Precision Convergence Webinar Series

STOP SHARING DATA: *Visiting Algorithms, Swarm Learning and Next Generation FAIR (Federated AI Ready) Principles and Practice*

By ***Prof. Dr. Barend Mons***
Leiden University

October 31, 2024, Thursday | 11 AM to 1 PM EST (2 hours in duration)

For Remote Participation, please register HERE

ABSTRACT: The rapid developments in the field of machine learning have also brought along some existential challenges, which are in essence all related to the broad concept of 'trust'. Aspects of this broad concept include trust in the output of any ML process (and the prevention of black boxes, hallucinations and so forth). The very trust in science is at stake, especially now that LLMs can generate 'good-looking nonsense' and paper mills come up in response to the perverse reward systems in current research environments. The other side of the same coin is that ML, if nor properly controlled will also break through security and privacy barriers and violate GDPR and other Ethical, Legal and Societal barriers, including equitability. In addition, the existence of data 'somewhere' by no means automatically implies its actual Reusability. This includes the by now well established four elements of the FAIR principles: Much data is not even **F**indable, if found, not **A**ccessible under well defined conditions, and if accessed not **I**nteroperable (understandable by third parties and machines) and this results in the vast majority of data and information not being **R**eusable without violation of copyrights, privacy regulations or the basic conceptual models that implicitly or explicitly underpin the query or the deep learning algorithm. Now that more and more data will also be 'independently' used by machines, all these challenges will be severely aggravated. This keynote will address how 'data visiting' as opposed to classical 'data sharing', which carries the connotation of data downloads, transport and loosing control, mitigates most, if not all, the unwanted side effects of classical 'data sharing'. For federated data visiting, the data should be FAIR in an additional sense or perspective, they should be '**F**ederated, **AI**-**R**eady', so that visiting algorithms can answer questions related to Access Control, Consent, Format, and can read rich (FAIR) metadata about the data itself to determine whether they are 'fit for purpose' and machine actionable (i.e. FAIR digital Objects, or Machine Actionable Units). The 'fitness for purpose' concept goes way beyond (but includes) information about methods, quality, error bars etc. The 'immutable logging' of all operation of visiting algorithms is crucial, especially when self learning algorithms in 'swarm learning' are being used. Enough to keep us busy for a while.
https://www.nature.com/articles/s41586-021-03583-3

PRESENTER: Pr. Dr. **Barend Mons,** FISC (born 1957, The Hague) is a molecular biologist by training and a leading FAIR data specialist. The first decade of his scientific career he spent on fundamental research on malaria parasites and later on translational research for malaria vaccines. In the year 2000 he switched to advanced data stewardship and (biological) systems analytics. He is most known for innovations in scholarly collaboration, especially nanopublications, and knowledge graph based discovery. In 2014 he organised the seminal FAIR conference at the Lorentz centre that led to the FAIR data initiative and GO FAIR. In 2012 he was appointed as full Professor in biosemantics in the Department of Human Genetics at the Leiden University Medical Center (LUMC) in The Netherlands.[1][2][3] In 2015 Barend was appointed chair of the High Level Expert Group on the European Open Science Cloud. Since 2017 Barend is heading the International Support and Coordination office of the GO FAIR initiative as director of the GO FAIR Foundation. From 2018-2023 he was the elected president of CODATA, the affiliated organisation on research data related issues of the International Science Council.. He has also been the European representative in the Board on Research Data and Information (BRDI) of the National Academies of Science Engineering and Medicine in the USA. In 2024, he was appointed as Fellow of the International Science Council. At his retirement in 2024 he was Knighted by the Dutch King in the 'Order of the Dutch Lion', the oldest and highest reward for cultural and scientific contributions to the international society. He is currently leading the Leiden Institute for FAIR and Equitable Science. He is a frequent keynote speaker about FAIR and open science around the world, and continues to participates in various national and international boards.

