교육분야 가명정보 처리 가이드라인

2020. 10.

교육부
Ministry of Education

가이드라인 활용 안내

□ 개 요

본 가이드라인은 교육 분야 개인정보를 가명처리 및 익명처리하는 절차와 방법에 대한 내용을 제공하는 자료로써 다음 사항을 참고하여 활용하시기 바랍니다.

1. 가이드라인 내용의 기본방향은 개인정보보호 관련 법령을 준수하여 학생, 학부모, 교직원 등 정보주체의 권리보호 및 개인정보처리자의 책임성 강화를 목적으로 합니다.

2. 본 가이드라인은 교육 분야의 특성을 고려한 최소한의 가명정보 처리 기준을 안내하고 있습니다.

3. 가명정보 처리 업무에 본 가이드라인을 활용하시기 바라며, 향후 법령의 변경 등에 따라 내용은 언제든지 수정·보완될 수 있습니다.

4. 본 가이드라인은 2020년 11월 기준으로 작성되었습니다. 항상 최신의 가이드라인은 교육부 개인정보보호 포털(자료실→참고자료)에서 확인하시기 바랍니다.

[제·개정 이력]

순 번	구 분	시행 일자	제정 · 개정 주요내용
1	제정	2020. 10.	1. ○○○○○ 2. ○○○○○ 3. ○○○○○
○	일부개정	○○○○. ○○. ○○.	1. ○○○○○ 2. ○○○○○ 3. ○○○○○
○	전부개정	○○○○. ○○. ○○.	1. ○○○○○ 2. ○○○○○ 3. ○○○○○

목 차

제 Ⅰ 편

가이드라인 개요

1 배경 및 목적

❍ 개정된 「개인정보 보호법」('20.8.5.시행)에서는 가명정보 처리에 관한 특례가 신설되어 개인정보처리자가 통계작성, 과학적 연구, 공익적 기록보존 등을 위해 개인정보를 가명처리하여 활용할 수 있는 기반 마련

❍ 4차 산업혁명 시대를 맞아 핵심 자원인 데이터의 이용 활성화를 통한 신산업 육성이 범국가적 과제로 대두되고 있고 교육정보 활용을 위한 가명정보의 처리 및 결합 등에 대한 수요 증가가 예상됨에 따라 교육분야 특수성을 고려한 가명정보 처리 체계 마련 필요

❍ 이에 본 가이드라인은 교육분야의 가명정보를 처리하는 과정에서 발생할 수 있는 개인정보 오·남용을 방지하고 안전한 가명정보 처리 방안을 제시하여 보다 풍부하고 다양한 교육정보 활용을 통한 효율적 교육정책 수립 등을 지원하고자 함

< 관 련 법 령 >

- 개인정보 보호법 제2조 및 제28조의2부터 제28조의6
  * 가명정보의 처리, 결합제한, 안전조치의무, 금지의무, 과징금 부과 등
- 개인정보 보호법 시행령 제29조의2부터 제29조의6
- 가명정보의 결합 및 반출 등에 관한 고시
- 가명정보 처리 가이드라인(개인정보보호위원회, '20.9.)

2 적용 범위

❍ **(우선순위)** 교육 분야의 개인정보 가명처리 및 결합 등에 관해서는 동 가이드라인을 우선 적용

※ 동 가이드라인에서 별도로 정하지 않은 사항은 공통가이드라인(개인정보보호위원회) 준용

❍ **(적용대상)** 교육 분야 데이터를 취급하는 교육행정기관*, 학교**, 연구자, 공공기관 등 개인정보처리자 전체

※ 가명정보를 보유·제공하고 활용하는 자·기관(법인) 모두 해당

* 교육부 및 그 소속 기관과 특별시·광역시·특별자치시·도 또는 특별자치도의 교육 관서(교육공무원법 제2조제4항)

** 유아교육법 제2조제2호에 따라 설립된 유치원 및 초·중등교육법 제2조·고등교육법 제2조에 따라 설립된 각급학교

3 용어 정리

❍ 개인정보 : 살아있는 개인에 관한 정보로서 다음의 정보를 포함

가. 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보

나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 함

다. 가목 또는 나목을 개인정보보호법 제1호의2조에 따라 가명처리함으로써 원래의 상태로 복원하기 위한 추가정보의 사용·결합 없이는 특정개인을 알아볼 수 없는 정보

※ 개인정보에 대한 판단기준은 개인정보처리자가 보유한 정보 또는 접근 가능한 권한 등 상황에 따라 상이

❍ 개인식별정보(식별자)

- 고유식별정보, 이메일주소, 휴대전화번호 등과 같이 그 자체로 특정 개인을 직접 식별하는 용도로 사용하는 정보 등

❍ 개인식별가능정보(준식별자 또는 간접식별자)

- 연령, 성별, 거주 지역, 국적 등과 같이 해당 정보만으로는 직접적으로 특정 개인을 식별할 수 없지만, 다른 정보와 결합하여 특정 개인을 전부 또는 일부를 드러낼 수 있는 정보

※ 개인식별가능정보는 다른 속성과 결합할 경우 개인 식별 가능성이 높고 낮음에 따라 가명처리 및 익명처리 수준 등을 달리할 수 있으며, 해당 속성이 개인 식별 가능성 여부는 구체적인 사례에 따라 달리 판단 가능

[주요 개인정보 분류 예시]

구분	분류	대상 개인정보
그 자체로 개인의 식별이 가능하거나 매우 민감한 개인정보 또는 관련 법령에 따라 처리가 엄격하게 제한된 개인정보	고유식별정보	• 주민등록번호, 여권번호, 운전면허번호, 외국인 등록번호
	민감정보	• 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 병력, 범죄 경력정보 등 사생활을 현저하게 침해할 수 있는 정보 • 개인의 신체적·생리적·행동적 특징에 관한 정보 • 인종이나 민족에 관한 정보
	인증정보	• 비밀번호, 바이오정보(홍채, 지문, 정맥 등)
	금융정보	• 신용카드번호, 계좌번호 등
	의료정보	• 학생건강기록부, 건강검진기록, 진료기록 등
	위치정보	• 개인 위치정보 등
자체로 식별될 가능성이 높거나 조합되면 명확히 개인의 식별이 가능한 정보	개인식별정보	• 이름, 상세주소, 핸드폰번호, 이메일주소, 회원번호 등
	개인식별 가능정보	• 시군구단위 주소, 생년월일, 성별, 학력, 직업, 키, 몸무게, 혼인여부, 가족상황, 취미 등
	자동생성정보	• IP주소, MAC 주소, 사이트 방문기록, 쿠키 등
	가공정보	• 통계성 정보, 가입자 성향 등
	제한적 본인 식별정보	• 학번, 내부용 개인식별정보 등

[표 1] 주요 개인정보 분류 예시

❍ 가명처리

- 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것

❍ 가명정보

- 개인정보를 가명처리 함으로써 원래의 상태로 복원하기 위한 추가정보의 사용 · 결합 없이는 특정 개인을 알아볼 수 없는 정보

※ 가명정보도 개인정보의 범주에 포함

❍ 추가정보

- 개인정보의 전부 또는 일부를 대체하는 데 이용된 수단이나 방식(알고리즘, Salt 값 등), 가명정보와의 비교·대조 등을 통해 삭제 또는 대체된 개인정보 부분을 복원할 수 있는 정보(매핑 테이블 정보, 가명처리에 사용된 개인정보 등)

※ 추가정보(원본정보와 알고리즘·매핑테이블 정보 등)와 가명정보는 시행령 제30조 또는 제48조의2에 따른 안전성 확보조치 및 각각 정보의 분리보관, 접근 권한의 분리를 하여야 함

❍ 특이정보

- 다른 데이터와 확연히 구분되거나 비정상적으로 데이터의 분포를 벗어나 측정이 되는 값으로서, 개인정보 식별과 관련하여 특정 개인의 식별 가능성이 매우 높은 정보

❍ 가명정보처리자

- 업무를 목적으로 개인정보를 가명처리하여 활용 또는 제공하는 공공기관, 법인, 단체 및 개인 등

❍ 가명정보취급자

- 가명정보를 처리하는 개인정보처리자의 지휘·감독을 받아 가명정보를 처리하는 임직원, 파견근로자, 시간제근로자 등

❍ 익명정보

- 시간 · 비용 · 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보

❍ 익명처리

- 개인정보의 전부 또는 일부를 데이터 값 삭제, 가명처리, 총계처리, 범주화 등 다양한 기술을 적용함으로써 더 이상 특정 개인을 알아볼 수 없도록 익명정보로 처리하는 것

❍ 익명정보처리자

- 업무를 목적으로 개인정보를 익명처리하여 활용 또는 제공하는 공공기관, 법인, 단체 및 개인 등

❍ 적정성 검토

- 본 가이드라인에서 제시하고 있는 절차를 기반으로 사전에 정의한 가명처리 기준에 따라 적절히 가명처리 되었는지 확인하는 절차

❍ 재식별

- 특정 개인을 알아볼 수 없도록 처리한 가명정보에서 특정 개인을 알아보는 것

❍ 정보집합물

- 정보를 체계적으로 관리하거나 처리할 목적으로 일정한 규칙에 따라 구성되거나 배열된 둘 이상의 정보들

○ 결합키(가명 ID)

- 결합 대상 가명정보의 일부로서 해당 정보만으로는 특정 개인을 알아볼 수 없으나 다른 정보주체와 구별할 수 있도록 조치한 정보

○ 결합키연계정보

- 동일 정보주체에 대해 가명정보를 결합할 수 있도록 서로 다른 결합신청기관 간의 결합키를 연계한 정보

○ 결합신청자

- 가명정보의 결합을 신청하는 개인정보처리자 또는 기관

※ 가명정보를 제공만 하는 자, 가명정보를 제공하고 결합한 정보를 이용하는 자 또는 가명정보의 제공 없이 결합한 정보를 이용하는 자를 모두 포함

○ 결합전문기관

- 개인정보 보호법 제28조의3제1항에 따라 서로 다른 개인정보처리자 간의 가명정보 결합을 수행하기 위해 개인정보보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관

○ 결합키관리기관

- 개인정보 보호법 시행령 제29조의3제2항에 따라 결합키연계정보를 생성하여 결합전문기관에 제공하는 등 가명정보의 안전한 결합을 지원하는 업무를 하는 한국인터넷진흥원 또는 보호위원회가 지정하여 고시하는 기관

○ 반출처리공간

- 결합신청기관이 결합된 정보를 반출하기 전 추가 가명·익명처리를 위해 안전성 확보에 필요한 기술적·관리적·물리적 조치가 된 공간

제 II 편

가명처리

1 개 요

❖ 가명정보는 법령에서 허용하는 정당한 처리 범위 내에서 통계 작성, 과학적 연구, 공익적 기록보존 등의 목적으로 정보주체의 동의 없이 처리 가능

□ 가명처리 목적

○ 통계작성

- 집단적 현상이나 수집된 자료의 내용에 관한 수량적인 정보를 작성하는 행위를 말함

※ 직접(1:1) 마케팅 등을 위하여 특정 개인을 식별할 수 있는 형태의 통계는 해당하지 않음

▸ (예시) 대학이 학생들의 취업지 활동 지원(직업군 및 교육과정 추천)을 위하여 가명처리된 졸업생들의 학습이력 분석과 취업기관 및 유형들에 대한 매칭 통계를 작성하는 경우

○ 과학적 연구

- 기술 개발, 실증, 기초연구, 응용연구, 민간투자연구 등 과학적 방법을 적용하는 연구를 말함

※ 자연과학적 연구뿐만 아니라 과학적 방법을 적용하는 역사적 연구, 학생 보건 분야에서 공익을 위해 시행되는 연구 등은 물론, 새로운 서비스의 개발, 시장조사 등 산업적 목적의 연구도 포함

▸ (예시) 교육부가 학생의 학생 학습 및 미래 보장을 위해 학생생활기록부, 건강기록부, 출결 정보 등을 심층 분석하여 위기징후* 탐지 알고리즘을 연구하여 새로운 서비스(시스템)을 구축·운영하려는 경우

* 중도 학업 포기, 학교폭력 등

○ 공익적 기록보존

- 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 기록정보를 보존하는 것을 말함

※ 공공기관이 처리하는 경우에만 공익적 목적이 인정되는 것은 아니며, 민간기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우도 공익적 기록보존 목적이 인정됨

▸ (예시) 학내 연구소가 현대사 연구 과정에서 수집한 개인정보 중에서 사료 가치가 있는 인물정보를 기록하여 보관하는 경우

□ 가명처리 원칙

○ 가명처리 대상은 법률에서 허용한 목적 내에서 개인정보를 정보주체의 추가적인 동의 없이 수집 목적 외로 이용 가능

- 민감정보와 고유식별정보(주민등록번호 제외)도 가명 처리할 수 있음
- 고유식별정보는 직접 식별자에 해당하므로 고유식별정보가 남아 있거나 역추적이 가능하도록 해서는 안 됨
- 주민등록번호는 법률 또는 시행령에 구체적인 근거가 없으면 사용이 안됨

○ 개인식별정보(식별자)는 삭제하여야 하나, 결합 등 데이터 이용 목적 상 필요한 경우 안전한 방식으로 대체값을 생성하여 개인식별정보를 대체

○ 개인정보를 가명처리하여 발생한 추가정보 삭제

- 다만, 불가피한 경우 추가정보는 법령에서 요구하는 분리보관 등 안전성확보조치를 취하여 보관 가능

○ 가명처리는 개인정보 보유부서 또는 총괄부서*에서 처리하도록 함

* 개인정보처리자는 가명처리 관련 업무의 총괄·관리 및 의사결정을 위한 총괄부서(또는 담당자)를 지정할 수 있음

※ 내부결합, 익명처리 등도 위에 준해서 처리

○ 가명처리를 수행하는 자 및 가명정보취급자는 가명처리 전 개인정보에 대한 취급권한 분리

- 취급권한을 분리할 수 없는 불가피한 사유가 있을 경우 보완통제 대책을 수립하여 관리자의 승인 하에 제한적으로 취급 가능

※ 해당 부서 소유 개인정보를 가명처리하여 해당 부서에서만 사용하는 등 가명정보를 취급할 자를 추가로 둘 여력이 없거나 분리가 어려운 정당한 사유가 있는 경우에는 업무 수행에 필요한 최소한의 취급권한 부여와 관련 기록을 보관 하는 등 예외적으로 보안대책을 수립한 후 허용 가능

○ 안전한 가명처리 방법을 적용하여 가명정보를 변환하도록 함

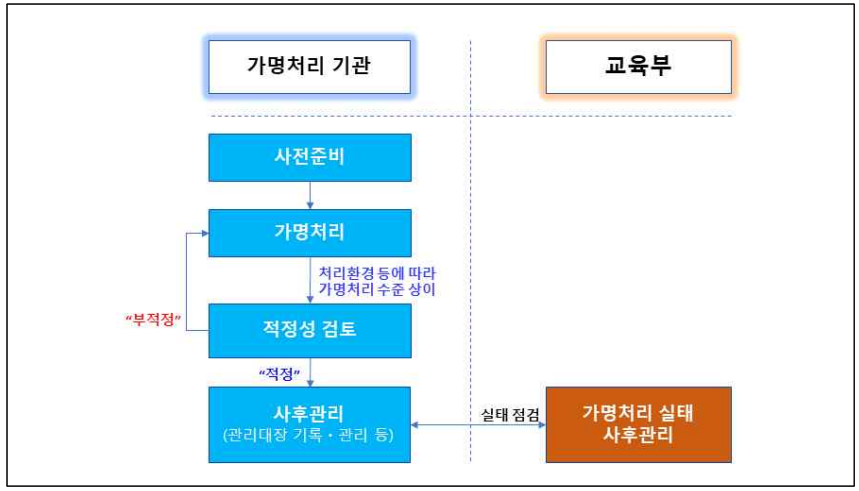
○ 가명정보를 제공받는 자는 해당 가명정보의 가명처리 참여 불가

○ 개인정보처리자는 가명처리를 제3자에게 위탁하는 경우 위탁계약서 등에 안전성 확보조치 관련 내용들을 포함하여 진행

※ 가명정보 기술지원을 한 자는 제공받은 가명정보의 취급권한 부여 금지 등 안전성 확보조치를 계약서에 포함하여 수행

○ 개인정보처리자는 가명처리를 수행하는 자 등에 대해 년 1회 이상 가명처리 관련 교육을 실시하거나 관련 교육 및 기술 세미나 등에 참석할 수 있도록 조치

□ 가명처리 절차 개념도



<그림 2> 가명처리 절차

○ **(사전준비)** 가명처리 목적을 명확히 정의하고 가명처리 대상 개인정보를 선정

○ **(가명처리)** 가명처리 수준을 정의하고 수준에 맞도록 가명처리 기법을 활용하여 개인정보를 가명처리

○ **(적정성 검토)** 가명처리 수준 등에 맞는 가명처리가 되었는지 여부를 확인하고 가명정보 내 개인정보 재식별 여부 및 재식별 가능성 검토

○ **(사후관리)** 가명처리된 가명정보와 추가정보에 대해 안전성 확보조치를 수행하고, 재식별 모니터링 및 재식별사고 발생에 관한 대책을 수립하고 이행

○ **(가명처리 실태 사후관리)** 교육부 등은 가명처리 현황에 대해 실태점검을 실시하여 미흡한 사항에 대한 시정조치 요구 가능

□ 가명처리 예시

[개인정보와 속성자 특징]

성명	연락처	성별	생년월일	혈액형		전공	학위
				ABO	RH		
김영희	090-1234-5678	여	19650512	A	Rh+	법학	박사
강순희	090-8525-4564	남	19671212	B	Rh+	컴퓨터	학사
최복례	090-8546-5456	여	19681015	O	Rh+	건축	학사
홍길동	090-5524-1325	남	19920721	AB	Rh-	보안	학사
이홍준	090-6974-1235	남	19930423	AB	Rh+	교육	학사
김신우	08-3456-7890	남	19940925	O	Rh+	음악	학사
〃	〃	〃	〃	〃	〃	〃	〃

[표 2] 개인정보 속성자 특징 예시

- 성별/혈액형별 나이와 전공 및 학위에 대한 상관관계 연구 목적으로 가명정보를 만들 경우 '성명', '연락처'는 직접적으로 개인이 식별 가능
- '성별', '생년월일'은 다른 정보와 조합하여 개인을 식별할 가능성이 높고 '혈액형'에서 Rh-의 경우 희귀 혈액형 유형으로 개인이 식별될 가능성이 매우 높음.
- '전공', '학위'는 다른 정보와 결합하여 개인이 식별될 가능성이 있으나 비교적 낮음

[가명처리]

- '성명', '연락처'는 특정 개인 식별이 가능하므로 삭제 처리를 하되, 데이터의 이용 목적상 개인을 특정할 필요가 있는 경우나 결합이 필요한 경우 특정 키(가명 ID)나 결합키의 입력값으로 활용
  ☞ 가명처리 기법 중 하나인 해시함수(SHA-256 이상)와 솔트값(salt)을 적용하여 가명 ID를 생성한다.

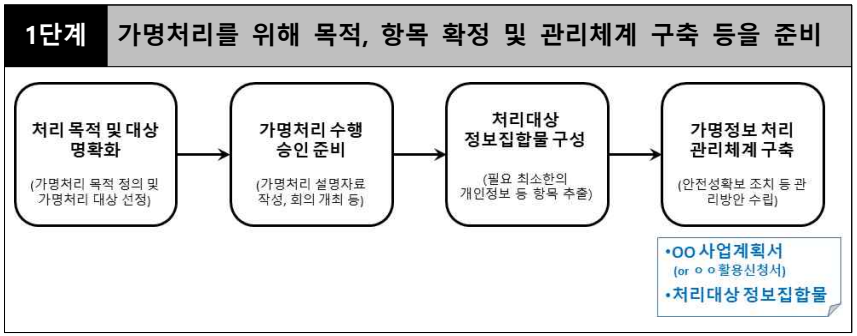
- '성별', '생년월일'은 다른 정보와 조합하여 개인을 식별할 가능성이 비교적 높고 '혈액형'에서 Rh-의 경우 희귀 혈액형 유형으로 개인이 식별될 가능성이 매우 높음. 따라서 혈액형에서 Rh-의 경우 특이정보로 판단하여 해당 레코드를 삭제 처리하고 생년월일은 출생년도로 범주화. 다만, 전체 데이터에서 목적에 따라 검증하여 특정인이 식별되는 경우 범주화를 연령대로 확대하거나 특정 레코드 삭제 등을 수행
- '전공', '학위'는 다른 정보와 조합하여 개인이 식별될 가능성이 있으나 특이 전공 등이 존재하는 지 확인하여 개인 식별 가능성이 낮은 경우 별도 가명 조치 없이 그대로 활용

가명 ID (특정키, 또는 결합키)	성명	연락처	성별	생년월일	혈액형		전공	학위
					ABO	RH		
725F8676075F7C0C5E6BAA0EAC76E26C1ED3A19A7AA74661162861D0573D4C334			여	1965~~0512~~	A	Rh+	법학	박사
9A00F1155584BA5DDFFBAA0EAC76E26C1ED3A19A7AA74661162861D04E5777AE4			남	1967~~1212~~	B	Rh+	~~컴퓨터~~	학사
27B339D75FF1DCED2CBAA0EAC76E26C1ED3A19A7AA74661162861D0CF37C17B8			여	1968~~1015~~	O	Rh+	건축	학사
11834268AF3110DB643BAA0EAC76E26C1ED3A19A7AA74661162861D021FDEED7E			~~남~~	~~19920721~~	~~AB~~	~~Rh-~~	~~보안~~	~~학사~~
6CE926B166980F9C5F0BAA0EAC76E26C1ED3A19A7AA74661162861D09E1185FA6C			남	1993~~0423~~	AB	Rh+	교육	학사
725F8676075F7C0C5E6BAA0EAC76E26C1ED3A19A7AA74661162861D011AAE952D			남	1994~~0925~~	O	Rh+	음악	학사
〃			〃	〃	〃	〃	〃	〃

[표 3] 개인정보 가명처리 예시

2 가명처리 세부 절차

2-1 사전준비 (1단계)



