2025 Vol.8

PRIVACY REPORT

개인정보 이슈 심층분석

의료분야 개인정보

KISA 한국인터넷진흥원

PRIVACY REPORT

2025년
개인정보 이슈
심층분석

CONTENTS

의료분야 개인정보 보호 강화 기술 현황

신수용
Kakao Healthcare I 선행기술연구소장 겸 개인정보보호책임자

1. 들어가며

의료 데이터는 개인정보보호법에서 명시하는 민감정보로 특별히 보호받아야 하는 정보로 분류되어, 데이터 활용에 엄격한 제한을 받고 있다. 하지만, 생성형AI로 대표되는 AI기술의 비약적인 발전에 힘입어 아주 많은 의료AI들이 개발되는데, 2025년 11월 기준으로 미국 식품의약국(Food and Drug Administration, FDA) 홈페이지 자료에 의하면 모두 1,247개의 AI의료기기가 인허가를 받았을 정도이다.[1] 의료 데이터는 철저히 보호받아야 하지만, 의료의 질을 향상시키기 위한 AI 의료기기를 개발하기 위해서는 대규모 의료 빅데이터를 활용해야 한다는 모순적인 상황에 부딪히게 된다. 이러한 상황을 해결하기 위해서 개인정보 보호 강화 기술(Privacy Enhancing Technology, PET)이 주목을 받고 있다. 본 글에서는 의료분야에서 많이 적용되고 있는 PET 기술에 대해서 기본적인 개념 정리부터 시작하여 실제 적용 사례를 소개하고자 한다.

2. PET 기술

현재 개발되고 있는 PET 기술들은 다양하나, 대표적인 PET 기술로는 기본적인 가명처리(Pseudonymization)부터 시작하여 차분 프라이버시(Differential Privacy), 합성데이터(Synthetic Data), 보안 다자간 연산(Secure Multiparty Computation, SMC) 연합학습(Federated Learning, FL), 동형암호(Homomorphic Encryption, HE) 등이 있다. 이를 데이터 변형과 데이터 활용·연산으로 구분해서 정리하면 다음과 같다.

1) 데이터 변형

데이터 변형은 원본 데이터를 직접적인 수정, 변형 또는 대체하여 개인을 식별할 수 없도록 만드는 기술로, 가명처리, 차분 프라이버시, 합성데이터 등의 기술이 대표적이다. 각 기술의 간단한 설명과 개인정보 보호 수준, 데이터 활용의 유용성 및 장단점을 비교하면 아래 표와 같다.

1) FDA, Artificial Intelligence-Enabled Medical Devices, 2025.07.10.

표1 주요 데이터 변형 기술

	가명처리	합성데이터	차분 프라이버시
개요	원본 데이터에서 식별 가능한 요소를 대체값(가명)으로 변경	원본 데이터의 통계적 속성을 모방하여 데이터를 합성해 내고, 원본 데이터 대신 활용	데이터 또는 분석 결과에 노이즈를 추가하여 결과로부터 특정 개인 식별 가능성을 통제
보호 수준	중	상	상
데이터 유용성	높음. 원본 데이터의 구조와 속성이 대부분 유지	중간 ~ 높음. 원본 데이터의 통계적 특성은 유지되나, 개별 데이터 정확성은 떨어짐	낮음 ~ 중간. 노이즈가 추가될수록 정확도가 떨어져 데이터 유용성이 떨어짐
주요 장점	기존 시스템 적용 용이	재식별 위험이 낮아 데이터의 공유 및 활용 가능하며, 민감정보 대체 가능	프라이버시 보호 수준을 수학적으로 명확하게 정의하고 보장
주요 단점	다른 데이터와의 결합 시 재식별 가능성 존재	데이터를 생성하는 모델이 복잡하며, 생성된 데이터의 통계적 정확성을 검증하기 어려움	데이터 유용성이 높지 않음

가명처리는 가장 대표적이면서도 기본적인 PET 기술이다. 하지만, 가명정보는 여전히 법적으로 개인정보에 포함된다는 것을 유념해야 한다. 구체적인 기술들은 개인정보위에서 발표한 “가명정보 처리 가이드라인[2]”에 상세히 설명되어 있고, 보건의료 분야에서 어떻게 처리하면 되는지는 보건복지부 “보건의료데이터 활용 가이드라인[3]”에 구체적인 예시와 함께 제시되어 있다. 가명처리는 개인정보 보호를 위해서 가장 기본적이고 필수적인 기술이기 때문에, 개인 동의를 받지 않은 경우에는 반드시 적용해야 한다. 현재 “개인정보 보호법”에서는 “가명처리”라는 표현을 쓰고 있고[4], 보건의료 분야 연구에 해당하는 특별법인 “생명윤리 및 안전에 관한 법률”에서는 “익명화”라는 표현을 쓰고 있다.[5] 이 부분은 법적 정의 일치를 위하여 추후 개정이 필요한 사항이며, 국제표준화기구(International Organization for Standardization, ISO)에서는 De-identification이 Anonymization과 Pseudonymization을 포함하는 개념으로 사용되고 있다.

합성데이터와 관련해서는 개인정보보호위원회가 24년 5월 “합성데이터 생성 참조모델” 가이드라인과[6] 24년 12월 “합성데이터 생성·활용 안내서”를 발표하여[7] 합성데이터 활용을 지원하고 있다. 하지만 생성된 합성데이터가 원래 데이터를 얼마나 잘 반영하고 있는지 평가에 대한 기준을 만

2) 개인정보보호위원회, 가명정보 처리 가이드라인, 2024.02.05.
3) 보건복지부, 보건의료데이터 활용 가이드라인, 2024.12.16.
4) 개인정보 보호법 제2조 1의2, “가명처리”란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 말한다.
5) 생명윤리 및 안전에 관한 법률 제2조 19, “익명화”(匿名化)란 개인식별정보를 영구적으로 삭제하거나, 개인식별정보의 전부 또는 일부를 해당 기관의 고유식별기호로 대체하는 것을 말한다.
6) 개인정보보호위원회, 합성데이터 생성 참조모델, 2024.05.30.
7) 개인정보보호위원회, 합성데이터 생성·활용 안내서, 2024.12.18.

들기 어렵고, 특히 생성형AI가 가지고 있는 환각현상(hallucination) 문제와 동일하게 현실적으로 존재할 수 없는 데이터(예를 들어, 실제 환자에게 존재할 수 없는 데이터)를 생성해 낼 수 있기 때문에, 합성데이터의 한계성을 분명히 인식하고 사용해야 한다.

데이터 변형을 위주로 하는 기술들은 기본적으로 원천 데이터를 왜곡하는 행위이기 때문에, 데이터의 품질이 떨어지거나 특정한 목적의 분석에는 사용할 수 없다는 한계를 가지고 있다. 이러한 특징으로 인하여 개인별 원천 데이터를 필요로 하는 개인맞춤형 의학 실현을 위해서는 기술 개발에 어느 정도 한계가 있다는 점을 유념해야 한다.

2) 데이터 활용·연산

데이터의 내용은 그대로 유지하면서, 암호화하거나 데이터를 외부로 반출하지 않으면서 필요로 하는 연산을 가능하게 하는 방식으로, 계산 환경 또는 방식에 변경을 주는 기술이다. 데이터 유출 및 비밀을 유지하면서 협력적 활용을 보장한다. 각 기술에 대한 설명과 주요 장단점은 아래 표와 같다.

