

# **AI for Healthcare (Facts & Limitations) & Privacy Protection for Healthcare**

**Soo-Yong Shin, PhD**  
**Department of Biomedical Informatics**  
**2016.5.9**

@ 소프트웨어정책연구소



**ASAN**  
Medical Center

# AI for Healthcare



**ASAN**  
Medical Center

# Sources of Healthcare Data

- 의료기관 (clinical data)
- 공공기관 (claim data)
  - 건강보험공단, 건강보험심사평가원
- 공공기관 (environmental data)
  - 기상청, 환경부, ...
- 개인 (Patient-Generated Health Data)
  - Wearables, home-monitoring devices
- 개인 (Lifelog)
  - Wearable data, SNS data, purchase data

# Types of Healthcare Data

- Text
- Image
- Video
- Code
- Sound
- ...

**Pathology Report (Text):** P1DBRM - 조기위암위험검역술(Mucosectomy) / Gastric cancer. DIAGNOSIS: Stomach, (antrum, greater curvature) endoscopic mucosal resection: - EARLY GASTRIC CARCINOMA, SINGLE, EGG TYPE I, c. TUBULAR ADENOCARCINOMA, POORLY DIFFERENTIATED, INTESTINAL TYPE, 0.6x 0.4x 0.2cm, with 1) confinement to mucosa (invasion to muscularis mucosae), 2) no involvement of lateral and deep resection margins, 3) lymphovascular invasion: not identified.

**ECG (Image):** A standard 12-lead electrocardiogram showing heart rhythm.

**Endoscopic Image (Image):** A grayscale image showing the internal view of the stomach during a mucosectomy procedure.

**ICD-10 (Text):** The International Classification of Diseases and Health Related Problems, Tenth Revision, Volume 1. Published by the Pan American Health Organization and the World Health Organization.

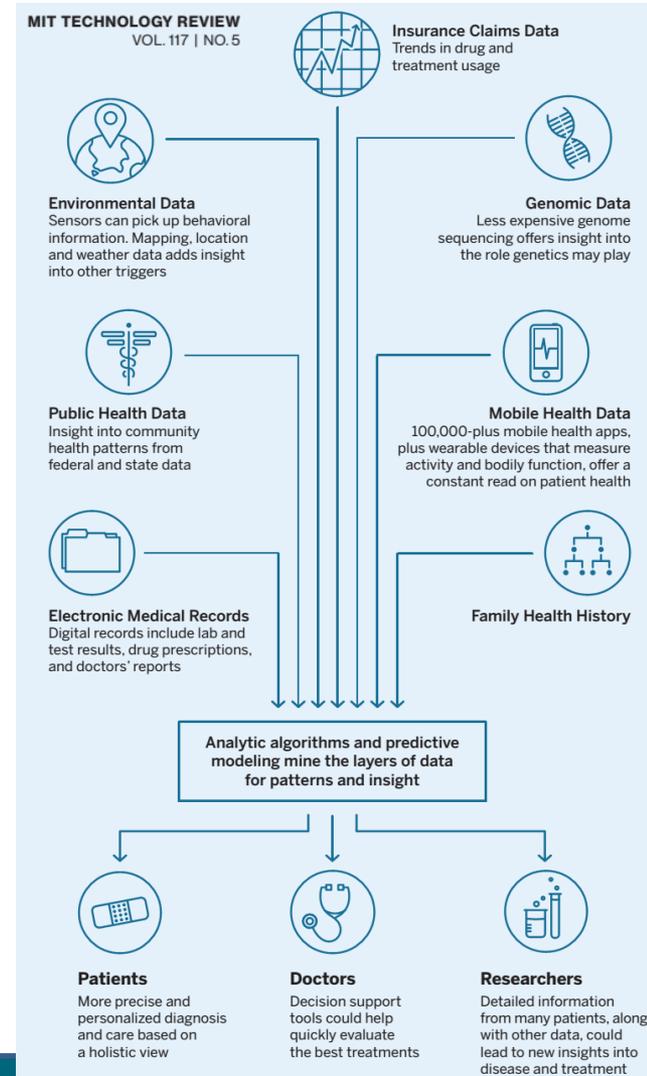
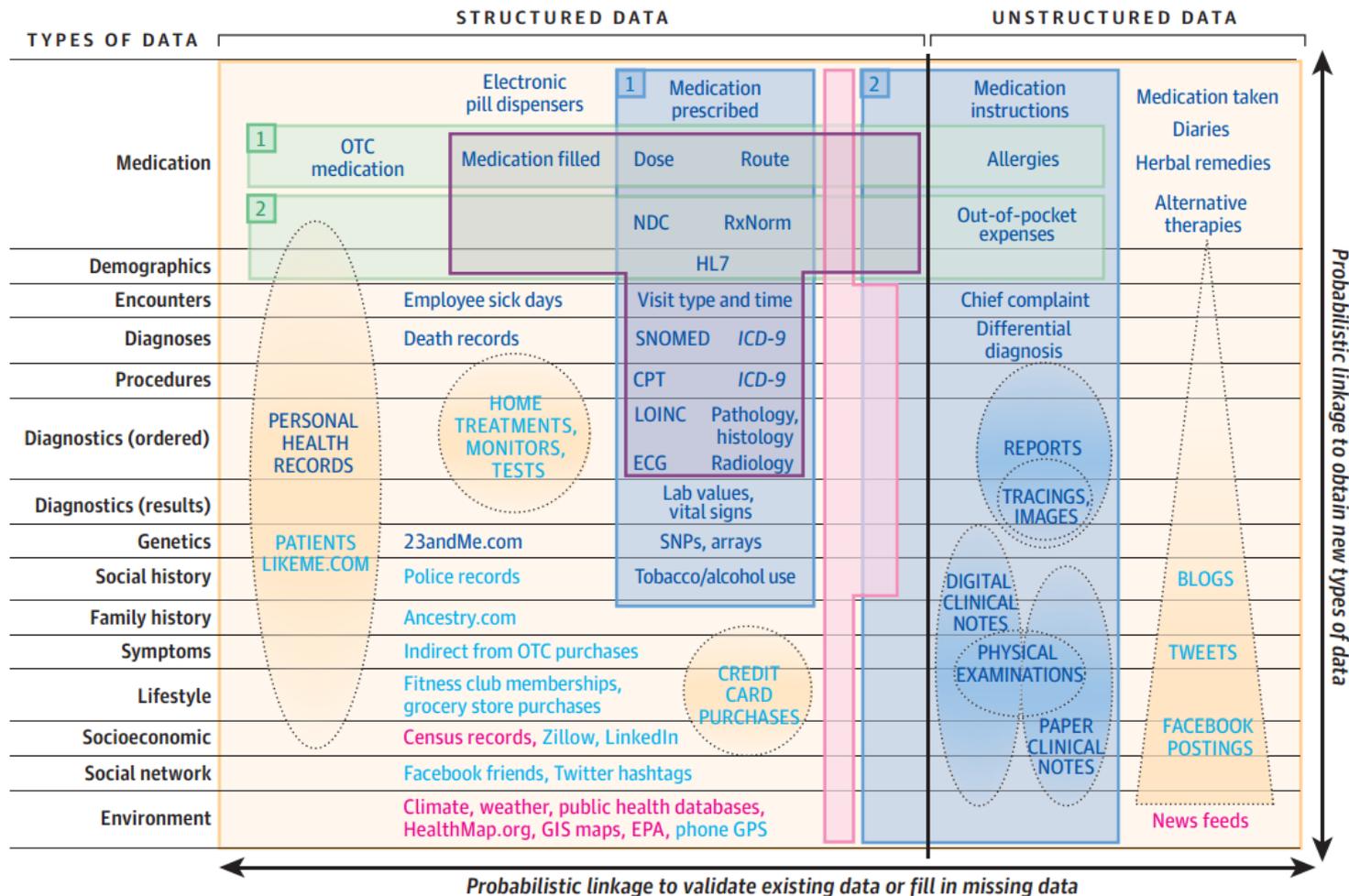


