|  |
| --- |
| 대 외 비 – 영구 |
| 지경정 22-01 |
| 보안위원회 |
| ’22. 10. 24. |

**가명정보처리 매뉴얼**

**22.10.26.**

**보안위원회**

**\* 본 문서는 (주)GS Retail의 영업비밀로 당사의 허락 없이 임의로 일부 혹은 전체를 사용하거나 전재하는 행위를 금합니다.**

**\* GS Retail Co.,Ltd. All Rights Reserved Proprietary and Confidential**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **문서 변경 이력** | | | | | | | | |
|  | | | | | | | | |
| **버전** | **승인일자** | **기안자** | | | **내용** | | | **비고** |
| 1.0 | ‘22.10.26 | 보안센터/서혁남 | | | 최초(안) 작성 | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
|  |  |  | | |  | | |  |
| **목 차** | | | | | | | | |
|  | | |  |  | |  |  | |
| **1. 개요** | | |  | **3. 위험도 측정 및 가명처리수준 정의** | |  | **5. 가명정보의 안전한 관리** | |
|  | | |  |  | |  |  | |
| 1.1 목적  1.2 적용범위  1.3 용어의 정의 | | |  | 3.1 위험도 측정  3.2 데이터 위험성 평가  3.3 처리 환경의 위험성 평가  3.4 가명처리 수준 정의 | |  | 5.1 가명(익명)정보 이용 및 제공  5.2 가명정보 제공 시 유의사항 | |
| **2. 가명정보 처리절차** | | |  | **4. 적정성 평가** | |  | **별첨 & 붙임** | |
|  | | |  |  | |  |  | |
| 2.1 내부활용/결합  2.2 외부결합 | | |  | 4.1 적정성평가  4.2 적정성평가단 구성 | |  | 별첨1. 가명처리계획서  별첨2. 가명정보 결합신청서  별첨3. 반출신청서  별첨4. 위험도측정결과\_개인정보속성정의  별첨5. 개인정보 속성 정의표  별첨6. 적정성평가 기초자료  별첨7. 적정성평가항목  붙임1. 가명처리기법 | |

# 1. 개요

## 1.1 목적

본 매뉴얼은 GS리테일(이하 "회사"라고 한다)의 가명정보 처리 시 기준 및 수행방법을 제시 함으로서, 가명정보와 개인식별 가능성을 방지하고, 관련된 업무 담당자들이 가명처리절차에 따라 수행할 수 있도록 안내하여 관련 법령을 준수할 수 있도록 하는데 목적이 있다.

- 개인정보 보호법 및 시행령

- 가명정보의 결합 및 반출 등에 관한 고시(개인정보보호위원회)

- 가명정보처리 가이드라인(개인정보보호위원회, 2022.04)

- 금융분야 가명•익명처리 안내서(금융위원회, 금융감독원, 2022.01 / 외부결합)

## 1.2 적용범위

본 매뉴얼은 『개인정보보호 업무지침』에 따라 회사 내 가명(익명)정보처리, 회사 내/외 가명정보결합, 제공 등 가명(익명)정보 처리 업무를 모두 포함한다.

## 1.3 용어의 정의

1. 개인정보 : 생존하는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등에 의하여 특정한 개인을 알아볼 수 있는 부호, 문자, 음성, 음향 및 영상 등의 정보(해당 정보만으로는 특정 개인을 알아 볼 수 없어도 다른 정보와 쉽게 결합하여 알아볼 수 있는 경우에는 그 정보를 포함)
2. 민감정보 : 개인정보 중 특히 수집에 유의하고 보호해야 하는 정보로 유전정보, 범죄경력정보, 인종 및 민족에 관한 정보와 개인의 신체적, 생리적, 행동적 특징에 관한정보로서 특정인을 알아볼 목적으로 생성한 정보
3. 가명정보 : 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 원래의 상태로 복원하기 위한 추가 정보의 사용 결합 없이는 특정 개인을 알아볼 수 없는 정보
4. 익명정보 : 시간, 비용, 기술 증을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보로, 해당 정보만으로 특정 개인을 알아볼 수 없으며 다른 정보와 쉽게 결합하여도 알아볼 수 없고, 개인을 식별할 수 없도록 비식별화(범주화, 통계화 등의 기법)한 정보
5. 추가정보 : 개인정보의 전부 또는 일부를 대체하는데 이용된 수단이나 방식(알고리즘 등), 가명정보와의 비교 및 대조 등을 통해 삭제 또는 대체된 개인정보 부분을 복원할 수 있는 정보(매핑테이블정보, 매핑테이블정보와 기타 속성자 정보, 가명처리에 사용된 원본 개인정보 등) 등을 의미
6. 특이정보 : 정보주체에게 고유하게 부여된 정보는 아니지만, 해당 정보주체에 대해서만 해당되는 정보로 누구든지 쉽게 식별이 가능한 정보
7. 식별자 : 해당 정보주체에게만 고유하게 부여되어 있는 정보로써 그 자체만으로 개인식별성이 강한 정보로, 성명, 사진, 주민등록번호, 전화번호, 이메일주소, IP주소, 차량번호 등
8. 준식별자 : 정보주체에게만 고유하게 부여된 정보는 아니지만 다른 정보와 결합하면 특정 개인을 식별하기 쉬운 정보로 생년월일, 사망일, 결혼기념일, 직업, 성별, 주소, 우편번호 등
9. 속성정보: 개인정보처리자만 보유하고 있어 개인정보처리자 이외의 자는 다른 정보와 결합해도 특정 개인을 식별하기 어려운 정보로 상품 구매이력, 멤버십 등급 등
10. 데이터셋 : 정보를 체계적으로 관리하거나 처리할 목적으로 일정한 규칙에 따라 구성되거나 배열된 둘 이상의 정보, 정보집합물과 같은 의미임
11. 메타데이터 : 정보자원이나 정보자원의 묶음을 설명(기술, describe)한 데이터, 식별자와 준식별자, 속성, 생성시기, 정보주체 수, 데이터 형태, 개인정보 활용 동의여부 속성 등이 포함
12. 결합대상정보 : 결합신청자가 결합을 위해서 가명처리한 정보로서 결합전문기관에 제공하는 가명정보 중 결합키를 제외한 정보
13. 결합키 : 결합키관리기관이 결합키 연계정보를 생성할 때 임시적으로 사용되는 정보
14. 결합결과물 : 결합대상정보가 결합된 정보를 의미하며, 내부 결합을 통해 생성된 결과물은 ‘내부결합결과물’, 외부 결합을 통해 생성된 결과물은 ‘외부결합결과물’로 사용
15. 재식별 : 추가정보 또는 취급자가 달리 보유하고 있는 다른 정보나 공개된 정보와의 결합 또는 대조 및 비교 등을 통해 특정 개인을 알게 되거나 알아보려고 하는 상태 또는 행위
16. 적정성 평가 : 가명처리 기준에 따라 적절히 가명처리 되었는지 확인하는 절차
17. 적정성 평가단 : 가명(익명)처리의 목적이 적합한지 검토하고, 목적 달성을 위해 적절한 수준의 가명처리 여부 및 재식별 가능성 여부 등을 판단하며 외부 결합 또는 외부 제공하는 경우 외부 전문가를 포함
18. 내부활용 : 정보주체에게 분석 등의 동의없이(약관 및 개인정보 수집 및 이용 동의서 기준) 가명처리한 데이터를 회사 내에서 통계처리, 과학적 연구 목적으로 사용하는 행위
19. 내부결합 : 정보주체에게 분석 등의 동의없이(약관 및 개인정보 수집 및 이용 동의서 기준)가명 처리한 데이터를 회사 내에서 연결하는 행위
20. 외부결합 : 회사의 가명데이터를 타사와 결합하기 위해 결합전문기관을 통해 신청 및 전송, 결합 등을 하는 행위
21. 외부제공 : 가명정보 또는 익명정보를 제3자에게 제공하는 행위
22. 결합대상기관 : 회사 외 가명정보 결합을 수행하고자 하는 외부 기관
23. 결합전문기관 : 개인정보처리 간의 가명정보 결합을 수행하기 위해 개인정보보호위원회 또는 관계 중앙행정기관의 장이 정하는 전문기관
24. 결합키관리기관 : 결합키연계정보를 생성하여 결합전문기관에 제공하는 등 가명정보의 안전한 결합을 지원하는 업무를 하는 기관(한국인터넷진흥원 또는 개인정보보호위원회가 지정하여 고시)
25. 추가정보처리시스템 : 가명처리 시 발생하는 매핑데이터 또는 결합키와 같은 추가정보의 보관, 원본데이터를 이용하여 가명처리를 수행하는 공간으로 가명처리에 의해 생성되는 가명정보는 적정성평가 후 가명처리시스템으로 이관한 다음 즉시 삭제
26. 가명정보처리시스템 : 내/외부 결합에 사용되는 결합대상정보, 내부집합결과물, 외부결합결과물 지정 및 통계처리, 과학적 목적으로 사용되는 공간
27. 가명정보처리자 : 업무 목적으로 개인정보를 가명처리하여 활용 또는 제공하는 법인 등을 의미
28. 가명정보취급자 : 개인정보처리자의 지휘 및 감독을 받아 가명정보를 처리하는 임직원, 협력업체 직원 등 가명정보를 취급/활용하는 자로 추가정보 취급의 권한을 요청할 수 없으며 수탁사, 협력업체 등 외부인력은 특정 개인정보처리시스템의 접근권한이 제한될 수 있다.
29. 추가정보취급자 : 원본 데이터를 가명처리하는 자로 개인정보보호 전담부서와 논의하여 최소한의 인원으로 제한해야 한다. 또한, 추가정보취급자가 가명정보 취급과 개인정보처리 권한이 필요한 경우, 추가정보취급 권한과 분리하여 발급받아야 한다.
30. 결합결과물취급자 : 결합전문기관을 통해 전달받은 결합결과물을 취급하는 자로 결합결과물 건 별에 각각 접근권한을 요청할 수 있다. 수탁사, 협력업체 등의 외부인력은 특정 가명정보처리시스템의 일부 데이터 분석 공간 또는 개인정보처리시스템의 접근권한이 제한될 수 있다.

