CogStack

Foundational Infrastructure for ‘Unlocking’ Electronic Health Record (EHR) Data for clinicians, academics and population health analysis

Tom Searle, King’s College London

Agenda

The Problem: why is EHR data useful but so hard to use

The Solution: CogStack – ecosystem of NHS home-grown technologies for data analytics, ad-hoc querying and structuring EHR data.

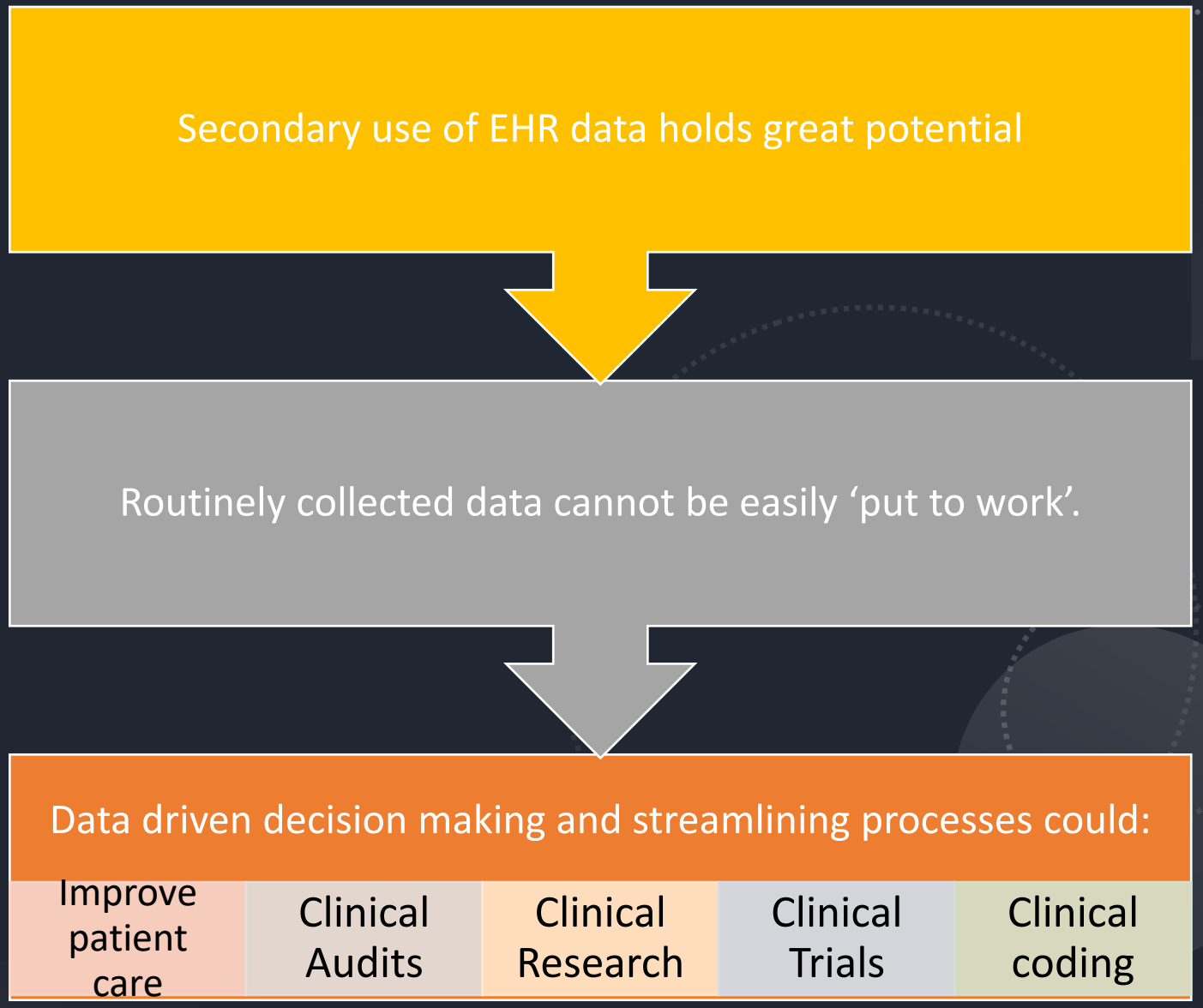
Deployment Site Case Studies:

- King's College Hospital Foundation Trust & Guy's and St Thomas' Foundation Trust
- South London and Maudsley NHS Foundation Trust
- University College London Hospitals Foundation Trust

Looking to the future:

- CogStack: NHS AI Lab (formely NHSx) AI Award
- Join our community of deployment sites.

The Problem

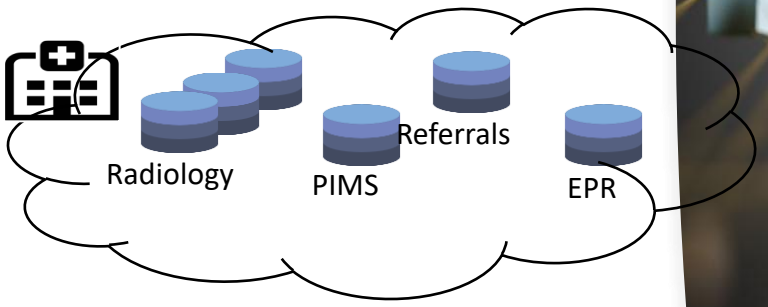

Secondary use of EHR data holds great potential
Routinely collected data cannot be easily ‘put to work’.
Data driven decision making and streamlining processes could:
Improve patient care
Clinical Audits
Clinical Research
Clinical Trials
Clinical coding

EHR Data

Data is:

- Structured (tabular)
- Unstructured (free-text narratives)

Scattered between systems



Stored in:

- different DBs,
- file formats, plain text

Clinical free-text is messy – spelling mistakes, synonyms, acronyms

The Solution: CogStack

An ecosystem of loosely coupled technologies

Open-source, developed in partnership with NHS Trusts, clinicians and domain experts.

CogStack Data: Real-time data ingestion, query layer.

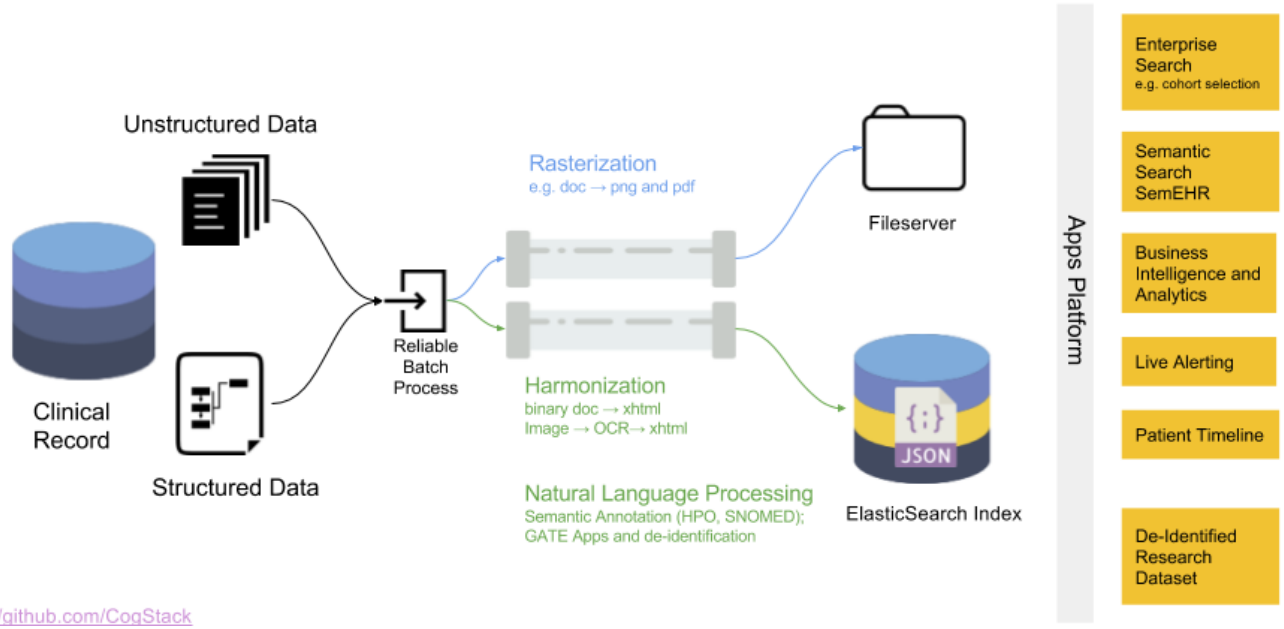
CogStack NLP: Machine learning models structure / organize free-text. Models are validated by experts, configurable, sharable.

Governance models replicated and localized according to needs.

CogStack: Data Availability

Real-time, fault tolerant, flexible, batch processing architecture

Offers:
- Querying across millions of records in seconds
- Alerting
- Visualisations
- Dashboards



https://github.com/CogStack

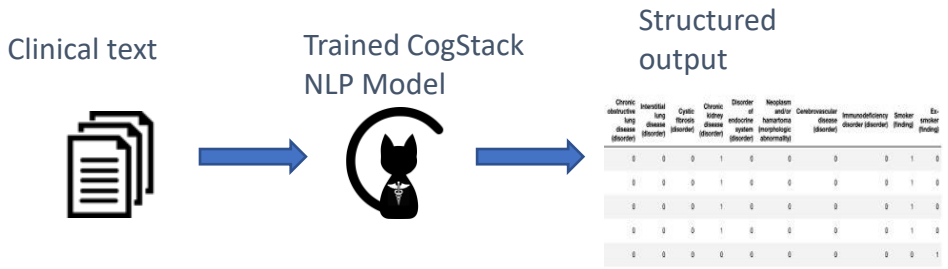
CogStack: Natural Language Processing

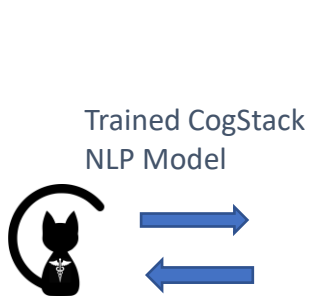
80% of EHR data is free text

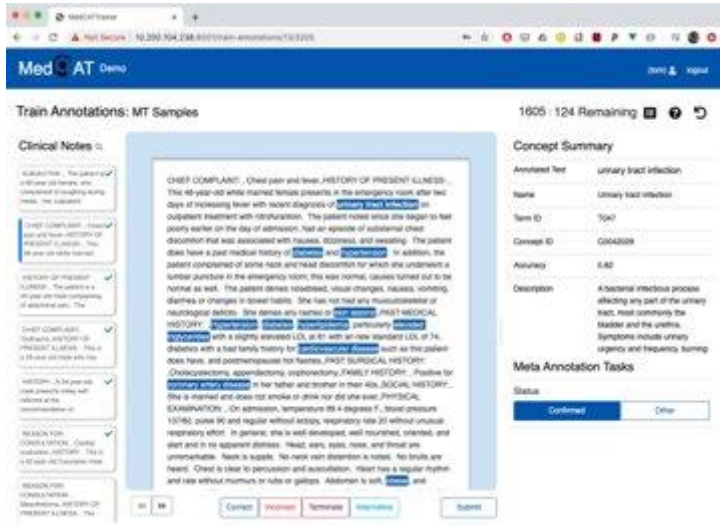
Structuring EHR data ready for downstream analysis

Configurable to any clinical terminology

Trainable by simply ‘running over text’







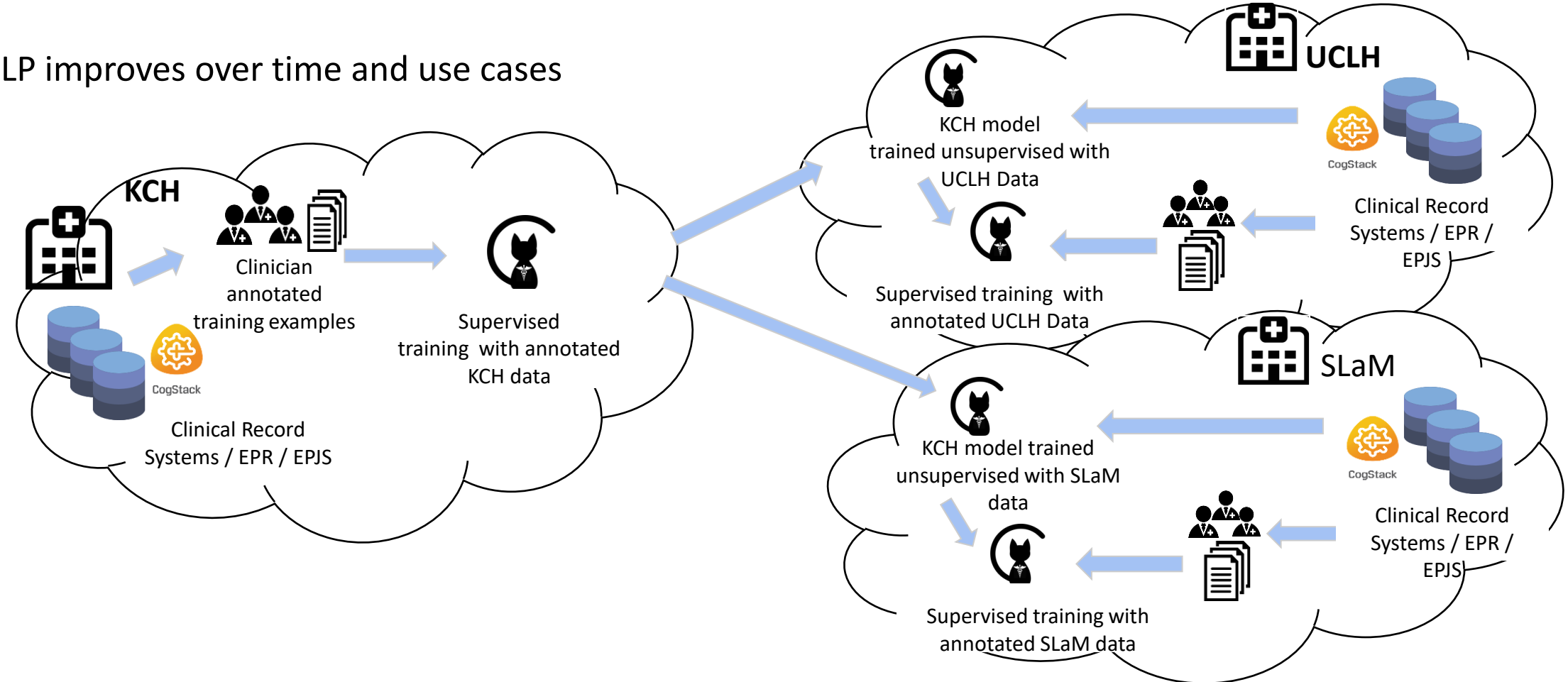
Clinical expert validation

Further fine-tuning of specific models via ’supervised training’

CogStack: Natural Language Processing

NLP models are sharable reusable between use cases and Trusts

NLP improves over time and use cases



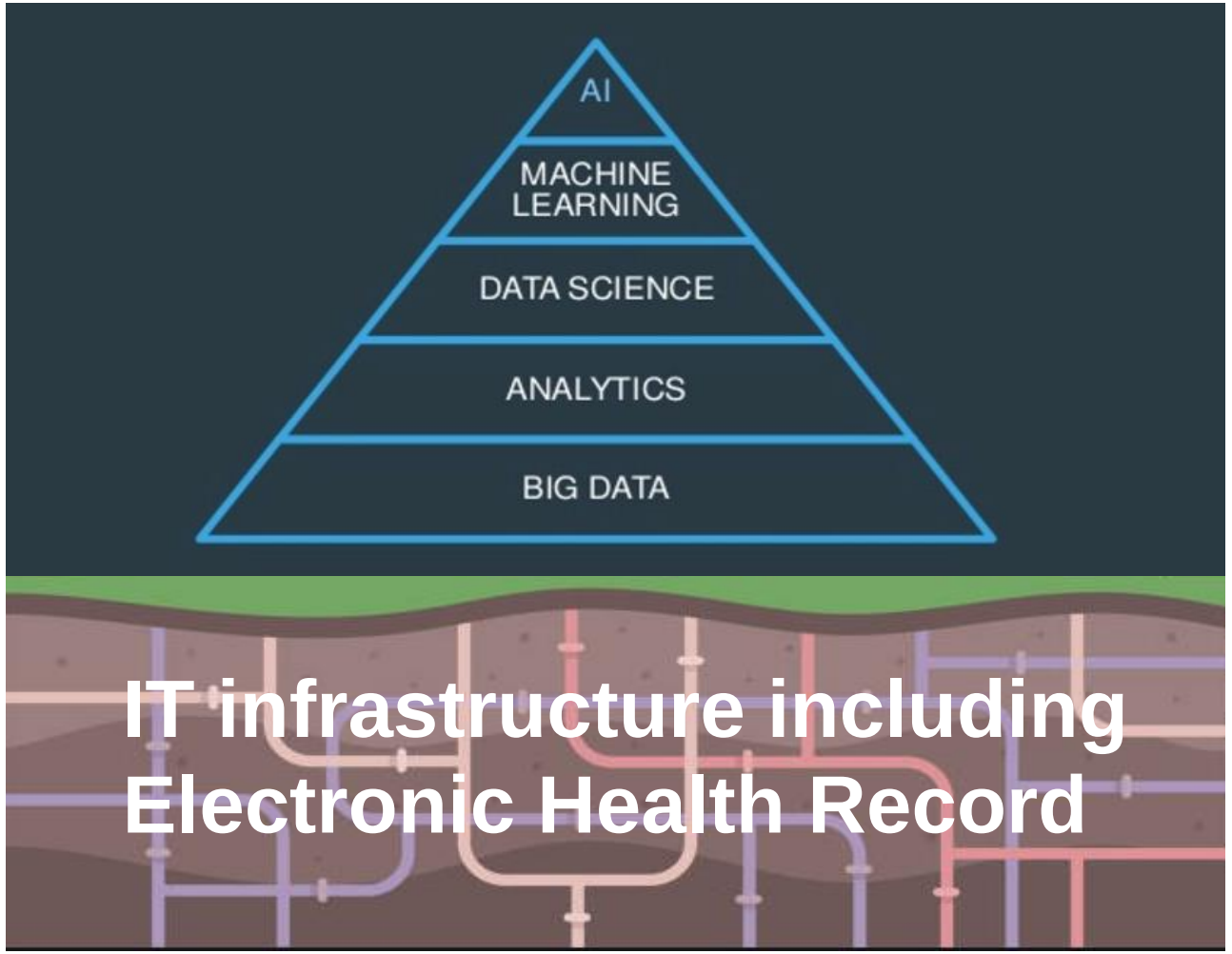
CogStack Deployment Case Studies

Cogstack

An open platform harnessing healthcare language

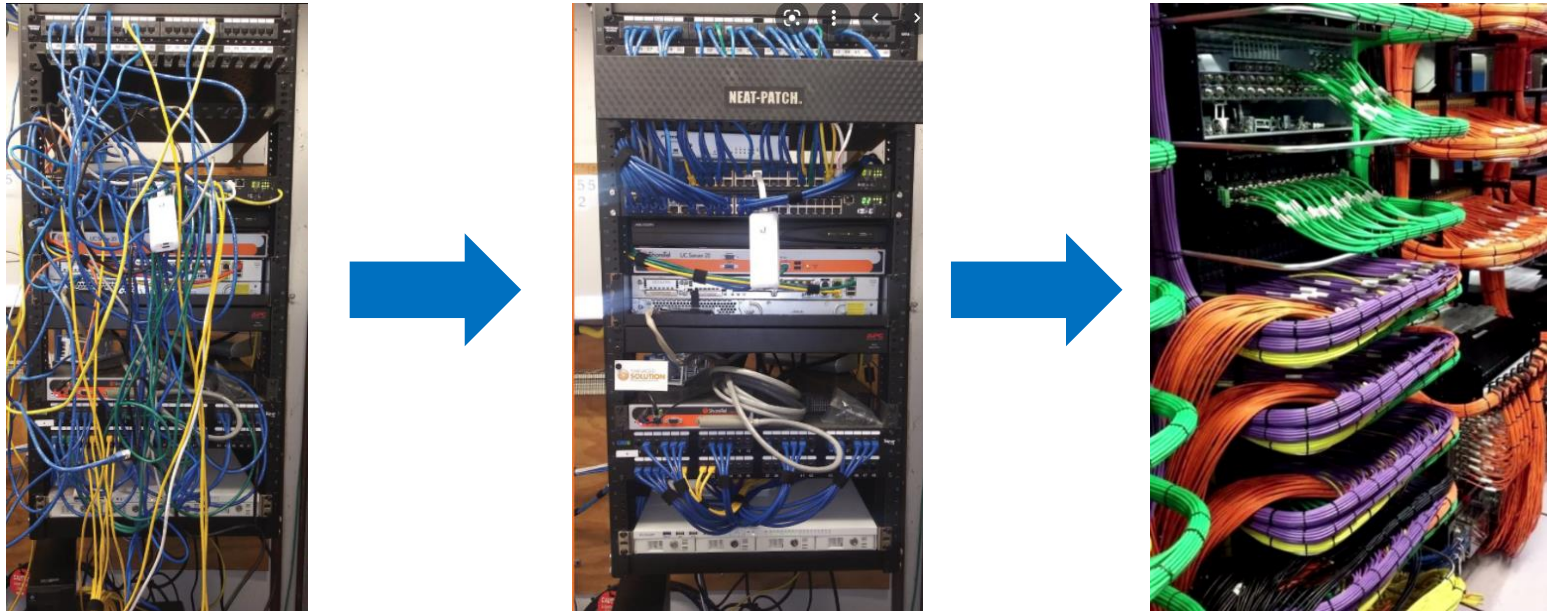
Prof James Teo
Joint Clinical Director of Data & AI
Professor of Neurology

