

# Preface

In 2016 the XVIII International Conference on Data Analytics and Management in Data-Intensive Domains (DAMDID/RCDL 2016) was held during October 11–14 in the Holiday Center, Ershovo (Moscow region).

By tradition, the Data Analytics and Management in Data-Intensive Domains Conference (DAMDID) is planned as a multidisciplinary forum of researchers and practitioners from various domains of science and research, promoting cooperation and exchange of ideas in the area of data analysis and management in domains driven by data-intensive research. Approaches to data analysis and management being developed in specific data-intensive domains (DID) of X-informatics (such as X = astro, bio, chemo, geo, medicine, neuro, physics, etc.), social sciences, as well as in various branches of informatics, industry, new technologies, finance, and business contribute to the conference content.

The program of DAMDID/RCDL 2016 alongside with the traditional data management topics reflected a rapid move into the direction of data science and data-intensive analytics. Three invited plenary talks of the conference emphasized the key problems of methods and facilities of data analytics and management in DID. In the keynote of Andrey Rzhetsky (Professor of Genetic Medicine of Chicago University) that opened the conference, the approaches for studying the mechanisms of cancer were presented, which are being developed by the University of Chicago consortium in which Andrey Rzhetsky acts as the PI according to the DARPA Big Mechanism program (<http://www.darpa.mil/program/big-mechanism>). These approaches are centered on the development of contemporary methods of data analysis in DID including cognitive methods based on the understanding of the natural language resembling those of IBM Watson, but focused on the hypothesis formation based on the causal relationships detected in the texts, cancer mechanism modeling to automatically predict therapeutic clues, and application of robot scientist methods for the organization of experiments. The invited talk of Sophia Ananiadou, Director of the National Centre for Text Mining (NaCTeM) at Manchester University, opened the second day of the conference. NaCTeM acts as part of the University of Chicago consortium supported by DARPA. In this talk the approaches for automatic reconstruction of the pathway models are considered. These approaches are based on the discovery of the various kinds of relationships between the concepts of arbitrary nature in the texts. Finally, the program on the third day was opened with the keynote of Dimitrios Tzovaras (Director at the Information Technologies Institute of the Centre for Research and Technology Hellas in Thessaloniki) in which the analytical survey of the European strategy in the area of research infrastructures was presented.

The conference Program Committee reviewed 57 submissions and accepted 27 of them as full papers, 16 as short papers, three as posters, two as demos, while nine submissions were rejected. According to the conference program, these 43 oral presentations (of the full and short papers) were structured into 13 sessions including

Semantic Modeling in DID, Data Analysis Methods, Knowledge Management, Learning Management, Semantic Search and Navigation, Pattern Analysis in Recommender Systems, Research Data Infrastructures (in Astronomy, Astrophysics, Material Sciences, Earth Monitoring), Data Extraction from Texts, Data Integration and Sharing, and Text Analysis Systems. Most of the presentations showcased the results of research conducted in the organizations located in the territory of the Russian Federation including Dubna, Ekaterinburg, Irkutsk, Kazan, Moscow, Novosibirsk, Obninsk, Puschino, Tomsk, Tula, Saint Petersburg, Yaroslavl, and Zvenigorod.

All accepted papers are published in the local proceedings; all full papers are published in the CEUR Workshop volume 1752 (<http://ceur-ws.org/Vol-1752>). This CCIS volume is a post-conference proceedings containing 19 selected, revised, and extended papers. The process of preparing nine of them for this volume included translation from the Russian language into English. It is worth mentioning that Sophia Ananiadou kindly agreed to transform the extended abstract of her invited talk into a full paper.

The contents of the volume are structured into the following sections reflecting the wide spectrum of topics related to the DAMDID conferences: Semantic Modeling in Data-Intensive Domains, Knowledge and Learning Management, Text Mining, Data Infrastructures in Astrophysics, Data Analysis, and Research Infrastructures. The position paper by Yannis Manolopoulos is included separately, providing an introduction into bibliometrics (considering such related fields as scientometrics, citation analysis, journal impact factor) with a warning that the widely used metrics and rankings should be applied with great skepticism.

DAMDID/RCDL 2016 would not have been possible without the support of the Russian Foundation for Basic Research, the Federal Agency of Scientific Organizations of the Russian Federation, the National Research Nuclear University MEPhI, and the Federal Research Center Computer Science and Control of the Russian Academy of Sciences.

Finally, we thank Springer for publishing this proceedings volume, containing the invited and selected research papers, in their CCIS series, The Program Committee of the conference appreciates the possibility to use the Conference Management Toolkit (CMT) sponsored by Microsoft Research, which provided great support during various phases of the paper submission and reviewing process.

February 2017

Leonid A. Kalinichenko  
Sergei O. Kuznetsov  
Yannis Manolopoulos

Data Analytics and Management in Data Intensive  
Domains

XVIII International Conference, DAMDID/RCDL 2016,  
Ershovo, Moscow, Russia, October 11 -14, 2016,  
Revised Selected Papers

Kalinichenko, L.; Kuznetsov, S.O.; Manolopoulos, Y.  
(Eds.)

2017, XII, 281 p. 64 illus., Softcover

ISBN: 978-3-319-57134-8