표2 주요 데이터 활용·연산 기술

	SMC	FL	HE
핵심 방식	여러 참여기관이 데이터를 공개하지 않으면서 공동의 연산을 수행	데이터의 중앙집중 없이 개별 기관에서 모델 학습 수행 후, 모델 파라미터만 중앙에서 취합하여 통합 모델 생성	데이터를 암호화된 상태 그대로 두고 복호화 없이 연산을 수행
보호 대상	참여기관 데이터에 대한 비밀성 보장	로컬 원본 데이터의 외부 유출 방지.	데이터 및 연산 결과의 기밀성 보장.
주요 장점	다양한 연산 적용 가능	원본 데이터 이동이 없으며, 높은 효율을 보임	복호화가 불필요하여 가장 강력함
주요 단점	참여 기관이 많아지면 복잡성 증가	모델 가중치를 통해 원본 데이터를 유추 가능성 존재	매우 느린 연산 속도

FL이 현재 데이터를 외부에 반출하지 않기 때문에, 의료기관들이 현재 가장 선호하는 방식이다. 동형암호의 경우 데이터가 암호화된 상태로 공유되기 때문에 개인정보 유출에 대한 우려가 없고, 개인적으로는 가장 확실한 개인정보 보호를 위한 PET가 될 것으로 기대하고 있다. 특히 기술적으로 한국이 선도하고 있는 분야이기도 하다. 하지만 아직까지 암호화된 데이터를 분석하는 기술이 많은 연산량을 필요로 한다는 한계가 있어 실제 상업적 적용 사례는 찾기 어렵다.

이 기술들은 단독으로 사용될 수도 있지만, 많은 경우 여러 개의 기술이 결합되어 서로의 단점을 보완하면서 사용되고 있고, 각 경우들을 정리하면 아래 표와 같다.

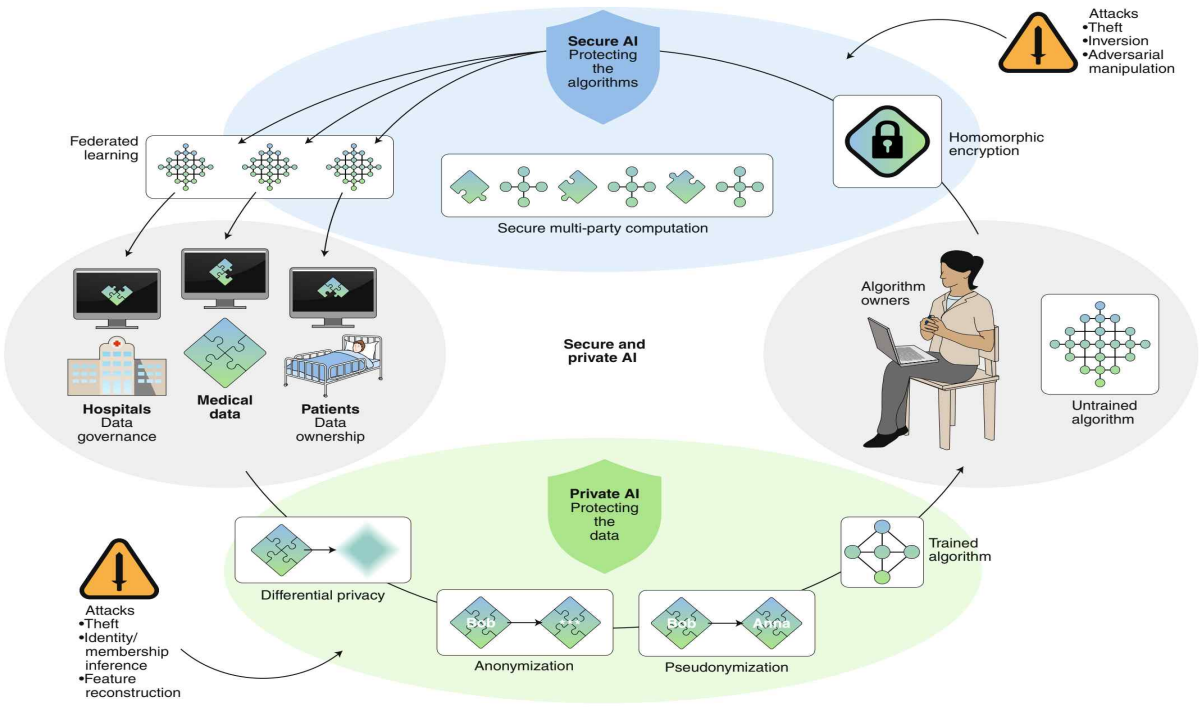
표3 FL 결합 기술 사례

결합 기술	주요 기능	장점
FL+SMC	모델 가중치 합산의 비밀성 보장	FL 과정 중 중앙 서버가 로컬 모델 업데이트 과정에서 전체 파라미터를 확인할 수 없음
FL+HE	암호화된 가중치 합산 및 복호화 불필요	민감한 파라미터나 데이터가 암호화되어 처리됨. HE의 덧셈 동형(Additive Homomorphism) 특성을 활용하여 가중치 합산에 매우 효율적
SMC+HE	복잡한 비선형 연산의 처리 및 효율화	MPC가 처리하기 어려운 비선형 함수나 비교 연산을 HE의 기능으로 처리하여 효율성 증가. HE로 암호화된 데이터 위에서 MPC를 사용하여 보안 레벨을 극대화
FL+SMC +HE	종합적인 프라이버시 보호 학습 환경	FL의 데이터 분산 이점과 MPC/HE의 암호화 연산 이점을 모두 결합 중앙 서버와 로컬 기기 모두를 신뢰하지 않는 환경(Zero-Trust)에서 안전하게 협력 학습을 수행

3. 보건의료 분야 적용 실제 사례

의료분야 PET 적용 사례는 아래 그림에 잘 정리되어 있다.[8)]

그림1 의료분야 PET 적용 개념도



8) Nature, Secure, privacy-preserving and federated machine learning in medical imaging, 2020.06.08.

전 세계 모든 의학 논문이 관리되는 미국 국립의학도서관의 학술 논문 검색 데이터베이스(PubMed)에 등록되어 있는 논문들을 대상으로 조사를 해 보면[9] “Pseudonymization”으로 검색하면 무려 1876년 논문부터 검색이 되고 있으며, “Pseudonymization” 또는 “de-identification”으로 검색하면 모두 144,994건의 논문이 발견된다. “Synthetic data”의 경우도 1971년 논문을 최초로, 모두 5,269건의 논문이 검색된다. “Differential privacy”로 검색을 하면 2011년 논문이 최초이며, 모두 479건의 논문이 발견된다. SMC의 경우 2014년이 시작이며 모두 80건의 논문, HE의 경우 2012년에 시작하여 311건의 논문, FL은 2017년 시작하여 1,879건의 논문이 검색된다. 즉, 가명처리가 압도적으로 많이 활용되고 있으며, 데이터 획득의 어려움으로 인하여 합성데이터도 많이 사용되고 있는 것을 확인할 수 있다. 그리고 최근에는 FL이 부각되고 있는 현상을 알 수 있다.

국내에서도 여러 정부 R&D 과제에서 FL이 핵심기술로 등장하고 있다. 대부분 의료·바이오 분야에 집중되어 있는데 이는 역시나 민감 데이터 활용과 관련하여 FL이 현재 상황에서 현실적인 대안 중 하나이기 때문이다. 대표적인 과제들로는 보건복지부, 과기정통부 주관의 “연합학습 기반 신약 개발 가속화 프로젝트(K-MELLODDY)”, 보건복지부 주관의 “K-AI 신약 개발 전임상·임상 모델 개발 사업” 등이 있다. 두 과제 모두 대형과제로 신약 개발에 초점을 두고 있고, 그 외에도 복지부에서 “다기관-멀티모달 연합학습 기반 의료 인공지능 기술 시범모델 개발 사업”을 진행하고 있고, 해당 과제는 전자 의무기록(Electronic Medical Record, EMR) 데이터를 연합학습으로 학습을 해 파운데이션 모델(Foundation Model)을 만드는 것을 목표로 진행되고 있다. 이런 정부 R&D 과제들 외에도 국내 여러 기업도 연합학습을 통하여 사업을 진행하고 있는데, 대표적인 기업들로 카카오헬스케어(Kakao Healthcare)와 에비드넷(Evidnet) 등이 있다. 국내에서 가장 큰 연합학습 사례는 2023년부터 카카오헬스케어가 주도한 대학병원 연합학습 플랫폼으로, 대형병원들의 임상데이터를 효과적으로 분석할 수 있는 클라우드 기반 헬스케어 데이터 플랫폼을 구축하고, 이에 연합학습 환경을 접목하여 병원 간 안전한 협력을 지원하고 있다.

