



**British Heart Foundation
Data Science Centre**

Led by Health Data Research UK

Workshop report – How can we use imaging data to better understand cardiovascular disease?

JULY 2022

Workshop report – How can we use imaging data to better understand cardiovascular disease?

Table of Contents

CONTENTS

Table of Contents.....	1
1 Executive summary.....	3
2 Background and aim	3
3 Workshop summary.....	4
3.1 Workshop organisation.....	4
3.2 Workshop sessions.....	4
3.3 Workshop participants.....	5
4 Workshop sessions in detail.....	7
4.1 What are the most important research questions that cardiovascular imaging can be used to address?	7
4.2 What are the biggest challenges for using imaging data in cardiovascular research?	9
4.3 What are the requirements for a pipeline to support cardiovascular imaging research? ...	12
4.4 How can federated analysis and federated learning be used to improve cardiovascular imaging research?	15
4.5 What are the views of patients and the public on using imaging data in research?	17
4.6 How can we best communicate imaging data research to patients and the public?	18
5 Next steps	18
5.1 Community engagement.....	19
5.2 Prioritisation exercise for cardiovascular imaging research	19
5.3 Communication of cardiovascular imaging research to patients and the public.....	19
5.4 Driver projects.....	19
5.5 Cardiovascular imaging based open challenges	19
6 Conclusions	20
7 Acknowledgements.....	21
8 References	22
9 Additional material	23
Appendix A – Agenda.....	23
Appendix B – Attendee list.....	25
Appendix C – Mural board images from session 1 “What are the most important questions that cardiovascular imaging can be used to address?”	28

Appendix D – Mural board images from session 1 “What are the top 3 criteria for scoring the research questions?”	33
Appendix E – Mural board images from session 2 “What are the biggest challenges for using imaging data in research?”	38
Appendix F – Mural board images from session 2 “What ideas do you have to resolve these challenges?”	43
Appendix G - Mural board images from session 3 “What are the requirements of a pipeline to support cardiovascular imaging research?”	48
Appendix H - Mural board images from session 4 “What types of research questions would a federated approach allow us to address that we can’t do now?”	53
Appendix I - Mural board images from session 4 “What are the advantages and challenges of a federated approach?”	58
Appendix J - Mural board images from session 4 “What ideas do you have to resolve these challenges?”	63
Appendix K - Mural board images from public and patient representatives’ session “What do you think are the benefits of imaging data being available/used in research?”	68
Appendix L - Mural board images from public and patient representatives’ session “What are your concerns, if any, about imaging data being used?”	69
Appendix M - Mural board images from public and patient representatives’ session “What do you think are good ways of explaining imaging data research? Can you suggest specific words to use or not use, and the kinds of formats we should consider for effective communication?”	70

1 Executive summary

Cardiovascular imaging research has huge potential to improve patient care and the UK is world-leading in its cardiovascular imaging research. However, the full realisation of this potential will require the harnessing of new large-scale data science resources, linkage to other health data, and novel analysis and machine learning techniques.

The British Heart Foundation (BHF) Data Science Centre Imaging Workshop brought together a wide range of stakeholders to explore the opportunities and challenges to improve cardiovascular imaging research, and address the central question of *How can we use imaging data to better understand cardiovascular disease?*

The insights gathered in the workshop have been used to guide our plans in this area, with the aim of advancing the use of imaging data in cardiovascular disease research to improve patient care. We plan to identify and obtain consensus on priorities in cardiovascular imaging research, to address key challenges in delivering these such as data access, storage, processing, and linkage to health data, and to engage with patients and the public to obtain and maintain trust.

2 Background and aim

Cardiovascular disease remains the leading cause of death globally [1] and in the UK 7.6 million people are living with heart and circulatory disease [2]. The use of cardiovascular imaging in patient care has increased dramatically over the past 20 years, and today there is widespread use of techniques such as echocardiography, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), electrocardiography (ECG), chest x-ray and diagnostic invasive coronary angiography (ICA). In addition, new techniques are emerging in cardiovascular disease assessment including retinal imaging and electroanatomic mapping. These modalities give detailed information on the structure and function of the heart, brain and blood vessels and are central to diagnosis and management of patients with cardiovascular diseases.

Cardiovascular imaging is widely used in cardiovascular disease research. It is used to better understand disease phenotypes, improve risk stratification, and assess the impact of treatments, with imaging markers used as endpoints in clinical trials. However, there is potential to improve how we use cardiovascular imaging in research. Combining imaging data with other healthcare information, such as healthcare utilisation, clinical outcomes, and medication use, would significantly augment its utility. Data science is a central component of these research activities, with machine learning (also called artificial intelligence or AI) showing great promise over recent years for the analysis of large datasets and direct analysis of images.

Establishing imaging datasets with linkage to healthcare records to support large-scale cardiovascular research presents many challenges. These include linking imaging data to other healthcare data, accessing, storing, processing, cleaning, and annotating imaging data, and developing and using machine learning techniques. Addressing these barriers through leveraging the experience of the UK's world leading cardiovascular imaging researchers, harnessing previous experience of linking other types of healthcare information, and developing cross disciplinary collaborations presents an opportunity to advance mechanistic understanding, prediction, prevention, and treatment of cardiovascular diseases through large-scale cardiovascular imaging research.

The British Heart Foundation (BHF) Data Science Centre organised a workshop to understand the key opportunities and challenges associated with using imaging data in cardiovascular research, and what steps can be taken to realise and address these. The objectives of the workshop were to:

- To identify the key research questions that cardiovascular imaging should be used to address
- To understand the key challenges of using imaging data in cardiovascular research, including technical, analytic, and regulatory issues
- To assess the potential for federated analysis and federated learning to enhance cardiovascular imaging research
- To create a UK wide network of diverse stakeholders interested in using imaging in cardiovascular research

3 Workshop summary

3.1 Workshop organisation

The workshop was held virtually on the afternoon of 17th November and the morning of 18th November 2021. It brought together representatives from a wide range of relevant stakeholder groups. Five patient/public representatives also attended the workshop.

3.2 Workshop sessions

The workshop was organised around four sessions, each centred on an important question (Table 1). Each session featured talks from invited speakers followed by an interactive breakout session to facilitate broad and open discussion and to gather input from all participants. During the breakout session, workshop participants were split into 5 groups, each with a chair and facilitator. Participant input was captured using virtual whiteboards, using an online tool called Mural. During interactive breakouts of Sessions 3 and 4, the patient/public representatives joined a separate breakout group to discuss questions of specific interest to the patients and public (Table 2). The final session summarised the discussions and featured a chaired panel discussion around identifying next steps. We also organised discussion sessions with the patient/public representatives before and after the workshop.

Table 1. Workshop session topics and breakout session questions.

Topic	Aim	Breakout session questions
1. What are the most important research questions that cardiovascular imaging can be used to address?	Build consensus, and define key research questions	What are the most important research questions that cardiovascular imaging can be used to address? What criteria should be used for prioritisation?
2. What are the biggest challenges for using imaging data in cardiovascular research?	Identify the biggest challenges for using imaging data in research	What are the most important research questions that cardiovascular imaging can be used to address? What criteria should be used for prioritisation?
3. What are the requirements for a pipeline to support cardiovascular imaging research?	Identify the challenges and barriers for an analysis pipeline supporting cardiovascular imaging research	What are the requirements for a pipeline to support cardiovascular imaging research? What could be done to deliver such a pipeline?
4. How can federated analysis and federated learning be used to improve cardiovascular imaging research?	Build consensus on how federated analysis/learning could be used to support cardiovascular imaging research	Which types of cardiovascular imaging research questions could be addressed with a federated learning/analysis approach? What are the advantages and challenges of a federated learning/analysis approach? How might we resolve these challenges?

Table 2. Public and patient representative breakout sessions (Sessions 3 and 4, Day 2).

Topic	Aim	Breakout session question(s)
3. What are the views of patients and the public on using imaging data in research?	To understand patient and public opinions on the benefits and concerns of using imaging data in research.	What are your concerns, if any, about imaging data being used? What do you think are the benefits of imaging data being available/used in research?
4. How can we best communicate imaging data research to patients and the public?	To understand patient and public opinions on the best ways to communicate imaging research.	What do you think are good ways of explaining imaging data research? Can you suggest specific words to use or not use, and the kinds of formats we should consider for effective communication?

3.3 Workshop participants

Workshop participants represented a wide range of stakeholder groups, including imaging, cardiovascular, data science and computer science researchers, NHS professionals, representatives from NHS organisations, relevant societies and companies, data custodians and patient/public representatives (Figure 1). Participants used a variety of established imaging techniques in clinical practice or research (Figure 2), as well as new and novel techniques such as retinal imaging and electroanatomic mapping.

Figure 1. Stakeholder groups represented by workshop participants.

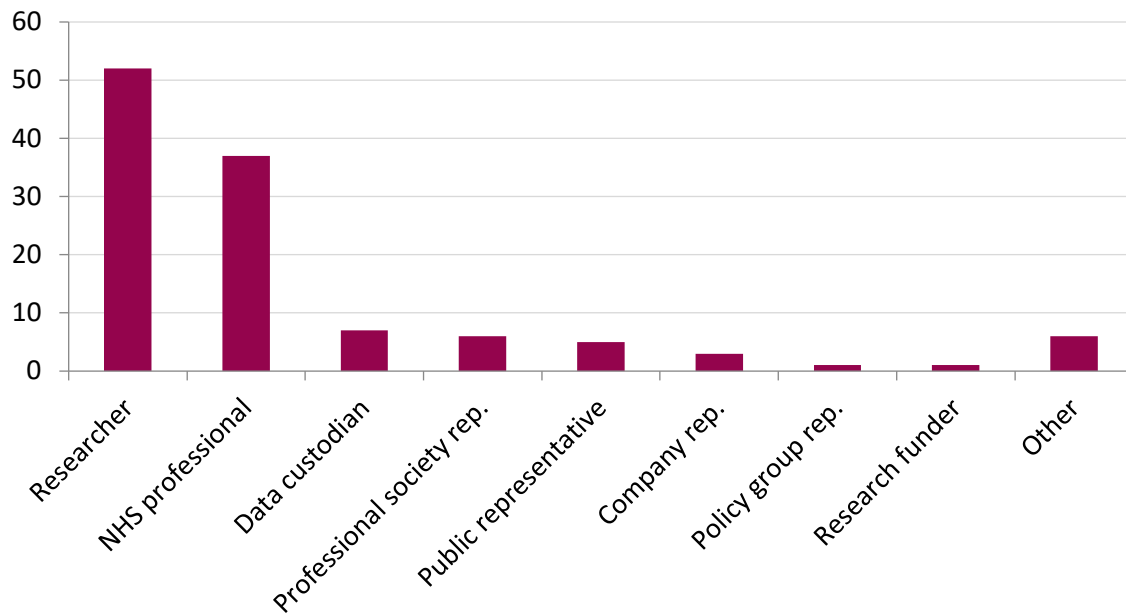
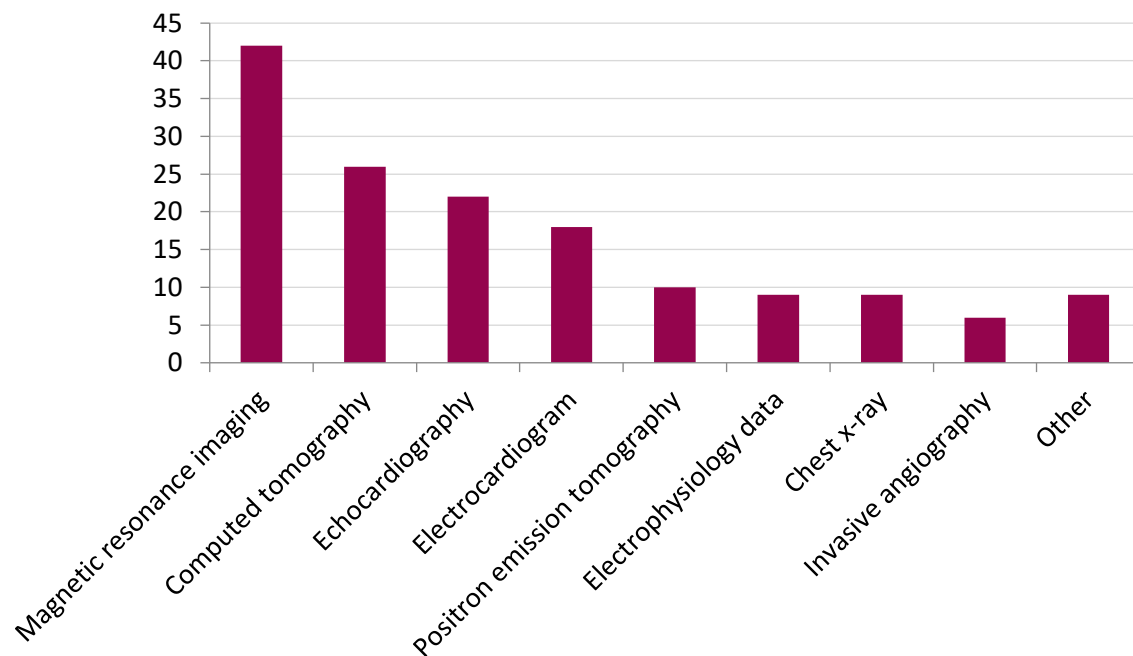


Figure 2. Imaging techniques used by participants in research and clinical practice.



4 Workshop sessions in detail

4.1 What are the most important research questions that cardiovascular imaging can be used to address?

The first session focused on identifying the most important research questions that cardiovascular imaging research should be used to address. Cardiovascular imaging includes anything that creates an image of the heart or blood vessels, or the brain in connection with stroke. This involves a wide variety of types of imaging and many potential research questions. This session aimed to begin to develop consensus regarding the most important research questions that will help guide the focus of the BHF Data Science Centre, and of other research groups and funders.

The UK's world leading position in cardiovascular imaging research was demonstrated by Professor Jeremy Pearson's (BHF Associate Medical Director) talk on "Cardiovascular research imaging in the UK – the BHF's contribution". He cited significant improvements in imaging technologies, clinical practice, patient care and NHS capacity due to cardiovascular imaging research that had been performed in the UK.

Professor Marc Dweck (University of Edinburgh) presented on the wide range of imaging modalities that are now available for cardiovascular imaging, providing detailed information on anatomy, function, and disease. Different characteristics of each imaging modality make them suited to tackle particular research questions, and combinations of imaging modalities can help develop a greater understanding of disease.

Dr Michelle Williams (BHF Data Science Centre Associate Director) introduced the BHF Data Science Centre's plan to carry out a prioritisation exercise to answer the question "*What are the most important research questions that cardiovascular imaging should be used to address?*". Building on the success of previous prioritisation exercises in this area [3], we will be using a modified Delphi technique to collect expert judgements and identify consensus among stakeholders. The first stage of this process involves gathering suggested research questions from a broad range of stakeholders, before two rounds of prioritisation. Prior to the workshop, participants were asked to submit potential research questions and over 150 diverse questions were submitted (Figure 3). During the breakout session, participants were prompted to provide and discuss potential additional questions and encouraged to think broadly in terms of imaging modalities, parts of the body and types of research. This will form the basis of the planned prioritisation exercise to be performed in 2022.

Several themes emerged across the breakout group discussions, including using imaging to improve diagnosis and risk prediction, maximising the use of existing data and linking to other datasets, ensuring space for blue-sky research and development of new imaging technologies, understanding disease mechanisms and normal cardiovascular health, developing improved imaging and image analysis techniques, developing new evidence based machine learning techniques, advancing the understanding of multisystem disease, and improving patient focused care. It was highlighted that national imaging datasets and information about what imaging is being performed would be valuable to improve research in this area. The patient/public representatives highlighted the importance of considering patient focused aspects of research such as the patient experience and making scans safer, shorter, and less frequent. The prioritisation exercise aims to help direct the activities of the centre, but it was highlighted that we should also make room for "blue-sky" research, as we never know where the next innovation in imaging may come from.

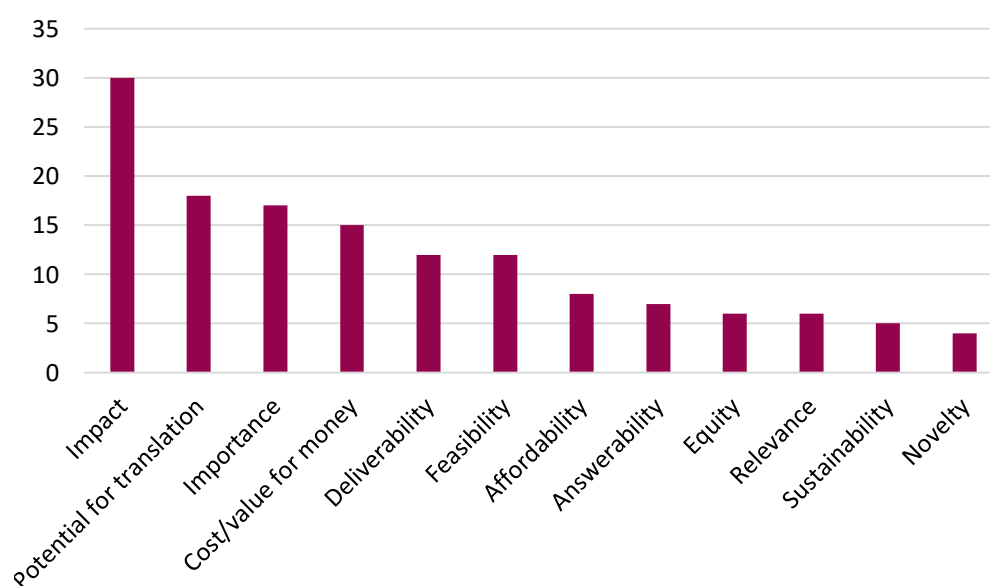
Figure 3. "Word cloud" generated based on the research questions submitted by workshop participants as part of the pre-workshop survey.



Participants were then asked to consider what domains should be used to prioritise the research questions. There are a wide variety of ways that research can be prioritised, and participants were presented with 12 commonly used domains: affordability, answerability, cost, deliverability, equity, feasibility, impact, importance, novelty, potential for translation, relevance, and sustainability. They were asked to suggest any additional domains that they thought were missing and to vote for the three they thought were the most important (Appendix D).

The number of votes for each prioritisation domain, summed across the five breakout groups, is shown in Figure 4. Impact, potential for translation and importance were the highest scoring domains. The additional domains of “cost effectiveness” and “value for money” were suggested. The votes for these domains were combined with those for “cost”, and together these were the fourth most voted for domain. The results from this workshop session will be considered when selecting the domains for use in the prioritisation exercise.

Figure 4. Prioritisation domains identified during interactive breakout groups.



4.2 What are the biggest challenges for using imaging data in cardiovascular research?

This session opened with Professor Steffen Petersen (Queen Mary University of London), Dr Joseph Jacob (University College London), and Professor Carola-Bibiane Schönlieb (University of Cambridge) presenting their experiences of using different types of imaging data in cardiovascular research and the challenges involved.

