KING'S HEALTH PARTNERS

NIHR | Maudsley Biomedical Research Centre

King's College Hospital NHS Foundation Trust

KING'S College LONDON

NHS Guy's and St Thomas' NHS Foundation Trust

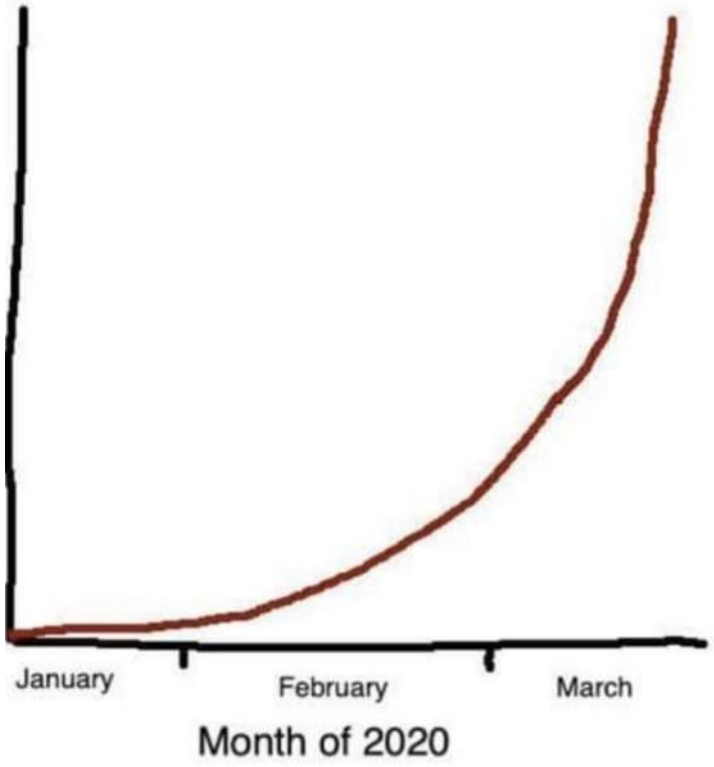
Healthcare technologies to improve Efficiency

