

# TRE-FX

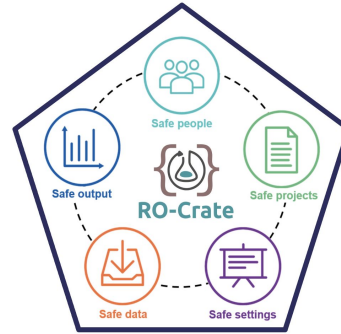
DELIVERING A FEDERATED NETWORK OF  
TRES TO ENABLE SAFE ANALYTICS

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

A guide by TRE-FX

# Contents

1. What is health data?
2. Health data and its uses
3. What is a Trusted Research Environment (TRE)?
4. What are the Five Safes?
5. What is federated analytics and why is it needed?
6. Advantages and disadvantages of federated analytics
7. What is TRE-FX and what is the project trying to do?



**TRE-FX**  
DELIVERING A FEDERATED NETWORK OF  
TRES TO ENABLE SAFE ANALYTICS

# What is health data?

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.

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 data can be used for research, how federated analytics can make data use safer than ever, and what choices we have about how our data is used.



## Health Data and the United Kingdom

The UK has a globally unique source of health data. The National Health Service (NHS - [www.nhs.uk](http://www.nhs.uk)) is a single publicly funded health service in the UK which is free at the point of health care need and available to the entire population. The NHS has been caring for us for the last 70 years, serves our entire population of 66 million people and NHS doctors and nurses see about 1 million patients every 36 hours.

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.

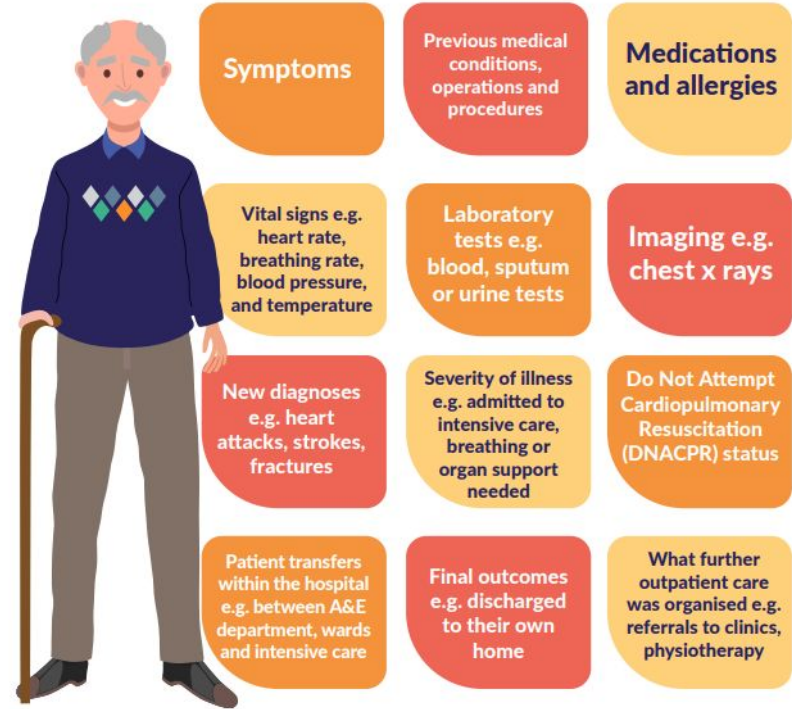
# Health data – its uses

Health data encompasses more than just information about diseases; it includes details about individuals affected by those diseases. This encompasses their age, gender, ethnicity, occupation, as well as their overall health, medical tests, procedures, and medications.

Health data plays a vital role in delivering personalised medical care. When health data is utilised for healthcare purposes, it is identifiable, meaning healthcare professionals have access to specific individuals' data and use it to provide tailored and targeted healthcare services for them.

This personalised approach enhances the quality and effectiveness of medical care.

## Health data includes



## Health data - its uses

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 most to different treatments.

When health data is used for research, it is de-identified, which is when personal information, such as names or addresses are removed. This means 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".



