



범죄수사학연구

ISSN 2466-1236



스마트 치안 기술 활성화를 위한 형사사법정보 활용 방안 : 개인정보보호와의 균형을 중심으로*

A Study on Utilization of Criminal Justice Information for Smart Policing with a Focus on Personal Information Protection

박성미** · 안정민***

* 이 논문은 2022년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: AI 기반 범죄수사 지원 / 연구개발과제번호: PR10-02-000-22]

** 고려대학교 정보보호대학원 박사과정(주저자)

*** 한림대학교 정보법과학전공 교수(교신저자)

ABSTRACT

빅데이터와 인공지능을 교통, 의료, 교육 등 다양한 분야에 접목하게 되면서 치안 분야 역시 막대한 연구개발 예산을 통하여 치안 관련 인공지능 개발을 추진하고 있다. 그러나 형사사법 절차상 수집되는 정보는 상당수 개인정보를 포함하게 되어 데이터 활용이 어려울뿐더러 활성화되고 있는 스마트 치안 분야에 비해 관련 개인정보보호정책은 미흡한 상황이다. 본 연구는 형사사법 절차상 생산되는 형사사법정보를 살펴보고 국내외 형사사법정보 활용 현황에 대해 알아보았다.

경찰청은 첨단기술과 치안현장의 융합을 위해 여러 연구개발과제를 시행하고 있으나 활용 가능한 데이터 확보 및 제공에 있어서 난관을 겪고 있으며, 민간기업과 협력, 범주화된 데이터 등을 개방하여 치안 데이터 시장 성장을 촉진하고 있으나 적극적으로 데이터를 학습에 활용하는 미국에 비해 다소 수동적인 모습을 보이고 있다. 이에 본 연구는 형사사법분야에 인공지능기술을 폭넓게 활용함에 장애가 되는 현행 체계의 문제점을 살펴보고, 포괄적인 도메인 중심의 개인정보보호의 관점에서 벗어난 정보의 보안 등급화와 개인정보 침해의 가능성을 고려해 개발 환경을 구축하는 대안방안을 제시하고자 한다.

KEYWORD

스마트 치안, 치안 데이터, 형사사법정보, 개인정보보호, 비식별화

I. 서론

IoT의 확산으로 다량의 데이터가 생성되면서 빅데이터와 인공지능 기술을 교통, 의료, 교육 등 다양한 분야에 접목하려는 움직임이 커지고 있다. 치안 분야에서도 자율주행 자동차의 책임소재 분석¹, 해킹 대응 추적기술개발², 양자기술 기반 보안기능이 탑재된 보안카메라 개발, 인공지능을 활용한 치안드론 등 스마트 치안 연구³를 통해 현장에 ICT 기술을 도입하기 위한 치안 기술 고도화가 진행 중이다. 나아가 첨단 기술을 악용하는 디지털 플랫폼을 통한 신종범죄는 기존의 인력투입 위주 치안활동의 한계를 부각시킴과 동시에 과학기술을 활용한 치안역량 강화의 필요성을 더욱 증가시키고 있다.

우리나라는 일찍이 2014년 경찰법을 개정하여 경찰청장으로 하여금 치안분야 과학기술 진흥을 위한 시책을 마련·추진하도록 하였고, 이를 위한 연구개발사업의 수행에 필요한 경비를 지원할 수 있도록 「치안분야 과학기술 진흥에 관한 규정」을 제정한 바 있다. 그동안 이를 통해 치안 분야의 과학기술 연구개발 사업이 추진되었고 스마트 현장대응·사건 관계망 분석 등 다양한 정부 연구개발(R&D) 과제가 진행되고 있다. 그러나 미국이나 유럽같이 치안분야 전문 연구기관을 두고 정부 주도 하에 세계 치안산업 시장에 대비하는 국가에 비해 우리나라의 치안분야 산·학·연 기반은 상대적으로 약하고 민간분야의 투자도 쉽게 이루어지지 못하여 성장에 어려움을 겪고 있는 상황이다.

2010년 「형사사법절차전자화촉진법」(이하 ‘형사절차전자화법’)의 제정을 통해 최첨단 IT 기술을 이용하여 형사사법기관의 모든 문서를 전자화하였고 형사사법 정보를 공동 활용하는 형사사법정보시스템(이하 ‘KICS’)을 구축하여 현재 모든 형사사법업무 처리기관에서 사용되고 있다. 2021년 9월에는 「형사사법절차에서의 전자문서 이용 등에 관한 법률」이 국회 본회의를 통과하게 되면서 형사사법정보의 전자화 및 데이터베이스 구축을 추진하는 등 치안 분야의 스마트화를 위해 많은 노력을 기울이고 있다. 사회 모든 분야에서 생성되는 각종 데이터가 빅데이터·인공지능을 통해 활용되고 발전을 거듭하고 있는 상황에서 범죄 예방과 수사를 위해서는 첨단과학기술의 활용이 필연적이다. 그러나 현재 경찰의 치안 기술 고도화를 위한 과학기술정책은 타부처에 비해서도 미흡하다는 평가를 받고 있다. 스마트 치안이라고 할 수 있는 치안 분야의 기술 고도화는 4차 산업혁명에 따른 신종 범죄나 메타버스 등의 가상세계에서 발생할 수 있는 각종 위협에 대응하기 위해서도 필수적이

1 조민제, 자율주행 ‘치안 안전’의 시작, 치안과학기술리뷰, 1권, 2021, 68면.

2 박광하, “민·관·군·경 협력해 랜섬웨어 대응역량 강화”, 정보통신신문, 2022. 9. 20.
<https://www.koit.co.kr/news/articleView.html?idxno=103123> (검색 2022. 11. 24.).

3 곽중희, “미래 안전, 스마트치안이 책임진다”, CCTV뉴스, 2022. 10. 13.
<https://www.cctvnews.co.kr/news/articleView.html?idxno=233661> (검색 2022. 11. 24.).

라 할 수 있다. 그럼에도 불구하고 그동안 경찰은 기본권 보장과 경찰권 남용을 우려하여 수사과정에서 발생한 중요한 데이터의 활용 및 공유에 소극적인 자세를 취해온 것이 사실이다.

형사사법 절차상 수집되는 정보가 상당수의 개인정보를 포함하고 있고 관련 당사자에게는 민감한 영역이 되기 때문에 이러한 경찰의 개인정보보호를 위한 노력을 폐쇄적이라고 비난할 수는 없는 일이다. 그러나 많은 정보를 제공하고 예측에 활용할 수 있는 보유데이터를 충분히 이용하지 못하고 빠른 속도로 발전하는 첨단기술을 활용하지 못한다면 자칫 치안 분야가 사회의 변화에 대응하지 못하게 될 우려가 있다. 이러한 관점에서 본 연구는 치안 기술의 고도화를 위해 활용할 수 있는 형사사법정보의 활용실태를 살펴보고 자동 비식별화 기술 등 개인정보를 보호하면서 치안 기술의 고도화를 추구할 수 있는 방안을 다양하게 모색하여 스마트 치안 활성화 방안을 제안해보고자 한다.

II. 형사사법정보의 정의 및 생성 정보

1. 형사사법정보의 개념 및 유형

2009년 형사절차전자화법의 제정으로 그동안 경찰·검찰·법원·법무부 등이 각각의 형사사건 처리 과정에서 수집하고 관리하던 형사사법정보를 통합적으로 관리·공유할 수 있는 형사사법정보시스템(이하 'KICS')이 구축되었다. KICS는 각 기관 간 업무처리과정을 연계하고 데이터를 활용함으로써 수사, 기소, 재판, 형의 집행에 이르는 일련의 형사사법절차를 보다 신속하고 투명하게 집행하고 대국민 형사사법서비스를 개선하기 위한 목적으로 만들어졌다.

형사절차전자화법 제2조는 형사사법정보를 “형사사법업무 처리기관이 형사사법업무 처리와 관련하여 형사사법정보시스템을 이용하여 작성하거나 취득하여 관리하고 있는 자료로서 전자적 방식으로 처리되어 부호, 문자, 음성, 음향 또는 영상 등으로 표현된 것”으로 정의하고 있다. 즉, 법원, 법무부, 검찰청, 경찰청 등의 형사사법업무를 처리하는 기관이 “수사, 공소, 공판, 재판의 집행 등 형사사건의 처리와 관련된 업무”와 관련하여 KICS를 이용하여 작성, 취득, 저장, 송신, 수신하는 자료라고 할 수 있겠다(제2항 제1호 내지 4호).

