

한국학 데이터 표준화 방안과 공공데이터 정책

서동신

한국학중앙연구원 부전문위원, 철학·인문정보학 전공

oriental.neo@gmail.com

I. 머리말

II. 데이터 표준화 체계와 한국학

III. 공공데이터의 적극적 표준화

IV. 맷음말: 상보적 역할의 지평

I. 머리말

디지털 인문학¹ 분야에서 정보화 사업을 담당해 보면 ‘표준화’라는 요건을 심심찮게 접할 수 있다. 그래서 질문한다.

선생님께서 관여하시는 사업에서는 데이터 표준화가 잘 진행되었습니까?

필자를 포함하여 대부분의 연구자나 담당자는 흠칫하며 소극적인 답변을 하게 될 것이다. 그런데 간혹 “네.” 하며 서슴없이 대답하는 경우도 있다.² 자신이 관여하고 있는 연구 과제 및 사업에서 ‘데이터 표준화’가 잘 이루어졌다고 생각하는 디지털 인문학 연구자나 담당자들 중 상당수는 실상 데이터 와 표준화에 대해 충분히 이해하지 않고서 정보화 사업의 요건으로 데이터 표준화가 요구되니까 용역업체에 표준화를 요구하고, 업체에서 표준화를 수행했다고 이야기하면 그냥 그렇게 표준화가 잘 이루어졌다고 생각하는 경우가 대부분이다. 하지만 속사정을 들여다보면, 표준화를 수행하되 최소한의 구색 맞추기 수준으로만 진행하고서 ‘요구사항을 이행’했다고 하고 이것을 받아들여 검수를 완료하는 경우가 많다.³

※ 이 글의 의의를 깊이 인정해 주시고 소중한 도움 말씀을 주신 익명의 심사위원분들께 감사의 말씀을 드린다. 아울러 이 글은 긴 호흡으로 수행할 ‘한국학 데이터 표준화 방안’ 연구의 일환으로서 필자의 후속 연구에서도 활용될 예정이다.

- 1 디지털 인문학의 개념 및 사례와 관련한 참고 정보는 제Ⅱ장 제2절(한국학 정보화의 터 다지기) 도입부의 각주 19, 각주 20에서 소개한다.
- 2 여기서의 대답이라는 것이 꼭 “선생님께서 관여하시는 사업에서는 데이터 표준화가 잘 진행되었습니까?”라는 질문을 많은 분들에게 일일이 한 뒤에 얻은 답변을 의미하는 것은 아니다. 정보화 사업에 관계된 분들과 이루어진 다양한 회의나 대화 속에서 그와 같은 맥락으로 의견을 피력한 경우를 집약해서 표현한 것이다.
- 3 사실 이것이 업계 실무에서는 당연한 것으로 인식되는 경향도 있지만, 빅데이터와 인공지능

인문학 데이터를 다루는 정보화 사업이나 그 데이터의 디지털화·지능화를 두고 고민해 본 경험이 있다면,⁴ 표준화의 필요성은 당연한 것이면서도⁵ 한편으로는 그것을 어떻게, 어느 수준까지 진행할 수 있을까 하는 것이 미제(謎題)로 남을 것이다.

학술적 지식에 대한 식견이 높은 인문학자들이 정보화 사업을 수행하는 IT

의 시대에 이러한 소극적 대처는 자칫 위험한 결과를 낳을 수도 있음을 간과할 수 없다. 후술 하겠지만 표준화 과정은 그리 간단하지도 않거니와 법정부 차원의 표준화 추진 여정도 힘겨웠고 현재 진행 중인 정책도 점점 더 확대되고 구체화되고 있다. 2019년 OECD 발표에 따르면, 한국은 공공데이터 개방 평가 조사 대상 34개국 중 1위를 차지함으로써 2015년 이래 3회 연속 세계 1위라는 기록을 세웠지만, 기실 행정안전부·한국지능정보사회진흥원은 각종 기관 업무와 시스템에서 표준화되지 않은 데이터로 인해 발생하는 정보 전달과 공유 문제 개선의 필요성을 강조한다. 조윤희·이용건, 「공공데이터 주요 성과와 제4차('23~'25) 기본계획 추진방향」, 『한국행정학회 학술발표논문집』2022-2(2022), 798쪽; 행정안전부(편), 『공공데이터베이스 표준화 관리 매뉴얼』(대구: 한국지능정보사회진흥원, 2023d), 2쪽, 6~7쪽, 30~31쪽에서 이러한 입장을 확인할 수 있다. 물론 근년 들어 이러한 추세에 급격한 변화가 일고 있기에, 긴 세월 동안 회의적이기만 했던 위와 같은 상황이 머지않아 옛이야기가 될 수도 있겠다는 생각이다. 여기서의 급격한 변화라는 것은 제Ⅱ~Ⅳ장에서 논한다. 그리고 이 연구 이후의 후속 연구까지 아우르는 '한국학 데이터 표준화 방안 연구' 전체를 통해 조금은 낙관적인 희망을 공유했으면 한다.

- 4 「인문학 데이터」란 말 그대로 인문학 분야의 데이터인데, '데이터'가 뜻하는 바는 제Ⅱ장 제1절(표준화, 데이터 다지기)에서 구체적으로 고찰한다. 디지털 인문학 분야의 정보화 사업은 아날로그 형태의 인문학 데이터를 디지털로 전환하는, 즉 디지털화하는 작업을 우선적으로 진행해 왔다. 그리고 이렇게 디지털화된 인문학 데이터는 지능화 단계로 도약하고 있으며, 디지털화와 지능화는 동시에 기획·진행되기도 한다. '지능화'란 「지능정보화 기본법」제2조의 5)에서 말하는 지능정보화, 즉 "정보의 생산·유통 또는 활용을 기반으로 지능정보기술이나 그 밖의 다른 기술을 적용·융합하여 사회 각 분야의 활동을 가능하게 하거나 그러한 활동을 효율화·고도화하는 것"을 집약한 표현이다. 해당 내용과 함께 더 구체적인 설명은 과학기술정보통신부, 「지능정보화 기본법」, 법제처 국가법령정보센터(<https://www.law.go.kr/법령/지능정보화기본법>)에서 확인할 수 있다.
- 5 데이터의 중복 및 조직·업무·시스템별 데이터 불일치, 데이터에 대한 의미 파악 지연으로 인한 정보 제공의 적시성 손실, 데이터 통합의 장애, 정보시스템 변경과 유지보수 곤란 등의 문제점을 개선하는 첫 단계가 데이터의 표준화·규격화라고 할 수 있다. 한국데이터산업진흥원(편), 『데이터아키텍처 전문가 가이드』(서울: 한국데이터산업진흥원, 2020), 244~245쪽.

기술자들을 잘 인도하려면 정보기술에 대한 이해가 뒷받침되어야 하는 것이 필수적이지만, 현실의 업계를 경험해 본 분들이라면 인정하듯이 이는 적잖은 부담과 희생을 감수해야 할 도전에 가까운 일이다.

각설하고 이 자리에서 데이터 표준화의 당위성을 일일이 논하기에 우리가 처한 이 시대와 앉은 자리는 너무 여유가 없다.⁶ 그리고 이내 합리화한다. 이미 ‘법’으로 규정되어 있다는 말로써 말이다. 여기서의 ‘법’이라 함은 「국가지식정보 연계 및 활용 촉진에 관한 법률」(약칭 「국가지식정보법」)과 시행령 및 관련 행정규칙, 「공공데이터의 제공 및 이용 활성화에 관한 법률」(약칭 「공

6 「너무 여유가 없다.」는 것은 정작 논하고 싶은 내용이 많지만 제한된 지면 사정과 논의의 분량상 이 글에 담을 수가 없기 때문에 궁색한 구실로나마 양해를 구하기 위한 의도에서 나온 표현이다. 지금까지 인문학 데이터 표준화의 필요성 및 당위성에 대한 논의는 주로 실용적 관점에서 이루어져 왔다. 기실 표준화라는 개념은 태생 자체가 다분히 실용적인 목적에서 출발한 것이기도 하다. 그럼에도 불구하고 표준화는 ‘동일성으로의 환원’과 논박을 위시한 철학적 담론 및 이해관계의 충돌과 호혜·협동을 포함한 사회학적 담론 등 다양한 관점의 성찰이 필요하다. 다만 이 글에서 거론하기에는 지면이 부족하므로 별도의 논고에서 상술하겠다. 경제학을 포함하여 실용적 관점에서 이루어진 논의로서 한국데이터산업진흥원(편), 앞의 책, 244~245쪽; 성태경, 『표준의 경제학: 이론·사례·정책』(파주: 한국학술정보, 2012), 17~18쪽, 33~34쪽, 37~39쪽, 47~48쪽, 231~232쪽 등을 참고할 수 있고, 인문학 데이터에 대한 논의로서는 다음을 참고할 수 있다. 김현, 『인문정보학의 모색』(성남: 북코리아, 2012), 271~273·317~319쪽; 류인태, 「인문학이 데이터와 만날 때: 재현-연결-표준으로서 인문학 데이터의 내포를 중심으로」, 『동서인문학』 65(2023), 28~33쪽. 더불어 ‘표준화’라는 것 자체가 갖는 맹점, 즉 혁신에 오히려 방해가 되는 경향도 있다는 점에 착안하여 데이터 표준화의 딜레마를 극복할 방안을 제시하는 다음의 논고들도 참고할 만하다. Yoshiaki Fukami, “Standardization Procedure for Data Exchange,” *Information*, Vol. 11, Issue 6(2020), p. 339; Michal S. Gal, Daniel L. Rubinfeld, “Data Standardization,” *New York University Law Review*, Vol. 94, Issue 4(2019), pp. 737-770.

7 「공공데이터법」은 2013년 7월 30일에 처음 제정되었고 그해 10월 31일 자로 동법 시행령과 함께 시행되었으며, 2023년 5월 16일에 5번째 개정 후 동년 11월 17일 자로 시행되었다. 현재 행정안전부 공공데이터정책과에서 주관한다. 행정안전부, 「공공데이터의 제공 및 이용 활성화에 관한 법률」, 법제처 국가법령정보센터(<https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>) 참조.

