Beating cancer with data

CANCER RESEARCH UK

Together we are beating cancer

Introduction

We're investing in research that uses the power of data to outsmart cancer and help people live longer, better lives.

This document explains:

- what research data is
- the role of research data in cancer research
- Cancer Research UK's involvement in this area
- the four priorities of our Research Data Strategy

We've created this document for the public and involved patients and the public in its development. For a more in-depth summary, see our **Research Data Strategy** launched in 2022.

Glossary

AI: Artificial intelligence. AI involves creating computer systems to carry out tasks that previously needed human brainpower. It's invaluable to researchers as it allows large amounts of data to be analysed at speed.

Anonymisation: The process of removing personal information (such as name, date of birth or address) so the details cannot be linked or traced back to a patient.

Big data: Extremely large or complex amounts of data that cannot be handled by humans and require special methods to be processed and analysed by computers.

Computer vision: An area of artificial intelligence that deals with how computers can be made to gain advanced understanding from digital images or videos. The computer can then make recommendations or take actions when it sees problems.

CT: Computerised tomography. CT scans use several X-ray images and computer processing to create detailed images of the body and organs.

CYP: Children and young people (used in the context of cancer for children and young people).

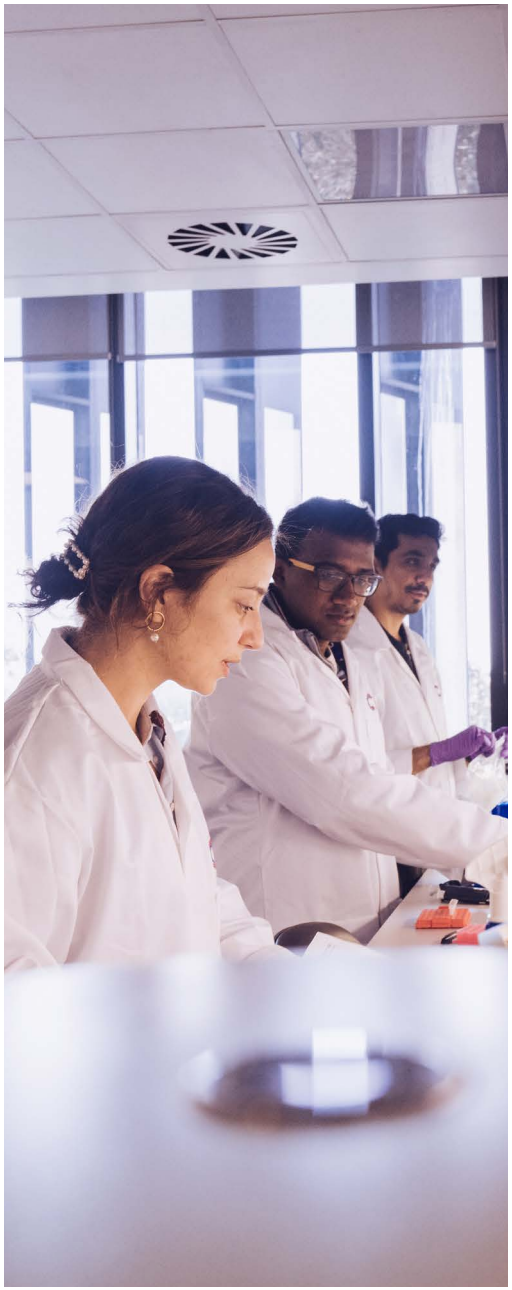
Data science: A form of scientific research that uses data to gain useful information and make real-life decisions. The process involves collecting, organising and analysing data, as well as presenting it in a way that's easier to understand. The process can involve different new technologies, such as artificial intelligence (AI).

Machine learning: An area of artificial intelligence which involves computer programmes learning for themselves from large amounts of data. In medicine, this is valuable for tasks like identifying tumours in scans that doctors might have missed.

MRI: Magnetic resonance imaging. MRI scans use strong magnetic fields and radio waves to produce detailed images of the inside of the body.

Patient data: Medical data about patients. This includes data about things like treatments, diagnoses and appointments. This data is anonymised so as not to reveal the patient's identity.

