Delivered by
Innovate UK and MRC

Data to Early Diagnosis and Precision Medicine Challenge:

Interoperability Recommendations

September 2023

Contents

Executive summary

The UK is well placed to play a major role in the revolution in healthcare being driven by the digitisation of pathology and the adoption of machine learning to diagnose disease in very early stages. The scale and diversity of data held across NHS Trusts provides the UK with a unique asset that is fundamental to the creation of AI diagnosis algorithms that are accurate across all patient demographics.

For AI developers to successfully create, train, validate and deploy these algorithms in UK settings, the landscape needs to be as 'interoperable' as possible. This means that developers can access and interact with data from any manufacturers' scanner, held in any archive and deploy their software within any NHS Trust without barriers caused by different technologies or policies.

The Digital Pathology, Radiology and Machine Learning Centres of Excellence established by the Data to Early Diagnostics and Precision Medicine (D2EDPM) Challenge, delivered by Innovate UK and the Medical Research Council, established a UK infrastructure across academia, NHS Trusts and industry to create AI algorithms between January 2019 and March 2023. The experience of building and operating this infrastructure, including the successful completion of dozens of exemplar projects, has informed the work of seven interoperability working groups that span the AI development cycle.

Broadly, the work falls into two areas:

- The standardisation of digital pathology processes to align them to the maturity of digital radiology practices established over many years.
- The development of AI algorithms based on both digital radiology images and / or digital pathology images within secure data environments.

These groups addressed seven high-level questions:

High level question	Workgroup	Pg
Will pathology images contain variation?	QA in digital pathology	17
Will digitisation of images increase variation?	Image standards for clinical use	22
How do we work with huge files in the cloud ?	Next generation file formats	27
How do we ensure patient privacy?	De-identification of data	32
Where is a safe and workable place to do this?	Approach to SDEs	37
How do we train AI across diverse patient datasets?	Federated learning	42
How do we know our algorithms really work?	Evaluation of AI algorithms	47

The following high-level recommendations have been developed through each workstream through first-hand feedback from those involved in algorithm creation, and practical experimentation:

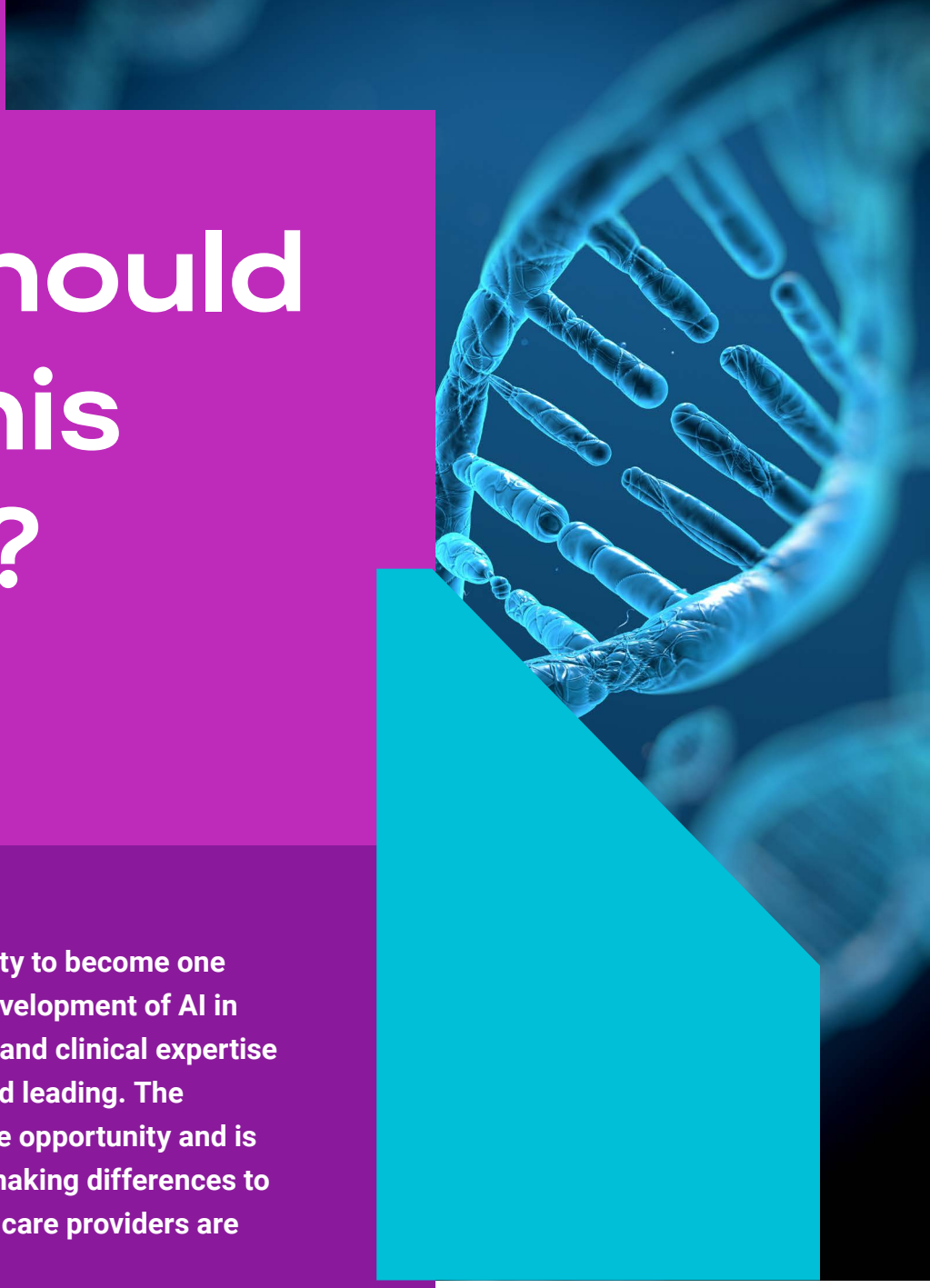
1. Establish guidelines for digital pathology lab processes utilising standardisation and normalisation concepts (Recommendations – pg 21).
2. Widen the adoption of DICOM standards to ensure digital pathology images are consistently codified across all input devices procured by NHS Trusts (Recommendations – pg 26).
3. Adopt a new file format to facilitate the processing of huge digital pathology image and metadata files across the cloud (Recommendations – pg 31).
4. Policy makers should ensure effective de-identification policies and technologies are applied consistently across all SDEs (Recommendations – pg 35).
5. NHS data owners should provide a high-quality managed service to allow algorithm developers to interact with patient data securely (Recommendations – pg 41).
6. Federated Learning has been shown to be technically feasible across two jurisdictions (England and Scotland). The objectives for future FL projects should align to NHS strategic priorities and provided with long term funding to give companies confidence to invest further (Recommendations – pg 46).
7. Policy makers should impose an evaluation framework for AI algorithms to ensure a safe and effective approach is followed by all developers (Recommendations – pg 51).

The work completed by the interoperability working groups spans the full life cycle of AI development from data gathering through to the evaluation of algorithms for clinical use and ultimately post-launch market surveillance.

Most of the issues faced by AI developers are centred on policy considerations rather than technical barriers.

To ensure that the NHS, its patients, and the UK economy benefit from the revolution taking place across healthcare, policy makers must adopt the above recommendations and intensify their engagement with AI development communities to comprehend and address the issues as they continue to evolve.

Who should read this report?



The UK has a unique opportunity to become one of the leading nations in the development of AI in healthcare. The research base and clinical expertise in the UK is recognised as world leading. The diagnostic base has risen to the opportunity and is developing solutions that are making differences to NHS patients and global healthcare providers are adopting their products.

However, large scale success and adoption of solutions cannot be assumed. It will take the collaborative efforts of all stakeholders within the sector to ensure that the respective needs of patients, NHS Trusts, and AI developers are met.

This report summarises the experience of the Centres of Excellence and is informative for the following groups:

- **Regulators:** The UK has an opportunity post-Brexit to create its own regulatory framework for the approval of healthcare AI, focused on safety, agility and efficiency. The UK should position this framework as a global benchmark that could be adopted by other countries. This would support UK companies to have their products adopted within both the NHS and export markets.

- **Data-controllers:** Agreeing common approaches to information governance across NHS Trusts would greatly simplify the landscape for developers, in addition to saving Trusts resource in re-inventing duplicative approaches.
- **Procurement and standards bodies:** Adopting nationally recognised standards as early as possible will future-proof NHS hardware and software implementations and allows for easier substitution of alternative provider technologies.
- **NHS Strategic Planning:** Interoperable architecture and protocols will deliver consistent, high-quality outcomes for patients. It also supports national workforce training protocols.
- **AI developers:** The experience of the 'first wave' of innovators in the UK health sector will inform those that follow, both in terms of where to focus resource and how long to expect the process steps to take.
- **NHS Trusts:** The operational teams within NHS Trusts need to know how to evaluate the effectiveness of AI solutions offered by vendors and form a deep understanding of how clinical pathways will be redesigned to successfully embed AI to provide earlier, more accurate diagnoses.
- **Hardware manufacturers:** Establishing common standards enables manufacturers to plan their future product strategies with certainty to minimise wasted costs and post implementation adaption.

Background

The Centres of Excellence in Digital Pathology, Digital Radiology and Machine Learning were established in 2019 by Innovate UK in response to the "ISCF Wave 2 Theme: Data to Early Diagnosis & Precision Medicine" competition. The competition was established in light of the market opportunity presented by the emerging revolution in healthcare presented by advancements in early diagnostics arising from developments in machine learning.

The objective of both the D2EDPM Challenge and the Centres of Excellence, was to build a network across the UK that could establish an innovation ecosystem to pave the way for the UK to be a global leader in this field.

The D2EDPM Challenge, brought together NHS Trusts, academic institutions, global diagnostic industrials and UK SMEs, to establish ways to safely utilise digitised patient data to develop advanced AI algorithms and software. These capabilities could help to detect disease far earlier than previously possible.

To fulfil the ambition to establish the UK as a leader in this field, it was recognised and articulated within the competition that solutions developed by the Centres should be easily navigable and accessible by all stakeholder groups across the UK and globally. In practical terms, this meant that solutions were vendor agnostic, images could be shared across platforms, and algorithms developed with one NHS Trust could be shared across all NHS Trusts.

At the heart of this ecosystem is the need for all stakeholders to be able to navigate across NHS Trusts and research databases easily without barriers caused by differing technology solutions, policies, and processes. The high-level requirements for each stakeholder group are depicted in the following graphic.

Challenges

- Regulatory control & assurance
- Patient concern & privacy
- Shared benefit & commercialisation
- Clinical acceptance & adoption
- Transformation of clinical process & practice

Technical challenges

- Availability of sufficient data for dev/test/validation
- Standards - many, inadequate, variably adopted
- Interoperability - variation across systems and data types
- User-centred design and integration into workflow

Figure 1: The vision for the UK healthcare AI ecosystem

Approach

The approach taken by the interoperability workstream of the D2EDPM Challenge followed the steps below:

- **Definition of Interoperability.**
- **Validation of the definition through industry research.**
- **Formation of workstreams to focus on key areas of concern.**
- **Detailed work within each workgroup to build on experience and expertise to provide recommendations for policy makers and future innovators.**

To prioritise efforts to focus on the areas of greatest need, the first phase of activity involved an assessment of where stakeholders across the Centres were experiencing the most issues or concerns in their development of algorithms across the network.

Definition of Interoperability

Interoperability means different things to each stakeholder group operating in the environment. The first step was to understand this complexity.

The Healthcare Information and Management Systems Society, Inc. (HIMSS) definition of interoperability provided a useful starting point:

Interoperability is the ability of different information systems, devices or applications to connect, in a coordinated manner, within and across organisational boundaries to access, exchange and cooperatively use data amongst stakeholders, with the goal of optimising the health of individuals and populations.

Health data exchange architectures and standards allow relevant data to be shared effectively and securely across the complete spectrum of care, within all applicable settings and with relevant stakeholders (including with the person whose information is being shared).

However, to break down the problem further, a detailed matrix was developed that explained the levels of interoperability (the stack) against the stages of the AI development journey:

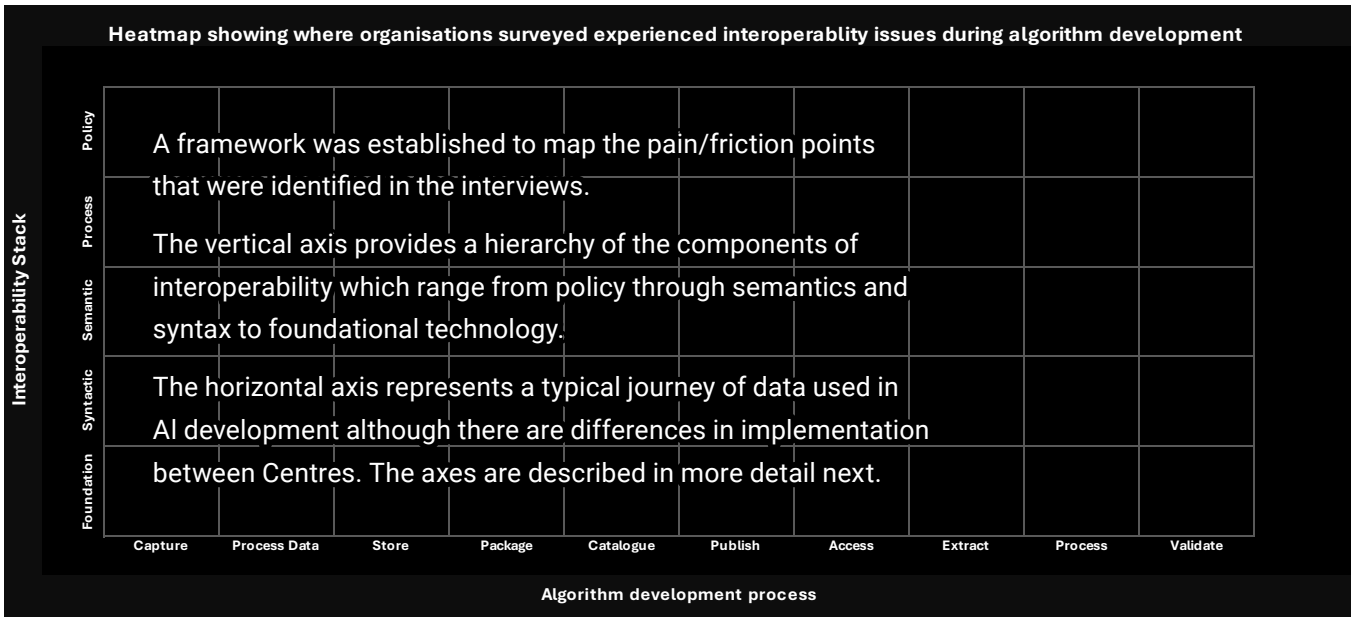


Figure 2: Interoperability framework

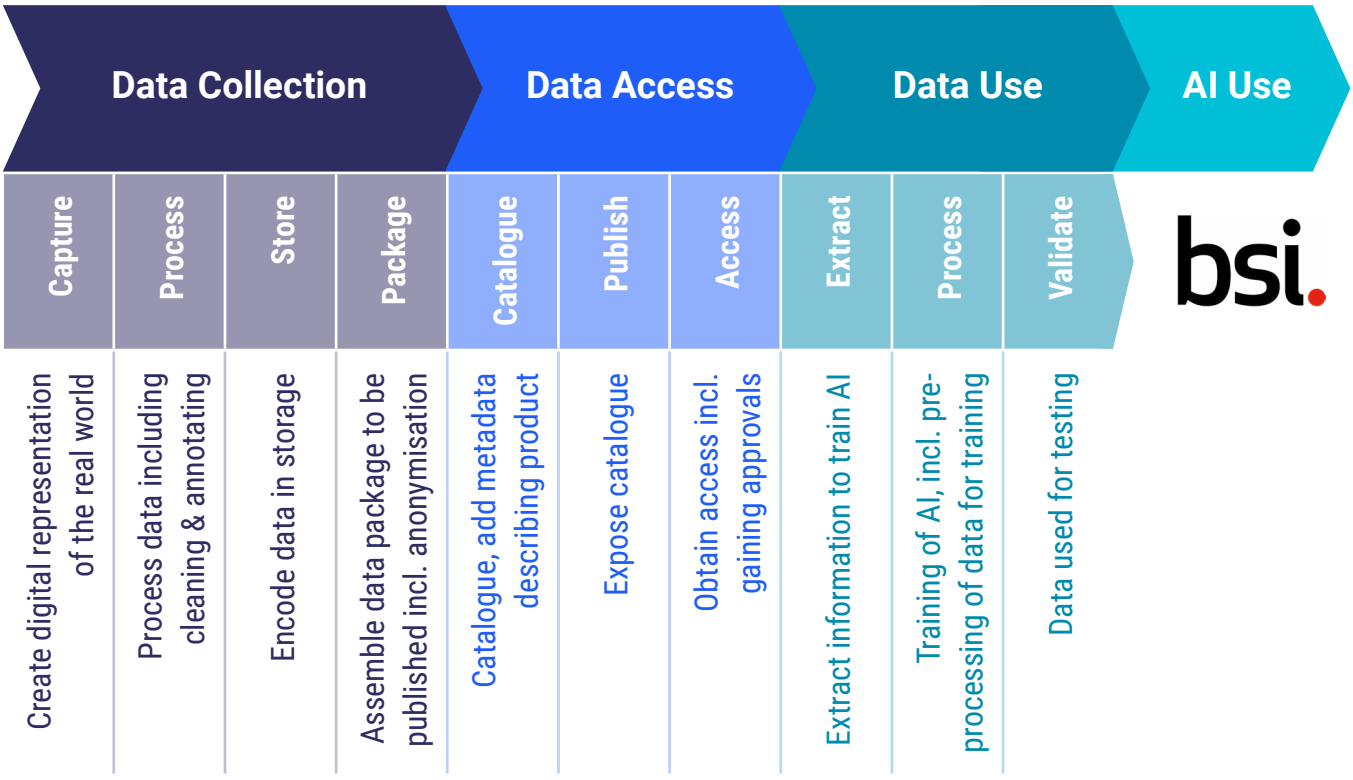


Figure 3: Description of AI algorithm development stages

Level	Description
Policy	Encompasses the clear policy, social and organisational components. These components facilitate the secure, seamless and timely communication and use of data within and between organisations and individuals. Inclusion of these non-technical considerations enables interoperability that is integrated into end-user processes and workflows in a manner that supports efficiencies, relationships and overall health and wellness through cooperative use of shared data both across and within organisational boundaries.
Process	• Functional standards (procedures, checklists, organisational rules to manage information for a specific use case) • Workflow standards (functional requirement analysis, evaluation of user needs) • Business process standards (practice standards, clinical pathways) • Safety standards (classification, risk mitigation)
Semantic	The ability of two or more systems to exchange information and to interpret and use that information. Semantic interoperability takes advantage of both the structuring of the data exchange and the codification of the data, including standard, publicly available vocabulary, so that the receiving information management systems can interpret the data. Semantic interoperability supports the electronic exchange of patient data and information among authorised parties via potentially disparate health information and technology systems and products to improve quality, costs, safety, efficiency, experience and efficacy of healthcare delivery.
Syntactic	Defines the structure or format of data exchange (i.e., the message format standards) where there is uniform movement of healthcare data from one system to another such that the clinical or operational purpose and meaning of the data is preserved and unaltered. Structural interoperability defines the syntax of the data exchange. It ensures that data exchanges between information technology systems can be interpreted at the data field level.
Foundation	Develops the building blocks of information exchange between disparate systems by establishing the inter-connectivity requirements needed for one system or application to share data with and receive data from another. It does not outline the ability for the receiving information technology system to interpret the data without interventions from the end user or other technologies.

