

# 보건의료 현장을 바꾸는 임상 빅데이터

가톨릭 빅데이터 통합 센터장 | 김대진

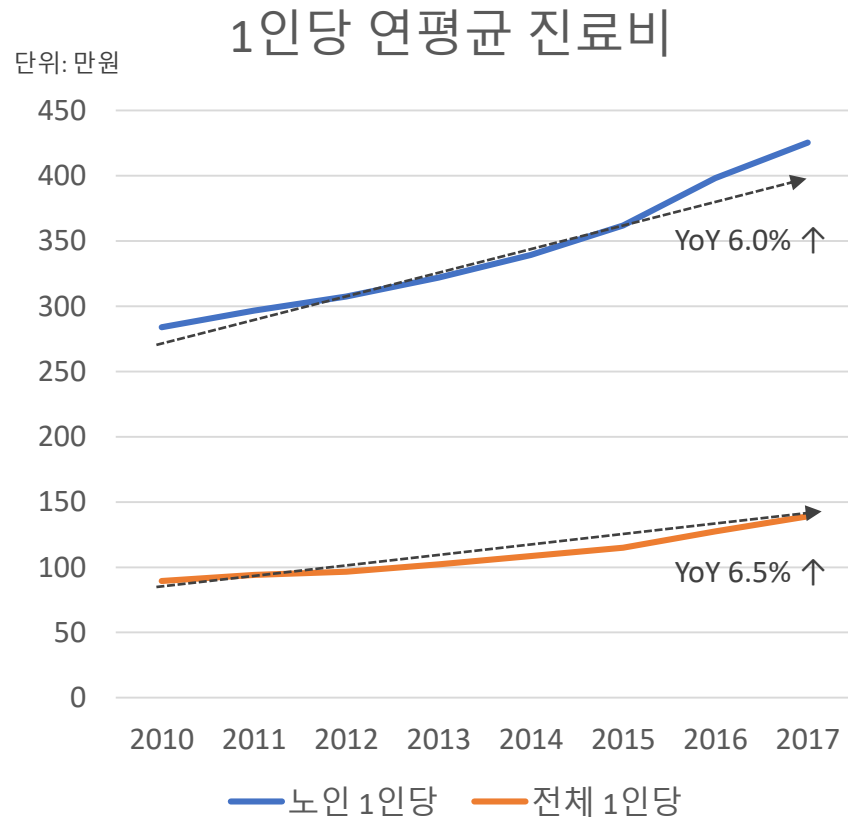
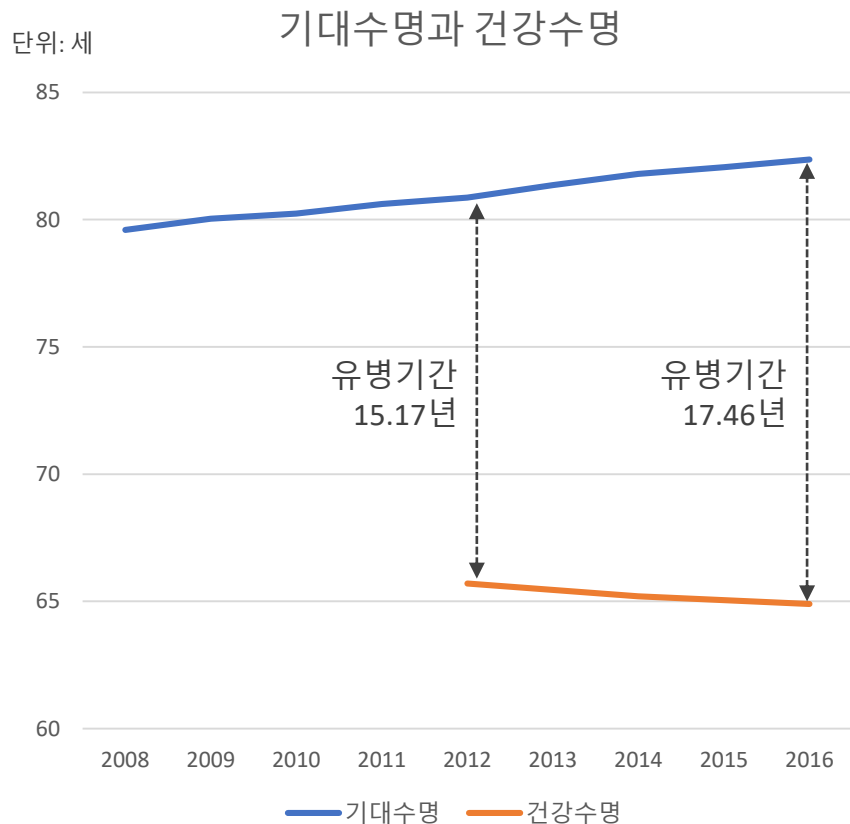
2018. 11. 21.

살면서 병원에 얼마나 많이 갈까요?

1년에 병원비로 얼마나 쓰시나요?

# 살면서 병원 신세 얼마나 질까?

대한민국 국민의 기대수명은 늘어나고 있으나 유병기간도 늘어나고 있으며, 진료비는 고령화 및 고가 진료행위의 영향으로 물가 상승률(YoY 1.8%)의 3배가 넘는 속도로 늘어나고 있음



출처: 통계청, 「생명표, 국가승인통계 제101035호」, 2017  
※ 2013년 및 2015년 건강수명은 추정치

출처: 건강보험심사평가원 | 국민건강보험공단, 2017 건강보험통계연보, 2018

# 보건의료 빅데이터 산업의 필요성

타 산업분야 대비 보건의료 분야의 빅데이터 사업은 산업적, 공익적 편익 뿐 아니라 국민 개개인의 의료비 절감으로 인해 체감 효과가 큰 사업으로, 삶의 질 향상에 크게 기여 가능

의료 서비스 수준 향상	의료비 절감	신산업 발현
<ul style="list-style-type: none"><li>▪ 현재는 <b>의료진간 실력의 격차가 상당히 크기</b> 때문에 투약, 처치, 수술 등 의료 행위의 정확도가 천차만별</li></ul> <p>↓</p> <ul style="list-style-type: none"><li>▪ 빅데이터 및 인공지능 기술 활용으로 <b>의료진 서비스 편차 축소 및 정확도 향상</b> 등으로 서비스 성과 30~40% 향상 기대**</li></ul>	<ul style="list-style-type: none"><li>▪ 최근 10년간(2007-2016) 대한민국 의료비는 GDP 성장률 대비 2.6배 이상 높은 증가율을 보임*</li></ul> <p>↓</p> <ul style="list-style-type: none"><li>▪ 빅데이터 및 인공지능 기술 본격 활용시 <b>의료비 50% 이상 절감 기대**</b></li></ul>	<ul style="list-style-type: none"><li>▪ 국내 보건의료 분야는 활용 가능한 데이터는 많지만 제대로 <b>접근하지 못하고 있는 것으로 분석***</b></li></ul> <p>↓</p> <ul style="list-style-type: none"><li>▪ 빅데이터 전문센터 사업을 통해 <b>데이터 전문 사업 및 연계 · 파생 신산업 활성화</b> 기대</li></ul>

보건의료 빅데이터를 활용한 국민 삶의 질 향상 필요성 증대

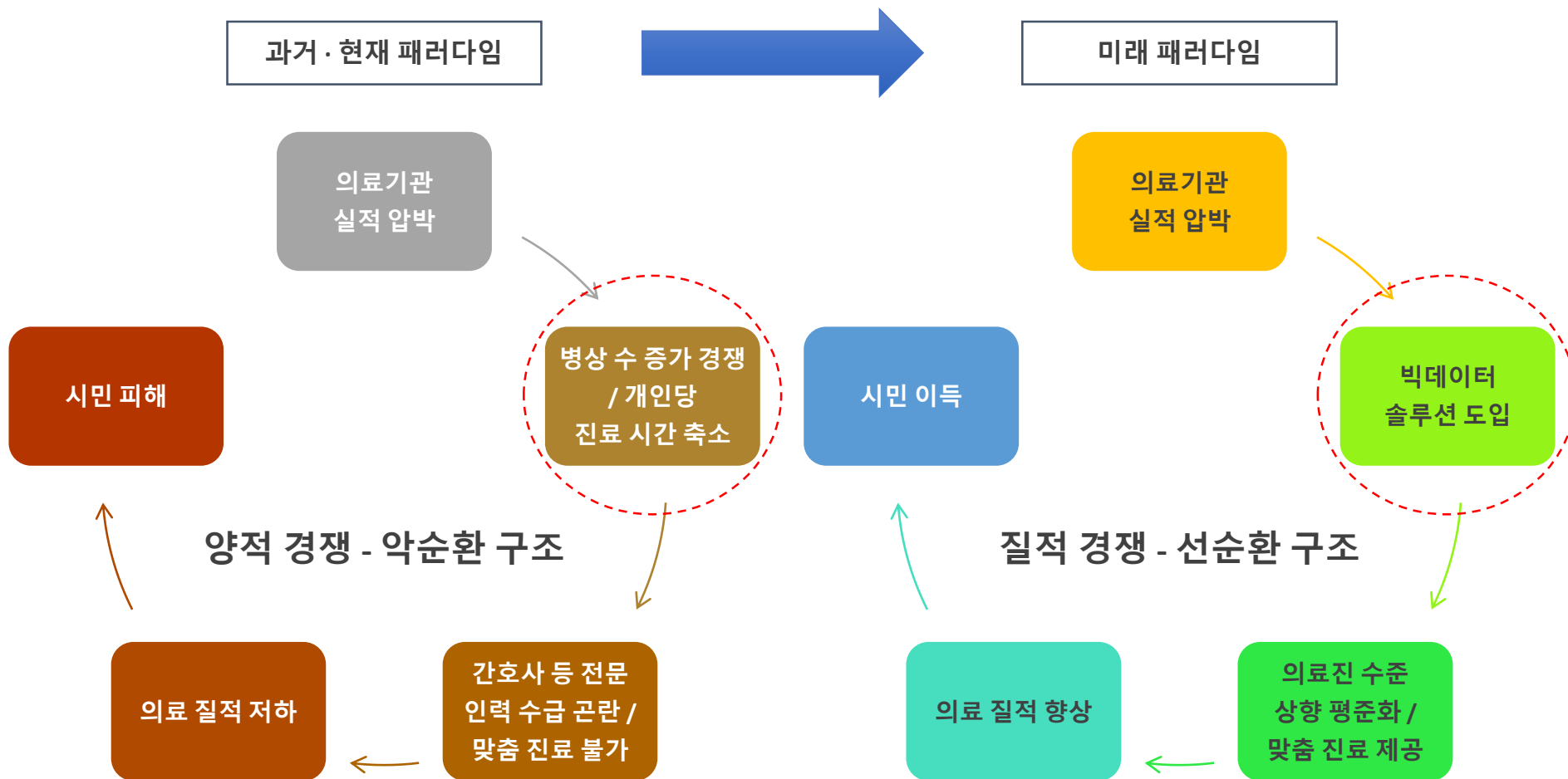
출처: \* 한국은행, 2018 / 산업연구원, 2017

\*\* ETRI, 2016

\*\*\* OECD, 2015 (산업연구원, 2017 자료에서 재인용)

