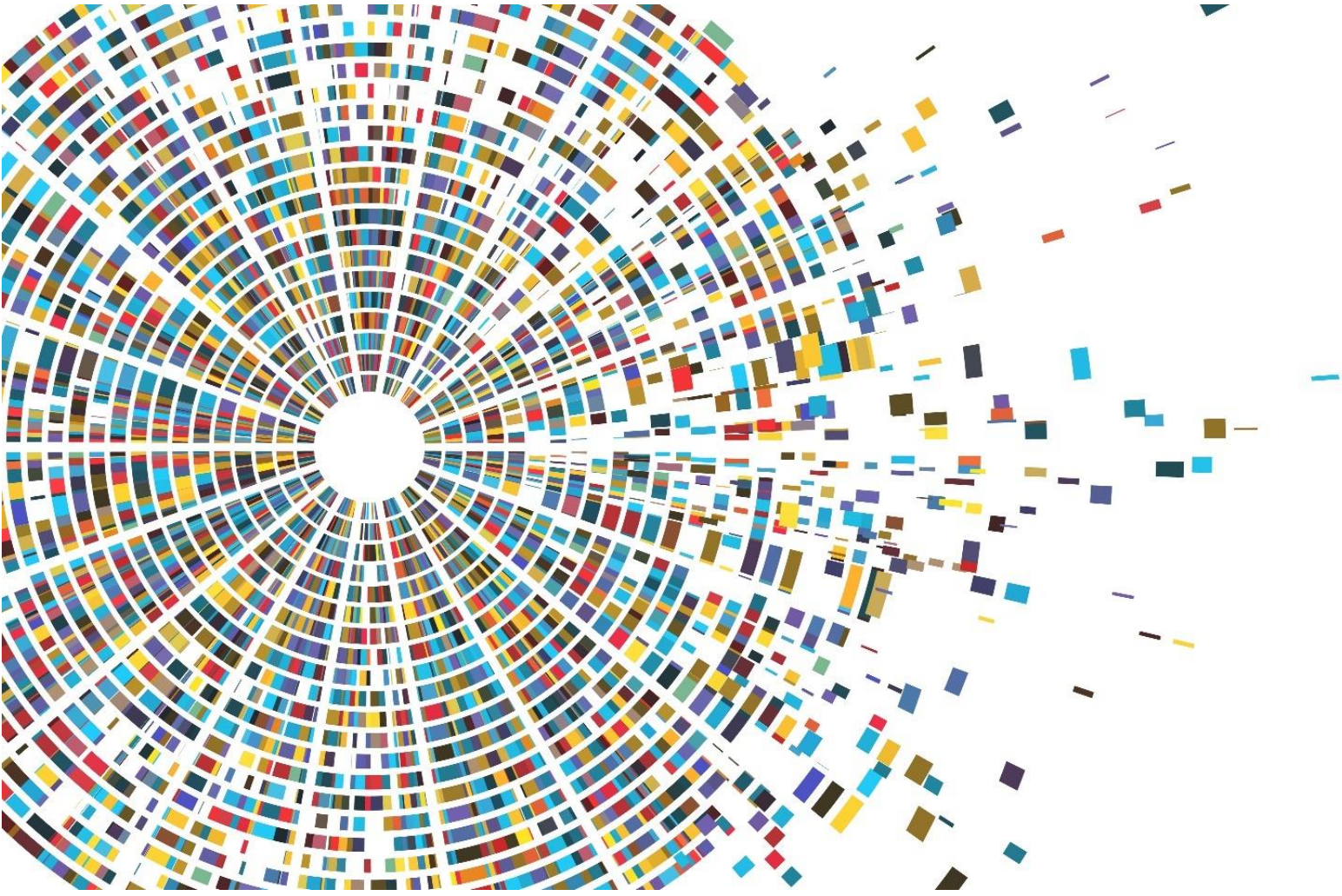
Public Policy Projects
Public Policy Institute

State of the Globe: Bringing the benefits of genome sequencing to the world

Roundtable Four Report

The Democratisation of Genomic Data



Public Policy Projects,
28 Queen Street
London, EC4R 1BB

Tel: 0207 839 9305
publicpolicyprojects.com

Registered in England & Wales: 05401243

This is the fourth roundtable in our series and so far the sessions have covered the general overall landscape (which identified key issues about equity of access and return), specific issues of the utilisation of genome sequencing in relation to pandemic preparedness and the last roundtable sought to address some of the issues relating to the legal and regulatory framework in which we are operating and role of cloud based technology to store the data sets that are currently being created at scale.

