

# 데이터 수집과 분석 관련 법제의 비교법적 연구

강백서\*†, 신민지\*

## A Comparative Study on the Legislation Related to Data Collection and Analysis

Baek-Seo Kang\*†, Min-Ji Shin\*

### 요 약

과학기술이 발달함에 따라 방대한 데이터를 수집 및 처리하기 용이한 기술들이 많이 생겨났다. 많은 기업들은 데이터를 수집하기 위해 데이터 수집 프로그램을 활용하고 있다. 데이터를 확보하기 위한 기업간 분쟁이 과열되고 있으나, 정립되지 않은 법체계는 산업 전반에 혼란을 야기하고 있다. 일부 판결에서는 문제의 소지가 있는 이용부분이 피고저작물에서 원고저작물에 차지하는 양적 또는 질적인 비중과 중요도를 맹목적으로 검토하고 있다. 데이터 혹은 데이터베이스 자체에 중점을 둔 법적 체계의 형성을 고려하는 것은 사실상 개념적 체계가 근간이 되어야 한다. 하지만 유동적이고 방대하며, 소유의 개념으로 분석하기 어려운 데이터 산업을 정립하는 것은 쉽지 않다. 따라서 데이터 수집 및 분석을 도모함과 동시에 보호도 함께 기여하고자 이용행위에 대한 법체계를 논의하는 방안을 검토하였다. 따라서 비교법적 고찰을 통해 이용행위에 대한 법적 체계를 형성하여 문제를 해결하고자 한다.

### Abstract

With the development of science and technology, many technologies have emerged that make it easy to collect and process large amounts of data. Many companies use data collection procedures to collect data. Disputes between companies to secure data are heating up, but the legal system, which has not yet been established, is causing confusion throughout the industry. In some judgments, the portion of the Defendant's work on the utilization of the problematic part is being blindly reviewed, accounting for the proportion and importance of the Plaintiff's work in terms of quantity and quality. Considering the formation of a legal system focusing on data or the database itself, it should be based on a conceptual system. However, it is not easy to create a data industry that is fluid, large, and difficult to analyze with the concept of ownership. Therefore, it is the solution to review utilization activities to contribute to protection while promoting data collection and analysis. Therefore, we try to form a legal system for crawling and analysing the data through comparative law examination to solve the problem.

**한글키워드** : 데이터 수집 프로그램, 크롤링, 데이터공유, 데이터산업, 저작권법

**keywords** : Data Collection Program, Crawling, Data Sharing, Data Industry, Copyright Act

\* 중앙대학교 일반대학원 융합보안학과 석사과정

† 교신저자: 강백서(email: cream8629@naver.com)

접수일자: 2022.11.18. 심사완료: 2022.12.07.

게재확정: 2022.12.20.

## 1. 서론

디지털 패권 경쟁으로 들어서게 됨으로서 데

데이터의 중요성이 높아지고 주요국들은 데이터 보호에 대한 문제를 국가안보관점에서 접근하고 있다. Marketsandmarkets 연구 보고서에 따르면 전 세계의 빅데이터 시장이 2025년 2,294억 2,300만 달러를 달성할 것으로 전망하고 있으며, 주요국들은 데이터를 기반한 신산업 육성 및 데이터 보호 등을 위한 다양한 정책을 수립하고 있다. 과학 기술이 발달함에 따라 데이터 수집 프로그램이 활성화되었으며 가장 대표적인 데이터 수집 프로그램인 '웹 크롤러'가 존재한다. 오늘날 웹 크롤링 기술은 검색엔진을 통한 마케팅 산업을 활성화 시키는 주된 이유가 되었으며, 이외에도 광범위하게 활용되고 있다. 많은 기업들은 데이터를 수집하기 위해 웹 크롤러와 같은 데이터 수집 프로그램을 활용하고 있으며, 이러한 방식은 현시대에서 불가피한 원천기술이라 볼 수 있다. 최근 데이터 산업에서는 무임승차라는 시각도 존재한다. 이러한 상반된 시각이 법적 분쟁을 만들어 내고 있으며, 법원에서는 저작권법상 데이터베이스제작자의 권리 등을 통해 판단하고 있으나, 데이터의 특성상 다양하고 방대한 양을 가져 모든 사건이 해당 조항만으로는 판단할 수 없는 실정이다. 따라서 이를 「부정경쟁방지 및 영업비밀보호에 관한 법률」(이하 부정경쟁방지법)과 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」(이하 정보통신망법)등으로 판단하고 있다. 이는 데이터에 대해 권리를 부여하는 것이 아닌 부정한 행위를 규제하는 행태를 보이고 있다. 상당량의 데이터의 수집하는데 들인 투자와 노력을 누군가 무임승차를 하는 행위를 제재하기 위해 데이터를 보호하는 방식을 택한 것이다[1]. 그러나 여전히 기업들은 정보 활용에 있어서 데이터 수집과 관련해 명확한 법 규정이 없고 일부 판례만 참고가 가능하기 때문에 위법성을 판단할 수 없다. 데이터 산업의 육성을 위해 구체적인 정책들이 제시되어야 하지만 한국의 경우에는 법

적 기준 및 균형있는 정책들이 미흡하다. 데이터 수집을 위한 행위의 위법성을 판단할 때 데이터의 양과 질의 상당성으로 판단하게 되며, 이는 매우 상대적인 기준으로 법원의 판단에 따라서 결정이 되는 사안이다. 따라서 데이터 수집을 위해서 데이터를 자발적으로 데이터 시장에 제공하여 산업을 활성화 해야 한다[2][3]. 따라서 데이터와 데이터베이스를 구별하고 이에 대한 체계를 논의하기 보다는 데이터 수집 및 분석에 중점을 두고 논의할 필요성이 있다. 필자는 관점의 전환을 통해 이용과 거래에 대한 중심적인 논의를 진행하여 비교법적 고찰이 필요한 문제라고 판단한다.

