

# Beating cancer with data



# Introduction

We're the world's largest charitable funder of cancer research. We want to harness 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.

## This document explains:

- what research data and data science are
- how data and AI can help cancer research
- our involvement in this area
- our data science in cancer research strategy

We've created this document for the public and involved patients in its development. For more detail, see our [data science in cancer research strategy](#), launched in 2026.



# Useful terms, explained

**AI:** Artificial intelligence.

AI involves creating computer systems to carry out tasks that usually need human brainpower. It's invaluable to researchers as it allows large amounts of data to be analysed more quickly, generates novel discoveries and uncovers new avenues of research. Because AI can influence important research decisions, it must be accurate, fair and keep patient data secure.

**Anonymisation:** The process of removing personal information (such as name, date of birth or address) so they can't be linked or traced to an individual. Sometimes data is only partly anonymised (known as pseudonymisation). This means identifying details are replaced or transformed, but the data could still be linked back to an individual using additional information that is held separately. For example, names and other identifiers might be replaced with codes so researchers can't directly identify people, but approved teams could reconnect the data to a patient if needed for important clinical reasons.

**Big data:** Extremely large or complex data that can't be handled manually and requires special methods to be processed and analysed by computers.

**Computer vision:** An area of AI that deals with how computers can be made to gain advanced understanding from digital images or videos. The computer can then make recommendations (reviewed by humans) or take actions when it sees patterns or problems.

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

**Data science:** The process of using data to learn new things, answer questions and solve problems. It involves collecting, organising and analysing data, then presenting the findings in ways that are easier to understand and use. Data scientists may use a range of tools and technologies, including AI.

**Machine learning:** An area of AI which involves computer programmes learning for themselves from large amounts of data. In medicine, this is valuable for tasks like recognising tumours in scans that might otherwise be difficult or impossible for humans to detect.

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

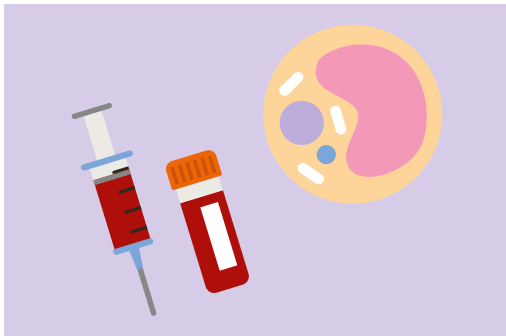
**Patient data:** Medical data about patients. This includes data about things like appointments, tests, diagnoses and treatments. This data is usually anonymised for use in research so as not to reveal the patient's identity and is protected by strict security controls.



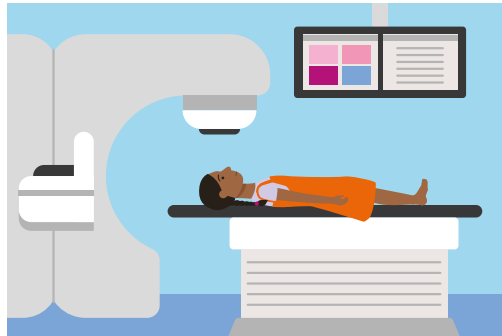
# How do research data and data science work together?

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

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



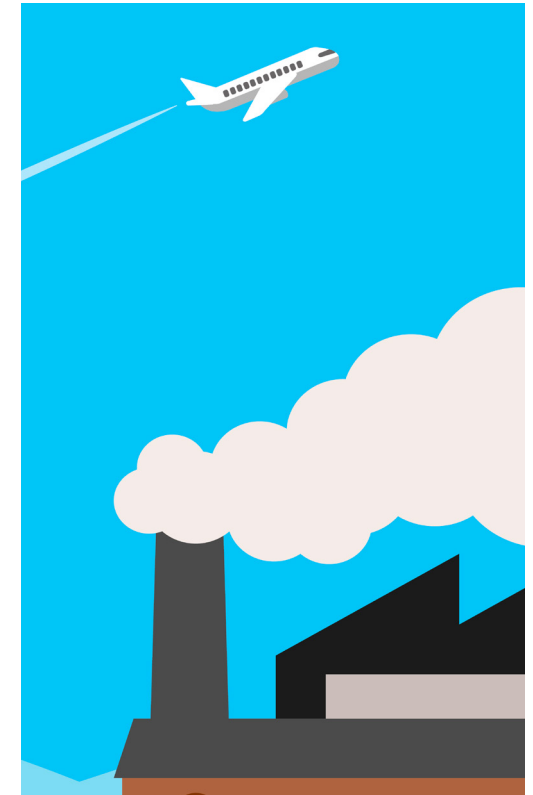
Samples from blood, cells or tissue



Imaging from X-ray, MRI and CT scans



Findings from clinical trials



Environmental data, such as air pollution levels



Data from health and fitness devices



NHS electronic health records



Data from public services such as schools or census responses

**Data science in cancer research is about using lots of research data to better understand the disease, predict what might happen next and improve the speed and accuracy of research.**

