 〈그림 2〉는 주요 개체를 도메인으로 하여 RiC-CM의 속성을 정의한 모델을 예시로 보여준다.

〈표 1〉 RiC-CM 주요 개체

최상위 기본 개체	RiC-E00 - Record Resource	RiC-E04 - Agent	RiC-E11 - Activity
하위 개체	RiC-E01 - Record RiC-E02 - Record Part RiC-E03 - Record Set	RiC-E05 - Person RiC-E06 - Group RiC-E07 - Family RiC-E08 - Corporate Body RiC-E09 - Position RiC-E10 - Delegate Agent	RiC-E12 - Function RiC-E13 - Process RiC-E14 - Action RiC-E15 - Rule

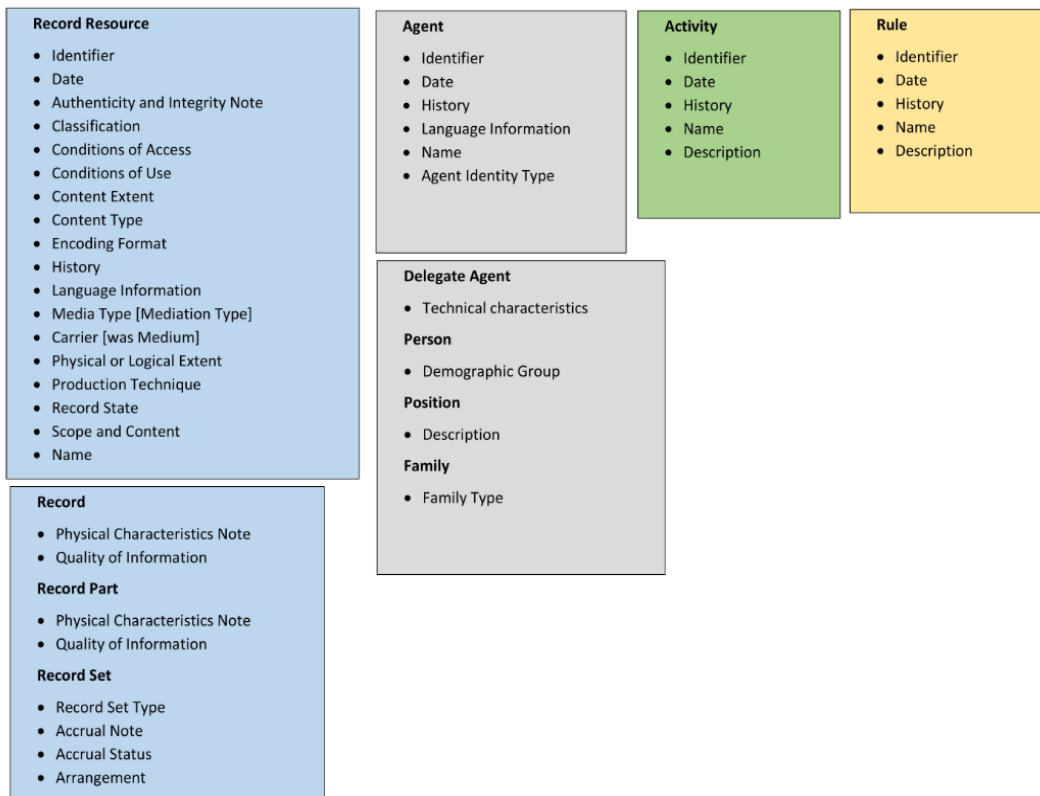
출처: Records in Contexts-Compendium 홈페이지

9) 객체지향 개발 방법론에서 클래스 간의 관계를 계층화하고 분류하는데 이런 개념이 바로 상속(inheritance)이다. 상위 클래스(super class)의 모든 것을 하위 클래스(sub class)가 물려받아 내 것처럼 사용함을 의미한다. 이때 물려주는 클래스를 상위 클래스 또는 부모 클래스라고 하고, 물려받는 클래스를 하위 클래스 또는 자식 클래스라고 한다.

〈표 2〉 RiC-CM 보조 개체

Appellation Entities	Classification Entities
RiC-E16 - Name RiC-E17 - Title RiC-E18 - Term RiC-E19 - Class Code RiC-E20 - Global Persistent Identifier RiC-E21 - Local Identifier	RiC-E22 - Occupation RiC-E23 - Activity Class RiC-E24 - Documentary Form RiC-E25 - Record Class
Date/Place	Concept/Thing
RiC-E26 - Date RiC-E27 - Single Date RiC-E28 - Date Range RiC-E29 - Date Set RiC-E30 - Place	RiC-E31 - Concept or Thing

출처: Records in Contexts-Compendium 홈페이지



〈그림 2〉 주요 개체와 각 개체의 속성을 정의한 모델 예시

출처: Records in Contexts-Compendium 홈페이지

관계는 RiC v0.2부터 개체의 구성 변화에 맞춰 좀 더 구체화되었는데, RiC v0.1에서 지적된 문제¹⁰⁾였던 관계 구조가 단순화되고 유연하게 개선되었다. <표 3>, <표 4>, <표 5>는 각각 Record

(RiC-E01)를 중심으로 정의된 RiC 관계를 보여준다. Record와 Record Set, Record 간 관계, Record와 주요 개체, 보조 개체 간 관계를 각각 정의하고 있다.

<표 3> Record(RiC-E01)의 관계: Record-Record, Record-Record Set, Record-Record Part, Record-Agent

<p>Record - Record Record has copy Record Record is copy of Record Record has draft Record Record is draft of Record Record has original Record Record is original of Record Record has subject Record Record is subject of Record Record is associated with Record Record is associated with Record Record was associated with Record Record was associated with Record Record is predecessor of Record Record is successor of Record</p>	<p>Record - Record Part Record had part Record Part Record Part was part of Record Record has part Record Part Record Part is part of Record Record is associated with Record Part Record Part is associated with Record Record Part was associated with Record Record was associated with Record Part</p> <hr/> <p>Record - Record Set Record is associated with Record Set Record Set is associated with Record Record was associated with Record Set Record Set was associated with Record Record is member of Record Set Record Set has member Record Record was member of Record Set Record Set had member Record</p>
<p>Record - Agent Record had rights held by Agent, Agent was rights holder of Record Record has rights held by Agent Agent is rights holder of Record Record has subject Agent Agent is subject of Record Record is associated with Agent Agent is associated with Record Record is held by Agent Agent is holder of Record Record is owned by Agent Agent owns Record Record was addressed to Agent Agent was addressee of Record Record was associated with Agent Agent was associated with Record Record was authored by Agent Agent authored Record Record was collected by Agent Agent collected Record Record was created by Agent Agent created Record Record was held by Agent Agent held Record Record was owned by Agent Agent owned Record Record was sent by Agent Agent sent Record Record was written by Agent Agent wrote Record</p>	

출처: Records in Contexts-Compendium 홈페이지

10) 박지영, RiC에 대한 기록공동체의 리뷰를 통해 본 기록물 기술표준 개선을 위한 제안, 기록학연구, (54), 81-109. 참고.

관련 개체의 속성 정의 미흡, 개체의 하위 레벨 정의에 대한 필요성, 다양한 분류나 색인체계를 표현할 수 있도록 해야 할 필요 등에 있어 속성이 개체로 모델링되거나 개체가 속성으로 모델링 되어야 하는 등 좀 더 명확한 정의가 필요하다는 의견(이탈리아 기록공동체 의견, 2017), 개념 모델인지 기술표준인지에 대한 명확한 제시, 핵심 개체 정의에 대한 미흡 등에 대한 의견(InterPARES Trust-Duranti, 2016), 누락된 개체와 속성(Artefactual Systems Inc., 2017), 기록관리를 위한 핵심 개체 외에 일부 개체는 속성의 성격이거나, 장소의 개체 표현이 장소 자체와 행위자 개체와 관련된 장소 표현과 겹쳐 혼동되는 점, 이벤트 개체를 추가하는 것에 대한 의견 등(Recordkeeping-Barbara Reed, 2017) 다양한 기록공동체 검토 의견이 있었다.

〈표 4〉 Record(RiC-E01)의 관계: Record-Activity, Record-Position, Record-Occupation, Record-Function, Record-Function(Abstract)

<p>Record - Activity Record has subject Activity Activity is subject of Record Record is associated with Activity Activity is associated with Record Record was associated with Activity Activity was associated with Record Record resulted from Activity Activity resulted in Record</p>	<p>Record - Function Record has subject Function Function is subject of Record Record is associated with Function Function is associated with Record Record was associated with Function Function was associated with subject of Record Record is evidence of Function Function is evidenced by Record</p> <p>Record - Function (Abstract) Record has subject Function (Abstract) Function (Abstract) is subject of Record Record is associated with Function (Abstract) Function (Abstract) is associated with Record Record was associated with Function (Abstract) Function (Abstract) was associated with Record</p>
<p>Record - Position Record has subject Position Position is subject of Record Record is associated with Position Position is associated with Record Record was associated with Position Position was associated with Record Record resulted from Position Position resulted in Record</p>	<p>Record - Occupation Record has subject Occupation Occupation is subject of Record Record is associated with Occupation Occupation is associated with Record Record was associated with Occupation Occupation was associated with Record Record resulted from Occupation Occupation resulted in Record</p>

출처: Records in Contexts-Compendium 홈페이지

〈표 5〉 Record(RiC-E01)의 관계: Record-Mandate, Record-Documentary Form, Record-Date, Record-Place, Record-Concept/Thing

<p>Record - Mandate Record has subject Mandate Mandate is subject of Record Record is associated with Mandate Mandate is associated with Record Record was associated with Mandate Mandate was associated with Record Record is evidence of Mandate Mandate is evidenced by Record</p>	<p>Record - Documentary Form Record has documentary form Documentary Form Documentary Form is documentary form of Record Record is associated with Documentary Form Documentary Form is associated with Record Record was associated with Documentary Form Documentary Form was associated with Record</p>
<p>Record - Date Record is associated with Date Date is associated with Record Record was associated with Date Date was associated with Record Record had creation date Date Date was creation date of Record</p>	<p>Record - Place Record is associated with Place Place is associated with Record Record was associated with Place Place was associated with Record Record has subject Place Place is subject of Record Record was created at Place Place was creation location of Record Record has holding location Place Place is holding location of Record Record had holding location Place Place was holding location of Record</p>
<p>Record - Concept/Thing Record has subject Concept/Thing Concept/Thing is subject of Record Record is associated with Concept/Thing Concept/Thing is associated with Record Record was associated with Concept/Thing Concept/Thing was associated with Record</p>	

출처: Records in Contexts-Compendium 홈페이지

주요 개체 간 관계는 <표 6>을 참조할 수 있다. RiC의 관계는 두 개의 RiC 개체를 연결한다. 관계는 개체 간 맺고 있는 의미의 연관성으로, 예를 들어 업무 처리에 대한 요구사항을 개체들을 이용해 하나의 문장으로 만들었을 때 ‘동사’에 해당하는 것이 관계라 할 수 있다. <표 3>, <표 4>, <표 5>와 <표 6>에서 살펴 본 바와 같이 RiC 관계로 제시된 요소는 ‘is associated with / was associated with’, ‘has subject / is subject of’, ‘resulted in / resulted from’과 같은 형식을 갖고 있다(<그림 3> 참조).

이렇듯, RiC은 개체(주요개체와 보조개체), 속성, 관계를 이용해 기술 요소를 데이터 모델링하는 개념 모델로 제시되었다. 따라서 RiC 개념 모델로 정보시스템 기능 요구사항에 맞춰 논리 모델로 구현하여 이에 따라 데이터 아키텍처를 설계할 수 있다. RiC을 이용하면 기존 기술표준에 내재된 계층구조를 계승할 수 있고

하나의 개념 모델로 여러 데이터 모델을 구현할 수 있기 때문에, 그래프 구조와 같은 새로운 데이터 모델로 전환할 수 있다. 이러한 RiC의 특징에 따라 기록을 유연하게 기술할 수 있고 기록물 데이터셋에 기술 요소를 쉽게 추가할 수 있어, 좀 더 효율적으로 데이터를 재사용하고 재활용할 수 있다(ANAI & ICAR, 2016).

2.2 데이터베이스 설계도구로서의 RiC-CM

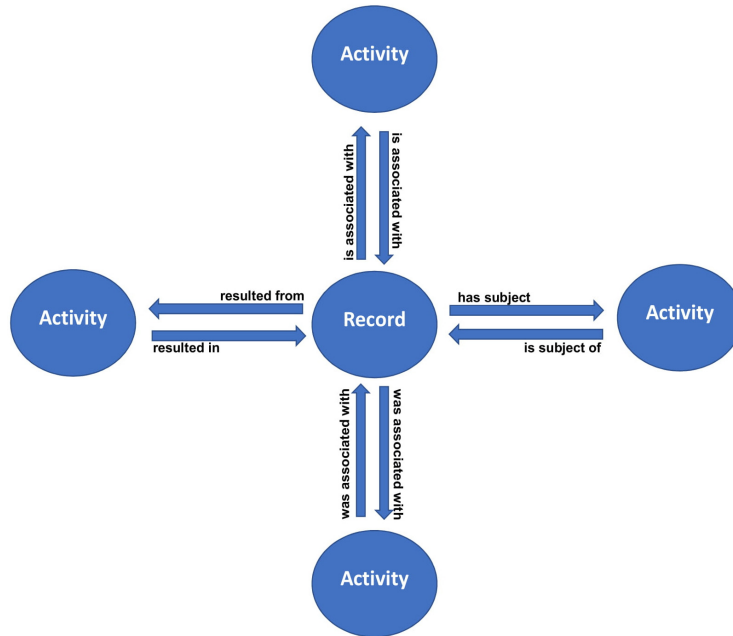
RiC이 개념 모델로 제시된 것은 근본적으로 데이터베이스와 상호관계를 고려한 것이라 할 수 있다. RiC을 이용하여 결과적으로는 아카이브 시스템 데이터베이스를 구성하여야 하기 때문이다. 따라서 본 장에서는 RiC이 데이터베이스와 어떻게 대응이 되고, 이를 통해 얻을 수 있는 장점이 무엇인지 살펴보고자 한다.

우리가 작성한 워드나 엑셀 파일이 컴퓨터에

<표 6> RiC-CM 주요 개체간 관계

Agent-Record Resource, is accumulator of	Agent-Agent, controls
Agent-Record Resource, is addressee of	Agent-Agent, has assumed identity
Agent-Record Resource, is annotator of	Agent-Agent, has functional relation with
Agent-Record Resource, is arranger of	Agent-Agent, is assumed identity of
Agent-Record Resource, is assembler of	Agent-Agent, is controlled by
Agent-Record Resource, is author of	Agent-Agent, is directed by
Agent-Record Resource, is collector of	Agent-Agent, is director of
Agent-Record Resource, is creator of	Agent-Agent, is subordinate of
Agent-Record Resource, is custodian of	Agent-Agent, is superior of
Agent-Record Resource, is holder of	
Agent-Record Resource, is rights holder of	Mandate-Agent, authorizes
Agent-Record Resource, is sender of	
Agent-Record Resource, manages	
Agent-Record Resource, owns	
Agent-Activity, has primary	Activity-Agent, is performed by
Agent-Activity, performs	Activity-Function, is performed to fulfill
Agent-Delegate Agent, uses	Activity-Mandate, is defined by
Agent-Function, fulfills	Activity-Place, is located in
Agent-Occupation, pursues	Activity-Records Resource, resulted in
Agent-Place, is located at	

출처: Records in Contexts-Compendium 홈페이지



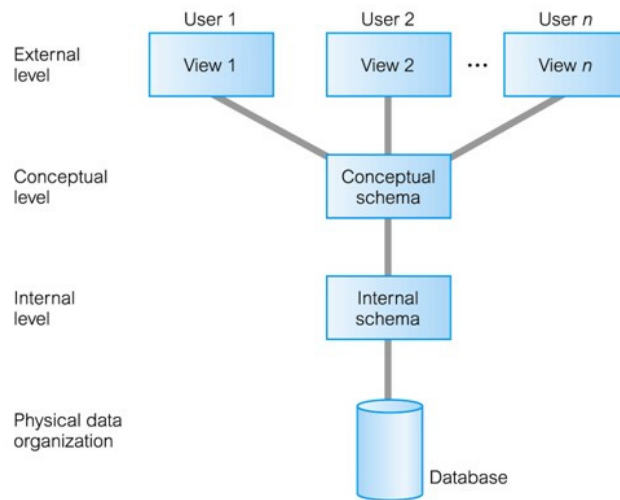
〈그림 3〉 Record(RiC-E01)와 Activity(RiC-E11)간의 관계 표현 다이어그램 예시

출처: Records in Contexts-Compendium 홈페이지

저장될 때는 파일 단위로 저장되지 않는다. 물리적으로 존재한다고 생각하는 파일은 실제 물리적 실체가 없다는 뜻이다. 데이터는 디스크에 할당된 블록에 저장되는데 심지어 그 블록도 파일크기에 맞춰 일정한 영역에 할당되지 않는다. 하지만 우리는 그런 디스크 블록에 대해 이해할 필요 없이 파일을 저장·복사·삭제할 수 있다. 데이터와 시스템 사이에 운영체제(OS)가 물리계층인 디스크와 논리계층인 파일을 연결해주기 때문이다(김상래, 2015). 데이터베이스의 저장구조도 디스크 저장구조와 비슷하다. 이를 데이터베이스의 '3단계 데이터 독립성 스키마 구조'(three-level architecture)라 하는데, 1978년 ANSI/SPARC에서 데이터베이스 관리 시스템(Database Management System, 이하 DBMS)과 그 인터페이스를 위한 아키텍

처로 제안되어 지금까지 대부분의 DBMS에 사용되고 있다.

