Grand Research Challenges in Computer Science in Brazil

Claudia Bauzer Medeiros
University of Campinas (Unicamp), Brazil

In May 2006, the Brazilian Computer Society proposed five Grand Research Challenges in Computer Science in Brazil. The society's goal was to foster long-term planning and research in computer science, enhance cooperation with other scientific domains, and provide input to public R&D policymakers in Brazil.

What do the following research topics have in common: analyzing the performance of a wireless network with millions of nodes, forecasting erosion effects of human occupation around a lake, studying a new medication's long-term effects on human metabolism, simulating crowd behavior in an emergency evacuation, and estimating the impact of the flight of a hummingbird in the pollination of flowers in a given area?

Answer: They're examples of possible research proposals within one of the five Grand Research Challenges in Computer Science proposed by the Brazilian Computer Society (SBC) in 2006.[1] This specific challenge concerns computational models of phenomena, whether artificial (the wireless network), natural (pollination), social (crowd behavior), or a result of human interaction with nature (erosion). Model design and implementation are key issues within these challenges.

Why did SBC lead the effort to identify the challenges? It detected the need for novel proactive actions to foster long-term planning and research in computer science in Brazil. Furthermore, SBC hoped to enhance cooperation with other scientific domains and provide input to public R&D policymakers.

How were the challenges identified? They were the result of a two-day Grand Challenge workshop organized by SBC in May 2006, with funding from the State of São Paulo Research Foundation (FAPESP) and the Ministry of Education foundation that promotes the development of graduate and research programs (CAPES).

What are these Grand Challenges? They represent five broad long-term research directions in computer science. They are intimately related with issues whose solution will advance science and have technological, social, and economic impact in Brazil. Table 1 lists the challenges, indicating for each a key research factor involved.

UNDERSTANDING THE HOW: THE GRAND CHALLENGES WORKSHOP

The SBC workshop followed the models of successful Grand Challenges initiatives undertaken in the US by the US National Science Foundation (www.cra.org/grand.challenges) and in the UK (www.ukcrc.org.uk/grand_challenges/index.cfm). A countrywide call for proposals resulted in 47 three-page position papers describing "visions of the future" for computer science research in Brazil for the decade

0018-9162/08/$25.00 © 2008 IEEE

Table 1. The five Grand Challenges and the main associated research dimension.

Challenge	Research factor
Impacts on computer science of the transition from silicon to new technologies	Impact of new architectures
Management of information over massive volumes of distributed multimedia data	Data
Computational modeling of complex systems: artificial, natural, sociocultural, and human–nature interactions	Models
Participative and universal access to knowledge for the Brazilian citizen	People
Technological development with quality: dependable, scalable, and ubiquitous systems	Reliable systems

2006-2016. Researchers in academia and industry from all regions of the country wrote these proposals, which covered a wide range of issues. A committee of five senior computer science researchers selected 21 proposals, using as criteria the proposal's scientific breadth, the vision described, and the authors' academic record.

At the workshop, four distinguished scientists from the philosophy, physics, and environmental and climate-monitoring domains joined the committee members and the researchers whose visions had been selected. The steering committee invited two of these scientists; the Brazilian Academy of Sciences identified the other two. Having these other views and practices extended the discussion of future directions beyond computer science issues.

The workshop was organized in two stages. First, participants briefly presented their position papers. Next, attendees split into six working groups to discuss the proposals and consolidated them into the five challenges. Discussions identified strategies and Brazilian research groups that could work on the challenges. In addition to scientific issues, the working groups also considered the social and economic impact of the challenges for Brazil.

A final session closed the workshop. In this session, participants presented their results to researchers representing Brazilian financing agencies and R&D companies. Scheduling this kind of session within the workshop motivated the participants, from the start, to be objective in their work. It was also useful for getting immediate feedback from decision makers.

ANALYZING THE WHAT: THE CHALLENGES AND ASSOCIATED ISSUES

None of the challenges can be tackled by a single research group. Each involves several open problems, requiring cooperation not only among computer scientists, but also from researchers in other scientific domains. Moreover, the challenges aim to explore how long-term research in computer science can help solve Brazilian problems. Finally, the problems are sufficiently vast to accommodate a wide range of research profiles and interests in computer science.

A complex R&D scenario—agricultural planning and monitoring—illustrates a few of the issues involved in the challenges. Figure 1 presents a high-level view of this scenario.