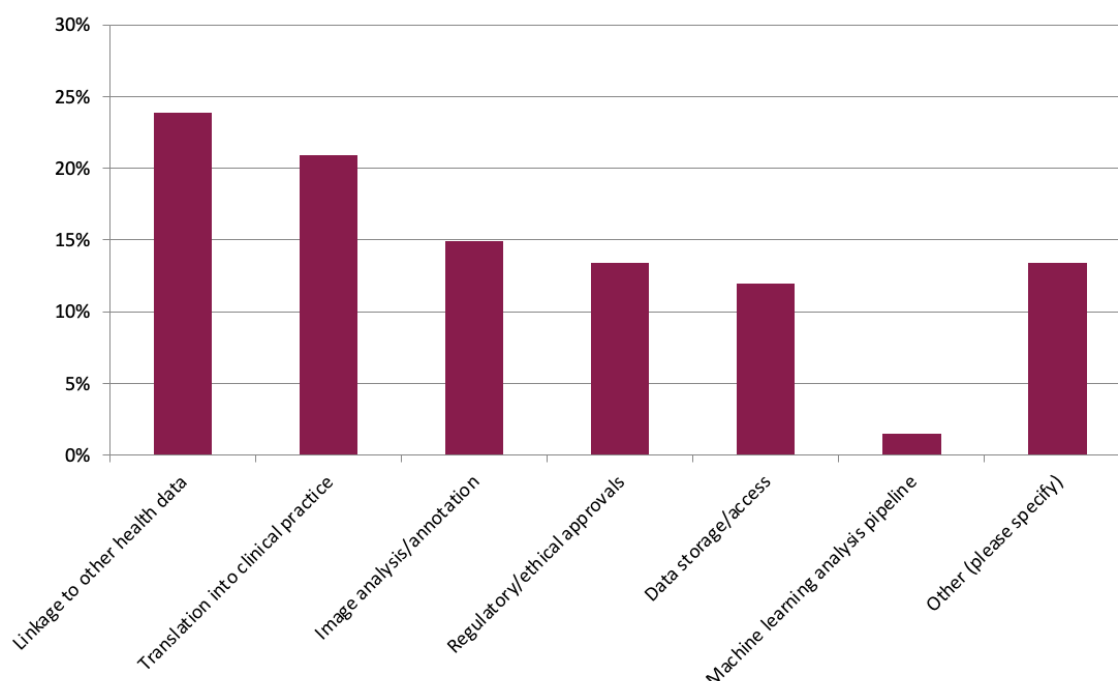
Professor Petersen presented his reflections from the UK Biobank cardiac imaging study [4], which used a standardised research protocol to acquire images from 100,000 participants across multiple imaging centres. Professor Petersen noted that the centralised organisation and consented prospective study design enabled the application of a standardised research protocol with consent to use clinical data for research. Important challenges included the feasibility of scanning 100,000 people, quality control, data processing and data access, and approaches to incidental findings in imaging research.

Dr Jacob presented his experience of setting up the National COVID-19 Chest Imaging Database [5] using routinely performed NHS imaging. An important driver behind this research was the need for access to representative, diverse, national data to support the training and testing of machine learning models during the COVID-19 pandemic. Important challenges included variation in patient identifiers, data curation, removing identifiable information, and the transfer and hosting of large volumes of data.

Professor Schönlieb presented the challenges of using machine learning in imaging research. Many of the challenges are due to real world imaging data being complex, diverse, messy, large, and highly variable. Machine learning needs to be able to cope with these real world data issues. Challenges that need to be overcome are the trade-off between performance and robustness/rigour, generalisability, and computational scalability.

Prior to the workshop participants were asked what they thought were the biggest challenges to using imaging data in research. The most frequent challenges were linking imaging data to other healthcare data, translating research into clinical practice, image analysis and annotation, and gaining regulatory/ethical approvals (Figure 5).

Figure 5. Important challenges highlighted by participants prior to the workshop in an online survey.



During the breakout session participants considered the question *What are the biggest challenges for using imaging data in cardiovascular research?* All statements captured on the virtual whiteboards were manually reviewed and grouped into themes (Appendix E). The nine themes which received the highest number of statements (each more than five statements) are shown in Table 3. Many of the challenges reflect broader challenges of using of routinely collected data in healthcare research. However, cardiovascular imaging specific challenges included standardisation of image acquisition and analysis, issues of diversity and biases in datasets, storage of and access to imaging data, developing and sharing annotations, and infrastructure requirements. The patient/public representatives highlighted important challenges, including improving communication around cardiovascular imaging research, improving the patient experience of imaging, and developing safer, quicker imaging tests that improve patient care.

Table 3. Key challenges for using imaging data for cardiovascular research raised during the breakout sessions.

Challenge theme	Number of statements	Examples
Consideration of patients and members of the public	22	Language used in research applications and grants makes them hard to understand
		Need for safer, quicker tests with low or no radiation dose
		Challenge of getting new techniques into clinical practice to benefit patients.
Standardisation	21	Variations in imaging protocols, imaging sequences, and imaging formats in different locations and at different time points
		Standardisation of image annotation
		Variable patient identifiers and meta-data
Regulatory hurdles and governance	19	Time consuming, repetitive, and poorly understood process and forms for ethics, governance, and regulatory requirements
		Lack of national guidelines covering the use of imaging data for research, particularly for routinely collected data
		Variations between different centres creating new inequalities
Data representativeness and validity	18	Understanding potential biases in imaging data and ensuring diversity
		Risk of under-representation of geographic regions and certain patient populations
		How to handle missing data
Data storage and access	16	Practicalities of image data sharing and storage
		Access to image files vs image derived phenotypes only
		Making research data accessible in an equitable manner
Leveraging AI	13	Developing and sharing image annotations and ground truth
		Testing and training models with real world data
		Developing and maintain AI expertise
Public perception, privacy, security, trust	13	Need to develop trust with all stakeholders, including patients and the public, and ensure good communication
		Consider public perception of sharing data with industry
		Ensure data security
Infrastructure	8	Transferring large volumes of data
		Developing infrastructure enable AI techniques
		Dealing with multiple storage systems and ways of storing images
Linkage and anonymisation	8	Linkage to other health records at scale
		How to capture encoded/encrypted identifiers to facilitate data linkage
		Consistent and thorough de-identification

The second question discussed during this breakout session was potential ideas to overcome these challenges. During the discussion the different types of imaging data were highlighted. These include imaging performed purely for research studies with no feed-back to participants, imaging performed primarily for research studies where the imaging is also part of clinical care, and imaging performed primarily for clinical care which could be used in research. Each of these types of imaging data has potential utility in cardiovascular imaging research. In addition, information about what imaging is performed, imaging reports and imaging meta-data provide valuable insights into cardiovascular disease.

In general, workshop participants found it challenging to provide suggestions to overcome these barriers to cardiovascular imaging research. However, the suggestions provided have broader benefits than just for cardiovascular imaging research and would benefit other types of imaging and data science research, and healthcare in general. Potential ideas included:

- Developing guidelines to help standardise all aspects of imaging research, including image acquisition protocols, post processing and annotation.
- Exploring ways to make the collection, storage, and annotation of imaging data easier, faster and more robust.
- Developing national and regional imaging datasets suitable for researcher access.
- Using information about what imaging is being performed across the country in research as well as the imaging data itself.
- Involving patients and the public in all phases of research, from study design to communication of the results, and beyond.
- Exploring the use of new privacy preserving techniques such as federated learning.
- Improving training and career development in relevant image analysis and machine learning techniques to maintain the highly skilled workforce.
- Developing informatics expertise in both the National Health Service (NHS) and academia.
- Exploring and developing new funding opportunities for this type of research.
- Organising imaging based open challenges on important cardiovascular imaging topics.

4.3 What are the requirements for a pipeline to support cardiovascular imaging research?

Cardiovascular imaging research involves a variety of components, some of which are similar to other data science research areas, but many of which are distinct to imaging research. This session focused on the aspects of a pipeline to support cardiovascular imaging research, and machine learning research in particular. It included talks and discussions around the challenges and potential solutions. This session opened with Professor Emily Jefferson (University of Dundee), Professor Alex Frangi (University of Leeds), and Dr Wenjia Bai (Imperial College, London) presenting their experiences of using imaging data in cardiovascular research and the challenges involved.

Professor Jefferson talked about curating NHS imaging cohorts within a trusted research environment. The use of AI research in clinical imaging has the potential to improve diagnostic accuracy, develop new disease detection methods and increase the efficiency of image analysis. However, at present most research studies use small numbers of research images for development and testing, which limits their generalizability to real world situations. Professor Jefferson talked about her experience with the PICTURES program, which has tackled this challenge by creating a copy of Scottish National Picture Archiving and Communication System (PACS) images from across Scotland's population of >5 million people, making these available within a Trusted Research Environment (TRE). De-identified cohorts of imaging data will be available for researchers within the TRE, with the potential to link to other relevant healthcare information.

Professor Frangi presented on infrastructure considerations for imaging research. He discussed the importance of co-creation of imaging research pipelines, so that they meet the needs of researchers and provide the potential for translation into clinical practice. Key considerations for an image analysis platform are what it will be used for, how it will be accessed, what software will be available and how

it will be maintained. Important challenges are interoperability between platforms and scalability. Starting with a pipeline for a small-scale, high-impact pilot would help develop an understanding of the central components, user requirements and limitations.

Dr Bai talked about pre-processing and image annotation for machine learning research. He presented work investigating how differences in the manually annotated images used in training will influence the performance of models. His talk highlighted the importance of having diverse training datasets and the importance of involving human assessment in the development of machine learning models.

During the breakout session participants were presented with an outline of a pipeline to support cardiovascular imaging research and were asked to consider potential challenges and solutions. Aspects of this pipeline that are common to other areas of health data science research include obtaining funding, ethical and regulatory authorisations, dissemination of results and translation into clinical practice. Imaging specific aspects include image acquisition, sharing imaging, data storage, pre-processing, image analysis, and image annotation. Application of machine learning techniques poses additional challenges, including need for training and testing datasets, and for diverse real world validation of machine learning techniques. Figure 6 illustrates the steps in the outline pipeline with suggestions gathered during the breakout session that would help solve potential challenges at each step. Suggestions from the virtual whiteboards are presented in Appendix G.

An important suggestion from workshop participants was that this process should be circular and iterative, rather than linear. Making sure that the machine learning research that shows benefits to patients becomes part of clinical practice is essential, but monitoring and ongoing training will also be important. Challenges highlighted by the breakout groups included ensuring multi-stakeholder buy-in from the start, communication with funders around the benefits of this type of research, issues of governance and regulatory approvals, methods to access data from PACS systems, developing and sharing image annotations, dealing with incidental findings, training, and career development to maintain the workforce, and considering the best approaches for collaborating with industry. It was also highlighted that annotations or reports performed for clinical practice may be different from those that are optimised for research and combining these can be challenging. Image annotations were highlighted as one of the most challenging, time-consuming, and expensive aspects of machine learning imaging research, and ways to share these annotations with other researchers could facilitate new research findings with benefits to patients. Potential solutions to these challenges are highlighted in Figure 6.

Figure 6. Potential solutions to key challenges at different points of the cardiovascular imaging research pipeline discussed during breakout sessions.

Funding	<ul style="list-style-type: none"> • Prioritisation of the most important research questions • Funding to facilitate better communication between engineers, imaging researchers and clinicians
Authorisations	<ul style="list-style-type: none"> • Approval processes that work across different centres and organisations, so only one application is needed • Standardisation of patient consent and opt-out options
Imaging	<ul style="list-style-type: none"> • Build protocols into imaging devices • Guidelines for standardisation of image acquisition, storage, annotation and pre-processing
Dataset curation	<ul style="list-style-type: none"> • Facilitate curation of well designed multi-centre datasets to be used for multiple research studies • Develop datasets that represent both disease and health
Data storage	<ul style="list-style-type: none"> • Ensure that data is stored in a way that facilitates imaging research, including annotation and machine learning • Explore new storage methods improve access e.g. cloud based
Pre-processing	<ul style="list-style-type: none"> • Keep data close to "real" clinical data • Avoid making datasets too "clean" and training too easy
Analysis, annotation	<ul style="list-style-type: none"> • Standardise and share annotations • Encourage crowd source annotations with incentives to improve engagement
Testing, training	<ul style="list-style-type: none"> • Consider both size and diversity of datasets, and consider potential biases • Explore new techniques such as federated learning
Validation	<ul style="list-style-type: none"> • Validate machine learning tools in real world clinical data • Develop hold-out validation data sets that can be used to test different machine learning tools
Dissemination	<ul style="list-style-type: none"> • Partner with stakeholder groups and professional organisations • Develop mechanisms to safely and securely share models
Clinical Practice	<ul style="list-style-type: none"> • Develop clear pathways for translation into clinical practice • Think about clinical translation from the start • Involve patients and public in communication strategies

4.4 How can federated analysis and federated learning be used to improve cardiovascular imaging research?

Cardiovascular imaging research will benefit from the use of large numbers of imaging tests from diverse populations. However, the challenges of creating large, centralised datasets including the size of the data, and the ability to share and transfer data between sites or countries. Federated approaches have the potential to address these challenges by leveraging the information in large datasets, without the requirement for moving the images themselves. However, these and other privacy preserving analysis techniques have yet to be applied to large scale cardiovascular imaging research.

Dr Jorge Cardoso (King's College, London) explained the background behind federated learning and federated analysis, and why they could be useful for cardiovascular imaging research. He explained that the creation of large, centralised imaging datasets is hampered by several challenges, including obtaining ethics, bureaucracy and anonymisation, and the logistics and infrastructure requirements of moving and hosting large volumes of data. Federated learning overcomes many of these challenges by running the algorithm over the data in locations where the data naturally exists, rather than bringing the data into a centralised location.

Dr Hans-Erik Aronson talked about the challenges of delivering such new analysis techniques. He is leading the UKRI Data and Analytics Research Environments (DARE) program which is aiming to create the next generation of trusted research environments to support research across multiple domains. Key challenges highlighted included data access and discovery, data governance, ethical and regulatory approvals, and hardware and software requirements.

The breakout sessions discussed the types of research questions and study designs where a federated approach could be useful and some of the central advantages and challenges of this approach.

A federated approach could facilitate types of research that are not currently possible and could improve the diversity of current research datasets. However, it was noted that a federated approach may not always be better, and there will be some types of research that are better performed in a single trusted research environment. Federated learning may permit large scale research, expanding beyond individual silos of data. Types of research which were highlighted that could benefit from a federated approach include:

- Enabling research across multiple centres or nations, particularly between countries where data sharing is not possible
- Providing access to more diverse data and a wider range of clinical endpoints
- Increasing the available size of the study population, which may be particularly useful for the study of rare diseases
- Combining healthcare information with other datasets, e.g. pollution data or socioeconomic datasets.

Potential advantages and challenges of a federated approach to cardiovascular imaging research which were discussed in the breakout groups are presented in Table 4. Interestingly, opinions differed between breakout groups on several of the topic areas. For example, some breakout groups thought that regulatory aspects would be easier with a federated approach compared to a trusted research environment, whereas others thought that this would create new regulatory challenges. The technical challenges were particularly highlighted as the cost to install hardware, deploy software, and maintain systems must be considered. Overarching UK wide governance structures and legislation would be required to make federated techniques useful more widely.

Table 4. Advantages and challenges of a federated approach

Theme	Advantages	Challenges
Ethics and governance	Data remains with data controller, potentially reducing regulatory and governance barriers in the future.	Lack of understanding of these techniques and the risks involved means that the complexity of approvals might not be reduced. Current regulatory barriers impact scalability.
Patients, carers and members of the public	Benefits of privacy preservation.	Obtaining and maintaining patient/public engagement and trust. Complicated terms and technologies that need properly explained.
Data sharing	Does not need data sharing between sites, but aspects of the models are shared.	Needs data in a common format with agreed standards.
Data discovery	Potential for more agile, reactive, and saleable identification of research cohorts. Potential improvements in research of rare diseases. Access to large scale diverse data.	Working with linked datasets is more difficult. May miss patient groups living in areas where there are not the resources to implement federated learning.
Data curation	Encourages collaborations across different groups, including international centres.	Curation, pre-processing and quality control are more challenging or not possible. Potential duplication of effort at individual sites.
Infrastructure	Does not need a central trusted research environment.	Resource implications include time, expertise, infrastructure, and network. Interoperability and risks associated with vendor lock-in.
Security and safety	Potential for greater transparency in machine learning tool development.	New security challenges need to be understood and safeguards put in place. Data security when shared across centres.
Training	Doing research without seeing the data requires a shift in thinking.	Requires high end technology and expertise
Diversity	Could increase size and diversity of research populations. Could help democratise research by engaging smaller centres. Could be used to test how biases in data affect models.	Will be challenging to onboard non-research orientated hospitals due to technical and resources requirements. Issue of equity, with academic centres being ready to use these techniques and other centres lacking resources and expertise.
Other	Potential to develop more accurate, fairer, representative machine learning models to support patient care.	Buy-in from industry is likely to be challenging. Ownership of mutually-generated intellectual property.

To develop a federated learning approach for cardiovascular imaging research several solutions were suggested. These included:

- Defining the need for federated learning clearly, with potential use cases.
- Learning from the use of this approach in other areas e.g. genomics.
- Engaging with patients and the public to obtain and maintain trust. This includes developing resources to explain the complicated terms and to co-create solutions to the research challenges.
- Developing multi-stakeholder buy-in for federated learning techniques, including from the NHS, researchers, patients, learned societies, and industry.
- Exploring options for funding of the required underlying infrastructure, which could have a benefit for local patient care.
- Developing new training materials to support researchers to develop in this area including a set of best practice techniques.
- Developing a national, once for all approach for ethics, governance, and intellectual property agreements.
- Demonstrating whether this approach provides tangible benefits over traditional approaches for certain research questions.
- Using driver projects: to demonstrate whether this approach is useful for cardiovascular imaging research; to work with regulators and other stakeholders to streamline authorisation processes; and to engage with stakeholders including the public in order to build trust.

4.5 What are the views of patients and the public on using imaging data in research?

We held a specific breakout session for patient/public representatives to enable them to discuss their views on using imaging data in research, which they might have been unable to raise or discuss in detail during the other sessions.

Our patient/public representatives recognised that there are important benefits to patients and the public of imaging data being made available for use in research. They were impressed by the wide range of research that could be performed using imaging data and felt that this was not always appreciated by patients and the public. Improving communication to patients about how research involving imaging could have benefits for themselves, other patients and the NHS was felt to be important. This information could be discussed when patients are invited to participate in research studies, and in routine letters when patients are invited to clinics or for imaging tests. Communication could be improved about the results of the research that was being performed in the UK, using varied and engaging and formats. Involving patients in the dissemination of the results of research studies would also be valuable.

Our patient/public representatives also highlighted the need for large datasets involving people from multiple groups, to enhance the research diversity and enable insights into rarer diseases. Using existing data collected as part of routine healthcare was suggested as a way of creating large-scale data resources, at no additional “burden” to patients. Many of the benefits mentioned related to the potential to improve patient care and patient outcomes, and that this would be applicable for cardiovascular diseases and other diseases. Potential wider benefits of cardiovascular imaging research were discussed, including making improvements to imaging technology (such as increased speed and accuracy) and potential benefits for NHS professionals (such as improved job satisfaction) and to the NHS (such as cost recovery, improved resources, and enhanced capacity).

Most of the concerns raised related to data privacy and security, including maintaining confidentiality and controlling who has access to the data. Our patient/public representatives emphasized the need for patients to feel in control of their data and what it is used for, and for this process to be transparent and understandable. Difficulties with opting in/out and understanding what this entails were highlighted. Concerns were raised about how accurate the data and its analysis were, and whether there might be potential for biases to be introduced. There were also concerns around how imaging research would be funded, with some questioning whether this is an efficient use of NHS funds and if other sources of funding could be found such as through industry. Commercial access to the data was discussed, with the overall feeling that this could be managed, but should be paid for. The importance of getting industry relationships right was discussed, particularly when developing new and expensive techniques.

Key suggestions for future work from the breakout discussion were that there should be improved clarity around the opt in and opt out processes, what imaging data could be used in research, what imaging data is being used for, and the differences between consented and unconsented imaging research. Central to the discussion was that cardiovascular imaging research should ultimately focus on improving patient care.

4.6 How can we best communicate imaging data research to patients and the public?

We held a breakout session specifically for the patient/public representatives to gather input on how we can best communicate imaging data research to patients and the public. The discussions of imaging research, machine learning and federated techniques can be complex and involve the use of technical language. In advance of the meeting, we provided the patient/public representatives with resources around the topics to be discussed in the meeting, as well as holding a separate discussion session with them. This breakout session provided an opportunity for our patient/public representatives to discuss what communication had worked well or could have been improved.