Figure 4: HIMSS definition of interoperability levels

This provided a useful framework to discuss with stakeholders to pinpoint where their experience during the early stage of the D2EDPM Challenge had shown that problems had arisen.

Research: Industry input

Against the framework outlined in the previous chapter, the key stakeholders were interviewed to understand where they had experienced issues in their development of AI algorithms.

All the global diagnostics companies investing in the programme were surveyed, in addition to many of the SMEs, and each of the central programme teams across the Centres of Excellence.

Interviews were held with a range of stakeholders revealing a wide set of shared challenges:

- Participants were asked questions based on their role in the data journey. Data controller, consumer of data, aggregator of data, national body.
- Questions focused on both issues across the data journey from image creation through to use in clinical practice.
- Friction points were extracted from the interview notes and referenced to an Interoperability Matrix.
- Comparing and contrasting different stakeholder types provides insight on the relative impact.

Small & medium enterprises	National bodies	Multinational industry	Centres of excellence
• K Heiron • Caristo • Sectra • Glencoe Software • Hetero Genaus • Bering	• HDR-UK • NHS-England • Gov.uk • NIHR • Royal College of Radiologists • Royal College of Pathologists	• Canon Medical Research Europe Ltd • Siemens-Healthineers • Leica Biosystems • Roche Diagnostics Limited • Philips • GE Healthcare	• Path Lake • National Consortium for Intelligent Medical Imaging (NCIMI) • Industrial Centre for Artificial Intelligence Research in Digital Diagnostics (iCAIRD) • NPIC • London Medical Imaging & AI Centre for Value-Based Healthcare

The key findings are shown below.

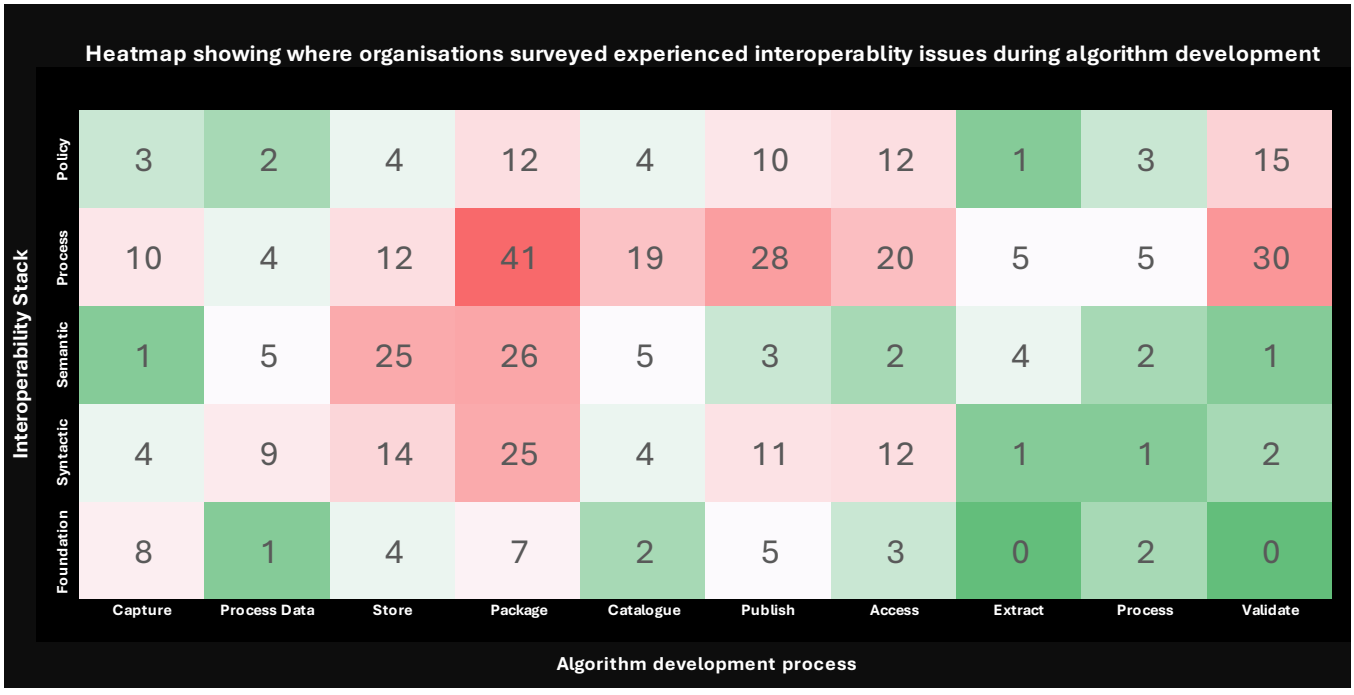


Figure 5: Industry concerns with interoperability

The outputs provided an informative picture of where the issues were concentrated for further exploration.

Workstream overview

The key questions spanned seven areas across the algorithm development journey as shown below.

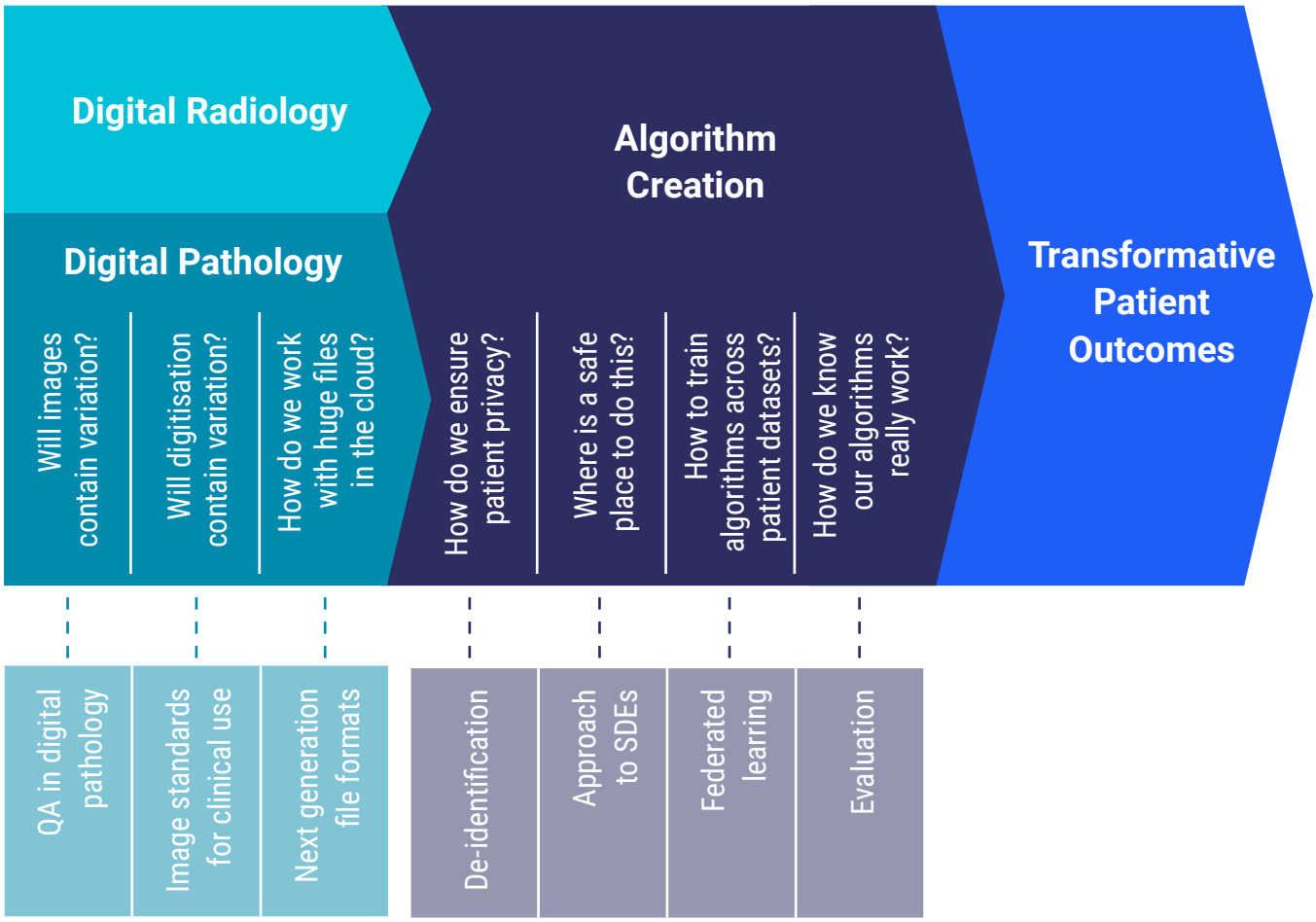


Figure 6: Workstream groups mapped to the algorithm development process

Through dialogue with the Centres, the team established seven workstreams that were focused on these issues as shown above mapped to the question areas.

Objectives were then agreed for the workstreams to direct the work for the remaining time of the D2EDPM Challenge.

The sections in the body of this report describe the work undertaken by the workstreams, their findings, insights, and recommendations.

Quality assurance of digital pathology

Background

The histopathology process has not changed significantly since it was first established in the 1800s. The process of fixing tissue, sectioning onto the slide, staining with histological stain and viewing the tissue on a microscope will be as recognisable then as it is now. The processes for quantifying the quality of each of the stages has also changed little. Clinical validation processes and schemes do exist, but physical quantification of the variability at each stage is largely subjective. To date, variability has been well tolerated by pathologists as humans are adept at coping with variation. Different laboratories also have effective 'signature' stain protocols, which can vary widely but are often readily recognisable especially within a regional setting. In digital pathology however the digital image loses this local context with images effectively becoming orphaned from the parent institution where they were generated upon being uploaded into a repository and shared wider.

The greater the variability of data, the more resilient algorithms must become in order to give consistent results and invariably to train resilient algorithms requires larger, more expansive, data sets. These algorithms would also be more portable to other institutions if they had a commonality of data consistency. In NPIC we have taken the stance that addressing issues of quality and consistency throughout the imaging chain will improve the quality of not only the digital data but also the resulting AI products. The interoperability work offers us a valuable platform to develop, test and scale at pace our quality solutions across the UK. This not only provides a realistic diversity of results to evolve our offerings, but also promulgates our ethos of 'quality in equals quality out' and we believe this quality ethos will drive up the 'value' of the UK digital pathology dataset. Ultimately, we believe the UK should lead the world in setting quality standards for all aspects of digital pathology and the interoperability work is an important step in realising this vision.

Objectives

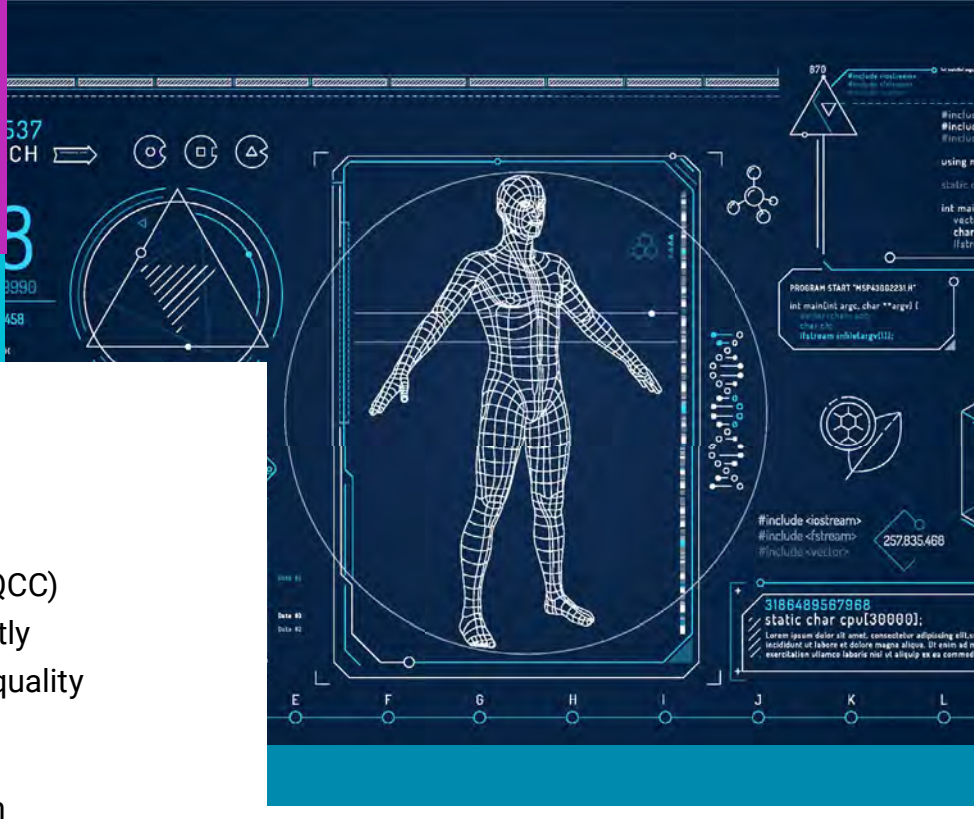
The main objective of the quality workstream is:

‘to establish a resilient quality network in digital histopathology, and equip it with relevant knowledge, skills, support and tools, to drive up quality in digital histopathology in the UK’.

At the outset we set out five clear requirements to deliver the objective:

1. Establish a national digital pathology quality coordination centre within NPIC with the aim of:
    - Developing digital pathology QA tools.
    - Providing expertise on digital histopathology quality issues.
    - Education and dissemination of best practice.
    - Establishing recommendations and standards.
2. Create a network of “Whole Slide Imaging” (WSI) quality co-ordinators across the UK.
3. Co-ordinate dissemination of information relating to WSI quality control (e.g. links to the latest product launches, research publications and white papers).
4. Assessment of inter and intra colour variability within scanners across the network.
5. Assessment of hematoxylin and eosin (H&E) staining variability within the network.

Objectives 3-5 were dependent on the successful completion of objectives 1 and 2.



Results

- A Quality Co-ordination Centre (QCC) has been established and currently 16 laboratories have nominated quality leads to work with the QCC.
- Education and training have been provided through webinars and frequent presentations at national and international conferences and our first newsletter is imminent. The impact of awareness of quality factors has been assessed and it has been shown education is changing attitudes to quality.
- New physical quality control tools have been developed and are at various stages of deployment:
  - The commercially available Sierra test tool for characterising of the colour profile of whole slide scanners will be circulated to the participating centres. (ffei.ai). with a distribution start-date of late February 2023.
  - A display quality assurance tool was launched during COVID to support home reporting and has been used over 3,500 times with an approximate 8% failure rate. (www.virtualpathology.leeds.ac.uk/research/systems/pouqa/pathology/) A screenshot of the web-based user interface is shown in Figure 7.
  - A new test tool, called Tango, allows the quantification of histological stains, it has been developed at NPIC and is in production ‘ramp-up’ to allow distribution to participating QCC centres, with a distribution start-date of late February 2023.
  - A Tango ‘pre-study evaluation’ of eight NPIC partner laboratories indicated a 30% reduction in variability between laboratories could be achieved by simply setting a standard target intensity for staining. Figure 8 is a representation of the wide range of stain colours measured by Tango from this pre-study.
  - Other physical test tools, not detailed here, which evaluate other elements of scanner performance, are also intended to be circulated along-side the stain assessment study. Allowing a unique opportunity within out network to measure the variation introduced across multiple parts of the Whole Slide Imaging pathway across the UK.

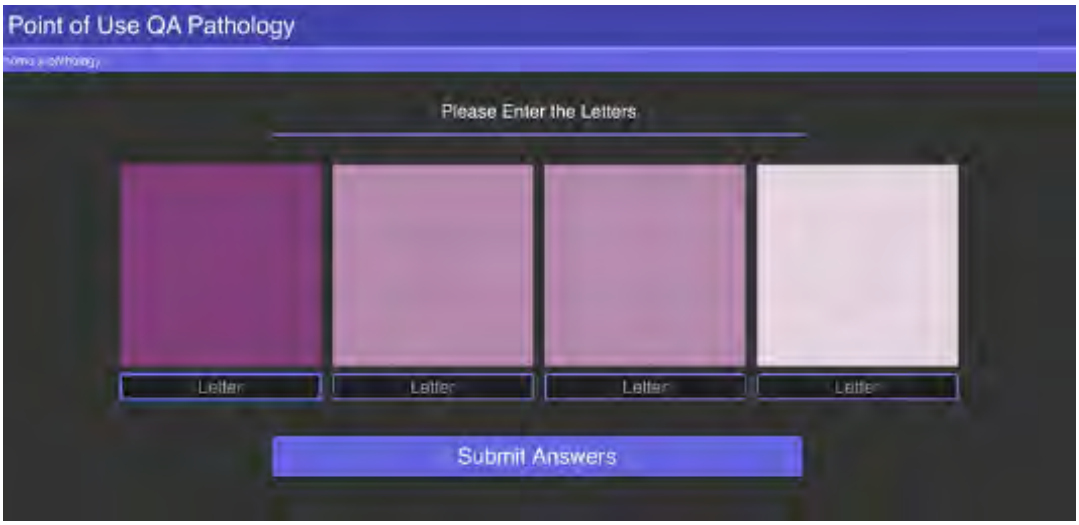


Figure 7: Point of Use QA is a simple point-of-use quality assurance tool for ensuring your display device allows you to see the minimum amount of visual contrast between colours, for scoring digital pathology slides. Data collected from this suggests an approximate 8% failure rate. The colours tested represent the most common pathological stain, H&E; pinks and purples

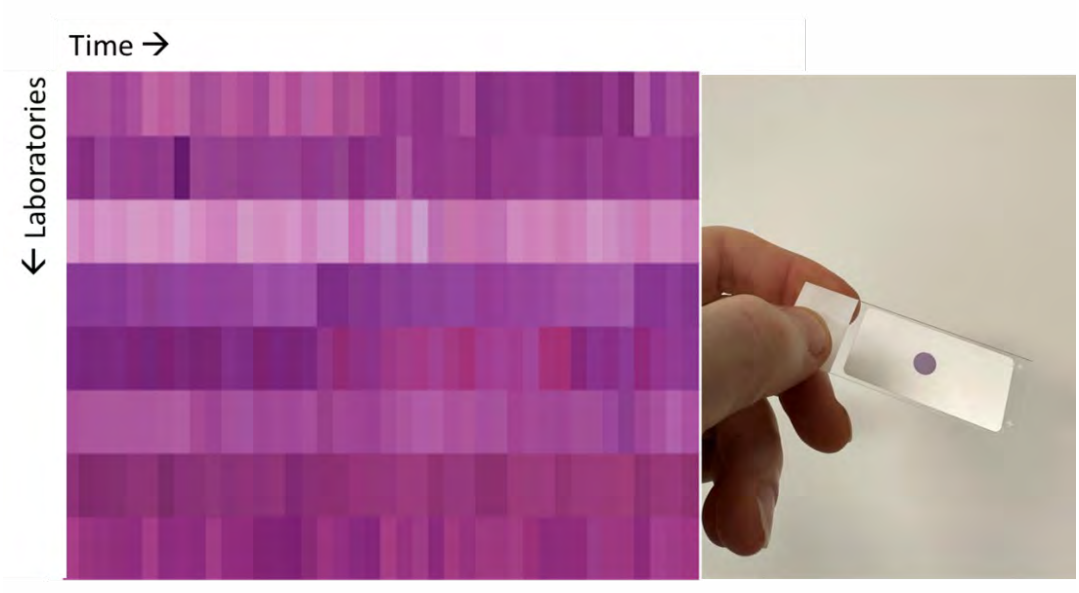


Figure 8: The Tango test object (shown on the right) can be used for quantification of histological stains including H&E, this figure shows a representation of the wide range of colours that the Tango test object was stained when used across 8 different NHS laboratories over time in a pre-study evaluation.

