

NLP use cases:

(i) King's Health Partners

(ii) The Maudsley CRIS dataset and NLP service



Examples from King's Health Partners



TNM staging in PET-CT reports for lung cancer

Stephen Barlow, Sugama Chicklore, Yulan He, Thomas Wagner, Anna Barnes, Gary Cook

Findings:

An FDG scan was acquired from skull base to upper thighs together with a low dose CT scan for attenuation correction and image fusion.

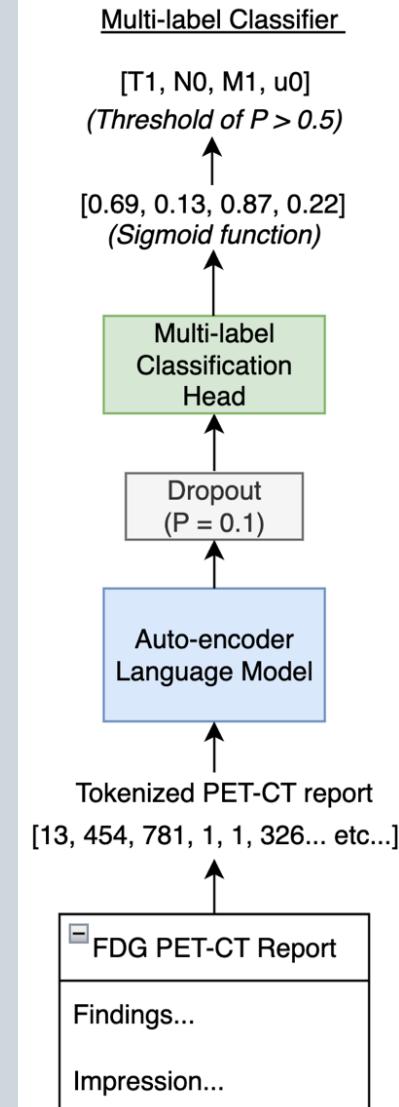
There is a 3.5cm right lower lobe mass which shows intense FDG uptake (SUV max 14.3).

There is focal intense uptake in a right hilar node and a smaller subcarinal node. The left adrenal gland is enlarged and is predominantly of low attenuation. It shows low grade abnormal uptake (SUV max 3.7). There is a left paravertebral soft tissue mass at the C7 level which shows intense uptake and is eroding the anterior edge of C7. There is an area of increased uptake in the midline of the anterior floor of mouth, which is not typical for the physiological muscle activity sometimes seen at this site. No definite underlying CT correlate is present.

Impression:

Scan findings are consistent with a malignant right lung tumour with right hilar and subcarinal nodal involvement. The findings also suggest a soft tissue metastatic mass in the left C7 paravertebral region. The level of uptake in the left adrenal gland in comparison to the lung mass is relatively low and it is felt more likely that the adrenal is benign in nature. Clinical correlation of the anterior floor of mouth is recommended to further evaluate whether this area of activity is pathological.

Key:
T – finding
N – finding
M – finding



Seizure frequency extraction from medical records

Ben Holgate, Shichao Fang, Mark Richardson

Clear example:

“We went through some of his seizures and in March he had two convulsions and three or four petit mal.”

Seizure diary example:

“Seizures: Partial seizures: July x 23, Aug x 0, Sept x 1, Oct so far x 7 (x1 daily 7th to 10th, 14th x1, 15th x 2, 18 x1.”

Ambiguous example:

“Louise and her mum confirm no seizures with her last seizure was possibly in November but they are not sure.”

Model prompted to be a “professional neuroscientist who is responding to fellow neuroscientists” and to give “succinct answers”.
Provided with 11 examples

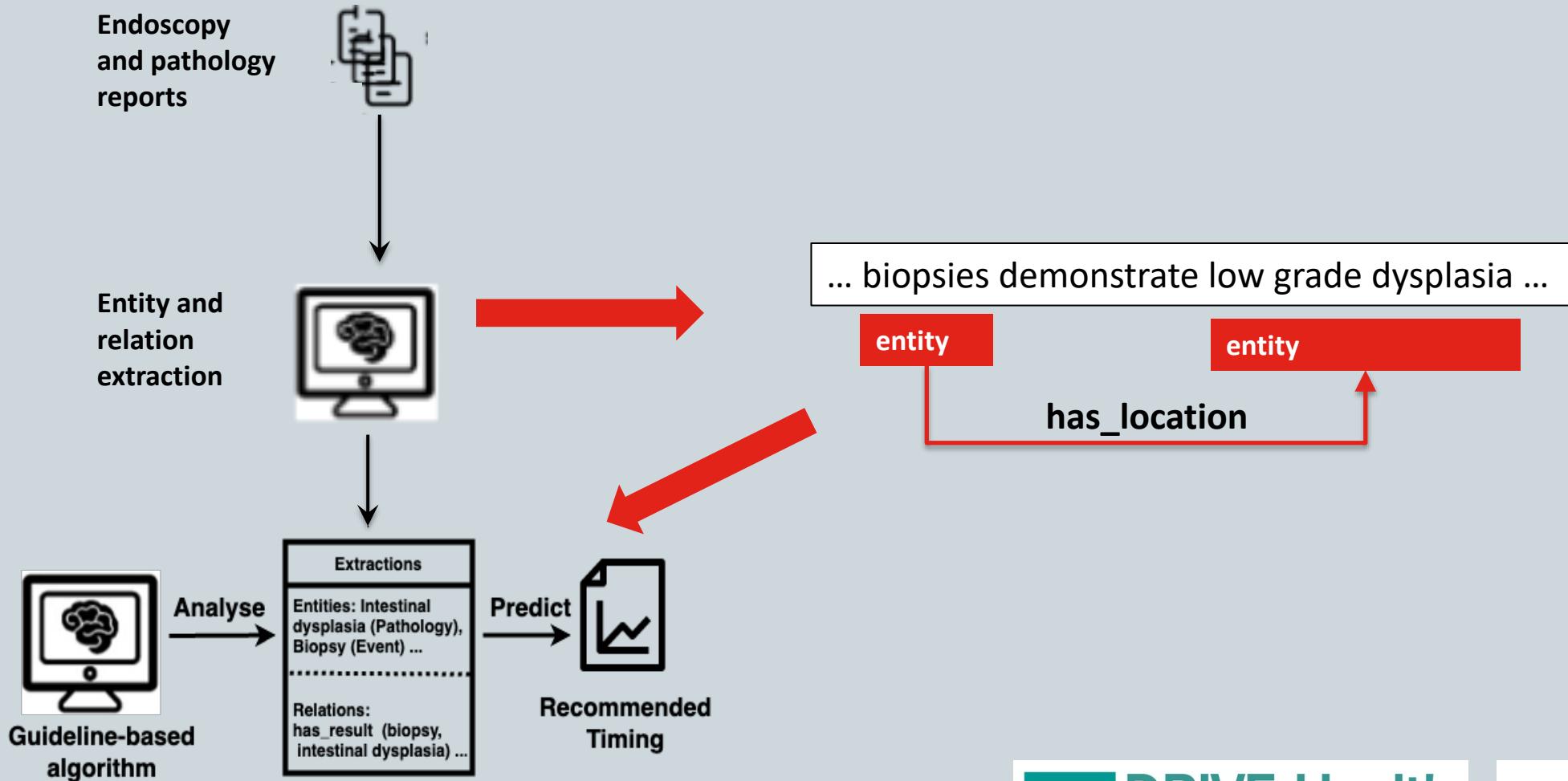
Read the following context then work through these 3 steps.

