Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

# What should AI agents know? Explicit knowledge for trustworthy, sustainable materials research

Simon Rihm[1], Markus Kraft[1,2,3,4]

released: November 10, 2026

[1] CMPG
GRIPS – Gründerinnenzentrum Pirmasens
Delaware Avenue 1–3
66953 Pirmasens
Germany

[2] Department of Chemical Engineering
and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[3] CARES
Cambridge Centre for Advanced
Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[4] CMCL
No. 9, Journey Campus
Castle Park
Cambridge
CB3 0AX
United Kingdom

Preprint No. 350



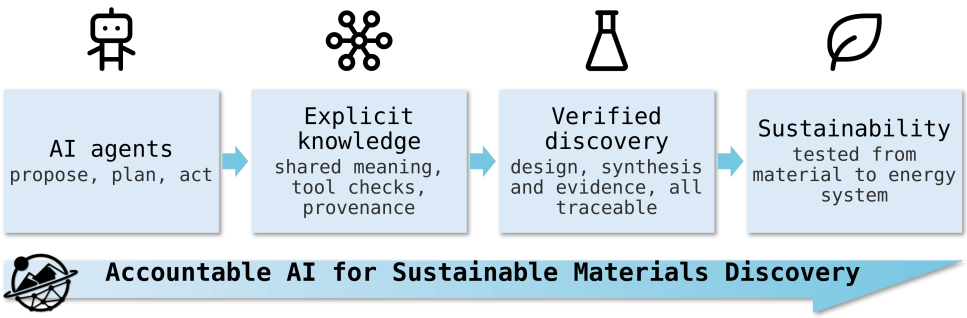

*Keywords:* The World Avatar; knowledge graphs; AI agents; self-driving laboratories; sustainable materials discovery

**Edited by**
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

**E-Mail:** mk306@cam.ac.uk
**World Wide Web:** https://como.ceb.cam.ac.uk/

**Abstract**

Artificial intelligence can propose materials, interpret measurements and increasingly coordinate experiments. Yet reliable scientific autonomy requires more than capable models: research systems must preserve what is known, where evidence came from and which actions are justified. We argue that machine-actionable scientific knowledge is essential for trustworthy autonomous research, and that ontology-grounded knowledge graphs are a particularly powerful way of achieving it when scientific meaning must be shared across disciplines, systems and scales. Using The World Avatar (TWA) as a case study, we examine how ontologies, dynamic knowledge graphs and software agents connect models, literature, experiments and laboratory assets, and how ontology-derived tools can constrain large language model (LLM) actions. A reported design-to-synthesis study of a new metal–organic polyhedron demonstrates the integration of previously separate tasks, but not yet an environmental benefit or fully autonomous discovery. A material-passport example further shows how shared semantics can support industrial circularity assessment. We distinguish structural validation from scientific verification, and circularity indicators from demonstrated environmental benefits. The priority is not to create a universal platform but to build evidence-preserving research infrastructure whose scientific and sustainability benefits can be tested.




## Highlights

- Scientific AI needs a shared, traceable record of what is known, how it was derived and which actions are justified.
- Ontology-grounded knowledge graphs share scientific meaning across disciplines, systems and scales, with automated provenance for registered computations.
- Ontology-derived tool contracts constrain LLM actions, but formal validity does not guarantee scientific correctness.
- A new porous cage material was designed, screened, synthesised and characterised within one connected knowledge graph.
- Faster AI discovery is not automatically sustainable: benefits must be tested from synthesis to full energy systems.

# Contents


1 Introduction 3

2 Explicit knowledge as research infrastructure 3

3 From descriptive ontologies to accountable agents 5

4 A materials discovery cycle: what has been demonstrated? 5

5 From material discovery to measurable sustainability 6

6 Challenges and a testable research agenda 8

7 Conclusion 9

References 10

# 1 Introduction

Batteries, electrolysers, catalysts and carbon-capture sorbents are central to energy transitions. Their performance depends critically on the materials they are made of, but a material's value cannot be established from a promising molecular property alone. Synthesisability, durability, process integration and environmental impacts ultimately determine whether a material is useful. Meanwhile, high-throughput computation, machine learning and self-driving laboratories are accelerating design–make–test–analyse cycles [1, 23, 38]. LLM-based agents add the ability to interpret literature, propose procedures and invoke scientific tools [7, 8].