이처럼 의료분야에서는 대상으로 하는 건강정보의 높은 민감성으로 인하여 가명처리는 필수적으로 사용되고 있으며 여러 최신 기술들을 도입하여 개인정보 보호를 위한 노력을 하고 있다. 앞으로도 데이터 활용과 개인정보 보호라는 두 가지 목적을 동시에 만족시키기 위해서 지속적인 기술 개발이 필요한 상황이다.

9) 2025년 11월 18일 검색 결과

참고 문헌 |

1. FDA, Artificial Intelligence-Enabled Medical Devices, 2025.07.10.
2. 개인정보보호위원회, 가명정보 처리 가이드라인, 2024.02.05.
3. 보건복지부, 보건의료데이터 활용 가이드라인, 2024.12.16.
4. 개인정보보호위원회, 합성데이터 생성 참조모델, 2024.05.30.
5. 개인정보보호위원회, 합성데이터 안내서 마련, 2024.12.18.
6. Nature, Secure, privacy-preserving and federated machine learning in medical imaging, 2020.06.08.

의료데이터의 안전한 활용을 위한 실제와 거버넌스
: 연세의료원 사례 심층분석

임준석
연세대학교 의료원 디지털헬스실 I 실장 I 교수

1. 요약

연세의료원은 국내 의료기관 중 선도적으로 의료데이터 2차 활용 거버넌스 체계를 구축하여 운영하고 있습니다. 데이터 3법 개정에 따른 가명정보 활용 기반 위에서 기관생명윤리위원회(Institutional Review Board, IRB)-데이터심의위원회(Data Review Board, DRB)-의료자산활용위원회의 3단계 심의 체계를 구축하고, 실질적인 데이터 보호와 활용의 균형을 달성하고 있습니다. 본 사례 분석은 법적 근거와 실무의 연결, 다층적 비식별화 전략, 국제 표준 기반 데이터 품질관리, 그리고 Honest Broker 시스템을 통한 안전한 데이터 결합 체계 등 의료데이터 거버넌스의 핵심 요소들을 실증적으로 제시합니다.

2. 서론: 의료데이터 2차 활용의 필요성과 도전

의료데이터의 2차 활용은 신약 개발, 질병 예측, 공공보건 연구 등 다양한 분야에서 혁신을 이끌고 있습니다. 그러나 개인정보 보호와 데이터 활용의 균형, 데이터 품질 및 표준화, 안전한 반출 및 결합 체계 등 복합적 과제가 동반됩니다. 연세의료원은 국내 의료기관 중 선도적으로 데이터 거버넌스 체계를 구축하고, 실제 운영 경험을 축적해 왔습니다.

연세의료원은 「데이터 3법」(「개인정보 보호법」·「정보통신망 이용촉진 및 정보보호 등에 관한 법률」·「신용정보의 이용 및 보호에 관한 법률」) 개정을 통해 마련된 가명정보 처리 및 활용 근거에 기반하여 의료데이터 2차 활용 체계를 운영하고 있습니다. 데이터 3법 개정은 가명정보를 통계작성, 과학적 연구, 공익적 기록보존 등의 목적으로 활용할 수 있도록 허용하였으며, 이에 따라 보건의료 분야에서도 진료기록 등 민감정보를 안전하게 활용할 수 있는 제도적 장치가 필수적으로 요구되고 있습니다.

보건복지부의 「보건의료데이터 활용 가이드라인」은 이러한 법적 근거를 구체화하여, 의료기관이 데이터 활용 시 준수해야 할 절차와 안전조치를 명시하고 있습니다.

3. 환자 데이터의 여정: 진료부터 연구 활용까지

그림1 환자 데이터 연구 활용 절차

1) 데이터 생성 단계

환자가 연세의료원에 내원하여 진료받는 순간부터 데이터의 여정이 시작됩니다. 의료진은 전자의무기록(Electronic Medical Record, EMR) 시스템을 통해 진단, 처방, 검사, 시술 등 모든 의료 행위를 실시간으로 기록합니다. 이때 생성되는 데이터는 구조화된 데이터(진단 코드, 약물 코드 등)와 비구조화된 데이터(경과 기록, 영상판독 소견 등)로 구분됩니다.

2) 표준화 및 품질관리

기록된 데이터는 즉시 표준화 과정을 거칩니다. 의료 용어는 SNOMED-CT, ICD-10, LOINC 등 국제 표준에 따라 매핑되며, 데이터 품질관리 시스템을 통해 오류나 누락이 실시간으로 검증됩니다. 이 과정에서 임상적 의미를 최대한 보존하면서도 연구 활용 가능성을 높이는 것이 핵심입니다.

3) 비식별화 처리

연구 목적으로 활용되는 모든 데이터는 개인정보 보호법에 따른 가명처리를 거칩니다. 환자의 성명, 주민등록번호 등 직접 식별정보는 완전히 제거되거나 암호화되며, 진료 일자, 검사 일자 등 간접 식별 가능성이 있는 정보는 연구 목적에 따라 적절한 수준으로 변환됩니다.

4) 통합 저장소 구축

비식별화된 데이터는 데이터 레이크(Data Lake) 기반의 통합 저장소에 적재됩니다. 여기에는 임상데이터뿐만 아니라 유전체 데이터, 영상 데이터, 생체신호 데이터 등이 환자별로 연결되어 저장되어, 연구자들이 다양한 관점에서 분석할 수 있는 토대가 됩니다.

4. 법적 기반과 제도적 프레임워크

1) 데이터 3법과 가명정보 활용

2020년 8월 시행된 데이터 3법은 가명정보의 활용 범위를 통계작성, 과학적 연구, 공익적 기록보존 등으로 확대하였습니다. 이를 통해 의료기관은 환자의 별도 동의 없이도 가명처리된 의

료데이터를 연구 목적으로 활용할 수 있는 법적 근거를 확보하게 되었습니다.

2) 보건의료데이터 활용 가이드라인

보건복지부는 2021년 「보건의료데이터 활용 가이드라인」을 발표하여 의료기관이 준수해야 할 구체적인 절차와 안전조치를 제시하였습니다. 이 가이드라인은 다음과 같은 핵심 원칙을 담고 있습니다.

① 가명처리의 적정성 확보
② 목적 외 사용 및 재식별 금지
③ 안전성 확보조치 의무
④ 영향평가 수행
⑤ 처리 현황 공개

3) 연세의료원의 법적 준수 체계

연세의료원은 데이터 3법과 보건의료데이터 활용 가이드라인을 바탕으로 자체적인 데이터 거버넌스 정책을 수립하고, 이를 기관생명윤리위원회(Institutional Review Board, IRB-DRB-의료자산활용위원회의 3단계 심의 체계를 통해 실질적으로 구현하고 있습니다.

5. 데이터 플랫폼과 비식별화: 기술적·관리적 조치의 실제

1) 통합 데이터 플랫폼

연세의료원은 통합연구플랫폼(Clinical Data Warehouse, CDW), 공통데이터모델(Common Data Model, CDM), 암 정밀 의료 DB(Yonsei Cancer Data Library) 등 다양한 임상·유전체 데이터를 데이터 레이크 기반으로 통합 관리합니다. 모든 데이터는 개인정보 보호법 등 관련 법령에 따라 비식별화(가명처리) 과정을 거치며, 연구 목적과 데이터 민감도에 따라 추가 보안 조치가 적용됩니다.