공데이터법」⁷과 시행령·시행규칙 및 관련 행정규칙,⁸ 「지능정보화 기본법」⁹과 시행령·시행규칙 및 관련 행정규칙 등을 말한다.¹⁰

이렇게 법으로 규정할 만큼 모든 분야에서 데이터 표준화가 의무적인 상황은 전 세계적으로 활발하게 진행되고 있는 다양한 분야별 데이터 표준화 연구 추세와 궤를 같이한다.¹¹

-
- 8 「공공데이터법」관련 행정규칙에는 각 공공기관별 규정이나 지침 외에 「공공데이터 이용활성화 지원 사업 관리지침」, 「공공기관의 데이터베이스 표준화 지침」, 「공공데이터 관리지침」 등이 포함된다.
- 9 인문학 정보화 연구의 일환인 이 연구의 입장에서 이 법령과 관련하여 특기할 만한 것은, 「지능정보화 기본법」이라는 명칭이 출현하기 전에 일시적으로 존속했던 「지식정보자원관리법」을 인문학 정보화의 입장에서 분석·정리하고 그 문제점에 대한 정책적 제안을 시도한 다음의 연구보고서, 그리고 뒤이어 이 법을 기반으로 구축된 「한국역사정보통합시스템」까지 함께 다루면서 한국 인문학 정보화를 조망한 다음의 논고를 눈여겨볼 만하다는 점이다. 더불어 후자는 『인문콘텐츠』 창간호에 게재되었다는 점도 나름의 의의를 더한다. 이남희·김석근, 『지식정보화 관련 법령 분석과 인문학 진흥을 위한 정책제안 연구』(서울: 한국교육개발원 인문사회연구회, 2002); 이남희, 「인문학과 지식정보화: 『지식정보자원관리법』과 『한국역사정보통합시스템』을 중심으로」, 『인문콘텐츠』 1(2003), 117~130쪽.
- 10 이 외에 행정안전부 디지털정부기획과에서 총괄하고 있는 「전자정부법」과 시행령·시행규칙 및 관련 행정규칙, 문화체육관광부 문화정책과와 교육부 학술연구정책과에서 주관하는 「인문학 및 인문정신문화의 진흥에 관한 법률」(약칭 「인문학법」)과 시행령 및 관련 행정규칙 등은 부가적 요소로만 감안하기로 한다. 다만 「전자정부법 시행령」의 행정규칙 중 하나인 「행정기관의 코드표준화 추진지침」은 분야를 막론하고 공통적 기반으로 감안되어야 한다.
- 11 이러한 상황이 전개되는 이유는 각주 5의 내용을 통해 추정할 수 있다. 다음의 논고들은 각각 미국, 중국, 일본, 독일, 우크라이나 등에서 다양한 분야별로 데이터 표준화에 중점을 둔 최근 논의들의 예이다. 특기할 만한 것은 여기서 제시하지 않은 많은 최근 논고들까지 함께 살펴볼 때 절대적인 다수가 의료 관련 분야에서 활발하게 발표되고 있다는 점이다. Michal S. Gal, Daniel L. Rubinfeld, *op. cit.*, pp. 737-770; 潘燕杰, “我国健康医疗数据治理的标准化路径: 以美国相关法律为借鉴,” *Journal of Nanjing Medical University: Social Sciences*, Issue 5(2023), pp. 423-429; Yoshiaki Fukami, *op. cit.*, p. 339; Antonie Fuhr, Andreas Kurtz, Christian Hiepen, Sabine Müller, “Organoids as Miniature Twins-Challenges for Comparability and Need for Data Standardization and Access,” *Organoids*, Vol. 1, Issue 1(2022), pp. 28-36; Δ. С. Кацу, Т. М. Александрова, П. С. Талапова, И. Б. Агсева, М. М. Всль, М. О. Трофименко, М. Р. Колесник, Т. С. Несміян, “СУЧАСНИЙ СТАН ТА ПЕРСПЕКТИВИ ВПРОВАДЖЕННЯ СТАНДАРТИЗАЦІЇ ДАНИХ У СИСТЕМУ

이러한 상황에서 국가적 정체성과 문화적 소통의 기반인 한국학¹² 분야에서의 데이터 표준화를 진지하게 고민하지 않을 수 없다.

하지만 아쉽게도, 한글과 한자를 주요 매개체로 하여 이루어지는 한국학 관련 데이터 표준화를 중점적으로 논한 연구나 그 현황을 조망하는 연구는 아직 보이지 않는다. 더불어 공공데이터를 포함하여 이 글에서 정리한 각종 데이터 표준화 관련 개념과 분류 체계에 대해서 총체적으로 알려주는 자료도 보이지 않는다.

그러면 자명해진다. 한국학 데이터 표준화를 어떻게, 어느 수준까지 진행 할 것인가를 고민하기에 앞서 우리가 해야 할 일은 각종 데이터 표준화 관련 개념과 분류 체계 및 정책·환경에 대해서 총체적으로 검토하고 한국학 분야의 데이터 표준화 현황이 어떠한지를 살펴보는 것이다.

다만 미리 유념해야 할 사항이 있다. ‘현황’이라고 하면 통상 중간점검 개념을 떠올리기 쉽지만, 한국학 데이터 표준화에서의 현황이라는 것은 이러한 개념을 투사하기에 무색한 느낌이 있다. 중간점검이라기보다는 기초공사를 위한 환경 여건 검토에 가깝다고 할 수 있기에 이 글에서는 데이터 표준화 관련 개념과 분류 체계 및 정책 환경에 대한 기반적인 검토를 수행하는 데에 의

ОХОРОНИ ЗДОРОВ'Я УКРАЇНИ (огляд літератури)," *Medični Perspektivi*, Vol. 28, Issue 3(2023), pp. 190-198. 한국의 통상적인 언어환경을 감안하여 부연 설명을 하자면, 마지막 논고는 ‘우크라이나 의료 시스템의 데이터 표준화 구현 현황과 전망’을 다룬 것이다.

12 ‘한국학’은 “한국 문화의 정체성과 한국인이 세계와 만나고 소통하면서 이룩한 성취를 연구하는 학문”으로 간명하게 정의할 수 있다. 한국학중앙연구원 홈페이지 ‘연구원 소개’(<https://www.aks.ac.kr/com/cmm/EgovContentView.do?menuNo=5010110000&lang=kor>) 참조. ‘한국학’의 개념과 성립에 관하여 최근까지 진행된 구체적인 논의는 황원구·정구복, *한국민족문화대백과사전*, ‘국학(國學)’(<https://encykorea.aks.ac.kr/Article/E0006552>); 전훈지, 「한국학의 정의와 방법에 대한 고찰: 김경일 저, 『한국의 근대 형상과 한국학 - 비교 역사의 시각』의 서평」, 『비교한국학』 29권 2(2021), 87~124쪽; 오영교, 「한국학·지역학의 발흥과 한국사연구」, 『한국사연구』 200(2023), 1~34쪽 및 이 글들 속 참고문헌을 넣을 캐드 하여 확인할 수 있다.

미를 둘 것이다. 그리고 이와 관련한 현 상황을 살펴보면서 그것이 취하고 있는 지형을 그려 보고 향후 전략을 모색해 볼 것이다. 더 나아가 이 글에서 정리한 체계를 토대로 후속 연구를 진행하는 과정에서 구체적인 ‘현황’의 수준이 성장해 가기를 바라 본다.

II. 데이터 표준화 체계와 한국학

‘한국학 데이터 표준화’를 살펴보기 위해서는 ‘한국학’ 및 ‘데이터 표준화’에 대한 기반적 검토가 필요하겠다. 우선 이렇게 시작하자. ‘데이터 표준화’란 무엇인가? ‘데이터’란 무엇이고, ‘표준화’란 무엇이며, ‘표준’이란 무엇인가?

1. 표준화, 데이터 다지기

‘한국학 데이터 표준화’를 이야기할 때의 ‘표준’은 “사물의 정도나 성격 따위를 알기 위한 근거나 기준”¹³에 상응하는 영어 ‘standard’¹⁴를 칭하는 것이다. 서기전 7000년 무렵 이집트에서 원통형 돌을 무게 측정 단위로 삼은 것

13 이러한 의미로 동아시아 한자 문화권에서 사용된 ‘標準(標準)’은 『荀子』 儒效篇의 “行有防表”에 대해 楊倞이 “謂有標準也”라고 주해한 것이나 『文選』에 실린 袁宏의 「三國名臣序贊」에서 “器範自然 標準無假”라고 한 것에서 확인할 수 있다. 국립국어원 표준국어대사전, ‘표준’ 「명사」 「1」(https://stdict.korean.go.kr/search/searchView.do?word_no=500503&searchKeywordTo=3); 단국대학교 부설 동양학연구소(편), 『한한대사전 7』(서울: 단국대학교출판부, 2004), 589쪽.

14 어원 측면으로는 프랑스어 standarde, estandard의 차용어이며 영어에서의 의미는 ‘단단하게 세워진 기치’ 정도로 축약해서 이해할 수 있다. Oxford University Press The Oxford English Dictionary(https://www.oed.com/dictionary/standard_n?tab=etymology#20823132).

이 그 기원으로 알려져 있으며, 다수에 의해 공식적으로 합의된 것일 수도 있고 암묵적 관례가 된 것일 수도 있다.¹⁵

‘표준화’란 표준의 이러한 두 가지 양상에 따라 달리 진행되는 두 종류의 과정으로 개념화된다. 하나는 집단적 합의에 의한 것으로서 각종 이해당사자들 간의 공통 기준 설정과 조정을 통해 이루어지는 과정이고, 다른 하나는 시장 과정에 의한 것으로서 시장점유율이 높아지고 지배제품이 되면서 이루어지는 과정이다.¹⁶

그러면 표준화의 대상이 되는 ‘데이터’란 무엇인가? 기본적으로 ‘data’는 ‘주어진’의 의미를 내포하는 라틴어 ‘datum’(dare, data)의 차용어로서 영어에서 여덟 가지 의미 맥락으로 구분할 수 있지만, 간략하게 세 가지로 대별해서 이야기해 보면 다음과 같다.¹⁷

정보화의 관점에서 표준화의 대상으로 가장 먼저 제시할 수 있는 것은 “컴

15 기본적으로 표준의 유형은 각종 표준화 기구에 의해 제정된 공적 표준(de jure standards, 데유레 표준)과 시장 논리에 의해 표준처럼 인정된 사실상 표준(de facto standards, 데파토 표준)으로 나눌 수 있다. 이러한 구분 외에도 제품표준·비제품표준, 공개표준·사적소유 표준, 자생적 표준·의도적 표준, 규제적 표준·비규제적 표준 등으로도 구분이 되고, 「국가 표준기본법」에는 국가표준·국제표준 및 측정표준·참조표준·성문표준 등이 설명되어 있는데, 이 글에는 논지상 일일이 거론하지 않는다. 성태경, 앞의 책, 14쪽, 25쪽, 30~32쪽; 이 강원·손호웅, 『지형 공간정보체계 용어사전』(서울: 구미서판, 2016), ‘Standard’(<https://terms.naver.com/entry.naver?cid=58439&docId=3482948&categoryId=58439>); 산업통상자원부, 「국가표준기본법」, 법제처 국가법령정보센터(<https://www.law.go.kr/법령/국가표준기본법>) 참조.

16 앞서 언급한 ‘공적 표준’은 전자에 해당하고, ‘사실상 표준’은 후자에 해당한다. 성태경, 앞의 책, 36쪽 참조.

17 Oxford University Press, Oxford English Dictionary, “Datum, N.”(December 2023, <https://doi.org/10.1093/OED/7571592234>); “Data, N.”(September 2023, <https://doi.org/10.1093/OED/7999740343>); 국립국어원 표준국어대사전, “데이터”(<https://stdict.korean.go.kr>); 두산 Doopedia, “데이터[data]”(https://www.doopedia.co.kr/doopedia/master/master.do?_method=view&MAS_IDX=101013000835838); 장석권, 『데이터를 철학하다 : 어떻게 데이터는 지혜가 되는가』(서울: 흐름, 2018), 19~23쪽.

퓨터가 처리할 수 있는 문자, 숫자, 소리, 그림 따위의 형태로 된 정보”(이하 ①)이다. 이는 사실상 근본적인 차원에서 보자면 이미 ‘표준화’가 진행되어 있다고 볼 수도 있다. 컴퓨터 저장장치에 기록된 데이터의 형태가 이미 세계적 표준에 근거한 이진 데이터를 기반으로 하기 때문이다. 다만 컴퓨터라는 기록매체를 포함하는 연산장치가 만들어지기 오래전부터 ‘data’라는 용어는 사용되어 왔다. 이는 다음과 같은 두 가지로 대별할 수 있다.

