

RESEARCH ARTICLE

A Study on Legal Standards for Collecting Open Data

Su-Yun Go

Senior Researcher, Institute of Comparative Legal Studies, School of Law, Kangwon National University, Republic of Korea

Corresponding Author: Su-Yun Go (rhtndbs@naver.com)

ABSTRACT

Open data have long been recognized as common resources. However, recent domestic and international disputes show that even their collection through automated technologies may exceed the scope of current laws. Relevant laws include the CRIMINAL ACT, the ACT ON PROMOTION OF INFORMATION AND COMMUNICATIONS NETWORK UTILIZATION AND INFORMATION PROTECTION, the PERSONAL INFORMATION PROTECTION ACT, the COPYRIGHT LAW, the UNFAIR COMPETITION PREVENTION AND TRADE SECRET PROTECTION ACT, and the MONOPOLY REGULATION AND FAIR TRADE ACT. Among these, the most relevant are the Rights of Database Producers under the COPYRIGHT LAW (Article 93) and the prohibition on infringing another's economic interests by unauthorized using the outcomes without permission under the UNFAIR COMPETITION PREVENTION AND TRADE SECRET PROTECTION ACT (Article 2(m)). Comparison of these provisions shows that the requirement of fair practice is a key factor in determining the legality of collecting open data. In this context, fair practice generally entails compliance with Terms of Use and Robot Exclusion Standard. Terms of Use may be legally construed as contractual terms, while the Robot Exclusion Standard can be construed as either contractual terms or the public notice. Conversely, prohibiting the collection of unique data may constitute an abuse of a dominant market position or unfair trade practices under the MONOPOLY REGULATION and FAIR TRADE ACT. Thus, the collection of unique data should be deemed as lawful. This study discusses the lawfulness of collecting open under the existing law; however, future discourse must also discuss the lawfulness of collecting new forms of open data outside the framework of the current law.

KEYWORDS

Crawling, Open data, UNFAIR COMPETITION PREVENTION AND TRADE SECRET PROTECTION ACT, COPYRIGHT LAW, Fair practice, Terms of use, Robots exclusion standard

Open Access

Received: June 09, 2025
Revised: June 15, 2025
Accepted: August 27, 2025
Published: September 30, 2025

Funding: The author received manuscript fees for this article from Korea Institute of Intellectual Property.

Conflict of interest: No potential conflict of interest relevant to this article was reported.

© 2025 Korea Institute of Intellectual Property



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

원저

공개 데이터 수집의 적법 기준에 대한 소고

고수윤

강원대학교 비교법학연구소 선임연구원

교신저자: 고수윤 (rhtndbs@naver.com)

차례

1. 서론
2. 공개 데이터 수집 기술 관련 법률의 검토
 - 2.1. 공개 데이터 수집과 관련된 자동화 기술
 - 2.2. 공개 데이터 수집시 저촉될 수 있는 법률
3. 공개 데이터 수집 기준 정립의 방향
 - 3.1. 현행 기준에 대한 평가
 - 3.2. 공개 데이터 수집 기준 정립의 방향
4. 결론

국문초록

공개된 데이터는 공유물로 인식되어 왔다. 하지만 최근 일어난 일련의 국내외 분쟁을 보면 자동화된 기술을 통해 공개 데이터를 수집하는 경우라도 현행법에 저촉될 수 있음을 알 수 있다. 관련된 법률에는 형법, 정보통신망법, 개인정보보호법, 저작권법, 부정경쟁방지법, 독점규제법 등이 있다. 그러나 이중 저작권법상 데이터베이스제작자의 권리와 부정경쟁방지법상 성과모용행위에 대한 규정이 가장 관련성이 깊다. 위의 두 규정을 비교하면, 특히 공정한 관행이라는 요건이 공개 데이터 수집에 대한 적법성 확보 측면에서 중요한 단서를 제공한다. 여기서 말하는 공정한 관행은 일반적으로 이용약관과 로봇배제표준을 준수하는 것으로 파악된다. 여기서 이용약관은 법적으로 약관계약으로 볼 수 있고 로봇배제표준은 약관계약 또는 공시방법으로 해석될 수 있을 것이다. 한편, 유일한 데이터에 대한 수집을 금지하는 행위는 독점규제법상 시장지배적지위의 남용이나 불공정거래행위가 될 수 있으므로 유일한 데이터의 수집은 언제나 적법한 것으로 판단해야 할 것이다. 이상에서는 현행법에 기초하여 공개 데이터의 수집에 대한 적법 기준을 논의하였지만 앞으로는 현행법 틀을 벗어나 새로운 공개 데이터 수집의 적법 기준에 대한 논의가 필요할 것이다.

주제어

크롤링, 공개 데이터, 부정경쟁방지법, 저작권법, 공정한 관행, 이용약관, 로봇배제표준

1. 서론

봉이 김선달은 대동강물을 팔아 최대의 사기꾼으로 회자되고 있다.¹⁾ 생수를 사먹는 것이 일상이 된 현대에 살았다면 봉이 김선달은 시대를 조금 앞선 것뿐인 것 같아 억울함을 가질 수도 있을 듯 싶다. 데이터 특히 공개된 데이터는 공유물로 인식되어 왔다. 하지만 빅데이터 시대에 원유로 여기는 데이터에 대하여 이제는 사유화의 필요성이 대두되고 있다. 이러한 흐름을 반영 하듯 우리나라는 2022년 4월 22일에는 데이터 거래 활성화를 목적으로 하는 데이터 산업진흥 및 이용촉진에 관한 기본법이 시행되었고 2018년 일본의 부정경쟁방지법상 신설된 한정제공 데이터 보호와 유사하게 데이터부정사용행위를 규제하는 규정을 신설하여 2022년부터 시행하고 있다. 이러한 법률 개정의 상황 뿐만아니라 실제 공개 데이터의 수집과 관련하여 리그베다 위키-엔하위키 미러 사건,²⁾ 잡코리아-사람인 사건,³⁾ 야놀자-여기어때 사건,⁴⁾ 네이버-다윈중개 사건⁵⁾ 등과 같은 분쟁사례들도 나오기 시작하였다.

이러한 사건들 중에는 이용자들이 대부분의 데이터를 생성하고 수정하는 사이트의 사업자에게 데이터베이스제작자의 권리를 인정한 사례⁶⁾가 있었고, 공개 데이터의 사용 자체가 타인의 권리침해가 된다는 판결도 있었다. 이러한 판결을 볼 때, 공개 데이터를 자동화된 기술로 수집하여 사용하는 사업자들은 법적 리스크에 대한 불안이 커질 수 밖에 없다.

이러한 불안을 증폭시키는 것은 현행법상 공개 데이터 수집에 침해 기준의 불명확성에 있다고 할 수 있다. 산발적으로 규정되어 있는 관련 법률 조문의 통일성 결여, 개별 기준의 불분명한 해석 여지, 충분한 판례 축적이 되지 않은 현상황, 실무적으로 적용할 수 없는 요건들이 수범자의 법적 안정성을 저해하고 있는 것이다. 이를 해결하기 위해서는 공개 데이터의 자동화된 수집이 적법성을 가질 수 있는 기준마련이 필요하다고 생각한다. 이에 본고는 현재 공개 데이터의 자동화된 수집을 데이터에 대한 권리 등의 침해를 야기할 수 있는 관련 법률을 검토하고 이를 토대로 공개 데이터 수집 기준 정립의 바람직한 방향에 대하여 제시하고자 한다.

2. 공개 데이터 수집 기술 관련 법률의 검토

2.1. 공개 데이터 수집과 관련된 자동화 기술

인력이 아닌 기술을 통해 자동화된 데이터 수집의 대표적인 방법에는 해킹, API, 크롤링, 스크래핑, 미러링이 있다. 해킹(security hacking)은 타인의 통신망에 정당한 접근 권한 없이 접근하거나 허용된 접근 권한의 범위를 초과하여 침입하는 행위를 말한다.⁷⁾ 해킹은 데이터를 복제하는 수집 단계 전에 일어나는 침입행위이며, 이는 법적으로 불법이다. 기술적 보호조치가 없는 ‘공개’ 데이터의 수집에는 상관없는 기술이므로 이를 제외하면 공개 데이터의 자동 수집과 관련된 기술에는 다음과 같은 방법이 있다.⁸⁾

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5B5A16078550).

1) 송종복, 「대동강물 사기꾼 봉이 김선달」, 경남매일, 2015. 5. 15자.

2) 서울고등법원 2016. 12. 15. 선고 2015나2074198 판결(대법원 2017. 4. 13. 선고 2017다204315 판결(심리 불속행)로 확정됨)

3) 대법원 2017. 8. 24 선고 2017다224395 판결.

4) 대법원 2022. 5. 12. 선고 2021도1533 판결.

5) 서울중앙지방법원 2024. 9. 27. 선고 2022가합528041 판결.

6) 대표적으로 리그베다 위키 대 엔하위키 미러 사건을 들 수 있다.

7) 위키백과, “해킹”, 위키백과, <<https://ko.wikipedia.org/wiki/%ED%95%B4%ED%82%B9>>, 검색일: 2025. 6. 6.

8) 데이터마이닝도 수집기술 중 하나로서 언급되기도 하는데 이는 공개 데이터를 수집하는 기술이 아니고 이미

2.1.1. API

API(Application Programming Interface)는 서로 다른 소프트웨어 응용 프로그램 간에 데이터를 교환하고 통신할 수 있도록 정의된 표준화된 인터페이스를 의미한다.⁹⁾ API는 서비스 간 효율적이고 안정적인 데이터 교환과 통합된 서비스 제공을 목적으로 한다. 식당에서 메뉴판을 보고 정해진 음식만 주문하는 행위처럼 특정 데이터를 제공하도록 하는 기술이며 실제로 구글 지도 API는 위치 데이터 및 지도 정보를 제공하는 것을 들 수 있다.

API는 일종의 공식적이고 체계적인 소프트웨어 계약으로서, 특정 서비스나 데이터를 요청하고 결과를 받기 위한 절차와 형식을 명확히 규정하고 있다. 일반적으로 서비스 제공자가 정의한 명확한 약관과 규칙을 따르게 된다.

2.1.2. 크롤링

크롤링(Crawling)은 웹사이트를 체계적이고 자동화된 방식으로 탐색하여 대량의 웹페이지 및 데이터를 수집하는 기술이다. 일반적으로 웹 크롤러(Web Crawler)라고 부르는 프로그램이 URL에서 시작하여 링크를 따라가면서 웹사이트의 구조를 따라 반복적으로 콘텐츠를 방문하고 수집한다.¹⁰⁾ 주로 웹페이지 내용을 통째로 수집하거나, URL 목록과 같이 넓은 범위의 콘텐츠를 빠르게 확보하는 데 목적이 있다. 이는 도서관에서 책 제목과 목차를 훑어보면서 자료를 찾는 것과 유사한 것으로, 대표적으로 구글, 네이버와 같은 검색엔진들이 검색결과를 제공하기 위해 정기적으로 인터넷 전반을 탐색하여 방대한 웹페이지를 인덱싱할 때 크롤링을 활용한다.

일반적으로 크롤링은 로봇배제표준(Robots.txt)과 서비스 이용약관에 명시된 규정을 따르며 비정상적 트래픽을 유발시키는 디도스(DDos) 공격이 되지 않도록 대상 서버에 과도한 트래픽과 부하를 일으키지 않도록 해야 한다.

2.1.3. 스크래핑

스크래핑(Scraping)은 웹페이지에서 필요한 특정 데이터를 자동으로 추출하여 구조화된 형태로 저장하고 활용하는 기술이다. 일반적으로 컴퓨터 프로그램간의 데이터 교환은 상호 합의된 프로토콜이나데이터 구조를 통해서 수행되나 스크래핑은 웹사이트페이지나 프로그램 화면과 같이 주로 사람이 읽도록 만들어진 콘텐츠에서 다른 프로그램이 자동으로 데이터를 추출하는 방법이다.¹¹⁾ 크롤링이 대량의 웹페이지를 넓게 탐색하는 데 초점을 둔다면, 스크래핑은 페이지 내부의 특정 정보만을 정밀하게 선택하여 뽑아내는 데 주력한다. 비유하자면 책에서 내가 필요한 문장만 필사하는 행위로, 전자상거래의 가격 비교 서비스, 시장 조사 데이터 수집, 뉴스 요약 서비스 등에서 사용된다.

