

## 한의 빅데이터 공유 활용을 위한 데이터 비식별화 파이프라인 및 심의지원 기능 개발

김익태 기술연구원, 장현철 책임연구원, 손윤희 기술연구원, 김태홍 책임연구원\*

한국한의학연구원 한의약데이터부

### Development of a Data De-identification Pipeline and Review Support Functions for Utilizing Korean Medicine BigData

Icktae Kim, Hyunchul Jang, Yunhee Son, Taehong Kim\*

KM Data Division, Korea Institute of Oriental Medicine

#### Abstract

In August 2020, South Korea underwent significant legal amendments. These changes affected key statutes governing personal data protection, including the Personal Information Protection Act, Credit Information Act, and the Act on Information and Communications Network. In response to these changes, government agencies have provided guidelines to facilitate data utilization while maintaining personal information protection. This paper delves into the intricate challenges of data sharing and utilization, addressing issues related to pseudonymization and the approval process, as well as managing diverse datasets and assessing the privacy risks associated with data post-pseudonymization. The proposed solution is the KMBIG system. This system dynamically adjusts pseudonymization conditions and levels, and modularizes pseudonym processing. KMBIG offers a framework to quantify the safety, personal identification risk, utility, and K-anonymity of pseudonymized data, thereby supporting the data evaluation committee's review process. The KMBIG system is designed to enhance efficient data management in sectors such as healthcare. It aims to improve data utility while ensuring the protection of personal information.

---

#### Correspondence: 김태홍(Taehong Kim)

1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054, Rep. of Korea

Tel: +82-42-868-9418, E-mail: thkim@kiom.re.kr

Received 2023-11-02, revised 2023-11-17, accepted 2023-11-23, available online 2023-11-27

doi:10.22674/KHMI-11-2-6



**Keywords:** Data Sharing and Utilization, Personal Data Protection Act, Pseudonymization Dynamic Processing, De-identification Pipeline Module, Privacy Data Risk Index

## 서론

데이터의 활용 요구가 높아지면서 2020년 8월 개인정보보호 관련 3개 법률(개인정보보호법, 신용정보법, 방송통신망법)이 개정되었다<sup>1-3)</sup>. 데이터 3법이라고 불리는 이 법을 통해 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보 주체의 동의 없이 가명정보 처리 후 개인정보의 사용이 가능해졌다<sup>4)</sup>. 의료분야에서도 데이터 3법은 중요한 이슈가 되고 있으며 데이터 활용 분야의 확대와 새로운 산업에 대한 기대가 높아지고 있다<sup>5-7)</sup>. 그러나 개인정보의 활용성이 증가하면서 개인정보 보호에 대해 중요성이 부각되고 표 1과 같이 정부에서도 각종 가이드라인을 제시하고 있다<sup>8)</sup>.

표 1. 개인정보 관련 일부 부처별 데이터 거버넌스 추진 현황표

	데이터기반행정활성화법	민간데이터	개인정보 보호	신용정보
주관 부처	행정안전부	과학기술정보통신부	개인정보보호위원회	금융위원회
근거 법령	데이터기반행정 활성화에 관한 법률('20.12)	지능정보화기본법('20.12)	개인정보 보호법('11.9)	신용정보의 이용 및 보호에 관한 법률('20.8)
목적·주요 내용	체계적인 데이터기반행정 수행을 위한 공공기관 간 데이터의 공동활용과 과학적 행정체계 구축	데이터 생산·수집 및 유통·활용, 표준화 및 품질제고 등	개인정보처리자가 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 목적으로 정보주체의 동의 없이 개인정보를 가명처리하여 활용할 수 있는 기반이 마련	신용정보 관련 산업을 건전하게 육성하고 신용정보의 효율적 이용과 체계적 관리를 도모하며 신용정보의 오용·남용으로부터 사생활의 비밀 등을 적절히 보호

2022년 보건복지부와 개인정보보호위원회에서 발표한 보건 의료데이터 활용을 위한 가이드라인에 따르면 개인정보를 활용하기 위해서는 데이터 심의위원회를 구성하고 가명정보 처리 목적의 적합성, 가명처리 적정성 검토, 기관 내 가명정보 활용, 기관 외부로 가명정보 제공 여부 및 방법 등에 대해 심의 절차를 진행해야 한다<sup>9-12)</sup>.