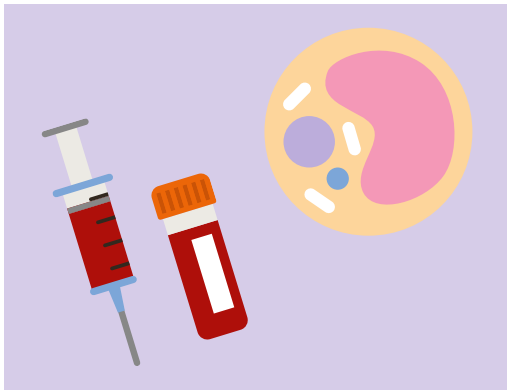
PPIE: Patient and public involvement and engagement.



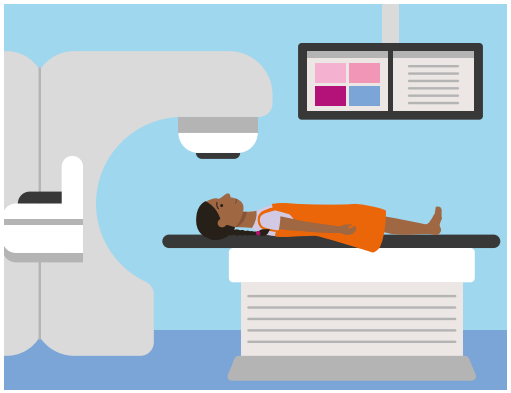
What is research data?

Research data is information collected, processed and studied for research.

There's a variety of different data that can help us learn more about cancer and how to beat it, including:

