

공 개

보건의료 빅데이터 플랫폼 시범사업 추진계획(안)

2017. 12.

보건복지부

- ◇ 본 문서는 자유롭게 전달, 게시할 수 있는 공개용입니다.
- ◇ 공개적인 의견수렴을 위해 작성된 추진계획 초안이며, 의견 반영을 거쳐 최종안으로 확정될 예정입니다.
- ◇ 본 추진계획에 대해 의견이 있으신 분은 아래 세 가지 방법 중 한 가지를 통해 의견을 접수하여 주시기 바랍니다.
 - 국민신문고 웹사이트(epeople.go.kr) > 민원·제안·참여 > 정책토론 > “보건의료 빅데이터 플랫폼 시범사업 계획(안)” 검색 > 댓글달기
 - * 국민신문고 웹사이트 로그인이 필요함
 - 의견 제출 웹사이트 (<http://bit.ly/2BUY6ZH>)
 - 이메일 (bigdatavoice2018@gmail.com)

목 차

1. 배 경	1
2. 기본 방향	2
3. 사업 내용	3
4. [가칭] 보건의료 빅데이터 특별법 추진 ...	7
5. 세부 추진 일정	8
붙임 1. 보건의료 빅데이터 정보보호 체계 개요	9
붙임 2. 주민등록번호 일방 암호화 개념	12
붙임 3. 보건의료 빅데이터의 공공적 활용 사례 (예시)	13
붙임 4. 주요 보건의료 빅데이터 데이터셋 사례 (예시)	16
붙임 5. 보건의료 빅데이터 플랫폼 예산 내역	17
붙임 6. 보건의료 빅데이터 추진경과	18
붙임 7. 보건의료 빅데이터 해외사례	19
붙임 8. 보건의료 빅데이터 플랫폼 주요 기능(안)	20

【 2018년 사업예산 개요 】

- ◇ (보건의료 빅데이터 플랫폼, 46억) 공공기관이 보유한 보건의료 정보를 연계하여 연구 등에 활용할 수 있는 기반 구축
⇒ 국회 예산심의 결과 77억→46억으로 감액(△31억)
- ◇ (기관 간 분석자료 공유·활용 네트워크, 24억) 의료기관·공공기관이 보유한 데이터의 분석결과를 취합*할 수 있는 기반 구축
* 개인정보 제공 없이 분석결과만을 취합, 연구에 활용
- ◇ (보건의료 빅데이터 연계·활용 강화 연구, 19억) 실제 플랫폼을 활용하여 데이터셋 및 공공서비스 개발을 위한 R&D 지원

□ 보건의료 빅데이터의 중요성

- 공공기관 내 보건의료 빅데이터를 기반으로, 건강상태 및 의료이용 흐름을 파악하여 정책개선, 의학연구의 재료로 활용
- 각 기관에 분산된 다양한 종류의 데이터 간 연계를 통해 기관별 데이터로는 할 수 없었던 다양한 종류의 연구가 가능해짐
 - * 예1) 건강검진(건강보험공단) ↔ 진료내역(심평원) ↔ 암등록(암센터)
⇒ 건강검진 시 발생한 의심증세와 진료내역 간 상관관계, 암종류 및 예후까지 연관분석
 - * 예2) 약 처방내역(심평원) ↔ 진료내역(심평원) ↔ 건강검진(건강보험공단)
⇒ 처방약별로 이상 진료패턴 및 건강검진 상 이상패턴 자동발견 → 부작용 조기감지
- 보건의료 전 분야에서의 근거기반 정책개선 및 의학연구 가능
 - 치료법·약제·의료기기 간 효과 비교, 제네릭 약제의 동등성 분석, 감염병 유행 조기탐지, 의료기관 종별 진료양상 분석 및 정책개선 등
 - 국민건강 증진에 필수적인 효율적·효과적 의료체계 구축에 기여

□ 보건의료 빅데이터 3대 추진원칙

- ① 보건의료 빅데이터는 공공적 목적으로 활용
- ② 시민참여·전문성에 기반한 논의구조 구축
- ③ 현행 법령에 근거하여 개인정보를 철저히 보호

□ 시범사업 운영 우선 추진

- 보건의료 빅데이터 사업의 안전성, 효용, 추진방향·내용 등에 대한 사회적 대화가 진행 중인 만큼, '18년에는 “시범사업”으로 추진
 - 보건의료 빅데이터의 사용 목적·대상 데이터·이용자 등을 제한
 - 시범적 R&D 연구 지원을 통해 플랫폼의 활용방안 검토 및 성과 평가

【 빅데이터 이용목적 및 이용자 등 】

- ◇ (목적) ① 보건의료 분야 정책연구 ② 의료정보보호 기술 연구
③ 보건의료기술 연구 ④ 건강 관련 학술연구 등 4개 분야
- ◇ (이용자) 중앙행정기관·지방자치단체, 공공기관·지방공공기관, 국내 의료기관·학계·연구기관(기관생명윤리위원회 승인필)
- ◇ (절차) 자료이용 신청 접수 → 이용목적 등을 살펴 자료 제공여부 건별 심의

- 시설투자는 우선 필요 최소한도로 구축하고, 향후 빅데이터 활용 방안·목적 등에 대한 논의와 함께 단계적으로 필요 장비 등 구축 확대
 - 현재 시급하지 않은 시스템 과부하 예방 장비 등은 제외(이중화 구조, 백업장비, 사용량통제 S/W 등)

3

사업 내용

□ 사업 개요

- (사업 단계) 준비단계('17) - 시범사업('18~'20) - 본사업('20~)으로 추진
 - 시범사업 종료 전 평가 및 본사업 여부를 결정하되, 위원회를 통해서 의견수렴 및 평가결과를 공개
- (추진전략) 향후 5개년에 걸친 보건의료 빅데이터 청사진 수립
- (거버넌스) 정책 전반에 걸쳐 토론, 심의 및 전문가 자문을 위한 협의체(위원회) 구성·운영
- (플랫폼) 빅데이터 활용을 위한 정보시스템은 단계적으로 확충
- (입법 추진) 보건의료 빅데이터의 활용 및 개인정보 보호를 위한 법적 근거를 강화 (특별법 제정 등 추진 검토)
- (R&D) 공공적 목적의 의료기술·정책·정보보호 기술 등 연구를 위한 R&D 지원 병행