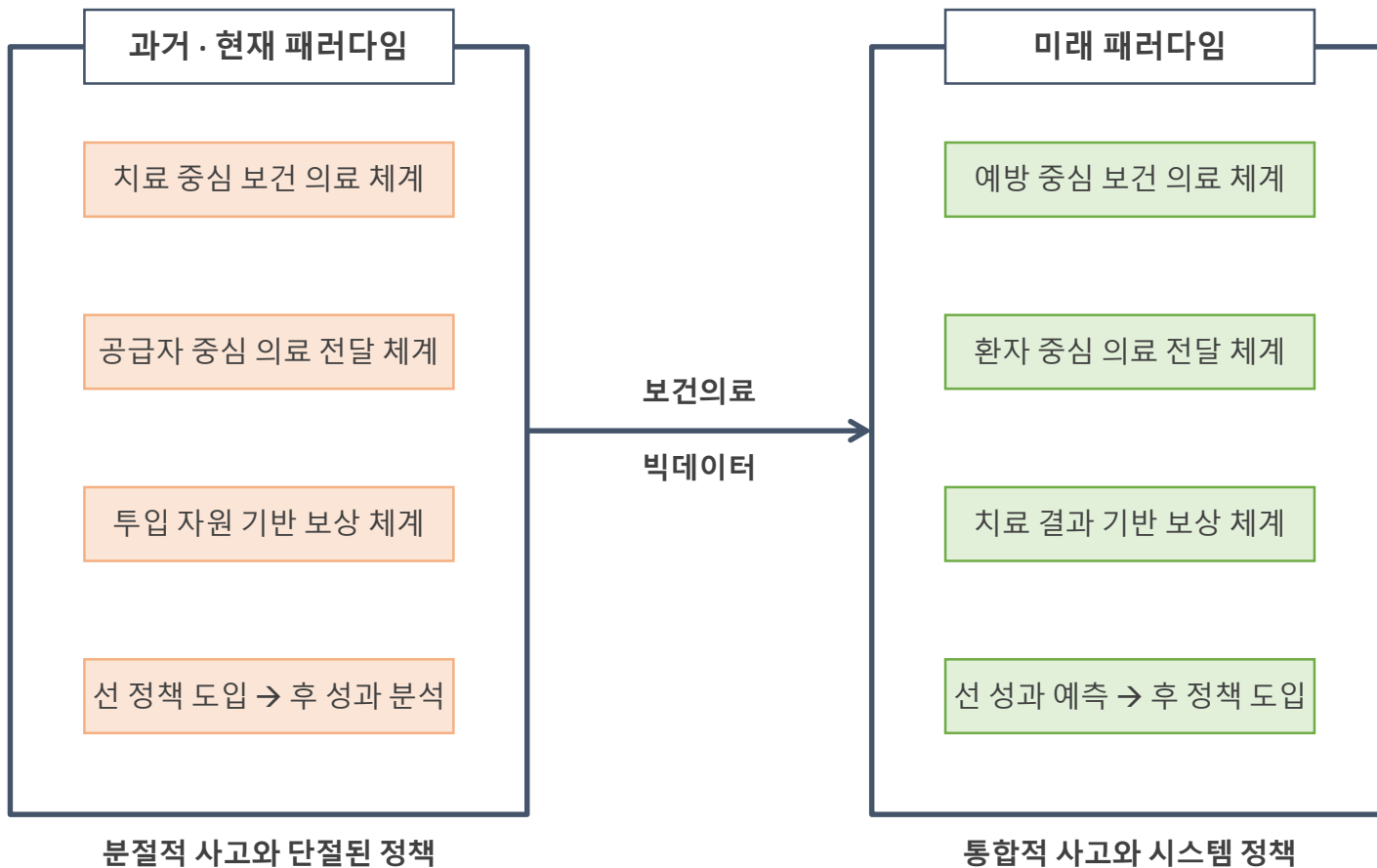
# 보건의료 빅데이터의 필요성

빅데이터는 중·대형 병원들이 병상 수 증가나 환자당 진료 시간 축소와 같은 방식의 양적 경쟁보다 빅데이터 분석을 통한 의료 서비스 고도화로 질적 향상 경쟁을 촉진할 수 있는 요인



# 보건의료 빅데이터의 패러다임 변화

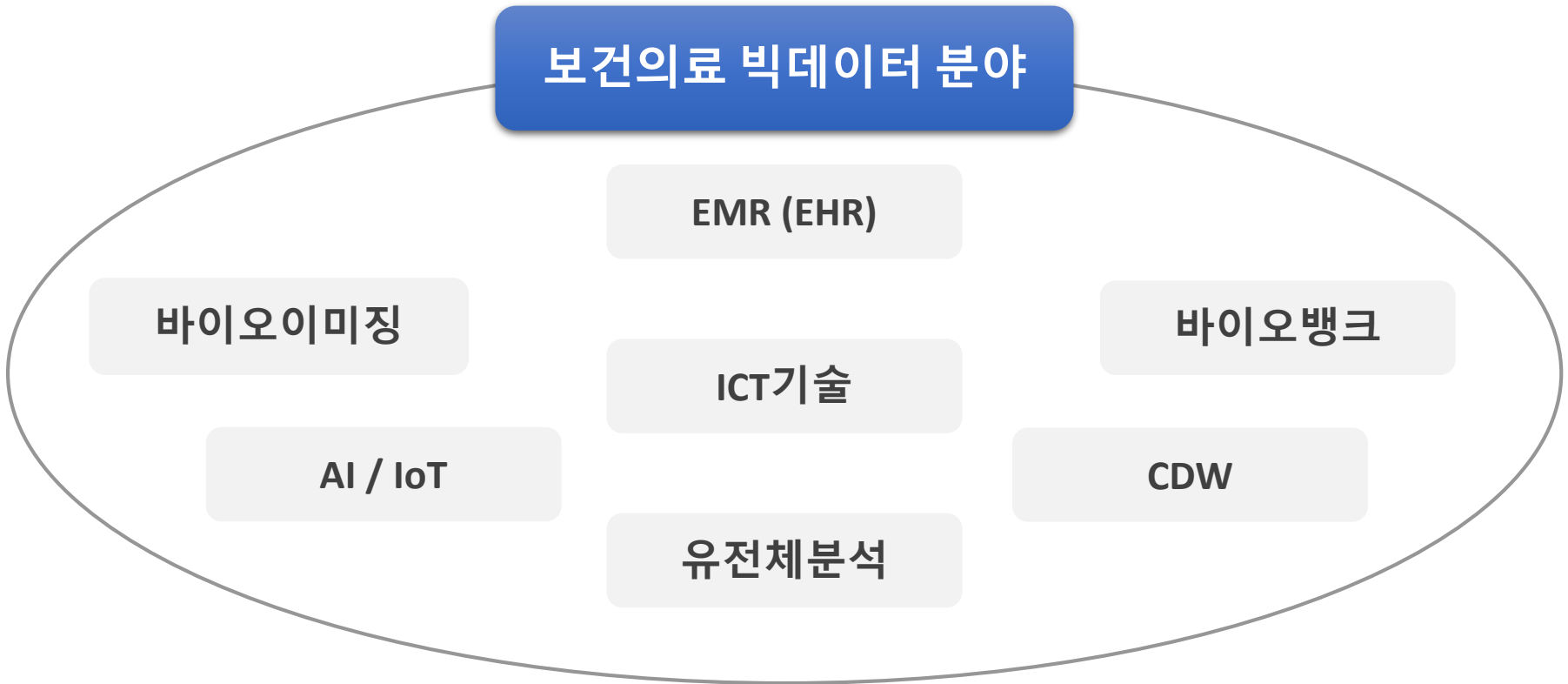
신기술을 통해 데이터 정밀 분석이 가능해지면서 정밀의학으로 패러다임이 변하였고, 개인의 건강 및 생활 환경 데이터를 통합하여 최적화된 의료서비스 제공과 비즈니스 모델 창출 가능



출처: 한국보건사회연구원, 2015

## 보건의료 빅데이터란?

의료 환경에서 빠른 속도로 발생하는 다양하고 방대한 규모의 의료 관련 데이터이자, 의료 빅데이터를 수집/저장/분석해 유의미한 가치를 제공하는 기술