Figure. The Tapestry of Potentially High-Value Information Sources That May be Linked to an Individual for Use in Health Care



<b>Examples of biomedical data</b> 1 Pharmacy data    1 Health care center (electronic health record) data 2 Claims data    2 Registry or clinical trial data Data outside of health care system		<b>Ability to link data to an individual</b> ■ Easier to link to individuals ■ Harder to link to individuals ■ Only aggregate data exists	<b>Data quantity</b>  More      Less
---	--	--	---

CPT indicates current procedural terminology; ECG, electrocardiography; EPA, US Environmental Protection Agency; GIS, geographic information systems; GPS, global positioning system; HL7, Health Level 7 coding standard; ICD-9, Institutional Classification of Diseases, Ninth Revision; LOINC, Logical

Observation Identifiers Names and Codes; NDC, National Drug Code; OTC, over-the-counter; SNOMED, Systematized Nomenclature of Medicine; SNP, single-nucleotide polymorphism.

<http://www.ncbi.nlm.nih.gov/pubmed/2485414>

# Size of Healthcare Data

## + Big Data: Healthcare

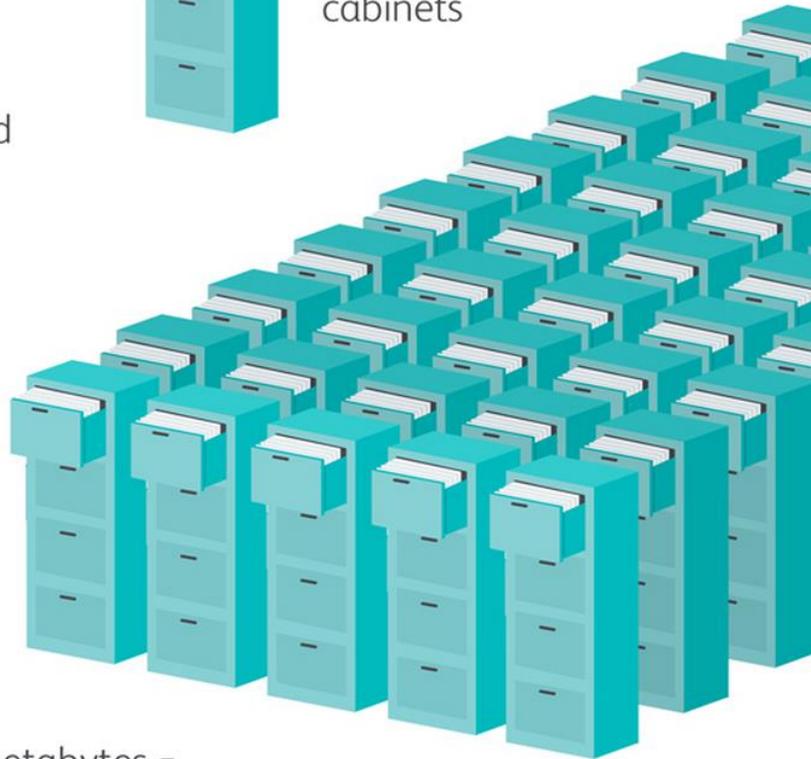
The amount of worldwide healthcare data is expected to grow to 50 times the current total to

25,000 petabytes



2012

500 petabytes =  
10 billion four-drawer  
cabinets



2020

25,000 petabytes =  
500 billion four-drawer cabinets

Source:

"Clouds Roll in to Handle Stratospheric Capacity Needs," Healthcare IT News, October 2011



ASAN  
Medical Center

# Size of Individual Healthcare Data

Industries dealing with data overload – Healthcare

## Exogenous data

(Behavior, Socio-economic, Environmental, ...)

**60%** of determinants of health  
*Volume, Variety, Velocity, Veracity*

**1100 Terabytes**  
Generated per lifetime

## Genomics data

**30%** of determinants of health  
*Volume*

**6 TB**  
Per lifetime

## Clinical data

**10%** of determinants of health  
*Variety*

**0.4 TB**  
Per lifetime

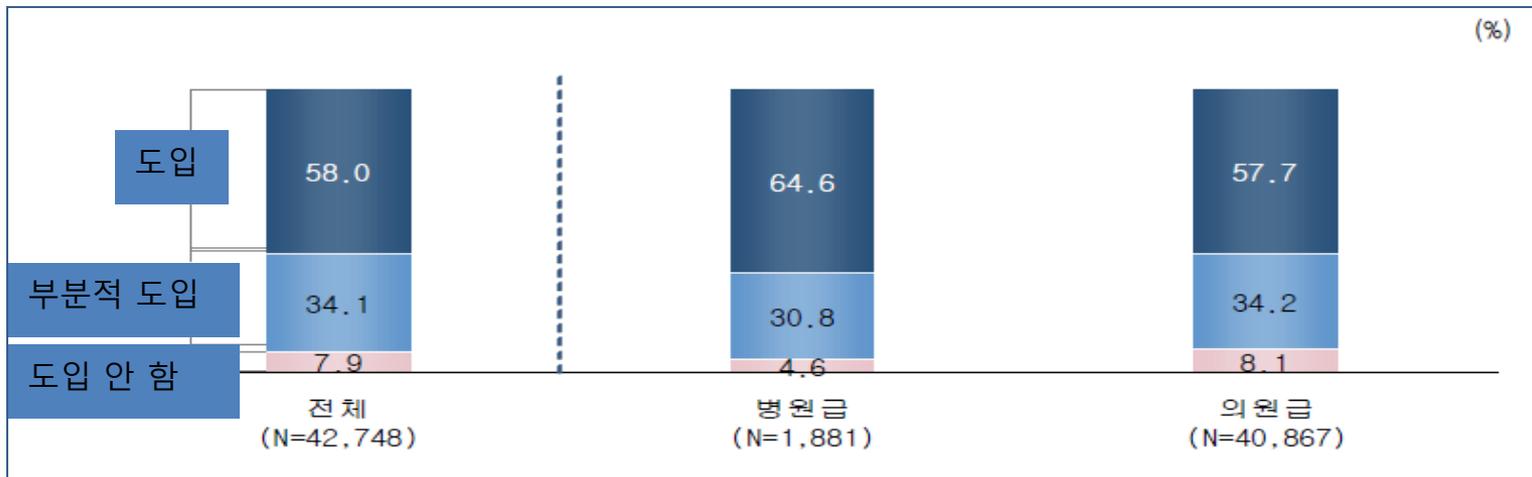


# AI for Healthcare!



# Facts & Limitations: EMR 보급률

- 국내 EMR 보급률은? (over 92%)



