2020 IT 21
Global Conference
Digital New Deal
Technology Essentials
디지털 뉴딜 기술 핵심

Session 2-1

헬스케어 AI & Big Data

신수용 교수 (성균관대학교)

[요약문]

최근 들어 데이터 3법이 통과되면서 의료AI, 보다 폭넓게는 디지털헬스에 대한 관심이 더욱 증가하고 있다. 이런 국내 상황에서 의료분야에 익숙하지 않은 청중들을 위하여 최신 연구 결과 위주가 아닌 data의 관점에서 헬스케어 빅데이터의 의미와 관련된 최신 트렌드, 그리고 국내외에서 추진되고 있는 mega project들에 대해서 소개한다. 즉, 데이터의 관점에서 헬스케어 빅데이터가 어떻게 정의되어야 하고, 다른 분야들과는 어떤 차이점들이 있는지에 대해서 설명하고자 한다. 이를 위해 헬스케어 빅데이터 프로젝트 중 가장 대표적인 미국의 All of Us를 통해 헬스케어 빅데이터 프로젝트들이 어떻게 진행되고 있는지도 상세히 소개하고자 한다.

마지막으로 의료 AI는 도구의 관점에서 몇가지 유의점들에 대해서 언급하고, 마지막으로 의료AI와 관련되어 필수적인 규제들을 정리하고자 한다.

[발표자 약력]

1998년 서울대 컴퓨터공학과 학사
2000년 서울대 컴퓨터공학부 석사
2005년 서울대 전기,컴퓨터공학부 박사
2006년~2008년 National Institute of Standards and Technology, Guest Researcher
2008년~2010년 서울대학교병원 의료정보센터 연구교수
2010년~2011년 삼성SDS Bioinformatics Lab 수석연구원
2011년~2016년 서울아산병원 의생명정보학과 연구조교수
2016년~2018년 경희대학교 컴퓨터공학과 조교수
2018년~현재 성균관대학교 디지털헬스학과 조교수
2020년~현재 삼성서울병원 빅데이터연구센터 센터장
관심분야 : 의료데이터 표준, 개인정보보호, 의료AI

Healthcare AI & Big Data
: In the perspective of Data

Soo-Yong Shin PhD

Department of Digital Health, SAIHST
SungKyunKwan University
2020. 09. 24

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

Big Data: Origin

July 1997

The Problem of Big Data

The term "big data" was used for the first time in an article by NASA researchers Michael Cox and David Ellsworth. The pair claimed that the rise of data was becoming an issue for current computer systems. This was also known as the "problem of big data". Source: Application-Controlled Demand Paging for Out-of-Core Visualization.

1997!!

1 Introduction

Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of *big data*. When data sets do not fit in main memory (*in core*), or when they do not fit even on local disk, the most common solution is to acquire more resources. This *write-a-check* algorithm

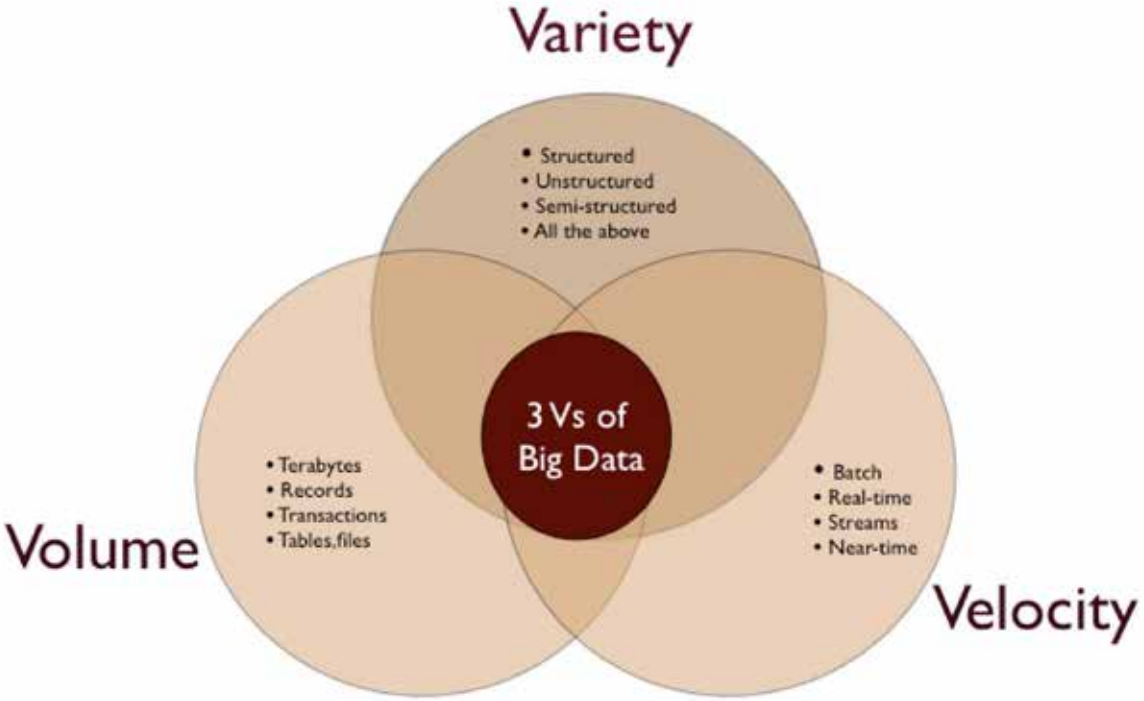
http://www.winshuttle.com/big-data-timeline/

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

3V



SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

하지만..

Application Delivery Strategies

META Group

Date: 6 February 2001

File: 949
Author: Doug Laney

3D Data Management: Controlling Data Volume, Velocity, and Variety. Current business conditions and mediums are pushing traditional data management principles to their limits, giving rise to novel, more formalized approaches.

META Trend: During 2001/02, leading enterprises will increasingly use a centralized data warehouse to define a common business vocabulary that improves internal and external collaboration. Through 2003/04, data quality and integration woes will be tempered by data profiling technologies (for generating metadata, consolidated schemas, and integration logic) and information logistics agents. By 2005/06, data, document, and knowledge management will coalesce, driven by schema-agnostic indexing strategies and portal maturity.

https://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

42 V's of Big Data

1. **Vagueness:** The meaning of found data is often very unclear, regardless of how *much* data is available.
2. **Validity:** Rigor in analysis (e.g., **Target Shuffling**) is essential for valid predictions.
3. **Valor:** In the face of big data, we must gamely tackle the big problems.
4. **Value:** Data science continues to provide ever-increasing value for users as more data becomes available and new techniques are developed.
5. **Vane:** Data science can aid decision making by pointing in the correct direction.
6. **Vanilla:** Even the simplest models, constructed with rigor, can provide value.
7. **Vantage:** Big data allows us a privileged view of complex systems.
8. **Variability:** Data science often models variable data sources. Models deployed into production can encounter especially wild data.
9. **Variety:** In data science, **we work with many data formats** (flat files, relational databases, graph networks) and varying levels of data completeness.
10. **Varifocal:** Big data and data science together allow us to see both the forest *and* the trees.
11. **Varmint:** As big data gets bigger, so can software bugs!
12. **Varnish:** How end-users interact with our work matters, and polish counts.
13. **Vastness:** With the advent of the **Internet of Things** (IoT), the "bigness" of big data [illegible]
14. **Vaticination:** Predictive analytics provides the ability to forecast [illegible] can be more or less accurate depending on rigor and [illegible] The future is pesky and never conforms to our March Madness brackets!
15. **Vault:** With many data science applications based on large and often sensitive data sets, data security is increasingly important.
16. **Veer:** With the rise of **agile data science**, we should be able to navigate the customer's needs and change directions quickly when called upon.
17. **Veil:** Data science provides the capability to peer behind the curtain and examine the effects of latent variables in the data.
18. **Velocity:** Not only is the volume of data ever increasing, but the rate of data generation (from the Internet of Things, social media, etc.) is increasing as well.
19. **Venue:** **Data science work takes place in different locations and under different arrangements**. Locally, on customer workstations, and in the cloud.
20. **Veracity:** Reproducibility is essential for accurate analysis.
21. **Verdict:** As an increasing number of people are affected by models' decisions, Veracity and Validity become ever more important.
22. **Versed:** Data scientists often need to know a little about a great many things: mathematics, statistics, programming, databases, etc.
23. **Version Control:** You're using it, right?
24. **Vet:** Data science allows us to vet our assumptions, augmenting intuition with evidence.
25. **Vexed:** Some of the excitement around data science is based on its potential to shed light on large, complicated problems.
26. **Viability:** It is difficult to build robust models, and it's harder still to build systems that will be **viable in production**.
27. **Vibrant:** A thriving data science community is vital, and it provides insights, ideas, and support in all of our endeavors.
28. **Victual:** Big data — the food that fuels data science.
29. **Viral:** How does **data spread** among other users and applications?
30. **Virtuosity:** If data scientists need to know a little about many things, we should also grow to know a lot about one thing.
31. **Viscosity:** Related to Velocity: how difficult [illegible] data to work with?
32. **Visibility:** Data science [illegible] visibility into [illegible] plex big data problems.
33. **Visualization:** [illegible] the [illegible] ways [illegible] ract with models.
34. **Vivify:** Data scienc[illegible] potent[illegible] nate all manner of decision making and business [illegible] **marke[illegible]** to **fraud detection**.
35. **Vocabulary:** Data science provides a vocabulary for addressing a variety of problems. Different [illegible] approaches tackle different problem domains, and different validation techniques harden these approaches in different applications.
36. **Vogue:** "Machine Learning" becomes "Artificial Intelligence", which becomes...?
37. **Voice:** Data science provides the ability to speak with knowledge (though not all knowledge, of course) on a diverse range of topics.
38. **Volatility:** Especially in production systems, one has to prepare for data volatility: Data that should "never" be missing suddenly disappears, numbers suddenly contain characters!
39. **Volume:** More people use data-collecting devices as more devices become internet-enabled. The volume of data is **increasing at a staggering rate**.
40. **Voodoo:** Data science and big data aren't voodoo, but how can we convince potential customers of data science's value to deliver results with real-world impact?
41. **Voyage:** May we always keep learning as we tackle the problems that data science provides.
42. **Vulpine:** **Nate Silver** would like you to be a fox, please.

How many V's?

https://www.elderresearch.com/blog/42-v-of-big-data

UNIVERSITY

Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG

SAMSUNG MEDICAL CENTER

Definition in ISO

- Extensive *datasets* — primarily in the *data* characteristics of volume, variety, velocity, and/or variability — that require a scalable technology for efficient storage, manipulation, management, and analysis
  - Note 1 to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

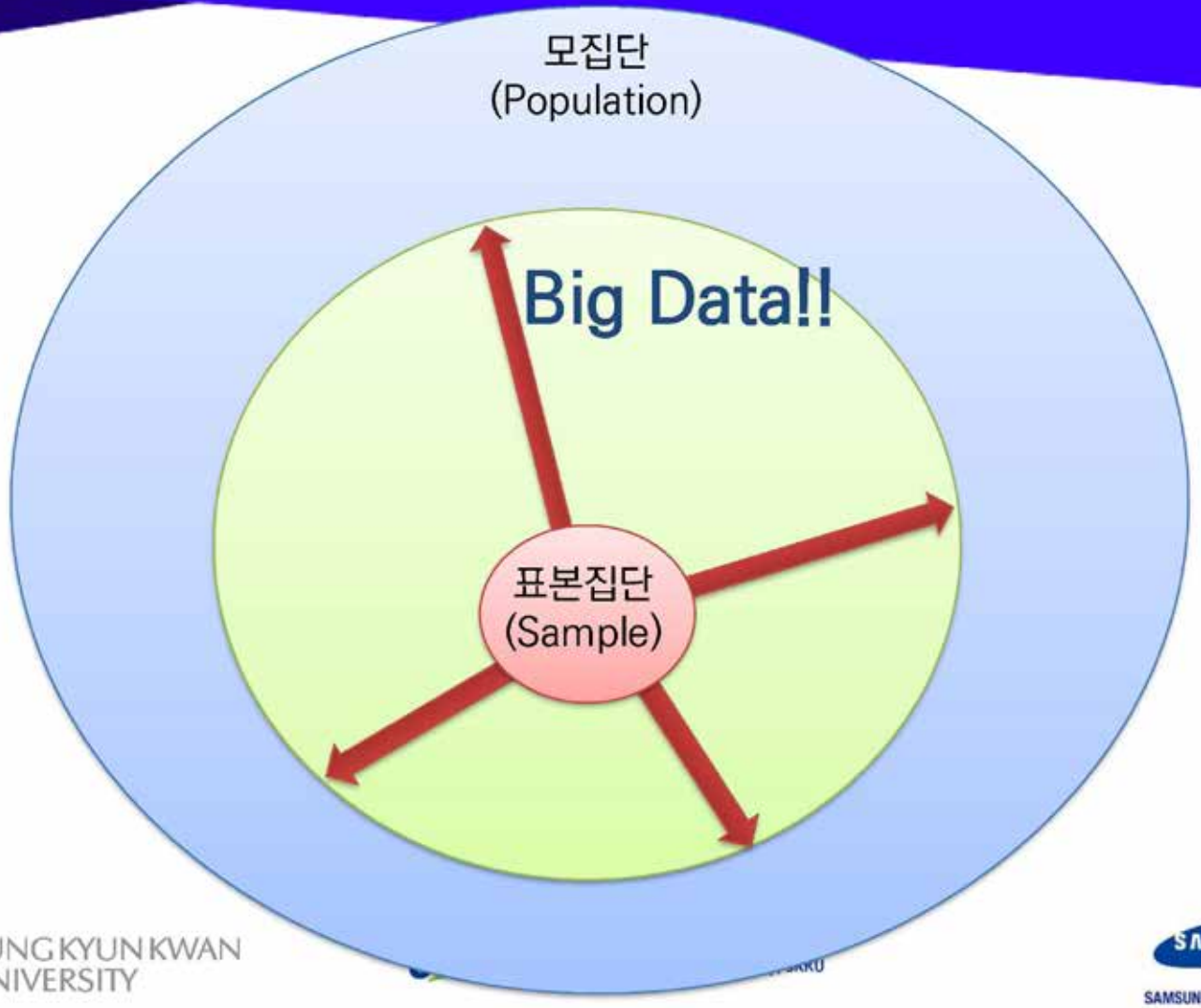
ISO/IEC 20546:2019 Information technology -- Big data -- Overview and vocabulary
https://www.iso.org/standard/68305.html