# 보건의료 분야에서의 빅데이터

보건의료 분야 빅데이터 전문센터는 인공지능을 활용한 정밀의료의 기반이며, 데이터 유통을 위한 필요 조건으로, 산업 발전 단계의 필수 요소



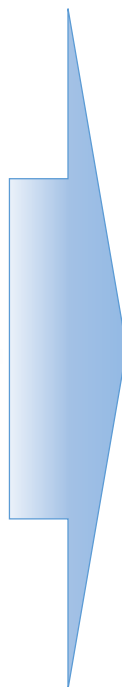
## 데이터 생산

- 진료 / 진단
- 임상 연구



## 데이터 웨어하우스

- 개별 의무 데이터 집적
- 제한적 분석 및 활용



## 빅데이터 센터

- 데이터 통합 / 연결
- 데이터 분석 / 시각화
- 데이터 공유 / 유통



## 인공지능 활용

- 이종 데이터 연계 분석
- 데이터 기반 예측
- 데이터 기반 정밀의료

데이터 전산화

빅데이터

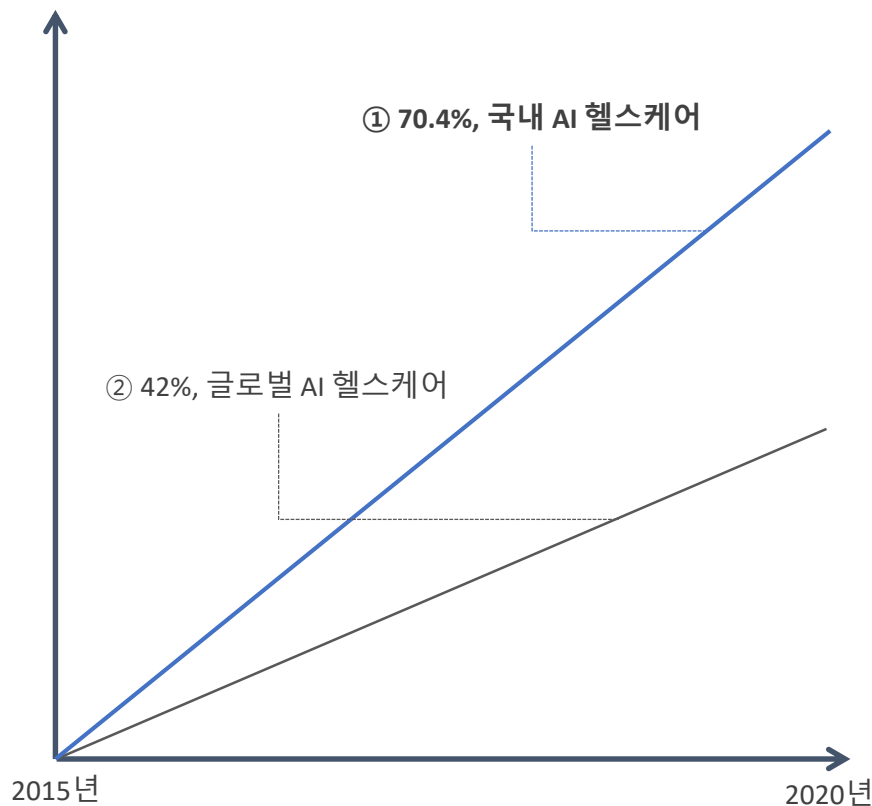
인공지능



# 보건의료 빅데이터 중요성 및 시장 전망

보건의료 분야 빅데이터 시장은 인공지능 기술과 결합시 높은 성장이 예상되고 있으나, 데이터 표준, 규제, 예산 등의 이유로 활성화되지 못하고 있음

연평균 시장 성장률 전망 (YoY)



출처: ① KISTI, 2016  
② Markets and Markets, 2016 – 2015-2021 전망치

표 6 빅데이터 도입 관심 수준(미도입 기업) (단위: 기업 수, %)

구분	사태수	논의된 적 없음		전산부서 차원		협업부서		CEO/CIO		기타	
		N	%	N	%	N	%	N	%	N	%
전체	734	452	61.6	193	26.3	48	6.5	37	5.0	4	0.5
업종	공공	170	95	55.9	62	36.5	8	4.7	5	2.9	-
	금융	33	13	39.4	12	36.4	5	15.2	3	9.1	-
	유통/서비스	120	77	64.2	18	15.0	11	9.2	14	11.7	-
	제조	312	206	66.0	82	26.3	12	3.8	10	3.2	0.6
	의료	47	33	70.2	11	23.4	2	4.3	1	2.1	-
중견원수	통신/미디어	52	28	53.8	8	15.4	10	19.2	4	7.7	3.8
	300명 미만	455	304	66.8	96	21.1	28	6.2	24	5.3	0.7
	300~999명	192	115	59.9	55	28.6	13	6.8	8	4.2	0.5
	1,000명 이상	72	33.3	33.3	38	52.8	5	6.9	5	6.9	-

표 7 빅데이터 미도입 이유(미도입 기업) (단위: %, 중복응답)

구분	사태수	빅데이터 라고 할 만한 데이터가 없음	빅데이터 관련 전문인력 없음	빅데이터를 분석할 만큼 큰 기업이 아님	지금만 아니지만 차후에 고려해 정립	시스템 구축비 관리비 등 예산부족	빅데이터 도입효과를 나타낼 업무가 없음	CEO/CIO (경영진의 무관심)	자사에 적합한 빅데이터 제품이 없음	빅데이터 자체가 애초부터 갈 오름	빅데이터 도입효과에 대한 불신
전체	775	37.2	32.1	24.8	23.2	21.5	19.5	9.5	8.3	7.4	6.3
업종	공공	182	41.8	31.9	15.4	27.5	29.1	19.2	9.9	6.6	7.7
	금융	34	29.4	35.3	8.8	35.5	14.7	17.6	20.6	11.8	5.9
	유통/서비스	132	42.4	36.4	28.8	25.8	15.9	15.9	6.1	6.1	8.3
	제조	321	36.1	30.2	28.0	19.9	18.4	20.9	10.0	9.7	8.7
	의료	50	24.0	38.0	36.0	20.0	36.0	16.0	18.0	8.0	12.0
중견원수	통신/미디어	56	32.1	26.8	26.8	17.9	19.6	25.0	-	8.9	3.6
	300명 미만	482	40.2	32.4	30.3	20.5	20.3	19.7	10.2	9.3	8.5
	300~999명	200	33.5	28.0	20.0	23.5	21.5	20.0	7.5	6.5	5.5
	1,000명 이상	75	26.7	40.0	5.3	38.7	28.0	20.0	13.3	6.7	2.7

출처: NIA, 2016

## 국내 현황 (규제 및 활용)

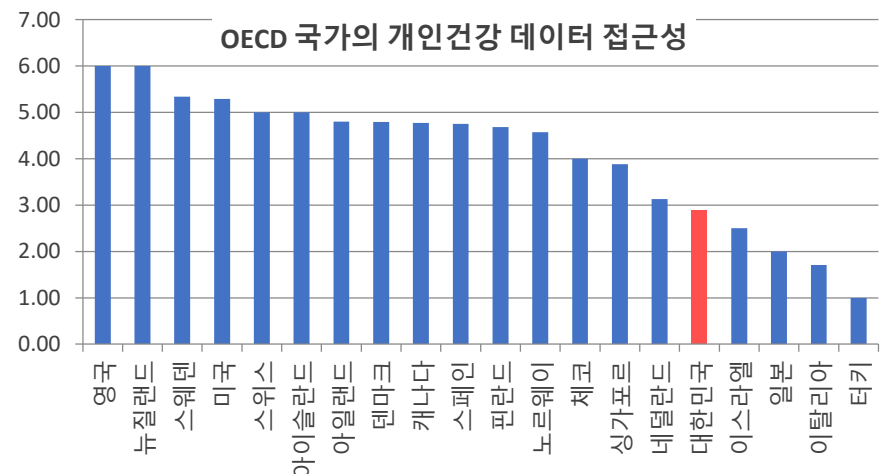
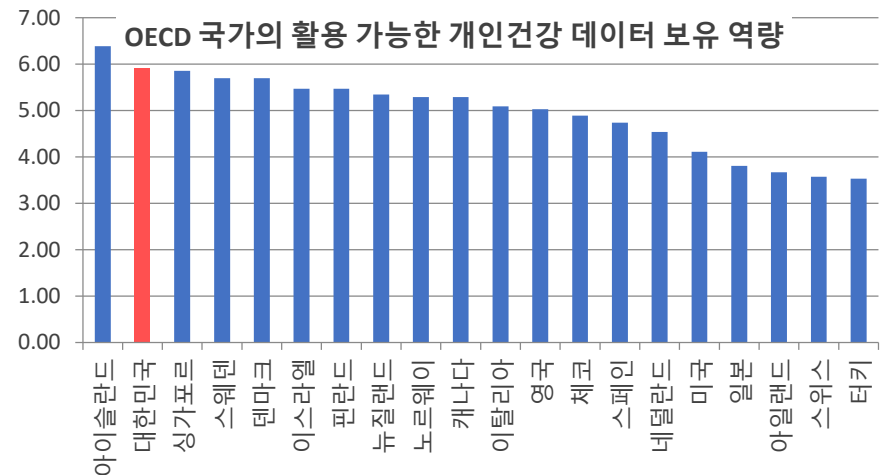
국내 병원의 EMR 도입 비율은 병원 등급에 따라 92~100%로 세계 최고 수준이지만, EMR 기술 수준 격차가 크며, 교류는 병원간만 8%에 정도에 불과하고 타 산업과의 연계는 사실상 없음

주요국 데이터 규제 수준

국가	데이터 수집	데이터 보관	융합 및 타목적 사용	프로파일링 차별화	데이터 교환	평균
대한민국	4	4	3	3	4	3.6
EU	3	2	3	3	3	2.8
미국	2	2	3	3	2	2.4
뉴질랜드	3	2	3	1	2	2.2
OECD	1	2	2	3	2	2.0
싱가포르	2	2	2	2	2	2.0
호주	1	2	2	3	2	2.0

(4: 장애 요인, 3: 상당한 수준의 장애 요인, 2: 약간의 장애 요인, 1: 규제가 약한 수준)

출처: Analysys Mason, 2014 (한국데이터진흥원, 2017 자료에서 재인용)



출처: OECD, 2015 (산업연구원, 2017 자료에서 재인용)

## 국외 현황

보건의료 빅데이터 센터 구축 및 데이터 공개에 대한 활발한 투자와 논의가 이루어지고 있음

지역	제도, 정책 및 법률	추진 현황
미국	<ul style="list-style-type: none"> <li>The Federal Big Data R&amp;D Strategic Plan (2016)</li> <li>Data to Knowledge to Action (2013)</li> <li>Big Data R&amp;D Initiative (2012)</li> </ul>	<ul style="list-style-type: none"> <li>국립보건원 (NIH)을 중심으로 NITRD (Networking and Information Technology Research and Development) 프로그램을 통해 R&amp;D 확대</li> <li>의료분야 빅데이터 컴퓨팅 전문센터 설립 추진</li> <li><b>정밀의료계획 추진을 통해 백만 명 이상의 코호트 구축 중</b></li> </ul>
영국	<ul style="list-style-type: none"> <li>빅데이터 역량 강화 전략 (2013)</li> <li>2010년 자유보호법 (2010)</li> </ul>	<ul style="list-style-type: none"> <li>개인 의료정보 수집·통합을 위한 보건의료 빅데이터 센터 기구축 (2013)</li> <li>보건의료 서비스 개발 지원 전담 독립기구 NHS Digital (구 HSCIC) 운영</li> <li><b>Google DeepMind Health 프로그램으로 NHS 보유 환자 데이터를 별도 동의 없이 비식별화를 거쳐 인공지능 프로젝트에 활용</b></li> </ul>
일본	<ul style="list-style-type: none"> <li>일본재흥전략 2016 (2016)</li> <li>Active Japan ICT (2012)</li> <li><b>헬스케어 데이터 활용을 위한 차세대의료기 반법(2017)</b></li> <li>빅데이터 활용을 통한 사회 문제 해결에 주력</li> </ul>	<ul style="list-style-type: none"> <li>보건의료 빅데이터 R&amp;D 강화, 해석기술 전문가 양성 추진</li> <li>테스트베드 개방을 통한 보건의료 빅데이터 지원 확대</li> </ul>
대한민국	<ul style="list-style-type: none"> <li>공공데이터의 제공 및 이용 활성화에 관한 법률 (2016)</li> <li>정부 3.0 추진기획 기본 (2013)</li> <li>빅데이터 마스터플랜 (2012)</li> <li>공공정보 공개 및 민간활용 활성화 추진</li> </ul>	<ul style="list-style-type: none"> <li>보건의료 빅데이터 활용을 통한 과학, 기술, 의료 선진화 추진</li> <li>보건의료 데이터의 융합분석을 통한 질병 예측, 건강정보 진단, 맞춤형 서비스 제공 추진</li> <li>의약품 정보 등 보건의료 공공 정보 공개 확대</li> </ul>