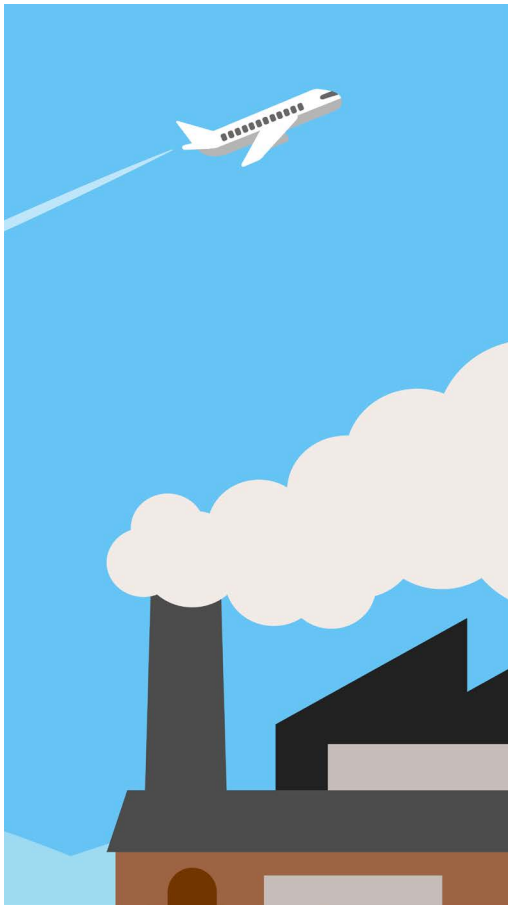


Samples from blood, cells or tissue

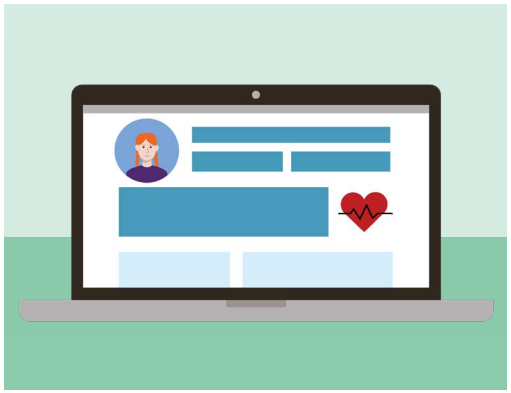


Imaging from X-ray, MRI and CT scans

Findings from clinical trials



Data from health and fitness devices



NHS electronic health records

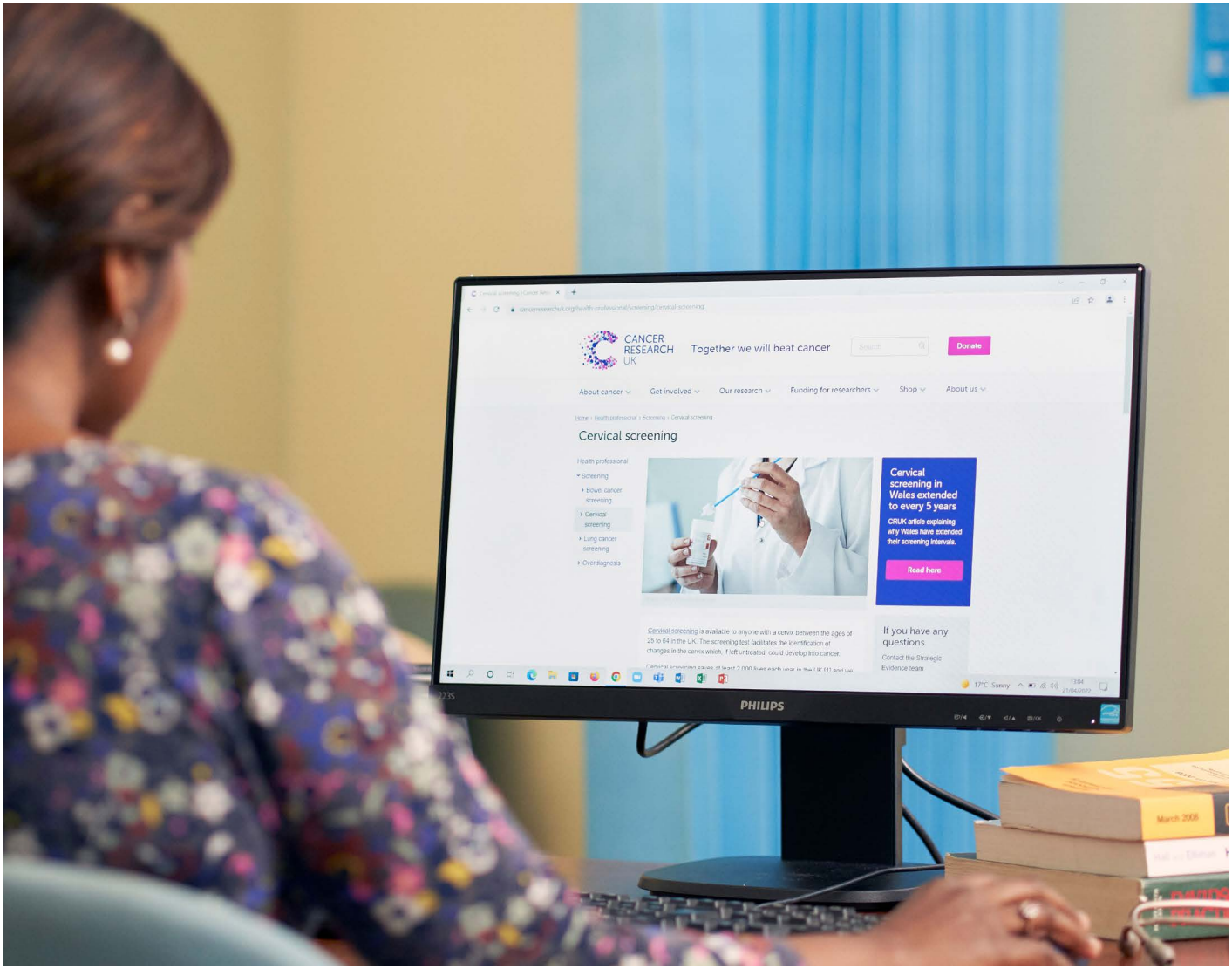
Data from public services such as schools or census responses

Environmental data, such as air pollution levels

How can patient data help cancer research?

Whenever you interact with the health service, for example during a GP appointment or hospital visit, they may collect data about you, your health and your lifestyle.

