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 Castelli[3], Francisco Fernando Ramírez[5,6], Johan Carlsson[4], Giovanni Pizzi[5,6,9], Simon Clark[7], Felix Hanke[8]*, Arghya Bhowmik[3]*, Helge S. Stein[1,2]*

[1]: Helmholtz Institute Ulm, Helmholtzstr. 11, 89073 Ulm, Germany
[2]: Institute for Physical Chemistry, Fritz-Haber Weg 2, Karlsruhe, Germany
[3]: Department of Energy Conversion and Storage, Technical University of Denmark (DTU), Anker Engelunds Vej, Kgs. Lyngby DK-2800, Denmark
[4]: Dassault Systèmes, Cologne, A-3, Germany
[5]: Theory and Simulation of Materials (THEOS), École Polytechnique Fédérale de Lausanne, CH-015 Lausanne, Switzerland
[6]: National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[7]: SINTEF Industry, Battery and Hydrogen Technologies, Sem Sælands Vei 12, 7034 Trondheim, Norway
[8]: Dassault Systèmes, 334 Science Park, Cambridge, CB4 0WN, United Kingdom
[9]: Laboratory for Materials Simulations (LMS), Paul Scherrer Institut (PSI), CH-5232 Villigen PSI, Switzerland

Correspondence should be addressed to: arbh@dtu.dk, felix.hanke@3ds.com, helge.stein@kit.edu

Tags: Cooperative research, fault-tolerant, autonomous experiments, multilocation, multimodal, materials acceleration platform

Abstract

The future of materials science is borderless, cooperative, and distributed across the globe. This necessitates flexible, reconfigurable software defined research workflows, which we herein demonstrate by integrating multiple disciplines and modalities. Our brokering approach to research orchestration exposes entire laboratories in a cooperative multi-tenancy platform that is asynchronous, modular, and resilient. To the best of our knowledge, this constitutes the first international materials acceleration platform (MAP) which is herein demonstrated through a battery electrolyte workflow that ran in five countries over two weeks. We discuss the lessons learned from multi-tenancy and fault tolerance and chart a way to a universal battery MAP with fully ontology-based schemas and cost-aware orchestration.

Introduction

Scientists have continuously innovated on the efficiency improvements of the research process over the last many decades[1]. Automation of research tasks[2] and their integration with data lineage tracking[3] was one of the driving accelerators in the past 20 years in biotechnology and materials science[4]. This research automation created the field of combinatorial materials science (CMS)[5] that is utilizing the paradigm of well-defined composition and processing variation to unravel the

underlying physicochemical relationships at a greater pace[6,7]. The near endless breath of the chemical space does, however, render any brute force exploration for discovery futile. Effective research therefore necessitates approaches that can predict how to design materials according to a target functional property. This inverse design[8–10] approach is rooted in the idea of rational design based descriptors[11,12] that define composition-structure-property relationships. Recent advancements in research instrumentation[13] and user interfaces have lowered the entrance barrier for high-throughput experimentation (HTE). Abundance of searchable data, enabled by data management[3,14], then sparked the proliferation of data-driven methods in synthesis[15], characterization[16], performance evaluation[17], and interpretation[18]. 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 called the "materials acceleration platform" (MAP)[19].

To date, there have been early demonstrations of the MAP[2] idea, but all have been constrained to a laboratory bench[20] or building[21]. Going beyond a single property or even transcending the laboratory is necessitated for a true battery map[22] as there are multiple functional properties to optimize[23,24], 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 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, nanoseconds to years and therefore cannot be mastered on a laboratory bench, a single institution, or domain alone. A complicating factor in all of this is the mutual exclusion of certain methods, i.e. density-functional theory (DFT) calculations cannot be run on the month scale[8,25] and manufactured batteries cannot easily be analyzed by e.g. XPS. This is in stark contrast to all of the early MAP-inspired demonstrations from other fields[26–28], where the entire research lifecycle 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 battery research value chain[29].

Herein we discuss how we envision the design of a cooperative and resilient MAP that is spatially distributed across multiple countries and involves scientists from the experimental and modeling domain. Based on these MAP design considerations we deploy a demonstration to a team of scientists involved in the BIG-MAP project[14], where a brokering software system was 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, United Kingdom), the machine learning optimizer was developed and run at the Technical University of Denmark (DTU, Denmark), data and ontology interfaces were managed at DTU, Stiftelsen for industriell og teknisk forskning (SINTEF, Norway) and École polytechnique fédérale de Lausanne (EPFL, Switzerland). This workflow is capable of determining the density, viscosity, and pair distribution functions and a machine learning based prediction of conductivity of aprotic electrolytes for Li-ion batteries. To the best of our knowledge, this is the first demonstration of an international MAP. After intensive design discussion, prior to the actual run, the real-world deployment of a MAP involving many partners was a valuable learning experience which we discuss at the end of this manuscript. We believe that sharing not just the success but also the failures of this new paradigm of performing research will help the community in avoiding some of the errors we made and accelerate the progress towards true autonomy.