2) 다층적 비식별화 전략

데이터의 원본성을 최대한 보존하면서도 재식별 위험을 최소화하기 위해 다음과 같은 다층적 비식별화 전략을 적용합니다.

3) 비식별화 조치 세부 사항

① 다층적 비식별화: 진료일, 수술일 등 식별 가능성이 높은 변수는 연·월 단위로 변환하거나, 차이값으로 대체
② 민감정보 보호: 희귀 질환, 마약류, 환자 및 의료진의 위치정보 등은 반드시 폐쇄망 환경에서만 분석이 가능
③ 암호화 적용: 환자 식별자는 일방향 암호화를 통해 연구용 ID로 변환

④ 데이터 마스킹: 자유 텍스트 내 개인정보는 자동 탐지 및 마스킹 처리

4) 보안 강화된 분석 환경

연구자들은 데이터 검색, 맞춤형 데이터 제작 지원, 클라우드 기반 분석 환경을 통해 원 내외 다기관 연구를 위한 기반을 제공받습니다. 민감정보가 포함된 연구의 경우 폐쇄망 클라우드 분석 환경에서만 분석이 가능하도록 제한하여, 데이터 활용의 유연성과 개인정보 보호의 균형을 실질적으로 구현합니다.

6. 데이터 품질관리와 표준화

1) EMR 시스템 기반 표준화 체계

연세의료원은 2005년 EMR 시스템 구축을 시작으로, 진단, 수술, 처치, 주호소, 간호 등 임상 용어의 마스터를 구축하고 표준화 활동을 체계적으로 수행해 왔습니다. 임상적 현상과 증상 등 의료 현장의 다양한 데이터를 최대한 원형 그대로 보유하기 위해 사용자 정의 용어를 등록하되, 각 용어별로 국내외 표준코드와 함께 매핑하여 적용하였습니다.

2) 국제 표준 코드 체계 적용

표1 국제 표준 코드 체계

분야	적용 표준	설명
진단	SNOMED-CT, ICD-10(KCD7), ICD-9-CM	국제 표준 진단 분류체계
수술/처치	ICHI, ICD-10-PCS	의료 행위 분류 표준
검사	LOINC	검사 및 관찰 식별자
약물	RxNorm	의약품 표준화 체계
분류	AHRQ CCS	임상 분류 소프트웨어

3) 데이터 거버넌스 조직체계 및 품질관리

표준화위원회를 비롯한 데이터 거버넌스를 구성하여 다음과 같은 지속적인 유지와 개선 활동을 수행하고 있습니다.

① 용어 분류체계 매핑
② 데이터 품질관리 대상 및 정제 기준 심의
③ 데이터 품질 모니터링 및 오류 개선
④ 공통 데이터모델 선정 및 적용 기준 정의

⑤ 데이터 품질인증 획득 및 지속적 관리
⑥ 품질관리 솔루션 고도화
⑦ 업무절차서 마련 및 정기 업데이트

이러한 체계적인 거버넌스와 품질관리를 통해 데이터 신뢰성을 확보하고 있으며, 연구자들이 신뢰할 수 있는 고품질의 표준화된 데이터를 활용할 수 있는 기반을 마련하였습니다.

7. 데이터 활용 거버넌스와 심의 체계

1) 3단계 심의 체계

연세의료원의 데이터 활용연구 심의 절차는 IRB-DRB-의료자산활용위원회 3단계로 운영됩니다.

표2 데이터 활용연구 심의 절차

위원회	주관 부서	주요 심의 내용
기관생명윤리위원회(IRB)	임상연구관리실 임상연구보호센터	연구 윤리와 연구 타당성
데이터심의위원회(DRB)	디지털헬스실 데이터서비스팀	연구 데이터 적정성
의료자산활용위원회	의과학연구처 기술사업팀	의료원 자산 보호

2) DRB 심의의 기본 원칙

① 가명 또는 익명 데이터 연구는 IRB 심의 및 의료원 DRB 심의 필수
② 연구자는 반드시 개인정보보호 대책(개인식별정보 삭제, 변경, 암호화 등)을 수립하여 데이터를 활용해야 하며, 그 적절성을 IRB에서 심의
③ 다기관 공동 연구 또는 대용량 데이터 사용 연구 진행 시, 의료자산활용위원회 상정
④ 가명 데이터의 제공 및 분석은 의료원 내에서 수행하는 것을 원칙으로 하고 있음
⑤ 연구 효율성과 접근성을 높이기 위해 보안 요건을 충족한 클라우드 기반 원외 접속 방식도 시험적으로 운영 중

3) DRB 시정 승인(조건부 승인) 사례: 비식별 강화(일자 데이터)

진료일, 수술일, 검사일 등 해당 일자로 환자를 특정할 수 있는 경우에는 해당 일자 데이터를 연, 월로 변경 후 일자 간 차이를 계산하는 방식으로 추가 비식별화를 시행합니다. 일 단위 이하 데이터를 반드시 이용해야 하는 경우는 사유를 확인한 후, 의료원 클라우드(폐쇄망) 사용 조건으로 DRB 시정 승인(조건부 승인)하고 있습니다.

8. 데이터 결합 및 반출 거버넌스

1) 중립적 데이터 중개자(Honest Broker) 시스템

연세의료원은 중립적 데이터 중개자(Honest Broker) 체계를 통해, 국민건강보험공단·심평원 등 결합전문기관과의 데이터 결합을 안전하게 수행합니다. 결합키 생성, 폐쇄망 전송, 결합 후 데이터 세트 제공 등 일련의 과정에서 개인정보 유출 위험을 최소화합니다.

Honest Broker는 정보를 제공하는 사람과 정보를 이용하는 사람 사이를 이어주는 제3자로, 정보 보안 조치를 수행하여 정보 이용자의 연구 목적에 맞게 최적화된 형태로 데이터를 제공하는 역할을 수행합니다.

데이터서비스팀에서는 결합전문기관(국민건강보험공단, 건강보험심사평가원, 국립암센터)과 연세의료원의 데이터를 연계하는 역할을 수행하고 있고, 이때 각 기관의 데이터 결합을 위한 결합키를 생성하고 연구자가 원하는 데이터 세트를 생성하여 결합전문기관의 폐쇄망 환경으로 전송합니다.

2) 데이터 반출 거버넌스

정부기관(통계청, 행정안전부 등)과의 데이터 연계 시에도, IRB 승인, 연구계획서, 보안 서약서 등 엄격한 절차를 거치며, 개인식별정보는 제공하지 않습니다. 데이터 반출 시에는 반출기관의 성격(공공/민간/결합전문기관 등)에 따라 제출 서류와 보안 조치가 차등 적용됩니다.

9. 성과지표

1) 정량적 지표

표3 정량적 성과지표

구분	2021	2022	2023	2024
DRB 심의(건)	152	139	136	187
단독:다기관(비율)	8 : 2	5 : 5	6 : 4	7 : 3
심의(IRB부터 DRB, 의료자산까지) 평균 소요일	62	58	59	63

2) 운영 현황

표4 운영 현황 상세

구분	2021	2022	2023	2024
DRB 심의건	152	139	136	187
승인	78	53	38	72
시정 승인	41	78	77	103
- 클라우드 사용 조건			26	50
보류(재심의)	6	3	17	9

10. 도전과제와 한계

1) 심의 대기시간과 연구 속도의 균형

철저한 심의 과정은 데이터 보호에는 효과적이지만, 연구자들의 연구 진행 속도에는 제약이 될 수 있습니다. 특히 긴급성이 요구되는 공중보건 연구나 임상시험의 경우 신속한 데이터 접근이 필요한 상황이 발생합니다.