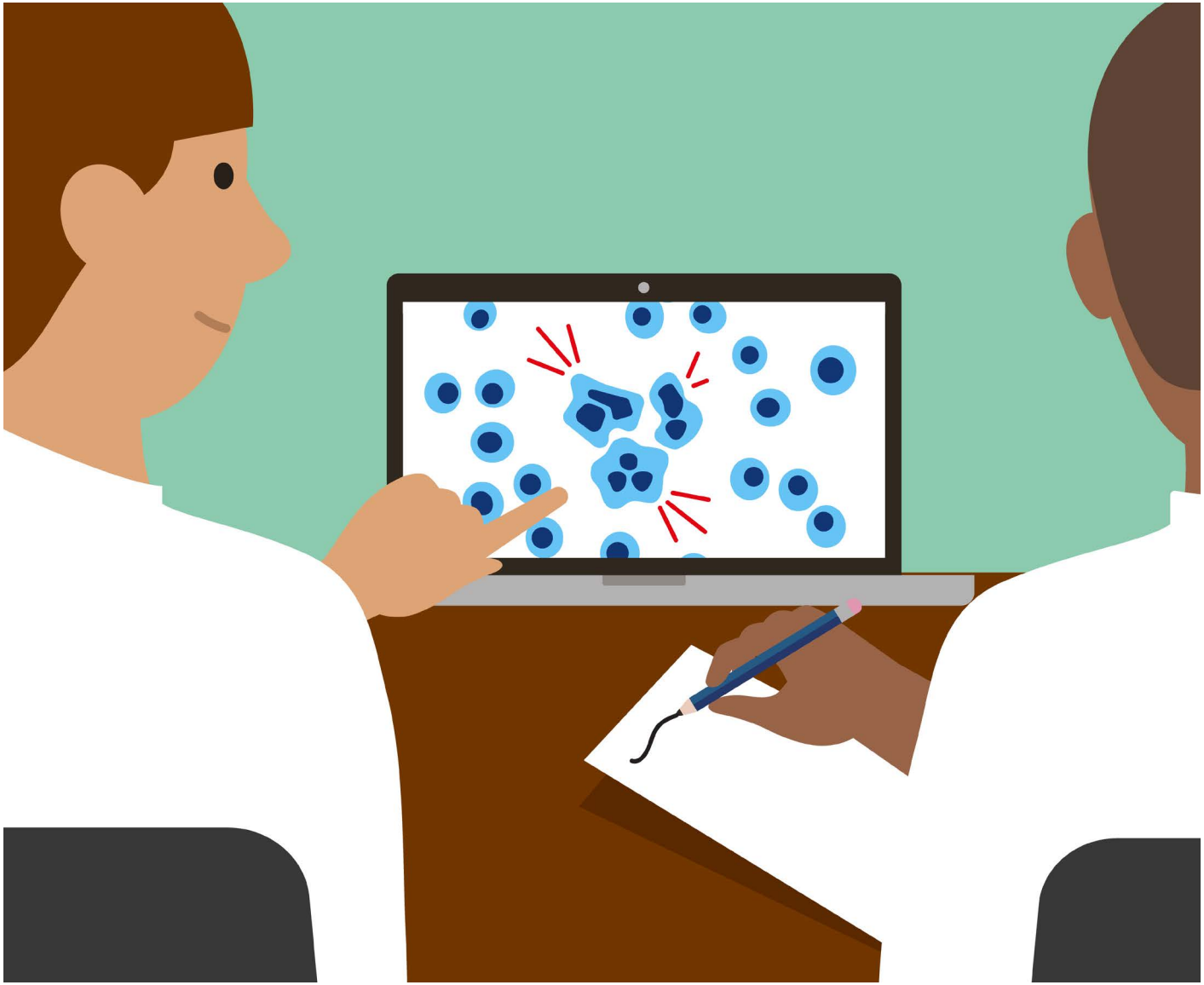
This may include your height, weight, whether you smoke, details of allergies or what medications you're taking. It could also include the results of blood tests, images or any procedures you've had.



Data from patients gives researchers and healthcare professionals vital information that can't be seen with the human eye – even through a microscope. By allowing researchers to understand what past treatments have achieved, patient data can help them to change the future by finding better ways to detect cancer at an earlier stage and develop better, more personalised treatments.