< 단계별 사업 추진(안) 개요 >

연도	2017년	2018년	2019년	2020년	2021년	2022년
단계	준비단계	시범 사업 단계			본 사업 단계	
전략	전략 수립		전략 추진			
거버넌스	위원회 준비		위원회 운영			
플랫폼	구축준비	정보화 계획	플랫폼 구축		본사업 전환, 이용자 집중 등 필요 시 시스템 확충	
			1년차	2년차		
사업	시범사업 기획	시범사업			본사업 (위원회 심의를 거쳐 추진)	
		착수, 체계수립		평가		
법	특별법안 마련		특별법 제정 추진			
R&D	보건의료 빅데이터 R&D 지원					

□ 활용 방안

○ 공공적 활용 원칙 아래, 4개 분야 연구로 한정

* 구체적 범주는 위원회 논의를 통해 결정

① 보건의료 분야 정책연구

- 보건의료 분야 정책 기획, 평가, 개선 등을 위한 연구
- 보건의료 전달체계 이용현황, 지역별 건강현황 등 보건의료 분야 정책환경을 살피기 위한 연구

② 의료정보보호 기술 연구

- 보건의료 빅데이터 정보보호 처리 기술의 발전을 위한 연구

③ 보건의료기술* 연구

* 보건의료기술진흥법 제2조 및 동법 시행령 제2조를 준용하여 연구범위 검토

- 법령상 정의와 같이 의과학, 치의학, 한의학, 의료공학·의료기기, 의료정보학, 의약품 분야 연구 중 **영리적*** 연구 제외

* 시장분석, 개발된 제품의 마케팅 등 의학적 진보와 무관한 연구

④ 건강 관련 학술연구

- 위의 분류에 속하지 않는 연구 중 건강과의 관계가 명확한 학술연구

○ 분야별 구체적 연구의 유형, 활용 가능한 데이터 등은 위원회 심의를 통해 결정, 보건복지부 고시를 통해 규범화

* 연구 유형 및 활용가능 데이터 예시 - 붙임 3, 4 참고

○ 보건복지부 R&D 사업('18년 19억)으로 공공적 목적의 의료기술 연구 및 정책연구 지원, 향후 사업 안정화에 따라 R&D 확대 추진

□ 연계 데이터

- 4개 공공기관(건보, 심평원, 질본, 국립암센터) 보유 데이터를 연계·활용
 - 건강과 연관된 전체 데이터 중에서, **활용·분석의 안전성이 검토된** 데이터에 한해서 연계·개방을 허용
 - 위원회 심의를 거쳐 **활용 가능 데이터**를 선정, 고시를 통해 규범화

< 각 공공기관 별 개방가능한 데이터 예시 - 붙임 4 참고 >

- ◇ (건강보험공단) 가입정보, 건강검진정보, 영유아·암검진, 요양기관, 장기요양 판정, 장기요양기관 등
- ◇ (건강보험심사평가원) 병원정보, 청구내역, 수가DB, 의약품 처방정보 등
- ◇ (국립암센터) 암등록정보, 암검진 코호트, 암환자 의료비지원 정보
- ◇ (질병관리본부) 유전체 정보, 각종 건강조사 정보

□ 개인정보 보호 조치

① 비식별화 조치

- 개인을 특정할 수 없도록, 주민번호·생년월일 등은 모두 삭제
- 특이한 병명이나 희귀한 질환 기록 등은 삭제하거나, 대분류만 기재하는 방식으로 특정인을 알아볼 수 없도록 처리
- 연구자에게 제공될 데이터를 사전에 분석, 어떤 사람의 데이터인지 알아낼 가능성을 사전에 평가하고, 위험이 클 경우 제공 거부
- * 위험평가, 제공여부 결정 등은 자문위 산하 데이터제공심의위원회에서 수행

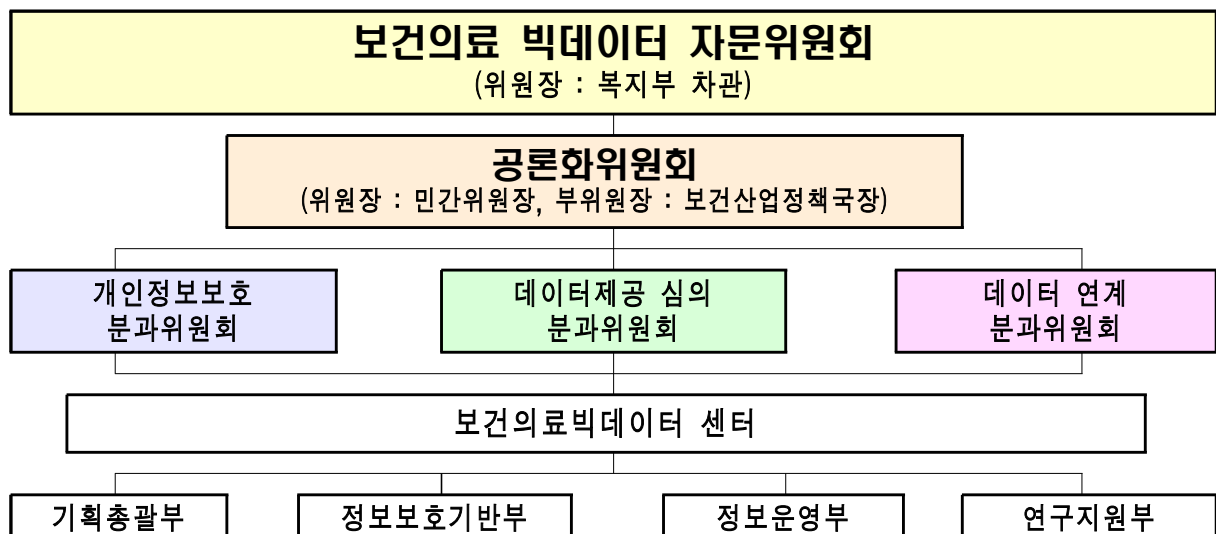
② 정보보안 기술 적용

- 인터넷에서 분리된 별도의 망(행정망)을 활용함으로써, 악의적 외부 해커로부터의 공격 등에서 원천적으로 방어 및 보호
- 정부통합전산센터에 서버를 위치함으로써 국가 중요 기간전산 시스템과 함께 방화벽 등 대규모 보안장비를 공유, 안전성 담보 가능
- 서버 간 통신, 자료제공 시 등 전송구간은 모두 암호화함으로써, 데이터 가로채기 공격을 예방하고, 위변조 공격 등에 대응
- * 행정안전부 정보화 검토, 국정원 보안검증, 개인정보영향평가 추진 예정

