빅데이터 활용을 위한

개인정보 비식별화 기술 활용 안 내 서

Ver 1.0

빅데이터 담당자들이 실무에 활용 할 수 있도록
비식별화 기술과 활용방법, 실무 사례 및 예제,
분야별 참고 법령 및 활용 Q&A 등 안내

미래창조과학부 NIA 한국정보화진흥원 KBIG 빅데이터 전략센터

| 작성 및 문의 |

미래창조과학부 : 양현철 사무관 / 김자영 주무관

한국정보화진흥원 : 김진철 수석 / 김배현 수석 / 신신애 부장

문의 : cckim@nia.or.kr / 02-2131-0216

| 자 문 |

고환경 변호사(광장) / 최재영 교수(성균관대) / 강장묵 교수(고려대) / 홍승필 교수(성신여대) / 박영우 팀장(KISA) / 최광선 본부장(솔트룩스) / 최현길 대표(메인라인) / 장홍성 팀장(SKT) / 최재원 이사(다음소프트) / 조기행 부장(코리아크레딧뷰로) / 김종민 이사(티크래이프) / 이강신 박사(김&장)

Contents | 빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서(Ver 1.0)

01
개 요

01 개 요

Ⅰ. 배경 및 필요성

● 빅데이터는 미래 신성장 동력이자 신산업으로, 데이터 개방 · 공유 패러다임의 변화와 함께 산업 · 경제 · 사회 전반에서 활용 기대

– 국내는 개인정보보호 관련 법제도상의 제약과 사생활 침해에 대한 막연한 우려로 빅데이터의 적극적 활용이 미흡한 실정

– 이에 따라 개인정보는 보호하면서 빅데이터 활용은 높일 수 있는 대안으로 비식별화 기술에 대한 관심이 높아지고 있음

● 미국, 영국 등 선도국에서도 개인정보를 보호하면서 빅데이터 활용을 위하여 '개인정보 비식별화 활용'을 권고

• **미국(FTC) :** '개인정보의 비식별화 가이드라인('12.03)'에서 특정한 이용자, 컴퓨터 및 기타 개인을 식별할 수 있는 장치들과 연관될 수 있는 것(reasonable link-ability)을 비식별처리 할 것을 권고

• **영국(IOC) :** '개인정보 비식별화 규약('12.12.)'에서 식별 항목을 업계에서 자율 판단하도록 규정

● 우리나라는 빅데이터 활용 노하우가 부족한 상황으로 비식별화를 기반으로 한 빅데이터 활용과 도입 확산이 필요한 시점

– 스마트폰, IoT, 의료, 제조 등 데이터 생산량이 많은 산업이 발달하였음에도 빅데이터 활용은 다소 저조한 상황

빅데이터 담당자가 실무에 활용 할 수 있도록 '빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서' 제공

Ⅱ. 용어 정리

- **공개된 개인정보** 이용자(정보주체) 및 정당한 권한이 있는 자에 의해 일반 공중에게 공개된 부호 · 문자 · 음성 · 음향 · 영상 등의 정보로서 생존하는 개인을 식별할 수 있거나 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보

- **이용내역정보** 정보통신서비스와 관련하여 이용자가 해당 서비스를 이용하는 과정에서 자동으로 발생되는 인터넷 접속정보파일, 거래기록 등의 정보로서 생존하는 개인을 식별할 수 있거나 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보

- **민감정보** 특정한 개인의 사상 · 신념, 노동조합 · 정당의 가입 · 탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 정보

- **정보 조합 · 분석 · 처리시스템** 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합, 분석, 처리하여 새로운 정보를 생성하는 시스템

- **생성된 개인정보** 정보 조합 · 분석 · 처리시스템 운용을 통해 생성된 정보로 개인을 식별할 수 있는 정보 또는 다른 정보와 결합하여 개인을 식별할 수 있는 정보

- **비식별화** 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 결합하여도 특정 개인을 식별할 수 없도록 하는 조치

- **재식별화** 비식별화된 정보를 조합, 분석 또는 처리하는 과정에서 개인정보가 재 생성되는 것

Ⅲ. 안내서 활용 방법

빅데이터 비식별화 기술과 활용방법, 실무 사례 및 예제, 분야별 참고 법령 및 활용 Q&A 등 안내

1. 빅데이터 수집 · 활용 시 개인정보 관련 법조항을 사전검토하고 안전 조치

- 27개 분야의 개인정보 관련 법령과 현행 법률상에서 제시하고 있는 개인 식별 정보 등 참조
  - '수집 → 저장 → 분석 → 이용 · 제공 → 파기' 단계별 안내하는 조치사항에 따라서 개인정보의 안전한 활용

2. 제시 된 다양한 비식별화 기술 활용 방법에 따라 개인정보 비식별 처리

- 18종의 비식별화 세부 방법과 분야별 적용사례 및 실무 예제* 등을 참고하여 개인정보 비식별 처리

  * 안전하고 효율적인 비식별화 처리를 위하여 안내서에서 제시 된 오픈 소프트웨어 등 활용 권고

[비식별화 세부 방법론]

처리 기법	주요 내용 및 처리 예	세부 기술
가명처리 (Pseudonymisation)	주요 식별요소를 다른 값으로 대체	① 휴리스틱 익명화, ② K-익명화, ③ 암호화, ④ 교환 방법
총계처리 (Aggregation)	데이터 총합 또는 부분 집계	⑤ 총계처리, ⑥ 부분집계, ⑦ 라운딩, ⑧ 데이터 재배열
데이터 값 삭제 (Data Reduction)	부분 또는 전체 삭제	⑨ 속성값 삭제, ⑩ 속성값 부분 삭제, ⑪ 데이터 행 삭제, ⑫ 식별자 제거를 통한 단순 익명화
범주화 (Data Suppression)	범주의 값으로 변환	⑬ 범주화, ⑭ 랜덤 올림, ⑮ 범위방법, ⑯ 제어 올림
데이터 마스킹 (Data Masking)	식별자가 보이지 않도록 부분 또는 전체 처리	⑰ 임의 잡음 추가, ⑱ 공백과 대체

3. 빅데이터 활용 Q&A를 참조하여 발생 가능한 문제들을 사전에 해소

- ‘수집 · 이용, 저장 · 관리, 제공 · 위탁, 파기, 사후관리’ 등 빅데이터 활용과정에서 발생 가능한 시나리오를 Q&A로 제공(10개 시나리오 26개 Q&A)

n2

분야별 개인정보 참고 법령 및 조치사항

빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서(Ver 1.0)

분야별 개인정보 참고 법령 및 조치사항

Ⅰ. 분야별 개인정보 참고 법령

● 각 단계별로「개인정보보호법」,「정보통신망 이용촉진 및 정보보호 등에 관한 법률」등 현행 개인정보 관계 법규*에 유의하여 데이터 활용

* 공공기관, 정보통신서비스제공자 등 정보처리 주체에 따라 다른 법규 적용

[분야별 개인정보보호 관련 법률]

구분		주요 내용	규제기관
일반	개인정보보호법	• 개인정보의 수집, 처리 및 보호에 관한 사항 – 개인정보보호일반법	행정자치부
공공부문	전자정부법	• 행정업무의 전자적 처리를 위한 기본원칙, 절차 및 추진방법 등을 규정 • 국가 등이 직접 국민에게 정보를 제공하는 서비스에 대한 근거 규정 – 민간에서 활용할 수 있는 표준화된 정보자원을 공유서비스로 지정	행정자치부
	주민등록법	• 주민을 등록하게 함으로써 주민생활의 편익 증진 및 행정 사무 처리 – 성명, 성별, 생년월일 등 개인정보 등록	
	공공기관의 정보공개에 관한 법률	• 공공정보에 대한 접근권을 부여 • 주체요건, 절차, 비용 등 구체적인 사항 규정 ※ 단, 공공정보에 접근하여 열람할 수 있지만, 정보 이용 · 활용 근거는 아님	
	공공기록물 관리에 관한 법률	• 공공기록물의 안전한 보존 및 효율적 활용 및 관리 – 기록물 원본에 비밀 보호기간 및 보존 기간 지정	
	민원사무처리에 관한 법률	• 민원사무 처리에 관한 기본적인 사항 규정 – 민원인의 신상정보 및 내용 누설 금지	

구분			주요 내용	규제기관
공공부문		국가정보화 기본법	• 국가정보화의 기본 방향과 관련 정책의 수립 · 추진에 필요한 사항을 규정 – 각종 지식과 정보가 공유 · 유통될 수 있도록 기반 마련 – 정보화 역기능을 방지하기 위한 정보보호, 개인정보 보호 등의 대책 마련	미래창조과학부
민간	정보통신	정보통신망 이용촉진 및 정보보호 등에 관한 법률	• 정보통신망의 이용 촉진 및 정보통신 서비스를 이용하는 자의 개인정보 보호 규정 – 빅데이터 처리 등 이용자에게 공개 – 민감정보 생성, 통신 내용 등 이용 금지 – 처리시스템에 기술적 · 관리적 보호조치	방송통신위원회
		위치정보 보호법	• 위치정보의 유노출로부터 사생활 비밀 등을 보호하고 위치정보의 안전한 이용환경 조성 – 개인위치정보 수집, 이용 · 제공, 파기 및 정보 주체의 권리 등 규정	
		전기통신사업법	• 전기통신사업의 운영 및 전기통신의 효율적 관리 규정 – 통신 비밀, 또는 타인 비밀의 침해 및 누설 금지	
		통신비밀 보호법	• 통신비밀을 보호하고 통신의 자유를 신장하기 위한 규정 – 동의 없는 감청, 녹음, 녹취 및 공개 금지	미래창조과학부 / 법무부
		정보통신기반 보호법	• 주요정보통신기반시설의 지정, 금융ISAC의 운영 – 주요정보통신기반시설 준수 법률	미래창조과학부
	상거래	전자문서 및 전자거래기본법	• 전자문서 및 전자거래의 안전성과 신뢰성 확보 – 전자거래이용자의 개인정보 수집/이용/제공 및 관리에 관한 사항	미래창조과학부 / 법무부

구분			주요 내용	규제기관
민간	상거래	전자상거래 등에서의 소비자보호에 관한 법률	• 전자거래시 소비자의 의사표시 확인 – 소비자에 관한 정보이용, 신원 및 거래조건에 대한 정보 제공 등	공정거래위원회
		전자서명법	• 전자서명에 관한 기본 사항 – 공인인증서, 인증업무의 안전성 및 신뢰성 확보	미래창조과학부
		산업기술의 유출방지 및 보호에 관한 법률	• 산업기술의 부정한 유출 방지 및 보호	산업통상자원부
	금융·신용	신용정보의 이용 및 보호에 관한 법률	• 신용정보 전산시스템의 기술적, 물리적 보안대책 – 금융분야 개인(신용)정보보호 법률	금융위원회
		금융실명거래 및 비밀보장에 관한 법률	• 실지명의에 의한 금융거래 및 비밀 보장	
		전자금융거래법, 전자금융감독규정	• 전자금융거래의 안전성과 신뢰성 확보 (안전성 확보 의무, CISO 지정 등) – 금융분야 IT 및 정보보호 법률	
		특정 금융거래 정보의 보고 및 이용 등에 관한 법률	• 자금세탁방지를 위한 금융거래 모니터링	금융위원회 / 법무부
	보건·의료	의료법	• 의료 개인정보의 보호 – 의료기관 특별법	보건복지부
		국민건강보험법	• 국민건강증진을 위한 보험급여를 정한 법률 – 요양급여비용의 청구와 지급 등	
		산업안전보건법	• 근로자의 건강진단 결과(질병 정보 및 건강 정보 등)에 대한 보호	
		후천성면역결핍증 예방법	• 후천성면역결핍증 환자에 대한 개인정보관리 및 정보 보호	
		감염법의 예방 및 관리에 관한 법률	• 법에서 정하는 감염병 환자에 대한 개인정보 관리 및 정보보호	
		응급의료에 관한 법률	• 응급환자 이송 시 개인정보 및 비용 청구 등에 관한 내용	
		장애인 차별금지 및 권리규제 등에 관한 법률	• 장애를 이유로 차별받은 사람의 권익을 효과적으로 규제하기 위한 법률 – 본인의 동의를 얻기 어려운 장애인의 개인정보 보호에 관한 내용	

Ⅱ. 현행 법률기반 개인 식별 정보

● 개인정보 침해 문제가 발생하지 않도록 현행 법규에서 제시하는 식별정보는 비식별화 활용

[법률 기반 개인 식별 정보]

구분		근거	개인 식별 정보 항목
일반		• 개인정보보호법 **제18조, 제23조, 제24조 제1항, 제24조 제2항, 제24조 제3항**	• 주체자의 사생활을 침해할 수 있는 식별정보 (ex. 의료정보, 정신적 성향 등) • 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 주소 등)
공공부문		• 전자정부법 **제42조**	• 정당한 사용자임을 인증하는 식별정보 (ex. 인증서 일련 번호, 유효기간 등)
		• 주민등록법 **제10조**	• 신분 확인정보와 가족구성원 정보를 통해 확인될 수 있는 식별정보(ex. 성명, 성별, 세대주와의 관계 등)
		• 공공기관의 정보공개에 관한 법률 **제18조** • 공공기록물 관리에 관한 법률 **제37조**	• 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 연락처 등)
		• 민원사무처리에 관한 법률 **제26조** • 국가정보화 기본법 **제39조**	• 본인 · 대리인 확인을 위한 식별정보 (ex. 주민등록번호, 대리인 신분증 등)
민간	정보통신	• 정보통신망 이용촉진 및 정보보호 등에 관한 법률 • 전자서명법 **제24조**	• 회원 관리를 위한 사용자 식별 정보 (ex. 이름, ID, PW 등) • 정당한 사용자임을 인증하는 식별정보 (ex. I-PIN인증, 단말정보, 휴대폰정보 등)
		• 전자금융거래법 **제25조**	• 휴대폰 결제 서비스 수행을 위한 식별정보 (ex. 결제수단별 개인정보, 카드번호, 비밀번호 등)
		• 전기통신사업법 **제83조**	• 주체자의 신분 정보 및 통신상의 사용자 정보에 대한 식별정보(ex. 이름, ID, 주민등록번호 등)
		• 위치정보보호법 • 통신비밀보호법	• 업무 수행 및 처리를 위한 통신상의 식별정보 (ex. 접속 IP정보, GPS 정보 등)
		• 청소년보호법 **제29조, 제16조**	• 제한된 연령 확인에 대한 식별정보 (ex. 법정 생년월일, 법정 대리인 정보 등)

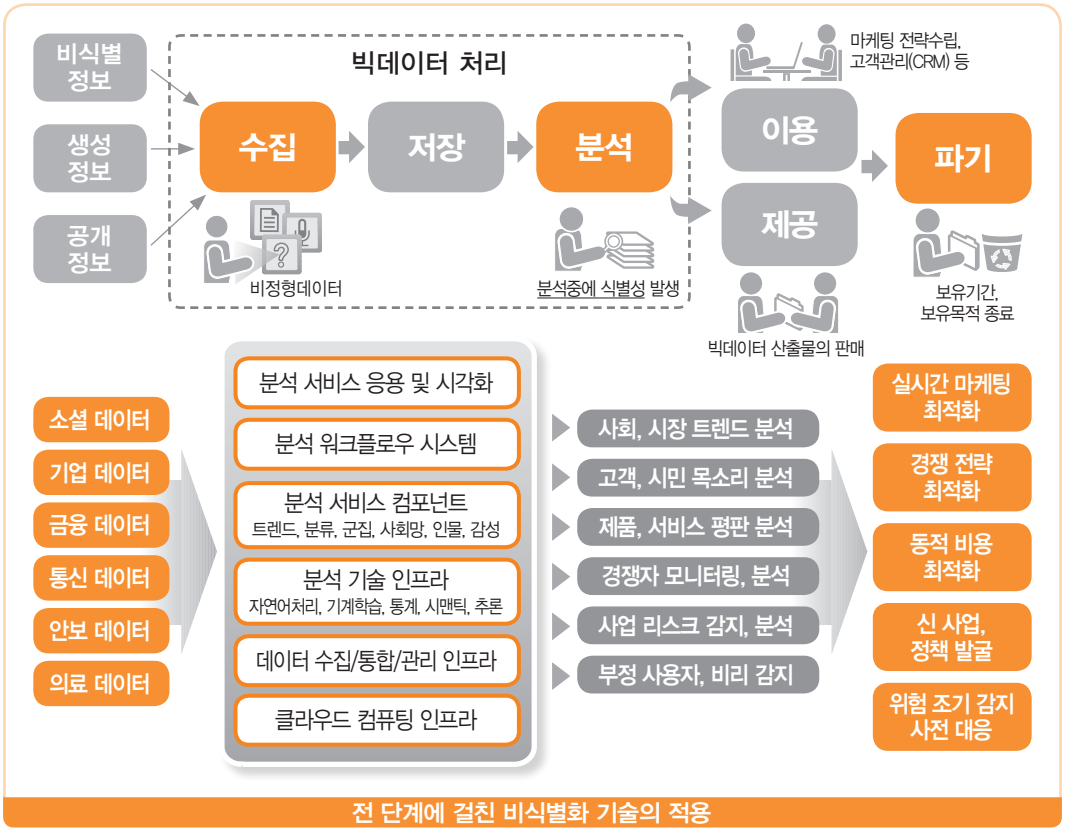
구분		근거	개인 식별 정보 항목
민간	상거래	• 전자문서 및 전자거래기본법 **제12조** • 정보통신망 이용촉진 및 정보보호 등에 관한 법률 **제23조, 제24조** • 전자상거래 등에서의 소비자보호에 관한 법률 **제12조**	• 전자문서 서비스를 위한 식별정보 (ex. 공인전자주소, 송신자, 수신자 등) • 통신의 안전한 조치를 위해 확인할 수 있는 식별정보 (ex. 비밀번호, 계좌번호, 주민등록번호 등) • 거래 기록 및 배송을 확인하기 위한 식별정보 (ex. 배송 주소지, 수령인 연락처 등)
		• 전자서명법 **제24조**	• 정당한 사용자임을 인증하는 식별정보 (ex. 가입자 이름, 전자서명검증정보, 인증서 일련번호)
	금융·신용	• 신용정보의 이용 및 보호에 관한 법률 **제33조** • 금융실명거래 및 비밀 보장에 관한 법률 **제4조**	• 신용정보 및 거래능력을 판단할 수 있는 식별정보 (ex. 재산, 소득, 대출 보증 등) • 금융기관의 거래내역을 판단할 수 있는 정보 (ex. 주민등록번호, 계좌번호, 거래실적 자료 등) • 이용자 및 거래내용의 정확성을 확인하기 위한 식별정보 (ex. 전자금융업자에 등록된 이용자번호, 이용자의 생체정보, 등)
		• 전자금융거래법 **제26조** • 전자금융 감독규정 **제5조의 3**	• 신용정보 및 거래능력을 판단할 수 있는 식별정보 (ex. 재산, 소득, 대출 보증 등) • 금융기관의 거래내역을 판단할 수 있는 정보 (ex. 주민등록번호, 계좌번호, 거래실적 자료 등) • 이용자 및 거래내용의 정확성을 확인하기 위한 식별정보 (ex.전자금융업자에 등록된 이용자번호, 이용자의 생체정보, 등)
		• 특정 금융거래 정보의 보고 및 이용 등에 관한 법률 **제5조의 3**	• 자금이체 수행을 위한 식별정보 (ex. 송금인 성명, 계좌번호, 수취인의 정보)
	보건·의료	• 의료법 **제21조** • 응급의료에 관한 법률 **제22조의 2항** • 산업안전보건법	• 정확한 환자의 진료를 위해 확인가능 한 식별정보 (ex. 주민등록번호, 의료기록, 가족력 등) • 신체의 질병정보를 통해 인지될 수 있는 식별정보 (ex. 감염병명, 혈액정보, 조직정보 등)
		• 후천성면역결핍증예방법 • 감염법의 예방 및 관리에 관한 법률 **제74조**	• 정확한 환자의 진료를 위해 확인가능 한 식별정보 (ex. 주민등록번호, 의료기록, 가족력 등) • 신체의 질병정보를 통해 인지될 수 있는 식별정보 (ex. 감염병명, 혈액정보, 조직정보 등)
		• 장애인 차별금지 및 권리규제 등에 관한 법률 **제22조**	• 신체 장애정보를 통해 확인 가능한 식별정보 (ex. 주민등록번호, 신체장애, 장애등급 등)
		• 국민건강보험법 **제5조**	• 가족구성원의 정보를 통해 확인할 수 있는 식별정보 (ex. 가족구성원의 이름, 출생지, 소득 등)

Ⅲ. 각 단계별 조치사항 및 관련 법조항

1. 빅데이터 활용 단계 정의

● 빅데이터는 '수집 → 저장 → 분석 → 이용 · 제공 → 파기' 단계를 거쳐 활용

[빅데이터 활용 단계 정의]



● **수집** 빅데이터 처리를 목적으로 다양한 경로를 통해 필요한 데이터를 모으는 과정

● **저장** 수집한 데이터를 장치에 저장하고 분석에 사용할 수 있도록 관리하는 과정

● **분석** 수집한 데이터를 다양한 방법을 통해 가공하여 새로운 정보를 생성하는 과정

● **이용·제공** 수집한 데이터와 이를 분석한 정보를 빅데이터 처리자가 직접 사용하거나 제3자가 사용하게 하는 과정

● **파기** 수집한 데이터 또는 이를 분석한 정보를 삭제하는 과정

2. 수집 단계

| 조치 사항 |

① 데이터 수집 대상이 특정 개인인 경우 법률의 허용 규정이 있거나 사전에 정보주체의 동의를 얻어야 함

② 개인정보가 포함된 공개된 데이터를 수집하는 경우 법률의 허용 규정이나 정보주체의 사전 동의가 없으면 개인정보를 비식별화 하여야 함

③ 서비스 제공에 필수적인 이용내역 데이터는 정보주체의 동의 없이 수집 · 이용할 수 있으나 수집 사실 등을 공개, 통지하는 것이 바람직함

④ 주민등록번호, 민감정보 등 수집 제한 데이터를 수집하지 않도록 유의하여야 함

● 데이터 수집 대상이 특정한 식별할 수 있는 개인인지 확인

– 데이터 수집 대상이 특정 개인 또는 특정할 수 있는 다수의 개인인 경우 법률의 허용규정이 있거나 사전에 정보주체의 동의를 얻어야 함

● (불특정 다수의 개인을 대상으로) 수집하는 데이터 중에 개인정보가 포함되어 있는지 확인

– 개인정보는 그 자체로 개인을 식별할 수 있거나 다른 정보(데이터)와 쉽게 결합하여 개인을 식별할 수 있는 정보(데이터)

