

## 인공지능 학습 데이터로서 공공연구데이터 활용 촉진을 위한 시론적 연구\*

### A Study on Promoting the Re-Use of Public Research Data as AI(Artificial Intelligence) Learning Data

이 재 훈\*\*  
Lee, Jae-Hoon

#### 목 차

- I. 처음에
- II. 인공지능 학습데이터로서 공공데이터와 공공연구데이터  
제공 법제
- III. 공공연구데이터 활용을 위한 공공데이터 법제 개선방향
- IV. 끝으로

#### 국문초록

다양한 현대 기술 분야의 경쟁력 제고를 위한 처음과 마지막은 우수한 인공지능 기술의 확보라고 할 수 있다. 우수한 인공지능 기술 확보가 중요한 현시점에서 인공지능 기술혁신을 위한 국가적 지원과 이를 규범적으로 구체화하고 있는 법제가 필요하다. 현재 우리나라에는 인공지능 기술 개발 지원을 위해 활용될 수 있는 법제가 존재한다. 대표적으로 지능형로봇법과 초고성능컴퓨터법이

논문접수일 : 2020. 07. 16.

심사완료일 : 2020. 08. 11.

게재확정일 : 2020. 08. 11.

\* 이 논문은 2020년 6월 12일(금) 한국재산법학회와 한국법제연구원이 공동 주최한 학술대회에서 이루어진 학술발표의 발표문을 수정·보완한 것이다.

\*\* 법학박사·한국법제연구원 부연구위원

인공지능 기술 개발 지원을 위해 활용될 수 있다. 하지만 인공지능 학습 데이터 확보가 인공지능 기술 개발의 핵심인 상황에도 불구하고, 지능형로봇법과 초고성능컴퓨터법에는 학습데이터 확보와 관련된 지원 방식이 규정되어 있지 않아 제도적 지원의 한계가 발생한다.

해외 주요 국가의 인공지능 전략과 우리나라의 인공지능 전략을 살펴보면 인공지능 기술 개발을 위한 학습데이터 제공이라는 관점에서 공공데이터를 개방하는 정책이 검토되고 있다. 그리고 유럽연합의 경우 오픈데이터 및 공공분야 정보의 이용에 대한 지침을 통해 고품질의 공공데이터를 통한 인공지능 기술 개발 촉진을 지원하려고 노력 중이다. 오픈데이터 및 공공분야 정보의 이용에 대한 지침에서 특히 눈여겨 볼 사항은 고품질의 데이터인 공공연구기관의 연구데이터를 공공데이터로 규정하고 이를 개방하도록 하고 있다는 점과 이를 위한 원칙과 조건을 규정하고 있다는 점이다.

우리나라도 공공데이터로서 성격을 갖고 있는 다양한 연구데이터를 갖고 있다. 다만 공공데이터로서의 성격을 갖고 있는 연구데이터가 현행 공공데이터의 제공 및 이용 활성화에 관한 법률을 통해서 적극적으로 개방되어 활용되고 있지 않다. 이러한 이유로는 법적으로 공공기관의 연구데이터가 공공데이터임이 명확하게 명시되고 있지 않은 점을 들 수 있다. 그리고 공공기관의 연구데이터를 체계적으로 관리하기 위한 제도가 충분하게 마련되어 있지 않은 점도 그 이유로 들 수 있다. 이러한 문제점들을 제도적으로 해소하기 위해서 본 연구는 유럽연합의 오픈데이터 및 공공분야 정보의 이용에 대한 지침과 우리나라의 공공데이터의 제공 및 이용 활성화에 관한 법률, 국가연구개발사업의 관리 등에 관한 규정 등을 분석하고, 인공지능 학습 데이터로서 공공기관의 연구데이터를 적극적으로 개방 및 활용할 수 있기 위해서는 어떠한 제도적 개선이 필요한지 검토한다.

**주제어** : 공공데이터, 연구데이터, 인공지능, 데이터관리계획, PSI-지침

## 1. 처음에

2016년 클라우스 슈밥(Klaus Schwab)이 세계 경제포럼에서 제4차 산업혁명이라는 용어를 사용한 이후, 제4차 산업혁명이라는 용어가 유행하여 광범위하게 사용되었다. 그리고 최근에는 제4차 산업혁명이라는 용어가 등장하게 된 배경 기술 및 활용분야를 중심으로 각론적인 논의가 풍성하게 이루어지고 있는 것으로 보인다. 인공지능, 빅 데이터로 대표되는 데이터 사이언스, 자율주행자동차, 서비스 로봇 등 구체적인 요소기술의 개발·구현 및 활용, 개발된 요소기술과 각종 기술을 연계한 새로운 산업생태계 구축, 이를 위한 정책적 지원 방향, 국제화된 경쟁체제 속에서 기술 주도권 확보를 위한 전략, 그리고 새롭게 등장한 기술로 인해 발생할 수 있는 부작용으로 인한 사회적 충격을 최소화하는 방식 등이 이에 해당한다.

이와 같은 사회적 배경 하에, 이하에서는 인공지능 기술개발을 지원하기 위한 법제개선방안을 시론적 관점에서 검토해보도록 한다. 제4차 산업혁명과 관련된 기술요소 중 연구대상을 인공지능으로 선택한 이유는 제4차 산업혁명을 대표하는 다양한 기술들의 핵심 요소이자 기술과 사회의 공진화<sup>1)</sup>를 앞당길 핵심 기술이 인공지능이기 때문이다.<sup>2)</sup> 빅 데이터가 의미를 갖게 된 이유가 폭발적으로 발전한 컴퓨팅 능력을 배경으로 한다면, 이러한 폭발적인 컴퓨팅 능력의 기반이자 이를 보다 강화하기 위한 소프트웨어적 기반이 인공지능 기술이다. 자

- 1) 기술발전과 사회구조의 관계를 바라보는 방식은 크게 세 가지 방식으로 나눌 수 있다. 기술결정론은 사회구조는 기술을 통해 결정된다는 입장이며, 사회구성론은 사회적 조건을 반영하면서 기술이 발전한다는 입장이다. 이들 두 가지 입장은 한쪽이 다른 한쪽에 일방적인 영향을 준다는 것을 전제로 하고 있다. 이와 달리 공진화이론은 기술발전과 사회구조는 서로 영향을 주고받으며 함께 발전한다는 입장이다. 정재림/김상욱, “시스템다이나믹스를 활용한 기술과 사회의 공진화 연구”, 「한국경영학회 통합학술발표논문집」, 한국경영학회, 2009, 2~3면.
- 2) 한세억, “제4차 산업혁명시대 기술과 제도의 공진화”, 「서울행정학회 동계학술대회 발표논문집」, 서울행정학회, 2018, 5~23면. 이외에도 국내 다른 학계의 동향을 살펴보면 인공지능과 공진화의 대상으로서 인간을 검토하는 문헌들도 존재한다. 그 예로는, 정재현, “AI와 인간의 공진화共進化와 관련해서 본 다산 철학”, 「다산학」 제35호, (재)다산학술문화재단, 2019, 209~240면; 노진아, “인간과 기계의 공진화 - 인공지능 로봇틱스 아트 “제페토의 꿈”을 중심으로”, 「Contents Plus」 제16권 No.1, 한국영상학회, 2018, 83~97면 등.

을주행자동차의 경우, 다양한 센서를 통해 수집된 주행관련 데이터를 실시간으로 분석하여 자율주행차량이 도로에서 안전하게 주행할 수 있도록 하는 핵심기술은 알고리즘에 기반한 인공지능 기술이다.<sup>3)</sup> 각종 로봇 기술도 마찬가지이다. 다양한 하드웨어로부터 수집된 데이터를 빠른 시간 내에 분석하고 이를 기반으로 시의적절한 구동이 이루어지도록 하기 위해서는 알고리즘에 기반한 인공지능의 도움이 필요하다.<sup>4)</sup> ICT 기술을 통한 행정의 자동화<sup>5)</sup>와 관련된 논의의 주요 쟁점 중 하나도 인공지능이라고 할 수 있다.<sup>6)</sup> 이처럼 일상생활에서 직·간접적 경험하게 될 수 있는 영역 이외에도 인공지능은 다양한 영역에서 활용될 수 있다. 대표적으로 순수과학 분야에서는 기초 데이터 분석 및 이를 기반으로 한 과학적 결론 도출, 그리고 이러한 과학적 지식을 현실에서 활용하는 과정에서 인공지능이 다양한 형태로 활용될 수 있다.<sup>7)</sup>

이처럼 우수한 인공지능 기술 확보가 중요한 현시점에서 인공지능 기술혁신을 위한 국가적 지원과 이를 규범적으로 구체화하고 있는 법제가 필요하다는

3) 자율주행 알고리즘에 대한 윤리적 설계기준 도입을 위한 「자율주행자동차 상용화 촉진 및 지원에 관한 법률」 개정안 관련 법제 동향에 대해서는 이재훈, “자율주행자동차 관련 국내 법제 동향”, 「경제규제와 법」 제12권 제2호, 서울대학교 공익산업법센터, 2019, 266면.

4) 관세부처 합동, 「2020년 지능형 로봇 실행계획」, 2020, 8~9, 13~14, 17면 외 다수.

5) 이와 관련된 주요 국내 문헌으로는 김중권, “행정에 인공지능시스템 도입의 공법적 문제점”, 「법조」 통권 740호, 법조협회, 2020, 53~77면; 김중권, “인공지능시대에 알고리즘에 의한 행위조종과 가상적 행정행위에 관한 소고”, 「공법연구」 제48집 제3호, 한국공법학회, 2020, 287~312면; 김중권, “인공지능시대에 완전자동화 행정행위에 관한 소고”, 「법조」 통권 723호, 법조협회, 2017, 146~182면; 이재훈, “전자동화 행정행위에 관한 연구”, 「성균관법학」 제29권 제3호, 성균관대학교 법학연구원, 2017, 142~192면; 이재훈, “전자동화 행정행위 도입 후 독일 행정절차법제 논의의 전개양상”, 「공법학연구」 제19권 제4호, 한국비교공법학회, 2018, 481~515면.

6) 최근 알고리즘 기반 인공지능 등을 활용한 행정의 변화상을 담고 있는 법률안들이 속속 등장하고 있다. 대표적으로 박선숙의원등이 발의했던 전자정부법 일부개정법률안(의안번호 2022748)은 전자정부법 제2조 제17호에 “행정기관 등이 행정업무의 전자적 처리에 대하여 중요한 판단, 결정, 평가 또는 자문 등을 수행하는 데이터기반학습기술을 사용하는 지능형 의사결정 시스템을 “지능형 행정시스템”이라고 정의하고 전자정부법 제5조의5를 통해 지능형 행정시스템을 관리하는 것을 규율하고 있었으나, 임기만료로 폐기되었다. 그리고 현재 정부입법의 형태로 입법예고(법제처 공고 제2020-28호) 중인 행정기본법(안)은 자동적 처분에 대한 근거 규정을 도입하기 위해 제23조를 통해 “행정청은 법령으로 정하는 바에 따라 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템을 포함한다)으로 처분을 할 수 있다. 다만, 처분에 재량이 있는 경우는 제외한다.”는 규정을 도입하려고 하고 있다.

7) 구체적인 사례는 이하 II. 1.을 참고할 것.