2) 과도한 비식별화의 리스크

개인정보 보호를 위한 과도한 비식별화는 데이터의 임상적 가치를 훼손할 수 있습니다. 특히 희귀 질환 연구나 정밀 의료 연구에서는 상세한 개인정보가 연구의 핵심일 수 있습니다.

3) 소규모 의료기관의 적용 가능성

연세의료원의 거버넌스 모델은 상당한 인적·물적 자원을 요구합니다. 소규모 의료기관이나 지역병원에서 동일한 수준의 시스템을 구축하고 운영하기에는 현실적 제약이 있습니다.

4) 확산 방안

① 단계별 도입 가이드 제공
② 클라우드 기반 공동 활용 플랫폼 구축
③ 지역 거점 병원 중심의 컨소시엄 모델
④ 정부 주도의 표준 플랫폼 제공

11. 시사점

1) 균형적 거버넌스의 중요성

데이터 활용과 개인정보 보호라는 상충 목표를 균형 있게 달성하는 거버넌스 모델이 필요합니다. 연세의료원의 사례는 이러한 균형이 실제로 달성할 수 있음을 보여줍니다.

2) 비식별화·가명화의 실질적 적용

단순한 기술적 비식별화가 아니라, 데이터 활용 목적·민감도에 따라 맞춤형 보안 조치와 심의가 병행되어야 합니다. 획일적 접근보다는 차등화된 전략이 더 효과적입니다.

3) 표준화와 품질관리의 선결 조건

데이터 결합·반출·다기관 연구의 활성화를 위해서는 표준화와 품질관리 체계가 필수적입니다. 이는 단순한 기술적 작업이 아니라 지속적인 거버넌스 활동이어야 합니다.

4) 심의 체계의 실효성 강화

DRB 등 심의 기구가 실제 데이터 활용의 위험·효익을 평가하고 조건부 승인·보류 등 실질적 통제권을 행사해야 합니다. 형식적 심의는 오히려 거버넌스의 신뢰성을 훼손할 수 있습니다.

5) 연구자 지원과 소통

심의 기준·절차의 투명성 확보와 연구자 대상 교육·지원이 병행되어야 데이터 활용 활성화와 개인정보 보호가 함께 달성될 수 있습니다. 규제가 아닌 지원의 관점에서 접근해야 합니다.

12. 미래 데이터 플랫폼의 로드맵

1) 환자 주도 동의(Patient-directed Consent) 플랫폼

환자가 직접 자신의 데이터 활용 범위와 조건을 설정할 수 있는 디지털 플랫폼을 구축합니다. 환자는 연구 분야별, 목적별로 세밀하게 동의 범위를 조절할 수 있으며, 자신의 데이터가 어떤 연구에 활용되고 있는지 실시간으로 확인할 수 있습니다.

2) 블록체인 기반 데이터 활용 이력 추적

블록체인 기술을 활용하여 데이터 활용의 전 과정을 투명하고 변조 불가능하게 기록하는 시스템을 구축합니다. 이를 통해 데이터 거버넌스의 신뢰성을 한층 더 강화하고, 규제기관과 환자에게 완전한 투명성을 제공할 수 있습니다.

3) 국가바이오빅데이터와의 안전한 연계 확대

국가바이오빅데이터, 한국인 유전체 정보, 국민건강정보 등과의 연계를 통해 통합 의료데이터 플랫폼을 구축합니다. 이를 위해 연합학습(Federated Learning) 기술을 도입하여 데이터를 직접 공유하지 않고도 통합 분석이 가능한 체계를 마련합니다.

4) AI 기반 합성 데이터 생성 체계

생성적 적대 신경망(Generative Adversarial Networks, GAN) 기반의 의료 합성 데이터 생성 플랫폼을 구축하여, 개인정보 보호와 데이터 활용성을 동시에 만족하는 솔루션을 제공합니다. 이를 통해 연구자들은 실제 환자 데이터의 통계적 특성을 보존한 안전한 합성 데이터를 활용할 수 있게 됩니다.

13. 결론

연세의료원 사례는 의료데이터 2차 활용의 전 과정에서 개인정보 보호와 데이터 활용의 균형, 데이터 품질 및 표준화, 안전한 반출 및 결합 체계 등 복합적 과제를 실질적으로 해결한 선진적 모델로 평가됩니다.

특히 법적 근거와 실무 운영의 완벽한 연결, IRB-DRB-의료자산활용위원회의 3단계 심의를 통한 실효적 거버넌스, 다층적 비식별화 전략을 통한 맞춤형 보안 조치, 그리고 Honest Broker 시스템을 통한 안전한 데이터 연계 등은 다른 의료기관이 참고할 수 있는 구체적인 모델을 제시합니다.

연세의료원의 경험에서 얻은 가장 중요한 교훈은 "완벽한 시스템보다는 실용적이고 지속 가능한 거버넌스"의 중요성입니다. 기술과 제도, 그리고 사람이 조화를 이루는 거버넌스만이 의료데이터의 잠재력을 안전하게 실현할 수 있습니다.

향후 의료데이터 활용이 더욱 확대될수록, 이러한 거버넌스와 실무적 경험이 국내외 정책 수립과 현장 적용에 중요한 참고점이 될 것입니다. 연세의료원은 앞으로도 혁신과 안전의 균형점을 찾아가며, 의료데이터 거버넌스의 새로운 표준을 제시해 나갈 것입니다.

의료데이터 거버넌스의 미래는 규제가 아닌 enablement, 제한이 아닌 responsible innovation에 있습니다. 연세의료원의 사례가 이러한 미래로 가는 길에 하나의 이정표가 되기를 희망합니다.

참고 문헌 |

1. 국가법령정보센터, 개인정보 보호법, 2025.04.01.
2. 국가법령정보센터, 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 2025.10.01.
3. 국가법령정보센터, 신용정보의 이용 및 보호에 관한 법률, 2024.02.13.
4. 보건복지부, 「보건의료데이터 활용 가이드라인」개정 안내, 2024.12.16.
5. 개인정보보호위원회, 가명정보 처리 가이드라인, 2024.02.
6. 한국인터넷진흥원, 데이터 3법 개정의 주요 내용과 전망, 2020 KISA REPORT 2월호_06, 2020.03.02.
7. 보건복지부, 비정형데이터 가명처리, 결합 데이터 제공 등 보건의료 데이터 활용 지원 강화, 2024.12.16.

EU의 GAIA-X와 Health Data Space
: EU 중심의 의료데이터 컴플라이언스 동향 및 이슈

김주한
서울대학교 의과대학 | 교수

1. 서론

의료 환경에서 데이터 기반 패러다임이 급부상하면서, 국가별·기관별 의료데이터의 "경계를 넘는 공동 활용"이 주요 정책 목표가 되었다. 한편 "연결"이 곧 "공유"를 의미하는 것은 아니며, 사람의 몸과 삶이 투영된 민감한 의료데이터는 유출 시 비가역적 피해를 줄 수 있음도 부각되었다.

유럽연합은 GDPR을 통해 최강의 개인정보보호 규범을 마련하여, 유럽 시민의 데이터를 누가, 어떤 목적으로, 어느 나라에서 활용할지 통제한다는 의지를 명확히 했다. 유럽보건데이터스페이스(European Health Data Space, EHDS)라는 의료데이터 특별 규범 추가로, EU는 "의료데이터를 잘 보호하면서도, 공익적 연구와 혁신을 위해 적극적으로 활용하겠다"는 이중 목표를 분명히 했다. 의료데이터는 기관의 벽을 넘지 않고, 분석 코드와 모델만 각 노드에 전송하여 연합·연계 분석을 수행하며, 전 과정을 투명하게 기록하는, 법과 기술, 신뢰와 거버넌스를 모두 요구하는 고난도 실험이지만, 향후 의료데이터 협력이 나아가야 할 방향을 선명히 제시한다. 원활한 국제 협력을 위해 EU는 GDPR, EHDS, Gaia-X, Dataspace4Health(DS4H) 구축으로 "데이터는 각국에 두고, 분석만 이동시키는 연합형 의료데이터 스페이스 모델"을 구축했다.

