Brokering between tenants for an international materials acceleration platform

Vogler, Monika; Busk, Jonas; Hajiyani, Hamidreza; Jørgensen, Peter Bjørn; Safaei, Nehzat; Castelli, Ivano E.; Ramirez, Francisco Fernando; Carlsson, Johan; Pizzi, Giovanni; Clark, Simon

Total number of authors:
13

Published in:
Matter

Link to article, DOI:
10.1016/j.matt.2023.07.016

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Matter

CellPress
OPEN ACCESS

Perspective

Brokering between tenants for an international materials acceleration platform

Monika Vogler,[1,2] Jonas Busk,[3] Hamidreza Hajiyani,[4] Peter Bjørn Jørgensen,[3] Nehzat Safaei,[4] Ivano E. Castelli,[3] Francisco Fernando Ramirez,[5,6] Johan Carlsson,[4] Giovanni Pizzi,[5,6,8] Simon Clark,[7] Felix Hanke,[9,*] Arghya Bhowmik,[3,*] and Helge S. Stein[1,2,10,*]

SUMMARY

The efficient utilization of resources in accelerated materials science necessitates flexible, reconfigurable software-defined research workflows. We demonstrate a brokering approach to modular and asynchronous research orchestration to integrate multiple laboratories in a cooperative multitenancy platform across disciplines and modalities. To the best of our knowledge, this constitutes the first internationally distributed materials acceleration platform (MAP) linked via a passive brokering server, which is demonstrated through a battery electrolyte workflow capable of determining density, viscosity, ionic conductivity, heat capacity, diffusion coefficients, transference numbers, and radial distribution functions that ran in five countries over the course of 2 weeks. We discuss the lessons learned from multitenancy and fault tolerance and chart a way to a universal battery MAP with fully ontology-linked schemas and cost-aware orchestration.

INTRODUCTION

Scientists have continuously innovated on the efficiency improvements of the research process[1] with notable driving accelerators such as automation of research tasks[2] and their integration with data lineage tracking[3,4] in biotechnology and materials science.[5] This research automation created the field of combinatorial materials science (CMS),[6] which utilizes the paradigm of well-defined composition and processing variation to unravel the underlying physicochemical relationships at a greater pace.[7,8] The combinatorial nature of chemical space does, however, render any brute force exploration for discovery ineffective. Efficient research therefore necessitates approaches that can predict how to design materials according to a target functional property. This inverse design[9–11] approach is rooted in the idea of descriptors based on rational design[12,13] that define composition-structure-property relationships. Recent advancements in research instrumentation[14] and user interfaces have lowered the entrance barrier for high-throughput experimentation (HTE). Abundance of searchable data, enabled by data management,[3,15] then sparked the proliferation of data-driven methods in synthesis,[16] characterization,[17] performance evaluation,[18] and interpretation.[19] The integration of accelerated research tasks in workflows[2] guided by data-driven methods was then conceived as the next evolutionary step in a 2018 workshop initiated by the UN Mission Innovation Initiative and was called the materials acceleration platform (MAP).[20]

PROGRESS AND POTENTIAL

New developments in modern technology often rely on specifically designed materials. Tailoring materials properties to meet the requirements of an application is a key task in materials research. The increasing pace of technology development requires faster discovery of optimized materials. Materials acceleration platforms (MAPs) automate materials optimization and therefore allow for a considerable acceleration of materials discovery. MAPs allow for a more efficient use of research infrastructure like laboratories, devices, and computational resources. We demonstrate a modular MAP design, which allows us to include automated as well as non-automated units. It can hence be used in the current research landscape already. Future development is envisioned to promote comprehensive recording of data and metadata and to expose laboratories as a service. This is expected to result in the usage of research equipment to capacity, maximizing the return on invest of research funding.

Here, we define MAPs as platforms that enable multiple units, which we call tenants, to connect using standardized communication schemas. A MAP should combine tenants providing physical services, like, e.g., experimental capabilities using multi-tenant instruments for measurements and simulations, as well as digital services for, e.g., data processing or machine learning. In the field of computer science, there is a nomenclature for multitenancy that relates to shared digital services serving multiple customers from a shared resource. We would like to avoid the competing nomenclature of an agent, as it semantically implies an entity with hidden or potentially malicious features as well as a hierarchy. To express a collaborative and community spirit, we choose to refer to all units within a MAP as tenants. MAPs should be designed such that tenants providing the same capabilities can be exchanged on the fly without affecting the platform. This implies that an operator or optimizer cannot *a priori* know who is going to fulfill a request at what time, requiring all tenants to be asynchronous. This asynchronous operation of tenants and the platform itself means that processing requests and result communication function independently. While the sharing of data should be as open as possible, resources need to be protected against malicious input, which renders security aspects a vital part of the MAP design.

The concept of connecting distributed resources through networking instances is well established in the field of computer science. In materials science, applications of this concept are, however, often focused on the processing and management of generated research data[21,22] or automation on a lab bench,[23] in a building,[24] or at a single research institution.[25] An example is large-scale research infrastructures and their data-processing pipelines and visions to integrate these into so-called superfacilities.[26] Using, for instance, the Globus[21] framework, single experimental setups are integrated into computational fabrics that enable the transfer and analysis of data at rates of gigabytes per second.

Going beyond a single materials property or even transcending the laboratory is necessary for a true battery MAP,[27] as there are multiple functional properties to optimize,[28,29] e.g., lifetime, energy capacity, energy density, voltage range, cost, power density, safety, embodied energy, etc. All these critically depend on high-dimensional parameter spaces comprising the chemistry, structure, and processing of the materials as well as the assembly, composition, and environmental conditions during operation of the cell system. The optimization of battery materials is therefore a truly multiscale challenge ranging from atoms to systems and from nanoseconds to years and therefore cannot be mastered by a single institution or domain alone. A complicating factor in all of this is the mutual exclusion of certain methods, i.e., high-throughput density functional theory (DFT) calculations are challenging to run on the month scale,[9,30] and manufactured batteries cannot easily be analyzed by, e.g., X-ray photoelectron spectroscopy (XPS). This is in stark contrast to all of the early MAP-inspired demonstrations from other fields,[31–33] where the entire research life cycle can be covered either using a single robotic setup or a single research group. A comprehensive MAP in battery research therefore needs to be distributed across different domains, spanning the entire set of battery research processes.[34]

Here, we describe the design of a cooperative MAP that is spatially distributed across multiple countries and involves scientists from the experimental and modeling domains. A graphical representation of our MAP concept is shown in Figure 1. The development involved partners across Europe associated with the BIG-MAP project.[15] A brokering software system called FINALES (fast intention-agnostic

[1]Helmholtz Institute Ulm, 89073 Ulm, Germany

[2]Institute for Physical Chemistry, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

[3]Department of Energy Conversion and Storage, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, Denmark

[4]Dassault Systèmes, 51063 Cologne, Germany

[5]Theory and Simulation of Materials (THEOS), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[6]National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[7]SINTEF Industry, Battery and Hydrogen Technologies, 7034 Trondheim, Norway

[8]Laboratory for Materials Simulations (LMS), Paul Scherrer Institut (PSI), 5232 Villingen, Switzerland

[9]Dassault Systèmes, 334 Science Park, Cambridge CB4 0WN, UK

[10]Present address: Department of Chemistry, School of Natural Sciences, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany

*Correspondence: felix.hanke@3ds.com (F.H.), arbh@dtu.dk (A.B.), helge.stein@tum.de (H.S.S.)

https://doi.org/10.1016/j.matt.2023.07.016

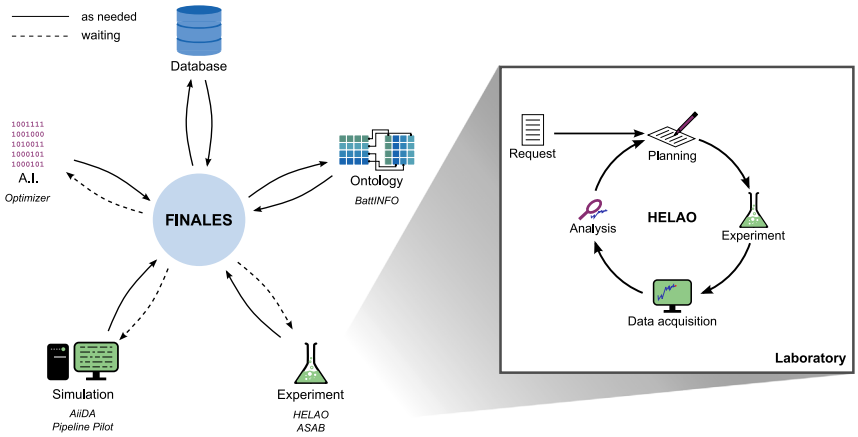

Figure 1. Schematics of conventional accelerated research labs and brokered multitenant MAPs

The left part of the figure shows the general layout of the FINALES paradigm to research orchestration in which multiple tenants, i.e., optimizers, experiments, simulations, and databases, exist in a cooperative environment in which each participant can be located anywhere on the globe. The communication within the MAP is based on pulling from the server in timed intervals, labeled "waiting," and pushing information to the server, labeled "as needed." The inset presents an emblematic workflow of a HELAO[25] run in which multiple instruments can be pooled together, but typically a run is limited to a single laboratory. The design of our MAP allows for the integration of a HELAO workflow as one of its tenants.

learning server) was developed and hosted at the Karlsruhe Institute of Technology (KIT, Germany), laboratory experiments were performed at the Helmholtz Institute Ulm (HIU, Germany), computer simulations were performed at Dassault Systèmes (3DS, Germany and UK), a machine-learning optimizer was developed and run at the Technical University of Denmark (DTU, Denmark), and ontology and data interfaces were prepared at DTU, Stiftelsen for industriell og teknisk forskning (SINTEF, Norway), and École polytechnique fédérale de Lausanne (EPFL, Switzerland).