형사사법정보는 성명, 주민번호, 외국인번호, 성별, 나이, 직장, 주소, 주거, 등록기준지, 전화번호, 이메일 주소와 같이 개인정보가 포함된 정보와 신고일시, 접수단서, 발생일시, 발생지 유형, 발생 관서, 관할 검찰청, 죄명, 관할법원, 관할 지구대와 같이 정보 주체를 식별할 수 없는 비식별 정보로 구분될 수 있다. 수사단계에서의 형사사법정보는 정보주체의 행위에 초점을 맞추기 때문에 필연적으로 많은 개인정보를 포함하고 있어 그 활용에 제한이 따르게 된다.

개인정보를 포함한 형사사법정보는 개인정보보호에 관한 일반법인 개인정보보호법의 적용을 받는다고 해석하여야 한다. 개인정보보호에 있어서 개인정보보호법과 형사절차전자화법을 일반법과

특별법의 관계로 보는 이유는 2011년 개인정보보호법 제정 당시부터 제6조(다른 법률과의 관계)에 있어 개인정보 보호에 관하여는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」 등 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는 바에 따르도록 규정하고 있기 때문이다. 또 형사사법절차 전자화 촉진법 시행령 제18조는 개인정보보호법 24조의 2가 정하는 주민등록번호 처리제한에 따라 개인정보보호법 제23조 상의 민감정보, 동법 시행령 제19조에 따른 주민등록번호, 여권번호, 운전면허의 면허번호, 외국인등록번호나 그 밖의 개인정보가 포함된 자료를 처리할 수 있는 근거를 마련해 두고 있는 것은 일반법으로서의 개인정보보호법을 상정하고 만들어진 것이라고 해석될 수 있기 때문이다.⁴

형사사법정보 상의 비식별 정보와 개인정보를 비식별화 또는 익명화시켜 치안 기술 고도화에 활용할 수 있는지를 판단하기 위해서는 두 법상의 관계 외에도 각각의 조문을 검토할 필요가 있다. 개인정보보호법 제18조는 개인정보의 목적 외 이용을 금지하고 있으며, 형사절차전자화법 제6조 제3항은 형사사법업무 처리 외의 목적으로 형사사법정보를 수집·저장 또는 이용”을 금지하고 있다. 여기서 금지된 “형사사법업무 처리 외의 목적”에서 “형사사법업무”란 동법 제2조 제1호에서는 수사, 공소, 공판, 재판의 집행 등 형사사건의 처리와 관련된 업무라고 정의한다. 이는 협의로 보면 “개별적인 형사사건 처리”를 의미하고, 광의로 보면 수사정책 수립, 통계작성·관리, 범죄정보 분석 등 수사 행정 업무까지 포함하는 형사사건의 처리와 “관련된” 업무로 본다.⁵ 법률에 “관련된 업무”라고 명시되어 있을 뿐만 아니라 이미 형사사법 업무가 형사사법정보시스템을 통해 처리되고 있는 현실을 고려할 때 광의의 개념으로 해석하는 것이 타당하며,⁶ 국가 사회의 안녕과 질서를 유지·보전하고자 하는 스마트 치안 기술개발에 개인을 식별할 수 없는 형사사법정보를 활용하는 것은 가명정보의 도입을 통해 개인정보의 활용을 도모하는 개인정보보호법의 개정 취지와도 부합된다고 판단된다.

경찰청은 형사사법정보시스템 구축 당시 치안 분야 과학기술 활성화를 위하여 정보주체의 별도 동의, 타법 특별 규정, 학술연구 등의 목적 등 개인정보보호법에 존재하는 목적 외 이용의 예외를 두는 것도 고려하였지만 최종적인 입법과정에는 개인정보의 오남용을 우려하여 포함되지 못하였다.⁷ 이에 따라 현재 정보 활용에 대한 명시적인 근거는 없고 해석이 필요한 상황이 되었다. 그 외에도 개인정보를 포함한 형사사법정보의 보호와 관련하여서는 개인정보보호법 제3조의 제원칙이 준수되어야 함은 물론이다.

4 동 취지의 해석으로 박민우, 형사절차전자화법 제6조 제3항의 해석과 형사사법정보의 목적 외 사용, 저스티스, 2017, 207-208면.

5 김기범, 형사사법정보의 이용·제공 실태 및 입법적 개선방안, 법학연구, 제16권 1호, 2016, 124-125면.

6 앞의 논문, 124-125면.

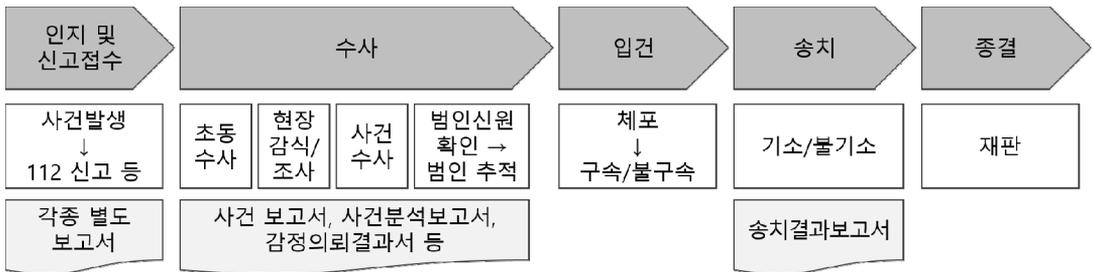
7 앞의 논문, 126면.

이미 피의자의 신상 및 범행 정보 등은 여죄 수사, 재범 방지, 범죄예방 등 수사기관 활동에 활용되고 있으며, 프로파일링 기법의 경우 목적 범위 내 ‘이용’에 포함된다고 해석되기도 하여⁸ 여기에 빅데이터 기술과 접목이 근본적으로 제외되는 것은 아니라고 할 것이다. 따라서 형사사법정보의 제약적 특수성을 고려하면서 스마트 치안 기술에 적합한 데이터 확보 방안의 모색이 필요하며, 이를 위해 이하에서 먼저 형사사법업무의 흐름에 따라 생성되는 정보의 속성을 파악해 보기로 한다.

2. 형사사법업무의 흐름도 및 생성 정보

현재 우리나라 경찰은 형사사법절차의 전자화를 촉진하기 위해 형사사법정보시스템인 형사사법포털(KICS)을 운영 및 관리하고 있다. KICS는 최근 단순한 조회, 검색, 저장의 목적이 아니라 고도화 사업을 통하여 프로파일링, 범죄예방 등 다양한 빅데이터 기술의 플랫폼으로의 활용을 목표로 운영되고 있다. 아래는 각 수사단계에서 각 수사관 또는 분석관을 통해 KICS에 등록되는 문서의 유형을 정리한 경찰청 형사사법절차의 흐름도이다.

〈그림 1〉 형사사법절차 흐름도



가장 먼저 피해자의 신고·고발, 범인의 자수, 수사기관의 자체인지 등을 통한 사건인지와 이와 관련된 보고서가 각 담당 수사관에 의해 생성된다. 이후 수사가 개시되어 수사 관련 보고서가 생성되며 발생한 범죄행위와 범인과 관련된 수많은 정보가 수집된다. 수사 중 작성하게 되는 수사보고서는 수사관들이 의견 및 수사 방향에 대해 작성한 의견 진술형, 새로운 증거 획득 및 분석 사항을 기재한 자료첨부형, 실황조사형 그리고 참고인 또는 피고인의 진술에 관하여 작성하는 진술 청취형 수사보고서가 있으며 감정인 등에 의한 감정서 역시 수사자료로 포함되게 된다.⁹ 입건 후 사법경찰이 확보한 수사기록과 증거물이 검찰에 송치되며 송치결과보고서가 작성된다.

KICS의 대상자 정보란은 피해자와 피의자의 성명, 성별, 직업, 주소, 성장배경 등을 구체적으로

8 김기범, 형사사법정보의 이용·제공 실태 및 입법적 개선방안, 134면.

