

〈공공데이터법〉과 인문데이터

- 공공기관 보유 인문데이터 공개 신청 사례를 중심으로*

김바로**

〈차례〉

1. 서론
2. 웹서비스와 RAWDATA
3. 인문 공공데이터 공개의 제 문제
 - 1) 개인정보보호 관련
 - 2) 데이터 업데이트 관련
 - 3) 저작권 관련
4. 결론

〈국문초록〉

본 논문에서는 공공데이터법에 따른 공공기관의 데이터 공개의 현재를 살펴보기 위하여, 우선 디지털 조선왕조실록을 통해, 공공기관에서 편찬한 인문 데이터의 온전한 활용을 위한 RAWDATA의 의미와 가치를 살펴보고, 공공데이터의 신청 절차를 통한 인문 데이터의 신청 과정에서 발생하는 제 문제를 탐색했다.

한국에서 대다수의 인문 데이터는 정부의 지원 하에 구축 및 운영되고 있다. 그리고 2013년 공공기관이 보유·관리하는 공공데이터를 국민에게 제공하도록 의무화하는 공공데이터법이 제정되었지만, 인문학 영역의 공공데이터는 아직 공개가 미진한 상황이다. 물론, 디지털 조선왕조실록과 같이 웹서비스로 공공데이터의 2차적 가공물을 제공하고는 있으나, 1차적 가공물인 RAWDATA의 공개는 많이 이루어지고 있지 않다. 공공데이터법에 따르면 디지털 환경에서의 새로운 인문학 연구인 디지털인문

* 본 논문에 꼼꼼한 검토와 함께 유익한 도움말을 해주신 익명의 심사자분들께 감사의 말씀을 드립니다. 아울러 본 논문에서 언급한 공공데이터 관련 문건은 <https://osf.io/96f38/>에서 모두 열람 가능합니다.

** 한국학중앙연구원 한국학대학원 인문정보학 교수 / ddokbaro@gmail.com

학 연구를 위해서 연구자에게 1차 사료에 해당하는 RAWDATA를 획득할 권리가 주어진다. 하지만 인문학 데이터를 보유하고 있는 대다수의 인문학 기관들은 공공데이터법에 따라서 주동적으로 공공데이터를 공개하지도 않고, 공공데이터법이 보장하는 국민의 공공데이터 공개 요청에도 개인정보보호, 데이터 업데이트, 저작권 보호 등의 사유를 들어서 인문 데이터를 공개하고 있지 않다.

이에, 학문의 인류 공영적인 이상과 디지털 시대의 새로운 인문학 연구를 위해서 공공데이터법이 보장하는 CSV, XML, RDF 등의 기계가독형 RAWDATA에 대한 공개에 대한 공공기관들의 태도 변화가 필요하다.

공공데이터, 인문학, 인문데이터, 공공기관, 디지털인문학

1. 서론

모든 연구자들은 언제나 자료에 굶주려 있다. 아무리 아름다운 이론도 튼튼한 근거가 없이는 사상누각에 불과하기 때문이다. 그리고 디지털 시대의 자료는 데이터¹⁾이다. 오늘날 그 누구나 온라인을 통해서 “조선왕조실록”의 원문 이미지, 원문 텍스트, 번역 텍스트 및 주석까지 손쉽게 접근할 수 있다. 도서관에서 두꺼운 조선왕조실록 영인본 책장을 넘기는 때로 돌아가기에는 디지털 기술의 발전으로 인한 편리함에 많은 연구자들이 익숙해져 있다. 또한 이제 대부분의 학술지들이 따로 종이 인쇄본을 만들지 않고 있다. 인류 지혜의 끝에서 한 발자국을 더 나아가기 위하여 몸부림치는 인류 지혜의 최전선인 논문조차도 이제는 종이가 아닌 데이터로 편찬되고 있다.

1) 본 논문에서 정의하는 데이터는 DIKW(Data, Information, Knowledge, Wisdom)에서의 데이터(data) 개념이 아니다. 본 논문의 데이터는 국립국어원 표준국어대사전의 “데이터 03(컴퓨터가 처리할 수 있는 문자, 숫자, 소리, 그림 따위의 형태로 된 정보)”의 개념을 토대로 작성되었다.

데이터의 중요성은 데이터를 온전히 편찬하지 못하는 정보공학의 영역에서 더욱 강조되고 있다. 데이터를 다루는 모든 영역의 불문율인 “Garbage-In Garbage-Out(쓰레기를 넣으면, 쓰레기가 나온다)²⁾”은 아무리 뛰어난 알고리즘도 좋은 데이터 없이는 아무 가치가 없음을 말하고 있다. 그렇기에 0부터 9까지의 손글씨 이미지의 모음인 “THE MNIST DATABASE of handwritten digit³⁾”은 4만개가 넘는 논문인용수를 보여주고 있다. 많은 연구는 데이터를 토대로 이루어지나, 데이터 자체를 만드는 연구는 학문적 성과로 높이 평가하지는 않는 경우가 많다.⁴⁾ 특히 인문학에서는 그 동안 자료를 독점하여, 학문적인 힘을 발휘하는 “문헌 권력”이 작동했기에, 인문학 연구의 과정 혹은 결과로서 발생하는 데이터를 공유하지 않았다. 그 결과 디지털 시대에도 대다수의 인문학 연구는 온전히 축적되지 못하고 백지부터 시작하는 경우가 다반사였다.

그럼에도 불구하고, 정부의 지원 하에 인문 데이터의 축적이 지속적으로 진행되었다. 민족문화추진회(현 한국고전번역원)의 1995년 “국역조선왕조실록 CD-ROM” 간행과 드라마 작가, 만화 작가 등의 창작자에 의한 활용을 통해서, 인문 데이터의 가치가 인정받기 시작했다. IMF체계를 벗어나고자 하는 정책의 일환으로 정보통신부에서는 공공 근로 정보화 사업을 추진하여 대량의 인문 데이터가 정부 주도 하에 만들어질 수 있었다.⁵⁾ 그

2) Corrales, David Camilo, Agapito Ledezma, and Juan Carlos Corrales, “A conceptual framework for data quality in knowledge discovery tasks (FDQ-KDT): A Proposal.” *Journal of Computers*, 10(6), 2015, pp.396~405.

3) Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, 86(11), pp.2278~2324, Nov. 1998, doi: 10.1109/5.726791.

4) 김광립, 「데이터역사과학 - 고대 문명의 보편 질문을 향한 관문」, 『東洋學』 79, 단국대학교 동양학연구원, 2020, 3쪽, <http://www.riss.kr/link?id=A106841155>

5) 김현, 『인문정보학의 모색』, 북코리아, 2012, 448~449쪽, <http://www.riss.kr/link?id=M13139452>

리고 현재까지도 대다수 양질의 인문 데이터는 정부 산하 기관들에서 주도적으로 편찬하고 서비스하고 있다.

대한민국 정부는 2013년 3월 정부 3.0의 슬로건 아래에서 <공공정보 민간활용촉진 종합계획>⁶⁾을 수립하고, 2013년 10월에 <공공데이터의 제공 및 이용 활성화에 관한 법률(“공공데이터법”)>⁷⁾을 제정하였다.⁸⁾ 그리고 공공데이터법을 토대로 하는 2022년 현재 대표적인 공공데이터 개방 플랫폼으로는 공공데이터포털⁹⁾과 AIHUB¹⁰⁾ 및 “모두의 말뭉치”¹¹⁾가 있다.

공공데이터법은 원칙적으로 공공기관이 보유·관리하는 공공데이터를 국민에게 제공하도록 하고 있다.¹²⁾ 그리고 대다수의 인문 데이터 편찬 프

6) 문화체육관광부, 대한민국 정책브리핑, “공공정보의 민간활용을 위해 정부가 나선다”, 2010.03.09., <https://www.korea.kr/news/pressReleaseView.do?newsId=155439672>

7) 법제처, 국가법령정보센터, “공공데이터의 제공 및 이용 활성화에 관한 법률[시행 2013. 10. 31.] [법률 제11956호, 2013. 7. 30., 제정]”, [http://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/\(11956,20130730\)](http://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/(11956,20130730))

8) 2013년 제정 이후에 총 4차례의 개정이 있었으나, 2016년 개정에서 공공데이터의 개념의 구체화와 민간 서비스 주도를 지향하는 방향으로의 변경 이외에는 큰 틀에서의 변화는 없다.