We chose agriculture as a running example for three reasons:

- relatively few computer scientists are working in this domain, although ample opportunities for exciting research exist;
- outcomes are tangible not only for researchers but also for the general population, affecting food production, distribution, and consumption.
- new results in this field can have enormous impact not only in Brazil, but also in the world.

Agricultural planning involves designing and implementing sophisticated models to simulate the evolution of one or more products in a given region, under different geophysical, social, economic, and cultural constraints. A large set of heterogeneous data feeds these models, such as soil chemical composition, weather information, product characteristics (crops or animals), socioeconomic variables, and historical data on the region's productivity for those products. Sensors provide a large amount of these data—ranging from satellite and aircraft-based sensors to those spatially distributed in ground networks. The models produce several kinds of outputs (such as reports, maps, and images) that experts analyze to decide what to grow, when, where, and how subject to health, legislation, market, and other restrictions.

After the plans are implemented, monitoring involves rerunning and calibrating the models for continuous use of the same kinds of data sources, now at real time, at different sampling time frames. Monitoring can also capture and control a given product's response to natural or artificial changes (for example, in weather or husbandry).

Figure 1 enumerates some of the many actors in this scenario. First, scientists from several domains cooperate in designing and implementing models (bottom left). Farmers must be active participants. Not only are they the plans' immediate end users and implementers, but they provide local expertise—for example, to assign realistic values to model variables, indicate critical factors, or describe work procedures. Second, government agencies and policymakers use model

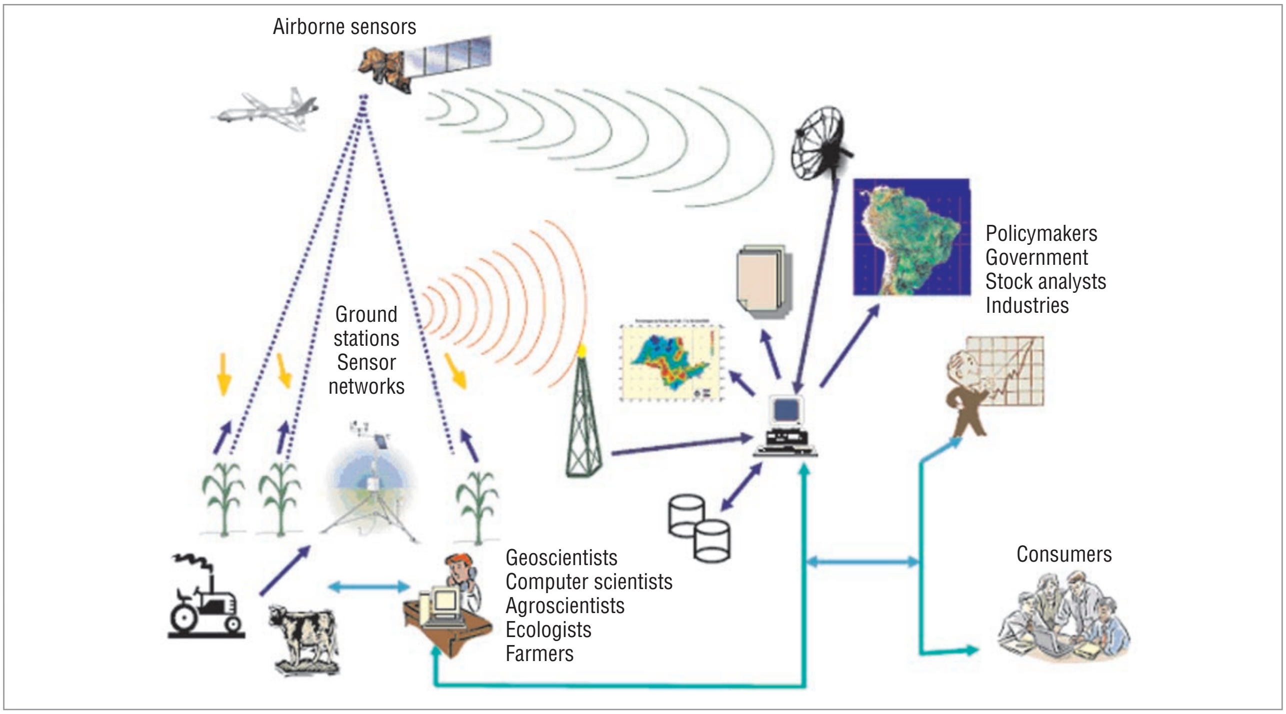

Figure 1. Data, models, people, and hardware involved in agricultural planning and monitoring. Dark blue arrows represent data flow, and light blue connections combine data, control, and human communication processes.

outputs and the plans derived from these outputs to draw long-term directives for a territory's agricultural management, with impact on the stock market (futures and commodities). Moreover, industries will process the products, which will be consumed at the end of the supply chain.