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

View point 1: Extension of existing data

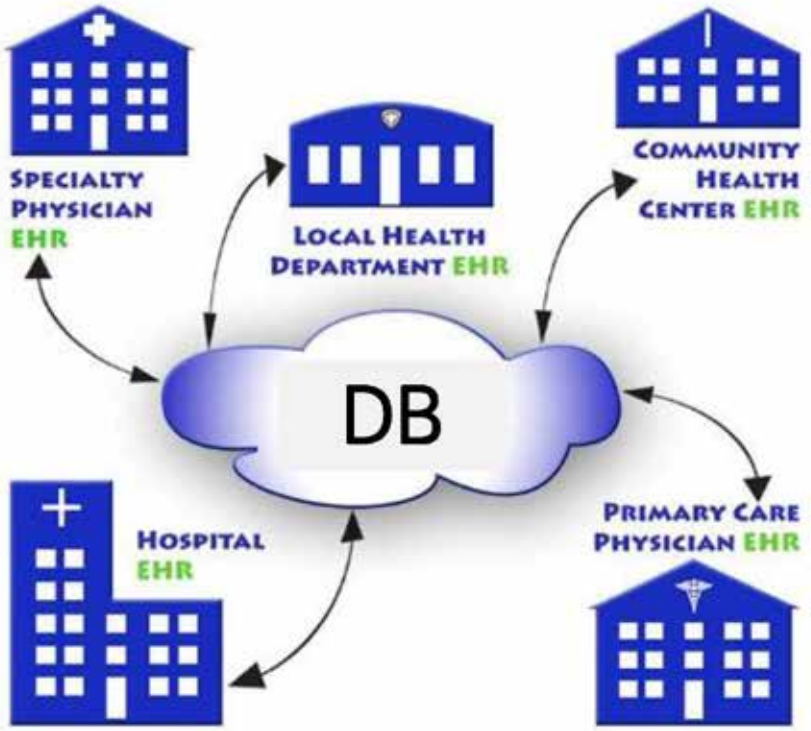


SUNG KYUN KWAN UNIVERSITY

SAMSUNG MEDICAL CENTER

Extension of existing data

- Multi-center research



SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

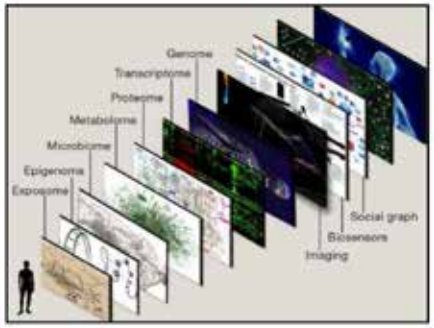

MORE DATA!
BIG DATA
Clinical Data
from multiple hospitals
or other related data
(Nation-wide claim data)
SUNG KYUN KWAN UNIVERSITY
SAIHST
Samsung Advanced Institute for Health Sciences & Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER

MORE DATA!
BIG DATA

Clinical Data
from multiple hospitals
or other related data
(Nation-wide claim data)

Genomics
Proteomics
Epigenomics
...

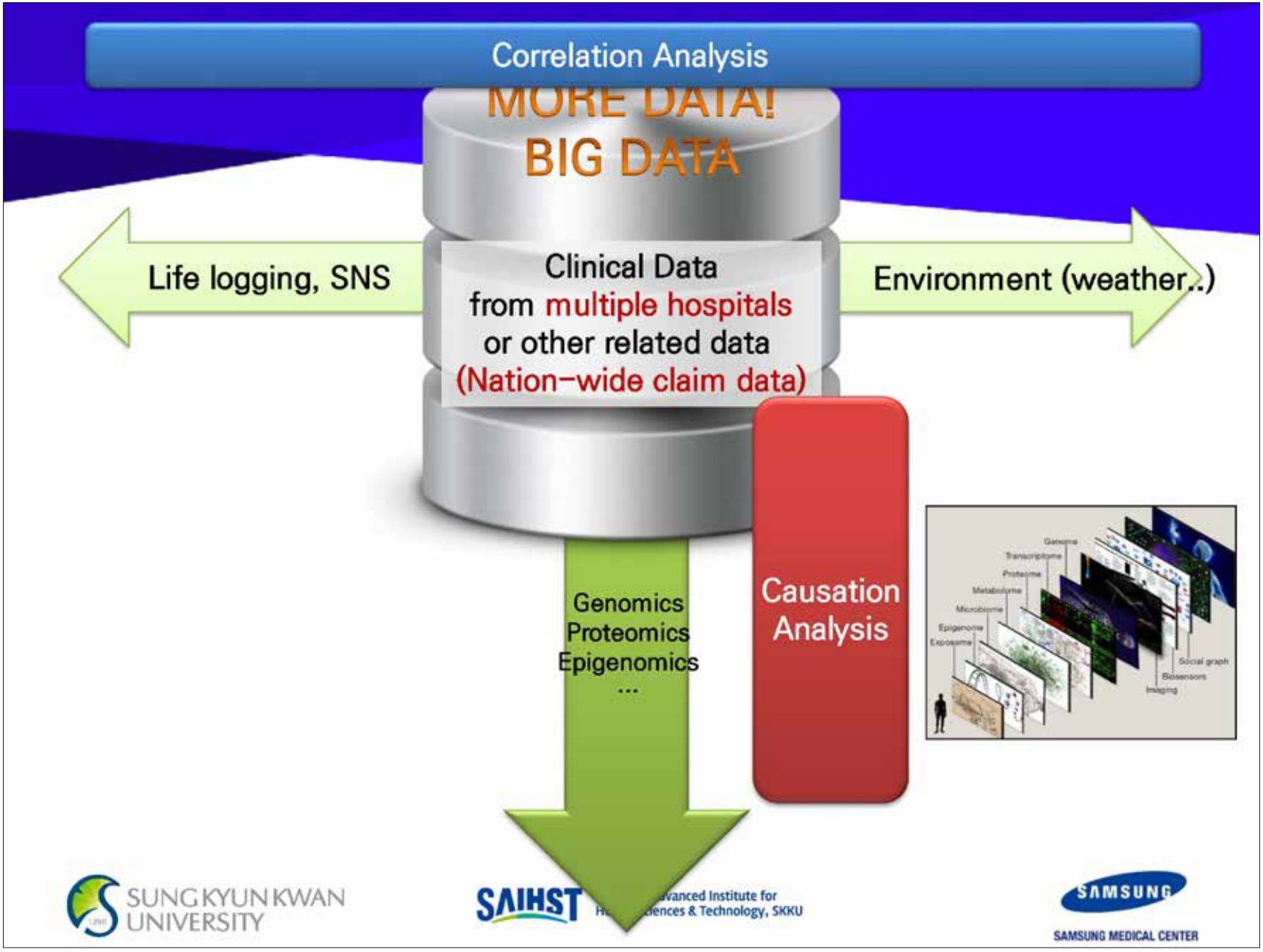
Causation Analysis



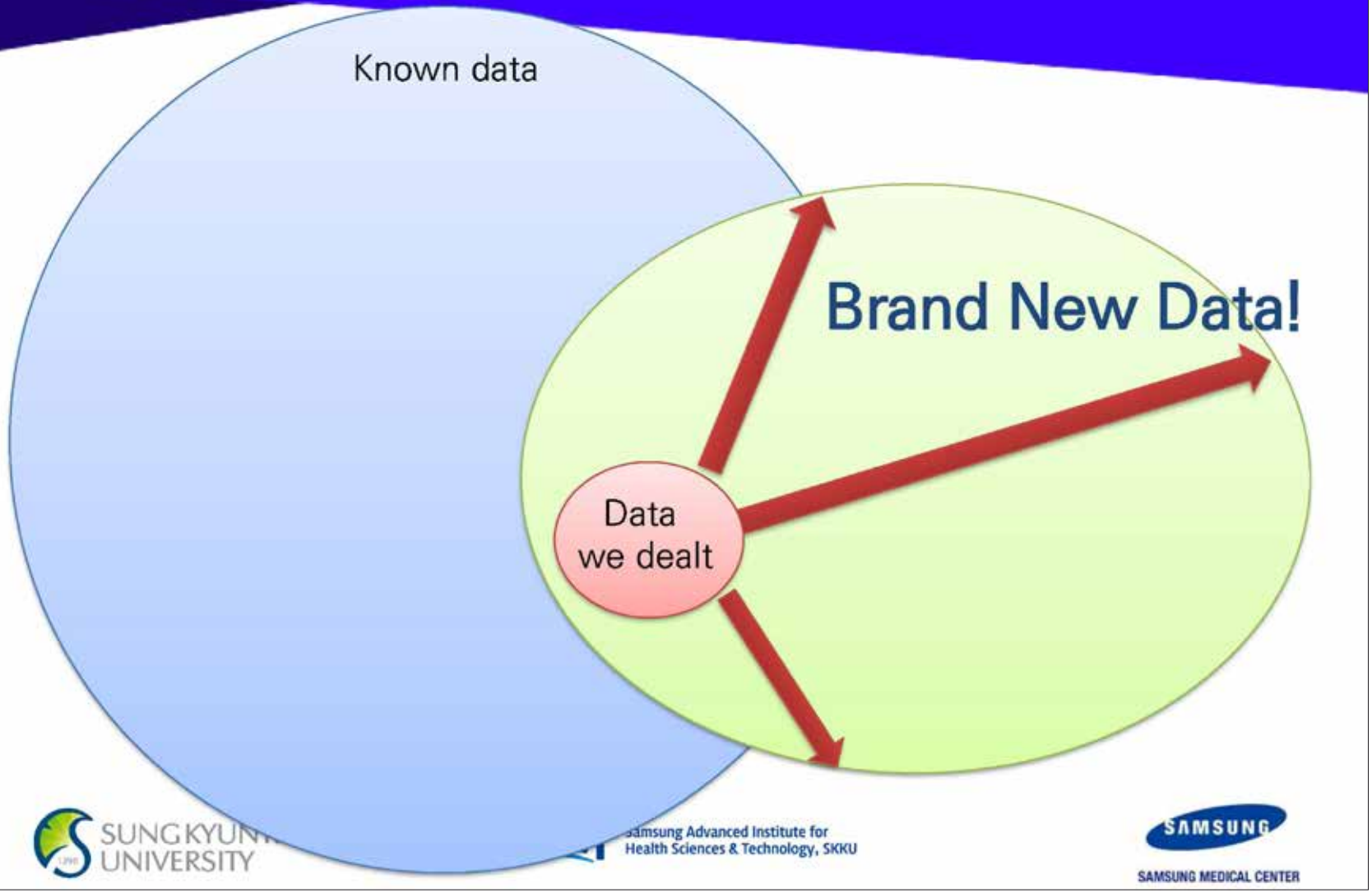
SUNG KYUN KWAN UNIVERSITY

SAIHST Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER


Correlation Analysis
MORE DATA!
BIG DATA
Life logging, SNS
Clinical Data
from multiple hospitals
or other related data
(Nation-wide claim data)
Environment (weather..)
Genomics
Proteomics
Epigenomics
...
Causation Analysis
Genome
Transcriptome
Proteome
Metabolome
Microbiome
Epigenome
Exposome
Social graph
Biosensors
Imaging
SUNG KYUN KWAN UNIVERSITY
SAIHST
Advanced Institute for
Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER

View point 2: Emerging data



SUNG KYUN UNIVERSITY

Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG

SAMSUNG MEDICAL CENTER

Digital Biomarker?! (by Rock Health)

ROCK HEALTH

WHAT IS A DIGITAL BIOMARKER?

