Dunia Innovations

Discovery Day 2026

Helmholtz-Zentrum Berlin | BESSY II · 19 May 2026

About this report

Discovery Day is where the frontier of AI research meets manufacturing and implementation. This annual invation-only summit unlocks the ecosystem to bridge the gap between discovery and deployment. The following report covers all six sessions of the 2026 programme, convened by Dunia Innovations and held at Helmholtz-Zentrum Berlin.

The event was held under the Chatham House Rule. Individual remarks are attributed by role and organisation type rather than by name. Where a company is named against a specific result, that result was presented publicly.

— About this report
— Executive summary
01 The programme
02 The room
03 Where the field stands
04 The barriers
05 The timescales
06 Where the room divided
07 Conclusions
08 Announcements
— Glossary
— Acknowledgements
— Endnotes

Dunia Innovations connects AI with automated laboratories to accelerate next-generation materials innovation.

Executive summary

AI models improved. Real-world validation lags behind.

One claim that went unchallenged all day: machine-learned interatomic potentials now work reliably for a wide range of material classes, replacing quantum simulation inside molecular dynamics at four to six orders of magnitude lower cost. In other words, these models work well when they encounter atomic structures similar to what they were trained on. Unfortunately, they become less reliable precisely where materials discovery requires exploration, in structures outside their training distribution.

Candidates are now designed faster than they can be made. The rate-limiting steps, regardless of improvements in AI models, are synthesis and characterisation.

The obvious fix, using the data industry already holds, was examined and rejected because of various shortcomings in such data. One laboratory has decades of microscopy stored without recoverable labels; nobody can now say what the images show. A senior figure at a major energy company said that, “Has this been done before?” remains unanswered inside their own organisation. The sharpest remark of the day was that “firms treat their datasets like a firstborn child, while quietly noticing that other people’s children are rather ugly.”

This report shows how the value and competitive grounds of inhouse datasets continue to be challenged and debated, leaving the ecosystem uncertain about how to price and defend any given organization’s proprietary data.

Some of it actively misleads. Round-robin testing across eleven laboratories shows a standard deviation roughly twelve times wider than a controlled procedure achieves on the same measurement (Figure 5). And published work shows rankings **inverting** between bench and plant: the bench winner is not merely weaker at scale, it can be the wrong molecule.

The analogy reached for was AlphaFold, and it is usually told backwards. The Protein Data Bank took twenty-four years to reach 100,000 structures and four decades to reach 240,000 (Figure 7): slow, curated, labelled, built by thousands of laboratories with no idea what it would enable. The largest open simulated dataset in the field now exceeds 110 million calculations[1]. It created the modern discipline, and it still cannot pick a material suitable for industry.

That's why three separate speakers reached the same conclusion, independently: **the binding constraint is experimental data, and autonomous laboratories are the mechanism to produce it.** The capture standard matters more here than throughput, a point made by a customer rather than a vendor.

Who builds that foundation and who owns it was argued on stage and left unresolved. This report doesn't try to settle it, but notes the one structural fact that may make the argument less decisive: data that does not yet exist has no incumbent owner.

One important caveat: six in ten companies are AI laggards (Figure 1), and almost none of them were in the room, since the event itself skewed toward companies that have already adopted AI.

MLIPs
10^4-$10^6\times$
cheaper than quantum simulation in dynamics

Data consostency
$12\times$
more variation in the same experiment run across multiple labs vs. in one controlled procedure

110M+
open simulated calculations from AlphaFold are still not enough

~40 yrs
to build the database AlphaFold stood on

>90%
of future training data has not yet been generated

What follows from this

01 **Machine-learned interatomic potentials work.** It was the only claim the room agreed on without qualification.

02 **Candidates are now designed faster than they can be made.** The bottleneck has moved into the physical world, and the gap will widen before it narrows.

03 **Industrial datasets are large, guarded, and largely not machine-learning ready.** Most have never been priced by the organisations holding them.

04 **Eleven laboratories testing one material disagree twelvefold, which highlights the reproducibility crisis.** A field in that condition cannot reliably tell whether it is improving.

05 **The database AlphaFold stood on took forty years to build.** Materials science has the generator and not the database.

06 **Who builds that database, and who owns it, was argued on stage and left unresolved.** The one structural fact that may make the argument less decisive: data that does not yet exist has no incumbent owner.