점은 별도의 이견이 없이 받아들여 질 수 있을 것이다. 법과 혁신을 주제로 이루어진 다양한 선행 연구들에서도 지적되고 있는 바와 같이 법이 혁신을 제한하는 걸림돌로 작용할 수도 있지만 법을 통해 혁신이 촉진될 수 있고,<sup>8)</sup> 법제도는 혁신정책의 도구로서 활용될 수 있기 때문이다.<sup>9)</sup> 다만 인공지능 기술개발을 지원하기 위한 법제적 방안을 검토하려는 본고의 방향성과 관련하여, 이미 인공지능 기술개발 지원에 활용될 수 있는 법제가 존재하는데, 이를 굳이 검토할 필요성이 있는가라는 의문이 생길 수 있다. 실제로 우리나라 현행 법제를 살펴보면 인공지능 기술개발과 관련한 지원법제로 활용될 수 있는 개별 법률들이 존재하고 있기 때문이다.

우선 「지능형 로봇 개발 및 보급 촉진법」(이하 지능형로봇법이라 함)은 지능형 로봇의 개발과 보급 촉진, 그리고 이와 관련된 기반조성을 위하여, 기업에 대한 자금지원, 지능형 로봇투자회사에 대한 「자본시장과 금융투자업에 관한 법률」의 예외 허용, 「조세특례제한법」 및 「지방세특례제한법」에 따른 조세감면, 로봇랜드 조성에 있어서 인허가 의제 등 행정적 지원 및 사인에 의한 공용수용 허용, 그리고 한국로봇산업진흥원 설립 및 지능형로봇전문연구원과 전문기업 지정을 통한 지원 방식 등 다양한 형태의 지원법제적 규율을 담고 있다.<sup>10)</sup> 특히 지능형로봇법 제2조 제1호는 지능형로봇을 “외부환경을 스스로 인식하고 상황을 판단하여 자율적으로 동작하는 기계장치(기계장치의 작동에 필요한 소프트웨어를 포함한다)”라고 정의하고 있어, 지능형 로봇의 소프트웨어로서 인공지능이 활용되게 된다면 지능형로봇법은 인공지능 기술개발과 관련된 지원법제로서의 역할을 담당할 수 있다. 이는 2020년 4월 관계부처 합동으로 발표된 「2020년 지능형 로봇 실행계획」에 인공지능에 대한 논의가 빈번히 등장한다는 점에서 간접적으로 확인할 수 있다.<sup>11)</sup>

8) Roßnagel, “Innovation als Gegenstand der Rechtswissenschaft”, in: Hof/Wengenroth(Hrsg.), *Innovationsforschung*, Hamburg: LIT, 2007, S. 11 ff.

9) Hoffmann-Riem, *Innovation und Recht - Recht und Innovation*, Tübingen: Mohr Siebeck, 2016, S. 270 ff.

10) 지능형로봇법 전반에 대한 개관으로는 고영미, “지능형 로봇 금융지원제도 개선방안”, 「선진 상사법률연구」 제86호, 법무부, 2019, 177~203면; 이규호, “지능형 로봇법 소고”, 「사법행정」 제61권 제1호, 한국사법행정학회, 2020, 46~52면.

그리고 「국가초고성능컴퓨터 활용 및 육성에 관한 법률」(이하 초고성능컴퓨터법이라 함) 또한 인공지능 기술개발 관련 지원법제로 활용될 가능성을 내포하고 있다. 초고성능컴퓨터법 제2조는 초고성능컴퓨팅을 “초고성능컴퓨터나 초고성능컴퓨터 기술을 이용한 고용량·고속의 전산망의 활용, 특수 목적의 실험시스템의 구축, 응용 및 시스템 소프트웨어, 대용량 데이터 관리 등을 포함하는 컴퓨팅, 통신 및 정보기술”이라고 정의하고 있으며(동조 제2호), 국가초고성능컴퓨팅을 “공공 및 산업적 목적으로 활용되는 과학기술·국방·교육·사회·문화·경제 등 국가차원에서 이루어지는 초고성능컴퓨팅시스템의 개발·구축·운영 및 활용”이라고 정의하고 있다(동조 제3호). 또한 국가초고성능컴퓨팅을 위한 각종 정책적 기반을 마련하고, 국가초고성능컴퓨팅센터를 설립·지정하여 산학연 협력을 통한 국가초고성능컴퓨팅 연구개발 수행을 해당 센터의 사업으로 규정하고 있다(초고성능컴퓨터법 제9조 제3항 제6호).<sup>12)</sup> 실제로 국가슈퍼컴퓨팅센터는 인공지능 R&D를 지원하고 있다.<sup>13)</sup>

지능형로봇법과 초고성능컴퓨터법과 같이 우리나라의 전형적인 지원법제적 성격을 갖고 있는 법률들이 존재함에도 불구하고,<sup>14)</sup> 본고에서는 인공지능 기술 발전을 위한 법제적 개선 필요성 및 방향성을 검토하려 한다. 이는 지능형로봇법이나 초고성능컴퓨터법상의 지원법제적 제도들이 인공지능 기술개발을 위해 활용될 수 있음에도 불구하고, 이들 법률에는 인공지능 기술개발과 관련된 중핵이라고 할 수 있는 데이터와 관련된 사항이 결여되어 있기 때문이다. 특히 이들 법률에는 다양한 형태의 인공지능 활용을 위해 필수적이라고 할 수 있는 다양한 분야에 속한 양질의 데이터 확보 방안 및 다양한 분야에 속한 양질의 데이터를 활용하기 위한 제도적 방안이 규율되고 있지 않다.

따라서 이하에서는 인공지능 기술개발을 지원하기 위한 제도적 방안으로, 공

11) 관계부처 합동, 「2020년 지능형 로봇 실행계획」, 2020, 8~9, 13~14, 17면 외 다수.

12) 초고성능컴퓨터법 전반에 대한 개관으로는 권성훈, “국가초고성능컴퓨터 활용 및 육성에 관한 법률의 입법영향분석”, 「한국정책학회 춘계학술발표논문집」, 한국정책학회, 2016, 491~508면.

13) “R&D 혁신지원 프로그램 안내”, 국가슈퍼컴퓨팅센터, <https://www.ksc.re.kr/mobile/jw/jg/hsjw/hsjwan>, 최종접속: 2020.6.8.

14) 우리나라의 지원법제 전반에 대한 검토로는 손현, 「진흥관련 법제의 입법 모델 연구」, 한국법제연구원, 2016을 들 수 있다.

공기관이 갖고 있는 다양한 유형의 연구데이터를 적극적으로 활용할 수 있도록 유도하기 위한 법제개선방안에 대해서 살펴보도록 한다. 이를 위해 우선 인공지능 기술개발에 있어서 데이터가 갖고 있는 함의 및 인공지능 기술개발을 지원하기 위해 국가적 차원에서 공공데이터를 공급하고자 하는 정책적 동향을 살펴본다. 그리고 이러한 동향들이 간과하고 있는 공공데이터적 성격을 갖고 있는 연구데이터의 활용 및 이와 관련된 사항을 규정하고 있는 유럽연합의 최근 법제 동향을 살펴보도록 한다(이하 II.). 그리고 우리의 법제상 공공데이터적 성격을 갖고 있는 연구데이터가 존재하는지, 그리고 존재한다면 이는 적극적으로 활용을 위해 제공되고 있는지, 그렇지 않다면 그 이유는 무엇이며, 이를 해소하기 위한 법제적 개선방향은 무엇인지를 검토하도록 한다(이하 III.).

다만, 본격적인 논의를 진행하기에 앞서 밝혀둘 점이 있다. 우선 본고는 공공데이터적 성격을 갖는 연구데이터 활용을 통한 인공지능 학습데이터 확보라는 관점에서 법제개선방안을 검토하기 때문에, 연구데이터에 대한 재산권 인정여부, 공공데이터인 연구데이터 활용으로 인해 발생할 수 있는 지적재산권의 충돌 등과 같은 사항을 직접적으로 다루고 있지 않다.<sup>15)</sup> 무엇보다 본고는 개인정보와는 관련성이 적은 과학기술 연구데이터를 인공지능 학습용으로 보다 적극적으로 활용할 수 있도록 하는 법정정책적 방안을 검토하는 것을 주요 목적으로 삼는다. 따라서 본고는 개인정보와 관련된 이슈를 다루지 않고 있음도 모두에 밝혀두는 바이다.<sup>16)</sup> 화학소재 또는 생체분자 물질 등의 물성정보, 다목적 방사광 가속기를 통해 수집한 연구데이터, 식물의 병충해 정보 등과 같은 과학기술 연구데이터는 개인정보와의 관련성이 적기 때문이다.

15) 이러한 논의는 공공연구데이터의 활용과 관련하여 동전의 양면과 같은 사항이며, 향후 공공연구데이터의 적극적 활용을 위해 반드시 법적으로 명확하게 규율되어야 할 사항이다. 이와 관련된 사항은 추후 후속연구를 통해 소개하도록 한다.

16) 공공데이터와 개인정보보호를 리스크 관리의 관점에서 고찰한 국내 연구로는 권은정, “공공데이터 영역 리스크 관리에 관한 법적 소고”, 「행정법연구」 제60호, 행정법이론실무학회, 2020, 165~190면.

## II. 인공지능 학습데이터로서 공공데이터와 공공연구데이터 제공 법제

### 1. 인공지능 학습용 데이터의 중요성과 수급의 애로

인공지능 기술개발과 관련하여 관심 있게 살펴보아야 할 점은, 학습시키는 데이터가 무엇인가에 따라 인공지능의 발전양상이 다양하게 나타난다는 점이다. 'IBM 왓슨'은 인간의 자연언어 처리를 위해 개발되었던 인공지능이었으나 의료 데이터 학습을 통해 인공지능 의료시스템으로 활용되고 있다. 이처럼 피부에 와 닿는 사례 이외에도, 실제로 인공지능을 활용하기 위해 노력하는 전문분야는 매우 광범위하다. 한국과학기술정보연구원은 20년 장기 전략 특허로서의 가능성을 갖고 있는 특허를 선별해내는 인공지능을 개발하였다.<sup>17)</sup> 「2020년 지능형 로봇 실행계획」을 살펴보면 농촌진흥청은 스마트팜 인공지능기반 병해판별 및 생육계측 기술을 개발 중이며, 인공지능 기반 병해판독 알고리즘을 개발하기 위해 작물의 병해 데이터를 수집하고 인공지능을 통해 이를 판독하는 기술과 작물 생육환경 및 생육량 데이터를 학습한 인공지능을 개발하기 위해 데이터를 정제하는 작업을 수행 중이다.<sup>18)</sup> 인공지능은 기초과학 분야에서도 각광을 받고 있다. 최근에는 소재구조와 물성 간 상관관계와 관련된 대규모 연구데이터를 기반으로 인공지능을 활용한 신소재 개발이 활발히 이루어지고 있다.<sup>19)</sup>

이러한 현상을 비추어보면, 성공적인 인공지능 개발을 위해서는 다양한 형태의 고품질의 학습데이터를 통해 인공지능을 학습시키는 것이 필요하다.<sup>20)</sup> 즉, 인공지능 기술개발 관련 핵심이슈는 데이터이다.<sup>21)</sup> 이처럼 인공지능 기술개발

17) "KISTI, AI로 '오래 가는 특허' 분석 예측 모형 개발", 전파신문(2019.11.27.일자), <http://www.jeonpa.co.kr/news/articleView.html?idxno=82872>, 최종접속일: 2020.6.8.

18) 관계부처 합동, 「2020년 지능형 로봇 실행계획」, 2020, 46~47면.

19) 한상수/김동훈, "인공지능, 소재 개발에 침투하다", 「융합연구리뷰」 vol. 5 no. 1, 융합연구정책센터, 2019, 1~34면.

20) "AI 완성의 속도 높은 데이터 가공", 공학저널(2020년 3월 23일자), <http://www.engjournal.co.kr/news/articleView.html?idxno=646>, 최종접속일: 2020.6.8.



