

Alexei Pozanenko · Sergey Stupnikov ·  
Bernhard Thalheim · Eva Mendez ·  
Nadezhda Kiselyova (Eds.)

Communications in Computer and Information Science

1620

# Data Analytics and Management in Data Intensive Domains

23rd International Conference, DAMDID/RCDL 2021  
Moscow, Russia, October 26–29, 2021  
Revised Selected Papers

 Springer



Editorial Board Members

Joaquim Filipe 

*Polytechnic Institute of Setúbal, Setúbal, Portugal*

Ashish Ghosh

*Indian Statistical Institute, Kolkata, India*

Raquel Oliveira Prates 

*Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil*

Lizhu Zhou

*Tsinghua University, Beijing, China*

More information about this series at <https://link.springer.com/bookseries/7899>

Alexei Pozanenko · Sergey Stupnikov ·  
Bernhard Thalheim · Eva Mendez ·  
Nadezhda Kiselyova (Eds.)

# Data Analytics and Management in Data Intensive Domains


23rd International Conference, DAMDID/RCDL 2021  
Moscow, Russia, October 26–29, 2021  
Revised Selected Papers

*Editors*


Alexei Pozanenko   
Space Research Institute of the Russian  
Academy of Sciences  
Moscow, Russia

National Research University Higher School  
of Economics  
Moscow, Russia

Bernhard Thalheim   
Christian-Albrecht University of Kiel  
Kiel, Germany

Nadezhda Kiselyova   
A. A. Baikov Institute of Metallurgy  
and Materials Science of RAS (IMET RAS)  
Moscow, Russia

Sergey Stupnikov   
Federal Research Center “Computer Science  
and Control” of RAS  
Moscow, Russia

Eva Mendez   
Universidad Carlos III de Madrid  
Getafe, Spain

ISSN 1865-0929 ISSN 1865-0937 (electronic)  
Communications in Computer and Information Science  
ISBN 978-3-031-12284-2 ISBN 978-3-031-12285-9 (eBook)  
<https://doi.org/10.1007/978-3-031-12285-9>

© Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

## Preface

This CCIS volume published by Springer contains the proceedings of the XXIII International Conference on 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, 2021. 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, chemo, geo, medical, neuro, physics, chemistry, material science, social science, etc.), as well as in various other branches of informatics, industry, new technologies, finance, and business, 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 data management topics. The program of this year included three keynotes. The keynote by Yibin Xu (Deputy Director of the Research and Services Division of Materials Data and Integrated System and leader of the Data-Driven Inorganic Materials Research Group at the National Institute for Materials Science, Japan) was devoted to the construction of an integrated materials data system for data-driven materials research. Emille E. O. Ishida (CNRS, Laboratoire de Physique de Clermont, Université Clermont-Auvergne, France) gave a talk on supervised (and especially active) and unsupervised machine learning and their application in astronomy for classification problems and the search for scientifically interesting anomalies. The keynote by Andrew Turpin (Associate Director of the Melbourne Connect and Director of the Melbourne Data Analytics Platform) discussed the development of a workforce of data and computer scientists that can support researchers at university to make use of digital technology in their research.

The workshop on Data and Computation for Materials Science and Innovation (DACOMSIN) constituted the first day of the conference on October 26. The workshop aimed to address the communication gap across communities in the domains of materials data infrastructures, materials data analysis, and materials in silico experiments. The workshop brought 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, comprised of members from 12 countries, reviewed 63 submissions. In total, 37 submissions were accepted as full papers and 15 as short papers and posters.

According to the conference and workshops program, 58 oral presentations were grouped into 13 sessions. Most of the presentations were dedicated to the results of research conducted in organizations located in Russia, including Kazan, Moscow, Novosibirsk, Obninsk, Tomsk, Tula, St. Petersburg, Petrozavodsk, and Voronezh. However, the conference also featured talks prepared by foreign researchers from countries such as Australia, Armenia, China, Finland, France, Germany, Japan, Italy, Sweden, and the UK.