2025년 7월 공식 발표된 국제헬스데이터스페이스 이니셔티브(International Health Data Space Initiative, IHDSI)는 이 모델을 유럽 밖(한·미·아시아)으로 확장하는 국제 의료데이터 스페이스 프로젝트다. 룩셈부르크 국립보건원과 한국 국립암센터, 국내외 기술 기업들이 참여한 IHDSI는 "데이터는 각국에 두고, 지식과 모델만 연합한다"라는 철학 아래, 암·희귀질환·약물감시·정밀의료를 위한 국제 분산 데이터 인프라 구축을 선언했다. 한국은 "빅데이터를 보유하면서도, 함부로 내보낼 수는 없는" 국가로 IHDSI가 지향하는 연합형 IHDSI의 자연스러운 파트너다. IHDSI는 단순 연구망이 아니라, GDPR, EHDS, Gaia-X 규범을 준수하며 글로벌 AI와 RWE 정밀의학의 가능성을 시험하는 '리빙랩'이다.

본 보고서는 IHDSI의 기술·정책 구조, EU의 GAIA-X 기반 규범 변화, 그리고 한국이 어떤 전략으로 참여해야 할지 분석한다. 또한 한국이 이 국제 이니셔티브에 참여할 때 어떤 기회와 위험이 동시에 존재하는지, 한국형 K-Health Data Space는 어떤 원칙과 구조로 설계되어야 하는지에 관한 개인정보 컴플라이언스를 분석한다.

2. 글로벌 규범 환경 변화

1) GDPR: “데이터는 시민의 권리”라는 철학의 탄생

GDPR은 단순한 개인정보보호 규정을 넘어 시민의 삶과 정체성을 구성하는 개인정보가 단순한 ‘자원’이 아니라 ‘권리’의 일부라는 유럽 정치철학을 법률로 구현한 형태다. 건강정보는 GDPR에서 “특수(민감) 정보”로 분류되어 매우 높은 보호 수준이 제안되어 있다. GDPR에서 가장 중요한 것은 개인정보 국외이전(GDPR 제44조부터 제49조까지) 규정이다. 원칙적으로 EU 시민의 건강정보는 충분한 보호체계가 없는 제3국으로 이동할 수 없다. 이는 오늘날 국제의료 AI, 다국가 정밀의료 연구를 가장 어렵게 만드는 규범이지만, 유럽이 오래 지켜온 ‘정보인권’의 집약체이기도 하다.

2) EHDS: ‘보호’와 ‘활용’을 동시에 껴안으려는 유럽의 두 번째 실험

GDPR이 지나치게 엄격하다는 비판 속에서도 EHDS는 의료데이터는 더 적극 활용할 수 있도록 특별규정을 만드는 결정을 담았다. “의료데이터는 보호되어야 하지만, 동시에 공익을 위해 사용되어야 한다.” EHDS는 의료데이터의 1차 사용(진료, 환자 이동성, 상호운용성)과 2차 사용(연구, AI, 약물감시, 정책·규제, 실세계 근거(Real-World Evidence, RWE)) 모두를 포괄하는 최초의 국제 규제다. EHDS는 GDPR보다 더 적극적이고, 더 명확하게 의료데이터 활용 기회를 열어준다. IHDSI가 이를 국제적으로 확장하기 위해서는 바로 이 EHDS의 사상과 절차를 이해해야 한다.

3) Gaia-X & Dataspace4Health: “데이터를 모으지 않고 연계하는” 기술적 대안

Gaia-X는 유럽이 데이터 경제에서 독점적 클라우드 기업에 의존하지 않고, 유럽 스스로 통제하는 분산형 데이터 생태계를 만들겠다는 야심 찬 시도다. 핵심 철학은 명확하다. 데이터는 제공자가 통제한다. 데이터는 기계가 읽을 수 있는 형태(machine readable)로 선언된다. 데이터는 중앙에 모이지 않는다. 신뢰할 수 있는 노드만 연결된다. 모든 행위는 투명하게 기록된다.

DS4H는 이 Gaia-X 원칙을 의료에 맞춰 구현한 실험장이다. 룩셈부르크 국립보건원이 EU GAIA-X Health의 헤드쿼터로 주도하여, EU 내 24,000여 개의 병원 및 연구기관과 정부를 데이터 스페이스로 연계한다. IHDSI는 DS4H의 “국제 확장판”이다. Gaia-X 기반의 신뢰·정책·기술 프레임 위에 전 세계 병원과 기관을 얹어 하나의 국제 의료데이터 스페이스를 만드는 것이다.

4) 미국(N3C·TEFCA)와 아시아 비교: 서로 다른 길, 서로 다른 철학

미국은 “데이터 흐름을 막기보다 관리한다”라는 철학을 견지한다. 그 결과로 두 가지 특징적 모델이 등장했다

① 미국 국가 코로나19 코호트 협력체(National COVID Cohort Collaborative, N3C): 중앙집중형 대규모 실세계 자료(Real-World Data, RWD) 허브는 코로나19 시기 중앙분석 허

브로 운영됐다. 병원들이 데이터를 "올리는" 방식으로 유럽식 분산 모델과 정반대였으나 연구 생산성은 높았다.

② 신뢰 교환 프레임워크 및 공통 합의(Trusted Exchange Framework and Common Agreement, TEFCA): 상호운용성 기반 의료정보 교환인 TEFCA는 진료목적 데이터의 교환을 촉진하는 규범으로, 데이터 주권보다 효율성과 접근성을 우선해 설계되었다.

5) **아시아의 '데이터 스페이스' 개념**이 아직 완성되지 않았지만, 한국은 강력한 공공 RWD(건강보험심사평가원, 국민건강보험공단)와 한국의약품관리원의 약물감시체계 기반의 분산연구 인프라를 보유했다. 즉 한국은 데이터 이동을 EU처럼 엄격히 막지는 않지만, 의료데이터를 강력히 보호하는 법제(개인정보 보호법·의료법·생명윤리법)로 유럽과 미국의 중간 입장을 취한다. 이는 한국이 IHDSI로 유럽-미국-아시아를 연결하는 허브 노드가 될 잠재력을 갖음을 보여준다.

3. 헬스데이터 스페이스 개념의 진화

의료데이터는 병원 서버실에 쌓여 있는 "부산물"에서 환자의 생애주기 전체를 관리하며, 유전체·영상·생활 습관·약물 반응을 포괄하는 정밀의료 시대의 주역으로 위상이 달라졌다.

1) **1세대: 데이터 수집 (central repository): "모으는 것이 힘이다."** 하지만 이 방식은 곧 한계에 부딪혔다. 복잡한 의료데이터는 통합하기 어렵고, 민감하여 외부 이동이 어렵고, 책임소재와 보안 문제, 소유권 등이 복잡하게 얽혔다. 중앙집중은 강력했지만 취약점도 컸다.

2) **2세대: 데이터 공유 (network transfer): "필요시 전송한다"**는 의료정보교류(Health Information Exchange, HIE), 기관 간 데이터 전송, 연구용 데이터 세트 교환 등이다. 쉽고 빠르지만, 데이터가 이동 절차는 느리고 이동 자체에서 문제가 발생하며, 이동 효율에 한계가 있었다.