Discussions

At the outset of this work there was little 'buy-in' for quality control in digital histopathology. This was confounded by the human tolerance of variation and the belief that 'AI will sort it all out'. However, as the project has continued attitudes to quality have changed. AI developers have been struggling with variability and this is impacting on the development and portability of algorithms. Clinical awareness of quality has also improved, we believe partly due to the work of NPIC, and recently we were awarded the clinical best poster award for Tango at an international digital pathology conference. The new quality tools we have developed are also gaining commercial interest and we soon hope to be able to establish a long-term outlet for them.

We believe that not only have we highlighted the need for quality controls in digital histopathology, but we have also established a viable route to sustain this through the QCC, and developed practical test tools to deliver quality metrics which ultimately can be used to establish standards.

Recommendations

We believe we have demonstrated the need, and value, of physical quality control in digital histopathology and have created a viable platform to continue this work. We recommend this work is prioritised and extended to deliver a national physical quality programme underpinning a high-quality UK digital pathology dataset.

We also recommend national quality standards are established with the involvement of manufacturers, professional bodies, and users.

Next Steps

Next step priorities are to complete the national study of stain and scanner variation, disseminate the results and lead on establishing physical quality standards in the UK. In addition, opportunities for the long-term sustainability of the quality work need to be explored.

Image standards for clinical use

Background

Currently there is a generally low uptake of digital pathology in the NHS but also an increasing need for high quality data (both digital images and associated metadata) across multiple sites to improve diagnostics and workflows and help address staff shortages in NHS histopathology departments. When procuring and implementing digital pathology systems for clinical use the importance of interoperability cannot be overstated, with the aim of avoiding isolated “silos” of data and the ability to independently change or upgrade individual elements of a whole slide imaging (WSI) system to meet any future requirements. Interoperability, at a regional and national level, is also vital in realising the full potential of digital pathology, for use cases such as referrals, interpretation by geographically distant subspecialists (e.g.,

national specialist tumour networks) and regional workload distribution. Many parallels can be drawn (and lessons learnt) from the digitisation of radiology services in the NHS, which now enjoy a seamless mixture of different brands and models of equipment and viewing / analysis software, allowing the most appropriate tools to be procured to meet the specific needs of departments ("best of breed" approach).

Digital Imaging and Communications in Medicine (DICOM) is the international standard for medical images and related information, being the recognised standard for radiology and many other enterprise medical imaging modalities. The DICOM standard format for WSI has been available since 2010. The National Pathology Imaging Co-operative (NPIC) aims to create a standards-based vendor neutral national digital pathology system and digital pathology archive for the whole NHS. From the outset of the project, NPIC industry partners were engaged and committed to the need for standardisation. DICOM compliance was stated as an entry criterion for NPIC, following our mantra of "any scanner, any software, any AI tool".

Since much of the metadata relevant to the interpretation of WSI (whether it be by human or machine) is managed in the Laboratory Information and Management System (LIMS), standardised interoperability between LIMS and slide scanners is essential. The Integrating the Healthcare Enterprise (IHE) organisation profiles existing standards like Health Level Seven (HL7) and DICOM to define transactions across this interoperability boundary. The result is to assure that scanners produce standard images that are identifiable, recognisable, and understandable.

Aims

The primary aims of interoperability within NPIC are as follows:

- Working with vendors to enable native production of DICOM standard images by scanners.
- Switchover from proprietary format images to DICOM standard images within the vendor neutral archive.
- Integration of scanners with LIMS to provide a richer set of patient / case and specimen / slide identification and preparation attribute values.
    - Deploy the IHE Pathology and Laboratory Medicine (PaLM) Digital Pathology Image Acquisition (DPIA) integration profile.
    - Use of this information to populate the DICOM 'header', making the scanned DICOM images a self-contained form suitable for off-site sharing, and allowing their use in downstream PACS, viewing and analysis tools.
- Wider education regarding the DICOM standard and its importance when procuring digital pathology systems within the NHS.
- Investigation into current / intended usage of DICOM.
    - Survey of NHS Trusts regarding awareness / use of the DICOM standard in existing / future digital pathology systems.
    - Sharing of DICOM file produced by NPIC scanners to explore real world compatibility between existing systems.
- Participation in ongoing enhancement of the DICOM standard in relation to whole slide imaging, particularly with respect to novel use cases.

Participants

Industry Partners	NHS Trusts	Others
• Leica Biosystems • Roche Diagnostics Limited • Sectra	• Participation of 18 NHS Trusts in survey	• Dr David Clunie • UKRI (Innovate UK)

Outcomes

- Our collaboration with Leica Biosystems has contributed to native production of DICOM whole slide images by their Aperio GT 450 DX scanners.
  - Early experience has resulted in successive and ongoing improvements to the DICOM compliance of the products.
- A number of educational activities have been undertaken to increase awareness of DICOM in digital pathology.
  - Panellist on Digital Pathology Association Webinar “Enabling Interoperability for Digital and Computational Pathology in the Age of Artificial Intelligence – Current Status and Future Directions”.
  - Presentation at Leica Digital Pathology Summit “Demystifying DICOM in Digital Pathology: Experiences from the National Pathology Imaging Co-operative”.
  - Presentation on the basics of DICOM in whole slide imaging, NPIC Webinar series.
- NPIC has been represented in the ongoing development of the DICOM standard.
  - Participation in DICOM Working Group 26.
  - Membership of DICOM Standards Committee.
  - Participation in IHE PaLM working group.
  - Contribution to publication “Integrating the Healthcare Enterprise (IHE) Pathology and Laboratory Medicine Guideline for Digital Pathology Interoperability”, J Pathol Inform. 2021 Mar 24;12:16.

Summary of investigation into current / intended DICOM usage:

- Responses from 18 NHS Trusts.
- 40% scanners reported as capable of producing images in DICOM format.
  - DICOM functionality being utilised in 75% of these.
- 55% viewing software reported as capable of viewing images in DICOM format.
  - DICOM functionality being utilised in 40% of these.
- Vast majority of respondents felt that DICOM compatibility was important / very important when procuring scanners / viewing software.
- Preliminary results from two sites have shown issues when attempting to view DICOM file / upload to VNA, highlighting non-conformance with DICOM of some existing systems.

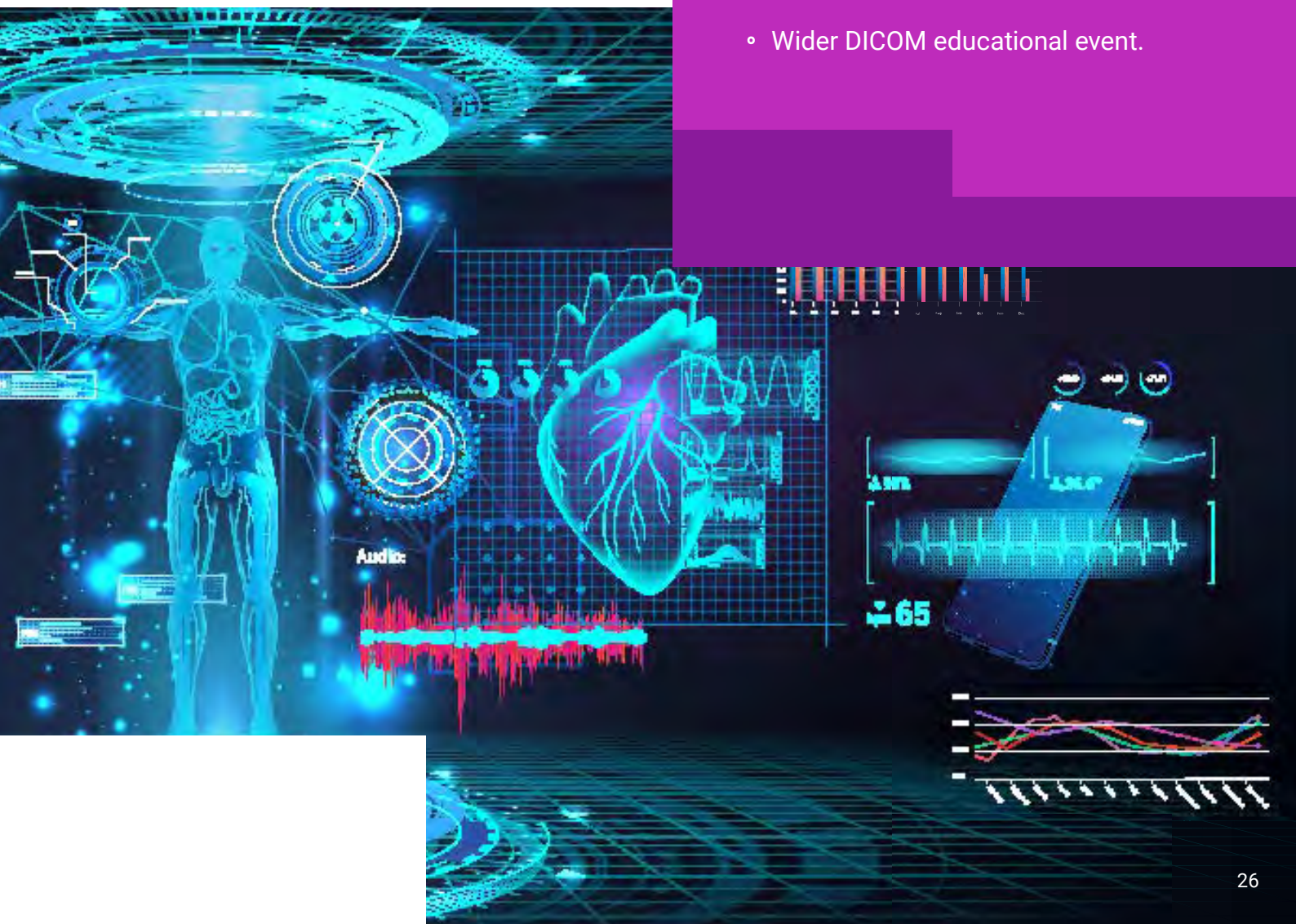
Recommendations

Our participation in educational events and survey of NHS Trusts confirms that there is increasing knowledge and appreciation of the DICOM format in whole slide imaging.

We recommend that conformance to the DICOM standard should be a prerequisite of digital pathology hardware and software procured for clinical use in the NHS to ensure interoperability and futureproofing of digital pathology image archives. DICOM is also recommended for routine basic science and clinical trial research use where the nature of the acquired images is amenable (e.g. immunohistochemistry and fluorescence imaging).

Next Steps

- Continue work on integration of scanners with LIMS to enable population of relevant information in the DICOM header using HL7 transactions and the IHE DPIA profile.
- Complete investigation into real world compatibility between existing systems.
- DICOM workshop planned for later in 2023.
    - To be headed by Dr David Clunie.
    - Discussion of DICOM conformance between scanner and software vendors.
    - Wider DICOM educational event.



Next generation file formats

Background / Aims of WG

Computational pathology has seen rapid growth in the recent past with the help of advanced AI algorithms. Most of these algorithms have been developed on public data from repositories such as TCGA (The Cancer Genome Atlas) and PathLAKE (PathLAKE: Computational Pathology Excellence). These repositories are built on open standards using open-source image formats, libraries, and APIs. One of the aims of PathLAKE include hosting a consortium data lake already populated with image data from multiple vendors on FAIR principles. The data lake needs to be in a

generic vendor-agnostic image data format and the solutions to be built on modern standards. This leads to the question, what is the best open-source vendor-agnostic whole slide image format which is suitable for developing AI algorithms and supports future technologies e.g., cloud-based platforms? Some of the main features of this format highlighted were support for large image sizes, high read / write performance, metadata storage, ability to convert from other formats, store multi-dimensional data (e.g. z-stacking), extensibility and modularity. As most of the modern AI libraries are built in Python, integration with Python is a major requirement along with fast and memory efficient (FME) region-wise random access with support for multi-resolution magnification reads. There should also be customisable cross-platform open-source software available for reading and writing the whole slide images (WSI) with support for state-of-the-art compression algorithms. Based on above the major aims of this workstream are defined as follows:

- Present an open file format which must be readable through open-source libraries with permissive license.
- Conversion tools must be readily available to allow conversion of file formats from various vendors to the proposed format.
- Should support state-of-the-art image compression methods.
- The proposed file format should be ready for widescale adoption by the industry.
- Should support advanced deep learning pipelines.

Methodology

Glencoe recently published their work in (Moore et al., 2021), where they compared different candidate formats such as TIFF, HDF5 and Zarr as possible file formats for storing bioimaging data. In this peer-reviewed work, Glencoe proposed low-latency, cloud-capable opensource next generation file format (NGFF) (i.e. ome.zarr) as a possible candidate to save bioimaging data in an opensource format. NPIC is also leading on an effort on defining standards for images for clinical use. This work is mainly based on extending DICOM standards to histology images which were previously developed for radiology images. Although DICOM can be used to achieve various goals defined in this workstream, but it lacks efficient support for modern cloud computing, time-series, multiplexed imaging and FME region wise and multi-resolution access required for development of AI algorithms and efficient deployment to cloud platforms.

Based on the aims and objectives, the following activities have been completed:

- Open file format.
  - We proposed **OME-NGFF** for this workstream as it has been shown to be cloud-friendly with extensive support for AI development. It also supports time-series, z-stack and multiplexed imaging data. Detailed comparison relative to existing formats **available** in (Moore et al., 2021).
  - Specifications for the file format are **available online**.
  - For any open file format it is necessary to provide public access to the example files to allow testing by the software developers. Examples files are **available online**.
- Availability of conversion tools.
  - Conversion tools through open-source libraries were made available under a permissive license, some of which are listed below. All the source codes are publicly available on GitHub.

    Isyntax (BSD)

    Bioformats (GPL)

    NGFF Converter (GPL)

    Wsic (MIT)
- Advanced Image Compression support.
  - Supports all the latest image compression techniques including JPEG2000.
  - Supports DICOM compatible codecs (i.e., DEFLATE, JPEG, JPEG 2000, JPEG-LOSSLESS, JPEG-LS completed).
- Evaluation of compression algorithms with the new format.
  - Preliminary results on this milestone were shared with the community in November in the workstream review meeting. This work is still under progress.

- Open-source reader.
  - Opensource reader is **available via GitHub** under a permissive license (BSD)
- Cloud Compute Support.
  - The file format has been optimised for cloud computing. **Example implementation is available.** Detailed comparison relative to existing formats on cloud-compute is available in (Moore et al., 2021).
- Widescale adoption is progressing.
  - Glencoe continues to engage with the open-source community on further development of this platform through **image.sc forum.** Open discussions are being held with the developers of AI and cloud computing tools for updating the technology to modern standards.
- End to End deep learning reference pipelines.
  - Reference AI pipelines were made available through TIAToolbox (Pocock et al., 2022) with the **opensource code under a permissive (BSD) license.** The source code provides support for open-source reader and for performing routine tasks for AI algorithm development along with reference pipelines.

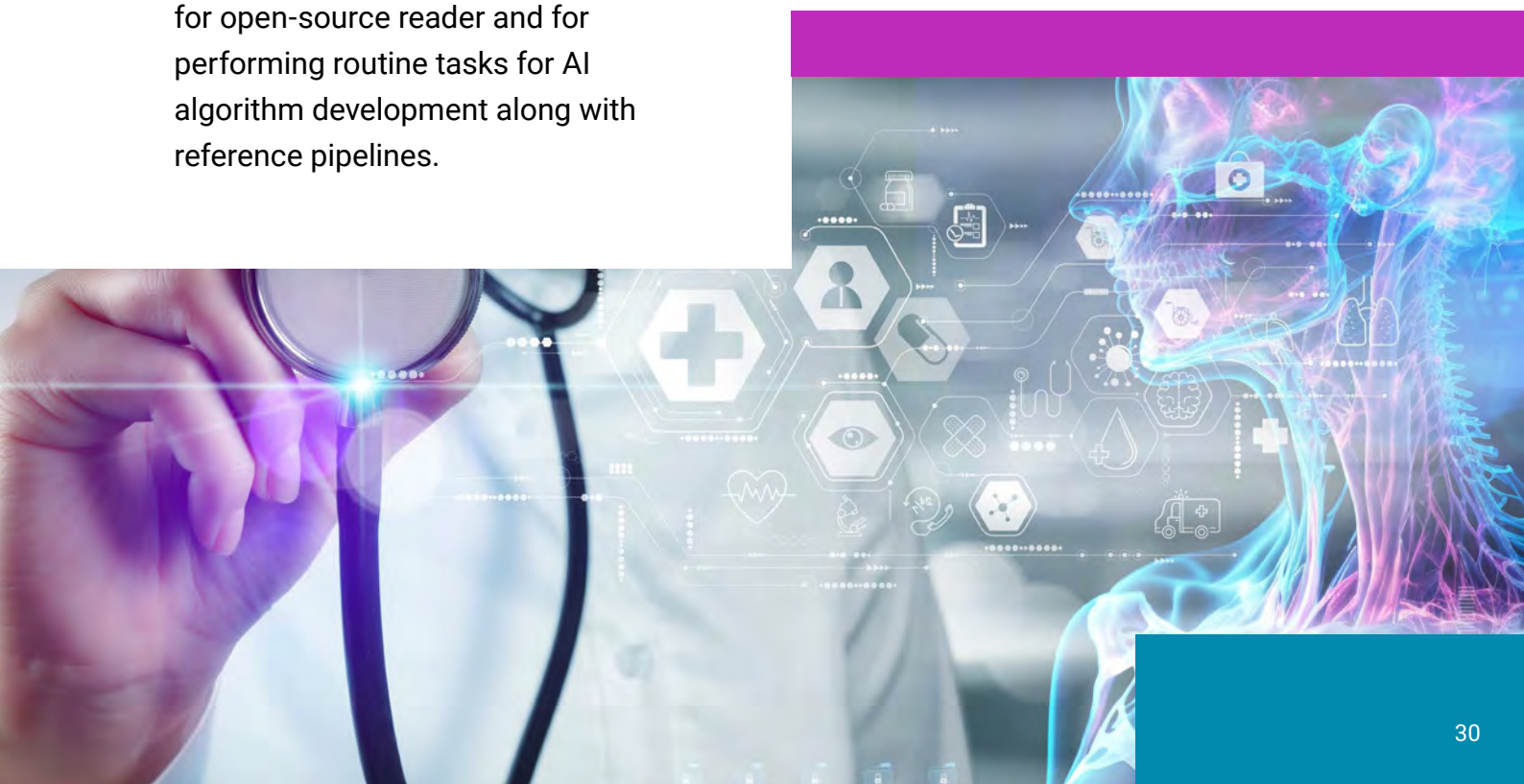
Participants

Whole slide image data from various scanner vendors was collected from partner hospitals in the PathLAKE consortium. The workstream was led by the teams at Warwick University and Glencoe software. All the source codes and data for conversion along with example AI pipelines were shared on public platforms.

