Delivering value with digital technologies
Briefing: March 2022

Using artificial intelligence to unlock health records

Case study

Supported by

NHS
Health Education England

Delivering value with digital technologies

Digital technologies such as digital medicine, genomics, artificial intelligence and robotics have a huge potential to transform the delivery of healthcare.[1]

These technologies can empower patients to participate actively in their care, with a greater focus on wellbeing and prevention. They also support the prediction of individual disease risk and personalise the management of long-term conditions.

The HFMA, supported by Health Education England, is delivering a 12-month programme of work to increase awareness amongst NHS finance staff about digital healthcare technologies, and enable finance to take an active role in supporting the use of digital technology to transform services and drive value and efficiency.[2]

As part of the programme, the HFMA is publishing a series of case studies. Working with organisations who have started on the digital transformation journey, we will identify examples of good practice and highlight the challenges that services face. This will include specific challenges relating to NHS finance.

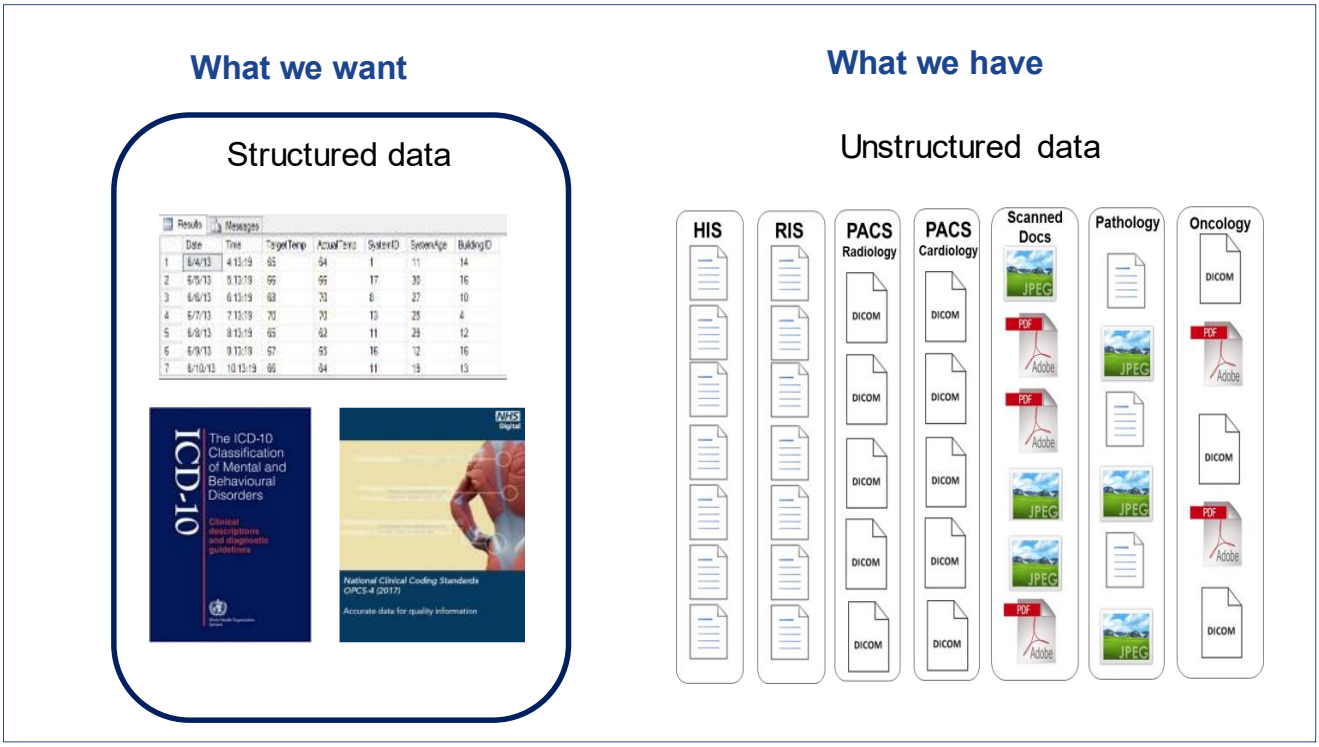
This case study describes how research teams from two biomedical research centres in London have used natural language processing - a form of artificial intelligence - to structure the large volumes of unstructured data in electronic patient records, so that the data can be used to support the delivery of high-quality care and clinical research.

[1] HFMA *Introduction to digital healthcare technologies*, July 2021
[2] HFMA Delivering value with digital technologies

The challenge of unstructured data

Data held in electronic patient records provides a wealth of information to support the delivery of high-quality care and clinical research. However, records are often difficult to access and about 80% of the data is unstructured, for example narrative text, pictures, emails, instant messages etc (**figure 1**).

Figure 1: Most data in the NHS is unstructured



Existing healthcare software systems are not designed to query unstructured data, which significantly restricts the use of the data. The result is that the wealth of information potentially available within health records is often inaccessible and underused.

In acute services, clinical coders read individual health records and assign standardised codes to the records based on particular words, conditions or treatments. In other sectors such as mental health, community physical health and primary care, health data is often coded by the clinician which transfers the administrative burden to patient-facing staff.