– 개인정보가 포함되어 있는 경우 해당 개인정보 수집을 위한 법률의 허용규정, 정보주체의 동의가 있는 경우를 제외하고 비식별화 조치

※ 우리법상 개인정보는 법률의 허용 규정 또는 정보주체의 동의가 있어야만 수집 · 이용 가능(「개인정보보호법」 제15조, 「정보통신망법」 제22조)

– 비식별화 조치를 한 정보는 개인을 식별할 수 없기 때문에 개인정보가 아니며, 그 이용 등에 대해 정보주체의 동의를 받을 필요가 없음

● 개인정보가 포함된 공개된 데이터를 수집하는 경우

– 개인정보를 비식별화 한 때에는 정보주체의 동의가 필요하지 않으나, 비식별화 하지 않고 이용하려면 법률의 허용규정이 있거나 정보주체의 동의를 얻어야 함(「빅데이터 개인정보보호 가이드라인」 제4조 제1항)

– 개인정보를 비식별화 하여 정보주체의 동의 없이 수집하는 경우에도 수집 출처, 조합 · 분석 · 처리 사실과 목적을 공개하고 이메일 · 쪽지 등 방법으로도 고지(「개인정보보호법」 제20조 제1항 · 제30조 제2항, 「정보통신망법」 제27조의 2, 「빅데이터 개인정보보호 가이드라인」 제4조 제2항)

● 수집 제한 데이터를 수집하지 않도록 유의

– 주민등록번호는 법령에서 수집 · 이용을 허용하는 경우에만 수집 가능(「개인정보보호법」 제24조 제1항, 「정보통신망법」 제23조 제1항)

– 민감정보는 법률에서 허용하거나 이용자의 별도 동의를 얻어야 수집 · 이용 가능(「개인정보보호법」 제23조, 「정보통신망법」 제23조 제1항)

※ 민감정보라도 개인을 식별할 수 없는 경우에는 개인정보로서 보호되지 않음

| 개인정보의 유형과 종류 |

① **고유식별정보** 주민등록번호, 여권번호, 운전면허번호, 외국인 등록번호(「개인정보보호법」제24조 제1항, 동법 시행령 제19조)

② **민감정보** 사상 · 신념, 노동조합 · 정당의 가입 · 탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보(「개인정보보호법」 제23조, 「정보통신망법」 제23조 제1항)

- 「개인정보보호법」은 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보를 유전자 검사 등의 결과로 얻어진 유전정보, 범죄경력자료에 해당하는 정보로 한정
- 「정보통신망법」은 가족 · 친인척관계, 학력(學歷) · 병력(病歷), 사회활동 경력 등도 포함하고, 개인의 권리 · 이익 및 사생활을 뚜렷하게 침해할 우려가 있는 개인정보로 폭넓게 규정

– 출처 : KISA 보호나라

[법률상 개인정보 수집 시 이용자 동의를 받지 않아도 되는 경우]

법률	주요내용
「정보통신망법」 (제22조 제2항)	• 정보통신서비스의 제공에 관한 계약을 이행하기 위하여 필요한 개인정보로서 경제적 · 기술적인 사유로 통상적인 동의를 받는 것이 뚜렷하게 곤란한 경우 ※ 과금정보, 통화내역, 접속로그 등의 정보가 생성된 경우 • 정보통신서비스의 제공에 따른 요금정산을 위하여 필요한 경우 ※ 서비스별 요금액, 납부, 미납사실 등의 정보를 수집하는 경우 • 「정보통신망법」 또는 다른 법률에 특별한 규정이 있는 경우
「개인정보보호법」 (제15조 제1항 제3호~제6호)	• 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우 ※ 국민건강보험공단이 보험급여관리 등을 위하여 진료내역 등을 수집 · 이용하는 경우 • 정보주체와의 계약의 체결 및 이행을 위하여 불가피하게 필요한 경우 ※ 회사가 취업지원자와의 채용 및 근로계약 체결 전에 지원자의 이력서, 졸업증명서 등 정보를 수집 · 이용하는 경우 • 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우 ※ 조난 · 홍수 등으로 실종되거나 고립된 사람을 구조하기 위하여 연락처, 주소, 위치정보 등 개인정보를 수집하는 경우 • 개인정보처리자의 정당한 이익을 달성하기 위하여 필요한 경우로서 명백하게 정보주체의 권리보다 우선하는 경우. 이 경우 개인정보처리자의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니하는 경우 ※ 고객의 물품 주문내역, 서비스 이용내역, 통신 사실 확인 자료 등과 같이 요금을 산출하고 과금하기 위한 자료를 수집하는 경우
「의료법」 (제21조 제2항 제4호 · 제5호)	• 급여비용 심사 · 지급 · 대상여부 확인 · 사후관리 및 요양급여의 적정성 평가 · 가감지급 등을 위하여 국민건강보험공단 또는 건강보험심사평가원에 제공하는 경우 • 의료급여 수급권자 확인, 급여비용의 심사 · 지급, 사후관리 등 의료급여 업무를 위하여 보장기관(시 · 군 · 구), 국민건강보험공단, 건강보험심사평가원에 제공하는 경우

3. 저장(관리) 단계

| 조치 사항 |

① 개인정보가 저장, 처리되는 정보 조합 · 분석 · 처리 시스템 (빅데이터 처리 시스템)에 대하여는 관계 법령에 따른 안전조치 또는 보호조치를 하여야 함

② 고유식별정보에 대하여는 관계 법령에 따라 송 · 수신, 전달 또는 저장 시 암호화를 하여야 함

③ 비식별 정보의 저장에 대하여는 자체적으로 판단하여 알맞은 안전조치 또는 보호조치를 취하면 됨

④ 수집단계에서 필터링 되지 않거나 저장된 데이터를 유형화하는 과정에서 추출된 고유식별정보는 암호화 또는 다시 비식별화 하는 등 보호조치를 하여야 함

● 빅데이터 처리 시스템에서 저장, 처리되는 고유식별정보를 비롯한 개인정보의 안전성 확보

– 빅데이터 처리 시스템에서 처리되는 개인정보가 분실 · 도난 · 유출 · 변조 또는 훼손되지 아니하도록 안전성 확보에 필요한 기술적 · 관리적 및 물리적 조치를 하여야 함(「개인정보보호법」 제29조, 「정보통신망법」 제28조)

[고유식별정보 및 개인정보에 대한 안전조치 법규]

법률	내용
「개인정보보호법」 제24조 제3항 · 제29조 / 시행령 제21조 · 제30조	• 내부 관리계획 수립 · 시행, 접근통제 및 접근권한 제한조치, 암호화기술 적용 등, 접속기록 보관 등, 보안프로그램 설치 · 갱신, 보관시설 마련 등 ※ 「개인정보의 안전성 확보조치 기준」(안행부 고시 제2011–43호)
「정보통신망법」 제28조 / 시행령 제15조	• 내부 관리계획 수립 · 시행, 접근통제장치 설치 · 운영, 접속기록 위 · 변조 방지, 암호화기술 이용 등, 백신 소프트웨어 설치 · 운영 등 ※ 「개인정보의 기술적 · 관리적 보호조치 기준」(방통위 고시 제2012–50호)

– 특히 고유식별정보를 정보통신망을 통하여 송 · 수신하거나 보조저장매체 등을 통하여 전달하는 경우, 내부망 등에 저장하는 경우 암호화 하여야 함(「개인정보보호법」 제24조 제3항, 동법 시행령 제21조 및 제30조 제1항 제3호, 「개인정보의 안전성 확보조치 기준」 제6조; 「정보통신망법」 제28조 제1항 제4호, 동법 시행령 제15조 제4항 제2호~제4호, 「개인정보의 관리적 · 기술적 보호조치 기준」 제6조 제2항~제4항)

- 고유식별정보 또는 개인을 식별할 수 있는 정보를 저장, 처리하지 않는 빅데이터 처리 시스템의 경우

  – 개인을 식별할 수 없는 정보 또는 개인정보를 비식별화한 정보는 개인정보가 아니므로 자체적으로 판단하여 알맞은 안전조치 또는 보호조치를 취하면 됨

  ※ (예) 사물인터넷(IoT) 등에 의해 자동 수집되는 정보, 로그 정보, 인터넷 게시판이나 소셜 네트워크 서비스 등에서 수집되는 행태 정보 등

  – 그러나 공개정보 및 이용내역정보를 비식별화 하여 빅데이터 처리 시스템에 저장 · 관리하는 경우에는, 접근 통제장치 설치 · 운영, 접속기록의 위조 · 변조 방지, 악성 프로그램에 의한 침해 방지 및 기타 안전성 확보를 위해 필요한 보호조치를 취해야 함(「빅데이터 개인정보보호 가이드라인」 제3조 제2항)

  ※ 분석에 의하여 재식별화될 수 있으므로 최소한의 보호조치를 취해야 함

- 수집 단계에서 필터링 되지 않았거나 저장된 데이터를 유형화하는 과정에서 추출된 고유 식별 정보

  – 암호화 또는 다시 익명화(비식별화) 조치를 하여 안전하게 관리(「개인정보보호법」 제29조 및 동법 시행령 제30조, 「정보통신망법」 제28조 및 동법 시행령 제15조)

  ※ 사업자의 업무환경에 따라 가능한 익명화 조치 방법을 선택할 수 있으며 익명화 조치에 대한 책임은 해당 사업자가 부담

4. 분석 단계

| 조치 사항 |

① 개인정보가 포함된 공개 정보, 이용내역 정보(이하 "공개된 개인정보 등"으로 약칭)는 비식별화 조치를 한 다음 조합, 분석 또는 처리하여야 함

② 비식별화 조치를 한 정보는 조합, 분석 또는 처리 과정에서 재식별화 되지 않도록 하여야 함

③ 공개된 개인정보 등의 조합, 분석 또는 처리 과정에서 생성된 개인정보는 조합, 분석 또는 처리 목적을 달성한 후 지체없이 파기 또는 비식별화 조치를 하여야 함

④ 정보주체의 동의 없이 또는 법률에 의하지 않고 공개된 개인정보 등을 조합, 분석 또는 처리하여 민감정보를 생성하지 않아야 함

⑤ 전송중인 이메일, 문자 메시지 등 통신내용을 조합, 분석 또는 처리하지 않아야 함

- 빅데이터 분석 및 활용은 법률상 허용되는 목적 및 범위 내에서만 가능하다는 점에 유의하여야 함
  - 개인의 사상 · 신념이나 성적 취향 등을 파악하기 위한 빅데이터 분석 및 활용은 허용되지 않음(「헌법」 제17조 · 제19조)
  - 법률 규정에 의하거나 정보주체의 동의를 얻고 개인정보를 수집한 경우 그 목적 및 범위를 벗어나지 않아야 함(「개인정보보호법」 제15조 제1항, 「정보통신망법」 제24조)

- 개인정보가 포함된 공개 정보, 이용내역 정보는 비식별화 조치를 한 다음 조합, 분석 또는 처리하여야 함(「빅데이터 개인정보보호 가이드라인」 제4조 · 제5조)
  - 비식별화 조치를 한 정보는 조합, 분석 또는 처리 과정에서 재식별화 되지 않도록 유의하여야 함
  - 다만, 이용자의 동의를 받거나 법령상 허용하는 경우 비식별화 조치를 하지 않을 수 있음(「빅데이터 개인정보보호 가이드라인」 제4조 제1항 단서 · 제5조 제1항 각호)
  - 비식별화 작업 단계 자체를 개인정보의 처리로 보고 정보주체의 동의를 요한다는 일부 시각도 있으나, 비식별화 처리에 대한 정보 주체의 동의가 필요하다고 볼 경우 빅데이터 사업의 육성 및 효과적인 데이터 분석을 저해하는 면이 있을 뿐만 아니라, 법적으로도 비식별화 자체에 대해 정보주체의 동의를 요한다고 볼 수 없음

● 분석 등 과정에서 생성된 개인정보는 목적을 달성한 후 파기 또는 비식별화 하여야 함

– 비식별화 조치를 하여 수집한 공개된 개인정보 등의 조합, 분석 또는 처리 과정에서 생성된 개인정보는 조합, 분석 또는 처리 목적을 달성한 후 지체없이 파기 또는 비식별화 조치를 하여야 함(「빅데이터 개인정보보호 가이드라인」 제6조 제1항 단서)

– 개인정보가 포함된 정보가 생성될 수 있다는 사실 및 그 처리 방법을 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개(「빅데이터 개인정보보호 가이드라인」 제6조 제2항)

● 민감정보의 생성을 목적으로 공개된 개인정보 등을 조합, 분석 또는 처리하지 않아야 함(「개인정보보호법」 제3조 제6항 · 제23조, 「정보통신망법」 제23조 제1항 본문, 「빅데이터 개인정보보호 가이드라인」 제7조 본문)

– 다만, 사전에 정보주체의 별도 동의를 받거나 또는 법령에서 허용하는 경우에는 가능(「개인정보보호법」 제23조 단서 각호, 「정보통신망법」 제23조 제1항 단서, 「빅데이터 개인정보보호 가이드라인」 제7조 단서)

[민감정보의 처리를 법령에서 허용하는 경우]

법률	주요내용
「의료법」 제21조 제2항	환자의 진료상 필요한 경우에 다른 의료기관에서 보유하고 있는 진료 기록을 수집
「보안관찰법시행령」 제6조 제1항	보안관찰처분대상자는 출소하기 전에 종교, 범죄기록 등을 포함한 신고서를 교도소 등의 장에게 제출
「보험업법」 제176조 제10항	보험요율 산출기관은 순보험요율을 산출하기 위해 교통법규 위반에 관한 개인정보를 수집
「병역법」 제11조의 2	지방병무청장은 징병검사와 관련하여 징병검사대상자의 진료기록 · 학교생활기록부 등을 수집

● 전송 중인 이메일, 문자메시지 등 통신 내용은 조합, 분석 또는 처리하지 않아야 함(「통신비밀보호법」 제3조 제1항,「빅데이터 개인정보보호 가이드라인」 제8조)

– 다만, 양 당사자의 동의를 받거나 또는 법률에서 허용하는 경우에는 가능

5. 이용 · 제공 단계

| 조치 사항 |

① 비식별화된 공개된 개인정보 등을 서비스 제공을 위하여 내부에서 이용할 때에는 정보주체가 이를 쉽게 확인할 수 있도록 공개하여야 함

② 공개된 개인정보 등의 분석 결과를 사생활 침해, 사회적 차별 조장 기타 사회질서에 반하는 목적으로 이용하지 않아야 함

③ 개인 식별 정보를 포함하는 분석 결과를 제3자에게 제공하는 경우에는 정보주체의 동의를 받아야 함

④ 통계, 학술 연구 등 공익 목적의 분석 결과를 일반에 공개하여 공유할 수 있음

- 공개된 개인정보도 개인정보에 해당되므로 수집 · 이용 및 제공을 위하여는 원칙적으로 정보주체의 동의를 얻어야 함
  - 다만, 법령의 규정에 의하거나 계약의 이행, 요금정산 등 서비스 제공업무 수행을 위해 비식별화된 공개된 개인정보 등을 내부에서 이용하는 경우에는 정보주체의 별도 동의 없이 이용 가능(「개인정보보호법」 제15조 제1항 제2호부터 제6호, 「정보통신망법」 제22조 제2항, 「빅데이터 개인정보보호 가이드라인」 제9조 제1항 본문)
  - 비식별화된 공개된 개인정보 등을 서비스 제공 목적으로 내부 이용하는 경우 해당 정보가 이용된다는 사실 및 그 목적을 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개(「빅데이터 개인정보보호 가이드라인」 제9조 제2항)

- 빅데이터 분석 결과의 사생활 침해 등 반사회적 목적에의 이용 금지
  - 빅데이터 분석 결과는 개인의 사생활 침해, 사회적 차별 조장, 기타 사회질서에 반하는 목적으로 이용하지 않아야 함

  ※ 개인의 사생활을 추적하는 스토킹, 여성 · 장애인 · 다문화 등 사회적 약자에 대한 사회적 편견 조장, 도박 등 불법행위에 이용 금지

- 개인정보가 포함된 공개 정보, 이용내역 정보, 생성 정보의 분석 결과인 경우, 정보주체의 동의를 얻어 제3자에게 제공 가능(「개인정보보호법」 제17조, 「빅데이터 개인정보보호 가이드라인」 제10조 본문)

– 다만, 비식별화 처리된 공개된 정보, 이용내역 정보, 생성 정보는 이용자 동의 없이 제3자 제공 가능(「빅데이터 개인정보보호 가이드라인」 제10조 단서)

- 공익 목적을 위한 빅데이터 분석 결과의 공개 및 공유
  – 통계, 학술 연구 등 공익 목적의 분석 결과는 특정 개인을 알아볼 수 없는 형태로 일반에 공개하여 공유할 수 있음(「통계법」 제31조, 「개인정보보호법」 제18조 제2항 제4호)

6. 파기 단계

| 조치 사항 |

① 정보주체가 개인정보 수집 · 이용 · 제공 등의 동의를 철회한 경우 분석 결과에 포함된 해당 정보주체의 개인 식별 정보를 즉시 파기하여야 함

② 개인 식별 정보가 포함된 분석 결과의 이용 목적이 달성되거나 보유기간이 경과한 경우 해당 개인 식별 정보를 즉시 파기하거나 익명화 (비식별화) 조치를 취하여야 함

● 정보주체가 빅데이터 분석을 위한 개인정보 수집 · 이용 · 제공 등의 동의를 철회한 경우

– 해당 정보주체의 개인 식별 정보가 빅데이터 분석에 이용되지 않도록 지체없이 삭제(「개인정보보호법」 제37조, 「정보통신망법」 제30조제1항 · 제3항)

▶ "지체없이"의 기간

• 정보주체가 서비스 이용을 거부하거나 자신의 개인정보 수집 · 이용 · 제공 등의 동의를 철회한 경우, 빅데이터 사업자는 정당한 사유가 없는 한 5일 이내에 조치하여야 함. (행정자치부 「표준 개인정보보호 지침」 제11조 제1항, 방송통신위원회 「온라인 개인정보 취급 가이드 라인」 12쪽)

– 동의철회 시에도 다음의 경우 개인 식별 정보를 보존할 수 있음

▶ 동의철회의 경우에도 보존할 수 있는 경우

• 사업자가 보존하여야 할 거래기록 및 그와 관련된 개인정보(「전자상거래 등에서의 소비자보호에 관한 법률」제6조)

• 후불제 서비스에서의 요금정산, 이용자의 요금 이의신청기간 운영, 법령에서 일정기간 보유하도록 규정한 경우 등(방송통신위원회 「온라인 개인정보 취급 가이드라인」 12쪽)

● 개인 식별 정보가 포함된 분석 결과의 이용 목적이 달성되거나, 보유기간이 경과하거나 또는 사업을 폐업하는 경우

– (수집 시 동의를 받은 개인정보) 동의 획득 당시 정한 이용목적이 달성되거나 보유기간이 경과한 경우 지체없이 파기(「개인정보보호법」 제21조, 「정보통신망법」 제29조 제1항)

– (수집 시 동의를 받지 않은 개인정보) 개인정보 취급방침에서 명시한 이용목적 또는 보유기간의 경과시 파기하거나 익명화(비식별화) 조치(「개인정보보호법」 제21조, 「정보통신망법」 제29조 제1항 제3호)

※ 〈붙임 1〉 법률상의 개인정보 보존기간 현황 참고

▶ 개인정보 유효기간제

- 3년* 또는 법령이나 이용자 요청으로 정한 기간 동안 정보통신서비스를 이용하지 않은 이용자의 개인정보는 즉시 파기 또는 다른 이용자 개인정보와 분리 보관하여야 함(「정보통신망법」 제29조 제2항 및 동법 시행령 제16조 제1항 · 제2항)
- 2015년 8월 18일부터는 1년

[붙임 1. 법률상 개인정보 보존기간 현황]

근거법령	개인정보의 종류	보존기간
통신비밀보호법 (통신사실확인자료)	로그기록자료, 접속지의 추적자료	3개월
	전기통신일시, 전기통신개시 · 종료시간, 사용도수, 상대방의 가입자번호, 발신기지국의 위치추적자료	12개월
정보통신망 이용촉진 및 정보보호 등에 관한 법률	본인확인정보	6개월
	통신과금서비스에 관한 기록 (건당 거래금액 1만원 이하)	5년 (1년)
전자상거래등에서의 소비자보호에 관한 법률 (거래기록)	소비자의 불만 또는 분쟁처리에 관한 기록	3년
	계약 또는 청약철회 등에 관한 기록, 대금 결제 및 재화 등의 공급에 관한 기록	5년
전자금융거래법 (전자금융거래기록)	건당 거래금액 1만원 이하 전자금융거래에 관한 기록 전자지급수단 이용과 관련된 거래승인에 관한 기록	1년
	전자금융거래 종류 및 금액, 상대방에 관한 정보 지급인의 출금 동의에 관한 사항 전자금융거래와 관련한 전자적 장치의 접속기록 전자금융거래 신청 및 조건의 변경에 관한 사항 건당 거래금액 1만원 초과 전자금융거래에 관한 기록	5년

근거법령	개인정보의 종류	보존기간
신용정보의 이용 및 보호에 관한 법률	신용정보 업무처리에 관한 기록	3년
의료법 (진료에 관한 기록)	처방전	2년
	진단서 등의 부본	3년
	환자 명부, 검사소견기록, 간호기록부, 방사선사진 및 그 소견서, 조산기록부	5년
	진료기록부, 수술기록	10년
국세기본법	국세 부과 제척기간(조세시효)	10년
	국세징수권 및 국세환급금 소멸시효	5년
상법	보험금액 청구권 소멸시효, 보험료/적립금 반환청구권 소멸시효	2년
	상사채권 소멸시효, 배당금 지급청구권 소멸시효	5년
	사채상환청구권 소멸시효	10년
제조물책임법	손해배상청구권 소멸시효	3년 / 10년

n3

비식별화 기술 실무활용 방법

빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서(Ver 1.0)

비식별화 기술 실무활용 방법

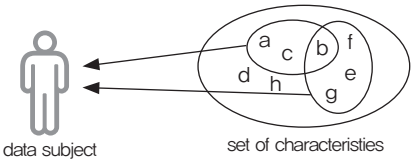
Ⅰ. 비식별화 개념

● (정의) 데이터 내에 개인을 식별할 수 있는 정보가 있는 경우, 이의 일부 또는 전부를 삭제, 또는 일부를 속성 정보로 대체 처리함으로써 다른 정보와 결합하여도 특정 개인을 식별하기 어렵도록 하는 조치

