Agenda

Workshop

Big Data and Systems Analysis

24-25 February 2020
Gvishiani Room,
International Institute for Applied Systems Analysis (IIASA),
Schlossplatz 1, 2361 Laxenburg, Austria

Monday, 24 February

09:00	**Minibus transfer from hotel (Novum Hotel Prinz Eugen) to IIASA (Laxenburg)**
10:00-10:15	**Welcome and introduction** *Chair: Albert van Jaarsveld* **Albert van Jaarsveld,** Director General and Chief Executive Officer, IIASA **John Broome**, Treasurer, CODATA **Fred Roberts** and **Igor Sheremet,** Co-chairs, Task Group on Advanced mathematical tools for data-driven applied systems analysis
10:15-11:25	**Session on Resilience and Sustainability – Part I** *Chair: Igor Sheremet*
10:15-10:50	**Fred Roberts**, Director Emeritus and Special Advisor, Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, USA **Resilience in complex linked systems** Today's society has become dependent on complex systems, enabled by increased digitization of our world, that have had a great impact on virtually all facets of our lives. The result is instant communication, ability to move money anywhere and quickly, ability to ask a machine to make our shopping list or turn on our favorite music, etc. Yet these changes have made us vulnerable: To natural disasters, deliberate attacks, just plain errors. In recent years, "resilience" of complex natural and social systems has become a major area of emphasis. One approach to resilience is to develop algorithms for responding to a disruption that will minimize the departure from the previous state when things settle down. This talk will illustrate this point with examples built around models using graphs and networks and apply them to disruptions involving spread of disease, fires, faults in the power grid, and damage to interdependent critical infrastructure.
10:50-11:25	**Nebojsa Nakicenovic**, Emeritus Research Scholar, Transition to New Technologies Program, IIASA, Austria **Digital revolution - a major transformation toward the world in 2050 (TWI2050)** The 2030 Agenda universally adopted in 2015 offers a vision of the future with its 17 Sustainable Development Goals. The objective of The World In 2050 (TWI2050) initiative is to develop transformational pathways toward achieving all 17 SDGs by using an integrated and systems approach. TWI2050 was established by the International Institute for Applied Systems Analysis (IIASA)

	to provide scientific foundations for the 2030 Agenda and policy advise for achieving SDGs to avoid potential conflicts among the 17 goals and reap the benefits of potential synergies of achieving them in unison. It is based on the voluntary and collaborative effort of more than 60 authors from about 20 institutions, and more than 100 independent experts from academia, business, government, intergovernmental and non-governmental organizations from all the regions of the world. The last TWI2050 report, launched at the 2019 UN High-level Political Forum, focuses on the Digital Revolution as the potential enabler of the 17 SDGs. It analyzes the potential integration of digital technologies into new systems and human activities and the ongoing advances in AI, connectivity, digitization of information, additive manufacturing (3D printing), virtual reality, Internet of things (IoT), machine learning, block chain, robotics, quantum computing and synthetic biology. A good example of the power of digitalization is the disruptive nature of information and communication technologies (ICT) and their gigantic potential effect in enabling fundamentally new human activities while reducing specific energy and materials requirements as well as greenhouse-gas emissions compared with the analogue devices they replace. We extend the analysis to include the potential effects of the new and advanced Digital Revolution that started with the ICT. Humanity may be moving toward new civilizational thresholds. Super-intelligent machines might even develop a life of their own, with the capacity to harm and benefit human agents. The digital transformation calls for a comprehensive set of regulatory standards and normative frameworks, physical infrastructure, and digital systems, to capture the benefits of the digital revolution while avoiding the many potential downsides. An essential priority should be to develop science, technology and innovation roadmaps to better understand the potential benefits and dangers of digitalization and how to leverage the digital revolution toward sustainable development for all. The principles of digital transformation for sustainable development have yet to be written as humanity embarks on the fundamental transformation. If the potential opportunities are seized, the transformational change would benefit people and the planet.
11:25-11:55	**Break**
11:55-13:05	**Session on Resilience and Sustainability – Part II** *Chair: Igor Sheremet*
11:55-12:30	**Mubbasir Kapadia**, Director, Intelligent Visual Interfaces Lab, Computer Science Department, Rutgers University, USA **Intelligent human-aware building management and design** In an increasingly populated world, civil engineers, architects, and urban planners must intrinsically account for the ergonomics of an environment with respect to the presence, occupancy, and movement of inhabitants who utilize these spaces. Building operations managers must factor in human occupancy and other key performance indicators to ensure that the function and wellbeing of inhabitants are considered in day-to-day operations. The next generation of buildings will be able to harness the maturity in sensing solutions, IoT, and robotics, to monitor the behavior of its inhabitants, and respond in real-time. For example, HVAC systems may dynamically alter the temperature conditions in a room to cater to the presence and occupancy of people. Smart signage systems can adapt to the relative distribution of people in a building, during an emergency evacuation. A key challenge is to harness the raw sensor data of human behavior, and transform it into actionable insights for building operations. In this talk, I will describe our ongoing work in developing data-driven computational models of human behavior, information-theoretic approaches for analyzing human behavior dynamics, and computer-assisted optimization of environments that intrinsically account for the movement and occupancy of its inhabitants, to inform building management and design.
12:30-13:05	**John Broome**, Treasurer, CODATA, Canada **CODATA: Leading the advancement of open science through FAIR data** The term “open data” first appeared in 1995, but during the last decade the concept has been refined and embraced by the scientific community as data-driven research has grown to address global scientific challenges such as the Sustainable Development Goals. As data became increasingly open

