Towards nationally curated data archives for clinical radiology image analysis at scale: Learnings from national data collection in response to a pandemic

Dominic Cushnan[1] (iD), **Rosalind Berka**[2], **Ottavia Bertolli**[2], **Peter Williams**[1],
Daniel Schofield[1], **Indra Joshi**[1], **Alberto Favaro**[2], **Mark Halling-Brown**[3,4],
Gergely Imreh[2] (iD), **Emily Jefferson**[5,6], **Neil J Sebire**[5], **Gerry Reilly**[5] (iD),
Jonathan C L Rodrigues[7], **Graham Robinson**[7], **Susan Copley**[8] (iD),
Rizwan Malik[9], **Claire Bloomfield**[10,11], **Fergus Gleeson**[10,11], **Moira Crotty**[12],
Erika Denton[13], **Jeanette Dickson**[14], **Gary Leeming**[15], **Hayley E Hardwick**[16] (iD),
Kenneth Baillie[17], **Peter JM Openshaw**[18], **Malcolm G Semple**[19],
Caroline Rubin[20], **Andy Howlett**[12], **Andrea G Rockall**[8,21], **Ayub Bhayat**[22],
Daniel Fascia[23], **Cathie Sudlow**[24], **NCCID Collaborative**[25]
and Joseph Jacob[26,27]

Abstract

[1]NHSX, UK
[2]Faculty, UK
[3]Scientific Computing, Royal Surrey NHS Foundation Trust, UK
[4]CVSSP, University of Surrey, UK
[5]Health Data Research UK, UK
[6]Health Informatics Centre (HIC), School of Medicine, University of Dundee, UK
[7]Department of Radiology, Royal United Hospitals Bath NHS Foundation Trust, UK
[8]Imaging Department, Hammersmith Hospital, Imperial College NHS Healthcare Trust, UK
[9]Department of Radiology, Bolton NHS Foundation Trust, UK
[10]National Consortium of Intelligent Medical Imaging (NCIMI), The Big Data Institute, University of Oxford, UK
[11]Dept of Oncology, University of Oxford, UK
[12]NHS England & Improvement, UK
[13]Norfolk and Norwich University Hospital Foundation Trust, UK
[14]The Royal College of Radiologists, UK
[15]Institute of Population Health, Faculty of Health and Life Sciences, University of Liverpool, UK
[16]National Institute of Health Research (NIHR) Health Protection Research Unit in Emerging and Zoonotic Infections, UK
[17]Roslin Institute, University of Edinburgh, UK
[18]National Heart and Lung Institute, Imperial College London, UK
[19]NIHR Health Protection Research Unit, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, UK
[20]Department of Radiology, University Hospital Southampton NHS Foundation Trust, UK
[21]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, UK
[22]NHS Arden & Greater East Midlands Commissioning Support Unit, UK
[23]Harrogate District Hospital, UK
[24]British Heart Foundation Data Science Centre Led by Health Data Research UK, UK
[25]NCCID Collaborative, UK
[26] Department of Respiratory Medicine, University College London, UK
[27]Centre for Medical Image Computing, Department of Computer Science, University College London, UK

Corresponding author:
Dominic Cushnan, NHSX, Skipton House, 80 London Road, London SE1 6LH, UK.
Email: dominic.cushnan@nhsx.nhs.uk

The prevalence of the coronavirus SARS-CoV-2 disease has resulted in the unprecedented collection of health data to support research. Historically, coordinating the collation of such datasets on a national scale has been challenging to execute for several reasons, including issues with data privacy, the lack of data reporting standards, interoperable technologies, and distribution methods. The coronavirus SARS-CoV-2 disease pandemic has highlighted the importance of collaboration between government bodies, healthcare institutions, academic researchers and commercial companies in overcoming these issues during times of urgency. The National COVID-19 Chest Imaging Database, led by NHSX, British Society of Thoracic Imaging, Royal Surrey NHS Foundation Trust and Faculty, is an example of such a national initiative. Here, we summarise the experiences and challenges of setting up the National COVID-19 Chest Imaging Database, and the implications for future ambitions of national data curation in medical imaging to advance the safe adoption of artificial intelligence in healthcare.

Introduction

Medical images have a central role in disease diagnosis, predicting prognosis and risk stratification. Recent advances in computing power and a growth in the availability of larger data repositories (including open-access datasets) have led to an increase in the medical applications of artificial intelligence (AI) systems, with a focus on image analysis to support clinical pathways of care.[1] These systems aim to deliver similar image analysis functions to human interpretation of medical imaging, as well as detect insights beyond the capability of the human eye. In response to the coronavirus SARS-CoV-2 disease (COVID-19), huge efforts have been made by the deep learning community analysing imaging datasets with the aim of enhancing the certainty of COVID-19 diagnosis[2,3] and improving disease outcome prediction[4,5] (such as ITU admission and death) with varying levels of success.[6,7,8] The main barriers to the development of successful AI models have been the variable quality of imaging data available for model training, and the limited amount of curated data that comprehensively encompasses the variability of real-world clinical cases necessary for robust model validation.

Until recently, the collection of imaging data to enable the development of AI solutions in the UK has occurred on a relatively limited scale, dependent largely on bespoke local academic and commercial partnerships within healthcare institutions.[9] Homogenous but small datasets are expensive to curate and limit clinical validation of AI tools using a wider study population.[10,11] In this regard, algorithms could be prone to biases[12] such as gender, ethnicity,[13] socioeconomic status,[14] and environmental factors, which show the need for a well-diverse dataset[15] such as a national-scale dataset.

Larger national-scale imaging datasets covering multiple geographic regions, using routinely collected data and accessible to the wider community, have the potential to enable development of more robust AI tools where the data can help evaluate clinical questions, validate AI algorithms and assess translational challenges. The general lack of such datasets is cited as one of the most significant challenges to the adoption of AI technologies in the NHS.[16]

In this paper, we discussed our experiences and challenges in setting up the National COVID-19 Chest Imaging Database (NCCID). The NCCID was established by the NHS AI Lab in May 2020 to combat the COVID-19 pandemic using AI technologies. It assimilates data of over 15,000 patients from 22 NHS Trusts and Health Boards to enable validation of AI algorithms.[17,18,1] The NCCID was endorsed by the Royal College of Radiologists and aimed to provide a central repository of chest X-rays, chest computed tomography (CT) scans and cardiac magnetic resonance images (MRIs) of COVID-19 diagnosed patients (through reverse transcription polymerase chain reaction (RT-PCR) testing) and controls.

