

발간등록번호

11-1790365-000029-01

# 가명정보 처리 가이드라인

2026. 3.



개인정보보호위원회

Personal Information Protection Commission



# 가명정보 처리 가이드라인

**[본권] 가명정보 제도 안내편**

2026. 3.



개인정보보호위원회  
Personal Information Protection Commission



발간 등록 번호

11-1790365-000029-01

# 가명정보 처리 가이드라인

**[본권] 가명정보 제도 안내편**

2026. 3.

# 가명정보 처리 가이드라인

## 안내 사항

### 발간 목적

본 안내서는 빅데이터, AI 등 다양한 융·복합 산업에서의 데이터 이용 수요가 급증하는 가운데, 데이터 활용의 핵심인 가명정보 활용을 위한 법적 근거가 마련됨에 따라 가명정보 활용에 필요한 가명정보 처리 목적, 처리 절차 및 방법, 안전조치에 관한 사항 등을 안내할 목적으로 마련되었습니다.

### 제·개정 이력

가명정보 관련 법제도 및 환경 변화를 반영하여 다음과 같이 개정하였습니다.

일자	주요 내용
'20 9. 발간	가명정보 처리 특례 활용 시 개인정보처리자가 참고할 수 있는 기준 안내
'21.10. 개정	「가명정보의 결합 및 반출 등에 관한 고시」 개정('21년) 사항 반영, 결합방식 개선사항 안내
'22. 4. 개정	가명정보 처리 절차 구체화, 결합유형 안내, 참고사례 및 FAQ 추가 수록
'24. 2. 개정	비정형데이터 가명처리 기준 마련, 「개인정보 보호법」 개정('23년) 사항 반영
'26. 3. 개정	대상 독자별 가이드 분리(본권, 별권), 표준화된 위험도 판단 기준 마련, 위험도 기반 가명처리 절차·서류 차등화, 가명처리 관련 서식 통합·간소화 등

### 재검토 기한

안내서의 최신성을 유지하기 위해 발간일(2026년 3월)을 기준으로 매 3년이 되는 시점(매 3년째의 12.31.까지를 말함)마다 보완 및 개선 등의 조치를 취할 예정입니다.

### 저작권 표시

본 안내서 내용의 무단전제를 금하며, 가공·인용할 때는 출처를 밝혀 주시기 바랍니다.

\* 출처 : 개인정보보호위원회, 「가명정보 처리 가이드라인」 2026.03.

### 문의처

안내서 내용 관련 문의는 소관 법령별로 다음의 연락처로 주시기 바랍니다.  
- 개인정보보호위원회 데이터안전정책과(☎02-2100-3088, 3079, 3049)

### 법적 근거

「개인정보 보호법」 제3장 제3절 '가명정보 처리에 관한 특례'

※ 법령 최신 자료는 국가법령정보센터(www.law.go.kr), 개인정보 보호 안내서 최신 자료는 개인정보보호위원회 누리집\*, 개인정보 포털\*\*을 참고

\* 개인정보보호위원회 누리집(www.pipc.go.kr) : 정책·법령 > 법령정보 > 안내서

\*\* 개인정보 포털(www.privacy.go.kr) : 자료 > 자료보기 > 안내서



... 가이드라인,

## 이렇게 활용하세요

본 가이드라인은 「개인정보 보호법」 제3장 제3절「가명정보 처리에 관한 특례」 규정을 누구나 쉽게 이해하고 현장에서 안전하게 적용할 수 있도록 돕기 위해 작성되었습니다.

특히 이번 개정판은 독자의 전문성과 필요에 따라 [본권 : 제도 안내편]과 [별권 : 처리 실무편]으로 분리하여, 데이터 활용의 문턱은 낮추고 실무의 전문성은 강화하였습니다.

### 1. 읽기 전에 확인하세요 : 가이드라인 적용 대상

- ◆ (적용 대상) 「개인정보 보호법」 제3장 제3절(가명정보 처리에 관한 특례)에 근거한 가명정보 처리를 대상으로 합니다.
- ◆ (우선 순위) 공공분야, 보건의료, 교육 등 개인정보보호위원회가 관계 부처와 공동 발간한 분야별 가이드라인\*이 있는 경우 해당 분야의 가이드라인을 우선 적용하며, 본 가이드라인은 일반적인 원칙과 기준을 제시합니다.
  - 예) 공공분야 가명정보 제공 실무안내서, 보건의료 데이터 활용 가이드라인, 교육분야 가명·익명정보 처리 가이드라인

### 2. 나에게 맞는 가이드는? : 독자별 맞춤 활용법

구분	대상 독자	주요 내용
[본권] 가명정보 제도 안내편	<input checked="" type="checkbox"/> 일반 국민 <input checked="" type="checkbox"/> 데이터 초급 실무자 <input checked="" type="checkbox"/> 가명정보 제도를 처음 접하는 자	<ul style="list-style-type: none"><li>● 가명정보의 기본 개념 및 법적 근거</li><li>● 주요 가명처리 절차</li><li>● 비정형데이터 가명처리 원칙·기준</li><li>● 자주 묻는 질문</li></ul>
[별권] 가명정보 처리 실무편	<input checked="" type="checkbox"/> 데이터 실무자 <input checked="" type="checkbox"/> 기관 담당자 <input checked="" type="checkbox"/> 적정성 검토위원 <input checked="" type="checkbox"/> 가명정보 업무 수행자 등	<ul style="list-style-type: none"><li>● 가명정보 결합 절차 및 세부 내용</li><li>● 가명처리 기술 및 기법</li><li>● 안전성 확보조치 이행 방법</li><li>● 서식 및 운영 문서 작성 방법</li><li>● 상황별 예시 시나리오</li></ul>

### 3. 궁금한 것만 골라보세요 : 한눈에 보는 목차 안내

궁금한 사항	관련 목차
● 가명정보와 익명정보는 무엇이 다른가요?	[본권] 제1장. 2. 가명정보란 무엇인가요?
● 가명처리는 어떤 순서로 진행되나요?	[본권] 제2장 가명처리 절차
● 이미지 영상, 텍스트 등은 어떻게 가명처리 하나요?	[본권] 제3장 비정형데이터 가명처리 기준
● 서로 다른 기관의 가명정보를 결합해서 활용하려면?	[별권] 제2장 가명정보의 결합
● 가명처리 후 안전성 확보는 어떻게 하나요?	[별권] 제3장 안전성 확보조치 이행 가이드
● 활용 가능한 가명처리 기술·기법은 어떤 것이 있나요?	[별권] 제4장 가명처리 기술 및 기법
● 실제 업무에 바로 쓸 수 있는 양식이 있나요?	[별권] 제5장 서식 모음
● 가명정보가 어떻게 쓰이는지 활용 사례가 궁금해요	[별권] 제8장 가명정보 활용 참고 시나리오
● 우리 기관과 유사한 사례가 있다면 참고하고 싶어요	

### 4. 이렇게 읽으면 가장 쉽습니다 : 내 상황에 딱 맞는 정주행 코스

Case	본권	별권
1 가명정보가 처음이라 감을 잡고 싶어요	제1장 개념파악 → 제2장 절차 훑기 → 제4장 Q&A로 궁금증 해결	
2 당장 실무에 적용해 데이터를 처리해야 해요	제2장 절차 숙지	제4장 기술·기법 선택 → 제5-6장 서식 및 운영문서 작성
3 다른 기관과 데이터 결합해 분석하고 싶어요	제1, 2장 개념·절차 이해	제2장 결합 규정 확인
4 영상, 음성 등 비정형데이터를 다루고 있어요	제3장 비정형데이터 기준 확인	제4장 비정형 가명처리 기법 선택 → 제8장 비정형 사례 참고
5 가명정보 처리 업무를 총괄하는 관리자(CPO 등)예요	제1장 특례 규정 및 법적 책임 확인	제3장 안전성 확보조치 이행 가이드 → 제6장 내부관리계획 등 운영문서 검토

### 5. 혼자 고민하지 마세요 : 가명정보 관련 지원 인프라·컨설팅 안내

- ◆ 가명정보 지원 플랫폼([dataprivacy.go.kr](http://dataprivacy.go.kr))
  - 가명정보 컨설팅, 기술지원, 교육·훈련, 전문가 풀, 수요·공급 연계, 결합신청 등 가명정보와 관련한 모든 사항을 종합적으로 지원합니다.
- ◆ 가명정보 활용지원센터
  - 전국 주요 지역별 거점에 설치된 센터에서 가명처리 인프라·솔루션, 컨설팅, 교육 및 실습 등을 지원받으실 수 있습니다. (가명정보 지원 플랫폼에서 방문 예약/신청)

# 목차 Contents



가이드라인, 이렇게 활용하세요

## 본권 가명정보 제도 안내편

제1장 가명정보 제도의 이해 3

제2장 가명처리 절차 23

제3장 비정형데이터 가명처리 기준 57

제4장 자주 묻는 질문 (Q&A) 77

# 목차 Contents



## 별권 가명정보 처리 실무편

제1장 용어 정리 3

제2장 가명정보의 결합 9

■  
■  
■  
■

제3장 안전성 확보조치 이행 가이드 43

제4장 가명처리 기술 및 기법 49

제5장 서식 모음 83

# 목차 Contents

---

**제6장** 가명정보 처리 관련 주요 운영문서 작성 가이드 97

**제7장** 가명정보 활용 위험도 판단 예시 113

■

■

**제8장** 가명정보 활용 참고 시나리오 155

■



# 1

## 가명정보 제도의 이해

가명정보 처리 가이드라인

가명정보 제도 안내편







# 제1장 가명정보 제도의 이해

## 1

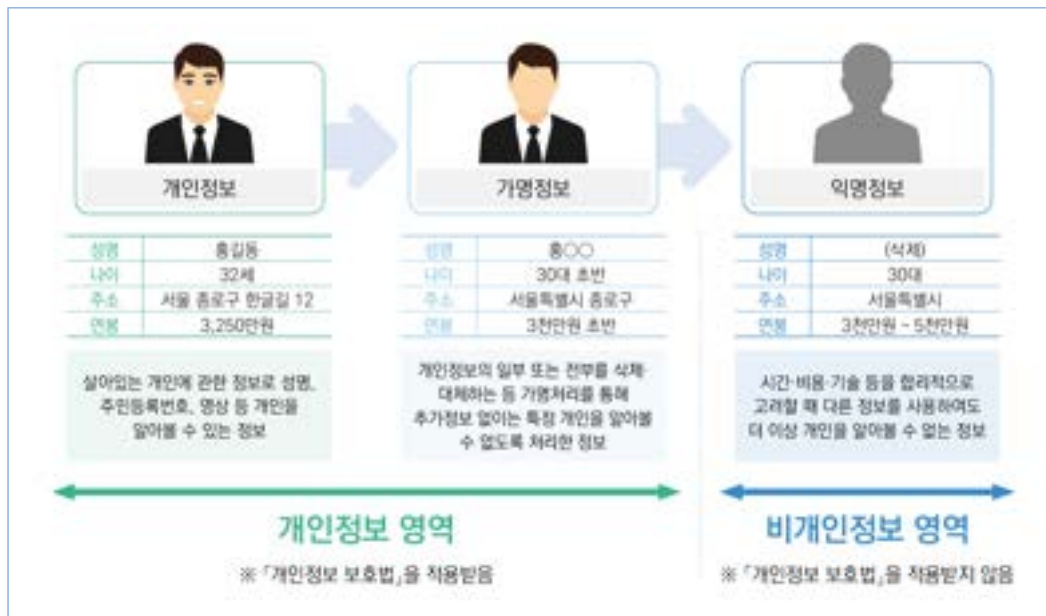
- ◆ 최근 빅데이터, 인공지능(AI), 클라우드 등 **신기술의 확산**으로 **데이터 활용**에 대한 **사회적·산업적 수요가 빠르게 증가**하고 있음
  - 이에 **데이터 활용에 대한 시대적 요구**를 반영한 **데이터 3법\***이 시행('20.8.5.)되어 **정보주체 동의 없이도 개인정보를 가명처리하여 2차적 목적으로 안전하게 활용**할 수 있는 기반이 새롭게 마련됨
 

\* 개인정보 보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 '정보통신망법'이라 함), 신용정보의 이용 및 보호에 관한 법률(이하 '신용정보법'이라 함)
- ◆ 「개인정보 보호법」(이하 '보호법'이라 함)에 도입된 **'가명정보 처리에 관한 특례'**(보호법 제3장 제3절)는 개인정보 보호와 활용을 조화롭게 달성할 수 있는 새로운 제도적 장치로서,
  - **특정 개인을 알아볼 수 없도록 가명처리**하고, **필요한 안전조치를 이행**하도록 하여 **재식별 위험을 적절히 관리**하는 것을 전제로 안전하게 데이터를 활용할 수 있는 환경을 조성하였음
- ◆ 이 가이드라인은 **가명정보 제도의 기본 개념과 취지**에 대해 누구나 **쉽게 이해할 수 있도록** 안내하고, **가명정보를 안전하게 처리·활용**하기 위한 **기본적인 방향**을 제시하기 위해 마련됨



## 2

◆ 보호법에서는 개인과 관련된 데이터를 다음과 같이 구분하고 있음



◆ “개인정보”는 살아 있는 개인에 관한 정보로, 해당 정보만으로 또는 다른 정보와 쉽게 결합하여 특정 개인을 알아볼 수 있는 정보를 의미함

- 예를 들어, 성명, 주민등록번호, 상세 주소, 연락처, 얼굴이 드러난 영상 등 개인을 직접 또는 간접적으로 식별할 수 있는 정보는 개인정보에 해당함
- 해당 개인정보는 원칙적으로 정보주체의 동의를 받거나 법률에서 허용한 수집 목적 범위 내에서만 처리·활용할 수 있음

◆ “가명정보”는 위 개인정보의 일부 또는 전부를 삭제하거나 대체하는 등의 방법으로 추가정보 없이는 특정 개인을 알아볼 수 없도록 처리한 정보를 의미함

- 예를 들어, 성명 중 성을 제외한 이름을 삭제하고 상세 주소를 시·군 단위로 축소하며, 나이를 ‘30대 초반’처럼 범주화한 경우, 추가정보가 없으면 특정 개인이 누구인지 바로 알기 어렵게 됨

- 이처럼 가명정보는 특정 개인을 식별하기 어렵도록 보호 수준을 높이면서도, 통계·학습·연구 등에 필요한 가치는 보존시킨 정보임
- 다만, 가명정보는 추가정보와 결합되어 특정 개인이 식별되는 등 재식별 위험을 관리할 필요가 있다는 점에서, 가명정보도 여전히 개인정보의 범주에 포함되며, 보호법의 적용을 받음

 관련 규정

개인정보 보호법

.

- ◆ “익명정보”는 시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보를 의미함
- 익명정보는 더 이상 개인정보에 해당되지 않기 때문에, 보호법의 적용을 받지 않음

 관련 규정

개인정보 보호법

.

개인정보							
이름	생년월일	핸드폰	주소	직업	가족	예금평균잔액	대출액
홍길남	67.01.03	010-9999-3333	창원시 마산회원구 상호로 63	국회의원	배우자, 아들 1, 딸1	4,567,900원	50,984,234원

잘못된 가명정보							
이름	생년월일	핸드폰	주소	직업	가족	예금평균잔액	대출액
		wd4e85s2r4qe	창원시 마산회원구	국회의원	배우자, 아들 1, 딸1	4,567,900원	50,984,234원
삭제	삭제	임호희	병주희	개인식별가능			

가명정보							
이름	생년월일	핸드폰	주소	직업	가족	예금평균잔액	대출액
		wd4e85s2r4qe	창원시 마산회원구		배우자, 아들 1, 딸1	4,567,900원	50,984,234원
삭제	삭제	임호희	병주희	삭제			

익명정보							
이름	생년월일	핸드폰	주소	직업	가족	예금평균잔액	대출액
		wd4e85s2r4qe	창원시 마산회원구		가족3	400-500만원	4천-5천만원
삭제	삭제	임호희	병주희	삭제	일반화	병주희	병주희

- ◆ 개인정보를 안전하게 가명처리하면, 보호법에서 정하는 목적\*에 한해 정보주체 동의 없이도 활용할 수 있음 (가명정보 처리 특례 제도)

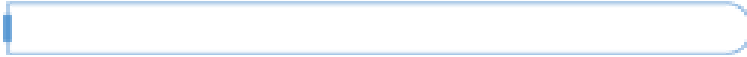
\* 통계작성, 과학적 연구, 공익적 기록보존 등

- ◆ 다음과 같은 상황에서 가명정보 제도가 특히 유용하게 활용될 수 있음

- 최초 수집 목적범위를 넘어 추가 데이터 활용 수요가 발생하였는데, 정보주체에게 일일이 동의를 받는 것이 사실상 어려운 경우
  - 데이터 규모가 매우 커 정보주체 수가 많은 경우
  - 과거에 수집된 데이터로, 현재 정보주체 연락이 어려운 경우
- 기존 동의받은 수집 목적이나 개별 법령상 허용된 범위만으로는 개인정보의 활용이 어려운 경우
  - 개별 법률에 규정된 공공기관의 소관 업무 목적 외로 통계작성, 정책연구 등 추가적인 활용 필요성이 발생한 경우
  - 각기 다른 목적으로 수집·보유 중인 여러 기관·기업의 정보를 결합하여 개별 정보로는 얻기 힘든 새로운 시사점을 얻고자 하는 경우
- 분석대상이 특정 개인이 아니라 집단의 경향이나 통계적 특성인 경우

여기서 질문 가명정보 특례 제도를 개인정보 '수집' 근거로 쓸 수 있나요?

- 
-



- ◆ 당초 수집 목적과 관련성이 없어도 과학적 연구 등 목적으로 활용할 수 있도록 한 **가명정보 특례 제도**에서의 가명처리 외에도 **보호법에서는 다양한 목적과 맥락에서 가명처리를 사용하고 있음**

⇒ 가명정보 특례 제도에서의 “가명처리”와 다음 맥락에서 수행되는 가명처리는 서로 구분하여 이해하여야 함

① 보호원칙 실현을 위한 가명처리(보호법 제3조 제4항, 제6항, 제7항)

- 개인정보 처리 시 정보주체 권리 침해 최소화, 보다 안전하게 개인정보를 관리하기 위한 수단 등 보호법에서 정하는 **개인정보 보호 원칙을 실질적으로 이행하기 위한 방법** 중 하나로 가명처리를 수행할 수 있음



관련 규정

개인정보 보호법 제3조(개인정보 보호 원칙)

② 안전성 확보 의무 이행을 위한 가명처리(보호법 제29조)

- 개인정보의 **안전성 확보**를 위한 **기술적·관리적 보호조치의 일환**으로 암호화 등의 가명처리를 수행할 수 있음



관련 규정

개인정보 보호법 제29조(안전조치의무)

· · · · ·

.



3

※ 보호법 제3장 제3절

◆ 가명정보의 활용

- 개인정보처리자는 원칙적으로 보호법 제18조제1항에 따라, **개인정보의 수집 범위를 초과하여 이용할 수 없으나**,
  - 그럼에도 불구하고 특례 규정에 따라 **통계작성, 과학적 연구, 공익적 기록보존** 목적으로는 **정보주체 동의 없이 개인정보를 가명처리하여 활용할 수 있음** (특례 규정 우선 적용)
- 이는 **적법하게 수집된 개인정보**를 제한된 목적 하에 안전하게 **2차적으로 활용할 수 있도록 허용한 것으로**,
  - 해당 목적으로 보유한 개인정보를 **직접 활용**하거나 **다른 부서에 제공**하거나, 내부에서 가명정보를 결합하여 활용하거나, **특정한 외부의 제3자에게 제공**하여 활용하게 하는 경우 등이 모두 포함됨
  - 이때, **가명정보**는 가명처리를 수행한 **당시의 처리 목적과 처리 환경(활용 형태, 처리 장소, 처리 방법)에 따라 이용하는 것이 원칙임**



관련 규정

개인정보 보호법

여기서 잠깐! “가명처리”와 “가명정보의 처리”는 어떻게 다른가요?

- 
- 
- 
- 

- 가명정보의 처리 기간은 처리 목적 등을 고려하여 필요한 기간으로 개인정보처리자가 별도로 정할 수 있음
  - 이는 당초 개인정보의 처리 기간과는 무관하게 가명정보 처리 특례 활용 목적에 필요한 기간으로 새롭게 정하는 것임

관련 규정

개인정보 보호법

여기서 잠깐! 목적 설정 없이 미리 가명처리하여 보유하고 있어도 되나요?

- 
-

### ◆ 가명정보의 제3자 제공

- 개인정보처리자는 원칙적으로 보호법 제18조제1항에 따라 **개인정보의 수집 범위를 초과하여 제3자에게 제공할 수 없으나**,
  - 그럼에도 불구하고 특례 규정에 따라 **통계작성, 과학적 연구, 공익적 기록보존** 목적으로는 **정보주체 동의 없이 개인정보를 가명처리하여 제3자에게 제공할 수 있음** (특례 규정 우선 적용)
- 가명정보를 제3자에 제공할 때에는, **가명정보를 제공받는 제3자의 상황을 종합적으로 고려하여 안전하게 가명처리 후 제공**하여야 함
  - 이 경우 **제공 목적, 제공 이후의 처리 방법·환경, 제3자가 보유·접근 가능한 정보 등을 종합적으로 고려하여 특정 개인을 알아보기 어렵도록 가명처리함으로써 재식별 위험을 합리적 수준으로 낮추기 위해 노력**하여야 함



#### 관련 규정

개인정보 보호법 제28조의2(가명정보의 처리 등)



**여기서 잠깐!** 가명정보를 제공 후 발생한 위법행위도 제공자 책임인가요?

■

■

**여기서 질문** 가명정보를 불특정 다수(제3자)에게 제공(공개)해도 되나요?

■

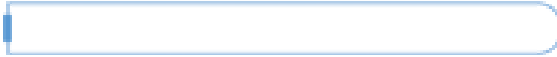
■

**여기서 질문** 가명정보를 다시 사용하거나 다른 기관에 재제공해도 되나요?

■ 「

■

■



### ◆ 가명정보 결합의 유용성

- 가명정보 제도는 여러 기관이 보유한 가명정보를 서로 결합하여 활용할 수 있도록 허용하고 있으며,
- 이종 산업간 결합 등으로 단일 정보만으로는 알아내기 어려운 새로운 인사이트를 창출하고 데이터의 활용 가치를 높일 수 있음

### ◆ 가명정보 결합 수행 주체

- 단일 기관·기업 내 부서 간 결합 등 동일 개인정보처리자가 보유한 가명정보를 결합하는 경우에는 자체적인 결합 수행이 가능함
- 한편, 서로 다른 기관(개인정보처리자)이 보유한 가명정보를 결합하려는 경우에는 반드시 결합 전문기관\*을 통해야 함

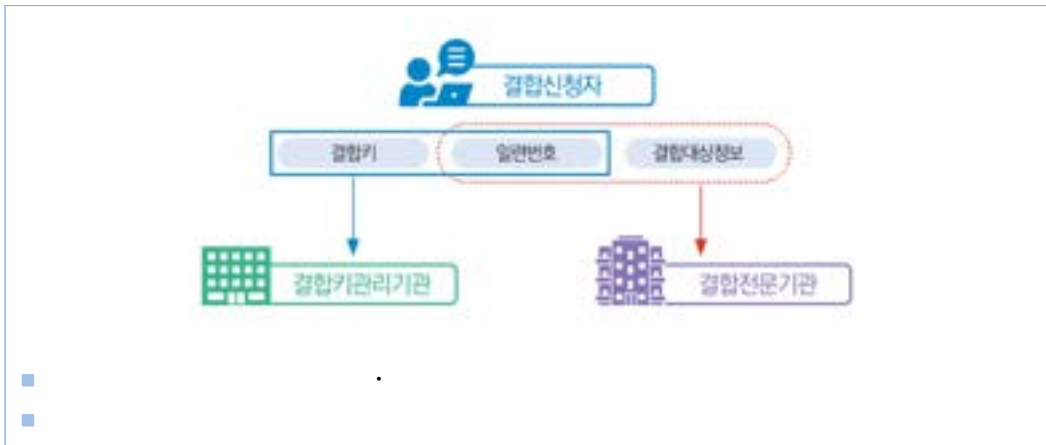
\* 결합전문기관은 개인정보보호위원회 또는 관계 중앙행정기관이 지정한 기관으로, '가명정보 지원 플랫폼 (dataprivacy.go.kr)' 에서 기관현황 확인 및 결합신청 가능

### ◆ 가명정보 결합 방법

- 가명정보 결합은 일반적으로 동일인을 서로 연결하기 위한 결합키를 안전한 방식으로 생성하고 해당 결합키를 기반으로 가명정보를 결합한 뒤, 결합키를 삭제하는 방식으로 수행됨



- 여러 기관이 보유한 정보를 결합하는 경우 **개인식별 위험성이 비교적 높으므로, '신뢰할 수 있는 제3자'인 결합전문기관을 통해 안전하게 결합을 수행하도록** 하고 있음
- 특히, 보호법상 가명정보 결합은 **단일 기관이 결합과 관련된 정보를 모두 처리하지 않게 역할을 분리하여, 결합 과정에서 누구도 특정 개인을 식별할 수 없도록 안전하게 설계되어** 있음

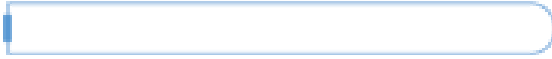


- 
- 

**관련 규정**

개인정보 보호법 제28조의3(가명정보의 결합 제한)

- 가명정보 결합의 세부 방법·절차는 별편 제2장(가명정보 결합) 참고



◆ 안전성 확보조치 의무 준수

- 가명정보도 개인정보 개념에 포함되므로, 개인정보에 대한 안전성 확보조치 의무(보호법 제29조)를 이행하여야 함

◆ 가명정보·추가정보의 별도 분리보관 및 접근권한 분리

- 가명정보는 추가정보\*와 비교·대조할 경우 재식별이 가능해지므로,

\* (추가정보) 가명정보를 원래의 상태로 복원하기 위한 정보로, 가명처리 과정에서 개인정보의 전부·일부를 대체 하는데 이용된 수단·방식(알고리즘 등), 삭제·대체된 개인정보 부분을 복원할 수 있는 정보(매핑테이블 등), 가명 처리에 사용한 개인정보 등이 포함될 수 있음

- 추가정보가 재식별에 악용되지 않도록 가명정보와 분리하여 별도로 저장·관리하고 접근권한을 분리하는 등 안전하게 보관·관리하여야 함

◆ 가명정보 처리 관련 기록 작성·보관

- 개인정보처리자는 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보 이용내역, 가명정보의 처리 기간, 제3자 제공 시 제공받는 자를 작성·보관하여야 함

관련 규정

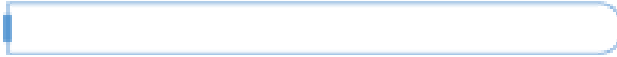
개인정보 보호법

· . . . .

·

· . . . .

·



◆ 가명정보 고의적인 재식별 시도 금지

- 누구든지 고의적으로 가명정보를 재식별하는 행위를 시도해서는 안 됨
- 특정 개인을 알아보기 위하여 가명정보를 처리하는 것은 해당 정보의 정보주체가 누구인지 알아볼 수 있도록 하는 행위 일체를 의미함
- 즉, 추가정보를 사용하여 원래 정보로 복원하는 행위는 물론이고, 보유하고 있는 다른 정보나 공개된 정보와의 결합 또는 대조·비교 등을 통해 특정 개인을 알아보기 위한 시도 등은 모두 금지됨

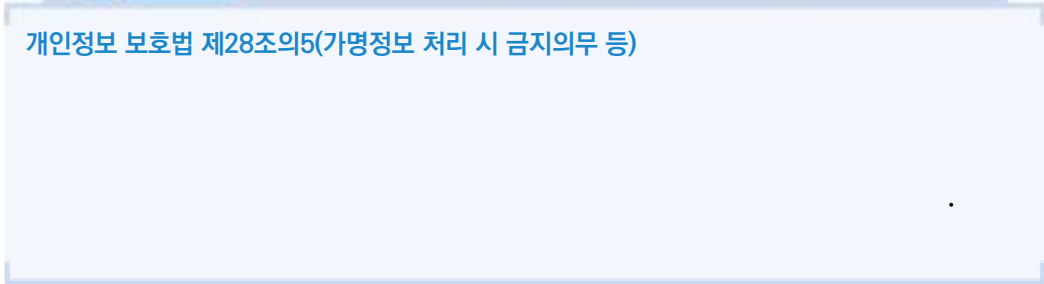
◆ 가명정보의 처리중지 및 회수·파기 의무

- 가명정보 처리 과정에서 우연히 특정 개인이 식별되는 경우에는 처리중지, 회수·파기 등과 같이 위험을 제거하기 위한 적절한 조치를 즉시 수행하여야 함



관련 규정

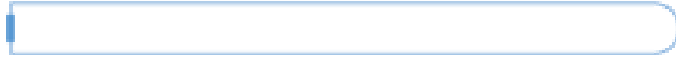
개인정보 보호법 제28조의5(가명정보 처리 시 금지의무 등)



여기서 질문! 가명정보 활용 중 의도치 않게 재식별이 된 경우에도 처벌을 받나요?

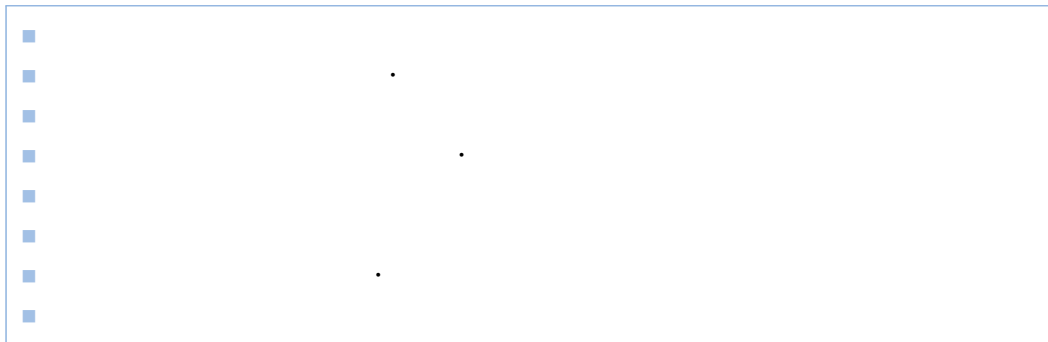
- 
- 





- ◆ 통계작성, 과학적 연구, 공익적 기록보존 등 목적으로 가명정보 제도를 통해 처리되는 가명정보에 대해서는 **아래 규정을 적용하지 않음\***

\* 이는 보호법 제28조의5 제1항에 따라 가명정보의 재식별 시도를 금지하고 있는데, 해당 규정들은 특정 개인을 식별하여야만 이행할 수 있는 사항이기 때문임



### 관련 규정

#### 개인정보 보호법 제28조의7(적용범위)

- ◆ 위 규정 외의 보호법 조항은 가명정보에도 동일하게 적용됨
  - 예컨대, 개인정보처리자는 가명정보를 처리하는 경우,
    - 개인정보 보호 원칙(제3조)에 따라서 처리 목적의 범위에 필요한 범위에서 가명정보를 처리하여야 하며,
    - 가명정보 처리 업무를 위탁하는 경우에는 보호법 제26조의 요건을 모두 준수하여야 함
    - 가명정보 처리 목적 달성 시 보호법 제21조에 따라 해당 가명정보는 파기하여야 함

**여기서 질문** 가명정보도 국외이전이 가능한가요?

- 
- 

**여기서 질문** 개인정보에 대한 가명처리 정지 요구가 가능한가요?

- 
- 

▶

**관련 규정**

개인정보 보호법

# 2

## 가명처리 절차

가명정보 처리 가이드라인

가명정보 제도 안내편





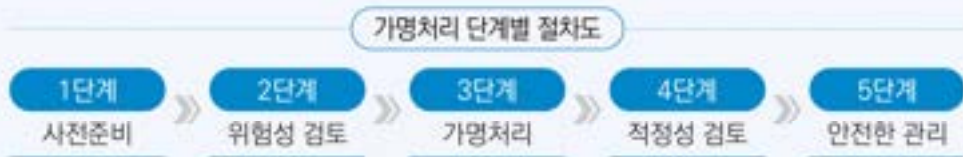


## 제2장 가명처리 절차

- ◆ 가명정보 특례 제도에 따라 수행하는 가명처리는 단순히 개인정보 일부를 가리는 기술적 처리만을 의미하는 것이 아니라,

- [가명처리 절차] 까지

- ◆ 개인정보의 가명처리 절차는 다음과 같은 [가명처리 단계별 절차도] 를 거쳐 이루어짐



- ① **사전준비** : 왜 가명정보를 활용하려고 하는지를 정리하는 단계
  - 처리 목적을 설정·검토하고, 해당 목적에 필요한 가명처리 대상을 선정
- ② **위험성 검토** : 가명처리 전, 가명정보 처리의 위험성을 미리 검토하는 단계
  - 가명정보 활용 시의 위험성을 먼저 검토하여 어떤 방식과 수준으로 가명처리를 수행할 것인지를 판단
- ③ **가명처리** : 위험성 검토 결과를 바탕으로 실제 가명처리를 수행하는 단계
  - 처리 목적을 달성할 수 있으면서도, 재식별 위험을 통제 가능한 수준으로 낮추는 방향으로 가명처리 수행
- ④ **적정성 검토** : 가명처리가 적절하게 수행되었는지 검토하는 단계
  - 처리 목적이 적절한지, 가명처리 이용항목이 잘 선정되었는지, 위험도 판단 및 가명처리가 적절하게 수행되었는지, 해당 가명정보로 목적 달성이 가능한지 등을 종합적으로 검토
- ⑤ **안전한 관리** : 가명정보를 활용하면서 재식별 위험을 지속 관리하는 단계

1

- 가명정보 처리 목적을 설정·검토하고 해당 목적에 필요한 가명처리 대상을 선정하는 단계

## ① 가명정보 처리 목적 설정

- ◆ 가명정보 제도는 **보호법에서 정한 목적에 한해 적용됨**
  - 따라서 가명정보 제도를 활용하고자 하는 개인정보처리자는 **통계작성, 과학적 연구, 공익적 기록보존 목적 중** 어떤 목적으로 가명정보를 활용할지를 먼저 **명확하게 설정**하여야 함
  - 이 때, 처리 목적은 “연구”, “통계”, “AI 학습”, “공익 활용” 등과 같이 추상적으로 설정해서는 안 되며, **무엇을 연구·분석·검증하려는지까지 구체적으로 정하는 것이 중요**

- ◆ 처리 목적을 **구체적으로 설정**해야지만 **목적에 필요한 정보 항목을 명확히 선별**할 수 있고,
  - 위험성 검토, 가명처리 방법·수준 결정, 적정성 검토 등 **후속 절차를 해당 목적에 맞추어 체계적으로 수행**할 수 있음

※ 이는 처리 목적을 명확하게 설정하고, 처리 목적에 필요한 범위에서 적당하게 개인정보를 처리하도록 한 개인정보 보호 원칙(보호법 제3조 제1항, 제2항)의 취지에 따른 것임

“통계작성” 목적이란?

- “통계”란 특정 집단이나 대상 등에 관한 수량적인 정보를 의미하며, 해당 통계를 작성하기 위해 가명정보를 이용, 분석, 결합, 제공하는 등의 행위가 통계작성 목적의 가명정보 처리에 해당함
- 시장조사를 위한 통계, 마케팅 전략 수립을 위한 통계 등 상업적 성격을 가진 통계도 통계작성 목적에 포함됨



여기서 질문! 가명정보로 통계작성 후 1:1 맞춤형 타겟 마케팅도 가능한가요?

- 
- 



### “과학적 연구” 목적이란?

- “과학적 연구”란 과학적 방법을 적용하는 연구\* 를 의미하며, 다양한 분야에서 이루어지는 기초·응용 연구를 모두 포함함

\* 과학적 방법을 적용하는 연구란 체계적이고 객관적인 방법으로 검증 가능한 질문에 대해 연구하는 것을 말함



### 관련 규정

#### 개인정보 보호법 제2조(정의)

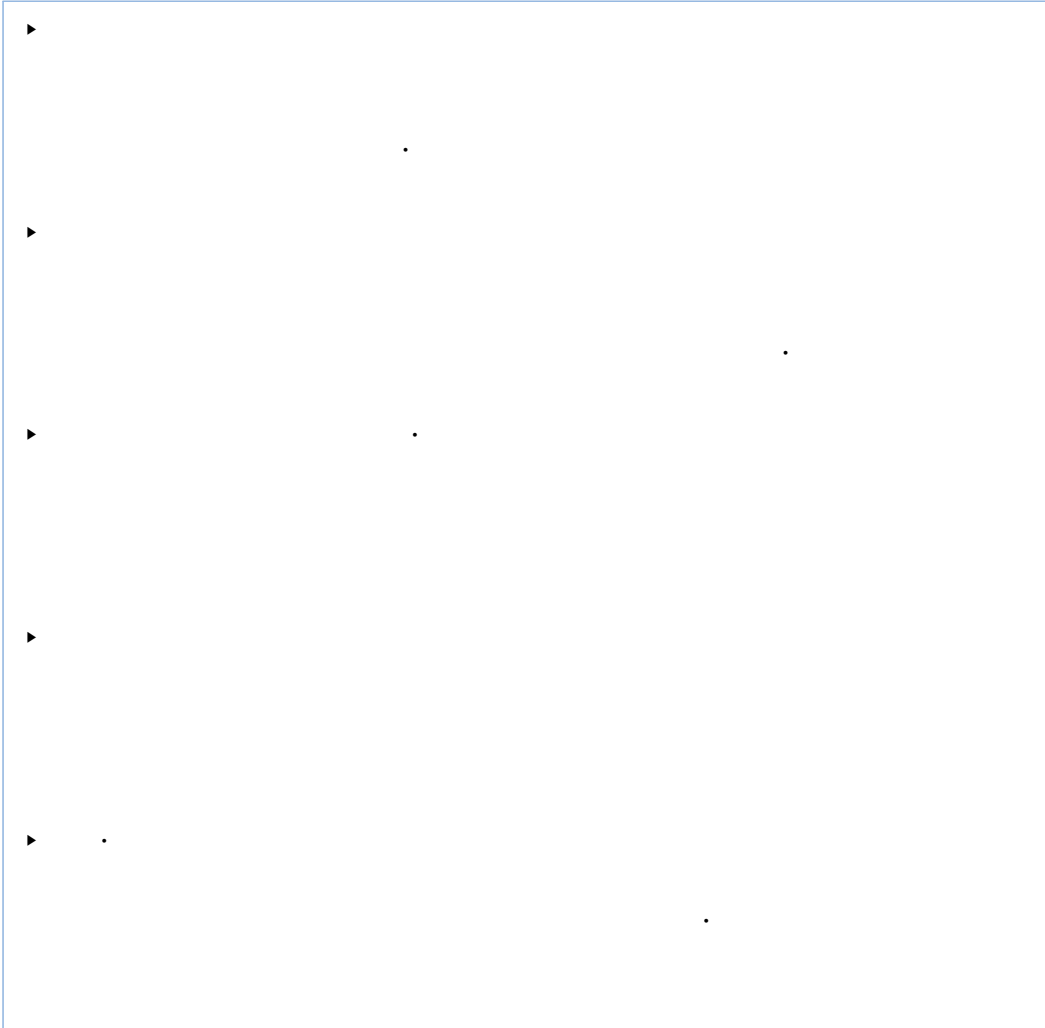
- 따라서, 새로운 기술·제품·서비스 개발 및 실증을 위한 연구, 데이터 분석을 기반으로 한 산업적 연구도 과학적 방법을 적용하는 경우 과학적 연구 목적에 포함됨
- 또한, 과학적 연구는 공적 자금으로 수행되는 연구에 한정되지 않으며, 민간의 투자나 자체 예산으로 수행되는 연구도 포함됨

### 여기서 질문! AI 기술·서비스 개발 및 고도화 목적도 “과학적 연구”에 해당하나요?

■

■

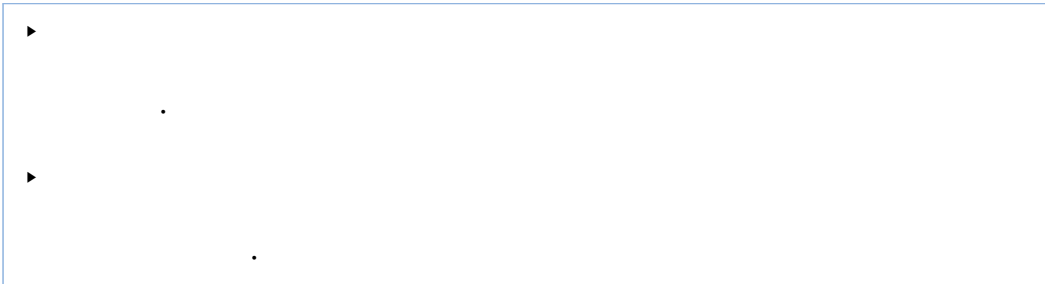
■



1. 가명정보의 범위

- ▶ 가명정보란 개인정보처리자가 개인정보의 식별성을 제거하여, 해당 개인정보가 특정 개인을 식별할 수 없도록 처리한 정보를 말한다.
- ▶ 가명정보는 개인정보처리자가 개인정보의 식별성을 제거하여, 해당 개인정보가 특정 개인을 식별할 수 없도록 처리한 정보를 말한다.
- ▶ 가명정보는 개인정보처리자가 개인정보의 식별성을 제거하여, 해당 개인정보가 특정 개인을 식별할 수 없도록 처리한 정보를 말한다.
- ▶ 가명정보는 개인정보처리자가 개인정보의 식별성을 제거하여, 해당 개인정보가 특정 개인을 식별할 수 없도록 처리한 정보를 말한다.
- ▶ 가명정보는 개인정보처리자가 개인정보의 식별성을 제거하여, 해당 개인정보가 특정 개인을 식별할 수 없도록 처리한 정보를 말한다.

- “공익적 기록보존”이란 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 정보를 기록하여 보존하는 것을 의미함
- 공익적 기록보존은 공공기관이 처리하는 경우에만 공익적 목적이 인정되는 것은 아니며, 기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우에도 공익적 기록보존 목적이 인정됨



❖ 목적 설정 시 ‘확장 가능한 목적’ 고려

- 가명정보 처리 목적을 사전에 단일 목적으로만 특정하기 곤란한 경우가 많고, 향후 유사 목적 활용 시 목적 변경에 따라 불필요한 가명처리·적정성 검토 수행이 반복적으로 요구되는 어려움 존재

예시 범용 목적 AI 개발 확산으로 당초 목적 외 용도로 응용·고도화, 미세조정 빈번



- 이에 최초 가명정보 처리 목적 설정 시, 단일 목적 외에 향후 ‘유사 목적으로 확장 가능한 목적 범위’ 까지 사전에 설정할 수 있음

예시



※ 단, 적정성 검토 시 확장 가능목적까지 고려하여 가명처리 방법·수준 검토 필요

- (제한조건) ①최초 목적과 직접적으로 연관된 추가 분석 목적 또는 ②동일연구 과정에서 파생되는 유사 목적의 응용연구로 한정
  - 또한, 확장 목적은 최초 목적과 동일한 데이터셋을 활용하여야 함
- 이 경우 가명처리 대상 선정, 가명처리, 적정성 평가 등 일련의 과정 또한 단일 목적 외에 확장 가능한 목적 범위까지 고려하여 수행

### ◆ 가명정보 처리 기간 설정 방법



#### 관련 규정

개인정보 보호법 제28조의4(가명정보에 대한 안전조치의무 등)

- 개인정보처리자는 가명정보 처리 목적 달성에 필요한 범위에서 가명정보의 처리 기간을 정하여 활용할 수 있음
- 이때, 가명정보의 처리 기간은 해당 가명처리의 대상이 된 기존 개인정보의 보유·이용 기간과는 별도로 설정 가능함
- ※ 다만, 가명정보는 처리 목적이 달성되면 지체 없이 파기하는 것이 원칙임

#### 예시 ①

▶

→

#### 예시 ②

장기간 분석이 필요한 연구를 수행하려는 경우 처리 기간 정하기

▶

→

제1장 가명정보 처리의 개요  
제2장 가명처리 절차  
제3장 가명처리 대상의 범위  
제4장 가명처리 방법 및 절차

예시 ③

- ▶
- 
- 

- 기존 활용 중인 가명정보를 다른 특례 목적으로 활용하려는 경우 기존에 처리 중인 가명정보가 파기되기 전이라면, 새로운 목적에 맞는 처리 기간을 별도로 설정하여 활용할 수 있음

※ 이 경우, 기존 목적에 따른 처리 기간과 새로운 목적에 따른 처리 기간은 각각 별도로 관리하여야 하며, 각 목적이 달성되면 해당 목적에 해당하는 가명정보는 파기하여야 함

예시 ④

- ▶
- 
-

## ② 가명처리 대상 선정

- ◆ 세부적인 처리 목적이 구체화되면, 이를 기반으로 해당 목적을 달성하는 데 필요한 정보의 종류와 범위를 정해야 함
  - 어떤 데이터셋을 쓸지, 어떤 항목(변수)이 필요한지, 어느 기간·범위의 데이터가 필요한지 등을 정리하여 가명처리 대상으로 선정



▶

▶

▶

**여기서 질문** '민감정보'나 '고유식별정보'도 가명처리해서 쓸 수 있나요?

■

■



※ 상세내용은 별권 제6장 참조

- **보호법 제29조(안전조치의무) 및 시행령 제30조에 따라 개인정보의 안전한 처리를 위해 내부관리 계획 수립 필요**
- 다만, 가명정보 처리를 위해 반드시 별도의 내부관리계획을 새로 작성·마련해야하는 것은 아니며,
  - 기 수립된 내부관리계획에 **시행령 제30조제1항제1호에서 요구하는 사항이 충실히 반영되어 있고, 가명정보 처리에 필요한 안전조치 사항을 충분히 규율할 수 있다면 기존 계획으로 갈음할 수 있음**
  - 기존 내부관리계획에 **가명정보 처리 특성에 맞는 사항이 충분히 반영되어 있지 않다면, 필요한 부분만 보완·개정**하거나, 가명정보 관련 내부관리계획만 **별도로 추가 마련**하는 것도 가능
  - 중요한 것은 형식적인 내부관리계획을 만드는 것 자체가 아니라, **실제 가명정보를 누가 어떻게 처리하고, 안전하게 관리할 수 있는지에 대한 내부 절차와 체계를 갖추는 것임**

**여기서 질문** 가명정보 제3자 제공 시 이용자의 내부관리계획 작성·제출 필요여부

- 
- 

※ 상세내용은 별권 제6장 참조

- 개인정보처리자는 개인정보의 **가명처리 업무** 또는 **가명정보를 활용한 연구, 분석, 서비스 개발·고도화 등 가명정보 처리 업무**를 외부 기관·기업 등에 위탁하여 처리할 수 있음
- 다만, 가명정보 처리 업무를 외부에 위탁하는 경우 **보호법 제26조에 따른 사항을 준수**하여야 함
  - 가명정보 처리에 대한 위탁 관련 계약문서를 작성하고, 처리방침에 가명정보 처리업무를 위탁받아 처리하는 자 등을 공개하여야 하며, 수탁자 관리·감독 등 위탁 시 준수해야 할 사항들을 확인해야 함

**여기서 질문!** 수탁자가 자체 목적을 위해 가명정보 제도를 활용하여도 되나요?

■

■



**관련 규정**

개인정보 보호법 제26조(업무위탁에 따른 개인정보의 처리 제한)

...

**여기서 질문!** 의료기관도 병원 외 기관·기업에 가명처리를 위탁해도 되나요?

■

「 」

「 」

「 」

「 」

## 2

가명처리 방법·수준을 결정하기 위해 가명처리 처리의 위험성을 검토하는 단계



### 1 기본 위험도 판단 (저·중·고)

활용 주체(내부 활용/제3자 제공)와 처리 환경(통제 가능 여부)을 기준으로 기본 위험도를 우선 판단

●	_____
▶	_____
▶	_____
●	_____
▶	_____
▶	_____
●	_____
▶	_____

## ② 사례별 위험도 조정

◆ 개별 사례에 따라 실질적 위험이 다를 수 있으므로 기본 위험도 판단 이후, 다음 요소를 고려해 최종 위험도를 상향·하향 조정할 수 있음

- 데이터 자체의 위험도

▶

- 가명정보 처리 맥락·상황

▶ 처리 과정, 기술적 구현 방식에 따라 활용된 가명정보가 유·노출, 재식별될 위험성 고려

- 기관 자체 지침

▶



### 기본 위험도의 사례별 조정(상향·하향 조정) 예시

◆ 기본 위험도(중위험·고위험)를 저위험으로 하향 조정 가능한 경우

- 이전에 처리했던 사례와 반복·유사 활용인 경우

▶

- 정부가 지정한 안전한 기관\*으로 제공하는 경우

\* 결합전문기관(데이터전문기관), 개인정보 이노베이션 존 등

▶

◆ 상향 조정이 권장

- 비정형데이터를 가명처리하여 활용하는 경우

▶

여기서 잠깐! '개인정보 이노베이션 존'이 무엇인가요?

■ [Blank area]

■ [Blank area]

■ [Blank area]

■ [Blank area]

제1장 개인정보 보호의 이해

제2장 가명처리 절차

제3장 개인정보처리방침과 개인정보처리방침

제4장 개인정보 보호 기술 (Anonymization)

### ③ 위험성 검토서 작성

◆ 가명처리에 앞서 수행한 위험도 검토 결과는 「**위험성 검토서**」 형태로 작성·보관하여야 함

※ 별권 제5장 서식③ 활용

- 「**위험성 검토서**」는 저·중·고 위험 여부를 단순 표시하는 것이 아니라, **왜 해당 위험도로 판단하였는지에 대한 근거를 남기는 문서**이며,
- 특히, 위험도 조정이 이루어진 경우 그 **사유와 관련 내용을 구체적으로 기재하여 사후 점검 시 판단 근거가 확인될 수 있도록** 하여야 함

◆ 「**위험성 검토서**」는 가명처리 방법과 수준을 결정에 참고가 되며, 적정성 검토 방법과 작성서류 범위를 차등 적용하기 위한 **판단 기준**이 됨

※ 개인정보처리자가 자체적으로 수립한 위험성 검토 기준이 있으면, 해당 「**위험성 검토서**」 대신 해당 기준으로 대체하여 검토·기록 보관 가능

#### 여기서 질문! 가명처리 시 위험도 산정이 왜 필요한가요?

■

■

■

■

...

...



여기서 잠깐! 정량적인 위험도 산정 기준(점수표)을 제시해줄 수는 없나요?

- 
-

### 3

#### ● 위험성 검토 결과를 바탕으로 실제 가명처리를 수행하는 단계

#### ◆ 가명처리의 핵심 목표

- 가명처리의 목표는 **재식별 위험을 0으로 만드는 것이 아님**
- 가명처리의 목표는 **처리 목적은 달성할 수 있도록 하면서도 계획하고 있는 처리환경 내에서의 재식별 위험을 통제 가능한 수준으로 낮추어 관리할 수 있도록 하는 것임**
- 따라서, “위험이 있는 항목은 모두 삭제한다”는 식의 접근은 **처리 목적 달성을 어렵게 만들고 가명정보 제도의 기본 취지와도 맞지 않음**

#### ① 항목별 가명처리 계획 수립

#### ◆ 가명처리 단계의 핵심 흐름

- 가명처리 단계는 **위험성 검토 결과를 바탕으로 항목별 가명처리 계획을 수립하고, 해당 계획에 따라 가명처리를 수행하는 방식**으로 진행함

#### ◆ 가명처리 계획을 왜 먼저 세워야 하나요?

- **동일한 정보 항목이라도** 데이터의 특성, 다른 정보와의 결합 가능성, 처리 환경, 활용 목적 등에 따라 **재식별 위험 수준은 달라질 수 있음**
- 따라서, 정보 항목 중 **어떤 항목을 가명처리할지, 항목별로 가명처리를 어떤 방식, 어느 수준까지 처리할지를 사전에 정리할 필요가 있음**

#### ◆ 개인정보처리자는 위험성 검토 결과를 기반으로 목적 달성에 필요한 가명처리 방법 및 수준을 정하여 항목별 가명처리 계획을 수립 **※ 상세 가명처리 기술 및 기법은 별권 제4장 참조**

- 식별 위험성이 있어 **가명처리가 필요한 정보 항목에 대해 위험성을 낮출 수 있는 가명처리 방법 및 수준을 선택**

※ **목적 달성에 필요한 항목 중, 해당 데이터 항목만으로 혹은 타 항목과 결합하여도 식별 가능성이 높지 않은 항목은 별도의 가명처리 없이 활용하여도 무방함**

예시

◆ 수립한 가명처리 계획을 기반으로 「가명처리 계획서」 작성 ※ 별권 제5장 서식④ 활용

No	개인정보 항목		가명처리 계획		
	항목명	예시	가명처리 방법	가명처리 수준	예시
1	고객ID	2428	<input checked="" type="checkbox"/> 대체	- 일련번호 대체	0032
2	나이	38세	<input checked="" type="checkbox"/> 범주화	- 10살 단위 범주화	30대
		18세	<input checked="" type="checkbox"/> 상하단 코딩	- 20세 미만 삭제	삭제
		86세		- 80세 이상은 80세 이상 경계치 입력	80세 이상
3	주소	서울시 종로구 한글길 12	<input checked="" type="checkbox"/> 부분 삭제	- 구단위 이하 삭제	서울시 종로구
4	성별	남	<input checked="" type="checkbox"/> 처리 없음 (분석에 반드시 필요한 항목)	-	남
5	식품류 구매액	984,450원	<input checked="" type="checkbox"/> 범주화	- 1만단위 라운딩 처리	980,000원

◆ 위험도 평가는 가명처리에 어떻게 반영하나요?

