NHS National Institute for Health Research

Sharing NLP and Compute Services: Cloud Based NLP at the Maudsley Biomedical Research Centre

Jyoti Sanyal, Matthew Broadbent, Angus Roberts, Anna Kolliakou, Ian Roberts, Robert Stewart

NHS
National Institute for
Health Research

Collaborators

About the SLAM Biomedical Research Centre (BRC)

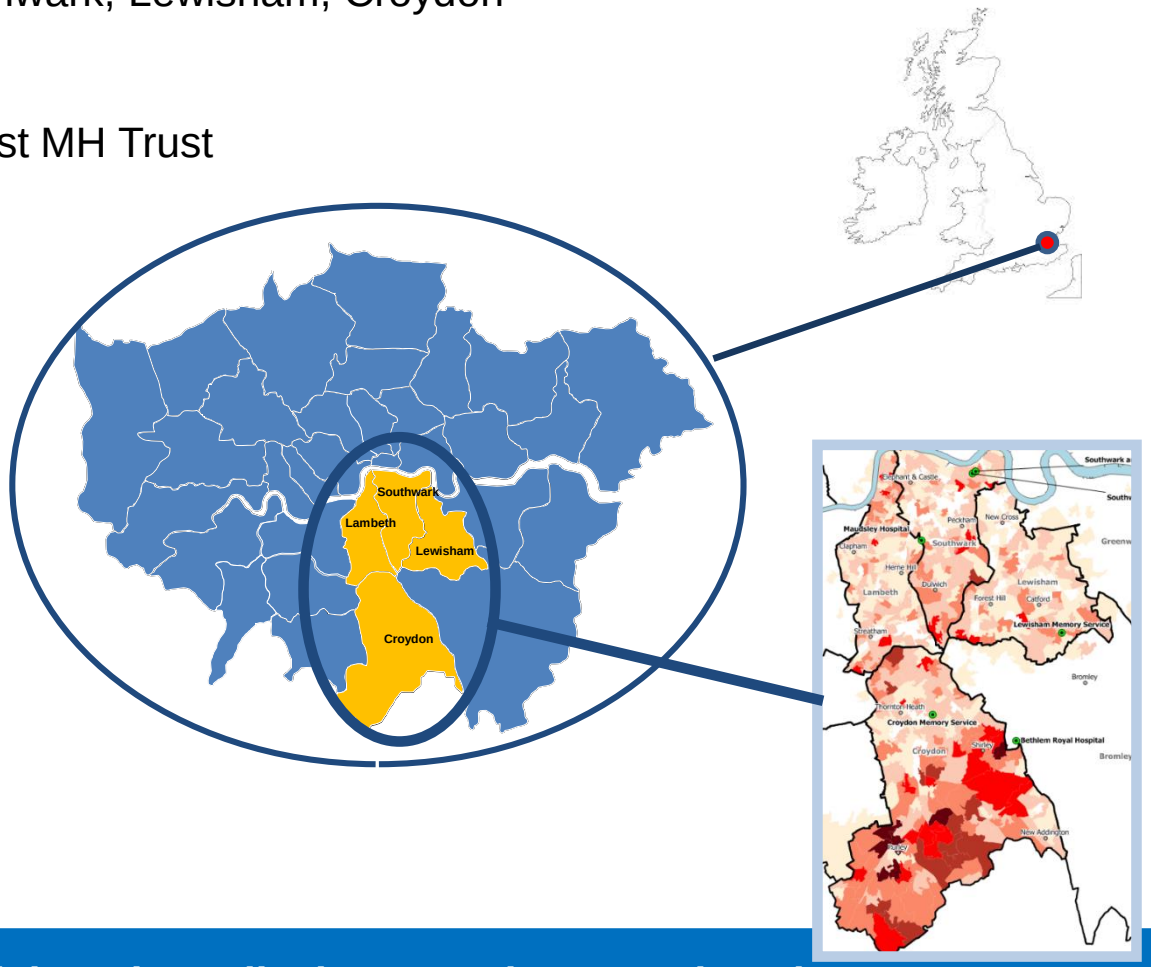
NHS National Institute for Health Research

Coverage – Lambeth, Southwark, Lewisham, Croydon

Base population – c.1.2m

Clinical services – specialist MH Trust

- CAMHS
- General adult psychiatry
- Older adult services
- Learning difficulties
- Addictions
- National
- IAPT
- Forensic