Prof James Teo
Clinical Director of Data & AI
Professor of Neurology

Kings College Hospital
Guys & St Thomas Hospital

~~Time spent looking at exponential graphs~~

Amount of medical data



PROBLEMS

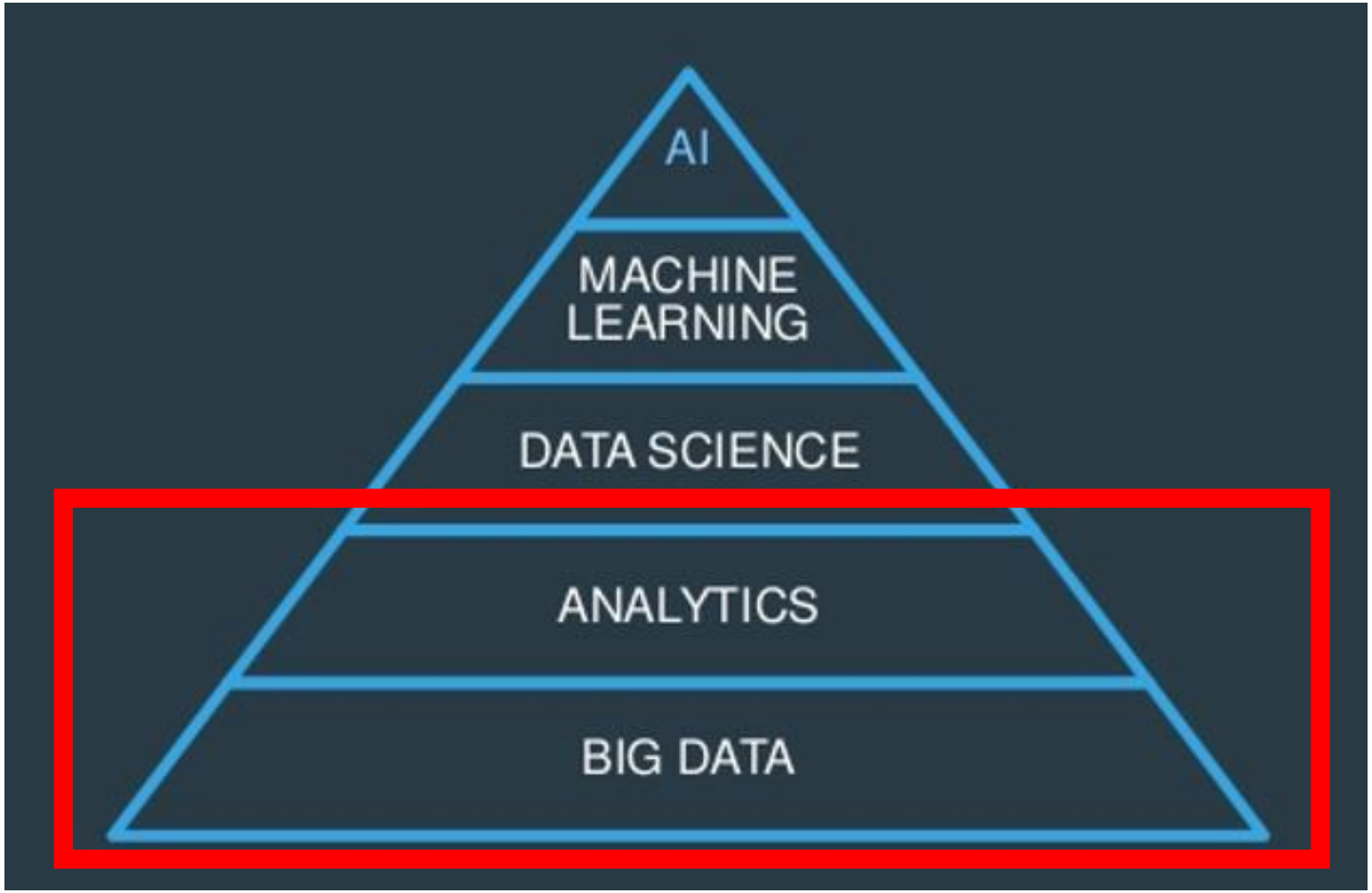
Too much health data

Health data is all over the place

Health data is not standardised

Data is stored inflexibly

HIERARCHY OF NEEDS

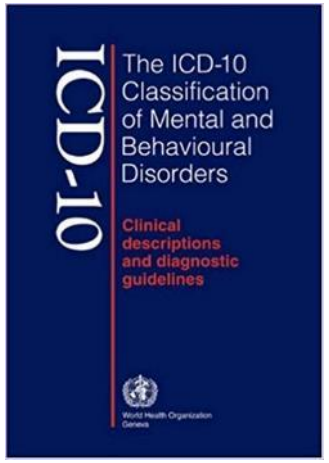


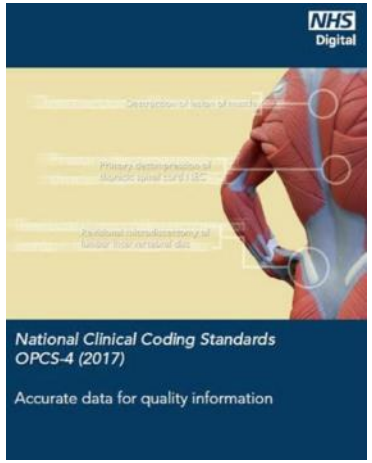
What We Want

Structured Data

Results | Messages

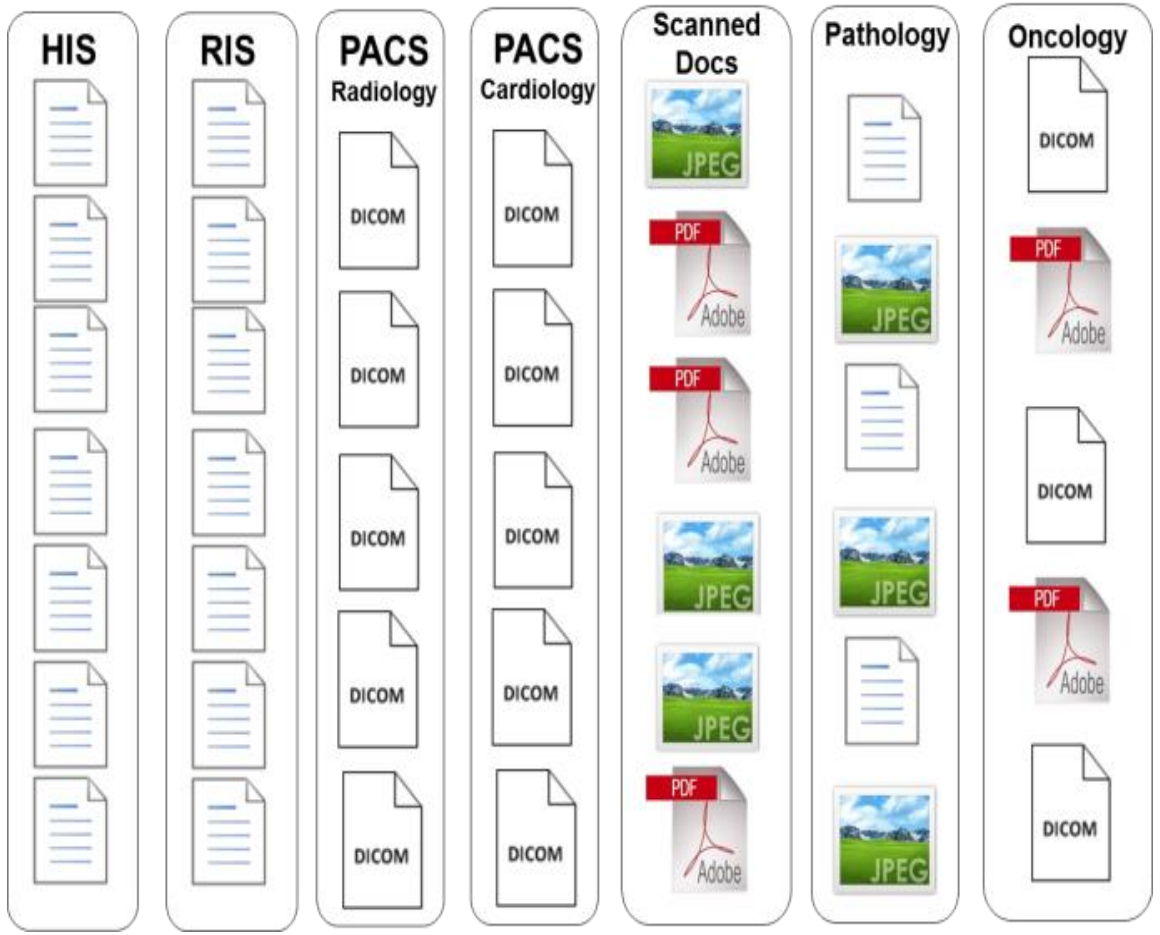