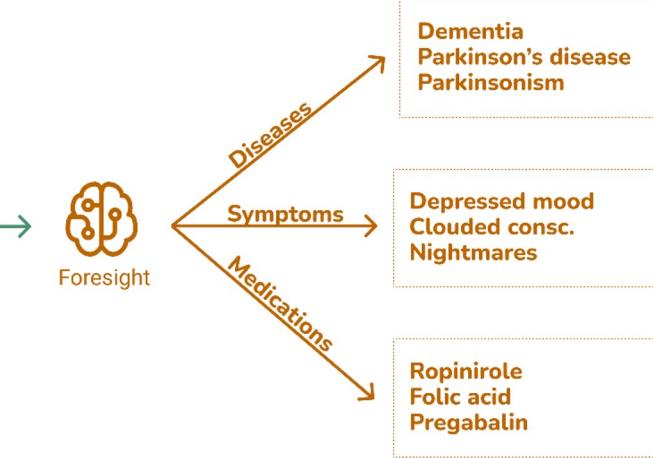
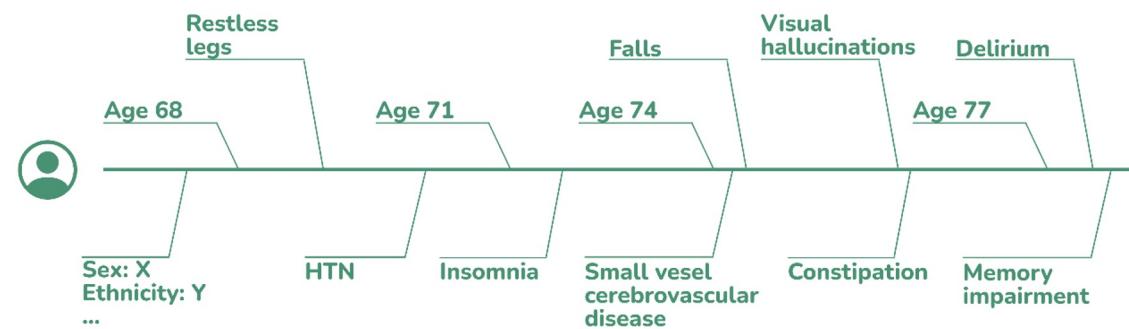
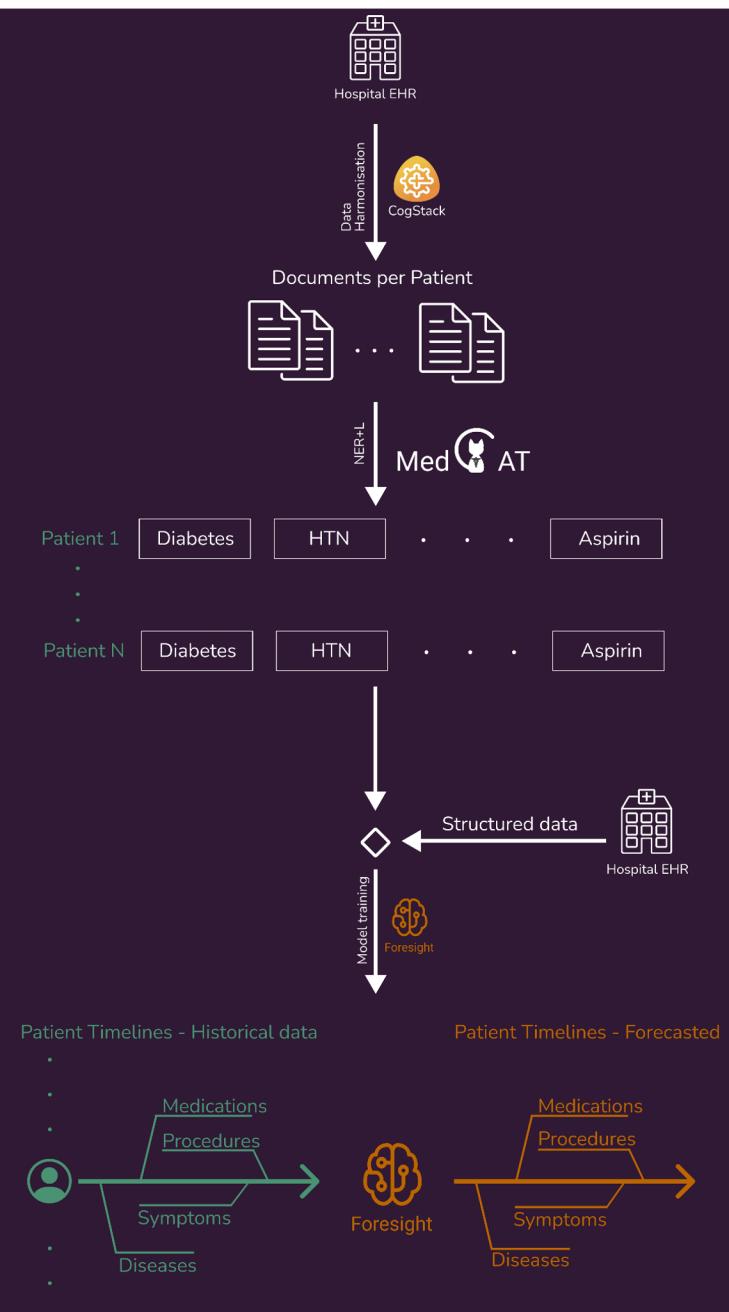
1. Determine whether the context has any information about the frequency of the epilepsy patient's seizures.
2. If the context does not have any information about the frequency of the epilepsy patient's seizures, then you answer: ‘I do not know.’
3. If the context does have information about the frequency of the epilepsy patient's seizures, then you estimate the frequency of the epilepsy seizures and express the frequency in terms of per year, per month, per week, or per day, whichever is most relevant.



Extracting information for Barrett's surveillance

Leo Zhang, Sebastian Zeki, Angus Roberts





	KCH (Precision)		SLaM (Precision)	
	TOP-1	TOP-10	TOP-1	TOP-10
Precision Overall	0.667	0.917	0.658	0.938
Precision Disorders	0.605	0.874	0.637	0.917
Precision Substances	0.716	0.950	0.731	0.958

Demo and Paper:
<https://foresight.sites.er.kcl.ac.uk/>



What's the difference between Levetiracetam and Keppra?



Levetiracetam and Keppra are exactly the same medication. Keppra is the brand name, and Levetiracetam is the generic name.



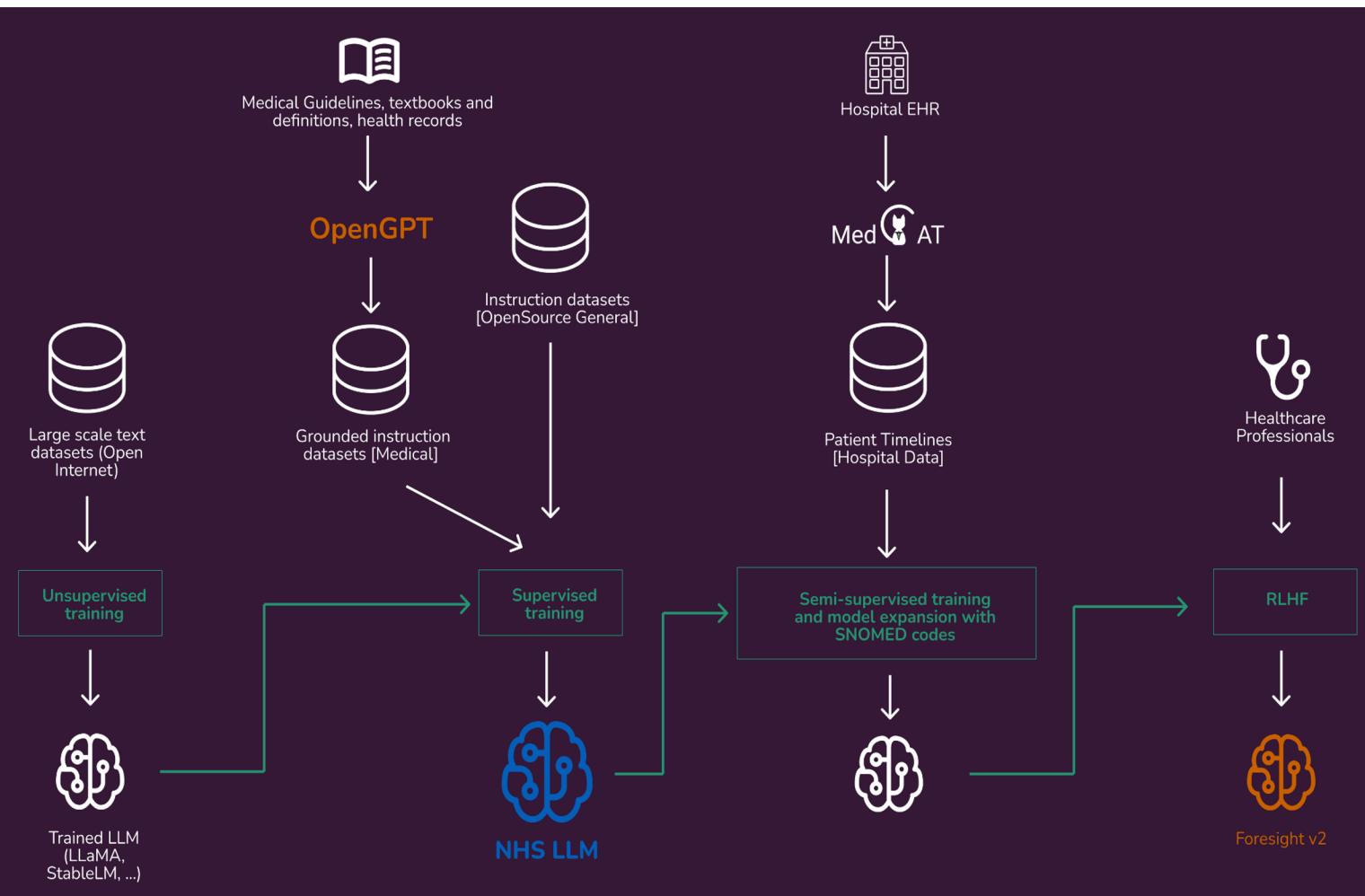
35 year old woman with abdominal distension, nausea and vomiting, amenorrhoea. What are the most likely diagnoses?



A 35-year-old woman presenting with abdominal distension, nausea, vomiting, and amenorrhea could have any of the following most likely diagnoses:

1. Pregnancy: The most likely diagnosis in this case, especially given her age, is early pregnancy. This needs to be confirmed by a pregnancy test and ultrasound.
2. Ectopic pregnancy: If a pregnancy test is positive, the embryo may have implanted outside the uterus (ectopic pregnancy), resulting in abdominal pain and other symptoms.
3. Ovarian cyst: A fluid-filled sac that develops on an ovary, causing pain and bloating.
4. Inflammatory bowel disease (IBD): Chronic inflammatory conditions like Crohn's disease or ulcerative colitis can cause abdominal pain, diarrhea, and weight loss.
5. Appendicitis: Inflammation of the appendix, causing abdominal pain, nausea, and vomiting.

Most of these conditions require a thorough clinical evaluation and diagnostic tests to identify the underlying cause of the symptoms.