## 2. 사안의 분석

### 2.1 주요 판례 분석

데이터 수집에 대한 분쟁이 지속적으로 발생하고 있으며 이러한 행위에 대한 객관적인 판단이 불명확하여 불필요한 사회적 분쟁이 지속적으로 발생할 것으로 보인다.

#### 2.1.1 리그베다위키 vs 엔하위키미러

원고 리그베다위키는 인터넷을 통해 각 주제별로 그에 관한 설명을 제공하는 온라인 백과사전의 일종으로, 이용자들이 특정 주제에 대해 자유롭게 게시물을 작성 및 수정하는 방식으로 운영하였으며, 피고 엔하위키미러가 2009년경부터 별지 목록 기재 각 도메인을 이용해 미러링 방식으로 원고의 사이트에 접속해 '위키' 게시물 전부를 복제한 뒤 인터넷 사이트를 개설한 사건이다. 해당 사건은 소재의 제작과 데이터베이스 제작을 구별하여 원고의 사이트를 제작 및 그 소재의 갱신, 검증 또는 보충을 위해 인적 또는 물적으로 상당한 투자가 있었다고 판단하였고 복제권 침해 및 전송권 침해가 있었다고 판단하였다[4].

### 2.1.2 잡코리아 vs 사람인

원고 잡코리아는 온라인 리크루팅 서비스와 인터넷을 기반으로 채용대행사업, 직업정보 제공사업, 직업소개업 등을 바탕으로 운영하고 있는 회사이다. 피고 사람인은 원고와 경쟁업체의 관계에 있으며, 2008년부터 수차례의 합의와 법원 조정 및 판결이 있었음에도 원고가 제공하는 채용정보를 무단으로 크롤링하여 자사 사이트에 게재하였다. 법원은 원고의 웹 사이트가 여러 구인업체들의 채용정보를 체계적으로 배열 정리한 점과 이용자가 각종 정보를 검색할 수 있는 점을 들어 데이터베이스에 해당한다고 판시하였으며 개별 채용정보를 분류 및 정리 하여 사이트에 올려 이용자들이 빠르게 열람 및 검색을 할 수 있도록 웹 사이트를 관리한 점을 들어 데이터베이스 제작자에 해당한다고 판시하였다[5].

### 2.1.3 야놀자 vs 여기어때

피해회사인 ‘야놀자’는 숙박정보를 제공하는 인터넷 웹 사이트 및 앱을 운영하는 회사이고, 피고인 ‘여기어때’는 피해회사의 경쟁업체의 관계이다. 2015년부터 피고인은 피해회사의 모바일 앱이나 웹 사이트에 지속적으로 접속하여 제휴 숙박업소의 목록이나 주소, 가격 등 숙박정보를 크롤링하여 내부적으로 공유한 사건이다. 최종적으로 법원은 데이터베이스에 관한 권리만을 강조하면 경쟁의 자유 또한 침해되고, 피고인들이 취득한 정보는 피해회사의 영업을 위해서 공개할 수 밖에 없는 정보이며, 이를 수집하는 것이 상당한 노력이나 비용이 들지 않는 점을 들며 무죄로 판단하였다. 또한 상당한 부분의 복제는 양적 상당성과 질적 상당성을 바탕으로 판단하였으며 복제의 양적 부분이 50개의 항목 중 8개의 항목이었다는 점과 질적 부분에서는 데이터 성질, 통상적인 이용, 데이터 사용 용도와 목적, 데이터 수집 난이도, 기간의 경과, 데이터베이스제작자의

이익, 공개 유무 등 제반 사정을 종합하여 평가하였다[6].

## 2.2 행위 객체의 특성

### 2.2.1 데이터의 영역

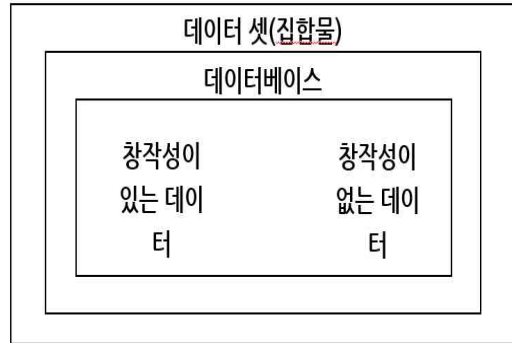


그림 1. 데이터의 영역  
Fig 1. Area of data

크롤러의 동작과정을 살펴보면 다음과 같다.

- 첫째. 웹 페이지 서칭 및 리소스 수집
- 둘째. 웹 페이지 파싱 및 추출 분석
- 셋째. 데이터 포매팅
- 넷째. 원하는 프로세스로 저장 및 사용

사실상 크롤링 된 데이터 셋과 데이터베이스는 크롤러의 동작과정을 보면 배열 정리 검색이 쉽게 가능하며 실제 불가분한 행위로서 이루어지고 있으므로 이에 대한 구별 실익이 없다.

### 2.2.2 보호영역에 대한 판단

법원은 체계적으로 배열을 정리하여 이용자가 각종정보를 검색할 수 있도록 하는지 판단한다. 하지만 수집된 데이터는 이용자의 이용목적에 따라 이용가치가 다분화되어 있다. 또한 양과 질에 대한 판단 역시 통일성 있지 않으며 필요성에 따

라 판단된다. 이용영역과 불가침영역에 대한 판단은 데이터 산업의 작지만 중요한 요소인데 이러한 사소한 판단부터 난관에 봉착하는 것은 산업부흥 큰 장애물이 된다.

### 3. 데이터 수집 행위에 대한 판단기준

#### 3.1 데이터 수집 프로그램의 개념

대량의 데이터를 수집하고 분석하는 컴퓨터 프로그램을 말하며 검색서비스, 스캐닝(Scanning), 스크래핑(Scraping) 기반의 금융서비스, 전자상거래상에서의 가격비교, 정보보호 서비스 등 다양한 분야에서 활용되고 있다[7]. 문서, 이미지, 동영상 등 형태와는 무관하게 데이터를 수집하고 축적하는데 용이하다는 점에서 데이터베이스·데이터셋 구축에 활용되고 있다[8]. 데이터가 재화가 되는 시대에서 경쟁력을 갖기 위해 이러한 프로그래밍 기법을 이용해 데이터를 수집한다.

