# Your data, your control

**Wednesday, 02 May 2018**
**10:00 to 16:00**

**No.11 Cavendish Square**
**London**
**W1G 0AN**

# Summary of the day

***"There is a responsibility for those holding the data to do something with it, because nobody wants to give their data, or their samples, and then nothing happen with it."***

**Patient Advocate & workshop delegate**

**Contents**

**Chair's welcome & overview of the day**

**Mike Birtwistle**
**Founding Partner, Incisive Health**

Mike opened the workshop, welcoming everybody. It was clear that many had attended previous use MY data events before and Mike welcomed new and returning delegates.

Mike is the founding partner of Incisive Health, a health policy consultancy that works with people who supply data (patients, charities) and also people who want to use data within the NHS (researchers, commissioners, providers of NHS services, suppliers to the NHS).

Our funders were thanked; the Leeds Institute for Data Analytics and Health Data Research UK and the speakers who have given their time for free. Most of all, thank you to the delegates.

The world of data is so topical at the moment and always in the news, for good or bad. As well as media coverage, there are also some big developments in data policy relating to the NHS coming too. use MY data is the sum of its parts: you are use MY data and decide the agenda.

**Session 1** **Patient data information part 1 – what we know**

**Chris Carrigan**
**Expert Data Adviser, use MY data**

**General Data Protection Regulation & the National Data Opt-out**

25 May is a key date for data protection rights as it sees the implementation of the General Data Protection Regulation (GDPR) for the UK and the National Data Opt-out for England.

There are lots of web resources for the General Data Protection Regulation, but two sensible places to start are the websites of the Information Commissioners Office and NHS Digital, as they both have clear and accessible information.

In advance of the new National Data Opt-out a number of roadshows are underway, organised by NHS Digital, with details available on their website, at https://digital.nhs.uk/services/national-data-opt-out-programme

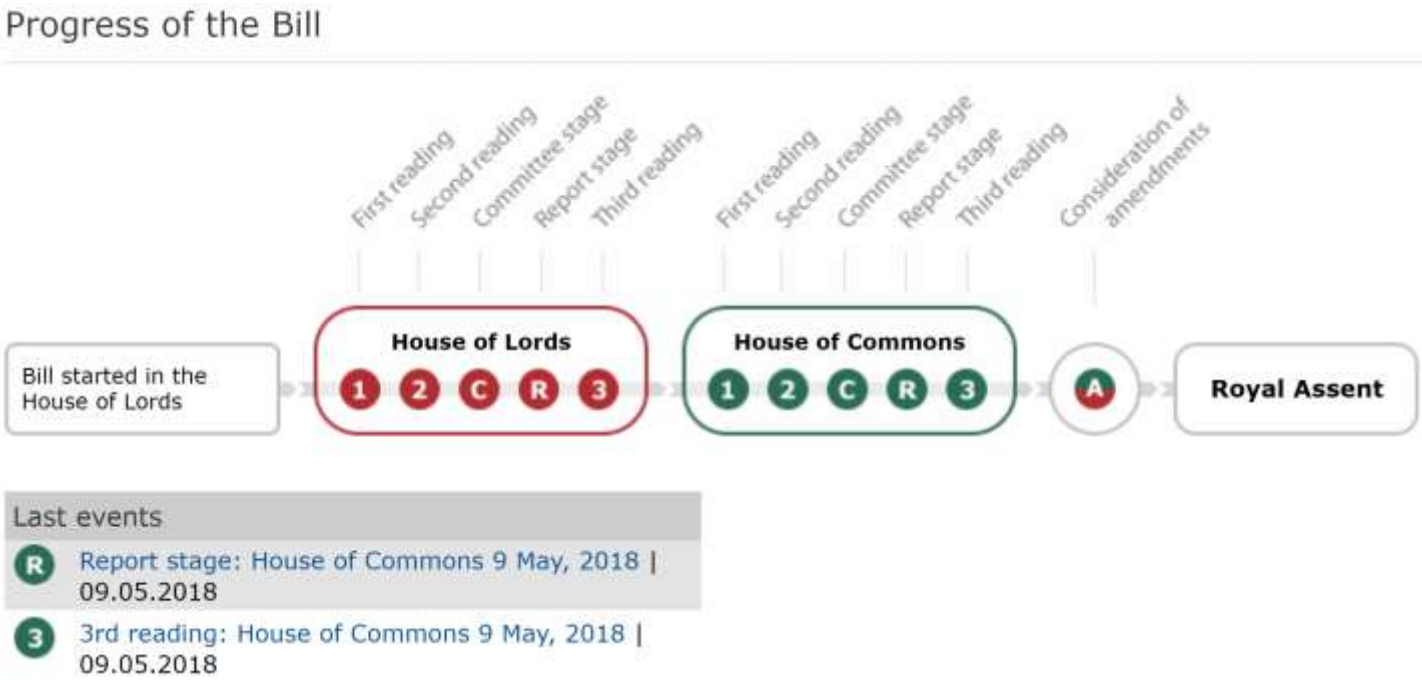
**Data Protection Bill**

The Bill will enact the General Data Protection Regulation into law for the UK.

There was a short consultation called 'Have your say' which encouraged responses on the consultation document. However, this time-period was quite short and the documents very long and technical.

The Bill, which will repeal the Data Protection Act 1998, sets out how the UK would apply the derogations available under the General Data Protection Regulation and updates the laws governing the processing of personal data by the intelligence services. It aims to ensure that the UK would be able to freely exchange data with the European Union post-Brexit.

A brief timeline of the current status of the Bill (as at 2 May) is:




**Association of British Pharmaceutical Industries**

At the use MY data workshop in October 2017 'Commercial access to patient data', the Association of British Pharmaceutical Industries (ABPI) and AstraZeneca jointly presented the session 'How does the pharmaceutical industry use patient data?'

The discussion around the presentations highlighted some communication problems with the pharmaceutical industry, around informing patients about the outcomes of projects where patient data had been used. Patients asked why they are refused entry to some drug company events and meetings. In response it was explained that some meetings are not for the 'public', which is why entry may have been refused. It is not permissible to 'promote' prescription only medicines to the public, as defined within the ABPI Code of Practice. A significant point was made by patients in follow-up: that as the patients who provide the data, they should not be seen as 'the public'.

At the 'Patients First: Pioneering Partnerships' conference, hosted by the Association of Medical Research Charities and the ABPI, there was a strongly supported theme about the need to involve patients at the start of research projects. It was recognised that patients are experts in their field, and that there were clear frustrations from patients at the barriers to engagement with pharma.

At the end of the Conference, Mike Thompson, the CEO of the ABPI said he was taking away two specific priorities.

1) The need for the way the industry does research to change
2) The need to look at the Code of Practice - the ABPI is in the process of recruiting for a role that will specifically work on what the Code can do, rather than can't do.

As a follow up to discussions with the ABPI, use MY data was invited to speak at ABPI's Annual Conference in April 2018, to highlight the benefits of patient inclusion and involvement.

## Data sharing between NHS Digital & the Home Office

There is a full chronology of the developments around the Memorandum of Understanding (MoU) between NHS Digital and the Home Office, on use MY data's website.

In response to the letter from the Department of Health and Social Care to the Health Select Committee (23 February) and as part of the campaign to raise awareness about the damage the MoU was doing to public confidence in data sharing, use MY data members were surveyed about whether they agreed with several of the statements contained within the letter.

A report was produced to summarise responses to the online poll, and this was sent to the follow-up consultation run by Public Health England.

A brief overview of the findings:

- There was unanimous agreement amongst members (either strongly agree or agree) that patient confidentiality should be respected by ensuring that medical information about a person is properly protected.
- Members largely agreed that sharing data was important, but that medical data should be properly protected.
- Members felt much less certain that patients would have a reasonable expectation that data should have been shared in this instance.
- Contrary to views stated by the Home Office and the Department for Health and Social Care, it was the strong view of members that sharing of data in the investigation of criminal offences had significant implications for public confidence.

*Post-meeting note: on the day following the workshop, the Government altered the terms of the MoU with immediate effect. It will now only be used to trace people who are being considered for deportation due to committing a serious crime.*

## Public Health England lung cancer data release to William E Wecker Associates

In January 2018 an article appeared in The Telegraph, raising concerns about the release of data from Public Health England to William E Wecker Associates, a company with links to the tobacco industry.

use MY data members submitted concerns and questions, which were collated and submitted to Public Health England for a response. A full summary of the details around the release, together with the responses to the questions raised by use MY data members, is available at the use MY data website at http://www.usemydata.org/newsx.shtml

A follow-on article on the data release appeared in the BMJ, and Paul Affleck and Chris Carrigan submitted an e-response, https://www.bmj.com/content/360/bmj.k293/rr

## UPD, Carnegie Trust & Involve – Data for Public Benefit

A new report has been published, 'Data for public benefit: balancing the risks and benefits of data sharing'. The report is based on the findings from a series of six workshops in different local authority areas across England. The workshops brought together over 120 professionals from the public and voluntary sectors (working in the fields of housing, criminal justice, health, social care and welfare) to explore how they

understand, define and value, the public benefits which can be derived from the use of personal data.

The research highlights that there are big differences in how public services define public benefit and risk with regard to data sharing. The report also includes a framework to help organisations better evaluate these benefits and risks. This framework will help professionals weigh up the purpose of sharing data against the potential for harm and help public service providers have conversations with the public about data sharing.

The report is available at https://understandingpatientdata.org.uk/news/data-public-benefit

**Local Health and Care Records Exemplar bids**

A small number of regional health and care collaborative communities across England have been invited to bid for national investment in shared health and care records.

The regional collaborations will compete to become one of five new Local Health and Care Record Exemplars (LHCRE), each potentially receiving up to £7.5m in national investment, which bidders will be expected to match fund.

Each regional LHCRE will build on existing local work on shared records to further develop joined up regional health and care information reference sites, focused on improving direct patient care.

use MY data was invited to review the proposals, but due to the timescale and size of the task involved, were unable to assist.