즉, “관찰이나 실험, 조사로 얻은 사실이나 정보”(이하 ②) 및 “이론을 세우는데 기초가 되는 사실. 또는 바탕이 되는 자료”(이하 ③)이다. 그리고 이 둘은 재귀적·순환적 인과 관계를 갖기도 한다. ②는 “어떠한 사실, 개념, 명령 또는 과학적인 실험이나 관측 결과로 얻은 수치나 정상적인 값 등 실체의 속성을 숫자, 문자, 기호 등으로 표현한 것”으로 구체화하여 표현할 수 있으며, 이는 다시 “사물, 현상, 사건, 인간관계 등에 관한 관찰 기록”으로 축약하고 현대화하여 표현할 수도 있다. 이러한 ②는 곧 ③으로 작용하여 또 다른 정보·이론을 도출하기도 하고 이렇게 도출된 정보는 다시 ②가 될 수 있는 것이다.

데이터의 표준화가 시대적·지속적 과제가 되는 것은 ①이 ① 자체에서 그치지 않고 과거의 본의, 즉 17세기에 이미 보이던 ②나 ③의 의미로까지 그 범위를 확장할 필요성이 커지기 때문이다.

그러면 ‘데이터의 표준화’는 무엇인가를 정리해 보자. 앞서 살펴본 개념어들의 의미를 조합해 보면 ①의 형태를 띤 ②나 ③의 데이터를 집단적 합의에 의해 표준화하는 과정, 또는 이 데이터가 시장 과정에 의해 표준화되는 과정이라고 할 수 있다. 이 두 가지 중 한국학 데이터 표준화를 논하는 관점에서는 전자가 해당된다. 이유는 한국학 정보화의 관점에서 데이터 표준화를 논하는 2절에서 확인할 수 있다.

이렇게 해서 데이터를 표준화하게 되면 ‘표준 데이터’가 된다. 그런데 표준 데이터는 두 가지 부류가 서로 다른 차원에서 거론된다. 하나는 ‘제공표준’이

라는 부류이고, 다른 하나는 ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부류이다.

두 가지 모두 행정안전부 소관의 「공공데이터법」 및 행정규칙에서 거론되는 표준 데이터이지만 그 성격은 사뭇 다르다. 각 부류가 의거하는 행정규칙이 다르고, 이 행정규칙을 담당하는 행정안전부 하위 과를 보면 그 차이를 비교적 명확하게 알 수 있다. 즉, 전자는 공공데이터정책과에서 담당하는 「공공데이터 관리지침」에 의거한 부류이고, 후자는 공공데이터관리과에서 담당하는 「공공기관의 데이터베이스 표준화 지침」에 의거한 부류이다. 전자는 연역적 개념체계의 소산이고, 후자는 귀납적 개념체계의 소산인 것이다.

이들에 대한 내용적 검토는 ‘공공데이터’ 개념의 서술이 우선되어야 하므로, 제Ⅲ장 1절에서 진행하기로 한다. 다만 이 글에서 다루는 ‘한국학 데이터 표준화’는 후자의 부류에 기반한 것임을 염두에 두고 후속 논의를 이어 간다.

그럼 이제 ‘데이터 표준화’라는 단순한 듯하면서도 막연할 수 있는 개념이 실제 어떠한 행위 집합으로 구성되어 있는지를 훑어본다는 차원에서 데이터를 표준화하는 과정을 간략하게 서술형으로 추려 보겠다. 물론 구체적인 내용은 제외하고 제목 수준으로 내세울 수 있는 사항만을 나열할 것이다.¹⁸

첫 단계로서 데이터 표준을 수립하는데, 이는 다음의 과정으로 진행된다. 우선 데이터 표준화의 원칙을 정의한다. 이를 위해 표준화 요구 사항을 수집한다. 그리고 현행 데이터 표준의 원칙을 분석한다. 그리하여 데이터 표준을 개선할 방안을 정의한다. 이를 토대로 데이터 표준의 원칙을 수립한다. 다음으로 데이터 표준을 정의할 차례다. 먼저 표준 단어 사전을 정의한다. 표준

18 한국데이터산업진흥원(편), 앞의 책, 261~294쪽; 행정안전부, 「공공기관의 데이터베이스 표준화 지침」, 법제처 국가법령정보센터([https://www.law.go.kr/행정규칙/공공기관의데이터베이스표준화지침/\(2023-18,20230403\)\)」; 행정안전부\(편\), 앞의 책\(2023d\), 14~164쪽 참조.](https://www.law.go.kr/행정규칙/공공기관의데이터베이스표준화지침/(2023-18,20230403))

도메인 사전도 정의한다. 표준 코드 사전도 정의한다. 그런 후에 표준 용어 사전을 정의한다. 이제는 데이터 표준을 확정할 차례다. 이를 위해 먼저 데이터 표준을 검토하는데, 일단 검토 계획을 수립하고 나서 본격적인 표준 검토가 진행되고 그 후에 데이터 표준을 보완 및 승인하는 과정을 거친다. 다음으로 데이터 표준을 공표하는데, 일단 전사 데이터 관리자의 승인을 받은 표준을 표준 관리 도구에 등록하여 배포한 후 사용자 및 운영자 교육까지 진행한다. 다음 단계로서 데이터 표준을 관리하는 절차가 진행되는데, 이는 다음의 과정들로 구성된다. 데이터 표준의 신규 요건이 발생했을 때의 처리 프로세스, 데이터 표준의 변경이나 삭제 시 표준화 요소·데이터 모델·데이터베이스·프로그램 영향도 분석 절차 및 처리 프로세스, 데이터 표준 준수 여부에 대한 수시 점검 프로세스 등이다. 그리고 이를 위해서는 업무 담당자, 데이터베이스 관리자, 데이터 관리자, 전사 데이터 관리자 등으로 역할과 담당 업무가 정의되어야 한다.

이렇게 열거 서술한 데이터 표준화의 행위집합은 크게 두 가지 단계로 대별할 수 있다. 즉, 데이터 표준 수립 단계와 데이터 표준 관리 단계로 구분할 수 있으며, 이 과정에서 데이터 표준 관리 도구도 활용된다. 그리고 이러한 작업은 데이터 품질관리와 밀접하게 연계된다. 제Ⅲ장에서 논의할 ‘공공데이터 품질관리 수준진단·평가’의 내용까지 어우러지면 감별이 용이해질 것이다.

여기까지 개관적인 ‘데이터 표준화’를 살펴보았고, 이제는 ‘한국학’ 및 그 것의 정보화 차원에서 데이터 표준화와의 연결고리를 살펴보겠다.

2. 한국학 정보화의 터 다지기

동아시아 인문학의 한 갈래인 ‘한국학’에서 ‘데이터 표준화’라는 것은 인문

학 정보화의 일환에서 상정되는 사안이다.¹⁹ 한국학 정보화는 연구의 계층적 관점에서 크게 두 갈래로 나누어 개념화할 수 있다. 하나는 데이터 체계화 연구이고, 다른 하나는 콘텐츠 큐레이션 연구이다. 최근에 진행되는 ‘디지털 인문학’ 연구는 대부분 콘텐츠 큐레이션에 무게중심이 실린 것으로 보인다. 데이터 표준화를 논하는 이 글은 데이터 체계화 연구에 해당하는 사례가 된다. 엄밀한 시각에서 보면 양자가 융합된 연구가 많고, 양자 사이에 존재하는 매개적 연구도 있다.²⁰

계층적 측면에서 데이터 체계화 연구는 콘텐츠 큐레이션 연구를 위시한 상위 계층의 연구를 위한 기반적 층위가 되기에, 이 중의 한 영역을 차지하는 한국학 데이터 표준화 연구는 한국학 정보화의 ‘터’를 다루는 역할을 한다. 그럼 이 시점에서 1절의 ‘데이터 표준화’ 개념 논의 중 남겨진 설명을 이어 가보자.

‘데이터의 표준화’는 요컨대 데이터를 집단적 합의에 의해 표준화하는 과

-
- 19 인문학 정보화를 연구하는 학문을 ‘인문정보학’이라고 부를 수 있는데. 이 용어가 2001년 처음 공식적으로 사용된 후로 점차 진화된 내역을 다음에서 확인할 수 있다. 김현, 인문정보학 개론 2024, “‘인문정보학’에 대한 정의”(http://dh.aks.ac.kr/Edu/wiki/index.php/인문정보학_개론_2024). 이곳에서는 2008년 이후로 미국에서 공식적으로 사용된 ‘디지털 인문학(digital humanities)’이라는 용어에 대해서도 함께 언급한다.
- 20 디지털 인문학(한국학)의 초기부터 요사이까지의 몇몇 대표적 사례는 김현·임영상·김바로, 『디지털 인문학 입문』(서울: 한국외국어대학교 지식출판원, 2016), 18~38쪽; 이남희, 앞의 논문, 118~120쪽, 125~128쪽; 양창진, 「정보화에세이: 정보통신 기술과 인문학의 만남, 그리고 상생」, 『지역정보화』 51(2008), 118~121쪽; 한국국학진흥원, [2022 전통 기록문화 창작 콘퍼런스] 발표 ② 한국학 자료의 큐레이션과 그 사례 - 양창진 관장(한국학중앙연구원)(<https://youtu.be/duiRHQmfjXY?si=2mXaBMPc8UAg05yD>); 최정도, 「인문학 위기 속의 국어정보학의 현황과 전망」, 『국어국문학』 201(2022), 35~44쪽; 지동은·SANTOSH KUMAR RANJAN, 「북한과 중국에서 조선어(한국어)정보화 현황과 과제」, 제51회 한중인문학회 국제학술대회(서울: 서울대학교, 2023.5.13.), 105~120쪽 등을 참고할 수 있다. 시맨틱 데이터 구축 연구는 데이터 체계화 연구와 콘텐츠 큐레이션 연구 사이를 매개하는 위상을 갖는 것으로 볼 수 있다.

정, 또는 데이터가 시장 과정에 의해 표준화되는 과정이라고 할 수 있다. 전자는 적극적 표준화, 후자는 소극적 표준화라고도 볼 수 있다. 한국학 데이터 표준화를 논할 때에는 전자가 해당된다. 후자가 배제되는 데에는 다음과 같은 저변의 사정이 있음을 상정할 수 있다.

인문학 분야에서 업계와 시장이 주도하는 데이터 표준화를 얼마나 기대하고 신뢰할 수 있을까? 이에 대해서는 회의적인 시각이 많을 것이다.²¹ 가시적인 경제적 이익을 보장하기 어려운 한국학이라는 인문학 분야에서 업계와 시장이 주도한다는 표현을 쓰는 것 자체도 어색하거나 표준화에 의한 경쟁 촉진·저해 여부, 기술혁신 위축·촉진 여부, 산출물의 신뢰성 여부 등의 경제적 효과를 거론하지 않더라도²² 이미 이 분야에서 대부분의 재원은 이공계 대비 영세한 수준의 정부 지원으로 이루어지고 정보화를 수행하는 업체와 인력은 인문학과 정보기술 양 방면의 전문성을 겸비하기 어려운 여건에서 한정적인 요건만을 충족시키는 수준으로 결과물을 산출하고 있다는 점에서 시장과정에 의해 데이터 표준화가 이루어질 것을 기대하기는 어렵다. 더불어 다양하게 분화된 학제적 성격이 제각각 존중·유지될 수밖에 없는 구조적 한계는 이를 더욱 실현하기 어렵게 만드는 요소로 작용할 수 있다.