출처: 국회도서관, 2018 / 한국데이터진흥원, 2017 / 매일경제, 2017 / 창의미래연구소, 2013 / DeepMind 웹사이트/과학기술정책연구원, 2018

## 가톨릭 중앙의료원의 보건의료 빅데이터 역량

가톨릭 중앙의료원은 국내 최대 규모 단일 의료원으로, 동일 EMR을 사용하는 8개 부속병원을 통해 7% 이상 국민의 환자 데이터를 보유하고 있어 국내 최대 민간 의료 빅데이터 구축 가능

### 보건의료 역량

- 국내 최대 단일 의료 네트워크
- 국내 최대 전국 네트워크
- 국내 최대 환자 수
- 국내 최대 1-2-3차 연계 네트워크

### 의료 빅데이터 역량

- 국내 최대 단일 EMR 시스템
- 동일한 데이터베이스
- 동일한 의료 용어
- 동일한 개발 업체

### 가톨릭 빅데이터 통합센터



## 빅데이터 전문센터 비전 및 핵심 가치

빅데이터 네트워크 전문센터 협의회의 일원이자 보건의료 분야 유일의 국비 지원 전문센터로서 빅데이터 산업 생태계 조성 역할을 충실히 수행하는 것을 목적으로 함



가톨릭대학교 컨소시엄은 「보건의료 빅데이터 전문센터 네트워크」 사업의 성공적 완수를 위해 데이터 관련 업무 뿐 아니라 거버넌스와 법·제도 관련 사항까지 종합적 추진

# 보건의료 빅데이터 전문센터 네트워크

### 거버넌스

- 관련 사업 통합 추진
- 전문위원회 구성
- 운영 정책

### 데이터 생산·가공

- 질환 3종
- 고지혈증
- 알코올 사용장애
- 당뇨
- 데이터 표준
- PHR 데이터

### 유통 플랫폼

- CKAN 플랫폼 (기관)
- PHR 플랫폼 (개인)
- 샌드박스 시스템 구축 (고가치, 맞춤형)

### 생태계 형성

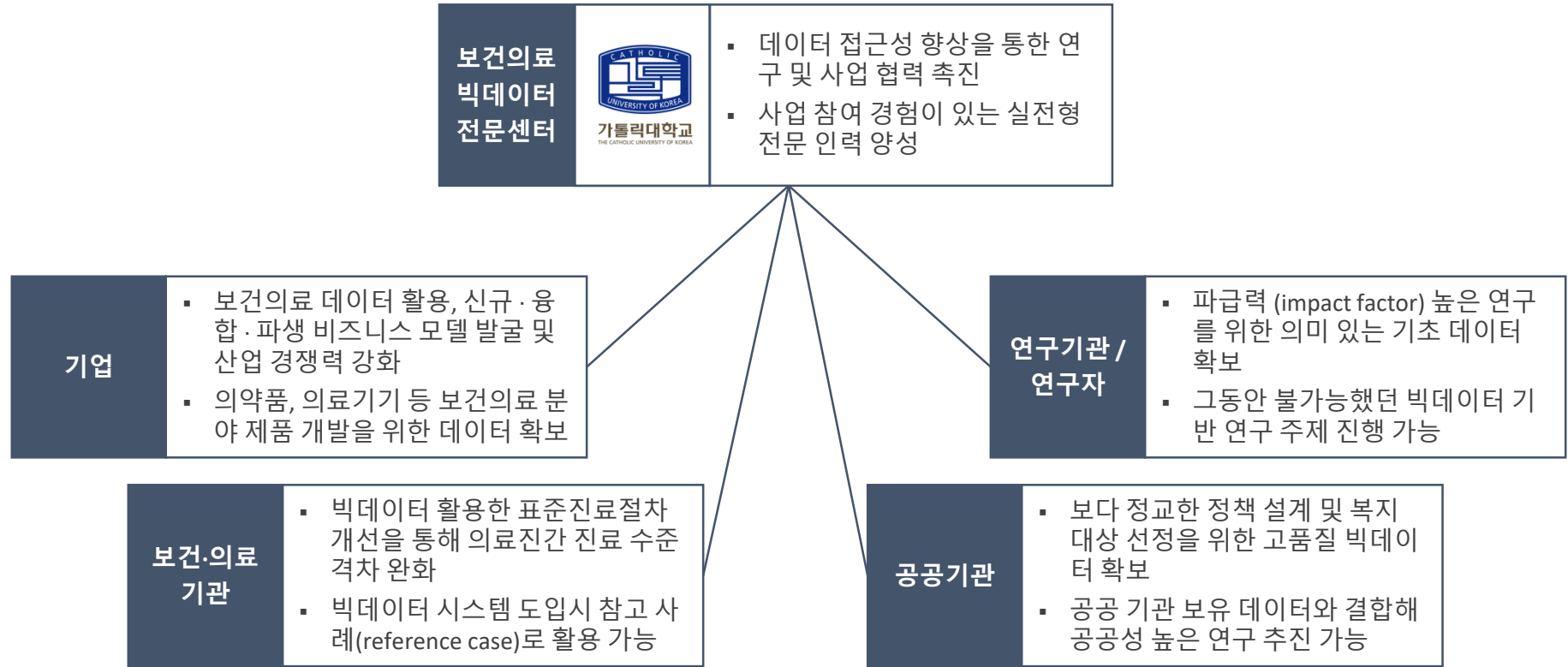
- 30개 기관 네트워크
- ODHSI국제네트워크
- 국내외 학술 네트워크
- 빅데이터활용경진대회
- 스타트업 지원센터
- 인력 양성 프로그램

### 법·제도

- 국내외 현황 파악
- 정책 동향 파악
- 당면 이슈 검토
- 개선안 도출

## 이용 대상 및 수혜자, 기대 효과

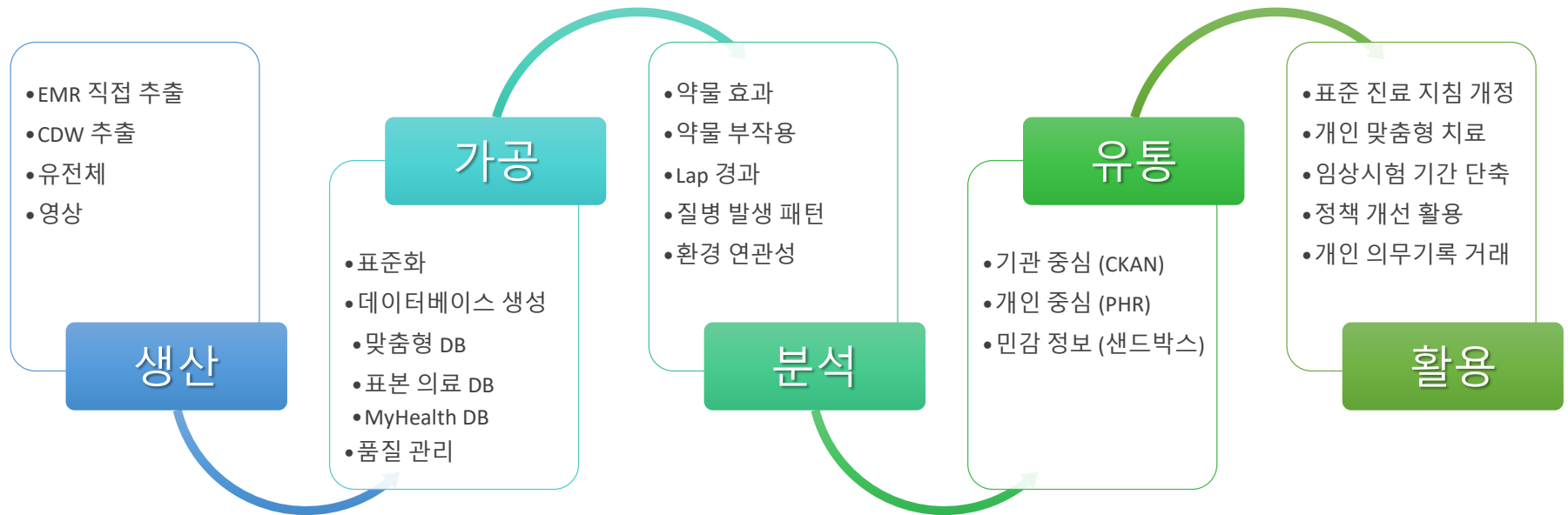
보건의료 빅데이터 전문센터를 통해 병원 뿐 아니라 바이오/제약, 의료기기, ICT 등 다양한 산업과 공공, 연구 분야에 걸친 파급 효과 및 국민 만족도 향상 기대 가능



**맞춤형 건강 관리 확산 및 의료비 절감으로 국민 삶의 질 향상**

# 빅데이터 가치사슬

빅데이터 가치사슬(value chain) 5단계를 기관 및 사업 컨소시엄 내에서 수행할 수 있는 체계를 갖추었으며, 기관 뿐 아니라 일반 국민도 사업의 성과로 이익을 얻을 수 있음