Methodology

The experiences from the NCCID have emphasised that there are opportunities to enhance the current UK health data infrastructure for the national collection of medical imaging data. This could be accelerated if led by a neutral national healthcare body, such as NHSX in England, in close collaboration with existing entities leading on data collection initiatives in the UK (several of which are

referenced in this paper) and the devolved administrations in Scotland, Wales and Northern Ireland.[19] These learnings are summarised in Table 1.

A diagram and description of the infrastructure set-up for the NCCID is also included in Figure 1.

Information governance: Rationalising data governance processes to reduce the barrier to participation in future national data collection exercises

The NCCID's data governance arrangements were established quickly as a consequence of regulations put in place for data collection during COVID-19. The notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 (COPI Notice), valid for England and Wales until end of September 2021, enabled the NCCID to collect pseudonymised patient data without seeking consent to support a national response to the pandemic.[20,21] Once the COPI Notice expires, additional ethical approvals will need to be acquired potentially through application to the Confidentiality Advisory Group (CAG).[22]

In addition to compliance with the terms of the COPI Notice, the NCCID has established robust data governance procedures. Using the IG expertise of NHSX advisors and project partners, in particular the National Consortium for Intelligent Medical Imaging (NCIMI), the NCCID initiative received formal ethical approval from the Health Research Authority's Research Ethics Committee to establish a research database for the collection and sharing of the NCCID data.[23] Indeed using a research database model was crucial in communicating the ethical basis of the initiative to hospital sites to encourage participation. Ongoing amendments became necessary as the data points collected in the database expanded to keep pace with the rapid progression of scientific understanding around COVID-19. These experiences highlight the comprehensive support and expertise required to establish data governance arrangements for national-level initiatives.

A Data Protection Impact Assessment (DPIA) is formally shared with each participating hospital site alongside a Data Sharing Agreement (DSA) for approval by Data Protection Officers (DPOs). The national data opt-out policy mandates that hospital sites should not upload data for patients who have registered to opt-out of research studies, and additional checks have been built in to the NCCID's data management procedures to ensure this data is automatically deleted in rare instances where it is mistakenly uploaded.[24] Before accessing the database, researchers must sign a Data Access Agreement and Data Access Framework contract.

Information governance (IG) for national data collection is further complicated by the fact that legislation is

Table 1. Learnings summary from the National COVID-19 Chest Imaging Database (NCCID) for future data collection initiatives.

|	Category	Learning for future data collection
1	Information governance (IG)	Data governance processes must be clarified and standardised to reduce barriers to NHS Trust participation in future national data collection exercises
2	Database linkages	Collaboration and linking with other databases improve the quality, completeness and coverage of the data collected, increasing opportunities for discovery
3	Automation	Incorporating automation is vital to enable mass data collection and reduce manual data capture and burden for hospital staff
4	Trusted research environments (TREs)	Building national infrastructure enables data to be accessed and analysed in a safe and secure way, facilitating research and innovation
5	Availability of validation datasets	Creation of large-scale high-quality validation datasets helps to accelerate the route to market for new artificial intelligence models
6	Funding	Defining a variety of funding mechanisms to support NHS Trust data collection activities and infrastructure engineering is key to the sustainability of any national programme
7	Patient and public engagement	Consulting patients and the wider public is important to ensure that concerns regarding how patient data is used and stored are addressed and that this is done in a safe, secure and ethical way
8	Benefit share models	Defining benefit share frameworks helps to ensure that the NHS benefits at the local site level, which can incentivise participation in national data collection exercises
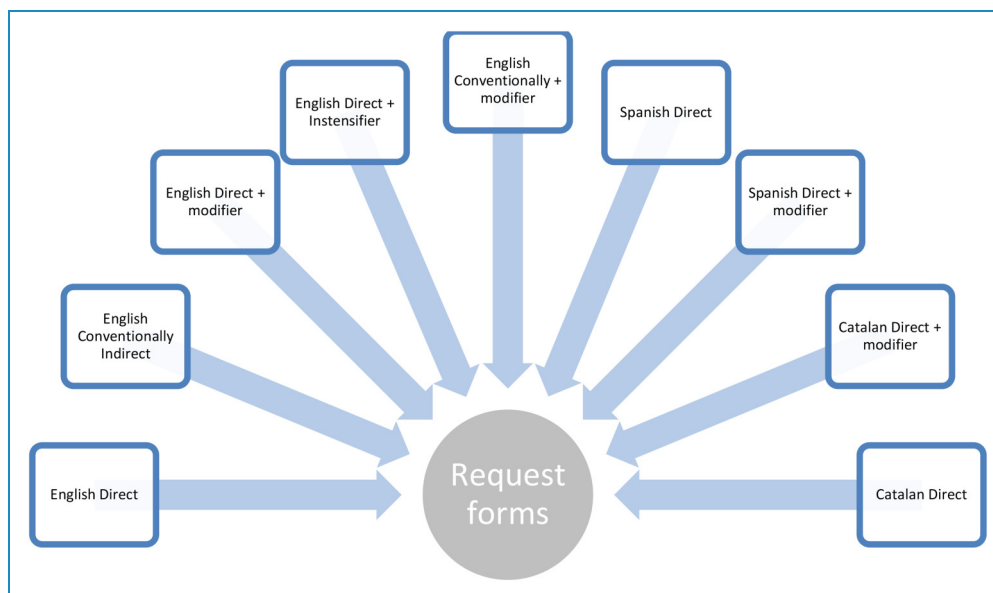
Figure 1. National COVID-19 Chest Imaging Database (NCCID) infrastructure and explanation.

Table 2. Database linkages for the National COVID-19 Chest Imaging Database (NCCID).

	Database linkage	Objective
1	International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) 4C repository[59]	To enhance the number of clinical variables that can be evaluated alongside clinical imaging and to reduce duplication of data gathering efforts for hospital sites during a period of stretched limited resources.
2	NHS England and Improvement	To provide comprehensive ethnicity data, which can be challenging to collect when relying solely on hospital records. Collecting reliable ethnicity data is essential given the disproportionate impact of COVID-19 on the Black, Asian and Minority Ethnic (BAME) populations, and to avoid introducing inherent biases during artificial intelligence model development.[60,61]
3	National Scottish Picture Archiving and Communications System (PACS) and Safe Haven Network[62,63]	To increase the geographic coverage of the database to the entirety of Scotland.[64]

not consistent across the devolved administrations. For example, the COPI Notice does not apply in Scotland, where approval must be sought through the Public Benefit and Privacy Panel (PBPP) instead. Navigating different regulatory procedures across UK territories inevitably delays the ability to facilitate UK-wide imaging data collection. In the short term, it would be beneficial for a neutral healthcare body to publish clear, available guidance for researchers to understand the necessary processes to be undertaken when setting up data collection studies. In the long term, collaboration across national organisations would help to review existing data governance models and identify how these can be streamlined to support researchers.[25]

Database linkages: Promoting discovery through collaboration with other databases

The NCCID has been collaborating with other UK database initiatives to improve the quality, completeness and coverage of the data collected. These are detailed in Table 2.

