

Singhealth-Duke Health Services Research Data Science Core

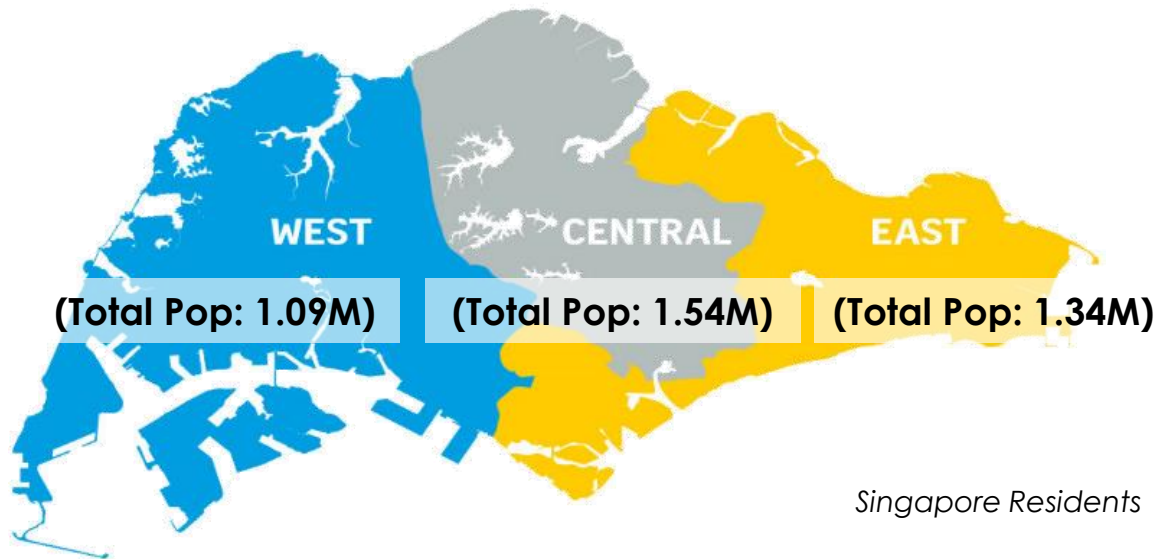
Dr Sean Lam

Head, Data Science, Health Services Research Center, SingHealth

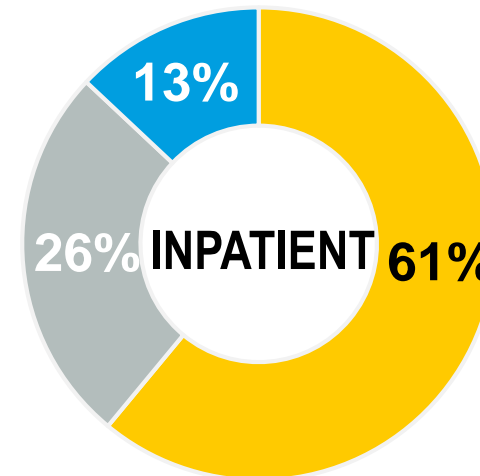
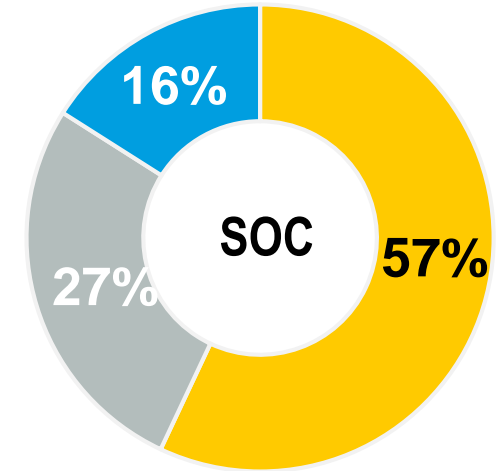
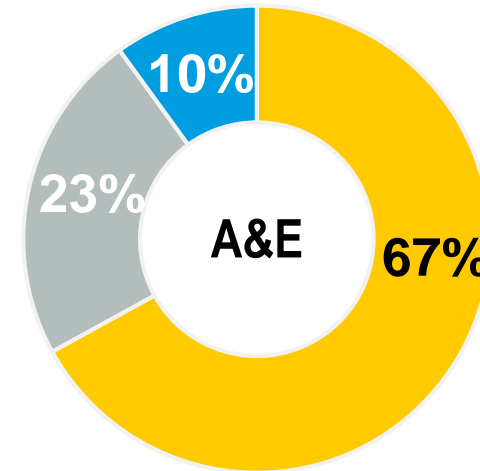
Assistant Professor, Health Services and Systems Research (HSSR), Duke-NUS Medical School

Adjunct Lecturer, Lee Kong Chian School of Business, Singapore Management University

Where are our Patients coming from



SingHealth's Patients



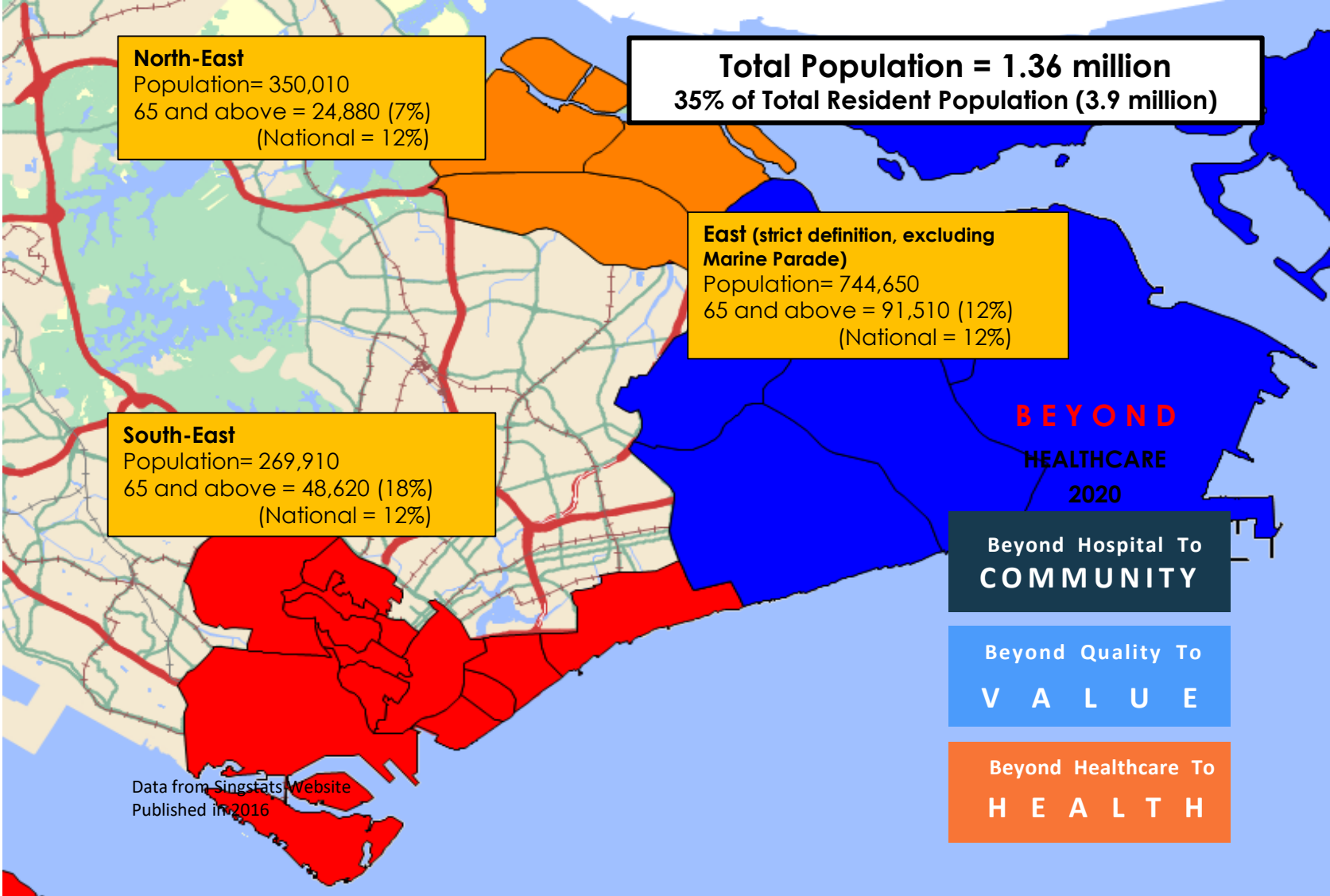
From the ...