〈그림 4〉의 데이터베이스의 3단계 스키마 구조 덕분에 실제 데이터가 저장되는 물리적 구조의 복잡함을 고려하지 않고 다루고자 하는 데이터를 논리적인 단위로 인식하여 데이터 모델링할 수 있게 된다. 3단계 스키마 중 최하단에 있는 내부 스키마(internal schema)가 바로 DBMS가 제공하는 영역이다. 내부 스키마는 DBMS가 데이터를 담은 논리적인 구조로 DBMS 종류에 따라 달라질 수 있다. 관계형 모델을 다루는 RDBMS(Relational DBMS)는 데이터 저장소를 2차원 표의 구조로 인식하고 데이터를 표현하며, 비관계형 데이터베이스는 각각의 저장소 구조를 다루는 종류에 따라 속성-그래프 모델, 키-값 모델, 트리플(주어-서술어-목적어) 저



〈그림 4〉 ANSI/SPARC DBMS 3단계 데이터 독립성 스키마 구조

출처: Pearson Education(2009)

장소 모델 등으로 표현한다. 이러한 점 외에 3단계 스키마 구조가 데이터 독립성을 보장한다는 장점도 있다.

최상단의 외부 스키마(external schema)는 사용자가 열람하는 정보 단위의 스키마이다. 데이터베이스에 접근하는 개별 사용자가 업무 내용과 사용 목적에 따라 필요한 데이터에 접근할 수 있도록 응용 프로그램의 데이터 설계에 종속적이다.

중간에 있는 개념 스키마(conceptual schema)는 데이터베이스의 논리적 구조로 개별사용자 관점을 통합하여 조직 전체의 관점에서 이해하고 표현한다. 즉, 외부 스키마의 통합 모델이라 할 수 있다. 역으로 외부 스키마는 개념 스키마를 기초로 하여 사용자의 이용목적에 맞게 만들어진 것이다(김연희, 2013).

내부 스키마의 논리적 구조를 논리 모델이라 할 수 있고 개념 스키마의 구조를 개념 모델이라 할 수 있는데, 여기에서 개념 모델은 데이터

모델링을 위해 주요 개체를 정의하고 개체의 키 속성(key attribute)과 주요 속성을 도출한 뒤 개체간의 관계까지 정리한 수준의 모델을 의미한다.

그렇다면 RiC의 데이터 모델은 이러한 데이터베이스 구조에 어떻게 적용될 수 있을까? RiC의 개념 모델은 키 속성에 대한 명시만 없을 뿐 데이터베이스의 개념 모델과 정확하게 일치한다. 키 속성도 RiC 보조 개체인 Global Persistent Identifier(RiC-E20)와 Local Identifier(RiC-E21)로 일부 제시된 것으로 볼 수 있다. 따라서 개체, 속성, 관계로 구성된 RiC-CM 아키텍처는 데이터베이스의 개념 모델에 대응된다.

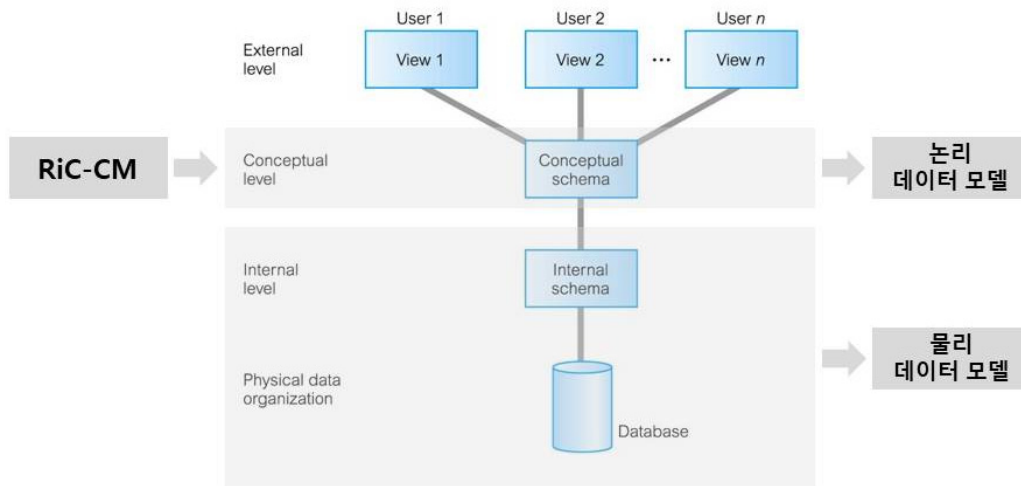
RiC은 기술 표준이기 때문에 개별 아카이브 시스템의 기능과 요구사항에 맞는 개체와 속성을 선별하고 관계를 재정의할 필요가 있다. 데이터베이스의 논리 모델도 마찬가지이다. 논리 모델은 개념 모델에 살을 붙여 상세하게 정의한 것인데, 주요 개체뿐만 아니라 모든 개체와 개별

개체의 모든 속성이 도출된 구체적이고 정규화(normalization)된 모델이다(김상래, 2015). 결과적으로 RiC-CM을 개별 아카이브 시스템에 적용하는 방식은 데이터베이스의 개념모델을 논리모델로 이행하는 과정으로 해석될 수 있다. 이러한 점을 <그림 4> 데이터베이스 3단계 데이터 독립성 스키마 구조를 바탕으로 RiC-CM이 어떻게 대응되는지를 살펴보면 <그림 5>와 같다. 즉, 데이터 모델링을 위해 개념 모델인 RiC-CM을 데이터베이스 개념 스키마(개념 모델)에 적용해 논리 모델을 도출하고 최종적으로는 물리 모델을 산출하는 것이다.

정보시스템으로서의 아카이브 시스템에서 데이터베이스의 중요성은 RiC의 적용이 아니라도 간과될 수 없다. 데이터베이스는 응용 프로그램의 데이터 종속성 문제를 해결하고, 데이터 중복을 최소화할 수 있으며 효율적으로

데이터를 검색, 저장할 수 있다. 또, 적절하게 데이터 접근을 제어하고 권한을 할당하여 보안성을 확보할 수 있고 장애 발생 시 회복 기능을 통해 백업과 복구를 지원한다. 이러한 점은 전자기록관리에서 반드시 준수되어야 하는 준칙이다.

그렇다면 RiC을 적용하는 데이터베이스가 기록관리에서 어떤 의미를 지니는가? 데이터베이스 자체의 장점 뿐만 아니라, 각 기관이 가지고 있는 기록들을 어떻게 연계하고 통합할 것인가에 대한 해결방안이 될 수 있다. 현재 각 기관의 기록관리 데이터베이스는 각각 다른 논리 모델을 가지고 있다 보니 기록을 통합하거나 연계하는 데 비용¹¹⁾이 많이 들고 제한적인 정보만을 대상으로 하게 된다. 기술표준으로서의 RiC은 이러한 한계를 넘을 수 있다. 이러한 점에서 RiC은 아카이브 시스템 데이터 모델의



<그림 5> RiC 개념모델을 데이터베이스 개념 스키마 단계에 대응
출처: Pearson Education(2009)에서 저자 편집

11) 이때의 '비용'은 각 기관 간 데이터 연계의 복잡함과 어려움으로부터 파생하는 시간과 노력도 포함된다.

핵심이 될 수 있으며, RiC이 정의한 개체, 속성, 관계 요소는 기록관리를 위한 충실한 메타데이터와 전거데이터를 통합 설계하기 위한 훌륭한 수단이 될 것이다.

2.3 시맨틱웹 검색 도구로서의 RiC-O

2.2에서 살펴본 데이터베이스로서의 RiC은 데이터 상호운용성, 확장 유연성과 메타데이터의 활용가능성을 염두에 둔 것이다. 이 지점에서 RiC이 단순히 데이터를 축적하는 것을 넘어 정보의 활용에 초점을 맞추었다는 것을 알 수 있다. 이를 위해 ICA는 데이터베이스의 개념 모델인 RiC-CM과 별도로 RiC-O를 시맨틱웹 기반 검색 도구로 제시했다.

본 연구의 범위에는 RiC 검색 도구 구현을 포함하고 있지 않지만, 시맨틱웹을 위한 RiC-O를 구현하면 검색 도구로서 어떤 기능을 할 수 있는지 알아보기 위해 ICA EGAD의 RiC-O 파일럿 프로젝트 사례를 살펴보고자 한다. 이 사례를 통해 RiC을 적용한 시맨틱웹 검색 도구를 제작하는 방법과 절차에 대한 시사점을 얻을 수 있을 것으로 본다.

ICA EGAD는 2015년부터 소그룹을 결성해 아카이브 기관 간 데이터를 RDF로 변환해 연계하기 위해 그래프형 데이터 모델을 구현하는 프로젝트를 수행해왔다. 그리고 그 결과 2018년 3월 RiC-O 개발 프로토타입을 완성했음을 발표했다(ICA EGAD, 2018). 프랑스 국립 아카이브(Archives Nationales(France)), 프랑스 국립 도서관(BnF), 프랑스 기록 보관소(Service interministériel des Archives de France, SIAF) 주도의 실험 프로젝트인 '프랑스

아카이브 기관들을 위한 상호운용성 프로젝트'(Pilote d'interopérabilité pour les autorités archivistiques françaises, 이하 PIAAF 프로젝트)가 바로 그것이며, RiC-O의 검색도구 구현 방법에 관한 내용을 온라인 사이트로 공개하고 있다(PIAAF 프로젝트 홈페이지).

RiC-O를 통해 기관 간 기록정보를 연계하기 위해서는 프로토콜로서의 데이터형식이 필요한데 이를 위해 PIAAF 프로젝트는 RDF를 활용했다. 데이터를 표현하는 방법으로 RDF 이전에는 HTML이 있었으나 HTML은 데이터의 서식과 의미가 혼용되어 있어 그 의미를 전달하는 데 한계가 있다. 이를 보완하기 위해 웹 페이지의 내용까지 포함하는 XML(Extensible Markup Language)이 탄생하고 XML을 기반으로 서로 다른 웹사이트가 일관된 형식으로 데이터를 게시하기 위한 방법으로 RDF가 탄생했다(마틴 클레프만, 2018).

그렇다면 RDF를 효과적으로 저장할 만한 데이터 저장소가 필요할 것이다. 이때 논리 모델로서의 데이터 저장소로 거론될 수 있는 것이 트리플 저장소(triplestore) 모델이다. 트리플 저장소는 본 논문에서 RiC 데이터 모델링에 활용한 속성 그래프(property graph)와 같은 그래프형 데이터 모델의 일종이다. PIAAF 프로젝트에서는 트리플 저장소 모델을 채택했으며, 그 구조를 살펴보면 트리플이라는 명칭처럼 모든 정보를 주어(subject), 서술어(predicate), 목적어(object)의 세 구문 형식으로 저장한다. 이와 달리 속성 그래프 모델에서는 개체와 관계를 저장한다.

그래프형 모델은 데이터를 정점(꼭짓점, vertex 또는 node)와 간선(변, edge)으로 표현하는데 속성 그래프 모델의 개체는 정점, 관계는 간선에 대

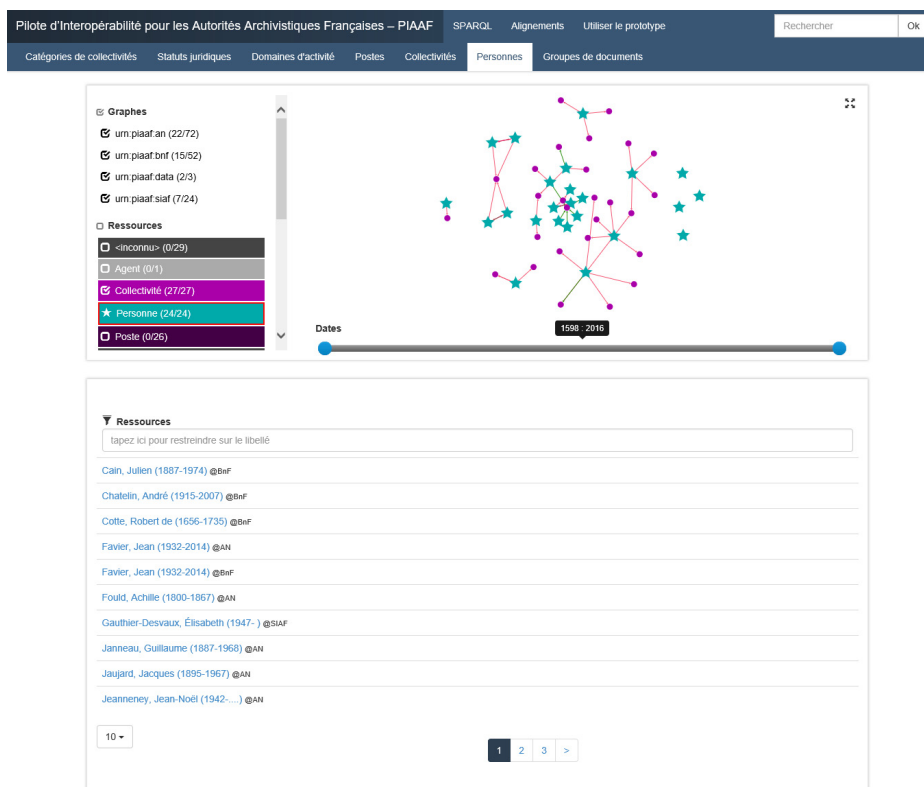
응되고, 트리플 저장소 모델은 주어와 목적어가 정점, 서술어가 간선에 대응된다(〈표 7〉 참조).

RiC 요소를 적용한 RDF와 트리플 저장소 모델을 활용하여 BnF, SIAF, Archives Nationales (France) 세 기관의 데이터를 상호 연결하고 시각화한 검색 화면은 〈그림 6〉과 같다. 〈그림 6〉은 RiC 주요 개체인 Person(RiC-E05)에 해

당하는 데이터의 연결 관계를 그래프 형태로 시각화해 보여주고 있으며, 그 관계의 구성요소(화면에서 Resources로 표시된 정보)를 검색 결과 하단에 나열하였다. 각 자원에 대한 상세 정보는 검색 결과의 항목을 클릭하거나 시각화된 그래프형 모델에서 그 정점(꼭지점)들을 클릭해 확인할 수 있다(〈그림 7〉 참조).