크롤링 보다 좁은 범위지만 보다 정밀하게 데이터를 추출하므로 웹사이트의 이용약관과 저작권, 개인정보보호 규정을 준수하며 데이터를 수집해야 한다.

저장되어 있는 데이터베이스에서 추출하여 분석하는 기술이라는 점에서 차이가 있다.

9) AWS, “애플리케이션 프로그래밍 인터페이스(API)란 무엇인가요?”, AWS, <<https://aws.amazon.com/ko/what-is/api/>>, 검색일: 2025. 6. 6.

10) 한국정보통신기술협회, “크롤링”, 정보통신용어사전, <https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=179640-1>, 검색일: 2025. 6. 6.

11) 한국정보통신기술협회, “스크래핑”, 정보통신용어사전, <https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=185196-1>, 검색일: 2025. 6. 6.

2.1.4. 미러링

미러링(Mirroring)은 특정 웹사이트 또는 온라인 콘텐츠의 전체적인 모습을 그대로 복제하여 별도의 저장소(서버)에 동일한 형태로 보존하는 기술이다. 미러링의 주요 목적은 서버의 부하를 분산시키고, 콘텐츠를 여러 지역에 분산 배치하여 사용자들에게 빠른 접근성을 제공하는 것이며, 원본 콘텐츠가 손실되거나 훼손될 경우를 대비한 백업 및 보존을 위해서도 활용된다.¹²⁾ 이는 마치 책 전체를 복사해서 집에 가져오는 것과 같이 미러링 기술은 원본 웹사이트와 완전히 똑같은 구조와 내용을 유지하며 데이터와 자원의 완전한 복제를 목표로 한다.

미러링은 전체 데이터의 복제를 하기 때문에 저작권, 개인정보보호 규정 등 법적 이슈가 있어 콘텐츠 복제 및 재배포시에는 사전 허가나 명시적 동의를 받아야 한다.

2.2. 공개 데이터 수집시 저촉될 수 있는 법률

공개 데이터를 수집할 경우 저촉될 수 있는 법에는 형법, 정보통신망법, 개인정보보호법, 저작권법, 부정경쟁방지법, 독점규제법이 있다.

2.2.1. 형법상 컴퓨터등장애업무방해죄 적용가능성

형법 제314조 제2항 컴퓨터등장애업무방해죄는 “컴퓨터등 정보처리장치 또는 전자기록등 특수매체기록을 손괴하거나 정보처리장치에 허위의 정보 또는 부정한 명령을 입력하거나 기타 방법으로 정보처리에 장애를 발생하게 하여 사람의 업무를 방해한 자”를 처벌한다.

‘허위의 정보 또는 부정한 명령의 입력’이란 객관적으로 진실에 반하는 내용의 정보를 입력하거나 정보처리장치를 운영하는 본래의 목적과 상이한 명령을 입력하는 것이고, ‘기타 방법’이란 컴퓨터의 정보처리에 장애를 초래하는 가해수단으로 컴퓨터의 작동에 직접·간접으로 영향을 미치는 일체의 행위를 말한다. 해킹과 같은 기술은 해당할 수 있지만 일반적으로 공개데이터를 수집하는 기술들은 허위의 정보 또는 부정한 명령의 입력을 하지 않게 된다.¹³⁾ 또한 동죄는 위와 같은 가해행위 결과 정보처리장치가 그 사용목적에 부합하는 기능을 하지 못하거나 사용목적과 다른 기능을 하는 등 정보처리에 장애가 현실적으로 발생하여야 한다. 단순히 접속 횟수가 증가한 것만으로는 적용되지 않는 것으로 보고 있다. 일반적으로 공개 데이터 수집 기술들은 대상 서버에 과도한 트래픽과 부하를 일으키지 않도록 사용되기에 본 요건이 적용될 가능성도 적고 특히 판례는 접속장애가 있다 하더라도 자연 이용자 증가에 따른 것일 수 있다는 점까지 고려하고 있기 때문에¹⁴⁾ 공개데이터의 수집기술이 사용될 시 동 조문이 적용될 여기가 거의 없다고 판단된다.

2.2.2. 정보통신망법상 정보통신망침입죄 적용가능성

정보통신망법 제48조 제1항 정보통신망침입죄는 “누구든지 정당한 접근권한 없이 또는 허용된 접근권한을 넘어 정보통신망에 침입하여서는 아니 된다”라고 규정하고 있다. 이는 ‘정당한 접근권한’이 없이 ‘침입’하는 행위를 하여야 적용되는 규정이다. 정당한 접근권한이 있는지 여부는 서비스제공자에 의해 정해지며¹⁵⁾ 서비스제공자가 자신의 웹 사이트 접근을 막기 위하여

12) 나무위키, “미러링”, 나무위키, <<https://namu.wiki/w/%EB%AF%B8%EB%9F%AC%EB%A7%81>>, 검색일: 2025. 6. 6.

13) 대법원 2022. 5. 12 선고 2021도1533 판결.

14) 서울중앙지방법원 2021. 1. 13 선고 2020노611 판결.

기술적 보호조치를 하고 있는지, 구체적으로 이용자의 사용범위나 접근권한을 이용약관에 명시하였는지와 같은 객관적 상황에 기초하여 판단하게 된다.¹⁶⁾ 일반적으로 공개 데이터 수집 기술은 이용자의 사용범위를 준수하게 된다.

한편, 침입행위가 인정되기 위해서는 접근통제 조치가 전제되어야 한다. 일반적으로 접근 통제조치가 없는 공개 데이터를 수집하는 경우에는 적용될 여지가 거의 없을 것이다. 실제로 크롤링 기술은 정보통신망 침입에 해당하지 않는다고 판단된다.¹⁷⁾

2.2.3. 개인정보보호법상 개인정보 보호 규정 적용여부

개인정보의 경우에는 정보주체의 동의 내지 법령상 규정 등 불가피한 경우가 아니면 제3자가 이를 임의로 수집할 수 없다(개인정보 보호법 제15조). 공개 데이터를 수집시 개인정보가 포함된다면 개인정보보호법에 적용을 받게 되며¹⁸⁾ 일반 데이터와 구별하여 개인정보보호법으로 침해여부를 판단하게 될 것이다. 개인정보는 일반 데이터와 개념 정의 면에서도¹⁹⁾ 정보의 보호 목적의 면에서도²⁰⁾ 상이하여 같은 궤를 가져가는 것이 적절하지 않을 수 있다. 특히 일반적인 공개 데이터의 경우에는 제공자와 이용자 양자간의 관계가 형성되지만 개인정보의 경우에는 정보주체와의 관계까지 3자 간의 관계를 고려해야 하는 형태라는 차이도 존재한다.

본고는 자동화 기술을 통한 공개 데이터 수집에 대한 것이므로 개인정보에 관한 것은 제외한다. 이와 관련하여 최근 개인정보위원회에서는 인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 공개하여 인공지능(AI) 개발·서비스에 이용되는 ‘공개 데이터’ 처리 기준 제시하고 있다.

2.2.4. 저작권법상 데이터베이스제작자의 권리²¹⁾ 침해 가능성

공개 데이터의 수집시 일반데이터 뿐만 아니라 저작물도 수집될 수 있다. 저작물의 경우에는 원칙적으로 사전에 이용허락을 받아야 한다. 공개된 저작물이라 하더라도 동일하다. 본고는 사전에 이용허락이 불가능한 자동화 기술을 통한 공개 데이터 수집에 대한 것이므로 저작물의 수집에 대한 것은 별론으로 하고 데이터베이스에 초점을 맞춘다.

15) 대법원은 2005. 11. 25. 선고 2005도870 판결에서 “위 규정에서 접근권한을 부여하거나 허용되는 범위를 설정하는 주체는 서비스제공자”라고 한다.

16) 선중수, “크롤링을 통한 데이터 수집과 형사책임 - 대법원 2022.5.12. 선고 2021도1533 판결을 중심으로”, 「형사법연구」, 제35권 제3호(2023), 322면.

17) 이에 대한 최초의 판결은 대법원 2022. 5. 12. 선고 2021도1533 판결.

18) 부정경쟁방지법 제15조에 따르면 데이터 내에 개인정보가 포함되어 있는 경우 「개인정보 보호법」이 우선 적용되도록 하고 있다.

19) 법적으로 데이터는 데이터의 생성 방식, 존재 상태같은 물리적 성질로 정의된 반면에 개인정보는 비물리적인 내재적 의미를 반영한 정의를 하고 있다. 구체적으로 “데이터”란 다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및 「소프트웨어 진흥법」 제2조 제1호에 따른 소프트웨어 등을 통하여 생성된 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보를 말하고(데이터산업법 제2조 제1호), “개인정보”란 살아 있는 개인에 관한 정보를 의미한다(개인정보보호법 제2조 제1호).

20) 개인정보는 기본적으로 보호하여 사생활의 비밀과 자유를 보장하는데 방점이 있지만, 데이터는 활용을 통해 얻게 된 경제적 가치를 보호하는데 방점이 있다. 관련 법의 목적을 살펴보면, 개인정보보호법은 사생활의 비밀과 자유에서 시작하여 발전된 것으로 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 한 반면, 데이터산업법의 경우 데이터로부터 경제적 가치를 창출하고 데이터산업 발전의 기반을 조성하여 국민생활의 향상과 국민경제의 발전에 이바지함을 목적으로 함을 알 수 있다.

21) “편집물”은 데이터베이스를 포함하는 개념으로 그 소재의 선택·배열 또는 구성에 창작성이 있다면 당연히 저작물로도 보호가 될 수 있다. 다만 본고는 저작물이 아닌 데이터의 수집에 초점을 두기 위해 이에 대한 설명은 생략한다.

2.2.4.1. 데이터베이스제작자의 권리

데이터베이스제작자는 그의 데이터베이스의 전부 또는 상당한 부분을 복제·배포·방송 또는 전송할 권리를 가지며 이때 데이터베이스의 구성부분이 되는 개별소재 즉 데이터 자체를 보호하지는 않는다(저작권법 제93조).²²⁾ 권리는 저작권법에 따라 준물권으로 보며 보호기간은 5년이다(저작권법 제95조).

한편, 동 권리는 공표된 저작물의 인용(저작권법 제28조)²³⁾ 등에 의해 제한되며 당연히 공정 이용 규정²⁴⁾도 적용된다(저작권법 제94조).

2.2.4.2. 데이터베이스와 데이터의 관계

데이터베이스제작자의 권리의 객체는 데이터베이스이며 저작권법에서는 “데이터베이스”를 ‘소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것을 말한다’라고 정의하고 있다(저작권법 제2조 제19조). 법률상 보통 데이터는 전자매체에 기록된 사항으로 표현되고 있는데 반해²⁵⁾ 저작권법상 데이터베이스는 컴퓨터에 의하여 전자적 형태로 자료를 저장·추출할 수 있는 것만이 아니라, 사전이나 전화번호부, 인명부, 주소록 등 비전자적 형태의 편집물도 포함되는 개념이다.²⁶⁾

한편 이는 데이터 자체를 보호하는 것이 아니라 데이터베이스를 보호함을 원칙으로 하고 있지만 자동화된 기술로 웹상의 데이터를 수집하는 것은 데이터베이스 수집과 다를 바 없다. 아날로그 데이터와 달리 디지털 데이터는 저장된 위치만 알 수 있으면 검색 가능한 상태이며, 모든 데이터는 컴퓨터 서버에 있는 데이터를 검색하여 불러온 것이므로 현대 사회에서 인터넷으로 검색할 수 있는 모든 데이터는 데이터베이스로부터 추출되는 것이나 마찬가지이기 때문이다.²⁷⁾

2.2.4.3 데이터베이스제작자의 의의와 광범위한 보호

권리의 주체는 데이터베이스제작자이다. “데이터베이스제작자”는 데이터베이스의 제작 또

22) 여기서 복제와 배포는 데이터베이스의 전부 또는 상당한 부분을 유형물로 복제하여 거래에 제공하는 것이며, 방송과 전송은 무형적으로 무선 또는 유선통신의 방법에 의하여 송신하거나 또는 다른 사람이 개별적으로 선택한 시간과 장소에서 접근할 수 있도록 이용에 제공하는 것이다(허희성, 「신저작권법 초조개설하」, 명문프릭컴, 2011, 477면).