Linking the NCCID to other databases was achieved one database at a time, and maintaining the necessary support from IG experts, project managers, technologists and clinicians over a prolonged period is difficult to sustain.[2] Establishing national-level infrastructure and processes to integrate and allow communication between existing datasets will be important to leverage the potential of data collected within the NHS. A number of initiatives in the UK have made progress on this. The ambition of Health Data Research UK (HDR UK) is to create a UK-wide approach for the development of secure and transparent data services through initiatives such as the Health Data Research

Table 3. Examples of existing initiatives facilitating collection of medical imaging data.

	Data initiative	Description of their work
1	National Consortium of Intelligent Medical Imaging (NCIMI)	Led from Oxford's Big Data Institute and brings together expertise across 14 NHS partners, academia and 13 industry partners to support the development, testing, validation and adoption of new clinical imaging AI tools into the NHS.
2	PICTURES (Interdisciplinary Collaboration for the Efficient and Effective Use of Clinical Images in Big Data Health Care Research)	Led by the University of Dundee in conjunction with NHS Scotland and other academic bodies. The study is collating 30 million images from the Scottish National PACS, across all 14 Scottish health boards, to support the development of AI technologies. PICTURES will allow researchers to work on vast amounts of data in a secure environment that protects individual patient information.
3	East Midlands Radiology Consortium (EMRAD)	Working with Kheiron Medical to develop and test a CE-marked Mammography Intelligent Assessment (MIA) tool that can help detect breast cancer, using the historic data and images provided by EMRAD.[65] Regional imaging networks are increasingly able to facilitate the rapid testing and application of machine learning technologies within large databases.[66]
4	Yorkshire Imaging Collaborative	Partnered with Leeds University to develop a Trusted Research Environment that extracts data from across the network to predict future capacity and demand requirements for each imaging modality across the region.[67]

Innovation Gateway, a common discovery portal to UK health research data for accredited researchers and innovators.[26,27] HDR UK is also working closely with other Trusted Research Environments (TREs[28]) and Safe Haven Environments to increase the discoverability of UK health datasets and is promoting accessibility to vital COVID-19 datasets via the COVID-19 National Core Studies programme.[29] Our experience in the NCCID has highlighted the national-level coordination required to overcome the barriers of a fragmented data landscape by establishing data access terms and processes that are common across UK initiatives.[303] Data federation is a potential solution, which refers to the aggregation of different data sources into a standard data format for users to access.[31] This is a long-term vision that will require national coordination to be fulfilled, yet success will be key to leveraging the potential of data collected in the NHS.

Automation: Building infrastructure to enable mass data collection and reduce manual data capture for hospital staff

The NCCID project aims to reduce the burden on staff at hospital sites by running a pilot exercise to automate imaging data collection. To rapidly commence data collection, the initial architecture of the NCCID pragmatically relied upon Sectra's Image Exchange Portal (IEP) as processes for transferring images via this network are well-known, resulting in minimal setup overhead at hospital sites.[32] To allow the automation of image transfer, Royal Surrey NHS Foundation Trust (RSNFT) has deployed SMART boxes at NCCID pilot sites, installed on either physical hardware or a virtual machine, that act as a data collection node at each site. The SMART box fulfils three roles: (a) management of cases and clinical data through an internal web portal, (b) image collection and de-identification through a DICOM server, and (c) data transfer to the central NCCID data warehouse. This infrastructure, previously used to create large-scale imaging databases such as the OPTIMAM Image Database project, is significantly reducing workloads on hospital PACS teams.[33]

The capture and storage of clinical variables are supported by technology and data standards that are less comprehensive and universal than those for images. The manual processes to capture these data points utilise local staff or informatics to collate the information from the local hospital systems. Automating the collection of these clinical data points by the NCCID has proved difficult due to the heterogeneity of the clinical systems and integration processes. The initiative is currently exploring possible

Table 4. Future aspirations to support the end-goal of implementing artificial intelligence (AI) technologies in the NHS.

	Future aspiration	Explanation
1	Central cloud-based infrastructure, managed by a neutral national body	To facilitate the automated collection of medical imaging data into comprehensive training and validation datasets.
2	Improved collaboration between the entrepreneurial community and the NHS	To encourage the identification of new life-saving technologies through more widely accessible data.
3	Greater engagement with the radiology community as users of AI	To ensure that the AI tools being prioritised for deployment are designed for optimal utility and address the needs of radiologists.
4	Further guidance and support on regulatory approvals and AI model evaluation criteria	For both technology developers and commissioners of AI technology – to ensure that AI products are deemed safe and effective for use in a clinical setting. NHSX has already published an 'AI Buyers Guide' to advise commissioners of AI technology on this matter, which is a useful starting point.[68]
5	A centralised, vendor-agnostic deployment infrastructure	To implement AI technologies for use in complex environment, and support the pathway from innovation to deployment

solutions to this challenge in addition to the data linkage with ISARIC 4C.

The experiences of the NCCID demonstrate that a coordinated UK-wide approach to automating components of the data collection process across multiple use cases of medical imaging could accelerate the development of AI technologies. Numerous regional and national initiatives are already underway including five new centres of excellence aiming to enhance digital pathology and medical imaging AI development and deployment,[34] initiatives such as PICTURES[35] led by the University of Dundee in conjunction with NHS Scotland and at a more local level, radiology imaging networks such as EMRAD (see Table 3).

Given the wide-ranging initiatives that already exist in the UK, there is an opportunity to establish national-level linkage and discoverability to support combining such networks and datasets to enable, promote and support public health research.

TREs: Building infrastructure to enable secure data access for research

Ensuring that data can be accessed and analysed in a safe and secure way is crucial to the successful development of AI technologies. Data access can be organised by bringing "the data to the users" or "the users to the data". The NCCID, due to the urgency presented by the pandemic and the need to make data available at pace, adopted the former approach, pushing data to users that are General Data Protection and Regulation (GDPR) and security compliant.