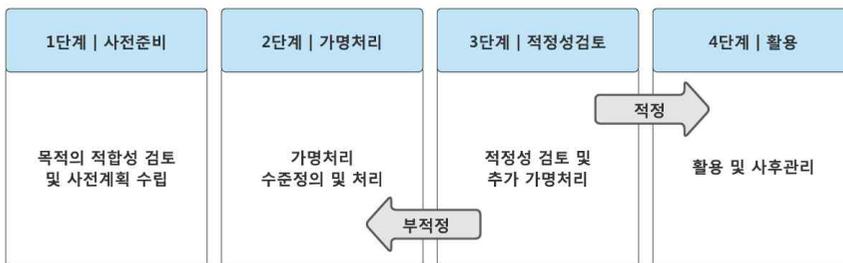


그림 1. 가명정보 처리와 데이터 활용 승인 절차



그림 1은 데이터 공유 활용에 필요한 심의 절차이다<sup>13)</sup>. 심의위원회는 사용자로부터 입력받은 데이터세트 활용 목적의 적합성을 검토한다. 가명처리된 후 개인정보 식별 위험성을 심의하고 부적정일 경우 가명처리 수준을 재조정하여 심의를 다시 진행한다. 또한 사용자는 여러 종류의 데이터를 재결합하고 분석하는 것을 반복해야 하기 때문에 새로운 데이터를 재결합할 때마다 데이터 공유 활용 신청을 반복해야 한다. 신청이 발생할 때마다 데이터 심의위원회는 해당 관련 전문가에게 심의를 요청하고 심의 관련 자료와 공유 활용할 데이터, 가명처리 후의 위험성 정보를 생성하고 제공해야 하는 절차를 반복하는 문제점이 발생하게 되며 이러한 문제점은 데이터 활용성을 낮추는 요인이 된다. 따라서 이를 해결하기 위해 사용자가 요청하는 데이터세트의 가명처리 조건과 수준을 즉시 반영하여 시각화하며 심의에 필요한 자료를 생성하고 적절한 심의 절차를 지원할 수 있는 시스템을 제안하고 이를 한국한의학연구원 데이터 공유 활용 서비스인 KMBIG 시스템에 적용하였다. KMBIG 시스템은 개인정보보호위원회의 가이드라인을 준수하는 가명처리 기능, 데이터 대체 기능, 범주화 기능, 재현데이터 생성 기능을 파이프라인 형태의 워크플로우 기능을 통해 모듈화되어 적용되었다. 동적으로 처리된 데이터의 안정성 및 개인식별 위험성과 유용성, K-익명성을 정량화하여 안정성 보고서를 생성 및 제공함으로써 처리된 데이터의 면밀하고 용이한 검토를 통해 데이터 심의위원의 사용성을 증대시키는 장점이 있다. 또한, 안정적인 시스템은 보건의료 데이터를 공유 활용하는데 효율적인 관리와 안정적인 개인정보보호를 기대할 수 있다.

논문의 구성은 다음과 같다. 본문 1절에서는 먼저 KMBIG 시스템의 적재 데이터인 한의 임상 데이터 분석에 관해 기술하고, 데이터의 형태와 구성요소를 분석하여 적절한 비식별화 방법을 제시하고, 2절에서는 데이터 심의위원회의 심의 절차와 기능에 대해 기술한다. 3절에서는 심의에 필요한 핵심 정보인 가명처리 후 개인정보 위험성과 데이터의 유용성, K-익명성을 포함하는 안정성 보고서를 제안하며 4절에서는 제안한 내용을 구현한 인터페이스를 기술한다.

## 본론

### 1. 데이터 분석

KMBIG 서비스의 공유 활용 데이터는 건강검진 중심의 임상 데이터, 질환 중심 임상 데이터, 개인 생성 데이터, 다국가 한의 정량 데이터가 있으며, 표 2와 같다.