**Questions and discussion**

**Q** I'm interested in the comment about the lung cancer data release – surely that was a bad thing? They want to sell cigarettes and we want to prevent/cure cancer. Surely a no-brainer?

**A** The data which was published was not identifiable; this type of tabular data is exactly the sort of thing that should be on the internet and freely available. Once data is published, you cannot prevent data being used by the company, or indeed anyone else. A question remains about overall transparency though and, specifically, why didn't people know about it?

**Q** For GDPR, is there some kind of on-line training we can do, to inform us? Useful for us to know how people (professionals) are being trained.

**A** Geoff Schrecker responded to this question. The Opt-out and GDPR do interact a little. The point we are trying to get across to people is the difference between consent for processing data under GDPR (as a health professional that will not be your basis for processing data) and consent for sharing data under the common law duty of confidence, which is a different standard of consent and a different process.

In terms of sharing data under the common law duty of confidence and consent, nothing has changed. In terms of processing data for health and care, you are actually using a different lawful basis under GDPR. You are not using consent as your basis for processing that information.

**Session 2 Primary care preparations for the National Data Opt-out**

**Dr Geoff Schrecker**
**Clinical Champion, Patient Data Choices**
**Royal College of General Practitioners**

In this session we heard from the Royal College of General Practitioners (RCGP) about the Patient Data Choices project and what is being done to support general practice teams to uphold the National Data Opt-out when it is implemented.

Geoff described the position when he started general practice in 1991, of all patient notes being on paper cards, which made the concept of wide-scale data sharing impossible. The practice Geoff trained in had just installed a computer system and, as was common at the time, the company that provided the computer system then harvested the prescribing data for use by the pharmaceutical companies. Being an advocate for using the data held in the practice system, Geoff conducted an audit on asthma and asthma prescribing. The results showed that that the data was meaningless, due to variation amongst the doctors in terms of whether they actually used a computer or not, and if they did what they recorded. Geoff's interest in IT and data in general practice continued to develop and he now works full time on the patient data side of primary care, including leading the RCGP's preparations for the new National Data Opt-out.

The Opt-out is really about how we access the huge amount of data that is held in the health service, not just primary care. It is not about using data to care for you as a patient, e.g. your summary care record / local record sharing system so that A&E can see your primary care information. Rather, it is about using data for improving care in general for patients; improving the systems that we use to deliver that care and improving efficiency within the system. The term 'research and planning purposes' doesn't really cover all that the data can achieve.

So, why the fuss? One factor is the previous attempt by the NHS to implement the programme called Care.data. The Care.data programme (now abandoned) highlighted that fact that we all have different views of what we know about our data and what we're willing to share. As a result of the concerns raised about the programme, patients were offered the opportunity to make either a Type 1 or Type 2 objection to their data being used.

- Type 1 objection – This stopped your GP sharing your identifiable patient information with anyone, including NHS Digital (previously the Health and Social Care Information Centre). Your identifiable data would not be used for any purpose other than your individual care.

- Type 2 objection – Allowed information about you to be shared by your practice with NHS Digital. But this data would not be onwardly shared with other organisations. This applied to nearly all uses of confidential patient information, with a few exceptions.


One factor that came about as a result of Care.data was an appreciation that people needed to be able to express an opinion about whether their data was available for use. It was also clear that there was a lack of understanding for people, about how their data was used. Geoff highlighted the great work of Understanding Patient Data in communicating the benefits of sharing data. It's really important for people to feel that there is some benefit from the sharing of their data.

Geoff emphasised that in the balance of making a decision about how you want your data to be used, trust is essential. The National Data Guardian has emphasised that if people are going to share their data, they have to trust "us" (the whole health research NHS community) with their data. Part of reinforcing trust is to say to people "if you don't trust us, you have the choice that we won't use your data".

Geoff highlighted other work that had taken place after Care.data, in particular the report of the National Data Guardian (commissioned by the Department of Health and Social Care) on issues of consent and security of data and how patients might opt-out if they do not want their data to be used. This, and the work of Understanding Patient Data, has emphasised the need for clear ways of describing different types and levels of data. The term 'confidential patient information' is now being used to describe identifiable patient data.

The National Data Guardian's report, a 'Review of Data Security, Consent and Opt-Outs', was published in July 2016. There was a delay before the Government responded in July 2017, publishing 'Your Data: Better Security, Better Choice, Better Care'. The Government's response accepted all of the National Data Guardian's recommendations in full.

Geoff contrasted the learning from Care.data with the approach being taken by the RCGP for the National Data Opt-out; a big part of what went wrong with Care.data was a failure to communicate what was being done. For the new Opt-out, NHS Digital is taking the lead role in wider communication to patients and the public, working with other agencies such as NHS England and the Department of Health and Social Care.



Geoff was clear that the RCGP's role is to work with primary care teams, to help them communicate more directly to patients. Information packs for practices, including a toolkit, eLearning resources and videos, will be aimed at the wider practice roles, including receptionists and practice managers, alongside GPs. Each practice needs to be able to tell a patient how they make their choice and how the choice will be implemented.

In terms of the timescales, Geoff highlighted the first notable date is 25 May 2018, when the new Opt-out will be introduced, to coincide with the implementation of the wider General Data Protection Regulation (GDPR). At that point the public will be able to directly set their preferences. This will be via either a dedicated website (which is being developed by NHS Digital) or the telephone (to an NHS call centre), or a paper form (sent to the individual after contact with the website or the call centre).

If you already have a Type 2 objection it will be automatically carried over. An important message to practices is that they have to know that from 25 May onwards they can no longer use the Type 1 or 2 codes. As currently is the case, the Opt-out will be applied to any data flows from NHS Digital outwards, and this will start on 25 May. The Opt-out will apply eventually apply across all of health and social care but, reflecting the difficulty in implementing this consistently and robustly across many organisations, the deadline for that is March 2020.

Geoff also highlighted that under GDPR the NHS has to be transparent about how we process patient data and patients must know how their data is being used.

Moving on to the details of the National Data Opt-out, Geoff highlighted:

- It covers confidential patient information that is being collected under certain legal bases.
- It does not cover data which is being collected with your explicit consent.
- The Confidentiality Advisory Group (CAG) of the Health Research Authority can already recommend use of the statutory power contained in the NHS Health and Social Care Act 2012, to ensure that NHS patient identifiable information needed to support essential NHS activity can be used without the consent of patients. This Section 251 approval sets aside the common law duty of confidentiality, so that confidential patient information can be transferred to an applicant without the discloser being in breach of the common law duty of confidentiality. However, if you do not want your confidential patient information to be used in this way you can opt-out from this.
- The Opt-out does not apply when there is a mandatory legal basis for data to be collected. Under Section 259 of the Health and Social Care Act 2012, NHS England or the Secretary of State can require NHS Digital to collect data from health organisations. This type of mandated data collection is not covered by the Opt-out.
- As another example, Geoff highlighted the National Diabetes Audit. This is collected by NHS Digital under a direction from NHS England, under the legal basis of Section 254 of the Health and Social Care Act 2012. You do not need to be asked for your consent, however any part of the service collecting your data has to do their best to make people with diabetes aware that their service is participating in the Audit and that they can opt out if they wish.

In summary, Geoff highlighted the key points:

- It is important for patients to understand the difference between 'care' and 'research and planning' when deciding to set a National Data Opt-out.
- Although the Opt-out is policy, not law, primary care has to understand how it works and the associated legalities.
- The RCGP resource is designed for all primary care staff, with specific resources for receptionists, practice managers, clinicians and the practice Caldicott Guardian.

**Questions and discussion**

**Q** How can you be sure a GP practice will not have a sense of bias, one way or another?

**A** All practices/staff are human and there will be bias. The thing is to be aware of the bias. By getting resources out there we hope there will be balance. The resource will be publicly available.

**Q** Will there an audit of practices?

**A** I'm not aware of intention to audit. This was undertaken for Type 1 & 2 objections and it would be a sensible thing to do.

Mike – use MY data might want to call for audit of primary care, to work out the impact of the Opt-out.

**Q** Care.data had a lot of GP practices opting out their entire patient community; will that roll over into the new system or do patients have to do it individually?

**A** I don't think there will be any practices left where patients have Type 2 objections on their record. If there were those would all roll over into an Opt-out. The big difference with the Opt-out is that it is set by individuals, not the practice. A practice will no longer be able to do a bulk up-load of opt outs.

Mike – Are there any plans in place to go back to those who made Type 2 objections?

Everyone with a Type 2 objection in place will receive a letter from NHS Digital explaining the changes, the conversion to the Opt-out and how to opt back in, should they wish to.

**Q** Does this apply to other parts of the UK? Do GPs have specific targets? And what steps are in place to look at homogeneity across the NHS?

**A** The Opt-out will apply in England only and not the other UK nations, though it affects data flowing between borders.

Geoff was not aware of any plans to monitor the consistency and numbers of opt-outs across the country at the present time.

**Q** NHS Digital does have limited information on those who have Type 1 and Type 2 objections.

1. Do we have any information/statistics on how often identifiable data in used in research/planning?

2. We know what the Opt-out will cover, but what it will not cover is quite a large list; audits and disease registries will not be included. Is there a full list somewhere?

**A** 1. On its website, the Confidentiality Advisory Group publishes in entirety all minutes, discussions and approvals. You can download information on everything that the Group has recommended Section 251 approvals for (recommendations to the Health Research Authority for research and the Secretary of State for non-research).

2. On the issue of things that aren't included; this is a fluid list. The National Data Guardian was asked to produce one simple solution to a complex situation. For Section 259 data provision notices; you cannot expect a patient or even a practice to understand the difference between a national audit that's being conducted under Section 251 and one that is under Section 259. This is a problem. Going back to the National Diabetes Audit; this was under Section 251 until 2107 and so would have been covered by the Opt-out. However, the Audit is now covered by a data provision notice under Section 259 and so is not covered by the Opt-out.

It is not easy to find one simple solution to such a complex situation.