We demonstrate the communication in our MAP during an explorative deployment using a setup consisting of the brokering software, the laboratory experiments, the computer simulations, and the machine-learning-based optimizer. This MAP can thus determine the density, viscosity, ionic conductivity, heat capacity, diffusion coefficients, transference numbers, and radial distribution functions for various formulations of lithium-ion battery electrolytes. To the best of our knowledge, this is the first demonstration of an internationally distributed MAP in the field of battery research.

The lessons learned from the development and real-world deployment of our MAP involving many partners are summarized in a separate section at the end of this perspective. We believe that sharing not only the success but also the shortcomings of these design considerations of this new paradigm of performing research will help to accelerate the progression toward truly autonomous and globally interconnected materials discovery.

METHODS AND TENANTS

Our MAP concept comprises various units collaborating via the MAP. We call each participating unit in our MAP a tenant and differentiate between intention-aware and intention-agnostic tenants. The optimization task, and hence the intention of the resulting requests, is known to intention-aware tenants, which act actively and thereby "inject intent" into the workflow. In contrast, intention-agnostic tenants

are not aware of the reasoning behind requests. They fulfill requests matching their capabilities without considering any optimization or experimental campaigns. In this context, it is essential that the brokering server FINALES is not just intention agnostic but purely passive to fully enable the communication in our MAP.

In this section, we will introduce the different tenants, which provide interfaces compatible with our broker server FINALES, namely the brokering server itself; the experimental setup for density and viscosity measurements ASAB (autonomous synthesis and analysis of battery electrolytes); the simulation orchestrator implemented in Pipeline Pilot; the AiiDA interface for data storage and machine-learning-based conductivity predictions; and the machine-learning optimizer based on a Gaussian process optimizer in combination with Chimera[35] to enable multiobjective optimization.

Broker server: FINALES

The FINALES tenant is designed as a broker server mediating between all the tenants in the MAP. It accepts requests from all the tenants through a web application programming interface (API), which requires requests and replies to be structured in defined Pydantic[36] schemas in JSON format. This structured representation of the data and requests facilitates the integration of an ontology such as BattINFO.[37] Furthermore, the type checking promoted by Pydantic improves the security of FINALES. The formalized and human-readable communication protocols allow us to include non-automated data acquisition in autonomously orchestrated loops, maximizing the inclusiveness and versatility of FINALES.

Tenants use the requests to order or report a measurement or a simulation or to retrieve data from the database connected to FINALES. The requests for measurements and simulations as well as reported results are stored in the FINALES database in an append-only scheme preserving all generated data. FINALES provides methods to request data from the database by its origin, meaning whether it was generated from an experiment or a simulation, or by the measured or simulated quantity. The respective request by the tenant contains a field specifying the method to use. The queries from the database are implemented as functions in FINALES, which take an element of an enumeration as an input. This way of working allows for type checking and enhances the security of the MAP against malicious database queries. FINALES performs all its actions in an asynchronous manner, allowing the MAP to continue its operation while requests are pending. The core functionality of FINALES is to act as a communication hub between the tenants while not actively controlling any tenant. This means that it is not actively triggering actions of individual tenants but rather keeps track of a queue of requests and replies, which can be queried by the various tenants in the MAP at any time. This passivity renders FINALES independent of the state of the individual tenants and distinguishes it from many other automation frameworks like HELAO,[25] Bluesky,[38] or ChemOS.[23] By design, instruments operated by these other frameworks or Globus[21] flows could be connected to FINALES as tenants.

A further difference between FINALES and HELAO is FINALES's hardware-independent operation by abstracting away from a sequence of hardware events to requesting desired outcomes. FINALES allows for maximum flexibility and redundancy by explicitly allowing the integration of several intention-agnostic tenants with the same capabilities as well as multiple intention-aware tenants, like optimizers, operating simultaneously and in parallel. This multitenancy design allows for several intention-aware tenants to explore and exploit the data and services available through FINALES.

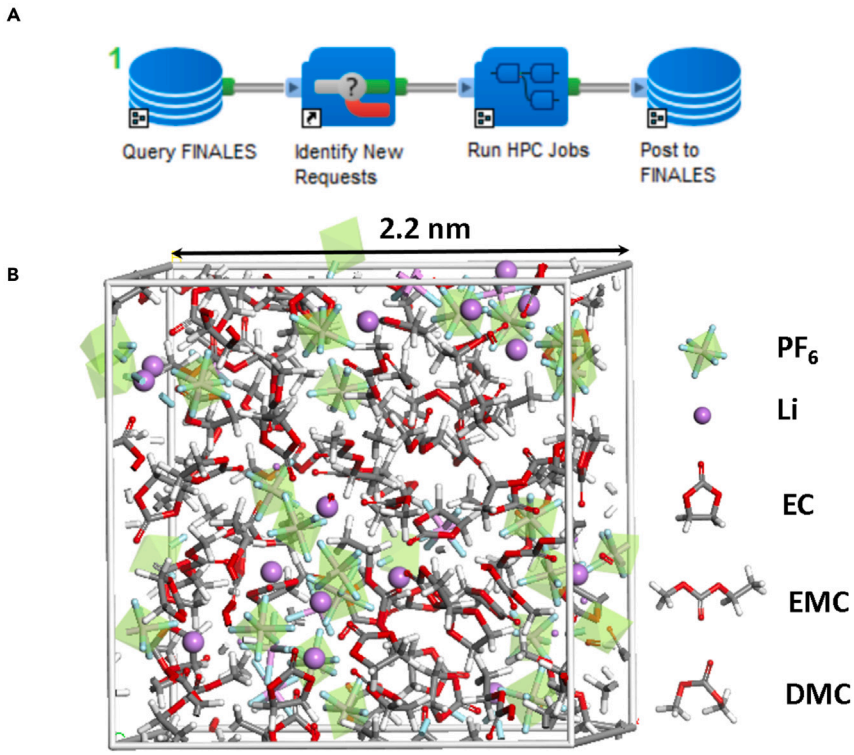

Figure 2. The workflow and the atomistic unit cell used by the simulation tenant

(A) The workflow architecture for automated simulations using BIOVIA Pipeline Pilot. The first component queries the FINALES broker server for simulation requests, which are read and input parameters to perform MD simulations, are extracted. The complete workflow for the MD simulations is contained in the "run HPC job" subprotocol. The results of the MD simulations are obtained asynchronously and posted back onto FINALES upon convergence via the "post to FINALES" component.

(B) The atomistic supercell model of an electrolyte. The ethylene carbonate (EC), ethyl methyl carbonate (EMC), and dimethyl carbonate (DMC) molecules are shown in a stick representation, the lithium ions are shown as large violet balls, and the hexafluorophosphate (PF_6^-) ions are represented as a stick model enclosed by a light green octahedron.

Simulation orchestrator: Pipeline Pilot

The intention-agnostic simulation tenant provides simulation capabilities to the MAP and is created using BIOVIA Pipeline Pilot.[39] This is a workflow engine and data analysis tool, which enables reading and writing of data from and to databases and the automation of simulations and experiments as well as the analysis, model building, visualization, and reporting of results. The manipulation of data is executed by components, which are connected via data pipelines into workflows called protocols. Figure 2A shows how the protocol used as the simulation tenant in our demonstration may be graphically represented. The first component in the protocol periodically queries the broker server for pending simulation requests using the FINALES API. The relevant input parameters required to set up a molecular dynamics (MD) simulation are extracted once a new simulation request is received and passed on to a high-performance computing (HPC) cluster for execution.