9) 행정안전부, “공공데이터포털”: www.data.go.kr

10) NIA 한국지능정보사회진흥원, “AIHUB”: <https://www.aihub.or.kr/>

11) 국립국어원, “모두의 말뭉치”: <https://corpus.korean.go.kr/>

12) 공공데이터 제공에 예외가 되는 특별한 경우는 다음과 같으며, 이를 토대로 하는 실제 인문 데이터 영역에서의 고려 사항은 3장에서 논의하도록 한다.

법제처, 국가법령정보센터, “공공데이터의 제공 및 이용 활성화에 관한 법률[시행 2020. 12. 10.] [법률 제17344호, 2020. 6. 9., 타법개정]”, <https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>

제28조(공공데이터의 제공중단) ① 공공기관의 장은 다음 각 호의 어느 하나에 해당 하는 경우 공공데이터의 제공을 중단할 수 있다.

1. 이용자가 제19조에 따라 공표된 공공데이터의 이용요건을 위반하여 공공기관 본래의 업무수행에 상당한 지장을 초래할 우려가 있는 경우
2. 공공데이터의 이용이 제3자의 권리를 현저하게 침해하는 경우
3. 공공데이터를 범죄 등의 불법행위에 악용하는 경우

로젝트가 정부의 직간접적인 지원을 받고 있기에 공공데이터의 범주에 포함된다고 할 수 있다. 대다수의 인문 데이터가 정부 주도로 편찬되고 있는 현실에서 공공데이터법은 다수의 인문 데이터가 자유롭게 공유될 수 있는 환경의 토대라고도 할 수 있다. 이에, 공공데이터 중에서 인문 데이터의 의미와 가치를 살펴보고, 실제 인문 데이터의 공개 신청 사례를 통해서, 앞으로 인문 데이터가 나아가야 할 방향에 대해서 탐색하고자 한다.

2. 웹서비스와 RAWDATA

공공데이터에서 규정하는 데이터(RAWDATA)는 “광(光) 또는 전자적 방식으로 처리된 자료 또는 정보”이다. 이는 비전자적 정보까지 포괄하는 “정보공개법¹³⁾”상 정보의 범주와는 차이가 있다.¹⁴⁾ “정보공개법”에서는 “전자매체를 비롯한 모든 형태의 매체 등에 기록된 사항”으로 전자적인 방식으로 처리된 자료 또는 정보뿐만 아니라, 도면, 사진, 필름, 테이프, 슬라이드와 같은 비 전자적 방식으로 처리된 자료 또는 정보도 포함하고

4. 그 밖에 공공데이터의 관리 및 이용에 적합하지 아니한 경우로서 제29조에 따른 공공데이터제공분쟁조정위원회가 정하는 경우

② 공공기관의 장은 제20조제1항 각 호에 따른 사유로 행정안전부장관에게 목록 제외 요청을 한 경우 전략위원회의 심의·의결 전에 제공중단 조치를 취할 수 있다. <개정 2014.11.19, 2017.7.26>

③ 제1항에 따라 공공데이터의 제공이 중단된 경우 이용자는 중단사유가 해소된 후 다시 제26조에 따라 제공받을 수 있다.

④ 그 밖에 공공데이터 제공중단의 절차·방법 등 필요한 사항은 대통령령으로 정한다.

13) 법제처, 국가법령정보센터, “공공기관의 정보공개에 관한 법률[시행 2021.12.23.] [법률 제17690호, 2020. 12. 22., 일부개정]”, <https://www.law.go.kr/법령/공공기관의정보공개에관한법률>

14) 권은정, 「공공데이터 영역 리스크 관리에 관한 법적 소고 — 데이터 개방에 따른 개인정보 관련 리스크를 중심으로 —」, 『행정법연구』, 행정법이론실무학회, 2020, 169쪽, <http://www.riss.kr/link?id=A106612515>

있다. 이에 반하여, 공공데이터로서 규정하고 있는 정보는 기계 판독이 가능한 형태인 기계가독형 데이터(Machine-readable Data)이다. “기계 판독이 가능한 형태”란 소프트웨어로 데이터의 개별내용 또는 내부구조를 확인하거나 수정, 변환, 추출 등 가공할 수 있는 상태”를 지칭하며, 공공데이터를 제공한다는 것은 “공공기관이 이용자로 하여금 기계 판독이 가능한 형태의 공공데이터에 접근할 수 있게 하거나 이를 다양한 방식으로 전달하는 것”을 의미한다.

공공데이터법 11조에 근거한 “공공데이터 관리지침¹⁵⁾” 제3조(기본원칙)에서는 공공기관은 누구든지 공공데이터를 편리하게 이용할 수 있도록 제공해야 하며, 보편적인 이용권 확대를 위한 필요한 조치를 취하도록 하고 있으며, 특별한 경우를 제외하면 영리적인 이용인 경우에도 이를 금지하거나 제한하지 못하도록 하고 있다. 또한 공공기관은 가공된 형태가 아닌 RAWDATA로 제공하도록 노력하여야 하며, 이용자가 공공데이터를 효율적으로 이용할 수 있도록 기계 판독이 가능한 형태로 정비하고 오픈포맷으로 제공해야 한다고 명시하고 있다.

기계 판독이 가능한 형태란 자유로운 수정, 변환이 가능한 형태의 데이터를 의미한다. 그렇기에 자유로운 수정, 변환이 불가능한 PDF는 공공데이터의 기준에 미달하는 형태인 1단계의 데이터이다. 또한 설령 기계 판독이 가능한 형태라고 하더라도 모든 소프트웨어에서 자유롭게 활용(수정, 편집 등)을 할 수 있는 형태의 오픈포맷이 아니면 연구 자료 활용에 있어 기존의 실제 문헌 자료와 차별점이 없다. 따라서 특정 소프트웨어에서만 읽기, 수정, 변환이 가능한 HWP, XLS 등의 데이터는 어디까지나 최소충족포맷이며, 모든 소프트웨어에서 읽기, 수정, 변환이 가능한 CSV¹⁶⁾,

15) 법제처, 국가법령정보센터, “공공데이터 관리지침[시행 2021. 10. 26.] [행정안전부고시 제2021-70호, 2021. 10. 26., 전부개정]”, [https://www.law.go.kr/행정규칙/공공데이터관리지침/\(2021-70,20211026\)](https://www.law.go.kr/행정규칙/공공데이터관리지침/(2021-70,20211026))

표 1. 기계 판독이 가능한 형태의 포맷 단계별 구분비교¹⁹⁾

구분	1단계	2단계	3단계	4단계	5단계
기계판독이 가능한 형태	미충족포맷 (포털등록불가)	최소충족포맷	오픈포맷	오픈포맷	오픈포맷
특징	특정 소프트웨어에서 읽을 수만 있는 데이터로 자유로운 수정, 변환 불가	특정 소프트웨어에서 읽고 수정, 변환 가능	모든 소프트웨어에서 읽고 수정, 변환 가능	URI를 기반으로 데이터 속성 특성 관계를 기술하고 있는 데이터 구조	웹상의 다른 데이터와 연결, 공유 가능
예시	PDF	HWP, XLS, JPG, PNG, WMV, MPEG, MP3, SWF	CSV, JSON, XML	RDF	LOD

JSON¹⁷⁾, XML¹⁸⁾ 형식의 데이터가 되어야만 온전한 오픈포맷이라고 할 수 있다. 더 나아가서 URI²⁰⁾를 기반으로 의미를 내재한 RDF²¹⁾와 이를

16) CSV는 Comma Separated Value의 약어로, 콤마를 통해서 각 항목이나 내용을 구분하는 형식을 의미한다. 보다 자세한 사항은 다음 링크를 내용을 참조할 수 있다. [https://ko.wikipedia.org/wiki/CSV_\(파일_형식\)](https://ko.wikipedia.org/wiki/CSV_(파일_형식))