**Q** How consistent are the records held by GPs across the country?

**A** The way in which GPs construct their digital records is highly variable. There are lots of different reasons for this, including the IT system, or the personal preferences of the GP. Some are obsessive coders, and some are more minimalist. So we already have an enormous amount of heterogeneity in studies. Managing free text information in terms of governance is difficult. One of the key things is that people using data understand its fallibility.

**Q** When people made Type 2 objections they did not necessarily understand what they were doing. When the letters are sent, will they include clear information so that people understand?

**A** Geoff had not seen a copy of the letter and, up to couple of days ago, the text had not yet been approved. He did know that the letter will direct people to the websites of NHS Digital and Understanding Patient Data. There is a real will to get people to understand the benefits, but Geoff recognised it is very difficult to get this across to all levels of the population. It is asking a lot of primary care colleagues to take this on, alongside current pressures. Understanding Patient Data has produced great animations which are helping to get the message across.

**Q** I'm part of the service user network of the Healthcare Quality Improvement Partnership (HQIP):

1. With regard to the national audits; is the Opt-out in relation to data going in or going out of the audits?

2. I understood from Public Health England that they are having considerable difficulties getting hold of End of Life Care data; will that resolve after 25 May?

**A** From 25 May the Opt-out applies to confidential patient information going *out* of NHS Digital. This does not apply to aggregated data (data this is anonymised in line with the Information Commissioner's Office standards).

For End of Life Care; this will depend on the lawful basis on which they collect data (historic or not). If they go through the Confidentiality Advisory Group and Section 251 then the Opt-out will apply. If it comes as a direction from NHS England or the Secretary of State through a data provision notice under S259 of the Health and Social Care Act, then the Opt-out will not apply.

If it comes under Section 251 a practice can decide not to take part and does not have to submit data. Under Section 259 a practice has to comply; this is a mandatory requirement that the practice cannot decline.

Mike suggested that a practical action for use MY data could be to ensure that all the national clinical audits have a plain English explanation of how they collect their data and on what basis.

**Q** Computer scientists are seeking to improve the quality of data on an aggregate level by taking into consideration how it is actually gathered in practices. Do you think that is feasible?

**A** Yes, the quality can be improved but as soon as you start using it you have to understand the mechanism used to provide the quality, and its limitations.

**Q** How will you raise the awareness of the general public? How will this include Patient Participation Groups, Healthwatch, people with learning difficulties? Will information be multi-faceted when it comes out, in terms of talking to people? If it will be written how will it be easy to read?

**A** We are a very small cog in the NHS Digital communications machine. NHS Digital has been working intensively with voluntary groups producing resources e.g. British Sign Language, groups with learning difficulties, hard to reach groups. The RCGP is a small part of the overall solution.

**Q** Any thoughts to how it will be on-going once the initial work has been done?

**A** This is currently subject to debate. The RCGP is working on the final draft of the e-learning resources and is waiting for policy decisions. The idea is to establish this type of resource as part of every hospital system, but it will be a challenge.

**Q** If I made an objection under Care.data and I didn't know if it was a Type 1 or Type 2, will I automatically be opted out?

**A** If it was a Type 2 objection you will receive a letter which will give you options.

If it was a Type 1 objection, then at the moment this is unclear, and it would be best to ask your GP which type of objection is registered for you. All Type 1 objections will remain until March 2020, when there will be another consultation.

Mike summarised the session, noting that he saw the National Data Opt-out as being a process rather than an event. The timelines are tight and use MY data will probably want to return to this issue.

It was suggested that some areas which use MY data could usefully seek to clarify would be:

- Auditing the impact on general practice behaviour to get a common and consistent approach
- Building public confidence in national and clinical audits

Mike noted that there were still some areas of big uncertainty, in particular people not knowing what they have previously objected to.

**Session 3 Machine learning and patient inclusion**

**Dr Antonio Criminisi**
**Principal Researcher**
**InnerEye, Microsoft Research**

**Dr Raj Jena**
**Academic Consultant Clinical Oncologist**
**InnerEye, Microsoft Research**

In this session we heard from Microsoft Research about Project InnerEye, how they use patient data and how they engage with patients. The session examined the practical applications of using big datasets to improve healthcare.

Raj introduced the session, saying it was a privilege for both he and Antonio to be part of the workshop, and to be able to talk about how they are using patient data and the progress they are making in the use of artificial intelligence techniques for medical image analysis.

There has been much discussion around the use of artificial intelligence in healthcare. The part that Project InnerEye particularly encounter is around the use of anonymized medical image analysis. Their interests/background is in computer vision; training computers to recognize objects in image data and that has a natural translation to medical image data.


There is a battle between hype and truth. Artificial Intelligence (AI) has a lot of potential but there are a lot of unanswered questions and there is a lot of work still to be done. The focus for Project InnerEye is on exploring one particular area of that, with a very pragmatic goal. The project started out focussing on radiation therapy.

Raj gave an overview of the session's content, outlining that it would cover why radiation therapy was chosen, the problems faced, what AI solutions are available now and how technologies might improve day-to-day working practice of clinicians. As well as exploring the current problems and how AI might help in future areas; can AI go beyond what's possible today?

Radiation therapy is a very important anti-cancer treatment and is used in the treatment of about 40% of patients who are cured of cancer. It is second only to surgery in terms of its effect/cancer cure. We know that overall 1 in 2 of us will be diagnosed with cancer and, of that group of people, about 1 in 2 patients will need radiotherapy, at some point in their cancer journey.


In 2015 (the last time there was a mature dataset across Europe) there were approximately 400,000 new cancer patients. There were a lot of discussions at a European level about the amount of radiotherapy needed in a timely fashion, up to 2025 (ten years on from the 2015 dataset). There are problems faced wherever you are. There is a difference between the availability of radiotherapy and the demand for radiotherapy.

There are well-resourced countries, such as Scandinavia, who want to use technologies to innovate and make radiotherapy treatment better. However, in large parts of Europe, including the UK, we are struggling just to treat patients and we need technologies to enable us to work faster and clear our backlog.

Delivering radiotherapy has become a very sophisticated digital process and imaging is at the heart. The first

part is going into a CT scanner so that a virtual representation of your insides can be built, to plan your radiation treatment. We have really precise machines that target radiation therapy, but they are no good unless the machines understand where the cancers are, and what healthy tissues/organs not to target, but to shield from the radiation.

Are there techniques that we already have, that can help us with this process? Humans are very good at diagnosing cancer. Project InnerEye is interested in computers being able to mark out how far the cancer has spread in the body and mark out where the healthy structures are. At the moment, there aren't any AI solutions for that in radiotherapy. It is really important in radiation therapy to be able to do that. Exactly the same tools that are being built for this work will be useful for all kinds of cancer treatment.

Generally, imaging is used to work out how well a treatment is working or not. It would be fantastic to be able to take a patient with a brain tumour for example, and have a computer look at the scans, and instead of trying to take measurements of the scans on the screen, receive quantitative information about exactly what is happening to different parts of the tumour.

Raj described the time involved in his work treating the tumour glioblastoma. It takes around 1.5 hours to do the radiotherapy preparation for each patient. This is additional time that does not appear in a clinician's job plan. Therefore, it always needs to be done as quickly as possible. With AI technology, the time may be reduced to 5 minutes per patient. This means Raj has time back to think about the patient's care, think about the radiation and how it can be improved and individualised for each patient, to minimise damage.

Raj handed over to Antonio who gave an overview of his presentation: computer science, how the data is used, a flavour of the algorithms, and a live demonstration of Project InnerEye.

Data in this case is anonymised medical images and these are reviewed by the application. Antonio showed an image with a glioblastoma. As Raj had outlined, the data preparation for the tumour is intensely time consuming and involves highlighting and outlining the contour of the tumour, in every one of the slices. Results can be variable.

In the application an automatic segmentation can be selected. This is one of many services (computer software services) being developed and which run on Microsoft Azure Cloud (the servers are in the UK). There are different services for different clinical work flows. Antonio gave a live demonstration of the system. Selecting the service for the glioblastoma, he demonstrated the rigorous version control. Once connected to the internet, the image is anonymised, compressed and encrypted and is sent to the software service/APIs (Application Programme Interface). The image is analysed and the results returned within seconds. Antonio can view the image in 3D and he can measure the tumour volume.

Clinicians do not have good tools to measure, quantify, delineate and segment tumours. This technology works on images other than Magnetic Resonance Imaging (MRI) too:

- Antonio demonstrated a CT angiogram, showing how different arteries can be automatically segmented – e.g. the aorta or renal arteries.

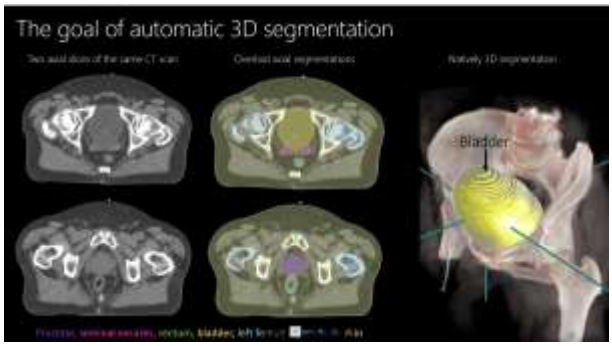
- Prostate cancer is another case; for the purposes of preparing the patient for radiotherapy treatment, the prostate and the healthy anatomy around the regions (organs at risk) need to be delineated. This would take an expert radiation oncologist up to 45 minutes; it took the computer approximately 20 seconds. Segmentation can be refined further by the human oncologist.

AI scare stories are far from the truth; Project InnerEye is a very practical tool that is designed to help human experts and amplify their ingenuity. It is a workflow acceleration tool and the human experts are always in control. Antonio prefers the term 'machine learning' (practical software tools) to AI.

How exactly does patient data help achieve the resource? Project InnerEye does not do diagnosis. Expert

doctors are much better than computers at diagnosing, in general. The work is to assist doctors with treatments and treatment planning in particular. The technology is focussed on segmentation - the delineation of healthy anatomy, tumours and other anomalies.

How is this achieved? If we look at images such as a CT scan, as demonstrated by Antonio, it is messy and unclear where the boundaries between different structures are.