Digital biomarkers are consumer-generated physiological and behavioral measures collected through connected digital tools that can be used to explain, influence and/or predict health-related outcomes. This excludes patient-reported measures (e.g., survey data), genetic information, and data collected through traditional medical devices and equipment.

TYPE OF DATA/MEASUREMENT	COLLECTION METHOD	BIOMARKER
Imaging	Connected digital tools	Wellness
Molecular	Medical device & equipment	Disease
Physiological	Observational	Drug
Behavioral	Molecular Assays	

http://rockhealth.com/reports/the-emerging-influence-of-digital-biomarkers-on-healthcare/

SUNGKYUNKWAN UNIVERSITY

Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

npj | Digital Medicine
www.nature.com/npjdigitalmed

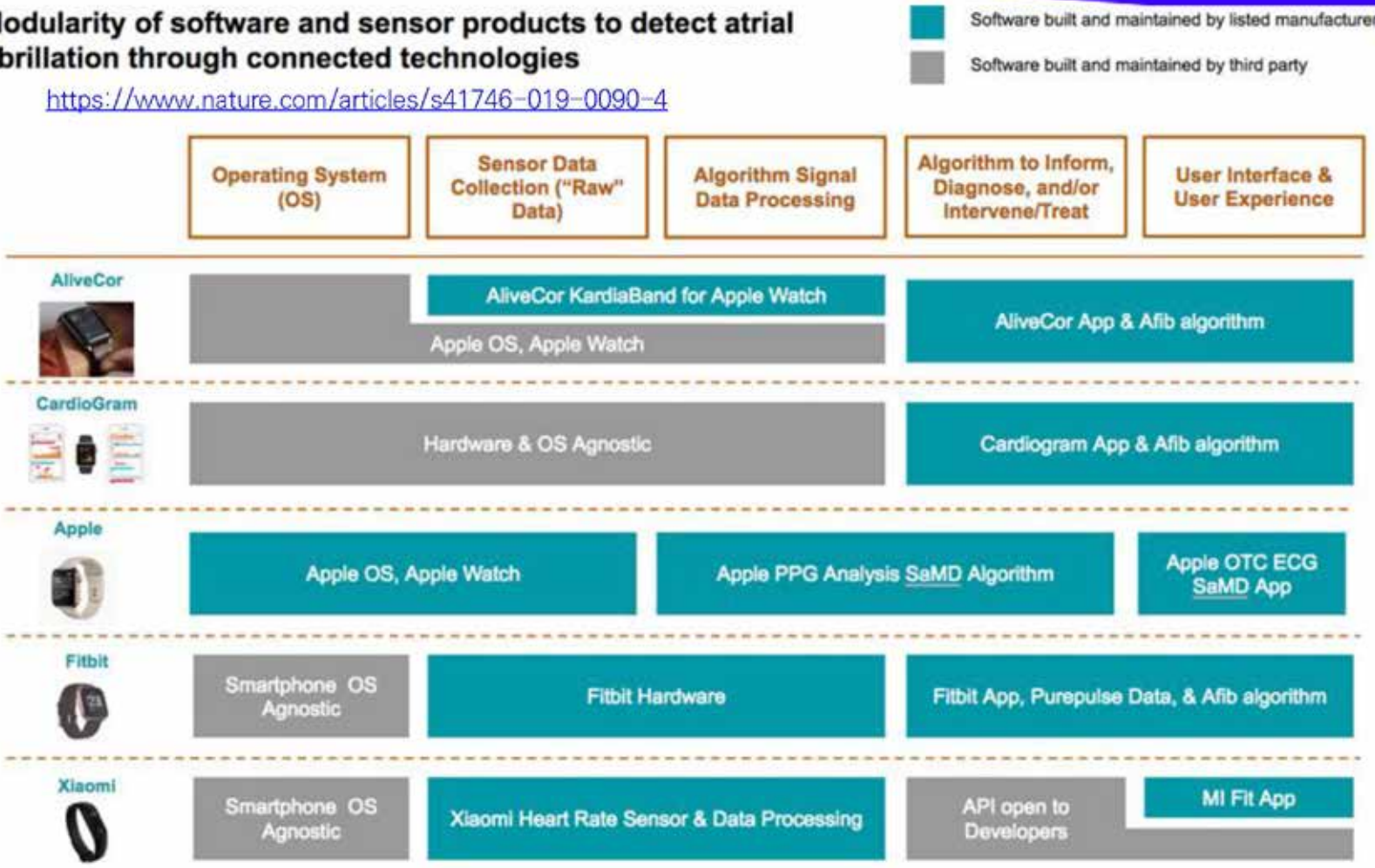
PERSPECTIVE OPEN

Developing and adopting safe and effective digital biomarkers to improve patient outcomes

Andrea Coravos[1], Sean Khozin[2] and Kenneth D. Mandl[1,3]

Modularity of software and sensor products to detect atrial fibrillation through connected technologies

https://www.nature.com/articles/s41746-019-0090-4



The Importance of Social Determinants of Health Data

https://www.ehidc.org/resources/importance-sdoh-data

LexisNexis RISK SOLUTIONS | Health Care

eHI eHEALTH INITIATIVE

SDOH

- Conditions in the environments in which people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks

SDOH to behavior change

Measuring Intervention Effectiveness

Low Interaction & Decreased Intervention - When a user has low interaction with the digital health solution, they receive decreased exposure to the intervention, there is reduced likelihood that the outcome will be achieved.

High Interaction & Ineffective Intervention - When a user has a high level of interaction with the digital solution but the intervention components are ineffective, there are diminished outcomes.

High Interaction & Effective Intervention - When a user has the right level of interaction with the digital solution and the intervention components are effective, there are increased and improved outcomes.

GOAL

SUNGK UNIVER

SAMSUNG

AMSUNG MEDICAL CENTER

SDOH to EHR

IOM Recommendation

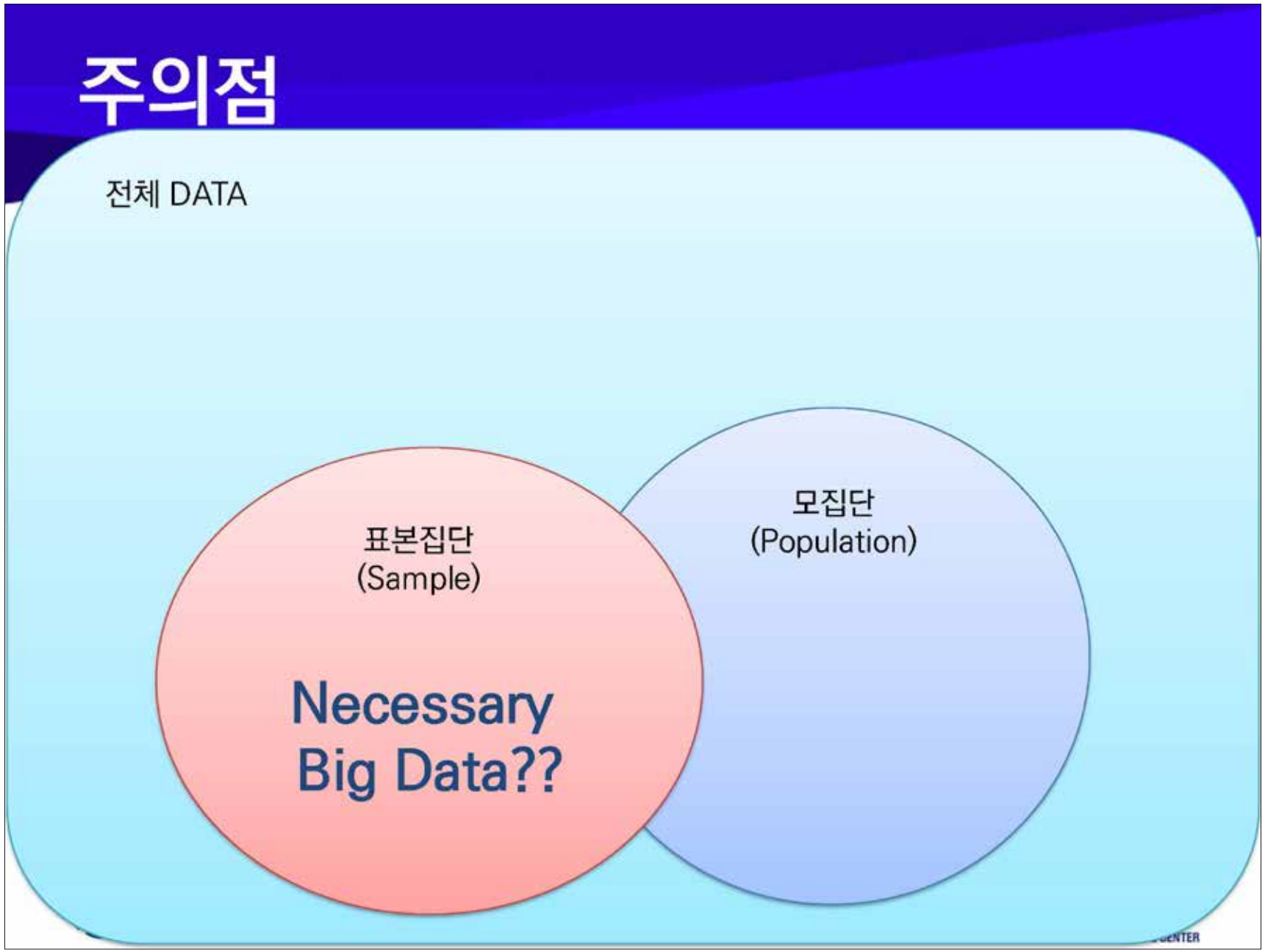
IOM Recommended Measures for EHRs[16]

Measure	# of questions	Frequency	
Alcohol Use**	3 questions	Screening & Follow Up	*Currently collected in clinical settings*
Race & Ethnicity	2 questions	At Entry	
Residential Address	1 question	Verify Every Visit	
Tobacco Use & Exposure	2 questions	Screening & Follow Up	
Census tract-median income	1 question	Update on Address Change	*Recommended Additions*
Depression**	2 questions	At Entry	
Educational Attainment**	2 questions	Screening & Follow Up	
Financial Resource Strain**	1 question	Screening & Follow Up	
Intimate Partner Violence**	4 questions	Screening & Follow Up	
Physical Activity**	2 questions	Screening & Follow Up	
Social Connections & Isolation**	4 questions	Screening & Follow Up	
Stress**	1 question	Screening & Follow Up	

***ONC Interoperability Standards Advisory identified these SDOH data elements*

SUNG KYUN KWAN UNIVERSITY

주의점




Healthcare Big Data?
SUNG KYUN KWAN UNIVERSITY
SAIHST
Samsung Advanced Institute for Health Sciences & Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER

What is Big Data?

Data Science Home / What is Big Data?

Biomedical Big Data is more than just very large data or a large number of data sources. Big Data refers to the complexity, challenges, and new opportunities presented by the combined analysis of data. In biomedical research, these data sources include the diverse, complex, disorganized, massive, and multimodal data being generated by researchers, hospitals, and mobile devices around the world.

Biomedical Big Data is diverse and complex. It includes imaging, phenotypic, molecular, exposure, health, behavioral, and many other types of data. These data could be used to discover new drugs or to determine the genetic and environmental causes of human disease.

Biomedical Big Data faces many challenges. The unwieldy amount of information, lack of organization and access to data and tools, and insufficient training in data science methods make it difficult for Big Data's full power to be harnessed.

Biomedical Big Data provides spectacular opportunities. Big Data methods allow researchers to maximize the potential of existing data and enable new directions for research. Biomedical Big Data can increase accuracy and supports the development of precision methods for healthcare.