23) 제28조(공표된 저작물의 인용) 공표된 저작물은 보도·비평·교육·연구 등을 위하여는 정당한 범위 안에서 공정한 관행에 합치되게 이를 인용할 수 있다.

24) 제35조의5(저작물의 공정한 이용) ① 제23조부터 제35조의4까지, 제101조의3부터 제101조의5까지의 경우 외에 저작물의 일반적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다.

② 저작물 이용 행위가 제1항에 해당하는지를 판단할 때에는 다음 각 호의 사항등을 고려하여야 한다.

1. 이용의 목적 및 성격

2. 저작물의 종류 및 용도

3. 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성

4. 저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향

25) 「데이터산업법」 제2조 제1호에 따르면 “데이터”란 다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및 「소프트웨어 진흥법」 제2조 제1호에 따른 소프트웨어 등을 통하여 생성된 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보를 말하고 있으며, 「공공데이터법」 제2조 제2호에 따르면 “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 말한다고 한다. 고수원, “빅데이터 기술환경에서 정보의 이용에 따른 침해판단방법론”, 강원대학교 대학원, 박사, 2019, 7면.

26) 오승중, 「저작권법」, 제5판, 박영사, 2021, 1061면.

27) 김현숙, “크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토”, 「강원법학」, 제61권(2020), 231면.

는 그 소재의 갱신·검증 또는 보충에 인적 또는 물적으로 상당한 투자를 한 자여야 한다(저작권법 제2조 제20호). 법률상 권리의 주체가 될 수 있는 인(人)이면 자연인이든 법인이든 누구나 될 수 있다. 이하 부정경쟁법과 같이 ‘업으로’라는 제한이 없으므로 개인이어도 된다.

한편, 데이터베이스제작자의 권리를 보장하게 된 배경은 “창작성이 없는 데이터베이스에 대해 배타적 권리를 부여하는 것이 비록 사회 전체의 정보공유를 저해하여 정보화 사회에 역행한다는 부정적 측면을 제기할 수 있으나, 제작 및 갱신 등에 상당한 투자가 있을 경우 그 투자 노력에 대해서도 법적으로 보호할 가치가 있고, 이는 데이터베이스 제작을 활성화시킨다는 점에서 궁극적으로 지식정보사회의 요구에 부합하는 조치라는 필요성에 대한 인식”이 있었다고 한다.²⁸⁾ 현대처럼 자동화된 정보처리 기술을 통한 데이터베이스 제작이 손쉽게 이루어질 수 없었고, 비전자적 정보를 수집하여 온라인 또는 오프라인상에서 체계화 시키는 작업에 많은 노력이 필요했던 시기에 해당 노력을 인정함으로써 데이터베이스 제작을 확대 시키고자 했던 권리이다. 그런데 근래에 와서는 저작권법상 정의된 데이터베이스 제작은 적어도 전자화된 방식에 있어서는 매우 쉬워졌음을 고려할 필요가 있다.

또한 정의 규정을 법문 그대로 적용할 경우, 데이터베이스 제작의 활성화라는 목적에 부합하지 않은 많은 유형의 비즈니스를 데이터베이스제작자로 보호할 여지도 있다. 데이터 산업의 비즈니스 유형²⁹⁾에는 ① 데이터 처리 및 관리 솔루션 개발·공급업³⁰⁾ ② 데이터 구축 및 컨설팅 서비스업³¹⁾ ③ 데이터 판매 및 제공서비스업³²⁾ 가 있으며 본래 목적에 따른다면 ② 데이터 구축 유형에 국한되어 인정되어야 마땅한데, 데이터베이스제작자의 정의가 제작 또는 그 소재의 갱신·검증 또는 보충에 투자한 자이므로, 이는 데이터 산업의 비즈니스 유형인 중에서 ② 데이터 구축 뿐만 아니라 데이터베이스를 가지고 ③ 데이터 판매 및 제공서비스업의 유형까지 포함할 수 있게 된다. 일명 엔하위키사건에서도 심지어 사이트에 집적된 20만 건 이상에 이르는 게시물 대부분은 이용자가 작성하거나 이를 수정하여 온 것이었음에도 데이터베이스제작자로 판단하고 있음을 알 수 있다.³³⁾

2.2.4.4. 권리침해 가능성 판단 기준

2.2.4.4.1. 상당한 부분을 복제한 경우

데이터베이스제작자의 권리 침해는 데이터베이스제작자의 허락 없이 데이터베이스의 전부 또는 상당한 부분의 복제 등이 일어날 경우 발생한다(저작권법 제136조 제2항 제3호). 상당한 부분의 복제 등에 해당하는지를 판단할 때는 양적 상당성과 질적 상당성을 모두 고려하여야 한다. 양적으로 상당한 부분인지 여부는 복제 등이 된 부분을 전체 데이터베이스의 규모와 비교해 판단해야 하며, 질적으로 상당한 부분인지 여부는 복제 등이 된 부분에 포함되어 있는 개별 소재

28) 문화관광위원회, “저작권법중개정법률안 심사보고서”, 국회, 2003, 7면(한지영, “데이터베이스의 법적 보호에 관한 연구”, 서울대학교 대학원, 박사, 2005, 23면에서 재인용).

29) 과학기술정보통신부·한국데이터산업진흥원, “2023 데이터산업 현황조사”, 과학기술정보통신부·한국데이터산업진흥원, 2023, 12면.

30) DBMS, DBMS관리, 데이터 모델링, 분석·시각화, 검색엔진, 품질 등 관련 솔루션 제품을 판매하는 등의 비즈니스를 의미하며, 주로 라이선스, 개발·커스터마이징, 유지보수를 통해 매출이 발생한다.

31) DB설계, 데이터 이행 등을 포함한 DB시스템 구축, 문서·음성·영상 등의 데이터를 DB로 변환, 정비, 가공, DW, Data Lake 구축 등의 데이터·DB구축·가공하거나 데이터 관련 컨설팅 비즈니스를 의미하며, 데이터 구축 및 컨설팅은 구축·개발, 유지보수·운영관리, 컨설팅을 통해 매출이 발생한다.

32) 데이터·DB를 원천데이터 형태나 분석 및 활용이 가능한 상태로 판매하거나, 중개, 데이터신디케이션 및 큐레이션 등을 통한 수요 맞춤형 데이터·DB판매, 데이터를 가공·활용·분석한 주제분야별 정보서비스, 분석 결과 정보 등을 온·오프라인으로 제공하는 비즈니스를 의미하며, 데이터 이용료·수수료 등의 직접매출과 광고료 등의 간접매출로 수익이 발생한다.

33) 대법원 2017. 4. 13. 선고 2017다204315 판결.

자체의 가치나 그 개별 소재의 생산에 들어간 투자가 아니라 데이터베이스제작자가 그 복제 등이 된 부분의 제작 또는 그 소재의 갱신·검증 또는 보충에 인적 또는 물적으로 상당한 투자를 했는지를 기준으로 제반사정에 비추어 판단된다³⁴⁾ 즉 전체 데이터베이스에서 복제된 부분의 규모의 상당성과 데이터베이스 제작 등에 한 투자의 상당성으로 판단된다.

2.2.4.4.2. (상당한 부분에 이르지 않은 경우라도) 반복적·체계적 복제한 경우

상당한 부분에 이르지 못하는 부분의 복제등이라 하더라도 반복적이거나 특정한 목적을 위하여 체계적으로 함으로써 해당 데이터베이스의 일반적인 이용과 충돌하거나 데이터베이스제작자의 이익을 부당하게 해치는 경우도 상당한 부분의 복제등과 같이 침해로 본다(저작권법 제93조 제2항). ‘반복적·체계적’이라는 것은 우리와 유사한 규정을 둔 EU 지침³⁵⁾의 의미를 염두할 때, 어떤 계획에 따라 반복적으로 이루어지는 것을 의미하며, 임의적이거나 우발적인 이용행위를 배제하는 것이라 할 수 있다. 우리의 학설은 종래 ‘통상적인 이용과 충돌’한다는 것은 자신의 데이터베이스의 개별 소재를 상당한 정도 복제 등의 방법으로 추출·제작하여 내놓은 상품이 시장에서 자신의 데이터베이스와 경쟁하는 상황에 놓인다거나 아니면 자신의 현재적이거나 잠재적 시장에 영향을 미칠 정도로 복제 등이 되는 경우를 의미하고, ‘이익을 부당하게 해치는 경우’는 그보다 정도는 약한 것으로서 제작자가 통상적으로 이용허락을 해서 얻을 수 있는 이익을 부당하게 상실하는 것을 의미한다고 보고 있다.³⁶⁾

2.2.5. 부정경쟁방지법상 데이터부정사용행위 및 성과모용행위

2022년 제정된 「데이터 산업진흥 및 이용촉진에 관한 기본법」(일명 ‘데이터 기본법’)은 데이터생산자가 인적 또는 물적으로 상당한 투자와 노력으로 생성한 경제적 가치를 가지는 데이터(‘데이터자산’)는 보호되어야 한다고 선언한 후(데이터 기본법 제12조 제1항), 누구든지 이러한 데이터자산을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 무단 취득·사용·공개하거나 이를 타인에게 제공하는 행위, 정당한 권한 없이 데이터자산에 적용한 기술적 보호조치를 회피·제거 또는 변경하는 행위 등 데이터자산을 부정하게 사용하여 데이터생산자의 경제적 이익을 침해하여서는 아니 된다고 규정하면서(데이터 기본법 제12조 제2항), 구체적인 데이터자산의 부정사용 등 행위에 관한 사항은 부정경쟁방지법에서 정한 바에 따른다고 규정하였다(데이터 기본법 제12조 제3항).³⁷⁾

사실 데이터와 관련하여서 종전에는 저작권 침해를 주위적으로 일반 부정경쟁행위인 성과모용행위규정(부정경쟁방지법 제2조 제1호 파목)이 보충적으로 주장되어 왔으나,³⁸⁾ 2022년 시행된 데이터부정사용행위 규정(부정경쟁방지법 제2조 제1호 카목)으로 데이터 부정사용에 대하여는 카목의 적용을 받을 수 있게 되었다. 이에 이하에서는 두 가지 행위에 대하여 살펴본다.³⁹⁾

34) 대법원 2022. 5. 12. 선고 2021도1533 판결; 양혜윤, 「[지식재산권] AI 데이터베이스 크롤링과 관련된 법적 쟁점과 판례 해설」, 법률신문, 2024. 3. 26자.

35) EU 지침 제7조 제5항.

36) 유럽은 우리나라보다 더 엄격하게 이를 해석하고 있는데 EU 지침은 ‘통상적인 이용과 충돌’ 또는 ‘이익을 부당하게 해치는 경우’는 데이터베이스제작자의 투자에 상당한 손해를 일으키는 경우로서, 상당하지 아니한 부분을 반복적이고 체계적으로 추출하는 등으로 결국 상당한 부분을 추출한 것과 동일한 결과를 가져오는 것을 의미하는 것으로 보고 있다(정현순, “저작권법상 데이터베이스제작자 권리 침해에 관한 연구 - 대법원 2022. 5. 12. 선고 2021도1533 판결을 중심으로 -”, 「계간 저작권」, 제35권 제3호(2022), 167-168면).

37) 이광욱 외 2인, 「크롤링, 최근 개정된 부정경쟁방지법이 바라보는 시각」, 법률신문, 2022. 4. 28자.

38) 잡코리아-사람인 사건(서울중앙지방법원 2016. 2. 17 선고 2015가합517982 판결)을 예로 볼 수 있다.