표 2. KMBIG 시스템 적재 데이터

종류	데이터 설명	특징
건강검진 중심 임상 데이터	- 5개 한방병원에서 수집된 한·양방 건강검진 데이터	- 한의건강검진 수집 프로그램(설문, 기기), eCRF, 모바일 APP 이용
질환 중심 임상 데이터	- 질환별 한의임상 데이터	- 질환별 한의임상 주요지표 측정 프로토콜 실증연구
개인 생성 데이터	- 건강검진 수집자 및 진료데이터 수집 연구대상자 데이터 및 일상 사용자 데이터	- 개인 생성 데이터 정의 및 수집 프로토콜 구축 - 건강검진 시스템과 PHR플랫폼 연계
다국가 한의 정량 데이터	- 사상체질, TCI 심리데이터	- 한중 체질/심리 임상데이터 취득을 위한 국제공동연구 SOP 개발

개인정보의 위험성을 갖고 있는 KMBIG 시스템 적재 데이터를 분석하고 적절한 가명처리를 위해 각 분야 개인정보보호 전문가로 구성된 전문가 집단과 데이터 분석이 가능한 데이터분석팀을 구성하여 개인정보보호 가이드라인을 준수하며 데이터의 공유 활용 방법을 검토했다.

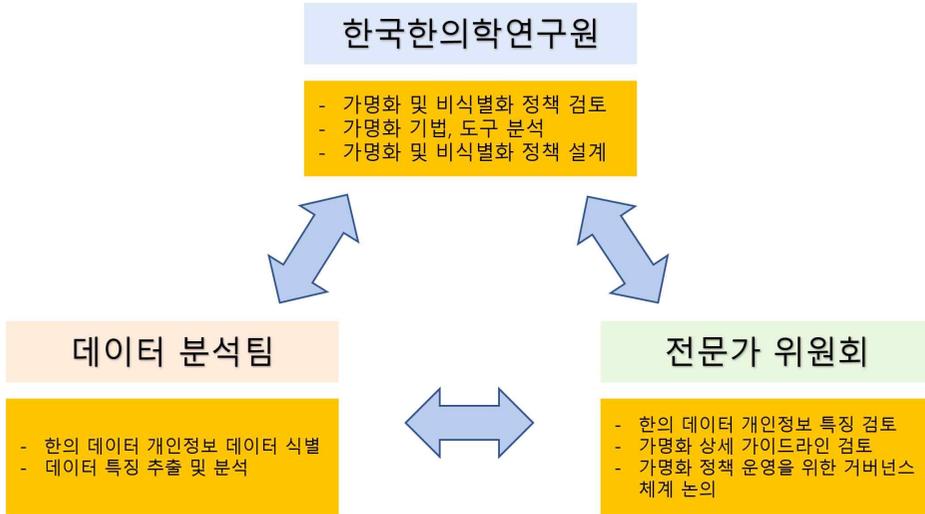


그림 2. 한의 데이터 분석 및 가명화 검토 추진체계

데이터 분석 방법으로는 타 데이터와의 재결합 가능 여부와 그에 따른 위험 수준(상중하)을 중심으로 분석하였고, 대상은 데이터베이스에 존재하는 모든 데이터의 속성 유형과 컬럼명(영·한), 정형 데이터와 비정형 데이터, 시계열 타입의 데이터와 파일로 존재하는 데이터의 파일명과 확장자까지 검토하였다. 분석 결과 파일 형태의 데이터는 TXT, CSV, PDF, XLSX, JPG, BMP, WAV 등으로 존재하며, 데이터의 값은 정형, 비정형(시계열) 데이터와 기계 측정에 의한 연속적인 수치데이터를 포함하고 있는 것을 확인하였다. 이를 바탕으로 표 3과 같이 데이터 유형에 따른 조치 가이드라인을 도출하였다.



