

The answers to so many pressing questions about cancer, and about how best to diagnose and manage the disease and care for patients, lie waiting to be revealed through analysis of data that is currently spread across tens of thousands of databases. That same data, which is collected and held for different purposes in different settings – from healthcare and social care to research, drug development, screening, public health, cancer registries, insurance, medical devices, and more – also constitutes the raw material needed for developing the AI software and apps needed to implement personalised cancer medicine.

Yet decades of efforts to find ways to unlock the potential of that data have been frustrated by the difficulties in overcoming a range of knotty problems, from security and interoperability of databases, to quality control, governance, vested interests and protection of patient privacy.

Reaching agreement on how to overcome these problems, to help Europe reap the benefits of better access to these valuable data resources, is now high on the list of the priorities for the European Health Union. The first steps were taken in May 2022, when the European Commission launched its proposal for setting up a ‘European Health Data Space’, which the Commission plans to be operational in 2025.

The intention of this Data Space is on the one hand to give citizens control over their own health data, and on the other, to create a single market for digital health services and digital health products, ensure interoperability between products and services working with health data, and set up a framework for the reuse of health data for research, innovation, regulation and policy.

Stella Kyriakides, EU Commissioner for Health, calls the Health Data Space not just a pillar of the European Health Union, but [‘its backbone’](#). Speaking at the launch on May 3rd, Kyriakides presented it as: “An initiative that brings us closer to making health data work for patients, people and innovation... For the first time, we will have a common EU data space connecting all Member States, for the benefit of all EU citizens.”

A contract with patients

Built into the fabric of the Health Data Space is an acknowledgement that, while patients are often happy to contribute their data to research that could benefit future patients, their primary concern is timely access to their own data – test results, imaging and so on – to ensure that their own evolving needs for treatment and care are met as quickly and effectively as possible. This is referred to as ‘primary use’, with use of the same (but anonymised) data for purposes of research, innovation, regulation, evaluation or other generalised applications coming under the heading of ‘secondary use’.

Citizens will be able to access their own health data, and share it with healthcare professionals, including across member states

The patient-centred credentials of the proposed Health Data Space were emphasised by Fulvia Raffaelli, head of Digital Health at the European Commission’s health policy unit, DG SANTE, addressing the European Cancer Summit in November 2022. “Primary use of health data will further empower individuals to control their health data,” she said. Under the current proposal, she explained, citizens will be able to access their own health data in electronic form, free of charge. They will be able to share their health data with healthcare professionals, including across member

states. They will be able to add information, correct wrong data, restrict who has access to their records or to specific information they don't want to share. They will also be able to get information about who has accessed their data and for which purposes.

The move towards a harmonised European Health Data Space will also help patients requiring healthcare in member states other than their own, via a new electronic cross-border health service [MyHealth@EU](#). This will require member states to adopt a common format for issuing - and accepting - ePrescriptions and patient summaries, as well as images and image reports, discharge reports and laboratory results. The aim is for 25 member states to be part of MyHealth@EU by 2025.

As an example of how primary data use in the Health Data Space will help patients, Kyriakides gave the example of ePrescriptions. "If a cancer patient has to stay longer than expected away from their country of residence, but is running out of the medicine prescribed by their oncologist, he or she will be able to take the prescription to a pharmacy in the country they find themselves in, and get the vital medication needed."

Even within a single health system, improving the ability of patients to access and share their data is expected to ensure more efficient delivery of care, for instance by reducing duplication of tests, saving patients time and the health system money. It will also assist patients to monitor their own condition and better manage their self-care.

Researchers, innovators, industry, as well as public institutions will be able to request access to health data through the Data Space

Turning to the provisions for secondary data use set out in the proposal for the Health Data Space, Raffaelli told delegates to the European Cancer Summit that it: "sets out a common EU framework on the use of health data for research, innovation, public health, policymaking, regulatory activities and personalised medicines." Researchers, innovators, industry, as well as public institutions will be able to request access to health data through the Data Space. Access will be controlled at a national level, with each member state being tasked with setting up a health data access body.

According to the current proposals, the data can be accessed and processed only in closed secure environments, with downloading prohibited for data that have not been anonymised. To access data, users require a permit, and only the data necessary for their research is accessible in the secure processing environment for the duration of the permitted project. Researchers, companies or public institutions can only access non-identifiable data, i.e., data that offers information about disease, symptoms and medication, but that doesn't reveal the identity of individuals. Users are also prohibited from re-identifying who the data came from. [According to the European Commission](#), "It will be forbidden to use the data to take decisions detrimental to individuals, or to market health products towards health professionals or patients."