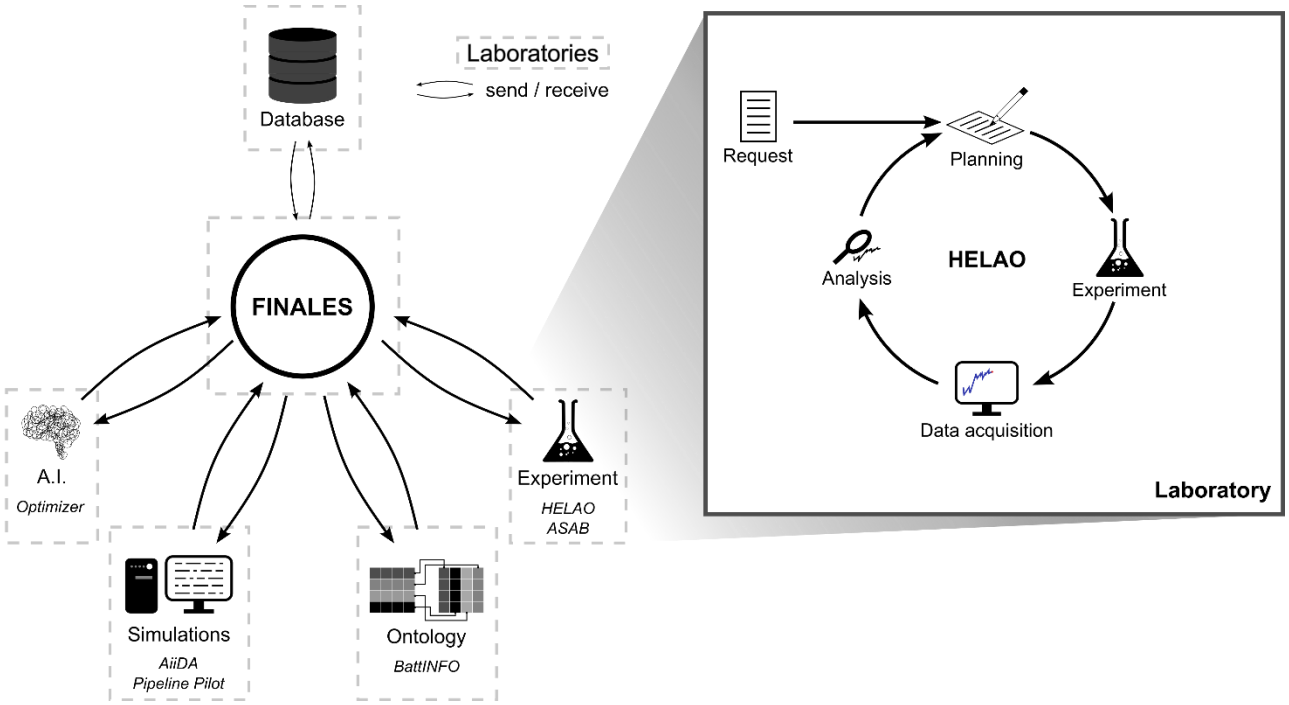

Figure 1: Schematics of conventional accelerated research labs and brokered multitenant materials acceleration platforms. The left part of the figure shows the general layout of the FINALES paradigm to research orchestration in which multiple tenants i.e. optimizers, experiments, simulation and databases exist in a cooperative environment in which each participant can be located anywhere on the globe. The inset presents an emblematic workflow of a HELAO[30] 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.

MAP design considerations

There have been numerous reviews and conceptual papers about the high-level design of a MAP, ranging from concept papers[7,9,14,19,20,31–41], to building a common language to describe the advancement towards autonomous research[2] and the announcement of the Nobel Turing Challenge[42]. The design of our MAP is motivated by a long-term vision of altering the way how we perform materials research[22], i.e. going beyond single-laboratory demonstrations and towards the creation of an ecosystem that enables atom-to-system innovation at global integration[43,44] today.

We believe that there is a fallacy in the current perception of the MAP paradigm. This misconception is in our view, that scientists often think of a single loop[35,37,38] optimizing for a specific target. Battery research, or energy conversion research in a broader sense, is inherently multimodal and multitarget and therefore requires a truly modular MAP, going beyond a single research instrument, group and modality. Consequently, a MAP for energy conversion requires a setup combining hardware and software components that overcome boundaries between the scales through multiple experimental probes, simulators and data driven model building to exploit complementary capacities. In the process, it needs to minimize repetitive human actions while increasing the efficiency of experiments and maximizing knowledge generation in a time- and cost-efficient manner. Funding allocation efficiency requires better utilization of equipment, which we believe to be a defining characteristic of the MAP idea as existing equipment can be deployed in new innovative workflows with no additional effort.

Emblematic for conventional optimization loops, that are not yet a MAP by our definition of a battery MAP, are demonstrations of optimizing fast charging capabilities[45,46] of a specific battery chemistry, prolonging the lifetime of a cell chemistry under fast charging conditions and numerous others[47–51]. In principle these achievements of fast charging and lifetime prolongation could also be achieved by changing the cell chemistry and not the charging procedure[52]. A next generation MAP would, however, need to be capable of optimizing both the chemistry and the protocol, whilst being agile in

the sense of allowing to investigate different manufacturing settings, such as changing the cell geometry[23]. The goal is therefore the establishment of a universal, highly modular approach that is not built for only one purpose.
We therefore introduce a new concept besides modularity in materials research and the MAP-Paradigm which is multi-tenancy, i.e. the same experimental or simulation capacity offered as well as multiple optimizers or operators with different intentions in the same MAP. Tenants in a MAP are all participating units, be it intention-aware tenants like operators and optimizers, or intention-agnostic tenants like orchestrators, experimentalists, theorists or data storage applications. The defining feature here is that there can be multiple copies of every tenant without racing conditions. Building shared MAPs, where instruments no longer belong to a single group, allow the utilization of research instruments for a variety of research questions and fields. This is not just important for ensuring fault tolerance and around-the-clock operation, but it also creates incentives for cost reduction. We also believe that the intention-agnostic sharing of research instrumentation will allow for significantly better utilization of instruments and allocation of institutional funding. By the collaborative nature of MAPs as described herein, community building is an integral part of the design process. The current funding schemes in 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 the need to create more cross-country funding lines that can help in building global research workflows.
With these design considerations in mind, we will describe the technical layout of our MAP, its tenants, results, and lessons learned in the following sections. A defining design decision is to disable direct communication between tenants, and to channel all communication through a brokering service we refer to as the fast intention agnostic learning server, or FINALES in short. FINALES is our take on interoperability and constitutes an open interface brokering between hard-, and software as well as humans. All simulations and experiments can be requested as a service through the FINALES APIs that enables abstraction with standardization in data, communication, and information retrieval through being fully ontology linked. Interoperability to FINALES is moved to a translation layer developed by the respective tenants such that no existing workflow needs to be altered upon addition of a new agent. To the best of our knowledge this is the first multi-tenancy, multi-modal and international demonstration of a MAP.

Demonstration of an international MAP

In the following section we will introduce the different agents compatible with 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 called Pipeline Pilot, the AiiDA interface for data storage and machine learning based conductivity predictions, and the optimizer based on a Gaussian process in combination with the Chimera[53] optimizer.

Tenants

Broker server - FINALES

The Fast INtention Agnostic LEarning Server (FINALES) is designed as a broker server mediating between the simulations, experiments, optimizers, and other tenants. It can handle incoming and outgoing requests asynchronously and acts as a central hub in the communication. In contrast to the laboratory orchestration system HELAO[30], FINALES acts as a brokering service on a hardware independent level by abstracting away from a sequence of hardware events to requesting desired

outcomes. FINALES allows for maximum flexibility and redundancy by explicitly allowing several hardware or software services with the same capabilities as well as a virtually unlimited number of simultaneous optimizers. This multi-tenancy design allows for several intention-aware tenants exploring and exploiting the data and services available through the server. The communication with FINALES is defined through Pydantic[54] schemas. This structured representation of the data facilitates the integration of an ontology such as BattINFO[55]. Furthermore, the type-checking, which Pydantic promotes, improves the security of FINALES. Type-checking and automatic validation of input data even allows to include non-automated data acquisition in autonomously orchestrated loops, maximizing the inclusiveness and versatility of FINALES. What most significantly distinguishes FINALES from HELAO or any other automation framework like Bluesky[56] or ChemOS[20] is FINALES' passivity as its operation does not depend on the state of the individual tenants. By design, HELAO, Bluesky or ChemOS operated instruments could be connected to FINALES as tenants. Data management in FINALES works in an append-only mode, which preserves all generated data. Datasets can be rated by the tenant providing it e.g. according to the success or failure of the measurement through a score related to the fidelity of the data. Deletion of requests is not allowed by design.