<그림 3> 가명처리 적정성 검토 단계 세부 절차

○ **(처리 목적 및 대상 명확화)** 가명처리 목적*을 명확히 정의하고 가명처리 대상 개인정보(개인정보파일, DB, 테이블 등) 선정

* 통계작성, 연구(산업적 연구 포함), 공익적 기록보존 목적 등

○ **(가명처리 수행 승인)** 계획수립 및 회의 개최 결과 등을 반영하여 기관장 또는 개인정보 보호책임자* 승인 후 가명처리 수행

* 개인정보처리자의 개인정보 처리에 관한 업무를 총괄해서 책임지거나 업무처리를 최종적으로 결정하는 자(법 31조, 시행령 제32조, 지침 21조)

※ 주요 처리 사항

- 기관 외부에 가명정보를 제공하는 경우 이용목적 및 방법, 재식별 위험관리, 가명정보의 안전성 확보를 위해 필요한 조치 마련, 목적외 사용 제한, 파기일자 명시 및 파기 시 관련 내용 통보 (사용기한에 제한을 둔 경우) 등을 포함하여 법적인 계약 체결

※ 타 기관과 가명정보를 결합 제공하는 경우도 동일

○ **(처리 대상 정보집합물 구성)** 가명처리 목적 달성을 위해 필요한 최소한의 항목을 추출하여 가명처리 대상 정보집합물 구성

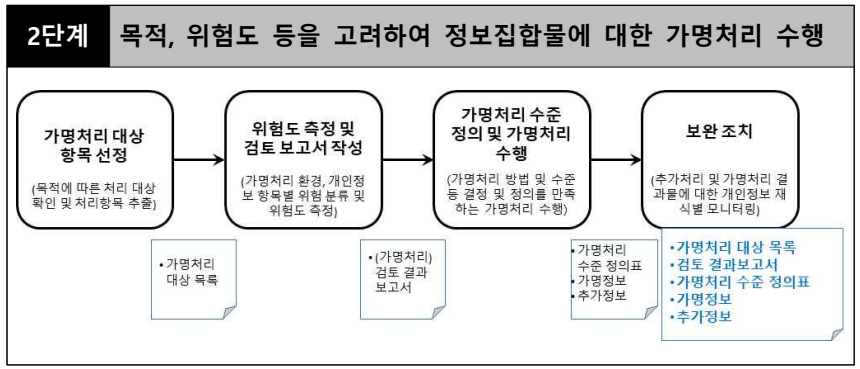
○ **(가명정보 처리 관리체계 구축)** 가명정보를 안전하게 보호할 수 있는 관리방안을 수립하고 추가정보 등에 대한 접근관리 체계 구축

※ 추가정보 안전성 확보, 접근관리, 가명정보취급자 지정 등

주요 산출물

- 사업계획서
- 처리대상 정보집합물(개인정보)

2-2 가명처리 (2단계)



<그림 4> 가명처리 적정성 검토 단계 세부 절차

□ 가명처리 대상 항목 선정

○ 사업계획서 등을 통해 정보집합물에서 처리대상 및 대상별 특성을 확인하고 가명처리에 필요한 최소한의 항목을 추출

▸ 정보집합물 항목 선정 (예시)
- 이름, 휴대폰 번호, 성별, 이메일, **주소**, 재학 학교명, 학년/반/번호

▸ 가명처리 목적 (예시) : 학교별 재학생의 성별 및 지역분포 통계
- 성별, **시군구 주소,** 재학 학교명

※ 분석 목적과 상관없는 개인정보 및 기타 정보는 대상 선정에서 제외

□ 위험도 측정 및 검토보고서 작성

○ **(처리환경별 위험도 측정)** 가명처리 환경과 개인정보 항목별 위험도를 분류하고 이를 기반으로 개인정보 위험도 측정

1. 가명처리 환경에 따른 위험도 분류

- 처리 목적에 따라 처리(제공)환경과 제공받는 자의 개인정보 보호수준 및 다른 정보 보유여부 등을 검토

※ 불특정 제3자(공개 등)에게 제공하는 경우 익명정보로 처리하는 것을 원칙

처리환경				기관의 개인정보 보호수준*	재식별수준 (다른정보 보유 등)**
활용장소	활용처	처리유형	기관유형		
내부	소유부서	활용	교육기관	상	상
외부	타부서	제공	민간기관	중	중
		결합		하	하

[표 4] 처리환경별 위험도 분류 고려사항 예시

* 가명정보처리자나 혹은 제공받는 자의 안전조치 수준

** 개인정보처리자가 보유한 다른 정보나 혹은 제공받는 자가 보유한 다른 정보와의 연결을 통해 특정 개인을 알아볼 가능성

- 처리(제공)환경과 정보집합물의 규모, 정보의 정확성 수준 등을 고려하여 다양한 요소를 가감할 수 있음

※ 예시) 활용처를 소유부서, 특정 타부서, 전체부서 등으로 추가, 개인정보 보호수준 및 재식별 수준 상-중-하, 1급~5급 등 단계 다양화 등

[처리환경별 위험도 분류 예시]

· ㅇㅇ교육청이 A부서의 A1 개인정보를 가명처리하여 B부서에서 사용하며 보호수준은 높고, 재식별 가능성은 낮은 상태라고 가정할 경우

처리환경				기관의 개인정보 보호수준**	재식별수준 (다른정보 보유 등)**
활용장소	활용부서	처리유형	기관유형		
내부(1)	소유부서(1)	활용(1)	**교육기관(1)**	**상(1)**	상(3)
외부(2)	**타부서(2)**	**제공(2)**	민간기관(2)	중(2)	중(2)
		결합(3)		하(3)	**하(1)**

[표 5] 처리환경별 위험도 분류 측정 예시

· 각각의 분류 기준의 점수를 합산하여 처리환경 1+2+2+1=6점, 보호수준 1점, 재식별수준 1점, 총 8점으로 산정하고 점수가 높을수록 위험도가 높은 처리환경으로 판정할 수 있음

· 처리환경별 위험도 점수별 가명처리 방안을 마련하여 해당 방안을 적용

2. 개인정보 항목별 위험도 분류

- 기관 자체적으로 개인정보 분류체계표(표1 참조)를 작성하고 해당 분류체계표에 따른 위험도를 분류하여 적용

[항목별 위험도 분류 예시]

분류	항목	위험도
고유식별정보	주민등록번호	사용불가
	여권번호, 운전면허번호, 외국인등록번호	10
개인식별정보	상세주소	6
〃	〃	〃

○ **(가명처리 검토 결과보고서 작성)** 가명정보처리자는 가명처리 환경 위험도와 항목별 위험도를 종합적으로 고려하여 가명처리에 대한 위험도 평가 결과 도출

- 가명정보 활용에 따른 처리 환경과 개인정보 분류를 통해 도출된 위험도를 기반으로 '가명처리 검토 결과보고서'(별지 1) 작성 관리

□ 가명처리 수준 정의 및 가명처리 수행

○ **(가명처리 수준 정의)** '가명처리 검토 결과보고서'(별지 1)를 기반으로 적절한 가명처리 방법과 수준을 결정하고 '가명처리 수준 정의표'(별지 2) 작성

○ **(가명처리 수행)** '가명처리 수준 정의표'(별지 2)를 기반으로 다양한 가명처리 기술(부록 5 참조)을 적용하여 가명처리 수행

- 가명처리 단계에서 생성되는 추가정보는 삭제하는 것을 원칙으로 하되, 시계열분석 등과 같이 불가피하게 저장하여야 하는 경우 분리 보관

※ 세부사항은 Ⅱ편 > 3. 가명처리의 안전한 관리 > 기술적 보호조치 참고

※ '가명처리 검토 결과보고서' 또는 '가명처리 수준 정의표'에는 프라이버시 보호모델 적용 여부 등을 추가할 수 있음

□ 보완조치

○ **(특이정보 처리)** 가명처리된 정보에서 개인이 식별될 가능성이 높은 특이정보를 확인하여 추가적으로 가명처리

- ▸ 수능 만점자, 국내 최고령, 최장신, 고액 체납금액, 고액 장학금기부자 등 전체적인 패턴에서 벗어나 극단값이 발생할 수 있는 정보
- ▸ 희귀 성씨, 희귀 혈액형, 희귀 눈동자 색깔, 희귀 병명, 희귀 직업 등 정보 자체로 특이한 값을 가지고 있는 정보 등

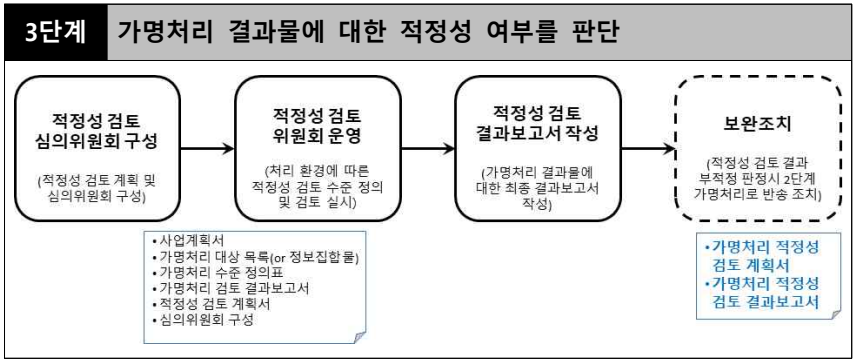
※ 특이정보 처리 사례는 개인정보보호위원회『가명정보 처리 가이드라인』>'참고자료 2. 특이정보 정의 및 처리사례'를 참조하여 처리

- 특이정보를 처리하는 과정에서 변경사항 발생 시 필요한 경우 기존에 작성된 '가명처리 검토 결과보고서'(별지 1)나 '가명처리 수준 정의표'(별지 2)에 해당 내용 반영

주요 산출물

- 가명처리 대상 목록
- 가명처리 검토 결과보고서
- 가명처리 수준 정의표
- 가명정보, 추가정보

2-3 적정성 검토 (3단계)



<그림 5> 가명처리 적정성 검토 단계 세부 절차

□ 심의위원회 구성

○ 가명정보처리자는 처리된 가명정보의 적정성 검토를 위한 심의위원회를 구성

- **(위원 자격 및 인원)** 개인정보보호, 가명처리 기법 등에 관한 학식과 경험이 풍부한 사람으로 다음 각 호에 해당하는 사람을 고루 포함하여 3명 이상 7명 이내의 위원으로 구성

1. 개인정보 보호와 관련한 업무 경력이 있거나 관련 단체로부터 추천을 받은 사람
2. 개인정보처리자로 구성된 단체에서 활동한 경력이 있거나 관련 단체로부터 추천을 받은 사람
3. 그 밖에 개인정보 보호와 관련한 경력과 전문성이 있는 사람

※『가명정보의 결합 및 반출 등에 관한 고시』(개인정보보호위원회 고시) > 별표 1 결합전문기관 지정 기준 > 2호의 자격 기준 준용 가능

- **(위원장)** 위원 중 호선으로 정하거나 가명정보 처리기관의 개인정보보호책임자 또는 그에 준하는 임직원

○ 내·외부 전문가로 위원 구성 시 비율은 기관별 업무 성격 등에 따라 탄력적으로 구성하되, 교육행정기관 및 학교* 이외에 가명정보 제공 시에는 반드시 외부 전문가를 포함하여 구성

* 교육 분야 가명정보 처리 가이드라인 적용대상 참고

□ 심의위원회 운영

○ **(자료 준비)** 가명정보처리자는 적정성 검토를 위한 기초자료* 준비

* 사업계획서, 가명처리 수준 정의표, 가명처리 정보집합물 샘플(필요시 전체 데이터), 가명처리 주요 검증 지표, 적정성 검토 계획서, 심의위원회 구성(안) 등

기초자료에 포함할 내용	기초자료명 예시
사용 목적	사업계획서
이용 환경(개인정보보호 수준, 보유 정보 등)	
가명처리 대상 정보집합물 및 가명정보 명세	가명처리 정보집합물
개인정보 항목별 적용 가명처리 기법	가명처리 검토 결과보고서, 가명처리 수준 정의표
가명처리 수준 기준	가명처리 수준 정의표
기타 프라이버시 보호모델 적용 여부 등이 기초자료에 포함이 가능	가명처리 수준 정의표 또는 가명처리 검토 결과보고서

[표 6] 기초자료에 포함될 내용별 예시

○ **(적정성 검토)** 안전한 물리적 또는 논리적 공간에서 기초자료를 기반으로 적정성 검토 실시

- 가명처리 자체의 적정성뿐만 아니라 목적 달성을 위한 최소한의 가명정보 만으로 생성되었는지 여부 검토
- 가명정보의 활용 목적을 달성할 수 있는 지와 재식별 가능성이 없는 지 여부 검토
- 가명정보를 제3자에게 제공하는 경우 가명정보 재식별 의도 및 능력과 가명정보 보호수준 및 신뢰도를 추가적으로 검토

※ 적정성 검토를 위한 자료는 개인정보에 해당하므로 반드시 심의위원회 위원들을 대상으로 서약서 등 개인정보 보호법 준수를 보장

○ **(자료 보완)** 항목별 위험도를 바탕으로 가명처리한 경우에도 특이 정보를 통해 개인식별이 가능할 수 있으므로 추가로 가명처리 필요

□ 적정성 검토 결과보고서 작성

○ 심의위원회는 적정성 검토 결과에 대해 ‘가명처리 검토 결과 보고서’(별지 1) 및 ‘가명처리 수준 정의표’(별지 2)를 참고하여 ‘적정성 검토 결과보고서’(기관별 자유 양식) 작성

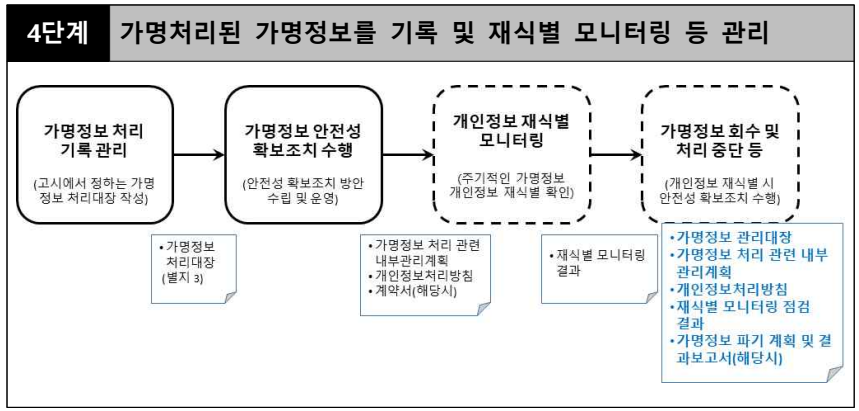
□ 보완조치

○ 적정성 검토 결과 “부적정” 판정 시 가명정보의 활용·제공 계획을 중단하거나, 가명정보 활용을 계속하고자 한다면 2단계인 “가명처리” 단계로 돌아가 추가적인 가명처리 조치를 해야 함

주요 산출물

- 가명처리 적정성 검토 계획서(위원회 구성안 포함)
- 가명처리 적정성 검토 결과보고서

2-4 사후관리 (4단계)



<그림 6> 가명처리 적정성 검토 단계 세부 절차

○ **(기록 관리)** 가명정보처리자(제공자)는 가명정보 처리에 관한 내용*을 기록으로 작성하고 안전하게 보관·관리(별지 3 참조)

* 가명정보의 처리 목적, 가명처리한 개인정보의 항목, 가명정보의 이용내역, 제3자 제공시 제공받는 자, 처리 및 보유기간, 추가정보의 이용 및 파기 등

○ **(안전성 확보조치)** 가명정보처리자(제공·활용하는 자)는 가명정보의 안전한 관리를 위한 내부관리계획의 수립 등 법령에서 요구하는 관련 조치수행

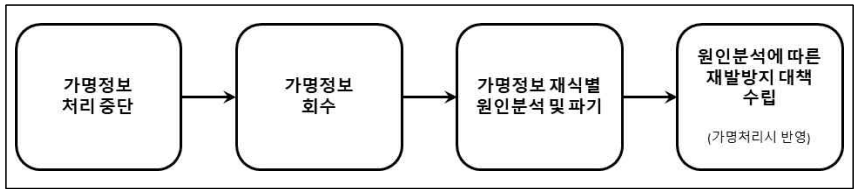
- 가명정보 처리 관련 주요 사항을 내부관리계획에 포함하여 수립하고 개인정보처리방침에 공개

※ 개인정보처리방침에 포함하여 가명정보 처리에 관한 사항이 공개될 경우 목적 달성이 어렵거나 비밀 등이 유출될 우려가 있는 경우에는 공개수준을 심의위원회를 거쳐 조정 가능

❍ **(재식별 모니터링)** 가명정보처리자(활용자)는 가명정보 보유기간 동안 개인정보 재식별 가능성이 증가하는 지 여부 등을 지속적으로 모니터링하고 정기적으로 점검(년 1회 이상 권고)

❍ **(가명정보 처리 중단 등)** 가명정보처리자(활용자)는 개인정보가 재식별된 경우 즉시 가명정보 처리 중단 및 즉시 삭제(파기) 등의 조치 수행

* 필요 시 개인정보처리자는 내부관리계획 또는 별도의 가명정보 처리 지침 등에 재발방지 방안 수립 가능



<그림 7> 가명정보 회수 및 처리 중단 절차

※ 교육부 실태점검 및 시정조치

- 필요 시 교육부는 교육 분야 가명정보처리자를 대상으로 가명처리 현황 조사 및 실태점검 등을 실시하여 미흡한 사항에 대해 해당 기관에 시정조치 요구 가능

주요 산출물

- 가명정보 관리대장
- 가명정보 처리 관련 내부관리계획 (기존에 존재할 경우 불필요)
- 개인정보처리방침 : 홈페이지에 해당 내용 공개
- 재식별 모니터링 점검 결과(연 1회 이상)

3 가명처리의 안전한 관리

□ 관리적 보호조치

❍ **(가명정보 처리 내부관리계획의 수립·시행)** 가명정보 처리 관련 내용을 내부관리계획에 포함하고 시행

※ 내부관리계획에 포함 사항

가. 가명정보 또는 추가정보의 관리책임자 지정에 관한 사항
나. 추가정보 별도 분리 보관 (개인정보-가명정보-추가정보 모두 분리)
다. 가명정보 또는 추가정보의 안전성 확보조치에 관한 사항
라. 가명정보취급자의 교육에 관한 사항
라. 가명정보 처리 기록 작성 및 보관에 관한 사항
마. 개인정보 처리방침 공개에 관한 사항
바. 가명정보의 재식별 금지 등 오남용 제한에 관한 사항
사. 가명정보 재식별 발생 시 대응조치
※ 가명정보처리자는 내부 가명정보의 처리 중단·회수·파기뿐만 아니라 제3자에게 제공한 가명정보도 관련 절차에 따라 파기할 수 있는 내용 포함

❍ **(개인정보처리방침 공개)** 가명정보 처리내용을 포함·공개

※ 개인정보처리방침에 포함 사항

가. 가명정보의 처리 목적
나. 가명정보의 처리 및 보유 기간
다. 가명정보의 제3자 제공에 관한 사항 (해당시)
라. 가명정보 처리의 위탁에 관한 사항 (해당시)
마. 처리하는 가명정보의 항목
바. 법 제28조의4에 따른 가명정보의 안전성 확보조치에 관한 사항

○ **(취급자 관리 및 교육)** 가명정보취급자 관리·교육 실시

- 가명정보취급자 직무 분리 : 가명정보의 원본 개인정보 및 추가정보 접근 금지 등 취급자 분리
- 가명정보취급자에 대한 교육계획 수립 및 교육 실시 등

○ **(가명처리 위탁 관리)** 가명정보 처리업무를 외부에 위탁하는 경우 수탁자 관리·감독을 위한 방안 마련 및 수행

- 법 제26조에 따라 위탁업무 수행 목적 외 개인정보의 처리 금지에 관한 사항 등을 포함한 수탁자 관리·감독 실시
- 개인정보 처리업무 위탁계약서에서 요구하는 주요 항목 이외에 추가 사항* 반영

* 가명정보 재식별 금지, 재식별 위험 발생 시 위탁사에 즉시 통지

○ **(제3자 제공시 계약 주의사항)** 가명정보를 제3자에게 제공시 보호대책을 마련하고 계약서에 주요 사항을 포함하여 시행

※ 가명정보 제3자 제공관련 계약서에 포함 사항

가. 가명정보 처리 목적
나. 가명정보 처리 및 보유 기간
다. 가명정보 재식별 금지 및 재식별 발생시 통지
라. 가명정보 제3자 제공 금지
마. 가명정보 안전성 확보조치 준수
바. 가명정보 위탁 제한 (금지는 아님)
사. 계약사항 위반시 손해배상 등 책임에 관한 사항
아. 가명정보 목적 및 기간 달성 시 즉시 파기에 관한 사항
자. 파기 후 파기내역 통보
차. 가명정보 제공자가 파기 요청시 즉시 파기 준수에 관한 사항(재식별 사고 및 계약 위반 사항에 한함)

○ **(가명정보 관리대장 기록·관리)** 개인정보처리자는 가명정보를 처리(추가정보 생성, 파기 등)하는 경우 가명정보 관리대장을 기록 관리(별지 3)하고, 년 1회 이상 점검

※ 추가정보는 가명정보 생성 즉시 파기하는 것을 원칙으로 하며 가명정보 생성 즉시 파기한 내역은 별도의 파기관리대장에 관리하지 않으나, 가명정보 관리대장의 비고란에 추가정보 파기를 기록함으로 파기대장을 대신함

□ 기술적 보호조치

○ **(추가정보의 분리 보관)** 가명정보 처리 시 생성되는 추가정보는 삭제하는 것을 원칙으로 하되, 불가피한 사유가 있는 경우 원본 개인정보, 가명정보와는 물리적으로 분리하여 보관

※ 추가정보를 물리적으로 분리하기 어려운 경우 DB 테이블 분리 등 논리적으로 분리하는 것도 가능하나 엄격한 접근권한 관리 및 접근통제가 적용되어야 함

○ **(접근 권한 관리)** 개인정보처리자는 가명정보 또는 추가정보에 접근할 수 있는 담당자를 가명정보 처리 업무 목적 달성에 필요한 최소한의 인원으로 지정

- 접근권한은 업무에 따라 차등부여 등

- 추가정보에 관한 보호조치를 강화하기 위해 법률상의 의무는 아니나 원본정보, 가명정보, 추가정보의 담당자를 각각 분리하여 운영하는 방안도 고려할 수 있음
- 가명처리를 수행하는 자, 가명정보 적정성을 검토하는 자, 가명정보 취급자는 관리적 또는 기술적으로 권한을 분리하여 운영하여야 함