Take radiotherapy, a common type of cancer treatment. Clinicians must pinpoint where to apply radiotherapy, which can be very time consuming if done manually. But technologies like machine learning and computer vision can take data from millions of previous patients and use it to predict where to target the radiotherapy for the best possible outcome.

Imaging from MRI and CT scans also helps to make radiotherapy more targeted, reducing the risk of side effects for patients by attacking the cancer cells while avoiding surrounding healthy cells. When patients share their imaging data after treatment, researchers can use this to develop more targeted radiation therapies and design future treatment plans that are kinder and more effective.



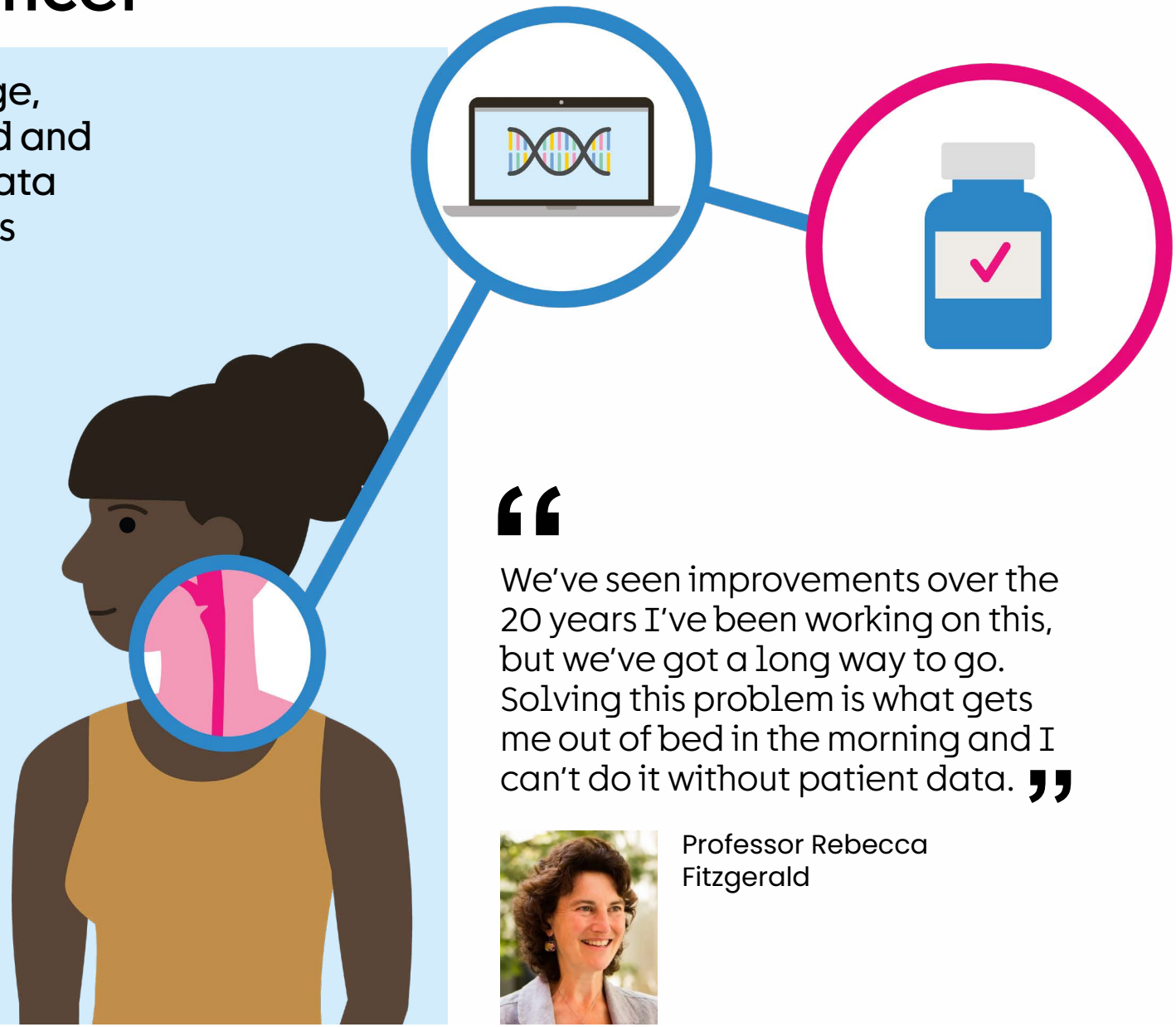
Using patient data to outsmart oesophageal cancer

At the University of Cambridge, Professor Rebecca Fitzgerald and her team are using patient data to try to unravel the mysteries of oesophageal cancer.