〈표 7〉 그래프형 데이터 모델로서의 속성 그래프와 트리플 저장소의 대응

그래프형 데이터 모델		속성 그래프	트리플 저장소
정점	↔	개체	주어, 목적어
간선	↔	관계	서술어



〈그림 6〉 PIAAF 프로젝트 데모 프로그램 화면1:
AN-BnF-SIAF 간 Person(RiC-E05)의 관계 표현과 검색 도구
출처: PIAAF 프로젝트 홈페이지

Interoperability Pilot for French Archival Authorities - PIAAF SPARQL Rosters Use the prototype Rechercher Okay

Categories of communities Legal status Business areas posts communities People Groups of documents

[Person] Favier, Jean (1932-2014)

http://piaaf.demo.logilab.fr/resource/FRBNF_person_0289 @ AN @ BnF

graphs

- urn: piaaf: bnf (4 / 4)

Resources

- <Unknown> (1 / 1)
- Community (1 / 1)
- Person (1 / 1)**
- Document group (1 / 1)

relationship

- identical to (1 / 1)
- leads (1 / 1)**
- is the archival provenance of (2 / 2)

attributes

ISNI identifier	000000120991284
Entitled	Favier, Jean (1932-2014)
birth date	1932
date of death	2014
history	Jean Favier is a senior official and French historian born April 2, 1932. He studied at the National School of Charters, and obtained the diploma of archivist paleographer in 1956. He is a member of the French School of Rome from 1956 to 1961. He received a degree in history in 1961. He was a doctor of letters in 1967. He became curator at the National Archives from 1968 to 1961. From 1962 to 1964, he was a teacher at the Lycée d'Orléans and then became attached to research at the CNRS (National Center for Scientific Research) between 1964 and 1964. He pursued a university career from 1965 (Brest, Rouen, Paris-Sorbonne). He is also director of studies at the Ecole pratique des hautes études from 1965 to 1997. He was appointed Director of the National Archives from 1975 to 1994. He then became the first President of the National Library of France between 1994 and 1997 and then President of the French National Commission for UNESCO. Jean Favier is also a member of the Académie des inscriptions et belles-lettres, president of the board of directors of the Ecole normale supérieure, administrator of the National Audiovisual Institute and honorary president of the International Council on Archives. He has published many historical books such as "The Pontifical Finances at the Time of the Great Schism of the West" (1966), "Finance and Taxation in the Late Middle Ages" (1971). "

relationship

type of resource	Person
similar to	http://data.bnf.fr/ark:/12148/cb11996420v#foaf:Person
predecessor of	Chronological relationship between the agents "Favier, Jean (1932-2014)" and "Angremy, Jean-Pierre"
has as a rule of thumb	Decree of 21 January 1994 appointing the President of the National Library of France
is described by	http://piaaf.demo.logilab.fr/resource/FRBNF_description_0289
occupies the position of	National Library of France. President
is the subject of	Collection of annual and multi-year activity reports of the Library
directed	[1994 - 1997] National Library of France More info
has for name	Favier, Jean (1932-2014) More information
is the archival provenance of	Collection of annual and multi-year activity reports of the Library More info

〈그림 7〉 PIAAF 프로젝트 데모 프로그램 화면2:
Person(RiC-E05)의 개별 항목별 관련 데이터 검색 페이지
출처: PIAAF 프로젝트 홈페이지

〈그림 7〉은 Person(RiC-E05)의 전체 관계 표현을 나타내는 상위 페이지에서 개별 항목 'Favier, Jean(1932-2014) @BnF'를 클릭해 그 상세정보로 이동한 화면이다. 'Favier, Jean(1932-2014)' 항목이 Person(RiC-E05) 주요 개체로 정

의되고 그 속성과 관계를 나타내는 정보가 모두 RiC의 요소로 표현되고 있다. 즉, 두 화면은 프랑스 국립 아카이브, 프랑스 국립 도서관, 프랑스 기록 보관소에 분산 관리되어 있는 'Favier, Jean'이라는 인물 정보가 Person(주요개체:RiC-E05)

를 통해 상호 연결되어 있으며, 인물 정보의 맥락을 RiC 그래프 데이터 모델링으로 시각화하고 검색할 수 있음을 보여준다.

이 사례를 통해 시맨틱웹 검색 도구로서의 RiC-O가 어떤 방식으로 기록정보 서비스를 제공할 수 있는지 알 수 있다. 지금까지 국가별 기관별 무수히 많은 기술 표준이 제정되고 활용된 것은 각각의 기관 목적에 맞는 기록을 생산·관리하며 이에 따라 기술 요소도 달라졌기 때문이다. RiC을 통해 데이터를 기술하고 검색 도구를 제공한다면, 서로 다른 기술 요소를 사용하는 기관 간 데이터를 상호 연결할 수 있게 된다.

RDF를 효과적으로 웹에 배포하기 위해서는 온톨로지 웹 언어(OWL, Ontology Web Language)를 사용하여야 하지만, PIAAF 프로젝트에서는 이를 활용하는 단계까지 구현하지는 않았다. 그러나 아카이브 데이터를 RDF로 표현하고 데이터 모델링 검색도구를 제공하여 시맨틱웹을 구축하는 데 RiC이 유용한 도구가 될 수 있음을 입증하고 있다. 또 기록 기술에 관여하는 모든 개체가 상호 연결될 수 있다는 점과 상호 연결을 통해 기록의 맥락을 다차원적으로 기술할 수 있다는 점에서 RiC의 목적과 의미를 확인할 수 있게 한다.

3. 데이터 모델링 방법과 RiC 모델링 개요

3.1 데이터 모델링 방법

이번 장에서는 데이터 모델링의 방법인 관계

형 데이터 모델링과 그래프형 데이터 모델링에 대해서 살펴보고자 한다. 본 연구의 주제인 RiC 개념 모델을 관계형 데이터 모델과 그래프형 데이터 모델로 설계하기 위함이다. 관계형 모델링은 2010년대 NoSQL이라는 비관계형 모델링이 나오기 전까지는 대체불가능하고 거의 유일한 모델링 기법이었다. 관계형 모델을 기반으로 하는 데이터베이스인 RDBMS는 막강한 ACID(Atomicity, Consistency, Isolation, Durability / 원자성, 일관성, 고립성, 지속성)의 트랜잭션과 조인 질의(SQL에서 두 개 이상의 테이블을 결합(join)하는 쿼리)로 인해 시스템과 데이터를 안정적으로 관리하려는 많은 사람들이 선호하는 도구가 되었다. 그러나 데이터가 테이블 두 개를 연결하는 2차원적 구조가 아니라 그 이상(다차원적)이라면 관계형 모델링에서는 복잡하게 표현할 수밖에 없다. 이를 극복하기 위해 NoSQL과 같은 비관계형 데이터 모델링 기법이 대두되었고 그 중 하나가 그래프형 데이터 모델링이다. 그래프형 모델링은 정보와 속성을 점과 선으로 각각 표현하고 이를 연결하면서 뻗어나가는 구조이기 때문에 다차원적인 구조를 갖는 데이터를 표현하기에 적합하다.

3.1.1 관계형 데이터 모델링

관계형 데이터 모델링은 데이터의 집합인 테이블 두 개가 공통적인 속성을 매개로 연결되면서('관계'를 맺으면서) 새로운 테이블로 구성되는 방식으로 모델링하는 기법이다. <그림 8>은 링크드인(Linked in) 프로필 정보를 관계형으로 데이터 모델링한 사례이다.

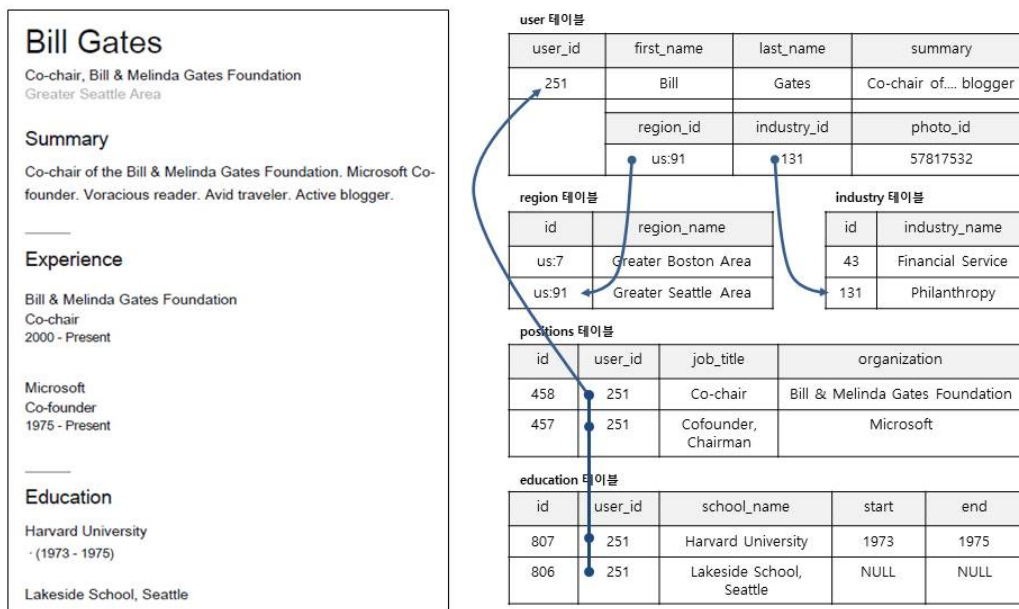
링크드인 프로필 정보의 가장 기본이 되는

것은 인적정보라 할 수 있다. 그 인적정보를 담은 테이블을 {user}라 한다면, {user}에는 이름(first_name), 성(last_name), 거주지(region), 업종(industry), 사진정보(photo), 학력(education)이 들어갈 수 있을 것이다. 그런데 개인에게는 이름과 성은 보통 한 개이지만, 사진, 학력, 거주지 등은 여러 개 있을 것이다. 하나의 테이블에 성명과 함께 사진 n개, 학력 m개, 거주지 x개를 저장한다면, 테이블에는 그 사람에 대한 정보가 $n \times m \times x$ 개 들어가게 된다. 이 중 사진 정보 하나만 바꾸려면 한 개의 정보를 바꾸는 것이 아니라 $n \times m \times x$ 개 모두 수정해야 하고 이 중 하나라도 수정하지 않으면 다른 사람으로 인식된다(무결성 위반이라 한다). 이러한 문제점을 해결하기 위해 <그림 8>처럼 테이블을 각각 쪼갠 뒤, 공통의 식별자(id)로

묶는 방법을 택하는 것이 관계형 데이터 모델이라 할 수 있다. 즉, {education} 테이블에 Bill Gates의 학력정보를 저장하고, 이 사람의 고유번호인 user_id로 {user}테이블과 연결한다('관계' 맺는다).

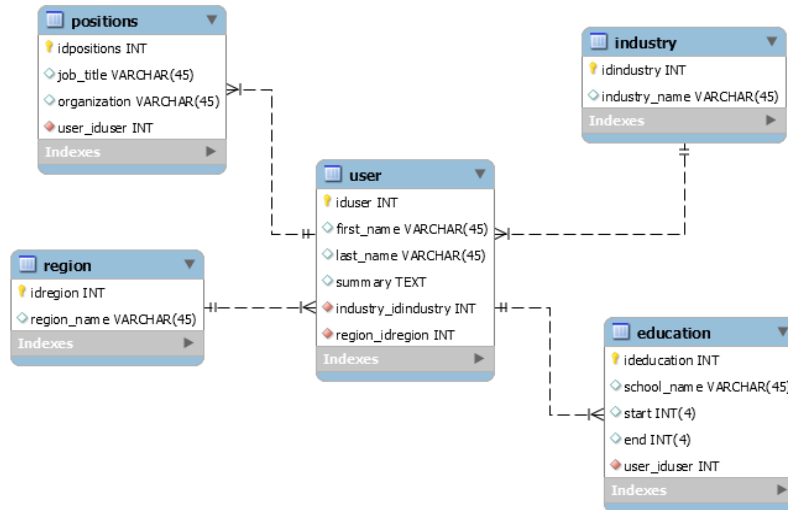
이러한 관계형 논리 모델은 데이터베이스 설계도구를 활용하면 물리 모델로 전환하기 쉽고, SQL 생성, 관리, 유지를 위한 개발환경을 제공할 수 있다. <그림 8>의 관계형 논리 모델을 데이터베이스 설계도구인 MySQL Workbench를 활용해 모델링하면 <그림 9>와 같다.

<그림 9>에서 상단 파란색 띠로 표시한 positions, user, industry 등은 테이블의 이름을 의미하고, {user} 하단의 iduser, first_name, last_name, summary, industry_id, region_id는 필드를 의미한다. iduser 왼쪽에 있는 노란



<그림 8> 관계형 모델을 사용해 링크드인 프로필을 표현

출처: 데이터 중심 어플리케이션 설계(2018)



〈그림 9〉 링크드인 프로필 정보를 MySQL Workbench를 활용해 설계한 관계형 데이터 모델

색 열쇠모양은 기본키(primary key)를 뜻하는 데, 기본키는 주민등록번호처럼 한 정보에 하나 밖에 없는 고유식별자이다. 빨간색 마름모는 외래키(foreign key)로서 다른 테이블(예를 들어 {industry})에서는 기본키(idustry_id)로, 또 다른 테이블(예를 들어 {user})에서는 참조되는 키를 뜻한다. 그리고 데이터베이스 물리적 설계를 위한 INT(정수), VARCHAR(가변형 문자열), TEXT(긴 문자열) 등의 데이터 형식이 표시되어 있다.

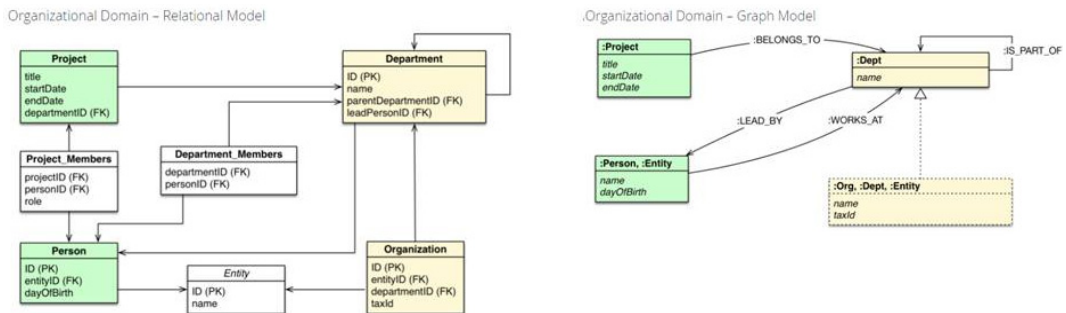
3.1.2 그래프형 데이터 모델링

앞서 설명한 바와 같이 NoSQL은 다양한 비관계형 데이터모델 기법을 채택하고 있는데, 이 중 하나가 그래프형 데이터 모델링이다. 본 장에서는 NoSQL과 NoSQL 데이터베이스 도구인 Neo4j로 그래프형 데이터 모델링을 설명하고자 한다.

그래프형 데이터 모델링은 정보(개체와 속성)를 정점(점)으로, 정보 간의 관계를 간선

(선)으로 표현하는 모델링 기법이다. 한 정보가 여러 세부 정보로 구성될 경우 한 정점(한 정보)과 여러 다른 정점(세부정보)이 여러 간선(관계)으로 연결될 수 있다. 또 세부정보는 각각 더 세부적인 정보로 구성될 수 있기 때문에 무한히 정점과 간선이 뻗어나가 트리구조를 만들 수 있다. 한편으로, 한 정점이 A라고 하고, 세부정보를 B라고 할 때 세부정보 B의 세부정보가 A인 순환구조를 가질 수 있기 때문에, 예를 들어 영화 '댓 씽 유 두(That Thing You Do!, 1996)'의 주연 배우 톰 행크스가 해당 작품의 감독이기도 한 정보를, 댓 씽 유 두(A) - "의 감독" - 톰 행크스(B) - "가 주연한 영화" - "댓 씽 유 두(A)"의 네트워크 구조도 만들 수 있다.