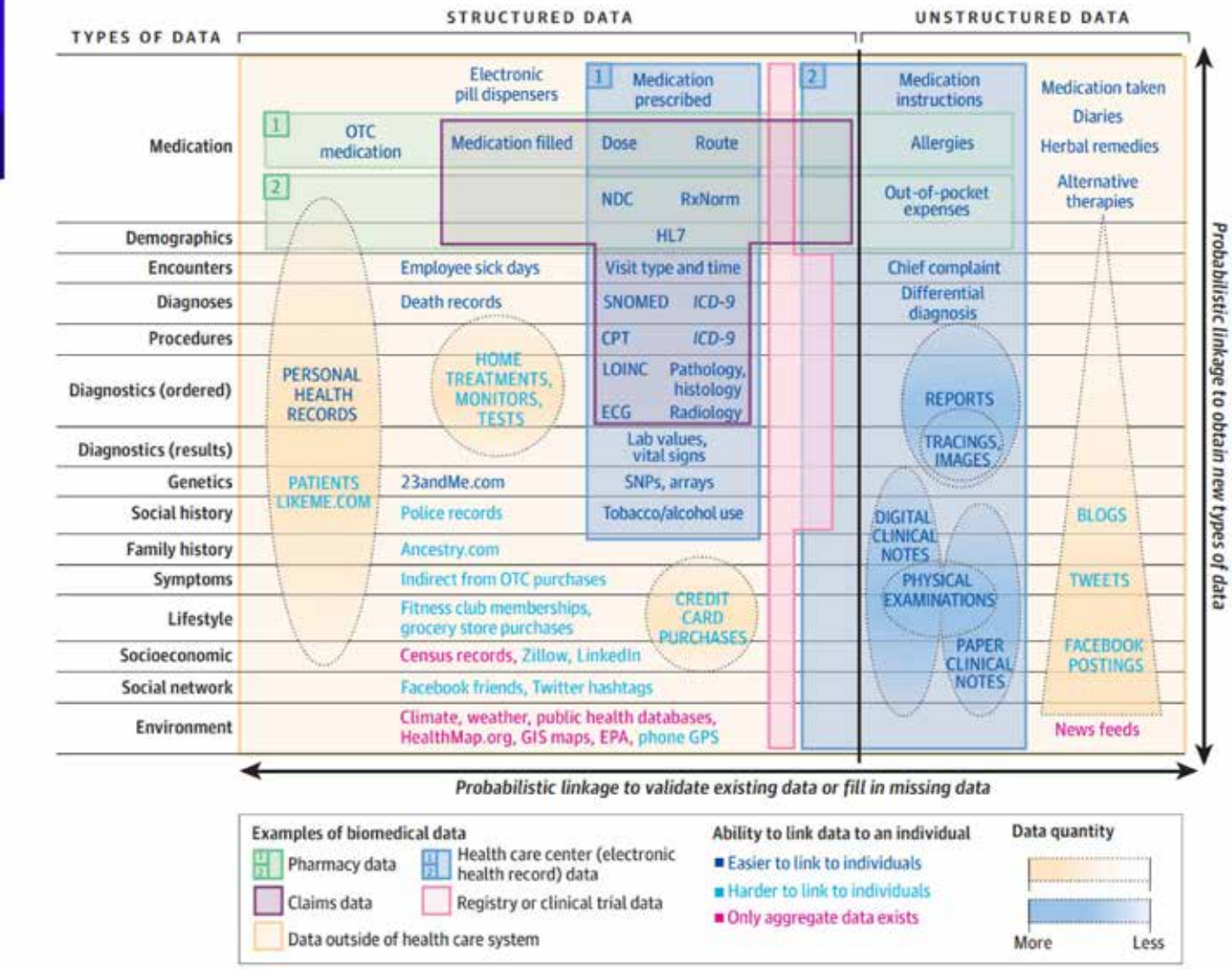
https://datascience.nih.gov/bd2k/about/what

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

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



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/24854141

의료계의 시각

- Victor Dzau, M.D., president of the Institute of Medicine.
  - when people discuss the concept of big data and healthcare **"they don't know what they're talking about."**
  - argues that big data is **"really about everything around the individual"** including social interactions and activities, and not just electronic health records.
  - not the collection of this vast amount of data

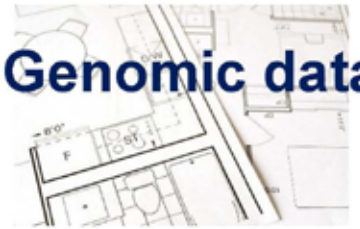
http://www.healthdatamanagement.com/news/IOM-President-Says-Big-Data-is-About-More-Than-Just-EHRs-50035-1.html

SUNGKYUNKWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

Unified View



Genomic data

Clinical Data

Patient-Generated Health Data

Social Determinants of Health

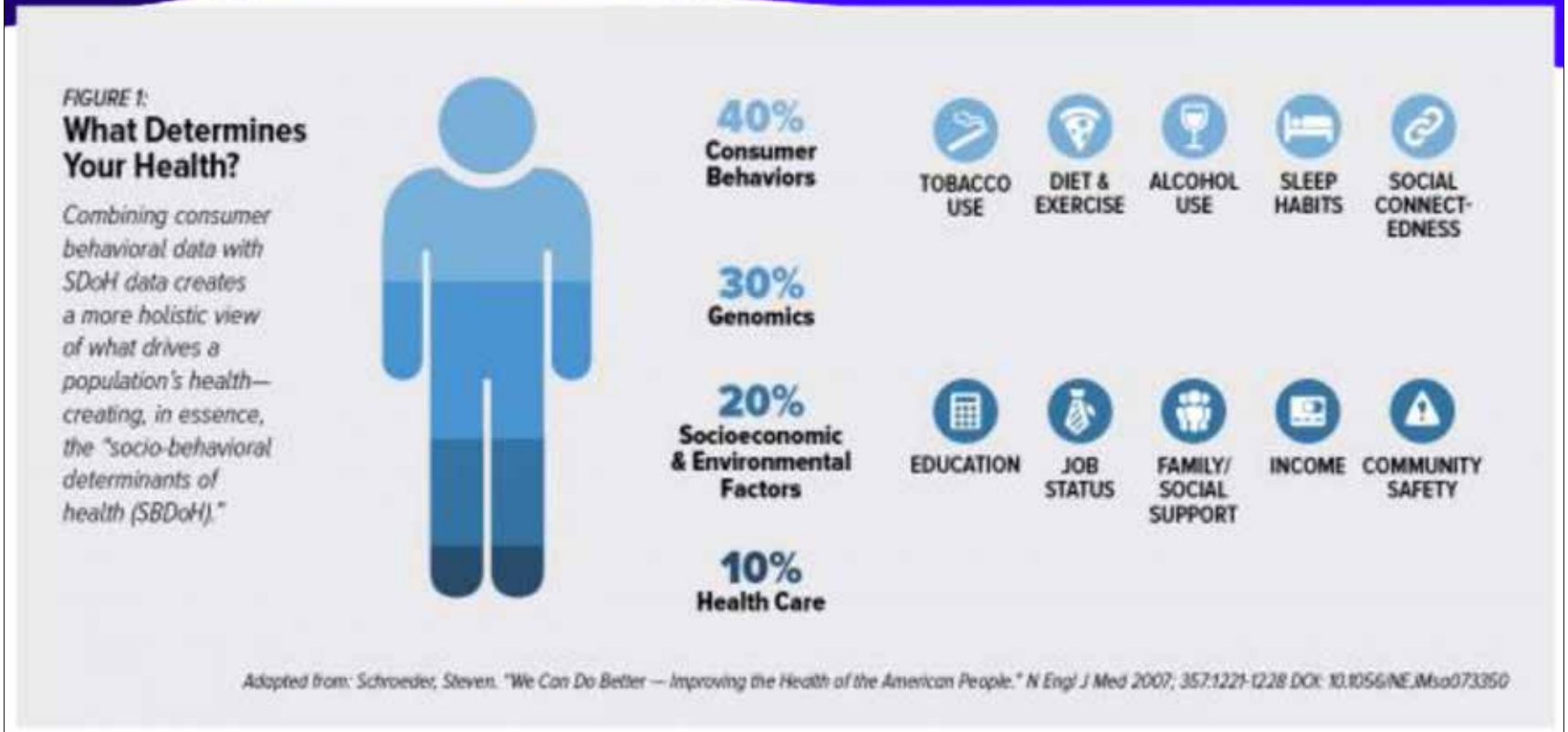
SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

https://doi.org/10.7599/hmr.2017.37.2.86

SAMSUNG MEDICAL CENTER

Determinants of Health



https://twitter.com/healthythinker/status/1095319553968623616

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

A quick guide to population health management

How can it help you?

Our health and care needs are changing: we are living longer with more multiple long-term conditions like asthma, diabetes and heart disease.

Much of this is down to lifestyle factors and where we live rather than the health and care services treating us.

Population health management helps us understand and predict future health and care needs so that we can better target support, make better use of resources and reduce health inequalities.

Which factors impact your health?

Our environment
5% housing quality and our built environment

Our behaviours
40% smoking, diet, alcohol use, poor sexual health

Healthcare
10% being able to access good quality care

Genetics
30% your genes can directly cause or increase your risk of developing a wide range of medical conditions

Social and economic circumstances
15% education, employment, income, family/ social support, community, safety

Each year lifestyle and environmental factors cost the NHS **£11 billion**

Impact of social and economic inequalities costs a further **£4.8 billion**

What is population health management?

The NHS and its public sector partners use expert analysis of data on factors like health, housing, employment and where people live as well as current medications.

They can find new insights about specific groups of people and whether they are being seen by the right professionals.

Expert medics, clinicians and public health professionals assess the data and can tailor care better to those patients.

Past public health research shows this personal care improves physical and mental health and reduces health inequalities.

Why does it matter and who is involved?

To solve wider issues impacting people's health, expert insights take a more complete snapshot of a person's needs and environment.

All our public services are involved • the NHS • the public • schools • fire service • councils • voluntary sector • housing associations • social services • police.

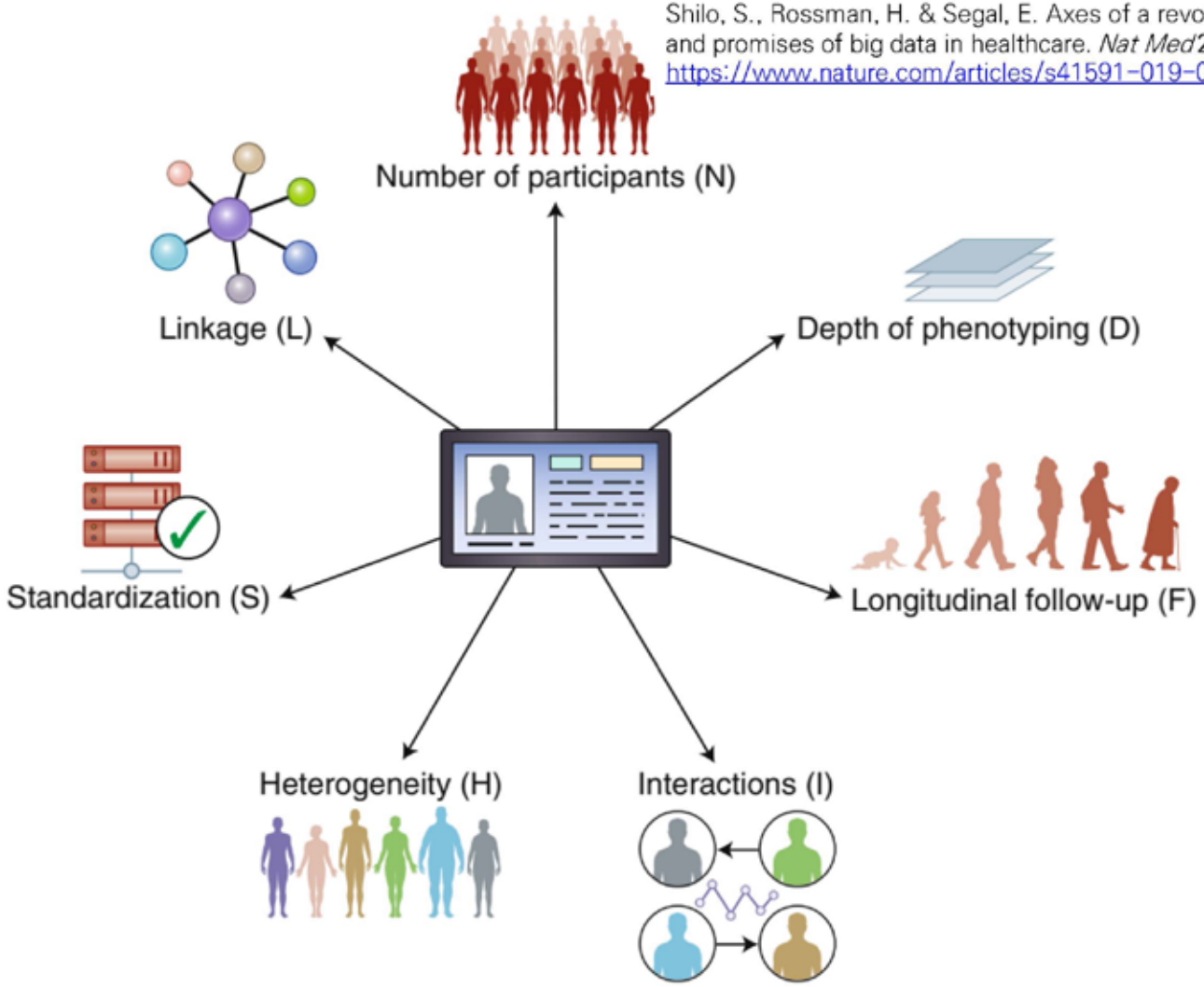
A flavour of what's been achieved so far

- In **Lancashire and South Cumbria** they used data on households with assisted bin collections to find frail patients in need of more proactive personalised care to keep them living well at home.
- In **Leeds** analysis pinpointed 80 patients with frailty at risk of further problems - they now get better personalised care to stay well, active and independent.