Glencoe Software has worked in collaboration with partners like the Open Microscopy Environment (OME). You can find more information about OME and their projects by visiting their **website.**

The format has already been adopted by developers of modern AI and cloud compute libraries which can be assessed through example platforms such as:

- **OME-NGFF in action**
- **OME-NGFF in Google Colab**
- **OME-NGFF and CellProfiler**



Insights

The digital histology images are relatively huge in size compared to natural images or radiology images (if we only consider medical images). There are existing standards which have been developed for other type of medical images, but these might not be best suited for digital microscopy images. The requirement for random data access with efficient read / write along with support for cloud-based platforms necessitates development of data formats based on modern standards. The NGFF provides a platform which has been built based on these requirements. Furthermore, opensource libraries have helped in accelerating research and development of AI algorithms. As part of this workstream, we provide opensource libraries with permissive license for reading, writing, conversion to NGFF along with example AI tools required for achieving the defined goals.

Recommendations

The NGFF is available with all the tools required for widescale adoption and deployment of the platform to the clinic. The format has already been adopted by developers of modern AI and cloud compute libraries which can be assessed through the platforms listed on page 30. It is now up to vendors to evaluate this platform for their scanners for widescale adoption and taking it to clinic. We would welcome any feedback through image.sc forum so it is available for the wider community to debate.

De-identification of data

Background

Healthcare systems digitised data is now viewed as a major asset in helping to transform healthcare services by extracting both operational and clinical insights embedded within patients' records. To this end, both public and private organisations are now exploring how to use real patient data in a safe and effective manner. Part of the requirement to gain access and process this data is to provide suitable mechanism to protect a patient's privacy. De-identification or anonymisation is one of the mechanisms that can be deployed to support such an approach.

The de-identification working group's remit has been to look not only at what partner Challenge organisations have been doing within the context of individuals centres' AI work relating to de-identification, but also what other private and public bodies have been doing to address such issues. The work of the group is inextricably linked to the SDE working group whose remit is much broader than just patient privacy protection but is concerned with offering a secure environment to work with sensitive data.

Approach

Workgroup meetings were held at regular monthly intervals from late 2021 until the end of the Centre of Excellence programme in early 2023. The initial remit was to identify the different approaches that the various D2EDPM Centres of Excellence took to address de-identification. What became clear at an early stage was that de-identification was just a narrow aspect of patient privacy protection and that the group would therefore need to widen its remit to explore other aspect of privacy protection.

The group outlined who the key stakeholders were related to patient privacy protection and in particular de-identification / anonymisation. An early attempt was made at identifying risks associated with de-identification and patient privacy protection. The group then discussed aspects of how the severity of the risk or the likelihood of it occurring could be established in a more standardised way. And finally ethical aspects related to de-identification were also discussed.

Participants

Participants in the group included representatives from industry, academia, and the NHS:

Industry Representation	Academic Representation	Healthcare Representation	Others
• Canon Medical Research Europe Ltd • Siemen Healthineers • CoE partner SME	• HDR UK • Challenge partner universities	• NHS Scotland boards • NHS England Trusts	• Regulatory bodies • UKRI (Innovate UK)

Insights

The insights noted below arose from the WG discussions related to actual experiences that individual partners have had and how they related to issues of data protection.

Privacy Protection with data at scale

When working with data at scale, the risk of identifying information not being successfully anonymised increases, particularly where some of the de-identification processes have a manual element. Fully automated de-identification systems can also sometimes miss potential patient identifiers – particularly in free text situations. Compounding this basic problem is the fact that using a larger data set for individual patients, be it longitudinal data or data that are linked across diverse data sources, means that the overall risk of re-identification increases. The more data points about individuals that are included increases the chance that re-identification techniques can be used to disclose their identity.

Risk assessment related to privacy protection in AI development

Data controllers at different data sites have varying appetites when it comes to the levels of risk that they will tolerate. Part of the issue is the current lack of suitable guidance at a national level offering advice on how to objectively assess risk and how these risks can then be mitigated. Partly this is related to local culture e.g. recent incidence in unauthorised data disclosure could result in more stringent controls and checks relating to data use; but in many cases controllers may simply lean towards erring on the side of caution when no guidance is on offer.

Other risks related to privacy protection

The number one privacy risk is identifying a patient from records not correctly de-identified or protected from re-identification attempts. However, it is also important to highlight the risk to AI development if data is not made available to researchers be they academic or from industry. This risk has nuances too – not getting access to the right quantity of data, restricted access to the required fields of data needed to develop an AI algorithm, and time it takes to make the data available for AI development and validation. We have heard many instances where projects are significantly delayed due to the delay in granting permissions to access data.

Mitigating risks

As AI development is a technically focussed endeavour it is natural to seek technical mitigations to alleviate potential risks. However, it has been noted that technical solutions are not a panacea at this point in time and suitable operational measures, such as definition of standard operating procedures related to de-identification, also need to be considered. It has also been highlighted that having legal recourse in situations of unauthorised data disclosure should also be considered as a mechanism to protect both private patient information as well as the reputation of data controllers.

Recommendations

The key recommendations are to provide support to data controllers to allow them to do their job as expediently as possible without compromising data privacy protection.

Support risk assessment related to AI development and evaluation

As mentioned above, risk assessment is not standardised in any way in this particular field, and most of the requests for access to data are currently assessed on a case-by-case basis. Having a curated set of information that outlines possible risks associated with large scale data types or AI development could potentially help data controllers to approach data governance in a more consistent way. How these would be curated has yet to be understood – possibly as a standard or even as a consultative (advisory) service. All of this has yet to be explored.

Regulatory bodies or organisation associated with them

It is important that regulatory bodies and associated organisations that have a vested interest in patient privacy protection are aware of each other's work in this space. From a data controller's perspective, having policies or guidance provided by different bodies that contradict each other will make data governance more problematic. Harmonisation of guidance or advice in this space is important.

Understand both patient and public attitudes towards data protection

It is important to regularly gauge both patient and public attitudes towards the use of their healthcare data for healthcare AI developments and deployments. This should help steer policy bodies to reflect these groups sentiments within the guidance and/or regulations relating to privacy protection.

Standards for privacy protection

Create a resource that offers data controllers a "standard" regarding data anonymisation that they should be conforming or at least aspiring to with regards to data being used for AI developments. As mentioned above if multiple agencies are involved then a coordination function would need to be established with regards to contributions and managing any conflicting views / advice.

Next steps

Curate current practices and associated risks

Explore the creation of a repository of information related to risks and mitigations associated with data privacy particularly with large scale data used for AI development. Consult with data controllers to understand their concerns and also to supply examples of risks that have occurred and how they were mitigated (both successfully and unsuccessfully). Explore how to best disseminate this information e.g. via an online resource or via some dedicated advisory service.

Extend work on mitigations

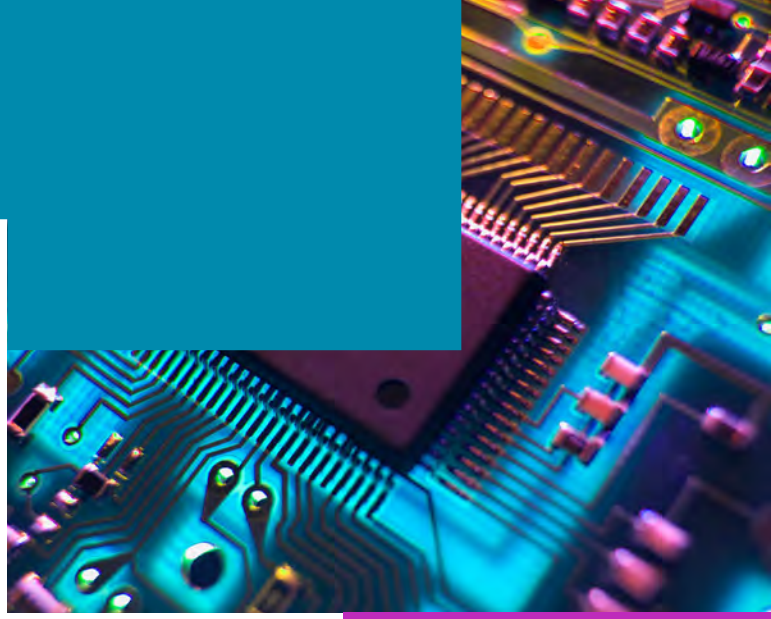
Curate information about mechanisms being used by data controllers and how it relates to the risk mentioned above e.g. HIPS (Hiding in Plain Sight) method used in iCAIRD (this has the benefit of providing realistic data as part of AI training rather than redacted data and avoids unexpected learnings inside the black box). Highlight what they do and what they don't do. Explore other "mechanisms" beyond pure technical solutions i.e. standard operating procedures, legal protections etc. Taking into account that different sites will have different infrastructure, resourcing, localised policies, so any mitigations need to be general purpose enough to accommodate these differences.

Approach to SDEs

Background

The recent policy sea-change concerning access to medical data will overturn the route by which academic and commercial data scientists access medical image and associated clinical datasets: away from the traditionally operated “data to researcher” provision to a new “researcher to data” model. The drivers underpinning this change include a desire to better control access to, and subsequent use of, sensitive private patient NHS data, the opportunity to provide access to larger datasets than controlled by an individual NHS Trust and the provision of a means for better value recognition of NHS data where its use contributes to the generation of new IP.

Improving the use of health data for research was a core theme of the Ben Goldacre report “Better, broader, safer: using health data for research and analysis”. The use of Secure Data Environments (SDEs; formerly Trusted Research Environments) is a central component to the delivery of this vision.



Funding was established in late 2022 for eleven sub-national SDEs operating across aggregate groupings of NHS Trusts across NHS England. These are being established over the next 5 years and will provide a single route of access to NHS health and social care data for analysis and research purposes.

This workgroup addressed the issues that commercial AI developers anticipate facing as they transition from their current in house "data to researcher" approach to an SDE centric model. It should be recognised that AI development activities currently happen within commercial organisations who provision appropriate compute, GPU, and storage capabilities along with the necessary technical support skills sets. The transition to the SDE model imposes greater business risks in that the required service levels to support AI development are insufficiently mature and stable. This elicits a concern that AI development within the UK will dwindle if companies who operate in a global healthcare context either offshore or develop AI solutions without reference to UK patient datasets.

Recognising that the AI industry has had little input to the evolution of SDE development, this workgroup sought to capture and articulate the specific requirements of commercial AI developers in respect of their future interactions with SDEs. The intent is to provide an insight into these requirements for SDE-service providers to address the concerns and capability requirements in a timely manner.

Approach

The workgroup held a series of workshops over Q4 2021 to Q3 2022. Its initial focus was to comprehend how the requirements of commercial AI developers with regards to complex imaging datasets differed from the needs of academic data scientist researchers. In particular, there is a need for NHS SDE providers to create and provide access to specialist environments that allow nascent AI tools to be trained and evaluated on NHS clinical images. The workgroup surveyed its industry partners for their needs with respect to SDE service provision and subsequently captured these as a set of formal requirements. In parallel, we contacted likely providers of SDE services to test their current capabilities or plans to support these requirements. The findings (listed at Appendix 1) were subsequently validated by both the Bioscience Industry Association and NCIMI's Industry Forum group. The overall findings of this workgroup were subsequently shared with NHS technical Directorate, DHSE, National Digital Diagnostics Senior Woking Group, the Bioscience Industry Association, and the Innovate UK challenge Executive.

Participants

The Approach to SDE workgroup operated for eleven months and comprised representatives from the commercial AI sector, from the NHS and associated data providers, from the Innovate UK funded Centres of Excellence and from Innovate UK.

Commercial AI developers	NHS & associated partners	IUK & CoEs
• Canon Medical Research Europe Ltd • Siemens-Healthineers • Leica Biosystems • Roche Diagnostics Limited • BC Platforms	• NHS-England (as NHS-X) • Royal Marsden Hospital • HDR-UK	• Innovate UK D2EDPM Challenge team • National Consortium for Intelligent Medical Imaging (NCIMI) • Industrial Centre for Artificial Intelligence Research in Digital Diagnostics (iCAIRD) • London Medical Imaging & AI Centre for Value-Based Healthcare

Additional validation on the findings of this workgroup was obtained from both the NCIMI Industry Forum and the Bioscience Industry Association with representation from:

NCIMI Industry Forum	Bioscience Industry Association
• Canon Medical Research Europe Ltd • GE Healthcare • Perspectum • Mirada • Caristo • Brainomix • RAIQC	• Workgroup delivering “Driving Growth and Patient Benefits Through SDEs” report

Insights

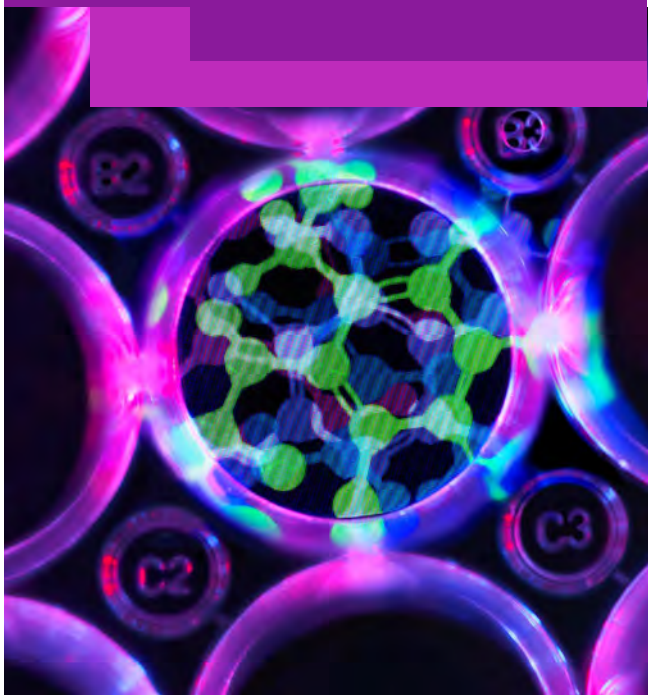
- Commercial AI software development for imaging analysis applications, which widens to multi-modal healthcare data, has an additional set of specific requirements for it to work effectively with SDEs. These are distinct from those of the more classical data scientist researcher (commercial, academic or NHS) and reflects the scale and complexity of the imaging datasets being employed and the AI tools' interactions with them.
- The requirements of the AI developer cover all phases of the AI product lifecycle. The requirements span initial project set-up, data validation and provisioning, model training and validation, model export and archiving (these are detailed at Appendix 1).
- There is a need for SDEs to provision highly performant remote access to AI developer teams to their unique project workspace. This will require provision of dedicated remote desktop spaces with a quality-of-service guarantee on latency, bandwidth, and issue resolution. Tiered profiles for remote access will need to be provided to ensure high-quality connections.
- SDEs will need to provide managed access to powerful ML training infrastructure. This is needed to provide sufficient processing power (GPU/CPU/RAM and storage) so that models can be trained effectively. Mechanisms will be needed to book / provide exclusive use or time slices of expensive resources such as GPU or fast storage / access for projects at different stages.
- SDE providers will need to provide specialist trained and experienced personnel to support the full lifecycle of AI research projects. There is an expectation that this will operate at a contractual service level across the lifespan of the project.
- There is a lack of clarity about the timelines over which access to NHS clinical data will change to a point that SDE-only access is mandated. The AI industry need clarity on this timeline in order to plan its migration to the SDE model.
- There is concern over the maturity of the SDE offerings to match the requirements of the AI industry and for SDE providers to stand up the service levels required.
- There is concern about the SDE providers to be able to access appropriately skilled and experience staff to support the data engineering requirements. Most of this workforce currently works in the AI industry.
- Expectations around future product royalty payments for access to NHS data (as compared to commercial fee-for-service charges) are seen as a showstopper, particularly for international organisations.

Recommendations

- The provision of the necessary staff and compute infrastructure is an investment that must be made upfront in advance of commercial AI development contracts. This investment needs central funding. It is likely that such investment will not be deployed across all eleven SDEs so agreement should be reached on a route of access for Industrial AI partners to NHS imaging data. This is potentially via a small number of SDE appropriately provisioned with compute capabilities and trained staff. This discussion needs to be entered in to.
- Prototyping of exemplar AI development projects needs undertaking with a subset of the nascent SDEs. This will inform on connectivity, access controls, data governance, protection of IP, compute, and service level requirements. The outcome of these exemplar projects will be a better definition of the service level capabilities needed to be provided on a commercial basis.
- Effort needs to be undertaken in advance to attract, recruit and train the workforce required to support AI development activities within SDEs. If this workforce cannot be attracted to the NHS, an outsourced commercial SDE provider may be required to support the AI development requirement.
- Discussions about the realities of value recognition to the NHS need to be concluded. If the current impasse continues, the risk to "UK PLC" is that the AI industry will offshore its product development activities outside of the UK and the NHS risks that future AI tools will not reflect the UK patient population or standards of healthcare delivery. This ultimately impacts the future benefits of AI to the NHS and to its patients.

Next steps

- During Q2 2023 representatives of this workgroup and Innovate UK to meet with NHS-E at a senior level to discuss the advancement of this report's findings and recommendations.
- Obtain commitment from NHS-E to the publication of the roll-out timeline for SDEs with trigger points as they impact data access requests.



Federated learning

Requirements Summary

The federated learning (FL) paradigm facilitates NHS Trusts in collaboratively developing AI models without exchanging the patient data. FL empowers NHS Trusts to preserve patient privacy by owning and controlling their data inside trusted hospital networks.

However, existing data anonymisation is often ineffective when dealing with large datasets or conducting experiments over diverse datasets. In addition, the technology stack at each hospital is owned and run by different stakeholders each complying with their internal organisational policies and operating procedures. Moreover, data ethics and governance requirements vary among participating institutions as discussed on page 32. Multiple FL platforms exist each of which use different technologies and communication protocols: each platform offers distinctive features with varying capabilities, site-specific model definitions, alternate communication protocols, and custom security models. Therefore, the development of a fully integrated and interoperable FL platform is a challenging task.

The FL workstream primarily undertook activities to:

- Minimise the interoperability gaps between different FL platforms by standardising the AI model development activities for all stakeholders and by training the FL models across multiple hospital infrastructures.
- Explore and understand the dependencies beyond FL stack.
- Develop a framework to comply with ethics and governance requirements of FL systems.

Example User Stories

Since FL is a collaborative model development activity among various stakeholders, (including model owners, data engineers, researchers, system engineers, and clinicians) each of the stakeholders should therefore be able to perform their intended activities.