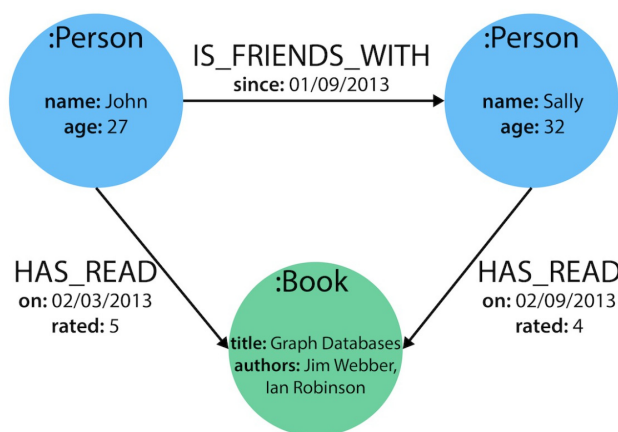
이와 같이 그래프 모델은 확장성이 좋고 데이터 구조 변경을 쉽게 수용할 수 있다. 〈그림 10〉에서 볼 수 있듯 관계형 데이터 모델(좌측)에서 복잡하게 표현되는 구조가 그래프형 데이터 모델(우측)에서는 간단하게 표현된다.



〈그림 10〉 동일한 Organizational Domain을 간략하게 관계형 모델과 그래프형 모델로 표현한 예시
출처: neo4j 블로그(rdbms vs graph datamodeling)

본 논문에서 데이터 모델링 도구로 사용할 Neo4j는 데이터 모델링은 물론 데이터 모델의 물리적 구현과 질의를 통합하는 속성 그래프 데이터베이스 설계 도구이다. 속성 그래프 모델은 정점인 노드(node)와 간선인 관계(relationship)로 구성하는데, 노드와 관계에 각 속성(property)을 포함할 수 있다. 〈그림 11〉에서 :Person :Book은 노드, IS_FRIENDS_WITH나 HAS_READ는 관계 :Person 내부의 name, age나 IS_FRIENDS_WITH 내부의 since는 속성을 나타낸다.

〈그림 11〉와 같은 도식을 생성하려면 Neo4j가 지원하는 선언형 질의언어 사이퍼(cypher)를 사용한다. 사이퍼 구문은 그래프 모델 표현과 거의 일치하는데, “A(Person) has read B(Book)”는 “(Person)-[:HAS_READ]->(Book)”로 표현된다. 속성은 키(name)와 값(John)으로 구성되는데 {키:'값'}으로 표시한다. 즉, “John이라는 Person의 이름은 John이고 나이는 27살이다”는 “(John:Person {name:'John', age:'27'})”로 표현된다.



〈그림 11〉 Neo4j를 통해 본 간략한 속성 그래프 모델 표현
출처: neo4j 홈페이지

좀 더 확장된 형태의 Neo4j 그래프 데이터 모델을 구현해보면, 그래프형 데이터 모델이 구현되었을 때 어떤 모습일지 확연히 알 수 있다. 다음 <표 8>은 Neo4j에서 제공하는 샘플 데이터 “Movie Graph”의 코드와 그래프 모델 구현 예시이다. 전체 코드 중 그 일부만 게시하였으

며, 전체 코드는 Neo4j 브라우저에서 “:play movie-graph” 실행으로 확인해볼 수 있다. “Movie Graph”가 표현하는 정보는 8개 영화와 9명의 관련 인물을 노드로 하여 그들 간 공동 작업관계 “ACTED_IN, DIRECTED”를 간선으로 나타내고 있다.

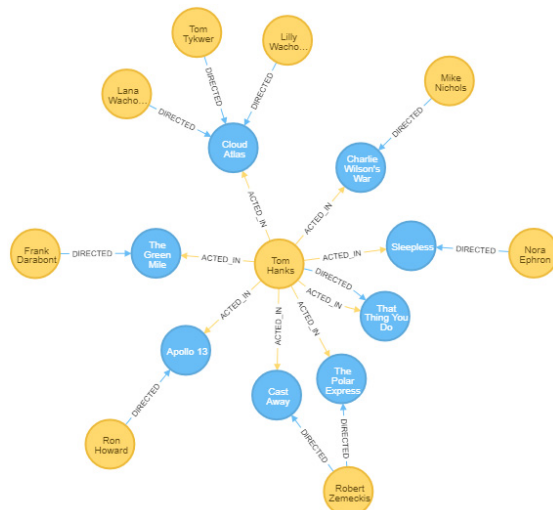
<표 8> Neo4j 그래프형 데이터모델(Neo4j Browser에서 “:play movie-graph” 실행)

```

(위) Neo4j 사이퍼 구문과 (아래) 해당 사이퍼를 구현한 그래프 데이터 모델
CREATE (TheMatrix:Movie {title:'The Matrix', released:1999, tagline:'Welcome to the Real World'})
CREATE (Keanu:Person {name:'Keanu Reeves', born:1964})
CREATE (Carrie:Person {name:'Carrie-Anne Moss', born:1967})
CREATE (Laurence:Person {name:'Laurence Fishburne', born:1961})
CREATE (Hugo:Person {name:'Hugo Weaving', born:1960})
CREATE (LillyW:Person {name:'Lilly Wachowski', born:1967})
CREATE (LanaW:Person {name:'Lana Wachowski', born:1965})
CREATE (JoelS:Person {name:'Joel Silver', born:1952})

CREATE
  (Keanu)-[:ACTED_IN {roles:['Neo']}]->(TheMatrix),
  (Carrie)-[:ACTED_IN {roles:['Trinity']}]->(TheMatrix),
  (Laurence)-[:ACTED_IN {roles:['Morpheus']}]->(TheMatrix),
  (Hugo)-[:ACTED_IN {roles:['Agent Smith']}]->(TheMatrix),
  (LillyW)-[:DIRECTED]->(TheMatrix),
  (LanaW)-[:DIRECTED]->(TheMatrix),
  (JoelS)-[:PRODUCED]->(TheMatrix)

//.....중간생략.....
//.....중간생략.....
//.....중간생략.....
WITH TomH as a
MATCH (a)-[:ACTED_IN]->(m)<-[:DIRECTED]->(d) RETURN a,m,d LIMIT 10;
    
```



출처: Neo4j Browser에서 “:play movie-graph” 실행

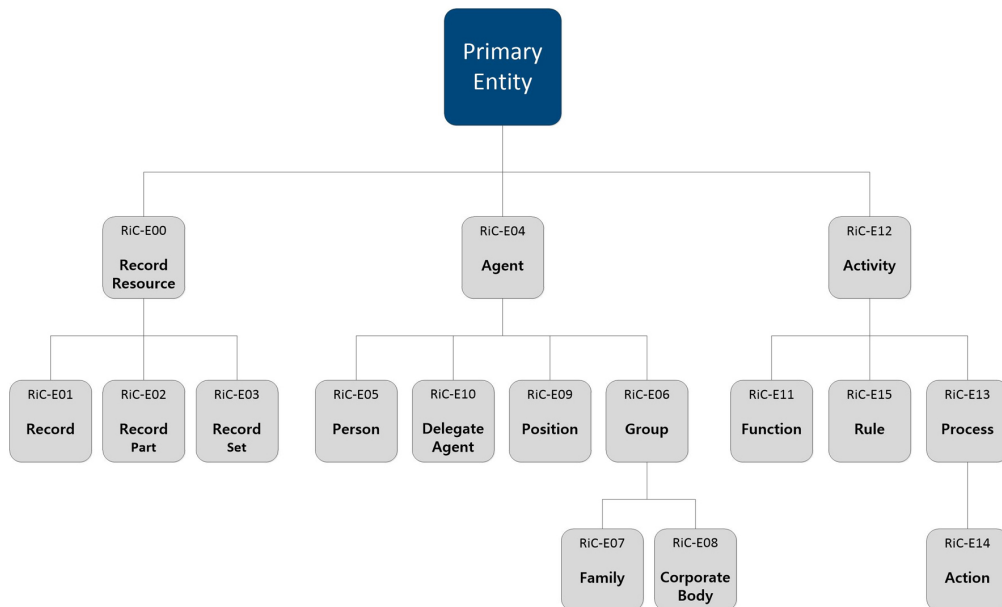
3.2 RiC 모델링 개요

앞서 2.1 RiC 아키텍처에서 이미 주요 개체, 보조 개체, 속성, 관계 등 RiC-CM의 요소를 살펴본 바 있다. 이번 장에서는 RiC요소를 구조적 측면에서 살펴보고자 한다. 구조적 측면이란 주요 개체와 보조개체를 도식화하여 RiC요소의 관련 개체 간 구조와 상속관계를 연결·파악하려는 것이다.

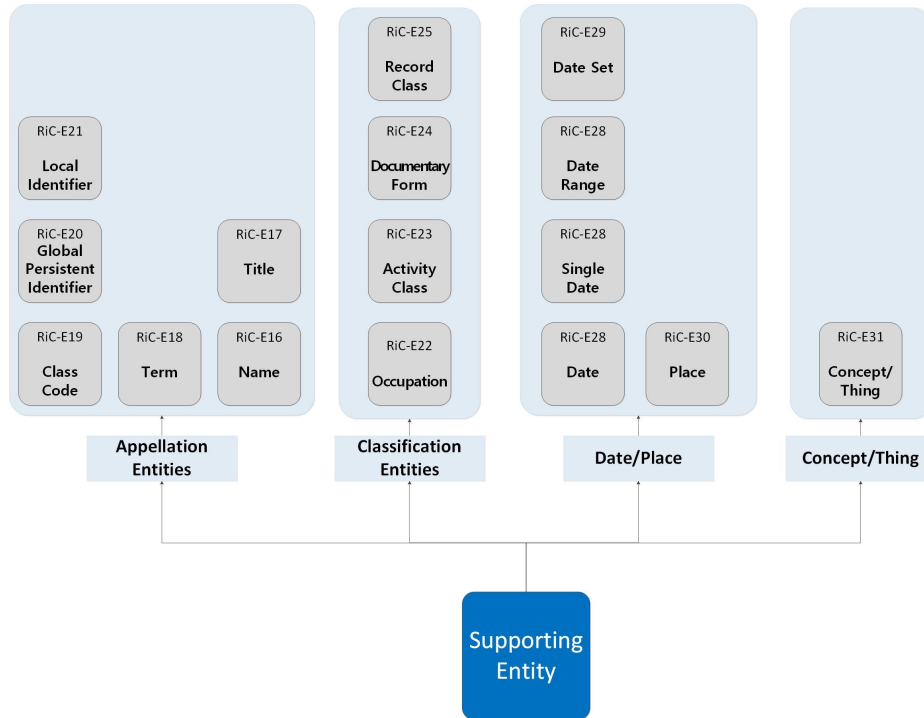
주요 개체 Record Resource(RiC-E00), Record(RiC-E01), Record Part(RiC-E02), Record Set(RiC-E03), Agent(RiC-E04), Person(RiC-E05), Group(RiC-E06), Family(RiC-E07), Corporate Body(RiC-E08), Position(RiC-E09), Delegate Agent(RiC-E10), Function(RiC-E11), Activity(RiC-E12), Process(RiC-E13), Action(RiC-E14), Rule(RiC-E15) 구성을 도식화하면

다음 <그림 12>와 같다. 주요 개체 중에서도 최상위 개체 Record Resource(RiC-E00), Agent(RiC-E04), Activity(RiC-E12)는 각각의 하위 개체를 연결하고, 연결된 개체는 최상위 개체를 통해 공통 속성을 상속받을 수 있다.

보조 개체의 구성을 도식화 한 내용은 <그림 13>과 같다. 보조 개체는 개체 각각의 구분과 식별을 위해 부여할 수 있는 요소인 Name(RiC-E16), Title(RiC-E17), Term(RiC-E18), Class Code(RiC-E19), Global Persistent Identifier(RiC-E20), Local Identifier(RiC-E21), 분류 요소인 Occupation(RiC-E22), Activity Class(RiC-E23), Documentary Form(RiC-E24), Record Class(RiC-25), 날짜와 장소를 표현하는 요소인 Date(RiC-E26), Single Date(RiC-E27), Data Range(RiC-E28), Date Set(RiC-E29), Place(RiC-E30), 주요 개체 전체를 아우



<그림 12> RiC-CM 주요 개체 도식화



〈그림 13〉 RiC-CM 보조 개체 도식화

르는 개념과 사물에 대해 표현할 수 있는 요소인 Concept/Thing(RiC-E31)으로 구성된다. 보조 개체는 주요 개체 간 관계를 부여·식별·구체화·구분·분류하기 때문에 보조 개체 사이에는 직접적인 연결 구조가 없다.

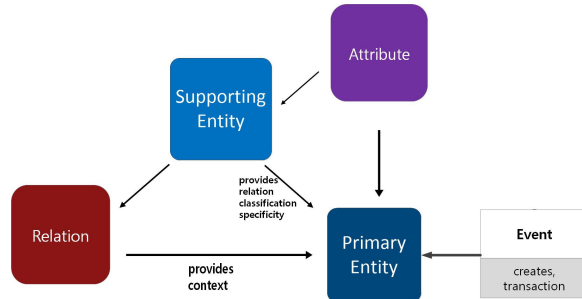
이러한 RiC 개체, 속성, 관계의 구조를 종합적으로 설명한다면 〈그림 14〉와 같이 표현할 수 있다. 주요 개체를 중심으로 보조 개체와 속성, 관계가 구조화되고 서로 연결된다. 속성은 주요 개체와 보조 개체에 각각 부여되고, 보조 개체는 주요 개체를 기술하며, 관계는 주요 개체의 맥락을 형성한다.

〈그림 14〉에서 추가한 'Event'는 데이터의 생성(creates)이나 데이터와 관련된 트랜잭션(transaction) 같은 '기록관리 업무 행위'로 볼

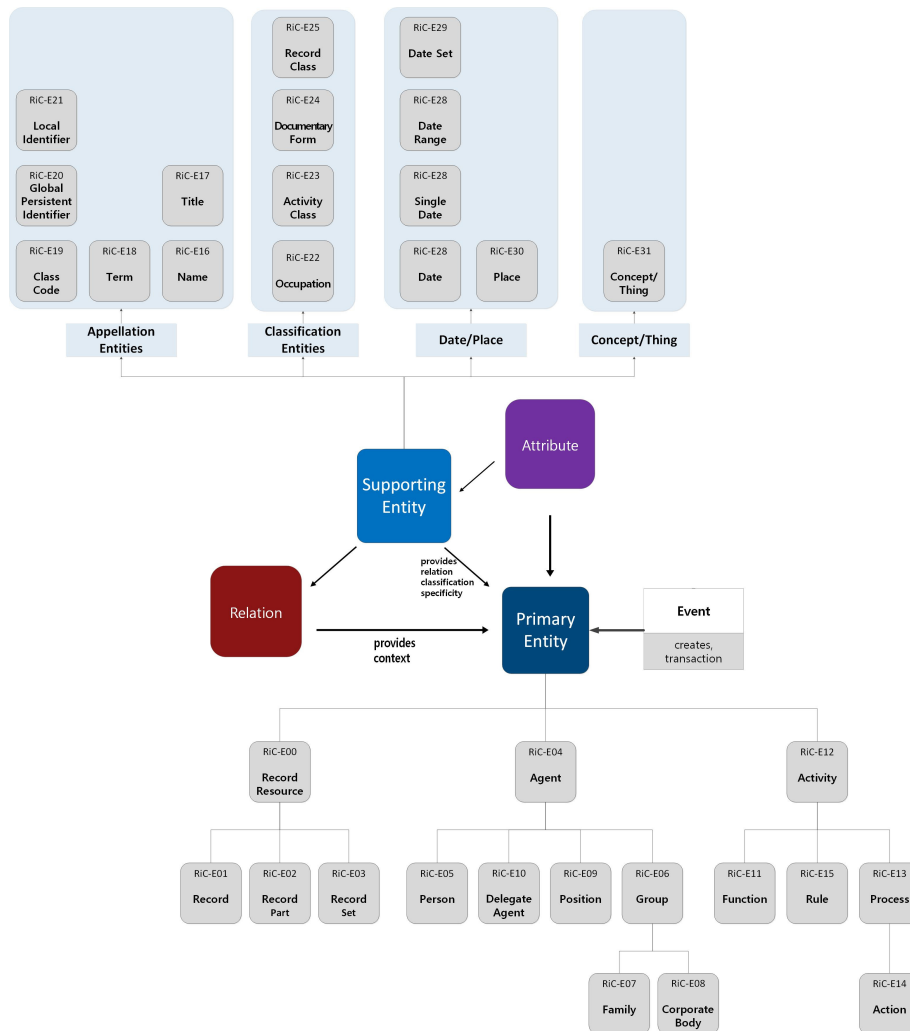
수 있으며 주요 개체와 관련된다.