About the series: The precision convergence series is launched to catalyze unique synergy between, on the one hand, novel partnerships across sciences, sectors and jurisdictions around targeted domains of real-world solutions,and on the other hand, a next generation convergence of AI with advanced research computing and other data and digital architectures such as PSC's Bridges-2, and supporting data sharing frameworks such as HuBMAP, informing in a real time as possible the design, deployment and monitoring of solutions for adaptive real-world behavior and context.

The McGill Centre for the Convergence of Health and Economics (MCCHE) is a virtual world network of scientist, action and policy leaders promoting the weaving of digital-powered interdisciplinary science into person-centered domain-specific solutions at scale to global challenges faced by traditional and modern economy and society worldwide. The MCCHE stimulates lasting collaborations that bridge the many divides in the market, economy, and society that are at the root of these most pressing modern challenges through collaborative of modular convergence innovation platforms.

The Pittsburgh Supercomputing Center is a joint computational research center between Carnegie Mellon University and the University of Pittsburgh. Established in 1986, PSC is supported by several federal agencies, the Commonwealth of Pennsylvania and private industry. PSC provides university, government, and industrial researchers with access to several of the most powerful systems for high-performance computing, communications, and data-handling available to scientists and engineers nationwide for unclassified research. PSC advances the state-of-the-art in high-performance computing, communications and informatics and offers a flexible environment for solving the largest and most challenging problems in computational science.

McGill Centre for the Convergence of Health and Economics | Carnegie Mellon University | PSC | University of Pittsburgh | DESAUTELS Faculty of Management | McGill

Chair:

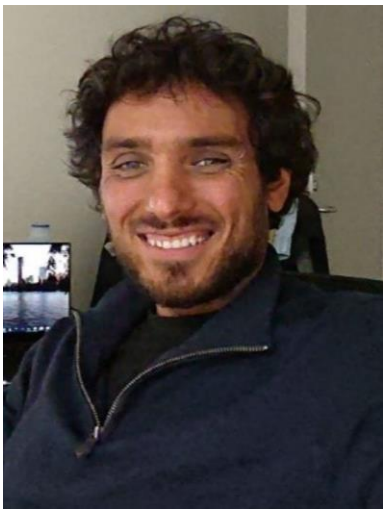
Laurette Dubé: Laurette Dubé, initially trained as a nutritionist, also holds degrees in finances (MBA), marketing (MPS), and behavioral decision-making/consumer psychology (PhD). Dr. Dubé is Full Professor at the Desautels Faculty of Management, McGill University. Her research focuses on the study of affects, behavioral economics, and neurobehavioral process-es underlying consumption, lifestyle, and health behavior. Her translational research examines how such knowledge can inspire effective interventions. She is also the founder and scientific director of the McGill Centre for the Convergence of Health and Economics, a unique initiative to push the boundaries of science to tackle societal and economic challenges and foster individual and collective health and wealth.

Sergiu Sanielevici: Sergiu Sanielevici, Ph.D. is Director of Support for Scientific Applications at the Pittsburgh Supercomputing Center, a joint project of Carnegie Mellon University and the University of Pittsburgh. He has served as the Deputy Director of the Extended Collaborative Support Service of the US NSF XSEDE project and as the manager of its Novel and Innovative Projects program, fostering nontraditional and interdisciplinary applications of advanced computing and data resources since 2011. He is currently the Principal Investigator of the Bridges-2 project and coPrincipal Investigator of the Neocortex project at PSC. Dr. Sanielevici is a proud alumnus of McGill University (Ph.D., Physics, 1986).

Panelists:

Dhabaleswar Panda: DK Panda is a Professor and University Distinguished Scholar of Computer Science and Engineering at the Ohio State University. He has published over 500 papers. The MVAPICH2 MPI libraries, designed and developed by his research group (http://mvapich.cse.ohio-state.edu), are currently being used by more than 3,200 organizations worldwide (in 89 countries). More than 1.56M downloads of this software have taken place from the project's site. This software is empowering many clusters in the TOP500 list. High-performance and scalable solutions for Deep Learning frameworks and Machine Learning applications are available from https://hidl.cse.ohio-state.edu. Prof. Panda is an IEEE Fellow and recipient of the 2022 IEEE Charles Babbage Award. More details about Prof. Panda are available at http://www.cse.ohio -state.edu/~panda.

