BECKER'S
HOSPITAL REVIEW

Unified Data architectures – Building the Instruments into the Operation rather than adding them later

Becker's Hospital Review
2nd Annual Health IT + Clinical Leadership Conference

May 2 - 4, 2019 at the Swissotel, Chicago, IL

Clemens M. Schirmer, MD, PhD, FAANS, FACS, FAHA

Director Comprehensive Stroke Centers
System Director Cerebrovascular and Endovascular Neurosurgery
Program Director, Neurosurgery Residency Program and CAST Neuroendovascular Fellowship Program
Department of Neurosurgery and Neuroscience Institute
Geisinger

Associate Professor in Neurosurgery
Temple University School of Medicine

Geisinger Caring

Big data is finally here...

By the numbers…

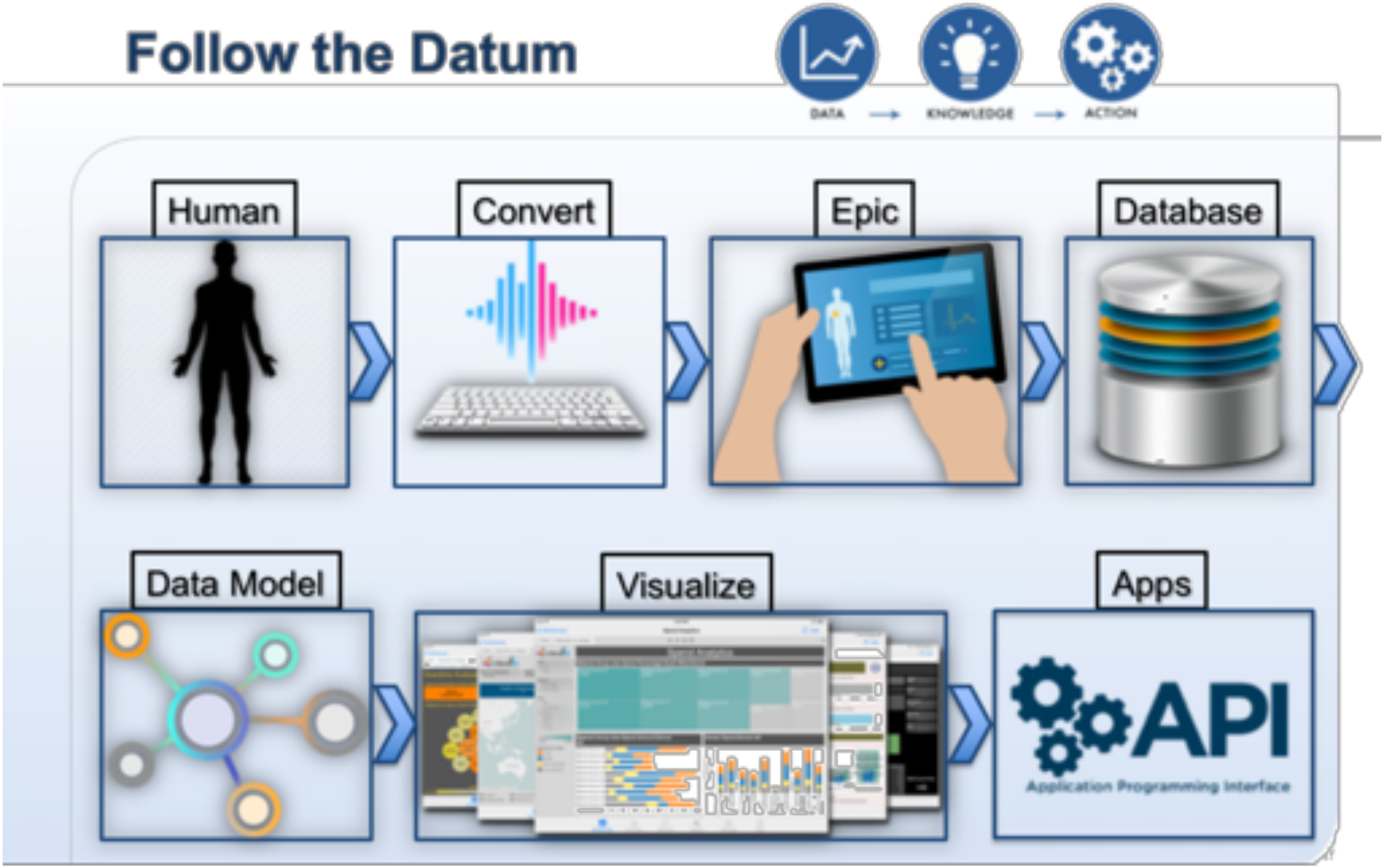
55 Data Sources
Sources Live

- EHR (Epic)
- Longitudinal record (Cerner)
- Genomics (Regeneron)
- Health Plan Claims (Trizetto)
- Radiology (PACS, Speech)
- Cardiology (EKG, Echo)
- Oncology (Oncolog)
- Pathology (Copath)
- Pulmonary (Breezesuite)
- Lab (Sunquest)
- Health Exchange (KeyHIE)
- Patient survey (DataStat)
- Secure text (TigerText)
- Real time tracking (Teletrack)