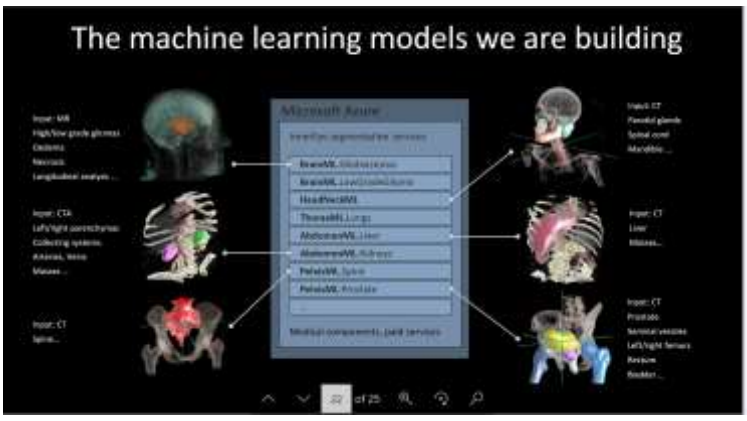

The machine learning scientists and computer vision researchers needed to design an algorithm that did not just use the pixels but uses spatial context too, to try and help with the classification of the pixels. Each pixel is assigned to a different class.

This is where patients came in, via a repository of images which are fully anonymised and consented. It is impossible to recover the identity of the patients from the images alone. It is the pixels that are essential and what the scientists are interested in. The experts, such as radiation oncologists, have also provided expert annotations (delineations). They have drawn these by hand, using their own clinical software.

In contrast to media stories, the InnerEye algorithms are actually economical in the use of patient data. As part of research efforts, the scientists are exploring algorithms that use much less data. The results demonstrated earlier come from machines that have been trained with a maximum of 100 patient images. This is very important for GDPR purposes too, which says that you must use the minimum amount of data necessary to achieve your goal. This rule is adhered to, even when using anonymised data.

How is the organ/tumour type segmented? For this probe voxels need to be examined. A particular machine learning technology is used, called Deep Decision Trees / Forests (sets of Deep Decision Trees). Each Tree analyses each voxel belonging to the three-dimensional image and for each of the voxels there is co-recorded map of probabilities e.g. the bladder.

As the results of the first iteration are not very accurate, ideas from ‘deep learning’ are borrowed. This is a way of reusing the results/output of previous computations: these are input into the algorithm to retrain it and make sure that it corrects its own mistakes. There is a chain of Forests, where the probabilistic results of previous computation are used as input to the next, and so different levels of contextual information are captured, amongst which are semantic context. Once you go through a number of these iterations you have built an accurate machine learning model, with a set of Decision Forests cascading different layers to capture more and more of the context.




Antonio highlighted further examples of test images, using images of patients that were *not* used in the training and optimisation of the algorithm.

With a high level of variability in the shapes/size of the patients, the algorithm managed to capture correctly where things are, on the first layer. The next levels are refining until there is a good accuracy and sharp outlines of the different organs.

Antonio handed back to Raj, who said that, after working together on the Project for the last seven/eight years,

it is over the last two years that they've been building something that could work commercially and be brought into the clinic. The technology is now integrated into Raj's department at Addenbrookes and it's amazing to see the power of what can be done, with just image data, to build tools that allow the day jobs to be done much more quickly.

It's now time to begin to dream of what could be done in the future; can this help us to do the next level – adaptive radiotherapy. It takes six weeks to deliver curative RT treatment and doctors have to close their eyes to the fact that in six weeks the tumour will shrink, the person will change shape underneath and things will move around. Some of the movement can be dealt with, but the same radiation treatment plan is applied despite the size of the tumour changing.

The goal would be to scan, identify and track the tumour every time the patient comes for treatment. The hardware is available for this, but not the manpower. Perhaps this new technology will change that. This would:

- benefit the patient by identifying tumour size, re-optimising the treatment plan and reducing toxicity
- lower costs significantly in terms of reducing follow-up care needed (to manage the side effects of treatment).

The only way to achieve this at scale is to use AI techniques; there is no other practical way.

Raj summarised the Project's approach to exploring this area of patient data. There is a very conservative approach to patient data. Patients who have been diagnosed with a malignancy are approached and asked for their permission to use the radiological images that have already been captured for their treatment. They have the strictest of anonymisation protocols. The anonymised data they receive is treated as though it is patient identifiable information and the same security arrangements are applied.

It is fascinating to see the acceptance of approaches like this, with Information Governance and the Ethics Committee in exploring this as a collaboration with the hospital. It's about empowering people to control the data, understand where it is going, understand that this is being used for a very defined and worthy project. The ultimate goals are to save time, get patients on to treatment more quickly, and reduce costs of the actual treatment. And, once patients are in treatment and/or have completed treatment, to be able to analyse what is happening to the tumour more accurately, in order to make predictions of outcomes and tailor treatments to a better degree.

**SHERLOC**

Raj gave a short second presentation about a consortium he has joined, SHERLOC, to work on a Cancer Research UK Artificial Intelligence (AI) Grand Challenge. This is a large challenge, worth about £20 million of funding, to look at using AI to achieve earlier detection of cancer. SHERLOC means Signs of Health, Early Recognition, Looking Out for Cancer.

The goal is to build an AI platform for the early detection of cancer, that will use both clinical and non-clinical data. The idea is to learn from established data such as clinical records and medication data. It will also use less traditional methods such as loyalty card data and internet search histories. The focus will be on patients diagnosed with lung, bowel, oesophagus or gullet, pancreas and brain cancers. As well as developing the AI, which is the part that Raj and Microsoft Research colleagues are involved in, there is a holistic research group looking at the ethical, legal and societal issues around this. Ultimately the project will link up with Manchester Connected Health Cities that has a learning health care system across primary and secondary care for research in consented patients, to trial the system.

These cancers were chosen as a large number of people present in emergency departments. The motivation is to reduce emergency presentation of these cancers, as an emergency usually means poorer outcomes for the patient. There are aspects around emergency presentations that relate to the tumour itself, and there are some behavioural factors too. The project will try to combine behavioural and clinical factors into the process.

A group of different AI applications has been built, with the hope that when everything is put together it will be greater than the sum of its parts. Patients who have an established diagnosis will be approached and, through consent, be asked about their early cancer history. Methods used will be questionnaires and interviews, to obtain clinical information. Patients will be asked for permission to look back at their internet search histories, to add details about the way that they find information. That will be both qualitative and then, if the patient agrees (and there is a differential consent platform), their internet search histories over the previous six months will be viewed.

Through a partnership with Sainsbury's Nectar, over the counter data will be explored and there are mechanisms for other stores' loyalty card systems too. This data will be combined with literature reviews around behaviour and risk factors for when people are diagnosed with cancer. A model has been built with cancer registry information, to show how much cancer to expect in a particular area.

All of this information will be built into a machine learning model. Larger anonymous datasets will be integrated and, once the final model is built, the pilot can be run. The integrated data will sit within a research platform and there are technologies that will make the data available to patients, including a technology being developed by Microsoft Research called 'Health Avatar'.

An interesting aspect is to think about engagement with patients and citizens. This is via interviews, focus groups and discrete choice experiments (if these are done at scale you start to build really interesting data).

If the system identifies a person at risk; what is the best way to bring information to that person in a manner that is bias-free, non-threating and non-judgemental? Different ways of these interventions will be explored: is it best to have contact via a GP, an alert, an application?

Thanking Antonio and Raj, Mike brought the session to a close commenting on the contrast between the datasets. At one end of the spectrum are small and anonymised datasets used to improve the care pathway and the clinician's work programme. At the other end is a dataset that spans all sorts of medical records, personal records, search histories and shopping patterns. There are then the ethics around what you do if you find something out. Two very different approaches with different implications.

**Questions and discussion**

**Q** A part that is missing in SHERLOC is the interaction with patients/ordinary citizens. There is a group who are always quantifying themselves, which you could immediately engage with. Most of those would not already be ill but would have a mass of information that you could accumulate and then find out, after they become ill, what things are correlating.

Everyone has a smart phone that tells you how much they are walking per day, as a minimum. That would be a pretty good indicator if they are fatigued or not, or whether it also relates to other illnesses. There's not enough relating directly from the patient inwards; it is going outwards to them, and this should be looked at more.

**A** Raj – that's a very valid point. To run the pilot we are looking for very small signals. Most of us are not developing symptoms that lead to cancer. When we run the pilot in Greater Manchester we will look to groups with risk factors and where there are known things being tracked in their medical history.

Dealing with smart phones/apps – our concerns are that for the group of people at risk of cancer, we didn't want to build a system that relied on all of that information. Many people coming into the cancer risk period at the time that SHERLOC is being tested won't be "digital natives". We wanted to design a platform that could work without requiring smartphone technology.

The interaction between citizens and potential patients is interesting e.g. sometimes with shopping data, the shopping may not be for the individual that you are mapping the risk for. So, we don't want to make it too much about the individual but want to take a slightly wider approach.

Timescales for SHERLOC are:

- Cancer Research UK Dragons Den in July 2018, with results announced later in the year
- it would open in 2019
- the pilot wouldl run from 2022 to 2024.

**Q** This is great work. What could you do if you had more data?

**A** Antonio – the unwritten rule of machine learning is the more data the better, but at the same time, we try to be as sensitive as possible and use the minimum data possible. Much of the research we are doing is on what algorithms work best, with as little data as possible. That works both ways, because if you do develop such algorithms you can train them with 100 images or 1000 images. Thus, economical algorithms trained with 1000 images could work as well as expensive algorithms trained with ten times as much. It is a good research direction to take.

What if you have more data not just in the sense of numbers, but in the sense of more diverse data e.g. genomics, patient history, shopping data? We greatly suspect that it will help get better accuracy in the results, but we are starting from a very practical point – what is the minimum viable product that will get us off the ground, get us traction. We can deliver something that, as it is, can save the NHS millions of pounds.

