# What is federated analytics and what might it mean for me and my health data?

A guide by FED-NET and PIONEER

PIONEER
Health Data Research Hub

# Contents



PIONEER
Health Data Research Hub

# What is health data and how can it be used?

Health data is the information that doctors, nurses and other healthcare professionals collect about you when you visit a hospital, GP or other healthcare provider.

It includes what symptoms people have, the medical conditions they are diagnosed with, what investigations or treatments people receive, and how people respond to those treatments. This health data is used so that healthcare professionals can provide you with the best possible medical care.

Health data can also be used for research – to develop new treatments for patients, and to improve NHS services.

For health data research to improve healthcare for all patients, it needs to reflect us all - including people from different ethnicities, backgrounds, and with different health problems.

This brochure was written by members of the public and patients for members of the public and patients. It will explore how NHS health data can be used for research, how federated analytics can make health data use safer than ever, and what choices we have about how our health data is used.


# Health Data and the United Kingdom

The UK has a globally unique source of health data.

Healthcare in the UK is mainly delivered by the National Health Service (NHS - www.nhs.uk ). The NHS is free when you need healthcare and available to the entire population. NHS doctors and nurses see about 1 million patients every 36 hours and the NHS performs over 10 million operations each year in England alone.

As NHS health data comes from our entire population, it has huge power to tell us about our health and to be used for research and improving NHS services. It is increasingly common for NHS health data to be held in electronic health records, held on computers rather than in paper records. This makes it much easier to use the data for research, but there are still some challenges to overcome.

This includes making sure data is used responsibly and that every patient's privacy is maintained.

# Health data and its uses

Health data includes information not just about diseases, but about the kinds of people who suffer with a disease. This includes information about their age, whether they are male or female and sometimes their ethnicity or the jobs they do. It includes the other health problems they have, what medical tests or procedures people have undergone and what medications they use.

Health data is used to provide medical care. When health data is used for medical care, the data is identifiable. This means that the healthcare professional knows exactly whose health data it is, and uses the data to provide healthcare specifically for this person.

Health data can also be used for research. This can include discovering new ways of diagnosing or treating a disease, developing new devices for monitoring illnesses and identifying which patients would respond best to different treatments. When health data is used for research, it is de-identified, meaning that the researchers cannot tell who the data is about.

During research, de-identified health data from many, sometimes thousands of people, are placed together and analysed to look for patterns which tell researchers about diseases. When you combine lots of people's data, we call this a "data set".

## Health data includes

- Symptoms
- Previous medical conditions, operations and procedures
- Medications and allergies
- Vital signs e.g. heart rate, breathing rate, blood pressure, and temperature
- Laboratory tests e.g. blood, sputum or urine tests
- Imaging e.g. chest x rays
- New diagnoses e.g. heart attacks, strokes, fractures
- Severity of illness e.g. admitted to intensive care, breathing or organ support needed
- Do Not Attempt Cardiopulmonary Resuscitation (DNACPR) status
- Patient transfers within the hospital e.g. between A&E department, wards and intensive care
- Final outcomes e.g. discharged to their own home
- What further outpatient care was organised e.g. referrals to clinics, physiotherapy

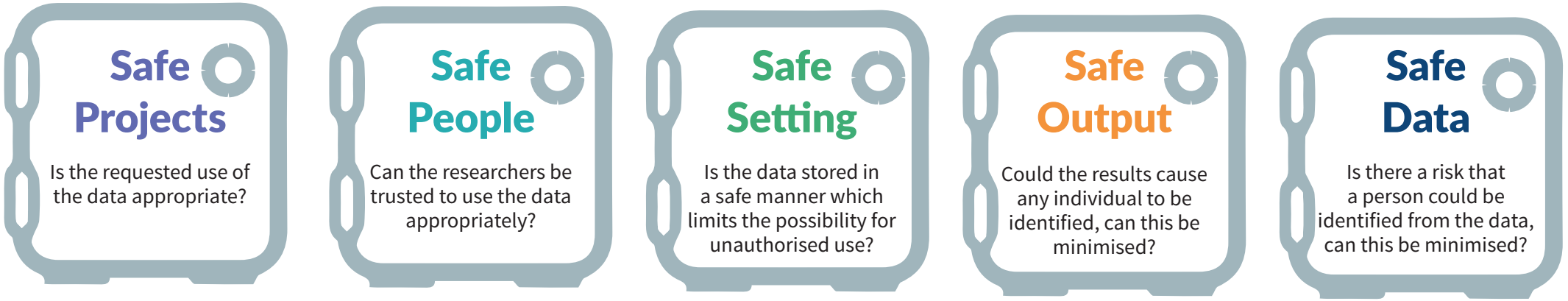
# The 5 Safes - what are they and why are they important?