The difficulty is no longer simply generating predictions. Scientific information is dispersed across papers, instrument logs, calculation outputs and local databases, often without comparable units, conditions or derivation histories. An agent may produce a plausible synthesis recommendation while confusing a calculated property with a measurement, overlooking a missing condition or forgetting a previous failure. Studies of LLM reasoning in closed-loop chemistry underline the continuing importance of domain expertise and verification [11, 13], and a recent perspective in this journal calls for investing as much in verification and accountability as in models [16]. The problem is most evident in sparse-data domains, where publication bias means the literature records mainly successes. Explicit representations can retain failed conditions and unviable candidates alongside successes, so that this evidence remains available to later reasoning and can help agents avoid repeating unproductive experiments [27, 32].

Alongside asking whether an AI system can suggest the next experiment, it is therefore worth asking *what evidence it draws on, what it is able to change, and how its results can be checked*.

We argue that machine-actionable scientific knowledge is essential for trustworthy autonomous research, and that ontology-grounded knowledge graphs are a particularly powerful way of achieving it when scientific meaning must be shared across disciplines, systems and scales. The World Avatar (TWA) offers a practical example through which to explore the opportunities and open questions [2].

# 2 Explicit knowledge as research infrastructure

For scientific reuse, a data value needs more than a label [41]. A measured adsorption capacity is meaningful only alongside the material identity, pressure, temperature, activation history, units, uncertainty and source, and linking it to a process model requires a shared account of what each quantity means. Explicit representations describe entities, relationships, assumptions and derivations in forms that people and software can inspect and query. They make provenance *recordable*; they do not guarantee that it is complete or correct.

TWA implements this approach through a dynamic knowledge graph (KG), domain ontologies and semantic software agents connected to computational and physical systems. For example, OntoSpecies describes chemical species [26]; OntoMOPs represents metal–

organic polyhedra and their building units [20]; and OntoSyn structures synthesis procedures [30]. The `twa` Python package and natural-language query interfaces offer access without direct graph programming [6, 31].

Crucially, TWA represents *agents themselves* semantically (Fig. 1). The OntoAgent ontology provides a schema for describing services: their capabilities, typed inputs and outputs, and how to invoke them. Each individual agent is then recorded as an entry in the KG. The agent-composition framework uses these descriptions to discover compatible agents and assemble executable workflows by matching input and output types, as demonstrated for cross-domain emissions modelling [42]. Trust in agents was also addressed early. A smart-contract-based marketplace scored agent outputs automatically and kept tamper-resistant records, so that composition could choose between functionally identical agents [43].

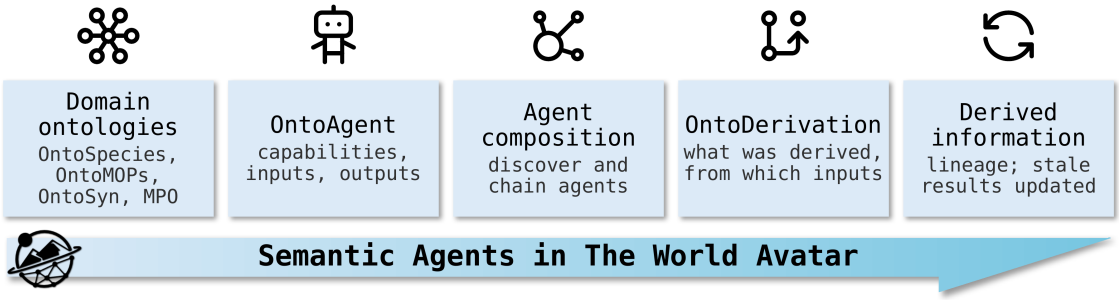



**Figure 1:** *Semantic agents and explicit knowledge representation in The World Avatar (TWA). Domain ontologies provide shared scientific meaning, while the OntoAgent ontology describes the capabilities, inputs and outputs of semantic software agents, enabling their discovery and automated composition. The OntoDerivation ontology describes how individual computations and their dependencies are recorded. TWA's Derived Information Framework records computational lineage and supports the automatic updating of outdated derived information. Together, these components connect scientific knowledge, computational models and physical systems through a dynamic knowledge graph.*

The complementary OntoDerivation ontology defines how *particular computations* are represented. TWA's Derived Information Framework links each derivation to its agent and input and output individuals, creating queryable dependency chains and enabling stale results to be recomputed upon access [4]. An automated laboratory application used this to manage the iterative workflow between two synthesis robots located in Singapore and Cambridge [5]. OntoAgent therefore describes *what an agent can do*, while OntoDerivation records *what it derived and from which inputs*. This provides automated computational lineage for registered workflows, though not completeness of external provenance or scientific correctness.

