



National Research  
Council Canada

Conseil national  
de recherches Canada

---

# **NRC · CNRC**

---

## ***Opening Address***

Dr. Arthur J. Carty  
President  
National Research Council

*“Scientific and Technical Data: Extending the Frontiers of Research”*

18<sup>th</sup> International CODATA Conference  
Montreal, Quebec

29/September/2002

## *Introduction*

Dr Rumble, distinguished guests, ladies and gentlemen.  
Good evening. It's a pleasure to be able to welcome you, on behalf of the National Research Council of Canada, to Montreal. Cette conférence du CODATA est la première à se tenir au Canada depuis plus de 15 ans et nous sommes extrêmement fiers et heureux de vous voir tous réunis ici!

As all of you here already know, CODATA plays a vital role within the international scientific community – facilitating international cooperation, and promoting best practices around the globe in the processing, management and analysis of scientific and technical data. With about 200 delegates expected at this conference, the next few days promise to be a tremendous opportunity for interdisciplinary exchange on a wide range of issues relating to data archiving, analysis and technology.

For our part, the National Research Council – or NRC -- is participating in CODATA in a number of ways. Our Canada Institute for Scientific and Technical Information sponsors the Canadian National Committee for CODATA. And our Institute for National Measurement Standards participates in CODATA's Task Group on Fundamental Constants.

Of course, NRC scientists are also engaged in a wide range of activities – from data mining to archiving – that are helping to extend the limits of data technology and analysis. Several NRC representatives will be presenting their work over the next few days.

My focus this evening is on your theme: "Frontiers of Scientific and Technical Data." I'd like to explore some of the ways that new data management techniques and tools are transforming the frontiers of science and technology. I'd also like to touch on some of the challenges facing those involved with data management – the Internet, interoperability and intellectual property issues, to name a few. I will also say a few words on the role CODATA can play in ensuring that the scientific community the world over continues to have access to the quality, reliable data it requires to advance the interests of humanity through scientific discovery.

***First of all let me say a few words about:***  
*Scientific Information and Data: The Basic Foundation for Scientific Research*

Before I begin, I'd like to make it clear that I am by no means a data specialist. Indeed, as president of a national research organization I am more likely to think of data in terms of the number of pieces of paper that cross my desk each day!

Joking aside, I consider myself fortunate to still be an active scientist with a small but very productive group of PDFs and students and my years as a chemist, combined with eight years in my current role at NRC, have given me a strong

appreciation of the many different ways data management and data technology contribute to the constant advancement of science.

Quite simply, sound science rests on the quality and reliability of its data. Without data we cannot test our hypotheses. And without reliable data – that has been soundly and rigorously evaluated – we cannot trust in the accuracy of our results.

Bertrand Russell – the great mathematician and philosopher -- used to refer to Aristotle to make this point. According to Russell: and I quote “*Aristotle maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify this statement by examining his wives’ mouths.*”<sup>1</sup>

Quality data of course are the backbone of the scientific method.

Where would we be, for example, without the reliability of the fundamental constants? How could we measure a nanosecond? How would we calculate gravity? Or vouch for the validity of our experiments?

Many of us take these quantifications of natural phenomena for granted, yet they are the product of years of painstaking research and revision by physicists, specialists in precision measurement, and pioneers in data technology. This is research that often paves the way for fundamental and significant breakthroughs in our ability to understand and manipulate the physical world around us.

Scientific breakthroughs would be impossible without the careful and innovative management, evaluation and analysis of data.

Science-fiction writer, and philosopher, Arthur C. Clarke, once wrote: “*The world needs uninhibited thinkers, not afraid of far out speculation; it also needs hard-headed conservative engineers who can make their dreams come true.*”

I would add that science needs data specialists who are a blend of both.

### **Fueling the Fires of Discovery: The Role of Data in 21<sup>st</sup> Century Science and Technology**

This is especially true at this time in our history. Science has entered a new age in the 21<sup>st</sup> century – an exciting new era of synergism, in which the quantum revolution, the computer revolution and the biomolecular revolution are converging to vastly accelerate the pace of discovery and increase our ability to manipulate matter, life and intelligence.

At the same time, scientists across the disciplines are moving away from reductionism -- embracing the cross-fertilization of ideas inherent in multi-

---

<sup>1</sup> Russell, Bertrand, *The Impact of Science on Society*, 1952.

disciplinary work, and taking advantage of the vast potential that advances in computing power and data technology promise for the collection, storage, sharing, manipulation and aggregation of huge amounts of data.