Population health management is a key building block in the development of integrated care systems.
To find out more visit www.england.nhs.uk/integratedcare #futureNHS #datasaveslives

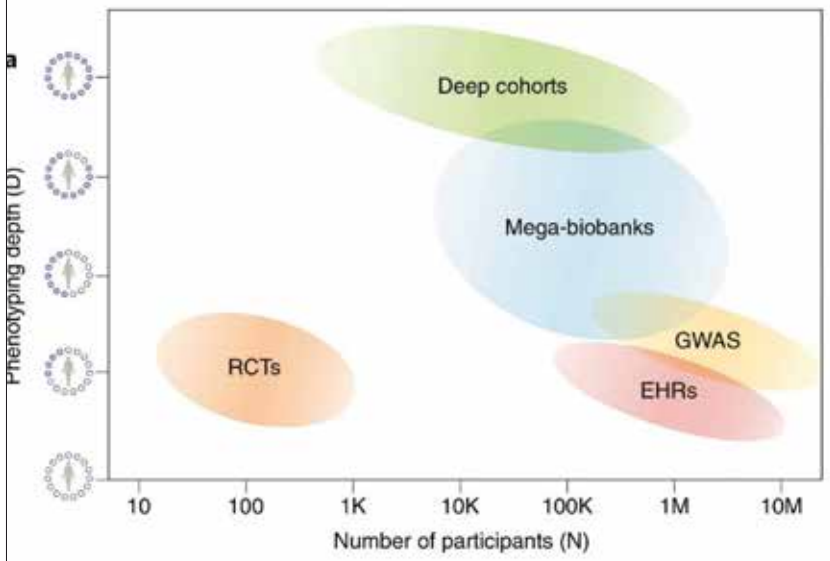
SUNG KYUN KWAN UNIVERSITY

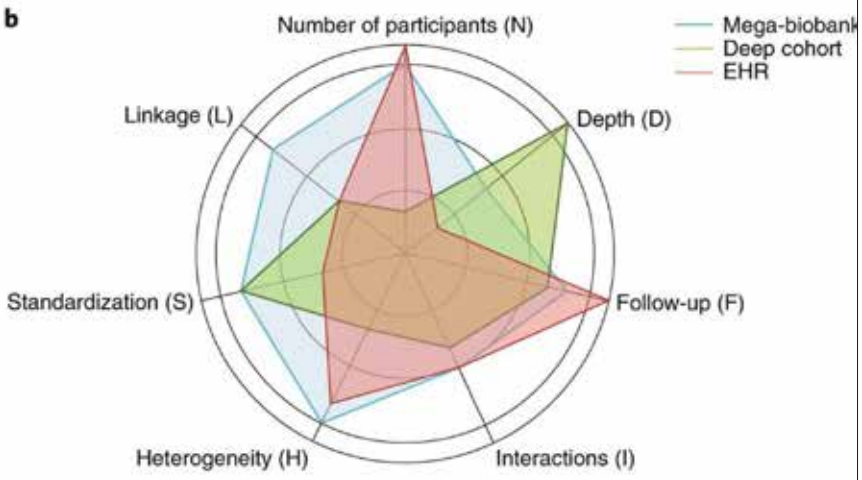
SAMSUNG MEDICAL CENTER



Shilo, S., Rossman, H. & Segal, E. Axes of a revolution: challenges and promises of big data in healthcare. *Nat Med* 26, 29–38 (2020). https://www.nature.com/articles/s41591-019-0727-5

Fig. 1 | The different axes of health data. The complexity of large health datasets can be represented by distinct axes, each encompassing a quantifiable property of the data.



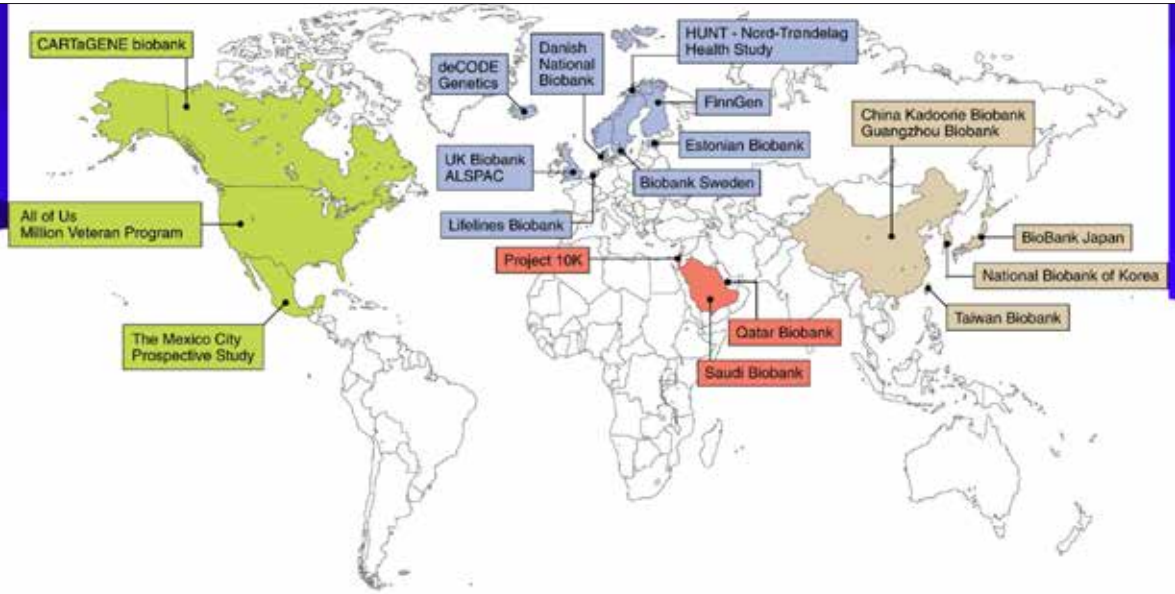


SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER


Mega Projects
SUNG KYUN KWAN UNIVERSITY
SAIHST
Samsung Advanced Institute for Health Sciences & Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER



Location	Biobank	N (goal)
Canada	CARTaGENE biobank[119]	43,000
USA	All of Us[33] Million Veteran Program[49]	1,000,000 > 600,000
Mexico	The Mexico City Prospective Study[52]	150,000
Iceland	deCODE Genetics	500,000
UK	UK Biobank[38] Avon Longitudinal Study of Parents and Children (ALSPAC)[20]	500,000 > 15,000
Netherlands	Lifelines Biobank[120]	> 167,000
Denmark	Danish National Biobank[121]	
Norway	HUNT - Nord-Trøndelag Health Study[122]	125,000
Sweden	Biobank Sweden	
Finland	FinnGen	500,000
Estonia	Estonian Biobank[123]	52,000
Israel	Project 10K	10,000
Saudi Arabia	Saudi Biobank	200,000
Qatar	Qatar Biobank[124]	60,000
China	China Kadoorie Biobank[51] Guangzhou Biobank[125]	> 500,000 30,000
Japan	BioBank Japan[126]	200,000
Korea	National Biobank of Korea[127]	500,000
Taiwan	Taiwan Biobank[128]	200,000

Shilo, S., Rossman, H. & Segal, E. Axes of a revolution: challenges and promises of big data in healthcare. *Nat Med* **26**, 29–38 (2020). https://www.nature.com/articles/s41591-019-0727-5


All of Us℠
The Precision Medicine Initiative®
THE FUTURE OF HEALTH BEGINS WITH YOU
NIH National Institutes of Health
All of Us Research Program
https://allofus.nih.gov/
JOIN NOW
Search
About
Get Involved
Funding and Program Partners
Protecting Data and Privacy
News and Events
The future of health begins with you.
The All of Us Research Program is inviting one million people across the U.S. to help build one of the most diverse health databases in history. We welcome participants from all backgrounds. Researchers will use the data to learn how our biology, lifestyle, and environment affect health. This could help them develop better treatments and ways to prevent different diseases.
JOIN NOW

US PMI Sub Projects

All of Us RESEARCH PROGRAM | The Precision Medicine Initiative — https://allofus.nih.gov

SYNC FOR SCIENCE — http://syncfor.science/

precisionFDA — https://precision.fda.gov/

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

All of Us

The *All of Us* Research Program

The cornerstone of PMI – led by the NIH

1,000,000+ volunteers reflecting the broad diversity of the U.S.

Opportunities for volunteers to provide data on an ongoing basis

Data shared freely and rapidly to inform many research studies

All of Us RESEARCH PROGRAM The Precision Medicine Initiative

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

The *All of Us* Research Program

All of Us will engage a group of **1 million or more** U.S. research participants:

Lay the scientific foundation

Identify new ways to treat and prevent disease

Test whether mobile devices can encourage healthy behaviors

Develop the right drug for the right person at the right dose

https://www.nih.gov/allofus-research-program

The Office of the National Coordinator for Health Information Technology

7

All of Us

Biobank

Communications and Engagement

Data and Research Center

Health Care Provider Organizations

Participant Center

Participant Technology Systems Center

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG

SAMSUNG MEDICAL CENTER

The NEW ENGLAND JOURNAL of MEDICINE
https://www.nejm.org/doi/full/10.1056/NEJMsr1809937

SPECIAL REPORT

The "All of Us" Research Program

Table 1. Data Available to Researchers from the All of Us Cohort.*

Data Source	Details
Current sources	
Health surveys	Initial surveys include information on sociodemographic characteristics, overall health, lifestyle, and substance use, with subsequent modules covering personal and family medical history and access to health care.
Physical measurements	Per-protocol measurements include blood pressure, heart rate, weight, height, body-mass index, and hip and waist circumferences.
Biospecimens†	Blood and urine samples are tested for DNA, RNA, cell-free DNA, serum, and plasma. If blood specimens cannot be obtained, saliva specimens are obtained.
Electronic health records	Initial capture of structured data includes billing codes, medication history, laboratory results, vital signs, and encounter records from health care provider organizations. Records will be expanded to include narrative documents. Pilot studies are testing data collection through Sync for Science and other health data aggregators.
Digital health information	Data can be captured from compatible participant-owned devices such as Fitbit. Pilot studies of other devices and linkage to health apps are being explored.
Future sources	
Health surveys	Additional modules, including surveys regarding social behavioral determinants of health, are under development.
Bioassays	Pilot studies for genotyping and whole-genome sequencing are expected to begin by early 2020. Additional pilot studies of bioassays are planned.
Health care claims data	Systems for the use of claims data, including billing codes and medication data, are under development.
Geospatial and environmental data	These data include geospatial linkage to measures such as weather, air quality, pollutant levels, and census data. Assays and sensor-based measurements of exposure are under consideration.
Other sources	Voluntary contributions of data from social networks (e.g., Twitter feeds) and additional biospecimen collections are under consideration.

* Additional information is available at https://www.researchallofus.org/data.
† Types of biospecimens are listed in Table S3 in the Supplementary Appendix.