- BUT.. 실제적으로는 ?
  - 영상EMR이라는 것도 있음.
    - 종이로 쓰고 스캔

# Facts & Limitations: EMR?

- **현재 EMR은 Word Processor**
  - 대부분 Text
  - 수없이 많은 약어
  - 복잡하고 전문적인 의학 용어
  - 약어의 경우 의사마다, 진료과마다 다름 (통일성 zero)
- NLP?

# Facts & Limitations: NLP?

- In US or other foreign countries
  - YES!
- In Korea
  - Probably **NO!!**
  - 통일되지 않은 약어
  - 전문 의학 용어
  - phrase not sentence
  - Bilingual (Korean + English) + symbols..

# Facts & Limitations: NLP?

- Personal experiences
  - 차라리 Regular expression이 잘 됨
  - 일부 검사보고서 (병리보고서)는 semi-structure
    - 패턴만 알면 99% 정도의 정확도로 data 추출 가능
  - NER (Named Entity Recognition)도 regular expression으로 대부분 가능
  - BUT, 의료진의 높은 기대치가 제약사항.
    - Precision > 95%, Recall > 90% (is it possible?)

# Facts & Limitations:

## incomplete data of EMR

- Clinical data의 정확성?
  - **NO!**
  - 성인 키가 15cm, 몸무게가 1000kg... (오타!)
  - 병원의 진단명은 환자의 Phenotype 중 아주 일부분만 나타냄
    - Ex) 당뇨병 환자 중 정말 당뇨병 진단명이 붙은 환자는 ?
  - Post-annotation이 반드시 필요
  - Phenotyping!

# Facts & Limitations:

## incomplete data

- **Claim data**
- 실제 환자 진료 데이터랑 불일치
  - 왜냐하면 claim data는 돈을 받기 위한 청구용 data
  - 돈을 벌기 위해서 진단명이 변경됨
- 아주 일부 데이터만 있음
  - 진단명, 투약력, 검사 시행 여부, 극히 일부 검사 결과 (돈을 받기 위한) 등

# Practical Direction of AI for Healthcare

- **Images/Video**에 집중!
  - 의사들 중 가장 컴퓨터에 익숙한 그룹이 영상을 판독하는 영상의학과 교수들
  - Blackbox algorithm에 대한 거부감이 적음. (어차피 통계기법으로 안 되기 때문에)
  - Deep learning이 가장 잘 할 수 있는 분야

# Practical Direction of AI for Healthcare

- **해석이 가능한 모델부터 시작!**
  - Decision tree와 같이 의료진이 이해할 수 있는 model부터 접근
  - 의사들은 통계분석을 선호함
  - 본인들이 이해할 수 없는 AI기법 (*통계도 이해하는지 의문이지만..*)에 대한 거부감이 큼
  - BUT, AlphaGo..?!

# Practical Direction of AI for Healthcare

- **PGHD**도 새로운 대안
  - Clinical data는 구하는 것이 어려움
  - IRB, 의료진의 비협조, ....
  - POCT 장비들의 발전으로 손쉽게 Data 획득 가능
  - 단, 정확도는 낮아서 의료용으로는 쓰기 어려움
  - 하지만, continuous monitoring이 가능

# Practical Direction of AI for Healthcare

- **의료기기 log data**를 분석하라!
  - 중환자실의 환자감시장치의 경우 생체신호를 real-time으로 생성
  - 하지만, 의료 기관에서는 무시되고 있음. 오로지 alarm용으로만 활용
  - 질환 사전 예측이 가능
- **SNS** 분석
  - Influenza outbreak 예측, 자살 예측...

# Remaining Hurdles

- 진료현장에 쓰기 위해서는 **임상시험이 필요**
- CDSS의 경우 **식품의약품안전처 승인이 필요**함!
- Common data model
  - 국내의 경우 **data integration**을 위한 **표준 준수 미비**
- Clinical data를 활용하기 위해서는 **개인 동의 필요**
  - 익명화가 대안

# What to do as engineers

- Health IT?
  - IT+BT+HT...
- 의학용어, Health IT 표준 review
- 의료진과의 COMMUNICATION!
  - 화성에서 온 남자, 금성에서 온 여자..

# What to do as engineers

- NLP를 이용한 Post-processing보다 SDE (structured data entry)를 통한 입력시점부터의 자료 구조화
  - Pre-processing!
- Cleansing, cleansing, cleansing!
- Annotating!
- PMI-cohort 프로그램을 하는 이유

# Final Q: AI Doctor?

- 의사는 면허직임!
- 컴퓨터에게 license를 줄 것이냐?
  - 의료 사고시 책임은 누가?