All actors provide feedback to each other, allowing the creation of new models or modification of industrial processes or agricultural practices. The latter might in turn affect the products themselves—for example, by establishing new sanitation controls on crops or animals, or by implanting precision farming procedures to improve soil conditions in specific plots.

Handling new hardware architectures

Challenge 1 concerns the impact that new nonsilicon hardware architectures—such as biological- or quantum-computing-based architectures—will have on how we design and run computational systems. The constant demand for processing power requires more processors on a chip, thus increasing miniaturization. However, more processors generate more heat, and scientists and the hardware industry agree that silicon is reaching its limits—hence the need for novel architectures.

All these technologies are still far from meeting the performance of silicon-based computers.[2] Although they might allow high parallelism and provide large integration capacity, they're unreliable and sensitive to noise. This challenge doesn't center on developing new architectures, but on how they will change computer science—and on how we teach it.

How does the scenario in Figure 1 reflect this challenge? We'll need to reimplement all software and models that run in the environment portrayed so they can run in the nonsilicon systems, using new programming paradigms and data structures. Compilers will have to both cope with chips that have hundreds of heterogeneous processors and compensate for faulty behavior. Satellite-based systems, sensor networks, and other data providers will produce even more data, aggravating the data deluge (see Challenge 2). Moreover, the new architectures will have to coexist with the old silicon systems. Thus, hardware architecture heterogeneity will compound the problem of data heterogeneity. On the positive side, software and models will execute much faster, allowing for more sophisticated model construction and testing, and presumably better planning and monitoring. Challenge 1 therefore involves research in areas such as compilers, languages, software engineering, and parallel computing.

Managing large volumes of heterogeneous data

Challenge 2 concerns managing the huge volume of heterogeneous multimedia data that's part of our daily lives. Data deluge is not a new phenomenon[3]; in the 16th century, savants already complained about the information overload produced by the invention of the printing press. Then, as now, the essential problem was

managing, filtering, and retrieving relevant information from a mass of available data sources. Five hundred years ago, the complaint was that no single person could organize all printed knowledge. Today, a much smaller ambition is to organize one's own data.

In the running example, this means dealing with both static and real-time data on multiple factors: weather, geographic variables, product properties, and market trends. For instance, crop management combines data on temperature, rainfall, and soil geophysical and chemical characteristics, as well as the husbandry procedures of the teams handling the actual planting and harvesting. This data is kept and updated in long historical series—in Brazil, at least 30 years for some crops, such as sugarcane or coffee. Experts must correlate this fine-grained information with data extracted from satellite images. At the end of the supply chain, consumers require information on food composition and quality, creating the need for research in provenance.

Those who access computing systems aren't just consumers, but partners in generating new information.

Issues to investigate thus include data fusion, interpretation, annotation, indexing, storage, querying, mining, cleaning, summarization, and provenance. Open problems require work in fields such as databases, software engineering, image processing, and artificial intelligence.

This challenge's emphasis on multimedia suggests that a large percentage of problems are still tackled at the textual data level, especially when it comes to Web mining. Fields such as content-based image retrieval or sensor network management must engender more cooperation within computer science—for example, involving experts in image processing, databases, and computer networks. Content engineering, video annotation, and personalized visual search are other activities that would benefit from work on this challenge, presenting a good opportunity for interaction with the arts. Still other associated topics are ensuring quality, privacy, security, and copyright management.

Developing models

Challenge 3 concerns developing the models—that is, designing, specifying, evaluating, modifying, composing, executing, managing, and exploring computational models for a wide variety of domains, dealing with hundreds of variables and data types. This challenge provides a large spectrum of examples of the benefits of collaborative research among computer scientists and experts from other areas. Interdomain cooperation always imposes additional learning tasks—of vocabularies, practices, and ways of understanding the world. End-user expertise must always be considered when testing and validating the models, which often complicates this process.