표 3. KMBIG 시스템 적재 데이터 유형별 조치 가이드라인

종류	데이터 유형별 가명화 조치 가이드라인
식별자	- 식별자는 삭제가 원칙이나 보건의료 데이터에서는 일련번호로 대체하는 것을 허용 - 시계열 데이터나 대량의 측정데이터에 대한 연속성을 보장하여 사용 가능함 - 한의학연구원에서도 일련번호로 대체하여 동일인의 측정 정보에 대한 연구가 가능함
주요 인적사항	- 주소는 읍면동 단위 범주화 - 연령, 생일 일자 삭제 - 개인 식별 가능성이 높거나 특이정보인 경우는 추가 처리 필요
속성값	- 측정 수치 정보 : 별도 조치 불요 - 의료인의 관찰입력 정보 : 별도 조치 불요 - 알고리즘이 생산한 건강정보 : 별도 조치 불요
체의 촬영 영상정보	- 신체 외양의 정지영상, 동영상은 처리 필요 - 눈, 코, 입, 문신, 기타 외양적 특징을 모두 삭제하거나 충분한 모자이크 처리 또는 마스킹 실시 - 영상 상에 환자 번호, 성명 등 식별자를 표시한 것은 삭제 또는 마스킹 - DICOM 헤더 등 메타데이터 상의 식별자 삭제
체내 촬영 영상정보	- 단층촬영 3D 이미지 정보 : 추가 처리 - 음성정보 : 유보 (본인 동의 기반 활용)
유전체 정보	- 아래 2가지 항목을 제외하고 유보 (본인 동의하에 활용) - 널리 알려진 질병에 관한 유전자 변이 유·무 또는 변이 유형 - 생식세포 변이 정보를 제거한 신생물 고유(neoplasm)의 신규변이 정보
유전체를 제외한 오믹스 정보	- 별도 조치 불요

## 2. 데이터 심의위원회 심의 절차

보건복지부와 개인정보보호위원회의 ‘보건의료 데이터 활용 가이드라인’을 보면 표 4와 같이 명시하고 있다<sup>9)</sup>.

표 4. 보건의료 데이터 활용 가이드라인

구분	설명
데이터 심의위원회	- 가명정보의 기관 내 활용, 기관 외 제공, 결합신청, 가명처리 적정성 검토 등을 실시할 수 있는 독립 위원회(심의 업무 외부기관 위탁 가능)로서 5인 이상 15인 이하의 위원으로 구성, 외부 위원이 과반수 차지해야 함
특이정보	- 가명처리를 통해 특정 개인을 알아볼 수 없게 처리했다라도 ‘특이정보’를 통해 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있으므로 특이정보의 유형 관찰 및 검토가 필요함

심사위원회에서 심사를 위해서는 가명정보 처리 목적의 적합성, 가명처리 적정성 검토, 기관 내 가명정보 활용, 기관 외부로 가명정보 제공 여부 및 방법 등을 심의하도록 규정하고 있다<sup>14)</sup>. IRB 조직이 심의위원회의 기능을 대체할 수 있지만 현실적으로 IRB 조직을 갖고 있는 단체나 기관은 많지 않으며 행정 절차가 복잡하고 어렵기 때문에 KMBIG 시스템에 데이터 심의위원회의 심의



기능을 구현하여 원활한 심의가 될 수 있도록 제안한다. 이에 따라 한의 빅데이터 공유 활용을 위한 개인정보의 전처리 (사전준비), 가명처리, 적정성 검토, 사후관리 프로세스를 만들고 그림 3과 같이 심의 절차를 도출하여 KMBIG 시스템에 적용하였다.



그림 3. 심의 신청 절차

### 3. 안정성 보고서

심의위원회의 심의를 지원하기 위해 제안한 안정성 보고서는 가명처리 후 개인정보 위험성과 데이터의 유용성, K-익명성을 포함하고 있으며 온라인으로 가명처리 데이터를 직접 확인하고 심사 절차를 편리하게 진행할 수 있다. 이 안정성 보고서 정보를 바탕으로 가명화의 수준을 판단하는 근거로 활용할 수 있다. 데이터의 위험성은 Differential Correct Attribution Probability (DCAP) 지수를 사용하여 계산했고, 유용성 지수는 CIO (Confidence Interval Overlap), ROE (Ratio of Estimates),  $p$ MSE (propensity score-based MSE) 값의 평균을 사용하였다. K-익명성은 데이터의 재식별성을 나타내는 데이터세트의 속성으로 각 개인의 유사 식별자가 동일한 데이터세트에 있는 최소 K-1명의 다른 사람과 동일한 경우 해당 데이터세트는 K-익명성을 가진다<sup>15-18)</sup>. KMBIG 시스템에서는 K-익명성을 계산하여 식별자 존재 가능성을 표시한다.

### 4. KMBIG 시스템 적용

보건 의료데이터 활용을 위한 가이드라인을 준수하는 KMBIG 시스템 적재 데이터의 유형별 가명처리 방법과 데이터 심의위원회의 심의 절차를 도출하고 KMBIG 시스템에 적용하였다.