- 앞 단계에서 수행한 위험성 검토 결과(저·중·고위험)는 항목별 가명처리 방법과 수준을 정할 때 참고하는 정보임
- 위험도가 높을수록, 해당 재식별 위험을 낮추기 위해 가명처리 수준을 높게 설정할 수 있음

**예시** 위험도에 따른 가명처리 방법·기준 예시

- ▶
- ▶
- ▶

- 다만, 위험도만으로 가명처리 방법·수준이 자동 결정되는 것은 아니며, **동일한 항목**이여도 처리 목적 및 상황·맥락 등에 따라 적절한 가명처리 수준은 달라질 수 있음
- 또한, 위험도가 높다고 하여 **항상 가명처리 수준을 높게 설정해야 하는 것은 아니며**, 목적 달성을 위해 필요한 항목은 가명처리 수준을 무조건 높이기보다 처리 환경에 대한 관리·통제 조치를 강화 하는 방식으로 **전체적인 재식별 위험을 낮추는 것도 가능함**

**가명처리 계획에 대한 사전 설명·협의 (선택)**

- 개인정보처리자가 가명처리 계획을 수립한 후에는 가명정보를 활용하려는 자와 해당 계획의 적정성에 대해 사전에 설명·협의할 수 있음
- 이는 가명처리 방법 및 수준이 처리 목적 달성에 적정한지, 활용 가능성을 지나치게 저해하지 않는지 등을 상호확인하기 위한 절차임
- 사전 설명은 가명정보 제공자, 가명정보 활용 예정자뿐만 아니라 가명처리 이후 적정성 검토를 수행할 인원까지 모두 참여하는 설명회 또는 협의회 형태로 운영할 수 있음
- 해당 과정을 통해 불필요한 재처리나 반복 작업을 줄이고, 향후 적정성 검토의 일관성과 예측 가능성을 높일 수 있음

## ② 가명처리 수행 ※ 상세 가명처리 기술 및 기법은 별권 제4장 참조

- ◆ 개인정보처리자는 수립한 **가명처리 계획**을 기반으로 가명처리를 수행하고, 「가명처리 결과서」에 가명처리 수행 결과를 작성 ※ 별권 제5장 서식⑥ 활용
- ◆ 가명처리 대상에 이미지, 영상, 음성, 텍스트 등 **비정형데이터**가 포함되어 있는 경우, 제3장 비정형 데이터 가명처리 기준을 참조하여 가명처리 필요
  - 비정형데이터 가명처리 시 비정형데이터의 특수성을 고려한 추가 검수가 필요하므로, 자체적인 추가 검수 후 「비정형데이터 추가검수 결과서」 작성·보관 필요 ※ 별권 제5장 서식⑥ 활용

### ◆ 특이정보\* 가명처리 방안

- 특이정보는 개인식별 위험을 높이는 대표적 항목이면서도, 과학적 연구 특히, AI 모델 성능을 높이기 위한 핵심 요소이기도 함
  - AI 모델 개발시 특이정보 학습은 데이터셋이 가진 정교한 패턴을 발견하고 예외적인 상황에도 높은 성능을 발휘할 수 있게 함

#### 예시

- AI의 공정성 차원에서도 특이정보를 충분히 학습하는 것은 AI 모델이 소수집단의 특성을 반영하고 편향을 줄이도록 하는데 기여
- 또한, 특이정보는 소수집단의 특성을 나타내는 경우가 많기 때문에, 해당 정보가 무시되는 경우 차별적·편향적 결과 초래 위험 존재

예시

⇒ 따라서 목적상 특이정보 활용이 필요한 경우, 무조건 삭제하는 것이 아니라 안전하게 가명처리하여 각별한 위험관리 하에 활용해야 함

● (활용조건) <sup>1)</sup>데이터 안전조치 + <sup>2)</sup>환경적 안전조치 병행

1) 데이터 안전조치 (→ 전체 데이터셋 위험도 조정)

- 특이정보 활용 시, 다른 식별가능 항목의 가명처리 수준을 강화해 ‘데이터셋 전체의 재식별 위험’을 낮추는 방식으로 관리

2) 환경적 안전조치 (→ 안전한 처리환경에서 활용)

- 특이정보 활용 시 위험 증가 우려가 있는 만큼, 기관 내부 분석공간·폐쇄형 연구망·SW반입 제한 구역 등 안전한 환경에서 활용할 것을 권장

## 4

1, 2, 3단계에서 수행한 가명처리 결과가 적정한지에 대해 검토하는 단계

◆ (검토 내용) 다음 항목을 검토하여 가명처리의 적정·부적정 여부를 결정

① 가명정보 처리 목적 적합성

- 보호법 제28조의2에서 정한 가명정보 처리 목적(통계작성, 과학적 연구, 공익적 기록보존)에 해당 하는지? 처리 목적을 구체적으로 설정하였는지?

▶ 「 . 」 「 」

② 가명정보 이용항목 적합성

- 처리 목적 달성에 필요한 항목만 처리 대상에 포함하였는지?

▶ 「 . 」 「 」

③ 위험도 판단의 적정성

- 활용 주체, 처리 환경, 개별 사례별 특수성 등을 고려하여 적정하게 위험도를 판단하였는지? 판단 근거가 적절한지?

▶ 「 」

④ 가명처리의 적정성(재식별 위험성)

- 항목별 가명처리 방법과 수준을 적절하게 설정하여 처리하였는지?
- 가명정보 데이터 자체의 위험성, 처리환경의 위험성을 종합적으로 고려하였을 때 재식별 위험성을 충분히 낮추었는지?

▶ 「 」 「 」 「 」 「 」

⑤ 목적 달성 가능성

- 당초 설정한 가명정보 처리 목적을 달성할 수 있는지?  
(과도한 가명처리로 처리 목적 달성이 어려워지지 않는지?)

▶ 「 」 「 」 「 」

◆ (검토 방법) 위험도에 따라 차등화된 적정성 검토방식 적용 가능

● 저위험

▶ \_\_\_\_\_

● 중위험

▶ \_\_\_\_\_

● 고위험

▶ \_\_\_\_\_

◆ 담당자 검토<sup>(저위험)</sup>, 내부 심의<sup>(중위험)</sup>, 적정성 검토위원회<sup>(고위험)</sup> 등을 통해 수행한 적정성 검토 결과는 「적정성 검토 결과서」에 작성하여 보관 ※ 별권 제5장 서식 7 활용

◆ 적정성 검토 결과 부적정으로 판단된 경우, 1, 2, 3단계의 절차를 다시 수행하거나 부분적으로 추가 가명처리를 수행해야 함



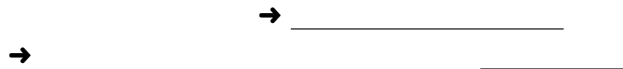
가명정보 처리 위험도별 작성 필요서류

가명처리 단계	서류양식 10종 ※ 별권 제5장 서식 참조	작성 필요서류		
		● 저위험	● 중위험	● 고위험
(1단계) 사전준비	① 가명정보 이용·제공 신청서	△ ※ 내부 활용 시 생략	✓	✓
	② 가명정보 활용 계획서	✓	✓	✓
(2단계) 위험성 검토	③ 위험성 검토서	✓	✓	✓
(3단계) 가명처리	④ 가명처리 계획서	△ ※ 반복·유사 활용 시 생략	✓	✓
	⑤ 가명처리 결과서	△ ※ 반복·유사 활용 시 생략	✓	✓
	⑥ 비정형데이터 추가검수 결과서	- ※ 비정형데이터 포함시 위험도 상향	△ ※ 비정형데이터 미포함시 생략	△ ※ 비정형데이터 미포함시 생략
(4단계) 적정성 검토	⑦ 적정성 검토 결과서	△ ※ 반복·유사 활용 시 생략	✓	✓
	⑧ 적정성 평가위원 서약서	-	△	✓
(5단계) 안전한 관리	⑨ 안전조치 이행확약서	-	△	✓
	⑩ 가명(추가)정보 관리대장	✓	✓	✓
		필수 서류 3종	필수 서류 7종	필수 서류 9종
		선택 서류 4종	선택 서류 3종	선택 서류 1종
✓ 필수 서류(반드시 작성)		△ 선택 서류(생략 가능)	- 작성 불필요	

※ 상세 서식 및 작성방법, 대체 가능 서류 등은 별권 제5장 참조

**여기서 질문** 반복·유사 활용의 경우 어떻게 적정성 검토를 수행하나요?

- 
- 
- 



- 
- 

**예시 ①** 기존 사례와 완전히 동일한 반복 활용

- ▶
- ▶

**예시 ②** 기간 연장 또는 연구자 일부 변경

- ▶
- ▶

**예시 ③** 기존 활용 건에 새로운 데이터 항목이 추가된 경우

- ▶

# 5

## 가명정보를 활용하면서 재식별 위험을 지속 관리하는 단계

### ◆ 재식별 가능성 모니터링

- 적정성 검토 이후 가명정보를 목적에 맞게 활용하는 과정에서 재식별이 되지 않도록 하기 위해 **보호법 및 내부관리계획**에 따라 기술적·관리적·물리적 안전조치 등 **안전조치를 이행**하여야 함

※ 구체적 내용은 별권 제3장 안전성 확보조치 이행 가이드 참조

- 또한, 개인정보처리자는 가명정보 활용 과정에서 특정 개인이 **재식별될 위험**이 있는지 여부를 **지속적으로 모니터링** 하여야 함

※ 가명처리 기술의 취약점으로 인한 재식별 가능성, 다른 정보와 결합 시 재식별 가능성이 있는 새로운 공개데이터의 발생 여부, 재식별 기술의 발전 등

### ◆ 가명정보 파기

- 개인정보처리자는 가명정보의 처리 목적을 달성하였거나 설정한 가명정보 **처리 기간이 도래**하였을 시 **가명정보를 파기**하여야 함
- 가명정보를 파기한 경우에는 **파기한 날부터 관련 기록을 3년 이상 보관**하여야 함

#### 🔑 관련 규정

개인정보 보호법 제28조의4(가명정보에 대한 안전조치의무 등)

<p>가명정보를 생성·처리하는 과정에서 재식별될 위험이 있는지 여부를 지속적으로 모니터링하여야 하며, 안전조치를 이행하여야 함</p>	<p>가명정보를 생성·처리하는 과정에서 재식별될 위험이 있는지 여부를 지속적으로 모니터링하여야 하며, 안전조치를 이행하여야 함</p>	<p>가명정보를 생성·처리하는 과정에서 재식별될 위험이 있는지 여부를 지속적으로 모니터링하여야 하며, 안전조치를 이행하여야 함</p>	<p>가명정보를 생성·처리하는 과정에서 재식별될 위험이 있는지 여부를 지속적으로 모니터링하여야 하며, 안전조치를 이행하여야 함</p>
--	--	--	--

### ◆ 가명정보 활용 기간 연장

- 최초 가명정보 활용을 위해 **최초 설정한 기간이 도래하였다**라도 **목적이 미달성되었거나 목적이 지속되어 활용이 계속 필요**한 경우, 개인정보처리자는 필요한 만큼 **처리기간을 연장하여 새로 정할 수 있음**
- 가명정보 처리기간 연장은 기존에 활용하던 건의 **반복·유사 활용인 경우에 해당\***되므로 **저위험으로 판단**할 수 있으며,

- 이에 따라, 가명처리·적정성 검토 관련 별도 서류작성 없이 **담당자가 ‘연장 사유’** 등 필요사항만 확인하는 방식으로 **간단히 갱신 가능**



#### 관련 규정

개인정보 보호법 제28조의4(가명정보에 대한 안전조치의무 등)



# 3

## 비정형데이터 가명처리 기준

가명정보 처리 가이드라인

가명정보 제도 안내편







# 제3장 비정형데이터 가명처리 기준

## 1

- AI 기술 발전과 컴퓨팅 자원 발달로 데이터 활용수요가 전통적 정형데이터(수치)에서 비정형데이터(이미지, 영상, 음성, 텍스트)로 변화

※ 전 세계 데이터 중 이미지, 영상, 음성, 텍스트 등 비정형데이터가 최대 90%를 차지 (IDC, '23)

### 참고 정형데이터와 비정형데이터의 차이점

▶	▶
---	---

- 비정형데이터도 가명정보 특례를 통해 과학적 연구 목적 등으로 정보주체 동의 없이 가명처리하여 AI 연구개발 등에 활용 가능

### 참고 비정형데이터의 가명처리·활용 예시

▶	▶
▶	▶
▶	▶

## 2

- ◆ **(개인식별성 판단의 어려움)** 개인식별 가능 정보와 그렇지 않은 정보의 구분이 상대적이며, 처리 목적·환경 등에 따라 다르게 판단될 수 있음

예시

예시

- ◆ **(가명처리 기술의 불완전성)** 비정형데이터 내 개인식별 위험성이 있는 모든 항목을 완벽하게 탐지·처리할 수 있는 기술이 부재

예시

예시

- ◆ **(재식별 공격 위험)** AI 및 데이터 복원기술 발달로, 다른 정보와의 연계·결합 없이도 개인 재식별 공격 위험성 증가

예시

예시

\* Kaur, J., Singh, W. Tools, “Techniques, datasets and application areas for object detection in an image: a review”, Multimed Tools Appl 81 (2022)  
† Deng, Jiangyi et al, “Catch You and I Can: Revealing Source Voiceprint Against Voice Conversion”, ArXivabs/2302.12434 (2023)  
‡ R. Dahl, M. Norouzi and J. Shlens, “Pixel Recursive Super Resolution”, 2017 IEEE International Conference on Computer Vision (ICCV) (2017)

### 3

1

- 연구목적에 맞춰 데이터 자체 훼손을 최소화하면서 관리적·환경적 통제 등 다양한 안전성 확보 방안 적용 가능
  - 가명처리시 연구목적 달성이 불가능한 필수항목을 남기는 대신, 그 외 항목에 대한 가명처리 수준을 높이거나, 타 정보·SW 반입 제한 등 충분한 안전조치를 보완하여 활용

2

- 가명처리 기술의 한계 보안을 위해 다음 조치를 이행할 것을 권고

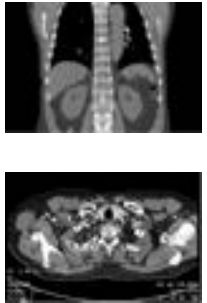

- ▶
- ▶

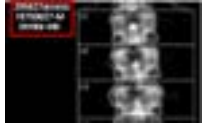

3



- AI 맥락에서 나타날 수 있는 다양한 위험을 사전에 완벽하게 제거하는 것은 불가능하므로, AI 서비스 제공·운영과정에서도 개인식별 위험성 및 프라이버시 침해 가능성 지속 모니터링 필요

주요 비정형데이터 가명처리 시나리오 예시

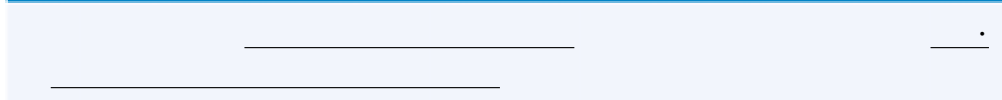
(사례1) 유방암·골밀도 감소 진단 시 개발



<p>〈흉부 CT사진〉</p> 	<p>개인식별 위험성 검토</p>	<ul style="list-style-type: none"> <li>- 흉부 CT사진만으로는 <b>개인식별 위험성 거의 없음</b></li> <li>- <b>개인당 200장씩 촬영된 CT사진이 활용</b>되는 연구로서 3차원 재건 기술 등을 통해 신체형상의 입체적 복원이 가능하고, 복원 시 특이한 외형·흉터 등이 있는 극히 일부 환자의 경우 낮은 확률로 <b>개인식별 위험성 존재</b></li> <li>- 클라우드 기반 폐쇄연구분석환경*을 이용하고 인가되지 않은 데이터·프로그램 반입을 철저히 통제하고 있어 <b>3차원 재건기술 적용 불가</b></li> </ul>	<p>(그대로 활용)</p> 
	<p>데이터 처리 방안</p>	<p>→ 3차원 재건으로 인한 개인식별 위험성이 존재하나, 환경적 통제로 인해 해당 위험이 발생하지 않으므로 <b>별도의 가명처리 없이 그대로 활용 가능</b></p>	

<p>〈CT사진 내 환자관련정보〉</p> 	<p>개인식별 위험성 검토</p>	<ul style="list-style-type: none"> <li>- 이미지 내 표시된 환자관련정보*는 타 정보와 결합되어 분석될 경우 <b>개인식별 위험성이 있음</b></li> <li>- 해당 정보는 <b>연구에 필요하지 않은 정보임</b></li> </ul>	<p>(블랙마스킹 처리)</p> 
	<p>데이터 처리 방안</p>	<p>→ 블랙마스킹 기법을 통해 <b>환자관련정보 삭제</b></p>	

<병리기록 텍스트> 	<b>개인식별 위험성 검토</b>	- 암병리기록지에는 연구에 불필요한 다양한 <b>개인식별가능정보가 정제되지 않은 형태로</b> 존재하여 개인식별 위험이 존재	(정형데이터로 변환 후 활용) 
	<b>데이터 처리 방안</b>	→ 자연어처리 기술을 통해 <b>정형데이터로 변환한 후에 활용</b> 하고 개인식별 위험성이 있는 항목이 있으면 가명처리 수행  → 자연어처리 기술 및 텍스트데이터 가명처리 기술의 정확도가 100%가 아니므로, 정형 데이터 변환 후 추가 <b>전수검사 등을 보완</b>	


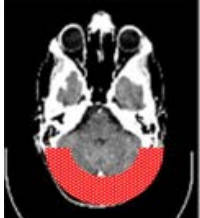
(사례2) 구강질환 진단 AI 개발



<구강 촬영사진> 	<b>개인식별 위험성 검토</b>	- 구강사진 자체로는 <b>개인식별 위험성 거의 없음</b> - 충치 영역 외 부분은 연구에 필요 없음 - 구강사진에 대한 <b>메타데이터(이름, 나이 등)는 구강사진과 결합되어 개인식별 위험성 존재</b>	(충치부분 : 그대로 활용) (그 외 : 블러링 처리) 
	<b>데이터 처리 방안</b>	→ 연구에 필요한 <b>충치 영역은 그대로 활용</b> 하고, 연구에 필요 없는 <b>그 외 영역은 블러링 처리</b>  → <b>메타데이터는 연구에 필요 없어 삭제</b>	



**(사례3) 안면골 골절 진단 AI 개발**

☞ 개인식별 위험성을 낮추기 위해 연구목적에 필요 없는 영역만 마스킹 처리하여 활용

<p>〈Facial CT사진〉</p> 	<p><b>개인식별 위험성 검토</b></p> <ul style="list-style-type: none"> <li>- CT사진 자체로는 <b>개인식별 위험성 거의 없음</b></li> <li>- 3차원 재건 기술을 통해 입체적 복원이 가능하며, 복원 시 <b>특이한 얼굴·외형, 알려진 얼굴</b> 등 극히 일부 환자의 경우 낮은 확률로 <b>개인식별 위험성 존재</b></li> <li>- 대용량 영상·이미지를 활용하는 연구로 3차원 재건이 가능하나, <b>가장자리 마스킹 기법</b>을 활용하여 <b>3차원 재건 공격 위험을 낮출 수 있음</b></li> <li>- <b>후두부(뇌 뒷부분) 영역은 연구에 필요 없음</b></li> </ul> <p><b>데이터 처리 방안</b></p> <p>→ 연구에 필요한 <b>안면부는 그대로 활용</b>하고, 연구에 필요없는 <b>후두부는 마스킹하여 3차원 재건위험을 낮추고 활용</b></p>	<p>(안면부 : 그대로 활용) (후두부 : 마스킹 처리)</p> 
--	---	--

**(사례4) 자율주행차 주행 시 비정상 상황인지 AI 개발**





☞ 연구 목적에 필요 없는 영역만 마스킹 처리하여 활용

<p>〈얼굴·차량번호판〉</p> 	<p><b>개인식별 위험성 검토</b></p> <ul style="list-style-type: none"> <li>- 사람의 얼굴이 선명히 보이는 경우, 차량 번호판이 그대로 노출되어 차량탑승자 유추가 가능한 경우 등에 <b>개인식별 위험성 존재</b></li> <li>- 연구목적상 사람·차량의 전체 형상과 움직임만 파악 가능하면 되므로 <b>얼굴·차량번호판은 마스킹 해도 무방</b></li> </ul> <p><b>데이터 처리 방안</b></p> <p>→ 얼굴·차량번호판 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 <b>마스킹하여 활용</b></p>	<p>(마스킹 처리)</p> 
---	--	---

**(사례5) 고속도로 다인승전용차로 단속 AI 개발**



**고속도로 통행차량 이미지** **다인승**  
**전용차로 위반\*을 단속하는 AI 개발**

☞ 연구 목적에 필요 없는 영역만 블러링 처리하여 활용



<p>&lt; 도로 통행차량 사진 &gt;</p> 	<p><b>개인식별 위험성 검토</b></p>	<ul style="list-style-type: none"> <li>- 탑승자의 얼굴이 선명하게 촬영된 경우, 차량 외부에 특이점이 있는 경우 등은 <b>개인식별 위험성 존재</b></li> <li>- 연구 목적상 특정한 식별·구분이 필요없고 (1) 사람인지 아닌지 여부, (2) 차량 탑승인원이 몇 명인지만 확인할 수 있으면 됨</li> <li>- AI가 사람인지 여부는 판단할 수 있도록 하되, 탑승자가 누구인지는 판별 불가능하도록 블러링 등 처리 필요</li> </ul>	<p>(① 탑승자 위치·범위 파악)</p> 
<p>&lt; 특이점 있는 차량 &gt;</p> 	<p><b>데이터 처리 방안</b></p>	<ul style="list-style-type: none"> <li>→ 특이점이 있는 차량 이미지는 삭제</li> <li>→ 블러링 수준(1~10단계)별로 데이터를 가명 처리한 후, 식별위험이 없으면서 AI 정밀도를 어느 정도 확보할 수 있는 <b>블러링 수준을 결정</b></li> </ul> <p>※ 적정성 검토 단계에서 목적 달성 가능성, 학습 데이터의 개인식별 위험성 등을 검증</p>	<p>(② 블러링 처리)</p> 

**(사례6) 한국어 대화가 가능한 AI 챗봇 개발**

☞ 개인식별 위험 항목을 엄격히 필터링하여 제거하고 메타데이터는 삭제하여 활용

<p>&lt; 대화텍스트 파일 &gt;</p> 	<p><b>개인식별 위험성 검토</b></p>	<ul style="list-style-type: none"> <li>- 일상대화 데이터(텍스트)에는 사생활 관련 정보 등 <b>개인식별 위험성이 높은 정보가 상당수 포함</b></li> <li>→ <b>메타데이터(대화 사용자 ID)를 삭제</b>하고 랜덤 ID로 대체하여 특정 개인과의 연결성을 제거</li> <li>→ 개인식별 위험이 있는 항목들을 엄격하게 <b>필터링하여 가명처리(치환*, 삭제)</b></li> <li>* 이메일 주소를 '[MAIL]'로 대체하는 등</li> </ul>	<p>(메타데이터 삭제, 개인식별정보 필터링·제거)</p> 
---	---------------------------	---	--




☞ 학습에 활용된 가명정보가 AI 챗봇 답변에 그대로 발화되지 않도록 조치

<챗봇 답변 시 위험성> 	<b>개인식별 위험성 검토</b> - 언어모델 학습에 활용된 <b>가명정보가 AI 챗봇의 답변으로 발화될 시</b> , 제대로 처리되지 않은 가명정보의 노출 등으로 인한 <b>개인식별 위험성이 높음</b>	(학습DB, 답변DB 분리) 
	<b>데이터 처리 방안</b> → 언어모델 학습을 위한 ‘ <b>학습데이터베이스</b> ’와 챗봇 답변을 위한 ‘ <b>답변데이터베이스</b> ’를 분리하여 학습에 활용된 문장이 그대로 노출되지 않도록 조치 ※ <b>답변데이터베이스에 개인 식별위험이 있는지 충분히 점검</b>	

**(사례7) 콜센터 직원 실습용 가상상담 시나리오 생성 AI 개발**

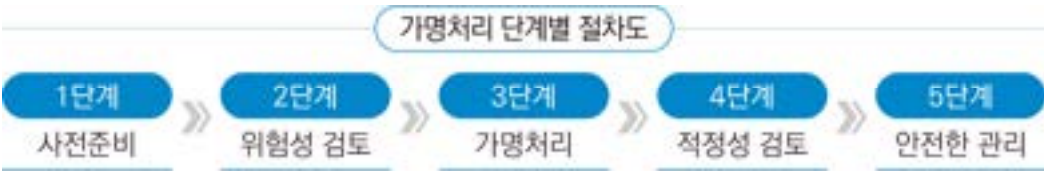
<b>AI 개발</b>	<b>음성 상담정보</b>	<b>상담 실습교육용</b>
--------------	----------------	-----------------

☞ 음성정보를 텍스트로 변환(STT, Speech To Text)한 뒤 가명처리하여 활용

<음성 상담파일> 	<b>개인식별 위험성 검토</b> - 상담음성파일에는 고객 및 상담사의 <b>실제 음성 데이터</b> (목소리, 음색, 억양, 발음 등)가 포함되어 있고, 대화내용에는 다양한 <b>개인식별가능정보가 정제되지 않은 형태로 존재</b> - 가상상담 시나리오 생성 AI 개발에는 상담목적 및 고객 특성에 따른 질의-응답과 대화의 흐름 파악이 중요하며 <b>실제 음성 자체는 필요하지 않음</b>	(① 텍스트로 변환)  ▼ (② 개인식별정보 치환·삭제) 
	<b>데이터 처리 방안</b> → 음성변환(STT) 기술을 통해 <b>텍스트로 변환</b> 한 뒤, 개인식별 위험성이 있는 항목들을 <b>가명 처리(치환·삭제)하여 활용</b> → 텍스트데이터에 대한 가명처리 기술의 정확도가 100%가 아니므로, 추가 <b>전수검사를 통해 식별 위험성이 있는 정보를 제거</b>	

# 4

■



## ① 사전준비 단계

·

- ◆ 비정형데이터 내 개인식별 가능성이 있는 항목들을 도출하고, 목적 달성에 필요한 항목의 종류·범위를 명확히 하여 가명처리 대상 선정

## ② 위험성 검토 단계

·

- ◆ 비정형데이터는 적절한 가명처리 방법 설정, 가명처리 기술의 한계 보완 등 데이터 특수성을 고려한 전문적인 처리 및 검토가 필요하므로,
  - 비정형데이터를 가명처리하여 활용하는 경우 활용 주체·처리 환경을 기준으로 판단한 기본 위험도의 상향 조정이 권장됨

·

▶ →

▶ →

- **비정형데이터의 특성을 고려하여 가명처리·활용 시 다양한 위험성을 종합적으로 검토하여 가명처리 방법·수준을 결정에 참고하여야 함**

(1) 데이터 자체의 위험성

- **(식별성)** 비정형데이터는 정형데이터와 달리 **식별정보\***와 **식별가능정보\*\***의 절대적 구분이 어렵기 때문에, 처리 목적·방법 등 데이터 처리 맥락을 고려하여 항목별 개인식별 가능성을 상대적으로 판단하여야 함

\* (식별정보) 특정 개인과 직접적으로 연결되어 다른사람과 구분되는 정보

\*\* (식별가능정보) 단일 항목으로는 식별 가능성이 없으나, 가명처리 대상이 되는 다른 항목과 결합하는 경우 식별 가능성이 높아지는 정보

예시

·

·

예시

- **(특이정보)** 개인의 특이한 외형적·행태적 특징 또는 개인과 연관된 객체·사물 등의 특이성으로 인해 식별 위험성이 높은 항목도 검토

예시

·

·

예시

예시

- (재식별 시 영향도) 가명처리된 비정형데이터가 재식별될 경우 특정 정보주체에게 사회적 파장 등 영향도가 높은 항목이 있는지 검토

예시

(2) 처리 환경의 위험성

- (활용 형태) 처리자가 보유하고 있는 정보 또는 접근·입수 가능한 정보와 이용 범위·유형을 고려하여 식별가능한 항목이 있는지 검토

예시

예시

예시

예시

- (처리 장소) 가명정보가 해당 가명정보 외에 다른 정보의 접근·입수, 재식별 기술의 접근이 제한된 환경·장소에서 처리되는지 검토

예시

예시

- (처리 방법) 가명정보가 다른 정보와 연계·결합되거나 반복 제공 등이 예정된 경우, 식별가능성이 높아지는 항목이 있는지 검토

예시

예시

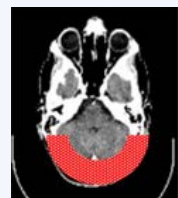
### ③ 가명처리 단계

- ◆ 가명처리가 필수적이지 않은 항목과 가명처리가 필요한 항목을 구분하고, 가명처리가 필요한 항목은 합리적인 가명처리 방법과 수준을 결정

→ 비정형데이터 항목 중 처리목적 달성을 위해 반드시 필요하지만, 개인식별 위험성이 낮은 정보는 가명처리하지 않고 그대로 사용 가능

→ 비정형데이터 항목 중 가명처리해도 처리목적 달성이 가능하고, 개인식별 위험성이 높은 정보는 가명처리하여 활용하여야 함

예시 CT 사진을 활용하여 얼굴뼈 골절 진단을 위한 AI 연구개발 시



**예시** 구강사진을 활용하여 충치를 분석·진단하는 AI 개발 시



❖ 비정형데이터 가명처리 후 추가검수 수행

- 비정형데이터 특성상 가명처리 기술이 불완전하여 해당 한계에 따른 잔존 위험(Residual Risk)이 남아있을 수 있으며,
  - 이를 제거하기 위해 가명처리 기술 적용 이후, 처리 결과에 대한 자체적인 추가검수가 요구됨
- 추가 검수는 비정형데이터의 특성과 위험 수준을 고려하여 다양한 검수방식 중 적절한 방식을 선택·혼용하여 검수 수행

**비정형데이터 추가검수 방식**

비정형데이터 추가검수 방식

- ▶
- ▶
- ▶
- ▶

[Redacted]

▶ . . .

▶ .

▶

[Redacted]

▶

▶

. . .

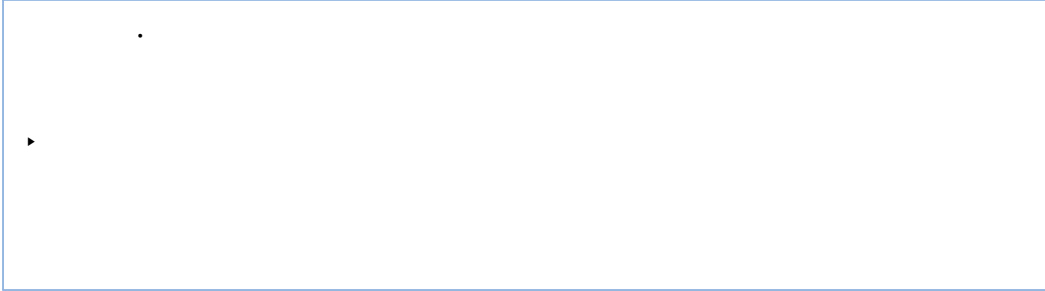
▶

[Redacted]

▶ . . .

▶

\* 예) 영상 내 얼굴 부분 가명처리 시 그림자, 마스크, 음영 등



**여기서 질문** 비정형데이터는 무조건 전수검수하는 것이 안전하지 않나요?

■

■

■

■

**◆ 추가검수 과정 및 결과 기록·보관**

- 비정형데이터에 대한 추가검수를 수행한 뒤에는 △적용한 검수 방식, △검수 과정, △검수 결과 등을 기록·보관할 것을 권고함 ※ 별권 제5장 서식⑥ 활용
- 이러한 기록은 비정형데이터 가명처리 시 기술적 한계로 일부 잔존 위험(Residual Risk)이 존재할 수 있다는 점을 전제로,
  - 해당 위험을 수용 가능한 수준으로 낮추기 위해 필요한 적절한 검수와 조치를 수행하였는지 확인할 수 있는 근거 자료가 됨
- 또한 가명정보 활용 과정에서 예기치 못한 식별 가능정보가 발견되는 경우에도, 관련 절차를 충실히 이행하고 충분한 검수 노력을 기울였는지 여부를 설명할 수 있는 합리적인 면책 근거로 활용될 수 있음

#### ④ 적정성 검토 단계

- ◆ 비정형데이터의 특성과 처리 목적·환경 등을 고려하여 합리적인 방법·수준으로 가명처리를 수행하였는지 검토
- ◆ 비정형데이터 가명처리에 활용한 기술의 적절성·신뢰성을 검토하고, 해당 기술의 한계 등으로 인한 잔존 위험(Residual Risk)을 충분히 제거하기 위한 안전조치\*를 수행하였는지 검토
  - \* AI 출력 데이터에 대한 엄격한 필터링 조치, 학습데이터 저장 및 출력 시 차분 프라이버시 기법 적용, 가명정보 처리자에 대한 보안서약서 징구, 내·외부 정보 및 복원기술 활용이 제한된 안전한 분석환경 내에서 활용 등
- ◆ 비정형데이터는 가명처리 시 데이터의 특성과 관련 기술 발전 수준, 재식별 위험을 종합 고려하여야 하고, 이를 위한 전문성이 필요하므로,
  - 외부전문가를 포함한 적정성 검토위원회를 구성하여 객관성·전문성 있는 검증을 받을 것을 권고

#### ⑤ 안전한 관리 단계

- ◆ 가명정보 특례를 활용하여 AI 학습·연구개발을 하고자 하는 자는 AI 기술·서비스의 특성을 고려하여 사전·사후적으로 발생할 수 있는 다양한 위험을 낮추기 위한 충분한 조치를 취해야 함
- ◆ 다만, AI 맥락에서 나타날 수 있는 다양한 위험을 사전에 완벽히 제거하는 것은 현재 기술상 불가능하므로,
  - 잔존 위험(Residual Risk)을 최소화시키기 위한 노력 정도에 따라 사후관리의 이행 수준을 판단하여야 함
- ◆ AI 서비스 운영 과정에서도 개인식별 위험성 및 프라이버시 침해 가능성이 높아지지 않는지 지속적으로 모니터링하여야 함
  - 또한, 개인식별 위험 증가 및 프라이버시 침해 등 관련 문제 발견 즉시, 해당 가명정보 처리를 중단하고 관련 위험을 제거하여야 함



# 4

## 자주 묻는 질문(Q&A)

가명정보 처리 가이드라인

가명정보 제도 안내편







## 제4장 자주 묻는 질문 (Q&A)

### 가명처리 관련

**Q** 정부가 일률적인 가명처리 방법·수준을 정해주면 안되는지?

**A**

- 
- 

**Q** 개인정보 가명처리 과정에서 일시적으로 가명정보와 추가정보가 같은 서버에 존재하게 되는 것이 보호법 제28조의4 제1항 및 동 법 시행령 제29조의5 제1항 제2호·제3호가 정하는 안전조치의무 위반인지 여부

**A**

## 적정성 검토 관련

**Q** 가명처리 적정성 검토는 반드시 외부전문가에게 받아야하는지?  
적정성 검토에 참여하는 자의 자격요건이 있는 것인지?

**A**

- 「개인정보 보호법」 제24조 제2항에 따라 적정성 검토는 반드시 외부전문가에게 받아야 하는 것으로 규정되어 있습니다.
- 「개인정보 보호법」 제24조 제2항에 따라 적정성 검토에 참여하는 자는 개인정보 보호법 제24조 제2항에 따라 적정성 검토에 참여하는 자의 자격요건이 있는 것인지를 확인하여야 합니다.

「개인정보 보호법」 제24조 제2항

## 가명정보 결합 관련

**Q** 가명정보 결합 시 어떤 결합전문기관에 신청해야 하는지?

**A**

**Q** 가명정보 자체결합(셀프결합)이 무엇인가요?

**A**

- 「개인정보 보호법」 제24조 제2항에 따라 적정성 검토는 반드시 외부전문가에게 받아야 하는 것으로 규정되어 있습니다.

**Q** 개인정보 가명처리 및 분석 지원의 범위 (ex. 모의결합 등)

**A**

- 
- 

**Q** 가명정보 결합 신청 중 신청 내용의 변경이 없는 기관의 데이터를 파기 하지 않고 변경된 데이터와의 결합에 다시 활용 가능한지 여부

**A**

### 가명정보 제공 관련

**Q** 가명정보 제공 시 수수료를 받아도 되는지?

**A**

- 

**Q** 가명정보 유상 판매가 가능한지 여부

**A**

-

## 타 법과의 관계

**Q** 개인정보가 포함된 공공데이터를 가명처리하여 제공해도 「공공데이터법」상 문제가 없는지?



「 」

「 」

「 」

「 」

**Q** 개별 개인정보 관계 법률에서 제3자 제공 등을 제한하는 경우에도 가명처리 시 제3자 제공이 가능한지?



「 」

「 」 「 」

「 」

## 기타

**Q** A부서에서 수집한 개인정보를 B부서에서 익명처리하여 수집 목적 외로 처리할 수 있는지 여부



- 
- 

**Q** 가명정보에서 추가정보를 삭제할 경우 익명정보로 볼 수 있는지 여부

**A** . . .

연번	내 용	페이지
1	가명정보 특례 제도를 개인정보 '수집' 근거로 쓸 수 있나요?	7
2	가이드라인이 적용되는 범위는?	9
3	"가명처리"와 "가명정보의 처리"는 어떻게 다른가요?	11
4	목적 설정 없이 미리 가명처리하여 보유하고 있어도 되나요?	11
5	가명정보를 제공 후 발생한 위법행위도 제공자 책임인가요?	12
6	가명정보를 불특정 다수(제3자)에게 제공(공개)해도 되나요?	13
7	가명정보를 다시 사용하거나 다른 기관에 재제공해도 되나요?	13
8	가명정보 활용 중 의도치 않게 재식별이 된 경우에도 처벌을 받나요?	17
9	가명정보도 국외이전이 가능한가요?	19
10	개인정보에 대한 가명처리 정지 요구가 가능한가요?	19
11	가명정보로 통계작성 후 1:1 맞춤형 타겟 마케팅도 가능한가요?	25
12	AI 기술·서비스 개발 및 고도화 목적도 "과학적 연구"에 해당하나요?	26
13	'민감정보'나 '고유식별정보'도 가명처리해서 쓸 수 있나요?	33
14	가명정보 제3자 제공 시 이용자의 내부관리계획 작성·제출 필요여부	35
15	수탁자가 자체 목적을 위해 가명정보 제도를 활용하여도 되나요?	36
16	의료기관도 병원 외 기관·기업에 가명처리를 위탁해도 되나요?	36
17	'개인정보 이노베이션 존'이 무엇인가요?	39
18	가명처리 시 위험도 산정이 왜 필요한가요?	40
19	정량적인 위험도 산정 기준(점수표)를 제시해줄 수는 없나요?	41
20	반복·유사 활용의 경우 어떻게 적정성 검토를 수행하나요?	50
21	비정형데이터는 무조건 전수검수하는 것이 안전하지 않나요?	71

# 가명정보 처리 가이드라인

**[별권] 가명정보 처리 실무편**

2026. 3.



개인정보보호위원회  
Personal Information Protection Commission



발 간 등 록 번 호
11-1790365-000029-01

# 가명정보 처리 가이드라인

**[별권] 가명정보 제도 실무편**

2026. 3.

# 1

## 용어 정리

가명정보 처리 가이드라인

가명정보 처리 실무편







# 제1장 용어 정리

구분	용어설명
개인정보	<p>살아있는 개인에 관한 정보로서 다음의 정보를 포함함</p> <ul style="list-style-type: none"> <li>- 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보</li> <li>- 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보</li> <li>※ 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 함</li> <li>- 가명처리를 거쳐 생성된 정보로서 그 자체로는 특정 개인을 알아볼 수 없도록 처리한 정보 (이하 '가명정보'라 함)</li> <li>※ 개인정보에 대한 판단기준은 개인정보처리자가 보유한 정보 또는 접근 가능한 등 개인정보 처리 상황에 따라 다르게 판단되어야 함</li> </ul>
가명처리	개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보(이하 '추가정보'라 함)가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것
개인정보처리자	업무를 목적으로 개인정보파일을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관, 법인, 단체 및 개인 등
익명정보	시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보
추가정보	<p>개인정보의 전부 또는 일부를 대체하는 가명처리 과정에서 생성 또는 사용된 정보로서 특정 개인을 알아보기 위하여 사용·결합될 수 있는 정보 (알고리즘, 매핑테이블 정보, 가명처리에 사용된 개인정보 등)</p> <p>※ 가명처리 과정에서 생성·사용된 정보에 한정된다는 점에서 다른 정보와 구분됨</p>
재식별	특정 개인을 알아볼 수 없도록 처리한 가명정보에서 특정 개인을 알아보는 것
가명정보처리시스템	개인정보를 가명처리하거나 가명정보를 처리할 수 있도록 체계적으로 구성한 시스템
결합키	결합 대상 가명정보의 일부로서 해당 정보만으로는 특정 개인을 알아볼 수 없으나 다른 결합대상정보와 구별할 수 있도록 조치한 정보로서, 서로 다른 가명정보를 결합할 때 매개체로 이용되는 값
결합키연계정보	결합키가 동일한 정보에 관한 가명정보를 결합할 수 있도록 서로 다른 결합신청자의 결합키를 연계한 정보
결합대상정보	결합신청자가 결합을 위해 결합전문기관에 제공하는 가명정보에서 결합키를 제외한 정보
결합정보	결합전문기관을 통해 결합대상정보를 결합하여 생성된 정보

구분	용어설명
반출정보	결합전문기관에서 결합된 결합정보 중 결합전문기관의 심사를 통해 반출된 정보
반복결합	시계열 분석 등을 위해 가명정보 결합을 반복하는 결합
반복결합연결정보	반복결합을 통해 반출된 반출정보의 반복적인 분석을 위해 필요한 정보로, 반출시 해당 정보를 포함하여 반출
결합신청자	가명정보의 결합을 신청하는 개인정보처리자 등 ※ 가명정보를 제공하거나 이용하는 자(공공기관, 법인, 단체, 개인 등)
결합전문기관	보호법 제28조의3 제1항에 따라 서로 다른 개인정보처리자 간의 가명정보 결합을 수행하기 위해 개인정보위 또는 관계 중앙행정기관의 장이 지정하는 전문기관
결합관리기관	보호법 시행령 제29조의3 제2항에 따라 특정 개인을 알아볼 수 없도록 가명정보의 안전한 결합을 지원(결합키연계정보를 생성하여 결합전문기관에 제공하는 등)하는 업무를 하는 한국인터넷진흥원 또는 개인정보위가 지정하여 고시하는 기관
적정성 검토	개인정보처리자가 개인정보를 가명처리한 뒤, 적정성 평가 위원회 등을 구성하여 처리 목적의 적합성, 위험성 검토 결과의 적정성, 가명처리 결과의 적정성, 목적 달성 가능성 등을 검토하여 적절히 가명처리가 되었는지 확인하는 절차
반출 심사	결합된 가명정보에 대해 결합 목적, 반출 정보의 관련성, 특정 개인식별가능성, 반출정보에 대한 안전조치 계획 등을 검토하여 가명정보를 활용하고자 하는 자에게 반출하여도 문제가 없는지에 대하여 심사하는 절차
비정형데이터	영상, 이미지, 음성, 텍스트 등 일정한 규격이나 정해진 형태가 없이 구조화되지 않은 데이터

연번	구분	내용	소관부처
1	법률	개인정보 보호법	개인정보위
2	법률	신용정보의 이용 및 보호에 관한 법률	금융위
3	시행령	개인정보 보호법 시행령	개인정보위
4	시행령	신용정보의 이용 및 보호에 관한 법률 시행령	금융위
5	고시	가명정보의 결합 및 반출 등에 관한 고시	개인정보위
6	고시	신용정보업감독규정	금융위
7	훈령	가명정보 관리 및 제공에 관한 규정	국무총리
8	가이드라인	가명정보 처리 가이드라인	개인정보위
9	가이드라인	보건의료데이터활용 가이드라인	보건복지부
10	가이드라인	교육분야 가명·익명정보 처리 가이드라인	교육부
11	가이드라인	공공분야 가명정보 제공 실무안내서	행정안전부
12	가이드라인	금융분야 가명·익명처리 안내서	금융위

※ 법제처 국가법령정보센터(www.law.go.kr) 및 각 정부부처 홈페이지를 통해 확인



# 2

## 가명정보의 결합

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제2장 가명정보의 결합

### 1

- ◆ **(내부 결합)** 개인정보처리자는 자신이 보유하고 있는 가명정보를 통계작성, 과학적 연구, 공익적 기록보존 등 목적으로 결합하여 활용할 수 있음 (부서 내 결합, 조직·부서 간 결합 등)
  - 내부 결합의 방법·절차가 정해져 있지는 않지만, 결합 과정에서도 보호법을 준수하고, 특정 개인을 알아볼 수 없도록 유의하여야 함
  - 내부 결합 시 안전한 결합을 위해 아래와 같이 결합키를 이용한 결합방법을 선택할 수 있음



- 개인정보처리자는 결합된 정보를 활용할 때 특별한 사유(시계열 분석 등)가 없는 한 결합키 등 결합을 위해 사용한 정보는 삭제하여야 함

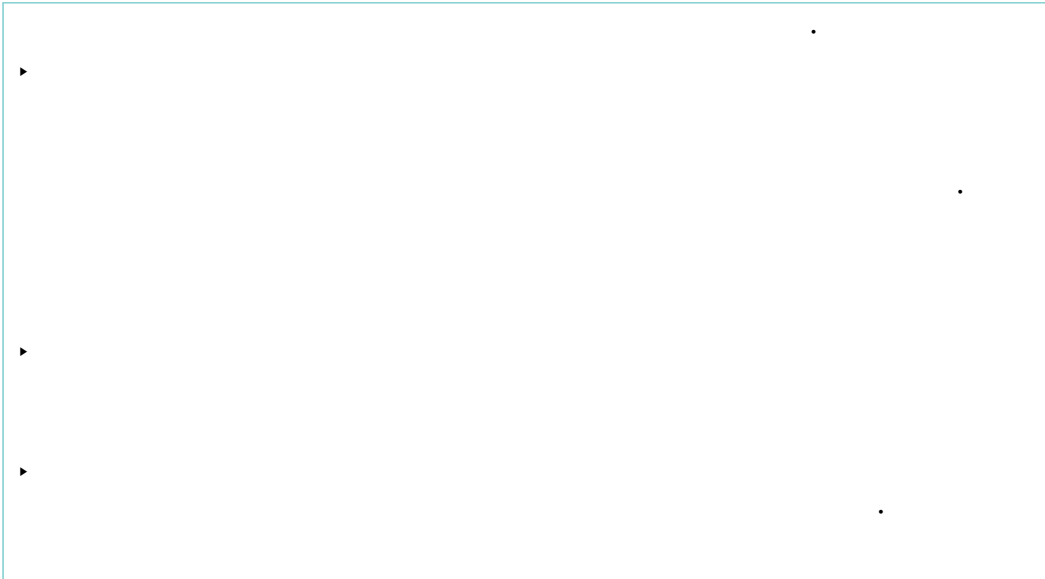
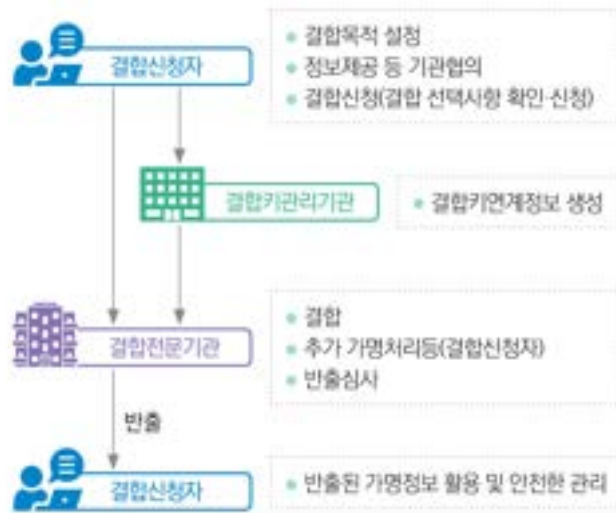
※ (주의) 결합키 생성에 이용된 알고리즘, 매핑테이블 등은 추가정보에 해당하므로, 결합된 가명정보와 분리하여 보관하여야 하고, 접근권한을 분리하여야 함



※ 결합키 생성, 결합 등의 구체적인 내용은 「제2장 가명정보 결합·반출 절차」 참고 가능

❖ **(외부 결합)** 서로 다른 개인정보처리자 간에 가명정보를 결합하여 활용하려는 경우 결합전문기관을 통해 통계작성, 과학적 연구, 공익적 기록보존 등의 목적으로 가명정보 결합이 가능함

- 서로 다른 개인정보처리자가 보유한 가명정보의 결합은 개인정보위 또는 관계 중앙행정기관의 장이 지정한 결합전문기관이 수행함(보호법 제28조의3 제1항)
- 가명정보 결합은 ①결합신청자의 결합신청, ②결합관리기관의 결합키연계정보 생성, ③결합전문기관의 가명정보 결합 및 반출, ④결합신청자의 반출정보 활용 및 관리 등으로 진행함



## 2

- 가명정보 결합·반출은 ① 결합신청, ② 결합 및 추가처리, ③ 반출 및 활용, ④ 안전한 관리의 총 4단계를 거쳐 진행함

- 결합신청자는 결합신청 시 모의결합, 결합률 확인, 가명정보 추출을 선택하여 신청할 수 있음



### ◆ 1단계: 결합신청

- 결합신청자는 신청자 간 결합신청에 필요한 사항\*의 협의, 결합신청서 작성 등 가명정보 결합에 필요한 사전 준비사항을 확인하고 결합전문기관에 결합을 신청함

\* 개인정보파일에서 가명정보 결합 목적 달성에 필요한 항목을 선정, 시계열 분석 여부, 모의결합/결합률 확인/가명정보 추출 신청여부, 결합키 생성항목 등

- 결합신청자는 결합전문기관과 결합일정, 전송방법 등을 협의함

### ◆ 2단계: 결합 및 추가처리

- 가명정보를 제공하는 결합신청자는 결합키관리기관으로부터 결합키 생성에 이용되는 정보(Salt 값)를 수신하여 결합키를 생성하고 결합신청 시 선택한 모의결합, 결합률 확인, 가명정보 추출 등이 완료되면 결합에 필요한 정보를 각 기관에 전송함



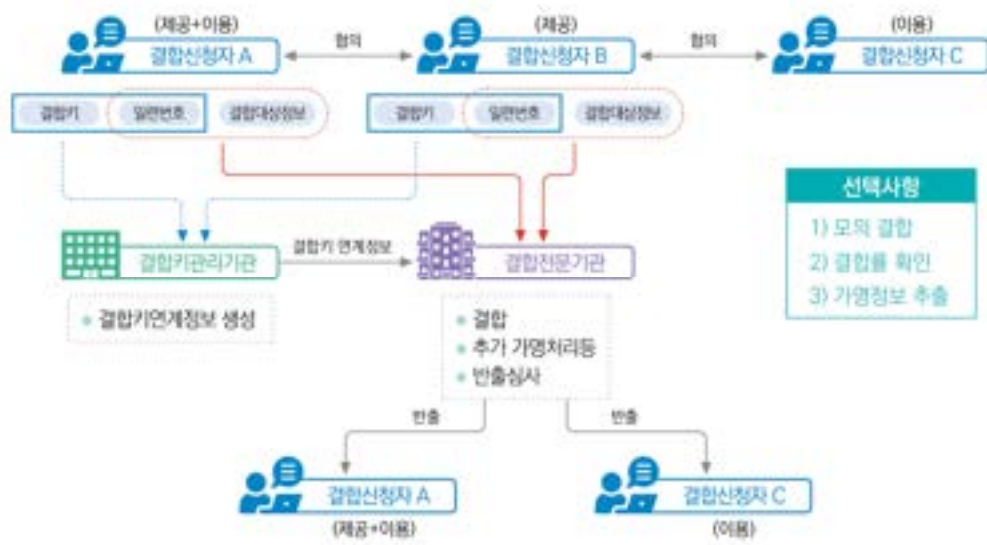
- 결합정보를 이용하는 결합신청자는 결합전문기관 공간에서 추가 가명·익명처리를 하거나, 결합전문기관이 지원하는 분석기능을 신청·이용하여 분석할 수 있음

### ◆ 3단계: 반출 및 활용

- 결합정보 또는 분석결과 등을 반출하려는 경우 결합전문기관에 반출을 신청함

### ◆ 4단계: 안전한 관리

- 결합정보를 이용하는 결합신청자는 반출한 결합정보(이하 반출정보)를 당초 결합신청서 및 반출신청서에 기재한 목적에 따라 처리하고 안전조치 의무 등을 준수하여야 함



절차	결합신청자	결합전문기관	결합관리기관	
1. 결합신청	①결합신청	②결합신청서 검토 및 접수 ③결합 일정·절차 등 협의(결합신청자)	-	
2. 결합 및 추가처리	1. 결합키 생성	①결합키 생성 협의 ③결합키 및 일련번호 생성	-	
	2. 모의결합 (선택)	①결합키 전송 (→결합전문기관)	②모의결합 가능성 검토 및 통지 ③모의결합 대상 결합키 선정 및 전송	-
		④모의결합 대상 가명처리 ⑤모의결합대상정보, 가명처리 내역 및 결합키 전송	⑥가명처리 수준 검토 (필요시 추가처리 요청) ⑦모의결합 수행	-
		⑧모의결합정보 분석(결합전문기관 내) * 반출 불가	⑨모의결합 관련 정보 파기	-
	3. 결합률 확인(선택)	①결합키 및 일련번호 전송 (→결합관리기관) ④결합률 확인	-	②결합키연계정보 생성 ③결합률 측정 및 통보
	4. 가명정보 추출(선택)	①결합키 및 일련번호 전송 (→결합관리기관) ③추출 요청	-	②추출 가능 여부 검토 및 통지 ④추출에 필요한 일련번호 선정 및 전송
	5. 가명처리 및 검토	①가명처리 대상 정보 확정 ②가명처리 ③결합대상정보, 가명처리 내역 및 일련번호 전송 (→결합전문기관)	★가명처리 지원(가능한 경우) ④가명처리 수준 검토 (필요시 추가처리 요청)	-
6. 결합	①결합키 및 일련번호 전송	③결합키연계정보 수신 및 가명정보 결합 * 반복결합의 경우 반복결합연계정보 포함	②결합키연계정보 생성 및 전송 * 반복결합의 경우 반복결합연계정보 포함	
7. 추가처리 (필요시)	①결합정보의 추가처리 및 분석 (결합전문기관 내) * 결합전문기관에 지원 요청 가능	★추가처리 및 분석 지원(가능한 경우)	-	
3. 반출 및 활용	1. 반출	①반출신청	②반출심사위원회 구성·개최 ③반출 승인 및 결합정보 반출 * 반복결합의 경우 반복결합연계정보 포함 ④결합키연계정보 파기	④결합키 및 결합키연계정보 파기 * 단, 반복결합의 경우, 결합키 생성방법 (Salt)·반복결합연계정보 생성방법 (Salt) 보관
	2. 활용	· 반출정보의 활용 원칙 준수 * 반복결합의 경우 반복결합연계정보를 통해 내부에서 연계하여 분석	-	-
4. 안전한 관리	· 안전성 확보 조치 이행 · 가명정보 처리 내역 기록·보관	★반출된 정보 분석 지원(가능한 경우) ★개인정보보호 교육 제공(가능한 경우)	-	

★ : 보호법 제11조의2에 따른 결합전문기관의 업무지원 사항으로, 결합신청자는 결합전문기관이 해당 업무에 대해 지원 가능한 경우 요청할 수 있음

※ 본 가이드라인에서는 결합신청자 기준으로 결합절차 안내

## ① 결합 신청

### 1. 목적 설정 및 결합 가능정보의 탐색

#### ◆ 결합 목적 설정

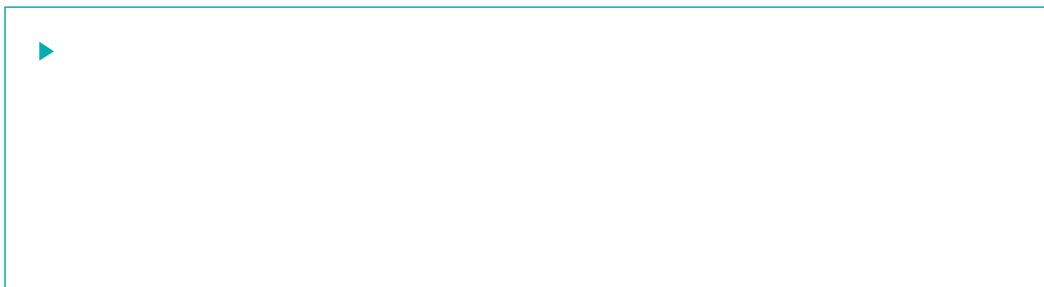
- 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등 결합의 목적을 명확히 설정하여야 함
  - 통계작성을 위한 결합이란, 특정 집단이나 대상 등에 대하여 수량적인 정보를 처리하여 통계 작성을 목적으로 가명정보를 결합하는 것을 말하며 상업적 성격의 통계 작성도 가능함
  - 과학적 연구를 위한 결합이란, 과학적 방법을 적용한 연구로서 자연과학, 사회과학, 기초연구, 응용 연구뿐만 아니라 새로운 기술·제품·서비스 개발 및 실증을 위한 산업적 연구를 포함한 과학적 연구를 목적으로 가명정보를 결합하는 것을 말함
  - 공익적 기록보존을 위한 결합이란, 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 정보를 기록하여 보존하는 것을 의미하며 공공기관뿐 아니라 기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우도 포함한 공익적 기록보존을 목적으로 가명정보를 결합하는 것을 말함