Figure 1 emphasizes the complexity of model construction for this domain. In addition to the variables already mentioned, the product's purpose also influences agricultural models. For instance, different uses of sugarcane require distinct models—whether it's to be planted to produce sugar, alcohol, or biodiesel. Associated issues are transportation and distribution logistics, involving work on combinatorics and optimization.

Computer scientists have been working for some time on computational models involving artificial interactions such as simulations of networks, natural interactions (such as photosynthesis), or even human-nature interactions (such as erosion or biodiversity studies). A less-explored aspect of this challenge concerns modeling sociocultural interactions, such as in virtual communities. This would open many research opportunities for collaboration with researchers in the social sciences, arts, and humanities. This kind of cooperation is still scarce in Brazil, and researchers might, for instance, leverage it to model interactions within the Orkut social-networking service (www.orkut.com), thus helping us to understand why Brazil houses more than 70 percent of all Orkut users in the world.

Furthermore, we need computing research not only on the algorithmic specification and implementation of the models, but also on the complex networked platforms to run them (such as grids), result materialization (and thus computer graphics and scientific visualization), and new methods for requirement analysis and testing of model runs (hence software engineering). Modeling the grid itself (an artificial object) also falls within this challenge. Because developers can adapt and reuse models, storing and maintaining them is also of interest—an example of an intersection with Challenge 2, because models can be conceived as another kind of complex data.

Providing universal and participative access

Challenge 4 concerns people—more specifically, each Brazilian citizen. The challenge statement refers to "participative access," which embeds two intermingled concepts:

- *Participation in knowledge generation.* Those who access computing systems aren't just consumers, but partners in generating new information.
- *Accessibility to computing systems.* We must account for people's distinct perspectives of the world, different social and economic living conditions, and possible physical or mental handicaps.

What kinds of participative access mechanisms should be considered? Figure 1 shows a few examples of the variety of participants and profiles that must be catered to. An agricultural monitoring system must, for instance, let agroscientists interpret collected data and market analysts project commodities trends. It must also let farmers make real-time decisions, such as last-minute modification of a harvesting schedule. This scenario involves scientists with distinct expertise, politicians, economists, engineers, and consumers (and thus the country's entire population). The figure stresses the two-way information flow between people and between people and systems (the participative aspect).

Accessibility to information systems is a multilayered concept related to usability. Although often associated with making computer systems available to users with disabilities, accessibility also covers educational, affective, gender, age, social, and economic factors. Accessibility isn't a recent problem—stained glass was used in medieval churches as a means to make the Bible accessible to the mostly illiterate European populations.

Accessibility to information systems is a multilayered concept related to usability.

Researchers in computer-human interactions are prime participants in Challenge 4. Indeed, issues to consider involve a wide range of concerns in interface studies—from choosing graphical elements in a screen and the associated semiotics, to interaction models and their implementation.

Work at the design level involves teams of computer scientists, graphical designers, psychologists, sociologists, and linguists, to name but a few. For effective implementation, computer scientists might need to combine their efforts with those of signal-processing researchers, mechanical engineers, physicists, or electrical engineers. Accessibility research also requires extensive end-user involvement and investigation into new testing procedures.

Research challenges aren't limited to interface issues. For instance, work in participation and accessibility requires visual content analysis using social interaction and community behavior models (and thus Challenge 2). Researchers can also mine cultural interactions from multimedia data, and hence tackle problems posed by Challenge 3.

Researchers must also consider these efforts in component and device design—such as processors that can meet the interaction requirements, or sensors to capture sound inflections or gesture modalities. For instance, voice interaction, an important issue for the visually impaired, requires more processing power to perform adequate physical signal processing, but also to run new generations of pattern-recognition algorithms. More processing power eventually sends us to deal with Challenge 1.

This challenge's emphasis on the Brazilian citizen is a consequence of the country's large socioeconomic, regional, and cultural differences. As the workshop report points out,[1] according to the 2003 census, 14 percent of Brazilians suffer from some kind of physical or mental disability—an overwhelming 26 million people. Another problem is heterogeneity in education—accessibility and participation should also consider the millions of functional illiterate, defined as individuals more than 15 years old with fewer than four years of schooling. The country's size and heterogeneous communication infrastructure aggravate this scenario. Some regions are adequately served by high-speed networks, whereas others rely at best on satellite links—how else would it be possible to provide access to the networked world to communities living in the Amazon area? Thus, the fourth challenge also prompts research in novel and less expensive computer networks and telecommunications infrastructure.

