D10.4 – Tiered theory and experiment screening pipeline as first test case for automatic reasoning calibration

VERSION

VERSION	DATE
1.0	14.12.2023

PROJECT INFORMATION

GRANT AGREEMENT NUMBER	957189
PROJECT FULL TITLE	Battery Interface Genome - Materials Acceleration Platform
PROJECT ACRONYM	BIG-MAP
START DATE OF THE PROJECT	1/9-2020
DURATION	3 years
CALL IDENTIFIER	H2020-LC-BAT-2020-3
PROJECT WEBSITE	big-map.eu

DELIVERABLE INFORMATION

WP NO.	10
WP LEADER	DTU, Tejs Vegge
CONTRIBUTING PARTNERS	KIT, DTU, 3DS, EPFL, SINTEF
NATURE	Report
AUTHORS	H.S. Stein (KIT* now: TUM), M. Vogler, F. Fernando Ramirez, E. Flores, L.H Rieger, L. Merker, S. Krarup Steensen, J.M. Carlsson, F. Liot, G. Pizzi, B. Zhang
CONTRIBUTORS	V. Kumar, for a list of all the contributors to the second demonstration of FINALES please see the author list in the BIG-MAP Archive entry: https://archive.big-map.eu/records/j5wjm-q0r39
CONTRACTUAL DEADLINE	31.12.2023
DELIVERY DATE TO EC	14-12-2023
DISSEMINATION LEVEL	PU

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. The project is part of BATTERY 2030+, the large-scale European research initiative for inventing the sustainable batteries of the future.

ABSTRACT

This report describes two demonstrations of the **F**ast **IN**tention-**A**gnostic **LE**arning **S**erver (FINALES). FINALES allows the user to request results from a MAP without the need to know the details of the respective data acquisition. Many units, which we call tenants, may be connected to FINALES to either offer their service to the MAP or to send requests. FINALES explicitly allows for several of the connected tenants to provide the same capabilities in terms of the data they can provide and/or the method they use. This multitenancy enables an efficient use of resources by rendering the MAP independent of the state of individual tenants. A very distinctive feature of FINALES is its passive mode of operation by not actively triggering or controlling the tenants. It rather serves as a marketplace fostering communication between the tenants. Hence, the tenants are free to operate according to their own schedule and the MAP is not impeded by individual tenants being temporarily unavailable. Furthermore, FINALES does not require the tenants to be fully automated instruments. Humans may interact with FINALES by pulling requests from the server and posting results in JSON format.

TABLE OF CONTENTS

1. FINALES

The **F**ast **IN**tention-**A**gnostic **LE**arning **S**erver offers access to experimental as well as computational capabilities to registered users. In contrast to similar frameworks, FINALES does not actively trigger actions in the MAP or control devices. This enables a wide variety of tenants with largely different capabilities as well as length and time scales to be connected to the MAP without blocking each other. Details about the design concept of FINALES may be found in deliverable D9.6 *Demonstrator of a simulation run by an experiment, and an experiment run by a simulation*.

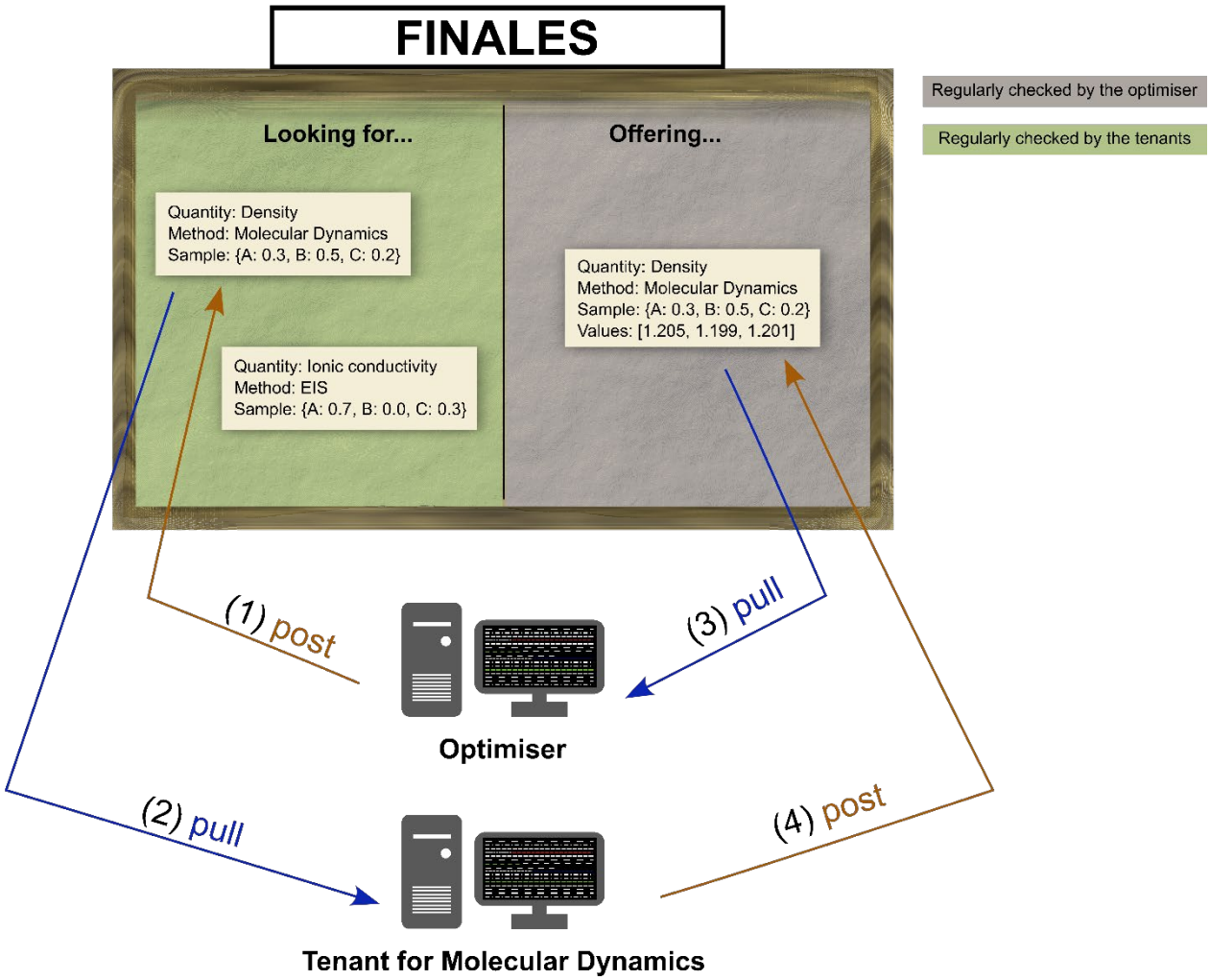


Figure 1. Graphical representation of the working principle of a FINALES run – The optimiser posts a request (1), which is picked up by a tenant, which can fulfil it (2). Once the request is processed, the tenant posts a result to FINALES (3). The optimiser collects all the available results in regular time intervals (4). After processing the available results, the optimiser posts a new request (1) and the next iteration starts.

1.1. Optimiser

The optimiser is a software tenant with modules responsible for retrieving and preprocessing data from FINALES, generating candidates for new measurement parameters within the limitations provided by the data producing tenants, applying a Bayesian optimisation algorithm and compiling and sending measurement requests back to FINALES. The optimiser code was developed in Python using the Torch and GPyTorch packages.