39) 현행 부정경쟁방지법 제2조 제1호 타목의 일반 부정경쟁행위는 다른 부정경쟁행위에 대한 보충적 일반행위

2.2.5.1. 카목의 데이터부정사용행위

2.2.5.1.1. 의의

데이터부정사용행위 규정은 공개 데이터 중에서도 비정형데이터를 보호할 수 없는 법적 공백을 해결하기 위하여 제시된 규정이다.⁴⁰⁾ 즉 비공개 데이터는 부정경쟁방지법상 영업비밀로 보호되고, 공개 데이터 중 정형데이터는 저작권법에 따른 편집저작물 또는 데이터베이스로 보호될 수 있으나 비정형 데이터는 명시적 보호규정이 없는 상황에서 이에 대한 보호의 필요성이 대두되며 규정된 것이다.

규정상 데이터를 부정하게 사용 행위는 ① 접근권한 없는 자의 부정취득 등의 행위 ② 접근권한 있는 자가 부정한 목적으로 데이터 사용 등을 하는 행위 ③ ①또는②임을 알면서 해당 데이터를 사용 등 하는 행위 ④ 데이터 보호를 위한 기술적 보호조치를 무력화하는 기술 등을 제공 등 하는 행위로서(부정경쟁방지법 제2조 제1호 카목) 직접적인 데이터 부정취득 뿐만아니라 이를 가능하게 하는 기술적 보호조치 무력화 기술을 제공까지도 폭넓게 제재하고 있음을 알 수 있다.

2.2.5.1.2. 데이터부정사용행위 규정의 적용가능성

데이터부정사용행위라는 부정경쟁행위에 의해 보호되는 데이터의 범위는 데이터기본법상 데이터의 범위보다 좁다.⁴¹⁾ 데이터부정사용행위 규정의 보호객체인 데이터는 「데이터기본법」 제2조 제1호에 따른 데이터 중에서, 업(業)으로서 특정인 또는 특정 다수에게 제공되는 것으로, 전자적 방법으로 상당량 축적·관리되는 (영업비밀을 제외한) 기술상 또는 영업상의 정보라고 규정하고 있다(부정경쟁방지법 제2조 제1호 카목). 이에 따르면 ① 전자화된 정보 즉 데이터 중에서 ② 업(業)으로 ③ 특정인 또는 특정 다수에게 제공되고, ④ 전자적 방법으로 ⑤ 상당량 축적·관리되는 ⑥ 기술상 또는 영업상의 정보로서 영업비밀이 아니어야 한다.

구체적으로 살펴보면, ② 업(業)으로 제공된다는 점에서 사회 통념상 사업의 수행·일환으로서 행하여지는 것이기에 데이터 부정사용행위에 대하여 주장할 수 있는 주체가 이에 한정될 수 있음을 알 수 있다. ③ ‘특정인 또는 특정 다수’라는 것은 상대방을 특정할 수 있는 데이터 거래로 생각된다. 그렇다면 일반적으로 불특정 다수에게 제공될 것이라 여기는 공개 데이터는 여기에 해당되지 않을 것이다. 물론 특정인 또는 특정 다수에게 제공된다는 것은 일정한 조건 하에서 데이터를 제공받는 자가 이용할 수 있는 상태에 둔다는 의미에서 공개 데이터 또한 포함될 여지가 있을 수 있다. 하지만 ④ 전자적 방법으로 관리된다는 요건은 데이터 보유자와 보유자로부터 데이터를 제공받는 자 외의 사람이 데이터에 접근할 수 없도록 아이디·패스워드를 통해 접

로 보는 것이 일반적이고, 하급심 판결이기는 하나 법원도 “특별한 사정이 없는 이상 (가) 내지 (자)목에서 정하고 있는 행위유형에는 해당하나 위 각 목에서 정하고 있는 부정경쟁행위로 인정되기 위한 요건을 갖추지 못한 행위에 대하여는 (차)목에 의한 부정경쟁행위(주: 현행 (타)목의 일반 부정경쟁행위)로 함부로 의율하여서는 아니 된다”고 판시하고 있으므로(서울중앙지방법원 2014. 8. 28. 선고 2013가합552431 판결, 확정), 전형적인 부정경쟁행위인 데이터 부정사용행위에 의해 보호되지 않는 불특정 다수에게 공개된 정보 내지는 데이터를 일반 부정경쟁행위에 의하여 보호받기는 쉽지 않을 것으로 보인다(이광욱 외 2인, 「크롤링, 최근 개정된 부정경쟁방지법이 바라보는 시각」, 법률신문, 2022. 4. 28자) 타목의 일반 부정경쟁행위는 열거된 행위는 아니지만 부정경쟁으로 볼 수 있는 행위를 적용하기 위한 보충적 규정이라는 점에서 여전히 적용가능해야 할 것이라 생각한다.

40) 산업통상자원중소벤처기업위원회, “부정경쟁방지 및 영업비밀보호에 관한 법률 일부개정법률안 검토보고(김경만의원 대표발의, 의안번호 제2107535호) 검토보고서”, 국회, 2021, 3면.

41) 데이터기본법 제2조 제1호에 따르면 “데이터”란 다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및「소프트웨어 진흥법」 제2조 제1호에 따른 소프트웨어 등을 통하여 생성된 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보를 말한다 하고 있으며 이는 전자화된 데이터를 모두 포괄하는 개념이다.

근 권한을 관리하는 등 전자적 조치를 취하는 것이기 때문에 전자적 조치가 취해 있지 않고 스크래핑 등의 기술적용이 가능한 공개 데이터에 대하여는 본 규정이 적용될 여지가 없을 것이다.⁴²⁾

2.2.5.2. 파목의 성과모용행위 규정의 적용가능성

2.2.5.2.1. 의의

이상에서 살펴본 바와 같이 데이터부정사용행위로는 공개데이터에 대한 침해를 주장할 여지가 없다. 그렇다면 공개 데이터의 수집에 대하여 파목의 성과모용행위 적용가능성에 대하여 살펴볼 필요가 있다.

부정경쟁방지법 제2조 제1호 파목은 그 밖에 타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위를 규제한다. 이에 해당하려면 ① 타인의 상당한 투자나 노력으로 만들어진 성과 등이어야 하고 ② 공정한 상거래 관행이나 경쟁질서에 반하는 방법을 사용해야 하며 자신의 ③ 영업을 위하여 ④ 무단으로 사용하고 이로써 ⑤ 타인의 경제적 이익을 침해해야 한다.

2.2.5.2.2. 적용 기준

구체적으로 파목이 보호하는 객체는 ① 상당한 투자나 노력의 성과 등이다. 보호대상인 ‘성과 등’의 유형에 제한을 두고 있지 않으므로, 유형물뿐만 아니라 무형물도 이에 포함되고, 종래 지식재산권법에 따라 보호받기 어려웠던 새로운 형태의 결과물도 포함될 수 있다.⁴³⁾ 이에는 투자나 노력의 상당성이 충족된다면 데이터도 이에 해당될 수 있다. ‘상당한 투자나 노력으로 만들어진 것’인지는 권리자가 투입한 투자나 노력의 내용과 정도를 그 성과 등이 속한 산업분야의 관행이나 실태에 비추어 구체적·개별적으로 판단하되, 성과 등을 무단으로 사용함으로써 침해된 경제적 이익이 누구나 자유롭게 이용할 수 있는 이른바 공공영역(public domain)에 속하지 않는다고 평가할 수 있어야 한다.⁴⁴⁾

성과 등은 ③ 영업을 위하여 사용되어야 하므로 데이터를 보유하여 침해를 주장하는 자는 누구나 가능하지만 성과모용의 침해행위를 하는 자는 적어도 영업 즉 영리목적의 활동을 하는 자여야 한다.

성과 등을 ② 관행이나 경쟁질서에 반하는 방법으로 ④ 무단으로 사용해야 한다. 무단이라는 의미는 권리자의 명시적·묵시적 동의없이 또는 법률적 정당성 없이 사용하는 것이다. 조문에 따르면 관행이나 경쟁질서에 반하는 방법을 사용하는 것과 무단으로 사용하는 것 모두를 충족시

42) 성립요건의 설명은 특허청·지식재산보호원, “부정경쟁방지법상 데이터 보호 제도 안내서”, 특허청 산업재산 보호정책과, 2024, 12-13면 참조.

43) 판례로서는 2020년 3월 26일의 이른바 ‘BTS’ 사건 대법원 결정(대법원 2020. 3. 26.자 2019마6525 결정)과 이른바 ‘골프존’ 사건 대법원 판결(대법원 2020. 3. 26. 선고 2016다276467 판결), 에르메스 눈알가방 형태 도용 사건(대법원 2020. 7. 9. 선고 2017다217847 판결), BBQ 영상 광고 콘티 사건(대법원 2020. 7. 23. 선고 2020다220607 판결), 구동장치 도면파일의 무단 이용(전송) 사례(대법원 2022. 4. 28. 선고 2021다310873 판결), ‘차량의 루프박스 구조’ 도용 사건(대법원 2020. 6. 25. 선고 2019다282449 판결) 등이 있다(법원 지적재산권법연구회·사단법인 한국특허법학회, “크롤링 프로그램을 이용하여 숙박업소 정보를 복제 및 영업상 이용한 행위가 성과 도용 부정경쟁행위에 해당 여부”, 차상욱(편), 부정경쟁방지법 판례백선, 박영사, 2024, 307-310면).

44) 법원 지적재산권법연구회·사단법인 한국특허법학회, “크롤링 프로그램을 이용하여 숙박업소 정보를 복제 및 영업상 이용한 행위가 성과 도용 부정경쟁행위에 해당 여부”, 차상욱(편), 부정경쟁방지법 판례백선, 박영사, 2024, 309면.

켜야 한다. 타인의 상당한 투자나 노력으로 만들어진 성과를 ‘허락 없이’ 사용했다고 하더라도, 그 방법이 공정한 상거래 관행이나 경쟁질서에 반하지 않는다면, 바꾸어 말해 ‘행위 태양’이 위법하지 않다면 부정경쟁방지법 제2조 제1호 (파)목[현행]의 부정경쟁행위에 해당하지 않는다.⁴⁵⁾ 대법원은 (파)목[현행]의 부정경쟁행위가 ‘공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용’한 경우에 해당하는지를 판단하는 기준으로 ‘권리자와 침해자가 경쟁 관계에 있거나 가까운 장래에 경쟁관계에 놓일 가능성이 있는지, 권리자가 주장하는 성과 등이 포함된 산업분야의 상거래 관행이나 경쟁질서의 내용과 그 내용이 공정한지, 위와 같은 성과 등이 침해자의 상품이나 서비스에 의해 시장에서 대체될 수 있는지, 수요자나 거래자들에게 성과 등이 어느 정도 알려졌는지, 수요자나 거래자들의 혼동가능성이 있는지’ 등을 제시하고 있다.⁴⁶⁾

마지막으로 이를 통해 타인의 경제적 이익을 침해해야 한다.

2.2.6. 독점규제법상 시장지배적 지위 남용 및 불공정거래행위 규정의 적용가능성

독점규제법은 시장지배적지위의 남용과 불공정거래행위를 금지한다. 사업자가 공개 데이터를 수집하는 행위를 하거나 수집을 제한하는 행위를 할 경우 독점규제법상 시장지배적 지위 남용 또는 불공정거래행위가 될 여지가 있다.⁴⁷⁾

2.2.6.1. 독점규제법상 금지되는 행위

2.2.6.1.1. 시장지배적지위의 남용금지

독점규제법에 따르면 시장지배적사업자는 다른 사업자의 사업활동을 부당하게 방해하는 행위를 하거나 새로운 경쟁사업자의 참가를 부당하게 방해하는 행위 등에 해당하는 경우 남용행위가 된다.⁴⁸⁾ 구체적으로 다른 사업자나 새로운 경쟁사업자의 상품 또는 용역의 생산·공급·판매에 필수적인 요소의 사용 또는 접근을 거절·중단하거나 제한하는 행위는 남용행위가 된다(독점규제법 시행령 제9조 제3항 및 제4항).