○ **(접속기록의 보관 및 점검)** 개인정보처리자는 가명정보 또는 추가정보 처리에 관한 접속 기록을 최소 1년 이상 보관·관리

※ 가명정보가 5만명 이상일 경우 최소 2년 이상 보관

- 접속기록에는 가명정보취급자가 가명정보처리시스템에 접속하여 처리한 업무내역을 알 수 있도록 주요 정보 등 모두 포함하고 월 1회 이상 주기적으로 확인

항목	접속기록 내용
계정	• 가명정보처리시스템에서 접속자를 식별할 수 있도록 부여된 ID 등 계정 정보
접속일시	• 접속한 시간 또는 업무를 수행한 시간
접속지 정보	• 가명정보처리시스템에 접속한 자의 컴퓨터 또는 서버의IP 주소 등
처리한 정보주체 정보	• 가명정보취급자가 어떠한 가명정보를 처리하였는지를 알 수 있는 식별정보 (가명정보 ID, 일련번호 등) • 가명정보는 일반적으로 대량의 정보를 처리하는 경우가 많으므로 대량의 정보를 처리시 해당 검색조건문(쿼리)을 정보주체 정보로 기록이 가능 • 가명정보의 특성상 추가정보의 사용 없이는 정보주체 식별이 불가능하므로 본 항목에는 실제 정보주체의 정보가 아니라 어떠한 가명정보를 처리했는지를 추적할 수 있는 정보를 기록할 수 있음
수행업무	• 가명정보취급자가 가명정보처리시스템을 이용하여 가명정보를 처리한 내용을 알 수 있는 정보(조회, 입력, 수정, 삭제, 다운로드, 출력 등)를 기록 • 가명정보의 재식별 행위를 파악할 수 있는 내용을 기록할 필요가 있음

[표 7] 접속기록 포함 내용 예시

□ 물리적 보호조치

○ **(가명정보 및 추가정보에 대한 출입통제)** 개인정보처리자는 가명정보 또는 추가정보의 안전한 관리를 위하여 물리적 안전조치 수행

- 전산실이나 자료보관실 등에 보관하는 경우 비인가자의 접근으로부터 보호하기 위해 출입통제 등의 절차를 수립 시행
- 가명정보 또는 추가정보가 물리적 공간을 이동하는 경우 이를 고려하여 내부관리계획에서 관리 절차를 수립 시행

※ 가명정보 공간으로의 기타 정보 반입, 기타 공간으로 가명정보의 반출, 가명정보 공간 간의 가명정보 이동 등

- 가명정보 또는 추가정보가 보조저장매체 등에 저장되어 있는 경우 잠금장치가 있는 안전한 장소에 보관하여야 하며, 이러한 보조저장매체 등에 대한 반·출입 통제를 위한 보안대책 마련

제 Ⅲ 편

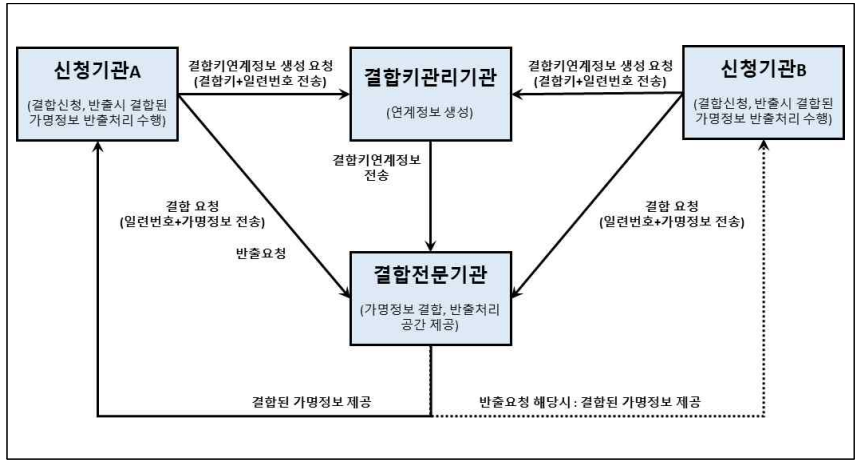
가명정보 결합

1. 가명처리 결합 · 반출
2. (참고) 가명정보 내부 결합

1 가명처리 결합·반출

□ 가명정보 결합 · 반출 절차

○ 서로 다른 개인정보처리자간의 교육분야 가명정보의 결합은 보호위원회 또는 교육부 장관이 지정하는 결합전문기관을 통해 수행



<그림 8> 결합전문기관을 통한 가명정보 결합 개념도

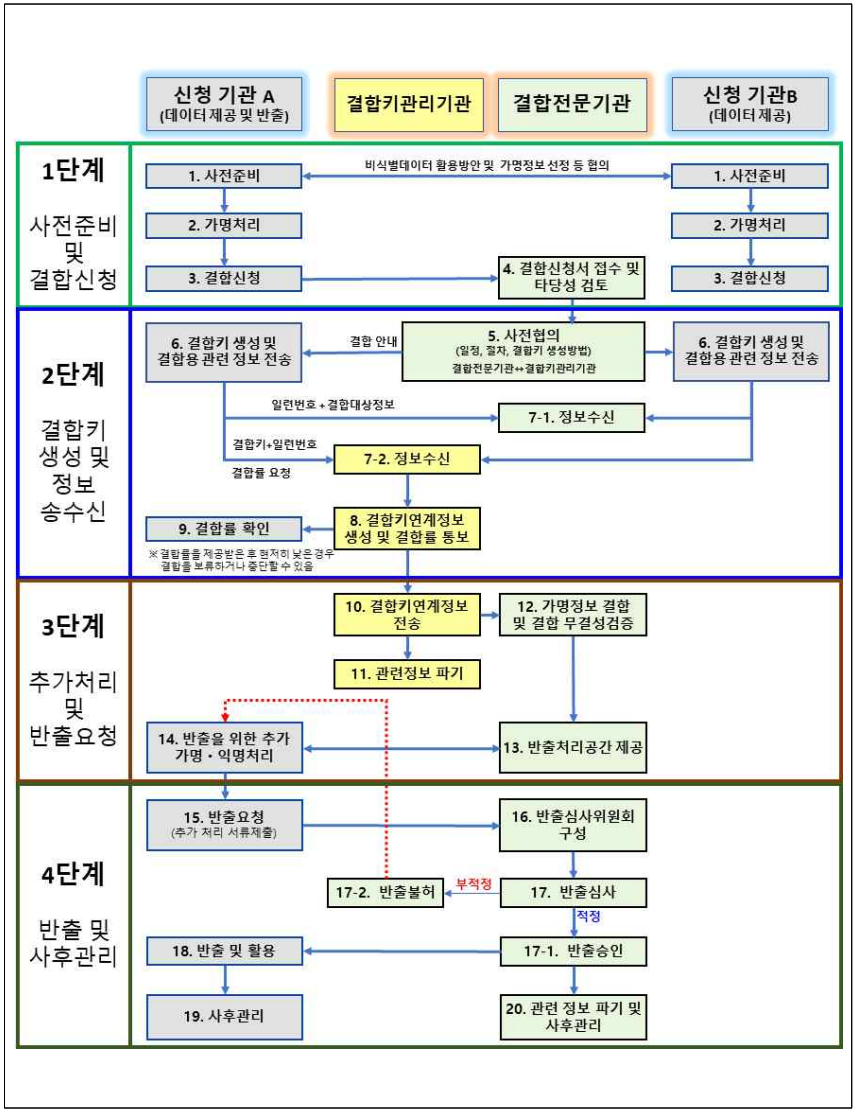
[가명정보 결합 기관 및 대상 예시]

서울시교육청이 등하교 시간과 학업성취도와의 상관관계를 연구하여 그 결과를 바탕으로 학생 배치, 노선 변경 요구 등을 진행하려는 경우

☞ 서울시교육청은 서울시교통공단이 보유한 교통카드 기록과 가명정보 결합

- 결합신청기관 : 서울시교육청(주관기관), 서울시교통공단(제공기관)
- 결합대상 : 교육청(학생성적), 교통공단(교통카드 기록)

□ 가명정보 결합 · 반출 세부 절차



<그림 9> 결합전문기관을 통한 가명정보 결합 절차 예시

[단계1 : 사전준비 및 결합신청]

1. 사전준비

○ 결합신청자(기관) A와 B는 상호 협의*를 통해 가명정보 결합에 대한 사전준비 수행

* 결합 목적에 부합하는 가명정보 선정 및 시계열 결합 여부 확인, 결합키 생성 항목 선정, 목적을 달성하기 위해 필요한 가명처리 수준 등

○ 결합신청자(기관)들은 공통으로 보유하고 있는 정보 중에서 결합키를 생성할 때 활용할 항목을 결정해야 함

[결합키 생성 항목 정의 예시]

- 신청자(기관)A : 성명, 휴대전화번호, 성별, 성적
- 신청자(기관)B : 성명, 휴대전화번호, 교통수단, 이동시간

☞ A와 B가 동일하게 가지고 있는 성명, 휴대전화번호를 결합키 항목으로 선정

※ 결합키 생성 시 주의사항

- 결합키 생성에 사용되는 항목들에 한글이 포함되어 있는 경우 동일한 인코딩 방식을 사용하는지를 확인해야 함
  - 동일한 결합키가 생성되는지 확인하기 위해 각자 동일한 데이터에 동일한 알고리즘을 적용하여 암호화 처리 후 결과 값을 확인해야 함
    ※ EUC-KR로 인코딩 된 데이터와 UTF-8로 인코딩된 데이터를 동일한 일방향 암호화하는 경우 서로 다른 결과 값이 생성될 수 있음
  - ☞ 결합률 0% 발생
- 데이터의 현행화 여부를 확인하여 상호 현행화 할 수 있도록 해야 함
  - ☞ 결합률 저하 발생

2. 가명처리

○ 결합신청자(기관)는 결합대상정보에서 정보주체별로 중복되지 않는 일련의 값(일련번호) 생성

[일련번호 생성방법 예시]

신청자(기관)A			
일련번호	성명	전화번호	…
A1	홍길동	090-2254-4565	
A2	김철수	090-1511-2545	
A3	이영희	090-5354-9178	
…	…	…	…

신청자(기관)B			
일련번호	성명	전화번호	…
B1	이영희	090-5354-9178	
B2	홍길동	090-2254-4565	
B3	김철수	090-1511-2545	
…	…	…	…

○ 결합신청자(기관)는 일련번호와 결합키 생성을 위한 항목을 제외한 나머지 정보들에 대해 가명처리 수행

[가명처리 대상 항목 선정 예시]

- 신청자(기관)A : 성명, 휴대전화번호, 성별, 성적
- 신청자(기관)B : 성명, 휴대전화번호, 교통수단, 이동시간

☞ A와 B가 동일하게 가지고 있는 성명, 휴대전화번호를 결합키 항목으로 선정하고 다른 항목인 성별, 성적, 교통수단, 이동시간 등이 가명처리 대상

※ 교통수단, 성별은 개인식별 가능성이 현저히 낮은 항목으로 가명처리 대상에서 제외 가능

- 가명처리 시 가명정보 처리환경을 고려하여 가명처리 수준 결정
- 결합신청자(기관)는 가명처리 수행에 대한 처리결과를 기록·보관

3. 결합신청

○ 결합신청자(기관)는 '가명정보 결합신청서(별지 5)'를 각자 작성하고, 관련 첨부서류를 함께 준비 후 교육부 장관이 지정한 결합전문기관에 결합 신청서를 제출

※ 작성 방법은 '가명정보 결합신청서 작성 예시(별지 6)' 참조

4. 결합신청서 접수 및 보완요청(결합전문기관)

○ **(결합전문기관)** 결합신청자(기관)의 신청을 접수하고 결합에 대한 타당성 검토를 실시한 후, 신청서와 첨부서류를 확인하여 미흡한 경우 부족한 서류 등에 대해 결합신청기관에 보완 요청

○ **(결합신청자)** 결합전문기관이 보완 요청한 서류 등에 대해 추가 제출하고 결합전문기관이 가명정보의 가명처리 수준에 대하여 권고하는 경우, 특별한 사정이 없으면 이를 반영

[단계2 : 결합키 생성 및 정보전달]

5. 사전협의(결합신청자, 결합키관리기관)

○ 결합전문기관 및 결합키관리기관은 결합신청자(기관)와 결합 및 결합키생성에 관한 사항에 대해 사전 협의 가능

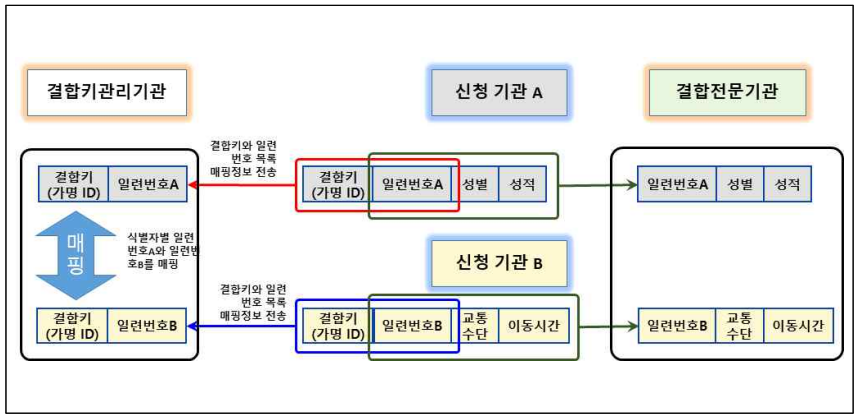
※ (결합전문기관) 결합일정 및 절차, 전송 파일 형태 등 결합에 필요한 사항, (결합키관리기관) 결합키생성 관련 기술지원, 사전 결합률 확인 여부 등

6. 결합키 생성 및 결합용 관련 정보 전송

○ **(결합키 생성)** 결합신청자(기관)는 사전준비 단계에서 합의한 항목과 결합키관리기관과 합의한 사항 등을 적용하여 결합키 생성

○ **(정보 전송)** 결합신청자(기관)는 결합키관리기관과 결합전문기관에 필요한 정보를 전송

- 생성한 결합키와 일련번호를 매핑한 정보를 결합키관리기관에 전송
- 결합전문기관이 제공하는 전송방법을 활용하여 결합전문기관에 결합대상정보와 일련번호를 전송(오프라인 포함)



<그림 10> 결합키 생성 및 결합용 관련 정보 전송 예시

- 결합신청자(기관)는 결합에 필요한 정보를 암호화 등 안전한 방법으로 결합전문기관 및 결합키관리기관에 전송

※ 정보통신망을 통하여 송·수신하거나 보조저장매체 등을 통해 전송하는 경우, 압축 및 암호화, 무결성 검증 등 보호조치 수행

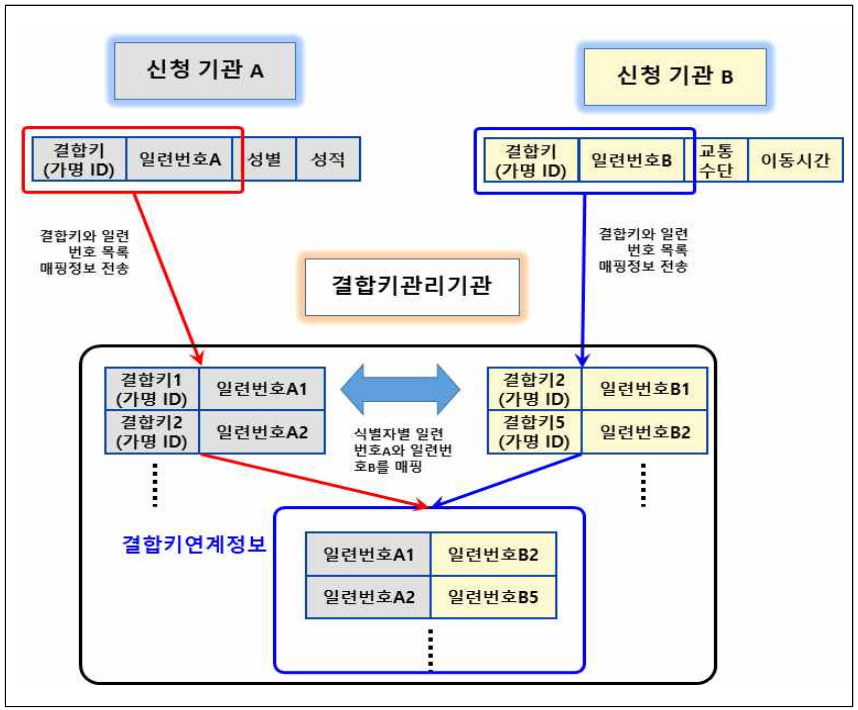
7. 정보 수신(결합전문기관, 결합키관리기관)

○ 결합전문기관과 결합키관리기관은 결합신청자(기관)이 전송한 정보 수신

8. 결합키연계정보 생성 및 결합률 통보(결합키관리기관)

○ 결합키관리기관은 결합신청자(기관)로부터 받은 '결합키+일련번호'를 기반으로 결합키연계정보 생성

○ 결합신청자(기관)가 결합키관리기관과의 협의 단계에서 사전 결합률 확인을 요청한 경우, 결합키관리기관은 결합키연계정보 생성과정에서 결합률을 확인하여 통보



<그림 11> 결합키연계정보 생성

9. 결합률 확인

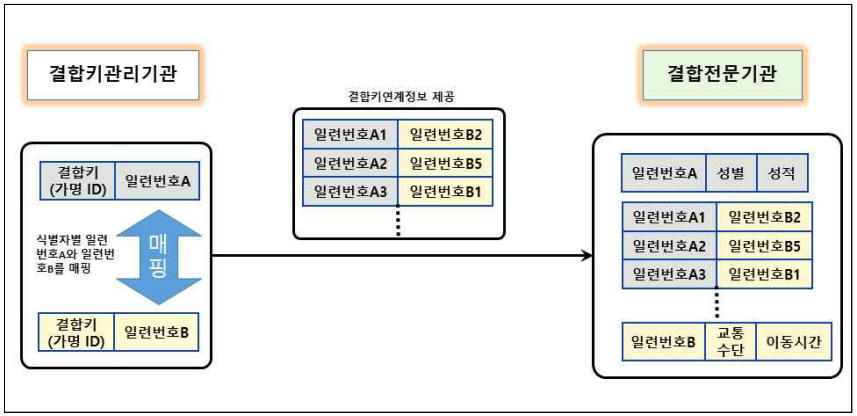
○ 결합신청자(기관)는 결합률을 확인한 후 필요한 경우 결합키를 재생성하거나 결합신청 자체 중단 가능

- 결합키에 문제(결합률 0% 등)가 있다고 판단하거나 효율성이 높은 가명정보 결합을 위해 결합키 생성 항목을 변경하는 경우 등

○ 단순히 결합키 생성 항목만 변경하는 경우 결합신청 절차를 생략할 수 있으나, 결합신청 내용의 변경(결합 목적, 항목수, 레코드 수 등)이 있는 경우 결합신청을 다시 해야 함

10. 결합키연계정보 전송(결합키관리기관)

○ 결합전문기관에게 결합을 위한 결합키연계정보 전송



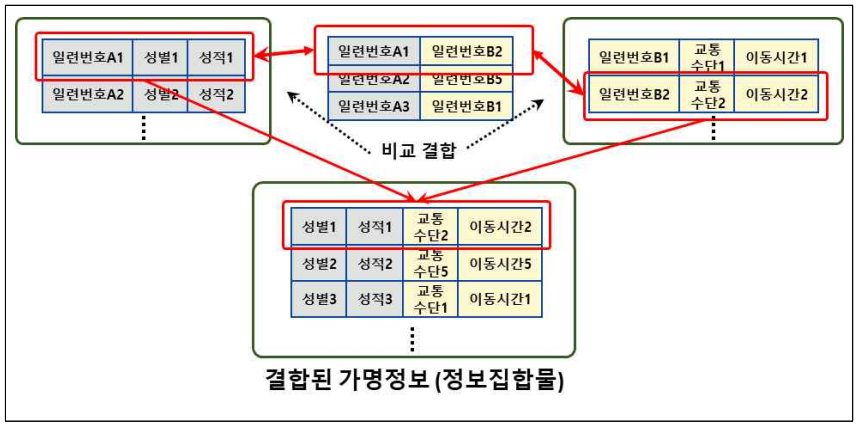
<그림 12> 결합키연계정보 전송 예시

11. 관련정보 파기(결합키관리기관)

○ 결합키연계정보가 결합완료(반출이후) 등 더 이상 불필요한 경우 즉시 파기

12. 가명정보 결합 및 결합 무결성 검증(결합전문기관)

○ 결합신청자(기관)로부터 제공받은 가명정보를 결합키관리기관으로부터 받은 결합키연계정보와 비교하여 결합을 수행하고, 결합 후에는 결합된 가명정보에 대한 무결성 등 결합에 대한 검증 실시



<그림 13> 결합키연계정보를 이용한 가명정보 결합 예시

[단계3 : 추가처리 및 반출심사 요청]

13. 반출처리공간 제공(결합전문기관)

○ 결합신청자(기관)가 결합된 가명정보 반출을 위해 추가적인 가명처리 및 익명처리를 할 수 있도록 반출처리공간 제공

※ 결합된 정보를 분석할 수 있는 시설, 장비 등을 갖춘 범위 내에서 결합된 정보를 결합전문기관 내에서 분석하고 분석결과물만 반출할 수 있음

14. 반출을 위한 추가 가명·익명처리

○ 결합신청자(기관)는 결합된 정보를 결합전문기관의 내에 설치된 반출처리공간에서 반출을 위한 추가 가명·익명처리를 해야 하며, 결합전문기관에 자문 등 필요한 지원 요청 가능

- 이때, 추가 가명처리를 할 수 있는 자는 결합신청자 중에서 결합된 정보를 반출하여 활용하고자 하는 자로 제한됨

○ 반출심사는 결합된 정보를 반출할 자의 처리 목적과 처리환경 등을 고려하여 이루어지므로, 추가 가명·익명처리 시 처리 수준은 결합된 정보를 활용하고자 하는 자의 처리환경 등을 고려하여 판단

○ 익명정보로도 목적 달성이 가능한 경우, 익명처리하여 반출

15. 반출요청

○ 결합신청자(기관)는 반출을 위한 추가처리가 완료되었다고 판단한 경우 결합전문기관에서 정한 규정에 따라 반출신청서를 작성하여 반출 요청

- 반출 요구 시 결합신청자는 반출 요구 조건에 충족하도록 추가적으로 수행한 가명·익명처리 등 관련 서류 제출

※ 신청자(기관)B도 반출을 요청하는 경우 <그림 15>에서 신청자(기관)A의 절차를 수행해야 함

- 반출심사는 결합신청기관별로 이루어지므로 추가처리가 완료되는 즉시 개별적으로 결합전문기관에 반출을 요청

16. 반출심사위원회 구성(결합전문기관)