The role of the optimiser tenant is to request new measurements from the data producing methods available in the MAP with the aim of optimising some predefined objective(s). In general, the optimiser is configured to iteratively consume previously observed measurement results from FINALES, use the data to train a machine learning model and apply a Bayesian optimisation procedure to propose new measurement parameters which are submitted to FINALES as new measurement requests. The specific optimisation task, including what data to consume and which quantities to optimise, is specified in a configuration file.

To fully utilise the capabilities of the MAP, the optimiser needs to be able to consider measurements of different fidelity and from different data producing methods, such as simulation and experiment. Data from different sources can have different noise and bias as well as different associated costs of evaluation which need to be considered. Additionally, it is sometimes desirable to co-optimise multiple objectives at the same time. Therefore, the optimiser needs to be capable of multi-source multi-objective optimisation to enable various optimisation tasks in the MAP.

In the Baysian optimisation procedure, we apply a Gaussian process (GP) regression model which is a popular choice of model in Bayesian optimisation because it is highly flexible, provides uncertainty estimates and can accommodate small datasets which is often the starting point in optimisation tasks. To handle data from multiple sources we consider a multi-task GP and treat each data source as a separate output. This allows for separate predictions for each data source while utilising correlations between data sources to improve the predictions. An acquisition function can then be applied to propose new measurement parameters. Multi-objective optimization can be achieved by transforming multiple objectives into a single objective with a scalarizing function which can then be optimised using standard acquisition functions.

1.2. Molecular Dynamics simulations

The molecular dynamics simulation tenant is created using BIOVIA Pipeline Pilot. This is a workflow engine and data analysis tool, which is programmed by connecting data processing components by data pipelines into workflows called protocols. The components may have various functionality like connecting to a database, performing data analysis or wrap and execute computational codes like the Forcite code for molecular dynamics. The components can also execute scripts in various languages like perl, python or the internal language Pilot script.

The role of the simulation tenant is to calculate the ionic conductivity of an electrolyte formulation suggested by the optimiser. The simulation tenant is complementary to the experimental measurement of the ionic conductivity by the ASAB tenant. The MD simulation tenant can perform

MD simulations at higher $LiPF_6$-salt concentrations, so it can cover a larger formulation space. We invested significant time to find simulation parameters in terms of equilibration and production run time to ensure that the values derived by MD simulation agree well with the ASAB measurements for a given formulation.

The protocol to calculate the ionic conductivity uses molecular dynamics (MD). The protocol starts by inquiring the FINALES Server for pending requests for MD simulations. The json file containing the requested formulation is parsed by converting the fractional concentrations into the number of molecules of each species and connecting the formulation with the request ID. The molecular structure for the electrolyte components, provided as SMILES, is inflated into 3D molecular structures. This information is used to generate atomistic models of the electrolyte formulation. These starting configurations are equilibrated under constant pressure and temperature molecular dynamics (NPT-MD) to find the correct density. Subsequent production runs under constant volume and temperature molecular dynamics (NVT-MD) provide statistics to calculate the diffusion coefficient for the Li^+ and PF_6^- ions, which form the basis to calculate the ionic conductivity.
The results of the analysis of the MD simulations are collected together with the request ID and formatted into a JSON file. The result file is submitted back to FINALES for storage under the request ID.

1.3. Overlort

The **Overl**ooking **Or**chestrating **T**enant (Overlort) or workflow tenant is a pure software-based tenant creating the workflow to enable the EOL Optimiser tenant's request.
When requested, it initiates the predefined workflow for this method. Therefore, it checks the defined workflow from top to bottom, which quantity is needed next, or starts the workflow. The Overlort saves all request and result parameters as soon as its finished processing in a JSON document and consequently is resistant to (un)expected stops. By saving all parameters from requests and results, the request for the next quantity can be built based on these parameters. Using the endpoint to request results via the request ID, it keeps track of new results coming in and processes them to create the next request.
The workflow here starts with requesting the cycler to reserve channels. After a successful reservation, ASAB is requested to mix the requested electrolyte, followed by a transportation of the electrolyte to the AutoBASS system. Then the assembly of cells is requested and afterwards another transportation of the cells to the respective channels reserved at the start. Now triggering the request for the cycler, the Overlort creates requests equal to the number of requested cells and after processing the result also the same amount of requests for the degradation model. It composes the result for the initially requested quantity and posts it to the server, as well as saving the whole workflow and its request and results locally in a JSON file. This tenant can handle multiple workflow requests at the same time, as each request/result is also linked to the request ID of the original workflow request.

1.4. ASAB

The system for the **A**utonomous **S**ynthesis and **A**nalysis of **B**attery electrolytes (ASAB) consists of six syringe pumps and ten eleven-port rotary valves, which are commercially available from Cetoni GmbH, Korbussen, a DMA 4100 M densimeter with an attached Lovis 2000 ME viscometer, both manufactured by Anton Paar GmbH, Graz, and an in-house built symmetrical electrochemical cell using stainless steel electrodes connected to a Palmsens 4 potentiostat by PalmSens B.V.. Its main purpose is to automatedly formulate electrolyte solutions from stock solutions and analyse them. It is connected to FINALES as a tenant, which can experimentally determine the density, viscosity, and ionic conductivity.

In preparation of a run, the operator provides the stock solutions to the system and inputs the information about the formulation of the stock solutions, its density, the mass of each chemical contained in the stock solution and metadata like a batch number, manufacturing date, etc. to the software of the system. This information enables ASAB to transform the specifications of formulations given in molar fractions in the requests to the volume fraction representation required for the actual formulation by ASAB. This transformation is necessary, since ASAB creates the electrolyte formulations by merging the flows of the different stock solutions in the ratio approximating the requested one with minimum error. Furthermore, a file containing information about the chemicals available in the stock solutions needs to be supplied by the operator.
Once the preparations of the hardware and the configuration of the software are done, the tenant script may be executed. During operation, the system fully automatically picks up requests from FINALES and processes them sequentially. The results include the measured values and actual values for input parameters possibly differing from the request as long as they are accessible. Alongside this information, relevant metadata reporting as for example the success of the procedure, a quality rating of the result and similar information is returned.

In the MAP demonstrated here, this tenant fulfils two purposes. The first one is to provide electrolyte formulations to the AutoBASS tenant, which assembles coin cells. In this use case, the flows of the stock solutions are merged and directed to an empty vial connected to an outlet of the system. After completion of the formulation, ASAB reports the successful completion of the task and the location of the respective vial to FINALES. This enables human operators to pick up the vial from ASAB and provide it to the AutoBASS tenant. In the second use case, ASAB can be requested to provide experimental conductivity values. In fulfilment of this task, ASAB also formulates an approximation of the requested electrolyte formulation, but the resulting formulation is directly pumped into a measuring cell, where an electrochemical impedance spectroscopy (EIS) measurement is performed. The measuring cell consists of a PTFE body with two oppositely positioned ports serving as an inlet and an outlet, respectively. Orthogonally to the fluidic channel, two stainless steel screws mounted in fittings are connected to the body part serving as electrodes. This setup results in a symmetrical electrochemical cell.