**Q** This is operational at Addenbrooke's now and presumably evaluations are taking place; when are you expecting it to spread?

**A** Antonio – it has already spread. We are working with a set of 10 to 20 hospitals around the globe. We have big traction in Latin America, we have a presence in New Zealand, Australia and Scandinavian countries. All of this is for evaluation purposes. It is really important in machine learning to avoid bias; we need to train algorithms with as much diverse data as possible and this means training the algorithms with global data.

**Q** For the use MY data community a recurring theme is that we want these practical examples, of how data is used. When would you expect this to be standard practice for a use MY data patient, no matter where they are in the country?

**A** Antonio – The answer depends on so many ifs. Will the NHS adopt it? Will we have a fair commercial model? Will we have robust enough implementation? How to turn into reality; as a global company Microsoft can help. A wide guess is that it would be operational in the UK in 2 years.

**Q** This is wonderful and am supportive of both projects. Where are you with health economics and the cost of this? How to model so everyone can benefit?

**A** Antonio – We have started to think about pricing models. We are very well aware that the NHS needs to reduce costs.

One of Project InnerEye's main goals is to make sure that institutions like the NHS provide better treatments to patients while at the same time reducing their costs. It is all about using machine learning technology to remove/ameliorate inefficiencies from existing clinical workflows.

Mike – to add the counterpoint and going back to Raj's first slide, we need to treat more patients with radiotherapy. It isn't just about saving money, it's about saving lives.

**Q** This is really exciting and interesting work. Raj, you mentioned that the adaptive radiation therapy might have the potential to reduce side effects for patients, due to reducing the radiotherapy. Are you studying that at the moment and already seeing that impact?

**A** Raj – yes, we've had a 5 year project running already, looking at computational aspects, funded by Cancer Research UK. We're about to embark on another phase of that project, which will actually have machine learning technology at its core.

There are different ways of doing adaptive radiotherapy and some of those are already out there. It is used as a valued resource that we can't implement across the board. For these sort of technologies, as we get the equipment to do it, we can implement for larger and larger numbers of patients. A good example of this is IMRT (Intensity-Modulated Radiation Therapy) which was the first technique that allowed radiation to be given much more sharply, aimed at the tumour and shielding normal tissues. Going back as little as 7/8 years ago, only 8% of patients were receiving IMRT in the NHS and now it's about 40% (at Addenbrookes it's higher than that).

From a health economic point of view, you can calculate the cost of long term toxicity that comes from curing someone using radiation therapy.

Implementing this kind of technology in the Cloud is timely; five years ago, the NHS was not ready, but it is now. The cost of adopting this as technology through the cloud for an individual health care provider is very small. This is very important when you think of the cost of radiotherapy machines, which are £1.5 million to £6 million each.

**Q** Very impressive, especially InnerEye. When talking about making it more widely available, you are obviously dependent on the quality of scans; is that a problem at all, and if so how widely? And what about the types of tumour? The ones you were showing seemed to be fairly well defined. Do you expect/hope this would be applicable for more diffuse tumours?

**A** Antonio – On the scan type and quality; this is a great concern of ours and it's one of the reasons why we are working with multiple hospitals. There are different qualities of scans and different characteristics of images coming from different manufacturers' scanners. It looks as though the technology can be trained and optimised to work across the spectrum of different types of scanners. It is not 100% correct and will probably never be 100% accurate, as humans aren't either and it is not clear what the 'gold standard' should be. We do expect variability in some cases in segmenting several structures e.g. the femur might not work well if there are implants. We track the issues and this is part of the regulatory process. This is another reason why the system is not completely automated and experts are in the room.

On the tumour types – we started with glioblastoma as that is Raj's speciality and he challenged us on this. More tumours are being worked on, though there is not enough experience yet in diffuse tumours. Gynaecological tumours are very difficult and more experience is needed.

**Q** SHERLOC and psycho/social behaviour; we are here talking about data and you want to use the minimum amount of data that you need. From the data that you would like from citizens that would contribute that kind of lived experience, what is the minimum data you would like from them?

**A** Raj – Very good point and we have tried deliberately to layer our system. One of the steers we were given was not to solve this problem by boiling the ocean of extant data. Our approach is much more focussed. Even if the psychological factors are just picked up from interview, hopefully that will give us enough of a clue. Although we would like to combine and layer with other information if possible. We are not looking to profile people through consumer datasets. We are looking to build our list of clues by talking to cancer patients and then project them forward into a new dataset.

**Q** The five cancers pinpointed (such as brain and oesophageal) are candidates for proton beam therapy – is your model (SHERLOC) applicable there?

**A** Raj – Ultimately, if SHERLOC worked well, what we would like is to adjust the stage distribution of those five cancers. As you mention proton beam therapy may be useful for certain brain and oesophageal tumours, but they need to be at an earlier stage. So, it's a big if, but if this works and we could actually demonstrate that by using this technology the cancers that come, come at an earlier stage, then it means more people might be able to explore more curative technologies.

Mike closed the session thanking Raj and Antonio. It was clear that there was a great deal of support for both Project InnerEye and SHERLOC and that this is exactly the sort of use of data use MY data wants to see, particularly with respect to:

- the recognition of privacy and confidentiality issues
- the recognition of ethical issues to be faced, for SHERLOC.

use MY data is always looking for practical examples of how patient generated information can make a difference and the projects are really good examples.

**Session 4 Use my data, use my DNA?**

**James Peach**
**Precision Medicine Lead, Medicines Discovery Catapult**

This session explored how patients, who want to make sure that their consented samples and data are used for medical research, can ensure that this happens.

James opened the session emphasising that research is being hampered due to a lack of access to data. He wants to understand how involved delegates (and specifically use MY data) wish to be in enabling projects involving samples and data to happen.

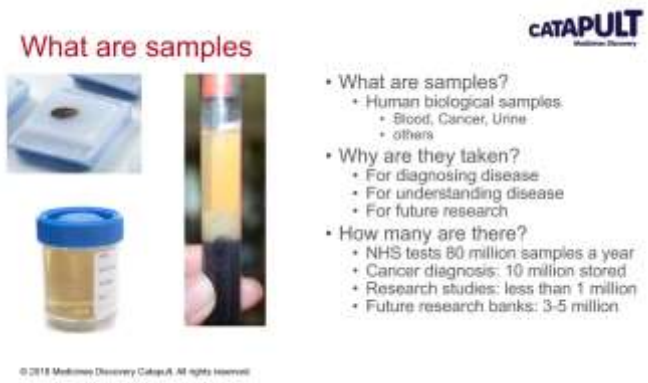
James outlined the areas his talk would cover:

- How samples are used in medical research
- Context of sample and data use in the UK (specifically use for the community of drug developers)
- How patients in the UK can ensure their samples are used.

Right now, there is *no way* to ensure that your samples, which you have provided consent for use, are used or indeed to check how they have been used.

James used the example of a lung cancer drug to emphasise the importance of using samples in research. The drug works because it was designed through combining samples and data. It is a drug that only works with people with lung cancer who have a certain genetic mutation. This was worked out by analysing multiple tumour samples, to understand whether the DNA change was in there. It was possible to work out, through the data collected, that people who had the genetic mutation did better on the drug.

- What are samples?
  Examples: solid tumour samples which are collected from the patient, put into preserving fluid and then stored in wax blocks. Other samples commonly taken in the NHS are urine and blood.




- Why do we take them?
  This is principally to help the patient's care. To diagnose a disease, to understand sub-types of disease you may have and to help research in the future.

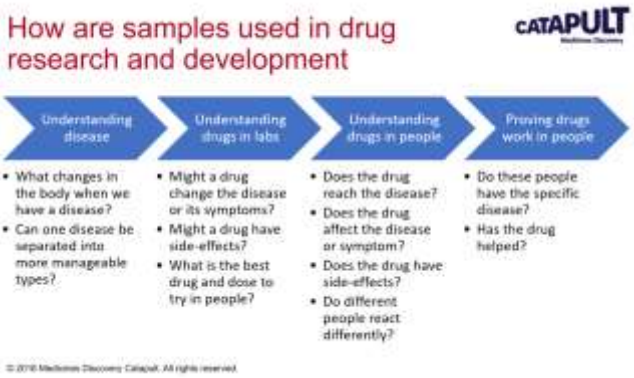
- How many samples are there?
  The NHS tests 80 million samples a year: that's 300,000 samples every day. In terms of cancer diagnosis, the samples are stored for a very long time; across the UK there are approximately 10 million cancer samples held in hospitals.

  There are also research studies: James estimated about 1 million patients have been consented to samples being used for research studies.

  Then there is the research area where patients have given consent for their sample to also be used in future studies. Samples are then stored in research tissue banks / biobanks. There are 3 to 5 million samples help in research tissue banks in the UK.

How are samples used in drug development?

- To understand disease
  What does the disease actually do inside your body; what are the changes in the body? Are there genetic changes, protein changes? Once you know that, you can separate one disease into sub-types, some of which might actually be more manageable. E.g. melanoma skin cancer was extremely difficult for drug development for a very long time. Once it was separated into sub-types, one type became more manageable.




- To understand drugs in a laboratory context
  This could involve taking a set of cells that are growing and have a disease, or the characteristics of the disease. You wash over many different types of drugs and work out which one helps the disease, or not. This can also be done to healthy tissue e.g. to try and understand whether the drug will have an impact on your liver, a common area for toxicity. The next step is to try and understand what might be the best dose to try on the disease.

- To understand drugs in people
  Does the drug reach the disease? You can look within the sample to find the drug. If you are targeting the eye or liver, for example, you might try to take a sample to see if the drug has reached the area.

  Does the drug affect the disease, or symptoms? Does it have side effects? It is important to find these when they are at a very small level. This can be done by finding chemical changes within the body that might predict a later side-effect.

- To prove that drugs work in people
  These are the final trials, where you think that a drug will work and need to test it on people.

  Increasingly, rather than trying to treat all people with one disease, the approach is to try and break the disease down into different diseases. The Institute of Cancer Research recently published a study that showed that people who had the BRCA mutation and triple negative breast cancer shouldn't have the normal drug, they should have a different drug. The only way they know this is because they took samples from the patients to understand who was BRCA positive and triple negative.