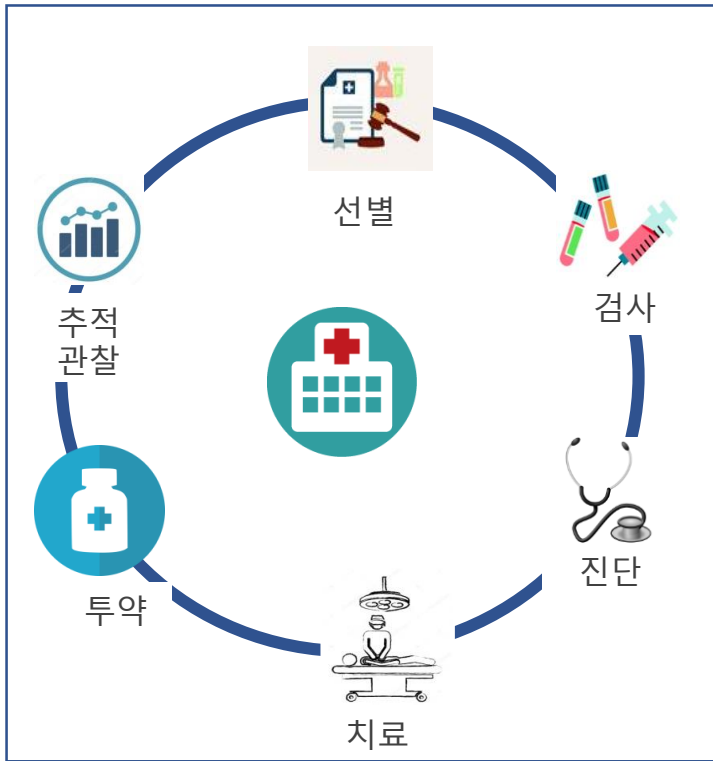




## 미래 데이터 생산 · 구축 범위

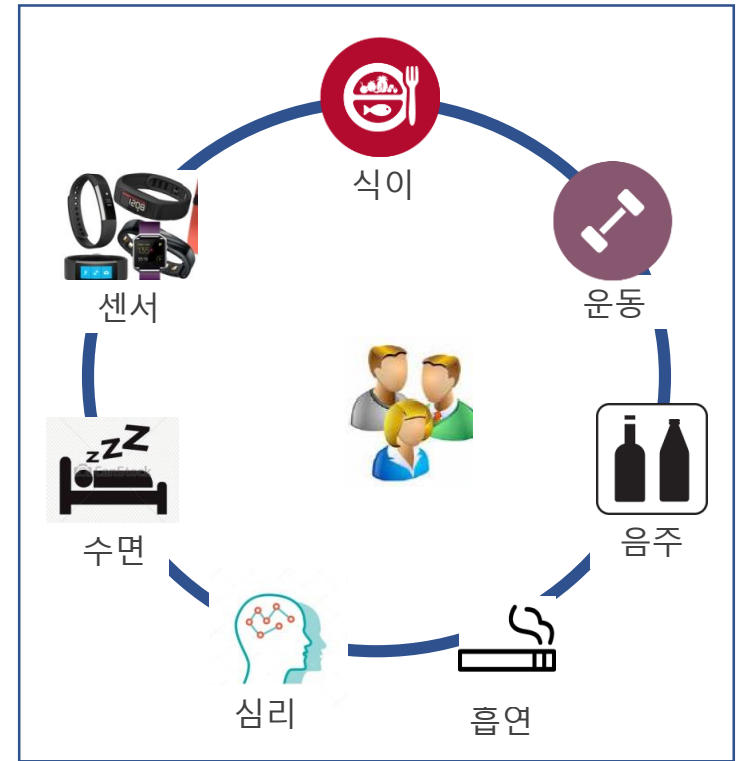
미래 보건의료 빅데이터는 병원 진료시 생성된 데이터와 개인 일상생활 생성 데이터를 통합하여 국가 보건연구, 인공지능, 신약개발 및 개인에 최적화된 건강관리 서비스에 활용함

### 병원 중심의 보건의료 빅데이터



공유

### 개인 중심의 보건의료 빅데이터



국가 보건 연구

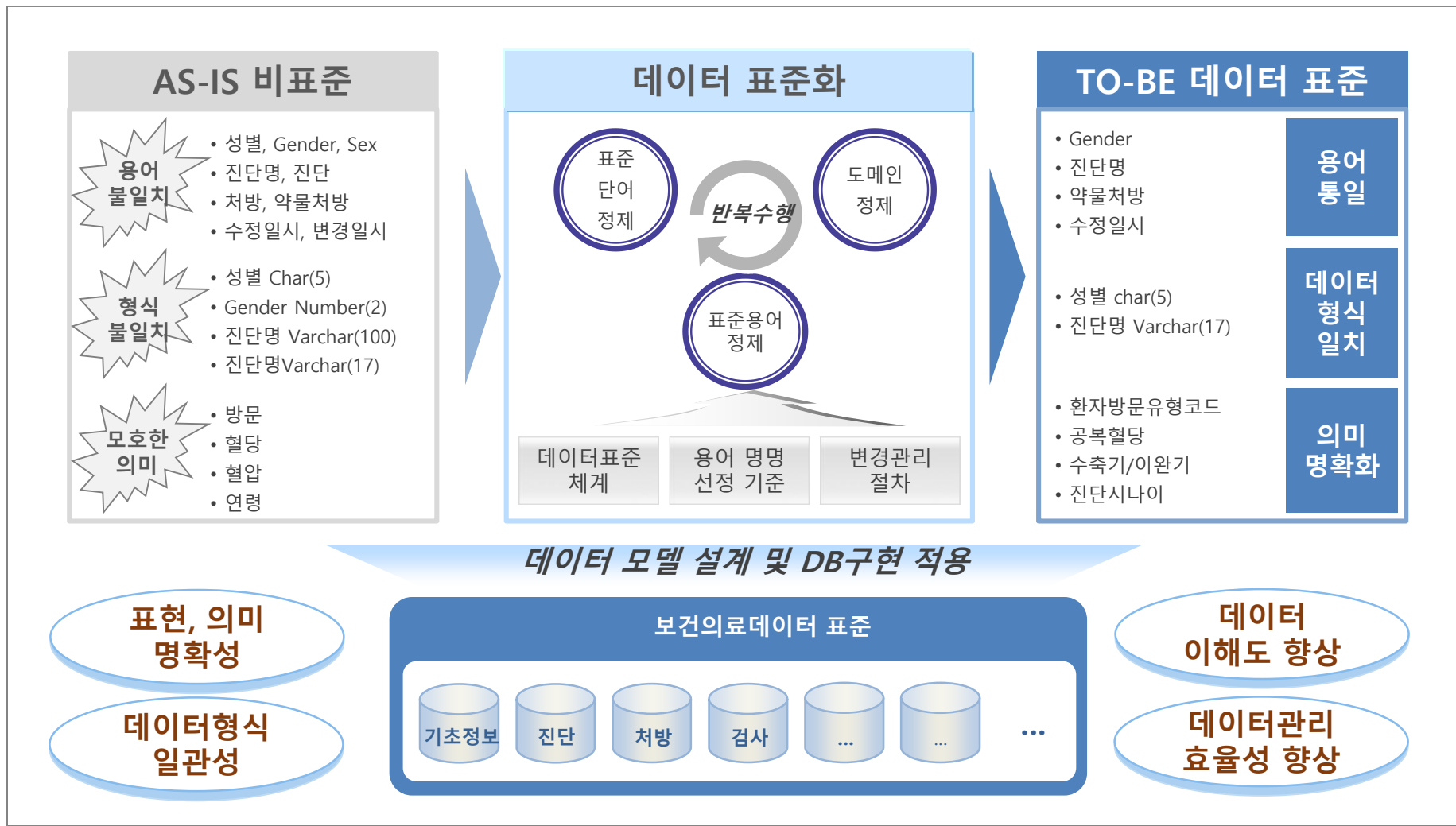
인공지능

신약개발

건강관리 서비스

# 데이터 가공처리 및 품질확보 (표준 개발)

데이터의 표현과 형식을 일관되고 명확하게 정의하여 데이터의 식별성과 이해도를 향상시킬 수 있도록 데이터 표준을 개발하고, 이를 유지할 수 있는 보건의료 데이터 표준 개발



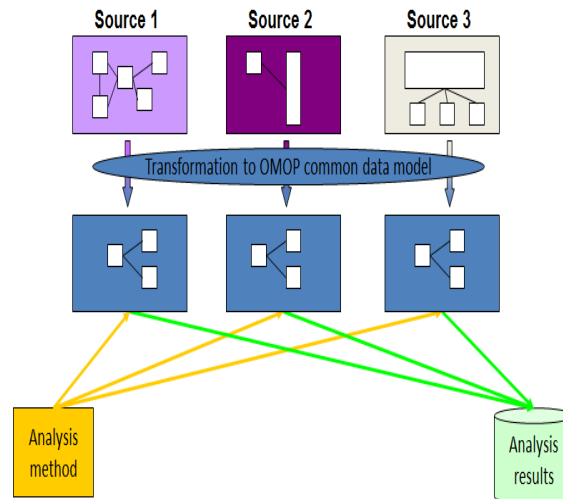
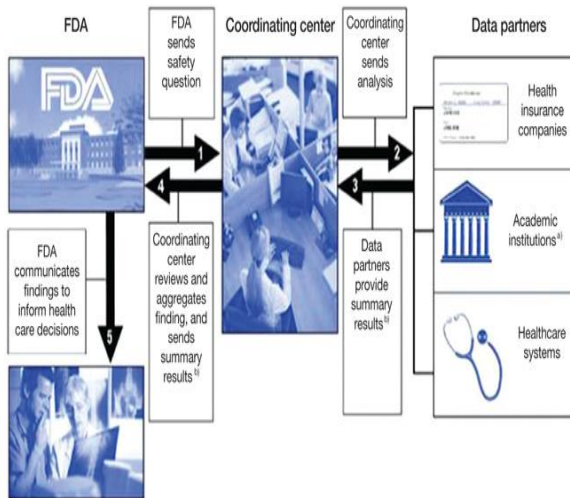
## 데이터 가공처리 및 품질확보 (표준 변환 예시)

글로벌 임상연구를 위해 진단, 약물, 검사, 처치에 대한 데이터를 국제 표준으로 현업 부서에서 매핑하며, 표준 매핑 기준으로 CDM 가공처리는 ETL 개발자가 담당

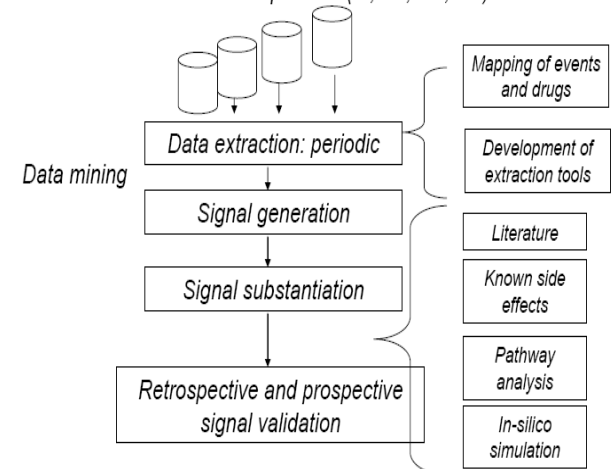
	비표준	표준화	국제 코드 매핑	국제 표준
환자 정보	남자, 남성, M, Male	M	8507	CDM
방문 정보	입원, 입원 방문, In patient	I	9201	CDM
진단 정보	알코올 중독, 알코올 사용장애, Alcohol Abuse	Alcohol Abuse	F101	ICD-10
약물 정보	졸피뎀, Zolpidem, DZPD10	Zolpidem	854873	RxNorm
검사 정보	간기능 검사, AST, Aspartate Transaminase	AST	1920-8	LOINC

