

# 형사사법정보의 빅데이터 활용방안 연구: 구조화 범주화 관점으로\*

## A Study on the Use of Criminal Justice Information Big Data in terms of the Structuralization and Categorization

김미령 (Mi Ryung Kim)\*\*

노윤주 (Yoon Ju Roh)\*\*\*

김성훈 (Seonghun Kim)\*\*\*\*

### 초 록

4차 산업혁명시대를 맞아 데이터의 중요성은 심화되고 있으나, 개인정보보호 등의 문제로 데이터의 활용이 쉽지 않은 경우가 많이 있다. 형사사법정보는 범죄 예측 및 예방, 범죄수사 과학화, 양형합리화 등 다양한 활용가치가 예상됨에도 현재 개인정보보호와 형사사법정보 관련 법률적 해석 문제로 활용이 상당히 제한되고 있다. 본 연구는 형사사법정보의 구조화·범주화를 통해 '범죄데이터'로 전환하여 빅데이터로서 활용하도록 제안하였으며, '범죄데이터' 활용시 법률적 문제, 활용가치, 데이터 생성 및 활용시 고려사항을 전문가를 통해 검증하고 향후 전략적 발전방안을 도출하였다. 연구결과, '범죄데이터'는 개인정보보호문제는 해결된 것으로 보여지나, 형사사법정보 관련법에 명시할 필요는 있으며, 빅데이터 활용을 위해 분석 가능하도록 표준화된 형태로 정리되는 것이 시급함이 밝혀졌다. 향후 진행방향으로는 데이터 요소 도출, 용어사전 시소러스 구축, 데이터 등급화를 위한 개인민감정보 정의 및 등급지정, 비정형데이터의 정형화를 위한 알고리즘 개발 등을 제시하였다.

### ABSTRACT

In the era of the 4th Industrial Revolution, the importance of data is intensifying, but there are many cases where it is not easy to use data due to personal information protection. Although criminal justice information is expected to have various useful values such as crime prediction and prevention, scientific investigation of criminal investigations, and rationalization of sentencing, the use of criminal justice information is currently limited as a matter of legal interpretation related to privacy protection and criminal justice information. This study proposed to convert criminal justice information into 'crime data' and use it as big data through the structuralization and categorization of criminal justice information. And when using "crime data," legal issues, value in use, considerations for data generation and use were verified by experts, and future strategic development plans were identified. Finally we found that 'crime data' seems to have solved the privacy problem, but it is necessary to specify in the criminal justice information related law and it is urgent to be organized in a standardized form for analysis to use big data. Future directions are to derive data elements, construct a dictionary thesaurus, define and classify personal sensitive information for data grading, and develop algorithms for shaping unstructured data.

키워드: 형사사법정보, 빅데이터, 범죄데이터, 데이터 구조화·범주화, 형사사법정보 빅데이터 활용전략  
criminal justice information, big data, crime data, data structuralization and categorization,  
criminal justice information big data utilization strategy

\* 본 연구는 2019년 경찰청 '형사사법정보의 빅데이터 활용을 위한 법적·기술적 개선 방안 연구' 연구용역 내용을 수정·보완한 것임.

\*\* 서울지방경찰청 사서(mrkimhy@police.go.kr) (제1저자)

\*\*\* 경찰청 사서(ballen3163@police.go.kr) (공동저자)

\*\*\*\* 성균관대학교 문헌정보학과 초빙교수(godwmaw@g.skku.edu) (교신저자)

■ 논문접수일자: 2019년 11월 17일 ■ 최초심사일자: 2019년 12월 17일 ■ 게재확정일자: 2019년 12월 23일  
■ 정보관리학회지, 36(4), 253-277, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.4.253>

## 1. 서론

### 1.1 연구배경 및 필요성

4차 산업혁명의 핵심요소로 꼽히는 인공지능(AI), 머신러닝(machine learning), 딥러닝(deep learning), 빅데이터(big data), BI(business intelligence)는 인간만이 가능했던 고도의 사고체계를 정밀한 알고리즘 기반의 기술을 통해 구현한다. 또한 고성능 컴퓨팅으로 데이터를 처리하는 기술이 뒷받침되어 공공 및 민간기관들의 의사결정에 반영되는 상호연계체계를 가지고 있다. Gartner는 그 핵심요소로 데이터의 중요성을 강조해왔으며, '데이터는 21세기의 원유다'라고 주장해 왔다. 급속하게 증가하는 데이터를 얼마나 잘 처리하고, 분석하고, 활용하느냐가 기업은 물론, 국가의 경쟁력을 좌우하는 핵심요소임에는 이견이 없으며, 4차 산업혁명의 심화로 데이터의 가치와 중요성은 더욱 높아지고 있다.

데이터의 가치와 중요성이 커져 가면서 데이터 공개에 대한 세계적인 흐름이 자명하며, 세계 주요 선진국 정부의 데이터 공개에 대한 인식은 급속도로 확산되었다. 이러한 흐름에 대응하여 우리나라 정부도 지속적으로 정부 시스템의 전자화를 추진함으로써 상당한 데이터를 확보하고 있으며, 정부 3.0, 공공데이터 포털(<https://www.data.go.kr/>) 구축 등 데이터 공개에 많은 관심을 기울이고 있다. 하지만, 세계적인 데이터 공개의 분위기와 그 파급효과에도 불구하고 국민의 세금으로 생

산된 데이터가 여전히 상당수 공개되지 않고 있음은 국내뿐 아니라 세계적으로도 이슈를 양산하고 있다.

형사사법기관은 「전자정부법」에 따라 기관별로 담당해 오던 전자화를 형사사법기관 간 협력하에 국민 편익 증진, 형사절차 투명성 제고, 업무 효율성 제고 및 예산 절감을 목적으로 형사사법기관 간 연계시스템인 KICS(Korea Information System of Criminal Justice Services, 이하 KICS)를 2010년 7월 전면 개통하였다. 이 시스템을 통해 기관 간 또는 부서·개인 간 종이 서류 전달에 걸리는 시간, 동일 정보 중복 입력에 걸리는 시간 등 오프라인 업무처리에 따른 비효율성이 대거 해소되었다. 또한 연간 290억 원의 예산 절감을 예상하였고,<sup>1)</sup> 2011년부터 2018년 기준으로 1,413만 건<sup>2)</sup> 이상의 형사사건에 대해 형사사법기관인 법원, 법무부, 검찰, 경찰, 해양경찰이 생산한 데이터들이 소위 빅데이터라고 지칭될 만큼 실시간으로 대량 생산되고 있다. 현재 데이터 활용에 대한 패러다임이 변화하면서 내·외부 관련 기관에서 형사사법정보의 데이터를 요구하지만 형사사법처리업무를 제외하고는 공개적으로 이용되지 못하고 있는 실정이다.

형사사법정보가 빅데이터로서 활용되지 못하는 이유는 기본적으로 KICS와 함께 2009년 발의되어 형사사법정보의 기본법 역할을 하고 있는 「형사사법절차 전자화 촉진법」의 강력한 법률 제약에 기인한다. 동법 제6조 제3항은 “형사사법업무 처리기관은 형사사법업무 처리 외의

1) 법무부 형사사법통합정보체계추진단. (2009.11.) 형사사법절차 전자화 촉진법(안) 등 설명자료.

2) 경찰청 통계. Retrieved from

[http://stat.kosis.kr/nsieu/view/tree.do?task=branchView&id=132\\_13204\\*MT\\_OTITLE&hOrg=132](http://stat.kosis.kr/nsieu/view/tree.do?task=branchView&id=132_13204*MT_OTITLE&hOrg=132)

목적으로 형사사법정보를 수집·저장 또는 이용할 수 없다”고 규정하고 있다. 이에 따라 형사사법기관 내·외부에서 형사사법정보의 데이터 조회가 업무 권한에 따라 제한적이거나 혹은 원천적으로 봉쇄되고 있으며, 빅데이터로서의 그 활용가치도 쉽게 파악할 수 없는 실정이다. KICS에서 관리하는 형사사법정보 내에는 고유 식별정보와 민감정보(범죄경력) 등 개인정보가 다수 포함되어 있어 공개하거나 목적 외 사용에 제약이 따르는 특수성이 있으므로 사용을 제약하는 법안은 타당하다고 할 수 있다. 하지만, 형사사법정보가 범죄예방을 비롯하여 과학적 수사 등에 활용되어 공익에 기여할 가능성이 예상됨에도 불구하고 개인정보의 비식별 처리 후 사용이나 개인정보와 상관없는 사용에 대한 조항이 전혀 없는 점은 형사사법정보의 빅데이터 활용에 있어 큰 문제로 지적되고 있다(나영민, 2016).

범죄 수법이 점점 지능화되고, 성범죄, 연쇄살인 등 사회불안요소가 증가함에도 불구하고 과학화·고도화 된 범죄예측과 범인검거에 기초가 되는 데이터를 활용할 수 없는 것은 현장과 학계에서 큰 문제점으로 지적되고 있다. 박민우(2017)는 빅데이터와 AI를 활용한 범죄예방은 시대적 흐름이며 범죄예방을 위해서는 이미 일어난 범죄 즉, 형사사법정보만큼 좋은 기반 데이터는 없음에도 활용되지 못하는 문제를 해결해야 함을 주장하였고, 이연욱(2019)은 경찰 내·외부에서 형사사법정보의 활용을 위해 데이터를 요구하지만 법적인 제약으로 인해 연구목적으로도 활용할 수 없기에 장기적인 관점

에서 문제해결을 위한 노력이 필요함을 주장하였다. 우종필(2019)은 “미국 FBI는 데이터와 범인예측 알고리즘을 공개하며 더 좋은 알고리즘을 내외부에서 개발하도록 자신감 있게 권장하는 반면, 우리나라는 범피데이터에 접근조차 할 수 없다”며 국제적 흐름과 거리감이 있는 국내 현실을 지적했다.

형사사법정보의 빅데이터 활용을 위한 해결책은 주로 법적인 개정을 통해 이뤄지도록 연구되어 왔으며(김기범, 2016; 김한균, 2018; 나영민, 2016; 박민우, 2017), 빅데이터로서의 활용가치도 연구되어 왔다(양종모, 2016; 권양섭, 2017; 김한균, 2018). 이처럼 4차 산업혁명 시대의 국가경쟁력 제고를 위해 형사사법정보의 활용에 대한 관심이 커지고 있다. 또한 데이터 산업 경쟁력 강화와 AI산업 활성화를 위해 데이터 3법(개인정보보호법·정보통신망법·신용정보법) 개정안에 관한 논의가 진행되고 있고, 이 중 개인정보보호법 개정안은 행정안전위원회 법안심사소위원회에서 의결했다. 이러한 시점에 기술적 한계를 극복하고 형사사법정보를 활용하기 위한 방안 연구가 필요하다고 보인다.

이에 본 연구는 형사사법기관이 형사사법처리를 위해 작성하여 관리하는 형사사법정보의 속성을 현재의 속성과 전혀 다르게 전환한 ‘범죄데이터’의 빅데이터 활용 가능성을 모색하고자 한다. 현재 형사사법정보가 전혀 사용되지 못하는 것은 아니며, 형사사법정보를 활용하여 범죄 관련 통계자료가 공개되어 활용되고 있다.<sup>3)</sup> 이는 형사사법정보가 다른 형태로 전환되어 활용되는 예라고 할 수 있다. 정보학적 관점에서 형사사법

3) 경찰범죄통계. Retrieved from <https://www.police.go.kr/portal/main/contents.do?menuNo=200529>

정보를 고찰하고, 형사사법정보의 근본적인 속성을 변경하여 법적인 제약에서 벗어나 활용방안을 모색하는 연구는 전무한 실정이다. 이에 형사사법정보의 근본적 속성을 전환할 경우, 법적 제약, 빅데이터로서의 활용가치, 고려사항에 관한 연구가 필요하다.