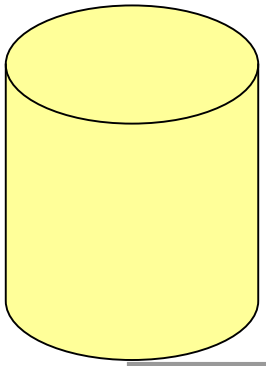


South London and Maudsley (SLAM)

EHR, search and NLP at the Maudsley BRC

NHS National Institute for Health Research

The Patient Journey System (ePJS)



Coverage: Four London boroughs

Local population: c. 1.1 million

Clinical area: specialist mental health

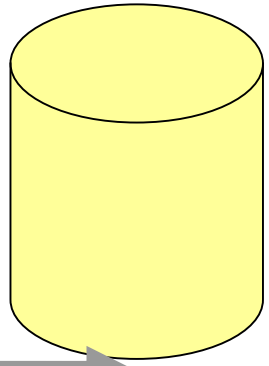
Active patients: c. 35 000

Total inpatients: c. 1 000

Total records: c. 392 000 people

Documents: 35 million

- **CRIS**
- **Interactive search**
- **FAST index**
- **SQL RDBMS**
- **GATE based NLP**



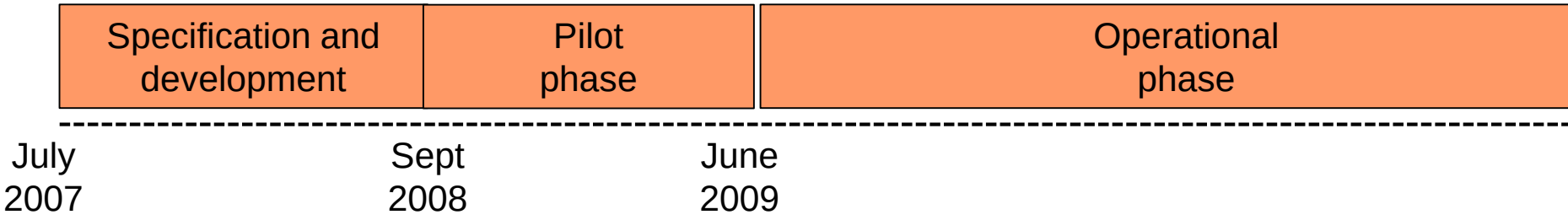
CRIS Project Summary

NHS National Institute for Health Research

Aims:

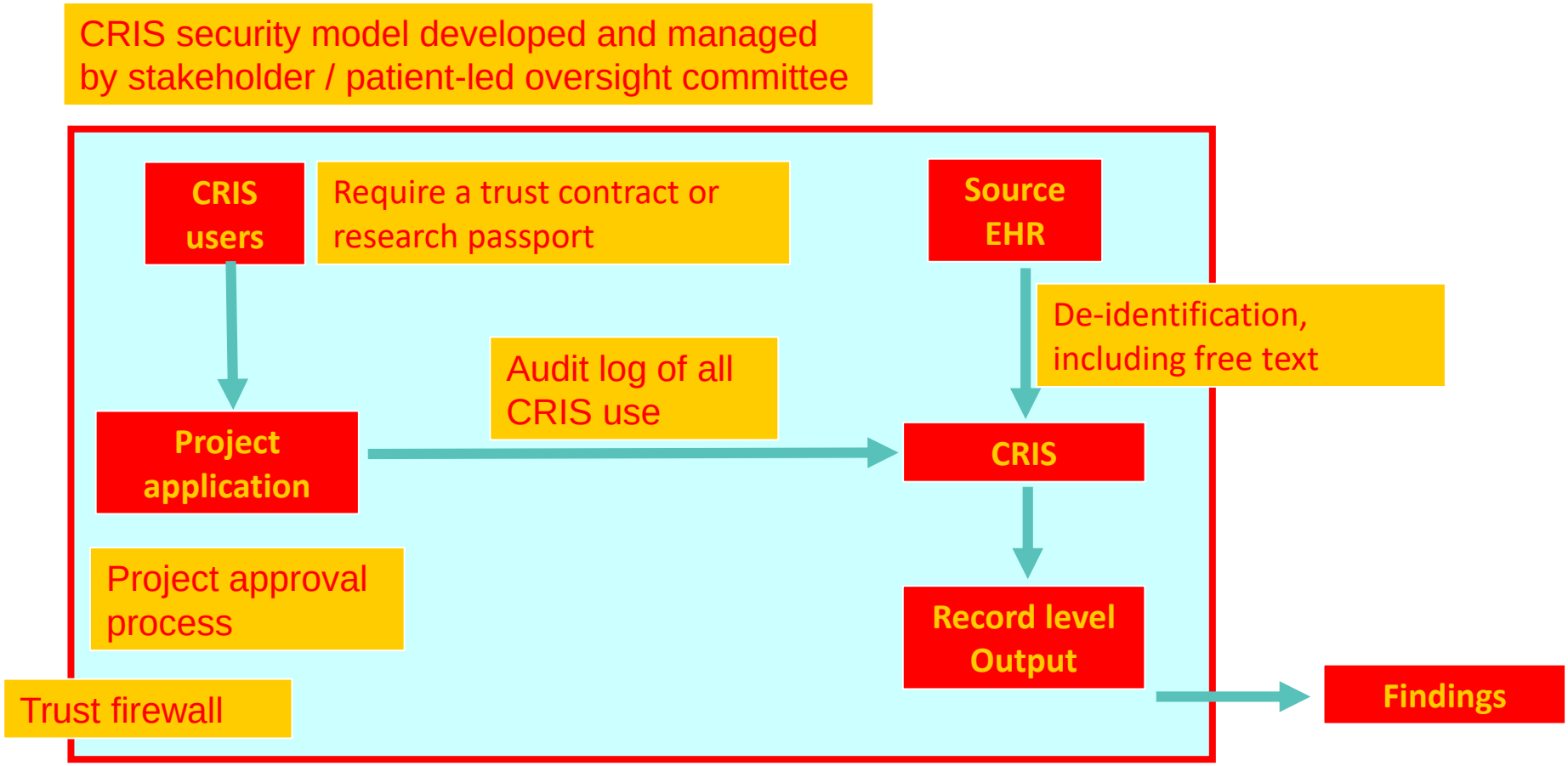
To provide searchable access to clinical records from the South London and Maudsley NHS Foundation Trust (SLAM) for research purposes

To ensure the legal and ethical rights of patients are protected



CRIS security model: service user led governance

NHS National Institute for Health Research



NLP at the Maudsley BRC

NHS National Institute for Health Research

- Largely GATE based NLP applications
  - Core of rule based applications
  - Large number (80+) of supervised ML applications, iteratively generated using TextHunter
- Python based NLP applications using deep learning and other approaches.
- Research output: over 100 research papers partly based on NLP generated data
- Use in clinical care at an early stage
  - Used to monitor physical health interventions

CRIS NLP applications

NHS National Institute for Health Research

Intervention indications

Symptoms

Psychosis
- Positive ☑
- Negative ☑
- Disorganisation ☑
- Manic ☑
- Catatonic ☑

Affective
- Mood ☑
- Biological symptoms ☑
- Instability ☑
- Anxiety ☑
- Obsessive/compulsive ☑
- Somatic ✍

Behaviour
- Agitation/withdrawal ☑

Other
- Insight ☑

Intervention context

Context

- Cognitive function ☑
- Social care ☑
- Living alone ☑
- Diagnosis ☑
- Social status ✍
- Education ☑
- Physical disorders ☑
- Investigations ✍
- Substances
  - Smoking ☑
  - Alcohol ✍
  - Cannabis ☑
  - Mephedrone ☑
  - Amphetamine ☑
  - Cocaine ☑
  - Others ✍

Intervention

Interventions

Pharmacotherapy ☑
- Polypharmacy ☑
- Adherence/compliance ☑

Psychotherapy
- CBT ☑ (receipt, offer, etc.)
- DBT ✍
- CAT ✍
- Family ✍
- Supportive/behavioural ✍