The core calculations of the protocol are performed in the "run HPC job" subprotocol component, which contains a complete workflow to calculate the ionic transport coefficients of a given electrolyte composition[40,41] at a given temperature. Briefly, this calculation starts by building an Amorphous Cell model of the molecular liquid

as shown in Figure 2B. This model is the starting configuration to run a sequential MD workflow containing an initialization stage MD run under constant volume and temperature (NVT-MD) followed by an MD run under constant pressure and temperature (NPT-MD) for 250 ns each to thermalize the system and to obtain an estimate for the density. This equilibrated amorphous model is used as a starting configuration for a 2,000 ns production MD simulation run. An automatic analysis of the MD trajectory provides diffusion coefficients, conductivities, and transference numbers for all ionic species in the system as described in detail by Hanke et al.[40] In addition, the radial distribution function (RDF) is computed, which provides detailed information about the ordering of the molecules in the liquid.

Finally, the results of the MD simulations are collected and formatted for reporting back to FINALES in the last component called "post to FINALES." The metadata, such as trajectories, can also be sent simultaneously to the Materials Cloud Archive. This protocol runs asynchronously in specific time intervals, and each request is treated independently for querying, simulation, and posting.

Experimental setup: ASAB

The ASAB system used to perform the experiments described in this study is shown in Figure 3. In the demonstration, it serves as an intention-agnostic tenant that comprises a base module for power supply, six syringe pumps, and ten eleven-port valve modules including a matching API commercially available from CETONI. Furthermore, an Entris II laboratory balance from Sartorius AG and a DMA 4100 M densimeter with an additional Lovis 2000 viscometer by Anton Paar Germany are connected to the system. A computer is used to control the system via an in-house-implemented Python package. The software package comprises servers for the densimeter and viscometer, the pump and valve system, and the balance, all based on the FastAPI framework.[42] These servers call functions in an actions level within the in-house-developed software, which run more complex tasks by calling functions in the driver level. The tubing and interconnections of pumps, valves, vials, and connected devices are digitally represented by a graph. This allows searching for the shortest connection between two nodes or paths passing defined nodes in a given order. Its graph implementation makes ASAB flexible regarding the tubing interconnections and enables processing of high-level requests even after reconfiguring the hardware. Prior to any mixing step, the system fills every syringe related to a reservoir containing a component included in the formulation to avoid gas being located at the top of the syringes, which results in flawed mixing ratios. The hardware layout is loaded from a configuration file, which needs to be supplied by the operator prior to starting an experiment. Requests are pulled from FINALES, and the corresponding functionality is triggered on the device requested. The setup offers functionalities to request mixing, providing a sample to a device, retrieving data, and draining the sample from the device. The deduction of detailed step-by-step procedures from these high-level requests is performed within the ASAB actions and drivers, exploiting its internal flexibility.

Since the simulations work on a molecular representation of the system, the optimizer is chosen to represent formulations as a set of molar fractions. Requests for formulations therefore need to be transformed to volume fractions by the ASAB software to enable volumetric dosing of the stock solutions. In case there is no exact solution to the transformation, the system selects the closest accessible formulation using a gradient descent method provided in the SciPy[43] Python package.

Based on the volume fractions calculated according to this procedure, the volume flow of each stock solution is determined by multiplication of the targeted total

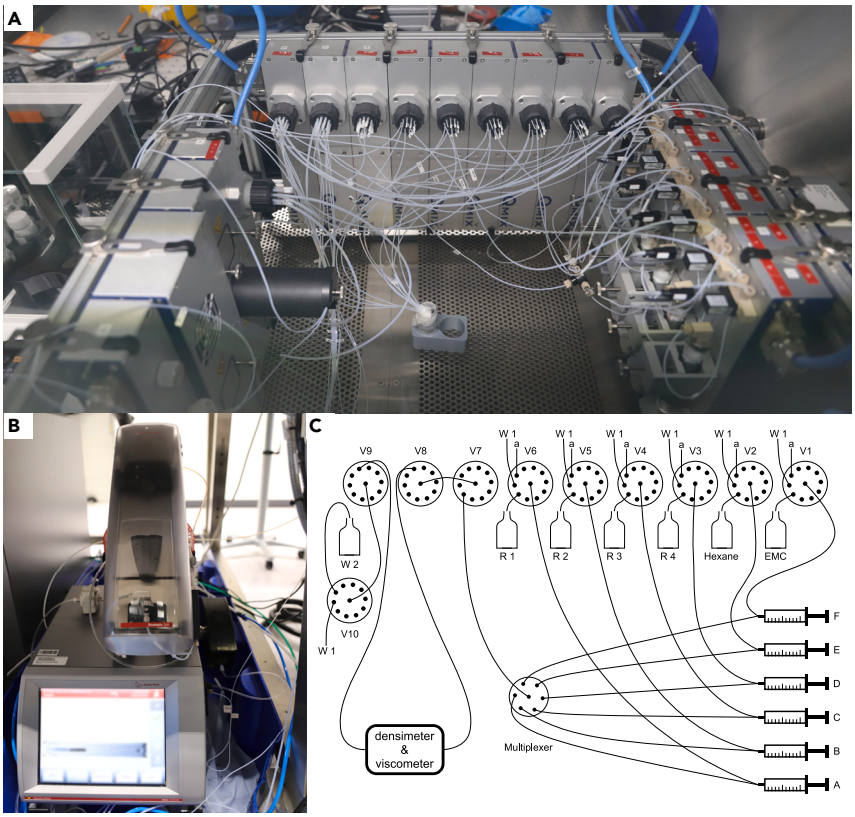

Figure 3. The hardware setup and a schematic overview of the ASAB tenant

(A) The pump and valve system of the experimental ASAB setup used to perform the experiments relevant for this study.

(B) The densimeter and the attached viscometer device used for the measurements.

(C) A schematic drawing showing the connections within the system.

The valves, reservoirs containing the stock solutions (R1–R4), cyclohexane, and EMC for cleaning are shown. Further, the syringes (A–F), the multiplexer for mixing, and the densimeter and viscometer device are displayed.

flow with the volume fraction of the respective stock solution. The flows of the stock solutions are merged in the multiplexer before a valve determines whether the flow is directed toward the waste while stabilizing the flows or toward the measuring device. After the measurement ends, the sample is drained to the waste using gas aspirated from the atmosphere in the glovebox, in which the pumps and valves are located. Prior to the subsequent experiment, gas is pumped through the measuring cells of the densimeter and viscometer several times to remove large residues of sample material.

Optimizer tenant

The intention-aware optimizer tenant software is designed in a modular and configurable way using the Python programming language. One module is responsible for all communication with FINALES, including authentication, data retrieval, and sending new measurement or simulation requests. Another module is responsible for data preprocessing and collating data in a table format appropriate for data analysis. A third module contains the optimizer algorithm, responsible for analyzing data and suggesting new measurement inputs. In the current version of the optimizer tenant, the optimizer algorithm is a Gaussian process optimizer based on the widely

used GP-UCB algorithm[44] extended for multiobjective optimization with the Chimera[35] scalarizing function. Finally, a central application module binds the other modules together in a main loop by continuously checking the current state of FINALES's database, retrieving available data, fitting the machine-learning model with the data, and using the optimizer algorithm to generate new measurement and simulation requests, which are finally sent to FINALES. The optimizer tenant can be configured to specify what to optimize, which quantities to request, and how to reach the broker server.

Low-fidelity conductivity predictor integrated through an AiiDA tenant

AiiDA[4,45,46] is a general-purpose workflow management platform for computational research projects, with a plugin interface to support external codes. Currently, the vast majority of the plugins cover simulations in the field of materials science. AiiDA provides a framework to codify and automate the different tasks involved in computational workflows and is able to seamlessly integrate with simulation codes running on HPC clusters. A generic intention-agnostic AiiDA tenant can thus grant FINALES users access to a wide variety of simulation tools without needing to implement specific infrastructure to interact with the computational server, similar to those offered by the Pipeline Pilot simulation orchestrator discussed earlier. Furthermore, it can also significantly simplify the management of the computing resources used to run these calculations, as well as the coordination between multiple simulation codes required to produce a given figure of merit (FOM), i.e., the implementation of a complete workflow.

The AiiDA-FINALES tenant demonstrates the interaction of a simulation managed by AiiDA with FINALES. A command line interface is also implemented to simplify the process of starting the tenant, and also provides an easy way of populating requests in the server, which is particularly useful for testing. Here, we demonstrate the use of the AiiDA tenant to estimate a low-fidelity value for the conductivity, based on the model reported by Rahmanian et al.[47] developed using one-shot active learning, in order to provide multifidelity data to the MAP. The model was designed for solutions of lithium hexafluorophosphate ($LiPF_6$) in mixtures of ethylene carbonate (EC), ethyl methyl carbonate (EMC), and propylene carbonate (PC), in a range of temperatures reaching from −30°C to 60°C. The original model was expressed as a polynomial on the mass ratios of the participating species, in which the coefficients were determined for specific temperatures within the range provided in increments of 10°C. In the implementation used here, any temperature is accepted, but processing is only possible for temperatures within −30°C to 60°C. The parameters are interpolated from those corresponding to the two nearest multiples of ten. A filter is implemented that only processes requests from the AiiDA internal queue, which are within the temperature range of the model. Once the tenant is started, it connects to the server, pulls any request for conductivity simulations, and submits any simulation that is not already queued to the AiiDA internal queuing system. It then monitors the state of the submitted calculations in the internal queue of AiiDA, and when they finish, it submits back the measurement results to FINALES. The overall data and process flow is analogous to the one discussed for the Pipeline Pilot simulation orchestrator in Figure 2A.