# Privacy Protection for Healthcare



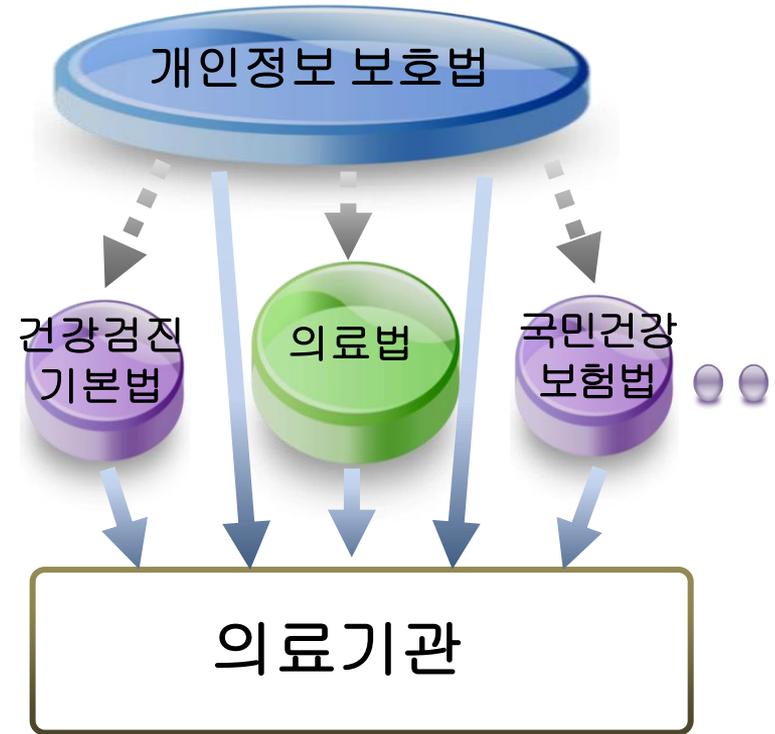
**ASAN**  
Medical Center

규제 때문에 모두  
법률 전문가화

# 의료와 관련된 개인정보보호 법률 체계

법령(예시)	주요내용
개인정보보호법	-개인정보보호에관한일반법 -공공및민간분야에모두적용
의료법	-환자의진료내역,병력등정보보호
건강검진기본법	-검진자료의활용및보호
국민건강보험법	-국민질병부상등에대한보험급여
보건의료기본법	-질병자의심자보고,신고의무
응급의료에 관한 법률	-응급환자이송
장기이식등에관한법률	-장기이식자등의비밀보호
생명윤리및안전에관한법률	-유전정보등의보호
인체조직안전 및 관리 등에 관한 법률	-조직기증자정보보호
감염병의 예방 및 관리에 관한 법률	-감염병환자의사환자등신고의무
후천성면역결핍증예방법	-감염인 진단검안사실 신고의무
혈액관리법	-특정수혈부작용신고의무

▶ 개인정보보호법은 일반법이며, 다른 법률에서 특별한 규정이 있으면 해당법률이 우선 적용



# 개인정보란?

개인의 사적 영역과 관련된, 개인을 식별할 수 있는 일체의 정보



# 개인정보 vs. 고유식별정보

## 개인정보

- 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보
- 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함

## 민감정보

- 사상/신념, 노동조합/정당의 가입/탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보
- 유전자검사 등의 결과로 얻어진 유전정보나 범죄경력자료에 해당하는 자료

## 고유식별정보

- 개인을 고유하게 구별하기 위하여 부여된 식별정보로 다음의 어느 하나에 해당하는 정보
- 주민등록번호
- 여권번호
- 운전면허번호
- 외국인등록번호

정보보호 강도

\* 개인정보 보호법상의 정의



ASAN  
Medical Center

# 의료기관 정보보호 관련 규제 현황

## 현

### 개인정보 보호법

- '11.3.29 제정, '12.3.30 시행
- 개인정보처리에 관한 기본원칙 규정
  - 표준개인정보보호지침
  - 개인정보의 안전성 확보조치 기준 고시
  - 의료기관 개인정보보호 가이드라인
  - 인사, 노무관리 업무 개인정보보호 가이드라인

### 생명윤리 및 안전에 관한 법률

- '15.12.29 일부 개정안 시행,
- 인간대상연구 및 인체유래물연구에 대한 윤리심사 의무화, 익명화 처리 후 이용 등

### 정보통신기반 보호법

- '12.2.23 신규지정 고시, '12.4.6 신규지정 통보
- 보건분야에서 3개 의료기관 지정
  - 서울아산병원, 삼성서울병원, 세브란스병원
- 매2년마다 취약점 분석, 평가 실시

### 의료법

- 비밀 누설 금지, 기록 열람 제한 등

### 정보통신망이용촉진 및 정보보호 등에 관한 법률

- '12.8.18 개정 시행, 6개월 계도기간
- 주민등록번호 수집, 이용 금지, 내부 업무망과 외부 인터넷망의 분리 등

### 근로기준법, 산업안전보건법 등



# 개인정보 보호법 & 생명윤리 및 안전에 관한 법률

## 개인정보 보호법

- 2011년 9월 30일 시행
- 2014년 8월 7일 개정안 시행
- 개인정보 이용 제한
  - 제18조 2항 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우” 예외로 인정
  - 제24조 고유식별정보\* 사용 원칙적 금지
  - 제24조 2항 **고유식별정보 중 주민등록번호는 다음의 경우를 제외하고는 처리 금지**
    - 법령에서 구체적으로 주민등록번호의 처리를 요구 또는 허용한 경우 (예: 의료법)
    - 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위한 경우
    - 안전행정부장관이 고시하는 경우
  - 보유하고 있는 주민등록번호는 법령에 구체적인 근거가 없는 경우 법 시행 후 2년 이내 파기 ('16년 8월까지 파기)

\* 고유식별정보: 주민등록번호, 외국인등록번호, 여권번호, 운전면허번호

## 생명윤리 및 안전에 관한 법률

- 2015년 12월 29일 일부개정안 시행
- 개인정보 이용 제한
  - 제18조, 제38조, 제43조: 연구를 위해서 반드시 개별 동의를 받거나 또는 익명화 처리를 하여 기관위원회의 심의를 받아야 함
- 인간대상연구
  - 사람을 대상으로 물리적으로 개입하는 연구
  - 설문조사, 행동관찰 등으로 자료를 얻는 연구
  - 개인을 식별할 수 있는 정보를 이용하는 연구
- 인체유래물연구

# 의료기관의 개인정보

## 환자로부터 수집한 정보

- ▶ 성명, 성별, 주소, 직업, 주민등록번호, 전화번호, 휴대폰번호, 이메일주소, 보호자 성명 및 주소, 병력 및 가족력, 주된 증상 등

## 의료인으로부터 생성되는 정보

- ▶ 의료행위에 따른 결과 및 의견정보인 진단결과, 검사결과, 진료경과 및 예견, 의료행위의 내용, 의료인의 성명, 종별면허번호, 서명날인, 진료일시 등

# 개인의료정보란?

- 개인정보보호포털 (<https://www.i-privacy.kr/intro/define1.jsp>)



명확한 정의 없음!

미국의 경우 HIPAA (Health Insurance Portability Accountability Act)에  
의해 명확히 정의

# 의료정보 활용

일차적 이용: 동의 획득 필요 없음

- 진료: 진료 예약, 진단, 검사, 치료, 납 등
- 병원 경영

이차적 이용: 개별 동의 필요 또는 익명화

- 연구, 홍보, 환자 교육 등

\* 관련 법령에서 요구하는 경우도 동의 없이 제공 가능

: 감염병예방법, 암관리법, 후천성면역결핍증예방법, 정신보건법, 결핵예방법 등

# 연구용 의료정보 활용

연구의 경우는 진료 외 목적인 2차 이용이기 때문에 반드시 개별 동의가 필요함

모든 연구 (인간대상연구 & 인체유래물연구)는 IRB의 심의를 받아야함

※ 심의면제 대상도 포함 (신속 심의 가능)

대안은?