3) **3세대: 분산연구네트워크(Distributed Research Network, DRN): "각자 데이터를 갖고, 분산처리 하는"** 분산연구망은 공통데이터모델로 표준화한 데이터 소유자들에게 분석코드를 전송해, 그 실행 결과들만 취합하는 방식이다. 각 기관 데이터 주권을 유지하며, 다기관 분석을 가능하게 하였으나, 분석의 자동실행 기전이 없어, 실행이 매우 느리고 느슨한 체계라는 한계를 보인다. 접근통제, 정책 자동화, 국제규범 준거성, AI 학습, Traceability 등의 구현이 꼭 필요했다.

4) **4세대: 상호운용성(interoperability): "형식"** 표준화인 구문적 상호운용성(Syntactic Interoperability)은 어느 정도 성취했으나, 실제 데이터 처리에 필요한 의미적 상호운용성(Semantic Interoperability)까지는 아직 요원하다.

5) **5세대: 의료데이터 스페이스: "연결된 주권"** 데이터는 각 기관·국가의 통제 안에 머물며, 각 기관은 자신의 데이터에 대해 누가 접근할 수 있는지, 어떤 목적으로 사용할 수 있는지, 어떤 분석은 허용되고 금지되는지를 명시적으로 선언하며, 사람뿐 아니라, 기계가 해석 가능한 정책(policy-as-code)으로 적용된다. 연합학습(Federated Learning), 분산 분석(Federated Analytics), Secure Compute Node, Self-Description, Gaia-X 신뢰 프레임, End-to-End Audit 등이 핵심 요소다.

6) **6세대: IHDSI 이니셔티브:** IHDSI는 의료데이터 스페이스의 개념을 처음으로 "국제적 규모"로 구현한다. 유럽의 Gaia-X/DS4H 철학을 바탕으로, 한·미·아시아의 노드가 연결되고, 각국의 법제와 규범을 충돌 없이 조화시키려 한다. 데이터는 국경을 넘지 않지만 지식은 넘나들고 모델은 공유된다. 결과는 집계되고, 의학적 통찰은 확장된다. "서로 다른 법, 다른 나라, 다른 시스템을 가진 기관들이 하나의 연구, 하나의 모델을 함께 만들고자 하는 공동실험"이다.

4. IHDSI: 설립 배경, 참여 기관, 기술 아키텍처, 운영 철학

1) **룩셈부르크 국립보건원:** DS4H를 주도해 온 기관으로, EHDS 구현 경험을 가진 Gaia-X 기반 헬스데이터 거버넌스의 규범적 기준점이다.

2) **한국 국립암센터:** 세계적 수준의 암 빅데이터·임상·유전체 데이터 역량을 가진 기관으로, 국민건강보험 기반 RWD(건강보험심사평가원·국민건강보험공단)와 유기적 연계 가능한 아시아 정밀의료 거점이다.

3) **한국·유럽·미국의 클라우드·기술 기업:**
   ① 사이퍼롬(Cipherome): 실리콘밸리 기업으로 EHDS의 Data Connector와 AI Analytic Layer 담당
   ② 오케스트로(Orkestro): 클라우드 관리와 가상화 기업으로 IHDSI Federated Orchestration Layer를 담당
   ③ 네이버 클라우드(Naver Cloud): 클라우드 인프라 및 보안·분산환경 제공

4) **IHDSI의 기술 아키텍처는 "데이터를 움직이지 않는 분산 분석기술"로 다섯 층으로 구성된다.**
   ① **데이터 소스 계층(Data Sources Layer)**: 다국가 노드의 집합으로 EU 24,000개 병원을 포함한다.
   ② **로컬 연합분석 계층(Local Federated Compute Layer)**: 공통데이터모델(Common Data Model, CDM) 및, FHIR(Fast Healthcare Interoperability Resources) 전송표준 기반의 폐쇄형 작업 공간이 수행되는 연합분석 클라이언트로 데이터는 각 기관에 분산되어 있고, 분석(요청 및 코드)이 기관을 방문한다. 사이퍼롬 사가 담당한다.
   ③ **거버넌스 계층(Orchestration & Governance Layer)**: IHDSI가 GDPR, EHDS, Gaia-X를 준수

할 수 있도록 분석 코드와 AI 모델을 각 로컬 연합분석 노드로 배포하고, 각국의 법·정책 위반 여부 등을 자동 점검하며, 분석 실행의 과정·결과·접근 정보를 모두 기록한다. 다국가 정책을 "기계가 읽을 수 있는 규칙"으로 구현하며 오케스트로(Orkestro) 사가 담당한다.

④ **임상 지능 계층(Clinical Intelligence Layer)**: IHDSI가 단순히 데이터 인프라가 아닌 이유는 여기에 있다. 약물 반응 예측(Pharmacogenomics), 항암제 독성 위험 모델, 희귀질환 AI 멀티오믹스(Multi-omics) 기반 정밀의료, 등의 임상 지능 솔루션이 이 계층에서 개발·검증된다. 사이퍼롬 사가 담당한다.

⑤ **산출 계층(Outputs Layer)**: 국제 공동의학지식 생산: 국제적 실세계 근거(Real-World Evidence, RWE), 약물감시 신호, 암 치료 반응 모델, 규제과학 보고서, 정밀의료 AI 등 단순한 데이터 소통이 아니라 새로운 의학 지식이 생산 출력된다.

5. GAIA-X 기반 의료데이터 컴플라이언스

GAIA-X는 단순 IT 플랫폼이 아니라, 데이터를 바라보는 철학으로 기술보다도 그 뒤의 규범적 구조, 거버넌스적 긴장, 철학적 고민을 그대로 드러낸다. 데이터를 안전하게 보호하면서도 기관 간 국가 간 협력을 가능하게 하기 위한 오래된 질문을 현대적 방식으로 다시 묻는 작업이다.

1) 법·정책 요구사항

① 첫째, 데이터 주권(Data Sovereignty)
② 둘째, GDPR의 원칙과 국외이전 금지
③ 셋째, EHDS의 2차 사용 (Secondary Use) 조건
④ 넷째, 정책의 기계 가독성 (Policy-as-Code) 조건

2) 기술 요구사항: "데이터는 움직이지 않고, 대신 신뢰가 움직인다"

① Secure Compute Node: 내부 데이터 보관 및 외부 분석코드를 실행하는 워크스페이스.
② Federated Learning / Analytics: 신뢰 실행 환경(Trusted Execution Environment, TEE)
③ Self-Description: 참여기관은 자신의 데이터의 유형, 품질, 사용목적 등을 메타데이터로 스스로 서술해야 한다. 메타데이터는 전체 데이터 스페이스를 서술하는 "언어"로 역할한다.
④ Zero-Trust IAM: 신뢰기반 접근 아님, "아무도 믿지 않음" "매 접근마다 인증·검증·감사"
⑤ End-to-End Audit Trail

3) 거버넌스 요구사항: 기술만으로는 신뢰가 확보되지 않으므로 제도·윤리·거버넌스가 기술보다 더 중요한 경우가 많다.

① 다층적 거버넌스: 정책 레벨 (GDPR, EHDS, 국가법), 기관 레벨 (병원·공공기관 데이터

위원회), 프로젝트 레벨 (기관생명윤리위원회(Institutional Review Board, IRB)·개인정보 영향평가(Data Protection Impact Assessment, DPIA), 기술 레벨 (Policy-as-Code 집행), 운영 레벨 (사이퍼롬, 오케스트로 같은 오케스트레이션 시스템)의 다섯 층이 서로 충돌하지 않고 정렬되어야 한다.

② 환자 신뢰: 데이터스페이스는 환자 정보주권과 투명성, 통제권, 접근권 보장을 요한다.

③ 기관 간 권한 배분: 거버넌스 분쟁은 의료데이터 협력의 가장 큰 장애 요인 중 하나다.