앞서 도식화한 RiC 주요 개체와 보조 개체를 〈그림 14〉의 구조에 더해보면 RiC의 개념 모델을 구성하는 개체 전체를 시각적으로 파악할 수 있다(〈그림 15〉 참조). 〈그림 15〉는 RiC 개념 모델을 ERD(개체-관계 다이어그램, Entity-Relationship Diagram)와 유사한 다이어그램 형태로 표현한 것이다. ERD는 개체-관계를 도식화하여 데이터 구조를 시각화하는 모델링 방법인데, 주로 관계형 데이터 모델을 작성하는 데 사용한다. 본 논문에서는 ERD를 그대로 사용하기보다는 개체-관계 구조를 간선으로 간략히 도식화함으로써, RiC 데이터 모델링에 사용할 것이다.

이렇게 도식화할 수 있는 RiC의 요소들이 실제로 어떻게 기록정보와 매핑될 수 있는지

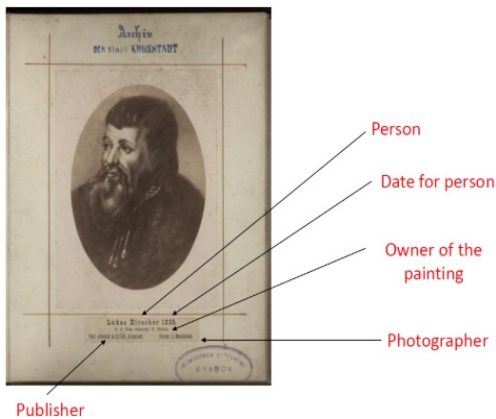


〈그림 14〉 RiC의 개체, 속성, 관계의 개괄적인 구조



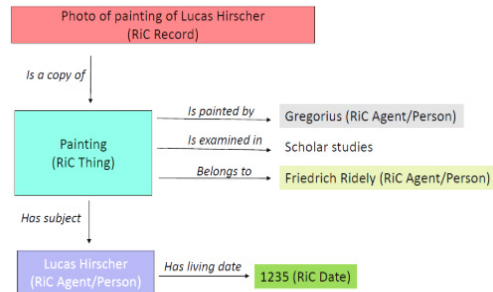
〈그림 15〉 RiC의 주요 개체와 보조 개체 구성

예시를 통해 살펴보자. 예시로 제시할 자료는 ICA EGAD가 2017년 발표자료에서 소개한 “Gregorius Painting(원본)을 사진촬영 한 Photo 사본: Photo of painting of Lucas Hirscher” 이다(〈그림 16〉 참조).



〈그림 16〉 Gregorius Painting(원본)을 사진촬영 한 사본 “Photo of painting of Lucas Hirscher” 출처: EGAD(2017)

〈그림 16〉의 “Photo of painting of Lucas Hirscher”의 기록정보를 RiC 요소에 적용해 보면 다음과 같다(〈그림 17〉 참조). “Photo of painting of Lucas Hirscher”는 한 개의 기록건으로 ‘RiC Record(주요개체:RiC-E01)’에 대응되고, 기록과 관련된 개체 간 맥락을 RiC 관계로 연결할 수 있다. 그 내용은 ‘작품(원작)과 사본의 관계(Is a copy of), 작품과 생산자(Is painted by)의 관계, 작품과 소유자(Belongs to)의 관계, 작품과 그 작품의 주제(Has subject, 여기에서는 Lucas Hirscher라는 인물)에 대한 관계, 작품의 주제(Lucas Hirscher)와 주제의 탄생일(Has living date)의 관계 등이 있다.



〈그림 17〉 Photo of painting of Lucas Hirscher(item)정보의 RiC 기반 맥락 출처: EGAD(2017)

관계는 업무의 흐름을 나타낸다는 점에서 중요하다. 따라서 관계를 모델링하는 것은 업무 분석 과정이라 할 수 있다. 〈그림 17〉에서 “Is a copy of”는 ‘개체(원본)는 또 다른 개체의 사본이다’라는 뜻인데, 이렇게 각 개체를 원본과 사본의 관계로 연결하려면 원본이 존재하고 그 사본을 수집·관리하는 등의 업무를 한다는 점을 알고 있어야 한다. “Is painted by”와 “Belongs to”는 작품 개체와 생산자 개체, 작품 개체와 소유자 개체를 연결하는 것인데, 누가 생산한 것이며 누가 소유하였는가 등의 업무와 관련성이 있다.

이와 같은 업무 분석을 통해 개념 모델링 단계에서 개체와 관계를 분석하고 RiC 요소와 매핑할 수 있다. 그 다음 개념 데이터 모델을 논리 데이터 모델로 전환한다. 이때 관계형 데이터 모델링에서는 2차원 표의 저장구조를 염두에 두고 무결성 제약조건(integrity constraints)에 따라 세부 속성을 도출하여 연결한다.

그래프형 데이터 모델링은 〈그림 17〉의 ‘생산자(Is painted by)’에 관련된 작품 개체와 제작자 개체는 정점으로, ‘생산되었다’라는 관계

는 간선으로 표현한다. 만일, 트리플 저장소 모델을 채택한 PIAAF 프로젝트였다면 해당 기록정보는 '(작품개체, 제작자개체, 생산되었다 관계)'로 개념 모델링 되고 '(Photo of painting of Lucas Hirscher의 원작, 화가 Gregorius, 그려졌다)'의 트리플 형식으로 저장되었을 것이다. 또 속성 그래프 모델이었다면 "(Photo of painting of Lucas Hirscher) [:IS_PAINTED_BY] -> (Gregorius)"로 표현될 것이다.

4. RiC 데이터 모델링

4.1 개념적 데이터 모델링

데이터 모델링은 현실 세계의 주요 데이터를 추출하여 개념 세계로 옮기는 개념적 모델링과 개념 모델을 데이터베이스에 저장하는 구조로 표현하는 논리적 모델링이 있다(김연희, 2013). 개념 모델링 단계에서는 관계형 모델링이나 그래프형 모델링 모두 동일한 방식으로 데이터를 분석할 수 있고, 관계형 또는 그래프형 중 어떤 모델을 선택하느냐는 논리적 모델링 단계에서 데이터를 사용하는 목적과 보유한(또는 보유할) 데이터베이스의 종류에 따라 결정된다.

본 장에서는 동일한 샘플 데이터를 사용하여 두 데이터 모델을 모두 만들어보면서 모델링

과정의 공통점과 차이점을 확인할 것이다. 다음 <표 9>는 데이터 모델링을 위한 샘플데이터이다. 샘플데이터의 기록정보는 임의로 선정한 클래식 음악 CD와 DVD 메타정보를 활용하였으며, 각 정보는 각각의 발행사(Opus Arte, UNIVERSAL MUSIC COMPANY/DG) 홈페이지에서 참조한 기본 메타정보와 검색 포털 사이트를 통한 관련 인물, 관련 단체 등의 추가 정보를 정리해 활용하였다.

데이터 모델링을 하기 전 가장 먼저 해야할 것은 어떤 기록에서 어떤 정보를 관리할지를 파악하는 것이다. 이 과정에서 최소 기록물 단위를 무엇으로 할지 정해야 한다. <표 9>의 샘플데이터에서 최소 기록물 단위는 음악CD와 영상DVD를 포괄하는 매체로 '앨범(album)'이라 할 수 있다.

이 앨범을 중심으로 기록정보를 관리하기 위해 관련된 주요 정보를 정리해보면 <표 10>과 같다. 앨범 (1)CD와 (2)DVD가 각각 '베토벤 CD컬렉션', '클래식음악 DVD컬렉션'에 속한다고 가정하였다. 컬렉션 정보에서는 컬렉션의 명칭, 설명정보를 앨범 정보에서는 그 앨범이 속한 소속컬렉션, CD/DVD명, 제작사, 발행사, 출연자, 연출자, 지휘자, 설명정보를 관리한다.

그 다음, <표 10>과 같이 정리한 정보를 RiC 개념 모델에서 정의한 개체와 속성에 매핑하는

<표 9> RiC 데이터 모델링을 위한 샘플데이터

	(1) John Eliot Gardiner 베토벤: 교향곡 전집 (Beethoven: The Symphonies)		(2) BBC Eroica, The Day that changed music forever
---	---	---	--

〈표 10〉 RiC 데이터 모델링을 위한 샘플데이터 기록정보 정리

정보	세부정보	값(value, 예시)
컬렉션	컬렉션명	베토벤 CD컬렉션, 클래식음악 DVD컬렉션
	설명정보	베토벤을 주제로 한 CD컬렉션, 클래식음악을 주제로 한 DVD컬렉션
앨범	소속 컬렉션	베토벤 CD컬렉션, 베토벤 DVD컬렉션
	CD/DVD명	John Eliot Gardiner 베토벤:교향곡전집, Eroica, The day that changed music forever
	제작사	ARCHIV, BBC
	발행사	Opus Arte, UNIVERSAL MUSIC COMPANY/DG
	출연자	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson (tenor), Gilles Cachemaille(bass)
	연출자	Simon Cellan Jones
	지휘자	Sir John Eliot Gardiner
	설명정보	Eroica, The day that changed music forever는 베토벤 영웅교향곡 그 이전의 흥미진진한 이야기들을 다룬다. BBC가 제작한 이 음악드라마는 1804년 6월 베토벤의 든든한 후원자였던 로브코비츠 공작의 저택에서 있었던 영웅교향곡의 비공식 초연을.....

단계이다. 가장 기본이 되는 앨범은 Record(RiC-E01)가 되고 컬렉션은 앨범의 집합이므로 Record Set(RiC-E03)이라 할 수 있다. 그리고 제작사·발행사는 업체정보이므로 Agent(RiC-E04), 출연자·연출자·지휘자는 인물정보이므로 Person(RiC-E05)이다. 다음으로 속성이 되는 정보를 찾는다면, 명칭·CD/DVD명은 Name(RiC-E16), 발행일은 Date(RiC-E26), 주요내용은 General note(RiC-P04)가 될 것이다. 여기서 Name과 Date는 RiC 보조 개체에 해당하는 것으로 주요 개체를 기술(description)한다는 면에서 속성에 해당하기도 한다.

여기서 Agent와 Person은 Record(RiC-E01)에 대한 세부정보가 아닌, 개체로서 별도의 전거정보(authority record)로 관리할 수 있다. 그렇다면 이에 대한 세부정보는 각각의 이름만 관리할 것인지 아니면 이름 이외에 (업체에 대한) 설명정보와 업종 같은 추가 정보도 관리할 것인지 고려할 수 있다. 본 논문에서는 Agent의

회사명과 설립연도, Person의 성명, 출생일, 직위, (인물에 대한) 설명정보를 관리한다고 가정하였다. 그렇다면 〈표 10〉에서 〈표 11〉의 Agent와 Person의 정보가 추가될 것이다.

그런데 인물을 분석하는 과정에서 직위 정보는 RiC의 Position(RiC-E09) 개체에 해당된다는 점을 발견하였다. 본 분석과정에서는 직위 개체를 활용하면서 직위의 세부정보로는 직위명과 설명정보가 있다고 가정하였다. 그렇다면 〈표 11〉에서 〈표 12〉의 Position 정보가 추가될 것이다.

지금까지 분석한 내용을 모두 종합하여 각각의 기록정보와 RiC 요소가 대응하는 것을 정리해보면 〈표 13〉과 같다.

다음 단계로 RiC 개체와 속성을 매핑한 결과를 토대로 관계를 분석해 볼 것이다.¹²⁾ 지금까지는 관계형 모델링이나 그래프형 모델링 모두 동일한 방법으로 기록정보를 분석했으나, 관계를 분석하는 방법은 차이가 있다.

〈표 11〉 샘플데이터 기록정보에서 RiC Agent와 Person 개체에 해당하는 정보

정보	세부정보	값(value, 예시)
업체	회사명	ARCHIV, Opus Arte
	설립연도	1947(ARCHIV), 1999(Opus Arte)
인물	성명	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Gilles Cachemaille(bass), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson(tenor)
	출생일	1964-10-08(Ian Hart), 1973-03-01(Jack Davenport), 1952-02-14(Anton Lesser)
	직위	지휘자, 출연자, 연주자
	설명정보	이안 하트(Ian Hart)는 2001년 해리포터와 마법사의 돌에서 Quirinus Quirrell을 연기했으며...

〈표 12〉 샘플데이터 기록정보에서 RiC Position 개체에 해당하는 정보

정보	세부정보	값(value, 예시)
직위	직위명	지휘자, 출연자, 연주자
	설명정보	"Eroica, The day that changed music forever"의 OST를 연주한 음악감독

〈표 13〉 RiC 데이터 모델링을 위한 개체와 속성 매핑

정보	정보-RiC개체 대응	세부정보	RiC개체 ~에 대응	RiC속성 ~에 해당
컬렉션	Record Set (RiC-E03)	컬렉션명		Name(RiC-E16)
		설명정보		Date(RiC-E26)
앨범	Record (RiC-E01)	소속컬렉션	Record Set(RiC-E03)	
		CD/DVD명		Name(RiC-E16)
		제작사	Agent(RiC-E04)	
		발행사	Agent(RiC-E04)	
		출연자	Person(RiC-E05)	
		연출자	Person(RiC-E05)	
		지휘자	Person(RiC-E05)	
	설명정보		General note(RiC-P04)	
업체	Agent (RiC-E04)	회사명		Name(RiC-E16)
		설립연도		Date(RiC-E26)
인물	Person (RiC-E05)	성명		Name(RiC-E16)
		출생일		Date(RiC-E26)
		직위	Position(RiC-E09)	
		설명정보		General note(RiC-P04)
직위	Position (RiC-E09)	직위명		Name(RiC-E16)
		설명정보		General note(RiC-P04)

12) 본 논문의 데이터 모델링에서 관계 분석은 RiC 아키텍처에서 제시한 관계를 (문자) 그대로 대입하지 않았다. 이는 데이터 모델링에서 관계를 파악하는 것이 데이터 자체의 특성뿐만 아니라 관련된 업무, 용어, 규정, 상황 등 여러 요인에 의해 영향을 받기 때문이다. 따라서 관계를 설정할 때는 기록정보(샘플데이터)에 '적합한 표현'을 사용하고, RiC 관계요소는 RiC의 개체와 속성을 파악하고 분석하는 도구로서 참고하였다.

* RiC Relation: 관계는 두 개의 RiC-CM 개체를 연결한다. 기록의 의미를 연결하여, 기록과 관련된 특정 기록집합(Record Set)의 역사와 내용을 이해하게 하고 서로 상호 보완적인 기록을 발견하게 하는데 활용된다. [cited 2018.12.31]. <https://web.esrc.unimelb.edu.au/ICAD/biogs/E000006b.htm>

관계형 데이터 모델링에서 관계는 속성을 기준으로 맺어진다. <표 13>에서 앨범과 업체는 어떤 관계가 있을까? 앨범의 제작사와 발행사가 업체이다. 그런데, 한 앨범의 제작사가 여러 개일 수 있고, 한 제작사가 여러 앨범을 제작할 수 있다. 한 앨범이 한 개의 데이터(행)이라고 할 때, 앨범 정보에 만일 제작사 정보가 포함되어 있다면 앨범의 제작사는 하나 밖에 표현할 수 없다. 따라서 앨범과 업체 사이에 이 둘을 연결해주는 별도의 정보(e.g. “앨범과 업체의 관계”)가 필요하게 된다. 그리고 앨범에서는 제작사와 발행사 정보가 더 이상 필요 없기 때문에 삭제한다(<표 14> 참조).