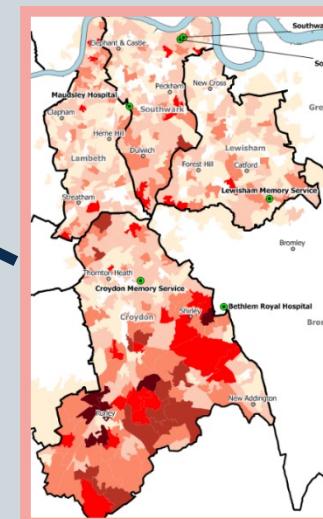
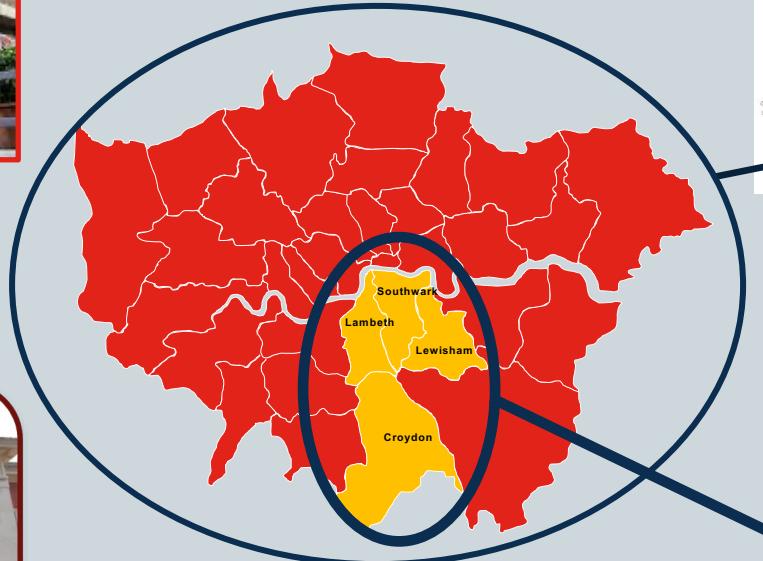
Examples from the Maudsley BRC

The Maudsley Biomedical Research Centre (BRC)

King's College London (KCL)



South London and Maudsley (SLaM)



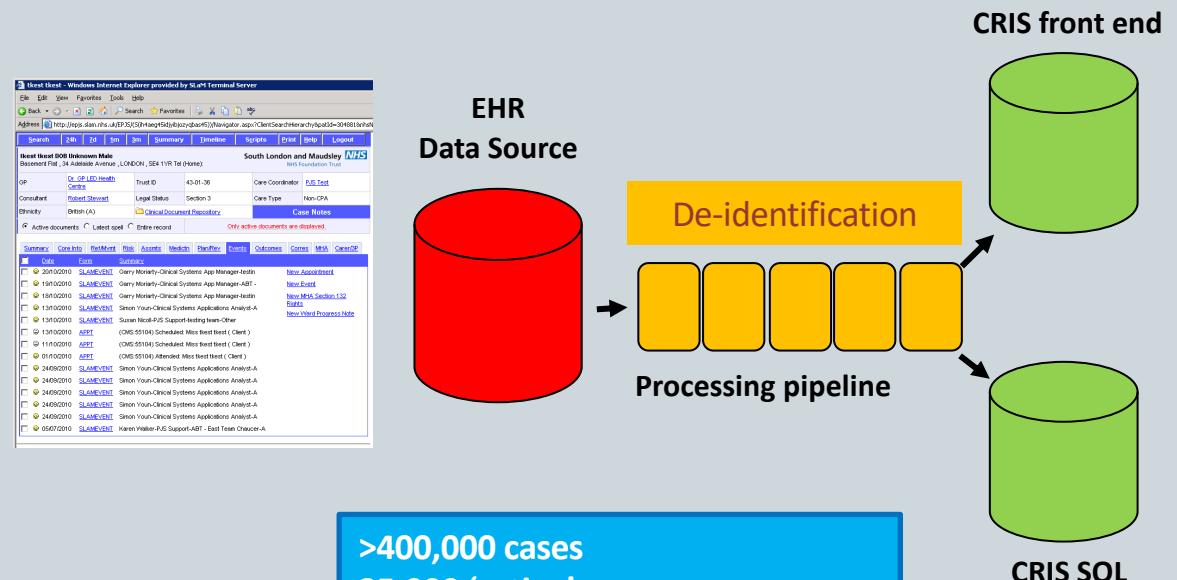
SLaM service provision
All aspects of specialist
mental health care

Defined catchment
1.36 million residents
Southwark, Lambeth,
Lewisham, & Croydon

**Complete EHRs since
2007**

The CRIS dataset and service

S92 CLINICAL A86G M31 RECORDS M1B LB6 INTERACTIVE 5C0 SEARCH RS90B



**>400,000 cases
35,000 ‘active’ cases
125 tables
6500 fields
30m documents**

**Set up in 2007-08 (NIHR funding)
Re-build and enhancement in 2017
Exported successfully to other UK Trusts
>300 research papers to date**

Researcher self-service

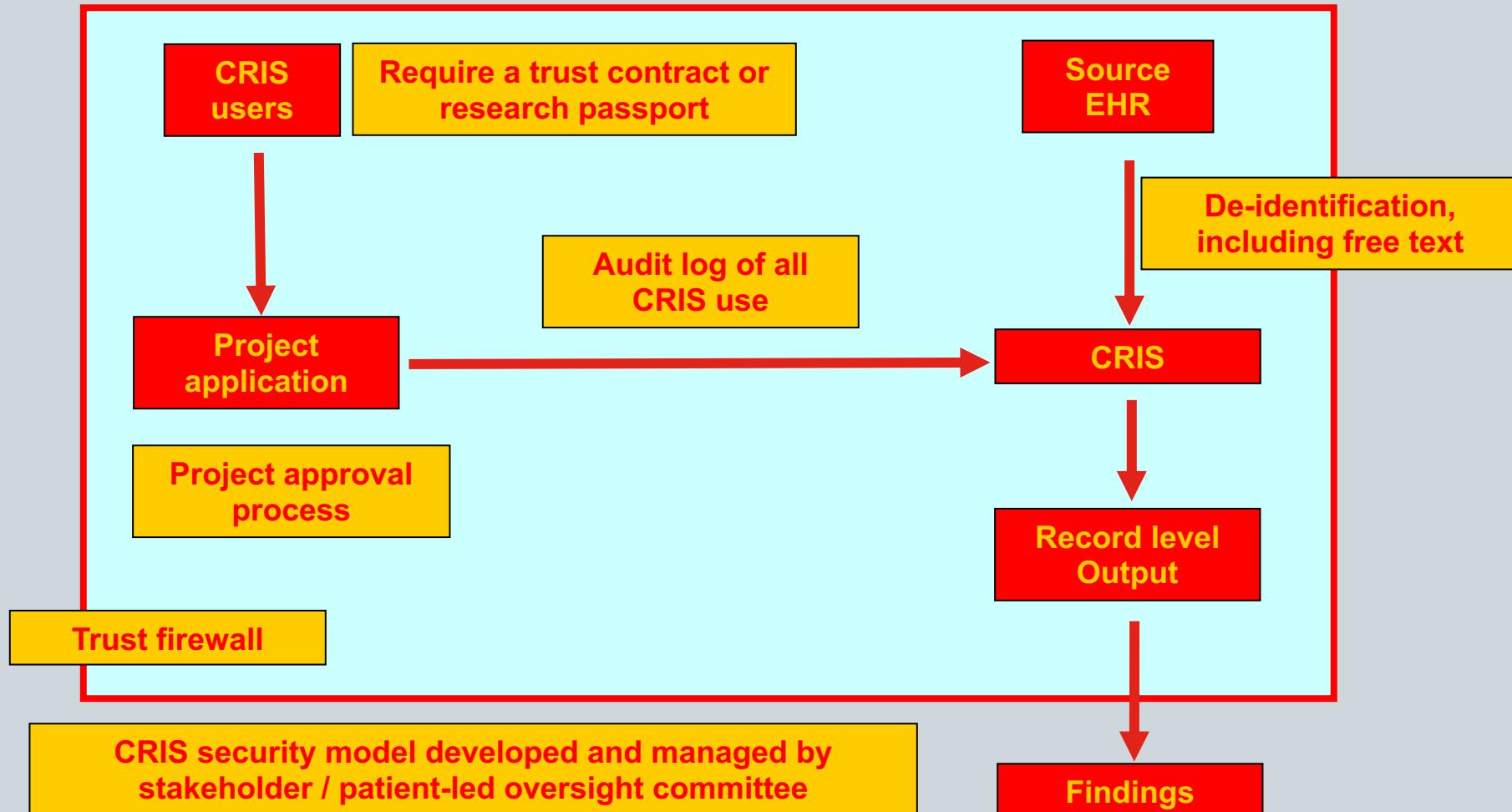
CRIS analyst service

The screenshot displays the Microsoft SQL Server Management Studio (SSMS) interface. The Object Explorer on the left shows a tree view of database objects, including tables like 'tbl_Inv', 'tbl_Cat', and 'tbl_Prod', and stored procedures like 'sp_InvCount' and 'sp_CatProd'. The Object Explorer Details pane provides detailed information about selected objects, such as column definitions for 'tbl_Inv'. The Results pane at the bottom shows the output of a query on 'tbl_Inv', listing items like 'Bread', 'Milk', 'Cheese', etc., with their respective InvID and InvName.

