Computational science

An Internet introduction

by Anna Keller Gold

From the Web pages of professional organizations[1] to job advertisements in scientific journals[2] to the pages of the *New York Times*,[3] there is evidence that, more and more, sciences are becoming computational. Computational science has become "an equal and indispensable partner, along with theory and experiment, in the advance of scientific knowledge and engineering practice."[4]

It is an important part of biology and medicine, chemistry, engineering, mathematics, physics, astronomy and space science, earth and environmental science, linguistics, and every other field in which data exist in massive quantities, and research problems exist that are "otherwise too big, too small, too expensive, too scarce, or too inaccessible to study."[5]

The subject of computational science is diffuse, and no single portal exists that brings together its many facets. Thus the goal of this guide is to offer a sampling of Web sites that introduce the subject or relate to the tools used in computational science, its diverse applications, and the variety of people and organizations involved in computational science instruction and research.

Computational science, informatics, and library professionals

Librarians in research libraries are continually faced with new disciplines and research programs styled as "informatics"; bioinformatics and medical informatics are among the most familiar of these.

Along with "computational science," "informatics" is a term used to describe the science of combining the use of computational techniques (algorithms), tools (specialized computers with various visualization systems, including the Web and databases, data warehousing, and other software), and mathematics (models from wavelets to partial differential equations to quantum lattices and cellular automata) in the analysis of scientific data.

Developments in computational science will place new demands on the information management skills of researchers and librarians alike and will continue to offer challenging opportunities for collaborations between scientists and information professionals.

Starting points, tutorials, and online courses

• **SIAM Working Group on CSE Education.** This site begins with an excellent definition of computational science as a distinct discipline, describes the research areas in which computational science plays an important role, and discusses graduate curricula and programs in CSE (Computa-

About the author

Anna Keller Gold is head of the Science and Engineering Library at the University of California, San Diego, e-mail: agold@ucsd.edu

tional Science and Engineering). *Access:* http://www.siam.org/cse/report.htm.

• **Partnership for Advanced Computational Infrastructure (PACI) Education, Outreach and Training.** This extensive site is devoted to education, outreach, and training. It includes online lesson plans, courses, tutorials, and educational simulations. *Access:* http://www.eot.org/.

• **Computational Science Education Project (CSEP).** Aimed at advanced undergraduates and higher-level students in science and engineering, this site includes an online textbook and tutorials for freely available networking and visualization software. *Access:* http://csep1.phy.ornl.gov/csep.html.

Hardware and software

Doing computational science requires expertise in a scientific discipline, mathematical modeling, and knowledge of computer science. A researcher also needs access to specialized hardware and software, from hardware arrangements in clusters, like AppleSeed and Beowulf, to Cray-like supercomputers to enormous distributed "virtual" machines or computing "grids."

Work on this hardware in turn commonly entails using the parallel computing algorithms that enable fast, high-volume processing of very large amounts of data, and software that allows scientists to develop models and simulations and to visualize research results. Together with the needed hardware, these computational tools provide opportunities to organize very large amounts of data in ways that both express and invite interpretation and understanding.

• **AppleSeed.** AppleSeed is all about "personal parallel computing" on a cluster of computers running Mac operating systems. *Access:* http://exodus.physics.ucla.edu/appleseed/appleseed.html.

• **Parallel Virtual Machine (PVM).** Thorough technical site for this free software package that allows a heterogeneous collection of Unix or NT computers (or both), connected by a network, to act as a single large parallel computer. *Access:* http://www.epm.ornl.gov/pvm/pvm_home.html.

• **IEEE Task Force on Cluster Computing.** Lots of organizational information here, but dig deeper for excellent links to newsletters, conferences, educational materials, e-print repositories in cluster computing (click on "CC archive"), and other parallel computing information portals, such as Beowulf. *Access:* http://www.ieeetfcc.org/.

• **Parallel COMputing Portal (PCOMP).** A peer-reviewed archive of essential hyperlinks for high-performance parallel computing. The site also includes lists of HPC newsletters and journals, meetings, tutorials, conferences, software, and benchmarks. *Access:* http://www.npaci.edu/PCOMP/.

• **Parallel Tools Consortium.** This site brings together tool users, developers, and researchers from academia, government, and industry to improve the usability and availability of parallel tools. *Access:* http://www.ptools.org/.

• **Parascope.** Hosted by IEEE, Parascope includes a comprehensive list of supercomputing and parallel computing centers, journals and conferences, HPC vendors, and federal agencies involved in HPC. Also lists the top 500 supercomputer sites and the "World's Most Powerful Computing Sites." *Access:* http://computer.org/parascope/.

• **Tera Grid (Distributed Terascale Facility).** A joint effort of supercomputing centers in Illinois and California, the Distributed Terascale Facility will offer more than 13.6 teraflops (trillions of calculations per second) of computing power and will manage and store more than 450 terabytes (trillion of bytes) of data. *Access:* http://www.npaci.edu/teragrid/ and http://www.teragrid.org/.