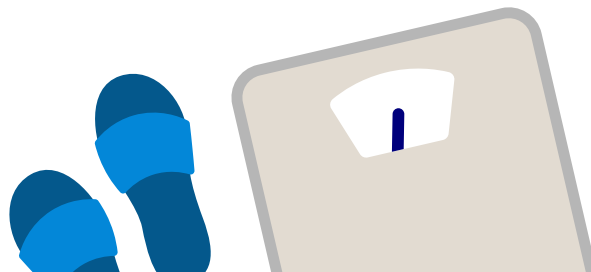
# How can patient data help cancer research?

When you have a GP appointment, attend hospital or take part in research, information may be collected about your health and lifestyle.

## This can include:

- height and weight
- whether you smoke
- allergies and medications
- results from blood tests
- scans and imaging
- details of treatments

Patient data can reveal patterns that can't be seen by the human eye, even through a microscope.



## By understanding what past treatments have achieved, researchers can use this data to:

- detect cancer earlier
- improve diagnosis
- develop more personalised treatments

Take radiotherapy, a common cancer treatment. Clinicians need to pinpoint exactly where to target it. Using data from previous patients, technologies like machine learning and computer vision can help predict the best place to aim treatment more quickly.

Data from MRI and CT scans is helping make radiotherapy more precise, reducing side effects by targeting cancer cells while avoiding healthy tissue and supporting the development of more effective, kinder treatments.



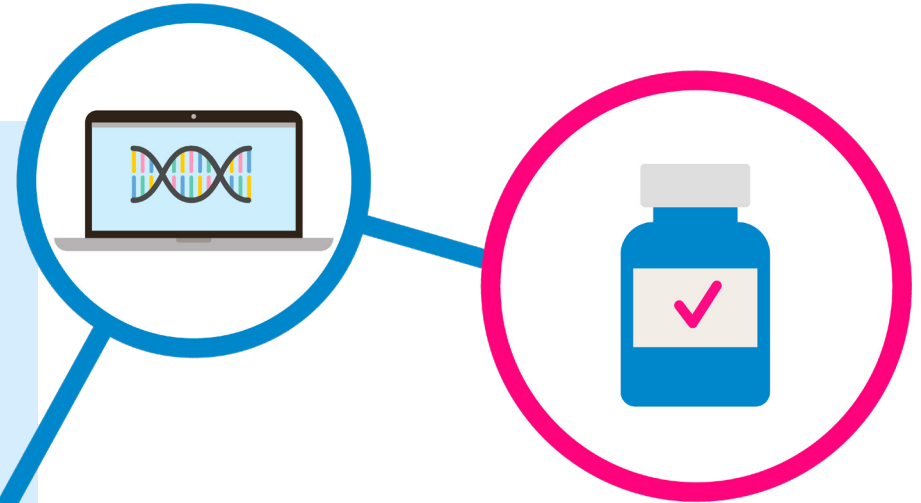
# Using patient data to outsmart oesophageal cancer

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 understand why current treatments aren't working as well as they could.

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.

They 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 behave differently, why some patients respond differently to treatment and the likelihood of 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.”