In addition, Brazilians consider digital inclusion important in fostering the country's development, and they intensely debate how to attain this inclusion. However, many decision makers are more concerned with the physical infrastructure, which is obviously a prerequisite. Challenge 4 emphasizes the harder and less visible issue—digital inclusion is about participation in knowledge production, and people must have access to this knowledge.

Ensuring ubiquity

Challenge 5 refers to the fact that our lives increasingly depend on ubiquitous computer systems. These systems must be dependable, robust, and secure. Furthermore, they must be scalable and available 24 hours a day. This list of desirable properties received a new name during the 2006 workshop—*omnivalence*—invented on the spot by one of the participating scientists. Had the participants adopted this neologism, the challenge would simply read, "Technological development of omnivalent systems." Of course, error-free, 100 percent reliable systems do not exist. This challenge thus consists of creating methodologies, tools, and techniques to support omnivalence at all stages of software design, development, and management.

Challenge 5 encompasses all of the elements in Figure 1. For instance, the people involved must always have access to adequate information, for distinct interaction possibilities. Mechanisms must exist to compensate for sensor network failures; faulty or missing data might lead to production losses. For example, if the sensors fail to signal sudden drops in temperature, they can compromise an entire crop. Errors in software in tractors will result in incorrect fertilizer application, which will affect not only the product concerned, but also underground water sources. Inadequate models can affect harvest schedule, producing inferior quality yields. For instance, sugarcane must be harvested when

the sugar content reaches a peak. Harvesting programs can accurately forecast the appropriate period, which lasts only for a few days, and provide the logistics for scheduling humans and machines. Inaccurate computations with apparently small errors—for example, "only two days" or "only 50 kilometers"—will produce considerable economic losses.

To tackle omnivalent systems, researchers must account for the fact that software development is now a global effort, involving teams that work cooperatively across continents. The impact of time zones, cultural diversity, distance, and customer requirements create software engineering challenges requiring novel work in areas such as requirements engineering, cooperative distributed environments, and autonomic computing. This theme is particularly relevant in Brazil, where official policies and programs aim to establish a powerful software industry.

Although specified with software in mind, this challenge also applies to hardware and microelectronics. As major data and service providers know, the only certainty is that computers will fail. The design and maintenance of reliable networks is yet another issue along this line.

All five challenges are inherently multidisciplinary, and so will require changes in computer science teaching and research practices in Brazil.

A socially relevant example

Social relevance was yet another reason for choosing agricultural planning to illustrate the challenges. Agriculture is responsible for almost 50 percent of Brazil's GNP. The country has large planted surfaces, and a farmer's decision on what to plant (and when, where, and how) depends not only on local geophysical and climatological factors, but also on the national and international markets. Indeed, the country produces approximately 3 percent of the world's agricultural GNP.[4] Hence, small mistakes in data, models, or software will have unexpected side effects. For instance, a 5 percent error in estimating Brazilian soybean production will provoke a 3- to 6-million-ton difference in the world market.

SPOTLIGHTING THE WHY: THE ROLE OF SBC IN THE BRAZILIAN R&D SCENARIO

SBC is the largest computer society in South America, occupying an important place in the Brazilian scientific scenario, with thousands of members all over the country. Its many institutional members include not only universities, but also IT companies, from hardware firms to software developers. It was thus the ideal promoter of the Grand Challenges initiative.

SBC's wide range of activities involves people from academia, industry, and the government. Every year, more than 40,000 people from all over the country participate in the events promoted by the society's 26 special interest groups and 15 regional chapters. In 2007 alone, SBC sponsored more than 30 conferences, as well as more than 20 regional events and thematic graduate and undergraduate courses, and cosponsored another 70 scientific meetings.

The society's education board sets up undergraduate curriculum recommendations for distinct kinds of course profiles, which practically all of Brazil's 700-odd IT-related courses follow. It also orchestrates joint initiatives of all computer science graduate courses in Brazil, promotes meetings of undergraduate officers, and offers activities for high school teachers.

All students who apply for admission to a computer science graduate program take SBC's nation-wide POSCOMP exam (a GRE-like test for computer science), which universities in Peru have also used in the past two years. The society also promotes the Computer Science Olympics (a national contest for elementary and secondary school students) and the regional ACM collegiate programming contest. Because of its activities' breadth and scope, SBC holds advisory seats on the boards of several government agencies.