The programme

09:00 Doors open / Networking

09:30 **Materials discovery: yesterday to tomorrow**
Bernd Rech, Scientific Director, Helmholtz-Zentrum Berlin

10:00 **The Foundations of Materials Discovery Playbook**
Peter Seeberger, Founding Director, CTC

PANEL **Legacy & high-throughput experimentation, industrial R&D today, and what is different now**
Patrick Maestro, ex-Scientific Director, Solvay
Hans-Joachim Müller, Non-Executive Director, AkzoNobel, LANXESS, and OMV
Jill S. Becker, CEO, 02139 Inc.
Stephan Schunk, Executive Expert - Vice President, hte GmbH & BASF
Virginie Bellière-Baca, Senior VP, Global Head of Technology and Innovation, Sulzer
Sascha Vukojevic, VP Commercial, Dunia Innovations

11:00 Break & Mini Demo

11:30 **State of the Industry Today**

TALK **Quantified Industry Overview: The state of AI across chemicals, batteries, and semiconductors**
Gabriela Schaeffer, Partner and Associate Director, BCG

TALK **Industry Case Study: AI-driven R&D in practice**
Axel Brait, Senior Manager | Power the Core R&D, Chevron

TALK **From Pattern Recognition to Physics: Building the AI Stack for Materials Discovery**
Ryan Reese, Solutions Architect, NVIDIA

12:30 Lunch + IRIS Tours

14:00 **The Frontier of Materials AI**

TALK **Sector Overview: Where AI for Materials is heading**
Georgia Channing, ML for Science Engineer, Hugging Face

PANEL **Frontier models AI methods Scientific infrastructure**
Tian Xie, Founding Member of Technical Staff, Project Prometheus
Daniel Zügner, Senior Researcher, Microsoft
Kristin Schmidt, Senior Manager | Research & Strategy Lead, IBM
Alexander Hammer, Co-founder & CEO, Dunia Innovations

TALK **Guest contribution: The AI Scientific Revolution will be Driven by Open-Ended Algorithms (pre-recorded talk)**
Jeff Clune, Professor, Computer Science, University of British Columbia

15:15 Break

15:45 **From Autonomous Labs to AI-Native Datasets**

TALK **Sector Overview & Updates**
Benjamin Burger, Robotic Discovery Engineer, Dunia Innovations

TALK **Dunia World Model**
Ruard van Workum, Machine Learning Engineer, Dunia, Innovations

PANEL **Fireside Chat**
Marcus Tze-Kiat Ng, Co-founder & CTO, Dunia Innovations
Nico Fleck, Associate Director | Head of HTE & Catalysis, Merck electronics
Gonzalo Abellán, Co-founder & CTO, Matteco

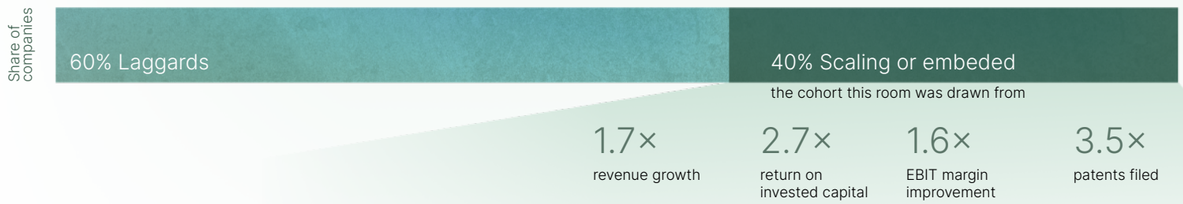
16:45 Closing

17:30 Reception + IRIS Tours

Figure 1

The adoption gap has stopped being about adoption

Four in ten companies are scaling AI or have embedded it. The gap between them and the rest now shows up in the financial statements rather than in the technology stack.



Source. Quantified industry overview, session three. Multiples are self-reported by surveyed firms and are associations rather than demonstrated causation. The four measure different things and are not comparable to one another.
BCG (2025). The Widening AI Value Gap. Boston Consulting Group, p. 4.

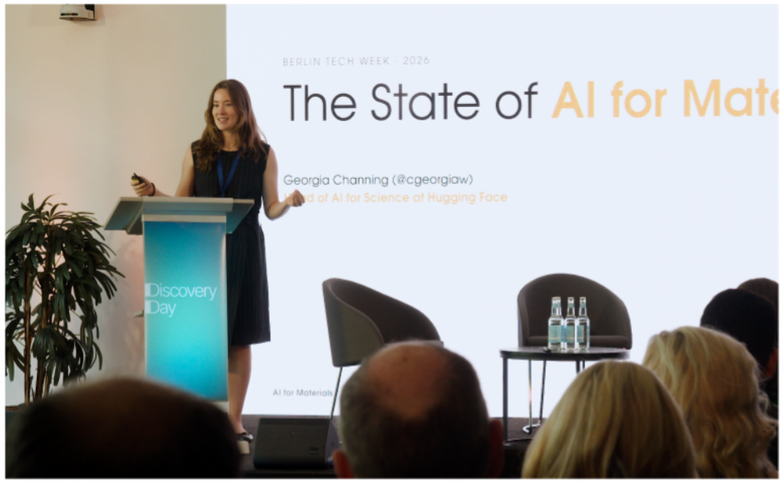
The room

Discovery Day is for the decision makers and executors, people who can actually act on what they hear. Its invitation-only format means a carefully cultivated audience from three worlds that rarely share a table: researchers with the instruments and a thirty-year view; industrial incumbents with decades of experimental data and the applications to prove it matters; and AI-native companies with a different way of thinking about innovation altogether. The disciplines in the room were just as mixed, so that chemistry, machine learning, robotics, data engineering and commercial strategy all had a seat at the table.