17) JSON은 JavaScript Object Notation의 약어로 데이터를 교환할 때 사용되는 형식이다. 보다 자세한 사항은 다음 링크를 내용을 참조할 수 있다. <https://ko.wikipedia.org/wiki/JSON>

18) XML은 eXtensible Markup Language의 약어로 국제 표준 마크업언어(Markup Language, 문서 처리를 지원하기 위해 문서에 추가되는 정보)다. 보다 자세한 사항은 다음 링크를 내용을 참조할 수 있다. <https://ko.wikipedia.org/wiki/XML>

19) 행정안전부, 공공데이터 관리지침 개정 고시(행정안전부고시 제 2019-71호), 2019. 09.04., https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000016&ntfId=72733

20) URI는 Uniform Resource Identifier의 약자로 인터넷에 있는 자원을 나타내는 유일한 주소이다. 그 하위개념으로 URL과 URN이 있다. 보다 자세한 사항은 다음 링크를 내용을 참조할 수 있다. https://ko.wikipedia.org/wiki/통합_자원_식별자

21) RDF는 Resource Description Framework의 약어로 범용적인 시맨틱 웹 기술 언어이다. 보다 자세한 사항은 다음 링크를 내용을 참조할 수 있다. https://ko.wikipedia.org/wiki/자원_기술_프레임워크

토대로 웹상에서 다른 데이터와 상호 연결하는 LOD²²⁾를 권장하고 있다.

인문학 영역에서도 기계 판독이 가능한 형태의 오픈포맷을 제공한다는 개념은 디지털 시대의 인문학 연구에서 매우 중요한 의미를 갖는다. 온라인을 통해서 접하는 인문학 관련 서비스는 이미 인문 데이터를 토대로 재가공을 거친 “2차 사료”이기 때문이다. 인문학 연구에서 불가피한 경우가 아니면 “2차 사료”를 기피하고, “1차 사료”를 근거로 하는 것은 “1차 사료”가 보다 정확하고 많은 정보를 내재하고 있기 때문이다. 기계 판독이 가능한 형태의 데이터는 컴퓨터가 보다 많은 지식을 이해할 수 있는 형태로 되어 있기에 보다 풍부한 정보를 컴퓨터가 처리할 수 있도록 한다. 이를 2차 사료에 대응하는 디지털 조선왕조실록 웹서비스와 1차 사료에 대응하는 디지털 조선왕조실록 RAWDATA의 컴퓨터 처리를 비교해서 살펴보면 다음과 같다.

디지털 조선왕조실록²³⁾은 국사편찬위원회에서 1995년에 민족문화추진위원회(현 한국고전번역원)에서 간행한 조선왕조실록 CD-ROM을 토대로 현재까지 끊임없는 수정 및 보완을 지속하고 있는 대국민 서비스이다. 디지털 조선왕조실록에서는 원문 이미지, 원문 텍스트, 번역 텍스트를 기반으로, 디렉토리 서비스와 검색 서비스 및 다양한 부가 서비스를 통해서 인문학 연구자들뿐만이 아니라, 콘텐츠 창작자들도 손쉽게 조선왕조실록의 내용을 살펴볼 수 있다. 그런데 디지털 조선왕조실록 RAWDATA는 XML 형식으로 이루어져 있으며, 웹 서비스에서 제공하는 것보다 훨씬 더 많은 정보가 내재되어 있다.

22) LOD는 Linked Open Data의 약어이며, 시맨틱 웹이 표방하는 데이터 웹(Data Web)을 구체적으로 구현하는 방법으로 인터넷 상의 각 사이트에서 RDF(Resource Description Framework) 형식의 데이터를 RESTful 프로토콜을 사용하여 정형화 데이터를 제공하는 방법이다. 보다 자세한 사항은 다음 링크의 내용을 참조할 수 있다. https://ko.wikipedia.org/wiki/링크드_데이터

23) 국사편찬위원회, “조선왕조실록”: <http://sillok.history.go.kr/>

현재 디지털 조선왕조실록에서는 인물, 장소, 관직에 대한 정보를 다양한 색으로 표시해주고 있다. 특히 인물을 클릭하면, 해당 인물의 기본 정보(성명, 신분, 자, 호, 본관, 생년, 몰년, 국가)를 제공할 뿐만 아니라, “상세 정보”를 통해서 조선왕조실록에 해당 인물이 등장하는 모든 실록 기사를 제공하고 있기에, 사용자들은 조선왕조실록에 기록된 인물의 활동을 손쉽게 입체적으로 탐색할 수 있다.

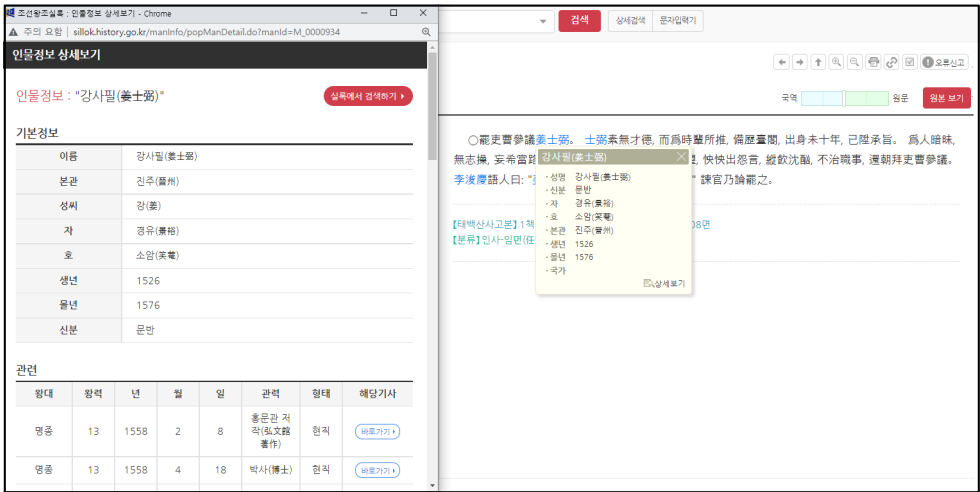


그림 1. 디지털 조선왕조실록의 인물 상세보기 서비스 샘플

그럼에도 불구하고, 온라인 서비스를 통해서 “조선시대에 이조좌랑(吏曹佐郎)을 지낸 사람들이 어떤 과정을 통해서 입사부터 최고관직까지 승진했는가?”와 같은 질문에 대해서 답하기 위해서는 조선 왕조에서 이조좌랑을 지낸 모든 인물을 찾고, 해당 인물의 이력을 하나하나 조사해야 할 것이다. 그에 반하여 디지털 조선왕조실록 RAWDATA에서는 데이터 탐색을 통해서 해당 내용을 손쉽게 처리할 수 있다.²⁴⁾ 다시 말해서 “2차 사료”인 디지털 조선왕조실록 웹서비스에서 구현하지 못한 서비스나 기능을

“1차 사료”인 디지털 조선왕조실록 RAWDATA에서는 처리가 가능하다.

현재 조선왕조실록의 RAWDATA는 공공데이터포털에 “교육부 국사편찬위원회_조선왕조실록 정보_실록원문²⁵⁾”과 “교육부 국사편찬위원회_조선왕조실록 정보_고순종실록 원문²⁶⁾”으로 원문 데이터가 XML 형태로 공개되어 있다. 또한 “교육부 국사편찬위원회_조선왕조실록 정보_부가정보 인물 데이터²⁷⁾”를 통해서 인물 부가정보가 CSV 형태의 데이터로 공개되어 있다. 모든 데이터는 로그인이 필요 없이 자유롭게 다운로드를 할 수 있으며, 출처 표기, 상업적 금지, 변경 금지 등 어떠한 제한도 없다.