Cancer of the oesophagus (the tube that connects the mouth to the stomach) is one of the hardest cancers to treat. To improve survival, we need to try and understand why current treatments aren't working.

Researchers and clinicians can't do this alone, which is why Rebecca and her team are using data from people affected by the disease to help them find answers.

Her team are using samples from patients to produce a complete map of the genetic sequences that contribute to oesophageal cancer. This information will help us understand what makes this cancer different, why some patients respond differently to treatment and the likelihood of people's cancer returning after treatment.



"We've seen improvements over the 20 years I've been working on this, but we've got a long way to go. Solving this problem is what gets me out of bed in the morning and I can't do it without patient data."

Professor Rebecca Fitzgerald

Our Research Data Strategy

As the world's largest charitable funder of cancer research, we aim to unleash the enormous potential of research data and data science to help bring about a world where everybody lives longer, better lives, free from the fear of cancer.

To do this, we've identified four key priorities to guide our current and future goals.

Our four priorities

▶ Priority 1

Build and retain public trust and involvement

Our Research Data Strategy puts people affected by cancer, as well as the wider public, at the heart of our activities.

Our patient and public communities tell us they want to be involved in discussions and decision-making about their data. So, we're focusing on four areas to build, gain and retain public trust that their data is being used responsibly and kept secure.

1 Partnerships:

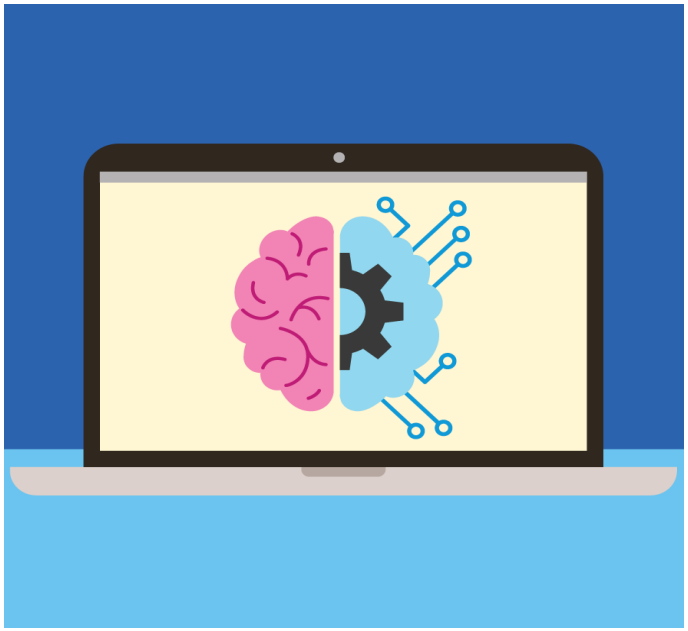
It's important that we collaborate with other patient and public-led organisations who work in this area and share our ambitions. We're proud to be part of the **Public Engagement in Data Research Initiative (PEDRI)** and to work closely with respected patient data organisations such as **use MY data**.

Priority 1 continued

2 Community:

We're bringing together all our patient and public involvement and engagement (PPIE) activities together as part of a Cancer Data Patient and Public Community. This large community of people will help shape the decisions we make and influence the work of our researchers.

3 Resources:



We're creating new resources to explain what we're doing and why. These will help the public understand how patient data and emerging technologies like artificial intelligence (AI) are being used in cancer research.

4 Embedded PPIE:

We'll continue to improve how we work with people affected by cancer and the wider public by trying new ways to involve people across the work we do. We'll also share learnings with researchers, research funders and data controllers so that they can improve their involvement too.