Bringing users to the data is a long-term ambition of the NCCID, as it can help reassure both controllers of healthcare data and the patient community that data is being managed in a secure environment. This model is increasingly adopted through the use of TREs, otherwise known as Data Safe Havens. TREs ensure that all data remains in infrastructure approved by the controller, thereby making it easier to audit and ensure compliance with IG standards. A number of initiatives in the UK have supported research into TREs in recent years, which have accelerated as a result of the data response to COVID-19. The Scotland National Safe Haven programme, Genomics England Research Environment and the NHS Digital TRE for England are all examples of enabling advances in TRE infrastructure design executed against the 'five safes' – safe people, projects, settings, data and outputs.[4]

However, progress remains to enhance TRE models to better cater for the requirements of researchers and technology developers, and adapting their functionality for scalability.[4] Users must be provided with the computational resource, data storage and utilities necessary to do useful work within the TRE itself, while preventing extraction of provided data. TRE users should also have access to statistical and data visualisation and annotation software tools, the ability to access large virtual machines equipped with graphical processing units (GPUs) and high performing, flexible compute power to manage large volumes of data, including imaging. Fair licensing and costing models are also necessary to make data more widely available to the public and commercial companies (the complexities of such commercial models are explored further in the Benefit share models: Defining frameworks to ensure that data sharing directly benefits the NHS system section – Benefit Share Models).[36,37] All these factors will require consideration in the future development of national

infrastructure to facilitate access to medical imaging data in the UK.

Availability of validation datasets: Promoting creation of large-scale high-quality validation datasets

Automation can support data collection for national-level validation datasets, which endeavour to capture an unbiased representation of the UK population. Validation datasets should ideally remain dedicated to the purpose of assessing the performance of established AI models. However, within the expansive research environment of the UK, it can be challenging to ensure that a subset of the data within validation datasets has never previously been utilised for training the same AI models that are to be validated.[38] The creation of an entirely segregated validation dataset to assess AI technologies was a core objective at the centre of the NCCID, which has yet to be fully realised. Due to the urgency presented by the pandemic and the requests from technology developers to share the NCCID data as soon as it became available, assignment to the NCCID validation dataset occurs via random sampling of patients. This initial pragmatic choice allowed users access to validation data while the NCCID was growing in scale, but once of a sufficient size, the aim remains the provision of comprehensive, tailored validation sets. This process will also need to include image annotation, which is crucial for curating high-quality datasets for medical image analysis.[2]

Hospital sites participating in the NCCID are asked to confirm if they are sharing the data with other entities as a way of understanding what data might have been used in the development of an AI algorithm for COVID-19 research. If data has been shared with other entities, it is excluded from the validation dataset. A consequence of this policy is that it takes longer and requires numerous participating hospital sites to build up a large enough validation sample that can be used to assess algorithms for robustness.[5] There is currently no easy solution to this problem. To request a hospital site to set aside data, which would have immediate value when used in local studies, for the express purpose of allowing long-term creation of a national AI model validation dataset, is a significant and probably unrealistic requirement. Hospitals may also be reluctant to share data that could be used for commercial purposes. There are a small number of examples of UK studies that have managed to curate hold-out unseen validation datasets, such as via the OPTIMAM study, NCIMI, and the PROSPECTS trial, but it is by no means common practice.[39,40] However, one could argue that the UK, with national oversight and systematic coordination of health and care networks in the NHS, would be in a unique position to develop hold-out validation datasets. A nationally coordinated effort to define and implement best

practices for generating validation datasets would significantly advance the path to independent robust validation, which is currently lacking for CE marked solutions. Robust validation in turn could speed up regulatory processes for the licensing and deployment of AI technologies in the UK.

Funding: Defining mechanisms to support data collection activities and infrastructure engineering as a research discipline

Understanding which funding mechanisms were the most appropriate for the NCCID to pursue was difficult because data collection as a pursuit in itself is not widely recognised as clinical research in the UK. As the NCCID required support from hospital Research & Development (R&D) departments to facilitate collection of clinical data, a number of sites recommended applying for inclusion on the National Institute of Health Research (NIHR) Urgent Public Health portfolio of studies, whereby hospital resources can be diverted towards supporting prioritised initiatives. However, the NCCID was deemed unsuitable for Urgent Public Health badging, resulting in some sites declining participation due to a lack of capacity among local research staff during the pandemic.

There are also challenges with sourcing research nurses or research coordinators for radiology studies specifically and with setting up such studies in radiology departments, as there is a significant shortage of radiology staff members.[41,42] To date, apart from hospital sites that reallocated or sourced in-house funding, the NCCID initiative has only received funding from NHSX and has predominantly relied on voluntary support from hospital sites to contribute data. This model has sufficed due to the urgency presented by the pandemic, but would be unsustainable for long-term mass collection of imaging data. Further support and collaboration at a national level are required to establish funding models that are specifically designed for national data collection infrastructure, including imaging, as a research discipline in its own right.

Patient and public engagement: Consulting patients and the wider public on the requirements and benefits of algorithmic development in medical imaging

The adoption of AI technologies in medical imaging, particularly with the input of imaging specialists such as radiologists, can have considerable benefit for patients and the healthcare sector where over 100,000 imaging studies are performed daily in the UK.[43,44] The 2019 Clinical Radiology Workforce census highlights increases in workforce shortages, with seven out of 10 clinical directors of UK radiology departments expressing the view that there

were insufficient numbers of clinical radiologists to be able to deliver safe and effective patient care.[6] AI tools may improve the accuracy of diagnosis and the estimation of likely prognosis for a patient in some settings, as well as potentially improve efficiencies by speeding up screening processes and reducing the number of human readers required.[7] However, despite data used in AI software development being de-identified, there are understandable concerns with how patient data is stored and used, particularly for purposes that result in profit for commercial organisations.

For the NCCID specifically, the Data Access Committee includes ethical and patient advisors, who participate in the review and approval of all applications to access the database.[45] This ensures that the patient viewpoint is always considered before a decision is made to share data with the requesting applicant. The public are also kept informed on the projects that the NCCID data is currently supporting through study publications on the NCCID website, which is important for patients to understand how their data is being used.[46]

The Patient and Public Involvement and Engagement (PPIE) methods adopted for the NCCID will be expanded in future. NHSX is working with the National Data Guardian (NDG) and the Department of Health and Social Care to use medical imaging as a case study for exploring public attitudes towards the sharing of health and social care data for data-driven research and innovation in England.[47] NHSX also plans to collaborate with the Ada Lovelace Institute to trial a model for algorithmic impact assessments. Such assessments will involve public engagement on AI solutions developed using medical imaging data and encourage innovators to review and refine their products based on discussions with citizens about its ethical and societal implications.[48] These mechanisms are useful for understanding the public perceptions of using data to develop and test algorithms, and the level of trust felt towards algorithms in clinical care supporting the role of human readers.[49] More targeted research and public consultation on the impact of this is needed to support the successful adoption of AI technologies in medical imaging at scale in the UK.