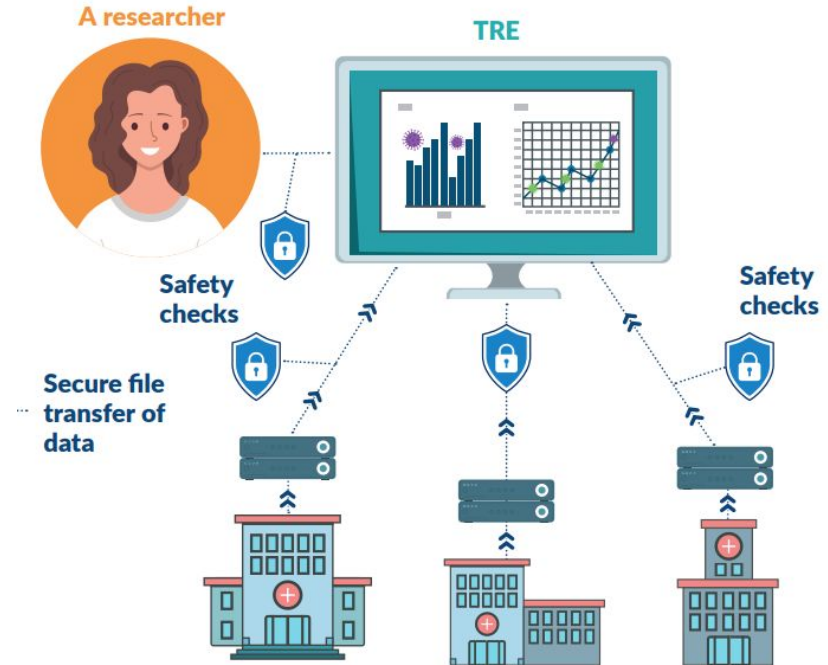
## What is a Trusted Research Environment (TRE)?

A Trusted Research Environment (TRE) is a highly secure computer system where health data is stored. Approved researchers can access this data remotely by using secure logins and passwords. TREs are designed to be safe, allowing only authorised individuals to access the data.

Data cannot be added or removed without proper permissions, ensuring transparency and accountability. Multiple sources of data can be combined in a TRE to create a comprehensive dataset for research.

Researchers can apply to access and analyse the data using computer programs they develop. While TREs are secure, organisations must still follow legal requirements when sending data to a TRE.

The figure below shows how this might happen.



# The Five Safes – what are they and why are they important?

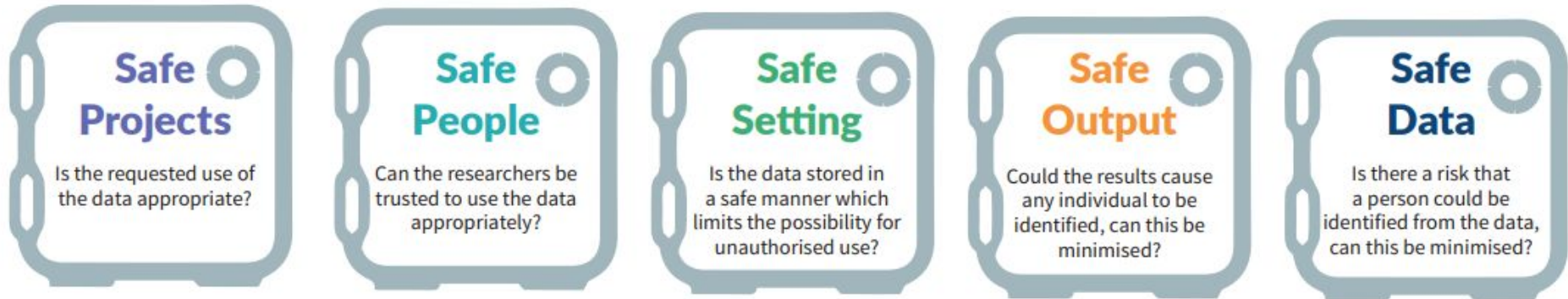
Health data contains personal and sensitive information about people. When this data is used for research, it's crucial to handle it safely, even if it has been made anonymous. This way, we can all benefit from important research findings while respecting individuals' privacy. To ensure safe data usage, a framework called the Five Safes has been introduced.

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.

Whenever a healthcare provider decides to share de-identified health data with a researcher, the Five Safes serve as a checklist to ensure the safety of data sharing in the best interest of the public.

The projects requesting access to the data are thoroughly evaluated to make sure that using the data will benefit and not harm the public. The individuals accessing the data undergo careful checks to ensure they have the required training and can be trusted with sensitive health information. The data itself is carefully examined to ensure that it cannot be used to identify any individual. Typically, there is a legal contract that specifies how the research results will be shared while maintaining privacy. Finally, agreements are reached on where and how the data will be accessed securely.

By following the Five Safes principles, health data can be used responsibly, leading to valuable research outcomes while safeguarding the privacy of individuals.