Selected Participants

NVIDIA · TOPSOE · BASF · Evonik (Leading Beyond Chemistry) · Henkel · ABB · LG · Hugging Face · IBM · TotalEnergies · Air Liquide · Merck · amazon · Project Prometheus · Saint-Gobain · HITACHI · SandboxAQ · Microsoft · Chevron · BCG · SULZER · covestro · hte · Center for the Transformation of Chemistry · SVENSQO · JX · CATL



60% Director & above

101 Attendees

67 Organizations

20 Speakers



Where the field stands

The models got cheap, so the bottleneck moved into the physical world

Figure 2

Where consensus ends

Each stage sits somewhere between settled and openly contested. The first was agreed by everyone in the room. The last was left unresolved on stage, by choice.



The models are cheap, so the bottleneck is physical. The archives cannot fix it, and some of them make it worse. The base has to be built, and nobody agreed who builds it.

Learned interatomic potentials work. That was said from the research side, the vendor side and the industrial side without qualification, and it is the only claim of the day that nobody contested. One model now covers every element at near-simulation accuracy. Generative models produce crystals targeted at named properties. Language models read the literature for recipes. Agents drive simulators, and closed-loop organic synthesis has been demonstrated. None of this existed three years ago.

These economics have already played out in production. Partner programmes produced more than a hundred candidate catalysts for oxygen evolution and more than two hundred immersion coolants, each within nine months, with roughly hundredfold acceleration in property prediction. Catalyst screening runs in production at BASF, Solvay and Dow. Semiconductor fabs use the same machinery for defect classification and recipe optimisation. In chemicals, AI-assisted carbon-capture solvent development has cut capital expenditure roughly tenfold and operating expenditure by five to ten times.

How you read the rest of this report depends on knowing where the field actually came from. It was built by a handful of public-good data releases: the Materials Project,[1] with over 170,000 openly licensed relaxed structures as of 2025 and the standard tooling built around them, then FAIR-Chem,[2] with more than 110 million calculations released alongside open models and a full stack. Every universal potential, every generative crystal model, every benchmark and most of the start-ups are downstream of those. Open models and datasets now see over a million downloads a day. **A small number of people gave away the thing everyone else is now arguing about keeping.**

> Consensus was available precisely where the work was already finished.

That consensus concerns one half of a five-step look: propose, screen, verify, make, measure. While the first three have collapsed in cost, the last two have not.

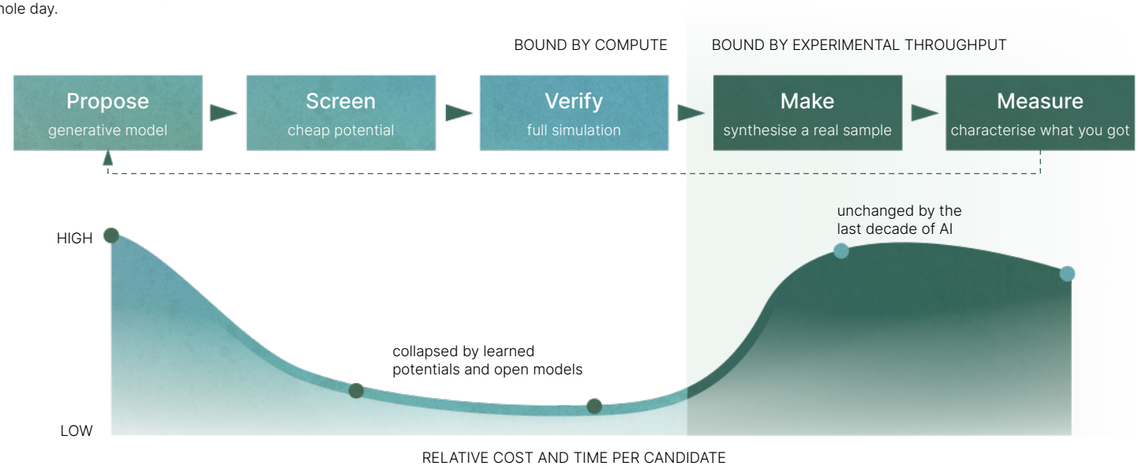
The gap was put at roughly threefold. The number measures throughput at campaign level, in validated candidates per unit time, comparing what a design pipeline generates and prioritises against what a conventional laboratory can synthesise and characterise. Raw compute speed against experimental speed would give a far larger ratio and a meaningless one, since most computationally proposed candidates are never worth making. On the measure that matters, the design half has stopped being the constraint. The physical half still is.

The consequence was not drawn on the day and belongs in the record. A research loop that only has to improve its own software gets faster with every cycle. A loop that has to pass through a furnace runs at the speed of the furnace instead. Those two loops don't just move at different speeds, the gap between them will keep widening before it narrows, and no amount of model quality can close it. The furnace sets the pace here.