For the CCIS conference proceedings, 16 peer-reviewed papers have been selected by the Program Committee (an acceptance rate of 25%), which are structured into four sections: Problem Solving Infrastructures, Experiment Organization, and Machine Learning Applications (three papers); Data Analysis in Astronomy (five papers); Data Analysis in Material and Earth Sciences (four papers); and Information Extraction from Text (four papers).

We are grateful to the Program Committee members, for reviewing the submissions and selecting the papers for presentation, to the authors of the submissions, and to the host organizers from the National University of Science and Technology MISIS. We are also grateful for the use of the Conference Management Toolkit (CMT) sponsored by Microsoft Research, which provided great support during various 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.

June 2022

Alexei Pozanenko  
Sergey Stupnikov  
Bernhard Thalheim  
Eva Mendez  
Nadezhda Kiselyova

# Organization

## Program Committee Co-chairs

Alexei Pozanenko	Space Research Institute, RAS, and National Research University Higher School of Economics, Russia
Eva Mendez	Universidad Carlos III de Madrid, Spain

## Program Committee Deputy Chair

Sergey Stupnikov	Federal Research Center “Computer Science and Control” of RAS, Russia
------------------	--

## DACOMSIN Workshop Co-chairs

Nadezhda Kiselyova	Baikov Institute of Metallurgy and Materials Science, RAS, Russia
Vasily Bunakov	Science and Technology Facilities Council, UK
Alexandra Khvan	National University of Science and Technology MISIS, Russia

## Organizing Committee Co-chairs

Mikhail Filonov	National University of Science and Technology MISIS, Russia
Victor Zakharov	Federal Research Center “Computer Science and Control” of RAS, Russia

## Organizing Committee

Igor Abrikosov	National University of Science and Technology MISIS, Russia
Alexandra Khvan	National University of Science and Technology MISIS, Russia
Marina Nezhurina	National University of Science and Technology MISIS, Russia
Alex Kondratiev	National University of Science and Technology MISIS, Russia
Nikolay Skvortsov	Federal Research Center “Computer Science and Control” of RAS, Russia



Vladimir Cheverikin	National University of Science and Technology MISIS, Russia
Irina Bajenova	National University of Science and Technology MISIS, Russia

## **Coordinating Committee**

Igor Sokolov (Co-chair)	Federal Research Center “Computer Science and Control” of RAS, Russia
Nikolay Kolchanov (Co-chair)	Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Sergey Stupnikov (Deputy Chair)	Federal Research Center “Computer Science and Control” of RAS, Russia
Arkady Avramenko	Pushchino Radio Astronomy Observatory, RAS, Russia
Pavel Braslavsky	Ural Federal University and SKB Kontur, Russia
Vasily Bunakov	Science and Technology Facilities Council, UK
Alexander Elizarov	Kazan (Volga Region) Federal University, Russia
Alexander Fazliev	Institute of Atmospheric Optics, SB RAS, Russia
Alexei Klimentov	Brookhaven National Laboratory, USA
Mikhail Kogalovsky	Market Economy Institute, RAS, Russia
Vladimir Korenkov	Joint Institute for Nuclear Research, Russia
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Vladimir Litvine	Evogh Inc., USA
Archil Maysuradze	Moscow State University, Russia
Oleg Malkov	Institute of Astronomy, RAS, Russia
Alexander Marchuk	Institute of Informatics Systems, SB RAS, Russia
Igor Nekrestjanov	Verizon Corporation, USA
Boris Novikov	St. Petersburg State University, Russia
Nikolay Podkolodny	Institute of Cytology and Genetics, SB RAS, Russia
Aleksey Pozanenko	Space Research Institute, RAS, Russia
Vladimir Serebryakov	Dorodnicyn Computing Center of RAS, Russia
Yury Smetanin	Russian Foundation for Basic Research, Moscow
Vladimir Smirnov	Yaroslavl State University, Russia
Bernhard Thalheim	Kiel University, Germany
Konstantin Vorontsov	Moscow State University, Russia
Viacheslav Wolfengagen	National Research Nuclear University “MEPhI”, Russia
Victor Zakharov	Federal Research Center “Computer Science and Control” of RAS, Russia