	Date	Time	TargetTemp	ActualTemp	SystemID	SystemAge	BuildingID
1	6/4/13	4:13:19	65	64	1	11	14
2	6/5/13	5:13:19	66	66	17	30	16
3	6/6/13	6:13:19	68	70	8	27	10
4	6/7/13	7:13:19	70	70	13	25	4
5	6/8/13	8:13:19	65	62	11	29	12
6	6/9/13	9:13:19	67	65	16	12	16
7	6/10/13	10:13:19	66	64	11	19	13



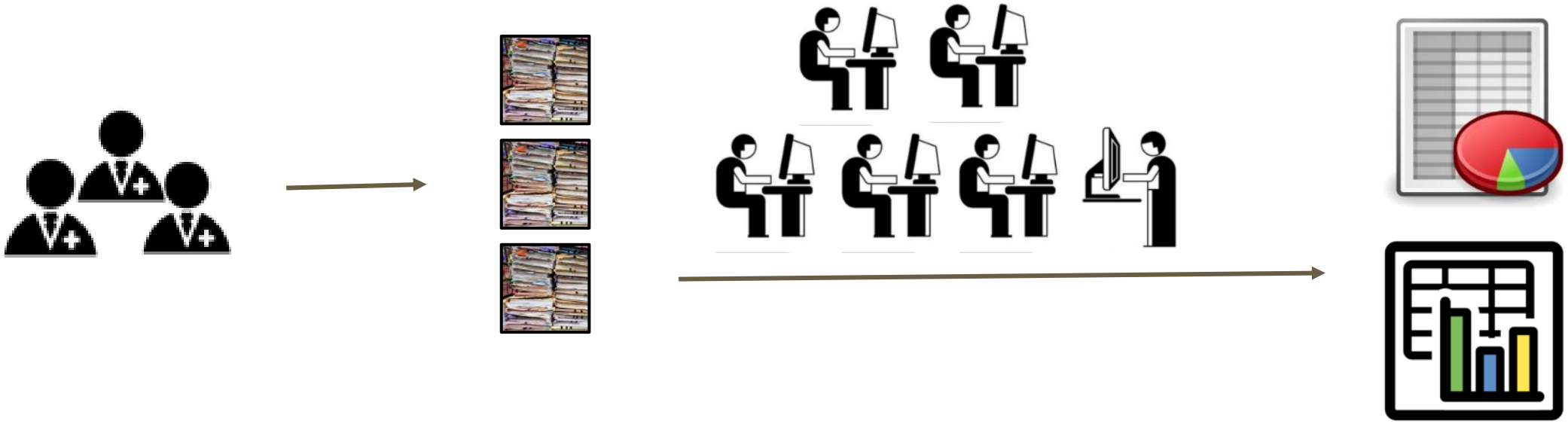


What We Have

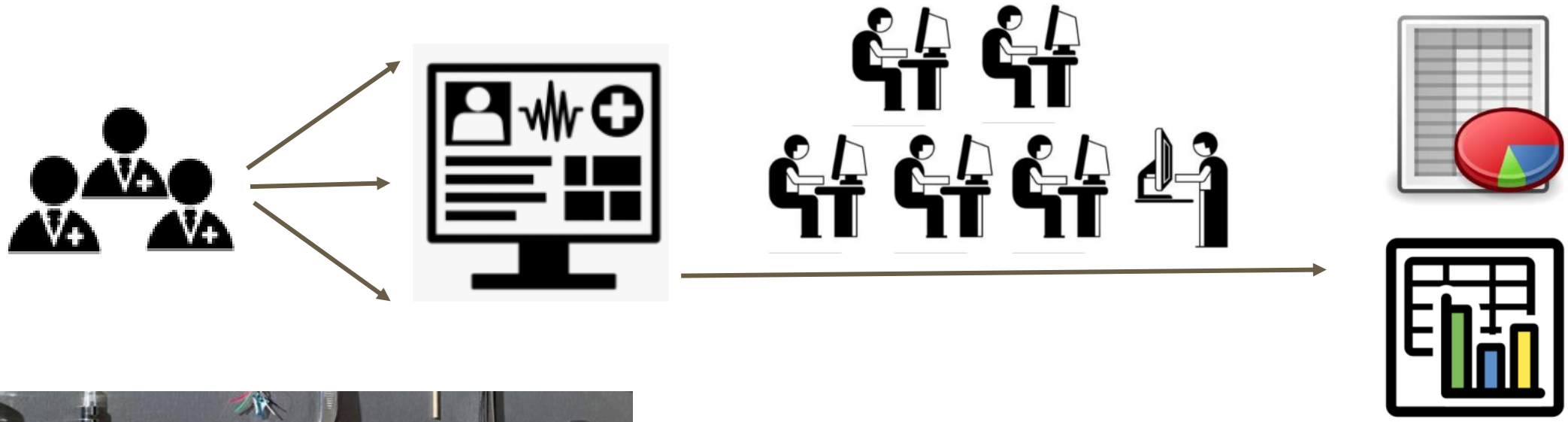
Unstructured Data

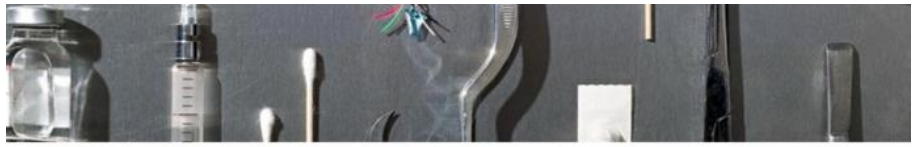


The original way of producing and cleaning data



Add a new Electronic Health Record System?