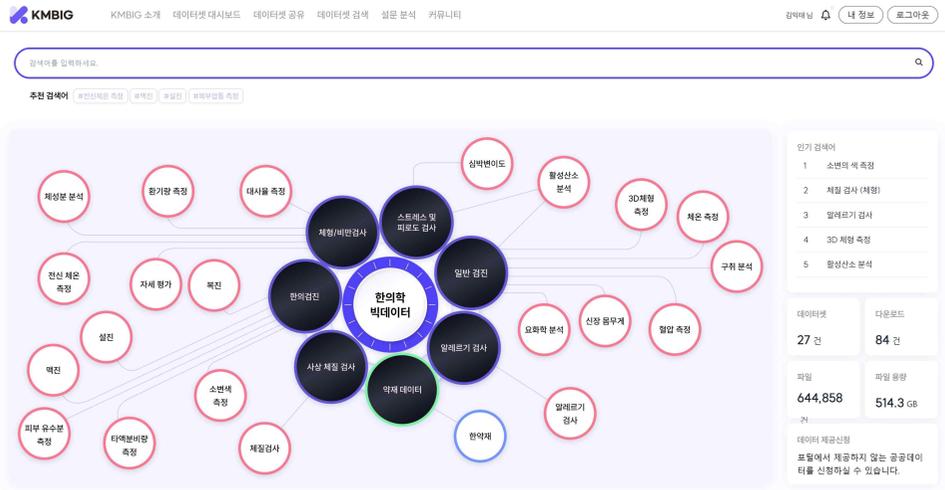


그림 4. KMBIG 메인화면\_데이터셋 대시보드화면

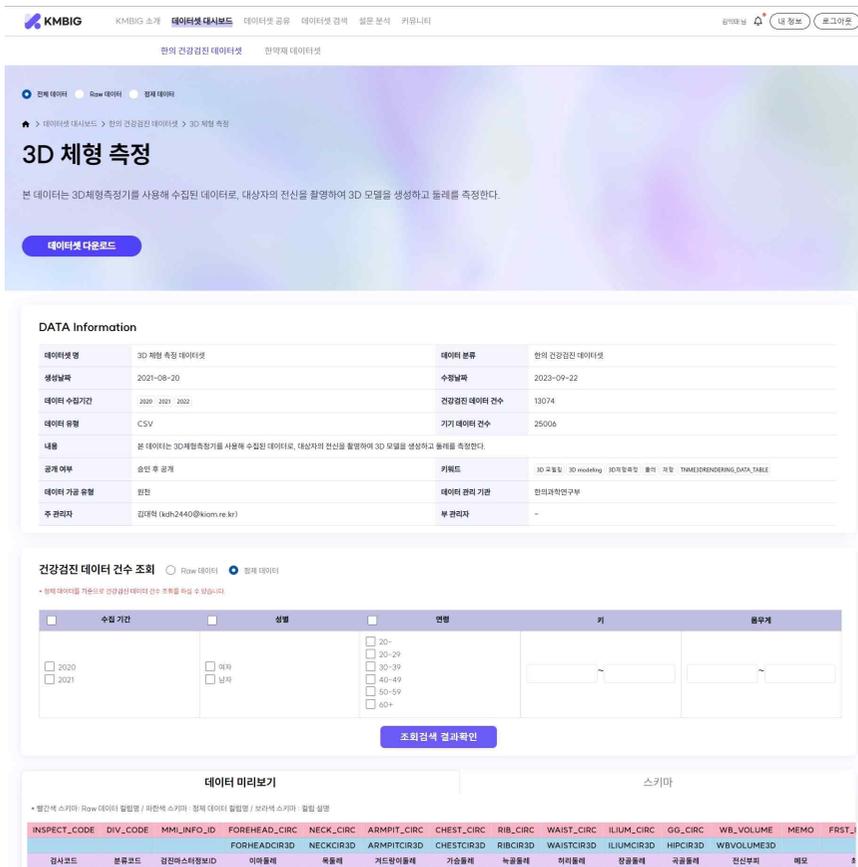


그림 5. 데이터셋 상세화면



그림 4, 그림 5는 사용자의 데이터 공유 활용 신청을 위한 데이터세트 검색 단계이며, 데이터 활용 목적, 공유 요청 목적을 입력하고 가명처리 후의 데이터 미리보기 기능을 통해 즉시 반영된 가명수준을 확인할 수 있게 한다.