A hierarchy of needs for digital health



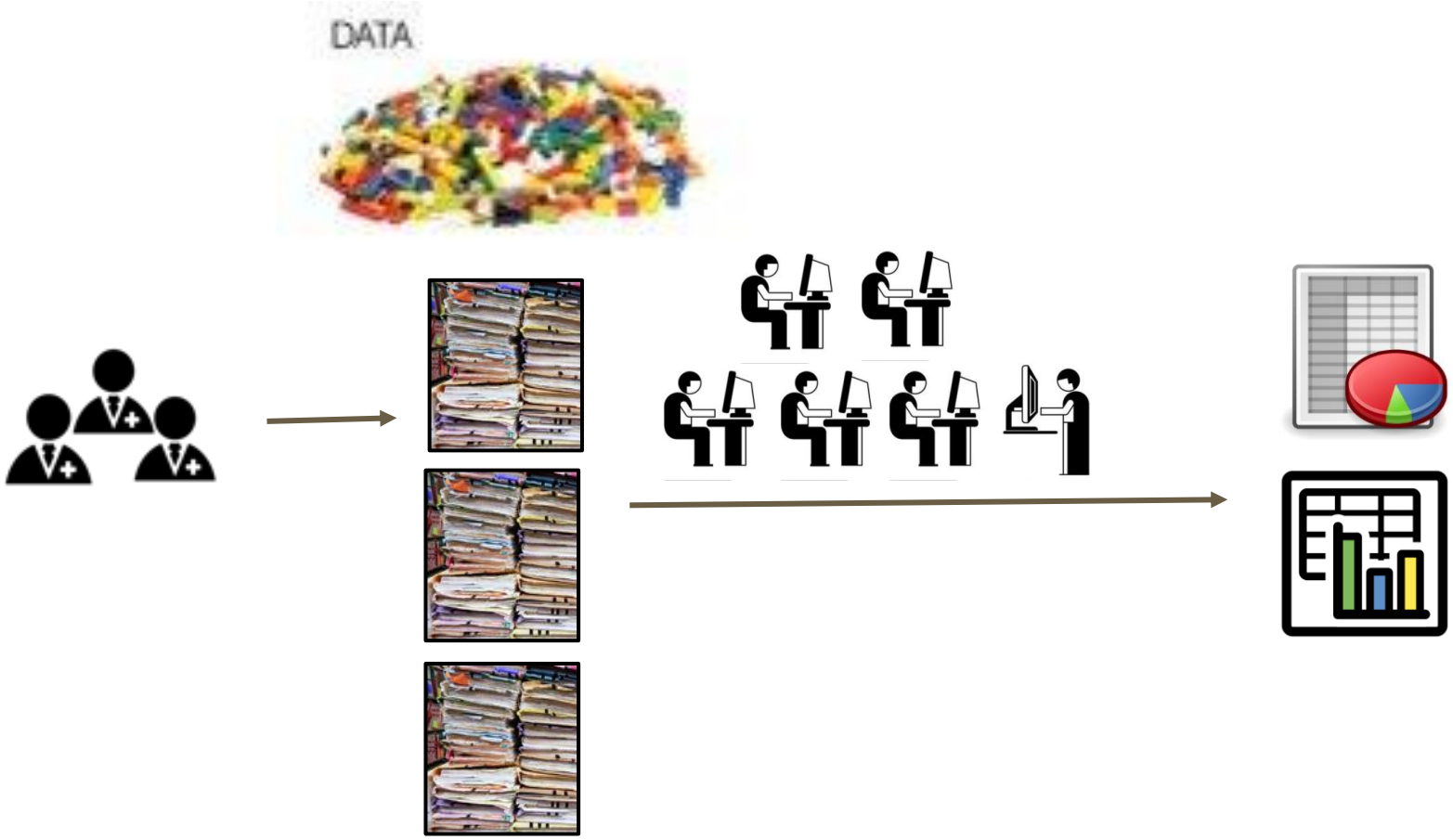
To ‘do’ data science and AI, you need cleaned, sorted and arranged data

But to ‘use’ AI in a production capacity, you need to be able to clean, sort and arrange **continuously and rapidly**



New interoperable infrastructure and EHR’s?

The original way of producing and cleaning data



Unstructured Data

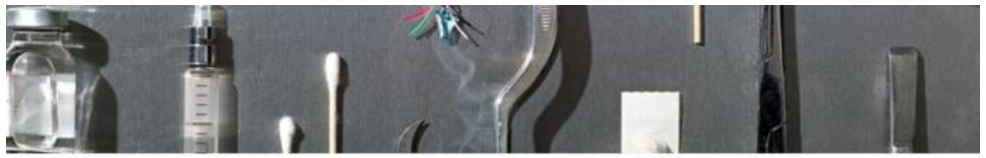
Easy for humans to input
Easy for humans to read
Contained in documents and variety of formats
Agnostic to ontologies and can capture non-health concepts
Particular to language

Problem: DIRTY & MESSY

“Mrs Smith is a 65 year old woman with atrial fibrilation had a CVA in March. She had a past history of a #NOF and OA. She has a family history of breast cancer. She has been prescribed apixiban. She has no history of haemorrhage.”

- Spelling / Typo
- Nomenclature
- Acronyms
- Negative terms
- Family history terms

Add a new Electronic Health Record System?

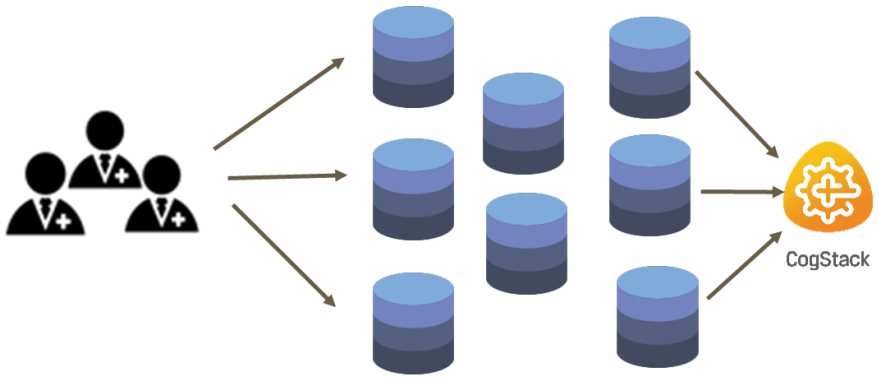


Death by a Thousand Clicks: Where Electronic Health Records Went Wrong

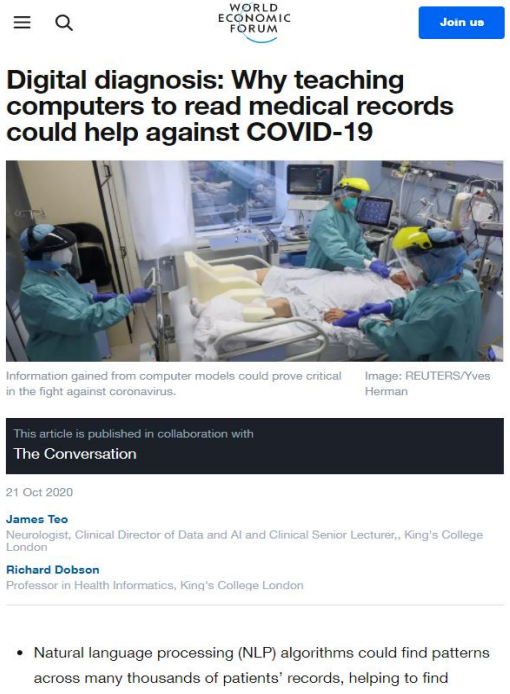
The U.S. government claimed that turning American medical charts into electronic records would make health care better, safer, and cheaper. Ten years and $36 billion later, the system is an unholy mess: Inside a digital revolution gone wrong. A joint investigation by Fortune and Kaiser Health News.

Cogstack platform for handling healthcare natural language

- Health data is extremely heterogenous
- Adoption of healthcare interoperable open standards is uneven
- Users will bypass structured data entry whenever there is any inadequacies in UX
- Natural Language AI is helping us make sense of all the Big Data



Heterogenous Clinical Record systems



WORLD ECONOMIC FORUM

Join us

Digital diagnosis: Why teaching computers to read medical records could help against COVID-19

Information gained from computer models could prove critical in the fight against coronavirus. Image: REUTERS/Yves Herman

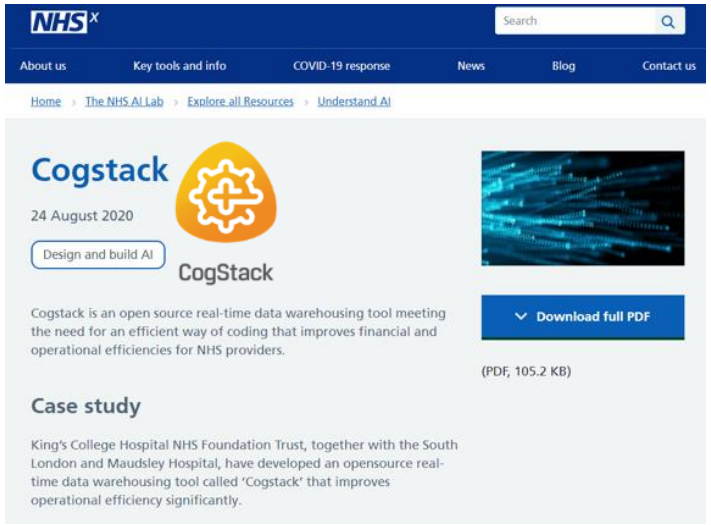
This article is published in collaboration with The Conversation

21 Oct 2020

James Teo
Neurologist, Clinical Director of Data and AI and Clinical Senior Lecturer,, King's College London

Richard Dobson
Professor in Health Informatics, King's College London

- Natural language processing (NLP) algorithms could find patterns across many thousands of patients' records, helping to find



NHS[X]

Search

About us | Key tools and info | COVID-19 response | News | Blog | Contact us

Home › The NHS AI Lab › Explore all Resources › Understand AI

Cogstack

24 August 2020

Design and build AI

CogStack

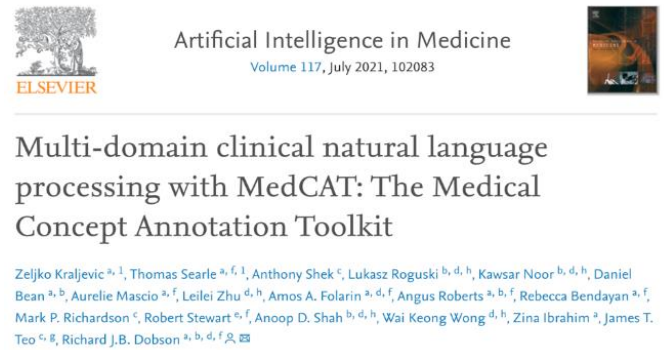
Cogstack is an open source real-time data warehousing tool meeting the need for an efficient way of coding that improves financial and operational efficiencies for NHS providers.

Download full PDF

(PDF, 105.2 KB)

Case study

King's College Hospital NHS Foundation Trust, together with the South London and Maudsley Hospital, have developed an opensource real-time data warehousing tool called 'Cogstack' that improves operational efficiency significantly.



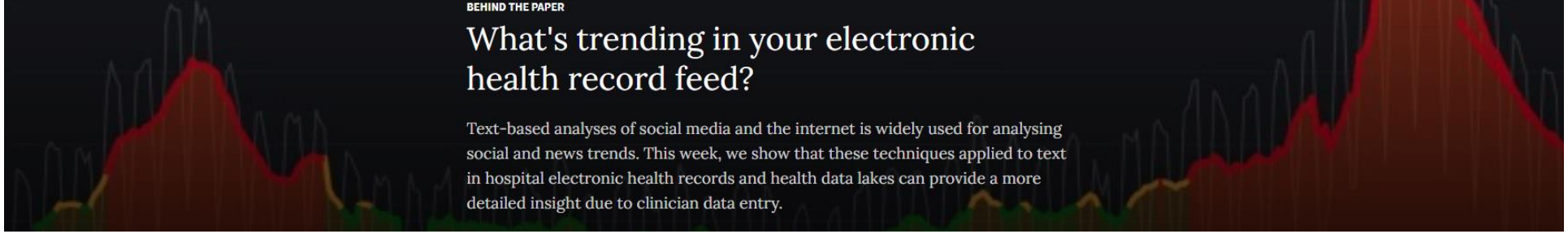
ELSEVIER

Artificial Intelligence in Medicine

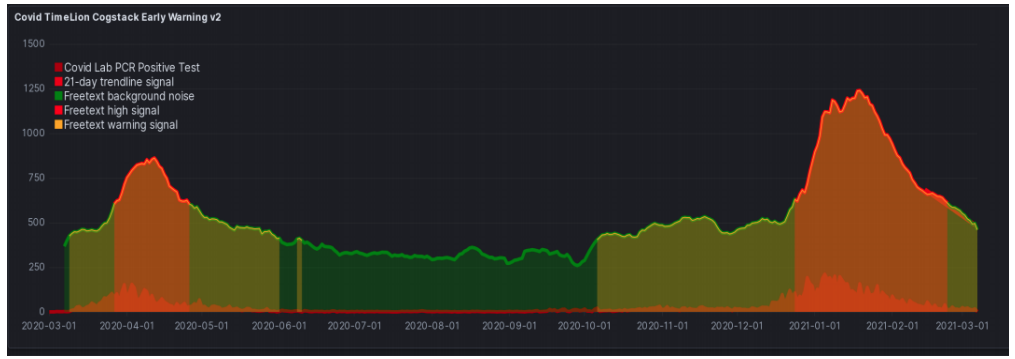
Volume 117, July 2021, 102083

Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit

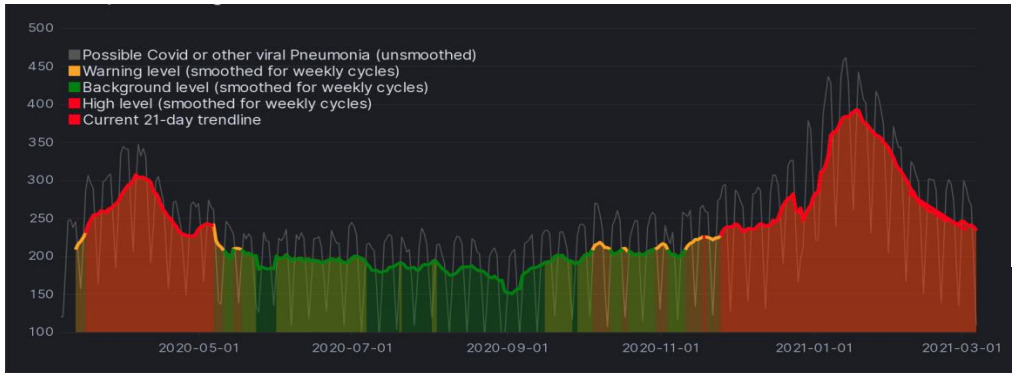
Zeljko Kraljevic [a, 1], Thomas Searle [a, f, 1], Anthony Shek [c], Lukasz Roguski [b, d, h], Kawsar Noor [b, d, h], Daniel Bean [a, b], Aurelie Mascio [a, f], Leilei Zhu [d, h], Amos A. Folarin [a, d, f], Angus Roberts [a, b, f], Rebecca Bendayan [a, f], Mark P. Richardson [c], Robert Stewart [e, f], Anoop D. Shah [b, d, h], Wai Keong Wong [d, h], Zina Ibrahim [a], James T. Teo [c, g], Richard J.B. Dobson [a, b, d, f]