- EAST
- CENTRAL
- WEST

SINGHEALTH GROUP

SGH & CGH A&E ATTENDANCES	210,590
SGH & CGH SOC ATTENDANCES	376,191
SGH/NHCS & CGH INPATIENT ADMISSIONS	103,549

One Regional Health System, Three Operating Sites



Transforming Healthcare with our Dual Roles

National Role

**Cutting Edge
Tertiary &
Quaternary
Care**

**Leverage
Research & Innovation**
to transform healthcare
and deliver new models
of care

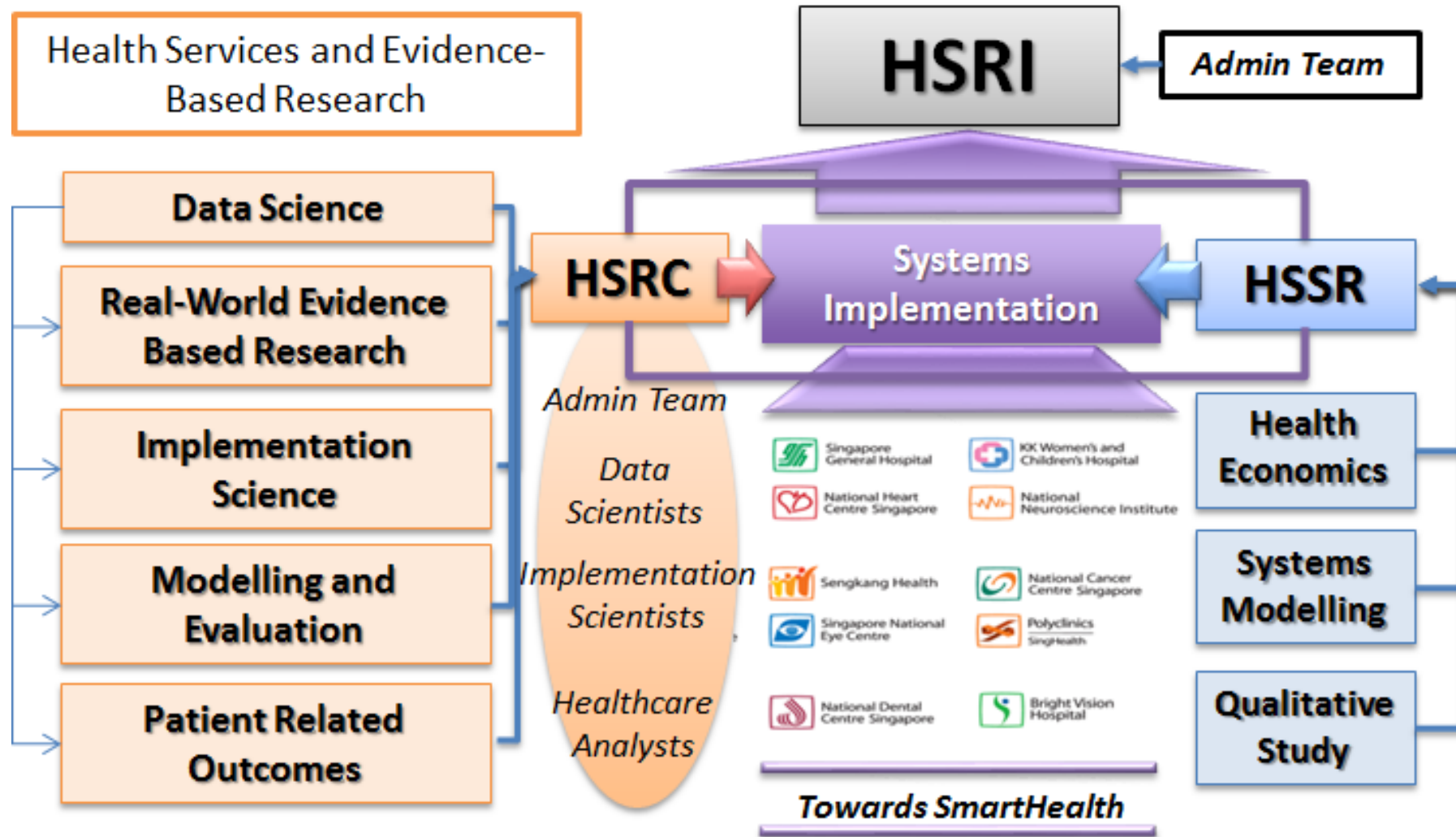


Regional Role

**Community &
Population
Health**

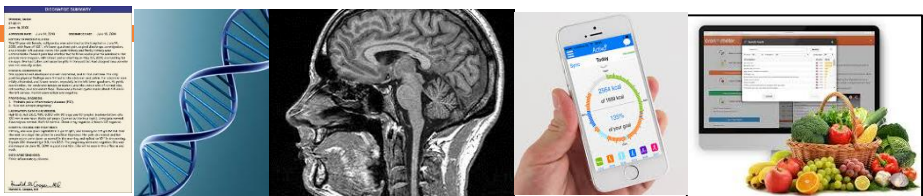
**Strengthen
Regional Health System**
to deliver integrated
care and promote
population health