#### ◆ 결합 가능정보의 탐색

- 결합을 추진하려는 결합신청자는 결합 목적을 달성하기 위한 결합 가능정보를 탐색·확인하고 해당 정보의 보유기관과 협의\*함

\* 결합목적을 달성하기 위한 항목의 보유 여부, 가명정보의 제공 가능 여부 등

#### 참 고



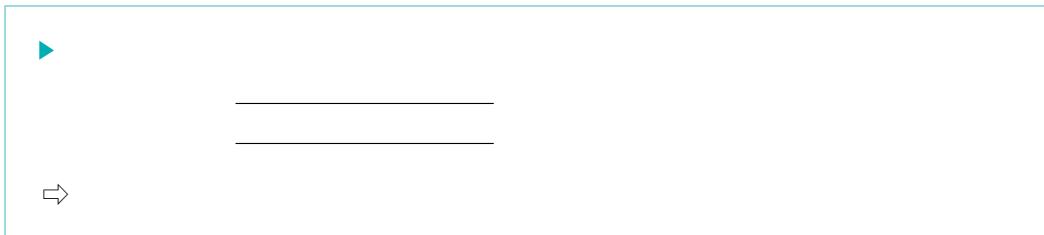
※ 정보보유기관은 결합을 위한 가명정보를 제공에 대한 의무가 있는 것은 아니므로, 해당 정보가 필요한 개인정보처리자 등은 정보보유기관과 협의를 진행하여야 함

## 2. 기관협의

- ◆ 결합신청자는 결합 목적을 설정하고 다른 결합신청자와 가명정보 결합에 대해 협의하는 등 필요한 사항을 사전준비 함

※ 협의사항 : 개인정보파일에서 가명정보 결합 목적 달성에 필요한 항목 선정, 시계열 분석 여부, 모의결합/결합률 확인/가명정보 추출 신청여부, 결합키 생성항목 등

- 결합신청자 간 개인정보의 공통항목 중에서 결합키 생성에 활용할 항목을 결정함



- 결합신청자는 가명정보 결합에 관한 별도의 내부승인절차 등을 진행할 수 있으며, 결합에 대한 계약 체결 등 필요한 조치를 할 수 있음

## 3. 결합 신청

- ◆ 결합을 위해 가명정보를 제공하는 개인정보처리자 또는 결합정보를 이용하려는 자\*(이하 결합신청자) 모두 결합신청서를 작성하여 신청하여야 함 ※ 세부 작성은 제6장 가명정보 결합신청서 작성방법 참고

\* 현재 가명정보를 보유하고 있지 않으나, 결합된 가명정보를 처리할 예정인 자 포함

※ 결합전문기관은 결합 및 반출 등에 필요한 비용을 결합신청자(결합정보를 이용하려는 자)에게 요청할 수 있음

- 결합신청자는 가명정보 지원 플랫폼(dataprivacy.go.kr)을 이용하여 결합전문기관\*에 결합을 신청함

\* 결합전문기관 현황은 가명정보 지원 플랫폼(dataprivacy.go.kr) > 가명정보결합 > 결합안내에서 확인



- 결합신청자는 '가명정보 결합 및 반출 등에 관한 고시(이하 '결합 고시'라 함)'의 제8조에 따른 [별지 제3호] 결합신청서와 첨부 서류\*를 결합 신청 시 제출함

\* 단, 결합신청자가 결합의 선택사항 진행 등의 사유로 결합대상정보에 대한 검토 및 확인을 완료하지 않은 경우, 결합전문기관과 협의하여 결합대상정보의 가명처리 내역에 관한 서류는 가명정보 전송 시 제출할 수 있음

- ◆ 결합신청자는 결합전문기관이 신청서 작성내용(결합 목적 적합성 등) 및 첨부 서류에 대한 보완을 요청한 경우 결합신청자는 해당사항을 보완하여 다시 제출하여야 함
- 결합전문기관은 서류 누락 등 신청서류에 더 이상 보완사항이 없는 경우 결합신청서를 접수하고 결합신청자에게 신청접수 사실을 통지함
- 결합신청을 접수한 이후에도 결합전문기관은 결합 목적, 결합대상 항목 등이 적절한지 여부를 추가로 확인하여 보완이 필요한 경우에는 결합 목적 증빙 자료 제출이나 결합대상 변경 등을 요청할 수 있음
- 결합신청자는 결합 신청 내역에 따라 결합 절차 및 필요한 정보 등을 결합전문기관 및 결합기관과 협의할 수 있음

### 여기서 한번 금융분야와 비금융분야간 가명정보 결합시 어디서 결합해야 하나요?

- 
- 

#### 4. 모의결합, 결합률 확인, 가명정보 추출 등의 선택 신청

- ◆ **(모의결합)** 결합 목적 달성을 위한 정보의 일부를 결합·분석하여 결합의 유용성을 확인하는 절차를 말함
- ◆ **(결합률 확인)** 결합대상정보의 가명처리에 앞서 결합률을 확인하는 절차로서 결합키와 일련번호로 결합률을 확인함
  - \* 결합률 확인 후 결합 절차의 진행여부를 결정할 수 있으며, 결합 절차 진행 결정 이후 가명처리를 수행할 수 있음
- ◆ **(가명정보 추출)** 결합대상정보 중 결합되는 정보를 가명처리하여 전송할 수 있도록 해당 정보를 추출\* 하는 절차를 말함
  - \* 결합키관리기관이 개인정보 침해 우려가 없도록 결합되지 않는 정보를 일부 포함하여 추출
  - 결합대상정보 전부를 가명처리하거나 전송하는 방식에 비하여 효율적이며 안전할 수 있음

신청 항목	처리기관	비고
모의결합	결합전문기관	결합신청자(결합키) → 처리기관(모의결합 대상 결합키) → 결합신청자(모의 결합대상정보) → 처리기관(모의결합) → 결합신청자(분석)
결합률 확인	결합키 관리기관	결합신청자(결합키, 일련번호) → 처리기관(결합률) → 결합신청자(결합률 확인)
가명정보 추출	결합키 관리기관	결합신청자(결합키, 일련번호) → 처리기관(추출된 일련번호) → 결합신청자(가명정보 추출)

## ② 결합 및 추가 가명처리

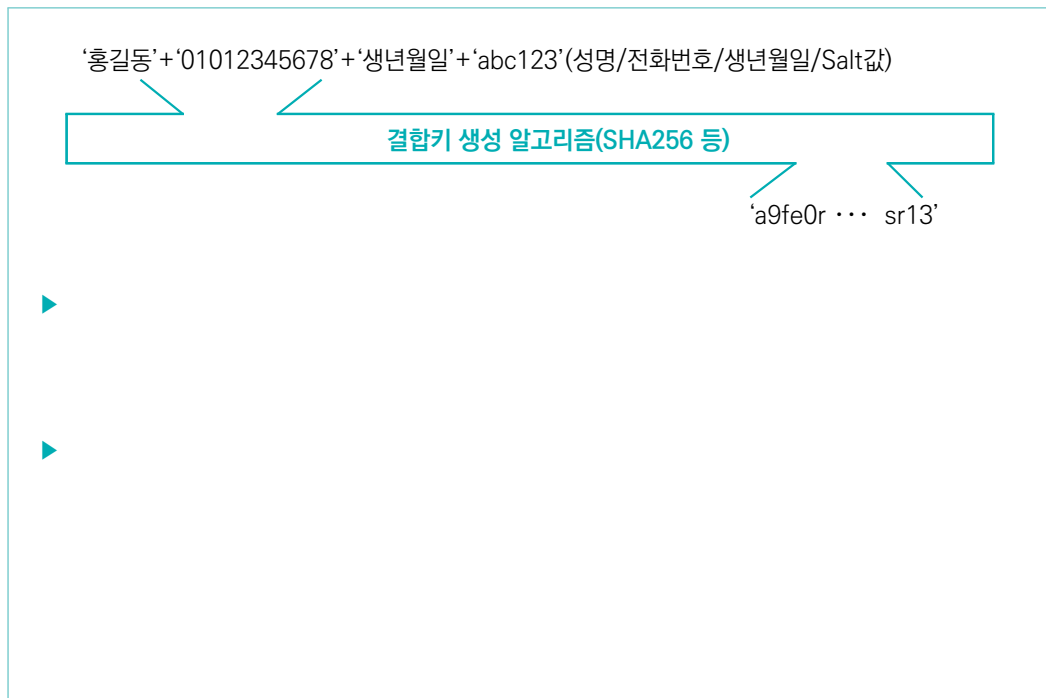
### 1. 결합키 생성

- 가명정보를 제공하는 결합신청자는 결합키관리기관과 결합키 생성에 관한 사항을 협의\*하고 결합키 관리기관으로부터 결합키 생성에 필요한 정보(Salt값)를 수신함

\* 결합키 생성 항목, 인코딩 방식, 결합키 생성 알고리즘

- 결합키 생성시에는 결합신청자 간 결합키 생성 항목, 인코딩\* 방식, 알고리즘을 동일하게 사용하여 함

\* 한글 인코딩 방식(EUC-KR, UTF-8)이 다를 경우, 동일한 일방향 암호화 알고리즘으로 데이터를 암호화 하여도 서로 다른 값으로 결합키가 생성되어 결합이 되지 않음(UTF-8 인코딩을 권고)



- 결합신청자는 가명처리 대상 정보에 정보주체별로 중복되지 않는 일련의 값(일련번호\*)을 생성함

\* 일련번호는 모의결합 시에는 활용되지 않으므로, 모의결합 절차가 종료된 이후 생성할 수 있음

결합신청자 A					결합신청자 B				
일련번호	성명	전화번호	생년월일	...	일련번호	성명	전화번호	생년월일	...
A1	강감찬	090-4562-7895	1947	...	B1	유관순	090-4567-9876	1982	...
A2	권율	090-7854-5689	1975	...	B2	권율	090-7854-5689	1975	...
A3	유관순	090-4567-9876	1982	...	B3	강감찬	090-4562-7895	1947	...
...	...	...	...	...	...	...	...	...	...

- 반복결합의 경우, 결합신청자는 추후 반출되는 정보와의 연계·분석을 위하여 결합키에 사용된 결합키 생성 항목, 인코딩 방식, 알고리즘(Salt값 제외\*)을 보관

\* 반복결합에 사용된 Salt값은 결합키관리기관이 보관하였다가, 추후 반복결합 진행시 재활용

## 2. 모의결합

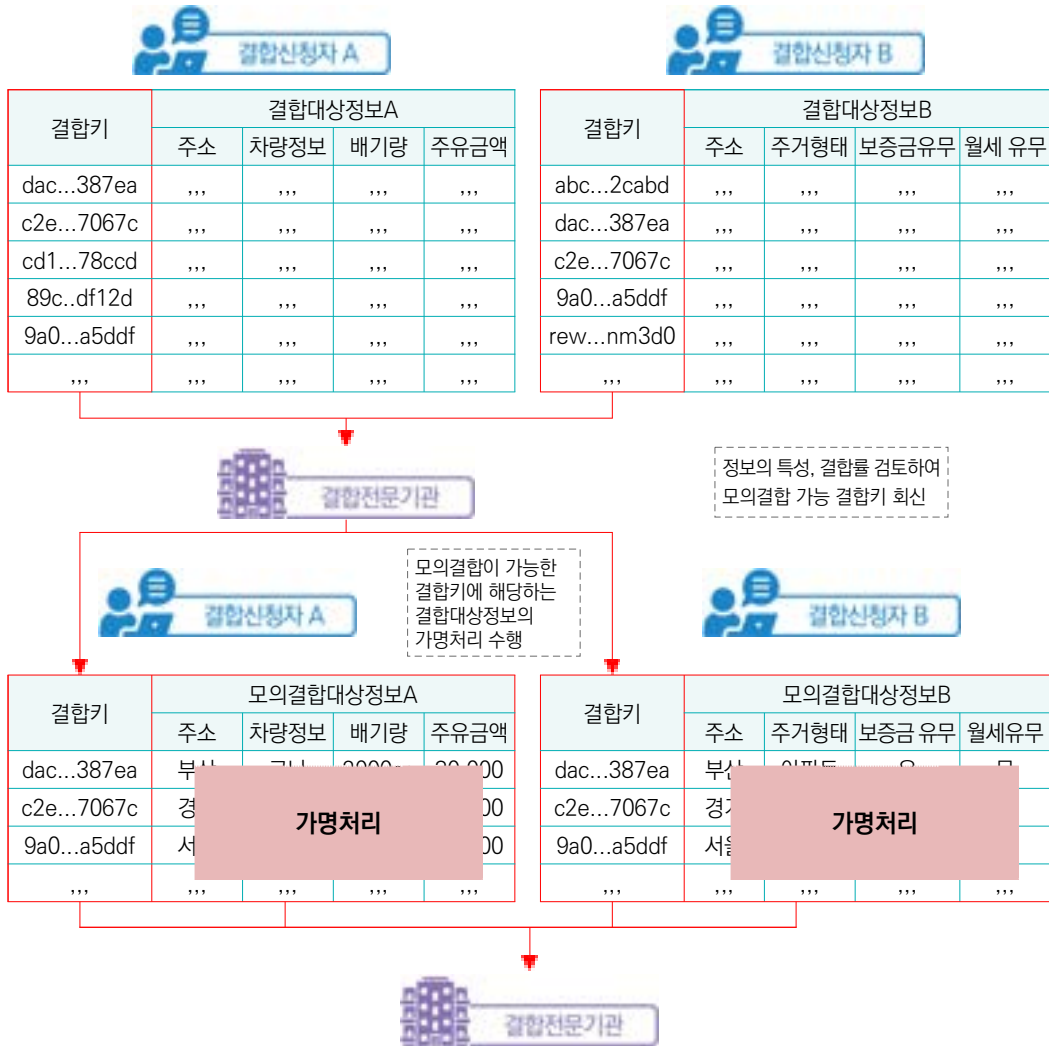
- ◆ 가명정보를 제공하는 결합신청자는 결합키관리기관 또는 모의결합을 수행할 결합전문기관과의 협의에 따라 생성한 모의결합키를 해당 결합전문기관에 전송

※ 결합신청자가 모의결합을 위한 결합키 생성을 위해 결합키관리기관과 협의하는 것은 선택사항

- ◆ 결합전문기관은 모의결합대상정보의 특성, 결합률 등을 고려하여 모의결합 가능여부를 판단하고, 모의결합이 가능한 경우 모의결합대상정보(결합키)를 선정하여 해당 결합키를 결합신청자에게 전송

※ 결합전문기관은 개인정보 침해의 우려가 없는 범위에서 결합신청자가 결합의 유용성을 확인 할 수 있도록 모의결합 대상 결합키를 선정

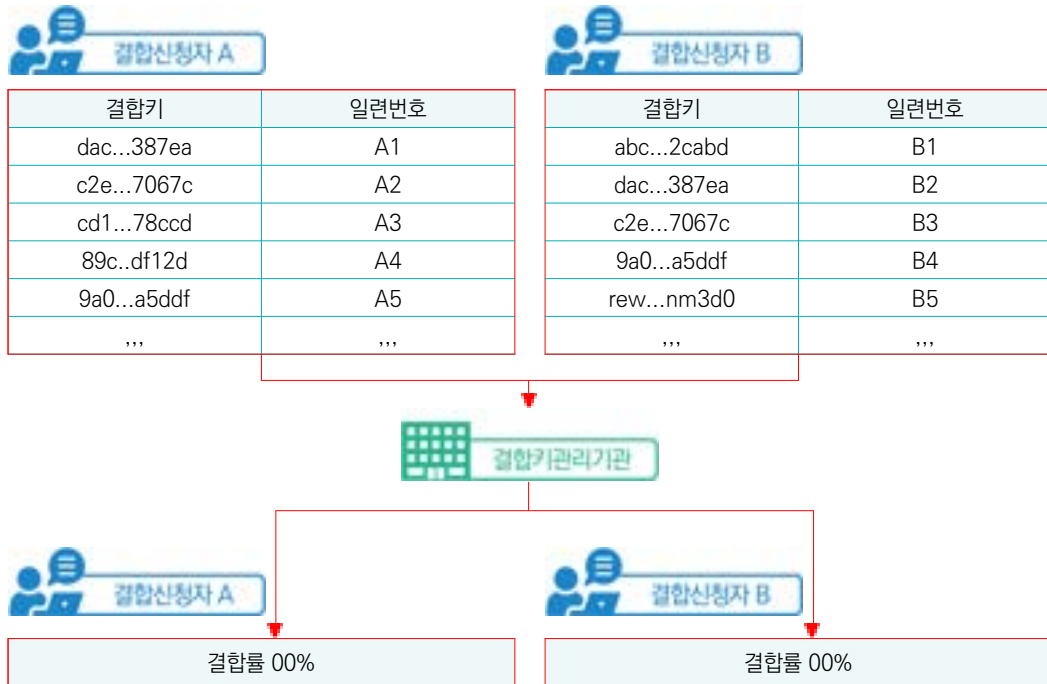
- ◆ 결합신청자는 결합전문기관으로부터 결합키를 제공받아 해당 모의결합대상정보를 가명처리하고 가명처리 내역과 함께 결합전문기관에 전송



- ◆ 결합전문기관은 결합신청자의 가명처리 내역을 확인하고(보완 필요시 보완 요청) 결합키를 사용하여 모의결합대상정보의 결합을 수행함
- ◆ 결합신청자는 결합전문기관(추가 가명처리 공간 등)에서 모의결합된 정보(이하 모의결합정보)를 분석할 수 있음
  - 결합신청자는 모의결합 분석 결과에 따라 본결합의 진행 또는 종료를 결정할 수 있음
  - ※ 단, 결합신청자는 분석한 결과물 및 모의결합 정보를 반출할 수 없음
- ◆ 결합전문기관은 결합신청자의 모의결합정보 분석이 완료되면 모의결합에 사용된 정보를 파기하여야 함

### 3. 결합률 확인

- ◆ 결합률 확인을 신청한 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함



- ◆ 결합키관리기관은 결합률을 측정하며 해당 결합신청자에게 해당 정보의 결합률을 통지함
  - 결합신청자는 결합률 확인 후 결합의 진행 또는 종료를 결정할 수 있음

### 4. 가명정보 추출

- ◆ 가명정보 추출을 신청한 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함
- ◆ 결합키관리기관은 추출 여부를 판단하는데 필요한 정보(결합 목적 등)를 결합신청자로부터 제공받아 추출 가능 여부를 검토하고, 추출이 가능한 경우 추출에 필요한 일련번호를 결합신청자에게 전송함



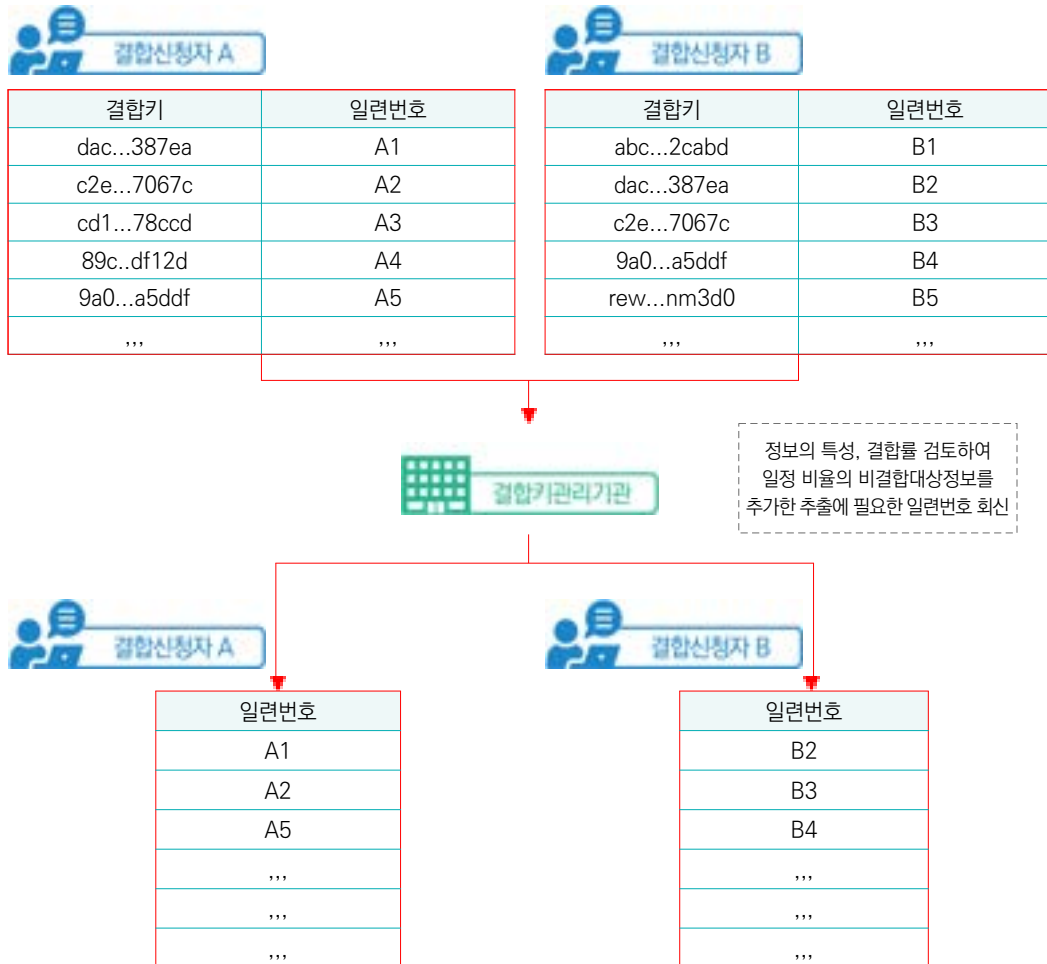
전체 정보(A)	결합되는 정보(B)	비결합대상정보(C) (결합되지 않는 정보 일부)	추출 대상 정보(D=B+C) (일련번호 수)
1,000,000	100,000	100,000	200,000

전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,000	2,000	4,000	1,000



전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,000	1,600	3,600	1,400

전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,700	2,300	(추출 불가)	-

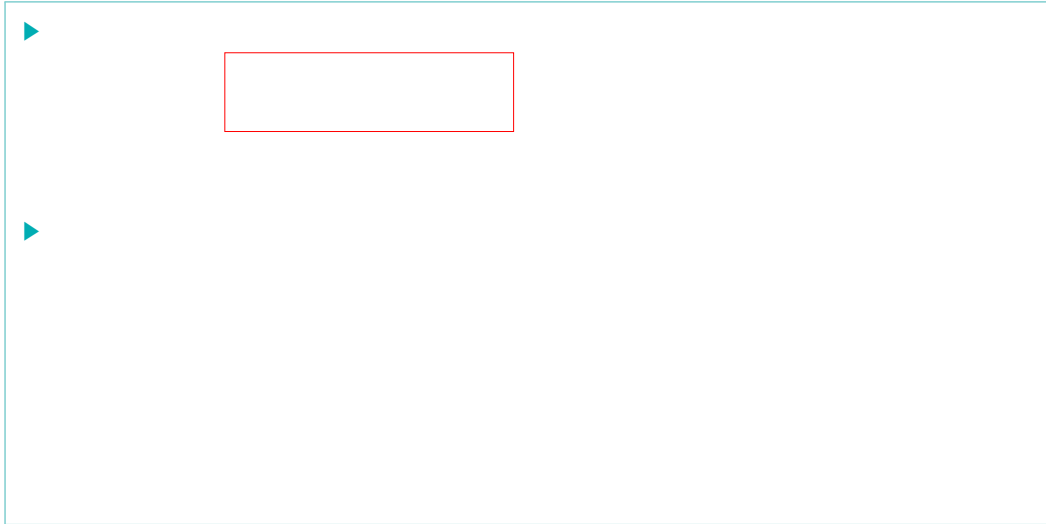


## 5. 가명처리 및 검토

◆ 결합을 진행하기로 결정한 결합신청자는 가명처리 대상 정보를 가명처리하여 결합대상정보를 결합 전문기관에 전송함

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 가명처리 지원을 요청할 수 있음

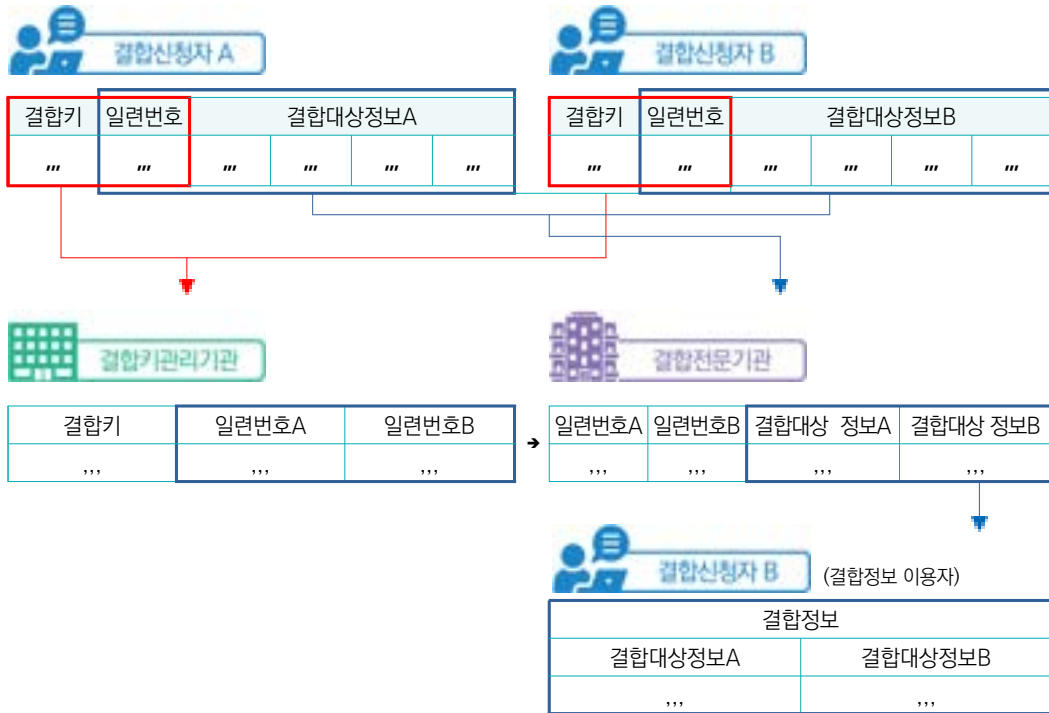
- 가명정보 추출을 신청한 결합신청자는 결합관리기관이 제공한 추출에 필요한 일련번호를 확인하고, 해당 일련번호의 결합대상정보의 가명처리 내역(결합대상정보, 가명처리 내역, 일련번호)을 결합 전문기관에 전송함



- ◆ 결합전문기관은 결합신청자가 제출한 결합대상정보 및 가명처리 내역을 검토함
  - 보완이 필요한 경우 결합전문기관은 결합신청자에게 보완 사항을 적시하여 보완 요청함

## 6. 결합

- ◆ 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함
  - ※ 결합을 확인 또는 가명정보 추출을 위해 결합키와 일련번호를 기 전송하고 결합대상정보의 변경이 없는 경우 전송을 생략할 수 있음
- ◆ 결합키관리기관은 결합키와 일련번호를 사용하여 결합키연계정보를 생성하고 결합전문기관에 결합키연계정보를 전송함
  - ※ 반복결합의 경우, 반복결합연계정보를 포함하여 결합키연계정보를 생성함
- ◆ 결합전문기관은 결합키연계정보와 일련번호, 결합대상정보를 사용하여 결합함



※ 필요시 추가처리 및 분석 수행, 반출심사 후 반출 가능

## 7. 추가처리 및 분석

- ◆ 결합정보를 이용하는 결합신청자는 결합전문기관(추가 가명처리 등 공간)에서 결합정보의 식별 위험성(가능성)을 확인하고, 보완이 필요한 경우 해당 부분에 대한 추가 가명처리를 수행함

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)으로부터 추가 가명처리에 대한 자문 및 지원을 받을 수 있음

- 결합신청자는 결합정보의 재식별 가능성이 없거나 추가 가명처리가 필요하지 않다고 판단하는 경우 추가 가명처리 없이 반출신청이 가능함

- ◆ 결합신청자는 결합전문기관에 마련된 분석에 필요한 시설, 장비를 갖춘 공간(추가 가명처리 등 공간)에서 결합정보를 분석할 수 있음

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 결합정보의 분석을 지원해줄 것을 요청할 수 있음

### ③ 반출 및 활용

#### 1. 반출신청 및 심사

- ◆ 결합정보를 반출하려는 결합신청자는 결합 고시 [별지 제4호] 반출신청서와 첨부 서류\*를 제출하며 반출을 신청함

\* 추가적인 서류 제출이 필요한 경우에 한하여 추가 처리 내역, 반출정보를 증명할 수 있는 서류, 반출정보에 대한 안전조치 계획을 제출 ※ 세부 작성은 제6장 결합정보 반출신청서 작성방법 참고

- ◆ 반출신청을 받은 결합전문기관은 반출신청서 및 첨부 서류를 확인하고 접수함
  - 보완이 필요한 경우 결합전문기관은 해당 사유를 적시하여 결합신청자에게 보완 요청하고 보완 사항을 확인한 이후 접수함
- ◆ 결합신청자가 반출을 요청하면 결합전문기관은 접수일로부터 영업일 기준 5일 이내 반출심사위원회 구성 등에 관한 사항을 결합신청자에게 통지함
  - 결합신청자는 결합전문기관으로부터 회의개최 일정 및 장소, 반출가능 예정 시기 등이 포함된 계획서를 받을 수 있음
  - 시계열 분석 등 반복결합의 반출심사는 2회부터 최초(첫회) 반출과 결합대상, 가명처리 방법 등이 거의 동일한 경우, 서면회의 등으로 간소화할 수 있음
  - 반출심사위원회는 3명의 위원으로 구성함. 단, 반출심사를 위해 필요한 경우 다른 결합전문기관에 소속된 전문가를 추가로 포함하여 구성할 수 있음
  - 반출심사위원은 개인정보 보호와 관련한 업무 경력이 있거나 관련 단체로부터 추천을 받은 사람, 개인정보처리자로 구성된 단체에서 활동한 경력이 있거나 관련 단체로부터 추천을 받은 사람, 그 밖에 개인정보 보호와 관련한 경력과 전문성이 있는 사람이어야 함(결합 고시 제11조제2항)
  - 반출심사위원회는 결합 목적과 반출정보의 관련성, 특정 개인의 식별가능성, 반출정보에 대한 안전 조치 계획 등을 심사하여야 함(보호법 시행령 제29조의3제4항, 결합 고시 제11조제3항)
- ◆ 결합신청자는 반출심사위원회의 요청에 따라 추가 서류를 제출하거나 직접 출석하여 설명할 수 있음

## 2. 반출

- ◆ 결합전문기관이 반출을 승인하면 결합신청자는 결합정보를 분석한 결과물을 반출하거나, 결합정보(데이터셋)를 반출할 수 있음

## 3. 활용

- ◆ 반출정보는 결합신청자가 반출심사 시 제출한 환경(가명정보 활용 형태, 처리 장소, 방법)과 목적범위에서 활용하는 것이 원칙임
  - ※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 반출정보에 대한 분석 지원을 요청할 수 있음
- 결합신청자가 반출정보를 반출심사 시와 다른 목적으로 활용하거나 제3자에게 제공하는 것이 금지되어 있지는 않으나(보호법 제28조의2 제1항), 반출심사 시 제출한 처리 상황의 변경이 있는 경우 해당 처리 상황에 맞게 가명처리하여 활용하여야 함
- ◆ 반복결합의 반출정보에는 반복결합연결정보가 포함되어 내부에서 연계하여 분석할 수 있음

## ④ 안전한 관리

### 1. 안전한 관리

- ◆ 결합신청자는 반출정보를 특정 개인을 알아보기 위한 목적으로 처리하여서는 아니 되며(보호법 제28조의5 제1항), 재식별되지 않도록 지속적으로 모니터링하여야 함
- ◆ 반출정보를 활용하는 결합신청자는 안전성 확보에 필요한 기술적·관리적·물리적 조치를 수행하여야 함
  - ※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 개인정보 보호 교육에 관한 지원을 요청할 수 있음

### 2. 결합전문기관 업무지원 사항

- ◆ **(모의결합)** 결합 전 모의결합 절차를 수행\* 할 수 있음
  - \* 모의결합 가능성 검토 및 통지, 모의결합대상정보 가명처리 수준 검토(필요시 추가처리 요청), 모의결합 수행
- ◆ **(결합 전 처리)** 결합 전 결합대상정보의 가명처리를 지원할 수 있음
- ◆ **(반출 전 처리)** 반출 전 결합정보의 추가 가명처리를 지원할 수 있음
- ◆ **(분석)** 반출 전 결합정보의 분석 및 반출 후 반출정보의 분석을 지원할 수 있음
- ◆ **(교육)** 가명정보를 반출하려는 결합신청자에 대한 개인정보 보호 교육\*을 지원할 수 있음
  - \* 반출정보의 안전조치에 관한 교육, 가명처리 지원 제도 안내 등

# 3

◆ 가명정보 결합은 결합신청자 간의 **공통되는 결합키**를 기반으로 이루어짐

◆ 결합신청자가 결합 후 **반출받아 활용할 수 있는 정보(반출정보)**에는






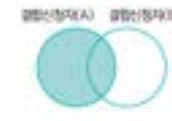



- 공통된 결합키를 기반으로 **결합된 정보(공통결합)**뿐만 아니라,
- 공통된 결합키를 통한 **결합을 통해 알게되는 그 외 정보\***도 포함될 수 있음

\* <확대단일결합> 반출정보 = 공통되는 정보(A∩B) + 결합되지 않은 정보(A-B)

<잔여결합> 반출정보 = 결합되지 않은 정보(A-B)

※ 예) 결합신청자(A)는 결합의 결과물로서 **공통된 정보(A∩B)** 외에도, **자신의 정보(A) 중 상대의 정보(B)와 결합되지 않은 정보(A-B)** 또는 **자신과 결합되지 않은 상대의 정보(B-A)**를 함께 반출받을 수 있음



공통결합 (INNER JOIN)			확대결합 (OUTER JOIN)	잔여결합 (ANTI-INNER JOIN)
단일 (INNER SINGLE)	다중 (INNER MULTI)	완전 (INNER FULL)	단일 (OUTER SINGLE)	단일 (ANTI-INNER SINGLE)
 	 		 	 

### ◆ 확대단일결합의 유용성

- “확대단일결합”이란 **공통된 결합키를 기반으로 결합된 정보와 결합되지 않은 정보까지 함께 제공** 받는 방식의 결합을 의미함
  - 두 기관이 가명정보를 결합하는 경우 반출정보에는, ‘**결합된 정보**’(양 기관 모두에 존재하는 결합대상 정보)와 ‘**내 기관에만 존재하는 정보**’ 또는 ‘**상대 기관에만 존재하는 정보**’도 함께 포함될 수 있음
- 통계작성, 과학적 연구에서는 **특정 집단과 비교 집단을 함께 분석**해야하는 **대조군 연구**가 필요한 경우가 많은데,
  - 이 경우, **확대단일결합을 활용**하면 분석에 필요한 **일반적인 비교 집단을 확보**할 수 있어 **연구의 통계적 타당성을 높일 수 있음**

#### 예시



⇒ 대조군을 통한 비교연구 등이 필요한 경우 **결합신청자는 결합전문기관과 협의하여 확대단일결합 추진 가능**






※ **결합되지 않은 다른 기관의 정보를 함께 제공**받는 확대단일결합의 경우에도 **결합전문기관이 △확대단일결합의 필요성, △정보의 제공 범위의 적절성, △재식별 위험성 등을 종합 고려하여 적절하다고 판단하는 경우 수행 가능**









### ◆ 확대단일결합 시 주의사항

- 확대단일결합은 결합되지 않은 다른 기관의 정보를 함께 제공받는 경우 다음과 같은 위험이 존재함



- 따라서 **결합전문기관은 확대단일결합 신청 건에 대한 반출심사시 다음 사항을 면밀히 검토할 필요가 있음**
  - 반출 신청한 정보가 결합 목적에 반드시 필요한지?
  - 결합되지 않은 상대 기관의 정보의 제공 범위가 과도하지 않은지?
  - 결합·반출정보가 축적·재식별되지 않고 안전하게 활용될 수 있는지?

유형	설명																																																																								
<p style="writing-mode: vertical-rl; text-orientation: upright;">공통 단일 결합 (INNER SINGLE JOIN)</p>  <p>결합신용재(A)    결합신용재(B)</p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2">결합신용재A</th> <th colspan="2">결합신용재B</th> <th colspan="2">결합신용재</th> </tr> <tr> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> </tr> </thead> <tbody> <tr> <td>dec_387ea</td> <td>A1</td> <td>abc_20abd</td> <td>B1</td> <td>A1</td> <td>B2</td> </tr> <tr> <td>c2e_7067c</td> <td>A2</td> <td>dec_387ea</td> <td>B2</td> <td>A2</td> <td>B3</td> </tr> <tr> <td>eff_78bcd</td> <td>A3</td> <td>c2e_7067c</td> <td>B3</td> <td>A3</td> <td>B4</td> </tr> <tr> <td>flk_dp12d</td> <td>A4</td> <td>flk_dp12d</td> <td>B4</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp5ad</td> <td>A5</td> <td>res_nm3bd</td> <td>B5</td> <td>...</td> <td>...</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	결합신용재A		결합신용재B		결합신용재		결합키	결합번호	결합키	결합번호	결합번호	결합번호	dec_387ea	A1	abc_20abd	B1	A1	B2	c2e_7067c	A2	dec_387ea	B2	A2	B3	eff_78bcd	A3	c2e_7067c	B3	A3	B4	flk_dp12d	A4	flk_dp12d	B4	...	...	flk_dp5ad	A5	res_nm3bd	B5	...	...	...	...	...	...	...	...																								
결합신용재A		결합신용재B		결합신용재																																																																					
결합키	결합번호	결합키	결합번호	결합번호	결합번호																																																																				
dec_387ea	A1	abc_20abd	B1	A1	B2																																																																				
c2e_7067c	A2	dec_387ea	B2	A2	B3																																																																				
eff_78bcd	A3	c2e_7067c	B3	A3	B4																																																																				
flk_dp12d	A4	flk_dp12d	B4	...	...																																																																				
flk_dp5ad	A5	res_nm3bd	B5	...	...																																																																				
...	...	...	...	...	...																																																																				
 <p>결합신용재(A) 결합신용재(B)    결합신용재(C)</p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2">결합신용재A</th> <th colspan="2">결합신용재B</th> <th colspan="2">결합신용재C</th> <th colspan="2">결합신용재</th> </tr> <tr> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> </tr> </thead> <tbody> <tr> <td>dec_387ea</td> <td>A1</td> <td>abc_20abd</td> <td>B1</td> <td>def_ave12</td> <td>C1</td> <td>A1</td> <td>B2</td> <td>C2</td> </tr> <tr> <td>c2e_7067c</td> <td>A2</td> <td>dec_387ea</td> <td>B2</td> <td>dec_387ea</td> <td>C2</td> <td>A2</td> <td>B3</td> <td>C3</td> </tr> <tr> <td>eff_78bcd</td> <td>A3</td> <td>c2e_7067c</td> <td>B3</td> <td>c2e_7067c</td> <td>C3</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp12d</td> <td>A4</td> <td>flk_dp5ad</td> <td>B4</td> <td>flk_dp12d</td> <td>C4</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp5ad</td> <td>A5</td> <td>res_nm3bd</td> <td>B5</td> <td>res_nm3bd</td> <td>C5</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	결합신용재A		결합신용재B		결합신용재C		결합신용재		결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2	c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3	eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...	flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	...	...	...	flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	...	...	...	...	...	...	...	...	...	...	...	...		
결합신용재A		결합신용재B		결합신용재C		결합신용재																																																																			
결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호																																																																		
dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2																																																																	
c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3																																																																	
eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...																																																																	
flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	...	...	...																																																																	
flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	...	...	...																																																																	
...	...	...	...	...	...	...	...	...																																																																	
<p style="writing-mode: vertical-rl; text-orientation: upright;">공통 다중 결합 (INNER MULTI JOIN)</p>  <p>결합신용재(A) 결합신용재(B)    결합신용재(C)</p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2">결합신용재A</th> <th colspan="2">결합신용재B</th> <th colspan="2">결합신용재C</th> <th colspan="3">결합신용재</th> </tr> <tr> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> </tr> </thead> <tbody> <tr> <td>dec_387ea</td> <td>A1</td> <td>abc_20abd</td> <td>B1</td> <td>def_ave12</td> <td>C1</td> <td>A1</td> <td>B2</td> <td>C2</td> </tr> <tr> <td>c2e_7067c</td> <td>A2</td> <td>dec_387ea</td> <td>B2</td> <td>dec_387ea</td> <td>C2</td> <td>A2</td> <td>B3</td> <td>C3</td> </tr> <tr> <td>eff_78bcd</td> <td>A3</td> <td>c2e_7067c</td> <td>B3</td> <td>c2e_7067c</td> <td>C3</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp12d</td> <td>A4</td> <td>flk_dp5ad</td> <td>B4</td> <td>flk_dp12d</td> <td>C4</td> <td>A5</td> <td>B4</td> <td>...</td> </tr> <tr> <td>flk_dp5ad</td> <td>A5</td> <td>res_nm3bd</td> <td>B5</td> <td>res_nm3bd</td> <td>C5</td> <td>A4</td> <td>...</td> <td>C4</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	결합신용재A		결합신용재B		결합신용재C		결합신용재			결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호	dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2	c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3	eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...	flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...	flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4	...	...	...	...	...	...	...	...	...
결합신용재A		결합신용재B		결합신용재C		결합신용재																																																																			
결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호																																																																	
dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2																																																																	
c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3																																																																	
eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...																																																																	
flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...																																																																	
flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4																																																																	
...	...	...	...	...	...	...	...	...																																																																	
 <p>결합신용재(A) 결합신용재(B)    결합신용재(C)</p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2">결합신용재A</th> <th colspan="2">결합신용재B</th> <th colspan="2">결합신용재C</th> <th colspan="3">결합신용재</th> </tr> <tr> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> </tr> </thead> <tbody> <tr> <td>dec_387ea</td> <td>A1</td> <td>abc_20abd</td> <td>B1</td> <td>def_ave12</td> <td>C1</td> <td>A1</td> <td>B2</td> <td>C2</td> </tr> <tr> <td>c2e_7067c</td> <td>A2</td> <td>dec_387ea</td> <td>B2</td> <td>dec_387ea</td> <td>C2</td> <td>A2</td> <td>B3</td> <td>C3</td> </tr> <tr> <td>eff_78bcd</td> <td>A3</td> <td>c2e_7067c</td> <td>B3</td> <td>c2e_7067c</td> <td>C3</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp12d</td> <td>A4</td> <td>flk_dp5ad</td> <td>B4</td> <td>flk_dp12d</td> <td>C4</td> <td>A5</td> <td>B4</td> <td>...</td> </tr> <tr> <td>flk_dp5ad</td> <td>A5</td> <td>res_nm3bd</td> <td>B5</td> <td>res_nm3bd</td> <td>C5</td> <td>A4</td> <td>...</td> <td>C4</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	결합신용재A		결합신용재B		결합신용재C		결합신용재			결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호	dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2	c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3	eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...	flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...	flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4	...	...	...	...	...	...	...	...	...
결합신용재A		결합신용재B		결합신용재C		결합신용재																																																																			
결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호																																																																	
dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2																																																																	
c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3																																																																	
eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...																																																																	
flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...																																																																	
flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4																																																																	
...	...	...	...	...	...	...	...	...																																																																	
<p style="writing-mode: vertical-rl; text-orientation: upright;">공통 완전 결합 (INNER FULL JOIN)</p>  <p>결합신용재(A) 결합신용재(B)    결합신용재(C)</p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2">결합신용재A</th> <th colspan="2">결합신용재B</th> <th colspan="2">결합신용재C</th> <th colspan="3">결합신용재</th> </tr> <tr> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합키</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> <th>결합번호</th> </tr> </thead> <tbody> <tr> <td>dec_387ea</td> <td>A1</td> <td>abc_20abd</td> <td>B1</td> <td>def_ave12</td> <td>C1</td> <td>A1</td> <td>B2</td> <td>C2</td> </tr> <tr> <td>c2e_7067c</td> <td>A2</td> <td>dec_387ea</td> <td>B2</td> <td>dec_387ea</td> <td>C2</td> <td>A2</td> <td>B3</td> <td>C3</td> </tr> <tr> <td>eff_78bcd</td> <td>A3</td> <td>c2e_7067c</td> <td>B3</td> <td>c2e_7067c</td> <td>C3</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>flk_dp12d</td> <td>A4</td> <td>flk_dp5ad</td> <td>B4</td> <td>flk_dp12d</td> <td>C4</td> <td>A5</td> <td>B4</td> <td>...</td> </tr> <tr> <td>flk_dp5ad</td> <td>A5</td> <td>res_nm3bd</td> <td>B5</td> <td>res_nm3bd</td> <td>C5</td> <td>A4</td> <td>...</td> <td>C4</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	결합신용재A		결합신용재B		결합신용재C		결합신용재			결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호	dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2	c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3	eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...	flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...	flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4	...	...	...	...	...	...	...	...	...
결합신용재A		결합신용재B		결합신용재C		결합신용재																																																																			
결합키	결합번호	결합키	결합번호	결합키	결합번호	결합번호	결합번호	결합번호																																																																	
dec_387ea	A1	abc_20abd	B1	def_ave12	C1	A1	B2	C2																																																																	
c2e_7067c	A2	dec_387ea	B2	dec_387ea	C2	A2	B3	C3																																																																	
eff_78bcd	A3	c2e_7067c	B3	c2e_7067c	C3	...	...	...																																																																	
flk_dp12d	A4	flk_dp5ad	B4	flk_dp12d	C4	A5	B4	...																																																																	
flk_dp5ad	A5	res_nm3bd	B5	res_nm3bd	C5	A4	...	C4																																																																	
...	...	...	...	...	...	...	...	...																																																																	

유형	설명
<p>확대 단일 결합 (OUTER SINGLE JOIN)</p> 	
<p>중첩 단일 결합 (INNER SINGLE JOIN)</p> 	
<p>잔여 단일 결합 (ANTI-INNER SINGLE JOIN)</p> 	
<p>중첩 단일 결합 (INNER SINGLE JOIN)</p> 	

다양한 가명정보 결합 응용 형태

(사례1) 인사이드 결합 ※결합기관리기관 협의

▶

인사이드 결합



결합기관	결합신청번호(A)
ABCFA	A1
SDVWZ	A2
DRFZB	A3
19QDN	A4
QWFZJ	A5
ZDWE	A6
ZDFWC	A7
LRWUV	A8
SDVWZ	A9
BDPQJ	A10
...	...
...	...

결합기관	결합신청번호(C,D)
ABCFA	B1
ZDFWB	B2
GDWE	B3
YBGA	B4
DRFWR	C1
Q1TDF	C2
SDVWZ	C3
ZDFWC	C4
BQDN	D1
DRFZB	D2
BQDN	D3
LRWUV	D4
...	...

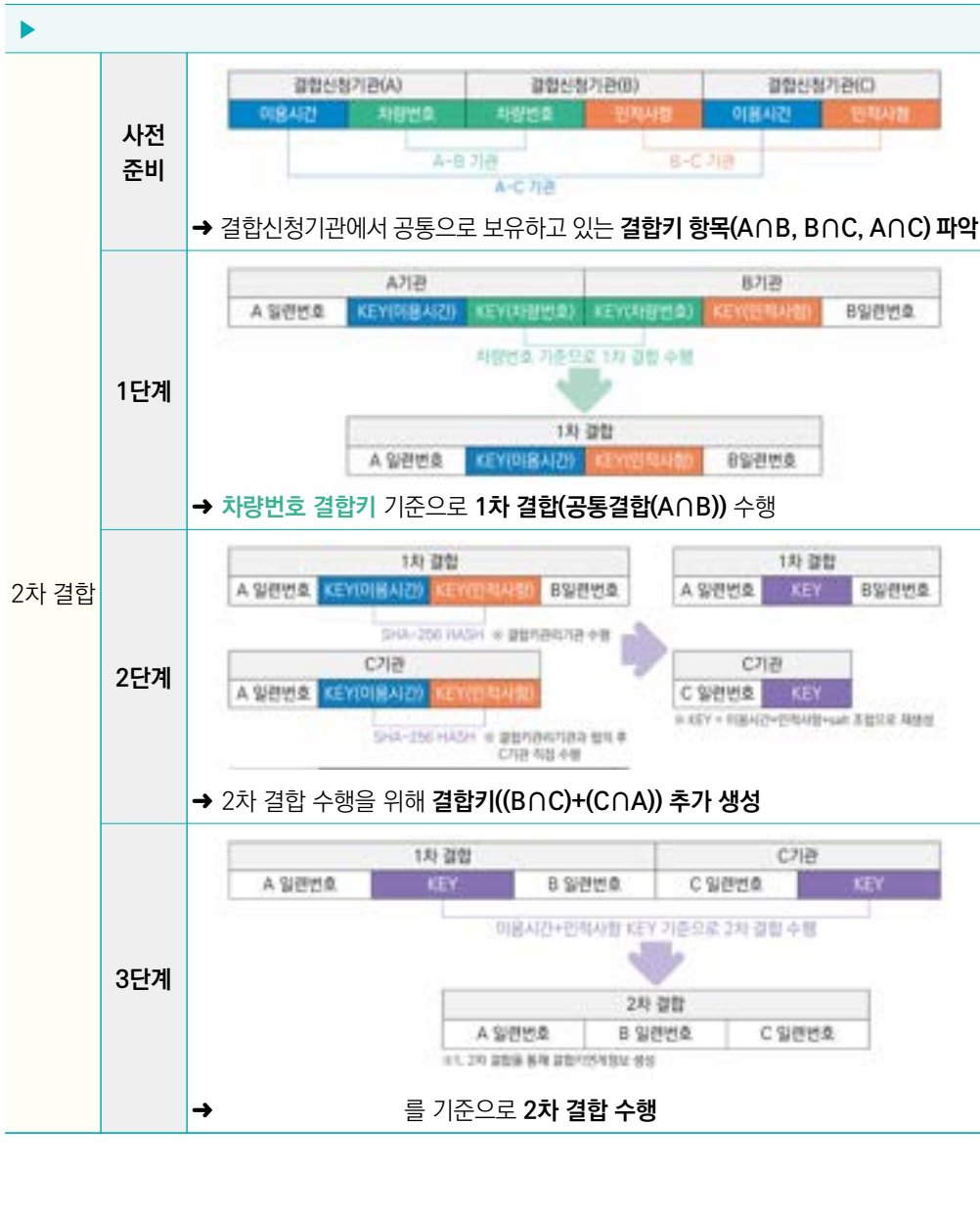
결합번호	
B1	A1
B2	A5
C3	A2
C4	A7
D2	A3
D4	A8
...	...
...	...
...	...
...	...
...	...
...	...
...	...