에 있어 학습데이터가 갖고 있는 중요성 때문에, 소규모 스타트업과 같이 기업의 규모가 작은 경우 데이터 확보 그 자체에 애로가 있어 대규모의 데이터를 확보하고 있는 기존의 대규모 기업에 의존하여 기술을 개발해야하는 상황이다.<sup>22)</sup> 다양한 다량의 데이터를 확보하고 싶은 자본력 있는 기업들은 데이터를 갖고 있는 기업들을 적극적으로 인수하는 현상도 발생하고 있다.<sup>23)</sup> 이와 같이 인공지능 개발 주체들이 겪고 있는 애로를 해결하기 위한 일환으로 정부의 역할에 대한 논의가 지속적으로 이루어지고 있으며, 이러한 문제에 대한 해결방안으로 정부가 갖고 있는 양질의 데이터를 공공데이터로 활용할 수 있도록 하는 것이 제시되기도 한다.<sup>24)</sup>

## 2. 공공데이터 개방 정책과 공공연구데이터로의 시선전환 필요성

이처럼 인공지능 기술개발의 처음이자 끝이며, 다양한 분야에서 활용 가능한 인공지능 개발에 있어서 애로라고 할 수 있는 다양한 유형의 학습데이터 확보 문제와 관련하여, 인공지능 기술개발 촉진 및 활용분야 확대를 위해 정부의 역할이 중요하며 정부가 공공데이터를 적극적으로 공개하여 인공지능 개발주체들에게 양질의 데이터를 제공하는 것이 필요하다는 인식이 최근 해외 주요국가의 인공지능 전략과 우리나라의 관련 전략에서 선명하게 나타나고 있다.

2018년 5월에 발표된 스웨덴의 인공지능전략(Nationell inriktning för artificiell

21) 알고리즘과 관련하여 데이터의 품질이 핵심이라는 지적으로는, Datenethikkommission der Bundesregierung(Hrsg.), *Gutachten der Datenethikkommission*, Berlin: Brandenburgische Universitätsdruckerei und Verlagsgesellschaft Potsdam mbH, 2019, S. 52; 이장재, “인공지능(AI) 데이터가 답이다.”, TePRI Insight, <http://tepri.kist.re.kr/?p=3675>, 최종접속일: 2020.6.8.; 정상조, “인공지능과 데이터법”, 법률신문(2020년 2월 10일자), <https://m.lawtimes.co.kr/Content/Info?serial=159295>, 최종접속: 2020.6.8.; 관계부처 합동, 「데이터 산업 활성화 전략」, 2018, 8면.

22) “국내 인공지능(AI) 스타트업 트렌드”, wow tale(2019년 10월 29일자), <https://wowtale.net/2019/10/29/ai-startups-trends-in-korea/>, 최종접속일: 2020.6.8.

23) “[ICT 4.0] 신성장 'AI', '데이터 부자' 네이버·카카오·SKT '독식'”, 뉴스핌(2020년 5월 7일자), <https://www.newspim.com/news/view/20200506000826>, 최종접속일: 2020.6.8.

24) 이상민/김은희, “글로벌 인공지능 정책과 전략적 과제”, 「KISTI Issue Brief」 제17호, 한국과학기술정보연구원, 2020, 14면.

intelligens)에는 데이터기반 확보를 위한 스웨덴 정부의 역할이 중요하다는 점과 함께, 데이터 접근과 관련된 원칙과 기준 등을 마련하고자 하는 스웨덴 정부의 노력이 드러나고 있다. 이러한 정책적 방향성과 함께 스웨덴 인공지능전략은 공공부분의 데이터 운용과 민간부분의 데이터 운용의 촉진을 강조하는 가운데, 특히 공공부분이 갖고 있는 데이터가 갖고 있는 고품질성에 대한 인식과 인공지능 발전을 위해서 국가가 데이터를 제공하는 것이 인공지능 역량 강화의 필요조건이라는 점을 명시하고 있다.<sup>25)</sup> 2018년 11월 공개된 독일 연방정부의 인공지능전략(Strategie Künstliche Intelligenz der Bundesregierung) 또한 스웨덴의 인공지능전략과 마찬가지로 양질의 데이터 확보가 인공지능기술 개발의 필요충분조건이라는 점을 파악하고 있으며, 이를 위해 공공데이터 공개 영역을 보다 확대할 것이라는 점을 명시하고 있다.<sup>26)</sup> 2019년 3월에 공개된 덴마크의 국가인공지능전략(National strategi for kunstig intelligens)에서는 덴마크의 공공데이터에 대한 접근성이 높다는 자체평가 하에 공공데이터 접근성을 보다 높이는 것을 인공지능 전략의 주요 내용으로 삼고 있다.<sup>27)</sup>

이러한 접근 방법은 우리 정부가 공개한 데이터 관련 전략과 인공지능 관련 전략에서도 찾아볼 수 있다. 2018년 발표된 「데이터 산업 활성화 전략」은, 비록 인공지능 전략으로서의 성격은 강하지 않지만, 국내적으로 인공지능 학습용 데이터 구축이 미흡하다는 점과 공공데이터의 개방 정책에도 불구하고 양질의 공공데이터가 적고 공공데이터의 품질이 낮다는 점을 문제점으로 제시하고 있다.<sup>28)</sup> 그리고 이를 극복하기 위한 전략으로 산업별 데이터 및 인공지능 학습데이터 구축과 원시 데이터 형태로 공공데이터를 최대한 수집하여 개방하는 방식을 제시하고 있다.<sup>29)</sup> 2019년 12월에 발표된 「인공지능 국가전략」에서도 인공지능 활용 활성화를 위해 공공데이터를 적극적으로 발굴·개방하는 것을 주요 추진과제로 제시하고 있으며,<sup>30)</sup> 인공지능 스타트업에게 공공데이터를 제공하고

25) 이재훈, “유럽 주요국 인공지능 정책동향 분석”, 「이슈페이퍼」 19-19-2, 한국법제연구원, 2019, 13~15면.

26) 이재훈, 전계서(각주 25), 21면.

27) 이재훈, 전계서(각주 25), 27면.

28) 관계부처 합동, 「데이터 산업 활성화 전략」, 2018, 6면.

29) 관계부처 합동, 「데이터 산업 활성화 전략」, 2018, 14면 및 17면.

활용할 수 있도록 하는 것 또한 추진 과제로 제시하고 있다.<sup>31)</sup>

이처럼 최근 발표되고 있는 인공지능 관련 정책은 공공데이터가 비교적 양질의 데이터라는 인식 하에 공공데이터를 인공지능 기술개발자들이 활용할 수 있도록 하는 방식을 공통적으로 담고 있다. 하지만 이러한 전략들의 경우 현재 공공영역에서 갖고 있는 데이터가 비교적 양질의 데이터라는 점과 공공데이터 제공이라는 점 자체에 집중한 나머지 간과하고 있는 지점이 있다. 이는 공공데이터로 생성·수집되는 데이터 그 자체의 품질이 좋다면 공공데이터의 품질 향상이 보다 수월하게 이루어질 수 있다는 점이다. 특히 다양한 연구개발단계에서 발생하는 연구데이터<sup>32)</sup>의 경우 특히 학습데이터로서 고품질의 성격을 갖고 있기 때문에 연구개발단계에서 발생한 연구데이터를 공공데이터로 수집하여 공개하게 된다면 보다 전문적인 분야에서의 인공지능 활용이 가능해질 수 있다. 따라서 연구데이터와 같은 고품질 데이터를 공공데이터화하고 이를 제공하여 활용하는 법정정책적 접근방식이 필요하다.<sup>33)</sup>

그리고 실제로 유럽연합은 이러한 방향성을 갖고 있는 지침(Richtlinie; Directive)을 마련하여 2019년 6월에 공포하였다. 이하에서는 인공지능 시대에 연구데이터의 공공데이터적 성격을 명확히 하고 이를 제공할 수 있도록 하는 제도를 담고 있는 유럽연합의 「오픈데이터 및 공공분야 정보의 이용에 대한 지침」(이하 PSI<sup>34)</sup>-지침이라 함)<sup>35)</sup>이 담고 있는 주요 내용을 살펴보도록 한다.

30) 관계부처 합동, 「인공지능 국가전략」, 2019, 13면.

31) 관계부처 합동, 「인공지능 국가전략」, 2019, 23면.

32) 연구데이터는 설문조사 등 사회과학적 방법론을 통해 생산되는 조사데이터, 실험 및 실험 장비를 통해 생산되는 실험데이터, 망원경, 전자현미경, 인공위성 등을 통해 생산되는 관찰데이터, 모델링 작업을 통해 생산되는 시뮬레이션 데이터, 이러한 원천 데이터들로부터 재생산되는 파생데이터, 그리고 전문적인 평가를 거쳐 신뢰성이 검증된 참조데이터로 구분된다. 최명석/김현우/이상환, “국가 연구데이터 공유의 시작, 데이터 관리계획(Data Management Plan)”, 「KISTI Issue Brief」 제13호, 한국과학기술정보연구원, 2019, 4면.

33) 연구데이터의 공개는, 인공지능 학습데이터 확보 및 제공을 위한 공공데이터로서의 공개라는 관점이 아닌, 중복 연구 수행 예방을 통한 연구의 효율성 증대, 분산화 된 연구 데이터로 인한 국가적 연구역량 약화 극복, 연구자 중심의 데이터 생태계 구축 등을 근거로 주창되는 오픈 사이언스 운동의 관점에서 논의가 이루어지고 있다. 이에 대해서는 문희진/강남규/이상환/정용일, “연구자 중심의 과학기술 데이터 생태계 활성화 전략”, 「KISTI Issue Brief」 제19호, 한국과학기술정보연구원, 2020을 참조할 것.

34) PSI는 Public Sector Information의 줄임말이다.

### 3. 유럽연합의 PSI-지침상 연구데이터 확보 및 공개 관련 법제

#### 가. PSI-지침 도입 배경과 입법의도

유럽연합은 공공분야에서 생산되고 수집된 정보를 활용할 목적으로 2003년 「공공분야 정보의 이용에 대한 지침」<sup>36)</sup>을 공포하였고, 2013년에는 해당 지침의 적용범위를 확장하기 위한 개정 지침<sup>37)</sup>이 공포되었다. 그러나 이와 같은 규범적 노력에도 불구하고 보다 많은 데이터가 개방되어야 한다는 요청과 고품질 데이터 제공과 관련한 규율이 미비하다는 문제점이 지속적으로 제기되었다.<sup>38)</sup> 이에 유럽연합 집행위원회는 이들 지침에 대한 검토 및 이해당사자들의 의견수렴을 진행하였고, 그 결과 기존 지침의 형태로는 인공지능과 같은 디지털 혁신을 지원하는 것에 한계가 있다는 결론에 도달하였다.<sup>39)</sup> 이러한 진단에 터 잡아 유럽연합의 입법자는 데이터의 개방에 보다 방점을 둔 PSI-지침을 채택하는 방식을 통해 「공공분야 정보의 이용에 대한 지침」을 전면 개정하였다.<sup>40)</sup>

유럽연합의 입법자는 인공지능을 통한 데이터의 활용이 모든 경제영역의 변환을 초래할 것으로 진단하면서,<sup>41)</sup> 인공지능 기술의 급속한 발전이 데이터 기반 서비스 범주의 확장과 새로운 제품 등장에 큰 영향을 미칠 것이므로 이와 관련한 법제의 대응도 필요하다는 점을 입법이유에 명시적으로 제시하고 있

35) Richtlinie (EU) 2019/1024 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors; Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

36) Richtlinie 2003/98/EG des Europäischen Parlaments und des Rates vom 17. November 2003 über die Weiterverwendung von Informationen des öffentlichen Sektors.