Health data includes some of the most sensitive information about people. When it is used for research, even when de-identified, it is important that the health data is used in a safe way, so that we can all benefit from the results of important research, but that the privacy of individuals is respected. A framework has been proposed to help make sure that the use of the data is safe. This framework is called the 5 Safes.

The 5 Safes consider what the data will be used for, who will have access to that data and how sure we are that a patient could not be identified from the health data. The 5 Safes also include how the data is accessed for research, where the data is stored, and how the learnings from the research can be shared without identifying any patients within the data.

Whenever a healthcare provider decides to share de-identified health data with a researcher, the 5 Safes act as a check list to make sure data sharing is as safe as possible and in the publics best interest.

The projects are checked to make sure they are important, that access to data will help, and not harm or disadvantage members of the public, and that all needed approvals are in place. The people accessing the data are checked, to make sure they have the necessary training to access health data and that they can be trusted with the data. The data is checked to make sure a person could not be identified from the data. There is usually a legal contract which checks how the results of the research will be shared. Finally, it has to be agreed where and how the data will be accessed.



# What is a Trusted Research Environment (TRE)?

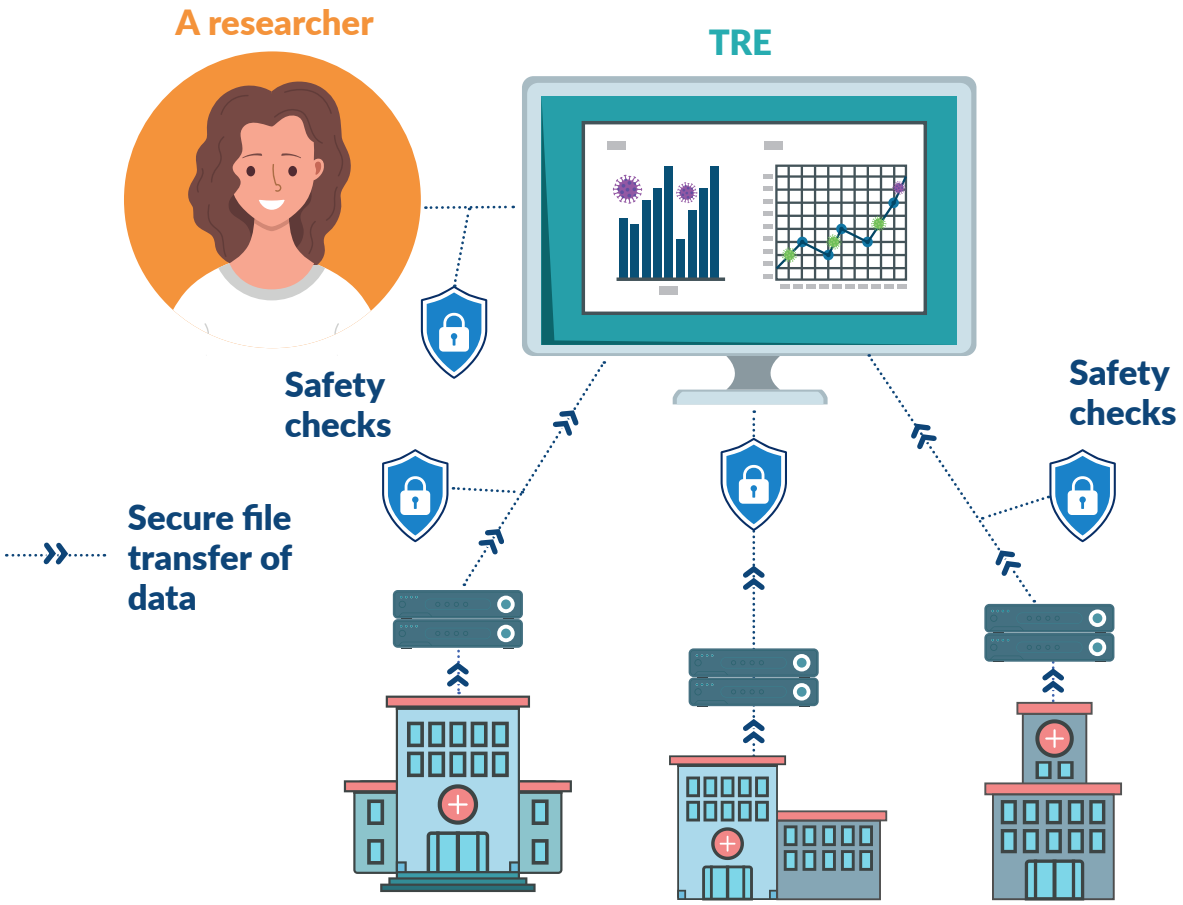
A TRE is a highly secure computer environment where health data can be placed, and approved researchers can access the health data remotely using secure log ins and passwords. TREs are specifically designed to be safe, and only allow specific people to access the data. Importantly, no data can be added or removed from a TRE without specific permissions and with proper checks and audits – meaning a trail is left which can be checked.