## 1.2 연구 목적 및 방법

본 연구는 형사사법정보의 구조화·범주화를 통한 빅데이터 활용방안을 모색하는 것을 목적으로 한다. 현재 형사사법정보는 형사사법처리기관의 제한된 인원만이 접근권을 가지기 때문에 실제로 형사사법정보를 구조화·범주화 하여 연구를 진행하기에는 제한사항이 많은 상황이다. 해외 형사사법정보의 구조화·범주화 사례를 통해 구조화·범주화된 형사사법정보인 '범죄데이터'를 제안하였으며, 이 '범죄데이터'의 법적 제약, 빅데이터로서의 활용가치, 고려사항에 관해 전문가의 의견을 수렴하여 분석하였다. 본 연구 목적에 따른 연구문제는 다음과 같다.

- RQ 1. 형사사법정보의 구조화·범주화를 통해 생성된 '범죄데이터'는 형사사법절차 전자화 촉진법, 개인정보보호법 등에 저촉

되지 않고 사용가능한 데이터인가?

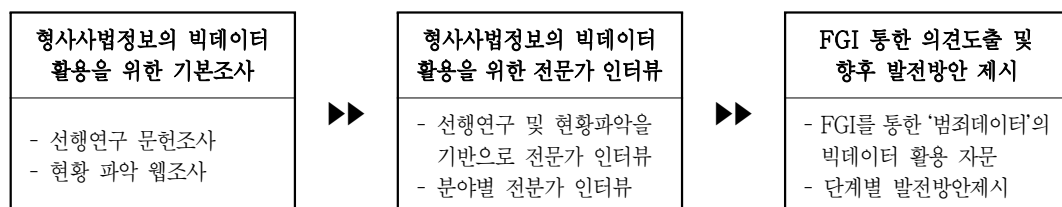
- RQ 2. 형사사법정보의 구조화·범주화를 통해 생성된 '범죄데이터'는 빅데이터로서 활용가치가 있는가?
- RQ 3. 형사사법정보의 구조화·범주화를 통해 생성된 '범죄데이터'의 생성 및 활용에 있어 고려사항은 무엇인가?

본 연구문제를 해결하기 위한 연구방법은 다음 〈그림 1〉과 같다.

현재까지 형사사법정보의 활용과 관련한 문헌을 조사하고, 형사사법정보를 활용한 해외사례를 조사하였다. 형사사법정보의 구조화·범주화를 통해 근본적 속성을 전환할 경우에 법적 제약 측면, 빅데이터로서의 활용가치 측면, 데이터 생성과 활용 시 추가적으로 고려할 사항에 대해 전문가 인터뷰 및 FGI를 실시하였다. 그리고 그 결과를 분석하여 단계별 발전전략을 제시하였다.

## 1.3 선행연구

형사사법정보의 빅데이터 활용을 위한 법적 개정에 관한 연구는 다음과 같다. 김기범(2016)은 형사사법정보의 수집목적보다 명확하게 규정하고, 범죄예방, 통계작성 및 학술연구 분야



〈그림 1〉 연구모델과 방법 도식화

에 한해 형사사법정보를 비식별화·익명화하여 목적 외 이용·제공할 수 있도록 입법방향을 제시하였다. 나영민(2016)은 형사사법정보와 개인정보 보호에 관한 이론적인 고찰, 관련제도의 현황과 문제점 고찰, 형사사법정보의 처리단계별 주요 쟁점들을 검토하여 형사사법정보의 이용과 보호에 대한 법률적 해결을 시도하였다. 박민우(2017)는 형사절차전자화법 제6조 제3항의 해석을 선언적 의미로 보고 부당한 목적 여부를 구체적 근거 법령의 존재, 「개인정보 보호법」과의 조화 등을 판단 기준으로 삼아 활용할 수 있음을 제안하였다.

형사사법정보의 빅데이터 활용 가치에 관한 연구는 다음과 같다. 양종모(2016)는 형사사법 절차에서 생성된 전자사법정보가 축적된 대량의 빅데이터를 활용·분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 범죄 예측과 능동적 대응, 그리고 양형 합리화 방안 모색과 같은 형사사법에 활용할 필요성이 그 어느 때보다도 커졌다고 주장하였다. 그 결과 형사사법정보 빅데이터를 활용할 경우 1) 범죄예측시스템, 2) 법률전문가 시스템, 전자법정, 전자조사의 구현, 3) 법관의 통상적 업무처리 과정에 적용할 경우 양형 데이터 자동 작성 및 분류, 저장될 수 있는 기반을 갖추 수 있다. 또한 전자사법정보를 데이터 마이닝 등 통계적 기법으로 양형과정에 적절히 활용하면 공정성 확보는 물론 투명성도 높여줄 것이며, 형벌의 범죄억제효과도 거양할 수 있을 것으로 예상하였다. 권양섭(2017)은 형사사법 안에서 빅데이터를 활용하기 위한 방안을 연구하였다. 이를 위해 국내외 빅데이터의 활용실태를 조사하여 빅데이터 수집과 분석이 갖는 법적 성격을 규명하였으며,

빅데이터 기술이 갖는 법률적 의미를 도출하여 기존의 형사사법 이론과 체계를 토대로 헌법적 가치를 보호할 수 있는 범위 내에서 활용방안을 제시하였다. 김한균(2018)은 형사사법정보의 빅데이터 활용을 위한 법적 문제와 정책적 활용가능성에 대한 방향을 정립하고자 하였다. 그 결과 빅데이터기반 형사정책 맥락에서 형사사법정보의 핵심요소인 형사사법 업무처리 관련성, 전자적 방식 처리로서의 시스템 특성, 개인정보가 포함된 정보자료의 통제필요성 등 세 가지 요소에 따라 각각 정당성과 효과성을 검증함으로써 빅데이터기반 형사정책 실현에 형사사법정보의 활용이 가능해 질 것으로 보았다.

선행연구 검토 결과, 형사사법정보의 기본법 역할을 하고 있는 '형사사법절차 전자화 촉진법'의 문제점 고찰과 그 개정을 통한 활용방안이 연구되고 있을 뿐, 속성전환을 통한 사용가능성이나 활용가치, 빅데이터 활용을 위한 고려사항 및 추진전략에 대한 연구가 부재하였다.

## 2. 이론적 배경

### 2.1 형사사법정보 및 형사사법정보시스템(KICS)

형사사법정보시스템(KICS)은 국가 차원에서 형사사법정보를 체계적으로 관리·공유하고 공개하는 세계 유일의 시스템이다(법무부 형사사법공통시스템 운영단, 2014). 4개 형사사법기관 즉, 법원, 법무부, 검찰, 경찰이 형사사법정보를 작성, 취득, 저장, 송수신 하는데 이용할 수 있도록 구축한 전자적 관리체계다. 각 기관의 형사

사법정보시스템은 하드웨어, 소프트웨어, 네트워크, 보안체제로 구성되고, 형사사법기관들이 공동 사용하는 법무부 형사사법정보공통시스템까지 아울러 형사사법정보시스템을 구성한다(김한균, 2018). 형사사법 절차에 IT기술을 이용, 형사사법기관의 문서 작성을 전자화하고 형사사법정보를 공동 활용하도록 만들어졌으며, 수사기관뿐만 아니라 이동통신사, 출입국관리사무소, 행정자치부 등 외부기관과도 연계되어 자료 정보 활용이 가능하다. 한 기관이 자신의 시스템을 자동화하는 기관별 전자정부 시스템과 달리 KICS는 4개 형사사법기관 간 협력과 소통을 통해 탄생했다.<sup>4)</sup>

이 시스템은 국민들이 자신의 사건정보를 실시간으로 확인해 볼 수 있는 형사사법포털, 음주·무면허운전 사건을 신속하게 처리하기 위

해 종이기록 없이 완전 전자화하여 처리하는 전자약식시스템, 4개 형사사법기관이 서로의 형사사건 정보를 업무적으로 주고받는 정보 공동 활용 등 크게 세 가지 서비스를 제공하고 있다(법무부 형사사법공통시스템 운영단, 2014). 2004년 초부터 시작된 프로젝트는 2010년 7월이 되어야야 마무리 되었으며, 시스템 개통 이래 2011년부터 2018년 기준으로 약 1,413만 건<sup>5)</sup>의 수사결과보고서, 형사사건에 대한 의견서, 현행범 인체포서, 확인서 등 형사사법정보들이 축적되고 있다.

현재 4개 기관은 KICS전용 서버를 설치·운영 중이며, 공통시스템 서버 운영과 기관 관 협의 업무는 법무부의 '형사사법 공통시스템 운영단(이하 형통단)'에서 수행하고 있다.<sup>6)</sup> 형사사법정보시스템의 연혁은 <표 1><sup>7)</sup>과 같다.

<표 1> KICS 연혁

날짜	내용
2004.4.	대검찰청 전자정부 과제 '검찰정보시스템 고도화를 통하여 형사법기관 간 정보공동활용체계 강화'에서 시작
2004.8.	정부혁신위원회에서 전자정부 로드맵 31대 과제로 '형사사법 통합정보체계 구축' 사업 선정
2004.12.3.	대통령 훈령(제135호)에 의거한 '통합형사사법체계구축기획단' 발족
2006.2.28.	형사사법 기관인 법원, 법무부, 검찰, 경찰이 합의를 거쳐 형사사법통합정보체계추진단으로 출발
2007.9.	1차 합의서에 법원, 법무부, 검찰, 경찰 등 4개 형사사법기관 대표 서명
2008.12.29.	2차 최종 합의서 서명 - 각 시스템은 4개 형사사법기관에서 독립적으로 운영·관리 - 공통시스템은 법무부 운영단에서 운영·관리 합의
2010.7.	형사사법정보시스템(KICS) 전면 개통
2010.10.18.	기획조정실 하에 '형사사법공통시스템운영단' 신설

4) 해외문화홍보원 (2017). Retrieved from <http://www.kocis.go.kr/koreanet/view.do?seq=7645>

5) 경찰청 통계. Retrieved from

[http://stat.kosis.kr/nsieu/view/tree.do?task=branchView&id=132\\_132\\_0\\_4\\*MT\\_OTITLE&hOrg=132](http://stat.kosis.kr/nsieu/view/tree.do?task=branchView&id=132_132_0_4*MT_OTITLE&hOrg=132)

6) 해외문화홍보원 (2017). Retrieved from <http://www.kocis.go.kr/koreanet/view.do?seq=7645>

7) 법무부 형사사법 공통시스템 운영단(국민을 위해 형통하라, 2014); 나영민, 박광민(형사사법절차 전자화의 쟁점과 그 해결방안, 2014) 자료를 재정리했음.

형사사법정보의 정의는 형사사법정보 전자화 촉진법<sup>8)</sup> 제2조(정의) 3호에 다음과 같이 규정하고 있다.

3. “형사사법정보”란 형사사법업무 처리기관이 형사사법업무 처리와 관련하여 형사사법정보 시스템을 이용하여 작성하거나 취득하여 관리하고 있는 자료로서 전자적 방식으로 처리되어 부호, 문자, 음성, 음향 또는 영상 등으로 표현된 것을 말한다.

‘형사사법업무’란 수사, 공소, 공판, 재판의 집행 등 형사사건의 처리와 관련된 업무를 의미하며, ‘형사사법업무 처리기관’이란 법원, 법무부,

검찰청, 경찰청, 해양경찰청 및 그 소속 기관과 그밖에 형사사법업무를 처리하는 기관으로서 대통령령으로 정하는 기관을 의미한다. 또한 “형사사법정보시스템”이란 형사사법업무 처리기관이 형사사법정보를 작성, 취득, 저장, 송·수신하는데 이용할 수 있도록 하드웨어, 소프트웨어, 데이터베이스, 네트워크, 보안요소 등을 결합시켜 구축한 전자적 관리체계를 의미하는데 구체적으로는 형사사법정보시스템(KICS)<sup>9)</sup>을 의미한다.