# 2. 가명정보 처리절차

가명정보는 관계 법령 및 개인정보보호 업무지침에 따라 통계처리, 과학적연구, 공익적 목적으로만 사용할 수 있다.

※ 예시

- 상권별 주 방문고객의 선호도, 주요 소비상품 들 소비패턴을 분석하여 지역상권별 고객 맞춤형 상품 제공 및 재고관리전략 수립

- 가구 구성형태, 연령대 및 성별, 소득 수준에 따라 주요 구매상품 및 품목정보, 구매금액을 분석하여 소비행태 확인(구매 상위고객 연령대, 구매 비중 등)

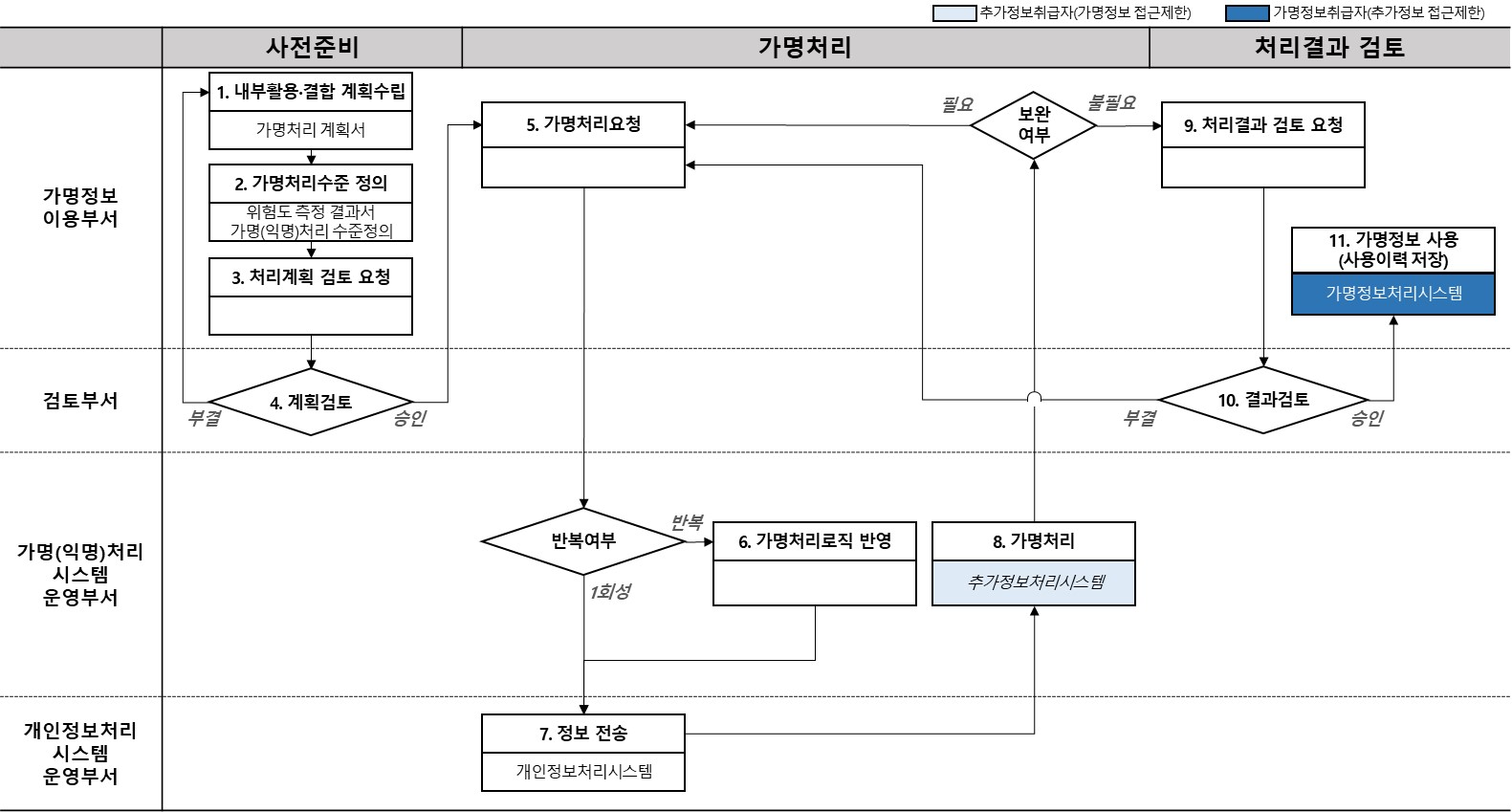
- 신규 상품의 구매 성향도(연령대, 성별 등) 분석하여 프로모션 진행 계획

가명정보 처리절차는 데이터 이용환경, 결합 및 제공 유무에 따라 유형을 분류한다.

|  |  |
| --- | --- |
| **사용 유형** | **수행내용** |
| **내부활용/결합** | * 내부활용: 회사 내부에서 개인정보처리시스템의 정보를 가명처리하여 사용 * 내부결합: 회사 내부의 가명처리된 정보간 결합   - 내부결합은 가명정보처리시스템에 적재된 가명정보간 결합을 전제로 한다. |
| **외부결합** | 외부기관과의 데이터 결합 수행 및 이용(결합전문기관, 데이터전문기관)  - 결합신청 이후의 절차는 결합전문기관 절차 준용  - 신용정보회사와의 데이터 결합 시, 데이터전문기관 통해서 수행  - 결합신청 이후의 절차는 데이터전문기관 절차 준용 |

## 2.1 내부활용/결합

내부활용 및 결합 시, 아래 흐름 및 표를 참고하여 수행하도록 한다.