Benefit share models: Defining frameworks to ensure that data sharing directly benefits the NHS system

NHS data used in the development and validation of AI imaging algorithms should result in improved patient care. At the same time, the NHS needs to receive fair value from the data assets that it holds. There has been much discussion of the forms this value might take, and can include clinical, social, economic development, environmental and commercial value. The commercial value can itself take a range of forms, including monetary, intellectual property (IP) ownership, and equity and royalties from any products developed, all of which can be achieved via payment or in the form of beneficial access terms for the NHS.[50] It is important to create the right frameworks to identify and realise the benefits for patients and the NHS where data underpins innovation, with an increasing focus on establishing a new 'social contract' to achieve this.[51]

Due to the severity of the COVID-19 pandemic, the NCCID has established a contractual arrangement with approved data users whereby data is provided to the user free of charge, but any related technologies developed as a result of this must be provided at zero-licence fee cost to the NHS for the duration of the Data Access Agreement signed by the data user. Importantly, all applicants must be able to demonstrate that their proposal will directly benefit the NHS to be granted access. This model aims to expedite the development of potentially useful tools, which is crucial in the context of COVID-19 but may not be appropriate for non-COVID-19-related technologies.

Another benefit share model that has the potential to accelerate the development of AI technologies[52] contractually requires data users to share derived data such as image annotations from their analysis for the benefit of all researchers. High-quality image labelling, often performed as an internal exercise within a research group, can be time-consuming and expensive to create. Trained experts may be necessary to annotate the images or check the annotations created by automated tools.[53] High-quality annotation is vital for supervised training, testing and validation of AI algorithms, and there may be inherent benefits in openly sharing the outputs of image annotation from multiple expert sources with the wider research community. The national influence and scale of the NHS mean that it is suitably positioned to encourage the open sharing of annotated datasets in the UK through collaborations with research consortiums and industry partners.

It is also important to consider a more direct form of benefit share model for hospital sites at a local level. Hospital sites may show greater willingness to commit resources and expertise to support commercial companies if they are able to directly benefit from the technology in return. Existing research databases have utilised these models through direct income share models, with hospital sites gaining income through commercial licensing of data and images. Collection sites sign an assignment agreement giving the study a non-exclusive licence to include the images and data in the research database and proportions of income shared is calculated based on contributions from the participating sites and the characteristics of the data submitted.[54]

NHSX is currently focusing on defining the terms of such benefit share models. The Life Sciences Industrial Strategy has set out guiding principles for sharing health

and care data that emphasises that data has to be used for the purpose of improving patient care and that any arrangements need to be clearly and transparently communicated.[55] Providing assurance to the public on the ethical basis for sharing data, involving them in a transparent dialogue and informing them of the direct public benefit are important strategies that need to be widely adopted.[56] The NHSX-led Centre for Improving Data Collaboration has been set up to provide direct support to the health and care sector to meet these guiding principles and establish frameworks for achieving fair value for the NHS.[57] Importantly, further research is needed to define the different types of benefit share models that should apply to commercial developers in the UK depending on the different use cases that the data is being shared for, for example, whether it is a public health emergency or a rare disease, and the intended scale of the deployment.[58]

Future aspirations

The experiences of the NCCID have resulted in valuable learnings to inform a national-level data collection approach for AI development in medical imaging. There are several areas that will need focus, which will be key to ensure that the benefits of data collection are realised through the implementation of AI technologies in real-world settings (see Table 4).

NHSX will be working closely with research groups, technology companies and NHS Trusts to progress the above focus areas through the work of the NHS AI Lab.

Conclusions

Establishing the NCCID during the COVID-19 pandemic has emphasised the challenges that exist when setting up national data collection infrastructure in the UK, but has also emphasised future opportunities for advancing the development and adoption of AI technologies in medical imaging. These opportunities relate to (a) improving access to high-quality data, (b) supporting impactful research objectives, and (c) facilitating better integration between the NHS and the entrepreneurial community to identify life-saving solutions. Robust and clear data governance structures are a requisite to support the legal and ethical basis for data collection. Automated data pipelines to build large-scale datasets for training and validating AI technologies will require national coordination and infrastructure expertise. TRE models will need to be enhanced through targeted support and closer integration to facilitate easier access to national datasets for research and technology development. All of this work will require consistent funding at a national level and support from the general public, along with patient advocacy groups, to realise impact. As one of the few countries to offer a publicly funded, single payer, nationwide healthcare system, the UK is optimally positioned to be a global leader in national data infrastructure for medical imaging to support AI development and deployment into clinical pathways and to demonstrate how care can be improved as a result.[8] Ultimately, the success of this endeavour will be dependent on national collaboration between NHS executives, clinicians, patients, researchers and technology companies, and the NHS AI Lab is well placed to coordinate these diverse stakeholders to drive this initiative forward.

If you work at a hospital site and are interested in contributing data to this initiative, please reach out directly to imaging@nhsx.nhs.uk.

If you are involved in research or technology development and would like to apply for access to the NCCID training dataset, please follow the instructions at this link: https://nhsx.github.io/covid-chest-imaging-database/.

Declarations of interest/conflicts

JCLR has received consultancy fees from NHSX for work completed on the NCCID.

CB and FG have received consultancy fees from NHSX for work completed on the NCCID.

NJS – nil.

JJ reports consultancy fees from NHSX for work on the NCCID and consultancy fees from Boehringer Ingelheim, Roche and GlaxoSmithKline unrelated to the submitted work.

Statement of contributorship

All authors have contributed significantly to the concept, design and writing of the work. All authors have been involved in drafting and revision, including all authors approving the article for submission.

Guarantor

Dominic Cushnan (dominic.cushnan@nhsx.nhs.uk)

NCCID collaborative:

Daniel Alexander, Melissa Alexander, Mr Nicholas Ashley, Hena Aziz, Dr Judith L Babar, Dr Ramona-Rita Barbar, Angela Bates, Oscar Bennett, Alison Bettany, Dr Bahadar Bhatia, Angela Bowen, Hannah Bown, Doreen Brookes, Stephen Buckle, Paul Burfield, Paul Burford, Sarah Cardona, Dr Harmeet Chana, Dr Madalina Chifu, Rachel Clark, Dr Jordan Colman, Hayley Connoley, John Curtin, Rachel Darrah, Mrs Janet Deane, Dr Dileep Duvva, Esme Easter, Dr Kieran Foley, Amy J Frary, Samantha Gan, Dr Nemi Gandy, Tara Ganepol, Avneet Gill, Fergus Gleeson, Natasha Greig, Peter Halson, Anisha Harrar, Rachael Harrison, Chris Heafey, Angela Heeney, Dr Benjamin Hudson, John Hurst, Zahida Hussain, Dr Mark Ingram,

