World view

By Paul Nurse

Biology must generate ideas as well as data

Data should be a means to knowledge, not an end in themselves.

Accepting a Nobel prize nearly two decades ago, my old friend Sydney Brenner had a warning for biology. "We are drowning in a sea of data and starving for knowledge," he said. That admonishment, from one of the founders of molecular biology, who established the nematode worm *Caenorhabditis elegans* as a model organism, is even more relevant to biology today.

Rather often, I go to a research talk and feel drowned in data. Some speakers seem to think they must unleash a tsunami of data if they are to be taken seriously. The framing is neglected, along with why the data are being collected; what hypotheses are being tested; what ideas are emerging. Researchers seem reluctant to come to biological conclusions or present new ideas. The same occurs in written publications. It is as if speculation about what the data might mean and the discussion of ideas are not quite 'proper'.

I have a different view: description and data collection are necessary but insufficient. Ideas, even tentative ones, are also needed, along with the recognition that ideas will change as facts and arguments accumulate.

Why are researchers holding back on ideas? Perhaps they are worried about proposing an idea that turns out to be wrong, because that might damage their chances of getting promotion or funding. But as Charles Darwin put it: "False facts are highly injurious to the progress of science, for they often endure long; but false views, if supported by some evidence, do little harm, for everyone takes a salutary pleasure in proving their falseness; and when this is done, one path towards error is closed and the road to truth is often at the same time opened." To wit, it's important to get the facts right, but new ideas are useful, as long as they are based on reasonable evidence and are amenable to correction.

Don't get me wrong. We need data produced from new technologies to advance understanding. The importance of 'hypothesis-free research' is well established: the philosopher Francis Bacon proposed it as part of his 'empirical method' in 1620. In his book *Novum Organum*, he argued that the first step in establishing scientific truth should be the description of facts through systematic observations. But this is only the first step. For example, it would have been rather a pity if Darwin had stopped thinking after he had described the shapes and sizes of finch beaks, and had not gone on to propose the idea of evolution by natural selection.

The next step is to extract knowledge from the data.

> **"It would have been a pity if Darwin had stopped thinking after describing the shapes and sizes of finch beaks."**

Paul Nurse is former president of the Royal Society and director of the Francis Crick Institute in London.
e-mail: paul.nurse@crick.ac.uk

To refocus on that goal, we must improve our working processes, placing a greater emphasis on theory and shifting our research culture.

How? Embed engineers and experimentalists who are developing new technologies and methods deeply into the biological problems. It is through deep familiarity with the biology – not simply a drive to collect more and more data – that important questions will be asked. Such questions will sustain the investigators' passion to keep probing data until patterns and knowledge emerge, and will also influence the data that are gathered.

There are other necessary steps. Develop appropriate analytical tools, including programs for data mining and machine learning. Make certain that data are usable, properly annotated and openly shared. Model the molecular and cellular components involved in a biological phenomenon, to allow analysis of dynamic behaviours and interactions. Sometimes just writing down the equations without seeking solutions can be helpful, simply because it imposes greater rigour on model building.

More theory is needed. My exemplars for this include the evolutionary biologists Bill Hamilton and John Maynard Smith, and the geneticists Barbara McClintock and Francis Crick. Their papers are permeated with richly informed biological intuition, which makes them a delight to read. This sort of thinking will accelerate a shift from description to knowledge. Theorists can find fertile ground in considering the flow of information through living systems, which can help them to make better sense of the flood of biological data.

Seeking to be led by theory and knowledge will probably require shifts in research culture. Theorizing should be encouraged, and theories should be included in experimental papers to put data in context. Attempts to do this should not be dismissed by editorial and funding processes as idle speculation. As Darwin said, it allows ideas to be attacked and either dismissed or modified. A sort of 'tyranny of the field' sometimes inhibits the generation of explanations different from the current consensus, but this is a mistake. If the new ideas are not satisfactory, then they will soon be eliminated and progress will be made.

False facts should not be tolerated, but journals and research funders should be open to reasonable new ideas and interpretations, particularly if they differ from the current consensus. Evaluation committees should be tolerant when some of the ideas of people they are considering for promotion or funding are shown to be incorrect.

Such approaches will advance not only research, but also teaching. Students will be better motivated and will feel more inspired if they are taught that biology has ideas – and that they are worth talking about.