Kings College Hospital
Princess Royal University Hospital



Guys & St Thomas (850 beds)



npj | Digital Medicine

www.nature.com/npjdigitalmed

BRIEF COMMUNICATION OPEN

Check for updates

Real-time clinician text feeds from electronic health records

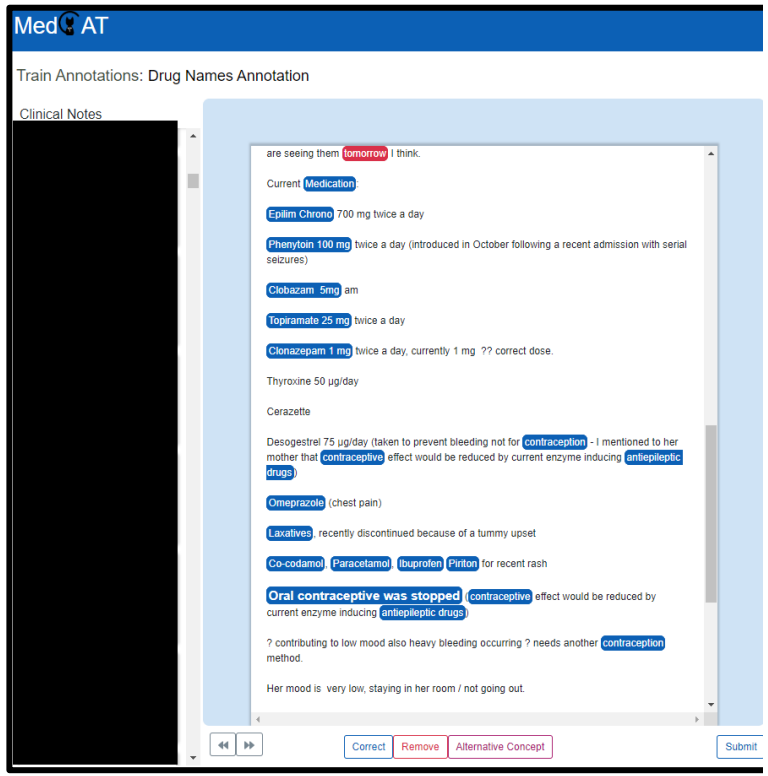
James T. H. Teo[1,2,3✉], Vlad Dinu[3], William Bernal[1], Phil Davidson[1], Vitaliy Oliynyk[2], Cormac Breen[2], Richard D. Barker[1] and Richard J. B. Dobson[3]

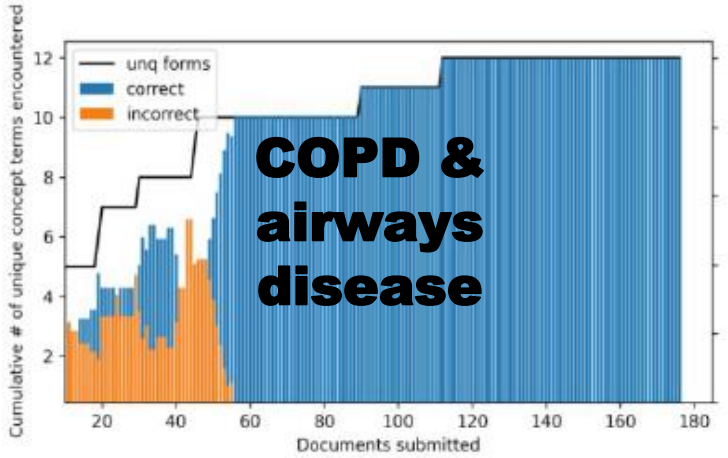
Analyses of search engine and social media feeds have been attempted for infectious disease outbreaks, but have been found to be susceptible to artefactual distortions from health scares or keyword spamming in social media or the public internet. We describe an approach using real-time aggregation of keywords and phrases of freetext from real-time clinician-generated documentation in electronic health records to produce a customisable real-time viral pneumonia signal providing up to 4 days warning for secondary care capacity planning. This low-cost approach is open-source, is locally customisable, is not dependent on any specific electronic health record system and can provide an ensemble of signals if deployed at multiple organisational scales.

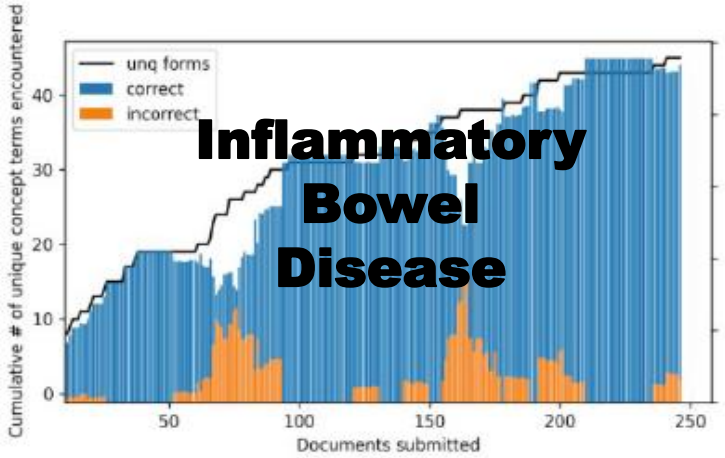
npj Digital Medicine (2021)4:35 ; https://doi.org/10.1038/s41746-021-00406-7

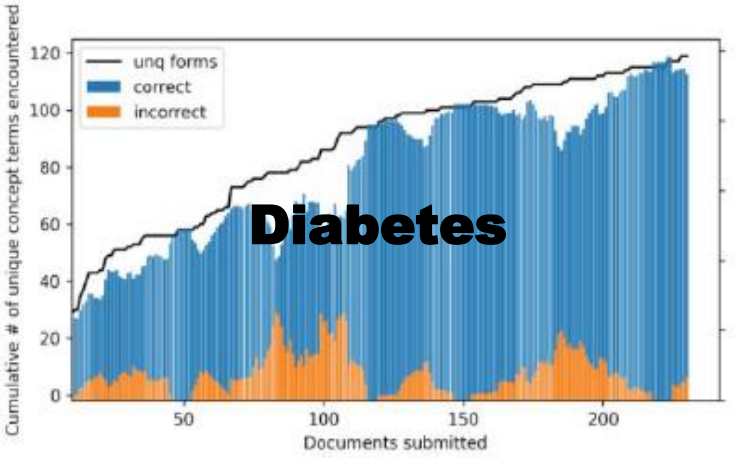
NLP that understands medical jargon and grammar

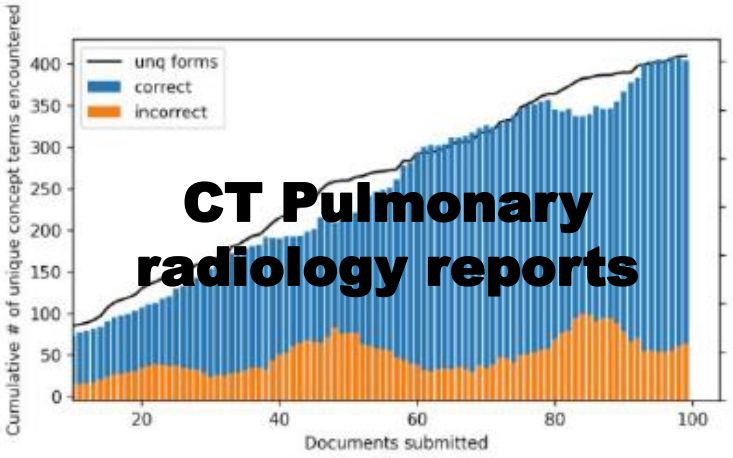
Targeted training to improve subspecialty understanding







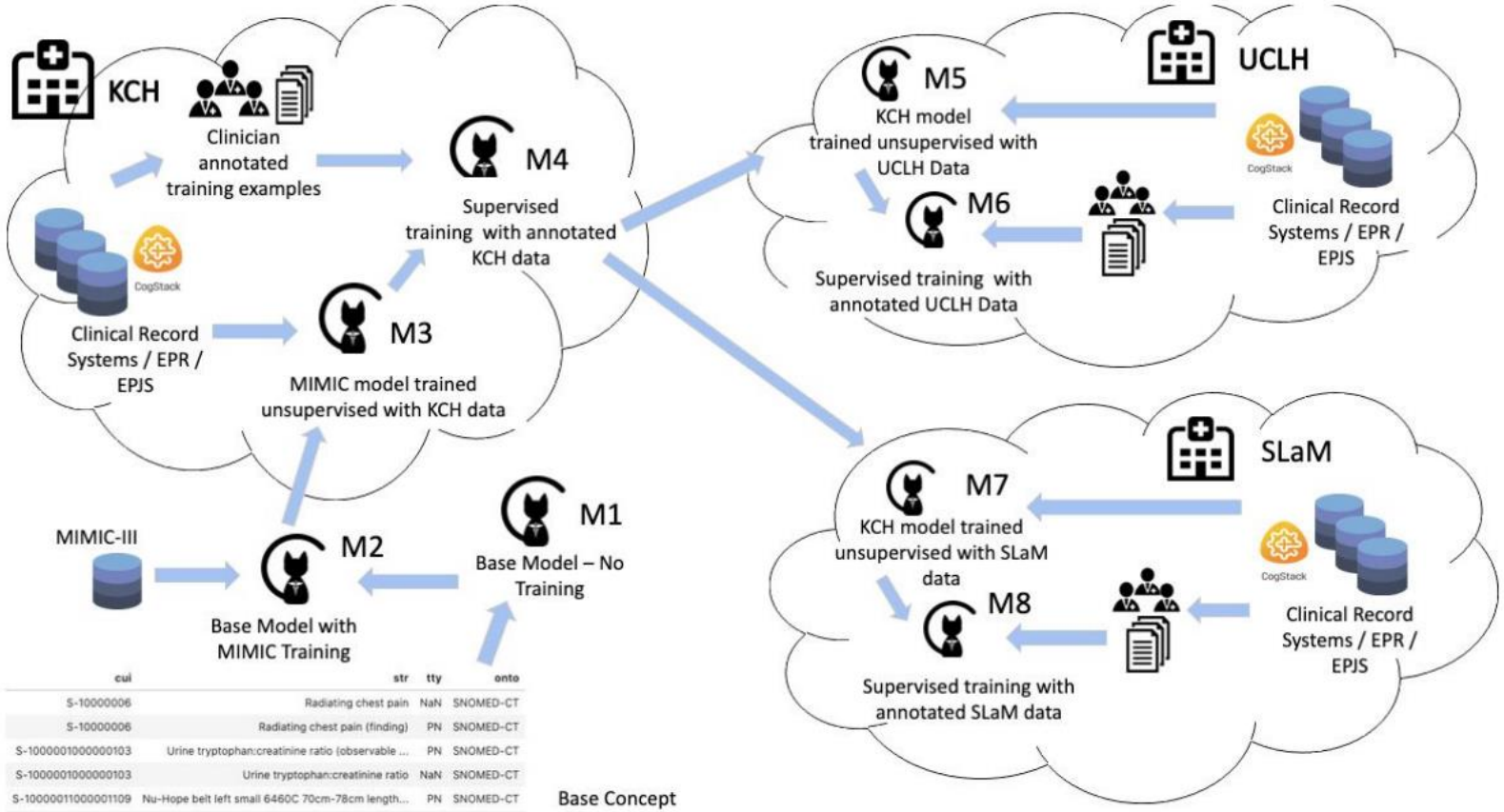




Top left to bottom right: MedCATtrainer annotation projects with respective numbers unique concepts seen throughout annotating and the number of configured concepts: Covid_COPD (5/2012), Covid_Gastro (8/679), Diabetes_Covid (15/864), Covid_CTPA_Reports (194/297280)

Cogstack-MedCat

Using a 'Rosetta Stone' AI to clean data by reading



cui	str	tty	onto
S-10000006	Radiating chest pain	NaN	SNOMED-CT
S-10000006	Radiating chest pain (finding)	PN	SNOMED-CT
S-1000001000000103	Urine tryptophan:creatinine ratio (observable ...	PN	SNOMED-CT
S-1000001000000103	Urine tryptophan:creatinine ratio	NaN	SNOMED-CT
S-10000011000001109	Nu-Hope belt left small 6460C 70cm-78cm length...	PN	SNOMED-CT
...	...	...	...
S-9999811000001102	Nu-Hope belt right large 6457 90cm-100cm lengt...	NaN	SNOMED-CT
S-999991000000109	Urine tyrosine:creatinine ratio (observable en...	PN	SNOMED-CT
S-999991000000109	Urine tyrosine:creatinine ratio	NaN	SNOMED-CT

Base Concept Vocabulary (SNOMED-CT)

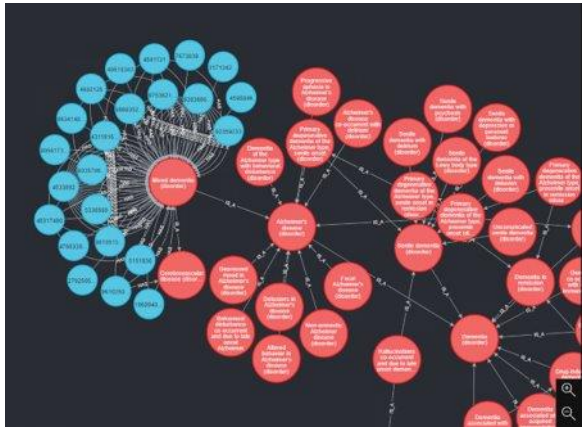
Code Availability

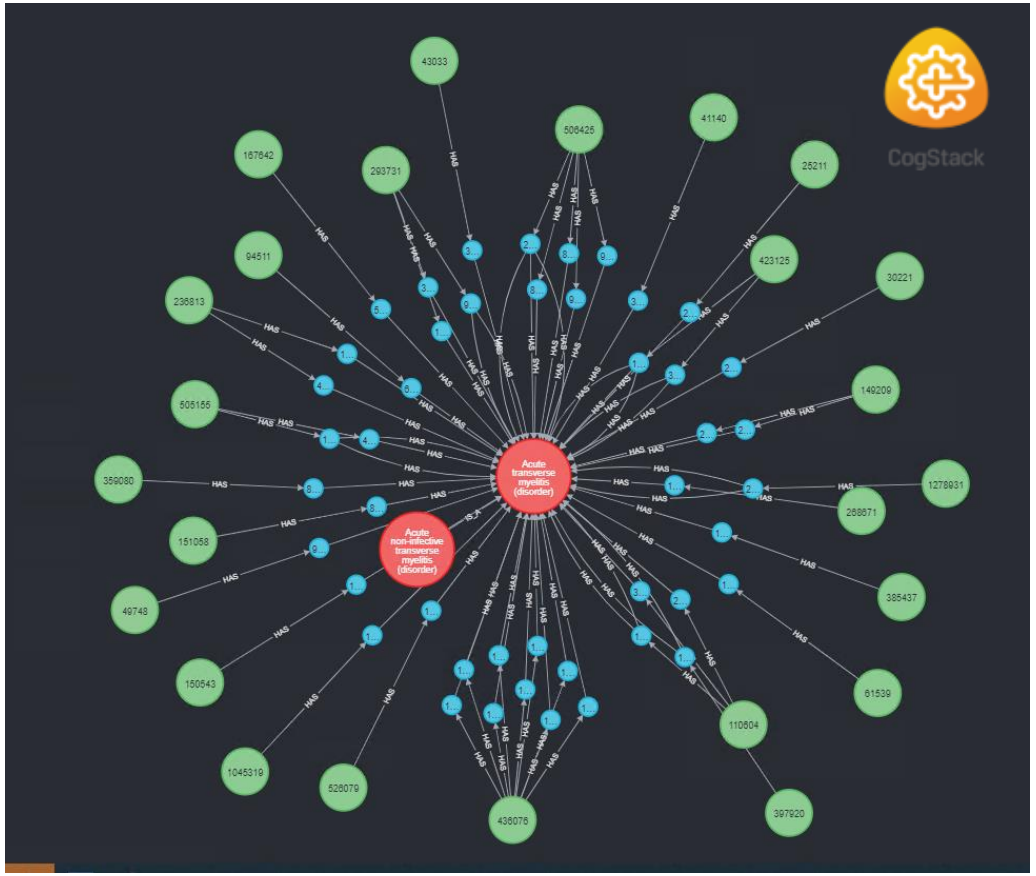
All code for running the experiments, the toolkit and integration with wider CogStack deployments are available here:

MedCAT: https://github.com/CogStack/MedCAT
MedCAT Tutorials/Example Code: https://github.com/CogStack/MedCAT/tree/master/tutorial
MedCATtrainer: https://github.com/CogStack/MedCATtrainer
MedCATtrainer Examples: https://github.com/CogStack/MedCATtrainer/tree/master/docs
MedCATservice: https://github.com/CogStack/MedCATservice
CogStack: https://github.com/CogStack/CogStack-Pipeline

Semantic maps to produce machine-readable meaning to allow ‘intelligent’ querying of data

GREEN = patient
BLUE = patient document
RED = disease concept





Semantic clinical NLP

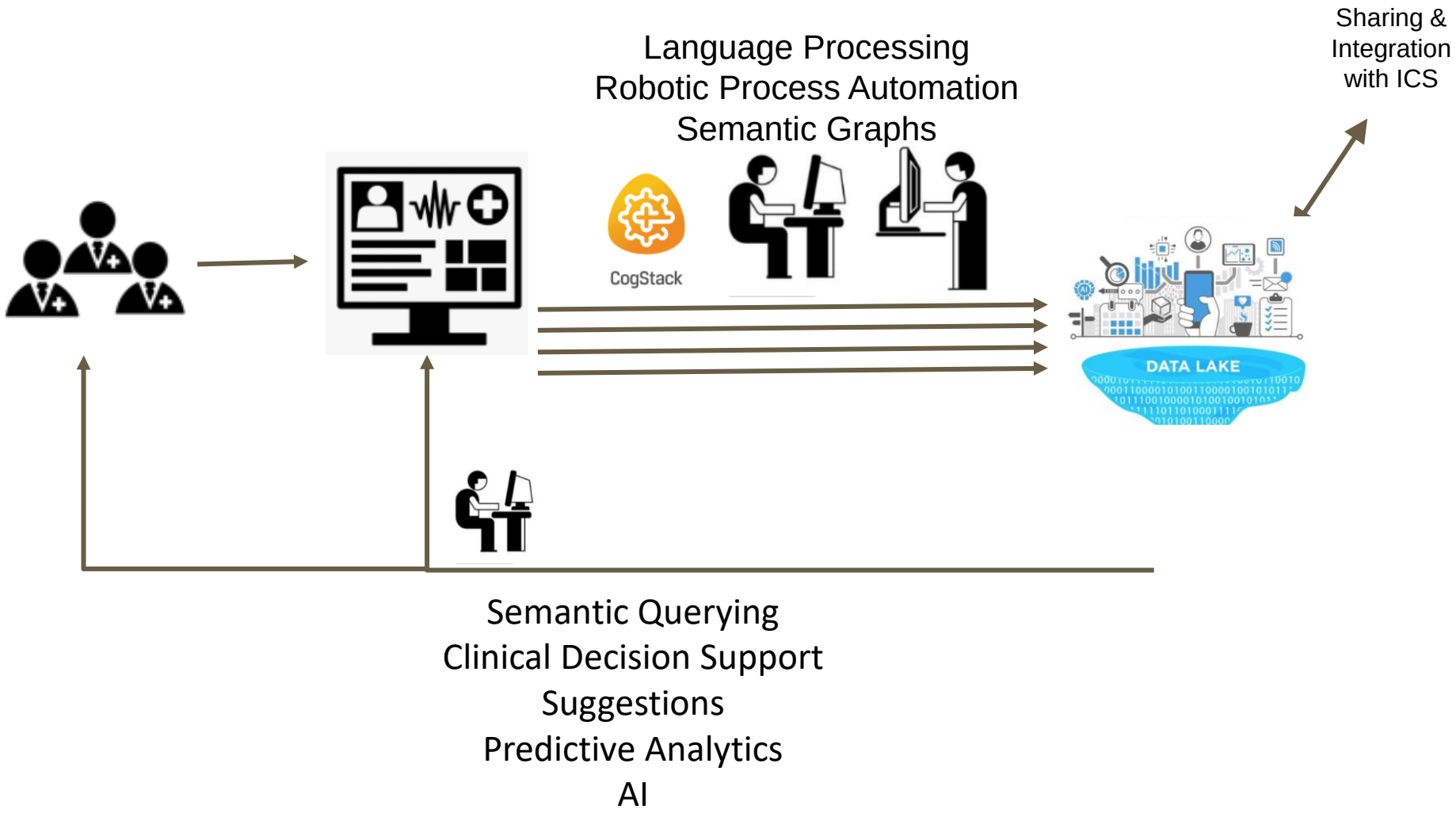
Language AI that can infer patterns of conditions and symptoms

Language AI that is learning what future medical problems a patient might expect

Figure 1: Importance of each token in the patient timeline for prediction of the right-most disorder using MedGPT. The weight was calculated using gradient-based saliency methods.

https://arxiv.org/abs/2107.03134

NLP AI to handle Big Data volume for speeding up data collection, enrichment, patient care and downstream data-driven technologies



VIEWER: A clinical-academic partnership enabling meaningful population health management (PHM)

Dr Robert Harland

Consultant in General Adult Psychiatry and Caldicott Guardian, South London and Maudsley NHS Foundation Trust

Clinical Director, King's Health Partners Psychosis Clinical Academic Group

Clinical Director, Lambeth Adult Mental Health

Challenges

1.3M

Local population served by SLaM

A large population with a high prevalence of mental ill health, particularly psychosis, with ongoing needs of healthcare.