Roles of HSRI and HSRC

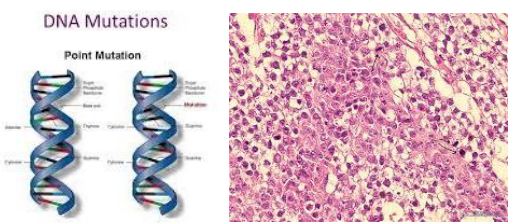


Data is the New Oil of Healthcare and Biomedicine

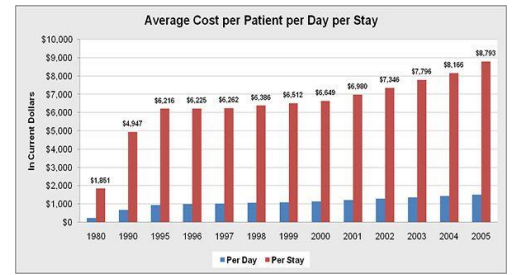
Data Generation



Harnessing and Using the Data



Disease and Biological Insights



Improve Hospital Efficiencies and Processes



New Tools for Healthcare



Improve Patient Outcomes and Experiences

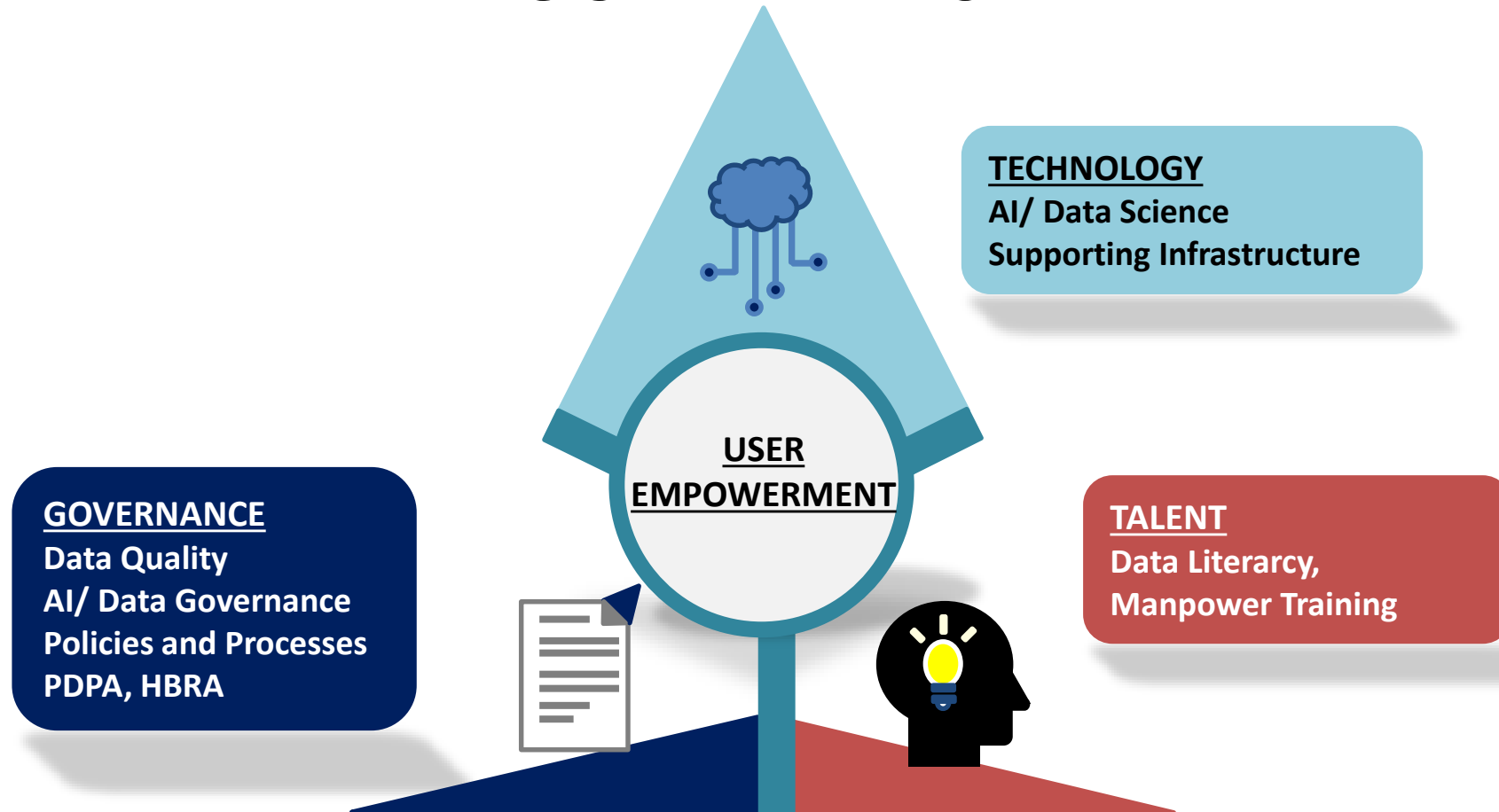


Lower Healthcare Costs



Towards an Empowered Data Science Ecosystem

TECHNOLOGY; TALENT ; GOVERNANCE



Data Science (not just) Data Analytics

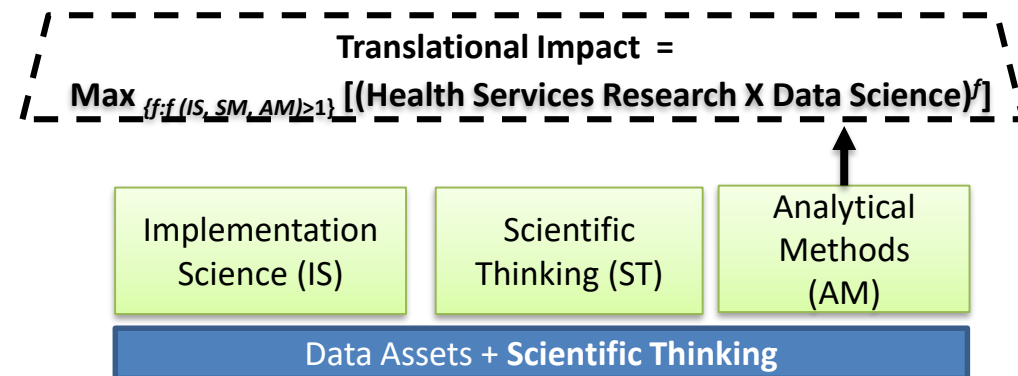
- Data Analytics can produce operational insights, but
- Data Analytics \neq Data Science!
- Data Science requires rigorous scientific thought processes
 - “Data-driven Science”
 - “Evidence-based”
 - “Statistical Science”
 - “Statistics” – Jeff Wu of Michigan U
 - Etc...

