The Collaboratory Opportunity

Wm. A. Wulf

Although computers have already had a significant impact on the conduct of science, there is now a special opportunity to use them to further leverage the entire scientific enterprise. A report from the National Research Council (NRC) points the way (*1*). Entitled "National Collaboratories: Applying Information Technology for Scientific Research," the report was produced by the NRC's Computer Science and Telecommunications Board.

The word "collaboratory" is a blend of the words collaboration and laboratory, and the NRC report explores the ways in which information technology can be used to support a much broader range of the activities in scientific inquiry. The unprecedented reduction in the cost-performance ratio of computing and communication, and the melding of these technologies into an information infrastructure, enables applications beyond the traditional analysis and modeling familiar to the scientific community. Important as these applications are, they are but one aspect of the process by which science proceeds.

As I wrote in 1989, the collaboratory is "a center without walls, in which the nation's researchers can perform their research without regard to physical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, [and] accessing information in digital libraries" (*2*). The physical infrastructure of the collaboratory is the worldwide collection of networked computers augmented by instrumentation interfaced to the network (such as that at Sondre Stromfiord in Greenland). The essence of the collaboratory, however, is not this physical infrastructure. Rather, it is the software that enables scholars to use remote libraries, collaborate with remote colleagues, interact with remote instruments, analyze data and test models—all with nearly the facility they now enjoy locally.

As prior use of computing has demonstrated, fuller exploitation of the power of information technology can not only increase the productivity of our researchers, it can qualitatively change the kinds of questions we ask and, hence, what we know about nature—indeed it has and it will. We cannot predict precisely how the information infrastructure will be used. However, just as computer-based modeling and visualization now allow us to test hypotheses about systems we could not otherwise control, such as galactic collisions, networked instruments and shared data should allow us to perform new kinds of interdisciplinary experiments.

The bottleneck to the achievement of such a vision is not hardware. A great deal of research needs to be done by computer scientists and engineers, working with client researchers in other disciplines, to build effective collaboratories. Some of this research is being done in other contexts, but it is neither coordinated nor, to my mind, given the priority it deserves relative to the immense payoff. Consider just a few examples:

1) *Scientific databases.* A common theme of many disciplines recently is huge databases; examples include the genome project, global seismic data, and space-based Earth measurements. The attributes of these data are generally the antitheses of those of the business data for which contemporary databases have been built: The content may be neither text nor numeric, may represent a temporal rather than instantaneous state, and may be effectively read-only, and individual records may be enormous. But perhaps the biggest difference is in the way that the data are used.

It is a truism in computing that when the parameters of a problem change by two orders of magnitude in any dimension, it becomes a new problem. In this case, there is change of many orders of magnitude along several dimensions simultaneously. Despite this, disciplinary scientists are mostly struggling with database technology designed for employee records and automated teller machines.

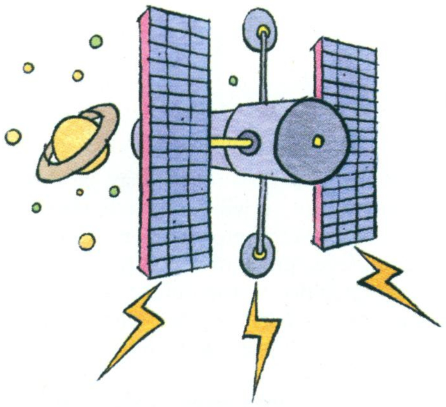
2) *Remote instruments.* Increasingly, instrumentation is remote from the investigator because the phenomenon to be measured is remote (as in deep space) or in a hostile environment (as in the deep ocean) or the instrument is simply too expensive to replicate (as a large telescope).

The more remote in both space and time, the greater the advantage in building autonomy (intelligence) into the instrument. The instruments not only must cope with their local environment (as in a planetary rover) but must be able to make smart decisions about the amount and kind of data to collect, store, and transmit. A deep-space probe, for example, will have limited storage and bandwidth for information to be beamed back to Earth, forcing long delays between instructions from the investigator; it will have to face possibly unforeseen observational opportunities (for example, a supernova or an unpredicted dark object flyby) and make real-time decisions about how to use its resources.

3) *Collaboration technology.* Much of the work on collaboration technology ("groupware") has either been targeted to business users or back to the computer science community. Yet, as most researchers will attest, the single most critical resource for innovation is access to stimulating colleagues, particularly those in different disciplines.

