

# re3data를 기반으로 한 인문사회 RDR 연구\*

## A Study on Analysis of Research Data Repository in Humanities and Social Sciences

조재인 (Jane Cho)\*\*

박종도 (Jong-Do Park)\*\*\*

### 초록

오픈데이터헌장을 계기로 연구데이터 공유에 대한 논의가 지속되는 가운데, 미국, 영국, 일본 등을 중심으로 연구지원 기관들은 연구자들에게 결과물을 신뢰할 수 있는 저장소에 기탁하도록 유도하고 있다. 인문사회분야는 생명, 자연과학 분야에 비해 연구데이터 공유 문화나 저장 인프라가 미성숙하지만 이 역시 데이터의 영속적 접근과 활용 보장을 위해 신뢰할 수 있는 저장 인프라의 구축과 운영이 필요하다. 본 연구는 이러한 배경에서 re3data에 등록된 305개 인문사회분야 데이터 레포지토리(RDR: Research Data Repository)를 대상으로 현황을 파악하고 운영 수준에 따라 4개의 군집으로 유형화해 분석하였다. 그 결과 전체의 70%를 차지하는 보편적 수준의 RDR이 유형화되었으며, 독일이 운영 주체이거나 언어학 분야가 두드러진 우수 군집도 확인되었다. 한편, 인문사회 하위주제 영역에 있어 아카이빙되는 데이터 유형에 차이가 존재하는지 대응일치 분석(Correspondence Analysis)을 통해 확인한 결과, 역사 및 예술분야는 이미지, 사회계열 전반은 통계데이터나 오피스문서, 언어학은 오디오, 텍스트, 코드 등이 관련성 있는 데이터 유형으로 분석되었다.

### ABSTRACT

As the discussions on sharing research data prevail by the chance of the inauguration of the International Open Data Charter, research support organizations in the United States, the United Kingdom, and Japan are encouraging researchers to deposit their findings in a credible repository. Humanities and social sciences field, in which research data sharing culture and storage infrastructure are immature compared to life science and natural science, also needs to establish and operate a reliable storage infrastructure to guarantee the continuous access and utilization of data. This study analyzed the overall operational status of 305 subject repositories registered in re3data for the humanities and social sciences and clustered them according to the operational level using 5 indicators. As a result, 70% of the population were identified as universal clusters, and 20% of the excellent cluster was found to have the largest number of linguistic fields and the German-operated. In addition, this study confirmed through correspondence analysis that there is a relation between the sub-theme fields of humanities and social sciences and the types of data to be archived. The history and art domains are related to images, and social studies are related to statistical data. Linguistics has also been analyzed to be related to audio, plain text, and code.

키워드: RDR, re3data, 인문사회, 연구데이터, 데이터 레포지토리

Research data repository, re3data, Humanities and Social Science, Open access

\* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A5A2A02065467).

\*\* 인천대학교 문헌정보학과 교수(chojane@inu.ac.kr) (제1저자)

\*\*\* 인천대학교 문헌정보학과 조교수(jdp@inu.ac.kr) (교신저자)

논문접수일자 : 2019년 5월 13일 논문심사일자 : 2019년 6월 3일 게재확정일자 : 2019년 6월 4일  
한국비블리아학회지, 30(2): 69-87, 2019. [http://dx.doi.org/10.14699/kbiblia.2019.30.2.069]

## 1. 서론

학술연구자들은 연구를 수행하는 과정에서 다양한 데이터를 수집하거나 생성한다. 그러나 지금까지 대다수의 연구데이터는 비공개 상태로 연구자 자신에 의해 개별적으로 관리되어져 왔기 때문에 이를 활용해 연구 결과를 검증하거나 재활용하는데 어려움이 존재하였다. 따라서 신뢰할 수 있는 데이터 레포지토리(RDR: Research Data Repository)를 통해 데이터가 장기간 보존되고 쉽게 접근될 수 있다면 새로운 과학적 발견을 유도하고 중복 작업을 줄여 연구 및 공공 사회에 여러 가지 이익을 가져올 수 있게 된다.

G8 과학장관 회의에서 합의된 오픈데이터헌장(Department For Business, Innovation & Skills Prime Minister's Office, 2013)을 계기로 전 세계적으로 데이터 개방과 공유에 대한 논의가 지속되는 가운데, 미국 백악관과학기술정책국은 연방 예산에 의한 연구 성과물이 공공에게 오픈 액세스될 수 있도록 하는 정책의 기초를 천명하였다. 그에 따라 연방정부기관에서는 데이터관리계획(Data Management Plan)의 제출을 의무화하고 있다(CENDI 2017). 영국의 RCUK(Research Councils UK), 일본의 과학기술진흥원(Japan Science and Technology Agency) 등에서도 유사 정책들이 개발되었으며, 우리나라도 최근 국가연구개발사업의 관리 등에 관한 규정(대통령령 제 29625 2018.3.19. 일부 개정)을 통해 연구개발계획서 작성 시 데이터관리계획을 수립하도록 유도하고 있다.

연구지원 기관들이 결과물을 신뢰할 수 있는 저장소에 기탁하도록 유도하면서 데이터 저장 인프라의 지속 가능성과 신뢰성 향상을 위한

노력도 요구되고 있다. 데이터 저장 인프라의 핵심 요소인 연구데이터 레포지토리(Research Data Repository, 이하 RDR)는 각 분야의 과학공동체가 사용하는 연구 기반 시설로(南山泰之 2017) 연구자가 자신의 데이터를 관리, 저장 및 제공하고 다른 출처의 데이터를 쉽게 찾을 수 있도록 하는 시스템이다. 그러나 아직까지 국가마다 주제 분야마다 각기 다른 운영 수준을 보이고 있다고 평가되고 있다. 특히 데이터 공유의 효과가 크고 직접적인 생명과학 분야에 비해 독자적인 작업과 해석을 통해 지식을 창출하는 인문사회분야는 연구데이터 공유 문화나 저장 인프라의 운영 수준이 저조하다. 그러나 인문사회 분야의 연구데이터 역시 자연과학 분야와 마찬가지로 후속 연구에 지속적으로 활용될 수 있는 원자료적 가치를 가지는 질적 양적 기초 자료이므로 이들의 영속적 접근과 활용 보장을 위해 신뢰할 있는 저장 인프라의 구축과 운영이 필요하겠다.

따라서 본 연구는 이러한 배경으로 re3data에 등록된 인문사회분야 주제 RDR의 운영 현황과 수준을 이해해 보고자 한다. 다양한 지표를 활용하여 하위주제영역과 운영 국가에 따라 RDR의 운영 수준을 파악하고 유사 수준의 RDR을 클러스터해 평가함으로써 보편적 수준의 인문사회 RDR의 모습을 도출해 내고자 한다. 더불어 연구데이터의 유형이 역사, 언어, 경제 등 하위주제분야에 따라 어떠한 차이를 보이는지 식별함으로써 질적자료를 포함해 원자료적 가치를 가지는 인문사회분야 데이터의 유형을 실증적으로 이해해 보고자 한다.