□ 거버넌스

- 보건의료 빅데이터 활용방안 및 정책방향 전반에 관해 “참여”와 “투명성”에 기반한 거버넌스를 구축·운영
 - (참여) 학계·의료계·전문가·시민사회 등이 참여하는 거버넌스를 구성하고, 일반 국민이 의견을 개진할 수 있는 통로 마련
 - (투명성) 빅데이터 정책 추진, 플랫폼 구축 및 제공 서비스 등에 대해 투명하게 공개하고 거버넌스를 통한 모니터링체계 확립
- 거버넌스 구축방안 (안)
 - (자문위원회) 보건의료 빅데이터 관련 최상위 논의기구로서 각계 대표로 구성*, 플랫폼 구축·빅데이터 활용 등 주요사항 심의
 - * 복지부 차관(위원장), 학계·의료계·전문가·시민사회·관계부처 등 20인 내외
 - (공론화위원회) 보건의료 빅데이터와 관련된 다양한 쟁점과 의견들을 수집·토의·정제하여, 자문위 심의활동 지원 및 대국민 소통을 담당
 - * 민간위원장 선임, 보건산업정책국장(부위원장) 및 각 분야 대표 30인 내외로 구성
 - (분과위원회) 데이터연계(연계기관간 협의), 데이터제공(제공 목적·내용 등 심의), 개인정보보호 3개 분야에 대해 분과위원회를 구성

< 보건의료 빅데이터 거버넌스(안) >



□ [가칭] 보건의료 빅데이터 특별법(안) 추진 검토

- 보건의료 속성, 데이터 활용 의학연구의 중요성 등 고려한 보건의료 분야 정보를 목적으로 하는 특별법 제정 추진
 - 현행 개인정보보호법은 “건강정보”를 하나의 분류로 규정, 종류별 처리 방법 및 안전수칙, 허용범위 등 세분화가 어려운 점 존재
 - * 예 : 키·몸무게 / 진료기록 / 유전자 등이 구별되지 않고 “건강정보”로 통칭
- 향후 법안 마련 시 중요 내용 및 방향성 등에 대해서는 위원회 논의를 반드시 거쳐 추진
 - * 법(안) 논의·검토가 광범위하고 장기적인 과제인 점을 고려, 공론화위원회를 적극 활용
 - 안전한 개인정보 연계·활용방법을 구체적으로 규정하고, 오남용에 대한 제재조치도 보다 강화할 필요

□ 특별법(안) 검토 필요사항

- 정보소유자(본인)의 권리 보호
 - 보건의료 정보의 원천적 권리 소유자인 개인을 보호하고, 오용·악용에 대비하여 사전적·사후적 조치를 할 수 있는 방안
- 적절한 보안조치, 악의적 이용에 대한 대책
 - 공공기관 보유 정보의 연계를 위한 방법, 활용 단계의 비식별조치 등
 - 비식별화된 정보에 대해 재식별화 시도, 성공, 자료유출 등 잘못된 이용에 대한 처벌조항 등 법적 안전장치 마련
- 활발한 정보연계 및 보호를 위한 정책 강화방안
 - 보건의료 빅데이터 정책위원회*, 기본계획 수립, 국가투자·지원 등 규정
 - * '17년 말~'18년 초 설치될 보건의료 빅데이터 자문위원회를 승계
 - 보건의료정보의 민감성을 고려하여, 전문가 및 시민사회단체 등으로 구성된 데이터 제공 심의 분과위원회 구성
- 보건의료 빅데이터 보유 기관 간 연계 시 정보보호의무 등

○ 2017년 4분기

- 시민사회단체 간담회 및 공청회 개최
- 보건의료 빅데이터 시범사업(안) 기획
- 자문위원회, 공론화 위원회 등 거버넌스 구성계획(안) 마련
- 보건의료 빅데이터 특별법안 마련 착수

○ 2018년 1분기

- 보건의료 빅데이터 자문위원회 등 거버넌스 구성
 - * 자문위원회에서 추진전략(안), 플랫폼 활용방향, 데이터·이용자 범주 등 논의
- 보건의료 빅데이터 추진전략 발표
- 정보시스템 구축사업 발주 협의 (조달청, 행안부, 국정원 등)

○ 2018년 2분기

- 보건의료 빅데이터 R&D 과제 착수
- 플랫폼 운영지침, 데이터제공계약서, 약관 등 개발 착수
- 보건의료 빅데이터 특별법 법안 논의

○ 2018년 3, 4분기

- 정보시스템 구축사업 착수, 수행

○ 2019년 상반기

- 플랫폼 운영지침, 데이터제공계약서, 약관 등 최종확정
- 플랫폼 1차 오픈 (비식별화, 기관 간 정보연계 기능 위주)

○ 2019년 하반기~2020년

- 플랫폼 2차 사업 추진, 2차 오픈 ('20)