2.2.6.1.2. 불공정거래행위

독점규제법 제45조에 따르면 부당하게 거래의 상대방을 차별하여 취급하는 행위, 부당하게 경쟁자를 배제하는 행위, 부당하게 다른 사업자의 사업활동을 방해하는 행위 등을 불공정거래행위로 보고 있다. 차별취급, 경쟁자 배제 뿐만 아니라 사업활동 방해행위까지 불공정거래행위가 될 수 있다. 다만 사업활동 방해행위는 기술의 부당이용 내지 거래처 이전방해와 동일시 할 수 있을 정도의 행위로서 자기의 능률이나 효율성과 무관하게 다른 사업자의 사업활동을 심히

45) 법원 지식재산권법연구회·사단법인 한국특허법학회, “웹 크롤링과 성과 무단 사용”, 권보원(편), 부정경쟁방지법 판례백선, 박영사, 2024, 334-335면.

46) 대법원 2020. 3. 26. 선고 2016다276467 판결, 대법원 2020. 7. 9. 선고 2017다217847 판결 등 참조.

47) 독점규제법은 공정하고 자유로운 경쟁을 촉진하기 위하여 사업자의 시장지배적지위의 남용과 과도한 경쟁력의 집중을 방지하고, 부당한 공동행위 및 불공정거래행위를 규제하고 있다.

48) 제5조(시장지배적지위의 남용금지) ① 시장지배적사업자는 다음 각 호의 어느 하나에 해당하는 행위(이하 “남용행위”라 한다)를 해서는 아니 된다.

1. 상품의 가격이나 용역의 대가(이하 “가격”이라 한다)를 부당하게 결정·유지 또는 변경하는 행위
 2. 상품의 판매 또는 용역의 제공을 부당하게 조절하는 행위
 3. 다른 사업자의 사업활동을 부당하게 방해하는 행위
 4. 새로운 경쟁사업자의 참가를 부당하게 방해하는 행위
 5. 부당하게 경쟁사업자를 배제하기 위하여 거래하거나 소비자의 이익을 현저히 해칠 우려가 있는 행위
- ② 남용행위의 유형 및 기준은 대통령령으로 정한다.

곤란하게 할 정도로 방해하는 행위를 의미한다.⁴⁹⁾

2.2.6.2. 적용 가능성

2.2.6.2.1. 공개 데이터를 수집하는 사업자의 경우

공개 데이터를 수집하는 사업자가 시장지배적지위를 가지게 된다고 볼 여지는 없을 것이다. 물론 넓은 범위의 콘텐츠를 수집하는 크롤링의 경우 대상 서버에 심한 부하를 주는 것을 지양하여야 한다는 것이 업계의 관행인 바, 이를 지키지 않아서 운영자의 웹 사용용량이 넘쳐서 비용이 발생하거나, 혹은 웹사이트의 호스트 서버에 과부하가 걸려 웹 속도가 느려지는 경우에는 불공정거래행위로서의 사업활동 방해행위가 해당할 여지는 있을 것이다.⁵⁰⁾ 그러나 일반적으로 웹사이트의 소유자가 허용한 범위 내라면 자동화된 기술을 통한 공개 데이터의 수집 그 자체로는 불공정거래행위가 될 소지가 크지 않다. 이는 제공자가 정의한 이용약관, 로봇배제표준(Robots.txt)을 따르고 저작권과 개인정보보호 규정을 준수하는 것을 전제로 하는 것이다. 결국 법과 관행을 준수하는 한 공개 데이터 수집은 불공정거래행위에 해당할 가능성은 크지 않다.

2.2.6.2.2. 공개 데이터를 보유하고 있는 사업자의 경우

반면에 유일한 데이터를 보유한 사업자가 시장지배적지위를 가진다고 볼 여지가 있다. 해당 데이터가 유일한 자료이거나 혹은 유일하지 않지만 다른 방식의 수집이 현저히 곤란하여 사실상 수집할 수 있는 유일한 자료와 다를 바 없는 데이터라면 이러한 데이터를 보유하고 있는 사업자는 시장지배적 사업자에 해당될 여지가 있다. 이러한 사업자가 공개 데이터에 대하여 이용약관, 로봇배제표준 등에 공개 데이터의 수집을 금지한다면 남용행위에 해당될 것이다.

한편, 공개 데이터 제공에 있어 특정 사업자를 차별취급하거나, 경쟁자를 배제하는 것, 사업에 필수적인 정보를 제공하지 않아 사업활동을 곤란하게 한 경우에는 불공정거래행위도 성립할 수 있을 것이다.⁵¹⁾

2.2.7 소결

이상에서 살펴본 바와 같이, 일반적으로 API, 크롤링, 스크래핑, 미러링 기술을 통한 공개 데이터 수집은 형법상 컴퓨터등장애업무방해죄와 정보통신망법상 정보통신망침해죄가 성립할 여지가 없다. 수집이전에 동의 또는 허락을 받는 것이 원칙인 개인정보와 저작물의 경우는 방대한 양의 공개 데이터에 대한 자동화된 수집행위와는 별도로 취급되어야 하고 각각 개인정보보호법과 저작권법을 따라야 할 것이다. 부정경쟁방지법상 데이터부정사용행위에 있어 데이터는 특정인 또는 특정 다수에게 제공된다는 점에서 불특정 다수에게 제공될 것이라 보이는 공개 데이터는 해당될 여지가 없으며 전자적 방법으로 관리되지 않는 점에서도 공개 데이터 침해를 주장할 수 있는 규정이 될 수 없다.

반면, 저작권법상 데이터베이스제작자의 권리를 침해하거나, 부정경쟁방지법상 성과모용행위에 해당할 여지가 있으며, 공개 데이터를 수집할 경우 현재 두 규정에 따른 적법성 확보가 요

49) 황태희, “웹 크롤링의 경쟁법적 고찰”, 『경쟁법연구』, 제43권(2021), 186-194면; 불공정거래행위 심사지침 [공정거래위원회예규 제477호, 2024. 12. 30., 일부개정] V. 8 참조

50) 특히 크롤링의 목적·범위 등에 비추어 보면 크롤링은 검색 등에서 검색의 질을 높이기 위해 사용되며, 그러한 검색결과를 통하여 고객들이 대상사이트 운영자의 사이트를 방문하여 다른 서비스를 이용할 수 있다는 이익이 있기 때문에 사업활동을 방해하려는 의도를 인정하기는 쉽지 않다(황태희, “웹 크롤링의 경쟁법적 고찰”, 『경쟁법연구』, 제43권(2021), 195면).

51) 구체적인 내용은 황태희, “웹 크롤링의 경쟁법적 고찰”, 『경쟁법연구』, 제43권(2021), 183-184면.

청된다고 할 수 있다.

저작권법상 보호되는 객체는 데이터베이스이지만 사실상 데이터를 의미하는 것으로 해석될 여지가 있다. 데이터베이스제작자는 데이터베이스 제작등에 인적 또는 물적으로 상당한 투자한 자인데, 법률상 권리의 주체가 될 수 있는 인(人)이면 누구나 될 수 있으며 정의 규정을 법문 그대로 적용할 경우, 데이터베이스 제작의 활성화라는 목적에 부합하지 않은 많은 유형의 비즈니스를 데이터베이스제작자로 포함될 여지가 있기 때문이다. 또한 부정경쟁방지법상 성과모용 행위 규제는 상당한 투자나 노력의 성과 등을 보호하는 바, 해당 공개 데이터가 상당한 투자나 노력의 성과로 인정될 여지가 있다.

한편, 공개 데이터를 수집하는 사업자의 경우에는 독점규제법의 규제를 받을 가능성이 크지 않고 오히려 공개 데이터를 보유하고 있는 사업자의 경우에 있어 해당 데이터가 유일한 자료이거나 혹은 유일하지 않지만 다른 방식의 수집이 현저히 곤란하여 사실상 수집할 수 있는 유일한 자료와 다를 바 없는 데이터에 대한 수집을 금지하는 행위를 시장지배적 지위 남용행위 또는 불공정거래행위로 규제할 여지가 있다는 점에서 적용가능성이 존재한다.

3. 공개 데이터 수집 기준 정립의 방향

3.1. 현행 기준에 대한 평가

3.1.1. 공개 데이터 수집과 관련한 현행 기준

공개 데이터의 수집시 저촉 가능성이 있는 두 규정 즉, 저작권법상 데이터베이스제작자의 권리 침해 기준과 부정경쟁방지법상 성과모용행위 성립 요건을 살펴보면, 우선 데이터베이스제작자의 권리 침해는 데이터베이스제작자의 허락 없이 데이터베이스의 전부 또는 상당한 부분의 복제 등이 일어날 경우 발생한다. 여기서 상당성은 전체 데이터베이스에서 복제된 부분의 규모의 상당성과 데이터베이스 제작 등에 한 투자의 상당성으로 판단된다. 상당한 부분에 이르지 않은 경우라도 반복적·체계적 복제한 경우 침해가 될 수 있다. 이는 임의적이거나 우발적인 이용행위를 배제하고 어떤 계획에 따라 반복적으로 행해져 제작자가 이용허락을 해 얻을 수 있는 이익을 상실하는 경우를 의미한다.

다음으로 성과모용행위의 성립 요건은 ① 상당한 투자나 노력의 성과 등 ② 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 ③ 자신의 영업을 위하여 ④ 무단으로 사용함으로써 ⑤ 타인의 경제적 이익을 침해하는 것이다. 성과 등을 ③ 영업을 위하여 사용되어야 하므로 데이터를 보유하고 침해를 주장하는 자는 누구나 가능하지만 성과모용의 침해행위를 하는 자는 적어도 영업 즉 영리목적의 활동을 하는 자여야 한다. 또한 침해행위는 성과 등을 ② 관행이나 경쟁질서에 반하는 방법을 사용하는 것과 ④ 무단으로 사용하는 것 모두를 충족해야 하며 이를 통해 ⑤ 타인의 경제적 이익이 침해되어야 한다.

이상의 두 규정을 비교해 공개 데이터 보호 또는 침해의 기준을 비교해 보면 다음과 같은 공통된 요건을 발견할 수 있다. 첫째, 상당한 투자이다. 저작권법은 인적 또는 물적으로 상당한 투자를 부정경쟁방지법은 상당한 투자나 노력을 요구하고 있다. 둘째, 허락을 요건으로 한다. 저작권법은 데이터베이스제작자의 허락이 없으면 침해가 되고 부정경쟁방지법은 무단으로 사용하는 것이 침해가 된다. 셋째, 두 법 모두 공정한 관행을 요건으로 하고 있다. 저작권법은 제28조에 따라 데이터베이스제작자의 권리가 있는 데이터베이스라고 하더라도 보도·비평·교육·연구 등을 위하여는 정당한 범위 안에서 공정한 관행에 합치되게 이를 인용할 수 있고 부정경쟁방지법은 타인의 성과를 허락없이 사용했다 하더라도 그 방법이 공정한 상거래 관행이나 경쟁질서

에 반하지 않는다면 부정경쟁행위가 아니다. 이 밖에도 부정경쟁방지법은 성과 등을 요건으로 하고 있고 저작권법은 복제한 데이터 양의 상당성을 요하고 있는 것으로 파악된다. 이와 같은 기준들에 대하여 다음과 같은 점에서 현행 규정에 대한 재고찰이 필요함을 생각해 볼 수 있다.

3.1.2. '투자'의 상당성 기준의 적절성 여부

첫 번째, '투자'의 상당성에 대한 기준이다. 저작권법과 부정경쟁방지법은 상당한 투자를 요건으로 하고 있고 부정경쟁방지법은 성과 등을 만들어져야 하지만 데이터베이스제작자는 창작성이라든지 신가치창출을 권리보호의 요건으로 보지 않는다.