이와 마찬가지로 출연자와 연출자, 지휘자(인물)도 한 앨범에 여러 명일 수 있으므로 “앨범과 인물과의 관계”가 필요하고, 출연자와 지휘자라는 직위도 한 사람에게 여러 지위가 부

여될 수 있으므로 “인물과 지위의 관계” 정보가 추가로 필요하다. 그런데, 어떤 인물이 A 앨범에서는 지휘를 B 앨범에서는 연출을 맡을 수 있으므로 “앨범과 직위의 관계”가 필요할 것이고, 그 정보는 “앨범과 인물의 관계”와 합쳐져서 “앨범, 인물, 직위의 관계” 정보가 될 것이다. 이를 정리하면 <표 15>와 같다.

<그림 18>은 <표 15>의 내용을 관계형 데이터 모델의 개념적 설계를 위해 다이어그램으로 나타낸 것이다.

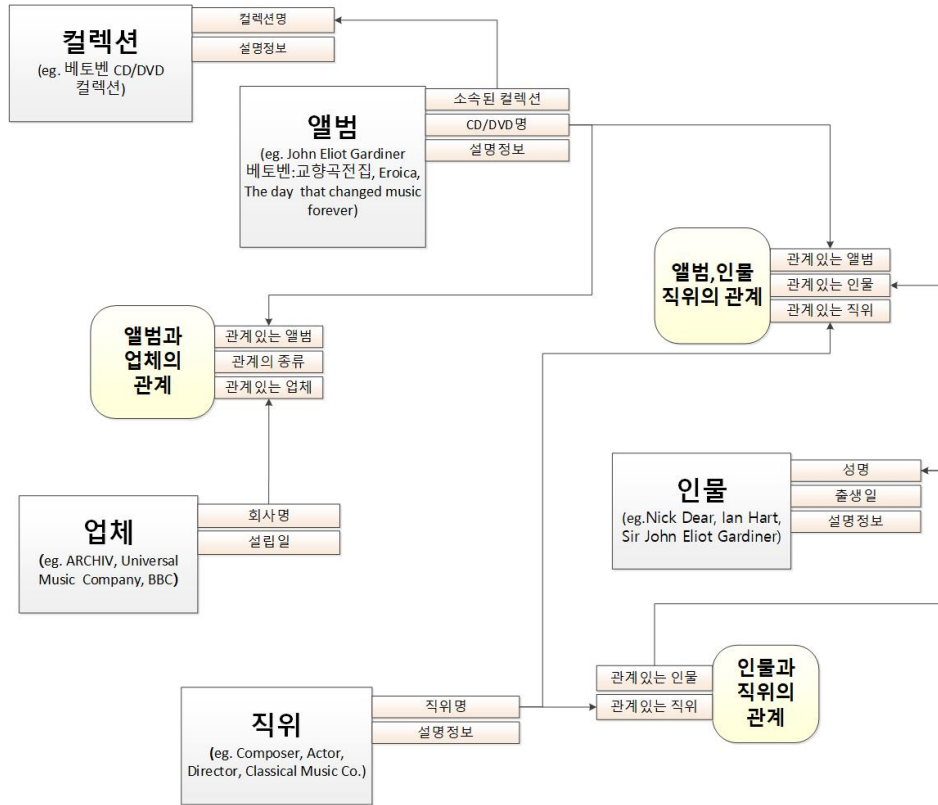
그래프형 데이터 모델링에서는 관계를 파악하기 위해 개체 간 연관성이 있는 말을 ‘동사’ 형태로 찾아낸다. 즉 ‘개체 A는 개체 B와 (C라는 관계이다)’라고 서술되는 형태이다. 이와 같은 방법으로 관계를 맺는 개체를 정리하면 <표 16>과 같다.

<표 14> RiC 관계형 데이터 모델링을 위한 ‘관계’ 분석(1)

정보	세부정보	값(value, 예시)
앨범	소속 컬렉션	베토벤 CD컬렉션, 베토벤 DVD컬렉션
	CD/DVD명	John Eliot Gardiner 베토벤: 교향곡전집, Eroica, The day that changed music forever
	제작사	ARCHIV, BBC
	발행사	Opus Arte, UNIVERSAL MUSIC COMPANY
	출연자	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson(tenor), Gilles Cachemaille(bass)
	연출자	Simon Cellan Jones
	지휘자	Sir John Eliot Gardiner
	설명정보	Eroica, The day that changed music forever는 베토벤 영웅교향곡 그 이전의 흥미진진한 이야기들을 다룬다. BBC가 제작한 이 음악드라마는 1804년 6월 베토벤의 든든한 후원자였던 로브코비츠 공작의 저택에서 있었던 영웅교향곡의 비공식 초연을.....
앨범과 업체의 관계	관계있는 앨범	John Eliot Gardiner 베토벤: 교향곡전집, Eroica, The day that changed music forever
	관계있는 업체	ARCHIV, BBC(이상 제작사), Opus Arte, UNIVERSAL MUSIC COMPANY(이상 발행사)
	관계의 종류	제작사, 발행사
업체	회사명	ARCHIV, Opus Arte
	설립연도	1947(ARCHIV), 1999(Opus Arte)

〈표 15〉 RiC 관계형 데이터 모델링을 위한 ‘관계’ 분석(2)

정보	세부정보	값(value, 예시)
컬렉션	컬렉션명	베토벤 CD컬렉션, 클래식음악 DVD컬렉션
	설명정보	베토벤을 주제로 한 CD컬렉션, 클래식음악을 주제로 한 DVD컬렉션
앨범	소속 컬렉션	베토벤 CD컬렉션, 베토벤 DVD컬렉션
	CD/DVD명	John Eliot Gardiner 베토벤: 교향곡전집, Eroica, The day that changed music forever
	제작사	ARCHIV, BBC
	발행사	Opus Arte, UNIVERSAL MUSIC COMPANY
	출연자	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe-Johnson(tenor), Gilles Cachemaille(bass)
	연출자	Simon Cellan-Jones
	지휘자	Sir John Eliot Gardiner
설명정보	Eroica, The day that changed music forever는 베토벤 영웅교향곡 그 이면의 흥미진진한 이야기들을 다룬다. BBC가 제작한 이 음악드라마는 1804년 6월 베토벤의 든든한 후원자였던 로브코비츠 공작의 저택에서 있었던 영웅교향곡의 비공식 초연을.....	
앨범과 업체의 관계	관계있는 앨범	John Eliot Gardiner 베토벤: 교향곡전집, Eroica, The day that changed music forever
	관계있는 업체	ARCHIV, BBC(이상 제작사), Opus Arte, UNIVERSAL MUSIC COMPANY(이상 발행사)
	관계의 종류	제작사, 발행사
업체	회사명	ARCHIV, Opus Arte
	설립연도	1947(ARCHIV), 1999(Opus Arte)
앨범, 인물, 직 위의 관계	관계있는 앨범	John Eliot Gardiner 베토벤: 교향곡전집, Eroica, The day that changed music forever
	관계있는 인물	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson(tenor), Gilles Cachemaille(bass) (이상 출연자) Simon Cellan Jones (이상 연출자) Sir John Eliot Gardiner (이상 지휘자)
	관계있는 직위	출연자, 연출자, 지휘자
인물	성명	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson(tenor), Gilles Cachemaille(bass) Simon Cellan Jones Sir John Eliot Gardiner
	출생일	1964-10-08(Ian Hart), 1973-03-01(Jack Davenport), 1952-02-14(Anton Lesser)
	직위	지휘자, 출연자, 연주자
	설명정보	이안 하트(Ian Hart)는 2001년 헤리포터와 마법사의 돌에서 Quirinus Quirrell을 연기했으며...
인물과 직위의 관계	관계있는 인물	Ian Hart, Jack Davenport, Tim Pigott-Smith, Claire Skinner, Anton Lesser, Luba Orgonasova(soprano), Anne Sofie von Otter(mezzo-soprano), Anthony Rolfe Johnson(tenor), Gilles Cachemaille(bass) (이상 출연자) Simon Cellan Jones (이상 연출자) Sir John Eliot Gardiner (이상 지휘자)
	관계있는 직위	지휘자, 출연자, 연주자
직위	직위명	지휘자, 출연자, 연주자
	설명정보	"Eroica, The day that changed music forever"의 OST를 연주한 음악감독



<그림 18> RiC 관계형 데이터 모델링의 개념적 모델 예시

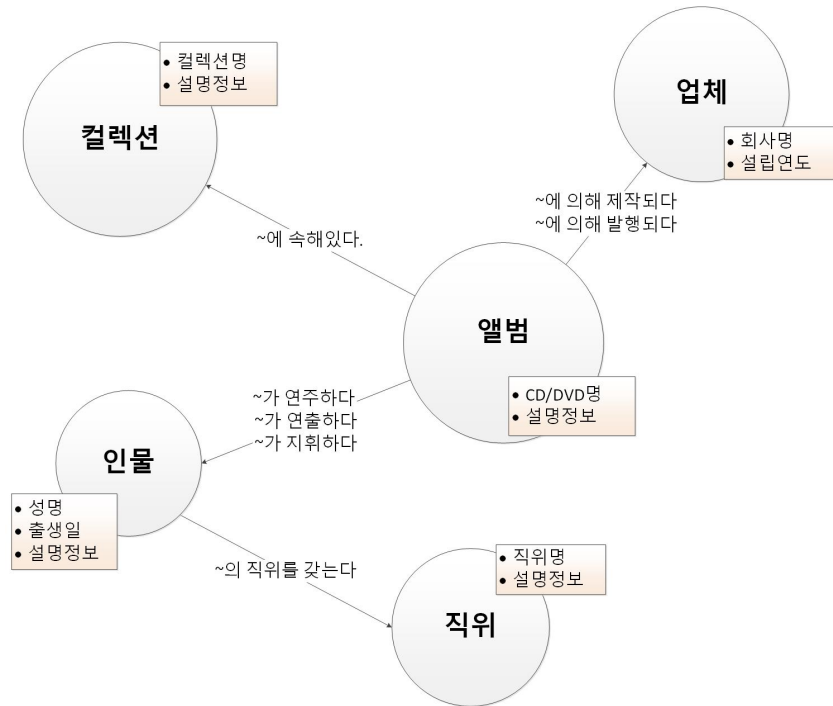
<표 16> RiC 그래프형 데이터 모델링을 위한 '관계' 분석

정보	개체 A	세부정보	개체 B	A와 B의 관계(C)
앨범	Record(RiC-E01)	소속 컬렉션	Record Set(RiC-E03)	A는 B에 속해있다.
		제작사	Agent(RiC-E04)	A는 B에 의해 제작된다.
		발행사	Agent(RiC-E04)	A는 B에 의해 발행된다.
		출연자	Person(RiC-E05)	A는 B가 출연하다.
		연출자	Person(RiC-E05)	A는 B가 연출하다.
		지휘자	Person(RiC-E05)	A는 B가 지휘하다.
인물	Person(RiC-E05)	직위	Position(RiC-E09)	A는 B의 직위를 갖는다.

이를 바탕으로 그래프형 데이터 모델을 다이어그램으로 표현하면 <그림 19>와 같다.

개념적 데이터 모델은 현실 세계를 사람들의 머릿속에 그릴 수 있는 개념적인 구조로 모델링

하는 데 사용하므로 데이터베이스 종류에 구애 받지 않고 작성된다. 데이터베이스의 종류에 따라 발생하는 차이는 데이터베이스(또는 데이터 모델)의 논리적 모델을 염두에 둔 관계를 설정



〈그림 19〉 RiC 그래프형 데이터 모델링의 개념적 모델 예시

하는 방법에서 시작된다. 앞서 살펴 본 것처럼, 관계형 데이터 모델링에서는 개체의 속성을 연결하고, 그래프형 데이터 모델링에서는 개체를 직접 연결하면서 관계를 설명한다.

이 결과 관계형 데이터 모델을 사용하는 RDBMS는 엄격한 무결성을 담보하고 데이터를 안전하게 관리할 수 있어 정확성과 일관성이 요구되는 영역에 활용하기 적합하다. 반면, 너무 엄격한 무결성은 확장가능성을 제약하기 때문에, 데이터 연계와 확장 측면에서는 상대적으로 비효율적이다. 한편, 그래프형 데이터 모델은 개체가 데이터의 유형과 관계없이 관계를 통해 설명되기 때문에 데이터의 구조를 개별적으로 보여주기에도 적합하고 시각화에 유용하다. 반면 이러한 데이터 유형에 대한 유연성 때문

에 엄밀하고 종합적인 데이터 분석이 어려워 집계나 평균 등의 수학적 연산에는 효과적이지 않다.

4.2 논리적 데이터 모델링

개념적 데이터 모델링을 마치면 논리적 데이터 모델링을 할 차례다. 논리적 데이터 모델은 개념적 데이터 모델을 사용할 DBMS에 따라 데이터베이스에 저장할 형태로 표현한 데이터베이스의 논리적인 구조다(김연희, 2013). 이때 관계형 모델과 그래프형 모델은 개체와 관계를 설정하는 방법이 서로 다르기 때문에, 논리적 데이터 모델도 다르게 구현된다.

먼저 관계형 데이터 모델에 대하여 논리적

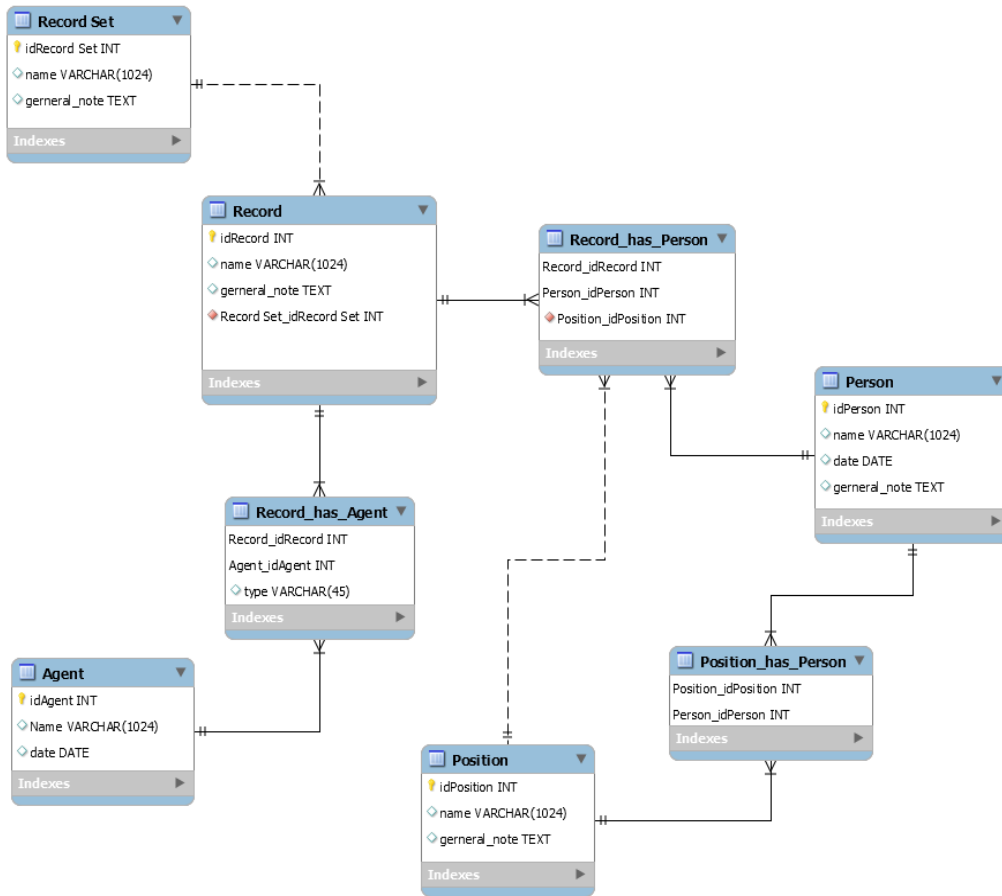
모델링을 해보도록 한다. 관계형 데이터 모델은 데이터베이스의 논리적 구조가 2차원 테이블 형태이다. 4.1의 관계형 데이터 모델 다이어그램(〈그림 18〉 참조)을 살펴보면, 관계가 이름(앨범의 CD/DVD명, 인물의 성명, 직위의 직위명, 업체의 회사명)으로 연결되어 있다. 이렇게 관계를 설정할 경우 중복될 가능성이 있으므로 고유식별자(기본키)가 필요하다. 이에 따라 각 정보(컬렉션, 앨범, 업체, 인물, 직위)마다 고유식별자를 설정하고(〈표 17〉의 'ID(고유식별자)'), 관계를 맺는 상대편 속성에 그 고

유식별자를 참조하도록 한다(〈표 17〉의 '참조되는 고유식별자').