REFLECTING ON THE AFTERMATH: THE IMPACT OF THE GRAND CHALLENGES

A few important factors permeate all five challenges. Perhaps the most relevant is that they're inherently multidisciplinary, and so will require changes in computer science teaching and research practices in Brazil. An associated outcome is the emergence of a new computer scientist profile, in which the focus is not only on the problem to be solved, but also on its impacts on other research domains and on those domains' contributions to our research.

For instance, in the scenario in Figure 1, they will also advance research in other domains, such as mathematical modeling, remote sensing, and food science. They'll also influence industrial processes and generate novel and better food production procedures, with positive consequences for everyone.

Another issue debated in the workshop is the need to create mechanisms to intensify the still tenuous interaction between industry and academia as a means to apply good-quality research to technological development. Last but not least, solving many of the problems evoked will bring social and economic benefits.

What consequences have the workshop and reports produced thus far? First, they've reinforced in the Brazilian scientific community the need for cooperation among the sciences. Members of the workshop steering committee have presented the challenges at meetings and conferences, organized by various scientific societies and by the Brazilian Academy of Sciences. They've also made presentations to several government R&D

funding agencies. Audiences are interested not only in the research involved but also in the procedures used to elicit the challenges—that is, the workshop, the mechanisms used to promote it, and its organization and discussion dynamics. Together with other computing societies, SBC is organizing a similar event that will involve all Latin American countries.

Second, the challenges have given computer scientists new opportunities to clarify to colleagues from other domains what computer science research might mean. Because it's a comparatively new science, it's still not well understood in the scientific milieu. This, in turn, complicates the lives of computing researchers in Brazil (for example, in competing for grants or evaluating graduate programs). Hence, disseminating the challenges might lead to a new perception of computer science research.

Third, the challenges are at the heart of two SBC initiatives. The first involves organizing scientific conferences around the challenges. The society's 2008 annual conference, to be held in July, is perhaps the main example (www.sbc.org.br/sbc2008). The second initiative is a nationwide program aimed at creating new graduate and undergraduate curricula to prepare future generations of scientists to cope with a multidisciplinary world. The first activity in this direction was a seminar that included suggestions on course content to involve students in the computing problems presented by the challenges. Approximately 150 undergraduate officers attended this conference, and several virtual meetings have continued this debate.

The Grand Challenges initiative succeeded in helping government policymaking. They were at the core of two large requests for proposals launched in 2007. The first, promoted by the Microsoft-FAPESP Virtual Institute for IT Research (www.fapesp.br/materia.php?data[id_materia]=2825), centered on the fourth challenge and awarded five grants. The second, promoted by the Brazilian National Council for Scientific and Technological Development (CNPq), covered all five challenges and awarded 49 grants. These RFPs are good indicators that the Brazilian research community has enough scientists to start attacking the challenges.

Much ground remains to be covered. Although proposed as having a horizon of 10 years of research in mind, the challenges' many scientific ramifications will result in other long-term research endeavors beyond this period. Furthermore, cooperation with other domains is still in its infancy, especially with the arts, humanities, and social sciences. Also, as in the rest of the world, the need for qualified manpower is growing, aggravated by the continuing decrease in the number of young people interested in computer science. In this sense, the visions of the future presented by the challenges can help attract more students to the area, helping to show that our research can be exciting and socially relevant. ■

References

1. A.C.P.L.F. de Carvalho et al, "Grand Challenges for Computer Science Research in Brazil 2006-2016," workshop report, 2006; http://sistemas.sbc.org.br/ArquivosComunicacao/Desafios_ingles.pdf.
2. T. Munakata, "Beyond Silicon: New Computing Paradigms," *Comm. ACM*, Sept. 2007, pp. 30-34.
3. D. Rosenberg, "Early Modern Information Overload," *J. History of Ideas*, vol. 64, no. 1, 2003, pp. 1-9.
4. *FAO's 2005-2006 Statistical Yearbook*; www.fao.org/statistics/yearbook.

***Claudia Bauzer Medeiros** is a professor of computer science in the Institute of Computing at the University of Campinas (Unicamp), Brazil. Her main research interests concern scientific databases, with applications in biodiversity and agroenvironmental studies. She received a PhD in computer science from the University of Waterloo, Canada, and a Doctor Honoris Causa from Universidad Privada Antenor Orrego in Peru. She is a member of SBC, the IEEE, and the ACM. Contact her at cmbm@ic.unicamp.br.*