37) Richtlinie 2013/37/EU des Europäischen Parlaments und des Rates vom 26. Juni 2013 zur Änderung der Richtlinie 2003/98/EG über die Weiterverwendung von Informationen des öffentlichen Sektors.

38) 김민호/이보옥, “공공 오픈데이터 EU Directive 개정 동향과 시사점”, 「성균관법학」 제32권 제1호, 성균관대학교 법학연구원, 2020, 5면.

39) EG(3) der RL (EU) 2019/1024.

40) EG(1) der RL (EU) 2019/1024; 김민호/이보옥, 전계논문, 16면.

41) EG(9) der RL (EU) 2019/1024.

다.<sup>42)</sup> 특히 유럽연합의 입법자는 공공데이터와 인공지능 기술 간의 관계를 주의 깊게 고려하고 있다는 점을 PSI-지침 입법이유에서 명시적으로 기술하고 있으며, 공공분야가 갖고 있는 정보와 공적 임무 수행과 관련하여 발생·수집한 정보, 그리고 공적 서비스 제공을 통해 발생·수집한 정보는 인공지능 기술 개발을 위한 고품질 데이터의 원천이 될 수 있다고 평가하고 있다.<sup>43)</sup>

공공분야의 데이터 중 고품질 데이터로서 이에 대한 확보 및 공유를 위해 유럽연합 입법자가 특히 주목한 분야는 연구데이터이다. 유럽연합 입법자는 연구데이터 개방을 통한 연구데이터 무료 사용 및 재사용은 불필요한 중복연구를 줄이고 기술발전을 촉진시킬 뿐 아니라 무료 사용 및 재사용 과정에서 데이터의 품질이 보다 향상될 수 있는 가능성을 갖고 있다고 판단하고 있다. 또한 이러한 연구데이터 유통·사용 및 품질향상은 혁신을 촉진한다고 평가한다. 그리고 이와 같은 인식 하에 연구데이터는 단순히 해당 연구가 속한 학문 분과에서만 의미가 있는 것이 아니며, 다양한 영역에서 발생한 다양한 데이터가 통섭적 형태로 적극 활용되어야 한다고 평가하고 있다.<sup>44)</sup>

#### 나. PSI-지침상 연구데이터에 대한 규율

다만 공개 대상이 되는 연구데이터와 관련하여 유럽연합 입법자의 입법 의도에 따르면, 모든 연구데이터가 공개 대상이 되는 것은 아니다. 오히려 PSI-지침의 입법이유에는 공개를 통한 무료사용의 대상이 되는 연구데이터는 공적 주체로부터 재정지원을 받았거나 민관합작의 형식으로 재정지원이 이루어진 연구에서 발생한 연구데이터에 한정된다고 명시되어 있다.<sup>45)</sup> 이하에서는 PSI-지침을 통해 새롭게 규율된 연구데이터와 관련된 법적 규율의 내용을 살펴보도록 한다.

##### (1) 연구데이터

42) EG(10) der RL (EU) 2019/1024.

43) EG(13) der RL (EU) 2019/1024.

44) EG(27) der RL (EU) 2019/1024.

45) EG(28) der RL (EU) 2019/1024.

PSI-지침 제2조 제9항은 학술적 연구과정에서 수집·생성되었으며 연구결과의 증명을 위해 활용되거나 학계에서 연구를 통해 발견한 사실과 연구결과를 검증하기 위해 필요하다고 인정되는 디지털 형태의 콘텐츠를 연구데이터라고 정의하고 있으며, 학술적 출판물은 연구데이터에서 배제하고 있다. 다만, 오픈데이터로서 활용될 수 있는 연구데이터는 PSI-지침 제10조의 조건을 충족하고 있어야 한다고 규정하고 있다(PSI-지침 제1조 제1항 c호).

그리고 PSI-지침 제10조 제2항에서는 연구데이터는 제3자의 지식재산권의 대상이 된다고 하더라도 공공데이터로서 타인을 위해 제공될 수 있다고 규정하고 있다. 이 지점에서 유럽연합 입법자가 연구데이터가 갖고 있는 고품질성으로 인해 개방필요성을 강하게 인정하고 있음을 발견할 수 있다. PSI-지침 제1조 제2항은 각호를 통해 PSI-지침의 적용대상에서 제외되는 사항들을 규율하고 있는데, 이 중 c호에서 제3자의 지식재산권이 관련 경우를 규정하고 있어, 원칙적으로 제3자의 지식재산권이 관련된 경우는 원칙적으로 PSI-지침의 적용대상이 아니기 때문이다.<sup>46)</sup>

## (2) 연구데이터 관련 원칙 및 이용조건

PSI-지침 제10조 제1항은 유럽연합회원국들에게 연구데이터와 관련하여 오픈 액세스 정책을 지원할 것을 요구하는 원칙적 사항을 규정하고 있다. PSI-지침에서 연구기관 및 연구지원기관에 대해 적용하도록 회원국에게 요청하는 오픈 액세스 정책이란, 유럽연합의 각 회원국들이 국가 전략 및 관련 조치(Maßnahmen)를 통해 공적으로 재정지원이 이루어진 연구데이터를 개방하여 활용할 수 있는 가능성을 기본적 개방 원칙(le principe d'ouverture par défaut; the principle of 'open by default'; Grundsatz der 'standardmäßig offenen Daten')과 FAIR 원칙<sup>47)</sup>에 부합하는 형태로 지원하는 것을 의미한다(PSI-지침 제10조 제1

46) 그렇다고 하여 유럽연합 입법자가 공공연구데이터 활용을 위해 지적재산권적 이슈를 백안시하고 있는 것은 아니다. 아래에서도 기술하고 있듯이 PSI-지침 제10조 제2항 제2문에서 회원국들은 연구데이터 개방 및 제공에 있어 지식재산권적인 사항을 고려하도록 요청하고 있다.

47) FAIR 원칙은 연구데이터 활용과 관련된 원칙으로 Findability(연구데이터의 발견가능성), Accessibility(연구데이터의 접근가능성), Interoperability(연구데이터의 상호운용가능성), Reusability(연구데이터의 활용가능성)의 각 첫 글자를 따서 조어되었다. 유럽연합 입법자는 PSI-지침

항 제1문). 그리고 오픈데이터 정책과 관련하여 회원국은 지적재산권, 개인정보 보호, 비밀유지, 질서유지, 정당한 영업이익과 관련된 요청을 ‘가능한 개방, 필요 시 폐쇄’(as open as possible, as closed as necessary; so offen wie möglich, so geschlossen wie nötig)라는 원칙에 부합하게 고려해야 한다(PSI-지침 제10조 제1항 제2문).

그리고 PSI-지침 제10조 제2항 제1문은 연구데이터가 상업적·비상업적 목적을 위해 제공될 수 있다는 것을 명시적으로 규정하고 있다. PSI-지침 제10조 제2항 제1문에 따르면 -제3자의 지식재산권이 관계가 되어 있다고 하더라도<sup>48)</sup> - 연구데이터는 원칙적으로 대가없이 활용될 수 있으며(PSI-지침 제10조 제2항 제1문 및 제6조 제1항 체계적 해석), 연구데이터 활용 조건은 비차별적이어야 하고(PSI-지침 제10조 제2항 제1문 및 제11조 제1항 체계적 해석), 공익적으로 특별한 사유가 존재하지 않는 한 제3자에게 연구데이터를 배타적으로 사용할 수 있는 권리를 부여하는 계약 내지 협약을 체결할 수 없다(PSI-지침 제10조 제2항 제1문 및 제12조 제1항·제2항 체계적 해석).

다만 이미 PSI-지침의 입법이유에 기술되어 있는 바와 같이 모든 연구데이터가 공공데이터로 개방되지는 않는다. 대가성 없이 비차별적이고 제3자에게 배타적 이용권을 부여할 수 없는 연구데이터가 되기 위해서는 우선 공적 재원에 의해 재정적 지원을 받아서 생성·수집된 연구데이터여야 하며, 또한 연구 데이터를 연구자, 연구기관 또는 연구지원기관이 기관 리포지터리 또는 주제 리포지터리를 통해 공적으로 활용할 수 있도록 만들어 놓은 상태여야 한다(PSI-지침 제10조 제2항 제1문 후단).<sup>49)</sup> 즉 연구데이터가 공공데이터로서 활용될 수 있기

입법이유(27)을 통해 FAIR 원칙에 맞게 연구데이터가 확산되는 것을 지원하기 위한 노력이 필요하다는 점을 제시하고 있다. 연구데이터 공개와 FAIR 원칙에 대해 보다 자세한 내용은 Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3, 160018. doi:10.1038/sdata.2016.18을 참고할 것.

48) PSI-지침 제1조 제2항은 각호를 통해 PSI-지침의 적용대상에서 제외되는 사항들을 규율하고 있는데, 이 중 c호에서 제3자의 지식재산권이 관련 경우를 규정하고 있어, 제3자의 지식재산권이 관련된 경우는 원칙적으로 PSI-지침의 적용대상이 배제된다.

49) 리포지터리는 디지털화 된 학술정보를 효율적으로 저장·관리하는 도구이다. 기관 리포지터리는 이를 구축한 기관에서 생성한 지식정보자원을 보존·관리하고 외부에 공개하는 형태의 리포지터리이며, 주제 리포지터리는 특정 분야의 주제와 한정되어 구축된 리포지터리로 학술단체회원, 해당 분야 연구자들의 자발적 운영 등의 방식으로 운영된다. 정영미/배정희, “기술확

위해서는 연구데이터 생성·수집 과정에 있어서의 공적 자원 투입 및 리포지터리를 통한 접근 가능성이라는 두 가지 요건을 갖추고 있어야 한다. 이와 같은 요건이 갖추어진 경우 연구데이터는 공공데이터로서 비대가성·비차별성과 같은 공공데이터 활용원칙에 따라 상업적·비상업적으로 자유롭게 활용될 수 있다.<sup>50)</sup>

### (3) 공공연구데이터 생성 연구기관의 법적 성격

PSI-지침 제10조 제2항 제1문은 공공데이터로서 연구데이터가 활용될 수 있다는 점과 연구데이터가 공공데이터로 활용될 수 있는 조건을 규정하고 있다. 그런데 법문상의 주어가 공공데이터로 되어 있기 때문에, 어떠한 연구주체가 생성한 연구데이터가 공공데이터로 활용될 수 있는지에 대해서는 불명확한 점이 있다. 다만 연구데이터가 공공데이터가 될 수 있는 조건으로 공적 재원이 투입되어야 한다는 점이 규정되어 있기 때문에, 공적 재원의 지원을 받아서 연구데이터를 생성·수집한 모든 연구주체가 연구데이터를 공공데이터로 공개해야하는 것처럼 해석될 소지가 있다.

그런데 PSI-지침의 입법이유(28)을 살펴보면, 연구기관 및 연구지원기관은 공공부문기관 또는 공기업일 수 있고, PSI-지침은 이러한 기관과 관련하여 이들 주체가 연구주체로서 갖고 있는 기능 및 이들이 보유한 연구데이터에만 적용된다고 기술되어 있어, 국가 또는 지방자치단체가 설립한 연구기관 내지 공기업과 같은 형태의 연구기관만이 연구데이터 개방과 관련된 연구기관인 것으로 해석된다.