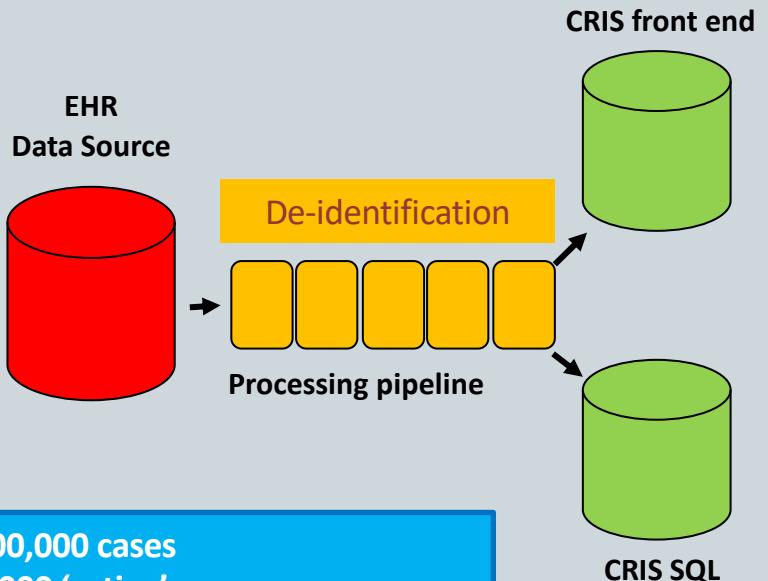
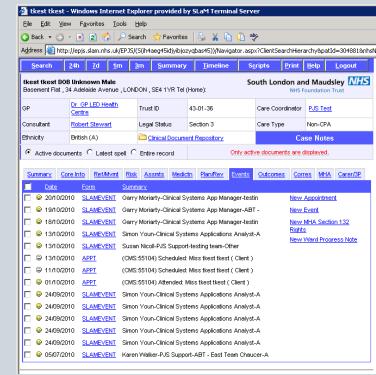
Data access: user led governance

Research ethics approval 2008, 2013, 2018

Numerous amendments for data linkages



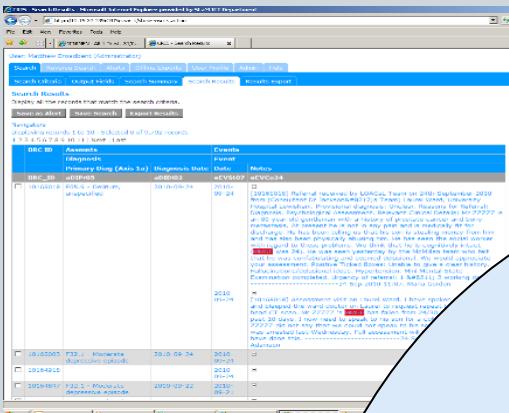
The CRIS dataset and service



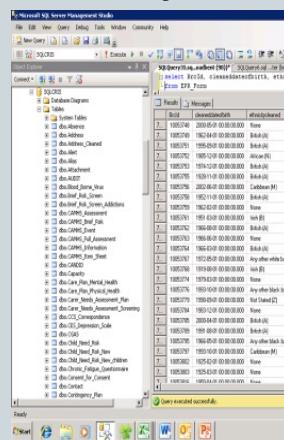
>400,000 cases
 35,000 'active' cases
 125 tables
 6500 fields
 30m documents

Set up in 2007-08 (NIHR funding)
 Re-build and enhancement in 2017
 Exported successfully to other UK Trusts
 >300 research papers to date

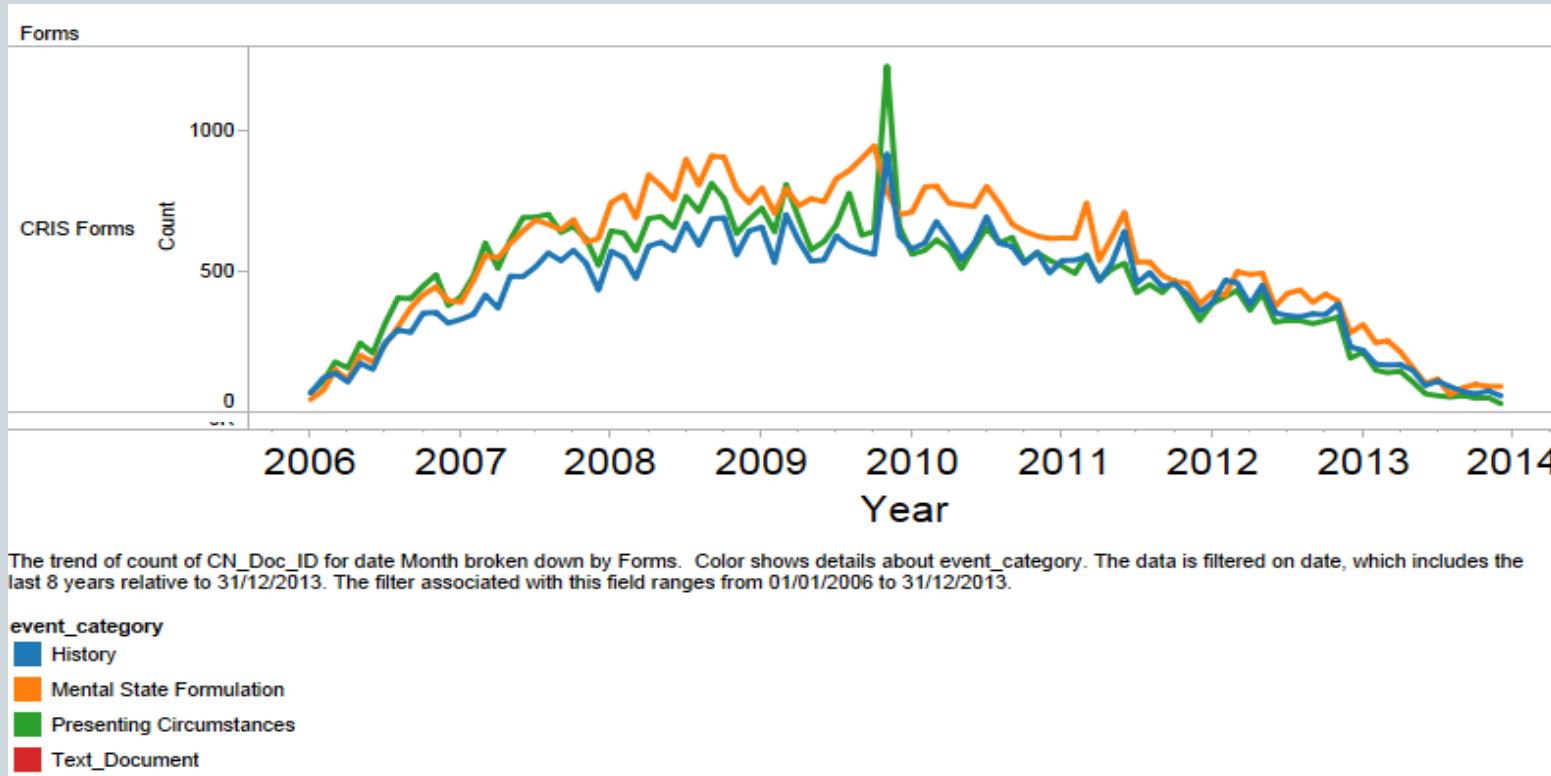
Researcher self-service



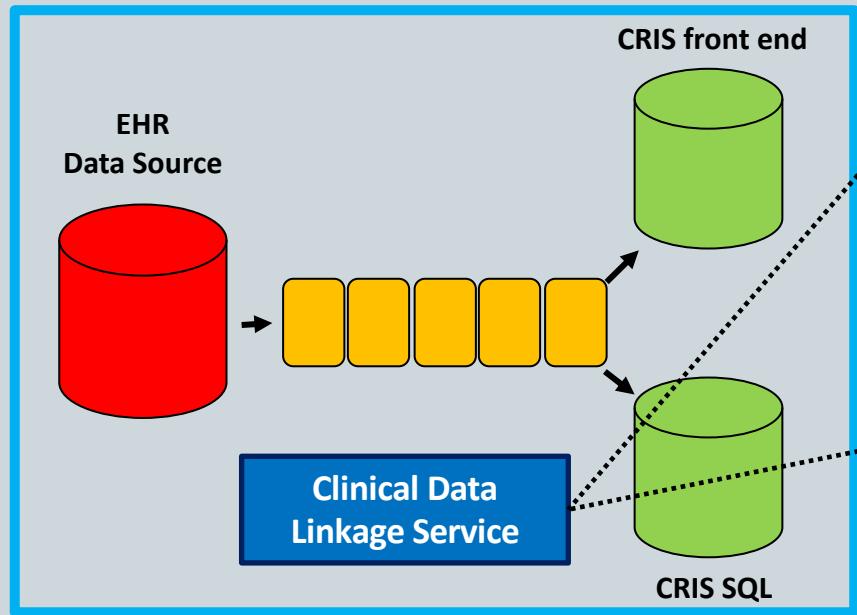
CRIS analyst service



Problem: structured data field use in CRIS



Data expansion 1 - database linkages



Internal linkages

Pharmacy dispensing data
 Research databases (e.g. GAP)
 Bioresource and imaging
 Psychological therapies (IAPT)
 Clozapine monitoring