9 조광훈, 각 유형별 수사보고서의 증거능력에 관한 검토, 형사법의 신동향, 제47호, 2015, 110면.

기재하고 있으며(〈표 1〉¹⁰ 참조) 공범 추적이나 여죄 추적을 위해서라도 수사대상자를 조사하며 알아낸 정보를 입력하도록 권장하고 있다. 단순히 성명이나 주민등록번호와 같은 식별정보를 넘어 종교, 과거 병력, 성적 습관 등 민감정보로 분류될 수 있는 사항을 작성하게 되어 형사사건절차에서 생산되는 문서 대다수가 개인정보보호법 상 민감정보 또는 고유식별번호를 포함하고 있다고 볼 수 있다.

〈표 1〉 KICS 생성정보에 속하는 개인정보 유형 비교표

	생성정보	KICS				그외	개인정보 유형
		KICS	사건접수	피해통보표	수법원지	Geo Pros	
사건정보	사건정보	○	○	○	○	○	x:
	피해자	○	○	○	○	○	●
	유류물	○	○	○	-	-	◎ (생체정보)
	피해품	○	○	○	○	○	x:
	범행도구	○	○	○	○	○	x:
	용의자	○	○	○	-	○	● (신원 확인)
	피의자	○	-	-	○	○	●
	유치정보	○	-	-	-	○	◎ (고유식별정보)
	검거정보	○	-	-	-	○	x:
행태정보	범죄행동	○	○	-	-	○	◎ (민감정보)
	수법정보	○	-	-	○	○	◎ (민감정보)
통계정보	발생원표	○	-	-	-	○	x
	피의자원표	○	-	-	-	-	●
	검거원표	○	-	-	-	-	x

○: 해당함, -: 해당 없음, x: 비개인형사법정보, ◎: 일반 + 고유식별번호/민감정보, ●: 일반 + 민감 + 고유식별번호

10 〈표 1〉과 〈표 2〉의 내용은 수사관과 인터뷰 내용을 토대로 작성하였음.

〈표 2〉 KICS 생성정보 설명 및 예시

	생성정보	설명	정보 예시
사건정보	사건정보	발생한 사건에 대한 기본 정보	관서코드, 수사단서, 범죄수법, 죄명, 발생지 주소
	피해자	피해자와 관련된 정보	피해자 성명, 성별, 주민번호, 신장, 과거병력, 외모특징 등
	유류물	범행현장에 남은 증거물 등	유류지문, 필적, 혈흔, 족적 등
	피해품	피해품에 대한 정보	피해품 코드, 모델명, 피해액 등
	범행도구	범행에 사용된 도구 정보	범행도구에 대한 상세 기술, 획득 방법 등
	용의자	용의자와 관련된 정보	성별, 국적, 연령, 인상(신장, 얼굴형, 두발 등)
	피의자	피의자와 관련된 정보	성명, 성별, 종교, 전과여부, 가정환경, 음주횟수, 평상시 이동수단, 성적습관 등
	유치정보	유치인관리를 위한 정보	유치인 성명, 주민번호, 체포일시, 입감일시 등
	검거정보	피의자 검거 여부 및 관련 정보	기수미수명, 검거일시, 검거단서, 범행시 정신상태, 마약류 사용, 검거인원 등
행태정보	범죄행동	범행전/중/후로 구분하며 범행 구성요건 외 구체적인 정황에 대한 정보	(전) 범행계획, 동기 (중) 침입여부, 공격횟수, 피해자와의 대화 내용 (후) 범행 후 신고 여부, 범행 후 도주방법
	수법정보	범행당시 주요수법 및 행동 특징 정보	범죄 수단(보이스피싱 등), 범행시간대, 경로 등
통계정보	발생원표	원표정보와 피해, 사건개요, 도구, 동기 등에 대한 통계원표	발생관서명, 사건번호, 발생지정보, 검거여부, 죄종, 피해정보(신체/재산), 상해정보, 사망자 수, 피해금액 등
	피의자원표	피의자 관련 통계원표	성명, 주민등록번호, 성별 코드, 직업코드, 국적코드, 혼인관계 코드, 부모 관계 코드, 자백여부 코드 등
	검거원표	검거 관련 통계원표	죄명, 범죄수법, 검거단서, 검거인원명수 등

Ⅲ. 형사사법정보의 국내외 활용 실태

1. 국내 형사사법정보의 공개 및 활용 현황

수사 관련 정보를 담고 있는 형사사법정보의 대부분은 개인과 관련된 정보를 포함하고 있고, 별

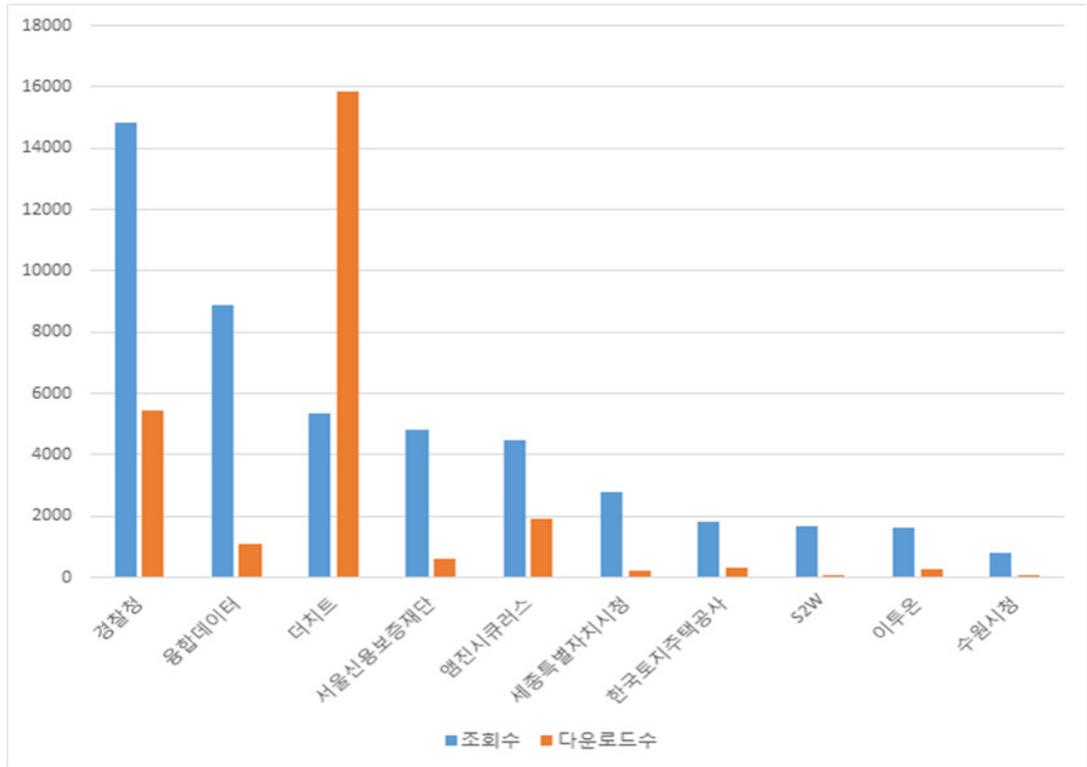
도의 예외 조항이 없는 상황에서 형사절차전자화법 제14조¹¹를 엄격하게 해석하여 형사사법정보를 적극적으로 활용하지 못하고 있다. 경찰청 역시 국제치안산업박람회 개최, 치안맞춤형 솔루션 개발을 위한 폴리스랩 구성, 형사수사업무를 지원하는 인공지능 연구과제를 공고하고 지원하는 등 스마트 치안 환경 마련을 위한 다양한 정책을 활발히 추진하고 있으나, 위와 같은 이유로 개발에 필요한 업무정보를 제한 없이 제공할 수 없는 딜레마에 빠져있다. 예를 들어 안면인식기술을 활용한 지능형 CCTV 개발이나¹² 사건 관련자들의 신용정보·통화 내역 등을 연결하여 사용하는 사회관계망 분석 도구의 활용, 금융거래 내역·의심 전화번호 등을 활용한 보이스포싱 대응 시스템의 개발¹³과 효용성 검증에는 반드시 실제 수사 데이터가 필수적임에도 불구하고 형사사법정보 활용에 대한 법 제도의 미흡으로 효율적인 개발이 어려운 상황인 것이다.