These capabilities matter most when independent tools or disciplines must reuse the same scientific meaning across changing schemas, organisations or scales; for a stable, local workflow, a conventional database with a well-designed interface may be simpler. Shared identifiers alone do not resolve contradictory definitions, uncertain measurements or differing modelling assumptions. Semantic interoperability is an ongoing agreement be-

tween communities, not a property conferred automatically by a graph database. TWA is one contribution among several: data infrastructures such as NOMAD and OPTIMADE [15, 33], materials and battery ontologies [12, 14] and workflow provenance in AiiDA [18] pursue complementary goals.

# 3 From descriptive ontologies to accountable agents

LLM-enabled agents extend TWA's established semantic-agent architecture by introducing flexible interpretation and planning, while relying on typed interfaces and explicit constraints for accountable execution (Fig. 1). The question is how far formal descriptions can constrain these actions without confusing scientific correctness with structural validity.

Literature information extraction shows both the promise and the associated challenges [39, 40]. Ontology-aligned LLM workflows have converted published MOP synthesis information into structured knowledge [30], but a plausible extracted procedure does not guarantee that the source was read correctly. In a recent study, ontology-to-tools compilation turns selected domain descriptions and constraints into tools that an LLM agent can call through the Model Context Protocol (MCP) [44]. Required fields, recognised entities and units are then checked whenever the agent tries to create a record, and structured feedback lets it correct errors. On a 30-paper synthesis-extraction benchmark, the approach achieved a micro-averaged F1 of 0.83; removing constraint feedback reduced synthesis-step F1 from 0.84 to 0.57.

This shows that constraint feedback improves the tested workflow, *not* that the resulting knowledge is necessarily correct. A well-formed value may come from a misread paper, and a plausible reaction may still fail in the laboratory. Independent source checks, uncertainty estimates, instrument interlocks and, where appropriate, human approval remain necessary [3, 21].

No single representation technology is required, however. Persistent optimisation campaigns can separate an LLM's interpretation from the state and execution of experiments [25], and relational databases or typed APIs may suffice for narrower tasks. The question is which combination preserves scientific meaning, evidence and control *at acceptable cost*, rather than whether all research can be placed in one graph.

# 4 A materials discovery cycle: what has been demonstrated?

Metal–organic polyhedra (MOPs) are discrete cage-like assemblies of metal-containing and organic building units. Their regular geometries suit structure-guided design, and their porosity motivates research into separations and other applications [20, 22]. In TWA, geometric and chemical building units are linked so that candidate assemblies can be explored systematically and passed to computational screening [9, 10].

A recent study reports a linked design-to-synthesis workflow [32] (Fig. 2). Candidate MOPs were generated from structural rules and screened by tight-binding density functional theory and crystal-structure prediction. Synthesis conditions were inferred by analogy with previously extracted procedures. A previously unreported Zr-EDB-MOP was synthesised accordingly; infrared spectroscopy, high-resolution mass spectrometry and powder X-ray diffraction provided evidence supporting its assignment. An experimentally unsuccessful candidate, identified as strained during screening, was also retained in the knowledge record.