Find out more about how we use patient data

To understand how your data is used for healthcare and research, visit **Understanding Patient Data** and **use MY data**

Priority 2

Support bold new data science projects

We aim to fund ambitious and high-quality data science projects with the potential to save and improve lives.

In particular, we're focused on cancers affecting children and young people.

1 Data4CYP programme

We've awarded up to £250,000 over a two-year period to develop new data-driven insights about children's and young people's (CYP) cancers. The aim is to establish new and improved solutions to common challenges encountered in CYP cancers and create improved data resources for the wider research community to use.

2 Cancer Data Driven Detection (CD3) programme

Led by Professor Antonis Antoniou at the University of Cambridge, this programme aims to transform our understanding of cancer risks, using a wide variety of data to help detect cancer at an earlier stage when treatment is more likely to be successful. A group of patients and the public were brought together to advise on the creation of the programme and discuss some of they key challenges around data and communicating risk.

Priority 3

Connect people doing data-driven research

Better connections within the cancer data science community will help researchers tackle common challenges together and share best practices.

By building a Research Data Community and providing new funding to solve these challenges, we'll help cancer researchers make discoveries more quickly. One area the Research Data Community is focusing on is improving public involvement and diversity.

Our past, present and future collaborations and initiatives include:

1 Conferences and webinars:

Connecting researchers with shared interests and expertise.

Priority 3 continued

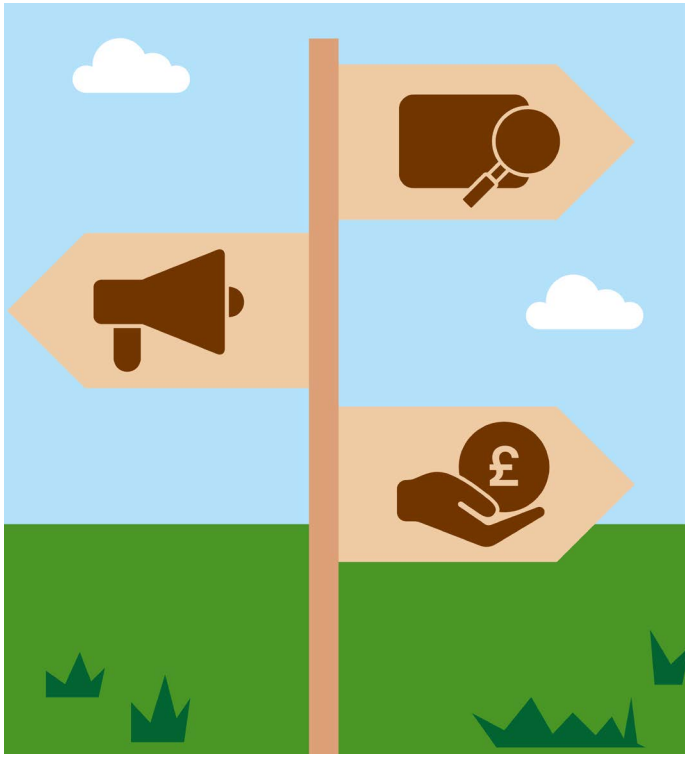
2 Data platforms and tools:

Supporting the research community to access patient data through the Cancer Data Collaborative and sharing new digital tools.

3 Guidance:

Offering advisory services on different themes such as commercial data partnerships and best practices in data science for health equity.

4 Signposting:



Directing the data science community to funding opportunities such as Data Innovation awards, open calls and grant calls.