다행히도 중앙정부·지방자치단체 및 공공기관이 보유·관리하는 공공데이터를 일반 국민이 활용할 수 있도록 다양한 형태로 데이터를 개방·제공하는 공공데이터의 제공 및 이용 활성화에 관한 법률은 경찰청의 이러한 문제점을 일부 해소해주고 있다. 현재 경찰청은 동법에 근거하여 디지털 뉴딜의 일부로 ‘스마트치안 빅데이터 플랫폼’을 운영하고 있다.

스마트치안 빅데이터 플랫폼은 경찰의 약 145억 건에 달하는 누적된 정형 데이터를 중심으로 공공, 민간이 협력해 시작한 치안을 위한 맞춤형 빅데이터 서비스다. 해당 플랫폼에서는 공공데이터 생산과 치안 분야 과학기술 활성화를 위해 무료 또는 일정의 사용료를 지불한 후 각 기관의 데이터를 받을 수 있으며 그 외 연구자가 필요한 데이터를 플랫폼을 통해 직접 신청할 수 있다. 아래 그림2는 경찰 데이터에 대한 다양한 수요와 높은 활용도를 보여주기 위해 플랫폼에 공개된 400여 건의 데이터 중 상위 10%(약 40개)를 뽑아 기관별로 각각 조회 수 및 다운로드 수의 합계를 정리하였다.

-
- 11 형사절차전자화법 제14조(형사사법정보의 보호 및 유출금지) ① 형사사법업무 처리기관은 형사사법업무를 처리할 때 형사사법정보가 분실, 도난, 유출, 변조 또는 훼손되지 아니하도록 안전성 확보에 필요한 조치를 하여야 한다.
 ② 형사사법업무에 종사하는 사람 또는 제8조제3항에 따라 시스템의 지원업무를 위탁받아 그 업무에 종사하는 사람은 권한 없이 다른 기관 또는 다른 사람이 관리하는 형사사법정보를 열람, 복사 또는 전송하여서는 아니 된다.
 ③ 형사사법업무에 종사하거나 종사하였던 사람 또는 제8조제3항에 따라 시스템의 지원업무를 위탁받아 그 업무에 종사하거나 종사하였던 사람은 직무상 알게 된 형사사법정보를 누설하거나 권한 없이 처리하거나 타인이 이용하도록 제공하는 등 부당한 목적으로 사용하여서는 아니 된다.
- 12 손재호·이창무, 치안 R&D의 생체정보 활용의 문제점과 개선방안, 한국경호경비학회, 제65호, 2020, 331면.
 13 장광호·김희두, 비정형 데이터 네트워크 기법의 수사 활용 연구:보이스피싱 대응 수사 지원을 중심으로, 경찰학연구, 제20권 제2호, 2020, 100면.

〈그림 2〉 스마트치안 빅데이터 플랫폼 기관별 조회수, 다운로드 수 합계 (상위 10% 기준)



〈그림 2〉에서 사용자들에게 가장 큰 관심을 받은 데이터는 경찰청에서 제공한 데이터이며, 스마트치안 빅데이터 플랫폼 측에서 치안 데이터를 기반으로 가공한 융합데이터가 두 번째로 높은 조회수를 기록하고 있다. 현재 플랫폼상에서 경찰청이 직접 제공하는 데이터는 전체 자료 중 약 7.8%(32/407)로 대부분이 경찰서별로 범주화된 통계데이터이다. 이 중 가장 조회 수와 다운로드 수를 기록한 데이터는 전국 112 신고 접수 데이터로 경찰서를 기준으로 신고건수, 신고자 성별 건수, 긴급유형별 건수, 연도 등에 따른 통계자료로 이러한 데이터가 활용될 수 있는 분야는 재난관리, 보건, 교통, 의료 등 치안에 국한되지 않는다.

데이터 활용도를 가늠할 수 있는 다운로드 수로 보면, 플랫폼에서 가장 높은 다운로드 수를 기록한 데이터는 민간기관에서 공개한 ‘스팸 SMS 내용’ 데이터다. 해당 데이터는 스팸 예방 사업을 하는 민간 기업이 직접 수집하고 있는 일별 스팸 문자열 데이터다. 직접 수집한 데이터기에 피해자의 원시데이터(raw data)를 받아 사용할 수 있어 특정지표를 기준으로 한차례 가공된 통계데이터 보다는 새로운 패턴을 발견하는 등 활용도가 크다는 장점이 있다.¹⁴ 민간기관 중 높은 다운로드 수

14 원데이터의 중요성과 관련하여 2021년 치안정책연구소에서 진행한 〈치안 관련 데이터의 수요 검증〉 전문가

를 보이는 연합뉴스의 ‘비리사건’ 데이터¹⁵의 경우 기사의 발행일자, 분야, 작성처 외 제목·부제목 데이터와 자동으로 추출한 키워드, 인물 등을 추가하여 데이터의 질을 한층 높였다고 볼 수 있다.

스마트치안 빅데이터 플랫폼은 경찰청에서 제공하는 통계데이터에 그치지 않고 여러 공공기관의 협약을 통해 사이버보안, 치안, 지역환경, 재난 등과 관련된 데이터를 공개하여 치안 데이터 시장 활성화에 기여하고 있다. 특히 민간기업과 협업으로 치안 관련 원데이터(raw data)를 생산하여 제공함으로써 양질의 치안 데이터에 대한 수요를 충족시키고 있다.

외부에 공개되지 않고 있지만 경찰청 내부적으로 치안 데이터를 활용하여 스마트 치안 기술의 잠재성을 보여주는 사례도 있다. 9대 범죄사건 정보와 지리적 정보를 연계하여 범죄위험지수 및 범죄다발지를 분석하는 지리적 프로파일링 시스템 GeoPros, 네트워크 분석 기법을 활용하여 범죄 핵심 개체를 선별, 데이터 간의 관계를 다차원 시각화 분석을 하는 IBM의 i2¹⁶, KICS에 입력된 데이터와 공공데이터를 융합하여 범죄 경향 분석, 유사 사건 추천을 위해 개발된 클루(이하 CLUE)에도 수사 데이터가 사용되었다.¹⁷

다만 i2의 경우, 우리나라 통신사와 금융기관에서 받은 데이터가 별도의 표준화 과정 없이 사용할 수가 없어 국내에서 사용하기 위한 전처리 도구(i1)의 개발이 추가적으로 진행되었으며, 이 전처리 과정은 기능 고도화를 위해 2021년 신규 치안 서비스 개발 과제로 채택되기도 하였다.¹⁸ CLUE의 경우, 빅데이터 분석 도구인 만큼 날씨, 인구 등 공공데이터와 대량의 치안 데이터를 분석하고자 했으나 개발 단계에서는 선별된 일부 수사 데이터만 비식별화를 하여 사용하거나 유사한 데이터를 자체적으로 생산하여 개발하는 등 데이터 사용에 있어 어려움을 겪었다. 위와 같은 현황과 사례를 통해서 결국 스마트 치안 기술 개발을 위해 형사사법정보가 필수적이라는 사실과 개발을 위해서는 학습 과정에서뿐만 아니라 분석 단계에서도 실데이터에 적용할 수 있는 정형화 방법 등을 위해 지속적인 형사사법정보 공급이 이루어질 필요가 있음을 볼 수 있다.

인터뷰의 한 관계자는 “어느 정도 원시수준의 데이터여야 기존의 IT서비스와 매핑할 수 있어 완전 가공이 아닌 추가적인 분석이 가능한 데이터가 제공되어야 함을 강조하였다. (치안정책연구소, 스마트 치안 빅데이터 플랫폼 로드맵과 실천전략 수립, 2021. 1., 121면.)