Simulation orchestrator - Pipeline Pilot

Pipeline Pilot is a workflow system with different applications in cheminformatics, laboratories, materials modeling, and simulations. Pipeline Pilot enables automation of calculations from the 3DS software family as well as the visualization and reporting of results. Research workflows are orchestrated similar to the way it is done by HELAO with the addition of being able to graphically compose a workflow from hundreds of different configurable components as shown in Figure 2. The workflow queries FINALES through the API and identifies pending requests not yet passed to the simulation cluster. The workflow extracts relevant input data for the molecular dynamics simulations from the simulation request and submits the calculations as jobs onto different high-performance computing (HPC) clusters. Finally, results are formatted for reporting them 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.

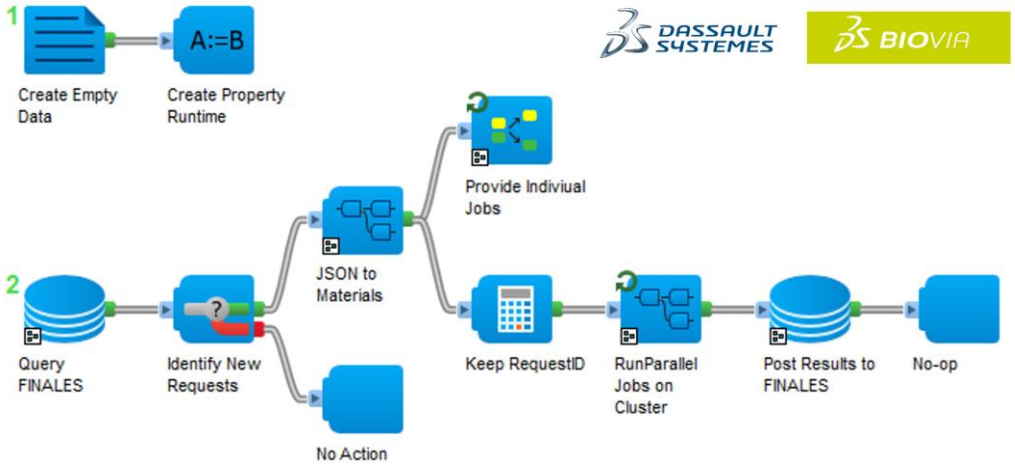

Figure 2: The workflow architecture for automated simulations using Pipeline Pilot in which FINALES is queried for new requests that are not yet in the simulation queue. Results are obtained asynchronously and posted onto FINALES upon convergence.

The core calculations of the workflow are performed in the *RunParallel Jobs on Cluster* component, which contains a complete workflow to calculate the ionic transport coefficients of a given electrolyte

composition[57,58]. Briefly, this calculation proceeds by building an *Amorphous Cell* of the molecular liquid. It then runs an initialization stage NVT molecular dynamics run followed by an NPT run for 250 ns each to thermalize the system and to obtain an estimate for the density. This is followed by a 2000 ns sampling simulation and an automatic analysis of the diffusion coefficients, conductivities, and transfer numbers for all ionic species in the system as described in detail in Hanke et al.[57]. In addition, the radial density function is computed, which provides a detailed picture of the molecular structure of the liquid.

Experimental setup - ASAB

The Autonomous Synthesis and Analysis of Battery electrolytes (ASAB) system used to perform the experiments described in this study is shown in Figure 3. It comprises a base module for power supply, six syringe pumps and ten eleven-position valve modules including a matching API commercially available from the company CETONI GmbH. 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 GmbH are connected to the system. A PC is used to control the system via an in-house implemented Python client. The software package comprises servers for the densimeter and viscometer, the pump and valve system and the balance, all based on the FastAPI framework[59]. 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 resulting 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. FINALES requests are pulled via the servers 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 mole 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[60] Python package. This may result in a significant difference between the requested and reported formulation, which poses the risk of blocking the optimization loop due to repeated request of the same, non-feasible formulation.

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 flow with the volume fraction of the respective stock solution. The flows of the stock solutions are merged in the multiplexer. Subsequent to the multiplexer, a valve determines, whether the flow is directed towards the waste while stabilizing the flows or towards the measuring device. After the measurement ended, the sample is drained to the waist using gas aspirated from the atmosphere in the glovebox. 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.

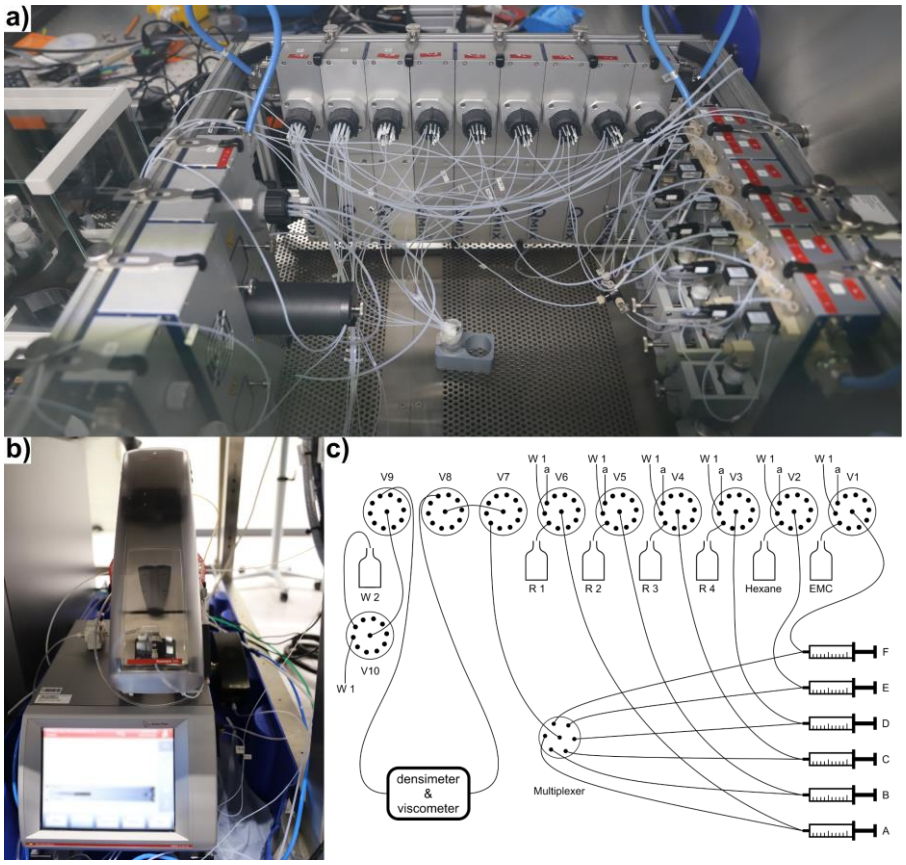

Figure 3: a) The pump and valve system of the experimental ASAB setup used to perform the experiments relevant for this study, b) the densimeter and the attached viscometer device used for the measurements, c) a schematic drawing showing the connections within the system. The valves, reservoirs containing 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.