Dr Leila Ismail, Mary Jones, Joanne Kellett, Rumi Kidwai, Dr Daniel Kim, Dr Fiona Kirkham, Dr Ross Kruger, Chinnoi Law, Francois Lemarchand, Emma Lewis, Dr Hannah Lewis, Gerald Lip, Katy Lomas, Berenice Lufton, Dr James MacKay, Peter Manser, Nigel Marchbank, Dr Giles Maskell, Liz Mathers, Violet Matthews, Dr Caroline McCann, Dr David McCreavy, Kate Milne, Jacqueline Monaghan, Annette Moreton, Andrew Moth, Dr Edward Neil-Gallacher, Dr Helen Oliver, Nnenna Omeje, Bethan Wyn Owen, Susan Palmer, Lisa Patterson, Dr Thomas Payne, Dr Emily Pearlman, Lindsay Van Pelt, Carla Pothecary, Philip Quinlan, Dawn Redwood, Rowena Reyes, Claire Ridgeon, Lisa Roche, Fiona Rotherham, Dr Timothy J Sadler, Dr Alexander Sanchez-Cabello, Anastasios Sarellas, Daniel Schofield, Simon Seal, Dr Aarti Shah, Dr Ban Sharif, Smita Shetty, Dr Melissa Sia, Dr Tze Siah, Marcelle de Sousa, Susanne Spas, James Sutherland, Andrew Swift, Dr Matthew Thorley, Joanna Tilley, Jenna Tugwell-Allsup, Keri Turner, Dr Debbie Wai, Gayle Warren, Janet Watkins, Dr Tom Welsh, Adele Wilson, Richard Wood, Dr Sarah Yusuf

ORCID iDs: Dominic Cushnan [ID] https://orcid.org/0000-0001-6168-327X
Gergely Imreh [ID] https://orcid.org/0000-0001-7892-0366
Gerry Reilly [ID] https://orcid.org/0000-0003-1427-3598
Susan Copley [ID] https://orcid.org/0000-0003-1856-9951
Hayley E Hardwick [ID] https://orcid.org/0000-0003-3421-6435

Notes

1. https://ieeexplore.ieee.org/abstract/document/8241753
2. https://www.nature.com/articles/s41467-020-18685-1
3. https://pubs.rsna.org/doi/full/10.1148/radiol.2020203511
4. https://www.nature.com/articles/s41467-020-18786-x
5. https://www.nature.com/articles/s41746-021-00453-0
6. https://www.nature.com/articles/s42256-021-00338-7
7. https://www.nature.com/articles/s42256-021-00307-0
8. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00039-X/fulltext
9. https://www.nature.com/articles/s41591-019-0447-x
10. https://www.bmj.com/content/369/bmj.m1328
11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6347576/
12. https://www.bmj.com/content/372/bmj.n304
13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6594166/
14. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233077/
15. https://link.springer.com/article/10.1007/s10994-020-05928-x
16. https://pubs.rsna.org/doi/10.1148/radiol.2020192224?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed
17. https://erj.ersjournals.com/content/56/2/2001809
18. https://www.nhsx.nhs.uk/ai-lab/
19. https://www.sciencedirect.com/science/article/pii/S0960977619305545
20. https://ico.org.uk/global/data-protection-and-coronavirus-information-hub/data-protection-and-coronavirus/health-social-care-organisations-and-coronavirus-what-you-need-to-know/#health3
21. https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information/coronavirus-covid-19-notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002-general
22. https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/
23. https://medphys.royalsurrey.nhs.uk/nccid/guidance/282365_REC_FO_Letter_17.04.2020.pdf
24. https://digital.nhs.uk/services/national-data-opt-out/compliance-with-the-national-data-opt-out#the-check-for-national-data-opt-outs-service-technical-solution
25. http://orca.cf.ac.uk/114102/1/432-Article%20Text-1967-2-10-20180921.pdf
26. https://www.hdruk.ac.uk/wp-content/uploads/2019/10/191031-Gateway-Phase-2-webinar-FINAL.pdf
27. https://www.healthdatagateway.org/
28. https://www.hdruk.ac.uk/wp-content/uploads/2021/04/Improving-UK-Health-Data-Impacts-from-Health-Data-Research-Hubs-v2.pdf
29. https://www.hdruk.ac.uk/covid-19/covid-19-national-core-studies/
30. https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf
31. https://www.nature.com/articles/s41746-020-00323-1
32. https://medical.sectra.com/case/national-network-for-image-exchange-connects-more-than-400-uk-sites/
33. https://medphys.royalsurrey.nhs.uk/omidb/
34. https://www.gov.uk/government/news/artificial-intelligence-to-help-save-lives-at-five-new-technology-centres
35. https://academic.oup.com/gigascience/article/9/10/giaa095/5912826
36. https://acmedsci.ac.uk/file-download/77652269
37. https://onlinelibrary.wiley.com/doi/full/10.1002/lrh2.10200
38. https://www.sciencedirect.com/science/article/pii/S0009926019301151
39. https://medphys.royalsurrey.nhs.uk/omidb/project-information/approval-ethics/

40. https://www.prospectstrial.org
41. https://www.researchgate.net/publication/330587099_The_National_Institute_for_Health_Research_making_an_impact_in_imaging_research
42. https://www.rcr.ac.uk/sites/default/files/clinical-radiology-uk-workforce-census-2019-key-findings.pdf
43. https://www.rcr.ac.uk/system/files/publication/field_publication_files/clinical-radiology-uk-workforce-census-2019-report.pdf
44. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6199205/
45. https://nhsx.github.io/covid-chest-imaging-database/
46. https://nhsx.github.io/covid-chest-imaging-database/project_summaries.html
47. https://www.gov.uk/government/organisations/national-data-guardian
48. https://www.adalovelaceinstitute.org/our-work/
49. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7385634/
50. https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology%23principle-10-define-the-commercial-strategy
51. https://onlinelibrary.wiley.com/doi/full/10.1002/lrh2.10191
52. https://www.clinicalradiologyonline.net/article/S0009-9260(19)30542-2/pdf
53. https://link.springer.com/article/10.1007/s10278-017-9976-3
54. OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data
    Mark D. Halling-Brown, Lucy M. Warren, Dominic Ward, Emma Lewis, Alistair Mackenzie, Matthew G. Wallis, Louise S. Wilkinson, Rosalind M. Given-Wilson, Rita McAvinchey, and Kenneth C. Young
    Radiology: Artificial Intelligence 0 0:ja
55. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf
56. https://www.bmj.com/content/368/bmj.l6927
57. https://www.nhsx.nhs.uk/key-tools-and-info/centre-improving-data-collaboration/
58. https://reform.uk/sites/default/files/2018-12/Making%20NHS%20data%20work%20for%20everyone%20WEB_1.pdf
59. https://isaric4c.net/
60. https://doi.org/10.1093/jamia/ocaa210
61. https://doi.org/10.1186/s12916-019-1426-2
62. https://www.rcr.ac.uk/system/files/rcr2016_image_sharing.pdf
63. https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/
64. https://www.informationgovernance.scot.nhs.uk/pbpphsc/wp-content/uploads/sites/2/2020/10/HSC-PBPP-Approved-COVID-19-applications-20201009-with-lay-summaries.pdf
65. https://www.emrad.nhs.uk/programmes/collaboration-and-transformation/398-transformation
66. https://improvement.nhs.uk/documents/6119/Transforming_imaging_services.pdf
67. https://yorkshireimaging.org/
68. https://www.nhsx.nhs.uk/media/documents/NHSX_A_Buyers_Guide_to_AI_in_Health_and_Care_KVjMIWP.pdf