데이터 미리보기 (전체 2801건 중 212건) - 정제 데이터 데이터 추가 가공    기기 데이터 추출

\* 데이터셋 공유 요청 필터

요구 사항				
<input checked="" type="checkbox"/> 소숫점 자리 제거	<input checked="" type="checkbox"/> 성별 제거	연령대 제거		
신장 제거	몸무게 제거	검진마스터정보 제거		

수집 연도	성별	연령	키	몸무게
2020	여자	20-	<input type="text"/> ~ <input type="text"/>	<input type="text"/> ~ <input type="text"/>
<input checked="" type="checkbox"/> 2021	<input checked="" type="checkbox"/> 남자	<input checked="" type="checkbox"/> 20-29		
		30-39		
		40-49		
		50-59		
		60+		

\* 데이터셋 미리보기

AGEGROUP	HEIGHT	WEIGHT	ABNMEAS	ABMEASND	REFDLTMM	REFDLTDKGF	REFDLTPAIN	ABNMEAS_REFDLTD	ABMEASND_REFDLTD	CV14MM
연령대	키	몸무게								
20-29	174	70	N		0.0	0.0	Y	N		0.0
20-29	171	76	N		0.0	0.0	Y	N		0.0

그림 6. 비식별화 처리 기능

그림 6은 사용자가 검색한 데이터에 가명처리를 진행하는 화면이다. 데이터 유형별 가명처리 방법에 따라 소숫점 자리 제거, 성별 제거, 연령대 제거, 신장 제거, 몸무게 제거, 검진 마스터 정보 제거를 선택하여 진행되며 사용자는 반영된 데이터를 실시간으로 확인하며 공유 신청을 진행한다.

그림 7은 다이제스트 기능의 화면으로써 관리자가 데이터 가명처리 항목을 추가하거나 암호화가 필요하거나 가명수준을 높여야 할 때 사용하는 화면이다. 다이제스트 기능은 가명처리 데이터 생성 기능을 파이프라인 형태의 워크플로우 기능을 통해 모듈화되어 적용하였으며 데이터 컬럼과 비식별 처리 방법을 변경하고 동적으로 데이터를 처리 후 시각적으로 확인할 수 있다. 다양한 가명처리 기능을 쉽고 직관적인 인터페이스 통해 사용자 친화적으로 지원하며 데이터 공유 심사 절차 간소화 및 효율성의 증대를 기대할 수 있다.



The figure displays four sequential screenshots of a data processing application interface, illustrating the steps to add a dataset and process variables.

**Top Left Screenshot: 사용자 공유 요청 정보 불러오기 (1.0.0)**  
 - Step 16: A 'Variables' table is shown with 'jsonFile' set to 'dataset/shareInfo.json'. The 'Input' column is 'undefined' and the 'Output' column is 'df'.  
 - Metadata: 데이터셋 ID: gdpOpwt7nywS2fr4SADImDF1lh, 데이터셋 명: 대사량 측정, 문서 ID: 76, 저장 파일 명: TNMEMETABOLISM(대사량측정)\_f.  
 - Data Table:

	SEX	AGEGROUP	HEIGHT	WEIGHT	MEI
0	M	30-39	173.0	82.0	211
1	M	30-39	177.0	76.0	232
2	M	30-39	163.0	72.0	177
3	M	30-39	165.0	62.0	150
4	M	30-39	173.0	80.0	138

**Top Right Screenshot: 숫자 범주화 (1.0.0)**  
 - Step 2: 'unit' is set to 2, and 'columns' is 'BROW\_SIZE'. The 'Type' is 'Text' and 'Desc' is '컬럼명'.  
 - Data Table:

MMI_INFO_ID	BROW_SIZE	NECK_SIZE	...
188	<100	40.6	...
190	<100	31.2	...
191	<100	34.2	...
192	<100	32.2	...
164	<100	31.4	...
182	<100	31.3	...
183	<100	32.0	...
184	<100	30.4	...