## Program Committee

Alexander Afanasyev	Institute for Information Transmission Problems, RAS, Russia
Ladjel Bellatreche	National Engineering School for Mechanics and Aerotechnics, France
Dmitry Borisenkov	RELEX Group, Russia
Pavel Braslavski	Ural Federal University, Russia
Vasily Bunakov	Science and Technology Facilities Council, UK
Kheeran Dharmawardena	Cytrax Consulting, Australia
Boris Dobrov	Lomonosov Moscow State University, Russia
Alexander Elizarov	Kazan Federal University, Russia
Alexander Fazliev	Institute of Atmospheric Optics, SB RAS, Russia
Evgeny Gordov	Institute of Monitoring of Climatic and Ecological Systems SB RAS, Russia
Valeriya Gribova	Institute of Automation and Control Processes, FEBRAS, Far Eastern Federal University, Russia
Maxim Gubin	Google Inc., USA
Sergio Ilarri	University of Zaragoza, Spain
Mirjana Ivanovic	University of Novi Sad, Serbia
Jeyhun Karimov	Huawei Research Center, Germany
Vitaliy Kim	National Research University Higher School of Economics, Russia, and Fesenkov Astrophysical Institute, Kazakhstan
Nadezhda Kiselyova	Institute of Metallurgy and Materials Science, RAS, Russia
Alexei Klimentov	Brookhaven National Laboratory, USA
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Giuseppe Longo	University of Naples Federico II, Italy
Evgeny Lipachev	Kazan Federal University, Russia
Natalia Loukachevitch	Lomonosov Moscow State University, Russia
Ivan Lukovic	University of Belgrade, Serbia
Oleg Malkov	Institute of Astronomy, RAS, Russia
Sergey Makhortov	Voronezh State University, Russia
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece
Archil Maysuradze	Lomonosov Moscow State University, Russia
Manuel Mazzara	Innopolis University, Russia
Mikhail Melnikov	Institute of Molecular Biology and Biophysics, SB RAS, Russia
Alexey Mitsyuk	National Research University Higher School of Economics, Russia
Alexander Moskvitin	Special Astrophysical Observatory, RAS, Russia

Xenia Naidenova	Kirov Military Medical Academy, Russia
Dmitry Namiot	Lomonosov Moscow State University, Russia
Boris Novikov	National Research University Higher School of Economics, Russia
Panos Pardalos	University of Florida, USA
Jaroslav Pokorny	Charles University in Prague, Czech Republic
Natalya Ponomareva	Research Center of Neurology, Russia
Roman Samarev	Bauman Moscow State Technical University, Russia
Vladimir Serebryakov	Dorodnicyn Computing Centre of RAS, Russia
Nikolay Skvortsov	Federal Research Center “Computer Science and Control” of RAS, Russia,
Manfred Sneps-Sneppe	Ventspils University of Applied Sciences, Latvia
Valery Sokolov	Yaroslavl State University, Russia
Kirill Sokolovsky	Michigan State University, USA
Alexander Sychev	Voronezh State University, Russia
Bernhard Thalheim	University of Kiel, Germany
Alexey Ushakov	University of California, Santa Barbara, USA
Pavel Velikhov	TigerGraph, Russia
Alina Volnova	Space Research Institute, RAS, Russia
Alexey Vovchenko	Federal Research Center “Computer Science and Control” of RAS, Russia
Vladimir Zadorozhny	University of Pittsburgh, USA
Yury Zagorulko	Institute of Informatics Systems, SB RAS, Russia
Victor Zakharov	Federal Research Center “Computer Science and Control” of RAS, Russia
Sergey Znamensky	Institute of Program Systems, RAS, Russia
Mikhail Zymbler	South Ural State University, Russia