Death by a Thousand Clicks: Where Electronic Health Records Went Wrong

The U.S. government claimed that turning American medical charts into electronic records would make health care better, safer, and cheaper. Ten years and $36 billion later, the system is an unholy mess: Inside a digital revolution gone wrong. A joint investigation by Fortune and Kaiser Health News.

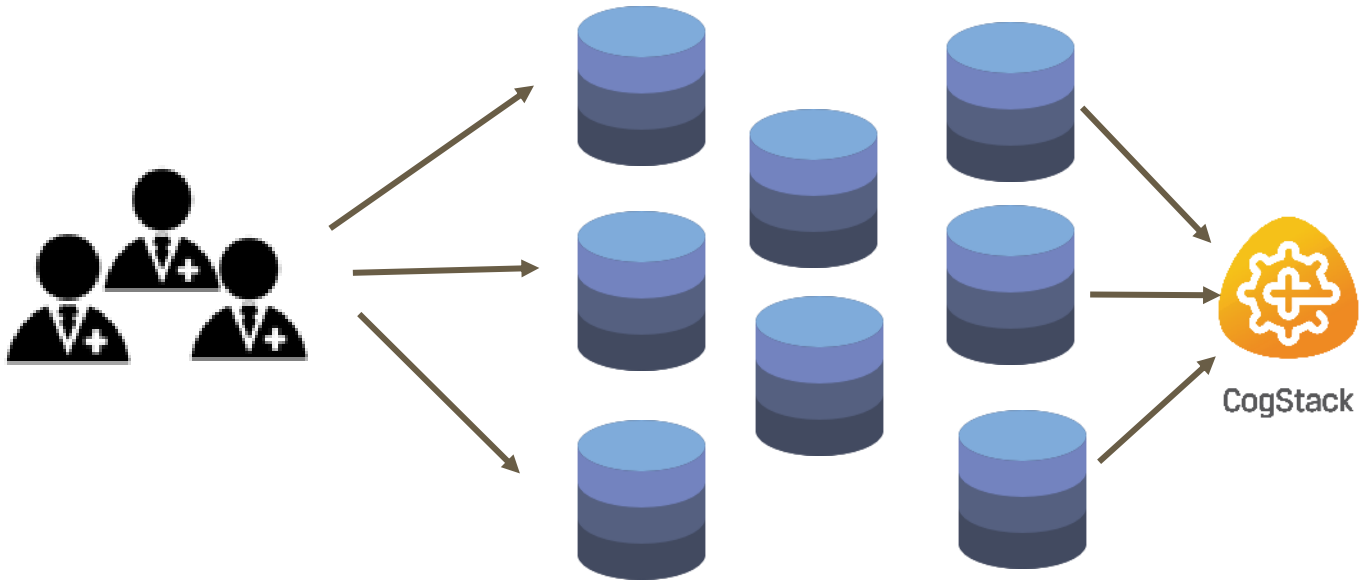
The Journey of Kings and Guys & St Thomas

BMC Medical Informatics and Decision Making

Home About Articles In Review Submission Guidelines

Software | Open Access | Published: 25 June 2018

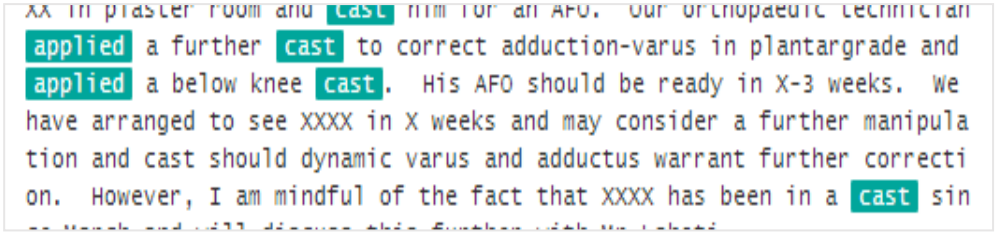
CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital