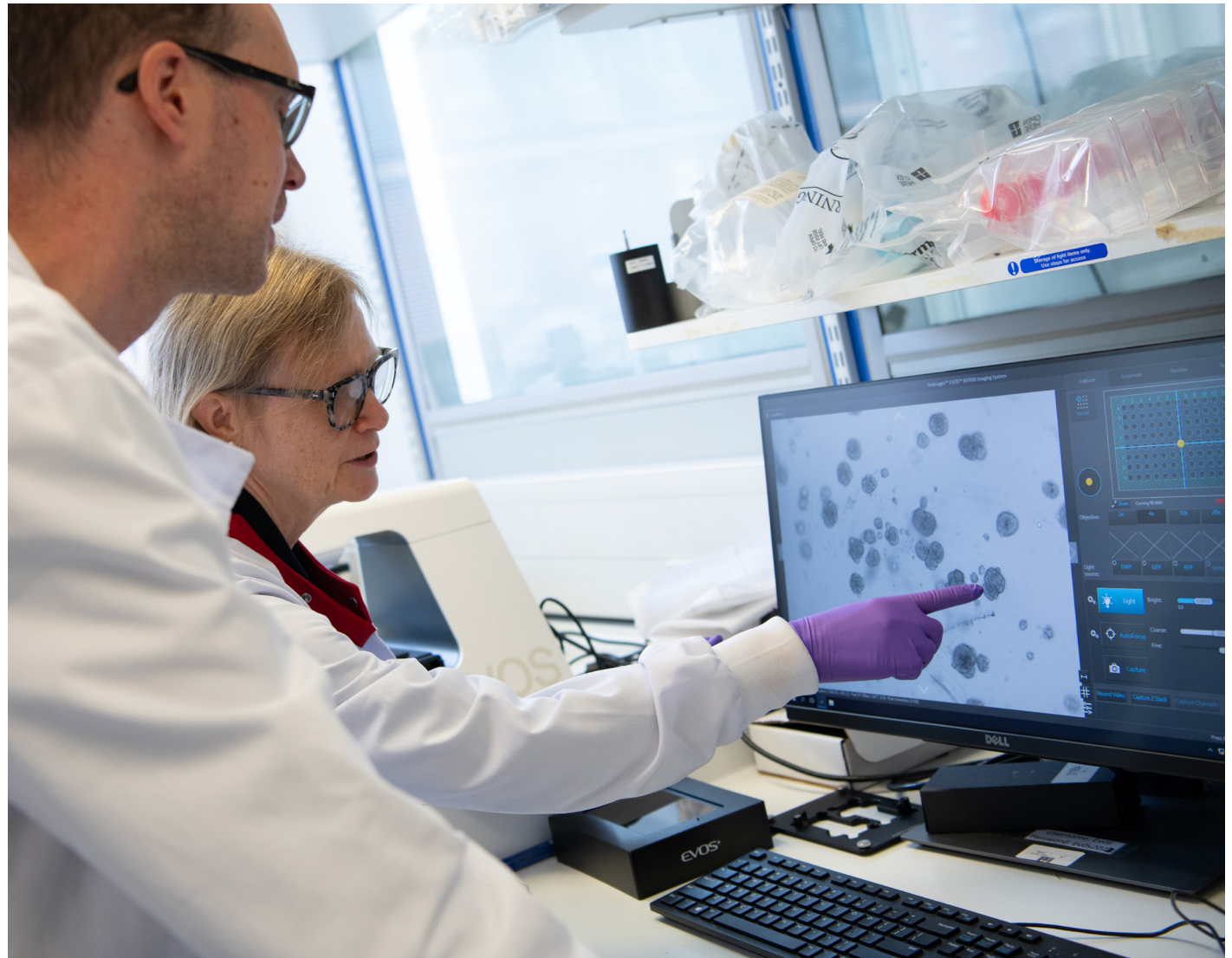
Professor Rebecca Fitzgerald

# How can AI help cancer research?

If AI is trained on high-quality data and research is carried out in a trustworthy way, there are opportunities for it to play a greater role in cancer research, from discovery science in the lab through to clinical research.

Alongside other types of data science, AI has already shown its potential to help researchers identify new avenues for discovery. Examples from cancer research projects show that AI is beginning to find cancer earlier, read scans and lab results more accurately, spot genetic changes and help provide evidence to design better treatments.

By speeding up some of the research process and completing tasks that are hard or even impossible for humans to do, we can advance discoveries in cancer research more quickly than before.



# Ethical and trustworthy data science and AI

In a rapidly changing world where AI is featuring more in day-to-day life, we need to ensure the ethical implications of data science and AI are understood and addressed in cancer research.

We have identified **ethics** as a key enabler for our Data Science in Cancer Research Strategy, alongside working in partnership with other organisations and influencing policy. The ethics enabler has four key areas: patient and public involvement and engagement (PPIE), reducing environmental impact, bias mitigation and responsible and lawful use of data.