- 15 2022년 11월 10일 기준 데이터별 다운로드 수로 4위(1,672건)를 차지하였다. (1위: 스팸 SMS 내용(더치트, 10,071건), 2위: 금융사기에 이용된 데이터_증권(더치트, 3,254건), 3위: 전국112신고접수(경찰청, 3,217건))
- 16 통계교육원, “범죄자를 추적하는 데이터: 사회연결망분석을 활용한 범죄수사”, 통계집중탐구, 2020, http://sti.kostat.go.kr/window/2020b/main/2020_win_07.html (검색 2022. 11. 10.).
- 17 IITP, 스마트 치안 최근 동향과 시사점, S&T GPS, 이슈분석 208호, 2022.
- 18 과학기술정보통신부, 인공지능 기반 아동학대 영상 분석 시스템 개발 등 과학기술을 활용한 치안서비스 고도화 추진, 2022. 7. 23. <https://www.korea.kr/news/pressReleaseView.do?newsId=156517781> (검색 2022. 11. 8.). ‘차세대 지능형 수사자료 분석 솔루션’은 i2의 이용 편의성을 높이고, 룰(Rule) 기반이기에 예외사항에 취약할 수 있는 i1의 한계를 넘어 지능형 분석 도구화를 목표로 하고 있다.

2. 국외 치안데이터 활용 사례

이상과 같은 국내 치안데이터의 활용과 현황에 많은 시사점을 제시해주는 해외 사례로는 치안 분야 기술과 기술 고도화 과정에 적극적으로 데이터 활용하고 있는 미국의 최대 경찰 조직인 뉴욕경찰(NYPD)이 있다. 뉴욕경찰(NYPD)은 2016년부터 머신러닝 기반 사건 패턴 인지 시스템(Patternizr)을 개발하여 예측 치안(Predictive Policing)과 스마트 범죄 분석에 적극 활용하고 있다.¹⁹

Patternizr는 강도 및 절도 사건이 발생하면 뉴욕 경찰국의 통합 데이터베이스에서 범죄수법 분석을 통한 유사 사건을 검색하여 범인 검거를 지원하는 프로그램이다.²⁰ 이 시스템을 구축하기 위하여 뉴욕경찰은 10년 치의 누적된 강도 및 절도 유형 분석 데이터를 시스템에 학습시켰고, 전문 분석가들을 두어 범죄정보 패턴 인지 시스템을 개발하고 정확도를 높이는 업무를 지속하고 있다.²¹

뉴욕경찰의 유형분석 툴인 Patternizr와 비슷한 시스템을 국내에서도 R&D사업 (CLUE 사업)의 일환으로 개발을 시도한 경험이 있다. CLUE사업은 수사 항목을 비교하여 과거의 유사사건을 찾아내고 치안데이터와 공공데이터를 융합하여 범죄분석 예측 시스템 개발을 목표로 하였다.²² 당시 CLUE는 2011~2015년에 발생한 침입절도 사건의 수사 결과보고서를 학습데이터 대상으로 두었는데²³, 개인정보보호 등의 문제로 수사결과보고서 중 범죄사실 부분만을 추출 후 전처리하여 학습데이터로 사용할 수밖에 없었다.²⁴ Patternizr는 (1) 대다수 범죄는 재범률이 높으며, (2) 범죄수법(modus operandi)에는 패턴이 있다는 범죄학 이론에 착안하여 동적으로 변하는 범죄수법의 패턴도 찾을 수 있는 인공지능 모델을 개발하고자 하였다. 개발자들은 분석 대상으로 강도(robbery)·절도(burglary)·중절도(grand larceny) 세 가지 범죄 유형 데이터를 선정하였는데 그 이유 중 하나로 충분한 패턴 데이터의 누적을 들었다.²⁵ 실제로 Patternizr의 개발자들은 각 범죄 유형에 맞춰 세 가지 결정트리(decision-tree) 모델을 학습하였으며, 2006년부터 2015년 동안

19 Chohlas-Wood, A., & Levine, E., *A recommendation engine to aid in identifying crime patterns*, INFORMS Journal on Applied Analytics, 49(2), 2019, 1면.

20 앞의 논문, 1면.

21 앞의 논문, 1면.

22 임기창, “빅데이터 수사’ 시대 열리나…경찰, 2019년 ‘클루’ 시범운영”, 연합뉴스, 2017. 12. 9. <https://www.yna.co.kr/view/AKR20171209026800004> (검색 2022. 11. 10.).

23 강명중, 침입절도 범죄 간 유사도 측정 알고리즘의 개발 및 응용에 관한 연구, 한국경찰학회보, 제20권 제6호, 2018, 60면.

24 임기창, “빅데이터 수사’ 시대 열리나…경찰, 2019년 ‘클루’ 시범운영”, 연합뉴스, 2017. 12. 9.

25 Chohlas-Wood, A., & Levine, E., *A recommendation engine to aid in identifying crime patterns*, 1면.

전문 분석가들이 구축한 각각 강도·절도·중절도 세 가지 범죄 유형에 대한 10,000건의 사건 패턴, 그리고 10년간 누적되었던 신고서(Complaint) 20,000건~40,000건 중 재범 발생률이 높은 30,000건을 사용하였다.²⁶ 종합적으로 Patternizr의 모델 학습에는 사건 발생 위치 정보, 일시, 범행 장소 정보, 용의자 정보 외에도 용의자의 진술, 피해자의 행동, 도주 방법 등의 정보가 포함된 비정형 데이터로 구성된 총 39가지의 패턴이 활용되었다.²⁷ 그뿐만 아니라 뉴욕경찰에서는 Patternizr의 지속적인 발전을 위해 기존의 연구원 외에도 100여 명의 민간인 분석 전문가를 추가 고용을 하였다. 이처럼 Patternizr에 투입된 학습데이터와 인적 자원을 비교하였을 때 CLUE사업에 활용된 학습데이터의 양과 품질은 매우 부족했다고 볼 수 있다.

또한, 뉴욕경찰은 2012년부터 Microsoft社와 함께 개발한 세계 최대 규모의 감시 시스템인 DAS (Domain Awareness System)을 통해 공공 CCTV, 차량번호, 911 신고접수 현황 등을 수집하여 경찰 데이터베이스와 연계하여 범죄예측에 활용하고 있다.²⁸ 우리나라에서도 DAS와 유사하게 지역별 범죄 위험도와 범죄 발생 건수를 예측하는 시스템인 프리카스(Pre-Cas)를 도입한 바가 있다.²⁹ 시범 운영 당시 프리카스는 112신고 건수, 경찰관 수, 유흥시설 수 등이 포함된 치안데이터 및 건물 유형, 지연인구 등 공공데이터를 활용하여 개발하였으나, 경찰청에서만 전체적 통계 데이터를 사용할 수 있어 지역맞춤형 치안이 목표임에도 불구하고 지역적 특성을 고려한 범죄예방 정책 수립이 어려운 문제점이 노출되기도 하였다.³⁰

26 앞의 논문, 2면.

27 앞의 논문, 2면. 수집했던 데이터 중 식별가능정보인 성별, 인종, 상세한 범행 발생 장소는 특정 지역, 인종에 대한 편향을 담을 수 있어 제외하였다. 단, 식별가능정보를 제거하는 것으로 다른 예측 서비스와 차이점을 두고 편향적이지 않다("immune to bias")는 주장은 논의될 필요성이 있어 보인다. Patternizr의 성능을 분석한 한 연구에서는 민감 데이터 제거 작업만으로 예측 모델의 공정성(Fairness)을 보장할 수 없으며 독립기구를 통한 감사 등을 통해 투명성을 확보해야 한다고 주장하였다 (Griffard, M., *A bias-free predictive too?: An evaluation of the NYPD's Patternizr*, Fordham Urban Law Journal, 47, 2019, 74면).

28 임운식, 스마트 치안과 관련된 해외 사례, 수사연구 8월호, 2018, 37면.

29 박홍용, "경찰 'AI로 범죄 예측' 내달 시행", 서울경제, 2021. 4. 29.
<https://www.sedaily.com/NewsView/22L9TJGOWU> (검색 2022. 11. 5.).