데이터가 공유물이라는 점에서 그리고 많은 데이터를 축적한 서비스에 락인(lock-in)되어 독점화가 일어나기 쉬운 성질을 가졌다는 점에서 데이터에 대한 권리를 인정하는 것은 보다 엄격한 기준이 적용될 필요가 있다.⁵²⁾ 게다가 데이터와 데이터베이스 간의 차이가 크게 없어져 버린⁵³⁾ 인터넷상에서의 데이터에 대하여 판단할 때는 더더욱 그러하다. 또한 투입된 자본이 많은 것이 곧 시장에서 교환가치나 사용가치 있는 결과물을 내는 것을 담보하지 않는다는 시장원리에도 배치된 기준이다. 이는 공유물이었던 데이터를 사유화하는 기준이 되기에 부적절한 기준이다. 상당한 투자라는 기준은 손해배상의 기초가 될 수 있을지언정, 데이터에 대한 권리 부여에 정당화 기준으로는 미흡한 것으로 생각된다.

3.1.3. '상당성' 기준의 적절성 여부

두 번째, '상당성'에 관한 기준이다. 이에 대하여는 세 가지 점에서 부적절한 기준이라고 생각된다. 첫째, 침해여부가 행위시가 아닌 판결시에 결정되는 기준이라는 점이다. 상당한 투자가 있을 때 보호되며 상당한 부분을 복제했을 때, 더 나아가 반복적·체계적 복제한 경우 상당성에 이른다고 까지 하고 있다. 데이터에 대한 권리를 주장하는 자도 얼마만큼까지 투자했을 때 권리가 있다고 볼 수 있는지 불분명하고 수집하는 자는 얼마만큼을 복제해야 상당한 복제로써 침해가 구성되는지 당해 권리가 형성되고 수집행위가 일어날 때는 판단할 수가 없다. 즉, '상당한' 투자나 '상당한' 부분과 같이 결과적으로 법원의 판단을 요구하는 기준으로 실무상 적용이 어려우며 현재는 법원의 판례 역시 충분히 축적되지 못한 상황이기⁵⁴⁾ 공개 데이터를 수집하는 경우 판결이 있기 전까지는 불확실한 위법의 리스크를 언제나 가지게 된다.

둘째, 데이터베이스가 저작물보다 더 강력하게 보호 될 여지가 있는 불균형을 초래하는 기준이라는 점이다. 구글북스 사건(Google Books)⁵⁵⁾에서 저작물은 전체가 복제되더라도 공정이용으로 판결되는 경우가 존재할 수 있는데, 현행법상 데이터베이스제작자의 권리는 상당한 양 이상의 데이터가 복제되면 언제나 침해가 성립된다. 이처럼 상당성으로 침해를 판단하는 것은 예외적으로 보호하는 데이터베이스에 대한 권리가 저작물에 대한 권리보다 더 강력하게 보호를 받을 수 있게 한다는 불균형을 초래한다. 물론 구글북스에서도 저작재산권을 부정하는 것이 아닌 제한규정인 공정이용을 적용하였다는 점에서 데이터베이스제작자의 권리도 공정이용 규정

52) 정보의 특성으로 인해 플랫폼은 하나로 몰리기 때문에 독점적 지위가 형성된다. 정보이용자는 필요한 정보를 신속하고 정확하게 얻는 것이 중요한데, 보다 많은 정보가 처리될수록 정교한 결과물이 나올 수 있고, 이용자가 늘어나면 제공되는 정보도 늘어나 더 정교한 결과물이 나올 수 있으므로 해당 플랫폼에 이용자가 더 늘어나는 현상이 생기는 것이다(고수운, "빅데이터 기술환경에서 정보의 이용에 따른 침해판단방법론", 강원대학교 대학원, 박사, 2019, 134면).

53) 김현숙, "크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토", 「강원법학」, 제61권(2020), 231면.

54) 정현순, "저작권법상 데이터베이스제작자 권리 침해에 관한 연구 - 대법원 2022. 5. 12. 선고 2021도1533 판결을 중심으로 -", 「계간 저작권」, 제35권 제3호(2022), 160면.

55) The Authors Guild v. Google, Inc., 578 U.S. 941, 136 S. Ct. 1658, 194 L. Ed. 2d 800 (2016).

의 적용을 통해 불균형이 생긴다는 주장은 틀렸다고 말할 수도 있다. 그러나 만약 이러한 적용이 가능하여 데이터도 전체를 복제한다 할지라도 공정이용이 될 수 있다고 한다면, 복제의 상당성이라는 권리침해 기준이 사실상 무용지물이기에는 새로운 기준이 필요하다는 반증이기도 할 것이다. 제4차 산업혁명이라 불리는 빅데이터 시대에 데이터는 양적 가치보다는 질적 가치가 더 중요시 되고 있다. 데이터를 보호하는 기준은 데이터베이스 제작의 확대라는 유인목적 등에 비추어 단순히 데이터 양을 기준으로 침해를 판단하는 것은 적절치 않다.

셋째, 기술의 생래(生來)와 맞지 않다는 점이다. 상당성을 컴퓨터 등에 명령어로 입력할 수 있는가? 그렇지 않다. 컴퓨터 등 광(光) 또는 전자적 방식으로 데이터를 수집하는 기술들은 0과 1로 데이터를 처리하며 불확정 개념인 상당하지 않은 만큼의 데이터를 수집하도록 요청하는 것은 불가능하다. 개인정보보호법 제15조를 보면 개인정보의 수집시 동의뿐만 아니라 처리자의 정당한 이익 달성이라든지 명백한 정보주체의 이익 등 다양한 기준을 제시하고 있지만 결국 산업에서는 거의 대부분 동의를만 통해 적법성을 확보하고 있는 실정이다. 불법성을 제거하는 다양한 기준 중 동의 또는 부동의라는 확정개념을 가진 사전동의 기준만이 기술에 적용하기 적합한 기준이었기 때문이다. 같은 논리에서 자동화된 기술을 통해 공개 데이터를 수집하도록 함에 있어 상당성이라는 기준은 적용하기 곤란하다. 이는 수범자의 행위규범으로 적용되기에 부적절한 기준이다.

3.1.4. '허락' 기준의 적절성 여부

세 번째, '허락'을 요구하는 기준이다. 저작권법은 데이터베이스제작자의 허락이 없으면 침해가 되고 부정경쟁방지법은 무단으로 사용하는 것이 침해가 된다. 그런데 이 허락이라는 기준은 기술의 내재적 성질에 반하며 이로 인해 공개데이터에 대한 자동화된 수집은 언제나 위법하게 된다는 점에서 적절하지 않다.

데이터를 수집하는 인터넷은 처음부터 개방형 시스템으로 설계되었으며 특정 IP 주소를 차단하는 등 액세스를 차단하는 적극적인 조치를 취하지 않는 한 기본적으로 자유롭게 웹 사이트를 탐색할 수 있는 것이 본래 성질이다.⁵⁶⁾ 이에 더하여 '자동화'된 기술로 방대한 양의 데이터를 수집하는 상황에서 일일이 데이터마다 사전 동의를 받는 것을 물리적으로 불가능하다.⁵⁷⁾ 게다가 이제 기술은 자동화에서 인공지능과 같은 자율화 기술로까지 발전하고 있기에 사전에 허락을 요구하는 것은 적어도 데이터 수집에 대한 적법한 기준으로 전혀 맞지 않다.

3.1.5. '공정한 관행' 기준의 적절성 여부

네 번째, 공정한 관행에 대한 기준이다. 저작권법 제28조에 따라 공정한 관행에 합치되게 데이터베이스는 인용될 수 있다. 판례는 온라인 서비스 제공자가 크롤링을 통하여 수집한 뉴스를 검색할 수 있도록 하는 서비스에 대하여 "피고가 사용한 로봇프로그램은 일반인들이 자유롭게 열람할 수 있는 웹사이트에서만 텍스트, 사진 등을 수집하는 점" 등을 이유로 공정한 관행에 합

56) 1960년대와 1970년대에 인터넷 프로토콜(IP)이 처음 개발되었을 때, 개발자들은 모든 사람이 참여할 수 있는 "개방형 아키텍처" 네트워크로 설계했다; Monika Isia Jasiewicz, "Copyright Protection in an Opt-Out World: Implied License Doctrine and News Aggregators", *The Yale Law Journal*, Vol.122(2014), p. 838(고수운, "기술발전에 따른 저작권 보호 및 이용방식에 관한 소고", 『계간 저작권』, 제37권 제1호(2024), 14-15면 재인용).

57) 만약 가능하다고 해도 사전동의로 인해 투입되는 비용이 증가하기 때문에 빅데이터의 활용을 통해 얻을 경제적 이익은 이전보다 월등히 높아 빅데이터 활용에 의미가 있게 된다. 경제적 이익이 보장되지 않는다면 빅데이터의 활용이 시도되지 않을 것이다(고수운, "빅데이터의 개인정보 이용으로 인한 법익침해의 私法적 해석", 『강원법학』, 제52권(2017), 521면).

치되게 인용한 경우에 해당한다고 판결한 바 있다.⁵⁸⁾

부정경쟁방지법상 성과모용행위가 되려면 ① 타인의 상당한 투자나 노력으로 만들어진 성과 등이어야 하고 ② 공정한 상거래 관행이나 경쟁질서에 반하는 방법을 사용해야 하며 자신의 ③ 영업을 위하여 ④ 무단으로 사용하고 이로써 ⑤ 타인의 경제적 이익을 침해해야 한다. 이 모든 요건을 충족시켜야 부정경쟁행위가 되기 때문에 타인의 성과를 허락없이 사용했다 하더라도 그 방법이 공정한 상거래 관행이나 경쟁질서에 반하지 않는다면 부정경쟁행위가 아니다. 이러한 점을 생각해 볼 때, 관행을 준수한다는 기준이 공개 데이터의 수집에 있어 적법성 확보에 중요한 기준이 될 여지가 있다.

3.2. 공개 데이터 수집 기준 정립의 방향

3.2.1. 적법 기준의 변화의 필요성

이상에서 살펴본 바와 같이 공개 데이터를 수집함에 있어 적법성 판단의 기준들은 데이터에 대한 권리를 부여함에 있어 너무 관대한 기준이며 저작물보다 데이터베이스를 보호할 가능성이 있는 형평에 어긋난 기준이다. 더욱이 문제되는 것은 침해 판단을 행위시가 아닌 판결시에 결정된다는 점과 수집 기술의 생래(生來)상 적용되기에 불가능하여 불법의 위험을 동반할 가능성이 있을 뿐만 아니라 나아가 자동화된 데이터 수집은 언제나 불법이 될 수밖에 없다는 점이다.⁵⁹⁾ 이러한 기준의 변화가 필요하다고 할 것이며 특히 행위시 적법성을 판단에 작동될 수 있는 기준이 바람직할 것이다. 다시 말해서 저작권법은 데이터베이스의 상당한 복제 행위, 부정경쟁방지법은 성과모용행위가 있을 때 침해이며 그렇지 않으면 적법하다는 방식을 두고 있는데, 반대로 (허락 이외에도) 무엇이 적법한 이용인가를 고려하여 규정하는 것이 필요하다는 생각이다.

이러한 점에서 현행 기준 중 공정한 관행이라는 기준이 주목할 만하다고 생각되며 이하에서는 자동화된 기술을 통한 공개 데이터 수집의 관행에 초점을 맞추어 공개 데이터 수집시 공정성을 확보할 수 있는 방안에 대하여 생각해 본다.

3.2.2. 자동화된 기술을 통한 공개 데이터 수집의 공정한 관행

앞서 살펴본 대로, 공개 데이터 수집과 관련된 자동화 기술에는 API, 크롤링, 스크래핑, 미러링이 있다. API의 경우 소프트웨어 계약으로서 서비스 제공자가 정의한 명확한 약관과 규칙을 따라야 한다. 크롤링은 로봇배제표준(Robots.txt)과 서비스 이용약관에 명시된 규정을 따라야 하며, 대상 서버에 과도한 트래픽과 부하를 일으키지 않도록 해야 한다.⁶⁰⁾ 스크래핑의 경우 정밀하게 데이터를 추출한다는 점에서 이용약관 뿐만 아니라 저작권, 개인정보보호 규정을 준수해야 한다. 전체 데이터를 복제하는 미러링의 경우 역시 이용약관, 저작권, 개인정보보호 규정을 따라야 한다. 정리하자면, 자동화된 기술을 통해 공개 데이터 수집시 공정한 관행은 데이터

58) 서울중앙지법 2006. 7. 21. 선고 2004가합76058 판결.