## 2. 이론적 배경

### 2.1 인문사회분야 연구데이터와 RDR 현황

#### 2.1.1 인문사회분야 연구데이터

OECD(Organization for Economic Co-operation and Development)에서는 연구데이터를 수치, 문자, 이미지나 음성 등의 사실적 데이터로 과학연구의 주요 원천일 뿐 아니라 연구결과 검증에 위한 필수 도구라고 규정하고 있다(OECD 2007). DataCite(2012)에서는 연구데이터를 관찰, 실험, 경험에 기반하는 사실 데이터 뿐 아니라 통계기록, 음원, 미디어, 이미지 등 다양한 매체에 기록된 부산물로 정의하고 있다. 더불어 데이터를 기술한 데이터페이퍼와 회색문헌까지도 연구데이터의 범주로 인정하여 고유 식별자를 부여하고 있다(DataCite 2012). 서울대학교 연구윤리 지침에서도 연구데이터란 실험의 재료나 과정 및 결과 뿐 아니라, 관찰이나 현장조사 결과와 같은 원자료를 포함한다고 정의하고 있다(서울대학교 2010). 신영란과 정연경(2007)은 인문사회 연구데이터를 후속 연구에 지속적으로 활용될 수 있으며 학술 보존의 가치가 있는 양적 질적 자료라고 언급하였다. 구체적으로 실험의 재료나 과정, 관찰, 설문조사 결과 등의 원자료로서, 센서스, 표본조사 자료, 통계 자료 뿐 아니라, 어문학자료, 문화예술·민속 자료 등을 포괄한다고 설명하였다.

연구데이터의 가치는 일반적으로 재사용 및 통합을 통해 새로운 과학으로 창출이 가능하고 연구결과에 대한 반복적 실험과 검증을 통해 과학 발전을 도모할 수 있으며, 중복 연구 비용을 절감시킬 수 있다는 것이다(조재인 2016).

인문사회분야의 연구데이터는 위에서 언급한 동일한 가치를 가지지만 후속연구 활성화와 교수 학습에 활용되어 연구 인력 양성에 기여하는 교육과 학습, 훈련을 위한 자원으로도 활용될 수 있다(Van den Eynden et al. 2011). 따라서 의료 혁신, 산업 개발 등의 가치가 강조되는 생명과학분야와는 달리 교육적, 연구적, 증거적 측면의 가치가 강조된다고 하겠다(신영란, 정연경 2012).

#### 2.1.2 RDR 현황

본 연구가 분석대상으로 하는 re3data.org는 인문사회분야 뿐 아니라, 생명공학분야를 포함해 2,000여개의 전분야 RDR에 대한 정보를 제공하고 있다. re3data는 연구조사자, 연구비지원기관, 출판업자 및 기타 이해 관계자가 RDR을 탐색, 평가 할 수 있도록 운영 주체, 규정 및 라이선스, 채택하고 있는 각종 기술 표준 등의 운영 정보를 제공하고 있다. European Commission과 National Science Foundation과 같은 연구비 지원 기관은 데이터 관리 및 공유와 관련된 지침에서 re3data 참조를 권고하고 있으며 Copernicus Publications, PeerJ, PLOS ONE 및 Nature's Scientific Data와 같은 출판사 및 저널도 데이터 기탁 장소 선정 도구로 re3data 활용을 권장하고 있다(Kindling 2017). 한편, re3data에 공개된 RDR의 현황을 요약하자면 <표 1>과 같다. 유형은 주제저장소가 절반 이상을 차지하며, 분야는 생명과학과 자연과학 분야가 지배적이다. 운영 주체 국가로서는 미국이 압도적으로 많으며, 독일, 영국이 그 뒤를 따르고 있다. 한편 WDS(World Data System)와 DSA(Data Seal of Approval)를 중심으로 인증 체계가 운영되

〈표 1〉 RDR 현황(re3data.org)

주제분야		저장소 유형		인증		국가	
				CLARIN certificate	24		
				CoreTrustSeal	12		
인문사회	305	주제 기관 기타	1724 508 196	DIN 31644	1	미국	954
생명과학	833			DSA	55	독일	318
자연과학	802			RatSWD	32	영국	282
공학	121			TRAC	1	EU	184
				WDS	71	캐나다	142
				other	14		

※ 복수 속성을 가지는 RDR은 중복 집계됨

고 있으며 WDS의 경우 주로 자연과학 분야, DSA의 경우 인문사회분야 RDR이 인증을 취득하고 있다.

## 2.2 선행연구

국내에서 인문사회분야 연구데이터에 관한 연구는 주로 KRM(Korea Research Memory)을 주제로 하고 있다. 구체적으로 살펴보면, 인문학분야 연구데이터의 수집 및 활용성 증진(십원식, 안혜연, 변제연 2015), 그리고 연구 과정에서 생산되는 기초학문자료에 대한 아카이빙 및 서비스 방안이 제시된 바 있다(이정연 외 2007; 고영만, 배경재 2011; 황정선, 윤유라 2016; 신영란, 정연경 2012).

한편, 신영란과 정연경(2012)은 KRM의 문제점과 개선과제를 도출하는 과정에서 해외 인문사회 아카이브 운영 사례를 집중적으로 조사하여 발표하였다. 영국의 AHDS, ESDS, 독일의 GESIS, 미국의 HRAF와 ICPSR 등 인문사회분야 RDR 조사를 통해, 해외 아카이브는 기관단위 협력 체계를 기반으로 국가적 수준으로 성장하고 있다고 보고하였다. 서비스 측면에서

는 표준 식별체계나 인용방법제시, 풍부한 메타데이터나 온라인 통계분석 또는 질적자료 내용분석 도구 등을 지원하고 있으며 DDI, DOI, OAI-OMH 등의 표준 적용, 생애주기 모델을 기반으로 한 큐레이션 활동을 토대로 운영되고 있다고 보고하였다. Gutmann 등(2004)은 ICPSR과 UKDA(UK Data Archive)의 운영 사례를 분석해 사회과학 데이터의 아카이빙과 장기보존에 대한 방안을 제시한 바 있으며, Rasmussen과 Blank(2007)는 인문사회분야 연구데이터 기술과 장기 보존을 위해 DDI(Data Documentation Initiative)를 소개하면서 그 가치와 유용성을 강조한 바 있다.

한편, 본 연구와 마찬가지로 re3data를 기반으로 RDR을 분석한 연구는 다음과 같이 수행된 바 있다. Pampel 등(2013)은 이기종 RDR의 유형을 제시하고 레지스트리를 통해 적합한 저장소를 식별하는 방법을 보여주었으며, Kindling 등(2017)은 2015년 말을 기준으로 re3data.org에 등록되어있는 1,381개의 RDR을 유형과 주제에 따라 분석하고 액세스 조건 등을 살펴보았다. Cho(2019)는 152개의 아시아 국가 RDR의 운영 수준을 분석하고 주제 키워드의 가중

네트워크 분석(PFNet)을 수행함으로써 아시아 국가가 주력하고 있는 연구데이터의 주제 분야를 발견하였다. 마지막으로 Kim(2017)은 re3data의 메타데이터를 수집하여 누락된 필수 요소(subject, dataUploadType)와 속성(policy Type), 불충분한 통계 어휘(institutionType, policyType, dataUploadRestriction, dataAccess Type, contentType) 등을 제시해 메타데이터 운영상의 개선 방향을 제안하였다.