The generated EIS data is analysed using the MADAP[1] software package to obtain the electrolyte resistance. The ionic conductivity of the electrolyte results from the EIS measurements as the inverse of the resistance. The resulting values, including the information regarding the

approximated composition of the sample and the metadata of the used solutions, are reported to FINALES. After the measurement, the samples are discarded.

1.5. AutoBASS

The tenant for the **A**utonomous **B**attery **AS**sembly **S**ystem (AutoBASS) provides the capability of assembling up to 64 CR2032 coin cells in one single batch. The setup is composed of 3 six-axis robotic arms (Mecademic meca500 rev.3), a precision linear rail (Jenny Science Linax LXS 1800), a digital coin cell crimping machine (MTI MSK-160E, China), a 200 µL dispensing module (Sartorius rLine) and a set of 3D printed end effectors. The robots will perform the whole assembly process of CR2032 coin cells, including pick-and-place of electrodes and separator, autocorrection of placement, dispersing electrolyte, transferring, sealing and placement in storage holder, etc. Traceability of the process is realised by collecting visual information using an integrated camera upon placement of components, which will later on be shared in the BIG-MAP Archive by manual upload, since image data can currently not be processed by FINALES.

Prior to running the assembly, the tenant constantly collects the pending request from FINALES, and transfers it into a setup file. Before initiating the assembly, the transportation of the electrolyte to the correct position and preparations of the necessary components (electrodes, separator, spacers, washer and casing parts) must be completed. When started, the tenant will stop gathering new requests, waiting for the assembly to be completed.

Once the assembly runs through, the generated data is posted to FINALES and the tenant will start to gather pending requests again. Since the generated data (pictures taken during image recognition) format and size is currently not compatible with FINALES server, they will be first stored locally and later uploaded into the BIG-MAP archive. Upon the completion of building cells, AutoBASS reports the successful completion and saves the results in a local file.

1.6. Cycler

The tenant for the Battery Cycler from Arbin Instruments provides the capability to reserve channels and cycle cells with a standardised cycling protocol. Cycling data is then analysed and serves as input for the degradation model. Based on in-house developed API servers for the cycler and analysis, further improvements and extensions were implemented for the integration into FINALES2. To increase the performance and multitasking, the analysis was set up as a separate server. The tenant will keep track of channels available and reserve space for incoming requests enabling optimal and dynamic use of the cycler. For each requested cell the tenant creates a unique protocol. Logging of errors and intelligent restart of channels for traceability is implemented as well as some manipulations of the protocol parameters to be flexible to user requirements. Therefore, the user can provide optional keys for number of cycles, cycling current, voltage range and more.

Automatic export of data during cycling or once it is done or even by request is possible. The data is then passed to the analysis and raw, as well as analysed data is saved in hdf5 format. An additional file with the request for the individual cell is created and saved together with the raw data.

In between, relevant data for the model is passed to the Overlort and further to the degradation model tenant. Due to the size, the data will be stored locally first and can then be uploaded into the BIG-MAP archive.

The analysis server splits up the raw data into cycles and saves raw as well as the split-up data in a hdf5 file. Furthermore, the data is used to calculate the difference of the coulombic efficiency, the voltage gap between charge and discharge and the capacity vector variance of the 10^{th} and 40^{th} cycle. The list of capacities for each cycle and the previously mentioned parameters serve as input for the degradation model request.

1.7. Degradation model

Experimental procedures often rely on workflows. In this demonstration of FINALES, several tenants rely on hardware or data generated by another tenant.
This tenant is used to preemptively predict the future EOL (End of Life) of a battery cell early in its lifetime based on data from some initial cycles that the battery cell has already gone through. It consists of a neural network trained on data acquired previously, specifically, an LSTM (Long Short-Term Memory) model ensemble comprising five LSTMs with the same architecture but trained with differently seeded initial weights. The model weights are part of the tenant.

In addition to the normal output predicting the next capacity, the model also outputs its uncertainty about the next capacity. By combining the ensemble uncertainty and the output uncertainty, we can predict the total uncertainty of the model over the capacity trajectory and the final lifetime of each battery.

The model takes as input the capacity of a minimum of 40 cycles along with coulombic efficiency, overpotential, and the variance of the QV-curve, as inspired by Severson et al.[2]. By outputting the uncertainty of the prediction, it is possible to judge whether the prediction is sufficient or if the battery should be cycled for a longer duration.

1.8. Transportation

The transportation tenant consists of two python scripts, which represent the server and a client based on sockets. It serves the need to be able to request the transport of physical samples between hardware setups. In the future, this service could be expanded to even allow the transport between different laboratories and partners. However, in the current demonstration it was only used for transportation within one laboratory.

To deploy the tenant, the server script starts a chat server accessible from within the same network, to which the host of the server script is connected. Users can connect to the server by running the client script. After the server is started, the transportation tenant logs in to FINALES in a pre-configured time interval, pulls pending requests for transportation and distributes the information about the origin and the destination of the transport to all clients in the form of a chat message. The human researchers operating the clients can receive these messages and perform the transport. Upon completion, they can send a reply message to the chat and the transportation tenant posts the result to FINALES. The syntax of the reply allows to either just mark the request as resolved with

the transport being performed as requested or if this was not possible, the actual destination may be provided and will be posted by the tenant.

1.9. Archiving

BIG-MAP Archive[3] serves as a restricted-access digital platform that enables storage and sharing of research data generated by the BIG-MAP project.

The archiving tenant consists of a command line client that interacts with the rich and well-documented application programming interface (API) exposed by the BIG-MAP Archive data repository. This tool named `big-map-archive-api-client`[4] can easily be installed on Linux machines, as a Python package hosted on the public repository PyPI (https://pypi.org/project/big-map-archive-api-client/).
One of its commands (`bma finales-db back-up`) backs up the SQLite database of the FINALES server into the BIG-MAP Archive by either creating a new entry or creating a new version of an existing entry. In both cases, the newly created record is immediately shared with all users of the data repository.
The whole back-up procedure is designed around two key ideas:

- avoiding major data loss in case that the FINALES' database gets corrupted.
- sharing results of calculations and experiments shortly after they were posted to FINALES.

Note that, if a data file remains unchanged from one back-up to the next, the file is uploaded only once. However, the corresponding file link appears in the two entry versions. This saves storage space and reduces back-up time.
A job scheduler is used to run the command automatically and periodically (e.g., every second day). Figure 1 and Figure 2 show the entry version that was created and shared on November 29, 2023.