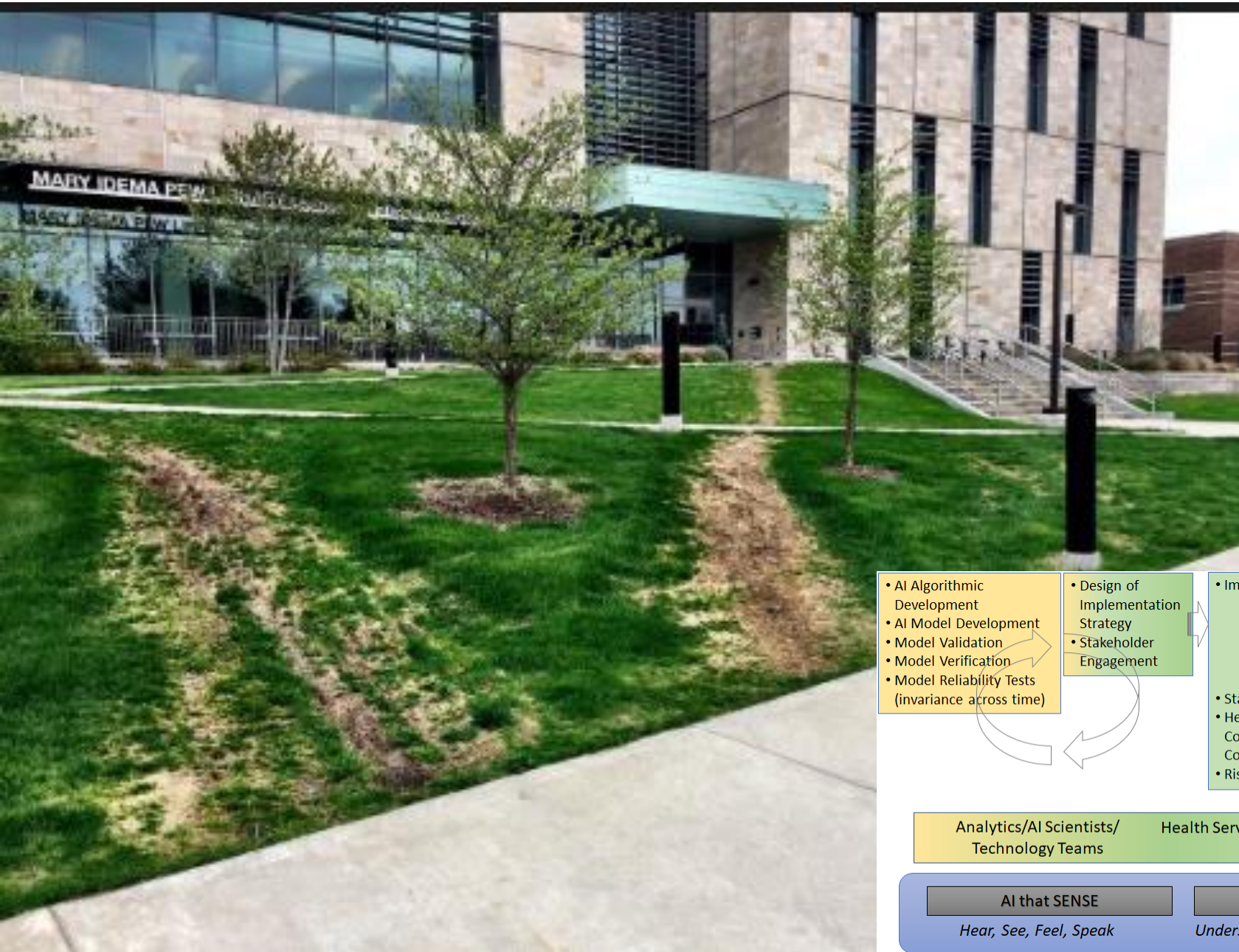
The key word in "Data Science" is not Data, it is Science

Jeff Leek 2013/12/12

One of my colleagues was just at a conference where they saw a presentation about using data to solve a problem where data had previously not been abundant. The speaker claimed the data were “big data” and a question from the audience was: “Well, that isn’t really big data is it, it is only X Gigabytes”.

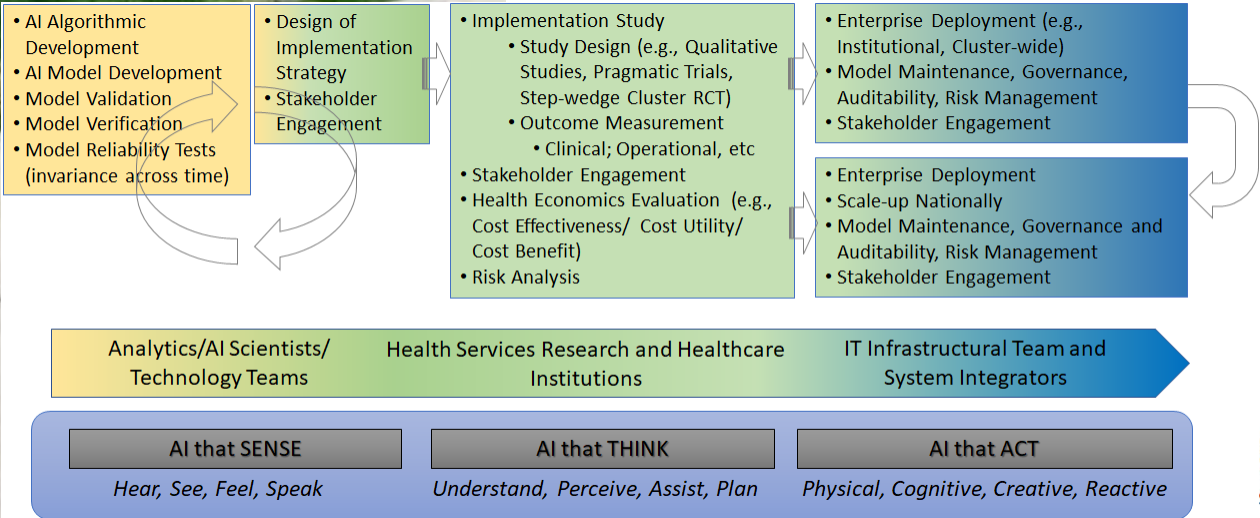
While that exact question would elicit groans from most people who work with data, I think it highlights one of the key problems with the thinking around data science. Most people hyping data science have focused on the first word: data. They care about volume and velocity and whatever other buzzwords describe data that is too big for you to analyze in Excel. This hype about the size (relative or absolute) of the data being collected fed into the second category of hype - hype about tools. People threw around EC2, Hadoop, Pig, and had huge debates about Python versus R.