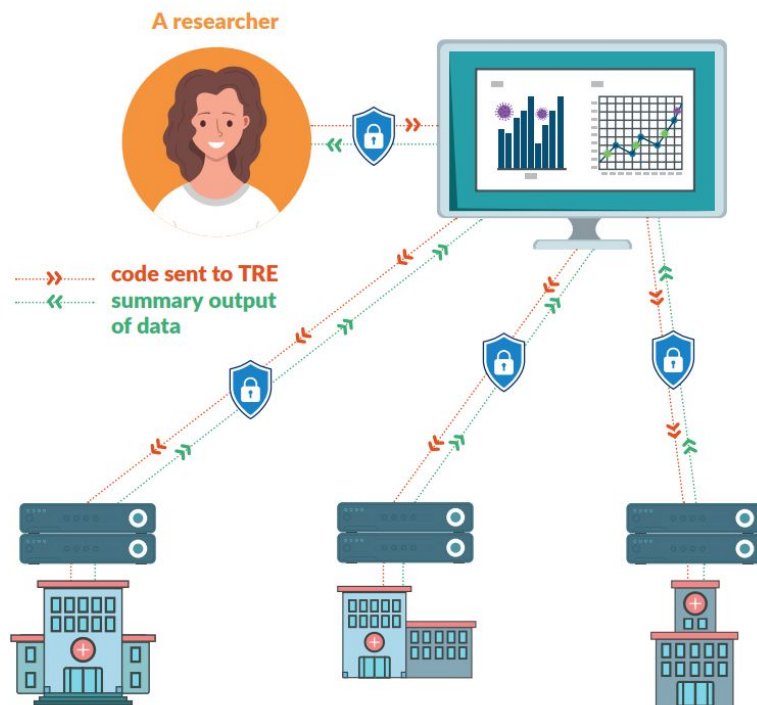
# What is federated analytics?

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 above), 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.

This figure shows how this might happen.



# Federated analytics – why is it needed?

Federated analytics is an innovative approach that enables data analysis and machine learning while respecting privacy and security. They allow organisations and individuals to collectively gain insights and train models without directly sharing their raw data.

The key reasons it is needed are as follows:

- This devolved approach reduces the need for large-scale data transfers, leading to faster and more efficient processing.
- Allows extraction of useful information like trends in a population's health without revealing any individual's specific details.
- Allows analysis on a larger and more diverse dataset, resulting in more accurate and comprehensive insights that represent a broader population or user base.

- Promotes privacy preservation, as data remains under the control of the data owners.
- Enables collaboration across organisations and individuals.
- Supports valuable insights and advancements in healthcare while safeguarding personal privacy.
- Data remains private while contributing to the collective knowledge embedded in the shared machine learning model.
- By enabling collaboration and knowledge sharing, federated analytics supports innovation and discovery.
- Machine learning models can be trained on local devices to capture unique insights.



# 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

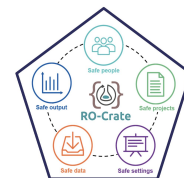
It is harder to know what the whole dataset looks like, as data is spread across organisations

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 TRE-FX and what is the project trying to do?



**TRE-FX**  
DELIVERING A FEDERATED NETWORK OF  
TRES TO ENABLE SAFE ANALYTICS

*It is hard for a researcher to perform analysis across multiple TREs, for example when data is to be analysed across geographical or governance boundaries, such as the devolved nature of healthcare in the United Kingdom. Yet this ability is urgently needed.*

Analysis across a federation of TREs would enable timely analysis of UK wide scattered data to answer urgent questions, as we needed in the COVID-19 pandemic. The technologies and standards we need to be able to do this are available now. They do not need to be invented.

TRE-FX is assembling leading technology providers from [ELIXIR-UK](#) and [HDR UK](#), with three TRE providers and two leading analysis platforms to show through a real reference implementation how we can use secure Research Objects to move between TREs while still supporting the **Five Safes principles** that govern and protect patient data; all overseen by patient representatives.

The impact will be a step change for how researchers can safely combine data from many sources, and for how data providers from any sector can safely implement this using technology and standards we already have today.

**DARE UK**

**UKRI** UK Research and Innovation

**MANCHESTER**  
1824



Swansea University  
Prifysgol Abertawe

**The University of Nottingham**

**University of Dundee**

**NHS**  
University Hospitals  
Birmingham  
NHS Foundation Trust

**UNIVERSITY OF LIVERPOOL**

**PIONEER**  
Health Data Research Hub

**HDRUK**  
Health Data Research UK

**Bitfount**

**DataSHIELD**  
Secure Bioscience Collaboration



For more information about TRE-FX please email:  
[contact@trefx.uk](mailto:contact@trefx.uk)

To learn more, visit our website at  
[www.trefx.uk](http://www.trefx.uk)



**TRE-FX**  
DELIVERING A FEDERATED NETWORK OF  
TREs TO ENABLE SAFE ANALYTICS



The University of  
**Nottingham**



University  
of Dundee



UNIVERSITY OF  
**LIVERPOOL**



Swansea University  
Prifysgol Abertawe



**PIONEER**  
Health Data Research Hub



**Bitfount**

**DARE UK**



**UK Research  
and Innovation**

**HDRUK**  
Health Data Research UK