# 보건의료정보 표준 분석

공통데이터모델(Common Data Model, CDM) 기반의 후향적 데이터베이스를 구축하고 여러 파트너 병원들이 분산형으로 분석 결과를 공유하는 연구 네트워크 인프라 확산






Medical databases: 30 Million persons (IT, NL, UK, DK)



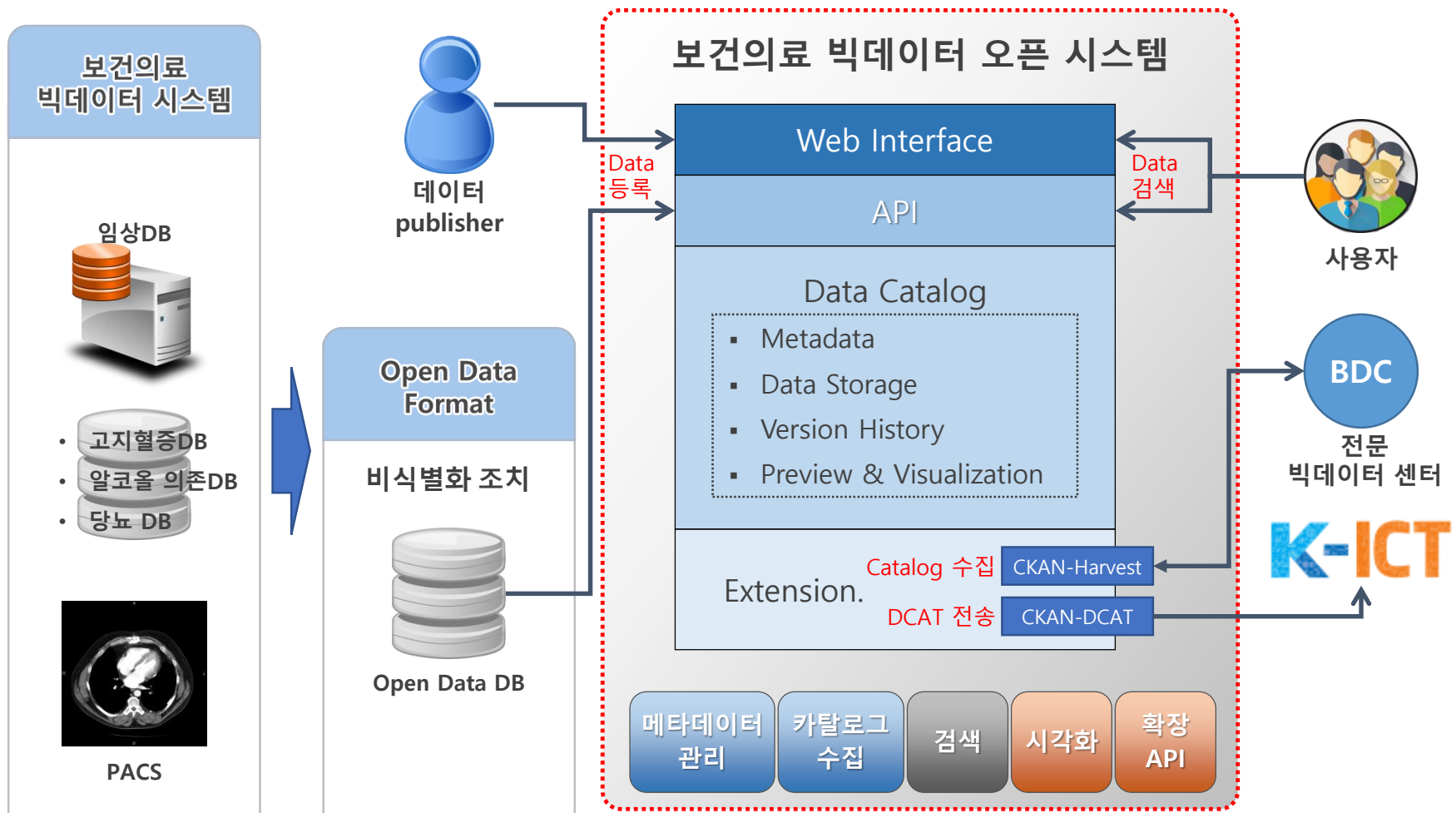
# 보건의료정보 표준 분석

공통데이터모델(Common Data Model, CDM) 활용

				
DB 형태		8개의 분산 DB	다기관 분산 DB	분산 DB
보안규칙		-	HIPPA Security Rule	HIPPA Security Rule
데이터 형태		EHR, Standardized Discharge Codes, Dispensation, Laboratory, Mortality, Prescription	Claims, Outpatient and inpatient EMR, Registries	Administrative claims, EHR
CDM 테이블		-	8	18
용어	Event	ICD-9-CM, ICPC, RCD	ICD-9-CM	ICD-9-CM
	Drug	ATC, BNF	NDC	ATC
	Prescription	-	-	SNOMED-CT
	Procedure	-	CPT	-
용어매핑시스템		UMLS	-	UMLS

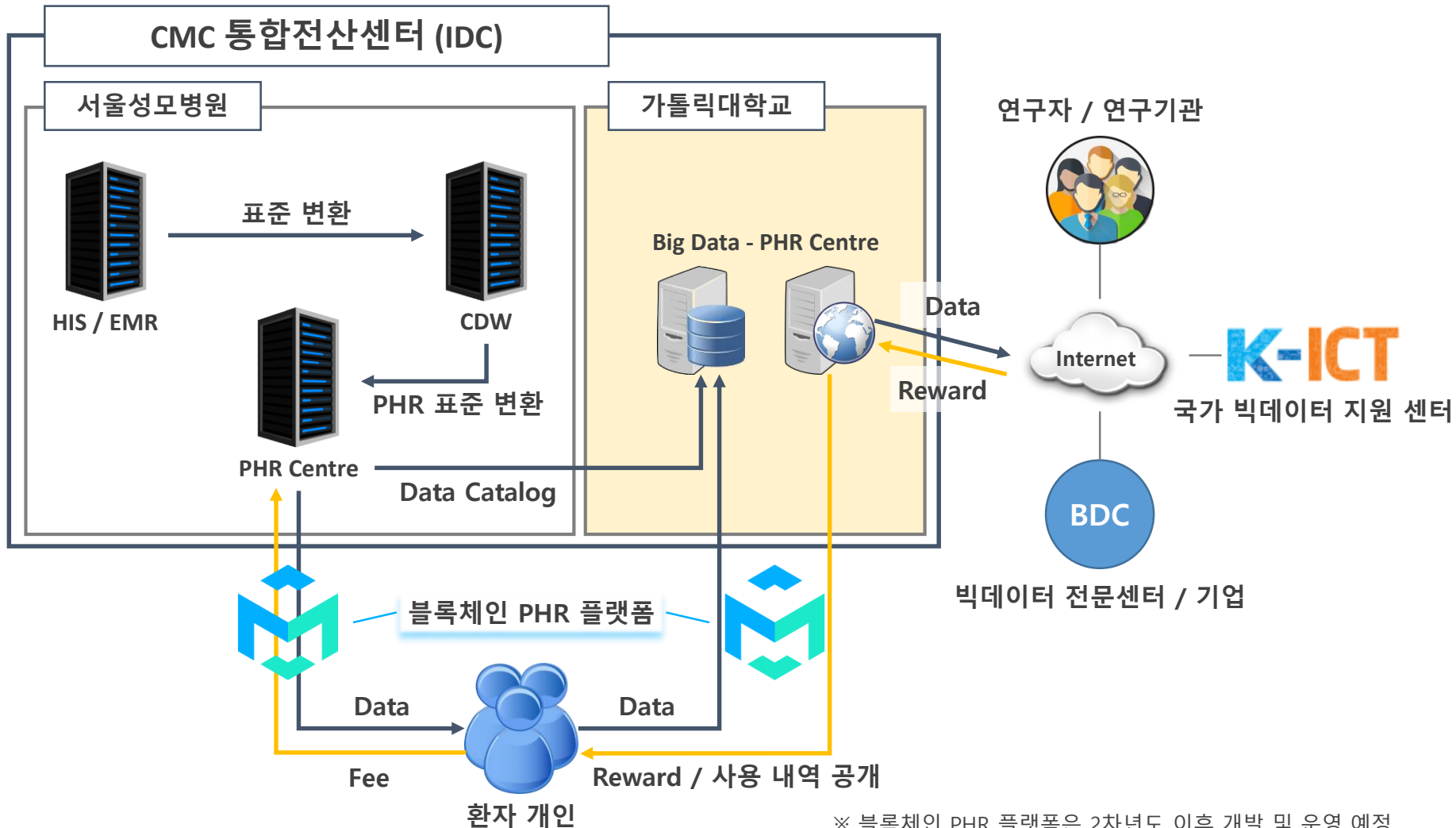
# CKAN 오픈 데이터 플랫폼

CKAN 기반의 보건의료 오픈 데이터 플랫폼을 구축하여 타 분야 오픈 데이터 카탈로그를 수집(CKAN-Harvest)하며 K-ICT 센터에 보건의료 오픈 데이터 카탈로그 전송(CKAN-DCAT)



## PHR 데이터의 개인 중심 유통을 위한 인프라

기관 중심 데이터 유통 채널 이외에 개인 중심 데이터 유통 채널 확보 및 개인 동의에 기반한 실명 데이터 사용을 위해 블록체인 PHR 플랫폼 및 활용 체계 구축



# 데이터 개방 – 법적 이슈

보건의료 분야 진료 데이터는 현행법에 비식별화 후 데이터 열람 및 반출에 관한 명확한 법 규정이 없으며, 비식별화 가이드라인도 법률이 아니기 때문에 제도적 사각지대에 있음