SAMSUNG
SAMSUNG MEDICAL CENTER

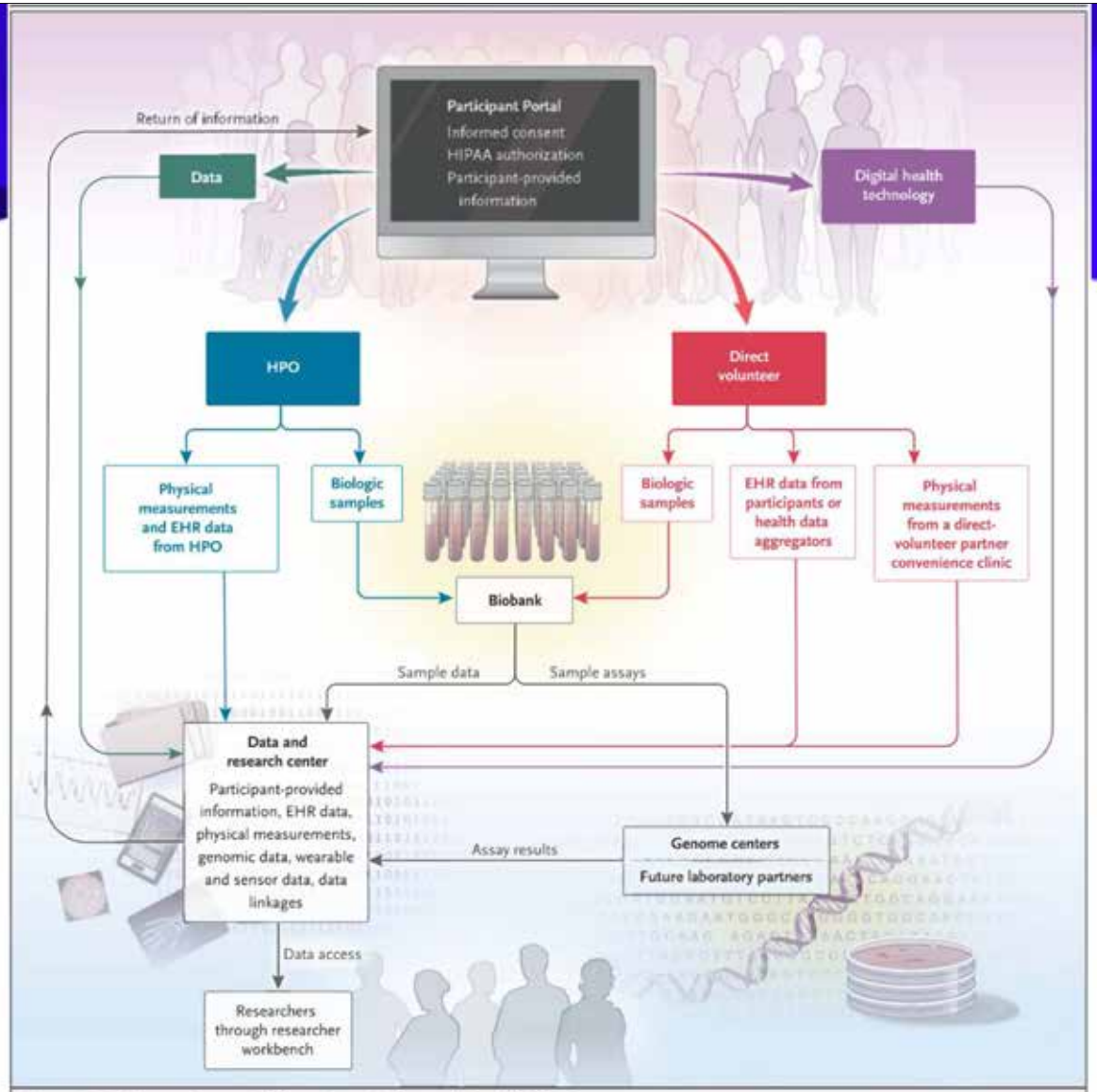


Figure 1. Participant, Data, and Biospecimen Pathways in the All of Us Program.

Genome centers in the All of Us program generate genomic data from biosamples. The researcher workbench is the platform for data analysis in the program. EHR denotes electronic health record, HIPAA Health Insurance Portability and Accountability Act, and HPO health care provider organization.

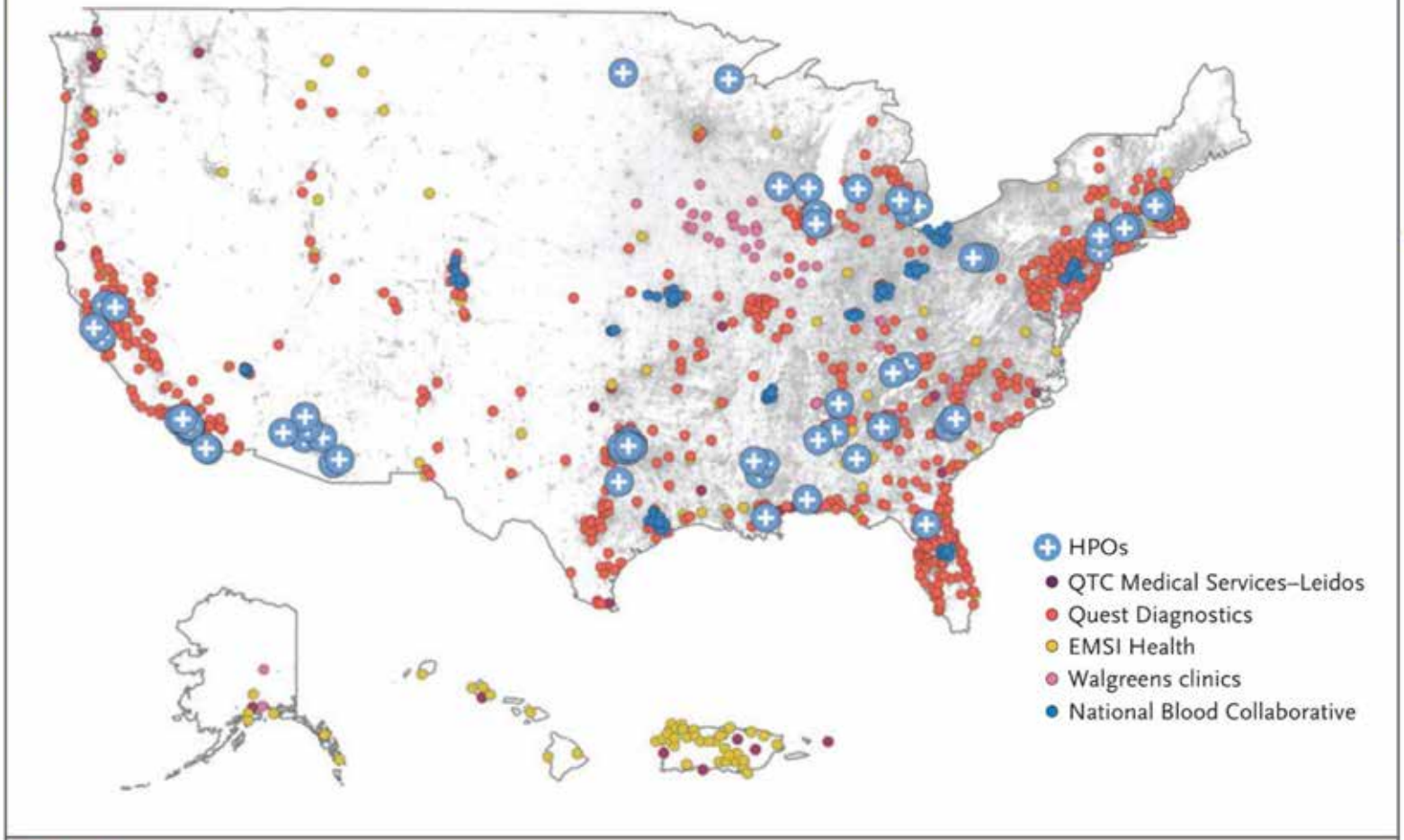


Figure 2. Planned Recruitment Sites and Network Partners in the All of Us Program.

Although most recruitment in the All of Us program is digital, in-person visits are required for physical measurements and collection of biospecimens. Currently, more than 340 recruitment sites have begun enrolling participants for these in-person assessments at HPOs and at the centers shown (QTC Medical Services–Leidos, Quest Diagnostics, EMSI Health, Walgreens clinics, and the National Blood Collaborative. The gray background of the map indicates population density, with one dot per person in the 2010 census.[12]

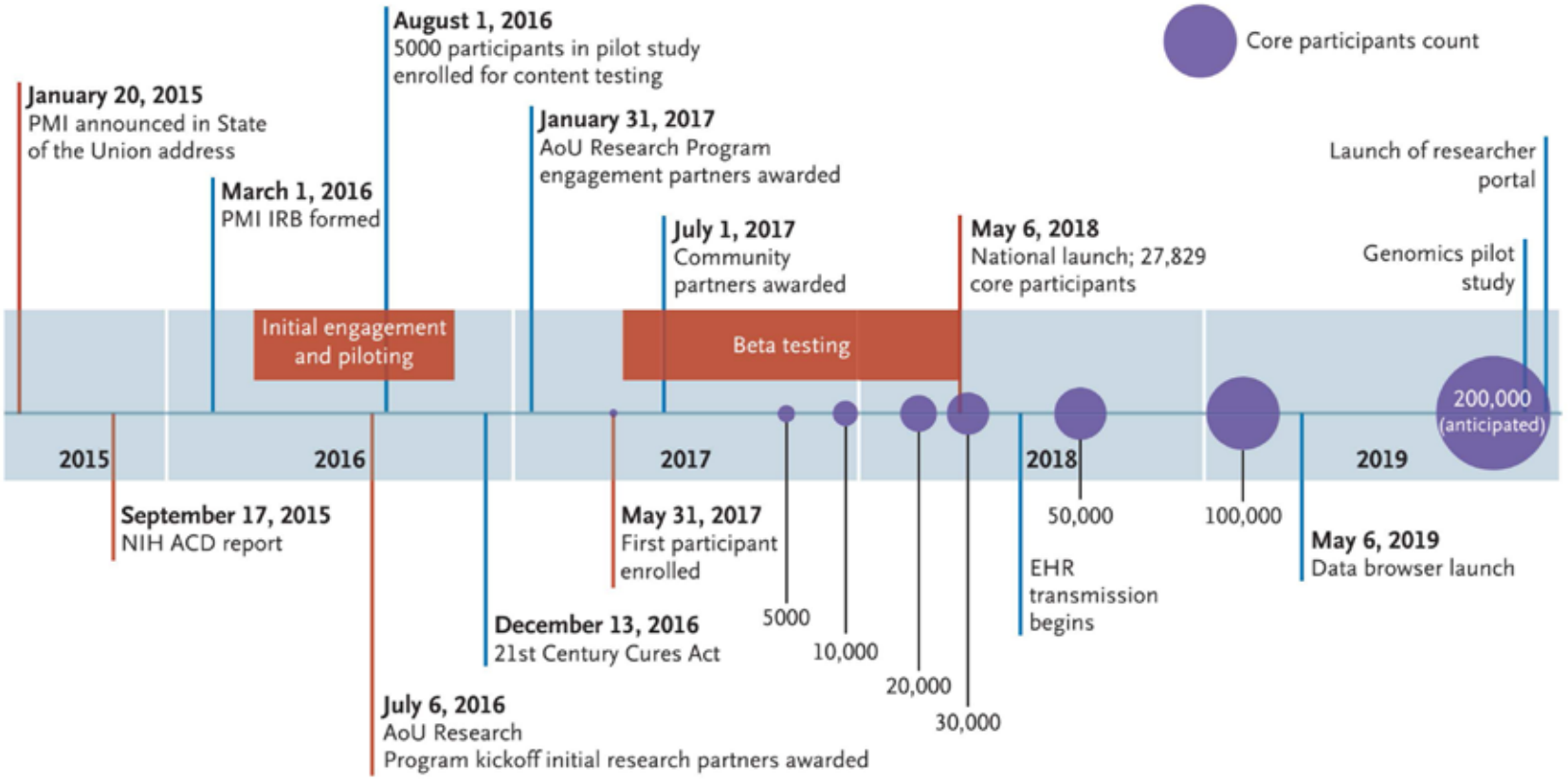


Figure 3. Timeline of the All of Us Program.

The Precision Medicine Initiative (PMI) Working Group of the Advisory Committee (ACD) to the National Institutes of Health (NIH) Director generated the initial blueprint for the structure and goals for the All of Us (AoU) program.[7] The core participants count includes persons who have completed health surveys, agreed to share their EHR data, had physical measurements taken, and contributed bio-specimens. Counts and dates after July 2019 are estimates. IRB denotes institutional review board.

Table 2. Scientific Goals of the All of Us Program and Expected Timelines.*

Goal	Years				
	End of 2018 (N=94,000)	End of 2019 (N=>200,000)	2020–2022 (N=<650,000)	2023–2027 (N=>1 million)	After 2027 (N=>1 million)
Return data to participants	+	+	+++	+++	++++
Establish demonstration projects†		+	+++	+	+
Discover genetic and environmental correlates with disease			++	+++	++++
Improve predictions of therapeutic safety and efficacy			++	+++	+++
Discover disease biomarkers			++	+++	+++
Connect mobile health, digital health, and sensor data with clinical outcomes			++	+++	+++
Develop new disease classifications			+	+++	++++
Support clinical trials			+	+++	+++
Enable machine-learning applications			++	+++	++++
Improve understanding of health disparities			++	+++	+++
Develop and test new therapeutic agents					++

* The expected number of participants in the cohort is shown for each time period. The number of plus signs in each cell indicates the anticipated relative degree to which each goal may be accomplished during the estimated timeline for focused research.

† Demonstration projects are scientific studies implemented by the All of Us program to show the quality, usefulness, validity, and diversity of the All of Us research data set and platform. In these projects, the population and data are further characterized, and the data are evaluated with a view to determining whether known associations can be replicated.

Learn about the program at AllofUs.nih.gov

All of Us Research Hub | NIH National Institutes of Health

ABOUT DATA TOOLS FAQ

Home > Data > Data Browser

Data Browser

The Data Browser provides interactive views of the publicly available *All of Us* Research Program participant data. Currently, participant provided information, including surveys and physical measurements taken at the time of participant enrollment, as well as electronic health record data are available. The *All of Us* Research Program data will include more data types over time.