→ 결합신청기관(B), 결합신청기관(C), 결합신청기관(D)는 서로 결합되는 항목이 존재하더라도, 분석 시 무의미하므로 결합신청기관(A)를 기준으로 공통단일결합 수행

예시

▶

**(사례2) 2차 결합** ※결합기관리기관 협의



예시

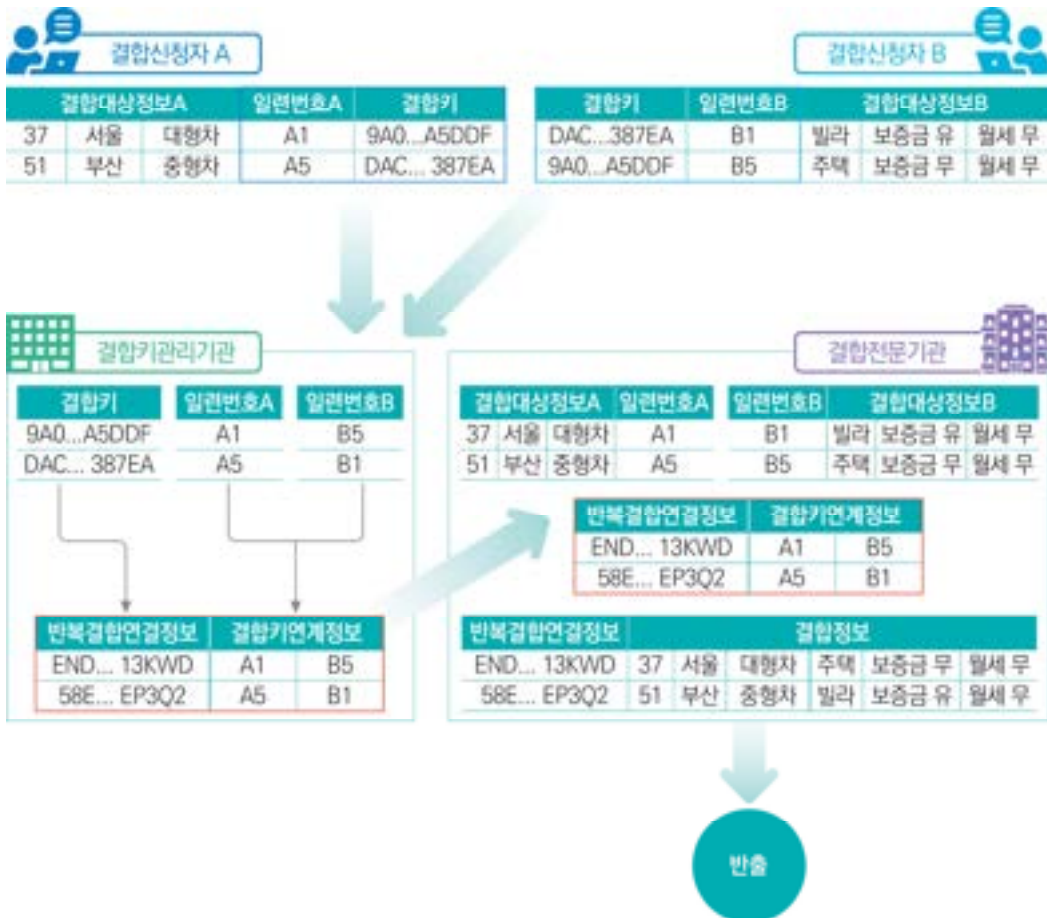


## 4

시계열 분석 등을 목적으로 가명정보를 결합할 때에는 동일한 서로 다른 개인정보처리자 간의 가명정보를 지속적·반복적으로 반복하며 결합할 수 있음

- 반복결합이 필요한 경우, 결합신청 시 반복결합을 선택하여 신청함

※ 반복결합의 경우, 반출정보에 반복적인 분석을 위해 필요한 정보(반복결합연결정보)가 추가 포함됨



- ① 결합신청자는 가명처리 대상 정보에 정보주체별로 일련번호를 생성하고, 결합키관리기관과 협의한 방법에 따라 결합키를 생성
  - 생성된 결합키와 일련번호는 결합키관리기관으로 송신
- ② 결합키관리기관은 결합신청자로부터 수신받은 결합키를 활용하여 1) 시계열 분석에 필요한 키 (반복결합키)와 2) 결합키연계정보를 생성
  - ※ 일반적인 결합의 경우, 추가적인 시계열 분석에 필요한 키를 생성하지 않음
- ③ 결합키관리기관에서 생성한 1), 2)의 정보를 결합전문기관에 송신하고, 결합전문기관은 수신받은 2)를 활용하여 결합신청자의 가명정보를 결합
- ④ 결합전문기관은 결합정보에 1)을 포함하여 반출하고, 결합신청자는 반출정보에 대한 안전조치 의무 수행
  - ※ 유의사항 : 반복결합 신청자는 추후 반출되는 정보와의 연계·분석을 위하여 결합키에 사용된 결합키 생성항목, 인코딩 방식, 알고리즘(Salt값 제외)을 보관하여야 함



※ 시계열 분석이 완전히 종료된 경우 이를 결합키관리기관에 통지하여야 하며, 결합키관리기관은 해당 결합 후 보관하고 있는 1)의 생성방법을 삭제



# 3

## 안전성 확보조치 이행 가이드

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제3장 안전성 확보조치 이행 가이드



### 1

◆ 개인정보처리자는 추가정보를 가명정보와 분리하여 별도로 저장·관리하고, 추가정보가 가명정보와 불법적으로 결합되어 재식별에 악용되지 않도록 접근권한을 최소화하고 접근통제를 강화하는 등 필요한 조치를 적용하여야 함

◆ 추가정보와 가명정보는 분리하여 보관하는 것을 원칙으로 하고, 불가피한 사유로 물리적인 분리가 어려운 경우 DB 테이블 분리 등 논리적으로 분리\*하는 것도 가능 함

\* 논리적으로 분리할 경우 엄격한 접근통제를 적용하여야 함

※ 추가정보의 활용 목적 달성 및 불필요한 경우에는 추가정보를 파기할 수 있으며, 이 경우 파기에 대한 기록을 작성하고 보관할 필요가 있음

### 2

◆ 개인정보처리자는 가명정보 또는 추가정보에 접근할 수 있는 담당자를 가명정보 처리 업무 목적달성에 필요한 최소한의 인원으로 엄격하게 통제하여야 하며, 접근권한도 업무에 따라 차등 부여 하여야 함

- 가명정보를 취급할 자를 추가로 둘 여력이 없는 경우 등 접근권한의 분리가 어려운 정당한 사유가 있는 경우\*에는 업무 수행에 필요한 최소한 접근권한 부여 및 접근권한의 보유 현황을 기록으로 보관하는 등 접근권한을 관리·통제하여야 함

\* 「소상공인 보호 및 지원에 관한 법률」 제2조에 따른 소상공인 등

- 가명정보를 처리하는 자가 가명처리를 수행하는 경우를 제외하고는 특정 개인을 알아볼 수 있는 개인정보처리시스템(가명정보처리시스템 제외)에 접근할 수 없도록 제한할 필요가 있음

◆ 가명처리 관련 업무 담당자의 분리

- 가명정보의 재식별 위험을 방지하기 위해서는 가명처리, 적정성 검토, 가명처리가 완료된 가명정보 처리를 수행하는 업무담당자를 각각 분리하고, 해당 업무별로 접근권한을 분리하여 운영하는 것이 안전함

※ 추가정보의 내용을 알고 있는 자가 가명처리의 적정성 검토를 수행하거나 가명정보를 처리(활용)하는 경우 특정 개인을 알아볼 우려가 있음

◆ 전보 또는 퇴직 등 인사이동이 발생하여 가명정보를 처리하는 자가 변경되었을 경우 지체 없이 가명정보 처리시스템의 접근권한을 변경 또는 말소하여야 함

- 가명정보처리시스템의 접근권한 부여, 변경 또는 말소에 대한 내역을 기록하고, 그 기록을 최소 3년간 보관하여야 함
- 가명정보처리시스템에 접속할 수 있는 사용자 계정을 발급하는 경우 가명정보를 처리하는 자 별로 사용자 계정을 발급하여야 하며, 다른 가명정보를 처리하는 자, 추가정보를 처리하는 자, 해당 가명정보 이외의 다른 개인정보취급자와 공유되지 않도록 하여야 함

◆ 가명정보를 처리하는 자가 안전한 비밀번호를 설정하여 이행할 수 있도록 비밀번호 작성규칙을 수립하여 적용하여야 함

- 가명정보에 대한 처리 권한이 있는 자만이 가명정보처리시스템에 접근할 수 있도록 계정정보 또는 비밀번호를 일정 횟수 이상 잘못 입력한 경우 접근을 제한하는 등 필요한 기술적 조치를 하여야 함

3

◆ 개인정보처리자는 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보의 이용내역, 제3자 제공 시 제공받는 자를 작성하여 보관하여야 함

- 관련 양식은 제5장 서식 ⑩ 「가명(추가)정보 관리대장」 활용

## 4

- 개인정보처리자는 가명정보 또는 추가정보의 안전한 관리를 위하여 물리적 안전조치를 취하여야 함
  - 개인정보처리자는 가명정보 또는 추가정보를 전산실이나 자료보관실에 보관하는 경우 비인가자의 접근으로부터 보호하기 위하여 출입 통제 등의 절차를 수립하여야 함
  - 또한 가명정보 또는 추가정보가 보조저장매체 등에 저장되어 있는 경우 잠금장치가 있는 안전한 장소에 보관하여야 하며, 이러한 보조저장매체 등의 반·출입 통제를 위한 보안대책을 마련하여야 함

# 4

## 가명처리 기술 및 기법

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제4장 가명처리 기술 및 기법

### 1

#### ① 개인정보의 가명처리 기술 종류

※ 아래 분류는 이해를 돕기 위해 ISO/IEC 20889, 그리고 EU ENISA에서 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)	• 특이정보를 해당 행 항목에서 삭제
개인정보 일부 또는 전부 대체	삭제기술	마스킹 (Masking)	• 특정 항목의 일부 또는 전부를 공백 또는 문자(' * ', ' _ ' 등)이나 전각 기호)로 대체
	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

\* EU ENISA(European Union Agency for Network and Information Security), Recommendations on shaping technology according to GDPR provisions, An overview on data pseudonymisation, November 2018 / EU ENISA(European Union Agency for Network and Information Security), Pseudonymisation and best practices, November 2019

분류	기술	세부기술	설명	
개인정보 일부 또는 전부 대체	일반화 (범주화) 기술	일반 라운딩 (Rounding)	<ul style="list-style-type: none"> <li>올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법으로, 일반적으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용</li> </ul>	
		랜덤 라운딩 (Random rounding)	<ul style="list-style-type: none"> <li>수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법</li> </ul>	
		제어 라운딩 (Controlled rounding)	<ul style="list-style-type: none"> <li>라운딩 적용 시 값의 변경에 따라 행이나 열의 합이 원본의 행이나 열의 합과 일치하지 않는 단점을 해결하기 위해 원본과 결과가 동일하도록 라운딩을 적용하는 기법</li> </ul>	
		상하단코딩 (Top and bottom coding)	<ul style="list-style-type: none"> <li>정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있음</li> <li>이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법</li> </ul>	
		로컬 일반화 (Local generalization)	<ul style="list-style-type: none"> <li>전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법</li> </ul>	
		범위 방법 (Data range)	<ul style="list-style-type: none"> <li>수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현</li> </ul>	
		문자데이터 범주화 (Categorization of character data)	<ul style="list-style-type: none"> <li>문자로 저장된 정보에 대해 보다 상위의 개념으로 범주화하는 기법</li> </ul>	
	암호화	암호화	양방향 암호화 (Two-way encryption)	<ul style="list-style-type: none"> <li>특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법</li> <li>암호화 및 복호화에 동일 비밀키로 암호화하는 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 비대칭키(Asymmetric key) 방식으로 구분</li> </ul>
			일방향 암호화 - 암호학적 해시함수 (One-way encryption - Cryptographic hash function)	<ul style="list-style-type: none"> <li>원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법</li> <li>키가 없는 해시함수(MDC, Message Digest Code), 솔트(Salt)가 있는 해시함수, 키가 있는 해시함수(MAC, Message Authentication Code)로 구분</li> <li>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</li> </ul>

분류	기술	세부기술	설명
개인정보 일부 또는 전부 대체	암호화	순서보존 암호화 (Order-preserving encryption)	<ul style="list-style-type: none"> <li>원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식</li> <li>암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능</li> </ul>
		형태보존 암호화 (Format-preserving encryption)	<ul style="list-style-type: none"> <li>원본 정보의 형태와 암호화된 값의 형태가 동일하게 유지되는 암호화 방식</li> <li>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</li> <li>암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용</li> </ul>
		동형 암호화 (Homomorphic encryption)	<ul style="list-style-type: none"> <li>암호화된 상태에서의 연산이 가능한 암호화 방식으로 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능</li> <li>암호화된 상태의 연산값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법</li> </ul>
		다형성 암호화 (Polymorphic encryption)	<ul style="list-style-type: none"> <li>가명정보의 부정합 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법</li> <li>정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨</li> </ul>
	무작위화 기술	잡음 추가 (Noise addition)	<ul style="list-style-type: none"> <li>개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법</li> </ul>
		순열(치환) (Permutation)	<ul style="list-style-type: none"> <li>분석 시 가치가 적고 식별성이 높은 열 항목에 대해 대상 열 항목의 모든 값을 열 항목 내에서 무작위로 순서를 변경하여 식별성을 낮추는 기법</li> <li>개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법</li> </ul>
		토큰화 (Tokenisation)	<ul style="list-style-type: none"> <li>개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 식별 위험을 제거하여 개인정보를 보호하는 기술</li> <li>토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 양방향 암호화, 형태보존 암호화 기법을 주로 사용</li> </ul>
		(의사)난수생성기 (P)RNG, (Pseudo) Random Number Generator)	<ul style="list-style-type: none"> <li>주어진 입력값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보와 대체</li> </ul>

분류	기술	세부기술	설명
가명·익명처리를 위한 다양한 기술 (기타 기술)		표본추출 (Sampling)	<ul style="list-style-type: none"> <li>데이터 주체별로 전체 모집단이 아닌 표본에 대해 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법</li> </ul>
		해부화 (Anatomization)	<ul style="list-style-type: none"> <li>기존 하나의 데이터셋(테이블)을 식별성이 있는 정보 집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술</li> </ul>
		합성데이터 (Synthetic data)	<ul style="list-style-type: none"> <li>원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법</li> </ul>
		동형비밀분산 (Homomorphic secret sharing)	<ul style="list-style-type: none"> <li>식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉼어(share)*로 대체 *기밀사항을 재구성하는데 사용할 수 있는 하위 집합</li> </ul>
		차분 프라이버시 (Differential privacy)	<ul style="list-style-type: none"> <li>특정 개인에 대한 사전지식이 있는 상태에서 데이터 베이스 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</li> <li>1개 항목이 차이나는 두 데이터베이스간의 차이(확률 분포)를 기준으로 하는 프라이버시 보호 모델</li> </ul>

## ② 개인정보의 가명처리 예시

※ 아래 모든 예시는 각 기법의 적용에 대한 예시이며 전체 데이터에 대한 가명처리에 대한 예시가 아닙니다.

### 1. 개인정보 삭제

- 삭제기술 : 선택된 항목을 제거하는 기술

수치형데이터

문자형데이터

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성별	나이	통신료	단말기금액	누적포인트
남	41세	98,700	1,198,700	356,800
여	61세	69,400	505,400	203,000
남	30세	104,400	1,604,400	198,000
여	57세	954,800	3,954,800	20,532,000
남	28세	83,600	883,600	400,900

수치형데이터    문자형데이터

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이운정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 여진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김	남	41세	8888	서울특별시 중구	98,700	1,198,700	356,800
이	여	61세	2222	부산광역시 북구	69,400	505,400	203,000
박	남	30세	7777	광주광역시 서구	104,400	1,604,400	198,000
이	여	57세	4444	전라남도 나주시	954,800	3,954,800	20,532,000
최	남	28세	6666	세종특별자치시	83,600	883,600	400,900

수치형데이터    문자형데이터

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이운정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 여진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
최동욱	남	28세	010-5555-6666	세종특별자치시 여진동	83,600	883,600	400,900

## 수치형데이터    문자형데이터

성명	성별	나이	핸드폰번호	주소	통신료	단말가금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이운정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 아진동	83,600	883,600	400,900

## 삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말가금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이운정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	
최동욱	남	28세	010-5555-6666	세종특별자치시 아진동	83,600	883,600	400,900

## 수치형데이터    문자형데이터

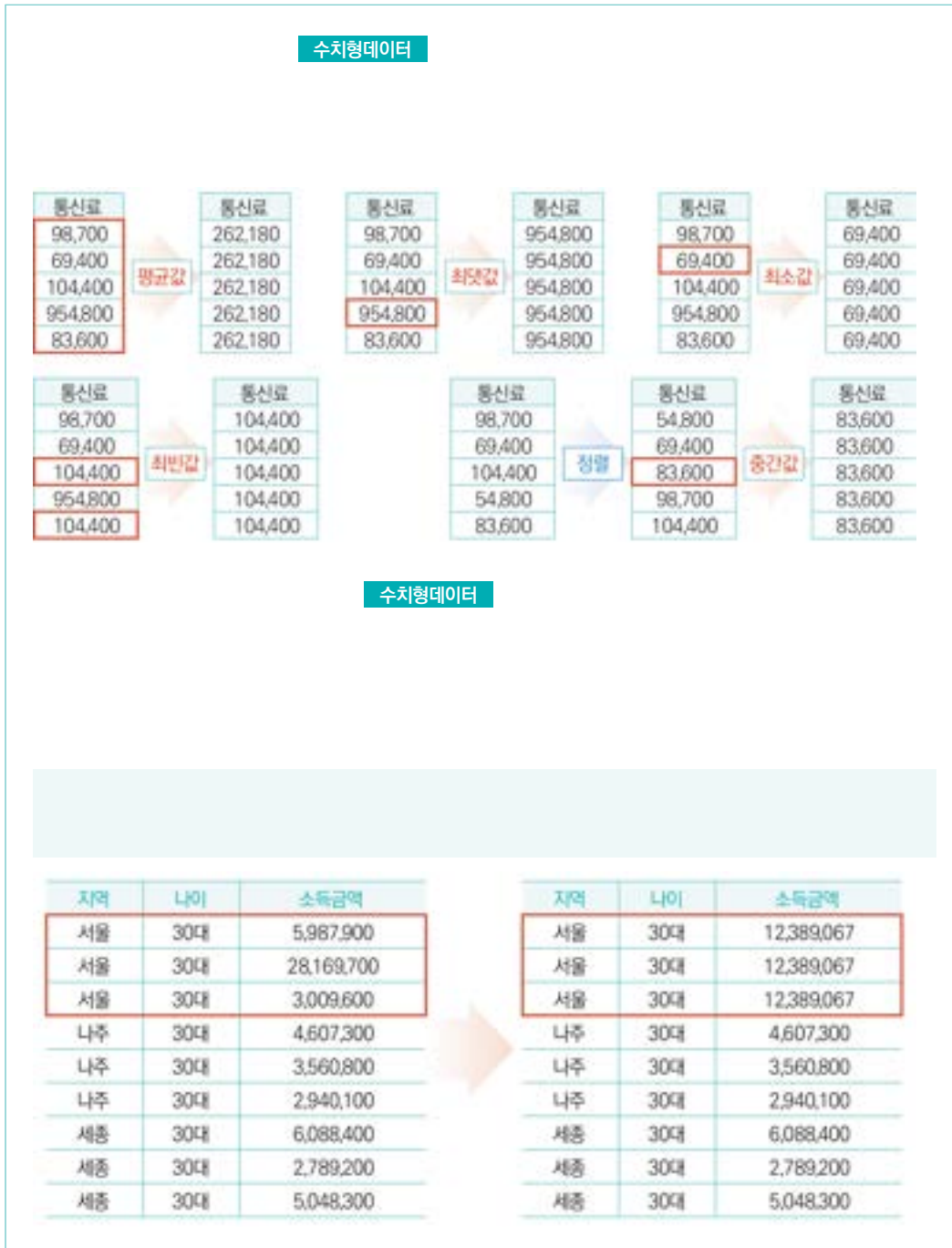
성명	성별	나이	핸드폰번호
김철수	남	41세	010-6666-8888
이영희	여	61세	010-9999-2222
박민호	남	30세	010-2222-7777
이운정	여	57세	010-3333-4444
최동욱	남	28세	010-5555-6666

마스킹

성명	성별	나이	핸드폰번호
김**	남	4*세	***-****-****
이**	여	6*세	***-****-****
박**	남	3*세	***-****-****
이**	여	5*세	***-****-****
최**	남	2*세	***-****-****

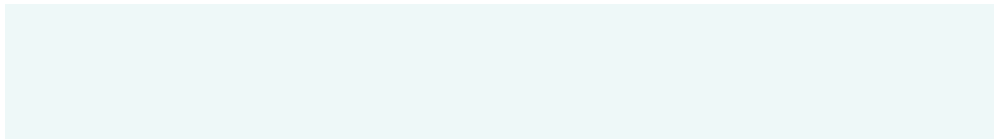
## 2. 개인정보 일부 또는 전부 대체

- 통계도구 : 데이터의 전체 구조를 변경하는 통계적 성질을 가진 기법



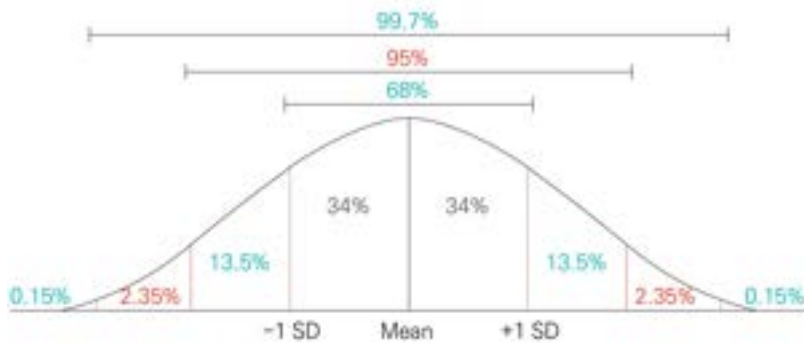


수치형데이터



원본(나이)	일반 라운딩	제어 라운딩
33세	30세	30세
61세	60세	60세
50세	50세	50세
72세	70세	70세
43세	40세	40세
44세	40세	50세
23세	20세	20세
67세	70세	70세
68세	70세	70세
49세	50세	50세
평균 : 51세	평균 : 50세	평균 : 51세
합계 : 510	합계 : 500	합계 : 510

수치형데이터



## 수치형데이터

지역	나이	소득금액
서울	30대	5,987,900
서울	30대	28,169,700
서울	30대	3,009,600
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300



지역	나이	소득금액
서울	30대	3,009,600-28,169,700
서울	30대	3,009,600-28,169,700
서울	30대	3,009,600-28,169,700
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

## 수치형데이터

~

## 문자형데이터

분류	분류
분류	육아용품
기저귀	육아용품
젖병	육아용품
시위타운	육실용품
육실화	육실용품

- 암호화 : 정보 가공 시 일정한 규칙의 알고리즘을 적용하여 대체

수치형데이터

문자형데이터

## - 무작위화기술 : 속성값을 수정하여 원래의 값과 다르게 변형

수치형데이터    문자형데이터

생년월일	값증가	값증가생년월일
2011-12-05	+3	2011-12-08
2016-08-09	-2	2016-08-07
2009-02-11	-5	2009-02-06
1998-05-27	-6	1998-05-21
1991-06-18	+9	1991-06-27

수치형데이터    문자형데이터

지역	나이	소득금액(원본)	소득금액(전체 재배열)	소득금액(동질집합 내 재배열)
서울	30대	5,987,900	2,789,200	3,009,600
서울	30대	8,169,700	4,607,300	5,987,900
서울	30대	3,009,600	5,987,900	8,169,700
나주	30대	4,607,300	2,940,100	2,940,100
나주	30대	3,560,800	8,169,700	4,607,300
나주	30대	2,940,100	5,048,300	3,560,800
세종	30대	6,088,400	3,009,600	2,789,200
세종	30대	2,789,200	3,560,800	5,048,300
세종	30대	5,048,300	6,088,400	6,088,400

원본 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967
전체 재배열 분석결과	지역	서울	나주	세종
	평균소득	4,461,467	5,048,300	4,219,600
동질집합 내 재배열 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967

수치형데이터    문자형데이터



수치형데이터    문자형데이터

## 3. 가명·익명처리를 위한 다양한 기술 (기타 기술)

수치형데이터    문자형데이터

수치형데이터    문자형데이터

Record ID	이름	성별	나이	월 납입금액	총 납부금액
1	조미선	F	33	817,250	66,300,000
2	홍길병	M	61	4,559,120	327,700,000
3	김영심	F	50	13,601,564	41,300,000
4	이미정	F	70	979,118	64,600,000
5	김경태	M	40	5,501,809	23,549,000
6	유영근	M	43	609,622	13,900,000

Record ID	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

Record ID	월 납입금액	총 납부금액
1	817,250	66,300,000
2	4,559,120	327,700,000
3	13,601,564	41,300,000
4	979,118	64,600,000
5	5,501,809	23,549,000
6	609,622	13,900,000



### ③ 특이정보 처리 사례

#### 1. 필요성

- ▶ 개인정보를 가명처리를 통해 특정 개인을 알아볼 수 없게 처리했다라도 ‘특이정보’를 통해 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 위험이 높아질 수 있음
  - 따라서 특이정보의 유형 등을 살펴보고 가명정보 내 해당 유형의 정보가 존재하고 있는지 검토할 필요가 있음
    - ※ 특이정보는 관측된 데이터의 범위에서 많이 벗어난 아주 작은 값이나 아주 큰 값을 의미

#### 2. 특이정보 사례

- ▶ 특정 기관의 급여가 2천만원에서 6천만원까지 고루 분포되어 있는데, 일부 고액 급여 수령자가 발생하는 경우
- ▶ 특정 직업의 소속인원이 전국에서 약 300명 정도로 추정되는데, 지역에 극소수(1~2인)만 존재하고 있는 경우
- ▶ 정보공개 규정에 따라 공개되는 정보에서 특정 나이대가 현저하게 적게 나타나는 경우

#### 3. 특이정보 관찰 방법

- ▶ 정보의 특이정보는 3시그마규칙 또는 도수분포표 등을 이용하여 검토할 수 있음
  - 3시그마 규칙 : 68-95-99.7규칙이라고도 하며, 정보의 분포의 3시그마(표준편차) 범위에 거의 모든 값들(99.7%)가 들어가는 것을 의미
  - 도수분포표 : 항목에 대한 값을 적당한 범위로 분류하고, 각 범위에 해당하는 수량을 조사하여 표로 나타내는 것을 의미

〈급여〉		〈지역, 직업〉			〈나이〉	
직원	급여(만원)	주소	직업	빈도	나이(세)	빈도
직원1	2,200	경기	국회의원	5	10-20	4
직원2	3,400	경기	국회의원	5	20-30	11
직원3	4,600	강원	국회의원	1	30-40	21
직원4	5,300	경기	국회의원	5	40-50	18
직원5	10,000	경기	국회의원	5	50-60	5
직원6	6,700	경기	국회의원	5	60-70	1

※ 3시그마 규칙을 이용 하여 표준 편차에 벗어난 특이정보 검토

※ 지역에 대한 도수분포(빈도)를 이용하여 특이정보 검토

※ 특정 나이에 도수분포(빈도)를 측정 하여 특이정보 검토

#### 4. 특이정보 처리 사례

▶ 삭제 기법을 활용한 목적별 사용 예시

- 분석 목적에 해당 정보가 없어도 분석에 크게 영향이 없는 경우에만 가능한 기법, 해당 특이정보를 삭제하여 개인 식별성을 제거

##### 가. 로컬 삭제(Local suppression)

일반적으로 특이정보 처리에 많이 사용되는 기법으로 도수분포표를 활용하여 빈도가 적은 항목을 삭제하여 처리하는 방법

(로컬삭제 기법 예시)

나이	주소	직업	월소득	나이	주소	직업	소득
35	서울	변호사	600 만원	35	서울	변호사	600 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	변호사	500 만원	35	서울	변호사	500 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	변호사	1,200 만원	35	서울	변호사	1,200 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	600 만원	35	경기	변호사	600 만원
35	경기	변호사	1,300 만원	35	경기	변호사	1,300 만원
35	경기	변호사	300 만원	35	경기	변호사	300 만원
35	경기	변호사	900 만원	35	경기	변호사	900 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	울릉도	변호사	200 만원	35	Null	변호사	200 만원

## 나. 행 삭제(Record suppression)

특이정보로 인해 개인의 식별가능성이 있는 경우 사용되는 기법으로 특이정보를 가지고 있는 행 전체를 삭제하여 처리하는 방법

※ 통계 분석에서 특이정보는 분석 목적을 달성하기보다 분석의 목적을 저해하는 요소로 작용하는 경우가 있으며, 이 경우 행 삭제 기법이 가장 적절한 기법이 될 수 있음

〈로컬삭제 기법 예시〉

나이	주소	직업	월소득	나이	주소	직업	소득
35	서울	변호사	600 만원	35	서울	변호사	600 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	변호사	500 만원	35	서울	변호사	500 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	변호사	1,200 만원	35	서울	변호사	1,200 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	600 만원	35	경기	변호사	600 만원
35	경기	변호사	7,300 만원				
35	경기	변호사	300 만원	35	경기	변호사	300 만원
35	경기	변호사	900 만원	35	경기	변호사	900 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	200 만원	35	경기	변호사	200 만원

▶ 통계적 기법의 종류와 목적별 사용 예시

- 분석 목적에 특이정보를 가지고 있는 해당 정보가 필요한 경우 활용하는 기법으로, 해당 특이정보를 통계적인 방법을 통해 통계값으로 변경하여 사용

가. 단일 속성으로 대체(Combining a set of attributes into a single attribute)

숫자형 정보가 아닌 경우(문자형 등) 주로 사용되는 방법으로 분류군의 상위로 묶어 처리하는 방법

※ 특정한 직업이 희귀하여 개인의 식별이 가능한 경우 상위의 분류로 변경하여 사용함으로써 희귀성을 제거

(단일속성 대체 예시)

나이	주소	직업	월소득	나이	주소	직업	소득
35	서울	변호사	600 만원	35	서울	변호사	600 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	변호사	500 만원	35	서울	변호사	500 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	판사	1,200 만원	35	서울	법조인	1,200 만원
35	경기	검사	800 만원	35	경기	법조인	800 만원
35	경기	변호사	600 만원	35	경기	변호사	600 만원
35	경기	변호사	1,300 만원	35	경기	변호사	1,300 만원
35	경기	변호사	300 만원	35	경기	변호사	300 만원
35	경기	변호사	900 만원	35	경기	변호사	900 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	200 만원	35	경기	변호사	200 만원

### 나. 로컬 일반화(Local generalization)

선택한 행에서 일부 특정 값을 일반화하여 활용하는 기법으로, 다른 행의 속성값은 수정하지 않고 희귀 값을 가진 속성값만 처리하여 사용

〈로컬 일반화(상단 코딩) 기법 예시〉

나이	주소	직업	월소득	나이	주소	직업	소득
35	서울	변호사	500 만원	35	서울	변호사	500 만원
35	서울	변호사	700 만원	35	서울	변호사	700 만원
35	서울	판사	1,200 만원	35	서울	변호사	1,200 만원
35	경기	검사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	600 만원	35	경기	변호사	600 만원
35	경기	변호사	1,300 만원	35	경기	변호사	1,300 만원
35	경기	변호사	300 만원	35	경기	변호사	300 만원
...	...	...	...	...	...	...	...
84	경기	변호사	800 만원	80초과	경기	변호사	800 만원
88	경기	변호사	200 만원	80초과	경기	변호사	200 만원

### 다. 부분 총계(Micro Aggregation)

부분 총계는 일부(특정그룹 값의 합)속성에서 정확한 통계적 값을 확인하는 기법으로, 로컬일반화 보다 일부 속성에서 정확한 값을 알 수 있음

〈부분 총계 기법 예시〉

나이	주소	직업	소득	나이	주소	직업	소득
35	경기	변호사	500 만원	35	경기	변호사	500 만원
35	경기	변호사	700 만원	35	경기	변호사	700 만원
35	경기	변호사	6,200 만원	35	경기	변호사	6,750 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	600 만원	35	경기	변호사	600 만원
35	경기	변호사	7,300 만원	35	경기	변호사	6,750 만원
35	경기	변호사	300 만원	35	경기	변호사	300 만원
35	경기	변호사	900 만원	35	경기	변호사	900 만원
35	경기	변호사	800 만원	35	경기	변호사	800 만원
35	경기	변호사	200 만원	35	경기	변호사	200 만원

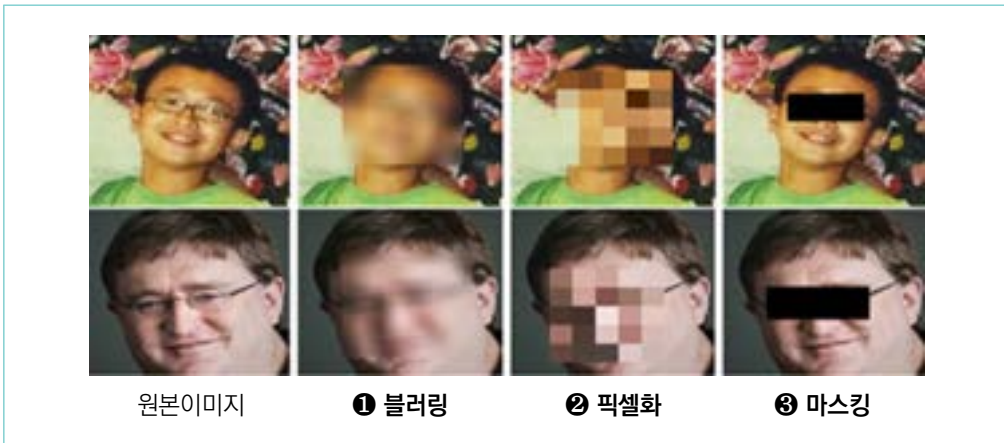
## 2



### 1. 영상정보 가명처리 기술

#### ① 이미지 필터링 기술

▶ 필터링 기술(① 블러링, ② 픽셀화, ③ 마스킹)로 처리된 영상 이미지



※ 출처 : Tao Li and Lei Lin, Natural Face De-identification with Measurable Privacy(CV-COPS, 2019)

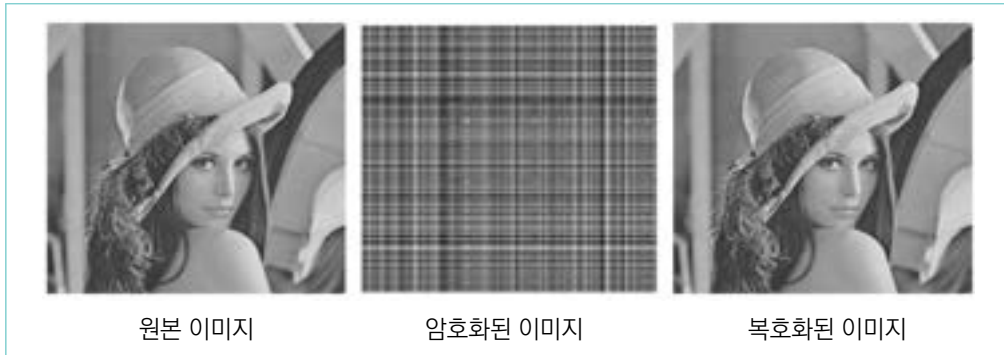
▶ 필터링 기술 : 블러링(Blurring), 픽셀화(Pixelization), 마스킹(Masking)

종류	설명
평균	<p>- 입력 이미지의 현재 위치에서 예를 들어 3×3 격자 범위의 주변 픽셀값의 평균을 구하여 원 픽셀값을 결과 이미지의 픽셀값으로 대체</p> <p>=&gt; <math>e = 1/9(a+b+c+d+e+f+g+h+i)</math></p>



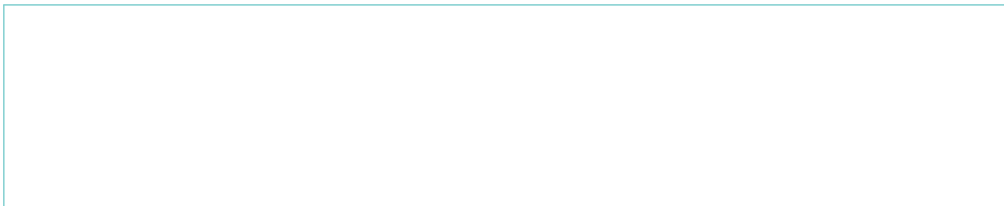
## ② 이미지 암호화

- 원본 이미지의 일부를 암호화하여 데이터 주체를 알아볼 수 없도록 하는 기법
- 이미지 암호화로 처리된 영상 이미지



※ 출처 : Lihua Gong 외 3인, Image compression-encryption algorithms by combining hyper-chaotic system with discrete fractional random transform(Optics and Laser Technology, 2018)

- 이미지 암호화 기술 : 이산코사인변환 기반 암호화, 픽셀 위치 기반 암호화



## ③ 얼굴 합성(프라이버시 보존형 데이터 마이닝)

- ▶ K-익명성 프라이버시보호 모델을 확장하여 K명의 얼굴을 합성한 기술로 **K-same 모델**로도 부름

※ 출처 : E.M.Newton 외 2인, preserving privacy by de-identifying face images(2005)

- ▶ K-same 기법으로 처리된 영상 이미지



※ 출처 : R. Gross 외 2인, Integrating utility into face de-identification(2005)

## ▶ K-same 모델의 개선

- K-same 모델을 보다 개선하여 개인정보 보호와 유용성의 균형\*을 맞추기 위한 **K-Same-Select 모델** 등장

\* 공개 위험(즉, 이미지 난독화 수준)과 분류 정확도 간의 균형

※ 출처 : R. Gross 외 2인, Integrating utility into face de-identification(2005)

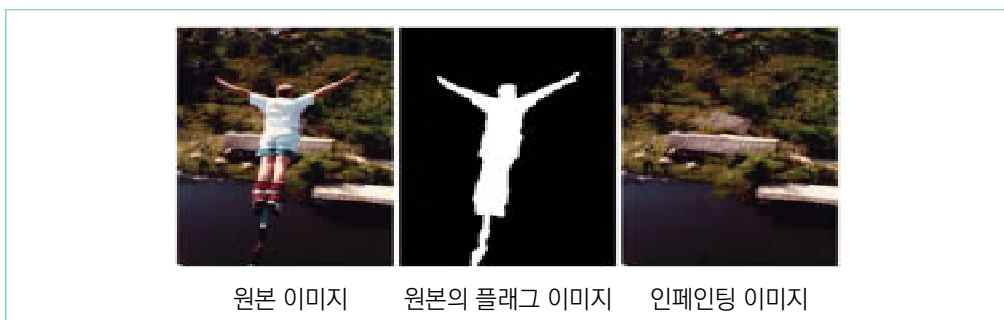
- **K-same 모델의 정확도를 유지하면서** 데이터 셋의 민감한 정보를 숨기기 위해 기존 샘플에 **무작위 노이즈를 추가하거나 새 샘플을 생성하는 난독화(obfuscate)\* 기능을 추가**로 설계

\* 안전성은 높으나 성별이 바뀌는 등 이미지 왜곡이 심함

※ 출처 : T. Zhang, Privacy-preserving machine learning through data obfuscation(2018)

## ④ 인페인팅(Inpainting)

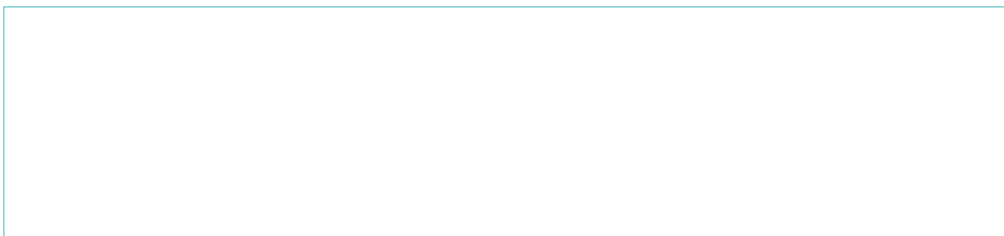
- ▶ 영상 내 개인 식별 영역을 제거한 후 다른 물체 또는 배경으로 대체하여 신원을 보호하는 기술
- ▶ 인페인팅 기법으로 처리된 영상 이미지



원본 이미지      원본의 플래그 이미지      인페인팅 이미지

※ 출처 : Takahiro Ogawa 외 1인, Image inpainting based on sparse representations with a perceptual metric(2013)

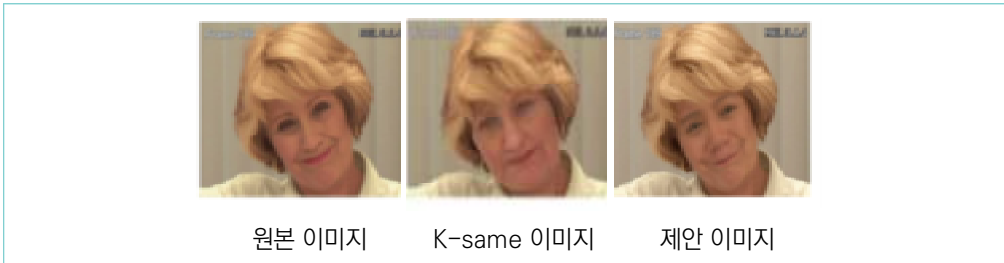
- ▶ 인페인팅 기술 : 패치 기반 인페인팅, 객체 기반 인페인팅



⑤ AI 이용 영상정보 가명처리

▶ 얼굴 보존형 가명·익명처리 기술(De-identification without losing faces)

- 원본 얼굴의 요소를 변경하거나 얼굴을 완전히 합성하는 대신, 훈련된 얼굴 속성 전달 모델을 사용하여 동义的한 대상의 소수(보통 2~3명)인 기증자의 얼굴에 비 신원 관련 얼굴 속성을 매핑

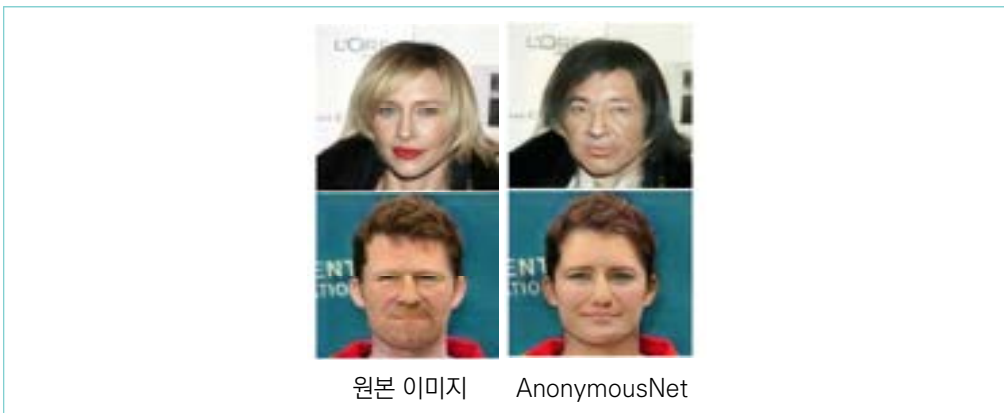


※ 출처 : Yuezun Li, De-identification Without Losing Faces(2019)

▶ AnonymousNet 프레임워크

- 얼굴 보존형 가명·익명처리 기술의 프레임과 비슷해 보이지만 다음의 4단계 절차를 통하여 아이덴티티가 완전히 다른 이미지를 생성

※ (1단계) 얼굴 특징 추출, (2단계) 의미 기반 속성 난독화, (3단계) 익명화된 얼굴 생성, (4단계) 적대적 교란 절차



※ 출처 : Tao Li 외 1인, AnonymousNet: Natural Face De-Identification with Measurable Privacy(CV-COPS 2019)

▶ AI 딥러닝 기반의 알고리즘을 활용

- 얼굴 및 차량 번호판 등을 추출한 후 각종 SW 라이브러리를 활용하여 블러링 처리

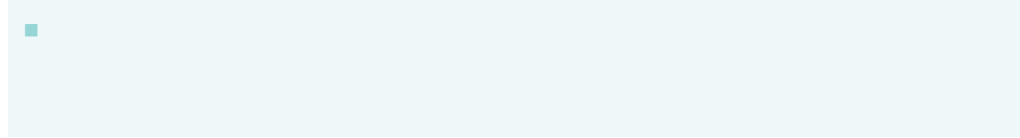
※ 미검출이나 오검출 얼굴 등에 대해서는 수작업을 통하여 추가 보정

▶ AI 적대적 생성 신경망 기반 모델(GAN, Generative Adversarial Networks)

- 이미지에서 보존해야할 영역을 수치화하여 연속형 데이터로 처리한 후 해당 개별 데이터에 노이즈를 추가하거나 재현 처리

## 2. 음성정보

### ① 음성정보 자체에 대한 가명처리



#### ▶ 규칙기반 개인정보 단순 삭제

##### - 음성 데이터 상에서 규칙\*을 정하여 개인식별가능 정보를 단순히 삭제

\* (예) 인명, 지명 등 개인식별가능 정보가 포함된 카드번호, 주민등록번호, 전화번호 등 삭제하고, 비속어, 성적 및 정치적 발언의 경우도 삭제하거나 다른 언어로 대체

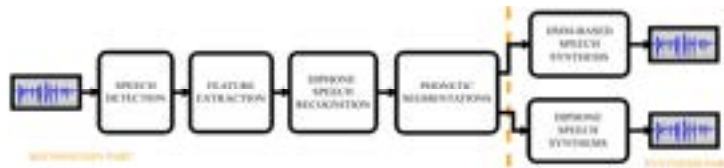
※ 이 방법은 완전 자동으로 수행이 어려운 측면이 있으므로 수작업으로 확인이 필요하여 시간과 비용이 많이 소요

##### - 음성 변형의 원리에 기반하여 음성 데이터 내용에 영향을 주지 않으면서 주어진 발화의 비언어적 특징들을 수정

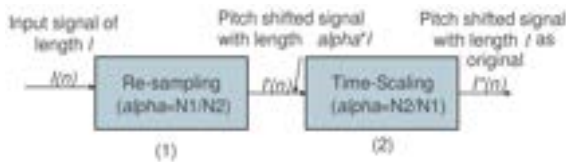
※ 소스로는 시간량(time-scale), 음조(pitch) 또는 음량(energy)을 변조할 수 있으며, 필터로는 음성 트랙 시스템에 기반하여 음폭을 변형 가능

종류	설명
시간량 (time-scale) 수정	- 원래 음성의 지각적(perceptual) 품질에 영향을 주지 않으면서 조음 속도를 변경
음조 (pitch, 음높이, 고저) 수정	- 짧은 시간 스펙트럼 엔벨로프(포먼트의 위치 및 대역폭)와 시간을 보존 하면서 스펙트럼의 하모닉 구성 요소 사이의 간격을 압축하거나 확장 하기 위해 기본 주파수를 변경
음량(energy) 변형	- 입력 음성의 인지된 크기를 수정
필터 수정	- 성도(vocal tract) 시스템의 주파수 응답의 크기 스펙트럼 수정 (예 : 여성의 목소리를 수정하여 아이처럼 들리게 함)
소스 수정과 필터 수정의 조합	- 일반적으로 사용되는 기법으로 예를 들어 화자의 음성을 다른 화자의 음성처럼 들리도록 수정하려면 운율(prosody)과 성도(vocal tract) 수정을 결합해야 함 → 이때 특정 대상 화자가 제공되는 경우 이를 음성 변환(voice conversion) 이라 부름

종류	설명
텍스트 종속적	- SI 학습 단계에서 병렬 말뭉치(동일한 텍스트를 발화하는 소스 및 대상 화자의 훈련 자료)가 필요 ※ 실제 응용 프로그램에서 음성 비식별화에 이러한 방식을 사용하는 것은 주요 제약 사항임
텍스트 독립적	- 학습 단계에서 병렬 말뭉치가 필요하지 않으며 발화자의 개인정보 보호 차원에서 보다 효과적이고 현실적임



※ 출처 : T. Justin 외 5인, Speaker de-identification using diphone recognition and speech synthesis(2015)



※ 출처 : Y. Stylianou, Voice transformation : a survey(2009)

## ② 음성을 텍스트로 변환(STT, Speech To Text) 후 가명처리

- ▶ 개인식별정보가 포함된 음성을 텍스트로 변환 후 변환한 텍스트에서 개인식별정보를 가명처리하고 다시 음성으로 변환하는 방식

### 3. 텍스트정보

#### ① 규칙기반 개인정보 단순 삭제 혹은 마스킹

▶ 사전에 텍스트 내 개인식별(가능)정보들을 정의하고 정의된 형태(포맷)에 기반하여 사람에 의한 수작업 검색을 통하여 개인정보를 단순 삭제나 마스킹 혹은 대체 처리한 후 검증

- 데이터량이 많지 않을 경우 전수 검증 필요
- 데이터량이 많은 경우 샘플링(5%, 10% 등)을 통한 검증 필요
- 휴먼 에러가 발생할 수 있으므로, 사후 검증 단계를 반드시 수행

※ 아래 소개된 ②~⑤의 방법들은 지난 2018년 발표된 ISO/IEC 20889 표준 문건을 참조하여 재구성한 것임

#### ② 스크러빙(Scrubbing)

▶ 원 텍스트의 내용과 구조를 보존하면서 즉석해서 파싱을 통하여 혹은 파싱 이후 개인식별(가능)정보만을 제거(마스킹 혹은 대체)하는 것으로 이 경우 다수의 정보 주체와 해당 속성들 사이의 명확한 연관성이 없어질 수도 있음

- 즉, 특정 정보 주체가 어느 속성을 지칭하는지 알 수가 없게 될 수도 있음
- 단순 삭제 혹은 마스킹 방법과 유사하며 수작업이 아닌 자동화 SW를 이용한다는 점이 다름

#### ③ 정규표현식(Regular Expression)

▶ 문자나 혹은 문자열의 일정한 패턴을 표현하는 일종의 형식 언어

	항목	텍스트 예시	정규표현식
1	전화번호	010-1111-2222	$^{\text{d}}\{3\}\text{-}\text{d}\{4\}\text{-}\text{d}\{4\}\$$
2	이메일주소	aa@bb.net	$^{\text{[0-9a-zA-Z]}([-_]?[0-9a-zA-Z])*\text{@[0-9a-zA-Z]}([-_]?[0-9a-zA-Z])*\text{[a-zA-Z]}{2,3}\$/i$
3	IP주소	111.111.111.111	$^{\text{d}}\{1,3\}\.\text{d}\{1,3\}\.\text{d}\{1,3\}\.\text{d}\{1,3\}\$$
4	주민등록번호	220101-1111111	$^{\text{d}}\{6\}\text{-}[1-4]\text{d}\{6\}\$$

※ 사용시 오타나 혹은 문자 숫자 혼합 등에 따라 누락되는 경우 존재하므로 수작업에 따른 별도 확인 작업이 필요

#### ④ 주석 달기(Annotation)

- ▶ 주어진 텍스트를 논리적으로 분할한 후 분할된 단어(들)에 주석을 첨가하는 기법

종류	설명
규칙 기반	- 구문 문법의 규칙에 따라 텍스트를 토큰(예 : 사전 정의된 수의 단어)으로 나누는 것으로 고급 규칙의 경우 정규표현식을 사용하여 정의
사전 기반	- 개인식별(가능)정보들을 미리 사전으로 정의한 후 개체명 인식(NER, Named Entity Recognition) 기술을 이용하여 주어진 텍스트(이름, 주소, 전화번호 등)와 일치시킴 * 사전기반의 경우 AI 기술을 활용하여 각 단어들을 자동으로 인식하지만 모든 단어(들)을 완벽하게 인식하는 것은 어려우며, 따라서 맥락에 의한 개인식별 가능성에 대한 조치는 어려울 수 있음

#### ⑤ AI 기반 텍스트정보 가명처리

- ▶ 규칙, 정규표현식 등을 통한 개인정보 검출 및 마스킹은 정확도 측면에서 한계가 있을 수 있으며, 이를 보완하기 위해 딥러닝 기술 등을 적용한 자연어 처리 언어 모델을 통해 사전에 정의되지 않은 패턴의 개인정보를 검출하고 마스킹할 수 있음
- ▶ 학습방법에 따라 다양한 형태의 인공지능 기반 개인정보 검출 기법 존재(HMM, MEM, CRFs, structural SVM, Deep-Learning)
- ▶ 규칙기반의 유연성 부족을 해결하기 위해 패턴이나 규칙을 수동이나 반자동으로 작성하고 인공지능을 통해 사전(dictionary)을 확장하여 규칙에 적용되는 개인식별가능정보를 새롭게 검출할 수 있음 (단계별 규칙을 순차적으로 적용하여 가중치에 따라 인명, 지명, 조직명 등으로 범주를 결정하는 등)

#### ⑥ 텍스트를 테이블 형식으로 변환

- ▶ 주어진 텍스트를 구문 문법의 규칙에 따라 파싱(parsing, 정해진 규칙에 따라 문장의 구문을 분할)한 다음 분할된 각 세그먼트들을 열과 행이 있는 테이블 형태로 정렬한 후 나머지 데이터들은 삭제
  - 변환된 테이블은 기존 정형 데이터에 대한 가명처리를 적용
  - 실제로는 텍스트 자체의 복잡성으로 인하여 구조화된 테이블로의 변환이 불가능할 수도 있음



# 5

## 서식 모음

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제5장 서식 모음

- ▶ 작성 주체 :
- ▶ 작성 목적 :
- ▶ 필수 여부 :
- ▶ 대체가능서류 :

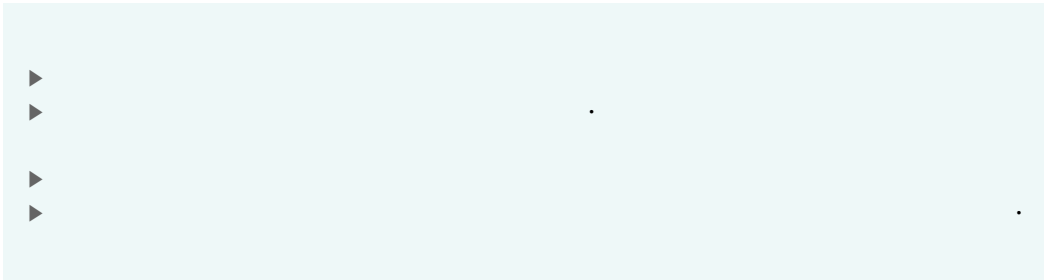
### 가명정보 이용·제공 신청서 ( 신규 변경 )

접수번호	기존 접수번호(변경시)		접수일
신청자	조직/부서명		
	담당자 직위	담당자 성명	
	전화번호	이메일 주소	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존 ※ 세부 활용 목적 얼굴 유형 데이터를 활용하여 고객의 구매 결정 과정과 구매 상품, 선호하는 커뮤니케이션 방식을 예측하기 위한 실시간 응대 전략 수립		
가명처리 대상 개인정보	개인정보 명칭		
	요청 개인정보 항목	고객 얼굴사진정보, 구매 상품 카테고리별 금액, 상담 내용 등	
결합 활용 여부	<input type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )		
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 (    )		
반복 제공 여부	<input type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정		
예상 이용기간	반출일로부터 ( )개월 이내 또는 반출일로부터 ~ 년 월 일 까지		
「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.			
			년 월 일
			신청인(부서장명) (서명 또는 인)
OOOOO장 귀중			
첨부서류	1. 가명정보 활용 계획서		

- ▶ 작성 주체 :
- ▶ 작성 목적 :
- ▶ 필수 여부 :
- ▶ 대체가능서류 :

### 가명정보 활용 계획서

신청 제목	(예시) 구매 상품과 얼굴 유형을 기반으로 고객 마케팅 전략 수립
책임자	소속                      성명                      (서명)
배경 및 목적	<p>(예시)</p> <p>전통적 마케팅의 한계 : 기존 마케팅 전략은 인구통계학적(연령, 성별) 또는 심리통계학적(라이프스타일, 가치관) 정보에 주로 의존. 이러한 정보는 고객이 매장에 들어서는 순간 즉각적으로 활용하기 어려움.</p> <p>인상경영학의 실무적 적용 가능성 : 인상학(Physiognomy)은 동서양에서 오랜 역사를 지닌 학문으로, 얼굴의 생김새가 개인의 성향과 관련이 있다는 관점을 제시. 본 연구는 이 관점을 현대의 소비자 행동론 및 마케팅 영역으로 가져와 과학적으로 검증하려는 시도</p> <p>가설 설정 및 검증 : 기사에서 제시된 역삼각형, 마름모형, 계란형, 삼각형 등 네 가지 얼굴 유형별 소비 성향(예: 가성비 추구, 개성 중시, 안정성 선호 등)의 통계적 유의성을 검증</p> <p>실무적 모델 개발 : 얼굴 유형 데이터를 활용하여 고객의 구매 결정 과정과 선호하는 커뮤니케이션 방식을 예측할 수 있는 실시간 응대 가이드라인을 제시함으로써, 서비스 및 판매 효율성을 극대화</p>
활용 데이터 (대상자 수)	(예시) 고객 기본정보(얼굴사진 포함), 구매 상품 카테고리별 금액, 상담 내용 정보
활용 계획 및 방법	<p>(예시)</p> <p>1. 얼굴 유형 분류 시스템 구축:</p> <ul style="list-style-type: none"> <li>• 인상경영학 전문가의 자문을 통해 네 가지 얼굴 유형을 정의하는 객관적인 **인면 측정 기준(Facial Metric Criteria)** 설정</li> <li>• 수집된 얼굴 이미지 데이터를 기반으로 딥러닝 기반의 **얼굴 유형 자동 분류 모델(CNN 또는 Vision Transformer)**을 개발하여 유형별 분류 정확도 확보</li> </ul> <p>2. 데이터 통합 및 가설 검증:</p> <ul style="list-style-type: none"> <li>• 얼굴 유형 분류 결과(범주형)와 구매 행동 데이터(연속형/범주형) 통합</li> <li>• 통계 기법을 사용하여 '얼굴 유형별 소비자 행동 특성'에 대한 가설 검증</li> <li>• 얼굴 유형이 '가성비 추구(역삼각형)'나 '단골 유지율(삼각형)' 등의 특정 행동에 미치는 영향력 측정</li> </ul> <p>3. 마케팅 전략 모델 최적화:</p> <ul style="list-style-type: none"> <li>• 통계적 검증 결과를 바탕으로, 각 얼굴 유형에 최적화된 판매 스크립트, 상품 추천 방식, 매장 분위기 조성 방안 등을 담은 실무형 **얼굴 유형 기반 고객 응대 매뉴얼**을 최종적으로 도출</li> </ul>
(예상) 활용 기간	(예시) 2025년 3월~ 2026년 2월까지 (12개월)
기대효과 및 활용방안	(예시) 얼굴 형태와 소비 심리 간의 관계를 실증적 데이터로 입증함으로써, 인상경영학적 관점을 현대 소비자 행동론에 통합하는 새로운 학문적 지평을 제시



## 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적			
가명처리 대상 개인정보			
활용하려는 자	(기관명)	(직위)	(성명)

(1) 기본 위험도 판단	
<b>〈Step 1〉 활용 주체</b>	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등 <input type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공
<b>〈Step 2〉 처리 환경</b>  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우 <input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
<b>기본 위험도 자동 도출</b>	〈Step 1〉 내부 활용 → <input type="checkbox"/> 저위험 〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input type="checkbox"/> 중위험 〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input type="checkbox"/> 고위험 ※ 판단 이유 :

(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장

(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input type="checkbox"/> 고위험
종합 검토 결과			
		검토일 :   년   월   일	
검토자 소속 :		성명 :   (서명 또는 인)	



▶	
▶	
▶	
▶	서류

### 가명처리 계획서

파일 개요						
No	파일관계	파일명	항목 수	데이터 건수	정보주체 수	비고
1	주	고객정보	5	10,000	10,000	
2	부	구매액집계정보	10	240,000	10,000	
3	부	상담정보	2	30,000	3,000	

파일별 상세 가명처리 계획						
No	파일명	항목명	유형*	예시	항목설명	가명처리 계획
1	고객정보	고객번호	KEY	19884312		대체-동일인 구별
2	고객정보	사진	이미지	증명사진	고객 얼굴 사진	얼굴 유형만 남김
3	고객정보	나이	범주형	45		일반라운딩-5세 내림
4	고객정보	주소	범주형	서울시 중구 ...		부분삭제-시도이하
5	고객정보	성별	범주형	M		미적용
6	구매액집계정보	고객번호	KEY	19884312		대체-동일인 구별
7	구매액집계정보	기준연월	범주형	202401		미적용
8	구매액집계정보	여행용품 구매액	수치형	10,000		상단코딩 99% 이상
9	구매액집계정보	식품류 구매액	수치형	20,000		상단코딩 99% 이상
10	구매액집계정보	의류 구매액	수치형	30,000		상단코딩 99% 이상
11	구매액집계정보	취미용품 구매액	수치형	40,000		상단코딩 99% 이상
12	구매액집계정보	생활용품 구매액	수치형	50,000		상단코딩 99% 이상
13	구매액집계정보	구매액 합계	수치형	150,000		상단코딩 99% 이상
14	구매액집계정보	선호제품군	범주형	생활용품		미적용
15	구매액집계정보	월별고객등급	범주형	VIP		미적용
16	상담정보	고객번호	KEY	19884312		대체-동일인 구별
17	상담정보	상담내용	텍스트	free text 형태	챗봇 상담 내용	식별정보 삭제

\* 정형데이터인 경우 범주형, 수치형 중 택일  
비정형데이터인 경우 음성, 텍스트, 이미지, 영상, 특히 보건의료 분야는 CT, MRI, 유전체 등 구체적으로 표기



서류 :

## 가명처리 결과서

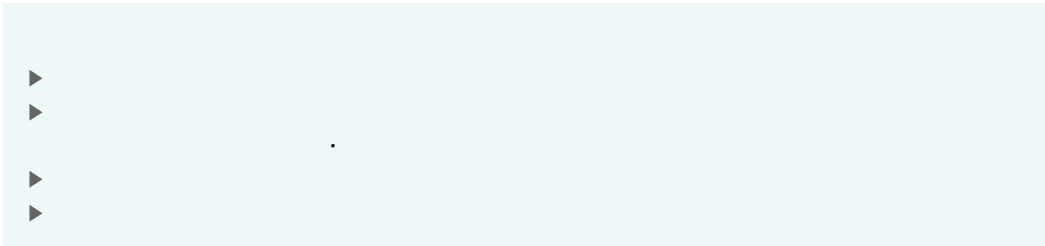
1. 공통적으로 적용한 가명처리 방법					
자료형	가명처리 순서				
	가명처리 방법 1	가명처리 방법 2		가명처리 방법 3	
결합기					
범주형					
수치형					
기타					

2. 예외적으로 적용한 가명처리 방법					
파일명	항목명	자료형	가명처리 방법1	가명처리 방법 2	가명처리 방법 3

3. 주요 범주형 항목 값 분포 (검토가 필요한 항목만 작성)					
파일명	항목명	값	빈도수	구성비	비고
		합계			

4. 주요 수치형 항목 값 분포 (검토가 필요한 항목만 작성)									
파일명	항목명	처리 전				처리 후			
		Min	빈도	Max	빈도	Min	빈도	Max	빈도

5. 비정형데이터 전후 비교 샘플링 (비정형데이터가 포함된 경우에만 작성)			
파일명	항목명(정보영역)	처리 전	처리 후
		※ 예시 사진	※ 예시 사진
		※ 설명	



### 비정형데이터 추가검수 결과서

검수 개요			
접수번호 (또는 처리명칭)	구매 상품과 얼굴 유형을 기반으로 고객 마케팅 전략 수립		
원본 데이터 규모	이미지 장수, 동영상 클립 수, 발화정보 개수, 글자 수, 기타(러닝타임, 해상도 FPS 등)	원본 데이터 크기(용량)	00GB, 00MB 등
원본 데이터 포맷	JPG, MP4, TXT 등	처리 결과 데이터 포맷	JPG, MP4, TXT 등
가명처리 적용 기술	사람얼굴 부분을 탐지하여 자동적으로 모자이크 처리		

검수 내용	
검수 기간	20 년 월 일 ~ 20 년 월 일
검수 장소	제공기관 가명처리실 내
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수 전체 데이터에서 10%를 무작위로 추출 후 추출된 데이터에 대한 검수 수행
처리된 결과 확인 사항	부분 검수 시 나타난 일정한 패턴의 미처리 부분을 전체에 반영하여 추가 가명처리 실시
검수 결과	재처리후 10% 부분 검수 결과, 재식별 가능성이 거의 없는 것으로 확인

검증자	소속 및 직위	성명	서명(인)

20 년 월 일

작성자 소속 :  
성명 : 서명(인)

## 적정성 검토 결과서

접수번호 (또는 처리명칭)					
검토위원 정보	성명		소속		직위
검토 대상	<input type="checkbox"/> 신규 <input type="checkbox"/> 보완				
검토 일자	년 월 일				
검토 결과	<input type="checkbox"/> 적정 (승인) <input type="checkbox"/> 조건부 승인 <input type="checkbox"/> 부적정 (반려)				
검토 사항	① 가명정보 처리 목적 적합성 - 보호법 제28조의2에서 정한 가명정보 처리 목적 (통계작성, 과학적 연구, 공익적 기록보존)에 해당하는가? - 처리 목적을 구체적으로 설정하였는가?	<input type="checkbox"/> 적정 <input type="checkbox"/> 미흡			
	② 가명정보 이용항목 적합성 - 처리 목적 달성에 필요한 항목만 처리 대상에 포함하였는가?	<input type="checkbox"/> 적정 <input type="checkbox"/> 미흡			
	③ 위험도 판단의 적정성 - 활용 주체, 처리 환경, 개별 사례별 특수성 등을 고려하여 적정하게 위험도를 판단하였는가? 판단 근거가 적절한가?	<input type="checkbox"/> 적정 <input type="checkbox"/> 미흡			
	④ 가명처리의 적정성 (재식별 위험성) - 항목별 가명처리 방법과 수준을 적절하게 설정하여 처리하였는가? - 가명정보 데이터 자체의 위험성, 처리환경의 위험성을 종합적으로 고려하였을 때 종합적인 재식별 위험성을 충분히 낮추었는가?	<input type="checkbox"/> 적정 <input type="checkbox"/> 미흡			
	⑤ 목적 달성 가능성 - 당초 설정한 가명정보 처리 목적을 달성할 수 있는가? (과도한 가명처리로 처리 목적 달성이 어려워지지 않았는가?)	<input type="checkbox"/> 적정 <input type="checkbox"/> 미흡			
검토 의견					
종합 검토 의견 (위원장)	※ 위원장일 경우에만 위원들 의견을 종합하여 최종 검토결과를 작성				
위와 같이 적정성 검토 결과를 통지합니다.					
작성자			서명(인)		
OOOOO장 귀중					

**【서식 설명】**

- ▶
- ▶
- ▶
- ▶

**적정성 평가위원 서약서**

<b>처리명칭</b>	
<b>신청기관</b>	

본인은 상기 적정성 검토와 관련하여 아래와 같이 적정성 검토 대상 가명정보 및 가명정보 이용신청자와 이해관계가 없음을 서약합니다.