- **As a model owner (a designated person from each collaborating Trust):**
  - I should be able configure the FL workflows and invite other collaborators to join the model development process.
  - I should also be able to track the experiments, allow/prevent other collaborators from participating, and control various versions of developed models.
- **As a data engineer:**
  - I should be able to access, query, organise, and transform local datasets from discrete sources.
  - I should also be able to collaborate with other data engineers from other collaborating trusts.
- **As an AI researcher:**
  - I should be able to submit my custom code complying with my domestic FL platform settings and standardise model collaboration process.
  - I should be able to train, validate, evaluate, and version local AI models complying with a the bare-minimum standard set by consortium.
  - I should be able to communicate with other Trusts in a standard way.
  - I should be able to collect and dispatch the same type and form of data across the training network.
- **As a system engineer:**
  - I should be able to provision my local training network, and equally access the aggregation servers provisioned by other Trusts.
- **As a clinician:**
  - I should be able to test the trained models and provide feedback on the quality of trained models.

Proposals to support the requirement

The interoperability experiments were conducted by deploying the federated learning and interoperability platform (FLIP) inside King's College London's (KCL) compute network, and iCAIRD-owned SHAIP nodes inside NHS Scotland (Glasgow and Aberdeen). Both the FLIP and SHAIP systems used NVIDIA's NVFLARE as the common communications component. Each participating Trust (re-) trained a global model using their local datasets and the model updates were transmitted back to the central server which aggregated all model updates and produced a new version.

Fig 9 depicts the primary experimental setup of our proof-of-concept whereby both participating trusts connected with a secure hosted environment provisioned on Amazon Web Services (AWS).

- Our AI researchers at both hospital data sources configured the same FL pipelines inside their FL networks, however, both FL platforms (i.e. FLIP and SHAIP) implemented additional components to enable bi-directional communication with the AWS-hosted environment. The connector component collects AI configurations and dispatches local model updates for aggregation.
- The FL hub at AWS aggregates all model updates and produces a new version of global model which is retransmitted to all FLIP and SHAIP nodes for subsequent training.
- The process continues until we reach a desired level of model accuracy.

The same setup is used to test and validate the trained models.

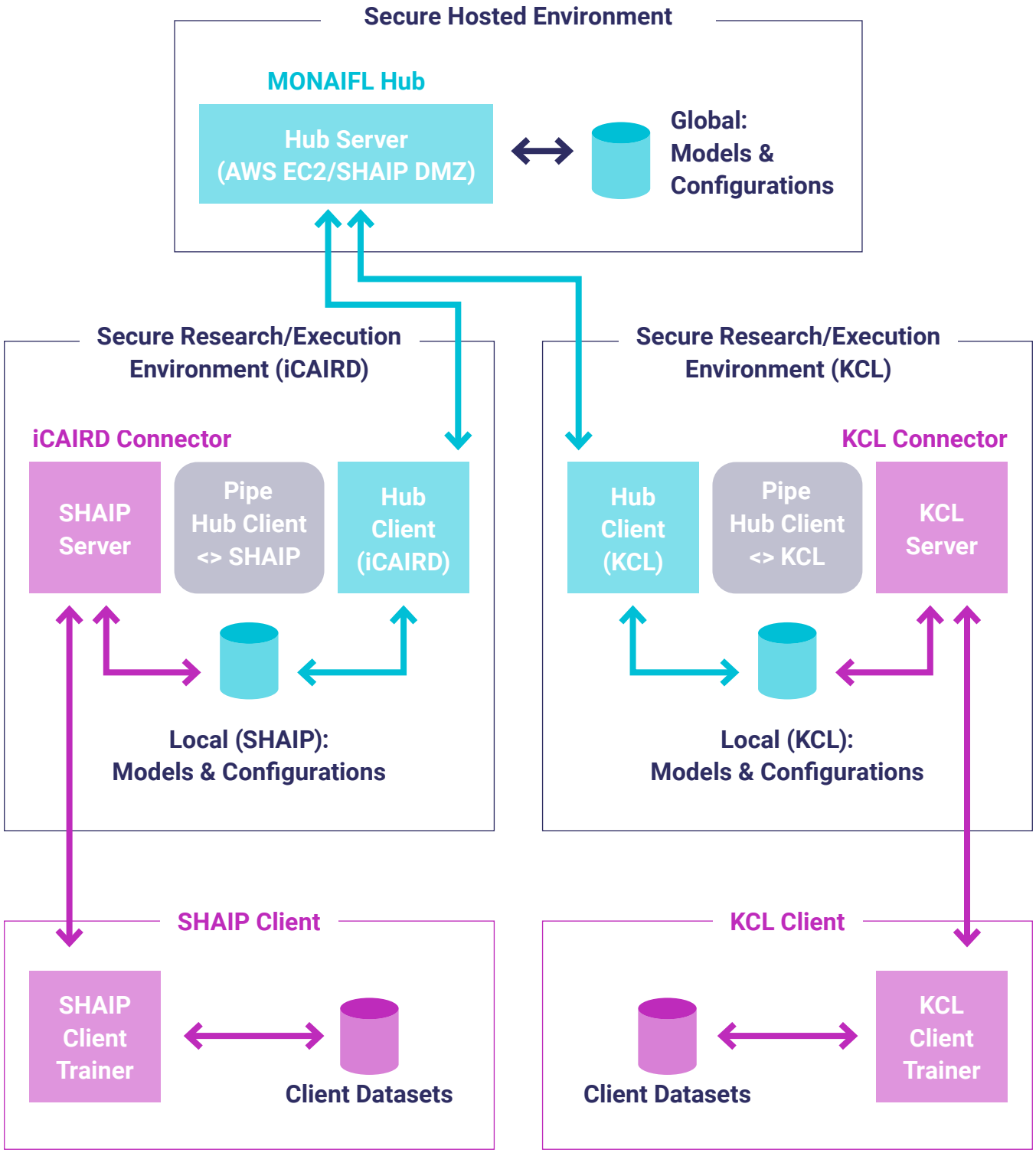


Figure 9: Main Experimental Setup

The Proof-of-Concept (PoC) project explored two different use cases, as two distinct experiments. The first used standard FL model definitions (named in MONAI-FL and accepted in the MONAI codebase) and MONAI bundle configurations. The hub that all sites connected to was hosted in AWS London, as described above. The second used custom python code and custom models (not predefined in the MONAI codebase) and moved the hub to being hosted at one of the sites (Glasgow NHS Health Board was used). We aimed to show different aspects of the flexibility of the FL paradigm in each of the experiments.

In each case, the proposed platform was tested by running live experiments across NHS-Scotland (using SHAIP as the training platform) and KCL sites (using FLIP), running the experiments against a well-known dataset (BRATS), and we found the results promising when compared to centralised model training.

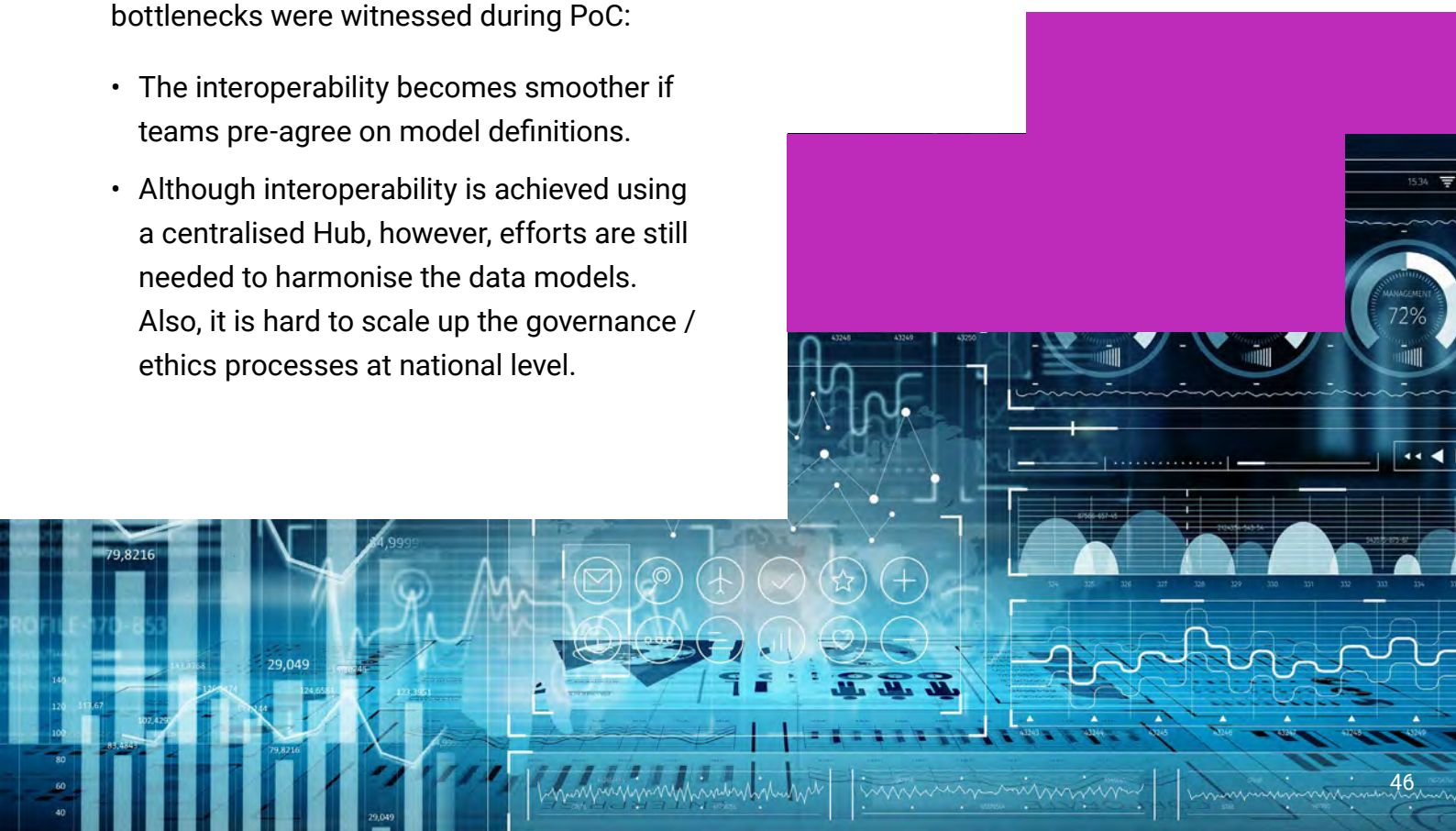
Despite successful implementation, a few bottlenecks were witnessed during PoC:

- The interoperability becomes smoother if teams pre-agree on model definitions.
- Although interoperability is achieved using a centralised Hub, however, efforts are still needed to harmonise the data models. Also, it is hard to scale up the governance / ethics processes at national level.

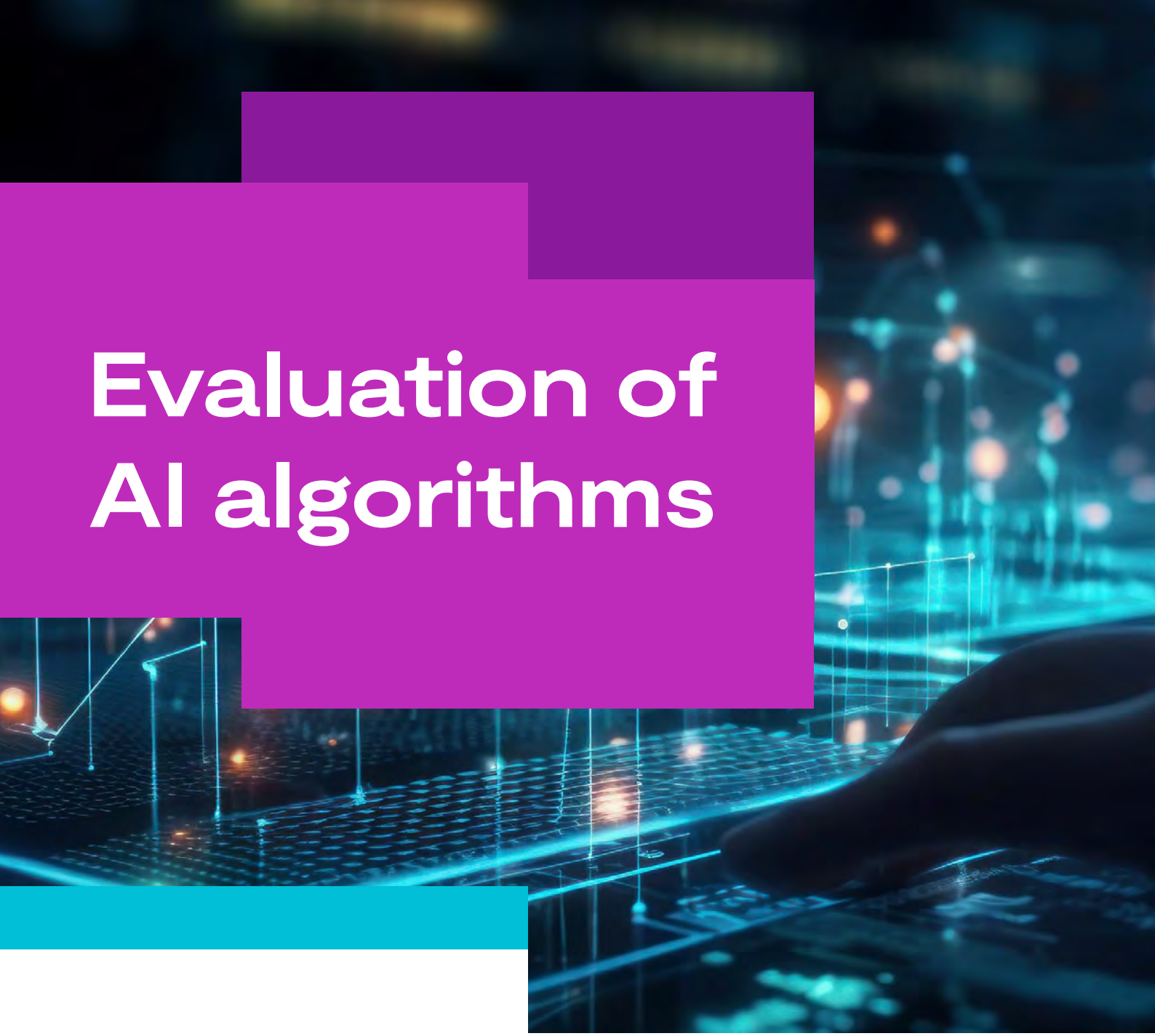
Policy Recommendations

Considering the utility and potential outcomes of these interoperability experiments, the following recommendations are being made for policy makers.

- There is a need to prioritise some high impact FL use-cases that can benefit most of the patient population and reduce variations in care and outcomes.
- FL studies should be encouraged by introducing new information governance policies that can lower the entry barriers.
- Investments should be made to acquire FL tools to meet the clinical requirements of high-quality of healthcare services.
- Large-scale FL literacy activities are needed in order to retrain staff and develop new skills and competencies, create awareness and engage all potential stakeholders.



Evaluation of AI algorithms



Background

From the start of the D2EDPM Challenge and in particular during the course of the COVID pandemic there has been an explosion in the use of digital health technologies to help patients and care providers manage various conditions. AI was (and is) being promoted as a mechanism to help health systems under pressure to cope with the additional workloads that have resulted from care backlogs that have been steadily growing over the past few years and that were exacerbated by the pandemic.

Today there are numerous regulatory approved AI-enhanced products available on the open market. However, what has become evident is that AI algorithms are not a plug and play technology and that careful assessment needs to be carried out on the technology to ensure a good fit for the care system planning on deploying it. What we are learning is that evaluation of AI needs to occur at the various stages of the technology's development pathway in order to deliver value in the healthcare context.

Approach

Workgroup meetings were held at regular monthly intervals from late 2021 until the end of the Centre of Excellence programme in early 2023. The remit of the evaluation working group was to determine what types of evaluation approaches have been carried or were being planned across the centres of excellence and to see if a “check list” could be created that could standardise a generic approach in setting up evaluations. In addition, the group discussed the different stages where evaluation would occur and determined what types of evaluation could, or should, be put in place for each of these stages.

The group discussed who the key stakeholders are with respect to evaluation. This included not only those that are involved in evaluations but also those who are impacted by or have a stake in the evaluation process.

Finally, the question of reproducibility of evaluations was addressed. Specifically, how should evaluations be re-run at various sites allowing like-for-like comparison of results.

Participants

Participants in the group included representatives from industry, academia and the NHS:

Industry Representation	Academic Representation	Healthcare Representation	Others
• Canon Medical Research Europe Ltd • Siemen Healthineers • CoE partner SMEs	• HDR UK • Challenge partner universities	• NHS Scotland Health Boards • NHS England Trusts	• Regulatory bodies (ICO, MHRA) • UKRI (Innovate UK)

Insights

The dominant insight we had from discussions in the group was that "evaluation of AI products", as a process, remains a nascent activity and therefore needs to be carefully nurtured and developed to ensure maximum benefit can be gained by all stakeholders involved.

Understand your stakeholders

It was evident that when you look at the various stakeholders in evaluation a lot of tension can arise. AI developers (both academic and industry) are keen to get access to representative data in a timely manner. Data controllers and their respective organisations, (NHS or academic institutions), are concerned about revealing personal information about individuals to unauthorised personnel and consequential repercussions. Clinicians are keen to understand exactly how an AI will benefit their practice and how it can impact the existing clinical workflow (what would need to change to accommodate this technology). NHS administrators want to understand what benefits such systems will offer from cost and patient outcomes perspectives. What is evident is that the scope of AI evaluation needs to engage with all the key stakeholder groups.

Can we "checklist" AI evaluations?

The group recommends that delivering evaluations needs to be carried out in a repeatable & consistent manner. The drivers are to ensure that such evaluations were:

1. **Fair:** ensuring that the same evaluation criteria could be applied to AI products addressing the same application area but from competing suppliers.
2. **Rigorous:** providing a measurable means of assessing AI product safety and efficacy in a real-world environment.
3. **Objective:** offering a way to simplify (or at least support the need to carry out) a value-based assessment of the technology.

The group concluded that it would be impossible to create a set of specific evaluation criteria that could be applied across the breadth and depth of AI application areas. Instead, the recommendation was to provide a checklist of potential evaluation areas that individual

projects could pick and choose depending on the relevance to their specific application area. The key aim being to consider the perspectives of the various stakeholders involved. This included addressing the context of use, what the safety criteria should be, what infrastructure would be needed for an evaluation, what data would be needed, what governance concerns and processes would need to be considered, who the stakeholders are, what QA processes should be considered and applied, what would the usability assessment include and finally, how would you go about determining the technology's value (e.g. carrying out a health economic assessment).

What evaluation at what stage?

The group identified the fact that different evaluations can occur at different stages of the technology's life cycle. These evaluation stages include:

1. **Evaluation of a proposal** the point at which an idea is evaluated for its clinical users need or desirability, determining if it has the potential to have a direct or indirect impact on patient outcomes, and whether it has the potential for a wider application beyond the evaluation site.

2. **Evaluation of an in-development system** the aim here being to provide feedback to the developer based on how well the system performs in the evaluation location across multiple location iterations. This would be used to deliver incremental improvements or pivot the development towards more valuable goals identified during the evaluation process.