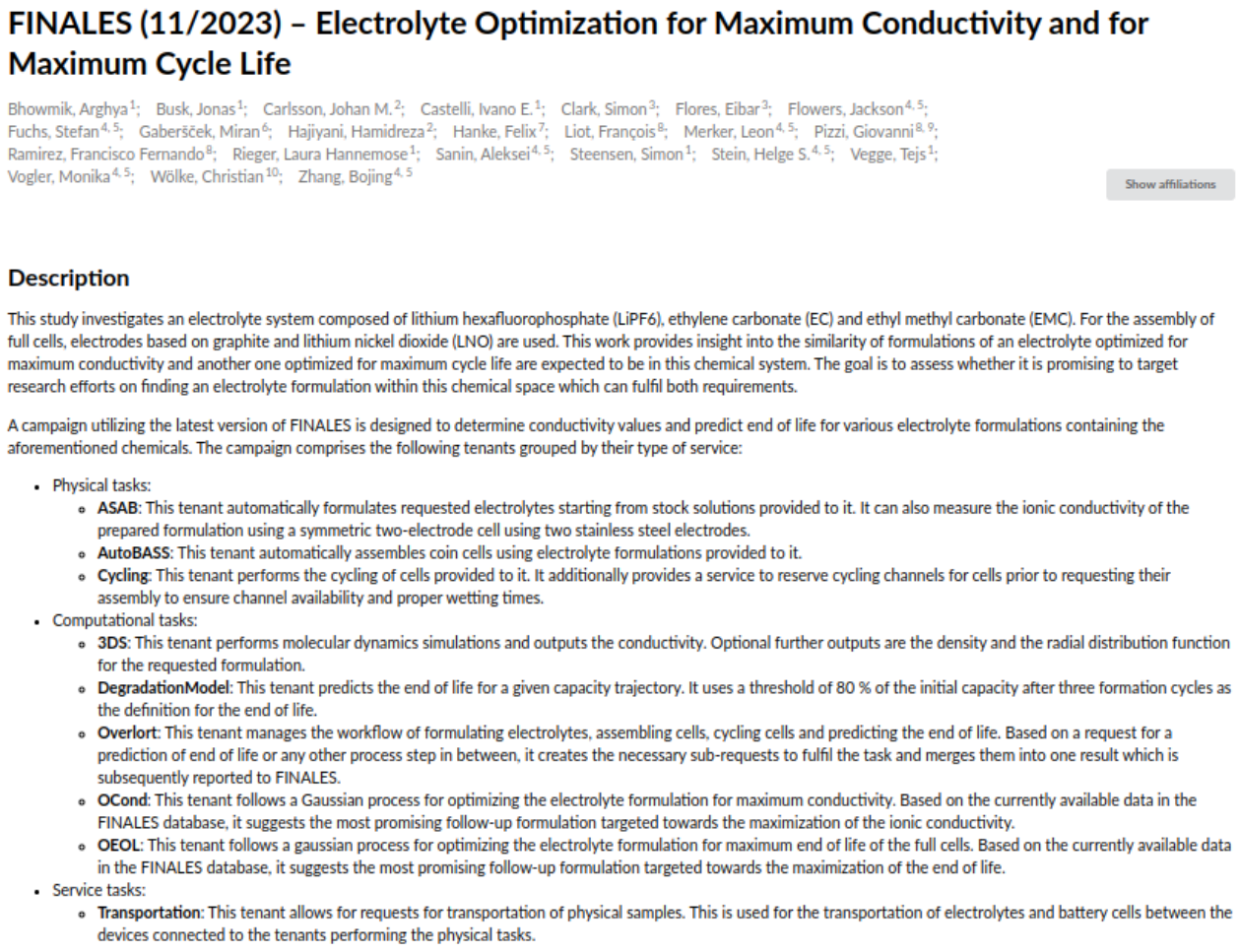
1.10. AiiDA

The AiiDA tenant performs the same task that was originally implemented for the first version of FINALES presented in Demonstration 1 (see section 3.1): a low-fidelity estimation for the conductivity, based on the model reported by Rahmanian et al.[5], developed using one-shot active learning. Improvements of the tenant were focused not on the underlying task, but on the supporting scaffolding surrounding it.

One major change is that the list of requests the tenant is servicing is no longer kept in memory. Instead, we rely on AiiDA's powerful querying capabilities to re-compute this in every iteration of the internal cycle of the tenant. This way the tenant can be restarted without losing track of any ongoing processes.

Another important difference is in the internal structuring of the tenant, where previously hard-coded sections that were specific to the task were removed from the core engine. Now the tenant just creates the input data nodes in the database with the full request, and it is the submitted process that will parse the data before starting to act on it. By keeping a mapping of the process

type that needs to be submitted for a given quantity and method (both keywords present in all requests), the AiiDA tenant just redirects the parameters and can therefore focus on the communication with the server and the tracking of ongoing calculations. This makes the tenant's capabilities very easy to extend, just by making a new AiiDA workflow and adding it to the previously mentioned mapping.



FINALES (11/2023) – Electrolyte Optimization for Maximum Conductivity and for Maximum Cycle Life

Bhowmik, Arghya[1]; Busk, Jonas[1]; Carlsson, Johan M.[2]; Castelli, Ivano E.[1]; Clark, Simon[3]; Flores, Eibar[3]; Flowers, Jackson[4, 5]; Fuchs, Stefan[4, 5]; Gaberšček, Miran[6]; Hajiyani, Hamidreza[2]; Hanke, Felix[7]; Liot, François[8]; Merker, Leon[4, 5]; Pizzi, Giovanni[8, 9]; Ramirez, Francisco Fernando[8]; Rieger, Laura Hannemose[1]; Sanin, Aleksei[4, 5]; Steensen, Simon[1]; Stein, Helge S.[4, 5]; Vegge, Tejs[1]; Vogler, Monika[4, 5]; Wölke, Christian[10]; Zhang, Bojing[4, 5]

Show affiliations

Description

This study investigates an electrolyte system composed of lithium hexafluorophosphate (LiPF6), ethylene carbonate (EC) and ethyl methyl carbonate (EMC). For the assembly of full cells, electrodes based on graphite and lithium nickel dioxide (LNO) are used. This work provides insight into the similarity of formulations of an electrolyte optimized for maximum conductivity and another one optimized for maximum cycle life are expected to be in this chemical system. The goal is to assess whether it is promising to target research efforts on finding an electrolyte formulation within this chemical space which can fulfil both requirements.

A campaign utilizing the latest version of FINALES is designed to determine conductivity values and predict end of life for various electrolyte formulations containing the aforementioned chemicals. The campaign comprises the following tenants grouped by their type of service:

- Physical tasks:
  - **ASAB**: This tenant automatically formulates requested electrolytes starting from stock solutions provided to it. It can also measure the ionic conductivity of the prepared formulation using a symmetric two-electrode cell using two stainless steel electrodes.
  - **AutoBASS**: This tenant automatically assembles coin cells using electrolyte formulations provided to it.
  - **Cycling**: This tenant performs the cycling of cells provided to it. It additionally provides a service to reserve cycling channels for cells prior to requesting their assembly to ensure channel availability and proper wetting times.
- Computational tasks:
  - **3DS**: This tenant performs molecular dynamics simulations and outputs the conductivity. Optional further outputs are the density and the radial distribution function for the requested formulation.
  - **DegradationModel**: This tenant predicts the end of life for a given capacity trajectory. It uses a threshold of 80 % of the initial capacity after three formation cycles as the definition for the end of life.
  - **Overlort**: This tenant manages the workflow of formulating electrolytes, assembling cells, cycling cells and predicting the end of life. Based on a request for a prediction of end of life or any other process step in between, it creates the necessary sub-requests to fulfil the task and merges them into one result which is subsequently reported to FINALES.
  - **OCond**: This tenant follows a Gaussian process for optimizing the electrolyte formulation for maximum conductivity. Based on the currently available data in the FINALES database, it suggests the most promising follow-up formulation targeted towards the maximization of the ionic conductivity.
  - **OEOL**: This tenant follows a gaussian process for optimizing the electrolyte formulation for maximum end of life of the full cells. Based on the currently available data in the FINALES database, it suggests the most promising follow-up formulation targeted towards the maximization of the end of life.
- Service tasks:
  - **Transportation**: This tenant allows for requests for transportation of physical samples. This is used for the transportation of electrolytes and battery cells between the devices connected to the tenants performing the physical tasks.