Clinical coding is under significant pressure due to the expanding volume of data, and capturing all the information from unstructured data can be challenging.

Using digital technology to structure NHS data

The problem of unstructured data is not just one that the NHS faces, and other industries, for example financial institutions, the legal system and journalism, have already started using natural language processing (NLP) to address the problem.

Natural language processing

Natural language processing (NLP) refers to the branch of computer science – and more specifically, the branch of artificial intelligence (AI) – concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics – rule-based modelling of human language – with statistical, machine learning[3], and deep learning[4] models. Together, these technologies enable computers to process human language in the form of text or voice data and to assign 'meaning', and thereby work towards 'understanding' the speaker or writer's intent and sentiment.

Everyday uses of NLP include chatbots on commercial websites, spam email filters and smart assistants like Apple's Siri and Amazon's Alexa.

A number of open source software[5] tools have been developed for NLP, but they need tailoring for healthcare as medical records are filled with medical terms, jargon, acronyms and local dialects.

CogStack

Researchers at the National Institute for Health Research Maudsley Biomedical Research Centre (BRC) and King's College Hospital, in partnership with the University College London Hospitals NHS Foundation Trust BRC, have started to tackle the challenge of unstructured data in the NHS with the development of CogStack[6].

CogStack uses clinical NLP to extract information from unstructured data sources in electronic patient records, where the majority of the information is in multiple formats of unstructured data (word documents, PDFs, images, text fields etc).

This innovative technology uses artificial intelligence to reveal important data locked in patients' health records to support clinical decision making and healthcare research. By providing automated clinical coding that is fast, efficient and accurate, the aim of CogStack is to provide better quality information and allow the NHS to release its skilled staff for other activities.

> *'When you visit your doctor or attend hospital a lot of information is collected about you on computers including your symptoms, tests, investigations, diagnosis, and treatments. Across the NHS this represents a huge amount of information that could be used to help us learn how to tailor treatments more accurately for individual patients and to offer them better and safer healthcare. The challenge we face is that most of the information held within these records is in written form which is difficult to use and learn from. We have developed the CogStack AI tools to read and understand this information.'*
>
> Professor Richard Dobson, group lead for CogStack and head of the Department of Biostatistics and Health Informatics at NIHR Maudsley BRC

[3] Machine learning is a subset of AI that focuses on learning, reasoning, and decision-making. The technologies use statistical models to make predictions (or decisions) without being explicitly programmed to perform the task. The computer 'learns' as it increases its data reference points – this is also referred to as predictive analytics.
[4] Deep learning is a type of machine learning in which multiple layers of processing are used to extract progressively higher level but more abstract features from data.
[5] Open source software is software with source code that anyone can inspect, modify and enhance.
[6] CogStack website

'Doctors' handwriting may be deemed illegible – but to a computer, their typing isn't much better! A lot of the information a patient tells their doctor is currently stored as typed text, which is difficult to access by other doctors, and often impossible to analyse by analysts and researchers. Patients therefore end up repeating the same story to different doctors which can be frustrating and time-consuming for everyone concerned. We, the CogStack team, are building artificial intelligence which will read and summarise doctors' notes so that they are easily understood by those that require them. In turn, this will reduce the administrative burden for the NHS and provide richer, standardised patient-centred data summaries.'

Dr James Teo, clinical lead for CogStack and clinical director of data science and consultant neurologist at King's College Hospital NHS Foundation Trust and King's Health Partners

'CogStack is an interrogation tool to interrogate existing data sets. A little bit like Google, it acts as a search engine that allows you to type in a very simple textual command.

You don't need to know the structure of the data, you don't need to know the exact term for what you're looking for. You say – I want to look for a particular disease, a drug, an adverse event and it searches everything.

It's interrogating data across multiple platforms, for example laboratory result data, clinical text data, and data that exists in hospitals and primary care.

CogStack has this fantastic ability to interrogate all of that data incredibly rapidly and provide a real time feed of what is happening.'

