EDITORIAL

Big Research Data and Data Science

Jianhui Li[1]

[1] Secretary-General, CODATA-China, China, Director, Scientific Data Center, Computer Network information Center, Chinese Academy of Sciences, China

CODATA China was very pleased to organize the 1st Scientific Data Conference – Scientific Research Big Data and Data Science, which was held in Beijing, China on 24–25 February 2014. Participants included more than 400 domestic experts, scholars and students from universities, research institutions and industry.

Group Photo of 1st Scientific Data Conference.

   The Conference aimed to improve understanding of the central issues in the era of Big Data, to promote multidisciplinary communication and collaboration, to help the development of young data scientists, to encourage the revitalization of traditional research approaches and to contribute to and support the Chinese national strategy to promote innovation. Around the world, there is talk of a 'data revolution' – this conference aimed to place China at the forefront of this revolution, providing the communication, skills and training to help Chinese scientists seize the opportunities of Big Data and 'ride the wave' of increasing data volumes, velocity and variety. Spanning two days, the conference featured two plenary sessions and fourteen breakout sessions. There were four major keynotes, three invited reports and three reports on projects and initiatives. The keynote lectures focused on the hot issues in the Big Data era, including integration and notation of Big Data, development opportunities and major challenges for science and technology. The breakout sessions also included technical sessions and open forums, with topics including:

· Big data applications in earth and spatial science
· Big data analysis and processing technology

· Life science and medicine data and
  applications
· Materials science data and applications
· Big data technologies and applications in agriculture and rural informatization
· Linked data and information recommendation
· Scientific data visualization
· Physics and chemistry data and applications
· Cloud computing and data discovery
· DOI registration and release
· Open forum: data science and data scientists

Academiacan GUO Huadong made keynote lecture.

Plenary Session.

Breakout Session.

At the closing ceremony Prof. LI Jianhui, secretary-general of CODATA-China, made a concluding speech thanking the participants and described CODATA-China's aspirations for the future. He also announced the forthcoming book *Scientific Discovery in Big Data Era* (now published) and that the 2nd Scientific Data Conference - Data, Science and the Silk Road Economic Belt would be held in Lanzhou in August 2015. The conference was one of the most successful academic activities to have been convened in the 30 years since CODATA-China was set up.

This collection of the *Data Science Journal* brings together the papers presented at this conference, providing insight into the current state of the art in China for data. The papers were selected and reviewed by the conference organizers as a fitting and representative collection of the discussions and insights presented at the conference. The organizers would like in particular to thank Diane Rumble for her assiduous editorial work for this special conference collection.

The selected papers consist of research work in three categories: 1) common technologies in data science, 2) data science research and 3) applications of data science in different domains.

Seven of the selected papers are concerned with the critical need for common technologies in data science:

- Cai and Zhu construct a dynamic assessment process of data quality in the big data era;
- Jiang et al. propose search strategies using synchronous search and asynchronous search to conquer the problems of search speed difference and to combine the two separate search results;
- Shen et al. introduce OpenCSDB, which is a solution of applying Linked Data in the Scientific Database, which will promote the sharing of scientific data and play a greater role in the "Twelfth Five-Year" program;
- Xiang et al. propose an invocation sequence of web service composition and its related invocation policies based on Petri net and analysis of structural relationships;
- An approach is proposed by Zhong et al. to increase the accuracy and efficiency of seeding algorithms of magnetic flux lines in magnetic field visualization;
- Xie et al. propose an ontology-based agricultural knowledge fusion method to enhance identification and fusion of new and existing data sets to make big data analytics more possible;
- Zhu and Xiong introduce the new discipline of Data Science, which provides a type of novel research method (data-intensive method) for natural and social sciences and takes research on data largely beyond computer science.

Data science has attracted the attention of researchers from different domains:

- Li and Lu propose a computational method to integrate both biomedical scientific data and literature for drug discovery and new uses of existing drugs;
- Meng et al. describe DarwinTree, which provides an integrated bioinformatics platform that supports all phases of the analytical pathway for phylogenetic study from data collection, phylogenetic

tree construction, visualization of the tree of life and web-based rendering, and to specific application service & data mining;
- Zhang and Zhao introduce data mining software and tools being applied in big data issues of Astrostatistics and Astroinformatics;
- Guo et al. introduce the National Rural Comprehensive Information Service Platform (NRCISP), supported by national science and technology support program.