○ 결합전문기관은 결합신청자(기관)가 반출을 요청하는 경우 지체 없이 반출심사위원회를 구성하여 반출 여부를 심사

- 반출심사위원회 구성은 개인정보 보호에 관하여 경험과 학식이 풍부한 사람으로서 다음 각 호에 해당하는 사람을 고루 포함하여 3인 이상 7인 이내의 위원으로 구성

1. 개인정보 보호와 관련한 업무 경력이 있거나 관련 단체로부터 추천을 받은 사람
2. 개인정보처리자로 구성된 단체에서 활동한 경력이 있거나 관련 단체로부터 추천을 받은 사람
3. 그 밖에 개인정보 보호와 관련한 경력과 전문성이 있는 사람

- 결합신청자(기관)는 반출요청을 한 경우 심사에 필요한 사항을 제출하고 이를 설명해야 함

- 결합신청기관의 임직원 등 이해관계자는 반출심사위원회 위원으로 참여 불가

17. 반출심사 : 반출 승인과 반출불허(결합전문기관)

○ 결합신청자(기관)가 반출심사를 위해 필요하다고 판단하거나 반출심사위원회의 요청이 있는 경우, 추가자료를 제출하거나 관련된 내용을 반출심사위원회에 출석하여 설명해야 함

- 결합신청자(기관)가 결합에 사용된 가명정보를 제공한 개인정보처리자인 경우 해당 정보와의 결합가능성 등이 반출 승인 과정에서 고려될 수 있음

- 심의를 득하지 못한 경우 결합신청자(기관)와 협의 하에 재처리를 요구하거나 반출심의 종결

※ 이 경우에도 반출승인을 위하여 결합전문기관에 자문 등 필요한 지원을 요청할 수 있음

[단계4 : 반출 및 사후관리]

18. 반출 및 활용

○ 결합신청자(기관)는 반출된 정보를 결합 목적에 맞게 활용 가능

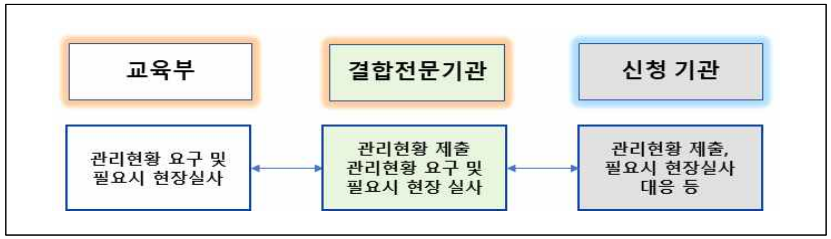
19. 사후관리

○ 결합신청자(기관)는 활용에 따른 자체 모니터링 및 사후관리에 대해 본 가이드라인 'Ⅱ편 가명처리 > 2-4. 가명처리 사후관리'에서 기술하고 있는 내용에 따라 관리를 해야 함

20. 관련정보 파기 및 사후관리(결합전문기관)

○ **(결합 이력 관리)** 결합이 종료되면 관련 이력을 관리하고 법령에서 별도로 보존하도록 한 정보를 제외하고 즉시 삭제처리(결합키 등)

○ **(사후관리)** 결합신청자(기관)에 대한 사후관리 및 감독을 실시하며, 이 경우 사후관리 및 감독을 위해 결합전문기관이 요청한 자료 등에 대해 결합신청자(기관)는 특별한 사유가 없으면 응해야 함



<그림 14> 결합전문기관의 관리현황 조사요구와 대응

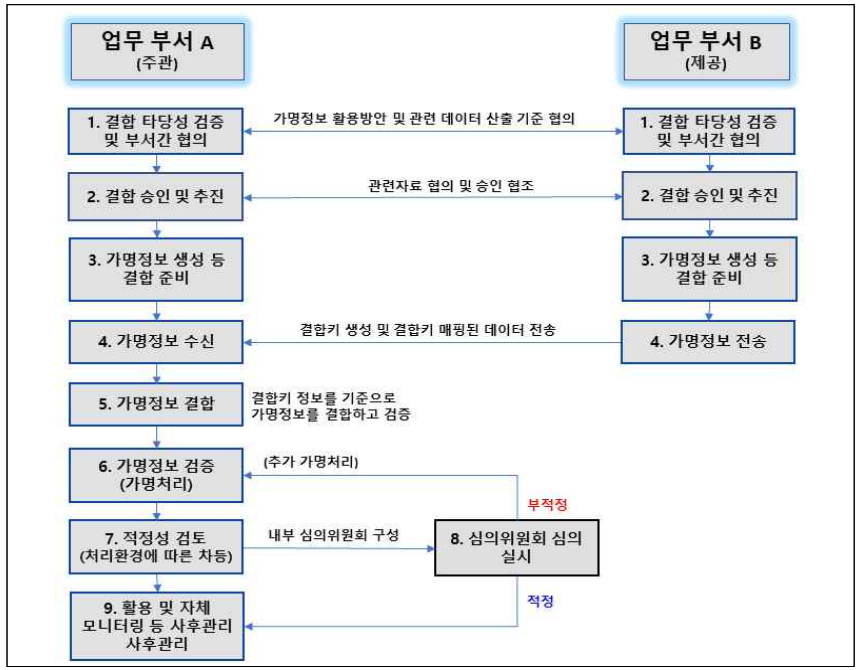
2 (참고) 가명정보 내부 결합

○ 개인정보처리자가 보유한 개인정보를 내부에서 가명처리하여 결합을 하는 경우 별도의 외부 결합전문기관을 거치지 않음

- 결합을 수행할 부서에서 결합할 가명정보를 모두 제공받아 결합키를 이용하여 자체적으로 결합 수행

○ 안전한 결합을 위하여 결합전문기관의 결합절차(Ⅲ편 1. 가명처리 결합·반출)와 유사하게 처리하는 것을 권고

- 내부 결합에 대하여는 법령에서 별도로 정하고 있지 않지만, 결합 과정에서 가명정보가 재식별되지 않도록 유의



<그림 15> 자체 가명정보 결합 절차 예시

○ 결합용 가명정보를 추출한 부서는 주관부서로 해당 가명정보를 전송(이동)하고, 전송 시 가명정보 안전성 확보조치 기준에 따른 안전성 확보

○ 주관부서는 결합키*를 기반으로 서로 다른 부서에서 받은 가명정보를 주관부서의 가명정보와 결합을 실시하고 결합된 가명정보에 대한 결합 무결성 검증 가능

* 결합방식에 따라 결합키 단독으로 결합을 하거나 결합키와 일련번호를 매핑한 결합키연계정보를 활용하여 결합할 수 있으며 결합의 방식은 기관의 선택에 따라 달리 할 수 있음

- 가명정보 결합과 관련된 정보(결합키 등)는 가명정보의 결합 및 적정성 검토가 완료되는 즉시 삭제
- 가명정보처리자는 결합된 가명정보를 처리하고자 하는 경우 별도로 안전한 장소에서 최소한의 권한이 부여된 최소한의 인원을 지정하여 운영
- 제한구역 이상의 안전한 지역에서 가명정보를 처리

※ 결합된 가명정보와 원본 개인정보를 동시에 취급하는 자가 없도록 권한 및 직무분리

○ 결합한 정보에 대해 결합된 가명정보의 처리 수준이 적정한지 판단하고 필요한 경우 추가 가명처리 수행 가능

제 Ⅳ 편

기타

부록 1 익명처리

❖ 익명정보는 더 이상 개인정보로 취급하지 않기 때문에 개인정보 보호법 등 관련 법령의 제한을 받지 않고 자유롭게 활용 가능

※ 단, 익명정보 활용을 위해서는 보다 명확하고 엄격한 처리와 객관적인 검증이 요구됨

□ 익명처리 원칙

○ 개인식별정보 및 개인식별 가능성이 있는 정보는 원칙적으로 삭제하도록 하되, 데이터 이용 목적상 꼭 필요한 경우 안전한 방식으로 익명처리해야 함

○ 익명처리 업무를 수행하는 자는 익명처리 대상 개인정보를 처리하는 업무 수행을 금지. 다만, 불가피한 사유가 있을 경우 보완통제 대책을 수립하여 관리자의 승인 하에 제한적으로 취급 가능

○ 개인정보처리자는 익명정보 적정성 검토를 수행하는 경우 가명정보 적정성 검토위원회와 동일한 수준으로 구성 운영

○ 익명정보처리자는 익명정보 관리대장을 기록·관리하고 개인정보처리자는 년 1회 이상 점검(해당 시)

※ 익명정보 관리대장에 포함되어야 할 사항

가. 익명처리한 날짜 나. 익명정보의 항목 다. 익명처리한 사유와 근거 라. 익명정보를 제3자 제공한 경우 제공받은 자와 제한사항

○ 개인정보가 포함 된 공공데이터는 특정 개인을 식별할 수 있는 요소를 삭제하거나 비식별 조치(익명처리 및 적정성 검토) 후 개방

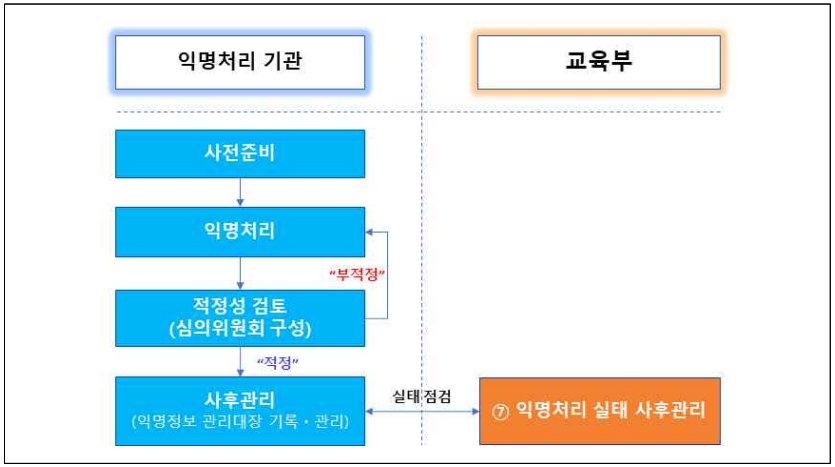
- 원칙적으로 그 자체로 개인을 식별할 수 있는 정보는 삭제
- 개방대상 정보에 이미 공개된 정보 등과 결합하여 개인 식별이 가능한 정보가 포함되어 있는지 여부 등을 사전 필터링
- 이미 개방한 데이터가 다른 정보와 결합하여 개인 식별이 가능한지 여부 등을 주기적으로 모니터링한 후 재식별이 되는 경우 해당 개인정보 삭제 또는 익명처리

개인정보 재식별 위험 주의 !

- 개인정보는 가명·익명처리하여 익명정보를 만들었다고 하여도 시간이 지남에 따라 다른 정보와 결합하여 개인정보가 재식별 될 수 있음
- 시간이 지남에 따라 결합용이성과 입수가능성이 증가하여 유형별 재식별 위험이 증가 (ISO/IEC 20889에 따른 분류 위험)

유형 구분	재식별 위험				
	식별가능성	복원가능성	특정가능성	추론가능성	연결가능성
가명정보	없음	없음	존재	존재	존재
익명정보	없음	없음	없음	없음	없음

□ 익명처리 절차 개념도



부록 1-<그림 1> 익명처리 절차

○ **(사전준비)** 익명처리 목적을 명확히 정의하고 익명처리 대상 개인정보 선정

○ **(익명처리)** 익명처리 수준을 정의하고 수준에 맞도록 익명처리 기법을 활용하여 개인정보를 익명처리

○ **(적정성 검토)** 익명처리 수준 등에 맞는 익명처리가 되었는지 여부 및 익명정보 내 개인정보 식별 여부 및 재식별 가능성 검토

○ **(사후관리)** 익명처리된 익명정보에 대해 기록·관리하고 재식별 발생 시 대응할 수 있도록 대책 수립 가능

○ **(익명처리 실태 사후관리)** 교육부는 익명처리 기록·관리 등에 관한 사항을 조사·점검할 수 있으며 미흡한 사항에 대해 해당 기관에 시정조치 요구 가능

□ 익명처리 예시

[개인정보와 속성자 특징]

성명	연락처	성별	생년월일	혈액형		전공	학위
				ABO	RH		
김영희	090-1234-5678	여	19650512	A	Rh+	법학	박사
강순희	090-8525-4564	남	19671212	B	Rh+	컴퓨터	학사
최복례	090-8546-5456	여	19681015	O	Rh+	건축	학사
홍길동	090-5524-1325	남	19920721	AB	Rh-	보안	학사
이홍준	090-6974-1235	남	19930423	AB	Rh+	교육	학사
김신우	08-3456-7890	남	19940925	O	Rh+	음악	학사
〃	〃	〃	〃	〃	〃	〃	〃

[익명처리 예시]

성별	혈액형		전공	학위	익명성
여	A	Rh+	법학	학사	동질집합 k = 100 이상
여	A	Rh+	법학	학사	
〃	〃	〃	〃	〃	
여	B	Rh+	법학	학사	〃
여	AB	Rh+	교육	학사	〃
여	O	Rh+	음악	학사	〃
여	A	Rh-	법학	학사	〃
〃	〃	〃	〃	〃	〃

성별	혈액형		전공	학위	통계
여	A		법학	학사	550
여	A		법학	석사	310
〃	〃		〃	〃	〃
여	B		법학	학사	251
여	AB		교육	석사	180
여	O		음악	박사	115
여	A	Rh-	법학	학사	151
〃	〃		〃	〃	〃

· 성별/혈액형별 전공과 학위 현황의 익명정보를 만들겠다는 목적으로 프라이버시 보호모델 k-익명성* 값은 100 이상으로 설정

* k-익명성이란 동일한 속성 또는 속성의 조합을 가지는 레코드가 최소한 k개 이상 존재하도록 하여 프라이버시를 보호하는 모델을 말함

· '성명', '연락처'는 특정 개인이 식별 가능하므로 삭제 처리

· '성별'은 활용 목적에 필요한 속성자이므로 유지

· '생년월일'은 활용 목적에 필요한 속성자가 아니므로 삭제

· '혈액형'은 활용 목적에 필요한 속성자이므로 유지. 다만, Rh-의 경우 희귀 혈액형 유형으로 개인이 식별될 가능성이 있는 경우 k-익명성에서의 k값은 100 이상이 나오도록 처리하되 100 미만인 경우 해당 레코드 삭제 (k값 100은 예시이며 상황에 따라 가감될 수 있음)

※ Rh-를 모두 레코드 삭제를 하게되는 상황이 발생하면 Rh+만 남게되고 이 경우 사실 Rh의 분류는 의미가 없어짐. 따라서 활용 목적에 따라 이런 경우 Rh 항목에 대해 전체 삭제하는 것으로 일부 레코드 삭제를 방지할 수 있음

· '전공' '학위'는 결과 도출을 위한 속성 정보이므로 유지. 다만, 개인이 식별될 가능성이 있는 경우 k-익명성에서의 k값은 100 이상이 나오도록 범주화하여 통계처리

부록 2

주요 산출물 및 처리방안

□ 가명처리

단계		내용	산출물	처리방안
1단계	목적 및 대상	가명처리 목적 및 가명처리 대상 개인정보 선정	사업계획서	기록물 관리에 따른 보존기관내 보관
	승인 준비	사업 설명자료 작성, 내부 회의 및 사업 추진 타당성 검토, 사업 승인 추진		
	정보집합물 구성	개인정보 항목 선정 및 추출하여 정보집합물 구성	정보집합물	가명처리 후 삭제
	관리체계 구축	가명정보 보호대책을 수립	내부관리계획	준영구 보관
2단계	가명처리 대상 재선정	정보집합물에서 처리대상 및 대상별 특성을 확인하여 최소한의 항목을 추출	가명처리 대상 목록 (또는 정보집합물)	가명처리 후 삭제
	위험도 측정 및 검토보고서 작성	위험도 측정 및 위험도에 따른 가명처리 검토 결과 보고서 작성	가명처리 검토 결과 보고서 (별지 1)	가명정보 이용기간 종료 후 3년간 보관
	가명처리 수준 정의 및 가명처리 수행	가명처리 검토 결과 보고서 기반으로 가명처리 수준 정의표 작성, 가명처리 수준 정의표에 따른 항목별 가명처리 수행	가명처리 수준정의표 (별지 2)	
			가명정보	기간 종료 후 삭제
			추가정보	즉시 삭제 권고 (보관시 가명정보와 같이 삭제)
	보완조치	특이정보 등 재식별 가능성	가명처리	가명정보

단계		내용	산출물	처리방안
		확인시 추가적으로 가명처리 (보완조치 수행 시 가명처리 수준 정의표 수정 반영)	수준정의표	이용기간 종료 후 3년간 보관
3 단계	적정성 검토 심의위원회 구성	심의위원회 구성(3~7명)안을 포함한 적정성 검토 계획 수립	적정성 검토 계획서(위원회 구성안 포함)	기록물 관리에 따른 보존기관내 보관
	심의위원회 운영	적정성 검토를 위한 기초자료 준비 및 적정성 검토	(기초자료)	기초자료가 포함된 각 산출물 처리방안 참고
	적정성 검토 결과보고서 작성	적정성 검토 결과에 대한 결과보고서 작성	적정성 검토 결과보고서 (자유양식)	가명정보 이용기간 종료 후 3년간 보관
	보완조치	적정성 검토 결과 부적정 판정 시 가명처리로 반송 조치		
4 단계	가명정보 처리 기록·관리	가명정보 관리대장 기록·관리	가명정보 관리대장 (별지 3)	준영구 보관
	가명정보 안전성 확보조치 수행	가명정보 등에 대한 안전성 확보 방안 수립 및 운영 (개인정보처리방침 공개, 내부관리계획 수립 등)	개인정보처리 방침	준영구 보관
			내부관리계획	준영구 보관
			계약서 (해당시)	법령에 따른 보존기관
	재식별 모니터링	가명정보를 처리 시 년 1회 이상 개인정보 재식별 여부 점검 수행	재식별 모니터링 (자유양식)	가명정보 이용기간 종료 후 3년간 보관
	가명정보 회수 및 처리 중단	개인정보 재식별된 경우 해당 가명정보 회수, 처리 중단 및 즉시 파기	가명정보 파기 계획 및 결과보고서	기록물 관리에 따른 보존기관내 보관

[표 8] 가명처리 단계별 산출물 및 처리방안 예시

□ 가명정보 내부결합

단계		내용	산출물	처리방안
1	결합 타당성 검증 및 부서간 협의	내부 회의 및 사업 추진 타당성 검토, 가명처리 목적 및 가명처리 대상 개인정보 선정	사업계획서 (결합키 생성 방안 등 포함)	기록물 관리에 따른 보존기관내 보관
2	결합 승인 및 추진	사업 설명자료 작성, 사업 승인 추진		
3	가명정보 생성 등 결합 준비	개인정보 항목 선정·추출	결합키	결합완료 후 삭제
		가명처리하여 가명정보 구성 (가명처리 단계의 모든 산출물 필요함. 단, 적정성 평가는 7번 및 8번에서 수행함으로 생략 가능)	가명정보 등	관련 산출물별 처리방안 참조
		가명정보 보호대책을 수립 (내부관리계획 포함 가능)	내부관리계획	준영구 보관
4	가명정보 전송 및 수신	제공부서는 주관부서로 가명정보를 전송하고 주관부서는 해당 가명정보 수신	가명처리 대상 목록 (또는 정보집합물)	결합완료 후 삭제
5	가명정보 결합	결합키 정보를 기준으로 둘 이상의 가명정보 결합 수행	결합된 가명정보	가명정보 이용 및 보존기간 종료 후 삭제
6	가명정보 검증	특이정보 등 재식별 가능성 발생 시 추가적으로 가명처리 (심의위원회 부적정 판단 시 추가 가명처리)	가명처리 수준정의표 (별지 2)	가명정보 이용기간 종료 후 3년간 보관
7	적정성 검토 (심의위원회 구성)	심의위원회 구성(3~7명)안을 포함한 적정성 검토 계획 수립 및 심의위원회 구성	적정성 검토 계획서(위원회 구성안 포함)	기록물 관리에 따른 보존기관내 보관
8	심의위원회 운영	적정성 검토 결과에 대한 결과보고서 작성	적정성 검토 결과보고서 (자유양식)	가명정보 이용기간 종료 후 3년간 보관
		적정성 검토 결과 부적정 판정 시 가명처리로 반송 조치		

단계		내용	산출물	처리방안
9	활용 및 사후관리	활용 및 가명정보 관리대장 기록·관리	가명정보 관리대장 (별지 3)	준영구 보관
		가명정보 등에 대한 안전성 확보 방안 수립 및 운영 (개인정보처리방침 공개, 내부관리계획 수립 등)	개인정보처리 방침	준영구 보관
			내부관리계획	준영구 보관
		가명정보를 처리 시 년 1회 이상 개인정보 재식별 여부 점검 수행	재식별 모니터링 (자유양식)	가명정보 이용기간 종료 후 3년간 보관
		개인정보 재식별된 경우 해당 가명정보 회수, 처리 중단 및 즉시 파기	가명정보 파기 계획 및 결과보고서	기록물 관리에 따른 보존기관내 보관

[표 9] 가명정보 내부결합 단계별 산출물 및 처리방안 예시

※ 위의 가명정보 처리 및 내부결합에 따른 산출물 및 처리방안은 예시로 제시한 것이며 개인정보처리자별 개인정보 처리 환경 및 상황 등에 따라 변경 활용