실제로 KICS를 이용해 작성하고 관리되는 형사사법정보의 서식은 <그림 2>, <그림 3>과 같으며, 개인정보는 \*기호로 비식별화 처리하였다.

**의 건 서**

1. 피의자 인적사항  
 이: \*\* (유점공(유점주식회사) 72세)  
 주민등록번호: 460914-2 \*\*\*\*\*  
 주 기: 서울 종로구 북촌로 \*\*\*\*\*  
 등록기준지: 서울 종로구 낙산길  
 전 화 번 호:

2. 범죄경력자료 및 수사경력자료  
 조회 중

3. 범죄사실  
 1. 범행 결의 및 범행 도구 준비  
 피의자는 울산 동구 (주소 1 생략)에 있는 공소의 1 주식회사에서 유점공으로 일하는 사람이다.  
 피의자는 2009. 7.경부터 피해자 오 \*\* (여, 27세)와 사귀어 왔는데, 동인은 2012. 7. 12.경 피고인에게 문자메시지를 보내어 백화점 정원 일을 하는 자신의 근무시간과 업무 스트레스 때문에 더 이상 피고인과 교제하기 힘들다며 결별 통보를 하였다. 피고인은 이에 피해자에게 계속 문자메시지를 보내어 다시 만나 줄 것을 요구하였으나 피해자가 이를 들어주지 않자, 2012. 7. 19.경 피해자를 살해하고 추가로 피고인과의 교제를 반대해 온 피해자의 동생 최 \*\* (여, 23세)도 살해하기로 마음먹었다.  
 피의자는 2012. 7. 19. 오전경 스마트폰을 이용하여 ‘불 붙는 기름’, ‘주방용 칼 파는 곳’, ‘울산 송 구할 수 있는 곳’을 검색하여 살해 도구를 구입할 수 있는

<그림 2> 의견서 예시

**경 찰 청**

제 2019-00006 호 2019. 7. 3.  
 수 신: 경찰청장  
 참 조: 수사기획과장  
 제 목: 수사결과보고

살인 피의사건에 관하여 다음과 같이 수사하였기에 결과 보고합니다.

1. 피의자 인적사항  
 이: \*\* (유점공(유점주식회사) 72세)  
 주민등록번호: 460914-2 \*\*\*\*\*  
 주 기: 서울 종로구 북촌로 \*\*\*\*\*  
 등록기준지: 서울 종로구 낙산길

2. 범죄경력자료 및 수사경력자료  
 조회 중

3. 범죄사실  
 1. 범행 결의 및 범행 도구 준비  
 피의자는 울산 동구 (주소 1 생략)에 있는 공소의 1 주식회사에서 유점공으로 일하는 사람이다.  
 피의자는 2009. 7.경부터 피해자 오 \*\* (여, 27세)와 사귀어 왔는데, 동인은 2012. 7. 12.경 피고인에게 문자메시지를 보내어 백화점 정원 일을 하는 자신의 근무시간과 업무 스트레스 때문에 더 이상 피고인과 교제하기 힘들다며 결별 통보를 하였다. 피고인은 이에 피해자에게 계속 문자메시지를 보내어 다시 만나 줄 것을 요구하였으나 피해자가 이를 들어주지 않자, 2012. 7. 19.경 피해자

<그림 3> 보고서 예시

- 8) 법제처, 국가법령정보센터, 형사사법절차 전자화 촉진법, Retrieved from <http://www.law.go.kr/%EB%B2%95%EB%A0%B9/%ED%98%95%EC%82%AC%EC%82%AC%EB%B2%95%EC%A0%88%EC%B0%A8%EC%A0%84%EC%9E%90%ED%99%94%EC%B4%89%EC%A7%84%EB%B2%95>
- 9) 형사사법정보시스템, Retrieved from <http://www.kics.go.kr/>

구체적인 내용 정보는 다음 <그림 4>에서 보는 것과 같이 피고인의 성장 배경, 범행 정보, 범행 현장 정황 등 형사사법처리를 위한 구체적인 개인정보와 민감한 정황 정보 등을 상당히 많이 포함하는 특징을 가진다. 피고인과 마찬가지로 피해자에 대한 정보 또한 상당히 포함하고 있다.

따라서 형사사법정보 전자화 촉진법 제2조(정의) 3호에서 정의하는 형사사법정보의 법적 정의를 정리하면, 법원, 법무부, 검찰, 경찰, 해양경찰 및 소속기관이 수사, 공소, 공판, 재판의 집행 등 형사사건의 처리와 관련하여 형사사법정보시스템을 이용하여 작성하거나 취득하여 관리하는 자료로서, 피의자의 성장 배경, 범행 정보, 범행 현장 정황 및 피해자의 개인정보 등 민감한 정보를 담고 있으며 전자적 방식으로 처리되어 부호, 문자, 음성, 음향 또는 영상 등으로 표현된 것을 말한다.

범죄 양상이 갈수록 다양화·지능화되어감에

따라 국내외 경찰기관은 지능적인 치안서비스를 위해 그동안 축적해온 데이터를 어떻게 활용할 것인지에 대해 많이 고민해 왔다. 해외의 경우 범죄 사건과 범죄자 정보 등을 기반으로 범죄예방을 위한 예측 시스템을 구축하여 치안 업무에 활용하고 있다. 이에 형사사법정보를 활용한 미국, 영국, 일본 등 해외 주요국의 사례를 살펴보면 다음 <표 2>와 같다.

<표 2>와 같이 형사사법정보를 통해 국민의 안전과 관련해 공익신장이 기대되지만, 국내에서는 KICS와 함께 발의된 「형사사법절차 전자화 촉진법」 제6조 제3항은 “형사 사법업무 처리기관은 형사사법업무 처리 외의 목적으로 형사사법정보를 수집·저장 또는 이용할 수 없다”는 내용으로 인해 활용에 문제를 겪고 있다. 개인정보보호법 제18조 제2항 제4호에 의하면 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인

피고인은 초등학교 6학년이 되었을 때부터 천안에서 여동생, 어머니와 함께 살게 되었으나 여전히 가정형편이 어려워 자주 이사를 다녔고, 중학교 3학년 때는 부산으로 이사하였다. 이곳에서의 학업성적 역시 하위권을 맴돌았으며 학교생활에 제대로 적응하지 못하였으나, 특별히 어떠한 문제행동으로 징계를 받은 일은 없었다.

#### 피고인의 성장배경

피고인은 2012. 7. 15.경 피해자 공소의 2의 집에 찾아가 동인과 한 시간 정도 대화를 나누었는데, 위 피해자가 무시하는 말투로 욕설을 섞어 피고인 부모님의 이혼, 피고인의 직업과 장래, 피고인의 경제력, 피고인의 집에 돈이 없어 결혼하지는 남자가 생기자 여동생을 일찍 결혼시킨 일 등에 대해 이야기하자, 마음이 크게 상하여 위 피해자와 헤어지는 것에 별 말 없이 동의하였다.

#### 범행동기

피고인이 배란다에서 거실로 통하는 문을 열자, 피해자 공소의 3이 그 소리에 깨어나 피고인을 보고 “이리로 왜 들어오느냐?”라는 등의 말을 하였는데, 자신은 위 피해자가 거실에서 잠을 자고 있을 줄은 몰랐기에 너무나 놀라 순간적으로 위 피해자의 목을 식칼로 찌른 후 다시 빼려고 했으나, 위 피해자가 손으로 위 식칼을 잡는 바람에 이를 빼려 힘을 주다 다시 한 번 더 찌르게 된 것이다.

#### 범행현장 상세진술

<그림 4> 형사사법정보의 범죄사실



〈표 2〉 형사사법정보 활용 해외 사례

국가	시스템명	설 명
미국	애리조나 투산경찰서 캡링크 (COPLINK)	<ul style="list-style-type: none"> <li>• 1997년 투산경찰서(Tucson Police Department)와 공동으로 애리조나 대학 인공지능연구소에 의해 처음 개발, 2011년에 IBM사에서 인수<sup>10)</sup></li> <li>• ‘인 캡링크 온 클라우드 솔루션(인 COPLINK on Cloud)’ 명칭으로 IBM에 의해 개발되었으며, 공간지리 정보 매핑 기능은 방화, 주거 침입절도, 빈집털이 등 특정 날짜나 시간, 장소에 따른 범죄유형을 지도상에 표시(M. Adan, 2016)</li> <li>• 자동으로 범죄 간 특정한 연계성을 찾을 수 있도록 지원하며, 정형 및 반정형 데이터를 활용하여 사람, 위치, 특징 등 다양한 요소 추출 가능<sup>11)</sup></li> </ul>
	산타크루즈 경찰서의 프레드폴 (PredPol)	<ul style="list-style-type: none"> <li>• 캘리포니아의 산타크루즈 경찰서(SCPD: Santa Cruz Police Department)는 LA 경찰청(LAPD)과 함께 미국에서 최초로 예측적 경찰 활동 시행, 기계학습 알고리즘을 이용하여 특정 범죄가 발생할 가능성이 가장 높은 시간과 장소(Hot Spot)를 예측한 후, 해당 지역을 순찰하여 범죄를 예방<sup>12)</sup></li> </ul>
	뉴욕 경찰청 (NYPD)의 DAS (Domain Awareness System)	<ul style="list-style-type: none"> <li>• 실시간 감시와 대응 체제를 구축한 시스템이며, 활용되는 빅데이터는 NYPD와 CIA, FBI 등이 보유한 데이터베이스(류연수, 2017), 뉴욕시의 공공 CCTV, 차량번호, 범죄경력, 911 신고접수 현황 등을 토대로 의심스러운 사람이나 차량, 물건의 위치를 추적하고 이동 경로 확인 가능</li> <li>• 범죄용의자의 생년월일과 전화번호, 주소, 신용카드나 금융거래 내역, 차량의 종류와 색상 및 번호뿐만 아니라, 신용정보, 판결기록 등의 빅데이터를 분석해서 용의자의 동선이나 생활패턴 등 다양한 정보 파악(Andrew Guthrie Ferguson, 2015)</li> </ul>
	루이지애나 Shreveport 경찰서의 파일럿 (PILOT)	<ul style="list-style-type: none"> <li>• 강도, 침입절도, 차량털이, 거리 및 차량절도를 “전술적 개입대상 범죄(tactical crime)”로 규정하고, 이들을 예측하기 위해 911 신고전화, 경범죄·청소년 검거기록, 계절적 지표 등 활용하여 각 구역 담당자들은 순찰차 두 대(각 2인 탑승)와 경사 한 명을 핫스팟 전담팀으로 구성해서 집중 순찰<sup>13)</sup></li> </ul>
	미네소타주 로체스터시의 범죄정보 분석 플랫폼	<ul style="list-style-type: none"> <li>• 빅데이터 기반의 범죄정보 분석 플랫폼 구축하였으며, 로체스터시 경찰이 보유한 모든 데이터베이스에서 사람, 장소, 휴대폰, 전화 기록, 차량 기록 간의 연관성을 분석해 수사 단서를 신속하게 찾는 시스템(안종욱, 이미숙, 신동빈, 2013)</li> </ul>
영국	웨스트미들랜즈주 경찰 국가데이터 분석시스템 NDAS	<ul style="list-style-type: none"> <li>• National Data Analytics Solution는 인공지능 딥러닝과 통계를 조합(활용)하여 범죄 예측 시스템을 구축하기 위한 프로젝트를 수행 중이며, 웨스트미들랜즈 경찰이 개발의 주축이 되어 런던, 맨체스터 등 8개 지역 경찰이 참여하였으며, 테스트 단계</li> <li>• 나이, 범죄나 구금 기록, 사건 첩보, 경찰 출동기록 등의 데이터를 분석하였으며, 그 결과 범죄 예측에 활용할 수 있는 1,400개 지표 발견<sup>14)</sup></li> </ul>
	런던 경찰청 (MPS, Metropolitan PoliceService)	<ul style="list-style-type: none"> <li>• 2014년 MPS와 Accenture Plc(NYSE: ACN)는 범죄조직원 예방 프로그램(Complete Analytics Pilot Program to Fight Gang Crime)을 구축하기 위해 20주간의 파일럿 연구 진행</li> <li>• 이 프로그램은 과거 5년간 범죄전력이 있는 범죄조직원의 범죄기록, 소셜미디어 게시물 분석 및 범죄 인텔리전스 시스템의 데이터를 병합하여 범죄자의 범행 가능성을 사전 예측<sup>15)</sup></li> </ul>
	ACE (Applied Cognitive Engine)	<ul style="list-style-type: none"> <li>• ‘중대사기수사국(SFO: Serious Fraud Office)’은 영국 인공지능 스타트업 ‘레이븐 시스템즈(Ravn Systems)’의 인공지능 솔루션 ‘ACE’ 도입</li> <li>• 인공지능 소프트웨어로 사람보다 1천만 배 빠른 속도로 많은 양의 비정형 데이터를 검색, 읽기, 해석 및 요약 가능</li> <li>• 특히 인공지능 기술은 사기나 부패 사건에 관한 기록 검토 작업을 위해 워드, 엑셀, 이메일 텍스트, 파워포인트, 이미지 등 다양한 유형의 문서 데이터 색인화를 함으로써 분류 및 관리 가능<sup>16)</sup></li> </ul>
일본	히타치 (Hitachi) PCA (Predictive Crime Analytics)	<ul style="list-style-type: none"> <li>• 웹 피드 같은 변수를 종합하여 범죄발생 가능성이 높은 장소와 시간을 예측하기 위해 범죄기록, 트위터, CCTV 카메라, 총소리탐지기, 교통시스템, 통신데이터, 실시간 기상데이터 등을 이용하여 데이터를 수집한 후, 범죄데이터와 연계시켜 범죄 발생 가능성이 가장 높은 지역을 보여주는 이른바 ‘히트맵(heat map)’을 생성하여 직관적으로 확인이 가능</li> <li>• 실시간으로 공공 및 민간데이터, 소셜미디어 등 다양한 데이터의 자연어 처리를 통해 범죄를 예측하는 최초의 솔루션<sup>17)</sup></li> </ul>