구분	이해 관계 내용	예	아니오
1	적정성 검토 대상 가명정보를 이용할 예정이 있는지		
2	적정성 검토 대상 가명정보 활용에 대한 경제적·비경제적 이익을 가지고 있는지		
3	가명정보 이용신청자와 고용관계(상근, 비상근/ 공식, 비공식 등)에 있는지		
4	가명정보 이용신청자로부터 본 적정성 검토 비용 외에 검토 결과에 영향을 미칠 수 있는 경제적·비경제적 이익을 제공받은 사실이 있는지		
5	본인 또는 배우자의 직계가족이 소속된 회사가 위에서 기술된 것과 같은 관계를 가지고 있는지		
6	그 밖에 적정성 검토 대상 가명정보 또는 가명정보 이용신청자와 이해관계가 있는지		

본인이 확인한 모든 내용은 정확히 기술되었으며 만약 평가 진행 중에 의뢰기관에 대한 이해관계가 변동되는 이해상충이 생기는 경우 이를 인지한 날로부터 5영업일 이내에 귀 기관에 통지 하였습니다.

또한 본인은 가명처리 적정성 검토 과정에서 획득한 모든 정보에 대하여 귀 기관의 허락 없이 외부에 공개하지 않을 것을 서약합니다.

년 월 일

소 속 :                      서 약 자 :                      (인)

000000장 귀중

## 【서식 설명】



## 안전조치 이행확약서

 안전조치 이행 확약

본 기관은 「개인정보 보호법」에서 규정하고 있는 가명정보에 대한 안전조치의무 등(제28조의4) 및 가명정보에 대한 안전성 확보 조치(동법 시행령 제29조의5)를 성실히 이행하고 관련 법령을 준수하겠습니다. 그리고 이를 이행·준수하지 아니하여 발생하는 모든 법적 책임을 부담할 것을 확약합니다.

 가명정보 취급자 서약

No	소속	직위	역할	성명	서명

위와 같이 서약함으로써, 본인은 정보의 비밀을 지키기 위해 합당한 역할과 완전한 책임을 다할 것에 동의합니다. 그리고 이를 이행·준수하지 아니하여 발생하는 관련 법적 책임을 부담할 것을 서약합니다.

20    년    월    일

책임자 :

서명(인)

신청기관 :

(직인)

000000장 귀중

【서식 설명】

- ▶
- ▶
- ▶
- ▶

가명정보·추가정보 처리대장 (제공자)

일련 번호	처리일자	관리 번호	정보 구분	처리 구분	처리내역 처리목적 및 내용 / 항목 / 이용기간 / 제공시 받는자를 기재	담당자	책임자
1	25-7-25	25-01	추가정보	생성	• Key + Salt 값으로 해시 생성		
2	25-7-29	25-01	가명정보	생성	• 처리목적 : 비정형데이터 상담내용 NER 처리		
3	25-9-12 ~ 9-19	25-01	가명정보	이용	• 처리목적 : 가명처리 결과에 대한 검증(전수)		
4	25-9-22	25-01	가명정보	제공	• 처리목적 : 과학적연구 (비정형 텍스트 가명처리 솔루션의 개인정보 검출 시 모델 성능 고도화) • 항목 : 상담식별번호, 상담일시, 상담내역 등 • 이용기간 : 2025.9.22.~12.31. • 제공받는자 : (주)OOOO		
5	25-9-30	25-01	가명정보 추가정보	파기	• 제공받는자에게 제공 후 내부 보관 가명정보는 파기함		

가명정보·추가정보 처리대장 (이용자)

일련 번호	처리일자	관리 번호	정보 구분	처리 구분	처리내역 처리목적 및 내용 / 항목 / 이용기간 / 제공시 받는자를 기재	담당자	책임자
1	25-9-30	25-가-01	가명정보	보관	• 제공기관A로부터 받은 가명정보를 폐쇄망 환경 B에 보관		
2	26-01-02	26-가-01	가명정보	파기	• 분석이 완료됨에 따라 가명정보 파기		

가명정보·추가정보 접근권한 관리대장 (제공자)

일련 번호	신청일자	관리번호	소속 (부서)	취급자	취급자P	처리 내역 및 사유	권한 범위	부여/말소 일자	담당자	책임자
1	25-7-20	25-01	A팀	홍길동	1.1.1.1	추가정보 처리 권한 부여	생성조회	25-7-25		
2	25-7-20	25-01	B팀	이순신	1.1.1.2	가명처리 권한 부여	생성 조회	25-7-25		
3	25-9-20	25-01	A팀	홍길동	1.1.1.1	추가정보 파기로 권한 회수	말소	25-9-30		
4	25-9-20	25-01	B팀	이순신	1.1.1.2	가명정보 제공으로 인한 권한 회수	말소	25-9-22		

가명정보·추가정보 접근권한 관리대장 (이용자)

일련 번호	신청일자	관리번호	소속 (부서)	취급자	취급자P	처리 내역 및 사유	권한 범위	부여/말소 일자	담당자	책임자
1	25-9-30	25-가-01	마케팅팀	김활용	2.1.2.3	가명정보 분석	조회	25-10-1		
2	25-12-31	25-가-02	마케팅팀	김활용	2.1.2.3	분석 완료로 인한 권한 회수	말소	25-12-31		

단계	작성 서류	작성 주체			비고
		가명정보 제공기관	가명정보 이용자	적정성 검토자	
(1단계) 사전준비	① 가명정보 이용·제공 신청서		✓		☞ 내부 활용 시 작성생략 가능
	② 가명정보 활용 계획서		✓		
(2단계) 위험성 검토	③ 위험성 검토서	✓			
(3단계) 가명처리	④ 가명처리 계획서	✓	상호 협의		☞ 제공기관이 작성하되, 가명정보 이용자의 활용 목적 달성을 위해 이용자의 요구수준을 파악하고 상호 간 협의하여 작성  ☞ 반복·유사 활용 시 작성생략 가능
	⑤ 가명처리 결과서	✓			☞ 반복·유사 활용 시 작성생략 가능
	⑥ 비정형데이터 추가검수 결과서	✓			☞ 비정형데이터 포함 시에만 작성
(4단계) 적정성 검토	⑦ 적정성 검토 결과서			✓	☞ 적정성 검토 시 ①~⑥번 서류 검토  ☞ 반복·유사 활용 시 작성생략 가능
	⑧ 적정성 평가위원 서약서			✓	☞ 적정성 검토위원회 미개최 시 작성생략 가능
(5단계) 안전한 관리	⑨ 안전조치 이행 협약서	협약 요청	✓		☞ 제공기관이 필요한 내용을 담아 이행협약을 요청하고, 이용자가 작성  ☞ 제공기관이 통제 가능한 환경에서 활용하는 경우, 작성생략 가능
	⑩ 가명(추가)정보 관리대장	✓	✓		☞ 제공기관은 제공 관련 사항을, 이용자는 활용· 파기 관련 사항 기록·관리

# 6

## 가명정보 처리 관련 주요 운영문서 작성 가이드

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제6장

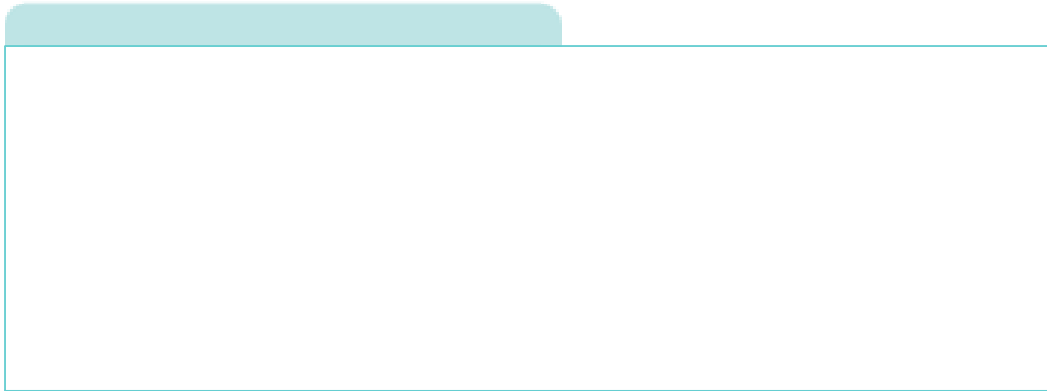
# 가명정보 처리 관련 주요 운영문서 작성 가이드

### 작성가이드 1

- ◆ 개인정보처리자는 가명정보 및 추가정보를 안전하게 관리하기 위한 내부 관리계획을 수립·시행하여야 함 (보호법 시행령 제29조의5 제1항 제1호)

※ 다만, 개인정보 개념에 가명정보 개념이 포함되므로, 개인정보의 안전한 관리를 위하여 수립·시행된 내부 관리 계획이 있을 경우, 가명정보의 처리에 관한 내용만 추가하여 수립·시행하는 것도 가능

- ◆ 내부 관리계획에는 추가정보의 별도 분리 보관 및 이에 대한 접근권한 분리에 대한 사항 등을 포함하여야 함



- ◆ 개인정보처리자는 내부 관리계획에서 정한 사항에 중요한 변경이 있는 경우 이를 즉시 반영하여 내부 관리계획을 수정·시행하고, 관리책임자는 연 1회 이상 내부 관리계획의 이행 실태를 점검·관리하여야 함

**제00조(가명정보 및 추가정보 관리책임자 지정)** ① 개인정보 보호책임자는 다음과 같은 역할을 수행한다.

1. 가명정보에 대한 내부 관리계획의 수립·시행
2. 내부 관리계획의 이행실태 점검 및 관리
3. 가명처리 및 적정성 검토 현황 관리
4. 가명정보 및 추가정보에 대한 관리·감독
5. 가명정보 처리 현황 및 관련 기록 관리
6. 가명정보를 처리하는 자 교육계획의 수립 및 시행
7. 가명처리 및 가명정보 처리 위탁 사항에 대한 관리·감독(해당 시)
8. 가명정보에 대한 재식별 모니터링 및 재식별 시 처리 방안의 수립·시행
9. 그 밖의 가명정보 처리에 대한 보호에 관한 사항

**제00조(가명정보 및 추가정보 분리 보관)** ① 가명정보는 가명처리가 완료되면 가명처리 전 개인정보와 분리·보관하여야 한다.  
 ② 가명처리의 과정에서 발생하는 추가정보는 원칙적으로 삭제한다. 다만, 불가피한 경우 가명정보와 분리·보관하여야 한다.  
 ③ 가명처리 전 개인정보, 가명정보 및 추가정보는 물리적으로 분리 보관하는 것을 원칙으로 하며 물리적 보관이 어려운 경우 논리적인 분리를 시행할 수 있다.  
 ④ 논리적으로 분리·보관하는 경우 엄격한 접근통제를 적용해야 한다.

**제00조(가명정보 및 추가정보에 대한 접근권한 분리)** ① 가명처리가 완료되면 가명정보 또는 추가정보의 접근권한은 최소한의 인원 으로 엄격하게 통제하여야 하며, 업무에 따라 차등적으로 부여 하여야 한다.  
 ② 추가정보에 대한 접근권한과 가명정보에 대한 접근권한은 분리하여 관리해야 한다.  
 ③ 가명정보 또는 추가정보에 대한 접근권한 부여, 변경 또는 말소에 대한 내역을 기록하도록 하고 이 기록은 최소 3년간 보관하여야 한다.

**제00조(가명정보 및 추가정보의 안전성 확보조치)** ① 가명정보와 추가정보는 법령에서 요구하는 안전성 확보조치를 수행하여야 한다.  
 ② 추가정보는 특별한 이유가 없는 한 생성 즉시 삭제하도록 한다. 단, 시계열 분석 등의 이유로 추가정보가 필요한 경우 저장 시 암호화하여 저장하여야 한다.

**제00조(가명정보를 처리하는 자의 교육)** ① 가명정보 관리책임자는 가명정보를 처리하는 자에게 필요한 가명정보 보호 교육계획을 수립하고 실시하여야 한다.  
 ② 가명정보 보호 교육은 다음과 같은 내용을 포함하여 시행하여야 한다.  
 1. 가명정보 처리에 관한 사항  
 2. 가명정보 및 추가정보의 안전조치에 관한 사항  
 3. 재식별 금지에 관한 사항  
 ③ 가명정보를 처리하는 자에 대한 교육은 개인정보 보호교육과 함께 수행할 수 있으며 교육을 실시한 결과 또는 이를 입증할 수 있는 관련 자료 등을 기록·보관하여야 한다.

**제00조(가명정보 기록의 작성·보관)** ○○(기관명)은 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보의 이용내역, 제3자 제공 시 제공받는 자, 가명정보의 처리기간(법 제28조의4제2항에 따라 처리 기간을 별도로 정한 경우에 한한다)에 대한 기록을 작성하여 보관하여야 하며, 가명정보를 파기한 경우에는 파기한 날부터 3년 이상 보관하여야 한다.

**제00조(개인정보 처리방침 공개)** ① 가명정보 처리와 관련하여 아래와 같은 내용을 개인정보 처리방침에 포함하여 공개하여야 한다.  
 1. 가명정보의 처리 목적  
 2. 가명정보 처리기간  
 3. 가명정보 제3자 제공에 관한 사항(해당 시)  
 4. 가명정보 처리 위탁에 관한 사항(해당 시)  
 5. 처리하는 가명정보의 항목  
 6. 가명정보의 안전성 확보조치에 관한 사항

**제00조(가명정보의 재식별 금지)** ① 가명정보를 처리하는 자의 가명정보에 대한 재식별 행위는 엄격하게 금지한다.  
 ② 가명정보를 처리하는 자는 가명정보를 처리하는 중 특정 개인에 대한 재식별이 발생하는 경우 즉시 처리를 중단하고 이를 가명정보 관리책임자에게 통보한 후 수립된 재식별 시 처리 방안에 따라 즉시 조치하여야 한다.

작성가이드 2

- ◆ 가명정보도 개인정보에 해당하므로, 개인정보처리자는 가명정보 처리업무를 외부에 위탁하는 경우, 보호법 제26조에 따른 위탁 관련 계약문서를 작성하여야 함
- ◆ 또한, 위탁자는 위탁하는 업무의 내용과 가명정보 처리업무를 위탁받아 처리하는 자를 공개하여야 하며,
  - 업무 위탁으로 인하여 가명정보가 분실·도난·유출·위조·변조·훼손 또는 재식별 되지 아니하도록 수탁자를 교육하고, 처리현황 점검 등 수탁자가 가명정보를 안전하게 처리하는지를 감독하여야 함

구분	위탁계약서에 포함되어야 할 사항
위탁업무 수행 목적 외 처리금지	가명정보를 위탁받은 범위 외로 처리하는 것을 금지하는 사항
가명정보의 안전조치 사항	가명정보와 추가정보의 분리 보관, 가명정보와 추가정보에 대한 접근권한 분리, 가명정보에 대한 안전조치 등에 대한 사항
위탁업무의 목적 및 범위	가명정보를 위탁하는 목적과 범위에 대한 사항
재위탁 제한	재위탁 가능한 범위에 대한 사항
관리·감독에 관한 사항	위탁업무와 관련하여 보유하고 있는 개인정보, 가명정보, 추가정보 등에 대한 안전성 확보조치에 관한 관리·감독사항
재식별 금지	가명정보를 제공받거나 처리를 위탁 받은 사업자 등은 다른 정보와 결합을 통해 재식별 시도가 금지됨을 명시
재식별 위험 발생시 통지	가명정보가 재식별 되었거나, 재식별 가능성이 높아지는 상황이 발생한 경우에는 가명정보 처리 중지 및 위탁자에게 통지 의무 명시

**제00조(위탁업무의 목적 및 범위)** “수탁자”는 계약이 정하는 바에 따라 “위탁자”의 ( ) 목적으로 다음과 같은 가명정보 처리 업무 범위 외 처리를 할 수 없다

- 1.
- 2.
- 3.

**제00조(재위탁 제한)** ① “수탁자”는 “위탁자”의 사전 승낙을 얻은 경우를 제외하고, “위탁자”와의 계약상의 권리와 의무의 전부 또는 일부를 제3자에게 양도하거나 재위탁할 수 없다.

② “수탁자”가 재위탁을 위해 제3의 수탁회사를 선임한 경우, “수탁자”는 해당 사실을 계약 체결 7일 이전에 “위탁자”에게 통보하고 협의 및 승인받아야 한다.

**제00조(안전성 확보조치)** “수탁자”는 「개인정보 보호법」 제29조 및 제28조의 4, 같은 법 시행령 제30조 및 「개인정보의 안전성 확보조치 기준」에 따라 개인정보의 안전성 확보에 필요한 기술적·관리적 조치를 취하여야 한다.

**제00조(가명정보의 처리 기록 작성 및 보관)** “수탁자”는 계약기간 동안 가명정보 활용과 관련한 사항은 기록 작성하여 보관하고, 보안을 갖춘 곳에서 수행하여야 한다.

**제00조(가명정보의 처리제한)** ① “수탁자”는 계약기간은 물론 계약 종료 후에도 위탁업무의 목적 및 범위를 넘어 가명정보를 이용하거나 이를 제3자에게 제공 또는 누설하여서는 안 된다.

② “수탁자”는 계약이 해지되거나 또는 계약기간이 만료된 경우 위탁업무와 관련하여 보유하고 있는 가명정보를 「개인정보 보호법」 시행령 제16조 및 「개인정보의 안전성 확보조치 기준」에 따라 즉시 파기하거나 “위탁자”에게 반납하여야 한다.

③ 제2항에 따라 “수탁자”가 가명정보를 파기한 경우 지체없이 “위탁자”에게 그 결과를 통보하고 파기확인서를 제출하여야 한다.

**제00조(수탁자에 대한 관리·감독 등)** ① “위탁자”는 “수탁자”에 대하여 다음 각 호의 사항을 감독할 수 있으며, “수탁자”는 특별한 사유가 없는 한 이에 응하여야 한다.

1. 가명정보의 처리 현황
2. 가명정보의 기술적·관리적 보호조치에 관한 사항 준수 여부
3. 가명정보의 접근 또는 접속기록의 보관·기록 여부
4. 가명정보의 접근통제 및 가명정보취급자 관리·감독 여부
5. 목적 외 이용·제공 및 재식별 금지 준수 여부
6. 재위탁 가능범위 준수 여부
7. 재식별 위험 발생 시 가명정보 처리 중지 및 통지 의무 준수 여부
8. 그 밖에 가명정보의 보호를 위하여 필요한 사항

② “위탁자”는 “수탁자”에 대하여 제1항 각 호의 사항에 대한 실태를 점검하여 시정을 요구할 수 있으며, “수탁자”는 특별한 사유가 없는 한 이행하여야 한다.

③ “위탁자”는 처리위탁으로 인하여 정보주체의 가명정보가 분실·도난·유출·변조 또는 훼손되지 아니하도록 「개인정보의 안전성 확보조치 기준」에 따라 1년에 1회 이상 “수탁자”를 교육할 수 있으며, “수탁자”는 이에 응하여야 한다.

④ 제1항에 따른 교육의 시기와 방법 등에 대해서는 “위탁자”는 “수탁자”와 협의하여 진행한다.

**제00조(재식별 금지)** ① “수탁자”는 “위탁자”로부터 제공받은 가명정보를 안전하게 처리하고, 이를 이용해서 개인을 재식별하기 위한 어떠한 행위도 하여서는 아니 된다.

② 재위탁을 하고자 하는 경우에도 “수탁자”는 재식별 방지를 위해 필요한 조치를 하여야 한다.

③ “수탁자”는 “위탁자”로부터 제공받은 정보가 재식별되거나 재식별 가능성이 현저하게 높아지는 상황이 발생하면 즉시 해당 정보의 처리를 중단하고 관련 사항을 “위탁자”에게 알리며, 필요한 협조를 하여야 한다.

④ “수탁자”는 제1항에서 제3항까지의 사항을 이행하지 않아 발생하는 모든 결과에 대해 형사 및 민사상 책임을 진다.

**제00조(가명정보의 파기)** ① “수탁자”는 제4조의 위탁업무기간이 종료되면 특별한 사유가 없는 한 지체 없이 가명정보를 파기하고 파기확인서를 “위탁자”에게 제출하여야 한다.

② “위탁자”는 파기 확인을 위해 “수탁자”에게 실제 확인 여부를 요청할 수 있으며, “수탁자”는 이에 적극적으로 협력(파기전문 프로그램 사용기록, 파기업무 수행기록(로그)의 공동 확인 등)하여야 한다.

**제00조(손해배상)** ① “수탁자” 및 “수탁자”의 수탁자가 이 계약에 의하여 위탁 또는 재위탁받은 정보를 취급함에 있어 이 계약에 따른 의무를 위반하거나 “수탁자” 또는 “수탁자”의 수탁자의 귀책사유로 인하여 재식별되어 “위탁자” 또는 개인정보주체 기타 제3자에게 손해가 발생한 경우 “수탁자”는 그 손해에 대하여 배상하여야 한다.

② 제1항과 관련하여 개인정보주체 기타 제3자에게 발생한 손해에 대하여 “위탁자”가 전부 또는 일부를 배상한 경우 “위탁자”는 이를 “수탁자”에게 구상할 수 있다.

작성가이드 3

- ◆ 개인정보처리자는 가명정보 처리와 관련하여 아래와 같은 내용을 개인정보 처리방침에 포함하여 공개하여야 함

※ 다만, 개인정보의 처리에 대하여 기 작성한 개인정보 처리방침이 있을 경우, 가명정보 처리에 관한 내용만 추가 가능



## 〈개인정보 처리방침 작성예시(일반형)〉

### (가명정보의 처리)

〈개인정보처리자명〉은(는) 수집한 개인정보를 「개인정보 보호법」 제28조의2에 따라 통계작성, 과학적 연구, 공익적 기록 보존 등을 위하여 특정 개인을 알아볼 수 없도록 가명처리하여 다음과 같이 활용하고 있습니다.

#### ▶ 가명정보의 이용에 관한 사항

구분	이용목적	이용항목	이용기간
과학적연구	○○알고리즘 성능 개선을 위한 □□와 △△간 연관성에 대한 과학적 연구	'25년 고객 구매데이터	결합데이터 분석 완료시까지

#### ▶ (해당 시) 가명정보의 제3자 제공에 관한 사항

제공받는 자	제공받는자의 이용목적	제공하는 항목	제공받는 자의 보유 및 이용기간
제공받는 자의 법인명 또는 명칭	○○알고리즘 성능 개선을 위한 □□와 △△간 연관성에 대한 과학적 연구	'25년 고객 구매데이터	연구목적 달성시까지

#### ▶ (해당 시) 가명정보 처리의 위탁에 관한 사항

위탁받는 자	위탁업무
수탁자 법인명 또는 명칭	○○와 △△간 상관관계 분석을 위한 통계작성
수탁자 법인명 또는 명칭	○○연구를 위한 가명처리

#### ▶ 가명정보의 안전성 확보조치에 관한 사항



## 〈 개인정보 처리방침 작성예시(통합형) 〉

### (가명정보의 처리)

〈개인정보처리자명〉은(는) 수집한 개인정보를 「개인정보 보호법」 제28조의2에 따라 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 특정 개인을 알아볼 수 없도록 가명처리하여 다음과 같이 활용하고 있습니다.

#### ▶ (해당 시) 가명정보의 이용·제공에 관한 사항

구분	제공 받는 자	이용·제공 목적	이용·제공 항목	이용·제공 기간
이용	-	OO지역 연령대별 이동패턴 및 주요 관심사항 분석 (통계 작성)	22년~24년 고객 이동정보 및 관심목록 정보	결합데이터 분석 완료시까지
제공	OOO(주)	OO데이터와 OO 사용자의 관심지수를 활용한 OOt렌드 분석 (과학적 연구)	22년7월 ~ 23년6월 고객 구매데이터	목적 달성 후 1개월 내

#### ▶ (해당 시) 가명정보 처리의 위탁에 관한 사항

위탁받는 자	위탁업무
수탁자 법인명 또는 명칭	OO와 △△간 상관관계 분석을 위한 통계작성
수탁자 법인명 또는 명칭	OO연구를 위한 가명처리

#### ▶ 가명정보의 안전성 확보조치에 관한 사항

## 결합 신청서

신청번호	
접수번호	

## ① 결합신청자

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

## ② 결합 개요

전체 결합신청기관명	총 기관 수	[ ] 개 기관
	기관명	
반복결합	[ ] 해당없음 [ ] 최초 [ ] 추가(결합접수번호 : )	
추가절차 신청	결합을 확인 [ ] 가명정보 추출 [ ] 모의결합 [ ]	

## ③ 가명정보 제공자

해당없음 [ ]

제출 방법	[ ] 온라인 [ ] 오프라인	
지원 요청 사항	[ ] 결합 신청에 필요한 가명처리	
가명정보 제공 담당자	이름	연락처 (전화, e-mail)

## ④ 결합정보 이용자

해당없음 [ ]

결합 목적	[ ] 통계작성 [ ] 과학적 연구 [ ] 공익적 기록보존 등		
세부 결합 목적			
분석공간 이용	[ ] 추가 가명처리 [ ] 결합정보 분석 [ ] 이용안함		
지원 요청 사항	[ ] 반출 전 처리 [ ] 분석		
결합정보 이용 담당자	이름	연락처 (전화, e-mail)	

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

년 월 일

## ⑤ 결합신청기관

(서명 또는 인)

(결합전문기관명)의 장 귀하

첨부 서류	<ol style="list-style-type: none"> <li>1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부</li> <li>2. 결합 목적을 증명할 수 있는 서류 1부(결합정보 이용자에 한함)</li> <li>3. 결합 대상 가명정보에 관한 서류(전체 항목명, 가명처리 대상 항목명*, 가명처리 내역 등**) 1부(가명정보 제공자에 한함) <ul style="list-style-type: none"> <li>* 결합키 생성에 사용된 항목 제외</li> <li>** 결합 대상 정보가 확정된 이후에 제출</li> </ul> </li> </ol>
----------	---

### ① 결합신청자

- 기관명, 사업자등록번호 또는 법인등록번호, 대표자명 : 결합신청자가 개인인 경우 공란 허용
- 담당자: 결합전문기관과 협의를 담당하는 자\*
- \* 예시) 가명정보 제공 : 데이터 보유부서 책임자, 결합정보 이용 : 연구 책임자
- ※ 결합신청서의 '담당자'와 반출신청서의 '담당자'는 다른 사람으로 작성해도 무관
- \* 예시) '가명정보 제공 + 결합정보 이용'의 경우 결합신청서에는 데이터 보유부서 책임자를, 반출신청서에는 연구책임자를 담당자로 작성 가능

### ② 결합 개요

- 전체 가명정보 제공자명(총수): 해당 결합을 신청하는 가명정보 제공자의 전체 기관명 및 전체 기관 수(총 00개)를 작성
- ※ 결합신청자간 서로 협의한 내용이 정확한지 확인 및 해당 결합 건의 전체 결합신청자 수를 파악하기 위한 목적 이므로, 기관이 너무 많은 경우 대표기관명만 나열 후 전체 기관수를 기재할 수 있음
- 반복결합: 추가일 경우 결합접수번호가 누락되지 않도록 주의
- 추가절차 신청: 모의결합을 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내하거나 모의결합 진행 없이 결합으로 진행할 수 있도록 안내

### ③ 가명정보 제공자

- 지원 요청 사항: '결합 전 가명처리'를 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내하거나 컨설팅 안내 등 추가적으로 협의를 통해 결합을 진행할 수 있도록 안내
- 가명정보 제공 담당자: 데이터 보유부서 책임자
- ※ 전체 담당자와 동일한 경우 공란 허용하며, '가명정보 제공 + 결합정보 이용'의 경우 데이터 보유부서 책임자 및 연구 책임자를 분리하여 작성 필요

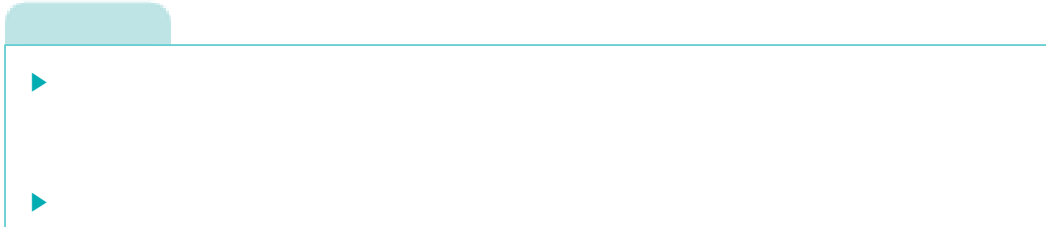
### ④ 결합정보 이용자

- 세부 결합 목적: 결합 목적 및 필요 정보를 알 수 있을 정도로 구체적으로 작성
- 분석공간 이용: 분석공간 이용기간에 대해서 결합 후 별도로 협의하여 이용할 수 있도록 안내

- 지원요청사항: '반출 전 처리', '분석'을 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내 하거나 컨설팅 안내 등 추가적으로 협의를 통해 결합을 진행할 수 있도록 안내
- 결합정보 이용 담당자: 연구 책임자
  - ※ 전체 담당자와 동일한 경우 공란 허용하며, '가명정보 제공 + 결합정보 이용'의 경우 데이터 보유부서 책임자 및 연구 책임자를 분리하여 작성 필요

#### ⑤ 결합신청자의 서명

- 결합신청자가 개인이 아닌 경우 개인정보처리자 직인이 원칙, 다만 서면(직인 날인) 외에 전자문서로 제출할 수 있음을 안내



작성가이드 5

① 반출신청서

②

반출접수번호	
결합접수번호	

결합신청자			
기관명		사업자등록번호 또는 법인등록번호	
주소		대표자명	
담당자		담당자 연락처 (전화, e-mail)	

③ 결합 유형

반복결합	<input type="checkbox"/> 최초	<input type="checkbox"/> 추가
------	-----------------------------	-----------------------------

④ 반출 개요

파일명			
반출 목적	<input type="checkbox"/> 통계작성	<input type="checkbox"/> 과학적 연구	<input type="checkbox"/> 공익적 기록보존 등
세부 반출 목적			
반출 정보 유형	<input type="checkbox"/> 가명정보	<input type="checkbox"/> 법 제58조의2에 해당하는 정보(익명정보)	
제공 받는 방법	<input type="checkbox"/> 온라인	<input type="checkbox"/> 오프라인	<input type="checkbox"/> 결합전문기관 내 분석공간
지원 요청 사항	<input type="checkbox"/> 반출된 정보의 분석	<input type="checkbox"/> 개인정보 보호 교육	

「개인정보 보호법」 제28조의3제2항 및 같은 법 시행령 제29조의3제3항·제6항, 「가명정보의 결합 및 반출 등에 관한 고시」 제10조제3항에 따라 결합된 정보를 반출하기 위하여 결합전문기관에 반출신청서를 위와 같이 제출합니다.

년 월 일

결합신청자

(서명 또는 인)

결합전문기관의 장 귀하

⑤ 첨부 서류	1. 반출 대상 정보에 관한 서류 1부(추가적인 서류 제출이 필요한 경우에 한함) 2. 반출 목적을 증명할 수 있는 서류 1부(추가적인 서류 제출이 필요한 경우에 한함) 3. 반출 정보의 안전조치계획 및 이를 증빙할 수 있는 서류 1부
---------	---

**① 반출신청서 작성**

- (작성 주체) 결합정보 또는 분석결과 등을 결합전문기관 외부로 반출하고자하는 자는 반출신청서를 작성하여 결합전문기관에 제출
- ※ 가명정보를 제공만 하는 자와 결합정보를 결합전문기관 내의 분석공간에서 분석만을 수행하는 자는 반출신청서를 작성하지 않아도 됨

**② 반출접수번호 및 결합접수번호**

- (반출접수번호) 결합전문기관이 제출된 반출신청서를 접수할 때 발행하는 번호, 반출신청서 제출 시 공란으로 제출
- (결합접수번호) 결합전문기관이 제출된 결합신청서를 접수할 때 발행하였던 번호

**③ 결합 유형**

- 반복결합을 신청한 자는 최초/추가 여부를 체크하며, 반복결합이 아닌 경우 공란

**④ 반출 개요**

- 결합정보를 반출하려는 결합신청자는 파일명(반출할 결합 결과물), 반출 목적, 반출정보 유형 등을 작성
- ※ 결합신청자는 반출심사 전인 경우 결합 목적과 반출 목적 변경 가능
- (반출정보 유형) 반출을 신청하려는 자가 정보의 형태 등을 고려하여 가명 또는 익명으로 판단하여 표기
- (제공받는 방법) 결합전문기관등이 제공하는 시스템을 통해 제공받는 경우는 온라인, USB등 저장장치를 이용해 반출하는 경우는 오프라인, 결합전문기관이 제공하는 분석공간을 이용하는 경우로 구분하여 표기
- (지원 요청사항) 정보의 분석을 위해 결합전문기관의 지원이나 가명정보의 처리에 관한 교육이 필요하면 표기

**⑤ 첨부 서류**

- 추가적인 서류 제출이 필요한 경우에 한하여 모든 서류를 제출
- (반출 대상 정보에 관한 서류) 분석공간을 통해 추가 가명처리가 수행되어 반출 대상 정보가 당초 제출한 결합 대상 가명정보와 상이한 경우에만 제출하고 동일한 경우에는 생략 가능
- (반출 목적 관련 서류) 반출 목적이 당초의 결합 목적과 달라진 경우(결합 목적과 반출 목적의 양립 가능성 검토 필요)만 제출하고 동일한 경우에는 생략 가능
- (안전조치 계획) 개인정보 처리방침, 내부관리 계획, 운영 지침 등 반출정보의 안전조치와 관련된 자료를 제출

# 7

## 가명정보 활용 위험도 판단 예시

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제7장 가명정보 활용 위험도 판단 예시

### 1

#### ① 가명정보 반복제공 사례



#### ◆ 개요

OO대학교 지역개발연구소가 2024년 △△공공기관의 농업경영체 등록정보를 가명정보 형태로 활용하여 지역별 농업인력 수급 현황을 분석하고, 2025년에도 동일한 항목과 처리수준을 유지한 데이터를 활용하여 전년도 결과와 비교·분석함으로써 농업인력 구조 변화의 시계열 추세를 도출

※ 이용기관, 제공기관, 활용목적이 일관된 형태로 매년 반복 수행되며 외부 폐쇄망 환경에서 분석 예정

#### ◆ 처리 목적

2024년 및 2025년 연도별 농업경영체 등록정보를 분석하여 농가 고령화율, 경영형태, 인력 수급 변화를 시계열 통계로 산출하고, 연차별 비교를 통해 지역 농업인력 정책의 변화 추세와 지원사업 효과를 평가하기 위한 통계작성 수행

◆ 데이터 흐름



◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	○○대학교 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	○○대학교 (가명정보 이용자)	작성	기존 계획서 그대로 활용
(2단계) 위험성 검토	서식③ 위험성 검토서	△△공공기관 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서		생략	기존 문서와 동일하므로 그대로 활용
	서식⑤ 가명처리 결과서		생략	기존 문서와 동일하므로 그대로 활용
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서		생략	기존 내용과 동일하므로 그대로 활용
	서식⑧ 적정성 평가위원 서약서		해당 없음	담당자 검토사항
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	이전 작성된 확약서로 같음
	서식⑩ 가명(추가)정보 관리대장	○○대학교 (가명정보 이용자), △△공공기관 (가명정보 제공기관)	작성	가명정보 제공 및 활용 과정에서 기록·관리 필요



〈서식②〉

가명정보 활용 계획서

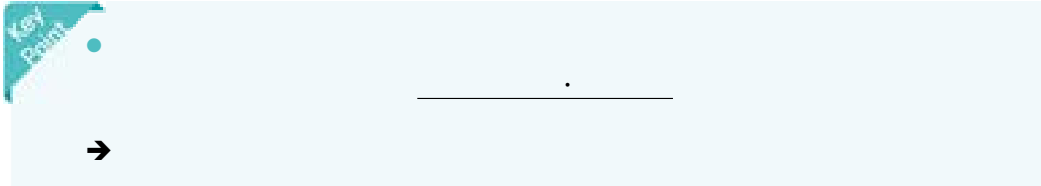
<b>신청 제목</b>	농업경영체 등록정보 기반 농업인력 수급 연차별 시계열 통계 분석
<b>책임자</b>	소속: OO대학교 지역개발연구소 농촌정책연구팀    성명: 김 민 수    (서명)
<b>배경 및 목적</b>	<p>농업인력의 고령화와 인력 수급 불균형이 지속됨에 따라, 지역별 농업경영체 구조 변화를 장기적으로 모니터링할 수 있는 시계열 통계의 필요성이 증대됨. 기존 표본조사 중심의 통계는 세부 지역 단위 비교가 어려워, 농업경영체 등록정보를 기반으로 동일 항목을 분석함으로써 정책적 대응 기반을 강화하고, 연도 간 변화를 정량적으로 검증하기 위함</p> <p>2024년도 농업경영체 등록정보를 활용하여 농가의 고령화율, 경영형태, 인력 수급현황 등의 변화를 분석하고, 전년 대비 농업인력 구조 및 경영 형태의 특성을 시계열로 파악함</p> <p>분석 결과를 토대로 지역별 농업인력 분포, 고령화 추세, 경영규모 변화 등의 주요 통계지표를 작성하여 2025년도 농업정책 수립 및 인력지원사업의 효과성 평가를 위한 근거자료로 활용</p>
<b>활용 데이터</b>	<p>2024년 1~12월 농업경영체 등록정보</p> <p>- 농업경영체 식별번호, 경영주 연령대, 성별, 경영형태, 경작면적, 인력 수급현황, 지역코드</p>
<b>활용 계획 및 방법</b>	<ol style="list-style-type: none"> <li>① △△공공기관이 보유한 2024년도 농업경영체 등록정보를 가명처리 후, OO대학교 내 폐쇄형 분석환경에서 지역개발연구소가 분석 수행</li> <li>② 동일 항목·동일 구조의 데이터를 기반으로 통계량 산출 및 전년 대비 비교 분석 진행</li> <li>③ 시계열 분석(변화율·추세선 도출) 및 군집분석을 통해 농가 유형별 인력 구조 변동 파악</li> <li>④ 분석 결과를 통계지표로 정리하여 정책평가 및 통계작성 근거자료로 활용</li> </ol>
<b>(예상) 활용 기간</b>	2025년 7월~ 2026년 6월까지 (12개월)
<b>기대효과 및 활용방안</b>	<p>전년(2024) 기준의 농업경영체 정보를 활용해 지역별 농업인력 구조를 정량적으로 파악하고, 농촌 고령화와 인력 수급 추세를 분석함으로써 농업정책의 실효성을 평가하는 근거 통계로 활용</p> <p>향후 연차별 누적 분석을 통해 농업인력 지원정책의 방향성과 우선순위 설정에 기여</p>

## 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요	
가명정보 활용 목적	농업경영체 등록정보 기반 농업인력 수급 연차별 시계열 통계 분석
가명처리 대상 개인정보	2024년 1~12월 농업경영체 등록정보
활용하려는 자	(기관명) OO대학교 지역개발연구소 농촌정책연구팀 (직위) 수석 (성명) 김민수
(1) 기본 위험도 판단	
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
기본 위험도 자동 도출	〈Step 1〉 내부 활용 → <input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(OO대학교)에 제공하며, 본 기관에서 통제할 수 없는 환경인 OO대학교 내의 분석환경에서 활용될 예정임
(2) 사례별 위험도 조정	
반복·유사 활용 건	<input checked="" type="checkbox"/> 해당 <input type="checkbox"/> 해당 없음    ※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음    ※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음    ※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정	
조정	<input type="checkbox"/> 유지 <input type="checkbox"/> 상향 <input checked="" type="checkbox"/> 하향
최종 위험도	<input checked="" type="checkbox"/> 저위험 <input type="checkbox"/> 중위험 <input type="checkbox"/> 고위험
종합 검토 결과	기존 접수되어 검토하여 반출하였던 건(NC-20250522001)과 활용 목적, 방식, 항목, 처리방법 및 수준이 동일한 반복제공 건이며, 폐쇄망 분석환경을 이용하는 등 안전조치가 확보되었으므로 저위험으로 판단하며, 적정성에 대한 검토는 담당자 검토로 처리함
검토일 : 2025년 5월 23일	
검토자 소속 : △△공공기관 데이터정보화과    성명 : 김능업 (서명 또는 인)	

## ② 가명정보 유사제공 사례 (연구자 추가·변경)



### ◆ 개요

OO공공기관은 학생 학업성과와 교육지원 프로그램 간의 연관성을 규명하기 위한 학습격차 연구를 추진함에 따라, △△교육청이 보유한 2024년도 학업성취도정보를 동일 항목·동일 처리수준으로 요청하여 연구 분석에 활용

이는 2023년에 □□연구소에서 수행된 전국 단위 학업성취도 분석 결과를 지역 단위(△△도)로 확장·심화하는 후속 연구로, 데이터 구조는 동일하나 연구 주체가 변경되어 연속적 연구 기반을 구축하기 위함

※ 기존에 적정성 검토하여 제공했었던 것과 비교하여, 이용기관만 **기존 □□연구소에서 OO공공기관으로 변경**. 나머지 목적·환경·데이터 항목 등은 동일

### ◆ 처리 목적

2024년도 학업성취도정보를 활용하여 교과별 성취 수준에 영향을 미치는 가정환경, 학습지원 프로그램, 학교요인 등을 통합 분석. 기존 연구의 성취도 비교분석을 확장하여, 학습격차 발생 요인을 규명하고 정책적 대응방안을 제시하기 위한 교육 연구 목적으로 활용

### ◆ 데이터 흐름



### 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	OO공공기관 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	OO공공기관 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	△△교육청 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서		생략	기존 문서와 동일하므로 그대로 활용
	서식⑤ 가명처리 결과서		생략	기존 문서와 동일하므로 그대로 활용
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서		생략	변경된 사항만 간단히 검토 후 기록
	서식⑧ 적정성 평가위원 서약서		해당 없음	담당자 검토사항
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		작성	변경된 이용자에게 확약 요청 필요
	서식⑩ 가명(추가)정보 관리대장	OO공공기관 (가명정보 이용자),  △△교육청 (가명정보 제공기관)	작성	가명정보 제공 및 활용 과정에서 기록·관리 필요

주요 서식 작성 예시

<서식 1>

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	GG-20250120002	기존 접수번호(변경시)		접수일	
신청자	조직/부서명	○○공공기관 빅데이터분석부			
	담당자 직위	주무관	담당자 성명	김 교육	
	전화번호	000-000-0000	이메일 주소	education@email.com	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 △△교육청이 보유한 2024년도 학업성취도 데이터를 가명정보 형태로 제공받아, 학생의 교과별 성취 수준에 영향을 미치는 가정환경·학교요인·학습지원 프로그램 등의 상관관계를 분석하고 교육격차 완화정책 수립을 위한 연구 수행				
가명처리 대상 개인정보	개인정보 명칭	2024년 1월~12월 학업성취도 정보, 출결이력 table, 상담이력 table, 학습지원 프로그램 참여정보, 학교 기본정보			
	요청 개인정보 항목	<ul style="list-style-type: none"> <li>▪ 학업성취도 정보(학생식별번호, 학년, 학교코드, 교과명, 과목코드, 평가점수, 성취등급, 학업성취도 총점)</li> <li>▪ 출결이력 table(학생식별번호, 결석일수, 지각일수, 출석률, 학기구분)</li> <li>▪ 상담이력 table(학생식별번호, 상담일자, 상담유형, 상담횟수, 상담결과 요약)</li> <li>▪ 학습지원 프로그램 참여정보(학생식별번호, 프로그램명, 참여기간, 참여횟수, 프로그램 유형)</li> <li>▪ 학교 기본정보(학교코드, 지역코드, 설립유형, 학생수, 교원수)</li> </ul>			
결합 활용 여부	<input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input type="checkbox"/> 기관 내부 분석 공간 <input checked="" type="checkbox"/> 기관 외부 반출				
	<input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	반출일로부터 ( 10 )개월 이내 또는 반출일로부터 ~ 2025 년 12 월 31 일 까지				
「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.					
2025 년 1 월 20 일 신청인 (부서장명) 김 교육 (서명 또는 인)					
△△ 교육청장 귀중					
첨부서류	1. 가명정보 활용 계획서				

## 〈서식②〉

## 가명정보 활용 계획서

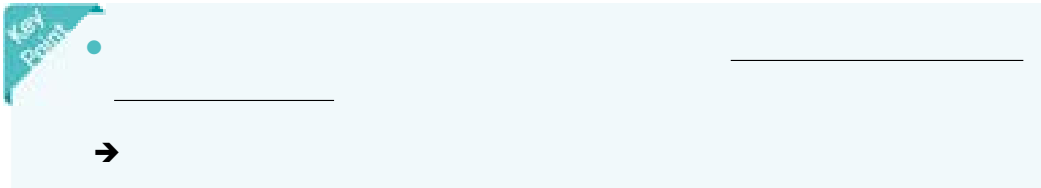
신청 제목	△△교육청 학업성취도정보를 활용한 학습격차 요인 분석 연구
책임자	소속: ○○공공기관 빅데이터분석부    성명: 김 교 육    (서명)
배경 및 목적	<p>최근 학생 간 학업 성취 격차가 지역·학교·가정환경별로 심화되고 있으며, 단순 성취도 통계로는 원인 분석에 한계가 존재함. 이에 ○○공공기관은 2023년에 □□연구소에서 수행한 전국 단위 학업성취도 분석을 바탕으로, 2025년에는 △△교육청이 보유한 동일 구조의 데이터를 활용하여 지역 단위 학습격차 요인을 심층 분석하고, 정책 기반의 연구체계를 구축하고자 함</p> <p>△△교육청이 보유한 2024년도 학업성취도 가명정보를 활용하여 교과 성취 수준에 영향을 미치는 가정환경, 학교요인, 학습지원 프로그램 등 주요 요인을 분석함. 기존 연구의 성취도 비교 분석 범위를 확장하여 학습격차 발생 원인을 규명하고, 교육정책 수립에 기초자료로 활용하기 위한 교육 연구 수행</p>
활용 데이터	<ul style="list-style-type: none"> <li>- 학업성취도 정보: 학생식별번호, 학년, 학교코드, 교과명, 과목코드, 평가점수, 성취등급, 학업성취도 총점</li> <li>- 출결이력 table: 학생식별번호, 결석일수, 지각일수, 출석률, 학기구분</li> <li>- 상담이력 table: 학생식별번호, 상담일자, 상담유형, 상담횟수, 상담결과 요약</li> <li>- 학습지원 프로그램 참여정보: 학생식별번호, 프로그램명, 참여기간, 참여횟수, 프로그램 유형</li> <li>- 학교 기본정보: 학교코드, 지역코드, 설립유형, 학생수, 교원수</li> </ul>
활용 계획 및 방법	<ol style="list-style-type: none"> <li>① △△교육청 On-premises 서버에서 2024년도 학업성취도 데이터를 가명처리</li> <li>② 가명처리된 데이터는 ○○공공기관 폐쇄망 환경에서 분석 수행</li> <li>③ 데이터 전처리 및 품질 검증 후, 학년·학교급·지역 단위로 표준화된 분석용 데이터셋 구축</li> <li>④ 기술통계와 상관분석을 통해 학업성취도 수준 분포 및 주요 변수 간 상관관계 파악</li> <li>⑤ 회귀분석·요인분석·경로분석(SEM) 등을 적용하여 교과 성취 수준에 영향을 미치는 가정·학교·지원요인 간 인과 구조 규명</li> <li>⑥ 지역·학교유형별 차이를 분석하기 위해 다층모형(HLM) 을 활용하여 교육격차의 구조적 요인 파악</li> </ol>
(예상) 활용 기간	2025년 3월~ 2025년 12월까지 (10개월)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>- 학업성취도 기반의 정량적 분석을 통해 지역별·학교별 학습격차 원인을 규명</li> <li>- 교육지원 프로그램의 효과성을 검증하고, 맞춤형 학습지원 정책 수립의 근거자료로 활용</li> <li>- 동일 구조 데이터의 연차별 활용을 통해 시계열 교육 연구 기반 구축 및 정책 지속성 확보</li> </ul>

〈서식③〉

위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	△△교육청 학업성취도정보를 활용한 학습격차 요인 분석 연구		
가명처리 대상 개인정보	2024년 1월~12월 학업성취도 정보, 출결이력 table, 상담이력 table, 학습지원 프로그램 참여정보, 학교 기본정보		
활용하려는 자	(기관명) ○○공공기관 빅데이터분석부 (직위) 주무관 (성명) 김교육		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(○○공공기관)에 제공하며, 본 기관에서 통제할 수 없는 환경인 ○○공공기관 내의 분석환경에서 활용될 예정임		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input checked="" type="checkbox"/> 하향
최종 위험도	<input checked="" type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input type="checkbox"/> 고위험
종합 검토 결과	기존 접수되어 검토하여 반출하였던 건(GG-20240120088)과 활용 목적, 방식, 항목, 처리방법 및 수준이 동일하고 활용기관만 다른 유사 활용 건이며, 폐쇄망 분석환경을 이용하는 등 안전조치 또한 기존 제공건과 크게 다르지 않아 저위험으로 판단함. 따라서 적정성에 대한 검토는 담당자 검토로 처리함		
검토일 : 2025년 1월 20일			
검토자 소속 : △△교육청 교육정보화과    성명 : 박재만 (서명 또는 인)			