The fourth roundtable sought to explore several key themes:

1. What do we mean by this data set – what data should be collected and how should it be structured to create a data set that will derive maximum benefit, both in a clinical and a research setting, in the longer term?
2. Building on the themes of equity of access and return, how can these data sets be used to bring the socio-economic benefits of genomic medicine to as many people globally as possible, ensuring that those who can make the data available share in the benefits of the research that is carried out on that data?
3. How to consent data that is contributed and the potential challenges of not getting that right from the outset.

Genomic Data

Genomic data in and of itself has limited utility. We start unlocking its value both for clinical practice and research when it is combined with other forms of biomedical data including our health records. And to have maximum benefit we need to have as close to real time health data as possible. Rightly or wrongly, however, people are very protective about genomic data, and we need to be aware of that.

If we want to have genomic data but use it for another purpose than that for which it was originally used, we need to have the information that is required for the next person or recipient to be able to determine its fitness for use for the next purpose. And to determine fitness for use you need to know some quality things – what are the features of the data or the provenance of it. Part of the problem, however, is that the providers of the raw data have not bought into the idea that they are part of the food chain of the data and the data they are producing needs to be in a format that is standardised across technologies as a package of interoperable health messaging information.

If we could get across the whole food chain and all the vendors buy into this need for interoperable data and really strive to get that in a format that is much more consumable by the health system, then we will have the information we need alongside the genomic data so that we can determine its fitness for use. Basic things like what transcript used, what genome used etc.

GA4GH provides part of the solution here, but not all the solution. There is ongoing work by the Medical Genome Initiative that is trying to create standards for medical genomes, for example. The revolution we need is in learning health systems – this is bigger than genomics. Every episode of care should be a potential to improve care for the next patient. Or that patient next time. GA4GH is focused on interoperability in relation to genomics research and increasingly works with ISO and HL7 which relate to wider health informatics accreditation.

If we do not get technical interoperability between GA4GH standards and FHIR then even if people want to share data and it is permissioned for sharing and there are huge benefits to sharing, it will not be able to be shared from a tech perspective.

Therefore, designing these systems with technical interoperability in mind so that when people have the opportunity and incentive and permission to share is where we want to get to. As opposed to – we have some great opportunities to learn but we cannot because the data is not reusable.

We need to bring the FAIR principles – Findable, Actionable, Interoperable and Reusable – to these systems. There are some health systems where patients are consented in, and they get their initial treatment knowing there is a secondary use for research. But that needs to be clear from the outset and is not the case everywhere.

In a sense once we have solved this agreement on what is an interoperable system – what is genomic data, how should it be stored and shared and in what format and to what depth – this may be the easiest part of what we are trying to achieve for health and for research and future health in that it is static. A lot of the other information we are going to want to start layering on top will be most useful if we can capture and assess it dynamically.

If we can get the public mindset, the government understanding, the systems, the funding and the trust and transparency issues resolved the technology will help us add in the other things.

Benefits

From a UK perspective, making commercial gain from a publicly funded dataset does not go hand in hand with the NHS. But is that an equally applicable point of view in other jurisdictions?

As the Genomics Thailand project is in its early stage there has not been much utilisation of data much from the public health system in terms of commercialisation yet. There are some private hospital systems that use data for either internal projects or further collaboration with other private companies, for example insurance schemes. However, it is anticipated that Genomics Thailand will see more and more private companies or biopharma wanting to access the data. Looking at the data itself Genomics Thailand consider 3 uses – for the clinical treatment of the patient; for public health use (and consent allows access to the data for this purpose); and the most challenging, private use with private hospitals and commercialisation. This is an area where support is required.

When we talk about benefit, we need to define what we mean by benefit and who is going to benefit. That depends on who is giving or sharing the data and whose data that is. There is also a controversial area where people by default opt into something - how do they benefit? Are they on board with it?

Broadly in terms of benefits, this is not a genomics question – this is a broader societal question. How should benefits be returned to a country which has used taxpayer funded resources to generate a dataset that has been accessed by organisations who have derived a benefit? Is that not a government level kind of decision? If it is not possible for individual organisations to force pharmaceutical companies to ensure that the products, they create are reasonably priced or that they are designing clinical trials well and incorporating different populations, regulation should step in.

Transparency and Consent

It is essential to be upfront with people about what their data is going to be used for and it is essential that organisations do only that which they have said with the data. People should also have the ability to change their opinion – what someone decides at 18 might not be what they decide at 28 or 50. However, it is crucial that it is clear that if someone withdraws their consent that means that their

data cannot be used in the future but the research that has already been done with their data cannot be retracted.

There are examples of surveys carried out on populations – in Australia and Qatar and through the "Your DNA, Your Say Survey" carried out by Professor Anna Middleton. In the examples from Australia and Qatar it was clear that monitoring public opinion is an important facet of understanding how to build public trust.