10) Eller College of Management, Retrieved from <https://eller.arizona.edu/departments-research/centers-labs/artificial-intelligence/research/projects/previous/copl原因>

11) IBM, Retrieved from <https://www.ibm.com/security/resources/demos/copl原因-demo/>

12) PREDPOL, Retrieved from <https://www.predpol.com/about/>

13) 장상진 (2012.8.10). 뉴욕시, 마이크로소프트, 최첨단 범죄감시시스템 공개, 조선닷컴, Retrieved from [http://news.chosun.com/site/data/html\\_dir/2012/08/10/2012081000252.html](http://news.chosun.com/site/data/html_dir/2012/08/10/2012081000252.html)

정보를 제공하는 경우”에는 개인정보의 목적 외 이용이 가능하나, 형사사법정보 영역에 직접 관련된 법률 제한으로 법적 해석이 필요한 상황이며, 활용에 많은 제약이 되고 있다. 그러나 형사사법정보를 활용하여 경찰통계, 범죄통계, 교통사고 현황 등 자료가 공개되어 활용되는 점을 고려할 때, 형사사법정보를 기준에 의해 가공함으로써 법률적 문제와 개인정보누출 등의 문제점들을 해결하여 활용가능성을 조사하는 것이 필요하다.

## 2.2 데이터 및 정보의 구조화범주화

데이터라는 용어는 라틴어 dare(‘give’의 의미)의 과거분사형으로 ‘주어진 것’이라는 의미를 가지고 있으며, 문헌에서는 1646년 영국에서 처음 사용되었다(한국데이터베이스 진흥원, 2014). 옥스퍼드 대사전에서는 데이터를 ‘추론과 추정의 근거를 이루는 사실(A thing given or granted: something known or assumed as fact, and

made the basis of reasoning or calculation: an assumption of premiss from which inferences are drawn, OED, Vol IV, 264)’로 정의하여 ‘객관적 사실’이라는 의미와 함께 ‘추론·예측·전망·추정을 위한 근거’라는 의미를 내포하는 것으로 보고 있다.

Borgman(2015)은 다양한 영역에서 데이터에 대한 높은 관심이 있음에도 불구하고 데이터에 대한 용어 정의가 불명확하게 정보와 혼용되어 사용되는 점과 정의를 명확하게 제시한 연구도 찾기 어렵다는 점을 지적하고 있다. 그러나 Rowley(2007)가 DIKW(Data Information Knowledge Wisdom)모델을 제시하며 데이터와 정보를 개략적으로 구분한 것은 참고할 만하다. Rowley에 따르면, 정보란 데이터를 기반으로 구체적 상황에 대해 시의 적절하게 전달되는 것을 의미한다. 즉, 데이터가 수많은 측정 및 관찰의 결과를 기록한 것이라면, 정보는 데이터에서 특정 의미를 추출하여 상황과 시기에 적절하게 보고한다는 의미이다. 이렇듯 데이터에서 정

〈표 3〉 OAIS 참조모델 데이터 및 정보의 정의(CCSDS, 2012)

구분	세부사항
데이터	<ul style="list-style-type: none"> <li>• 소통과 해석, 처리에 적합하게 공식화된 방식으로 재해석 할 수 있게 한 정보의 표현체</li> <li>• 데이터 용례는 비트배열, 난수표, 페이지의 글자, 말소리의 녹음이나 월석의 표본 등이 있음</li> </ul>
정보	<ul style="list-style-type: none"> <li>• 교환될 수 있는 모든 형태의 지식</li> <li>• 교환되면서 데이터로 표현</li> <li>• 정보의 용례는 섭씨온도(표현정보)로 측정된 숫자가 표현하는 것처럼 어떻게 해석되는지 설명이 붙어있는 (데이터의) 비트열</li> </ul>

- 14) 광래건 (2019.9.27). AI수사관이 지목했다 “뒷골목 가방 멘 저 남자, 흥기 휘두를 확률 70%.” 조선닷컴. Retrieved from [http://news.chosun.com/site/data/html\\_dir/2019/09/27/2019092700242.html](http://news.chosun.com/site/data/html_dir/2019/09/27/2019092700242.html)
- 15) DBR (2017). Retrieved from [https://dbr.donga.com/article/view/1203/article\\_no/8347](https://dbr.donga.com/article/view/1203/article_no/8347)
- 16) techworld. Retrieved from <https://www.techworld.com/startups/heres-how-ravn-systems-is-harnessing-power-of-ai-3638980/>
- 17) HITACHI Inspire the Next. Retrieved from <https://www.hitachivantara.com>

보로 방향이 발전해가는 것으로 이해할 수 있으나, <표 3>에서 보는 것 같이 OAIS 참조모델(Reference Model for Open Archival Information System)을 제시한 우주 데이터 시스템 자문위원회(Consultative Committee for Space Data Systems, CCSDS 2012)는 비교적 구체적인 데이터와 정보의 정의와 함께 그 생성을 Rowley(2007)의 관점과 다르게 조망하고 있다.

CCSDS(2012)가 제시한 데이터의 정의는 “소통과 해석, 처리에 적합하게 공식화된 방식으로 재해석 할 수 있게 한 정보의 표현체”로 비교적 구체적으로 언급하고 있다. 데이터의 용례에서 볼 수 있듯이, 소통, 해석, 처리는 비트 배열이나 말소리의 녹음 등 어떤 형태여도 가능할 수 있으나, 이에 적합한 공식화된 방식과 재해석이 가능한 표현체 라는 부분을 주목할 필요가 있다. 즉, 정보가 모든 형태의 지식을 지칭한다면, 데이터란 이를 교환하는 과정에서 공식적인 형태로 표준화되고 재해석에 적합한 형태로 표현된 것을 의미한다고 볼 수 있다. 본 연구에서는 공식적인 형태로 표준화되고 재해석에 적합한 형태로 표현되는 한 가지 방법으로서 구조화·범

주화를 제시하였다.

Zeng과 Qin(2016)은 메타데이터를 구조화된 데이터의 예로 제시하는데, 속성(property)과 값(value)으로 구성되어 구조화되고 특정한 기준을 통해 값이 표현되어 표준화되고 재해석에 적합한 형태로 표현된다. 메타데이터의 어원을 살펴보면 Meta는 회랍어에서 파생된 말로서 변화, 추상, 초월 등의 의미를 지닌다. 즉, 정보자원을 구조화하여 새로운 형태로 변경 및 추상화시켜 이전의 상태보다 초월적인 활용가치를 제공하는 것이라고 할 수 있다. 범주화는 일반적으로 값의 범위를 특정 기준으로 제한하여 분석을 용이하게 하는 것으로 볼 수 있으나, 형사사법정보의 경우, 포함된 다량의 개인식별정보로 인해 개인을 식별할 수 없도록 범주화하는 것을 의미한다. 윤미숙(2014)과 나영민(2016)의 연구에 의하면, 개인정보보호법 상 개인정보분류체계에 따라 형사사법정보를 다음 <표 4>와 같이 분류하고 있다.

고유식별정보, 민감정보, 일반정보는 비식별 처리가 필요하며, 그 방안으로서 범주화하여 특정 개인을 식별할 수 없도록 할 수 있다(한국정

<표 4> 개인정보보호법 상 개인정보분류체계에 따른 형사사법정보 분류

구분		유형 구분 내용
고유식별정보		• 피의자, 피해자 등 사건관계인의 주민등록번호, 운전면허번호, 여권번호, 외국인 등록번호
민감정보		• 피의자, 피해자 등 사건관계인의 사상·신념, 노조·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등 관련정보 • 범죄경력자료(형실효법)
일반정보	식별가능성이 높은 정보	• 피의자, 피해자 등의 성명, 나이, 직업, 주소, 연락처 등 • 상훈, 병역, 교육, 경력, 기호 정보 등 • 사건번호, 죄명, 범죄사실, 송치의견, 공소사실, 판결내용 등 • 경찰관, 검사, 판사 등 형사사법기관 종사자의 인적사항 등
	식별가능성이 낮은 정보	• 사건접수·체포·구속·선고·처분 등의 일시 및 장소 등

보화진흥원, 2014; 서병조, 2015). 또한 범주화된 고유식별정보, 민감정보라고 해도 ODI의 Data Spectrum<sup>18)</sup>처럼 데이터의 성격에 따라 등급을 지정하여 사용목적에 따라 공개범위를 지정하는 방안도 고려할 수 있다.

### 3. 형사사법정보의 구조화 및 범주화

형사사법정보는 개인식별정보 및 민감정보(피의자, 피해자 등의 개인정보), 신고/발생 일시 및 장소, 관할, 범행 유형, 범행 도구, 범행 수법, 범죄사실 등 특정 사건에 대한 정황 정보를 구체적이고 사실적으로 담고 있으며, 관련 비문서(음성, 음향 또는 영상, 부호)를 포함하고 있다. 개인정보가 다수 포함된 점에서 형사사법정보는 공공기관의 단편적인 데이터와는 그 속성이 다르다 할 수 있다.

형사사법정보 내 개인정보 및 민감정보에 대한 문제를 해결하고 빅데이터로 활용하기 위해서 형사사법정보를 구조화·범주화하여 속성을 변환하는 것이 필요하다. 형사사법정보를 구조화하고 범주화한다는 것은 형사사법정보 내 핵심요소들을 추출하여 요소들에 대한 값을 일목요연하게 정리하되, 값 작성 시 정확한 실제 값이 아닌 값의 범위를 말해주는 것을 의미한다. 예를 들면, 나이는 '연령대'로, 구체적인 사건 발생 장소 및 주소는 '지역'으로 전환하는 것으로 이는 개인정보 비식별 조치에 해당한다(서병조, 2015; 정상조, 2018; 한국인터넷진흥원, 2019).