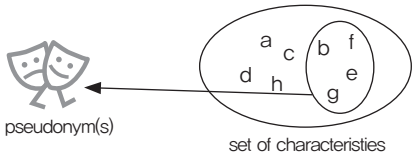
비식별화 처리 예시

▶ 정보 내 식별 가능한 특징을 제거하거나 변형시킴으로써 데이터 집합과 데이터 대상(정보 이용자)과의 유일한 연관관계를 제거

– 개인과 여러 정보를 연결시켜 개인의 정보가 드러나지 않게 하거나 하나의 특징 정보를 여러 개인과 연결시켜 개인 식별 방지



Source : ISO/IEC DTS 25237



Source : ISO/IEC DTS 25237

▶ 이름을 '김수철'(유명 가수 이름 등), 김삿갓(역사적 인물 등)으로 바꾸어 누군지 알 수 없도록 함

▶ 특정인의 몸무게를 20대 서울 거주 여성의 평균 몸무게로 처리하여 누구의 몸무게인지를 구분할 수 없도록 함

▶ 991202–1234567과 같은 주민번호를 99년생 남성으로 변환하여 개인을 식별할 수 없도록 함

▶ 이명박, 73세라는 특정인을 구분할 수 있는 경우에는 이씨 성을 가진 70대로 바꾸어 개인정보를 보호함

▶ 이명박, 안암1동 거주, 고려대학교 재학 이라는 특정인을 구분할 수 있는 경우에는 이○○, ○○대학 재학, ○○ 거주 식으로 처리함

Ⅱ. 비식별화 대상 및 기준

1. 적용대상

● 그 자체로 개인을 식별할 수 있는 정보 및 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보들을 대상으로 함

[비식별화 적용 대상 예시]

구분	주요 내용
① 그 자체로 개인을 식별할 수 있는 정보	• 쉽게 개인을 식별할 수 있는 정보 : 이름, 전화번호, 주소, 생년월일, 사진 등 • 고유식별정보 : 주민등록번호, 운전면허번호, 의료보험번호, 여권번호 등* • 생체정보 : 지문, 홍채, DNA 정보 등 • 기관, 단체 등의 이용자 계정 : 등록번호, 계좌번호, 이메일 주소 등 • 기타 유일 식별번호 : 군번, 사업자등록번호 특성(별명), 식별코드(아이디, 아이핀 값(cn, dn)) 등
② 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보	• 개인특성 : 성별, 생년, 생일, 연령(나이), 국적, 고향, 거주지, 시군구명, 우편번호, 병역여부, 결혼여부, 종교, 취미, 동호회 · 클럽, 흡연여부, 음주여부, 채식여부, 관심사항 등 • 신체 특성 : 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔, 신체검사 결과, 장애유형, 장애등급, 병명, 상병코드, 투약코드, 진료내역 등 • 신용 특성 : 세금 납부액, 신용등급, 기부금, 건강보험료 납부액, 소득분위, 의료급여자 등 • 경력 특징 : 학교명, 학과명, 학년, 성적, 학력, 직업, 직종, (전 · 현)직장명, 부서명, 직급, 자격증명, 경력 등 • 전자적 특성 : PC사양, 비밀번호, 비밀번호 질문/답변, 쿠키정보, 접속일시, 방문일시, 서비스 이용기록, 위치정보,I 접속로그, IP주소, MAC주소, HDD Serial 번호, CPU ID, 원격접속 여부, Proxy 설정여부, VPN 설정 여부, USB Serial 번호, Mainboard serial 번호, UUID, OS 버전, 기기 제조사, 모델명, 단말기 ID, 네트워크 국가 코드, SIM Card 정보 등 • 가족 특성 : 배우자, 자녀, 부모, 형제 여부, 가족정보, 법정대리인 정보 등 • 위치 특성 : GPS 데이터, RFID 리더 접속 기록, 특정 시점 센싱기록, 인터넷 접속, 핸드폰 사용기록, 사진 등

2. 적용시기

● 빅데이터 수집 · 활용의 전(全) 단계에서 개인정보가 식별되는 경우 혹은 이후 정보의 추가 가공 등을 통하여 개인이 식별되는 경우 등

※ 예시) ① 개인 정보의 수집 및 저장 시, ② 개인 정보가 포함되어 있을 수 있는 데이터의 활용 시, ③ 다른 기관(정보)과의 정보 공유 시, ④ 기관내의 서로 다른 부서간의 정보 공유 시

3. 적용 기준 및 준수 사항

① 그 자체로 개인 식별이 가능한 정보는 삭제

– 단, 수집 시에 개인정보에 대한 자체이용, 제3자 제공 등 활용에 대한 이용자 동의를 받았을 경우 비식별화 없이 활용 가능

② 다른 정보와 결합에 따른 재식별 위험 최소화

보유 개인정보의 분석을 위한 동의 등이 곤란한 경우 분석 목적을 달성할 수 있는 한도에서 비식별화 처리

– 적절한 비식별화 기법을 적용하여 개인 정보를 삭제하는 대신 일부 또는 전체를 속성 정보로 대체 처리

– 연결정보가 있는 경우, 기본적으로 URL 등을 삭제 처리하여 링크된 정보로 인해 재식별되는 위험요인을 제거

– 비식별화 기술은 재식별에 대한 일정한 한계를 가지므로 비식별화 목적을 명확히 하여 재식별에 대한 위험요소 최소화

– 정보의 유용성 수준 등을 감안하여 가능한 재식별에 의한 개인정보 침해가 없도록 사전에 고려

※ 전체 데이터중 분석 목적에 불필요한 정보들은 모두 삭제 또는 비식별화 처리 등

| 비식별화 적용시 사전 고려사항 |

- 빅데이터 자료 내에 어떠한 유형의 개인 정보가 포함되어 있는가?
- 누가 정보에 접근할 것이며, 어떠한 목적으로 활용할 것인가?
- 빅데이터 자료 내에 유일하거나 흔하지 않은 특징(유사 식별자)으로 인하여 재식별 가능성이 있는가?
- 수집된 정보가 누군가 혹은 무엇인가와 연관시키기 위하여 재식별 목적으로 쓰일 수 있는가?
- 비식별된 자료나 정보를 재식별하기 위한 다른 정보들이 존재할 가능성이 있는가?
- 비식별 정보 혹은 자료가 재식별 되었을 경우의 피해 결과는 어떠한 것이 있는가?

③ 정보가 식별 될 수 있는 리스크를 고려하여 사후관리 철저

– 주기적으로 재식별에 대한 리스크를 검토하고 리스크를 통제 할 수 있는 메커니즘을 확보

– 분석에 사용한 비식별화 처리 자료가 기술발전 또는 관련 정보의 추가 공개 등으로 재식별화 가능한지 정기 점검

– 빅데이터 분석 등의 과정에서 불필요한 개인정보가 새로 생성되거나 비식별화 처리된 정보가 재식별화 된 경우에는 지체없이(통상 5일이내) 삭제하거나 비식별화 처리

– 비식별화 자료가 불필요하거나 더 이상 활용도가 없는 데이터는 폐기

– 재식별 사례를 분석하여 주기적으로 비식별화 처리 기법 개선에 반영

| 참고 : 재식별 가능성이 높은 정보(예시) |

- 소수 집단에 관한 정보(90대 이상 연령자, 도서산간 거주자, 희귀질병감염자 정보 등)
- 연속하여 공개되는 패널 데이터 등(분기별 공개하는 환자진료 및 처방에 따른 회복 관련 정보 등)
- 링크정보를 가지고 있는 집단에게 정보를 공유 · 개방하는 경우(자동차 번호별 소유자를 알고 있는 처리자에게 자동차 번호를 제공하는 경우 등)
- 집단과 그 구성원이 알려져 있는 경우로서 동일 속성을 가진 집단에 관한 정보

Ⅲ. 18가지 비식별화 기술 활용 방법

[비식별화 주요 기술 및 활용 예]

처리기법	세부 기술	주요 내용 및 처리 예
가명처리 (Pseudonymisation)	① 휴리스틱 익명화 ② K-익명화 ③ 암호화 ④ 교환 방법	개인정보 중 주요 식별요소를 다른 값으로 대체하여 개인 식별을 곤란하게 함 (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대 서울 거주, 국제대 재학
총계처리 (Aggregation)	⑤ 총계처리 ⑥ 부분집계 ⑦ 라운딩 ⑧ 데이터 재배열	데이터의 총합 값을 보임으로써 개별 데이터의 값을 보이지 않도록 함 (예) 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm
데이터 값 삭제 (Data Reduction)	⑨ 속성값 삭제 ⑩ 속성값 부분 삭제 ⑪ 데이터 행 삭제 ⑫ 식별자 제거를 통한 단순 익명화	데이터 공유 · 개방 목적에 따라 데이터 셋에 구성된 값 중에 필요 없는 값 또는 개인식별에 중요한 값을 삭제 (예) 홍길동, 35세, 서울 거주, 한국대 졸업 → 35세, 서울 거주 (예) 주민등록번호 901206-1234567 → 90년대 생, 남자 (예) 개인과 관련된 날짜 정보(자격 취득일자, 합격일 등)는 연 단위로 처리
범주화 (Data Suppression)	⑬ 범주화 ⑭ 랜덤 올림 방식 ⑮ 범위 방법 ⑯ 제어 올림	데이터의 값을 범주의 값으로 변환하여 명확한 값을 감춤 (예) 홍길동, 35세 → 홍씨, 30-40세
데이터 마스킹 (Data Masking)	⑰ 임의 잡음 추가 ⑱ 공백과 대체	공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 주요 개인식별자가 보이지 않도록 처리하여 개인을 식별하지 못하도록 함 (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학

1. 가명처리(Pseudonymization)

- **개념** 개인 식별이 가능한 데이터에 대하여 직접적으로 식별 할 수 없는 다른 값으로 대체 하는 기법
- **처리대상 식별정보** 성명, 기타 고유특징 (출신학교, 근무처 등)

- **장점** 그 자체로는 완전 비식별화가 가능하며 데이터의 변형, 변질 수준이 적음
- **단점** 일반화된 대체값으로 가명 처리함으로써 성명을 기준으로 한 분석에 한계 존재

| 비식별화 실무 적용 방법 |

• 휴리스틱 익명화(Heuristic Pseudonymization)

– 식별자에 해당하는 값들을 몇 가지 정해진 규칙으로 개인정보를 숨기는 방법이다. 혹은 사람의 판단에 따라 가공하여 자세한 개인 정보를 숨기는 방법이기도 함
예를 들어, 성명을 홍길동, 임꺽정 등 몇몇 일반화된 이름으로 대체하여 표기하거나 소속 기관명을 화성, 금성 등으로 일반적 명칭을 쓰지 않는 몇몇 대명사로 대체하도록 사전에 규칙을 정하여 수행함
이 방법은 식별자의 분포를 고려하거나 수집된 자료의 사전 분석을 하지 않고 모든 데이터를 동일한 방법으로 가공하기 때문에 사용자가 쉽게 이해하고 활용할 수는 있음. 반면 휴리스틱 익명화 적용 이후의 데이터 유용성이 떨어지고 활용할 수 있는 대체 변수의 한계가 있음.
다른 값으로 대체하는 일정한 규칙이 노출되는 취약점이 있음. 따라서 개인을 쉽게 식별할 수 없도록 규칙의 세심한 고려가 필요함

– **적용정보** 성명, 사용자 ID, 소속(직장)명, 기관번호, 주소, 신용등급, 휴대전화 번호, 우편번호, 이메일 계정

• K-익명화(K-anonymity)

– 동일한 속성 값을 가지는 데이터를 k개 이상으로 유지하여 데이터를 공개하는 방법임. 지정된 속성이 가질 수 있는 값을 일정 수준(k개) 이상으로 유지함으로써 프라이버시 누출을 방지함.
예를 들어, 30개의 데이터에서 3-anonymity를 수행함. 이때 최소한 3개 이상의 데이터들끼리 같은 속성값으로 대체되어 전체 자료가 10개의 대표 데이터로 표현될 수 있는 기법임. 같은 속성값으로 대체하기 위하여 범주화(suppression), 일반화(generalization) 등의 방법을 사용 가능함.

– **적용정보** 나이, 신장, 주소, 우편번호, 소속(직장)

• 암호화(Encryption)

– 정보의 가공에 있어서 일정 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 방법임. 통상적으로 다시 유용하게 사용하기 위해서 복호가 가능하도록 암호화 / 복호화 값(key)을 가지고 있어서 key에 대한 보안방안도 함께 필요함. 활용 목적에 따라

단방향 암호화(one-way encryption 또는 hash)를 사용할 수 있으며 이 경우 이론상 개인정보로의 복호화가 원천적으로 불가능함. 단방향 암호화는 개인정보의 식별성을 완전히 제거하는 것으로, 양방향 암호화에 비해 더욱 안전하고 효과적인 비식별화 기술에 해당함. '중간영역(DMZ)'에 저장할 경우 암호화하고 있으나 강력한 비식별화 조치로서 암호화 기법을 적용하여 사용가능함.

– **적용정보** 주민등록번호, 여권번호, 의료보험번호, 외국인등록번호, 사용자 ID, 신용카드번호, 기관번호, 디바이스 ID, 생체정보, 민감정보

• 교환 방법(Swapping)

– 추출된 표본 레코드에 대하여 이루어짐. 미리 정해진 변수(항목)들의 집합에 대하여 데이터베이스의 레코드와 연계하여 교환함.
총계처리(aggregation)의 데이터 재배열(rearrangement)과 구분되어, 데이터 재배열은 레코드 값들간의 교환이 이루어진다. 반면 교환 방법은 사전에 정의된 외부값으로 대체됨.
따라서, 이는 민감한 속성이 있는 경우 그룹 내에서만 교환이 이루어질 경우 전체 그룹을 식별할 수 있는 위험성을 내포하고 있을 때 사전에 정해진 외부값으로 대체하여 민감 정보를 비식별화 함.

– **적용정보** 주민등록번호, 요양기관번호, 사용자 ID, 기관번호, 나이, 성별, 신체정보(신장, 혈액형 등), 소득, 휴대전화, 주소

2. 총계처리 (Aggregation)

- **개념** 개인정보에 대하여 통계값(전체 혹은 부분)을 적용하여 특정 개인을 판단할 수 없도록 함
- **처리대상 식별정보** 개인과 직접 관련된 날짜 정보(생일, 자격 취득일), 기타 고유특징(수입 지출, 신체정보, 진료기록, 병력정보, 특정소비기록 등의 개인 민감 정보) 등
- **장점** 민감한 정보에 대하여 비식별화가 가능하며 다양한 통계분석(전체, 부분)용 데이터 셋 작성에 유리함
- **단점** 집계 처리된 데이터를 기준으로 정밀한 분석이 어려우며 집계 수량이 적을 경우 데이터 결합 과정에서 개인정보 추출 또는 예측이 가능

| 비식별화 실무 적용 방법 |

• 총계처리(Aggregation) 기본방식

– 수집된 정보에 민감한 개인정보가 있을 경우 데이터 집단 또는 부분으로 집계(총합, 평균 등) 처리를 하여 민감성을 낮춤
예를 들어, 특정 나이 값이 있는 경우 집단의 평균 나이값(대표값)을 구한 후 각 개인정보 속성값을 구해진 대표값으로 대체하거나 해당 집단의 소득을 전체 평균을 구한 뒤 일정규칙의 오차를 가감하여 각 개인정보의 소득 속성값을 변환
단, 데이터의 평균이나 총합 등으로 표현할 경우 전체가 유사한 특정 속성을 지닌 개인으로 구성되어 있을 경우 단체의 대표 속성이 개인의 정보를 그대로 대변할 수도 있으므로 주의해야 함

– **적용정보** 나이, 신장, 소득, 카드사용액, 유동인구, 사용자수, 제품 재고량, 판매량

• 부분집계(Micro Aggregation)

– 분석 목적에 따라 부분 그룹만 비식별 처리. 즉, 다른 속성값에 비하여 오차 범위가 큰 항목이나 속성값에 대하여 통계값(평균 등)을 활용하여 값을 변환
예를 들어, 다양한 연령대의 소득 분포에 있어서 40대의 소득 분포 편차가 다른 연령대에 비하여 매우 크거나 특정 소득 구성원을 포함하고 있을 경우, 40대의 소득만 선별하여 평균값을 구한 후 40대에 해당하는 각 개인정보의 소득 속성값을 해당 평균값으로 대체함으로써 식별이 가능한 소득을 가진 40대 일부를 비식별 처리

– **적용정보** 나이, 신장, 소득, 카드사용액

• 라운딩(Rounding)

– 집계 처리된 값에 대하여 라운딩(올림, 내림, 사사오입) 기준을 적용하여 최종 집계 처리
일반적으로 총계 처리하는 기본방식에서 많이 쓰이는 값으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용
예를 들어, 23, 41, 57, 26, 33 등 세세한 나이의 속성값을 20, 30, 40, 50 등의 각 대표 연령대로 표기하거나, 3,576,000원, 4,210,000원 등의 소득 표기를 십만원 혹은 백만원 단위 이하를 절삭하여 3백만원, 4백만원 등으로 집계 처리하는 방식
범주화의 랜덤 올림 방법(random rounding)과도 방식이 유사하여 같은 의미로 사용하기도 함

– **적용정보** 나이, 신장, 소득, 카드지출액, 유동인구, 사용자 수

• **데이터 재배열(Rearrangement)**

– 기존 정보값은 유지하면서 개인정보와 연관이 되지 않도록 해당 데이터를 재배열. 즉, 개인의 정보가 타인의 정보와 뒤섞임으로써 전체 정보의 손상없이 개인의 민감 정보가 해당 개인과 연결되지 않도록 하는 방법
예를 들어, 여러 개인정보 중에서 나이, 소득 등의 특정 속성을 개인별로 서로 교환하여 재배치하게 되면 개인의 실제 나이와 소득과는 차이가 발생하는 비식별 자료를 얻게 되지만, 전체적인 통계적 분석 등에 있어서는 자료의 손실없이 분석을 할 수 있는 장점이 있음

– **적용정보** 나이, 신장, 소득, 질병, 신용등급, 학력

3. 데이터 값 삭제(Data Reduction)

- **개념** 개인정보 식별이 가능한 특정 데이터 값 삭제 처리
- **처리대상 식별정보** 쉽게 개인을 식별할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진 등), 고유식별정보(주민등록번호, 운전면허번호 등), 생체정보(지문, 홍채, DNA 정보 등), 기관 · 단체 등의 이용자 계정(등록번호, 계좌번호, 이메일 주소 등)임
- **장점** 민감한 개인식별 정보에 대하여 완전한 삭제 처리가 가능하여 예측, 추론 등이 어려움
- **단점** 데이터 삭제로 인한 분석의 다양성, 분석 결과의 유효성, 분석정보의 신뢰성을 저하시킴

| 비식별화 실무 적용 방법 |

• **속성값 삭제(Reducing Variables)**

– 원시 데이터에서 민감한 속성값 등 개인식별 항목을 단순 제거하는 방법
예를 들어, 주민번호, 나이, 성명이 나열되어 있는 경우 분석 목적에 따라 주민번호를 나이만으로도 대체 가능하다면 주민번호 속성값은 삭제, 이때, 남아 있는 정보 그 자체로도 분석의 유효성을 가져야 함과 동시에 개인을 식별할 수 없어야 하며 인터넷 등에 공개되어있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 함

– **적용정보** 주민등록번호, 성명, 전화번호, 계좌번호, 요양기관번호, 사용자 ID, 기관번호, 이메일 주소, 위치정보

• 속성값 부분 삭제(Reducing Partial Variables)

– 민감한 속성값에 대하여 전체를 삭제하는 방식이 아닌 해당 속성의 일부값을 삭제함으로써 대표성을 가진 값으로 보이도록 하는 방법
예를 들어, 상세 주소의 경우 부분 삭제를 통하여 대표지역으로 표현 가능(예: 서울특별시 중구 무교동 77번지 → 서울시 중구)하며 이러한 경우 범주화(suppression)의 경우와 유사할 수 있으나 범주화 방법은 주로 수치데이터에 적용하는 경우가 일반적인데 반하여 속성값 부분 삭제는 수치데이터를 포함하여 텍스트 데이터 등에도 폭넓게 활용 가능

– **적용정보** 주소, 위치정보(GPS), 전화번호, 계좌번호, 주민등록번호

• 데이터 행 삭제(Reducing Records)

– 타 정보와 비교하여 값이나 속성의 구별이 뚜렷하게 식별되는 정보 전체를 삭제. 즉, 특정하게 민감한 속성값 하나가 아닌 해당 정보를 가진 개인의 내용 전체를 제거하는 방법
예를 들어, 소득이 다른 사람에 비하여 뚜렷이 구별되는 값을 가진 정보는 해당 개인정보 전체를 삭제
이 방법은 통계분석에 있어서 전체 평균에 비하여 오차범위를 벗어나는 자료를 제거할 때에도 사용 가능

– **적용정보** 키, 소득, 질병, 카드지출액

• 준식별자 제거를 통한 단순 익명화(Trivial Anonymization)

– 단순 익명화 방법은 식별자뿐만 아니라 잠재적으로 개인을 식별할 수 있는 준식별자를 모두 제거함으로써 프라이버시 침해 위험을 줄이는 방법
예를 들어, 연예인 · 정치인 등의 가족 정보(관계정보), 판례 및 보도 등에 따라 공개되어 있는 사건과 관련되어 있음을 알 수 있는 정보 등 잠재적 식별자를 사전에 제거함으로써 연관성 있는 정보의 식별 및 결합을 예방
개인정보 유출 가능성을 최대한 줄일 수 있지만 데이터 활용에 필요한 정보까지 사전에 모두 없어지기 때문에 데이터의 유용성이 낮아지는 문제 발생

– **적용정보** 나이, 소득, 키, 몸무게 등 개별적으로는 단순한 개인정보이지만 분석 목적에 따라 사후 개인식별이 가능한 속성값이 될 수 있다고 판단되는 정보

4. 범주화(Data Suppression)

- **개념** 단일 식별 정보를 해당 그룹의 대표값으로 변환(범주화)하거나 구간값으로 변환(범위화)하여 고유 정보 추적 및 식별 방지
- **처리대상 식별정보** 쉽게 개인을 식별할 수 있는 정보(주소, 생년월일 등), 고유식별정보(주민등록번호, 운전면허번호 등), 기관 · 단체 등의 이용자 계정(등록번호, 계좌번호)임
- **장점** 범주나 범위는 통계형 데이터 형식이므로 다양한 분석 및 가공이 가능
- **단점** 범주, 범위로 표현됨에 따라 정확한 수치 값에 따른 분석, 특정한 분석 결과 도출이 어려우며, 데이터 범위 구간이 좁혀질 경우 추적, 예측이 가능

| 비식별화 실무 적용 방법 |

• 범주화(Data Suppression) 기본방식

– 은폐화 방법이라고도 하며 명확한 값을 숨기기 위하여 데이터의 평균 또는 범주의 값으로 변환하는 방식
단, 데이터의 평균이나 범주로 전체를 표현할 경우 특정 속성을 지닌 개인으로 구성된 단체의 속성 정보 공개는 그 집단에 속한 개인의 정보를 공개하는 것과 마찬가지의 결과가 나타나므로 이 경우는 비식별화 처리로 볼 수 없음 (예: 에이즈 환자 집단임을 공개하면서 특정인물 '갑'이 그 집단에 속함을 알 수 있도록 표시하는 것은 '갑'이 에이즈 환자임을 공개하는 것과 마찬가지임)