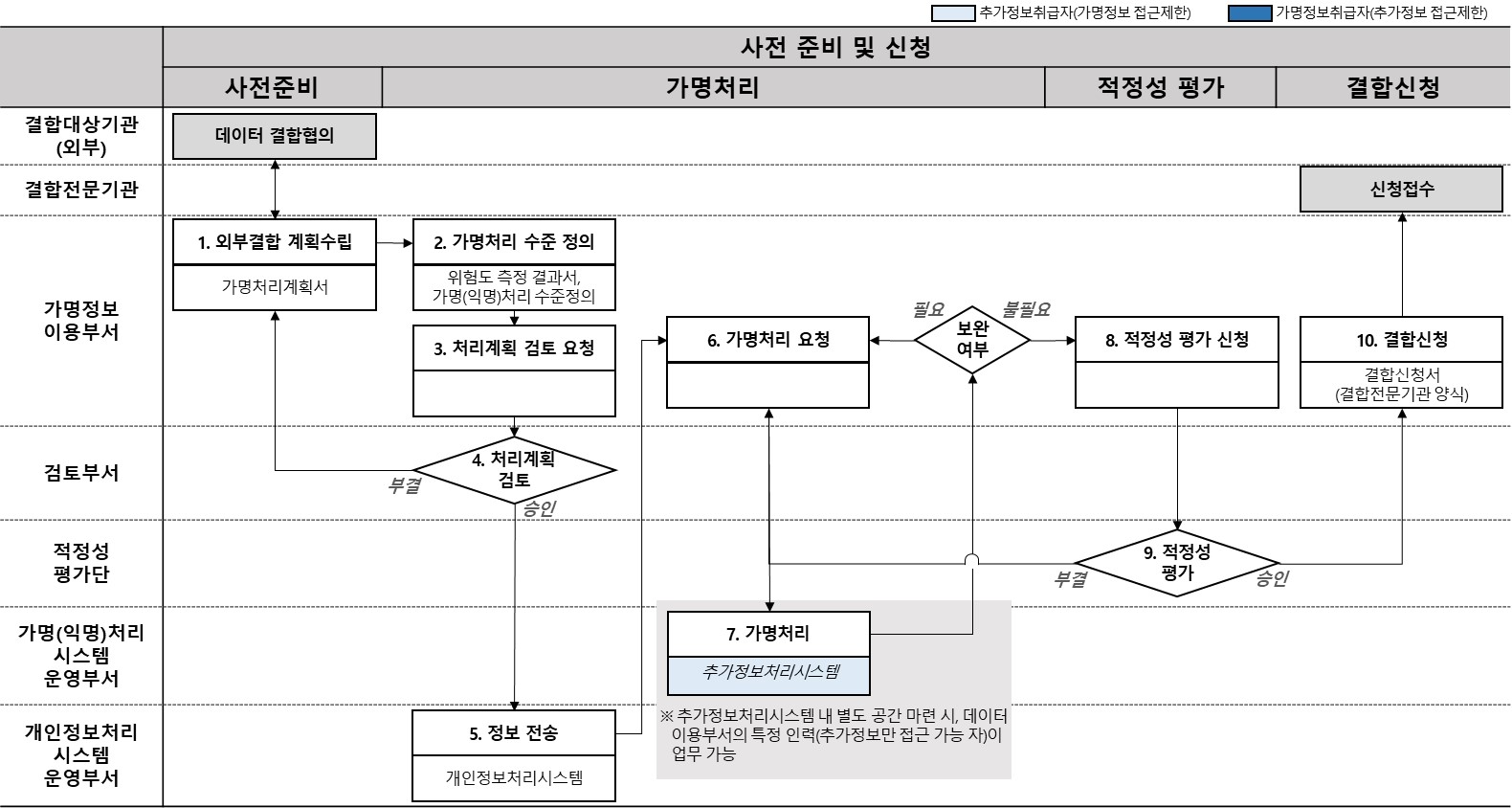


**<가명정보 처리 프로세스 내부활용/결합>**

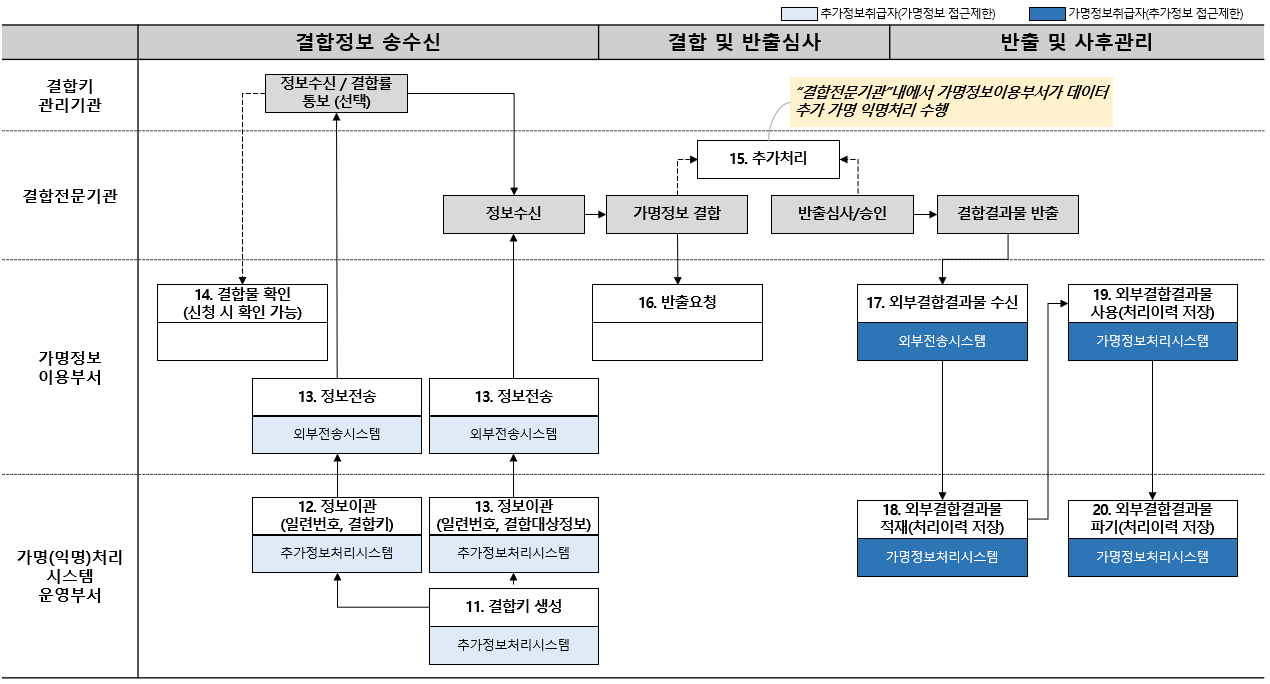
| **수행업무** | **수행업무 내용** | **주관부서** |
| --- | --- | --- |
| **1. 내부활용 및 결합 계획수립** | * **가명처리 목적 정의, 가명처리 대상 데이터 신청** * 이용자 동의 없는 개인정보의 가명처리는 통계, 과학적 연구, 공익적 목적으로만 사용 가능 * 이용자 동의 목적 내 사용 가능 * **이용 또는 제공 환경 정의** * 가명정보취급자: 가명정보를 취급(이용)하는 자 대상 확인 * **추가정보 사항 정의** * 추가정보 항목 생성 방법 (결합키 등) * 추가정보 파기 여부 정의 | * 가명정보이용부서 |
| **2. 가명처리수준정의** | * **가명처리 대상 개인정보 항목별 가명처리 방법 정의** * 가명(익명)처리 속성 정의 및 위험도 측정 결과 작성 (별첨4) * 개인정보 속성 정의표 참고 (별첨5) | * 가명정보이용부서 |
| **3. 처리계획 검토 요청** | * **요청에 필요한 문서 작성 후 처리계획 검토 요청** * (결재)소속팀장, (합의) 검토부서, (합의)개인정보처리시스템 운영부서 | * 가명정보이용부서 |
| **4. 처리계획 검토** | * **내부활용/결합 처리계획 진행여부 검토** * 내부활용 및 내부결합 목적에 적합한 가명정보 사용여부 검토 * 가명정보 처리계획에 대한 검토는 보안센터에서 진행 * 단, 가명정보 이용부서가 데이터플랫폼본부일 경우 본부 내 지정 인원이 병행 검토 진행 | * 검토부서 |
| **5. 가명처리 요청** | * **가명(익명)처리시스템 운영부서에 개인정보 가명처리 요청** * 가명(익명)처리 속성 정의 및 가명처리계획 제공 * 가명정보 이용부서와, 가명/익명 시스템 운영부서가 동일할 수 있음 * **보완이 필요한 경우 추가처리 요청** | * 가명정보이용부서 |
| **6. 가명처리 로직 반영** | * **반복적인 가명처리의 경우 처리 로직을 시스템 반영** * 가명(익명)처리 속성 정의 및 가명처리 계획 기준 로직 반영 | * 가명(익명)처리시스템 운영부서 |
| **7. 정보 전송** | * **반복적인 가명처리의 경우 가명처리 대상 정보 전송** * **1회성 가명처리의 경우 사용될 정보 제공** | * 개인정보처리시스템 운영부서 |
| **8. 가명처리** | * **요청 받은 내용으로 가명처리 수행** * 추가정보와 가명정보에 대해 동시 접근 불가 (별도 인원 배정) * 사용기간 경과 또는 작업 완료 시 원본 데이터 파기 * 추가정보의 경우 별도 salt값을 적용하여 해쉬(hash)함수 적용   (salt값의 경우 DB접근제어 운영 환경 내 보관)   * **가명정보 적재 전/내부결합결과물 적재 시 메타 데이터 저장** * 가명처리 결과물(데이터셋 별) 사용목적, 항목별 개인정보 유형, 속성(고유식별자, 준식별자 등), 파기 예정일 등 관리 * **가명정보(내부결합결과물) 적재이력 저장** * 처리일시, 처리유형, 가명처리 결과물명, 처리 항목 수, 처리 데이터 건수, 처리근거, 처리자 등 * 주기적·반복적으로 처리되는 경우, 매 적재 시 마다 이력 저장 | * 가명(익명)처리시스템 운영부서 |
| **9. 처리결과 검토 요청** | * **관련 문서 작성 및 처리결과 검토 요청** * 처리계획과 실 반영에 대한 차이점 확인 요청 * (결재)소속팀장, (합의) 검토부서, (합의)개인정보처리시스템 운영부서, (결재)소속부문장 | * 가명정보이용부서 |
| **10. 결과 검토** | * **처리결과 검토** * 처리목적의 적합성 검토 * 항목별 가명처리의 적정성 검토 * 재 식별 가능성 검토 * **가명처리가 적절하게 수행되지 않은 경우, 보완 요청** | * 검토부서 |
| **11. 가명정보 사용** | * **가명정보이용부서의 가명정보 취급자가 가명정보 사용** * 가명정보별 데이터 접근권한 차등 적용 * **가명정보 사용 이력 저장** | * 가명정보이용부서 |

## 2.2 외부결합

외부결합 시, 아래 흐름 및 표를 참고하여 수행하도록 한다.