Figure 2. Record created in the BIG-MAP Archive by the archiving tenant on November 29, 2023. The record's title, the list of authors with their affiliations, and the description remain unchanged from the previous version.

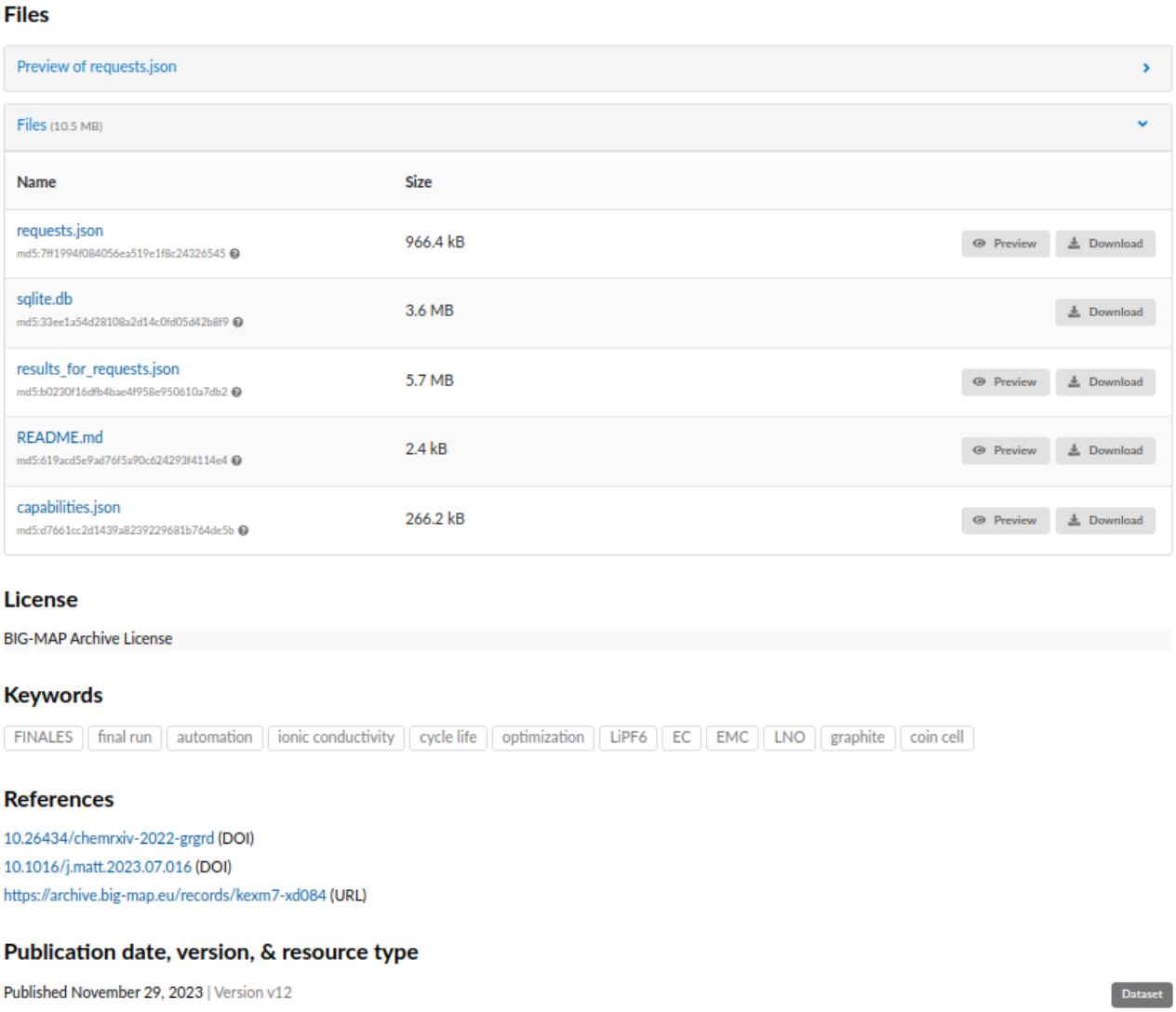


Figure 3. Record created in the BIG-MAP Archive by the archiving tenant on November 29, 2023. The `sqlite.db` file is a copy of the database file on the FINALES server. `capabilities.json`, `requests.json`, and `results_for_requests.json` contain data obtained from various API endpoints of the FINALES server. The content of these four files may change from one entry version to the next.

2. Demonstrations

FINALES was demonstrated in two demonstrations with different aims. Demonstration 1 was geared towards proving the design concept and functionality of the MAP. The Demonstration 2 aimed to showcase that the MAP is capable of optimising electrolytes and gain insights into composition-property relationships in a battery electrolyte system.

2.1. Demonstration 1

In this demonstration, the optimiser was configured to maximise the density and minimise the viscosity of battery electrolytes. The optimisation was conducted in the chemical space of electrolytes containing lithium hexafluorophosphate ($LiPF_6$), ethylene carbonate (EC), ethyl methyl carbonate (EMC) and dimethyl carbonate (DMC). The optimiser created suggestions for the next formulation to test by tuning the molecular fraction of each of these components.

The tenants involved were the ASAB tenant for formulating the electrolytes from the stock solutions 1 M $LiPF_6$ in EC:EMC 3:7 by weight, and 1 M $LiPF_6$ in EC:DMC 1:1 by weight, EC:EMC 3:7 by weight and DMC. The stock solutions were ordered from E-Lyte Innovations GmbH and used as received. For the prepared electrolyte formulations, the density and viscosity are automatically measured by ASAB and reported to FINALES. Furthermore, the molecular dynamics tenant was connected to provide computational results for density, heat capacity, ionic conductivity, diffusion coefficients and, transference numbers and radial distribution functions for the molecules contained in the formulations. Additionally, the optimiser tenant was integrated to guide the campaign by processing the data available in the database and suggesting the most promising follow-up formulation. Apart from the actively involved tenants, an additional tenant deployed a model reported by Rahmanian et al.[5] using the AiiDA[6] framework to provide low-fidelity estimates of electrolyte ionic conductivity. This tenant demonstrated that an AiiDA workflow can be included as a tenant into a MAP running through FINALES, but it was not involved in the optimisation task.

A more detailed description of Demonstration 1 including the results may be found in deliverable D9.6 *Demonstrator of a simulation run by an experiment, and an experiment run by a simulation*, which is currently in preparation and will be delivered to the European Commission by the end of January 2024.

2.1.1. Overview

Table 1 provides an overview of the key aspects of the first demonstration.

Table 1. Overview of Demonstration 1.

Optimisation task	Components of the electrolytes	Involved Tenants	Results	Future
Minimise density and maximise viscosity	EC, EMC, DMC, $LiPF_6$	• Molecular dynamics • ASAB • Optimiser	• A MAP can be operated using FINALES. • Improvements in the experimental setup are required. • A different optimisation target is needed for future demonstrations. • Data structures need to be interoperable and all tenants must stick to the use of the data structures. • Quality ratings are valuable for the client consuming the data to filter or process them as appropriate. • The communication of cost of a service, the feasible range for a method and the units, in which results are reported, to the whole MAP is crucial to collaborate efficiently. • Design decisions like having FINALES operate passively, asynchronous operation of the tenants, collecting and storing as much data as possible and not deleting data as well as most of the tenants being agnostic about the intention of requests proved beneficial.	• Expansion of data lineage tracking • Creation of more elaborate data structures • Improved user management • Communication of limitations within the MAP • Continued work towards multitenancy

2.2. Demonstration 2

The second demonstration of running an internationally distributed MAP using FINALES focused on the generation of electrochemical insights. Pursuing this goal, the optimiser was configured to find formulations, which maximise ionic conductivity using experimentally and computationally determined conductivity values. In a second task, another optimiser was requested to find electrolyte formulations, which achieve a long cycle life when used in coin cells. Both mutually independent tasks were run in parallel on the same instance of FINALES.