In order to protect participant privacy, the data is limited to aggregate counts and summary demographic information. For more information, please visit our FAQ page.

https://databrowser.researchallofus.org/

Participants at a Glance

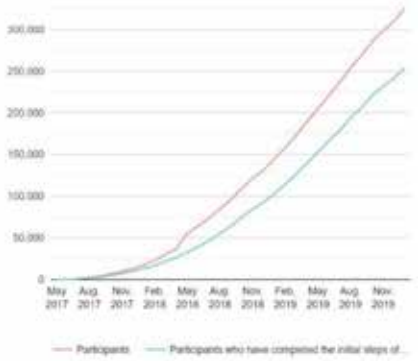
325,000+
Participants

254,000+
Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants (individuals who have consented to join the program) and participants who have completed all initial steps of the program (i.e., those who have consented, agreed to share their electronic health records, completed the first three surveys, provided physical measurements, and donated at least one biospecimen stored at the biobank).

The following numbers are approximated to protect participants' privacy. Numbers are updated as of February 1, 2020.



SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Health Sciences & T

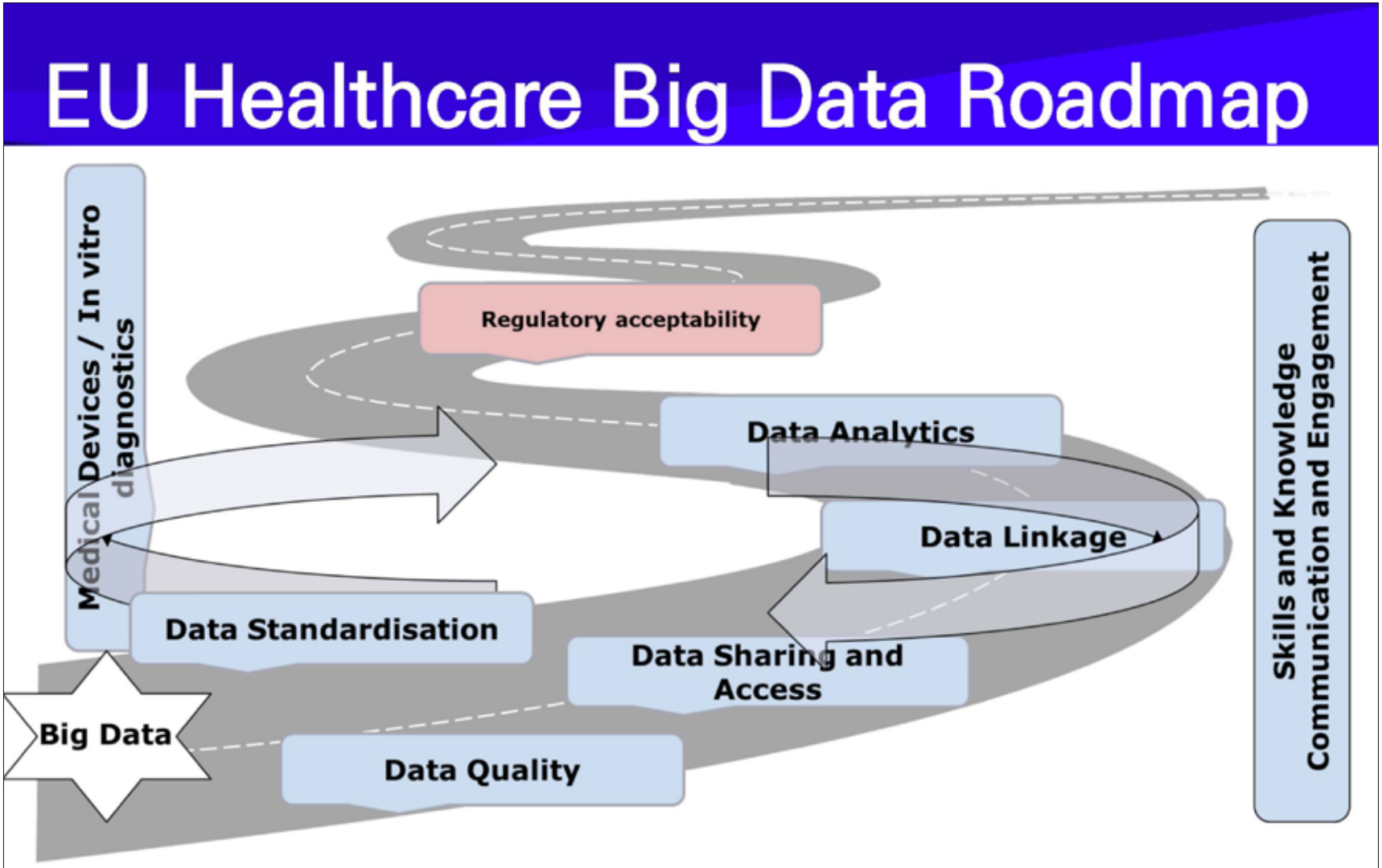
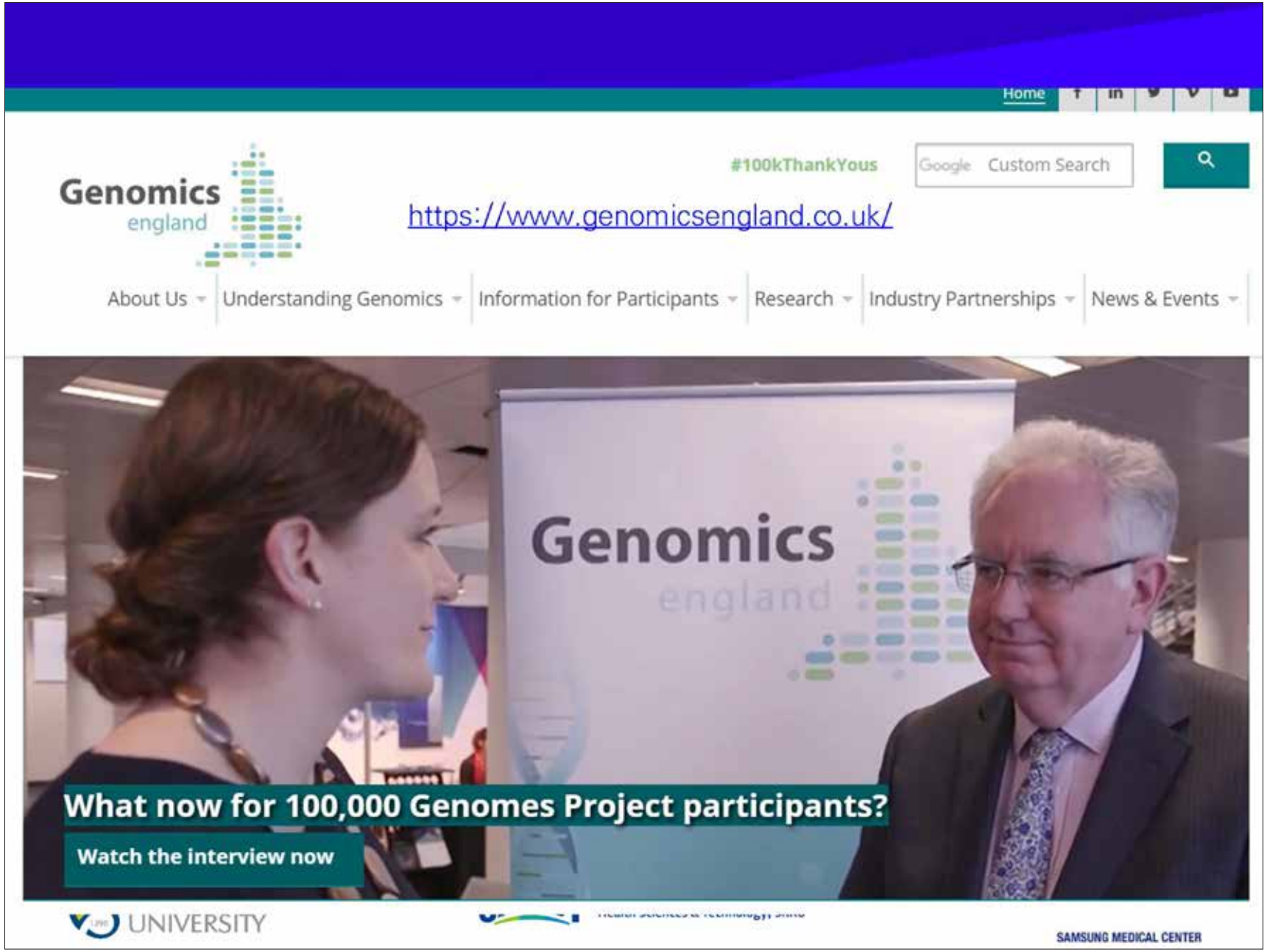
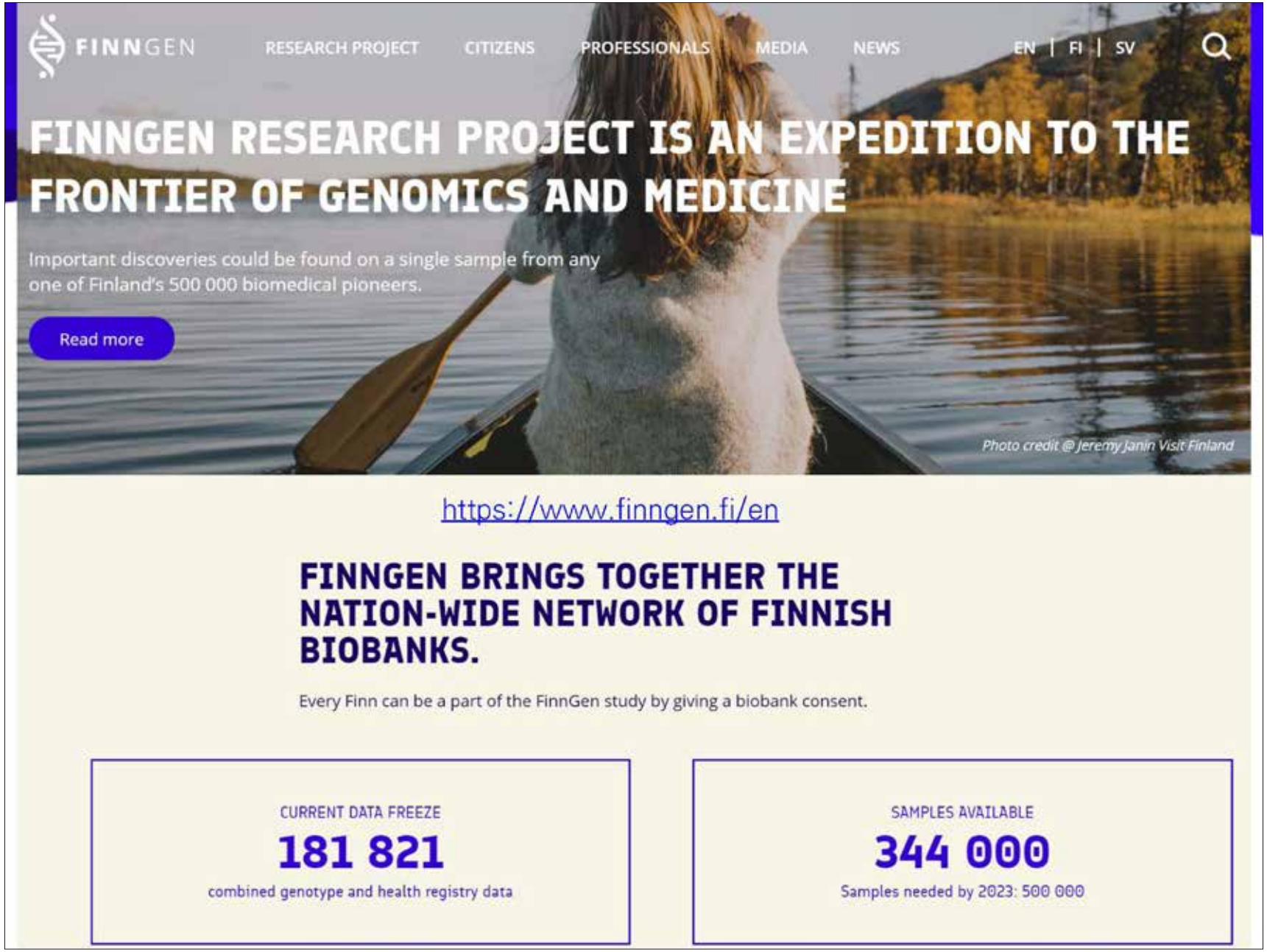


Figure 1: The Road to Regulatory acceptability: an integrated strategy reflecting core recommendations to support the use of Big Data in the assessment and monitoring of medicinal products in Europe. The individual steps are not necessarily sequential, may not be required across all datasets, many are interdependent and all will require active and iterative communication between all stakeholders.