< 정보보호 대책 수립 경과 >

- ☐ 보건의료 빅데이터 추진단 구성('17.3) 시 정보보호 분야 중요성을 인식, 정보보호·기술 분과를 별도 설치하고 정보보호·암호화 분야 전문가로 구성
 - (위원장) 고려대학교 정보보호대학원 한근희 교수
 - (위원) 서울대학교 법전원 고학수 교수, 경북대 컴퓨터학부 김일곤 교수, 이지서티 이재훈 연구소장, 삼성서울병원 이병기 교수, 한국 CPO포럼 심상현 사무국장, 아주대학교 곽진 교수, 한라대학교 김순식 교수, 고려대학교 이인혜 연구원, 신시웨이 김광열 상무 등
- ☐ 서비스·데이터 분과 등에서 제안된 시스템 기능 및 구성도를 검토하고, 각 시스템별 보안요구사항 및 보안조치·대책을 규정
 - 기술적으로 적절하지 않은 시스템 기능·구성은 대안을 제시하여 위원회 검토를 거쳐 최종 확정
- ☐ 시민단체 간담회 등 의견수렴 시 정보보호 대책에 대해 중점 논의 추진

< 개인정보 유출 방지 >

☐ 개인정보 비식별화 등을 통한 정보보호 강화

- 보건의료 빅데이터 제공 기관별 데이터 연계를 위한 비식별 대상 개인건강정보 표준 마련
- 보건의료 빅데이터 비식별 조치 기준 및 가이드라인 마련
- 각 공공기관에서 보건의료 빅데이터 공유 플랫폼으로 송부하거나, 최종 이용자 제공 시점에 개인정보 비식별화 실시
- 데이터 제공심의위원회를 통해 비식별화 평가 수준을 검토하고 (k-익명성 등) 위원회 심의 결과에 따라 이용자에게 제공
- 데이터 제공 방법, 이용자 군, 법적근거 등에 따라 상이한 데이터 비식별화 기준 적용

□ 주민번호 일방 암호화(붙임 2 참고)를 통한 안전한 결합키 체제 활용

- 주민번호 또는 생년월일 등 주민번호의 일부 자리 및 기타 정보를 일방 암호화*함으로써, 두 개 이상의 원천기관의 자료 결합에 활용
 - 단순히 주민번호만을 암호화하는 기존 방식과 달리, 무작위 숫자(이하 암호키)와 함께 암호화*하는 방식을 채택

* 암호키를 이용하는 일방 암호화 방식. 美 연방정부 공무원 패스워드 저장 방식

□ 정보유출·재식별화 방지를 위한 보안관제 시스템 구축

- 일정 수준 이상의 자유도를 갖는 자료 활용 시 클라우드 등 폐쇄적 환경을 활용하여 데이터에 접근, 유출·재식별 등 방지
 - 조회 프로그램 작성 및 실행 기록, 결과 등을 중앙화 하여 관리하는 등 자료 활용에 있어서 책임성을 강조하는 기술적 조치 이행
- 개인정보 열람, 조회기록 관제 시스템을 통해 특정인 정보 조회 등 확률 모델링 기반 위험 사례 자동탐지 기능 도입*

* 보건복지부 개인정보통합관제센터와 협조

□ 분양 데이터 유출경로 탐지를 위한 기술 도입

- 제공 데이터 중 일부에 대해 미세한 무작위 가감을 가하여, 전체적인 통계 영향은 없되, 제3자 제공 경로를 파악

* 예 : 키 180.3 → 180.2, 몸무게 75.2 → 75.3

< 기술적 정보보안체계 구축 >

□ 공개키 기반 전송구간 암호화, 망분리 등 보안체계 완비

- 공개키 기반 전송구간 암호화(PKI*)를 통한 암호키, 데이터 전송 보호

* Public Key Infrastructure : 데이터 전송 시 별도 암호화 인증기관을 활용, 가로채기 공격 등 각종 보안공격으로부터 보호하는 암호화 방식.

행정망 내에서는 행정안전부에서 암호화 인증기관 역할을 수행(G-SSL)

- 발송기관(예: 건강보험공단), 수신기관(공유 플랫폼) 양자만 내용을 알 수 있도록 PKI 구조를 활용하여 데이터 전송 암호화
- 행정망을 사용하여, 인터넷망과 물리적으로 분리된 보안체계 완비
 - DDoS 등 외부자의 공격을 원천적으로 차단하고, 망연계 서버를 통한 외부로부터의 접속은 별도의 보안조치를 통해 2중 관리

□ 보건의료 정보보호 기반기술 연구개발 추진

- 블록체인, 정보보호 방법, 안전한 자료교류 등 향후 기술발전에 대응하기 위한 정보보호 기반기술 개발 및 적용연구 추진

< 제도적 정보관리체계 강화 >

□ 오용 방지를 위한 제도적 장치 보완

- 이용자에게 데이터 제공 시 정보제공 계약을 통해, 정보제공 시 가장 큰 위협인 제3자 유출, 재식별화 등을 금지
 - 당사자의 제3자 유출, 재식별화 시도 등을 금지하는 표준 정보 제공 계약서를 개발하고, 각 제공채널별로 적용
 - 사고 발생 시 원인 이용자 및 소속기관 이용제한 페널티 부여
- 데이터 분양 신청서 제출 시 기관 정보보호 인증 또는 데이터 관리체계 서술, 기관별 책임성 있는 사후관리체계 보완 추진

□ 보건의료 빅데이터시스템 정보보호 관리체계 연구

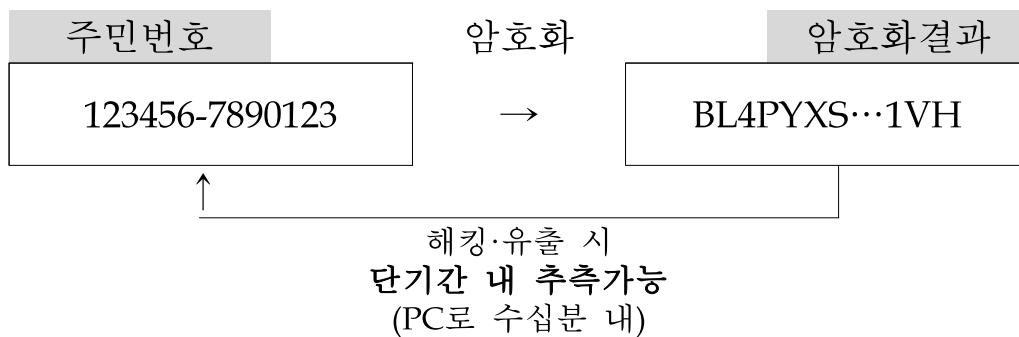
- 보건의료 빅데이터 보유·활용 기관을 위한 정보보호 프레임워크 마련
 - 보건의료정보 및 빅데이터 유통 환경에 적합한 위험관리 모델 개발
 - * 관련 국제표준(ISO 31000(위험관리), ISO 27005(정보보호 위험관리) 참조
- 보건의료 빅데이터 시스템을 위한 클라우드 센터 보안적합성 검증