Our implementation serves to demonstrate AiiDA's capabilities that can be enabled by connecting AiiDA as a tenant to FINALES. Further features, such as a persistence mechanism to allow the tenant to be restarted or error handling, will be implemented in the future, also leveraging future features of the FINALES broker.

In the following sections, we describe the design of our MAP, the reduced setup used to demonstrate the operability of the communication, the results obtained from the demonstration run, the lessons learned, and future goals.

MAP design

Current implementations of MAPs often limit themselves to loops[48–50] optimizing for a specific target by using workflows, which are predefined to a certain degree, i.e., a fixed sequence of events is given, and instrument or research actions are predetermined. Battery research, or energy conversion research in a broader sense, is, however, intrinsically multimodal and multitarget. Consequently, our MAP allows us to combine hardware and software components, providing complementary capabilities and aiming to minimize error-prone human interventions while increasing the experimental and computational efficiency and maximizing information generation in a time- and cost-efficient manner.

A defining design decision, which is specific to our approach, is to disable direct communication between tenants and to channel all communication through a brokering service, which we implemented. FINALES constitutes an open interface brokering between hardware and software as well as humans. Simulations and experiments can be requested as a service through the FINALES API. By implementing an interface to the FINALES API, new tenants can easily be connected to our MAP without necessitating alterations of existing workflows in the MAP or the tenant.

Further, it is designed to be highly modular with respect to possible workflows, which, in the future, is expected to allow for various independent optimizations running simultaneously on a common MAP. For this to work, the design and mode of operation of our MAP must be independent of the target of any optimization running on the MAP. This necessitates a separation of concerns among all the tenants, which results in the aforementioned distinction between intention-aware and intention-agnostic tenants. Moreover, FINALES is prepared to allow for multitenancy. Several intention-agnostic tenants offering the same physical or computational capabilities and multiple intention-aware tenants with different intentions are planned to operate in the same MAP. This will allow the MAP to be utilized for multiple purposes at once and may dramatically increase its utility as a general research framework.

This concept requires the central instance, here FINALES, to be passive, i.e., it does not actively trigger actions in the MAP. This allows for asynchronous operation of the individual tenants by enabling them to fetch new tasks from the server according to their own schedule. This passive operation of FINALES allows us to continue optimization independent of any single tenant's online availability or load. If any equipment or compute cluster is removed from or added to the MAP, the queuing system of FINALES remains unaffected. Requests, which cannot be served by any tenant in the MAP, will remain in the queue until a tenant with the required capabilities connects to the MAP and picks up the request. This behavior allows for the handling of errors on the tenant level without affecting the further operation of the remaining MAP. If a tenant does not report results due to an internal error, the request remains pending, and the MAP will proceed its operation. Hence, there is no need for FINALES to deal with tenant-internal errors. Errors in tenants, which can occur and still allow for results to be reported, need to trigger an entry in the quality field of the response to enable intention-aware tenants to withdraw the data to treat these data according to their standards. Fault tolerance toward a failure of FINALES can be achieved in principle by running redundant servers and to switch to another instance as soon as one instance cannot be reached.

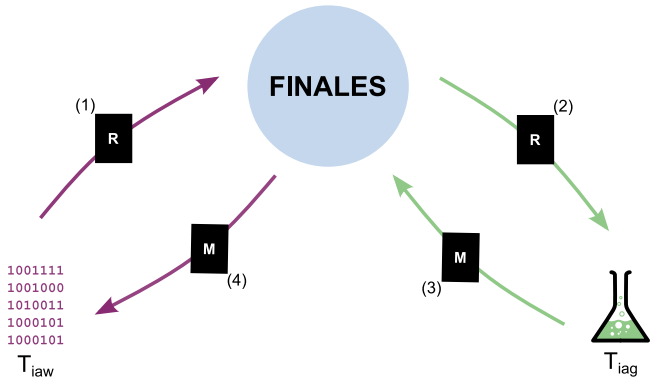

Figure 4. The workflow for performing an experiment or a simulation in our MAP

The intention-aware tenant (T_{iaw}) posts a request (R) to FINALES (1), where it is picked up by the intention-agnostic tenant (T_{iag}) (2), which processes the request and posts back the results (M) to FINALES (3). The intention-aware tenant collects the data from FINALES (4) and operates on them according to its task. Links to the BattINFO[37] ontology are prepared for all applicable fields in the JSON-formatted schema.

The general workflow of the MAP is shown in Figure 4. Initially, an intention-aware tenant (T_{iaw}) posts a request to FINALES. From there, an intention-agnostic tenant (T_{iag}) can pick it up and process it according to its own schedule. Hence, the requests are processed asynchronously within the MAP, and different durations of experimental, simulation, or other processes do not block the whole MAP. Once results are available, the T_{iag} posts them back to FINALES, from which the T_{iaw} can withdraw them by posting an appropriate request to FINALES. This request will be served by FINALES itself, and the results matching the request will be sent.

Figure 5 exemplarily shows the structure of a request for a density measurement and a corresponding reply. The two schemas are identical except that the field "pending" must be "true" to identify a request or "false" for reported results, and the field "fom_data" may not be empty for reported results. The field "failed" is used to mark failed measurements or simulations. The "message" field allows us to freely add additional information. In the example, this contains the composition in a volumetric representation and in the molar representation resulting from this. In the example, it is further used to convey a rating of the fidelity of the data by the tenant providing it, e.g., according to the success or the failure of the measurement. In this example, the system reported viscosity and density in the reply, as these two quantities are measured in a single measurement.

DEMONSTRATION OF OPERABILITY

To demonstrate the operability of the current implementation of our MAP, we deployed a minimal configuration for the task of optimizing the composition of a battery electrolyte by maximizing viscosity while minimizing density. The primary aim of the demonstration was to show the basic operability of the MAP design in an actual physical manifestation rather than drawing any scientific conclusions regarding battery electrolytes from the generated data.

The setup included the ASAB system and the simulation orchestrator created using Pipeline Pilot for performing the experiments and simulations, respectively. The optimizer tenant was used to guide the optimization procedure and send requests to FINALES based on the experimental and simulated density, viscosity and ionic conductivity data available in the database and retrieved through FINALES. The

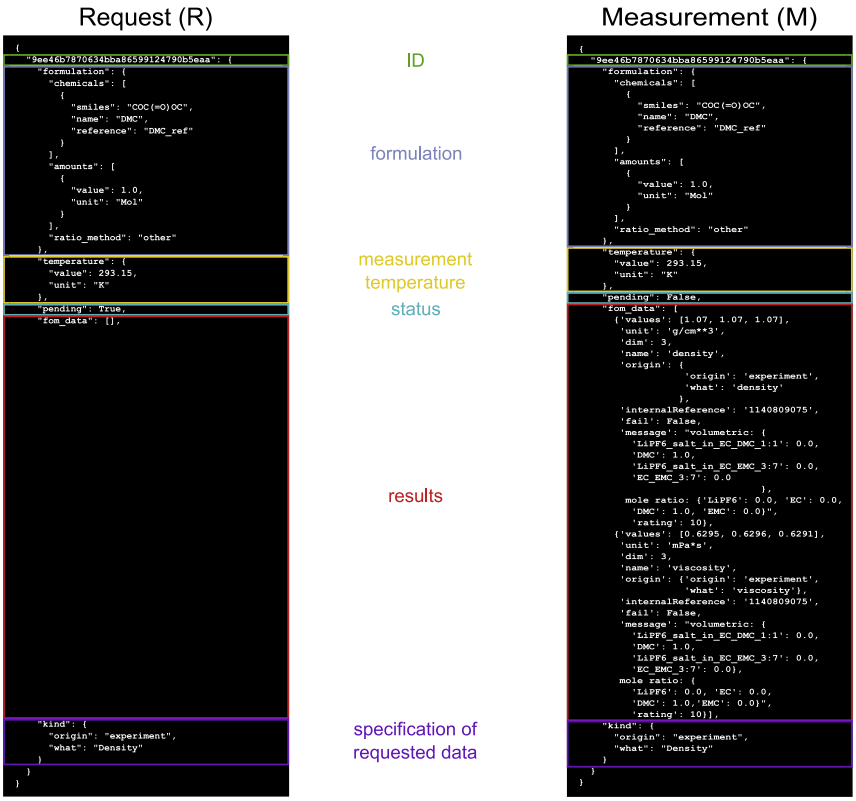

Figure 5. The format of a request and a measurement

The schemas for requests (R) and measurements (M) is the same. For measurements, the field "fom_data" must not be empty, while it may be empty for requests. The main difference is the field "pending," which is "true" for the request and "false" for the measurement.

distributed MAP was tested incrementally in several trials during development. During periods in which the optimizer was not running, requests were generated manually, demonstrating FINALES's ability for human-in-the-loop operation. Most recently, an autonomous run of the MAP, in which FINALES orchestrated the optimizer tenant, the Pipeline Pilot-based simulation orchestrator, and ASAB, ran successfully for approximately 4.5 h without any human intervention and performed several iterations of requests, simulations, and measurements. The current version of the optimizer did not distinguish whether the data originated from experiments or from simulations and treated all values provided by both methods the same. The data presented in the following section were retrieved from the database after the end of the demonstration. It therefore contains the results requested by manual input as well as the results obtained during the autonomous period of the run.