이 외에도 데이터베이스 테이블을 구성하기 위해 속성의 데이터형(e.g. 가변형 문자열 varchar, 정수형 int)이나 관계차수(mapping cardinality, 일대일·일대다 등)를 추가하여 논리적 데이터 모델링을 수행한다. 이를 반영하여 MySQL Workbench의 도구로 작성한 것이 〈그림 20〉이다. 관계형 데이터베이스 논리적 설계 단계에서는 완전한 논리적 스키마를 구성하기 위해 테이블(릴레이션) 명세서를 작성할 수 있는데,

〈표 17〉 RiC 관계형 데이터 모델링의 2차원 테이블 구조 저장을 위한 ID 설정

정보	RiC개체	세부정보	RiC개체와 속성	관계형 모델의 속성	참조되는 고유식별자
컬렉션	Record Set (RiC-E03)	ID(고유식별자)		idRecordSet	
		컬렉션명	Name (RiC-E16)		
		설명정보	General note (RiC-P04)		
앨범	Record (RiC-E01)	ID(고유식별자)		idRecord	
		소속 컬렉션			idRecordSet
		CD/DVD명	Name (RiC-E16)		
		설명정보	General note (RiC-P04)		
앨범과 업체 관계	-	관계있는 앨범			idRecord
		관계있는 업체			idAgent
		관계의 종류		type	
업체	Agent (RiC-E04)	ID(고유식별자)		idAgent	
		회사명	Name (RiC-E16)		
		설립연도	Date (RiC-E26)		
앨범, 인물, 직위의 관계	-	관계있는 앨범			idRecord
		관계있는 인물			idPerson
		관계있는 직위			idPosition
인물	Person (RiC-E05)	ID(고유식별자)		idPerson	
		성명	Name (RiC-E16)		
		출생일	Date (RiC-E26)		
		설명정보	General note (RiC-P04)		
인물과 직위의 관계	-	관계있는 인물			idPerson
		관계있는 직위			idPosition
직위	Position (RiC-E09)	ID(고유식별자)		idPosition	
		직위명			
		설명정보	General note (RiC-P04)		



〈그림 20〉 RiC 관계형 데이터 모델링의 논리적 모델 예시

이때 속성이름, 데이터형, 널(NULL) 허용여부, PK, FK, 제약조건 등의 세부 설계 내역을 정한다(김연희, 2013). 논리적 설계를 MySQL Workbench로 구현하는 과정은 이러한 테이블 명세서의 내용을 반영하고, 물리적 설계를 위한 SQL 코드를 자동 작성하여 제공한다.

다음으로 그래프형 데이터 모델에 대하여 논리적 모델링을 해보도록 한다. 앞서 설명한대로, 그래프형 데이터 모델에서는 개체를 정점으로, 관계를 간선으로 표현한다. 이 때 두 개체를 시작 정점(Start Node)과 끝 정점(End Node)으

로 지정하여 인과관계를 명확하게 한다. 4.1의 그래프형 데이터 모델의 개체의 관계를 분석한 〈표 16〉에서 개체를 시작 정점과 끝 정점으로 정하고 그 인과관계를 정리한 것이 〈표 18〉이다. 예를 들어, 〈표 16〉의 개체 A(Record)와 개체 B(Agent)의 관계를 'A는 B에 의해 발행된다.'로 분석한 것은 개체 A(e.g. Eroica, The Day that changed music forever)를 시작정점으로 하고 개체 B(e.g. Opus Arte)를 끝 정점으로 하여 관계를 PUBLISHED_BY로 나타낸다.

〈표 18〉 RiC 그래프형 데이터 모델링을 위한 샘플데이터 시작 정점-끝 정점 설정

시작 정점	관계	끝 정점	임의의 예시 값
Record	PUBLISHED_BY	Agent	Eroica, The day that changed music forever PUBLISHED_BY Opus Arte
Record	PRODUCED_BY	Agent	Eroica, The day that changed music forever PRODUCED_BY BBC
Record	ASSOCIATED_WITH	Agent	Eroica, The day that changed music forever ASSOCIATED_WITH Sir John Eliot Gardiner
Record	IS_PART_OF	Record Set	Eroica, The day that changed music forever IS_PART_OF 클래식음악 DVD컬렉션
Agent	IS_OCCUPIED_BY	Position	Opus Arte IS_OCCUPIED_BY Classical Music Inc.
Position	IS_OCCUPIED_BY	Person	Composer IS_OCCUPIED_BY Sir John Eliot Gardiner
Record	IS_ABOUT_TO	Thing	Eroica, The day that changed music forever IS_ABOUT_TO Beethoven: Symphony No. 3 'Eroica'

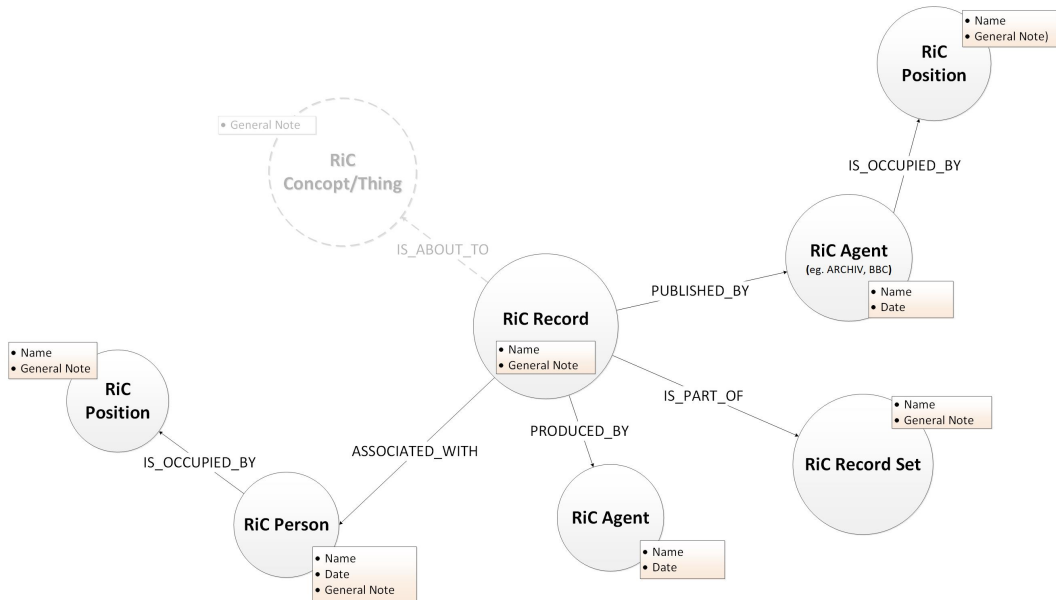
〈표 18〉의 시작 정점-끝 정점의 관계를 정리한 것을 그래프형 다이어그램으로 작성해 〈그림 21〉과 같이 논리적 모델을 만들 수 있다.¹³⁾

〈그림 21〉과 같은 논리 모델을 그래프형 데이터 모델링 도구인 Neo4j로 구현한다면, 샘플 데이터는 〈그림 22〉와 같이 도출될 것이다. 논리적 데이터 모델인 〈그림 21〉과 실제 그래프형 데이터 결과물이 매우 비슷한데, 이는 그래프 데이터 모델이 현실 세계를 직관적으로 표현하기 때문이다. 그렇기 때문에 데이터 모델링이 쉽고 빠르다. 그래프형 데이터 모델링은 개념 모델 수준에서 논리 모델로 전환할 때 관계형 데이터 모델링과 같은 엄격한 데이터 구조(즉, 무결성 제약조건인 데이터베이스 스키

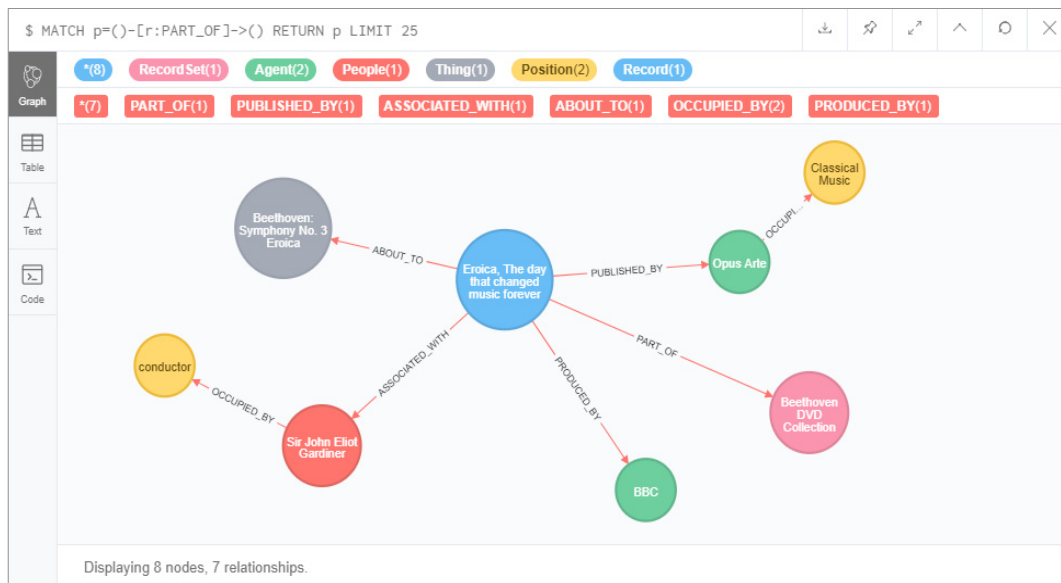
마)를 요구하지 않으며, ‘객체 + 속성’을 ‘관계’를 통해 연결하는 구조이기 때문에 그 자체로 객체 지향적이다.

이와 같이 본 연구에서 수행한 RiC 데이터 모델링 과정은 기록정보를 RiC 개체, 속성과 관계로 분석하여 개념적 구조로 표현하고 이를 사용할 목적과 DBMS에 따라 데이터베이스 논리 모델로 구현하는 절차를 수행한다. 관계형 데이터 모델링은 RiC 요소가 포괄하는 기록관리 핵심 정보를 안정적으로 관리하고 분석하는 정형화된 데이터 처리에 뛰어나며, 그래프형 데이터 모델링은 다양한 기록 유형을 RiC에 적용하기 쉽고 데이터 연계와 확장 측면에서 유용하다.

13) 〈그림 21〉에서 점선으로 표시한 개체 RiC-Concept/Thing과 관계 IS_ABOUT_TO는 〈표 18〉의 내용에는 없지만, 예를 들어 Record가 어떤 주제나 대상에 대한 것임을 표현하고자 할 때 점선과 같이 개체와 관계를 쉽게 추가할 수 있음을 설명하고자 표시하였다.



<그림 21> RiC 그래프형 데이터 모델링의 논리적 모델 예시



<그림 22> 그래프형 데이터 논리 모델에 샘플데이터를 적용해 Neo4j로 구현한 결과

이밖에도 문서 저장소(document object 모델)¹⁴⁾ 키-값 저장소(Key-value store) 모델¹⁵⁾

등 다양한 데이터 모델이 존재한다. 이는 어느 한 기술이 다른 기술을 대체하는 것이 아니라

서로 공존하며 상호보완적인 조화를 이뤄야 함을 시사한다고 볼 수 있다. 앞으로는 다양한 데이터 유형을 포괄하기 위해 그 용도에 맞는 데이터 저장소를 사용하고 각각을 연계하는 다중 데이터베이스 모델을 활용할 수도 있을 것이다.

5. 맺음말

지금까지 RiC의 실체는 무엇이고 실무에 어떻게 적용할 수 있을지에 대한 방안을 찾기 위해 RiC을 분석하고 RiC 개념 모델을 데이터 모델링의 도구로서 다뤄보았다. RiC 개념 모델은 기술 요소가 개체, 속성, 관계로 정의된 다차원적 데이터 참조 모델이다. RiC을 데이터 모델링하기 위해서는 기록정보를 RiC 개체, 속성과 관계로 분석하여 개념적 구조로 표현하고 이를 사용할 목적과 DBMS에 따라 데이터베이스 논리 모델로 구현하는 절차를 수행한다. 이러한 과정을 통해 본 연구는 기록관리시스템 설계 과정에서 핵심적인 부분이라고 할 수 있는 데이터 모델링을 강조하고 나아가 데이터 모델링의 핵심 요소로서 RiC을 적용하는 방안을 모색하고자 하였다. 그러나 본 연구에서 수행한 데이터 모델링은 간략한 샘플 데이터를 대상으로 구현했기 때문에 기록관리에 필수적인 주요 속

성을 모두 제시하지 못한 한계가 있다. RiC을 실제 기관의 아카이브 시스템에 적용하는 데이터 모델링을 위해서는 데이터와 관련한 기록관리 업무를 분석하고 그 흐름을 개체 간 관계 모델링에 포함해야 할 것이다.

아직까지 RiC은 초안(draft version) 단계이기 때문에 전 세계적으로도 실무 적용에 이르지 못했다. 하지만 이미 RiC은 많은 국가의 기록공동체들이 주목하고 있는 기술표준이며 앞으로 국제 표준으로 제시될 것이기 때문에 국내 논의와 연구는 계속되어야만 한다. 국내 레거시 시스템이 대다수 종이기록을 분류하고 기술하던 방식을 유지하고 있으며 기록관리 절차 중심으로 구성되어 있기 때문에, 향후 RiC과 같은 데이터 중심 기술표준을 도입하기 위해서는 파일럿 프로젝트와 같은 시도가 계속 필요하다. 이미 명세가 공개된 RiC과 유사한 많은 데이터 중심 기술표준, 메타데이터 표준이 존재한다. RiC의 명세가 모두 완성되기를 기다리기보다, RiC에 대한 적극적인 연구와 참여를 통해 아카이브 시스템 패러다임을 바꾸려는 노력들이 지속되어야 한다. 이러한 의미에서 기록관리분야와 IT분야의 매개로서 본 연구의 의의를 찾을 수 있겠다. 기록전문가가 RiC을 도입하려면 IT 전문가와 소통하고 협업해야 하기 때문이다.

RiC은 분산되어 있는 기술요소를 통합하여

-
- 14) 문서 저장소란 데이터를 JSON, BSON 또는 XML 문서에 저장하는 비관계형 데이터베이스이다. 유연한 스키마가 특징. 사용자가 데이터를 입력하기 전에 테이블의 스키마를 선언해야 하는 SQL 데이터베이스와 달리 문서 구조를 강요하지 않는다. 문서에는 원하는 어떤 데이터도 넣을 수 있다(RDB부터 검색엔진까지 내게 꼭 맞는 DB 고르기). [cited 2018.12.31]. <http://www.ciokorea.com/news/38041>
 - 15) 키-값 저장소란 각 값이 특정 키와 연관된 일종의 비관계형 데이터베이스다. 연관 배열(associative array)이라고도 한다(RDB부터 검색엔진까지 내게 꼭 맞는 DB 고르기). [cited 2018.12.31]. <http://www.ciokorea.com/news/38041>

메타데이터를 충실히 생산·축적하고, RiC 데이터 모델링으로 사용자 요구사항에 맞는 다양한 아카이브 데이터베이스를 구성할 수 있다. 온톨로지 검색도구로서의 RiC은 서로 다른 기관 간 기록 정보를 연계하고 서비스할 수 있는 데이터 확장성을 보장한다. 이를 바탕으로 다양한 기술적 논의가 확장되어 아카이브가 전자 기록 시대에 맞는 이용 환경과 정보 서비스를 제공할 수 있기를 기대해본다. 레코드 컨티뉴엄(Record Continuum) 개념은 기록관리의 생

애주기(life cycle)라는 사고의 틀을 변화시켰다. RiC은 데이터의 확장과 상호운용성 면에서 탁월하기 때문에 레코드 컨티뉴엄 개념을 기록 관리체계에 도입하는 데 유용하다. 기록의 수집과 보존을 중심으로 하는 업무에서 기록의 가치와 활용에 초점을 두는 업무로 패러다임을 전환해야 하는 시점에서, RiC은 이를 실현할 수 있는 충분한 대안이 될 수 있으며 정보 공유를 위한 엔진으로서 기록의 의미를 확장하고 연결을 지속하게 하는 이정표가 될 것이다.