이렇게 특정 사건을 처리하도록 수많은 개인정보와 민감정보를 포함하고 있는 형사사법정보를 구조화·범주화한다면, 형사사건의 처리와 관련하여 형사사법정보시스템을 이용하여 작성하거나 취득하여 관리하는 자료로서 특정 개인 및 사건과 연결되어 있는 형사사법정보가 아니라 '범죄데이터'로 전환되어 그 민감한 속성이 바뀌게 될 것으로 파악된다. 이러한 방식은 범죄 수사에 빅데이터를 활용하고 있는 미국 FBI에서도 활용하고 있는 방식으로, 그 사례는 다음의 <그림 5>에서 보이는 바와 같다.

<그림 5>에서 나타나듯이 각각의 행과 열에 포함된 데이터는 구조화되어 각 데이터를 구분하고 있고 데이터의 값은 범주화하였다. 구조화한 요소를 구체적으로 살펴보면, 레코드ID, 담당기관이름, 담당 기관 유형, 도시, 주, 발생 연도, 범죄유형, 해결 여부, 피해자 성별, 피해자 연령, 범죄자 성별, 범죄자 연령, 인종, 범인과 피해자 관계, 범죄 도구 등의 요소를 중심으로 구조화하였다. 또 값의 범주화는 그림에서 보듯이, 담당 기관을 보안관, 주 경찰, 자치주 경찰, 시 경찰 등으로 범주화했고, 피의·피해자 관계도 지인, 이웃, 의붓딸, 무관계, 부인 등으로 범주화했다. 이렇게 특정 사건과 연결된 정보를 구조화·범주화하여 변환한다면 특정 개인이 특정 사건과 연결되어 형사사법업무 처리를 위한 형사사법정보가 아니라 '범죄데이터'로 전환된다. 또한 RDBMS(Relational DataBase Management System)에 적합한 형태로 데이터가 전환됨으로써 타 데이터와 연계하여 빅데이터를 활용할 수 있는 가능성도 커지게 된다.

18) Open Data Institute. Retrieved at <https://certificates.theodi.org/en/about/badgelevels>  
ODI Data Spectrum. Retrieved at <https://theodi.org/about-the-odi/the-data-spectrum/>

**담당기관 유형** **구조화**

Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Type	Crime Solved	Victim Sex	Victim Age
1	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	January	1	Murder or Manslaughter	Yes	Male	14
21	AK00118	North Slope Borough	County Police	North Slope	Alaska	1980	August	1	Murder or Manslaughter	Yes	Female	31
27	AKASPO0	Alaska State Police	State Police	Juneau	Alaska	1980	February	1	Murder or Manslaughter	No	Male	27
48	AL00100	Jefferson	Sheriff	Jefferson	Alabama	1980	February	1	Murder or Manslaughter	Yes	Male	58
24	AK00118	North Slope Borough	County Police	North Slope	Alaska	1980	July	1	Manslaughter by Negligence	Yes	Female	44
7	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	May	1	Murder or Manslaughter	Yes	Male	30
7	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	May	2	Murder or Manslaughter	Yes	Female	42

**<값 범주>** **<값 범주>**

Perpetrator Sex	Perpetrator Age	Perpetrator Ethnicity	Relationship	Weapon
Male	15	Native American/Alaska Native	Unknown	Acquaintance
Male	29	Unknown	Unknown	Blunt Object
Unknown	0	Unknown	Unknown	Rifle
Male	43	Black	Unknown	Handgun
Male	41	Unknown	Unknown	Rifle
Male	36	White	Unknown	Blunt Object
Male	34	Black	Unknown	Rifle
Male	37	Unknown	Unknown	Knife

**피의 · 피해자관계**

**<값 범주>**

- ☒ County Police
- ☒ Municipal Police
- ☒ Sheriff
- ☒ State Police

**<값 범주>**

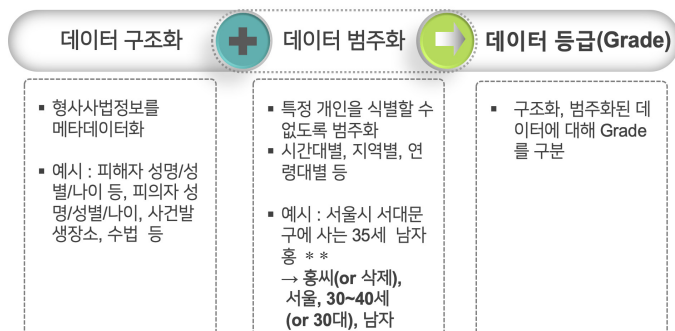
- ☒ Acquaintance
- ☒ Neighbor
- ☒ Stepdaughter
- ☒ Unknown
- ☒ Wife

〈그림 5〉 미국 FBI 범죄데이터의 구조화 · 범주화 예시<sup>19)</sup>

본 연구에서 제시하는 형사사법정보의 구조화 · 범주화를 통한 빅데이터 활용은 〈그림 6〉에서 보는 것과 같이 데이터의 구조화와 범주화를 통해 형사사법정보의 속성을 변환하여 새로운 ‘범죄데이터’를 구성하는 것을 근간으로 하며, 실제 활용영역에서는 데이터 등급을 지정하여 이루어진다.

데이터 등급은 ‘범죄데이터’를 구성하는 요소, 범죄유형에 따라 설정할 수 있으며, 이는 데이

터를 제공하는 각 형사사법기관의 내규에 따라 원칙을 설정하여 정할 수 있다. 〈그림 7〉은 형사사법정보의 구조화 · 범주화 예시로서 데이터 제공기관은 데이터 사용 용도와 목적에 따라 등급이 높은 개인정보에 해당하는 요소를 제거하고 제공하게 된다. 그리고 범죄유형에 대해서도 연쇄살인과 같이 등급이 높은 데이터에서부터 단순한 과실에 해당하는 데이터까지 등급에 따라 제공할 수 있다.



〈그림 6〉 형사사법정보의 구조화 · 범주화, 등급선정

19) FBI 공식데이터 공개 페이지. Retrieved from <https://ucr.fbi.gov/crime-in-the-u.s>

이\*\* (460914-1\*\*\*\*\*)은 서울 종로에 있는 용접주식회사에서 용접공으로 일하고 있는 사람이다. 피의자는 2009년 7월 경부터 피해자 오\*\* (여, 57세)과 사귀어 왔는데, 피해자는 2012년 7월 12일 경 피의자에게 문자를 보내 더이상 피고인과 교제하기 힘들다며 결별을 통보하였다... 이에 앙심을 품은 피의자는 2012년 7월 19일 피해자와 피해자의 동생 오\*\* (여, 54세)을 살해하기로 마음먹었다. 피의자는 2012년 7월 19일 오전 스마트폰을 이용해 불붙는 기름, 주방용 칼 파는 곳, 종로 총 구할 수 있는 곳 등을 검색하여 같은날 부엌칼(증제7호, 총 길이 33cm, 칼날길이20cm)를 구입하였다. 피의자는 2012년 7월 20일 03시 13분경 서울 중구 명동2가 3\*-\* 빌라 201호 피해자들의 주거지에서 주변에 사람이 없는 것을 확인하고 가스배관을 타고 ...

사건번호	피고인 이름	주민번호	성별	직업	사건발생 일시	발생장소	범행도구	범행동기	...
A0001	이**	460914-1*****	남	용접공	2012.7.20	서울	칼	보복성	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...

개인정보  
삭제

개인정보  
삭제

상세정보  
고려

범주화  
색인어

범주화  
색인어

〈그림 7〉 형사사법정보의 구조화 · 범주화 예시

#### 4. 전문가 자문 및 FGI를 통한 검증

형사사법정보의 구조화 · 범주화를 통한 빅데이터 활용방안을 모색하기 위해 ‘범죄데이터’의 법적 사용가능성, 빅데이터로서 활용가치, ‘범죄데이터’의 생성 및 활용에 있어 고려사항을 확인하고 도출하기 위해 전문가 자문 및 포커스 그룹 인터뷰(FGI)를 수행하였다. 본 연구에 참여한 전문가들은 빅데이터, 법 해석, 정보화 및 공공데이터, 정보보안 전문가들로 구성되었다. 전문가들의 소속, 직위 및 각각의 인터뷰 장소와 시간은 〈표 5〉에서 제시하였다.

##### 4.1 ‘형사사법정보’의 빅데이터 활용을 위한 법적 사용가능성

형사사법정보를 빅데이터로 활용하는데 있어서 제약이 되는 가장 핵심적인 이유는 형사사법 정보 전자화 촉진법 제6조 3항에서 이 정보의 형사사법처리 목적 외 사용을 금지하고 있기 때

문이다. 이 부분과 관련하여 관련 전문가에게 ‘형사사법정보’를 구조화 · 범주화하여 ‘범죄데이터’ 형태로 전환하여 빅데이터로 활용하기 위한 법적 가능성에 대하여 자문을 구했다. 내용은 〈표 6〉과 같다.

전문가 자문에 따르면, 현재 형사사법정보 시스템인 KICS내에 형사사법 처리를 목적으로 수집한 자료에는 개인정보 및 민감정보가 포함되어 있으므로 형사사법정보를 빅데이터로 활용하기 위해서는 그 정보를 ‘범죄데이터’로 구조화하여 속성을 변환한다면 가능할 수 있을 것이라는 견해를 제시하였다.

이를 위해서는 법률에 ‘범죄데이터’에 대한 조작성 정의가 제시되어야 함을 언급하였다. 예를 들어 ‘범죄데이터’의 정의, 대상 범위, 활용 범위 등을 법률에 명시해야 한다는 것이다. 법률에 ‘범죄데이터’에 대한 정의가 명시되어 있지 않다면 추후 이의 활용과 범위에 대한 논란 가능성이 농후하다고 보고 지속적인 활용을 위해서는 반드시 필요하다는 의견을 제시하였다. 뿐만 아니라 형사사법정보 내 개인정보 및 민감정보를

〈표 5〉 전문가 인터뷰 장소 및 시간

전문가 code	소속기관	담당업무	면담장소	면담시간
A	S대학교	빅데이터 주임교수	경찰청	2019. 07. 15 17:00~21:00
			경찰청	2019. 08. 30 16:00~18:00
B	데이터마케팅 기업	본부장	본사 지하 강연장	2019. 07. 23 16:00~18:00
C	법 제 처	법령해석	세종시 법제처	2019. 07. 25 17:00~21:00
			경찰청	2019. 08. 30 16:00~18:00
D	K대학교	문헌정보학과 교수	경찰청	2019. 07. 31 12:00~14:00
			경찰청	2019. 08. 30 16:00~18:00
E	S대학교	빅데이터학과 교수	S대학교 교수회관	2019. 08. 12 11:00~13:00

〈표 6〉 ‘형사사법정보’의 빅데이터 활용을 위한 법적 사용가능성에 관한 전문가 의견 요약

전문가	면담 내용
C	<ul style="list-style-type: none"> <li>• ‘범죄데이터’에 대한 조작적 정의 필요</li> <li>• 형사사법정보의 정의만으로는 부족, 형사사법정보의 가명정보나 익명정보에 대한 법률 필요</li> <li>• ‘범죄데이터’의 활용 및 활용 범위에 대한 법적 근거를 마련해야 함</li> <li>• 형사사법정보 내 개인정보·민감정보를 가상정보로 만든다면 가능</li> <li>• 특별한 경우 범죄데이터 내 개인정보·민감정보 활용에 대한 합의 필요</li> <li>• 익명정보의 수집·저장·이용은 형사사법정보에 있는 개인정보에 포함된다 할 수 없음</li> </ul>

개인을 알아볼 수 없는 형태로 전환하여 가상정보로 만든다면 활용 가능성이 있으나 이 또한 가명정보나 익명정보에 대한 법률이 필요하다고 보았다. 이러한 익명정보나 가명정보의 수집·저장·이용은 형사사법정보 내에 있는 개인정보에 포함된다고 볼 수 없어 개인정보보호법 적용 대상이라 보기 어렵다는 의견이었다. 또한 특별한 경우 범죄데이터에 포함된 개인정보나 민감정보를 활용하는 것에 대한 관련 기관 간 합의가 필요함을 덧붙였다.