**Bottom Left Screenshot: 개인정보 마스크 (1.0.0)**  
 - Step 3: 'method' is 'a\*\*\*a' and 'columns' is 'MMI\_INFO\_ID'.  
 - Data Table:

CT_CODE	DIV_CODE	MMI_INFO_ID	BROW_!
	M8	1*8	<100
	M8	1*0	<100
	M8	1*1	<100
	M8	1*2	<100
	M8	1*4	<100
	M8	1*2	<100
	M8	1*3	<100
	M8	1*4	<100

**Bottom Right Screenshot: 개인정보 암호화 (1.0.0)**  
 - Step 4: 'length' is 8 and 'columns' is 'BIRTHDAY'.  
 - Data Table:

NECK_SIZE	...	SEX	BIRTHDAY	HEIGHT	WEIG
.6	...	M	XZUZxyZp	0	0
.2	...	F	YyWhaQMK	164.0	55.0
.2	...	F	B09MaH8j	0	0
.2	...	F	Bj6DMONO	0	0
.4	...	M	KV9iWrMA	178.0	76.0
.3	...	F	Rfaq9HgN	160.0	50.0
.0	...	F	uQLLu2av	0	0
.4	...	F	UmvIWKAU	158.0	54.0

그림 7. 다이어그램 추가 가명처리 기능



**공유 심사 확인**
심사중
목록
데이터셋 정보

---

**요청 번호**  
72

**요청 데이터셋**  
질진

**활용 목적**  
약술연구

**Utility**  
52

**요청자**  
김익태

**머감 날짜**  
기법발전

**데이터셋 공유 요청 목적**  
약술연구

**Risk Measure**  
34

**K-anonymity**  
식별자가 존재할 가능성이 높습니다.

---

**데이터 미리보기 (전체 2680건 중 162건) - 정제 데이터**

\* 데이터셋 공유 필터

요구 사항

소숫점 자리 제거

성명 제거

연령 제거

신장 제거

몸무게 제거

잠깐/스터잡보 제거

수집 연도	성별	연령		키	몸무게
<input type="checkbox"/> 2020 <input type="checkbox"/> 2021 <input type="checkbox"/> 2022	<input type="checkbox"/> 여자 <input type="checkbox"/> 남자	<input type="checkbox"/> 0~9세 <input type="checkbox"/> 10~19세 <input type="checkbox"/> 20~29세 <input type="checkbox"/> 30~39세 <input type="checkbox"/> 40~49세 <input type="checkbox"/> 50~59세 <input type="checkbox"/> 60~69세 <input type="checkbox"/> 70~79세 <input type="checkbox"/> 80~89세 <input type="checkbox"/> 90~99세 <input type="checkbox"/> 100세 이상	<input type="checkbox"/> 10대 <input type="checkbox"/> 20대 <input type="checkbox"/> 30대 <input type="checkbox"/> 40대 <input type="checkbox"/> 50대 <input type="checkbox"/> 60대 <input type="checkbox"/> 70대 <input type="checkbox"/> 80대 <input type="checkbox"/> 90대	<input style="width: 50px; height: 20px;" type="text"/>	<input style="width: 50px; height: 20px;" type="text"/>

\* 데이터셋

WEIGHT	TENMEAS	TEMEASND	LIGHTRED	COATED	BLUEPURPLE	TOOTHMARK	TEETC
82.0	N		0.32	0.32	0.3	12.0	
84.3	N		0.64	0.18	0.28	8.0	
71.9	N		0.46	0.9	0.46	22.0	
68.6	N		0.23	0.9	0.51	21.0	
94.9	N		0.45	0.62	0.27	11.0	
83.2	N		0.38	0.9	0.54	6.0	
87.8	N		0.24	0.83	0.45	8.0	
71.8	N		0.38	0.72	0.37	19.0	
86.4	N		0.41	0.49	0.31	15.0	
72.4	N		0.41	0.6	0.41	8.0	

**활용 승인 여부**

심사중

**코멘트**

문단

[Rich Text Editor]

거절
승인

그림 8. 안정성 보고서 심의화면



그림 8은 데이터 심의위원의 심의 화면이며, 심의를 지원하는 안정성 보고서를 포함한다. KMBIG 시스템은 사용자의 공유 신청 데이터 상세 정보와 3절에서 설명된 안정성 보고서를 생성하고 제공하여 원활한 심의를 지원한다.