디지털 조선왕조실록에서 색상으로 표시한 인명, 지명 등에 INDEX 태그를 부여하고, 인명에는 TYPE 속성을 “이름”으로, 지명에는 TYPE 속성을 “지명”으로 서술한 것을 볼 수 있다. 무엇보다 중요한 것은 개별 INDEX의 NUM 속성은 INDEX의 고유향으로서 지속적으로 증가하지만, REF 속성은 객체의 고유향으로 객체를 식별하는 데 사용하고 있다. 예를 들어서, 강사필(姜士弼)의 고유향은 “M_0000934”로 설정되어 있으며, 해당 텍스트가 정확하게 강사필(姜士弼)로 쓰여 있지 않아도, 다시 말해서 “士弼(사필)”로 적혀 있다고 하더라도, 객체를 강사필(姜士弼)로 식별할 수 있으면, REF 속성에 강사필(姜士弼)의 고유향인 “M_0000934”를 기입하고 있다. 이를 통해서, 컴퓨터는 姜士弼, 士弼 혹은 강사필의 자인 경유(景裕)나 호인 소암(笑菴)으로 서술되어 있다고 하더라도, 주민등록번호와 같이

24) 물론 RAWDATA를 토대로 데이터 탐색을 하기 위해서는 이를 위한 데이터 조작 능력이 필요하다. 따라서 인문학자의 디지털 리터러시 교육에 대한 논의도 필요하지만, 이는 본 발표의 주제 범위를 벗어나기에 본 발표에서는 해당 내용은 생략한다.

25) 공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_실록원문”:

<https://www.data.go.kr/data/15053647/fileData.do>

26) 공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_고순종실록 원문”:

<https://www.data.go.kr/data/15053646/fileData.do>

27) 공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_부가정보 인물 데이터”: <https://www.data.go.kr/data/15053645/fileData.do>

표 2. 실록 원문 웹서비스와 XML 비교 샘플
- 선조수정실록 2권, 선조 1년 2월 1일 辛巳 2번째 기사²⁸⁾

○罷吏曹參議姜士弼。士弼素無才德，而爲時輩所推，備歷臺閣，出身未十年，已陞承旨。爲人暗昧，無志操，妄希當路。及拜忠清道觀察使，乃大失望，怏怏出怨言，縱飲沈酗，不治職事，還朝拜吏曹參議。李浚慶語人曰：“姜士弼不被劾，是，朝無清議也。”諫官乃論罷之。
<pre> <paragraph align="center">○罷吏曹參議(index num="1088411_0" ref="M_0000934" sort="K" type="이름")姜士弼</index>。 <index num="1088412_0" ref="M_0000934" sort="K" type="이름")士弼</index>素無才德，而爲時輩所推，備歷臺閣，出身未十年，已陞承 旨。爲人暗昧，無志操，妄希當路。及拜(index num="1088413_0" sort="K" type="지명")忠清 道</index>觀察使，乃大失望，怏怏出怨言，縱飲沈酗，不治職事，還朝拜吏曹參議。 <index num="1088414_0" ref="M_0005159" sort="K" type="이름")李浚慶</index>語人曰：“<index num="1088415_0" ref="M_0000934" sort="K" type="이름")姜士弼</index>不被劾，是，朝無清 議也。”諫官乃論罷之。</paragraph> </pre>

REF 속성값을 기본으로 강사필이라는 동일 객체로 식별할 수 있다.

또한 “교육부 국사편찬위원회_조선왕조실록 정보_부가정보 인물 데이터”에서는 “교육부 국사편찬위원회_조선왕조실록 정보_실록부가정보_인물.csv” 파일을 통해서, 조선왕조실록에 등장하는 인물 115,861명의 성, 명, 자, 호, 본관, 생몰년 등의 인물 정보 및 한국학중앙연구원 한국역대인물종합정보시스템²⁹⁾의 인물고유ID를 연결해 놓았다. 뿐만만이 아니라, 인물들의 현직 혹은 전직, 관직 등을 정리하였으며, 해당 관직 식별의 근거가 된 조선왕조실록 기사에 고유햄(PARA_ID)을 부여했다.

조선왕조실록에 등장하는 모든 인물들이 식별되어 있으며, 각각의 인물들의 관직 기록이 모두 망라되어 있는 RAWDATA를 통해서, 연구자는 단순히 특정 개인 혹은 10명 이하의 소수 집단에 대한 연구뿐만이 아니라, 조선시대의 모든 인물들의 관직 관련 연구를 수행할 수 있게 된다.³⁰⁾

28) 국사편찬위원회, 조선왕조실록, 선조수정실록 2권, 선조 1년 2월 1일 辛巳 2번째기사, http://sillok.history.go.kr/id/wnb_10102001_002

29) 한국학중앙연구원, “한국역대인물 종합정보시스템”: <http://people.aks.ac.kr/>

표 3. “교육부 국사편찬위원회_조선왕조실록 정보_부가정보 인물 데이터” 데이터 샘플

MAN_ID	KING_YEAR	KING_START	YEAR	MONTH	DAY	OFFICE_LEVEL	TYPE	PARA_ID
M_0000934	21	명종	1566	11	16	청홍도 관찰사 (淸洪道觀察使)	현직	kma_12111016_003
M_0000934	22	명종	1567	6	6	이조 참의 (吏曹參議)	현직	kma_12206006_001
M_0000934	1	선조	1568	2	1	이조 참의 (吏曹參議)	현직	knb_10102001_002
M_0000934	6	선조	1573	4	22	참의(參議)	현직	kna_10604022_001

이는 디지털 조선왕조실록의 RAWDATA가 공공데이터법의 시행을 계기로 공개되었기에 가능한 일이다. 디지털 조선왕조실록의 연구결과물 외에도 인문학 연구 지원을 하고 있는 한국연구재단 토대 사업 등의 최종결과물은 기계가독형 데이터 형태로 제출하도록 되어 있다.³¹⁾ 이러한 법적인 근거와 여러 연구의 RAWDATA 공개 사례에도 불구하고, 아직도 많은 인문 데이터가 공공데이터로서 공개되지 못하고 있다.

3. 인문 공공데이터 공개의 제 문제

공공데이터법은 공공기관이 주동적으로 보유·관리하는 공공데이터를

30) 물론 해당 내용은 어디까지나 조선왕조실록에 등장하는 인물과 관직만을 다루었기에, 엄밀한 의미에서 조선시대의 “모든” 인물이라고 말하기는 어렵다. 하지만 조선왕조실록 RAWDATA를 활용한다면, 전통적인 방법을 적용했을 때 대비 분석 가능 범위가 늘어나는 것은 분명하다. 또한 대상 문헌의 확장에서도 “교육부 국사편찬위원회_조선왕조실록 정보_실록부가정보_인물.csv”에서 한국학중앙연구원 한국역대인물종합정보시스템의 인물고유ID를 활용한 것과 같은 방식을 사용함으로써 인간 개인으로서 축적하여 처리하기 어려운 범위까지도 확장이 가능하다.

31) 한국연구재단에서 데이터 제출로 인정 받는 것은 결과물의 원형(엑셀, 액세스, XML, SQL, MYSQL 등의 RDB)과 멀티미디어 파일 등의 원자료가 포함된 연구결과물을 제출했을 때이다.

국민에게 제공하도록 규정하고 있다. 그러나 현실적으로는 다수의 공공기관이 공공데이터를 주동적으로 개방하지 않기에 공공데이터 제공 신청 절차를 두고 있다. 대한민국 국민이라면 누구나 공공데이터포털을 통해서 공공데이터 제공 신청 절차에 따라 공공데이터를 신청할 수 있다.

공공데이터 제공 신청은 공공데이터포털에 가입 및 로그인하여 공공데이터 제공 신청서를 제출해야 하며, 공공데이터 포털의 가입은 대한민국 국적을 가지고 있는 내국인으로 제한한다.³²⁾ 공공데이터 제공 신청서는 신청인 정보와 신청 내용으로 구분할 수 있다. 신청인 정보는 1) 성명(단체명 및 대표자 성명), 2) 생년월일, 3) 주소(소재지), 4) 사업자(법인, 단체) 등록번호, 5) 전화번호, 6) 이메일 주소이며, 이 중에서 상세주소와 사업자(법인, 단체) 등록번호를 제외하면 모두가 필수 항목이다. 신청 내용은 1) 공공데이터 명, 2) 기관명, 3) 공공데이터 내용, 4) 활용목적(선택)³³⁾, 5) 데이터 활용 목적(자유기술)으로 구성되어 있다.