In Qatar, the research evidence demonstrates that repeatedly there is a very favourable outcome when the public is surveyed (up to 90 per cent) for sharing of data to support Qatar Genome or the Qatar Biobank for the good of science. Continued monitoring of public opinion is important to help shape regulations, guidelines, and policy – and importantly if such surveys serve to identify that people who donate their samples are quite willing for them to be shared, are scientists and policy makers perhaps sometimes over-protective as data custodians?

The experience in Australia would also suggest an initial strong degree of support from patients at a basic level but more concerns are identified as one delves deeper into the issues – there are concerns about data being shared with life insurance companies, about what the data is going to get used for and who will get access to it. And there is further concern about sharing of data outside of Australia where there is concern that standards may not be as high – ethically and legally – as within Australia.

More broadly concerns are also raised about sharing of data with pharmaceutical companies and, in some cases, governments. Consistently concerns are raised with respect to whether data can be identified. And so, in general terms, whilst there is a general willingness to share data this is eroded the more questions that are asked.

Can there be a technological solution to some of this – a modular approach which draws on the benefit of looking outside the sphere in which we are working and look at how these problems – or similar problems – have been solved elsewhere? For example, there are examples of how people interact with their credit scores or credit reports where there have been several drives to make this more accessible to people. They can see who has accessed their data, when did they access it and what they accessed it for. Or examples of how to opt in and out of being a blood donor in the UK where there is an app for that. Or even how Cancer Research can, transparently, show where donations went and which study they helped.

If that level of transparency has been solved in other areas, how can we pick out those pieces and make them applicable in this context? Building on these examples can a modular approach be developed to support consent and patients being able to have really good visibility over their own data and where it is being used – the ability to opt in and the ability to opt out and perhaps even individual toggles against particular studies or institutions. Technology can be the enabler here.

Within Genomics Thailand, which is at the beginning of their journey, the data sharing policy, compliant with GDPR and PPPA, provides that the person who provides the data has total ownership and Genomics Thailand is granted permission to use that data. It was acknowledged that the way that the consent is written, and how participants are communicated with right at the outset, would be a huge advantage in democratising the data and data sharing policy. The intention was that getting this right at the outset would make it easier to do further research and share the data with other international groups in the future. But there is a balance between being a data custodian and trying to make sure participants don't get exploited and making sure that the data custodian does not become an obstacle to the utilisation of the data for the greater good by becoming too strict and not allowing science to advance because there are too many restrictions on usage.

In Qatar they started with a broad consent and Qatar Biobank acts as the custodian of the data. The participants are the owner, but they relinquish rights to the Biobank for future uses of the data. Participants are also asked whether the Biobank can do genetic analysis and access their medical records, and this is their choice whether to agree to this or not.

Qatar has a law which means they are aligned with GDPR for the protection of personal data and whilst international data sharing is permitted in accordance with that regime, from the outset the specific consent stated that samples would not be send abroad. With the advent of specialist sample analysis which is not present in Qatar itself this means that specialist collaborators cannot analyse samples outside Qatar.

In Australia, where there is a devolved health system, they have to centralise all of their data which is clinically generated, and this has forced greater thinking about how genomic data is used. Firstly, has the data been clinically generated or generated through research. Secondly, is it being used for clinical purposes or for research.

Using this 2x2 table of the main 4 different areas where data gets used leads on to thinking how it can be managed. A pure research environment can be constructed and managed. Clinically generated data exists in an environment with more legislative and regulatory parts around it and that is where the levers are and that is the environment which is going to have more widespread impact in the longer term. Alongside getting governance of data right in the clinical setting it is really important to think about patients needing to provide dynamic consent, fully understanding where their data is going and what it is being used for and we need to be having more conversations about the way patients' data is being used for ongoing clinical use.

There is an interaction between consent for large sample stores of databases and the legal aspects of the debate with respect to the standards GDPR may be setting for consent for processing data and two do not marry up very well. In the EU, the standard is not specific to health sector processing and requires quite a high level of informed specific consent. There is potentially some flexibility around genomic research and medical research, but the position is not completely clear and there is almost an implication that some sort of a route consent process is required as plans develop.

Another layer of complexity is the extent to which GDPR has an impact in other jurisdictions, for example, where a patient is being treated in a country outside the EU. How should contradictory laws be managed? There can be no doubt that the extra-territorial reach of GDPR is potentially huge and there is a sub-group at GA4GH focusing on GDPR. Guidance is required for science and genomic research.

There are multiple different consent models in play here. We are talking about different forms of broad consent; we can have specific consent and dynamic consent. What is key is that this is just one aspect of governance and certain trade-offs may demand you may be able to have broad consent as long as you have rules around specific decision making you have feed in from patients and participants. Consent is but one part of the equation.

Could we utilise this group to do some sort of global pilot around a piece of technology that we could test out how patients react to having those kinds of choices for different settings around the world?