#### 3.2 데이터 수집 프로그램의 유형

##### 3.2.1 크롤링

크롤링(Crawling)은 원격에 위치한 리소스를 범용적인 TCP/IP 통신, 특히 HTTP(S) 통신을 통해서 가져오는 기법을 의미한다. 크롤링 기법의 대중화로 인해 웹 사이트를 옮겨다니며 정보를 수집하는 로봇인 ‘크롤러’가 개발되었으며 웹 사이트를 방문하여 생산요소로서 이용이 가능한 데이터를 신속하게 복제·수집하는 방식[9]으로 데이터 수집의 활성화에 기여하고 있다. 크롤러가 웹 문서 내에서 원하는 데이터를 특정하여 수집할 수 있으며 다양한 형태로 존재한다. 집중크롤러(focused crawling)의 경우 크롤링한 정보를 분류기(classifier)에서 어떤 범주에 포함시킬지를

결정한 후 분배기(distiller)를 통해 크롤링한 페이지를 현출하는 결과의 선후관계를 결정하여 소비자가 필요로 하는 검색결과를 빠른 순위별로 나타낼 수 있다[10].

크롤링은 탐색, 분류, 저장의 과정을 거쳐 이루어진다. 중소기업자나 개인들도 구글 스프레드시트 함수, 엑셀 또는 크롤러 개발 전용 라이브러리를 통해 쉽게 제작하고, 인터넷에서 구하거나[11] 전문 크롤링 업체[12]를 통해 필요한 정보를 수집할 수 있다.

##### 3.2.2 웹 마이닝과 TDM

모든 분야에서 데이터는 필수적인 요소가 되며 데이터를 수집하고 분석하여 새로운 정보를 추출하는 과정이 더욱 중요시되고 있다. 웹으로부터 필요한 정보를 찾고 이를 추출하여, 원하는 서비스를 제공할 수 있는 정보를 도출하는 기술이 바로 웹 마이닝이다[13]. 웹 마이닝은 머신러닝을 이용한 정형화되지 않은 텍스트 데이터를 유형화하고, 기술적 방법을 통해 필요한 정보만을 가져오는 텍스트마이닝과 데이터들 간 유의미한 상관관계를 통해 필요한 정보만을 추출하는 데이터마이닝을 말한다. 이를 Text and Data Mining(이하 TDM)으로 정의한다[14]. 필요한 정보를 가져오는 것이 크롤링과 유사하지만, TDM은 대규모 데이터로부터 중요한 흐름이나 정보를 분석하고 해석하는 과정이다[15].

##### 3.2.3 소결

형태와 관계없이 우리가 사용하는 자료나 정보가 데이터라고 한다면 해당 데이터를 수집하고 분석하는 작업을 통해 축적된 것은 데이터베이스이다. 앞서 언급한 유형 이외에도 데이터 수집 및 분석을 통해 필요한 정보만 추출하는 것 모두 데이터 수집 프로그램이라고 볼 수 있다.

### 3.3 프로그래밍을 통한 데이터 수집행위에 대한 법원의 태도

3.3.1 저작권법상 데이터베이스제작자의 권리 다수의 판례에서 데이터 수집행위에 대한 위법성을 판단할 때에 저작권법 제93조(데이터베이스 제작자의 권리)를 통해 법원에서는 판단하고 있다. 데이터베이스제작자의 권리는 전부 또는 상당한 부분을 복제·배포·방송 또는 전송할 권리를 가지며 데이터 개별소재는 상당한 부분으로 간주 되지 않는다. 하지만 반복적이거나 특정한 목적을 위해 체계적으로 함으로써 통상적인 이용에 충돌하거나 데이터베이스제작자의 이익을 부당하게 해치는 경우에는 상당한 복제로 본다.

3.3.2 부정경쟁방지법상 무단 성과 도용 행위 현 부정경쟁방지법이 개정되기 이전에 법원에서 데이터의 부정사용행위를 판단할 때 다수의 판례들은 부정경쟁방지법 제2조 제1항 (파)목으로 판단하였으며, 이는 타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위를 부정경쟁행위로 판단하였다. 하지만 이는 미래에 데이터의 이용 및 유통 행위를 재제하기에는 어려움이 있어 현 부정경쟁방지법에서는 제2조 제1호 (카)목을 신설하여 데이터 부정사용행위를 부정경쟁행위로 보게 되었다.

## 4. 비판적 고찰

### 4.1 판단기준의 모호성 : 상당성

저작권법상 타인의 데이터베이스에서 데이터를 추출하는 사안이 문제된다. 데이터베이스의 제작자는 데이터베이스의 전체 혹은 상당한 부분

표 1. 사안의 분석  
Table 1. An analysis of an issue

리그베다위키 vs 엔하위키미러 (서울고법 2016.12.15. 선고 2015나2074198 판결.)
미러링 방식으로 위키 게시물 전부를 복제하여 불법행위로 인정
잡코리아 vs 사람인 (대법원 2017.8.24. 선고 2017다224395 판결.)
채용정보를 허락없이 크롤링 하여 자사 사이트에 불법행위로 인정
야놀자 vs 여기어때 (대법원 2022.5.12. 선고 2021도1533 판결)
숙박정보를 크롤링하여 영업을 위해 내부적으로 공유하여 합법

을 복제, 배포, 방송 또는 전송할 권리를 갖는다. (저작권법 제93조 제1항) 데이터베이스로부터 소재가 되는 데이터들을 추출하는 것부터 추출한 소재로 새로운 데이터베이스를 마련하거나 새로운 방식으로 이용하는 것 또한 포함될 수 있다 [16]. 데이터베이스는 단순한 데이터의 집합체를 조직하는 것에 그치는 것이 아니라, 정보화가 진행되는 분야에서 주된 역할을 차지하고 있다[17].