- Improve the health of an entire population
- Enable data-driven care planning and delivery at scale

15-20Y

Die earlier in people with severe mental illness[1]

2/3 of these premature deaths are from avoidable physical illnesses, including heart disease and cancer.

- Improve both physical and mental health outcomes
- Reduce inequalities
- Use of stratification of population to design targeted interventions

£11.8B

Cost of psychosis for society per year in England[2]

Direct healthcare costs, lost productivity due to unemployment or death and informal costs to families and carers.

- Improve productivity
- Move care from being reactive to proactive to reduce cost
- Inform resource allocation

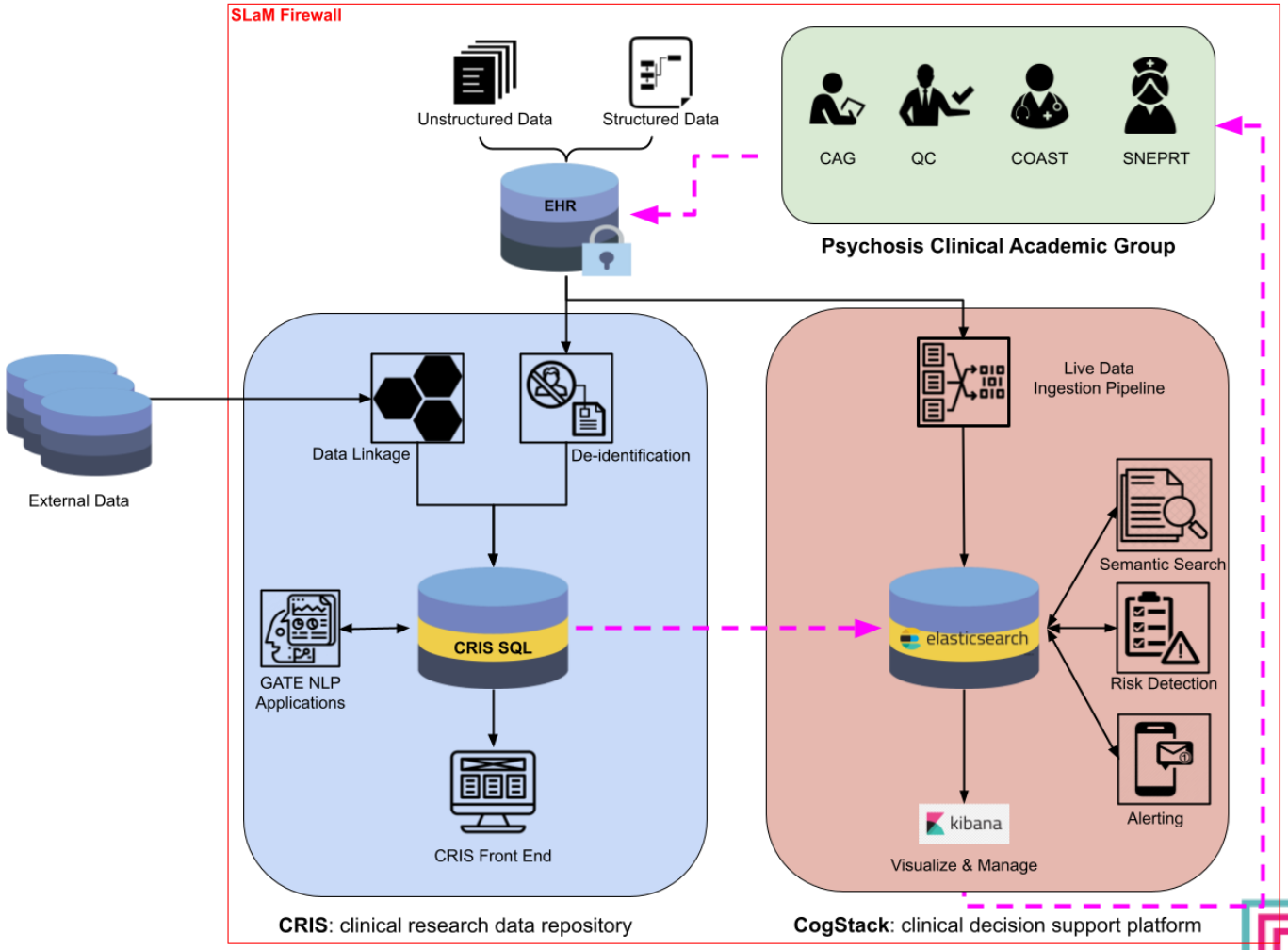
How to give right care to right service users at right time on a large scale?

1.NICE, Implementing the Early Intervention in Psychosis Access and Waiting Time Standard: Guidance, 2016
2.The Schizophrenia Commission. The abandoned illness: a report from the Schizophrenia Commission. London: Rethink Mental Illness; 2012

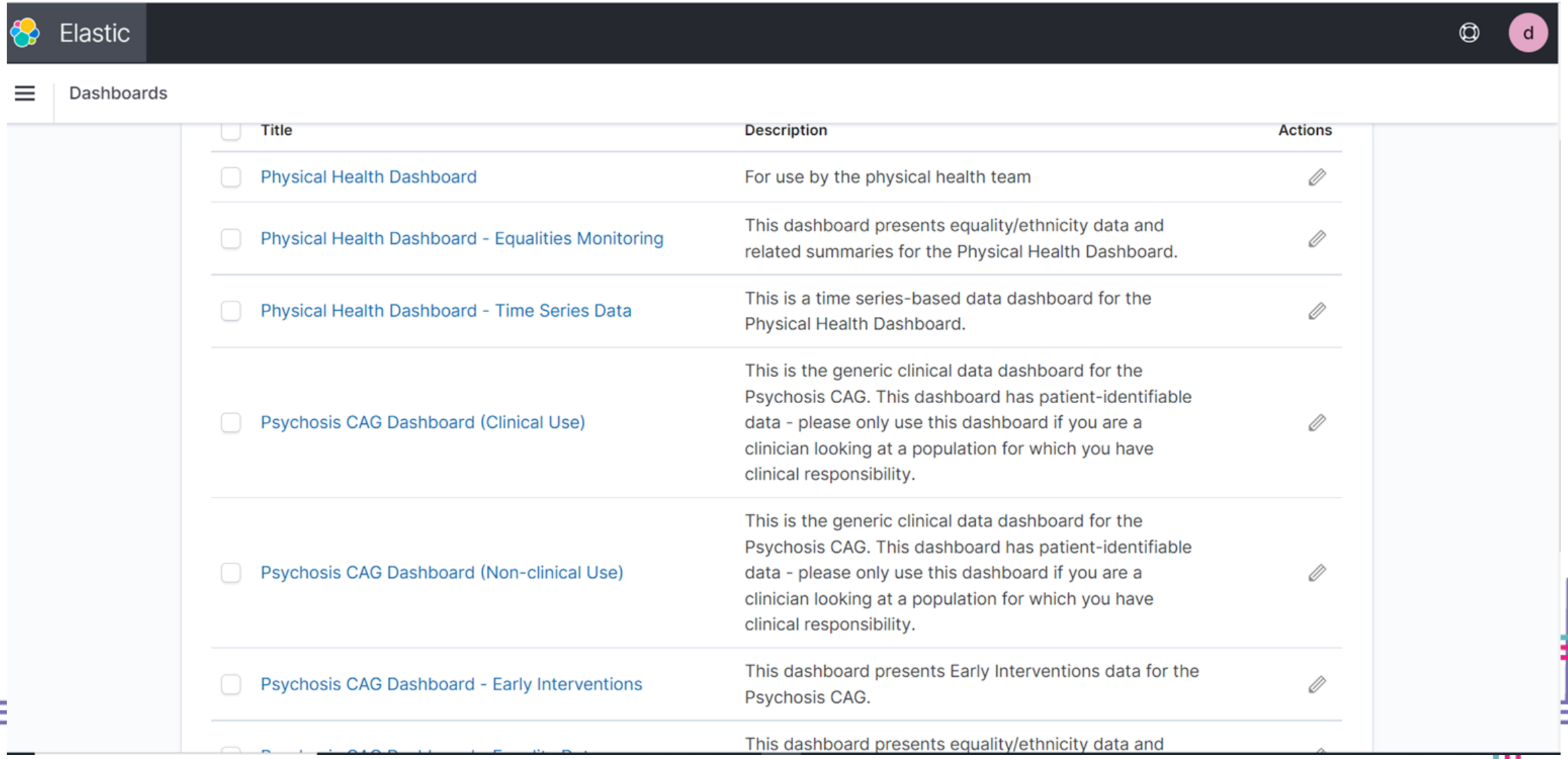
VIEWER: An informatics approach

Collaboration between clinicians, informaticians and computer scientists, bringing innovative solutions to real-world clinical problems.

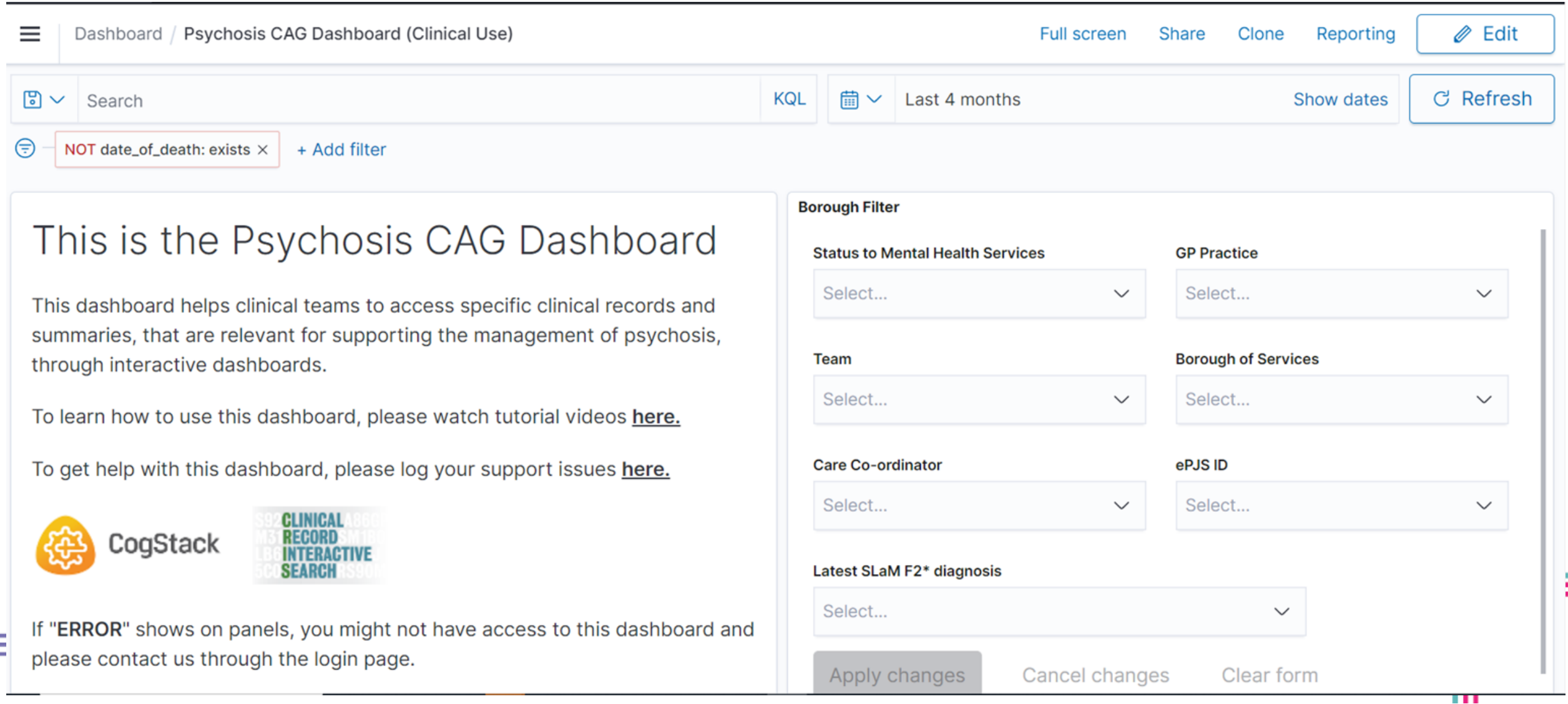
- Collaborative working to develop information models that meet clinicians' needs;
- Using cutting-edge research including automated natural language processing (NLP) techniques to extract insights from data at scale;
- Innovative and thoughtful interactive visualisation to identify patients with unmet needs from multiple perspectives;
- Support with roll-out and responding to feedback to engage staff and refine the platform.



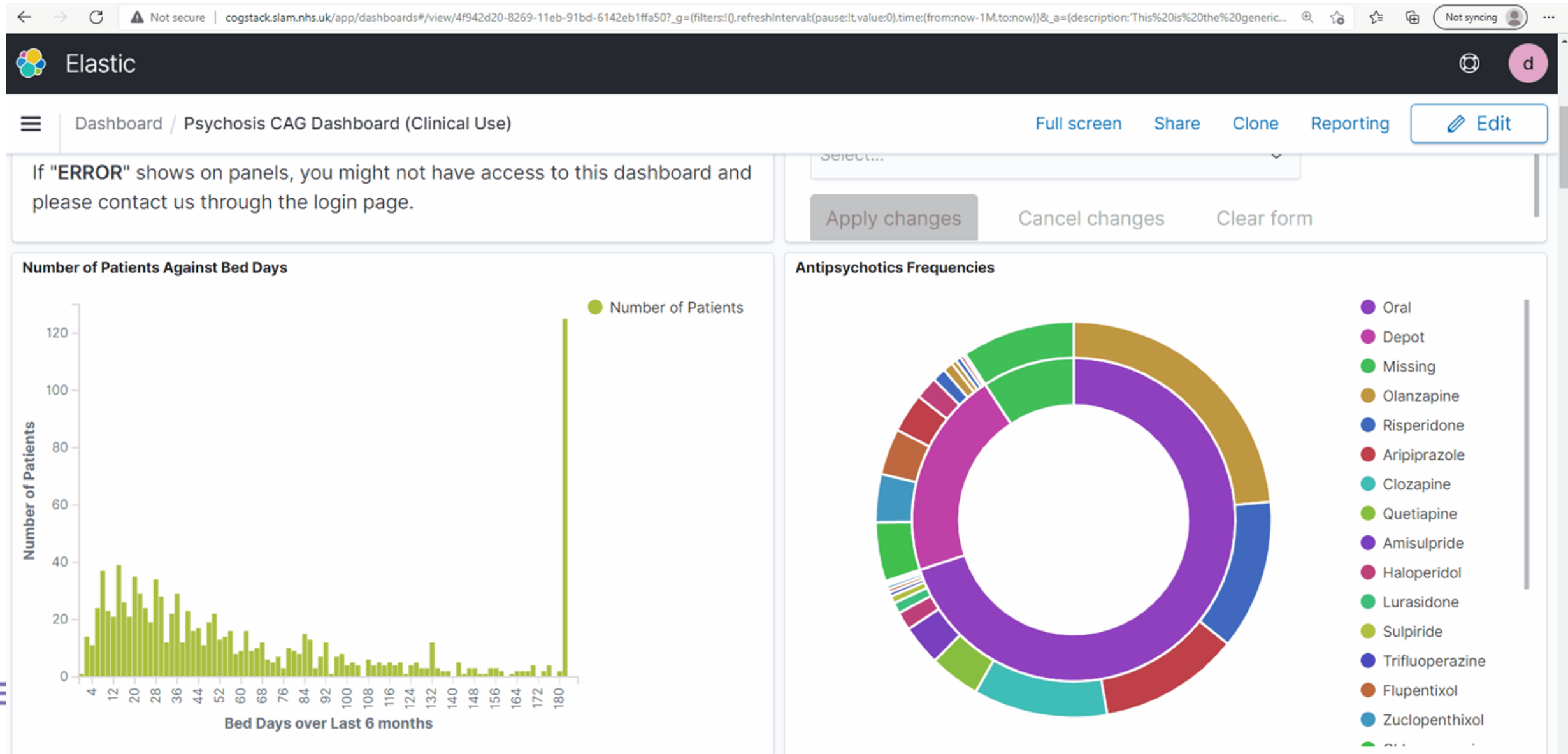
Unlock data for diverse clinical use cases