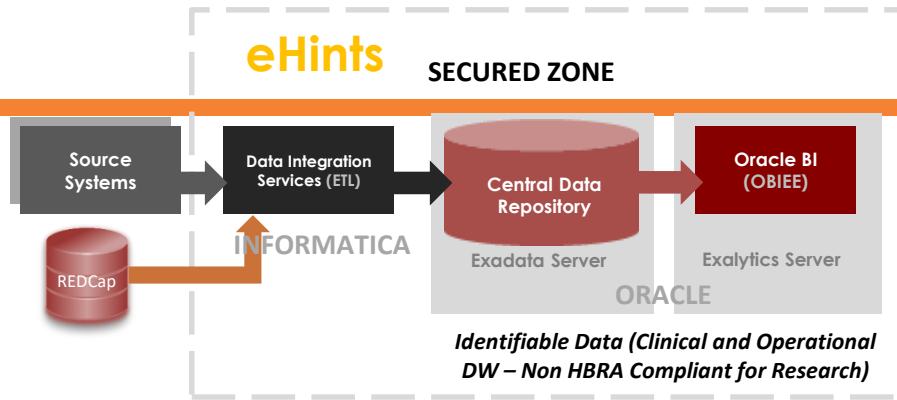


Design for Implementation!

From “Data Science” to “Implementation Science”



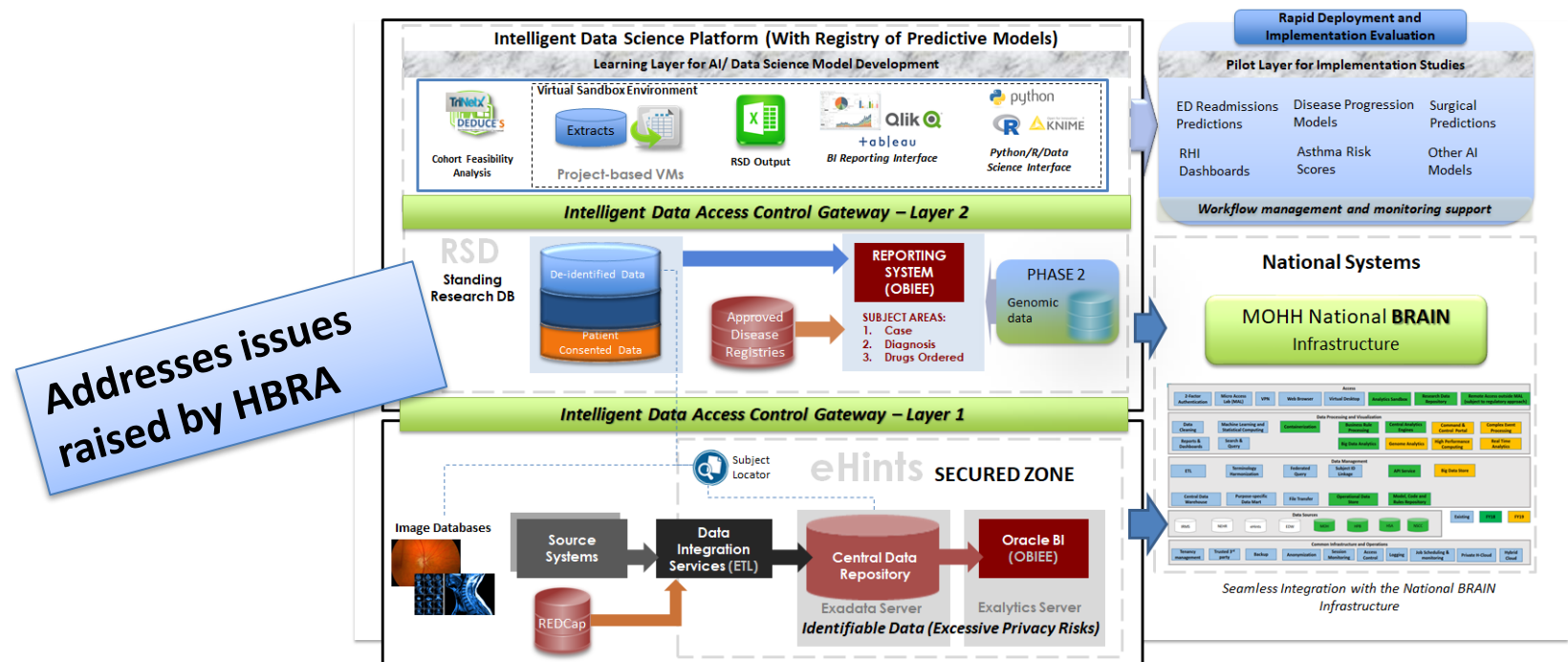
Clinical Data Warehouse (DW) at SingHealth



eHINTS - SingHealth Data Warehouse

- Sample Data Sources ingested:
 - LAB, OAS, OPEC; SAPISH; MAXCARE; SCM-ED; OTMS; RIS; REDCap
- Structured DW that facilitates logical data consumption from disparate data sources

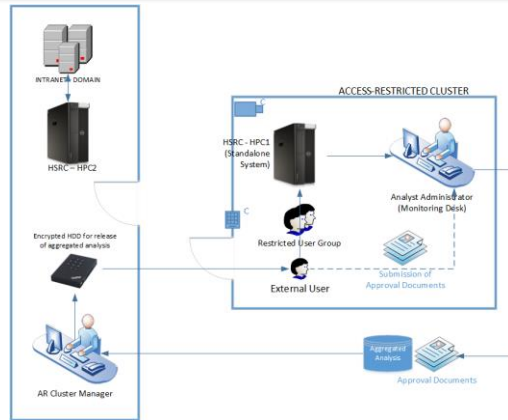
Research Standing Database



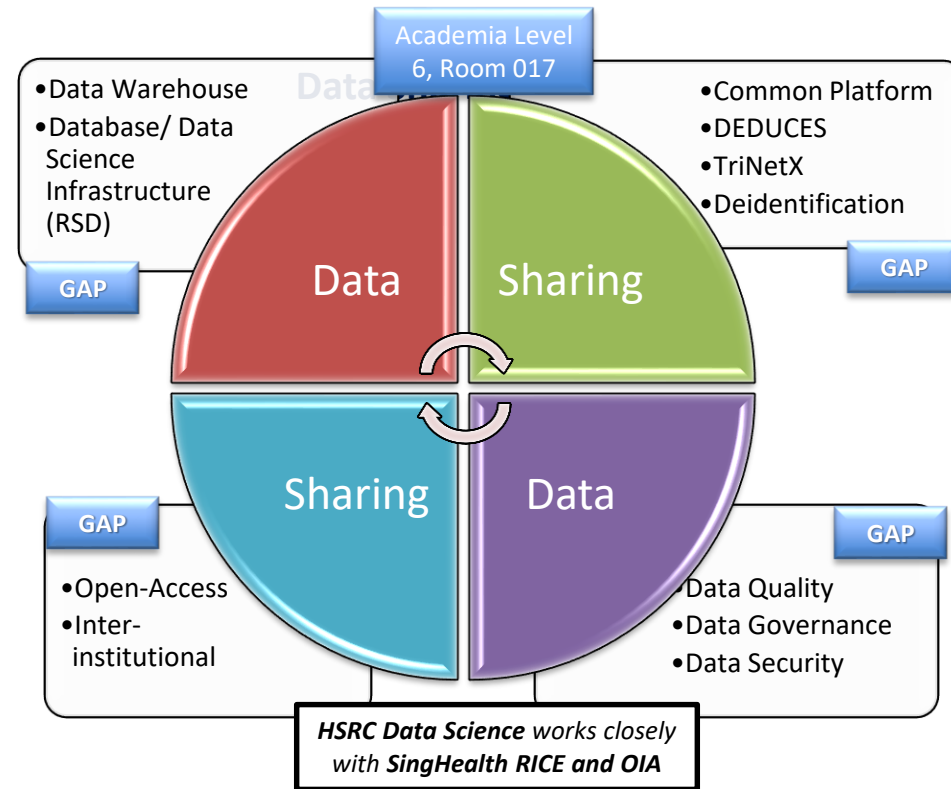
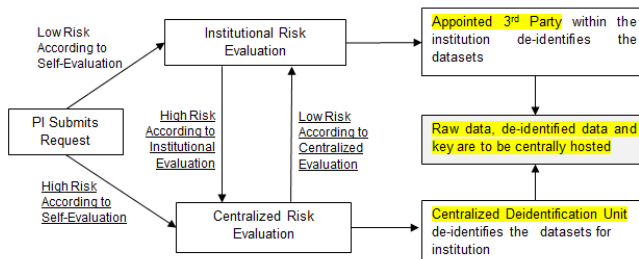
Addresses issues raised by HBRA