Samples are used all the way across research and development and are absolutely vital to it.

The UK has very good infrastructure for collecting samples and data. We have first rate pathology units across the UK that can handle samples in the right way for complex analysis, including genetics. We have the infrastructure to collect data and transfer this. Connected Health Cities is an excellent example of how data can be shared.

Many millions of samples are taken. Lots of data is taken within the NHS. Although James has searched to try and find out just how much he has not found the amount. However, King's College Hospital alone has stated that they store 180 terabytes of data. Worked out on a per patient basis for the Hospital, that is 50 bibles per patient of data. Some of the data has patient consent for research and development.

The Genomics England programme, which James previously worked on, is a good example of using samples and data. The programme set out to try and improve the understanding and management of cancer and rare genetic diseases by asking a large number of patients with these diseases, if their DNA could be studied. This was with the aim of trying to understand the causes of the disease and potentially how the type of disease they have, interacts with their treatment and outcomes.

We have great samples and data, we have groups of patients who want their data to be used and in some areas we are managing to do it. But there is another side.

The Medicines Discovery Catapult is trying to help small companies in the UK do research and development. At the beginning of last year, a study of the Chief Executives of 100 small biotech companies (people who are developing drugs) gave the following results:

- Access to samples for commercial development (new drugs) is hugely important
  There was overall agreement from survey respondents

- It is easy to access high quality NHS samples for commercial research
  There was overall disagreement

  Everyone knows that this is something that really needs to be done, it really helps the research, but it's very difficult.

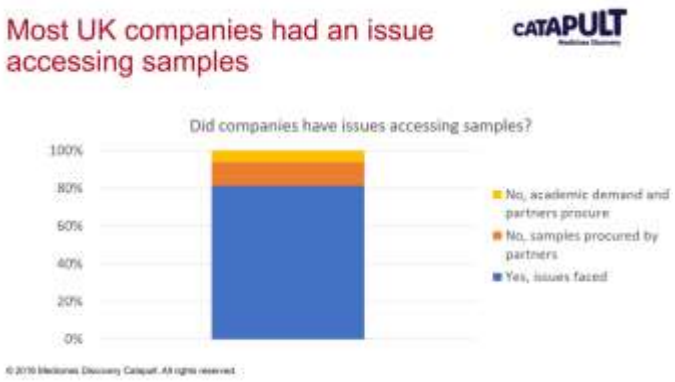
- Access to the right health data
  There was overall agreement

- Is it easy to access the right health data, including registries and activity
  There was overall disagreement

There is something going wrong here, where we have a group of people who are meant to be developing treatments for us in 10 or 20 years' time, who cannot get the information that they want.

Medicines Discovery Catapult followed up with a deep dive into diagnostic companies; people who are developing new ways of understanding what disease you have. The first question asked was:

Did you have a problem accessing samples?




Over 80% said yes. The other 20% that were able to access samples, had not directly accessed them, instead others had done this on their behalf.

Medicines Discovery Catapult then worked with the diagnostics companies to find out what was causing the problem. You might have thought that the problem would have been patient consent. But that wasn't the case generally. A third of respondents said that consent and governance was a problem. Other significant factors were the cost of access, the speed of access and then another big issue was around finding samples. Consent, data quality, sample quality, sample volume were also in the mix. However, the most common issue at just under 70% was procurement. This means the process, after we've worked out who has a sample and who needs that sample for research, of getting the agreement done and finished is problematic.

We don't know exactly why this is happening, but James presented a series of ideas.

There are good reasons why samples might not be used:

- There are laws to prevent someone's samples and data being used without consent

- There are laws that say when you handle samples and data you have to do it responsibly

- Doctors, who are treating patients, don't have time to respond to research requests to use their previous patients' samples and data.

Also, some potentially bad reasons:

- Sample and data access is seen as an advantage for the owner (the researcher who has collected the sample/data to do their own research and therefore may be reluctant to share in case they use up the samples that they need for their own research)
- Ensuring the laws aren't broken is difficult, so doing this inefficiently costs money and takes time
- Offering other people access to the samples and data you have collected for your study is not rewarded scientifically.

Many people give samples for use in their own care without broad consent for research, so these are not on the national directory. They will never be seen by companies looking for them, for example searching sample directories on the web. The only way to find them is if you understand that a specific person has taken the sample and has the sample, and you go to speak with them. There is a national directory, the UKCRC Tissue Directory and Coordination Centre, that those who want to use samples can access. The directory shows where samples with broad consent for research are. But if you have given samples as part of your care or for only one specific piece of research, they are not likely to be recorded in the directory and this means they will never be found.

The current estimate from the biobanking community is that only 15% of samples are used, leaving a huge 85% never used. When people give samples to a research tissue bank, they sign up via an explicit and detailed consent process, which involves training the person taking consent, creating and getting approval for the consent materials, and having a discussion to bring out and respond to any concerns the patient has. So after all this, only 15% of samples being used feels unacceptably low.

James highlighted that 50% of the sample collections in the UK are not findable on a web directory.

Of those samples that have been given for general research reuse (kept on a research tissue bank consent), 85% have not been used.

What can we do? Do we need to create more samples, or do we need to make it easier to find samples and access samples? James suggested some possible steps to increase sample usage.

1) Encourage an efficient system of broad consent within the NHS for sample use

2) Ensure samples and data are visible
   How do we make the samples and data easy to find? We need to make sure that all samples are on the national directory.

3) Ensure samples and data are shared
   In terms of access this could be having one or two specialist support teams in the UK that facilitate sample and data access. Large pharmaceutical companies have these specialist teams in-house.

How do we know if our samples and data are actually being used? We just don't know unless we can find details of specific studies. How can this be monitored? How can we ensure that our samples and data are visible and they are shared?

Questions that patients can ask a hospital or researcher:

1) Why doesn't the patient information form (that patients are required to complete) contain a broad consent clause?

2) Are my samples & data findable on a web directory?

3) How many of the samples & data have ever been used?
Genomics England and UK Biobank do this well and work really hard to get the data used. All of the data within UK Biobank has been used at least once.

4) How do you make it more likely that my samples & data will be used?

The question is not, 'how many samples does a biobank have?' it should be 'how many samples has a biobank used?

James concluded by saying that he would very much like patient feedback and involvement to take this work forward. Thanking James for this talk, Mike asked delegates what they thought about data not being used.

**Questions and discussion**

**Q** I work for a bioresearch centre at Guy's & St Thomas' and for every £1 that is put into the research centre, they get £6 back for the NHS. We are not selling data but go into partnership to find cures. I can see that this data is valuable. Big pharma is involved but it's a real problem getting it out to small researchers and universities, who don't have the same level of funding. Perhaps we could have a central place for researchers to go to. Because there is funding available, we have funding to go in with researchers - but how do they know where to go / whom to approach?

Mike – If you are collecting samples and involving patients, and presumably involving tax payers' money in the collection and storage, then surely there should be an onus on the researcher to use the samples/data? Or to stop collecting the samples, if they will not be used.

**A** I would like to see that and there are some moves towards it. Some big non-governmental research funders, Medical Research Council, Cancer Research UK, Wellcome Trust, have now said that if they fund research that involves collection of tissues, details of the samples have to be put on the registry. That is a good start, but it is not enough as the registry is just the first step to getting the samples reused. I would like researchers to proactively report on this and to give details of plans to increase sample use, if they are not being used.

**Q** Antonio - Very interesting. On the usage point, we haven't discussed enough the reasons why the data doesn't get used. There are similarities to what we see in medical images. Organisations, such as the National Institutes of Health (NIH) in America, has an archive of freely available anonymised medial images of patients; some of these get used, many of the images don't get used. The question is why does the data not get used for research purposes. It is good to collect masses of images and make them available for people to use. If those images are not structured in the right way they become much less useful. For a radiotherapy project, I want to work on images which I used for the planning of radiotherapy. For other things, such as monitoring the evolution of a low grade, less aggressive tumour, then I want to make sure that those images were captured for monitoring a low-grade tumour. There might be different imaging protocols used, or different stages of the disease. If

I'm not sure how the images were acquired and for what purpose, then the data becomes a lot less useful to me.

**A** Absolutely right, and I'd add as well that the older samples get, often the less useful they become. For example, this is because of changes in the standard of care, which means the drug that a new drug needs to do better than, or the radiotherapy that you need to improve on. So, if the data was collected twenty years ago, you may be able to say that your drug/treatment is better than what was

there twenty years ago, but if the standard of care has changed, the results won't prove your drug/treatment is better than today's standard of care.

**Q** Difficult to envisage a situation where anyone involved in the collection of data/permission to use data would actively, positively refuse. I can't imagine this. This suggests a lack of awareness and understanding. Is there not room for a campaign addressed to both the donors of data and the protectors/users to make them aware of the overall activity and to promote the idea?

**A** Mike – There is probably something cultural here as well. I imagine if your life's work is to collect samples, that could be your focus, creating a block to thinking that the goal isn't to have a magnificent freezer full of samples, it's to find a new drug, or an AI algorithm.

**Q** Have you seen anything from other countries, how this might have been done exceptionally well and the rates of use of biosamples increased?

**A** In terms of changing the rates of use of biosamples, I haven't seen anything. But I haven't looked that hard as the MDC is focussed on the UK. If anyone has international examples, I would love to hear from them. In the US there is a presumption of consent for excess surgical tissues, so that if tissue is removed for clinical analysis, any left-over tissue can be used for research (it is anonymised). This is not yet the case in the UK but I believe initiatives are starting to do this.

**Q** Chris – Thinking about why use MY data started, which was because lots of data was being collected and it was inaccessible to researchers; people were up in arms about this and wanted their data to be used. The amount of samples not being used is a disgrace, and not something I was aware of. Alongside support for data being used, I think there is strong support for tissue usage. We need advice about how best to make this happen; where do we have to have a voice and who with?