#### 4.2 ‘형사사법정보’의 빅데이터로서 활용 가치

형사사법정보를 구조화·범주화하여 ‘범죄데

이터’로 데이터화 한다면 빅데이터로서 활용 가치가 무엇인지에 대하여 각 분야별 전문가에게 자문을 구했다. 자문 내용은 〈표 7〉과 같다.

‘형사사법정보’의 빅데이터로서 활용가치에 대한 자문위원들 대부분의 공통된 의견은 ‘데이터화’가 시급하다는 것이었다. 현재 형사사법정보는 텍스트, 사진, 영상, 녹음 형태의 비정형 형태이거나 데이터 형태라 하더라도 빅데이터로 활용할 수 있는 형태는 아니라는 것이다. 형사사법정보가 빅데이터로서 활용할 가치가 있는가를 논하기 전에 구조화된 데이터 형태인가에 대한 검토가 필요하다는 것이다. 이 부분에서 형사사법정보가 데이터인가 혹은 데이터 형태가 되려면 어떻게 해야 하는가와 관련한 정보학적 측면의 기본적인 연구가 필요함을 확인할 수 있었다.

〈표 7〉 ‘형사사법정보’의 빅데이터 활용가치에 관한 전문가 의견 요약

전문가	면담 내용
A	<ul style="list-style-type: none"> <li>• 범죄 예방·예측뿐만 아니라 범죄 간 연결고리, 범행 대상, 범행 동기 등 분석 및 역추적 가능</li> <li>• 누적된 범죄데이터는 분석 알고리즘의 정교화를 가능하게 함</li> <li>• 범죄 유형간 연결고리, 인과관계, 범죄 원인, 범죄 패턴에 대한 연구 가능</li> <li>• 축적된 범죄데이터를 공유함으로써 범죄 추론 및 압목지의 데이터화 가능</li> </ul>
B	<ul style="list-style-type: none"> <li>• 사회정의 구현 및 데이터 공개를 통한 양형 기준에 대한 국민의 신뢰 확보</li> <li>• 범행장소, 범행시간, 범죄유형, 나이, 피해자 유형 등 패턴화 후 활용 가능</li> <li>• 형사사법정보의 빅데이터 활용을 통한 사법기관의 신뢰성 제고</li> <li>• 구속영장 발부 등 형사사법 업무 관련 교육에 도움이 될 수 있음</li> <li>• 빅데이터 분석을 통해 패턴화하면 예방 차원에서 대국민 정보 제공 가능</li> </ul>
D	<ul style="list-style-type: none"> <li>• 빅데이터 활용은 데이터로 변환 후 가능, ‘범죄데이터’의 데이터화가 시급</li> </ul>
E	<ul style="list-style-type: none"> <li>• 빅데이터 분석을 통해 범죄 간 연계성 등 다양한 패턴 정보 유추 가능</li> <li>• 범죄 예방을 위한 선택과 집중이 가능</li> <li>• 범죄 시간, 장소, 인물 등을 파악하여 제한적인 자원을 집중할 수 있음</li> </ul>

형사사법정보는 범죄수사와 관련한 정황과 사실 정보를 담고 있고 범죄 원인을 분석하기 위해 현상과 사실에 입각한 구체적인 내용을 포함하고 있다. 이러한 각각의 정보와 정황을 분석하면 범죄 예방이나 예측뿐만 아니라 범죄 간 연결고리, 범행 대상, 범행 동기 등을 분석하여 그 원인에 대한 역추적이 가능하다는 것이다. 예를 들어 특정 시간과 관련한 범죄 유형이나 범행 대상을 역추적하여 범행 동기 등을 과학적으로 추론하거나 패턴을 안다면 범죄의 환경적 요인을 제거하여 범죄를 예방할 수 있는 가능성이 높아진다는 것이다. 또한 이러한 과학적인 분석 결과를 바탕으로 제한된 자원과 인력을 선택적으로 집중 할 수 있으며 이는 선진 사례에서도 증명된 바 있다. 이러한 과학적인 데이터 분석은 형사사법기관의 신뢰성 제고에도 기여할 것이다.

누적된 수십만 건 규모의 형사사법정보는 범행장소, 범행시간, 범죄유형, 대상 등 범죄와 관련한 패턴을 분석함으로써 다양한 정보를 유추할 수 있는 가능성이 높다. 뿐만 아니라 형사사

법기관의 형사사법처리와 관련한 구속영장 발부나 양형에 대한 과학적이고 합리적인 기준 마련 등은 국민적 신뢰 확보에도 기여할 것으로 보았다. 더욱이 범죄 분석 및 범죄 처리와 관련한 노하우 등 압목지를 데이터화 할 수 있어 교육이나 전문성 제고에도 활용할 수 있으므로 활용가능성 및 적용 분야가 상당히 다양하다는 의견이었다.

#### 4.3 ‘범죄데이터’의 생성 및 활용에 있어 고려사항

형사사법정보를 ‘범죄데이터’로 전환하기 위하여 고려해야 할 점은 무엇인지, 이를 활용함에 있어서 고려해야 할 부분에 대한 전문가 자문을 수행하였다. 자세한 내용은 다음 〈표 8〉에서 제시하였다.

‘형사사법정보’의 빅데이터로서 활용 가치 부분에서 언급한 것처럼 이를 위해서는 우선 분석이 가능한 데이터 형태로 존재해야 한다. 범죄 패턴과 유형을 분석할 수 있는 데이터 구조가



〈표 8〉 ‘범죄데이터’의 생성 및 활용에 있어 고려사항에 관한 전문가 의견 요약

전문가	면담 내용
A	<ul style="list-style-type: none"> <li>• 범죄 패턴·유형 등을 분석할 수 있는 데이터 구조가 되어야 함</li> <li>• 빅데이터 분석은 형사사법정보를 정형화해야 가능함</li> <li>• 범죄데이터에 포함되어 있는 특정 지명 등을 특정할 수 없도록 범주화 필요</li> </ul>
B	<ul style="list-style-type: none"> <li>• 빅데이터로 활용하기 위해서는 분석이 가능한 데이터 형태여야 함</li> <li>• 정제된 데이터와 원시 데이터를 이원화해야 함</li> <li>• 내부적으로 규정과 절차를 만들고 감사체계를 통해 검증하는 처리 매우 중요</li> </ul>
C	<ul style="list-style-type: none"> <li>• 형사사법정보의 가명·익명정보 필요</li> <li>• 특별한 경우 범죄데이터 내 개인정보·민감정보 활용에 대한 합의 필요</li> <li>• 경찰청 내 데이터 관리 조직 필요</li> </ul>
D	<ul style="list-style-type: none"> <li>• 형사사법정보를 분석 가능한 데이터로 만드는 것이 우선시 되어야 함</li> <li>• ‘범죄데이터’에 대한 경찰청 내 뿐만 아니라 유관기관 간 합의가 있어야 함</li> <li>• 공개 가능한 데이터 선별 및 ‘범죄데이터’의 등급화 필요</li> <li>• 공공데이터로 공개하는 것이므로 공공의 목적이 무엇인가를 생각해야 함</li> <li>• 빅데이터 활용의 목적과 가치가 위배 되지 않아야 함</li> <li>• ‘범죄데이터’에 대한 경찰청 내 혹은 유관기관 간 합의가 선행되어야 함</li> </ul>
E	<ul style="list-style-type: none"> <li>• 매우 민감한 정보여서 관리적, 기술적, 물리적 보안 필요 <ul style="list-style-type: none"> <li>- 관리적 절차: 타 정보와 조합, 비교, 분석을 금지한다는 조항 필요</li> <li>- 기술적 절차: 정보사용을 추적하여 정보사용이 적절한지 관리</li> </ul> </li> <li>• 형사사법정보가 저장된 데이터인지 네트워크를 통해 전송해야 하는 데이터인지에 대한 단계적 검증을 거친 후 적절한 조치가 적용될 수 있음</li> <li>• 형사사법정보는 독립성이 요구되는 정보이므로 다른 사건 정보에 영향을 줄 우려가 있으므로 적절한 수준의 공개가 필요</li> </ul>

되어야 한다는 것이며 이를 위해서는 정형화해야 가능하다고 보았다. 예를 들어 사건과 관련한 지역, 시간, 피해자나 피의자의 성별이나 나이 같은 인구통계학적 내용 등 분석할 수 있는 형태의 구조로 정형화해야 한다. 이때 범죄데이터에 포함되어 있는 특정 지명이나 장소를 특정할 수 없도록 범주화하여 특정 지역 주민이나 피해자나 피의자 가족들의 기본권이 침해되지 않도록 해야 한다고 강조하였다. 이러한 부분은 형사사법기관 간 합의가 있어야 가능할 것으로 보았다.

범죄데이터는 민감한 정보여서 타 정보와 조합하거나 비교 또는 분석을 금지한다는 조항을 두어야 한다. 기술적 절차로는 정보사용을 추적

하여 정보를 사용한 것이 적절한지를 관리하여야 한다. 또한 형사사법정보가 저장된 데이터인지 네트워크를 통해 전송해야 하는 데이터인지를 구분하여 여러 단계의 검증을 거친 후에야 적절한 조치가 적용될 수 있도록 해야 한다고 하였다.

기본적으로 형사사법정보는 각각의 형사사법기관의 독립성이 요구되는 정보이므로 다른 사건 정보에 영향을 줄 우려가 있을 수 있으므로 공개 한다면 적절한 수준으로 해야 할 필요가 있다는 의견이었다. 또한 형사사법정보를 빅데이터로 활용하기 위해 공개한다면 공공의 목적이 무엇인가를 생각하여 활용 목적과 가치가 위배되지 않도록 해야 함을 강조하였다. 이를 위

해서는 범죄데이터 중 무엇을 공개할 수 있는지 선별해야 하기 때문에 ‘범죄데이터’를 등급화 할 필요가 있으며 정제된 데이터와 원시 데이터를 이원화할 것을 제안하였다.

또한 빅데이터 활용의 목적과 가치가 위배되지 않도록 하기 위하여 데이터 형태 및 이용을 위한 내부 규정과 절차를 제도화하고 감시체계를 통해 데이터를 검증하는 것도 매우 중요하다고 하였다. 이에 관하여 유관기관간의 합의를 통하여 데이터를 등급화하고 공개할 수 있는 데이터를 선별하여 그 수준을 정하고 이를 관리 감독할 조직이 필요하며 이에 관한 법적 근거를 마련해야 함을 언급하였다. 형사사법기관이 내부적으로 규정과 절차를 만들어 감시체계를 통해 범죄데이터를 검증하는 처리가 매우 중요하다고 하였다.

전문가들은 형사사법정보를 빅데이터로 활용하기 위해 고려해야 할 점으로 크게 세 가지를 언급하였다. 형사사법정보를 데이터로 구조화해야 한다는 것, 형사사법정보 활용의 목적과 가치가 위배되지 않도록 해야 한다는 것, 빅데이터 활용을 위해 범죄데이터를 등급별로 검증하고 관리하는 부분에 관한 것이었다.

#### 4.4 전문가 의견 기반 형사사법정보의 빅데이터 활용 추진방안

앞 절에서 형사사법정보를 ‘범죄데이터’로 전환하여 빅데이터 활용이 이루어지기 위하여 법, 활용가치, 고려사항 영역에서 전문가들의 의견을 검토하였다. 이를 기반으로 형사사법정보의 빅데이터 활용을 위해 추진할 사항들을 정리하였다.