4.4M Million
Unique patients on platform

- **595k** Health Plan member
- **22,900** Active Providers
- **1M** Surgical cases
- **198M** Encounters
- **300M** Clinical notes
- **4M** Pathology specimens
- **2.4M** Patient billed
- **89M** Encounters billed
- **75.4M** Health Plan claims

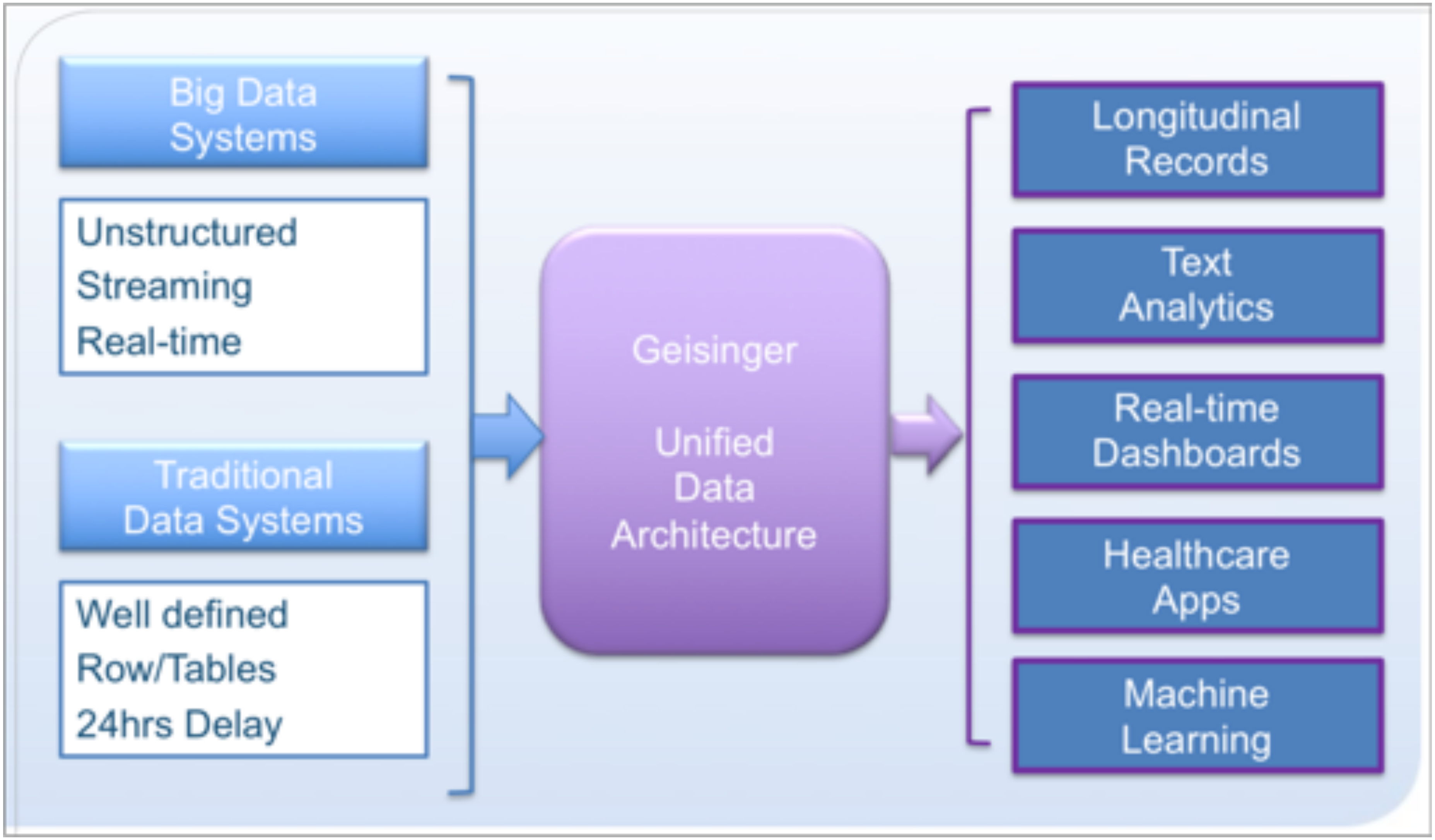
Follow the Datum



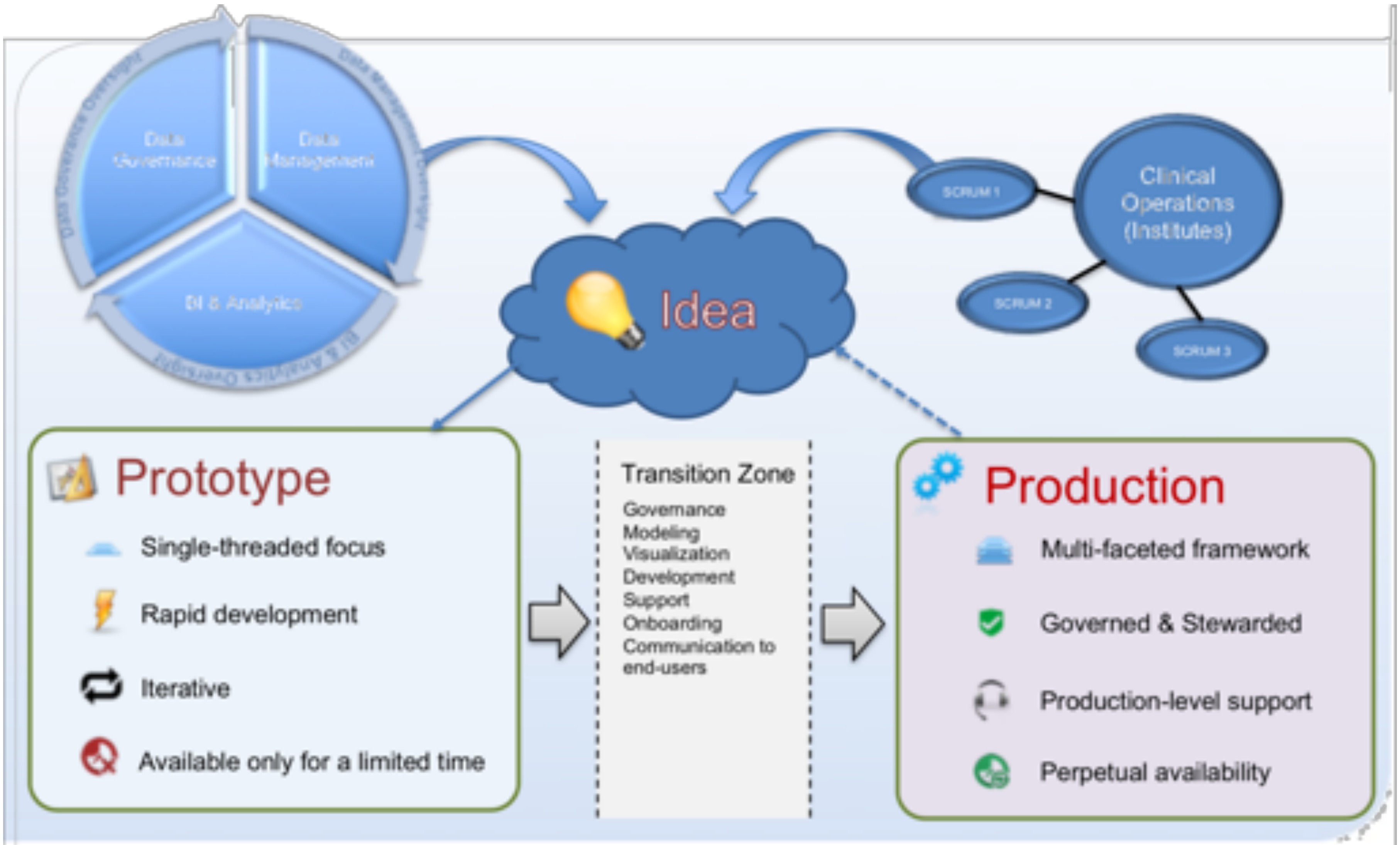
A short history of analytics

- **Paper**
  - Excel
- **EHR**
  - Vendor specific
- **EDW**
  - Relational
- **UDA**
  - Hedged hybrid