30 장수인, "프리카스 시스템 정확도 등 보완 이뤄져야", 전북도민일보, 2021. 11. 2. https://www.domin.co.kr/news/articleView.html?idxno=1359313&sc_section_code=S1N8 (검색 2022. 11. 10.). 현재는 마약·전세사기 등 사회적으로 문제가 되는 분야의 예방에도 확장 도입되기도 하였다. 이종현, "경찰 범죄예측 분석 시스템 '프리카스'로 마약·전세사기 잡는다", 조선비즈, 2022. 10. 18., https://biz.chosun.com/topics/topics_social/2022/10/18/FJJD7DMJWZGJXPCVQTJ74VCGJM/

IV. 형사사법정보활용상의 문제점

1. 처리 가능한 형사사법정보량의 부족

앞서 확인한 바와 같이 경찰청에서 직접 수행하는 연구과제임에도 불구하고 연구 및 개발을 위해 필수적인 양질의 치안 데이터 공급은 어려운 상황이다. 이는 형사사법정보에 대한 분명한 정의나 이용에 대한 명확한 법적 근거 부족으로 데이터 활용에 소극적으로 대응하는 이유도 있지만, 기본적으로 비식별화하여 제공할 수 있는 데이터의 수량이 많지 않다는 것도 문제라고 볼 수도 있다. 현재 형사사법정보는 서면, 이미지, 영상, 녹음 형태 등 다양한 형태의 기록물로 보관되고 있으면서³¹ KICS에 보고서를 올리게 됨에도 불구하고 실무에서는 여전히 사건마다 종이 기록이 수백에서 수천 페이지씩 생성되고 있는 실정이다. 전자화되지 않는 서면자료에 대해 비식별화 조치는 진행이 어려울뿐더러 이미지나 녹음 자료의 비식별화 역시 별도의 난독화 기술이 요구되기 때문에 현재 기술로 처리할 수 있는 형사사법정보는 실제 생산되는 정보량에 비해 극히 적다. 기술적인 측면에서의 치안 데이터 활용은 치안 데이터 전면 전자화를 통해 전반적인 데이터량 확보를 전제로 하는데 현재로서는 처리 가능한 형사사법정보가 부족하다는 문제의 개선도 필요하다.

2. 형사사법도메인(domain)의 원천적 접근 봉쇄

현재 경찰은 데이터별로 접근 권한에 차별을 두는 것이 아니라 형사사법도메인 전체를 보호대상으로 보고 제한하는 방식을 채택하고 있다. 즉, 특정데이터가 범죄수사의 영역(domain)에 해당되기만 하면 데이터의 성질과 관계없이 보호 또는 접근금지의 대상이 되는 것이다. 이러한 정책은 기본적으로 데이터의 신뢰성(trustworthiness) 확보보다는 보안성(security) 확보에만 중점을 두기 때문이며 보호 업무를 담당하는 주체도 개방형보다는 폐쇄성의 관점에서 데이터를 관리하게 된다. 원칙적으로 접근을 제한하기에 형사사법절차에서 개인정보보호는 별도의 과정을 거친다기보다는 형사사법정보 처리절차에 내포된 견제와 감시 장치로 확보되고 있다.³² 즉, 데이터가 경찰, 검찰, 법원을 거치게 될 때 증거 관리 연속성과 책임의 주체와 접근목적이 명확히 기록됨으로써 개인정보

31 김미령 외, 형사사법정보의 빅데이터 활용방안 연구: 구조화 범주화 관점으로. 정보관리학회지, 제36권 제4호, 2019, 267면. 이 논문에서 진행된 전문가 자문에서는 형사사법정보가 대부분 비정형 데이터이거나 데이터 형태인 자료도 구조화되지 않아 빅데이터로 활용하는 데 어려움이 있으며, '형사사법정보의 빅데이터로서 활용가치'가 있으려면 데이터화가 시급하다고 밝혔다.

32 이만형, 형사재판절차에서의 개인정보 보호에 관한 연구, JPRI 연구보고서 2021-04, 사법정책연구원, 2021, 74면.

의 오남용을 예방한다고 보는 것이다. 그러나 이러한 막연한 내부 운영방식은 치안 기술 고도화 지원을 위한 외부로의 데이터 제공이 이어지면서 더이상 의미 있는 방침이라고 볼 수 없다.

형사사법분야의 도메인 단위의 접근 통제 정책은 초연결 시대에 들어서기 위한 ‘데이터 중심’의 현 시대적 상황과 정면으로 충돌한다.³³ 4차산업혁명위원회의 보고서에 따르면 ‘도메인 중심의 정책’은 과거 인터넷을 민간·공공·금융 등의 영역으로 구분해 관리하는 방식으로 그 대표적인 사례로 2007년에 시작된 ‘망 분리 정책’이 있다.³⁴ 동 보고서에 의하면 “망 분리란 공공 및 금융 기관의 업무용 망(내부망 또는 인트라넷)을 인터넷과 물리적으로 단절시킴으로써 외부 해커의 침입을 원천 봉쇄하는 것으로, 해킹 사고를 감소시키는 효과가 있는 것은 사실이다.”라고 하면서도³⁵, 이러한 망 분리 정책이 주요 데이터의 생산·유통 등을 해당 기관의 도메인으로만 제한하게 되어 “모든 것이 네트워크에 연결(초연결)되어 있고, 데이터는 활발하게 공유·활용돼야 한다.”는 4차 산업혁명의 기본 철학과 상충됨을 밝히고 있다.³⁶

이러한 견지에서 4차산업혁명위원회는 도메인 중심정책은 ‘인공지능, 빅데이터 등 4차 산업혁명 시대 관련 산업 육성의 심각한 걸림돌’이라는 문제점을 인식하고,³⁷ 최근 소위 ‘데이터 3법’ 개정을 통해 걸림돌을 제거하기도 하였다. 이와 관련하여 4차산업혁명위원회는 생산되는 데이터를 중요도에 따른 전자데이터 분류 체계를 마련하고 데이터 중요도에 따라 망 분리 등의 보안 정책 및 컨트롤 타워 체계를 재정립할 것을 권고하고 있기도 하다.³⁸

그러나 형사사법 분야에서는 이러한 데이터 정책 패러다임의 변화에 대한 논의가 없는 것으로 보인다. 연구과제의 성공적인 도입에 데이터의 확보 및 이용이 절실한 상황임에도 데이터 사용 허가가 느려 연구가 지연되는 경우가 발생할 뿐만 아니라, 연구과제에 명시된 자료조차 반출이 불가하여 연구기관에서 직원을 파견하여 현장에서 분석해야 하는 사례도 있다. 그 외에도 데이터 제공이 이뤄지지 않아 별도의 예산을 투자하여 가상의 시나리오를 제작 후 학습데이터로 활용하는 경우도 빈번하게 발생한다. 이미 추진하고 있는 수많은 치안 R&D과제들과 현장에 도입된 도구들이 형사사법 데이터 없이 성공적인 진행 및 발전을 할 수 없음에도 불구하고 별다른 조치가 마련되지 못하고 있어 논의를 통한 개선방안의 마련이 필요하다.

33 4차산업혁명위원회, 4차 산업혁명 대정부 권고안 보고서(사이버보안), 2019, 70면.

34 앞의 보고서, 70면.

35 앞의 보고서, 70면.

36 앞의 보고서, 70면.

37 앞의 보고서, 70면.

38 앞의 보고서, 70면.

3. 비식별화 도구 고도화 필요성

그 외의 문제로 연구를 위해 전처리를 할 때 사용하는 비식별화 도구의 고도화가 있다. 현재 판결문을 판결문 인터넷 열람 사이트 등 다양한 경로를 통해 공개하고 있는 법원 데이터의 예를 들면 원칙적으로 사건관계인의 성명을 비롯하여 (1) 아이디, 닉네임과 같이 성명에 준하는 정보, (2) 연락처, (3) 금융정보, (4) 차량등록번호 등 개인을 특정할 수 있는 정보를 비실명처리하고 있다.³⁹ 이러한 정보는 알파벳 대문자 A, B, C 등으로 대체 표기를 하고 있으며 이는 다음 심급 판결이 있으면 동일하게 유지하도록 권장되고 있다(비실명예규 제6조).