DATA SHARING AND RESEARCH

Access Restricted Cluster (External Collaborations)



Cluster Deidentification Process and Training

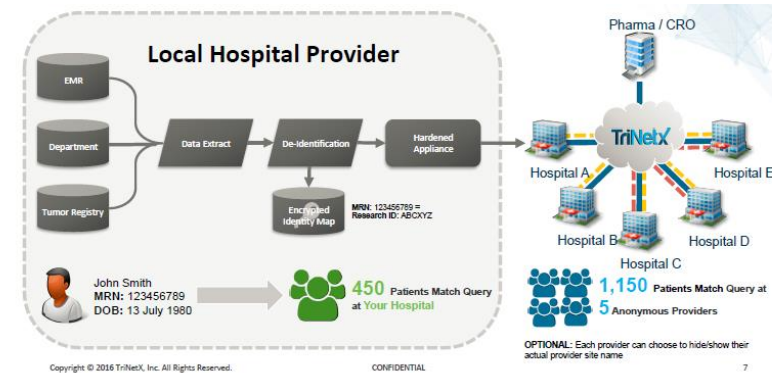


Within GOVERNANCE:
Training – Systems – Policies and Processes

DATA SCIENCE INFRASTRUCTURE for RESEARCH

TriNetX - 46.2M patients with 10.4B clinical facts in TriNetX network

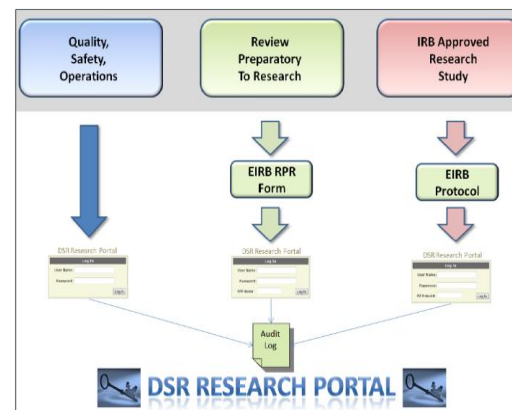
- SingHealth joined the TriNetX Consortium in 2017
- HSRC currently hosts the internal TriNetX hardware servers and software
- Pending feasibility studies and IT Security clearance for full pilot
- **Only aggregate results** can be obtained by Pharma / CRO
- **Collaborative research** is also possible between hospitals
- Hospitals' data always **remains within the internal system**



DEDUCES – Collaboration between SingHealth – Duke NUS – Duke Health

The screenshot shows the 'DEDUCE Research Portal' interface. The main section is the 'Cohort Manager' with a table of cohorts. The table has columns for 'Select', 'Line #', 'Cohort Name', 'Phy. Enc. Counts', 'Source', and 'Updated On'. The table lists several cohorts, including 'All Patients', 'ICD Diagnosis Code', 'Age on the Day of the Visit', 'Test Name', and 'Type II Diabetes, over 50, HgA1c in range, seen in last year'. The 'All Patients' cohort has 3,077,000 patients. The 'Type II Diabetes, over 50, HgA1c in range, seen in last year' cohort has 11,560 patients.

Select	Line #	Cohort Name	Phy. Enc. Counts	Source	Updated On
<input type="checkbox"/>	1	All Patients	3,077,000	all	202018-01-22 PM
<input type="checkbox"/>	1.1	ICD Diagnosis Code	250.20,250.22,250.10,...	all	202018-01-22 PM
<input type="checkbox"/>	1.1.1	Age on the Day of the Visit, in years	40-50 122,545	all	202018-01-22 PM
<input type="checkbox"/>	1.1.1.1	Test Name	GLYC HDG (A1C) G	all	202018-01-22 PM
<input type="checkbox"/>	1.1.1.1.1	Numeric Result	Between 6.4 & 8.1	all	202018-01-22 PM
<input type="checkbox"/>	2	Type II Diabetes, over 50, HgA1c in range, seen in last year	11,560	all	202018-01-22 PM
<input type="checkbox"/>	3	Patients with Vascular Disease	62,203	all	202018-01-22 PM
<input type="checkbox"/>	4	7, 2 and 3	2,262	all	202018-01-22 PM
<input type="checkbox"/>	5	Patients with Carotids or Kidneys	68,187	all	202018-01-22 PM
<input type="checkbox"/>	6	4 minus 5	2,102	all	202018-01-22 PM

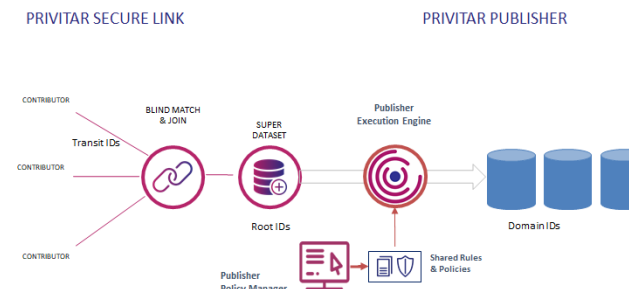
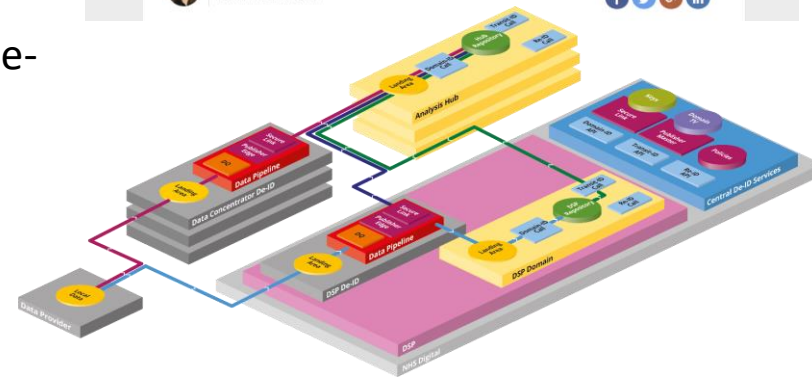


DATA GOVERNANCE INFRASTRUCTURE

Privacy Preserving System (IT Evaluation)

- Technology in Privacy Preserving Systems is available and rapidly evolving
- NHS recently called for a tender to develop a Data Services Platform (DSP)
 - DSP includes various components, including De-ID services
 - Key objectives:
 - Enhance safety and security
 - Improve timeliness and utility
 - Remove duplication and drive efficiency
 - Etc ...