익명화한다면 심의면제, 동의면제 대상



## 익명화

- **WHAT?**
- **HOW?**



**De-identification?**  
**Anonymization?**  
**Pseudonymization?**

# ISO/TS 25237

## Health informatics - Pseudonymization

- Principles and requirements for privacy protection using pseudonymization services for the protection of personal health information.

TECHNICAL  
SPECIFICATION

ISO/TS  
25237

First edition  
2008-12-01

---

---

Health informatics — Pseudonymization



# IHE IT Infrastructure Handbook: De-identification

- Explains the PROCESS for removing individually identifiable information from healthcare data.
- Includes de-identification, pseudonymization, re-linking, design considerations, techniques, and risks.

Integrating the Healthcare Enterprise

# IHE

**IHE IT Infrastructure  
Handbook**

**De-Identification**

**Published**

Date: June 6, 2014  
Author: IHE IT Infrastructure Technical Committee  
Email: ITI@ihe.net

# **NISTIR 8053:**

## **De-Identification Personal Information**

- Provides an overview of de-identification issues and terminology.
- Summarizes significant publications to date involving de-identification and re-identification.

**NISTIR 8053**

### **De-Identification of Personal Information**

Simson L. Garfinkel

This publication is available free of charge from:  
<http://dx.doi.org/10.6028/NIST.IR.8053>

**NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

\*2015 October

<http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>

## De-identification

- **General term for any process of removing the association** between a set of identifying data and the data subject

## Anonymization

- Process that **removes the association** between the identifying data set and the data subject
- **One-way de-identification** (cannot be re-identified):  
definition from IHE handbook

## Pseudonymization

- **Particular type of anonymization** that both **removes the association** with a data subject and **adds an association** between a particular set of characteristics relating to the data subject and one or more **pseudonyms**

- **Cryptography**: the transformation of data **in order to hide its information content, prevent its undetected modification and/or prevent its unauthorized use**

# Three levels of the pseudonymization

Level 1:

the risks associated with the **person identifying data elements**



Level 2:

the risks associated with **aggregating data variables**



Level 3:

the risks associated with **outliers** in the populated database

# Outlier Variables

Variables
Rare diagnoses*
Uncommon procedures
Some occupations (e.g. tennis professional)
Certain recessive traits uncharacteristic of the information resources
Distinct deformities

\* 유병률 기준 2만명 이하 또는 적절한 치료방법과 대체의약품이 개발되지 않은 질환  
(보건복지부 "희귀의약품 지정에 관한 기준")

# Genetic information

~~The genetic information is different from other medical information ("genetic exceptionalism")~~

vs.

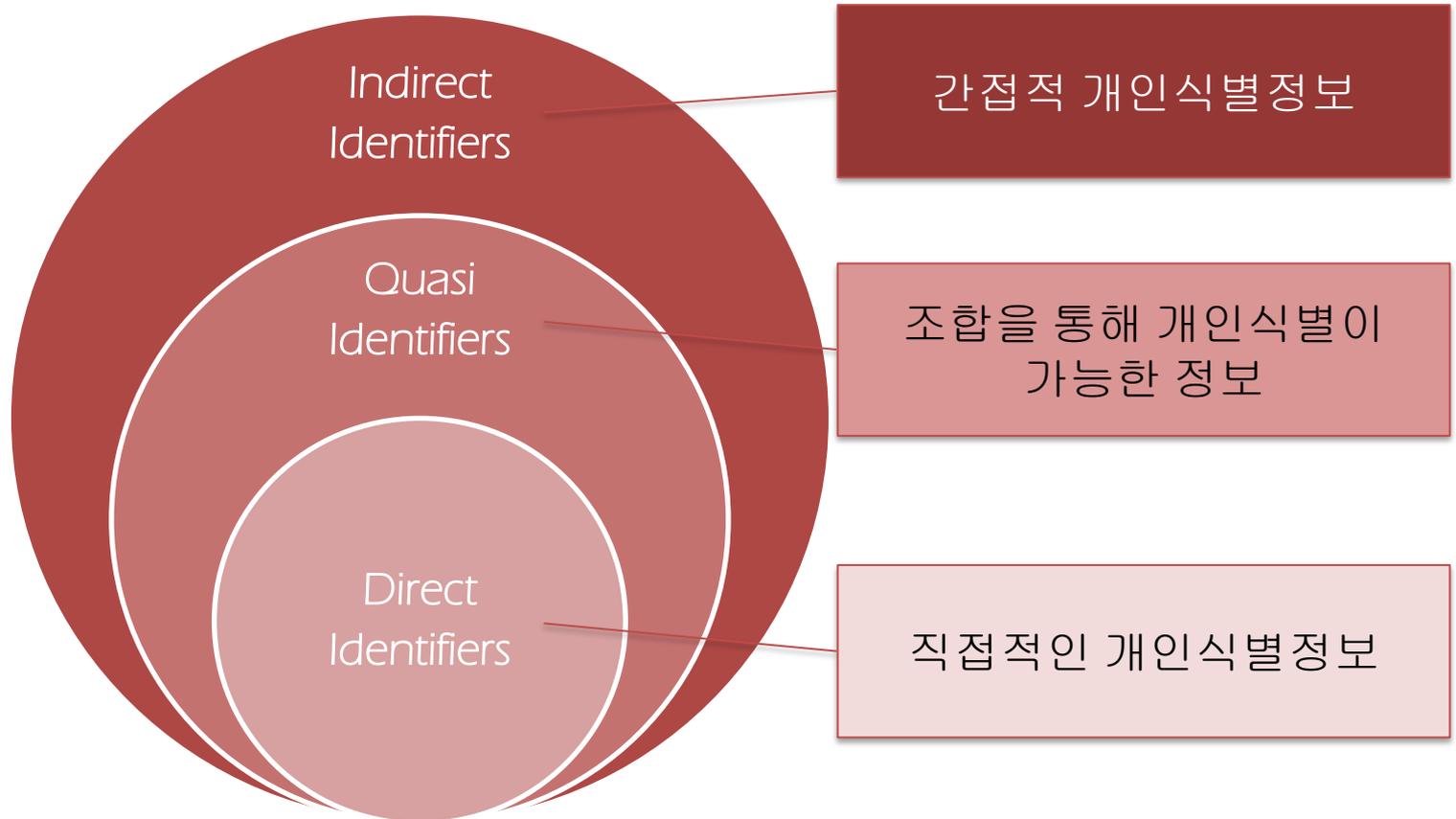
The genetic data is the same as any other medical information



## 익명화

- WHAT?
- HOW?

# Type of Identifiers



# Types of Identifying Data



# HIPAA

- Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule
  - <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/guidance.html>
- By the HITECH Act, HHS must issue **guidance annually** on the "most effective and appropriate technical safeguards for use in carrying out" the HIPAA security standards.