## 4. 소결

지금까지 인공지능 개발의 주요 전제조건 중 하나인 데이터와 관련된 사항을 살펴보았다. 특히 인공지능 기술 분야에서 다양한 유형의 데이터와 고품질의 데

산 통합모델을 통한 개방형 기관 리포지터리 수용의 영향요인 분석”, 「한국도서관정보학회지」 제46권 제4호, 한국도서관정보학회, 2015, 531면 참조.

50) 다만 이와 관련해서 정당한 상업적 이익, 지식이전활동, 이미 성립된 제3자의 지식재산권이 전적으로 무시되는 것은 아니며, 연구데이터의 활용과 관련하여 이와 같은 사항이 고려되어야 한다(PSI-지침 제10조 제2항 제2문).



이터가 필요함에도 불구하고 이를 수급하는데 애로가 있다는 점과 이를 극복하기 위한 방식으로 정부가 갖고 있는 공공데이터를 활용하려는 정책적 인식 및 방향성을 살펴보았다. 그리고 다양한 고품질의 공공데이터를 확보하기 위한 법정책적 접근 방식으로서 공적인 연구주체들에 의해 수집·생성되어 리포지터리로 저장된 연구데이터의 공개를 규율하고 있는 유럽연합의 PSI-지침의 내용을 살펴보았다.

이하에서는 우리나라 법제상 연구데이터가 공공데이터로 활용될 수 있는지 살펴보고, 이와 관련하여 어떠한 법제적 개선이 필요한지를 검토하도록 한다.

### III. 공공연구데이터 활용을 위한 공공데이터 법제 개선방향

유럽연합의 PSI-지침과 유사한 성격과 기능을 갖고 있는 우리의 법률은 「공공데이터의 제공 및 이용 활성화에 관한 법률」(이하 공공데이터법이라 함)이다. 실제로 공공데이터법 제정 관련 입법 관련 자료를 살펴보면 유럽연합의 2003년 「공공분야 정보의 이용에 대한 지침」으로부터 일정 부분 영향을 받아 공공데이터법이 제정된 것임을 확인할 수 있다.<sup>51)</sup>

이하에서는 우리의 공공데이터 법제 하에서 PSI-지침이 규정하고 있는 바와 마찬가지로 공적 자금을 지원 받아 생성·수집되어 관리되는 연구데이터가 공공데이터로 개방되어 인공지능 개발을 위해 사용될 수 있는 것인지 검토하기 위해, 공공데이터법의 수범주체 및 공공데이터의 유형을 분석하여 검토한다. 그리고 공공데이터 법제가 갖고 있는 규범적 틀과는 별개로 현실에서 공공기관이 생성·수집한 연구데이터가 공공데이터로서 공유되고 있는지 그 현황과 이와 관

51) 국회 행정안전위원회, 「공공데이터의 제공 및 이용 활성화에 관한 법률안 검토보고서」, 2012, 9~10면. 공공데이터법 제3조는 공공데이터와 관련된 기본원칙을 정하고 있다. 동조 제1항에 따르면 공공기관은 누구든지 공공데이터를 편리하게 이용할 수 있도록 노력해야 하며, 공공데이터의 이용권 확대를 위해 필요한 조치를 취해야 한다고 규정하고 있다. 그리고 동조 제2항에서는 공공데이터 접근에 대한 평등의 원칙을 규율하고 있다. 또한 동조 제4항은 별도의 예외 사유를 제외하고는 공공데이터의 영리적 이용이 허용됨도 정하고 있다. 이러한 일반원칙을 살펴보면, PSI-지침상 연구데이터의 활용과 관련하여 부수적으로 검토되었던 공공데이터 이용의 원칙과 우리의 공공데이터법상 원칙이 실제로 매우 유사함을 확인할 수 있다.

런된 문제점을 검토한다. 그리고 공공기관이 갖고 있는 연구데이터가 공공데이터로서 적극적으로 활용될 수 있기 위해 필요한 제도적 장치가 무엇인지 검토하도록 한다.

## 1. 공공데이터법의 수범주체, 그리고 제공대상으로서 공공데이터

### 가. 공공데이터법의 수범주체 - 검토범위의 한정

공공데이터법 제2조 제1호는 ①국가기관, ②지방자치단체 및 ③「국가정보화 기본법」 제3조 제10에 따른 공공기관을 공공데이터법이 적용되는 공공기관으로 정의하고 있다. 그리고 「국가정보화 기본법」 제3조 제10호는 ①「공공기관의 운영에 관한 법률」에 따른 공공기관, ②「지방공기업법」에 따른 지방공사 및 지방공단, ③특별법에 따라 설립된 특수법인, ④「초·중등교육법」, 「고등교육법」 및 그 밖의 다른 법률에 따라 설치된 각급 학교, ⑤기타 대통령령으로 정하는 법인·기관 및 단체를 공공기관으로 정의하고 있다. 다만 대통령령으로 정하고 있는 법인·기관 및 단체와 관련해서는 아직 대통령령으로 구체화하고 있지는 않다.

이들 수범주체 중 연구주체로서 직접 연구를 수행하며 연구데이터를 생성·수집하는 수범자는 ①「공공기관의 운영에 관한 법률」에 따른 공공기관, ②「지방공기업법」에 따른 지방공사 및 지방공단, ③특별법에 따라 설립된 특수법인일 수 있다. 다만, 이하에서는 원고의 분량 등을 고려하여 「공공기관의 운영에 관한 법률」에 따른 공공기관을 중심으로 논의를 전개하도록 한다.

### 나. 공공데이터의 범주 및 연구데이터의 포섭 가능성

공공데이터법 제2조 제2호에 따르면, 공공데이터는 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 ①「전자정부법」 제2조 제6호에 따른 행정정보, ②「국가정보화 기본법」 제3조 제1호에 따른 정보 중 공공기관이 생산한 정보, ③「공공기록물 관리에 관한 법률」 제20

조 제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물, ④기타 대통령령으로 정하는 자료 또는 정보를 말한다. 하위법령으로 위임된 2가지 유형 중 「공공기록물 관리에 관한 법률」 제20조 제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물에 대해서만 공공데이터법 시행령 제1조의2에서 규율하고 있는데, 이는 웹기록물 및 행정정보 데이터세트 등의 기록정보자료를 의미한다.

이처럼 공공데이터법은 공공데이터의 유형을 광 또는 전자적 방식으로 처리된 자료 또는 정보라고 규정하고 있을 뿐 그 자료 또는 정보가 담고 있는 구체적인 내용이 무엇인지까지 세세하게 규율하고 있지는 않다. 공공데이터법상 자료 또는 정보가 갖고 있는 내용적 속성을 추측할 수 있는 규범적 기준점은 공공데이터법의 수범자가 만들어낸 정보라는 부분이다.

그리고 이와 관련하여 주목할 점은 위에서도 살펴본 바와 같이 「공공기관의 운영에 관한 법률」에 따른 공공기관이 공공데이터법의 수범자로 되어 있다는 점이다.<sup>52)</sup> 따라서 연구를 수행하는 공공기관이 생성·수집한 연구데이터가 광 또는 전자적 방식으로 처리된 자료 또는 정보의 형태를 갖는 경우 이러한 연구데이터는 공공데이터에 포섭될 수 있다.

## 2. 공공데이터로서 공공기관의 연구데이터, 그리고 실무상 비공개 경향성

이와 같은 해석론을 전제로 한다면 우리나라의 경우 광범위한 주체에 의해 대규모의 연구데이터가 공공데이터로서 생성·수집되었고 현재도 공공데이터의 생성·수집이 진행 중이라고 평가할 수 있다.

실제로 공공데이터법의 수범자인 공공기관과 관련하여, 「공공기관의 운영에 관한 법률」 제4조 제1항에 따라 기획재정부장관이 공공기관으로 지정한 공공기관은 현재 340개 기관에 달한다.<sup>53)</sup> 그리고 이들 공공기관의 면면을 살펴보면

52) 공공데이터법상 공공기관 중 국가와 지방자치단체를 제외한 공공기관은 「국가정보화 기본법」 제3조 제10호의 공공기관인데, 「국가정보화 기본법」 제3조 제10호는 공공기관의 유형 중 하나로, 「공공기관의 운영에 관한 법률」에 따른 공공기관을 들고 있다. 따라서 공공데이터법상 공공기관에는 「공공기관의 운영에 관한 법률」상의 공공기관이 포함된다.

53) “공공기관 현황”, <http://www.alio.go.kr/alioPresent.do>, 최종접속일: 2020.6.8.

경제·인문사회연구회 소속 국책연구기관 뿐 아니라 국가과학기술연구회 소속 국책연구기관들이 포함된다.

특히 국가과학기술연구회 소속 국책연구기관들은 과학기술과 관련된 기초연구를 수행하면서 다양한 연구데이터를 생성·수집하는 기관이라고 할 수 있다. 실제로 국가과학기술연구회 소속 출연연구기관이 수행한 국가연구개발과제에서 발생한 연구데이터와 관련된 설문조사를 살펴보면 국가연구개발과제에서 연구데이터를 생산한다고 응답한 연구자는 전체 설문자 301명 중 75%에 해당하는 226명이었다.<sup>54)</sup> 그리고 연구데이터의 생산용량과 관련하여 응답한 전체 응답자 175명 중 68%이상인 119명이 기가바이트 이상의 연구데이터를 생산한다고 응답하였으며, 이중 29명은 테라바이트 수준의 연구데이터를, 3명은 페타바이트 수준의 연구데이터를 생산한다고 응답하였다.<sup>55)</sup> 그리고 생산한 연구데이터와 관련하여 개인정보가 포함되었는지 여부에 대해 응답한 158명의 응답자 중 본인인 생산한 연구데이터에 개인정보가 포함되어 있지 않다고 응답한 연구자는 총 140명으로 전체 응답자의 89%에 해당하였다.<sup>56)</sup>

이와 같이 연구 분야에 속한 기타 공공기관 뿐 아니라 시장형 공기업, 준시장형 공기업 및 위탁집행형 준정부기관 또한 다양한 형태의 연구데이터를 생성·수집할 수 있다.

시장형 공기업의 예로 한국가스공사를 살펴보면, 「한국가스공사법」 제11조에서 한국가스공사의 사업범위를 규율하고 있는데, 이 중 제5호는 도시가스, 천연가스, 액화석유가스, 수소에너지와 관련된 연구 및 기술개발을 한국가스공사의 사업으로 명시하고 있다. 즉 한국가스공사는 공공데이터법상 공공기관으로서 공공데이터의 성격을 갖고 있는 연구데이터의 생성·수집을 하고 있다. 준시장형 공기업의 예로 한국조폐공사를 살펴보면, 「한국조폐공사법」 제11조에서는 화폐 및 유가증권 제조와 관련된 물품 생산 및 기술개발을 위한 국내외 관련 기관·기업 및 단체와의 협력을 한국조폐공사의 업무로 규정하고 있다. 따라서 한국조폐

54) 한국과학기술정보연구원, 「출연(연) 연구데이터 관리·활용 방안 연구」, 국가과학기술연구회, 2019, 66면.

55) 한국과학기술정보연구원, 전계서, 69면.

56) 한국과학기술정보연구원, 전계서, 72면.

공사 또한 공공데이터법상 공공기관으로서 공공데이터에 해당할 수 있는 연구데이터의 생성·수집을 하고 있다고 할 수 있다. 위탁집행형 준정부기관 중 하나인 한국방송통신전파진흥원 또한 이들과 마찬가지로 할 수 있다. 「전파법」 제66조 제4항은 전파이용 촉진에 관한 연구 및 각종 기술 수집·조사·분석을 한국방송통신전파진흥원의 사업영역으로 규정하고 있기 때문에, 한국방송통신전파진흥원 또한 공공데이터법상 공공기관으로서 공공데이터인 연구데이터를 생성·수집한다.