Usually, data from many sources is sent to a single TRE to build a dataset that is required for research. For example, de-identified health data from a number of hospitals may be sent to one TRE and combined together to build a dataset.

When the data from different settings is sent to a TRE and combined, checks occur to make sure the data has been collected in the same way. For example, the height of a person can be measured in centimetres (for example 160cm) or metres (1.6m) or feet and inches (5'4"). By physically combining the data in one place, these checks are easy to manage.

Researchers can then apply to access and analyse all of the data in that single TRE. To analyse data, researchers write mathematical computer code (a computer programme) which they apply to the health data to answer questions or identity patterns in health and disease.

The figure below shows how this might happen.

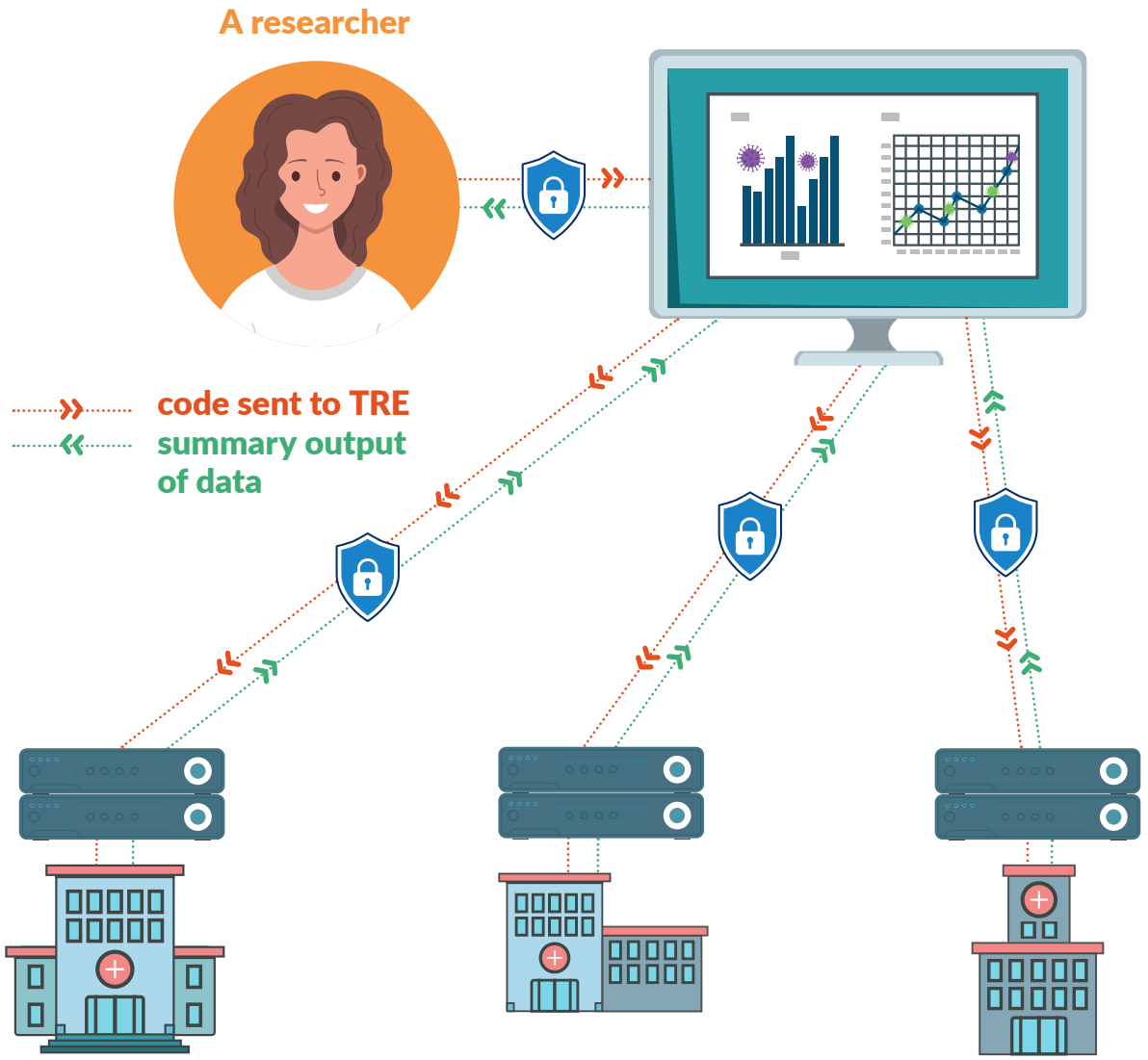


# What is federated analytics?

Federated analytics is a system where the data does not move, and instead the computer code the researchers write is sent to the data.