Optimizer application

The optimizer application is designed in a modular and configurable way. One module is responsible for all communication with the broker server, including authentication, retrieving data and sending new requests. Another module is responsible for data preprocessing and collecting 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 application, the optimizer algorithm is based on a Gaussian process optimizer extended for multi-objective optimization by using Chimera[53]. Finally, a central application module binds the other modules together by checking the state of the broker server, retrieving data, generating and sending new requests in a main loop. The optimizer application 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 client

AiiDA[61–63] is a general purpose workflow management platform for computational research projects, with a plugin interface to support external code. 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 AiiDA client 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. 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 complete workflow.

The AiiDA-FINALES client 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 client, and also provides an easy way of populating requests in the server, which is particularly useful for testing. Once the client is started, it connects to the server, pulls any request for conductivity measurements, and submits a simulation that estimates a low-fidelity value for this physical property. 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 calculation performed for the low-fidelity conductivity value is based on the model reported by Rahmanian et al.[64] developed using one-shot active learning. 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 ºC to 60 ºC. 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 10ºC. In the implementation used here, any temperature is accepted and the parameters are interpolated from those corresponding to the two nearest multiples of ten.
The implementation serves as a proof of concept of the capabilities that can be enabled by connecting AiiDA as a client to FINALES. Further features, such as a persistence mechanism to allow the client to be restarted or error handling will be implemented in the future, also leveraging future features of the server.

Demonstration of Operability

To demonstrate the operability of the system, we chose to involve the ASAB system and Pipeline Pilot for performing the experiments and simulations, respectively. The optimizer application was used to guide the optimization procedure and send requests to FINALES. Maximizing viscosity while minimizing density by variation of the formulation was chosen as the objective for the run. The suggestions of the optimizer for adequate follow-up experiments were based on the experimental as well as the simulated data available in the database interfaced through FINALES.

The samples are formulated based on stock solutions provided in vials prior to the start of the experiment. For the purpose of the demonstration reported herein, commercial solutions of 1 M $LiPF_6$ in EC:EMC 3:7 by weight, and 1 M $LiPF_6$ in EC:DMC 1:1 by weight, as well as EC:EMC 3:7 by weight and DMC, all ordered from E-Lyte Innovations GmbH are used as received and provided as stock solutions to the setup. In preparation of the demonstration run, density measurements for each stock solution needed to be performed as these values are required as inputs to calculate the volume fractions from the mole fractions to achieve the correct formulation. Each measurement originating from the experimental setup comprises three individual measurements. The value for each individual measurement is reported to FINALES within one measurement. Hence, it is left up to the user of the data, whether the data shall be averaged or used as is.

Figure 4 shows the averaged data for each quantity and each measurement in the sequence, in which they were retrieved from the server. This sequence does not necessarily correspond to the order, in which the experiments were performed or requested. Therefore, it is not possible to draw any conclusions regarding the effectivity of the optimization. The demonstration revealed some shortcomings in the current status of the development of the agents. In the experimental setup, cross-contaminations resulted in the mixing ratios not matching with the calculated ones as accurately as expected. Further, it is possible that the simulation agent reaches boundaries regarding the composition when the mole fraction of the salt exceeds 50 %. Also, our results show

a strong correlation between the viscosity and density in the system under investigation as shown in Figure 5. This renders minimum density and maximum viscosity a non-ideal target for optimization in this formulation space.

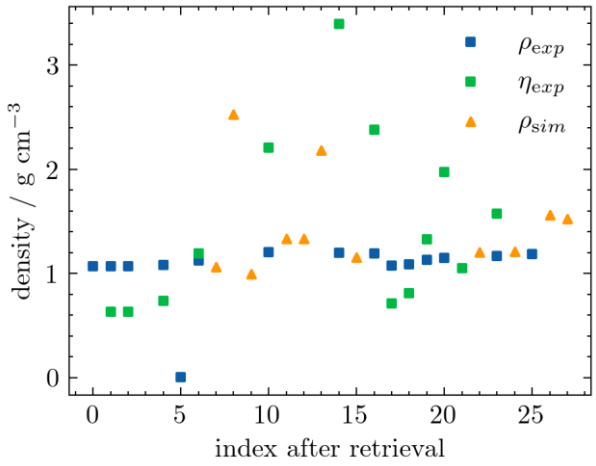

Figure 4: Data obtained from the server after the run. The experimentally determined densities, ρ_{exp}, and viscosities, η_{exp}, are shown as well as the density obtained from the simulations, ρ_{sim}. The data are presented in the sequence, in which they are obtained from the server. This does not necessarily correspond to the sequence, in which they were recorded or requested. Zero-valued experimental results arise because it was decided to report zero, if all the individual measurements failed.

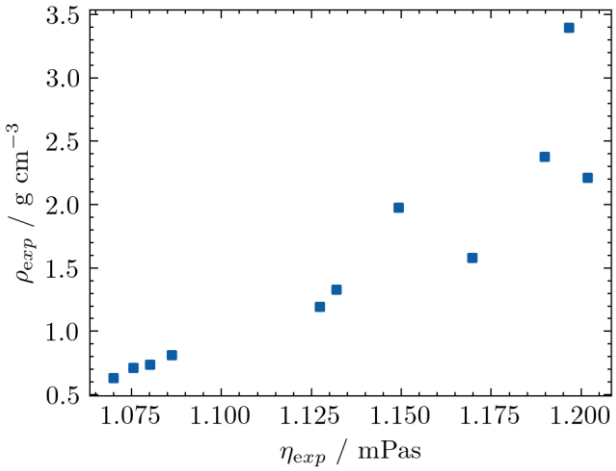

Figure 5: This graph shows the correlation between the density ρ_{exp} and the viscosity η_{exp} as observed from the formulations yielding both, successful density and viscosity measurements.

