

CODATA and the World Summit on the Information Society

Introduction

The UN General Assembly, on 21 December 2001, adopted a Resolution endorsing the organization of the World Summit on the Information Society (WSIS), to be convened under the patronage of the United Nations Secretary-General, Kofi Annan, with the International Telecommunication Union taking the lead role in its preparation along with interested UN organizations and the host countries. The World Summit on the Information Society will be held in two phases: in Geneva from 10-12 December 2003 and in Tunis in 2005.

The Summit will address the broad range of questions concerning the Information Society and move towards a common vision and understanding of this societal transformation. The Summit is expected to adopt a Declaration of Principles and an Action Plan to facilitate the effective growth of the Information Society and to help bridge the Digital Divide. It aims to bring together representatives from the highest levels of government, the private sector, civil society and NGOs. It will offer a unique opportunity for the world community to discuss and give shape to the Information Society. More information can be obtained from <http://www.itu.int/wsis/>.

Importance of Scientific Data to Society

Scientific data are the quantitative information used to communicate the results of science. Experiments, observations, theory, models and simulation all generate scientific data. In the past, data usually were presented as tables of numbers. Today scientific data are most often stored in databases and involve numbers, text, images, diagrams, pictures, and equations. They may range in size from a simple number; e.g. the speed of light equals 2.9999 centimeters per second, to vast computer stores of satellite images.

The primary feature that differentiates scientific data from scientific information is that the data are the results of science, not a description of what was done, how it was done and why it was done.

These latter items plus data plus more comprise scientific information. To have a full understanding of science, full information is required. To use scientific results to do new science require the data themselves.

Today our ability to generate data is unprecedented. New instruments such as the Hubble Space Observatory make possible detailed observations on an immense scale. Electronics and computers simplify experimental operations and make repeat measurements much easier. Computers, application software and advances in numerical algorithms and modeling techniques allow calculations of virtually every physical phenomenon. Fortunately information technology provides us with the capability to capture, store, manage and use these large volumes of data. Indeed, during the last twenty years, every area of science has become dependent on scientific data activities to capture and exploit the advances in their disciplines and almost every scientist has become a data practitioner (creating and using databases).

A less obvious but perhaps more important reason for the growing importance of scientific data is the evolving multi-disciplinary nature of research and development. Advances in one discipline become critical for research in another area, and data are the mechanism to make those linkages quantitatively. For example, a medical researcher will design a potential gene therapy using the data from molecular biology. Environmental modeling requires basic chemical reaction rate and solubility data.

Society in general uses the results of science in many ways. Society takes advantage of a better understanding of our physical world to improve the quality of life. Society uses research results to solve problems. Society accepts technological advances based on an understanding of their impact on our physical world. Society uses science to control the future based on accurate predictions and informed decisions. In these and other ways, society takes uses scientific results, usually irrespective of the original intent of the scientific investigation.

Example: Use and misuse of medicine

There is no doubt of the tremendous impact modern pharmacology has had on health quality. From the simplest antibiotic to complex chemotherapy, almost everyone has felt the benefits. Yes the abuse of modern drugs has become a major health issue itself, with the advent of antibiotic resistant germs to inappropriate use of behavior modifying drugs for borderline hyperactive children. Sound scientific data exist that clearly demonstrate these problems, yet those data are either not communicated effectively or just plain ignored.

Example: The change to environmentally friendly refrigerants
In the 1980s, scientific evidence accumulated of the negative impact chloro-fluoro-hydrocarbons had on various regions of the earth's atmosphere. These CFCs were prevalent in modern household and industrial products. However the scientific data were clearly and effectively presented not only to scientists and engineers, but also to society at large. Society's acceptance of these data was instrumental to the many changes and comprises that were required to change to new propellants and refrigerants.

Because data are a primary mechanism for communicating scientific results, scientific data are important to society. Health, safety, social, environmental, security and other issues are illuminated by these data, and for society to take advantage of them, data quality accessibility and utility are important.

Data quality is perhaps of most significance for a decision based on bad data is a bad decision. Understanding data quality is very hard for non-scientists. Many scientists are aggressive in proclaiming the correctness of their data, and non-scientists are often persuaded of that correctness not because of scientific arguments but rather by the biases or persuasion of the advocate. In contrast, many scientists are reticent to assert the quality of their data because of their modesty or recognition of uncertainty. So in one case people believe too much, in the other too little. The growing importance of scientific data to society make it mandatory for the scientific community to make data quality more understandable to society at large.

Data accessibility is equally important because if data are not accessible, it is the same as if they did not exist. Today it seems unimaginable that data are not accessible. The Internet, world wide web, search engines, PCs on everyone's desk and web pages by the millions appear to have solve the accessibility problem. That is not the case. Substantial economic barriers are appearing to society's

access to data. Companies want to exploit data for profit motives. Nations consider protective intellectual property regimes to foster economic development. Data vendors want to make money from providing data access. And large amounts of existing data are being ignored because of the costs associated with computerizing older paper data collections. Society and science must work together to make sure that investments in science are not constrained by lack of accessibility. Data themselves are rarely profitable; their use is. Economic rewards should not be based on data scarcity, but rather on innovation in data exploitation.

Data management is also of concern. What do you do with a terabyte of data? An exabyte of data? What do you do with data from ten different data sources, each in a different format and perhaps in different languages? How do you take advantage of incomplete data sets, or to compare one data set to another when the two sets have different amounts and kinds of data? These issues are of concern to scientists themselves, and are of greater concern to non-expert users. Tools such as standards, visualization and analysis, aimed at non-expert users will be critical to society taking advantage of scientific data. It is the job of science to help society manage and use data.

It should be clear of the great importance of scientific data to society. One can quickly list numerous issues that society must address – for example, the use of genetically modified organisms, energy shortages, HIV, global climate change – to realize that clear communication and widespread availability of high quality scientific data is critical for informed decisions. The upcoming World Summit of the Information Society presents a remarkable opportunity to educate scientists and society about scientific data and their importance. As **ICSU** addresses the role of scientific information in the emerging information society, **CODATA** is pleased to work with **ICSU** and **UNESCO** in including the role of scientific data in the discussions.