### 4.2 보호의 필요성 증대

#### 4.2.1 지속적인 분쟁의 발생

네이버가 부동산 중개 플랫폼인 다윈중개에 대해 ‘크롤링 금지’ 민사소송을 제기했다. 다윈중개가 네이버 부동산의 매물을 크롤링하여 ‘아웃링크’ 방식으로 자사의 플랫폼에 노출하였다. 여기서 아웃링크 방식이란 기존의 출처 웹페이지로 넘어가게 되는 방식이며, 다윈중개 플랫폼에서 매물을 클릭하면 네이버 부동산으로 넘어가는 시스템이다. 네이버 부동산은 부동산 매물정보를

상당한 투자와 노력을 하였으며, 이는 보호되어야 한다고 하였다. 그러나 다윈중개는 이미 공개된 데이터를 활용한 것이고 누구나 접근 가능한 데이터를 아웃링크 방식으로 연결하여 오히려 네이버 부동산에 도움을 줬다는 입장이다. 수원지방법원 성남지원에서는 양측의 입장을 고려하여 다윈중개 매물란에 자사의 플랫폼으로 유도하는 광고문구를 삭제하라는 화해권고를 내렸다.

#### 4.3. 데이터 산업 시장의 변화

데이터베이스 보호제도는 약 13년 동안 운영되고 있다. 법률 문헌상 불확정 개념을 사용하여 해석의 여지를 둠으로써 해당 영역에서의 축적된 판례를 기반으로 법리의 형성을 기대하였으나 현재 연구진이나 실무진들의 관심을 받지 못하고, 제도적으로도 큰 변화 없이 현 상태에 머물러 있다. 특히 데이터베이스의 보호를 강변했던 산업계도 해당 제도를 대부분 활용하지 않았으며, 이 때문에 중요한 기반이 될 수 있는 판례들 또한 축적되지 못한 상황이다. 데이터베이스 보호가 동 산업의 육성과 보호라는 정책에 의해 탄생한 범영역이라 볼 수 있다. 불확정적인 개념의 모호성이 극복되지 못한다면, 그만큼 해당 제도는 실효성을 갖지 못하게 될 가능성이 있다 [18]. 또한 데이터의 보호에 대해서는 지식재산권법(특허법·부정경쟁방지법·저작권법을 포함한 산업재산권법)이 적용되는데 이러한 법체계는 혼란을 야기할 위험이 있다. 이처럼 데이터 관련 입법의 필요성이 지속적으로 제시되고 있지만, 그 보호대상에 대한 모호성과 유동적 데이터의 성격 때문에 여전히 논의만 지속되고 있다. 디지털 산업의 안정적인 기반을 마련하기 위해서는 데이터 보호체계에 대한 고찰이 필요하다.

데이터베이스 소유자는 데이터베이스 개발 및 유지, 관리하는데 발생하는 비용과 노력이 크게 증가했으며 과학기술의 발달로 인해 데이터베이스

를 복제·공유하는 기술 또한 크게 향상되었다. 이를 바탕으로 데이터베이스에 대한 투자비용 회수를 위해 강력한 데이터베이스 보호가 필요하다고 주장한다. 이들은 데이터베이스 보호를 위한 실효성 있는 제도가 마련되지 않을 시 데이터베이스 개발자들은 개발 의욕을 상실하게 될 것이며 산·학·연이 모두 힘들어 질 것이라고 주장했다. [19] 결국 데이터베이스는 개발비용이 높은 반면에 복제비용은 낮기 때문에 실질적으로 동일한 물품을 산출할 수 있다. 따라서 데이터베이스 복제자는 연구 개발비용이 없기 때문에 데이터베이스 개발자보다 현저하게 싸게 가격을 책정할 수 있다. 이것은 곧 데이터베이스 개발자의 투자비 회수를 어렵게 하기 때문에 시장실패를 초래한다[20].

### 5. 해외의 대응방안 검토

#### 5.1 영국 개정 저작권법 제29조A

2014년 6월 1일부터 영국은 저작권법의 개정을 하여 과학 및 기술 연구의 새로운 발전을 위해 비상업적인 연구를 통한 데이터 분석, 즉 TDM을 할 수 있도록 제29A 규정을 신설하였다 [21]. 신설된 규정은 다음과 같다.

표 2. 영국저작권법  
Table 2. Copyright, Designs and Patents Act

구분	내용
영국 저작권법 (CDPA)	제29A조(비상업적 연구 목적의 텍스트 및 데이터 분석을 위한 복제) (1) 저작물에 대하여 적법한 접근 권한을 가지고 있는 사람이 해당 저작물의 복제물을 작성할 때 이

<p>하의 조건을 모두 충족하는 경우 저작권을 침해하지 않는다.</p> <p>(a) 복제는 저작물에 대하여 적법한 접근 권한을 가지고 있는 사람이 비상업적 연구만을 목적으로 해당 저작물에 기록되어 있는 것을 컴퓨터에 의하여 분석하기 위하여 작성하는 것</p> <p>(b) 복제물에 충분한 출처 명시를 하는 경우(단, 실현가능성 등의 이유로 불가능한 경우는 제외한다)</p> <p>(2) 이 조에 근거하여 작성된 저작물의 복제물은 이하의 경우 저작권을 침해한다.</p> <p>(a) 그 복제물이 타인에게 양도되는 경우(단, 그 양도가 저작권자에 의하여 허락된 경우는 제외한다.)</p> <p>(b) 그 복제물이 (1)(a)에서 언급된 것 이외의 목적을 위하여 사용되는 경우(단, 그 사용이 저작권자에 의하여 허락된 경우에는 제외한다.)</p> <p>(3) 이 조에 근거하여 작성된 복제물이 그 후에 이용되는 경우에는, (a) 그 복제물은 그 이용의 목적상 침해 복제물로 취급한다.</p> <p>(b) 그 이용이 저작권을 침해하는 경우, 그 복제물은 그 후의 모든 목적상 침해 복제물로서 취급된다.</p> <p>(4) 제3항의 ‘이용’이란 판매 또는 임대, 판매와 임대를 위한 청약과 전시를 포함한다.</p>	
--	--