Figure 3

The five-step loop, and where the cost now sits

Propose, screen and verify have collapsed. Make and measure have not. The shape of this curve is the argument of the whole day.



The barriers

If the bottleneck is physical experimentation, the obvious response is to use the data industry already holds. The room examined that and rejected it, for three reasons that compound.

The datasets cannot be queried

Everyone assumes the data exists and the difficulty is getting hold of it. On the contrary, Discovery Day's conversations revealed blind spots and obstacles that make using even existing data difficult or impossible.

One refiner, for example, cannot establish whether a given experiment has been run before, because decades of results collected under changing methods across nine orders of magnitude of reactor volume cannot be reconciled well enough to search. A national laboratory, in the meantime, holds decades of microscopy without labels. The labels are unrecoverable by machine and unreconstructable by an expert. Even language models digitising historic records have their limits. While these models recover text well, they cannot recover the metadata that was never written down in the first place.

Nevertheless, a panel of incumbents observed that one's own data always looks more valuable than anyone else's.

Nobody can be at fault for laboratories in 1995 not recording metadata for a nonexistent model. Recoverability, on the other hand, remains an operative question. For the unlabelled cases, better tooling unlocks nothing because the information was never there. This points to the next step being an inventory more than a disclosure policy.

> An organisation that cannot query its archive has never tested what the archive is worth.

Three questions R&D leaders should ask their team

01 **How long does it take to find out whether a given experiment has already been run here?**
Where the answer depends on a long-serving colleague remembering, the knowledge lives in people rather than in the company archive, and it leaves when they do.

02 **What fraction of our spectra and performance tests carry the process conditions that produced the sample?**
Supplier, batch, operator, temperature curve, atmosphere, timing, recorded as the run actually went rather than as the protocol specified.

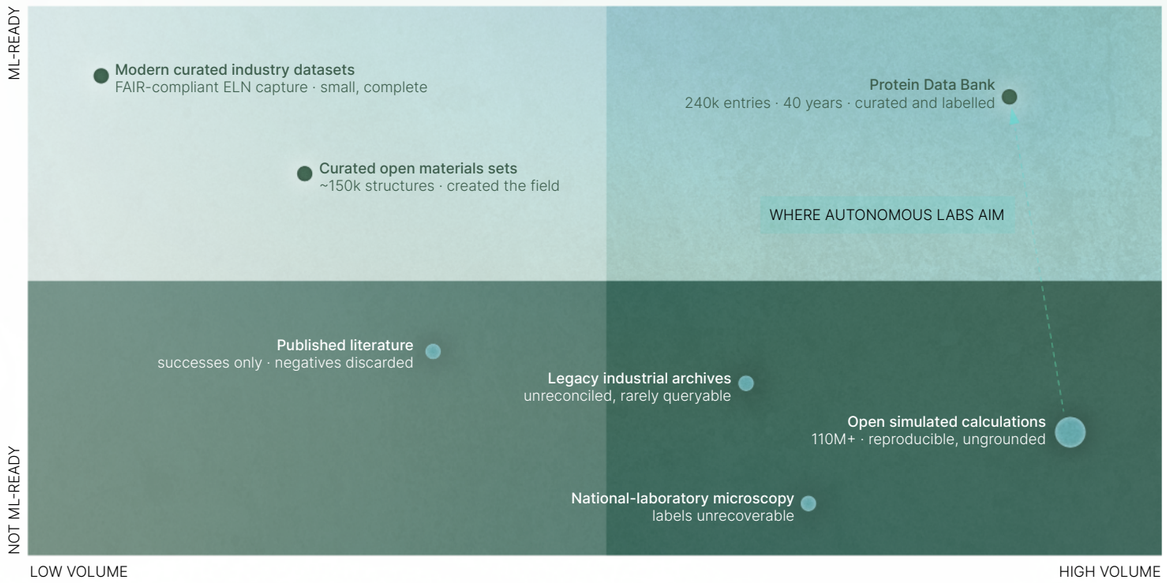
03 **What happens to a failed run?**
Where it goes unrecorded, or lands somewhere nobody can query, the most valuable training data the organisation generates is being discarded daily.

Three negative answers do not mean the data is worthless. They mean its value is unknown, which is more urgent, because an unknown value cannot be defended, priced, traded or improved.

Figure 4

Volume is not the axis that matters

Plotted against readiness as well as size, most industrial and simulated data sits in the two quadrants that cannot train a trustworthy model.



Vertical axis combines three properties that fail independently: grounding in real conditions, completeness of metadata, and reproducibility of the underlying measurement. The simulated set scores high on the third and low on the first, which is why volume alone does not place it.

The measurements cannot be trusted

A round-robin across eleven laboratories worldwide, testing the same material to the same protocol, reports a standard deviation roughly twelve times wider than a controlled procedure achieves. [3] This holds true across findings from multiple published sources that never referred to one another. Teams in chemicals, batteries, and semiconductors who consider themselves unrelated fail at the same three things, namely, solid handling, liquid handling and atmosphere control.