3. **Evaluation to gather regulatory evidence** so that a developed product can build an evidence file that can be submitted to a regulatory approval body.

4. **Pre-deployment evaluation** to determine the appropriateness of a regulatory-cleared AI product for a specific deployment. The aim is to determine how well an AI product performs at the target deployment location before going live. This could include determining actual efficacy against advertised capability, testing areas of potential safety concerns, establishing how well the AI product performs with local demographic data and local edge-case data.

5. **Post deployment evaluation (surveillance)** part of the discussion, particularly post-pandemic, is how much pre-deployment evaluation is enough? The question is - can a robust surveillance mechanism be put in place that could allow pre-deployment evaluation to be scaled down without compromising a product's safety in terms of end-user or patient outcome impact? How much manual intervention or human oversight would a surveillance system need?

Recommendations

Define exactly what you mean by evaluation

It is important to clearly identify the type, or the stage, the evaluation in question is at. This will dictate what resources will be required to carry out the work. Each stage will require engagement with different stakeholders although some, such as clinical experts, may be involved across all the stages.

Each stage will have different infrastructure requirements so a clear definition of what is needed in terms of personnel and equipment is imperative.

Each stage should also clearly define what its outputs or intentions are. As an example, at earlier stages evaluation outputs are product development guidance (i.e. what features are useful), while at later stages the evaluations are more likely to be assessing product suitability for a site.

The one thing that all evaluation stages should have in common is a focus on the "value" of the technology (i.e. in terms of cost effectiveness, predicted or actual patient outcome improvements and benefits to the care systems themselves). Understanding infrastructure requirements (e.g. what platforms could be used to facilitate the setup and teardown of evaluations) is another.

Do not re-invent wheels

Effort should be made to identify who already has a stake in defining what evaluations are required and what the expected outputs should be. A good example of this is MAAS which brings together NICE, MHRA, HRA and CQC expertise.

One concern expressed by industry and academic AI researchers is the inconsistency between regional centres related to gaining access to data to be used for evaluations (i.e. data governance procedures and rules). AI developers need to understand exactly what the requirements are with regards to evaluations so that they can assess what needs to be done from their perspective and if that effort will result in valuable outputs.

Related to the "check list" for AI evaluations above is the need to review the work being done in the development of AI evaluation protocols. We already know that initial efforts are in place such as SPIRIT-AI and CONSORT-AI relating to extending existing clinical trial protocols to accommodate AI interventions. One of the overriding issues here is to balance evaluation costs (time and money) as compared to satisfying the end-user requirements in a timely manner.

Learn from those with experience

Learning from teams already carrying out AI evaluations would be time well spent. Irrespective of outcome, these evaluators have been testing mechanisms to determine protocols and methods to deliver evaluations as effectively as possible. An example of this type of effort is NHSX AI labs Artificial Intelligence in Health and Care Awards Technology Specific Evaluation Teams (TSETs) that were established to evaluate technologies funded through the associated award scheme. Learning from the teams involved in these evaluations would be an invaluable exercise.

Next steps

Harmonise work on “check lists” for evaluations

Ensure that bodies with a stake in AI evaluation, such as MAAS, deliver mechanisms to “standardise” how evaluations are set up and reported on. It is important that developers are clear regarding what needs to be done and when in order to satisfy end user requirements for the eventual deployment of AI technologies in a healthcare environment.

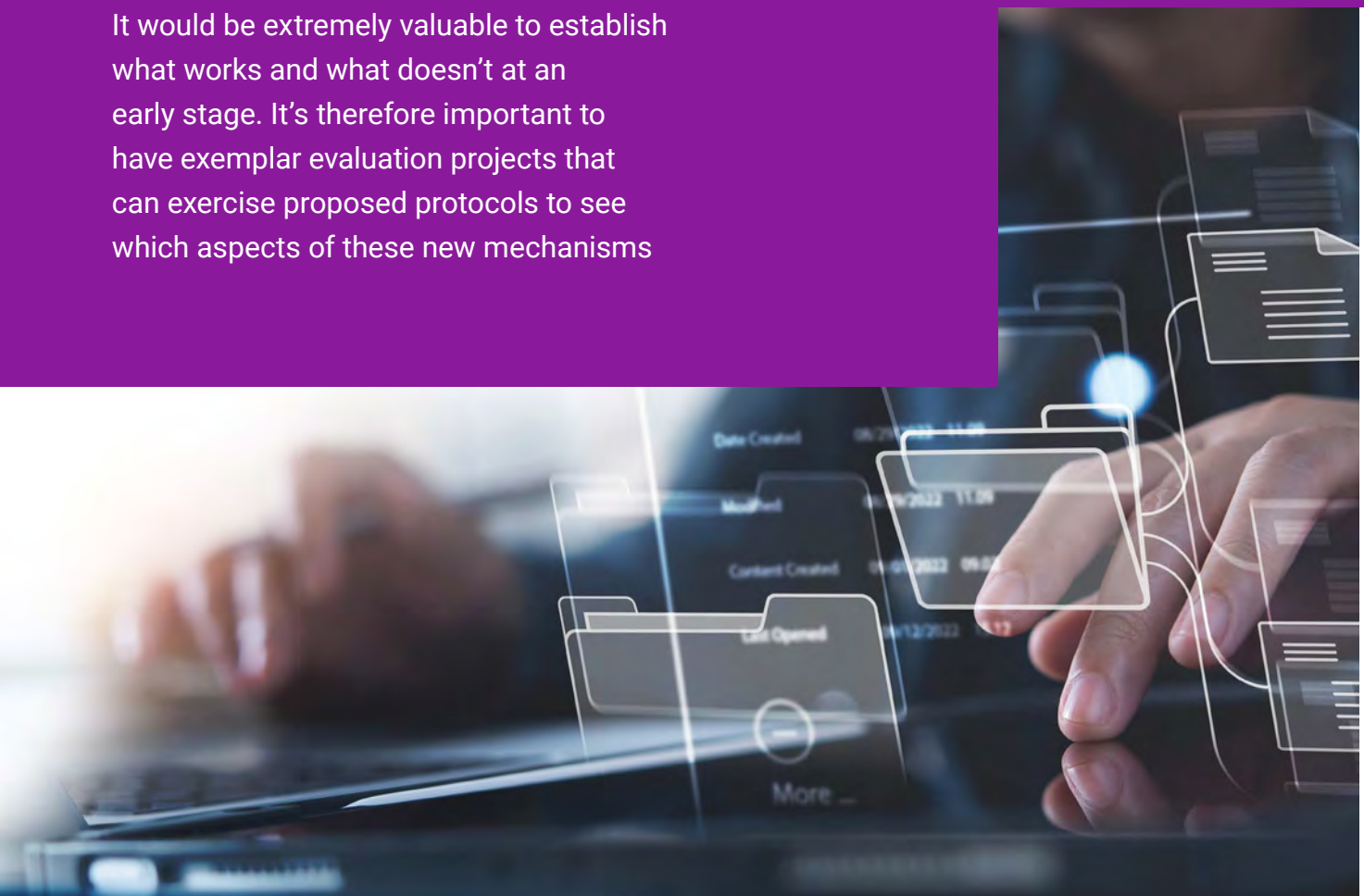
Tracking an effort on protocol development such as SPIRIT-AI, CONSORT_AI and DECIDE-AI would be valuable in order to avoid unnecessary repetition of work.

Establish what types of evaluation are effective by running exemplar evaluations

It would be extremely valuable to establish what works and what doesn’t at an early stage. It’s therefore important to have exemplar evaluation projects that can exercise proposed protocols to see which aspects of these new mechanisms work or do not work in various contexts. Having exemplar evaluations exercise AI development at the various stages would need to be prioritised by the key stakeholders (healthcare providers and AI developers).

Define the requirements for evaluation stages

Understanding the development stage at which an AI application or product is currently at is important to determine what resources will be required. Creating checklists for each of the evaluation stages could be a valuable resource for teams embarking on such an endeavour.



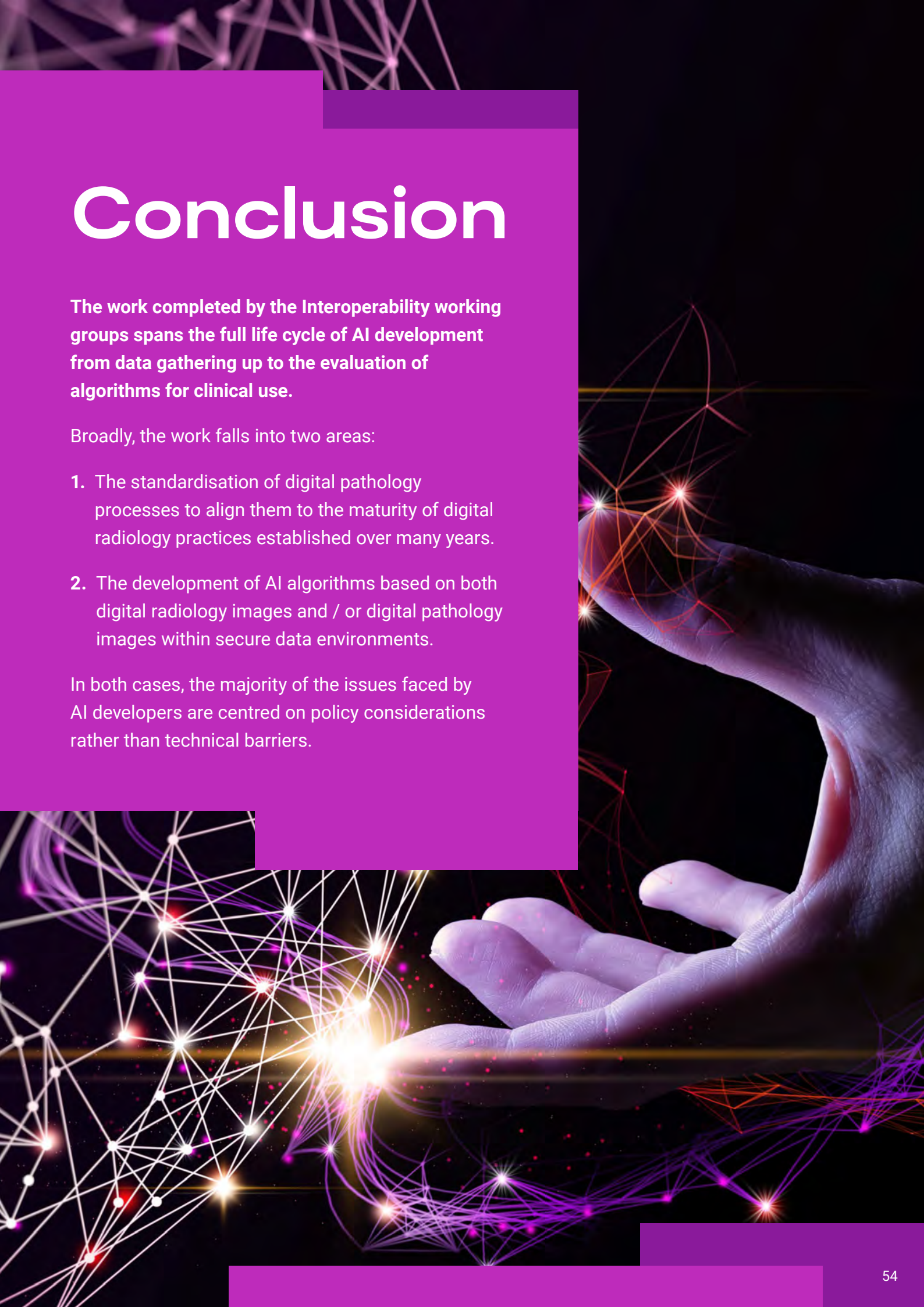
Conclusion

The work completed by the Interoperability working groups spans the full life cycle of AI development from data gathering up to the evaluation of algorithms for clinical use.

Broadly, the work falls into two areas:

1. The standardisation of digital pathology processes to align them to the maturity of digital radiology practices established over many years.
2. The development of AI algorithms based on both digital radiology images and / or digital pathology images within secure data environments.

In both cases, the majority of the issues faced by AI developers are centred on policy considerations rather than technical barriers.



The recommendations are broad ranging and encompass the following points that should be recognised by policy makers to ensure that the UK can establish the right conditions to foster future investment from international diagnostics companies and support future UK SME growth.

The key conclusions include:

- Establish guidelines for digital pathology lab processes utilising standardisation and normalisation concepts.
- Widen the adoption of DICOM standards to ensure digital pathology images are consistently codified across all input devices procured by NHS Trusts.
- Adopt a new file format to facilitate the processing of huge digital pathology image and metadata files across the cloud.
- NHS data owners should provide a high-quality managed service to allow algorithm developers to interact with patient data securely. This needs to meet, or exceed, performance levels currently being achieved by AI developers in house organisations.
- Policy makers should ensure effective de-identification policies and technologies are applied across all SDEs.
- Federated Learning has been shown to be technically feasible across two jurisdictions (England and Scotland). The objectives for future FL projects should align to NHS strategic priorities and provided with long term funding to give companies confidence to invest further.
- Policy makers should impose an evaluation framework for AI algorithms to ensure a safe and effective approach is followed by all developers.

Acknowledgements

The D2EDPM Challenge team at Innovate UK would like to extend their thanks to all of the interoperability working group members for their valuable input to this activity over the time of the challenge. In particular, the following individuals were instrumental to the delivery of the workstreams and this report:

John Zurowski ICAIRD
Mark Beggs NCIMI
Jorge Cardosa LMIAIC
Shan Reza PathLAKE
Craig Sawyers NPIC
David Brettle NPIC
Alan Pinder Canon Medical Research Europe Ltd

Finally, we would like to acknowledge the work of our late colleague, **Chris Geary,** who initiated the D2EDPM Challenge interoperability work establishing the conceptual frameworks and energised networks of individuals that enabled the successful outcomes of the working groups.

Appendices

Appendix 1

Summary of key requirements for AI research in a SDE / TRE

[SDE-R1]

Provide services for AI research project approval and onboarding pathways

Requirement Summary

The SDE should operate within a standard, repeatable framework, and process for approving and then onboarding AI projects, to ensure a consistent experience for researchers across projects and across SDEs. The process should be streamlined, and support iterative discussions where needed.

Notes:

Existing healthcare specific research protocols can work well for AI – but AI data scientists may not be experienced in them, slowing the onboarding of projects, and causing frustrations and delays in generating new downstream research results. Existing project approval processes (which themselves can vary by Centre), and associated IG reviews of project data requests, may both be improved with better understanding of AI training techniques and processes, to help streamline the request approval process: see also [SDE-R12].

AI often requires large data volumes but increasingly may expect to also require large amounts of data points as part of their project requests, which, in innovative work, may change as understanding of the problem space is refined – and fear of the difficulty or delays to approve amendments can influence scientists to initially over-request to their possible needs – SDEs should support and raise awareness of streamlined data amendment processes.

AI using Federated Learning as per [SDE-R8] may require dedicated processes to support project onboarding.

Example User Stories

As an AI Researcher, I want a consistent framework and common industry practice for project governance and approvals, so that my company can efficiently start new projects in any SDE.

As a SDE Projects Manager, I want efficient processes for reviewing and approving AI projects, so that our SDE can onboard a large number of projects. I also want these processes and practices to be commonly applied, so that they do not form a differentiator between SDEs or a barrier to companies or academic partnerships working with our SDE.

Proposals to support the requirement

- Work with relevant parties in the UK and international SDE space to promote common processes and practices for AI project governance, including efficient data amendment processes.
- Proactively target the new SME / AI scientist community with education resources in established health research protocols and processes.
- Proactively consider issues such as public acceptance, monetisation, rights, public-private partnerships, so that competent discussions are promulgated.

[SDE-R2]

Provide curated de-identified data at scale to project workspaces

Notes:

Definitions and scope of curation may vary. Data linkage between discrete data sets in both simple and complex ways may be expected. Healthcare AI has a larger focus on imaging and free text than normal analytics / public health projects workspaces.

Data storage and transfer sizes per project may be large e.g., 20,000 CT studies may consume 1TB or more of raw storage – plus potentially a similar amount of disk storage allowance for intermediate artefacts / versions of the images prepared for training, and further allowance for backups.

Researchers need to review the provided data to complete curation of the data prior to training: see also [SDE-R5] – Remote access to Workspace. Some curation activities by researchers are better performed earlier e.g. against statistics drawn from the data before the full data is provided.

Example User Stories

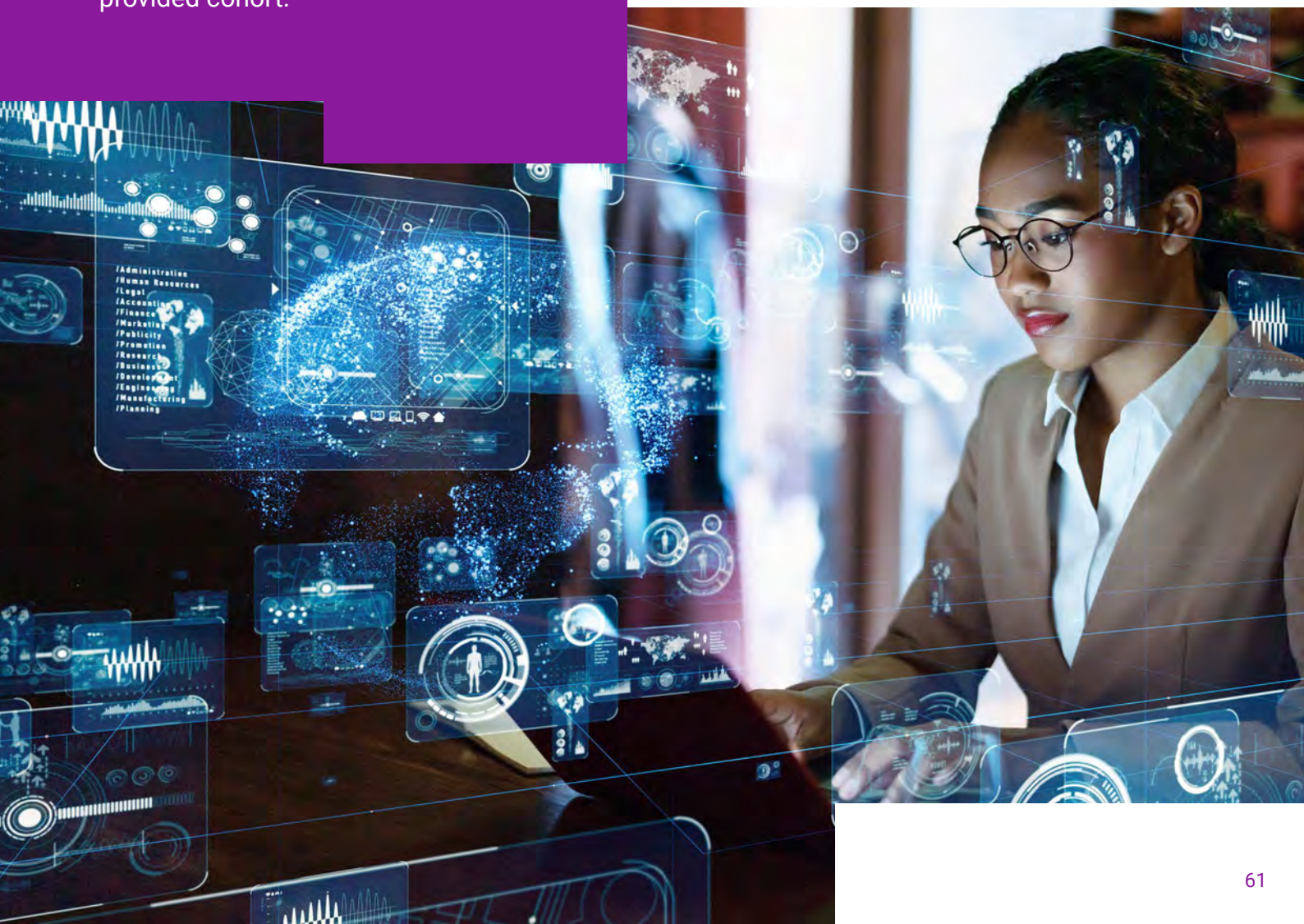
As a SDE Projects Manager, I want tools to create curated de-identified datasets for projects in a repeatable way, so that I can provide data for AI projects at scale with the expected data privacy rules applied.