Outcome

Outcomes

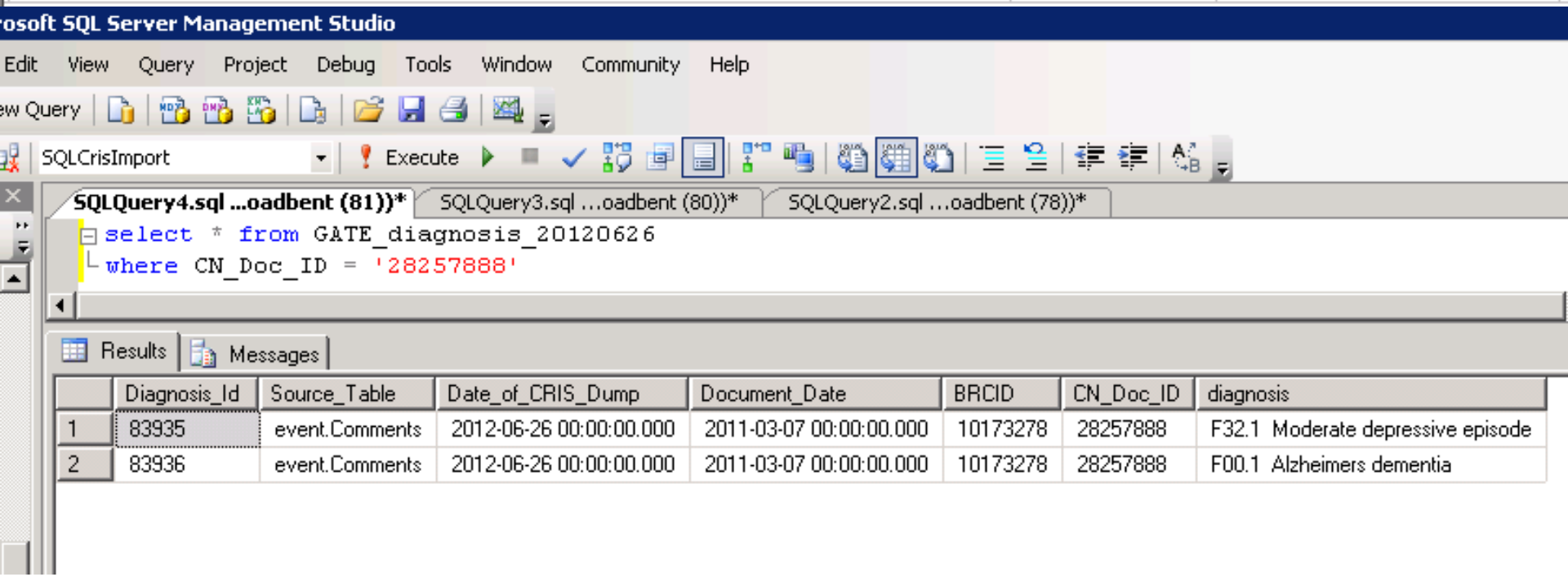
Adverse drug events
- Extrapyramidal ☑
- Other ✍

Symptom trajectories ✍
- Improvement
- Deterioration

General mental health ✍
- Improvement
- Deterioration

☑ = complete
✍ = in progress


Events
Assmnts
Event
Diagnosis
Notes
Document ID
Primary Diag (Axis 1a)
eEVCo33
eEVCN51
aDIPr05
28257888
F32.1 - Moderate depressive episode
[10173278] 22222 22222 D O B 22222 Age 87 Female Aspen (David South Diagnosis F32.1 Moderate depressive episode F00.1 Alzheimers dementia Action from ward round 1 Increase Mirtazapine to 30mg . 2 MMSE when possible (was 19/30) 3 Request investigations to exclude NG as cause of decreased appetite. 4 Awaiting Dietician's opinion. 5 Monitor for pain relief. Dr Amanda Thompsell Consultant in Liaison Older Age Psychiatry ----------09 Mar 2011 11:15, Julie Errington


rosoft SQL Server Management Studio
Edit View Query Project Debug Tools Window Community Help
ew Query
SQLCrisImport
Execute
SQLQuery4.sql ...oadbent (81))*
SQLQuery3.sql ...oadbent (80))*
SQLQuery2.sql ...oadbent (78))*
select * from GATE_diagnosis_20120626
where CN_Doc_ID = '28257888'
Results
Messages
Diagnosis_Id Source_Table Date_of_CRIS_Dump Document_Date BRCID CN_Doc_ID diagnosis
1 83935 event.Comments 2012-06-26 00:00:00.000 2011-03-07 00:00:00.000 10173278 28257888 F32.1 Moderate depressive episode
2 83936 event.Comments 2012-06-26 00:00:00.000 2011-03-07 00:00:00.000 10173278 28257888 F00.1 Alzheimers dementia