# HIPAA Guideline

The Office of the National Coordinator for  
Health Information Technology



## Guide to Privacy and Security of Electronic Health Information

Version 2.0  
April 2015

*The information contained in this Guide is not intended to serve as legal advice nor should it substitute for legal counsel. The Guide is not exhaustive, and readers are encouraged to seek additional detailed technical guidance to supplement the information contained herein.*

Putting the I in HealthIT  
HealthIT.gov

<http://www.healthit.gov/sites/default/files/pdf/privacy/privacy-and-security-guide.pdf>



ASAN  
Medical Center

# HIPAA 18 Protected Health Information

(2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names (excluding physician's name)	
(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census: <ol style="list-style-type: none"> <li>(1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and</li> <li>(2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000</li> </ol>	
(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older	
(D) Telephone numbers	(L) Vehicle identifiers and serial numbers, including license plate numbers
(E) Fax numbers	(M) Device identifiers and serial numbers
(F) Email addresses	(N) Web Universal Resource Locators (URLs)
(G) Social security numbers	(O) Internet Protocol (IP) addresses
(H) Medical record numbers	(P) Biometric identifiers, including finger and voice prints
(I) Health plan beneficiary numbers	(Q) Full-face photographs and any comparable images
(J) Account numbers	(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section; and
(K) Certificate/license numbers	

(ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.

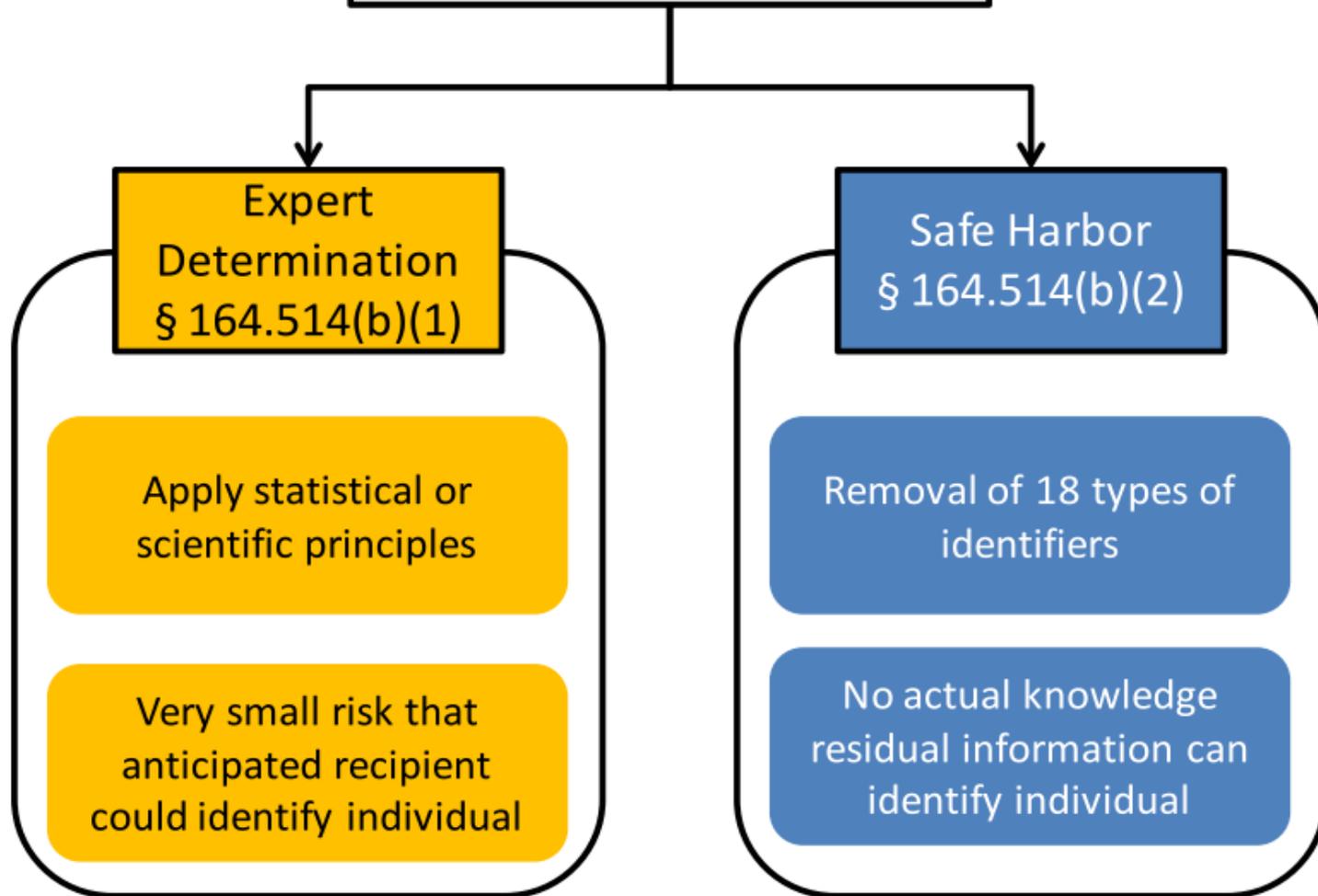
# Limited Data Set

- Refers to **PHI that excludes 16 categories of direct identifiers** and may be **used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual's Authorization or a waiver** or an alteration of Authorization for its use and disclosure, with a data use agreement.

[http://privacyruleandresearch.nih.gov/pr\\_08.asp](http://privacyruleandresearch.nih.gov/pr_08.asp)