The contrast between data and tissues; data is the gift that gives on giving, it doesn't run out and will be there forever. Obviously tissue does diminish over time, in terms of both quantity used, and the use-by date. Supermarkets have a sale when items get to their use-by date. Perhaps we need to encourage the people who hold tissue samples to have an active push-out.

**A** I agree, we need an efficient mechanism so it is not difficult, to make it easy for people to get onto the directory, to make sure it's easy for them to make contracts and to ensure those applying for data are legitimate research and development organisations. We need an understanding what is not being used and what might go to waste.

**Q** Is it that there aren't enough researchers available to use the tissue samples?

**A** It doesn't seem to be the case and there seems to be enough researchers. Perhaps a research project may not match the tissue or data available.

**Q** Raj – This is absolutely fascinating, working at the point where I consent patients into clinical studies and to understand that it's not consent that is the limiting step. If you jump domain into high energy physics, when there is a publication about the Higgs boson there are 6,000 names on the publication; it includes absolutely everyone involved. Researchers of tissue data think that the curation of the samples adds value, that they can monetise the data and they can increase their scientific impact through that process of curation. So, do we need to push at changing those perceptions amongst the people who are effectively acting as the gatekeepers of the data?

Can we use technologies from Cryptocurrency / electronic ledgers, to actually expose this to patients, so that if your dataset that you gave (at considerable discomfort to yourself) to go into a biobank, is not being used you are aware of it and can start making noise. Are there other ways of looking at the problem?

Mike – What is the feedback loop to patients about how their tissue is used?

**Q** Presumably samples are taken for a specific reason and patients consent for a specific research study. There are times when research can't happen; if so would it not be good to go back to those in charge of the data, to say patients want to know what is happening and if you are not going to use their samples, perhaps the patients would like them to be destroyed. There is a responsibility for those holding the data to do something with it, because nobody wants to give their data, or their samples, and then nothing happen with it.

**A** Completely agree as an option. What is causing this, to a certain extent, is that lots of organisations have been set up to ensure that no one breaks the law, such as the Human Tissue Authority, Health Research Authority, Caldicott Guardians. This is right and correct and protects our privacy and rights, but this has made an imbalance as there aren't any corresponding organisations set up to ensure that samples are used in research. We need to ensure that research is being done. One way could be to try and get the communication method back with the donor, to hear their choices and inform them of progress. Twenty years ago that would have been an amazingly expensive postal operation. Now, with email, it would be cheap and easy to do. If every different research study tried to do it off their own back it would be a nightmare, but should we be thinking of something a bit more central that could help people through.

**Q** I presume that most samples are taken with a view to have a diagnosis for the patient. So, the sample has already served a purpose for me. What I'm not sure of is what I can do to enhance your idea to get samples reused by other researchers. I can't see what I can do here.

**A** At the moment the only consent form you get is the one you are offered. As an individual, you can't go into a hospital and offer your samples for general use. The answer must come from us/patients in this room. There is an organisation in the US called Patients Like Me – groups of patients who volunteer to share information about their condition, including samples, as an attempt to get research and development working quicker in that area. There is a lot we can learn from this approach in the UK.

**Q** Do you know how many samples we are currently sampling from the EU, and is there an impact on the demand for samples procured and stored in the UK?

**A** I don't know how many samples come from the EU, but I have asked those in the diagnostic industry what happens if you can't get your samples in UK. 66% of them said that they buy the samples internationally. So, UK samples sit in the freezer doing nothing and research money is wasted buying samples abroad.

**Q** I have given samples and I don't mind doing that, but is there a slight medical risk to the patient having samples taken? And is it expensive if samples that are taken are not used? It seems pointless and inefficient.

**A** The risk depends on the type of sample. Urine no, but liver and lung biopsy do have a clinical risk.

We need to find a way to do this efficiently, so that we don't collect the samples that we don't need. How do we make sure we use the samples that we take?

Mike closed the session by highlighting that there was much appetite and a clear steer from use MY data to get involved in the problem of low usage of samples

| Session 5 | Patient data information part 2 – what do we still need to know<br><br>Chris Carrigan<br>Expert Data Adviser, use MY data |
|---|---|

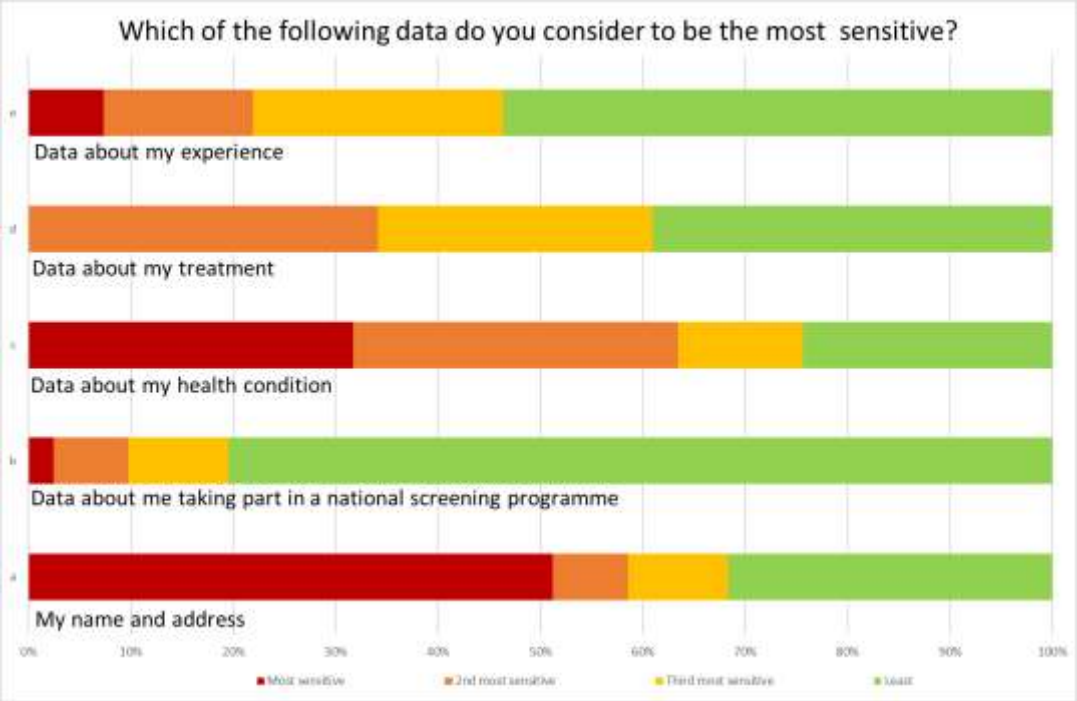
In this interactive session, the audience were asked to consider a range of questions and scenarios around uses of data via facilitated table discussions. The results were discussed briefly in plenary.

**Which types of data do you consider to be the most sensitive?**

We asked delegates to reflect on which of a defined list of data they considered to be the most "sensitive" to them. They were asked to select three from the list, with the most sensitive showing as 1, then 2 and 3 as the sensitivity decreases.

The results below show the overall collated scores for the types of data.

Notably, a significant number of people in the workshop felt that their most sensitive data is their name and address. This is clearly in contrast to the Home Office's stance on name and address being at the lower end of the data confidentiality spectrum.

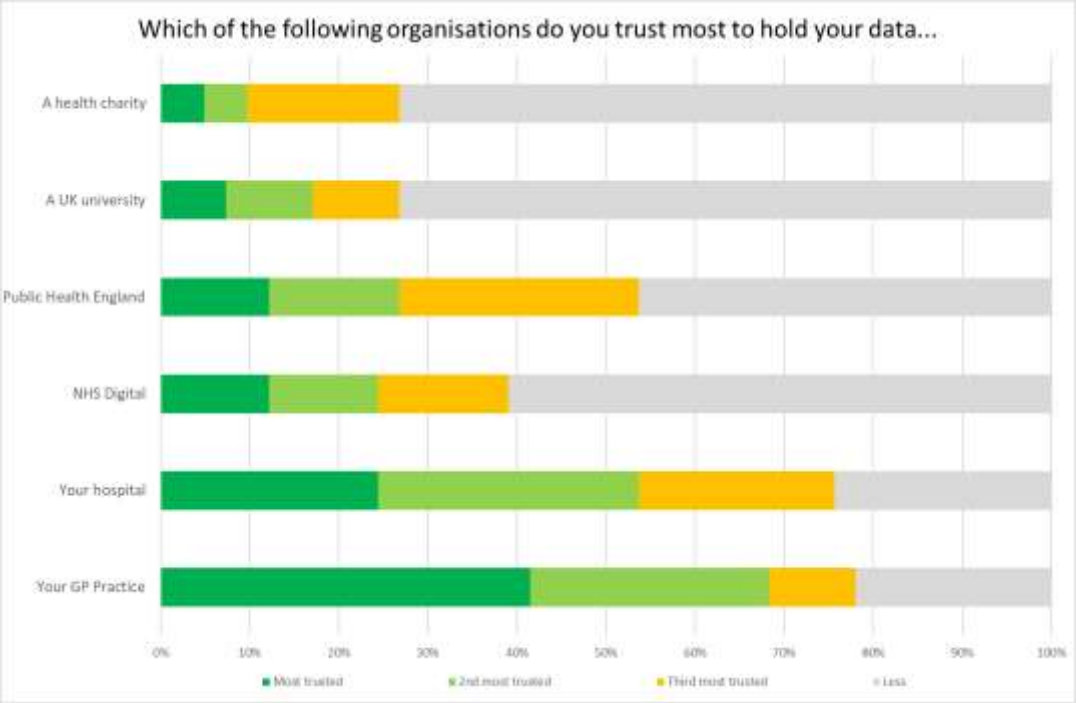



**Which of the following organisations do you trust most to hold your data?**

Delegates were asked to rank organisations according to the level of trust they had in the organisations managing their data. They were asked to rank three from the list, with the most trusted showing as 1, then 2 and 3 as the level of trust decreases.

The results below show the overall collated scores for the types of data. The grey elements of the plot indicate the frequency with which each organisation did not appear in the top 3. As might be expected, those organisations with the most direct contact with patients were seen as the most trusted, with GP practices coming out on top.