Events		Assmnts
Event		Diagnosis
Notes	Document ID	Primary Diag (Axis 1a)
eEVCo33	eEVCN51	aDIPr05
[10173278] ZZZZZ ZZZZZ D O B ZZZZZ Age 87 Female Aspen (David South) Diagnosis F32.1 Moderate depressive episode F00.1 Alzheimers dementia Action from ward round 1 Increase Mirtazapine to 30mg . 2 MMSE when possible (was 19/30) 3 Request investigations to exclude NG as cause of decreased appetite. 4 Awaiting Dietician's opinion. 5 Monitor for pain relief. Dr Amanda Thompsell Consultant in Liaison Older Age Psychiatry ----------09 Mar 2011 11:15, Julie Errington	28257888	F32.1 - Moderate depressive episode

```
      ,[When]
      ,[Route]
  FROM [SQLCrisImport].[dbo].[GATE_medication_20120626]
where CN_Doc_ID = '28257888'
```

Results | Messages

document_date	brcid	CN_Doc_ID	Drug	Drug_Type	Status	Tense	Dose	Dose_Value	Dose_u
2011-03-07 00:00:00.000	10173278	28257888	Mirtazapine	Antidepressants	continuing	present	30mg	30.0	mg

Events		Assmnts
Event		Diagnosis
Notes	Document ID	Primary Diag (Axis 1a)
eEVCo33	eEVCN51	aDIPr05
[10173278] ZZZZZ ZZZZZ D O B ZZZZZ Age 87 Female Aspen (David South) Diagnosis F32.1 Moderate depressive episode F00.1 Alzheimers dementia Action from ward round 1 Increase Mirtazapine to 30mg . 2 MMSE when possible (was 19/30) 3 Request investigations to exclude NG as cause of decreased appetite. 4 Awaiting Dietician's opinion. 5 Monitor for pain relief. Dr Amanda Thompsell Consultant in Liaison Older Age Psychiatry ----------09 Mar 2011 11:15, Julie Errington	28257888	F32.1 - Moderate depressive episode

ent (81))* | SQLQuery3.sql ...oadbent (80))* | SQLQuery2.sql ...oadbent (78))*

```
GATE_mmse_cleaned_20120626
ID = '28257888'
```

es

_table	Date_of_CRIS_Dump	BRCID	CN_Doc_ID	Document_Date	Numerator	Denominator	MM
Comments	2012-06-26 00:00:00.000	10173278	28257888	2011-03-07 00:00:00.000	19	30	201

Why NLP is important for SLAM ?

- **MMSE score and date**

	MMSE Structured	MMSE from NLP
N instances	10467	121283
N patients (at least 1 instance)	6944	27486
Precision		97%
Recall		98%