Sometimes the dataset is made from lots of pots of data in different settings. For example, a dataset studying asthma might include data from several hospitals around the country. Instead of the hospitals sending all that data to one TRE (as described on the previous page), in federated analytics each hospital would build their own TRE and keep their data within that TRE. In this situation, the researcher writes their computer code and this is then sent to each TRE, analysing only the data kept in that one TRE. The results from all the different TREs are then returned to the researchers and combined.

The figure below shows how this might happen.

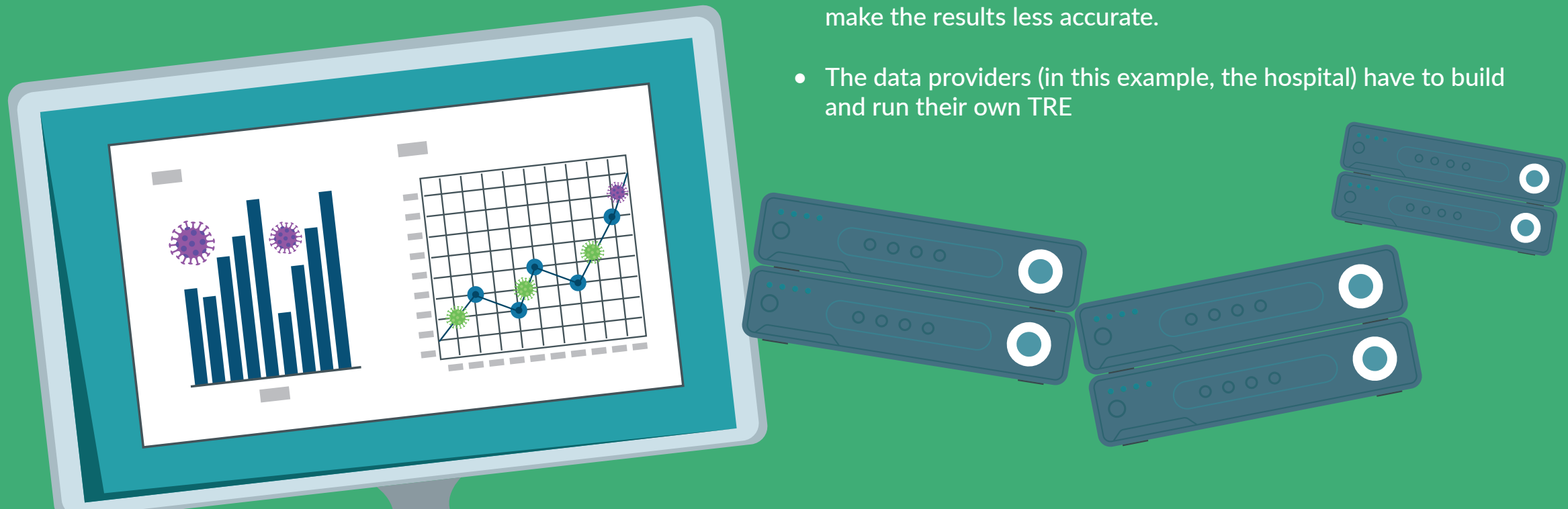


# Benefits of federated analytics

- The data providers (in this example, hospitals) always hold and control their own data. This is better for security.
- It is easier to identify which data has been used for which projects, and this can be done at a local level.