먼저 법적 영역에서, 개인정보보호법 제18조 제2항 2호에 따르면 ‘통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우’ 개인정보를 목적 외 이용하거나 제3자에게 전달할 수 있다. 이러한 근거에 따라 형사사법정보도 일부 통계데이터로 전환되어 공개되고 있다. 이와 같은 맥락에서 형사사법정보 내 개인 민감정보를 구조화·범주화를 통해 개인을 식별할 수 없도록 ‘범죄데이터’로 전환하여 공익을 위해 공개하는 방안에 관해 법 전문가의 핵심적인 의견은 개인을 식별할 수 없도록 하면 빅데이터 활용가능성이 있지만, 「형사사법절차 전자화 촉진법」의 강력한 법률 제약으로 인해 법적인 해석의 문제가 남을 수 있으므로 법률적인 개정은 불가피 함을 제시하였다. 즉, 공익을 위해 형사사법정보를 활용하고 개인을 식별하지 못하더라도 「형사사법절차 전자화 촉진법」 제6조 제3항의 해석에 따라 범법행위가 될 수 있는 위험이 있다는 것이다. 이에 따라, 「형사사법절차 전자화 촉진법」 내에 ‘범죄데이터’의 정의, 활용 목적 및 범위, 위반시 처벌 등의 내용이 기술되거나 혹은 형사사법정보의 목적 외 사용에 대한 정의가 기술되는 방향으로 개정되는 것이 필요하다고 판단된다. 이를 위해 단계적인 추진전략이 필요할 것으로 예상된다.

다음으로 형사사법정보의 빅데이터 활용가치에 있어서는 다양한 해외사례에서 볼 수 있듯, 범죄 예측을 통한 범죄발생 억제, 누적된 데이터를 통한 범인 추정, 그리고 데이터 기반 양형 합리화와 형사사법기관의 신뢰 확보 등 다양한 활용방안이 논의 되었다. 형사사법정보의 빅데이터 활용을 통해 공익에 기여하는 것은 매우

환영할 만한 것이지만, 그 활용가치가 충분히 발현되기 위해서는 고려사항이 다수 존재한다. 첫째, 형사사법정보 분석, 국내외 활용현황, 기술동향 등을 충분히 고려하여 국내 상황에 적절한 활용이 이뤄지도록 ISP(Information Strategy Planning)와 같은 체계적인 방법론에 의한 서비스 및 시스템 구축계획이 선행되어야 한다. 둘째, 형사사법정보의 구조화·범주화를 위해 빅데이터 분석에 필요한 메타데이터 요소 도출과 값에 대한 데이터 사전(dictionary) 정의가 필수적이다. 이를 위해서는 형사사법정보 구조와 국제 표준 및 모범실무(best practice) 조사와 함께 범죄현장 전문가가 함께 참여하여 학술적 관점과 실무적인 관점이 어우러진 폭넓은 체계를 구축할 필요가 있다.

이 외에도 현재 형사사법정보가 비정형데이터로 구성되어 있는 문제를 분석에 용이한 형태로 자동화하기 위한 작업을 고려해야 한다. 머신러닝을 활용하여 자동화 하더라도 학습데이터가 필요하기 때문에 정교한 패턴 도출 및 시소러스 구축이 선행되어야 할 것으로 판단된다. 아울러 활용목적에 따라 원칙있는 데이터 접근이 이뤄지도록 데이터 등급선정 및 내규 마련 또한 필요할 것으로 판단된다.

## 5. 형사사법정보의 빅데이터 활용을 위한 단계별 추진전략

형사사법정보는 범죄 현상을 분석하고 패턴을 찾아내는 등 우리 사회의 공익을 위해 중요한 원자료로 파악된다. 앞 장에서 제시한 관련 전문가들의 의견을 기초로 형사사법정보의 빅데이터

활용 방안을 종합적으로 고려해 볼 때, 형사사법정보가 빅데이터로서 활용이 가능하려면 각 영역별로 단계별 추진이 필요하다. 이를 위한 주요 고려사항은 다음과 같다.

첫째, 형사사법정보의 속성전환이 필요하다. 정보학적 관점에서 형사사법정보의 내포는 특정 사건에 대한 정황 정보를 구체적이고 사실적으로 담고 있는 정보이다. 이는 특정 사건을 형사사법 처리할 목적으로 생성한 정보로 법적인 관점(형사사법절차 전자화 촉진법 제2조(정의))에서와 다르지 않다. 이 정보에는 개인정보, 민감정보, 범행관련 사실 정보 등을 담고 있다. 이 정보의 외연은 형사사법정보시스템에 의한 형사사법처리 목적으로 수집·저장 관리하는 모든 자료라 할 수 있다. 따라서 형사사법정보의 근본적인 속성전환을 통해 개인정보보호문제를 해결해야 한다. 단, 전문가 자문 결과에서 보여지듯이 '범죄데이터'의 정의와 활용범위 및 대상에 대한 법적근거를 마련하는 것은 필요하므로 단계적 개선 전략이 필요하다.

둘째, 형사사법정보의 구조화는 형사사법정보의 속성전환에 관한 문제뿐만 아니라 어떤 요소를 분석할 것인지 정의하는 것으로 빅데이터 활용과 밀접한 관련이 있다. 어떤 데이터 요소를 분석해야 하는가를 선결해야 한다. 이는 데이터로서의 신뢰성 및 품질 확보 측면에서도 필요하다 할 수 있다. 데이터 요소의 누락 혹은 오류는 데이터에 대한 신뢰성 문제와 직결되며 이로 인한 분석 결과의 오차는 의사결정이나 정책에 치명적인 영향을 줄 수 있기 때문이다. 또한 분석이 용이하도록 표준화 된 형태로 전환되도록 진행해야 한다.

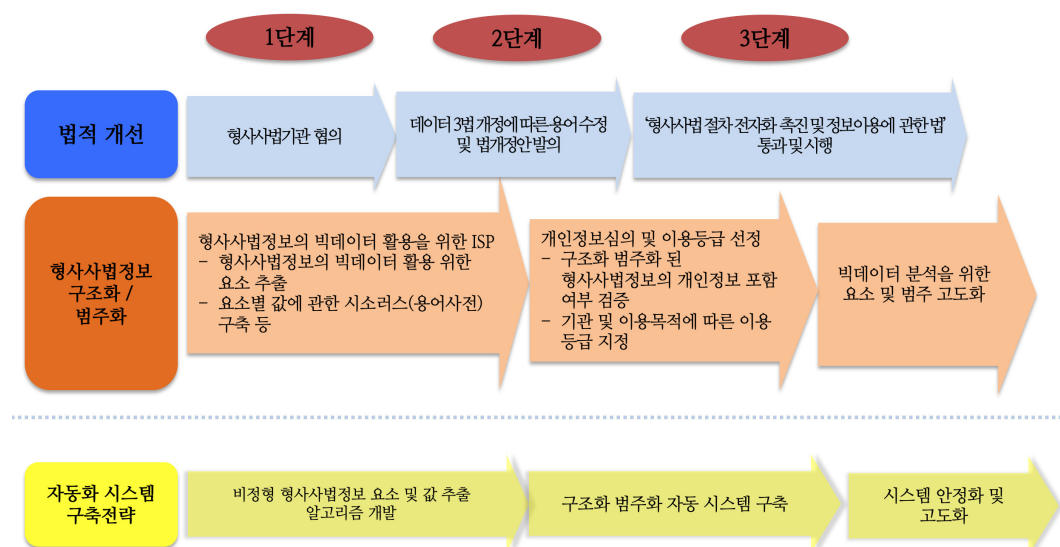
셋째, 각 데이터가 포함하고 있는 내용 요소

별로 범죄와 관련한 지역이나 사건을 특정할 수 없도록 비식별화 처리기능을 포함하여 범주화해야 한다. 즉, 범죄 분석에 필요한 각각의 요소를 추출하여 이를 상위 개념으로 범주화함으로써 사건과 관련한 연관 정보를 특정할 수 없도록 해야 한다. 뿐만 아니라 데이터를 등급별로 규정함으로써 활용목적과 대상에 따라 공개와 활용범위에 대한 정의가 사전에 필요하다.

이상의 연구결과를 토대로 향후 1~3단계로 구분하여 전략적인 수행이 가능하다. 크게 법적 개선, 형사사법정보 구조화·범주화, 자동화 시스템 구축전략으로 구분하여 단계별 진행으로 정리하였다. <그림 8>에서 로드맵을 제시하였으며, 향후 형사사법정보의 빅데이터 활용을 위해 참고할 수 있다.

먼저 법적 개선 영역에서는 제1단계에서 형

사사법기관 간 협의 과정이 필요할 것으로 판단된다. 『형사사법절차 전자화 촉진법』이 2009년 입법된 이후, 본 법이 형사사법정보에 관한 기본법의 역할을 해왔으며 5개의 형사사법기관이 모두 관여된 법이라는 특징을 갖는다. 따라서 모든 형사사법기관의 의견을 수렴할 필요가 있을 것으로 예상된다. 제2단계에서는 현재 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)이 개정 진행 중이고, 이 법들이 통과될 경우 본 연구에서 제시된 법안이 수정될 필요가 있다. 이 중 더불어민주당 인재근 의원이 대표발의(의안번호 제16621호)한 개인정보보호법 개정안이 행정안전위원회 법안심사소위원회에서 의결돼 규제 개선의 첫 삽을 떴다.<sup>20)</sup> 그러나 신용정보법과 정보통신망법이 소관 상임위원회 법안심사소위원회 심사를 마치지 못했다는 이



<그림 8> 1~3 단계별 진행전략

20) 박종진, 송혜영 (2019.11.14). 데이터3법 연내 통과 청신호...개인정보보호법 법안소위 의결, 전자신문. Retrieved from <http://www.etnews.com/20191114000652#>

유 등으로 본회의에는 상정조차 되지 못했다.<sup>21)</sup> 하지만, 데이터 3법 개정과 구체적 개선 방향에 대한 간담회가 개최되는 등 활발한 논의가 진행되고 있다. 이에 향후 데이터 3법 개정안이 통과될 경우 방향은 바뀌지 않지만 전담기관이나 평가단과 같은 용어가 보다 구체적으로 표현될 수 있을 것으로 예상된다. 제3단계에서는 1~2단계를 거쳐 의견이 수렴되고 보완된 법안이 통과되는 모든 과정을 포함한다.

다음으로 형사사법정보의 구조화·범주화 영역에서는 4.4절에서 제시한 것과 같이 형사사법정보의 효과적인 활용을 위한 기반 작업으로서 ISP를 시행하는 것이 바람직하다고 판단된다. 이 과정에서 보편적인 ISP 방법론에 따라 내부와 외부 환경조사, 목표 모델에 대한 제시가 이뤄짐과 동시에 집중적으로 두 가지 연구가 필요할 것으로 보인다. 첫째, 형사사법정보의 빅데이터 활용을 위해서는 핵심 요소 추출에 관한 연구가 진행되어야 한다. 형사사법정보의 빅데이터 활용이 목적에 맞게 이뤄지기 위해 어떤 요소를 추출하여 데이터를 구축할지가 관건이다. 이를 위해 현장 수사전문가와 데이터 전문가, 빅데이터 분석 전문가가 TF팀을 이루어 실제적으로 활용될 수 있는 요소를 추출하는 것이 필요하다. 둘째, 추출된 요소별 값에 관한 시소러스(용어사전)를 구축해야 한다. 예를 들어 이웃집 사람, 옆집 사람, 아랫집 사람 등 자연어에서는 다양한 용어가 유사어로 등장할 수 있는데, 이에 대한 대표어를 지정하여 구조화·범주화할 때, 대표어로 통제해야 빅데이터 분석이 용이해진다. 이 과정에서 요소마다 범주화할 값들

을 지정하는 것이 동시에 필요하다. 제2단계에서는 구조화·범주화한 형사사법정보가 개인정보를 포함하지 않는지 심의과정을 거친다. 그리고 개인 이용자, 공공기관 등 이용대상 및 목적에 따른 이용등급을 선정하여 기준과 원칙이 있는 데이터 활용 규정을 제작하여야 한다. 제3단계에서는 1~2단계에서 추출된 형사사법정보의 요소, 값, 범주 등을 평가하여 빅데이터 분석을 위한 요소 및 범주의 고도화를 통해 활용 범위 및 용이성을 높일 필요가 있다.

