

# From molecules to the Milky Way: dealing with the data deluge

November 7 2007

---

Most people have a few gigabytes of files on their PC. In the next decade, astronomers expect to be processing 10 million gigabytes of data every hour from the Square Kilometre Array telescope.

And with DNA sequencing getting cheaper, scientists will be data mining possibly hundreds of thousands of personal human genome databases, each of 50 gigabytes.

CSIRO has a new research program aimed at helping science and business cope with masses of data from areas like astronomy, gene sequencing, surveillance, image analysis and climate modelling.

The research program, which began this year, is called ‘Terabyte Science’ and is named for the data sets that start at terabytes (thousands of gigabytes) in size, which are now commonplace.

“CSIRO recognises that, for its science to be internationally competitive, the organisation needs to be able to analyse large volumes of complex, even intermittently available, data from a broad range of scientific fields,” says program leader, Dr John Taylor, from CSIRO Mathematical and Information Sciences.

One aspect of the problem is that methods that work with small data sets don’t necessarily work with large ones.

An aim of the program is to develop completely new mathematical

approaches and processes for scientists in a range of disciplines to further their research and boost Australia's position as a world science leader.

“Large and complex data is emerging almost everywhere in science and industry and it will hold back Australian research and business if it isn't dealt with in a timely way,” Dr Taylor says.

Countries like the US also recognise the challenges, as Dr Taylor has seen first hand in his ten years' working in laboratories there.

“This will need major developments in computer infrastructure and computational tools. It involves IT people, mathematicians and statisticians, image technologists, and other specialists from across CSIRO all working together in a very focussed way,” he says.

After a workshop in September, specific research areas have been identified and projects are progressing in advanced manufacturing, high throughput image analysis, modelling ocean biogeochemical cycles, situation analysis and environmental modelling.

Source: CSIRO Australia

Citation: From molecules to the Milky Way: dealing with the data deluge (2007, November 7) retrieved 19 April 2024 from <https://phys.org/news/2007-11-molecules-milky-deluge.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.