참 고 문 헌

- 국가기록원 영구기록물 기술규칙 [NAK 13:2011(v2.0)]. (행정안전부 고시 제2011-42호).
- 김상래 (2015). 프로젝트 성패를 결정짓는 데이터 모델링 이야기. 서울: 한빛미디어.
- 김익한 (2004). 기록의 속성과 메타데이터 표준을 통해 본 한국의 기록·기록기술. 기록학연구, 10, 3-26.
- 김치수 (2015). 쉽게 배우는 소프트웨어 공학. 서울: 한빛아카데미.
- 랜 실버스톤 (2018). 데이터 모델 리소스 북 vol.1(Revised Edition): 앞서가는 모델러와 모든 회사를 위한 유니버설 라이브러리. (김기창 역). 서울: 스포트라잇북.
- 마이클 j. 헤르난데즈 (2014). 가장 쉬운 데이터베이스 설계 책: 적절한 데이터베이스 디자인을 위한 지침서. (송현호, 황규용 공역). 서울: 비제이퍼블리(BJ퍼블릭).
- 마틴 클레프만 (2018). 데이터 중심 애플리케이션 설계: 신뢰할 수 있고 확장 가능하며 유지보수하기 쉬운 시스템을 지탱하는 핵심 아이디어. (정재부, 이도경 역). 파주: 위키북스.
- 마틴 파울러 (2015). 엔터프라이즈 애플리케이션 아키텍처 패턴: 엔터프라이즈 애플리케이션 구축을 위한 개체지향 설계의 원리와 기법. (최민석 역). 파주: 위키북스.
- 명지대학교 디지털아카이빙연구소 (2017a). 차세대 기록관리 모델 재설계연구 개발: 1권. http://www.archives.go.kr/next/news/viewPublicationList.do?bg_no=428.
- 명지대학교 디지털아카이빙연구소 (2017b). 차세대 기록관리 모델 재설계연구 개발: 2권. http://www.archives.go.kr/next/news/viewPublicationList.do?bg_no=429.
- 박지영 (2008). 문화유산 자원 통합 활용을 위한 CRM 기반 FRBR 응용 온톨로지 적용에 관한 연구. 한국비블리아학회지, 19(2), 45-62.
- 박지영 (2016). 차세대 기록물 기술표준에 관한 연구. 한국기록관리학회지, 16(1), 223-245.

- <http://dx.doi.org/10.14404/JKSARM.2016.16.1.223>
- 박지영 (2017a). ISAD(G)에서 RiC-CM으로의 전환에 관한 연구. 한국기록관리학회지, 17(1), 93-115.
<https://doi.org/10.14404/JKSARM.2017.17.1.093>
- 박지영 (2017b). RiC에 대한 기록공동체의 리뷰를 통해 본 기록물 기술표준 개선을 위한 제안. 기록학 연구, 54, 81-109.
- 박지영 (2017c). FRBRoo 분석을 통한 FRBR 개념모형의 확장과 개선. 정보관리학회지, 34(4), 201-225.
<https://doi.org/10.3743/KOSIM.2017.34.4.201>
- 백옥인 (2013). 디지털 데이터 · 정보 · 지식. 서울: 커뮤니케이션북스.
- 샘 뉴먼 (2017). 마이크로서비스 아키텍처 구축: 대용량 시스템의 효율적인 분산 설계 기법. (정성권 역). 서울: 한빛미디어.
- 오진관, 임진희 (2018). 차세대 기록관리시스템 재설계 모형 연구. 한국기록관리학회지, 18(2), 163-188.
<https://doi.org/10.14404/JKSARM.2018.18.2.163>
- 오쿠노 미키야 (2016). 관계형 데이터베이스 실전 입문. (성창규 역). 파주: 위키북스.
- 원종관 (2008). 레코드 컨티뉴엄의 속성을 통해 본 증거와 기억의 조화에 관한 연구. 석사학위논문. 한국외국어대학교 대학원, 정보 · 기록학과.
- 이유빈, 이해영 (2017). 온톨로지 기반의 기록물 검색 시스템을 위한 인터페이스 제안. 한국기록관리학회지, 17(1), 217-244. <https://doi.org/10.14404/JKSARM.2017.17.1.217>
- 이춘식 (2005). 데이터베이스 설계와 구축(성능까지 고려한 데이터 모델링). 서울: 한빛미디어.
- 이춘식 (2008). 아는 만큼 보이는 데이터베이스 설계와 구축. 서울: 한빛미디어.
- 이해영 (2013). 기록조직론: 한국국가기록연구원 교육총서-1. 서울: 선인.
- 이현실, 한성국 (2006). 기록 관리 메타데이터의 개념 모델링. 정보관리학회지, 23(3), 23-48.
<https://doi.org/10.3743/KOSIM.2006.23.3.023>
- 임진희, 이대욱, 김은실, 김익한 (2012). 영구기록물관리를 위한 기록물 데이터베이스 스키마 개발 방향. 기록학연구, 34, 57-105.
- 한국기록학회 (2014). 기록학용어사전. 서울: 역사비평사.
- 현문수 (2014). FRBRoo/CIDOC CRM 기반의 로컬리티 정보자원 구조화 연구. 한국비블리아학회지, 25(4), 265-290. <https://doi.org/10.14699/kbiblia.2014.25.4.265>
- 현문수, 설문원 (2018). 차세대 공공 전자기록의 조직 모형 개발을 위한 방향 탐구. 기록학연구, 56, 183-212.
- 황진현, 임진희 (2012). 시각예술기록정보 관리를 위한 데이터모델 설계. 기록학연구, 33, 155-206.
- ANAI & ICAR (2016). Records In Contexts: A conceptual model for archival description(draft v0.1, september 2016): Il contributo italiano.
http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n2.pdf

- Belinda Battley, Elizabeth Daniels & Gregory Rolan (2014). Archives as multifaceted narratives: linking the 'touchstones' of community memory. *Archives and Manuscripts*, 42(2), 155-157.
- DB-Engines Ranking of Graph DBMS, 31 systems in ranking, December 2018 Available at: <https://db-engines.com/en/ranking/graph+dbms> (last visited 2018-12-31).
- DOREMUS: DOing REusable MUSical Data = Données en REutilisation pour la Musique en fonction des Usage. Web site. Available at: www.doremus.org/ (last visited 2018-12-31).
- EGAD (2016). International Council on Archives. Expert Group on Archival Description. 2016. Record In Contexts: A Conceptual Model For Archival Description, Consultation Draft v.0.1.
- EGAD (2017). Presentation of the standard "Records in Context." ICA Expert Group on Archival Description (EGAD): The ALA-ICA annual Conference 2017 Mexico, 27 to 29 November 2017.
- EGAD (2018). Presentation of the standard "Records in Context" Records in Contexts (RiC): a standard for archival description developed by the ICA Experts Group on Archival Description. International Council on Archives, Paris. Retrieved 12 January 2018.
- Fredric. M. Miller (2002). 아카이브와 매튜스크립트의 정리와 기술. (한국국가기록연구원 감수/조경구 역). 서울: 진리탐구.
- Gretchen Gueguen, Vitor Manoel Marques da Fonseca, Daniel V. Pitti, and Claire Sibille-de Grimoüard (2013). Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description. *American Archivist* 76, no. 2 (Fall/Winter 2013): 567-584.
- Ian, Robinson, Jim, Webber, & Emil, Eifrem (2015). *Graph Databases(2nd Edition): New Opportunities for Connected Data*. O'Reilly Media.
- ICA EGAD (2018). Semantizing and visualising archival metadata: the PIAAF French prototype online. Florence Clavaud (Archives nationales de France | EGAD) | 4 May 2018.
- ICA Experts Group on Archival Description 2017 Annual Meeting - Rome, Italy 23-25 October 2017.
- ICA Experts Group on Archival Description, Records in Context: A Conceptual Model for Archival Description, Consultation Draft v0.1 September 2016, International Council on Archives, Paris, September 2016, pp. 116.
- Ketelaar, E. (2011). Cultivating Archives, *Archival Science* 12 (2012) 19-33.
- KROENKE (2009). *Database Processing 11/E: Fundamentals, Design, and Implementation* (Paperback). Pearson Education.

- Llanes-Padrón, D., Pastor-Sánchez, J.A. (2017). Records in contexts: the road of archives to semantic interoperability. *Program*, Vol. 51 Issue: 4, 387-405.
- M. Achichi, R. Bailly, C. Cecconi, M. Destandau, K. Todorov, & R. Troncy. DOREMUS: Doing Reusable Musical Data. In 14th International Semantic Web Conference (ISWC), 2015.
- McKemmish, S. (2017). Recordkeeping in the continuum: An Australian tradition. In A. J. Gilliland, S. McKemmish, & A. J. Lau (Eds.), *Research in the Archival Multiverse* (pp. 122-160). Clayton, Victoria: Monash University Publishing.
- Pearson Education (2009). *Database Processing 11/E: Fundamentals, Design, and Implementation*.
- Reed, Barbara (2017). New conceptual model for recordkeeping description, *Records in Contexts*. (2017. 2. 1).

[웹사이트]

- neo4j 블로그(rdbms vs graph datamodeling). 검색일자: 2018.12.31.
<https://neo4j.com/blog/rdbms-vs-graph-data-modeling/>
- neo4j 홈페이지. 검색일자: 2018.12.31. <https://neo4j.com/developer/guide-data-modeling/>
- PIAAF 프로젝트 홈페이지. 검색일자: 2018.12.31. <http://piaaf.demo.logilab.fr/>
- Records in Contexts - Compendium 홈페이지. 검색일자: 2018.12.31.
<https://web.esrc.unimelb.edu.au/ICAD/biogs/E000166b.htm>: 'RiC Primary Entities: What are they?'
- Records in Contexts - Compendium 홈페이지. 검색일자: 2018.12.31.
<https://web.esrc.unimelb.edu.au/ICAD/biogs/E000396b.htm>: 'RiC Supporting Entities: What are they?'
- RiC-IM HOME: ICA Records in Context - Compendium. 검색일자: 2018.12.31.
<https://web.esrc.unimelb.edu.au/ICAD/index.html>

• 국문 참고자료의 영어 표기

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

- Baek, Wook-in (2013). *Digital data · information · knowledge*. Seoul: CommunicationBooks, Inc.
- Fowler, Martin (2002). *Patterns of Enterprise Application Architecture*. Boston: Addison-Wesley Professional.
- Hernandez, M. J. (2013). *Database Design for Mere Mortals (3rd): A Hands-on Guide to Relational Database Design*. Boston: Addison-Wesley Professional.

- Hwang, Jin-hyun & Yim, Jin-Hee (2012). A study on Multiple Entity Data Model Design for Visual-Arts Archives and Information Management in the case of the KS X ISO 23081 Multiple Entity Model. *The Korean Journal of Archival Studies*, 33, 155-206.
- Hyun, Moon Soo & Seol, Moon-won (2018). An Exploration of the Direction of Development of the Next Generation Conceptual Model for Organizing Public Digital Records. *The Korean Journal of Archival Studies*, 56, 183-212.
- Hyun, Moon Soo (2014). A Study on the Conceptualization of Information Resources for Localities Based on the FRBRoo/CIDOC CRM. *Journal of the Korean Biblia Society for Library and Information Science*, 25(4), 265-290. <https://doi.org/10.14699/kbiblia.2014.25.4.265>
- Institute of Digital Archiving Myongji University (2017a). A Study on Conceptual Redesign of the next generation record management: Part 1.
http://www.archives.go.kr/next/news/viewPublicationList.do?bg_no=428.
- Institute of Digital Archiving Myongji University (2017b). A Study on Conceptual Redesign of the next generation record management: Part 2.
http://www.archives.go.kr/next/news/viewPublicationList.do?bg_no=429.
- Kim, Chi Soo (2015). *Easy-to-learn Software engineering*. Seoul: Hanbit Media, Inc..
- Kim, Ik-han (2004). Evaluating Records and Their Descriptive Elements in the Records Management of Korea on the Basis of the Characteristics of a Record and Recordkeeping Metadata Standards. *The Korean Journal of Archival Studies*, 10, 3-26.
- Kim, Sang Rae (2015). *Data Modeling Stories Determining Project Success or Failure*. Seoul: Hanbit Media, Inc.
- Kleppmann, Martin (2016). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Sebastopol, California: O'Reilly Media.
- Korean Society of Archival Studies (2014). *Dictionary of Records and Archival Terminology*.
- Lee, Chun Sik (2005). *Database design and implementation (data modeling with performance considerations)*. Seoul: Hanbit Media, Inc.
- Lee, Chun Sik (2008). *Design and build a database that seems to know*. Seoul: Hanbit Media, Inc.
- Lee, Hyun Sil & Han, Sung Kuk (2006). Conceptual Modeling of Record Management Metadata. *Journal of the Korea Society for Information Management*, 23(3), 23-48.
<https://doi.org/10.3743/KOSIM.2006.23.3.023>
- Lee, Yu-Been & Rieh, Hae-young (2017). A suggestion of interface for ontology-based record

- retrieval system. *Journal of Korean Society of Archives and Records Management*, 17(1), 217-244. <https://doi.org/10.14404/JKSARM.2017.17.1.217>
- Newman, Sam (2015). *Building Microservices: Designing Fine-Grained Systems*. Sebastopol, California: O'Reilly Media.
- Oh, Jin-Kwan & Yim, Jin-Hee (2018). A Study on Designing a Next-Generation Records Management System. *Journal of Korean Society of Archives and Records Management*, 18(2), 163-188. <https://doi.org/10.14404/JKSARM.2018.18.2.163>
- Okuno, Mikiya (2015). *Introduction to relational database practice(理論から學ぶデータベース實踐入門)*. Japan: Gijutsuhyoronsha.
- Park, Zi-Young (2008). A Study on the Application of a CRM-based FRBR Ontology for Cultural Heritage Information: Based on the FRBRoo(object-oriented FRBR). *Journal of the Korean Biblia Society for Library and Information Science*, 19(2), 45-62.
- Park, Zi-Young (2016). Analyzing the Next-generation Archival Description Standard: "Record in Context" of ICA EGAD. *Journal of Korean Society of Archives and Records Management*, 16(1), 223-245. <http://dx.doi.org/10.14404/JKSARM.2016.16.1.223>
- Park, Zi-Young (2017a). Transition of Archival Description from ISAD(G) to Record in Context Conceptual Model. *Journal of Korean Society of Archives and Records Management*, 17(1), 93-115. <https://doi.org/10.14404/JKSARM.2017.17.1.093>
- Park, Zi-Young (2017b). Improving Archival Descriptive Standard Based on the Analysis of the Reviews by Archival Communities on RiC-CM Draft. *The Korean Journal of Archival Studies*, 54, 81-109.
- Park, Zi-Young (2017c). Expanding and Improving FRBR Conceptual Model through FRBRoo. *Journal of the Korea Society for Information Management*, 34(4), 201-225. <https://doi.org/10.3743/KOSIM.2017.34.4.201>
- Rieh, Hae-young (2013). *Record Organization*. Seoul: SuninBook.
- Silverston, Len (2008). *The Data Model Resource Book, Vol. 1: A Library of Universal Data Models for All Enterprises(Revised Edition)*. Sebastopol, California: O'Reilly Media.
- Won, Jong-Kwan (2008). *A Study on Reconciliation of Evidence and Memory Based on the Characteristics of the Records Continuum*. Master's Thesis, Department of Information and Records Management, The Graduate School of Hankook University of Foreign Studies.
- Yim, Jin-Hee, Lee, Dae-Wook, Kim, Eun-Sil, & Kim, Ik-Han (2012). Directions for Developing Database Schema of Records in Archives Management Systems. *The Korean Journal of Archival Studies*, 34, 57-105.