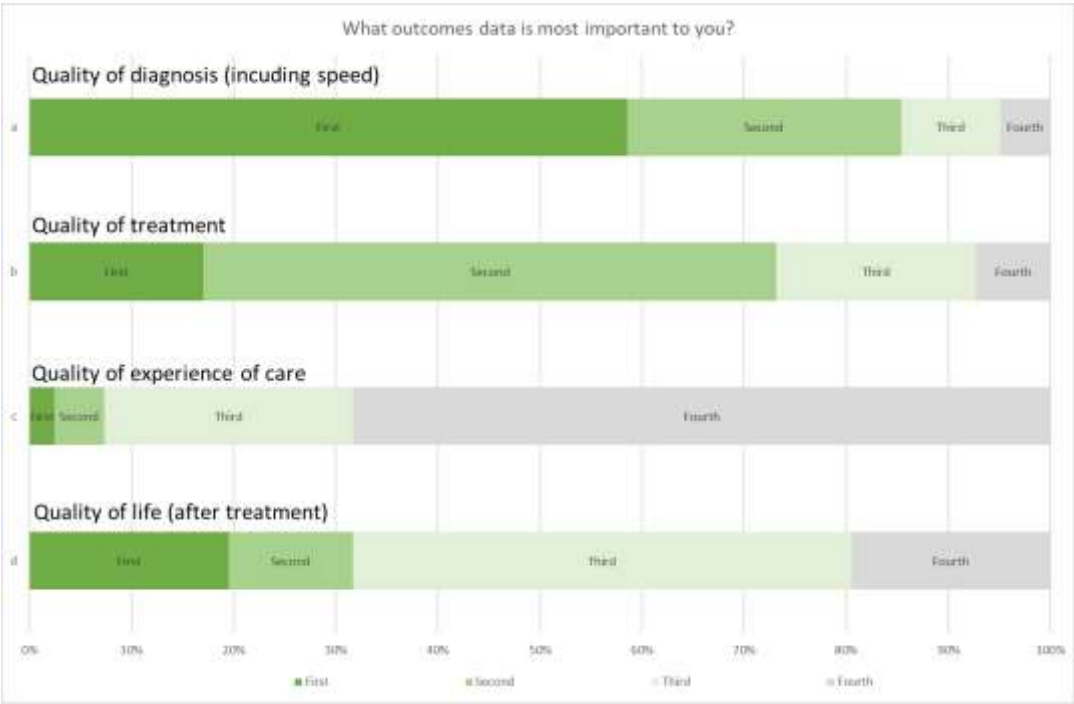
The two large health-data organisations, NHS Digital and Public Health England, scored similarly.




**What outcomes data is most important to you?**

We asked delegates to reflect on what types of outcomes data they considered to be the most 'important' to them. They were asked to select three from a defined list, with the most sensitive showing as 1, then 2 and 3 as the importance decreases.

The results show the overall collated scores for the types of data, with outcomes data for diagnosis and treatment scoring the highest, followed by the life-after-treatment outcomes.



**What is the data that you want to be collected?**

Following the initial questions, the audience were asked to discuss the types of data that they felt needed a greater focus for collection. There were no constraints on the types of data that could be suggested.

For presentation purposes, the (unedited) list of responses below has been grouped:

| (Group) | Responses: What is the data you want collected? |
|---|---|
| Choices | consent - for use of samples and data |
| Choices | consent - for approach to take part in research |
| Choices | whether offered a clinical trial |
| Choices | willingness to participate in research |
| Diagnosis | data around second opinions |
| Diagnosis | mis-diagnoses |
| Diagnosis | pre-diagnosis data |
| Diagnosis | symptoms |
| Family | impact on family |
| Family | family history |
| General | everything |
| General | longitudinal |
| General | radiation (Chernobyl) |
| Lifestyle | behavioural |
| Lifestyle | driving - black-box |
| Lifestyle | driving - medication, alcohol |
| Lifestyle | daily activity |
| Lifestyle | exercise - smartphone |
| Lifestyle | prescription drugs |
| Lifestyle | psychosocial data |
| Lifestyle | personal characteristics |
| Lifestyle | occupation |
| Living with | late-effects (a lot not collected) |
| Living with | PROMS - collect more consistently |
| Living with | post-discharge data (living with and beyond) |
| Living with | lived experiences (ups and downs) |
| Pathway | delays |
| Quality of Life | quality of life |
| ‘big data’ | data linkage |
| ‘big data’ | free text |

In addition to the types of data people felt should be collected, several noted that the real significance comes when data is linked together, and that this theme should not be forgotten. Respondents commented that “no individual dataset will solve everything”.

**Organisations you think hold health data about you**

Having discussed the types of data that are used, and then discussed other data which may not yet be collected, delegates were asked to discuss (and then list) organisations that they thought would hold health data. These are listed below, in alphabetical order.

In the discussions it was recognised that many of these organisations would only have data which was supplied (with consent). However, others are listed whom members felt may be able to derive health information from the data that they hold about an individual.

| | |
|---|---|
| Amazon (books that you have bought) | Opticians |
| Apple | Patient groups |
| Banks | Pensions |
| Beautician | Pharmacists |
| Charities | Podiatrist |
| Clinical Commissioning Group | Police |
| Dentists | Private providers |
| Department for Work and Pensions (DWP) | Public Health England |
| Driver and Vehicle Licensing agency (DVLA) | Registries |
| Employer | School |
| Facebook | Scouts |
| General Practitioners | Social care |
| Google (search history) | Social media |
| Gyms | Social services |
| Hospitals | Specific studies |
| Insurance companies | Store loyalty cards |
| Local Authorities | Supermarkets (what you've bought) |
| Marketing companies | Telephone companies |
| NHS | Twitter |
| NHS Digital | University |
| NHS England | Voluntary organisations |

The list of organisations was longer than delegates had anticipated. Delegates noted that there would be other organisations who hold data, but which had not been listed.

**The final question was about the patient pathway. More outcomes are focused on the pathway approach and delegates were given a typical pathway and asked to identify the other parts of a patient pathway which can be missed.**

In addition to the proposed pathway of pre-diagnosis, screening, diagnosis, therapeutic, progression, survivorship, death, experience and infrastructure, the following elements were also identified by delegates.

- Awareness
- Bereavement care
- Carers, dependency, patient context
- Co-morbidities
- Complementary therapies
- Dependency
- Diagnosis not simple
- Experience - different at different stages
- Family history
- Family impact/support
- Family risk-factors
- Growing group of patients with a chronic condition; they have a treatable disease – are they survivors?
- Lifestyle
- Living with (non-curable and chronic)
- Monitoring/watchful waiting – people under observation
- Multiple pathways
- Palliative care
- Patient feelings – they change throughout the pathway
- Pre-diagnosis – about awareness – the first step
- Pre-birth – the impact of parental experience
- Quality of life
- Self-treatment/medication
- Social care
- Support at home
- Support team (Clinical Nurse Specialist, key workers)
- The impact on families
- The point at which you become of support team available to you
- Treatment compliance

## Chair's closing summary

**Mike Birtwistle**
**Founding Partner, Incisive Health**

Mike closed the day, noted how fascinating discussions throughout the day had been. It's easy to be weighed down by the whys and wherefores of data protection and regulation, so it is good to be reminded of the positives around using patient data.

Some practical ideas have emerged that use MY data can follow-up on:

- The importance of an audit for the implementation of the National Data Opt-out, ensuring there is a common national approach
- The importance of making what data collection does real. A good example is the national clinical audits and making sure that they explain, in plain English, what they are collecting, why they are collecting it, how they are using it, who they are collecting if from and what the basis for them being allowed to collect it is.
- A focus on the very low number of tissue samples that are used and ensuring that when people have consented for their data to be used, it is used.

Rather than giving consent for the use of our data, we are instructing people to use our data. People expect their data to be used and are quite shocked when they find out it is not. There is accountability to put on those collecting data, to ensure it is not collected and just polished, but that it is actually used and that something good comes from it.

Mike thanked all of the speakers and delegates for their contributions to the day and the use MY data coordinating team. Mike acknowledged once more the role of patients, relatives and carers and the contribution of supporters, who all enabled the workshop to take place.

use MY data is not an organisation as such, it is a collective. And as such, is only as good as those willing to turn up and give their time – a big thank you.


## Acknowledgements

For their contributions to the workshop, use MY data gratefully acknowledges and thanks:

- Our workshop chair and speakers, for so generously giving their time
- Our delegates for their participation and time
- Trish Gray for her assistance on the day
- Our funders who enabled the workshop to take place:
  - Leeds Institute for Data Analytics
  - Health Data Research UK

| **Appendix 1 Organisations to whom delegates will feed back** |
| --- |
| Bloodwise |
| Bowel Cancer Intelligence UK (BCI-UK) |
| Carers together |
| Clinical Informatics Research Group at the University of Cambridge. |
| Chronic Myeloid Leukaemia (CML) Cancer Support Group |
| Disability Forum |
| East of England Cancer Alliance Patient Advisory Board |
| Group concerned with Early Phase clinical trials and the use of patient data including genomics in these |
| Guise & Dolls Head & Neck Cancer Support Group |
| Macmillan Cancer Support patient group |
| Healthcare Quality Improvement Partnership (HQIP) |
| Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London |
| Independent Cancer Patients' Voice |
| Patient Educator Programme, Kings College London |
| Leeds Centre for Personalised Medicine & Health |
| Local groups |
| Medicines Discovery Catapult |

National Cancer Research Institute (NCRI) patient liaison group

NHS Digital Programme teams (Information Representation Service and WiFi)

NHS PPF

Patient & Public Involvement Advisory Group (PPIAG), Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London

Quora.com

South East London Consumer Research Panel for Cancer, Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London

Speak Easy

University College London Hospitals (UCLH) Cancer Patient and Public Advisory Group

Understanding Patient Data

**www.useMYdata.org.uk**

**getinvolved@useMYdata.org.uk**

**@useMYdata**