Sean Hill: Dr. Sean Hill is a neuroscientist, professor, and CEO of SENSCIENCE, an AI company advancing open and FAIR data for research. He previously founded the Krembil Centre for Neuroinformatics (KCNI) at CAMH, applying data science, machine learning, and modeling to improve brain disorder treatment. He is a Professor at the University of Toronto and holds a Titular Professorship at EPFL, Switzerland. Dr. Hill earned his Ph.D. in computational neuroscience from the Université de Lausanne and completed postdocs at The Neurosciences Institute and the University of Wisconsin, Madison. He joined IBM and co-directed the Blue Brain Project, focusing on brain modeling and led the neuroinformatics strategy in the EU Human Brain Project. He also served as Executive Director of the International Neuroinformatics Coordinating Facility (INCF). His research focuses on brain circuitry and states like sleep and wakefulness. He leads the Canadian Youth Mental Health Insight (CYMHI) Platform, using open data and AI for youth mental health outcomes. Dr. Hill has authored over 100 publications and holds multiple patents.



Pierpaolo Sorrentino: Pierpaolo Sorrentino is a tenured associate Professor and principal investigator at the French National Institute of Health (Inserm) and at Sassari university. Pierpaolo is a neurologist and neuroscientist with expertise in neuroimaging and brain network dynamics, in particular in the context of neurodegenerative diseases. He holds a Ph.D. in engineering. His research focuses on understanding how diseases disrupt brain activities on a large scale, applying complex systems theory and dynamical models to guide the analysis of multimodal neuroimaging data. In particular, his most recent work has focused on the design of virtual patients to personalize both disease monitoring and therapeutic choices. Previously, Pierpaolo co-founded the Magnetoencephalography facility at the National Research Council of Italy, where he has been in charge of the creation of multiple large-scale cohorts for patients with neurodegenerative diseases. His interdisciplinary work is an attempt to blend methods derived from physics and applied mathematics, multimodal data, artificial intelligence, and clinical data into a coherent framework, aimed at the delivery of personalized treatments in clinical practice.

Lise Korsten: Prof. Lise Korsten is the Co-Director of the Department of Science and Innovation (DSI), National Research Foundation (NRF) Centre of Excellence in Food Security in Pretoria, South Africa, overseeing the food safety research program. Her research focuses on postharvest pathology, plant health, food safety, market access, and international trade. She has contributed to the development of a national food safety authority and a food control framework in South Africa. Prof. Korsten has supervised 87 MSc and PhD students, 60 BSc Hons students, and published over 200 peer-reviewed papers, chapters, and books, alongside 150 popular papers. With an H-index of 50, i10-index of 154, and over 10,387 citations, she ranks in the top 2% of scientists globally in her field. Her awards include the Water Research Commission's Human Capital Development Award and the MT Steyn Prize from the South African Academy of Science. She is a Fellow of the South African Society for Plant Pathology and the Academy of Science of South Africa. She also chairs the International Society for Plant Pathology Task Force on Global Food Security and is the newly elected President of the African Academy of Science, aiming to promote inclusivity, gender balance, and excellence across the continent.

Mark Daley: Mark Daley is the chief digital information officer at Western University, a full professor in the Department of Computer Science at Western University, with cross-appointments in multiple departments, The Rotman Institute of Philosophy, and The Western Institute for Neuroscience. He is also a faculty affiliate of Toronto's Vector Institute for Artificial Intelligence. Mark has previously served as the Vice-President, Research at the Canadian Institute for Advanced Research(CIFAR), and Associate Vice-President (Research) and Special Advisor to the President at Western.Mark was the institutional administrative lead for the successful Canada First Research Excellent Fund proposal “BrainSCAN: Healthy Canadian Brains for Life”. $66M. He has published over 100 research articles and has been awarded U.S, European, and Chinese patents for metabolomics profiling of concussion and has US patents pending for COVID-19 diagnostics.