The chemical space covered during the run is shown in Figure 6. The squared markers represent experimentally measured data, triangles indicate results obtained from simulations. The round markers with blue outlines represent requested formulations. A significant deviation between the formulations targeted in the experiments and the formulations requested by the optimizer can be observed. This most probably depends on the stock solutions, which do not allow to cover the full chemical space spanned by the individual chemicals. Furthermore, the approximation required during the transition from mole fractions to volume fractions and back during the experiment adds another source for the deviation.

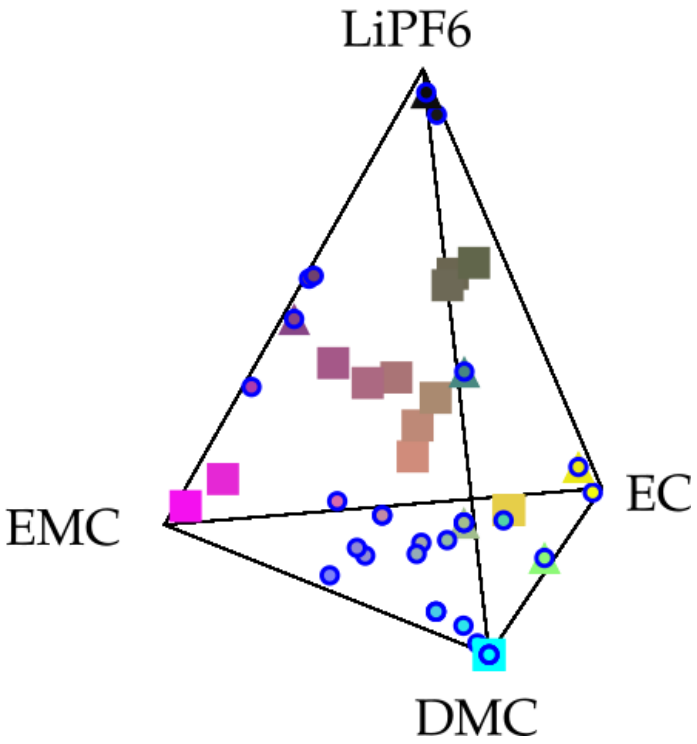

Figure 6: The compositional space covered during the demonstration run. Squared markers represent target formulations of the experiments, triangles mark formulations covered by simulations and the circles with blue outlines represent the requested formulations.

Discussion and Lessons Learned

Our demonstration revealed several crucial aspects to be considered when designing and running an international MAP. In the following, we will summarize the lessons learned from this demonstration.

The experience gained from the demonstration presented herein shows the major importance of a binding agreement between all parties regarding input and output data formats for successful operation and interaction. The specified data format implemented in FINALES allows for the integration of an ontology, which can improve interoperability of the generated data. This is further fostered by storing the data in an open machine-readable format as opposed to proprietary formats. Further, data from various sources also outside of FINALES may be incorporated for analysis with the help of an ontology. Existing data can therefore be complemented by additional data for more detailed investigations.

The addition-only data storage implemented in FINALES allows traceability of the requests and responses. Additionally, the metadata stored with each dataset enables e.g. the identification of failed measurements, which the user may want to handle differently from successful measurements during data analysis. However, an agreement on the handling of missing values proved to be crucial even outside the optimizer agent. Missing values can originate from failures during data generation or upon merging data from various sources, where some quantities are by design not provided by all services. Compliance with the schemas offered by FINALES requires the necessary fields to be filled with an appropriate data type. However, it needs to be possible to filter for invalid or unavailable data. Therefore, either filling values need to be chosen, which are universally applicable for all agents, or each individual value needs to be labelled regarding its validity. Both approaches have limitations for some kinds of results. So, care must be taken when deciding for one solution or the other. A mixed approach could also be worth to be considered.

The ability of handling side-results, which are typically generated in any experiment or simulation, is an important aspect to maximize knowledge gain from each individual activity. Our approach on this is to record a comprehensive number of parameters not only focusing on the data of primary interest. A strong interconnection between the various data, corresponding metadata, requests, and reports is crucial to allow for versatile entry points of data access. Unique identifiers corresponding each part of the data and timestamps showed to be very important for tracking the origin and sequence of results. Data lineage takes the collection of metadata one step further. This is not yet included in the current implementation of our MAP, as the full data lineage in the MAP cannot be provided by a central service. Full data lineage tracking requires each agent to provide the data lineage for its workflow and to report it to FINALES. In our framework, AiiDA, HELAO and Pipeline Pilot offer data lineage tracking functionalities. In the future, FINALES needs to keep track of data lineage during the overall workflow and needs to collect all the data lineage from the connected agents.

Due to the intention agnostic approach of FINALES, several intention-aware tenants can request experiments, simulations or other services available in the MAP. The various intention-aware tenants may access all the data, comprising even the ones not requested by the respective tenant. Hence, all intention-aware tenants can benefit from all the available data while being able to pursue very different or even opposed objectives. Efficient usage of the hardware equipment and data is therefore maximized.

In a large-scale application, a billing system will be required to avoid accumulation of cost for some of the tenants as the cost of the infrastructure may vary significantly depending on the service to be offered. A subscription model could offer a solution. However, the pricing might be volatile, if new services are added continuously.

The communication of limitations of tenants through FINALES to other tenants showed to be very important to limit requests to a feasible range. In our demonstration, this would apply e.g. to the composition of the stock solutions or limitations arising due to the limited number of molecules, which can be represented in the simulations. The limitations of each tenant are relevant for the generation of requests. The future development of the experimental setup should include a check of the actual chemical composition used for the experiments, as this is not yet covered in the current state.

Overall, the demonstration reported here proved the operability of a MAP on an international scale, even though, a variety of challenges is to be tackled during further development and integration of an increasing number of agents.

Acknowledgement

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 programme 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. MV acknowledges the KIT Graduate School Enabling Net Zero - ENZo.

References

(1) Maier, W. F. Early Years of High-Throughput Experimentation and Combinatorial Approaches in Catalysis and Materials Science. *ACS Comb. Sci.* **2019**, *21* (6), 437–444. https://doi.org/10.1021/acscombsci.8b00189.

(2) Stein, H. S.; Gregoire, J. M. Progress and Prospects for Accelerating Materials Science with Automated and Autonomous Workflows. *Chem. Sci.* **2019**, No. 10.1039/C9SC03766G.

(3) Soedarmadji, E.; Stein, H. S.; Suram, S. K.; Guevarra, D.; Gregoire, J. M. Tracking Materials Science Data Lineage to Manage Millions of Materials Experiments and Analyses. *Npj Comput. Mater.* **2019**, *5* (1), 79. https://doi.org/10.1038/s41524-019-0216-x.

(4) Amis, E. J.; Xiang, X.-D.; Zhao, J.-C. Combinatorial Materials Science: What's New Since Edison? *MRS Bull.* **2002**, *27* (4), 295–300. https://doi.org/10.1557/mrs2002.96.