그러므로 데이터를 집단적 합의에 의해 표준화하는 과정으로서의 한국학 데이터 표준화는 적극적 조치가 될 수밖에 없다. 적극적인 조치라고 하면 결국 통제적인 성격을 가질 수밖에 없고 어느 정도의 범위 내에서 구속을 가할 수 있는 체계가 성립되어야 한다. 머리말에서 언급한 바와 같이 데이터 표준

21 데이터 과학 분야에서 시장이 이끄는 표준화의 주도적 입장은 신뢰할 수 있는지 여부와 정부가 데이터 표준화를 활성화할 때 수행해야 하는 역할에 대한 논의로 다음을 참고할 수 있다. 이 글에서는 데이터 표준화의 필요성과 관련하여 이식용이성(portability)과 상호운용성(interoperability)의 측면에 비중을 둔다. Michal S. Gal, Daniel L. Rubinfeld, *op. cit.*, pp. 737-770.

22 이에 대한 거론은 다음을 참고할 수 있다. 성태경, 앞의 책, 43~54쪽.

화가 이미 법으로 규정되어 있는 상황은 바로 이 조치의 일환이며, 이는 제Ⅲ장에서 구체적으로 논한다.

아울러 또 하나의 적극적인 조치로서 온정적인 지원과 함께 표준화를 유도하는 방법도 있다. 이는 제Ⅳ장에서 간략히 언급한 후 후속 논고로 배턴을 이어 간다. 어느 쪽이 되건 이 양자는 모두 한국학 정보화의 터가 되는 데이터를 다지는 작업이 된다. 그리고 그것이 나아갈 방향은 한국학 데이터에 대한 ‘세계적 표준’을 지향해야 한다. 이를 위해서는 한국학과 인접한 국내외 학문 영역에서의 표준과 조화를 이룰 수 있어야 하고, 한국학만이 가진 영역에서 는 독자적인 표준을 정의해야 한다.

III. 공공데이터의 적극적 표준화

공공데이터 정책은 매년 점점 확대되고 구체화되고 있다.²³ 이로 인해 해가 갈수록 인문학 정보화 용역사업 수행업체들은 공공데이터 정책의 기준에 따르기 위해 기존에 해 보지 않았던 요구사항을 이행하느라 고심을 거듭하고 있다.

‘데이터 표준화’만을 보아도 관련된 정책 및 가이드 자료가 다양하며, 어느 하나만을 보아서는 제대로 이해하기가 어렵고 모두를 함께 검토해야 한다. 그런데 데이터 표준화를 포함한 데이터 품질관리를 비롯하여 데이터에 대한 활용과 메타적 관리 체계까지 함께 서술될 수밖에 없는 것이 이 분야의 특성 이기 때문에 이 분야에 대해 어느 정도 갈피를 잡기 위해서는 상당한 기간과

23 행정안전부(편), 『2023년 공공데이터 품질관리 수준진단·평가 매뉴얼』(대구: 한국지능정보사회진흥원, 2023c), 9쪽; 행정안전부(편), 앞의 책(2023d), 7쪽, 30~31쪽, 165쪽; 조운휘·이용건, 앞의 논문, 801쪽에서 이러한 양상을 확인할 수 있다.

노력을 투여해야 한다.

하지만 한국학이라는 전문 영역에서 진행되는 데이터 표준화 현황을 알기 위해서는 복잡한 미로 속에서 논점에 필요한 줄기만을 속아 내어 최소한의 분량으로나마 나름의 체계를 만들어 낼 필요가 있다. 이를 수행한다는 측면에서도 이 글은 적잖은 의의가 있다.

1. 공공데이터 논의를 위한 속아 내기

제Ⅱ장에서 ‘데이터 표준화’에 대해 가지치기를 한 바 있다. 그리고 데이터 표준화를 위한 적극적인 조치의 하나로서 법적인 통제 체계를 언급했다. 이 절에서는 해당 내용을 개괄하여 더 체계적인 관점에서 더 집중적인 논의를 위하여 복잡한 개념과 체계를 다듬고 속아 내려 한다.

데이터 표준화를 위한 법적인 통제 체계로서 「국가지식정보법」과 시행령 및 관련 행정규칙, 「공공데이터법」과 시행령·시행규칙 및 관련 행정규칙, 「지능정보화 기본법」과 시행령·시행규칙 및 관련 행정규칙, 「전자정부법」과 시행령·시행규칙 및 관련 행정규칙, 「인문학법」과 시행령 및 관련 행정규칙 등을 간략하게 요약 제시할 수 있다. 이 중 데이터 표준화의 관점에서 가장 중심에 있는 핵심 법규는 「공공데이터법」과 시행령·시행규칙 및 관련 행정규칙이다. 그리고 이는 다시 다양하게 세분화된 체계로 시행되고 있다. 일단 ‘공공데이터’에 대한 정의부터 살펴보자. 「공공데이터법」에서는 ‘공공데이터’를 다음과 같이 정의한다.

“공공데이터”란 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 다음 각 목의 어느 하나에 해당하는 것을 말한다.

- 가. 「전자정부법」 제2조제6호에 따른 행정정보
- 나. 「지능정보화 기본법」 제2조제1호에 따른 정보 중 공공기관이 생산한 정보
- 다. 「공공기록물 관리에 관한 법률」 제20조제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물
- 라. 그 밖에 대통령령으로 정하는 자료 또는 정보

공공데이터는 협업에서 크게 공공데이터 활용과 공공데이터 관리의 두 갈래로 나뉘어서 처리되는데, 전자는 공공데이터 제공의 공개·개방 관련 영역이고²⁴ 후자는 공공데이터 제공의 품질관리 영역으로 편의상 구분할 수 있다.

데이터 표준화는 상술한 공공데이터의 양 측면 모두에서 핵심적 기반이 된다. 공공데이터 활용의 측면에서는 이식용이성(portability)을 중심으로 데이터 표준화가 기반적 역할을 하고, 공공데이터 관리의 측면에서는 상호운용성(interoperability)을 중심으로 데이터 표준화가 요구되는 것이다.

앞(II-1)에서 ‘표준’과 ‘표준화’, 그리고 ‘데이터’와 ‘데이터 표준화’를 다루었다. 그리고 ‘표준 데이터’의 두 가지 부류를 구분하면서 그 내용적 검토를 보류한 바 있다. 이제 그것을 이어 갈 차례다.

‘제공표준’이라는 부류는 공공데이터 활용의 영역에서 취급되는 표준 데이

24 이용자들은 공공데이터 포털을 위시하여 공공기관에서 공개한 데이터를 자신의 필요에 맞게 활용할 수 있다. 인문학 분야에서 공공데이터를 활용하는 과정과 사례들을 소개한 논고로 다음을 참고할 수 있다. 김바로, 「〈공공데이터법〉과 인문데이터: 공공기관 보유 인문데이터 공개 신청 사례를 중심으로」, 『한국고전연구』, 57(2022), 167~192쪽. 아울러 2022년에 보고된 공공데이터 ‘추진 실적 - 활용’ 부문에서 전 분야 중 가장 많이 활용한 데이터(파일 부문)로 한국고전번역원의 ‘고전원문’이 5위를 차지한 점도 눈여겨볼 만하다. 참고로 2위는 소상공인 시장진흥공단의 ‘상가/상권 정보’, 3위는 공공데이터활용지원센터의 ‘전국 종량제 봉투가격표준데이터’, 4위는 국민건강보험공단의 ‘건강검진정보’였다. 조윤휘·이용건, 앞의 논문, 795쪽 참조.

터이고, ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부류는 공공데이터 관리의 영역에서 취급되는 표준 데이터이다. 그 구체적 내용을 간략히 훑어보면 다음과 같다.

‘제공표준’은 공공데이터 중 “민간의 활용수요 등을 고려하여 동일한 항목과 형식으로 제공해야 하는 데이터의 묶음 및 집합”에 적용해야 하는 표준으로서 공공데이터 포털에 공개된다. 이렇게 작성된 표준 데이터는 ‘제공표준 데이터 셋 검색 서비스’ 및 목록등록관리시스템을 통해 정책적으로 관리된다.²⁵

다음으로 ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부류는 공공데이터베이스(“공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 데이터베이스”)에 일관되게 적용할 목적으로 수립한 “코드, 용어, 도메인, 메타데이터, 데이터셋 등의 표준”으로서 메타데이터 관리 시스템을 통해 정책적으로 관리된다. ‘공통표준’은 공통표준용어·공통표준 단어·공통표준도메인을 통칭한 것이고, ‘기관표준’은 기관표준용어·기관표준단어·기관표준도메인·기관표준코드를 통칭한 것으로서 각 기관 내부에서 공동으로 활용된다. ‘DB표준’은 DB표준용어·DB표준단어·DB표준도메인·DB표준코드를 통칭한 것으로서 기관표준을 준용하여 개별 DB 구축·운영 시 정해진다.²⁶

제Ⅱ장 1절에서 언급하기를 이 글에서 다루는 ‘한국학 데이터 표준화’는

25 행정안전부(공공데이터정책과), 「공공데이터 관리지침」 제26·27조, 법제처 국가법령정보 센터([https://www.law.go.kr/행정규칙/공공데이터관리지침/\(2021-70,20211026\)](https://www.law.go.kr/행정규칙/공공데이터관리지침/(2021-70,20211026))) 참조. 이 지침에서 데이터의 묶음 및 집합은 ‘데이터셋’으로 약칭되는데, 「공공기관의 데이터베이스 표준화 지침」에서는 “데이터베이스, 스프레드시트 등과 같이 구조화된 데이터의 묶음 및 집합”으로 데이터셋을 더 구체적으로 정의한다. ‘제공표준’은 과거에 ‘개방표준’으로 칭해졌다.

26 행정안전부(공공데이터관리과), 「공공기관의 데이터베이스 표준화 지침」 제2·4·5·9조 참조. 메타데이터 관리시스템은 행정안전부에서 구축·운영하는 중앙 메타데이터 관리시스템과 기관별로 구축·운영되는 기관 메타데이터 관리시스템으로 구분된다.

후자, 즉 ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부류에 기반한 것이라고 하였다.

전자의 경우는 2023년 10월에 개정된 제17차 『공공데이터 제공 표준』의 내역을 볼 때 총 203개 표준데이터셋 중 통상적인 관점에서 한국학과 관련지 을 수 있을 만한 것으로 9개(4.4%)를 찾을 수 있다.²⁷ 해당 표준데이터셋 명칭 은 법정구역(읍면동)정보, GIS건물통합정보, 산정보, 도시철도역사정보, 지 역특화거리, 박물관미술관정보, 관광지정보, 향토문화유적, 도서관 도서 등 이다. 이 정도를 갖고서 “민간의 활용수요 등을 고려하여” 제공하는 한국학 분야의 표준 데이터 현황을 논하기에는 아쉬움이 크다.²⁸ 이에 이 글에서는 한정된 분량으로 한국학 데이터 표준화 현황을 논하는 대상으로서 보다 체계 적으로 논의할 여지가 있는 후자로만 한정하기로 한다. 그리고 이는 뒤에 이 어질 제Ⅲ장의 2절과 3절에서 구체적으로 논의한다.