**<가명정보 처리 프로세스 외부결합 1/2>**



**<가명정보 처리 프로세스 외부결합 2/2>**

| **수행업무** | **수행업무 내용** | **주관부서** |
| --- | --- | --- |
| **1. 외부결합 계획수립** | * **외부결합 대상 기관과 결합관련 사항 협의** * 결합목적, 결합항목 선정 및 가명처리 수준 협의 * 공통으로 보유한 항목 중 결합키 대상 항목 선정 * 결합률 확인여부, 시계열 분석여부 등 | * 가명정보이용부서 |
| **2. 가명처리 수준정의** | * **가명처리 대상 개인정보 항목별 가명처리 방법 정의** * 가명(익명)처리 수준 정의표와 위험도 측정 결과 작성 (별첨4) | * 가명정보이용부서 |
| **3. 처리계획 검토 요청** | * **요청에 필요한 문서 작성 후 품의 진행** * (결재)소속팀장, (합의) 검토부서, (합의)개인정보처리시스템 운영부서 | * 가명정보이용부서 |
| **4. 처리계획 검토** | * **외부결합 처리계획 진행여부 검토** * 외부결합 목적에 적합한 가명정보 사용여부 검토 * 가명정보 처리계획에 대한 검토는 보안센터에서 진행 * 단, 가명정보 이용부서가 데이터플랫폼본부일 경우 본부 내 지정 인원이 병행 검토 진행 | * 검토부서 |
| **5. 정보전송** | * **외부결합에 사용할 정보 제공** | * 개인정보처리 시스템운영부서 |
| **6. 가명처리 요청** | * **가명(익명)처리시스템운영부서에게 개인정보에 대한 가명처리 요청** * 가명(익명)처리 수준정의표 제공 * 적정성 검토결과 추가처리가 필요한 경우 시스템운영부서에게 추가처리 요청 * 가명정보이용부서가 추가처리 대상 및 방법 제시 | * 가명정보이용부서 |
| **7. 가명처리** | * **가명정보이용부서가 요청한 방식으로 가명처리 수행** * 결합대상정보에 중복되지 않는 일련번호 생성 * 결합키 생성항목을 제외하고, 나머지 정보들에 대해 가명처리 수행 * 결합키는 결합키관리기관과 협의 후 생성 * **가명정보이용부서가 추가처리 요청 시 수행** * **가명처리(추가처리) 환경 및 데이터 파기** * 시스템운영부서의 추가정보취급자가 수행 * 본 작업은 임시공간에서 수행(사용기간 설정) * 사용기간 경과 시 또는 작업 완료 시 데이터 파기 | * 가명(익명)정보처리 시스템운영부서 |
| **8. 적정성 평가 신청** | * **관련 문서 작성 및 적정성 검토 신청** * 위험도 측정 결과서 (별첨4) * 개인정보 속성 정의 및 가명처리 계획 (별첨1, 별첨5) * 적정성평가 기초자료 (별첨6) | * 가명정보이용부서 |
| **9. 적정성 평가** | * **적성성평가 외부결합의 적정성 검토** * 실제 데이터에 대한 가명처리가 잘 수행되었는지 검토 * 재식별 가능성 검토 * **가명처리가 적절하게 수행되지 않은 경우, 추가처리 요청** * **적정성 평가 결과를 기반으로 내부 품의 진행** * (결재)소속팀장, (합의)적정성평가단, (합의)개인정보처리시스템 운영부서, (결재)소속부문장 | * 적정성평가단 |
| **10. 결합신청** | * **결합전문기관에 결합신청** * 결합전문기관 선정 후 "가명정보 결합 종합지원 시스템"을 통해 결합신청서 및 첨부서류 준비 후 제출 * 결합전문기관과 결합 일정, 전송 방법 등을 협의 * 결합키관리기관으로부터 결합키 생성에 관한 정보 수신 | * 가명정보이용부서 |
| **11. 결합키 생성** | * **결합키관리기관과 결합키 생성 방법 협의** * 결합키 생성에 사용되는 salt값, 결합키 생성에 관한 기술지원, 사전 결합률 확인여부 협의 * **결합키 생성** * 결합키 생성 항목 중 한글이 포함되어 있는 경우 인코딩 방식 확인 | * 가명(익명)정보처리 시스템운영부서 |
| **12. 정보이관** | * **결합키관리기관, 결합전문기관으로 정보를 전송하기 위해 추가정보처리시스템에서 외부전송시스템으로 전송대상 정보 이관** | * 가명(익명)정보처리 시스템운영부서 |
| **13. 정보 전송** | * **결합키관리기관에 결합키, 일련번호 전송** * "가명정보 결합 종합지원시스템”을 통해 제출 * **결합전문기관에 결합대상정보, 일련번호 전송** * 결합기관과 사전 정의된 방식으로 제출 * **정보통신망 등을 통하여 송수신 또는 보조저장매체를 통하여 전송하는 경우 압축 및 암호화, 무결성 검증 등의 보호조치 수행 필요** * **전보전송이 완료되면, 전송한 정보 파기** | * 가명정보이용부서 |
| **14. 결합물 확인** | * **결합키관리기관을 통해 결합물 확인**(신청 시 확인 가능) | * 가명정보이용부서 |
| **15. 추가처리** | * **“결합전문기관” 내에서 가명정보이용부서가 데이터 추가 가명 익명처리 수행** | * 결합전문기관 |
| **16. 반출요청** | * **결합된 가명정보 반출요청** (별첨3) | * 가명정보이용부서 |
| **17. 외부결합결과물 수신** | * **외부전송시스템을 통해 외부결합결과물 수신** | * 가명정보이용부서 |
| **18. 외부결합결과물 적재** | * **외부결합결과물 적재 시 메타 데이터 저장** * 가명처리 결과물(데이터셋 별) 사용목적, 항목별 개인정보 유형, 속성(고유식별자, 준식별자 등), 파기 예정일 등 관리 * **외부결합결과물 적재이력 저장** * 처리일시, 처리유형, 가명처리 결과물명, 처리항목 수, 처리데이터건수, 처리근거, 처리자 등 * 주기적·반복적으로 처리되는 경우, 매 적재 시 마다 이력저장 | * 가명(익명)정보처리 시스템운영부서 |
| **19. 외부결합결과물 사용** | * **가명정보이용부서의 가명정보 취급자가 가명정보 사용** * 가명정보 별 데이터 접근권한 차등 적용 * **가명정보 사용이력 저장** | * 가명정보이용부서 |
| **20. 외부결합결과물 파기** | * **목적 달성이 완료된 외부결합결과물 파기처리** * 가명정보이용부서가 목적달성을 완료 후 파기 요청 시 * 데이터 파기 예정일 도래 시 * 보안센터에서 정기점검 결과로 파기 요청 시 * **외부결합결과물 파기 이력 저장** | * 가명(익명)정보처리 시스템운영부서 |

# 3. 위험도 측정 및 가명처리 수준 정의

## 3.1 위험도측정

위험도 측정은 데이터 이용 부서에서 수행하며, 가명처리 대상 데이터의 식별 위험성을 평가하여 이에 따라 적절한 가명처리 수준을 결정하고자 한다.

위험도 측정 시, 데이터 위험성과 처리 환경의 위험성 평가로 구분한다.

위험도 측정 결과는 가명정보 처리계획 검토 요청 시, 첨부하여야 하며 검토 의견은 가명정보 처리 시 반영하여야 한다.

## 3.2 데이터 위험성 평가

데이터 위험성은 내부활용 및 외부제공은 가명처리 전 데이터, 결합의 경우 결합 전 데이터를 대상으로 평가한다.

데이터 위험성평가는 식별 요소 유무, 식별 가능성, 특이정보, 재 식별 시 영향도 항목으로 구분하여 작성자는 상세 내용을 작성해야 한다.

| **점검항목** | **점검기준** |
| --- | --- |
| **식별자(식별정보)** | 그 자체만으로 개인식별성이 강한 정보 포함여부  - 예시) 고유식별정보(주민등록번호, 여권번호, 외국인번호, 운전면허번호), 성명, ID, CI, 개인휴대전화번호, 개인이메일, 거주지주소(상세) IP, 디바이스ID, 차량번호 등 |
| **준식별자(식별 가능 정보)** | 2개 이상의 항목(컬럼)과 조합하거나, 다른 내/외부(공개된 데이터 포함)데이터와 결합하면 특정 개인을 식별하기 쉬운 정보 포함여부  - 예시) 성별, 나이, 생년월일, 거주지역, 국적, 직업, 위치정보, 결혼기념일, 디바이스 모델명, 고객등급(VIP) 등  - 예시) 데이터셋 내 계층적 특성(가족관계, 직책관계 등)을 가진 정보가 포함되어 개인 식별가능성이 높아질 수 있는 경우(회사내 정보 분석 시 해당 데이터셋에 소수직책이 포함되어 있는 경우) 등  - 예시) 시간, 위치, 행위 등 정보를 함께 분석하는 경우 등 |
| 데이터셋의 크기가 적어 식별 가능성 여부  - 예시) 분석 대상이 특정 집단으로 한정된 인원이라서 식별 가능성이 높아지는 경우 등 |
| 원본 데이터 전체가 아닌 일부 데이터 처리(샘플링) 여부  - 예시) 상품 구매이력 분석에서 특정 고객층의 전체 중 일부만 샘플링하여 분석하는 경우 등 |
| 시계열 성격을 가진 데이터 포함여부  - 예시) 시간, 일, 주, 월, 반기 등의 주기로 분석하는 경우 |
| **특이정보** | 가명처리 대상 데이터에 식별 가능성이 있는 고유(희소) 정보 포함여부  - 예시) 희귀 성씨, 희귀 병명 등 정보자체로 특이한 값을 가지고 있는 정보  - 예시) 데이터 대상정보 중 직업에 개인 식별성이 매우 높거나(지방자치단체장, 국가대표감독/선수 등), 도서/산간지역 주민의 만14세 미만의 정보 등 특정 데이터 분석 집단의 희소한 값  - 예시) 일반적인 문자형 데이터 중 특정 값으로 현저히 낮은 정보를 가지고 있는 경우 등 |
| 가명처리 대상 데이터 중 편중된 분포를 가지고 있는 정보 여부  - 예시) 전체 데이터 분포의 양 끝의 정보가 현저히 낮은 정보(연령 분포상 14세미만, 90세 이상 등) |
| **재 식별 영향도** | 데이터가 지니는 특성만으로 재 식별 시, 특정 정보주체에게 사회적 파장 등 영향도가 높은 정보 포함여부  - 사회통념상 차별정보 등으로 정보주체가 피해 또는 불이익을 받을 수 있는 정보 등 |

## 3.3 처리 환경의 위험성 평가

처리 이용환경은 가명정보 이용•제공 형태, 처리환경, 처리방법에 대해 위험성 평가를 실시한다.

작성자는 항목별 해당되는 사항을 가능한 상세히 작성하여야 하며, 평가 시 필요할 경우 보안서약서, 계약서, 협약서 등을 통해 파악 가능한 범위의 정보를 고려해야 한다.

| **점검항목** | **점검기준** |
| --- | --- |
| **이용•제공 형태** | 가명정보 이용 유형 및 처리주체 구분  - 예시) 가명정보 이용 유형 : 내부 활용/결합, 외부결합, 외부제공 등  - 예시) 처리주체 : 회사임직원(개인정보처리 동일부서/타 부서 등), 외부 제공(제3자 제공) |
| 처리주체가 보유하고 있는 정보 또는 접근/입수 가능한 정보 등 다른 정보와 가명정보 활용 범위 및 유형을 고려했을 때, 식별 가능한 항목 있는지 여부  - 예시) 가명정보를 반복적으로 이용/제공 시, 반복되는 가명정보를 통해 식별 위험이 높아 있는 항목이 있을 경우  - 예시) 가명정보 취급자가 보유하고 있는 정보 또는 접근/입수 가능한 정보 등 다른 정보가 처리하는 가명정보 집합물에 포함되어 있는 경우 |
| 추가정보를 삭제하지 않고 보관 여부  - 추가정보는 가명정보와 분리하여 보관해야 하며, 삭제하거나 암호화하여 보관 필요 |
| (외부결합/제공 시) 처리주체가 외부 처리자/취급자 일 경우, 처리주체 기관의 개인정보보호 수준(신뢰 가능한)  - 예시) ISMS/ISMS-P인증, ISO27001/ISO27701 등 |
| **처리환경** | 가명정보 처리하는 장소/시스템이 다른 정보를 접근/입수할 수 있는지 여부  - 예시) 누구나/내부인원만/가명정보처리/취급자만 접근 가능한 장소/시스템/네트워크  ※ 장소/시스템/네트워크 등의 환경은 회사 내 자산 또는 외사 외 자산인지 구분 필요 |
| **처리방법** | 다른 정보와의 연계 분석이 예정되어 있는지 여부 |
| 처리주체가 보유하고 있는 정보 또는 접근/입수 가능한 정보 등 다른 정보와 연계/결합하여 식별 가능한 항목 있는지 여부  - 예시) 처리한 가명정보 데이터 + 기존 업무의 데이터를 통해 식별 위험이 높아지는 항목이 있는 경우  - 예시) 내/외부 이용기관 또는 처리자가 보유하고 있는 과거 유사정보에 대한 수행 경험이나 지식 등  ※ 가명정보 처리할 담당자에게 사전 확인이 필요하며, 확인이 어려울 경우, 가명처리 수준을 높여야 함 |

## 3.4 가명처리 수준 정의

가명처리 수준은 개인정보항목별 개인정보속성, 데이터유형 등에 따라 달라진다.