Figure 5

Eleven laboratories, one material, twelvefold disagreement

Both measurements use the same material and standardized test protocol, making the comparison meaningful. What changes is the laboratory environment and how the procedure is implemented.



Source. Round-robin study after Appelhaus et al.[3]. Controlled comparison as presented on the day. Closing line is this report's inference.

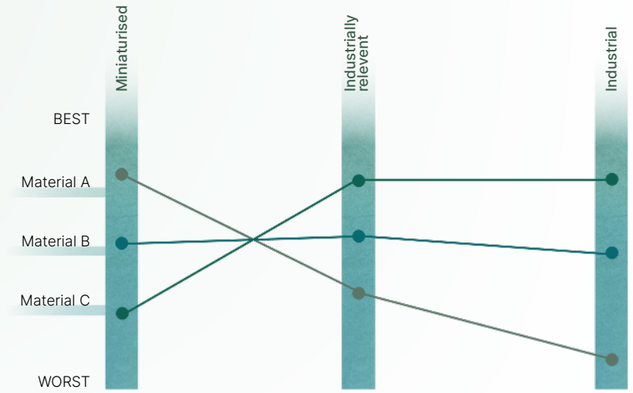




Figure 6

The best material on the bench can lose at industrial scale.

If the ordering inverts, screening more candidates on the bench sharpens confidence in the wrong candidate rather than improving the shortlist.



Illustrates an effect reported in Browne et al.[4] and Xu et al.[5]. Not plotted from a dataset.

Instead of degrading gracefully with scale, the rankings reorder, resulting in qualitative failure. Miniaturised setups omit mass and heat transport, morphology, durability and system effects, and the literature records bench winners failing at industrial conditions where candidates discarded early would have won.[4][5]



Teams report against their own metrics, and one experienced operator put the question directly: how do you benchmark, and against what? No clear answer was available in the room, because no common benchmark exists in the field.

Every acceleration claim in circulation, including the credible ones, rests on an unvalidated ruler. Collecting more data on the same footing does not resolve it: the additional data carries the same limitations, and can sharpen confidence in a candidate whose ranking may not hold under industrial conditions. This is the gap the benchmark initiative in the ‘Announcements’ section is intended to address.

> A field that disagrees twelvefold with itself, and whose rankings reorder between bench and plant, cannot reliably tell whether it is improving.

The timescales

Forty years behind, seven years ahead

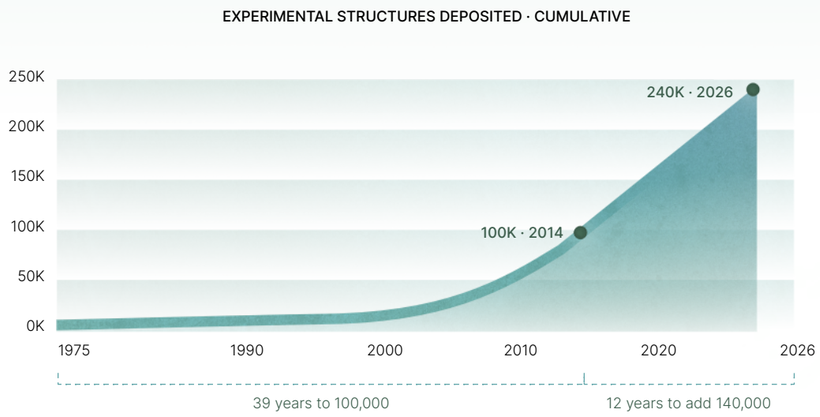
Two timelines sit on either side of discovery. Before it, comes the time needed to build a trustworthy data foundation. After it, comes the time it takes to qualify a material for industrial use. For faster discovery, the answer turns out to be moving work earlier: better data must be captured sooner, and more industrial constraints need to be brought into the design process.

The AlphaFold analogy is often used here, but it can hide an important part of the story. The model did not appear in isolation, it was built on decades of accumulated data and scientific work. The curve puts the model at the end of a much longer story.(Figure 7)

Figure 7

Forty years of measurement, then three years of generation

Left: the experimental archive, on a linear scale, accelerating. Right: what became possible once it existed.



ONCE THE BASE EXISTED

240,000
experimental structures, built over 40 years

Predicted structures released, 2021 to 2024

214M

roughly 900 predicted structures for every experimental one, generated in three years rather than forty

The second number depends on the first.

Materials science has the generator and not the database.

Sources. Cumulative archive holdings from the milestone series; ~170,000 at the 2020 model training cutoff; >240,000 as of January 2026. Predicted database ~365,000 at launch in 2021 to >214 million by 2024.[6] [7]

Alphafold involved four decades of patient, standardised, labelled measurement by thousands of laboratories, none of whom knew what it would be used for. Only afterwards did the generative expansion happen, which was trustworthy precisely because of the strength of the base underneath it. [6][7]

In materials, the AI frameworks are well established, but we do not have that kind of data to train them on. Without it, even 110 million open calculations cannot pick an industrially scalable solution. If more than ninety per cent of future training data in materials does not yet exist, there's a strong argument to be made for funding the slow, institutional work of building that base for materials, unglamorous as it may be.