In the experimental setup, the electrolyte samples were formulated based on stock solutions provided in vials prior to the start of the experiment. For the purpose of the demonstration reported here, commercial solutions of 1 M $LiPF_6$ in EC:EMC 3:7 by weight and 1 M $LiPF_6$ in EC:dimethyl carbonate (DMC) 1:1 by weight, as well as EC:EMC 3:7 by weight and DMC, all ordered from E-Lyte Innovations, were used as received and provided as stock solutions to the experimental setup. In preparation of the demonstration run, density measurements for each stock solution needed to be performed, as these values were required as inputs to calculate the volume fractions from the molar fractions. Each measurement originating from the experimental setup comprised three individual measurements. The value for each

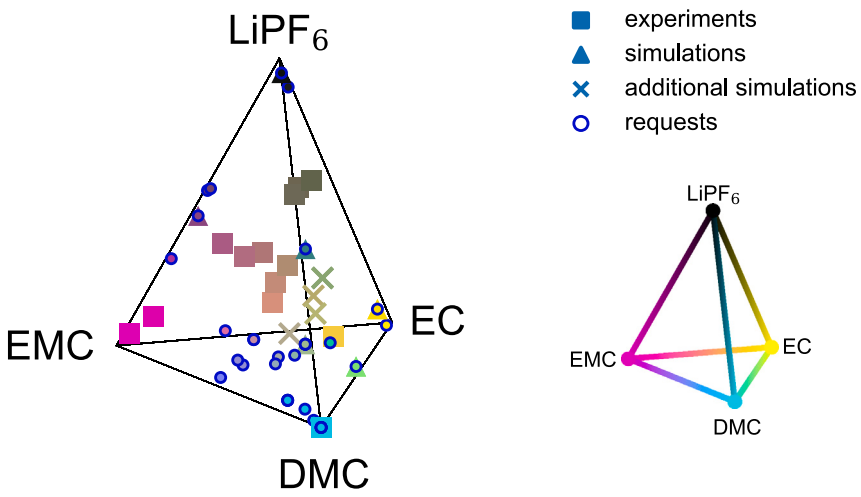

Figure 6. The compositional space covered during the demonstration run

Squared markers represent calculated target formulations of the experiments, triangles mark formulations covered by simulations, and the circles with blue outlines represent the requested formulations. Markers in an x shape denote additional simulations performed during test runs and saved to an instance of FINALES not connected to the optimizer. The set composition in molar fractions is represented by the color. This color scheme is also used in the following figures. A severe deviation between the requested and experimentally targeted formulations is observed. This is likely to be caused by the limitations due to the stock solutions and the approximation of the requested molar fractions by volume fractions.

individual measurement was reported to FINALES within one measurement object. Hence, it is left to the user of the data whether the data shall be averaged or used as is.

RESULTS

The chemical space covered during the entire run is shown in Figure 6. The squared markers represent experimentally measured data, and triangles indicate results obtained from simulations. The round markers with blue outlines represent requested formulations. Simulations for formulations marked by x were added to the server after the demonstration run, as they were saved during testing to a second instance of the server not connected to the optimizer. A significant deviation between the molecular formulations requested by the optimizer and the formulations resulting from the calculated volume fractions targeted in the experiments can be observed. This most probably depends on the stock solutions, which are not allowed to cover the full chemical space spanned by the individual chemicals. Furthermore, the approximation required during the transition from molar fractions to volume fractions and back during the experiment adds another contribution to the deviation if no exact transformation is possible.

Figure 7 shows a plot of the ionic conductivity obtained from simulation results versus the concentration of $LiPF_6$. The values shown originate from simulations requested manually as well as optimizer-requested simulations, and it is also important to note that the $LiPF_6$ content is not the only compositional variation between the data points. The formulations corresponding to the conductivity data presented in this figure also differ in the composition of the solvent. However, the strongest correlation in our data seems to be observed for the $LiPF_6$ content, which is in line with the literature reporting a stronger influence on conductivity by the salt concentration than the composition of the solvent in various electrolyte systems.[47,51,52] Figure 7

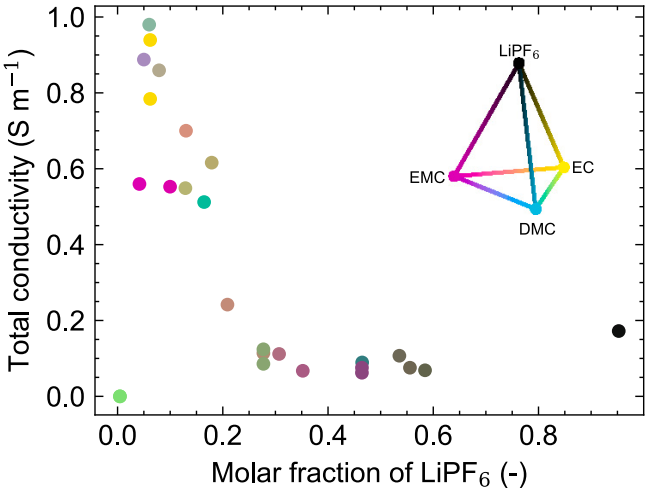

Figure 7. The ionic conductivity data calculated in the course of the MD workflow

The color scheme of Figure 6 was used to denote the composition. The graph comprises data points, which were requested for simulation during the demonstration run and, as such, were generated for the formulations requested for experimental measurement during the demonstration. The simulations for the latter were run after the demonstration was finished. It must be noted that the data points correspond to formulations differing not only in their salt content but also in the composition of the solvent. Further, data originating from simulations triggered by manually posted requests as well as optimizer-generated requests are presented. It seems that a maximum in conductivity for a certain $LiPF_6$ concentration is found, assuming no significant influence of the solvent composition on conductivity.

shows an increase of ionic conductivity for small concentrations of $LiPF_6$ up to a maximum followed by a decrease. The presence of a maximum with respect to the salt concentration is reported in the literature based on experimental results[47,51,53,54] and calculations[40] for similar electrolyte systems. These correlations to the literature suggest that the simulations triggered during this demonstration captured some general trends regarding the ion conductivity in the electrolyte system that has been investigated.

The RDF presented in Figure 8 shows the interatomic distances of lithium ions to relevant atoms within molecules contained in the electrolyte mixture. The peaks at 2 and 4 Å in the Li-O RDF denote oxygen atoms in the carbonate groups of the electrolyte, with corresponding peaks in the Li-C function. Together, these peaks indicate the first solvation shell around the lithium ions. The ordering becomes stronger with increasing EMC content as the relative height of peaks increases compared to the liquid background. In contrast, the RDF between lithium ions shows multiple peaks at very high EMC content. This suggests long-range ordering of the lithium ions, which is likely to be related to the low dielectric constant of pure EMC (ε = 2.5–3)[55] and the short screening length arising from it.

Our experimental results show a strong correlation between the viscosity and the density averaged per measurement in the system under investigation with a Pearson correlation coefficient of approximately 0.91. The correlation can also be seen in Figure 9. Due to this strong correlation, the density and viscosity cannot be truly individually optimized, which renders the minimization of the density and the maximization of the viscosity non-ideal targets for multiobjective optimization in this formulation space.