Promoting research, protecting rights

With the proposal now up for scrutiny and amendment by the European Parliament and Council, one aspect that will be closely looked at is the tension between protection of patients' privacy and the level of control over how their data is used on the one hand, and the goal of giving greater access for research and innovation on the other.

It's a tricky balance to strike, says Bettina Ryll, leading melanoma patient advocate and member of the first EU Cancer Mission board. "Cancer patients hope for options that increase their chances of survival and their quality of life; this includes anything from better management of disease, access to treatments and trials as well as research."

Yet the vulnerability that puts them in such need of novel treatments is also a reason to ensure that their rights and interests are well protected in the way the Health Data Space operates, Ryll adds. "Many cancer patients, including some patient organisations, will say: just take my data and do whatever you need if it only increases my chances to survive," she says. "As carer, I felt like that. However, as a patient advocate, I feel differently - just because people like us are desperate does not give others the right to exploit that desperation, so we need research *and* protection of the individual. It is not right that, for instance, children of a cancer patient get punished for the desperation of their parents."

That said, she supports the aspiration to find safe ways to allow researchers to make greater use of the huge amount of patient data that is out there. "We cannot on one side ask for products and services that perfectly address patient need, but on the other side not want our data to be used for commercial development," she argues. She gives the example of apps developed to scan moles for signs of malignancy, where the underlying algorithms are poor, as they are trained on insufficient data sets. "So if we want better mole mapping apps we need better training sets - and these sets have to come from *somewhere*. So there are commercial applications that patients want, while obviously being protected against malicious attempts." Governance will be crucial to ensure the level of protection and trust that will be needed for the system to function well, she adds.

"We cannot ask for products that perfectly address patient need, but then not want our data to be used for commercial development"

Convincing patients to share their data also puts an onus on everyone using patients' data to show how the European Health Data Space could benefit patients. "Altruism is no business model, and cancer patients have many unaddressed needs," says Ryll. "If we started caring about truly addressing these needs rather than using moral argumentation, we would move faster. Let's face it, everyone has a clearly stated - usually financial - interest in the European Health Data Space. And then the patients are supposed to hand over their data for the greater good of all?"

That's not to say that access should be given purely to research questions of immediate importance to patients, she adds. "I absolutely would like to see a mechanism by which patients can drive the research they want to see, but not at the exclusion of everything else... I believe that society has valid interests in e.g., optimising its healthcare resources to provide the best possible care for as many people as possible - this is something that many patients will not have at the top of their minds, but how do we know what care to provide if we cannot measure it?"

Ryll also emphasises recognising and rewarding those who contribute and maintain high-quality data. "I very much like the approach Estonia took in their [genome project](#)," she says. "They had a proactive strategy in place, earning the trust of the population to share their data, rather than trying to talk people into it - and it worked. I find this a much more respectful way towards citizens as well." Effective sanctions against abuse are also important, she adds, "Having ways to prosecute those who misuse data would increase the confidence in the system."

Informing cancer policy

Speaking from the perspective of healthcare providers and researchers, Mark Lawler, Professor of Digital Health at Queen's University Belfast, emphasises that co-creation of secondary use projects with patients would be key for achieving impact. As an example, he cites [DATA-CAN](#), the UK's Health Data Research Hub for Cancer, where Lawler is Scientific Director. "Over the last three years, we have embedded patients within all our practices, also from a governance point of view. Patients are members of our steering committee, our management committee, each project committee and have power of veto. We have also worked closely with the European Cancer Organisation to give a Europe-wide resonance to our work. This has allowed us to do work with significant impact; for example, we highlighted that approximately a million cancer diagnoses may have been missed in Europe during the Covid-19 pandemic."

Lawler emphasises that working together with patients empowers patients, as it means answering questions that are relevant to them. "We also find we produce better research – the research that comes out of that work is much more impactful, because it has a wider perspective. It is critically important to not only to use data intelligence to highlight the challenges, but also to implement the results of the data, because there is significant patient relevance associated with these data."

He argues that patients who contribute data for research should receive feedback on what happened with their data, how it was used. "Patients should be informed, in general terms, what the results of research are that use their data. This also allows researchers to gain patient trust in data sharing. And of course, patients, as data donors, should also have access to their own data and be able to decide how their data is used most effectively."

"In relation to cancer, that can mean more accurate and earlier diagnosis, better treatment, improved patient safety"

Lawler has big hopes for what the proposed European Health Data Space will be able to deliver for cancer patients, cancer professionals and cancer research. "In relation to cancer, that can mean more accurate and earlier diagnosis, better treatment, improved patient safety. From the health service point of view, it will also lead to improved healthcare efficiencies. "It will allow us to look at developing policy that can be driven by the use of data. So not only can we have better data that's relevant to patients, we can also have a better health policy that is more data informed, with greater opportunities for research and innovation and most importantly implementation."