The strategic importance of data in science and engineering is on the rise. In the words of CODATA President, John Rumble, Jr.: “... *large collections of data are rapidly becoming sources of new discovery, the basis for new scientific insight and understanding.*”<sup>2</sup> The day is swiftly coming when every scientist and engineer will need a working knowledge of the tools available to collect, store, mine, evaluate, analyze, visualize, archive and disseminate data.

### *Data At The Frontiers of Science and Technology*

So what are some of the breakthroughs? How is data science really making a difference?

All of us are familiar with high profile, global-scale projects such as the Hubble Space Telescope, the Human Genome Project and the International Geosphere-Biosphere Programme. New data and information technologies have made the collection, sharing and analysis of massive amounts of data possible for these major undertakings. But countless other research projects around the globe are also using new data tools and techniques to extend the frontiers of science and technology.

Indeed, just a quick flip through your conference program reveals a wealth of examples where advances in data science are leading to new discoveries in fields as diverse as biotechnology, materials science and environmental management.

And there are many others.

Take, for example, a new Canadian initiative, known as the Biomolecular Interaction Network Database – or BIND. Sponsored by a variety of Canada’s leading public and private research institutions –including the National Research Council Canada<sup>3</sup> -- BIND provides public access to one of the world’s most comprehensive sources of biomolecular interactions. Researchers can use the

---

<sup>2</sup> Rumble, John Jr., “The Fuel for 21<sup>st</sup> Century Science and Technology,” *CODATA Newsletter*, Committee on Data for Science and Technology, No. 83, March 2002.

<sup>3</sup> BIND’s sponsors include: Aventis Pasteur Canada, the Canada Foundation for Innovation, the Canadian Genetic Diseases Network, the Canadian Institutes of Health Research, the Canadian Protein Engineering Network, the Centre for Molecular Medicine and Therapeutics, IBM Canada, MDS Proteomics, MDS Sciex, Mount Sinai Hospital (Samuel Lunenfeld Research Institute, the Ontario Challenge Fund, the National Research Council Canada, the Natural Sciences and Engineering Research Council of Canada, and the University of British Columbia (Centre for Molecular Medicine and Therapeutics).

database to study networks of interactions, map pathways across taxonomic branches and generate information for kinetic simulations.

Intended to act as a resource that will rapidly increase our understanding of human health and speed the development of new medicines, BIND also anticipates the coming influx of interaction information from high throughput proteomics, including detailed information about post-translational modifications, from mass spectrometry.

Another fascinating example of data technology helping to push the limits of knowledge can be found in the field of astronomy. Here, too, there is a Canadian connection.

Scientists at NRC's Herzberg Institute of Astrophysics are part of an international effort to create the "Virtual Observatory" -- a collaborative project that uses advanced data and information management tools to make some of the world's leading ground- and space-based observations readily accessible to astrophysical scientists around the globe. Scientific and technical staff have developed the means to integrate datasets from different wavelength regimes, as well as to add value to datasets through additional processing. Dr. David Schade of the HIA's Canadian Astronomy Data Centre will be making a presentation on the project tomorrow.

And then there is the exploding field of data mining. New techniques in this area are accelerating research across almost all the scientific disciplines. In Canada, for instance, To the Information Systems -- an NRC spin-off company -- is using new data mining techniques to facilitate research and experimentation in the development of advanced materials. The company is using ab initio quantum mechanical methods to mine crystal structure and property databases to calculate and predict the properties of new ternary and quaternary materials. Each new experiment expands the database and provides new insights into the laws of physics and chemistry.

In other areas, the interactive computer programs pioneered by Professor A.D. Pelton and his team at École Polytechnique here in Montreal are also attracting international attention. Dr. Pelton -- winner of this year's CODATA prize -- and his researchers use their new computer programs to perform thermochemical calculations on single inorganic substances and solutions. Together they have also developed the Facility for the Analysis of Chemical Thermodynamics -- or FACT -- one of the largest database computing systems in chemical thermodynamics in the world and the only one of its kind in North America.

And finally, I'd like to mention NRC's own Canadian Bioinformatics Resource (CBR). This powerful, new distributed computing network provides Canadian researchers with high-speed access to over 70 of the world's most extensive genomics and biotechnology-related databases and resources. Such networks and the tools to mine the databases are becoming critically important as the

amount of genomics information available from high throughput sequencing, proteomics and structural biology increases exponentially.