부록 3

가명처리 및 익명처리 관련 양식

(별지 1) 가명처리 검토 결과보고서 예시

가명정보 활용목적	◦ ○○교육청이 거주지별 학생정보와 성적 정보들을 가명처리하여 A기관에 제공하여, 코로나로 원격학습에 따른 지역별 학업 성취도 격차를 파악하기 위한 연구 수행	
가명정보 항목	◦ 학생개인번호, 과목, 학년, 성별, 성적(점수), 시도, 시군구, 읍면동, 지번, 거주지 유형 ※ 항목을 나열하지 못하는 경우 '별지'로 추가하여 사용 가능	
처리(제공) 환경 검토	처리 환경	◦ 특정 제3자(A기관) 제공 ◦ ○○교육청은 A기관과 교육분야 가명정보 처리 가이드라인에 따라 필요한 항목들이 모두 포함된 계약체결을 통해 가명정보를 제공
	제공 받는 자의 보호 수준	◦ A기관은 개인정보처리시스템에 대한 ISMS-P 인증을 취득하고 있으며, 새롭게 가명정보처리시스템에 대해 ISMS-P 인증을 진행하고 있어 보호수준이 상대적으로 높음 ◦ 내부관리를 통해 관리적·기술적·물리적 보호조치를 수행하고 있음
	재식별 수준	◦ 가명정보를 제공받은 A기관은 학생 관련 다른 (개인)정보들을 보유하고 있지 않음
항목별 위험도 분석	◦ 학생개인번호, 학생성적정보(과목, 학년, 성별, 성적)가 결합할 경우 식별가능정보 및 특이정보 가능성 존재 ◦ '지번'은 식별가능정보	
최종 검토의견*	◦ 해당 연구는 특정 제3자와의 계약서 체결을 통해 가명정보를 활용하는 경우에 해당하며, 제공받는 자가 별도의 다른 (개인)정보를 통해 가명정보를 재식별 할 가능성이 낮음 ◦ 학생개인번호는 그 자체로 또는 결합시 식별될 가능성이 매우 높으므로 반드시 필요하지 않은 경우 삭제하고 과목별, 학년별 성적 추적을 위해 필요한 경우 학생개인정보를 그대로 사용하지 않도록 가명처리 필요 ◦ 학생성적정보는 결합시 식별될 가능성이 높으므로 반드시 필요한 경우가 아니라면 특정 항목은 삭제 필요. 다만, 목적달성에 필요한 경우에 한하여 최소한 익명성(k) 값을 10 이상으로 하여 처리 ※ '성적(점수)'는 특이정보 가능성이 존재하므로 범주화 등의 가명처리가 필요 ◦ '지번'의 경우 다른(공개된 정보 등) 정보를 통해 재식별 가능성이 있어 삭제 또는 목적 달성에 필요한 경우 가명처리 필요 ◦ 그 외의 정보들은 재식별 가능성이 낮으며 목적달성을 위해 필요하다고 판단되므로 가명처리하지 않음	

* 최종 검토의견은 외부전문가를 활용하여 자문 및 작성을 요청할 수 있음

(별지 2) 가명처리 수준 정의표 예시

○ '가명처리 검토 결과보고서'에 분류한 개인정보에 대한 가명처리 수준 정의

순번	항목명	처리수준	비고
1	학생개인번호	◦ 가명처리 (암호화 : SHA2 + Salt)	◦ 학년별, 과목별 추적 분석 등을 위해 가명처리 수행
2	과목	◦ 처리하지 않음	◦ 처리하지 않는 항목을 작성 (과목, 성별 등이 결합하여 특이정보가 발생할 경우 삭제 등의 가명처리 필요)
3	학년		
4	성별		
5	성적(점수)	◦ 가명처리 (범주화 : 100점 만점 기준으로 2~5점 단위)	◦ 특이정보 발생시 범주화 기준 상향 필요
6	시도	◦ 처리하지 않음 (항목이 다수여서 작성이 어려운 경우 '별지'를 활용하여 목록만 제시)	◦ 처리하지 않는 항목을 작성
7	시군구		
8	읍면동		
9	거주지 유형		
10	지번	◦ 가명처리(삭제)	◦ 세부 지번의 정보는 분석 목적에 필요하지 않음

(별지 3) 가명정보 관리대장

순번	기간	목적	항목	처리 및 보유기간	제공받는 자 등	위탁 사항	제한사항	처리 구분	기타 (추가정보)	가명처리 수행자	책임자

1. 순번 : 가명정보 처리에 대한 순서를 작성한다.
2. 기간 : 가명처리를 시작한 날부터 끝난 날까지의 기간을 작성한다.
3. 목적 : 가명정보를 처리하는 목적을 작성한다.
4. 항목 : 처리하는 가명정보의 항목을 작성한다.
5. 처리 및 보유기간 (필요시) : 가명정보의 사용 및 보존기간을 작성한다.
   예시) 처리기간(2년), 보존기간(1년) or 처리기간(1년), 보존기간(없음) 등
6. 제공받는 자 등 : 제3자 제공시 제공받는 자의 정보를 작성한다.
   가명정보 결합시 결합대상 가명정보처리자의 정보를 작성한다.

7. 위탁사항 : 가명정보 처리를 위탁한 경우 해당 사항을 작성한다.
8. 제한사항 : 가명정보 처리 유형에 따라 제한사항을 둔 내용을 작성한다.
9. 처리구분 : 가명정보 처리 유형을 작성한다. 예시) 생성, 결합, 제공, 공개 등
10. 기타 : 추가정보의 경우 가명정보 생성즉시 파기한 경우 이 란에 '추가정보 동시 파기'를 작성한다. 기타 추가적인 사항에 대해 이 란에 작성한다.
11. 가명처리를 수행한 자 : 가명처리를 수행한 자의 이름을 작성하고 서명한다.
12. 책임자 : 가명정보 처리 책임자 이름을 작성하고 서명한다. (ex. 가명처리 부서의 책임자)

(별지 4) 익명정보 관리대장

순번	기간	항목	사유/근거	제공받은 자	제한사항	기타	익명처리 수행자	책임자

1. 순번 : 익명정보 처리내용에 대한 순서를 작성한다.

2. 기간 : 익명처리를 시작한 날부터 끝난 날까지의 기간을 작성한다.
4. 항목 : 익명정보의 항목을 작성한다.
5. 사유/근거 : 익명정보 처리 사유 및 근거를 작성한다.
6. 제공받은 자 : 익명정보를 제공한 경우 제공받은 자의 정보를 작성한다.(기업명[또는 성명), 담당자 연락처)
   예시) 처리기간(2년), 보존기간(1년) or 처리기간(1년), 보존기간(없음)
7. 제한사항 : 익명정보 처리 유형에 따라 제한사항을 둔 내용을 작성한다.(제3자 제공시 계약서 준수 여부 등)
8. 기타 : 익명정보 재식별시 파기 여부 등 추가적으로 필요한 내용을 작성한다.
9. 익명처리를 수행한 자 : 익명처리를 수행한 자의 이름을 작성하고 서명한다.
10. 책임자 : 익명처리 책임자의 이름을 작성하고 서명한다.

(별지 5) 가명정보 결합신청서

가명정보 결합 신청서

신청번호	

결합신청자			
기관명		사업자등록번호 또는 법인등록번호	
주소		대표자명	
담당자		담당자 연락처 (전화, e-mail)	
신청자 구분	[ ] 개인 [ ] 공공기관 [ ] 비영리법인 [ ] 민간기관		

가명정보 제공		해당없음 [ ]
파일명		
제공방법	[ ] 온라인 [ ] 오프라인	
제출예정일	년 월 일	
전체 가명정보 제공 기관명(총수)		
제공정보 요약	컬럼 수	
	전체 레코드 수	
	전체 파일 크기	

결합 결과물 이용		해당없음 [ ]
시계열 분석	[ ] 해당없음 [ ] 시계열(신규) [ ] 시계열(추가, 결합접수번호 :)	
결합목적	[ ] 통계작성 [ ] 과학적 연구 [ ] 공익적 기록보존 등	
세부결합목적		

「개인정보보호법」 제28조의3제1항 및 같은 법 시행령 제29조의3제1항에 따른 결합을 위하여 결합전문기관에 결합신청서를 위와 같이 제출합니다.

년 월 일

신청인 (서명 또는 인)

결합전문기관의 장 귀하

첨부 서류
1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부
2. 결합 대상 가명정보에 관한 서류(전체 항목명 가명처리 대상 항목명, 가명처리 기법 및 예시 등) 1부(해당 경우에 한함) * 결합키 생성에 사용된 속성 제외
3. 결합 목적을 증명할 수 있는 서류 1부
4. 안전조치계획 및 이를 증빙할 수 있는 서류 1부

(별지 6) 가명정보 결합신청서 작성 예시

개인 홍길동과 신청기관 A사와 신청기관 B사가 신청하는 경우

[개인정보처리자 개인 홍길동의 신청서 예시]

가명정보 결합 신청서

신청번호	A-2020-10-001

결합신청자			
기관명	해당없음	사업자등록번호 또는 법인등록번호	1976년 10월 10일
주소	서울시 강남구 강남대로 2길 ㅇㅇ	대표자명	해당없음
담당자	홍길동	담당자 연락처 (전화, e-mail)	010-123-1234 abc@cdefg.com
신청자 구분	[√] 개인 [ ] 공공기관 [ ] 비영리법인 [ ] 민간기관		

가명정보 제공		해당없음 [ ]
파일명	2019년도_종합	
제공방법	[√] 온라인 [ ] 오프라인	
제출예정일	2020 년 10 월 22 일	
전체 가명정보 제공 기관명(총수)	총 2개 : 홍길동(2019년도 종합), 결합신청기관 A사(파일명1) ※ 결합신청기관 B사는 별도의 가명정보를 제공하지 않고 활용만 함	
제공정보 요약	컬럼 수	25개
	전체 레코드 수	10,500개
	전체 파일 크기	25 MB

결합 결과물 이용		해당없음 [√]
시계열 분석	[ ] 해당없음 [ ] 시계열(신규) [ ] 시계열(추가, 결합접수번호 :)	
결합목적	[ ] 통계작성 [ ] 과학적 연구 [ ] 공익적 기록보존 등	
세부결합목적		

「개인정보보호법」 제28조의3제1항 및 같은 법 시행령 제29조의3제1항에 따른 결합을 위하여 결합전문기관에 결합신청서를 위와 같이 제출합니다.

2020 년 10 월 22 일

신청인 홍 길 동 (서명 또는 인)

결합전문기관의 장 귀하

첨부서류	
	1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부 2. 결합 대상 가명정보에 관한 서류(전체 항목명 가명처리 대상 항목명, 가명처리 기법 및 예시 등) 1부(해당 경우에 한함) * 결합키 생성에 사용된 속성 제외 3. 결합 목적을 증명할 수 있는 서류 1부 4. 안전조치계획 및 이를 증빙할 수 있는 서류 1부

[개인정보처리자 결합신청기관 A사의 신청서 예시]

가명정보 결합 신청서

신청번호	A-2020-10-001

결합신청자			
기관명	해당없음	사업자등록번호 또는 법인등록번호	1975년 09월 19일
주소	서울시 서초구 서초대로 11길 ㅇㅇ	대표자명	해당없음
담당자	김결합	담당자 연락처 (전화, e-mail)	010-123-4567 abc@caaa.com
신청자 구분	[ ] 개인 [√] 공공기관 [ ] 비영리법인 [ ] 민간기관		

가명정보 제공		해당없음 [ ]
파일명	파일1	
제공방법	[√] 온라인 [ ] 오프라인	
제출예정일	2020 년 10 월 22 일	
전체 가명정보 제공 기관명(총수)	총 2개 : 홍길동(2019년도 종합), 결합신청기관 A사(파일명1) ※ 결합신청기관 B사는 별도의 가명정보를 제공하지 않고 활용만 함	
제공정보 요약	컬럼 수	30개
	전체 레코드 수	21,00개
	전체 파일 크기	46 MB

결합 결과물 이용		해당없음 [ ]
시계열 분석	[√] 해당없음 [ ] 시계열(신규) [ ] 시계열(추가, 결합접수번호 :)	
결합목적	[√] 통계작성 [ ] 과학적 연구 [ ] 공익적 기록보존 등	
세부결합목적	(구체적 목적 설명.... 통계작성 .....)	

「개인정보보호법」 제28조의3제1항 및 같은 법 시행령 제29조의3제1항에 따른 결합을 위하여 결합전문기관에 결합신청서를 위와 같이 제출합니다.

2020 년 10 월 22 일

신청인 김 결 합 (서명 또는 인)

결합전문기관의 장 귀하

첨부서류	
	1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부 2. 결합 대상 가명정보에 관한 서류(전체 항목명 가명처리 대상 항목명, 가명처리 기법 및 예시 등) 1부(해당 경우에 한함) * 결합키 생성에 사용된 속성 제외 3. 결합 목적을 증명할 수 있는 서류 1부 4. 안전조치계획 및 이를 증빙할 수 있는 서류 1부

[개인정보처리자 결합신청기관 B사의 신청서 예시]

가명정보 결합 신청서

신청번호	A-2020-10-001

결합신청자			
기관명	해당없음	사업자등록번호 또는 법인등록번호	1978년 03월 05일
주소	서울시 서초구 효령로34길 ㅇㅇ	대표자명	해당없음
담당자	최가명	담당자 연락처 (전화, e-mail)	010-456-7890 abc@cbbb.com
신청자 구분	[ ] 개인 [ ] 공공기관 [ ] 비영리법인 [√] 민간기관		

가명정보 제공		해당없음 [√]
파일명		
제공방법	[ ] 온라인 [ ] 오프라인	
제출예정일	년 월 일	
전체 가명정보 제공 기관명(총수)		
제공정보 요약	컬럼 수	
	전체 레코드 수	
	전체 파일 크기	

결합 결과물 이용	해당없음 [ ]
시계열 분석	[√] 해당없음 [ ] 시계열(신규) [ ] 시계열(추가, 결합접수번호 :)
결합목적	[ ] 통계작성 [√] 과학적 연구 [ ] 공익적 기록보존 등
세부결합목적	(구체적 목적 설명.... 과학적 연구 .....)

「개인정보보호법」 제28조의3제1항 및 같은 법 시행령 제29조의3제1항에 따른 결합을 위하여 결합전문기관에 결합신청서를 위와 같이 제출합니다.

2020 년 10 월 22 일

신청인 최 가 명 (서명 또는 인)

결합전문기관의 장 귀하

첨부 서류
1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부
2. 결합 대상 가명정보에 관한 서류(전체 항목명 가명처리 대상 항목명, 가명처리 기법 및 예시 등) 1부(해당 경우에 한함)
* 결합키 생성에 사용된 속성 제외
3. 결합 목적을 증명할 수 있는 서류 1부
4. 안전조치계획 및 이를 증빙할 수 있는 서류 1부

부록 4

가명처리 및 익명처리 QnA

질의	**가명처리 또는 익명처리를 한 경우 해당 결과물이 적절하게 되었음을 어떻게 판단 하나요?**
답변	● 가명처리 및 익명처리의 적정성 여부는 별도의 심의위원회를 구성하여 판단하도록 하고 있습니다. ● 개인정보 보호법에서는 개인정보처리자의 판단에 따라 외부전문가를 포함하여 적정성 검토를 수행할 수 있다고 명시하고 있습니다. ● 교육분야 가명정보 처리 가이드라인에는 처리환경 등에 따라 적정성 평가를 위한 심의위원회를 구성하되, 교육기관 이외의 기관에 가명정보를 제공할 경우 반드시 외부전문가를 포함한 적정성 검토 심의위원회를 구성하여 적정성 여부를 검증할 수 있습니다. ● 익명처리를 한 경우에는 가명처리 절차와 유사하게 진행 할 수 있습니다.

질의	**개인정보를 암호화 하는 경우 가명정보에 해당하나요?**
답변	● 개인정보를 가명처리하는 여러 기법 중에는 암호화도 해당이 됩니다. ● 하지만 개인정보를 암호화하는 경우 가명정보에 해당할 수 있으나 가명처리가 적절하다고 판정하기에는 구체적 사안에 따라 달라질 수 있으니 반드시 적정성 검토가 이루어져야 할 것입니다. ● 예를 들면 일방향 암호화 알고리즘 중에 MD5 기법을 활용한 암호화는 안전하지 않아 암호화로 인정하지 않습니다. ● 따라서 MD5로 암호화 했다고하여 가명정보라고 판정할 수는 없을 것입니다.

질의	**민감정보나 고유식별정보를 정보주체의 동의없이 가명처리하여 활용할 수 있나요?**
답변	● 개인정보는 통계, 연구, 공익적 기록보존을 목적으로 가명처리 할 수 있습니다. ● 가명처리관련 특례의 내용은 법률에서 허용한 세 가지 목적 내에서 가명처리를 통해 정보주체의 별도 동의없이 수집 목적 외로 이용할 수 있도록 한 것입니다. ● 법령상에서는 민감정보와 고유식별정보에도 가명처리에 관한 특례조항이 적용됩니다. ● 다만, 고유식별정보 중 주민등록번호는 법정주의로 처리하기 때문에 활용할 수 없습니다. ● 민감정보는 그 자체만으로 개인이 식별되는 경우가 많으므로 가명처리를 적용할 때 주의를 기울여야 합니다.

질의	**가명정보에서 추가정보의 연결고리를 끊으면 익명정보에 해당하나요?**
답변	● 가명정보에서 추가정보의 연결고리를 끊으면 익명정보에 해당할 가능성이 있습니다. ● 하지만 추가정보의 수준이 일상에서 쉽게 습득할 수 있는 부분이 되어 법 제2조의 제1호 나목에서 이야기 하는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 요소들의 수준과의 차이에 따라 익명정보에 해당할 수도 있고 가명정보에 해당할 수도 있습니다. ● 또한 익명정보라 판단하여 외부 공개될 경우 다양한 수단과 방법에 의해 식별될 가능성이 있으면 개인정보로 판단할 수 있습니다. ● 따라서 해당 가명정보의 세부 고려사항에 따라 다르게 판단이 될 수 있을 것입니다.

질의	**당초 수집 목적과 합리적으로 관련된 범위 내에서 동의없이 개인정보를 처리하는 경우 가명처리 하지 않고 사용할 수 있나요?**
답변	● 개인정보 보호법 제15조 제3항에서 당초 수집 목적과 합리적으로 관련된 범위에서 동의없이 개인정보를 이용할 수 있다고 정의하고 있고, 시행령 제14조의2(개인정보의 추가적인 이용·제공 기준 등)에서 고려하여야 하는 사항들을 규정하고 있습니다. ● 따라서 시행령에서 요구하는 고려사항들을 만족하는 경우 가명처리 하지 않고 사용이 가능하도록 법령은 변경되었습니다. ● 다만, 적절한 수준의 안전성 확보 등 시행령에서 요구하는 내용에 대한 입증 책임은 개인정보처리자에게 있습니다.

질의	**신용정보법에는 가명정보를 상업적 목적의 통계 작성 사용이 가능하다고 명시되어있는데 개인정보 보호법에서는 명시되어 있지 않지만 상업적 목적으로 사용이 가능한가요?**
답변	● 개인정보 보호법에서는 명시적으로 해당 목적을 위해 사용이 가능하다고 규정하고 있지 않지만 법령(안)의 '제안이유'에 "산업적 목적을 포함하는 과학적 연구,..." 등으로 기술함으로써 산업적 목적으로 연구활동에 사용이 가능하도록 열어놓고 있습니다. ● 또한 가명정보 처리 가이드라인에서도 해당 내용에 대해 "시장조사와 같은 상업적 목적의 통계 처리도 포함"으로 사용이 가능하다고 설명하고 있습니다. 다만 1:1 마케팅과 같은 특정 개인을 식별할 수 있는 형태의 통계는 해당하지 않습니다. ● 다만 최초의 목적이 통계작성, 과학적 연구, 공익적 기록보존을 위해서 이루어져야 하며 그 결과 나온 결과물에 대해서 상업적 목적의 활용이 가능하다는 취지로 생각하시면 더 이해가 쉬울 것으로 판단됩니다.

질의	**개인정보 보호법 제28조의2(가명정보의 처리 등)을 준수하여 A의 가명정보A를 B에게 제공된 경우, 제공받은 B는 보유하고 있던 개인정보(또는 가명정보)와 제공받은 가명정보A와 자유롭게 결합이 가능한가요?**
답변	● A로부터 가명정보A를 받은 B는 제공받은 가명정보A의 소유권이 B에게 있으므로 해당 가명정보A는 B의 가명정보라고 판단할 수 있습니다. ● 이는 개인정보 보호법 제28조의3(가명정보의 결합 제한)의 서로 다른 처리자간의 가명정보에 해당하지 않습니다. ● 다만, 제공된 가명정보A에는 결합에 필요한 식별자(가명 ID)가 없으므로 결합이 불가능하며, 결합이 가능한 경우 제대로 가명처리 되지 않은 정보를 제공받은 것으로 위법의 소지가 있습니다. ● 참고로 개인정보보호법 제28조의2 제2항에서는 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니된다고 명시되어 있는 등, 가명정보를 제3자에게 제공하는 경우 결합이 가능한 정보는 포함하지 않도록 특별히 주의하여야 합니다.

질의	**기관이 가명정보 내부 결합을 하고자 하는 경우 결합전문기관을 통하지 않고 가능한데 이 결합한 데이터에 대한 적정성 검토는 누가 어떤 기준으로 하나요?**
답변	● 현행 법령상 기관 내부의 가명정보 결합은 결합전문기관을 통하지 않고 수행할 수 있습니다. ● 다만, 서로 다른 처리자간 가명정보 결합은 반드시 결합전문기관을 통하여야 합니다. ● 결합한 가명정보의 적정성 검증은 본 가이드라인을 참고하여 기관 내부적으로 심의위원회를 구성하여 적정성 검토 실시하면 됩니다. ● 보다 세부적인 내용은 교육분야 가명정보 처리 가이드라인의 내용을 확인하시면 됩니다.

질의	**가명정보가 너무 쉽게 만들어져서 가명정보를 역으로 추적하여 개인정보를 알아낼 수 있게 된다면 법적으로 어떻게 되나요?**
답변	● 개인정보 보호법 제28조의5(가명정보 처리 시 금지의무 등)에서 재식별을 금지하고 있고 혹시라도 가명정보를 처리 중에 특정 개인이 식별된다면 즉시 처리를 중지하고 회수 및 파기하도록 되어 있습니다. ● 가명정보를 역으로 추적하는 행위는 그 자체가 재식별 행위에 해당하며 이를 위반할 경우 처리자에게는 과징금(매출의 3% 이하에 해당되는 금액. 다만, 매출액이 없거나 산정이 곤란한 경우에는 4억원 또는 자본금의 3% 중 큰 금액 이하에 해당되는 금액)이 부과되고 개인에게는 최대 5년 이하의 징역 또는 5천만원 이하의 벌금에 해당됩니다.