References

1. Ker J, Wang L, Jai R, et al. Deep learning applications In medical image analysis. *IEEE Access* 2017; 6: 9375–9389.
2. Jin C, Chen W, Cao Y, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun* 2020; 11: 1–14.
3. Wehbe RM, Sheng J, Dutta S, et al. DeepCOVID-XR: an artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large US clinical data set. *Radiology* 2021 Apr; 299: E167–E176.
4. Feng Z, Yu Q, Yao S, et al. Early prediction of disease progression In COVID-19 pneumonia patients with chest CT and clinical characteristics. *Nat Commun* 2020; 11: 1–9.
5. Shamout FE., Shen Y., Wu N., et al. 2021. An artificial intelligence system for predicting the deterioration of COVID-19 patients In the emergency department. *NPJ digital medicine*, 4, pp.1–11.
6. DeGrave AJ, Janizek JD and Lee SI. AI For radiographic COVID-19 detection selects shortcuts over signal. *Nat Mach Intell* 2021; 3: 610–619.
7. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell* 2021; 3: 199–217..
8. Jiao Z, Choi JW, Halsey K, et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. *The Lancet Digital Health* 2021 May 1; 3: e286–e294. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00039-X/fulltext.
9. Ardila D, Kiraly A, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; 25: 954–961.
10. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *Br Med J* 2020; 369: m1328.
11. Gianfrancesco MA, Tamang S, Yazdany J, et al. Potential biases In machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018; 178: 1544–1547.
12. Leslie D, Mazumder A, Peppin A, et al. Does "AI" stand for augmenting inequality In the era of covid-19 healthcare? . *bmj* 2021; 372, https://www.bmj.com/content/372/bmj.n304.
13. Rajkomar A, Hardt M, Howell MD, et al. Ensuring fairness In machine learning to advance health equity. *Ann Intern Med* 2018; 169: 866–872.
14. Ghassemi M, Naumann T, Schulam P, et al. A review of challenges and opportunities In machine learning for health. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science* 2020; 191–200. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233077/.
15. van der Schaar M, Alaa AM, Floto A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Mach Learn* 2021; 110: 1–14.
16. Willemink M, Koszek W, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology* 2020; 295: 4–15.
17. Jacob J, Alexander D, Baillie JK, et al. Using imaging to combat a pandemic: rationalefor developing the UK national

COVID-19 chest imaging database. *Eur Respir J* 2020; 56: 2001809.

18. NHSX. *National COVID-19 Chest Imaging Database (NCCID)*. https://www.nhsx.nhs.uk/covid-19-response/data-and-covid-19/national-covid-19-chest-imaging-database-nccid/ [Accessed 07 December 2020].

19. NHSX. *NCCID Collaborative*. https://nhsx.github.io/covid-chest-imaging-database/NCCID_Collaborative.html (2020, Accessed 07 December 2020).

20. Information Commissioner's Office. *Health, social care organisations and coronavirus - what you need to know*. https://ico.org.uk/global/data-protection-and-coronavirus-information-hub/data-protection-and-coronavirus/health-social-care-organisations-and-coronavirus-what-you-need-to-know/#health3 (2020, Accessed 08 December 2020).

21. UK Gov. *Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – general*. https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information/coronavirus-covid-19-notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002-general (2020, Accessed 08 December 2020).

22. Health Research Authority. *Confidential Advisory Group*. https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/ (2021, Accessed 08 December 2020).

23. Health Research Authority. *Approval letter from Research Ethics Committee for National COVID-19 Chest Imaging Database*. https://medphys.royalsurrey.nhs.uk/nccid/guidance/282365_REC_FO_Letter_17.04.2020.pdf (2020, Accessed 08 December 2020).

24. NHS Digital. *Compliance with the national data opt-out*. https://digital.nhs.uk/services/national-data-opt-out/compliance-with-the-national-data-opt-out#the-check-for-national-data-opt-outs-service-technical-solution (2021, Accessed 08 December 2020).

25. Lugg-Widger FV, Angel L, Cannings-John R, et al. Challenges In accessing routinely collected data from multiple providers In the UK for primary studies: managing the morass. *International Journal of Population Data Science* 2018; 3, https://doi.org/10.23889/ijpds.v3i3.432.

26. ISARIC 4C. *ISARIC4C (Coronavirus Clinical Characterisation Consortium)*. https://isaric4c.net/ (2020, Accessed 08 December 2020).

27. Röösli E, Rice B and Hernandez-Boussard T. Bias at warp speed: how AI may contribute to the disparities gap In the time of COVID-19. *J Am Med Inform Assoc* 2020; ocaa210. https://doi.org/10.1093/jamia/ocaa210.

28. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17: 1–9.

29. Royal College of Radiologists. *It's good to share: medical image and report exchange between UK healthcare providers*. https://www.rcr.ac.uk/system/files/rcr2016_image_sharing.pdf (2016, Accessed 08 December 2020).

30. Scottish Government, Health and Social Care. *Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics*. https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/ (2015, Accessed 08 December 2020).

31. HSC Scotland. *HSC-PBPP APPROVED COVID-19 applications: 09 October 2020*. See application reference 2021-0018. https://www.informationgovernance.scot.nhs.uk/pbpphsc/wp-content/uploads/sites/2/2020/10/HSC-PBPP-Approved-COVID-19-applications-20201009-with-lay-summaries.pdf (2020, Accessed 08 December 2020).

32. HDR UK. *An introduction to the Innovation Gateway Technology Partnership*. https://www.hdruk.ac.uk/wp-content/uploads/2019/10/191031-Gateway-Phase-2-webinar-FINAL.pdf (2020, Accessed 08 December 2020).

33. Health Data Research Innovation Gateway. https://www.healthdatagateway.org/ [Accessed 08 December 2020].

34. HDR UK. Improving UK Health Data: Impact from Health Data Research Hubs. https://www.hdruk.ac.uk/wp-content/uploads/2021/04/Improving-UK-Health-Data-Impacts-from-Health-Data-Research-Hubs-v2.pdf (2021, Accessed 17 August 2021).