There are novel applications in data science and related research areas:

- Peng et al. review recent advances of geophysical data and geophysical informatics developed in China;
- Guo et al. provide comfortable bus routes recommendation methods for passengers using an approach that combines multi-objective programming and a genetic algorithm in personalized information recommendation services;
- Huang et al. propose a novel trigger model with data mining techniques for sales prediction;
- Liu et al. offer an information filtering method and an aggregation model to provide a solution for how to choose appropriate experts for peer review;
- Wang Zhongya et al. introduce a novel personalization-oriented academic literature recommendation method to meet the user's preference in multiple dimensions simultaneously;
- Wang Haitao et al. propose an analysis architecture to make full use of data on natural environment corrosion of materials: the approach includes grey relational analysis, artificial neural network, fracture mechanics calculation;
- A knowledge model for literature data mining is proposed by Zhang et al. and it is applied to analyze the correlation between earthquake events and multidisciplinary data types.

These papers introduce the development of data science from a range of different perspectives. They help us get an overall understanding of the common or the most popular techniques used in data science research and applications. I hope you find them interesting and illuminating.

Prof. Li Jianhui

Secretary-General, CODATA-China
Director, Scientific Data Center, Computer Network information Center, Chinese Academy of Sciences

References

Cai & Zhu (2015) The Challenges of Data Quality and Data Quality Assessment in The Big Data Era. *Data Science Journal*, **14**:2. DOI: http://dx.doi.org/10.5334/dsj-2015-002

Guo et al. (2015) A Study of the Application of Big Data in a Rural Comprehensive Information Service. *Data Science Journal*, **14**:12. DOI: http://dx.doi.org/10.5334/dsj-2015-012

Guo et al. (2015) How to Find a Comfortable Bus Route – Towards Personalized Information Recommendation Services. *Data Science Journal*, **14**:14. DOI: http://dx.doi.org/10.5334/dsj-2015-014

Huang et al. (2015) A Novel Trigger Model for Sales Prediction Data Mining Techniques. *Data Science Journal*, **14**:15. DOI: http://dx.doi.org/10.5334/dsj-2015-015

Jiang et al. (2015) The Development of a Combined Search for a Heterogeneous Database. *Data Science Journal*, **14**:3. DOI: http://dx.doi.org/10.5334/dsj-2015-003

Li & Lu (2015) An Integrative Approach for Discovery of New Uses of Existing Drugs. *Data Science Journal*, **14**:9. DOI: http://dx.doi.org/10.5334/dsj-2015-009

Liu et al. (2015) How to Choose Appropriate Experts for Peer Review: An Intelligent Recommendation Method in a Big Data Context. *Data Science Journal*, **14**:16. DOI: http://dx.doi.org/10.5334/dsj-2015-016

Meng et al. (2015) DarwinTree: A Molecular Data Analysis and Application Environment for Phylogenetic Study. *Data Science Journal*, **14**:10. DOI: http://dx.doi.org/10.5334/dsj-2015-010

Peng et al. (2015) From Geophysical Data to Geophysical Informatics. *Data Science Journal*, **14**:13. DOI: http://dx.doi.org/10.5334/dsj-2015-013

Shen et al. (2015) OpenCSDB: Research on the Application of Linked Data in Scientific Databases. *Data Science Journal*, **14**:4. DOI: http://dx.doi.org/10.5334/dsj-2015-004

Wang et al. (2015) A Personalization-Oriented Academic Literature Recommended Method. *Data Science Journal*, **14**:17. DOI: http://dx.doi.org/10.5334/dsj-2015-017

Wang et al. (2015) Effective Utilization for Data of Natural Environment Corrosion of Materials. *Data Science Journal*, **14**:18. DOI: http://dx.doi.org/10.5334/dsj-2015-018

Xiang et al. (2015) The Executable Invocation Policy of Web Services Composition with Petri Net. *Data Science Journal*, **14**:5. DOI: http://dx.doi.org/10.5334/dsj-2015-005

Xie et al. (2015) Research on Agricultural Knowledge Fusion Method for Big Data. *Data Science Journal*, **14**:7.

Zhang & Zhao (2015) Astronomy in the Big Data Era. *Data Science Journal*, **14**:11. DOI: http://dx.doi.org/10.5334/dsj-2015-011

Zhang et al. (2015) Correlation Analysis Model on Multidisciplinary Data for Disaster Research. *Data Science Journal*, **14**:19. DOI: http://dx.doi.org/10.5334/dsj-2015-019

Zhong et al. (2015) An Improved Seeding Algorithm of Magnetic Flux Lines Based on Data in 3D Space. *Data Science Journal*, **14**:6. DOI: http://dx.doi.org/10.5334/dsj-2015-006

Zhu & Xiong (2015) Towards Data Science. *Data Science Journal*, **14**:8. DOI: http://dx.doi.org/10.5334/dsj-2015-008