and abundant, it became clear that openness alone would not result in data that could be efficiently integrated and analyzed cross-discipline as required for productive open science. To be truly useful, data must be *findable*, *accessible*, *interoperable* and *reusable* - or "FAIR". Each of the 4 components of FAIR pose a different set of challenges that can be resolved only through a partnership approach involving many global organizations each with their network of experts and unique skill sets.

CODATA exists to promote global collaboration to advance open science and to improve the availability and usability of data for all areas of research. Over many years, as the *Data Committee of the ISC*, CODATA has worked on developing, sharing, publishing and implementing solutions to many of the challenges that must be addressed to establish a global FAIR data infrastructure. A few examples of supporting CODATA activities include; the Data Policy Committee, International Data Week conferences, the Data Science Journal, Data Science Summer Schools, and many task groups focused on a diverse range of data issues and opportunities.

With this background, CODATA has become one of the leading organizations supporting adoption of the FAIR principles. CODATA is actively engaged in defining policies, technology, and practices that will contribute to creation of a global data infrastructure that is both FAIR and multidisciplinary. Looking to the future, CODATA is now partnering with the ISC and other interested organizations, to embark on a *Decadal Programme* that will establish the ecosystem of resources required to facilitate and accelerate the advanced human and machine-assisted analysis of cross-domain data required to understand complex global systems.

13:05-14:35 **Lunch at IIASA Schloss Restaurant Club Room**

14:35-15:45 **Session on Data and Decisions – Part I** *Chair: Fred Roberts*

14:35-15:10

Tatiana Ermolieva, Research Scholar, Ecosystems Services and Management Program, IIASA, Austria

Non-smooth stochastic optimization, quantile-based performance function and iterative stochastic quasigradient procedure for robust machine learning and decision making: application for robust distributed models' linkage and stochastic crop yield generator

Tatiana Ermolieva, Yuri Yermoliev, Petr Havlik, Elena Rovenskaya, Esther Boere, Juraj Balkovic, Rastislav Skalský