④ 계약과 법적 책임: 데이터 손상, 재식별, 알고리즘 오류 등의 발생 시 GAIA-X는 이러한 책임을 계약 기반으로 명확히 정의할 것을 요구한다.

4) 난점: "이론적으로는 이론과 실제가 같아야 하지만, 실제로는 이론과 실제가 같지 않다."

① 첫째, 규범 간 충돌: GDPR vs. EHDS, 국가 간 해석 차이, EU, 한, 미 법제 철학 차이

② 둘째, 정책·법률의 "코드화" 문제: Policy-as-Code는 미래 지향적이지만, 구현은 어렵다.

③ 셋째, 재식별 가능성: 데이터 결합 등을 통한 재식별화의 완전 차단은 불가능하다.

④ 넷째, 의료기관의 준비 부족, 자원 부족

⑤ 다섯째, 기술·운영 인력 부족: 한 기관이 요구되는 다양한 전문성을 다 충족하기 어렵다.

6. IHDSI 기반 의료데이터 컴플라이언스

IHDSI의 아키텍처는 '정치적·윤리적 질서'를 담은 5계층이다. Federated Node Layer (환자와 기관에 대한 존중), Orkestro Layer (국가 간 조율), Cipherome Layer (의학적 가치의 생산), Data Sovereignty Layer (데이터 주권), Audit Layer (신뢰의 확보)가 데이터는 움직이지 않지만, 의학 지식은 더 멀리, 더 넓게, 더 공정하게 확산되는 것을 지원한다.

IHDSI에서 가장 중요한 문제는 한국의 의료데이터 규범이 유럽의 규범과 어떻게 만나는가이다. 두 체제의 공통점인 "강력한 데이터 보호, 민감정보 중시" 법 제도는 결합가능성을 높인다. 가장 큰 차이점은 국외이전 규제의 강도다. EU GDPR은 제3국으로의 데이터 이동을 거의 금지하다시피 엄격하지만, 한국은 조건과 절차를 갖추면 데이터 해외 이전이 가능하다. IHDSI는 이 차이를 연합 학습과 연계분석으로 해결한다. 데이터가 이동하지 않기 때문에 한국과 EU 모두의 법제를 동시에 만족할 수 있다.

EHDS의 "2차 사용 규범"은 의료데이터의 2차 활용(연구·정책·AI·약물감시)을 공식적으로 인정한 최초의 국제규범인 반면 한국에서는 연구 목적은 생명윤리법, 정책 목적은 공공기관 운영, AI 목적은 개별기관 결정, 약물감시는 식약처 규정처럼 여러 법률에 흩어져 있다. IHDSI는 한국이 "보건데이터 2차 사용의 공통 기준"이 무엇인지 다시 고민하게 만드는 중요한 계기다.

EU는 데이터 스페이스라는 “분산형 규범” 모델을 택했고 미국은 N3C 모델을 택했다. 한국은 그 중간에 있다; 건보공단과 심평원은 중앙집중형, 병원 CDM과 의약품안전정보원은 분산형, 법제 구조는 EU식 “민감정보 보호형”이라는 독특한 위치는 한국이 IHDSI에서 단순 참여자가 아니라 “유럽과 미국을 잇는 구조적 중간 허브 노드가 될 수 있음”을 뜻한다.

7. 정책제언 – “보호와 활용 사이에서 한국이 선택해야 할 미래”

한국이 IHDSI와 같은 국제 협력에 참여하는 순간, 우리는 단지 새로운 기술을 도입하는 것이 아니라, 데이터와 의료를 바라보는 국가적 관점을 새롭게 정립하는 과정에 들어서게 된다.

1) “데이터는 환자의 것이다. 그러나 지식은 인류의 것이다”라는 원칙

한국의 개인정보 보호법은 강력하고, 의료법은 엄격하며, 생명윤리법은 보수적이다. 이는 환자 보호에는 유리하지만, 데이터 기반 연구·AI·정밀의료에는 높은 장벽이다. IHDSI가 보여준 미래는 “데이터를 내보내지 않고도 지식을 공유할 수 있다”는 새 모델을 받아들여야 한다.

2) 규범의 현대화를 통해 ‘데이터 스페이스 시대에 맞는 한국형 법제’를 준비해야 한다.

기술이 법을 초월하는 속도로 진화하고 있다. 필요한 것은 단순한 규제 완화가 아니라 의료데이터 2차 활용에 대한 단일 프레임의 도입이다.

3) 공공데이터(HIRA·NHIS)와 의료기관 데이터를 ‘서로 이어지는 두 다리’로 바라봐야 한다.

K-Health Data Space에서 중앙집중형 공공 데이터는 전 국민 규모의 질병 패턴을 보여주고, 분산형 의료기관 데이터는 질병의 깊이와 정밀도를 제공하며, 두 데이터는 서로 이동 없이 조율된다. 정책은 이 두 데이터 강을 하나의 생태계로 엮기 위한 제도적 다리 역할을 해야 한다.

4) 기술을 도입하는 능력보다 ‘기술을 해석하는 능력’을 키워야 한다.

IHDSI가 보여주는 기술적 구조—연합학습, 정책 자동화, GAIA-X 신뢰 프레임—는 매우 고도화되어 있다. 정책 전문가, 기술학 전문가, 윤리 전문가, 의료 정보학자의 네 전문가가 함께 논의하는 문화를 정책적으로 만들어야 한다.

5) 의료기관·공공기관·정부가 새로운 역할을 받아들여야 한다.

IHDSI는 각 기관이 기존의 역할을 넘어서도록 요구한다. “내 데이터만 보호하면 되는” 폐쇄 역할을 넘어 국제 데이터 생태계의 구성원으로서 공동 책임을 지는 역할 정립이 필요하다.

6) 국민 신뢰를 지키는 것은 기술이 아니라 투명성이다.

모든 단계에서 국민을 협력의 ‘객체’가 아닌 ‘참여자’로 인정하는 정책적 접근이 필요하다. 신뢰는 설명 가능성에서 생기고, 설명 가능성은 투명한 정책에서 온다.

7) 한국은 이제 '데이터 수출국'이 아니라 '의학 지식의 수출국'이 되어야 한다.

IHDSI는 한국의 가능성을 실현할 기회다. 한국이 국제규범을 지키면서 국제연구의 중심이 되는 환경이 필요하다. 정부의 역할은 데이터의 이동통제가 아닌 지식 이동의 촉진이다.

미래는 규제할 수 없다. 정책은 미래를 가능하게 만드는 수단이다.

참고 문헌 |

1. 세계법제정보센터, European Health Data Space (EHDS) 규정 (Regulation (EU) 2025/327), 2025.03.19.
2. Gaia-X Association, Gaia-X Architecture Document, 2023.10.
3. Arthur Kari & Tim Schurig & Martin Gersch, European Health Data Space (EHDS), Gaia-X and Health-X dataLOFT, 2023.11.
4. 전자신문, 글로벌 의료 AI 데이터 연합 출범…“임상 데이터 공유·공동 연구로 정밀의료 가속화”, 2025.07.22.
5. Dataspace4Health, Luxembourg launches Dataspace 4 Health: A Pioneering Dataspace and Governance Framework for Secure and Compliant Health Data Exchange, 2025.03.28.

PRIVACY REPORT

개인정보 이슈 심층분석

「2025 개인정보 이슈 심층분석 보고서」는
개인정보보호위원회의 출연금으로 수행한 사업의 결과물입니다.

본 보고서의 내용은
한국인터넷진흥원의 공식 견해가 아님을 알려드립니다.

발행일 2025년 12월
발행처 한국인터넷진흥원 개인정보제도팀
전라남도 나주시 진흥길 9

KISA 한국인터넷진흥원