	<p>(5) 계약규정이 이 조에 의하여 저작권 침해가 되지 않는 복제물의 작성을 금지 또는 제한하는 것을 의도하는 경우 그 범위에 관하여 해당 규정은 집행 불능인 것으로 한다[22].</p>
--	--

이는 TDM을 위한 예외규정이 만들어진 것이며, 비상업적인 연구목적으로 TDM을 할 시에는 저작물 등을 복제하는 행위가 위법이 아니라는 점과 저작권자의 사전동의를 받을 필요가 없다. 그러나 허용되지 않는 것도 있다. 첫째, 상업적인 복제는 허용되지 않는다. TDM 목적으로 한 복제라 하더라도 어떠한 방식으로 판매되거나 공중에 제공될 수 없다. 둘째, 불법적인 접근은 허용되지 않는다. 합법적인 접근권을 가지는 연구자가 개인적으로 혹은 기관에 갈음하여 라이선스가 부여되거나 구독 되어진 데이터나 콘텐츠에 한해서만 TDM을 활용할 수 있다. 셋째, 기술적 조치의 회피는 불가하다. 불법적인 방법으로 기술적 보호조치를 회피한다면 민사소송이나 형사소추를 당할 위험이 있다. 넷째, TDM의 결과물의 공유는 불가하다[23].

### 5.2 독일 개정 저작권법 제60조d조

2017년 독일 저작권법 개정을 통해 데이터마이닝 면책규정인 제60d조 규정을 신설하였다. 텍스트 및 데이터마이닝에 관한 저작권 규정을 처음 신설하였으며, 신설된 규정은 다음과 같다.

표 3. 독일 저작권법  
Table 3. Urheberrechtsgesetz

구분	내용
<b>독일 저작권법</b>	제60d조(문서와 데이터 마이닝) (1) 다음의 경우 학술적인 연구

<p>(Urheber rechtsge setz)</p>	<p>를 위한 다수 저작물(원본자료)을 자동으로 분석하기 위하여 허용된다.</p> <p>1. 특히 정규화, 구조화 및 범주화를 통하여 분석할 코퍼스(말뭉치, corpus)를 만들기 위하여 원본자료를 자동적 및 시스템적으로 복제하는 경우</p> <p>2. 공동의 학술적 연구를 위하여 특정 범위로 제한된 사람 또는 학술적 연구의 품질을 검증하기 위한 제3자인 개인이 원본자료에 공개적으로 접근하는 경우 이와 관련해서 이용자는 비상업적 목적을 추구하여야 한다.</p> <p>(2) 제1항의 기준에 따라 데이터베이스저작물이 이용되는 경우에는 제55a조 제1문에 따른 통상의 이용에 해당한다. 제1항의 기준에 따라 데이터베이스의 비본질적 부분에 이용되는 경우에는 데이터베이스의 통상적인 활용 및 제87b조 제1항 제2문과 제87e조에서 의미하는 데이터베이스 제작자의 권한 있는 이익에 부합한다.</p> <p>(3) 자료 수집과 원본자료의 복제는 연구행위의 종료 후에 폐기한다. 또한 공중접근도 종료한다. 그러나 원본자료의 수집과 복제는 제60e조와 제60f조에서 명시된 영구적 보관시설에 인도하는 것이 허용된다.</p>
--	--

위 조항들을 살펴보면 먼저 60d조의 제1항에서 학술적인 연구를 위한 수많은 저작물(원본자

료)의 TDM이 허용되는 경우는 다음과 같다. 1. 정규화, 구조화 및 범주화를 통해 분석할 코퍼스를 만들기 위해 원본자료를 자동적 및 시스템적으로 복제한 것. 2. 공동으로 학술적인 연구를 위해 특정 범위로 한정된 인원들에게 접근하게 하며 학술연구의 질을 검증하기 위해 제3자에게 접근을 가능케 하는 것. 3. 비상업적인 목적을 가진 경우. 또한 제2항에서는 독자적인 데이터베이스에 대해 위의 3가지 방식은 정당한 것으로 간주되고 권리자에 의해 금지가 불가능하다. 마지막으로 3항에서는 코퍼스 및 원본자료의 사본은 연구행위의 종료 후에 삭제하고 공중에 대한 접근을 종료한다. 하지만 원본자료의 수집과 복제는 도서관 및 기록보관소 등 영구적 보관시설에 영구보존은 가능하다. 따라서 독일 저작권법은 원자료의 삭제와 보존 및 보상 의무까지 규정하여 가장 구체적인 규범으로 평가받고 있다[24].

### 5.3 일본 저작권법 제47조의7

과거 일본은 데이터베이스를 저작권법상 편집 저작물 규정에 포함하여 편집저작물로 보호하였다. 하지만 1986년 저작권법 개정으로 인해 기존에 규정된 편집저작물과는 다른 데이터베이스에 대해 규정하였다. 일본 저작권법에서의 데이터베이스는 “논문, 수치, 도형 등의 정보의 집합체로서, 그러한 정보를 전자계산기를 이용하여 검색할 수 있도록 체계적으로 구성한 것”이다.(일본 저작권법 제2조 제1항 제10호의3) 또한 “정보의 선택 혹은 체계적 구성에 창작성을 가지는 것”을 데이터베이스저작물로 규정하고 있다.(동법 제12조의2) 우리나라와 달리 일본은 컴퓨터에서 검색할 수 없는 상태는 데이터베이스로 인정하지 않으며 데이터베이스저작물로 인정되기 위해서는 소재의 선택과 배열에 창작성이 인정되어야 한다. 좁은 보호범위는 데이터 이용에 많은 한계점을 야기한다. 1990년대 후반에 인터넷의 보급이

활성화가 되면서 저작물의 이용형태가 달라졌고, 저작물의 이용환경 변화에 대응하기 위한 일환으로 일본은 2009년 저작권법의 일부개정을 통해 저작권법 제47조의 7을 신설하였다.