59) 크롤링이 문제가 된다면 데이터 수집 업체는 불법적인 사업을 하는 것이나 마찬가지가 된다. 실제로 크롤링과 관련해 어느 정도 수준까지 합법 또는 불법인지 모호하여 기업들은 언제 범죄자가 될지 모르는 불안에 시달린다고 한다. 박현익, 「데이터 수집부터 겁나는 기업들… “불법인지 합법인지 모르겠다”」, Chosun Biz, 2020. 9. 23자. (김현숙, “크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토”, 「강원법학」, 제61권(2020), 242면 재인용).

60) 여기서 대상 서버에 과도한 트래픽과 부하라는 것은 자연 이용자 증가에 기인한 것일 수 있는 단순한 접속횟수가 증가에 대하여는 적용되지 않을 가능성이 높으며 서버 설계상의 동시접속자 수용량 등을 고려하여 판단될 것이다. 이와 유사한 내용은 형법상 컴퓨터등장애업무방해죄 관련 판결 서울중앙지방법원 2021. 1. 13. 선고 2020노611 판결의 정보처리에 장애 발생을 판단한 부분에서 확인할 수 있다.

보유자가 명시한 이용약관, 로봇배제표준을 따르는 것, 저작권과 개인정보보호 규정을 준수하는 것, 그리고 대상 서버에 과도한 트래픽 부하를 일으키지 않는 것으로 요약될 수 있다.

그렇다면 저작물과 개인정보에 대하여는 별론으로 하고 일반적으로 공개 데이터 수집에 있어 공정한 관행에 따른다는 것은 이용약관과 로봇배제표준을 준수하는 것으로써 이를 통해 공개 데이터 수집에 적법성을 확보할 여지가 있다.

3.2.3. 이용약관과 로봇배제표준의 법적 성격

이용약관과 로봇배제표준을 준수하는 것이 공정한 관행에 따른다는 요건에 따라 공개 데이터 수집에 적법성을 부여할 여지는 있지만, 구체적으로 이용약관과 로봇배제표준의 준수가 공정한 관행이라는 기준의 요건으로 작용하기 위해서는 이를 준수하지 않은 것이 법률상 왜 제재 받아야 하는지에 대한 법적 근거는 이용약관과 로봇배제표준의 법적 성격을 어떻게 볼 것인가를 통하여 결정될 것이다.

이용약관의 법적 성격은 계약의 한 유형인 약관으로써 데이터 보유자의 이용약관 표시는 청약의 의사표시로 이해될 수 있다. 그렇다면 약관에 따라 공개 데이터를 수집하는 행위는 청약에 따른 승낙의 의사표시로, 이러한 계약에 기초하여 미준수에 대한 법적 제재가 가능하다고 설명될 수 있다.

한편, 로봇배제표준의 법적 성격은 현재 불분명하다. 로봇배제표준은 웹 사이트에 로봇이 접근하는 것을 방지하기 위한 규약으로 1994년 6월에 처음 만들어졌다. 이는 권고안에 불과하기 때문에 접근 제한에 대한 설명을 robots.txt에 기술했다고 하더라도 지켜지지 않을 수 있다.⁶¹⁾ 하지만 법원은 데이터 수집을 거부하였는데도 기계적으로 수집한 것을 악의적인 행위로 판단하고 있다.⁶²⁾ 로봇배제표준의 준수 여부를 실질적으로 불법적인 데이터 수집의 판단 기준으로 적용시킨 것이다. 로봇배제표준의 법적 성격에 대하여는 약관계약이거나 혹은 공시방법 중 하나로 고려될 수 있을 것이다. 로봇배제표준에 대한 법적 근거를 약관계약으로 본다면 데이터 보유자의 의사표시에 따른 수집을, 공시방법으로 본다면 표시된 권리의 내용을 침해하지 않는 수집을 통하여 적법성을 확보할 수 있을 것이다.

3.2.3.1. 약관계약설

만약 데이터 보유자가 접근 금지 또는 제한표시를 한 것이 계약상 의미가 있다면 향후 데이터를 수집한 것에 대해 계약위반을 문제 삼을 수 있을 것이다.⁶³⁾ 한쪽 당사자가 미리 마련했다는 점에서 계약 중에서도 약관으로 볼 수도 있어 보인다.

약관규제법상 “약관”이란 그 명칭이나 형태 또는 범위에 상관없이 계약의 한쪽 당사자가 여러 명의 상대방과 계약을 체결하기 위하여 일정한 형식으로 미리 마련한 계약의 내용을 말한다(약관규제법 제2조). 약관에 해당하기 위해서는 ① 계약이어야 하고 ② 여러 명의 상대방과 체결할 목적이 있어야 하며 ③ 일정한 형식을 갖추고 ④ 일방당사자가 미리 마련해 두어야 한다. robots.txt에 기술한 내용의 성격이 계약이라면 나머지 요건은 충족된다.

그렇다면 계약의 의미가 중요한데 계약이 무엇인지에 대해 현행법은 명확한 정의 규정을 두고 있지 않으나, 학자들은 대개 “일정한 사법상의 법률효과의 발생을 목적으로 하여 서로 상대

61) 나철원·온병원, “최신 웹 크롤링 알고리즘 분석 및 선제적인 크롤링 기법 제안”, 「인터넷정보학회논문지」, 제20권 제3호(2019), 45면.

62) 서울중앙지방법원 2020. 2. 11. 선고 2019고단1777판결.

63) 김현숙, “크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토”, 「강원법학」, 제61권(2020), 243-244면.

하는 당사자의 의사표시의 합치”라고 계약을 설명하고 있다. 즉, 계약이 성립되기 위해서는 서로 상대하는 당사자 간에 일정한 법률효과를 발생시키고자 하는 의사표시인 청약과 이에 대한 승낙이라는 의사표시의 합치가 필요하다.⁶⁴⁾

물론 robots.txt 상에는 단순히 금지 범위에 대하여 전부, 전무 또는 상세 목록이 제공될 뿐이며 이것에 대한 상대방의 동의의 의사표시인 승낙을 요구하지 않는다. 그러나 민법 제532조의 의사실현에 의한 계약의 성립을 적용해 데이터 수집행위를 통해 승낙을 인정할 여지는 있다.⁶⁵⁾ 또한 이를 인정하면 약관규제법에 따른 불공정약관심사를 통해 빅테크 기업의 데이터 독점을 규제할 수 있게 된다는 이점도 있다.

한편, 이것이 부적절하다는 견해도 있다. robot.txt를 통한 금지표시가 매우 빈번하며 수집을 허용하는 페이지와 제한하는 페이지를 선별하여 금지표시를 하는 것이 아니라, 단순히 모든 크롤링을 금지한 경우도 많은데 만약 이러한 경우 데이터 수집은 원천적으로 불가능해 진다는 문제가 있다는 것이다.⁶⁶⁾ 하지만 검색엔진 크롤러가 웹페이지 수집을 통해 해당 사이트 정보나 링크를 제공하는 것을 통해 방문자가 유도된다. 이러한 이유로 크롤링 허용이 오히려 필수적일 수 있다. robot.txt를 통해 서비스제공자가 금지표시를 할 경우 이러한 점이 당연히 고려될 부분이기 때문에 여전히 robots.txt상의 기술은 유의미하다고 생각된다.

3.2.3.2. 공시방법설

한편, 데이터에 대한 권리를 준물권으로 인정하고 이에 기초하여 로봇배제표준을 자신의 권리를 공시하는 방법으로 볼 수도 있다. 데이터는 통상적인 재화와 달리 복제가 용이하며 복제를 해도 원본이 파괴되지 않는 성질을 가진다. 따라서 정보의 복제와 전달에 드는 비용을 제외하면 정보의 이용에 관한 사회적 한계비용은 제로이며 한번 생산된 정보는 무료로 공개되는 것이 사회적 효율을 극대화하는 방법이다. 그러나 이러한 데이터의 공공성만을 지나치게 강조하여 데이터의 무상공유를 전제한다면 데이터 생산자의 투자자본의 회수가 곤란해져 사적 주체에 의한 정보생산의 인센티브가 현저하게 줄어들 수밖에 없다. 자신의 권리와 이익이 법으로 인정되지 않으면 ‘무임승차(free ride)’를 조장하고 동일한 서비스를 제공하는 경쟁자 간에 서비스를 개선할 유인이 부족해지기 마련이다. 데이터의 공공성이 주는 사회적 효율성을 감안하더라도 아무런 제한 없는 데이터 공유는 시장에 그 이상의 새로운 가치는 제공하지 않으면서 혁신의 동기를 약화시킬 수 있다. 이는 궁극적으로 소비자의 후생을 감소시키고 소비자 이익 보호라는 경쟁법의 목적과 배치되는 결과를 초래한다.⁶⁷⁾ 이처럼 데이터에 대한 법적 권리 인정에 대한 당위성이 인정되고 있다.

현행법으로는 물권법정주의에 따라 데이터에 대한 물권을 인정할 수는 없다.⁶⁸⁾ 그러나 저작권을 준물권으로 인정하는 저작권법에서 데이터베이스에 대한 권리를 인정하고 있으며⁶⁹⁾ 데이

64) 의사표시의 합치란 서로 상대하는 (당사자에게는 다른 사회적·경제적 의의를 가지는) 두 개 이상의 의사표시가 객관적·주관적으로 합치하는 것을 뜻한다(이정운, “개인정보 처리방침의 법적성격과 활용방안 - 개인정보 처리방침의 약관성에 대한 비판적 고찰을 중심으로 -”, 「가천법학」, 제10권 제1호(2017), 56면).

65) 제532조(의사실현에 의한 계약성립)청약자의 의사표시나 관습에 의하여 승낙의 통지가 필요하지 아니한 경우에는 계약은 승낙의 의사표시로 인정되는 사실이 있는 때에 성립한다.

66) 약관계약이라고 보기 어렵다는 견해는 김현숙, “크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토”, 「강원법학」, 제61권(2020), 243-244면.

67) 강정희, “데이터 수집과 그 제한의 한계 - 야놀자 판결과 링크드인 판결 사이에서-”, 「경쟁법연구」, 제46권(2022), 135면.

68) 데이터는 소유권, 점유권, 용익물권, 담보물권의 대상이 되지 않기 때문에 저작권 등의 지식재산권이 발생하는 경우를 제외하면, 현행법상 데이터에 소유권 기타 물권적인 권리는 성립할 수 없다(한국지식재산학회, “데이터 거래 가이드라인”, 한국데이터산업진흥원, 2019, 6면).

69) 저작권법은 데이터베이스를 저작물로 보호하는 것이 아니라 일종의 독자적 권리(sui generis rights)로 보

터에 대한 준물권성 인정이 향후에 가능할 여지가 있다고 생각한다. 실제로 미국에서는 데이터라는 개념이 없었던 1918년 저작권성이 부정된 사실 전달에 불과한 기사에 대한 제한적인 준물권(quasi-property)이 인정되기도 했다.⁷⁰⁾

만약 데이터에 대한 준물권성이 인정될 수 있다면 로봇배제표준은 제3자에게 데이터에 대한 자신의 권리를 식별할 수 있도록 하는 공시방법으로 해석될 여지도 있다고 생각한다.

3.2.4. 유일한 데이터의 독점 배제

한편, 공개 데이터를 보유하고 있는 사업자의 경우에 데이터에 대한 수집을 금지하는 행위를 시장지배적 지위 남용행위 또는 불공정거래행위로 규제할 여지가 있다는 점에서 적용가능성이 존재한다는 점은 상기에 설명한 바와 같다.