질의	**가명정보가 법 적용대상이라고 하면, 개인정보와 동등한 수준으로 관리가 되어야 하는 건가요? 또 가명정보의 활용에 대해 정보주체는 자기결정권을 주장할 수 있나요?**
답변	● 가명정보는 개인정보 보호법 제2조(정의)에서 '개인정보'에 해당하므로 개인정보와 동등한 수준으로 관리가 되어야 합니다. ● 다만, 법 제28조의7(적용범위)에서 예외 규정을 두어 관리하지 않아도 되는 부분이 있습니다. ● 제20조(수집출처 고지), 제21조(파기), 제27조(이전 제한), 제34조제1항(유출통지), 제35조~제37조(정보주체 권리보장)이며 정보통신사업자의 경우 제39조의3~4, 제39조의6~8을 추가적으로 두고 있습니다. ● 이러한 적용 예외 규정을 둔 이유는 가명정보이기 때문에 정보주체를 특정할 수 없어서 해당 법령을 준수할 수 있는 방법이 없기 때문입니다. ● 따라서 정보주체는 자기결정권을 확인 할 수 있는 방법이 없습니다.

질의	**개인정보에 성명, 연락처, 연령, 거주지 등이 있다면 성명과 연락처 정도만 삭제하면 가명정보라고 할 수 있나요?**
답변	● 가명정보는 개인정보 보호법 제2조(정의)에서 원래의 상태로 복원하기 위한 추가정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보라고 정의하고 있습니다. ● 따라서 성명과 연락처가 삭제 후 남아있는 정보가 해당 개인정보처리자 내부에서 해당 정보를 취급하는 사람들에게 특정 개인이 식별되지 않는다면 가명정보라고 할 수 있습니다. ● 하지만 거주지가 특별하거나 그 자체만으로 개인이 식별되거나 해당 거주지에 연령을 결합할 경우 특정 개인이 식별이 된다면 가명정보라고 할 수 없습니다. ● 이런 경우 특정 개인에 대한 추가 가명처리(삭제)를 통해 가명정보로 변경하여야 합니다.

질의	**가명정보 결합 신청자는 결합전문기관의 반출처리공간에서만 가명 및 익명처리를 할 수 있는지?**
답변	● 반출처리공간은 결합신청기관이 결합된 가명정보를 반출하기 위해 가명정보 또는 익명정보로 추가 처리하는 곳입니다. ● 따라서 결합전문기관에 의해 결합된 가명정보를 반출하기 위한 가명 및 익명처리는 반출처리공간에서만 가능합니다. ● 또한 결합전문기관 내에서 분석하고 분석결과물만 반출할 수 있습니다. ● 반출처리공간 이외에서 결합된 가명정보를 처리 등 분석하고자 하는 경우 결합전문기관에게 반출 요청하고 반출 승인이 난 경우에 한하여 반출할 수 있습니다.

질의	**가명정보만 유출되었을 경우, 또는 가명정보와 가명정보의 추가정보가 유출되었을 경우 등 정보 유출에 대한 사후 법적 문제에 대한 내용이 궁금합니다.**
답변	● 개인정보 보호법 제2조(정의)에서 가명정보는 개인정보로 규정하고 있어 안전한 관리를 위한 안전성 확보조치 등을 요구하고 있습니다. ● 따라서 유출이 발생한 경우 법 제29조(개인정보 안전성 확보조치) 및 법 제28조의4(가명정보에 대한 안전조치의무 등) 위반에 해당하여 2년 이하의 징역 또는 2천만원 이하의 벌금에 해당한다고 볼 수 있습니다. ● 가명정보만 유출된 경우 가명정보 처리 적용 범위 특례에 따라 정보주체에 대한 유출 통지 등은 하지 않습니다. 하지만 가명정보와 추가정보가 모두 유출된 경우 개인이 식별되므로 가명정보 처리 특례 적용이 어려울 것으로 보이며 유출 통지 등이 추가로 이루어져야 할 것입니다.

질의	**가명정보와 개인정보가 매칭이 되는 정보를 기관이 저장할 수 있나요?**
답변	● 가명정보와 개인정보가 매칭이 되는 정보(가명처리시 발생하는 **추가정보**-매핑테이블 등)은 법령에서 요구하는 안전성 확보조치와 함께 기관에서 저장할 수 있습니다. 다만, 교육기관은 본 가이드라인에 따라 원칙적으로 추가정보는 삭제하도록 하고 있습니다. ● 하지만 가명처리 과정에서 나오는 추가정보가 아닌 별도로 만들어지거나 입수한 매칭 정보는 가명정보 재식별 금지 위반이기 때문에 원천적으로 불법이 됩니다. ● 이를 위반하면 5년 이하의 징역 또는 5천만원 이하의 벌금에 해당하며 기관은 매출액 3% 이내 등의 과징금에 처하게 됩니다.

질의	**가명정보를 제3자에게 제공하는 경우 목적내 제3자 제공인가요? 목적외 제3자 제공인가요? 또한 제3자 제공에 따른 후속조치(공지 및 대장 기록 등)는 무엇이 있나요?**
답변	● 법 제2조(정의)에 가명정보도 개인정보로 규정하고 있습니다. ● 가명정보의 처리는 특례로 일부 목적으로 허용했지만 일반적으로 목적외 처리에 해당하며 질문하신 내용에 따라 목적외 제3자 제공에 해당한다고 볼 수 있습니다. ● 가명정보를 처리하는 경우 법 제28조의4(가명정보에 대한 안전조치의무 등) 제2항에서 요구하는 기록을 작성하여 보관하여야 합니다. ● 또한 가명정보를 목적외 제3자 제공 함에 따라 홈페이지 등에 공지를 하시고 목적외 제3자 제공대장에도 기록 하셔야 합니다. ● 교육분야 가명정보 처리 가이드라인에서는 가명정보를 제3자에게 제공하는 경우 계약서를 기반으로 주요 제한사항을 포함하여 제공하도록 하고 있으며 세부적인 사항은 본 가이드라인을 참고하시면 됩니다.

질의	**데이터 3법 개정 이후 가명정보 활용을 위해 타 기관에서 우리 기관이 소유한 정보를 연구 목적으로 사용하고자 요청하는 경우 우리 기관이 취할 수 있는 방법은 무엇인가요?**
답변	● 타 기관에서 가명정보 처리를 목적으로 가명정보를 요청하는 경우 교육기관은 요청 사항의 적정성을 검토하여 적절하다고 판단하면 가이드 절차에 따라 조치하고 제공하면 됩니다. ● 다만, 타 기관에서 요청한다고 무조건 제공하여야 하는 것은 아니며 적정성 검토 단계 없이 바로 거부할 수도 있습니다. ● 정보공개법에 따라 요구하는 경우에도 가명정보는 개인정보에 해당하므로 가명정보를 함부로 제공해서는 안됩니다.

질의	**우리 기관은 교육부의 사업을 법령에 따라 위탁받아 개인정보를 보유하고 있습니다. 이 개인정보들을 우리 기관이 가명처리하여 활용할 수 있나요?**
답변	● 개인정보 및 가명정보의 처리 권한은 해당 정보의 소유자에 한정합니다. ● 위탁을 받아 보유하고 있다고 소유권이 있는 것으로 판단하여 무단으로 가명처리하여 활용할 수 없습니다. 해당 개인정보의 처리 권한과 책임은 소유권자에게만 있으므로 무단 사용은 불가능합니다. ● 위탁을 받은 경우 수탁자는 위탁자에게 반드시 계약서 등 공식적인 문서를 통해 해당 처리의 허가를 득하고 법령이 허용하는 범위 내에서만 개인정보의 처리가 가능합니다. ● 대표적인 예로 한국교육학술정보원은 위탁받은 NEIS 정보를 무단으로 가명처리가 불가능하며, 한국교육개발원은 위탁받은 교육통계 정보를 무단으로 가명처리가 불가능합니다.

질의	**개인정보처리자, 가명정보처리자, 익명정보처리자라는 용어가 나오는데 차이점은 무엇인가요? 그리고 A기관이 B기관으로 가명정보A를 제공 목적으로 생성하고 제공한 이후 가명정보A를 삭제한 경우 A기관은 어디에 해당하나요?**
답변	● 개인정보처리자는 개인정보를 처리하는 기관, 단체 및 개인 등을 의미합니다. 가명정보처리자는 가명정보를, 익명정보처리자는 익명정보를 처리한다는 점이 다릅니다. ● 다만, 가명정보도 개인정보에 해당하므로 가명정보처리자를 개인정보처리자로 말할 수 있습니다. ● 하지만 개인정보처리자는 가명정보처리자라고 할 수는 없습니다. 개인정보를 처리하더라도 가명정보를 처리하지 않는다면 별도로 가명정보처리자라고 할 수 없습니다. ● A기관이 제공을 목적으로 가명정보A를 만들고 삭제한 경우이므로 A기관은 가명정보A를 생성하고 삭제한 순간까지 가명정보처리자로 볼 수 있습니다. 하지만 삭제한 이후부터는 가명정보처리자로 볼 수 없습니다. 따라서 A기관은 가명정보 관리대장에 가명정보A에 대한 해당 내용을 기록·관리를 수행하며, 더 이상 가명정보가 없으므로 별도의 안전성 확보조치는 수행하지 않습니다. 만약 A기관이 가명정보A를 제공뿐만 아니라 내부 활용을 할 경우 안전성 확보조치 등 관련 사후관리를 수행해야 합니다. ● 가명정보A를 제공받은 B기관은 가명정보를 받아 처리하는 입장으로 B기관은 가명정보처리자에 해당하며, 가명정보처리자가 수행해야 하는 사후관리 내용을 동일하게 진행해야 합니다. 또한 제공받은 경우도 가명정보 "생성"에 해당하므로 가명정보 관리대장에 기록·관리해야 합니다.

부록 5

가명처리 및 익명처리 기법[1)]

1. 가명화 기법

□ 식별자의 대체값 생성 방법

○ 매핑 테이블(Mapping Table)

원래의 정보와 가명처리된 정보를 매핑하여 저장하는 테이블을 의미한다.

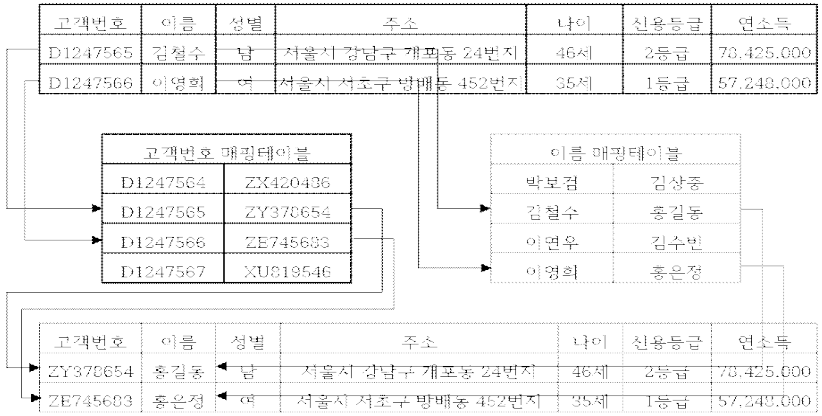
매핑 테이블은 암호화와 같이 알고리즘을 이용하여 처리되는 기법을 포함한 모든 가명처리기법에서 다시 원래의 정보로 복원하기 위한 방법이며 이를 사용하는 경우 이 매핑테이블은 추가정보로 분리보관의 대상이 된다.

원래의 정보와 가명처리된 정보를 매핑하여 저장하는 테이블을 의미한다.

원 이름	가명처리된 이름
김건우	강감찬
이규리	유관순
박혜자	신사임당
윤기태	이순신

부록 4-[표 1] 대체 기법을 이용한 가명처리의 매핑테이블 예시

1) 본 가이드라인에서 제시한 기술은 가명처리 및 익명처리에 대한 기법의 최소한의 안내를 기술하였으며 이 기술을 적용한다고 하여 가명처리 및 익명처리가 안전하게 되었다고 보장할 수 없습니다. 반드시 적정성 검토를 통해 안전성을 확보하셔야 합니다. 세부적인 처리 기법은 EU ENISA에서 발간한 보고서와 국제표준 ISO/IEC 20889 등을 참조하기 바랍니다.



부록 4-<그림 1> 매핑테이블의 적용 예시

○ 카운터(Counter)

가장 간단한 가명화 기술로 0에서 시작하여 1씩 증가하는 숫자를 발생시켜 이를 원본에 대체값으로 대체하는 기법이다. 카운터는 중복되는 값이 생성되어서는 안되며 기본적으로 원본의 값과 관계없이 생성되는 값이기 때문에 기본적으로 되돌릴 수 없는 가명화 기법이다. 이를 되돌릴 수 있게 하기 위해서는 '매핑테이블'을 유지하여야 하며 이에 대한 비용이 발생할 수 있다.

또한 Counter를 사용할 경우 대상 정보가 순서성(규칙성)이 존재하는 경우 Counter로 발생한 수를 기반으로 원본 정보와 결합할 수 있는 위험성이 존재하므로, 대상 정보의 순서를 섞는 등 규칙성이 발생하지 않도록 하여야 한다.

고객번호	Counter로 발생한 수
FS0158	1
FS0132	2
FS0058	3
FS0085	4
FS0093	5
FS0072	6
FS0415	7
FS0282	8
FS0234	9
FS0025	10
FS0351	11
FS0196	12

부록 4-[표 2] Counter의 적용 예시

○ PRNG(유사난수 생성기, Pseudo Random Number Generator)

난수를 발생시켜 그 난수를 이용하여 식별자를 대체하는 가명화 기법. 난수는 특정한 배열 순서가 규칙적인 의미가 없는 임의의 수를 뜻하며 컴퓨터의 출력은 입력을 기반으로 다양한 계산을 통해 결과가 얻어지게 되며 생성조건과 입력이 동일하면 항상 동일한 출력을 발생하기 때문에 완전한 난수의 생성은 불가능하다.

이에 따라 컴퓨터의 계산으로 얻어지는 난수는 완전한 난수가 아닌 난수의 특성을 일부 보유하고 있는 수이며 이를 유사 난수라고 한다.

유사난수 생성 알고리즘에는 선형합동법, 일방향 암호화를 이용한 유사난수 생성법, 대칭키 암호화나 공개키 암호화를 사용한 유사난수 생성법 등 다양한 알고리즘이 사용된다. 최근에는 소프트웨어적으로 생성된 유사난수가 아닌 하드웨어를 이용한 난수생성기를 이용하여 난수를 생성하는 경우도 증가하고 있으며 이러한 난수생성기를 기존의 의사난수생성기와 비교하여 순수 난수 생성기(TRNG, True Random Number Generator)라고 한다.

○ 해시 알고리즘

해시 알고리즘을 이용한 가명처리 기법은 크게 다음의 세 가지로 구분한다.

- 키가 없는 일반 가명처리 알고리즘(MDC, Message Digest Code)
- 솔트(Salt)를 추가한 일방향 암호화
- 키를 사용하는 일방향 암호화(MAC, Message Authentication Code)

키가 없는 일방향 암호화를 이용한 가명처리 기법은 추가정보를 보관해도 원본으로 되돌릴 수 없는 알고리즘을 이용하여 암호화를 적용하는 방식으로 시계열 분석 등의 필요에 의해 이전 가명 처리한 데이터와 결합하여 분석을 하여야 할 필요가 있는 경우 많이 사용되는 가명화 기법이다. 주요 알고리즘으로는 SHA 256, SHA 384, 512, 384/256, 512/256, LSH 256 등이 존재하며, 특히 MD5, SHA1 등 안전하지 않은 알고리즘은 사용하지 않도록 하여야 한다.

일방향 암호화는 원칙적으로 암호화된 해시값에서 원본을 추출할 수 없으나 원본을 파악하기 위한 다음의 공격이 존재한다.

- 무작위 대입 공격(Brute force attack) : 무작위로 대입된 값을 동일한 알고리즘으로 일방향 암호화 처리하여 그 결과값을 비교하여 원문을 찾아내는 공격기법으로 입력 가능한 모든 값을 무작위로 입력하여 공격하는 방법이다. 온라인상에서는 이를 막기 위해 지연처리 등의 방법을 사용할 수 있으나 데이터의 제공시에는 이러한 방어가 불가능하다.
- 레인보우 테이블 공격 : 미리 원문에 해당하는 모든 값을 동일한 해시 알고리즘으로 만들어놓은 테이블을 이용하여 원문을 찾아내는 공격으로 무작위 대입공격의 시간이 많이 소요되는 약점을 해결한 공격으로 매번 해시를 통해 대조하는 것이 아니라 전체 입력 범위에 대한 해시테이블을 미리 만들어 놓고 단순 비교만으로 찾아내는 방법이다. 무작위 대입공격에 비해 매우 빠른 속도의 공격이 가능해진다.

- 추측 공격 : 가명처리된 값이 추측을 통한 다양한 정보를 입력하여 원래의 값을 알아내려는 기법이다.

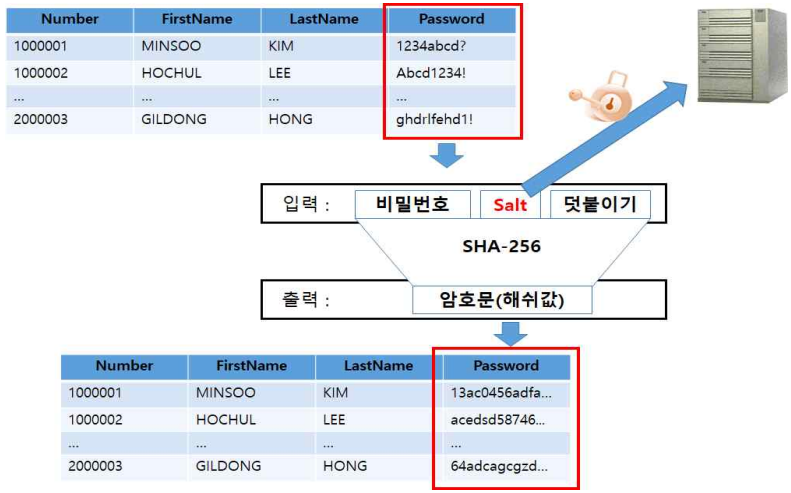
이름	성별	나이	수신 총잔액	신용등급
조미선	F	33	817,250	3
홍길병	M	61	4,559,120	2
김영심	F	50	13,601,564	1
이미정	F	70	979,118	4
김경태	M	40	5,501,809	2

⇩

가명처리 이름	성별	나이	수신 총 잔액	신용등급
F67FED7C74D09511CB9A75A219308021A216D3FD2A65716298ED018D966BF0E4	F	33	817,250	3
B040B101DC52F1EF33F180E6CB8FF369ED254D9DDA92DB9963327ADACA217E0A	M	61	4,559,120	2
EB9F054A348ADB9376D5D1678B98D4C32C92A69D4FAE40562FE433819942FA7B	F	50	13,601,564	1
5d2e387415514d6a30861f353fabd7eb3be682522f3f60fc9ccca6b717392bbc	F	70	979,118	4
c0d7b34289aad5a23b1355ade4f3accddb6dff69976f6706b804402b4628ac71	M	40	5,501,809	2

부록 4-[표 3] 일방향 암호화를 이용한 가명처리 예시

이러한 공격을 방어하기 위해 솔트(Salt, 일부에서는 Pepper이라고도 함)를 사용하거나 키를 사용하는 일방향 암호화를 사용한다.



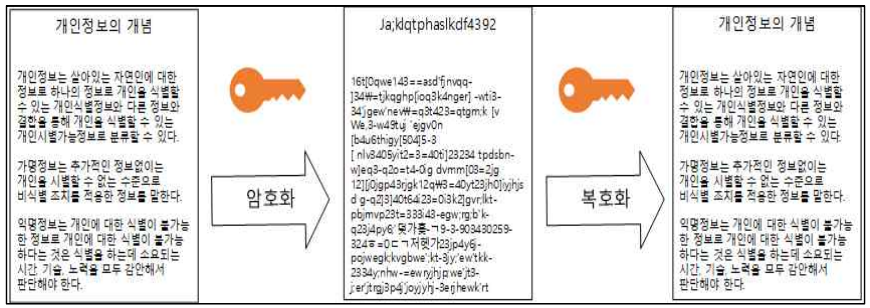
부록 4-<그림 2> Salt를 사용하는 단방향 암호화

솔트를 추가한 일방향 암호화의 경우 솔트값을 모르는 경우 위의 무작위 대입공격이나 레인보우 테이블 공격이 불가능하다. 그러나 동일한 솔트값을 계속해서 사용하는 경우 솔트값을 유추할 수 있는 문제가 발생할 수 있어 일정한 길이 이상의 솔트값(128bit 이상)의 사용과 계속해서 솔트 값을 변경하는 동적 솔트값을 사용하는 것이 더욱 안전한 Salt의 사용방법이다.

마지막으로 단방향 암호화 중 메시지 인증에 사용되는 MAC(Message Authentication Code)가 있으며 MAC의 경우 대칭키를 사용한다. 일반적인 해시 암호화가 키를 사용하지 않기 때문에 전송되는 암호문을 탈취하여 이를 위의 두 가지 공격 방법으로 원문을 알아낼 가능성이 존재하지만 MAC의 경우 키와 원문을 함께 이용하여 해시를 처리하기 때문에 단순히 전달되는 메시지만으로는 원문을 찾아낼 수 없어 안전하게 된다.

○ 양방향 암호화

양방향 암호화 알고리즘을 이용하여 원본 데이터를 암호화 한 후 암호화된 값을 원본데이터와 치환하여 사용하는 가명화 기법이며, 비밀키암호화와 공개키 암호화 기법으로 분류할 수 있다.



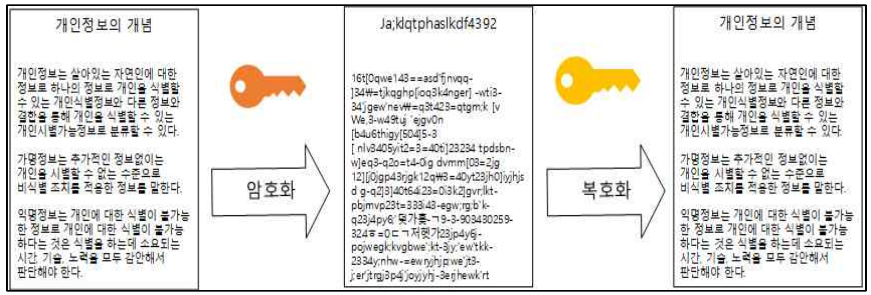
부록 4-<그림 3> 비밀키 암호화 기법

비밀키 암호화 기법은 대칭키 암호화 기법이라고도 하며 암호화와 복호화에 하나의 키를 이용하여 암호화를 하는 기법이며 암호문과 비밀키가 있으면 원문을 복원해 낼 수 있어 키의 관리가 매우 중요하다. 비밀키 암호화 시스템에서는 암호화된 정보를 주고 받는 사람이 많을수록 키의 개수가 서로의 정보를 보호하기 위해 n*(n-1)/2개가 필요하게 되어 키의 관리가 매우 어려운 방식이다.