Big Data and information are generated continuously. The information may come with errors, from different sources, with missing and inconsistent values. Therefore, efficient adequate optimization methods are needed to solve the resultant large-scale machine learning (ML) problems maximizing the performance of an estimator model. In general, optimization problems in ML correspond to the optimization of an appropriate performance (loss, goodness-of-fit, error) function that measures the quality of learning solution. Stochastic Gradient method (SG) of smooth stochastic optimization has been traditionally used for large-scale ML problems characterized by smooth performance functions. However, it is often important to minimize non-smooth quantile-based performance functions. They enable to achieve robust systems' performance with respect to endogenous and exogenous systemic shocks and derive a desirable joint risk distribution of systemic responses. The shocks can result from climate- or weather-related extreme events or threats due to a lack of systemic planning, appropriate regulations and deregulations. For example, the quantile-based performance functions and an iterative Stochastic Quasigradient (SQG) "learning" procedure of non-smooth stochastic optimization have been used for robust linking (RL) of distributed sectorial and regional systems (Intelligent Agents (IA)) under uncertainty and asymmetric information. Without the RL, one-by-one individual myopic decisions of distributed systems (regarding e.g. capacity expansion, resource utilization, risk mitigation and sharing) can be suboptimal to the overall system and lead to disastrous violation of regional constraints, cause lock-in situations and irreversibilities. The iterative SQG-based RL can be considered as a tool for negotiations between decentralized IAs to adopt robust decisions maximizing both individual and global short- and long-term goals. The iterative SQG procedures also allow for joint treatment (learning, modeling, optimization) of probabilistic and non-probabilistic marginal and multivariate probability distributions (say, distributions of flood losses, crop yield losses, precipitation, incomes, etc.). By using appropriate similarity measures (kernel estimators) and advanced support-

	vector machines, the SQGs and quantile-based indicators update robust solutions towards newly arriving information in a minimum cost path.
15:10-15:45	**Konstantin Mischaikow**, Professor, Department of Mathematics, Rutgers University, USA **Data, nonlinear dynamics, and algebraic topology** A fundamental challenge in data-driven systems analysis is that knowledge of the system is derived from a finite set of observations that may or may not constitute a representative sample of the behavior of the system and yet many of the variables of interest are of a continuum nature, e.g.\ space, time, temperature, or are conceptualized as continua, e.g.\ cost, population. In essence, in the context of applications we are being asked to make justifiable decisions based on limited knowledge of the model being used to make the decision. I will highlight how algebraic topology in the form of homology can be used to extract consistent information about the structure of spaces, e.g.\ the geometry of nonlinear data sets, and actions on these space, e.g.\ nonlinear dynamics, from finite data. Time permitting, I will explain these ideas in the context of complex spatio-temporal patterns from convective flow and regulatory networks coming from systems biology.
15:45-16:15	**Break**
16:15-16:50	**Session on Data and Decisions – Part II** *Chair: Fred Roberts* **Nadejda Komendantova**, Research Scholar, Advanced Systems Analysis Program, and **Love Ekenberg**, Senior Research Scholar, Risk and Resilience Program, IIASA, Austria **Towards compromise-oriented policy solutions: application of Multi Criteria Decision Analysis in energy transition and media policy** Deployment of changed socio-economic conditions must lead to transitions and transformations of entire sectors. Such transitions are complex processes, which has political, social, economic and technical dimensions and involve a multitude of stakeholders. Therefore, a holistic, inclusive and comprehensive governance approach to such transformation is essential, since unguided significant socio-technical transition process will lead to many frictions and conflicts. Such changes will thus lead to a socio-technological transition processes, which are combined with and emphasised by shifts in technologies, business models, governance structures, consumption patterns, values and worldviews. Thus, such multi-stakeholder, multi-criteria situations are typical for the planning and decision processes involved herein and a significant issue is of course what methodologies to use. A multitude of methods for analysing and solving decision problems with multiple criteria have been suggested during the last decades. A common approach is to make preference assessments by specifying a set of attributes that represents the relevant aspects of the possible outcomes of a decision. Value functions are then defined over the alternatives for each attribute and a weight function is defined over the attribute set. One option is to simply define a weight function by fixed numbers on a normalised scale and then define value functions over the alternatives, where these are mapped onto fixed values as well, after which these values are aggregated and the overall score of each alternative is calculated. In this contribution we are discussing the multi-criteria decision analysis (MCDA) method and its application through various projects such as energy transition involving compromise solutions for various alternatives or media policy involving new sets of decision-support tools and instruments to deal with existing and emerging factors such as misinformation or impacts of social media.
16:50-18:00	**Discussion of next steps for the Task Group** *Chair: Igor Sheremet*

18:00-20:30	**Workshop dinner at Heuriger Restaurant GaumenPunkt** Neudorferstraße 10, 2340 Mödling.