전술한 바와 같이 ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부 류에 대한 논의는 「공공기관의 데이터베이스 표준화 지침」에 의거한다. 이 내용에서는 모든 사업 종류와 범위에 공통되는 부분이 핵심이며, 이는 바로 ‘데이터 표준화 및 상위표준 준수’라는 한마디로 집약할 수 있다. 그리고 이

27 구체적인 공공데이터 제공 표준 내역은 행정안전부(편), 『공공데이터 제공 표준』(대구: 한국 지능정보사회진흥원, 2023e) 참조.

28 “한국의 전통문화 속의 역사적 사실 관계 및 그 사실의 문헌적 근거에 관한 지식을 데이터화 하기 위해 개발한 온톨로지 스키마”인 EKC(Eencyves of Korean Culture) 데이터 모델의 최신 버전에는 Actor·Event·Place·Architecture·Clothing·Food·Object·Record·Work·Concept·Heritage·Multimedia·WebResource·Bibliography·Text·Story·Index 등 열일곱 가지 기본 클래스가 정의되어 있다. 이들과 단순 비교만 해 보아도 현재의 제공표준이 한국학 데이터 표준화를 논하기에 많이 아쉬운 수준임을 볼 수 있다. EKC 데이 티 모델은 2016년 처음 제정된 후 계속 확장되고 있으며, 첫 제정 당시 필자도 핵심 역할을 담당했다. 한국학중앙연구원 디지털인문학연구소 Ontology:EKC 2022(https://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022) 참조.

것이 이 절에서 속아 낸 결과로 남는 논점이 된다.

2. 공공기관의 데이터베이스 표준화

그럼 「공공기관의 데이터베이스 표준화 지침」을 집약하는 논점 ‘데이터 표준화 및 상위표준 준수’는 어떠한 체계를 갖고 있으며, 한국학 분야에서는 어느 정도로 진행이 되고 있을까? 이를 살펴보도록 하자.

2절에서 공공데이터 논의를 위한 속아 내기를 진행했지만, 실상 이 작업은 좀 더 이어질 수밖에 없다. 「공공기관의 데이터베이스 표준화 지침」뿐만 아니라 그 상위 법령인 「공공데이터법」 제22조(공공데이터의 품질관리) 및 동법 시행령 제17조(공공데이터 품질 진단 및 개선)까지 연계된 편람·매뉴얼 등 다수의 공문서와 두꺼운 책자들에 복잡하게 얹혀 있는 데이터 표준화 체계 속에서 다시금 핵심을 추려서 한국학 데이터 표준화 현황 파악을 위한 기준으로 삼아야 하기 때문이다. ‘데이터 표준화 및 상위표준 준수’라는 기치 아래에서 추릴 수 있는 핵심사항 세 가지는 다음과 같다.

데이터 표준사전 관리항목

공통표준 필수 관리항목

행정표준코드

앞에서 ‘공통표준’-‘기관표준’-‘DB표준’ 유형으로 구성되는 부류를 거론해 왔는데, 이것이 어떤 형태를 갖고 있는지 가장 간명하게 확인할 수 있는 자료가 「공공기관의 데이터베이스 표준화 지침」의 [별표 제1호] ‘데이터 표준사전 관리항목’(표1)이다.

데이터 표준사전 관리항목에서는 공통표준, 기관표준, DB표준 모두에 적

용되는 관리항목들의 기본 패턴을 보여준다. 개별 정보시스템에 적용하는 DB표준, 다수의 개별 정보시스템을 운영하는 각 기관 내부에서 공동으로 활용하는 기관표준, 국내 모든 공공기관이 공동으로 활용·준용하는 공통표준 모두가 이러한 데이터 표준사전의 패턴으로 작성된다. 이 중 가장 상위에 있는 것이 공통표준이고, 공통표준에서 필수적으로 관리하는 항목이 두 번째 핵심사항이 된다. 공통표준은 공통표준단어·공통표준도메인·공통표준용어로 구성되는데 이들은 다시 다음과 같은 관리항목들로 구성된다.²⁹

공통표준단어에는 공통표준단어명·공통표준단어영문약어명·공통표준단어영문명·공통표준단어설명·형식단어 여부·공통표준도메인 분류명·이음동의어 목록·금칙어 목록 등의 관리항목이 있고, 이들은 모두 필수이다.

공통표준도메인에는 공통표준도메인 그룹명·공통표준도메인 분류명·공통표준도메인명·공통표준도메인 설명·데이터 타입·데이터 길이(자릿수)·소수점 길이·저장형식·표현형식·단위·허용값·행정표준코드명 등의 관리항목이 있고, 이 중 행정표준코드명은 필수가 아니다.

공통표준용어에는 공통표준용어명·공통표준용어 설명·공통표준용어영문약어명·공통표준도메인명·허용값·저장형식·표현형식·데이터 타입·데이터 길이(자릿수)·소수점 길이·단위·행정표준코드명·소관 기관명·이음동의어 등의 관리항목이 있고, 이 중 데이터 타입·데이터 길이(자릿수)·소수점 길이·단위·이음동의어는 필수가 아니다. 즉, 공통표준용어명·공통표준용어 설명·공통표준용어 영문약어명·공통표준도메인명·허용값·저장형식·표현형식·행정표준코드명·소관기관명 등이 핵심 항목이 된다.

다만 DB표준이나 기관표준과 달리 공통표준의 경우, ‘공통코드’라는 구분이

29 행정안전부, 「공공기관의 데이터베이스 표준화 지침」[별표 제3호] ‘공통표준용어 관리항목’; 행정안전부(편), 앞의 책(2023d), 95~98쪽 참조.

표1 - 데이터 표준사전 관리항목

구분	항목명	항목 정의 및 작성 지침
	표준용어명	<ul style="list-style-type: none"> • 엔터티, 속성 등 데이터 요소의 명명에 사용될 용어를 기재 <ul style="list-style-type: none"> — 동음이의어와 약어는 중복을 허용하거나 사용 또는 작성은 최소화 • 한글을 원칙으로 하며, 특수 문자와 공백은 사용 불가 <ul style="list-style-type: none"> — 숫자의 사용은 가능하나, 표준용어의 직관적 의미 파악이 가능하도록 사용을 최소화 • 관용적으로 널리 알려진 영문약어에 한하여 표준용어로 사용 가능(예: URL, IP 등)
	영문명	<ul style="list-style-type: none"> • 표준용어의 영문명으로, 대문자를 사용하여 영문전체명(full name)으로 기재 <ul style="list-style-type: none"> — 예: 표준용어명 “가공식품”의 경우 영문명은 “PROCESS FOOD”로 기재
표준 용어	영문약어명	<ul style="list-style-type: none"> • 해당 표준용어에 대한 영문약어명을 기재 • 대문자로 작성하고, 불가피한 경우를 제외하고는 특수문자는 사용하지 않도록 유의 • “용어구분”이 동의어인 경우 해당 용어의 표준어와 동일한 영문약어를 기재 • 아래 예시를 참조하여 자체적인 영문약어 명명규칙을 작성하여 약어명을 정의할 것을 권고 <ul style="list-style-type: none"> — 예: 모음을 제외한 자음의 결합으로 영문 약어 작성(PRICE → PRC) — 예: 같은 자음이 2개 이상 연속되면 1개만 선택하여 작성(MAPPING → MPNG) — 예: 용어 영문명의 앞부분만 사용하여 작성(ADDRESS → ADDR) — 예: 용어 영문명의 첫 글자와 마지막 글자를 조합하여 작성(YARD → YD) — 예: 용어 영문명이 긴 경우는 첫 자음과 마지막 자음 사이에 특징 있는 자음들을 조합하여 작성(NETWORK → NW, SOFTWARE → SW, MESSAGE → MSG)
	용어 설명	<ul style="list-style-type: none"> • 표준용어의 의미나 내용을 상세히 서술 • 표준어, 동의어 이외에 금칙어를 제공할 경우에는 금칙어 목록을 포함하여 기재할 수 있으며, 1개 용어에 대하여 다수의 금칙어가 존재할 수 있으므로 별도의 관리항목으로 정의하여 사용 가능 <ul style="list-style-type: none"> — 예: ‘휴대폰’이 표준어로 정의된 경우 ‘핸드폰’, ‘모바일폰’ 등을 금칙어로 규정하여 표준용어로 사용하지 못하도록 할 수 있음
	표준도메인명	<ul style="list-style-type: none"> • 도메인정의서에 정의된 용어의 경우 해당 도메인명을 기재

구분	항목명	항목 정의 및 작성 지침
표준 단어	허용값	<ul style="list-style-type: none"> 해당 표준도메인이 가질 수 있는 최댓값/최솟값이나 유효값을 기재 예: 월 : 01~12, 여부 : 'Y', 'N' <p>※ 행정표준코드는 행정표준코드관리시스템(https://www.code.go.kr)을 참조</p>
	관리부서명	<ul style="list-style-type: none"> 해당 용어를 생성 및 관리하는 주체인 부서명을 기재
	표준코드명	<ul style="list-style-type: none"> 해당 용어가 표준코드에 해당하는 경우 해당 표준코드명을 기재
	업무 분야	<ul style="list-style-type: none"> 해당 테이블의 내용과 관련된 BRM 분류체계를 기재 해당 용어가 다수의 업무분야에서 일상적으로 사용되는 경우에는 “공통 영역”으로 기재, 그 외에는 정부업무분류체계(BRM) 1단계 수준으로 분류하여 기재
	표준단어명	<ul style="list-style-type: none"> 표준단어에 대한 한글명을 기재
	단어 영문명	<ul style="list-style-type: none"> 표준단어에 대한 영문전체명(full name)을 기재
	단어 영문약어명	<ul style="list-style-type: none"> 표준단어에 대한 영문약어명을 기재
	단어 설명	<ul style="list-style-type: none"> 표준단어에 대한 정의 및 설명을 기재
	형식단어 여부	<ul style="list-style-type: none"> 해당 표준단어가 도메인 특성을 지닌 형식단어(분류어)인지 여부를 기재
	도메인 분류명	<ul style="list-style-type: none"> 형식단어(분류어)인 경우 해당되는 표준도메인의 분류명을 기재
표준 도메인	이음동의어 목록	<ul style="list-style-type: none"> 해당 표준단어에 대한 이음동의어 목록을 기재 이음동의어 : 소리는 다르나 의미가 동일한 단어 (예: 연도/년도, 비율/율) 이음동의어는 공통표준단어와 한글명 이외에는 동일하여 한글명만 관리
	금칙어 목록	<ul style="list-style-type: none"> 해당 표준단어에 대한 금칙어 목록을 기재 금칙어 : 사용이 허락되지 않거나 일정기간 사용하다가 특정 시점 이후 사용이 중지된 단어 예: 파일 (표준어) ↔ 화일 (금칙어. 사용X) 금칙어는 표준단어 정의대상이 아니며 금칙어의 한글명만 관리
표준 도메인	표준도메인 그룹명	<ul style="list-style-type: none"> 표준도메인의 상위 구분을 기재 예: 날짜/시간
	도메인 분류명	<ul style="list-style-type: none"> 표준도메인의 세부 구분을 기재 예: 연월