The tenants connected to the MAP for this demonstration were the Molecular dynamics tenant, ASAB, AutoBASS, Overlort, Cycler, Degradation model, Archiving, and the two instances of the Optimiser, OCond for the conductivity optimisation and OEOL for the EOL optimisation. The MAP was run using the chemical space spanned by $LiPF_6$, EC and EMC. The ASAB tenant used 1.5 M $LiPF_6$ in EC:EMC 1:1 by weight, 1.5 M $LiPF_6$ in EMC, EC:EMC 1:1 by weight and EMC as stock solutions all ordered from E-Lyte Innovations GmbH and used as received. The Optimiser was configured to disregard all results, which are marked as *deleted* or have a quality rating below 3.

2.2.1. Results

The MAP was run in the time from 26.09.2023 until 01.12.2023 in two parts and the second part is still ongoing as of 01.12.2023. The first part dealt with the conductivity task. The second part of the run used a new, empty database and the tenants related to the EOL task, namely the AutoBASS, Overlort, Cycler, Degradation model and the second Optimiser were connected to the MAP. The second part also included a restart of the conductivity task. The analysis of the obtained results is reported for each part of the run in the following sections.

First part

The first part focused on the conductivity task only and took place between 26.09.2023 and 28.10.2023. It generated 81 data points in total, out of which 43 originated from experiments and 38 from computations. These total numbers are reduced by nine data points, which were manually marked as deleted due to hardware malfunction and those, which were insufficient in quality and therefore not considered by the Optimiser (OCond). This results in 69 valid data points at the end of the first part of the run. Figure 3 presents the number of datapoints stored in the database after the end of the first part.

Figure 4 shows the trend of ionic conductivity versus the molality of $LiPF_6$. This trend indicates maximum conductivity around 1 m $LiPF_6$, which is in good agreement with the empirical model reported by Ding et al.[7]. The model and the values computed during the run of our MAP deviate increasingly for higher molality of the salt. A possible reason for this may be that such high molalities may not be reached in experiments and phase changes may occur. Furthermore, this graph shows a focus of the experimental data towards the lower end of the molalities, while the molecular dynamics method is rather exploited in the high molality range. Most likely, this is due to the limited chemical space, which can be covered by ASAB based on the selected stock solutions.

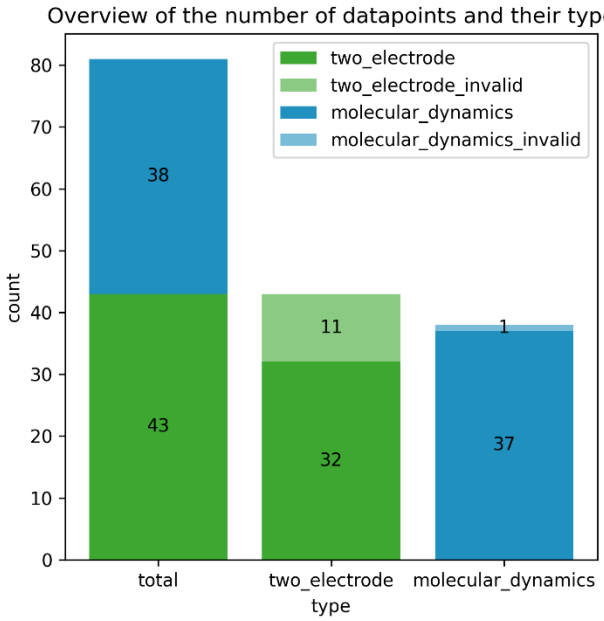


Figure 4. The number of data points in the database after the first part of the run – The data is grouped by the method and its validity according to the criteria of a successful finish of the method and a quality rating above 3. Deleted data points are also considered invalid.

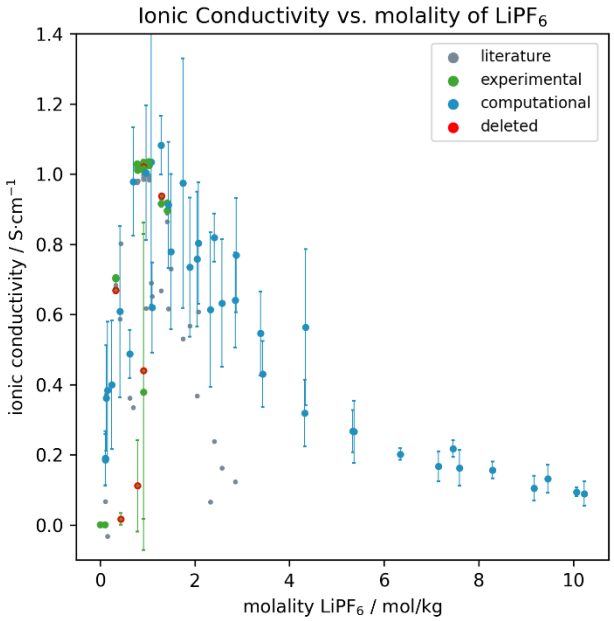


Figure 5. Ionic conductivity vs. molality of $LiPF_6$. A trend of the ionic conductivity with the molality of the LiPF6 is observed, which shows a good match with the one obtained from the empirical model by Ding et al.[7] except for high molalities.

The time series shown in Figure 5 reveals that the predictions for the experiment converged very soon during the run. The simulations started with a delay due to preparations, which were still in progress. However, once a computational datapoint was posted, the optimiser also requested formulations showing a different ionic conductivity from the experiments.

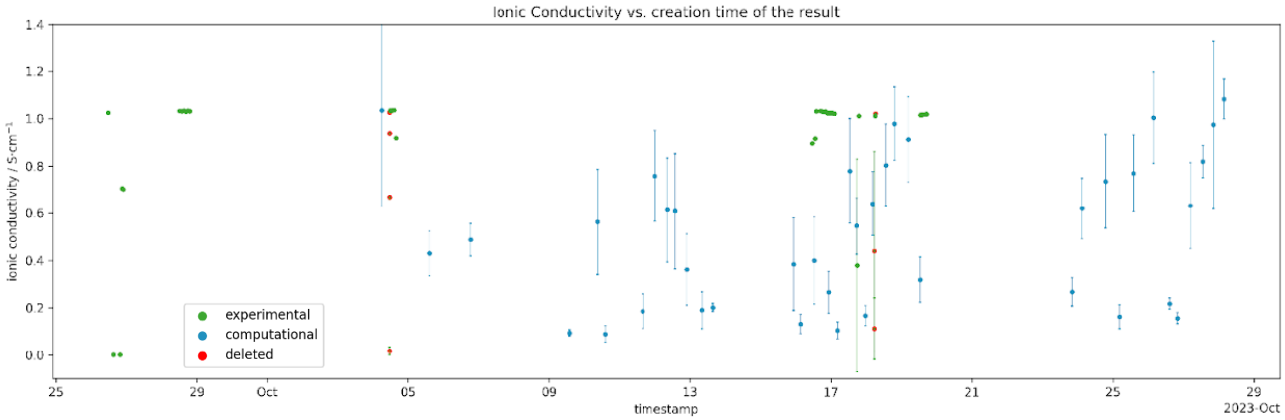


Figure 6. Time series of the conductivity data. Convergence in the experiment is visible for a conductivity around 1 S cm^{-1}. Newly incoming computational data affects subsequent requests to the experimental tenant.

