



National University of Singapore and University of Toronto launch the Materials Data Foundry to fast-track discovery of functional materials with AI and robotics

NUS Institute for Functional Intelligent Materials (I-FIM) and U of T's Acceleration Consortium (AC), in collaboration with Nvidia and VeChain industrial partners, will build an open autonomous lab to create the world's most comprehensive experimental materials property dataset. This data will power AI models to accelerate advances in electronics, energy and infrastructure.

The Institute for Functional Intelligent Materials (I-FIM) at the National University of Singapore and the Acceleration Consortium (AC) at the University of Toronto have partnered to develop the Materials Data Foundry (MDF), responding to a critical bottleneck in materials innovation: the lack of large, high-quality, standardised datasets that is impeding the effective use of artificial intelligence (AI) to accelerate the discovery and deployment of high-value materials.

The MDF fixes the missing link by capturing recipes and results together, so AI can recommend not just *what* to make, but *how* to make it — reliably and at scale. This dataset will help shorten the path from idea to industrial use in various high-value sectors, from beyond-silicon electronics to clean energy and durable infrastructure.

Through the creation and development of an open autonomous lab in Singapore that integrates multi-modal AI with high-throughput robotics and in-situ measurements, the MDF will create the largest real-world dataset of synthesis routes and properties for solid-state materials.

The MDF has received a grant award of close to S\$10 million from Singapore's National Research Foundation (NRF), building on initial seed funding from U of T and NUS. The MDF is among the eight inaugural projects awarded under the S\$120 million AI-for-Science (AI4S) Initiative. The official launch of these projects was announced by Professor Tan Chorh Chuan, Permanent Secretary (National Research and Development), at the AI4X-Accelerate Conference 2026. The Initiative, which supports the National AI Strategy 2.0 (NAIS 2.0) in building up a vibrant and innovative AI research ecosystem, brings together teams of AI researchers and scientific domain experts from Singapore and top international institutions to develop AI methods and tools that can raise research productivity and accelerate discovery.

"Synthesis pathways of materials encode their composition, structure and morphology, thus determining their realistic properties," said Professor Sir Konstantin Novoselov, Director of I-FIM, who won the 2010 Nobel Prize in Physics for his world-changing experiments regarding the two-dimensional material graphene. "The MDF turns abstract predictions into practical recipes by feeding real synthesis data into AI models."

I-FIM and AC will join forces with Nvidia and VeChain industrial partners to integrate the computational and experimental materials data efficiently. The industrial partners will also help

develop reaction-landscape datasets and the AI- and blockchain-enabled algorithms to identify the optimum reaction pathways and conditions.

Why this matters

Modern technologies lean on complex solids, for example, two-dimensional semiconductors for energy-efficient computing, mixed-metal oxides for green hydrogen, and corrosion-resistant alloys for sustainable infrastructure. However, materials discovery is a major bottleneck.

Existing computer simulations assume ideal conditions, assuming thermodynamically stable states, while current AI tools often predict properties for simplified crystal structures while ignoring defects, entropy and, crucially, synthesizability. All this leads to thousands of promising, computer-suggested materials but far too few can actually be made to work in the real world.

The MDF fills a critical gap by linking synthesis recipes with their outcomes, enabling AI to recommend not only which materials to make, but also the best way to make them — reliably and at scale. In particular, the MDF addresses two core problems at once:

1. Data scarcity and bias: high-fidelity experimental data is limited and scattered.
2. Design without a recipe: most models do not encode synthesis pathways — the step-by-step conditions that create phases, defects and morphologies that determine performance.

“Our goal is to accelerate science for solid-state materials,” said Professor Alán Aspuru-Guzik, Director of the Acceleration Consortium. “Predicting the structures of materials is simply not enough now. We are closing the loop between AI suggestions, robotic synthesis and in-situ feedback — tracing the paths that lead to new working materials that make an impact in the real world.”