A comprehensive tender should be called to provide a **rigorous evaluation of privacy-preserving software and technologies** with conformance to detailed requirements.



TALENT DEVELOPMENT (CURRENT)



SingHealth – Duke NUS Joint HSR and Data Science Training Programs

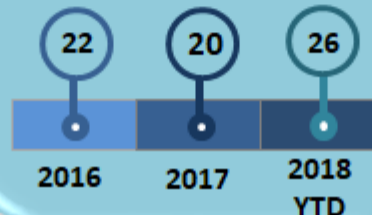
- Deep Learning for Image Processing
- Health Systems Modeling Using Systems Dynamics
- Advanced REDCap User Training
- Tableau Training
- De-identification Training
- Practical Data Science Primer



Data Science Internships

Identifying Talents that can contribute to our journey to becoming a data-driven LHS

Interns (OIA / HSRC)



Citizen Data Science Training

Empowering SingHealth users to be equipped with advanced data literacy and data analytical skills

In collaboration with:

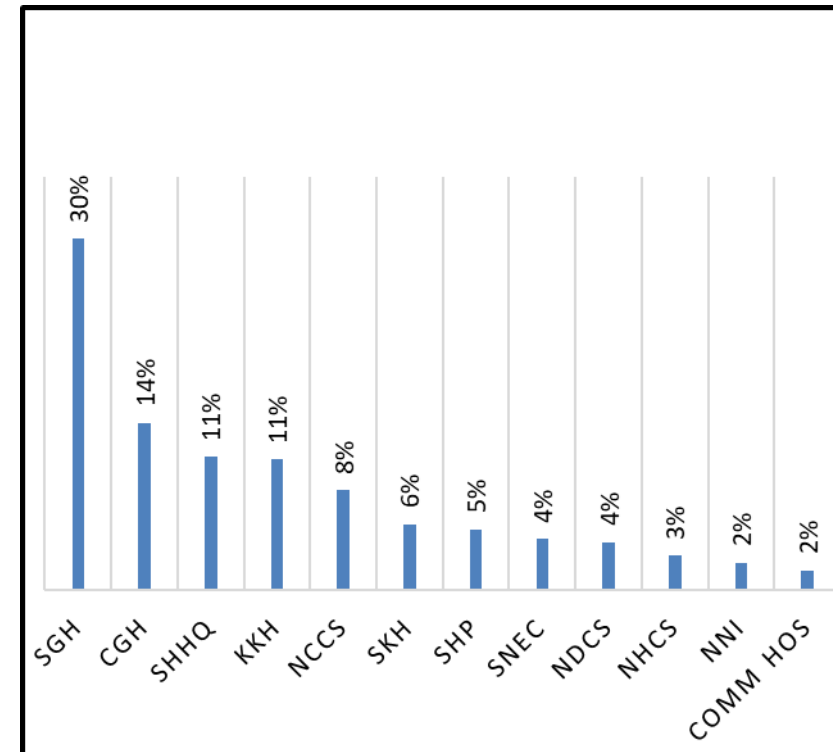


TALENT DEVELOPMENT

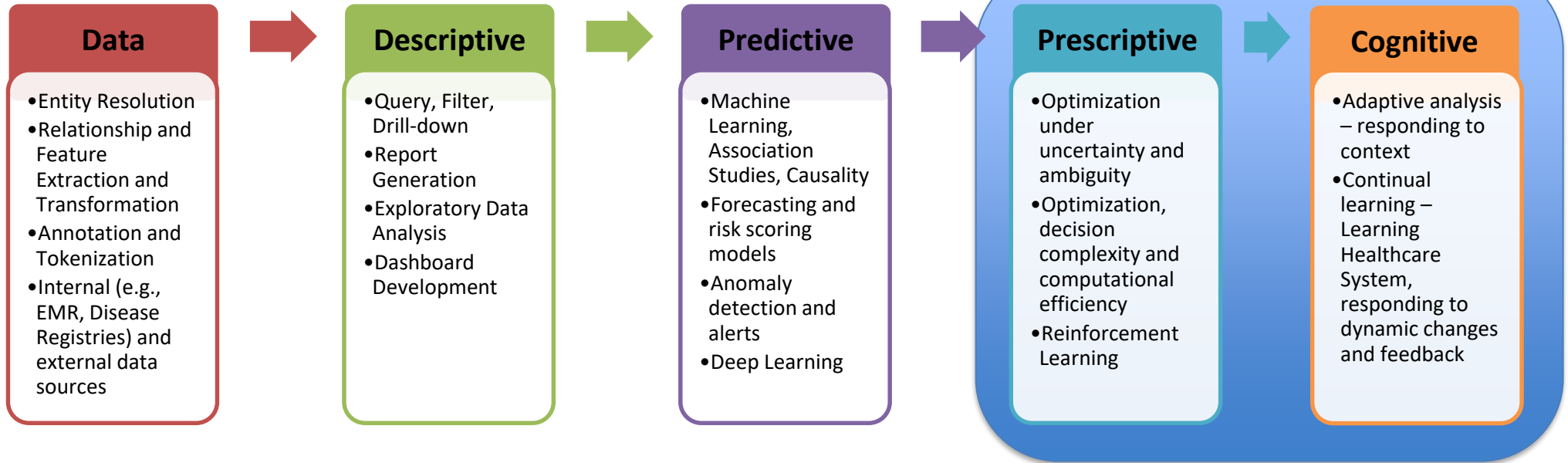
- Survey was conducted by HSRC and OIA in Nov 2018 to evaluate the Data Science Needs of the Cluster

List of Important Data Science Modules:

1. Data Visualization And Story Telling
2. Design Thinking And Agile Methodology
3. Predictive Modelling And Deep Learning
4. Capstone Project: Developing A Data Analytics Model
5. Optimization And Simulation Modelling (Prescriptive Modelling)
6. Data Preprocessing Using R Or Python