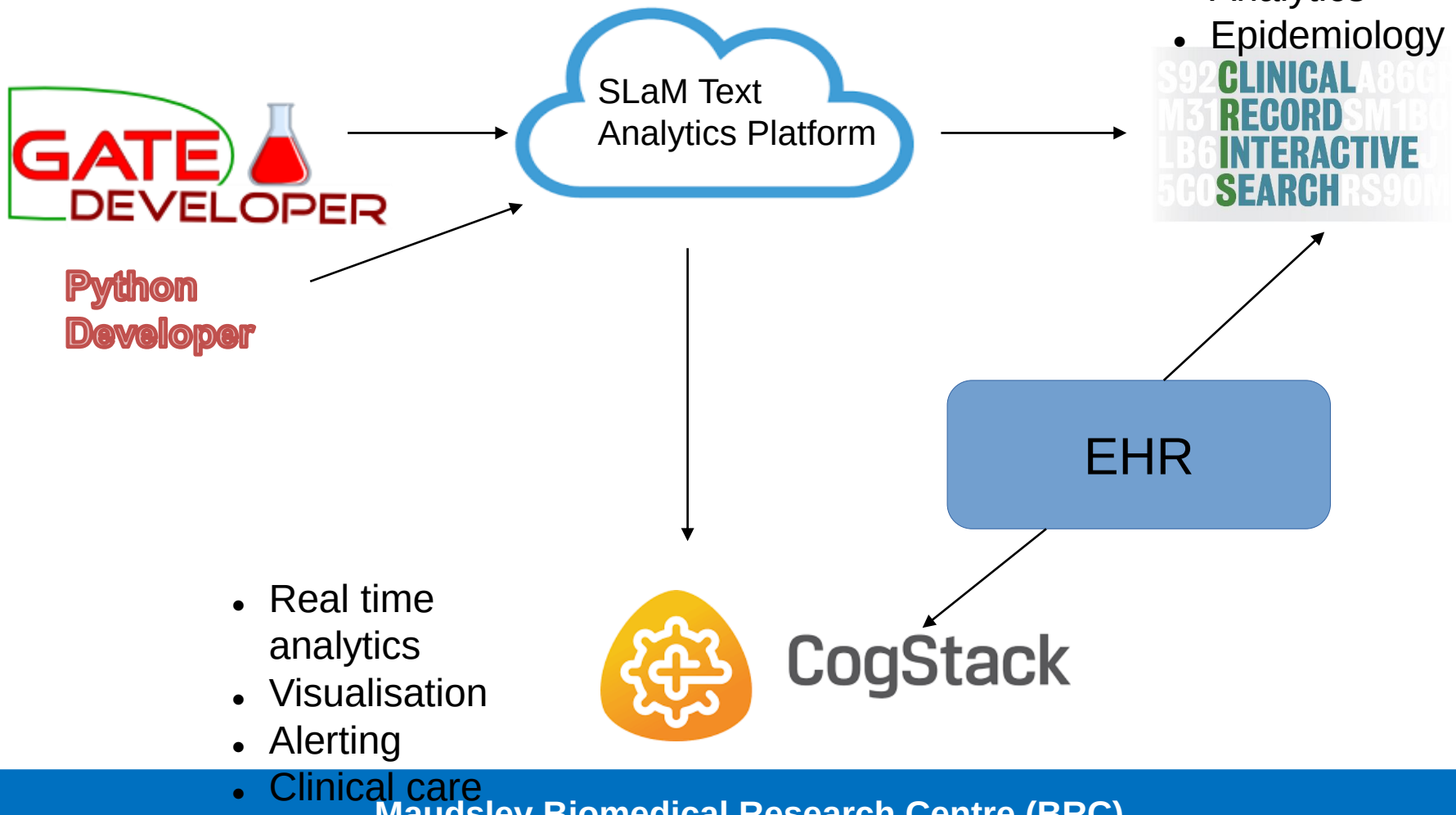
The BRC's need for cloud-based NLP and compute services

NHS National Institute for Health Research

- Scaling
  - Scaling beyond desktop servers
  - Implementation of Azure infrastructure at SLaM
- Self service NLP
  - Streamline access to NLP, textual data, compute resources
  - Simplify routine running of applications by analysts
  - End user access to individual applications
- Collaboration
  - Sharing NLP services and compute resources with collaborators
  - Delivering NLP to other trusts - anyone interested?

The Maudsley text analytics ecosystem

NHS National Institute for Health Research



The Maudsley BRC Text Analytics platform

NHS National Institute for Health Research

- Based on GATE Cloud – a public Amazon backed NLP cloud service run by University of Sheffield GATE Team
- Adapted to Azure for SLaM by the GATE Team
- This demonstration
  - is not running on the SLaM Azure tenancy
  - contains a partial set of applications
  - does not use SLaM data
    - MTSamples
    - MIMIC III

Demonstration

https://nhsta.slam-services.gate.ac.uk/

Data - MIMIC III

- https://mimic.physionet.org/
- Deidentified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012
- Available for research
- 2 083 180 free text event notes
- Encrypted in our database
  - We cannot view content
  - We can show counts of tables and limited extracted structured data

Data - MTSamples

- http://www.mtsamples.com/
- 2 377 medical transcription sample reports and examples
- Publicly available
- American English
- Some are low quality
- As this is publicly available, we can show you the content

Demo hardware

NHS National Institute for Health Research

- Running in Azure West Europe region
- Database: 100 DTU SQL Server
- Front end: B2ms burstable (up to 2 cores and 8GB RAM, 7.3p/hr)
- Processing
  - Configurable and expandable, today we are running one of:
  - 6 x dynamically provisioned DS2v2 (2 cores, 7GB RAM, 10.2p/hr)
  - 1 x DS2v2 “always on”, plus 5 x dynamically provisioned D4v3 (4 cores, 16GB RAM, 17.9p/hr)

Processing software

- Nodes run open source Gate Cloud Paralleliser - GCP
  - Database rows split in to batches
  - Multi-threaded processing of each batch by nodes
  - Join
- Each batch is around 15000 database rows
- If several jobs are running, processing nodes will round robin between the jobs, rather than jobs monopolising nodes