본 매뉴얼에서는 기본적인 가명(익명)처리 수준을 제시하는 것으로 개인정보 항목별, 개인정보속성, 데이터 유형 등 기본적인 가명처리 수준은 ‘별첨5. 개인정보 속성 정의표’ 참조하여 작성한다.

가명정보이용부서는 1차적으로 개인정보 속성에 따라 가명(익명)처리 수준을 정의하고, 업무상 필요에 따라 가명처리 수준을 낮추거나 높이는 것은 ‘적정성 평가’ (내부활용/결합 시 처리결과 검토)결과에 따른다.

* 1. 가명(익명)처리 시 민감정보는 사용할 수 없으나, 업무상 사용이 필요한 경우 ‘적정성평가’결과(내부활용/결합 시 처리결과 검토 결과)에 따라 예외처리 하여 사용할 수 있다.
  2. 익명처리 시, 식별자는 삭제가 원칙이나, 업무상 사용이 필요한 경우 ‘적정성평가’결과(내부활용/결합 시 처리결과 검토 결과)에 따라 예외처리 하여 사용할 수 있다.
  3. 가명(익명)처리 시, 본 매뉴얼에서 제시하는 가명(익명)처리 수준보다 낮게 적용하고자 하는 경우 ‘적정성 평가’결과(내부활용/결합 시 처리결과 검토 결과)에 따라 예외처리 하여 사용할 수 있다.

가명처리 시, 민감정보는 삭제를 원칙으로 한다.

1. 민감정보 : 사상/신념, 노동조합, 정당 가입/탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보로서 그 밖에 정보주체의 사생활을 현저히 침해할 수 있는 정보(개인정보 보호법 제23조)

※ ‘개인정보 영향도 등급표’의 1등급에 해당되는 민감정보, 의료정보, 인증정보-바이오정보(지문, 홍채, 정맥 등)가 해당됨

1. 장애인여부, 신용등급, 신용불량여부 등은 민감정보로 가명처리 시 삭제
2. 민감정보로 지정하지 않았으나 정보주체의 위치정보, 종교적/정치적 서향을 파악할 수 있는 내용 등은 가명처리 시 주의가 필요

개인정보속성은 다음과 같다.

| **개인정보속성** | **설명** | **개인정보 영향도 등급표** |
| --- | --- | --- |
| **식별자 (식별정보)** | * 정보주체에게 고유하게 부여되는 정보로, 개인을 식별할 수 있는 정보 * 주민등록번호, 이름, 사진, 전화번호, 이메일주소, IP 등이 식별자에 포함됨 * 회사 내부에서 부여한 고객번호, 멤버십번호, 가입번호 등 내부에서 개인을 바로 식별할 수 있는 정보도 식별정보에 포함됨 * 가명처리 시 삭제 원칙 * 업무 필요에 따라 사용 필요 시, 재식별이 불가능하도록 적용(일방향 암호화 등)하여 가명처리 | * 1등급: 고유식별정보, 인증정보(ID), 금융정보 * 1.5등급: 개인식별정보 * 2등급: 개인식별정보(이름, 이메일) * 3등급: 자동생성정보, 제한적 본인 식별정보 (회원번호) |
| **준식별자**  **(식별가능정보)** | * 개인을 식별할 수 있는 정보는 아니지만, 다른 정보와 결합하면 특정 개인을 식별하기 용이한 정보 * 생년월일, 직업, 성별, 주소(상세주소는 식별자 임) 등 준식별자에 포함됨 | * 1등급: 위치정보 * 2등급: 개인식별정보(생년월일, 성별), 개인관련정보 |
| **속성정보** | * 회사에서 보유하고 있는 정보 중 식별자, 준식별자 외의 정보로서 임직원 이외의 자는 다른 정보와 결합해도 특정 개인을 식별하기 어려운 정보 * 멤버십 등급/유형, 가입일시, 판매금액 등이 속성정보에 포함됨 | - |

데이터 유형은 다음과 같다.

| **데이터 유형** | **설명** |
| --- | --- |
| **고유식별정보** | * 개인정보 보호법 제24조에 의해 개인을 고유하게 구별하기 위해 부여된 식별정보 * 주민등록번호, 운전면허번호, 여권번호, 외국인등록번호 * 가명처리 대상이 될 수 없음. 삭제 원칙 적용 |
| **고유식별정보 외** | * 식별자 중 고유식별정보를 제외한 데이터 |
| **binary** | * 값의 카테고리(범위)가 2가지를 가지고 있는 데이터 * 성별, 외국인여부, 가족유무 등 * 해당 값만으로는 개인 식별하기 어려워 그대로 가용 가능하나 외부 정보 노출 시 위험성이 있을 경우 코드화 적용 가능 * 예시) 성별 여 🡪 A, 성별 남 🡪 B |
| **명목형** | * 데이터 값의 카테고리(범위)가 정의되어 있는 데이터 * 멤버십 등급, 직업, 통신사, 학력 등 * 가명처리 시 특정 데이터의 빈도가 낮은 경우, 추가처리 대상이 될 수 있음 * 예시) 멤버십 등급 VIP에 해당되는 데이터가 가명처리대상 데이터 중 1% 미만인 경우 |
| **문자** | * 한글, 영문, 숫자의 형태들이 복합적으로 형성된 데이터 * 생년월일+성별, 주소, 우편번호, 직장전화번호 등 |
| **숫자** | * 수치형 데이터 * 나이, 구매금액, 자녀 수 등 * 평균적인 데이터 값을 벗어나는 최소/최대 값은 추가처리 대상이 될 수 있음 * 예시) 나이 95세 이상 |
| **날짜** | * 년, 월, 일, 시간으로 표현되는 데이터 * 예시) YYYY-MM-DD, YYYY-MM-DD HH:MM:SS 등 |
| **비정형** | * 사람이 직접 입력하여 형태와 구조가 정의되지 않은 데이터 * 민원처리내용, 검색어 등 * 비정형 데이터 내부에 저장된 모든 개인정보를 식별하여 삭제하는 것은 현 기술 수준으로 적용하기 어려움으로 가명처리 시 삭제 원칙임 * 단, 비정형데이터 내의 식별자, 준식별자를 가명처리하고 기간별 데이터의 재 식별 가능성에 대한 추가 평가가 완료된 경우 이용 가능함(보안센터 협의 하에 검토 및 예외처리 필요) |

개인정보 속성 데이터 유형에 따른 기본적인 가명처리 수준은 다음과 같으며, 상세내용 및 예시는 ‘별첨5. 개인정보 속성 정의표’ 참조한다.

| **개인정보속성** | **데이터 유형** | **가명처리** | **익명처리** |
| --- | --- | --- | --- |
| **식별자(식별정보)** | 고유식별정보 | 삭제 | 삭제 |
| 고유식별정보 외 (회원번호 포함) | 삭제 원칙, 필요 시 암호화 등 식별불가 조치 | 삭제 |
| **준식별자(식별가능정보)** | binary | 유지 | 유지 |
| 명목형 | 유지 | 유지 |
| 문자 | 마스킹(내부 마스킹 기준 적용) | 마스킹(내부 마스킹 기준 적용) |
| 숫자 | 범주화 또는 라운딩 | 범주화 또는 라운딩 |
| 날짜 | 마스킹(일 단위) | 마스킹(월 단위) |
| **속성정보** | binary | 유지 | 유지 |
| 명목형 | 유지 | 유지 |
| 문자 | 유지 | 유지 |
| 숫자 | 유지 | 유지 |
| 날짜 | 유지 | 유지 |
| 비정형 | 삭제 | 삭제 |

가명(익명) 처리 수준이 결정되면, ‘별첨4. 위험도측정결과\_개인정보속성정의’에 개인정보 항목별 가명처리 방법을 작성한다.

# 4. 적정성평가

## 4,1 적정성평가

적정성평가는 외부결합/제공 시 개인식별 가능성과 가명처리의 적정성에 대한 검토를 수행하는 것으로 내부 인원을 활용하여 자체적으로 검토하거나 외부 전문가를 통하여 검토할 수 있다.

내부 활용/결합 시, 가명정보 처리계획 검토를 수행한 부서의 인원이 적정성평가 수준의 가명정보 처리결과 검토를 진행하여야 한다.

적정성평가 시, 다음을 포함하여 검토하여야 한다.

1. 가명처리의 적정성 : 가명처리 수준 정의한 것에 따라 적절히 가명처리 되었는지 확인(가명정보 항목 전체 검토)
2. 목적달성 가능성 : 생성한 가명정보가 초기 가명정보 활용 목적에 달성할 수 있는지 확인
3. 특이정보 처리 : 항목별 위험도 바탕으로 가명처리 시에도 특이정보를 통해 개인 식별이 가능여부 확인

데이터를 가명처리 하여 활용하고자 하는 주관부서에서 기초자료를 작성하여 평가 요청하며, 평가는 각 담당자 및 전문가를 통해 검토 진행되고 의견을 취합하여 최종 평가 결정을 한다.

1. 기초 자료는 ‘별첨6. 적정성평가 기초자료’을 참고하여 작성한다.

평가를 요청하는 주관부서는 가명처리한 데이터를 전문기관을 통한 결합 없이 대외 제공하는 경우, 가명정보 수요기관에 대한 가명정보 이용 및 관리환경을 검토한 의견을 추가로 첨부하여야 한다.

평가 의견에 따라 가명처리 절차를 재 수행하거나 추가 가명처리 하여 적정성평가를 재 요청하여야 한다.