구분	항목명	항목 정의 및 작성 지침
표준 도메인	도메인명	<ul style="list-style-type: none"> 해당 표준도메인명으로 “표준도메인분류명 + 데이터 타입 + 길이(선택)” 형태로 정의 <ul style="list-style-type: none"> 예: 연도C4
	도메인 설명	<ul style="list-style-type: none"> 해당 표준도메인에 대한 정의 및 설명을 기재
	데이터 타입	<ul style="list-style-type: none"> 해당 표준도메인이 가지는 데이터 타입을 기재 <ul style="list-style-type: none"> 예: 문자형(CHAR, VARCHAR), 숫자형(NUMERIC), 날짜형(DATE-TIME)
	데이터 길이	<ul style="list-style-type: none"> 해당 표준도메인이 가지는 데이터 값의 길이, 소수점이 존재하면 소수점 자리까지 포함한 길이를 기재 <ul style="list-style-type: none"> ※ 바이트(Byte) 단위의 길이로 한글은 2바이트 기준
	소수점 길이	<ul style="list-style-type: none"> 해당 표준도메인의 데이터 값이 가질 수 있는 소수점 이하의 최대 자릿 수를 기재(소수 값이 존재하고, 데이터 유형이 숫자형일 때만 정의)
	저장 형식	<ul style="list-style-type: none"> 해당 표준도메인의 데이터 값을 저장하는 형식을 기재 <ul style="list-style-type: none"> 예: 1999년 9월 → YYYYMM (199909)
	표현 형식	<ul style="list-style-type: none"> 해당 표준도메인의 데이터를 제공할 때 표현하는 형식을 기재 <ul style="list-style-type: none"> 예: 1999년 9월 → YYYY-MM (1999-09)
	단위	<ul style="list-style-type: none"> 해당 표준도메인의 데이터 값을 측정하는 단위를 기재 <ul style="list-style-type: none"> 예: m, kg, 도 등
표준 코드	허용값	<ul style="list-style-type: none"> 해당 표준도메인이 가질 수 있는 최댓값/최솟값이나 유효값을 기재 <ul style="list-style-type: none"> 예: 월 : 01 ~ 12, 여부 : 'Y', 'N'
	관리부서명	<ul style="list-style-type: none"> 코드의 생성, 변경 등을 관리하는 기관의 부서 정보를 기재
	한글코드명	<ul style="list-style-type: none"> 표준용어를 준수하여 부여할 표준코드의 한글명을 기재
	영문코드명	<ul style="list-style-type: none"> 표준용어를 준수하여 부여할 표준코드의 영문명을 기재
	코드 설명	<ul style="list-style-type: none"> 코드의 적용 범위, 설명, 예외사항 등 코드의 정보를 기재
	데이터 타입	<ul style="list-style-type: none"> 코드 값을 표현하기 위한 데이터 타입을 기재 <ul style="list-style-type: none"> 예: 문자형(고정길이)은 CHARACTER, CHAR 등으로 기재 예: 숫자형(정수)은 NUMERIC(또는 NUM), DECIMAL(또는 DEC), INTEGER(또는 INT), SMALLINT(또는 SINT) 등으로 기재
	데이터 길이	<ul style="list-style-type: none"> 코드 값을 표현하기 위한 데이터의 최대 길이를 숫자로 표현
	코드값	<ul style="list-style-type: none"> 해당 코드가 가질 수 있는 허용 가능한 값의 집합이나 범위를 기재
	코드값 의미	<ul style="list-style-type: none"> 대상 코드값의 의미를 기재 <ul style="list-style-type: none"> 예: 코드명 : 가족관계, 코드값 : 001, 코드값 의미 : 본인

없다. 머리말에서 주석으로 언급한 「행정기관의 코드표준화 추진지침」의 행정 표준코드가 있기 때문이다. 따라서 DB표준코드와 기관표준코드는 행정표준코드를 준수해야 한다. 그리고 이것이 곧 세 번째 핵심사항이 되는 것이다.

2013년 「공공데이터법」이 시행되면서부터 데이터 표준화는 각 공공기관의 의무사항이었고 이후로도 데이터 비표준화 상황의 개선을 촉구하는 목소리가 이어져 왔지만, 10년이 지난 아직도 한국학 분야는 말할 것도 없고 각 분야 대다수 기관의 데이터 표준화 상황은 미흡하기만 하다.³⁰

현재 데이터 표준화를 추진하는 정책은 ‘공공데이터 품질관리 수준진단·평가’의 일환으로 진행되고 있는데, 위와 같은 상황을 감안하여 이 평가는 사전에 ‘예방적 품질관리 진단’이라는 선제적 품질관리 시책을 펴고 있다.³¹ 즉, 2단계에 걸친 진단·평가가 이루어지는 것이다. 이는 예방적 품질관리 진단을 수행한 후 그 결과(점수)를 본격적인 품질관리 수준진단·평가의 일부 관련 항목에 반영하는 방식으로 진행된다.³² 그리고 2022년부터 이 품질관리 수준진

30 행정안전부, 「공공데이터의 제공 및 이용 활성화에 관한 법률」[시행 2013. 10. 31.], 법제처 국가법령정보센터([https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/\(11956,20130730\)](https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률/(11956,20130730))); 한억수, 「공공데이터 개방 현황 및 이용 활성화 방안」, 『한국콘텐츠학회 종합학술대회 논문집』, 2018-5(2018), 78쪽; 한상규·김은진·김희웅, 「공공데이터의 표준 비표준 비교를 통한 표준화 방안 연구」, 『한국경영정보학회 학술대회논문집』, 2019-5(2019), 63쪽; 조윤희·이용건, 앞의 논문, 798쪽; 행정안전부(편), 앞의 책(2023d), 2쪽, 6~7쪽, 30~31쪽 참조.

31 예방적 품질관리의 필요성을 더 구체적으로 살펴보면 다음과 같다. 이미 구축된 DB는 여러 가지 이유로 데이터 품질 저하 요인의 근본적 원인을 해소하거나 즉각적으로 품질을 개선하기에 한계가 있다. 이러한 이유에서 고품질 공공데이터 제공 및 활용을 위해서는 정보시스템 계획 단계부터 선제적인 품질관리를 시행하여 양질의 데이터를 확보하고 관리할 수 있도록 점검을 추진할 필요가 있다. 행정안전부, 「공공데이터법」 제22조(공공데이터의 품질관리); 행정안전부(편), 『공공데이터 품질관리 중장기계획』(서울: 공공데이터전략위원회, 2019), 10쪽; 행정안전부(편), 앞의 책(2023d), 26쪽 참조.

32 더 넓게 조망하자면 공공데이터 품질관리 수준진단·평가 또한 ‘공공데이터 제공 운영실태 평가’의 일환이 된다. 운영실태 평가는 시스템이나 데이터가 아닌 그것들을 운영·관리 주체

단·평가 결과는 각 공공기관의 경영평가에 반영되기 시작했다.

예방적 품질관리 진단은 4개 검토 영역에서 진행이 되는데, 그중 첫 번째가 ‘데이터 표준’이다. 이 영역의 검토 항목은 두 가지인데, ‘데이터 표준화 및 상위표준 준수’와 ‘데이터 표준 관리’이다. 전자의 세부 기준으로는 ‘데이터 표준 코드, 용어, 단어, 도메인 등 데이터 표준화 사항 반영 여부’와 ‘공통 표준 및 기관표준 준용 여부’가 있고, 후자의 세부 기준으로는 ‘데이터 표준 관리방안 반영 여부’와 ‘데이터 표준 변경이력 관리방안 반영 여부’가 있다.³³

이후 품질관리 수준진단·평가에서 예방적 품질관리 진단은 정보시스템(DB) 구축 계획 수립 단계, 사업 추진(발주) 단계, 사업 완료 단계 등 세 단계에 대해 진단·평가가 이루어진다.³⁴

그럼 3절에서 데이터 표준화와 관련하여 예방적 품질관리 진단 사항까지 반영된 총체적 ‘공공데이터 품질관리 수준진단·평가’의 체계를 핵심적으로 간추리면서 한국학 분야에서 관찰할 수 있는 실정을 살펴보도록 하자.

3. 공공데이터 품질관리 수준진단·평가

‘공공데이터 품질관리 수준진단·평가’는 「공공데이터법」 제22조(공공데이터 품질관리) 및 동법 시행령 제17조(품질진단·개선) 등에 근거하여 “공공

인 기관을 대상으로 공공데이터 제공 운영실태를 평가하는 체계이다. 운영실태 평가지표 5개 영역 중 4번째 영역으로 품질관리 수준진단·평가 결과가 반영되는 형태이다. 행정안전부(편), 『2023년 공공데이터 제공 운영실태 평가 편람』(세종: 행정안전부, 2023b), 6쪽; 행정안전부(편), 『2023년 공공데이터 제공 운영실태 평가 결과』(서울: 공공데이터전략위원회, 2024), 38쪽 참조.

33 행정안전부(편), 앞의 책(2023c), 75~81쪽 참조.

34 위의 책, 4~17쪽; 행정안전부(편), 『공공데이터 제공·관리 실무 매뉴얼』(세종: 행정안전부, 2023f), 54~55쪽 참조.

기관이 생성·취득하여 관리하는 공공데이터의 품질을 확보하기 위하여 체계적인 품질관리 활동을 수행하는지 여부를 진단”할 목적으로 700개 기관(2023년 기준)을 대상으로 행정안전부가 총괄하여 실시하는 조치이다.³⁵

평가 대상 기관은 현행 보유 DB 중 평가 대상이 될 DB를 선정하고, 이들을 대상으로 기관별 자체진단·평가를 진행한 후 각 평가지표에 대한 증빙자료를 등록한다. 이후 각종 오류를 진단·조치한다. 그에는 중에 한국지능정보사회진흥원(NIA)의 공공데이터활용지원센터에서는 기관별 자체진단·평가 결과를 확인·점검하고, 필요 시 기관 자체평가 및 품질관리 활동을 지원하거나 개선 방향을 안내한다. 그리고 나면 수준평가 심의위원회는 진단·평가 결과에 대한 검증심의 후 결과를 확정한다. 그리고 이 결과는 공공데이터 제공 운영실태 평가에 반영되고, 공공데이터전략위원회 및 국무회의에도 보고된다.³⁶

진단·평가항목은 데이터 관리체계 및 데이터 값 관리라는 2개 영역의 11개 지표로 구성되는데, 데이터 표준화와 관련된 영역은 데이터 관리체계 영역이다. 이 영역의 평가지표 여섯 가지 중 데이터 표준화와 직결되는 것이 바로 네 번째 평가지표인 ‘데이터 표준 확산’이다. 그리고 이 평가지표는 두 가지 세부지표, 즉 ‘데이터 표준 정의’와 ‘데이터 표준 적용률’로 구성된다.³⁷

‘데이터 표준 정의’ 지표는 “기관의 품질관리 수준진단·평가 대상 DB에 대하여 공통표준 및 기관표준을 준용하여 데이터 표준을 정의하였는지 여부”를 측정한다. 여기에서의 데이터 표준 정의란 앞에서 제시한 데이터 표준사전 관리항목을 말한다. 그리고 ‘데이터 표준 적용률’ 지표는 “품질관리 수준 진단·평가 대상 DB(컬럼)에 데이터베이스 표준(DB표준)을 적용하였는지 여

35 행정안전부(편), 앞의 책(2023c), 2쪽, 7쪽.

