Genomics England and *The Turing Way*: a path to open science in a sensitive data landscape

Stuart Gillespie ⊙   Augusto Rendon ⊙   Maxine Mackintosh ⊙
Natalie Banner ⊙   Raphael Sonabend ⊙   Alexandra Araujo Alvarez ⊙
Kirstie Whitaker ⊙   Malvika Sharan ⊙   Vicky Hellon ⊙

11 December 2023

What does 'working in the open' mean to you?

Is it sharing your code, your methods or your data? Is it a state of mind? It may be any of those things at any one time – and, in an organisation like Genomics England (GEL) that deals with sensitive and personal data, it may not be without its challenges.

But by taking a broad interpretation of openness, and by following best practices set out in guides like The Alan Turing Institute-supported *The Turing Way* handbook, even a privacy-conscious company such as GEL can reap the benefits of open science.

That's the view of Dr Raphael Sonabend, an Expert in Residence of The Turing Way Practitioners Hub's first cohort and a previous member of GEL's Diverse Data initiative, an initiative that aims to reduce health inequalities and improve patient outcomes in genomic medicine for underrepresented communities. They believe that open access resources and open source tools can help researchers and practitioners around the world make sure that genomic medicine equitably benefits all people.

"Part of my role at GEL involved talking to people about how open science and its associated best practices can be beneficial to their work," says Raphael. With previous work experience in data science and technology management at GEL, Imperial College London and the Wellcome Trust, Raphael brings a cross-sectoral perspective in open source. "There's a real interest at all levels at GEL, both in open ways of working and in acquiring the skills needed to do that in the most effective way possible. People are open to 'working open'."

The benefits of open working

"Public-facing organisations have a duty to be transparent and to work openly wherever possible," says Raphael. "But beyond that, there is good evidence for the benefits of working in that way." Those benefits can include raised standards internally: developers may take more care over their work in the knowledge their code will be scrutinised externally. And, by the same token, that external scrutiny can pick up any bugs that do remain, as well as enabling code to be improved upon more broadly.

As alluded to earlier, the term 'openness' can be applied to a broad spectrum of activities in an organisation: examples of open working at GEL can range from simple actions like making calendars visible for others and using shared documents to be open for input, to large-scale engagement and communication such as crowdsourcing translated consent forms and publishing findings in quick-turnaround blog posts ahead of longer-term formal publication in scientific journals.

"Openness is more than just publishing your code," says Dr Maxine Mackintosh, Programme Lead for GEL's Diverse Data initiative. Maxine says: "There are some great examples internally at GEL. My colleague Dr Sam Tallman did an analysis of data from the 100,000 Genomes Project to find out if there were differences in clinical outcomes or diagnoses by ancestry. We published the findings as an interim technical blog post rather than waiting for a scientific paper to be accepted. We ended up receiving lots of inbound enquiries about this research and offers of collaboration, so the approach was hugely beneficial."

Another open-adjacent example is a toolkit to support individuals working in diversity and genomics to be more aware of the language they use. We co-designed with people with lived experience, clinicians and geneticists and is currently in a public sandbox. We hope that in its current form, it supports people to use more precise and appropriate language, and also inspires others to identify and help define more terms that we often use in this area.

"More generally," adds Maxine, "I believe that adopting a more open approach will improve the quality of science conducted on GEL's data and the trust of those communities we aim to serve, and that in many cases it will be riskier not to take this approach."

Adopting best practice

The Diverse Data team is exploring the potential of open source in genomics research and practice and understanding how projects dealing with sensitive data and on sensitive topics (like eugenics) can effectively adapt to accepted best practices. A project called link23 was piloted under Diverse Data in 2022 and 2023 to create, find and showcase community developed tools that could help researchers promote and use more equitable practices in genomics research. One of the go-to resources for the team is *The Turing Way*, which provides a working example for how to engage volunteer members in co-creating community resources they can use. As part of this team, Raphael was exploring ways for researchers, developers and data scientists to use open source tools and share them publicly in platforms like link23. They also encouraged the use of '*The Turing Way* handbook in helping plan and support the implementation of more open projects – taking considerations of the nuts and bolts of code quality, documentation and licence types to bigger-picture topics such as ethics, community-building and collaboration.

Dr Augusto Rendon is Chief Bioinformatician at GEL. "For the bioinformatics teams I work with at GEL," he says, "best practice means starting with a well-defined problem, knowing what a good outcome looks like, and making sure that the solution is reproducible and that if people come back to the project in a year's time they'll still be able to understand it. Best practice is also about our overall standards at GEL: we're a clinically accredited organisation, and all of our tools have to be well-built, validated and extensively documented. That's the most important thing."

The challenges of open working

"One of our limiting factors to making things open at GEL is time," says Augusto. "There's no lack of willingness, but sometimes the realities and priorities on the ground make it challenging. There is also a limit to what we can publish openly: while code and tools can be shared, the data has to be protected."

Other potential barriers to openness in many organisations include unease over market competition, security concerns and culture change. "But there are ways of mitigating these things," says Raphael, "and of reassuring people that the risks and burdens may not be as high as they think."

Maxine adds: "I think we have a gap where we could do more to work with GEL's external research community on subjects like reproducibility and making things open, to help get the most value from our data asset. That's one of the areas and audiences where *The Turing Way* could be beneficial."

PanelApp: an open source GEL success story

The best way to convince people to work more openly, says Raphael, is to come armed with examples that demonstrate its benefits. At GEL, the best example is PanelApp.