- 주민번호만을 암호화하는 기존 방식과 달리, 무작위 숫자와 함께 암호화*하는 방식을 통해 해킹·유출 시에도 주민번호 추측 불가능

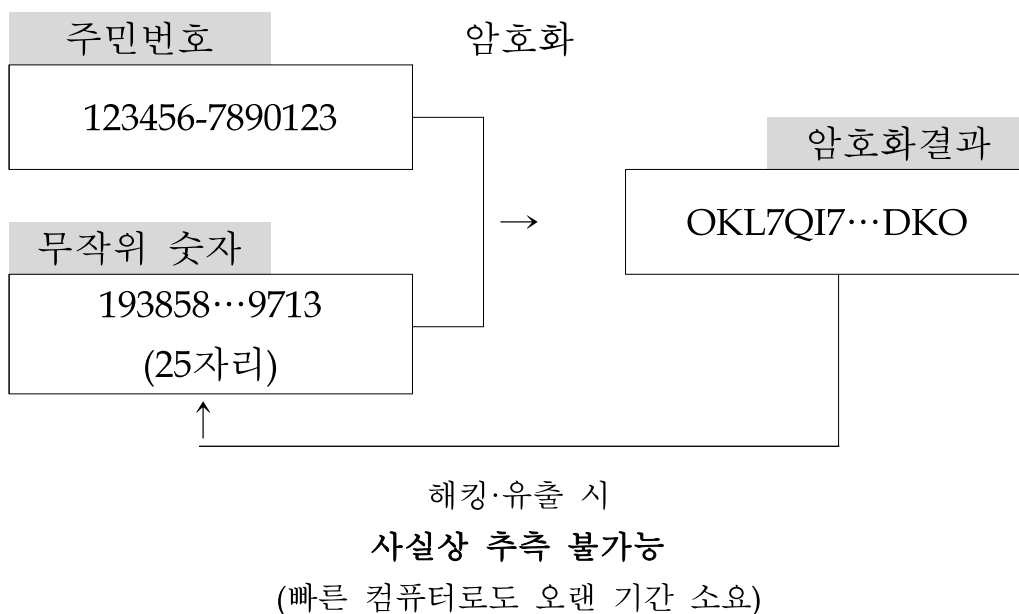
* Salt + Hash 방식, bcrypt, scrypt 등 활용

- 기존 방식에 비해 추측에 소요되는 시간이 현저히 길어져, 암호화된 주민번호가 유출되어도 사실상 해독이 불가능

기존방식



개편방식



* 주요 분야별 대표적 1~2가지 유형 연구 소개

① 취약계층 건강권 관련 연구

- 여성·장애인·아동·노인 등 사회적 취약계층의 건강불평등 연구
 - 취약계층이 많이 앓고 있는 질환을 데이터로부터 도출하고, 환경 및 사회적 요인과의 상관관계를 살펴 대응책을 도출
 - 특히 빈도가 비 취약계층과 유사하더라도, 완치에 소요되는 기간 또는 그 비율 상 차이 등 데이터기반 건강격차 연구
- 사회적 취약계층의 질병악화 경로 연구
 - 사회적 취약계층에서 두드러지는 질병악화 양상 및 이유(적은 치료기회, 낮은 의료 접근성 등)를 분석, 제도적인 보완책 마련
 - 질병악화에 따라 치료비용이 급증하는 등 가능한 시나리오를 도출, 의료진·사회복지사 등으로 하여금 대비할 수 있도록 정보 제공

② 사회적 건강위험 관련 연구

- 지역별 건강 유해요인 도출
 - 환경, 교통, 주거, 산업, 식수관리 등 다양한 지역별 요인과 건강상태 간의 비교를 통해 지역별 건강 유해요인 도출
 - 특히 유해한 사례의 경우 행정조사 등을 통해 구체화할 수 있도록 후보군을 도출, 전국단위 건강유해요소 스크리닝 연구
- 직역별, 직장별 건강 유해요인 도출
 - 특정 직역 또는 사업소 등에서 특별히 빈번하게 발생하는 질환 등을 연구하여 산업안전 차원에서의 건강연구 수행
 - 필요 시 시계열 연구를 통해 근로기간 및 제도 변화에 따른 직업재해 및 건강개선 등 추적조사 가능

③ 질환 양태 및 대책 관련 연구

- 감염성 질환 확산경로 연구 / 확산 실시간 감지
 - 그간의 진료기록을 분석, 모든 감염성 질환의 확산경로·속도를 파악하고 확률모형化 하여 향후 감염확산에 대비
 - 약제 처방 기록(DUR) 등을 활용, 준 실시간 수준으로 감염성 질환 확산 모니터링 및 양상 예측을 실시할 수 있도록 기술 개발
- 질환별 다빈도 세부유형 및 악화경로 연구
 - 질환 악화 경로 분석을 통해 사회 전체의 건강 수준을 개선하기 위한 건강검진·진료 정책 개선
 - 대규모 연구를 통해 한국인 호발성 질환에 대해 그 병리를 구체화하고 실제 의료 현장에서의 활용방안을 연구

④ 보건의료 전달체계 관련 연구

- 데이터기반 전달체계 유형별 문제분석 및 대책마련
 - 보건의료 전달체계 상 쏠림·비효율·중복투자 등 문제점을 발견하고, 이에 대한 대책을 마련
 - 특히 상시적 보건의료 전달체계 모니터링 체계로서 수가정책, 의료정책 변화에 따른 결과 관찰 및 개선책 도출
- 현 전달체계에서 해결하지 못한 새로운 건강문제 연구
 - 현재 보건의료 전달체계에서 적절하게 대응하지 못하고 있는 건강 문제(예:만성질환의 과도한 급성기 외래진료 등) 대응책 마련
 - 지역별, 의료기관별 비교를 통해 개선방안 도출

