

A two-way street to science's future

To view the relationship between computing and science as a one-way street is mostly untrue today, argues Ian Foster, and will be even less true by 2020.



A growing number of sciences, from atmospheric modelling to genomics, would not exist in their current form if it were not for computers. A simplistic analysis of this relationship focuses on hardware, and sees

science as largely a passive beneficiary of the computing industry's relentless innovation, acquiring and applying to its own ends the fastest computers, largest disks and most capable sequencing machines. In this view, science and computing (as an intellectual discipline) have little to say to each other: it is the computer industry that drives the advances that have an impact on science.

A more sophisticated narrative says that science is increasingly about information: its collection, organization and transformation. And if we view computer science as "the systematic study of algorithmic processes that describe and transform information"¹, then computing underpins science in a far more fundamental way. One can argue, as has George Djorgovski, that "applied computer science is now playing the role which mathematics did from the seventeenth through the twentieth centuries: providing an orderly, formal framework and exploratory apparatus for other sciences."²

Information overload

Why this shift? Of course, there is more information, as technology allows us to collect, store and share vastly more data than before. Equally significant is that science is becoming less reductionist and more integrative, as researchers attempt to study the collective behaviour of larger systems. To quote Richard Dawkins: "If you want to understand life, don't think about vibrant, throbbing gels and oozes, think about information technology."³

Such system-level approaches are emerging in fields as diverse as biology, climate and seismology. A frequent goal is to develop high-fidelity computer simulations as tools for studying system-level behaviour. Computer science, as the 'science of complexity', has much to say about how such simulations — which can be considered a new class of experimental apparatus — should be constructed, and how their output should be analysed and compared with experiment⁴. Similarly, information theory provides formidable insight



The sciences rely on computers, but the benefits are two-way; each is driving the other forwards.

into how biological systems encode, transform and transmit information.

Both the data deluge and system-level science demand computing technology in all its forms — hardware, software, algorithms and theory. The growing importance of computing has several implications for the science of 2020, of which I explore three here.

First, the scientist of 2020 will be adept in computing: not only will they know how to program, but they will have a solid grounding in, for example, the principles and techniques by which information is managed; the possibilities and limitations of numerical simulation; and the concepts and tools by which large software systems are constructed, tested and evolved. This knowledge has been picked up on the job by many pioneering scientists and will hopefully be instilled in the next generation by more formal training. The idea that you can be a competent scientist without such training will soon seem as odd as the notion that you need not have a solid grounding in seventeenth-century mathematics (such as algebra).

Fruitful partnerships

Second, successful science collaborations of 2020 will include computer scientists as key members. All scientists will be adept at applying existing computational techniques, but they will also understand that progress in their fields will require innovation in computing technology. So they will work with computer scientists to identify computational problems, much as today's experimentalists and theorists understand the strengths and weaknesses of their favoured methods and know to partner with others when new techniques are needed. Indeed, this fruitful interplay is

already occurring: for example, the communication challenges inherent in far-flung physics collaborations inspired the development of the World Wide Web, and the need for efficient indexing of terabytes of digital astronomy data has spurred new approaches to organizing spatial data in relational databases⁵. Elsewhere, the scientific opportunities that arise from using wireless sensor networks for, say, continuous habitat monitoring are driving innovations in network protocols and algorithms.

Third, the scientific disciplines and institutions of 2020 will need to train, attract and reward researchers whose focus is on producing the computing innovations required for science to advance: what we might term 'applied computing'. Thus, we see new organizational structures, such as the Computation Institute at the University of Chicago and Argonne National Laboratory (for which I work) and Harvard's Institute for Innovative Computing, that aim to bridge the distinct concerns of computer science and other sciences. Academic departments are hiring faculty with strong computational inclinations. National laboratories have established computational directorates. It will be interesting to see which of these interdisciplinary structures work best.

The growing importance of applied computing also has implications for computer science. Indeed, just as during the early days of the sciences, scientific concerns drove mathematics forward (think of the origins of the calculus), so the many challenging problems posed by modern science can help to focus and motivate research in computing. In my view, it is no accident that some of the most vibrant areas in computing today are those tightly coupled to scientific problems. These dynamics occur in fields as diverse as sensor networks, data integration and grid computing. It's a two-way street, and always has been.

Ian Foster is director of the Computation Institute at the University of Chicago and Argonne National Laboratory, Illinois, USA.

Acknowledgements To Anne Rogers for her comments.

- Denning, P. et al. *Commun. ACM* **32**, 9–23 (1989).
- Djorgovski, G. S. *Virtual Astronomy, Information Technology, and the New Scientific Methodology*. 7th Int. Workshop Comput. Arch. for Machine Percept. 125–132 (2005).
- Dawkins, R. *The Blind Watchmaker* (Norton, 1986).
- Cohen, J. *PLoS Biol.* **12**, 2017–2023 (2005).
- Szalay, A. S. et al. *SIGMOD Rec.* **29**, 451–462 (2000).