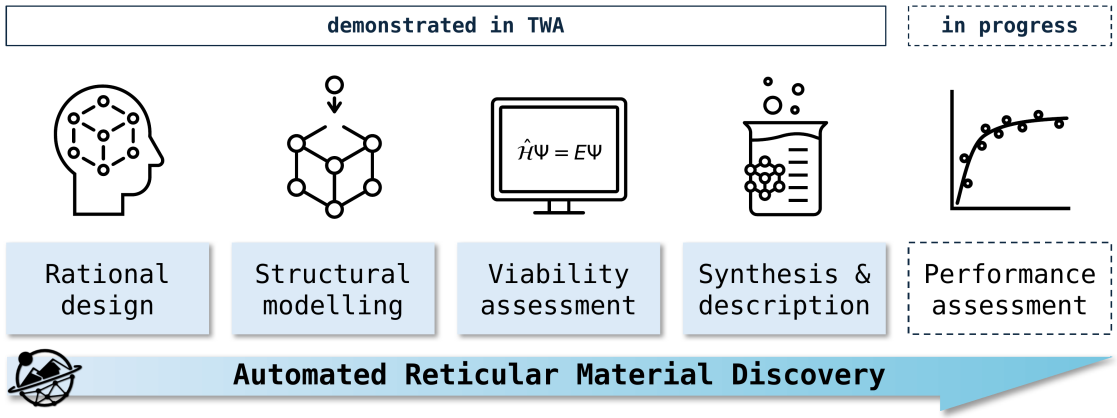



**Figure 2:** *Automated reticular material discovery using The World Avatar (TWA). The workflow integrates five stages: rational design of candidate materials from chemical building units and assembly models; structural modelling of the proposed architectures; computational assessment of their viability; prediction and structured description of synthesis procedures; and experimental assessment of the resulting materials. Explicit knowledge representation connects the individual stages, enabling the reuse of scientific information and the integration of computational predictions with experimental results. Adapted from [32].*

This is a meaningful demonstration of joining design, calculation, precedent, proposed synthesis and experimental evidence. It does not establish that every experimental decision and instrument action was autonomous, and the $CO_2$ capture performance and environmental impact of the new MOP remain to be established; $CO_2$ adsorption measurements are under way. The advance is the traceable connection among stages usually managed through separate files and workflows [36]; the actual efficiency and efficacy increase over a conventional workflow remains to be tested in a controlled comparison.

# 5 From material discovery to measurable sustainability

Faster discovery is not synonymous with sustainable discovery. Autonomous campaigns consume solvents, reagents, computing power and laboratory energy. An apparently superior sorbent may need energy-intensive regeneration, scarce elements or frequent replacement. Two outcomes therefore need separate evaluation: the resources used by the *research process*, and the life-cycle consequences of the *technology discovered*.

The research process itself can be measured. Green-chemistry metrics such as process mass intensity (PMI) and reaction mass efficiency (RME) [24] require quantitative records of material consumption and product yields. PMI accounts for all process materials, including solvents, whereas RME quantifies the efficiency with which reactants are incorporated into the desired product. Structured synthesis representations such as OntoSyn can capture the required data if experimental records are complete. An agentic optimisation campaign has computed these metrics alongside yield and identified more mass-efficient conditions than a human-directed campaign [25]. Solvent typically dominates input mass and must be tracked. Reporting energy and emissions per validated discovery would extend such accounting to computation [16].

TWA already has a concrete circular-economy application: the Material Passport Ontology (MPO) represents product composition, provenance and end-of-life flows for circularity assessment [34]. Industrial partners created ten passports for automotive components and wind-turbine blades, checked automatically against formal consistency rules written in the Shapes Constraint Language (SHACL). Graph queries computed circularity indicators, and calculated recoverable fibre yields were reported as comparable with recycling experiments. This shows how structured product information supports circularity calculations across stakeholders, but not reduced emissions; reliable end-of-life data remained difficult to obtain, and ten passports cannot establish generalisability.

Such work suggests a route from materials knowledge to sustainability assessment (Fig. 3). Adsorption data for a $CO_2$ sorbent could be linked to regeneration energy, operating lifetime and process emissions; battery performance to manufacturing and end-of-life flows; and catalyst energetics to reactor and process models. Cross-domain, cross-scale cou-

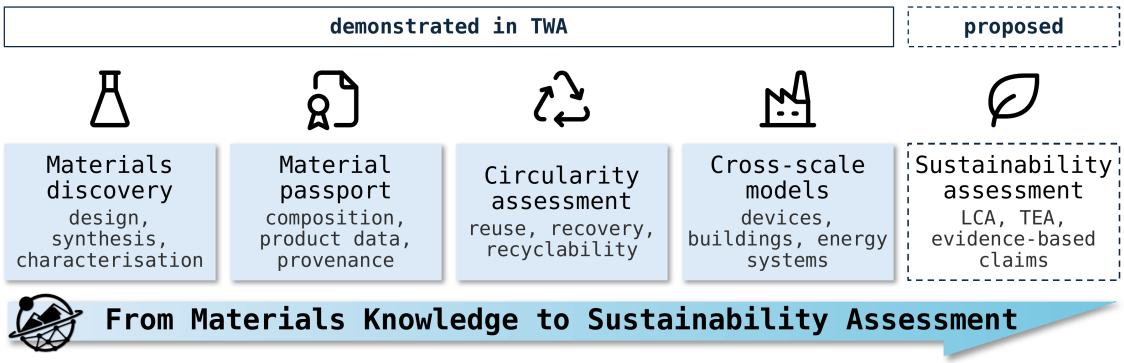



**Figure 3:** *From materials knowledge to sustainability assessment using The World Avatar (TWA). Explicit knowledge representation provides a foundation for connecting materials discovery, product composition, circularity assessment and system-level sustainability evaluation. The Material Passport Ontology demonstrates how structured product information can support circularity indicators and resource-recovery assessments. TWA has already coupled models across domains and scales at building and city level [37]; extending this coupling to link material properties to devices, industrial processes and energy systems offers a pathway towards integrated life-cycle assessment (LCA) and techno-economic analysis (TEA).*