비록 전수조사의 형식으로 모든 공공기관 관련 수치화된 데이터가 존재하지는 않지만, 기획재정부에 의해 지정된 공공기관의 수 및 주요 공공기관의 기능을 살펴보면 이미 우리나라는 굉장히 다양하고 광범위한 영역에서 공공데이터적 성격을 갖고 있는 연구데이터를 확보하고 있다고 평가할 수 있다.

하지만 국가과학기술연구회 소속 출연연구원이 수행한 국가연구개발과제에서 발생 연구데이터에 국한해서 보자면, 연구데이터를 외부와 공유한 경험이 있는지에 대한 응답한 226명의 연구자 중 87.6%에 해당하는 198명이 연구데이터를 외부와 공유한 경험이 없다고 응답한 것에서 알 수 있듯이,<sup>57)</sup> 연구데이터는 해당 연구를 수행한 연구자만이 활용하는 경향이 존재한다. 실제로 지금까지 공공데이터법을 통해 공개된 공공데이터의 유형들의 면면을 살펴보면 연구데이터로서의 성격을 갖고 있는 데이터가 많지 않은 상황이다.<sup>58)</sup>

### 3. 공공연구데이터 비공개 경향성 원인 분석

이처럼 공공데이터로서의 성격을 갖고 있는 연구데이터 생성·수집 주체가 다양하고 이들을 통해 대량의 다양한 연구데이터가 생성·수집되어 있음에도 연구데이터가 적극적으로 개방되어 활용되지 않는 점이 무엇인지에 대한 의구심이 생길 수밖에 없다. 이러한 의구심과 관련하여, 국가과학기술연구회 소속 출연연구기관이 수행한 국가연구개발과제에서 발생한 연구데이터 관련 설문조사는 유의미한 시사점을 제공한다. 특히 주의 깊게 살펴볼 설문결과는 연구데이터의 관

57) 한국과학기술정보연구원, 전게서, 79면.

58) 신은정/손수정/서지영/김영린, 「공공부문 연구데이터의 소유·활용제도 개선방향」, 과학기술정책연구원, 2019, 109면.

리방식 및 연구데이터가 개방되지 않고 다른 연구자에 의해 활용되지 못하고 있는지, 그리고 이러한 현상의 원인이 무엇인지에 대해 연구자들이 대답한 설문 의 내용이다.

국가과학기술연구회 소속 출연연이 수행한 국가연구개발과제에서 발생한 연구데이터의 관리 방식 설문에 응답한 219명 중 28명의 응답자만이 본인 소속 기관에서 연구데이터를 관리하고 있다고 하였으며 응답자의 59%에 달하는 129명의 연구자는 연구자 개인이 자체적으로 연구데이터를 관리하고 있다고 응답하였다.<sup>59)</sup> 그리고 연구데이터 관리방식과 관련해서는 218명의 응답자 중 75%로에 해당하는 164명이 단순저장장치(서버, 외장하드, PC 등)에서 연구데이터를 관리한다고 응답하였다.<sup>60)</sup>

국가연구개발과제를 수행하면서 다른 연구자가 생성한 연구데이터를 왜 활용하지 않는지에 대한 설문에서 약 28%의 응답자는 연구데이터가 미공개 되었기 때문이라고 응답하였으며 22.1%의 응답자는 연구데이터가 존재하고 있는지조차 알 수 없기 때문이라고 응답하였다.<sup>61)</sup>

그리고 이처럼 연구데이터가 공유가 어려운 이유가 무엇인지에 대한 복수응답설문에서 34%의 응답자가 정책·제도·보안의 문제로 인해 연구데이터의 공유가 어렵다고 응답하였으며, 24%의 응답자는 개인 또는 기관의 자산이므로 연구데이터를 공개할 수 없다고 응답하였다.<sup>62)</sup>

이러한 설문조사 내용은 왜 공공기관에 의해 생성된 연구데이터가 공공데이터로 개방되어 활용되고 있지 못한지에 대한 실마리를 담고 있다. 우선 이 설문을 통해 드러난 첫 번째 문제점은 공공기관<sup>63)</sup>이 수행한 연구개발과 관련하여 애초에 연구데이터가 존재하는지 여부에 대해서 해당 분야 연구자들도 인식하지 못하고 있다는 점이다. 이러한 연구데이터 관리의 비체계성은 생성된 연구데이터를 개인 연구자 스스로 관리한다는 응답이 59%라는 점에서 눈에 띄게 드러난다.

59) 한국과학기술정보연구원, 전계서, 73면.

60) 한국과학기술정보연구원, 전계서, 73면.

61) 한국과학기술정보연구원, 전계서, 88면.

62) 한국과학기술정보연구원, 전계서, 83면.

63) 앞서 검토한 설문의 내용은 국가과학기술연구회 소속 출연연구기관, 즉 기타 공공기관에 해당하는 연구원에 종사 중인 연구자들을 응답자로 설정한 설문조사였음을 상기할 필요가 있다.

이 설문을 통해 발견할 수 있는 두 번째 문제점은 공공기관이 생성한 연구데이터가 공공데이터에 해당한다는 점을 연구자들이 크게 인식하고 있지 못하는 점이다. 이는 연구데이터는 개인 또는 기관의 자산이라서 공개할 수 없다고 응답한 24%의 입장을 통해서 알 수 있다. 그리고 연구자들이 연구데이터에 대해 이와 같은 관점으로 접근하는 것은 개별 연구가 갖고 있는 속성에 따라서도 달라진다. 연구개발의 목적 자체가 데이터 확보인 연구를 통해 발생한 데이터와는 달리 데이터 확보 자체는 연구의 목적이 아니라 연구데이터가 단지 연구과정에서 발생하게 된 부산물적인 성격을 갖고 있는 경우, 이러한 데이터는 각 연구자의 개인적 자산이라고 생각하는 경우가 빈번하기 때문에 이러한 연구자들의 경향 또한 연구데이터를 공공데이터로 활용하는데 있어 장애요인이 된다는 분석도 존재한다.<sup>64)</sup>

세 번째 문제점은 연구데이터의 공유가 어려운 이유로 연구자들이 제도 및 정책을 문제점으로 인식하고 있다는 점이다. 이는 두 번째 문제점인 공공기관의 연구데이터가 공공데이터에 해당한다는 사실을 연구자들이 인식하지 못하고 있는 점과도 일정 부분 연관이 된 지점이라고 할 수 있다. 특히 이와 관련하여 현행 공공데이터법은 공공기관의 연구데이터가 명확하게 공공데이터에 해당하는지 여부와 관련하여 해당 연구기관 종사자들이 명확하게 인식할 수 있도록 하는 규범구조를 갖고 있지 못한 점을 상기할 필요가 있다. 다른 한편, 제도 및 정책이 문제점이라고 지적되는 것은 첫 번째 문제점인 연구데이터가 체계적으로 관리되고 있지 못하다는 점과도 연관을 맺을 수 있다. 연구데이터를 체계적으로 관리할 수 있도록 하는 제도가 도입되어 있었다면 어떠한 연구데이터가 공공데이터로서 어떠한 주체를 통해 어떻게 생성되고 있으며 그 데이터가 담고 있는 내용은 어떠한 것인지가 파악될 수 있음에도 불구하고, 현행 우리 공공데이터 법제에는 이와 같은 사항이 부재하기 때문이다. 물론 공공데이터법 제17조 제1항에 따르면 공공기관의 장은 공공기관이 보유·관리하는 공공데이터를 국민에게 제공해야 한다고 규정하고 있고, 제18조 제1항에서 공공기관의 장은 해당 공공기관의 소관 데이터 목록을 행정안전부장관에게 등록해야한다고 규정

64) 최명석/이승복/이상환, “국내 과학기술분야 연구기관의 과학데이터 관리 현황”, 『한국콘텐츠학회논문지』 vol. 17 no. 12, 한국콘텐츠학회, 2017, 120면.

하고 있으나, 연구데이터를 공공데이터라고 인식하고 있지 못한 (사실적·규범적) 상황에서 이러한 법제가 실효성을 거두는 것이 쉽지 않다. 이는 공공기관에 해당하는 연구기관의 자료 중 공공데이터로 공개된 자료의 대부분이 해당 연구기관의 시설, 자원보유, 이용과 관련된, 즉 연구개발과는 관련성이 적은 단순한 행정데이터라는 사실이 반증하고 있는 바이다.<sup>65)</sup>

#### 4. 공공연구데이터 활용성 증대를 위한 법제개선 방안

앞서 살펴본 바와 같이 국가연구개발과제를 통해서 생성·수집된 연구데이터는 다양한 분야의 대용량 데이터로서 다양한 형태의 인공지능 학습을 위해 활용될 수 있는 공공데이터임에도 불구하고 적극적으로 공개되지 않고 있는 상황임을 알 수 있다. 이는 비단 국가연구개발과제에서만 나타나는 현상은 아니며, 각종 공공기관이 수행하여 수집·연구한 연구데이터 전반에서도 나타나는 현상이라고 할 수 있다. 그리고 이와 같은 현상이 발생하게 된 원인을 -법적·정책적 한계점을 제외하고- 분석하면 ①공공기관의 연구데이터가 공공데이터라는 것을 연구자와 연구기관이 명확하게 인식하고 있지 않은 점, ②공공기관의 연구데이터의 생성·관리 차원에서 체계적 관리가 이루어지고 있지 않다는 점으로 요약될 수 있다. 따라서 이하에서는 이와 같은 문제점들을 해결하기 위한 법제적 개선방안을 검토하도록 한다.

##### 가. 연구데이터의 공공데이터성 명시

우리의 공공데이터법상 공공데이터의 개념정의 및 공공데이터법의 수범자 범위를 살펴보면, 다양한 형태의 공공기관이 수행한 연구를 통해 수집된 연구데이터는 공공데이터의 개념에 포섭될 수 있다. 따라서 공공기관의 연구데이터는 공공데이터법에 따라 공공기관의 장이 공공데이터 목록에 등록한 후, 공공데이터 전략위원회가 제공대상에 대한 심의·의결하여, 행안부 장관 및 공공기관의 장에 의한 공공데이터 목록 및 이용조건 공표 및 공공데이터 포털 등록을 통해 공공

65) 신은정/손수정/서지영/김영린, 전계서, 109면 참조.



데이터를 이용하고자 하는 자에게 제공될 수 있는 법적 시스템 하에 놓여있다. 그럼에도 불구하고 다양한 연구데이터가 공공데이터로 공개되지 못하는 이유는 연구데이터가 공공데이터인지 아닌지 수범자 입장에서 명확하지 않았던 점, 연구자들이 데이터를 개인 내지 기관의 소유물로 인식하고 있는 점 등에 기인한다고 볼 수 있다.

따라서 공공기관이 생성·수집한 연구데이터 자체가 공공데이터법상 공공데이터에 해당함을 공공데이터법에서 명확히 규율함으로써 연구수행자 및 연구의 주관기관인 공공기관이 연구데이터가 공공데이터임을 명확하게 인식하도록 하는 공공데이터법 개정이 필요할 것으로 생각된다. 그리고 이를 위해 공공데이터로서 공개되는 공공연구데이터의 개념,<sup>66)</sup> 그리고 이러한 공공연구데이터가 갖고 있는 속성과 범주, 공공연구데이터 공개와 관련된 FAIR 원칙, 공공연구데이터의 리포지터리화 등이 공공데이터법에 종합적으로 반영 및 도입되어야 할 것으로 보인다. 이와 같은 규율방식은 앞서도 검토한 PSI-지침에서도 그 선례를 찾아볼 수 있다.