As an AI Researcher, I want access to data that is curated (organised, linked, filtered for suitability), so that I can start exploratory data analysis, and AI model training.

As a Clinical Researcher, I want the ability to ensure the data is suitably curated for the project, so that the AI model we are working on has the best chance to train effectively against the provided cohort.

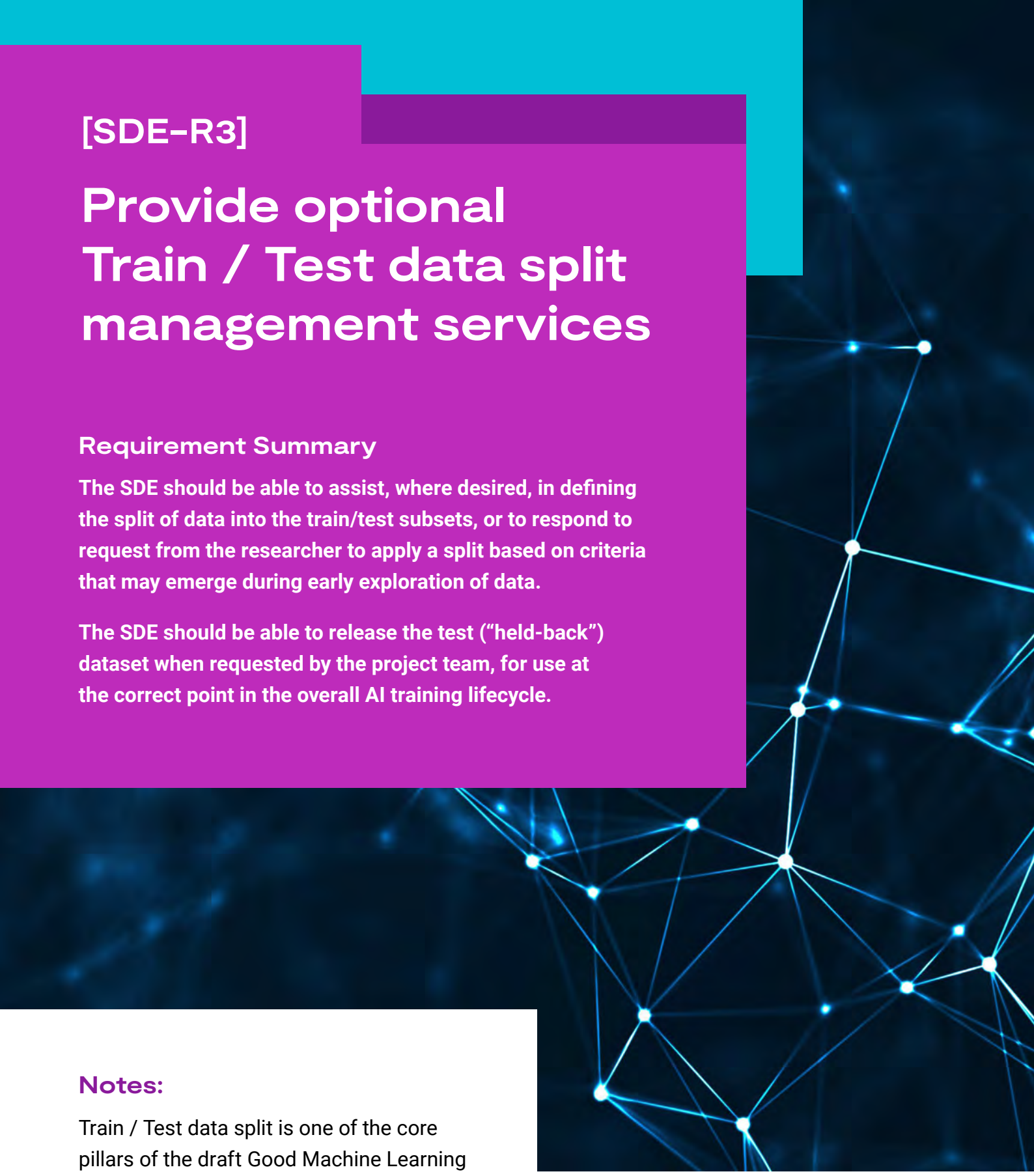
Proposals to support the requirement

- Procure / re-use, develop or collaborate with other healthcare IT stakeholders to create a rich and re-usable set of interfaces or 'pipelines' where healthcare data from diverse sources such as PACS, VNA, EMR, EHR, Lab systems, Prescribing, Public Health datasets etc. can be imported, linked, and curated.
- Use (procure / re-use, develop or collaborate to acquire) robust de-identification tools and services which automatically remove obvious PII such as patient name, but also indirect PII such as visit locations, photos, occupations etc. while ensuring that patient data received from multiple sources maintain correct links and context.



[SDE-R3]

Provide optional Train / Test data split management services



Requirement Summary

The SDE should be able to assist, where desired, in defining the split of data into the train/test subsets, or to respond to request from the researcher to apply a split based on criteria that may emerge during early exploration of data.

The SDE should be able to release the test ("held-back") dataset when requested by the project team, for use at the correct point in the overall AI training lifecycle.

Notes:

Train / Test data split is one of the core pillars of the draft Good Machine Learning Practices from MHRA, Health Canada and the FDA.

A "blind" separation of the data up-front by the SDE is not always effective in ML training: the project scientists and clinical researchers may need to further curate the provided data, and then potentially request the SDE to remove / hold back some of that data, for later release at the appropriate time during the training lifecycle.

Example User Stories

As a Clinical Researcher, I want my definition of data splits to be applied to the SDE data prepared for me so that I am ready to begin data curation immediately when the project is ready to start.

As an AI Researcher, I want a clinically driven split of data, so that the best statistical evaluation of new models can be performed when the training is complete using a hold-out test set and so that I can ensure the training has not been overfitted to the training data.

As a SDE Projects Manager, I want to assist projects in separating train / test data that I provide them, so that AI training results generated in our SDE are understood to have used scientifically valid evaluation methods.

Proposals to support the requirement

- Include processes and software to support and evidence the holding back of a test data set for use in the “run once” end-of-training stage of the project.

[SDE-R4]

Provide fully isolated project workspaces for each research team

Notes:

AI Researcher innovation and IPR are contained within many aspects of their research, including the code they use and the data they select for inclusion in their project. AI Researchers expect full isolation of their work from other researchers in most cases – but also need means to collaborate in broader project teams when needed.

SDEs must enable good working environments for their researchers, but also ensure researchers cannot breach their workspace isolation and affect any other projects or processes within the SDE.

The storage disk supporting each project workspace may be large – in addition to access to large cohorts of data, processing the source data for training may itself consume significant working space on disk – as well as significant time, depending on the compute and storage / IO performance provided. Similarly, checkpointing of large models can also require significant additional storage during the AI project's lifetime. See also [SDE-R7].

Example User Stories

As an AI Researcher, I want my workspace to be restricted to project members, so that I can be sure that processed data, AI models, etc. are not accidently corrupted or inappropriately shared and that any IP in my code is kept private.

As a SDE Projects Manager, I want to have a solution that ensures complete segregation of researcher's workspaces from each other and from other SDE related systems, so that I have confidence that researchers IPR and other SDE related systems are all fully protected.

Proposals to support the requirement

- Use appropriate technologies, such as containerisation, virtualisation, network controls, authentication, and secure hosting to isolate workspaces.
- Ensure audit trails are available on request for workspace accesses, imports, and exports, including accesses made by SDE personnel.
- Provide tiered storage at different performance levels, to ensure maximum cost/benefit ratio from large storage vs fast access storage. Factor in storage for original data, artefacts of training, and audit log of model exports.

[SDE-R5]

Provide highly performant remote access for research teams to their workspace

Notes:

AI Research is a time consuming and iterative process of inspection, adjustment and review which requires a high-quality remote access service. Increasingly, AI researchers will use model statistic rendering visual tools that require 3D viewing and manipulation in real time.

AI Research on medical data usually includes a process of review and annotation of existing cases by skilled clinicians, who will be remote to the SDE. In some cases, the remote access needs to be of the same quality as is required for diagnostic image reporting (e.g., for 3D imaging modalities).

Example User Stories

As a Clinical Annotator / Clinical Evaluator, I want to be able to view and manipulate moving images (e.g., CT or MR series), so that I can undertake the project without distractions and frustrations from poor quality or intermittent access.

As a SDE Projects Manager, I want to offer a QoS guarantee for the bandwidth and latency of our remote access solution, so that our projects can progress smoothly through their Clinical Annotation and Evaluation stages.

Proposals to support the requirement

- Provision dedicated Remote Desktop Services with QoS guarantees on latency, bandwidth, and resolution.
- Offer tiered profiles for remote access services, to ensure users who require it have (or can budget extra to receive) the necessary high-quality connections.

[SDE-R6]

Enable routine import of code for ML experiments

Requirement Summary

SDEs must provide means to allow import of frequently changing Experiment code (e.g., containers, GitHub – high rate of change).

Notes:

This approach aligns well with container technologies, which are themselves well adopted in the AI research community, although other technical solutions can also provide this capability.

There may be technical difficulties in linking GitHub to a SDE in an IG compliant way – GitHub resolves conflicts between code versions on the client.

Example User Stories

As an AI Researcher, I want the ability to frequently and rapidly (potentially a few times every hour, ideally taking no more than 1 minute) change the code running within the SDE, so that I can develop code locally if I wish, and fix bugs and problems quickly to allow experiments to continue.

As a SDE Projects Manager, I want to provide practical, yet safe ways for researchers to import code, so that researchers can use the SDE without incurring significant productivity loss compared to working outside of a SDE.

Proposals to support the requirement

- Establish an unsupervised, one-way method of importing raw or compiled code (binaries) and other development environment components (software tools, editors, compilers etc.) into the SDE so that AI researchers can develop code in the SDE with the tools of their choice. The environment in which this code and tools can run needs to be air-gapped from other SDE systems.
- Ideally: establish a two-way link between the software development area of the SDE and an external source repository, so that standard source control practices can be followed. Ensure that only the researcher's code, and no sensitive data, can leave the SDE in this way.

[SDE-R7]

Provide managed access to powerful ML training infrastructure

Notes:

The infrastructure needs to enable access to compute as supported by GPU, CPU, or both (varied requirements). Fast storage and disk access / IO throughput are necessary to load data quickly enough to support typical Deep Learning projects.

Access may be required for long training periods (e.g., for several weeks), as well as enabling access to powerful compute for shorter experiments.

SDEs may use a mix of on-prem dedicated hardware, private cloud-hosted instances, or a hybrid approach to satisfy the cost / performance trade-offs that come with this requirement.

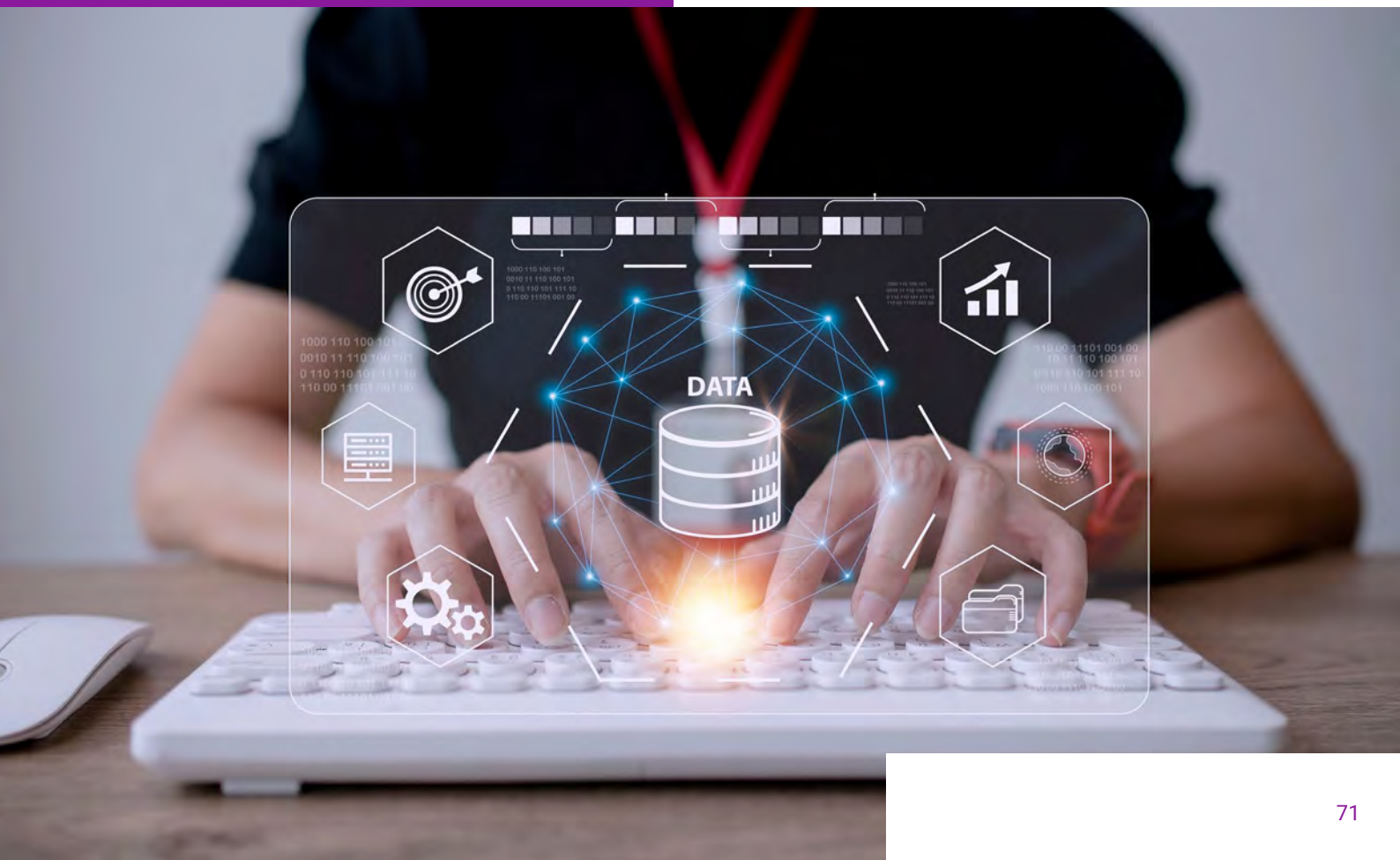
Example User Stories

As an AI Researcher, I want access to sufficient processing power (GPU and/or CPU), RAM, and storage/throughput, so that I can train ML models effectively.

As a SDE Projects Manager, I want to be able to manage and share our available compute and storage fairly across the current research projects, so that the projects running within the SDE can all progress effectively in parallel and within the cost profile I expect for their resource usage.

Proposals to support the requirement

- Provide a mechanism to enforce resource limits for project's access to resources such as CPU, memory, and storage, aligned to their expected usage / funded profile.
- Provide a mechanism to book exclusive use of, or time-sliced access to, discrete expensive resources such as GPUs or fast storage/access.
- Utilise different tiers of hardware (GPU, CPU, storage, transfer IO) to facilitate effective use of resources for different projects and for projects at different stages.
- See also: related [SDE-R4] about maintaining isolation and separation of data between different projects, and associated data storage sizes.



[SDE-R8]

Support Federated Learning projects and processes with other SDEs

Notes:

Federated Learning is a powerful technique that allows a Machine Learning model to be trained concurrently but asynchronously on different subsets of data and iteratively combined to yield a model with approximately the same quality as if exposed to the entire data at once. This allows AI algorithms to be developed using data that is spread around different universities, hospitals, or other facilities without the need to pool such data. It is a key benefit of the SDE model to enable AI development at scale.

Federated Learning is an emergent standard within Machine Learning, that is expected to evolve in the same time frame as a network of international SDEs.

[SDE-R9]

Enable AI Evaluation by clinical partners and export of evaluation results

Notes:

The SDE needs to include a means to transfer AI algorithms in development regularly and securely from a research space to an evaluation space, where they can be tested with e.g., full patient data and assessed by clinicians, all while maintaining appropriate privacy to avoid disclosing patient data to the project's non-Clinician researchers.

Example User Stories

As an AI Researcher, I want to be able to demonstrate AI algorithms in use to Clinical partners, so that I can get feedback from clinicians on the AI accuracy and areas for improvement, as well as quantitative metrics of the accuracy of AI as verified by clinicians, and so that the AI under development can be objectively improved.

As a SDE Projects Manager, I want an easy and safe way to present AI algorithms that are in development for evaluation by clinicians on real-world patient data, so that clinicians can provide the valuable services of evaluation and expert guidance.

As a Clinical Researcher I want to export high level evaluation statistics, plots, and a few examples of AI performance so that I can use them in reports and publications.

Proposals to support the requirement

- Organise the SDE as two areas separated by technical and process controls to ensure good data hygiene: a project research area containing only de-identified data to which AI researchers have access, and a project evaluation area where AI algorithms in development can be applied to patient data (potentially with PII) and assessed by clinicians.
- Provide a means to transfer AI algorithms in development to clinical evaluation, without requiring lengthy certification as medical devices and while observing the guarantee that AI researchers are not allowed (by their role) to see patient data with PII.

Example User Stories

As an AI Researcher, I want to initiate and take part in multi-site projects where the cohort data is split among multiple facilities, so that AI can be developed with the benefit of a cohort size and demographic coverage that's beyond any one facility.

As a SDE Projects Manager, I want SDE infrastructure to support Federated Learning projects in a vendor neutral and interoperable way, so that my facility can be part of multi-site projects without compromising privacy, and so that my choice of SDE software solutioning does not limit the facilities I can inter-operate with.