표 4. 일본 저작권법  
Table 4. Japanese Copyright Law

구분	내용
일본 저작권법	제47조의7(정보해석을 위한 복제 등) 저작물은, 전자계산기에 의한 정보해석(다수의 저작물 기타의 대량의 정보로부터, 당해 정보를 구성하는 언어, 음, 영상 기타의 요소와 관련된 정보를 추출, 비교, 분류 기타의 통계적인 해석을 행하는 것을 말한다. 이하 이 조에서 같다)을 하는 것을 목적으로 하는 경우에는, 필요하다고 인정되는 한도에서 기록매체에 의 기록 또는 번안(이에 의하여 창작한 2차적저작물의 기록을 포함한다)을 할 수 있다. 다만, 정보해석을 하는 자의 이용에 제공하기 위해 작성된 데이터베이스 저작물에 대하여는 그러하지 아니하다.

저작권에게 미칠 영향을 고려하여 컴퓨터를 이용하여 정보해석에 필요한 범위에 한해서 저작물을 기록매체에 기록하거나 번안을 할 수 있도록 저작재산권제한 규정을 두었다[25]. 이러한 규정은 현 시대에서 정보해석의 기술의 사회적 의의를 고려해서 일정한 요건 아래에서 전자계산기를 활용하여 정보해석을 하는 것을 목적으로 할 경우에는 저작물의 기록매체에 기록 및 번안을

허용하는 규정이라 볼 수 있다.

하지만 초기 미국의 경우와 유사하게 너무 넓은 범위의 허용은 데이터 산업의 부흥과 성장에 큰 도움을 주지만, 데이터 유출과 개인정보 등 사생활에 대한 문제, 데이터의 품질에 대한 문제 등의 한계점을 가진다. 넓은 허용범위를 도입하는 것은 숙고해야 할 문제이다.

## 6. 비교법적 논의

### 6.1 외국의 입법례를 통한 해석론적 해결

현행 법은 데이터베이스의 보호목적은 산업에 대한 투자로 본다. 하지만 투자라는 보호법의 자체가 모호한 성질을 가지고 있어 상황에 따라 보호의 범위가 지나치게 확장되거나 축소될 위험이 있다[26]. 저작권법은 입법 목적을 저작자의 권리와 그에 인접한 권리의 보호로 하고 있으며, 저작물이용의 공정성을 추구한다.

데이터와 데이터베이스에 대한 쟁의가 지속적으로 발생하는 것은 보호와 이용의 합의점을 찾는 데 장애물이 된다. 컴퓨터 프로그램을 통한 데이터 수집 행위를 분명하게 판단 내릴 수 없는 것도 이러한 연유이다.

데이터 영역은 신산업 분야로 빠른 변화라는 특성과 분석의 부족으로 법체계를 마련하여 권력적인 규제를 하는 것은 선부른 판단이다. 이러한 환경적 요인은 데이터 산업이 '사적 보호의 경향'으로 치우치게 하여 사회에 혼란을 야기한다. 자율적 규제로 수집행위에 대한 자유를 부여한다면 데이터의 자산성을 확보하기를 원하는 대기업들은 기술적 보호조치로 합법적 자산화의 토대를 만들고 있으며, 이용료를 통한 기업들의 횡포를 막을 수 없다.

따라서 외국의 입법례를 살펴본 것처럼 이에 대한 산업상의 올바른 관습을 형성할 수 있는 가

이드라인이 필요하다. 현재 노력과 투자의 상당성은 그 데이터를 형성하는데 투여된 자본을 중심으로 판단하는 것으로 데이터의 가치 창출 구조를 반영하지 못하고 있다[27]. 이러한 문제를 뛰어넘기 위한 노력이 필요한 실정이다[28][29].

## 6.2 입법론적 해결

데이터베이스가 컴퓨터를 이용하여 정보추출이 가능한 경우 보호의 문제가 발생한다. 이러한 경우, 추출·복제 및 재이용이 매우 쉽기 때문에 침해가 발생할 위험성이 매우 높다. 비정형 데이터의 유의미한 반복성을 찾아, 기술을 이용하여 가공하고 새로운 가치를 도출한다. 시장은 점점 데이터에 대한 독점적 권리를 추구하고 있으며, 법적 권리 체계가 불안정한 상황에서 아무리 산업의 발전을 위해 노력한다 하더라도 시장의 안정성이 위태롭게 된다. 우리는 데이터 보호 체계를 쌓아가야만 하는 시기이다. 결국 데이터베이스의 활용을 위하여 데이터베이스에 대한 배타적 권리를 확대하는 것에 반대하지만 기술적 보호조치 등 자율적 보호조치를 자아내어 발전에 큰 기여를 할 수 있을 것이라는 기대를 하기 어렵다. 미국의 경우에는 기술적 보호조치의 발전이 현재의 수준에 이르기 전 자율적인 통제를 지향하였지만 최근 한국데이터산업진흥원에서 발표한 조사보고서에 따르면, 미국은 최근 ‘미국 데이터 프라이버시 및 보호법’을 발표하여 이전과 달리 보호적 관점의 논의가 필수불가결함을 보여주고 있다.

데이터 산업의 특성상 변화의 속도가 빠르고 법률로써 이를 모두 규정하기는 힘든 실정이다. 따라서 환경적·시대적 상황을 반영하는 탄력적 적용이 필요하므로 위임입법을 통해 하위법률에 규정할 사항이다. 데이터 산업 분야를 활성화하기 위한 정책이 실효성을 갖추기 위해서는 고도의 전문성을 갖추어야 할 것이다. 이를 위해 국

가데이터정책위원회를 통해 심도있는 입법안을 제시할 수 있는 법체계의 마련이 필요하다.

## 6.3 자율적 해결

연구 목적으로 약관을 통한 데이터 이용계약에 대하여 합리적 근거규정을 두는 방법이 가장 많이 이뤄지고 있다. 공공영역이 정보에 접근하는데 장애가 되는 기술적 보호조치를 금하는 입법, 또는 사용용도에 따라 사용자의 권리를 우선 시하여 사용자의 통보를 이용한 제도도 마련되었다. 통보를 받은 권리자는 정해진 기간내에 허용된 범위에 대한 법률의 근거에 따라 허용여부를 결정하고, 허용할 경우 통보를 한 사용자에게 한하여 접근 통제를 해제하고 사용할 수 있게 하는 방법도 존재한다[30].