이에 향후 데이터에 대한 권리가 만일 준물권으로까지 인정 하더라도 데이터의 독점을 허용해서는 안 될 것이다. 즉, 사실상 수집할 수 있는 유일한 자료와 다를 바 없는 데이터를 보유한 자가 이용약관, 로봇배제표준 등을 통해 이를 금지하는 행위는 독점규제법상 시장지배적지위의 남용이나 불공정거래행위가 될 수 있으므로 이를 배제하는 것을 요건으로 해야 한다. 이용약관과 로봇배제표준 등에 명시적으로 금지된 수집이라고 하더라도 유일한 공개 데이터는 수집하는 것이 언제나 적법한 것으로 판단해야 한다.

4. 결론

일반적으로 공유물로 인식되던 데이터에 대하여 최근 일어난 일련의 국내외 분쟁으로 인하여 자동화된 기술을 통해 공개 데이터를 수집하는 경우라도 현행법에 저촉될 수 있음이 분명해졌다. 공개 데이터를 수집할 경우 저촉될 수 있는 법으로는 형법, 정보통신망법, 개인정보보호법, 저작권법, 부정경쟁방지법, 독점규제법이 있지만 이 중 형법상 컴퓨터등장애업무방해죄와 정보통신망법상 정보통신망침입죄가 성립할 여지가 없고 개인정보와 저작물의 경우 별도로 동의를 받아야 한다는 점에서 별도로 취급되어야 하며 부정경쟁방지법상 데이터부정사용행위는 공개 데이터에 해당될 여지가 없다. 공개 데이터를 수집할 경우 적법성 확보에 중점이 되는 것은 저작권법상 데이터베이스제작자의 권리를 침해하거나, 부정경쟁방지법상 성과모용행위에 해당하는 기준일 것이다.

저작권법과 부정경쟁방지법을 비교해 데이터 보호 또는 침해의 기준을 정리해 보면 공개 데이터를 수집함에 있어 적법성 판단의 기준들은 데이터에 대한 권리를 부여함에 있어 너무 관대한 기준이며 저작물보다 데이터베이스를 보호할 가능성이 있는 형평에 어긋난 기준이라는 것을 알 수 있다. 더욱이 문제되는 것은 침해 판단을 행위시가 아닌 판결시에 결정된다는 점과 수집 기술의 생래(生來)상 적용되기에 불가능하여 불법의 위험을 동반할 가능성에서 나아가 자동화된 데이터 수집은 언제나 불법이 될 수밖에 없다는 점이다. 이러한 기준에 대하여는 변화가 필요할 것이며 특히 행위시 적법성을 판단에 작동될 수 있는 기준이 바람직 할 것이다. 현행 기

호하고 있다.

70) 대표적으로 *International News Service v. Associated Press* 사건은 사실정보를 보호하는 방안으로 저작권보다는 부정경쟁방지법의 관점으로 정보의 무임승차 방지 및 시장 경쟁에 대한 보호를 도모하였는데 미국 연방대법원은 사실 전달에 불과한 기사의 저작권성은 부정하였지만, AP와 INS의 직접적인 경쟁 관계 및 언론 산업에서 정보의 신속성이 미치는 영향을 고려하여 정보를 얻기까지 들어간 비용과 노력을 회수할 때까지 제한적인 준물권(quasi-property)을 인정하였고 INS의 행위는 AP 재산의 부정이용(misappropriation)으로 부정경쟁방지법 위반이라고 판시하였다. *International News Service v. Associated Press*, 248 U.S. 215 (1918); 심석찬, “데이터 크롤링 보호 대상에 관한 소고 - 채용정보 크롤링 사례를 중심으로”, 『IP & Data 法』, 제3권 제2호(2023), 97면.

준 중 공정한 관행이라는 기준이 이러한 범에서 주목할 만하다고 생각된다.

공개 데이터 수집과 관련된 자동화 기술에는 API, 크롤링, 스크래핑, 미러링이 있으며 이러한 자동화된 기술을 통해 공개 데이터 수집이 이루어지고 있는 공정한 관행은 데이터 보유자가 명시한 이용약관, 로봇배제표준을 따르는 것, 저작권과 개인정보보호 규정을 준수하는 것, 그리고 대상 서버에 과도한 트래픽 부하를 일으키지 않는 것으로 요약될 수 있다. 저작물과 개인정보에 대하여 사전동의를 받는 것은 별론으로 하고 관행에 따르는 것은 일반적으로 이용약관과 로봇배제표준을 준수하는 것으로써 이를 통해 공개 데이터 수집에 적법성을 확보할 여지가 있다. 이용약관은 법적으로 약관계약이며, 로봇배제표준은 현재 불분명하지만 약관계약 또는 향후에는 공시방법으로 볼 여지가 있으며 약관계약의 경우 데이터 보유자의 의사표시에 따른 수집을, 공시방법으로 본다면 표시된 권리의 내용을 침해하지 않는 수집을 통하여 적법성을 확보할 수 있을 것이다. 이와 더불어 사실상 수집할 수 있는 유일한 자료와 다를 바 없는 데이터에 대한 수집을 금지하는 행위는 독점규제법상 시장지배적지위의 남용이나 불공정거래행위가 될 수 있으므로 이러한 데이터의 수집은 언제나 적법한 것으로 판단되어야 한다. 이상에서는 현재 존재하는 법률에 기초하여 공개 데이터의 수집에 대한 적법성 확보 방안을 논하여 보았다. 추후에는 신기술에 발맞추어 현행 법률에 국한되지 않은 적절한 기준들이 논의될 필요성이 있을 것이라 생각한다.

참고문헌

단행본(국내 및 동양)

- 법원 지식재산권법연구회·사단법인 한국특허법학회, “웹 크롤링과 성과 무단 사용”, 권보원(편), 부정경쟁방지법 판례백선, 박영사, 2024.
- 법원 지식재산권법연구회·사단법인 한국특허법학회, “크롤링 프로그램을 이용하여 숙박업소 정보를 복제 및 영업상 이용한 행위가 성과 도용 부정경쟁행위에 해당 여부”, 차상욱(편), 부정경쟁방지법 판례백선, 박영사, 2024.
- 오승중, 「저작권법」, 제5판, 박영사, 2021.
- 허희성, 「신저작권법 축조개설」, 명문프리컴, 2011.

학술지(국내 및 동양)

- 강정희, “데이터 수집과 그 제한의 한계 - 야놀자 판결과 링크드인 판결 사이에서-”, 「경쟁법연구」, 제46권(2022).
- 고수운, “기술발전에 따른 저작권 보호 및 이용방식에 관한 소고”, 「계간 저작권」, 제37권 제1호(2024).
- 고수운, “빅데이터의 개인정보 이용으로 인한 법익침해의 私法적 해석”, 「강원법학」, 제52권(2017).
- 김현숙, “크롤링을 이용한 공개데이터 수집·활용의 법적쟁점에 대한 비판적 검토”, 「강원법학」, 제61권(2020).
- 나철원·온병원, “최신 웹 크롤링 알고리즘 분석 및 선제적인 크롤링 기법 제안”, 「인터넷정보학회논문지」, 제20권 제3호(2019).
- 선종수, “크롤링을 통한 데이터 수집과 형사책임 - 대법원 2022.5.12. 선고 2021도1533 판결을 중심으로”, 「형사법연구」, 제35권 제3호(2023).
- 심석찬, “데이터 크롤링 보호 대상에 관한 소고 - 채용정보 크롤링 사례를 중심으로”, 「IP & Data 法」, 제3권 제2호(2023).
- 이정운, “개인정보 처리방침의 법적성격과 활용방안 - 개인정보 처리방침의 약관성에 대한 비판적 고찰을 중심으로 -”, 「가천법학」, 제10권 제1호(2017).
- 정현순, “저작권법상 데이터베이스제작자 권리 침해에 관한 연구 - 대법원 2022. 5. 12. 선고 2021도1533 판결을 중심으로 -”, 「계간 저작권」, 제35권 제3호(2022).
- 황태희, “웹 크롤링의 경쟁법적 고찰”, 「경쟁법연구」, 제43권(2021).

학술지(서양)

- Monika Isia Jasiewicz, “Copyright Protection in an Opt-Out World: Implied License Doctrine and News Aggregators”, *The Yale Law Journal*, Vol.122(2014).

학위논문(국내 및 동양)

- 고수운, “빅데이터 기술환경에서 정보의 이용에 따른 침해판단방법론”, 강원대학교 대학원, 박사, 2019.
- 한지영, “데이터베이스의 법적 보호에 관한 연구”, 서울대학교 대학원, 박사, 2005.

판례

- 대법원 2005. 11. 25. 선고 2005도870 판결.
- 대법원 2017. 4. 13. 선고 2017다204315 판결.
- 대법원 2017. 8. 24. 선고 2017다224395 판결.
- 대법원 2020. 3. 26. 선고 2016다276467 판결.
- 대법원 2020. 3. 26.자 2019마6525 결정.
- 대법원 2020. 6. 25. 선고 2019다282449 판결.

대법원 2020. 7. 9. 선고 2017다217847 판결.
 대법원 2020. 7. 23. 선고 2020다220607 판결.
 대법원 2022. 4. 28. 선고 2021다310873 판결.
 대법원 2022. 5. 12. 선고 2021도1533 판결.
 서울고등법원 2016. 12. 15. 선고 2015나2074198 판결.
 서울중앙지방법원 2006. 7. 21. 선고 2004가합76058 판결.
 서울중앙지방법원 2014. 8. 28. 선고 2013가합552431 판결.
 서울중앙지방법원 2016. 2. 17. 선고 2015가합517982 판결.
 서울중앙지방법원 2020. 2. 11. 선고 2019고단1777판결.
 서울중앙지방법원 2021. 1. 13. 선고 2020노611 판결.
 서울중앙지방법원 2024. 9. 27. 선고 2022가합528041 판결.
 International News Service v. Associated Press, 248 U.S. 215 (1918) .
 The Authors Guild v. Google, Inc., 578 U.S. 941, 136 S. Ct. 1658, 194 L. Ed. 2d 800 (2016).

신문기사

박현익, 「데이터 수집부터 겁나는 기업들… “불법인지 합법인지 모르겠다”」, Chosun Biz, 2020.09.23.자.
 송종복, 「대동강물 사기친 봉이 김선달」, 경남매일, 2015. 5. 15자.
 양혜윤, 「[지식재산권] AI 데이터베이스 크롤링과 관련된 법적 쟁점과 판례 해설」, 법률신문, 2024. 3. 26자.
 이광옥 외 2인, 「크롤링, 최근 개정된 부정경쟁방지법이 바라보는 시각」, 법률신문, 2022. 4. 28자.

인터넷 자료

나무위키, “미러링”, 나무위키, <<https://namu.wiki/w/%EB%AF%B8%EB%9F%AC%EB%A7%81>>, 검색일: 2025. 6. 6.
 위키백과, “해킹”, 위키백과, <<https://ko.wikipedia.org/wiki/%ED%95%B4%ED%82%B9>>, 검색일: 2025. 6. 6.
 한국정보통신기술협회, “크롤링”, 정보통신용어사전, <https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=179640-1>, 검색일: 2025. 6. 6.
 한국정보통신기술협회, “스크래핑”, 정보통신용어사전, <https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=185196-1>, 검색일: 2025. 6. 6.
 AWS, “애플리케이션 프로그래밍 인터페이스(API)란 무엇인가요?”, AWS, <<https://aws.amazon.com/ko/what-is/api/>>, 검색일: 2025. 6. 6.

기타 자료

과학기술정보통신부·한국데이터산업진흥원, “2023 데이터산업 현황조사”, 과학기술정보통신부·한국데이터산업진흥원, 2023.
 문화관광위원회, “저작권법중개정법률안 심사보고서”, 국회, 2003.
 산업통상자원중소벤처기업위원회, “부정경쟁방지 및 영업비밀보호에 관한 법률 일부개정법률안 검토보고 (김경만의원 대표발의, 의안번호 제2107535호) 검토보고서”, 국회, 2021.
 특허청·지식재산보호원, “부정경쟁방지법상 데이터 보호 제도 안내서”, 특허청 산업재산보호정책과, 2024.
 한국지식재산학회, “데이터 거래 가이드라인”, 한국데이터산업진흥원, 2019.