That does not mean building the dataset has to take another forty years. Autonomous labs can now run thousands of experiments per day, generating a similarly sized dataset much faster. That said, experiments automated without standardising how their data are captured simply produce unusable data faster. The most important contribution of autonomous labs may be its metadata recording capability, a system that ensures every experiment in this high throughput environment is annotated according to thorough, consistent standards.

Unfortunately, even spectacularly faster discovery doesn't necessarily make commercialisation happen at unprecedented speeds. There's still a whole qualification process required to prove that it actually works, is safe, is insurable, can meet commercial agreements, etc. For a refinery catalyst or an electrolyser, that process often takes another three to seven years after the scientific discovery is essentially finished.

The process can be shorter where urgency is high, as vaccines and energy security have both demonstrated at scale. But what sets its length in ordinary conditions is the amount of unknowns that arrive with the candidate. If industrially relevant testing conditions are introduced early in the pipeline, those unknowns diminish.

Here, AI can help incorporate real-world constraints into the design and selection phase where changing a decision is much cheaper. It can reject problematic candidates early and send more mature candidates into qualification, candidates that require less time, money and troubleshooting. ●

> Every constraint found at the design stage is one that does not have to be found three years into a qualification programme.

Where the room was divided

Three things the day could not settle

Should the data foundation required by materials AI be proprietary or shared?

Proprietary data is too valuable to share

The industrial position

Data is expensive to generate, competitively sensitive, and one of the last strongholds of industrial advantage. It takes substantial work to create genuinely comparable structured data at scale, making proprietary datasets even more valuable.

The datasets will not be open-sourced, and there is no interest in doing so. Protecting data is exactly what doing business means.

Shared proprietary data is foundational to better AI

The technology position

Pooled data accelerates discovery for everyone and builds the infrastructure that better models will depend on. Releasing a dataset benefits the community without disadvantaging the company that releases it.

A handful of public releases created this entire filed, and everything built since has grown from them. That history alone shows the benefit in practice.

The debate around existing legacy datasets reached a standstill, but there's still 90% of future training data that has not yet been generated and has no incumbent owner. There's an opportunity here to create a new system in which new experimental data is generated according to an agreed common standard from the beginning. If enough high-quality new data is generated under common standards, access to the existing proprietary data may become far less consequential.

Figure 8

Three paths forward

The options discussed differ less in technology than in who absorbs the risk and who keeps the upside.

Each firm alone Build internal capability, keep everything.	**Open source community** Shared datasets, released and maintained openly.	**Each firm alone** Shared facilities generating new data to one standard.
RISK: FULL	RISK: SHARED	RISK: POOLED
UPSIDE: FULL	UPSIDE: DILUTED	UPSIDE: NEGOTIATED
SPEED: SLOW	SPEED: FAST	SPEED: FAST
Twenty years of attempts with limited success. Least defended in the room.	Created this field once already. Industrial participants said plainly they will not.	New data carries no legacy owner. Governance is unsolved.

The argument on stage ran between the first two columns.
The third came up as a way through, moving the question from who should share to who should generate.

Transferable data across applications

Beyond data ownership and creation debates, the question of knowledge transfer, though far less discussed through the day, may be at least equally significant. After all, this question determines what kind of dataset should be built in the first place. An AI trained on data for one kind of material or problem is not automatically useful for other material problems. Nobody currently knows how far knowledge transfer goes in practical industrial applications.

With AlphaFold, structural biology benefited from researchers repeatedly measuring and problem solving on one unit operation. That said, the industry is moving toward foundation models across catalysis, batteries, semiconductors and photovoltaics, betting on breadth. In the meantime the best available model performed worse when trained across multiple materials problems at once than one trained specifically for a given task.

The debate about the best way to create transformative AI says, from the broad foundation model perspective, to collect an enormously diverse dataset spanning lots of materials and applications, assuming that learning from one area will improve performance elsewhere. On the other side, the deep specialist model would amass enormous quantities of extremely consistent data about one well-defined task.

Currently, models can learn general things about how atoms interact that apply across the periodic table. Some transfer plainly happens, as they don't necessarily need an entirely separate conception of physics for every single material. The open question sits further up, whether a universal simulation capability carries through to applied industrial cases. We don't know, for example, whether the model will be able to accurately predict which catalyst will work in a refinery, which battery material will survive thousands of cycles, or which photovoltaic material will perform well in a manufactured device.

The elephant in the room

Rules built before AI-driven discovery existed and which shape how experiments get designed, how results get published, and how fast teams are expected to move, are now the most persistent obstacles in the room. Research institutions, incumbent panels, and operators alike repeatedly expressed frustrations related to these rules.