다만 이처럼 공공기관의 연구데이터가 공공데이터로 명확하게 규율하게 되면 공공기관의 연구주체들의 연구가 침해된다거나, 공공기관 연구주체들의 내부적 사정 등으로 인해 연구데이터를 공개할 수 없는 상황임에도 불구하고 연구데이터가 공공데이터로서 공개되는 것은 아닌가와 같은 다양한 형태의 우려가 제기될 수도 있다.<sup>67)</sup>

66) 참고로 연구데이터 자체와 관련한 법적 정의 방식은 「국가연구개발사업의 관리 등에 관한 규정」 제2조 제18호를 참조할 수 있을 것으로 생각된다. 이에 따르면, 연구데이터란 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터”이다. 따라서 공공데이터법에 공공연구데이터의 개념을 도입함에 있어, 이를 참조하여 ‘공공기관이 업무범위 내에서 수행한 연구 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통해 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터’로 공공연구데이터를 정의하는 방식을 검토해볼 수 있다.

67) 이 뿐 아니라 공공연구데이터 개방과 관련하여 저작권 등 기타 제3자의 권리가 문제될 수 있기 때문에 공공연구데이터의 개방이 적합하지 않다는 반론이 제기될 수 있다. 하지만 현행 공공데이터법 제17조는 제1항에서 저작권법 및 그 밖의 다른 법령에서 보호하고 있는 제3자의 권리가 포함된 것으로 해당 법령에 따른 정당한 이용허락을 받지 아니한 정보를 비제공대상 공공데이터로 규율하고 있다. 그리고 동조 제2항에서는 이러한 내용을 기술적으로 분리할 수 있는 경우에는 이를 분리하여 공공데이터를 제공하는 방식을, 그리고 동조 제3항에서는 제3자의 권리를 포함하고 있더라도 이를 공공데이터로 활용할 수 있도록 하기 위한 정당

하지만, 이미 우리의 공공데이터법은 비공개 대상인 공공데이터가 존재할 수 있다는 것을 법적으로 인정하고 있다. 공공데이터법 제9조 제1항은 비공개대상 공공데이터의 유형으로 「공공기관의 정보공개에 관한 법률」 제9조에 따른 비공개대상정보를 명시하고 있다. 그리고 「공공기관의 정보공개에 관한 법률」 제9조 제5호에는 공개될 경우 연구·개발에 현저한 지장을 초래한다고 인정할 만한 상당한 이유가 있는 정보를 비공개 대상정보로 규정하고 있다. 그리고 경우에 따라서는 공공기관의 장이 공공데이터 목록에서 제외하는 것이 필요하다고 판단하는 경우에는 공공기관의 장이 해당 공공데이터를 공공데이터 목록에서 제외할 것을 요청할 수 있고, 이러한 요청을 공공데이터 전략위원회의 심의·의결을 거쳐서 제외할 수 있다(공공데이터법 제20조). 이와 같은 공공데이터 비공개 제도 및 공공데이터 목록 제외 가능성으로 인해 공공기관이 생성한 연구데이터의 공공데이터성 명료화로 인해 발생할 수 있는 예상 밖의 부작용은 해결될 수 있을 것으로 생각된다.<sup>68)</sup>

#### 나. 공공연구데이터 관리 법제 구축

앞서도 살펴본 바와 같이 연구데이터가 활용되지 못하는 주요 이유 중의 하나는 어떠한 내용의 연구데이터를 누가 생산하였는지 체계적으로 알 수 없다는

---

한 이용허락 확보 방안 관련 조치에 대한 사항을 규율하고 있다. 이러한 법적 시스템은 저작권 등 제3자의 권리 보호와 공공연구데이터 개방 간 충돌을 조율할 수 있는 제도적 기제로 작용할 수 있다. 하지만 보다 적극적인 공공연구데이터 개방 및 활용을 촉진하기 위해서는 해당 제도 운영과 관련된 기본권적·지적재산권적·데이터법적 법이론, 제도 운영 관련 세부 기준 및 절차 등이 보다 발전되어야 할 것으로 판단된다.

68) 이와 같은 법적 시스템의 존재와는 별개로 실제 이러한 법적 시스템 하에서 공공연구데이터를 공개할지 비공개할지 결정하는 과정과 관련한 제도적 보완이 필요할 것으로 생각된다. 우선 공공연구데이터를 공개할 것인지 비공개할 것인지를 판단하는 주체들의 전문성 향상이 필요하다. 물론 공공데이터법 시행령 제2조 및 제4조와 제5조에서 공공데이터 전략위원회 및 공공데이터전략실무위원회 구성과 관련하여, 공공데이터관련 전문가를 위원으로 참석하도록 규율하고 있다. 하지만 연구데이터의 경우 해당 연구 분야의 특성 및 전문성도 고려되어야 하므로, 공공연구데이터와 관련된 비공개 여부 판단이 필요한 경우에는 해당 연구 분야의 전문가의 견해가 반영될 수 있는 제도적 장치가 필요할 것으로 생각된다. 따라서 공공연구데이터의 비공개 여부 판단과정에는 해당 데이터를 수집·축적하는데 기여한 개별 연구자 및 해당 연구분야의 연구주체들의 의견수렴과정도 제도화될 필요가 있다. 그리고 공공연구데이터 비공개 결정과 관련한 판단기준의 명료화 등도 제도화될 필요가 있다.

점에 있다. 따라서 고품질의 공공데이터가 될 수 있는 연구데이터가 존재하는지 여부 그 자체가 불분명한 경우도 존재하고, 혹시 관련 데이터가 있다는 사실을 인지하고 있더라도 실제로 어떤 연구자가 어떠한 형식으로 해당 연구데이터를 관리하고 있는지 그리고 해당 데이터가 실제 데이터수요자가 원하는 내용을 담고 있는 것인지 등이 불분명한 문제도 존재한다.

이러한 문제점을 해결하기 위해 국가연구개발법제에서는 2019년 9월 1일부터 ‘데이터관리계획’(Data Management Plan) 제도가 실시되고 있다. 데이터관리계획은 연구개발과제 수행과정에서 산출된 사실자료로서 연구결과의 검증에 필수적인 연구데이터를 체계적으로 관리하기 위해 「국가연구개발사업의 관리 등에 관한 규정」(이하 국가연구개발사업관리규정이라 함)에 도입되었다.<sup>69)</sup>

국가연구개발사업관리규정 제2조 제19호에 따르면 데이터관리계획은 연구데이터를 체계적으로 생산·보존·관리 및 공동활용을 위해서 수립하는 데이터관리계획 제도를 의미한다. 국가연구개발사업관리규정 제6조 제4항에 따르면 국가연구개발사업을 수행하거나 참여하려는 자는 데이터관리계획이 포함된 연구개발계획서를 작성하여 연구개발과제를 신청해야 한다. 그리고 중앙행정기관의 장은 연구개발과제를 선정함에 있어서 데이터관리계획에 따른 연구데이터 생산·보존·관리의 충실성 및 공동활용 가능성을 검토해야 한다(국가연구개발사업관리규정 제7조 제3항 제10호). 그리고 과제협약을 체결함에 있어 협약의 내용으로 데이터관리계획이 포함되어야 하며(국가연구개발사업관리규정 제9호 제1항 제1의2호), 연구개발성과의 최종보고서에도 이러한 데이터관리계획이 포함되어 있어야 한다(국가연구개발사업관리규정 제15조 제2항 제10호). 데이터관리계획에 포함된 연구개발정보는 국가연구개발사업과 관련된 모든 연구개발정보를 관리하기 위한 과제지원시스템을 통해 관리된다(국가연구개발사업관리규정 제25조 제4항 및 제16항 제5호).

그런데 국가연구개발사업과 관련한 연구데이터를 체계적으로 관리하기 위해

69) 국가연구개발사업의 관리 등에 관한 규정[대통령령 제29625호, 2019.3.19. 일부개정] 재정·개정이유, <http://law.go.kr/lstInfoP.do?lsiSeq=208071&ancYd=20190319&ancNo=29625&efYd=20190319&nwJoYnInfo=N&efGubun=Y&chrClsCd=010202&ancYnChk=0#0000>, 최종접속일: 2020.6.8.

도입된 데이터관리계획 제도는 중앙행정기관의 장이 필요하다고 인정하는 연구 개발과제의 경우에만 적용되는 제도이다(국가연구개발사업관리규정 제7조 제3항 제10호, 제9호 제1항 제1의2호, 제15조 제2항 제10호 및 제25조 제4항 및 제16항 제5호). 즉, 중앙행정기관의 장이 인정한 몇몇 국가연구개발사업에 대해서만 연구개발 착수부터 종료 시까지 발생하는 연구데이터의 체계적 관리가 가능한 상황이다.<sup>70)</sup>

따라서 양질의 연구데이터 확보를 위한 연구데이터 전처리 과정의 체계적 관리를 실현하기 위해 데이터관리계획이 적용되는 연구의 범위를 확장할 필요가 있는 것으로 판단된다. 무엇보다 적용대상을 중앙행정기관의 장이 인정하도록 하되, 그 구체적인 기준이나 절차 등을 명확하게 하지 않고 있는 현행 국가연구개발사업관리규정의 규율 방식은 보다 정치한 형태로 개선되어야 할 필요가 있다. 또한 데이터관리계획 제도는 현행 법제상 아직 국가연구개발사업관리규정에만 도입되어 있는 제도이므로, 공공기관을 통해 수행되는 연구데이터 전반에 적용되지 않는다는 제도적 한계도 존재한다. 따라서 연구데이터 생애주기 전과정을 체계적으로 관리할 수 있는 데이터관리계획 제도는 국가연구개발사업 분야를 넘어 공공기관이 생성하는 연구데이터 전반에 도입될 필요가 있다.

다만, 이러한 사항이 국가연구개발사업과 관련된 「과학기술기본법」 및 그 하위 법령의 수준에서 규율되는 방식은 적합하지 않다. 앞서도 검토한 바와 같이 「과학기술기본법」 및 그 하위법령은 국가연구개발사업과 관련된 사항을 규율하고 있기 때문에, 국가연구개발사업이 아닌 형태로 연구를 수행하여 연구데이터를 생성·수집·관리하는 공공기관은 이 법령에 적용을 받지 않기 때문이다. 또한 개별 공공기관의 설립 및 운영을 규율하고 있는 개별 법률에 데이터관리계획 제도를 개별적으로 규율하는 것은 입법기술상 비효율을 초래하기 때문에 타당한 방식은 아닐 것이다.

따라서 공공기관이 갖고 있는 연구데이터 전반에 대해서 적용될 수 있는 형

70) 2019년 데이터관리계획 작성 의무과제는 한국연구재단이 주관기관인 4개 분야 11개 과제와 정보통신기획평가원이 주관기관인 4개 분야 10개 과제에 국한되어 있다. “연구데이터와 DMP 작성”, DGIST CURATION, <https://curation.dgist.ac.kr/curation/w/1555>, 최종접속: 2020.6.8.

태로 규범을 구축하기 위해서는 원칙적으로 공공기관 전체를 수범자로 포섭할 수 있으며, 더 나아가 데이터와 관련된 규율을 담고 있는 공공데이터법을 통해 공공기관이 연구를 수행함에 있어 데이터관리계획을 수립·운영하고, 이를 체계적으로 관리할 수 있는 제도를 구축하는 방식이 타당한 것으로 보인다.