The need for clear communication with patients and the public was repeatedly and consistently highlighted throughout the workshop. The use of appropriate language and the avoidance of technical jargon is important. Where technical jargon cannot be avoided, clear definitions of technical terms are needed. Examples that our patient/public representative highlighted were words and abbreviations like *DICOM*, *PACS*, *machine learning* and *federated analysis*. The use of images, graphics and video could all improve explanations. Peer support was highlighted as very useful, for example from others who have experienced imaging tests or been involved in research. It was stressed that an appropriate amount of time would be needed to enable good communication and that “one size does not fit all” when it comes to communication.

The suggestions from our patient/public representatives on how to best communicate imaging research can be grouped into three key areas: the need to consider carefully the language used, to seek and listen to feedback from patients, and to use accessible and engaging formats. Patients should be involved in all aspects of research, from developing the initial ideas all the way to disseminating the results, and beyond.

5 Next steps

The last session of the workshop discussed potential future activities of the BHF Data Science Centre to improving cardiovascular imaging research. Key topics from throughout the workshop were community engagement, conducting a prioritisation exercise for cardiovascular imaging research, communicating cardiovascular imaging research to the public, coordinating imaging based driver projects and developing cardiovascular imaging based open challenges.

5.1 Community engagement

During the workshop we brought together a diverse range of stakeholders with an interest in cardiovascular imaging research. We will maintain engagement across this diverse community through: inclusion of cardiovascular imaging research at the BHF Data Science Centre monthly webinars; creating an advisory group for this thematic area of the BHF Data Science Centre; coordinating working groups around key topics such as incorporating imaging data into the proposed BHF Data Science Centre cohorts platform and demonstrating whether federated learning has utility in cardiovascular imaging research.

5.2 Prioritisation exercise for cardiovascular imaging research

We plan to perform a prioritisation exercise to obtain consensus on priorities in cardiovascular imaging research, using a modified Delphi technique. This workshop has completed our exploratory phase for the prioritisation exercise, including assessing the range of potential research questions that may be considered, important themes in the research questions, and domains that should be considered for prioritisation. The next stage of the prioritisation question will involve an open call for further research questions followed by two rounds of prioritisation. We will involve the community further in this important exercise.

5.3 Communication of cardiovascular imaging research to patients and the public

This workshop has helped us explore how we can improve the communication around cardiovascular imaging research to the patients and public. We will use this information to co-produce resources with our patient/public representatives which will be made available on the BHF Data Science Centre website, aimed at patients and the public, but also useful for researchers and clinicians. This will include how different imaging modalities can be used in research and explanations of key terms. It will also include links to other relevant resources such as those on the BHF and HDRUK websites.

5.4 Driver projects

Central to the next steps of this theme will be a series of driver projects that can act as exemplars for future research by the community. **Particularly relevant topics highlighted by this workshop will include research involving cross-nation data collection, federated learning, novel types of imaging data, imaging datasets and/or annotation that can be made available to other researchers, and bringing together imaging data with other BHF Data Science Centre themes such as wearable data and phenomic or genomic data.** The BHF Data Science Centre will support funding applications for potential projects and provide project advice/support. Shared knowledge from these projects will include best practices, ethical and regulatory approval requirements, data curation and analytic techniques. We encourage the community to explore potential driver projects with the BHF Data Science Centre team.

5.5 Cardiovascular imaging based open challenges

In the past 5 years imaging based open challenges have advanced machine learning analysis techniques in a variety of fields, including imaging [6]. These open challenges galvanise a community around a particular image analysis challenge. A dataset is provided along with ground truth annotations, and researchers are invited to submit their solutions. We will explore with the community the potential to organise a relevant challenge to drive forward our work to support and enable cardiovascular imaging research.

6 Conclusions

This workshop brought together a broad range of stakeholders and patient/public representatives to discuss the important topic of *How can we use imaging data to better understand cardiovascular disease?* The insights provided will help improve the already world leading cardiovascular imaging research in the UK. Many of the discussions are generalisable to other types of imaging research or broader health data science research. Central to the discussions was the importance of involving patients and the public at all stages and keeping the impact of cardiovascular imaging research on the health and wellbeing of the public at the forefront of discussions. The discussions at this workshop are now being used to guide the BHF Data Science Centre activities in this thematic area, which will centre around community engagement, prioritisation of cardiovascular imaging research questions, communication with patients and the public, driver projects and cardiovascular imaging based challenges.

7 Acknowledgements

This workshop and report were produced by the BHF Data Science Centre (grant No SP/19/3/34678, awarded to Health Data Research (HDR) UK), with the following contributions: Jackie MacArthur provided input on workshop design, organised workshop, drafted and revised report; Michelle Williams designed, co-organised and chaired workshop, drafted and revised report; Cathie Sudlow provided input on workshop design, co-chaired, revised report; Lynn Morrice provided input on workshop design, revised report; Debbie Ringham organised recruitment and orientation of patient/public contributors; Tammy Watchorn (<https://tammywatchorn.com/>) designed breakout sessions; Lydia Martin provided administrative support.

In addition, the following people are acknowledged for their engagement and rich contributions to the workshop: workshop speakers, breakout session chairs and facilitators (listed as contributors), patient/public contributors, and workshop participants (see Appendix B).

8 References

1. WHO. Cardiovascular diseases (CVDs). 2021 [updated 16 August 2021; cited 2021 16 August 2021]; Available from: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
2. BHF. UK Factsheet. 2021 [16 August 2021]; Available from: <https://www.bhf.org.uk/what-we-do/our-research/heart-statistics>.
3. Yong GL, Weir-McCall J, Wilson M, Roditi GH, Bull R, Williams MC, et al. Research priorities in cardiovascular imaging. *Open Heart*. 2020;7(2):e001389.
4. Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance*. 2013;15(1):46.
5. Cushnan D, Bennett O, Berka R, Bertolli O, Chopra A, Dorgham S, et al. An overview of the National COVID-19 Chest Imaging Database: data quality and cohort analysis. *Gigascience*. 2021;10(11).
6. Prevedello LM, Halabi SS, Shih G, Wu CC, Kohli MD, Chokshi FH, Erickson BJ, Kalpathy-Cramer J, Andriole KP, Flanders AE. Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions. *Radiology: Artificial Intelligence* 2019;1(1):e180031.

9 Additional material

Appendix A – Agenda

Day 1 – Wednesday 17th November 2021

All times are in GMT

Time	Item	Speakers
13:00	Welcome and overview	Michelle Williams, Cathie Sudlow
Session 1 – What are the most important research questions that cardiovascular imaging can be used to address?		
13:10	Cardiovascular imaging research in the UK – the BHF's contribution	Jeremy Pearson
13:20	What can we learn from cardiovascular imaging?	Marc Dweck
13:30	Identifying priorities in cardiovascular imaging research	Michelle Williams
13:40	Questions	
13:50	Introduction to breakout sessions	Tammy Watchorn
14:00	Breakout session	
	<i>Introductions</i>	
14:10	<i>What are the most important research questions that cardiovascular imaging can be used to address?</i>	<i>15 people per breakout, 1 person invited before event to chair each group</i>
	<i>What criteria should be used for prioritisation?</i>	
14:40	Break	
14:55	Facilitators of each breakout group report back	
Session 2 – What are the biggest challenges for using imaging data in cardiovascular research?		
15:15	Challenges of using imaging in a consented multi-centre study	Steffen Petersen
15:25	Challenges of using routine NHS imaging in research	Joseph Jacob
15:35	Challenges of using artificial intelligence/machine learning in imaging research	Carola-Bibiane Schönlieb
15:45	Questions	
15:55	Breakout session	<i>15 people per breakout, 1 person invited before event to chair each group</i>
	<i>What are the biggest challenges for using imaging data in research?</i>	
	<i>How might we resolve these challenges?</i>	
16:25	Facilitators of each breakout group report back	
16:45	Wrap up of the key themes and explain next steps	Michelle Williams
17:00	End of day 1	

Day 2 – Thursday 18th November 2021

All times are in GMT

Time	Item	Speakers
9:00	Welcome	Michelle Williams, Cathie Sudlow
	Session 3 – What are the requirements for a pipeline to support cardiovascular imaging research?	
9:05	Curating NHS imaging cohorts within a trusted research environment	Emily Jefferson
9:15	Infrastructure considerations for imaging research	Alex Frangi
9:25	Pre-processing and annotating images for machine learning research	Wenjia Bai
9:35	Questions	
9:45	Breakout session <i>What are the requirements for a pipeline to support cardiovascular imaging research?</i> <i>What could be done to deliver such a pipeline?</i>	<i>15 people per breakout, 1 person invited before event to chair each group</i>
10:15	Facilitators of each breakout group report back	
10:35	Break	
	Session 4 - How can federated analysis and federated learning be used to improve cardiovascular imaging research?	
10:45	What are federated learning and federated analysis and why are they useful?	Jorge Cardoso
10:55	What are the challenges of delivering federated learning and federated analysis?	Hans-Erik Aronson
11:05	Questions	
11:15	Breakout session <i>Which types of cardiovascular imaging research questions could be addressed with a federated learning/analysis approach?</i> <i>What are the advantages and challenges of a federated learning/analysis approach?</i> <i>How might we resolve these challenges?</i>	<i>15 people per breakout, 1 person invited before event to chair each group</i>
11.45	Facilitators of each breakout group report back	
12.05	Break	
	Session 5 – What should be done next?	
12.20	Expert panel discussion and questions	
12:50	Closing remarks	Michelle Williams, Cathie Sudlow
13.00	Meeting ends	

Appendix B – Attendee list

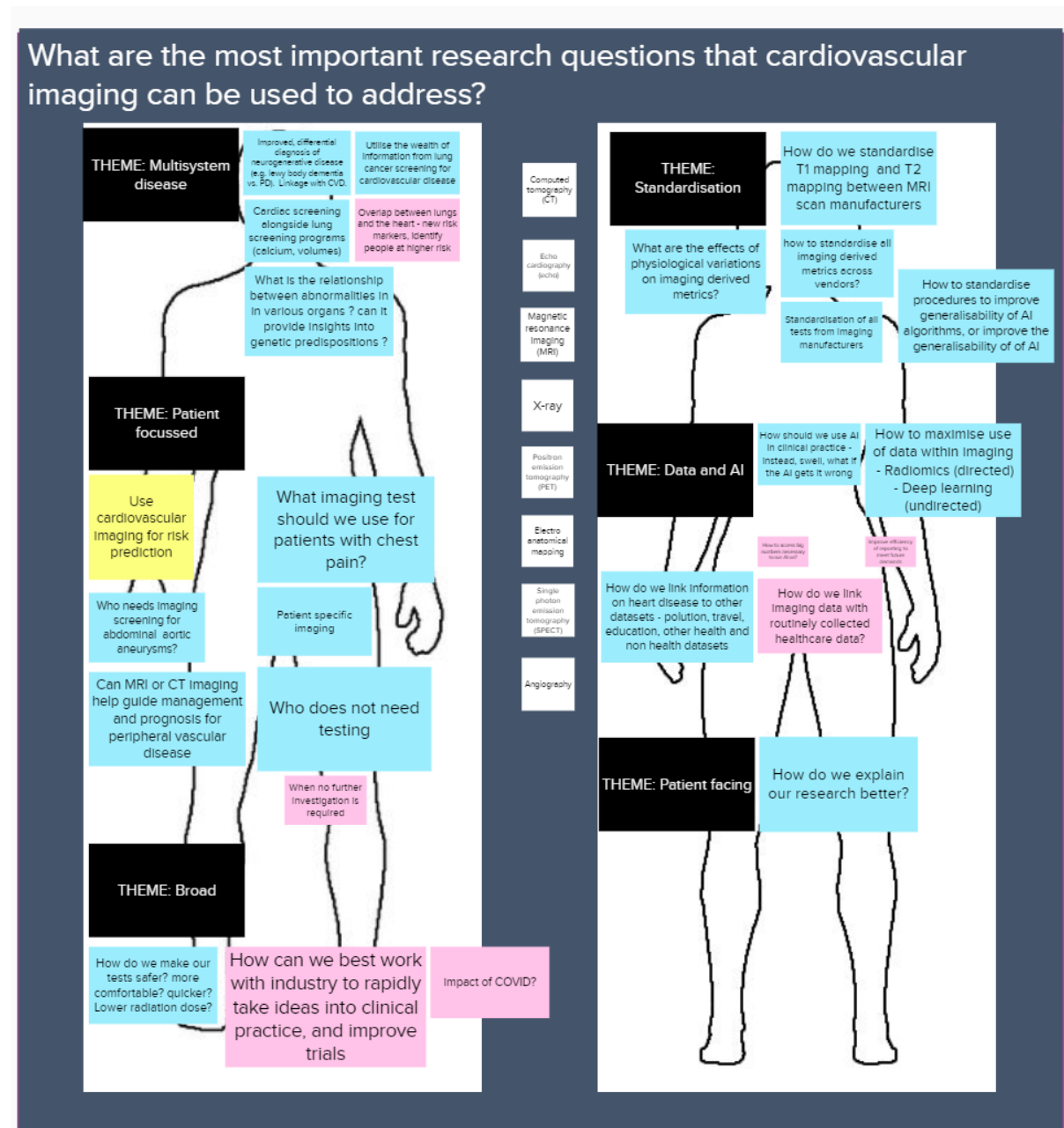
Role	Name	Affiliation
Organisers	Cathie Sudlow	BHF Data Science Centre, HDR UK
	Debbie Ringham	BHF Data Science Centre, HDR UK
	Jackie MacArthur	BHF Data Science Centre, HDR UK
	Lydia Martin	BHF Data Science Centre, HDR UK
	Lynn Morrice	BHF Data Science Centre, HDR UK
	Michelle Williams	BHF Data Science Centre, HDR UK and University of Edinburgh
	Tammy Watchorn	https://tammywatchorn.com/
Speakers	Alejandro F Frangi	University of Leeds
	Carola-Bibiane Schönlieb	University of Cambridge
	Emily Jefferson	PICTURES
	Hans-Erik Aronson	National Digital Research Infrastructure Director at Health Data Research UK
	Jeremy Pearson	British Heart Foundation
	Jorge Cardoso	King's College London
	Joseph Jacob	University College London
	Marc Dweck	University of Edinburgh
	Steffen Petersen	Queen Mary University of London
	Wenjia Bai	Imperial College London
Breakout session chairs and facilitators	Alastair Moss	Department of Cardiovascular Sciences, University of Leicester
	Caroline Roney	Queen Mary University of London
	Esther Puyol Anton	King's College London
	Jason Tarkin	University of Cambridge
	John Nolan	BHF Data Science Centre, HDR UK
	Jonathan Weir-McCall	British Society of Cardiovascular Imaging
	Marianna Fontana	University College London
	Matthew Thakur	HDR UK
	Thomas Bolton	BHF Data Science Centre, HDR UK
Workshop participants	Aaron Lee	Queen Mary University of London
	Allyson Arnold	BHF Clinical Research Collaborative
	Amitava Banerjee	University College London
	Andrew King	King's College London
	Angelica Aviles-Rivero	University of Cambridge
	Anish Bhuvu	University College London, HDR UK
	Becky Harmston	Public contributor
	Betty Raman	Oxford Center for Magnetic Resonance Research, University of Oxford

Role	Name	Affiliation
	Charis Antoniades	University of Oxford
	Charlie Mayor	NHS Greater Glasgow & Clyde
	Charlotte Manisty	University College London and Barts Heart Centre
	Chiara Bucciarelli-Ducci	Royal Brompton and Harefield Hospitals, Guys and St Thomas NHS Trust
	Chim Lang	University of Dundee
	Chris Cole	Health Informatics Centre, University of Dundee
	Chris Miller	University of Manchester
	Colin Berry	BHF Glasgow Cardiovascular Research Centre
	Danielle Paul	University of Bristol
	Darrel Francis	National Heart and Lung Institute, Imperial College London
	David Oxborough	Liverpool John Moores University
	Elizabeth Le	University of Cambridge
	Gabriella Captur	University College London
	Gerald Mccann	University of Leicester, Glenfield Hospital
	Gerry Reilly	HDR UK
	Giorgos Papanastasiou	Pfizer
	Jackie Caldwell	Scottish Medical Imaging, eDRIS, Public Health Scotland
	James Moon	University College London and Barts Heart Centre
	Julian Meldrum	Public contributor
	Kenan Direk	University College London
	Lucie Duluc	British Heart Foundation
	Mark Halling-Brown	Royal Surrey NHS Foundation Trust
	Mathew Watt	NHSX AI LAB
	Matthew Sydes	BHF DSC (and MRC CTU at UCL)
	Mustapha Koriba	Public contributor
	Naaman Tammuz	Bitfount
	Nishant Rabikumar	University of Leeds
	Patrick Donnelly	South Eastern Health and Social Care Trust
	Pedro Sanchez	University of Edinburgh
	Phil Blakelock	Public contributor
	Rachel Gerrard	Public contributor
	Rani Buskell	BHF Clinical Research Collaborative
	Rhodri Davies	University College London, University Hospital of Wales, Barts Heart Centre
	Rob Baxter	EPCC, University of Edinburgh
	Robert Wallace	Scottish Medical Imaging, eDRIS, Public Health Scotland
	Shaun Robinson	British Society of Echocardiography
	Sotirios Tsaftaris	University of Edinburgh

Role	Name	Affiliation
	Stefan Neubauer	University of Oxford
	Steven Niederer	King's College London
	Steven Williams	The University of Edinburgh
	Susan Krueger	University of Dundee
	Sven Plein	University of Leeds
	Tareen Dawood	King's College London
	Tim Fairbairn	Liverpool Heart and Chest Hospital, Liverpool Centre for Cardiovascular Science (LCCS)
	Vanessa Ferreira	University of Oxford
	Vesna Vuksanovic	Swansea University Medical School and HDR UK
	William Bradlow	University Hospitals Birmingham NHS Foundation Trust
	William Whiteley	University of Edinburgh
	Winok Lapidaire	University of Oxford
	Xiao Liu	University of Edinburgh
	Zohya Khaliq	Imperial

Appendix C – Mural board images from session 1 “What are the most important questions that cardiovascular imaging can be used to address?”