Academic publishing rewards clean, successful results and discards failed experiments, even though failures are exactly the data models need to learn from. Machine-learning teams iterate in days, while lab protocols are built around years of consistency. Publishing a paper and shipping a material are goals built on completely different systems that respectively pull people in different directions. More than one person who had tried to digitalise a pipeline inside their own company said the hard part was getting colleagues to change how they worked far more than writing the software itself.

These issues came up incidentally in almost every session, without being the explicit focus of any one session and with no candidate solutions amidst the rest of the day's discussions. This space for institutional change offers the clearest opening for next year's Discovery Day, ideally led by people who have already succeeded in changing scientific workflows, incentives, data practices, or organisational structures inside big companies and institutions.

Conclusions

01

Progress shifted the bottleneck

The computational constraint is gone, now that simulation and modeling have become fast and cheap (Figure 3). What's left runs into physical limits instead: furnace time, trained technicians, and calibration standards that can be trusted. **Cheaper simulation means the real work, the physical bottleneck, can finally begin.** Money spent on better models buys diminishing returns against a fixed bottleneck, while spending it on the bottleneck itself is what actually speeds up the field.

02

Companies defend data they haven't valued

The archives are large and tightly guarded, but most companies don't actually know how valuable their own data is. One panel admitted that companies tend to overvalue their own data. Another described a company that couldn't even tell whether it had already run a given experiment before. Nobody has measured what their archive is actually worth. **Companies are debating whether to share industrial data before anyone has figured out what that data is actually worth.**

03

Evaluation, not volume, is the first unsolved problem

Testing one material across eleven independent laboratories produced results twelve times more inconsistent than a single lab repeating its own test (Figure 5). Rankings get worse than noisy at industrial scale, they flip: a material that wins in small lab tests can lose under real production conditions (Figure 6). And there's still no agreed way to measure whether a result is actually good. **The industry decides which materials are worth pursuing with a ranking method that different labs can't even agree on.** Collecting more data under the same conditions builds confidence in results that might still be wrong.

04

Generating data at scale only works once the foundation is solid

The Protein Data Bank took forty years of careful, labelled data collection before AlphaFold could use it to generate five hundred times more structures in just three years (figure 7). Materials science hasn't built that foundation yet, and most of the data future models will need doesn't exist. **Rushing to generate data before that foundation exists would just recreate today's messy archives, faster.** A realistic next step: build one small, carefully labelled dataset in a single application area, and use simulation to help fill it in.

05

The room was not a sample of the industry

Roughly six in ten companies in this industry are still behind on AI adoption, and almost none of them were in the room. **Every finding in this report comes from the roughly 40% of companies that have already adopted AI, not from the 60% still behind.** The real open question is what happens to the majority still behind, and what would actually get them to change.

What stayed open

Four questions the day raised and did not answer.

What would make an industrial company willing to share, and how large would the benefit actually be?

Before anyone can answer that, they'd need to know what their data is actually worth, and most companies can't, since they've never been able to query their own archive. What they'd stand to gain by sharing it is just as unmeasured.

Whose job is it to make sure today's experiments don't become tomorrow's unreadable archive?

Right now, nobody's. The comparison people kept reaching for was the instruments themselves, standardized, maintained, trusted. Metadata has none of that, and no one has claimed responsibility for building it.

Does a universal simulation capability carry through to applied prediction across scales?

Some transfer happens at the level of shared representation. Whether it reaches applied predictive performance, and by how much, is unsettled, and the one comparison presented favored a model trained on a single material task over one trained broadly across many.

What happens to the AI laggards who were not in the room today?

This stayed unaddressed, even though the gap between leaders and laggards is already visible in revenue, returns and patent output, and it is widening rather than closing.



Announcements

01 · Open initiative

An end-to-end benchmark for industrial materials innovation

Existing benchmarks measure model accuracy against curated datasets. They do not measure whether a proposed candidate can be made, or whether it holds its ranking under industrial conditions, which is where most of the value and most of the error live. As mentioned in the "Barriers" section: a field that cannot measure reproducibly cannot tell whether it is improving.

Because of this, a small group has formed to build such a benchmark. It is deliberately composed of parties with opposing interests on the basis that a benchmark authored by any single constituency encodes that constituency's assumptions. The work is collaborative and the group intends to publish at a leading machine learning conference next year.

Contributing

- Open to further models, datasets and evaluation protocols
- Particularly sought: experimental datasets with full process conditions, including failed runs
- Target: publication at a leading machine learning conference next year

To join the working group or contribute a dataset, get in touch.

Alexander Hammer
Co-founder & CEO
alex@dunia.ai

02 · Open to founding members

The Gigalab consortium

A facility designed to generate a trillion training tokens of materials science data, under industrially relevant testing conditions, for training materials foundation models. Modular by construction, solve-once by design. The consortium exists so the workflow and metadata standard is set by the organisations that will have to live with it. Joining early buys a governance position in establishing the standard.