국사편찬위원회의 조선왕조실록 원문 데이터를 신청하고자 한다면, 공공데이터 명은 “국사편찬위원회 조선왕조실록 원문”으로, 기관명은 “교육부 국사편찬위원회”로 기입한다. 또한 공공데이터의 내용과 데이터 활용 목적은 간략하게만 작성해도 무방하다. 예를 들어서, 조선왕조실록 RAWDATA에 대한 공개 신청을 진행한다면, 공공데이터 내용은 “국사편찬위원회에서 서비스 중인 조선왕조실록(<http://sillok.history.go.kr/>)의 원문 RAWDATA에 대한 정보공개를 요청드립니다”로, 데이터 활용 목적

32) 공공데이터포털에 공개된 데이터 중에서는 로그인에 따라 필요하지 않은 데이터가 대다수이기에, 회원가입을 하지 않더라도 이미 공개된 공공데이터를 다운로드 받아서 사용하는 데에는 큰 문제가 없다. 다만, “실명으로 가입하지 않을 경우 공공데이터 제공신청, 분쟁조정신청 등 일부 서비스에 대해 불이익이 있을 수 있다.” 이는 공공데이터법의 핵심 개념이 “국민에 대한 의무”로서의 공공데이터 제공이기 때문이다.

33) 선택항목은 (1) 앱개발, (2) 창업 및 사업, (3) 학술연구(논문,과제 등), (4) 출판, (5) 기타로 구성되어 있다.

은 “조선왕조실록 원문을 토대로 한 데이터 분석 연구를 수행하고자 합니다.”로 간략하게 기입하면 된다.

신청을 받은 공공기관의 장은 신청인에게 접수에 대해서 고지하고, 접수 부서는 이를 담당부서 또는 소관기관에 이송하여 제공여부를 심의하게 된다. 제공여부의 심의는 “해당 데이터가 비공개 대상정보(「공공기관의 정보공개에 관한 법률」), 「저작권법」 및 그 밖의 다른 법령에서 보호하고 있는 제3자의 권리가 포함된 것으로 해당 법령에 따른 정당한 이용 허락을 받지 아니한 정보에 포함되는지 여부를 검토하는 요청을 받은 날로부터 10일 이내에 제공 여부를 결정하여야 하며, 부득이한 경우 10일의 범위 내에서 연장할 수 있”도록 규정하고 있다.³⁴⁾ 해당 기간 내에 공공기관의 담당자가 신청인에게 구체적인 공공데이터의 범위와 제공 방법 등의 내용을 상의하기 위하여 전화 혹은 메일로 연락을 취하기도 한다.

공공기관의 장은 제공여부 심의를 거쳐 공공데이터의 제공을 결정한 때에는 지체 없이 신청인에게 제공방법·절차 등을 통지하고, 공공데이터포털에 공공데이터 목록을 등록해야 한다. 그리고 신청인이 제공 신청한 공공데이터를 제공 거부하기로 결정한 경우, 공공기관의 장은 지체 없이 거부 결정의 내용과 사유를 신청인과 행정안전부장관에게 통보해야 한다. 다만, 처리 과정에서 완전한 제공 거부가 아닌 부분 제공이 이루어지기도 한다.

신청인은 공공기관의 공공데이터 제공거부결정에 대하여 불복이 있을 때에는 제공거부 결정 통지를 받은 날로부터 60일 이내에 공공데이터제공분쟁조정위원회(www.odmc.or.kr) 또는 공공데이터 포털에서 분쟁조정을 신청할 수 있다. 분쟁조정위원회는 분쟁 사건 관련 자료를 수집하고 사건에 대한 사실 조사를 실시하여, 이를 토대로 사건을 위원회에 회부한다.

34) 10일은 공휴일을 제외한 근무일을 기준으로 하며, 일반적으로 해당 기관에서 기간을 연장하여 20일이 소요되기에, 실제로는 신청부터 제공까지 약 1달의 시간이 소요되는 경우가 많다.

공공데이터제공분쟁조정위원회는 정식 조정 절차에 들어가기 앞서 사건 당사자들 간의 자율적인 노력으로 원만히 분쟁이 해결될 수 있도록 합의를 권고하고, 상호간의 합의가 성립하면 사건이 종결된다. 만약 조정 전 합의가 이루어지지 않으면 위원회를 통해 조정절차가 개시된다. 조정절차가 진행되면 당사자의 의견 청취, 증거수집, 전문가의 자문 등 필요한 절차를 거쳐 쌍방에게 합당한 조정안을 제시하고 이를 받아들일 것을 권고하며, 이 경우 사건의 신청자나 상대방은 위원회의 회의에 참석하여 자신의 의견을 개진할 수 있다. 공공데이터제공분쟁조정위원회의 조정을 통하여 내려진 결정에 대하여 조정결정일로부터 15일 이내에 신청인과 상대방이 이를 수락한 경우에는 조정이 성립된다. 당사자가 위원회의 조정안을 수락하고자 하는 경우 위원회가 송부한 조정서에 기명 날인하여 위원회에 제출한다. 양 당사자가 모두 조정안을 수락하면 조정이 성립되어 조정서가 작성되고 조정절차가 종료된다. 만약 당사자 중 일방이 조정안을 수락하지 않을 경우 소송을 제기하거나 포기할 수 있다.

공공데이터의 제공 불가 사유는 국가안전보장부터 제3자권리침해까지 비교적 광범위 하지만, 인문학 영역의 공공데이터에서의 제공 불가 사유는 크게 1) 개인정보보호, 2) 데이터 업데이트, 3) 저작권보호가 있다.

1) 개인정보보호 관련

인문 데이터에서는 개인정보보호가 사실상 제공 불가 사유로 인정되지 않는다. 개인정보 보호법 제2조(정의) 제1항³⁵⁾에서는 “개인정보란 살아있는 개인에 관한 정보”로 대상이 생존해있는 경우로 한정하고 있다. 따라서, 대다수의 근대 이전의 인물에 대한 개인 정보는 개인정보 보호법의 적용 대상이 아니다. 설령 현재 생존해 있는 인물이라도 공적인 존재에 대해서

35) 법제처, 국가법령정보센터, “개인정보 보호법[시행 2020. 8. 5.] [법률 제16930호, 2020. 2. 4., 일부개정]”, <https://www.law.go.kr/법령/개인정보보호법>

는 공공성이 있는 정보로서 “알권리”를 비롯한 법적 이익이 그와 같은 정보처리를 막음으로써 얻을 수 있는 정보주체의 인격적 법익에 비하여 우월하다.³⁶⁾ 비록 공인의 범위에 대한 논란이 없는 것이 아니나³⁷⁾, 인문학의 연구 대상의 대다수가 보편적으로 공인(公人)으로 인정하는 인물이라는 점을 고려하면, 개인정보보호를 근거로 정보 제공을 하지 않는 것은 타당하지 않다. 특히 해당 개인정보가 해당 기관의 웹서비스로 제공되고 있다면 더욱더 제공 불가 사유가 되지 못한다.

