data sets can be useful? How can they be managed? What's the best way to win the confidence of public and regulators? And, crucially, is academia training enough mathematicians and medical-data scientists, who will have to develop and harness all this new potential? The last of these questions at least has a clear answer: no.

Big data sets in medicine include genomics, transcriptomics and proteomics (which, respectively, describe our genomes, identify which of our genes are being expressed and catalogue all of the proteins in a tissue sample). Genomic data sets alone have already shown their value. The presence or absence of a particular gene variant can put people in high- or low-risk groups for various diseases and identify in some cases which people with cancer are likely to respond to certain drugs.

But a single molecular data set will not contain sufficient information to tell the whole story of an individual's medical fate. Integration of different types of molecular data might tell more, but that remains a computational challenge. Even more would emerge if an individual's molecular data were placed in the context of their physiology, behaviour and health. Electronic health records, the numbers of which are skyrocketing, could be useful here. So could disease registries, hospital and health-insurance records, as well as research publications and clinical-trial data. New types of data set come from wearables and apps, which collect health data directly from individuals, and from genome sequences of volunteers. Some researchers are even trying to extract medically relevant data from social-media platforms such as Twitter.

This adds up to a mind-boggling volume of data. According to one estimate at the workshop, clinical data from a single individual will generate 0.4 terabytes of information per lifetime, genomics data around 6 terabytes and additional data, 1,100 terabytes. By 2020, the amount of health-related data gathered in total will double every 73 days. Health professionals will confront more data than do those in finance.

To collect and hold all of this information within strict privacy regulations — non-negotiable for medical data — is another challenge. Some tech firms, such as IBM Watson Health and Hewlett-Packard, are building systems that keep data local — algorithms can dip into them, but the data are not transmitted anywhere else. Google, unsurprisingly, thinks that all these data are safer on the Cloud.

The big question for scientists is how to take the next step to convert these artistic sketches of potential into scientific knowledge. Data sets vary in their reliability. Those derived from skimming social media will be very messy and will have to prove their usefulness. And large data sets, reliable or not, inevitably throw up spurious correlations, and so recognizing meaningful patterns requires deep understanding of biology, something that software developers don't generally have. Future biologists — funders and universities should note — will need much more training in mathematics and data science.

*"Health professionals will confront more data than do those in finance."*

Big data has the exciting potential to allow clinical trials to be conducted partly *in silico* — which would mean using fewer animals in drug testing as well as recruiting fewer patients to the actual trials. Yet the field is developing at a time when public trust in experts is at an all-time low.

Regulators are ready to be persuaded to accept computational information in clinical trials. On occasion, the European Medicines Agency and the US Food and Drug Administration already accept it in pharmacokinetics, one of the simplest data sets that drug developers must supply agencies with. But the trust of doctors, patients and regulators in the abstract informatics and mathematical know-how that go into developing scientific and clinical predictions cannot be taken for granted. As medics and researchers nudge the field gradually from poetry and seduction to delivery, they must engage the next generation of scientists — and the public — and inform and educate them in the art and science of what is possible. ■

# Source of hope

*More funding for the Middle East X-ray factory could give it a cafe — and sow the seeds of peace.*

Even as the United States bombed Vietnam and the Red Army quashed an uprising in Czechoslovakia, there was a place where citizens from the two cold-war superpowers could meet: the CERN cafeteria. To this day, the same spot at Europe's particle-physics laboratory near Geneva, Switzerland, is a rare venue for Pakistani and Indian physicists to have a coffee, or for engineers from Israel and Iran to share ideas — no doubt fuelled by the huge lunches and inspiring views.

There is no better cure for cross-border acrimony than meeting someone from the other country and realizing that they are not so different from yourself. And major science facilities, whose inhabitants already share many beliefs, such as in openness and the scientific method, are natural places to kindle such mutual understanding.

What a chance, then, for SESAME. Synchrotron-light for Experimental Science and Applications in the Middle East is a venture that brings together Bahrain, Cyprus, Egypt, Iran, Israel, Jordan, Pakistan, the Palestinian Authority and Turkey, to build the region's first synchrotron-light facility near Amman in Jordan. Proposed in 1997, modelled on CERN and created under the auspices of the United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2002, the remarkable collaboration is now set to open (see page 475).

SESAME has struggled at every turn. Several member states have made only partial payments, and its opening has been delayed by more than five years owing to financial difficulties. Other hurdles were less predictable: 2009 and 2010 saw Iranian council members assassinated, and in 2013, the facility's roof fell in after unprecedented snowfall.

But SESAME has also ridden on a wave of generosity and perseverance. Its director and various advisers contribute their time for free. Facilities in Germany, France, the United Kingdom, Sweden, Spain and Switzerland have donated crucial parts of the machine, and Jordan, which provided land for the site at no cost, has continued to cough up funding, even as the neighbouring Syrian war has brought huge numbers of refugees across its borders.

To get this far is a terrific success. But when it opens next year, the facility will have just two beamlines, rather than a planned four, and will be without an administration building, dormitory, major library, auditorium or visitor centre. Nor, sadly, will SESAME have its hub of scientific life, a cafeteria. Experimental equipment takes priority, but in a facility where time is so precious that researchers allotted 24 hours of beam time barely sleep, some worry that, with nowhere to grab a coffee and a seat, scientists from conflicting countries may pass like ships in the night.

A group from the Sharing Knowledge Foundation in Geneva is trying to raise €30,000 (US$32,000) through public donations by the end of this year towards creating a SESAME cafeteria. But funds for both the cafe and the science could also come from another source, which has been conspicuously lacking over the years: the United States.

A Congress ruling in 2011 cut funding to UNESCO, after the organization formally recognized Palestine as a member state. This is a public-relations difficulty for SESAME, even though there should be no official block because it is independent from UNESCO. A bigger question could be whether the incoming US president meddles with the Iran nuclear deal, which could restore sanctions.

Still, brokering peace in the Middle East is something that Donald Trump said on the campaign trail he'd "really like to do". For less than the cost of a return flight to Tel Aviv on Air Force One, his administration could help SESAME to develop from a bare-bones scientific laboratory into a true beacon of peace and cooperation. ■