External linkages

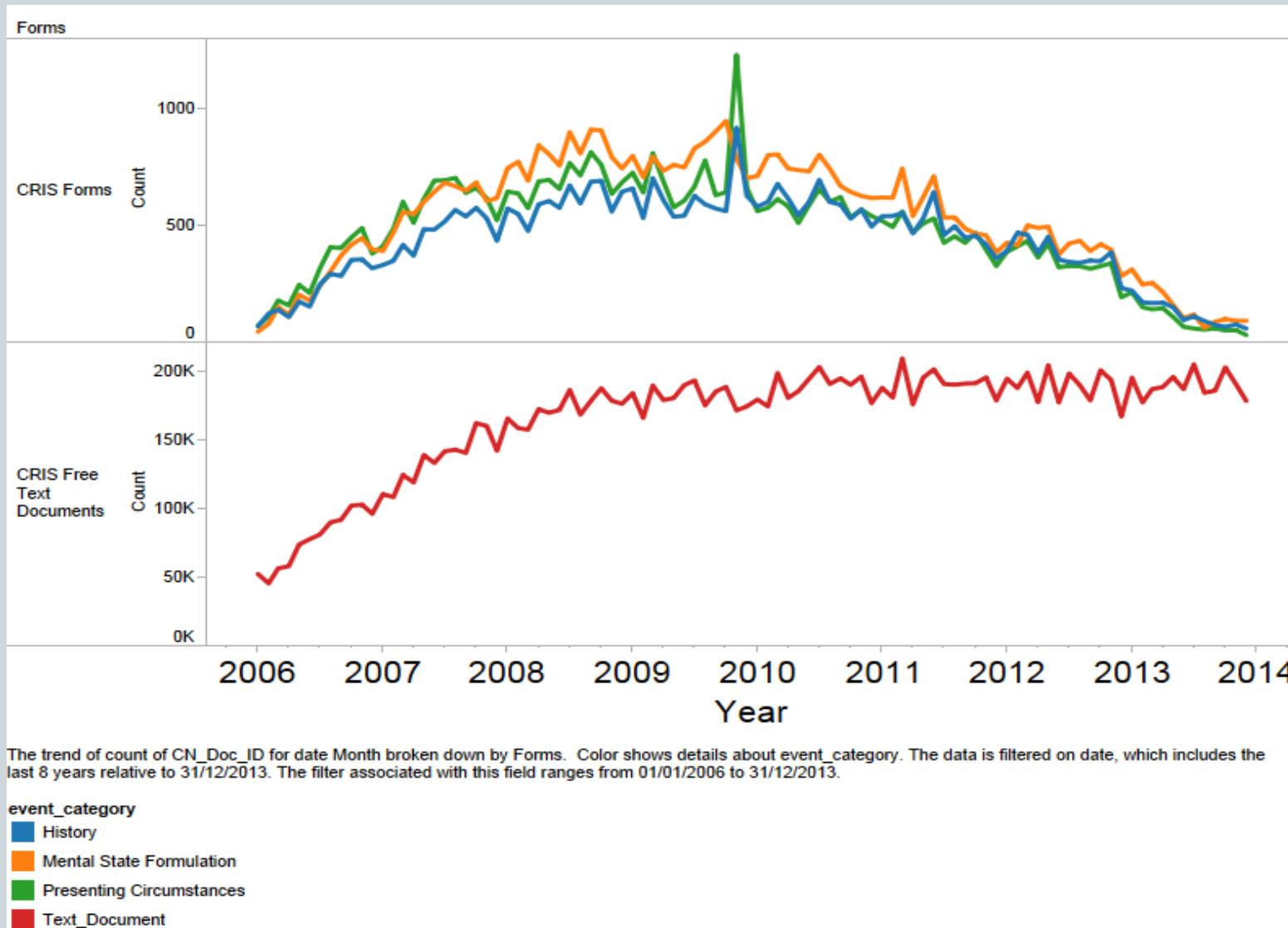
ONS (death certification)
 Local Primary Care (LDN)
 NHS Digital (HES)
 PHE (National Cancer Registry)
 Local Acute Care (maternity, neonatal)
 DfE/ONS (National Pupil Database)
 'Me and My School'
 DWP (benefits receipt)
 ONS (UK Census records)

- Technical elements (identifiers and matching)
- Procedural elements (data flow, documentation)
- Underlying governance (REC, HRA CAG – s251)
- Rationale (scientific question)
- Goodwill, trust and luck

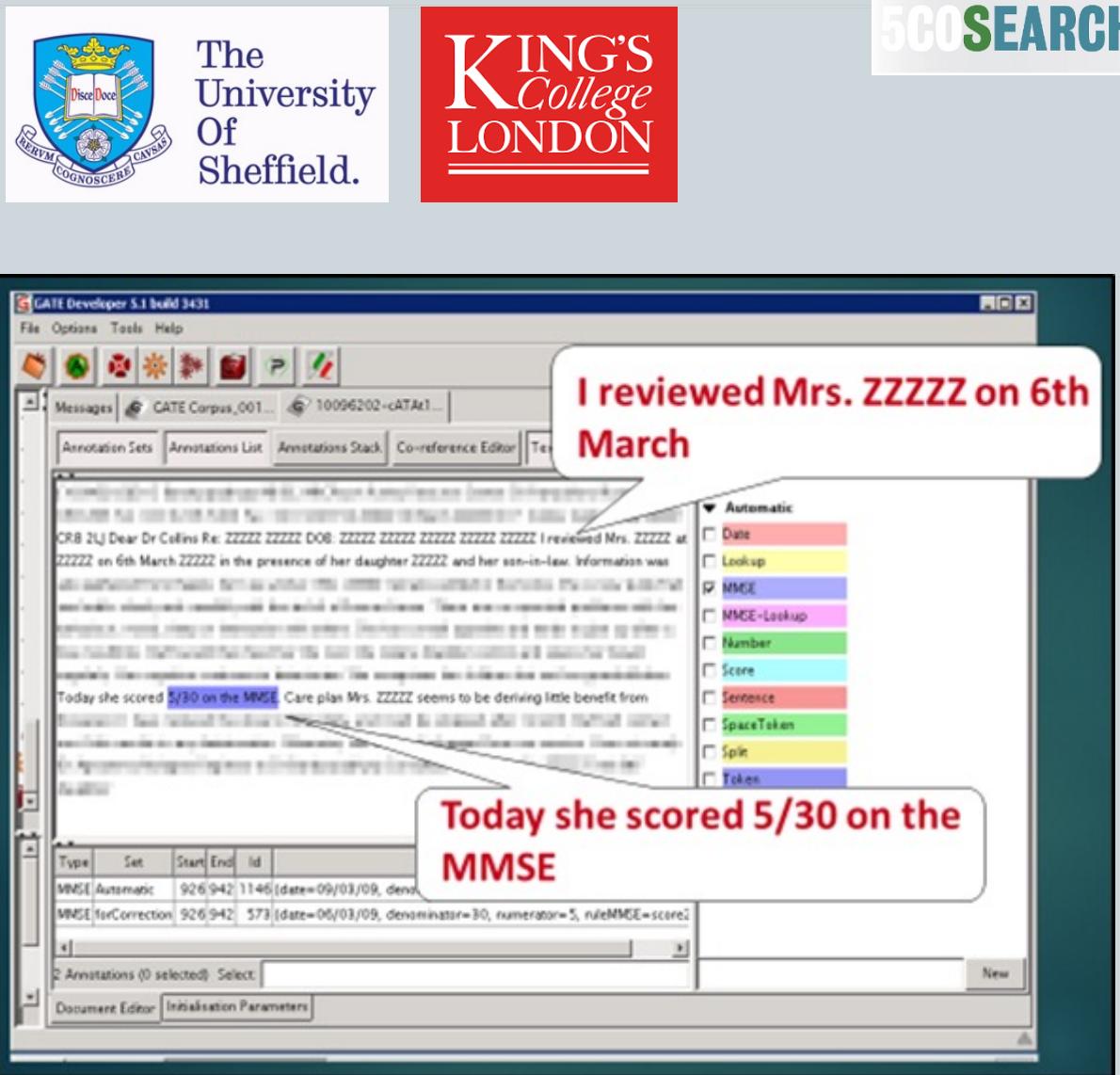
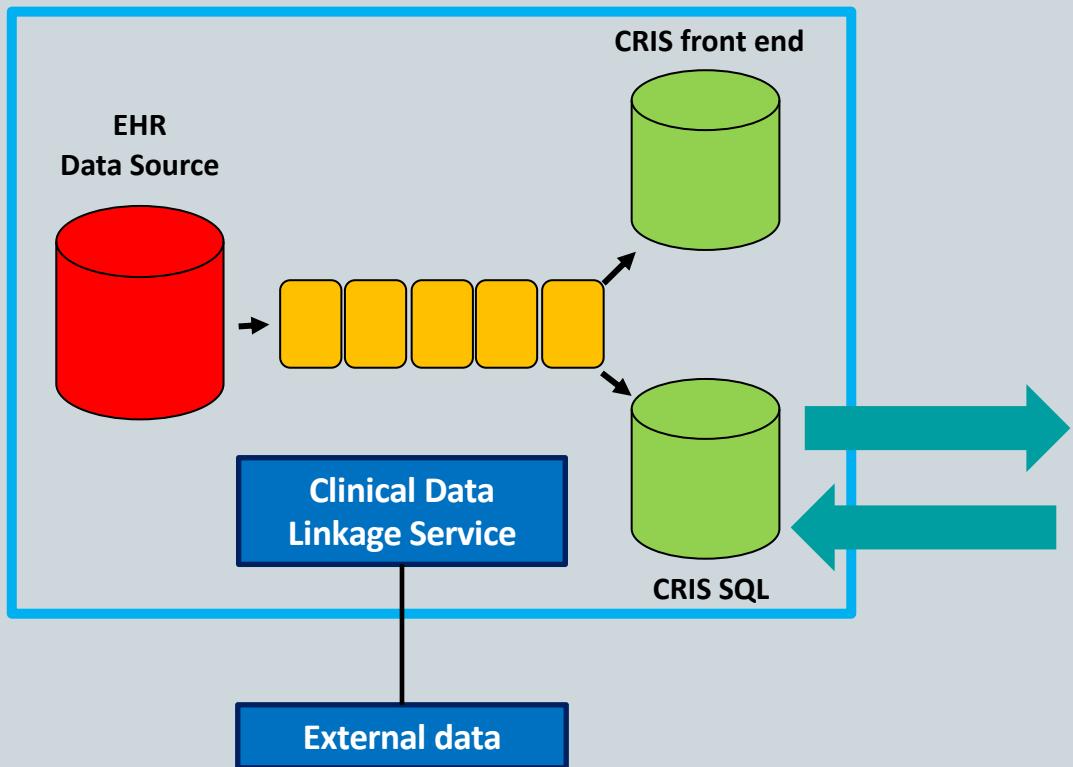
'Context' / spatio-temporal

Local environment (SELCoH)
 Social media (PHEME)
 Geospatial data (pollution)
 Temperature/weather