### 3. 연구의 방법

본 연구의 분석 대상 데이터는 re3data에 등록된 인문사회분야의 주제 RDR로 한다. 2018년을 기준으로 필터 기능을 이용해 subject에서 Humanities and Social Sciences로 제한하고 더불어 Repository types에서 disciplinary로 한정해 얻은 305개의 RDR를 분석 대상으로 선정하였다. 분석 대상 RDR의 메타데이터에서 일반사항, 정책, 이용 및 접근 조건, 라이선스, 표준에 관한 사항을 추출해 다음과 같은 분석에 활용하였다.

첫 번째, 인문사회 분야 RDR의 전반적인 현황을 이해하기 위하여 추출한 세부 속성을 대상으로 통계적인 분석을 수행하였다. 운영국가, 세부 주제, 콘텐츠 유형, 인증여부 등의 일반사항과, 정책, 이용 및 접근조건과 라이선스에 관한 사항, 그리고 메타데이터 및 영구식별자등 표준에 관한 사항을 대상으로 분석을 수행하였다.

두 번째, 인문사회 분야 RDR을 군집화하여 그 운영 수준을 평가해 보았다. 운영 수준 평가를 위한 변수로 오픈액세스여부, 영구식별자제

공여부, 이용조건 및 라이선스부여여부, 정책제 공여부, 인증여부를 채택해 계층형 군집 분석(Hierarchical Clustering)을 수행하였다. 이렇게 생성된 군집단위로 국가 및 세부 주제 분야의 분포 정도를 비교함으로써 인문사회 영역에 있어 어떠한 하위주제 분야, 그리고 어떠한 국가에서 운영되고 있는 RDR의 운영 수준이 우수한지 살펴 보았다. 더불어 보편적 수준의 RDR의 모습과 분포 정도도 검토하였다.

세 번째, 경제, 역사, 언어 등 인문사회 영역의 하위주제분야와 데이터 유형간에 유의미한 관계가 존재하는지 파악하기 위하여 대응일치 분석(Correspondence Analysis)을 수행하고 그 관계값을 2차원 지도에 플로팅하였다. 대응일치분석은 다차원척도법의 일종으로 행범주와 열범주에 대한 최적의 수량화 값을 계산하여 자료를 2차원 평면상에 플롯팅함으로써, 변수간의 관련성을 파악하는 기법이다(고봉성 2001). 기술통계 분석을 포함해 위에서 열거한 모든 통계적 분석은 SPSS 32를 활용하여 수행하였다.

## 4. 분석 결과

### 4.1 일반 사항 분석

RDR은 대부분 여러 국가가 연합한 컨소시엄 형태로 운영되는데 인문사회분야의 RDR도 마찬가지이다. 참여 빈도가 가장 높은 국가는 미국이며, 독일과 영국도 주요 운영 국가로 집계되었다. 일본 2개, 한국은 1개가 포함되어 아시아 국가는 많지 않았다. 한국에서는 KLIPS 한국노동패널조사가 인문사회분야의 RDR로 등

록되어 있는 것으로 나타났다. RDR의 언어는 영어가 가장 압도적이었으며, 그 다음 독일어로 나타났다.

두 번째, 인문사회과학 RDR은 독립된 주제 분야를 가지기 보다 여러 분야에 걸쳐 있는 경우가 대부분이다. 각 RDR에 부여되어 있는 주제명 중 앞서서부터 세계를 선별하여 빈도 분석을 해본 결과, 경제 분야(20.2%)가 가장 많았으며, 언어(15.7%), 사회학(65%), 역사(9.6%), 고고학(6.6%), 예술(6.2%), 교육(5.7%), 정치(5.1%) 순을 보였다. 문학, 철학, 신학관련 주제 분야의 RDR은 많지 않은 것으로 나타났다. 한편 분야별로 RDR을 예시하면 언어학 분야에서는 프랑스 Aix-Marseille 2 대학 등이 컨소시엄으로 운영하는 SLDR(Speech and Language Data Repository)이 있으며, 사회학분야에서는

미시간대학의 ICPSR, 역사 및 고고학은 미국의 요크대, 영국의 AHRC(Arts and Humanities Research Council), JISC(Joint Information Systems Committee), NERC(Natural Environment Research Council)가 참여하는 ADS(Archaeology Data Service)를 들 수 있겠다 (<표 2> 참조).

세 번째, RDR이 아카이빙하고 있는 데이터 유형도 여러 종류를 망라하지만, 각 RDR에 부여되어 있는 데이터 유형 중 앞서서부터 다섯 개를 선별하여 집계해 보았다. 그 결과 <표 3>과 같이, 오피스문서(Standard office documents 19.8%), 플레인 텍스트(Plain text 13.5%)가 가장 많았으며 그 다음 통계데이터(Scientific and statistical data 12%), 이미지(Images 8.4%), 오디오비주얼데이터(Audiovisual data 8.1%)

<표 2> 인문사회 분야 RDR의 주제 분야

주제분야	빈도	비율
Economics	95	20.2
Linguistics	74	15.7
Empirical Social Research	65	13.8
History	45	9.6
Ancient Cultures	31	6.6
Fine Arts, Music, Theatre and Media Studies	29	6.2
Education Sciences	27	5.7
Political Science	24	5.1
Non-European Languages and Cultures, Social and Cu	18	3.8
Psychology	18	3.8
Jurisprudence	17	3.6
Communication Science	9	1.9
Literary Studies	9	1.9
Theology	7	1.5
Philosophy	2	0.4
Total	470	100

※ 복수 속성을 가지는 경우 중복 집계됨

〈표 3〉 인문사회 분야 RDR의 데이터 유형

데이터 유형	빈도	비율
Standard office documents	224	19.8
Plain text	153	13.6
Scientific and statistical data	136	12.0
Images	95	8.4
Audiovisual data	92	8.1
Structured text	84	7.4
Raw data	76	6.7
Structured graphics	62	5.5
other	55	4.9
Archived data	52	4.6
Databases	50	4.4
Software applications	34	3.0
Networkbased data	7	0.6
Source code	7	0.6
Configuration data	2	0.2
Total	1,129	100

※ 복수 속성을 가지는 경우 중복 집계됨

순을 나타냈다. 그 밖에 소프트웨어(Software applications), 네트워크 데이터(Network based data), 아카이브데이터(Archived data), 소스코드(Source code) 등이 존재하는 것으로 나타났다.