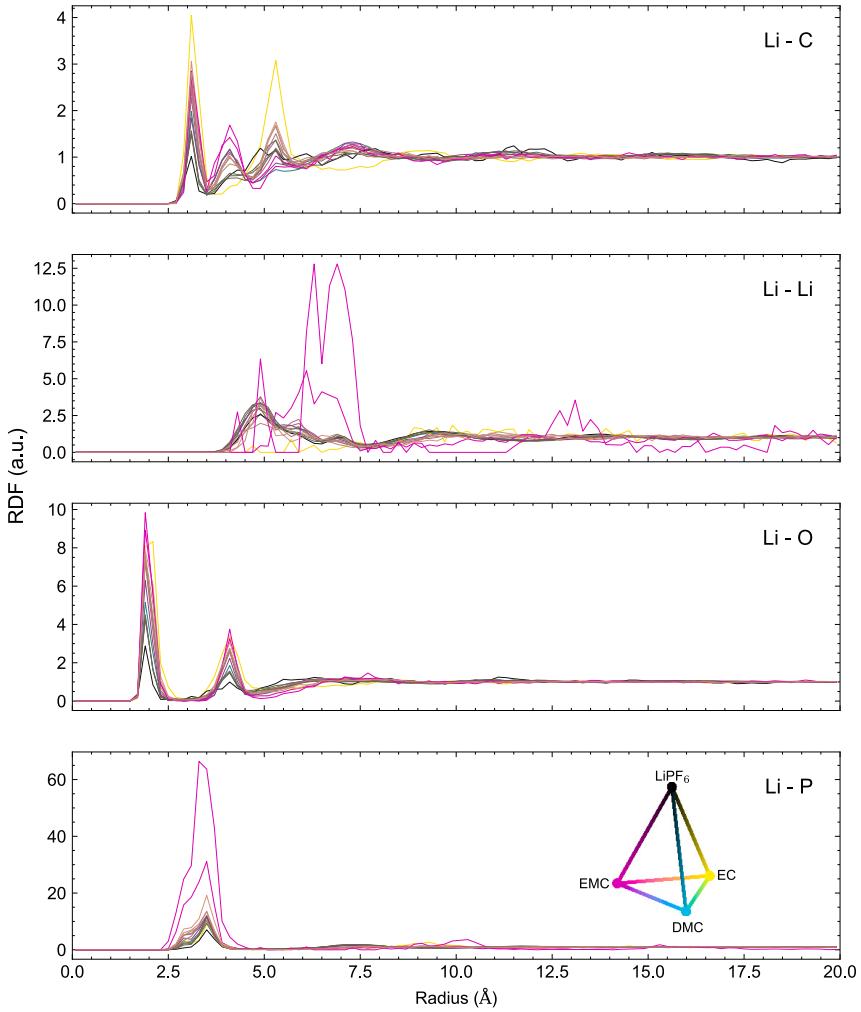

Figure 8. The radial distribution function for lithium ions as obtained from the simulations run on the data generated during the demonstration run based on manual and optimizer-generated requests

The solvation shell of the lithium ions can be observed. The color coding representing the various formulations corresponds to that used in Figure 6.

LESSONS LEARNED

The first run of our international MAP with a passive FINALES brokering instance generated some valuable lessons, which we believe are of great interest for the community and of impact for MAP design.

Data structures

A general insight from this deployment is the realization that the actual implementation of a truly autonomous platform is critically dependent on binding, well-documented, and interoperable data structures. Our MAP requires unambiguous communication. In order to be applicable to such MAPs, data management plans (DMPs)[15] need very strict structures and definitions. During the development of our MAP, a lot of communication between the scientists was required to agree on how to correctly and usefully populate the fields in the schemas due to the various needs of different tenants. Schemas developed for future MAPs need clear

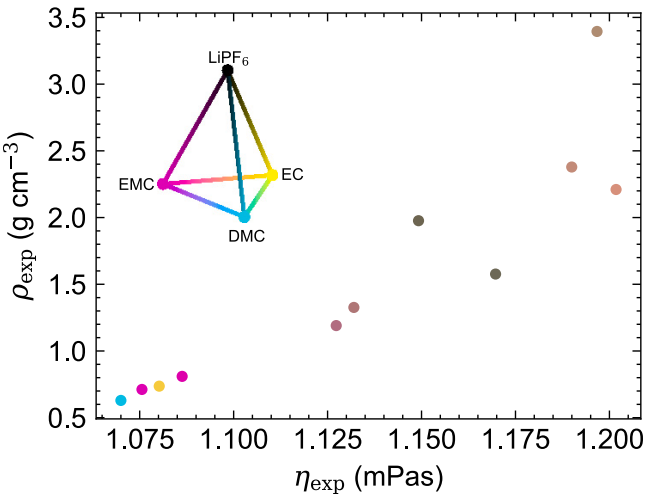

Figure 9. The correlation between the density ρ_{exp} and the viscosity η_{exp} as observed from the formulations yielding both successful density and viscosity measurements

The formulations are colored according to the coloring scheme introduced in Figure 6. A correlation between ρ_{exp} and η_{exp} can be seen. The Pearson correlation coefficient amounts to approximately 0.91.

definitions of the type and format of data contained in each field, which go well beyond the FAIR criteria.[56] A strong mapping from the fields in the schemas to an ontology can help to precisely describe the information conveyed with each field. Further, an ontology can provide a common semantic for the MAP, enabling specific requests and replies.

Data quality

Due to the multimodality of our map, it turned out to be important to rate the quality of data reported to FINALES on the level of the tenant providing them. If, e.g., a measurement with several subsamples is requested, it is valuable information how many data points were valid and can be included in the analysis. This information enables, e.g., an optimizer to decide whether data with a certain rating should be included in the training of the machine-learning model or not. This concept also applies to other tenants, which could rate the fidelity based on different measures. Hence, there is a need for a tangible and measurable data quality rating for all possible data types and tenants.

Sharing limitations, costs, and units

The demonstration revealed the significance of the limitations of individual tenants on the effective operation of our MAP. In the experimental setup, the range of feasible formulations is mainly limited by the stock solutions provided to the system. Requests by the optimizer can exceed these limitations because the current implementation of the optimizer defines formulations on a molecular level without consideration of the stock solutions. In several cases, this results in rough approximations of the formulation when determining the volume fractions of the stock solutions. Analogously, limitations in the simulation tenant arise due to the limited number of molecules, which can be simulated with reasonable use of computational resources. In the case of our demonstration, a molar fraction of the salt exceeding 0.5 is the limit due to the size of the unit cell. These examples emphasize the necessity of dealing with the limitations of each tenant. We suggest sharing the limitations upon connection of a tenant to the MAP so that other tenants can adjust their requests

accordingly. For costs arising for the use of tenants, this concept of sharing can be extended to include the price, which can be used to enforce a budget on optimizations. This does not mean restricting the search space to the limits of the most restrictive tenant but rather installing a preventive measure for cost-excessive or impossible-to-fulfill requests. The implementation of a billing system could further add a new direction to research by offering laboratories as a service like existing biotechnology services.

The example of the formulations also shows the importance of finding a common way of specifying crucial quantities shared by several tenants and transforming them according to the specific needs of a tenant, where necessary. Such common representations and units need to be defined for the whole MAP during setup.

MAP design

In our demonstration, asynchronicity and the append-only way of saving requests and results proved to be valuable features in orchestrating our MAP. As each tenant processes its tasks independently from the MAP, no tenant is being blocked by the state of another. This fulfills the requirement of a modular and hardware-independent operation. The intention-agnostic operation of some tenants also proved beneficial regarding the flexible selection of the optimization targets. The T_{iaw} is able to access all the data available in the database of the MAP and to exploit the full dataset while pursuing its objective.

For future development, the implementation of full data lineage tracking including timestamps and the complete collection of metadata for the exploitation of additional results will be targeted. Full data lineage tracking requires each tenant to provide the data lineage tracking for its workflow and to report it to the MAP or at least to save it in tenant-specific storage for future reference. Overall, we recommend not restricting data recording solely to the quantities of primary interest but to record all data and metadata possible.

Our proposed MAP architecture, including the centralized passive FINALES brokering service, fulfills the requirements of interchangeability of tenants, asynchronous and distributed operation, and inclusion of experimental and computational setups. This constitutes a significant portion of our definition of a MAP. Hence, we regard this as a possible concept for the future development of international MAPs.

Improvements need to be made to the standardization of the schemas, the user authentication, and the communication of limitations of tenants. Further development and a demonstration of multitenancy and data lineage tracking are part of the future work.

CONCLUSION

The demonstration reported here showed an operation principle for a distributed MAP on an international scale. We believe that this orchestration paradigm of implementing a marketplace of requests to be fulfilled by available tenants could be a universal approach to the acceleration of research. This first run of our internationally distributed MAP allowed us to identify a variety of challenges, which need to be addressed in the further development and integration of an increasing number of tenants. We believe that efforts toward the further development and deployment of MAPs are worth undertaking, as building shared MAPs, where instruments are no

longer used solely by a single group, allows the simultaneous usage of research equipment for a variety of research questions and fields. This is important for ensuring fault tolerance and around-the-clock operation and, simultaneously, creates incentives for cost reduction and efficiency enhancement. By the collaborative nature of MAPs, community building is an integral part of the design process. The current funding schemes in, e.g., the European Union (EU) and associated national or bilateral funding agencies in fact foster this collaborative approach to research, and we believe that there is a need to create more cross-country funding lines that can help in building global research workflows. We believe that MAPs can improve the efficient utilization of equipment in new innovative workflows with little additional effort to ultimately enable a more efficient allocation of funding. We hope that our insights shared in this perspective will support the further development of international MAPs and the acceleration of research to overcome the material limited challenges of our time.