Priority 4

Help our researchers to quickly adopt new data, technologies and improved practices

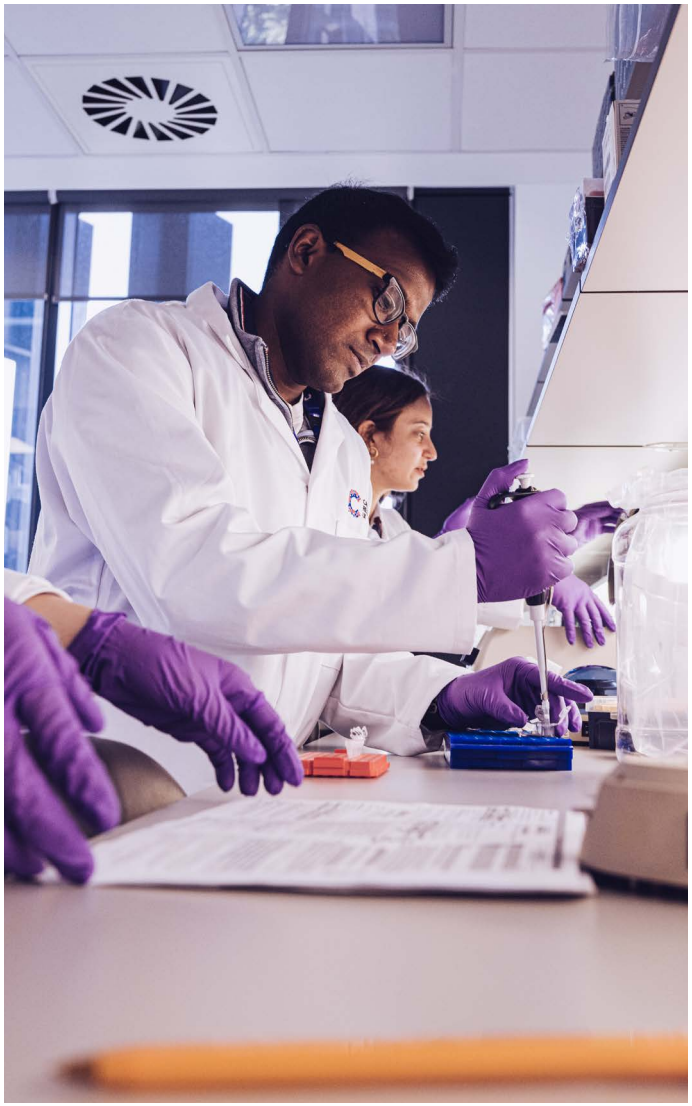
New technologies and tools such as artificial intelligence (AI) mean the potential for data science is rapidly expanding.

We want to prepare our researchers to adopt new innovations as soon as they become available. This includes emerging technologies such as quantum (advanced) computing. Quantum computers are able to solve certain types of problems faster than normal computers by taking advantage of quantum mechanical effects.

This advanced computing uses a lot of energy, so we also want our researchers to consider the environmental impact that data science has by using more environmentally sustainable practices.

Find out more

Kelly Gleason, a Cancer Research UK research nurse at Imperial College London, has co-developed this fantastic resource to help people **learn about data and Artificial Intelligence (AI)**



"Data is central to the research we're doing. The more tumours we study, the more we understand how they differ from each other and how we can use this information to improve outcomes for people with cancer. I'm incredibly grateful to the patients who donate their data and tissue. I couldn't do my work without it."

Professor Crispin Miller, Head of the Computational Biology group at the Cancer Research UK Scotland Institute who are using advanced data science to explore tumour evolution.

Have your say on the use of patient data in our research

We actively encourage the public to get involved in our work and would love to hear your thoughts and perspectives on the use of patient data in research.

In 2025, we launched our Cancer Data Patient and Public Community, which brings together different opportunities to support the work of our Research Data Strategy. To find out how you can participate, email **researchdata@cancer.org.uk**

If you've been personally affected by cancer, you can share your experiences to help shape our work to better meet the needs of patients and their families. Find out about our **Involvement Network**.

We're proud to support and partner with **use MY data** – an independent movement of patients, relatives and carers in the UK focused on the use of patient data to save and improve lives. Although use MY data aren't cancer-specific, you can still access lots of information and opportunities which may be of interest by **becoming a member**.

CANCER RESEARCH UK

Registered with FUNDRAISING REGULATOR

Registered address: 2 Redman Place, London, E20 1JQ. Registered charity in England and Wales (1089464), Scotland (SC041666), the Isle of Man (1103), and Jersey (247). Registered company limited by guarantee in England and Wales (4325234) and registered in the Isle of Man (5713F).

Together we are beating cancer