## PPIE



An essential part of making data-driven research trustworthy is making it understandable and focused on the needs and concerns of people affected by cancer and the wider public. We're working with cancer researchers to improve how they involve the public in the design, development, evaluation, implementation and dissemination of research, as well as how they talk about their research to the public in easier ways to understand.

## Reducing environmental impact



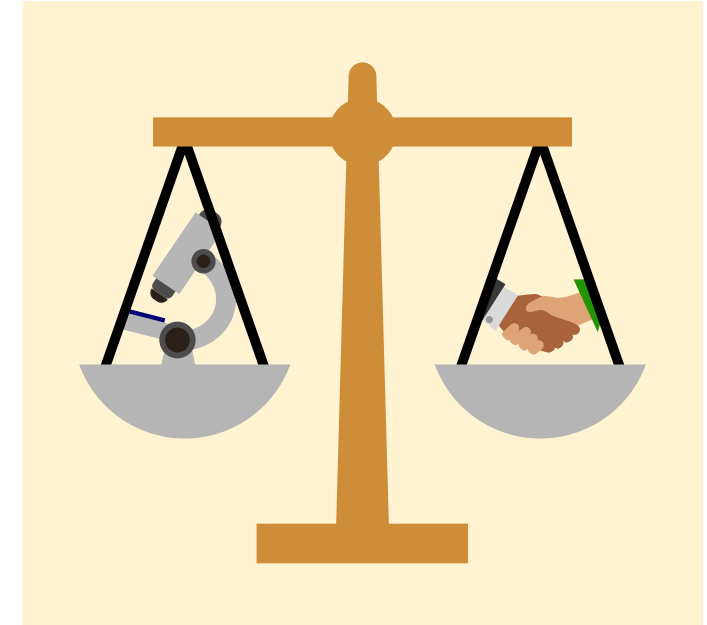
We're working with our researchers to find ways to reduce the environmental impact of data science and AI and take steps to minimise it. This includes making AI models more efficient, reducing unnecessary computer usage and enabling research data to be reused where appropriate so that valuable data can support multiple studies.

## Bias mitigation



A lack of diversity in patient data can make it less representative of the population in the UK and make the findings of research less effective for some people. We're working with researchers, as well as the public, to identify where this might be a problem and address it in cancer research. This also includes making sure AI tools are designed and tested to avoid unfair or inaccurate results for different groups of people.

## Responsible use



We support researchers to navigate legal, regulatory and ethical challenges that can slow or complicate research. In many cases, it isn't the law itself, but the complexity of applying laws across different countries to new technologies and ways of doing science. By supporting researchers in this way, we hope to achieve the right balance of protecting ethical safeguards and preventing research collaborations from being delayed or stopped.

# Theme 1: Funding data-driven research at scale

Alongside our existing funding for data-driven cancer research through our core funding schemes and initiatives such as **Cancer Grand Challenges**, we'll continue to invest in new opportunities that keep pace with advances in data science and AI.

We're bringing together world-class expertise in cancer and data science to establish a new AI Research Alliance to create new teams of experts from a variety of disciplines, sectors and backgrounds, who will use the new technologies and techniques of data science and AI to solve the greatest challenges in cancer research. Our AI Research Alliance is expected to lead the way in demonstrating ethical approaches so that the research isn't just high quality, but also trustworthy.



# Theme 2: Unlocking the value of data

We invest approximately £400m in research a year, but much of the data produced through this research isn't reused after the initial project. Through this theme, we'll address the issues of making the data easier to find, more complete and establish better, more trustworthy processes for making it accessible.

By creating a Cancer Research UK Data Hub, we'll help cancer research datasets become more reusable for other research projects and maximise the value that they provide for research.



# Theme 3:

## Building a future-ready workforce

We want the cancer research workforce to be able to make the most of the new technologies and possibilities that data science and AI provide. To do this, we need to invest in developing our workforce to keep pace with the rapidly changing technology landscape and attract more talented people from other sectors, for example, mathematicians and computer scientists.

Our Data Science Skills Catalyst will include training, secondments and opportunities for researchers to collaborate and build skills together. We'll also increase funding to support future leadership in AI for cancer research through a new group of AI-focused clinical future leaders.



# 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 on the use of patient data in research.

In 2025, we launched our Cancer Data Patient and Public Community, which brings together opportunities to support the work of our Research Data Strategy. To find out how to take part, email [researchdata@cancer.org.uk](mailto: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 by [becoming a member](#).





---

Together we are  
beating cancer



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).