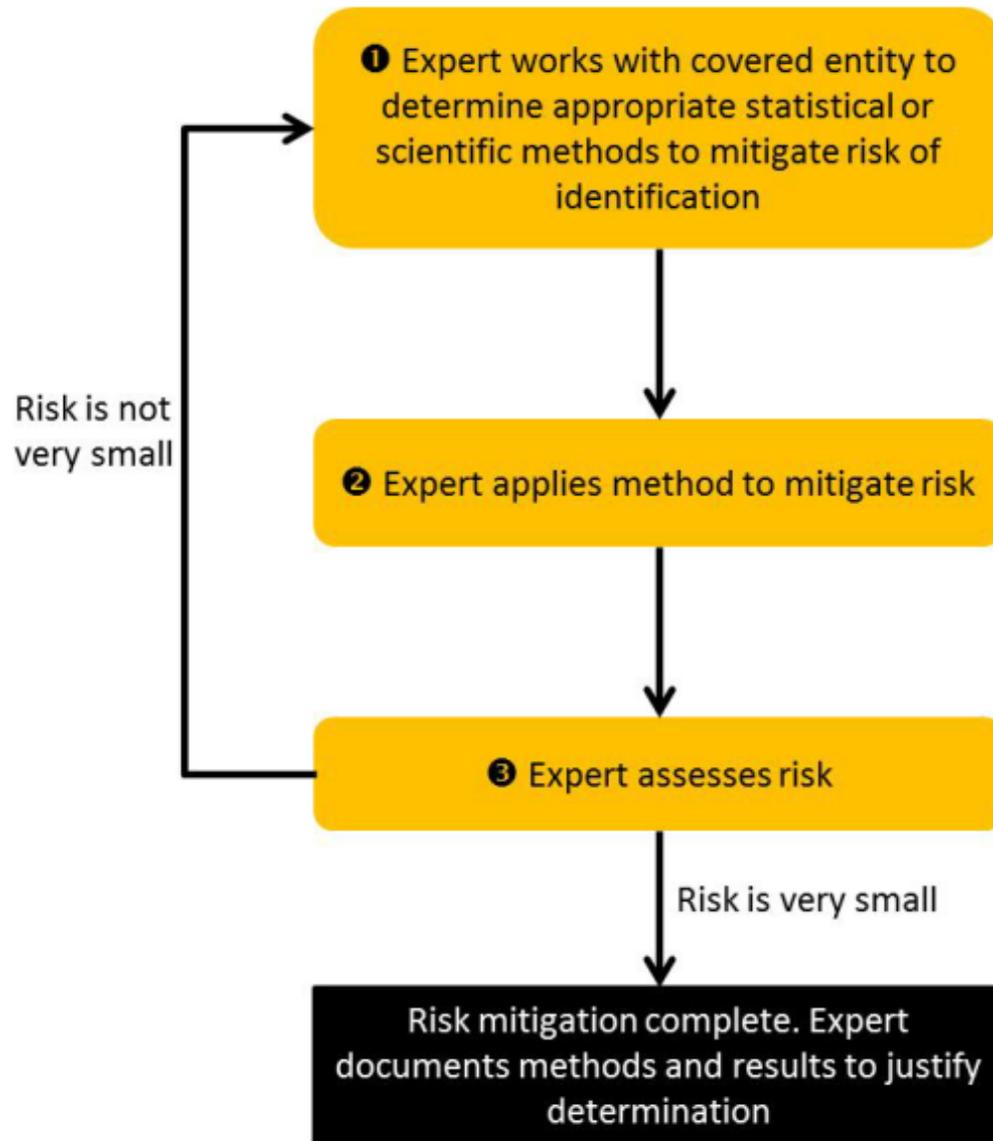
18 PHI	16 direct identifiers of Limited Data Set
Names	Names
All geographic subdivisions smaller than a state,	Postal address information, other than town or city, state, and ZIP Code
<b>All elements of dates (except year)</b>	
Telephone numbers	Telephone numbers
Facsimile numbers	Facsimile numbers
Electronic mail addresses	Electronic mail addresses
Social security numbers	Social security numbers
Medical record numbers	Medical record numbers
Health plan beneficiary numbers	Health plan beneficiary numbers
Account numbers	Account numbers
Certificate/license numbers	Certificate/license numbers
Vehicle identifiers and serial numbers, including license plate numbers	Vehicle identifiers and serial numbers, including license plate numbers
Device identifiers and serial numbers	Device identifiers and serial numbers
Web universal resource locators (URLs)	Web universal resource locators (URLs)
Internet protocol (IP) address numbers	Internet protocol (IP) address numbers
Biometric identifiers	Biometric identifiers
Full-face photographic images and any comparable images	Full-face photographic images and any comparable images
<b>Any other unique identifying number, characteristic, or code</b>	

# HIPAA Privacy Rule De-identification Methods



# Expert Determination Method

- Have expert determination been applied outside of the health field?
  - YES
- Who is an “expert?”
  - There is no specific professional degree or certification program
- What is an acceptable level of identification risk for an expert determination?
  - There is **no explicit numerical level** of identification risk. An expert will define an acceptable “very small” risk.



**Figure 2. Process for expert determination of de-Identification.**

# 개인정보보호 가이드라인

- 방송통신위원회, 빅데이터 개인정보보호 가이드라인, 2014.12.23  
(붙임1)

## 「빅데이터 개인정보보호 가이드라인」 주요내용

- ① 수집 시부터 개인식별 정보에 대한 철저한 비식별화 조치(제3조·제4조·제5조·제10조)
  - 개인정보가 포함된 공개된 정보 및 이용내역정보는 비식별화 조치를 취한 후 수집·저장·조합·분석 및 제3자 제공 등 가능
- ② 빅데이터 처리 사실·목적 등의 공개를 통한 투명성 확보(제4조·제5조·제9조)
  - 개인정보 취급방침을 통해 비식별화 조치 후 빅데이터 처리 사실·목적·수집 출처 및 정보 활용 거부권 행사 방법 등을 이용자에게 투명하게 공개

· (개인정보 취급방침) 비식별화 조치 후 빅데이터 처리 사실·목적 등을 이용자 등에게 공개하고 '정보 활용 거부 페이지 링크'를 제공하여 이용자가 거부권을 행사할 수 있도록 조치