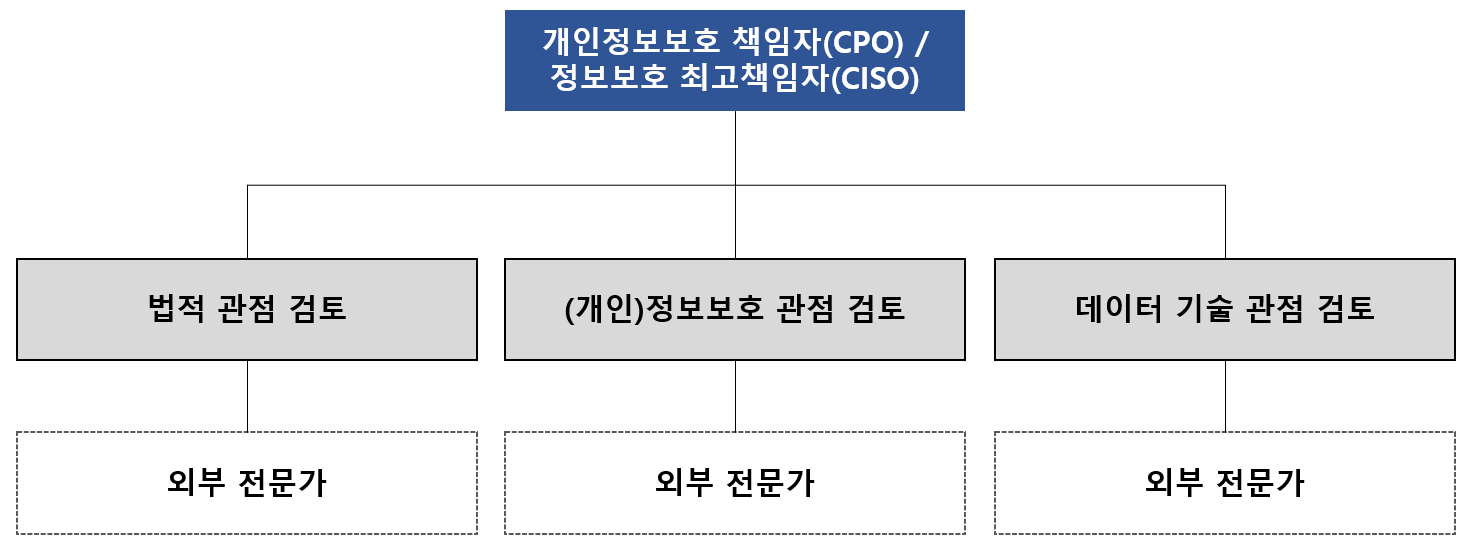
## 4.2 적정성평가단 구성

개인정보 보호책임자 또는 정보보호 최고책임자의 지휘 하에 데이터 적정성평가단을 구성한다.

1. 3명 이상의 관련 분야(법률/보안/기술)에 따라 선임하며, 외부 전문가를 포함하여 구성할 수 있다.

- 법적 관점의 검토는 평가단 선정하지 않고 회사 법무팀의 자문을 받은 것으로 대체할 수 있다.

1. 평가단은 데이터 이용 목적과 직접적 이해관계가 없는 자로 구성한다.
2. 평가단의 장은 평가단 운영과 관련된 전반적인 사항을 관장하며, 각 분야별 검토 의견을 취합하여 최종 결정을 통보한다.



# 5. 가명정보의 안전한 관리

## 5.1 가명(익명)정보 이용 및 제공

적정성평가에 한하여 내부 활용 및 외부 제공이 가능하며, 관련 법 및 개인정보보호 업무지침에 따라 안전한 관리를 이행하여야 한다.

익명정보는 더 이상 개인정보로 추정하지 않더라도 불특정 다수에게 공개하는 것은 제한하여야 한다.

| **관리유형** | **관리방안** |
| --- | --- |
| **추가정보** | * 추가정보는 가명정보와 분리하여 별도 저장 및 관리   - 물리적 분리가 어려울 경우, 논리적 분리 가능하며 논리적 분리 시 엄격한 접근통제 적용  - DB접근제어가 적용되는 환경 내 저장   * 추가정보는 목적 달성 시, 삭제 또는 암호화   - 삭제 시, 삭제 내역을 기록하여야 함 |
| **접근 통제** | * 가명정보, 추가정보에 접근할 수 있는 인원을 가명정보 처리업무에 필요한 최소한 인원으로 제한   - 업무에 따라 접근 권한 차등 부여  - 가명정보에 접근 가능한 인원은 원본 데이터에 대한 접근권한 미 부여  - 휴직/퇴직 등의 인사이동으로 가명정보 처리하는 인원이 변경되었을 경우, 가명정보처리시스템의 접근권한 변경/삭제  - 가명정보처리시스템의 접근 가능한 계정은 개인정보취급자와 공유되지 않도록 유의   * 이외 가명정보처리시스템 접근통제 정책은 시스템 운영 및 개인정보처리시스템 규칙 준수 |
| **재 식별 통제** | * 가명정보 처리 부서는 재 식별 가능성이 존재 시, 가명정보 처리를 중단하고, 재 식별 가능한 가명정보 파기   - 가명정보를 제공받은 업체가 재 식별 가능성 존재 판단 시, 회사의 정보제공 부서에 해당 사실을 즉시 통보하고 제공부서에서 요청하는 조치 사항 처리   * 재 식별 경위와 내역, 조치 내역을 정보보호 부서에 통보   - 파기 수행한 결과를 주관부서의 장 승인을 받아야 하며, 정보보호 부서는 파기 확인 및 점검 결과를 개인정보보호 책임자에게 보고 |
| **보관** | * 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보의 이용내역, (제3자 제공 시) 제공받은 자 기록 보관 * 가명정보처리시스템의 접근권한 부여, 변경, 삭제 내역 최소 3년간 보관 |
| **정기 점검** | * 가명정보 및 추가정보의 안전한 관리를 위해 정기적 점검 수행 |

## 5.2 가명정보 제공 시 유의사항

가명정보 또는 결합정보 제공 시, 가명정보 제공 및 활용 관련 협약서와 보안관리 약정서를 체결하도록 해야 하며, 아래의 내용을 포함하여야 한다.

1. 업무 목적 외 사용 및 제3자 제공 금지

- 제공받은 가명정보를 목적에 맞는 업무 외에 다른 용도로 사용될 수 없으며, 제3자에게 임의 제공 또는 누설할 수 없다.

- 제공받은 가명정보를 회사의 사전 승인 없이 업무 목적 외 별도 저장, 출력, 복사, 가공할 수 없다.

1. 적정성평가

- 가명정보가 적절하게 가명처리 되었는지 확인될 수 있도록 적정성평가를 시행해야 하며, 개인식별요소가 있을 경우 제공받은 정보의 처리를 즉시 중단하고 이를 파기하여야 한다.

1. 재 식별 금지

- 업무 목적으로 안전하게 이용하여야 하며, 개인 식별하기 위한 어떠한 행위도 해서는 안 된다.

- 가명정보가 재 식별되거나 재 식별 가능성이 높아지는 상황 발생 시, 회사로부터 제공받은 가명정보의 처리 중지하고 관련사항을 즉시 통지하여야 한다.

- 제공한 가명정보의 재 식별 가능성을 정기적 모니터링 해야 한다.

- 재 식별 가능성이 발견 시 추가적인 가명 조치 이행을 위해 회사에 요구할 수 있으며, 회사는 합리적 사유가 없는 한 이를 거절할 수 없다.

1. 재 제공/위탁 제한

- 재 제공/위탁 허용 시, 회사 제공 정보에 대한 또는 관리상 어려움(재 위탁 사실 누락 등)이 존재할 수 있음으로 재 제공/위탁을 금지한다.

- 부득이하게 필요 시, 관련 부서와 협의하여야 하며 재 제공/위탁 가능 범위 및 재 식별 방지 조치 등을 계약서에 명시하여야 한다.

1. 안전성 확보조치

- 가명정보의 안전성 확보 및 가명정보 유출 시 보호에 필요한 관리적, 기술적 보호조치를 취해야 한다.

1. 이용자에 대한 관리감독

- 가명정보를 제공하는 회사의 가명정보 처리현황 등의 사항을 관리하도록 요구할 수 있으며, 제공받은 회사는 특별한 사유가 없는 한 이에 응해야 한다.

- 제공 회사는 가명정보 처리 실태를 점검 및 결과 시정 요구를 할 수 있으며, 특별한 사유가 없는 한 시정 요청 사항을 이행해야 한다.

(가명정보 처리현황, 가명정보 접근 또는 접속현황, 가명정보 접근 인원, 목적 외 이용제공 및 재 위탁 금지 준수, 재 식별 시도 금지 준수, 안전성 확보조치 준수, 그 외 개인정보 및 가명정보 보호를 위하여 필요한 사항)

1. 가명정보 파기

- 제공받은 가명정보에 대하여 이용 목적 달성하거나 해당 가명정보 불필요하게 된 경우 지체없이 해당 가명정보 파기하여야 한다.

- 가명정보 파기 시 파기 확인서를 작성하여 가명정보를 제공한 회사에 송부하여야 한다.

1. 재 식별 시 제재조치 및 손해배상

- 협약에 규정된 의무 위반 또는 관련 법률 규정 위반하여 회사 또는 제3자에게 손해 발생한 경우 손해배상에 관한 사항을 포함하여야 한다.