36 위의 책, 8~9쪽 참조.

37 위의 책, 3~6쪽 참조.

부”를 측정한다. 여기에서의 적용율은 백분율(%)로 산정된다.³⁸

앞서 ‘데이터 표준화 및 상위표준 준수’라는 집약적 핵심사항을 거론했는데, 이는 아직 시작 단계라고 해도 과언이 아니다. 일례로 데이터 표준사전 관리항목 네 가지 중 기본 단위라고 할 수 있는 표준단어의 정의 여부는 2022년까지만 해도 공공데이터 품질관리 수준진단·평가에 포함되지 않았다.³⁹

아직 시작 단계나 다름없는 데이터 표준화의 정책 기조와 현 실정에서, 한국학 데이터 표준화의 현황을 파악하기 위해서는 공통표준에 대한 세밀한 분석과 더불어 기관표준과 DB표준 차원에서 데이터 표준화를 위한 지침 수준의 요구사항과 현재 이루어지고 있는 공공데이터 부문의 품질관리 평가 수준 및 그 결과, 그리고 실제 정의된 기관표준과 DB표준 내역들을 상세하게 비교해 볼 필요가 있다. 하지만 이 글에서 정리한 내용을 넘어서는 구체적·수치적 데이터를 제시하는 것은 「공공기관의 정보공개에 관한 법률」(약칭 「정보공개법」)상 불가하거나 기관 간의 입장 문제로 인하여 민감한 사안이 될 수 있기 때문에 다음과 같은 정도에서 현황을 간단히 언급하는 것으로 대신하려 한다.

〈표2〉는 데이터 표준화가 포함되어 있는 데이터 품질관리 부문에서 나온 공공데이터 제공 운영실태 평가의 최근 3개년 결과이다.⁴⁰ ‘기타 공공기관’ 속에는 한국학 관련 몇몇 기관들도 속해 있다. 이 중에는 상당히 우수한 점수를 받은 기관도 확인된다. 그럼에도 불구하고 평균점수가 이 정도임을 감안

38 위의 책, 17~20쪽; 행정안전부(편), 앞의 책(2023f), 53쪽 참조.

39 행정안전부(편), 『2022년 공공데이터 품질관리 수준진단·평가 매뉴얼』(대구: 한국지능정보사회진흥원, 2022b), 12쪽 참조.

40 행정안전부(편), 앞의 책(2024), 본문 5쪽 참조.

표2-공공데이터 제공 운영실태 평가 결과(데이터 품질관리 부문)

기관 범위	평균점수(100점 만점)		
	2021년	2022년	2023년 ⁴¹
전체 평가 대상 기관 ⁴²	54.9	61.6	58.1
기타 공공기관 ⁴³	37.4	51.9	50.4

하면 여타 기관의 경우 어느 정도 수준일지 가늠이 된다.

나아가야 할 길이 혐난함을 알고 있기에 아직은 다음과 같은 융통성의 여지도 마련되어 있다. “원칙적으로 모든 정보화사업을 대상으로 공공데이터 베이스 표준의 적용 및 관리가 이루어져야 하나, 표준 적용이 어려운 경우에는 표준의 적용 여부 및 범위를 자체적으로 검토하여 선별적으로 적용할 수 있”기 때문이다.⁴⁴ 그리고 이것은 현재 수준에서 우리가 그나마 기댈 수 있는 언덕이고, 이것을 잘 활용하여 수준진단·평가를 그런대로 무난히 받도록 하

41 2022년 대비 2023년 점수가 하락한 데에는 데이터 품질관리 대상 DB의 범위가 기관 전체 DB의 40%(2022년)에서 50%(2023년)로 확대되면서 품질관리 난이도가 상승한 것이 주요 원인으로 꼽힌다. 행정안전부(편), 『2023년 공공데이터 제공 운영실태 평가결과』(서울: 공공데이터전략위원회, 2024), 5쪽 참조.

42 2023년도에는 696개 기관(중앙행정기관 45, 광역자치단체 17, 기초자치단체 226, 공공기관 391, 교육행정 17), 2022년도에는 570개 기관(중앙행정기관 45, 광역자치단체 17, 기초자치단체 226, 공공기관 282), 2021년도에는 548개 기관(중앙행정기관 45, 광역자치단체 17, 기초자치단체 226, 공공기관 260)이 해당된다. 행정안전부(편), 앞의 책(2024), 본문 1쪽; 행정안전부(편), 『2022년 공공데이터 제공 운영실태 평가결과』(서울: 공공데이터전략위원회, 2023a), 본문 1쪽; 행정안전부(편), 『2021년 공공데이터 제공 운영실태 평가결과』(서울: 공공데이터전략위원회, 2022a), 1쪽 참조.

43 2023년도에는 391개 공공기관 중 258개가 기타공공기관이고, 2022년도에는 282개 공공기관 중 152개가 기타공공기관 및 지방공기업이다. 2021년도에는 260개 공공기관 중 128개가 기타공공기관 및 지방공기업이다. 행정안전부(편), 앞의 책(2024), 본문 1쪽, 25쪽; 행정안전부(편), 앞의 책(2023a), 20쪽; 행정안전부(편), 앞의 책(2022a), 7쪽 참조.

44 행정안전부(편), 앞의 책(2023d), 92쪽 참조.

는 방안이 현실적 최선책이다. 실제 최근 수준진단·평가를 높게 받은 기관들의 상당수는 이 방식을 전략적으로 채택한 것이 유효하게 작용했다고 볼 수 있다.

하지만 이러한 전략이 언제까지나 유효할 것이라고 안도할 수는 없다. 앞서도 언급했지만 요 몇 해 사이의 추세만 보아도 해가 다르게 정책과 평가의 기조가 엄격해지고 있다. ‘작년까지는 대략 이 정도로 조치해도 무리 없이 잘 넘어갔는데, 금년에 요구되는 사항은 어떻게 대처해야 할지 난감하다.’는식의 말이 해마다 여기저기에서 되풀이되고 있다. 이는 비단 기관의 업무담당자나 실무담당자만의 토로가 아니다.⁴⁵ 공공데이터활용지원센터의 지원 인력들도 각 기관들의 다양한 상황에서 제기되는 문제점과 질의에 대처하느라 분투하고 있다.⁴⁶ 즉, 위와 같은 전략도 그 융통성의 범위가 점차 줄어들 것이라는 점은 자명해 보인다.

요컨대, 정부의 의지는 그야말로 드세다. 반면 인문학 연구기관의 의욕은 망연자실에 버금가는 수준이다. 마냥 이렇게 있을 수도 없고, 있어서도 안 되는 게 현 상황이다.

다만 이는 전제한 바와 같이 적극성의 측면에서 말한 것이고, 그렇게 주도적인 과단성을 발휘하기에 적절치 않은 경우를 상정하지 않을 수 없다. 그래서 ‘통합플랫폼의 온정적 표준화’가 필요하다.

45 공공데이터 부분에서는 ‘실무담당자’와 ‘업무담당자’가 서로 다른 층위의 용어로 쓰이고 있다. 개별 정보시스템의 DB 및 공공데이터를 담당하는 사람을 ‘업무담당자’로 칭하고, 여러 정보시스템을 운영하는 기관에서 모든 정보시스템 DB들을 총괄하여 공공데이터 처리를 담당하는 사람을 ‘실무담당자’라고 칭한다. 기관별 사정에 따라서 ‘업무담당자’와 ‘실무담당자’를 겸하는 경우도 있다.

46 “공공데이터의 효율적인 제공 및 이용 활성화 지원을 위하여” 공공데이터활용지원센터가 한국지능정보사회진흥원(NIA)에 설치되어 운영 중이다. 행정안전부, 「공공데이터의 제공 및 이용 활성화에 관한 법률」 제13조.

IV. 맷음말: 상보적 역할의 지평

데이터 표준화를 위한 법적인 통제 체계 외에 또 다른 편의 적극적인 조치로서 온정적인 지원과 함께 표준화를 유도하는 방법이 있고, 이는 ‘통합플랫폼’이라는 서비스 미디어를 통해서 시도해 볼 수 있다.

다만 여기에서 말하는 ‘통합플랫폼’은 여러 곳에서 ‘통합플랫폼’이라는 표현을 넣어 만든 모든 서비스 채널을 포괄해서 이르는 것이 아니라 ‘국가지식 정보 통합플랫폼’(디지털집현전)과 ‘한국학자료통합플랫폼’의 두 가지만을 아우르는 것이다.⁴⁷

통합플랫폼은 공공데이터 정책에 비해 보다 사용자 친화적인 방법으로 데이터 표준화를 유도할 수 있다. 통합플랫폼 양자 간의 온도차는 있지만, 일단 공공데이터 정책처럼 진단·평가와 같은 부담스러운 체계로 만나는 것이 아닐 뿐만 아니라 사이트 간 데이터 연계의 과정에서 배려·포용·포섭을 통해 데이터를 집단적 합의에 의해 표준화하는 과정이 자연스럽게 진행될 수 있기 때문이다. 이러한 과정이 곧 ‘통합플랫폼의 온정적 표준화’이며 그 구체적인 체계는 별도의 후속 논고에서 논의될 것이다.

한편 한국학자료통합플랫폼과 디지털집현전은 상호 연관성이 깊어 보이지만, 서비스 기획과 구축의 과정이 다를 뿐만 아니라 조직과 체계도 전혀 다른 기반에서 구축된 서비스 플랫폼들이다. 서비스 연계 방법 차원에서도 이 둘은 확연히 구분되는 스탠스를 취하고 있다. 이 점은 뒤이어질 논고에서 ‘온정적 표준화’의 온도차를 드러내는 이랑·고랑이 된다.

47 이 글을 집필하는 현 시점 기준으로 고려할 수 있는 범위가 그러하다. 과학기술정보통신부 디지털집현전(<https://k-knowledge.kr>); 한국학중앙연구원 한국학자료통합플랫폼 (<https://kdp.aks.ac.kr>). 서비스 개시 시점은 한국학자료통합플랫폼이 2023년 4월, 디지털집현전이 2024년 1월이다.

마지막으로 이 글의 논지와 더불어 후속 연구까지 포괄하는 조감도를 간략히 제시하는 것으로 논의를 일단락 지으려 한다. 이 글은 ‘한국학 데이터 표준화 방안’ 연구의 일환이다. 이 연구는 한국학 데이터 표준화의 현황을 탐색하는 과정에서 도출된 ‘통합플랫폼과 공공데이터의 상보적 역할’이라는 지형도에서 시작했고,⁴⁸ 그 논고 중 일부로서 공공데이터 정책을 중심으로 하는 시론을 내게 되었다.