네 번째, NSF, RCUK 등 전 세계적으로 공공기금을 부여한 연구지원기관에서는 연구 결과물을 신뢰할 수 있는 저장소에 기탁하도록 하는 정책을 강화하고 있다. 따라서 신뢰할 수 있는 연구데이터 저장소의 조건을 수용해 국제 인증을 획득하기 위한 노력이 확산되고 있다. 대표적인 인증체계로는 DSA와 WDS가 있다. DSA는 데이터의 장기적 발견 및 활용을 목적으로 네덜란드를 중심으로 한 국제위원회에서 관리되는 기준이며, WDS는 ICSU(International Council for Science)의 학제간조직으로 데이터 저장소에 관한 평가 인증 기준을 가진다. 최근 ICSU-WDS와 DSA 양자가 협력하여 보

다 핵심적인 요구 사항을 찾는 노력을 하게 되었는데, 그 결과 CoreTrustSeal을 출범해 “신뢰할 수 있는 데이터 저장소를 인증하기 위한 핵심적인 통일 요구 사항(Core Trustworthy Data Repositories Requirements)”을 바탕으로 새로운 인증 체계를 운영하게 되었다. 2018년을 기준으로 인문사회분야에서는 DSA의 인증을 받은 기관이 29개, WDS의 인증을 받은 기관이 2개 존재하였으나 WDS와 DSA가 CoreTrustSeal로 대체되어 향후 숫자적 변동이 있을 것으로 예측된다. 그 밖에 Rat für Sozial-und Wirtschaftsdaten의 인증을 받은 기관이 22개 존재하였으나 78%에 이르는 대부분의 기관은 미인증 상태인 것으로 분석되었다. Rat für Sozial-und Wirtschaftsdaten은 독일 연방교육연구부(BMBWF: Bundesministerium für Bildung und Forschung)의 재정지원을 받

아 데이터 센터의 인증 및 평가를 주관하는 기관으로 독일이 참여하는 RDR 중 다수가 여기를 통해 인증을 받는 것으로 나타났다. 한편, 인증된 RDR이 많은 국가를 살펴보면 자체적인 국가 인증 시스템을 가진 독일과 유럽 연합이 가장 많았으며, 아시아권 국가는 한곳도 없는 것으로 나타나고 있다(〈표 4〉 참조).

다섯 번째, 영구식별자를 부여하고 있는 여부를 살펴본 결과, 〈표 5〉와 같이 142개인 45%가 식별자를 부여하고 있는 것으로 확인되었다. 식별자를 부여하고 있는 경우는 DOI(82개, 26.9%)가 가장 많이 채택되고 있는 것으로 나타났다. 그밖에 독자적 handle 시스템(49개, 16.1%)이나 URN(7개, 2.3%) 등이 이용되는 것으로 나타났다. RDR 시스템은 알려지지 않은 시스템을 사용하는 경우가 전체의 절반 이상을 차

지하였으나, 알려진 시스템인 Fedora, dSpace, Nesstar, DataVerse 등을 사용하는 경우도 눈에 띄었다. 메타데이터의 경우는, 표준을 채택하지 않는 경우가 대부분이었으나, 알려진 표준으로는 DDI(20개, 6.6%)가 가장 많았으며, DC와 DataCite Metadata Schema가 그 뒤를 이었다. DDI는 북미와 유럽을 중심으로 통계 데이터용 메타데이터 필요성이 증가하면서 데이터 기록화와 관련된 국제 단체 DDI(Data Documentation Initiative)가 설립되어 제안된 통계 메타데이터 표준이다(최윤경, 최유성 2015). 현재 DDI-L(DDI-Lifecycle)과 DDI-C(DDI-Codebook) 두 개의 규격으로 개발되어 있으며, XML 기반의 DDI-L 3.1 버전(2009)과 DDI-C 2.5 버전(2012)이 최신 버전으로 공개되어 있다(신영란, 정연경 2012).

〈표 4〉 인문사회 분야 RDR 인증 취득 현황

인증 종류	빈도	비율	인증을 취득한 국가	빈도	비율
없음	239	78.4	Germany	33	48.5
Data Seal of Approval	29	9.5	European Union	10	14.7
Rat für Sozial-und Wirtschaftsdaten	22	7.2	UnitedStates	5	7.4
CLARIN certificate B	7	2.3	Netherlands	3	4.4
other	3	1.0	UnitedKingdom	3	4.4
CoreTrustSeal	2	0.7	Sweden	2	2.9
World Data System	2	0.7	CzechRepublic	2	2.9
DINI Certificate	1	0.3	Spain	2	2.9
Total	305	100.0	SouthAfrica	1	1.5
			RussianFederation	1	1.5
			Poland	1	1.5
			Norway	1	1.5
			Lithuania	1	1.5
			Ireland	1	1.5
			Denmark	1	1.5
			Australia	1	1.5
			Total	68	100.0



〈표 5〉 인문사회 분야 RDR PID, 시스템, 메타데이터 현황

PID	빈도	비율	시스템	빈도	비율	메타데이터	빈도	비율
none	163	53.4	unknown	152	49.8	none	240	78.7
DOI	82	26.9	none	56	18.4	DDI	20	6.6
hdl	49	16.1	other	40	13.1	Dublin Core	19	6.2
URN	7	2.3	Fedora	17	5.6	DataCite Metadata Schema	11	3.6
PURL	3	1.0	DSpace	10	3.3	ISO 19115	3	1.0
ARK	1	0.3	Nesstar	9	3.0	other	3	1.0
Total	305	100.0	DataVerse	8	2.6	Repository-Developed Metadata	3	1.0
			CKAN	7	2.3	OAI-ORE	2	0.7
			MySQL	3	1.0	ABCD	1	0.3
			dLibra	1	0.3	FGDC/CSDGM	1	0.3
			Eprints	1	0.3	Metadata standard name	1	0.3
			eSciDoc	1	0.3	RDF Data Cube Vocabulary	1	0.3
			Total	305	100.0	Total	305	100.0

#### 4.2 RDR 운영 수준 분석

re3data에서는 연구자들이 효과적으로 데이터를 기탁할 RDR를 선정할 수 있도록 지원하기 위하여 RDR 식별 시스템을 제공하고 있다. 여기에 사용되고 있는 주요 지표는 오픈액세스, 라이선스, 영구식별자, 인증, 정책수립여부의 다섯 가지이다. 본 연구에서는 이 다섯가지 지표들을 이분형(Binary) 데이터로 변환해 계층형 군집분석(Hierarchical Clustering)을 수행함으로써 300여개의 인문사회분야 RDR의 클러스터를 생성해 보았다. 그 결과 클러스터는 5가지 조건의 수용 여부에 따라 4개의 군집을 형성하였으며, 그 중에 모든 조건을 수용하는 우수형 클러스터와 가장 많은 분포를 보이는 보편형 클러스터를 식별할 수 있었다.

클러스터 결과를 제시하기 이전에 먼저 각 지표에 따른 기초 통계를 살펴보도록 하자. 그 결과 〈표 6〉과 같이 98%의 RDR이 공개 서비스를 하고 있으며 이용조건이나 규약에 대한

라이선스를 가지고 있는 것으로 나타났다. 또한 92%의 대부분이 보존 및 이용, 공개 접근, 정보 보호, 소유권 및 책임, 데이터 보호, 윤리 강령, 데이터 공유 등의 정책을 공개하고 있는 것으로 분석되었다. 그러나 영구식별자를 가지는 RDR의 규모는 전체의 46%로 절반에 미치지 않았으며, 인증을 받은 RDR도 66개로 21%에 그치는 것으로 나타났다.