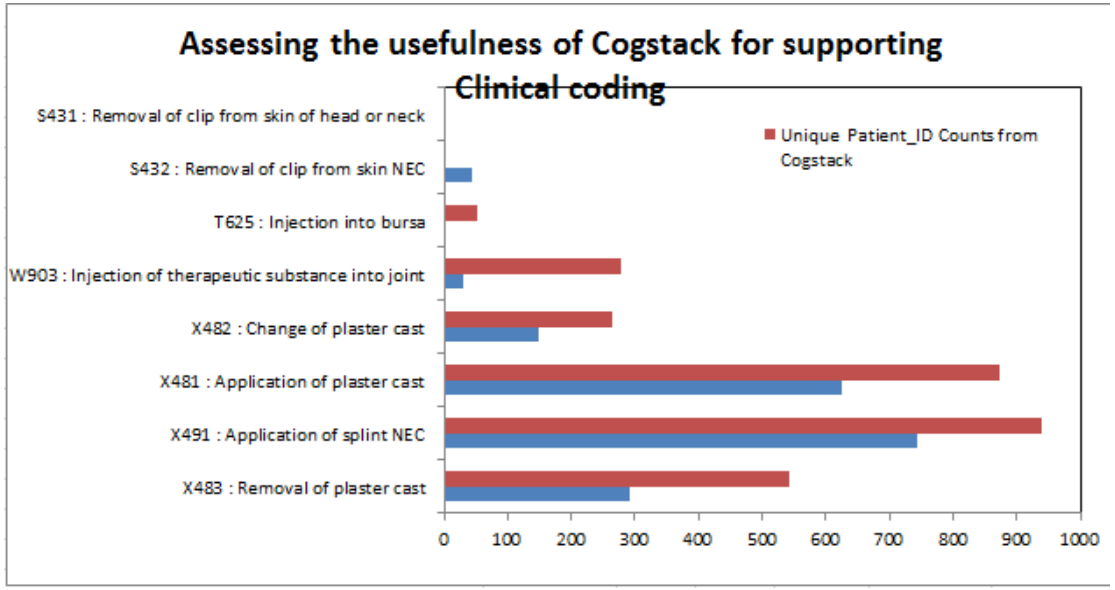
Heterogenous Clinical Record systems

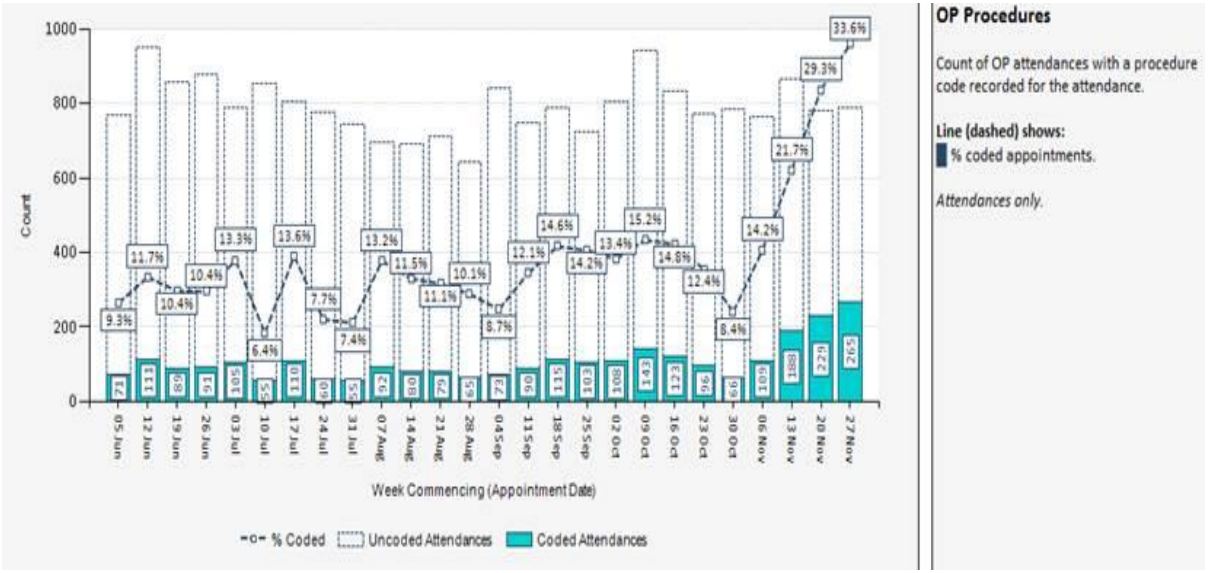
Scoping the level of under-coding found in unstructured text

Mismatch of patients having OPD procedures documented in letters (RED) versus what is coded on appointment systems (BLUE)



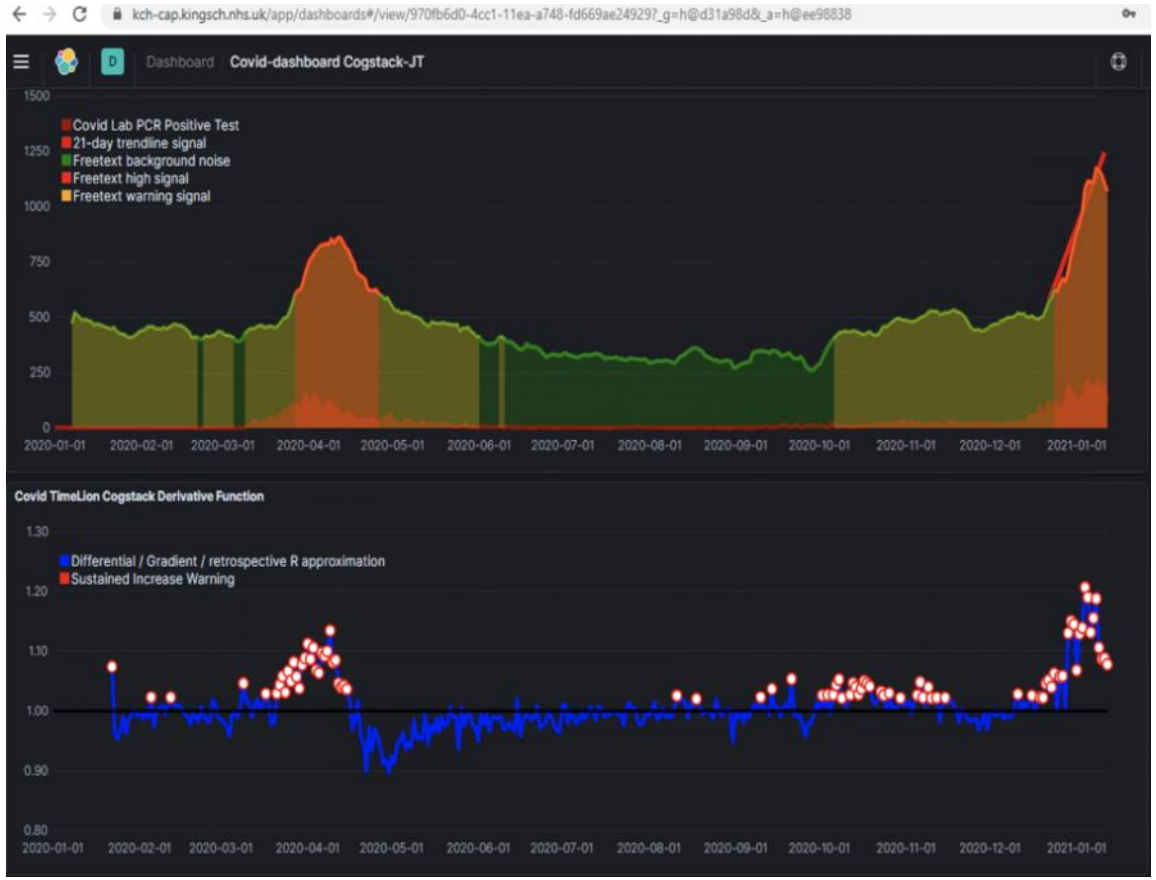
2x increase in coding in just 3 weeks





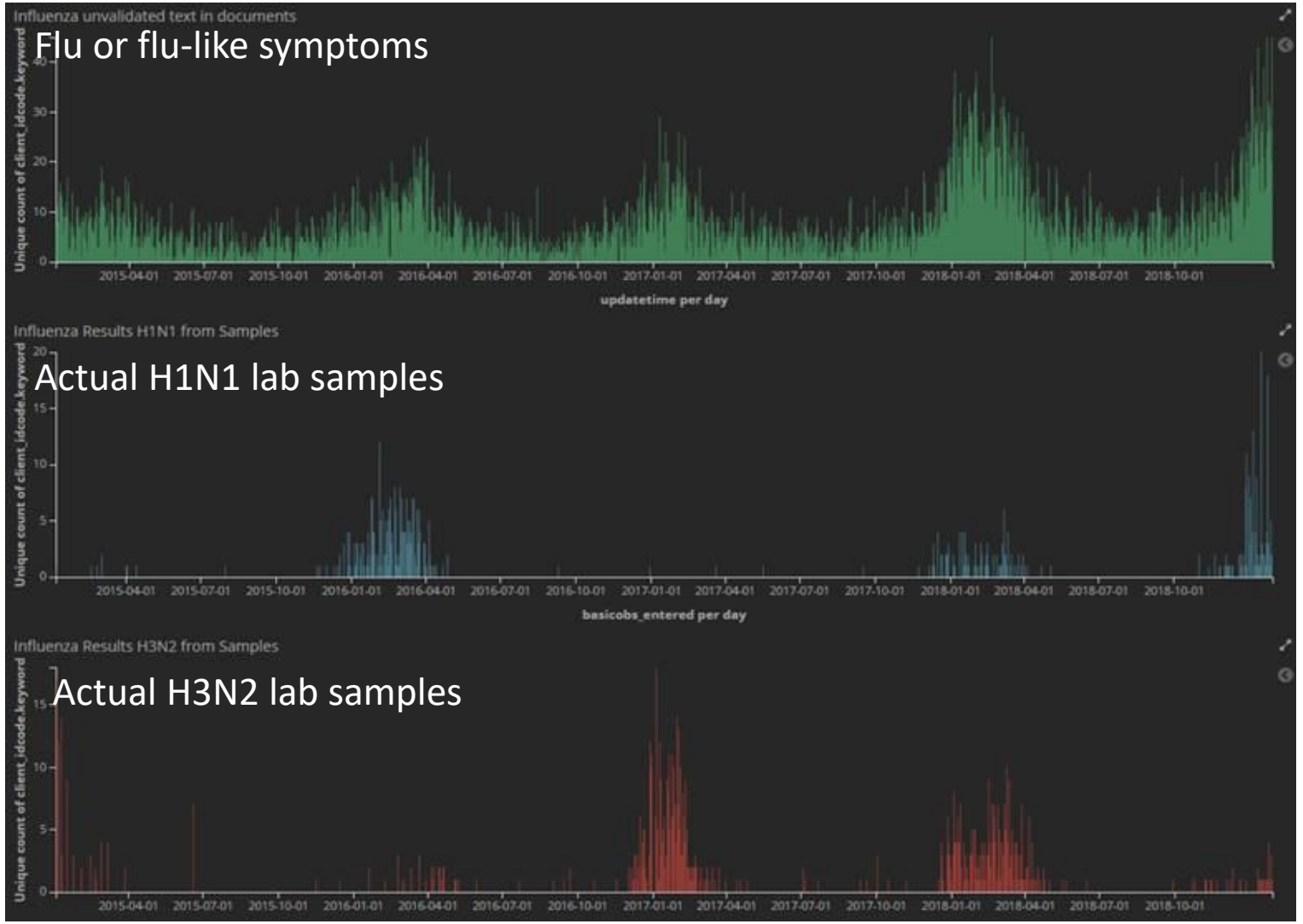
Streaming unstructured data feeds of the hospital