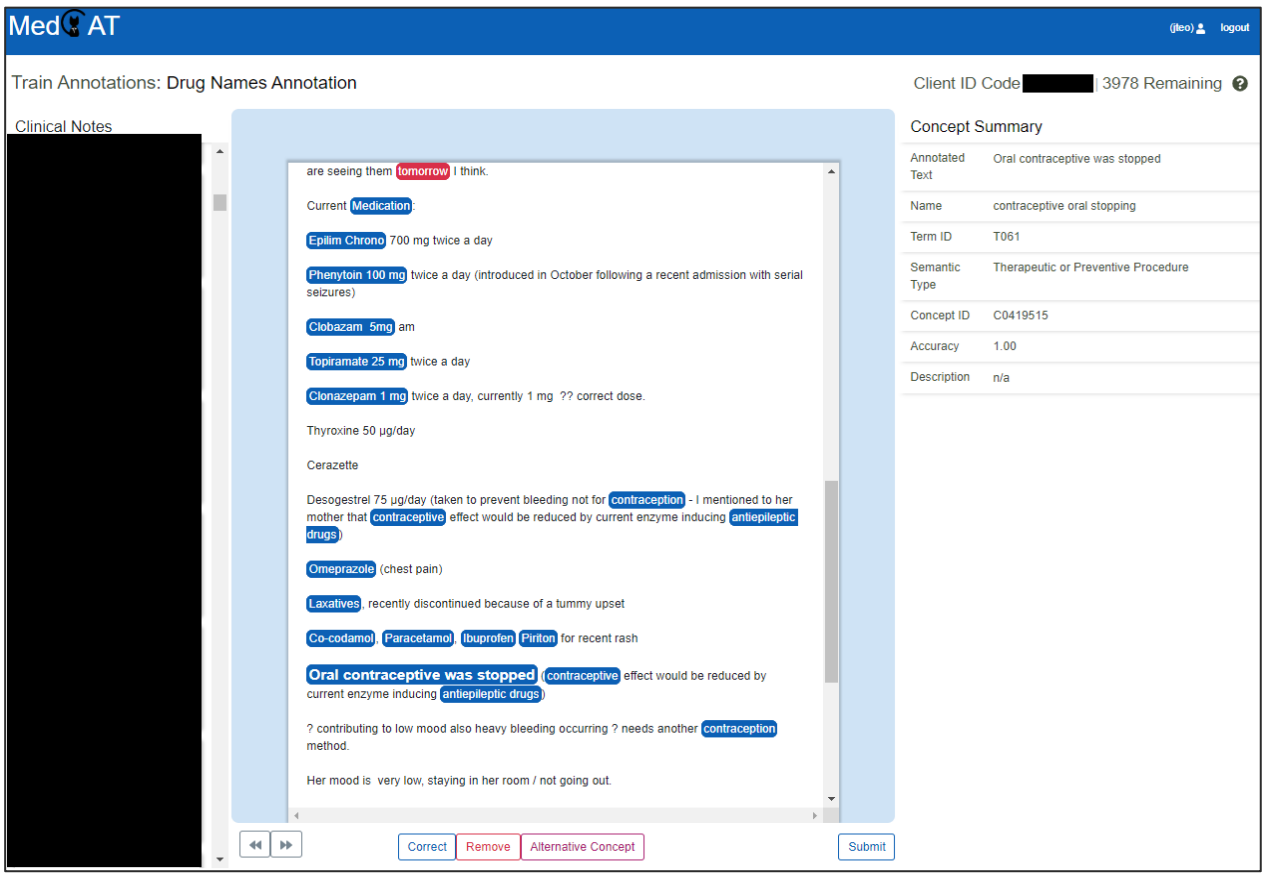
Dr James Galloway, consultant rheumatologist, King's College Hospital NHS Foundation Trust

Training and testing CogStack

The team use clinical NLP to code diagnoses, symptoms, events and procedures using algorithms.

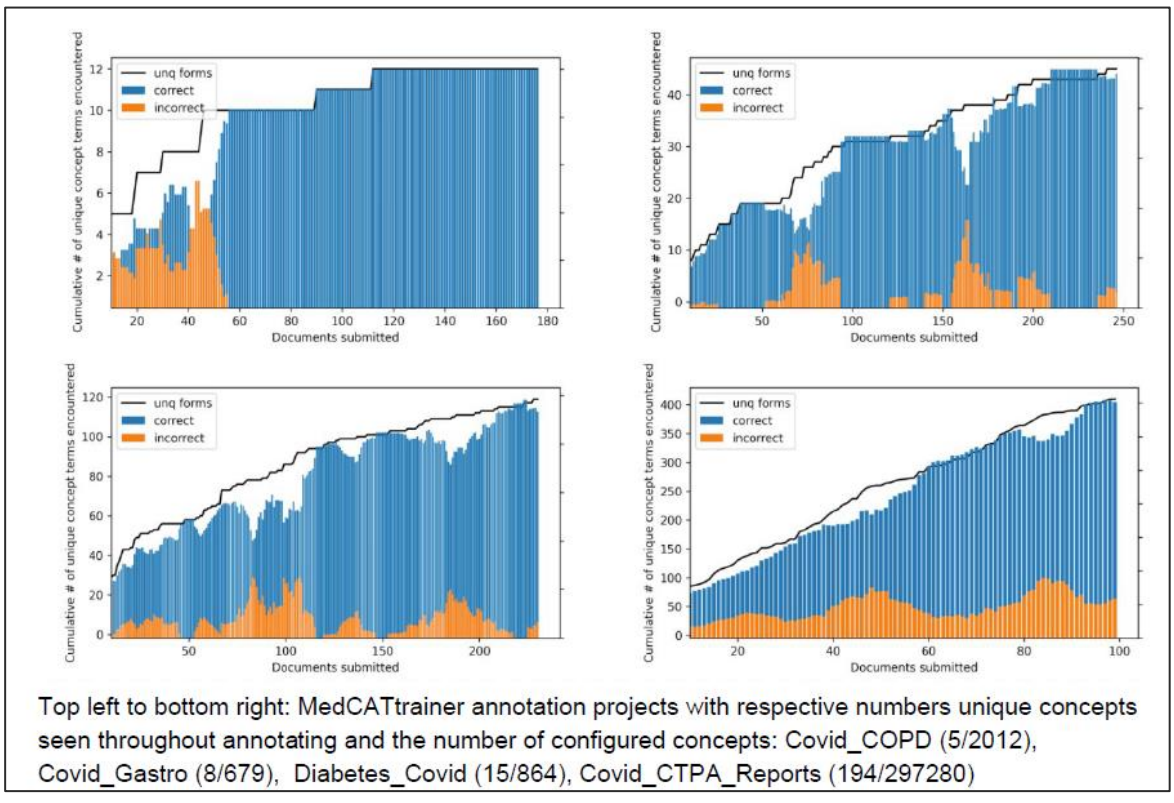
Junior doctors act as volunteers to train the AI, correcting its mistakes (**figure 2**). Once the AI has been trained, clinical staff don't have to spend significant amounts of their time collating data as they have done in the past.