Tuesday, 25 February

08:30	**Bus transfer from hotel (Novum Hotel Prinz Eugen) to IIASA (Laxenburg)**
09:30-10:45	**Session on Big Data and Fairness** *Chair: Fred Roberts*
09:30-10:05	**Alexei Gvishiani**, Chief Scientist and Chair of the Scientific Council, Geophysical Center, Russian Academy of Sciences, Russia **Big data, FAIR data and open data for systems analysis** The talk is devoted to formalization of the notions "Big Data", "Open Data" "FAIR Data" and "Systems analysis". These notions are nowadays widely used in planning and reporting scientific, business and media activities. How are they really linked to each other? Are we in the position and how to compare them? The talk makes an attempt to give answers to these questions. Some geosciences examples are given to illustrate the point. Mathematics is always being used as a major tool to explore and exploit the data. The talk addresses similarities and differences in mathematical statuses in IXX – XX – XXI centuries, as far as data are concerned. Physics was a source of the IXX century math. Mathematical modeling gave birth and intensive development to theoretical mathematics in XX century. Big Data studies are giving motivation and basis to systems analysis as new discrete mathematics of XXI century.
10:05-10:40	**Igor Sheremet**, Deputy Director for Science, Russian Foundation for Basic Research, Russia **Multigrammatical framework for systems assessment and optimization** Socio-technological system (STS) in the most general case may be considered as a collection of its organizational structure, technological base, resource base, and input flow of messages to be processed (orders to be completed). The most essential problems concerned with STS are assessment of their producing capabilities and resource consumption, optimal planning and scheduling for low-cost completion of orders, as well as assessment of STS sustainability to the destructive impacts and recoverability after such impacts. A unified mathematical toolkit called filtering multiset grammars (FMG) and developed for the description and solution of the aforementioned problems is considered. FMG were designed as a widely applied knowledge representation model integrating best features of classical mathematical programming, modern knowledge engineering and scheduling theory. Techniques of FMG application and key implementation issues are described. A Big Knowledge concept being the next natural step of the evolution of Big Data is discussed.
10:40-11:10	**Break, group photo**
11:10-12:40	**Session on FAIR Data Stewardship – Part I** *Chair: Alena Rybkina*
11:10-11:45	**Simon Hodson**, Executive Director, CODATA, France **Good practice for FAIR data stewardship** This talk will discuss the drivers, benefits and recent developments in good practice for data stewardship. Ultimately, data stewardship is an essential part of the scientific process, underpinning

	transparency and reproducibility. There is ample evidence for the scientific, innovation and economic benefits of good data stewardship that ensures data is as open as possible (and as closed as necessary) and FAIR. Good data stewardship requires a systemic approach and reposes on a number of concepts which will be discussed. These include Open Science, FAIR, the FAIR ecosystem, community interoperability frameworks, the research data lifecycle and data stewardship competence centres.
11:45-12:20	**Barend Mons**, President, CODATA, Netherlands **The internet of FAIR Digital Objects** Science and innovation are increasingly dependent on complex data and on machines to analyze it. There is a need to design and develop large integrative infrastructures for the digital artifacts that are populating the Internet. Such integrative infrastructures will facilitate findability, accessibility, interoperability and reusability (FAIR) of data and other objects, especially for machines. For at least two decades, scientists doing data-driven research have been aware of the need to improve this integration and have been taking a variety of measures. This has led to an emergence and mixing (creolization) of standards, tools, interfaces, etc. This fragmentation and heterogeneity have been aggravated by new technologies and commercial offerings, which are increasingly seen as hampering cross-discipline data-driven science and requiring urgent measures for harmonization and convergence. An analysis of earlier examples of building large integrative infrastructures, including the internet as we know it, shows that convergence can often be achieved by employing the well-known hourglass strategy, i.e., the convergence mechanisms need to be minimal, allowing maximum flexibility for both implementations and applications. Years of discussions led to the FAIR Principles, which are now agreed to be an essential element of this convergence process. Years of discussions also led to the concept and definition of Digital Objects (DO). It also became evident that the DO concept was reasonably compliant with the FAIR principles and that we could build on it by adding principles of semantic relationships, such as employed by the Linked Data Platform, to get to a FAIR Digital Object Framework (FDOF). FDOF has the potential to work as the center of the emerging Internet of FAIR Data and Services, allowing designers to choose alternative implementation paths and yet remain interoperable.
12:20-12:30	**Albert van Jaarsveld,** Director General and Chief Executive Officer, IIASA, Austria **IIASA Data Strategy moving forward**
12:30-12:40	**Michaela Rossini,** Head, Library and Knowledge Resources Unit, IIASA, Austria **Open access to IIASA research**
12:40-14:10	**Lunch at IIASA Schloss Restaurant Club Room**
14:10-16:00	**Session on FAIR Data Stewardship – Part II** *Chair: Alena Rybkina*
14:10-14:45	**Daniel Huppmann**, Research Scholar, Energy Program, IIASA, Austria **IIASA as a FAIR data hub for the energy systems & integrated assessment communities** For more than a decade, IIASA has supported the research communities working on energy systems, integrated assessment and climate change mitigation by providing infrastructure and support as a data curation hub. High-profile use cases include several reports by the Intergovernmental Panel on Climate Change (IPCC), successive rounds of the Energy Modeling Forum (EMF) and numerous projects funded by the EU's Horizon 2020 program. Supporting the IPCC's Special Report on 1.5°C (SR15), IIASA and the Integrated Assessment Modeling Consortium (IAMC) compiled and curated a large scenario ensemble of quantitative emissions pathways. Many of the headline statements widely reported in the media, like the imperative to reduce greenhouse gas emissions by 50% in a decade, are based on this scenario resource. This talk will use the scenario ensemble for the SR15 and related tools as an illustrative example of the work done at IIASA towards open science. It highlights the many steps taken to ensure the