Founding members

What founding members shape

- The workflow and the metadata standard the data is captured to
- Which application domains the facility runs first
- Proposed best practice on data and IP, through a working group formed at this event and reporting at the next Discovery Day

The facility

- A trillion training tokens of materials data, generated under industrially relevant conditions
- Modular by construction, solve-once workflows
- Capture standard published for adoption beyond the consortium

To discuss founding membership, get in touch.

Alexander Hammer
Co-founder & CEO
alex@dunia.ai

Glossary

Terms used across the day.
In the form they were introduced.

DFT (Density Functional Theory)
Quantum-mechanical simulation. The standard ground truth for materials property data. Accurate, slow and computationally expensive.

GNN and equivariance
Graph neural networks that respect the rotation and translation symmetries of physics. The dominant architecture for atomistic machine learning.

Sim2Real
The gap between what a simulator predicts and what a real synthesised sample does. Where most pipelines break.

Closed-loop / self-driving lab
Model proposes, robot synthesises, instrument measures, result feeds back. End-to-end automation.

MLIP (Machine-Learning Interatomic Potential)
A model that predicts forces and energies, replacing DFT inside molecular dynamics at 10^4 to 10^6 times lower cost.

Foundation model (for science)
One large pretrained model, broadly applicable across tasks. In materials: universal potentials and multimodal chemistry models.

Inverse design
"Give me a material with property X." Conditional generation in place of brute-force screening.

Synthesizability
Whether a proposed material can realistically be produced. Distinct from thermodynamic stability, and substantially harder to predict.

Acknowledgements

With thanks

Discovery Day 2026 was hosted at Helmholtz-Zentrum Berlin, and we are grateful to the centre and to the BESSY II team for opening the facility to the event. Our thanks to every speaker, panellist and moderator who prepared material for the day, and to the participants who argued with them.

Under the Chatham House Rule, contributors are not attached to the findings in this report, and no participant should be taken to endorse its conclusions. The organisations represented are listed in "The room" section; attendance implies neither endorsement of this report nor participation in anything announced in the announcements.

Endnotes

01 Jain, A., Ong, S. P., Hautier, G., et al. 'Commentary: The Materials Project: A materials genome approach to accelerating materials innovation.' APL Materials, vol. 1, no. 1, 2013, 011002. https://doi.org/10.1063/1.4812323

Horton, M. K., Huck, P., Yang, R. X., et al. 'Accelerated data-driven materials science with the Materials Project.' Nature Materials, 24(10), 2025, pp. 1522–1532. https://doi.org/10.1038/s41563-025-02272-0

02 Barroso-Luque, L., Shuaibi, M., Fu, X., Wood, B. M., Dzamba, M., Gao, M., Rizvi, A., Zitnick, C. L., & Ulissi, Z. W. 'The Open Materials 2024 (OMat24) inorganic materials dataset and models.' Nature Computational Science, 2026. https://doi.org/10.1038/s43588-026-00996-w (preprint: arXiv:2410.12771, 2024).

03 Appelhaus, S., Ritz, L., Pape, S.-V., et al. "Benchmarking performance: A round-robin testing for liquid alkaline electrolysis." International Journal of Hydrogen Energy, vol. 95, 2024, pp. 1004–1010. https://doi.org/10.1016/j.ijhydene.2024.11.288

04 Browne, M. P., Dodwell, J., Novotny, F., Jaśkaniec, S., Shearing, P. R., Nicolosi, V., Brett, D. J. L., & Pumera, M. (2021). "Oxygen evolution catalysts under proton exchange membrane conditions in a conventional three electrode cell vs. electrolyser device: a comparison study and a 3D-printed electrolyser for academic labs." Journal of Materials Chemistry A, 9(14), 9113–9123. https://doi.org/10.1039/d1ta00633a

05 Xu, D., Stevens, M. B., Cosby, M. R., et al. 'Earth-Abundant Oxygen Electrocatalysts for Alkaline Anion-Exchange-Membrane Water Electrolysis: Effects of Catalyst Conductivity and Comparison with Performance in Three-Electrode Cells.' ACS Catalysis, vol. 9, no. 1, 2019, pp. 7–15. https://doi.org/10.1021/acscatal.8b04001

06 Vallat, B., Rose, Y., Piehl, D. W., et al. 'RCSB Protein Data Bank: Delivering integrative structures alongside experimental structures and computed structure models.' Nucleic Acids Research, 54(D1), 2026, pp. D489–D498. https://doi.org/10.1093/nar/gkaf1187

RCSB Protein Data Bank. "PDB Statistics: Overall Growth of Released Structures Per Year." RCSB.org. https://www.rcsb.org/statistics/growth/overall-structures (accessed January, 2026)

07 Varadi, M., Bertoni, D., Magana, P., et al. 'AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences.' Nucleic Acids Research, 52(D1), 2024, pp. D368–D375. https://doi.org/10.1093/nar/gkad1011

Discovery Day 2026
Report

Dunia Innovations GmbH
Magnusstraße 11
Haus 4, Stock 3
12489 Berlin Germany

dunia.ai

Dunia Innovations