EXPERIMENTAL PROCEDURES

Resource availability

Please refer to the "Methods and tenants" section to obtain information about the experimental procedures and how to reproduce them.

Lead contact

Requests for further information should be directed to the lead contact, Helge S. Stein (helge.stein@tum.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data reported in this perspective will be shared by the lead contact upon request. All original code regarding the FINALES broker server has been deposited at https://github.com/BIG-MAP/finale under https://zenodo.org/record/8009625 and is publicly available under the MIT license as of the date of publication. Any additional information required to reanalyze the data reported in this perspective is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work contributes to the research performed at CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was funded by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). This project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 957189. The authors acknowledge BATTERY 2030PLUS, funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 957213. M.V. acknowledges the KIT Graduate School Enabling Net Zero - ENZo.

AUTHOR CONTRIBUTIONS

M.V. built and designed ASAB and organized the FINALES run; J.B. and P.B.J. programmed and ran the optimizer tenant; H.H., N.S., J.C., and F.H. programmed and ran the Pipeline Pilot tenant; F.H. organized and J.B., H.H., and M.V. implemented regression tests; I.E.C. and S.C. developed the DMP and ontology link; F.F.R. and G.P. developed the AiiDA tenant; A.B. and H.S.S. wrote the first draft of the

manuscript; H.S.S. developed the FINALES software, hosted and maintained the FINALES broker server, and conceived the initial idea and MAP design.

DECLARATION OF INTERESTS

H.H., N.S., J.C., and F.H. may be beneficiaries of shares in Dassault Systèmes through an employer scheme.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Maier, W.F. (2019). Early Years of High-Throughput Experimentation and Combinatorial Approaches in Catalysis and Materials Science. ACS Comb. Sci. *21*, 437–444. https://doi.org/10.1021/acscombsci.8b00189.

2. Stein, H.S., Gregoire, J.M., and Gregoire, M. (2019). Progress and prospects for accelerating materials science with automated and autonomous workflows. Chem. Sci. *10*, 9640–9649. https://doi.org/10.1039/C9SC03766G.

3. Soedarmadji, E., Stein, H.S., Suram, S.K., Guevarra, D., and Gregoire, J.M. (2019). Tracking materials science data lineage to manage millions of materials experiments and analyses. npj Comput. Mater. *5*, 79. https://doi.org/10.1038/s41524-019-0216-x.

4. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N., and Kozinsky, B. (2016). AiiDA: Automated Interactive Infrastructure and Database for Computational Science. Comput. Mater. Sci. *111*, 218–230. https://doi.org/10.1016/j.commatsci.2015.09.013.

5. Amis, E.J., Xiang, X.-D., and Zhao, J.-C. (2002). Combinatorial Materials Science: What's New Since Edison? MRS Bull. *27*, 295–300. https://doi.org/10.1557/mrs2002.96.

6. Koinuma, H., and Takeuchi, I. (2004). Combinatorial solid-state chemistry of inorganic materials. Nat. Mater. *3*, 429–438. https://doi.org/10.1038/nmat1157.

7. Ludwig, A., Zarnetta, R., Hamann, S., Savan, A., and Thienhaus, S. (2008). Development of multifunctional thin films using high-throughput experimentation methods. J. Mater. Chem. *99*, 1144–1149. https://doi.org/10.3139/146.101746.

8. Ludwig, A. (2019). Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. npj Comput. Mater. *5*, 70–77. https://doi.org/10.1038/s41524-019-0205-0.

9. Bhowmik, A., Castelli, I.E., Garcia-Lastra, J.M., Jørgensen, P.B., Winther, O., and Vegge, T. (2019). A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning. Energy Storage Mater. *21*, 446–456. https://doi.org/10.1016/j.ensm.2019.06.011.

10. Sanchez-Lengeling, B., and Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. Science *361*, 360–365.

11. Noh, J., Kim, J., Stein, H.S., Sanchez-Lengeling, B., Gregoire, J.M., Aspuru-Guzik, A., and Jung, Y. (2019). Inverse Design of Solid-State Materials via a Continuous Representation. Matter *1*, 1370–1384. https://doi.org/10.1016/j.matt.2019.08.017.

12. Isayev, O., Oses, C., Toher, C., Gossett, E., Curtarolo, S., and Tropsha, A. (2017). Universal fragment descriptors for predicting properties of inorganic crystals. Nat. Commun. *8*, 15679–15712. https://doi.org/10.1038/ncomms15679.

13. Peterson, A.A., and Nørskov, J.K. (2012). Activity Descriptors for CO_2 Electroreduction to Methane on Transition-Metal Catalysts. J. Phys. Chem. Lett. *3*, 251–258. https://doi.org/10.1021/jz201461p.

14. Curtarolo, S., Hart, G.L.W., Nardelli, M.B., Mingo, N., Sanvito, S., and Levy, O. (2013). The high-throughput highway to computational materials design. Nat. Mater. *12*, 191–201. https://doi.org/10.1038/nmat3568.

15. Castelli, I.E., Arismendi-Arrieta, D.J., Bhowmik, A., Cekic-Laskovic, I., Clark, S., Dominko, R., Flores, E., Flowers, J., Ulvskov Frederiksen, K., Friis, J., et al. (2021). Data Management Plans: the Importance of Data Management in the BIG-MAP Project. Batter. Supercaps *4*, 1803–1812. https://doi.org/10.1002/batt.202100117.

16. Ling, J., Hutchinson, M., Antono, E., Paradiso, S., and Meredig, B. (2017). High-Dimensional Materials and Process Optimization using Data-driven Experimental Design with Well-Calibrated Uncertainty Estimates. Integr. Mater. Manuf. Innov. *6*, 207–217. https://doi.org/10.1007/s40192-017-0098-z.

17. Ament, S.E., Stein, H.S., Guevarra, D., Zhou, L., Haber, J.A., Boyd, D.A., Umehara, M., Gregoire, J.M., and Gomes, C.P. (2019). Multi-component background learning automates signal detection for spectroscopic data. npj Comput. Mater. *5*, 1–7. https://doi.org/10.1038/s41524-019-0213-0.

18. Schwarting, M., Siol, S., Talley, K., Zakutayev, A., and Phillips, C. (2017). Automated algorithms for band gap analysis from optical absorption spectra. Mater. Discov. *10*, 43–52. https://doi.org/10.1016/j.md.2018.04.003.

19. Umehara, M., Stein, H.S., Guevarra, D., Newhouse, P.F., Boyd, D.A., and Gregoire, J.M. (2019). Analyzing machine learning models to accelerate generation of fundamental materials insights. npj Comput. Mater. *5*, 34. https://doi.org/10.1038/s41524-019-0172-5.

20. Mission Innovation (2018). Materials Acceleration Platform—Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods with Artificial Intelligence (Mission Innovation).

21. Vescovi, R., Chard, R., Saint, N.D., Blaiszik, B., Pruyne, J., Bicer, T., Lavens, A., Liu, Z., Papka, M.E., Narayanan, S., et al. (2022). Linking scientific instruments and computation: Patterns, technologies, and experiences. Patterns *3*, 100606. https://doi.org/10.1016/j.patter.2022.100606.

22. Chard, R., Pruyne, J., McKee, K., Bryan, J., Raumann, B., Ananthakrishnan, R., Chard, K., and Foster, I.T. (2023). Globus automation services: Research process automation across the space–time continuum. Future Generat. Comput. Syst. *142*, 393–409. https://doi.org/10.1016/j.future.2023.01.010.

23. Roch, L.M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., Yunker, L.P.E., Hein, J.E., and Aspuru-Guzik, A. (2018). ChemOS: Orchestrating autonomous experimentation. Sci. Robot. *3*, eaat5559. https://doi.org/10.1126/scirobotics.aat5559.

24. Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al. (2020). A mobile robotic chemist. Nature *583*, 237–241. https://doi.org/10.1038/s41586-020-2442-2.

25. Rahmanian, F., Flowers, J., Guevarra, D., Richter, M., Fichtner, M., Donnely, P., Gregoire, J.M., and Stein, H.S. (2022). Enabling Modular Autonomous Feedback-Loops in Materials Science through Hierarchical Experimental Laboratory Automation and Orchestration. Adv. Mater. Interfac. *9*, 2101987. https://doi.org/10.1002/admi.202101987.

26. Antypas, K., Canon, S., Dart, E., Fagnan, K., Gerhardt, L., Jacobsen, D., Lockwood, G.K., Monga, I., Nugent, P., Ramakrishnan, L., et al. (2020). Superfacility: The Convergence of Data, Compute, Networking, Analytics and Software. In Handbook on Big Data and Machine Learning in the Physical Sciences (World Scientific Series on Emerging Technologies), pp. 361–386. https://doi.org/10.1142/9789811204579_0017.