한국학 데이터 표준화는 학계의 시장 과정에 의한 표준화보다는 집단적 합의에 의한 표준화의 일환으로¹⁶ 정부의 공공데이터 정책에 따라 최소한의 형식적 표준화(표1 및 상위표준 준수)가 종용되고 있지만 아직은 시작에 불과한 실정이다.⁴³

공공데이터의 적극적 표준화와 통합플랫폼의 온정적 표준화 간의 상보적 역할은 이 시점에서 유효한 구도를 형성함과 동시에 명시적 의도의 유무와는 무관하게 필연적으로 거쳐 갈 수밖에 없는 불가피한 체계일 수 있다. 그리고 그것은 한국학 데이터에 대한 ‘세계적 표준’을 지향해야 한다. 이를 위해서는 한국학과 인접한 국내외 학문 영역에서의 표준과 조화를 이룰 수 있어야 하고, 한국학만이 가진 영역에서는 독자적인 표준을 정의해야 한다. 공공데이터 정책 관련해서는 ‘데이터 표준사전 관리항목’(표1)과 관련 상위표준의 준수에서 더 나아가 상위표준을 보완하는 단계까지 나아가야 하고, 통합플랫폼 관련해서는 공공데이터 정책의 준수가 용이하도록 협조하되 정책 측면에서 취급하기 어려운 부문의 데이터에 대한 데파토 표준을 형성하는 단계까지 나아가야 한다.¹⁵

48 이 양자가 어우러진 방식을 필자는 ‘상보적 양방 몰이’라고 표현한다. 이러한 내용이 실제 기획되어 있었는지 여부는 명시된 바가 없기에 확인할 수가 없다. 이 방식은 그저 필자가 현 상태를 관찰하고 정책 및 이론을 분석하면서 방안을 연구하다가 도출된 해석 결과로서의 지형도일 뿐이다.

이러한 과정은 기존에 통용되어 온 한국학 데이터를 표준화한다는 차원에서 찾을 수 있는 의의를 넘어, 현재 성장하고 있는 차세대 한국학 데이터를 표준화하기 위한 전략을 수립하는 데에 일조할 수도 있다는 점에서 그 잠재적 의의를 기대할 수 있다.⁴⁹

49 차세대 한국학 데이터란 시맨틱 데이터와 인공지능·메타버스 등 첨단 영역까지 아우르는 기술 로드맵과 더불어 사용될 지능화된 한국학 데이터라고 할 수 있다.

참고문헌

1. 1차 자료

- 행정안전부(편), 『(편리한 활용과 용·복합 촉진을 위한) 공공데이터 품질관리 중장기계획』, 서울: 공공데이터전략위원회, 2019.
- 행정안전부(편), 『2021년 공공데이터 제공 운영실태 평가결과』, 서울: 공공데이터전략위원회, 2022a.
- 행정안전부(편), 『2022년 공공데이터 품질관리 수준진단·평가 매뉴얼』, 대구: 한국지능정보사회진흥원, 2022b.
- 행정안전부(편), 『2022년 공공데이터 제공 운영실태 평가결과』, 서울: 공공데이터전략위원회, 2023a.
- 행정안전부(편), 『2023년 공공데이터 제공 운영실태 평가 편람』, 세종: 행정안전부, 2023b.
- 행정안전부(편), 『2023년 공공데이터 품질관리 수준진단·평가 매뉴얼』, 대구: 한국지능정보사회진흥원, 2023c.
- 행정안전부(편), 『공공데이터베이스 표준화 관리 매뉴얼』, 대구: 한국지능정보사회진흥원, 2023d.
- 행정안전부(편), 『공공데이터 제공 표준』, 대구: 한국지능정보사회진흥원, 2023e.
- 행정안전부(편), 『공공데이터 제공·관리 실무 매뉴얼』, 세종: 행정안전부, 2023f.
- 행정안전부(편), 『2023년 공공데이터 제공 운영실태 평가결과』, 서울: 공공데이터전략위원회, 2024.

2. 논저

- 김바로, 「〈공공데이터법〉과 인문데이터: 공공기관 보유 인문데이터 공개 신청 사례를 중심으로」, 『한국고전연구』 57, 2022.
- 김현, 『인문정보학의 모색』, 성남: 북코리아, 2012.
- 김현·임영상·김바로, 『디지털 인문학 입문』, 서울: 한국외국어대학교 지식출판원, 2016.
- 단국대학교 부설 동양학연구소(편), 『한한대사전 7』, 서울: 단국대학교출판부, 2004.
- 류인태, 「인문학이 데이터와 만날 때: 재현-연결-표준으로서 인문학 데이터의 내포를 중심으로」, 『동서인문학』 65, 2023.
- 성태경, 『표준의 경제학: 이론·사례·정책』, 파주: 한국학술정보, 2012.

양창진, 「정보화에세이: 정보통신 기술과 인문학의 만남, 그리고 상생」, 『지역정보화』 51, 2008.

오영교, 「한국학·지역학의 발흥과 한국사연구」, 『한국사연구』 200, 2023.

이강원·손호웅, 『지형 공간정보체계 용어사전』, 서울: 구미서관.

이남희, 「인문학과 지식정보화: 『지식정보자원관리법』과 『한국역사정보통합시스템』을 중심으로」, 『인문콘텐츠』 1, 2003.

이남희·김석근, 『지식정보화 관련 법령 분석과 인문학 진흥을 위한 정책제안 연구』, 서울: 한국교육개발원 인문사회연구회, 2002.

장석권, 『데이터를 철학하다: 어떻게 데이터는 지혜가 되는가』, 서울: 흐름, 2018.

전훈지, 「한국학의 정의와 방법에 대한 고찰」: 김경일 저, 『한국의 근대 형상과 한국학 - 비교 역사의 시각』의 서평, 『비교한국학』 29-2, 2021.

조윤휘·이용건, 「공공데이터 주요 성과와 제4차('23~'25) 기본계획 추진방향」, 『한국행정학회 학술발표논문집』 2022-2, 2022.

지동은·SANTOSH KUMAR RANJAN, 「북한과 중국에서 조선어(한국어)정보화 현황과 과제」, 제51회 한중인문학회 국제학술대회, 서울: 서울대학교, 2023.5.13.

최정도, 「인문학 위기 속의 국어정보학의 현황과 전망」, 『국어국문학』 201, 2022.

한국데이터산업진흥원(편), 『데이터아키텍처 전문가 가이드』, 서울: 한국데이터산업진흥원, 2021.

한상규·김은진·김희웅, 「공공데이터의 표준 비표준 비교를 통한 표준화 방안 연구」, 『한국경영정보학회 학술대회논문집』 2019-5, 2019.

한여수, 「공공데이터 개방 현황 및 이용 활성화 방안」, 『한국콘텐츠학회 종합학술대회 논문집』 2018-5, 2018.

潘燕杰, “我国健康医疗数据治理的标准化路径: 以美国相关法律为借鉴,” *Journal of Nanjing Medical University: Social Sciences*, Issue 5, 2023, pp. 423~429.

Fuhr, Antonie, Andreas Kurtz, Christian Hiepen, Sabine Müller, “Organoids as Miniature Twins-Challenges for Comparability and Need for Data Standardization and Access,” *Organoids*, Vol. 1, Issue 1, 2022.

Fukami, Yoshiaki, “Standardization Procedure for Data Exchange,” *Information*, Vol. 11, Issue 6, 2020.

Gal, Michal S., Daniel L. Rubinfeld, "Data Standardization," *New York University Law Review*, Vol. 94, Issue 4, 2019.

Кацу, А. С., Т. М. Александрова, П. С. Талапова, І. Б. Агєєва, М. М. Вєль, М. О. Трофименко, М. Р. Колесник, Т. С. Несміян, "СУЧАСНИЙ СТАН ТА ПЕРСПЕКТИВИ ВПРОВАДЖЕННЯ СТАНДАРТИЗАЦІЇ ДАНИХ У СИСТЕМУ ОХОРОНИ ЗДОРОВ'Я УКРАЇНИ (огляд літератури)," *Medični Perypektyvi*, Vol. 28, Issue 3, 2023.

3. 기타

과학기술정보통신부 디지털집현전(<https://k-knowledge.kr>).

국립국어원 표준국어대사전(<https://stdict.korean.go.kr>).

김현, 인문정보학 개론 2024(http://dh.aks.ac.kr/Edu/wiki/index.php/인문정보학_개론_2024).

두산 Doopedia(<https://www.doopedia.co.kr>).

법제처 국가법령정보센터(<https://law.go.kr>).

한국국학진흥원, [2022 전통 기록문화 창작 콘퍼런스] 발표 ② 한국학 자료의 큐레이션과 그 사례: 양창진 관장(한국학중앙연구원), <https://youtu.be/duiRHQmfjXY?si=2mXaBMPc8UAg05yD>.

한국학중앙연구원 홈페이지(<https://www.aks.ac.kr>).

한국학중앙연구원 한국민족문화대백과사전(<https://encykorea.aks.ac.kr>).

한국학중앙연구원 한국학자료통합플랫폼(<https://kdp.aks.ac.kr>).

한국학중앙연구원 디지털인문학연구소 Ontology:EKC 2022(https://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022).

Oxford University Press Oxford English Dictionary(<https://www.oed.com>).

국문초록

법으로 규정된 데이터 표준화의 의무적 상황은 전 세계적으로 활발하게 진행되고 있는 각 분야별 데이터 표준화 연구 추세와 궤를 같이한다. 이 글은 ‘한국학 데이터 표준화 방안’ 연구의 일환으로서, ‘통합플랫폼과 공공데이터의 상보적 역할’이라는 지형도를 기반으로 후속 연구로 이어질 예정이다.

한국학 데이터 표준화는 학계의 시장 과정에 의한 표준화보다는 집단적 합의에 의한 표준화의 일환으로 정부의 공공데이터 정책에 따라 최소한의 형식적 표준화가 종용되고 있지만 아직은 시작에 불과한 실정이다. 이 시론은 이러한 실정의 저변을 체계적으로 집약 정리하는 차원에서 각종 데이터 표준화 관련 개념과 분류 체계 및 정책 환경에 대해서 검토하고 한국학 분야의 데이터 표준화 현황을 공공데이터 정책과 함께 살펴본 후 이를 더욱 효과적으로 전개시킬 수 있는 역할 체계의 일면을 제시한다.

통합플랫폼과 공공데이터의 상보적 진형(陣形)은 한국학 데이터에 대한 ‘세계적 표준’을 지향해야 한다. 이를 위해서는 한국학과 인접한 국내외 학문 영역에서의 표준과 조화를 이룰 수 있어야 하고, 한국학만이 가진 영역에서는 독자적인 표준을 정의해야 한다.

투고일 2024. 3. 17.

심사일 2024. 4. 16.

제재 확정일 2024. 5. 9.

주제어(keywords) 데이터 표준화(data standardization), 디지털 인문학(digital humanities), 공공데이터(public data), 한국학자료통합플랫폼(Korean Studies Data Platform), 디지털집현전(K-knowledge Platform)

Abstract

Approaches to Standardizing Korean Studies Data: Policy of Public Data

Seo, Dong Shin

This work is part of a study on “Methods for standardizing Korean studies data.” This study began with the topographical map of “the complementary role of integrated platforms and public data” derived in the process of exploring the current status of Korean studies data standardization. As part of the study, this work focuses on the role of public data policy.

In line with the trend of data standardization research in each field that is actively underway around the world, this work reviews various data standardization-related concepts, classification systems, and policy environments, examines the current status of data standardization in the field of Korean studies along with public data policy, and then presents a role system that can develop this more effectively.

To determine the status of data standardization in the specialized field of Korean studies, one’s own system needs to be created by narrowing down only the stems necessary for the issue in the complex maze. In terms of carrying out this task, this work also has considerable significance.