한편 계층형 군집분석을 통해 생성된 4개 클러스터의 특성은 〈표 7〉을 통해 요약해 볼 수 있다. 218개의 대다수 RDR이 포함된 가장 큰 클러스터(CL2)가 형성되었으며, 5가지의 지표를 거의 모두 수용하는 우수형 클러스터(CL1)가 생성된 결과가 확인되었다. 또한 비공개형 RDR이 군집된 CL3, 다소 열악한 운영 조건을 보이는 RDR이 군집된 CL4로 구분되었다. 각각 클러스터의 특징을 구체적으로 살펴보면 다음과 같다.

첫 번째, 〈표 8〉의 CL1은 전체의 21%를 차지해 66개의 RDR이 군집되었다. 이 군집의 특

〈표 6〉 5가지 지표에 관한 기초 통계

지표		빈도	비율
open access	Y	303	98.3
	N	2	0.7
license	Y	305	100
pid	Y	142	46.5
	N	163	53.4
certification	Y	66	21.6
	N	239	78.4
policy	Y	286	93.7
	N	19	6.3

〈표 7〉 클러스터의 특성

Cluster	N	%	Open access	License	DOI	Certification	Policy	비고
CL1	66	21.6	100%	100%	87%	100%	100%	우수형
CL2	218	71.5	100%	100%	38%	0%	100%	보편형
CL3	2	0.7	0%	100%	50%	0%	100%	비공개형
CL4	19	6.2	100%	100%	0%	0%	0%	열악형

〈표 8〉 CL1의 주제분야 및 운영 국가

주제분야	빈도	비율	대표 운영 국가	빈도	비율
Linguistics	25	37.9	Germany	33	50.0
Empirical Social Research	18	27.3	European Union	10	15.2
Economics	6	9.1	United States	5	7.6
History	5	7.6	Netherlands	3	4.5
Ancient Cultures	4	6.1	Czech Republic	2	3.0
Education Sciences	4	6.1	Sweden	2	3.0
Fine Arts, Music, Theatre and Media Studies?	1	1.5	United Kingdom	2	3.0
Jurisprudence	1	1.5	Australia	1	1.5
Political Science	1	1.5	Denmark	1	1.5
Psychology	1	1.5	Ireland	1	1.5
Total	66	100	Lithuania	1	1.5
			Norway	1	1.5
			Poland	1	1.5
			Russian Federation	1	1.5
			South Africa	1	1.5
			Spain	1	1.5
			Total	66	100

성은 라이선스를 선언하고 있는 공개서비스 기반의 RDR로 영구식별자와 정책을 공개하고 있을 뿐 아니라 인증까지 획득해 5개의 조건을 모두 충족하고 있다는 것이다. 다시 말해 우수한 운영 수준을 가진 RDR들이 군집된 클러스터라고 할 수 있겠다. <표 8>을 보면 이 군집에 속하는 RDR 운영 국가는 독일이 절반을 차지하고 주제 분야는 언어학(37.9%)이 가장 두드러진 것을 확인할 수 있다. 이 군집에 속하는 RDR를 예시하면 다음과 같다. CLARIN-D(Common Language Resources and Technology Infrastructure - D) BMBF Federal Ministry of Education and Research 등이 운영하는 독일의 언어학 RDR IDS Repository, 미국 미시간대가 주도하고 있는 사회과학 분야의 Inter-university Consortium for Political and Social Research, 호주의 멜버른, 시드니, 호주국립대 등이 연합해 태평양지역 언어 관련 디지털 리소스를 아카이빙하는 RDR인 PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures) 등이다. 아시아 권에서는 이 군집에 포함되는 RDR이 부재한 것으로 분석되었다.

두 번째는 218개의 RDR이 포함된 클러스터(CL2)로 전체의 70%를 점유하고 있어 인문사회 영역에 있어 가장 보편적인 수준을 보이는 군집이라고 설명할 수 있겠다. 이 군집의 특징은 공개서비스되고 있으며, 라이선스와 정책을 제공하고 있으나 일부만이 영구식별자를 가지고 있다는 것이다. 또한 대부분이 인증을 받지 못한 상태이다. 여기에는 경제, 사회, 언어, 고대문화, 예술 등 다양한 주제 분야를 포괄하며, 미국이 압도적이지만 한국, 일본, 파키스탄 등

을 포함해 여러 국가들이 포함되는 것으로 조사되었다(<표 9> 참조).

세 번째, CL3는 단 3개의 RDR만을 포함하고 있는 클러스터이다. 다른 요소들은 보편적인 형태를 보이는 CL2와 유사하지만 비공개되어 있다는 특징을 가진다. 여기에 속하는 RDR은 고대 문화 및 역사 분야로 독일과 미국에서 운영되고 있는 것으로 확인되었다(<표 10> 참조).

마지막 네 번째는 19개(6.2%)의 RDR이 포함되어 있는 상대적으로 열악한 운영 조건을 가진 클러스터이다. 공개서비스되고 있으며, 라이선스를 게시하고 있으나 영구식별자와 인증을 부여받지 못하였으며, 운영 및 이용 정책까지 공개하고 있지 않은 경우가 대부분이다. 예술, 비유럽권 언어문화 분야 RDR이 여기에 포함되어 있는 것으로 나타났다(<표 11> 참조).

위에서 제시한 바와 같이 RDR을 운영 조건에 따라 군집화한 결과, 가장 보편적인 형태의 인문사회 분야 RDR이 유형화되었다. 70% 이상을 점유하는 가장 일반적인 모습을 가진 인문사회 분야의 RDR은 다른 조건들은 모두 수용하고 있으나 식별자를 부여하지 않으며, 국제적 인증을 취득하지 못한 상태의 운영 수준을 보이는 것으로 해석되었다. 그러나 모든 조건을 충족하고 있는 우수한 수준의 RDR이 20%를 차지하고 있으며 거기에는 국가내 자체 인증 체계를 갖춘 독일의 RDR이 가장 많으며 언어학 분야가 두드러진 것으로 요약할 수 있겠다.

〈표 9〉 CL2의 주제분야 및 운영 국가

주제분야	빈도	비율	대표 운영 국가	빈도	비율
Economics	43	19.7	United States	96	44.0
Empirical Social Research	35	16.1	Germany	35	16.1
Linguistics	32	14.7	United Kingdom	19	8.7
Ancient Cultures	25	11.5	Canada	10	4.6
Fine Arts, Music, Theatre and Media Studies	16	7.3	European Union	9	4.1
History	16	7.3	France	7	3.2
Education Sciences	15	6.9	International	6	2.8
Political Science	10	4.6	Switzerland	6	2.8
Psychology	10	4.6	Australia	5	2.3
Non-European Languages and Cultures, Social and Cu	7	3.2	Netherlands	3	1.4
Communication Science	4	1.8	Austria	2	0.9
Jurisprudence	3	1.4	Greece	2	0.9
Philosophy	1	0.5	Italy	2	0.9
Theology?	1	0.5	Japan	2	0.9
Total	218	100	Norway	2	0.9
			Belgium	1	0.5
			Cote d'Ivoire	1	0.5
			Estonia	1	0.5
			Ghana	1	0.5
			Ireland	1	0.5
			Korea, Republic of	1	0.5
			Mexico	1	0.5
			New Zealand	1	0.5
			Pakistan	1	0.5
			Poland	1	0.5
			Russian Federation	1	0.5
			Sweden	1	0.5
			Total	218	100