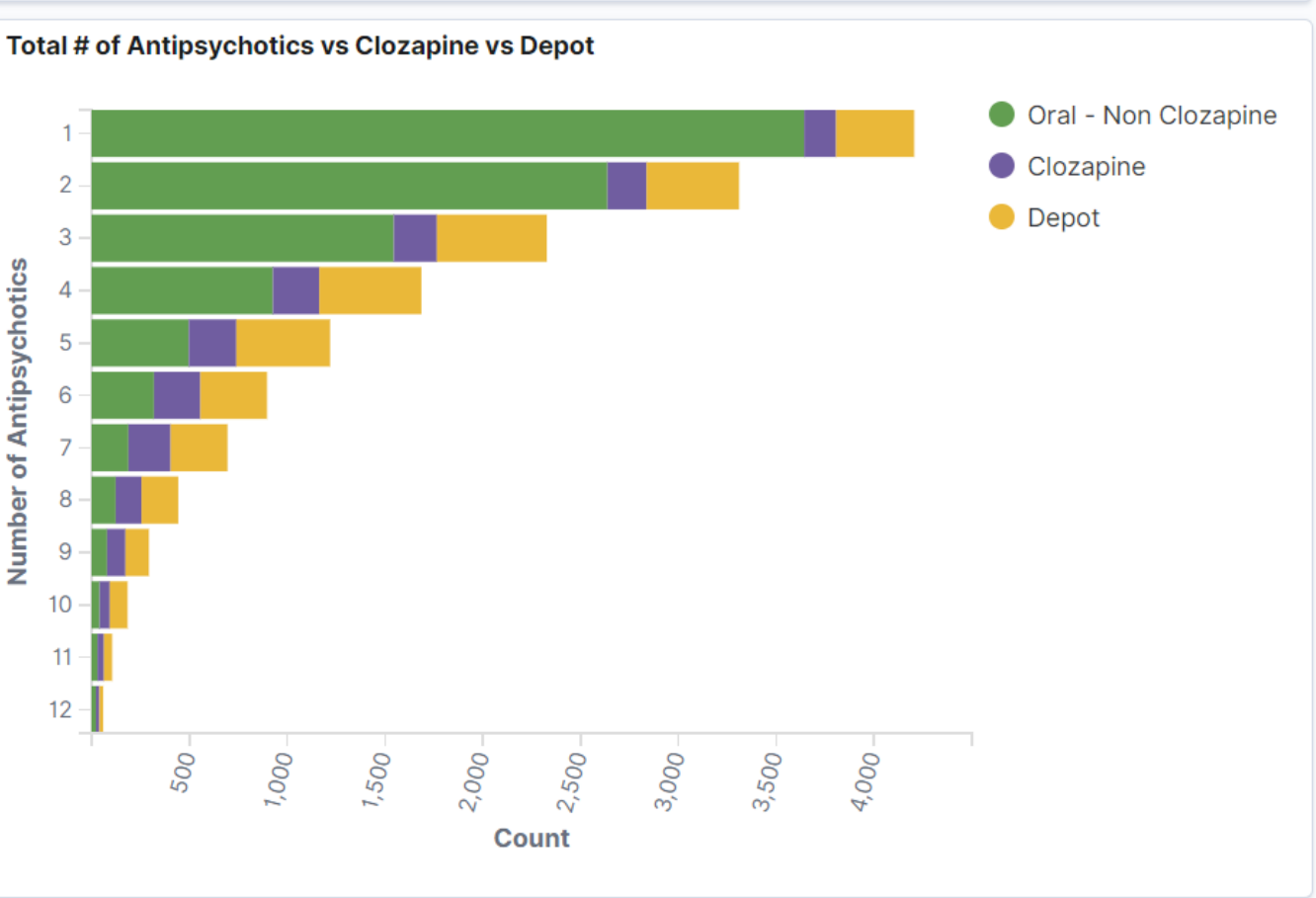
Intelligent filters



Interactive visualizations



Facilitate medication review

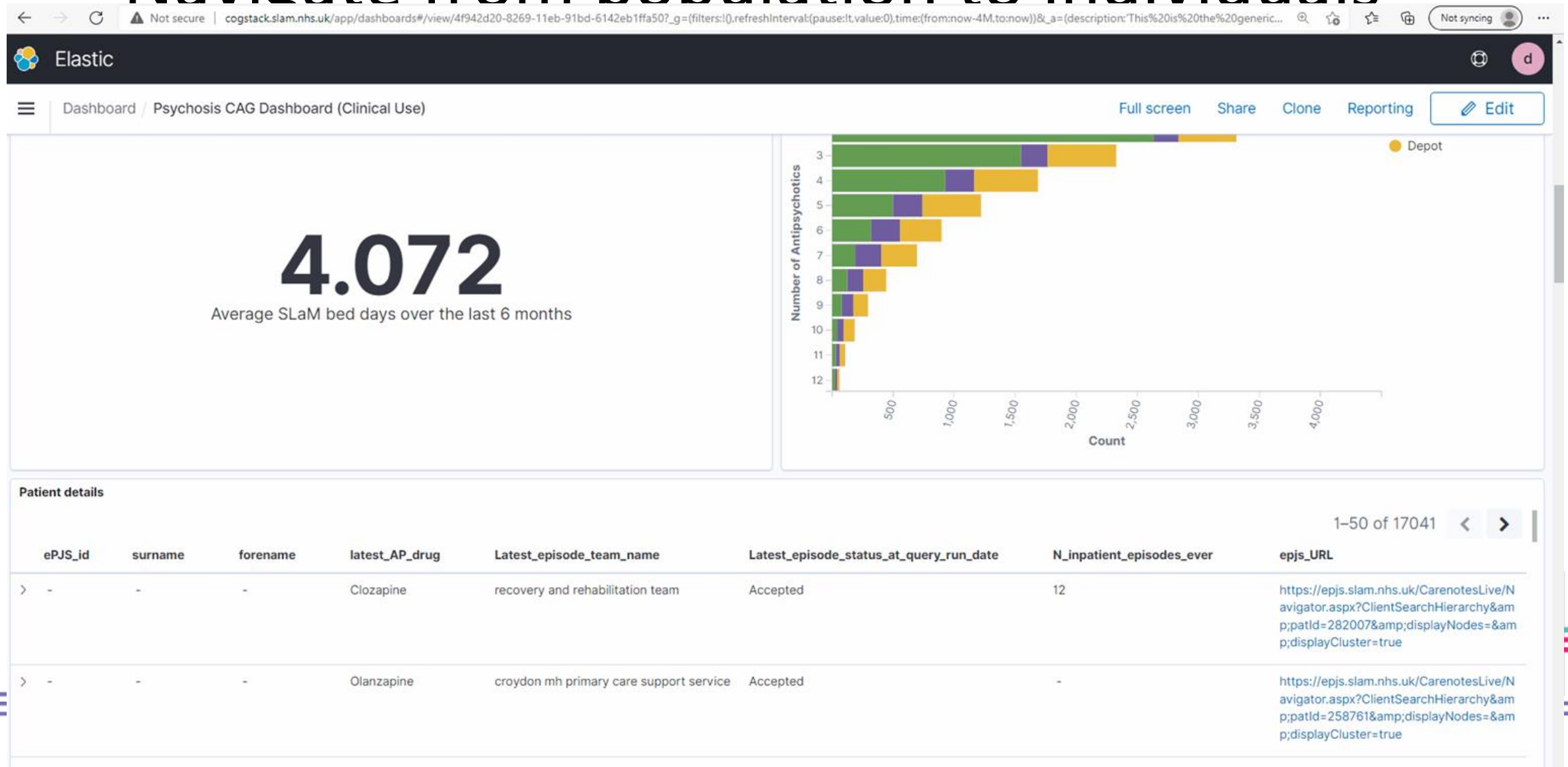


Medications Mentioned

1–50 of 51250

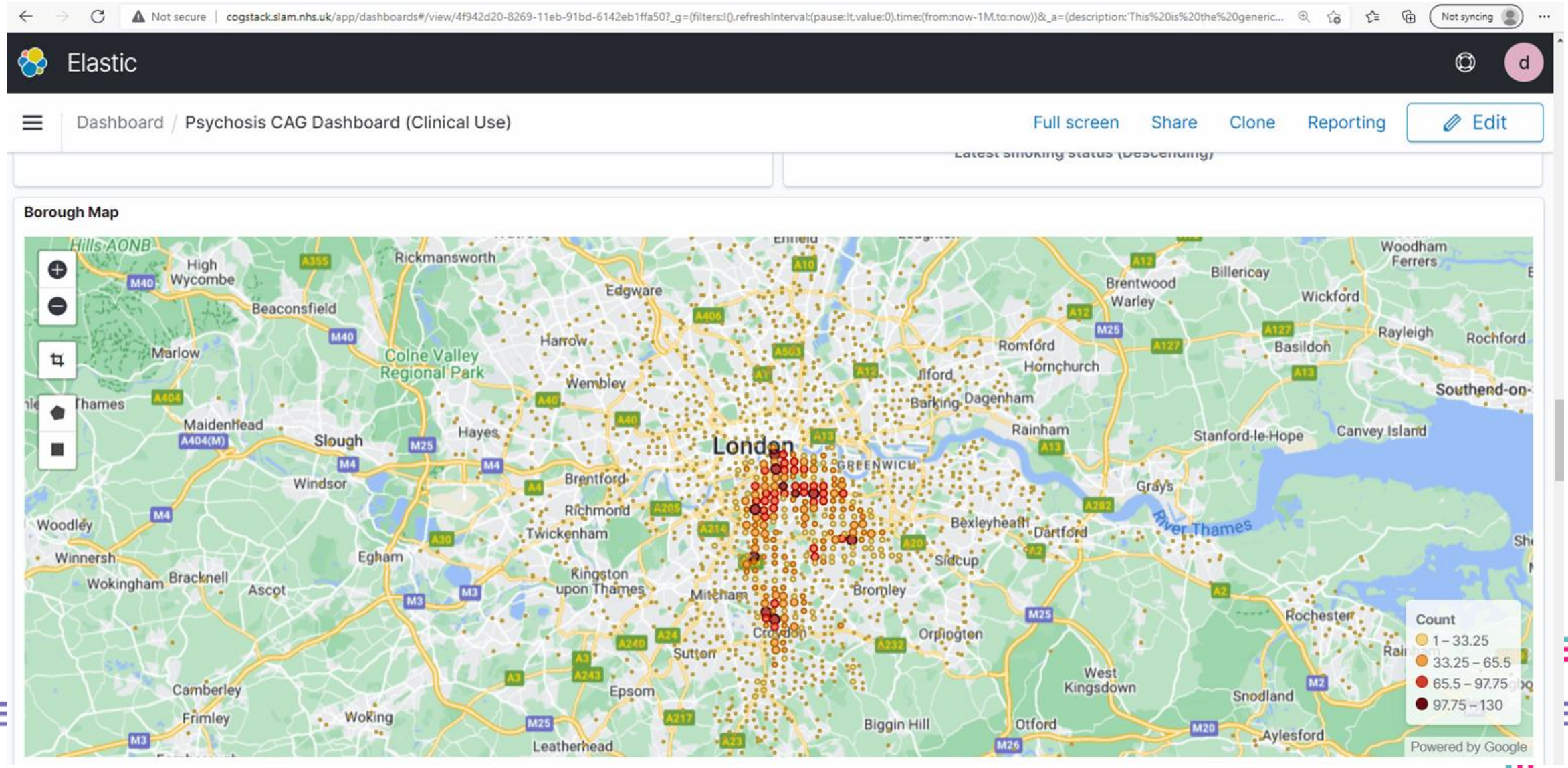
ePJS_id	medication	number_of_prescriptions	date_of_latest_prescriptio
-	Aripiprazole	20	Mar 8, 2022 @ 00:00:00.00
-	Flupentixol	46	Mar 8, 2022 @ 00:00:00.00
-	Aripiprazole	37	Mar 7, 2022 @ 00:00:00.00
-	Flupentixol	51	Mar 3, 2022 @ 00:00:00.00
-	Olanzapine	53	Feb 27, 2022 @ 19:48:57.87
-	Aripiprazole	40	Feb 27, 2022 @ 19:31:12.783
-	Aripiprazole	12	Feb 27, 2022 @ 15:57:17.220
-	Aripiprazole	202	Feb 27, 2022 @ 10:44:15.80

Navigate from population to individuals

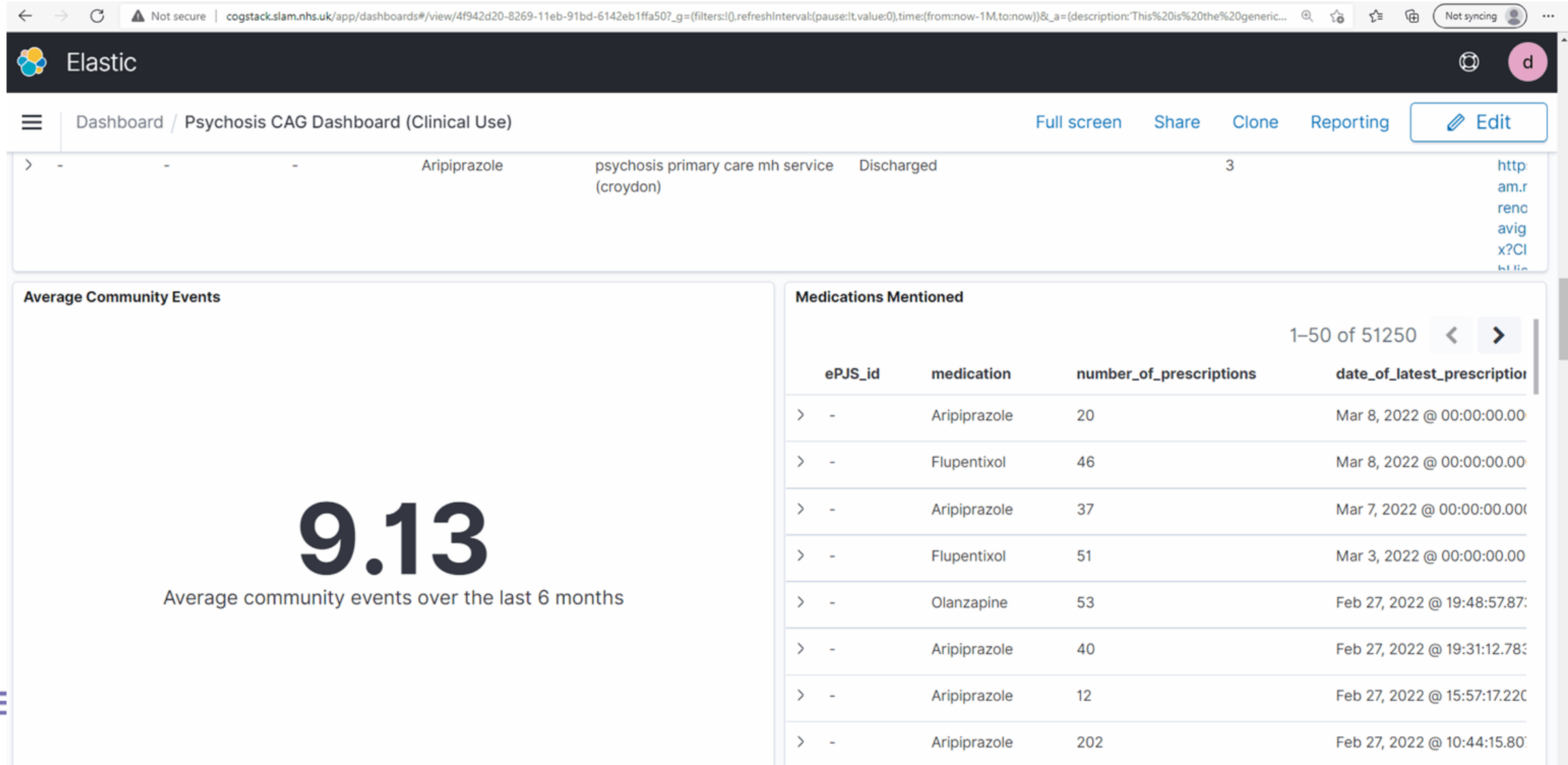


ePJS_id	surname	forename	latest_AP_drug	Latest_episode_team_name	Latest_episode_status_at_query_run_date	N_inpatient_episodes_ever	epjs_URL
-	-	-	Clozapine	recovery and rehabilitation team	Accepted	12	https://epjs.slam.nhs.uk/CarenotesLive/Navigator.aspx?ClientSearchHierarchy&patId=282007&displayNodes=&displayCluster=true
-	-	-	Olanzapine	croydon mh primary care support service	Accepted	-	https://epjs.slam.nhs.uk/CarenotesLive/Navigator.aspx?ClientSearchHierarchy&patId=258761&displayNodes=&displayCluster=true

Support resource allocation



Inform evidence-based care decisions

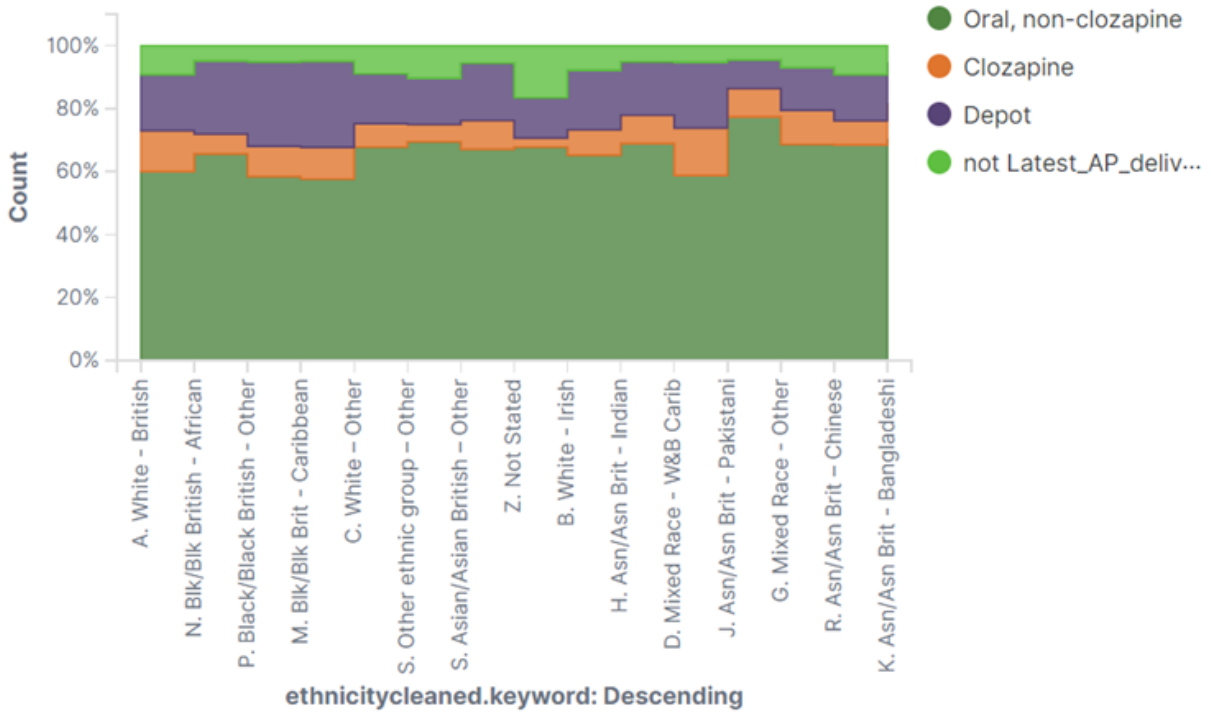


Assess inequality

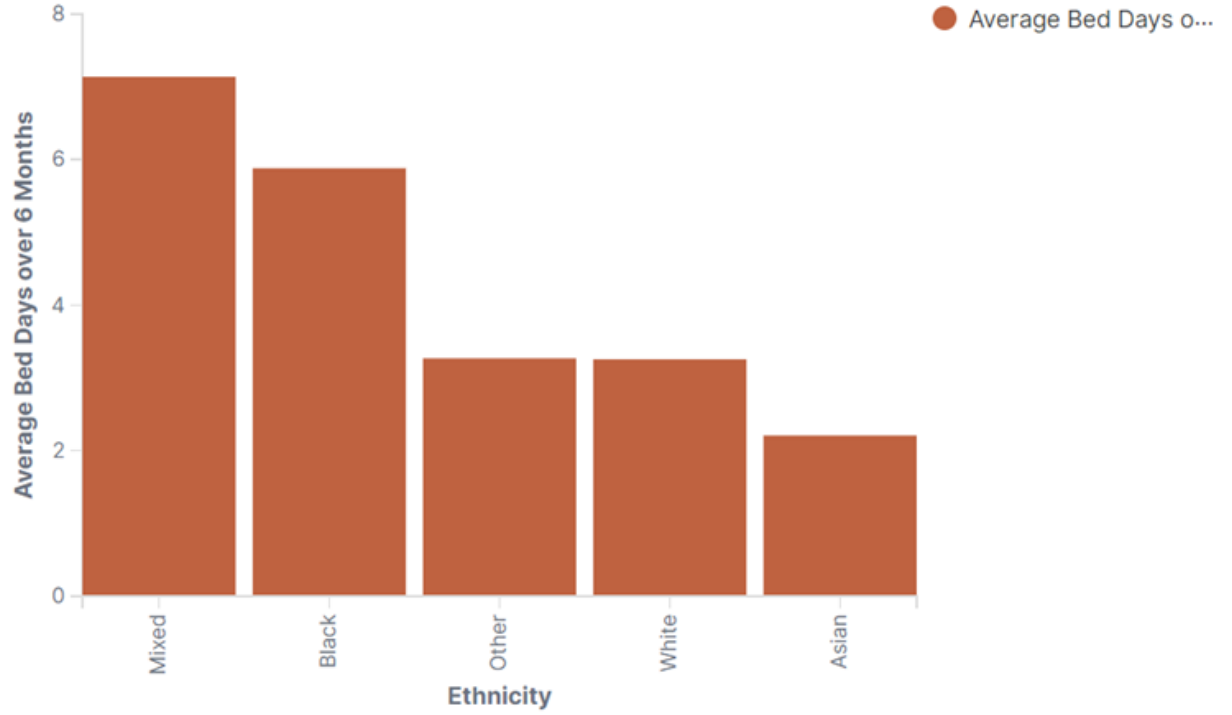
Dashboard / Psychosis CAG Dashboard - Equality Data