### ③ 내부 부서 간 가명정보 제공·활용 사례



#### ◆ 개요

○○과학원은 국민환경보건기초조사를 통해 수집한 조사참여자의 건강검진 결과(혈압, 폐활량 등)와 같은 기간의 대기오염 측정자료(PM2.5, NO<sub>2</sub>, SO<sub>2</sub> 등)를 연계하여, 대기오염 노출 수준과 개인 건강지표 간의 관계를 규명하고, 장기적 노출이 건강에 미치는 영향을 분석하는 연구를 수행하고자 함

#### ◆ 처리 목적

국민환경보건기초조사를 통해 수집된 참여자의 건강검진 결과(혈압, 폐활량 등)와 동일 기간의 대기오염 측정자료(PM2.5, NO<sub>2</sub>, SO<sub>2</sub> 등)를 연계하여 거주지·근무지·체류시간을 반영한 개인별 대기오염 노출 수준을 산정하고, 오염물질 노출이 혈압·폐기능 등 주요 건강지표에 미치는 영향을 시계열적으로 분석함

이를 통해 생활권 특성과 노출시간에 따른 건강영향 차이를 규명하고, 환경보건정책 및 취약계층 건강 보호 대책 수립의 근거자료로 활용

#### ◆ 데이터 흐름



❖ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서		생략	내부 공문으로 대체
	서식② 가명정보 활용 계획서	기획정책과 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	환경연구과 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	환경연구과 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	환경연구과 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	환경연구과 (가명정보 제공자)	작성	
	서식⑧ 적정성 평가위원 서약서		해당 없음	담당자 검토사항
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	별도 협약 없이 내부 지침에 따라 관리
	서식⑩ 가명(추가)정보 관리대장	기획정책과 (가명정보 이용자)  환경연구과 (가명정보 제공자)  ※ 기관 내 개인정보 보호 부서에서 총괄 작성·관리하는 것도 가능	작성	가명정보 제공 및 활용 과정에서 기록·관리 필요

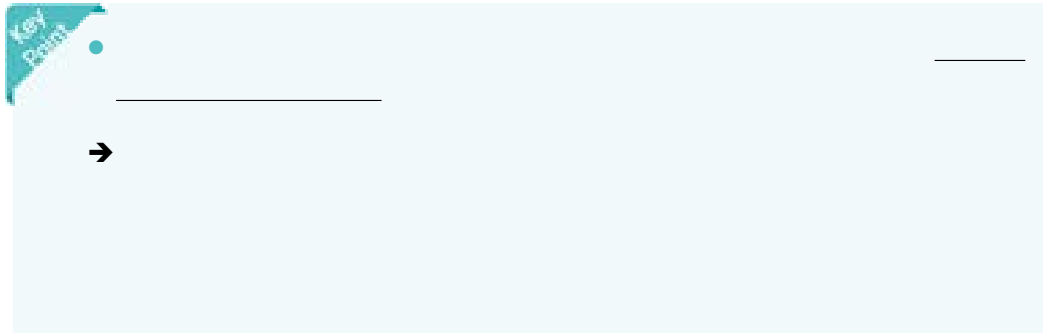


〈서식③〉

위험성 검토서

◆ 가명정보 활용 개요		
가명정보 활용 목적	2020~2024년 대기오염 노출수준 기반 개인 건강지표 시계열 분석 및 정책활용 연구	
가명처리 대상 개인정보	2020년 1월 ~ 2024년 12월 국민환경보건기초조사 참여자 건강검진 및 생활패턴 정보	
활용하려는 자	(기관명) OO과학원 기획정책과 (직위) 책임 (성명) 손청명	
(1) 기본 위험도 판단		
〈Step 1〉 활용 주체	<input checked="" type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등	
	<input type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공	
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우	
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우	
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→ <input checked="" type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→ <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→ <input type="checkbox"/> 고위험
	※ 판단 이유 : 기관 내 타 부서에서 활용하는 내부 활용 사례임	
(2) 사례별 위험도 조정		
반복·유사 활용 건	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정		
조정	<input checked="" type="checkbox"/> 유지 <input type="checkbox"/> 상향 <input type="checkbox"/> 하향	
최종 위험도	<input checked="" type="checkbox"/> 저위험 <input type="checkbox"/> 중위험 <input type="checkbox"/> 고위험	
종합 검토 결과	내부 기관에서 안전하게 활용될 예정이고, 데이터 자체의 위험도도 높지 않은 편이며, 외부 기관에 제공하지 않는 내부 활용 건은 저위험으로 분류하는 「OO과학원 내부 가명정보 제공·활용 지침」에 따라 저위험으로 판단함. 따라서 적정성에 대한 검토는 담당자 검토로 처리함	
검토일 : 2025 년 4 월 12 일 검토자 소속 : OO과학원 환경연구과    성명 : 홍민성 (서명 또는 인)		

## ④ 결함을 위해 결합전문기관에 제공하는 사례



### ◆ 개요

△△기관은 보유한 국민체력검증사업 참여자 데이터를 가명처리하여 ○○공단의 보험급여지출정보와 결합하고, 체력수준·참여빈도·성별·연령 등 체력 관련 요인과 진료과·상병코드 등 의료비 지출 항목 간의 관계를 분석함으로써 체력 향상이 질병 발생률과 의료비 절감에 미치는 효과를 검증하기 위한 연구를 수행함

결합은 결합전문기관인 ○○공단에서 수행하며, 분석은 결합전문기관 내 폐쇄망 분석환경에서 진행

### ◆ 처리 목적

국민체력검증사업 참여자의 체력수준 변화와 국민건강보험 급여지출 내역을 연계·분석하여, 체력 향상이 질병 발생률과 의료비 지출 감소에 미치는 영향을 파악하고, 연령·성별·참여횟수 등 개인 특성별로 건강증진 효과의 차이를 규명하며, 예방 중심 보건정책 수립의 근거를 마련함

### ◆ 데이터 흐름



❖ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서		생략	기관 내 타부서 제공이므로 내부 공문으로 대체
	서식② 가명정보 활용 계획서	△△기관 국민체력사업팀 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	△△기관 디지털서비스팀 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	△△기관 디지털서비스팀 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	△△기관 디지털서비스팀 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	△△기관 디지털서비스팀 (가명정보 제공자)	작성	
	서식⑧ 적정성 평가위원 서약서		해당 없음	담당자 검토사항
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	결합전문기관 관리·통제 하에 활용되므로 생략
	서식⑩ 가명(추가)정보 관리대장	△△기관 디지털서비스팀 (가명정보 제공자)  ※ 기관 내 개인정보 보호 부서에서 총괄 작성·관리하는 것도 가능	작성	가명정보 활용 과정에서의 관리대장은 결합전문기관 관리·통제 하에 활용되므로 작성 불필요



〈서식③〉

위험성 검토서

◆ 가명정보 활용 개요	
가명정보 활용 목적	국민체력검증사업 참여자 체력정보를 활용한 건강증진 및 의료비 절감 효과 분석
가명처리 대상 개인정보	2020~2024년 국민체력검증사업DB
활용하려는 자	(기관명) △△기관 국민체력사업팀 (직위) 과장 (성명) 박강한

(1) 기본 위험도 판단	
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
기본 위험도 자동 도출	〈Step 1〉 내부 활용 → <input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 기관 내 타 부서가 결합정보를 활용하는 건이나, 본 기관이 제공하는 가명정보가 타 개인정보처리자의 가명정보와 결합되고 내부 기관에서 통제 불가능한 외부 환경에서 분석될 예정임

(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장

(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input checked="" type="checkbox"/> 하향
최종 위험도	<input checked="" type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input type="checkbox"/> 고위험
종합 검토 결과	정부가 이미 처리환경의 안전성을 검증하여 지정한 결합전문기관에서 결합될 예정이며, 결합전문기관이 결합 후 가명처리 적정성에 대한 반출심사를 추가로 수행하게 되므로 저위험으로 판단하여 담당자 검토로 진행하도록 함.		
검토일 : 2025년 6월 27일			
검토자 소속 : △△기관 디지털서비스팀    성명 : 오상민 (서명 또는 인)			

## ⑤ ‘개인정보 이노베이션 존’에 가명정보를 제공하는 사례



### ◆ 개요

본 연구는 A부동산연구원이 임차인 보호 정책 개선 연구를 위해 B지자체로부터 임대차 계약 현황 데이터를 OO센터에서 운영하는 ‘개인정보 이노베이션 존’에서 활용하는 건

### ◆ 처리 목적

보증금·월세·증액보증금·전세/월세 구분·확정일자·전입일 등 계약 항목을 활용해 금액 수준별로 계약이 어떤 특성을 갖는지 패턴을 분석하고, 금액대별 임차인 보호장치(전입·확정일자 등) 차이를 파악하여 정책 개선 기초자료로 활용

### ◆ 데이터 흐름



❖ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	A부동산연구원 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	A부동산연구원 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	B지자체 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	B지자체 (가명정보 제공기관)	작성	
	서식⑤ 가명처리 결과서	B지자체 (가명정보 제공기관)	작성	
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	B지자체 (가명정보 제공기관)	작성	
	서식⑧ 적정성 평가위원 서약서		해당 없음	담당자 검토사항
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	'개인정보 이노베이션 존' 운영기관의 관리·통제 하에 활용되므로 생략
	서식⑩ 가명(추가)정보 관리대장	B지자체 (가명정보 제공기관)  ※ 기관 내 개인정보 보호 부서에서 총괄 작성·관리하는 것도 가능	작성	가명정보 활용 과정에서의 관리대장은 '개인정보 이노베이션 존' 관리·통제 하에 활용되므로 작성 불필요

## 주요 서식 작성 예시

### <서식 1>

#### 가명정보 이용·제공 신청서 ( 신규 변경 )

접수번호	AA-2025177	기존 접수번호(변경시)		접수일	
신청자	조직/부서명	A부동산연구원 / 정책연구팀			
	담당자 직위	차장	담당자 성명	김 임 대	
	전화번호	070-000-0000	이메일 주소	house@happy.or.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 보증금·월세·증액보증금·전세/월세 구분·확정일자·전입일 등 계약 항목을 활용해 금액 수준별로 계약이 어떤 특성을 갖는지 패턴을 분석하고, 금액대별 임차인 보호장치(전입·확정일자 등) 차이를 파악하여 정책 개선 기초자료로 활용				
가명처리 대상 개인정보	개인정보 명칭	2021년 1월 ~ 2024년 12월 간 임대차 계약 현황 데이터			
	요청 개인정보 항목	임대인 생년월일, 임대인 성별, 임대인 개인/법인 구분, 임차인 연령, 임차인 성별, 신규재계약구분, 계약체결일, 계약시작일, 계약종료일, 전세/월세 구분, 보증금, 증액보증금, 월세, 전입일, 확정일자, 퇴거일			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input checked="" type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input checked="" type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	반출일로부터 ( 12 )개월 이내 또는 반출일로부터 ~ 년 월 일 까지				

「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.

2025년 2월 15일

신청인(부서장명) 박 보 호 (서명 또는 인)

#### B지자체 시장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

<서식②>

**가명정보 활용 계획서**

<b>신청 제목</b>	보증금·월세 수준에 따른 임차인 보호제도 차이 분석 연구
<b>책임자</b>	소속: A부동산연구원 정책연구팀      성명: 김 임 대      (서명)
<b>배경 및 목적</b>	<p>최근 전세사기, 보증금 미반환 등 임대차 관련 피해가 사회적 문제로 대두됨에 따라, 임대차 계약구조와 보호제도의 실제 작동 간의 차이를 실증적으로 분석할 필요성이 커지고 있음</p> <p>이에 임대차계약 데이터를 기반으로 임대료 수준(보증금·월세)과 계약 특성에 따른 임차인 보호수준의 차이를 분석하여, 제도적 보완 필요성과 정책 방향을 제시하고자 함</p> <p>임대차계약 관련 정보(보증금, 월세, 계약기간 등)를 활용하여 임대차 유형별(전세·월세·보증부월세 등) 보호제도 적용 현황과 실효성을 비교·분석함으로써, 임차인의 주거 안정성 및 제도적 보호 수준의 차이를 규명하고, 향후 주거정책 개선을 위한 근거자료를 마련하고자 함</p>
<b>활용 데이터</b>	2021년 1월 ~ 2024년 12월 간 임대차 계약 현황 데이터
<b>활용 계획 및 방법</b>	<ol style="list-style-type: none"> <li>1. 데이터 정제 및 가명처리 : 이름, 주소 등 직접식별정보 제거 및 계약 단위 난수 키 부여</li> <li>2. 계약유형 분류 : 보증금 및 월세 수준을 기준으로 전세형·월세형·혼합형 구간별 분류</li> <li>3. 보호수준 분석 : 확정일자, 전입신고, 임대인 유형, 계약기간 등을 기준으로 보호제도 적용 여부 및 빈도 분석</li> <li>4. 회귀 및 군집분석 : 계약 특성과 임차인 보호수준 간의 상관성 분석, 지역별/계약형태별 패턴 도출</li> <li>5. 시각화 및 정책 도출 : 계약금 수준과 보호조치 현황의 관계를 시각화하고 정책적 개선방향 제시</li> </ol>
<b>(예상) 활용 기간</b>	2025년 3월~ 2026년 2월까지 (12개월)
<b>기대효과 및 활용방안</b>	<ul style="list-style-type: none"> <li>- 임대차 유형별(전세·월세·보증부월세 등) 임차인 보호수준 차이를 정량적으로 분석하여 정책 근거자료로 활용</li> <li>- 보증금 규모 및 계약형태에 따른 보호 사각지대 규명으로 주거안정성 제고</li> <li>- 향후 임대차 제도 개선 및 피해예방 정책 수립을 위한 실증 기반 마련</li> </ul>

## 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요	
가명정보 활용 목적	보증금·월세 수준에 따른 임차인 보호제도 차이 분석 연구
가명처리 대상 개인정보	2021년 1월 ~ 2024년 12월 간 임대차 계약 현황 데이터
활용하려는 자	(기관명) A부동산연구원 정책연구팀 (직위) 차장 (성명) 김임대
(1) 기본 위험도 판단	
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
기본 위험도 자동 도출	〈Step 1〉 내부 활용 → <input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(A부동산연구원)에 제공하며, 본 기관에서 통제할 수 없는 환경인 ○○센터 내의 분석환경에서 활용될 예정임
(2) 사례별 위험도 조정	
반복·유사 활용 건	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음    ※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input checked="" type="checkbox"/> 해당 <input type="checkbox"/> 해당 없음    ※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당 <input checked="" type="checkbox"/> 해당 없음    ※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정	
조정	<input type="checkbox"/> 유지 <input type="checkbox"/> 상향 <input checked="" type="checkbox"/> 하향
최종 위험도	<input checked="" type="checkbox"/> 저위험 <input type="checkbox"/> 중위험 <input type="checkbox"/> 고위험
종합 검토 결과	개인정보보호위원회가 이미 처리환경의 안전성을 검증하여 지정한 ○○센터의 '개인정보 이노베이션 존'에서 활용될 예정이며, 개인정보 이노베이션 존은 반입 전·후의 전문심의위원회의 심의를 수행하게 되므로 저위험으로 판단하여 담당자 검토로 진행하도록 함.
검토일 : 2025년 2월 07일	
검토자 소속 : B지자체 정보보안과    성명 : 김주택 (서명 또는 인)	

## 2



**Key Point**

→

### ◆ 개요

보험설계사 연결 플랫폼인 A업체가 고객 특성 및 보험금 지급 패턴 연구를 위해 B화재보험 청구·지급 데이터를 가명정보로 제공받아 B화재보험 내부 분석환경에서 이용하는 건

### ◆ 처리 목적

보험사의 가명처리된 청구·지급 데이터를 활용하여, 설계사별 상품추천 유형과 고객의 보험금 청구 빈도·금액·사유 간의 연관성을 분석함으로써, 영업활동의 품질관리 및 고객 맞춤형 상품 설계 개선 방안을 연구

### ◆ 데이터 흐름



### 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	A업체 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	A업체 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	B화재보험 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	B화재보험 (가명정보 제공기관)	작성	
	서식⑤ 가명처리 결과서	B화재보험 (가명정보 제공기관)	작성	
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	비정형데이터 미포함
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	B화재보험 (가명정보 제공기관)	작성	
	서식⑧ 적정성 평가위원 서약서		생략	외부평가위원이 없는 내부심의 이므로 생략함
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	제공기관 관리·통제 하에 활용되므로 생략
	서식⑩ 가명(추가)정보 관리대장	B화재보험 (가명정보 제공기관)	작성	가명정보 활용 과정에서의 관리대장은 제공기관 관리·통제 하에 활용되므로 작성 불필요

주요 서식 작성 예시

〈서식 ①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	KL-20250305	기존 접수번호(변경시)		접수일	
신청자	조직/부서명	A업체 / 빅데이터 분석팀			
	담당자 직위	매니저	담당자 성명	김보험	
	전화번호	010-000-0000	이메일 주소	bhkim@insurance.co.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 보험사의 가명처리된 청구·지급 데이터를 활용하여, 설계사별 상품추천 유형과 고객의 보험금 청구 빈도·금액·사유 간의 연관성을 분석함으로써, 영업활동의 품질관리 및 고객 맞춤형 상품 설계 개선 방안을 연구				
가명처리 대상 개인정보	개인정보 명칭	2021년 ~ 2024년 동안 자사의 보험 청구 데이터 중 보험금 지급이 완료된 데이터			
	요청 개인정보 항목	<ul style="list-style-type: none"> <li>계약/상품 정보 : 피보험자D, 상품명, 보험료, 가입·만기일, 특약, 갱신/해지 이력</li> <li>청구/지급 정보 : 피보험자D, 사고유형, 사고접수일시, 청구일시, 청구금액, 지급일시, 지급금액, 지급결정내용, 심사담당자</li> </ul>			
결합 활용 여부	<input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	반출일로부터 ( 12 )개월 이내 또는 반출일로부터 ~ 년 월 일 까지				

「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.

2025년 3월 30일

신청인 (부서장명) 김 신 청 (서명 또는 인)

B화재보험사장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	보험 청구·지급 데이터 기반 설계사 맞춤형 상품 설계 개선 연구
책임자	소속: A업체/빅데이터 분석팀      성명: 김 보 험      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>보험설계사 연결 플랫폼을 중심으로, 고객의 보험 청구·지급 데이터가 대규모로 축적되고 있으나 이를 영업 품질 관리나 상품 설계 개선에 활용하는 체계는 미비함</li> <li>가명정보 기반 분석을 통해 설계사별 추천 유형, 고객 청구 빈도·금액·사유 간의 상관성을 파악함으로써 맞춤형 보장 설계 및 상담 효율화를 도모하고자 함</li> <li>보험 청구·지급 이력 데이터를 활용하여 설계사별 상품추천 패턴과 고객 청구특성 간의 연관성을 분석</li> <li>영업활동 품질관리 지표 도출 및 고객 맞춤형 상품 추천 알고리즘 개선방안 제시</li> <li>데이터 기반 보험설계사 매칭 서비스 모델의 검증 및 실증적 개선안 마련</li> </ul>
활용 데이터	<p>2021년 ~ 2024년 동안 자사의 보험 청구 데이터 중에서 보험금 지급이 완료된 데이터</p> <ul style="list-style-type: none"> <li>계약/상품 정보 : 피보험자ID, 상품명, 보험료, 가입·만기일, 특약, 갱신/해지 이력</li> <li>청구/지급 정보 : 피보험자ID, 사고유형, 사고접수일시, 청구일시, 청구금액, 지급일시, 지급금액, 지급결정내용, 심사담당자</li> </ul>
활용 계획 및 방법	<ol style="list-style-type: none"> <li>전처리 <ul style="list-style-type: none"> <li>개인정보 직접식별 항목(이름, 연락처, 계약번호 등)은 전면 삭제</li> <li>고객·설계사 구분키는 난수 기반 일회용 식별자로 치환</li> <li>보험사·상품명은 범주형 코드화, 금액은 구간화 및 로그변환</li> </ul> </li> <li>모델링 전략 <ul style="list-style-type: none"> <li>연관분석(Apriori, FP-Growth)을 통한 추천유형-청구패턴 간 규칙 탐색</li> <li>XGBoost/LightGBM 기반 고객 청구빈도 예측 및 설계사 성과모델 학습</li> <li>클러스터링(K-Means, DBSCAN)을 통한 설계사 그룹 세분화 및 특성 도출</li> <li>시계열분석(ARIMA, Prophet)을 통한 청구금액 추이 예측</li> </ul> </li> <li>검증 및 시각화 <ul style="list-style-type: none"> <li>피처 중요도, 추천 정확도(Precision@K), 청구패턴 유사도(코사인 유사도) 등 성능지표 활용</li> </ul> </li> </ol>
(예상) 활용 기간	반출일로부터 12개월 이내
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>설계사별 영업활동 품질관리 지표 개발 및 플랫폼 서비스 개선</li> <li>고객 맞춤형 상품 추천 알고리즘 고도화를 통한 보장분석 서비스 경쟁력 강화</li> <li>보험사-플랫폼 간 협력형 데이터 생태계 조성 및 가명정보 활용모델 확립</li> </ul>

〈서식③〉

### 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	보험 청구·지급 데이터 기반 설계사 맞춤형 상품 설계 개선 연구		
가명처리 대상 개인정보	2021년 ~ 2024년 동안 자사의 보험 청구 데이터 중에서 보험금 지급이 완료된 데이터		
활용하려는 자	(기관명) A업체 빅데이터 분석팀 (직위) 매니저 (성명) 김보험		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input checked="" type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input checked="" type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(A업체)에 제공하며, 본 기관에서 통제할 수 가능한 환경인 기관 내 안전한 분석환경에서 활용될 예정임		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input checked="" type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input checked="" type="checkbox"/> 중위험	<input type="checkbox"/> 고위험
종합 검토 결과	위험도를 조정할 특이사항이 없으므로 기본 위험도인 중위험을 최종 위험도로 채택하고 내부심의를 통해 적정성 검토 진행		
검토일 : 2025 년 4 월 4 일			
검토자 소속 : B화재보험 정보보호팀    성명 : 박정보 (서명 또는 인)			

## 〈서식④〉

## 가명처리 계획서

파일 개요						
No	파일관계	파일명	항목 수	데이터 건수	정보주체 수	비고
1	주	청구지급정보	5	10,000	10,000	
2	부	계약상품정보	4			

파일별 상세 가명처리 계획						
No	파일명	항목명	유형	예시	항목설명	가명처리 계획
1	청구지급정보	성별	범주형	여	성별	미적용
2	청구지급정보	주소	범주형	서울시 중구...	거주지 주소	부분삭제-시도이하
3	청구지급정보	직업	범주형	교사	직업	빈도수 3이하 로컬 삭제
4	청구지급정보	청구액	수치형		보험 청구액	상단코딩 99% 이상
5	청구지급정보	지급액	수치형		보험 지급액	상단코딩 99% 이상
6	계약상품정보	보험료	수치형		계약상품 보험료	상단코딩 99% 이상
...	...	...	...	...	...	...

〈서식⑤〉

### 가명처리 결과서

1. 공통적으로 적용한 가명처리 방법					
자료형	가명처리 순서				
	가명처리 방법 1	가명처리 방법 2	가명처리 방법 3		
범주형					
수치형	상단코딩 99% 이상				
기타					

2. 예외적으로 적용한 가명처리 방법					
파일명	항목명	자료형	가명처리 방법1	가명처리 방법 2	가명처리 방법 3
청구지급정보	주소	범주형	부분삭제-시도이하		
청구지급정보	직업	범주형	빈도수 30이하로컬 삭제		
...	...	...	...	...	...

3.1 범주형 변수 분포 현황 (필요한 항목만 작성)					
파일명	항목명	값	빈도수	구성비	비고
청구지급정보	주소	서울	6,000	30%	
청구지급정보	주소	부산	3,000	15%	
...	...	...	...	...	...

3.2 범주형 변수 조합 분포 현황 (필요한 항목만 작성)					
파일명	항목명	값	빈도수	구성비	비고
청구지급정보	성별, 주소, 직업	여-제주-교사	3	0.04%	
청구지급정보	성별, 주소, 직업	남-강원-철도기관사	3	0.04%	
청구지급정보	성별, 주소, 직업	여-경북-외환딜러	5	0.18%	
...	...	...	...	...	...

4.1 수치형 변수 분포 현황 (필요한 항목만 작성)									
파일명	항목명	처리 전				처리 후			
		Min	빈도	Max	빈도	Min	빈도	Max	빈도
청구지급정보	청구액	20,000	256	1,100,000,000	1	20,000	256	150,000,000	43
청구지급정보	지급액	20,000	256	1,100,000,000	1	20,000	256	80,000,000	78
계약상품정보	보험료	3,000	150	12,570,000	1	3,000	150	1,200,000	345
...	...	...	...	...	...	...	...	...	...

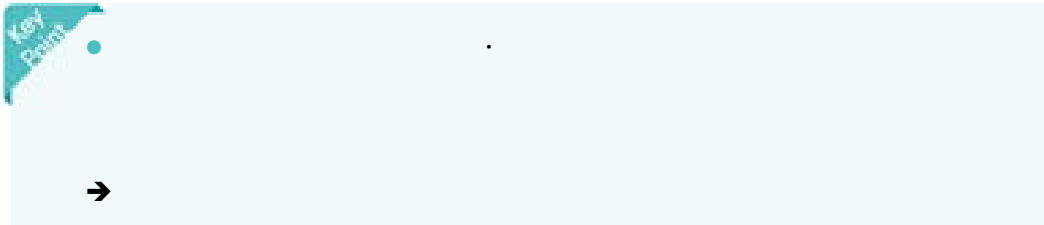
4.2 수치형 변수 사분위수 현황 (필요한 항목만 작성)						
파일명	항목명	최소값	1분위수(25%)	2분위수(중앙값)	3분위수(75%)	최대값
청구지급정보	청구액	20,000	350,000	900,000	2,500,000	150,000,000
청구지급정보	지급액	20,000	300,000	750,000	2,000,000	80,000,000
...	...	...	...	...	...	...

## 〈서식 7〉

## 적정성 검토 결과서

접수번호 (또는 처리명칭)	보험 청구·지급 데이터 기반 설계사 맞춤형 상품 설계 개선 연구					
검토위원 정보	성명	이검토	소속	B화재보험 데이터융합팀	직위	부장
검토 대상	<input checked="" type="checkbox"/> 신규 <input type="checkbox"/> 보완					
검토 일자	2025년 5월 12일					
검토 결과	<input checked="" type="checkbox"/> 적정 (승인) <input type="checkbox"/> 조건부 승인 <input type="checkbox"/> 부적정 (반려)					
검토 사항	① 가명정보 처리 목적 적합성 - 보호법 제28조의2에서 정한 가명정보 처리 목적 (통계작성, 과학적 연구, 공익적 기록보존)에 해당하는가? - 처리 목적을 구체적으로 설정하였는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	② 가명정보 이용항목 적합성 - 처리 목적 달성에 필요한 항목만 처리 대상에 포함하였는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	③ 위험도 판단의 적정성 - 활용 주체, 처리 환경, 개별 사례별 특수성 등을 고려하여 적정하게 위험도를 판단하였는가? 판단 근거가 적절한가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	④ 가명처리의 적정성 (재식별 위험성) - 항목별 가명처리 방법과 수준을 적절하게 설정하여 처리하였는가? - 가명정보 데이터 자체의 위험성, 처리환경의 위험성을 종합적으로 고려하였을 때 종합적인 재식별 위험성을 충분히 낮추었는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	⑤ 목적 달성 가능성 - 당초 설정한 가명정보 처리 목적을 달성할 수 있는가? (과도한 가명처리로 처리 목적 달성이 어려워지지 않았는가?)				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
검토 의견	- 김영희 대리 : 처리된 가명정보의 결과로는 특정 개인을 식별할 가능성은 희박하고 기관 내부 분석환경 내 철저한 관리적·기술적 안전조치로 인해 안전하게 관리될 것으로 보임					
종합 검토 의견 (위원장)	- 특정 개인이 식별되지 못할 정도로 충분히 가명처리가 수행되었으며, 내부 분석환경 내에서 활용될 경우 다른 정보와의 결합 가능성·용이성도 상당히 낮은 수준이기에 해당 분석환경 으로의 가명정보 반출을 승인함					
위와 같이 적정성 검토 결과를 통지합니다.						
<div style="display: flex; justify-content: space-around;"> <span>작성자</span> <span>이 검 토</span> <span>서명(인)</span> </div>						
<b>B화재보험사장 귀중</b>						

3



◆ 개요

A업체가 AI 학습을 통한 가명처리 솔루션 성능 개선 및 고도화를 위해 B공공기관의 상담데이터를 제공 받아 활용

B공공기관 내부 분석공간에 A업체가 출입하여 활용 수행

◆ 처리 목적

개인정보가 포함된 B공공기관의 상담 데이터(음성 및 자유형 텍스트 데이터)를 활용하여 AI학습을 통해 A업체가 개발 중인 가명처리 솔루션의 개인정보 탐지 AI 모델 성능을 개선·고도화하고자 함

◆ 데이터 흐름



### 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	A업체 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	A업체 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	B공공기관 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	B공공기관 (가명정보 제공기관)	작성	
	서식⑤ 가명처리 결과서	B공공기관 (가명정보 제공기관)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	B공공기관 (가명정보 제공기관)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	A업체 (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	A업체 (가명정보 이용자)  B공공기관 (가명정보 제공기관)	작성	가명정보 제공 및 활용 과정에서 기록·관리 필요

◆ 주요 서식 작성 예시

〈서식 ①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	KS-20250622	기존 접수번호(변경시)		접수일	
신청자	조직/부서명	A업체 기술연구소			
	담당자 직위	수석연구원	담당자 성명	김 학 습	
	전화번호	02-000-0000	이메일 주소	hskim@data.co.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 가명처리 솔루션 내 비정형 데이터(음성, Free Text) 가명처리 성능 개선 (처리율, 오탐율) 및 고도화				
가명처리 대상 개인정보	개인정보 명칭	2023년 6월 ~ 2023년 12월간 (6개월)간 접수된 상담 데이터 (음성)			
	요청 개인정보 항목	통화 시간, 상담자, 피상담자, 시각, 상담 내용, 발신자 전화번호			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	반출일로부터 ( )개월 이내 또는 반출일로부터 ~ 2025년 12월 31일 까지				

「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.

2025년 6월 18일

신청인 (부서장명) 김 학 습 (서명 또는 인)

B 공공기관장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

<b>신청 제목</b>	상담(비정형, 음성)데이터 내 개인정보 탐지 AI 모델 개발
<b>책임자</b>	소속: A업체 기술연구소      성명: 김 학 습      (서명)
<b>배경 및 목적</b>	<p>인공지능 기반 개인정보 탐지 기술은 개인정보 비식별화 및 가명처리의 핵심 기반 기술임에도 불구하고, 비정형데이터(상담 텍스트·음성 등)에 대한 학습 데이터 부족으로 인해 고정밀 탐지가 어려운 한계가 존재함. 이에 상담센터에서 축적된 비정형 텍스트 데이터를 활용하여, 개인정보 탐지에 특화된 AI 모델을 고도화함으로써 향후 공공·민간 분야에서의 안전한 데이터 가명처리 및 AI 기반 데이터 활용을 촉진하고자 함</p> <p>개인정보가 포함된 비정형 텍스트 데이터(상담센터 데이터)를 활용하여 개인정보 탐지에 특화된 고정밀 AI 모델을 개발하고 학습시켜 현재 당사가 개발 중인 개인정보 탐지 AI 모델의 성능 향상</p> <ul style="list-style-type: none"> <li>고정밀 AI 개인정보 탐지 모델 개발 및 성능 향상</li> <li>B공공기관의 상담센터 데이터셋을 활용하여 개인정보 탐지에 최적화된 AI 모델을 개발하고 F1-score 0.95 이상의 탐지 성능 달성 <ul style="list-style-type: none"> <li>기존 모델과 개발될 AI 모델 각각의 개인정보 탐지 수행 결과를 객관적으로 비교 분석하여 성능 향상을 실증</li> </ul> </li> <li>공공데이터의 AI 산업 활용 가능성 입증 <ul style="list-style-type: none"> <li>공공기관으로부터 제공받은 실제 상담 데이터를 활용한 AI 모델 개발 및 솔루션 성능 향상 사례를 구축하여, 민감한 공공 데이터의 AI 산업 활용 가능성과 그 가치를 실증</li> </ul> </li> </ul>
<b>활용 데이터</b>	<ul style="list-style-type: none"> <li>B공공기관의 상담센터 데이터(비정형 텍스트 데이터) <ul style="list-style-type: none"> <li>해킹, 스팸, 개인정보침해 등에 대한 상담내용 데이터로 신고내용 및 신고처리를 위해 개인정보인 이름, 연락처 등을 포함하고 있음</li> </ul> </li> </ul>
<b>활용 계획 및 방법</b>	<ol style="list-style-type: none"> <li>데이터 활용에 앞서 데이터 정제 및 개인정보 태깅(라벨링)을 포함한 데이터 전처리 작업 수행</li> <li>비정형 데이터 내 개인정보를 정확하게 탐지하는 고성능 AI 모델을 개발 <ul style="list-style-type: none"> <li>자연어 처리(NLP) 모델 기반의 AI 개발</li> <li>Transfer Learning, Fine-tuning 기법 활용하여 모델 최적화</li> </ul> </li> <li>내부분석 공간에서 AI 모델 학습 진행 <ul style="list-style-type: none"> <li>학습된 모델에 대해서는 Precision, Recall, F1-score, Accuracy 등 지표를 활용하여 성능 평가 및 비교</li> </ul> </li> <li>개발된 AI 모델을 탑재한 비정형데이터 비식별처리 SW 프로토타입 개발</li> </ol>
<b>(예상) 활용 기간</b>	반출일~ 2025년 12월 31일까지
<b>기대효과 및 활용방안</b>	<ul style="list-style-type: none"> <li>고정밀 AI 개인정보 탐지 기술 역량 확보 <ul style="list-style-type: none"> <li>B공공기관 상담센터 데이터셋을 활용하여 비정형 데이터 내 개인정보 탐지에 최적화된 고정밀 AI 모델을 개발함으로써, 기술 경쟁력 확보 및 데이터 처리 역량 강화</li> </ul> </li> <li>AI 모델 개발 및 성능 고도화 프로세스의 효율성 극대화 <ul style="list-style-type: none"> <li>학습 데이터를 통해 AI 모델 학습 시간을 단축하고, 반복적인 실험 및 모델 튜닝 과정을 효율적으로 수행함으로써, 연구 개발 기간 단축 및 비용 절감</li> </ul> </li> <li>공공기관-민간기관 협력 모델 구축 및 확산 기여 <ul style="list-style-type: none"> <li>공공기관(B공공기관)과 민간기관(A업체) 간의 데이터 제공, 안전 활용, 기술 개발 연계를 통한 실질적인 협력 모델을 구축하여, 향후 다양한 분야에서 공공 데이터의 안전한 활용과 AI 산업 발전을 촉진할 수 있는 협력 모델 제시</li> </ul> </li> </ul>

### 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	상담(비정형, 음성)데이터 내 개인정보 탐지 AI 모델 개발		
가명처리 대상 개인정보	B공공기관의 상담센터 데이터(비정형 텍스트 데이터)		
활용하려는 자	(기관명) A업체 기술연구소 (직위) 수석 (성명) 김학습		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input checked="" type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input checked="" type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(A업체)에 제공하며, 본 기관에서 통제할 수 없는 환경인 기관 내의 분석환경에서 활용될 예정임		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	내부 분석환경에 제공하여 비교적 안전함에도 불구하고 비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 고위험으로 판단하고 외부전문가 인원을 포함한 적정성검토 위원회를 별도로 구성하여 검토 후 반출하도록 하는 것이 적절		
검토일 : 2025 년 5 월 23 일			
검토자 소속 : B공공기관 데이터정보화팀 성명 : 김기동 (서명 또는 인)			

## 〈서식④〉

## 가명처리 계획서

가명처리 대상 파일			
No	파일관계	대상 정보	설명
1	주	전화상담 음성 정보	23년 6월~12월까지의 전화상담 음성정보(MP3)
2	변환	전화상담 텍스트 정보	전화상담 음성 정보를 텍스트로 전환한 정보

## 1. 전화상담 음성 데이터 (비정형-음성)

정보주체 수	117,729건	항목수	1개 항목	데이터건수	117,729건
No	항목명	유형	예시	항목설명	가명처리 계획
1	전화상담 음성정보	음성	“안녕하십니까?전화상담센터이전수입니다.네여보세요.안녕하세요.죄송한데,혹시이계경찰청선생님이연결해주신건데,여기가정확하게무슨기관인거죠?여기는B공공기관입니다.요약해서말씀드리면,SNS를통해서여자BJ사칭을받아서카톡대화까지이어졌거든요.”	<ul style="list-style-type: none"> <li>○ 음성정보</li> <li>○ 통화 건별 발생</li> </ul>	<ul style="list-style-type: none"> <li>○ 5분 미만 데이터 삭제 후 텍스트 파일 변환</li> <li>○ 텍스트 파일 생성 후, 음성 정보는 미활용 예정임에 따른 삭제</li> </ul>

## 2. 전화상담 텍스트 데이터 (비정형-텍스트)

정보주체 수	117,729건	항목수	1개 항목	데이터건수	117,729건
No	항목명	유형	예시	항목설명 및 데이터량	가명처리 계획
1	전화상담 음성정보	텍스트	“안녕하십니까?상담센터홍길동입니다.네여보세요.안녕하세요.죄송한데,혹시이계경찰청선생님이연결해주신건데,여기가정확하게무슨기관인거죠?여기는B공공기관입니다.요약해서말씀드리면,SNS를통해서여자BJ사칭을받아서카톡대화까지이어졌거든요.”	<ul style="list-style-type: none"> <li>○ 주 파일인 음성정보를 텍스트로 변환한 파일</li> <li>○ 117,729건</li> </ul>	<ul style="list-style-type: none"> <li>• STT 변환 후 원본 삭제</li> <li>• 데이터 활용 가능한 3만건 외 나머지 약 8만건 삭제(전처리)</li> <li>• 치환(이름-홍길동, 숫자-0, 영어-a)</li> <li>• 주요 개인정보 삭제</li> </ul>

〈서식 5〉

가명처리 결과서

5. 비정형데이터 전후 비교 샘플링			
파일명	항목명	처리 전	처리 후
상담정보	상담내용	<p>안녕하세요. B공공기관 상담센터 상담사 <b>홍길동</b>입니다.</p> <p>안녕하세요. 제가 지금 보이스피싱 연락을 받았는데, 어떻게 해야할지 모르겠어서 전화했어요. 제 전화번호는 <b>일이삼사일이삼사</b>거든요, 보이스피싱 전화가 왔어요. 집 주소인 <b>한국아파트 백일동101호</b>며, 제 이름 <b>김대한</b>모든 정보를 알고 있어요. 또 한국은행이 주거래은행인데, 거기 계좌번호인 <b>123-12-123456</b>도 알고 있어서, 뭔가 의심스러웠지만 자꾸 제 개인정보를 말해서 순간적으로 보내준 사이트 주소로 들어가서 주민등록번호로 본인인증을 했어요. 이거 보이스피싱 맞는 거죠? 본인인증 한 사이트 주소는 알고 계실까요? 사이트 확인 후 안내 도와드리겠습니다.</p>	<p>안녕하세요. B공공기관 상담센터 상담사 <b>김가명</b>입니다.</p> <p>안녕하세요. 제가 지금 보이스피싱 연락을 받았는데, 어떻게 해야할지 모르겠어서 전화했어요. 제 전화번호는 <b>77777777</b>거든요, 보이스피싱 전화가 왔어요. 집 주소인 <b>한국아파트 7동7호</b>며, 제 이름 <b>김가명</b>모든 정보를 알고 있어요. 또 한국은행이 주거래은행인데, 거기 계좌번호인 <b>777-77-7777</b>도 알고 있어서, 뭔가 의심스러웠지만 자꾸 제 개인정보를 말해서 순간적으로 보내준 사이트 주소로 들어가서 주민등록번호로 본인인증을 했어요. 이거 보이스피싱 맞는 거죠? 본인인증 한 사이트 주소는 알고 계실까요? 사이트 확인 후 안내 도와드리겠습니다.</p>
		<p>• 미처리된 데이터 추가 처리 수행</p> <ul style="list-style-type: none"> <li>- 계획 대비 실행 검증: '항목별 가명처리 계획서'의 처리 규칙과 실제 처리된 샘플 데이터를 상호 비교하여 계획 준수 여부를 전수 확인.</li> <li>- 식별자 잔존 여부 검증: 이름, 전화번호, 주민등록번호 등 주요 직접식별자 패턴에 대한 자동화된 스크립트를 실행하여 마스킹 누락 여부를 검증하고, 수동으로 비정형 텍스트 내 간접 식별자 처리 적절성을 검토.</li> <li>- 특이정보(Outlier) 분석: 상담 내용이 매우 독특하여 특정한 추론 가능성이 높은 데이터 100건을 별도 추출하여, 문맥 삭제 및 요약 처리가 재식별 위험을 충분히 완화했는지 심층 검토</li> </ul>	

## 〈서식⑥〉

## 비정형데이터 추가검수 결과서

가명처리 개요			
접수번호 (또는 처리명칭)	상담(비정형, 음성)데이터 내 개인정보 탐지 AI 모델 개발		
원본 데이터 규모	음성정보 117,729개 파일	원본 데이터 크기(용량)	30GB
원본 데이터 포맷	WAV	처리 결과 데이터 포맷	TXT
가명처리 적용 기술	데이터 활용 가능한 3만건 외 나머지 약 8만건 삭제(전처리) 후 치환(이름-홍길동, 숫자-0, 영어-a)		
추가검수 내용			
검수 기간	2025년 7월 5일 ~ 2025년 7월 12일		
검수 장소	B공공기관 내부 가명처리실		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input checked="" type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수		
	<ul style="list-style-type: none"> <li>가명정보 처리 담당 부서장의 주도 하에 가명정보 처리 담당 부서 직원 6인이 함께 진행</li> <li>부서장과 직원이 함께 전체 3만 건의 데이터 중 약 300개(10%) 파일을 무작위 샘플링 하여 육안 검사 (각 50개 파일씩 검수)</li> </ul>		
처리된 결과 확인 사항	<ol style="list-style-type: none"> <li>전화상담데이터의 텍스트 변환정보 내 아래의 정보가 계획대로 처리되었는지 확인             <ul style="list-style-type: none"> <li>계획 준수: '항목별 가명처리 계획'에 명시된 모든 기술이 계획대로 100% 적용되었음을 확인</li> <li>식별자 제거율: 패턴으로 탐지 가능한 직접 식별자는 99.9% 이상 마스킹 처리되었으며, 누락된 소수 사례는 STT 변환 오류로 인한 오타로 확인되어 추가 조치 완료</li> <li>원본 데이터 파기: STT 변환 후 원본 음성 파일이 영구 삭제되었음을 서버 접근 및 파일 시스템 로그를 통해 교차 확인</li> </ul> </li> <li>위 사항 중 하나라도 만족하지 않았을 경우 추가처리 완료             <ul style="list-style-type: none"> <li>⇒ 누락하여 재처리한 사례 중 개인식별 위험이 특별히 높은 건은 발견되지 않았으며, 대화맥락에 따른 식별 위험도 거의 없는 것으로 판단되었음. 또한 활용 기관이 통제·관리하는 안전한 환경에서 제한된 인원에 의해서만 처리되므로 위험이 실질적으로 발생되지 않을 것으로 판단되어 추가 표본검수·재처리 없이 해당 잔존 위험은 감수하기로 결정</li> </ul> </li> </ol>		
검수 결과	가명처리 결과, 특정 개인을 알아볼 수 있는 정보가 합리적인 수준 이하로 제거되었으며, 과학적 연구 목적 달성을 위한 데이터의 통계적 유용성은 충분히 보존되었음을 확인함. 따라서 본 가명처리 결과는 '적정한 것으로 판단함'		
검증자	소속 및 직위	성명	서명(인)
	B공공기관 수석	양길동	
	B공공기관 책임	임순신	
	B공공기관 선임	김몽룡	
	B공공기관 주임	김춘향	
B공공기관 주임	신철수		
2025년 7월 13일			
작성자 소속 : B공공기관 데이터정보화팀			
성명 : 김기동      서명(인)			

# 8

## 가명정보 활용 참고 시나리오

가명정보 처리 가이드라인

가명정보 처리 실무편







## 제8장 가명정보 활용 참고 시나리오

### 1

#### ◆ 개요

- (주)한국유통은 상품군별 판매추이에 대한 통계작성을 위한 데이터 제공을 요청받아, (주)나이스통계 업체에게 다음과 같이 판매 데이터를 제공하려 한다.

#### ◆ 처리 목적

- 코로나 이전과 코로나 이후의 상품군별 판매추이에 대한 통계작성으로 지속적인 분석을 위해 2019년 1월부터 12월까지의 상품 매출 데이터와 2021년 1월부터 12월까지의 상품 매출 데이터를 제공하기로 함

#### ◆ 데이터 특징

- 동일 기간의 고객 구매 데이터를 제공 목적에 맞게 원본데이터를 가공하여 제공
- 전체 820만명의 고객 중 50%를 샘플링하여 410만건을 제공

#### ◆ 데이터 이용 환경

- 데이터 이용자는 (주)나이스통계이며, 데이터 제공에 대한 계약을 맺어 제공함
- (주)나이스통계는 다양한 통계분석을 통해 통계정보를 만들어 판매하는 회사
- 개인정보의 분석환경은 인터넷과 분리된 환경에서 별도의 네트워크를 운영하고 있음
- 관련 연구원들은 10년이상 통계를 생성하는 분석업무를 수행한 전문인력

◆ 데이터 흐름



◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	(주)나이스통계 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	(주)나이스통계 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	(주)한국유통 (가명정보 제공기관)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	(주)한국유통 (가명정보 제공기관)	작성	
	서식⑤ 가명처리 결과서	(주)한국유통 (가명정보 제공기관)	작성	
	서식⑥ 비정형데이터 추가검수 결과서		해당 없음	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	(주)나이스통계 (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	(주)나이스통계 (가명정보 이용자)  (주)한국유통 (가명정보 제공기관)	작성	

## 〈서식①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	OZ-20220325	기존 접수번호(변경시)		접수일	2022.3.22.
신청자	조직/부서명	(주)나이스통계 물류분석팀			
	담당자 직위	과장	담당자 성명	이 통계	
	전화번호	02-000-0000	이메일 주소	stat1234@nicestat.com	
가명정보 활용 목적	<input checked="" type="checkbox"/> 통계 작성 <input type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존 ※ 세부 활용 목적 코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계 작성을 위한 가명정보의 제공				
	가명처리 대상 개인정보	개인정보 명칭	구매부에서 보관하고 있는 자사 고객 중 회원가입이 2019년 1월 이전인 고객의 구매정보 중 2019년 1월 ~ 2019년 12월, 2021년 1월 ~ 2021년 12월 까지의 구매정보데이터		
	요청 개인정보 항목	고객ID, 나이, 주소, 성별, 카테고리별 상품 구매액, 회원등급 등 (전체 222컬럼, 전체 대상인원 820만명 중 50% 무작위 샘플링한 410만명의 정보)			
결합 활용 여부	<input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : ) <input checked="" type="checkbox"/> 해당 없음				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input type="checkbox"/> 기관 내부 분석 공간 <input checked="" type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	반출일로부터 ( 7 )개월				

「개인정보 보호법」 제28조의2에 따라 위와 같이 가명정보 이용·제공을 신청합니다.

2022년 3월 22일

신청인 (부서장명) 이 통계 (서명 또는 인)

(주)한국유통 대표이사 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

〈서식②〉

가명정보 활용 계획서

신청 제목	고객 구매정보 통계분석
책임자	소속: (주)나이스통계                      성명: 이 통 계                      (서명)
배경 및 목적	COVID19는 전세계적으로 팬데믹을 불러왔다. 이에 따라 각종 산업이 위축되고 개인에 대한 거리두기로 인해 다양한 산업에서 변화를 가지고 왔으며 특히 개인의 소비 생활에도 많은 영향을 미치게 되었다. 이에 따라 코로나 이전과 코로나 이후의 개인별 판매 데이터를 분석하여 COVID19로 인해 개인에게 미치는 영향에 대한 파악 등 다양한 방면에 사용하기 위한 통계를 작성하려 한다.
활용 데이터	410만명 (2019년 1월~2021년 12월까지 구매내역이 있는 고객 중 50%를 무작위 샘플링)
활용 계획 및 방법	동일 집단의 COVID19발병 이전과 발병이후의 제품의 판매 추이를 통해 COVID19가 개인의 소비 패턴의 변화에 어떤 영향을 주었는지를 파악하기 위한 통계를 작성
(예상) 활용 기간	2022년 4월 1일 ~ 2022년 10월 31일(7개월)
기대효과 및 활용방안	코로나와 같은 질병이 발생할 때 소비의 변화에 따라 지원금의 처리 방법, 지원 규모와 생산 산업군에 대한 영향 등을 파악하여 이를 보완하기 위한 정책 연구 등에 사용

## 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	고객 구매정보 통계분석		
가명처리 대상 개인정보	2019년 1월~2021년 12월까지 구매내역이 있는 고객정보		
활용하려는 자	(기관명) (주)나이스통계 (직위) 과장 (성명) 이통계		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2도) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관(주)나이스통계)에 제공하며, 본 기관에서 통제할 수 없는 환경인 해당 기관 분석환경에서 활용될 예정임		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input checked="" type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	위험도를 조정할 특이사항이 없으므로 기본 위험도인 고위험을 최종 위험도로 채택하고 외부전문가를 포함한 적정성검토위원회 심의를 통해 검토 진행		
검토일 : 2025 년 5 월 23 일			
검토자 소속 : (주)한국유통 개인정보보호팀    성명 : 김한호 (서명 또는 인)			

〈서식④〉

가명처리 계획서

가명처리 대상 파일			
No	파일관계	대상 정보	설명
1	주	고객 정보	
2	부	구매 정보	

1. 고객 정보

정보주체 수	4,100,000건	항목수	4	데이터건수	4,100,000건
No	항목명	유형*	예시	항목설명	가명처리 계획
1	ID	식별자	hjkim2001	고객 구분값	Salt 추가 해시처리
2	나이	범주형	45	생년월일에서 추출한 정보	5세 단위로 범주화 20미만, 80이상 경계치
3	주소	범주형	서울시 송파구 잠실2동 7-101	고객의 주소	동단위 이하 삭제
4	성별	범주형	M		처리없음

2. 구매 정보

정보주체 수	4,100,000건	항목수	10	데이터건수	98,400,000건
No	항목명	유형	예시	항목설명	가명처리 계획
1	ID	식별자	hjkim2001		Salt 추가 해시처리
2	연월	범주형	2019-01	구매 기준 연월	처리없음
3	여행용품 구매액	수치형	10,000		상단 99.9% 초과시 계값으로 변경 후 천원 단위 절사
4	식품류 구매액	수치형	10,000		
5	의류 구매액	수치형	20,000		
6	취미용품 구매액	수치형	10,000		
7	생활용품 구매액	수치형	10,000		
8	구매액 합계	수치형	90,000		상단 99.9% 초과시 계값으로 변경 후 만원 단위 절사
9	선호제품군	범주형	의류		처리없음
10	고객등급	범주형	VIP		다음과 같이 범주화 처리 VIP, S, A → 1등급 B, C → 2등급 D, E, F → 3등급

\* 정형데이터인 경우 범주형, 수치형 중 택일  
비정형데이터인 경우 음성, 텍스트, 이미지, 영상, 특히 보건의료 분야는 CT, MRI, 유전체 등 구체적으로 표기

## 〈서식⑤〉

## 가명처리 결과서

1. 공통적으로 적용한 가명처리 방법				
자료형	가명처리 순서			
	가명처리 방법 1	가명처리 방법 2	가명처리 방법 3	
식별자	Salt 추가 해시처리			
범주형	빈도 3 이하 로컬삭제			
수치형	상단코딩(99.9%) 경계값 변경	천원단위 절사		
기타	해당없음			

2. 예외적으로 적용한 가명처리 방법			
자료형	적용 항목	가명처리 방법	사유
수치형	구매액 합계	상단코딩 99.9% 적용 후 십만단위 절사	
	.....		

3. 주요 범주형 항목 값 분포 (검토가 필요한 항목만 작성)				
항목명	값	빈도수	구성비	비고
나이	80	5	0.0%	
	75	34,124	0.0%	
	...	...	...	
	25	639,235	0.1%	
	20	422,304	0.0%	
주소	경상북도 울릉군 북면	5	0.0%	
	전라남도 신안군 흑산면	17	0.0%	
	제주특별자치도 제주시 추자면	22	0.0%	
	경상북도 봉화군 소천면	32	0.0%	
	강원특별자치도 양구군 양구읍	43	0.0%	

4. 주요 수치형 항목 값 분포 (검토가 필요한 항목만 작성)								
항목명	처리 전				처리 후			
	Min	빈도	Max	빈도	Min	빈도	Max	빈도
취미용품 구매액	0	340,000	138,023,000	1	0	340,000	8,900,000	2,345
.....								

〈서식 7〉

적정성 검토 결과서

접수번호 (또는 처리명칭)	고객 구매정보 통계분석					
검토위원 정보	성명	김한호	소속	(주)한국유통 개인정보보호팀	직위	팀장
검토 대상	<input checked="" type="checkbox"/> 신규 <input type="checkbox"/> 보완					
검토 일자	2022년 3월 31일					
검토 결과	<input checked="" type="checkbox"/> 적정 (승인) <input type="checkbox"/> 조건부 승인 <input type="checkbox"/> 부적정 (반려)					
검토 사항	① 가명정보 처리 목적 적합성 - 보호법 제28조의2에서 정한 가명정보 처리 목적 (통계작성, 과학적 연구, 공익적 기록보존)에 해당하는가? - 처리 목적을 구체적으로 설정하였는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	② 가명정보 이용항목 적합성 - 처리 목적 달성에 필요한 항목만 처리 대상에 포함하였는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	③ 위험도 판단의 적정성 - 활용 주체, 처리 환경, 개별 사례별 특수성 등을 고려하여 적정하게 위험도를 판단하였는가? 판단 근거가 적절한가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	④ 가명처리의 적정성 (재식별 위험성) - 항목별 가명처리 방법과 수준을 적절하게 설정하여 처리하였는가? - 가명정보 데이터 자체의 위험성, 처리환경의 위험성을 종합적으로 고려하였을 때 종합적인 재식별 위험성을 충분히 낮추었는가?				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
	⑤ 목적 달성 가능성 - 당초 설정한 가명정보 처리 목적을 달성할 수 있는가? (과도한 가명처리로 처리 목적 달성이 어려워지지 않았는가?)				<input checked="" type="checkbox"/> 적정	<input type="checkbox"/> 미흡
검토 의견	- (주보호 팀장) 이용기관으로 제공하는 고객정보와 구매정보의 항목이 가명정보 처리목적을 달성하기 위한 필수 항목들로 구성되어 있고, 불필요한 항목은 없음을 확인함. 가명처리 수준도 적절하게 처리되었음 - (외부위원: 박보안 변호사) 가명처리 목적이 보호법 제28조의2의 통계작성 목적에 적합하고, (주)한국유통이 통제 불가능한 외부로 제공하는 건이므로 고위험 판단이 적절하며 전반적인 가명처리가 적정하게 수행되었음 - (외부위원: 심동수 CPO) 실제 이용자와 협의 하에 가명처리 계획이 적정하게 작성되어 목적 달성 가능성이 충분하며, 폐쇄망 분석환경에서 재식별 시도 또는 목적 외 활용 가능성을 충분히 통제할 수 있을 것으로 기대됨					
종합 검토 의견 (위원장)	- 위원들의 의견을 종합한 결과, 이용기관이 구축한 폐쇄망에서 안전하게 처리할 계획으로 다른 정보의 입수가능성과 결합가능성을 배제한다면 현재 가명처리된 수준으로 특정 개인을 식별할 가능성은 희박함					
위와 같이 적정성 검토 결과를 통지합니다  작성자    김 한 호    서명(인)  (주)한국유통 대표이사 귀중						

## 2

## ① 유방암·골밀도 감소 진단 AI 개발 사례

## ◆ 개요 (이미지, 영상, 텍스트 정보)

- 한국대학교병원은 병원 내부 연구자에게 과학적 연구 수행을 위한 데이터 제공을 요청받아 유방암 환자의 이미지·영상 데이터(병리조직, 흉부CT, 골밀도검사(DEXA) 기록 등)와 임상 데이터(외과 병리 기록지(텍스트), 정형데이터)를 제공하려 한다.