## **DACOMSIN Workshop Program Committee**

Igor Abrikosov	Linköping University, Sweden
Toshihiro Ashino	Toyo University, Japan
Keith Butler	Science and Technology Facilities Council, UK
Victor Dudarev	HSE University, Russia
Jennifer Handsel	Science and Technology Facilities Council, UK
Martin Horsch	STFC Daresbury Laboratory, UK
Natalie Johnson	The Cambridge Crystallographic Data Centre, UK
Francesco Mercuri	Consiglio Nazionale delle Ricerche, Italy
Igor Morozov	Joint Institute for High Temperatures of RAS, Russia
Artem Oganov	Skoltech, Russia

Sergey Stupnikov

Federal Research Center “Computer Science and Control” of RAS, Russia

Irina Uspenskaya

Lomonosov Moscow State University, Russia

Yibin Xu

National Institute for Materials Science, Japan

## **Supporters**

National University of Science and Technology MISIS, Moscow, Russia

Thermo-Calc Software AB

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, Russia

Moscow ACM SIGMOD Chapter

# Contents

## **Problem Solving Infrastructures, Experiment Organization, and Machine Learning Applications**

MLDev: Data Science Experiment Automation and Reproducibility Software .....	3
<i>Anton Khritankov, Nikita Pershin, Nikita Ukhov, and Artem Ukhov</i>	
Response to Cybersecurity Threats of Informational Infrastructure Based on Conceptual Models .....	19
<i>Nikolay Kalinin and Nikolay Skvortsov</i>	
Social Network Analysis of the Professional Community Interaction—Movie Industry Case .....	36
<i>Ilya Karpov and Roman Marakulin</i>	
<b>Data Analysis in Astronomy</b>	
Cross-Matching of Large Sky Surveys and Study of Astronomical Objects Apparent in Ultraviolet Band Only .....	53
<i>Aleksandra S. Avdeeva, Sergey V. Karpov, Oleg Yu. Malkov, and Gang Zhao</i>	
The Diversity of Light Curves of Supernovae Associated with Gamma-Ray Bursts .....	74
<i>Sergey Belkin and Alexei Pozanenko</i>	
Application of Machine Learning Methods for Cross-Matching Astronomical Catalogues .....	92
<i>Aleksandra Kulishova and Dmitry Briukhov</i>	
Pipeline for Detection of Transient Objects in Optical Surveys .....	104
<i>Nicolai Pankov, Alexei Pozanenko, Vladimir Koupryanov, and Sergey Belkin</i>	
VALD in Astrophysics .....	135
<i>Yury Pakhomov and Tatiana Ryabchikova</i>	

**Data Analysis in Material and Earth Sciences**

Machine Learning Application to Predict New Inorganic Compounds – Results and Perspectives .....	151
<i>Nadezhda Kiselyova, Victor Dudarev, and Andrey Stolyarenko</i>	

Interoperability and Architecture Requirements Analysis and Metadata Standardization for a Research Data Infrastructure in Catalysis .....	166
<i>Martin Horsch, Taras Petrenko, Volodymyr Kushnarenko, Bjoern Schembera, Bianca Wentzel, Alexander Behr, Norbert Kockmann, Sonja Schimmler, and Thomas Bönisch</i>	

Fast Predictions of Lattice Energies by Continuous Isometry Invariants of Crystal Structures .....	178
<i>Jakob Ropers, Marco M. Mosca, Olga Anosova, Vitaliy Kurlin, and Andrew I. Cooper</i>	

Image Recognition for Large Soil Maps Archive Overview: Metadata Extraction and Georeferencing Tool Development .....	193
<i>Nadezda A. Vasilyeva, Artem Vladimirov, and Taras Vasiliev</i>	

**Information Extraction from Text**

Cross-Lingual Plagiarism Detection Method .....	207
<i>Denis Zubarev, Ilya Tikhomirov, and Ilya Sochenkov</i>	

Methods for Automatic Argumentation Structure Prediction .....	223
<i>Ilya Dimov and Boris Dobrov</i>	

A System for Information Extraction from Scientific Texts in Russian .....	234
<i>Elena Bruches, Anastasia Mezentsseva, and Tatiana Batura</i>	

Improving Neural Abstractive Summarization with Reliable Sentence Sampling .....	246
<i>Daniil Chernyshev and Boris Dobrov</i>	

<b>Author Index</b> .....	263
---------------------------	-----