## 7. 결론

디지털 대전환시대를 맞아 데이터에 대한 쟁점들은 다양하게 나타나고 있다. 데이터의 활용과 권리, 보호 체계가 아직 안정되지 않아 관련 분쟁이 증가하고 있다. 데이터는 우리사회에서 아주 빠른 속도로 주요한 자산으로 자리잡고 있으며 데이터를 수집, 분석, 활용하는 등의 과정에서 경쟁우위를 차지하기 위한 사회 전반의 움직임이 나타나고 있다. 최근 판례들 역시도 이러한 사회적 분위기를 담고 있다. 데이터베이스나 빅데이터를 형성하기 위해 사람들은 공공데이터뿐만 아니라 민간데이터를 수집하여 활용하고 있다. 더 나아가 수집행위의 효율을 위하여 데이터 수집 프로그램을 통해 대량의 데이터를 짧은 시간에 쉽게 수집한다.

공개되어 있는 상태이지만, 자신이 형성하고 수집한 데이터베이스에서 데이터를 동의없이 가져가는 행위에 대해 무임승차라는 시각이 만연하

다. 이를 저작권법상 권리침해로 주장하며 법적 분쟁으로 이어지는 경우가 빈번하다. 데이터와 데이터베이스에 대한 보호영역을 중심으로 논의되고 있다. 하지만 유동적인 형태와 네트워크상에 존재하는 것으로 물건으로 볼 수 없는 대상을 법률적으로 정의하고 보호하는 것은 쉽지 않은 문제이다. 이러한 접근의 어려움 때문에 법체계는 산업 전반에 혼란을 야기하고 있다.

대부분 판결에서는 이용부분의 양적·질적인 유사도와 중요도를 검토하고 있으나 이는 전체적 유사성을 비교하고 있으며 중요도라는 상대적 개념은 불필요해 보인다. 저작권법에 공정이용 제도를 도입하여 적극적 항변요건이 존재함에도 저작권법상 침해를 판단하는 기준으로써 활용되며 생기는 문제이다.

따라서 필자는 비교법적 고찰을 통해 데이터 수집 및 분석에 관련 법체계에 대한 논의가 필요하다는 입장이다. 영국의 허용범위 및 규율영역에 대한 구체성은 차용할 수 있을 것이다. 데이터 수집의 경우라 하더라도 판매되거나 공중에 제공할 권리는 부여되지 않으며, 이용할 권리만을 규율하고 있다. 또한 TDM을 통해 도출한 결과물은 공유할 수 없도록 하여 데이터 수집 및 분석에 대한 행위의 허용범위를 명확화하고 있는 것으로 보인다. 독일은 데이터 수집 및 분석에 대한 자료와 사본에 대한 관리 규정을 더하고 있다. 이러한 경우 해당 자료를 이용할 수 있는 기한을 정하는 것이 선행 절차가 될 것이다. 하지만 도서관 및 기록보관소 등 영구적 보관시설에 영구보존은 가능하다. 보관 및 보상 의무와 원본의 삭제까지 규정하여 가장 구체적인 규범으로 평가받고 있다. 이처럼 데이터와 관련한 문제는 그 분야가 방대하여 경우별로 차이가 발생한다. 따라서 완벽한 해답을 찾을 수 없으며 상황에 따라 결론이 바뀔 수 있음을 고려할 필요가 있다 [31]. 우리나라 역시 급변하는 데이터의 영역에

질서를 정립하기 위해서는 보호객체에 대한 관점에서 행위규범에 대한 논의로 태도의 변화가 필요하다.

이러한 접근은 데이터 수집 프로그램을 활용에 대한 기업들의 소극적 태도를 개선하여 데이터 산업에 대한 투자를 이끌어 낼 수 있을 것이다[32]. 데이터 산업의 진보적 발전을 이루기 위해서는 법체계의 제정 및 개정시에 규율영역에 대한 구별적 개념 정립이 필요할 것이다. 하지만 부정법 상 부정경쟁 행위 및 저작권법상 데이터베이스 제작자의 권리 등 법률간의 보호영역의 충돌이 발생할 수 있다는 한계점이 존재한다.

본 논문은 교육부 및 한국연구재단의 BK21 4단계(사이버-물리공간 청정화 연구사업단)로 지원된 연구임(5199990314137)

## 참 고 문 헌

- [1] S. W. Son. (2022). [public opinion] Data transaction and distribution 'protection path' will be held..., The Korean economy. <https://www.hankyung.com/opinion/article/2022041975581>.
- [2] K. H. Park. (2019). A Study on the Public Interest of Collected Information. *Informatization Policy* 26(1), 25-45. DOI : 10.22693/NIAIP.2019.26.1.025
- [3] S. P. Yoon & H. Y. Kwon. (2021). Analysis of the Global Data Law & Policy and its Implications: Focusing on the cases of the United States, the United Kingdom, and the European Union. *Informatization Policy*, 28(2), 98-113. DOI : 10.22693/NIAIP.2021.28.2.098
- [4] Seoul High Court, 2016.12.15. sentencing 2015나2074198 judgement.