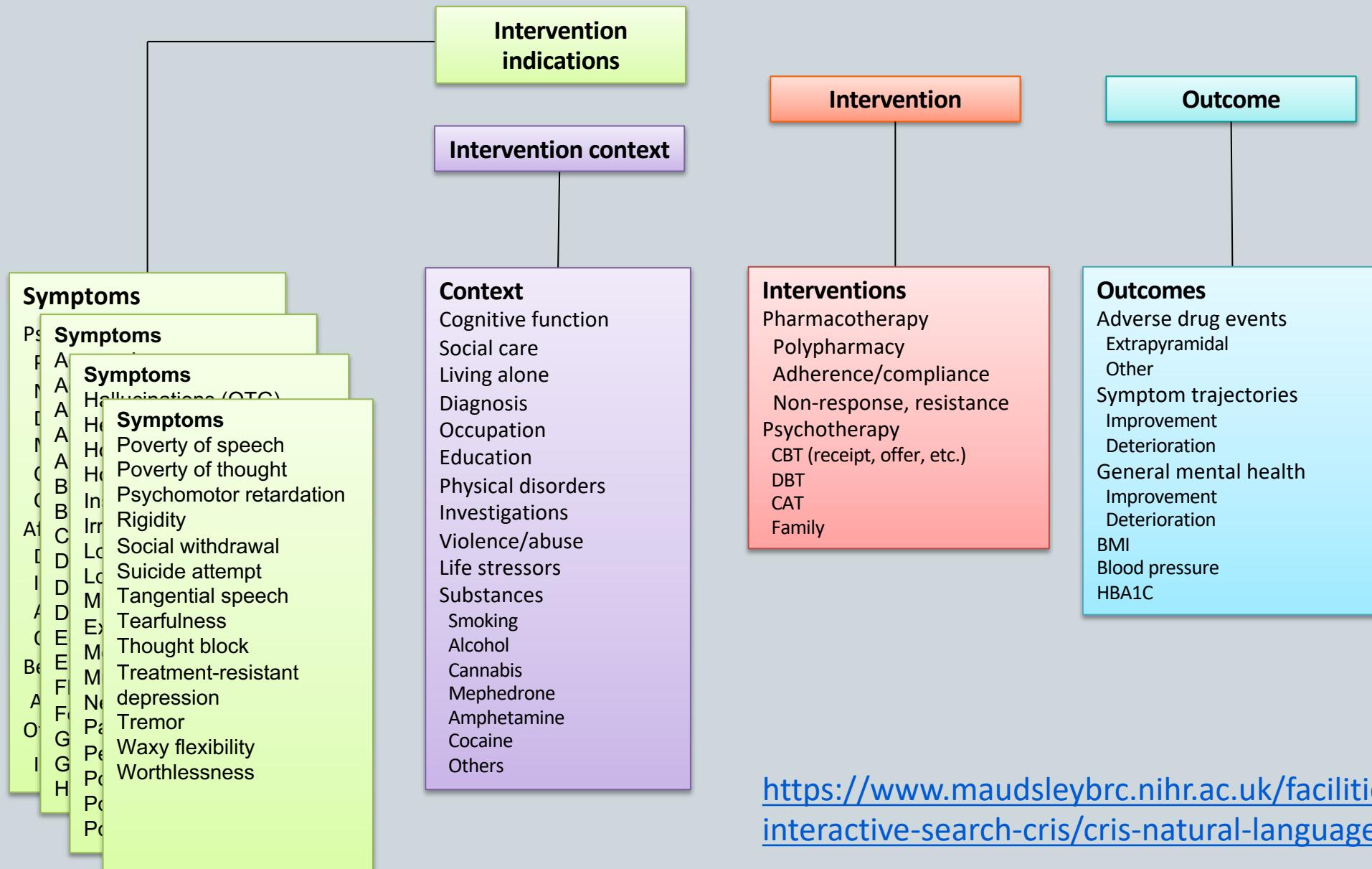
Unstructured data fields in CRIS – a solution?



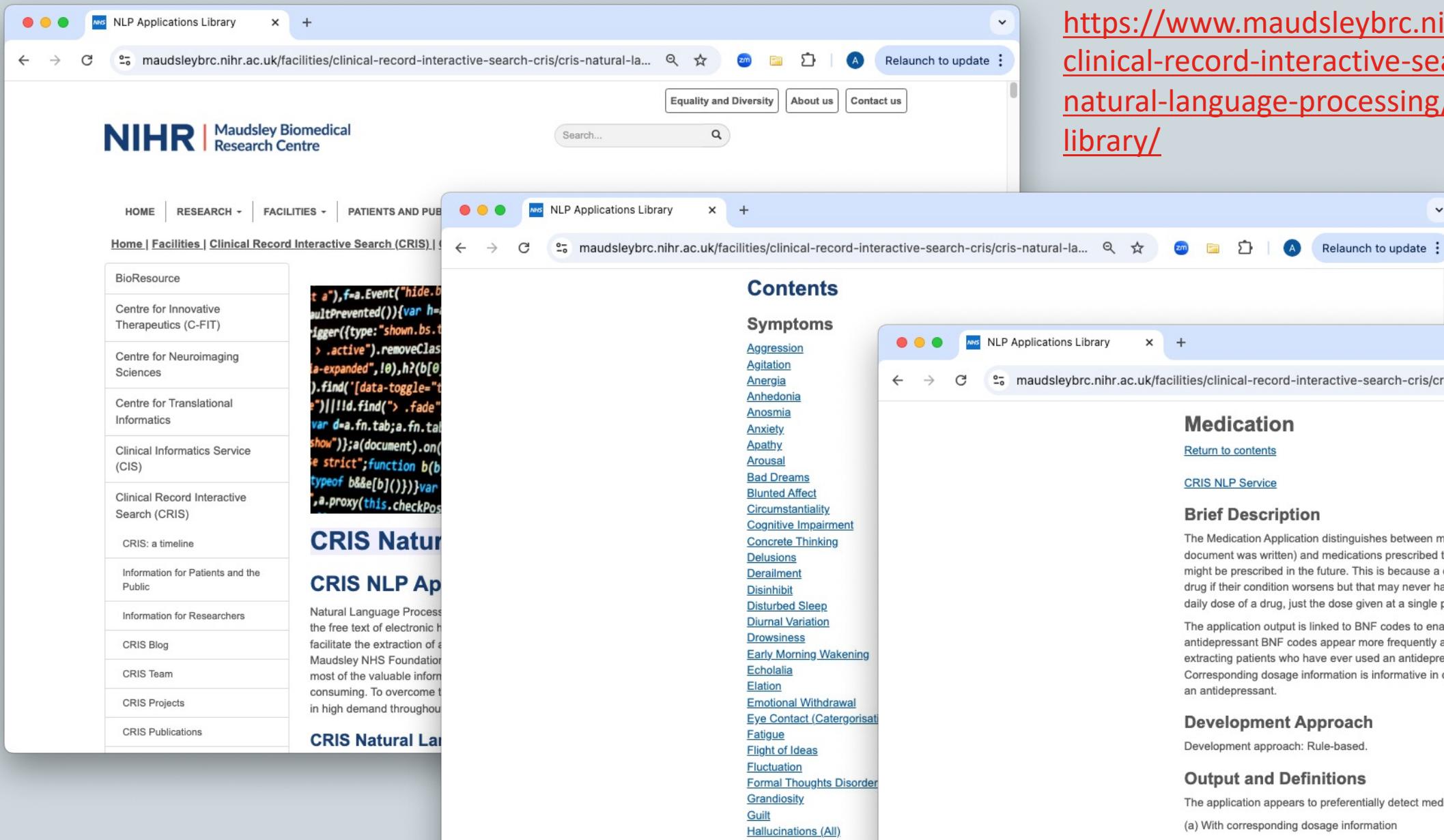
Data expansion 2 – text analytics



Cohort studies using EHRs – rich textual data



CRIS NLP Applications Catalogue



<https://www.maidsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/cris-natural-language-processing/nlp-applications-library/>

Self-service cloud-based analytics

The diagram illustrates the self-service cloud-based analytics process across three main stages:

- 1. Choose a service**: A screenshot of the "NHS Text Analytics" service catalog on the GATE Cloud platform. It shows various service cards, including "Smoking Status Annotator" and "Lives Alone". A yellow callout points to the top service card.
- 2. Test the service**: A screenshot of the "Smoking Status Annotator" service interface. It includes a "Test this pipeline" section with a text input field and file upload options, and a "Data Bundles" section showing two pre-defined bundles.
- 3. Run services on pre-defined data bundles**: A screenshot of the "Angus Roberts: Your Dashboard" page. It displays annotation jobs and data bundles. A yellow callout points to the "Data Bundles" section.

GATE CLOUD and **Azure** are shown in clouds at the top right, indicating the cloud-based nature of the services.

NHS Text Analytics

Smoking Status Annotator

Lives Alone

Education

BioYODIE Named Entity Disambiguation

Test this pipeline

Choose file No file chosen

Output JSON

Document plain text

Name Created At State

BioYODIE Named Entity Disambiguation (pipeline 12) 24 April 2017 13:43:55 BST reserved but not yet defined

Smoking Status Annotator (pipeline 6) 24 April 2017 13:57:04 BST reserved but not yet defined

Bundle ID **Name** **Created At** **Price per month**

D-000003 Small test 24 April 2017 13:43:03 BST £0.00

D-000004 J-000002 results (2017-04-24T13:48:42.701+01:00) 24 April 2017 13:48:42 BST £0.00

Upload your own data

Pipelines and services x Secure | https://slam-cloud.gate.ac.uk/shopfront

Angus Roberts's account | Admin | Log out

Home Services Dashboard

Angus Roberts's account | Admin | Log out

Angus Roberts's account | Admin | Log out

Angus Roberts: Your Dashboard

NHS Text Analytics

Annotation Jobs

Filter view...