• **Terascale Computing System (TCS).** Nicknamed "Lemieux" (in honor of the Pittsburgh Penguins star), this research grid, capable of six teraflops per second, is funded by the National Science Foundation and housed at the Pittsburgh Supercomputing Center. *Access:* http://www.psc.edu/machines/tcs/status/.

Disciplinary portals and gateways

Portals in computational science, as in other endeavors, serve as gateways to information. In computational science, "portal" also may mean an integrated, remotely accessed platform or "scientific workbench widely accessible through networking and optimized for a particular set of problems or a problem-solving paradigm."[6] Portals are therefore focused on a research domain or discipline and provide a place where researchers can deposit, access, and manipulate data relevant to their work.

Biology and medicine

• **Biology Workbench.** From this site biologists can search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modeling tools, all within a point and click interface. *Access:* http://workbench.sdsc.edu/.

• **Biomedical Informatics Research Network (BIRN).** Originally the "Brain Imaging Research Network," this huge national initiative has been broadened to encompass biomedical informatics generally. It will allow researchers nationwide to share high-resolution animal and human brain images to enable analysis and comparison at many different scales. *Access:* http://birn.ncrr.nih.gov/.

• **Computational Neuroscience.** This is a comprehensive, annotated information portal maintained by Jim Perlewitz. There are links to modeling and simulation software, major laboratories, researchers, conferences, education, and funding for theoretical neurobiology. *Access:* http://home.earthlink.net/~perlewitz/.

• **National Center for Biotechnology Information (NCBI).** Since 1988, NCBI has been a crucial national resource for molecular biology information, producing public databases, conducting research, and developing software tools for analyzing genome data. This site includes links to molecular and genomic databases, literature, books, software, and tutorials. *Access:* http://www.ncbi.nlm.nih.gov/.

• **Protein Data Bank (PDB).** This is the single worldwide repository for macromolecular structure data, operated by University of California San Diego and the Research Collaboratory for Structural Bioinformatics. *Access:* http://www.rcsb.org/pdb/.

Chemistry

• **Australian Computational Chemistry via the Internet Project (ACCVIP).** Aimed at developing educational modules in computational chemistry for use in undergraduate and postgraduate teaching, this joint effort by faculty at several Australian universities includes a list of Web links to computational chemistry resources. *Access:* http://www.chem.swin.edu.au/chem_ref.html.

• **ChemViz.** A chemistry visualization program, ChemViz uses a Web interface to generate images of atoms, molecules, and atomic orbitals. Resources are also available for teachers to take advantage of ChemViz in secondary schools. *Access:* http://chemviz.ncsa.uiuc.edu/.

• **GAMESS: General Atomic and Molecular Electronic Structure System.** This site supports computational research in quantum chemistry. *Access:* http://gridport.npaci.edu/GAMESS.

Engineering

• **Computational Fluid Dynamics Online.** A comprehensive information portal maintained since 1994 by Jonas Larsson, with links to academic sites, organizations, newsgroups, projects, software, hardware, jobs, and introductions to the field. *Access:* http://www.cfd-online.com/.

Environment and geosciences

• **CCS Climate Research Community.** An information portal at the Oak Ridge National Laboratory, this site consists of links to research efforts such as the Parallel Climate Model. *Access:* http://www.ccs.ornl.gov/climate.html.

Linguistics

• **ACL NLP/CL Universe.** An information portal of the Association for Computational Linguistics on Natural Language Processing and Computational Linguistics, with links to people, conferences, academic departments, jobs, introductory materials, linguistic corpora,

and more. *Access:* http://perun.si.umich.edu/~radev/u/db/acl/.

Mathematics

• **Netlib.** A repository of mathematical software, papers, and databases relevant to computational mathematics, hosted by the University of Tennessee and Oak Ridge National Laboratories. *Access:* http://www.netlib.org/index.html.

Physics and astronomy

• **GriPhyN: Grid Physics Network.** The GriPhyN (Grid Physics Network) collaboration plans to implement the first Petabyte-scale[7] computational environment for data intensive science in the 21st century; it is funded by the National Science Foundation. *Access:* http://www.griphyn.org/.

• **Particle Physics Data Grid.** A collaboration of U.S. federal laboratories and several U.S. universities, enabling a worldwide distributed computing model of current and future high-energy and nuclear physics experiments. *Access:* http://www.ppdg.net/.

• **Digital Sky Survey.** This site is an effort to federate data from multiple large-scale sky surveys to provide a perspective of the universe that is statistical and data-focused, in contrast to traditional work with individual stellar objects. *Access:* http://www.cacr.caltech.edu/SDA/digital_sky.html.

Who does computational science?