Second part

The second part of the run started on 06.11.2023 and is still ongoing as of the 01.12.2023. The data presented in this section was obtained from FINALES on the 01.12.2023. In total 196 entries were created in the database during this time. These split into 75 entries related to the conductivity task and 45 entries for the EOL task up to now. Since the workflow to generate EOL data requires several entries in the database per data point, the aforementioned number of database entries relates to five data points relevant for the optimiser. The conductivity results are further composed of 44 experimental data points and 31 computational results. Applying the same filtering rules as in the first part, this leaves 29 valid data points from experiments and 31 from computations for conductivity and 4 valid results for the EOL task.

Conductivity task

Figure 6 shows the number of data points generated for the conductivity task during the demonstration. In total 75 data points were generated, out of which 60 are valid in the sense of the success and quality rating criteria.

As shown in Figure 7, the trend of the ionic conductivity with the molality of $LiPF_6$ as obtained from the results recorded in the FINALES database is in good agreement with the empirical model reported by Ding et al.[7]. The maximum conductivity is reached around 1 mol kg^{-1} of $LiPF_6$ for the data obtained from experiments, computational methods and the empirical model. Besides that, the same deviations between the empirical model and our data are observed as in the first part.

The time series of the data as presented in Figure 8 reveals that once the optimiser finds a promising formulation, it continues requesting similar formulations, which indicates a convergence of the model. It further shows that incoming new computational data points also cause the optimiser to request different formulations from the experiment. Therefore, the experimental tenant does not stick with close to optimal formulations for as long as it was observed in the first part.

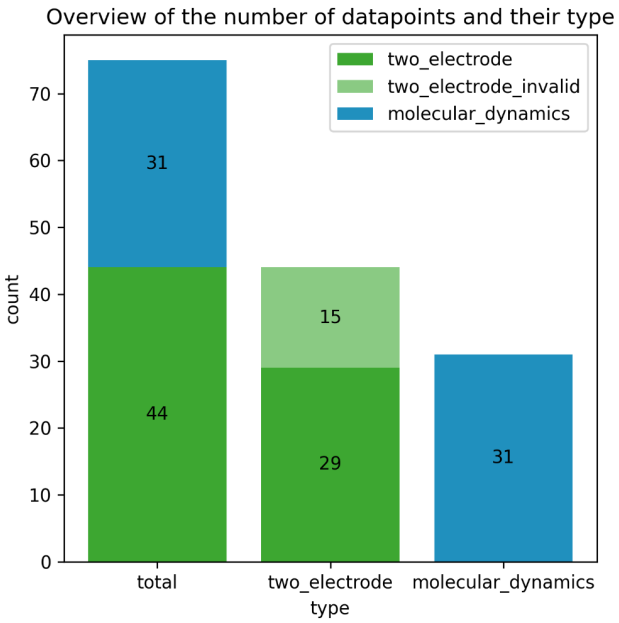


Figure 7. Number of datapoints for the conductivity task relevant for the optimiser. The data is grouped by the method, from which they originate and their validity according to the criteria of the success of the method and the threshold of 3 for the quality rating. Results marked as deleted are considered as invalid.

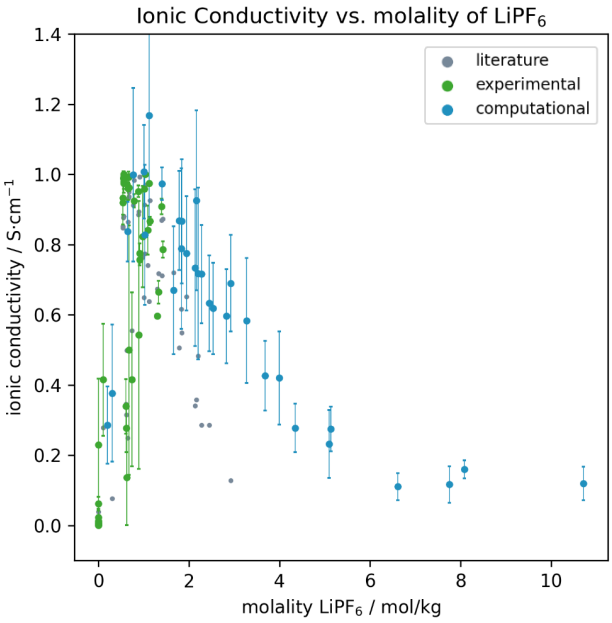


Figure 8. Ionic conductivity vs. molality of $LiPF_6$. The trend of ionic conductivity with the molality of $LiPF_6$ as observed from experiments, molecular dynamics and an empirical model by Ding et al.[7].

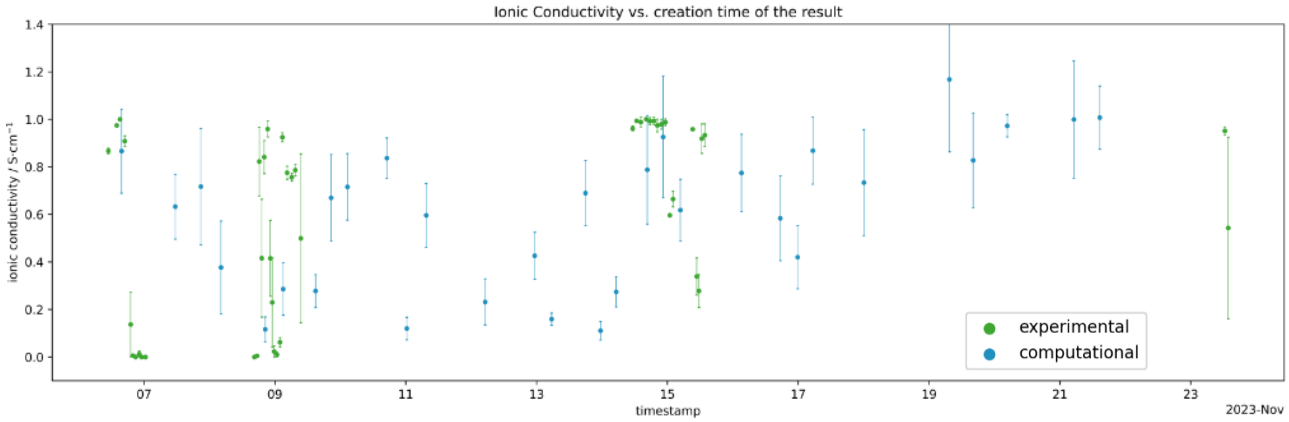


Figure 9. Time series of the conductivity data. Convergence in the experiment is visible for a conductivity around 1 S cm^{-1}. Newly incoming computational data affects subsequent requests to the experimental tenant**.**

EOL task

Figure 9 shows the valid data points per cell available for the EOL task in the database available on the 01.12.2023. Since the duration for a single iteration in this task is significantly longer than in the conductivity task due to the long duration of cycling experiments, there are only four valid batches resulting in eight data points each and one deleted result available. Furthermore, two of the valid batches are submitted manually to FINALES to include one formulation optimised for conductivity and two formulations, which were identified not to show optimal conductivity, into the database. The remaining two requests related to valid results were randomly generated by the optimiser.