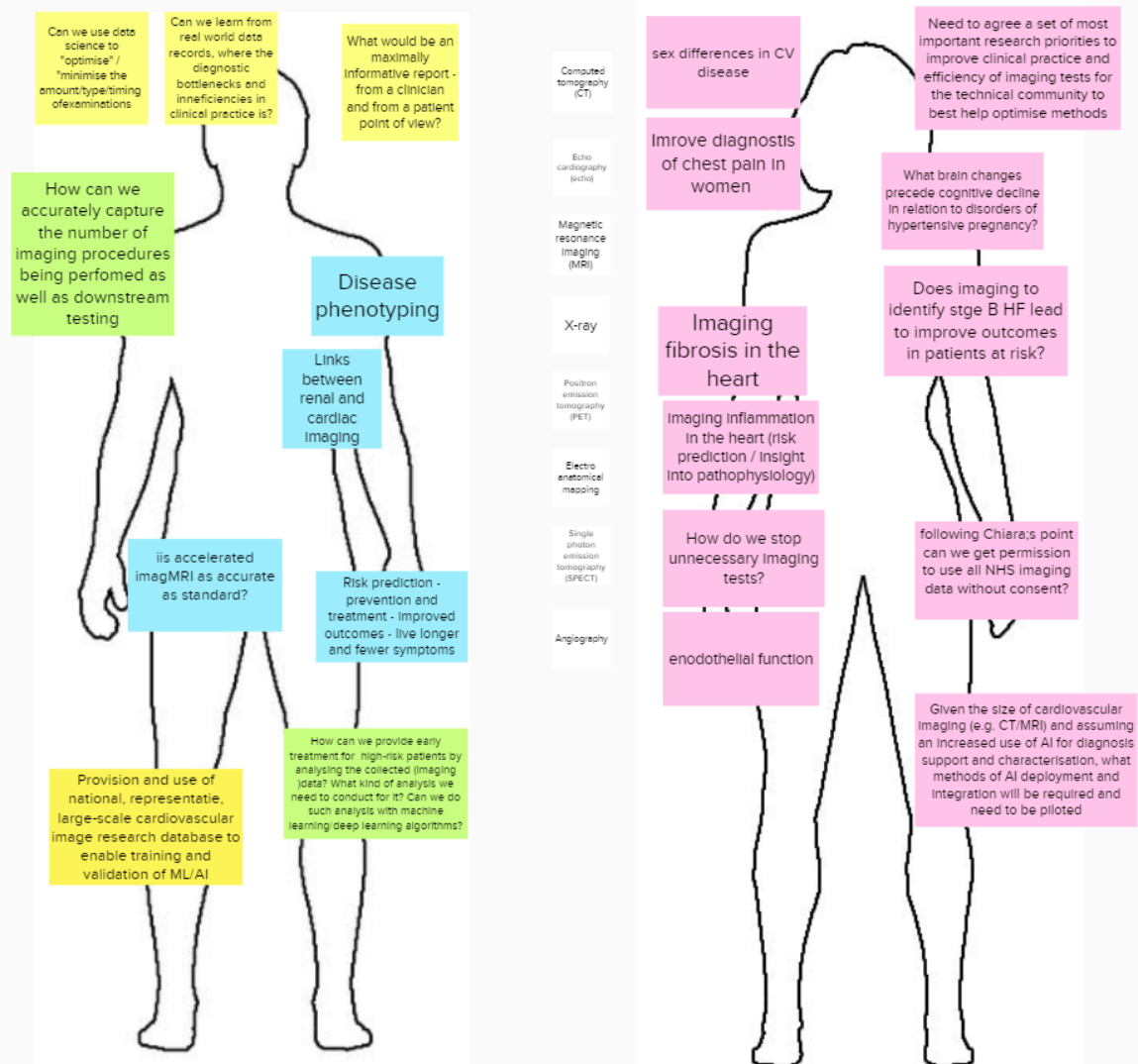
Group 1



Group 2

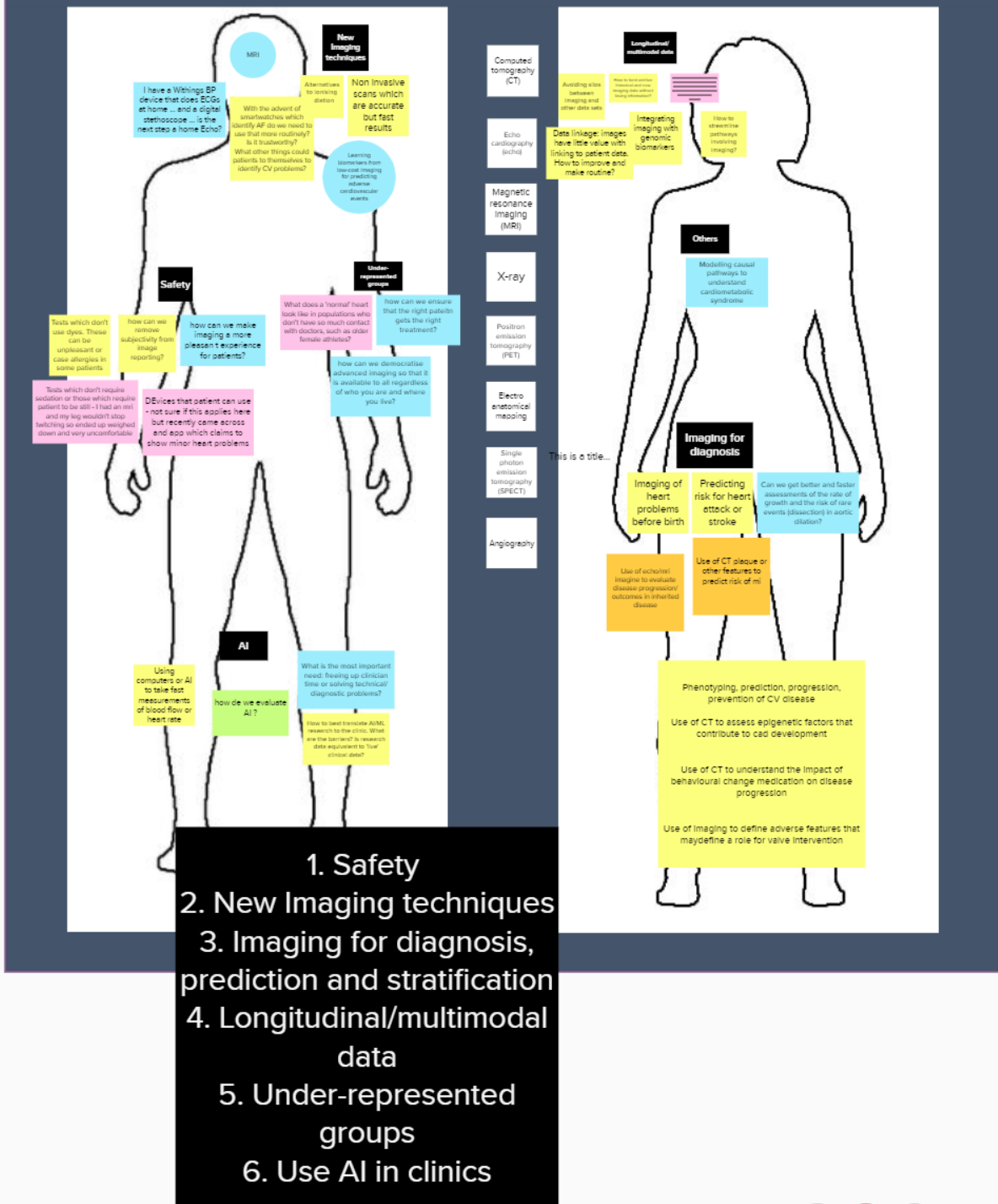
What are the most important research questions that cardiovascular imaging can be used to address?

Research opportunities

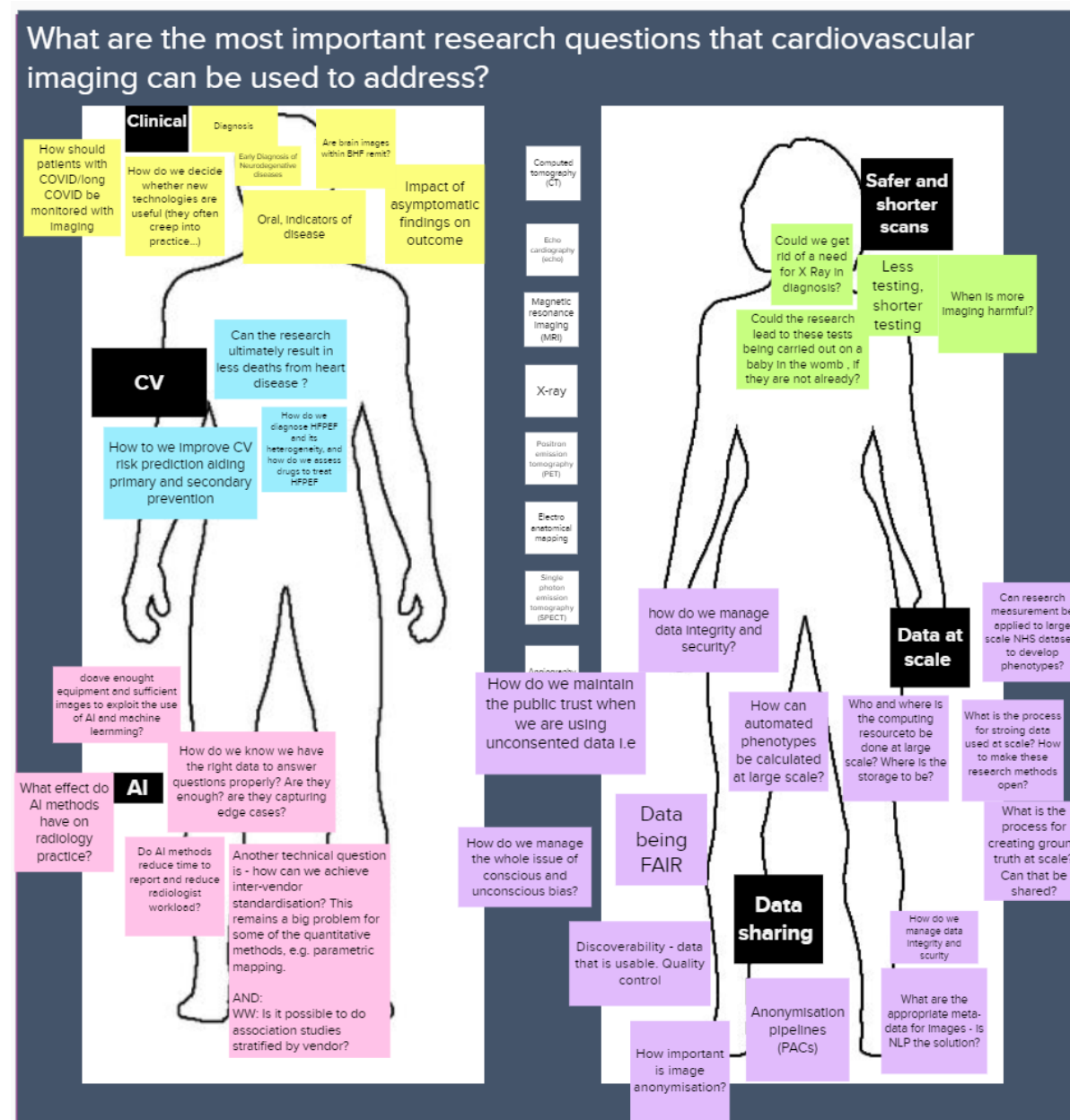


Group 3

What are the most important research questions that cardiovascular imaging can be used to address?

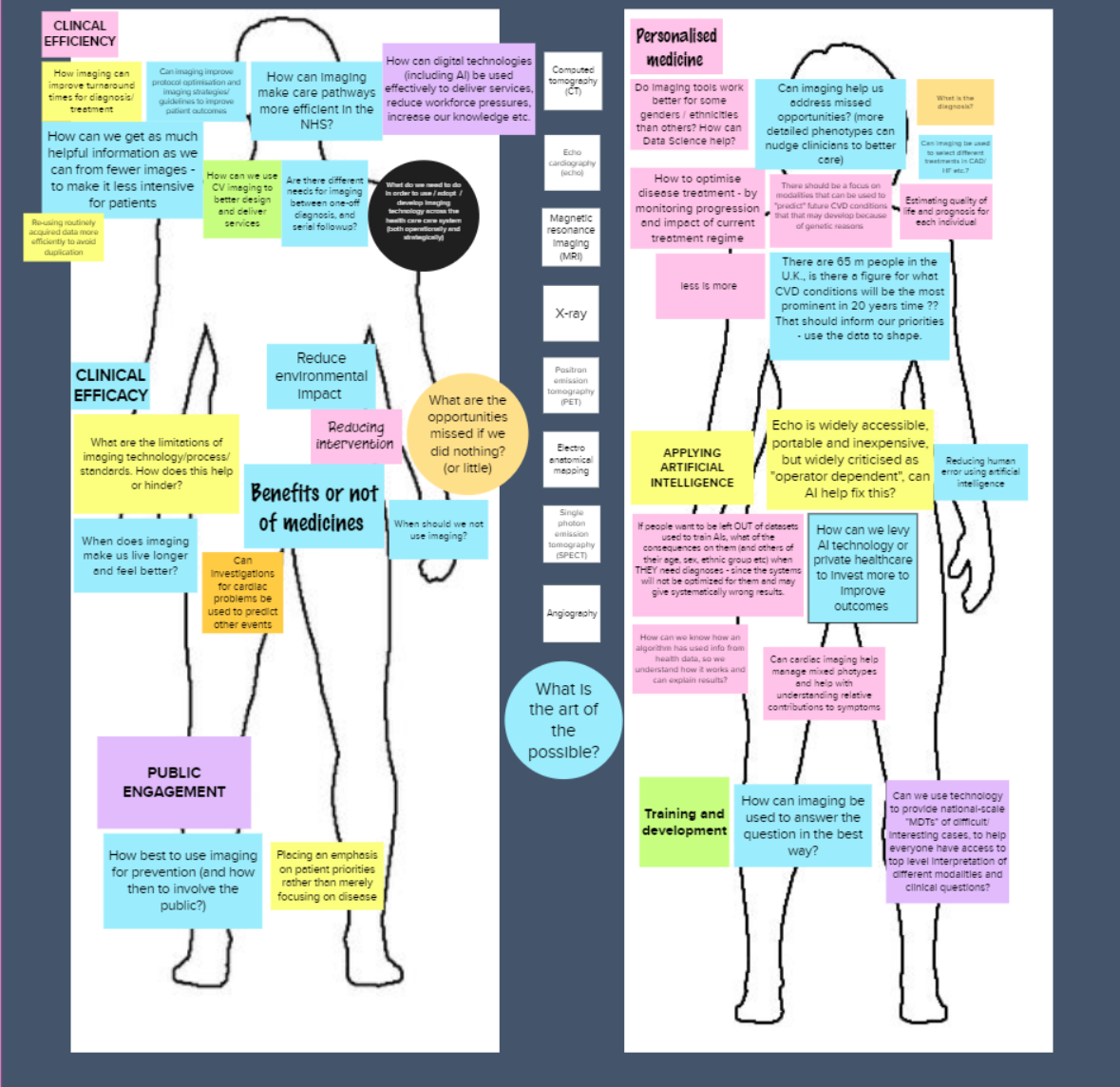


Group 4



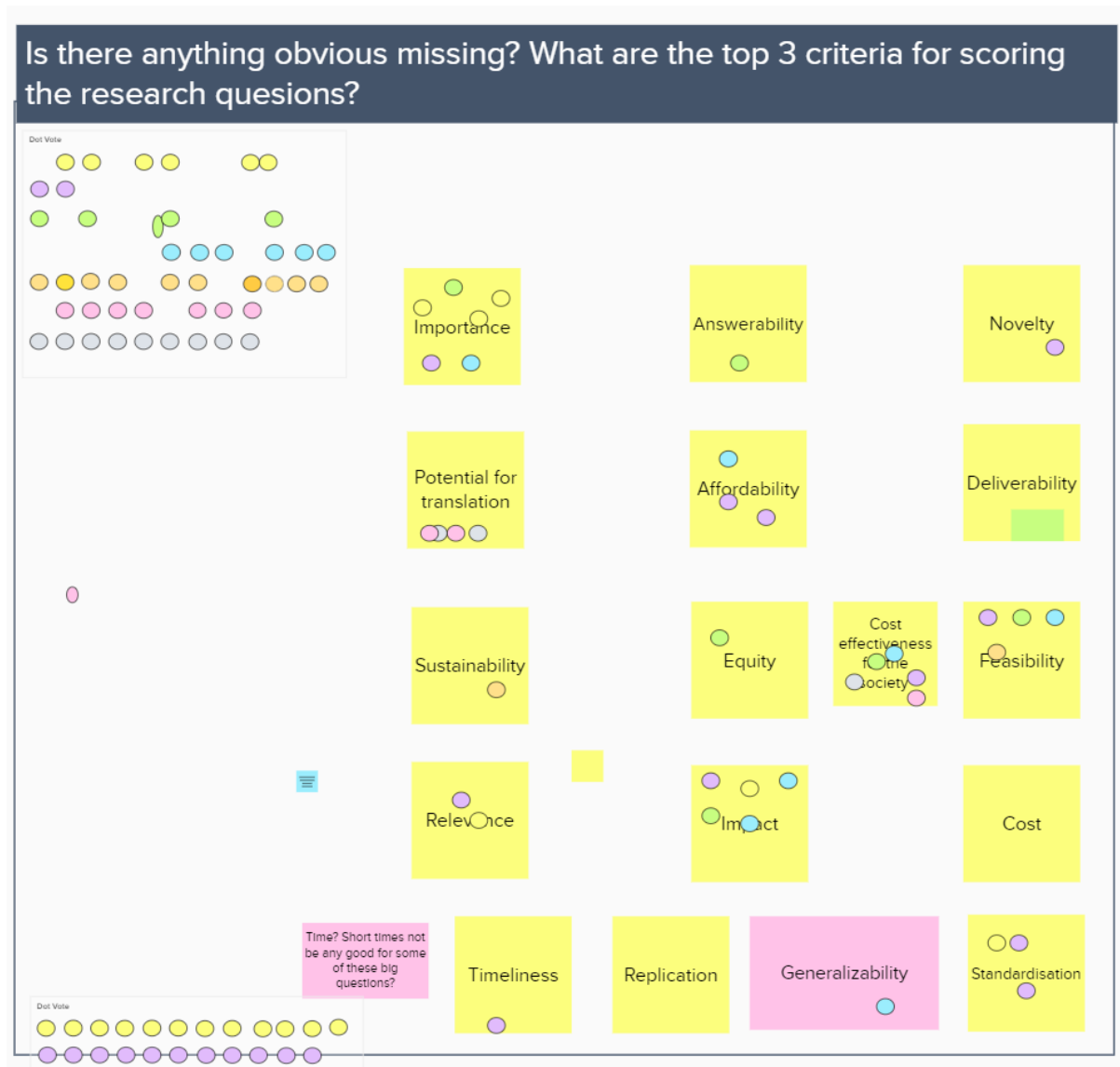
Group 5

What are the most important research questions that cardiovascular imaging can be used to address?

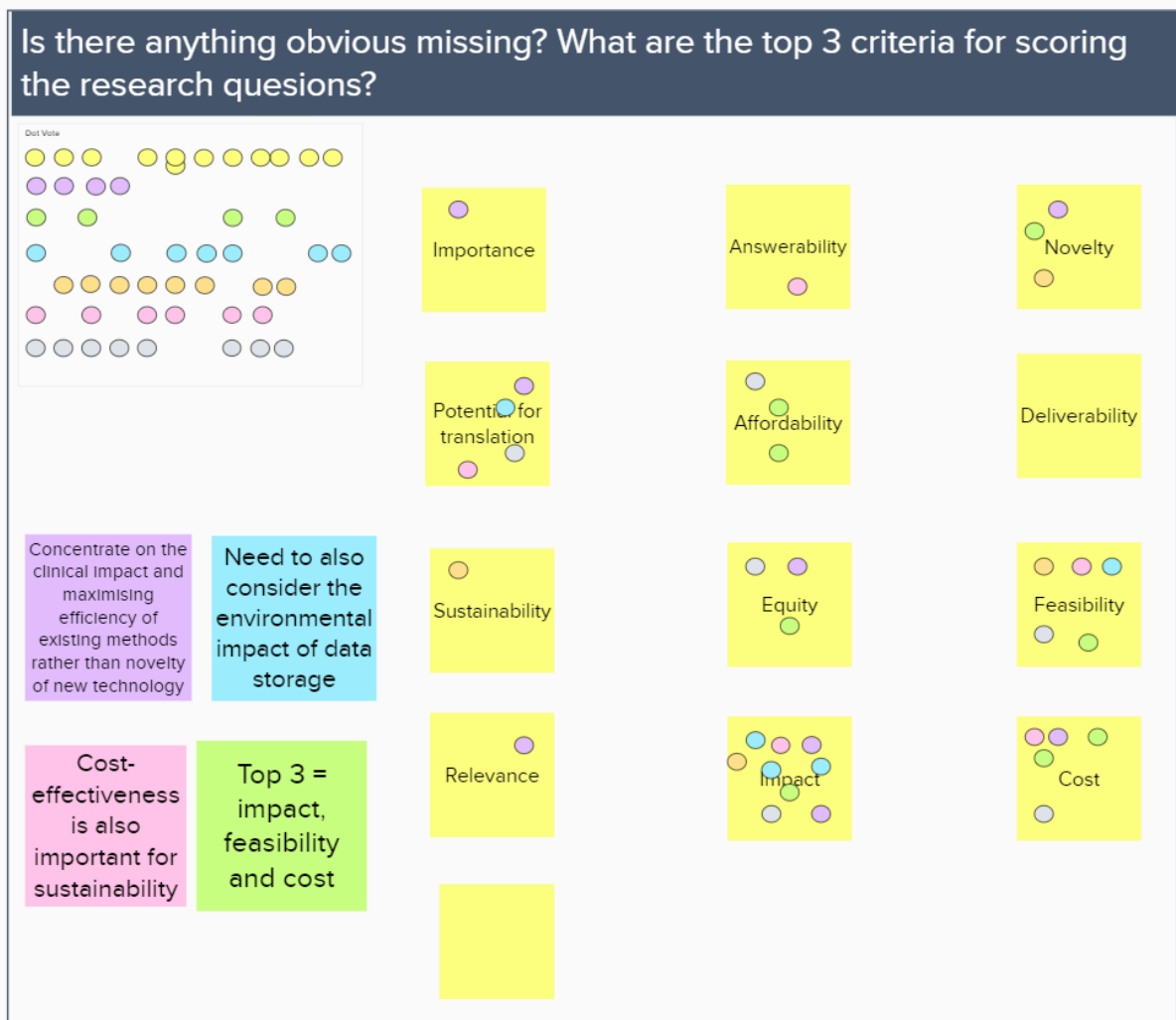


Appendix D – Mural board images from session 1 “What are the top 3 criteria for scoring the research questions?”