Artificial Intelligence in Medicine

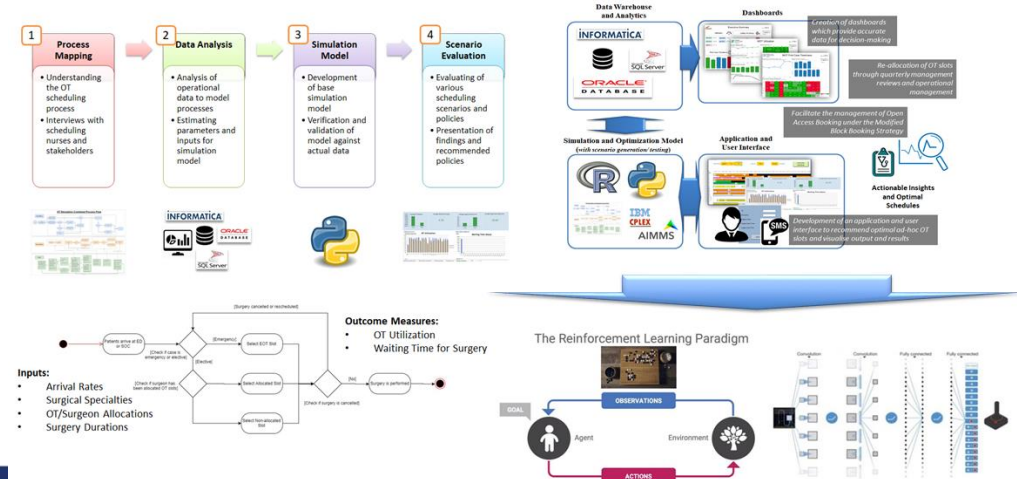
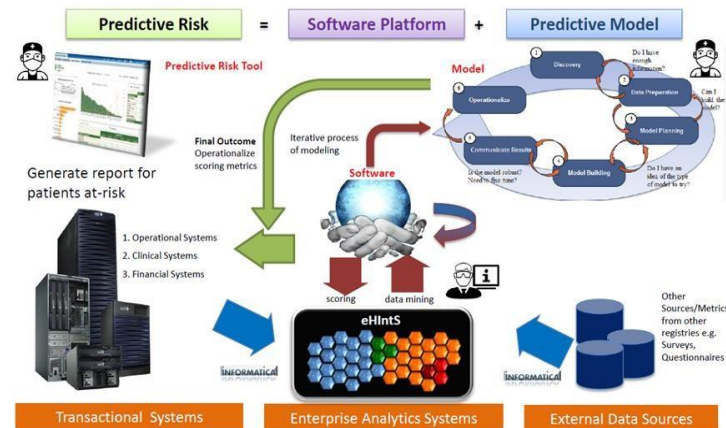


Electronic Research Databases

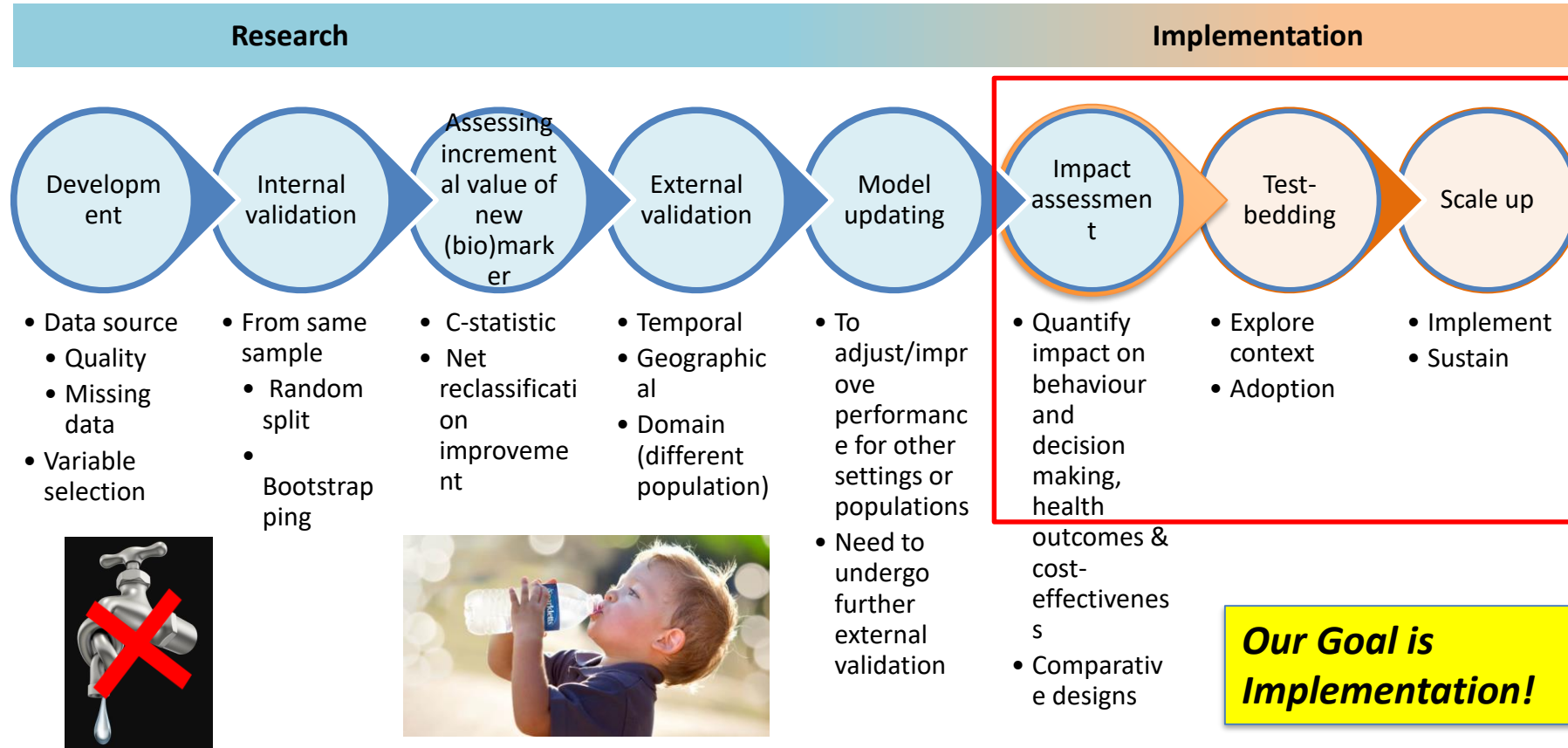
REDCap
Research Electronic Data Capture

Disease Registries

RoD
Registry of Databases



AI/ Data Science models need to go beyond validation to IMPLEMENTATION



Our Goal is Implementation!

Data Rich with INformation and Knowledge (DRINK!)

1. Moons *et al. Heart.* 2012;98(9):691-698
2. Moons *et al. Heart.* 2012;98(9):683-690
3. Amarasingham *et al. Health affairs* 2014;33(7):1148-54

SingHealth HSRC Data Science Core

- General Enquiries
- Email: hsr@singhealth.com.sg

- HSRC Data Science Core:
 - Dr Sean Lam
 - Email: lam.shao.wei@singhealth.com.sg