이에 따라, 국사편찬위원회의 “일제감시대상 인물카드”³⁸⁾는 제공불가 대상이 되지 못한다. 일제감시대상 인물카드는 일제 경찰에서 제작한 4,858명에 대한 출생일, 출생지, 주소지, 신장 등의 기본 신상 정보와 각종 활동 기록 그리고 검거 기록을 입력한 인물 신상 카드 6,264건의 데이터이다. 근대 역사인물 관련 온톨로지 설계 및 데이터 편찬을 위해서, 2016년 4월 25일 공공데이터 제공 신청을 하였고, 2016년 5월 9일 국사편찬위원회에서는 일제감시대상 인물카드가 “개인정보를 포함”하고 있기에 “제공불가” 입장을 보였다.³⁹⁾ 일제감시대상 인물카드의 인물 모두에 대한 현재 생존 여부는 명확하지 않기에 개인정보보호 원칙에 위배될 소지가 있을 수도 있다. 하지만 일제감시대상 인물카드는 국사편찬위원회에서 온라인으로 서비스를 하고 있기에, 만약 개인정보보호를 사유로 공공데이터로 제

36) 법제처, 국가법령정보센터, 부당이득금반환(공개된 개인정보를 수집하여 제3자에게 제공한 행위에 대하여 개인정보자기결정권의 침해를 이유로 위자료를 구하는 사건) [대법원 2016. 8. 17., 선고, 2014다235080, 판결], <https://law.go.kr/precInfoP.do?precSeq=183259>

37) 심석태, 「공인 개념의 현실적 의의와 범위에 대한 고찰 = 법무부 ‘수사공보준칙’에 나타난 공인 개념을 중심으로」, 『언론과 법』, 한국언론법학회, 2011, <http://www.riss.kr/link?id=A60302545>

38) 국사편찬위원회, 한국사데이터베이스, “일제감시대상인물카드”, <http://db.history.go.kr/item/level.do?itemId=ia>

39) 국사편찬위원회, 공공데이터 부분제공 결정통지서(2016-33), <https://osf.io/wtdmv/>

공불가하다면, 온라인 서비스도 개인정보보호 위반이 되어 버린다. 따라서, 개인정보보호는 국사편찬위원회의 “일제감시대상 인물카드”의 공공데이터 제공 불가의 정당한 사유로 인정받지 못했다. 이에, 2015년 10월 5일 공공데이터 포털을 통해서 해당 데이터를 공개하였다.⁴⁰⁾ 그리고 “일제감시대상 인물카드” 데이터는 박사논문을 기반으로 출판한 “시맨틱 데이터 아카이브의 구축과 활용 : 제도와 인사의 관계에 대한 근대(1895~1910) 학교 자료를 중심으로”⁴¹⁾에서 관보의 인사 관련 데이터, 직원록 데이터 등과 상호 연계하는 구한말 교원 임용 데이터에서 활용할 수 있었다.

2) 데이터 업데이트 관련

방대한 데이터를 오랜 시간 동안 쌓아가며, 끊임없이 오탈자를 비롯한 다양한 오류를 수정 및 보완하는 인문 데이터의 특성상 현재 개발 중이라는 사유로 제공 불가 사유를 제시하는 경우가 있다. 하지만 이는 공공데이터법 제26조(공공데이터의 제공) 제2항에서 명시한 “공공기관의 장은 해당 기관이 개발·제공하고 있거나 개발 예정인 서비스에 관련 공공데이터가 포함되어 있다는 사유로 공공데이터의 제공을 거부하여서는 아니 된다.”에 따라서, 이는 공공데이터법상 정당한 제공거부 사유에 해당하지 않는다.

실제 국사편찬위원회의 한국근대지리정보⁴²⁾는 서비스 구축 중이라는 사유로 제공불가 대상이 되지 못했다. 한국근대지리정보는 삼일운동 GIS 맵을 사용하여, 교육, 교통, 군사, 사법, 금융 등 다양한 시설의 공간

40) 공공데이터포털, “교육부 국사편찬위원회_한국사데이터베이스 정보_일제감시대상 인물카드”, <https://www.data.go.kr/data/15053627/fileData.do>

41) 김바로, 『시맨틱 데이터 아카이브의 구축과 활용: 제도와 인사의 관계에 대한 근대(1895~1910) 학교 자료를 중심으로』, 북코리아, 2018, 1~409쪽, <http://www.riss.kr/link?id=M14988324>

42) 국사편찬위원회, 한국사데이터베이스, “한국근대지리정보”, http://db.history.go.kr/hgis/mod_g1/main.do

정보를 편찬한 서비스이다. 근대 인물과 공간 정보의 연계를 위하여, 2020년 4월 27일 공공데이터 제공 신청을 하였고, 2020년 5월 13일 국사편찬위원회에서는 현재 시범 서비스로 데이터 정비 중이라는 이유로 제공 신청을 반려하였다.⁴³⁾ 하지만 2020년 8월 12일 분쟁 조정 결정⁴⁴⁾에서는 국사편찬위원회에서 이미 웹페이지를 통해서 이를 공개하고 있기에 데이터 정비를 정당한 제공거부사유로 여기기 어렵다고 보았다. 다만, 해당 데이터를 활용한 결과물을 공표할 때에는 데이터 제공시기 및 데이터의 오류가능성 등을 표시하도록 하였다. 이에, 2020년 8월 18일에 이메일 전송의 형태로 해당 데이터를 수령했으며, 현재는 공공데이터 포털에도 등록되어 있다.⁴⁵⁾

3) 저작권 관련

저작권 부분은 공공데이터 공개에서 가장 큰 분쟁의 사유이다. 현재 공공데이터는 “제3자의 권리를 현저하게 침해하는 경우” 공공데이터를 제공하지 않을 수 있으며, 해당 내용의 핵심 관계 법령이 저작권법이다. 저작권법 제39조(보호기간의 원칙)에 따라서 저작권은 저작자가 생존하는 동안과 사망한 후 70년간 존속⁴⁶⁾한다. 공공기관이 제공하는 대다수의 서비스는 국비를 토대로 편찬되기에 공공데이터로 공개할 수 있는 저작권 계약이 되어 있는 경우가 많다. 그러나 2012년 공공데이터법이 제정되기 이전에는 공공데이터로서 제공 관련 사항이 명문화되어 있지 않아서 문제가 되는

43) 국사편찬위원회, 한국근대지리정보 제공신청 반려, <https://osf.io/2394z/>

44) 공공데이터제공분쟁조정위원회, 공공데이터제공분쟁조정사건(2020-015), <https://osf.io/89aw3/>

45) 공공데이터포털, “교육부 국사편찬위원회_근대지리정보”, <https://www.data.go.kr/data/15080850/fileData.do>

46) 저작권법 제5조(2차적저작물)에 따라, 조선시대 고문헌에 대한 영인본과 같이 저자 사망 후 70년이 경과하였지만 번역·편곡·변형·각색·영상제작 등의 방법으로 작성한 창작물은 2차적 저작물로서 독자적인 저작물로 보호된다.

경우도 분명 존재하며, 2012년 이후라고 하더라도 다양한 사정에 의해서 저작권을 온전히 확보하지 못하는 경우도 생긴다.

2009년에 DB로 구축된 한국고전번역원의 한국문집총간 편목색인⁴⁷⁾은 저작권을 이유로 하는 공공데이터 제공불가 대상이 되지 않는다. 한국문집총간 편목색인은 「한국문집총간(韓國文集叢刊)」 제1집~제350집과 「한국문집총간속(韓國文集叢刊續)」 제1집~제150집에서 전기적(傳記的) 내용의 편목(篇目)을 선별하여 피전자(被傳者)별로 모아 배열함으로써 한국문집총간을 보다 효율적으로 활용할 수 있도록 기획된 것⁴⁸⁾이다.

역사인물들의 사회적인 관계를 연구하기 위해서, 본 연구자는 2018년 3월 17일 공공데이터 제공 신청을 하였고, 2018년 3월 30일에 한국고전번역원에서는 1) 공공데이터포털에 등록하지 않은 자료로 현재는 제공 불가능한 자료이며, 2) 2차적 저작물로 사업 참여자의 저작권 보호를 위해서 현재는 제공 불가하다고 제공 거부 사유를 밝혔다.⁴⁹⁾ 이에, 공공데이터 분쟁조정위원회에서 조정 절차를 밟게 되었다. 2018년 4월 30일에 공공데이터제공분쟁조정위원회에 출석하여 양측의 입장을 다시 한 번 밝혔고, 그 결과 2018년 5월 11일에 조정결정을 받게 되었다. 조정 결정에서는 한국문집총간 편목색인이 1) 공공데이터 포털에 등록되지 않은 사유는 공공데이터법에 따라서 적합한 거부 사유가 아니라고 보았고, 해당 2) 저작권에 대해서는 한국고전번역원의 제출자료 및 진술, 위원회의 사실조사결과를 종합⁵⁰⁾하여, 한국고전번역원이 해당 데이터 관련 사업참여자로부터 저작권