• 랜덤 올림 방법(Random Rounding)

– 개인식별 정보에 대한 수치데이터를 임의의 수 기준으로 올림(round up) 또는 절사(round down)하는 기법으로서 민감성이 높은 정보에 대하여 대표값(범주화)으로 처리. 총계처리(Aggregation)의 라운딩(rounding) 방법과 같은 의미로 사용하나 수치데이터 이외의 데이터에도 확장 적용 가능
예를 들어, 나이, 우편번호 등과 같은 수치 정보로 주어진 식별자는 일의 자리, 십의 자리 등 뒷자리 수를 숨기고 앞자리 수만 나타내는 방법 (나이 : 42, 45, 49, 43, 42 → 40 혹은 40대로 대표값 표현)

– **적용정보** 나이, 소득, 카드지출액, 우편번호, 유동인구, 사용자 수

• **범위 방법(Data Range)**

– 개인식별 정보에 대한 수치데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로서 해당 값의 분포(범위(range), 구간(interval)로 표현
예를 들어, 소득의 경우 3,300만원은 3,000만원~4,000만원으로 대체 표기

– **적용정보** 서비스 이용 등급, 처방정보(횟수, 기간 등), 위치정보, 유동인구, 사용자 수, 분석 시간 / 기간

• **세분정도 제한 방법(Subdivide Level Controlling)**

– 개인정보 중 단일 항목으로 개인식별이 될 수 있는 항목을 민감(sensitive) 항목 또는 높은 시각(high visibility) 항목이라 하는데, 이와 같은 민감한 항목을 상한(top), 하한(bottom) 코딩, 구간 재코딩(recoding into intervals) 방법을 이용하여 정보노출 위험을 줄일 수 있는 기법
예를 들어, 소득 항목 상한을 7,000만원, 하한을 3,000만원으로 정할 경우 해당 구간에 속하는 값은 '〉7,000만원'과 '〈3,000만원'으로 각각 표현할 수 있음
그 사이의 속성값들에 대하여 구간 재코딩 방법을 적용하면 3,000만원~3,999만원 등으로 나누어 대체 표기할 수 있음

– **적용정보** 나이, 소득, 카드지출액, 서비스 이용 등급

• **제어 올림 방법 (Controlled Rounding)**

– 랜덤 올림 방법에서 어떠한 특정 속성값을 변경시킬 때 행과 열의 합이 일치하지 않는 단점을 해결하기 위해 행과 열이 맞지 않는 것을 제어하여 일치시키는 기법, 그러나 컴퓨터 프로그램으로 구현하기 어렵고, 복잡한 통계표에는 적용하기 어려우며 해결할 수 있는 방법이 존재하지 않을 수 있어 아직 현장에서는 잘 사용하지 않는 방법임

– **적용정보** 나이, 키, 소득, 카드지출액, 위치정보

5. 데이터 마스킹(Data Masking)

● **개념** 개인 식별 정보에 대하여 전체 또는 부분적으로 대체값(공백, '*', 노이즈 등)으로 변환

● **처리대상 식별정보** 쉽게 개인을 식별할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진 등), 고유식별정보(주민등록번호, 운전면허번호 등), 기관 · 단체 등의 이용자 계정(등록번호, 계좌번호,

이메일 주소 등) 등

- **장점** 완전 비식별화가 가능하며 원시 데이터의 구조에 대한 변형이 적음
- **단점** 과도한 마스킹 적용 시 필요한 정보로 활용하기 어려우며, 마스킹의 수준이 낮을 경우 특정한 값의 추적 예측 가능함

| 비식별화 실무 적용 방법 |

• 임의 잡음 추가 방법(Adding Random Noise)

– 소득과 같은 민감 개인식별 항목에 대해 임의의 숫자 등의 잡음 추가(더하거나 곱)하여 식별 정보 노출을 방지하는 기법
예를 들어, 생년월일의 경우 실제 생년월일(DoB)에 사전에 정의한 6개월의 잡음을 추가한다고 한다면 원래의 생년월일 데이터에 1일부터 최대 6개월의 날짜가 추가되어 기존의 자료와 오차를 가질 수 있게 적용
이 방법의 특징은 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않음. 하지만 경우에 따라 잡음값은 속성을 유지하면서 추적, 예측을 방지하기 위한 마스킹 값으로 적용되어 데이터 값과는 무관하기 때문에 분석이나 유효한 데이터로 활용할 수 없음

– **적용정보** 주민번호, 사용자 ID, 성명, 생년월일, 키, 나이, 민감상병의 주상병/부상병 코드, 전화번호, 주소

• 공백(blank)과 대체(impute) 방법

– 빅데이터 자료로부터 비식별 대상 데이터를 선택한 후, 선택된 항목을 공백으로 바꾼 후에 대체법(imputation)을 적용하여 공백부분을 채우는 기법
공백 이외에도 특수문자(‘ * ’, ‘ _ ’ 등이나 전각 기호)로 처리하는 경우가 많음

– **적용정보** 주민번호, 사용자 ID, 성명, 민감상병의 주상병 / 부상병 코드, 전화번호, 주소

Ⅳ. 실무 적용 사례

1. 국민건강 주의 예보 시범서비스 구축

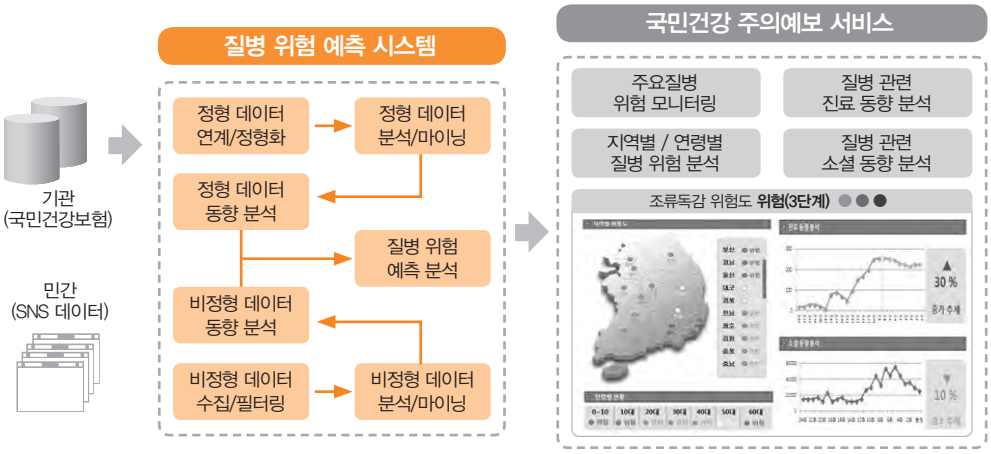
시범사업 개요

▶ **주관 / 참여기관**

- **주관기관 :** 국민건강보험공단
- **참여기관 :** 다음소프트

▶ **주요내용**

- 주요 유행성 질병에 대한 건강보험 정보와 소셜미디어 정보를 융합하여 질병 예측모델 개발
- 주요 유행성 질병에 대한 지역별, 연령별 진료동향, 위험동향, 소셜동향 등 종합정보 제공



개 발 내 용

▶ **활용데이터**

- **국민건강보험공단 :** 국민건강정보(진료내역 데이터)
- **다음소프트 :** SNS 데이터(트위터, 카페)

▶ **분석기법**

- **비정형 빅데이터 분석 :** 자연어처리기술을 이용한 텍스트마이닝 기법 사용
- **정형 데이터 분석 :** 통계분석, 기계학습 기술을 이용한 데이터마이닝 기법 사용
- **분석 결과 시각화 :** 자바스크립트 기반 시각화툴을 이용하여 웹상에 구현

  ※ 텍스트마이닝을 위한 지식구축 인력과 예측모델 튜닝을 위한 통계 전문가 운용

▶ **서비스 내용**

- 주요 유행성 질병의 위험도와 동향을 한눈에 파악할 수 있는 대시보드 서비스 제공
- 지역별 주요 유행성 질병 위험도 정보, 지역 내 질병 관련 진료 동향 및 연령별 진료 현황정보 제공
- 주요 유행성 질병 진료현황, 과거 진료통계, 질병 동향 및 고위험 지역 정보 제공
- 소셜 데이터에 발현된 주요 유행성 질병 관련 키워드의 동향, 연관 키워드/문서, 질병에 대한 관심도, 인식 상태, 주요 내용을 직관적으로 파악할 수 있게 제공

▶ **기대효과**

• 주요 감염병 유행을 예측하여 관련 기관과 국민 개개인이 유행에 대비하고 예방할 수 있도록 지원함으로써 국민건강 증진 및 사회적 편익 극대화

비식별화 대상 및 방법

▶ **비식별화 조치 필요 정보**

• **개인정보 :** 주민등록번호, 연령, 주소, 요양기관기호 • **사생활정보 :** 소득, 민감상병

▶ **비식별화를 위한 처리 기법**

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓	✓	✓	

구체적 사례

▶ **비식별화 조치 필요 정보**

• 국민건강보험공단에서 수집 · 분석의 대상이 되는 정보는 개인정보 및 민감한 사생활정보를 포함하고 있는 경우가 많아 고의적 · 우발적 개인정보 유출을 방지하기 위한 방안이 필요했다. 이에 수집 · 분석 대상에 포함된 개인정보를 텍스트마이닝, 패턴매칭 기술을 통해 검증 및 대체문자로 치환하고 있다. 계좌번호, 성명, 이메일, 전화번호, 주민등록번호, 주소, 휴대전화번호 등의 개인식별정보를 탐지 및 치환하며 탐지 가능한 개인식별정보를 추가 · 수정 · 삭제할 수 있는 기능을 제공한다.

▶ **적용 예시**

1. **가명처리 :** (식별번호 대체)
   – 요양기관기호(8자리) → 요양기관대체번호(6자리) 예) 31100678(일산병원) → 123456
2. **삭제 :** (전부 또는 일부삭제)
   – 주민등록번호(13자리) → 삭제 예) 110011–1479712 → “”
   – 주소 → 16개 시도 예) 11110(서울특별시 종로구 삼봉로 43) → 11(서울특별시)
3. **범주화 :** (그룹화)
   – 연령(0~80세이상) → 18개층(5세 단위 구간) 예) 53세 → 12(50~54세 구간)
   – 소득 → 보험료분위(전체 대상자(세대)를 20분위 균등분할) 예) 보험료 103,530원 → 14분위
4. **마스킹 :** (특수문자 대체)
   – 공단에서 규정한 민감상병의 주상병, 부상병코드
   1) 상병기호의 대분류만 표시 : 예) A**** (A : 특정감염 및 기생충성 질환, 콜레라)
   2) 전체 상병기호 표시하지 않음 : 예) **** (D : 남성 생식기관의 양성 신생물)

2. 보건의료 빅데이터 활용 시범사업

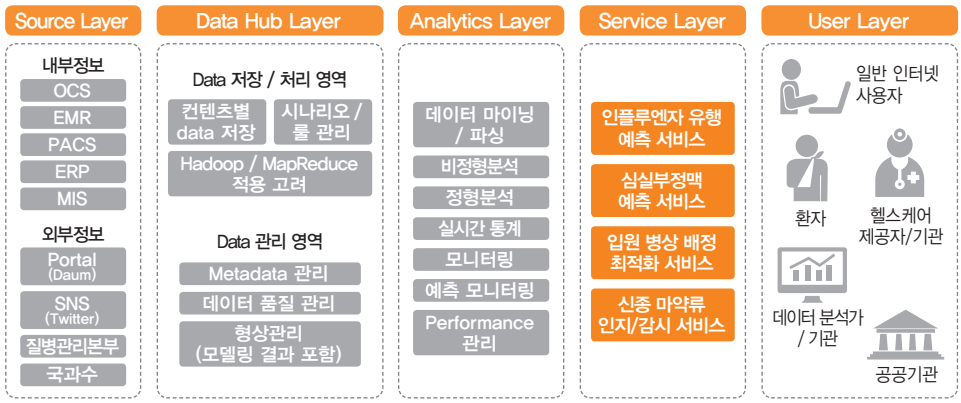
시범사업 개요

▶ **주관 / 참여기관**

- **주관기관 :** 서울아산병원
- **참여기관 :** 한국전자통신연구원, 한국마이크로소프트(유), (주)테크아이, 켐아이넷(주), 한국쌔스 소프트웨어(유)

▶ **주요내용**

- 보건의료 질 향상 및 비용 절감을 위한 보건의료 빅데이터 활용 서비스 개발



개 발 내 용

▶ **활용데이터**

- 검색 데이터, SNS 데이터, 입 / 퇴원 기록, 병원 경영자료, 실시간 환자 심전도 / 심박수, 진료 데이터, 국과수 마약류 관련 DB, SNS 데이터

▶ **분석기법**

- **비정형 / 정형 빅데이터 분석 :** 데이터 크롤링 / 필터링 기법, 자연어처리 기법 활용한 텍스트 마이닝 기법, 통계분석 및 기계학습 기술을 이용한 데이터 마이닝 및 모델링 기법, 패턴 분석 / 비교 및 인공신경회로망 알고리즘 사용
- **분석 결과 시각화 :** MS Windows Azure 활용

▶ **서비스 내용**

- 인플루엔자 현황 및 예측 동향 웹서비스 형태로 제공
- 입원 병상 배정 최적화 모델 병원 내 시스템에 구축/활용
- 심실부정맥 예측 모델 개발 후 대시보드 형태로 병원 내 시스템에 구축/활용
- 마약류 인지 현황 제공, 위험등급별 분류 후 관련정보 제공, 출현현황 뉴스레터 제공

▶ **기대효과**

- **공중보건 분야 :** 전염병 발생 및 불법 약물 전파와 같은 긴박한 순간에 미리 대비하고, 빠른 의사결정을 도움으로써 의료대응와 관련한 사회적 비용 대폭 절감

비식별화 대상 및 방법

▶ 비식별화 조치 필요 정보

• **개인정보 :** 성명, 시/군/구 보다 작은 단위의 지역정보 (읍/면/동 이하 상세주소), 전화번호 (자택, 직장, 이동전화, Fax 모두 포함), 이메일주소, 주민등록번호, 외국인등록번호, 여권번호, 등록번호, 건강보험증번호, 은행계좌번호, 자격/면허번호, 차량번호, 바이오정보 (지문, 얼굴, 홍체, 정맥, 음성, 필적 등), 유전자정보, 홈페이지 회원 ID, 사번, 비밀번호

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓		✓	

구체적 사례

❶ 본 기관에서는 개인식별 정보를 내부적으로 정의하기 위해 HIPAA, ISO/TS 25237:2008을 검토하여 최종적으로 20가지 개인식별정보를 정의하였고, 이에 대해 익명화를 실시함

❷ 의료정보는 영어와 한글이 혼용되고 있으며, 다양한 약어와 전문용어들이 많기 때문에 일반적인 자연언어처리 방법을 적용하기 어려움. 본 기관에서는 구조화된 정보는 테이블 내 컬럼 정보를 삭제하고, 비구조화된 정보는 regular expression rule을 작성하여 text 정보 중 개인식별정보들이 있으면 masking 처리함

❸ 더불어, 조합을 통해 개인식별이 가능한 quasi identifier도 방지하기 위해서 5명 미만의 개인 의료정보는 제공하지 않음

▶ 적용 예시

1. **가명처리 :** (등록번호 대체)
   – 환자등록번호 (8자리) → 임의로 생성된 번호 (8자리)
   예) 11111111 → 92429988

2. **삭제 :** (테이블 컬럼)
   – 성명 → 삭제
   예) 홍길동 → “ ”
   – 주소 → 삭제
   예) 서울특별시 송파구 풍납2동 388–1 → “ ”

3. **마스킹 :** (특수문자 대체)
   – 주소
   예) 서울특별시 송파구 풍납2동 388–1 → ***** *** **** *****
   – 각종 연락처
   예) 전화: 010–111–1111 → 전화: *** – *** – ****

3. 빅데이터 기반 의약품 안전성 조기경보 서비스 구축

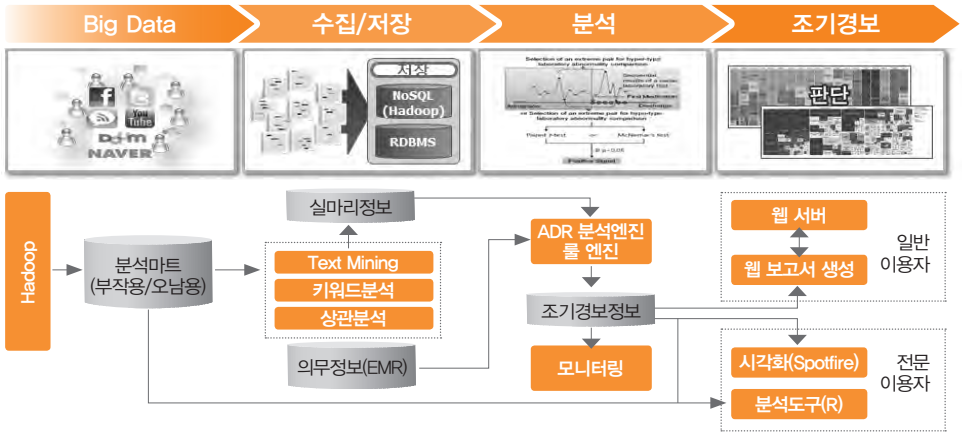
시범사업 개요

▶ **주관 / 참여기관**

- **주관기관 :** 에스지에이(주)
- **참여기관 :** 한국의약품안전관리원, (주)와이즈넛

▶ **주요내용**

- 빅데이터에서 의약품 부작용 및 오남용 사례를 수집 및 분석
- 의약품 부작용 가능성을 병원 의무정보를 기반으로 확인하여 조기 인지
- 의약품 오남용 사례를 파악하여 조기 대응을 위한 정보 제공



개 발 내 용

▶ **활용데이터**

- **와이즈넛 :** SNS (블로그, 지식인, 카페, 트위터), 뉴스(웹), 문헌(웹) 데이터
- **아주대학교병원 :** EMR(전자의무기록)

▶ **분석기법**

- **비정형 빅데이터 분석 :** 자연어처리기술을 이용한 텍스트마이닝 기법 사용
- **정형 데이터 분석 :** 통계분석과 OLAP 기술을 이용한 데이터마이닝 기법 사용
- **부작용 검증분석 :** 코호트 기반 연구방법, 환자 / 대조군 비교 알고리즘 사용
- **분석 결과 시각화 :** Spotfire를 이용한 웹기반의 Drill-Down 분석화면 제공

▶ **서비스 내용**

- 의약품 부작용 검증을 위해 추출된 실마리정보 제공
- 의약품 부작용에 대한 검증 결과 제공
- 의약품 오남용 사례 및 분석결과 제공
- 오남용 사례에 대한 지역별, 성별 등의 다면 분석결과를 시각적으로 제공
- 빅데이터에서 수집된 정보와 시스템 결과를 상세 분석할 수 있는 전문가 분석환경 제공

▶ **기대효과**

- 빅데이터를 활용, 국민복지와 건강을 위협하는 의약품 부작용 및 오남용을 조기에 발견하여, 선제적 대응을 통한 "안전한 사회"구현에 기여

비식별화 대상 및 방법

▶ 비식별화 조치 필요 정보

- **개인정보 :** 나이, 생년월일
- **사생활정보 :** 아이디, 진단명, 약처방 날짜, 진단검사 날짜, 검사수행 날짜

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓	✓		

구체적 사례

에스지에이에서 수집 · 분석의 대상이 되는 정보는 개인정보 및 민감한 사생활 정보를 포함하고 있는 자료가 많아 고의적 · 우발적인 개인정보 유출을 방지하기 위한 방안이 필요하다. 이에 수집 · 분석 대상에 포함된 개인정보를 랜덤키 생성, 패턴매칭 기술을 통해 대체문자로 치환 및 쉬프트 처리, 민감데이터 삭제처리를 하고 있다.