Especially in the case of interdisciplinary efforts, the probability that exactly the right collection of expertise will be collocated is small. Saying it another way, if collaboration between noncollocated scientists were made easy, the probability that interdisciplinary problems would be successfully attacked would increase. But we

The author is in the Department of Computer Science, University of Virginia, Charlottesville, VA 22901.



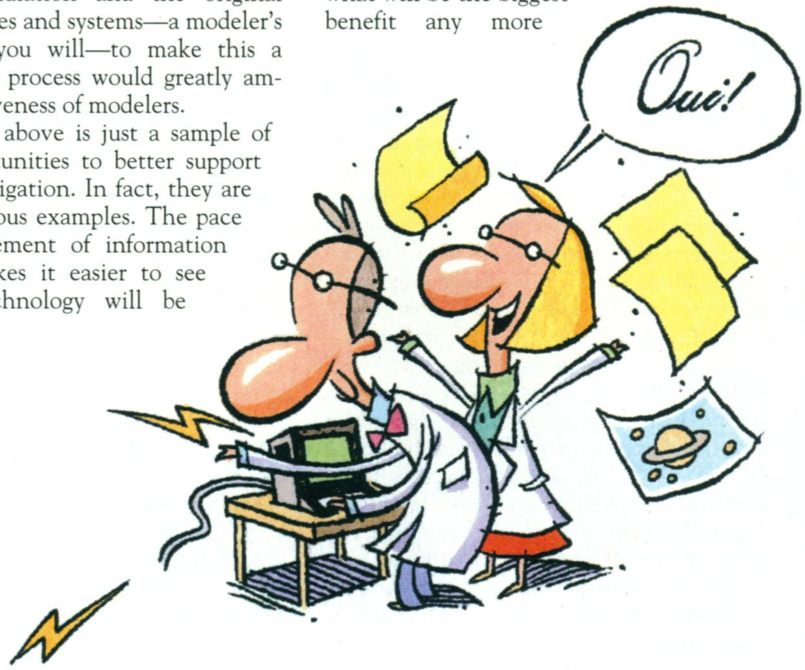
have a great deal to learn about how to support collaboration among researchers. Although tools that support things like joint authoring and brainstorming may apply to all disciplines, the real and perceived needs of disciplines will differ.

4) *Data fusion.* The mere existence of large amounts of on-line data will encourage the analysis of phenomena observed by multiple sensors in multiple locations—whether the use of the data in that way was planned or not. This requires that the data not only be accessible and manipulable, but understood. Techniques need to be developed for data description that facilitates use by disparate applications and enables integration from heterogeneous sources in a form that leads to an understanding of the overall problem.

5) *The modeler's workbench.* Currently there is a gap between the development of a mathematical model and the creation of the computer program to simulate it. That discontinuity is a potential source of error and an opportunity for a divergence between the simulation and the original model. Languages and systems—a modeler's workbench if you will—to make this a single, seamless process would greatly amplify the effectiveness of modelers.

Each of the above is just a sample of the rich opportunities to better support scientific investigation. In fact, they are the rather obvious examples. The pace of the improvement of information technology makes it easier to see where the technology will be rather than how it will be used. The tendency is to assume that the applications will be the same as the present ones, but better. Although that is partially true, history suggests that new, unanticipated applications will also emerge—and somehow, these are usually the important ones.

Consider, for example, that within a few years, the typical personal computer will have graphic capability comparable with that of today's highest priced graphic workstation. Does that mean merely a crisper desktop? Certainly that's one implication, but I doubt that will be the major benefit. However, I cannot predict what will be the biggest benefit any more than I predicted the success of electronic mail when the ARPAnet was installed in the late 1960s.



Alas, although research served as the rationale for the initial development of information technology, it is a minuscule market. We should not expect the private sector alone to fully exploit the opportunity to use the technology to leverage research. The recommendations in the NRC report point the way, and although tiny by "big science" standards, I believe they have the potential for an immense positive impact and should have the highest priority.

ILLUSTRATIONS BY TERRY E. SMITH

References

1. National Research Council, "National Collaboratories: Applying Information Technology for Scientific Research" (National Academy Press, Washington, DC, 1993).
2. W. A. Wulf, "The National Collaboratory—A White Paper," Appendix A in "Towards a National Collaboratory," an unpublished report of a National Science Foundation invitational workshop held at Rockefeller University, New York, 17 to 18 March 1989.