“data lake” AKA Unified Data architecture

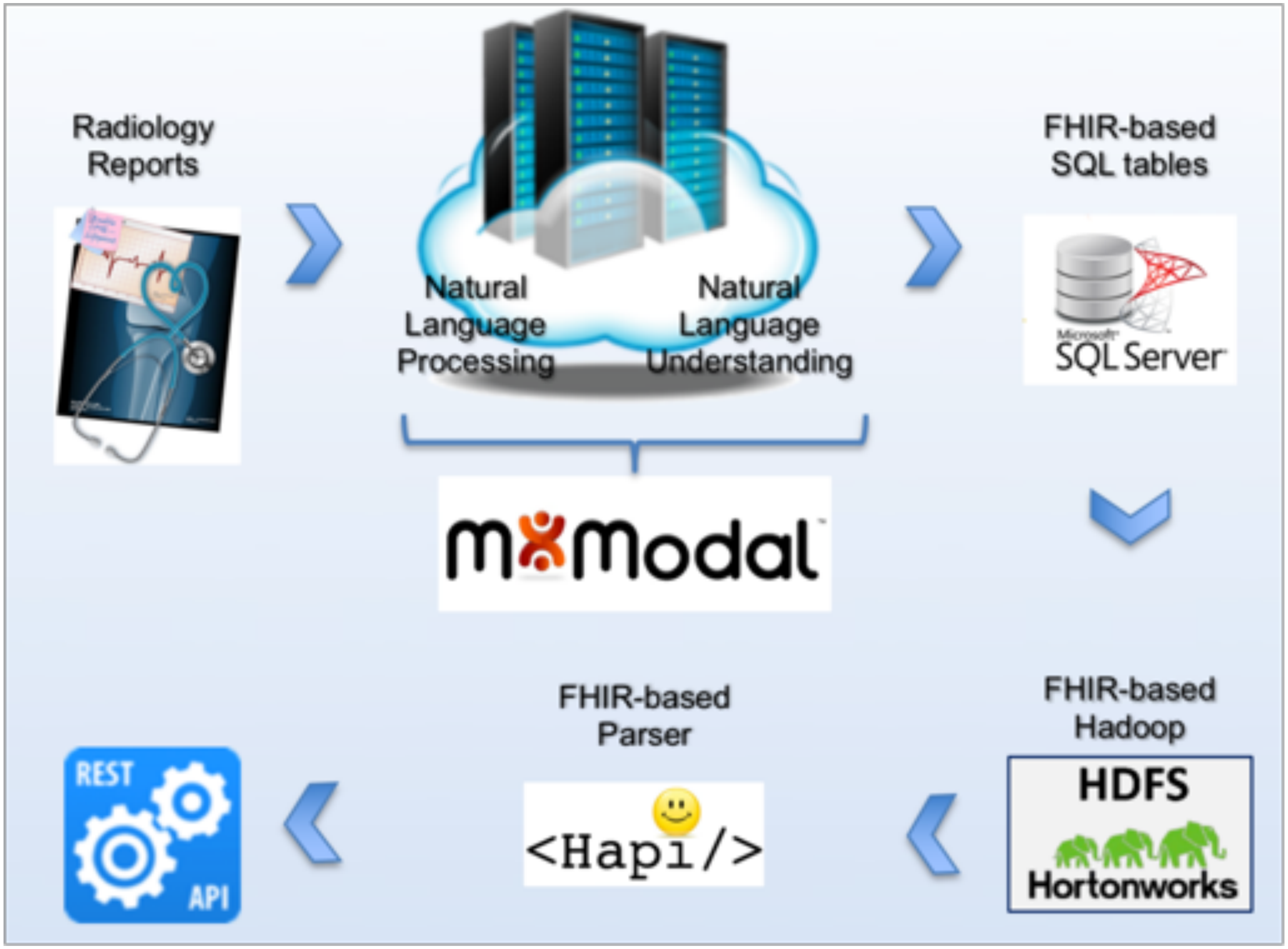


Hub and spoke – different than what you might think (AGILE)



Visualize and reporting

Understand constraints

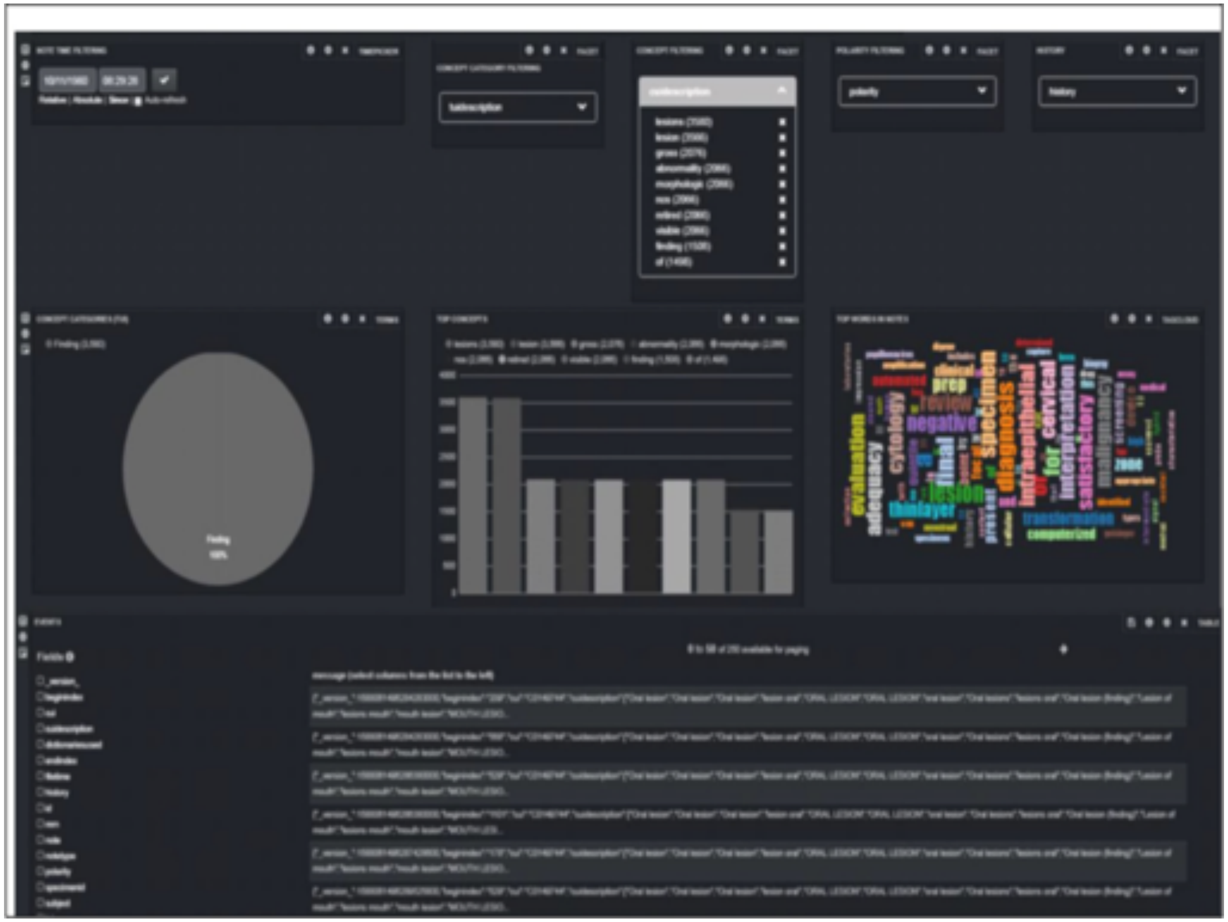


Enhance clinical activities

- AAA “close the loop” program
- Bundled payment assessment in real-time
- Smartphone text analytics
- “check please” OR process cost
- Actual vs scheduled OR cases
- Prioritizing CT-Scan using Artificial Intelligence bot