〈표 10〉 CL3의 주제분야 및 운영 국가

주제분야	빈도	비율	대표 운영국가	빈도	비율
Ancient Cultures	1	50	Germany	1	50
History	1	50	United States	1	50
Total	2	100	Total	2	100

〈표 11〉 CL4의 주제 분야 및 운영 국가

주제 분야	빈도	비율	대표 운영국가	빈도	비율
Economics	4	21.1	United States	6	31.6
Fine Arts, Music, Theatre and Media Studies?	3	15.8	Germany	5	26.3
Non-European Languages and Cultures, Social and Cu	3	15.8	European Union	2	10.5
Empirical Social Research	2	10.5	France	1	5.3
Linguistics	2	10.5	Israel	1	5.3
Psychology	2	10.5	Netherlands	1	5.3
Ancient Cultures	1	5.3	South Africa	1	5.3
History	1	5.3	Switzerland	1	5.3
Political Science	1	5.3	United Kingdom	1	5.3
Total	19	100	Total	19	100

#### 4.3 하위주제 영역과 데이터 유형간의 대응일치 분석

한편 인문사회 분야의 하위주제 영역에 따라 데이터의 유형에 차이가 존재하는지 분석한 결과를 살펴보도록 한다. 앞에서 설명한 바와 같이 본 연구에서는 개별 RDR에 부여된 세부 주제 영역 및 데이터 유형에 대한 메타데이터 요소를 대상으로 변수간의 관계를 해석하였다.

분석 결과를 제시하면 〈표 12〉와 같다. 주제와 데이터 유형간의 교차 분석 결과에서 유의한 차이( $p=.000$ )가 존재하였으며 〈그림 1〉에서 제시한 대응 분석 결과에서 2차원 누적 설명력이 81%로 나타나 제시된 플로팅 결과는 신뢰할 수 있는 수준인 것으로 나타났다. 대응일

치분석에서는 변수간의 연관성을 2차원 좌표에서 점들의 분포를 보고 판단하므로 플로팅 결과를 통해 데이터의 유형과 주제 간의 관계를 설명해 보도록 하겠다.

〈그림 1〉을 보면, 2-3분면에 거쳐 경제, 심리, 정치, 교육, 사회 등 사회과학 전분야가 상하로 넓게 퍼져있다. 그리고 해당 분면에는 통계데이터(Scientific and statistical data)나, 오피스 다큐먼트(Standard office documents) 등의 데이터 유형이 근접하게 위치하고 있다. 이는 SPSS 형식의 통계데이터나 데이터를 설명하기 위한 오피스 다큐먼트 등이 사회과학 전 분야에 있어 가장 관련성이 높은 데이터 유형임을 의미한다. 두 번째, 1분면에는 사회과학분야는 부재하며 고고학, 역사, 예술, 철학 분야만이 근접한 위치

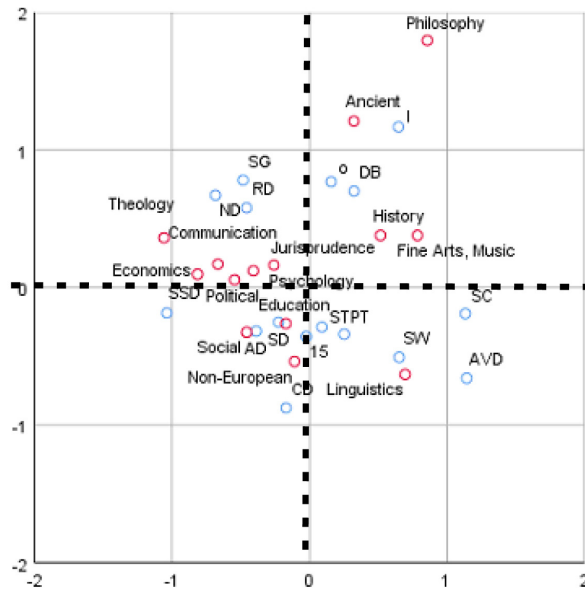
〈표 12〉 데이터와 주제 유형에 관한 교차분석 결과

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	356.6194	182	.000
Likelihood Ratio	343.3541	182	.000
Linear-by-Linear Association	1.480289	1	0.224
N of Valid Cases	1,129		

〈표 13〉 데이터와 주제 유형에 관한 교차표

주제	SD	PT	SSD	I	AVD	ST	RD	SG	O	AD	DB	SA	ND	SC	CD	계
Economics	39	18	41	7	0	16	15	16	8	8	8	4	1	0	1	182
Linguistics	44	43	11	12	44	25	13	4	7	9	9	13	0	3	1	238
Empirical Social Research	50	32	35	9	9	14	11	9	7	16	4	2	1	0	0	199
History	16	12	4	13	7	8	4	3	4	1	9	3	0	0	0	84
Ancient Cultures	15	11	5	27	7	5	11	13	11	5	10	1	1	1	0	123
Fine Arts, Music, Theatre and Media Studies	11	13	2	13	10	3	2	4	7	2	3	4	1	1	0	76
Education Sciences	16	10	10	4	5	3	7	3	3	4	0	4	1	1	0	71
Political Science	9	3	8	2	2	5	4	4	2	1	1	1	0	0	0	42
Non-European Languages and Cultures	8	3	4	1	4	1	1	0	1	3	1	0	0	0	0	27
Psychology	8	4	10	5	3	2	5	3	1	1	3	1	0	1	0	47
Jurisprudence	4	1	3	1	1	1	2	2	1	0	1	0	0	0	0	17
Communication Science	3	3	2	0	0	0	1	1	1	2	1	1	2	0	0	17
Theology	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	3
Philosophy	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	3
Total	224	153	136	95	92	84	76	62	55	52	50	34	7	7	2	1129

SD: Standard office documents, PT: Plain text, SSD: Scientific and statistical data  
 I: Images, AVD: Audiovisual data, ST: Structured text  
 RD: Raw data, SG: Structured graphics, O: Others  
 AD: Archived data, DB: Databases, SA: Software applications  
 ND: Networkbased data, SC: Source code, CD: Configuration data



〈그림 1〉 데이터 유형과 주제에 관한 2차원 플로팅 결과

에 자리 잡고 있다. 그리고 데이터 유형으로는 이미지(Images), 데이터베이스(Database) 등이 위치하고 있다. 통계데이터와 오피스 문서가 주로 아카이빙된 경제, 정치 분야와는 달리, 이미지가 인문예술 분야의 가장 지배적인 데이터 형식임을 의미한다. 마지막으로 4분면에는 주제 영역 중 유일하게 언어학만이 존재하며 여기에는 2-3분면과는 다른 데이터 유형이 위치해 주목된다. 오디오비주얼데이터(Audiovisual data), 구조화된 텍스트(Structured text), 플레인 텍스트(Plain text), 코드(Source code), 소프트웨어(Software applications)가 그것이다. 녹음녹화 데이터나 언어 자체를 아카이빙한 텍스트, 이를 전자화하는 과정에서 사용된 코드와 소프트웨어가 대표적인 언어학 분야의 데이터 유형이라는 것이다. 예를 들어 대표적인 언어학 관련 RDR인 PARADISEC(Pacific and Regional Archive for Digital Sources in Endangered Cultures)의 아카이빙 콘텐츠를 보면, 토착언어권의 노래, 방언, 구비문학에 대한 오디오 비주얼 또는 텍스트 아카이브가 주를 이룬다. 멸종 위기 소규모 문화권의 기록이 오디오, 비디오 데이터, 문자기록 등의 질적 자료 형태로 아카이빙되어 있다. 이에 따라 메타데이터도 통계데이터를 다루는 DDI 등이 아니라, OLAC(Open Language Archives Community)의 권고에 따라 결정된 고유한 메타데이터 셋이 사용되고 있다.