(5) Koinuma, H.; Takeuchi, I. Combinatorial Solid-State Chemistry of Inorganic Materials. *Nat. Mater.* **2004**, *3* (7), 429–438. https://doi.org/10.1038/nmat1157.

(6) Ludwig, A.; Zarnetta, R.; Hamann, S. Development of Multifunctional Thin Films Using High-Throughput Experimentation Methods. *J. Mater. Chem. A* **2008**, *99* (10), 1144–1149. https://doi.org/10.3139/146.101746.

(7) Ludwig, A. Discovery of New Materials Using Combinatorial Synthesis and High-Throughput Characterization of Thin-Film Materials Libraries Combined with Computational Methods. *Npj Comput. Mater.* **2019**, *5* (1), 1–7. https://doi.org/10.1038/s41524-019-0205-0.

(8) Bhowmik, A.; Castelli, I. E.; Garcia-Lastra, J. M.; Jørgensen, P. B.; Winther, O.; Vegge, T. A Perspective on Inverse Design of Battery Interphases Using Multi-Scale Modelling, Experiments and Generative Deep Learning. *Energy Storage Mater.* **2019**, *21*, 446–456. https://doi.org/10.1016/j.ensm.2019.06.011.

(9) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361* (6400), 360–365.

(10) Noh, J.; Kim, J.; Stein, H. S.; Sanchez-Lengeling, B.; Gregoire, J. M.; Aspuru-Guzik, A.; Jung, Y. Inverse Design of Solid-State Materials via a Continuous Representation. *Matter* **2019**, *1* (5), 1370–1384. https://doi.org/10.1016/j.matt.2019.08.017.

(11) Oses, C.; Toher, C.; Gossett, E.; Tropsha, A.; Isayev, O.; Curtarolo, S. Universal Fragment Descriptors for Predicting Properties of Inorganic Crystals. *Nat. Commun.* **2017**, *8*, 1–12. https://doi.org/10.1038/ncomms15679.

(12) Peterson, A. A.; Nørskov, J. K. Activity Descriptors for CO_2 Electroreduction to Methane on Transition-Metal Catalysts. *J. Phys. Chem. Lett.* **2012**, *3* (2), 251–258. https://doi.org/10.1021/jz201461p.

(13) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12* (3), 191–201. https://doi.org/10.1038/nmat3568.

(14) 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.; others. Data Management Plans: The Importance of Data Management in the BIG-MAP Project. *Batter. Supercaps* **2021**, *4* (12), 1803–1812.

(15) Ling, J.; Hutchinson, M.; Antono, E.; Paradiso, S.; Meredig, B. High-Dimensional Materials and Process Optimization Using Data-Driven Experimental Design with Well-Calibrated Uncertainty Estimates. *Integrating Mater. Manuf. Innov.* **2017**, *6* (3), 207–217. https://doi.org/10.1007/s40192-017-0098-z.

(16) Ament, S. E.; Stein, H. S.; Guevarra, D.; Zhou, L.; Haber, J. A.; Boyd, D. A.; Umehara, M.; Gregoire, J. M.; Gomes, C. P. Multi-Component Background Learning Automates Signal Detection for Spectroscopic Data. *Npj Comput. Mater.* **2019**, *5* (1), 1–7. https://doi.org/10.1038/s41524-019-0213-0.

(17) Schwarting, M.; Siol, S.; Talley, K.; Zakutayev, A.; Phillips, C. Automated Algorithms for Band Gap Analysis from Optical Absorption Spectra. *Mater. Discov.* **2017**, *10*, 43–52. https://doi.org/10.1016/j.md.2018.04.003.

(18) Umehara, M.; Stein, H. S.; Guevarra, D.; Newhouse, P. F.; Boyd, D. A.; Gregoire, J. M. Analyzing Machine Learning Models to Accelerate Generation of Fundamental Materials Insights. *Npj Comput. Mater.* **2019**, *5* (1), 34. https://doi.org/10.1038/s41524-019-0172-5.

(19) *Report of the Clean Energy Materials Innovation Challenge Expert Workshop January 2018, Mission Innovation Http://Mission-Innovation.Net/Wp-Content/Uploads/2018/01/Mission-Innovation-IC6-Report-Materials-Acceleration-Platform-Jan-2018.Pdf*.

(20) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: Orchestrating Autonomous Experimentation. *Sci. Robot.* **2018**, *3* (19). https://doi.org/10.1126/scirobotics.aat5559.

(21) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2.

(22) Amici, J.; Asinari, P.; Ayerbe, E.; Barboux, P.; Bayle-Guillemaud, P.; Behm, R. J.; Berecibar, M.; Berg, E.; Bhowmik, A.; Bodoardo, S.; Castelli, I. E.; Cekic-Laskovic, I.; Christensen, R.; Clark, S.; Diehm, R.; Dominko, R.; Fichtner, M.; Franco, A. A.; Grimaud, A.; Guillet, N.; Hahlin, M.; Hartmann, S.; Heiries, V.; Hermansson, K.; Heuer, A.; Jana, S.; Jabbour, L.; Kallo, J.; Latz, A.; Lorrmann, H.; Løvvik, O. M.; Lyonnard, S.; Meeus, M.; Paillard, E.; Perraud, S.; Placke, T.; Punckt, C.; Raccurt, O.; Ruhland, J.; Sheridan, E.; Stein, H.; Tarascon, J.-M.; Trapp, V.; Vegge, T.; Weil, M.; Wenzel, W.; Winter, M.; Wolf, A.; Edström, K. 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.* **2022**, *12* (17), 2102785. https://doi.org/10.1002/aenm.202102785.

(23) Fichtner, M. Recent Research and Progress in Batteries for Electric Vehicles. *Batter. Supercaps* **2022**, *5* (2). https://doi.org/10.1002/batt.202100224.

(24) Fichtner, M.; Edström, K.; Ayerbe, E.; Berecibar, M.; Bhowmik, A.; Castelli, I. E.; Clark, S.; Dominko, R.; Erakca, M.; Franco, A. A.; Grimaud, A.; Horstmann, B.; Latz, A.; Lorrmann, H.; Meeus, M.; Narayan, R.; Pammer, F.; Ruhland, J.; Stein, H.; Vegge, T.; Weil, M. Rechargeable Batteries of the Future—The State of the Art from a BATTERY 2030+ Perspective. *Adv. Energy Mater.* **2021**, 2102904. https://doi.org/10.1002/aenm.202102904.

(25) Bhowmik, A.; Berecibar, M.; Casas-Cabanas, M.; Csanyi, G.; Dominko, R.; Hermansson, K.; Palacin, M. R.; Stein, H. S.; Vegge, T. Implications of the BATTERY 2030+ AI-Assisted Toolkit

on Future Low-TRL Battery Discoveries and Chemistries. *Adv. Energy Mater.* **2022**, *12* (17), 2102698. https://doi.org/10.1002/aenm.202102698.

