MEETING REPORT

Big and open data*

Unprecedented explosion in the human capacity to acquire, store and manipulate data together with instant communication globally has transformed research from an era of data scarcity to data deluge, where experts see it as a 'second revolution of discovery'. Data citation being one of the ways of giving attribution, its standards and good practices can form the basis for increased incentives, recognition and rewards for research data activities, that in many cases are currently lacking.

A workshop was conducted to take stock of what is being done in the data citation area, the existing culture of data sharing and citation practices along with the needs and interests of the respective disciplines. Around 300 eminent scientists, domain experts from learned societies, funding agencies, policy makers, researchers and research librarians from across the country and international organizations participated in the workshop. Structured in the form of theme-based panel discussions, the workshop dealt with data science, technology, research, applications and development in the Indian context; data attribution and citation practices; and data management, sharing and services: current and future roles of data stakeholders.

Usha Mujoo Munshi (IIPA) spelt out the concept behind this international workshop and also the key deliberations at the end of the workshop. She stated that the recommendations arising from the deliberations are expected to provide a valuable input towards formulation and adoption of a formal data citation protocol for science and social science communities.

Chandrima Shaha (National Institute of Immunology (NII)), emphasized the need for taking action on discipline-wise specific big data and citation attributions, particularly relating to large datasets of life sciences, biomedical data, genomic studies and the like. She underscored the need for formulation of a national-level policy in India to establish standards in data citation practices. Krishan Lal (Past President, INSA) highlighted the need for creation, use, reuse and repurposing of data for R&D globally. He appraised the role of CODATA in promoting such activities for global good and summed up the recommendations for framing policy guidelines. Tishyarakshit Chatterjee (IIPA) emphasized the importance of reliable data and mentioned about the data monopoly by multinationals and countries.

Jan Brase (CODATA-ICSTI Data Citation Task Group) pointed out that efforts of the data generator are usually unnoticed and felt that one should be able to cite the data/dataset itself in a manner similar to citation of articles. T. C. A. Anant (MoSPI) in his keynote address, stated that the Government of India (GoI) by itself is the largest user and producer of data in social sciences. He emphasized the need for integration of data generated by the Government in administrative silos and of data across the silos. He opined that though more openness is required, the security and safeguards of individuals and the country should always be kept in mind.

R. Chidambaram (Principal Scientific Advisor to GoI and Chairman, SAC-C), while delivering the inaugural address highlighted the significance of big and open data. Big data and open data are not the same and data related to national security and integrity cannot be made open. There is a great deal of knowledge about data storage, management and access in the scientific community. He mentioned that it would be worthwhile to have a brainstorming session to see if these techniques can be used in governance.

K. R. Murali Mohan (DST) specified the challenges of evolving data management landscape, data standards, big data, curation, storage, security and privacy. He also informed the audience about the efforts of DST to motivate entrepreneurs to conduct research in big data and create analytics as service.

Yadati Narahari (IISc, Bengaluru) stressed the need for theory, technology as well as tools for transforming raw data into actionable knowledge leading to decision making and the need for data analytics to be integrated with big data ecosystem for India. Ajit Kembhavi (IUCAA, Pune) shared his own experience working with astronomical data, their quality and distribution. He explained the data-driven initiatives in astrophysics and bioscience programmes using data management techniques. B. B. Prahlada Rao (C-DAC, Bengaluru) stressed on the need of associating domain experts in big data analytics.

Alok Bhattacharya (JNU, New Delhi) highlighted that there are many experimental data available for which data analytics may be implemented. He suggested that at first, data should be available in specific format for fast and efficient analysis. Santanu Chaudhury (IIT, Delhi) mentioned that for benchmarking of the research work carried out in a specific domain, there is need for building ontological framework using citation network. Suresh Misra (Centre for Consumer Studies) explained how the data collected can help in understanding the trends of complaints, and recommended that access to the consumer databases be offered for quality services. Vinod K. Sharma (IIPA) stressed on the importance of reliable datasets for cyclone tracking, rainfall and earthquake predictions, etc. to minimize loss of lives and also for economy. Durga Toshniwal (IIT, Roorkee) mentioned that data privacy should be maintained. She said that there is a need to evolve techniques for privacy protection of data during data mining, analysis and knowledge discovery.