위에서 제시한 내용을 요약하자면, 경제, 심리학과 같은 사회과학 분야는 텍스트 기반의 문서나 통계 데이터셋, 고대문화, 역사, 예술 분야는 이미지, 그래픽 등의 유형, 그리고 언어학 분야는 말뭉치 등의 구조화된 텍스트, 오디오비주얼 데이터가 관련성 있는 유형으로 나타났다.

RDR에 아카이빙된 인문사회 연구데이터는 양적자료 뿐 아니라 문화, 민속 등을 관찰, 채록한 원자료적 형태의 질적 자료 등을 망라하고 있으며, 하위 주제 분야의 특성에 따라 각기 다른 형태를 가진다. 인문학은 대체로 내용 분석이 필요한 질적 기초자료, 사회과학분야는 통계 데이터 등의 양적 자료에 관련성을 보인다고 정리할 수 있겠다.

## 5. 결론

비교적 독자적인 작업을 수행해 개인적 해석이 담긴 연구 결과를 도출하는 인문학은 정보를 모으고 선택 해석하는 과정을 통해 지식으로 변환되므로 연구데이터를 규정 분류하고 관리하는 것이 쉽지 않다. 그러나 인문사회학 분야의 연구 과정에서 산출되는 양적 질적 자료는 후속 연구에 지속적으로 활용될 수 있는 원자료적 가치를 가진 자료로 생명 및 자연과학 분야의 연구데이터와 마찬가지로 보존과 공유, 재활용되어야 한다(심원식 외 2015). 국제 사회가 요구하는 수준으로 RDR 운영 수준을 제고하기 위해서는 접근의 영속성, 라이선스, 데이터의 검색과 식별 등을 신뢰할 수 있는 수준으로 향상시켜야 한다. 구체적인 발전 방향 모색을 위해서는 우선 유사 목적을 가진 해외 RDR의 실증 분석이 선행되어야 하는데 기초분석을 수행한 본 연구는 그러한 측면에서 의의를 가진다고 할 수 있겠다. 이 연구에서 발견된 내용을 요약하면 다음과 같다.

첫 번째, 인문사회분야의 RDR은 미국과 독일 등 서구를 중심으로 컨소시엄 형태로 운영

되고 있으며, 경제, 사회 분야 등을 주요 대상으로 하고 있다. 인문학 영역에서 언어, 역사, 고고학 분야를 다루는 RDR은 다수 존재하였으나 문학, 철학, 신학분야를 다루는 경우는 많지 않았다. 콘텐츠의 유형은 오피스 문서(Standard office documents)와 통계데이터(Scientific and statistical data)가 가장 많았으나 이미지(Images), 오디오비주얼데이터(Audiovisual data), 플레인텍스트(Plain text) 등도 존재한다. RDR은 국제 인증을 취득한 경우가 많지 않았지만, 10% 정도 DSA 인증을 취득하고 있으며 독일의 경우 연방교육연구부(BMBF)의 자체 인증을 취득한 경우가 많은 것으로 분석되었다. 그 밖에 영구식별자는 절반 정도만이 부여하고 있었으며, 메타데이터는 DDI를 가장 많이 채택하고 있는 것으로 분석되었다. 두 번째, RDR의 유형은 운영 수준에 따라 4개의 그룹으로 군집화되었다. 70%의 RDR이 분포된 가장 보편적 모습의 인문사회 RDR 군집이 형성되었는데, 이 군집에 속하는 RDR은 공개 서비스되고 있으며, 라이선스와 자체 운영 정책을 공개하고 있으나, 대부분 영구식별자를 부여하지 않으며, 인증을 취득하지 못했다는 특징을 가진다. 한편, 전체의 20%가 분포된 우수 RDR 군집은 인증을 포함해 다섯까지 운영 조건을 모두 충족하고 있었는데 언어학이 두드러지며 독일이 운영 주체인 경우가 가장 많았다. 마지막으로 인문사회영역의 세부 주제 분야에 따라 데이터 유형에는 통계적으로 의미 있는 차이가 나타났다. 사회과학 전반은 문서나 통계데이터가 주류를 이루나, 역사 및 예술분야는 이미지, 언어학은 오디오비주얼, 텍스트 등의 질적 자료가 관련성 있는 데이터 유형으로 분석되었다.

오픈데이터현장(Department For Business, Innovation & Skills Prime Minister's Office, 2013)을 계기로 연구지원 기관들이 결과물을 신뢰할 수 있는 저장소에 기탁하도록 하고 있다. 우리도 연구 결과물 기탁을 위한 저장소 인프라의 지속 가능성과 신뢰성 향상을 위한 인증 체계(Core Trustworthy Data Repositories Requirements)에 관심을 가질 필요가 있겠다. DSA와 ICSU-WDS에서 책정한 신뢰할 수 있는 데이터 저장소의 핵심적인 통일 요구 사항에는 조직기반, 디지털 개체 관리, 기술로 구성되며, 데이터의 발견 가능성, 명확한 권리와 라이선스 부여를 통한 접근성, 데이터의 재사용성과 신뢰성, 참조 가능성 등을 평가 요소로 포함하고 있다(ICSU 2016). WDS나 DSA의 인증을 중심으로 미국은 947개의 RDR 중 30개, 중국은 36개의 RDR 중 9개, 일본은 57개의 RDR 중 8개가 인증을 받고 있으나 등록된 한국 RDR 중에는 인증을 취득한 경우를 찾을 수 없다. 국제 공동연구가 증가하는 상황에서 연구 데이터를 신뢰할 수 있는 데이터 저장소에 기탁하는 것이 의무 요건이 된다면, 우리는 내국인 생산 연구 데이터의 해외 유출을 방조하는 것이 된다(Center for Open Data in the Humanities, 2017). 이 연구는 비록 인문사회 분야에 한정해 분석을 수행하였으나, 아시아 국가를 비롯한 국내 RDR이 전반적으로 얼마나 국제적 수준에 미치지 못하고 있는지도 가늠할 수 있었다. 하루 빨리 신뢰할 수 있는 수준의 연구데이터 저장소의 설치와 운영이 시급한 상황에서 본 연구 결과는 국내에서 관련 논의를 촉발하고 정책 수립 시 유용한 기초자료로 활용될 수 있을 것이다.