Longitudinal record

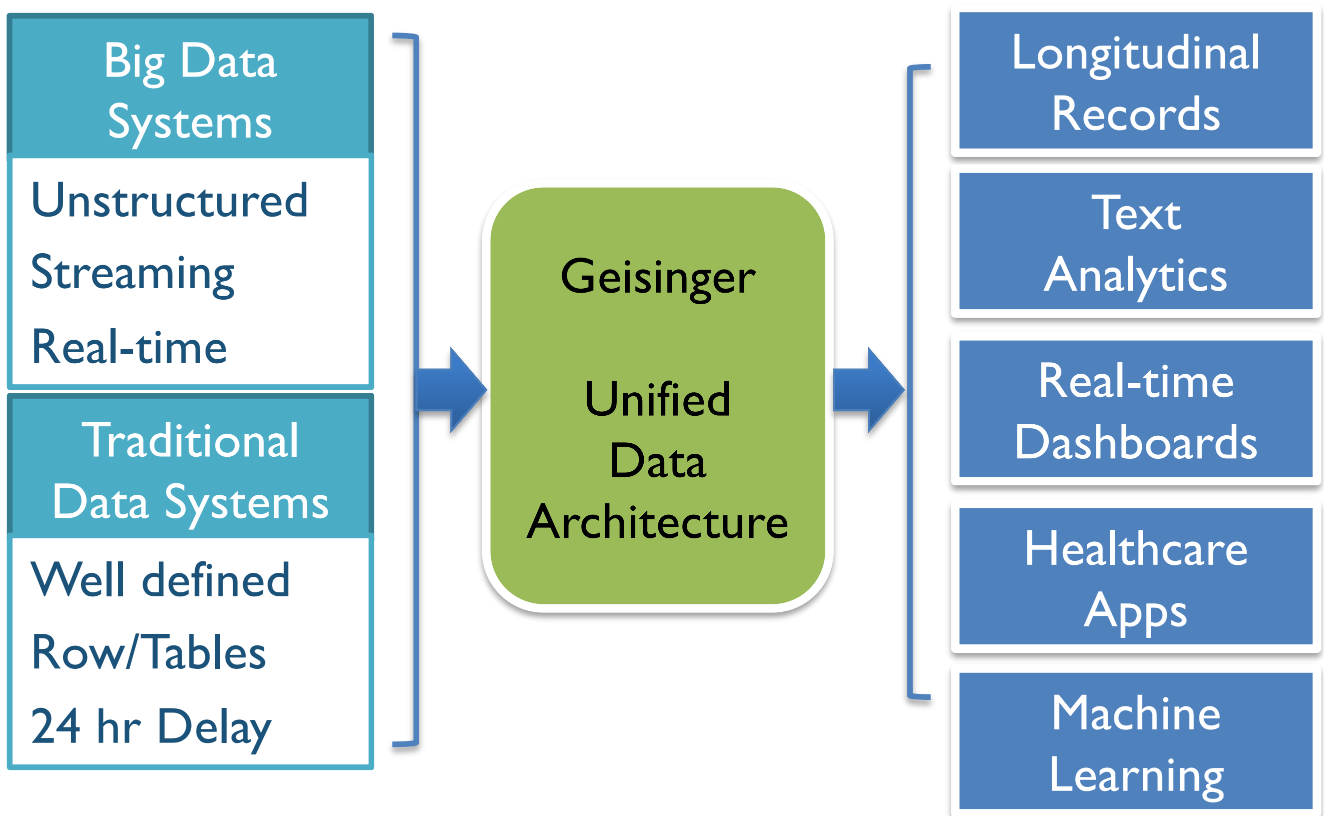
Search – Unstructured (pre-annotated) data



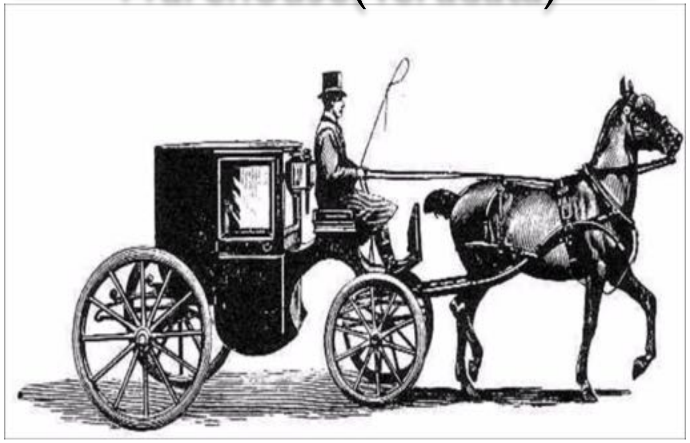
- Notes Annotated for Conditions
- Facets Based Search –
  - Category (Findings, body part etc.),
  - Concept (Lesions, cancer etc.),
  - Polarity (positive, negative)
  - History (yes, no)

- Relevant Results- 100’s of results instead of 1000’s

Unified Data Architecture(UDA)



The Differences?