## 개인정보보호법

- 비식별화 처리시 동의 없이 사용 가능하나, 다른 정보와 쉽게 결합하여 재식별 가능할 경우 개인정보로 간주하여 명시적 동의 없이 사용 불가능
- 제23조(민감정보의 처리 제한) ①개인정보처리자는 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, **건강**, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보로서 대통령령으로 정하는 정보(이하 "민감정보"라 한다)를 처리하여서는 아니 된다. → 보건의료(건강) 정보는 민감정보로 간주될 뿐 아니라 구체적 정의도 없는데, 산업적 빅데이터 처리를 위해 사용하는 것이 **예외조항에 해당되지 않음**

## 의료법

- 진료기록의 열람이나 송부는 **본인의 동의가 필수적**이며, 의식이 없거나 응급환자인 경우 또는 환자의 보호자가 없어 동의를 받을 수 없는 경우에만 예외적으로 동의 없이 송부 또는 전송 가능
- 진료 데이터를 산업적 목적으로 사용하는 것에 관한 규정이 없으며, 비식별화된 진료기록이 법적인 진료기록인지 불분명

## 생명윤리 및 안전에 관한 법률 (생명윤리법)

- 개인의 서면 동의를 받고 기관위원회의 심의와 익명화** 절차를 거친 후 제3자 제공 가능
- 익명화 조치를 취함에도 불구하고 개인의 서면 동의와 기관위원회의 심의를 모두 받아야 한다면, 현행 개인정보보호법 테두리 안에서도 개인의 서면 동의를 통해 충분히 제3자 제공 가능하므로 무의미한 조항

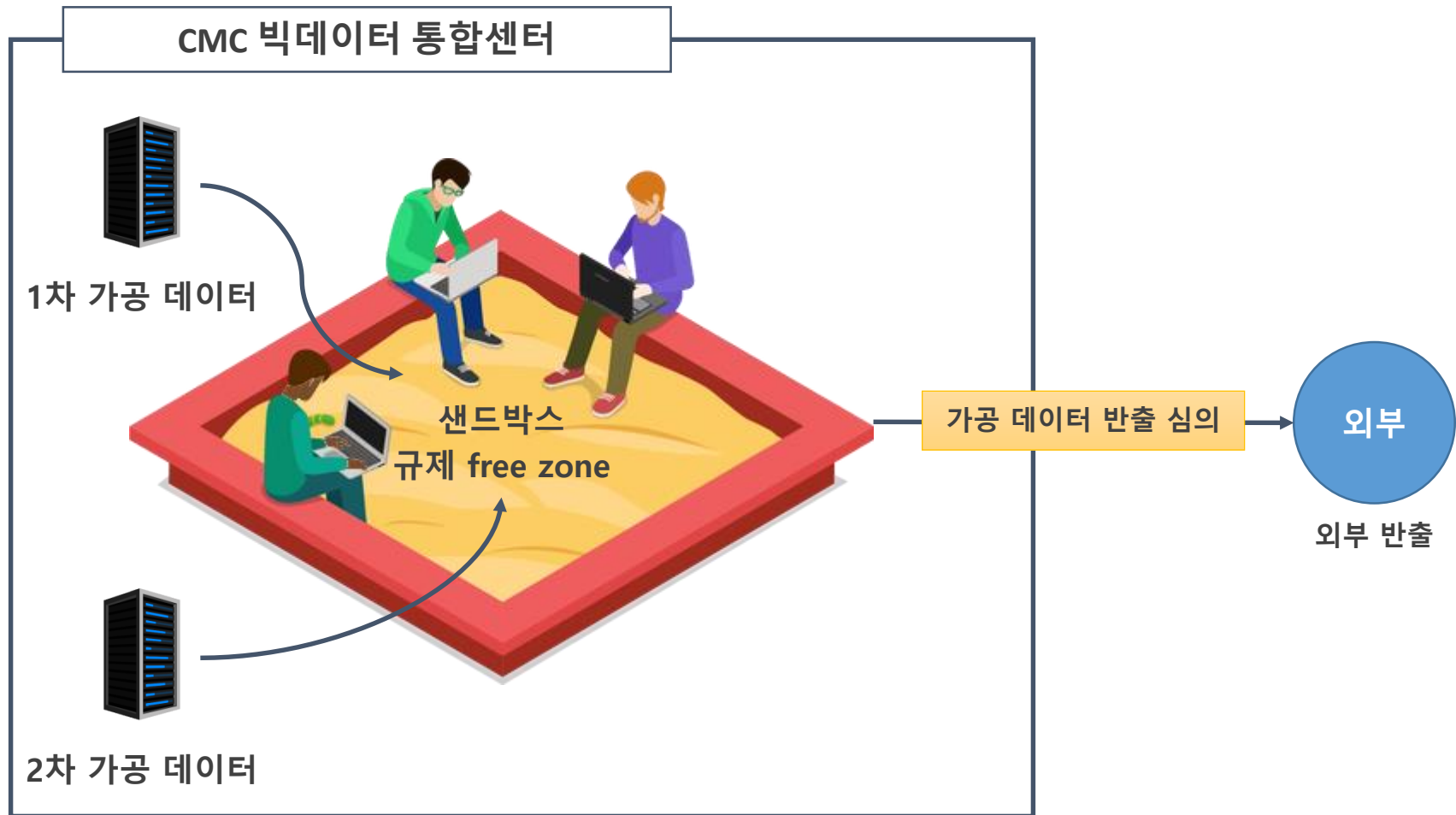
## 개인정보 비식별 조치 가이드라인

- 적용해 볼 수 있는 일부 지침을 주고 있으나 보건의료 데이터에 대해 특화된 지침이 아니라 적용에 한계가 있음
- K-익명성 평가 등 가이드라인이 제시하는 수준을 모두 만족시키면 **산업적으로 의미 없는 데이터** 산출
- 법령이 아니므로 가이드라인을 **지켜도 법적 분쟁 가능성**



## 샌드박스

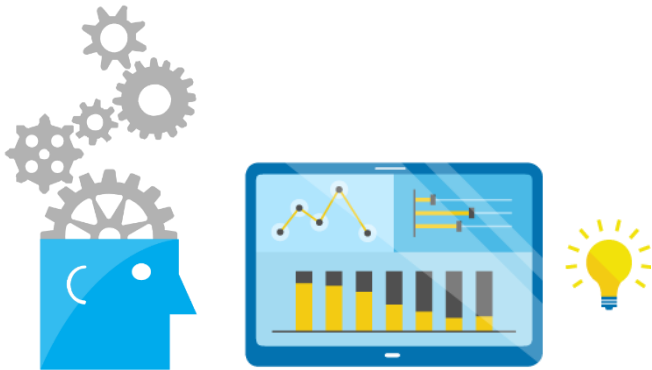
빅데이터 전문센터에 샌드박스(sandbox) 개념을 도입, 내·외부 이용자가 센터 내방시 최소한의 제한으로 보유 데이터 접근 및 분석 등 활용 업무를 수행할 수 있도록 정책 및 공간 마련 추진



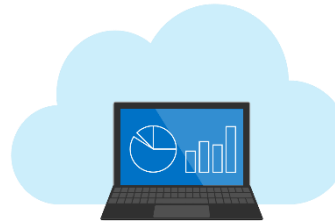
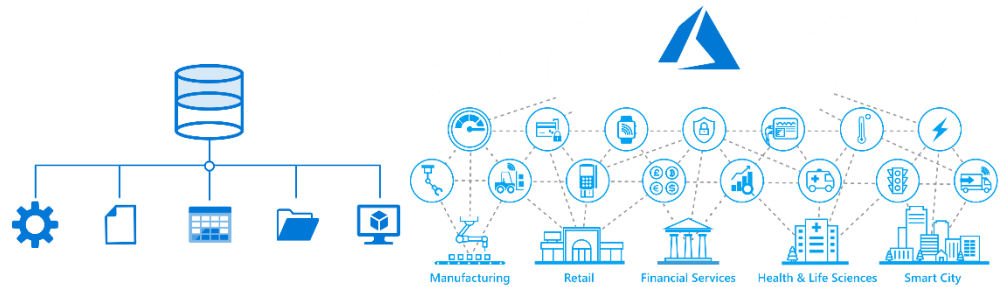
※ 샌드박스 시스템은 2차년도 이후 개발 및 운영 예정

## 교육 프로그램

가톨릭대학교 및 K-ICT 빅데이터 센터의 인프라 및 인적 자원을 적극 활용, 정규 교과 과정을 제외한 대부분의 교육 과정을 개방형, 실무 중심형 분석 실습 과정으로 기획 및 제공



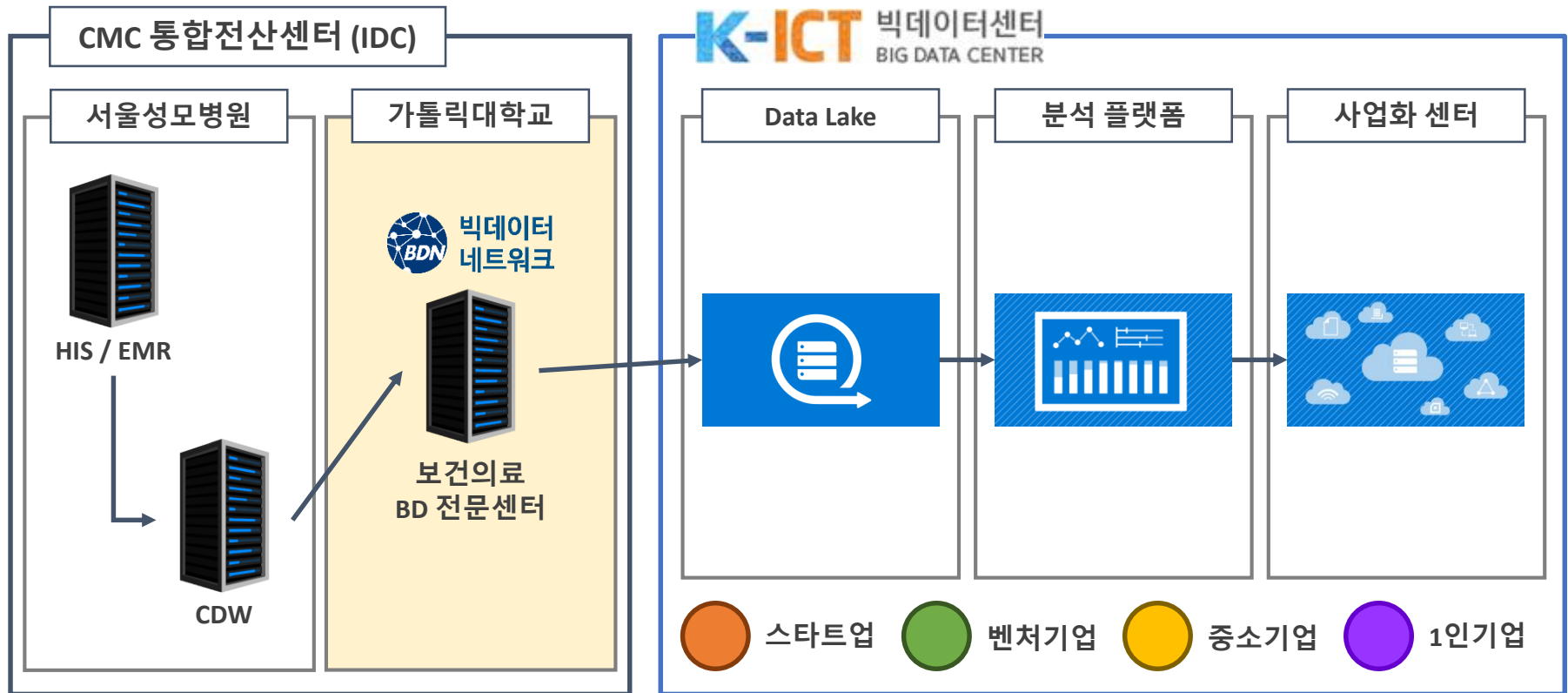
- 빅데이터 분석 워크숍
- 임상 빅데이터 정규 교과
- 분석 / 아이디어 경진대회 (내부)