27. Amici, J., Asinari, P., Ayerbe, E., Barboux, P., Bayle-Guillemaud, P., Behm, R.J., Berecibar, M., Berg, E., Bhowmik, A., Bodoardo, S., et al. (2022). A Roadmap for Transforming Research to Invent the Batteries of the Future Designed

within the European Large Scale Research Initiative BATTERY 2030+. Adv. Energy Mater. *12*, 2102785. https://doi.org/10.1002/aenm.202102785.

28. Fichtner, M. (2022). Recent Research and Progress in Batteries for Electric Vehicles. Batter. Supercaps *5*, e202100224. https://doi.org/10.1002/batt.202100224.

29. Fichtner, M., Edström, K., Ayerbe, E., Berecibar, M., Bhowmik, A., Castelli, I.E., Clark, S., Dominko, R., Erakca, M., Franco, A.A., et al. (2021). Rechargeable Batteries of the Future—The State of the Art from a BATTERY 2030+ Perspective. Adv. Energy Mater. *12*, 2102904. https://doi.org/10.1002/aenm.202102904.

30. Bhowmik, A., Berecibar, M., Casas-Cabanas, M., Csanyi, G., Dominko, R., Hermansson, K., Palacin, M.R., Stein, H.S., and Vegge, T. (2022). Implications of the BATTERY 2030+ AI-Assisted Toolkit on Future Low-TRL Battery Discoveries and Chemistries. Adv. Energy Mater. *12*, 2102698. https://doi.org/10.1002/aenm.202102698.

31. Nikolaev, P., Hooper, D., Perea-López, N., Terrones, M., and Maruyama, B. (2014). Discovery of Wall-Selective Carbon Nanotube Growth Conditions *via* Automated Experimentation. ACS Nano *8*, 10214–10222. https://doi.org/10.1021/nn503347a.

32. Kusne, A.G., Gao, T., Mehta, A., Ke, L., Nguyen, M.C., Ho, K.-M., Antropov, V., Wang, C.-Z., Kramer, M.J., Long, C., and Takeuchi, I. (2014). On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. Sci. Rep. *4*, 6367. https://doi.org/10.1038/srep06367.

33. Li, Y.J., Savan, A., Kostka, A., Stein, H.S., Ludwig, A., and Ludwig, A. (2018). Accelerated atomic-scale exploration of phase evolution in compositionally complex materials. Mater. Horiz. *5*, 86–92. https://doi.org/10.1039/C7MH00486A.

34. Stein, H.S., Sanin, A., Rahmanian, F., Zhang, B., Vogler, M., Flowers, J.K., Fischer, L., Fuchs, S., Choudhary, N., and Schroeder, L. (2022). From materials discovery to system optimization by integrating combinatorial electrochemistry and data science. Curr. Opin. Electrochem. *35*, 101053. https://doi.org/10.1016/j.coelec.2022.101053.

35. Häse, F., Roch, L.M., and Aspuru-Guzik, A. (2018). Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. Chem. Sci. *9*, 7642–7655. https://doi.org/10.1039/C8SC02239A.

36. Colvin, S. (2022). Contributors to pydantic (pydantic). https://github.com/pydantic/pydantic.

37. Clark, S., Bleken, F.L., Stier, S., Flores, E., Andersen, C.W., Marcinek, M., Szczesna-Chrzan, A., Gaberscek, M., Palacin, M.R., Uhrin, M., and Friis, J. (2022). Toward a Unified Description of Battery Data. Adv. Energy Mater. *12*, 2102702. https://doi.org/10.1002/aenm.202102702.

38. Allan, D., Caswell, T., Campbell, S., and Rakitin, M. (2019). Bluesky's Ahead: A Multi-Facility Collaboration for an *a la Carte* Software Project for Data Acquisition and Management. Synchrotron Radiat. News *32*, 19–22. https://doi.org/10.1080/08940886.2019.1608121.

39. Dassault Systèmes Americas Corporation (2022). BIOVIA Pipeline Pilot release 2022. https://www.3ds.com/products-services/biovia/products/data-science/pipeline-pilot/.

40. Hanke, F., Modrow, N., Akkermans, R.L.C., Korotkin, I., Mocanu, F.C., Neufeld, V.A., and Veit, M. (2019). Multi-Scale Electrolyte Transport Simulations for Lithium Ion Batteries. J. Electrochem. Soc. *167*, 013522. https://doi.org/10.1149/2.0222001JES.

41. Schaarschmidt, J., Yuan, J., Strunk, T., Kondov, I., Huber, S.P., Pizzi, G., Kahle, L., Bölle, F.T., Castelli, I.E., Vegge, T., et al. (2022). Workflow Engineering in Materials Design within the BATTERY 2030+ Project. Adv. Energy Mater. *12*, 2102638. https://doi.org/10.1002/aenm.202102638.

42. Ramírez, S.. Contributors to FastAPI FastAPI. fastapi. https://fastapi.tiangolo.com/.

43. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

44. Srinivas, N., Krause, A., Kakade, S., and Seeger, M. (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In Proceedings of the 27th International Conference on Machine Learning ICML, *10*Proceedings of the 27th International Conference on Machine Learning ICML (Omnipress), pp. 1015–1022.

45. Huber, S.P., Zoupanos, S., Uhrin, M., Talirz, L., Kahle, L., Häuselmann, R., Gresch, D., Müller, T., Yakutovich, A.V., Andersen, C.W., et al. (2020). AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. Sci. Data *7*, 300. https://doi.org/10.1038/s41597-020-00638-4.

46. Uhrin, M., Huber, S.P., Yu, J., Marzari, N., and Pizzi, G. (2021). Workflows in AiiDA: Engineering a high-throughput, event-based engine for robust and modular computational workflows. Comput. Mater. Sci. *187*, 110086. https://doi.org/10.1016/j.commatsci.2020.110086.

47. Rahmanian, F., Vogler, M., Wölke, C., Yan, P., Winter, M., Cekic-Laskovic, I., and Stein, H.S. (2022). One-Shot Active Learning for Globally Optimal Battery Electrolyte Conductivity. Batter. Supercaps *5*, e202200228. https://doi.org/10.1002/batt.202200228.

48. Flores-Leonar, M.M., Mejía-Mendoza, L.M., Aguilar-Granda, A., Sanchez-Lengeling, B., Tribukait, H., Amador-Bedolla, C., and Aspuru-Guzik, A. (2020). Materials Acceleration Platforms: On the way to autonomous experimentation. Curr. Opin. Green Sustain. Chem. *25*, 100370. https://doi.org/10.1016/j.cogsc.2020.100370.

49. Saikin, S.K., Kreisbeck, C., Sheberla, D., Becker, J.S., and Aspuru-Guzik, A. (2019). Closed-loop discovery platform integration is needed for artificial intelligence to make an impact in drug discovery. Expet Opin. Drug Discov. *14*, 1–4. https://doi.org/10.1080/17460441.2019.1546690.

50. Häse, F., Roch, L.M., and Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories. Trends Chem. *1*, 282–291. https://doi.org/10.1016/j.trechm.2019.02.007.

51. Ding, M.S., Xu, K., Zhang, S.S., Amine, K., Henriksen, G.L., and Jow, T.R. (2001). Change of Conductivity with Salt Content, Solvent Composition, and Temperature for Electrolytes of LiPF6 in Ethylene Carbonate-Ethyl Methyl Carbonate. J. Electrochem. Soc. *148*, A1196. https://doi.org/10.1149/1.1403730.

52. Chen, H.P., Fergus, J.W., and Jang, B.Z. (2000). The Effect of Ethylene Carbonate and Salt Concentration on the Conductivity of Propylene Carbonate|Lithium Perchlorate Electrolytes. J. Electrochem. Soc. *147*, 399. https://doi.org/10.1149/1.1393209.

53. Nyman, A., Behm, M., and Lindbergh, G. (2008). Electrochemical characterisation and modelling of the mass transport phenomena in LiPF6–EC–EMC electrolyte. Electrochim. Acta *53*, 6356–6365. https://doi.org/10.1016/j.electacta.2008.04.023.

54. Kondo, K., Sano, M., Hiwara, A., Omi, T., Fujita, M., Kuwae, A., Iida, M., Mogi, K., and Yokoyama, H. (2000). Conductivity and Solvation of Li^+ Ions of $LiPF_6$ in Propylene Carbonate Solutions. J. Phys. Chem. B *104*, 5040–5044. https://doi.org/10.1021/jp000142f.

55. Hall, D.S., Self, J., and Dahn, J.R. (2015). Dielectric Constants for Quantum Chemistry and Li-Ion Batteries: Solvent Blends of Ethylene Carbonate and Ethyl Methyl Carbonate. J. Phys. Chem. C *119*, 22322–22330. https://doi.org/10.1021/acs.jpcc.5b06022.

56. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018. https://doi.org/10.1038/sdata.2016.18.