Traditional Data Warehouse(Teradata)	UDA
	
Features: Structured Data Support, Longitudinal Records **Cost**: $5 Million, $600K yearly **Storage & Data:** 12 TB & 15 Sources	**Features:** Structured & Unstructured Data Support, Longitudinal Records, Real-time data, Machine Learning(ML),NLU, AI **Cost**: $1 Million, $350K yearly **Storage & Data:** 200 TB & 45 Sources

Longitudinal Record example : CheckPlease!

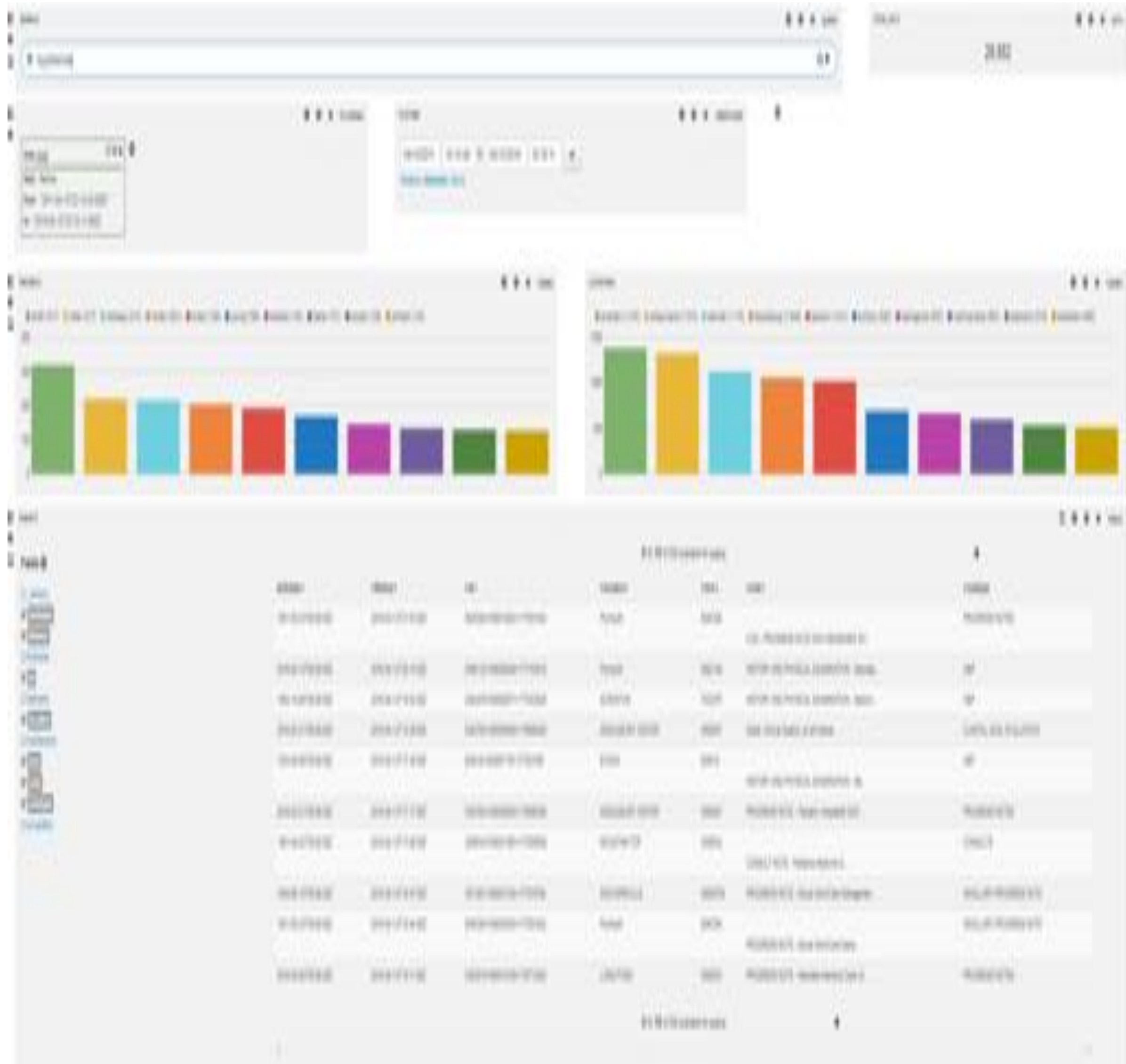
Longitudinal Record Radiology TurnAround Time

Longitudinal Record Sepsis 3 and 6 hour bundles

- Cohorts
  - Patients having diagnosis or DRG of septicemia (for training)
  - Patients coming into ED with a possible infection (for prediction)