Name Created At State

BioYODIE Named Entity Disambiguation (pipeline 12) 24 April 2017 13:43:55 BST reserved but not yet defined

Smoking Status Annotator (pipeline 6) 24 April 2017 13:57:04 BST reserved but not yet defined

Data Bundles

You have 2 data bundles totalling 113.3 kB. The approximate monthly cost of this data is £0.00 (help)

Bundle ID Name Created At Price per month

D-000003 Small test 24 April 2017 13:43:03 BST £0.00

D-000004 J-000002 results (2017-04-24T13:48:42.701+01:00) 24 April 2017 13:48:42 BST £0.00

Upload your own data

powered by GATE © South London and Maudsley NHS Foundation Trust, King's College

This website is licenced under the CC-BY-NC-SA

MedCAT - multimorbidity

Mapping multimorbidity in individuals with schizophrenia and bipolar disorders: evidence from the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register

Rebecca Bendayan ^{1 2}, Zeljko Kraljevic ³, Shaweena Shaari ², Jayati Das-Munshi ⁴, Leona Leipold ², Jaya Chaturvedi ³, Luwaiza Mirza ², Sarah Aldelemi ², Thomas Searle ³, Natalia Chance ², Aurelie Mascio ³, Naoko Skiada ³, Tao Wang ³, Angus Roberts ^{3 2}, Robert Stewart ^{2 4}, Daniel Bean ^{3 5}, Richard Dobson ^{3 2 6}

Affiliations + expand

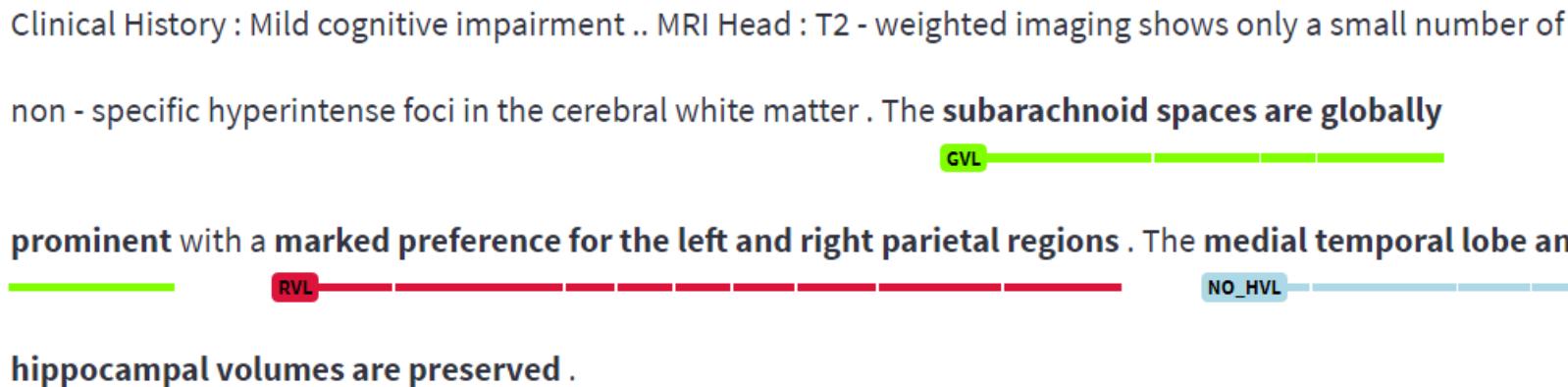
PMID: 35074819 PMCID: [PMC8788233](#) DOI: [10.1136/bmjopen-2021-054414](https://doi.org/10.1136/bmjopen-2021-054414)

Cerebrovascular accident
Epilepsy
Diabetes mellitus
Chronic kidney disease
Psoriasis
Parkinson's
Multiple sclerosis
Eczema
Hypertensive disorder
Transient ischaemic attacks
Migraine
Chronic obstructive lung disease
Arthritis
Heart failure
Asthma
Ischaemic heart disease
Irritable bowel disease
Atrial fibrillation
Chronic liver disease
Chronic sinusitis
Coronary arteriosclerosis

Extracting dementia related volumetric assessments from radiology reports

Adam Mayers

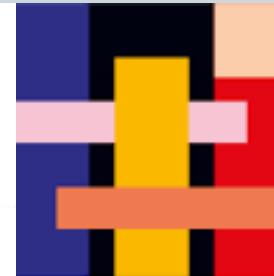
Clinical History : Mild cognitive impairment .. MRI Head : T2 - weighted imaging shows only a small number of non - specific hyperintense foci in the cerebral white matter . The **subarachnoid spaces are globally prominent** with a **marked preference for the left and right parietal regions** . The **medial temporal lobe and hippocampal volumes are preserved** .



	Category	Label
	Global Volume Loss – Present	GVL
0	Global Volume Loss – Absent	NO_GVL
1	Regional Volume Loss – Present	RVL
2	Regional Volume Loss – Absent	NO_RVL
	Hippocampal/Medial Temporal Lobe Volume Loss – Present	HVL
	Hippocampal/Medial Temporal Lobe Volume Loss – Absent	NO_HVL

Extracting mentions of interpersonal violence

- People with mental illness are more likely to experience violent victimisation compared to the general population- 15–45% of female patients report experiences of victimization in the past year, and 40–90% reporting lifetime victimization.
- Similar patterns have been observed for domestic violence, sexual violence, violence perpetration, and witnessing violence.
- Potential for use in research of electronic health records (EHRs) kept by mental health services



Extracting mentions of interpersonal violence

Annotation label	Training set (average score on 10-fold cross-validation)			Blind test set
	Precision	Recall	F1-score	
Violence presence	93%	93%	93%	95%
Patient status: Perpetrator	89%	89%	89%	85%
Patient status: Victim	91%	89%	91%	90%
Violence type: Domestic	94%	94%	94%	93%
Violence type: Physical	91%	92%	91%	98%
Violence type: Sexual	98%	97%	97%	93%

> BMJ Open. 2022 Feb 16;12(2):e052911. doi: 10.1136/bmjopen-2021-052911.

Can natural language processing models extract and classify instances of interpersonal violence in mental healthcare electronic records: an applied evaluative study

Riley Botelle ¹, Vishal Bhavsar ², Giuliana Kadra-Scalzo ³, Aurelie Mascio ³,
Marcus V Williams ⁴, Angus Roberts ^{5 6}, Sumithra Velupillai ³, Robert Stewart ^{3 7}

Affiliations + expand

PMID: 35172999 PMCID: PMC8852656 DOI: 10.1136/bmjopen-2021-052911

VIEWER – bringing NLP to the clinic

Credit: Tao Wang, Yamiko Msosa, David Codling



Not secure | cogstack.slam.nhs.uk/app/dashboards#/view/4f942d20-8269-11eb-91bd-6142eb1fffa507_g=filters[0].refreshInterval(pause:0,value:0).time:(fromNow-1M,toNow)]&_a=(description:This%20is%20the%20generic...)

Elastic

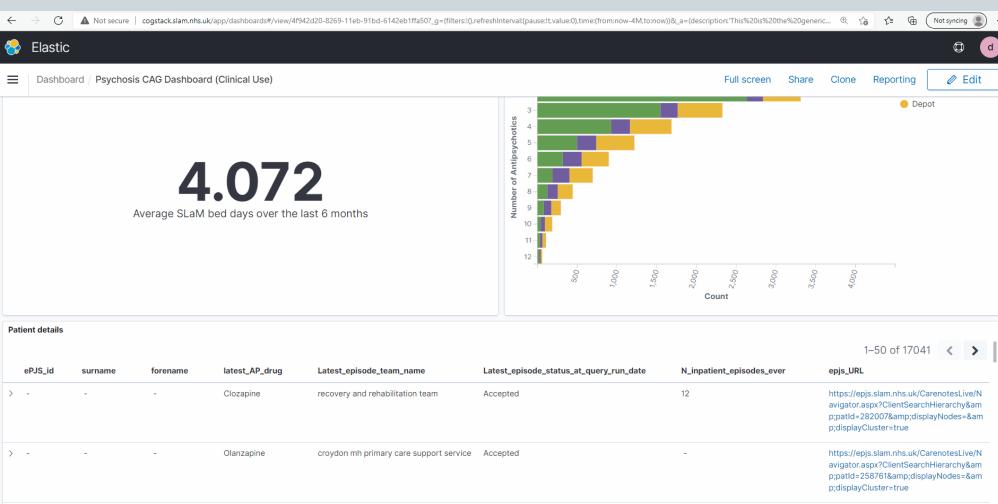
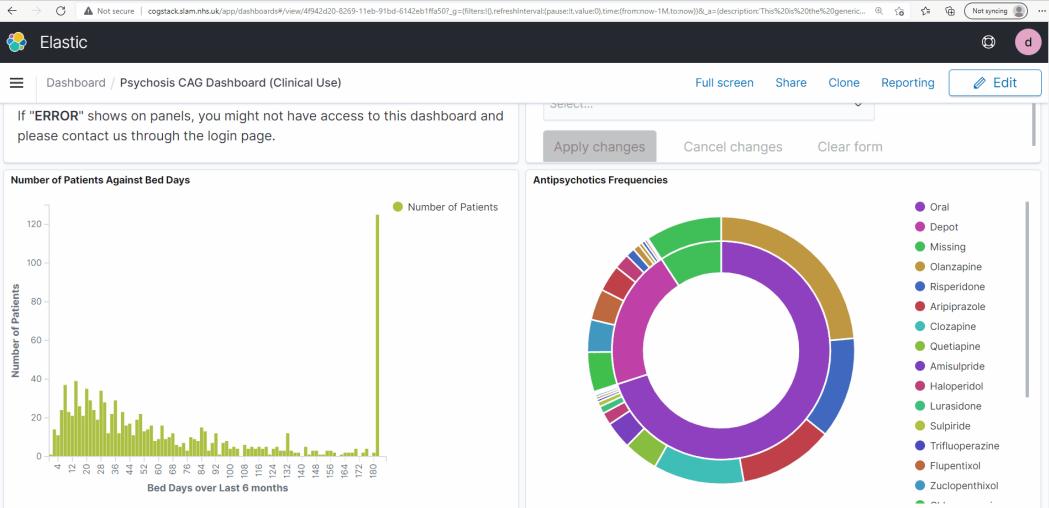
Dashboard / Psychosis CAG Dashboard (Clinical Use)