현재 법원에서 사용하는 비실명처리 방법은 반자동화라고 볼 수 있다. 1차적으로 지능형 비식별 시스템에서 추출 대상을 정의하여 패턴 자동화를 통해 처리하게 되며, 2차 검수에서 매년 새롭게 선정되는 비실명화 사업소를 통해 민간업체 종사자와 법원 인원 약 90명이 투입되어 오류 교정 및 처리하고 있다.⁴⁰ 판결문은 공공데이터 대상인만큼 공개가 원칙이며 이에 따라 사건 관계자의 개인정보보호를 위해 비실명처리가 필수적이라고 볼 수 있으나, 공개할 의무가 없는 형사사법정보의 경우 검수 시 필요한 인원 투입이 불필요한 부담으로 다가올 수 있다. 따라서 검수 인력을 최소화할 수 있는 비식별화 도구가 필요하지만, 현재 사용 중인 비식별처리 시스템은 특정한 규칙으로 키워드를 추출하고 단순한 기계학습⁴¹으로 작동하는 구조라 문맥의 의미를 파악하지 못하여 소송상 지위를 구별하거나 심급별 지위 구분 또는 생략처리가 어렵다.⁴² 경찰청에서 i2의 이용 편의성을 높이기 위해 국내 금융·통신 데이터를 표준화하는 전처리 도구를 개발하게 되었듯이 비식별화 과정에서도 최적화된 지원 도구가 없다면 적극적인 데이터 유통을 통해 발전해나가야 할 스마트 치안 기술개발에 걸림돌이 될 수 있다.

V. 결론: 개인정보보호를 고려한 정보 활용 접근 방안

앞서 제기된 형사사법정보 활용상의 문제는 다양한 방법으로 해소될 수 있을 것이다. 여러 정책

39 이민형, 형사재판절차에서의 개인정보보호에 관한 연구, 351면.

40 앞의 논문, 351면. 2020년 기준 50만 건의 판결문 비식별화를 위해 약 31억 원의 예산이 투입되었다.

41 정해진 패턴과 일치하는 문자를 탐색하는 정규표현식(regular expression), 명사·동사·조사 등을 추출하는 형태소 분석기 등을 조합하여 규칙(Rule)을 만든 후 키워드를 추출, 이러한 단어의 특징(feature)을 기반으로 학습된 분류모델을 말한다. 이러한 모델은 단어 단위의 학습만 가능할 뿐, 전체적인 문장의 구조, 의미 등을 학습하지 못해 문맥이나 의도에 따라 달라지는 핵심 정보를 추출하지 못하는 한계가 있다.

42 이민형, 형사재판절차에서의 개인정보보호에 관한 연구, 357면.

적, 기술적 방법 중에서 어떤 방식이 가장 효율적이고 효과적인지는 법과 실무, 현실적인 기술 실현가능성 등의 판단이 요구된다. 그럼에도 불구하고 현재 스마트 치안 빅데이터 플랫폼, 경찰청의 범죄통계시스템(CSS)에서 공개되는 형사사법정보의 형태를 고려하였을 때 전자화된 치안 데이터를 현행법 내에서 최대한 활용할 수 있는 방안은 다양한 범주화 데이터를 생성하여 사용하는 것이라고 볼 수 있다. 범주화는 각 행과 열의 데이터를 구조화하여 그 값을 일정한 범위로 묶어 개인정보를 비식별화하는 방식이다. 범주화 데이터는 데이터의 특성에 따라 범주화하는 범위를 조정할 수 있으며 그 범위는 보안 등급에 따라 구체화하여 제공할 수도 있다.⁴³ 특히 데이터를 연구 목적에 따라 범주화를 하게 되면 특정 개인을 식별하지 않으면서도 연구에 필요한 특성들을 뽑아 사용할 수 있다. 예를 들어 살인사건에서 피의자의 동기(motivation)에 관해 연구할 경우, 피해자와의 관계나 환경에 대한 상세한 설명이 들어간 텍스트를 제거하고 ‘치정’이라는 카테고리로 분류하여 다른 변수와 함께 사용하는 방식이다. 이처럼 연구 목적에 따라 정의된 값의 범위로 분류된 정보를 다양한 요소들과 연계하여 활용한다면 유의미한 연구결과를 얻을 수 있다.

두 번째 제기한 도메인 문제에 대해서는 모든 데이터를 영구히 관리하겠다는 ‘도메인 중심’ 정책의 보안성은 데이터의 활용이라는 이익과 상충하여 법의 목적에 어긋나는 광의적 해석으로 이어질 수 있으므로 정확한 규제를 통해 허용할 것은 허용하고 제한할 것은 제어하는 접근 통제 정책이 요구된다. 앞서 KICS의 생성 데이터에 포함된 개인정보 비교(〈표 1〉)를 통해 보았듯이 모든 데이터가 동일한 보호가치를 가지는 것은 아니다. 나아가 4차산업혁명위원회 역시 데이터의 사용용도, 정련도, 개인정보보호 상 중요도 등에 따라 등급을 지정하고, 데이터의 등급에 따라 기밀이 아닌 것으로 분류된 공공·금융분야 데이터에 대해서는 유통 및 활용을 최대한 장려하여 글로벌 데이터산업 육성의 기반을 마련할 것을 제안한 바 있다.⁴⁴ 형사사법 분야에서의 데이터도 ‘도메인 중심’이 아닌 ‘데이터 중심’의 정책을 채택하여 데이터 중요도에 따라 등급화하고 공개 및 활용 여부를 차등화해야 할 것이다. 이때 데이터 중요도에 따라 공개되는 일부 데이터는 인공지능 관련 분석가 혹은 개발자가 직접 전처리 등을 통해 사용 목적에 맞게 활용할 수 있도록 최대한 원시데이터(raw data) 형태로 제공될 필요가 있다. 그동안 공개된 형사사법데이터는 주로 통계분석 목적으로 수정되거나 가공된 형태로 제시됐는데 이러한 방식의 데이터만 제공되는 것은 딥러닝이나 머신러닝용 학습데이터 구축의 장애 요인으로 작용하기 때문이다.

마지막으로 비식별화 도구 고도화와 관련해서는 개체명 인식(NER, Named Entity Recognition) 작업에 학습된 인공지능 모델에 대한 연구가 선행될 필요가 있다고 본다. 개체명 인식이란 사람, 장소, 시간, 기관 등의 기본식된 항목들을 자동으로 추출하는 작업으로 비식별화 시스템에서는 학습된 인공지능 모델이 주어진 텍스트에서 개체를 인식하고 비식별화 규칙에 따라 교체하는 방식으로 진행된다.

43 김미령 외, 형사사법정보의 빅데이터 활용방안 연구: 구조화 범주화 관점으로, 264면.

44 4차산업혁명위원회, 4차 산업혁명 대정부 권고안 보고서(사이버보안), 70면.

최근 자연어 처리 분야에서는 사전학습된 모델의 성능이 활발히 연구되고 있는데 이제는 문장 인식을 넘어 문장 이해(Understanding)를 추구하고 있다. 이와 같이 문맥, 어휘, 의도를 이해하는 기술을 기반으로 개인정보 검출 모델을 개발한다면 사법경찰과 같은 수사기관에서도 개인정보의 침해를 최소화할 수 있다. 더 나아가 이러한 개체명 인식 기술을 비식별화에 사용하는 것을 넘어 개인정보를 제외한 형사사법관련 정보를 추출 및 분석하거나 정보간의 결합을 통한 새로운 형태의 정보를 제공하는 등 관련 정보 전체를 자동으로 조망하도록 구성할 수 있어 형사사법관련 담당자의 전문성을 높이는데 기여할 수 있다. 이러한 방법은 개인정보에 대한 침해위험을 줄이면서도 연구 자료의 가치를 유지할 수 있어 스마트 치안 기술 진흥에 도움이 될 수 있다.

현재 경찰청에서 구성한 치안현장 맞춤형 연구개발사업(폴리스랩 2.0.)을 보면 일부 과제제안요구사항(RFP)에 개인정보 유출 방지 항목이 포함되어 있기는 하지만 기타 기술적 요구사항에 비해 높은 가중치를 가지지는 않고 있다. 이는 최근 유럽연합(EU) 집행 위원회에서 2023년에 시행될 'EU AI 규제안'에 대비해 요청한 대규모 파일럿 과제의 요구사항과 대조적이다.⁴⁵ 디지털 유럽(Digital Europe) 사업의 일부인 동 파일럿 과제는 오픈소스 우선의 기술개발·실제 데이터에 사용 적합성과 EU 개인정보보호 표준 준수를 핵심 목표로 가지며 최종 솔루션의 개인정보처리 방침 준수 입증은 필수요구사항에 포함되어 있다. 이처럼 개인정보보호 준수를 연구의 일부로 돕으로써 과제의 특징을 반영한 구체적인 개인정보보호 방침이 생성될 것이라 예상된다.