Conceived in 2015, PanelApp collates gene panels – collections of genes grouped together because of their association with a particular condition – from the NHS National Genomic Test Directory and the 100,000 Genomes Project. It allows researchers and clinicians to access the latest consensus around genetic causality, aiding patient diagnosis. PanelApp is entirely open source and benefits from an engaged global community of expert contributors who constantly review and update the evidence around particular genes and gene panels.

"PanelApp has proved to be very successful internationally, and a lot of people around the world now use it as their main reference point," says Augusto. "Back in 2015 there was little consensus in the community about the pathogenicity of particular genes. Seeking one person's opinion wouldn't be enough, and endlessly consulting a formal panel of experts would be too slow, so we created this crowdsourcing approach. It allows the community to keep the database up to date themselves, although the project still needs to be managed and maintained. PanelApp also has the benefit of sitting separately from our other systems, which makes it easier for us to publish openly, decoupled from the things we may need to keep closed."

Exploring the ethics of openness

As a government owned organisation, GEL also grapples with challenges and limitations when generalising open practices in the context of healthcare. Dr Natalie Banner is GEL's Director of Ethics, responsible for helping the organisation navigate the tricky ethical and regulatory waters of cutting-edge genomic medicine and sensitive patient data.

"There are essentially two aspects to what I do," says Natalie. "One involves the compliance element of GEL's activities, and the other is what I call the 'wayfinding' aspect. We're working with technology that's advancing at a rapid pace and may have controversies attached to it. There are questions that law and regulation can't answer, so I think of ethics as trying to do the best for

our participants within that uncertainty and trying to find innovative ways of thinking about the benefits, risks, opportunities and challenges of our technology."

In the context of open science and open source, that means striking a balance between safeguarding participants' genomic data and maximising the research potential of that data. And when it comes to open tools and resources, developers need to think carefully about whether they could be used in unforeseen ways that exacerbate rather than alleviate existing inequalities.

Natalie adds: "I don't think we have a sophisticated enough understanding yet of what's possible in terms of open science and open source – partly because of the focus on stringency around access to data and the need to protect it. There is an unavoidable 'closedness' to what we do, but I think we could really reap the benefits of working openly if we had a more nuanced view of what it means. That could be greater use of open source code, better showcasing of our huge amounts of available data, fostering collaborations through our Genomics England Research Network (GERN) or new programmes, like Diverse Data – from my perspective – more transparency for participants around how data could potentially be used by researchers. Of course, all of that needs to be done while fully respecting the privacy and confidentiality of our participants.

"But in this field there is no alternative to open science: researchers are often looking for the proverbial needle in a haystack, and therefore we need to work collaboratively on an international scale."

Open Source in Genomic Medicine: a disruptive force for good

Ultimately, says Maxine, Diverse Data's aim can be greatly enhanced by promoting open source and citizen involvement for improving equity in genomic medicine. Proof-of-concepting link23 demonstrated the real-world need for adopting open source in genomics research through the provision of helpful tools, handbooks and other resources. When applied effectively, open science practices are likely to increase the diversity of both the users and the beneficiaries of these genomic tools. "Open science and equity in genomics are closely related," she says. "Open ways of working can lower the barrier to entry, encourage collaboration at all levels – particularly internationally – and broaden the reach of tools and resources by making them modifiable and more widely applicable. My hope is that people at GEL will engage with open source, as trailed by link23, and find them fun and valuable additions to their day-to-day work."

Reproducibility is a key aspect of data science best practice, and open source could have an important role to play in this area. Raphael says: "For instance, projects like *The Turing Way* and link23 engage different stakeholders and provide a collection of useful open source tools and practices to enable reproducibility and reusability worldwide." As open source becomes a common approach in health research, Raphael says "Hopefully, people within GEL will be able to make use of those in their own work to save time and at zero cost. Open source doesn't just mean doing the hard work and putting it out there – you can also benefit from things other people have already made, which is a big selling point." Augusto agrees: "Our teams will be developing tools for Diverse Data and they should also benefit from things like catalogues of accepted best practice and knowing what tools are already available and how to use them. That will save having to start from scratch every time and wasting resources. In the end, despite all the technology, it's about people talking to people and

empowering and helping each other."

"I think open source in genomic medicine could be quite disruptive," adds Natalie. "It's a different way of thinking, and the explicit focus on equity and creativity means it's not about that academic paper mindset." Alluding to the new possibilities that open source practices in genomic medicine bring, she says "I'm excited to see where it goes."

Key takeaways

- Open working takes on diverse meanings for different individuals and teams. At GEL, this concept ranges from collaborative documentation within teams to open communication via publicly accessible blogs, as well as leveraging large-scale open source software development.

- Organisations like GEL, dealing with sensitive data and topics, face unique challenges in implementing open practices. A nuanced approach to open science helps promote the use of open source software, standard data practices, collaborative efforts, and transparent reporting, all while safeguarding participant privacy and data confidentiality.

- Institutionally supported open source projects, like PanelApp, ensure the long-term use and sustainability of resources. Similar investment can extend the benefits and opportunities for future users by fostering collaborative development and maintenance of broadly applicable solutions.

- The Diverse Data initiative at GEL, in its endeavours to reduce health inequalities and improve patient outcomes in genomic medicine for minoritised communities, is adopting and raising awareness of open tools, practices, and resources.

- By actively adopting best practices and collaborating with open science projects like *The Turing Way*, GEL's Diverse Data is striving to reduce barriers to participation for patients and the public in genomic medicine.

Important resources

- Diverse Data Initiative of Genomics England

- link23 Sandbox - Genomic tools that work for everyone

- Mind the Gap: Stories of Health Data Equity

Acknowledgements

Cite this publication