Full screen | Share | Clone | Reporting | Edit

Ethnicity Clozapine Depot New Ratios



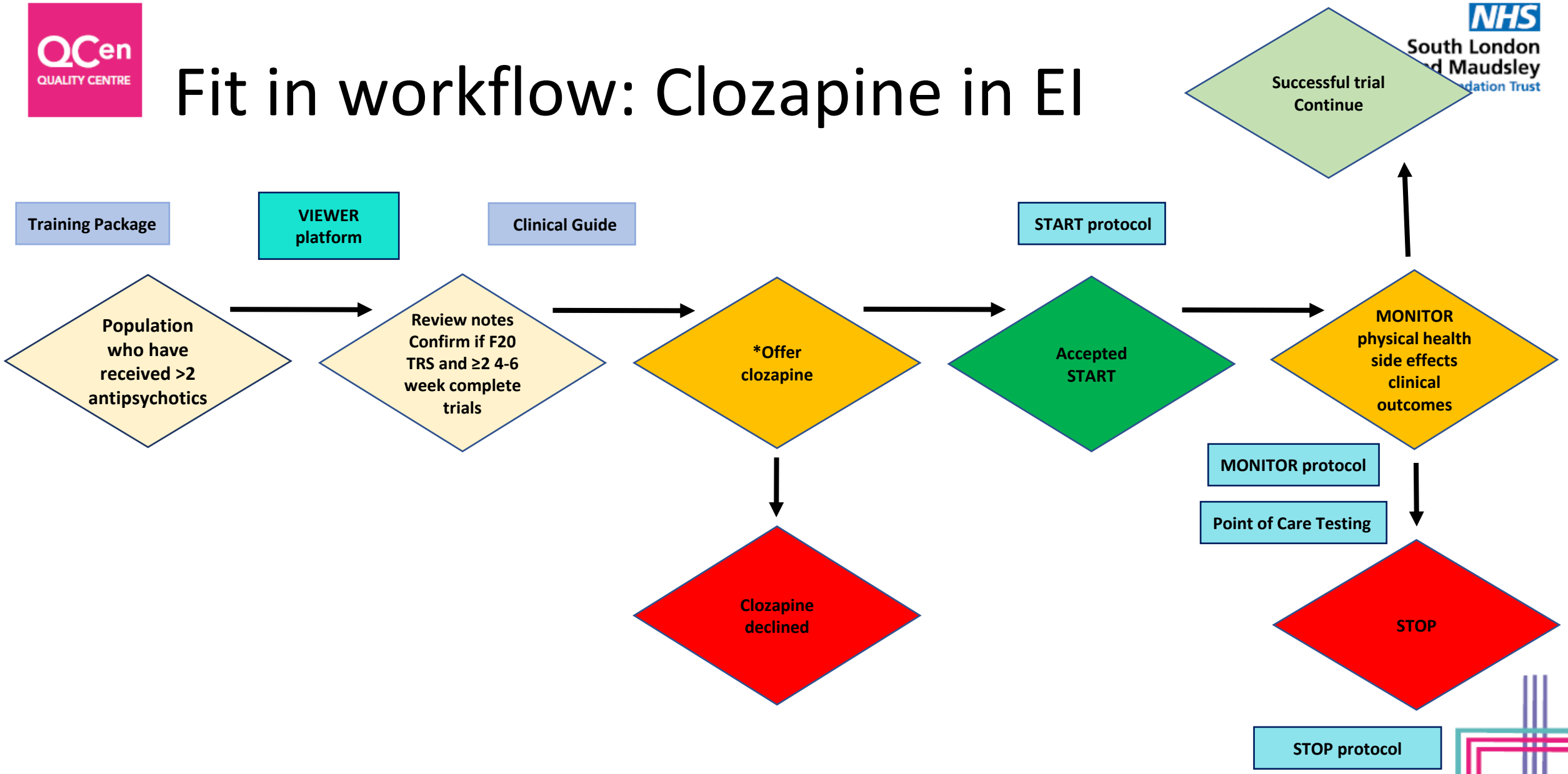
Bed Days by Ethnicity



Psychotherapy by Ethnicity

BMI by Ethnicity

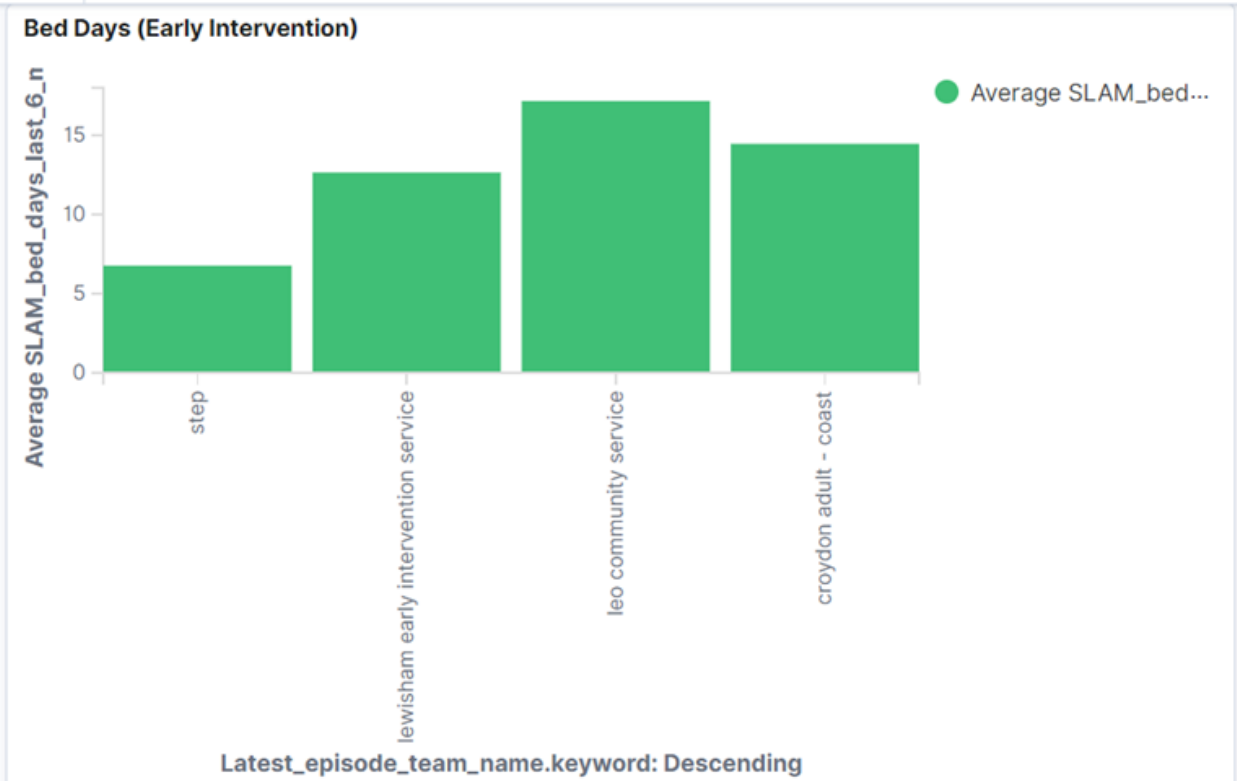
Fit in workflow: Clozapine in EI

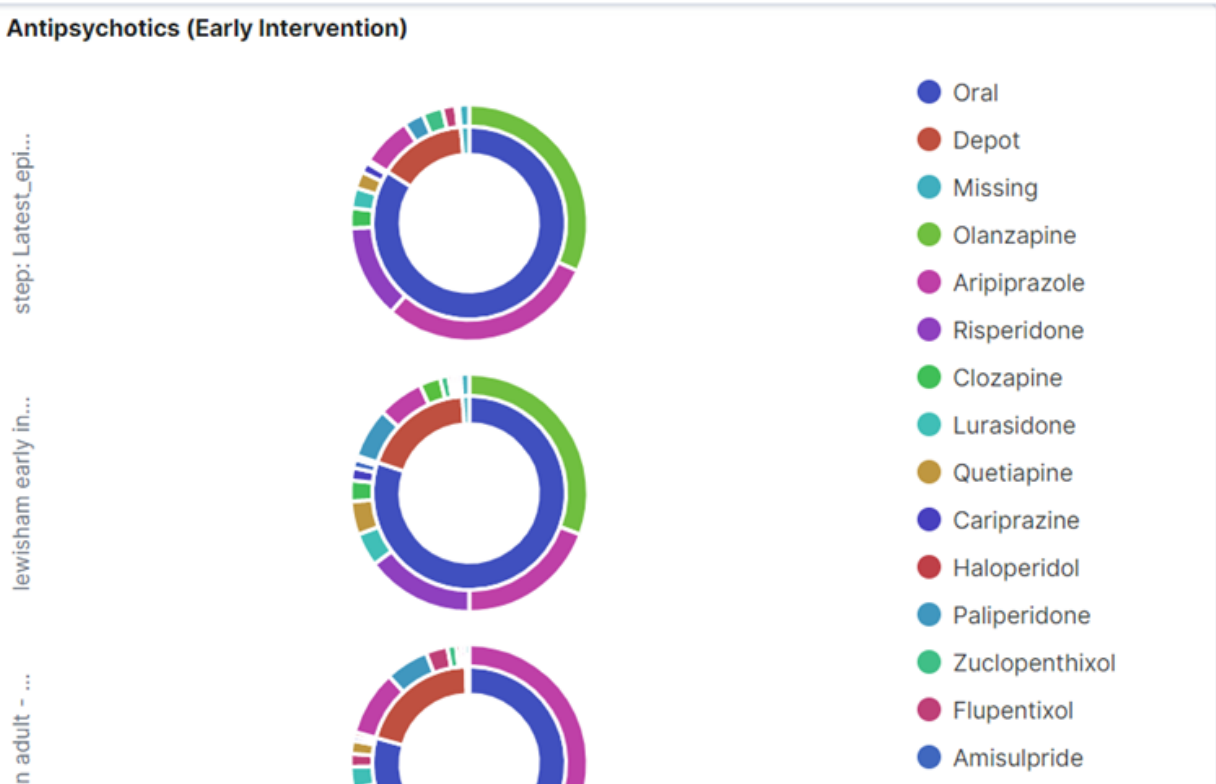


EI dashboard

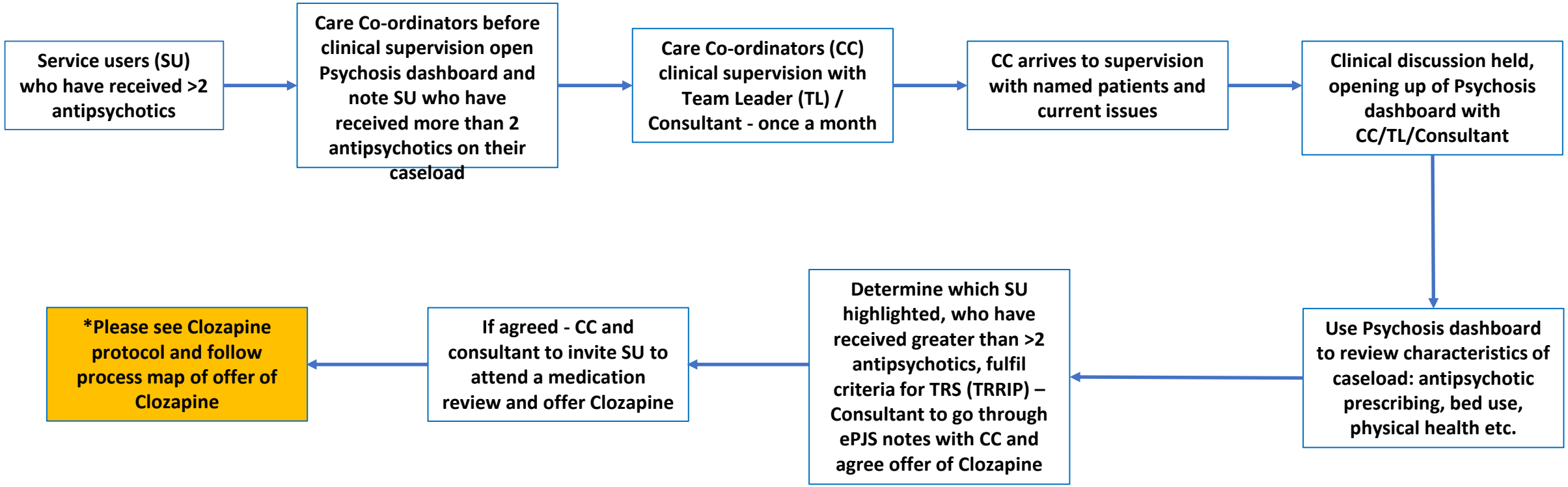
Dashboard / Psychosis CAG Dashboard - Early Interventions

Full screen | Share | Clone | Reporting | Edit





Fit in workflow: medication and physical health supervision

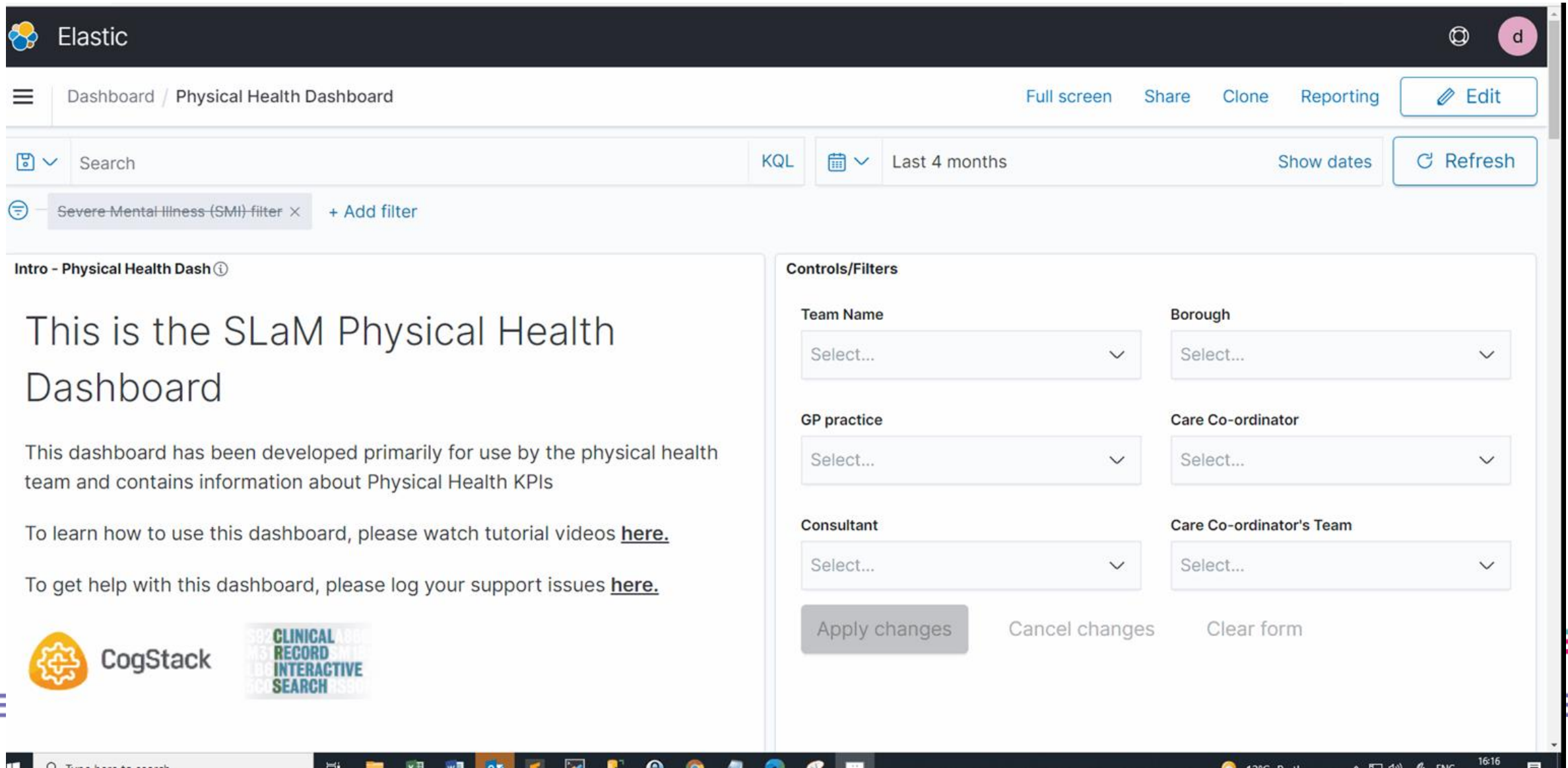


Traditional approach for physical health management

MINIMON (Minimum Monitoring Protocol)