35. HDR UK. *COVID-19 National Core Studies*. https://www.hdruk.ac.uk/covid-19/covid-19-national-core-studies/ (2020, Accessed 08 December 2020).

36. UK Health Data Research Alliance. *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data science needs p1-31*. https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf (2020, Accessed 08 December 2020).

37. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ digital medicine* 2020; 3: 1–7.

38. Sectra. *National network for image exchange connects more than 400 UK sites*. https://medical.sectra.com/case/national-network-for-image-exchange-connects-more-than-400-uk-sites/ (2020, Accessed 08 December 2020).

39. OPTIMAM. *OPTIMAM Mammography Imaging*. https://medphys.royalsurrey.nhs.uk/omidb/ (2020, Accessed 08 December 2020).

40. UK Gov. *Artificial Intelligence to help save lives at five new technology centres*. https://www.gov.uk/government/news/artificial-intelligence-to-help-save-lives-at-five-new-technology-centres (2018, Accessed 08 December 2020).

41. Nind T, Sutherland J, McAllister G, et al. An extensible big data software architecture managing a research resource of real-world clinical radiology data linked to other health data from the whole scottish population. *GigaScience* 2020; 9 p.giaa095. https://doi.org/10.1093/gigascience/giaa095.

42. EMRAD - East Midlands Imaging Network. *Use of artificial intelligence (or computer-assisted technology)*. https://www.emrad.nhs.uk/programmes/collaboration-and-transformation/398-transformation (2020, Accessed 08 December 2020).

43. NHS England and NHS Improvement. *Transforming imaging services in England: a national strategy for imaging networks*. https://improvement.nhs.uk/documents/6119/Transforming_imaging_services.pdf (2020, Accessed 08 December 2020).

44. Yorkshire Imaging Collaborative. *Mission*. https://yorkshire-imaging.org/ (2020, Accessed 08 December 2020).

45. The Academy of Medical Sciences. *Artificial Intelligence and Health: Summary report of a roundtable held on 16 January*

2019. https://acmedsci.ac.uk/file-download/77652269 (Accessed 08 December 2020).

46. Tully MP, Hassan L, Oswald M, et al. Commercial use of health data—A public "trial" by citizens' jury. *Learn Health Sys* 2019; 3: e10200.

47. Ho CWL, Soon D, Caals K, et al. Governance of automated image analysis and artificial intelligence analytics In healthcare. *Clin Radiol* 2019; 74: 329–337.

48. OPTIMAM. *Approval & Ethics.* https://medphys.royalsurrey. nhs.uk/omidb/project-information/approval-ethics/ (2019, Accessed 09 December 2020).

49. PROSPECTS. *About the Prospects Trial.* https://medphys. royalsurrey.nhs.uk/prospects/trial-inforamtion/about/ (2020, Accessed 09 December 2020).

50. Goh V, Smye SW, Allen L, et al. The national institute for health research: making an impact In imaging research. *Clin Radiol* 2019; 74: 242–246.

51. The Royal College of Radiologists. *Radiologist shortages are negatively impacting clinical care.* https://www.rcr.ac.uk/ sites/default/files/clinical-radiology-uk-workforce-census-2019-key-findings.pdf (2019, Accessed 09 December 2020).

52. The Royal College of Radiologists. *Clinical radiology UK workforce census 2019 report.* https://www.rcr.ac.uk/system/ files/publication/field_publication_files/clinical-radiology-uk-workforce-census-2019-report.pdf (2019, Accessed 09 December 2020).

53. Pesapane F, Codari M and Sardanelli F. Artificial intelligence In medical imaging: threat or opportunity? Radiologists again at the forefront of innovation In medicine. *Eur Radiol Exp* 2018; 2: 35.

54. NHSX. *National COVID-19 Chest Image Database (NCCID).* https://nhsx.github.io/covid-chest-imaging-database/ (2020, Accessed 09 December 2020).

55. NHSX. *NCCID project summaries.* https://nhsx.github.io/ covid-chest-imaging-database/project_summaries.html (2020, Accessed 09 December 2020).

56. UK Gov. *National Data Guardian.* https://www.gov.uk/ government/organisations/national-data-guardian (2020, Accessed 09 December 2020).

57. Ada Lovelace Institute. *Our work.* https://www.adalovelaceinstitute. org/our-work/ (2020, Accessed 09 December 2020).

58. Gao S, He L, Chen Y, et al. Public perception of artificial intelligence In medical care: content analysis of social Media. *J Med Internet Res.* 2020; 22: e16649.

59. UK Gov. *Code of conduct for data-driven health and care technology* (see *Principle 10 - Define the commercial strategy*) https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology#principle-10-define-the-commercial-strategy (2020, Accessed 09 December 2020).

60. Ford E, Boyd A, Bowles JKF, et al. Our data, our society, our health: a vision for inclusive and transparent health data science In the United Kingdom and beyond. *Learning Health Systems* 2019; 3: e10191.

61. Gilbert FJ, Smye SW and Schonlieb C-B. Artificial intelligence In clinical imaging: a health system approach. *Clin Radiol* 2020; 5: 3–6..

62. Kohli MD, Summers RM and Geis JR. Medical image data and datasets In the era of machine learning—whitepaper from the 2016 C-MIMI meeting dataset session. *J Digit Imaging* 2017; 30: 392–399..

63. Halling-Brown MD, Warren LM, Ward D, et al. OPTIMAM Mammography image database: a large-scale resource of mammography images and clinical data. *Radiology: Artificial Intelligence* 2020; 3: e200103.

64. Bell J. *Life sciences industrial strategy.* Office for Life Sciences, 2017, pp 1-75. https://assets.publishing.service. gov.uk/government/uploads/system/uploads/attachment_ data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf [Accessed 09 December 2020].

65. Vollmer S, Mateen BA, Bohner G, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *Br Med J* 2020; 368, https://doi.org/10.1136/ bmj.l6927.

66. NHSX. *The Centre for Improving Data Collaboration.* https:// www.nhsx.nhs.uk/key-tools-and-info/centre-improving-data-collaboration/ [Accessed 09 December 2020].

67. Harwich E and Lasko-Skinner R. *Making NHS data work for everyone.* REFORM, 2018, https://reform.uk/sites/default/ files/2018-12 /Making%20NHS%20data%20work%20for% 20everyone%20WEB_1.pdf [Accessed 09 December 2020].

68. Joshi I and Cushnan D. *A Buyer's Guide to AI In health and care.* NHSX, 2020, https://www.nhsx.nhs.uk/media/ documents /NHSX_A_Buyers_Guide_to_AI_in_Health_ and_Care_KVjMIWP.pdf [Accessed 09 December 2020].