



Avalanche of data

Advances in eScience are changing the traditional way of conducting science

Yuri Vasconcelos

PUBLISHED IN NOVEMBER 2014

There was a time when it was a problem for scientists to obtain the necessary data required for progress in their research. However, in many fields of knowledge, recent advances in information technology, along with the democratization of computing, the expansion of computer networks and the proliferation of information sources, have directly resulted in massive production of data. This is occurring in fields as diverse as astronomy, which is inundated daily with thousands of images and data from celestial bodies captured by powerful telescopes; to molecular biology, which has benefited from the emergence of high-performance genetic sequencing instruments; to ecology, which is aided by a variety of technologies and sensors that can precisely document changes occurring in different biomes. All of these advances have left researchers with a new problem: how to process, organize and view the avalanche of data that is obtained through such diverse means. In response to this dilemma, a new branch of science has gained much attention. Known as eScience, it uses mathematical models and computational tools to analyze information and increase research speed in other areas of knowledge.

“The idea of connecting traditional scientific practice with the access, use and processing of large amounts of data will change the way we do science and increase its potential. FAPESP is at the forefront of this process; at the end of 2013 we launched the eScience Program,” said Carlos Henrique de Brito Cruz, the Scientific Director of the Foundation, during the Microsoft eScience Workshop 2014, held October 20-22, 2014

in Guarujá, on the coast of São Paulo. The objective of the eScience Program is to organize or integrate groups that are involved in research on algorithms, computational modeling and data infrastructure with teams of scientists working in other fields of knowledge, such as biology, social sciences, medicine and the humanities.

GLOBAL CHALLENGE

“One of the principal barriers we could face is communication problems among scientists on the teams needed to do science in this way, which is heavily based on data or large amounts of data. This requires very effective communication between researchers in the computer science field and scientists in other fields. It’s a challenge in Brazil, as it would be anywhere,” Brito said. He was a participant in the roundtable on “The Strategic Importance of eScience,” which also included scientists Jason Rhody, Senior Program Officer in the Office of Digital Humanities at the National Endowment for the Humanities, and Chris

Mentzel, Program Director at the Gordon and Betty Moore Foundation, two U.S.-based organizations that operate programs supporting science.

“At the present time, every field of research is affected by the modern scale of data production,” said Mentzel, emphasizing the importance of data scientists, which is the name given to those professionals who pour over the enormous volumes of data generated by researchers, and use it as a foundation to produce new knowledge. “They are researchers who work between disciplines. They are bridge builders,” he said. At the Gordon and Betty Moore Foundation, Mentzel heads a \$60 million program designed to incentivize eScience initiatives. Rhody believes that scientists are witnessing a paradigm shift. “We are moving from a culture of data scarcity to a culture of data abundance.”

eScience, named as such in 1999 by John Taylor, Director of the Office of Science and Technology of the United Kingdom, is also known by other names, such as data-driven science or data-in-



tensive computing. Some countries, such as the United States and England, already have government-supported programs focused on developing this new area of science. In Brazil, the Center for eScience Research at the University of São Paulo (USP), which was formally established in 2012, is worth special mention. The center has 20 researchers under the coordination of Roberto Marcondes Cesar Junior of the Institute of Mathematics and Statistics (IME) and one of FAPESP’s Supervising Panel on Exact Sciences and Engineering of the Scientific Directorate.

The Microsoft eScience Workshop 2014 was held in conjunction with the 10th IEEE International Conference on eScience, organized by the Computer Society of the Institute of Electrical and Electronics Engineers (IEEE), which was founded in the United States by electrical and electronics engineers. During the event, a panel discussion was held with researchers who have received grants from the FAPESP-Microsoft Virtual Research Institute and who connect computer science applications to the challenges posed by basic science in areas related to climate change and other fields associated with the environment. One of the studies presented explores innovative solutions for monitoring plants in the tropics, combining computer science research and phenology. Phenology, which is one of the oldest branches of science, is an area of ecology that studies the cyclical phenomena of plants, such as the emergence of leaves, buds, flowers and fruit,

Connecting computer science applications to phenology, which studies cyclical phenomena in plants



and the ways in which these phenomena are related to environmental conditions.