	FAIRness, transparency and reproducibility of the IPCC assessment and provides an overview of the suite of open tools to facilitate exploration of the scenario ensemble and communication of its findings (and the underlying data and methods) to various stakeholder groups. It also showcases how these tools were implemented following the FAIR principles for open, collaborative research.
14:45-15:20	**Steffen Fritz**, Acting Program Director, Ecosystems Services and Management Program, IIASA, Austria **Monitoring SDGs using non-traditional datasets** Traditional data sources are not sufficient for measuring the United Nations Sustainable Development Goals. New and nontraditional sources of data are required. Citizen science is an emerging example of a non-traditional data source that is already making a contribution. In this Perspective, we present a roadmap that outlines how citizen science can be integrated into the formal Sustainable Development Goals reporting mechanisms. Success will require leadership from the United Nations, innovation from National Statistical Offices and focus from the citizen-science community to identify the indicators for which citizen science can make a real contribution.
15:20-16:00	**Discussion**
16:00-16:30	**Break**
16:30-18:15	**Session on CODATA Decadal Program** *Chair: Albert van Jaarsveld*
16:30-17:05	**Simon Hodson**, Executive Director, CODATA, France **CODATA Decadal Programme: Making data work for cross-domain grand challenges** The major global scientific and human challenges of the 21st century can only be addressed through cross-domain research that seeks to understand complex systems through machine-assisted analysis at scale. Our capacity for such analysis is currently constrained by the limitations in our ability to access and combine heterogenous data within and across domains. Major consensus-building effort will be needed to enable an ecosystem of resources that would enable data to be FAIR for humans and machines, in particular to gain agreement about the core technologies and semantic solutions which will allow data to be combined for cross-domain research. CODATA, with the support of and on behalf of ISC, proposes a major, global, decadal programme — "Making data work for cross-domain grand challenges" — to address these challenges. The programme will take a three-pronged approach (though engagement and cross-fertilisation between these areas of activity will be essential), Theme 1) Enabling Technologies and Good Practice for Data-Intensive Science, Theme 2) Mobilising Domains and Breaking Down Silos, Theme 3) Advancing Interoperability Through Cross-Domain Case Studies. The impact of the programme will be to assist the scientific and innovation communities to accelerate scientific understanding through a step change in the application of interdisciplinary data-intensive methodologies and thereby enable more efficient and transparent science to address global challenges. CODATA is now charged to put in place the core funding, capacity and partnerships in order to launch a Decadal Programme at the ISC General Assembly in Oman in October 2021. As part of this process of partnership building, CODATA invites discussions with IIASA to explore the potential for collaboration on this initiative.
17:05-18:15	**Discussion**
18:15	**Transfer from IIASA to Airport / Hotel**