KCH (900 beds)
PRUH (600 beds)



Guys & St Thomas (850 beds)

Real-time Data feed of flu or flu-like illness



Human-lite Data Processing Pipeline

Pooled → **Structured** → **Semantic** → **Rationalised** → **Applied**

Pooled	Structured	Semantic	Rationalised	Applied
Data needs to be ‘laked’ in one place “Data lakes” Heterogeneous data types, agnostic	Structured form entry Metadata automation Avoid overdesigning structure	Terminologies (SNOMED CT) Attach meaning to the unstructured data	Reduces the noise for a more specific purpose, and applies expert knowledge De-duplicated Contextualised Mismatch adjudication	Supports and enhances clinical applications: ·Clinical Audit ·Analytics - AI

Natural Language Processing

- Designed for free-form digital text
- Language AI for common languages already very mature
- Clinical language NLP is in rapid development and past proof-of-concept phase

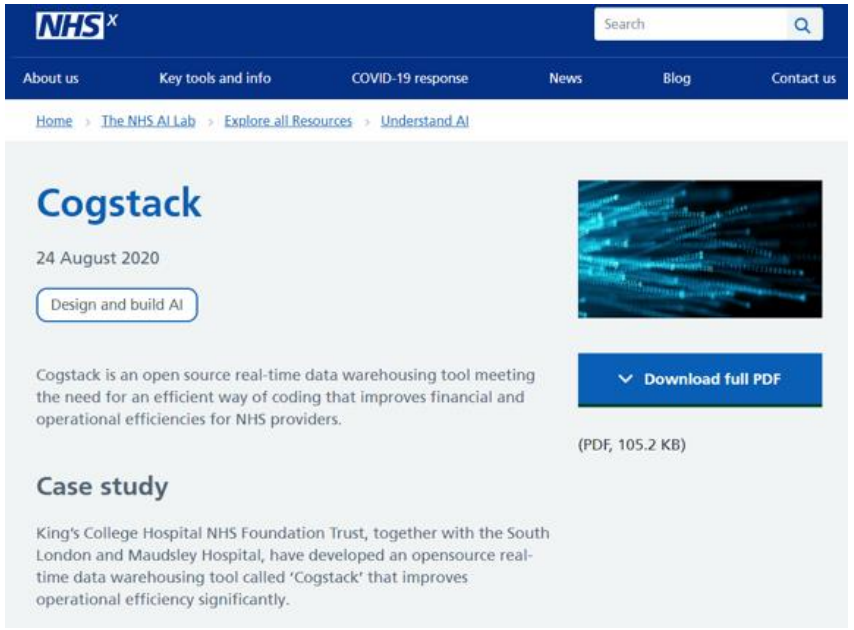
CogStack

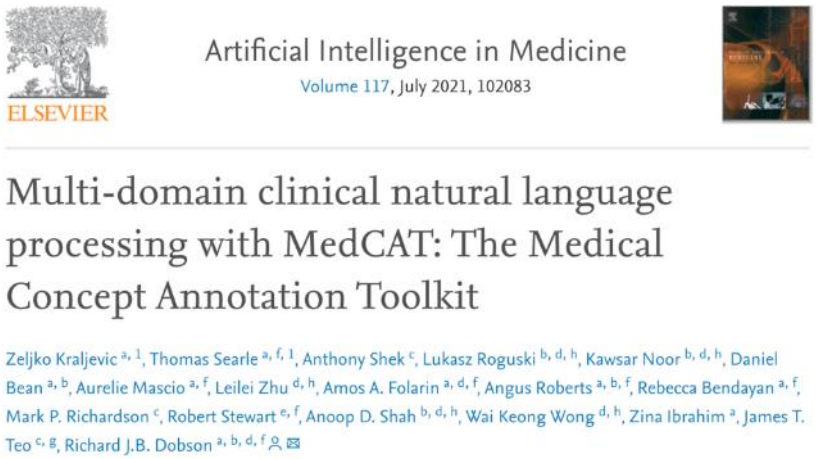
AI-based language processing

Documentation (Wiki):
https://cogstack.atlassian.net/wiki/spaces/COGDOC/overview

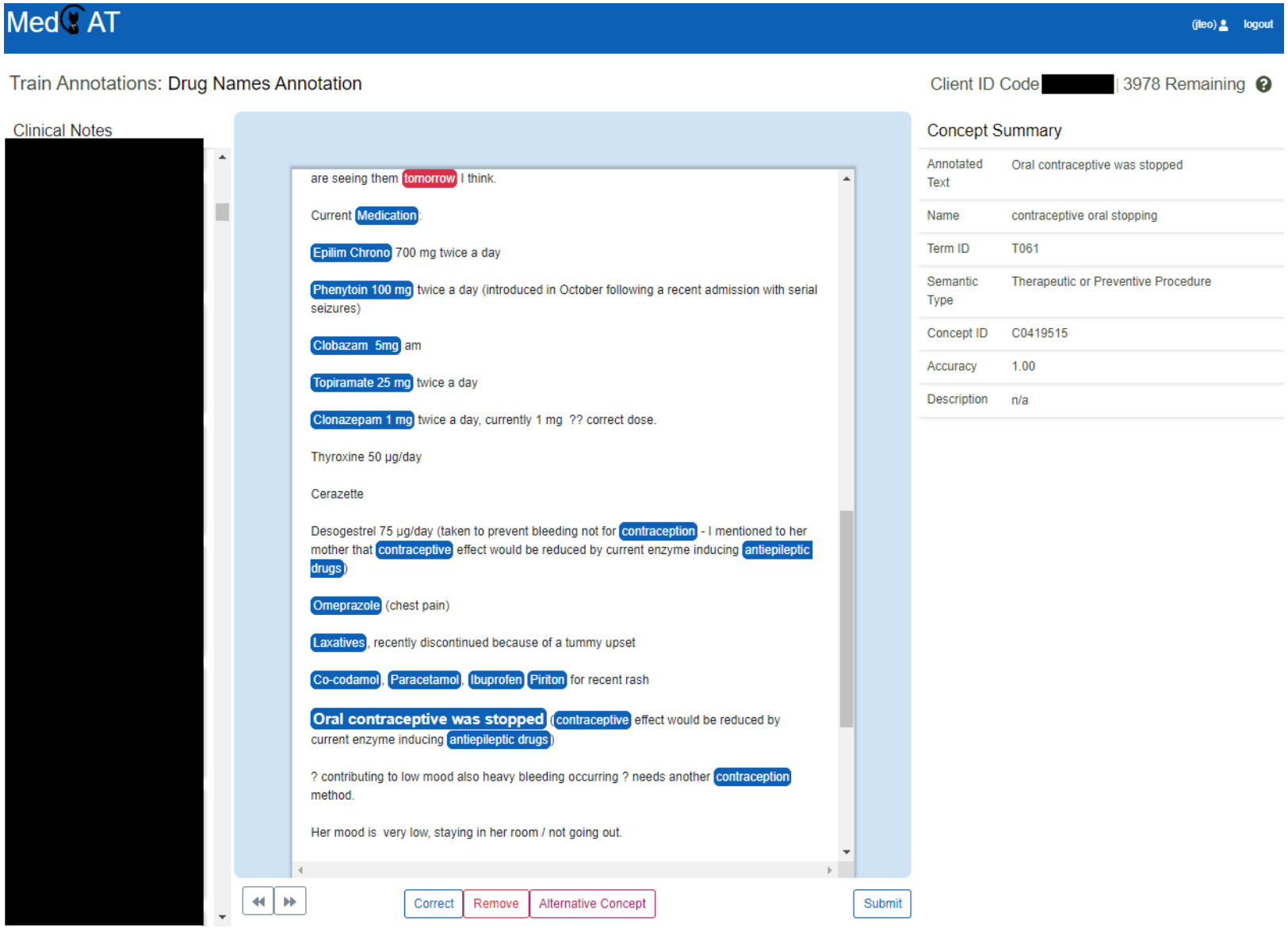
Slack:
cogstack-dev.slack.com

GitHub:
https://github.com/CogStack/

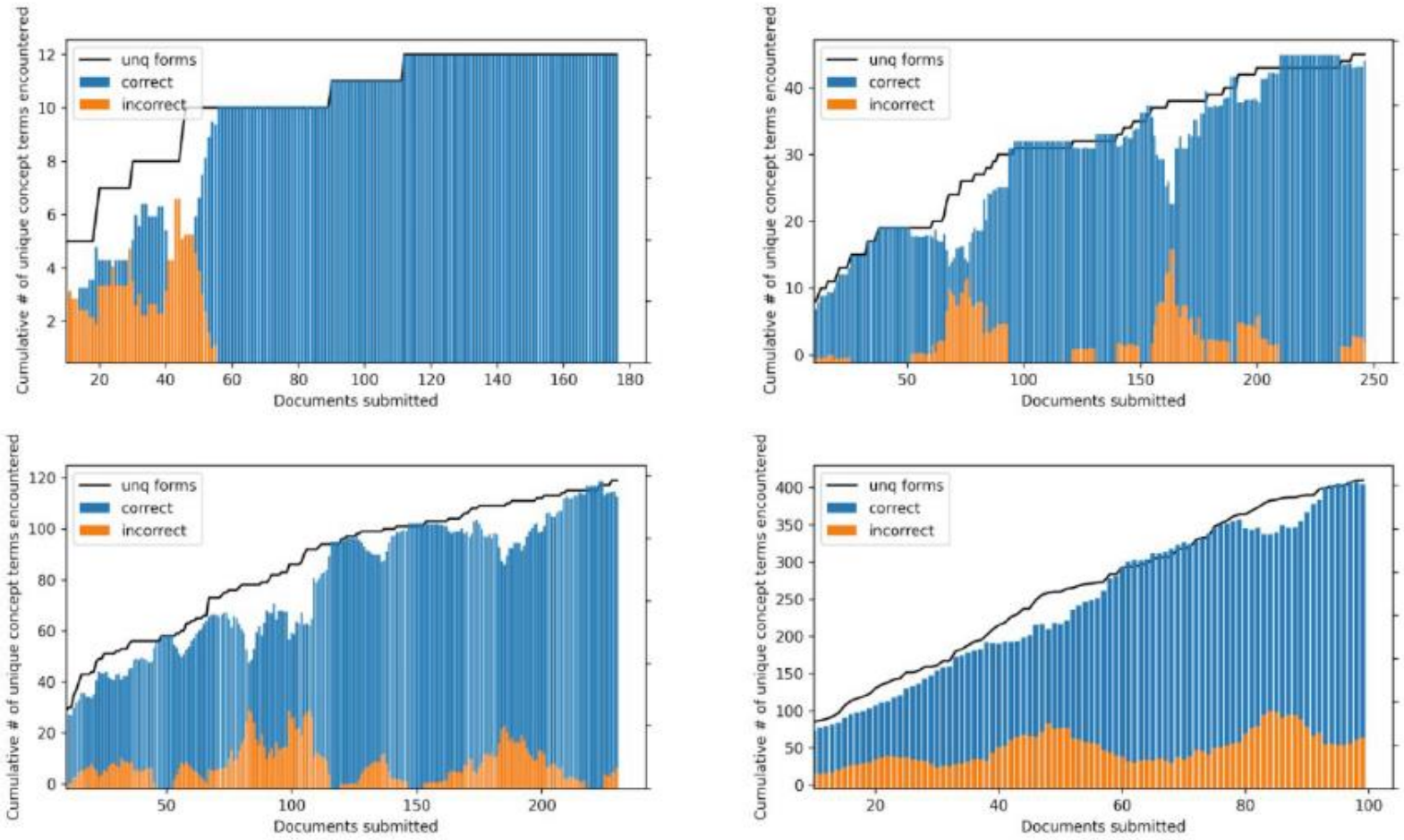




Training AI to recognise medical words and sentences