Figure 2: Training AI to recognise medical words and sentences



The team have now trained and tested AI across many diseases (**figure 3**). After the doctors have 'marked' about 150 to 200 documents, which takes about a day, the AI becomes 80 to 90% accurate.

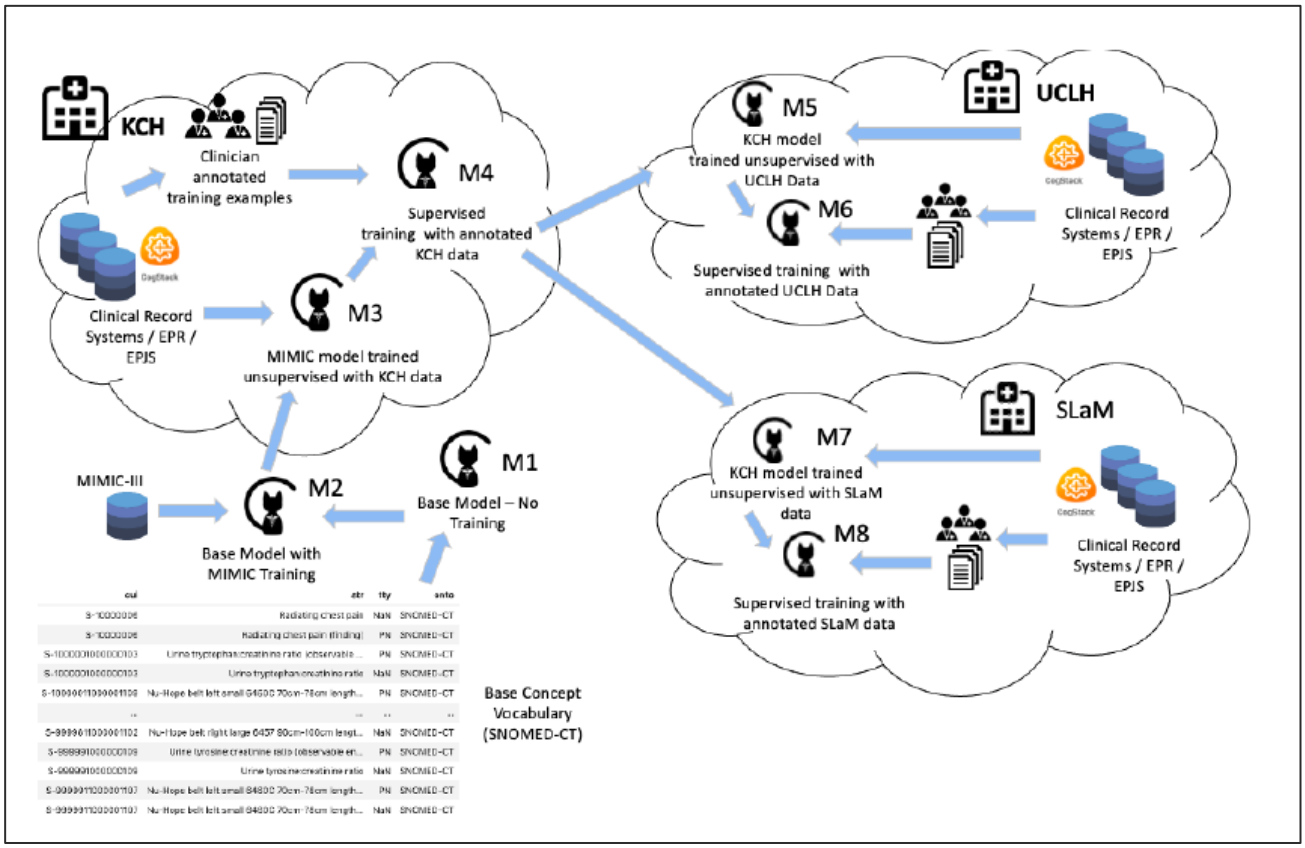
Figure 3: Training and testing AI across a range of diseases



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)

The AI is also trained and tested in different hospitals, each who will have their own language nuances (**figure 4**).

Figure 4: Training and testing AI in different 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

Making a difference with CogStack

NHS teams are already seeing the benefits of CogStack's ability to turn unstructured data into structured data. This section describes a few examples.