pling has already been demonstrated from building to city level, linking buildings, energy, flooding, transport and air quality [37], while connected laboratory digital twins offer complementary infrastructure [28, 29]. Circularity indicators are not substitutes for life-cycle assessment, and any claim of environmental benefit requires explicit functional units, system boundaries, comparison technologies and uncertainty [17].

# 6 Challenges and a testable research agenda

Four questions should shape community investment:

1. *How much formalisation is enough?* Comprehensive ontologies for batteries, catalysts and polymers are costly to build, align and maintain; small, versioned modules with governed meanings and mappings may deliver more value than a universal schema.

2. *What is actually verified?* Tool validation can reject malformed requests, but not false measurements or unsupported inferences. Derivation records need source evidence, calibration, model versions and human oversight for consequential actions.

3. *Who bears the integration cost?* Legacy notebooks, instrument controllers and proprietary systems are difficult to connect [19], and federation requires security, permissions, reliable identifiers and resolution of conflicting interpretations.

4. *Who maintains the infrastructure?* Negative results, metadata curation and reusable software need lasting institutional support and academic credit [35].

Scientific knowledge also changes, and laboratories may use different definitions or uncertainty conventions. A trustworthy record therefore needs versioned claims and corrections that preserve the history of decisions rather than treating the latest graph state as unquestionable truth. Agent permissions should depend on the consequences of an action: querying a database is not equivalent to modifying a synthesis protocol or actuating equipment.

These difficulties call for discriminating benchmarks rather than platform demonstrations alone. A multi-laboratory study should compare an ontology-grounded workflow with well-engineered database/API alternatives on the same scientific tasks, reporting extraction accuracy, recoverability of evidence, interoperability costs, human interventions, experimental success, resource use and curation effort. Independent replication would be more persuasive than a claim of complete autonomy. The practical route is incremental: agree on minimal metadata for high-value experiments, keep original evidence accessible, expose validated actions through defined tools, and add cross-domain links only where they answer a scientific or sustainability question. The aim is interoperable and auditable research, not mandatory adoption of any one ontology, agent framework or storage technology.

# 7 Conclusion

Scientific AI will be most useful when its suggestions become inspectable decisions and its actions leave reusable evidence. What, then, should AI agents know? What is known and how it was derived; which actions they may take, through typed and checked tools; and what has already failed. The World Avatar shows that such knowledge can be built and used today, while the MOP study reveals what remains unproven, and where shared semantics across disciplines and scales add most value. The next standard for agentic materials science should be not simply a new material or a faster experiment, but a justified result whose provenance, limitations and consequences for sustainability can be independently assessed.

## Acknowledgements

This research project ("SM3") is part of the CREATE Thematic Programme in Decarbonisation and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. M.K. gratefully acknowledges funding from the European Union's Horizon Europe research and innovation programme under grant agreements 101188248 (CLIMATE-ADAPT4EOSC), 101226137 (TOGETHER), and 101309693 (RESILIACT), as well as support from the Alexander von Humboldt Foundation and the Massachusetts Institute of Technology. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript arising.

## Declaration on the use of generative AI

During the preparation of this manuscript, the authors used large language model tools for writing assistance, editing and proofreading. All output was reviewed and revised by the authors, who take full responsibility for the content of the publication. The figures were composed by the authors from open-source line icons; no generative-AI imagery was used.

## Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study. The World Avatar software and ontologies discussed here are openly available at https://github.com/TheWorldAvatar.

## Competing interests

M.K. is a director of Computational Modelling Cambridge Ltd. (CMCL) and Computational Modelling Pirmasens GmbH (CMPG); S.D.R. is Head of Research at CMPG. Both companies develop and use TWA. The authors declare no other competing interests.