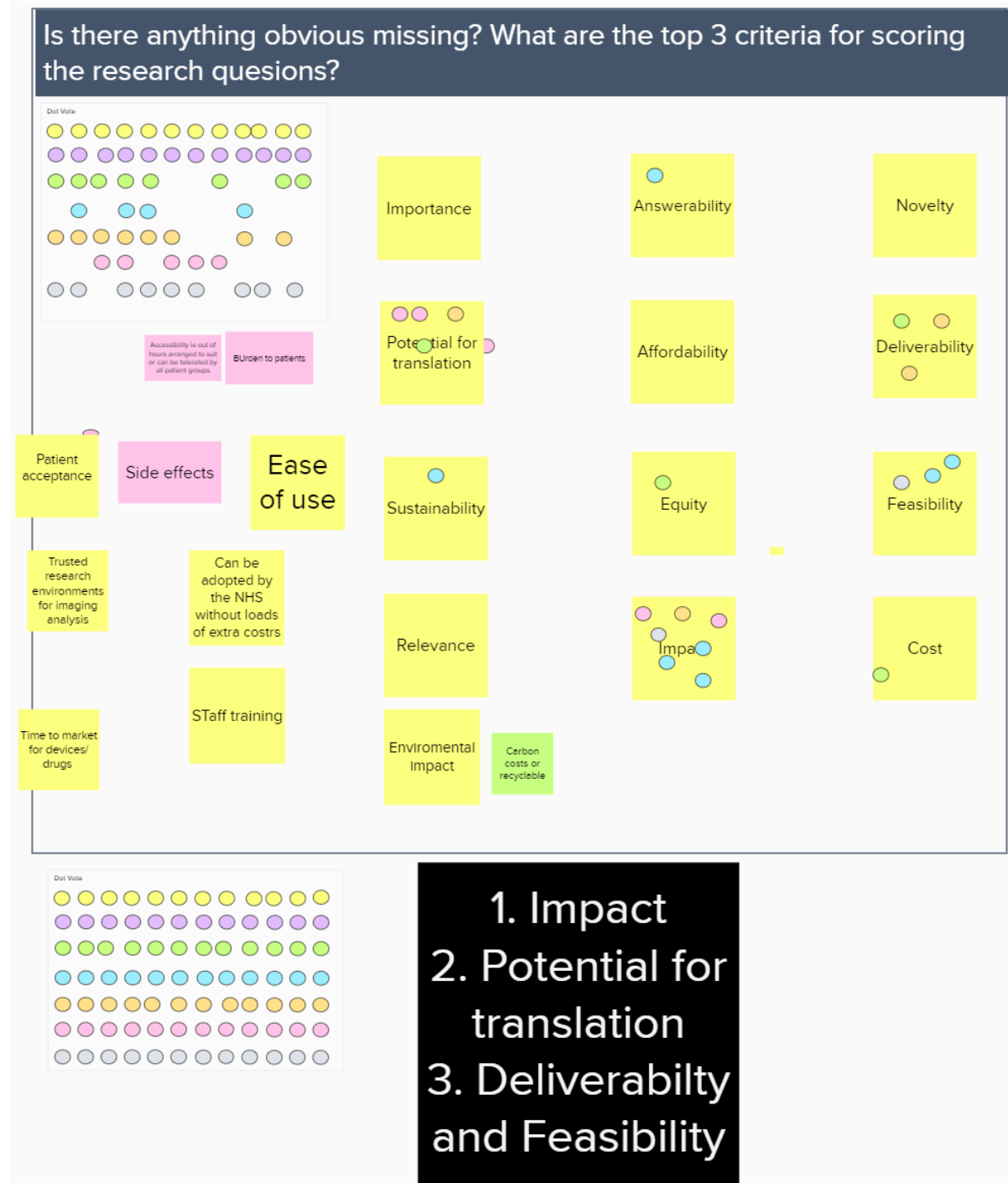
Group 1



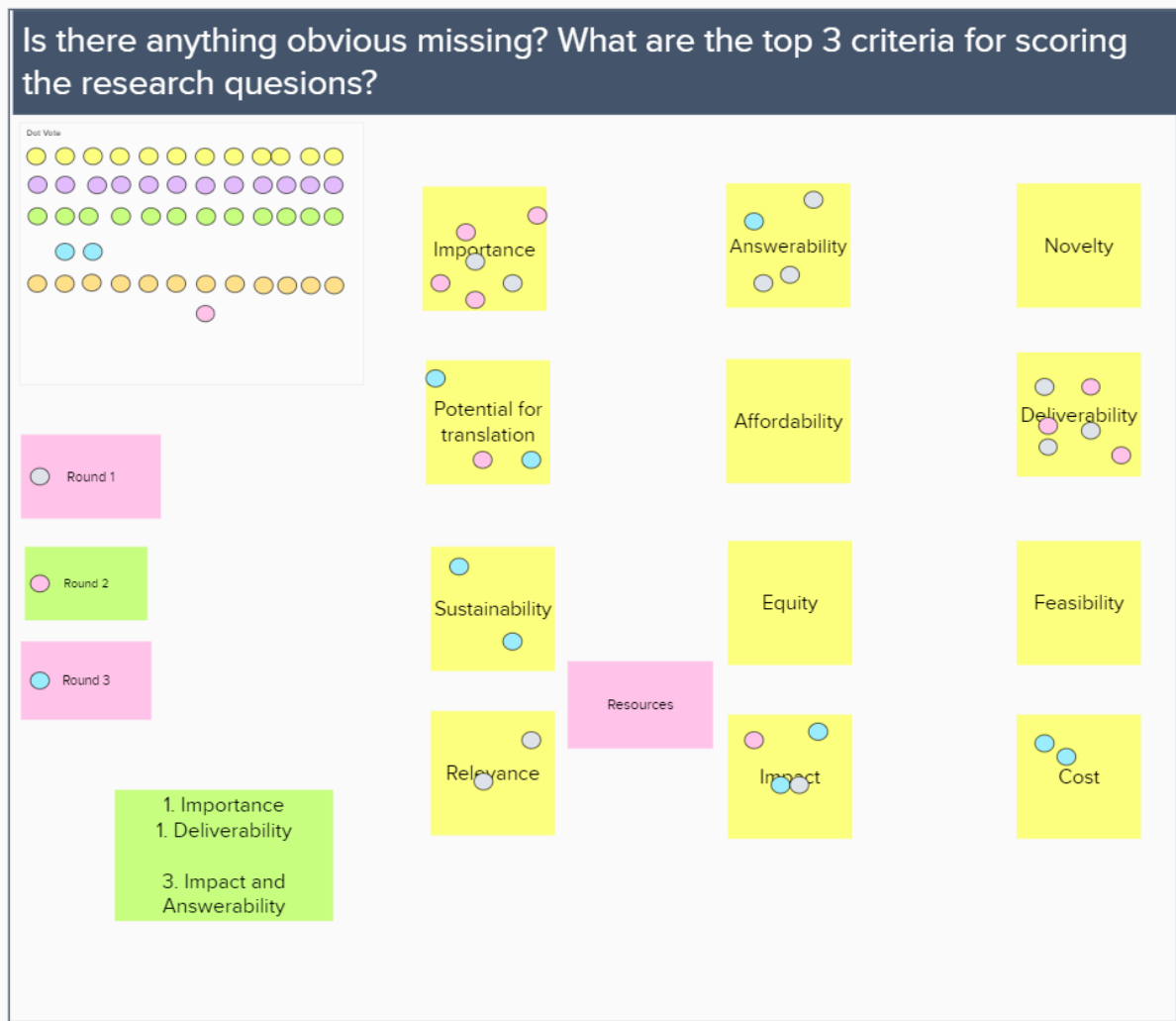
Group 2



Group 3



Group 4



Group 5

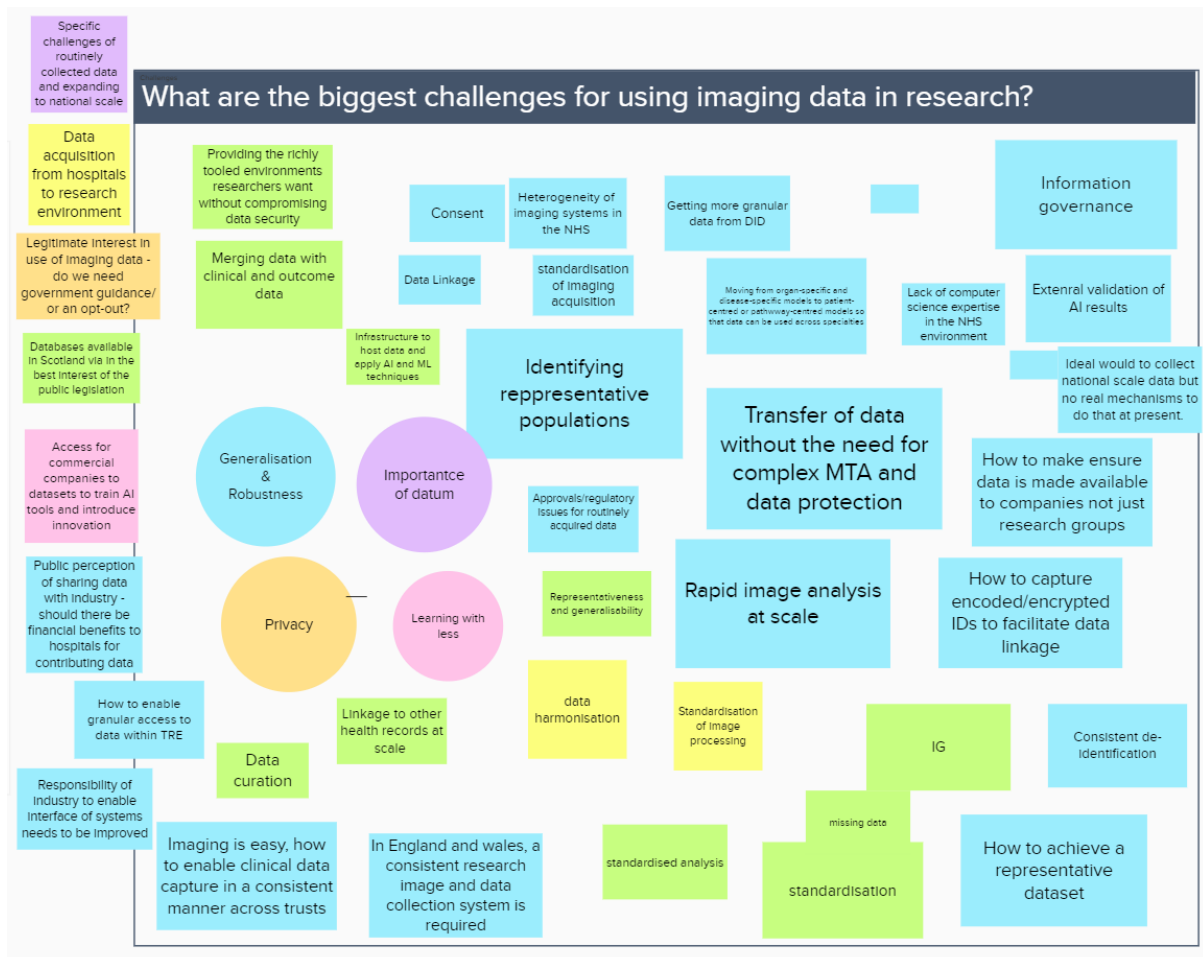


Appendix E – Mural board images from session 2 “What are the biggest challenges for using imaging data in research?”