(26) Nikolaev, P.; Hooper, D.; Perea-López, N.; Terrones, M.; Maruyama, B. Discovery of Wall-Selective Carbon Nanotube Growth Conditions *via* Automated Experimentation. *ACS Nano* **2014**, *8* (10), 10214–10222. https://doi.org/10.1021/nn503347a.

(27) 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.; Takeuchi, I. On-the-Fly Machine-Learning for High-Throughput Experiments: Search for Rare-Earth-Free Permanent Magnets. *Sci. Rep.* **2014**, *4* (1), 6367. https://doi.org/10.1038/srep06367.

(28) J. Li, Y.; Savan, A.; Kostka, A.; S. Stein, H.; Ludwig, A. Accelerated Atomic-Scale Exploration of Phase Evolution in Compositionally Complex Materials. *Mater. Horiz.* **2018**, *5* (1), 86–92. https://doi.org/10.1039/C7MH00486A.

(29) Stein, H. S.; Sanin, A.; Rahmanian, F.; Zhang, B.; Vogler, M.; Flowers, J. K.; Fischer, L.; Fuchs, S.; Choudhary, N.; Schroeder, L. From Materials Discovery to System Optimization by Integrating Combinatorial Electrochemistry and Data Science. *Curr. Opin. Electrochem.* **2022**, *35*, 101053. https://doi.org/10.1016/j.coelec.2022.101053.

(30) Rahmanian, F.; Flowers, J.; Guevarra, D.; Richter, M.; Fichtner, M.; Donnely, P.; Gregoire, J. M.; Stein, H. S. Enabling Modular Autonomous Feedback-Loops in Materials Science through Hierarchical Experimental Laboratory Automation and Orchestration. *Adv. Mater. Interfaces* **2022**, *9* (8), 2101987.

(31) Hattrick-Simpers, J. R.; Gregoire, J. M.; Kusne, A. G. Perspective: Composition–Structure–Property Mapping in High-Throughput Experiments: Turning Data into Knowledge. *APL Mater.* **2016**, *4* (5), 053211–053212. https://doi.org/10.1063/1.4950995.

(32) Suram, S. K.; Pesenson, M. Z.; Gregoire, J. M. High Throughput Combinatorial Experimentation + Informatics = Combinatorial Science. In *Information Science for Materials Discovery and Design*; Springer International Publishing: Cham, 2015; pp 271–300. https://doi.org/10.1007/978-3-319-23871-5_14.

(33) Rohr, B.; Stein, H. S.; Guevarra, D.; Wang, Y.; Haber, J. A.; Aykol, M.; Suram, S. K.; Gregoire, J. M. Benchmarking the Acceleration of Materials Discovery by Sequential Learning. *Chem. Sci.* **2020**, *11* (10), 2696–2706. https://doi.org/10.1039/C9SC05999G.

(34) Aykol, M.; Hummelshøj, J. S.; Anapolsky, A.; Aoyagi, K.; Bazant, M. Z.; Bligaard, T.; Braatz, R. D.; Broderick, S.; Cogswell, D.; Dagdelen, J.; Drisdell, W.; Garcia, E.; Garikipati, K.; Gavini, V.; Gent, W. E.; Giordano, L.; Gomes, C. P.; Gomez-Bombarelli, R.; Balaji Gopal, C.; Gregoire, J. M.; Grossman, J. C.; Herring, P.; Hung, L.; Jaramillo, T. F.; King, L.; Kwon, H.-K.; Maekawa, R.; Minor, A. M.; Montoya, J. H.; Mueller, T.; Ophus, C.; Rajan, K.; Ramprasad, R.; Rohr, B.; Schweigert, D.; Shao-Horn, Y.; Suga, Y.; Suram, S. K.; Viswanathan, V.; Whitacre, J. F.; Willard, A. P.; Wodo, O.; Wolverton, C.; Storey, B. D. The Materials Research Platform: Defining the Requirements from User Stories. *Matter* **2019**, *1* (6), 1433–1438. https://doi.org/10.1016/j.matt.2019.10.024.

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

(36) Talirz, L.; Kumbhar, S.; Passaro, E.; Yakutovich, A. V.; Granata, V.; Gargiulo, F.; Borelli, M.; Uhrin, M.; Huber, S. P.; Zoupanos, S.; Adorf, C. S.; Andersen, C. W.; Schütt, O.; Pignedoli, C. A.; Passerone, D.; VandeVondele, J.; Schulthess, T. C.; Smit, B.; Pizzi, G.; Marzari, N. Materials Cloud, a Platform for Open Computational Science. *Sci. Data* **2020**, *7* (1), 299. https://doi.org/10.1038/s41597-020-00637-5.

(37) Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Becker, J. S.; A., A.-G. Closed-Loop Discovery Platform Integration Is Needed for Artificial Intelligence to Make an Impact in Drug Discovery. *Expert Opin. Drug Discov.* **2019**, *14* (1), 1–4. https://doi.org/10.1080/17460441.2019.1546690.

(38) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends Chem.* **2019**, *1* (3), 282–291. https://doi.org/10.1016/j.trechm.2019.02.007.

(39) Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabeca, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multi-Component Systems. *ArXiv190903511 Phys.* **2019**.

(40) MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Häse, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6* (20), eaaz8867. https://doi.org/10.1126/sciadv.aaz8867.

(41) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part II: Outlook. *Angew. Chem. Int. Ed.* **2020**, *59* (52), 23414–23436. https://doi.org/10.1002/anie.201909989.

(42) Kitano, H. Nobel Turing Challenge: Creating the Engine for Scientific Discovery. *Npj Syst. Biol. Appl.* **2021**, *7* (1), 29. https://doi.org/10.1038/s41540-021-00189-3.

(43) Stach, E.; DeCost, B.; Kusne, A. G.; Hattrick-Simpers, J.; Brown, K. A.; Reyes, K. G.; Schrier, J.; Billinge, S.; Buonassisi, T.; Foster, I.; Gomes, C. P.; Gregoire, J. M.; Mehta, A.; Montoya, J.; Olivetti, E.; Park, C.; Rotenberg, E.; Saikin, S. K.; Smullin, S.; Stanev, V.; Maruyama, B. Autonomous Experimentation Systems for Materials Development: A Community Perspective. *Matter* **2021**, *4* (9), 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.

(44) Seifrid, M.; Hattrick-Simpers, J.; Aspuru-Guzik, A.; Kalil, T.; Cranford, S. Reaching Critical MASS: Crowdsourcing Designs for the next Generation of Materials Acceleration Platforms. *Matter* **2022**, *5* (7), 1972–1976. https://doi.org/10.1016/j.matt.2022.05.035.