물론 현행 공공데이터법의 경우 공공기관의 장이 공공데이터 목록을 등록하는 방식을 취하고 있어(공공데이터법 제18조 제1항), 굳이 데이터관리계획 제도의 도입은 불필요한 것으로 보일 소지가 있다. 실제로 공공데이터법 제18조 제2항에 따르면 공공데이터 목록에 누락된 공공데이터의 경우 행정안전부장관이 공공기관의 장에게 이를 등록할 것을 요청할 수 있도록 되어 있기도 하다. 그런데 어떠한 연구데이터가 있는지 체계적으로 관리가 되고 있지 않은 상황에서 공공데이터 목록에서 누락된 연구데이터를 공공데이터로 등록하라고 요청하는 것은 사실상 불가능하다. 따라서 공공데이터법에서 연구데이터관리 제도를 도입하여, 공공기관의 장에 의한 공공데이터 목록 등록과는 별개로, 공공데이터에 해당하는 연구데이터의 생성부터 폐기까지 전주기를 체계적으로 관리하는 제도를 도입·운영하는 것이 타당할 것으로 생각된다.

#### IV. 끝으로

지금까지 인공지능이 다양한 분야에서 활용되기 위해서는 다양한 형태의 고품질 데이터가 필요하며, 이를 마련하기 위한 지원 정책적 방안으로 공공데이터를 제공하는 국내외 정책적 동향을 살펴보았다. 그리고 인공지능 학습데이터로 공공데이터를 제공하는 법정정책적 방식이 간과하고 있는 공공기관이 생성한 연구데이터의 활용에 대해서 살펴보았다. 특히 이를 명시적을 규정하고 있는 유럽 연합의 PSI-지침의 주요 내용을 검토한 후, 우리나라의 공공데이터법이 공공기관이 갖고 있는 연구데이터를 공공데이터로 활용할 수 있는 법적 장치를 마련하고 있는지 진단하였다. 그리고 공공기관의 연구데이터가 공공데이터로 활용될 수 있는 법적 제도가 공공데이터법에 구축되어 있음에도 불구하고 공공기관의 연구데이터가 활용되지 않고 있는 이유는 무엇인지 살펴보았으며, 이를 개선하

기 위한 법제적 대안을 검토해 보았다.

그 구체적인 대안으로는 첫 번째로, 공공기관의 연구데이터가 공공데이터임을 명시적으로 규율하여 공공기관의 장 및 공공기관의 연구자가 연구데이터가 공공데이터라는 점을 명확히 인식할 수 있도록 하는 것이 필요하다는 점과 공공연구데이터와 관련하여 연구데이터의 개념 및 그 속성 및 관리 방식, 공공연구데이터에 대한 FAIR 원칙 등이 공공데이터법에 명시적으로 도입되어야 한다는 점을 제시하였다. 두 번째로, 공공기관의 연구데이터를 보다 적극적으로 활용하기 위한 제도적 보완책으로 현행 국가연구개발사업관리규정에 도입되어 있는 데이터관리계획 제도의 확장 필요성과 제도적 투명성 제고 필요성, 그리고 공공데이터법에 데이터관리계획 제도를 도입하는 것이 필요하다는 것을 제시하였다.

앞서 인공지능 기술동향의 예시에서 언급된 사례들에서 알 수 있듯이, 고품질의 연구데이터를 학습한 인공지능은 단순히 향상된 서비스를 제공하는 수준의 인공지능을 넘어 순수과학분야의 새로운 발견을 도출해내는 인공지능으로 발전할 수 있는 가능성을 갖고 있다. 따라서 인공지능을 활용한 기술발전을 선도하기 위해서는 양질의 연구데이터가 개방되어 다양한 연구 분야에 활용되어야 한다. 지능정보사회에서 데이터의 가치가 중시되고 있다는 논의가 우리나라에서도 다양한 영역에서 제기되고 있으나, 실상 그 내용을 살펴보면 상업적으로 즉시 활용될 수 있는 데이터, 특히 개인과 관련된 데이터에 그 초점이 놓여 있다는 것을 알 수 있다. 하지만 이러한 즉물적 접근방식으로부터 시선을 돌려 기초연구데이터가 인공지능에 접목되면 보다 진일보한 과학적 발견을 가능하게 한다는 점을 명확하게 인식할 필요가 있다. 특히 다양한 기초연구데이터를 활용한 인공지능 개발은 기초과학적 연구역량과 인공지능 연구역량 동시에 향상시킬 수 있는 가능성을 갖고 있다는 점이 강조되어야 하는 시점이다. 국제적으로 인공지능 기술 개발 경쟁이 심화되는 가운데, 개인 연구실 또는 연구기관 내에 숨겨진 원석인 연구데이터를 적극 활용할 수 있도록 하는 법정정책적 접근방식을 토대로 국가 기술 경쟁력 향상을 적극적으로 모색해야 하며, 이를 위한 첩경은 공공데이터 법제 개선을 통한 공공기관의 연구데이터 개방성 향상 및 연구데이터 관리 제도의 개선이라고 할 수 있다.

## 참고문헌

### [국내문헌]

- 관계부처 합동, 「2020년 지능형 로봇 실행계획」, 2020.
- 관계부처 합동, 「인공지능 국가전략」, 2019.
- 관계부처 합동, 「데이터 산업 활성화 전략」, 2018.
- 국회 행정안전위원회, 「공공데이터의 제공 및 이용 활성화에 관한 법률안 검토 보고서」, 2012.
- 권성훈, “국가초고성능컴퓨터 활용 및 육성에 관한 법률의 입법영향분석”, 「한국정책학회 춘계학술발표논문집」, 한국정책학회, 2016.
- 권은정, “공공데이터 영역 리스크 관리에 관한 법적 소고”, 「행정법연구」 제60호, 행정법이론실무학회, 2020.
- 김민호/이보옥, “공공 오픈데이터 EU Directive 개정 동향과 시사점”, 「성균관법학」 제32권 제1호, 성균관대학교 법학연구원, 2020.
- 김중권, “행정에 인공지능시스템 도입의 공법적 문제점”, 「법조」 통권 740호, 법조협회, 2020.
- 김중권, “인공지능시대에 알고리즘에 의한 행위조종과 가상적 행정행위에 관한 소고”, 「공법연구」 제48집 제3호, 한국공법학회, 2020.
- 김중권, “인공지능시대에 완전자동적 행정행위에 관한 소고”, 「법조」 통권 723호, 법조협회, 2017.
- 노진아, “인간과 기계의 공진화 - 인공지능 로보틱스 아트 “제페토의 꿈”을 중심으로”, 「Contents Plus」 제16권 No.1, 한국영상학회, 2018.
- 문희진/강남규/이상환/정용일, “연구자 중심의 과학기술 데이터 생태계 활성화 전략”, 「KISTI Issue Brief」 제19호, 한국과학기술정보연구원, 2020.
- 손현, 「진흥관련 법제의 입법 모델 연구」, 한국법제연구원, 2016.
- 신은정/손수정/서지영/김영린 「공공부문 연구데이터의 소유·활용제도 개선방향」, 과학기술정책연구원, 2019.

- 이상민/김은희, “글로벌 인공지능 정책과 전략적 과제”, 「KISTI Issue Brief」 제 17호, 한국과학기술정보연구원, 2020.
- 이재훈, “유럽 주요국 인공지능 정책동향 분석”, 「이슈페이퍼」 19-19-2, 한국 법제연구원, 2019.
- 이재훈, “자율주행자동차 관련 국내 법제 동향”, 「경제규제와 법」 제12권 제2호, 서울대학교 법학연구소, 2019.
- 이재훈, “유럽연합 규범체계에 대한 체계적 고찰을 위한 소고”, 「성균관법학」 제30권 제4호, 성균관대학교 법학연구원, 2018.
- 이재훈, “전자동화 행정행위 도입 후 독일 행정절차법제 논의의 전개양상”, 「공법학연구」 제19권 제4호, 한국비교공법학회, 2018.
- 이재훈, “전자동화 행정행위에 관한 연구”, 「성균관법학」 제29권 제3호, 성균관대학교 법학연구원, 2017.
- 정영미/배정희, “기술확산 통합모델을 통한 개방형 기관 리포지터리 수용의 영향요인 분석”, 「한국도서관정보학회지」 제46권 제4호, 한국도서관정보학회, 2015.
- 정재림/김상욱, “시스템다이내믹스를 활용한 기술과 사회의 공진화 연구”, 「한국경영학회 통합학술발표논문집」, 한국경영학회, 2009.
- 정재현, “AI와 인간의 공진화共進化와 관련해서 본 다산 철학”, 「다산학」 제35호, (재)다산학술문화재단, 2019.
- 최명석/이승복/이상환, “국내 과학기술분야 연구기관의 과학데이터 관리 현황”, 「한국콘텐츠학회논문지」 vol. 17 no.12, 한국콘텐츠학회, 2017.
- 최명석/김현우/이상환, “국가 연구데이터 공유의 시작, 데이터 관리계획(Data Management Plan)”, 「KISTI Issue Brief」 제13호, 한국과학기술정보연구원, 2019.
- 한국과학기술정보연구원, 「출연(연) 연구데이터 관리·활용 방안 연구」, 국가과학기술연구회, 2019.
- 한상수/김동훈, “인공지능, 소재 개발에 침투하다”, 「융합연구리뷰」 vol. 5 no. 1, 융합연구정책센터, 2019.
- 한세억, “제4차 산업혁명시대 기술과 제도의 공진화”, 「서울행정학회 동계학술대회 발표논문집」, 서울행정학회, 2018.



**[해외문헌]**

- Datenethikkommission der Bundesregierung(Hrsg.), *Gutachten der Datenethikkommission*, Berlin: Brandenburgische Universitätsdruckerei und Verlagsgesellschaft Potsdam mbH, 2019.
- Hof, Hagen/Wengenroth, Ulrich(Hrsg.), *Innovationsforschung*, Hamburg: LIT, 2007.
- Hoffmann-Riem, Wolfgang, *Innovation und Recht – Recht und Innovation*, Tübingen: Mohr Siebeck, 2016.
- Wilkinson, Mark D., et al., “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3, 160018. doi:10.1038/sdata.2016.18.

**[Abstract]**

**A Study on Promoting the Re-Use of Public Research  
Data as AI(Artificial Intelligence) Learning Data**

**Lee, Jae-Hoon**

*Korea Legislation Research Institute, Research Fellow*

The core of modern technology is artificial intelligence technology. Therefore, national support for AI innovation and legal framework for supporting AI innovation are needed. There are some legislation in Korea that can actually be used to support the development of artificial intelligence technology. For example, the Intelligent Robot Act and Ultra High Performance Computer Act could be used to support development of artificial intelligence technology.

However, although artificial intelligence learning data is the core of the development of artificial intelligence technology, the Intelligent Robot Act and the Ultra High Performance Computer Act do not have any support method related to artificial intelligence learning data.

Looking at artificial intelligence strategies in foreign countries and artificial intelligence strategies in Korea, policies to open public data for re-use are being considered from the perspective of providing learning data for the development of artificial intelligence technologies. And in the case of the European Union, it is trying to support the promotion of artificial intelligence technology through high-quality public data. What should be noted in the PSI-Directive is that research data from public research institutes, which are of high quality, are defined as public data and are open to the public for re-use. The PSI-Directive stipulate principles and conditions for opening public research data and its re-use.

Korea also has various research data that have characteristics as public data. However, research data, which has a character as public data, is not actively being opened and re-used. The reason for this is that it is not clearly stated that research data from public institutions are public data. And the lack of sufficient systems to systematically manage research data from public institutions is another reason. To solve these problems, this study analyzes the PSI-Directive of European Union, the Act on Promotion of the Provision and Use of Public Data, and the regulations on the management of national R&D projects, and examines legal system under which research data of public institutions can be opened as artificial intelligence learning data.

**Key words** : public data, research data, artificial intelligence, data management plan, PSI-Directive