Search Unstructured Data

- 300 Million Notes
- Refreshed Daily
- Sub-second Query Time



Search Unstructured Annotated Data

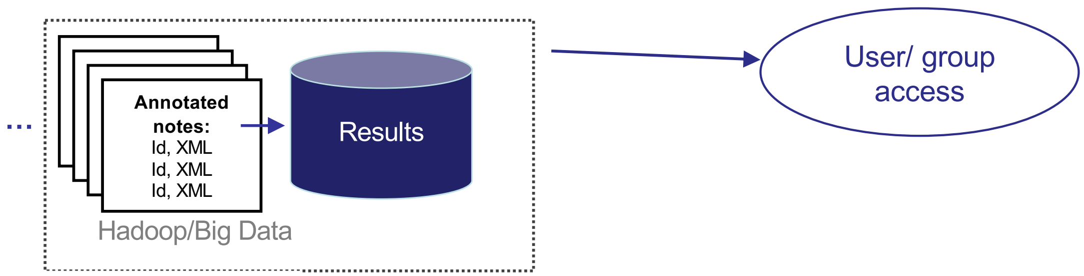
- Notes Annotated for Conditions
- Facets Based Search –
  Category(Findings, body part etc.), concept(Lesions, cancer etc.), polarity(positive, negative), history(yes, no)
- Relevant Results- 100’s of results instead of 1000’s

Natural Language Processing(NLP)

- Extract following information from the sentence describing nodule.
  - nodule sizes.
  - description keywords(e.g.: “stable”, “increase”, “large”, “new” etc.)
  - site location(e.g.: “left upper lobe” etc.)

e.g.:

“There is a lobular **pulmonary nodule** in the **right upper lobe** measuring **11 x 10 mm** (image 65/291), previously measuring **9 x 6 mm**, mildly **increased** in size.”



Results

MRN	ENCOUNTER_ID	Note time	RADS Score	length	width	units	description	location	Is Past Reading?
1	22	12/19/2016 13:20	LUNG RADS 4B	11	10	MM	“increase”	Right upper	No
2	34	11/18/2016 22:33	LUNG RADS 1	3	0	MM	“tiny stable subcm”	Left lower	No

Natural Language Processing(NLP) AAA & Lung Nodules

Lung Nodules

10 million
radiology notes

300,000
notes with nodules

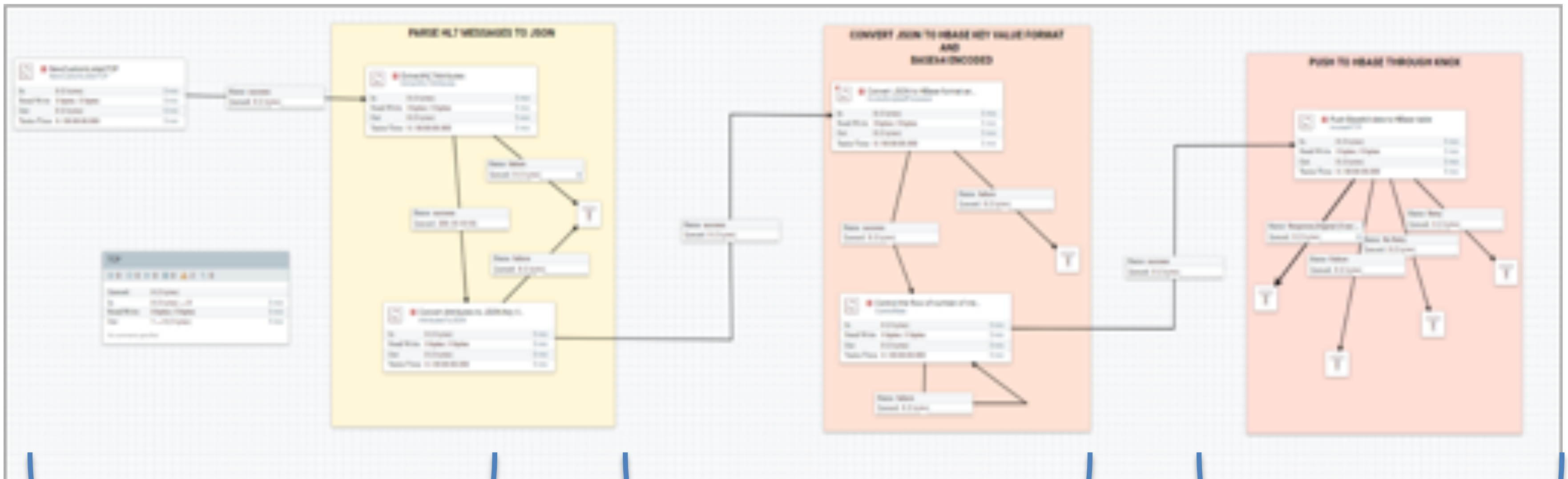
210,000
unique patients

82,594
care gaps

AAA

23
Lives Saved

Data From Data Captor Devices



Capture

Process

Store and Report

AI- Machine Learning, What is It?

Types of Learning

- Branch of Artificial Intelligence (AI)
- Strategic approach to expanding organizational architecture – advanced analytics

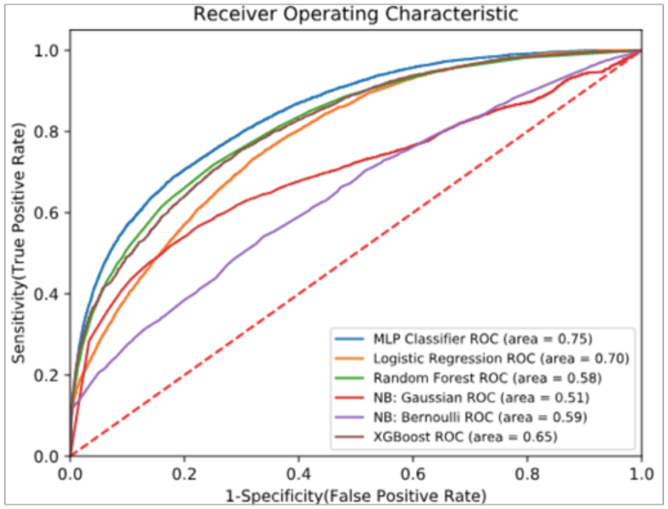
Supervised (Structured)	Guiding outputs by tagging inputs with the desired outcomes
Unsupervised (Semi-structured)	Recognition of patterns that meet the desired outcome
[illegible]	[illegible] outcomes