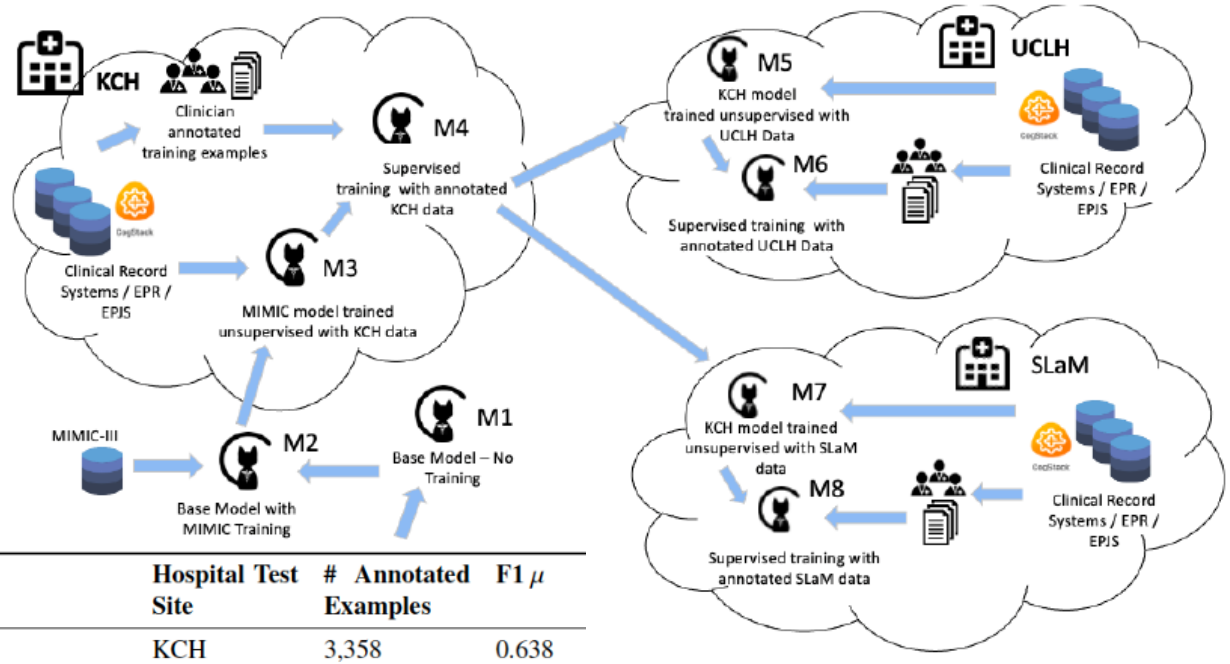


Training and testing AI across many diseases



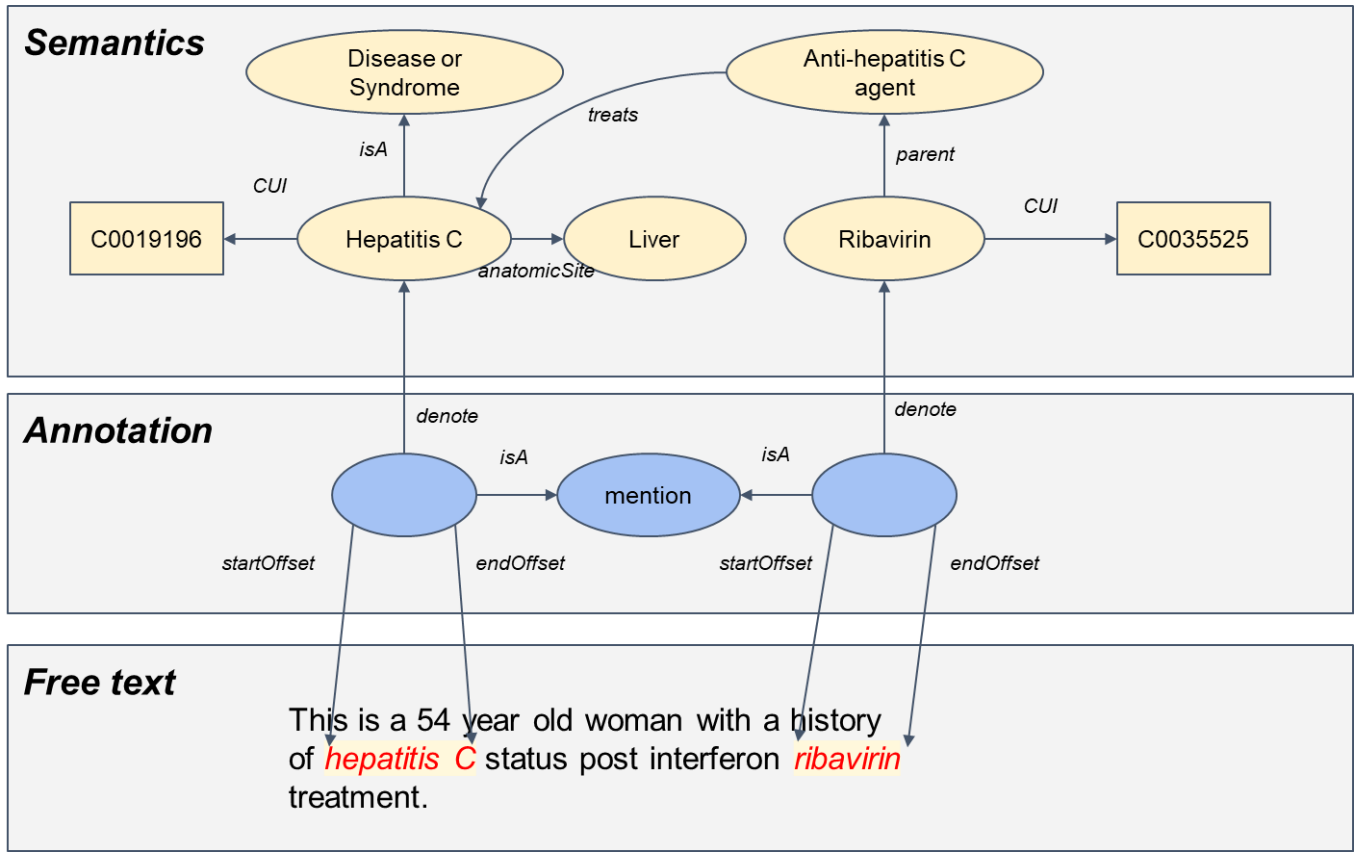
Top left to bottom right: MedCATtrainer annotation projects with respective numbers unique concepts seen throughout annotating and the number of configured concepts: Covid_COPD (5/2012), Covid_Gastro (8/679), Diabetes_Covid (15/864), Covid_CTPA_Reports (194/297280)

Training and testing AI by rotating through hospitals



Model	Training Configuration	Hospital Test Site	# Annotated Examples	F1 μ
M1	Base - No Training	KCH	3,358	0.638
M2	Base + Self-Supervised MIMIC-III	KCH	3,358	0.840
M3	Base + Self-Supervised KCH	KCH	3,358	0.889
M4	KCH Self-Supervised + KCH Supervised	KCH	3,358	0.947
M4	KCH Self-Supervised + KCH Supervised	UCLH	499	0.903
M5	KCH Self-Supervised + KCH Supervised + UCLH Self-Supervised	UCLH	499	0.905
M6	KCH Self-Supervised + KCH Supervised + UCLH Self-Supervised + UCLH Supervised	UCLH	499	0.926
M4	KCH Self-Supervised + KCH Supervised	SLaM	1,425	0.885
M7	KCH Self-Supervised + KCH Supervised + SLaM Self-Supervised	SLaM	1,425	0.907
M8	KCH Self-Supervised + KCH Supervised + SLaM Self-Supervised + SLaM Supervised	SLaM	1,425	0.945

Natural Language Understanding
Mapping Language to Meaning

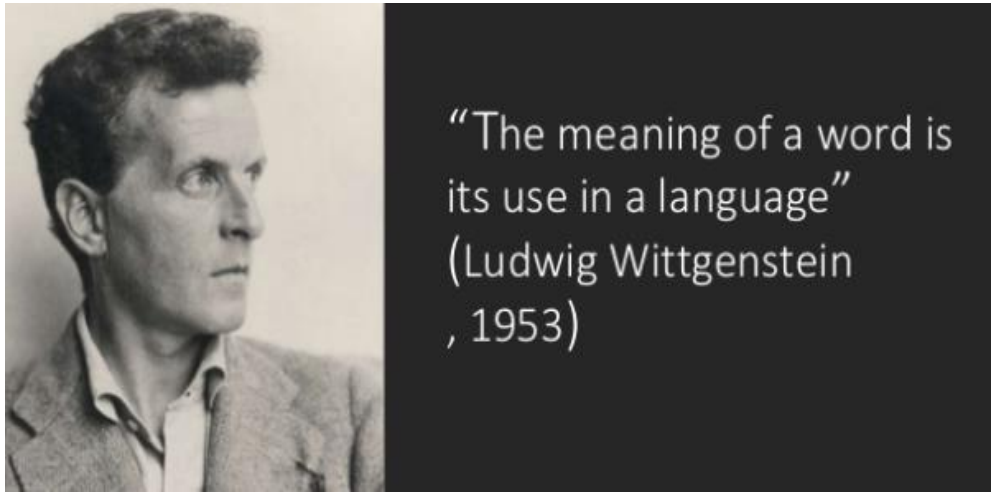


Representation of meaning (Semantics)