Rapid availability of data will be important to have an impact on policy, adds Lawler, who again cites DATA-CAN as an example. "We produced data on the impact of Covid-19 and national lockdowns very rapidly, made it available to the community, and contacted the chief medical officers in the four UK nations to let them know the impact of Covid on cancer services and cancer patients. That allowed for rapid decisions to be made in relation to reopening cancer services, for example."

To strike the balance between rapid availability and complete data, Lawler suggests, for example, that registries make a smaller number of parameters available quickly, with a more detailed deep-dive into the data taking longer. "If we have data that's two years old, that's not going to help you address what's happening now. You need to have access to near real-time data to make appropriate decisions based on what the data are telling you."

Control v access

Like Ryll, Lawler stresses the need to strike the right balance between governance that protects the privacy and rights of patients and promoting access to allow the data to be put to legitimate and productive use. It's a regulatory area that created widespread consternation within the medical research community when the general data protection regulation was introduced in 2016, though its impact has varied according to how member states have chosen to implement the regulation, and the general principle of patients maintaining tight control over access to their data [has some support among patient advocates](#).

The European Cancer Organisation has flagged up the need for further reflection about how to achieve the right balance within the framework of the European Health Data Space. Speaking at the 2022 European Cancer Summit, Wim Oyen, head of Nuclear Medicine at the Royal Marsden Hospital in London, and co-chair of the European Cancer Organisation's Digital Health network, highlighted issues of data security, trust, respect for privacy, and transparency regarding the use of and access to health data as key areas where the organisation will be seeking amendments to the European Health Data Space proposal, as it goes through scrutiny by the European Parliament and European Council.

In its [submission to the consultation](#), the European Cancer Organisation said, "There is a careful balance to be met in respect to ensuring good governance and protection of data privacy, whilst also not overregulating the cancer research environment. Many stakeholders have publicly raised the difficulties that GDPR implementation has created for cancer research in Europe as an example. Any additional rules on conditions for access to health data for research, innovation, policy making, and regulatory decisions needs to be formed on a clear evidence-base and carefully constructed to achieve proportionality in approach."

It also pointed out that "better application of existing rules/regulations" can be more effective than creating new rules and regulations.

"There is a careful balance to be met in ensuring good governance, whilst also not overregulating the cancer research environment"

The European Cancer Organisation is campaigning for the European Commission to commission a study on the impact of the GDPR on cancer research in Europe, so the lessons can be applied in formulating the final version of the European Health Data Space.

One ingenious approach that could help strike a good balance between control and access is a model of data sharing that was developed in the Netherlands, in the context of a project to implement a structure for federated data sharing within the country. Dubbed the 'Personal Health Train', the model flips the traditional approach whereby researchers apply for access data sets to be sent to them, for them to interrogate, explains André Dekker, Professor of Clinical Data Science at Maastricht UMC+, who played a leading role in its development. "Rather than sending all the data that's obtained in hospitals around the world to the researcher, can we send the research questions to where the data is?"

The advantage of the Personal Health Train model is not just avoiding all the security issues involved

in transferring data sets around the internet and having multiple copies stored on multiple systems. It also means that the people and institutions governing the data base retain control over how it is used. Each 'station', or data source, on the Personal Health Train can apply its own rules on what a 'train' or research question/analysis can do with the data. These rules could reflect the terms of the consent given by the patients whose data is requested for the project. Such a set-up could be crucial in helping to overcome barriers of trust and control that have frustrated efforts to interrogate pooled data sets, for so long.

“Rather than send all the data that’s obtained around the world to the researcher, can we send the research questions to where the data is?”

This approach to accessing data may also change how researchers approach a project. Rather than investigating the real world impact of a new technology by asking what information can be gained, for instance, from 200 patients treated at one’s own hospital, researchers could investigate 200 patients in different hospitals. “This is really doable with the Personal Health Train,” says Dekker.

The implications could be quite far reaching. For one thing, that sort of data would make it quicker and easier to assess the real world value of new technologies, which is important both for clinical decision making and for making best use of health budgets. The data would be more robust – because of being measured across many different treatment settings – and would also be generated much faster because of the wider patient base.

Secondly, measuring impact across different hospitals and even different countries could highlight potential disparities in the impact of the technologies in question, and facilitate research and investigation of potential issues, and about the most effective way to introduce the technology into the clinic.

Though the model is as yet largely untested, at least at a supranational level, it does make intuitive sense, and seems to have been well received among both patient and research communities. It is also a great illustration of how success in achieving the objectives of the European Health Data Space rests primarily on stakeholders finding creative ways to build the confidence needed to collaborate over shared objectives, rather than being subjected to new regulations and laws.