Materials Data Foundry: Bridging recipes and results for faster materials discovery

Above all, the MDF is a unified, multi-fidelity data engine. The Foundry will fuse heterogeneous inputs — high-throughput experiments, in-situ spectra, phase diagrams, microscopy and multi-scale simulations (such as DFT, MD, Monte Carlo) — into standardised, open datasets that explicitly record synthesis steps, intermediates and outcomes. This becomes training fuel for foundation models that predict both properties and viable routes to achieve them, including out-of-equilibrium phases and defect distributions.

The MDF is also an autonomous, open experimentation lab. It will deploy robotic synthesis tuned not merely to cover composition space, but to span alternative synthesis routes (gas/solid precursors, temperature–pressure histories, growth rates), while logging intermediates via mass spectrometry, gas chromatography, Raman and PL. Downstream testing modules will characterise electronic, optical and catalytic performance and feed results back to the models in real time, forming a backbone for MDF operation through digital lab integration and active-learning loops for AI-driven experimentation.

In addition, the MDF will be an algorithmic navigator of synthesis. New AI frameworks — combining Markov models of reaction steps, stochastic differential equation surrogates of macroscopic dynamics and Monte Carlo Tree Search — will explore the “synthesis landscape” under constraints such as energy use, cost or environmental impact. Learning from success

and failure in situ enables the frameworks to avoid dead ends and converge on robust, scalable recipes.

A key novelty lies in using machine learning interatomic potentials (MLIPs) near saddle points, coupled with advanced saddle-point search methods and uncertainty quantification. This will enable the creation of the first large-scale database of reaction landscapes and AI-assisted optimisation and control algorithms that operate directly on reaction-route graphs.

This integrated approach to reaction landscape modelling and control is genuinely novel and may lead to significantly shortened paths from lab discovery to deployable materials in various high-value sectors, from beyond-silicon electronics to clean energy and durable infrastructure. Furthermore, Kinetics- and transition-aware generative frameworks represent a significant step forward in our capabilities. Digital lab integration and active-learning loops: This highlights the practical application and dynamic nature of our research.

“Linking routes to results is the missing layer,” noted Associate Professor Kedar Hippalgaonkar, Principal Investigator at I-FIM. “The MDF will make standardised schemas and APIs openly available so other researchers can plug their data and tools into the Foundry — building a commons for practical materials discovery.”

Key priorities

In its first phase, the partnership focuses on three areas where better materials enable near-term gains:

1. Low-power and quantum-ready electronics: routes to beyond-silicon and topological materials for cooler, more efficient chips and spintronic devices.
2. Clean hydrogen production: mixed-metal oxide electrocatalysts engineered for >10,000 hours at 2 A/cm² and full-cell voltages around 1.7 V — translating to weeks of industrial operation without performance drop.
3. Long-life coatings: corrosion-resistant high-entropy alloy coatings for lightweight magnesium alloys that outlast commercial Zn–Ni on steel.

Across these testbeds, the MDF will deliver:

1. The world’s largest open materials dataset built from 50,000 high-fidelity experiments.
2. Multi-modal, multi-fidelity foundation models that generalise across compositions and conditions and propose candidates under cost and energy constraints.
3. Open software and APIs to query, visualise and integrate the MDF with external data and self-driving labs.
4. Best-practice protocols for data fusion and for building scalable automated materials labs.
5. Pathways to pilot manufacturing with industry partners.

“The MDF is designed for impact and not just publications,” said Professor Jason Hatrick-Simpers at the University of Toronto. “The Foundry’s outputs — data, models and validated recipes — are built to shorten the road from lab discovery to deployable materials.”

The MDF project benefits from strong existing collaborations with industry partners like Nvidia, SEA, and VeChain, facilitating access to top talent and joint PhD programs. Additionally, this initiative leverages I-FIM and AC’s extensive international collaborations with institutions such as Berkeley, Imperial College, RIKEN, University of Tokyo, and ETH-Zurich, to further strengthening research and discovery.

This new approach for creating a synthesis database and navigating reaction landscapes is highly scalable and applicable across a wide range of materials, including metal alloys, oxides, organic molecules, MOFs, and hybrid materials. By using advanced modelling methods to learn from smaller datasets and expand to much larger materials spaces, it provides a strong foundation for long-term impact and sustained global collaboration.