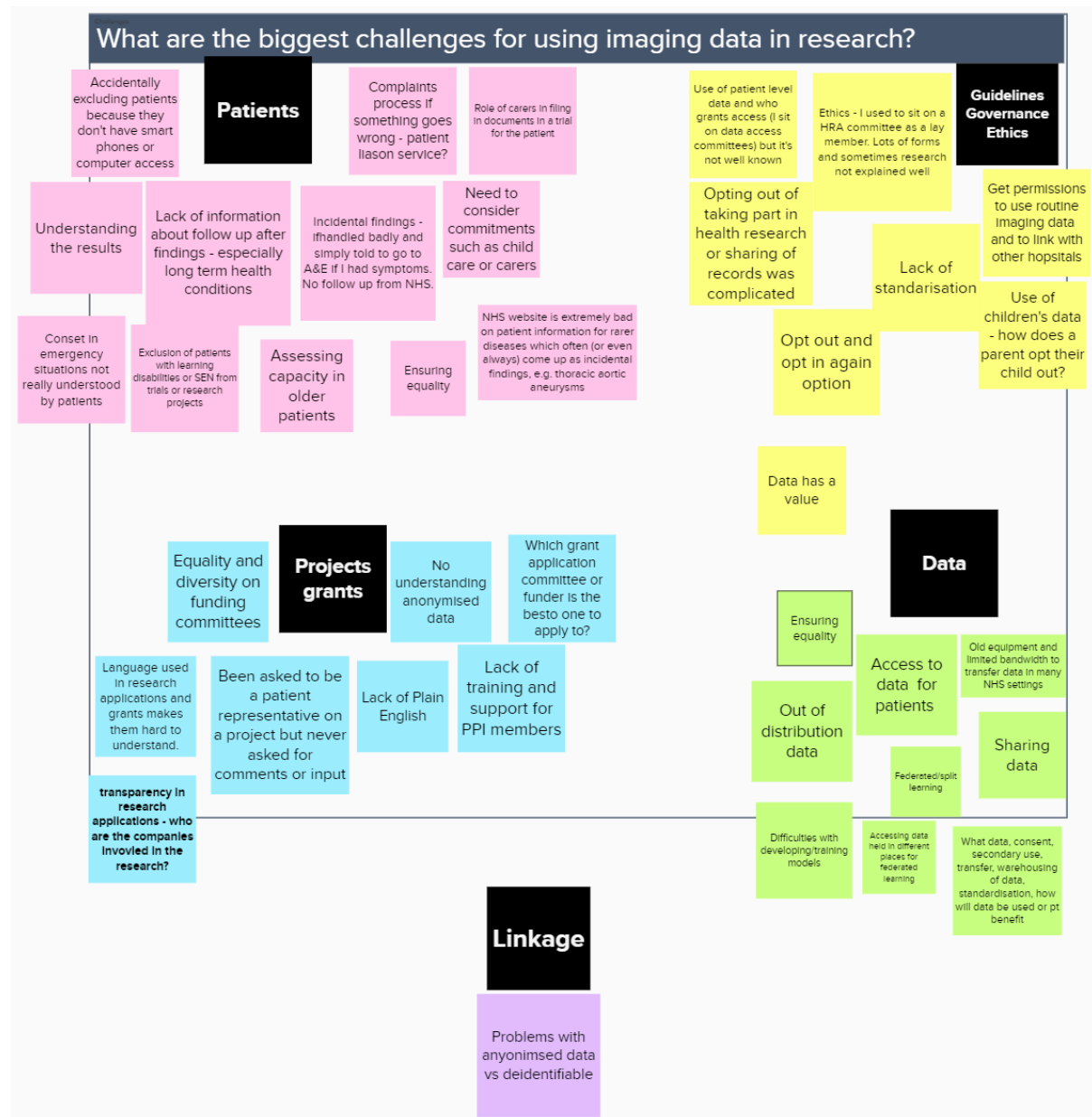
Group 1



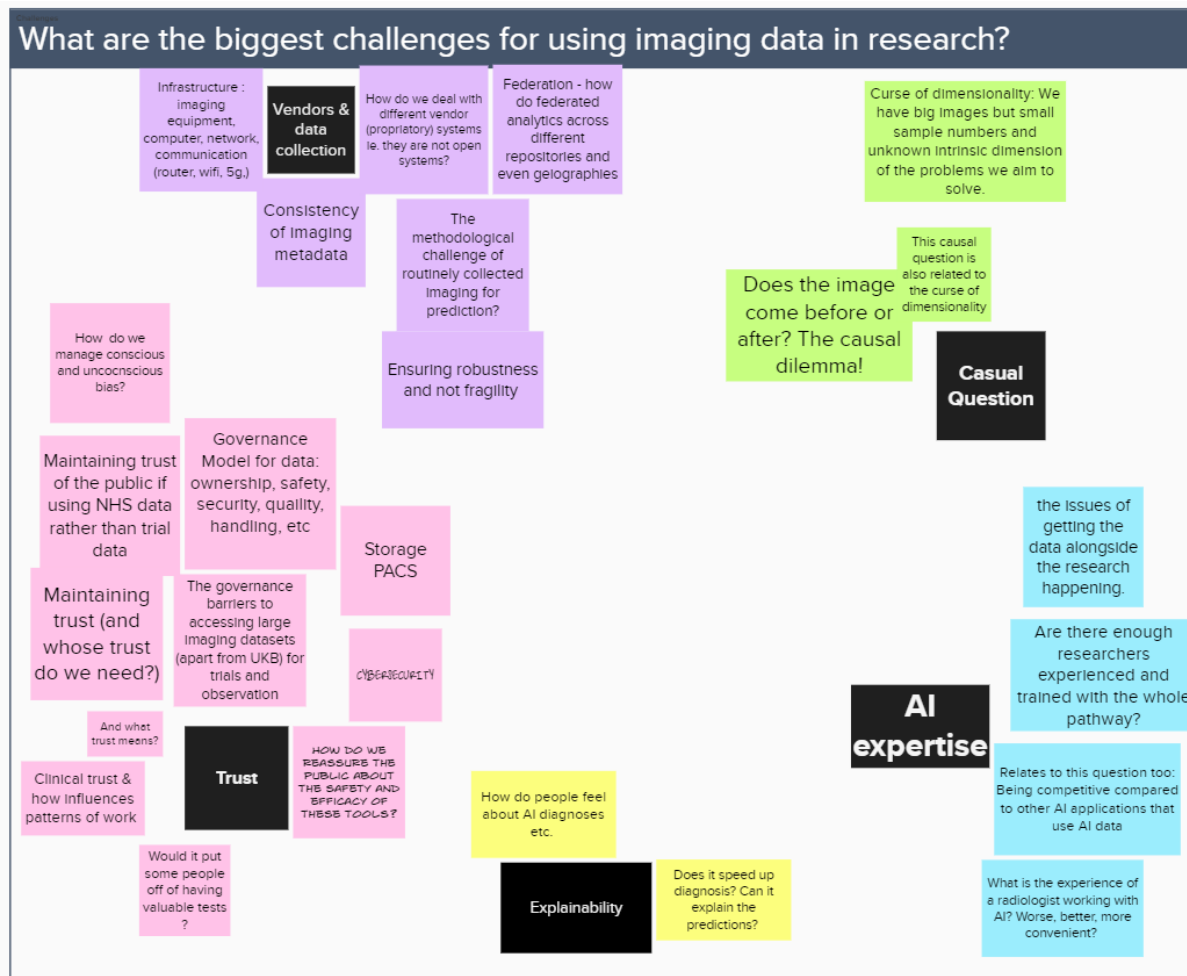
Group 2



Group 3

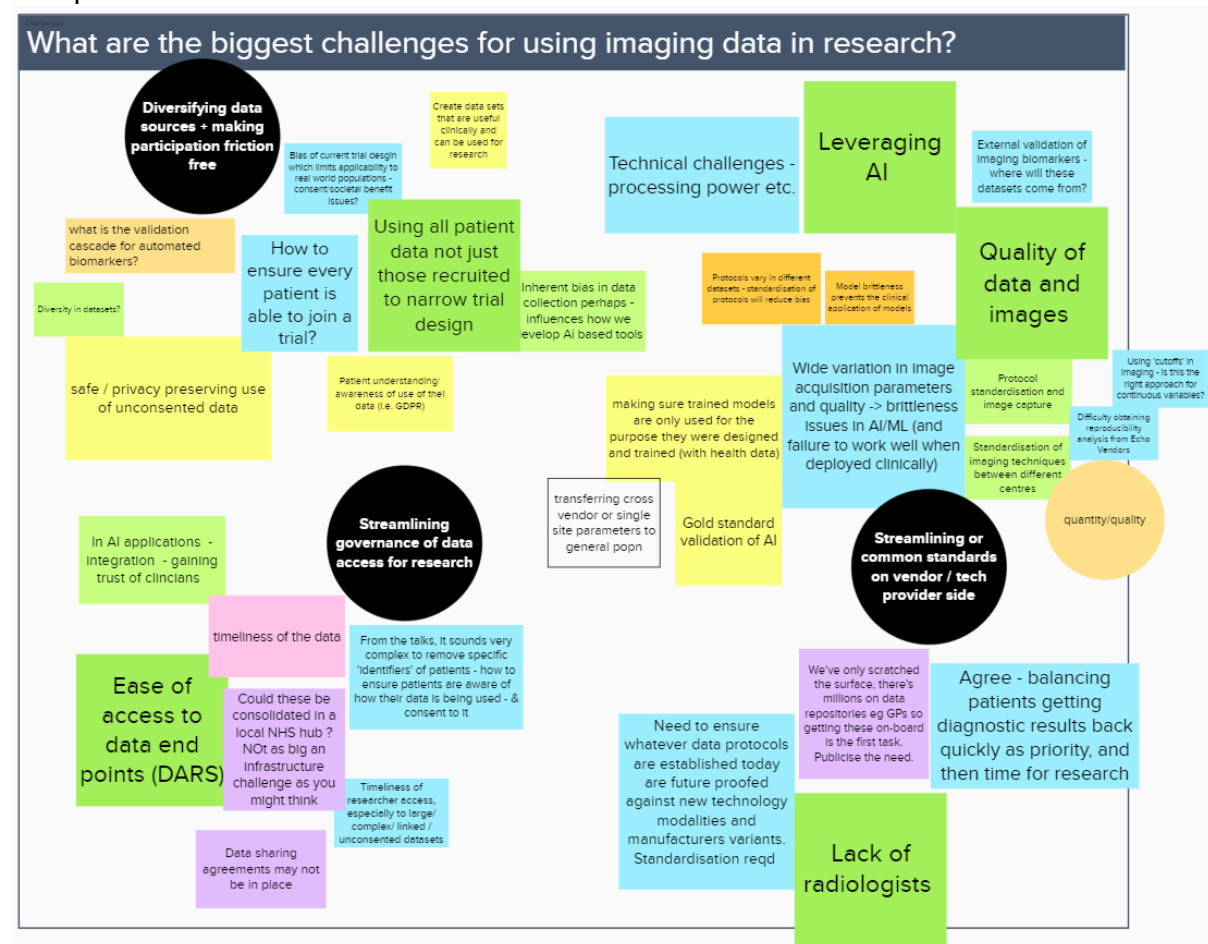


Group 4





Group 5



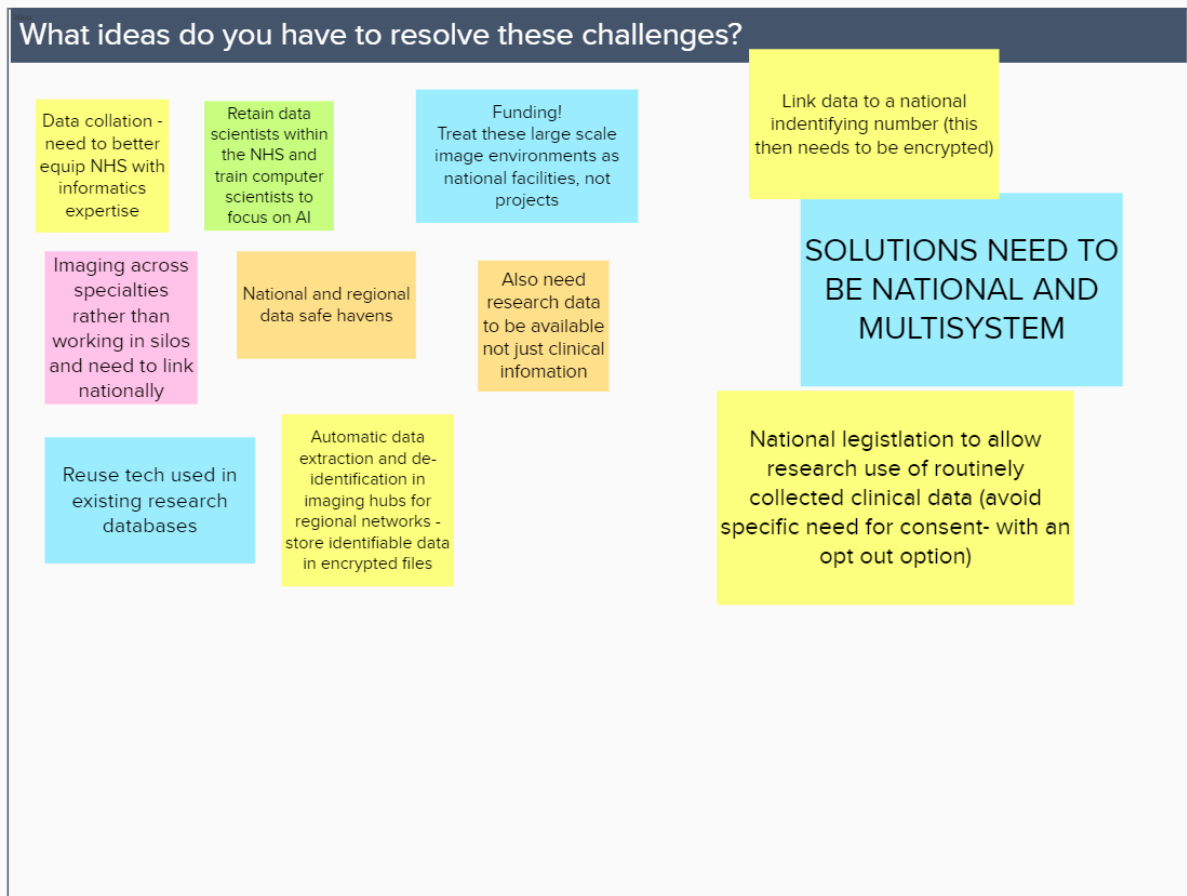


Appendix F – Mural board images from session 2 “What ideas do you have to resolve these challenges?”

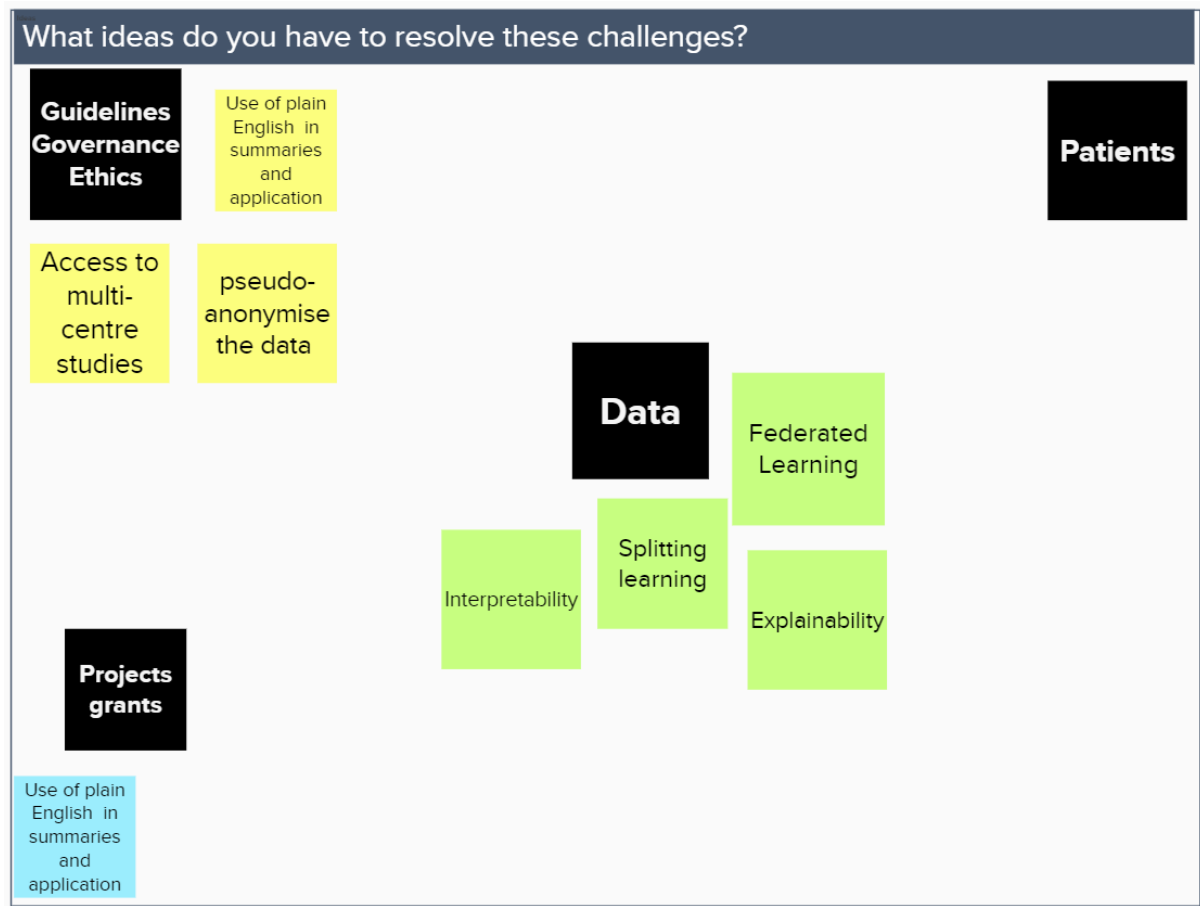
Group 1



Group 2



Group 3



Group 4

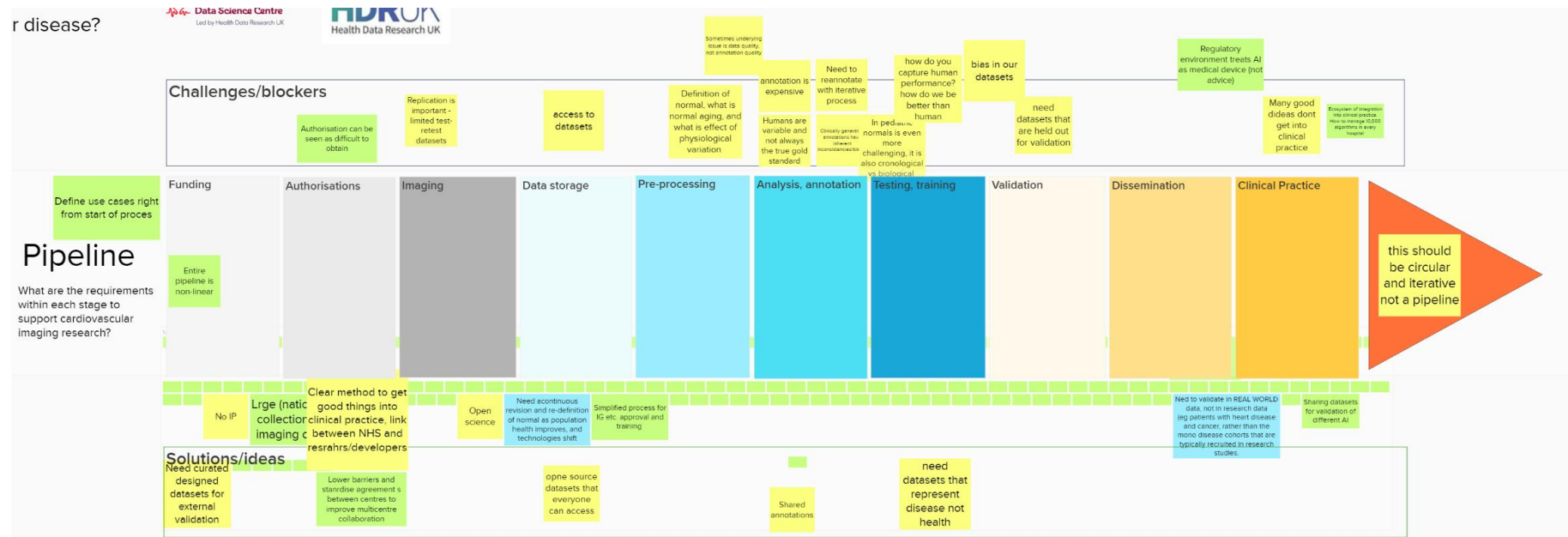


Group 5



Appendix G - Mural board images from session 3 “What are the requirements of a pipeline to support cardiovascular imaging research?”

Group 1

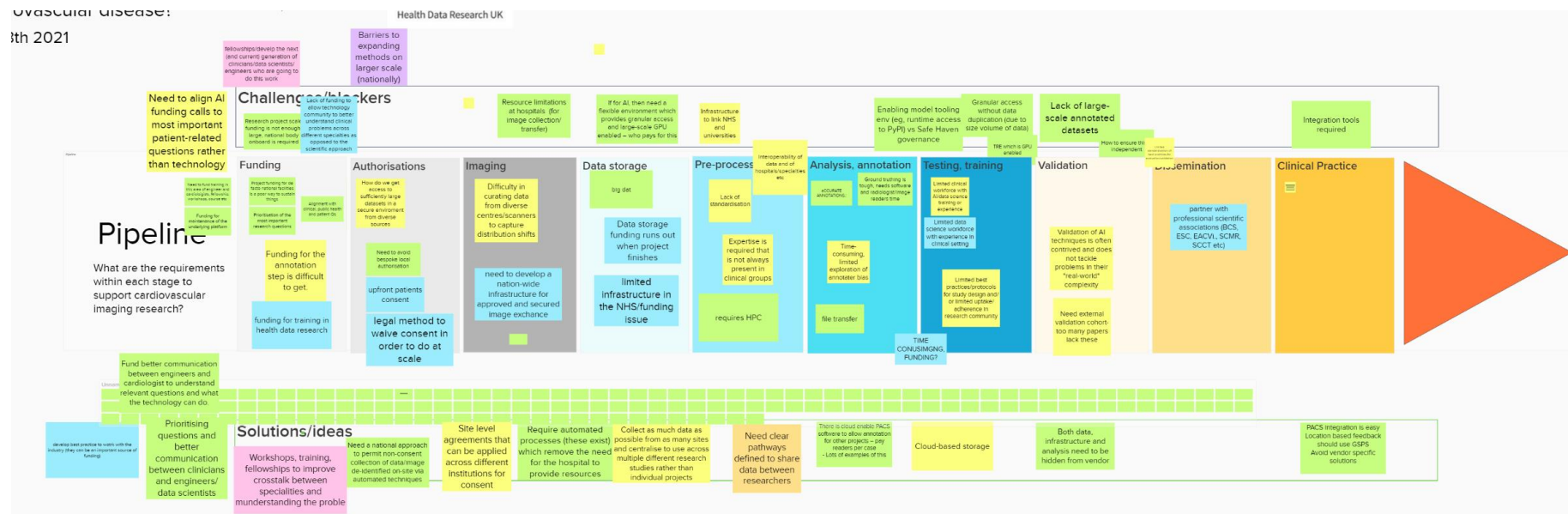


BHF Data Science Centre Workshop report – How can we use imaging data to better understand cardiovascular disease?

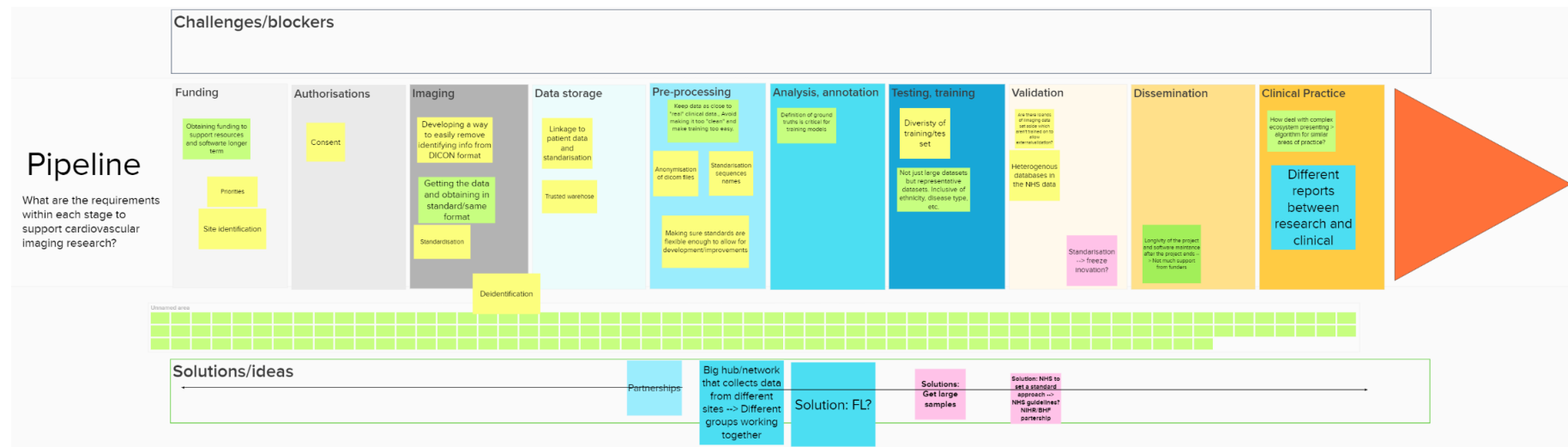
Group 2

Cardiovascular disease:

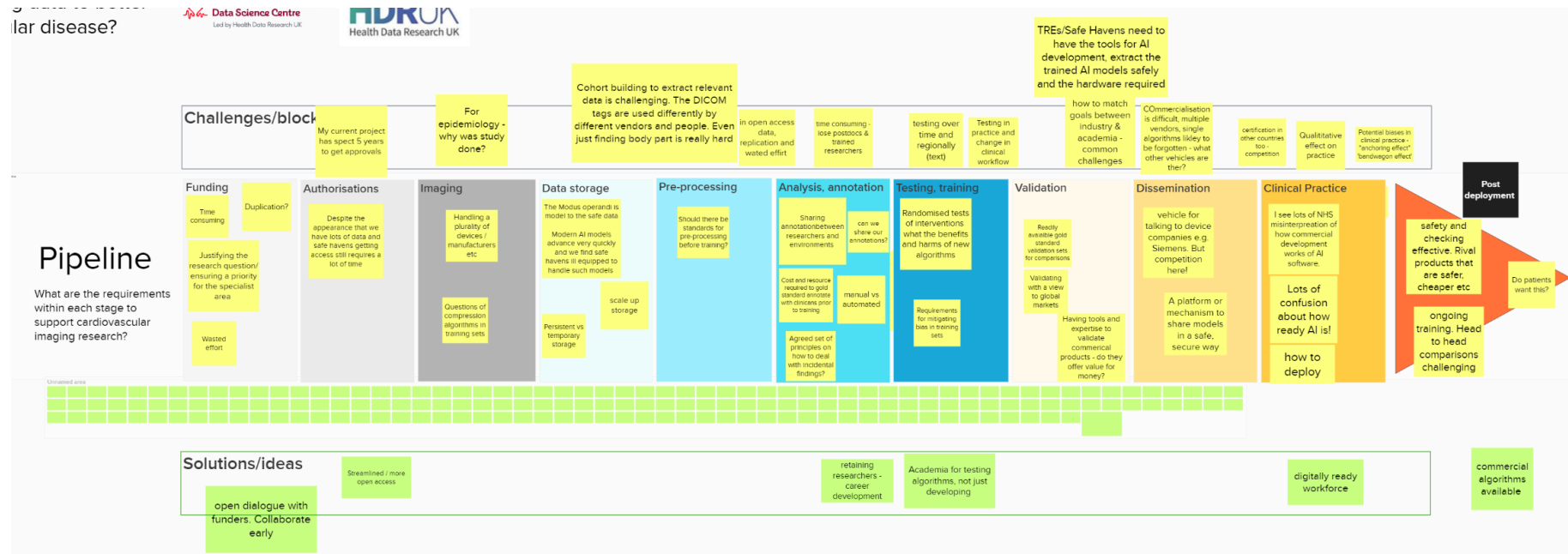
15th 2021



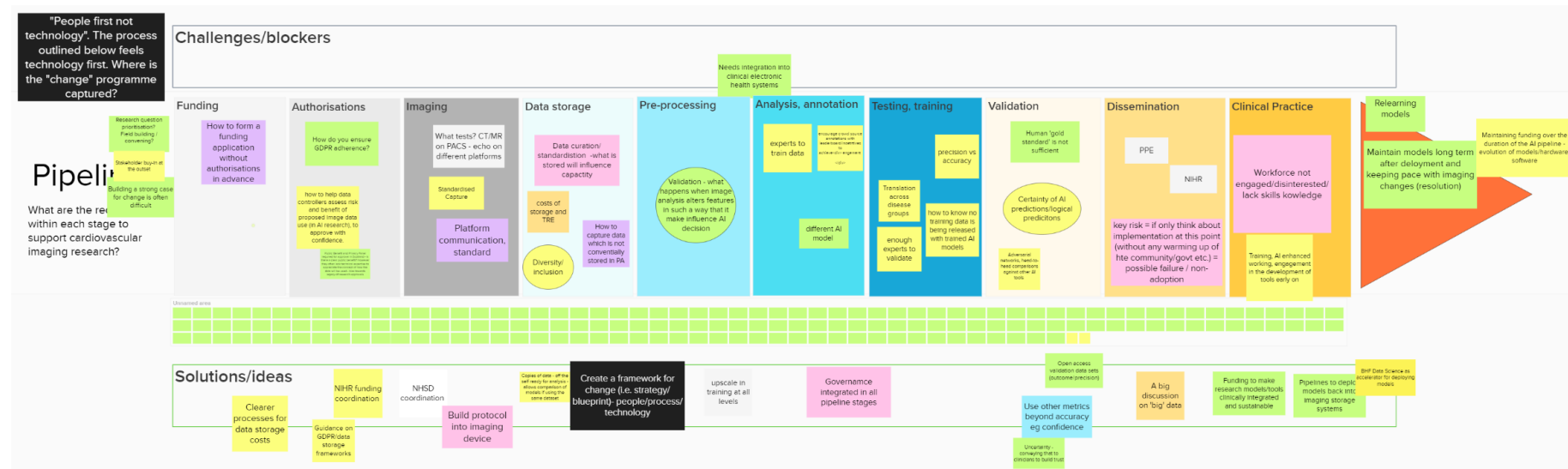
Group 3



Group 4

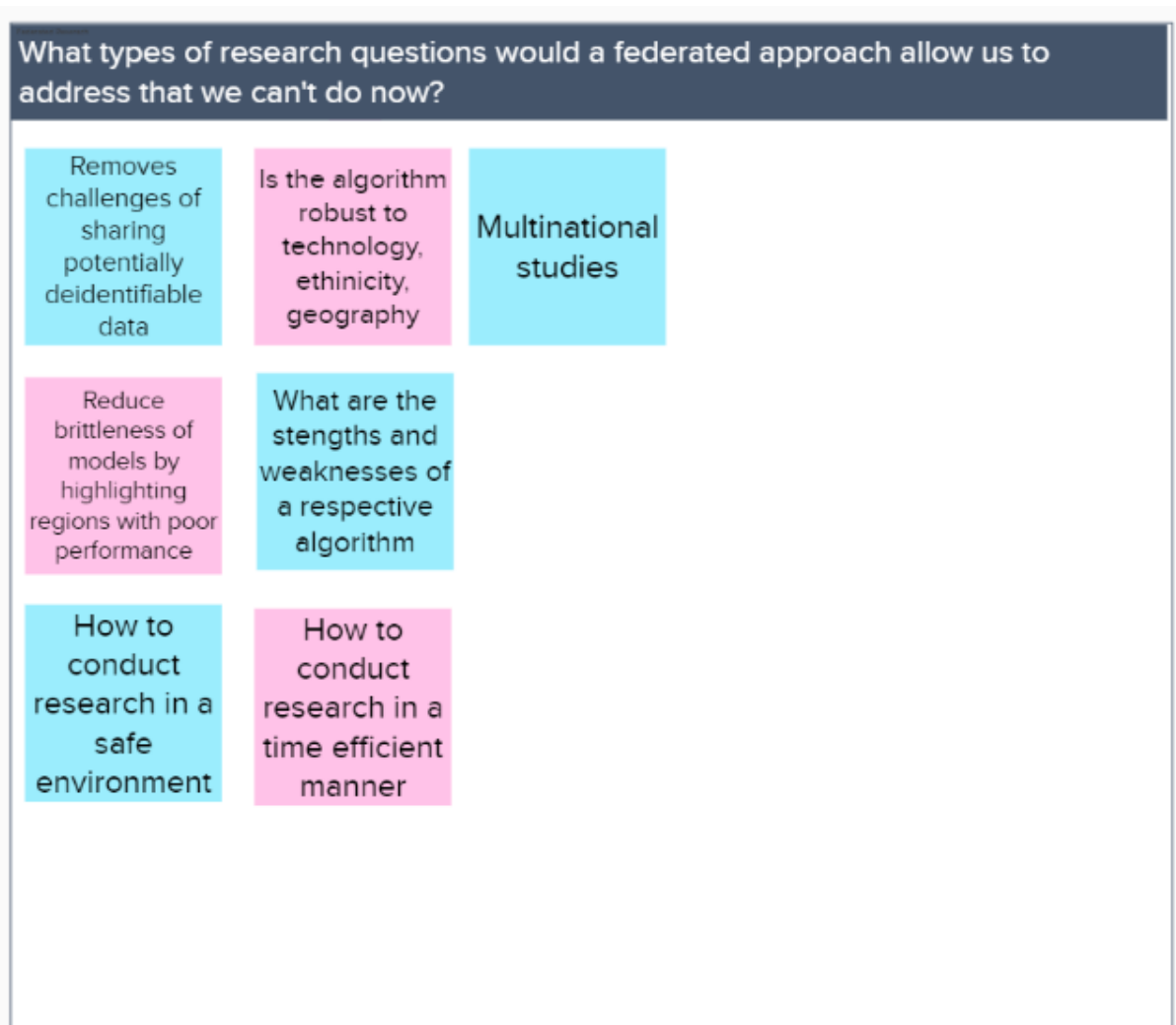


Group 5



Appendix H - Mural board images from session 4 “What types of research questions would a federated approach allow us to address that we can’t do now?”

Group 1



Group 2



Group 3

What types of research questions would a federated approach allow us to address that we can't do now?

Group 4

What types of research questions would a federated approach allow us to address that we can't do now?

Ability to create more diverse data where diversity is required.

Simply - looking at associations with image derived phenotypes

Effects of ageing and environment on cardiovascular health

Testing how biases in data affect models.

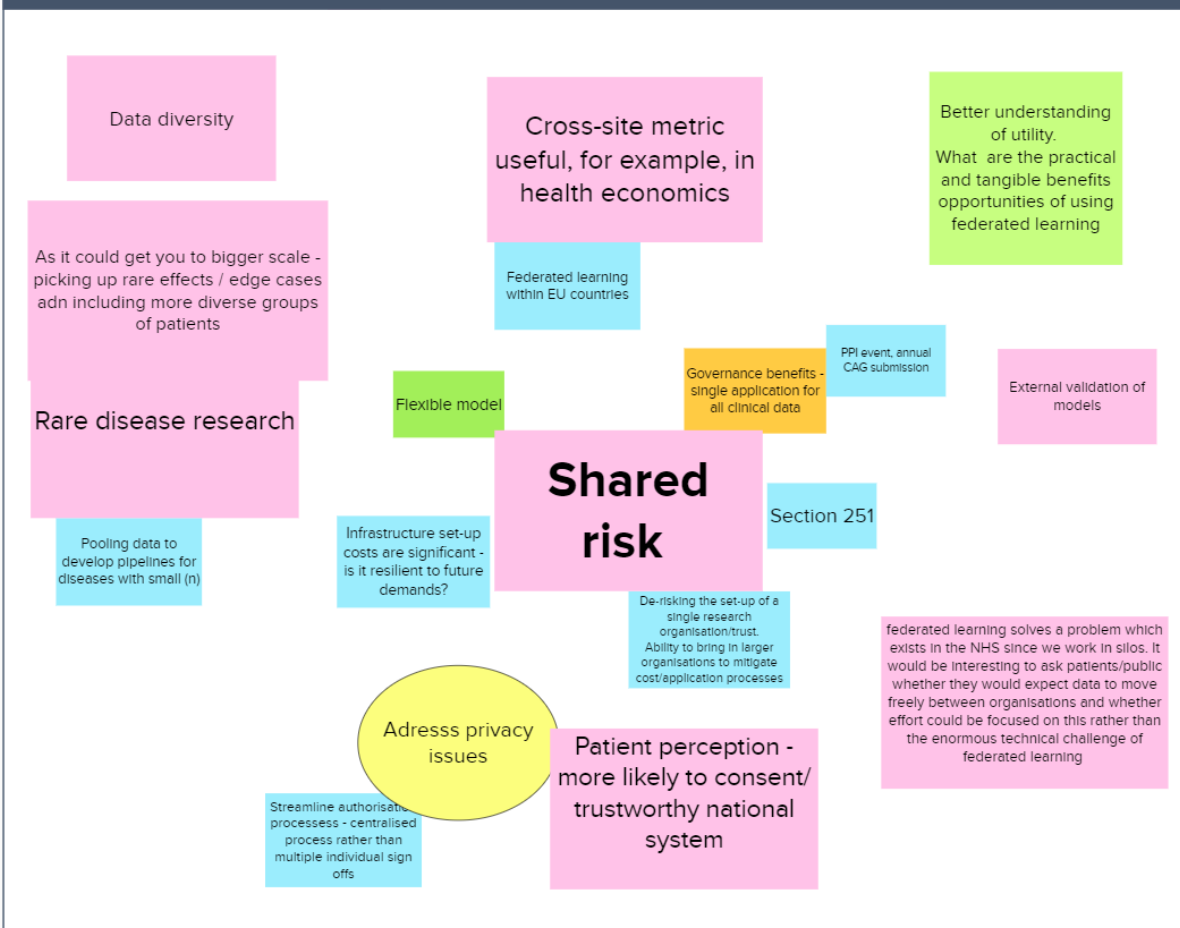
e.g. effects of ethnicity

How to handle outliers / rare phenotypes

variations - statistical shape models and outliers

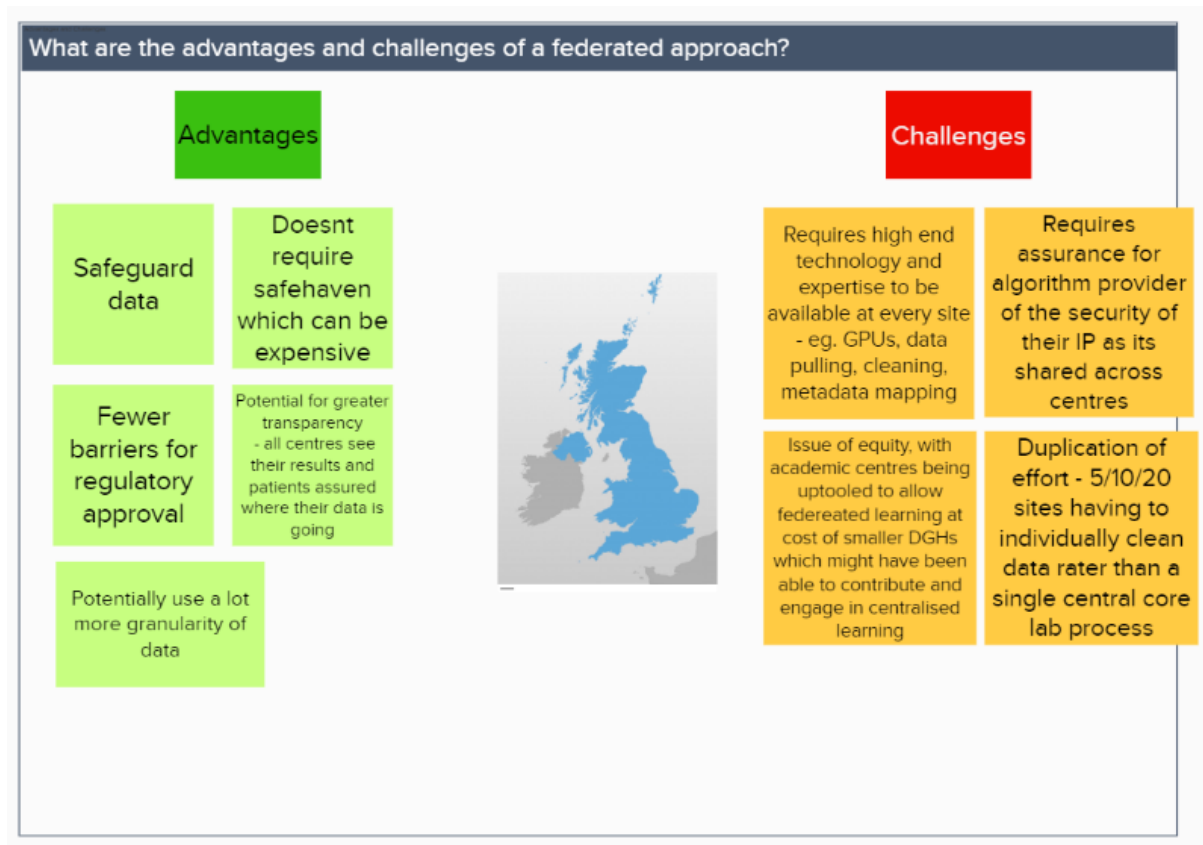
Group 5

What types of research questions would a federated approach allow us to address that we can't do now?

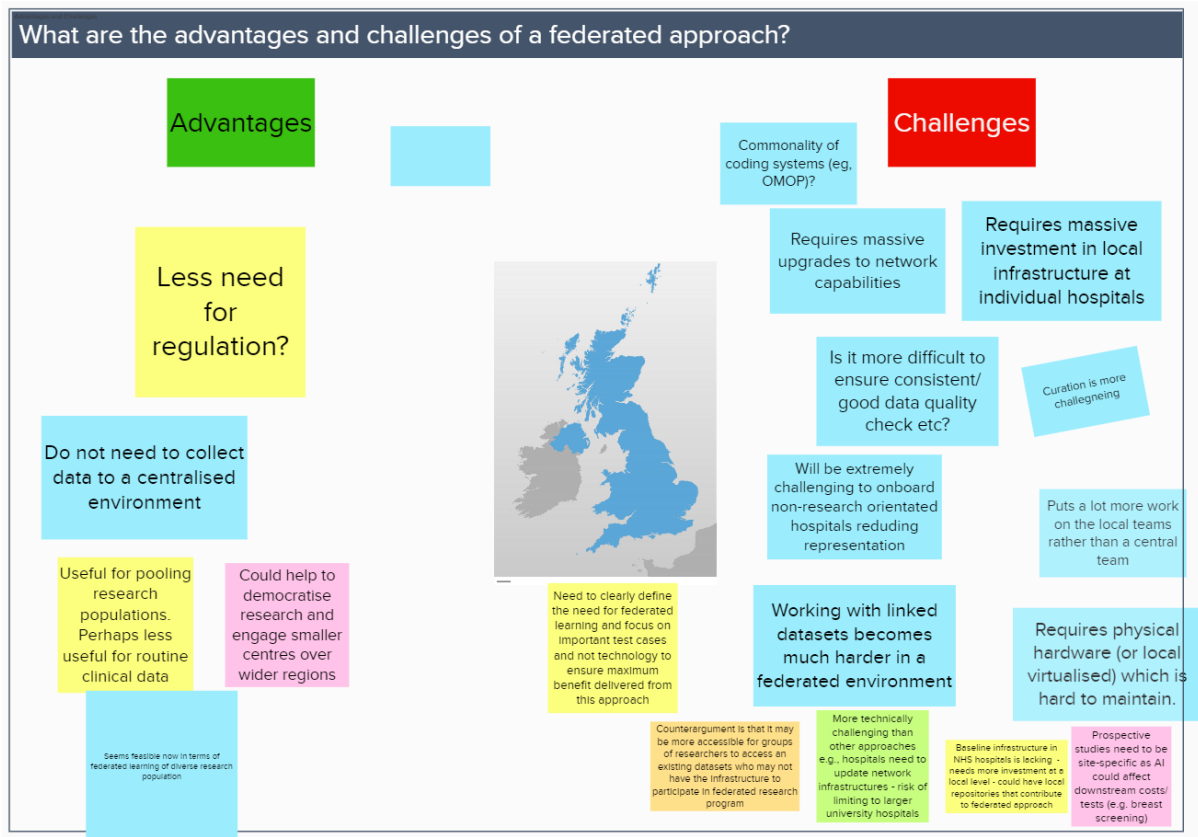


Appendix I - Mural board images from session 4 “What are the advantages and challenges of a federated approach?”