Under the coordination of researcher Leonor Patricia Morellato of the Phenology Laboratory at the Institute of Biosciences at São Paulo State University (Unesp) in Rio Claro, the project seeks to combine technologies to monitor the long-term changes undergone by plants native to the Cerrado, the Atlantic Forest, the rupestrian grasslands, and even the Caatinga. The central area of research is in Itirapina, in inland São Paulo State. “In addition to directly observing the plants at ground level, we installed a camera on top of an 18-meter tower to take daily photographs of the vegetation, and set up a meteorological station. We’re also going to have an unmanned aerial vehicle (drone) equipped with a hyperspectral sensor and a camera to add a spatial scale to the data collection,” the researcher says. With high spatial resolution, hyperspectral sensors can provide details about the physical and chemical properties and physiological responses of the plants shown in the images. Morellato views phenology as one of the best tools for understanding the effects of climate change on plants. “This has already been established in temperate regions, where the phenological triggers are ambient temperature and length of day. But we know little about what happens in tropical plants. With the data from the cameras and the hyperspectral sensor, we want to determine what the triggers are for phenology in the tropics, i.e., what causes flowers, fruits and leaves to emerge at certain times,” she says.

ANALYZING IMAGES

According to Morellato, without the help of computer science researchers and resources, it would be impossible to conduct the research. “The volume of data we will collect is enormous. One digital camera alone takes 60 photos per day. We have 11 cameras monitoring six types of vegetation, and we need to observe developments for at least one growing season in order to go back and connect them with the climate data. Then we need to process and analyze all the images, which would be impossible to do with a simple electronic spreadsheet. We need help to work with this big data. That is why we enlisted a masters student to create a database especially for the project, and

Wireless sensors installed in forests generate large amounts of data on natural processes

a post-doc to work on software for viewing and organizing the images.”

The research collaborator for the Unesp professor is scientist Ricardo Silva Torres, Director of the Institute of Computing at the University of Campinas (Unicamp), who is also working on a project under the FAPESP-Microsoft Research agreement. He is heading a study aimed at developing new analytical techniques and computational tools for processing remote sensing images to enable scientists to analyze the dynamics of some biomes on regional and continental scales. The research is being conducted in partnership with Professor Marina Hirota of the Department of Physics at the Federal University of Santa Catarina (UFSC), and it focuses on South American tropical biomes.

Another study presented at the event in Guarujá is led by Unicamp ecologist Rafael Silva Oliveira, collaborating with researchers Antonio Alfredo Ferreira Loureiro of the Computer Science Department at the Federal University of Minas Gerais (UFMG) and Stephen Burgess of the University of Western Australia. “The goal of our study is to investigate the water and carbon dynamics in cloud forests, pastures, and the transition areas between them,” says Oliveira. Cloud forests are found at the tops of tropical mountains. “We want to understand how key processes, such as carbon

absorption and storage, transpiration of trees and the way plants capture water from fog, are affected by changes in land use and climate variations.”

The field studies are being conducted in a section of forest along the Mantiqueira Range, in the region of Campos do Jordão in inland São Paulo State. According to Oliveira, a network of wireless sensors is being implemented there to monitor three layers of the ecosystem, the atmosphere, vegetation and soil, to determine the microclimatic parameters of plant metabolism and water dynamics in the soil. “These data could improve the forecasting of environmental impacts caused by changes in land use, and at the same time will make it possible to develop hydrological models and models of biosphere-atmosphere circulation with greater predictive capability,” Oliveira explains. ■

Projects

1. Towards an understanding of tipping points within tropical South American biomes (Nº. 2013/50169-1); **Grant mechanism** Research Partnership for Technological Innovation (PITE) and FAPESP-Microsoft Agreement; **Principal Investigator** Ricardo da Silva Torres (Unicamp); **Investment** R\$384,838.38 (FAPESP).
2. Combining new technologies to monitor phenology from leaves to ecosystems (No. 2013/50155-0); **Grant mechanism** FAPESP Research Program on Global Climate Change – Research Partnership for Technological Innovation (PITE) and FAPESP-Microsoft Agreement; **Principal Investigator** Leonor Patrícia Cerdeira Morellato (Unesp); **Investment** R\$1,115,752.48 and \$535,902.72 (FAPESP).
3. Soil-plant-atmosphere interactions in a changing tropical landscape (No. 2011/52072-0); **Grant mechanism** Research Partnership for Technological Innovation (PITE) and FAPESP-Microsoft Agreement; **Principal Investigator** Rafael Silva Oliveira (Unicamp); **Investment** R\$644,800.74 and \$663,429.82 (FAPESP).