▶ 적용 예시

1. **가명처리 :** (식별번호 대체)
   – 환자 아이디를 고유 아이디로 생성하여 문자 형식으로 저장하여 개인식별을 곤란하게 함
   예) 환자 아이디(6자리) → 고유 아이디 생성(36자리)
   0001012 → E214F58E–9E3F–44B7–B3A3–9854BF439216
   – 날짜 데이터를 환자별 랜덤 숫자 −90 ~ 90 사이의 랜덤 숫자를 발생하여 식별이 곤란하게 함
   예) 날짜 : 랜덤 숫자 처리
   20101010 → 20110118

2. **삭제 :** (민감데이터 삭제)
   – 환자의 진단명중 민감한 정보를 삭제하여 사생활 정보의 식별을 곤란하게 함
   성명, 에이즈 감염, 비정상적인 염색체 이상, 낙태 등 삭제(민감데이터 삭제)
   예) 진단명 : AIDS → 해당 자료 삭제
   진단코드 : B20* → B20으로 시작하는 코드 삭제

3. **범주화 :** (그룹화)
   – 80세 이상 나이를 80으로 고정하여 명확한 나이를 감춤.
   예) 나이 : 80이상 → 80으로 고정
   86세 → 80세

4. 점포 평가 서비스

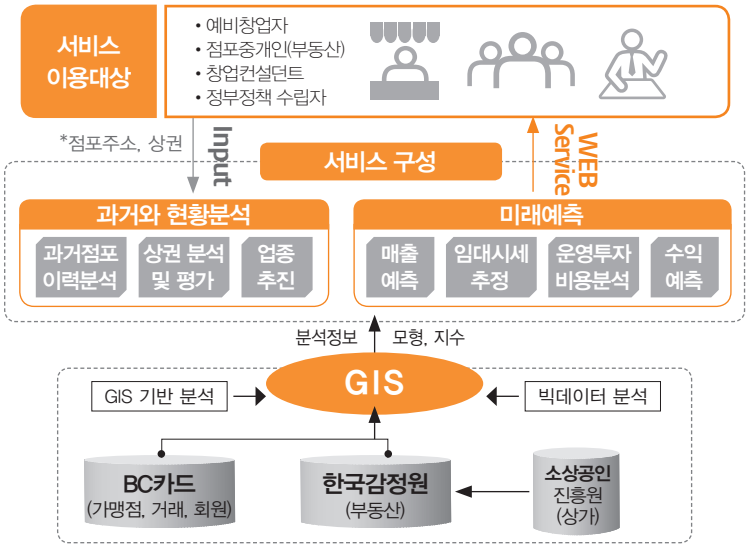
시범사업 개요

▶ **주관 / 참여기관**

- **주관기관 :** ㈜오픈메이트
- **참여기관 :** 비씨카드㈜, 한국감정원

▶ **주요내용**

- 약 1억건의 상가업소데이터, 6억건 이상의 카드사용 트래픽 데이터 기반 분석
- 동일지점의 최근 3년간 점포 개폐업 이력 추적
- 점포별 추정 매출/입지평가/상권평가



개 발 내 용

▶ **활용데이터**

- **비씨카드 :** 월 2억건씩 누적되는 카드거래 데이터 중 서울지역 2억건(3년치)
- **한국감정원 :** 부동산 임대시세 및 건축물 대장 데이터
- **소상공인진흥원 :** 월 300만건씩 누적되는 소상공인 상가정보(점포명, 업종, 주소, 전화번호) 총 1억건(3년치)
- **인문사회지리데이터 :** 인구, 가구세대, 36만 블럭

▶ **분석기법**

- **텍스트마이닝, 지오코딩 :** 월단위로 단절된 업소이력을 추적
- **통계분석 :** 다중회기분석, huff확률모형, 상권 / 업종평가모형, 입지진단모형
- **공간분석 :** 공간가중회기, 보간법

▶ **서비스 내용**

- 점포별 수년간의 개 · 폐업 이력을 추적하고, 업종변화별로 해당 점포의 매출 추정

• 창업자가 어떤 업종으로 창업하면 가장 높은 매출을 낼 수 있을지, 어떤 업종일 때 영업기간이 짧고 폐업율이 높았는지, 점포의 입지는 어떤 수준인지 등 창업결정을 위한 지표정보 제공
• 임대시세, 추정매출, 점포진단평점 등 입지상권분석에 필요한 기초정보 제공

▶ **기대효과**

• 입지특성에 맞지 않는 부적절한 업종의 개업을 예방하여 창업 실패율 감소
• 경험적, 계약우선의 점포거래 관행에서 데이터를 기반한 과학적인 창업컨설팅 유도

비식별화 대상 및 방법

▶ **비식별화 조치 필요 정보**

• **개인정보 :** 개별 상가업소 매출

▶ **비식별화를 위한 처리 기법**

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓	✓		✓	✓	

구체적 사례

상가업소의 매출정보는 창업자에게 매우 중요한 정보이나 개별업소의 매출을 개인정보 및 세원 노출에 대한 법적규제로 인해 제공되지 못하고 있음. 이에 대해 카드사에서는 개별업소에 대한 매출을 보정한(현금비율, 타사카드 비율 반영) 추정값으로 산정한 뒤 지역별, 업종별 유형화, 업종 그루핑(업종분류), 지역단위별 5개 단위 이하 업소매출 제거 등을 통해 통계화 한 뒤 정보서비스로 제공하고 있음

▶ **적용 예시**

1. **가명처리 :** (식별번호 대체)
   – 업소명 → 업소ID　　예) 김가네김밥 : 업소명 → 업소ID B3231123_23

2. **총계처리 :** (총합 집계)
   – 추정매출액 = 원시매출액 * 추정현금비율 * 카드사시장점유비(MS)
   지역별업종추정평균매출액 = Σ(업소별추정매출액) / 업소수
   매출범위로 환산 : 5,314,000원 → 추정매출 5,000천원~6,000천원 (3. 범주 / 범위화 기법과 혼용)

3. **범주화 :** (그룹핑)
   – 300만개 업소 → 1,500개 업종으로 그루핑할 수 있도록 유형화
   김가네김밥 → 음식 〉 분식　/　김밥천국　→ 음식 〉 분식

4. **범주마스킹 :** (특수문자 대체)
   – 업소전화번호가 핸드폰 번호일 경우 마스킹처리
   010–4333–1234 → 010 – **** – ****

5. 빅데이터 분석을 통한 심야버스 노선정책 지원

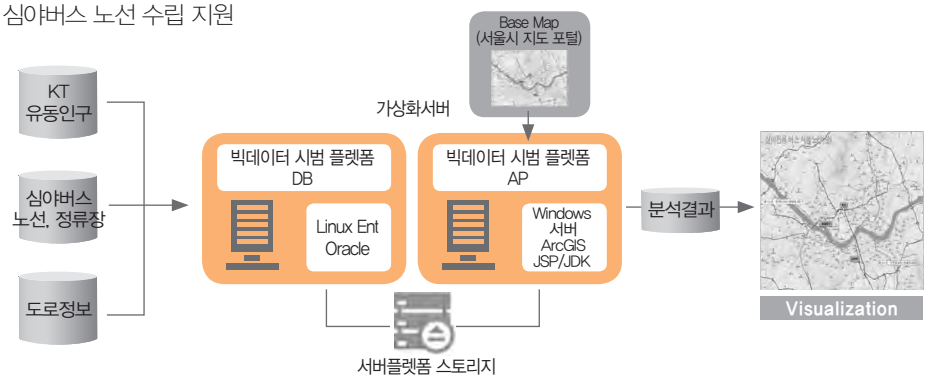
시범사업 개요

▶ **주관 / 참여기관**

• **주관기관 :** ㈜ KT　　• **참여기관 :** 서울특별시

▶ **주요내용**

• KT–서울시가 보유하고 있는 데이터와 분석기술을 활용, 심야시간대 서울 시민들이 활용하게 될 심야버스 노선 수립 지원



개 발 내 용

▶ **활용데이터**

• **KT :** CDR(call detail record) 데이터 ⇨ 유동인구파악
　　고객 통계데이터 ⇨ 목적지 및 이용대상 파악

• **서울특별시 :** 공공 데이터 ⇨ 최적 정류소 위치 선정

▶ **분석기법**

• **블록 단위 공간분석 :** 유동인구 파악 – 일정 크기 분할, 상권 · 인구수 기반 분할

• **거리 기반 알고리즘 적용**

– 정류소와의 거리, 유동인구 빈도별 가중치 부여를 통해 통행량 높은 정류장 추적

▶ **서비스 내용**

• KT의 위치정보기반 유동인구 데이터와 서울시 공공 교통 데이터를 융합하여 시민들에게 최적의 심야버스 혜택 제공

– 1일 1억건 이상의 CDR데이터를 통계 분석하여 심야시간(24:00–05:00) 서울시 유동인구가 많은 지역 분석

– 서울시가 보유 하고 있는 버스 정류장 위치, 도로정보 등 교통정보 데이터 활용하여 버스 경로 선정

▶ **기대효과**

• 효율적 노선서비스의 제공으로 서울 시민들에게 심야 최적의 교통서비스 제공

• 상대적으로 소득이 낮은 심야 경제 활동인구에게 경제적으로 기여

• 범죄에 취약한 심야 및 새벽시간대 심야버스 운행으로 범죄 예방 효과

비식별화 대상 및 방법

▶ 비식별화 조치 필요 정보

• **개인정보 :** 연령, 청구지 주소

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부			✓	✓		

구체적 사례

KT와 서울시의 사업에서 고객의 개인 정보라고 할 수 있는 데이터는 KT 내부에서 제거한 후에 그 외의 데이터를 분석에 활용

▶ 적용 예시

1. **삭제 :** (일부 삭제)
   – 업소명 → 업소 ID
   김가네김밥 : 업소명 → 업소 ID B3231123_23

2. **범주화 :** (헥사곤 형태 가공 및 통계처리)
   – 개인 정보에 가까운 데이터인 연령대 같은 경우 특정 지역(1km 핵사곤 형태)으로 가공 및 통계처리함, 개인을 식별하는 분석이 아니기 때문에 통계처리로 충분함

   예) 핵사곤 A : 연령 10대 1000명
   연령 20대 500명
   연령 30대 2000명
   …

   – 청구지 주소의 경우 우편번호 단위를 핵사곤에 매핑 후 지역별 통계로 보기 때문에 개인의 집 위치를 파악하지 않음

   예) 핵사곤 A 에 속하는 우편번호 : 690 – 022
   120 – 200 → 200명
   200 – 120

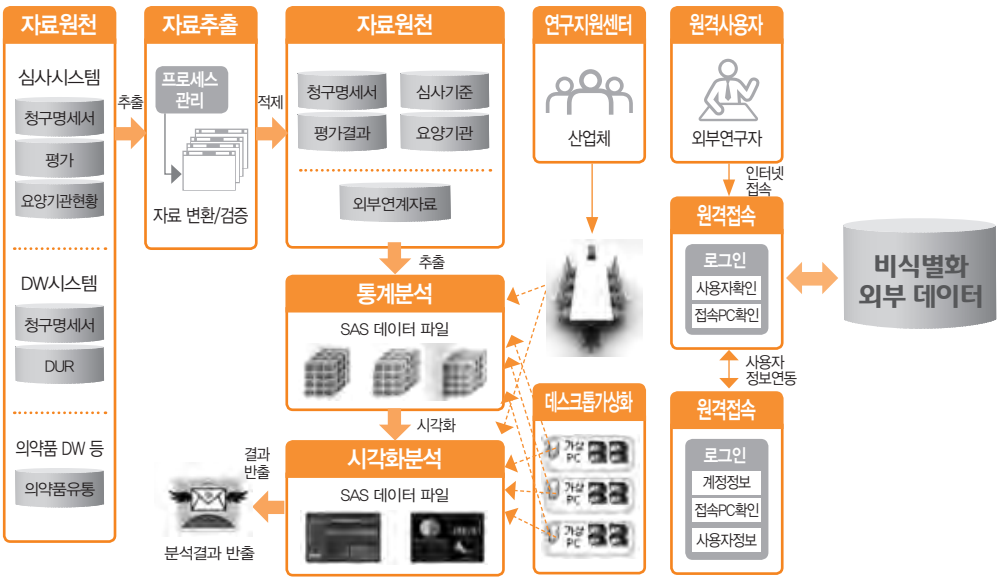
6. 병원정보 분석을 통한 맞춤형 의료정보 서비스

개 요

- 심평원에서 보유하고 있는 BigData를 활용한 의료기관 정보 제공 서비스
- 의사 및 의료진 정보제공, SNS 연계를 통한 정보의 공유
- 고객정보의 활용을 통한 맞춤형 진료과목 및 선택 진료 유도
- 진료과목별, 상병별 환자수에 따른 명의 찾기
- GPS정보를 통한 가까운, 저렴한, 진료 잘하는, 친절한 등의 검색을 통해 병원 찾기
- 의료보험 청구 패턴/경향의 장/단기적 모니터링 및 예측

서비스 내용

- 심평원에서 제공 받은 빅데이터를 활용한 웹 서비스



- 심평원 서비스를 원격에서 지원함
- 향후, 추가적인 개인화 서비스 개발에 활용함

비식별화 대상 및 방법

▶ 비식별화 조치 필요 정보

- **개인정보 :** 요양기관정보, 의사 정보, 간호사 정보, 주소, 요양기호
- **요양기관정보 :** 월환자수, 집중진료과목

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓	✓	✓	

구체적 사례

요양기호, 요양기관명, 수진자 성명, 상병정보, 진료과목, 전화번호, 주민등록번호, 휴대전화번호, 의사정보, 간호사정보, 처방통계정보 등의 비식별화 정보를 제공받아 SNS 및 웹서비스가 가능한 기능을 제공함

▶ 적용 예시

1. **가명처리 :** (식별번호 대체)
   – SNS와 연동시 주민등록번호 연계 불가능하도록
   주민등록번호 → 개인일련번호
   예) 820101–1000000 → 12345678
   – 요양기관기호 → 요양기관대체번호
   예) 31100678(일산병원) → 0000001

2. **삭제 :** (일부 삭제)
   – 일별 GPS 정보 → 위도 : 37. 29.32076(27도, 29분, 32.076초)
   경도 : 126.5533138(126도, 55분, 33.138초)
   예) 37. 29.32076 → 37. 29.00000

3. **범주화 :** (그룹화)
   – 월 감기로 처방건수(0~10세 이상)
   – 월 전문의 상시상주기간(15~21일 내외)

4. **마스킹 :** (특수문자 대체)
   – 의사, 전문의 번호 등 식별가능 의료정보
   1) 의사변호번호 : 32***
   2) 주민등록번호 : ******_*******

7. NFC / LBS 기반의 소액 결제 정보와 마케팅 트렌드 제공 서비스

개 요

- ▶ 카드 결제 등을 추적하여 고급 마케팅 정보로 활용
- ▶ 소지역 즉 LBS기반의 로컬 상권에서 카드, 금융 서비스를 통한 마케팅 트렌트 추이 분석 및 세분화된 표적 마켓에 대한 고객 응대 서비스

서비스 내용

- ▶ 초단거리 무선통신 기술로 대략 10cm 이내의 기기 간에 통신을 가능하게 해줌. NFC는 기본적으로 휴대폰에서 사용할 목적으로 활용되는 LBS를 통한 위치 서비스임
- ▶ NFC와 LBS를 통한 결제 지점, 소비 품목, 실시간 결제액 추이 등
- ▶ 결제 시점의 맥락을 추론하여 결제 전후의 이용자 행태나 동일 집단의 이용자 추이를 분석

비식별화 대상 및 방법

▶ 비식별화 대상

- **개인정보 :** 주민등록번호, 연령, 주소, 소득, 직업, 금융거래내역, 신용정보

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부			✓	✓	✓	

구체적 사례

개별 고객과 계약한 서비스를 제공하기 위한 목적으로 각 금융사는 실시간 고객의 신용 카드 사용 등의 결제 정보와 해당 개인정보를 내부 통제망 안에서 목적별 제한 규정에 따라 상세히 처리할 수 있으나, 빅데이터 서비스 관점에서는, 아파트 동(棟) 혹은 단지(소지역) 단위의 통계치로 묶어 제공하며, 아래의 3가지 방법을 주로 사용함

▶ 적용 예시

1. **삭제 :** (일부 삭제)
   - – 성명 : 삭제
   - – 주민등록번호: 삭제(연령, 성 추출)
   - – 주소 : 상세주소 → 소지역 단위

예) 서울특별시 종로구 삼봉로 43 삼봉아파트 101동 101호 → 서울특별시 종로구 삼봉로 43 삼봉아파트 101동

2. 범주화 : (그룹화)

– 연령 : 실제 연령 → 5세 단위 구간 혹은 Baby Boomer, 10대, 골드미스 등
예) 53세 → 베이비 부머, 50~55세

– 소득 : 실제수입(자기기입, 증빙, 추정) → 일정 급간화
예) 연수입 34,100,000원 → 3천~3500만원

– 카드사용금액 : 승인금액 → 일정 급간화
예) 일시불 카드소비 금액 1,553,400원 → 150만원~200만원

4. 총계 : (합계, 평균, 빈도, 추세, 추정)

– 소득: 개인의 소득 → 소지역 단위로 총계화
예) 〈삼봉아파트 101동 구역〉 홍길동 연수입 34,100,000원 / 성춘향 연수입 27,000,000원 / 심학도 연수입 41,000,000원 → 삼봉아파트 101동 연수입(총계, 평균, 등록 고객수, 작년대비 성장, 대푯값)

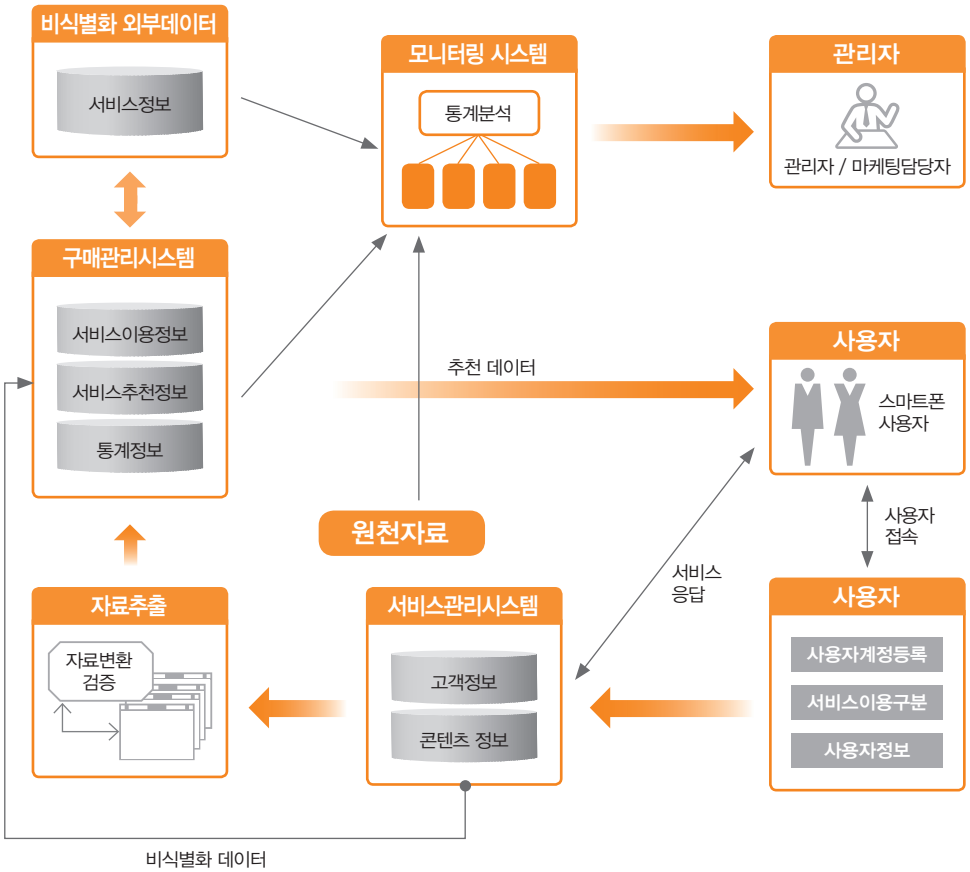
8. 도서 추천 및 유통 서비스 제공

개 요

- 물류 이동 정보 제공, SNS 연계를 통한 도서 구매 정보의 공유
- 고객 정보의 활용을 통한 맞춤형 도서 추천 및 유통 관리
- 구매 도서별, 사용자 성향별 선호 작가 찾기
- GPS 정보를 통한 구매 도서의 이동 상태 및 배달 정보 제공

서비스 내용

- 온라인 서점에서 보유하고 있는 도서 구매 및 유통 빅데이터를 활용한 도서 추천 및 유통 정보 제공 서비스



비식별화 대상 및 방법

▶ 비식별화 조치 필요 정보

- **개인정보 :** 사용자ID, 주소, 전화번호, 주민등록번호, 휴대전화번호, 수신자성명
- **물류기관정보 :** 유통담당기관, 물품배달기록

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓	✓	✓	

구체적 사례

사용자ID, 수신자성명, 전화번호, 주민등록번호, 주소, 휴대전화번호 등의 비식별화 정보를 제공받아 물류센터, SNS, 구매추천 서비스 및 웹 서비스가 가능한 기능을 제공한다.

▶ 적용 예시

1. **가명처리 :** (식별번호 대체)

   – 물류센터, SNS 등과 연동시 주민등록번호 연계 불가능하도록
   주민등록번호(13자리) → 개인일련번호(8자리)
   예) 820101–1000000 → 12345678

2. **삭제 :** (일부삭제)

   – 물류의 이동위치에 대한 GPS 정보 → 위도 : 37. 29.32076(27도, 29분, 32.076초)
   경도 : 126.5533138(126도, 55분, 33.138초)
   예) 37. 29.32076 → 37. 29.00000

3. **범주화 :** (그룹화)

   – 물류도착지에 대한 지역 범주화(서울, 종로)

4. **마스킹 :** (특수문자 대체)

   – 개인에 대한 식별 가능 정보
   1) 주민등록번호 : 70****–*******
   2) 주소 : 서울시 종로구 종로1가 *******

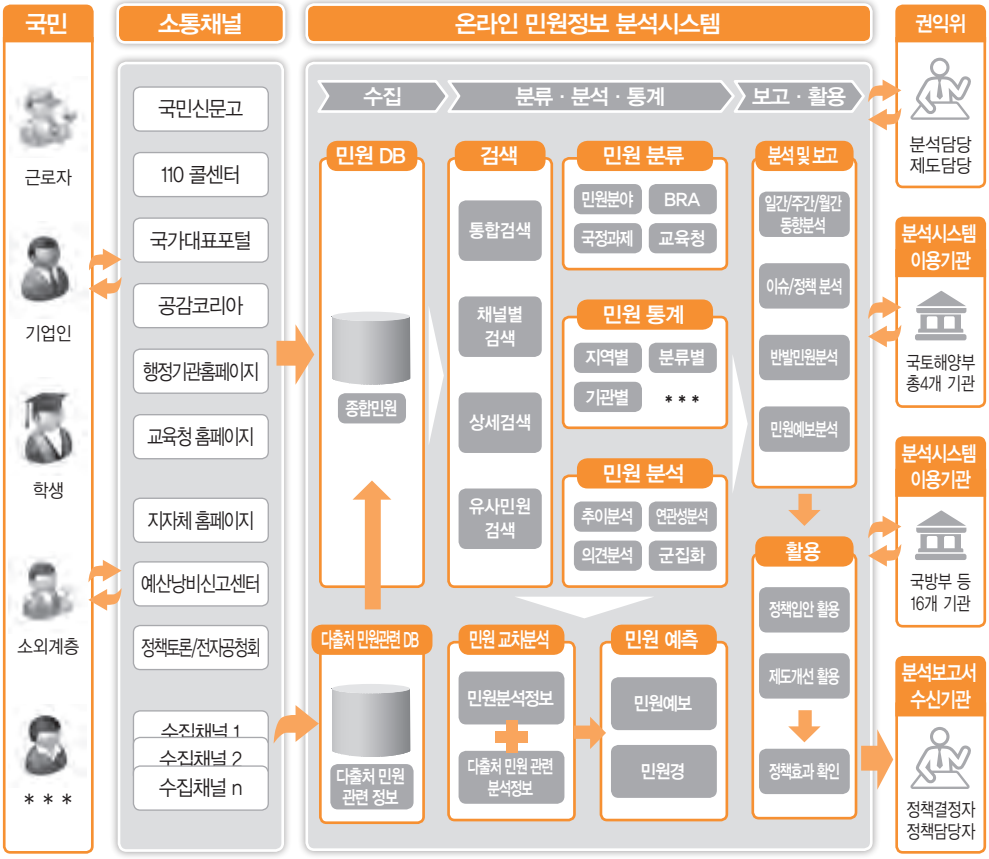
9. 민원처리 분석 활용

개 요

- 민원, 제안, 콜센터 상담을 통해 축적된 민원 데이터를 종합적, 체계적으로 분석하여 정책에 환류할 수 있도록 지원
- 중앙행정기관, 교육청, 지방자치단체 게시판 및 언론기사 등을 수집 및 분석
- 국민의 소리 획득 및 분석시간 단축 등을 통한 정책 개선 및 제도개선 지원

서비스 내용

- 일일 1만 5천건 이상 민원 및 상담 정보 수집 및 중복 제거
- 핵심키워드별 / 문서분류별 특성정보와 정형통계정보 간의 다차원 분석 및 추이분석, 정형 보고서 제공
- 통계적 모델링을 통해 지역별, 기관별, 주제별 등 예측 및 급등에 대한 경보 기능 제공



비식별화 대상 및 방법

▶ 비식별화 대상

• **개인정보 :** 성명, 주민등록번호, GPS, 주소

▶ 비식별화를 위한 처리 기법

처리기법	가명처리	총계처리	삭제	범주화	마스킹	기타
적용여부	✓		✓	✓	✓	

구체적 사례

분석의 대상이 되는 민원정보는 민감한 개인정보를 포함하고 있는 경우가 많기 때문에 고의적 · 우발적 개인정보 유출을 방지하기 위한 방안이 필요하여, 문서 추출 및 비식별화 모델을 이용하여 민원정보 분석 데이터를 제공함.