Nafees Meah (Research Council, UK) highlighted the need to develop a joint approach for data collection, data hubs for different sectors, low-cost approaches for data visualization, provide training in data sciences, machine learning and low-cost methodologies for data collection. S. Ganesan (Homi Bhabha National Institute, BARC, Mumbai), while highlighting the challenging issues in multi-scale modelling in physics, emphasized data science activities in nuclear science tailored to the Indian atomic energy

applications and stressed the need for horizontally sharing the data science expertise, across disciplines. Paul Meller (Economic and Social Science Research Council, UK) pointed out the challenges of open science and open data, while emphasizing on security concerns. He suggested that appropriate policies and procedures be adopted with incentives and rewards through capacity building and training.

Neeta Verma (NIC) talked about the open data initiative of GoI and open data platform developed by NIC. Over 18,000 datasets available on this platform (http://data.gov.in) can be accessed by anyone for academic, research and development purposes. Each dataset published on this platform has a unique URI/UUID, which can be used for citation purposes. She also discussed technologies and tools that can be deployed for big data analytics of government data and highlighted the need for cloud-enabled data infrastructure.

Jagdish Arora (INFLIBNET) discussed the current status of ICSSR data repository–policy framework and shared the procedure to cite data from ICSSR Data Service. M. Haridas (Centre for Land Warfare Studies) spoke on execu-

tion challenges of big data applications by the defence and security forces.

H. K. Kaul (DELNET) discussed the need for training in big data in library science and also making Ph D research data accessible to users for further analysis from different dimensions. Debasisa Mohanty (NII) spoke on the development of structural bioinformatics methods for analysis of protein interaction network as the nucleic acids research datasets are to be released to the public. Roshan Lal Raina (JK Lakshmipat University, Jaipur) discussed the increasing credibility and effectiveness of data, and suggested provision of data for citing practices. Harpreet Singh (ICMR) discussed the medical data that can be collected from hospitals, the sequencing format, research paper citation, data reliability and data standards. Biplav Srivastava (IBM Research, India) opined that data must be integrated, while pointing out the challenges of applying analytics and linked data.

In the concluding session, a detailed presentation on the deliberations clearly brought out how the presentations of speakers helped spell out a set of recommendations for further consideration. Some of these were:

- Emphasis on the need for new courses on data science and technology like M Tech must be ensured to strengthen S&T activities in this domain.
- There is an urgent need to organize international schools in this field with experts from across the globe, covering all aspects.
- The scientific data which emerge out of projects sponsored by Government agencies like DST, DBT and other funding bodies should be put in the public domain, preferably in a national database.
- Continuous international collaboration will greatly benefit the data community. These collaborations could be inter-academy as well as inter-governmental. South–south collaboration is important.
- Collaboration between S&T sector and social science sector should be initiated and strengthened.

Usha Mujoo Munshi*, Indian Institute of Public Administration, New Delhi 110 002, India; **Neeta Verma,** National Informatics Centre, New Delhi 110 003, India. *e-mail: umunshi@gmail.com

MEETING REPORT

Management of *Phytophthora* – a deadly plant pathogen*

Stramenopile pathogens of the genus *Phytophthora* cause devastating diseases

on a wide range of agricultural and horticultural crops, natural vegetation and forestry worldwide. There are over 140 species in the genus and many have wide host range. *Phytophthora infestans*, which caused the great Irish Potato Famine during late 1840s, still remains the most destructive pathogen of potatoes and tomatoes, causing crop losses of up to US\$ 6.7 billion annually. This famine caused by *P. infestans* in 1845–57, changed the history of many countries. Other notable species that have emerged in more recent times are *P. ramorum* on oak, *P. alni* on alders, *P. kernoviae* on ornamentals, *P. cinnamomi* on forest crops, *P. agathis* on kauri, *P. cactorum* on hardwood trees, *P. capsici* on solanaceous and cucurbitaceous vegetables, *P. fragariae* on strawberries, *P. megakarya* on cocoa, *P. palmivora* on palms, and *P. meadii* on arecanut, small cardamom

and rubber from different parts of the world.

Despite the above records, *Phytophthora* continues to be a major pathogen. One of the major reasons for the spread of pathogens is the difficulty in implementation of quarantine regulations in the open trade regime. The knowledge base on *Phytophthora* research and extension methodologies needs proper reorientation to meet the demands of farming community. To address these issues, an international symposium was organized.

The symposium began with a welcome address by M. Anandaraj (ICAR-Indian Institute of Horticultural Research (ICAR-IIHR), Bengaluru). P. Chowdappa (ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI), Kasaragod) highlighted the recognition of 140 species of *Phytophthora*, challenges