## 참 고 문 헌

- 고봉성. 2001. 대응일치 분석을 이용한 사이버 쇼핑몰의 조사연구. 『The Journal of Korean Data Analysis Society』, 3(4): 419-427.
- 고영만, 배경재. 2011. 상이한 데이터 구조의 데이터베이스간 통합 운영방안 연구. 『한국문헌정보학회지』, 45(3): 69-85.
- 南山泰之. 2017. 信頼できるデータリポジトリの中核的な統一要件. カレントアウェアネス-E, No. 320. [online]. [cited 2018.11.2]. <<http://current.ndl.go.jp/e1888>>.
- 서울대학교. 2010. 서울학교 연구윤리 지침. [online]. [cited 2019.1.5]. <[http://www.snu.ac.kr/research/images/down/research\\_08.pdf](http://www.snu.ac.kr/research/images/down/research_08.pdf)>.
- 신영란, 정연경. 2012. 국내 인문사회 연구데이터 아카이브의 개선방안에 관한 연구. 『한국기록관리학회지』, 12(3): 93-115.
- 심원식, 안혜연, 변제연. 2015. 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구. 『한국문헌정보학회지』, 49(3): 155-183.
- 이정연, 김정민, 최석두, 김이경. 2007. 기초학문자료 메타데이터 설계 분석 및 온톨로지 적용 방안 연구. 『한국문헌정보학회지』, 41(2): 291-316.
- 조재인. 2016. Data Citation Index를 기반으로 한 연구데이터 인용에 관한 연구. 『문헌정보학회지』, 50(1): 189-207.
- 최윤경, 최유성. 2015. 메타데이터 표준에 관한 연구(DDI\_2 vs DDI\_3), 제5회 국가통계방법론 심포지엄. [online]. [cited 2019.1.5]. <[http://kostat.go.kr/edu/sri\\_kor\\_new/3/6/5/index.board?bmode=read&aSeq=349108&pageNo=2&rowNum=10&amSeq=&sTarget=&sTxt=](http://kostat.go.kr/edu/sri_kor_new/3/6/5/index.board?bmode=read&aSeq=349108&pageNo=2&rowNum=10&amSeq=&sTarget=&sTxt=)>.
- 황정선, 유아라. 2016. 기초학문자료센터(KRM)의 시각화 서비스에 관한 연구. 『한국정보관리학회 학술대회 논문집』, 97-100.
- CENDI. 2017. Implementation of public access programs in federal agencies. [online]. [cited 2019.3.2]. <[https://cendi.gov/projects/Public\\_Access\\_Plans\\_US\\_Fed\\_Agencies.html](https://cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html)>.
- Center for Open Data in the Humanities. 2017. [online]. [cited 2019.3.5]. <<http://codh.rois.ac.jp/seminar/coretrustseal-20171204/>>.
- Cho, J. 2019. "Study of Asian RDR based on re3data." *The Electronic Library*, 37(2): 302-313.
- DataCite. 2012. Business models principles. [online]. [cited 2019.3.5]. <[https://www.datacite.org/documents/Business\\_Models\\_Principles\\_v1.0.pdf](https://www.datacite.org/documents/Business_Models_Principles_v1.0.pdf)>.

- Gutmann, M., K. Schürer, D. Donakowski, and Hilary Beedham. 2006. "The Selection, Appraisal, and Retention of Social Science Data." *Data Science Journal*, 3: 209-221.  
DOI: <http://doi.org/10.2481/dsj.3.209>
- ICSU. 2016 Core Trustworthy Data Repositories Requirements [online]. [cited 2019.3.5].  
<[https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf)>.
- Kim, S. and M. Choi. 2017. "Registry Metadata Quality Assessment by the Example of re3data.org Schema." *International Journal of Knowledge Content Development & Technology*, 7(2): 41-51.
- Kindling, M. et al. 2017. The Landscape of Research Data Repositories in 2015: A re3data Analysis: D-Lib Magazine. 23(3/4). [online]. [cited 2018.11.5].  
<<http://www.dlib.org/dlib/march17/kindling/03kindling.html>>.
- OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris, France: OECD Publishing.
- Pampel, H. et al. 2013. "Making Research Data Repositories Visible: The re3data.org Registry." *PLOS ONE*, 8(11): 1-10. DOI: 10.1371/journal.pone.0078080
- Rasmussen, K. B. and G. Blank. 2007. "The Data Documentation Initiative: A Preservation Standard for Research." *Archival Science*, 7(1): 55-71.
- Van den Eynden, V., L. Corti, M. Woollard, L. Bishop, and L. Horton. 2011. *Managing and sharing data: A Best Practice Guide for Researchers*. [online]. [cited 2018.11.5].  
<<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>>.

• 국문 참고자료의 영어 표기

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

- Cho, J. 2016. "Study about Research Data Citation Based on DCI (Data Citation Index)." *Journal of the Korean Society for Library and Information Science*, 50(1): 189-207.
- Choi, Y. and Y. Choi. 2015. Study on Metadata Standards (DDI\_2 vs DDI\_3). Symposium on the 5th National Statistical Methodology. [online]. [cited 2018.11.5].  
<[http://kostat.go.kr/edu/sri\\_kor\\_new/3/6/5/index.board?bmode=read&aSeq=349108&pageNo=2&rowNum=10&amSeq=&sTarget=&sTxt=](http://kostat.go.kr/edu/sri_kor_new/3/6/5/index.board?bmode=read&aSeq=349108&pageNo=2&rowNum=10&amSeq=&sTarget=&sTxt=)>.
- Ko, B. 2001. "Research on Cyber Shopping Mall using Correspondence Analysis." *The Korean Data Analysis Society*, 3(4): 419-427.

- Ko, M. and G. Bae. 2011. "A Study of the Integrated Operation for Databases with Different Data Structures." *Korea Society for Library and Information Science*, 45(3): 69-85.
- Lee, J., J. Kim, S. Choi, and I. Kim. 2007. "A Study on Design and Analysis of Metadata and Ontology based on Humanities and Social Sciences." *Korea Society for Library and Information Science*, 41(2): 291-316.
- Minamiyama, Y. 2017. "Core Unified Requirement of Reliable Data Repository." *Current Awareness-E*, 320. [cited 2018.11.2]. <<http://current.ndl.go.jp/e1888>>.
- Seoul National University. 2010. Research Ethics Guidelines for Seoul School. [online]. [cited 2018.11.5]. <[http://www.snu.ac.kr/research/images/down/research\\_08.pdf](http://www.snu.ac.kr/research/images/down/research_08.pdf)>.
- Shim, Y., H. Ahn, and J. Byun. 2015. "Strategies for Improving the Collection and Use of Research Data in the Humanities." *Korea Society for Library and Information Science*, 49(3): 155-183.
- Shin, Y. and Y. Chung. 2012. "Study on the Improvement Plans of the Humanities and Social Sciences Research Data Archives in Korea." *Korea Society for Archive and Record Management*, 12(3): 93-115.
- Whang, J. and A. Yoo. 2016. "Study of Korean Research Memory Visualization Services." *Proceedings of Korea Society for Library and Information Management*, 97-100.