⑤ 보건·복지 정책개선 관련 연구

- 지역별 보건·복지 분야 시범사업의 건강결과 연구
 - 지역별 시범사업에 대해 미시행 지역과의 건강지표 비교를 통해 정책효과를 평가하고, 정책의 실효성 분석 및 개선방향 도출

- 주요 취약계층별 보건·복지 정책 효과 모니터링
 - 취약계층 대상 보건·복지 정책 시행 전후, 수급률과의 비교 등을 통해 건강지표에의 영향을 분석, 개선방안 도출

⑥ 희귀·난치 질환 관련 연구

- 희귀·난치 질환 임상양상 연구
 - 희귀·난치 질환 환자의 진료양상 등을 분석, 기존에 파악되지 않았던 약에 대한 반응 또는 임상양상을 연구
 - 자료 분석 과정에서 희귀·난치 질환의 기전 이해 및 치료제 개발의 실마리 발견 기대

⑦ 의약품 부작용 관련 연구

- 알려지지 않은 의약품 부작용 연구
 - 의약품 처방 자료와 진료결과 간의 관계 분석을 통해 품목허가 단계에서 밝혀지지 않았던 부작용 발견 가능
 - 장기추적, 대규모 추적이 필요한 부작용, 특정 환자군에서만 발생하는 부작용 등의 발견에 유효할 것으로 기대

⑧ 의료기술 간 효과비교 연구

- 치료법, 치료제, 의료기기 등 의료기술 간 효과비교 연구
 - 동일 질환에 대해 상이한 치료방법을 적용하였을 경우, 효과 측면에서의 상호 비교 연구 가능
- 비용효과성 측정
 - 진료결과의 분석 평가를 통해 특정 의료기술의 비용효과성을 분석, 보험급여 수가 결정 과정에서 활용
 - 특히 유사 효과 치료법 또는 유관 치료법과의 효과 비교결과 등을 종합적으로 고려한 비용효과성 산출 가능

* 향후 구체적인 범위는 위원회 검토를 거쳐 확정 추진

☐ **병의원·약국 등 기관 데이터셋**

- 요양기관 현황 상세내역 (심평원)
- 건강증진센터 이용현황, 개인 맞춤형 프로그램 현황 (건보공단)

☐ **유전·신체계측 등 환자 데이터셋**

- 국내 보유 생물테러가능 고위험병원체의 전장 유전체 염기서열 해독, 비교유전체 (질병관리본부)
- 시스템 생물학 기반 병원체 오믹스 표준 데이터 (질병관리본부)
- 유방암, 위암 환자군 유전정보 DB (질병관리본부)
- 장기요양DB 인정조사결과 DB (질병관리본부)

☐ **질환·역학 등 질병 데이터셋**

- 임신출산 여성, 장애인 등 취약계층, 경찰 등 특정 직업군과 같이 특정 대상자들로 구성된 목적용 DB (질병관리본부)
- 주요 질병 발생, 특정 질병 환자-대조군 등 특정 연구를 위한 목적용 DB (심평원, 질병관리본부)
- COPD 외 천식 등 관련 호흡기 질환 코호트 데이터 (심평원)
- 질병 소분류별 다빈도 상병 급여 현황(심평원) 및 시도별 신규 중증 암등록환자의 진료현황 데이터(암센터)

☐ **의약품정보·처방내역 등 의약품 데이터셋**

- 원외처방 약제 통계자료/의약품 상위 성분 청구현황 (심평원)
- 희귀난치성질환자 처방건수 및 약품비 정보 (심평원)
- 국가중점 의약품 처방(건보)

☐ **급여·보험료 등 행정 데이터셋**

- 보험급여 청구내역(심평원)
- 의료기관별 진료비 현황 정보(심평원), 건강보험가입정보(건보)

붙임 5
보건의료 빅데이터 관련 예산 내역
□ 세부 예산내역

(단위 : 백만원)

구분	사업명	정부안	최종 심의액	비고
정보화 사업	플랫폼구축	7,731	4,601	31.30억 (40.4%) 감액
	분석자료 공유네트워크	2,425	(좌동)	
	사업관리비	77	(좌동)	
	소계	10,233	7,103	31.30억 (30.58%) 감액
R&D	보건의료 빅데이터 시범R&D	1,134	(좌동)	
	보건의료 빅데이터 본R&D	1,900	(좌동)	
	평가연구, 거버넌스지원	200	(좌동)	
	소계	3,234	(좌동)	
합 계		13,467	10,337	31.30억 (23.2%) 감액

□ 보건의료 빅데이터 추진전략 수립을 위한 선행연구 실시 ('15~'16)

- 기본계획 수립 연구('15.11), 정보화 전략계획(ISP) ('16.1)

□ 보건의료 빅데이터 추진단 구성 ('17.3월~)

- (구성) 관계 부처, 공공기관, 기업·전문가 등 총 130여명으로 구성

* 학계·의료계·연구계 등 민간 전문가, 정부부처 관계자, 공공기관(건강보험공단, 심사평가원, 질병관리본부, 국립암센터), 5개 분과위원회 130여명, 회의 총 40여회

- (분과위원회) 추진단 산하에 5개 분과위원회를 구성, 분야별 논의

* (5개 분과) ① 총괄 분과 ② 서비스 개발 분과 ③ 데이터 개방·연계 분과
④ 정보보호·기술 분과 ⑤ 국민소통 분과

□ 보건의료 빅데이터 추진전략 심포지움 개최 ('17.6.12)

- 보건의료 빅데이터 추진전략 수립 배경, 서비스 개발 경과, 시스템 구축계획 및 데이터셋, 정보보호전략·국민소통전략 개요 등 논의

□ 시민사회단체 의견청취 ('17.8~, 진행 중)

