

Preface

The XXIII International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL 2021) that was set to be held at the National University of Science and Technology MISIS, Moscow, Russia during October 26-29. However, because of the worldwide COVID-19 crisis, DAMDID/RCDL 2021 had to take place online.

DAMDID is 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, geo, med, neuro, physics, chemistry, material science, etc.), social sciences, as well as in various branches of informatics, industry, new technologies, finance, and business are expected to contribute to the conference content.

Previous DAMDID/RCDL conferences were held in St. Petersburg (1999, 2003), Protvino (2000), Petrozavodsk (2001, 2009), Dubna (2002, 2008, 2014), Pushchino (2004), Yaroslavl (2005, 2013), Pereslavl (2007, 2012), Kazan (2010, 2019), Voronezh (2011, 2020), Obninsk (2016), and Moscow (2017, 2018).

The program of DAMDID/RCDL 2021 was oriented towards data science and data-intensive analytics as well as on data management. The program of this year included three keynotes.

The keynote by Yibin Xu (deputy director of Research and Services Division of Materials Data and Integrated System, and the group leader of Data-Driven Inorganic Materials Research Group in National Institute for Materials Science, Japan) was devoted to the construction of integrated materials data system for data-driven materials research. Emille E. O. Ishida (CNRS/Laboratoire de Physique de Clermont (LPC), Université Clermont-Auvergne (UCA), Clermont Ferrand, France) gave a talk on supervised (and especially active) and unsupervised machine learning and their application in astronomy for classification problems and search for scientifically interesting anomalies. The keynote by Andrew Turpin (associate director of the Melbourne Connect, and director of the Melbourne Data Analytics Platform) discusses the development of a workforce of data and computer scientists that can support researchers at our university to make use of digital technology in their research.

The workshop on Data and Computation for Materials Science and Innovation (DACOMSIN) constitutes the first day of the conference on October 26. The workshop is aimed to address communication gap across communities in the domains of materials data infrastructures, materials data analysis, and materials in silico experiment. The workshop brings together professionals from across research and innovation to share their experience and perspectives of using information technology and computer science for materials data management, analysis and simulation.

The conference program committee consisted of scientists from 12 countries. The committee reviewed 63 submissions, and 37 submissions were accepted as full papers

or demos, 15 as short papers or posters, whereas 9 submissions were rejected or cancelled after reviewing.

According to the conference and workshops program, 58 oral presentations were structured into 13 sessions. Most of the presentations were dedicated to results of researches fulfilled in the research organizations located in Russia, including Kazan, Moscow, Novosibirsk, Obninsk, Tomsk, Tula, St. Petersburg, Petrozavodsk, Voronezh. Though the conference featured talks by the foreign researchers from Australia, Armenia, China, Finland, France, Germany, Japan, Italy, Sweden, United Kingdom.

For the sake of better visibility of the conference publications by the international scientific community two post-proceedings volumes are submitted to Springer's Communications in Computer and Information Science (CCIS) and CEUR Workshop Proceedings (<http://ceur-ws.org/>). Sixteen full papers are accepted into CCIS volume of the DAMDID/RCDL 2021 conference proceedings, and 33 papers were included into this CEUR volume are structured into eight sections: Conceptual modeling, Data Integration, Ontologies and Applications; Problem Solving Architectures and Infrastructures, Experiment Organization; Machine Learning Applications; Advanced Data Analysis Methods; Data Analysis in Astronomy; Ontologies, Databases and Data Analysis in Materials Science; Information Extraction from Texts; Scientific, Educational, and Literary Texts Analysis.

The chairs of Program Committee express their gratitude to the Program Committee members for reviewing of the submissions, to the authors, and to the host organizers from the NUST MISIS. The Program Committee appreciates the possibility of using the Conference Management Toolkit (CMT) sponsored by Microsoft Research, which provided great support during all phases of the paper submission and reviewing process. The Organizing Committee wants to gratefully acknowledge the sponsor of the conference Thermo-Calc Software AB for their generous support. Thermo-Calc Software's mission is to develop computational tools that allow engineers to generate the materials data they need in their daily decision making to drive innovation and improve product performance.

December 2021

Alexei Pozanenko
Sergey Stupnikov
Bernhard Thalheim
Eva Mendez
Nadezhda Kiselyova