Aripiprazole psychosis primary care mh service Discharged 3 http://am.r/reno/avix?CI

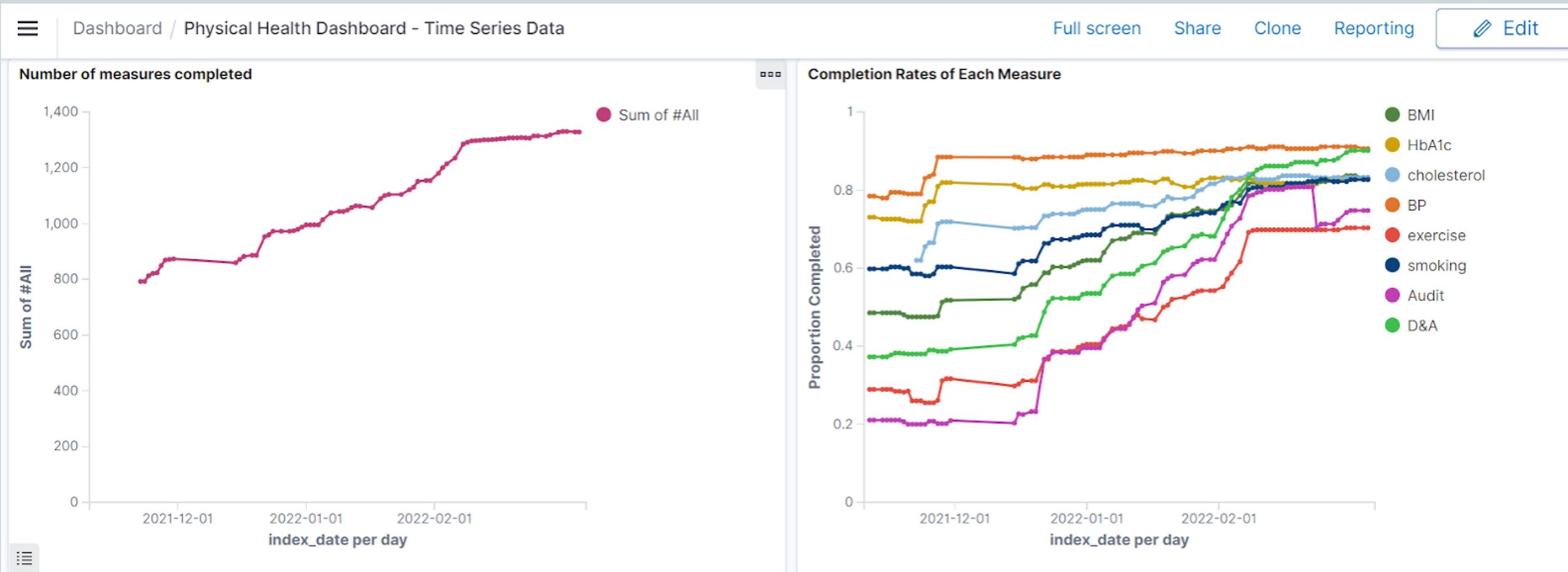
Average Community Events 9.13 Average community events over the last 6 months

Medications Mentioned 1-50 of 51250

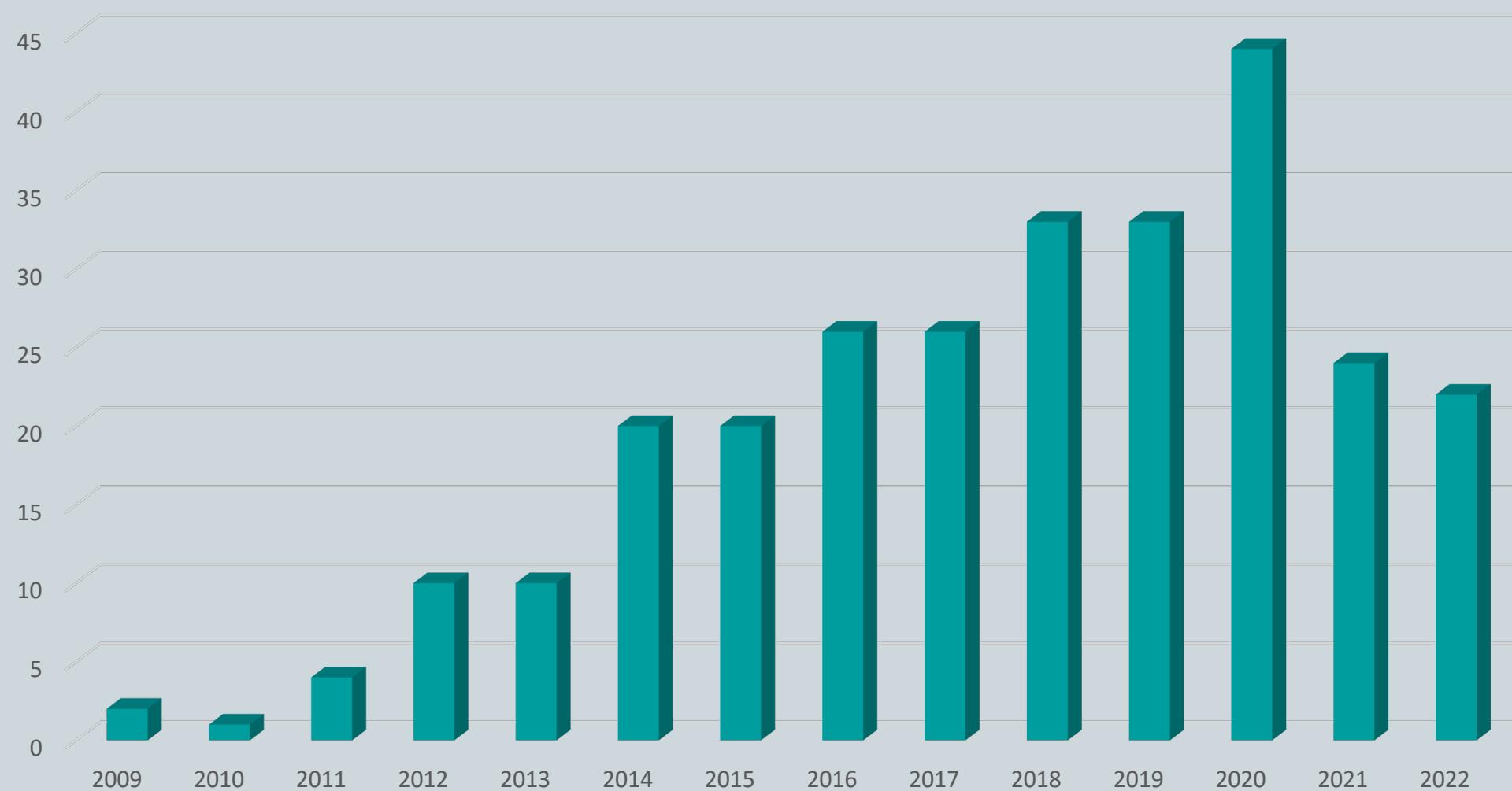
ePJS_id	medication	number_of_prescriptions	date_of_latest_prescription
-	Aripiprazole	20	Mar 8, 2022 @ 00:00:00
-	Flupentixol	46	Mar 8, 2022 @ 00:00:00
-	Aripiprazole	37	Mar 7, 2022 @ 00:00:00
-	Flupentixol	51	Mar 3, 2022 @ 00:00:00
-	Olanzapine	53	Feb 27, 2022 @ 19:48:57.87
-	Aripiprazole	40	Feb 27, 2022 @ 19:31:12.78
-	Aripiprazole	12	Feb 27, 2022 @ 15:57:17.22
-	Aripiprazole	202	Feb 27, 2022 @ 10:44:15.80



e.g. Impact on care: improved physical health monitoring



CRIS papers per year (estimated)



CRIS Acknowledgements - a multi-disciplinary team

Leads

Matthew Broadbent, Richard Dobson, Stuart MacLellan, Rob Stewart

Administration / management

Debbie Cummings, Amelia Jewell, Daisy Kornblum

Technical team

Lara Brooks-Simpson (data linkage), Pampa Chakravarti (clinical informatics service), Risha Govind (clinical informatics service), Shanmukha Gudiseva, Satyendra Katneni (database administration), Hitesh Shetty (data extraction), Jyoti Sanyal (NLP support)

Epidemiology / Clinical

David Codling, Johnny Downs, Rina Dutta, Sophie Epstein, Richard Hayes, Giouliana Kadra, Ruimin Ma, Christoph Mueller, Rashmi Patel, Gayan Perera, Kate Polling, Eugenia Romano, Brendon Stubbs

Bioinformatics / Computer Science / NLP

Craig Colling, David Chandran, Zina Ibrahim, Zeljko Kraljevic, Lifang Li, Yamiko Msosa, Angus Roberts, Tao Wang, Lukasz Zalewski

PhD studentships

Norah Alothman, Nikeysha Bell, Delia Bishara, Jaya Chaturvedi, Natasha Chilman, Katrina Davies, Emel Yorganci

Oversight and Governance

Christina Armoogum, Eva Broeckelmann, Claire Delaney-Pope, Patrick Green

SLAM partners

Barbara Arroyo, Fiona Gaughran, Justin Sauer, David Taylor

University of Sheffield (Computer Science) collaborators

Kalina Bontcheva, Ian Roberts

Thank you

angus.roberts@kcl.ac.uk

<https://www.kcl.ac.uk/people/angus-roberts>