- [5] Supreme Court, 2017.8.24. sentencing 2017 다224395 judgement.
- [6] Supreme Court, 2022.5.12. sentencing 2021 도1533 judgement.
- [7] H. K. Kim. (2021). Suggestion to utilizing crawling methods by purpose of data collection, The Korea Contents Association, 19(2), 13-18. 1598-9437 (pISSN)
- [8] W. S. Lee & J. M. Shin & J. H. Lim, D. I. Kim & K. I. Cho. (2017). Subject oriented crawling method, Korea Information Processing Society, 24(2). 371-373. DOI:10.3745/PKIPS.y2017m11a371.
- [9] 竹添直樹. (2017). 島本多可子, 田所駿佑, 萩野貴拓, 川上桃子, “クローリングハック”, 翔泳社. ISBN : 9784798154732
- [10] S. Chakrabarti & M. van. den. Berg & B. Dom. (1999). Focused crawling: a new approach to topic-specific Web resource discovery, Computer Networks 31, 11-16, DOI : 10.1016/S1389-1286(00)00052-3
- [11] United States Congress Senate. (2011). The Power of Google: Serving Consumers Or Threatening Competition?, Committee on the Judiciary. Subcommittee on Antitrust, Competition Policy, and Consumer Rights.
- [12] T. H. Hwang.(2021). An Analysis of Web Crawling from a perspective of Competition Law, Journal of Korean Competition Law. 43, 174-208 DOI : 10.35770/jkcl.2021.43..174
- [13] Dinuca & Ciobanu. (2012). Web content mining, Annals of the University of Petroşani, Economics 12(1).
- [14] S. J. Oh, (2019) Limitations on Copyright for Data and Text Mining, Hongik Law Review, 20(2) DOI : 10.16960/jhkr.20.2.201906.455
- [15] C. H. Shin. (2019). Mass Digitization of Copyrighted Works in the Era of Artificial Intelligence (AI), SungKyunKwan Law Review 31(1). DOI : 10.17008/skklr.2019.31.1.005
- [16] S. J. OH.(2016). copyright law, pybook.
- [17] S. H. Bae Lee & M. S. Seo, I. G. Shin, M S. Lee & J. J. Lim & J. S. Kim & S Y. KIM.(2014). A Research on the Korean Database Industry Promotion Act, Kangwon Law Review, 43, 309-337 DOI : 10.18215/kwlr.2014.43..309
- [18] I. H. Lee & K H Kim. (2016). To Whom Does Big Data Belong?: Protection of Big Data Under Korean Copyright Law and Its Effects in the Public Sector. Journal of Korean Association for Regional Information Society. 19(4). 37-62 DOI : 10.22896/karis.2016.19.4.002
- [19] D. H. Gu.(2006). Copyright Protection of Databases in the EU and Korea. Seoul National University Law Review, 19(1) 111-155 DOI : 10.34122/jip.2006.06.1.1.111
- [20] D. H. Gu.(2006). Copyright Protection of Databases in the EU and Korea. Seoul National University Law Review, 19(1) 111-155 DOI : 10.34122/jip.2006.06.1.1.111
- [21] UK CDPA §29A <http://www.legislation.gov.uk/ukdsi/2014/9780111112755>, Oct. 2014>
- [22] B. G. Kim & H. C. Shin & C. W. Ahn.(2017). Copyright Limitations and Exceptions for ‘Big Data’ Analysis and ‘Data Mining’, Copyright Quarterly, .30(1) 29-61 UCI : G704-SER000010581.2017.30.1.001
- [23] S. Y. Cha.(2018). A study on the protection of big data under intellectual property law, Lawyers association journal. 67(2), 71 - 148 DOI : 10.17007/klaj.2018.67.2.002
- [24] H. J. Ahn.(2018). Text and Data Mining for Big Data Utilization and Artificial Intelligence Development, lawtimes , <https://m.lawtimes.co.kr/Content/Info?serial=149591&kind=CC02>,
- [25] 中山信弘.(2014).著作權法. 第2版, 有斐閣.
- [26] I. H. Lee & K H Kim. (2016). To Whom

Does Big Data Belong?: Protection of Big Data Under Korean Copyright Law and Its Effects in the Public Sector. Journal of Korean Association for Regional Information Society. 19(4). 37-62 DOI : 10.22896/karis.2016.19.4.002

- [27] S. H. Bang.(2012). Conflicts Issues of Scientific Data Exploitation and Copyright Protection in the Big Data Era, Hongik Law Review, 13(4). ISSN : 2765-5148
- [28] J.H. Reichman & Jonathan A. Franklin, Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Goods Uses of Information, 147 U. PA. L. Rev. 930 (1999); Law No. 9620 of 19 February 1998, on Copyright and Neighboring Rights, Consolidated with the Bill in Public Consultation since 14 June 2010 ([http://www.vgrass.de/wp-content/uploads/2010/07/Brazilian\\_Copyright\\_Bill\\_Consolidated\\_June\\_2010.pdf](http://www.vgrass.de/wp-content/uploads/2010/07/Brazilian_Copyright_Bill_Consolidated_June_2010.pdf)) (English translation); "reverse notice and takedown" - Jerome H. Reichman, Graeme Dinwoodie & Pamela Samuelson, A Reverse Notice and Takedown Regime to Enable Public Interest Uses of Technically Protected Copyrighted Works, 22 Berkeley Tech. L.J. 981, 983-85, 2007.
- [29] S. H. Bang.(2012). Conflicts Issues of Scientific Data Exploitation and Copyright Protection in the Big Data Era, Hongik Law Review, 13(4). ISSN : 2765-5148
- [30] S. H. Bang.(2012). Conflicts Issues of Scientific Data Exploitation and Copyright Protection in the Big Data Era, Hongik Law Review, 13(4). ISSN : 2765-5148
- [31] Korea Data Agency.(2022). Data Industry Trends Issue Brief.
- [32] W. S. Lim.(2020). A review on the copyright infringement types and standardfor infringement decision. Copyright Quarterly, 33(3) 111-151 DOI : 10.30582/kdps.2020.33.3.111

저 자 소 개



강백서(Baek-Seo Kang)

2022.2 경기대학교 지식재산학과 졸업  
2022.2-현재 중앙대학교 일반대학원 융합  
보안학과 석사과정  
<주관심분야> 지식재산 (특허, 저작권, 상표)



신민지(Min-Ji Shin)

2022.2 단국대학교 법과대학 법학과 졸업  
2022.2-현재 중앙대학교 일반대학원 융합  
보안학과 석사과정  
<주관심분야> 지식재산 (특허, 저작권, 상표)