· 주요 알고리즘으로는 AES, SEED, ARIA등이 있다.

하나의 키를 사용하는 비밀키 암호화 기법의 키관리의 단점을 해결하고자 만들어진 암호화가 공개키 암호화 기법이다. 공개키 암호화 기법에서는 암호화 할 때 사용하는 키와 복호화 할 때 사용하는 키가 서로 다르게 되며 그 중 공개하여 누구나 사용할 수 있는 키를 공개키, 개인이 보관하고 개인만 사용하는 키를 개인키라고 한다. 또한 공개키 암호화의 경우 공개키로 암호화한 것은 개인키로 복호화가 가능하고 개인키로 암호화한 것은 공개키로

복호화가 가능하며 보호를 위해서는 공개키(Public key)로 암호화 하고 개인키(Private key)로 복호화 하는 과정을 거치게 되며 인증을 목적으로 할 때에는 반대로 개인키로 암호화하고 공개키로 복호화를 한다. 양방향 암호화에 비해 암/복호화에 소요되는 시간이 매우 많이 필요하며 주요 알고리즘으로는 RSA, ECC 등이 있다.



부록 4-<그림 4> 공개키 암호화 기법

○ 마스킹(Masking)

마스킹은 정보에 대해 전체 또는 일부를 다른 문자·숫자 또는 기호로 대체하여 개인의 식별을 방지하는 방법으로 정보가 가지고 있는 개인정보 속성(식별자, 개인식별가능정보 등)과 관계 없이 모든 정보에 대해 적용할 수 있는 가명처리 기법으로 다른 가명처리 기법과 달리 마스킹 적용의 범위를 선택할 수 있기 때문에 일부에만 마스킹을 적용하는 경우 마스킹을 하지 않은 부분에 대한 정보의 유용성을 유지할 수 있다는 장점이 있다 아래의 [표 4]를 보면 동일한 정보에 서로 다른 마스킹을 적용하여 데이터의 의미를 다르게 부여하는 사례를 볼 수 있다.

원 데이터	마스킹 된 데이터	설명
홍길동	홍XX	직접 식별자에 대해 마스킹을 적용하여 식별성을 낮춘 사례
25살	2X살	식별가능정보 등에 대해 마스킹을 적용하여 범주화한 사례

원 데이터	마스킹 된 데이터	설명
F35(질병코드)	FXX(질병코드)	민감 질병에 대해 마스킹을 적용하여 식별성을 낮춘 사례
124-86-23875	124-XX-23875	사업자 등록번호에 대해 마스킹을 적용하여 법인의 종류를 알 수 없도록 한 사례
124-86-23875	124-86-XXXXX	사업자 등록번호에 대해 마스킹을 적용하여 법인의 종류만 구분이 가능하도록 한 사례
abc@def.com	XXX@def.com	이메일 주소에 대해 마스킹을 적용하여 메일 발신 기관만 구분이 가능하도록 한 사례
abc@def.com	aXX@XXX.com	이메일 주소에 대해 마스킹을 적용하여 메일 발신자와 발신 기관을 모두 확인할 수 없도록 한 사례

부록 4-[표 4] 다양한 마스킹 적용 예시

○ 스크램블링(Scrambling)

스크램블링(셔플링)은 대체의 방법 중 데이터 내부의 문자열에 대해 문자열 자체의 순서를 랜덤하게 변경하는 것을 의미한다. 아래 표5는 카드번호에 대해 스크램블링을 적용한 예로 스크램블링은 원데이터의 값을 크게 변화시키지 않음으로 인해 데이터의 형식을 그대로 유지하게 된다. 데이터의 수집 시 가명화를 적용할 때 자주 사용된다.

이름	성별	나이	신용카드번호	신용등급
조미선	F	33	9410-2124-3268-4878	3
홍길병	M	61	5365-1002-1324-7689	2
김영심	F	50	4032-9380-0230-7913	1
이미정	F	70	6250-0370-2481-9079	4

⇩

이름	성별	나이	신용카드 번호	신용등급
조미선	F	33	1246-8842-9283-0144	3
홍길병	M	61	1326-3019-8705-6524	2
김영심	F	50	3083-3904-0319-2720	1
이미정	F	70	3679-0952-8047-2100	4

부록 4-[표 5] 신용카드 번호에 대한 스크램블링 적용 예시

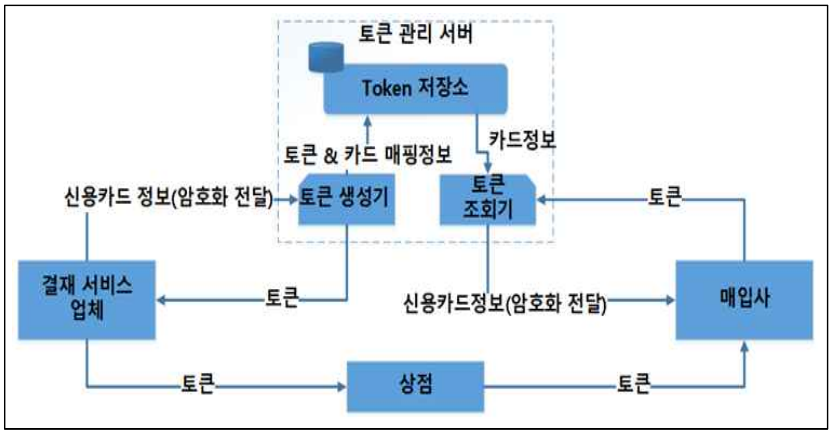
○ 블러링(Blurring)

일반적으로 블러링은 이미지의 픽셀의 정확도를 낮춰 이미지의 식별성을 낮추는 기법으로 이미지에 적용하는 비식별 기법의 하나이다. 하지만 이 기법을 가명화 기법으로 사용하는 경우 원 데이터의 정확도를 낮춰 식별성을 낮추는 것으로 일정정도의 잡음을 추가하여 정확한 값을 파악하지 못하게 하는 형태로 적용이 된다.

블러링의 경우 추가되는 잡음의 크기에 따라 값이 가지고 있는 통계적인 특성을 그대로 유지하도록 하는 것이 중요하다.

○ 토큰시스템

가명처리에서 사용하는 토큰시스템은 개인을 식별할 수 있는 정보를 토큰으로 변환하여 사용함으로써 개인정보를 직접 사용하여 발생하는 개인에 대한 식별 위험을 제거하여 개인정보를 보호하게 된다. 토큰 시스템은 앞에서 매핑테이블, 유사난수 생성기, 암호화 등을 조합하여 시스템으로 구현한 것으로 다음과 같이 구현된다.

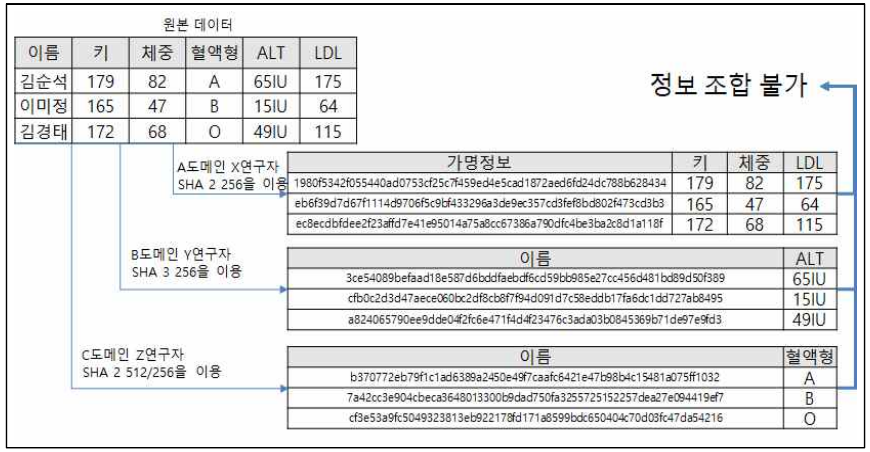


부록 4-<그림 5> 토큰 시스템

토큰 시스템 중 가장 널리 사용되는 시스템은 신용카드 정보에 대한 토큰 시스템으로 EMV표준[2]으로 제정되어 있다.

○ 다형성 암호화

다형성 암호화는 동일한 정보에 대해 제공 시 다른 방식의 가명처리를 적용하여 제공하는 기법으로 제공되는 도메인별로 다른 가명을 가지게 된다.



부록 4-<그림 6> 다형성 암호화 적용 예시

다형성 암호화는 가명정보의 유통으로 인한 재식별을 방지할 수 있다.

2) https://www.emvco.com/wp-content/uploads/2017/08/EMV-Payment-Tokenisation-Technical-Framework-v2-General-QA-FINAL.pdf

2. 통계기법

○ 표본추출(Sampling)

- 데이터 주체 즉, 각 개인별로 전체 모집단이 아닌 표본에 일반화 또는 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법으로 전체 데이터를 분석하는 비용 대비 적은 비용을 가지고 전체를 분석하는 것과 유사한 결과를 얻을 수 있는 기법
- 모집단의 대표성이 있는 표본을 확보하는 것이 중요
- 샘플데이터는 모집단 데이터에 비해 식별에 대한 위험이 상대적으로 적음
  예시) 전국민이 모집단인 경우 특정 개인식별가능정보의 조합이 유일하면 이는 곧 식별이 될 가능성이 매우 높아지지만 10%샘플에서 특정 개인식별가능정보의 조합이 유일한 경우 이는 원데이터에서 유일하지 않을 가능성이 더 높으며 이에 따라 재식별에 대한 위험도 감소됨
- 표본 추출은 확률적 표본 추출방법과 비확률적 표본 추출 방법으로 나눠짐
- 확률적 표본 추출방법에는 무작위 표본 추출, 계통적 표본 추출, 층화 표본 추출, 집락 표본 추출 등이 있음
- 비확률적 표본 추출 방법에는 임의 표본 추출, 판단 표본 추출, 할당 표본 추출, 누적 표본 추출 등이 있음
- 통계적으로 분석에 많이 사용되는 표본 추출은 확률적 표본 추출을 많이 사용함

○ 총계처리

총계처리는 특정 컬럼을 통계적으로 처리하는 것으로 다음의 5가지 중 하나로 처리를 한다.

- 평균값, 중간값, 최빈값, 최대값, 최소값

총계처리 종류

- 평균값 : 전체의 평균값으로 대체
- 중간값 : 모든 값을 일렬로 세울 때 정확하게 중간에 있는 값으로 대체
- 최빈값 : 대상 중 가장 많은 빈도로 나타난 값으로 대체
- 최대값 : 모든 값 중 가장 큰 값으로 대체
- 최소값 : 모든 값 중 가장 적은 값으로 대체

이 5가지 중 어떤 기법을 사용하여 처리를 하는지는 데이터의 활용 목적 및 데이터 특성에 따라 구분되어야 한다. 일반적인 통계분석에서는 평균값, 중간값을 주로 사용하며 특정 분류에 대한 경계값을 확인하여야 하거나 크게 군집을 나누는 분석의 경우에는 최대값과 최소값을 사용한다.

월간 교육비 (원본)		월간 교육비 (평균값)
268,057		2,048,620
1,029,984		2,048,620
1,376,470		2,048,620
2,408,994		2,048,620
8,541,986		2,048,620
1,502,390	→	2,048,620
639,105		2,048,620
2,408,994		2,048,620
2,672,396		2,048,620
4,574,763		2,048,620
775,202		2,048,620
716,392		2,048,620
361,038		2,048,620

부록 4-[표 6] 총계처리 – 평균값 예시

총계처리 방식에서 평균으로 대체하는 경우 평균의 오류를 주의해서 사용하여야 한다.

평균의 오류

평균의 오류는 평균을 내는 대상에 특이치가 포함되어 평균값이 일반적인 평균치에 비해 많이 올라가거나 내려가는 현상을 의미한다. 우리나라 통계청에서 소득 등의 통계를 발표할 때 이런 문제를 해결하기 위해 평균값과 중간값을 같이 발표하여 평균의 오류에 대한 확인을 쉽게 할 수 있도록 하고 있다.

월간 교육비 (원본)		월간 교육비 (최댓값)
268,057		8,541,986
1,029,984		8,541,986
1,376,470		8,541,986
2,408,994		8,541,986
8,541,986		8,541,986
1,502,390	→	8,541,986
639,105		8,541,986
2,408,994		8,541,986
2,672,396		8,541,986
4,574,763		8,541,986
775,202		8,541,986
716,392		8,541,986

부록 4-[표 7] 총계처리 – 최대값 예시

최소값은 대상 중 가장 적은 값으로 대체한다

월간 교육비 (원본)		월간 교육비 (최소값)
268,057		268,057
1,029,984		268,057
1,376,470		268,057
2,408,994		268,057
8,541,986		268,057
1,502,390		268,057
639,105	→	268,057
2,408,994		268,057
2,672,396		268,057
4,574,763		268,057
775,202		268,057
716,392		268,057
361,038		268,057
1,404,903		268,057

부록 4-[표 8] 총계처리 – 최소값 예시

최빈값은 대상 중 가장 많은 빈도로 나타난 값으로 대체한다

월간 교육비 (원본)		월간 교육비 (최빈값)
268,057		2,408,994
1,029,984		2,408,994
1,376,470		2,408,994
2,408,994		2,408,994
8,541,986		2,408,994
1,502,390		2,408,994
639,105	→	2,408,994
2,408,994		2,408,994
2,672,396		2,408,994
4,574,763		2,408,994
775,202		2,408,994
716,392		2,408,994
361,038		2,408,994
1,404,903		2,408,994

부록 4-[표 9] 총계처리 – 최빈값 예시

중간값은 대상을 나열하여 정확하게 중간에 있는 값으로 대체한다

월간 교육비 (원본)		월간 교육비 (나열값)		월간 교육비 (중간값)
268,057		268,057		1,390,687
1,029,984		361,038		1,390,687
1,376,470		639,105		1,390,687
2,408,994		716,392		1,390,687
8,541,986		775,202		1,390,687
1,502,390		1,029,984		1,390,687
639,105	→	1,376,470	→	1,390,687
2,408,994		**1,390,687**		1,390,687
2,672,396		1,404,903		1,390,687
4,574,763		1,502,390		1,390,687
775,202		2,408,994		1,390,687
716,392		2,408,994		1,390,687
361,038		2,672,396		1,390,687
1,404,903		4,574,763		1,390,687
1,390,687		8,541,986		1,390,687

부록 4-[표 10] 총계처리 – 중간값 예시

3. 암호화 기법

ISO/IEC 20889의 암호화에 대한 분류는 일반적인 암호학에서의 암호화의 분류와 약간 다른 분류법을 사용하고 있다. 일반적인 암호화에 대한 분류는 가명처리 기법에서 살펴본 것과 같이 양방향 암호화, 단방향 암호화, 그리고 양방향 암호화에서 비밀키 암호화, 공개키 암호화 등으로 분류를 하지만 여기서의 암호화 기법의 분류는 암호화된 정보를 활용하는 관점에서 분류를 하고 있다. (일반적인 암호화의 분류는 가명처리 기술 부분을 참조)

ISO/IEC 20889의 암호화 기법에 대한 분류는 아래와 같다.

- 결정성 암호화
- 순서보존 암호화
- 형태보존 암호화
- 동형 암호화

○ 결정성 암호화

동일한 알고리즘과 동일한 키로 동일한 값을 암호화한 경우 암호화된 값이 항상 일정한 값으로 생성되는 암호화 기법이다.

- 분석 시 기존 값과 동일하다, 동일하지 않다 정도의 비교만 필요한 경우 결정성 암호화는 안전한 데이터 사용을 가능하게 한다.
- 암호화 된 열에 대한 지점조회, 동등조인, 그룹화 및 인덱싱 가능
- 암호화 된 값 집합이 작은 경우 암호화 된 값에 대한 정보의 추측 가능

이름	SHA 256 암호화
조미선	F67FED7C74D09511CB9A75A219308021A216D3FD2A65716298ED018D966BF0E4
홍길병	B040B101DC52F1EF33F180E6CB8FF369ED254D9DDA92DB9963327ADACA217E0A

이름	SHA 256 암호화
김영심	EB9F054A348ADB9376D5D1678B98D4C32C92A69D4FAE40562FE433819942FA7B
이미정	5d2e387415514d6a30861f353fabd7eb3be682522f3f60fc9ccca6b717392bbc
김경태	c0d7b34289aad5a23b1355ade4f3accddb6dff69976f6706b804402b4628ac71
유영근	ff0a8815dc0615c98ce55f1dd2891dacf0b6e543f0dc529e76cea3d2301979ce
박을규	a944576344a31ba20e1b882f3659dd86baf1b5b9f40fe6273ac0aadaa6713ef6
문정은	71b026571df2fd990befb3c8fc4e75330df0155dfcdb260d39e22160e285a58e
오한근	4ac5d9d3fc5da58762b7056f3f465767f8fbd5786f7b1184e466927051fe29f3
전태홍	fb6519f6fcf0f9995b6045c57b6e8187c6733f40e573b76f993c11ff93f916da
이현주	611e5aaa81ae738d2c530f15e55a12cd79bb9c3873374aaa3b6affd79b3b16f2
백지연	e3f5e91c327536c898e8d1c37c5aeabcf2d2bb03d3864e55e8f10f695a491168
민영기	d978992c87838a39c6d9222a6142ece2a405cd7530afed84c501cb9f2677619a
김수복	8326718173db5d40107b9ac186a08da59c6f4ab6f5defed2953af424a293878d

부록 4-[표 11] 결정성 암호화 예시(단방향 암호화 적용 예시(SHA 256 no Salt))

○ 순서보존 암호화

원데이터의 순서와 암호화된 암호값의 순서가 동일하게 유지되는 암호화 방식이다.

암호화된 상태에서도 원본의 순서가 유지되어 검색속도 저하를 극복하고 분석 시 값들 간의 크기에 대한 비교만 필요한 경우 순서보존 암호화를 통해 원본사용시보다 안전한 분석이 가능하다.

상대적으로 암호화 강도가 낮으며 순서 자체가 중요한 정보가 되어 식별성을 가질 수 있다.

값	Order-preserving encryption 적용 값
1	0x0001102789d5f50b2beffd9f3dca4ea7
2	0x000422a6142ece2a405cd7530afed84

값	Order-preserving encryption 적용 값
3	0x009fb3c8fc4e75330df0155dfcdb48sd
4	0x010ce5aaa81ae738d2c530f15e55az7
5	0x04jo294lkksjlrops2365t9lk2jkml4m22
6	0x04zwoekrjfl3422ljfsadnml8s903ljkasjj
7	0x05e75330df0155dfcdb230afed84c5a
8	0x06002b6e8187c6733f40e573b76f99a
9	0x07234e75330df0155dfcdb260d39eac
만약 x<y, Enc(x) < Enc(y)	

부록 4-[표 12] 순서보존 암호화 예시

○ 형태보존 암호화

원데이터의 형태와 암호화된 암호값의 형태가 동일하게 유지되는 암호화 방식이다.

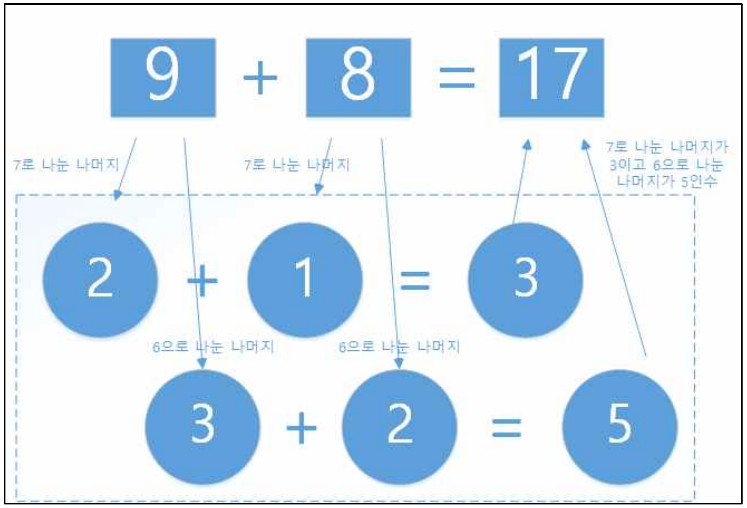
원데이터와 동일한 크기의 동일한 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장공간의 스키마 변경 이슈가 없어 저장공간의 비용 증가를 해결할 수 있으며 또한 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용된다. 이전에는 암호화의 강도에 대한 논란이 있었으나 2016년 미국NIST는 AES기반의 FPE에 대해 미국 표준으로 승인되면서 안전성에 대한 논란이 적어지고 있다.

부록 4-<그림 7> FPE 암호화를 이용한 신용카드 번호에 대한 암호화

○ 동형 암호화

1970년대 처음 이론연구가 시작된 이후 2009년에 IBM의 연구원인 Gentry에 의해 기술적 가능성이 증명되었으며 2011년 미국 MIT 대학의 기술보고서에서 10대 유망기술로 선정된 4세대 암호체계를 말한다.

암호화된 상태에서의 연산이 가능한 암호화 방법으로 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용이 가능하다.



부록 4-<그림 8> 동형 암호화 예시

연산 중 덧셈연산만 가능하거나 곱셈연산만 가능한 기법을 반동형 암호화 또는 부분동형암호화라고 하고 덧셈과 곱셈이 모두 가능한 기법을 완전 동형암호화라고 한다.

현재 전세계적으로 5개의 구현그룹에서 연구가 진행되고 있으며 서울대학교의 HeaAn, IBM의 HELib, Micorsoft의 SEAL, European HEAT 프로젝트의 PALISADE, cuHE 등의 알고리즘이 있다.

동형 암호화 기술은 데이터의 정확한 값을 사용할 필요가 상대적으로 높은 의료분야에서 매우 높은 요구사항을 가지고 있으며 생체 인식 분야에서도 생체 정보를 복호화하지 않고 인증을 수행할 수 있어 많은 연구가 진행되고 있으나 현재까지는 처리 및 저장에 들어가는 비용의 문제, 국제 표준에 대한 문제, 동형암호화 인프라간의 호환성 결여 등으로 활용에 걸림돌이 되고 있다.