Improved coding in fracture outpatient clinic

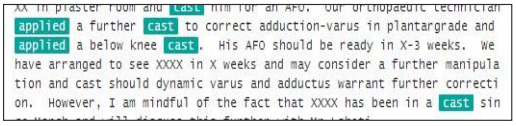
One of the first things the team used CogStack for was to test the depth of clinical coding in a fracture outpatient clinic. The data showed a low rate of procedures for new casts and splints, which seemed unusual for a fracture clinic**.**

Searching the data using CogStack took half an hour and revealed that large numbers of patients were missing codes **(figure 5)**. The blue columns on the chart below show the number of patients per procedure according to the data generated by clerks and clinical coders. The red columns with the higher patient numbers were generated by CogStack which searched for verbs and nouns. This shows that the information is in the records - dictated into letters or typed into free text - but is hard to access in the traditional way.

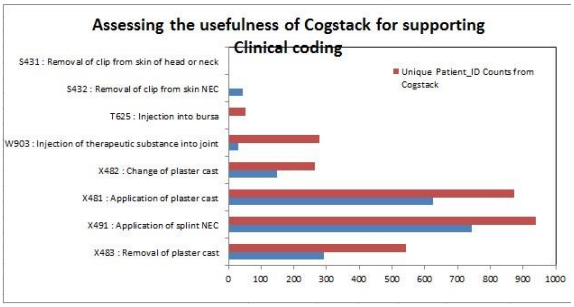
Figure 5: Improving the coding of patients

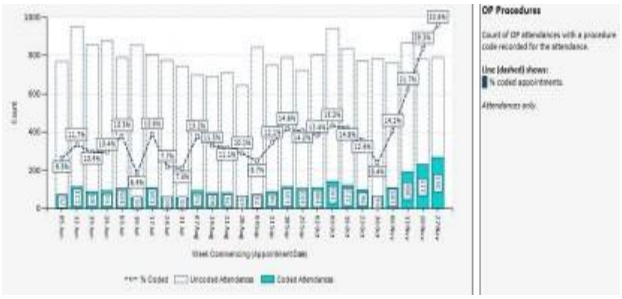
Scoping the level of under-coding found in unstructured text

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



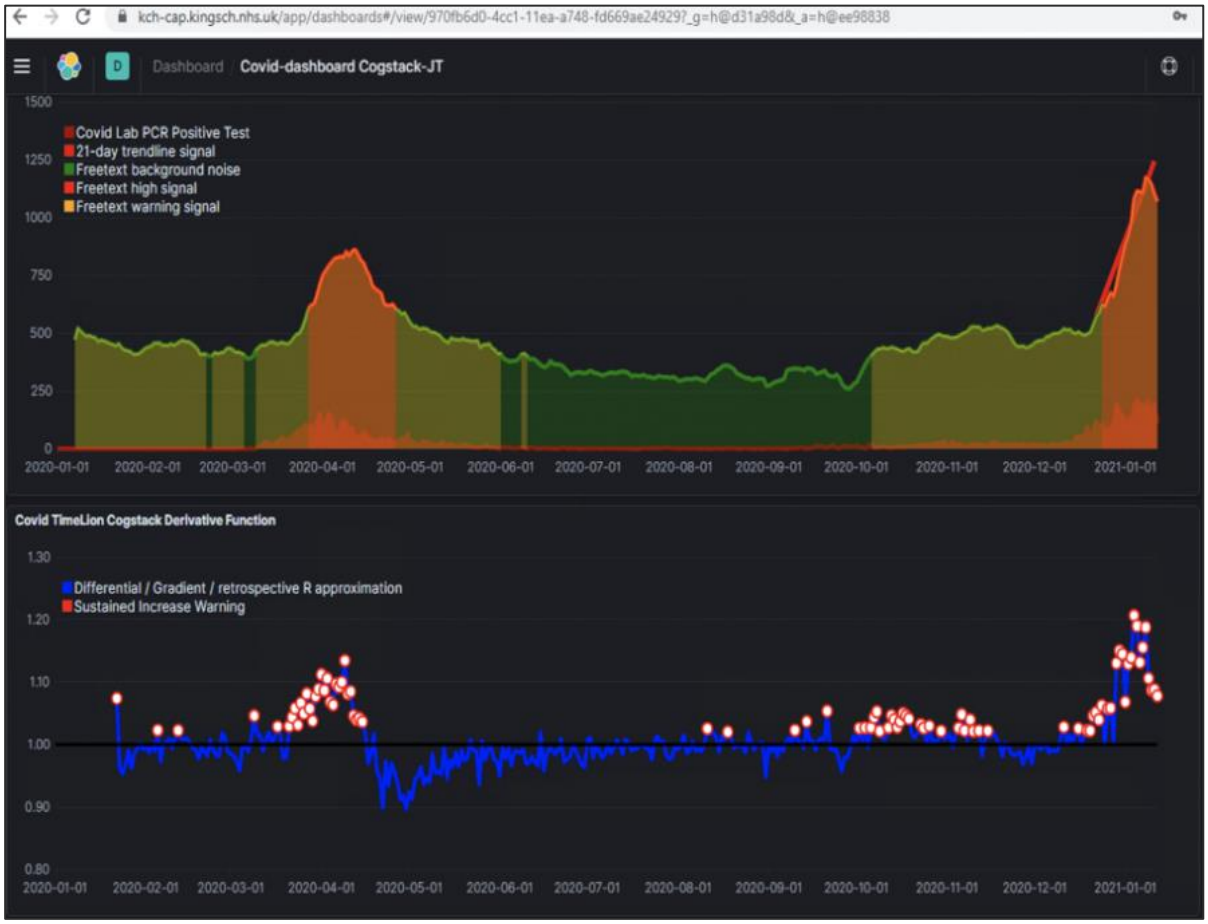


Identifying patients with Covid-19 in real time

At the start of the pandemic, when a range of terms were being used to describe Covid-19, CogStack was used to track patients who potentially had Covid in real time. The AI searched the health records for terms such as coronavirus, Covid-19 and Ncov, as well as phrases like fever and shortness of breath occurring in same sentence, and generated real time information within minutes of staff entering the data in the health record.