- 시민사회단체 간담회 3회 개최 (12월 중 추가 개최 예정)

- ◇ 국가별 기술·제도는 다소 상이하나, 데이터 기반 건강증진 정책은 공통적으로 추진
- ◇ 주요 선진국은 ①국가 전략 수립, ②전담 기관(플랫폼) 신설, ③서비스·데이터 개발, ④제도 마련, ⑤R&D 지원의 사이클로 효율적인 활용 체계 구축

□ (미국) 연방정부 중심의 전략·법제·표준·데이터 등 보건의료 빅데이터 활용 기반 마련 추진('04~)

- * [전략] 의료정보기술추진계획('04), 빅데이터 추진계획('12), 정밀의료 추진계획('16)
- [법제] 건강정보기술법(HITECH Act, '09), 21세기 치유법('16)
- [표준] 보건부 산하 건강정보표준관리국(ONC) 설치('04), 표준화 활동 지속

- 연방정부에서 보유한 **Medicare·Medicaid** 데이터를 연구자와 의료기관에게 제공하고, 연구·본인열람·정책개선 등에 활용 중

□ (일본) 첨단 의료 기술·서비스를 실현하고 의료·의약품·의료기기산업을 육성하기 위해 법제화, 데이터 정비, 활용 연구 추진

- * [전략] 일본재흥전략('13), 건강·의료전략('13) [법제] 차세대의료기반법('17)
- * 익명화 및 정보제공을 민간에까지 허용, 사후철회 제도 도입('17)

□ (호주) 데이터 결합·제공 서비스('95~)의 활용 및 이를 기반으로 국가적 전략을 수립('13)하고, 연구개발·정책개선을 지원

- 의료이용자료, 암등록자료, 인구동태자료, 출생·사망, 부모정보 등 개인 단위 자료를 결합 후 비식별화하여 연구자에게 제공

□ (英 잉글랜드) 보건의료 데이터 활용 사업*을 추진 중이었으나, 개인 정보처리동의 미흡, 본인통지 누락 등을 이유로 폐지('16.7)

- * NHS Digital의 care.data 프로그램

□ (英 스코틀랜드) 제3자 신뢰방식(TTPI)이라는 방법을 활용, 연구 목적으로 빅데이터를 결합·제공하는 프로그램(SILC) 운영('14~)

- * (제3자 신뢰방식) 공급기관·연계기관 외 독립적인 제3자를 둬으로써, 데이터 원천기관·결합기관 모두 최소한의 정보만으로 자료결합을 할 수 있는 기술

○ 보건의료 빅데이터 플랫폼

개요	주요 공공기관과의 온라인 연계를 통해 기관보유 빅데이터를 실시간으로 제공받아 결합, 처리 등을 거쳐 이용자에게 제공
이용자	정책입안자·연구자, 의학연구자, 일반국민 등
핵심기능	
<p>○ (자료연계) 기관별 데이터공개 로드맵에 기반하여 공개용 데이터마트 구축하고, 플랫폼 수준에서의 협약을 통한 연계체계 구축</p> <ul style="list-style-type: none"> - 플랫폼에서 연계요청시 데이터 추출 프로세스가 동작하고, 추출된 데이터는 네트워크로 전송하여 연계 데이터 셋을 신속하게 생성 - 비식별화 및 암호화된 데이터를 기관으로부터 전송받아 수집하고 연계하고, 연계에 사용된 개인식별정보는 삭제 후 단순 일련번호로 대체 <p>○ (정보보호) 개인식별정보 및 민감정보는 기본적으로 기관별 비식별화-플랫폼 비식별화의 이중 비식별화를 거쳐 이용자에게 제공함</p> <ul style="list-style-type: none"> - 비식별화는 자동전략수립→비식별화→평가(보고서 생성) 절차로 구성 - 주민번호는 플랫폼과 별도로 존재하는 암호화 키 발급시스템에서 주민번호 암호화 방식을 요청시마다 기관으로 재송부하여 암호화함 <p>○ (이용자제공) 각 기관으로부터 수합하여 처리한 자료는 활용 플랫폼으로 전송, 방법에 따라 활용할 수 있도록 함</p>	
플랫폼 개념도	
<pre> graph TD subgraph 원천자료 [원천자료] A[건강보험] B[심평원] C[질본] D[암센터] end subgraph 시스템 [시스템] E[보건의료 빅데이터 플랫폼 가칭] E1[데이터 완전공개 웹사이트] E2[맞춤형 데이터셋 요청 웹사이트] E3[폐쇄형 자료활용 공간] end subgraph 사용자 [사용자] F[이용자] end A --- E B --- E C --- E D --- E E --- F E --- E1 E --- E2 E --- E3 </pre>	

○ 데이터 완전공개 웹사이트

개요	①미리 잘 정리된 파일 형태의 공개 데이터 다운로드기능 제공 ②OpenAPI 방식으로 업데이트된 최신 자료를 제공
이용자	정책입안자, 연구자
핵심기능	
<ul style="list-style-type: none"> ○ (수집) 수요조사, 기 제공된 데이터 등을 통해 제공 데이터셋 개발 ○ (정보보호) 공공영역으로 제한 없이 제공됨에 따라 강력한 비식별화 및 비식별화 검증 ○ (제공) 웹사이트를 통한 제공 또는 API를 통한 외부 앱에서 활용할 수 있도록 제공 	
플랫폼 개념도	
<p>The diagram illustrates the platform concept. On the left is a box labeled '공공기관' (Public Institution). In the center is a box labeled '포탈' (Portal) which contains two sub-sections labeled 'REST'. On the right are two boxes labeled '브라우저' (Browser) and 'Apps'. Arrows indicate the flow of data: from '공공기관' to '포탈' (indicated by two right-pointing arrows labeled '--->'), and from '포탈' to both '브라우저' and 'Apps' (indicated by two left-pointing arrows labeled '<---').</p>	