### *Opportunities and Challenges*

As these examples indicate, data science is coming into its own. One advance is building upon another -- opening up incredible opportunities for discovery across all the scientific disciplines, as well as the possibility of many exciting new applications in medicine and industry beyond.

Of course – as is so often the case – these opportunities are accompanied by challenges.

The Internet, for instance, is proving itself a double-edged sword for the data science community. Yes, the Internet can be a tremendous tool for the collection and exchange of information, best practices and vast quantities of data. But it is also becoming increasingly congested and its popular use raises issues about authentication and evaluation of information and data.

Interoperability is also a significant challenge. The growing number and volume of data sources, together with the high-speed connectivity of the Internet and the increasing number and complexity of data sources, are making interoperability and data integration an important research and industry focus. Incompatibilities between data formats, software systems, methodologies and analytical models are barriers to the easy flow and creation of data, information and knowledge.

Similarly, as the volume of data and nature of its use expands and changes, the pressures around the preservation and archiving of scientific and technical data are mounting. Without concerted action, we risk losing many potentially significant collections of data – a loss that will be felt for generations to come. Developing countries, in particular, face significant technical, financial and organizational hurdles in this area.

And then there are the complex legal, economic and ethical issues associated with the creation, management and exchange of increasingly large and sophisticated databases. Who should have access to data and information about biological materials? world species diversity? or genome information for plants, animals or humans? Who should be able to look at your medical records? Ethics in cyberspace are proving as thorny as they are elsewhere.

Proprietary issues are equally tough. How do we balance intellectual property rights with the concerns of those engaged in public interest or not for profit research and education? Will the European Directive – which provides extensive copyright protection to the creators of scientific and technical databases in the European Union – put a chill on scientific exchange and inquiry?

Finally, data collection and management are part of the basic infrastructure of science. And yet, from country to country, the funding provided for this kind of

activity is uneven. How can we overcome inequities in political support and funding for data-related activities – particularly in developing countries? Given the vital link between science, innovation and economic and social development this is as much an issue of social justice as it is of scientific progress.

These are all fundamental issues – not just for those involved with data management and data technologies – but for science in general. Every discipline must come to grips with these data-related challenges.

### *The Role of CODATA*

So where does the search for solutions begin?

In my view, it begins – at least in part – with CODATA. CODATA is already a strong force in promoting the exchange and dissemination of the latest information on data technology and techniques, and it is proving itself an increasingly strong and able player in the bid to influence policy-making around data-related issues.

We are extremely fortunate to have an interdisciplinary forum for international co-operation already in place. Standards development; training and education; access to data and intellectual property issues – these are all areas where CODATA have the international networks and credibility to make a difference.

I'd like to commend CODATA on its breadth and its vision -- particularly in its efforts to foster cooperation and exchange between researchers in developing nations and other members of the international data community. Both the Survey of Data Sources in Asian-Oceanic Countries and the Task Group on Scientific Data Sources in Africa are proving to be highly successful in facilitating regional and inter-regional cooperation.

The Senegal River Basin project—for instance – is an impressive example of multi-disciplinary, international cooperation. The project -- which brings together the work of specialists in biodiversity, agronomy, environment and health, geography, biology, information and communications, sociology and social sciences, chemistry, toxicology, computer science and evaluation – is one of the first of its kind on the African continent. The resulting database is proving to be an outstanding resource for decision-makers considering issues relating to sustainable development in Africa and elsewhere.

### *Conclusion*

In closing, I urge you to keep your collective voice – in the form of CODATA – strong.

In this age of increasing globalization of knowledge and innovation, the creation and maintenance of vibrant – yet politically neutral -- scientific networks such as CODATA are more important than ever.

Allow me to leave you with a quote from the poet, T.S. Eliot. More than half a century ago, Eliot asked: and I quote "*Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?*"

These words can be taken as a caution for the information age. But to me, they also speak to our responsibility as scientists.

As scientists – with our bent to explore and to understand – our task is to sift through the data and the information, so that we may draw out the knowledge and the wisdom, and apply our creativity to the glory and betterment of the human spirit and condition. As scientists, we must be part of the solution.

I trust you will make the most of the next few days to renew old acquaintances, to forge new contacts, and to lay the seeds to work together to push the frontiers of science and technology, again and again.

Je sais que vous profiterez des prochains jours pour renouer des liens, pour établir de nouveaux contacts et pour jeter les bases de nouveaux partenariats afin de reculer encore et toujours les frontières de la science et de la technologie.

I wish you all the best with your conference!

Je vous souhaite à tous et à toute une bonne conférence!