Figure 6 shows how words and phrases used by doctors and nurses to describe Covid in the text of records matched the laboratory samples peaks and troughs. Essentially, Covid and words related to it became a 'trending' phrase in the health record, just the same way words can 'trend' on social media. This 'trending' was detectable in real-time and in fact preceded the laboratory results by one to two days, meaning that this approach was both faster and less burdensome for staff.

Figure 6: Streaming unstructured data feeds



Improving safety for patients with rheumatoid arthritis

One of the most common medicines used for people with rheumatoid arthritis is methotrexate. It is recognised as a very good medicine, but there is a small risk of serious harm if people continue to take it if they develop renal (kidney) problems.

The rheumatologist team at King's College Hospital NHS Foundation Trust were keen to establish an approach for identifying people with rheumatoid arthritis who develop renal failure or renal impairment, so that they could stop using methotrexate before any harm happened.

Using CogStack, they were able to interrogate the electronic health records, both the structured and unstructured data, to identify all patients taking methotrexate. They also looked at the records for laboratory results, both within the hospital and also outside the hospital in the wider network, for evidence of renal problems. The resulting information was used to create a real time reporting system.

> *'Recently I got an email mid-morning alerting me that one of my patients had been admitted. Her renal function had deteriorated, and she was still taking methotrexate. I was able to immediately contact one of my registrars, who went straight to the ward and stopped the medicine, therefore pre-empting the potential problem and harm.*
>
> *We get an alert like this about once or twice a month. If the person is an inpatient, we can see them on the ward. If they are in the community, we contact them directly to stop the medication.'*
>
> Dr James Galloway, consultant rheumatologist, King's College Hospital NHS Foundation Trust

Integrating mental health and physical healthcare

The local mental health and acute trusts have developed an approach to improve the support available for patients who are experiencing psychological distress such as anxiety, low mood or depression, and who are also receiving treatment for a physical health condition. It is called IMPARTS (Integrating Mental and Physical Healthcare: Research, Training and Services)[7].

The IMPARTS approach includes collecting patient-reported outcomes. Patients attending the acute trusts are asked to answer a series of questions about their health before their appointment, using an iPad. The questions cover topics such as smoking, and depression and anxiety symptoms. The results are added to the patient's electronic health record, and used by clinicians to guide the consultation and treatment plan.

CogStack allows the team to put the patient-reported outcomes into context, so that they have a better understanding of the clinical state of the patients entering the IMPARTS programme. It tells the team whether doctors have previously recognised a problem like anxiety or depression, whether the patient is receiving any treatment for mental heath problems, and whether there are any changes as a result of IMPARTS.

'I think in the future CogStack is going to be revolutionary in the way in which clinical information is used for decision-making because it does three really important jobs:

- *improves patient safety by alerting clinicians to problems before they've recognised they are happening*
- *potentially allows for early intervention because you can spot a pattern or deterioration in clinical state and intervene earlier, and thereby improve patient lives, but also potentially save money*
- *because it makes information available about the whole of the patient's journey and their care, you can provide more holistic care - you can bring mind and body together.'*

Professor Matthew Hotopf, director, NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London