HMA (Heads of Medicines Agencies)–EMA (European Medicines Agency)Joint Big Data Taskforce summary report, 2019. 02. https://www.hma.eu/506.html?&L=0


Home
Genomics england
#100kThankYous
Google Custom Search
https://www.genomicsengland.co.uk/
About Us
Understanding Genomics
Information for Participants
Research
Industry Partnerships
News & Events
Genomics england
What now for 100,000 Genomes Project participants?
Watch the interview now
UNIVERSITY
SAMSUNG MEDICAL CENTER


FINNGEN
RESEARCH PROJECT
CITIZENS
PROFESSIONALS
MEDIA
NEWS
EN | FI | SV
FINNGEN RESEARCH PROJECT IS AN EXPEDITION TO THE FRONTIER OF GENOMICS AND MEDICINE
Important discoveries could be found on a single sample from any one of Finland's 500 000 biomedical pioneers.
Read more
Photo credit @ Jeremy Janin Visit Finland
https://www.finngen.fi/en
FINNGEN BRINGS TOGETHER THE NATION-WIDE NETWORK OF FINNISH BIOBANKS.
Every Finn can be a part of the FinnGen study by giving a biobank consent.
CURRENT DATA FREEZE
181 821
combined genotype and health registry data
SAMPLES AVAILABLE
344 000
Samples needed by 2023: 500 000

[그림 1] 핀란드의 유전체 정보 및 국민건강 데이터 구조(Finnish health data exploitation infrastructure)

http://www.koreabio.or.kr/cms/cmsView.do?menu1=8&menu2=2&type1=4&viewCheck=true&bbsseq=45284&bbsid=bbs9

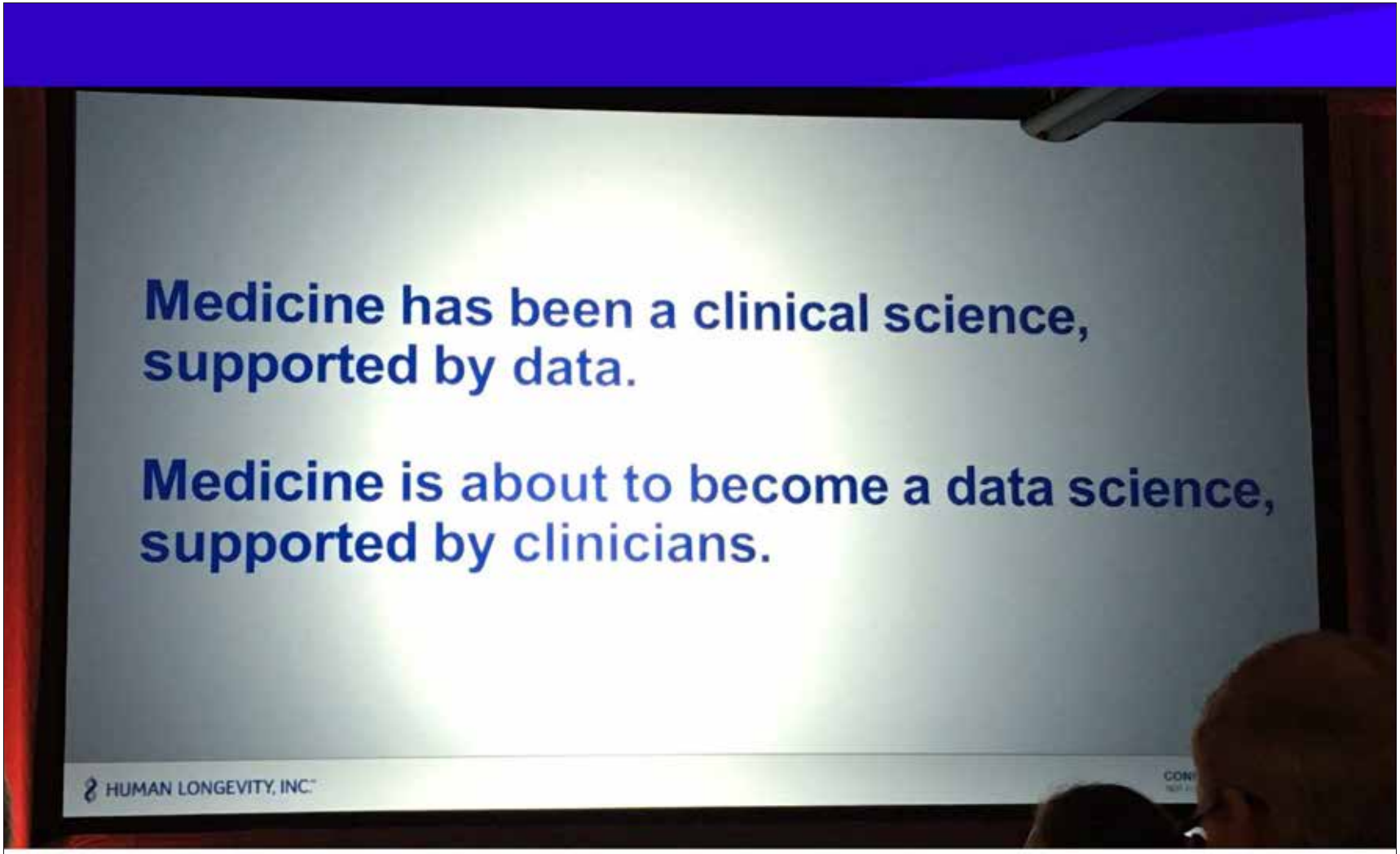
SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER


보건의료 빅데이터 플랫폼
소개
데이터 카탈로그
데이터 신청
빅데이터 활용 제안
고객의 소리
https://hcdl.mohw.go.kr/BD/Portal/Enterprise/DefaultPage.bzr
보건의료 빅데이터 플랫폼
보건의료 분야 연구자들이 공공적 목적으로 연구할 수 있도록
공공기관이 보유한 보건의료 데이터를 연계 결합하여 제공합니다
시범사업 소개
데이터 신청
데이터 카탈로그 조회
데이터 신청 현황
신청접수 11 건
심의 중 84 건
처리완료 5 건
공지사항
• 제 5, 6차 연구평가소위원회 결... 2019/12/18
• 데이터 이용신청서 출력 방법 안... 2019/11/04
• 플랫폼 서비스 정기점검안내 (매... 2019/10/31
• (개통안내) 국민이 원하는 보건... 2019/07/25
데이터 제공 심의 일정
정책심의위원회
연구평가소위원회
5차 04.15 5차 12.09
6차 09.06 6차 12.10
FAQ
• R&D과제와는 별개로 보건의... 2019/12/27
• 4개 기관 별 데이터 제공 기간... 2019/12/05
• 데이터 이용 신청 시 기관생명... 2019/11/08
• 폐쇄환경에서 사용할 수 있는 ... 2019/11/08
UNIVERSITY
SAMSUNG MEDICAL CENTER


의료 AI
SUNG KYUN KWAN UNIVERSITY
SAIHST
Samsung Advanced Institute for Health Sciences & Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER



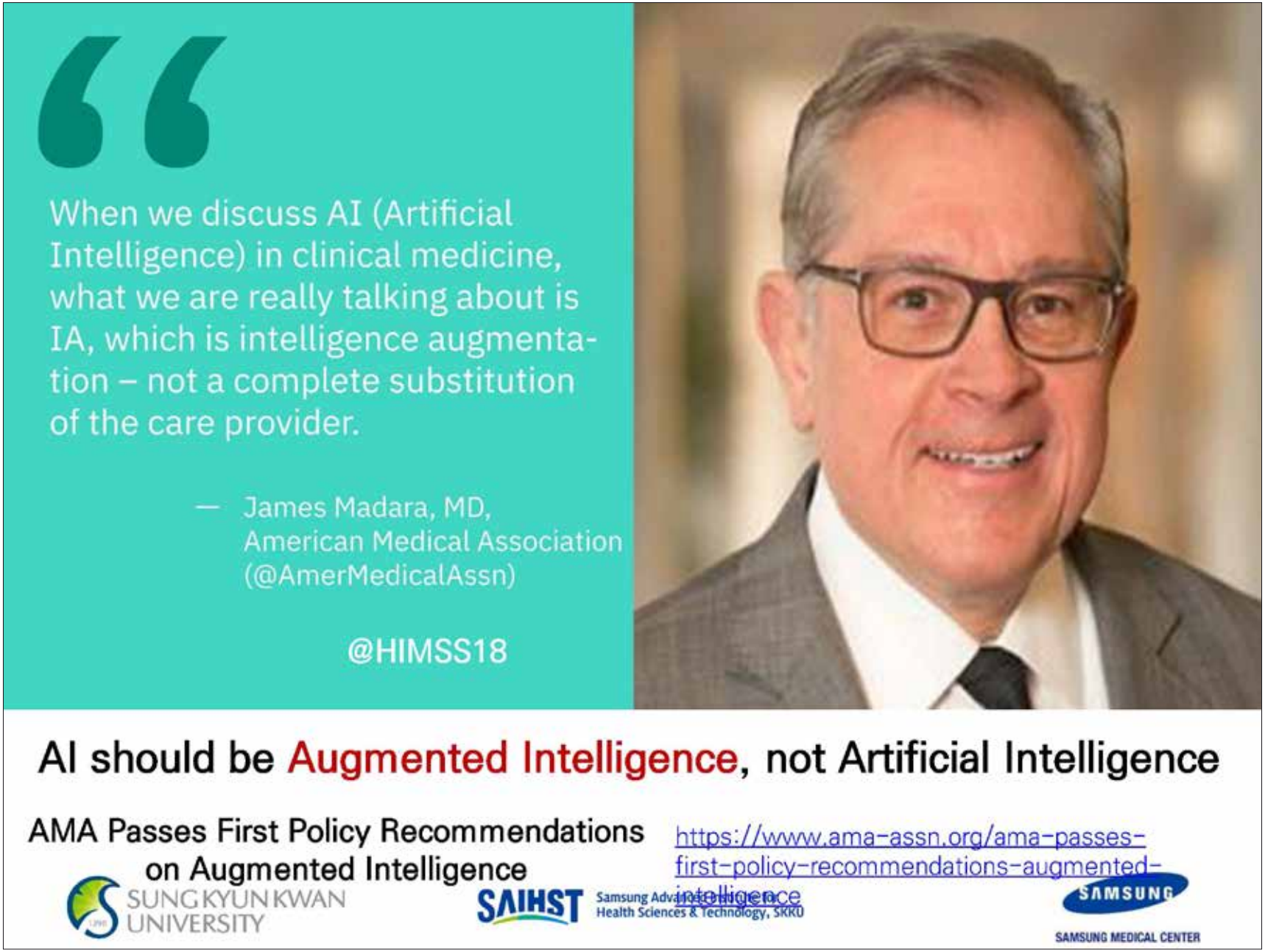
https://twitter.com/pafournier/status/798582105760415748

http://www.humanlongevity.com/wp-content/uploads/HLI-FactSheet.pdf

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER


When we discuss AI (Artificial Intelligence) in clinical medicine, what we are really talking about is IA, which is intelligence augmentation – not a complete substitution of the care provider.
— James Madara, MD, American Medical Association (@AmerMedicalAssn)
@HIMSS18
AI should be Augmented Intelligence, not Artificial Intelligence
AMA Passes First Policy Recommendations on Augmented Intelligence
https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence
SUNG KYUN KWAN UNIVERSITY
SAIHST
Samsung Advanced Institute for Health Sciences & Technology, SKKU
SAMSUNG
SAMSUNG MEDICAL CENTER

Other Important Points

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

Data!, Not Algorithm!

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough:		**3 years**	**18 years**

http://www.kdnuggets.com/2016/05/datasets-over-algorithms.html

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

Isaac Kohane
@zakkohane

Following

‘Our problem is not sample size, it’s distinguishing clinically significant from only statistically significant results’

Isaac Kohane @zakkohane
Health care systems that really use their data to help their patients: Great talk by Dr Noa Dagan-winner of @NEJM data challenge challenge.nejm.org/pages/home -discusses her work/research @ClalitResearch today @HarvardDBMI

2:37 AM - 12 Oct 2018

https://twitter.com/zakkohane/status/1050440139858735104

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

규제는 피할 수 없다

Multiple spheres of regulation

phg foundation making science work for health

- Machine learning ***as data*** | the GDPR **개인정보보호법, 생명윤리 및 안전에 관한 법률**
- Machine learning ***as a medical device*** | the MDR, IVDR, and harmonised standards **의료기기법, 의료기기산업 육성 및 혁신의료기기 지원법**
- Machine learning ***as intellectual property*** | patents, copyright, trade secrets, and licensing
- Machine learning ***as a source of liability*** | negligence and product liability **의료법,…**

https://twitter.com/PHGFoundation/status/1171030245102759936

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER

Q & A

sooyong.shin@gmail.com
http://bmi.skku.edu

SUNG KYUN KWAN UNIVERSITY

SAIHST Samsung Advanced Institute for Health Sciences & Technology, SKKU

SAMSUNG MEDICAL CENTER