· (수집 출처 고지) 이용자 이외의 자로부터 수집한 개인정보 처리 시 '수집 출처·목적, 개인정보 처리 정지 요구권'을 이용자에게 고지

# 개인정보 활용 방법: 비식별화

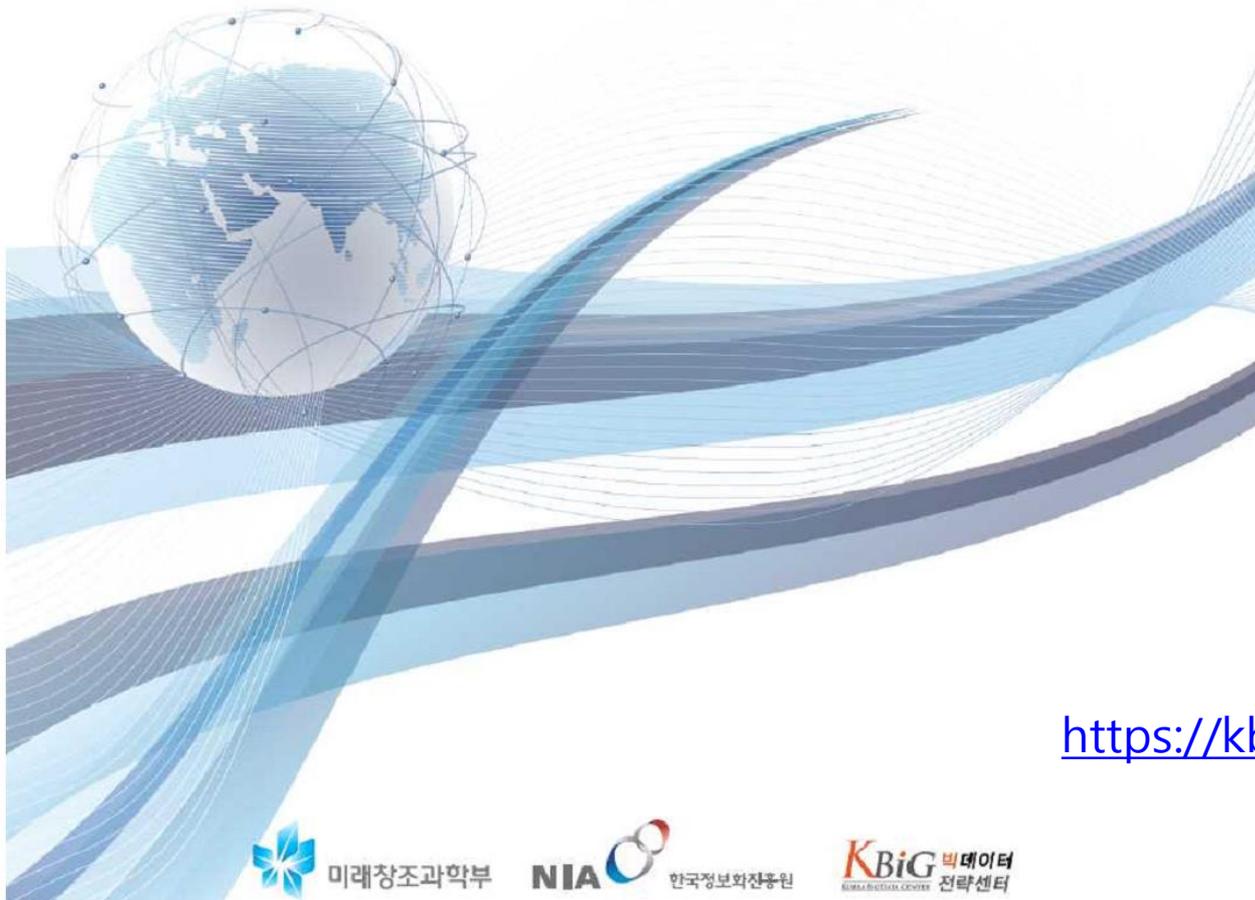
4. “비식별화”란 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없도록 하는 조치를 말한다.

제7조(민감정보 생성의 금지) 특정한 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 이용자의 사생활을 현저히 침해할 우려가 있는 정보의 생성을 목적으로 공개된 개인정보 등을 수집·저장·조합·분석 등 처리하여서는 아니 된다. 다만, 이용자의 사전 동의를 받거나 법률에 따라 허용된 경우에는 그러하지 아니하다.

**비식별화 혹은 익명화를 하더라도 상업적 목적은 불가**

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

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



<https://kbig.kr/?q=지식자료실/15596>



미래창조과학부

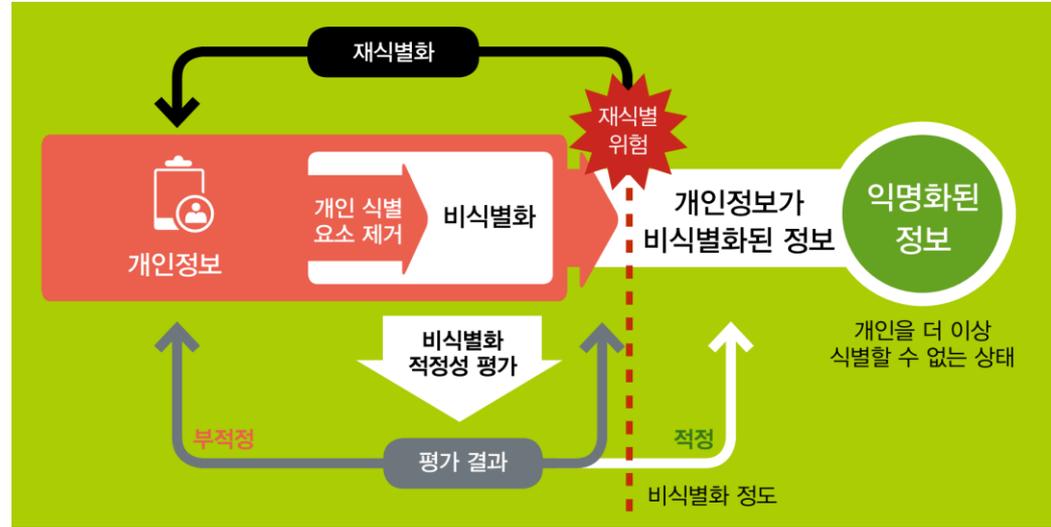
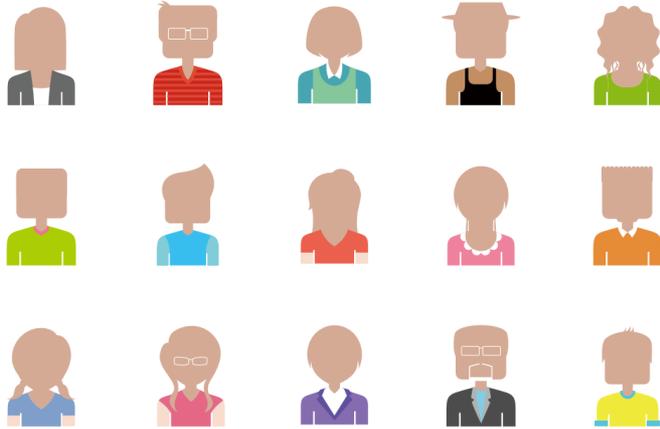


한국정보화진흥원



ASAN  
Medical Center

# 개인정보 비식별화에 대한 적정성 자율평가 안내서



## ◎ 평가위원회 구성

- 개인정보처리자의 개인정보보호책임자(CPO)가 지정하는 3인 이상의 평가위원으로 구성
- ※ 평가위원의 과반수 이상은 외부의 전문가로 지정하고, 업무영역 전문가 1인, 개인정보 비식별화 전문가 1인, 법률전문가 1인은 필수적으로 포함하여 구성

## ◎ 평가 수행

- 개인정보처리자에서 작성한 '기초자료'와 '세부 평가 방법'을 기반으로 평가를 수행하고 '적정' 또는 '부적절' 의견 제시
- ※ 적정: 다른 정보와 결합해도 특정 개인을 식별할 수 없는 상태를 의미
- ※ 부적절: 추가적인 개인식별요소 제거 조치가 필요한 상태를 의미

## ◎ 추가 비식별화 조치

- 평가위원회의 의견에 기초하여 평가 대상 데이터에 대해 개인식별요소 제거 조치 등을 추가적으로 수행
- ※ 추가 비식별화 조치 이후 비식별 조치가 적절하게 수행되었는지 평가 수행

[http://www.privacy.go.kr/inf/rfr/selectBoardArticle.do?nttId=5951&bbsId=BBSMSTR\\_00000000044](http://www.privacy.go.kr/inf/rfr/selectBoardArticle.do?nttId=5951&bbsId=BBSMSTR_00000000044)



sooyong.shin@amc.seoul.kr

 @likesky3



**ASAN**  
Medical Center