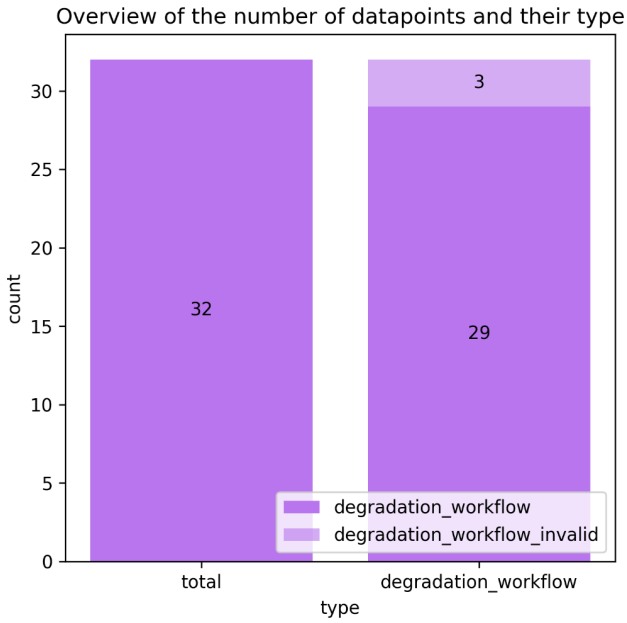


Figure 10. Number of datapoints for the conductivity task relevant for the optimiser. The data includes four valid batches with eight cells each. The invalid batch is not shown. Each data point counted into the histogram represents an individual cell.

As can be seen from Figure 10, the predicted EOL increases for higher concentration of $LiPF_6$ and the spread within the batches also increases. This may be due to the very short EOL for low $LiPF_6$ rarely allowing for a significant variation between the cells. It can further be observed that the predicted EOL for the conductivity optimised formulation at approximately 0.91 m $LiPF_6$ is higher than for the non-optimized formulation containing about 0.11 m $LiPF_6$. However, based on the available data, it cannot yet be determined whether the 0.91 m formulation is an optimum in EOL.

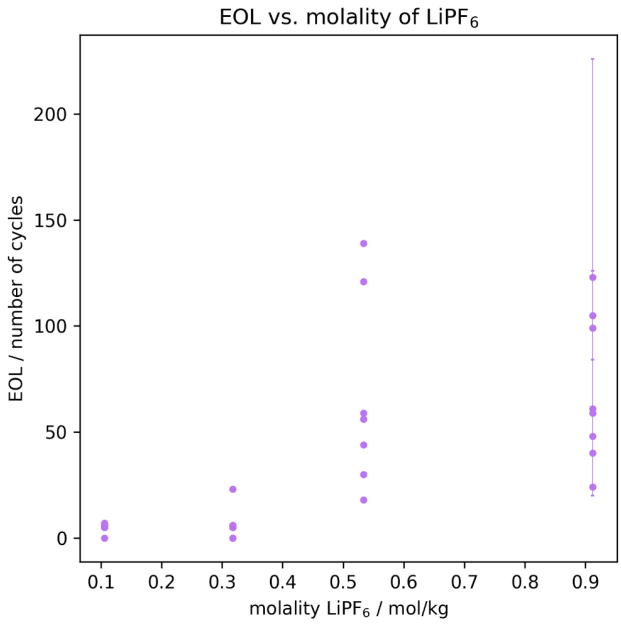


Figure 11. EOL data versus molality of $LiPF_6$. EOL increases for higher molality of $LiPF_6$ and the spread between the cells within a batch increases for increasing molality.

2.3. Data availability

The data generated during Demonstration 2 is published in the BIG-MAP Archive in the following entries:

- First part: https://archive.big-map.eu/records/kexm7-xd084
- Second part: https://archive.big-map.eu/records/j5wjm-q0r39

2.3.1. Overview

Table 2 provides an overview of the key aspects of the second demonstration.

Table 2. Overview of Demonstration 2.

Optimisation task	Components of the electrolytes	Involved Tenants	Results	Future
Maximise ionic conductivity	EC, EMC, $LiPF_6$	• Molecular dynamics • ASAB • Optimiser (OCond) • Archiving	• The uncertainty of simulation results is higher than for the experimental results for the contributions to the experimental error considered in the analysis. • The trend of conductivity with $LiPF_6$ concentration as reported in the literature[7] can be observed in the data generated by the MAP.	• Applying the FINALES MAP concept to novel research questions and more complex optimisation tasks • Running larger campaigns on larger MAPs
Maximise cycle life		• Molecular dynamics • Overlort • Cycler • ASAB • AutoBASS • Degradation model • Optimiser (OEOL)	• The cycle life increases with higher molality of $LiPF_6$. • The formulation optimised for conductivity seems to show higher EOL than the non-conductivity-optimised formulation.	

3. Conclusion

We herein demonstrate the worldwide first MAP that generated scientific insights beyond mere single tasks optimisation and for a complex task involving a multitude of tenants. Naturally this MAP run demonstrated FINALES operability in a passive mode of operation. Furthermore, it was shown that a European-wide distributed MAP can generate datasets to investigate battery related research questions. It was specifically demonstrated that a known trend between the conductivity and the concentration of the salt in the electrolyte solution was re-discovered. Furthermore, investigations involving the use of an optimised result in a second optimisation task was demonstrated in the EOL task, which is still ongoing as of 01.12.2023. The currently available data suggest that the electrolyte optimised for conductivity seems to result in a higher EOL than the non-optimised formulation. The demonstrated tasks covered research questions from the bulk properties of the electrolyte to the system level of a coin cell. The second demonstration showed FINALES being able to conduct continent wide orchestration of complex research workflows that even allow tests of optimal components' influence onto systems.

4. Outlook

For future development, the application of FINALES-based MAPs for the investigation of novel battery related research questions is intended. The further development may include improvements in the management of users and research campaigns in the system. This could allow control over the access to certain data.
Moreover, the deployment of FINALES in larger MAPs comprising a larger number and variety of tenants is envisaged.

5. References

1 Rahmanian, F. Modular and Autonomous Data Analysis Platform (MADAP). (2023) doi:10.5281/ZENODO.7374382.
2 Severson, K. A. *et al.* Data-driven prediction of battery cycle life before capacity degradation. *Nat Energy* **4**, 383–391 (2019).
3 BIG-MAP Archive. https://archive.big-map.eu/.
4 Liot, F. big-map-archive-api-client. (2023).
5 Rahmanian, F. *et al.* Enabling Modular Autonomous Feedback-Loops in Materials Science through Hierarchical Experimental Laboratory Automation and Orchestration. *Adv Materials Inter* **9**, 2101987 (2022).
6 Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science* **111**, 218–230 (2016).

7 Ding, M. S. *et al.* Change of Conductivity with Salt Content, Solvent Composition, and Temperature for Electrolytes of LiPF[sub 6] in Ethylene Carbonate-Ethyl Methyl Carbonate. *J. Electrochem. Soc.* **148**, A1196 (2001).