(45) Attia, P. M.; Grover, A.; Jin, N.; Severson, K. A.; Markov, T. M.; Liao, Y.-H.; Chen, M. H.; Cheong, B.; Perkins, N.; Yang, Z.; Herring, P. K.; Aykol, M.; Harris, S. J.; Braatz, R. D.; Ermon, S.; Chueh, W. C. Closed-Loop Optimization of Fast-Charging Protocols for Batteries with Machine Learning. *Nature* **2020**, *578* (7795), 397–402. https://doi.org/10.1038/s41586-020-1994-5.

(46) Herring, P.; Balaji Gopal, C.; Aykol, M.; Montoya, J. H.; Anapolsky, A.; Attia, P. M.; Gent, W.; Hummelshøj, J. S.; Hung, L.; Kwon, H.-K.; Moore, P.; Schweigert, D.; Severson, K. A.; Suram, S.; Yang, Z.; Braatz, R. D.; Storey, B. D. BEEP: A Python Library for Battery Evaluation and Early Prediction. *SoftwareX* **2020**, *11*, 100506. https://doi.org/10.1016/j.softx.2020.100506.

(47) Benayad, A.; Diddens, D.; Heuer, A.; Krishnamoorthy, A. N.; Maiti, M.; Cras, F. L.; Legallais, M.; Rahmanian, F.; Shin, Y.; Stein, H.; Winter, M.; Wölke, C.; Yan, P.; Cekic-Laskovic, I. High-Throughput Experimentation and Computational Freeway Lanes for Accelerated Battery

Electrolyte and Interface Development Research. *Adv. Energy Mater.* **2021**, 2102678. https://doi.org/10.1002/aenm.202102678.

(48) Borodin, O.; Olguin, M.; Spear, C. E.; Leiter, K. W.; Knap, J. Towards High Throughput Screening of Electrochemical Stability of Battery Electrolytes. *Nanotechnology* **2015**, *26* (35), 1–15. https://doi.org/10.1088/0957-4484/26/35/354003.

(49) Ding, M. S.; Xu, K.; Zhang, S. S.; Amine, K.; Henriksen, G. L.; Jow, T. R. Change of Conductivity with Salt Content, Solvent Composition, and Temperature for Electrolytes of LiPF6 in Ethylene Carbonate-Ethyl Methyl Carbonate. *J. Electrochem. Soc.* **2001**, *148* (10), A1196. https://doi.org/10.1149/1.1403730.

(50) Dalavi, S.; Guduru, P.; Lucht, B. L. Performance Enhancing Electrolyte Additives for Lithium Ion Batteries with Silicon Anodes. *J. Electrochem. Soc.* **2012**, *159* (5), A642–A646. https://doi.org/10.1149/2.076205jes.

(51) Wang, Y.; Meng, X.; Sun, J.; Liu, Y.; Hou, L. Recent Progress in “Water-in-Salt” Electrolytes Toward Non-Lithium Based Rechargeable Batteries. *Front. Chem.* **2020**, *8*, 595. https://doi.org/10.3389/fchem.2020.00595.

(52) Zhang, B.; Fischer, L.; Sanin, A.; Stein, H. S. Robotic Cell Assembly to Accelerate Battery Research. *chemRxiv* **2022**.

(53) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Chimera: Enabling Hierarchy Based Multi-Objective Optimization for Self-Driving Laboratories. *Chem. Sci.* **2018**, *9* (39), 7642–7655. https://doi.org/10.1039/C8SC02239A.

(54) Colvin, S.; Contributors to pydantic. *pydantic*. https://github.com/pydantic/pydantic (accessed 2022-08-31).

(55) Clark, S.; Bleken, F. L.; Stier, S.; Flores, E.; Andersen, C. W.; Marcinek, M.; Szczesna-Chrzan, A.; Gaberscek, M.; Palacin, M. R.; Uhrin, M.; Friis, J. Toward a Unified Description of Battery Data. *Adv. Energy Mater.* **2022**, *12* (17), 2102702. https://doi.org/10.1002/aenm.202102702.

(56) Allan, D.; Caswell, T.; Campbell, S.; Rakitin, M. Bluesky’s Ahead: A Multi-Facility Collaboration for an *a La Carte* Software Project for Data Acquisition and Management. *Synchrotron Radiat. News* **2019**, *32* (3), 19–22. https://doi.org/10.1080/08940886.2019.1608121.

(57) Hanke, F.; Modrow, N.; Akkermans, R. L. C.; Korotkin, I.; Mocanu, F. C.; Neufeld, V. A.; Veit, M. Multi-Scale Electrolyte Transport Simulations for Lithium Ion Batteries. *J. Electrochem. Soc.* **2019**, *167* (1), 013522. https://doi.org/10.1149/2.0222001JES.

(58) Schaarschmidt, J.; Yuan, J.; Strunk, T.; Kondov, I.; Huber, S. P.; Pizzi, G.; Kahle, L.; Bölle, F. T.; Castelli, I. E.; Vegge, T.; Hanke, F.; Hickel, T.; Neugebauer, J.; Rêgo, C. R. C.; Wenzel, W. Workflow Engineering in Materials Design within the BATTERY 2030+ Project. *Adv. Energy Mater.* **2022**, *12* (17), 2102638. https://doi.org/10.1002/aenm.202102638.

(59) Ramírez, S.; Contributors to FastAPI. *FastAPI*. fastapi. https://fastapi.tiangolo.com/ (accessed 2022-08-17).

(60) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

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

(62) Huber, S. P.; Zoupanos, S.; Uhrin, M.; Talirz, L.; Kahle, L.; Häuselmann, R.; Gresch, D.; Müller, T.; Yakutovich, A. V.; Andersen, C. W.; Ramirez, F. F.; Adorf, C. S.; Gargiulo, F.; Kumbhar, S.; Passaro, E.; Johnston, C.; Merkys, A.; Cepellotti, A.; Mounet, N.; Marzari, N.; Kozinsky, B.; Pizzi, G. AiiDA 1.0, a Scalable Computational Infrastructure for Automated Reproducible Workflows and Data Provenance. *Sci. Data* **2020**, *7* (1), 300. https://doi.org/10.1038/s41597-020-00638-4.

(63) Uhrin, M.; Huber, S. P.; Yu, J.; Marzari, N.; Pizzi, G. Workflows in AiiDA: Engineering a High-Throughput, Event-Based Engine for Robust and Modular Computational Workflows. *Comput. Mater. Sci.* **2021**, *187*, 110086. https://doi.org/10.1016/j.commatsci.2020.110086.

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