▶ 적용 예시

개인식별정보 검증 전 민원내용	개인식별정보 검증 및 치환 후 민원내용
… 서울시 강북구 수유3동 123-4의 담벼락에 붙어 있는 통신주가 첨부한 사진과 같이 곧 넘어갈것 같이 옆으로 쓰러져 있길래 KT에 연락하여 신고 하였더니 KT 도봉지사 특별기동팀의 홍길동 과장 (010-1234-5678, 02-3456-7890)이 2013년 1월 28일 오후 2시경 현장 답사하더니 2~3일 내에 새것으로 교체해준다 하였습니다. …	… 서울시 ***주소***의 담벼락에 붙어 있는 통신주가 첨부한 사진과 같이 곧 넘어질 것 같이 옆으로 쓰러져 있길래 KT에 연락하여 신고하였더니 KT 도봉지사 특별기동팀의 ***이름***(***휴대전화번호***, ***전화번호***)이 2013년 1월 28일 오후 2시경 현장 답사하더니 2~3일 내에 새것으로 교체해준다 하였습니다. …
… 이런 비윤리적인 행동과 비도덕함을 고발합니다. 이에 따른 피해보상과 민원, 법률제기 및 후속조치에 관한 방법을 질의합니다. 성명 : 홍길동, 010-1234-1234(abc@abc.co.kr) 업무담당자 : SH공사 보상본부 보상기준팀 이순신 대리 02-1234-4321 현재 본인들도 업무과오와 실수를 인정하지만 이주대책 책임에 대해서는 내부규정규칙만 재차 강조하는 상황입니다. …	… 이런 비윤리적인 행동과 비도덕함을 고발합니다. 이에 따른 피해보상과 민원, 법률제기 및 후속조치에 관한 방법을 질의합니다. ***이름***, ***휴대전화번호***(***이메일***) 업무담당자 : SH공사 보상본부 보상기준팀 ***이름***, ***전화번호*** 현재 본인들도 업무과오와 실수를 인정하지만 이주대책 책임에 대해서는 내부규정규칙만 재차 강조하는 상황입니다. …
… 이름 : 홍길동(551111-1111111) 010-5555-6666 031-666-8888 애드컴 188-81-4444 / 110001-1100000 본사 : 경기도 남영주시 도농동 123-4 진관일반산업단지 12-3 블럭 저희는 LH공사에서 조성중인 일반산업단지를 분양받고자 11년 12월 25일 남영주시 기업지원과 홍길동계장의 심사를 거친 후 업주계약체결을 하고 11년 12월 30일 LH공사와 분양계약을 하였습니다. …	… ***이름***(***주민등록번호***) ***휴대전화번호***, ***전화번호*** 애드컴 ***사업자등록번호*** / ***법인등록번호*** 본사 : 경기도 ***주소*** 저희는 LH공사에서 조성중인 일반산업단지를 분양받고자 11년 12월 25일 남영주시 기업지원과 ***이름***의 심사를 거친 후 업주계약체결을 하고 11년 12월 30일 LH공사와 분양계약을 하였습니다. …

Ⅴ. 활용 예제

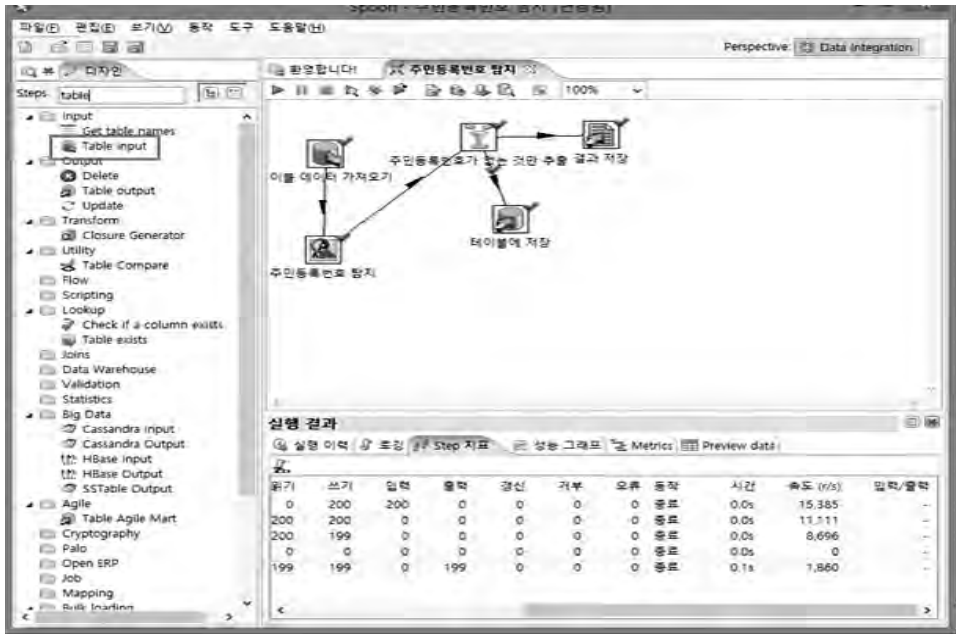
● (시나리오) 오픈소프트웨어인 Pentaho Data Integration을 활용하여 식별정보인 주민등록번호를 탐지 및 비식별화 처리하는 과정을 설명

① 빅데이터가 포함된 DB 연동 및 워크 플로우 작성
② 주민번호 탐지를 위한 정규식 등록
③ 주민번호 탐지 및 추출
④ 주민번호 비식별 처리
⑤ 데이터 저장

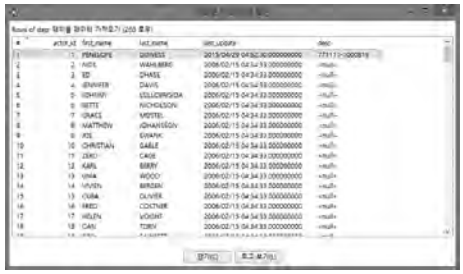
- **Pentaho Data Integration :** 데이터 가공처리하는 ETL 도구로 다양한 DB 연동을 통한 개인정보 탐지, 비식별화, 필터링 가공 등 다양한 기능 제공(단, 개인정보 탐지관련 패턴은 자체 작성활용 필요)
- **다운로드 :** http://sourceforge.net/projects/pentaho/files/Data%20Integration/5.3/pdi-ce-5.3.0.0-213.zip/download

1단계 : 빅데이터가 포함 된 DB 연동 및 워크 플로우 작성

● 테이블 연결 후 주민번호를 탐색하여 주민번호가 없는 레코드만 추출하여 테이블에 저장하는 과정을 워크플로우로 구성



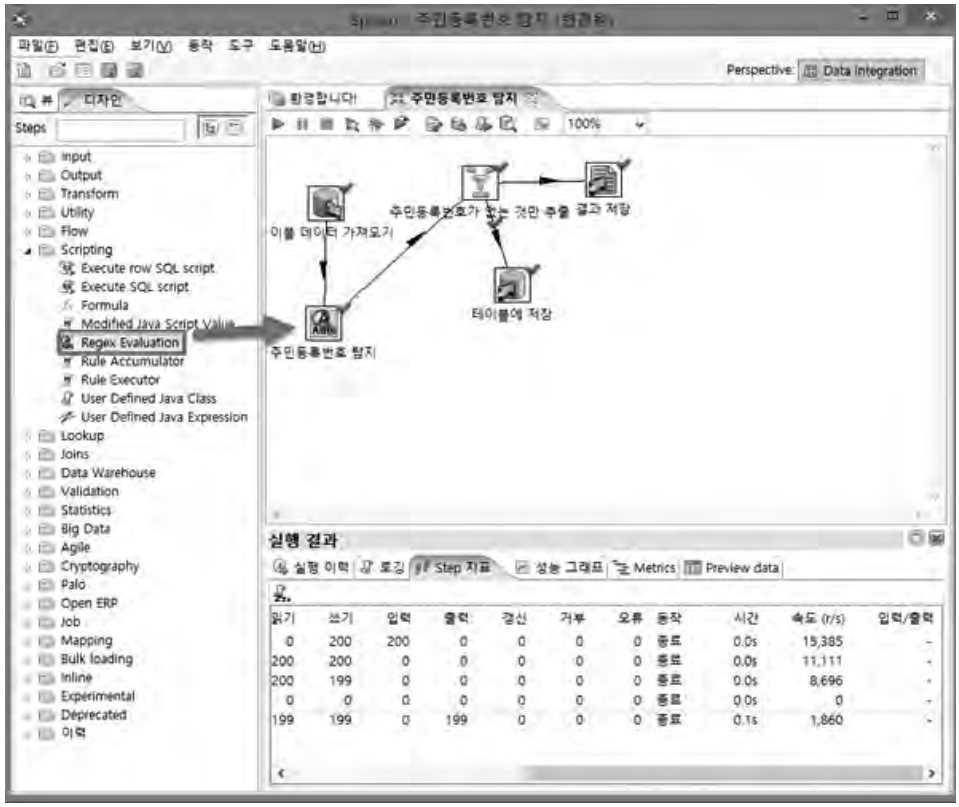
● 데이터가 포함 된 해당 테이블을 설정(연결)

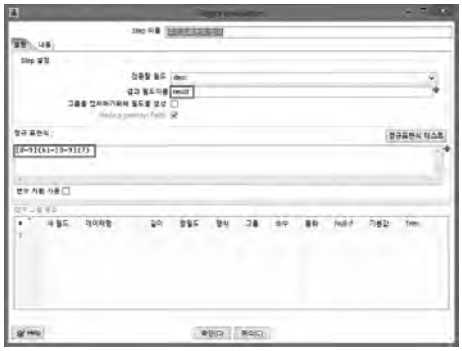


● 테이블에 있는 데이터 가져오기

2단계 : 주민번호 탐지를 위한 정규식 등록

● 주민등록번호가 들어있는 컬럼을 탐지하기 위해서 정규 표현식(Regular Expression)을 사용하므로 이를 위해서 아래와 같이 Regex Evaluation 모듈을 드래그 앤 드롭



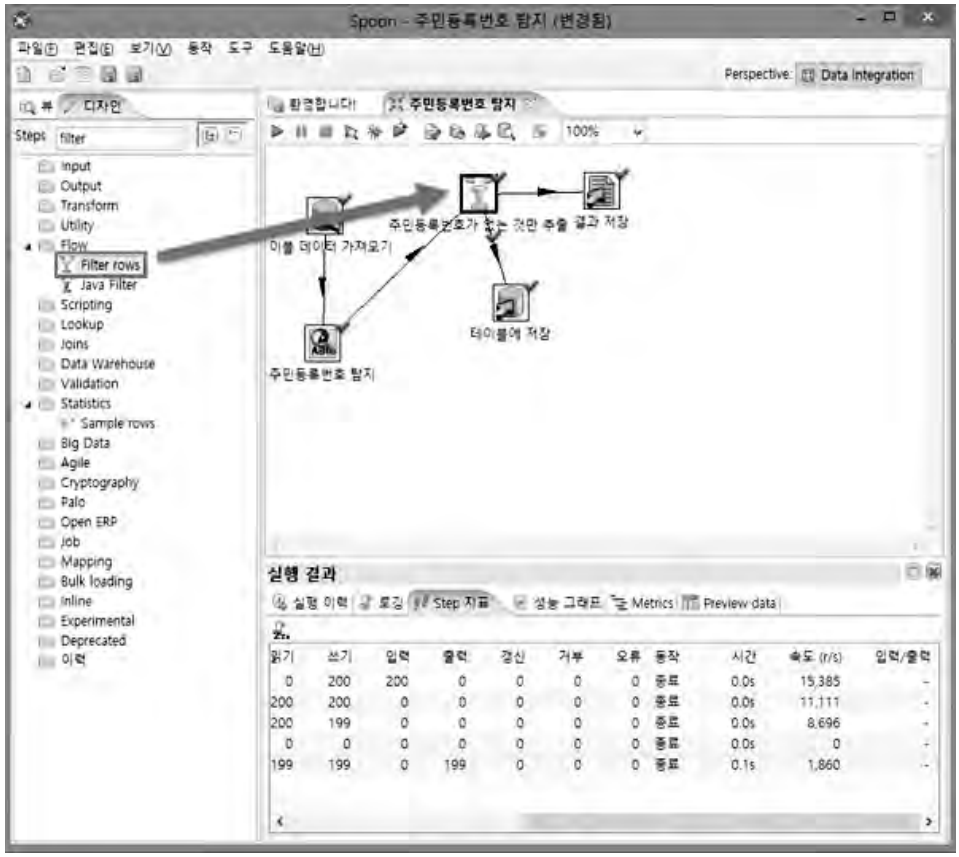


- (이제 주민등록번호를 탐지하기 위한 정규 표현식을 아래와 같이 [0–9]{6}–[0–9]{7} 입력

※ 이것은 0부터 9까지의 숫자가 6개가 순서대로 나오고 다시 7개가 순서대로 나온다는 의미의 패턴입니다. 이렇게 정규 표현식과 일치여부를 result 컬럼에 추가하도록 설정

3단계 : 주민번호 탐지 및 추출

- 주민등록번호가 들어있는 컬럼을 탐지

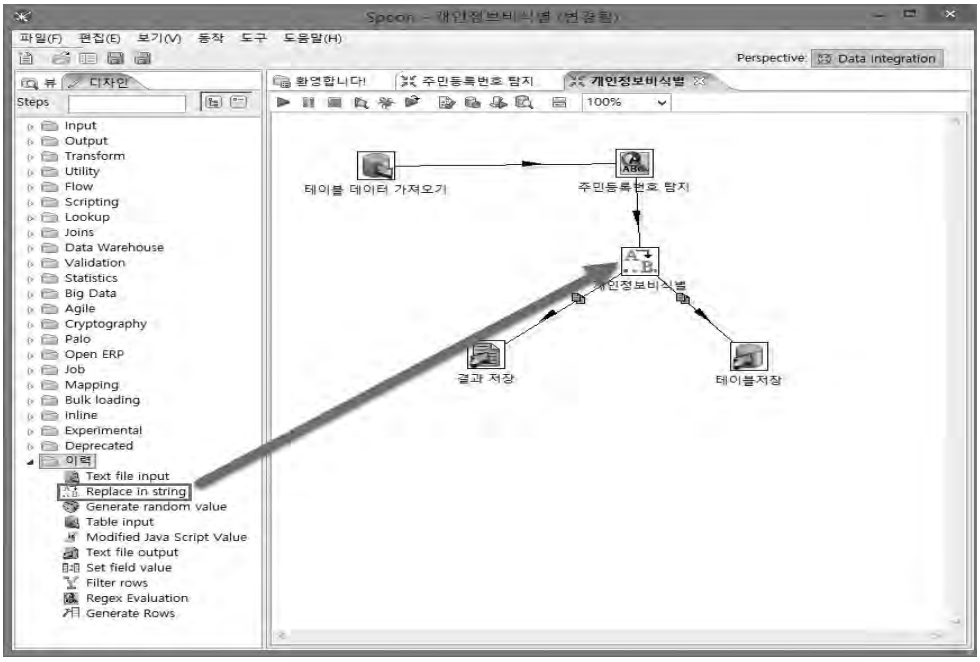


● 만약 주민등록번호라고 판단이 되면 result 컬럼에 Y가 설정 됨

● 주민등록여부를 식별하는 컬럼인 result가 N인것만 필터링할 수 있도록 옵션을 추가

● 주민등록번호를 포함하는 ROW가 삭제된 것을 확인

4단계 : 주민번호 비식별화

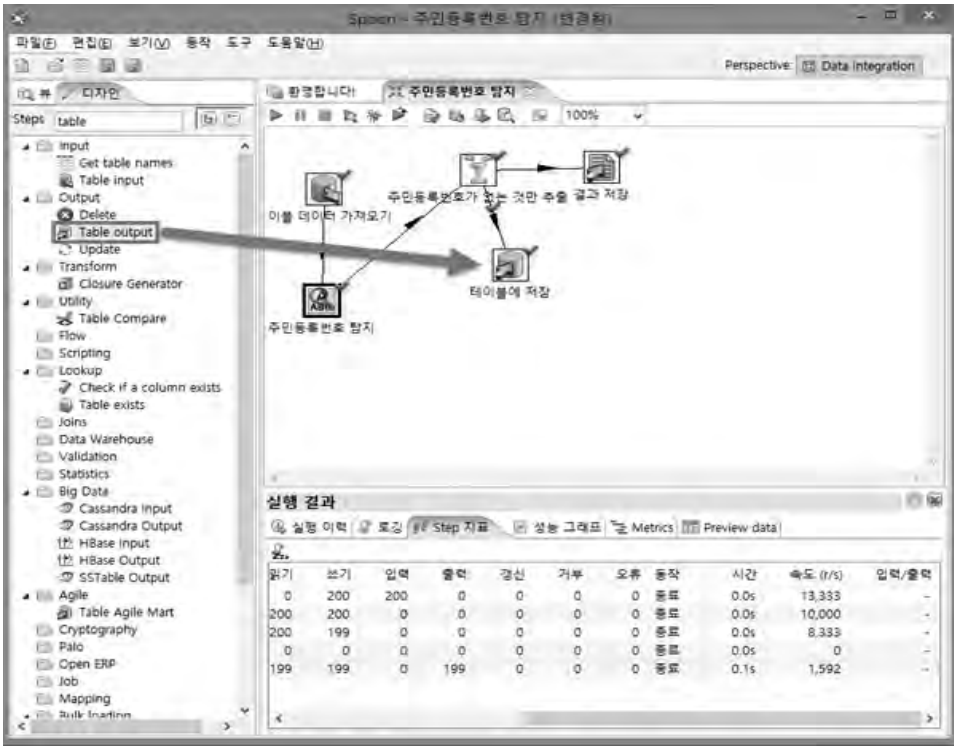


● 정규식으로 주민번호를 임의값 바꿈

● 미리보기 테이블 확인

5단계 : 데이터 저장

● 아래와 같이 테이블에 저장기능을 드래그 앤 드롭



● 필터링된 데이터는 filtered_actor 테이블에 저장

● 미리보기 테이블 확인

빅데이터 활용 Q&A | 부록

빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서(Ver 1.0)

빅데이터 활용 Q&A

● 빅데이터 활용 담당자가 빅데이터 처리 과정에서 고민할 수 있는 사항들을 시나리오 기반 Q&A 구성

[주요 단계 및 Q&A]

주요 단계	시나리오	Q&A
수집 · 이용	정보주체의 동의를 받고 개인정보를 빅데이터로 활용	1. 수집 대상 데이터 목록 선정은? 2. 수집 시 이용자 고지사항은?
	추가 동의후, 개인정보를 빅데이터로 활용	3. 추가 동의 절차는? 4. 정보주체가 14세 미만인 경우 조치? 5. 활용 목적이 변경된 경우 조치?
	개인식별이 가능한 데이터를 비식별화하여 활용	6. 정보 일부가 개인식별이 가능한 경우 조치? 7. 정보 일부가 다른 정보와 결합하여 개인식별이 가능한 경우 조치?
	재식별된 개인정보를 활용	8. 재식별에 대한 조치 방법은?
저장 · 관리	수집된 개인정보에 대한 안전한 관리	9. 주요 관리항목은? 10. 안전 관리를 위한 암호화 적용 방법은? 11. 사후 모니터링 방법은?
	빅데이터 처리 시스템의 안전한 운영	12. 주요 관리항목은? 13. 보안 고려사항은? 14. 저장 · 관리 효율화 방안은? 15. 업무용 PC 관리는? 16. 시스템 담당자 변경 시 조치사항은?
제공 · 위탁	개인정보가 포함된 빅데이터를 제 3자 제공하는 경우	17. 개인식별 정보에 대한 조치? 18. 동의받은 범위 외 데이터를 제공하는 경우 조치? 19. 재식별성의 여지가 있는 경우 조치?
	소유한 빅데이터 분석을 제 3자에게 위탁하는 경우	20. 제3자 위탁 절차는? 21. 개인 식별정보에 대한 처리? 22. 정보통신망 이용 시 보안은?
파기	빅데이터 활용 후 파기	23. 파기 방법은? 24. 사후 관리는?
사후관리	식별정보 노출 시 대응	25. 침해사고 대응단계 및 행동지침은?
		26. 개인정보보호 위반사항에 대한 처벌 규정은?

수집 · 이용

1. (시나리오 1) 정보주체의 동의를 받고 개인정보를 빅데이터로 활용

▶ **Q&A 1 |** 수집 대상 데이터 목록에 대한 선정은?

● 개인정보를 수집하는 경우, **서비스에 꼭 필요한 정보만을 최소화하여 수집하고 수집하는 개인정보에 대해 반드시 동의를 받은 후** 활용

| 예 시 |

● 서비스 제공에 필요한 **최소한의 개인정보 수집을 위한 고려사항**

1) 서비스에 필요한 개인정보 종류 조사

2) 필수 동의 항목 검토 구성

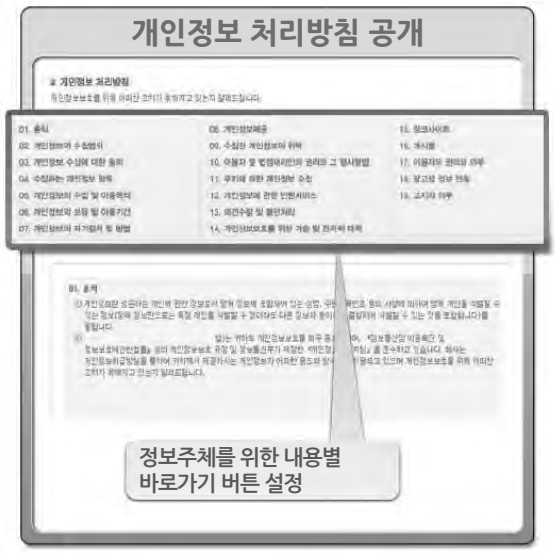
3) 필수 / 선택 동의 사항을 구분하여 수집

현 행	개 선
• 필수 · 선택 기재사항에 대한 명확한 구분없이 정보 수집 • 사업과 직접 관련이 없는 개인정보보호 항목도 일부 포함	• **공통필수 기재사항** 성명, 주소, 연락처 등 5개 항목 • **연결필수 기재사항** 계약 전 알릴 의무 사항 등 5개 항목 • **선택사항**
평균 약 30개 항목	총 10여개 항목

▶ **Q&A 2 |** 개인정보를 수집하는 경우 이용자 고지사항은?