## 결론

개인정보보호 관련 3개 법률이 개정되면서 데이터의 활용과 개인정보보호 기술이 중요해지고 있다. 정부기관과 개인정보보호위원회에서 데이터 활용을 위한 비식별화 가이드라인을 제시하고 있다. 본 연구에서는 한국한의학연구원에서 수집 중인 한의 임상 빅데이터를 대상으로 데이터 타입(시계열, 정형), 샘플값, 개인정보 여부, 가명정보 처리 필요 여부, 위험 수준(상중하), 재결합 가능 여부, 데이터 속성 유형을 검토하였다. 검토 결과를 바탕으로 한의 임상 데이터 유형별 조치 가이드를 제안하고 가명처리 기능을 KMBIG 시스템에 구현하였다. 또한, 데이터 심의위원회의 심의 처리 절차를 정립하였고 가명처리 후의 개인정보 위험성, 데이터의 유용성, K-익명성 정보를 포함한 안정성 보고서를 생성 및 제공하여 다수 사용자의 다양한 데이터세트 공유 심사를 효율적으로 진행할 수 있도록 하였다.

향후 연구에는 KMBIG 시스템을 일반인들과 전문가들로 구성된 사용성 평가 설문조사를 실시하여 만족도와 장단점을 분석하고 보완할 계획이다.

## 감사의 글

본 연구는 2023년도 한국한의학연구원 'AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축(KSN1922110)' 과제의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

## 참고문헌

1. 행정안전부(대한민국). 행정안전부 개인정보보호 지침. 행정안전부훈령 제 167 호. 공포 2020-11-10. 전부개정. 시행 2020-11-10.
2. 한국인터넷진흥원. 2023 Privacy Report : 개인정보보호 월간동향분석 (3 월호). 한국인터넷진흥원. 2023.
3. 문화체육관광부(대한민국). 대한민국 정책브리핑:데이터 3 법. 2021-09.
4. 개인정보보호위원회. 개인정보 보호법 및 시행령 개정사항 안내(초안). 2023-09-27.
5. Bae YI, Shin HR. Makes a Better Future for Korea : 데이터 3 법, 데이터경제의 시작. 경기연구원. 2020-02.
6. Son YH. Revision of Data 3 Law and Utilization of Personal Healthcare Data in Platform Economy Era. Journal of Industrial Property. 2021;(67):437-78.
7. Lee HY. 데이터 3 법 개정이 산업계에 미치는 영향(2). The Law Times. Available from: <https://www.lawtimes.co.kr/news/161573/> (accessed 2023-10-31).
8. 개인정보보호위원회. 생체정보 보호 가이드라인. 2021-09.
9. 보건복지부(대한민국), 개인정보보호위원회. 보건의료데이터 활용 가이드라인. 2022-12.

10. Lee K-H, Kim K-H. A Study on the Contents and Limitation of Guidelines for Utilization of Healthcare Data. Institute for Law of Science & Technology. 2020;26(4):89-118.
11. Lee JB. "A Legal study on the scope and limitations of health · medical big data utilization". Institute for Legal Study. 2021;45(1):67-99.
12. Jung HW, Cho YS, Ko GW, Song J-I, Yu DH. Comparison study of synthetic data generation methods for credit card transaction data. Journal of the Korean Data And Information Science Society. 2023;34(1):49-72.
13. 개인정보보호위원회. 가명정보 처리 가이드라인. 2022-04.
14. Choi YH. 산업경제 산업포커스 : 의료데이터의 활용과 정책 시사점. 2023-05.
15. Yoo SJ, Park NR. Synthetic Data Generation for Individual Credit Data Using CART. Journal of the Korean Official Statistics. 2020;25(1):1-30.
16. Kim KW, Min HK. Mobility Creation of Reproduction Data (Synthetic Data) and Statistical Verification. Korean Society of Transportation. 2023;20(2):6-13.
17. Ryu SH, Hong YK, Ko GH, Yang HD, Kim JW. Privacy Model Recommendation System Based on Data Feature Analysis. Journal of the Korea Society of Computer and Information. 2023;28(9):81-92.
18. Choi YJ, An SH, Bae JY. Trends in Research on AI-Based Face Image De-Identification Technology. The Journal of The Korean Institute of Communication Sciences. 2022;39(12):25-31.