개인정보 침해위험에 대한 대책을 마련하지 않고 기술발전만을 우선시하는 연구가 진행되게 되면 보호되어야 할 개인정보가 부수적인 것으로 밀리는 상황이 발생함과 동시에 결과적으로 국내법 체계에서 일원화된 개인정보보호 체계 수립이 어려워질 수 있다. 장기적인 관점에서 개인정보보호는 서비스 완성 후 문제가 발생할 때마다 대응하는 사후관리방식 대신 사전예방방식인 소위 '보안 내재화(Security by Design)'가 효율적이다. 이러한 점들을 고려했을 때 국내에서도 유럽연합의 파일럿 과제처럼 개인정보보호 표준 준수를 중시하는 연구 환경을 구축하고 데이터 활용에 있어 개인정보보호에 대한 방침을 고려하여 형사사법분야에 인공지능기술이 폭넓게 활용될 기반을 닦아야 한다.

ACCEPTANCE INFO

논문접수: 2022. 11. 10. / 심사개시: 2022. 11. 18. / 게재확정: 2022. 12. 5.

⁴⁵ European Commission, Digital Europe Programme - Security (Law enforcement): AI-based pilot, 2022.

참고문헌

I. 논문

1. 국내 논문

- 강명종, 침입절도 범죄 간 유사도 측정 알고리즘의 개발 및 응용에 관한 연구, 한국경찰학회보, 제20권 제6호, 한국경찰학회, 2018.
- 김기범, 형사사법정보의 이용·제공 실태 및 입법적 개선방안. 법학연구, 제16권 1호, 한국법학회, 2016.
- 김미령 외, 형사사법정보의 빅데이터 활용방안 연구: 구조화 범주화 관점으로, 정보관리학회지, 제36권 4호, 한국정보관리학회, 2019.
- 박민우, 형사절차전자화법 제6조 제3항의 해석과 형사사법정보의 목적 외 사용, 저스티스, 제161권, 한국법학원, 2017.
- 손재효·이창무, 치안 R&D의 생체정보 활용의 문제점과 개선방안, 한국경호경비학회, 제65권, 한국경호경비학회, 2020.
- 장광호·김희두, 비정형 데이터 네트워크 기법의 수사 활용 연구:보이스피싱 대응 수사 지원을 중심으로, 경찰학연구, 제20권 제2호, 경찰대학, 2020.
- 조광훈, 각 유형별 수사보고서의 증거능력에 관한 검토, 형사법의 신동향, 제47권, 대검찰청, 2015.

2. 국외 논문

- Chohlas-Wood, A., & Levine, E., *A recommendation engine to aid in identifying crime patterns*, INFORMS Journal on Applied Analytics, 49(2), 2019.
- Griffard, M., *A bias-free predictive too?: An evaluation of the NYPD's Patternizr*, Fordham Urban Law Journal, 47, 2019.

II. 기타

1. 보고서

- 4차산업혁명위원회, 4차 산업혁명 대정부 권고안 보고서(사이버보안), 2019.
- IITP, 스마트 치안 최근 동향과 시사점, S&T GPS, 이슈분석 208호, 2022.
- 오동수 외, 스마트 치안 빅데이터 플랫폼 로드맵과 실천전략 수립, 치안정책연구소 연구용역보고서, 2021. 1.
- 이민형, 형사재판절차에서의 개인정보보호에 관한 연구, JPRI 연구보고서 2021-04, 사법정책연구원, 2021.

2. 국내 언론 보도 및 기타

- 과학기술정보통신부, “인공지능 기반 아동학대 영상 분석 시스템 개발 등 과학기술을 활용한 치안서비스 고도화 추진”, 2022. 7. 23. <https://www.korea.kr/news/pressReleaseView.do?newslid=156517781> (검색 2022. 11. 8.).
- 곽중희, “미래 안전, 스마트치안이 책임진다”, CCTV뉴스, 2022. 10. 13. <https://www.cctvnews.co>.

- kr/news/articleView.html?idxno=233661 (검색 2022. 11. 24.).
- 박광하, “민·관·군·경 협력해 랜섬웨어 대응역량 강화”, 정보통신신문, 2022. 9. 20. <https://www.koit.co.kr/news/articleView.html?idxno=103123> (검색 2022. 11. 24.).
- 박홍용, “경찰 'AI로 범죄 예측' 내달 시행”, 서울경제, 2021. 4. 29. <https://www.sedaily.com/NewsView/22L9TJGOWU> (검색 2022. 11. 5.).
- 이종현, “경찰, 범죄예측·분석 시스템 ‘프리카스’로 마약·전세사기 잡는다”, 조선비즈, 2022. 10. 18. https://biz.chosun.com/topics/topics_social/2022/10/18/FJJD7DMJWZGJXPCVQTJ74VCGJM/ (검색 2022. 11. 10.).
- 임기창, “‘빅데이터 수사’ 시대 열리나…경찰, 2019년 ‘클루’ 시범운영”, 연합뉴스, 2017. 12. 9. <https://www.yna.co.kr/view/AKR20171209026800004> (검색 2022. 11. 9.).
- 임운식, 스마트 치안과 관련된 해외 사례, 수사연구 8월호, 치안정책연구소, 2018.
- 장수인, “[경찰청 AI 프리카스 들여다보니…] (하) 프리카스 시스템 정확도 등 보완 이뤄져야”, 전북도민일보, 2021. 11. 2. https://www.domin.co.kr/news/articleView.html?idxno=1359313&sc_section_code=S1N8 (검색 2022. 11. 10.).
- 조민제, “자율주행 ‘치안 안전’의 시작”, 치안과학기술리뷰, 1권, 2021. 7.
- 한민경, “범죄자를 추적하는 데이터: 사회연결망분석을 활용한 범죄수사”, 통계의 창 2020 겨울호, 통계교육원, 2020, http://sti.kostat.go.kr/window/2020b/main/2020_win_07.html (검색 2022. 11. 10.).

2. 국외 언론 보도 및 기타

- European Commission, Digital Europe Programme – Security (Law enforcement): AI-based pilot, 2022.
- Charles, J., NYPD's Big artificial-intelligence reveal, Governing, 2019. 3. 19. <https://www.governing.com/archive/gov-new-york-police-nypd-data-artificial-intelligence-patternizr.html> (검색 2022. 11. 24.).

A Study on Utilization of Criminal Justice Information for Smart Policing with a Focus on Personal Information Protection

Park, Sungmi* · Ahn, Jungmihn Jamie**

* Korea University, School of Cybersecurity, Ph.D. Candidate

** Hallym University, Legal Informatics & Forensic Science, Professor

ABSTRACT

As the interest in big data and artificial intelligence(AI) grows, the field of "Smart Policing" has received increased attention. However, the information collected during the criminal justice process inevitably contains a lot of personal information, and thus its usage for machine learning datasets is limited. The Korean Personal Data Protection Laws have recently adopted major amendments to free up the use of pseudonymized data and ease the way for the expansion of Big Data-driven services. Contrary to this, the strict privacy protection compliance by the police remains the same - causing passive usage of the Criminal Justice Information.

This study seeks ways to utilize data on the Korea Information System of Criminal Justice Services(KICS), an electronic work system that the police and three other criminal agencies jointly use to perform investigations. The National Police Agency is implementing various R&D tasks to converge high-tech and public security sites. However, it is having difficulties providing this usable Criminal Justice Information due to the strict interpretation and application of vague legal provisions. By examining the Criminal Justice Information produced during the criminal justice process and the current status of domestic and overseas use of criminal justice information, this study finds and suggests alternative solutions such as data classification, data-oriented personal information policy, and ways to establish a privacy-compliant environment as countermeasures to pursue both privacy protection and innovative policing development.

KEYWORD

smart policing, police data, criminal justice information, personal information, de-identification