# Challenges with federated analytics

- It is not as easy to check that the data has been collected in the same way.  The data is not combined in one place, each data pot does not “see” the other data pots, and the researchers do not see the data at all, they just write and send code.
- It is unclear if analysing the data in “pots” and then combining the results alters how useful the results will be as this might make the results less accurate.
- The data providers (in this example, the hospital) have to build and run their own TRE



# What are the advantages and disadvantages of federated analytics?

## Advantages

- The local organisation keeps full control of the data and only has to allow access to its own data
- The data stays put and is not sent anywhere. Computer code to perform the research is sent to the data
- The local organisation has control over who accesses the data
- The local organisation can audit and check which data is used for which project

## Disadvantages

- As each organisation only "sees" its own data, it is harder to check that the data has been collected in the same way
- It is harder to know what the whole dataset looks like, as data is spread across organisations
- The researcher does not see the data, and cannot check if the dataset has everything they need for the research project
- It is unknown whether analysing data in "pots" and combing the results loses any meaning or accuracy
- The local organisation has to know how to build and run a TRE

# What is FED-NET and what is the project trying to do?

## FED-NET is a partnership between 4 organisations

NHS University Hospitals Birmingham NHS Foundation Trust

UNIVERSITY OF BIRMINGHAM

NHS Nottingham University Hospitals NHS Trust

The University of Nottingham

- FED-NET was funded by DARE, Innovate UK and is being led by the HDR-UK Hub called PIONEER (see following page).
- FED-NET was designed to build and test federated analytics.

- FED-NET will build two TREs, one in University Hospitals Birmingham NHS Foundation Trust and one in Nottingham University Hospitals NHS Trust, build a dataset (about asthma), and then will ask researchers from University of Birmingham and University of Nottingham to analyse the data in two ways.

1. **When the data is combined and held together in one TRE**
2. **When the data is split across two TREs using federated analytics.**

The questions FED-NET will explore are

1. How secure are federated TREs from hacking and other cyber attacks?
2. How easy is it to describe what the data contains and check it has been collected in the same way when the data has been split over two sites?
3. How accurate are the results, comparing federated analytics compared to the data being held in one TRE?

# What is PIONEER and how does PIONEER make data sharing decisions?

PIONEER is a Health Data hub which brings together health data collected across the UK, and enables researchers to access this data to improve our understanding of how diseases develop, can be detected and should be treated. This has led to better health care locally, nationally and around the world.

PIONEER is focused on acute care. Acute care is any unplanned health care. This can include going to the Emergency Department at a hospital, calling for an ambulance or seeing a GP or pharmacist using an emergency appointment. It also includes any unplanned event during care, such as developing an infection in hospital, having a fall on a hospital ward or experiencing a complication after surgery.

The PIONEER team understands that patients and the public generally support access to de-identified health data if it improves the lives of others, but that there are also concerns about privacy, data security and data misuse.

Members of the public play a key role in all the decisions PIONEER makes about data sharing. Every single data request is run past our Data Trust Committee, a group of people who are patients and members of the public, each with different health experiences and of different ages and backgrounds.

This group review data requests and decide on whether they believe the work is in the public's best interests. If the Data Trust Committee do not feel a data request is of benefit to the public, that data request is not supported. If the Data Trust Committee agree that data access is in the public's interest, data is only shared using a specific legal contract, and data sharing follows the 5 Safes.

## How can I learn more?

PIONEER is improving patient care by making health data available to healthcare staff and researchers. Data access is only allowed with a legal contract. This ensures the data is used to benefit peoples' health.

PIONEER works to improve healthcare for patients.

You can learn more about PIONEER, the work that we do, and the projects we have supported by visiting the PIONEER website 🌍 www.pioneerdatahub.co.uk

Or by contacting the team ✉ pioneer@uhb.nhs.uk

The website also includes information about the type of data PIONEER holds and examples of how PIONEER has improved our knowledge of disease and the care the NHS provides to patients.

## What are my choices?

If you are a patient who has received acute care from a PIONEER data partner, your de-identified health data could have been used in our research. Your anonymised data may have helped improve health care for everyone.

If you want your data to be used to help answer important health research questions, or plan better services for you, your family and others, then you do not need to do anything. Your data is automatically used for these purposes.

However, you can choose to stop your patient information being used for research and planning. The option to opt out applies to anyone over the age of 13 who lives in England. If you live elsewhere in the UK, your data may be handled differently.

Further information is available at 🌍 www.nhs.uk/your-nhs-data-matters

For more information about PIONEER, please email:
pioneer@uhb.nhs.uk.

To learn more, visit our website at
www.pioneerdatahub.co.uk

NHS
University Hospitals Birmingham
NHS Foundation Trust

UNIVERSITY OF BIRMINGHAM

NHS
Nottingham University Hospitals
NHS Trust

The University of Nottingham