		MONITORING SCHEDULE						
		Pre-treatment	Week 1 + 2 (daily)	Week 1 - 4 (weekly)	Month 1	Month 3	Month 6	Annual
PHYSICAL ASSESSMENT	Blood pressure	✓	✓	✓				✓
	Heart rate	✓	✓					
	Temperature		✓	✓				✓
	ECG	✓						✓
	Weight*	✓		✓				✓
	Physical exam[ø]	✓						✓
BLOOD TESTS	FBC[#]	✓						
	LFT	✓						✓
	U & Es	✓						✓
	CRP	✓						
	Lipids	✓				✓		✓
	Troponin	✓						
	Glucose (or HBA1c)	✓			✓	✓	✓	✓
SIDE EFFECT (E.G., GASS-C)					✓			✓
MEDICATION REVIEW[†]								✓
CLINICAL OUTCOMES[‡]		✓					✓	✓

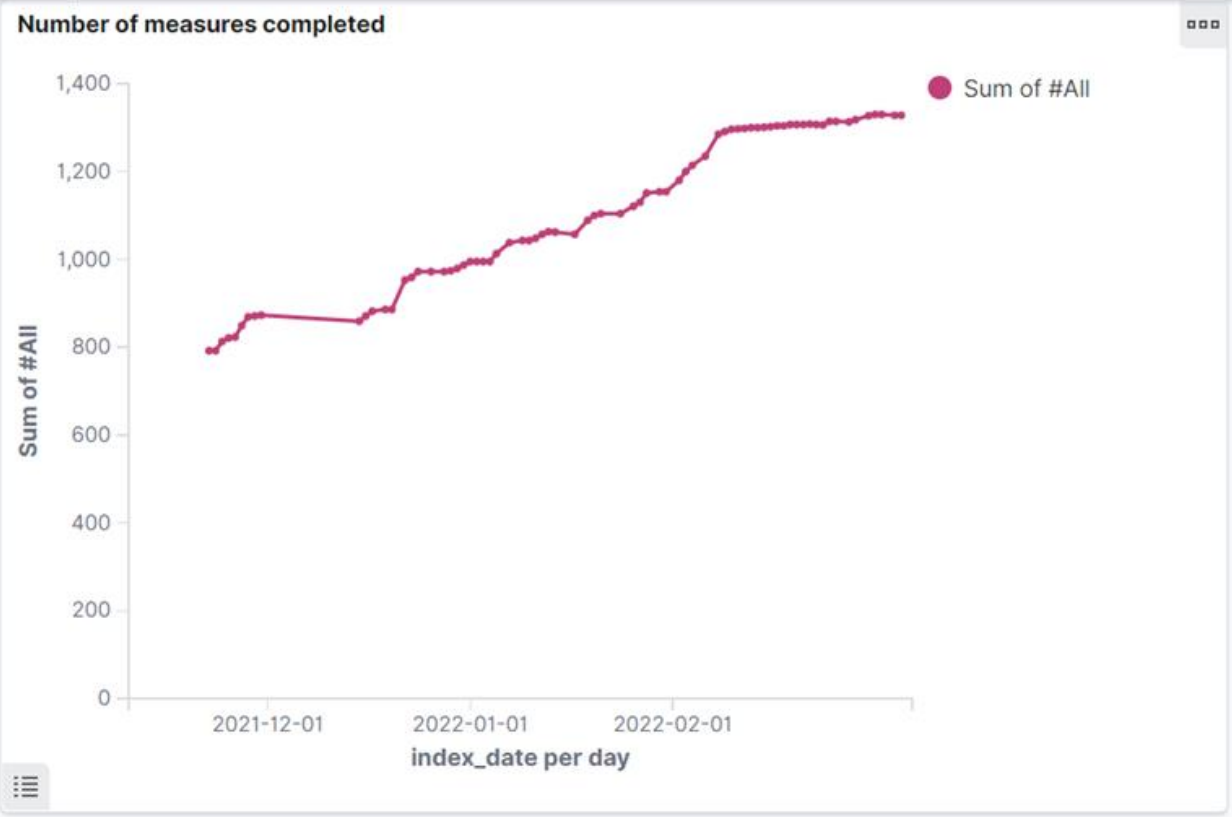
Manage physical health via VIEWER

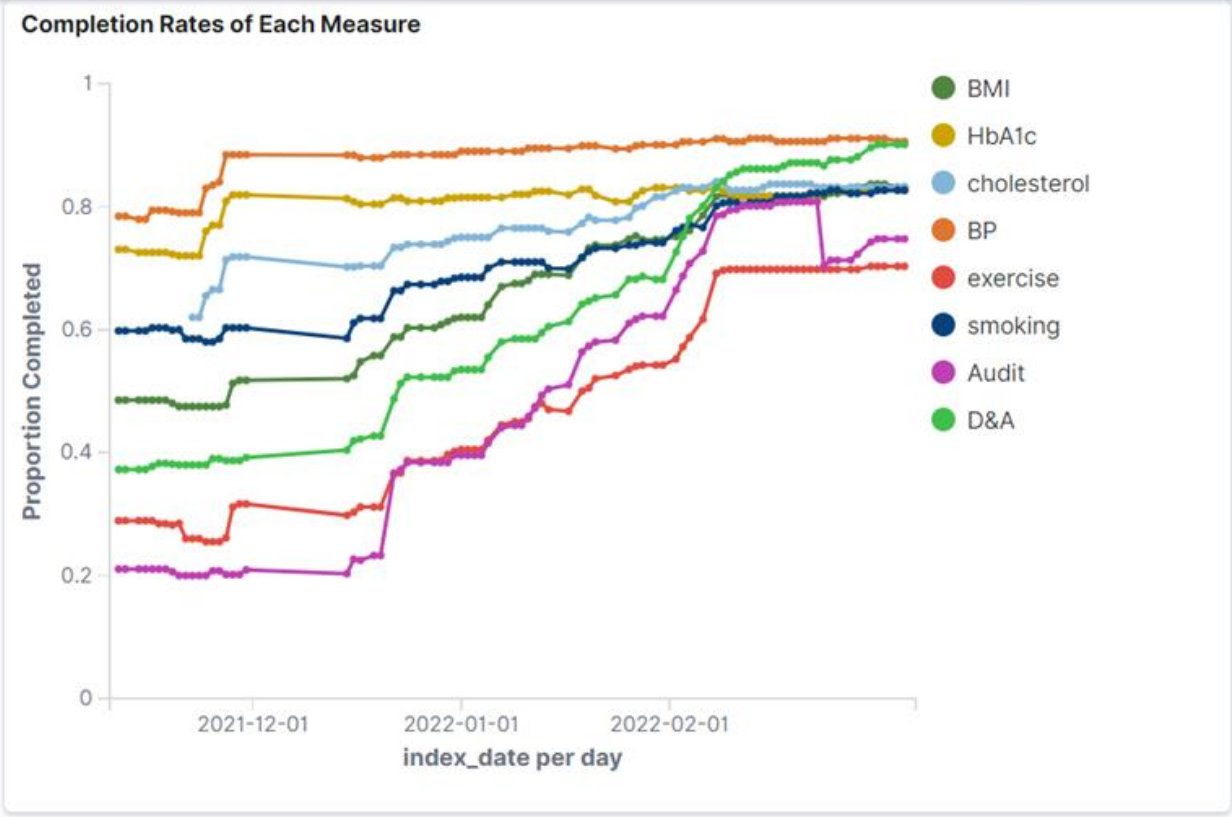


Improve physical health over time

Dashboard / Physical Health Dashboard - Time Series Data

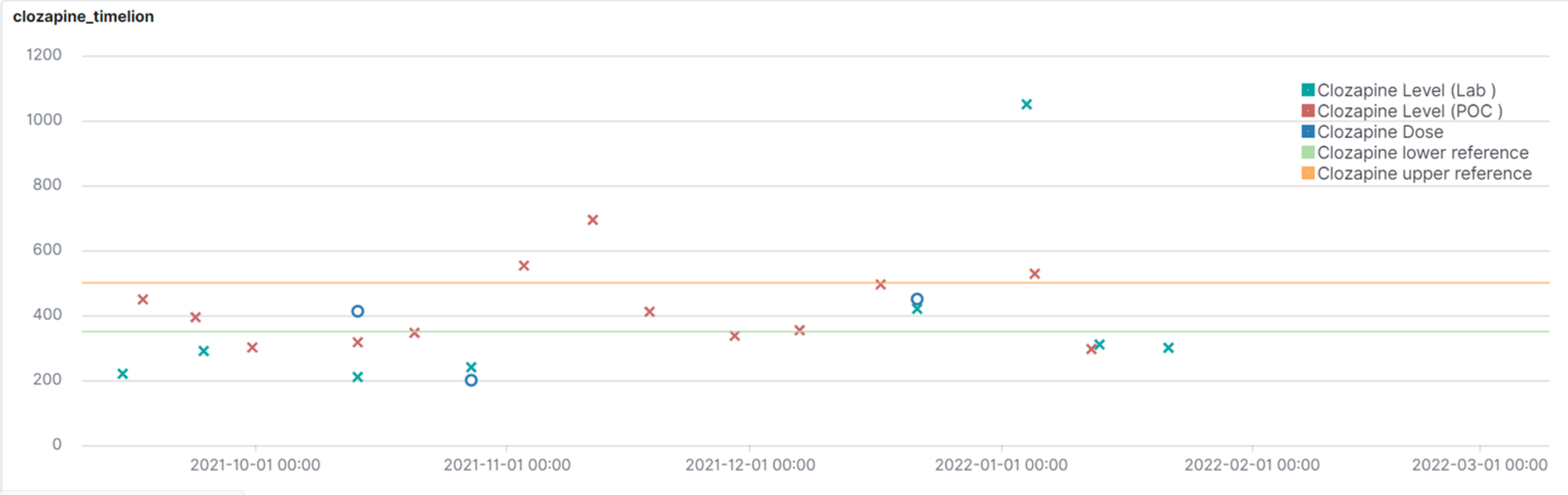
Full screen | Share | Clone | Reporting | Edit



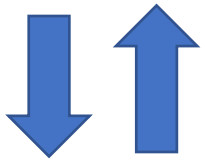


Point of care (POC) devices for clozapine monitoring

Integrate POC and Lab results

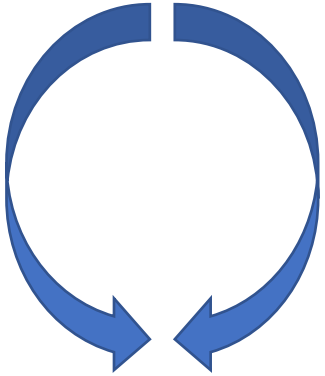


From Reactive Care to Proactive Care based on Need



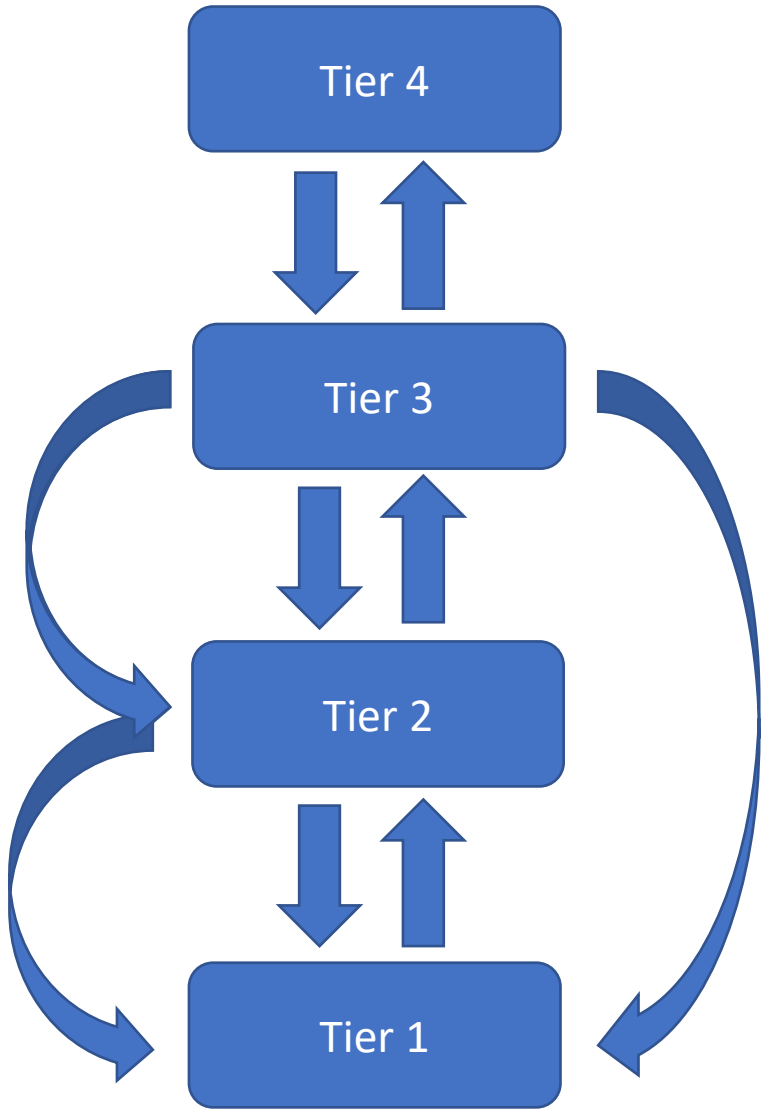
Traditional Pathway:

- Where patients meet criteria for a higher tier than they are currently in, they are 'stepped up' to be case-managed in that tier.
- When patients meet criteria allowing management in a lower tier, they are 'stepped down'.



New pathway:

- Higher tiers use population health approaches to identify discreet 'unmet needs' in lower tier.
- Using informatics, identify patients in lower tiers that may benefit from these interventions and engage with their lead clinician around delivery.
- Aim is to more proactively deliver interventions to patients in lower tiers to prevent deterioration.



Acknowledgement

VIEWER

Psychosis CAG
David Codling
Aaron Brewer
Rashmi Patel
Philip McGuire
Robert Harland

CRIS
Matthew Broadbent
Megan Pritchard
Daisy Kornblum
Robert Stewart

CogStack
Tao Wang
Yamiko Msosa
Angus Roberts
Richard Dobson

Quality Centre
Barbara Arroyo
Serena Lai
Claire Delaney-Pope
Nicola Byrne
Michael Holland

Care Teams
Southwark North East Promoting Recovery Team (SNEPRT);
Croydon Early Intervention In Psychosis (COAST);
Physical Health Team

Special thanks to Vicky Baughen (from COAST), Martin Ford (SNEPRT) Richie Morton (Service user).

MiADE
Medical information AI Data Extractor

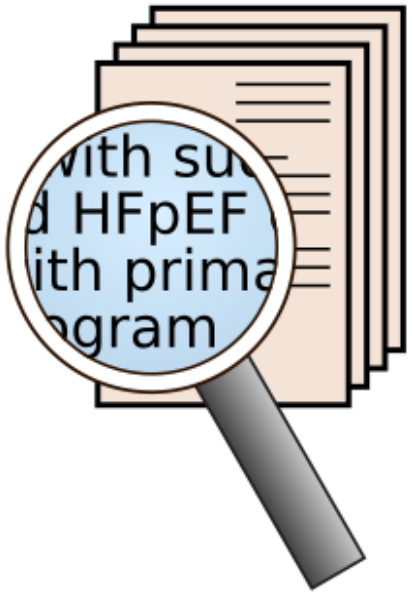
Natural language processing to support point of care structured documentation in electronic health records

Dr Anoop D. Shah, Associate Professor, UCL

Dr Wai Keong Wong, Chief Research Information Officer, UCLH

The need for structured health records

Paper



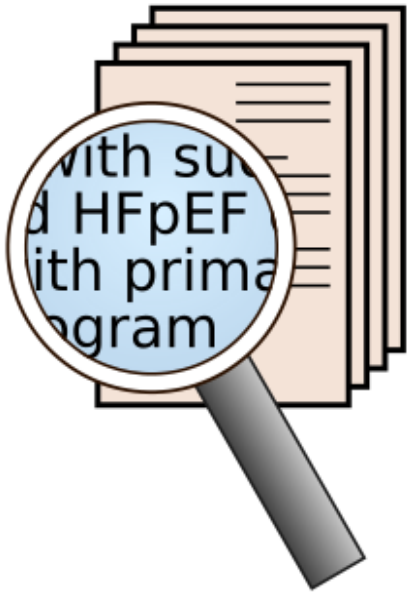
Electronic paper
(unstructured)