# References


[1] M. Abolhasani and E. Kumacheva. The rise of self-driving labs in chemical and materials sciences. *Nature Synthesis*, 2(6):483–492, 2023. doi:10.1038/s44160-022-00231-0.

[2] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal digital twin – a dynamic knowledge graph. *Data-Centric Engineering*, 2:e14, 2021. doi:10.1017/dce.2021.10.

[3] A. Aspuru-Guzik and V. Bernales. Reaction: Building trust infrastructure for agentic experimental science beyond fragile eloquence. *Chem*, 12(8):103163, 2026. doi:10.1016/j.chempr.2026.103163.

[4] J. Bai, K. F. Lee, M. Hofmeister, S. Mosbach, J. Akroyd, and M. Kraft. A derived information framework for a dynamic knowledge graph and its application to smart cities. *Future Generation Computer Systems*, 152:112–126, 2024. doi:10.1016/j.future.2023.10.008.

[5] J. Bai, S. Mosbach, C. J. Taylor, D. Karan, K. F. Lee, S. D. Rihm, J. Akroyd, A. A. Lapkin, and M. Kraft. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nature Communications*, 15:462, 2024. doi:10.1038/s41467-023-44599-9.

[6] J. Bai, S. D. Rihm, A. Kondinski, F. Saluz, X. Deng, G. Brownbridge, S. Mosbach, J. Akroyd, and M. Kraft. twa: The World Avatar Python package for dynamic knowledge graphs and its application in reticular chemistry. *Digital Discovery*, 4(8): 2123–2135, 2025. doi:10.1039/D5DD00069F.

[7] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023. doi:10.1038/s41586-023-06792-0.

[8] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, and P. Schwaller. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 6(5):525–535, 2024. doi:10.1038/s42256-024-00832-8.

[9] P. W. V. Butler, S. D. Rihm, S. Mosbach, J. Akroyd, and M. Kraft. Inverse design of metal-organic polyhedra through molecular fragmentation and evolutionary optimisation. *Journal of Chemical Information and Modeling*, 66(7):3933–3943, 2026. doi:10.1021/acs.jcim.5c02956.

[10] P. W. V. Butler, A. Subramanian, S. D. Rihm, A. F. Fischer, T. S. Choksi, A. Kondinski, S. Mosbach, J. Akroyd, and M. Kraft. Computational modelling workflows for metal-organic polyhedra in The World Avatar. *Physical Chemistry Chemical Physics*, 28(24):14828–14838, 2026. doi:10.1039/d5cp04871k.

[11] A. Cissé, M. E. Cooper, M. Zhu, X. Evangelopoulos, and A. I. Cooper. Can we automate scientific reasoning in closed-loop experiments using large language models? *Digital Discovery*, 5(3):1132–1160, 2026. doi:10.1039/D5DD00520E.

[12] S. Clark, F. L. Bleken, S. Stier, E. Flores, C. W. Andersen, M. Marcinek, A. Szczesna-Chrzan, M. Gaberscek, R. Palacin, M. Uhrin, and J. Friis. Toward a unified description of battery data. *Advanced Energy Materials*, 12(17):2102702, 2022. doi:10.1002/aenm.202102702.

[13] A. I. Cooper, Q. J. Yeow, Z. Xiao, M. Zhu, A. Cissé, X. Evangelopoulos, and C. Boott. Large language models do not replace chemists in a closed-loop catalysis experiment. *Research Square preprint*, 2026. doi:10.21203/rs.3.rs-10032842/v1.

[14] P. Del Nostro, J. Friis, E. Ghedini, G. Goldbeck, O. Holtz, O. Roscioni, F. Zaccarini, and D. Toti. Elementary multiperspective material ontology: Leveraging perspectives via a showcase of EMMO-based domain and application ontologies. In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD)*, pages 135–142, 2024. doi:10.5220/0012910200003838.