## ◆ 처리 목적

- 유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구

## ◆ 데이터 특징

- (이미지·영상 데이터) 암 진단 및 치료 관련 병리조직이미지, 흉부CT, 골밀도 검사(DEXA) 기록
- (임상 데이터) 암 진단 시 발생한 암 관련 외과병리보고서(텍스트), 암등록정보 등 정형데이터

## ◆ 데이터 이용 환경

- (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가
- (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)
- (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

## ◆ 데이터 흐름



❖ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	한의료 교수 연구팀 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	한의료 교수 연구팀 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	빅데이터 처리팀 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	빅데이터 처리팀 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	빅데이터 처리팀 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	빅데이터 처리팀 (가명정보 제공자)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	내부 규정에 따라 관리
	서식⑩ 가명(추가)정보 관리대장	개인정보보호팀 (개인정보보호 담당 부서)	작성	

## 주요 서식 작성 예시

### 〈서식①〉

#### 가명정보 이용·제공 신청서 ( 신규 변경 )

접수번호		기존 접수번호(변경시)		접수일	2023년 1월 10일
신청자	조직/부서명	한국대학교병원 외과			
	담당자 직위	교수	담당자 성명	한의료	
	전화번호	03-1234-5678	이메일 주소	research_han@corea.ac.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존 ※ 세부 활용 목적 유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구				
	가명처리 대상 개인정보	개인정보 명칭	임상데이터, 병리 및 CT, DEXA 영상 이미지		
	요청 개인정보 항목	2012년부터 2019년까지 유방암 병리 및 CT, DEXA 검사를 받은 유방암 환자(500명)의 임상데이터, 병리 및 CT, DEXA 영상 이미지			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 5월 1일 ~ 2025년 4월 30일(2년)				

2023년 1월 10일

신청인 (부서장명) 한 의료 (서명 또는 인)

한국대학교병원장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구
책임자	소속: 한국대학교병원 외과      성명: 한 의 료      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>• 유방암은 2020년 기준 우리나라 여성에서 가장 많이 발생하는 암으로 정확한 진단을 위해 조직병리검사를 수행하고 있으며, 암 수술 후 전이, 재발 등을 검진하기 위해 정기적으로 CT 등 검사를 수행하고, 치료 중 골밀도 감소 여부 확인을 위해 DEXA 검사를 수행하고 있음</li> <li>• 유방암 병리조직이미지를 이용한 AI 분석을 통해 유방암 조직학적진단의 민감도를 높이고자 함</li> <li>• 또한 유방암 치료 중 골밀도 감소증 발생이 흔하여 주기적인 추적 검사가 필요한바, CT 및 DEXA 이미지 학습을 통해 CT 검사를 통한 골밀도 감소 사전 예측을 수행하고자 함</li> </ul>
활용 데이터	2012년부터 2019년까지 유방암 병리 및 CT, DEXA 검사를 받은 유방암 환자(500명)의 임상데이터, 병리 및 CT, DEXA 영상 이미지
활용 계획 및 방법	<ul style="list-style-type: none"> <li>• 병리 이미지와 조직 병리 검사결과(기록지)를 이용한 AI 학습을 통해 유방암 진단 결과 도출</li> <li>• 450명은 학습용 데이터셋으로, 50명은 테스트 데이터셋으로 활용</li> <li>• 흉부CT 이미지와 DEXA 검사결과를 이용한 AI 학습을 통해 CT에서 골밀도 감소 여부 결과 도출</li> </ul>
(예상) 활용 기간	2023. 5. 1. ~ 2025. 4. 30.(2년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>• 조직 병리 진단 자동화를 통해 유방암 진단의 정확도를 높이고 예후와의 관련성을 확인</li> <li>• 유방암 환자의 재발 등을 확인하기 위해 검사한 CT 이미지를 이용하여 골밀도 감소 진단을 수행함으로써 유방암 환자들에게 필요한 DEXA 검사 등의 부담을 감소시킬 수 있음</li> </ul>

## 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구		
가명처리 대상 개인정보	유방암 병리 및 CT, DEXA 검사를 받은 유방암 환자(500명)의 임상데이터, 병리 및 CT, DEXA 영상 이미지		
활용하려는 자	(기관명) 한국대학교병원 (직위) 교수 (성명) 한의료		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input checked="" type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용		→ <input checked="" type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경		→ <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경		→ <input type="checkbox"/> 고위험
	※ 판단 이유 : 내부 부서에서 자체 활용 예정		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	내부 분석환경에 제공하여 비교적 안전함에도 불구하고 비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 고위험으로 판단하고 외부전문가 인원을 포함한 적정성검토 위원회를 별도로 구성하여 검토 후 활용하는 것으로 결정		
검토일 : 2023 년 1 월 12 일			
검토자 소속 : 빅데이터처리팀 성명 : 김국현 (서명 또는 인)			

〈서식④〉

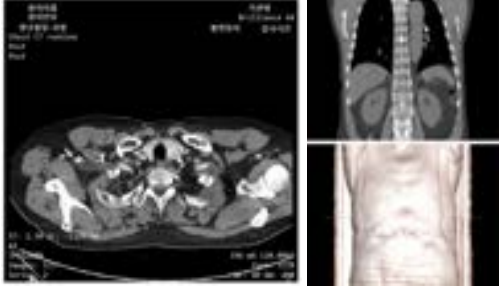

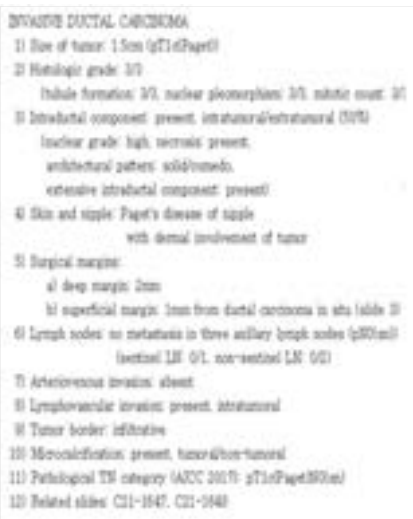
가명처리 계획서

【가명처리 대상 파일】

가명처리 대상 파일(요약)				
연번	대상 정보	데이터 유형	설명	비고
1	병리조직 데이터	이미지	2,000장 (500명*4장)	비정형데이터
2	흉부CT 데이터	영상·이미지 (DICOM*)	100,000장 (500명*100장*2회 촬영)	
3	골밀도 검사 (DEXA) 데이터	영상·이미지 (DICOM*)	1,000장 (500명*2장*1회 촬영)	
4	외과병리 보고서	텍스트(비정형, 관찰입력정보)	500건	
4-1	Organ(장기명)	텍스트(정형)	500건	비정형데이터 ⇒ 정형데이터 변환
4-2	Location(유방 수술의 위치)	텍스트(정형)	500건	
~	~	~	500건	
4-17	Intraductal_comp(상피내암 유무)	텍스트(정형)	500건	정형데이터 (환자 임상데이터)
5	gender(성별)	텍스트(정형)	500건	
6	death(사망일자)	숫자(정형)	500건	
~	~	~	500건	
32	MIEX_YMD(영상검사 일자)	숫자(정형)	500건	

\* DICOM(Digital Imaging and Communications in Medicine) : 의료용 디지털 영상 및 통신 표준

가명처리 대상 파일(상세)					
연번	항목명	데이터 유형	데이터 규모	구분	예시
1	병리조직 데이터	이미지	2,000장 (500명* 4장)	① 비율슬라이드 사진	
				② 병리슬라이드 번호 사진	
				③ 검체 슬라이드 사진	
				④ 검체 확대 사진	

가명처리 대상 파일(상세)					
연번	항목명	데이터 유형	데이터 규모	구분	예시
2	흉부CT 데이터	영상·이미지 (DICOM)	100,000장 (500명 * 100장 * 2회 촬영)		 <p>* 이미지 내 환자이름, 생년월일, 성별, 환자번호 존재</p>
3	골밀도 검사 (DEXA) 데이터	영상·이미지 (DICOM)	1,000장 (500명 * 2장 * 1회 촬영)		 <p>* 이미지 내 환자 생년월일, 성별, 환자번호, 촬영기관명, 촬영일자·시간 존재</p>
4	외과병리 보고서	텍스트 (관찰입력정보, 자유입력텍스트)	500명 (11.5MB)		

→ 4. 외과병리보고서(비정형 텍스트데이터)는 자연어 처리기술 및 수기작업을 통해 정형데이터로 변경하여 활용하고자 함

**예시** 비정형 외과병리보고서의 정형화 데이터 변환

〈 비정형 외과병리보고서의 정형화 데이터 변환 과정 예시 〉



가명처리 대상 파일 (상세: 비정형데이터 ⇒ 정형데이터 변환)				
연번	항목명(영문)	항목명(한글)	예시	비고
<b>&lt;외과병리 정보&gt;</b>				
4-1	Organ	장기명	Breast	
4-2	Location	유방 수술의 위치	Right, left	
4-3	Tumor_size	종양의 크기	2.2cm	
4-4	Histologic_type	조직학적 유형	Invasive ductal carcinoma	
4-5	Histologic_grade	조직학적 등급	2/3	
4-6	Surgical_margins	수술절제면	deep margin	
~	~	~	~	~
4-17	Intraductal_comp	상피내암 유무	present	

가명처리 대상 파일(상세) (상세: 정형데이터)				
연번	항목명(영문)	항목명(한글)	예시	비고
<b>&lt;기본정보&gt;</b>				
5	gender	성별	F	
6	death	사망일자	19910525	
<b>&lt;암 등록정보&gt;</b>				
7	fdx	초진연월일	20200505	
8	age	암진단당시 나이	50	
~	~	~	~	
<b>&lt;수술 정보&gt;</b>				
20	sur_Date	수술일자	20200605	
21	sur_Date1	처방일자	20200605	
~	~	~	~	
<b>&lt;외과병리 정보&gt;</b>				
25	hist_date1	병리검사일자	20200605	
26	hist_date2	처방일자	20200605	
<b>&lt;DEXA 검사 결과&gt;</b>				
27	MIEX_YMD	검사일자	2017-04-21	
28	MIEX_CD	검사코드	RX211	
29	MIEX_NM	검사명	Dexa Bone Densitometry	
30	RGN_NM	검사부위명	L1-L4	
<b>&lt;CT영상 검사 결과&gt;</b>				
31	ORD_YMD	영상검사 처방일자	2020-08-24	
32	MIEX_YMD	영상검사 일자	2020-09-27	

【가명처리 계획】

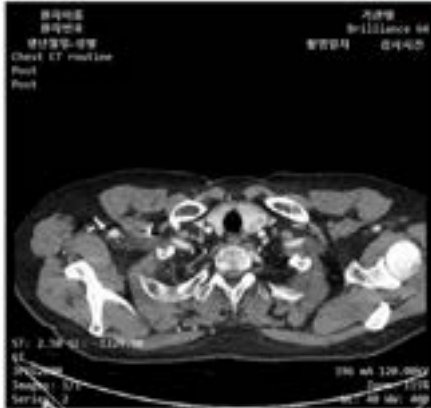


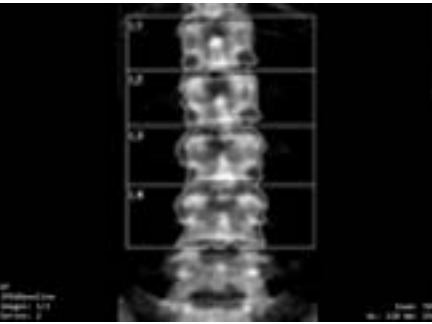
연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	병리조직 데이터	1) 비올슬라이드 사진	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		2) 병리슬라이드 번호 사진	<input checked="" type="checkbox"/> 삭제	분석에 필요 없고, 식별 가능성 있으므로 삭제
		3) 검체 슬라이드 사진	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		4) 검체 확대 사진	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
2	흉부 CT 데이터	1) 흉부 촬영부분	<input type="checkbox"/> 표면 가장자리 삭제 <input checked="" type="checkbox"/> 그대로 사용	개인당 200장의 사진이 촬영되어, 3차원 재건 등의 기술을 활용하면 신체 이미지를 복구할 수 있고 복원시 특이한 외형·흉터 등이 있는 환자의 경우 낮은 확률로 식별 가능성 존재 3차원 재건으로 인한 개인식별 가능성을 더욱 낮출 수는 있으나 이 경우, 데이터 손실로 인해 연구목적 달성이 어려움 이미지를 그대로 사용하되, 3차원 재건 기술을 적용할 수 없도록 처리 환경을 통제 * 클라우드 기반 폐쇄연구 분석환경에서만 활용, 철저한 외부 데이터·프로그램 반입 관리 수행
		2) 이미지 내 환자관련정보 * DICOM 형식 그대로 연구에 활용이 필요하기 때문에 블랙마스킹 처리가 불가하며, 이미지 내 표시된 DICOM 메타데이터를 손쉽게 일괄 변경할 수 있도록 자체개발한 도구를 활용하여 가명처리 수행		
		2-1) 환자이름	<input checked="" type="checkbox"/> 대체	연구에 필요 없는 정보이므로, 자체개발한 DICOM 데이터 변경 도구를 활용하여 Anonymized 값으로 대체 (블랙마스킹으로 삭제해도 무방)
	2-2) 생년월일	<input checked="" type="checkbox"/> 대체(일부)	“일자” 정보는 연구에 필요 없으므로 자체개발한 DICOM 데이터 변경도구를 활용하여, “00”으로 대체 (블랙마스킹으로 삭제해도 무방)	
	2-3) 환자성별	<input checked="" type="checkbox"/> 삭제·대체	여성환자만 분석에 사용하므로 남성 환자가 있다면 데이터를 삭제하고 이후 성별값은 구분에 의미가 없으므로, 자체개발한 DICOM 데이터 변경도구를 활용하여, “F”(Female) 값을 “S”(Sex)로 대체 (블랙마스킹으로 삭제해도 무방)	
	2-2) 환자번호	<input checked="" type="checkbox"/> 대체	환자구분만 되면 되므로 자체개발한 DICOM 데이터 변경도구를 활용하여, 단순일련번호로 대체	
	3	골밀도 검사 (DEXA) 데이터	1) 골밀도 촬영부분	<input checked="" type="checkbox"/> 그대로 사용
2) 이미지 내 환자관련정보 * PNG 포맷으로 연구에 활용할 계획으로, 파일 포맷 변환(DICOM→PNG) 후, 블랙마스킹 기법을 통해 이미지 내 환자관련정보 삭제				
2-1) 생년월일			<input checked="" type="checkbox"/> 마스크	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
2-2) 환자성별			<input checked="" type="checkbox"/> 마스크	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
2-3) 환자번호			<input checked="" type="checkbox"/> 마스크	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
2-4) 촬영기관명			<input checked="" type="checkbox"/> 마스크	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
2-5) 촬영일자·시간			<input checked="" type="checkbox"/> 마스크	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
4	외과병리 보고서	검사결과 내용	<input checked="" type="checkbox"/> 삭제	정형데이터로 변환시켜 분석에 필요한 주요 검사결과만 추출 (아래 4-1~17 참조)한 뒤 삭제

연번	항목명(영문)	항목명(한글)	처리방법	세부방법 및 처리수준
<b>&lt;외과병리 보고서&gt;</b>				
4-1	Organ	장기명	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-2	Location	유방 수술의 위치	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-3	Tumor_size	종양의 크기	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-4	Histologic_type	조직학적 유형	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-5	Histologic_grade	조직학적 등급	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-6	Surgical_margins	수술절제면	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
~	~	~	~	~
4-17	Intraductal_comp	상피내암 유무	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음

연번	항목명(영문)	항목명(한글)	처리방법	세부방법 및 처리수준	가명처리 전	가명처리 후
<b>&lt;개인 신상정보&gt;</b>						
5	gender	성별	<input checked="" type="checkbox"/> 삭제	분석에 필요 없으므로 삭제 처리		
6	death	사망일자	<input checked="" type="checkbox"/> 부분삭제	연도만 필요하므로 월일은 삭제 처리	20230404	2023
<b>&lt;암 등록정보&gt;</b>						
7	fdx	초진연월일	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
8	age	암진단당시 나이	<input checked="" type="checkbox"/> 범주화	39세 이하, 40-84세(1세단위), 85세 이상으로 구분	29세 50세 93세	39세 이하 50세 85세 이상
~	~	~	~	~		
<b>&lt;수술 정보&gt;</b>						
20	sur_Date	수술일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
21	sur_Date1	처방일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
~	~	~	~	~		
<b>&lt;외과병리 정보&gt;</b>						
25	hist_date1	병리검사일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
26	hist_date2	처방일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
<b>&lt;DEXA 검사 결과&gt;</b>						
27	MIEX_YMD	검사일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리		
28	MIEX_CD	검사코드	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
29	MIEX_NM	검사명	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
30	RGN_NM	검사부위명	<input checked="" type="checkbox"/> 그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
<b>&lt;CT영상 검사 결과&gt;</b>						
31	ORD_YMD	영상검사 처방일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리	20200610	202006
32	MIEX_YMD	영상검사 일자	<input checked="" type="checkbox"/> 부분삭제	연월만 필요하므로 일자는 삭제 처리	20200927	202009

〈서식⑤〉

가명처리 결과서 (비정형데이터 부분만 발췌)

비정형 데이터 전후 비교 샘플링		
항목명	처리 전	처리 후
흉부CT 이미지	 <p>1) 흉부촬영부분 2) DICOM 이미지 내 환자관련정보 2-1) 환자이름 2-2) 생년월일 2-3) 환자성별 2-4) 환자번호</p>	 <p>1) 흉부촬영부분 : 그대로 유지 2) DICOM 이미지 내 환자관련정보 2-1) 환자이름 : Anonymized로 대체 2-2) 생년월일 : 연도정보로만 변경 2-3) 환자성별 : 성별값을 "S"로 대체 2-4) 환자번호 : 환자 구별을 위한 단순 일련번호로 대체</p> <p>* 자체개발한 DICOM 데이터 변경도구 활용</p>
골밀도 검사 DEXA 이미지	 <p>1) 골밀도 촬영부분 2) 이미지 내 환자관련정보 2-1) 생년월일 2-2) 환자성별 2-3) 환자번호 2-4) 촬영기관명 2-5) 촬영일자·시간</p>	 <p>1) 골밀도 촬영부분 : 그대로 유지 2) 이미지 내 환자관련정보 2-1) 생년월일 : 블랙마스킹 2-2) 환자성별 : 블랙마스킹 2-3) 환자번호 : 블랙마스킹 2-4) 촬영기관명 : 블랙마스킹 2-5) 촬영일자·시간 : 블랙마스킹</p> <p>* PNG 포맷으로 연구에 활용할 계획으로, 파일 포맷 변환 (DICOM→PNG) 후, 이미지 블랙마스킹 기법을 통해 이미지 내 환자관련정보 삭제</p>

## 〈서식⑥〉

비정형데이터 추가검수 결과서 (흉부CT 데이터<sup>영상·이미지</sup>)

가명처리 개요			
원본 데이터 규모	100,000장 (500명 * 100장 * 2회)	원본 데이터 크기(용량)	80GB
원본 데이터 포맷	DICOM	처리 결과 데이터 포맷	DICOM
가명처리 적용 기술	한국대학교병원에서 자체 개발한 DICOM 데이터 변경 도구를 활용하여 이미지 내 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호) 대체처리		

추가검수 내용	
검수 기간	2023년 4월 1일 ~ 2023년 4월 7일
검수 장소	회의실 (원격으로 분석실의 가상컴퓨터 접속)
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input checked="" type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input checked="" type="checkbox"/> 휴리스틱 표본 검수
	- 검수는 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행 - 민감성이 높은 암진단 표본만 선별(500장)하고, 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행
처리된 결과 확인 사항	1. DICOM 이미지 내 환자관련정보가 다음과 같이 처리되었는지 전수 확인 ① 환자이름 : Anonymized 로 대체 ② 생년월일 : 연도정보만 표시되고, 일자는 '00'으로 대체 ③ 환자성별 : 환자성별값('M', 'F')을 'S'로 대체 ④ 환자번호 : 환자구별을 위한 단순일련번호로 대체  2. 위 4가지 사항 중 하나라도 만족하지 않았을 경우 DICOM 데이터 변경도구 또는 수작업을 통해 다시 한 번 추가처리를 수행  3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리
검수 결과	확인 결과 이상 없음 (검수 결과, 35장이 인식되지 않았으나, 수작업을 통해 최종 가명처리 완료)

검증자	소속 및 직위	성명	서명(인)
	한국대학교 개인정보보호부장	가관리	
	한국대학교 의료정보관리실장	나정보	
	한국대학교 개인정보보호부	다보호	
	한국대학교 개인정보보호부	라보호	
	한국대학교 개인정보보호부	마보호	

### 비정형데이터 추가검수 결과서 (골밀도 검사 데이터<sup>영상·이미지</sup>)

가명처리 개요			
원본 데이터 규모	1,000장 (500명 * 2장 * 1회)	원본 데이터 크기(용량)	1GB
원본 데이터 포맷	DICOM	처리 결과 데이터 포맷	PNG
가명처리 적용 기술	한국대학교병원에서 수집한 골밀도 검사(DEXA) 데이터를 DICOM 포맷에서 PNG 포맷으로 변환 후, 영상이미지 내 환자관련정보(생년월일, 환자성별, 환자번호, 촬영기관명, 촬영일자·시간)을 블랙마스킹 프로그램을 활용하여 마스킹처리		

추가검수 내용	
검수 기간	2023년 4월 10일 ~ 2023년 4월 11일
검수 장소	회의실 (원격으로 분석실의 가상컴퓨터 접속)
검수 방식 설명	<input checked="" type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input checked="" type="checkbox"/> 휴리스틱 표본 검수 - 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행 - 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행
처리된 결과 확인 사항	1. PNG 이미지 내 환자관련정보가 문제 없이 블랙마스킹 처리되었는지 전수 확인 - 생년월일, 환자성별, 환자번호, 촬영기관명, 촬영일자·시간 2. 처리가 완벽히 수행되지 않았을 경우 수작업으로 블랙마스킹 처리 수행
검수 결과	확인 결과 이상 없음 (검수 결과, 22장이 제대로 처리되지 않아, 수작업을 통해 최종 가명처리 완료)

검증자	소속 및 직위	성명	서명(인)
	한국대학교 개인정보보호부장	가관리	
	한국대학교 의료정보관리실장	나정보	
	한국대학교 개인정보보호부	다보호	
	한국대학교 개인정보보호부	라보호	
	한국대학교 개인정보보호부	마보호	

비정형데이터 추가검수 결과서 (외과병리보고서<sup>텍스트</sup>)

가명처리 개요			
원본 데이터 규모	500건	원본 데이터 크기(용량)	11.5MB
원본 데이터 포맷	DB	처리 결과 데이터 포맷	DB
가명처리 적용 기술	암 조직 관련 외과병리보고서를 자연어 처리 기술을 적용하여 정형데이터로 변환하여 활용		
추가검수 내용			
검수 기간	2023년 4월 13일 ~ 2023년 4월 14일		
검수 장소	회의실 (원격으로 분석실의 가상컴퓨터 접속)		
검수 방식 설명	<input checked="" type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수		
	- 검수는 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 2인이 함께 진행 - 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행		
처리된 결과 확인 사항	1. 17가지의 데이터 항목이 문제없이 인식되어 정형화되었는지 확인 2. 처리가 완벽히 수행되지 않을 경우 수작업으로 DB에 데이터 입력		
검수 결과	확인 결과 이상 없음 (검수 결과, 57건의 표본이 제대로 처리되지 않아, 수작업을 통해 최종 가명처리 완료)		
검증자	소속 및 직위	성명	서명(인)
	한국대학교 개인정보보호부장	가관리	
	한국대학교 의료정보관리실장	나정보	
	한국대학교 개인정보보호부	다보호	

## ② 구강질환 진단 AI 개발 사례

### ◆ 개요 (이미지, 영상 정보)

- 한국대학교병원 임상시험센터는 헬스케어 분야 AI개발 스타트업인 (주)헬스케어소프트로부터 과학적 연구 수행을 위한 데이터 제공을 요청받아 구강 건강검진 촬영 이미지 데이터를 제공하려 한다.

### ◆ 처리 목적

- 구강 내 충치 및 치주염 관련 질환을 진단하는 AI 개발 연구

### ◆ 데이터 특징

- (이미지 데이터) 구강 건강검진 촬영 사진
  - 5,000명의 구강 건강검진 촬영영상에서 각 10장씩 이미지를 추출하여 저장한 데이터 (JPEG포맷)
  - AI 모델 학습을 위해 각 구강 이미지의 구개, 혀, 위 아래 잇몸, 상악치, 하악치, 절치, 구치, 충치영역, 치주염 영역을 각각 라벨링처리하였음
  - 각 이미지에 촬영된 자에 대한 메타데이터(성별, 이름, 나이, 촬영 날짜) 포함

### ◆ 데이터 이용 환경

- (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가
- (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)
- (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

### ◆ 데이터 흐름



### 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	(주)헬스케어소프트 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	(주)헬스케어소프트 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	한국대학교병원 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	한국대학교병원 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	한국대학교병원 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	한국대학교병원 (가명정보 제공자)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	(주)헬스케어소프트 (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	한국대학교병원 (가명정보 제공자)  (주)헬스케어소프트 (가명정보 이용자)	작성	

◆ 주요 서식 작성 예시

〈서식 ①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	SG-2023011002	기존 접수번호(변경시)		접수일	2023년 3월 2일
신청자	조직/부서명	(주) 헬스케어소프트 데이터연구팀			
	담당자 직위	팀장	담당자 성명	홍길동	
	전화번호	03-1234-5678	이메일 주소	kildong@healthcaresoft.com	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발				
가명처리 대상 개인정보	개인정보 명칭	구강 건강검진 촬영 영상 및 이미지 데이터			
	요청 개인정보 항목	2022년 구강 건강검진 및 질환 치료를 위해 촬영한 영상 데이터 5만건을 대상으로 학습 데이터로 필요한 단층 이미지로 변환해둔 데이터를 가명처리·제공하여 구강 질환 진단 AI 모듈을 개발하는데 활용			
결합 활용 여부	<input type="checkbox"/> 결합전문기관 결합 예정 <input checked="" type="checkbox"/> 해당 없음 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출				
	<input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 4월 1일 ~ 2026년 4월 1일 (3년)				

2023년 1월 5일

신청인 (부서장명) 홍길동 (서명 또는 인)

한국대학교병원장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발 연구
책임자	소속: ㈜헬스케어소프트 데이터연구팀      성명: 홍길동      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>- 구강검사 시 단순한 기초 검사에도 대기와 진행시간이 소요되고 질환 치료의 의사결정까지 오래 걸림</li> <li>- 대형병원의 충치 및 치주염 등 기초검사 진행시간 단축 및 정확도 향상 필요</li> <li>- 다양한 구강질환의 진단, 초기 발견 등 치료에 필요한 구강검진 기록입력 자동화 등을 지원 가능</li> <li>- 국민의 Self 구강검진을 통한 조기검진 및 예방관리 지원 기술 개발 필요</li> </ul>
활용 데이터	<p>한국대학병원에서 구강 건강검진을 받은 5,000명의 내원자 데이터</p> <ul style="list-style-type: none"> <li>- 연구대상자 선정기간 (2012.1.1. ~ 2022.12.31.)</li> </ul>
활용 계획 및 방법	<ul style="list-style-type: none"> <li>- (데이터 전처리) 데이터를 전처리하여, 이미지 크기 조정, 노이즈 제거, 이미지 보정 등의 과정 수행</li> <li>- (데이터 라벨링) 데이터에 대해 라벨링 작업 수행, 라벨링 작업은 전문가들이 수행하며, 충치와 치주염 여부, 질환의 위치, 크기 등을 라벨링</li> <li>- (모델 개발) 데이터를 바탕으로 구강 질환 판별 모델 개발</li> <li>- (모델 평가) 개발한 모델을 평가(정확도, 재현율, F1-score 등의 지표 검증)</li> </ul>
(예상) 활용 기간	2023년 4월 1일 ~ 2026년 4월 1일 (3년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>- 구강진단 AI 모듈 개발로 진료 정확성 및 효과성 향상</li> </ul>

〈서식③〉

위험성 검토서

◆ 가명정보 활용 개요	
가명정보 활용 목적	구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발 연구
가명처리 대상 개인정보	한국대학병원에서 구강 건강검진을 받은 5,000명의 내원자 데이터
활용하려는 자	(기관명) ㈜헬스케어소프트 (직위) 팀장 (성명) 홍길동

(1) 기본 위험도 판단	
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input checked="" type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
기본 위험도 자동 도출	〈Step 1〉 내부 활용 → <input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input checked="" type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청(헬스케어소프트)에 따라 병원 내부 환경에 제공 예정

(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장

(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 최종 위험도를 고위험으로 상향 설정하고 외부전문가 인원을 포함한 적정성검토위원회를 별도로 구성하여 검토 후 해당 결과에 따라 제공하기로 결정		

검토일 : 2023 년 1 월 6 일

검토자 소속 : 빅데이터처리팀      성 명 : 나상호 (서명 또는 인)

## 〈서식④〉

## 가명처리 계획서

## 【가명처리 대상 파일】

가명처리 대상 파일(요약)					
연번	항목명	데이터 유형	데이터 규모	예시	비고
1	구강 촬영데이터	이미지 (JPEG 포맷)	이미지 50,000장 (5,000명*10장) * 개인 영상 당 단층 촬영물 10개의 이미지 파일 추출 * 개인당 평균 약 6MB, 총 300GB		비정형데이터 (메타데이터 포함)





가명처리 대상 파일(상세)					
연번	항목명	구분	설명	예시 (해당 부분 강조)	비고
1	구강 촬영데이터	① 구개	입천장		이미지 (비정형데이터)
		② 혀	구강의 바닥에 위치한 근육 조직		
		③ 윗잇몸	윗잇몸(상악점막)		
		④ 아랫잇몸	아랫잇몸(하악점막)		
		⑤ 상악치	윗니		
		⑥ 하악치	아랫니		
		⑦ 충치 영역	절치와 구치 중 일부		
		⑧ 치주염 영역	상악점막과 하악점막 중 일부		
1-1	성별		'남성', '여성'으로 구분	여성	구강 촬영데이터의 메타데이터
1-2	이름		내원자의 이름	홍길동	
1-3	나이		12세~89세까지 분포	27세	
1-4	촬영날짜		2012.1.1.~2022.12.31.까지 분포	2021.8.28.	

**【가명처리 계획】**

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	구강 촬영데이터	① 구개	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		② 혀	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		③ 윗잇몸	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		④ 아래잇몸	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		⑤ 상악치	<input checked="" type="checkbox"/> 블러링 처리	충치 영역 외 부분 블러링 처리
		⑥ 하악치	<input checked="" type="checkbox"/> 블러링 처리	충치 영역 외 부분 블러링 처리
		⑦ 충치 영역	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		⑧ 치주염 영역	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
1-1	성별	<input checked="" type="checkbox"/> 삭제	연구에 필요 없으므로 삭제	
1-2	이름	<input checked="" type="checkbox"/> 삭제	연구에 필요 없으므로 삭제	
1-3	나이	<input checked="" type="checkbox"/> 삭제	연구에 필요 없으므로 삭제	
1-4	촬영날짜	<input checked="" type="checkbox"/> 삭제	연구에 필요 없으므로 삭제	

**〈서식⑤〉**

**가명처리 결과서 (비정형데이터 부분만 발췌)**

비정형 데이터 전후 비교 샘플링		
항목명	처리 전	처리 후
구강 촬영 데이터	<p>⑤ 상악치</p> 	
	<p>○ 충치 영역 외 부분 블러링 처리 - A사의 '이미지 가명처리 솔루션' 활용 (블러링 수준 3단계 적용)</p>	
구강 촬영 데이터	<p>⑥ 하악치</p> 	
	<p>○ 충치 영역 외 부분 블러링 처리 - A사의 '이미지 가명처리 솔루션' 활용 (블러링 수준 3단계 적용)</p>	

## 〈서식⑥〉

## 비정형데이터 추가검수 결과서

가명처리 개요			
접수번호 (또는 처리명칭)	구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발 연구		
원본 데이터 규모	50,000장 (5,000명 * 10장)	원본 데이터 크기(용량)	300GB
원본 데이터 포맷	JPEG	처리 결과 데이터 포맷	JPEG
가명처리 적용 기술	- 한국대학교병원에서 구매한 A사의 '이미지 가명처리 솔루션'을 활용하여 각 이미지의 상악치·하악치로 라벨링된 부분을 블러링 처리(3단계 수준)		
추가검수 내용			
검수 기간	2023년 1월 20일 ~ 2023년 1월 27일		
검수 장소	병원 내 임상시험센터 분석PC		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input checked="" type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input checked="" type="checkbox"/> 휴리스틱 표본 검수		
	- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행 - 특히 구강구조(부정교합) 및 특이 케이스(다량의 충치/치주염 등)의 표본만 선별(350장)하고 실장 및 직원이 데이터를 나누어 블러링 처리 정상 수행 여부를 육안으로 검수한 뒤, 이상처리가 발생한 표본만 선별하여 전체 인원이 추가 합동검수·처리 수행		
처리된 결과 확인 사항	1. 구강검진 이미지 내 상악치·하악치가 다음과 같이 처리되었는지 전수 확인 ① 충치영역으로 라벨링된 부분을 제외한 상악치·하악치 모든 영역이 블러링처리 되었는지? ② 충치영역으로 라벨링된 부분까지 블러링처리되지는 않았는지? 2. 위 사항 중 하나라도 만족하지 않았을 경우 수작업으로 영역을 지정한 뒤, 솔루션을 활용하여 블러링 처리를 수행 3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리		
검수 결과	확인 결과 이상 없음 (검수 결과, 최초 26장이 제대로 처리되지 않았으나, 수작업을 통해 최종 가명처리 완료)		
검증자	소속 및 직위	성명	서명(인)
	한국대학교 개인정보보호부장	가관리	
	한국대학교 의료정보관리실장	나정보	
	한국대학교 개인정보보호부	다보호	
	한국대학교 개인정보보호부	라보호	
한국대학교 개인정보보호부	마보호		

### ③ 안면골 골절 진단 AI 개발 사례

#### ◆ 개요 (이미지, 영상 정보)

- 한국대병원 임상시험센터는 한국대학교병원 내 영상의학과 연구자와 헬스케어 분야 AI개발 스타트업인 (주)퓨처비전데이터(공동연구 컨소시엄)로부터 과학적 연구 수행을 위한 데이터 제공을 요청받아 안면골(얼굴뼈) 골절 관련 Facial CT 이미지·영상 등의 데이터를 제공하려 한다.
- 한국대학교병원 영상의학과 / (주)퓨처비전데이터 데이터연구팀 공동연구

#### ◆ 처리 목적

- 안면골 골절을 진단하는 AI 솔루션 개발 연구

#### ◆ 데이터 특징

- (영상·이미지 데이터) 안면골 골절 증상으로 CT검사를 받은 80,000명(2007년~2020년)의 환자에 대한 Facial CT 영상·이미지 파일 (DICOM포맷)
  - DICOM 헤더정보 포함 (환자번호, 환자이름, 성별, 생년월일 등)
- (증례기록지 데이터정형데이터) 안면골 골절 증상으로 Facial CT를 촬영한 환자에 대한 기본정보 (차트번호, 검사일자, 성별, 나이)와 진단명(진단코드)

#### ◆ 데이터 이용 환경

- (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가
- (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)
- (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

#### ◆ 데이터 흐름



### 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	AI 개발팀 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	AI 개발팀 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	한국대학교병원 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	한국대학교병원 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	한국대학교병원 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	한국대학교병원 (가명정보 제공자)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	AI 개발팀 (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	한국대학교병원 (가명정보 제공자)  AI 개발팀 (가명정보 이용자)	작성	

주요 서식 작성 예시

〈서식①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	SG-2023011058	기존 접수번호(변경시)		접수일	2023년 2월 1일
신청자	조직/부서명	한국대학교병원 영상학과			
	담당자 직위	연구원	담당자 성명	강진단	
	전화번호	03-1234-5678	이메일 주소	kjd@hankook.ac.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존 ※ 세부 활용 목적 안면골 골절에 대한 진단 AI 및 영상분석 솔루션 개발				
	가명처리 대상 개인정보	개인정보 명칭	Facial CT 영상·이미지(비정형) 및 증례기록지(정형)		
	요청 개인정보 항목	2007년부터 2020년까지 안면골절의 진단을 위해 Facial CT 검사를 받은 사람 중 안면골절 환자 40,000명과 정상인자 40,000명의 CT 영상·이미지 및 증례기록			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 ) <input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출 <input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 3월 1일 ~ 2025년 3월 1일 (2년)				

2023년 2월 1일

신청인 (부서장명) 유 관 순 (서명 또는 인)

한국대학교병원장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	안면골 골절에 대한 진단 AI 및 영상분석 솔루션 개발
책임자	소속: (주)퓨처비전데이터      성명: 박 비 전      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>• 2020년 심사평가원에서 제공한 보건의로 통계에 따르면 연간 70만명에 이르는 안면골 골절 환자가 발생하고 있음</li> <li>• 안면골 골절은 사고나 상해 등의 이유로 매우 빈번하게 일어나므로 CT 촬영이 가능한 모든 의료기관에서의 진단이 필요하나 많은 병원에서 실시간 판독이 가능한 전문의가 없는 경우가 많음</li> <li>• 안면골 골절 진단이 가능한 영상의학과, 성형외과, 구강외과 등의 전문의가 없더라도 당직 의사에게 빠른 의사결정을 위한 안면골 골절을 진단할 수 있는 AI솔루션 개발 필요</li> </ul>
활용 데이터	2007년부터 2020년까지 한국대학교병원에서 안면골 골절 의심 증상으로 Facial CT검사를 받은 환자의 CT촬영 영상 80,000건
활용 계획 및 방법	<ul style="list-style-type: none"> <li>• 본 연구는 후향적 연구로, 안면골 골절에 대한 AI를 이용한 판독을 지원하는 AI솔루션을 개발하는 것을 목표로 함</li> <li>• AI 기계학습과 대조군을 이용한 검증 수행</li> <li>• 안면골 골절 증상으로 CT검사를 받은 80,000건의 환자 데이터 활용 <ul style="list-style-type: none"> <li>- 안면골 골절진단을 받은 40,000건을 학습데이터 사용하고 대조군의 40,000건을 대상으로 검증 시행 예정</li> </ul> </li> <li>• 기계학습을 통해 S02의 다양한 진단에 대해 학습하고 검증분석할 예정임</li> </ul>
(예상) 활용 기간	• 2023년 3월 1일 ~ 2025년 3월 1일 (2년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>• Facial CT 분석을 통해 성형외과, 정형외과, 영상의학과 전문의가 없으나 CT를 보유하고 있는 병원에서 안면골 골절에 대한 진단에 도움을 받을 수 있음</li> <li>• 안면골 골절의 정확한 영상 판독을 통한 적절한 치료로 후유증 발생률 최소화 가능</li> </ul>

〈서식③〉

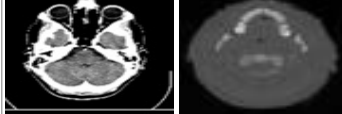
위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	안면골 골절에 대한 진단 AI 및 영상분석 솔루션 개발		
가명처리 대상 개인정보	Facial CT 영상·이미지(비정형) 및 증례기록지(정형)		
활용하려는 자	(기관명) (주)퓨처비전데이터 (직위) 책임 (성명) 박비전		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input checked="" type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input checked="" type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청(퓨처비전데이터)에 따라 병원 내부 환경에 제공 예정		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 최종 위험도를 고위험으로 상향 설정하고 외부전문가 인원을 포함한 적정성검토위원회를 별도로 구성하여 검토 후 해당 결과에 따라 제공하기로 결정		
검토일 : 2023 년 2 월 3 일			
검토자 소속 : 빅데이터처리팀    성명 : 나상호 (서명 또는 인)			


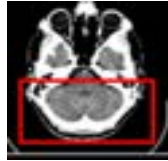
## 〈서식④〉

## 가명처리 계획서

## 【가명처리 대상 파일】

가명처리 대상 파일(요약)					
연번	항목명	데이터 유형	데이터 규모	예시	비고
1	Facial CT 데이터	영상·이미지 (DICOM*)	80,000건 (97TB) - 안면골 골절 판정 데이터 (40,000건) - 정상 판정(안면골 골절X) 데이터 (40,000건)		비정형데이터 (DICOM 헤더정보 포함)
2	증례기록지	텍스트·숫자 등 (정형)	80,000건 (52MB)	차트번호, 검사일자, 환자성별, 환자나이, 진단코드 포함	정형데이터

\* DICOM(Digital Imaging and Communications in Medicine) : 의료용 디지털 영상 및 통신 표준

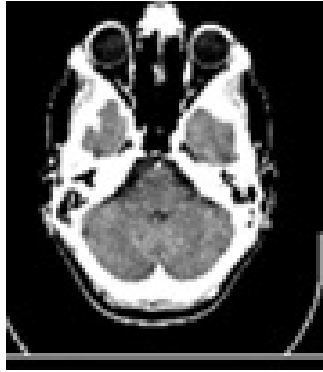
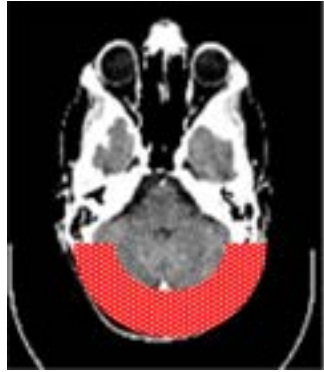
가명처리 대상 파일(상세)				
연번	항목명	구분	설명	예시
1	Facial CT 데이터	① 안면부	안면골 골절 여부 판단에 필요한 안면부 촬영부분	
		② 후두부 (뇌 뒷부분)	안면골 골절 여부 판단에 필요없는 후두부(뇌 뒷부분) 촬영부분	
1-1	DICOM 헤더정보	환자 번호	환자 구분을 위해 고유하게 부여되는 번호	P158687
1-2		환자 이름	환자의 이름	김철수
1-3		환자 성별	1: 남자, 2: 여자	1
1-4		환자 생년월일	환자의 생년월일 정보	1968.5.26.
2-1	증례기록지	① 차트번호	환자진료기록 차트에 고유하게 부여되는 번호로, 각 환자의 Facial CT 사진에 매핑하여 활용	01768062
2-2		② 검사일자	안면골 골절에 대한 Facial CT 검사일자	2019-06-23
2-3		③ 환자 성별	1: 남자, 2: 여자	2
2-4		④ 환자 나이	1세 단위, 12세~98세까지 분포	58
2-5		⑤ 진단명 (진단코드)	<ul style="list-style-type: none"> <li>▪ S02 두개골 및 안면골의 골절</li> <li>▪ S02.0 두개원개의 골절</li> <li>▪ S02.1 두개저의 골절</li> <li>▪ S02.2 비골의 골절</li> <li>▪ S02.3 안와바닥의 골절</li> <li>▪ S02.4 광대뼈 및 상악골의 골절</li> <li>▪ S02.5 치아의 파절</li> <li>▪ S02.6 하악골의 골절</li> <li>▪ S02.7 두개골 및 안면골을 침범한 다발골절</li> <li>▪ S02.8 기타 두개골 및 안면골의 골절</li> <li>▪ S02.9 두개골 및 안면골의 상세불명 부분의 골절</li> </ul>	S02

**【가명처리 계획】**

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	Facial CT 데이터	① 안면부	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
		② 후두부 (뇌뒷부분)	<input checked="" type="checkbox"/> 마스크	연구에 필요 없으므로, 마스크 처리하여 안전하게 활용
1-1	DICOM 헤더정보	환자 번호	<input checked="" type="checkbox"/> 삭제(포맷변환)	연구에 필요 없으므로 삭제
1-2		환자 이름	<input checked="" type="checkbox"/> 삭제(포맷변환)	- DICOM 포맷을 TIFF 포맷으로 변경하여 저장(헤더 삭제)
1-3		환자 성별	<input checked="" type="checkbox"/> 삭제(포맷변환)	증례기록지를 통해 확인 가능한 정보이므로 삭제
1-4		환자 생년월일	<input checked="" type="checkbox"/> 삭제(포맷변환)	- DICOM 포맷을 TIFF 포맷으로 변경하여 저장(헤더 삭제)
2-1	증례기록지	① 차트번호	<input checked="" type="checkbox"/> 대체	환자 구분을 위해 단순일련번호로 대체
2-2		② 검사일자	<input checked="" type="checkbox"/> 범주화	연 단위로 범주화
2-3		③ 환자 성별	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음
2-4		④ 환자 나이	<input checked="" type="checkbox"/> 범주화 (10세 단위)	10세 단위 범주화 적용, 90세 이상은 90대로 상단코딩 적용
2-5		⑤ 진단명 (진단코드)	<input checked="" type="checkbox"/> 그대로 사용	별도 처리하지 않음

〈서식⑥〉

**가명처리 결과서 (비정형데이터 부분만 발췌)**

비정형 데이터 전후 비교 샘플링			
연번	항목명	처리 전	처리 후
1	Facial CT 데이터		

## 〈서식⑥〉

## 비정형데이터 추가검수 결과서

가명처리 개요			
원본 데이터 규모	Facial CT 영상·이미지 80,000건	원본 데이터 크기(용량)	97TB
원본 데이터 포맷	DICOM	처리 결과 데이터 포맷	TIFF
가명처리 적용 기술	한국대학교병원에서 구매한 B사의 '이미지 마스크 솔루션'을 활용하여 각 환자의 후두부 영역만 마스크 처리		
추가검수 내용			
검수 기간	2023년 2월 20일 ~ 2023년 2월 27일		
검수 장소	회의실 (원격으로 분석실의 가상컴퓨터 접속)		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input checked="" type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수		
	- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행 - 전체 표본 중 10% 표본만 선별(8,000장)하고 실장 및 직원이 데이터를 나누어 마스크 처리 정상 수행 여부를 육안으로 검수한 뒤, 이상처리가 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행		
처리된 결과 확인 사항	1. Facial CT 후두부 영역이 다음과 같이 처리되었는지 전수 확인 ① 후두부 영역 전체가 확인 불가능하도록 마스크처리 되었는지? ② 안면부 영역까지 마스크처리되지는 않았는지?  2. 위 사항 중 하나라도 만족하지 않았을 경우 수작업으로 영역을 지정한 뒤, 솔루션을 활용하여 마스크 처리 수행  3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리		
검수 결과	8,000개 파일 중 6개 파일에 대해 오류 발견되어 추가 가명처리 실시 완료 ⇒ 처리가 미흡하여 재처리한 사례 중 개인식별 위험이 특별하게 높은 건은 발견되지 않았음(일부 후두부 영역만 처리). 안전한 내부 환경에서 분석되므로 추가 표본검수·재처리 없이 해당 잔존 위험은 감수하기로 결정		
검증자	소속 및 직위	성명	서명(인)
	한국대학교 개인정보보호부장	가관리	
	한국대학교 의료정보관리실장	나정보	
	한국대학교 개인정보보호부	다보호	
	한국대학교 개인정보보호부	라보호	
한국대학교 개인정보보호부	마보호		

## ④ 자율주행차 주행 시 비정상 상황인지 AI 개발 사례

### ◆ 개요 (이미지, 영상 정보)

- 한국영상연구원은 국가 R&D 사업의 일환으로 자율주행차를 통해 한국대학교 교내 도로 주행영상을 촬영하여 보유하고 있다(근거법률: 「개인정보 보호법」 제25조의2 제2호).  
한국영상연구원은 자율주행차의 주행환경에 대한 비정상 상황인지 AI 기술개발을 수행하고자 하는 (주)한국자율테크에 해당 영상을 제공하고자 한다.

### ◆ 처리 목적

- 자율주행자동차의 주행환경에 대한 비정상 상황 인지 AI 기술 연구
- 과학적 연구 목적으로 가명처리 후 AI 기술개발을 위한 인공지능 신경망 학습용 데이터로 활용

### ◆ 데이터 특징

- (이미지·영상 데이터) '22년 1월~2월 동안 자율주행차가 한국대학교 교내 도로 주행 상황을 촬영한 영상
  - 차량 탑재 카메라로부터 MP4데이터를 취득하고 이를 PNG파일로 생성(327GB, 5만장)
  - 이미지·영상 내 객체는 정적 객체(신호등, 도로 등)와 동적 객체(사람, 차량, 현수막 등)으로 구성
  - 이미지·영상과 관련된 메타데이터는 별도로 생성·저장되지 않음

### ◆ 데이터 이용 환경

- (한국영상연구원 내부 분석실 이용) USB보안 기능이 설치된 PC에서 관리하며, 이용자는 외부망이 차단된 PC에서 이용 및 활용
  - \* 물리적 망분리, 가상 단말기(PC) 보안 준수
- (자료 반입) 자료 반입(오프라인으로 전달)시 한국영상연구원 관리자에게 요청(관리자가 자료 확인 후 반입), 사전 특정된 연구원 외 데이터 접근 불가
- (자료 반출) 분석결과 반출 시 한국영상연구원 관리자에게 요청(관리자 자료 확인 후 제공)

### ◆ 데이터 흐름



### ◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	(주)한국자율테크 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	(주)한국자율테크 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	한국영상연구원 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	한국영상연구원 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	한국영상연구원 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	한국영상연구원 (가명정보 제공자)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	(주)한국자율테크 (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	한국영상연구원 (가명정보 제공자)  (주)한국자율테크 (가명정보 이용자)	작성	

주요 서식 작성 예시

〈서식 ①〉

이용·제공 신청서 (  신규  변경 )

접수번호	IMAGEINSTI_001	기존 접수번호(변경시)		접수일	2023년 1월 3일
신청자	조직/부서명	(주)한국자율테크			
	담당자 직위	책임	담당자 성명	이테크	
	전화번호	03-1234-5678	이메일 주소	lee_korea@imageinti.re.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 자율주행차의 비정상 주행 상황인지 AI 개발				
가명처리 대상 개인정보	개인정보 명칭	도로 주행 영상 데이터			
	요청 개인정보 항목	한국대학교 교내에서 주간과 야간, 새벽 시간대에 취득한 도로 주행 영상 데이터			
결합 활용 여부	<input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출				
	<input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 02월 01일 ~ 2023년 12월 31일(11개월)				

2023년 1월 3일

신청인 (부서장명) 이 테 크 (서명 또는 인)

한국영상연구원장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	자율주행차의 비정상 주행 상황인지 AI 개발
책임자	소속: 한국자율테크      성명: 이 테 크 (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>• (주)한국자율테크에서는 2023년 1월부터 산업부 R&amp;D 사업의 일환으로 클라우드 기반 자율주행차 인지 기술 개발을 진행하고 있음</li> <li>• 상기 과제에서 본 기관은 자율주행차의 주행 경로상에서 발생할 수 있는 비정상 상황을 인지하여 자율주행차가 원활한 주행을 할 수 있도록 상황 인지 SW 개발을 추진하고 있음</li> <li>• 본 데이터를 자율주행차의 주행 경로상에서 발생할 수 있는 비정상적인 상황을 인지할 수 있는 인공지능 SW 개발의 학습데이터로 활용하고자 함</li> </ul>
활용 데이터	2022년 1월부터 2월까지 한국대학교 교내 캠퍼스에서 새벽, 주간, 야간에 수집한 자율주행 영상(20시간)·이미지(50,000장) 데이터
활용 계획 및 방법	<ul style="list-style-type: none"> <li>• 데이터의 학습에 필요한 대상 객체를 가공(라벨링)하여 활용</li> <li>• 상황 인지를 위한 딥러닝 인공지능 알고리즘의 학습용 데이터를 활용하며, 일부 데이터는 테스트용으로 활용함</li> <li>• 자율주행차 주행환경 비정상 상황인지 기술개발             <ul style="list-style-type: none"> <li>① 주행 경로 동적, 정적 객체인지                 <ul style="list-style-type: none"> <li>- 자율주행차의 OEDR와 OEDR에 명시된 상황을 통해 예외 상황을 정의</li> <li>- 예외 상황을 구성하는 객체(사람, 자동차, 쓰레기 더미 등)를 분류하고, 가명처리 데이터에서 학습용으로 활용할 수 있는 객체들을 구분함</li> <li>- 상기 분류된 객체들을 가명처리 데이터에서 라벨링하고, 인공지능 학습데이터로 활용</li> </ul> </li> <li>② 비정상 상황인지 기술 개발                 <ul style="list-style-type: none"> <li>- 상기에서 인지한 객체 정보를 활용하고, 객체 간의 관계와 시간, 장소, 날씨 등의 연관 관계를 통해 OEDR에 명시된 예외 상황 인지</li> </ul> </li> </ul> </li> </ul>
(예상) 활용 기간	2023년 02월 01일 ~ 2023년 12월 31일 (11개월)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>• 자율주행차의 상용화를 위한 인지 학습데이터로 활용함으로써 향후, 자율주행 모빌리티 산업의 활성화에 기여할 수 있음</li> <li>• 가명처리 데이터는 비정상 상황인지와 더불어 자율주행차의 동적 객체 인지에 활용될 수 있음</li> </ul>

〈서식③〉





위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	자율주행차의 비정상 주행 상황인지 AI 개발		
가명처리 대상 개인정보	도로 주행 영상 데이터		
활용하려는 자	(기관명) 한국자율테크 (직위) 책임 (성명) 이테크		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input checked="" type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input checked="" type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청(한국자율테크)에 따라 연구원 내부 환경에 제공 예정		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	내부 분석환경에 제공하여 비교적 안전함에도 불구하고 비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 고위험으로 판단하고 외부전문가 인원을 포함한 적정성검토 위원회를 별도로 구성하여 검토 후 활용하는 것으로 결정		
검토일 : 2023 년 1 월 10 일			
검토자 소속 : 개인정보보호팀    성명 : 박검토 (서명 또는 인)			

## 〈서식④〉

## 가명처리 계획서

## 【가명처리 대상 파일】

가명처리 대상 파일(요약)			
항목명		데이터 유형	데이터 규모
자율주행차 촬영 영상 및 이미지		영상(MP4), 이미지(PNG)	영상: 20시간 분량 이미지 : 50,000장(327GB)
연번	항목명	설명	예시
1	사람 전체형상	사람의 신체·형상 전체	
2	사람 얼굴	사람의 정면과 측면 얼굴	 ※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음
3	차량 전체형상	차량의 형상 전체	
4	차량 번호판	차량의 번호판 영역	 ※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음

**【가명처리 계획】**

연번	항목명	처리방법	세부방법 및 처리수준
1	사람 전체형상	<input checked="" type="checkbox"/> 그대로 사용	• 연구목적 달성을 위해 반드시 필요하며, 처리환경의 안전성을 고려하였을 때 개인식별 위험이 높지 않으므로 그대로 사용
2	사람 얼굴	<input checked="" type="checkbox"/> 마스크	• 얼굴 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 마스크 처리
3	차량 전체형상	<input checked="" type="checkbox"/> 그대로 사용	• 연구목적 달성을 위해 반드시 필요하며, 처리환경의 안전성을 고려하였을 때 개인식별 위험이 높지 않으므로 그대로 사용
4	차량 번호판	<input checked="" type="checkbox"/> 마스크	• 차량 번호판 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 마스크 처리

〈서식⑤〉

**가명처리 결과서 (비정형데이터 부분만 발췌)**

비정형 데이터 전후 비교 샘플링			
연번	항목명	처리 전	처리 후
2	사람 얼굴		
4	차량 번호판		

※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음

## 〈서식⑥〉

## 비정형데이터 추가검수 결과서

가명처리 개요			
원본 데이터 규모	영상 : 20시간 분량 이미지 : 50,000장	원본 데이터 크기(용량)	327GB
원본 데이터 포맷	MP4, PNG	처리 결과 데이터 포맷	MP4, PNG
가명처리 적용 기술	'임베디드 장치 기반 영상데이터 개인정보 비식별화 시스템'을 활용하여 영상의 사람 얼굴과 차량번호판을 인식하고, 마스킹 기법을 적용하여 가명처리		
추가검수 내용			
검수 기간	2023년 01월 10일 ~ 2023년 01월 20일		
검수 장소	한국영상연구원 내(외부 망 차단) 보안이 적용된 PC(보안 USB)		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input checked="" type="checkbox"/> 휴리스틱 표본 검수		
	- 가명처리 SW 적용 후 사람의 움직임이 탐지되는 구간 중 사람이 지나다니는 시간대(08:00~20:00)에 촬영된 영상·이미지만 선별하여 5명의 검수자가 10일간 검수 → 객체가 인식되지 않아 마스킹 처리가 안된 경우 검토 → 마스킹 처리가 제대로 되지 않아 얼굴·차량번호판이 식별 가능한 경우 검토 - 전수조사 이후 마스킹 처리되지 않은 파일은 마스킹 SW로 재처리하였으며, 재처리되지 않는 파일은 수작업을 통해 마스킹 처리함		
처리된 결과 확인 사항	- 조사 보고서에 대해 책임자가 검토 후 조사 완료		
검수 결과	선별한 이미지·영상 중 미흡하게 처리된 부분 모두 개별 가명처리 실시 완료 ⇒ 처리가 미흡하여 재처리한 사례 중 개인식별 위험이 특별하게 높은 건은 발견되지 않았으며, 안전한 내부 환경에서 분석되므로 추가 표본검수·재처리 없이 해당 잔존 위험은 감수하기로 결정		
검증자	소속 및 직위	성명	서명(인)
	자율주행연구실 책임	가관리	
	자율주행연구실 선임	나정보	
	자율주행연구실 선임	다보호	
	자율주행연구실 주임	라보호	
	자율주행연구실 주임	마보호	

## 5 고속도로 다인승전용차로 단속 AI 개발 사례

### ◆ 개요 (이미지 정보)

- A지자체의 교통정보센터는 교통단속 및 교통정보의 수집·분석 등을 위하여 고속도로에 CCTV를 설치하고 고속도로 통행차량 영상·이미지를 촬영하여 보관하고 있다(근거법률: 「개인정보 보호법」 제25조 제1항 제4호, 제5호). A지자체는 AI솔루션 개발 전문기업 COREA-AI사에 과학적 연구 수행을 위한 데이터 제공을 요청받아 해당 고속도로 통행차량 영상·이미지를 제공하려 한다.