● 정보주체가 자신의 어떠한 정보가 수집되어 활용되는지 명확하게 인지할 수 있도록 **담당자가 상세하게 고지**

[개인정보 처리방침 공개 예]



출처 : PIMS 구축 및 운영교육, KISA, 2011. 10.

2. (시나리오 2) 추가동의 후, 개인정보를 빅데이터로 활용

▶ **Q&A 3 |** 개인정보(공개, 이용내역 정보) 추가 동의 절차는?

● 개인식별정보의 수집 · 이용으로 **정보주체 추가동의**가 필요한 경우 동의를 획득하고 빅데이터로 활용

– 정보주체 및 대리인에게 동의를 받는 절차

1) **개인정보를 제공받은 단계에서 식별된 정보가 포함**되어 있을 때 정보주체의 추가동의를 받은 후 활용

2) 제공받은 정보를 분석 및 가공하는 단계에서 **기존 정보와 결합하여 다시 식별성을 갖게 되었을 때** 정보주체의 추가동의를 받은 후 활용

3) 가입, 등록 등 수집단계에서 **정보주체에게 사전 고지를 통해** 식별 또는 재식별될 수 있는 빅데이터 활용 사항을 구체적으로 알려주고 정보주체에 추가동의를 받은 후 활용

※ 해당 정보가 수집, 조합, 분석 등 정보주체가 쉽게 확인할 수 있도록 공개

▶ Q&A 4 | 정보주체가 14세 미만일 경우 조치사항은?

- 법정 대리인에게 동의 받은 후 빅데이터로 활용

| 예 시 |

- 가입약관 및 서비스 이용규칙을 통해 만14세 미만의 어린이들은 법정대리인의 동의가 필요함을 설명하고, 가입시 부모님 동의 여부를 확인

[만 14세 미만 아동의 정보 수집 온라인 / 오프라인 예]

□ 만 14세 미만 아동의 개인정보를 수집 하기 전에, 부모의 본인확인 후 개인정보 수집이 진행될 수 있도록 개발

※ 14세 미만 가입 버튼과 법정대리인의 본인 확인을 위한 i-Pin, 핸드폰 인증 등

□ 만위와 같이 개인정보를 수집 · 이용 하는데 동의하십니까?

법정대리인의

성명 :_____ **연락처 :**_____

※ 14세 미만 이동인 경우 법정대리인이 서명 하여 주시기 바랍니다.

출처 : 개인정보보호 종합포털, 2013.12

▶ Q&A 5 | 빅데이터 수집 · 이용 목적이 변경된 경우 조치사항은?

- **빅데이터에서 필요한 정보 이외**에 수집하거나 **빅데이터 활용 목적이 변경**된 경우에는, 정보주체에게 추가동의를 획득하고 빅데이터로 활용

| 예 시 |

- 개인임을 알 수 없는 SNS 글이나 인터넷 쿠키 같은 단순한 이용 내역 정보는 개인 동의없이 이용하되, 빅데이터를 활용하고자 하는 기업은 해당 정보를 제공받는 자, 이용목적, 제공하는 정보의 항목, 해당정보를 제공받는 자의 보유 및 이용 기간 등을 정보주체에게 고지

3. (시나리오 3) 개인식별이 가능한 데이터를 비식별화하여 활용

▶ Q&A 6 | 수집한 정보가 개인 식별이 가능한 경우 조치사항은?

- 개인 식별 정보는 재동의 받거나 **삭제**하고 활용할 수 있도록 함
- 주민등록번호, 운전면허번호 등 고유식별정보가 포함된 내용에서 개인정보를 수집하는

경우에는 **고유식별정보를 삭제하거나 데이터 마스킹(Data Masking)**을 통해 명시한 후 빅데이터 활용

- 불가피하게 수집하는 식별정보에 대해서는 안전하게 **암호화***하여 처리하거나 **토큰화(Tokenization)****하여 빅데이터에 활용

(*) 수집되는 개인정보를 안전하게 보호하기 위해 적합한 암호화 기술 (알고리즘, 키길이 등)을 선택

(**) 데이터를 토큰(Token)으로 치환한 뒤에 원본 데이터를 대신하는 기술

| 예 시 | 토큰 예시

- 이동통신사 가입 시 요금할인 및 결제를 위해 불가피하게 수집하는 카드번호 등 개인식별정보는 토큰화 방식을 통해 다른 데이터로 처리하여 저장

– 카드번호는 원본데이터에 대해 전체(또는 부분)를 Token 암호화하여 적용

원본 데이터	Token	Token 활용	설명
4567–7894–4567–7895	RANDOM	4567RAMDON7895	원본데이터의 앞뒤를 보존하여 부분 치환

▶ Q&A 7 | 다른 정보와 결합하여 개인 식별이 가능한 경우 조치사항은?

- 정보가 가공을 통해 기존 정보와 결합되어 개인을 알아볼 수 있는 정보로 재식별 되어 생성되었다면 기술적 분리 가능여부를 확인하여 식별된 정보를 삭제 처리한 후, 빅데이터에 활용할 수 있도록 함

- 결합되어 생성된 개인 식별정보가 필요한 정보일 경우,

1) **데이터 마스킹 등 비식별화 처리 기법을 통해 식별**이 불가능하도록 하여 빅데이터로 활용

2) 개인 식별정보 활용에 대해 **정보주체** 추가동의 획득

- 단, 정보주체에게 사전 동의를 받은 경우 비식별화 없이 활용 가능

| 예 시 |

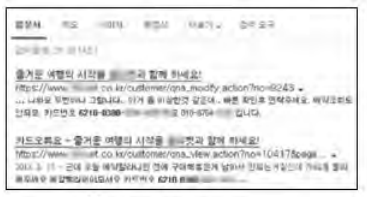
- 이동통신사 가입 시 **주요 개인정보는 데이터 마스킹 기술을 통해 수집**하며, 내부 담당자에 의해 고객정보 조회 시 주민등록번호와 같은 개인정보를 뒷자리만 '*'으로 치환하는 마스킹 기능으로 표시제한 조치를 취하도록 함

[데이터 마스킹 적용 예]

781120 – 1624633 홍길동 ➡ 781120 – ****** 홍길동
781120 – 1 – 60 – 46 – 33 홍길동 ➡ 781120 – 1624633 홍길동

● SNS 등 비정형 데이터는 데이터 마스킹으로 식별 위험이 있는 개인정보를 비식별화 조치하여 담당자가 정보를 식별할 수 없도록 한 후 활용

[크롤링을 통해 외부에서 수집한 정보의 비식별화 조치 예]



노출된 정보 크롤링

DBMS

담당자 조회 화면

이름	주민등록번호	전화번호	주소
KIM	700324 – *******	010 – 1234 – 5678	서울시 ***** ******* ****
LEE	751212 – *******	010 – 9876 – 5432 010 – 9876 – 5432	수원시 ***** ******* ****

비식별화 조치

4. (시나리오 4) 재식별된 개인정보를 빅데이터로 활용

▶ **Q&A 8 |** 비식별화 조치 후에도 다른 정보와 결합하여 재식별되는 경우 조치사항은?

● **재식별된 정보는 다시 비식별화 처리를 수행**하여 빅데이터에 활용

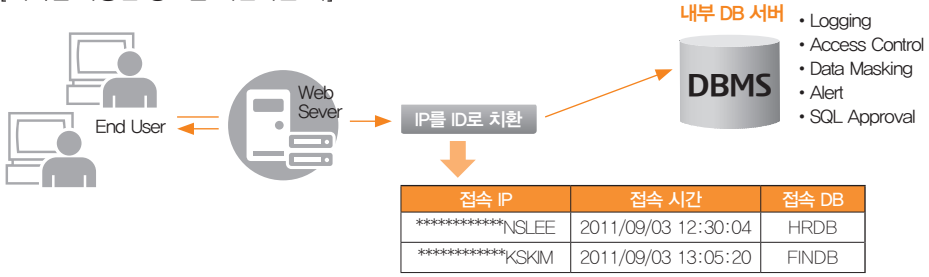
– 재식별된 정보를 이용하는 경우 정보주체의 추가동의 획득

| 예 시 |

● 웹정보, SNS에서 크롤링(Crawling) 등 자동화된 시스템을 통해 데이터를 수집할 수 있다. 이때 빅데이터 활용으로 가공 및 분석을 통해 식별가능한 정보가 파생 또는 생성되었다면 담당자는 정보주체에게 별도동의(또는 사전동의)를 획득한 후 정보를 활용

● 기존 정보와 결합 시 개인정보 침해가 가능한 IP 정보는 비식별화 조치가 이루어질 수 있도록 하며, 담당자는 고의 또는 실수로 생성된 개인 식별정보에 대해서 즉시 파기하거나, 비식별화 조치를 취하도록 지속적으로 확인

[재식별 가능한 정보를 치환하는 예]



접속 IP	접속 시간	접속 DB
**********NSLEE	2011/09/03 12:30:04	HRDB
**********KSKIM	2011/09/03 13:05:20	FINDB

접속자 정보를 식별 불가능한 정보로 치환하여 저장

5. (시나리오 5) 수집된 개인정보 데이터를 안전하게 관리

▶ Q&A 9 | 수집된 빅데이터를 직접 관리하는 경우 조치항목은?

- 수집된 데이터 유형에 따라 **저장 계획을 수립**하고 적절한 시스템을 구축한 후 데이터 저장 · 관리
- 빅데이터의 안전한 활용을 위해 저장된 정보가 고의 및 실수로 외부에 유출되지 않도록 **데이터 접근 제어와 보안 감사를 수행**
- 빅데이터 환경 내에서 발생할 수 있는 피해 · 사고를 최소화하기 위해 업무 담당자들의 직무 분리 및 접근권한을 구분하는 등 **빅데이터의 안전한 활용을 위한 내부운영관리정책 마련**

| 예 시 |

- 입력 · 승인 · 확인 등 모든 거래과정이 동일 직원에 의해 처리되지 않도록 분리
- 프로그램 개발자와 시스템 운영자 분리
- 데이터베이스 관리자와 시스템 및 운용 프로그래머 분리
- 시스템 프로그래머의 작업에 대한 자체 감사자의 통제
- 오퍼레이터에 대한 자체 감사자 또는 시스템 프로그래머의 통제
- Batch 프로그램에 의한 온라인 데이터베이스 접근 통제
- 업무개발자와 관리자의 분리
- 내부감사자 및 보안 관리자의 타 업무 겸임 금지

| 예 시 |

- 개인정보처리시스템에 보관된 개인정보는 **지정된 관련 업무 담당자만 열람**할 수 있도록 하고, 영업부서의 접근을 제한하여 **외부 영업목적(텔레마케팅 등)으로 이용할 수 없도록 제한**하여야 하며, **필요에 따라 최소한의 정보에만 접근**할 수 있도록 하여야 한다.

▶ Q&A 10 | 빅데이터의 안전한 관리 측면에서 암호화 알고리즘 적용이 필요한 경우 조치사항은?

- 적용하는 암호화 방식이 **시스템 성능 저하를 최소화**할 수 있도록 빅데이터의 특성, 제약사항 등을 고려하여 암호화 방식을 선택
- 빅데이터 활용 측면에서 안전한 암호화 알고리즘 적용방법은,

1) 개인 식별정보 저장 시 상용 암호화 프로그램 또는 국내외 전문기관(KISA, NIST, ECRYPT, CRYPTREC 등)을 중심으로 제시하고 있는 안전한 암호화 알고리즘을 적용

2) 개인정보가 포함된 일반 데이터 저장 시 문서 프로그램 등에서 제공하는 암호화 적용

| 참 고 |

[국내외 전문기관 권고 암호화 기술]

암호기술 / 권고기관	Token	Token 활용	설명
NIST (미국)	2TDEA, 3TDEA, AES	DH, RSA, MQV, DSA, ECDH, ECDSA, ECMQV	SHA-1/224/256/384/512
ECRYPT (유럽)	AES, 2TDEA, 3TDEA, KASUMI 등	RSA-PKCS#1, DSA, RSA-PSS/ OAEP/KEM, ECDSA 등	RIPEMD-128/160 SHA-1/224/256/384/512, Whirlpool
CRYPTREC (일본)	AES, Camellia, Cipherunicom-A, SC2000 등	DH, ECDH, RSA-OAFP, RSAES-PKCS#1, DSA, ECDSA 등	RIPEMD-160, SHA-1/256/384/512
암호검증기준 (국내)	SEED, ARIA, HIGHT	ACE-KEM, PSEC-KEM, RSA-KEM, RSAES-OAEP, RSASSA-PSS, RSASSA-PKCS1(V1.5), DSA, ECDSA	HAS-160, SHA-1/224/256/384/512

출처 : KISA, 암호정책수립기준안내서, 2013.12.

▶ Q&A 11 | 보유하고 있는 빅데이터를 안전하게 활용하고 있는지에 대한 사후 모니터링 방법은?

● **내부규정, 가이드라인 등에서 제시한 빅데이터 활용 보안기준을 따르지 않아 발생하는 취약점**으로 인해 개인정보 DB가 노출될 수 있으므로 지속적으로 취약점을 점검 및 조치

| 참 고 | 취약점 진단 시 참고 자료

참고자료	웹 페이지 주소
OWASP Top 10	https://www.owasp.org/index.php/Top_10_2013-Top_10
SANS Top 25	http://www.sans.org/top25-software-errors/
주요정보통신기반시설 기술적 취약점 분석평가 방법 가이드	http://www.mospa.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000012&nttId=41297
진단 · 제거 가이드	http://www.kisa.or.kr/public/laws/laws3.jsp

6. (시나리오 6) 빅데이터 처리 시스템의 안전한 운영

▶ Q&A 12 | 빅데이터 정보시스템 운영을 위한 기본 조치사항은?

- **빅데이터 처리 시스템의 안정성 확보를 위해** 개인정보 보관시 기술적 · 관리적 · 물리적 조치 마련으로 개인정보를 취급하는 주요 시스템에 **데이터베이스 접근통제장치, 서버접근제어 등 설치 · 운영**
- **네트워크 구성**은 내 · 외부망을 물리적으로 분리하여 업무망과 인터넷망으로 구분하며 방화벽, 침입방지시스템(IPS: Intrusion Prevention System), ACL(Access Control List)등 구축하여 네트워크 접근통제

▶ Q&A 13 | 빅데이터 처리 시스템 운영 시 보안 관련 조치사항은?

- 빅데이터가 이용되는 정보통신망에서 **불법적인 접근을 탐지하는 관제활동**이 적절히 이루어질 수 있도록 함
- **암호화된 데이터베이스 접속 프로그램**을 통해 빅데이터가 저장된 정보에 접근함으로써 보호대상을 안전하게 보호하고, 필요에 의해 데이터의 목적, 내용에 따라 데이터베이스(또는 테이블)를 분리하여 설치 · 운영
- 안전하게 빅데이터 보관하기 위하여 개인정보를 취급하는 서버* 대상으로 접근제어를 적용 (*) 서버접근제어는 System(Server) Access Control, SecureOS 등을 포함하며 서버에 접근하는 담당자(또는 이용자)를 관리

| 예 시 |

- 고객정보를 보유하고 있는 데이터베이스 서버에 **서버보안 소프트웨어를 설치**하여 네트워크상 우회접근을 차단하고, 파일접근, 수행명령어 등 모두 감사로그 기록을 하여 사고 발생 시 추적이 가능하도록 한다.
- 고객정보를 취급하는 담당자는 별도의 비밀번호를 부여하고 정기적으로 갱신함으로써 안전성 확보 조치를 준수한다.

[패스워드 관리대장 작성 예]

(5)월 비밀번호 관리대장

결제	보안 담당자	정보보호관리자
	* * *	* * *

번호	시스템명	계정	비밀번호	변경일자	정보시스템운영담당자	비고
01	NCC	administrator	*********	2007-05-04	* * *	완료

출처 : 위치정보의 관리적 · 기술적 보호조치 권고 해설서, 2011.9.

- 수집된 빅데이터의 안전한 관리 및 운영을 위해 **담당자가 시스템에 접속하여 수행한 업무 내역**에 대하여 식별자, 접속일시, 접속자를 알 수 있는 정보, 수행업무 등 **접속한 사실을 기록하여 관리**

▶ Q&A 14 | 빅데이터 저장 · 관리의 효율성 제공 방안은?

- 빅데이터의 분류와 저장은 ROI 관점에서 중요한 이슈이므로, **데이터를 비용대비 효과적으로 저장 및 관리할 수 있는 체계 구축**

[빅데이터 저장 방식]

구현방식	설명
분산 파일 시스템	컴퓨터 네트워크를 통해 공유하는 여러 호스트 컴퓨터의 파일에 접근할 수 있게 하는 파일 시스템
NoSQL	데이터 모델을 단순화해서 분산의 기본 개념을 쉽게 정의하고, ACID 요건을 완화하거나 제약하는 형태의 새로운 저장 시스템을 통칭
병렬 DBMS	다수의 마이크로프로세서를 사용하여 여러 디스크에 대한 질의, 갱신, 입출력 등의 데이터베이스 처리를 동시에 수행하는 데이터베이스 시스템
네트워크 구성 저장 시스템	서로 다른 종류의 데이터 저장장치를 하나의 데이터 서버에 연결하여 총괄적으로 데이터를 저장, 관리
클라우드 파일 저장 시스템	클라우드 컴퓨팅 환경에서 가상화 기술을 활용한 분산 파일 시스템

▶ Q&A 15 | 업무용 PC에 대해 보안관리가 필요한 경우 조치사항은?

- **빅개인식별정보가 포함된 빅데이터 처리시스템 또는 업무용 PC**에 불법소프트웨어 설치를 정기적으로 확인하며, P2P 등과 같이 정보가 외부로 유출될 수 있는 프로그램을 사용할 수 없도록 보안설정 확인
- 공유 폴더를 사용할 경우 **드라이브 전체 또는 불필요한 폴더가 공유되지 않도록 조치**하고, 공유폴더에 개인정보 파일이 포함되지 않도록 정기적으로 점검
- 또한, 빅데이터 개인 식별정보를 관리하는 업무용 PC의 로그인 비밀번호는 **안전한 비밀번호 작성 규칙을 활용**하여 이용 하도록 하고, 안전한 사용을 위해 운영체제 및 **백신 · 보안 프로그램은 정기적으로 업데이트**

▶ Q&A 16 | 빅데이터 운영을 위해 담당자 변경 시 조치사항은?

- 담당자의 인사이동 또는 휴직, 퇴사 발생 시 담당자가 사용한 PC 내 활용 된 빅데이터 개인식별정보가 오 · 남용 되지 않도록 **개인정보보호 보안 프로그램 등을 도입하여 관리**
- 개인식별정보를 취급하던 담당자의 추가, 변경, 휴직 및 퇴사 발생 시 **담당자의 권한을 변경(또는 말소)하고 접근권한 관리 대장에 기록**

  1) 이때, 부여받은 접근권한 계정은 다른 담당자와 공유하지 않도록 주의하며 안전한 빅데이터 운영 마련

  2) 또한 정당한 권한이 없이 또는 허용된 권한을 초과하여 개인정보를 훼손 · 멸실 · 위조 · 변경 또는 유출하지 않도록 주의

| 예 시 | 빅데이터 처리 담당자 접근권한 관리대장 예

사용자명	부서명	개인정보 업무현황	개인정보 접근권한	변경사유	변경사항

7. (시나리오 7) 개인정보가 포함된 빅데이터를 제3자에게 제공

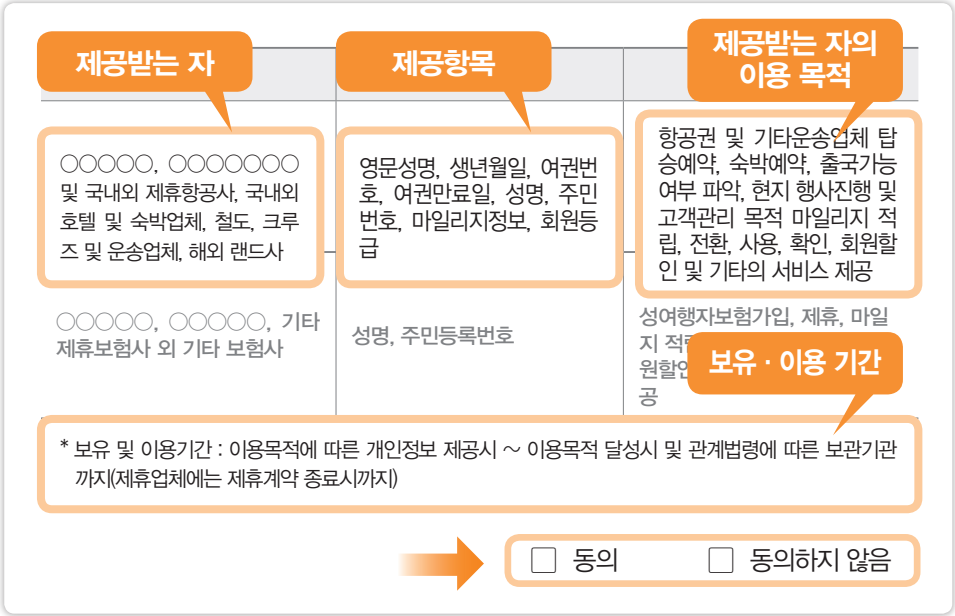
▶ **Q&A 17 |** 빅데이터 활용을 위해 제공되는 정보 내 개인 식별정보를 포함하고 있는 경우 조치사항은?

- 개인정보를 빅데이터 활용 목적으로 제3자에게 제공하는 경우 이용약관 내 제공받는 자, 제공목적, 보유기간 등을 명시하고 정보주체에게 동의 획득
- 정보주체가 빅데이터 활용을 위한 제 3자에의 **개인정보 제공을 거부하는 경우에도** 서비스 이용이 제한되지 않도록 처리 함
- 단, 미동의로 인한 빅데이터 활용 서비스가 어려울 경우 **정보주체에게 사전 동의를 받은 후 활용**

| 예 시 |

- 빅데이터 활용 목적의 개인 식별정보 제공 고지 예

  • **[고지사항]** 개인정보 제공 시, 개인정보를 제공받는 자, 제공받는 자의 이용 목적, 개인정보 항목, 보유 및 이용 기간 등에 대해 별도 동의



출처 : 개인정보 위반 사례와 현장 점검결과, 안전행정부, 한국정보화진흥원, 2014.10

▶ **Q&A 18 |** 본래의 목적 범위 외 개인정보가 포함된 빅데이터를 제공해야 하는 경우 조치사항은?