[15] M. L. Evans, J. Bergsma, A. Merkys, C. W. Andersen, O. B. Andersson, D. Beltrán, E. Blokhin, T. M. Boland, R. Castañeda Balderas, K. Choudhary, A. Díaz Díaz, R. Domínguez García, H. Eckert, K. Eimre, M. E. Fuentes Montero, A. M. Krajewski, J. J. Mortensen, J. M. Nápoles Duarte, J. Pietryga, J. Qi, F. d. J. Trejo Carrillo, A. Vaitkus, J. Yu, A. Zettel, P. Baptista de Castro, J. Carlsson, T. F. T. Cerqueira, S. Divilov, H. Hajiyani, F. Hanke, K. Jose, C. Oses, J. Riebesell, J. Schmidt, D. Winston, C. Xie, X. Yang, S. Bonella, S. Botti, S. Curtarolo, C. Draxl, L. E. Fuentes Cobas, A. Hospital, Z.-K. Liu, M. A. L. Marques, N. Marzari, A. J. Morris, S. P. Ong, M. Orozco, K. A. Persson, K. S. Thygesen, C. Wolverton, M. Scheidgen, C. Toher, G. J. Conduit, G. Pizzi, S. Gražulis, G.-M. Rignanese, and R. Armiento. Developments and applications of the OPTIMADE API for materials discovery, design, and data exchange. *Digital Discovery*, 3:1509–1533, 2024. doi:10.1039/D4DD00039K.

[16] G. Giannopoulos, M. Chhowalla, A. Bielefeld, M. Chi, Y. Gogotsi, J. Nanda, and E. A. Kocs. Accelerating sustainable energy storage and conversion with AI and ML: A perspective on emerging directions. *MRS Energy & Sustainability*, Nov. 2026. doi:10.1557/s43581-026-00171-8.

[17] S. Hellweg and L. Milà i Canals. Emerging approaches, challenges and opportunities in life cycle assessment. *Science*, 344(6188):1109–1113, 2014. doi:10.1126/science.1248361.

[18] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, F. F. Ramirez, C. S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky, and G. Pizzi. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific Data*, 7:300, 2020. doi:10.1038/s41597-020-00638-4.

[19] H. Joress, B. DeCost, K. Jones, A. G. Kusne, A. McDannald, Z. Trautt, and F. Tavazza. Toward a composable, modular laboratory ecosystem for au-

tonomous materials research and development. *Matter*, 9(5):102756, 2026. doi:10.1016/j.matt.2026.102756.

[20] A. Kondinski, A. Menon, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Automated rational design of metal-organic polyhedra. *Journal of the American Chemical Society*, 144(26):11713–11728, 2022. doi:10.1021/jacs.2c03402.

[21] S. X. Leong, C. E. Griesbach, R. Zhang, K. Darvish, Y. Zhao, A. Mandal, Y. Zou, H. Hao, V. Bernales, and A. Aspuru-Guzik. Steering towards safe self-driving laboratories. *Nature Reviews Chemistry*, 9(10):707–722, 2025. doi:10.1038/s41570-025-00747-x.

[22] H. Lyu, Z. Ji, S. Wuttke, and O. M. Yaghi. Digital reticular chemistry. *Chem*, 6(9): 2219–2241, 2020. doi:10.1016/j.chempr.2020.08.008.

[23] B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette. Self-driving laboratory for accelerated discovery of thin-film materials. *Science Advances*, 6(20):eaaz8867, 2020. doi:10.1126/sciadv.aaz8867.

[24] C. R. McElroy, A. Constantinou, L. C. Jones, L. Summerton, and J. H. Clark. Towards a holistic approach to metrics for the 21st century pharmaceutical industry. *Green Chemistry*, 17(5):3111–3121, 2015. doi:10.1039/C5GC00340G.

[25] M. Müller, J. Bai, W. Gottstein, A. Mandal, M. Nazeri, E. Savino, Y. Fang, S. Das, S. P. García Carrillo, Y. Kang, J. B. Pérez-Sánchez, S. Pilon, M. Fitzner, T. Noël, F. Gu, V. Bernales, and A. Aspuru-Guzik. La Agente Óptima: Towards agentic self-driving laboratories. *arXiv preprint arXiv:2609.04564*, 2026. doi:10.48550/arXiv.2609.04564.

[26] L. Pascazio, S. D. Rihm, A. Naseri, S. Mosbach, J. Akroyd, and M. Kraft. Chemical species ontology for data integration and knowledge discovery. *Journal of Chemical Information and Modeling*, 63(21):6569–6586, 2023. doi:10.1021/acs.jcim.3c00820.

[27] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist. Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601):73–76, 2016. doi:10.1038/nature17439.

[28] S. D. Rihm, J. Bai, A. Kondinski, S. Mosbach, J. Akroyd, and M. Kraft. Transforming research laboratories with connected digital twins. *Nexus*, 1(1):100004, 2024. doi:10.1016/j.ynexs.2024.100004.