[7] IMPARTS website

Looking to the future

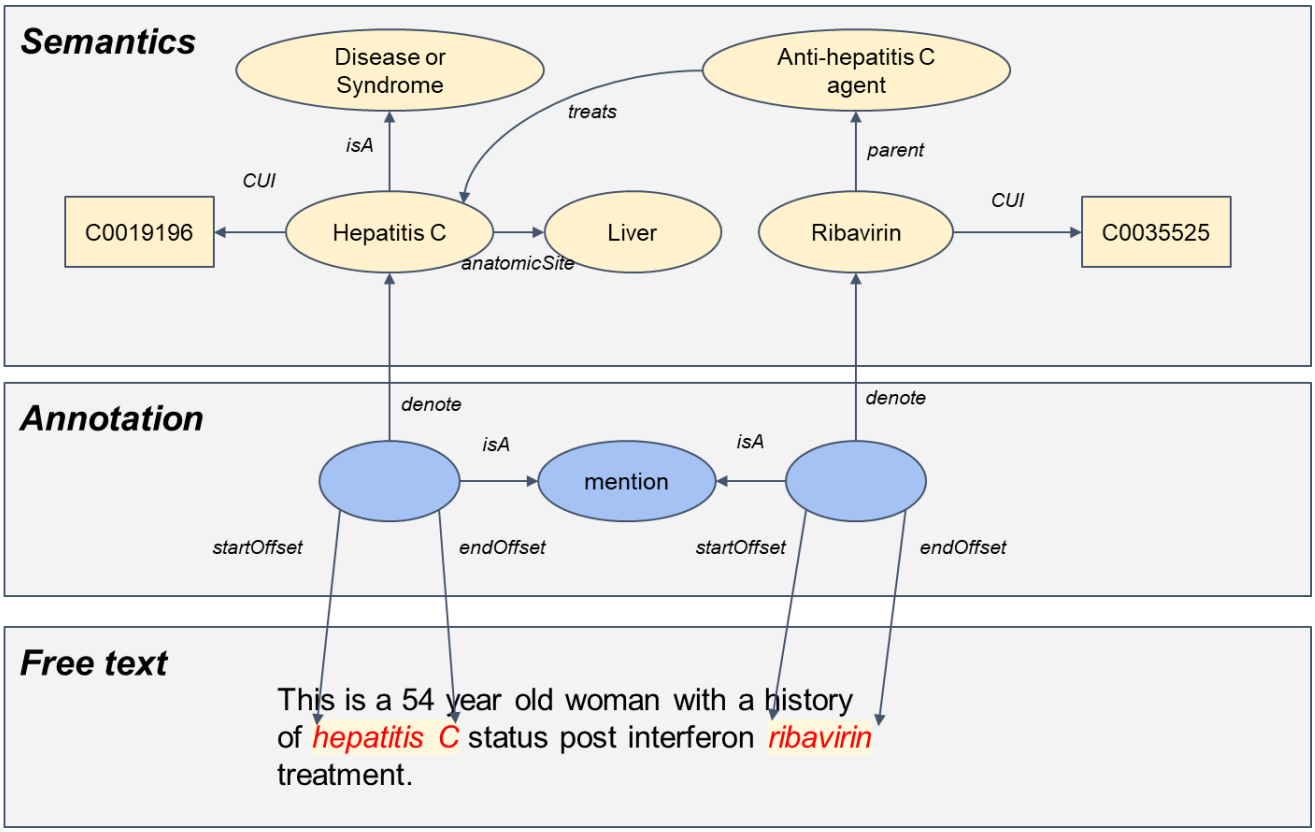
Introducing meaning into the data

The team are now starting to introduce meaning into the data. **Figure 7** shows how modern computational linguistics operate:

- A dictionary of meaning (**Semantics**) is created which maps out relationship between words. This is shown in the top layer.
- The **free text** is labelled with annotations or tags by a reader to mark out where the meaning is in the sentences. This is shown in the middle layer.
- An AI is then trained to recognise these mappings from bottom layer to middle layer to top layer. This allows translation of text with unstandardised vocabularies into standardised meaning.

This process is essentially teaching an AI grammar and vocabulary through lots and lots of examples.

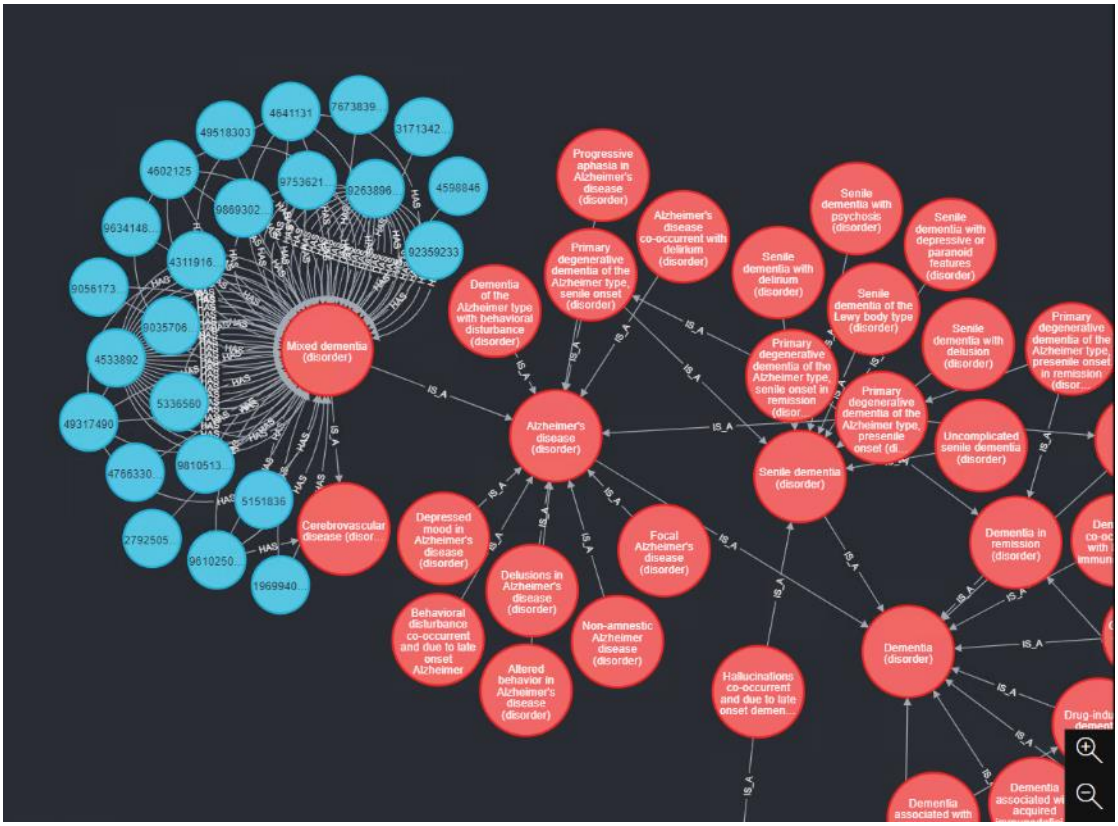
Figure 7: Using AI to attach meaning to text