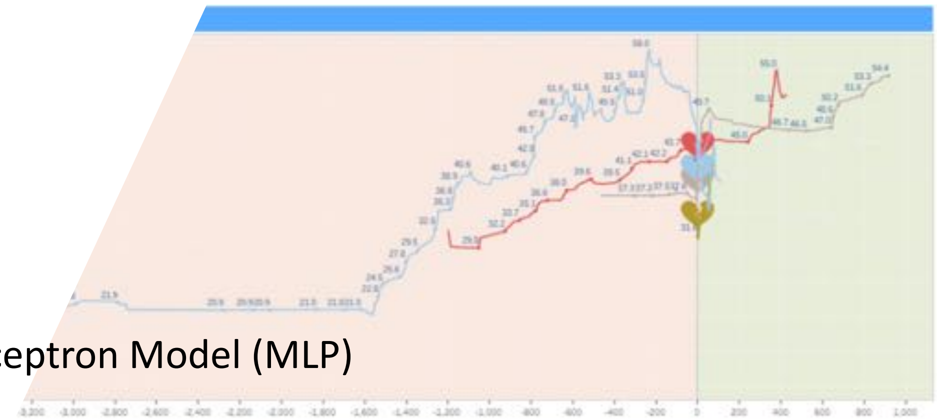
Machine Learning - Sepsis to Septic Shock Prediction

- Cohorts
  - Patients having diagnosis or DRG of septicemia (for training)
  - Patients coming into ED with a possible infection (for prediction)
- Prediction of patients
  - Sepsis to septic shock
  - Time interval between sepsis to septic shock
  - Confidence score on the prediction
- Features for prediction
  - SIRS (systemic inflammatory response syndrome) – vitals and labs
  - Organ dysfunction – vitals and labs
  - Source of infection – patient notes
  - Antibiotics time
  - End time of Fluid Bolus
  - Hypotension Values

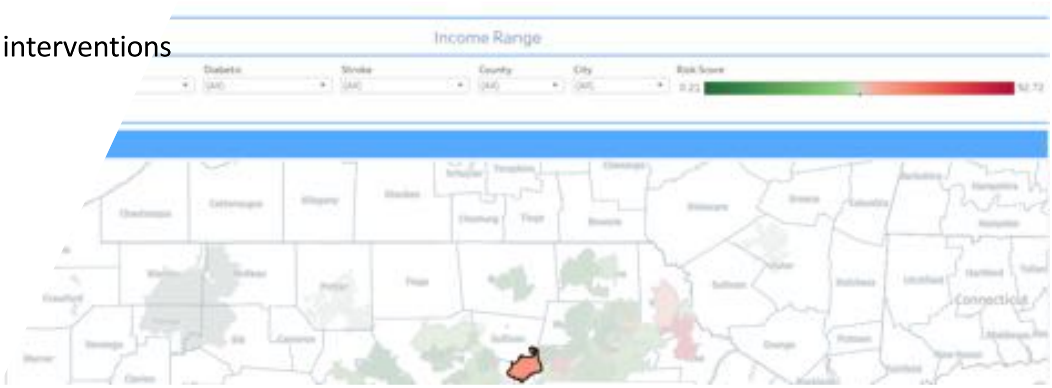
ASVCD Risk Score card

Primary/secondary risk prediction and managment





- Latest generation Machine Learning Model – Multi Layer Perceptron Model (MLP)
  - Current AUC 0.94 – up from 0.75
- Based on concurrent EPIC data WITHOUT manual abstraction
- Two tiered approach
  - Give clinicians immediately actionable temporal/trending information
  - Generate geographic heatmap
  - Generate top 100 lists to identify care gaps and guide targeted interventions
    - high yield
    - Per service area



Model	TP	TN	FP	FN	Accuracy	Sensitivity	Precision	Specificity	AUC	F-Score	Kappa Coeff	Matthews Coeff
MLP Classifier	3200	69189	14920	1541	81.47	67.5	17.66	82.26	74.879	0.28	0.2134	0.2775
Logistic Regression	3299	59111	24998	1442	70.24	69.58	11.66	70.28	69.932	0.1997	0.1192	0.1923
Random Forest	740	83554	555	4001	94.87	15.61	57.14	99.34	57.474	0.2452	0.2275	0.2803
NB: Gaussian	4511	6068	78041	230	11.91	95.15	5.46	7.21	51.182	0.1034	0.0027	0.0207
NB: Bernoulli	1470	73434	10675	3271	84.3	31.01	12.1	87.31	59.157	0.1741	0.1054	0.1198
XGBoost	1239	82680	1429	3502	94.45	26.13	46.44	98.3	62.217	0.3345	0.3079	0.3218


Caring