U.S. federal laboratories affiliated with the Department of Energy have long had a critical role in developing and using high-performance computing capabilities. The National Science Foundation has also invested in large-scale partnerships that make supercomputing facilities available for academic scientific research. Private industry, scholarly associations, and educational programs and university research institutes also have key roles in carrying out computational science, as do rare individuals like David and Gregory Chudnovsky, who built their computational mathematics hardware and software from scratch.[8]

Government agencies

• **Partnerships for Advanced Computational Infrastructure (PACI).** Funded by the National Science Foundation, PACI is the umbrella under which two major national computational partnerships operate: NPACI and the Alliance. PACI partners develop, apply, and test the software, tools, and algorithms to create a national grid of interconnected, high-performance computing systems. *Access:* http://www.paci.org/.

• **National Partnership for Advanced Computational Infrastructure (NPACI).** Based in San Diego at the San Diego Supercomputer Center (http://www.sdsc.edu), NPACI publishes an excellent newsletter (*Online*), holds workshops, provides research information, and develops resources developed to enable computational science. *Access:* http://www.npaci.edu/.

• **The Alliance (National Computational Science Alliance).** Based in Illinois at the National Center for Supercomputing Applications, the Alliance is a partnership of more than 50 academic, government, and business organizations. Their site features an excellent newsletter (*Access*) and educational and research events (including Web casts). *Access:* http://www.ncsa.uiuc.edu/.[9]

Federal agencies and research centers

• **Federal National Laboratories.** Many federal national laboratories operate significant computational science programs. This site provides a comprehensive directory by laboratory name to all U.S. federal labs. *Access:* http://www.federallabs.org/laboratories_national.html.

• **National Energy Research Scientific Computing Center.** Funded by the Department of Energy, and home of the "world's largest unclassified supercomputer," this center supports research in astrophysics, climate research, and materials science. News, technical

help, publications, and more. *Access:* http://www.nersc.gov/.

• **NASA Advanced Supercomputing Division (NAS).** Located at NASA Ames Research Center, the mission of NAS is to "develop, demonstrate, and deliver innovative, distributed heterogeneous computing capabilities to enable NASA projects and missions." Also publishes *Gridpoints* magazine. *Access:* http://www.nas.nasa.gov/.

Academic programs, research centers, and consortia

• **Northwest Alliance for Computational Science and Engineering (NACSE).** A coalition of Pacific Northwest institutions and individuals that provides scientists, engineers, and students with Web-based training materials and tools for learning about computational techniques and parallel programming. *Access:* http://www.nacse.org/.

• **SUNY Brockport's Guide to Computational Science and Engineering Programs and Societies.** A comprehensive listing of degree-granting programs in the United States, Canada, and other nations, as well as professional societies active in computational research. *Access:* http://www.cps.brockport.edu/cps-links.html.

Associations and conferences

• **Supercomputing Conference.** Billed as "The International Conference for High Performance Computing and Communication," this annual conference is sponsored in part by IEEE and ACM. *Access:* http://www.supercomp.org/.

• **International Conference on Computational Science (ICCS).** An international, multidisciplinary scientific computing conference aimed at bringing researchers and scientists together with software developers and vendors. The next conference will be held April 2002 in the Netherlands. *Access:* http://www.science.uva.nl/events/ICCS2002/.

Notes

1. SIAM Working Group on CSE Education, "Graduate Education for Computational Science and Engineering," http://www.siam.org/cse/report.htm.

2. See advertisements for positions at Caltech in the August 2001 issue of *Communications of the ACM* and the September 2001 issue of *Physics Today*: "At present, a significant fraction of scientific research at Caltech relies on large-scale distributed computing; scalable algorithms and data structures; the recording, handling, processing, and visualization of large data sets; and the organization of, and access to, large databases. These and related endeavors in applied mathematics and computer science are recognized as an area of research in its own right: Computational Science and Engineering (CSE)."

3. George Johnson, "All Science is Computer Science," March 25, 2001, *New York Times*. ". . . as research on so many fronts is becoming increasingly dependent on computation, all science, it seems, is becoming computer science."

4. SIAM Working Group on CSE Education, op cit.

5. The chair of SUNY Brockport's computational science and engineering department, quoted by Florence Olsen, "Computational Science Spans Barriers Between Traditional Disciplines," *Chronicle of Higher Education*, November 24, 2000.

6. http://www.computer.org/computer/articles/einstein_1299_2.htm. See also Mary Thomas, "Computational Science Portals: The SDSC Grid Portal Toolkit (GridPort)," NPACI Parallel Computing Institute 2000.

7. A petabyte is equal to 1,024 terabytes.

8. "Mountains of Pi," *New Yorker*, March 2, 1992.

9. Logos for the NCSA and the Alliance are reproduced with permission of the National Center for Supercomputing Applications (NCSA), http://www.ncsa.uiuc.edu/. Copyright 2002, University of Illinois Board of Trustees. ■