47) 한국고전번역원, “한국문집총간 편목색인”, <http://index.itkc.or.kr/>

48) 한국고전번역원, 한국문집총간 편목색인, 범례, <http://index.itkc.or.kr/legendPop?mode=legend>

49) 공공데이터포털, 한국고전번역원 한국문집총간 편목색인 제공신청반려, <https://osf.io/wtdmv/>

50) 유념할 부분은 저작권 관련 사항에 대한 증명의 의무는 신청자에게 있지 않고 피신청자에게 있다는 점이다. 다시 말해서, 공공기관은 자신이 서비스하고 있는 데이터에

을 양도 받았음을 확인하였기에, 데이터에 대한 제공을 권고하였다. 다만, 사업참여자로부터 2차적저작물작성권을 양도받지 않았고, 이용 허락을 받은 사실도 없었으므로 연구목적으로는 해당 데이터를 활용할 수 있지만, 2차 저작물은 작성할 수 없다고 밝혔다.⁵¹⁾ 이에, 2018년 6월 22일 한국문집총간 편목색인 데이터를 공공데이터 포털이 아닌 연구자의 개인 메일로 제공받을 수 있었다. 그리고 “한국문집총간 편목색인” 데이터는 조선 후기 여성 전기문학을 통한 인물 관계 데이터 모델을 연구한 석사학위논문인 “조선 후기 여성 전기문학을 통한 인물 관계 데이터 모델 연구: 『한국문집총간』 수록 여성 전기문학을 중심으로”⁵²⁾에서 토대 데이터로 활용되었다.

그런데 공공데이터임에도 불구하고 저작권으로 인하여 사용에 제한이 발생하는 경우가 분명히 존재하고 있다. 국립국어원 “모두의 말뭉치”는 2020년부터 4차 산업혁명 시대의 인공지능 개발의 기반 확보를 목적으로 편찬한 말뭉치들의 공개플랫폼이다.⁵³⁾ “모두의 말뭉치”에서 제공하는 말뭉치는 국어와 언어 정보 처리 분야에 응용하기 위한 연구 및 기술 개발용으로 활용할 수 있으며, 승인된 이용 목적에 한하여 말뭉치를 분석하거나 처리하여 사용할 수 있다. 하지만 대다수 말뭉치의 출처가 신문, 현대 문학 등의 저작권이 설정되어 있는 텍스트이기에, 말뭉치를 활용한 결과물을 외부에 공개(복제, 전송, 배포, 기타 방법 등)하고자 할 때에는 그 결과물에 말뭉치와 말뭉치가 포함된 언어 자료의 원문 일부 또는 전부가 포함되어서

대한 저작권 관련 사항을 제출하고 공공데이터로서 공개를 하지 않을 사유를 명확하게 제시할 수 있어야 한다.

51) 공공데이터제공분쟁조정위원회, 공공데이터제공분쟁조정사건(2018-011), <https://osf.io/etfkg/>

52) 김선미, “조선 후기 여성 전기문학을 통한 인물 관계 데이터 모델 연구: 『한국문집총간』 수록 여성 전기문학을 중심으로”, 한국학중앙연구원 석사학위논문, 2021, 1~87쪽, <http://www.riss.kr/link?id=T15804059>

53) 유연석, “국립국어원, 중단됐던 ‘국가 말뭉치 구축사업’ 10년 만에 재개”, 노컷뉴스, 2018.12.06., <https://www.nocutnews.co.kr/news/5072089>

는 안 되며, 공개 전 국립국어원의 사전 승인을 받아야 한다.⁵⁴⁾

이에 따라 “모두의 말뭉치”의 말뭉치 데이터를 통한 데이터 분석 혹은 인공지능 학습 결과를 공개할 수는 있으나, 실제 데이터 분석 혹은 인공지능 학습을 위해서 수행하는 오탈자 수정을 포함하는 데이터 전처리 결과물은 공개가 불가능하게 된다. 따라서 연구자에 의해서 수정 보완된 RAWDATA를 공개할 수 없게 되기에, 디지털 분석 방법에서 보장되는 연구 재현성 확보가 어려워지며, 서로 다른 연구자들이 RAWDATA 상의 문제를 반복적으로 수정해야 하는 연구 성과 축적 측면에서의 아쉬움이 있다.

4. 결론

본 논문에서는 공공데이터법에 따른 공공기관의 데이터 공개의 현재를 살펴보기 위하여, 우선 디지털 조선왕조실록을 통해, 공공기관에서 편찬한 인문 데이터의 온전한 활용을 위한 RAWDATA의 의미와 가치를 살펴보고, 공공데이터 신청 절차를 통한 인문 데이터 신청 과정에서 발생하는 제 문제를 탐색했다.

중이를 중심으로 하는 전통적인 인문학 자료는 문화유산으로서 소중히 보존해야 하며, 인문학 연구에서 원본 자료에 대한 탐색을 게을리해서는 안 될 것이다. 하지만 디지털 환경을 통해서 수행 가능한 새로운 탐색 방법이 있다면, 이를 무시하고 방기하는 것도 연구자로서 온당하지 않은 자세라고 생각한다. 그리고 디지털 환경에서의 새로운 연구를 수행하고자 하는 디지털 인문학에서는 웹서비스를 통한 2차적인 데이터가 아닌 CSV, XML, RDF 등의 기계가독형 RAWDATA가 반드시 필요하다. 한국에서

54) 문화체육관광부 국립국어원, 모두의 말뭉치, 자주 찾는 질문, “말뭉치를 상업적인 용도로 사용할 수 있나요?”, <https://corpus.korean.go.kr/boards/faqList.do>

는 정부의 지원 하에 수많은 인문 데이터가 축적되고 있는데, 그 성과가 온전히 국민에게 환원되지 못하여 결과적으로 인문학의 이상과 가치가 훼손되고 있는 실정이다.

그럼에도 불구하고 대다수의 인문학 데이터를 보유하고 있는 공공기관들은 공공데이터법에 따라 주동적으로 RAWDATA를 공개하고 있지도 않으며, 공공데이터법에 의거한 데이터 제공 신청에서도 개인정보보호, 데이터 업데이트, 저작권 보호 등의 사유로 RAWDATA의 공개를 지체하고 있다. 이는 인류공영의 학문적 이상을 거론할 필요도 없이, 공공데이터법이 규정한 데이터 공개에서조차 소극적인 모습을 보이는 것이기에, 공공기관들은 자신의 데이터가 기관의 것이 아닌 국민의 것이라는 점을 분명히 인지하고, 국민에게 적극적으로 데이터를 공개해야 할 것이다. 다만, 공공기관의 공공데이터에 대한 태도 변화 이전까지는 개인 연구자들 스스로가 공공데이터법이 보장하는 공공데이터 공개 신청 절차를 통해 연구에 필요한 공공데이터를 확보할 수밖에 없을 것으로 보인다.

다만 인문학의 미래를 생각하면 공공기관뿐만이 아니라, 연구자 개개인도 인문 데이터를 주동적으로 공유하여야만, 디지털 시대의 인문학 연구의 양적·질적 변화가 가능해질 것이며, 인문학이 그 가치를 새로운 시대에 온전히 인정받을 수 있을 것이라고 생각한다. 그러나 현재 인문 데이터를 편찬한 연구자의 노력에 대한 정당한 평가가 이루어지고 있다고 하기는 어렵다. 이에, 후속 연구에서는 논문수를 중심으로 하는 현재의 연구 업적 평가 시스템부터 시작하여, 인문 데이터의 편찬에 대한 평가 시스템 구축까지 인문 데이터 편찬을 중심으로 하는 인문학 연구 평가 시스템 전반에 대한 개혁 방안에 대해서 다루어 보고자 한다. 또한 디지털인문학 연구의 현장에서 RAWDATA의 공유를 통해서 확보되는 연구 재현성과 연구 축적에 관한 논의에 대해서도 별도의 논문을 통해서 따로 다룰 예정이다.