[29] S. D. Rihm, Y. R. Tan, W. Ang, H. Y. Quek, X. Deng, M. T. Laksana, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. The Digital Lab Facility Manager: Automating operations of research laboratories through "The World Avatar". *Nexus*, 1(3):100031, 2024. doi:10.1016/j.ynexs.2024.100031.

[30] S. D. Rihm, F. Saluz, A. Kondinski, J. Bai, P. W. V. Butler, S. Mosbach, J. Akroyd, and M. Kraft. Extraction of chemical synthesis information using The World Avatar. *Digital Discovery*, 4(10):2893–2909, 2025. doi:10.1039/D5DD00183H.

[31] S. D. Rihm, D. N. Tran, A. Kondinski, L. Pascazio, F. Saluz, X. Deng, S. Mosbach, J. Akroyd, and M. Kraft. Natural language access point to digital metal-organic polyhedra chemistry in The World Avatar. *Data-Centric Engineering*, 6:e22, 2025. doi:10.1017/dce.2025.12.

[32] S. D. Rihm, P. W. V. Butler, K. Elumalai, A. F. Fischer, P. C. Lim, X. Rong, T. S. Choksi, A. Kondinski, and M. Kraft. Prediction and synthesis of novel metal-organic polyhedra in The World Avatar. *ChemRxiv preprint*, 2026. doi:10.26434/chemrxiv.15007397/v1. Also available as c4e-Preprint 347, Cambridge Centre for Computational Chemical Engineering.

[33] M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz, and C. Draxl. NOMAD: A distributed web-based platform for managing materials science research data. *Journal of Open Source Software*, 8(90):5388, 2023. doi:10.21105/joss.05388.

[34] M. H. Seddiqui, F. Farazi, F. Giunchiglia, M. Bagchi, S. Bocca, J. Akroyd, S. Mosbach, F. Arifin, R. Ahmed, M. Arman, L. Di Cesare, A. Pipino, U. Halaçoğlu, Z. Korkmaz, T. Islam, K. Kižys, D. Gendvilas, A. Berkay, S. Yildiz, and M. Kraft. A material passport ontology for a circular economy. *Computers in Industry*, 178:104465, 2026. doi:10.1016/j.compind.2026.104465.

[35] E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C. P. Gomes, J. M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S. K. Saikin, S. Smullin, V. Stanev, and B. Maruyama. Autonomous experimentation systems for materials development: A community perspective. *Matter*, 4(9):2702–2726, 2021. doi:10.1016/j.matt.2021.06.036.

[36] H. S. Stein and J. M. Gregoire. Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chemical Science*, 10(42): 9640–9649, 2019. doi:10.1039/C9SC03766G.

[37] Y. R. Tan, M. Hofmeister, S. Z. Phua, G. Brownbridge, K. Rustagi, J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Beyond connected digital twins – Can digital twins really deliver sustainable cities? *Sustainable Cities and Society*, 131:106596, 2025. doi:10.1016/j.scs.2025.106596.

[38] G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid, and A. Aspuru-Guzik. Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 124(16):9633–9732, 2024. doi:10.1021/acs.chemrev.4c00055.

[39] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, 571(7763):95–98, 2019. doi:10.1038/s41586-019-1335-8.

[40] A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller, and T. Laino. Automated extraction of chemical synthesis actions from experimental procedures. *Nature Communications*, 11(1):3601, 2020. doi:10.1038/s41467-020-17266-6.

[41] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 2016. doi:10.1038/sdata.2016.18.

[42] X. Zhou, A. Eibeck, M. Q. Lim, N. B. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator – a knowledge graph for the process industry. *Computers & Chemical Engineering*, 130:106577, 2019. doi:10.1016/j.compchemeng.2019.106577.

[43] X. Zhou, M. Q. Lim, and M. Kraft. A smart contract-based agent marketplace for the J-Park Simulator – a knowledge graph for the process industry. *Computers & Chemical Engineering*, 139:106896, 2020. doi:10.1016/j.compchemeng.2020.106896.

[44] X. Zhou, P. W. V. Butler, C. Yang, S. D. Rihm, T. Angkanaporn, J. Akroyd, S. Mosbach, and M. Kraft. Ontology-to-tools compilation for executable semantic constraint enforcement in LLM agents. *arXiv preprint arXiv:2602.03439*, 2026. doi:10.48550/arXiv.2602.03439. Also available as c4e-Preprint 343, Cambridge Centre for Computational Chemical Engineering.