○ 맞춤형 데이터셋 요청 웹사이트

개요	가상 환경에 설치된 통계·시각화 패키지(R, 태블로 등) 및 데이터 공유 플랫폼에서 제공된 데이터를 활용, 통계처리 등 가공
이용자	정책입안자, 연구자
핵심기능	
<ul style="list-style-type: none"> ○ (신청) 연구목적, 필요 데이터 등을 기재 후 신청 ○ (심의) 사회편익, 안전성 등을 고려하여 심의 ○ (제공) 비식별화한 데이터 제공 ○ (관리) 데이터 사용 후 폐기 또는 지속 사용 시 폐기 여부 및 보안체계 유지 등 사후 관리 실시 	
플랫폼 개념도	
<p>The diagram illustrates the platform concept as a linear process flow. It consists of five boxes connected by arrows: '연구자' (Researcher) → '제공기관' (Providing Institution) → '심의기구' (Review Body) → '연구자' (Researcher) → '활용' (Utilization). Above the arrows, the stages are labeled: '신청' (Application) above the first arrow, '심의' (Review) above the second arrow, '제공' (Provision) above the third arrow, and '관리' (Management) above the fourth arrow.</p>	

○ 메타데이터 카탈로그 웹사이트

개요	정부가 보유한 보건의료 데이터의 세부 메타데이터를 검색할 수 있도록 제공(통상 “완전공개 웹사이트”와 통합운영 추세)
이용자	정책입안자, 연구자
핵심기능	
<ul style="list-style-type: none"> ○ (카탈로그) 미국, 영국, 호주 등의 정보공개포털에서 사용되는 CKAN을 적용하여 보건의료 빅데이터 카탈로그 제공 <ul style="list-style-type: none"> * Comprehensive Knowledge Archive Network, 세계최대 오픈소스 데이터 카탈로그 솔루션 ○ (정보제공) 포털을 통한 정보 제공 및 REST 인터페이스를 활용하여 API 정보 제공 병행 	
플랫폼 개념도	
<p>The diagram illustrates the platform architecture. On the left, a box labeled '공공기관' (Public Institution) has three arrows pointing right towards three stacked components: '포털' (Portal), 'REST CKAN카탈로그 REST', and 'CKAN수집기' (CKAN Collector). The '포털' component has an upward arrow pointing to the 'REST CKAN카탈로그 REST' component, which in turn has an upward arrow pointing to the 'CKAN수집기' component. On the right, two boxes labeled '브라우저' (Browser) and 'Apps' have arrows pointing left towards the '포털' and 'REST CKAN카탈로그 REST' components respectively.</p>	

○ 빅데이터 활용 선진사례 웹사이트

개요	빠르게 변화하는 보건의료 빅데이터 기술 및 시장 대응하기 위하여 새로운 정부 서비스 제안 및 정부 서비스를 활용한 좋은 사업례 등 등록
이용자	일반국민, 플랫폼 이용자
핵심기능	
<ul style="list-style-type: none"> ○ (서비스제안) 데이터를 활용한 서비스 개발/제공 신청 창구를 일원화 하여 기관별 중복 사업 방지 및 효율적인 신규 사업 개발 <ul style="list-style-type: none"> * 신규 서비스, API 개발, 제공 통계 등에 활용 ○ (사업현황) 정부 제공 서비스 활용 사례 공유를 통해 피드백을 통한 사업 개선 등 선순환 구조 구축 	
플랫폼 개념도	
<p>The diagram shows the platform concept. At the top center is a box labeled '정부' (Government). Below it is a box labeled '플랫폼' (Platform). To the left of the '플랫폼' box is a box labeled '국민' (Citizen), and to the right is a box labeled '이용자' (User). Arrows indicate interactions: a double-headed vertical arrow between '정부' and '플랫폼' is labeled '신규사업개발' (New Business Development); a horizontal arrow from '국민' to '플랫폼' is labeled '제안' (Proposal); a horizontal arrow from '플랫폼' to '이용자' is labeled '제안' (Proposal) and '이용지원' (User Support).</p>	

○ 분석자료 공유·활용 네트워크

개요	민간 의료기관 및 공공기관 내 보건의료 데이터 간 “분석자료 공유·활용 네트워크” 구축					
이용자	정책입안자, 연구자					
핵심기능						
<div>○ (개요) 연구자가 분석을 의뢰하면, 각 병원이 보유한 의료기록(EMR) 데이터에서 개인정보 이동 없이, 통계처리·분석한 결과만을 취합하여 제공<ul style="list-style-type: none">* 병원에서 의료기록이 반출되지 않음- 현재 각 병원에 분석을 의뢰하면 개별적으로 제공이 가능한 자료이나, 모든 병원에 제각각 요청하는 불편을 덜 수 있음</div> <div>○ (연구지원) 데이터 분석을 위한 소프트웨어 제공 및 인력 지원 등을 통해 의료기관의 빅데이터 연구 지원</div> <div>○ (분석제공) 수요기업의 요청에 따른 각 기관의 분석 결과 데이터를 통합 분석하여 효율적인 연구 데이터 제공</div>						
플랫폼 개념도						
<div><div><div>의료연구자</div><div>정책수요자</div></div><div>이용자</div><div>요청</div><div>분석자료 공유·활용 네트워크</div><div>SW제공</div><div>A병원</div><div>B병원</div><div>병원</div><div>통합데이터</div><div>분석데이터</div></div> <div>< 플랫폼 활용 예시 ></div> <div><div>* 예) 분석요청 : A약을 복용한지 3개월 내 부작용을 보인 사례의 비율</div><div>병원별 분석 : <table><tr><td>X병원 1/100(1%)</td><td>Y병원 2/100(2%)</td><td>Z병원 1/200(0.5%)</td></tr></table></div><div>취합·합산 : <table><tr><td>합산 : 4/400(1%)</td></tr></table></div><div>분석제공 : <table><tr><td>4/400(1%)</td></tr></table></div></div>		X병원 1/100(1%)	Y병원 2/100(2%)	Z병원 1/200(0.5%)	합산 : 4/400(1%)	4/400(1%)
X병원 1/100(1%)	Y병원 2/100(2%)	Z병원 1/200(0.5%)				
합산 : 4/400(1%)						
4/400(1%)						