More and more NHS Trusts now have electronic health records

– but without structure, they are just an electronic pile of paper

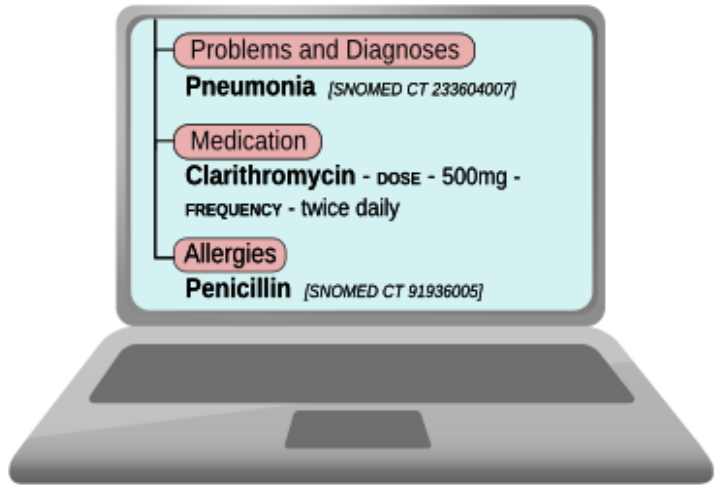
The need for structured health records

Paper

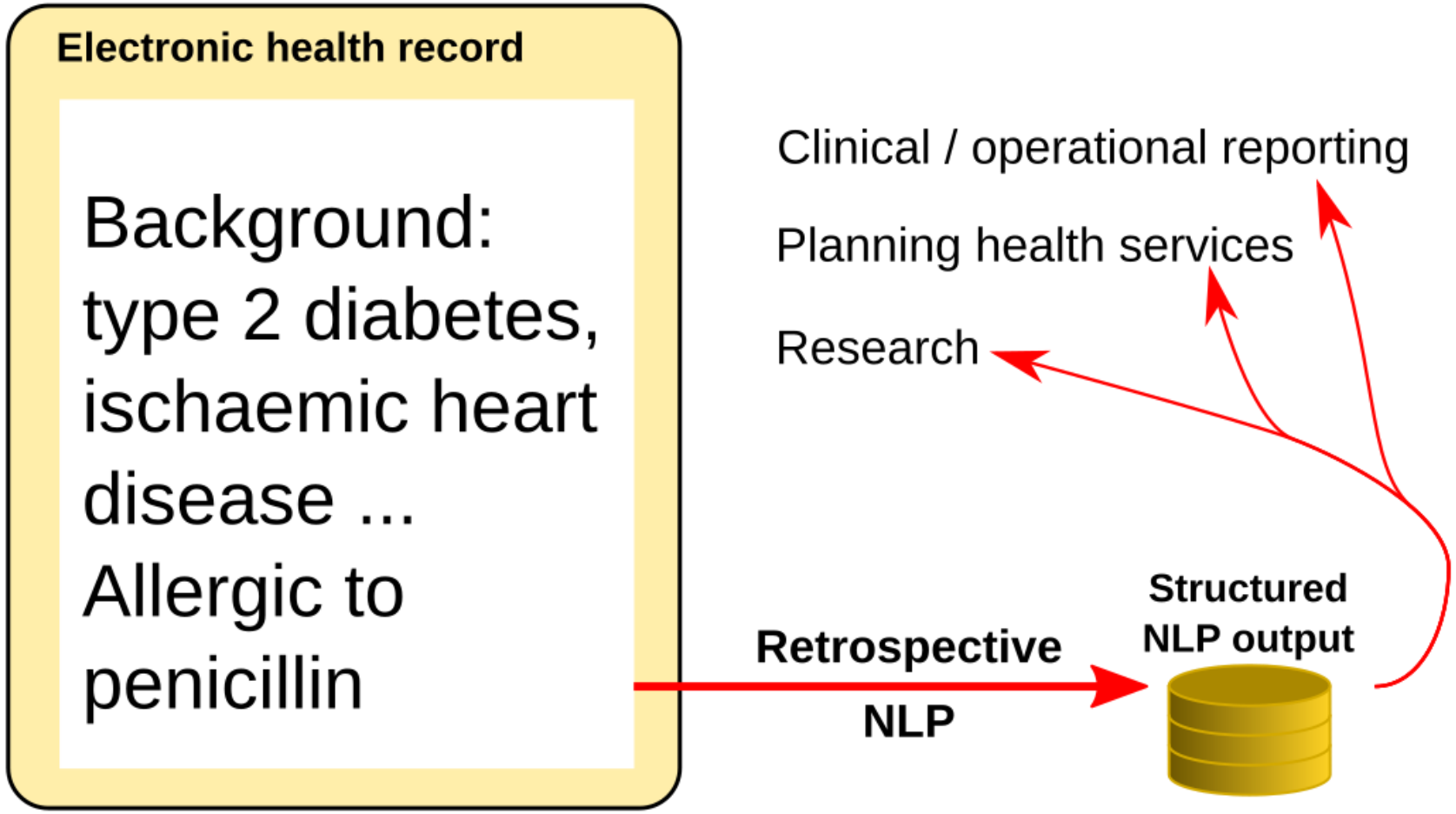


Electronic paper (unstructured)

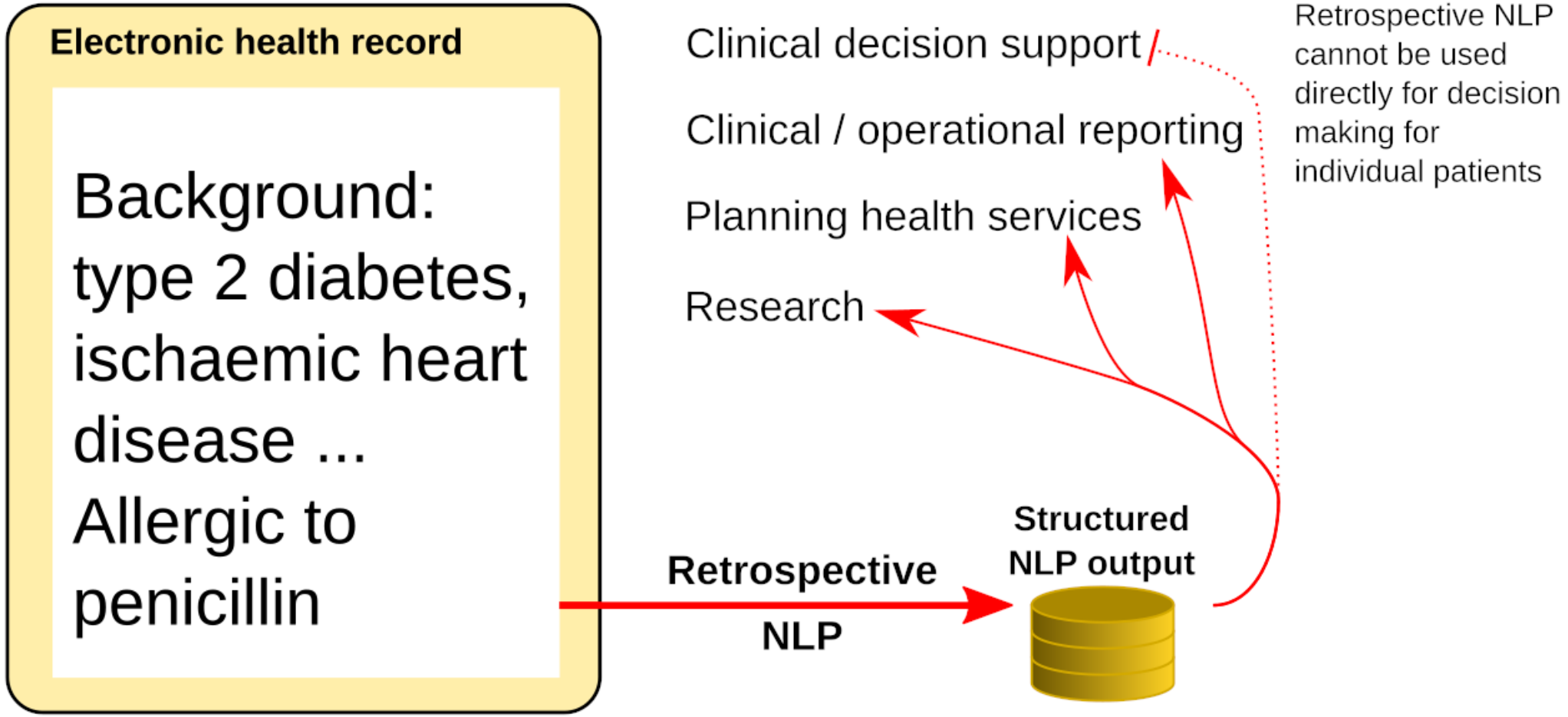
Structured electronic health record



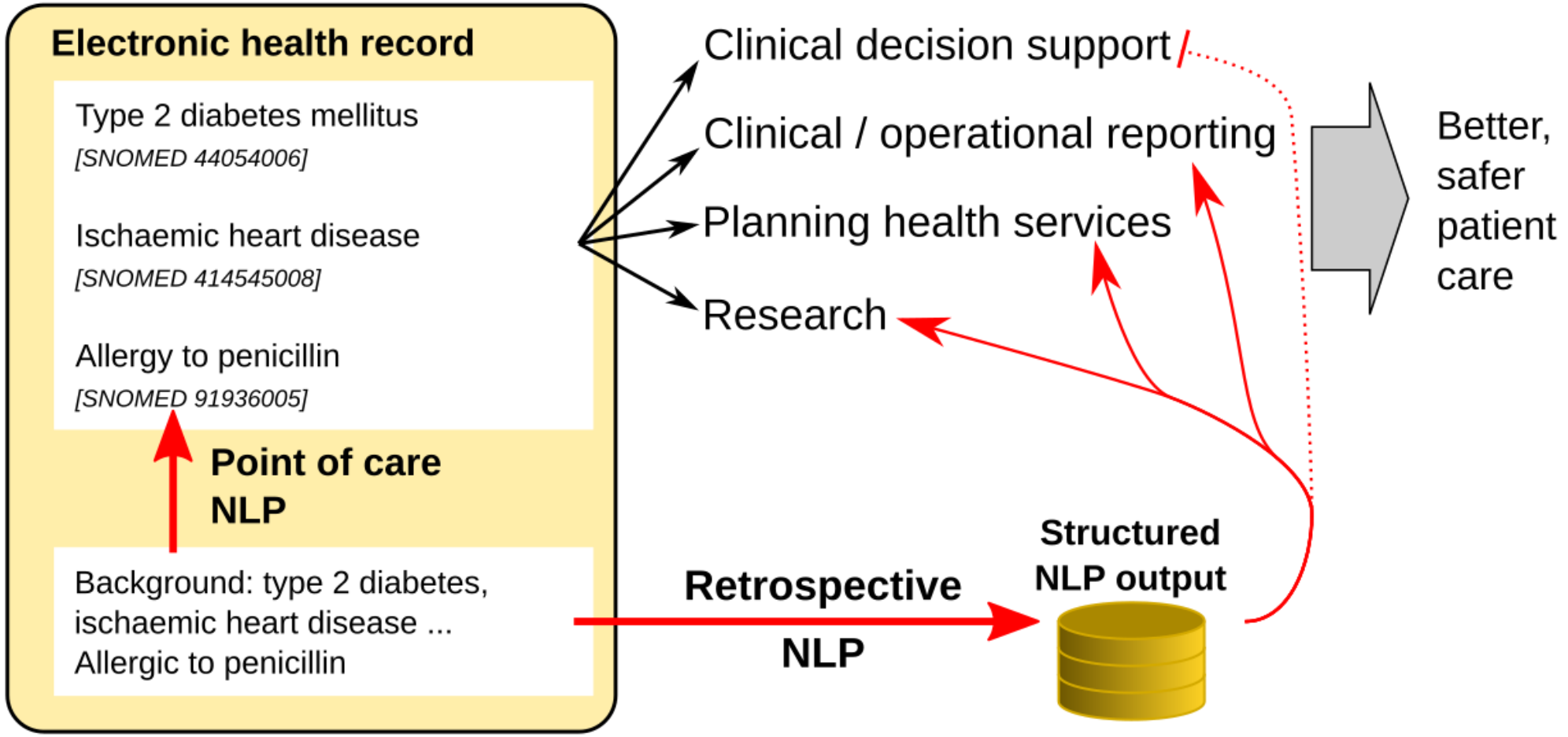
Converting free text to structured data for secondary uses



Converting free text to structured data for secondary uses



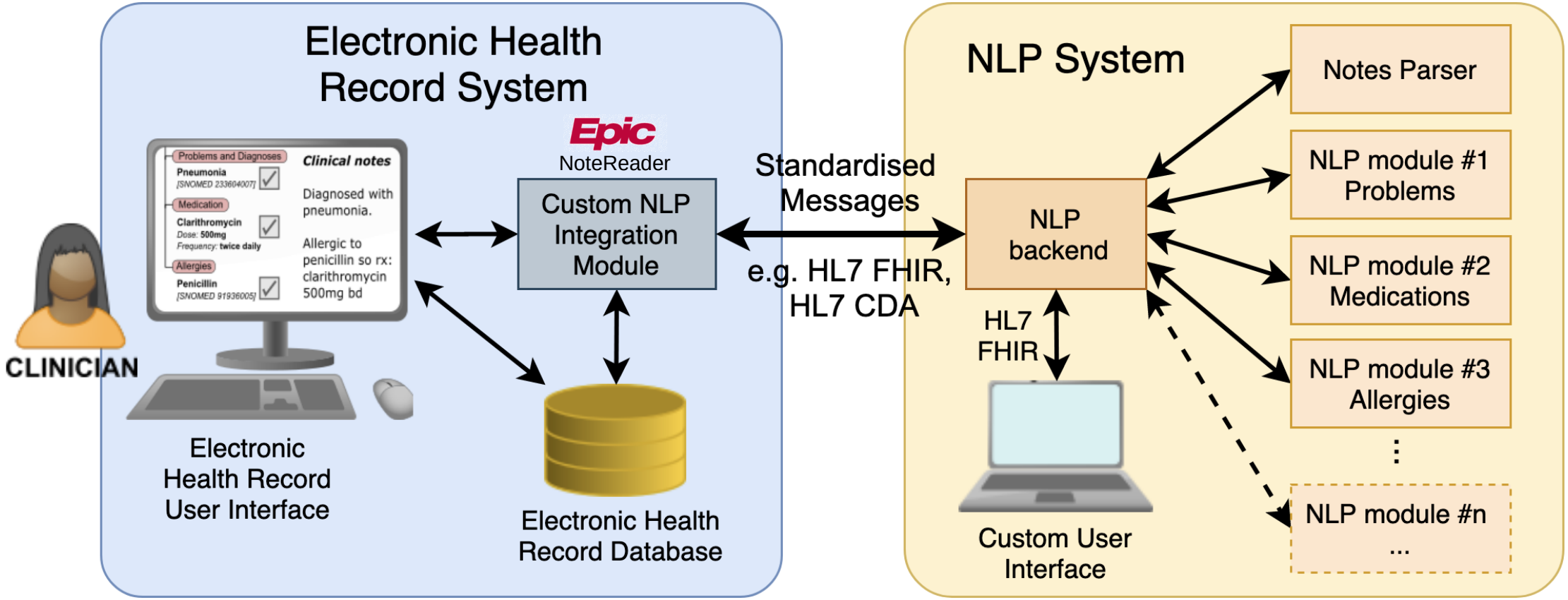
Natural language processing (NLP) at the point of care



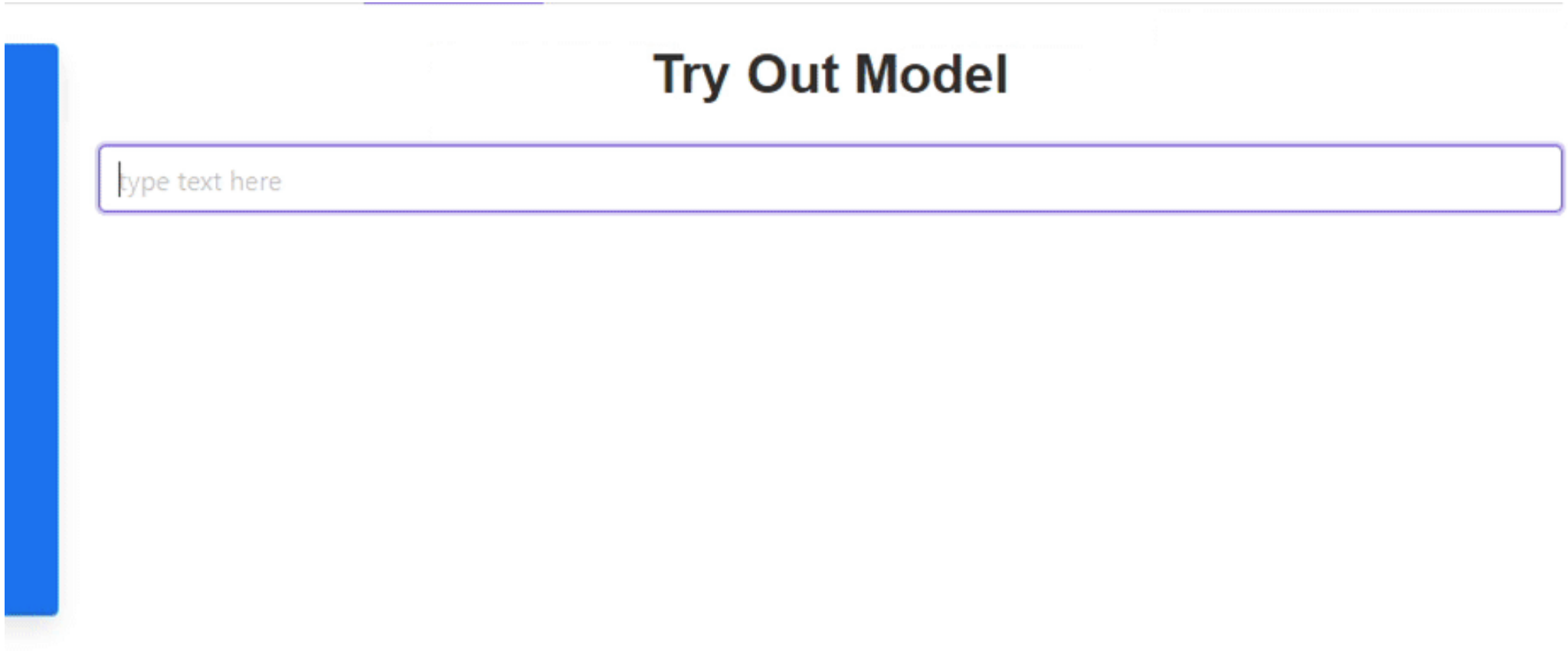
Principles of our approach

- Optimise **technology** and **human factors**
- **Open source** software
- Use **data standards** (HL7 CDA and FHIR) to be **interoperable** with standards-compliant electronic health record system
- Continuous **improvement**
  - Collection of clinical responses to NLP suggestions
  - Ongoing development to refine NLP algorithms and user interface
- **Patient involvement**
  - Co-develop evaluation study and dissemination strategy
- Enable **widespread adoption** to maximise patient benefit

Design of NLP system



Demonstration



Development of MiADE

- Problems (diagnoses), medication and allergies
- Based on the open source 'MedCAT' named entity recognition tool
  - Recognises medical terms in text based on unsupervised and supervised learning
  - Meta-annotations for context (e.g. negation, suspected, historic)
  - Filtering of output to omit entries already in structured data
- Timeline:
  - Currently 6 months into year 1
  - Clinical evaluation study and hackathon planned for year 2

Advantages of NHS-led design

- Model created using NHS medical text
- Ability to share models between NHS sites
- Tailor to clinical settings
  - inpatient vs outpatient
  - different specialities
- To extract new items of information
  - Add concepts / synonyms
  - Train on additional data
  - Collect data

Acknowledgements

- The team:
  - James Brandreth, software developer
  - Jennifer Jiang, software developer
  - Kawsar Noor, software developer
  - Yogini Jani, UCL School of Pharmacy and UCLH
  - Leilei Zhu, UCLH clinical data standards lead
  - Enrico Costanza, UCL Interaction Centre
  - Neil Sebire, Great Ormond Street Hospital
  - Richard Dobson, UCL and King's College London
  - Folkert Asselbergs, UCL and UMC Utrecht
- Lay members of the steering committee
- CogStack Community
- Funding: NIHR AI award, with support from the UCL/UCLH BRC

CogStack: Looking to the future

Stage 3 AI Award winner

Next 18 months:

Mature CogStack deployments at select Trusts: KCH, GSTT, UCLH, SLaM, UHB

Demonstrate Clinical Coding exemplar use case

Plan for broader rollout and funding sustainability

CogStack: Get Involved

Join our growing community of deployment sites.

Get in touch to discuss your use case: contact@cogstack.org

Source-code: https://github.com/CogStack

Questions

https://cogstack.org/

contact@cogstack.org

Supported by:

NIHR | Maudsley Biomedical Research Centre

UKRI Innovate UK

Office for Life Sciences

King's College London

NHS South London and Maudsley NHS Foundation Trust

NHS King's College Hospital NHS Foundation Trust

NHS Guy's and St Thomas' NHS Foundation Trust

NHS University College London Hospitals NHS Foundation Trust