Ref. Ares(2024)6192428 - 02/09/2024

Horizon Europe Work Programme

BIG LEAP

Next Generation of Battery Management Systems to increase Interoperability, bridge the Gap between 1st and SL-BESS, Extend Adaptability and emPower battery value chains

D3.1 - Methodologies for SOX, Self-Diagnosis and RUL algorithms development and data collection

Lead Contractor: IKERLAN

Project Coordinator: Image Worighi

Authors and organisation(s): Josu Olmos (IKERLAN), Markel Azkue (IKERLAN), Carmen Leticia Castrejón Barrón (BRING), Pankaj Saha (VTT), Prashant Singh (VTT), Joel Wooden (BFH), Priscilla Caliandro (BFH), Stefan Waldhoer (FHG), Marco Rodrigues (INEGI) and Rui Abreu (INEGI)

Date: 27/08/2024

Doc. Version: 2

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Co-funded by the European Union under grant agreement 101137815. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

PROJECT DETAILS			
Project name	Next Generation of Battery Management Systems to increase Interoperability, bridge the Gap between 1st and SL-BESS, Extend Adaptability and emPower battery value chains		
Start	01/01/2024	**Duration**	42M
Project acronym	BIG LEAP	**GA number**	101137815
Topic identifier	HORIZON-CL5-2023-D2-01-04	**Call identifier**	HORIZON-CL5-2023-D2-01
Type of Action	Horizon IA	**Coordinator**	BRING
Project Coordinator	Imane Worighi	Imane.worighi@bringvzw.be	
Project Manager	Costantino Laureanti	costantino.laureanti@bringvzw.be	
Website	https://bigleapproject.eu/		

DELIVERABLE DETAILS			
Number	3.1		
Title	Methodologies for SOX, Self-Diagnosis and RUL algorithms development and data collection		
Short description	This deliverable focuses on the definition of the specifications and methodologies for the SOX, RUL and self-diagnosis algorithms developed within Big Leap project. The data collection process necessary for the development of the algorithms is also presented		
Work Package	WP3	**Task**	T3.1
Dissemination level	PU - Public		
Due date (M)	M8	**Submission date (M)**	M8
Deliverable responsible	Josu Olmos (IKERLAN)	**Contact**	jolmos@ikerlan.es

DELIVERABLE CONTRIBUTORS			
	Name	**Organisation**	**E-mail**
Deliverable Author	Josu Olmos	IKERLAN	jolmos@ikerlan.es
Contributing Author(s)	Markel Azkue Pankaj Saha Prashant Singh Joel Wooden Priscilla Caliandro Stefan Waldhoer Marco Rodrigues Rui Abreu Carmen Leticia Castrejón Barrón	IKERLAN VTT VTT BFH BFH FHG INEGI INEGI BRING	mazkue@ikerlan.es pankaj.saha@vtt.fi prashant.singh@vtt.fi joel.wooden@bfh.ch priscilla.caliandro@bfh.ch stefan.waldhoer@iisb.fraunhofer.de mrodigues@inegi.up.pt rabreu@inegi.up.pt carmen.barron@bringvzw.be
Internal Reviewer(s)	Joel Wooden Priscilla Caliandro	BFH BFH	joel.wooden@bfh.ch priscilla.caliandro@bfh.ch
Internal Reviewer(s)	Carmen Leticia Castrejón Barrón	BRING	carmen.barron@bringvzw.be
External Reviewer(s)	Oleksij Chumak	EATON	oleksijchumak@eaton.com
External Reviewer(s)	Tata Shanmukha	TATA POWER	shanmukha.tata@tatapower.com
Final review and quality approval	Imane Worighi	BRING	Imane.worighi@bringvzw.be

DOCUMENT HISTORY			
Date	**Version**	**Name**	**Changes**
27/03/2024	0.1	20240327_BIGLEAP_D3.1_v0.1	First version of document
20/05/2024	0.2	20240520_BIGLEAP_D3.1_v0.2	Document adapted to template v4
30/05/2024	0.3	20240530_BIGLEAP_D3.1_v0.3	Document adapted to template v5
25/07/2024	1.0	20240725_BIGLEAP_D3.1_v1.0	Document finished and internal review completed
23/08/2024	2.0	20240823_BIGLEAP_D3.1_v2.0	External review completed

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Executive Summary

The BIG LEAP project focuses on developing solutions for the Second Life Batteries (SLBs) Battery Management System (BMS) and its reconfiguration process. Technology breakthroughs will be made in its BMS, as a new three-layer architecture will be designed to ensure interoperability, safety, and reliability. It will be complemented with an adaptable Energy Storage System (ESS) design to ensure BMS integration and expand the SLB's potential applications. Additionally, the BIG LEAP project intends to optimize the battery reconfiguration process by making it cost-effective, faster, and standardized.