- 데이터큐브 교육실습 콘텐츠
- 분석/아이디어 경진대회 (공개)
- 이종분야 결합 데이터 활용 경진대회

## 스타트업, 중소기업 지원

K-ICT 빅데이터 센터의 인프라 및 지원 프로그램을 적극 활용, 스타트업 및 중소기업이 가톨릭대학교 생산 보건의료 데이터를 활용하여 사업화까지 진행 가능하도록 지원



## 동종 · 이종 산업 데이터 연계 · 결합

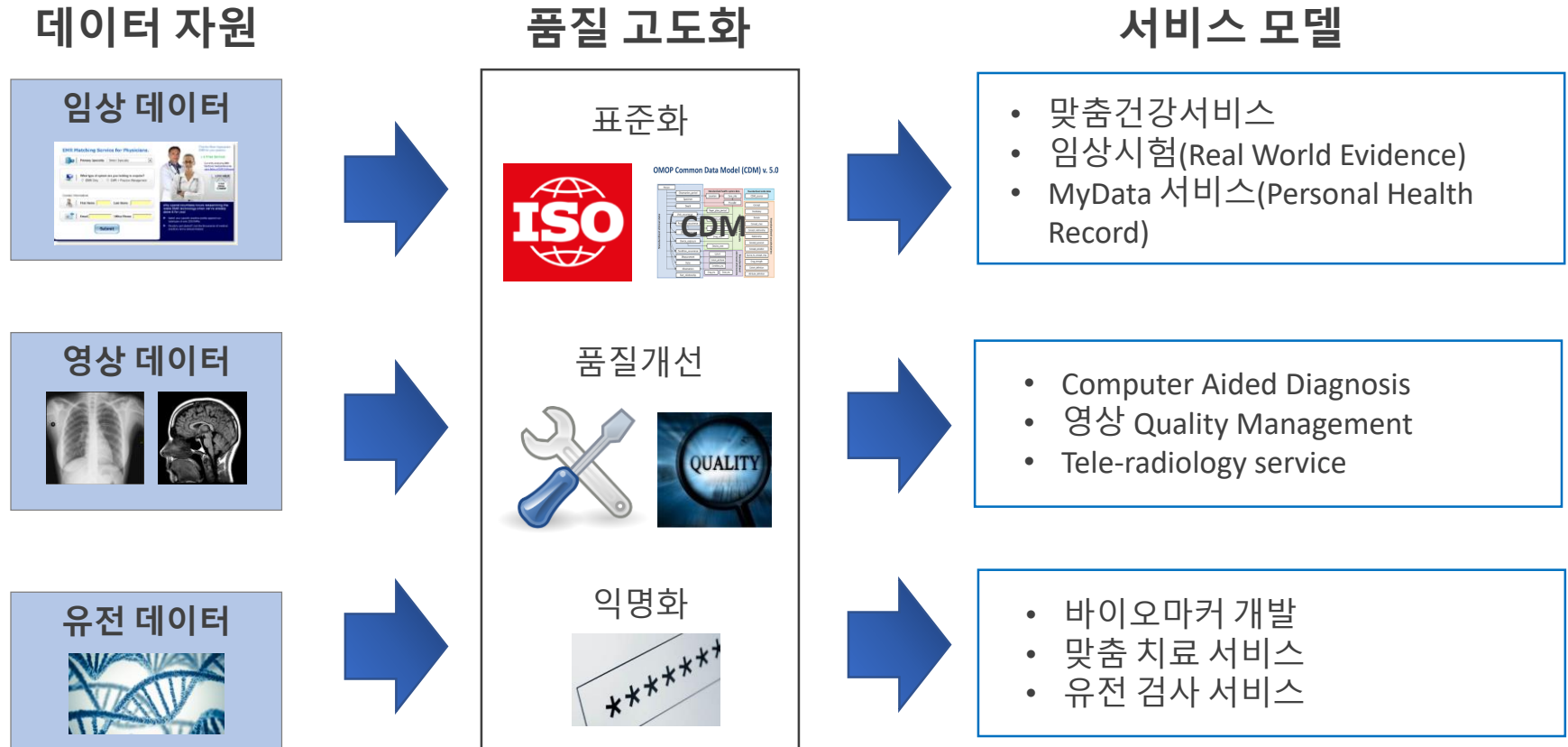
100여개로 확대될 빅데이터 전문센터들 뿐 아니라 외부 기관 · 기업과도 협업, 동종 및 이종 산업 데이터를 CMC 산하 기관 데이터와 연계 · 결합하여 새로운 산업적 가치 및 연구 성과 창출



※ 협업 대상 기관은 추후 사업 추진 상황에 따라 변경될 수 있음

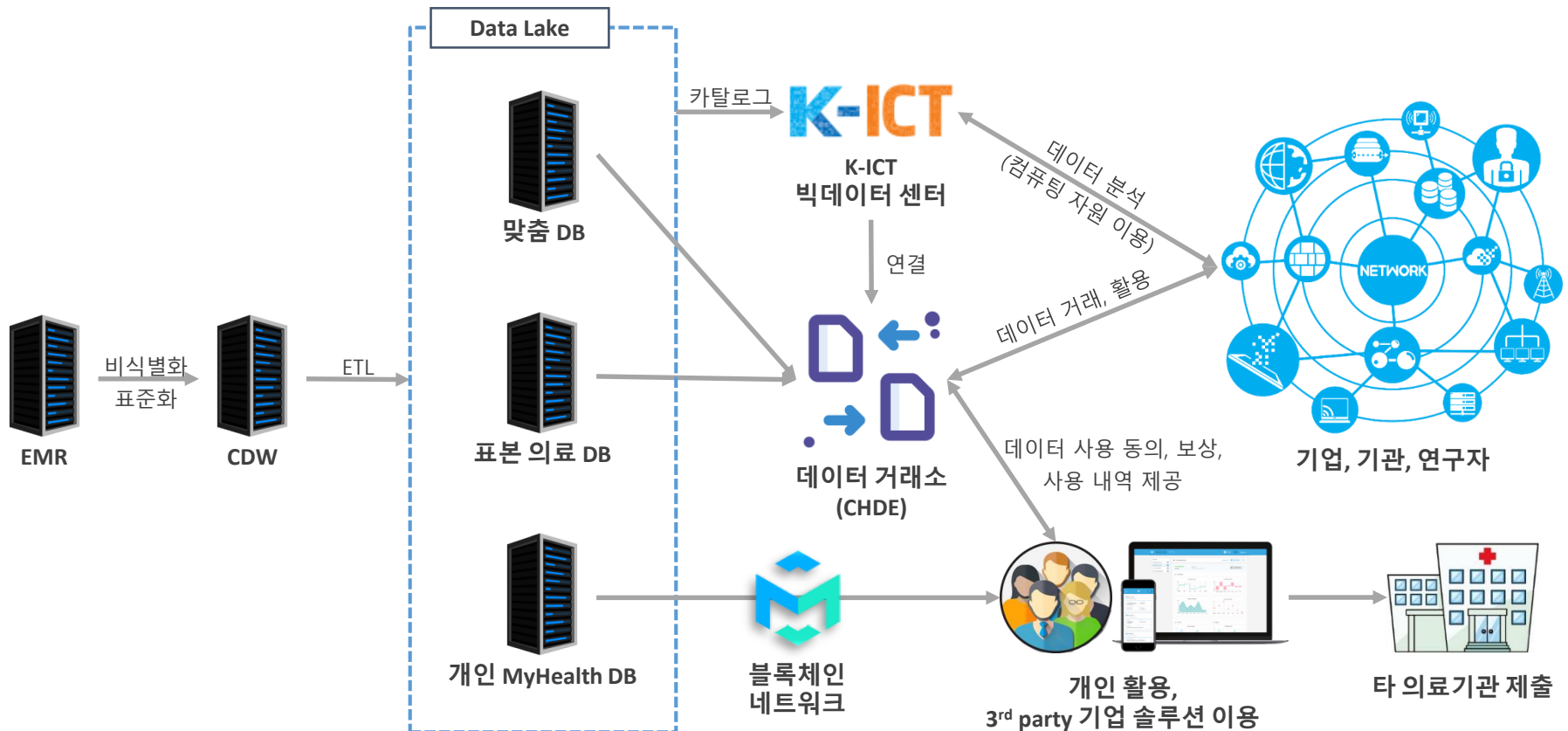
# 가톨릭 빅데이터 통합센터 서비스 모델 예시

가톨릭 중앙의료원 및 산하 기관들이 보유한 다양한 데이터 자원을 연구자 및 시장의 필요에 맞게 가공, 이전에 불가능했던 연구 및 혁신적 서비스 출현에 기여



# 미래 데이터 흐름 개념도

추후 보건의료 데이터 거래소(CHDE)를 설립, 가톨릭 중앙의료원 뿐 아니라 국내외 보건의료 기관에서 생산된 데이터를 유통할 수 있는 시스템 구축 및 데이터 기반 경제 확립



※ CHDE: Catholic Healthcare Data Exchange (가톨릭 보건의료 데이터 거래소)



# 빅데이터 네트워크



가톨릭대학교  
가톨릭중앙의료원

가톨릭빅데이터통합센터  
Catholic Big Data Integration Center

## 감사합니다