# 별첨.

| **No.** | **목차** | **별첨** | **자료** |
| --- | --- | --- | --- |
| 1 | 2. 가명정보 처리절차 | 별첨1. 가명처리계획서 |  |
| 2 | 2.3 외부결합 | 별첨2 가명정보 결합신청서 |  |
| 3 | 2.3 외부결합 | 별첨3. 반출신청서 |  |
| 4 | 3.1 위험도측정 | 별첨4. 위험도측정결과\_개인정보속성정의 |  |
| 5 | 3.4 가명처리 수준 정의 | 별첨4. 위험도측정결과\_개인정보속성정의 |  |
| 6 | 3.4 가명처리 수준 정의 | 별첨5. 개인정보 속성 정의표(참고용) |  |
| 7 | 4.1 적정성평가 | 별첨6. 적정성평가 기초자료 |  |
| 8 | 4.1 적정성평가 | 별첨7. 적정성평가항목 |  |

# 붙임1. 가명처리 기법

## 개인정보의 가명·익명처리 기술 종류

[가명정보처리 가이드라인(2022.04, 개인정보보호위원회) 참고]

* 아래 분류는 이해를 돕기 위해 IS0/IEC 20889, 그리고 EU ENISA(European Union Agency for Network and Information Security)에서 발간한 보고서 등 국내·외 자료들을 참고하여 작성했으며 표준이 아님

| **분류** | **기술** | **세부기술** | **설명** |
| --- | --- | --- | --- |
| 개인정보 삭제 | 삭제기술 | 삭제  (Suppression) | * 원본정보에서 개인정보를 단순 삭제 |
| 부분삭제  (Partial suppression) | * 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제 |
| 행 항목 삭제  (Record suppression) | * 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제 |
| 로컬 삭제  (Local suppression) | * 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제 |
| 개인정보 일부 또는 전부 대체 | 삭제기술 | 마스킹  (Masking) | * 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제 |
| 통계도구 | 총계처리  (Aggregation) | * 평균값, 최댓값, 최솟값, 최빈값, 중간값 등으로 처리 |
| 부분총계  (Micro aggregation) | * 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리. 즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체 |
| 일반화(범주화) 기술 | 일반 라운딩  (Rounding) | * 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법으로, 일반적으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용 |
| 랜덤 라운딩  (Random rounding) | * 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법 |
| 상하단코딩  (Top and bottom coding) | * 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있음 * 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법 |
| 로컬 일반화  (Top and bottom coding) | * 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상 의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법 |
| 범위 방법  (Data range) | * 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현 |
| 문자데이터 범주화  (Categorization of  character data) | * 문자로 저장된 정보에 대해 보다 상위의 개념으로 범주화하는 기법 |
| 암호화 | 양방향 암호화  (Two-way encryption) | * 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법 * 암호화 및 복호화에 동일 비밀키로 암호화하는 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 비대칭키(Asymmetric key) 방식으로 구분 |
| 일방향 암호화·암호화적 해시함수  (One-way encryption- Cryptographic hash function) | * 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법 * 키가 없는 해시함수(MDC, Message Digest Code), 솔트(Salt)가 있는 해시함수, 키가 있는 해시함수(MAC, Message Authentication Code)로 구분 * 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력 값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음 |
| 순서보존 암호화  (Order-preserving encryption) | * 원본정보의 순서와 암호 값의 순서가 동일하게 유지되는 암호화 방식 * 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능 |
| 형태보존 암호화  (Format-preserving encryption) | * 원본 정보의 형태와 암호화된 값의 형태가 동일하게 유지되는 암호화 방식 * 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음 * 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용 |
| 동형 암호화  (Homomorphic encryption) | * 암호화된 상태에서의 연산이 가능한 암호화 방식으로 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능 * 암호화된 상태의 연산 값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법 |
| 다형성 암호화  (Polymorphic encryption) | * 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법 * 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨 |
| 무작위화 기술 | 잡음 추가  (Noise addition) | * 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법 |
| 순열(치환)  (Permutation) | * 분석 시 가치가 적고 식별성이 높은 열 항목에 대해 대상 열 항목의 모든 값을 열 항목 내에서 무작위로 순서를 변경하여 식별성을 낮추는 기법 * 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법 |
| 토큰화  (Tokenisation) | * 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 식별 위험을 제거하여 개인정보를 보호하는 기술 * 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 양방향 암호화, 형태보존 암호화 기법을 주로 사용 |
| (의사)난수생성기  ((P)RNG, (Pseudo) Random Number Generator) | * 주어진 입력 값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보와 대체 |
| 가명·익명처리를 위한 다양한 기술  (기타 기술) | | 표본추출  Sampling) | * 데이터 주체별로 전체 모집단이 아닌 표본에 대해 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법 |
| 해부화  (Anatomization) | * 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술 |
| 재현데이터  (Synthetic data) | * 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법 |
| 동형비밀분산  (Homomorphic secret sharing) | * 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)\*로 대체 \*기밀사항을 재구성하는데 사용할 수 있는 하위 집합 |
| 차분 프라이버시  (Differential privacy) | * 특정 개인에 대한 사전지식이 있는 상태에서 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법 * 1개 항목이 차이나는 두 데이터베이스 간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델 |

## 개인정보의 가명·익명처리 예시

※ 아래 모든 예시는 각 기법의 적용에 대한 예시이며 전체 데이터에 대한 가명·익명처리에 대한 예시가 아님

## 2.1 개인정보 삭제

**• 삭제기술: 선택된 항목을 제거하는 기술**

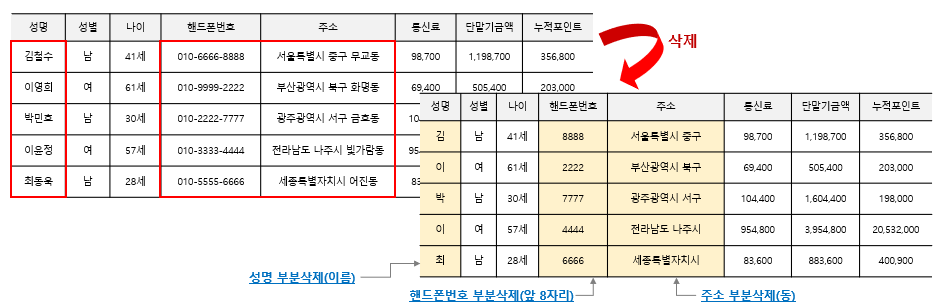
* 1. 삭제(Suppression)

※ 이때 남아 있는 정보 그 자체로도 분석의 유효성을 가져야 함과 동시에 개인을 식별할 수 없어야 하며, 인터넷 등에 공개되어 있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 함



* 1. 부분삭제(Partial suppression)

- 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제



* 1. 행 항목 삭제(Record suppression)

- 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제

- 통계분석에 있어서 전체 평균에 비하여 오차범위를 벗어나는 자료를 제거할 때 사용



* 1. 로컬 삭제(Local suppression)

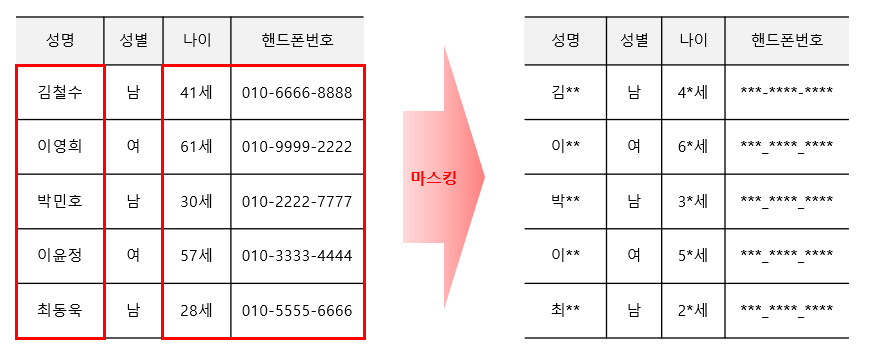
- 특이정보를 해당 행 항목에서 삭제



* 1. 마스킹(Masking)

- 특정 항목의 일부 또는 전부를 공백 또는 문자(‘\*’ , ‘\_ ’ 등이나 전각 기호)로 대체

※ 분류는 개인정보 일부 또는 전부 대체로 분류되지만, 기술적으로 마스킹된 부분은 데이터로써의 가치가 없어져 일부 문건에서는 삭제로 분류되기도 함



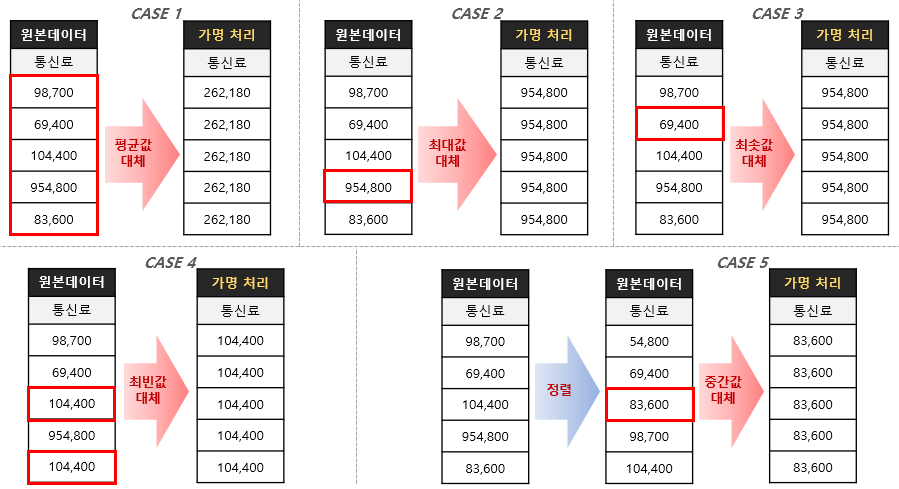
## 2.2 개인정보 일부 또는 전부 대체

**• 통계도구: 데이터의 전체 구조를 변경하는 통계적 성질을 가진 기법**

* 1. 총계처리(Aggregation)

- 평균값, 최댓값, 최솟값, 최빈값, 중간값 등으로 처리

※ 단, 데이터 전체가 유사한 특징을 가진 개인으로 구성되어 있을 경우 그 데이터의 대푯값이 특정 개인의 정보를 그대로 노출시킬 수도 있으므로 주의 필요



* 1. 부분총계(Micro Aggregation)

- 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리 즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체

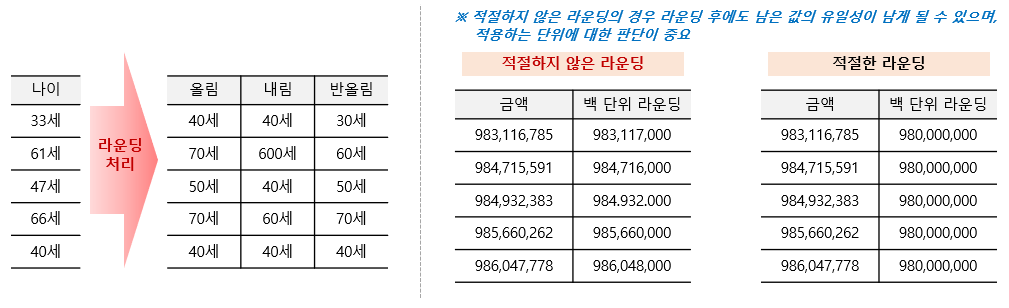
- 동질 집합 내의 특정 항목을 총계처리 하거나 특정 조건에 너무 특이한 값이 있어 개인의 식별 가능성이 높지만 분석에 꼭 필요한 값인 경우 처리