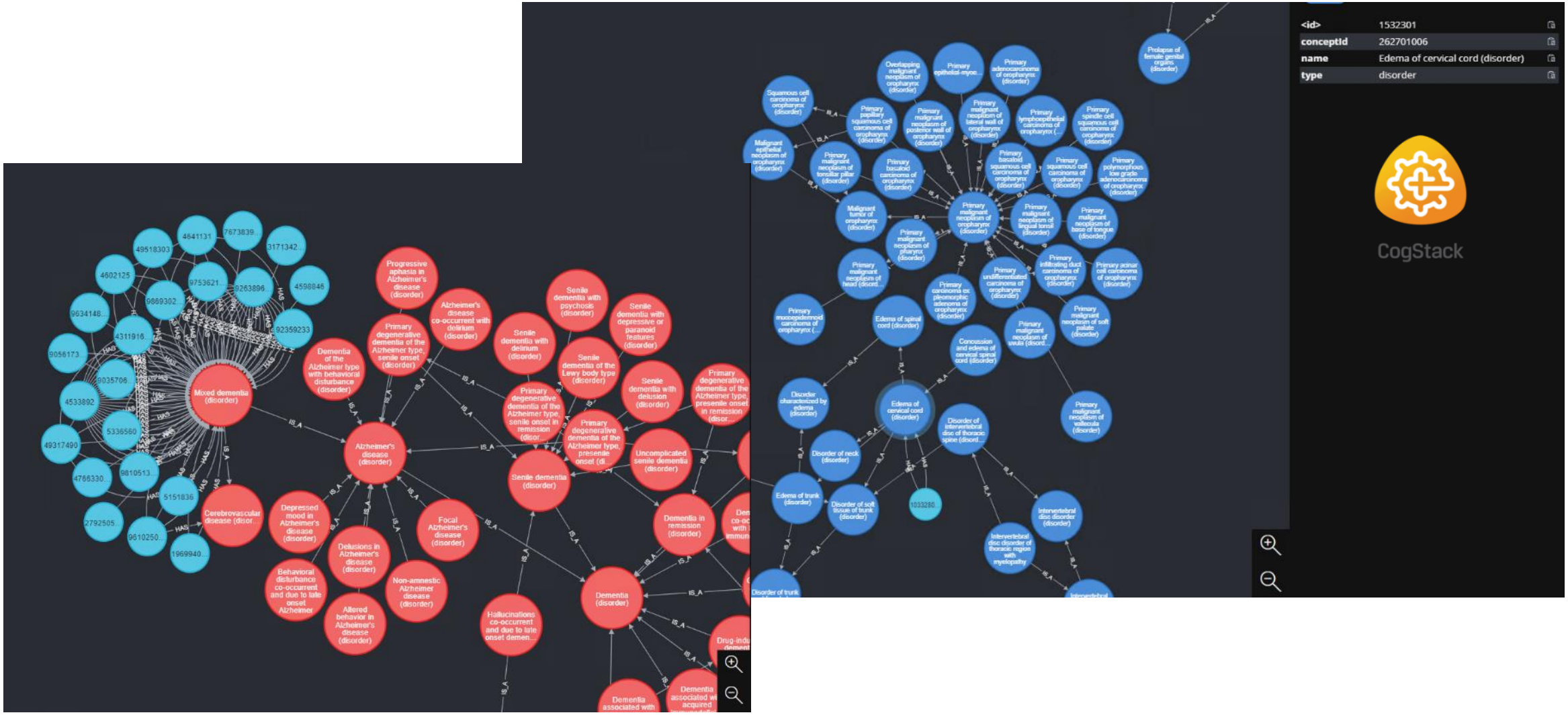
Like a dictionary but with mappings for meaning

Meaning as defined by context

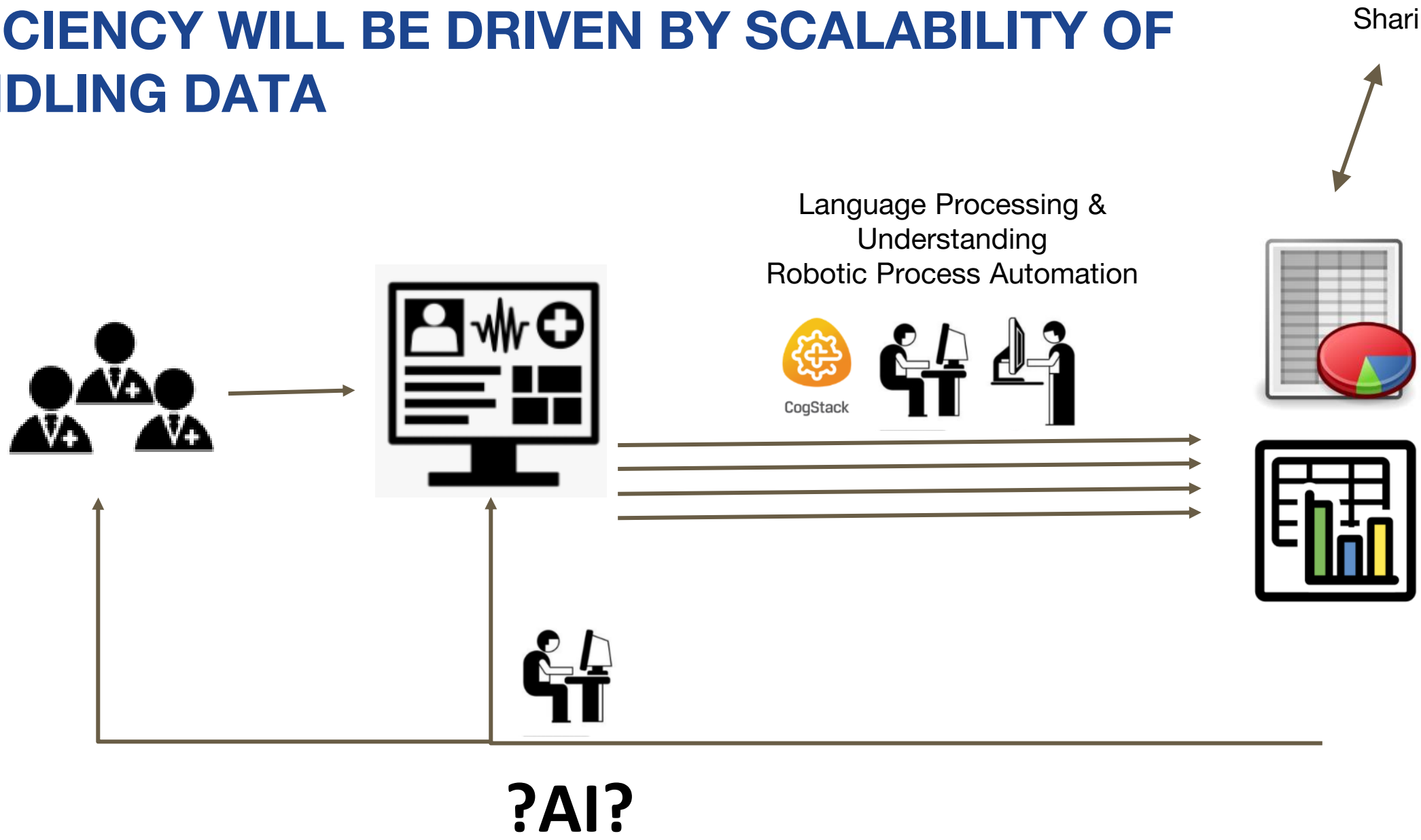
Allows machines to assign human-based meaning to words



Semantic maps to produce machine-readable meaning



EFFICIENCY WILL BE DRIVEN BY SCALABILITY OF HANDLING DATA



EFFICIENCY WILL BE DRIVEN BY SCALABILITY OF HANDLING DATA

To use data technologies, one needs to be able to clean, sort and arrange data **continuously and rapidly**

EHR's which structure data at source is helpful, but overdesigned EHR's pass the load of cleaning and structuring data to frontline staff which is the **opposite of efficiency**

NLP and Semantic technologies are past the Proof-of-Concept stage and deployments have started in some organisations

Further maturation is necessary to integrate them into workflows and then feed the data back for actionable changes

Prof James Teo
Clinical Director of Data & AI
Professor of Neurology

Kings College Hospital
Guys & St Thomas Hospital