Proposals to support the requirement

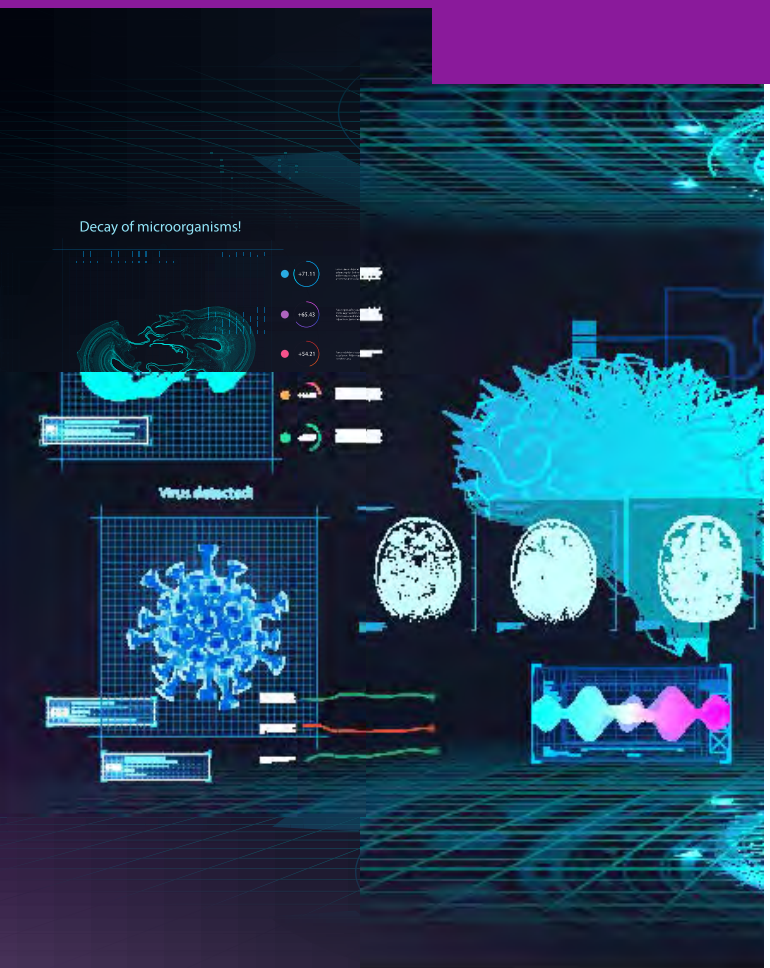
- Include Federated Learning capabilities as a core and interoperable requirement in all SDEs.
- Make minimal technical assumptions to avoid limiting the APIs, frameworks, development tools, or types of AI that can be used with Federated Learning.
- Work with existing or emerging APIs to support Federated Learning and ensure their compatibility and openness.
- Consider also: [SDE-R1] as it relates to necessary agreements between SDE operators; [SDE-R7] to schedule Federated Learning sessions at compatible times on the site's cloud or local infrastructure; and [SDE-11] to retain the relevant archives for a federated project they are participating in or hosting.

[SDE-R10]

Enable AI Model and AI Training results export to project researchers including inspection / release process

Requirement Summary

AI researchers will, at various stages of the training project, want to export one or more artefacts from training, including models (statistics and model 'weights'), charts or graphs, tabular reports, or latest code, for their continued use outside of the SDE.



Notes:

Models can be large and are not always easy to interpret and review.

Ground Truth artefacts created during the project lifecycle are valuable – but not without the data that they refer to (e.g., the original image, original document). Therefore, they are not appropriate for inclusion in normal export processes but may need to be retained by the SDE on behalf of the researcher. See also [SDE-R11] regarding retention of these artefacts.

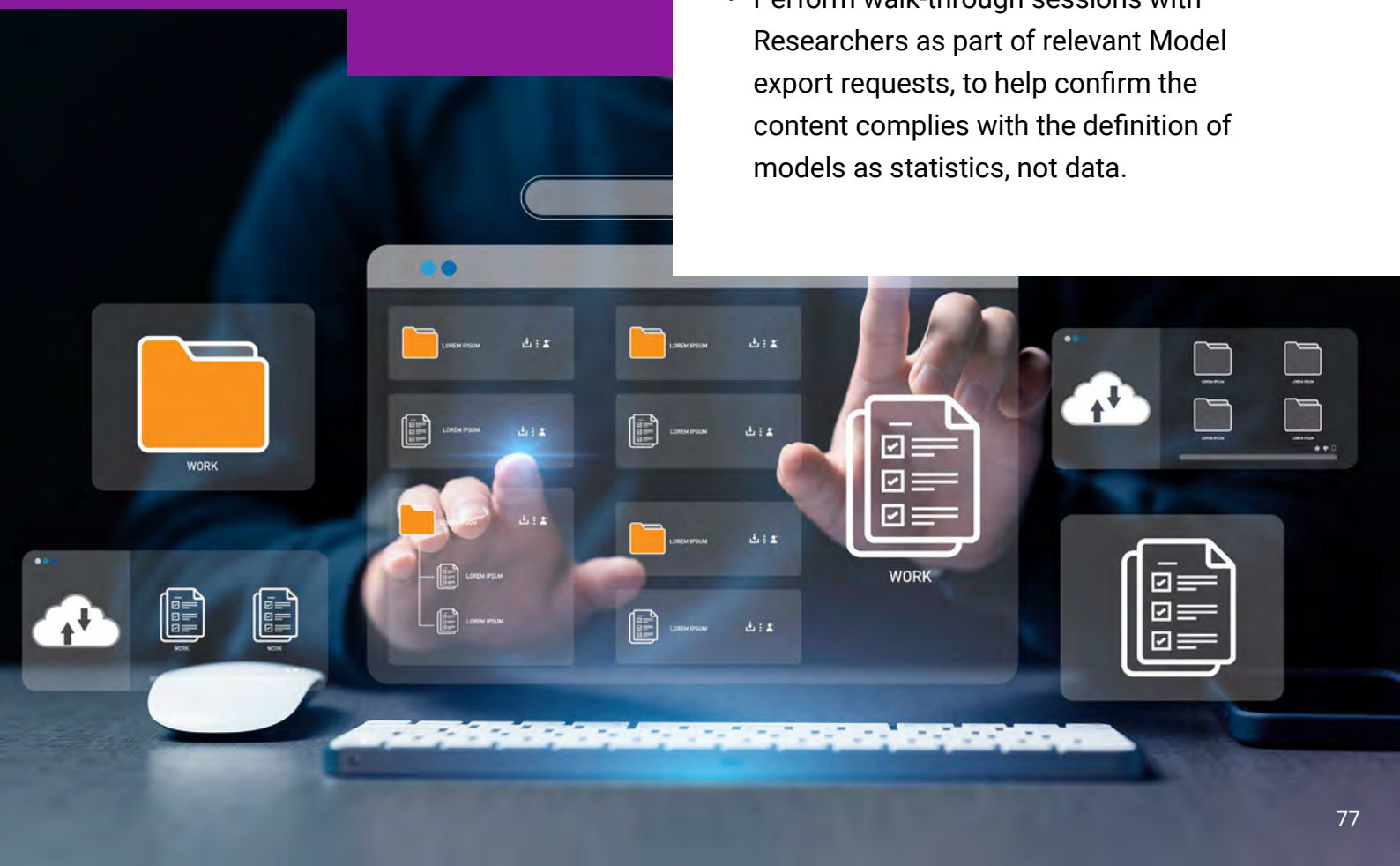
Example User Stories

As an AI Researcher, I want a process to export training outputs such as models, plots, and updated code, so that I can use my SDE training results to create new AI for science, commercial products, and clinical solutions. Additionally, I want to be able to export evidence of good Machine Learning practice, training and verification records, and accuracy metrics so that I can support the process of certification of AI as a medical device.

As a SDE Projects Manager, I want a process to allow me to assure project outputs as safe to release to the project team from a data privacy perspective, so that I can guarantee that the outputs of projects undertaken at my site comply with agreed privacy rules.

Proposals to support the requirement

- Provide a rigorously specified export process that must be followed strictly to export any kind of data from the SDE. Items that are candidates for export may be inspected by artefact type, and by automatic inspection tools as well as, in relevant cases, manual inspection by SDE personnel.
- As much as possible: ensure that all outputs requested from the SDE are of an open format that facilitates inspection rather than use an opaque format. For supporting evidence such as training results, clinician feedback, etc. use simple open formats where data hygiene can be easily verified.
- Offer supporting evidence to researchers of the training performed within the SDE as additional content for their subsequent Regulatory Compliance technical files.
- Perform walk-through sessions with Researchers as part of relevant Model export requests, to help confirm the content complies with the definition of models as statistics, not data.



[SDE-R11]

Provide long term archive for key project data

Notes:

For reasons that may include the support of regulatory compliance, business needs of companies developing AI algorithms commercially, economies such as re-using original cohort data and ground truth, and to avail of technical improvements in Machine Learning or cohort availability over time, the SDE needs to ensure availability of ‘hot’ and ‘cold’ data of various types, and according to a retention policy agreed in advance with the stakeholders.

Example User Stories

As an AI Researcher, I want to access old project data if I need to so that, for example I can make further improvements on previously developed AI algorithms, I can expand or merge cohorts to develop AI at greater scale, I can investigate or troubleshoot scenarios that require revisiting past data, and generally pursue business and academic goals that do not start and end within one AI development project.

As a SDE Projects Manager, I want to be able to offer a clear and detailed data retention policy and have it agreed with stakeholders, so that I can plan and provision resources, and so that stakeholders such as data owners and AI development organisations can be assured that the SDE will continue to meet their future business needs.

Proposals to support the requirement

- Develop and publish a clear data retention policy in consultation with stakeholders (data owners, AI researchers, companies). Although this remains the responsibility of each SDE site, establishing common norms across the sector will help set expectations and further the acceptance of the SDE model.
- Data retention may be segmented by data type and purpose (cohort source data, project-generated ground truth, metrics, and results, etc.) and may stipulate that data be immediately available ('hot') or made available by arrangement ('cold').
- Consider also: a retention policy may evolve into a long-term guarantee of service agreement covering other SDE provided resources beyond the actual data.

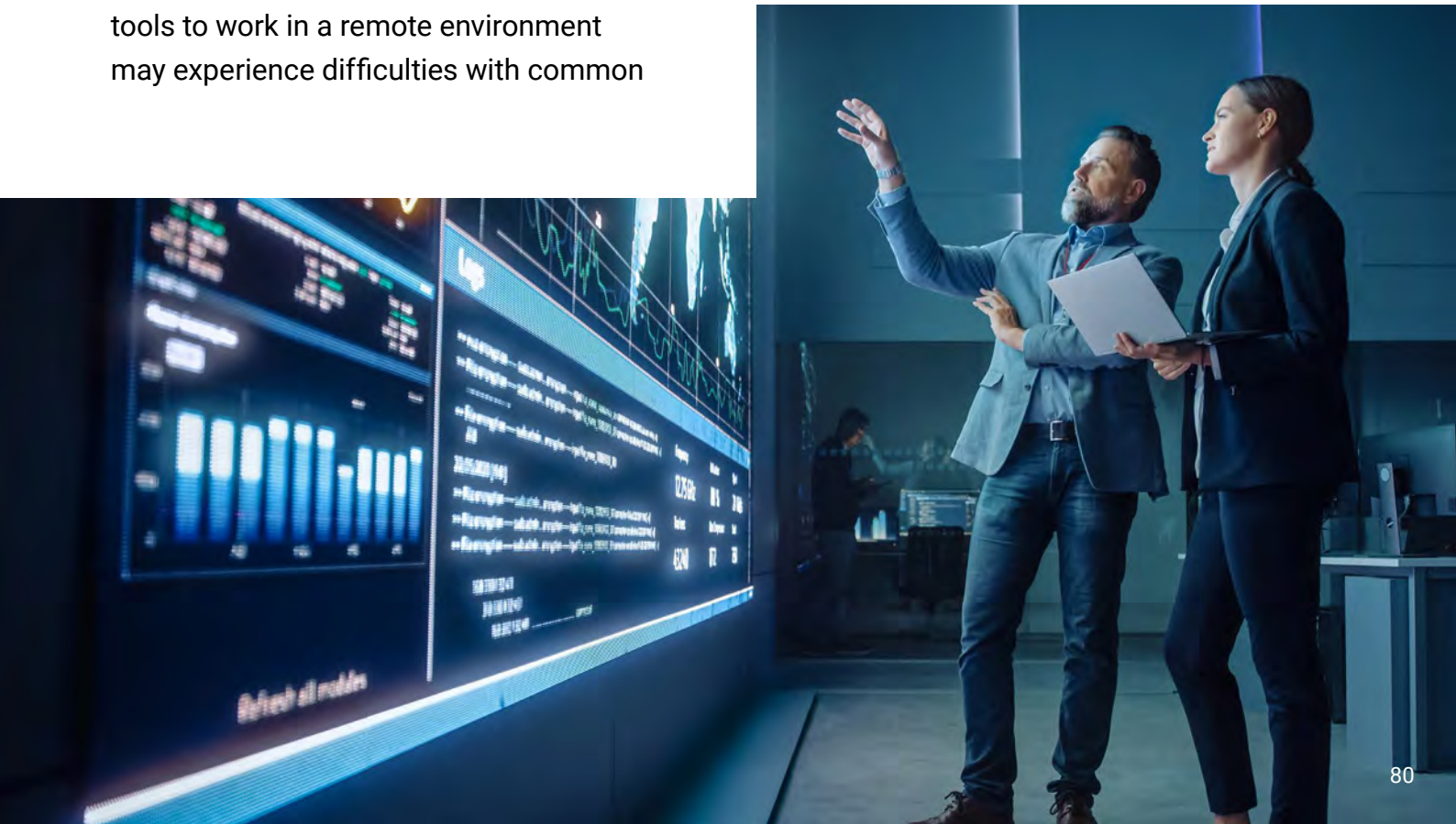
[SDE-R12]

Provide specialist personnel to support the full lifecycle of AI research projects

Notes:

The SDE service provider needs to make available skilled personnel to carry out project related tasks, such as overall project screening and approval, data curation, data de-identification review, and approval of (inspection of) artefacts for export.

Researchers translating their code and tools to work in a remote environment may experience difficulties with common ML tools and libraries that they use, or with specific e.g. CUDA or other GPU-related drivers associated with the SDE's infrastructure, and specialist knowledge of 'ML Ops' may sometimes be required to help unblock such projects within the SDE.



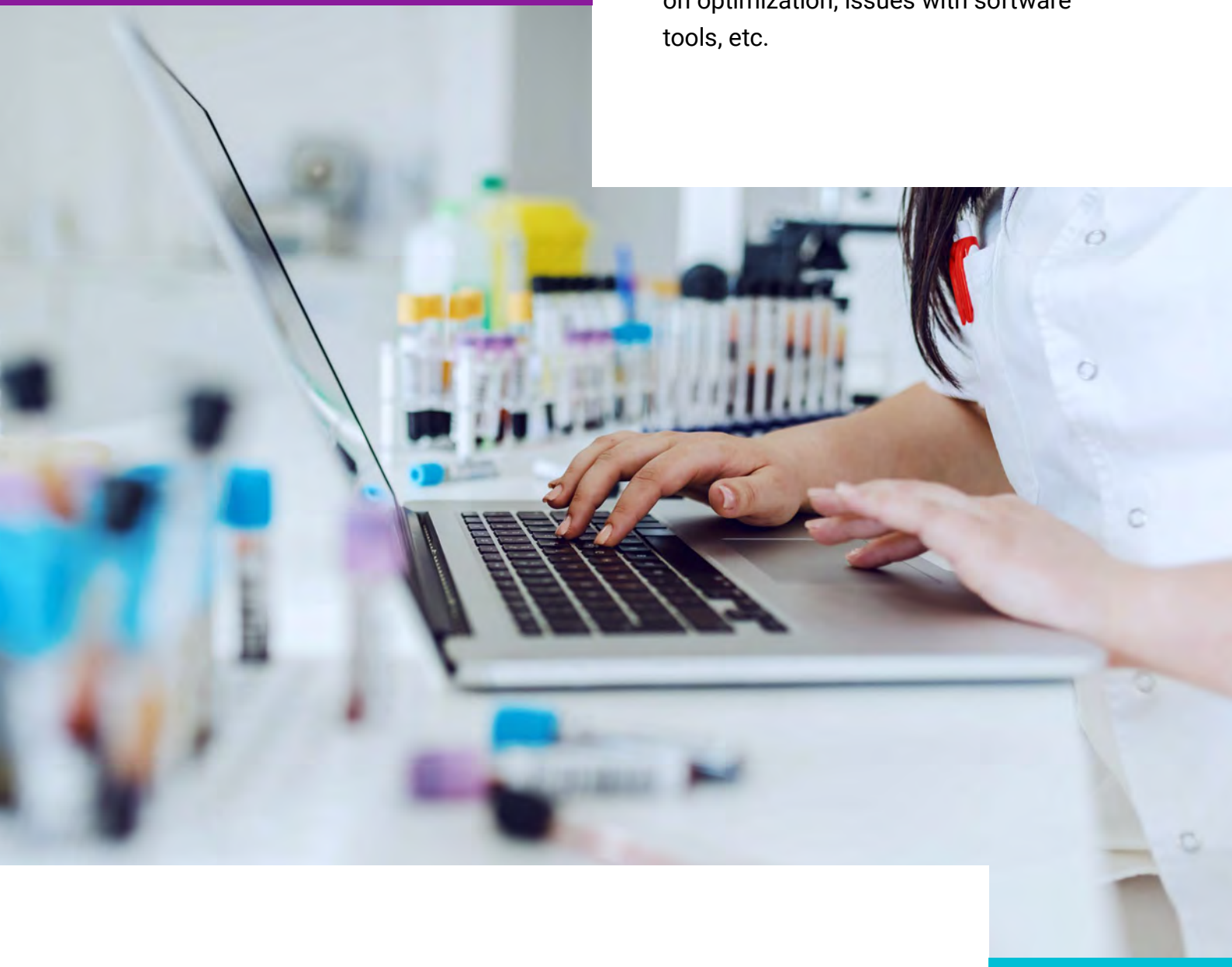
Example User Stories

As an AI Researcher, I want to liaise with knowledgeable staff at the SDE, so that we can resolve any training, export, or data curation issues I meet during my project.

As a SDE Projects Manager, I want to ensure all aspects of my project are running well with researchers, so that the SDE can deliver valuable results from the projects it runs.

Proposals to support the requirement

- Ensure trained subject-matter experts are retained as part of the SDE with knowledge of AI / ML training principles and techniques and an understanding of the scientific methodologies and approaches commonly used, to facilitate the initial review and approval of projects, and the eventual export and release of trained models from the SDE to the researchers.
- Provide technical support services by scientists and engineers knowledgeable in machine learning to advise with topics such as resource provisioning, guidance on optimization, issues with software tools, etc.



Delivered by
Innovate UK and MRC

For further information on the contents of this report contact:

Overall report, D2EDPM Challenge background

Tim Padgett tim.padgett@iuk.ukri.org

Matt Hoare matthew.hoare@iuk.ukri.org

For any information relating to the workstreams please contact Matt or Tim for an introduction to the workstream leads below:

Approach to SDEs Mark Beggs

Deidentification of data John Zurowski

Federated Learning Jorge Cardosa

Evaluation of AI algorithms John Zurowski

Quality Assurance of Digital Pathology David Brettle

Image Standards for Clinical Use Craig Sawyers

Next Generation File Formats Shan Reza

ukri.org/councils/innovate-uk

03003 214357 • support@iuk.ukri.org • @innovateuk

Innovate UK, Polaris House, North Star Avenue, Swindon, SN2 1FL