마지막으로 자동화 시스템 구축에 대한 연구를 수행할 필요가 있다. 형사사법정보가 대부분 비정형데이터와 상용 포맷 문서로 작성되어 있는 실정이다. 형사사법정보의 빅데이터 활용이 현재 법적인 제한이 있지만, 향후에는 비정형데이터의 가공문제가 반드시 활용에 제약사항이 될 것이다. 따라서 단계적으로 그 대비를 할 필요가 있다. 형사사법정보의 요소와 값들 자동으로 추출할 수 있는 알고리즘을 개발 및 시스템을 구축하고, 장기적으로 이를 고도화하여 최대한 인적소요가 줄어들 수 있도록 진행해야 한다.

## 6. 결론

영국의 경우, 2017년 4월 형사사법정보 처리를 전면 전자화했으며, 독일을 비롯한 선진 국가들도 최근 몇 년 전부터 형사사법정보 자동화 시스템 구축에 나서고 있다. 한국 형사사법기관이 유수의 선진국보다도 한발 앞서 형사사법정

21) 백봉삼 (2019.11.25). 스타트업얼라이언스, 데이터 3법 개정안 국회 토론회 연다. ZDNet Korea. Retrieved from <http://www.zdnet.co.kr/view/?no=20191125083511>

보 전자화를 시작한 점과 경찰, 검찰, 법원, 법무부 4개 형사사법기관 간에 협력과 소통을 통해 기관마다 고유 영역, 기득권 등 견해 차를 극복하고 공통시스템을 구축한 사례는 최초의 사례로 전 세계가 주목하고 있다.<sup>22)</sup> 이러한 과정에서 축적된 노하우를 배우고자 페루, 우즈베키스탄 등의 국가가 협력사업을 요청하여 진행 중이며, 오스트리아, 인도네시아, 몽골 등에서 우리 형사사법정보시스템을 벤치마킹하고 있는 것은 어쩌면 당연한 일일 수 있다.

그러나 한국이 우수한 경험을 앞서 축적하여 형사사법정보시스템을 구축하였음에도 불구하고 현재까지 여러 제약으로 인해 형사사법정보 처리 목적 외 활용이 원천 봉쇄되고 있는 것은 참으로 안타까운 일이라 할 수 있다. '형사사법정보'를 빅데이터로 분석함으로써 범죄 현상에 대한 패턴을 제시하고 이를 예측할 수 있다는 것은 미국의 'COPLINK', 'PredPol' 등의 사례에서처럼 많은 부분이 검증되어 활용되고 있다. 물론 범죄 사실과 관련한 구체적인 정황적 사실을 묘사하고 있는 정보와 개인정보를 포함하고 있는 형사사법정보에 대한 접근을 권한이 있는 사람으로 제한하고 보수적으로 철저히 관리하는 것은 마땅한 일이다. 하지만 이미 수집, 저장된 형사사법정보를 범죄 예방 및 과학적이고 합리적인 양형 기준 제시 등을 위한 '범죄데이터'의 활용 가능성을 모색하는 것이 필요한 시점이다.

본 연구는 지금까지 학계에서 이뤄졌던 형사사법정보 관련법과 활용의 문제점 분석 차원을 넘어, 형사사법정보를 구조화·범주화한 '범죄데이터'가 활용에 있어 법적 제한점, 활용가치,

활용 시 고려사항을 전문가 자문을 통해 조사한 최초의 연구라는 점에서 의의를 가진다. 본 연구결과를 종합하면 다음과 같다.

빅데이터 활용에 관한 선행연구 대부분은 주로 데이터의 기술적인 활용 방법 및 가치와 예측에 관한 내용을 주로 하고 있다. 그러나 본 연구에서는 형사사법정보의 구조화·범주화를 통해 '범죄데이터'로 전환하여 빅데이터로서 활용하도록 제안하였으며, '범죄데이터' 활용시 법률적 문제, 활용가치, 데이터 생성 및 활용시 고려사항을 전문가를 통해 검증하고 향후 전략적 발전방안을 도출하였다. 연구결과, '범죄데이터'는 개인정보보호문제는 해결된 것으로 보이나, 형사사법정보 관련법에 명시할 필요는 있으며, 빅데이터 활용을 위해 분석 가능하도록 표준화된 형태로 정리되는 것이 시급함이 밝혀졌다. 향후 진행방향으로는 데이터 요소 도출, 용어사전 시소러스 구축, 데이터 등급화를 위한 개인민감정보 정의 및 등급지정, 비정형데이터의 정형화를 위한 알고리즘 개발 등이 이루어져야 할 것이다. 형사사법정보를 '범죄데이터'로 데이터화 하는데 있어 속성별(attribute)로 구조화·범주화하고 이를 등급화 하는 구체적인 방법에 관한 연구는 추후 과제로 삼고자 한다.

현재까지 형사사법정보의 이용에 있어 가장 큰 제한점은 개인정보의 보호 문제로 인한 개인의 피해 우려였다. 그러나 개인정보를 보호받을 권리도 중요하지만, 국민들이 범죄로부터 안전하고 자유롭게 살아갈 수 있는 권리도 매우 중요하다. 따라서 개인의 권리를 보호할 수 있도록 구조화·범주화된 형사사법정보를 빅데이터

22) 윤소정 (2017.3.). 세계와 공유한다, 전자정부 시리즈 8: 형사사법정보시스템. 코리아넷뉴스.  
Retrieved from available at <http://www.kocis.go.kr/koreanet/view.do?seq=7645>

로 활용하여 정보주체의 이익보호 뿐 아니라 보다 안전하고 살기 좋은 사회가 되도록 공익의 발전에도 진보하게 되기를 기대한다. 아울러 본 연구 결과로 형사사법정보의 빅데이터 활용의 단초가 되고, 향후 타분야 빅데이터, 인공지능, 영상 및 음성인식 등 최첨단 정보기술을 형사사법 업무에 활용하여 지속적인 발전이 이루어지기를 기대한다.

## 참 고 문 헌

- 권양섭 (2017). 범죄예방과 수사에 있어서 빅데이터 활용과 한계에 관한 연구. 법학연구, 17(1), 179-198.
- 김기범 (2016). 형사사법정보의 이용제공 실태 및 입법적 개선방안. 법학연구, 16(1), 121-142.
- 김한균 (2018). 형사사법정보의 빅데이터 활용을 위한 법·정책적 과제. 법조, 67(4), 249-284.
- 나영민 (2016). 형사사법정보의 이용과 보호에 관한 연구. 박사학위논문, 성균관대학교 법학전문대학원.
- 나영민, 박광민 (2014). 형사사법절차 전자화의 쟁점과 그 해결방안. 성균관법학, 26(2), 1-22.
- 류연수 (2017). 스마트치안 국내외 사례와 향후 과제. 행정포커스, 3+4(126), 16-21.
- 박민우 (2017). 형사절차전자화법 제6조 제3항의 해석과 형사사법정보의 목적 외 사용. 저스티스, 통권 (161), 200-229
- 법무부 형사사법 공통시스템 운영단 (2014). 국민을 위해 형통하라. 한국표준협회미디어.
- 서병조 (2015). 빅데이터 활용을 위한 개인정보 비식별화 기술활용 안내서 Ver 1.0. 한국정보화진흥원, 1-90.
- 안종욱, 이미숙, 신동빈 (2013). 공간빅데이터 개념 및 체계 구축방안 연구. 한국공간정보학회지, 21(5), 43-51.
- 양종모 (2016). 형사사법절차 전자화와 빅데이터를 이용한 양형합리화 방안 모색. 홍익법학, 17(1), 419-448.
- 우종필 (2019). 2019년 7월 15일, 동년 8월 30일 서울에서 개인면담 및 FGI에서의 담화.
- 윤미숙 (2014). 형사사법정보 오·남용 방지를 위한 정보 분류 체계 구현: 약식사건을 중심으로, 석사학위논문, 고려대학교 정보보호대학원.
- 이연욱 (2019). 2019년 8월 30일 서울에서 FGI 중 경찰청 현황 관련 면담.
- 정상조 (2018). 비식별개인정보의 보호 및 활용에 관한 연구. 정보통신부 최종보고서, 1-132.
- 한국데이터베이스진흥원 (2014). 데이터 분석 전문가 가이드. 서울: 한국데이터베이스진흥원.
- 한국인터넷진흥원 (2019). 2019년 개인정보 비식별 조치 기본교육, 2019.9.18. 1-100.
- Borgman, C. (2015). Big Data, Little Data, No Data Scholarship in the Networked World. MIT Press.
- 심원식, 현은희 공역. (2019). 빅데이터, 새로운 깨달음의 시대. 서울: 성균관대학교출판부.

- CCSDS (2012). Space data and information transfer systems - Open archival information system (OAIS) - Reference model. ISO 14721:2012(en)
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information and Communication Science*, 33(2): 163-180. doi:10.1177/0165551506070706.
- Zeng, M. L., & Qin, J. (2016). *Metadata 2nd Edition*. Chicago: Neal-Schuman.

<p>• 국문 참고문헌에 대한 영문 표기 (English translation of references written in Korean)</p>
--

- Ahn, Jong Wook, Yi, Mi Sook, & Shin, Dong Bin (2013). Study for spatial big data concept and system building. *Journal of Korea Spatial Information Society*, 21(5), 43-51.
- Jung, Sang Jo (2018). Legal review on protection and use of the personally non-identifiable information. Ministry of Information and Communication Final Report, 1-132.
- Kim, Gi Bum (2016). A study on current status of USE or provision of criminal justice information and legislative improvement. *LAW REVIEW*, 16(1), 121-142.
- Kim, Han-kyun (2018). Big data-based criminal justice policy and personal information protection. *Korean Lawyers Association Journal*, 67(4), 249-284.
- Korea Database Agency (2014). *Data analysis expert guide*. Seoul: Korea Database Agency.
- Korea Internet & Security Agency (2019). Basic training on personal information de-identification in 2019. 2019, 18 Sep 2019. 1-100.
- Kwon, Yangsub (2017). Study on the application and legal limits of big data for crime prevention and investigation. *LAW REVIEW*, 17(1), 179-198.
- Lee, Yeon Wook (2019, August 30). Interview with the Korea national police agency during FGI in Seoul.
- Ministry of Justice: Criminal justice common system operation group (2014). *Integrate the Criminal Justice Information for the People*, KSAM.
- Na, Young Min & Park, KwangMin (2014). A study on limitations to computerization of the criminal process. *SungKyunKwan Law Review*, 26(2), 1-22.
- Na, Young Min (2016). A study on the treatment and protection of the criminal justice information. doctoral dissertation, SungKyunKwan University, Department of Law the Graduate School.
- Park, Min Woo (2017). Interpretation of article 6 section 3 of act on promotion of the digitalization of the criminal justice process and other purpose use of criminal justice information. *JUSTICE*, 161, 200-229.



- Ryu, Yeon Su (2017). Domestic and international cases of smart security and future tasks. *Public Administration Focus*, 3+4(126), 16-21.
- Seo, Byeong Jo (2015). Personal information de-identification technology utilization guide for big data utilization Ver1.0. National Information Society Agency, 1-90.
- Yang, Jong Mo (2016). Digitalization of the criminal justice procedure and applying big data analytics in rationalization of criminal sentencing. *The Law Research Institute of Hongik Univ.*, 17(1), 419-448.
- Yoon, Mi Sook (2014). Implementation of information classification system to prevent misuse of criminal justice information-Mainly on abbreviated events. Master's thesis, Korea University Graduate School of Information Security.
- Yu, Jong Pil (2019). 2019.7.15., 8.30. Interview in Seoul and FGI discussion.