Traditional dictionaries like the Oxford English Dictionary are lists of words and meanings. The dilemma facing an NLP AI is how do you use a dictionary if you can't read? The solution is to produce a dictionary that an AI can read, known as a semantic map of meaning or a knowledge graph (**figure 8**). They map out the relationship of words to each other, and the meaning of a word is derived from its relationship with other words.

This knowledge is explicitly taught to clinical coders as lists in official coding manuals, and is implicitly acquired by doctors in medical training. Semantic maps produce this knowledge for computers. Once a machine can derive meaning, it is much easier for a human to interact with it since the human doesn't need to learn how to speak in 'computer language' for the machine to understand it.

Figure 8: Semantic maps produce machine-readable meaning



CogStack win AI in Health and Care Award

In 2021 the CogStack team won an Artificial Intelligence in Health and Care Award from the NHS AI Lab[8]. With the government funding received from the award, the team will extend the training of the NLP AI, developed at King's College Hospital and King's College London, to multiple NHS trusts to widen the exposure and usability of the AI for clinical coding and extracting standardised information from text.

The work will build on the CogStack technology to establish a more efficient way to read and code records in collaboration with five NHS trusts: South London and Maudsley, King's College Hospital, Guy's and St Thomas', University Hospitals Birmingham and University College London Hospitals.

Making the case for investment

Accuracy in using CogStack for clinical coding is going to be evaluated for health economic benefits, and to support increasing automation of burdensome data-collecting and data-parsing processes in the NHS.

Findings of the health economic evaluation, technical documentation, and template business plans will be constructed and shared openly to any fast follower[9] healthcare organisations.

Conclusion

The aim of this case study is to make people aware of how researchers are exploiting and developing the capabilities of NLP in the NHS. CogStack is still in its research phase, but the potential for unlocking the wealth of information held in patient records to support the NHS in delivering patient care and clinical research is exciting.

[8] NHS England and NHS Improvement, The Artificial Intelligence in Health and Care Award

[9] A fast follower is an organisation that quickly picks up the innovations from other organisations

How to find out more about CogStack

Dr Thomas Searle, King's College London

Professor Richard Dobson, King's College London and University College London

Professor James Teo, King's College Hospital and Guy's & St Thomas' Hospital

Email: contact@cogstack.org

Twitter: @Cogstack

Website: https://cogstack.org/

Open-source code is available at: https://github.com/CogStack

This case study is part of the *Delivering value with digital technologies* programme that the HFMA is undertaking, supported by Health Education England. The programme aims to increase awareness amongst NHS finance staff about digital healthcare technologies, and enable finance to take an active role in supporting the use of digital technology to transform services and drive value and efficiency. For more information click here.

About Health Education England

Health Education England (HEE) is part of the NHS, and we work with partners to plan, recruit, educate and train the health workforce. HEE exists for one reason only: to support the delivery of excellent healthcare and health improvement to the patients and public of England by ensuring that the workforce of today and tomorrow has the right numbers, skills, values, and behaviours, at the right time and in the right place.

HEE's Digital Readiness Programme, commissioned by NHSX, aims to uplift digital skills, knowledge, understanding and awareness for all our health and care workforce. This includes:

- Supporting the right culture and environment, for example by ensuring digital is understood, embedded and championed at trust and ICS board level.
- Professionalising the digital workforce through support for professional bodies, regional Informatics Skills Development Networks, and collaborative community networks.
- Establishing learning and development through the NHS Digital Academy and specific learning and development initiatives, for example the Florence Nightingale Digital Nurse Scholarship, and through access to tailored, appropriate online learning for all.
- Building our future digital workforce by undertaking workforce analysis and demand forecasting, and sustainable models to recruit talent, for example through graduate schemes, as well as opportunities for nurturing existing talent, for example through the Topol Digital Health Fellowships.

For more information visit the Digital Readiness Programme website or follow the programme on Twitter @HEE_DigiReady.

About the HFMA

The Healthcare Financial Management Association (HFMA) is the professional body for finance staff in healthcare. For over 70 years, it has provided independent and objective advice to its members and the wider healthcare community. It is a charitable organisation that promotes best practice and innovation in financial management and governance across the UK health economy through its local and national networks.

The association also analyses and responds to national policy and aims to exert influence in shaping the wider healthcare agenda. It has particular interest in promoting the highest professional standards in financial management and governance and is keen to work with other organisations to promote approaches that really are 'fit for purpose' and effective.

HFMA

HFMA House - 4 Broad Plain, Bristol, BS2 0JP

T 0117 929 4789

E info@hfma.org.uk

W hfma.org.uk