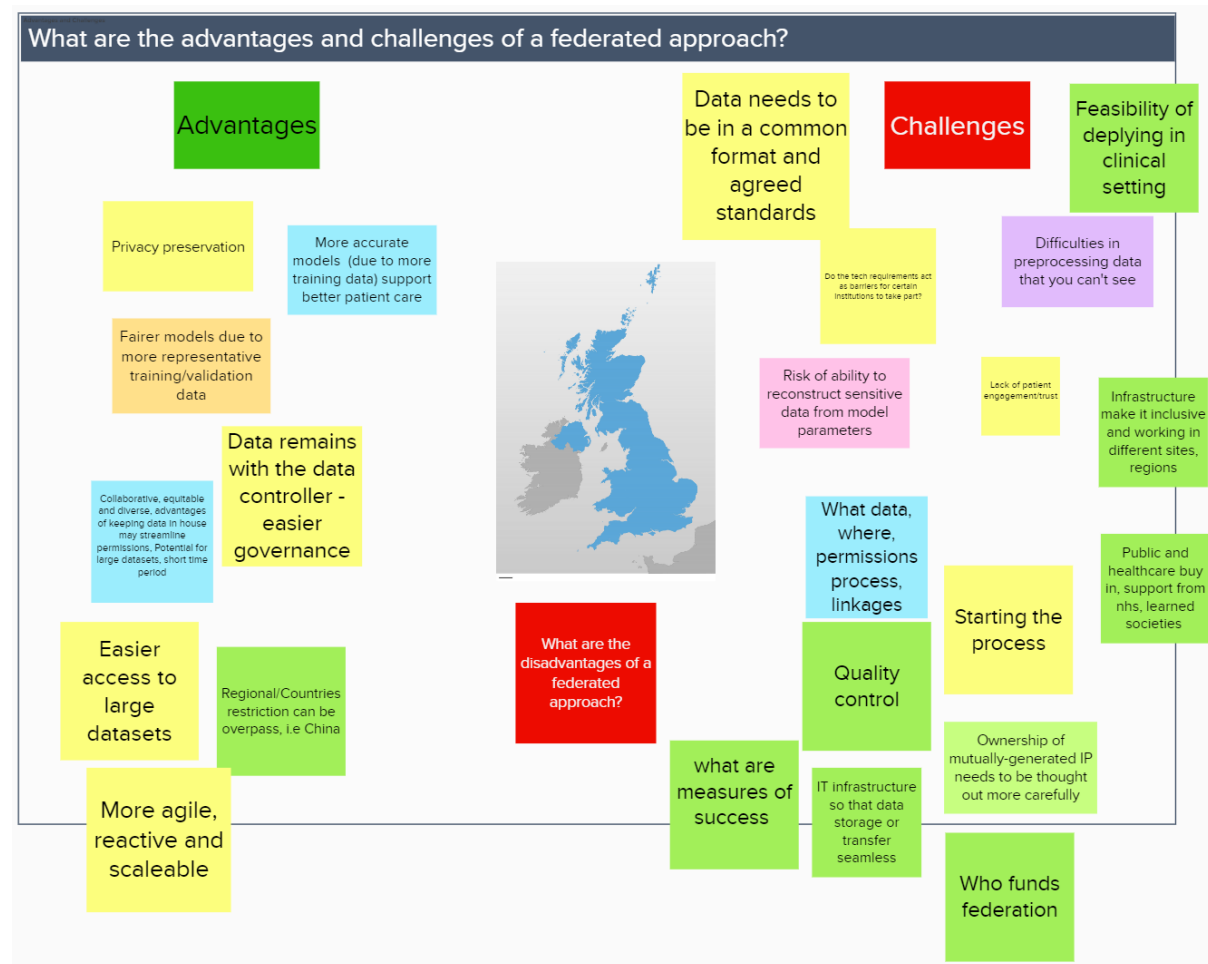
Group 1



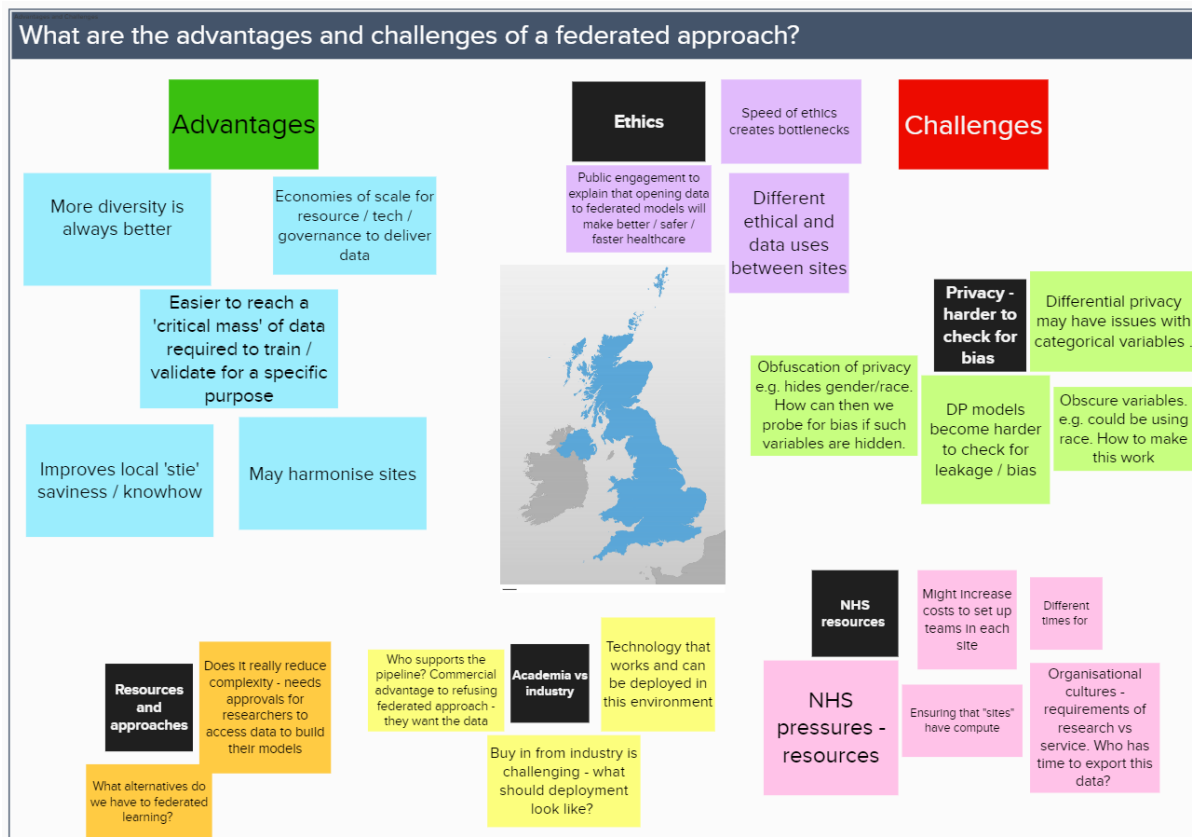
Group 2



Group 3



Group 4

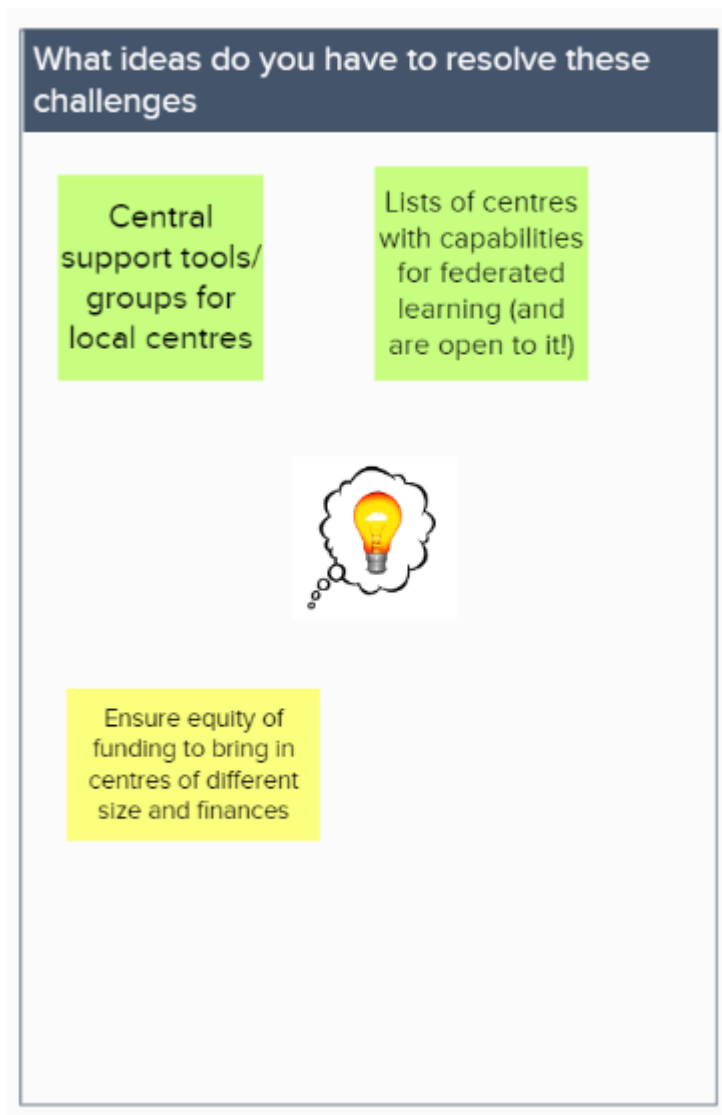


Group 5



Appendix J - Mural board images from session 4 “What ideas do you have to resolve these challenges?”

Group 1



Group 2

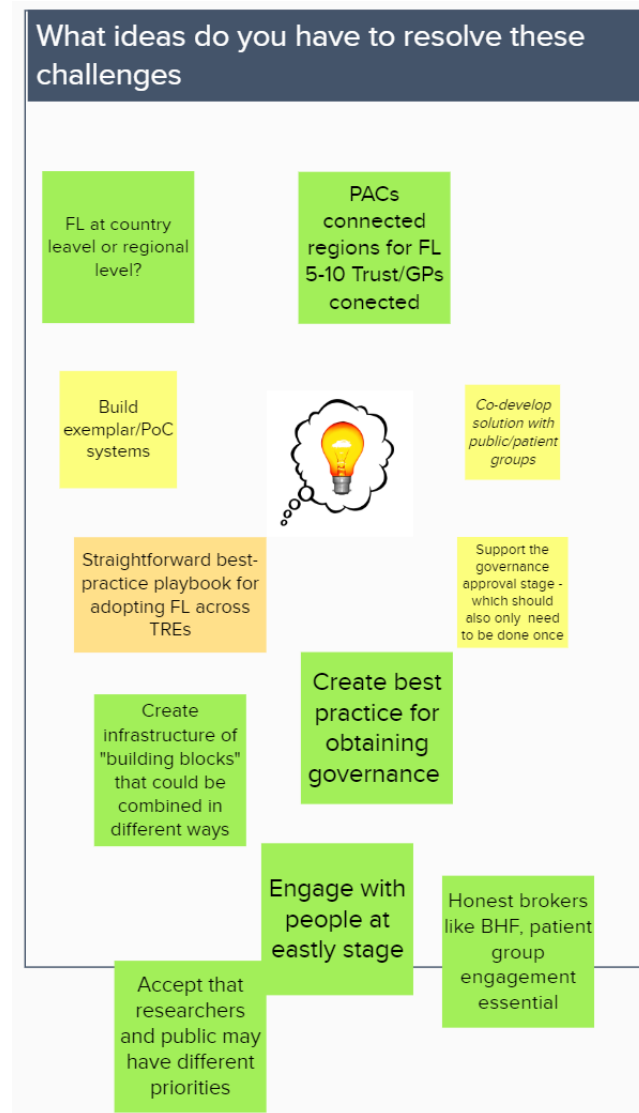
What ideas do you have to resolve these challenges

Can money
invested in
federated learning
be used to
improved hospital
infrastructure?

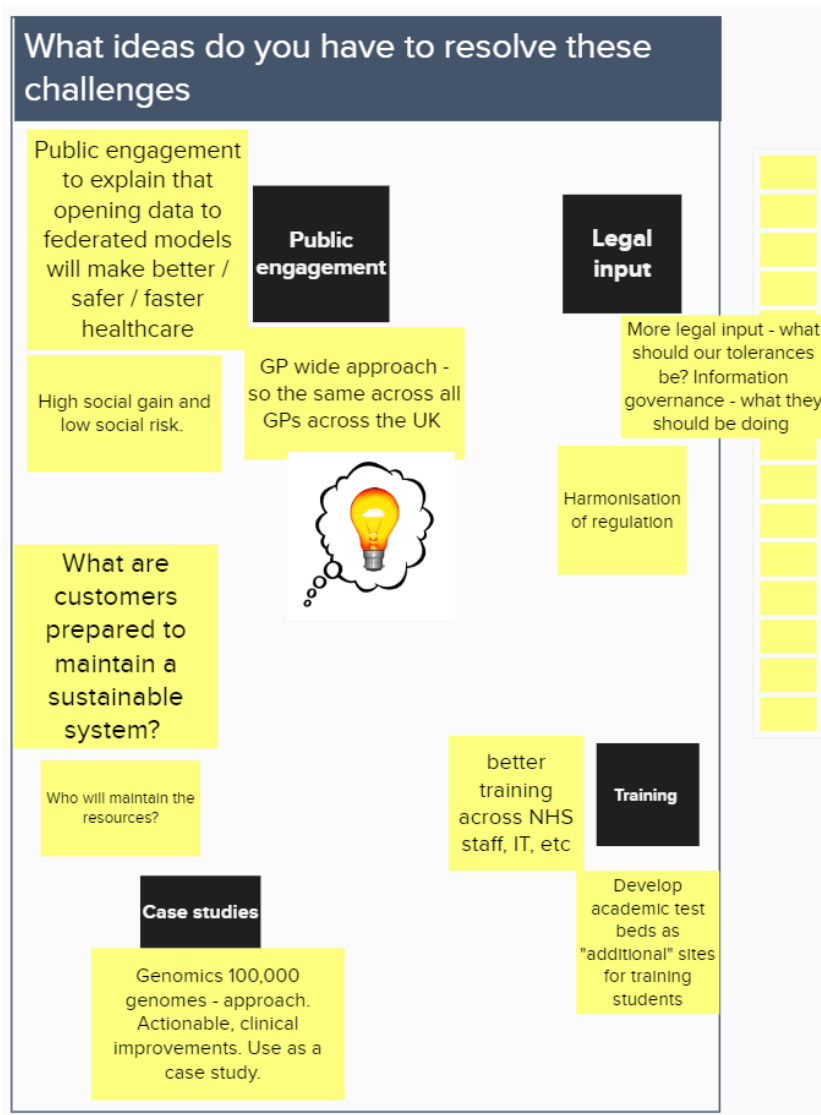
Need more
investment in IT
infrastructure for
hospitals across
the board,
regardless of
specialty



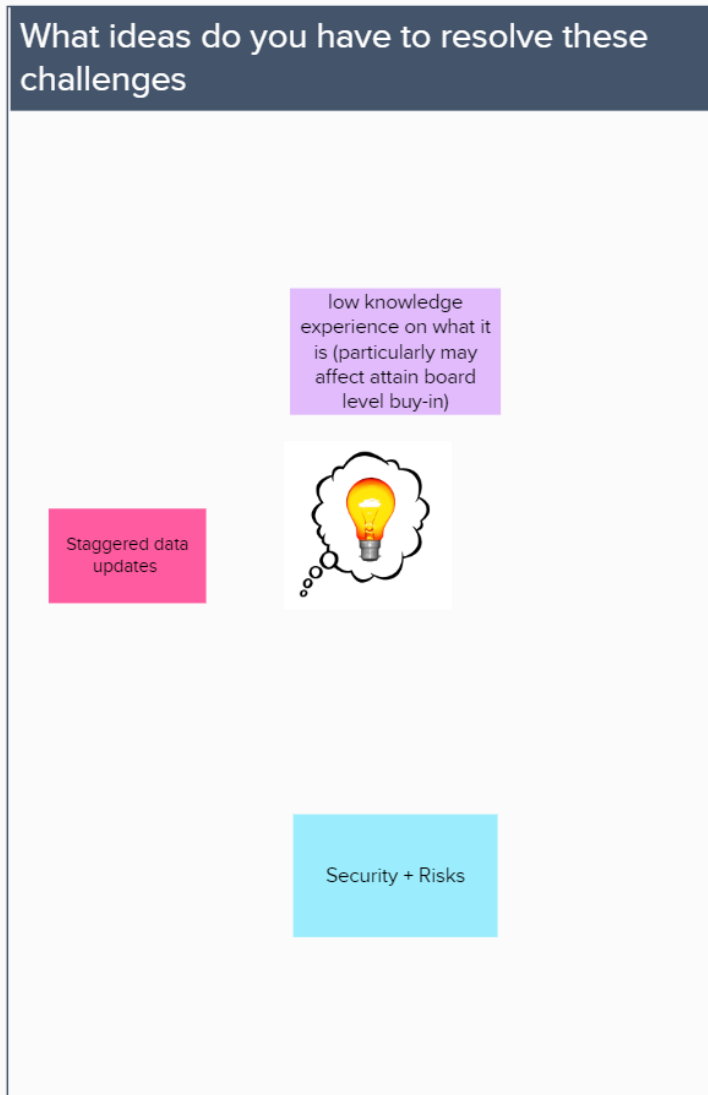
Group 3



Group 4

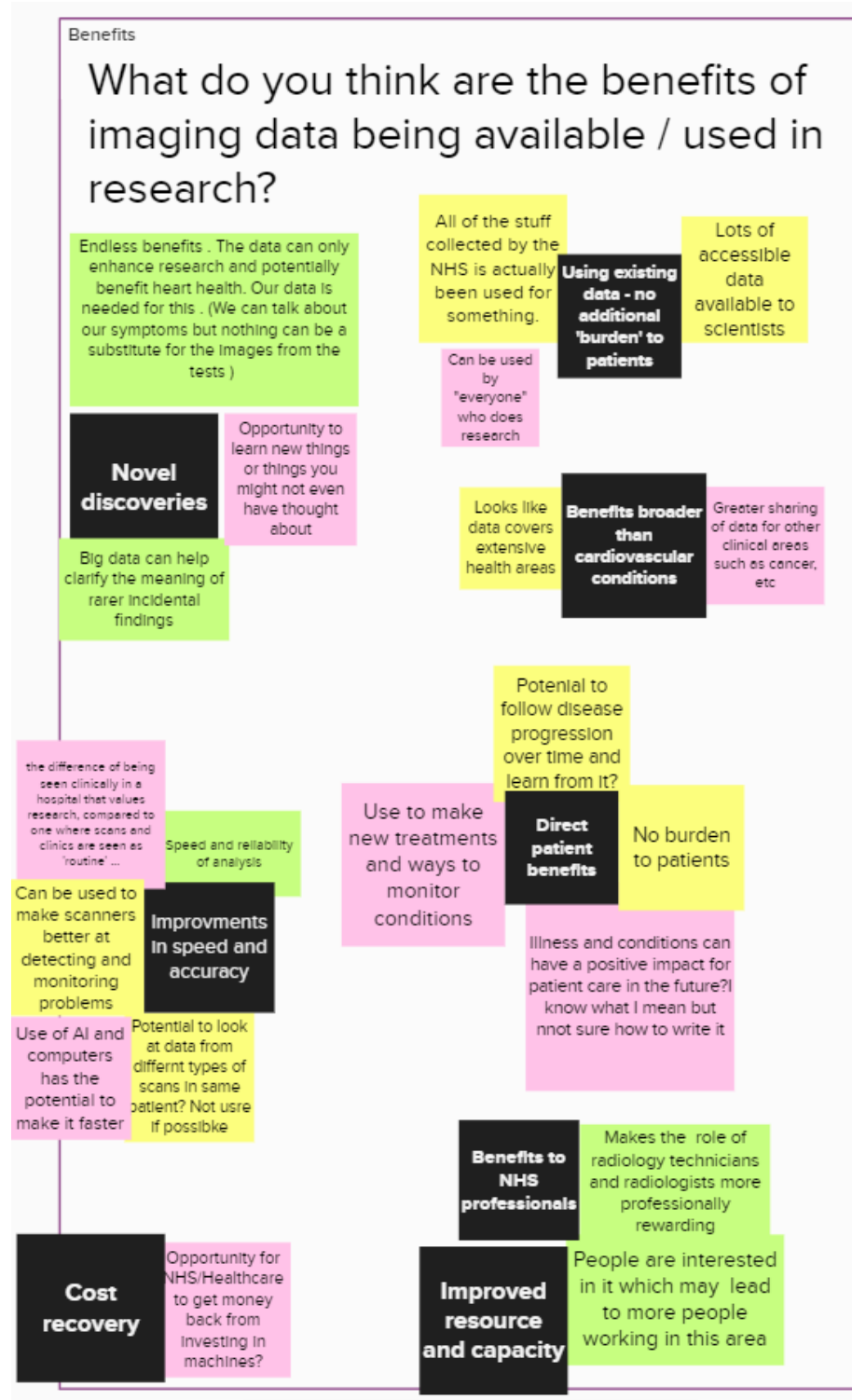


Group 5



Appendix K - Mural board images from public and patient representatives' session

“What do you think are the benefits of imaging data being available/used in research?”





Appendix L - Mural board images from public and patient representatives' session

“What are your concerns, if any, about imaging data being used?”





Appendix M - Mural board images from public and patient representatives' session

“What do you think are good ways of explaining imaging data research? Can you suggest specific words to use or not use, and the kinds of formats we should consider for effective communication?”