참고문헌

【논저】

- 권은정, 「공공데이터 영역 리스크 관리에 관한 법적 소고 - 데이터 개방에 따른 개인정보 관련 리스크를 중심으로 -」, 『행정법연구』, 행정법이론실무학회, 2020, 165~190쪽, <http://www.riss.kr/link?id=A106612515>
- 김광림, 「데이터역사과학 - 고대 문명의 보편 질문을 향한 관문」, 『東洋學』 79, 단국대학교 동양학연구원, 2020, 191~219쪽, <http://www.riss.kr/link?id=A106841155>
- 김바로, 『시맨틱 데이터 아카이브의 구축과 활용 : 제도와 인사의 관계에 대한 근대(1895~1910) 학교 자료를 중심으로』, 북코리아, 2018, 1~409쪽, <http://www.riss.kr/link?id=M14988324>
- 김선미, “조선 후기 여성 전기문학을 통한 인물 관계 데이터 모델 연구 : 『한국문집총간』 수록 여성 전기문학을 중심으로”, 한국학중앙연구원 석사학위논문, 2021, 1~87쪽, <http://www.riss.kr/link?id=T15804059>
- 김 현, 『인문정보학의 모색』, 북코리아, 2012, 1~829쪽, <http://www.riss.kr/link?id=M13139452>
- 심석태, 「공인 개념의 현실적 의의와 범위에 대한 고찰 = 법무부 ‘수사공보준칙’에 나타난 공인 개념을 중심으로」, 『언론과 법』, 한국언론법학회, 2011, 207~236쪽, <http://www.riss.kr/link?id=A60302545>
- Corrales, David Camilo, Agapito Ledezma, and Juan Carlos Corrales, “A conceptual framework for data quality in knowledge discovery tasks (FDQ-KDT): A Proposal.” *Journal of Computers*, 10(6), 2015, pp.396~405.
- Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, 86(11), pp.2278~2324, Nov. 1998, <https://doi.org/10.1109/5.726791>

【법률, 신문, 공보물】

문화체육관광부 국립국어원, 모두의 말뭉치, 자주 찾는 질문, 말뭉치를 상업적인 용도로 사용할 수 있나요?

<https://corpus.korean.go.kr/boards/faqList.do>

문화체육관광부, 대한민국 정책브리핑, “공공정보의 민간활용을 위해 정부가

www.kci.go.kr

- 나선다”, 2010.03.09.,
<https://www.korea.kr/news/pressReleaseView.do?newsId=155439672>
- 법제처, 국가법령정보센터, “개인정보 보호법[시행 2020. 8. 5.] [법률 제16930호, 2020. 2. 4., 일부개정]”, <https://www.law.go.kr/법령/개인정보보호법>
- 법제처, 국가법령정보센터, “공공기관의 정보공개에 관한 법률[시행 2021. 12. 23.] [법률 제17690호, 2020. 12. 22., 일부개정]”, <https://www.law.go.kr/법령/공공기관의정보공개에관한법률>
- 법제처, 국가법령정보센터, “공공데이터 관리지침[시행 2021. 10. 26.] [행정안전부고시 제2021-70호, 2021. 10. 26., 전부개정]”, [https://www.law.go.kr/행정규칙/공공데이터관리지침/\(2021-70,20211026\)](https://www.law.go.kr/행정규칙/공공데이터관리지침/(2021-70,20211026))
- 법제처, 국가법령정보센터, “공공데이터의 제공 및 이용 활성화에 관한 법률[시행 2013. 10. 31.] [법률 제11956호, 2013. 7. 30., 제정]”,
[http://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/\(11956,20130730\)](http://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/(11956,20130730))
- 법제처, 국가법령정보센터, “공공데이터의 제공 및 이용 활성화에 관한 법률[시행 2020. 12. 10.] [법률 제17344호, 2020. 6. 9., 타법개정]”,
<https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>
- 법제처, 국가법령정보센터, 부당이득금반환(공개된 개인정보를 수집하여 제3자에게 제공한 행위에 대하여 개인정보자기결정권의 침해를 이유로 위자료를 구하는 사건)[대법원 2016. 8. 17., 선고, 2014다235080, 판결],
<https://law.go.kr/precInfoP.do?precSeq=183259>
- 유연석, “국립국어원, 중단됐던 ‘국가 말뭉치 구축사업’ 10년 만에 재개, 노컷뉴스, 2018.12.06., <https://www.nocutnews.co.kr/news/5072089>
- 한국고전번역원, 한국문집총간 편목색인, 법령,
<http://index.itkc.or.kr/legendPop?mode=legend>
- 행정안전부, 공공데이터 관리지침 개정 고시(행정안전부고시 제 2019-71호), 2019.09.04.,
https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_00000000016&ntId=72733

【웹서비스】

- 국립국어원, “모두의 말뭉치”: <https://corpus.korean.go.kr/>
국사편찬위원회, “조선왕조실록”: <http://sillok.history.go.kr/>

국사편찬위원회, 조선왕조실록, 선조수정실록 2권, 선조 1년 2월 1일 辛巳 2번
제기사, http://sillok.history.go.kr/id/wnb_10102001_002
국사편찬위원회, 한국사데이터베이스, “일제감시대상인물카드”:
<http://db.history.go.kr/item/level.do?itemId=ia>
국사편찬위원회, 한국사데이터베이스, “한국근대지리정보”:
http://db.history.go.kr/hgis/mod_g1/main.do
한국고전번역원, 한국고전종합DB, “한국문집총간 편목색인”:
<http://index.itkc.or.kr/>
한국지능정보사회진흥원, “AIHUB”: <https://www.aihub.or.kr/>
한국학중앙연구원, “한국역대인물 종합정보시스템”: <http://people.aks.ac.kr/>
행정안전부, “공공데이터포털”: www.data.go.kr

[RAWDATA]

공공데이터제공분쟁조정위원회, 공공데이터제공분쟁조정사건(2018-011),
<https://osf.io/etfkg/>
공공데이터제공분쟁조정위원회, 공공데이터제공분쟁조정사건(2020-015),
<https://osf.io/89aw3/>
공공데이터포털, “교육부 국사편찬위원회_근대지리정보”:
<https://www.data.go.kr/data/15080850/fileData.do>
공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_고순종실록 원
문”: <https://www.data.go.kr/data/15053646/fileData.do>
공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_부가정보 인물
데이터”: <https://www.data.go.kr/data/15053645/fileData.do>
공공데이터포털, “교육부 국사편찬위원회_조선왕조실록 정보_실록원문”:
<https://www.data.go.kr/data/15053647/fileData.do>
공공데이터포털, “교육부 국사편찬위원회_한국사데이터베이스 정보_일제감시
대상인물카드”: <https://www.data.go.kr/data/15053627/fileData.do>
공공데이터포털, 한국고전번역원 한국문집총간 편목색인 제공신청반려,
<https://osf.io/xky82/>
국사편찬위원회, 공공데이터 부분제공 결정통지서(2016-33),
<https://osf.io/wtdmv/>
국사편찬위원회, 한국근대지리정보 제공신청 반려, <https://osf.io/2394z/>

ABSTRACT

Public Data Act and humanities data

- Applying for open to humanities data held by public institutions

Kim, Ba-ro

To examine the present state of data disclosure by public institutions under the Public Data Act, I first analyzed the humanities data compiled by public institutions and explored problems in requesting open to them.

In Korea, humanities data are mostly stored and managed with government support. In 2013, the Public Data Act was enacted, mandating that public institutions provide public open to the public data they hold and manage. However, public data in the humanities sector frequently have yet to be disclosed. Although secondary workpieces of public data are provided as a web service, such as the Annals of the Digital Joseon Dynasty, it is necessary to obtain RAWDATA corresponding to the primary feed for digital humanities research, a new form of study of the humanities in the digital environment. However, most humanities institutions that store humanities data do not disclose public data under the Public Data Act, citing reasons such as personal information protection, data updates, and copyright protection despite public data disclosure requests.

Therefore, it is necessary to change public institutions' attitudes toward the disclosure of machine-only RAWDATA such as CSV, XML, and RDF guaranteed by the Public Data Act for the study of the humanities in the digital era.

Key Words

Public data, humanities, humanities data, public institution, digital humanities

논문투고일 : 2022.03.28.
심사완료일 : 2022.04.28.
게재확정일 : 2022.05.02.