**• 일반화기술: 범주화로도 불리며, 특정한 값을 상위의 속성으로 대체**

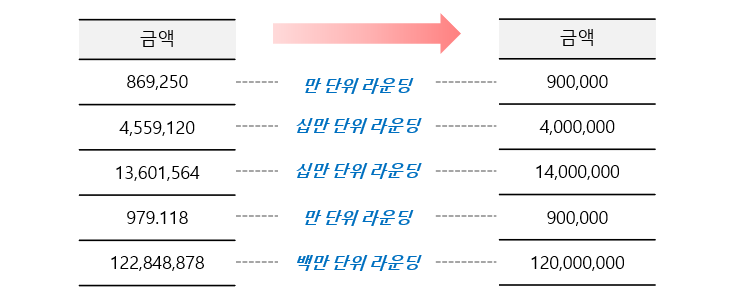
* 1. 라운딩(Rounding)
  2. 일반 라운딩

- 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법



* 1. 랜덤 라운딩(Random Rounding)

- 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법



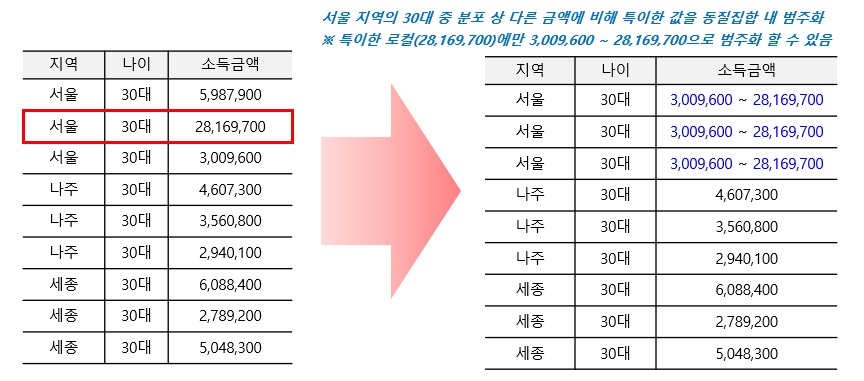
* 1. 상하단코딩(Top and bottom coding)

- 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있으며, 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법

|  |  |
| --- | --- |
|  |  |

* 1. 로컬 일반화(Local generalization)

- 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법



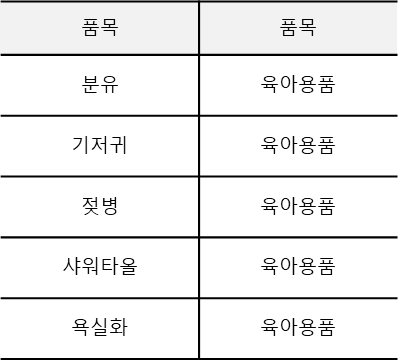
* 1. 범위방법(Data range)

- 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현

(예시) 소득 3,300만원을 소득 3,000만원~4,000만원으로 대체 표기

* 1. 문자데이터 범주화(Categorization of character data)

- 문자로 저장된 정보에 대해 상위의 개념으로 범주화하는 기법



**• 암호화: 정보 가공 시 일정한 규칙의 알고리즘을 적용하여 대체**

* 1. 암호화(Encryption)

※ 암호화에 따른 세부적인 내용은 한국인터넷진흥원 암호이용활성화 관련 안내서 참조

* 1. 양방향 암호화(Two-way encryption)

- 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법

- 암호화 및 복호화에 동일한 비밀키로 암호화하는 AES, ARIA 등 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 RSA 등 비대칭키(Asymmetric key) 방식으로 구분되며, 키(key) 관리에 주의 필요

* 1. 일방향 암호화 – 암호학적 해시함수(One-way encryption-Cryptographic hash function)

- 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법

- 키가 없는 해시함수(MDC, Message Digest Code), 키가 있는 해시함수(MAC, Message Authentication Code), 솔트(Salt)가 있는 해시함수로 구분

- 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음

* 1. 순서보존 암호화(Order-preserving encryption)

- 원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식

- 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능

* 1. 형태보존 암호화(Format-preserving encryption)

- 원본 정보의 형태와 암호화된 암호값의 형태가 동일하게 유지되는 암호화 방식

- 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음

- 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용

* 1. 동형 암호화(Homomorphic encryption)

- 암호화된 상태에서의 연산이 가능한 암호화 방식

- 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능

- 암호화된 상태의 연산한 값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법

* 1. 다형성 암호화(Polymorphic encryption)

- 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법

- 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨

**• 무작위화기술: 속성의 값을 원래의 값과 다르게 변경**

* 1. 잡음 추가(Noise addition)

- 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법

- 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않으나, 잡음값은 데이터 값과는 무관하기 때문에 유효한 데이터로 활용하기 곤란하여, 중요한 종적정보는 동일한 잡음을 사용해야함  
(예시: 입원일자에 +3이라는 노이즈를 추가하는 경우 퇴원일자에도 +3이라는 노이즈를 부여해야 전체 입원일수에 변화가 없음)



* 1. 순열(치환)(Permutation)

- 기존 값은 유지하면서 개인이 식별되지 않도록 데이터를 재배열하는 방법

- 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법

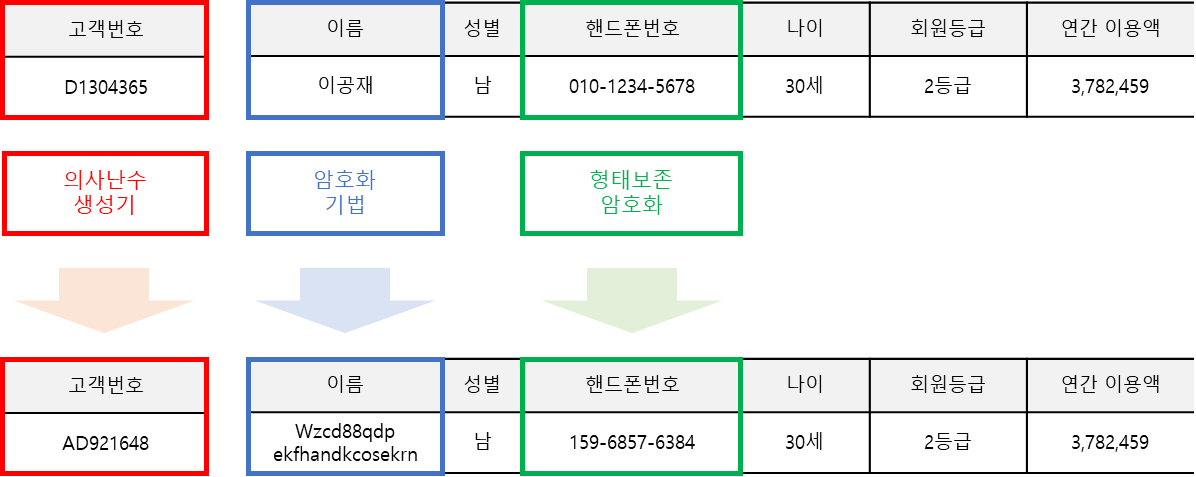
※ 데이터의 훼손 정도가 매우 큰 기법으로 무작위로 순서를 변경하는 조건 선정에 주의 필요



* 1. 토큰화(Tokenisation)

- 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 개인에 대한 식별 위험을 제거하여 개인정보를 보호하는 기술

- 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 일방향 암호화, 순서보존 암호화 기법을 주로 사용



* 1. (의사)난수생성기((P)RNG, (Pseudo) Random Number Generator)

- 주어진 입력 값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보에 할당

※ 난수는 원칙적으로 규칙적인 배열순서가 없는 임의의 수를 의미하며 컴퓨터는 원천적으로 입력에 의한 처리 결과를 반환하는 것으로 처리의 방법과 입력이 동일하면 항상 동일한 출력이 발생하기 때문에 완전한 난수의 생성은 불가능

## 2.3 가명·익명처리를 위한 다양한 기술(기타 기술)

* 1. 표본추출(Sampling)

- 데이터 주체별로 전체 모집단이 아닌 표본에 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법

- 확률적 표본추출 방법과 비확률적 표본추출 방법으로 나누어지며, 확률적 표본추출이 통계적 분석에 많이 사용

- 확률적 표본추출 : 무작위 표본추출(복원 표본추출, 비 복원 표본추출), 계통적 표본추출, 층화 표본추출, 집락 표본추출 등

- 비확률적 표본부출 : 임의 표본추출, 판단 표본추출, 할당 표본추출, 누적 표본추출

* 1. 해부화(Anatomization)

- 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술



* 1. 재현데이터(Synthetic data)

- 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법

※ 원본 데이터 포함 여부에 따라 완전 재현 데이터(Fully Synthetic Data), 부분 재현 데이터(Partially Synthetic Data), 하이브리드 재현 데이터(Hybrid Synthetic Data)로 구분

* 1. 동형비밀분산(Homomorphic secret sharing)

- 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)**\***로 대체

**\*** 기밀사항을 재구성 하는데 사용할 수 있는 하위 집합

※ 재식별은 가명·익명처리된 데이터의 쉐어를 소유한 모두가 동의하는 경우만 가능

* 1. 차분 프라이버시(Differential privacy)

- 특정 개인에 대한 사전지식이 있는 상태에서 해당정보가 포함된 데이터베이스와 포함되지 않은 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법

- 1개 항목이 차이나는 두 데이터베이스간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델

※ 질의응답 값을 확률적으로 일정 크기 이하의 차이를 갖도록 함으로써 차이에 따른 차분 공격 방지