- 빅데이터 활용을 위해 이용약관에 명시 또는 고지하고 동의를 받은 개인정보 범위를 넘어 제3자에게 제공 및 위탁하지 않도록 수행 함
- **동의 받은 범위 외 빅데이터 활용 목적으로 개인정보를 활용**해야 할 경우, 이에 대한 **상세내용을 고지하고 별도의 추가 동의 획득 후 빅데이터로 활용**

| 예 시 | 동의를 얻어야 하는 경우 예

- 빅데이터 개인 식별정보 활용 기반 제공목적 외 새로운 상품소개, 서비스 혜택, 이벤트 할인행사 등의 홍보를 해야 하는 경우
- 정당한 정보 제공자가 아닌 자에게 빅데이터 활용 개인 식별정보를 열람 및 이용을 허용하는 경우
- 전략적으로 상품판매 및 개발을 위해 고객의 관심분야 또는 주문정보를 빅데이터에 활용해야 하는 경우

▶ **Q&A 19 |** 식별 정보가 자동 추출되거나 비식별화 된 정보와 결합되어 특정 개인으로 식별 될 가능성이 있는 경우 조치사항은?

- 고도화된 분석 기술을 통해 개인정보가 자동 추출되어 훼손 · 침해 · 무단 제공 등 **발생할 수 있는 사고에 대비하여 빅데이터 활용 정책 및 적용기술 등 안전 조치 마련하여 활용**
- 제공되는 데이터가 개인 식별이 가능한 경우 **기술적 분리 가능여부를 확인하여 식별된 정보를 삭제 처리한 후, 부분제공으로 빅데이터에 활용**할 수 있도록 함
- 빅데이터 활용 목적으로 제공하는 개인정보 처리 중 **비식별화 된 정보와 결합되어 특정 개인으로 식별 될 가능성이 있는 경우** 해당 정보는 **비식별화 처리 기법을 적용** 한 후 활용하도록 함

8. (시나리오 8) 소유한 빅데이터의 분석을 제3자에 위탁

▶ **Q&A 20 |** 빅데이터 활용 단계 중 업무 위탁이 발생하는 경우 절차는?

- 동의 받은 범위 내 빅데이터로 **활용 시 필요한 최소한의 정보**만을 **위탁**하도록 수행
- **빅데이터 활용 목적으로** 개인정보 제공 · 위탁 시 **개인정보 관리대장**을 통해 기록 · 보존하고 **이용 목적, 보유기간, 담당자 정보** 등의 상세내역을 명확히 **기록 후 위탁하도록 함**

| 예 시 | 빅데이터 내 개인 식별정보 제공 관리 대장 예

개인 식별정보의 제공 관리 대장			
제공 목적	[ ] 위탁 제공 [ ] 제 3자 제공 [ ] 양도 · 양수 [ ] 합병		
제공 기관명		처리 담당자	소속
개인정보 항목			성명
이용일시			전화번호
제공방법			
보유기간			

- 빅데이터 활용 목적으로 개인정보 위탁 시 개인정보 보호조치를 위한 **구체적인 고려사항을 계약서로 작성 후 보존**

| 예 시 | 안전한 빅데이터 활용을 위한 계약서 포함내용 예

- 빅데이터 활용을 위해 개인 식별정보 취급 시 보관, 처리, 관리, 파기의 전(全) 과정에서 발생할 수 있는 고려사항 작성
- 기술적 · 관리적 보호조치 의무, 개인 식별정보에 관한 비밀유지의무, 빅데이터 목적 외 이용 제한, 개인정보침해로 인한 손해배상 책임 등의 사항 및 절차 규정 포함

- 정보주체의 개인정보가 분실 · 도난 · 유출 · 변조 또는 훼손되지 아니하도록 **수탁자를 정기적으로 교육**하고, **수탁자가** 관련 법규에 따른 **의무 · 이행사항 준수 여부에 대해 지속적으로 관리 감독**하여 **위반에 대한 점검**

▶ Q&A 21 | 개인 식별정보가 포함된 빅데이터를 제공하는 경우 조치사항은?

- 서비스 제공에 대한 계약 이행의 목적으로 필요한 정보인 경우 **정보주체에게 고지 및 별도 동의 획득 후 빅데이터에 활용**하도록 함
- 개인 식별정보의 처리를 위탁받은 처리자(수탁자)는 **제공받은 개인정보의 활용 사실**과 **목적을 정보주체가 쉽게 확인**할 수 있도록 사이트 게재, 이메일 고지 등을 통해 알림
- 계약 이행 목적으로 **분석 및 개인 식별정보가 포함 된 빅데이터 제공 시**, 제공 항목 및 용도를 **기록 후 활용**

| 예 시 | 빅데이터 내 식별정보 제공 시 관리 대장 작성 예

접속ID	처리자명	자료종류	처리일	분석항목/식별정보	처리 목적	파기(예정)일	파기 책임자

▶ Q&A 22 | 빅데이터 제공 시 정보통신망을 이용하는 경우?

- 개인 식별정보가 포함된 빅데이터를 전송하는 구간에서는 **SSL/TLS 등의 암호화 방식을 적용**하여 안전하게 전송
- 빅데이터 처리시스템 간 **안전한 개인정보 파일 전송은 VPN 기술을 활용**하여 수행하도록 함

| 예 시 | 윈도우에서 IPsec VPN를 통한 암호화 설정 방법

- Windows 7의 제어판 메뉴에서 [윈도우 방화벽] → [고급설정] → [로컬 컴퓨터 고급 보안이 포함된 [윈도우 방화벽] → [속성] → [IPsec 설정] → [사용자 지정]을 선택
- [IPsec 설정 사용자 지정]과 같은 대화창이 나타나면 [키교환] → [사용자 지정]을 선택하여 [고급 키 교환 설정 사용자 지정]에서 IPsec VPN 방식에 사용할 암호 알고리즘을 변경

파기

9. (시나리오 9) 빅데이터 활용 후 파기

▶ **Q&A 23 |** 빅데이터 활용의 처리 목적이 달성 된 후 파기방법은?

- 빅데이터 활용 목적을 달성한 경우, 사업을 폐업하는 경우, 이용자가 동의를 철회한 경우에는 **지체없이(5일 이내) 개인정보를 복구 · 재생할 수 없는 방법으로 파기**
- **개인정보의 일부만을 파기하는 경우**에도 개인정보를 삭제한 후 복구 및 재생되지 않도록 관리 및 감독
- **종이문서철에 포함된 개인정보**는 각각의 개인정보를 선택적으로 파기하는 것이 현저히 곤란하므로, **분기 또는 반기 단위로 별도의 점검 및 파기절차를 마련**한 후 시행

▶ **Q&A 24 |** 이용 및 보유 기간이 끝난 경우 파기?

- 개인정보 이용 · 제공목적이 달성된 경우 또는 이용약관에 명시된 이용 · 제공 보유기간이 만료된 경우 기술적 방법을 사용하여 즉시 파기 후 **기록 관리**

| 예 시 | 개인정보 파기 관리 대장 예

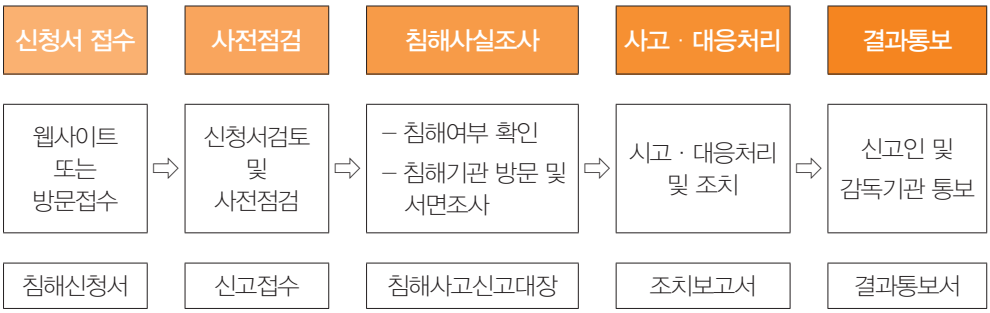
기관명	개인정보 파일명	자료종류	생성일	파기일	파기사유	처리담당자	담당자 연락처

사후관리

10. (시나리오 10) 빅데이터 개인정보 식별에 대한 사후관리

▶ **Q&A 25 |** 개인식별 정보 노출 시 대응절차는?

[식별정보 노출발생 대응절차 예시]



① 신청서 접수

– 긴급한 사항인 경우, 전화를 이용하여 접수

– 일반적인 경우, 전자우편 또는 팩스를 이용하여 접수

– 이때, 접수자를 확인할 수 있도록 이름, 전화번호, 이메일주소 전달

② 사전 점검

– 침해사고에 대해 수사기관에 신고할 것인지를 사전점검을 통해 수행하고 심각정도를 고려하여 침해 신고 접수

③ 침해 사실 조사

– 침해 위협에 대한 신속한 정보 수집 및 분석

④ 사고 · 대응 처리

– 수집된 정보를 바탕으로 구체적으로 위협내용 및 위협수준 분석

– 보고된 자료에 대한 정확한 판단이 어려운 경우 추가적인 정보 획득 및 초기 분석 재실시 (필요한 경우 외부기관을 통해 침해위협 진단)

– 긴급시 "긴급대책본부"구성하여 사고 처리 실시

– 침해사고 범위에 정의된 사고를 중심으로 대응조치(보안조치는 언제 할것인지, 복구조치는 어떻게 수행할 것인지 등)결정

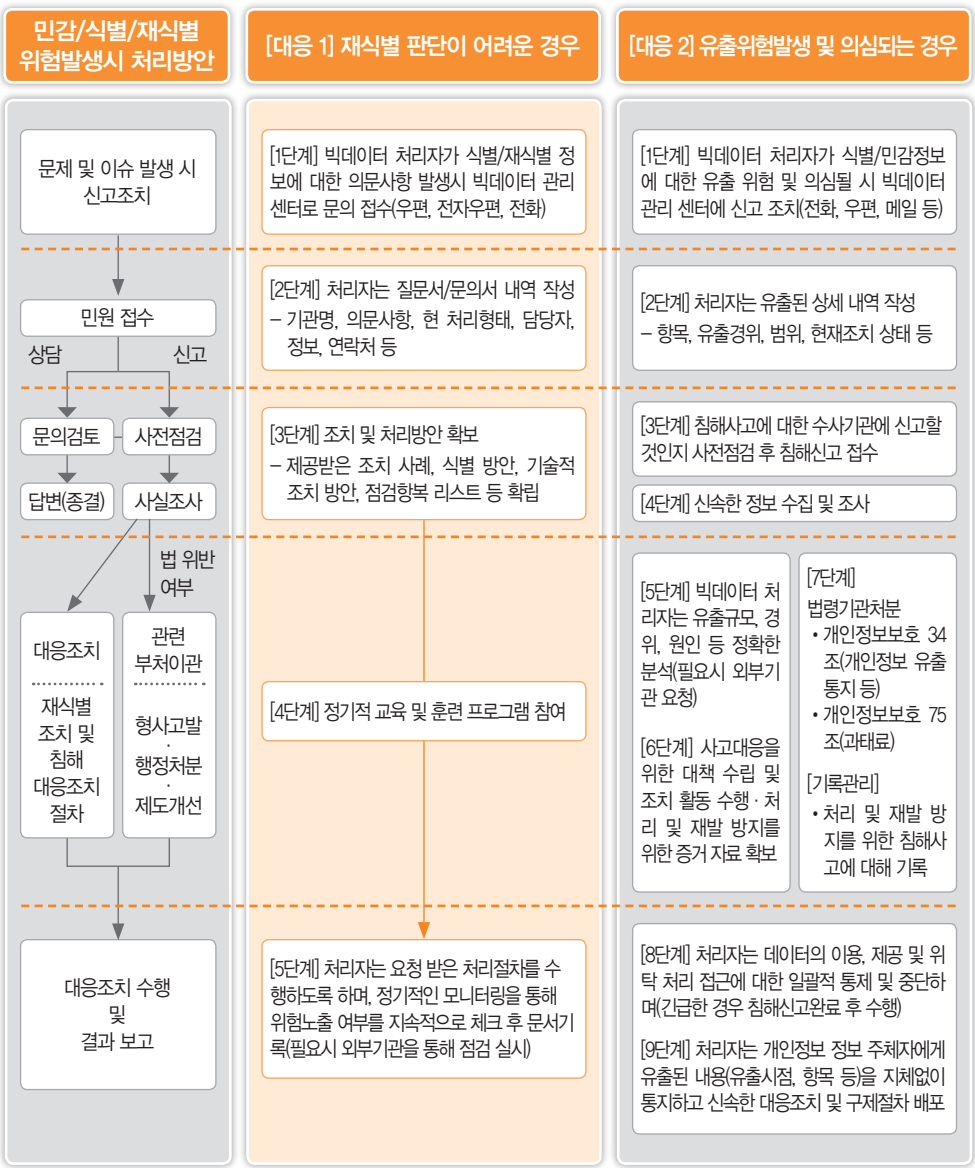
– 현행 법률 기반 법적대응 조치 수행

⑤ 결과 통보

– 빅데이터 처리자는 데이터 이용 · 제공 · 위탁 처리되고 있는 모든 식별정보에 대한 일괄적 통제 수행

– 개인정보 주체자에게 유출된 내용(유출시점, 항목, 대처방안 등)을 통지하고 신속한 대응조치가 가능한 구제절차 배포

⑥ 재식별정보 의심 및 유출 위험 발생 시 대응절차



참고 : 정보주체의 권리보장

| 이용자의 권리 보호 |

● 책임자는 정보주체가 자신의 개인정보에 대한 열람, 정정 · 삭제, 처리정지를 요구할 경우 이에 대한 대응조치가 이루어질 수 있도록 함

[참 고]

- 개인정보 열람 요구 대응조치(「개인정보보호법 시행령」 제41조)
  - 내용적으로 10일 이내에 열람할 수 있도록 조치,
  - 열람할 수 없는 정당한 사유가 있을 경우 정보주체에게 그 사유를 알리고 열람 연기(사유가 소멸하면 지체없이 열람)
- 개인정보 정정 · 삭제, 처리정지 요구 대응 조치(「개인정보보호법 시행령」 제43조)
  - 10일 이내에 조치, 단 열람 거부사유가 있는 경우 거부 통지 가능

※ 위반 시 3천만원 이하의 과태료 부과 (「개인정보보호법」)

● 개인정보 사고 시 처리 및 대응 절차 수립

– 담당자는 개인정보 유출시 지체없이(5일 이내) 정보주체에게 유출사실 통지

[참 고]

- 개인정보 유출 시 통지 사항(「개인정보보호법」 제34조)

※ 통지 · 신고 미이행시 3천만원 이하 과태료(「개인정보보호법」)

- 유출된 개인정보 항목
- 유출 시점 및 경위
- 유출 피해 최소화를 위해 정보주체가 할 수 있는 방법
- 개인정보처리자 대응조치 및 피해구제 절차
- 신고접수 담당부서 및 연락처

● 책임자는 빅데이터에 포함된 개인정보의 유 · 노출, 변경, 삭제 시 보고 및 대응 절차, 사고 대응 조직의 구성을 포함한 개인정보사고 처리 및 대응 체계를 수립 · 시행하여야 함

▶ Q&A 26 | 개인정보보호 위반 사항에 따른 처벌 규정은?

벌칙	행위내용	금지조항	벌칙조항
10년 이하 징역 또는 1억원 이하 벌금	✓ 개인정보 변경 등 업무 방해		제70조
5년이하 징역 또는 5천만원 이하 벌금	✓ 정보주체의 동의를 받지 않은 개인정보를 제공한 자와 제공 받은 자	제17조 제1항	제71조 제1호
	✓ 정보주체의 동의 범위를 초과하여 이용하거나 제3자에게 제공한 경우	제18조 제1항, 제2항	제71조 제2호
	✓ 제공받은 정보를 목적 외의 용도로 이용하거나 제3자에게 제공한 경우	제19조	제71조 제2호
	✓ 수탁자가 해당 업무 범위를 초과하여 개인정보를 이용하거나 제3자에게 제공한 경우	제26조 제5항	제71조 제2호
	✓ 이전된 개인정보의 본래 목적 외로 이용하거나 제3자에게 제공한 경우	제27조 제3항	제71조 제2호
	✓ 민감정보 처리에 대한 별도의 동의를 받거나 법령에서 허용한 경우 외에 민감정보를 처리한 경우	제23조	제71조 제3호
	✓ 정보주체의 동의 또는 법률에 정한 바 없이 고유식별정보를 처리한 경우	제24조	제71조 제4호
	✓ 개인정보처리자가 업무상 알게된 개인정보를 누설하거나 권한 없이 제3자가 이용하도록 제공한 경우 ✓ 개인정보처리자가 권한을 초과하여 타인의 개인정보를 훼손, 멸실, 변경, 위조 또는 유출한 경우	제59조 제2호, 제3호	제 71조 제5호, 제6호
3년이하 징역 또는 3천만원 이하 벌금	✓ 영상정보처리기기를 설치목적과 다른 목적으로 임의조작 또는 다른 곳을 비추거나 녹음기능을 사용한 경우	제59조 제1호	제72조 제1호
	✓ 부정한 수단 또는 방법으로 개인정보를 취득하거나 동의를 받은 경우 및 그 정보를 제공받은 자	제59조 제1호	제72조 제2호
	✓ 개인정보보호위원회의 업무, 영향평가 업무, 분쟁조정업무로서 알게된 비밀을 누설하거나 직무상 목적 외로 이용한 경우	제60조	제72조 제3호

벌칙	행위내용	금지조항	벌칙조항
2년이하 징역 또는 1천만원 이하 벌금	✓ 개인정보처리자가 고유식별정보의 암호화 등 안전성 확보조치를 취하지 않아 개인정보의 분실, 도난, 유출, 변조 또는 훼손을 당한 경우	제24조 제3항	제73조 제1호
	✓ 영상정보처리기기 운영자가 안전성 확보조치를 취하지 않아 개인정보의 분실, 도난, 유출, 변조 또는 훼손을 당한 경우	제25조 제6항	제73조 제1호
	✓ 개인정보처리자가 내부관리계획, 접속기록 보관 등의 조치(시행령 제30조 미준용)를 취하지 않아 개인정보의 분실, 도난, 유출, 변조 또는 훼손을 당한 경우	제29조	제73조제1호
	✓ 정보주체의 요구에도 불구하고 개인정보의 정정 및 삭제 조치를 취하지 아니하고 개인정보를 계속 이용하거나 제3자에게 제공한 경우	제36조 제2항	제73조 제2호
	✓ 정보주체의 요구에도 불구하고 개인정보의 처리를 정지하지 아니하거나 제3자에게 제공한 경우	제37조 제2항	제73조 제3호
	양벌 규정		제74조

참 고 문 헌

1. 한국인터넷진흥원, “美 행정부, '빅데이터 R&D 이니셔티브'발표”, 한국인터넷진흥원, 2012. 4.
2. 미래창조과학부, 한국정보화진흥원, “빅데이터 활용을 위한 개인정보 비식별화 사례집”, 2014. 5.
3. 방송통신위원회, “빅데이터 개인정보보호 가이드라인”, 2014. 12.
4. 최대선, 이윤호, “공공정보 개방 · 공유에 따른 개인정보보호 기술”, 정보과학회논문지 : 시스템 및 이론 제41권 제3호, 2014. 6.
5. 안전행정부, “공공정보 개방 · 공유에 따른 개인정보 보호 지침”, 2013. 9.
6. 국회입법조사처, “공공데이터 개방 및 빅데이터 활용 지원 서비스 현황과 과제”, NARS 현장조사보고서 제 35호, 2014. 12
7. 행정자치부, “공공데이터 관리지침(행정자치부고시 제2014-1호)”, 2014. 3.
8. 한국인터넷진흥원, “2011년 개인정보보호관리체계(PIMS) 구축 및 운영 교육”, 2011.
9. 김선남, 이환수, “빅데이터 개인정보보호 가이드라인(안)의 개선 방향에 관한 연구”, 정보화정책 제21권 제4호, 2014년 겨울, pp.20-39, 2014. 12.
10. 개인정보보호위원회, “빅데이터 환경에서 개인정보보호 강화를 위한 법 · 제도적 대책 방안 연구”, 2012. 12.
11. 미래창조과학부, 한국정보화진흥원, “빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼(Version 1.0), 2014. 5.
12. 관계부처 합동, “개인정보보호 정상화 대책 – 기본이 지켜지는 개인정보보호 – ”, 2014. 7.
13. 한국인터넷진흥원, “개인정보 영향평가 수행 안내서”, 2011. 12.
14. 행정자치부, 한국인터넷진흥원, “개인정보의 안전성 확보조치 기준 해설서”, 2015. 2.
15. 행정자치부, 한국인터넷진흥원, “시스템 개발 · 운영자를 위한 개인정보보호 가이드라인”, 2015. 3.

16. 방송통신위원회, “개인정보 최소수집 · 보관을 위한 온라인 개인정보 취급 가이드라인”, 2014. 11.

17. 한국인터넷진흥원, “정보통신 환경변화에 따른 개인정보 법제도 개선방안 연구”, 2013. 10.

18. 한국정보화진흥원, “2014 개인정보보호 트렌드 전망” 2014.

19. 정보통신정책연구원, “빅데이터 2.0 시대 주요 이슈와 정책적 시사점, 2014. 12.

20. 방송통신위원회, “비식별 개인정보의 보호 및 활용에 관한 연구”, 2010. 8.

21. ICO (Information Commissioner’s Office, 영국), “Anonymisation: managing data protection risk code of practice,” 2012.

22. Bernhard Riedl, Thomas Neubauer, Gernot Goluch, “A secure architecture for the pseudonymization of medical data,”IEEE Computer Society, 2007.

MEMO

MEMO

빅데이터 활용을 위한

개인정보 비식별화 기술활용 안내서 Ver 1.0

발행인 | 서병조

발행일 | 2015년 6월 10일

발행처 |

NIA 한국정보화진흥원

무교청사 : 100-775 서울특별시 중구 청계천로(무교동 77번지) 14 NA빌딩
등촌청사 : 157-715 서울특별시 강서구 공항대로 489(등촌동)
www.nia.or.kr

편집 · 디자인 | 전우용사촌(주)