동형암호의 난제성은 매우 높은 수준으로 양자컴퓨팅 환경에서도 안전하다는 것이 어느 정도 증명되고 있어 양자컴퓨팅 환경에서의 암호화 기법으로도 각광받고 있다.

○ 동형비밀분산

식별자 또는 기타 민감정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)로 대체하는 기법이다. 이 기법은 계산에 관한 성능 오버헤드가 상대적으로 낮지만, 쉐어 소유자와 쉐어를 교환할 때 추가적인 오버헤드가 발생하며 이용 기법에 따라 상당한 성능 비용이 발생할 수 있으나 공유 데이터의 통제된 재식별화는 비식별화된 데이터의 쉐어를 소유한 쉐어 소유자가 정해진 수만큼 재식별화에 모두 동의할 경우만 가능하게 되는 기법이다. 관련 기법과 연산에 대한 설명은 ISO/IEC 19592와 ISO/IEC 29101에 표준화 되어 있다.

4. 삭제 기법

삭제 기법은 식별가능성이 높은 직접 식별자 또는 개인식별정보를 전체 또는 부분 삭제하여 개인의 식별성을 낮추는 기법이다.(로컬 삭제의 경우 민감속성에 대해서도 삭제를 하기도 함)

ISO/IEC 20889에는 마스킹*, 로컬삭제, 레코드 삭제의 세가지 삭제 기법을 소개하고 있으나 실제 사용할 때는 마스킹, 컬럼 삭제, 부분삭제, 레코드 삭제(로컬삭제 포함), 식별요소 전부 삭제 등 다양한 삭제 기법이 사용된다.

삭제 기법은 분석 목적에 필요 없는 컬럼이나 중복된 정보를 다루고 있는 컬럼을 삭제하여 식별성을 낮추는 것을 포함한다.

* 식별요소를 'ㅇㅇ' 또는 '**' 등으로 대체하여 식별되지 않도록 하는 기법으로 마스킹은 삭제 기법 및 대체 기법 모두에 해당

○ 레코드 삭제

특이한 값을 가지고 있어 식별성이 높거나 또는 아주 민감한 정보라 정보주체에 대한 프라이버시 침해의 정도가 큰 경우 분석에 크게 문제가 되지 않을 때 레코드를 삭제하는 기법이다.

성별	주소	나이	주상병	신장	체중
F	경남	50	J01	168	72
F	서울	70	J00	156	51
M	서울	40	I21	169	56
M	경기	43	F02	174	78
M	경기	23	B20	172	70
F	경기	67	K35	162	61

⇩

성별	주소	나이	주상병	신장	체중
F	대전	33	A15	163	43
F	서울	70	J00	156	51
M	서울	40	I21	169	56
F	경기	67	K35	162	61

부록 4-[표 13] 민감한 질병코드를 가진 레코드에 대한 레코드 삭제 예시

○ 컬럼 삭제

직접 식별자나 식별 가능성이 높은 식별가능정보, 중복된 정보 등을 삭제하는 것으로 일반적으로 대상 컬럼을 삭제한다.

환자번호	이름	성별	생년	주소	나이	주상병	신장	체중
10010785	조미선	F	1985	대전 동구 용운동	33	A15	163	43
10011953	홍길병	M	1957	경북 안동시 용상동	61	C15	172	84
10012231	김영심	F	1968	경남 진주시 옥봉동	50	J01	168	72
10012598	이미정	F	1948	서울 강서구 가양3동	70	J00	156	51
10013649	김경태	M	1978	서울 은평구 역촌1동	40	I21	169	56
10014221	유영근	M	1975	경기 고양시 고양동	43	F02	174	78
10015665	박을규	M	1995	경기 수원시 고색동	23	E11	172	70

⇩

환자번호	이름	성별	생년	주소	나이	주상병	신장	체중
		F		대전 동구 용운동	33	A15	163	43
		M		경북 안동시 용상동	61	C15	172	84
		F		경남 진주시 옥봉동	50	J01	168	72
		F		서울 강서구 가양3동	70	J00	156	51
		M		서울 은평구 역촌1동	40	I21	169	56
		M		경기 고양시 고양동	43	F02	174	78
		M		경기 수원시 고색동	23	E11	172	70

부록 4-[표 14] 환자번호, 이름, 생년 컬럼에 대한 컬럼삭제 예시

○ 부분 삭제

부분 삭제는 컬럼의 일부를 삭제하여 데이터의 식별성을 낮추는 기법으로 주소의 일부를 삭제하거나 날짜의 일부를 삭제하는 등 다양한 방법이 있다. 부분삭제는 대부분의 경우 범주화 기법으로 자주 사용된다.

삭제 시 삭제 범위가 너무 적은 경우 데이터의 식별성을 낮추지 못할 수 있으므로 주의하여야 한다.

성별	주소	나이	주상병	신장	체중
F	대전시 동구 용운동	33	A15	163	43
M	경북 안동시 용상동	61	C15	172	84
F	경남 진주시 옥봉동	50	J01	168	72
F	서울시 강서구 가양3동	70	J00	156	51
M	서울시 은평구 역촌1동	40	I21	169	56
M	경기 고양시 고양동	43	F02	174	78
M	경기 수원시 고색동	23	E11	172	70

성별	주소	나이	주상병	신장	체중
F	대전	33	A15	163	43
M	경북	61	C15	172	84
F	경남	50	J01	168	72
F	서울	70	J00	156	51
M	서울	40	I21	169	56
M	경기	43	F02	174	78
M	경기	23	E11	172	70

부록 4-[표 15] 주소 컬럼에 대한 부분삭제 예시

○ 식별요소 전부 삭제

- 식별성이 있는 요소를 전부 삭제하는 방법이다.
- 실제 사용되는 사례는 HIPAA의 세이프 하버(Safe Harbor)기법이 대표적인 사례이다.

①이름	⑦사회보장번호	⑬각종 장비 식별변호
②주소 정보*	⑧의료기록번호	⑭인터넷 주소(URL 정보)
③날짜 정보*	⑨건강보험번호	⑮IP주소
④전화번호	⑩계좌번호	⑯생체정보(지문, 음성 등)
⑤팩스번호	⑪자격취득번호	⑰전체 얼굴사진 및 유사 이미지
⑥이메일 주소	⑫자동차번호 (차량식별번호, 등록번호 등)	⑱기타 특이한 식별번호 또는 코드

* ② 주소정보 : 주(State)보다 작은 지리적 구획으로 세부주소(Street), 시(City), 군(County), 구역(Precint), 우편번호 및 그와 대등한 지역번호(Geocode)를 포함하나, 인구조사국의 공개적으로 이용 가능한 데이터를 따를 때 다음에 해당하는 우편번호 첫 3자리의 경우 예외로 함(우편번호 첫3가지로 구성되는 지리적 단위가 2만명 이상을 포함 등)
* ③ 날짜정보 : 연도는 제외하고 개인과 직접적으로 관계된 날짜와 관련된 모든 요소(출생일자, 입원일자, 89세를 초과하는 나이는 90세로 단일 분류 등)

부록 4-[표 16] 미국 HIPAA Safe harbor방식의 식별자 유형

5. 해부화

하나의 테이블을 두 개 이상의 테이블로 분할하여 개인의 식별성을 낮추는 기법이다. 특정 테이블에 정보의 식별성이 발생하는 이유는 식별과 관련된 컬럼과 개인의 속성과 관련된 컬럼이 함께 존재하기 때문이라고 생각하고 모든 컬럼을 분석하여 식별과 관련있는 컬럼과 분석 대상의 속성과 관련된 컬럼을 분할하여 실제 데이터 분석 시에는 분석대상의 속성들만으로 분석을 진행하게 하는 것이다.

- 일반적으로 해부화를 적용할 때 식별관련 컬럼과 분석 대상 컬럼을 분할한다.

임시 일련번호	이름	성별	나이	주상병	신장	체중
1	조미선	F	33	A15	163	43
2	홍길병	M	61	C15	172	84
3	김영심	F	50	J01	168	72
4	이미정	F	70	J00	156	51
5	김경태	M	40	I21	169	56
6	유영근	M	43	F02	174	78

식별관련 컬럼 속성관련 컬럼

임시 일련번호	이름	성별	나이
1	조미선	F	33
2	홍길병	M	61
3	김영심	F	50
4	이미정	F	70
5	김경태	M	40
6	유영근	M	43

임시 일련번호	주상병	신장	체중
1	A15	163	43
2	C15	172	84
3	J01	168	72
4	J00	156	51
5	I21	169	56
6	F02	174	78

부록 4-[표 17] 해부화 적용 예시

또한 하나의 코드를 두 개 이상의 코드로 분할하여 개인의 식별성을 낮추는 기법도 포함한다.(코드화라고 함)

일련번호	이름	성별	해부화 적용 성별
10010785	조미선	F	3
10012231	김영심	F	9
10012598	이미정	F	7
10020238	박화순	F	9
10022529	김경태	M	15
10035485	오한근	M	16
10036391	전태홍	M	18

부록 4-[표 18] 성별 컬럼에 대한 코드화 적용 예시

코드화의 경우 분석자에게 있어 개인의 식별성을 낮추는 것은 불가능하다. 데이터의 분석 목적을 달성하기 위해 어떤 코드가 어떤 값을 가지고 있는지를 알려줘야 하는 경우가 대부분이기 때문이다. 그럼에도 코드화 역시 하나의 비식별기법으로 포함하는 것은 데이터의 유출 시 개인의 식별성을 낮추는 데는 큰 효과를 나타내기 때문이다.

6. 일반화 기술

○ 라운딩

라운딩은 수치 데이터에 가장 일반적으로 사용하는 일반화 기술로 일반 라운딩, 랜덤라운딩, 제어라운딩 등의 다양한 기법이 포함된다.

가) 일반 라운딩

- 일반 라운딩은 반올림, 올림, 내림으로 사용된다.

나이		반올림	올림	내림
33		30	40	30
61		60	70	60
50		50	50	50
70		70	70	70
40	→	40	40	40
43		40	50	40
23		20	30	20
67		70	70	60
66		70	70	60
47		50	50	40

부록 4-[표 19] 일반 라운딩의 종류별 예시

나) 랜덤 라운딩

랜덤 라운딩은 라운딩의 자리수와 기준이 되는 수를 자유롭게 지정할 수 있는 라운딩 기법이다.

- 라운딩 자리수 : 만단위는 천단위에서 반올림, 백만단위는 만단위에서 반올림, 1억 단위는 백만 단위에서 라운딩 등
- 기준이 되는 수 : 반올림에만 적용 가능한 방법으로 일반적인 사사오입이 아니라 기준점을 다르게 주고 그 기준점 미만은 모두 버리고 그 이상은 올리는 방법

소득금액		자리수 기반 랜덤 라운딩
869,250		869,000
4,559,120		4,560,000
13,601,564		13,600,000
979,118		979,000
5,501,809	→	5,500,000
609,622		610,000
3,885,329		3,890,000
23,992,801		23,990,000
185,878,354		186,000,000
274,489		1) 274,000

부록 4-[표 20] 자리수를 기반으로 하는 랜덤 라운딩 예시

다) 제어 라운딩

제어 라운딩은 원본의 행, 열의 합과 라운딩 적용 후 행, 열의 합이 동일하게 만드는 라운딩 기법이다. 일반적인 라운딩에는 특정한 수를 기준으로 라운딩을 적용하지만 제어 라운딩의 경우 계산에 의해 적절한 수에서 라운딩을 적용한다. 통계적으로 라운딩 처리를 통한 값의 왜곡을 막을 수 있는 기법이다. 그러나 처리하여야 할 수의 개수가 증가할수록 처리하는데 필요한 컴퓨팅 파워와 시간이 아주 많이 소모되는 라운딩 기법으로 일반적으로 잘 사용되지 않는다. 다만 적절한 알고리즘을 이용하여 처리하는 경우 수치 연산을 중심으로 하는 데이터의 분석에 높은 효과를 줄 수 있으나 매우 어렵다.

나이		반올림	제어라운딩
33		30	30
61		60	60
50		50	50
72		70	70
43		40	40
44	→	40	50
23		20	20
67		70	70
68		70	70
49		50	50
합계 510		**합계 500**	**합계 510**

부록 4-[표 21] 제어 라운딩 예시

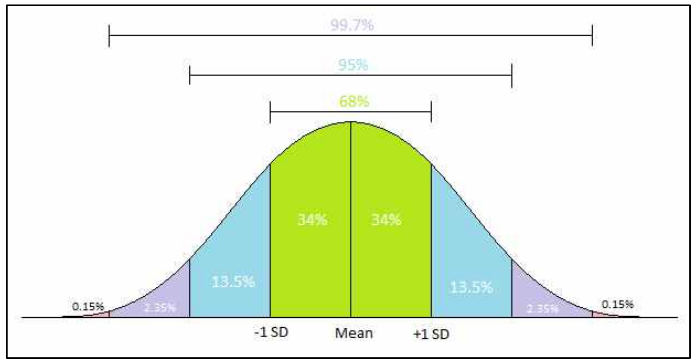
○ 로컬 일반화

로컬 일반화는 전체 데이터 중 특정 부분에 대해서만 일반화를 적용하는 것으로 특정 부분이 분포상 특징 또는 값의 특이성으로 인해 식별성이 높아지는 경우 사용된다.

○ 상하단 코딩

정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 개인의 식별성을 가질 수 있으므로 이를 해결하기 위해 적은 수의 분포를 가진 양끝단의 정보를 범주화 등의 비식별 기법을 적용하여 개인의 식별성을 낮추는 기법이다.

이러한 특이치는 때로는 평균의 오류를 나타내기도 하고 때로는 분석에 있어 노이즈의 역할을 하여 분석의 결과를 오히려 부정확하게 할 수 있다. 이러한 특이치에 대해 처리하는 것이 상하단 코딩이다. 즉 데이터 중 특정 값의 빈도가 아주 적거나 매우 특이한 값을 가지는 경우 이의 비식별처리를 위해 사용하는 기법이 상하단 코딩이다.



부록 4-<그림 9> 정규분포에서의 Σ에 따른 분포 비율

상하단 코딩에서 가장 중요한 것은 어디까지를 정상적인 값을 판단하여 분석에 사용하고 어디까지를 특이한 값으로 판단하여 특정한 처리를 하느냐에 있다.

통계학적으로는 3시그마, 캡핑 등의 방법을 사용하여 특이치를 판별한다.

○ 속성결합(범주화)

문자나 숫자의 하위의 공통된 특성을 찾아 하나의 상위 개념으로 묶는 기법을 의미한다. 또는 특정 정보를 해당 그룹의 대표값이나 구간값으로 변환하는 방법으로 적용하기도 한다. 속성결합을 적용하면 특정 정보의 명확한 값을 숨길 수 있기 때문에 감추기라고도 한다. 개인식별가능정보의 경우 대부분 인구통계학적 특성을 가지고 있으며(나이, 주소, 성별 등) 이러한 특성은 분석 시 일정 정도의 범주화를 분석 목적에 의해 필요로 하게 된다. 속성결합과 범위방법, 일반화는 대부분의 경우 개인식별가능정보에 사용하면 분석의 목적에 따른 결합으로 사용된다.

아래 부록 4-[표 22] 는 숫자형 데이터에 대해 상위 개념으로 묶는 범주화를 적용한 것이다.

나이 (원데이터)	5살 단위 범주화	10살 단위 범주화	20살 단위 범주화
33	30~34	30~39(30대)	20~39
61	60~64	60~69(60대)	40~59
50	50~54	50~59(50대)	40~59
70	70~74	70~79(70대)	60~79
40	40~44	40~49(40대)	40~59
43	40~44	40~49(40대)	40~59
23	20~24	20~29(20대)	20~39

부록 4-[표 22] 숫자형 데이터에 대한 범주화 적용 예시

숫자형 데이터의 경우 범주화를 하면 데이터의 Type이 변경될 수 있어 이에 대한 주의가 필요하다.

아래 부록 4-[표 23] 은 문자형 데이터에 대해 상위 개념으로 묶는 속성결합을 적용한 것이다.

상세주소	시군구단위	시도단위	광역시, 도 구분
대전 동구 용운동	대전 동구	대전	광역시
경북 안동시 용상동	경북 안동시	경북	도
경남 진주시 옥봉동	경남 진주시	경남	도
서울 강서구 가양3동	서울 강서구	서울	광역시
서울 은평구 역촌1동	서울 은평구	서울	광역시
경기 고양시 고양동	경기 고양시	경기	도
경기 수원시 고색동	경기 수원시	경기	도

부록 4-[표 23] 문자형 데이터에 대한 속성결합 예시

7. 무작위화 기술

○ 잡음추가

원래의 데이터에 아주 적은 잡음을 추가하여 데이터의 식별성을 낮추는 기법이다.

이때 사용하는 잡음(노이즈)는 원 데이터의 분석에 영향을 주지 않을 정도의 노이즈를 추가하여야 하며 종방향 일관성(Longitudinal Consistency)을 보장하여야 한다. 종방향 일관성은 분석의 영향과 오류를 최소화하기 위해 연관 컬럼에 대해 동일한 노이즈를 추가하여야 한다는 것으로 다음의 예를 통해 살펴볼 수 있다. 예를 들어 입원일자에 노이즈를 +3일을 추가하였다면 연관컬럼인 퇴원일자에도 동일한 +3일의 노이즈를 추가하여 입원일수라는 데이터에 변경이 없도록 적용하여야 한다는 뜻이다. 연관 컬럼에 대한 종방향 일관성을 적용하지 않는 경우 데이터를 분석하는 중요 컬럼에 잡음을 추가하면 데이터의 분석결과에 많은 오류가 포함될 수 있다.

퇴원일자	노이즈	노이즈퇴원일자
2001-11-05	3	2001-11-08
2007-09-27	-1	2007-09-26
2002-06-11	-5	2002-06-06
2002-10-27	-6	2002-10-21
2006-01-18	3	2006-01-21
2007-06-17	4	2007-06-21
2005-10-10	-4	2005-10-06

부록 4-[표 24] 퇴원일자에 대한 잡음 추가 예시

마지막으로 잡음의 크기가 예측될 수 있어서는 안된다는 것이다. 잡음의 크기가 예측이 가능하거나 일률적인 잡음을 주는 경우 잡음을 추가한 이유인 식별성을 낮추는데 도움이 되지 못할 수 있다. 이에 따라 잡음을 추가하는 경우 특정 범위 내에서 무작위로 발생한 잡음을 추가하여야 한다.

○ 순열(Permutation, 재배열), 1:1 교환(Swapping)

순열은 특정 컬럼의 데이터를 무작위로 순서를 변경하는 기법(일명 재배열이라고도 부름)으로 데이터의 훼손 정도가 매우 크기 때문에 무작위로 순서를 변경하는 조건 선정에 주의가 필요하다. 분석 목적이 대상 컬럼의 값이 다른 컬럼의 값들과 연관성이 없는 경우에만 적용하여야 한다.

특정 조건 하(전체 컬럼에 적용이 아니라 일부 컬럼에만 제한적으로 적용)에서 재배열을 하는 경우 식별성은 낮추면서 분석 결과에 영향을 최소화할 수 있다.

시도	나이	월소득 (원본)	월소득 (전체 재배열)
인천	20대	64,167	10,169,741
인천	20대	10,169,741	61,667
인천	20대	1,054,167	5,987,923
인천	20대	3,009,167	23,333
인천	20대	5,987,923	1,054,167
울산	20대	1,904,819	3,009,167
울산	20대	10,395,123	8,031,667
울산	20대	45,833	10,395,123
서울	20대	8,031,667	2,671,137
서울	20대	2,671,137	1,904,819
서울	20대	7,048,333	7,048,333
서울	20대	61,667	45,833
서울	20대	1,820,267	1,820,267
서울	20대	23,333	64,167

부록 4-[표 25] 전체 재배열 예시

○ 부분 총계

부분총계는 특정 조건에 따라 부분적으로 총계처리를 적용하는 것을 의미한다.

시도	나이	월 소득 (원본)	비고
인천	20대	4,057,033	
인천	20대	4,057,033	
인천	20대	4,057,033	} 동질집합
인천	20대	4,057,033	
인천	20대	4,057,033	
울산	20대	4,115,258	
울산	20대	4,115,258	} 동질집합
울산	20대	4,115,258	
서울	20대	3,276,067	
서울	20대	3,276,067	
서울	20대	3,276,067	} 동질집합
서울	20대	3,276,067	
서울	20대	3,276,067	
서울	20대	3,276,067	

부록 4-[표 26] 동질집합 내의 특정 컬럼을 총계처리 한 예시

나이	주소	월 소득	월 소득 부분총계 처리
인천	20대	64,167	7,657,033
인천	20대	28,169,741	7,657,033
인천	20대	1,054,167	7,657,033
인천	20대	3,009,167	7,657,033
인천	20대	5,987,923	7,657,033
울산	20대	1,904,819	1,904,819
울산	20대	2,395,123	2,395,123
울산	20대	45,833	45,833
서울	20대	8,031,667	8,031,667
서울	20대	2,671,137	2,671,137
서울	20대	7,048,333	7,048,333
서울	20대	61,667	61,667
서울	20대	1,820,267	1,820,267
서울	20대	23,333	23,333

부록 4-[표 27] 특정 조건에 너무 특이한 값이 있을 때의 부분 총계예시

부분총계의 사용 목적은 순열의 사용 목적과 동일하다. 즉 특정 동질집합 또는 레코드가 식별성이 높거나 민감도가 높을 때 이를 완화하기 위해

사용한다. 부분총계에 사용하는 값의 변경은 총계처리에서 사용하는 것과 같이 평균값, 중간값, 최빈값, 최대값, 최소값 중 하나로 변경하게 된다.

8. 재현데이터

원 데이터의 활용도를 유지하며 가상의 데이터를 합성하여 만드는 기술로 활용목적은 달성이 가능하나 개인의 특성이 배제되어 있어 안전한 활용이 가능한 기술이다. 적용 범위에 따라 완전 재현 데이터, 부분 재현 데이터, 하이브리드 재현 데이터의 세 가지 유형이 있다.

- 완전 재현 데이터 : 모든 데이터가 재현데이터로 구성한 데이터
- 부분 재현 데이터 : 민감한 속성을 가진 일부 컬럼만 재현데이터로 구성하는 경우
- 하이브리드 재현 데이터 : 재현데이터와 실제 데이터를 모두 사용하는 유형, 가장 높은 유용성을 가지나 메모리등에 소요되는 비용과 처리 시간이 많이 필요함