### ◆ 처리 목적

- 재차인원 검지기반 다인승전용차로 단속 AI 개발 (학습데이터로 활용)

### ◆ 데이터 특징

- (이미지 데이터) 고속도로 통행 차량을 촬영한 이미지 (120,000장)
  - 재차인원 검지만을 위해 별도 제작된 카메라를 활용해 촬영된 사진으로, 차량번호판이 포함된 차량 앞부분은 촬영되지 않음
  - 탑승된 인원의 얼굴이 함께 촬영된 경우가 많음
- \* 얼굴이 선명하게 촬영된 경우, 흐릿하게 촬영된 경우, 그림자·장애물 등으로 절반만 보이는 경우 등 다양
  - 각 이미지마다 메타데이터 존재(촬영일시, 촬영장소)

### ◆ 데이터 이용 환경

- (망분리 폐쇄형 개발 환경 활용) COREA-AI 개발본부에서 구축한 Private 클라우드 기반의 망분리된 폐쇄형 AI 개발 환경에서 데이터를 이용 및 활용
  - 취급자에게만 접근 권한 부여, 접근통제 솔루션을 통해 비인가자의 접근 통제
- (자료 반입) COREA-AI 관리자의 검토 및 승인 절차를 수립하여 운영(관리자가 자료 확인 후 반입)
- (자료 반출) 분석결과 반출 시 COREA-AI 관리자에게 요청(관리자 자료 확인 후 제공)

### ◆ 데이터 흐름



### ◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	COREA-AI (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	COREA-AI (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	A지자체 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	A지자체 (가명정보 제공자)	작성	
	서식⑤ 가명처리 결과서	A지자체 (가명정보 제공자)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	A지자체 (가명정보 제공자)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서	COREA-AI (가명정보 이용자)	작성	
	서식⑩ 가명(추가)정보 관리대장	A지자체 (가명정보 제공자) COREA-AI (가명정보 이용자)	작성	

주요 서식 작성 예시

〈서식 ①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	SG-202305081	기존 접수번호(변경시)		접수일	2023년 5월 8일
신청자	조직/부서명	COREA-AI 개발본부 인공지능개발팀			
	담당자 직위	수석	담당자 성명	박채주	
	전화번호	03-1234-5678	이메일 주소	chaejoo@corea-ai.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 재차인원 검지기관 다인승전용차로 단속에 활용될 AI 기술개발				
가명처리 대상 개인정보	개인정보 명칭	고속도로 통행 차량 사진			
	요청 개인정보 항목	2022년 4월부터 2023년 4월까지 1년간 A지자체 OO고속도로 통행 차량에 대한 스냅사진 약 12만장			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input type="checkbox"/> 기관 내부 분석 공간 <input checked="" type="checkbox"/> 기관 외부 반출				
	<input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 6월 1일 ~ 2025년 6월 1일(2년)				

2023년 5월 8일

신청인 (부서장명) 박 채 주 (서명 또는 인)

A지방자치단체장 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	재차인원 검지기반 다인승전용차로 단속 AI 개발
책임자	소속: COREA-AI 개발본부 인공지능개발팀      성명: 박 채 주      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>고속도로의 수용 효율을 증대하기 위하여 다인승전용차로 제도 시행 중(1995년 2월~) - 9인승이상 승용자동차 및 11,12인 승합자동차의 경우, 6인 이상이 실제 승차하였을 때만 다인승전용차로 이용 가능</li> <li>그러나 다인승전용차로 이용자의 상당 수가 불법으로 이용 중인 것으로 파악됨</li> <li>불법이용을 단속하고 교통흐름을 개선하기 위해 차량통행사진을 기반으로 재차인원을 파악하고 불법 이용자를 단속할 수 있는 시스템 개발 필요</li> </ul>
활용 데이터	2022년 4월부터 2023년 4월까지, A지자체 OO고속도로 통행 차량을 촬영한 이미지(12만장)
활용 계획 및 방법	<ul style="list-style-type: none"> <li>A지자체가 고속도로 통행 차량 촬영 사진을 가명처리하여 COREA-AI에 제공</li> <li>COREA-AI 개발본부 내 담당부서에서 라벨링 진행</li> <li>라벨링 결과를 받아서 AI 학습 수행</li> <li>수행 결과를 바탕으로 인원계수 알고리즘 개발</li> <li>다인승전용차로 시행효과 극대화를 위해 재차인원 검지기술 개발 및 실증 수행</li> <li>다인승 전용차로 단속기능을 구현하기 위해               <ol style="list-style-type: none"> <li>영상촬영 기술 개발</li> <li>학습데이터 수집, 라벨링 및 학습</li> <li>차량 내 인원계수 알고리즘 개발</li> </ol> </li> <li>다인승 전용차로 단속시스템 관련 AI 모듈 개발</li> </ul>
(예상) 활용 기간	2023년 6월 1일 ~ 2025년 6월 1일(2년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>고속도로 버스전용차로 운영 효율화 및 단속에 따른 불필요한 지체를 감소</li> <li>재차인원 검지기반 기술을 통해 재해재난 시 조난 대상자 파악 등 사회적 현안 해결에 활용 가능</li> </ul>

〈서식③〉


### 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	재차인원 검지기반 다인승전용차로 단속 시 개발		
가명처리 대상 개인정보	A지자체 OO고속도로 통행 차량을 촬영한 이미지		
활용하려는 자	(기관명) COREA-AI 개발본부 인공지능개발팀 (직위) 수석 (성명) 박채주		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input checked="" type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input checked="" type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input checked="" type="checkbox"/> 고위험
	※ 판단 이유 : 외부 요청기관에게 기관 외부 반출		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input checked="" type="checkbox"/> 유지	<input type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 기본 위험도 판단을 유지하고 외부전문가 인원을 포함한 적정성검토위원회를 별도로 구성하여 검토 후 활용하는 것으로 결정		
검토일 : 2023 년 5 월 8 일			
검토자 소속 : 개인정보보호팀    성명 : 윤태훈 (서명 또는 인)			

## 〈서식④〉

## 가명처리 계획서

## 【가명처리 대상 파일】





가명처리 대상 파일(요약)					
연번	항목명	구분	설명	예시 (해당 부분 강조)	비고
1	고속도로 동행차량 촬영 데이터	① 차량 외형	특수차, 차종 등을 판단하기 위해 차량 외형이 촬영된 부분		이미지 (비정형데이터)
		② 차량 탑승자	차량 내에 탑승한 사람이 촬영된 부분		
		③ 주변 환경	차량, 탑승자 영역 외 도로 주변환경이 촬영된 부분		
1-1	촬영 일시	사진이 촬영된 날짜 및 시간	2023.1.2.14:00	각 사진에 대한 메타데이터	
1-2	촬영 장소	사진이 촬영된 주소 정보	00시 00동 00로 123-4		

## 【가명처리 계획】

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	고속도로 동행차량 촬영 데이터	① 차량 외형	<input checked="" type="checkbox"/> 특이정보 파일 표본 제외	• 차량 외형이 일반적이지 않고, 특이하여 탑승자의 개인식별 위험성이 비교적 높은 경우(회사로고가 크게 적혀있는 경우, 소방차·구급차 등 특수차량 등)해당 파일을 표본에서 제외
		② 차량 탑승자	<input checked="" type="checkbox"/> 블러링 처리	• 탑승자의 위치 및 범위를 파악하고 얼굴·상체 등 차량 탑승자 영역 전체를 블러링 처리 (블러링 수준 5단계 적용)
		③ 주변 환경	<input checked="" type="checkbox"/> 그대로 사용	• 연구목적 달성에 필요한 정보이며 차량 외형(특이한 경우 연구표본 에서 삭제), 차량 탑승자(블러링 처리), 메타데이터(삭제)를 가명 처리하므로 타 정보와의 결합을 통한 개인식별 가능성이 거의 없어 별도 처리하지 않고 그대로 사용
1-1	촬영 일시	<input checked="" type="checkbox"/> 삭제	• 연구에 필요 없으므로 삭제	
1-2	촬영 장소	<input checked="" type="checkbox"/> 삭제	• 연구에 필요 없으므로 삭제	

〈서식 5〉

가명처리 결과서 (비정형데이터 부분만 발췌)

비정형 데이터 전후 비교 샘플링				
연번	항목명	구분	처리 전	처리 후
1	고속도로 통행차량 촬영 데이터	① 차량 외형		<p>연구 표본에서 제외 (연구에 활용하지 않음)</p> <p>※ 특수차량(응급차, 경찰차, 회사차 등), 특이한 스티커·로고·광고 등이 포함된 차량 등</p>
		② 차량 탑승자		 <p>(탑승자 위치·범위 파악)</p> <p>▼</p>  <p>(1~10단계 중 5단계 수준 블러링 처리)</p>

## 〈서식⑥〉

## 비정형데이터 추가검수 결과서

가명처리 개요			
원본 데이터 규모	120,000장	원본 데이터 크기(용량)	18.31GB
원본 데이터 포맷	JEPG	처리 결과 데이터 포맷	JEPG
가명처리 적용 기술	<ul style="list-style-type: none"> <li>- A지자체에서 자체개발한 '이미지 블러링 프로그램'을 활용하여 각 이미지 내 차량탑승자 영역을 블러링 처리(5단계 수준)</li> <li>- 차량특이치는 직접 검수하여 삭제</li> </ul>		
추가검수 내용			
검수 기간	2023년 5월 12일 ~ 2023년 5월 15일		
검수 장소	A지자체 사무실		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input checked="" type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수		
	<ul style="list-style-type: none"> <li>- 가명처리 SW 적용 후 1% 표본(1,200장)만 선별하여 4명의 검수자가 3일간 검수               <ul style="list-style-type: none"> <li>① 차량의 색상, 특수차량, 회사로고 등 특이한 차량인가?                   <ul style="list-style-type: none"> <li>→ 해당 시 표본 삭제 (연구 대상에서 제외)</li> </ul> </li> <li>② 탑승자를 식별할 수 없도록 블러링 처리가 되었는가?                   <ul style="list-style-type: none"> <li>→ 제대로 처리되지 않았을 시 표본 삭제 (연구 대상에서 제외)</li> </ul> </li> </ul> </li> <li>- 검수 이후 파일대상 진단 툴 조사 및 보고서 작성 (처리된 이미지의 영역정보를 기준으로 원본 이미지와 가명처리된 이미지의 영역을 비교하여 진단)</li> </ul>		
처리된 결과 확인 사항	차량외형의 특이치가 존재하여 표본을 삭제한 경우 : 252장 탑승자 블러링처리가 제대로 안되어 표본을 삭제한 경우 : 131장 나머지 이미지는 정상 처리되어 연구 활용 가능		
검수 결과	선별한 이미지 중 미흡하게 처리된 부분 모두 개별 가명처리 실시 완료 ⇒ 처리가 미흡하여 재처리한 사례 중 개인식별 위험이 특별하게 높은 건은 발견되지 않았으며, 안전한 내부 환경에서 분석되므로 추가 표본검수·재처리 없이 해당 잔존 위험은 감수하기로 결정		
검증자	소속 및 직위	성명	서명(인)
	개인정보보호팀 팀장	박지원	
	개인정보보호팀 과장	최우선	
	AI 품질팀 인턴	김용두	
	AI 품질팀 인턴	이용미	

## ⑥ 한국어 대화가 가능한 AI 챗봇 개발

### ◆ 개요 (대화·검색분야 - 텍스트 정보)

- 인공지능 챗봇 전문기업 (주)한국데이터테크는 그간 채팅업을 통해 수집한 고객 간의 다양한 대화 텍스트 데이터를 가명처리하여 한국어 대화가 가능한 AI 챗봇 개발에 활용하고자 한다.

### ◆ 처리 목적

- 한국어 대화가 가능한 AI 챗봇 언어모델 개발

### ◆ 데이터 특징

- 채팅업을 통해 수집된 고객들 간 일상 대화 데이터(텍스트)
  - 고객 1,500명이 대화한 대화 데이터셋 총 20,000개 파일
  - 대화 당사자 간의 일상 대화 내용이 포함되어 있으며 데이터에 정해진 형식과 법칙성이 없고 대화 주제가 다양함
    - \* 격식을 갖추지 않은 일상의 구어체 대화가 대부분이며 한국어의 문법, 어법, 철자 등을 철저히 준수하기 보다는 편리함에 초점을 둔 대화 데이터임
  - 대화 당사자의 이름, 연락처, 계좌번호, 사생활의 영역까지 개인식별위험이 있는 다양한 정보가 포함되어 있음
  - 대화파일마다 대화에 참여한 사용자 계정정보가 메타데이터로 포함

### ◆ 데이터 이용 환경

- (폐쇄형 내부 개발환경 활용) (주)한국데이터테크 내부 공간에 마련한 철저한 폐쇄망 환경에서만 데이터 활용
  - \* 취급자에게만 접근권한 부여, 접근통제 솔루션으로 비인가자 접근 통제
- (자료 반입) 내부 보안팀의 검토 및 승인 절차에 따라 운영(관리자가 자료 확인 후 반입)
- (자료 반출) 분석결과 반출 시 관리자에게 요청(관리자 자료 확인 후 제공)

### ◆ 데이터 흐름



### ◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	AX팀 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	AX팀 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	서비스운영팀 (가명정보 제공자)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	개인정보보호팀 (가명처리 부서)	작성	
	서식⑤ 가명처리 결과서	개인정보보호팀 (가명처리 부서)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	개인정보보호팀 (가명처리 부서)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	내부 규정에 따라 관리
	서식⑩ 가명(추가)정보 관리대장	개인정보보호팀 (개인정보보호 담당 부서)	작성	

주요 서식 작성 예시

〈서식 ①〉

가명정보 이용·제공 신청서 (  신규  변경 )

접수번호	SG-20230110083	기존 접수번호(변경시)		접수일	2023년 3월 2일
신청자	조직/부서명	한국데이터테크 AX팀			
	담당자 직위	책임	담당자 성명	이신뢰	
	전화번호	03-1234-5678	이메일 주소	security@kor_datatech.kr	
가명정보 활용 목적	<input type="checkbox"/> 통계 작성 <input checked="" type="checkbox"/> 과학적 연구 <input type="checkbox"/> 공익적 기록보존				
	※ 세부 활용 목적 채팅 앱 내 고객 간 일상대화 데이터를 정제, 분석하여 한국어 대화가 가능한 AI 챗봇 언어모델 개발				
가명처리 대상 개인정보	개인정보 명칭	자사 채팅앱 내 대화데이터셋			
	요청 개인정보 항목	2018~2020년동안 자사 채팅앱 내에서 1,500명의 고객들 간 이루어진 대화 데이터셋 20,000개 파일			
결합 활용 여부	<input checked="" type="checkbox"/> 해당 없음 <input type="checkbox"/> 결합전문기관 결합 예정 (결합대상정보 : ) (이용예정인 결합전문기관 : )				
처리 환경	<input type="checkbox"/> 정부가 지정한 안전한 처리환경 ( <input type="checkbox"/> 개인정보 이노베이션 존 <input type="checkbox"/> 결합전문기관 )				
	<input checked="" type="checkbox"/> 기관 내부 분석 공간 <input type="checkbox"/> 기관 외부 반출				
	<input type="checkbox"/> 기타 ( )				
반복 제공 여부	<input checked="" type="checkbox"/> 1회 제공 <input type="checkbox"/> 시계열 분석 등을 위한 반복 제공 ( ) 회 예정				
예상 이용기간	2023년 4월 1일 ~ 2026년 4월 1일 까지 (3년)				

2023년 3월 2일

신청인 (부서장명) 이신뢰 (서명 또는 인)

주식회사 한국데이터테크 대표이사 귀중

첨부서류	1. 가명정보 활용 계획서
------	----------------

## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	한국어 대화가 가능한 AI 챗봇 언어모델 연구
책임자	소속: AX팀      성명: 이 신 례      (서명)
배경 및 목적	(주)한국데이터테크는 사람과 한국어로 자유롭게 일상적인 대화를 할 수 있는 AI 챗봇 서비스를 제공하고자 하며, 이를 위해 한국어 대화의 맥락과 행간을 깊이 있게 이해하고 실제 사람과 대화하는 것 같은 대화 경험을 제공하기 위한 한국어 AI 언어모델을 개발하고자 함
활용 데이터	2018~2020년동안 자사 채팅앱 내에서 1,500명의 고객들 간 이루어진 20,000개의 대화 텍스트 데이터셋 파일
활용 계획 및 방법	<ul style="list-style-type: none"> <li>• 학습용 데이터베이스를 통해 상대방의 발화를 이해하고 적절한 답변을 선택하기 위한 모델, 문맥을 파악하고 부적절한 답변을 걸러내는 모델을 개발</li> <li>• AI 챗봇의 대답은 외부에 공개되기 때문에 가명정보를 통해 학습된 대화가 그대로 발화되지 않도록, 답변에 활용 되는 데이터베이스는 별도로 구축하여 가명정보 노출 가능성을 근본적으로 차단</li> <li>• 답변 데이터베이스는 (주)한국데이터테크 자체 생성 AI 모델이 만들어낸 문장과 내부 AI 데이터구축팀이 직접 작성한 문장으로 구성</li> </ul>
(예상) 활용 기간	2023년 4월 1일 ~ 2026년 4월 1일 (3년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>• 쉽고 자연스럽게 대화할 수 있는 AI 챗봇을 개발하여 사회 소외계층, 디지털 소외계층도 함께 누릴 수 있는 챗봇 생태계 조성</li> <li>• 한국어 대화 맥락 파악 고도화를 통한 맞춤형 상담·검색 기능 확대</li> </ul>

〈서식③〉

위험성 검토서

◆ 가명정보 활용 개요	
가명정보 활용 목적	한국어 대화가 가능한 AI 챗봇 언어모델 연구
가명처리 대상 개인정보	자사 채팅앱 내 대화데이터셋
활용하려는 자	(기관명) AX팀 (직위) 책임 (성명) 이신뢰

(1) 기본 위험도 판단	
〈Step 1〉 활용 주체	<input checked="" type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등
	<input type="checkbox"/> 제3자 제공 (→ Step 2도) - 타 개인정보처리자에 제공
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우
기본 위험도 자동 도출	〈Step 1〉 내부 활용 → <input checked="" type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경 → <input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경 → <input type="checkbox"/> 고위험
	※ 판단 이유 : 내부 부서에서 자체 활용 예정

(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장

(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	내부 분석환경에 제공하여 비교적 안전함에도 불구하고 비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 고위험으로 판단하고 외부전문가 인원을 포함한 적정성검토 위원회를 별도로 구성하여 검토 후 활용하는 것으로 결정		
검토일 : 2023 년 3 월 5 일			
검토자 소속 : 개인정보보호팀    성 명 : 나상호 (서명 또는 인)			

## 〈서식④〉

## 가명처리 계획서

가명처리 대상 파일			
No	파일관계	대상 정보	설명
1	주	사용자 계정정보 (정형)	
2	부	채팅앱 내 고객간 일상대화 데이터 (비정형)	20,000개의 파일로 구성되어 있음

## 1. 사용자 계정정보 (정형)

정보주체 수	1,500	항목수	4	데이터건수	20,000
No	항목명	유형	예시	항목설명	가명처리 계획
1	아이디	식별자	hkim2001	서비스 이용 계정 정보	동일한 사용자를 구분할 수 있도록 대체 일련번호 부여
2	채팅시작일시	범주형	2022-12-31 12:23:11	채팅이 시작된 일시	분 이하 삭제
3	채팅종료일시	범주형	2022-12-31 14:3:11	채팅이 종료된 일시	분 이하 삭제
4	파일ID	식별자	20221231_1	대화 내용을 저장한 파일명	Salt 추가 해시처리

## 2. 채팅앱 내 고객간 일상대화 데이터 (비정형)

정보주체 수	1,500	항목수	19	파일수	20,000
No	항목명	유형	예시	항목설명	가명처리 계획
1	파일ID	식별자	20221231_1	대화 내용을 저장한 파일명	Salt 추가 해시처리
2	이름	텍스트		대화에서 언급됨	인공지능 모델(Personal Name Recognition) 기반으로 문장 내에서 이름이라고 판단되는 단어가 발견되면, 해당 이름을 관련 통계에 따른 한국인 이름 분포에 따라 랜덤하게 생성한 임의의 이름으로 치환
3	생년월일	텍스트		대화에서 언급됨	생년월일 패턴이 발견되면 해당 정보를 랜덤하게 생성한 임의의 생년월일로 치환함
4	주민번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제
5	연락처	텍스트		대화에서 언급됨	전화번호 패턴(00 또는 000-000-0000 등)을 포함하는 문장 전체를 삭제 숫자를 한글로 표현(예: 일이삼, 영, 공 등)한 경우까지 고려하여 삭제하며, 한글 숫자와 아라비아 숫자가 병용된 패턴 역시 문장 삭제
6	카드번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제
7	계좌번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제
8	차량번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제
9	학번	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제
10	여권번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제

정보주체 수		1,500		항목수	19	파일수	20,000
No	항목명	유형	예시	항목설명	가명처리 계획		
11	운전면허번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제		
12	외국인등록번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제		
13	건강보험증번호	텍스트		대화에서 언급됨	해당정보 포함 전체문장 삭제		
14	상세주소	텍스트		대화에서 언급됨	지역명 주소로 구성된 읍·면·동·리 상세주소 및 도로명 주소로 구성된 길 이름이 포함된 문장 전체를 삭제 아파트 동호수 등의 패턴이 포함된 문장 전체를 삭제 길안내, 네비게이션 앱 메시지 삭제		
15	아이디	텍스트		대화에서 언급됨	숫자, 영문, 특수문자 등으로 구성된 7자리 이상의 단어가 포함되면 해당 문장 전체를 삭제 '아이디', 'ID' 등의 문구가 포함되면 해당 문장 전체를 삭제		
16	비밀번호	텍스트		대화에서 언급됨	숫자, 영문, 특수문자 등으로 구성된 7자리 이상의 단어가 포함되면 해당 문장 전체를 삭제 '비밀번호' '패스워드' 등의 문구가 포함되면 해당 문장 전체를 삭제 아파트 비밀번호 등에 자주 쓰이는 특수문자(#, *) 등으로 시작하거나 끝나는 패턴이 있으면 해당 문장을 삭제		
17	이메일	텍스트		대화에서 언급됨	'@' 기호가 포함되어 있는 이메일 패턴을 토큰 [MAIL]로 치환		
18	URL	텍스트		대화에서 언급됨	http:// 및 .com 등 URL 패턴 부분을 토큰 [URL]로 치환		
19	나이	텍스트		대화에서 언급됨	5세 단위 범주화 (16~20세, 21~25세, 26세~30세, 31세~35세 등)		



데이터 처리목적 및 활용 방식을 고려한 가명처리 대상 선정

▶	
▶	

## 〈서식⑤〉

## 가명처리 결과서

1. 공통적으로 적용한 가명처리 방법				
자료형	가명처리 순서			
	가명처리 방법 1	가명처리 방법 2	가명처리 방법 3	
식별자	Salt 추가 해시처리			
범주형	일자는 분 이하 삭제	빈도 3 이하 로컬삭제		
수치형	해당없음			
기타	해당없음			

2. 예외적으로 적용한 가명처리 방법			
자료형	적용 항목	가명처리 방법	사유
식별자	아이디	대체 일련번호	
	.....		

3. 주요 범주형 항목 값 분포 (검토가 필요한 항목만 작성)				
항목명	값	빈도수	구성비	비고
채팅시작일시	2020-12-31 12	7	0.0%	
	2019-05-24 15	10	0.0%	
	2018-06-16 13	14	0.0%	
	2019-11-09 18	64	0.0%	
	2020-02-19 14	95	0.0%	

4. 주요 수치형 항목 값 분포 (검토가 필요한 항목만 작성)								
항목명	처리 전				처리 후			
	Min	빈도	Max	빈도	Min	빈도	Max	빈도
.....								

5. 비정형데이터 전후 비교 샘플링		
항목명	처리 전	처리 후
주민등록번호	<ul style="list-style-type: none"> <li>민준: 자기야, 우리 가족관계등록부 발급받아야 하는데, 어떻게 하는 건지 알아?</li> <li>지수: 음, 예전에 한 번 해봤는데, 인터넷으로 발급받는 게 제일 편해.</li> <li>민준: 아, 인터넷으로도 되는구나.</li> <li>지수: 직접 방문해서 발급받고 싶으면 주민센터나 구청, 시청에 가면 돼. 신분증만 가져가면 바로 발급받을 수 있어. 자기 주민번호로 내가 발급 받아 놓을게. 자기 주민번호가 010101-3123456 맞지?</li> <li>민준: 응 맞아 고마워 집에가서 봐.</li> </ul>	<ul style="list-style-type: none"> <li>민준: 자기야, 우리 가족관계등록부 발급받아야 하는데, 어떻게 하는 건지 알아?</li> <li>지수: 음, 예전에 한 번 해봤는데, 인터넷으로 발급받는 게 제일 편해.</li> <li>민준: 아, 인터넷으로도 되는구나.</li> <li>지수: 직접 방문해서 발급받고 싶으면 주민센터나 구청, 시청에 가면 돼. 신분증만 가져가면 바로 발급받을 수 있어. 자기 주민번호로 내가 발급 받아 놓을게.</li> <li>민준: 응 맞아 고마워 집에가서 봐.</li> </ul>
	주민등록번호 정규표현식으로 해당 정보를 1차로 추출하였고, 혹시 누락이 있을 수 있으므로 학습된 LLM 모델을 통해서 추가 추출하여 해당 문장 삭제처리함	

〈서식⑥〉

비정형데이터 추가검수 결과서

가명처리 개요			
접수번호 (또는 처리명칭)	한국어 대화가 가능한 AI 챗봇 언어모델 연구		
원본 데이터 규모	고객 1,500명이 대화한 일상대화 데이터 20,000개 파일	원본 데이터 크기(용량)	24GB
원본 데이터 포맷	TXT	처리 결과 데이터 포맷	TXT
가명처리 적용 기술	- (주)한국데이터테크에서 자체 개발한 필터링 솔루션을 활용하여 주요 개인식별 위험 항목을 검출한 뒤 치환, 삭제, 범주화 처리 - 일반대화정보 내 개인식별 가능성이 있는 단어, 문장은 직접 검수하여 삭제 또는 대체		
추가검수 내용			
검수 기간	2023년 3월 23일 ~ 2023년 3월 25일		
검수 장소	(주)한국데이터테크 데이터센터		
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input checked="" type="checkbox"/> 통계적 표본 검수 <input type="checkbox"/> 휴리스틱 표본 검수		
	개인정보보호팀 담당 팀장 주도하여 총 2명의 검수자가 Workday 3일간 검중 • 전체 2만 건의 데이터 중 약 200개(1%) 파일을 무작위 샘플링 하여 육안 검사		
처리된 결과 확인 사항	① 치환되어야 할 항목(이름, 생년월일, 이메일, URL)들이 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환되었는지 확인 ⇒ 치환되지 않았을 시 수작업으로 변환 실시 ② 삭제되어야 할 문장(주민번호, 연락처 등 포함)들이 삭제되었는지 확인 ⇒ 삭제되지 않았을 시 수작업으로 삭제 실시 ③ 나이정보가 범주화되어 토큰으로 대체 되었는지 확인 ⇒ 대체되지 않았을 시 수작업으로 범주화 및 토큰화 실시 ④ 가명처리되어야 할 항목들이 인식되지 않아 제대로 처리가 되지 않은 문장이 존재하는지 확인 ⇒ 해당 문장 발견시 수작업으로 추가 가명처리 실시 ⑤ 가명처리 항목 외에도 일반대화정보에서 개인식별 가능성이 있는 단어나 문장이 존재하는지 확인 ⇒ 확인 시 팀장 주도 최종회의에서 삭제·대체여부 결정하여 추가 가명처리 실시		
검수 결과	200개 파일 중 14개 파일에 대해 오류 발견되어 추가 가명처리 실시 완료 ⇒ 누락하여 재처리한 사례 중 개인식별 위험이 특별히 높은 것은 발견되지 않았음. 대화맥락에 따른 식별 위험은 극히 일부 존재하였으나, 가명정보를 통해 학습된 대화가 그대로 발화되지 않도록, 답변에 활용되는 데이터베이스는 별도로 구축하여 가명정보 노출 가능성을 근본적으로 차단하였으므로 추가 표본검수·재처리 없이 해당 잔존 위험은 감수하기로 결정		
검증자	소속 및 직위	성명	서명(인)
	개인정보보호팀 팀장	나보호	
	개인정보보호팀 직원	우담당	
2023년 3월 25일 작성자 소속: 개인정보보호팀 성명: 나보호      서명(인)			

## 7 콜센터 직원 실습용 가상상담 시나리오 생성 AI 개발

### 개요 (음성, 텍스트 정보)

- 인터넷통신 기업 (주)코리아인터넷은 인터넷·통신 상품 관련 고객상담 시 분쟁방지 및 서비스 품질 향상 목적으로 통화내용이 녹음됨을 고객에게 공지하고 상담 음성정보를 수집·보관하고 있다. (주)코리아인터넷은 해당 음성정보를 가명처리한 뒤, 자사 콜센터 상담직원들에게 상황별·업무별 상담 실습 교육을 진행하기 위한 목적으로 가상상담 시나리오를 생성하는 AI를 개발하고자 한다.

### 처리 목적

- 콜센터 직원 상담 실습교육용 가상상담 시나리오 생성 AI 모듈 개발

### 데이터 특징

- 상담사와 고객 간 질의-응답으로 이루어진 음성 데이터(WAV 포맷)
  - 상담 목적(①단순 문의, ②A/S, ③결제, ④교환, ⑤설치, ⑥반품, ⑦기타), 성별, 연령, 주거 지역별로 샘플링한 상담음성 녹취파일 총 10,000개 파일
  - 각 파일 당 평균 3분 분량으로 총 500시간 (60GB)
  - 고객 및 상담사의 음성이 녹음되어 있으며 대화 내용에 고객의 이름, 생년월일, 주소, 연락처 등 개인식별성 있는 정보가 포함

### 데이터 이용 환경

- (망분리 폐쇄형 개발 환경 활용) (주)코리아인터넷 빅데이터본부에서 구축한 Private 클라우드 기반의 폐쇄형 AI개발 환경에서만 데이터를 활용
  - 취급자에게만 접근 권한 부여, 접근통제 솔루션으로 비인가자 접근 통제
- (자료 반입) (주)코리아인터넷 보안팀의 검토 및 승인 절차에 따라 운영
- (자료 반출) 분석결과 반출 시 관리자에게 요청(관리자 자료 확인 후 제공)

◆ 데이터 흐름



◆ 작성 서류

가명처리 단계	서식명	작성 주체	작성 여부	비고
(1단계) 사전준비	서식① 가명정보 이용·제공신청서	AI개발팀 (가명정보 이용자)	작성	
	서식② 가명정보 활용 계획서	AI개발팀 (가명정보 이용자)	작성	
(2단계) 위험성 검토	서식③ 위험성 검토서	데이터팀 (가명처리 부서)	작성	
(3단계) 가명처리	서식④ 가명처리 계획서	데이터팀 (가명처리 부서)	작성	
	서식⑤ 가명처리 결과서	데이터팀 (가명처리 부서)	작성	
	서식⑥ 비정형데이터 추가검수 결과서	데이터팀 (가명처리 부서)	작성	
(4단계) 적정성 검토	서식⑦ 적정성 검토 결과서	검토위원	작성	
	서식⑧ 적정성 평가위원 서약서	검토위원	작성	
(5단계) 안전한 관리	서식⑨ 안전조치 이행확약서		생략	내부 규정에 따라 관리
	서식⑩ 가명(추가)정보 관리대장	개인정보보호팀 (개인정보보호 담당 부서)	작성	



## 〈서식②〉

## 가명정보 활용 계획서

신청 제목	콜센터 직원 실습교육용 가상상담 시나리오 생성 AI 모듈 개발
책임자	소속: 빅데이터본부 AI 개발팀      성명: 박 안 전      (서명)
배경 및 목적	<ul style="list-style-type: none"> <li>• (주)코리아인터넷에서는 자사 콜센터 직원을 대상으로 상담 업무 능력 향상을 위한 교육에 활용할 교육용 음성생성 AI 기술개발을 진행하고 있음</li> <li>• 다양한 상담 목적, 고객 유형에 따라 단순 상담부터 난이도 높은 상담까지 문제없이 대응할 수 있도록 시를 통해 상황별 고객상담 음성을 생성하여 상담직원 교육에 활용하고자 함</li> </ul>
활용 데이터	다양한 발화자(성별, 연령, 지역 등)의 특징을 고려하여 샘플링한 10,000건의 녹취 파일
활용 계획 및 방법	<p>언어모델 개발과 언어모델 학습을 통한 상담 데이터 생성 모듈 개발</p> <ul style="list-style-type: none"> <li>• 상담음성을 텍스트로 변환(STT)한 뒤, 개인식별성이 있는 단어, 문장 등을 제거 또는 치환하여 개인식별 위험성을 제거</li> <li>• 가명처리된 텍스트 데이터를 기반으로 고객 질문 의도 분류 모델 및 상담사 답변 추천 모델 개발 <ul style="list-style-type: none"> <li>- 고객 질문 입력 이후 고객 질문 의도를 분류하는 합성곱 신경망 기반 분류 모델</li> <li>- 모델이 분류한 결과를 실제 고객 질문의도 범주와 비교하여 오차를 계산하고, 오차 역전파(backpropagation) 기법을 통해 모델 학습 진행</li> <li>- 성능평가 지표를 통해 학습된 모델을 평가</li> </ul> </li> <li>• 개발된 AI 모델에 의해 새롭게 생성된 콜센터 상담 시나리오(텍스트) 생성</li> <li>• 해당 상담 시나리오를 음성파일로 재변환(TTS)하여 교육실습자료로 활용</li> </ul>
(예상) 활용 기간	2023년 3월 15일 ~ 2024년 3월 16일 (1년)
기대효과 및 활용방안	<ul style="list-style-type: none"> <li>• 다양한 분야의 상담 데이터와 상담 사례별 데이터를 통해 상담 지원 대상 교육 효과 증대</li> <li>• 상담 만족도 향상을 위한 상담업무 효율 및 비즈니스 효과 증대</li> </ul>

## 〈서식③〉

## 위험성 검토서

◆ 가명정보 활용 개요			
가명정보 활용 목적	콜센터 직원 실습교육용 가상상담 시나리오 생성 시 모듈 개발		
가명처리 대상 개인정보	다양한 발화자(성별, 연령, 지역 등)의 특징을 고려하여 샘플링한 10,000건의 녹취 파일		
활용하려는 자	(기관명) 시개발팀 (직위) 책임 (성명) 박안전		
(1) 기본 위험도 판단			
〈Step 1〉 활용 주체	<input checked="" type="checkbox"/> 내부 활용 - 동일 개인정보처리자의 내부 활용 (외부제공 없음) - 직접 활용, 내부 타 부서에 제공, 내부 부서 간 결합하여 활용하는 경우 등		
	<input type="checkbox"/> 제3자 제공 (→ Step 2로) - 타 개인정보처리자에 제공		
〈Step 2〉 처리 환경  (제3자 제공인 경우만)	<input type="checkbox"/> 제공기관이 통제 가능한 환경 - 제공기관이 운영하는 내부 분석공간 등 제공기관의 관리·통제 범위 내에 있는 경우		
	<input type="checkbox"/> 제공기관이 통제 불가능한 환경 - 가명정보를 제3자에게 외부로 반출하여 제공기관이 처리환경을 통제·관리할 수 없는 경우		
기본 위험도 자동 도출	〈Step 1〉 내부 활용	→	<input checked="" type="checkbox"/> 저위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 가능 환경	→	<input type="checkbox"/> 중위험
	〈Step 1〉 제3자 제공 + 〈Step 2〉 통제 불가능 환경	→	<input type="checkbox"/> 고위험
	※ 판단 이유 : 내부 부서에서 자체 활용 예정		
(2) 사례별 위험도 조정			
반복·유사 활용 건	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
정부가 지정한 안전한 처리환경으로 제공	<input type="checkbox"/> 해당	<input checked="" type="checkbox"/> 해당 없음	※ 저위험으로 하향 조정 가능
비정형데이터 포함	<input checked="" type="checkbox"/> 해당	<input type="checkbox"/> 해당 없음	※ 기본위험도 상향 조정 권장
(3) 최종 위험도 확정			
조정	<input type="checkbox"/> 유지	<input checked="" type="checkbox"/> 상향	<input type="checkbox"/> 하향
최종 위험도	<input type="checkbox"/> 저위험	<input type="checkbox"/> 중위험	<input checked="" type="checkbox"/> 고위험
종합 검토 결과	내부 분석환경에 제공하여 비교적 안전함에도 불구하고 비정형데이터의 가명정보 제공 건임에 따라 보다 신중하고 전문적인 검토가 필요하므로 고위험으로 판단하고 외부전문가 인원을 포함한 적정성검토 위원회를 별도로 구성하여 검토 후 활용하는 것으로 결정		
검토일 : 2023 년 1 월 10 일			
검토자 소속 : 데이터팀    성 명 : 황국장 (서명 또는 인)			

**연구목적에 따른 연구방법 설정 및 데이터 전처리 (예시)**

●	.
●	.

**데이터 전처리(STT) 완료된 텍스트 데이터셋 (예시)**

발화자	대화 텍스트
(고객)	아 저희가 지금 인터넷TV 그 사용하고 있는데 속도가 좀 너무 안 나와서요
(상담원)	아 설치 시보다 인터넷 속도가 너무 느리시다는 말씀이신 거죠
(고객)	네
(상담원)	이용 많이 불편하실 텐데 너무 죄송하구요 우선 고객 확인 먼저 하겠습니다. _____ 고객님의 본인 맞으실까요?
(고객)	맞습니다
(상담원)	가입해주신 고객님의 혹시 _____ 맞으실까요?
(고객)	네 맞습니다
(상담원)	확인 감사드리구요. 인터넷 속도저하 부분이 와이파이나 인터넷 모두 동일한 현상인가요?
(고객)	와이파이가 조금 많이 느려요.. 그래서 TV가 자주 끊기거든요
...(중략)...	
(상담원)	_____ 되셨는데, _____ 과 가족할인 요금제로 사용하고 계시거든요. _____ 하시면서 _____ 도 가족결합할인으로 추가 하시겠다는거 맞으시죠? 상품먼저 배송해드릴게요. 주소가 어떻게 되시죠?
(고객)	_____ 입니다.
(상담원)	네 고객님의 _____ 로 배송해드리겠습니다
(고객)	찾기가 좀 어려우실텐데 _____ 돼요
(상담원)	네 기사분께 전달드려놓겠습니다. 기사분 내방전에 전화드리고 갈 건데 _____ 이니까 모르는 연락처라도 통화 잘 부탁드립니다
(고객)	네 알겠습니다. 혹시 안 받으면 저희 회사쪽으로 연락주시면 됩니다. _____ 입니다.
(상담원)	네 혹시 또 다른 문의사항 있으실까요?
(고객)	아니요 없어요.
(상담원)	하하하 네 이상 상담원 _____ 이었습니다

## 〈서식④〉

## 가명처리 계획서

## 【가명처리 대상 파일】

가명처리 대상 파일(요약)				
연번	항목명	구분	설명	비고
1	고객상담 음성파일	고객-상담사 간 상담녹취음성	▪ 상담사와 고객 간 질의-응답으로 이루어진 음성 데이터	음성 (비정형데이터) ⇒ STT 변환후 삭제

(변환: STT)



※ 음성변환(STT, Speech To Text) 기술을 통해 텍스트로 변환하여 활용  
각 음성파일에 고객 관련 메타데이터(상담목적, 성별, 연령, 주거지역) 포함

2	고객상담 텍스트 파일	① 이름	▪ 고객 및 상담사의 이름	텍스트 (비정형데이터)
		② 생년월일	▪ 고객의 생년월일과 관련된 정보	
		③ 주소정보	▪ 고객의 상세 주소와 관련된 정보	
		④ 거주정보	▪ 상세 주소는 아니나 고객이 거주하는 공간과 관련된 정보	
		⑤ 일반 전화번호	▪ 고객의 집·회사 연락처	
		⑥ 휴대폰번호	▪ 고객의 휴대폰 연락처	
		⑦ 일반대화정보	▪ 고객과 상담사 간 질의-응답 내 대화의 맥락을 통해 고객, 상담사, 또는 다른 특정인을 식별할 위험이 있는 대화정보	
2-1	상담 목적	고객의 상담 목적 (1.단순문의, 2.A/S, 3.결제, 4.교환, 5.설치, 6.반품, 7.기타)	각 상담파일의 메타데이터 (정형데이터)	
2-2	성별	고객의 성별 (1. 남성, 2. 여성)		
2-3	연령	고객의 연령 (1세대위, 17세~74세까지 분포)		
2-4	주거지역	고객의 주거지역 (1.서울, 2.부산, 3.대구, 4.인천, 5.광주, 6.대전, 7.울산, 8.세종, 9.경기, 10.강원, 11.충북, 12.충남, 13.전북, 14.전남, 15.경북, 16.경남, 17.제주)		

**【가명처리 계획】**

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	고객상담 음성파일	고객-상담사 간 상담녹취음성	☑ 삭제	• 음성변환(STT)을 통해 텍스트로 변환한 후 원본 음성파일은 삭제

(변환: STT)



※ 음성변환(STT, Speech To Text) 기술을 통해 텍스트로 변환

2	고객상담 텍스트 파일	① 이름	☑ 치환	<ul style="list-style-type: none"> <li>• 고객 이름은 [NAME_CUSTOMER] 토큰 정보로 변환한 뒤 '김행복'으로 변경</li> <li>• 상담원 이름은 [NAME_CENTER] 토큰 정보로 변환한 뒤, '김신리'로 변경</li> </ul>
		② 생년월일	☑ 치환	<ul style="list-style-type: none"> <li>• 생년월일은 [BIRTH] 토큰 정보로 변환한 뒤, 메타데이터 중 연령대(10~20대, 30~40대, 50~60대, 70대 이상) 범주 내에서 랜덤한 값으로 치환</li> </ul>
		③ 주소정보	☑ 치환	<ul style="list-style-type: none"> <li>• 주소는 [ADDR] 토큰 정보로 변환한 뒤, 메타데이터 내 주거 지역(시단위) 내 가상의 주소로 대체</li> </ul>
		④ 거주정보	☑ 삭제 (또는 대체)	<ul style="list-style-type: none"> <li>• 검수자(시나리오 작가)가 직접 가상의 문장으로 각색하고, 대체가 어려운 경우 해당 질의-응답 문장 삭제</li> </ul>
		⑤ 일반 전화번호	☑ 치환	<ul style="list-style-type: none"> <li>• 일반전화번호는 [PHONE] 토큰 정보로 변환한 뒤, 실재하지 않는 가상 전화번호로 대체</li> </ul>
		⑥ 휴대폰번호	☑ 치환	<ul style="list-style-type: none"> <li>• 휴대폰번호는 [CELLPHONE] 토큰 정보로 변환한 뒤, 실재하지 않는 가상 휴대폰번호로 대체</li> </ul>
		⑦ 일반대화정보	☑ 그대로 사용 (또는 삭제)	<ul style="list-style-type: none"> <li>• 연구 목적에 필요하므로 그대로 사용하는 것을 원칙으로 하되, 검수자가 전수 검사하여 대화 맥락이나 특이정보 등을 고려하여 특정 개인이 식별될 가능성이 있다고 판단되는 문장은 삭제 처리</li> </ul>
2-1	상담 목적	☑ 그대로 사용	<ul style="list-style-type: none"> <li>• 연구 목적 달성을 위해 별도 처리하지 않음</li> </ul>	
2-2	성별	☑ 그대로 사용	<ul style="list-style-type: none"> <li>• 연구 목적 달성을 위해 별도 처리하지 않음</li> </ul>	
2-3	연령	☑ 범주화 (20세 단위)	<ul style="list-style-type: none"> <li>• 고객 연령대 별 분석이 필요하므로, 20세 단위 범주화 처리하고 70세 이상은 상단코딩 적용 (10~20대, 30~40대, 50~60대, 70대 이상)</li> </ul>	
2-4	주거지역	☑ 그대로 사용	<ul style="list-style-type: none"> <li>• 연구 목적 달성을 위해 별도 처리하지 않음</li> </ul>	

## 〈서식⑤〉

## 가명처리 결과서 (비정형데이터 부분만 발췌)

비정형 데이터 전후 비교 샘플링				
연번	항목명	구분	처리 전	처리 후
2	고객상담 텍스트 파일	① 이름	<ul style="list-style-type: none"> <li>홍길동 고객님 본인 맞으실까요?</li> </ul> <p>〈고객이름 '김행복'으로 치환〉</p>	<ul style="list-style-type: none"> <li>'[NAME_CUSTOMER]' 고객님 본인 맞으실까요?</li> </ul> <p>⇒ '김행복' 고객님 본인 맞으실까요?</p>
			<ul style="list-style-type: none"> <li>네 감사합니다. 이상 상담원 <u>이선정</u>이었습니다</li> </ul> <p>〈상담원 이름 '김신뢰'로 치환〉</p>	<ul style="list-style-type: none"> <li>네 감사합니다. 이상 상담원 '[NAME_CENTER]'이었습니다</li> </ul> <p>⇒ 네 감사합니다. 이상 상담원 '김신뢰'였습니다</p>
		② 생년월일	<ul style="list-style-type: none"> <li>가입해주신 고객님 혹시 <u>1987년 7월 22일</u>생 맞으실까요?</li> </ul> <p>〈메타데이터의 범주화된 연령(30~40대) 내 랜덤값으로 치환〉</p>	<ul style="list-style-type: none"> <li>가입해주신 고객님 혹시 '[BIRTH]'생 맞으실까요?</li> </ul> <p>⇒ 가입해주신 고객님 혹시 '1991년 6월 8일'생 맞으실까요?</p>
		③ 주소정보	<ul style="list-style-type: none"> <li>고객님 서울시 신뢰동 신뢰아파트 백일동 구백삼호로 배송해드리겠습니다</li> </ul> <p>〈메타데이터의 주거지역(서울) 내 가상주소로 치환〉</p>	<ul style="list-style-type: none"> <li>고객님 '[ADDR]'로 배송해드리겠습니다</li> </ul> <p>⇒ 고객님 '서울시 신뢰동 신뢰아파트 1동 1호'로 배송해드리겠습니다.</p>
		④ 거주정보	<ul style="list-style-type: none"> <li>(질의) 찾기가 좀 어려우실텐데 신뢰역 3번 출구 앞에서 오른쪽으로 쪽 올라오시면 돼요</li> <li>(응답) 네 기사분께 전달드리겠습니다.</li> </ul>	<p>〈삭제〉</p> <p>⇒ 문맥상 다른 문장으로 대체가 어려우므로, 질의-응답문 세트 전체를 삭제</p> <p>* 시나리오 작가가 가상의 문장으로 각색하는 것도 가능</p>
		⑤ 일반 전화번호	<ul style="list-style-type: none"> <li>저희 회사쪽으로 연락주시면 됩니다. 공이 일이삼사 오욱칠팔입니다</li> </ul> <p>〈일반전화번호를 가상 전화번호로 치환〉</p>	<ul style="list-style-type: none"> <li>저희 회사쪽으로 연락주시면 됩니다. '[PHONE]'입니다</li> </ul> <p>⇒ 저희 회사쪽으로 연락주시면 됩니다. '03-111-1111'입니다</p>
		⑥ 휴대폰 번호	<ul style="list-style-type: none"> <li>기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 팔천오백이니까</li> </ul> <p>〈휴대폰번호를 가상 전화번호로 치환〉</p>	<ul style="list-style-type: none"> <li>기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 '[CELLPHONE]'이니까</li> </ul> <p>⇒ 기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 '일일일일'이니까</p>
⑦ 일반 대화정보	<ul style="list-style-type: none"> <li>인터넷 가입하신지 12년 5개월 되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요</li> <li>다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠?</li> </ul>	<p>〈그대로 사용〉</p> <ul style="list-style-type: none"> <li>인터넷 가입하신지 12년 5개월 되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요</li> <li>다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠?</li> </ul> <p>* 단, 전수검사하여 대화 맥락이나 특이정보를 고려하여 개인식별 위험이 있는 문장은 삭제·변경 필요</p>		

〈서식⑥〉

비정형데이터 추가검수 결과서

가명처리 개요			
원본 데이터 규모	녹취파일(WAV) 10,000개를 텍스트로 변환한 데이터셋	원본 데이터 크기(용량)	각 3분 분량, 총 500시간, 60GB
원본 데이터 포맷	WAV ⇒ JSON	처리 결과 데이터 포맷	JSON
가명처리 적용 기술	- A사 '텍스트 비식별화 솔루션'을 활용하여 이름, 생년월일, 주소, 연락처 정보를 검출하여 토큰 정보로 변환한 뒤 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환 - 일반대화정보 내 개인식별 가능성이 있는 단어, 문장은 직접 검수하여 삭제 또는 각색		

추가검수 내용	
검수 기간	2023년 2월 18일 ~ 2023년 2월 20일
검수 장소	(주)코리아인터넷 녹취통합시스템 분석PC
검수 방식 설명	<input type="checkbox"/> 전수 검수 <input type="checkbox"/> 부분 검수 <input type="checkbox"/> 통계적 표본 검수 <input checked="" type="checkbox"/> 휴리스틱 표본 검수 개인정보보호팀 담당 팀장 주도하여 총 2명의 검수자가 Workday 3일간 검증 • 전체 1만 건의 데이터 중 분량이 길거나, 식별가능항목이 3개 이상 있다고 탐지된 약 1,000개(10%) 파일을 선별하여 육안 검사 • 각색이 필요한 문장을 선별하여 시나리오 작가 2인이 각색 수행
처리된 결과 확인 사항	① 치환되어야 할 항목들이 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환되었는가? → 치환되지 않았을 시 수작업으로 변환 실시  ② 치환대상 외에 일반대화정보 내 개인식별 가능성이 있는 단어, 문장이 포함되어 있는가? → 발견 시 팀장과 논의 하에 삭제할지 가상의 내용으로 각색할지 결정 → 각색 대상으로 선별된 문장은 2명의 시나리오 작가가 각색 수행
검수 결과	• 수작업으로 이름, 생년월일, 주소, 연락처 정보 추가 변환 : 82개 파일 • 가상의 내용으로 일부 문장 각색 : 23개 파일

검증자	소속 및 직위	성명	서명(인)
	개인정보보호 팀장	가관리	
	음성 시팀 인턴	나정보	
	음성 시팀 인턴	다보호	
	음성 시팀 시나리오 작가	라작가	
	음성 시팀 시나리오 작가	마작가	

2023년 2월 20일

작성자 소속 : 개인정보보호팀

성명 : 가 관리      서명(인)

### 3

#### 가. 동일 개인정보처리자 내 이용

##### 잘못된 내부이용(동일 부서 내 이용) 사례 ①

→

▶

✓

▶

▶

##### 잘못된 내부이용(동일 부서 내 이용) 사례 ②

→

▶

▶

▶

### 잘못된 내부이용 (타 부서 내 이용) 사례 ①

→

▶

✓

▶

▶

### 나. 다른 개인정보처리자 제공(제3자 제공)

#### 잘못된 외부제공 사례 ①

→

▶

✓

▶

▶

### 잘못된 외부제공(위탁) 사례 ②

The diagram illustrates a process flow for external provision. At the top, a box contains a right-pointing arrow and a dot. Below this, a vertical list of steps is shown on the left side of a large light blue area. The steps are: a right-pointing triangle, a checkmark, another right-pointing triangle, and a right-pointing triangle followed by a red rectangular box. This red box represents an error or a point of concern in the process.



## 가명정보 처리 가이드라인

발행일	2026년 3월
발행처	개인정보보호위원회
지원기관	한국인터넷진흥원
디자인·인쇄	한결엠

- 본 가이드 내용의 무단전재를 금하며, 가공·인용할 때는 출처를 밝혀 주시기 바랍니다.

본 가이드라인은 2026년 3월 기준으로 작성되었습니다.  
항상 최신의 가이드라인은 개인정보보호위원회 누리집([www.pipc.go.kr](http://www.pipc.go.kr))  
또는 가명정보 지원 플랫폼([dataprivacy.go.kr](http://dataprivacy.go.kr))에서 확인하시기 바랍니다.