The methodology for the development of these innovations includes the collection of Electric Vehicle (EV), maritime E-Vessel, and ESS batteries that will be dismantled, and the data collected will serve as the basis for the BMS architecture development. It will contain adaptable State-of-X (SOX) algorithms for accurate battery measurement, a Digital Twin (DT) for real-time monitoring, and a standardization roadmap. The new BMS will be integrated into the batteries, alongside the ESS and will be tested in three demo sites. Two physical demos will be in Paris and Prague, and a virtual demo will be in Morocco. They aim to validate the novel BMS and ESS, proving their optimization and interoperability.

This document is the BIG LEAP deliverable No. 3.1, being the first deliverable of WP3. This work package is focused on the development of adaptive SOX and Remaining Useful Life (RUL) estimators, enhanced battery models and the cloud-based software layer. In this context, D3.1 serves as a baseline for the development of the mentioned algorithms and breaks ground for the main tasks to be developed within WP3. Specifically, this deliverable reports the specific methodologies and specifications for each of the SOX and RUL algorithms to be developed. Additionally, data collection is addressed and guidelines for data standardization are given. The data collection is divided into the identification of open-access databases and the collection of data from battery OEMs participating in BIG LEAP.

Acronyms and abbreviations

A	Ampere
Ah	Ampere-hour
APE	Absolute Percentage Error
BIG LEAP	Next Generation of Battery Management Systems to increase Interoperability, bridge the Gap between 1st and SL-BESS, Extend Adaptability and Empower battery value chains.
BFH	Berner Fachhochschule (Associated partner n°15 - BIG Leap project)
BMS	Battery Management System
CA	Consortium Agreement
CNN	Convolutional Neural Network
CPU	Central processing unit
DL	Deliverable
DoD	Depth of Discharge
DT	Digital Twin
ECM	Equivalent Circuit Model
EIS	Electrochemical Impedance Spectroscopy
EOL	End of Life
EO1L	End of First Life
EO2L	End of Second Life
ESS	Energy Storage System
EUCAR	European Council for Automotive
EV	Electric Vehicles
FHG	Fraunhofer (Beneficiary partner n°2 - BIG LEAP project)
HW	Hardware
Hz	Hertz
IP	Intellectual Property
IPR	Intellectual Property Rights
KPIs	Key Performance Indicators
LAB	Laboratory
LCO	Lithium Cobalt Oxide
LFP	Lithium Iron Phosphate
LMO	Lithium ion manganese oxide
LSTM	Long short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
NCA	Nickel Cobalt-Aluminium Oxide
NN	Neural Network
NMC	Nickel Manganese Cobalt
OCV	Open-Circuit Voltage
OEM	Original Equipment Manufacturer
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
RU1L	Remaining Useful First Life
RU2L	Remaining Useful Second Life
RUL	Remaining Useful Life
SDF	Statistical Data Frame
SLB	Second Life Battery
SOA	Safety Operation Area
SOC	State of Charge
SOE	State of Energy
SOH	State of Health
SOP	State of Power
SOR	State of Resistance
SOS	State of Safety
SOX	State of X
SW	Software
SYN	Synthetic
TF	Transfer Learning
VPN	Virtual Private Network
WP	Work package

Introduction

In the rapidly evolving field of ESSs and lithium-ion batteries, the effective management of these systems is understood as fundamental to ensure their longevity, reliability, and performance. Central to this management are the State of X (SOX) and Remaining Useful Life (RUL) estimators. These algorithms are essential for accurately monitoring and predicting the health and performance of batteries, which in turn can significantly enhance their operational efficiency and lifecycle. For the SLB applications currently under study in BIG LEAP project, this becomes even more crucial, as their longevity and reliability might be even more uncertain due to their previous use in a first life application.

In order to tackle this issue, WP3 of BIG LEAP project focuses on the development of adaptive SOX and RUL estimators, enhanced battery models, and the cloud-based software layer. On the one hand, SOX algorithms include a range of state estimations such as State of Charge (SOC), State of Energy (SOE), State of Health (SOH), State of Safety (SOS) and State of Power (SOP). These estimations provide comprehensive insights into various aspects of battery performance and condition: SOC indicates the current charge level of a battery, SOH assesses its overall health and capacity, SOS specifies its safety level, and SOP determines its ability to deliver power under specific conditions. On the other hand, RUL algorithms predict the remaining operational lifespan of a battery before it requires replacement or significant maintenance. These predictions are vital for planning maintenance schedules, reducing downtime, and minimizing operational costs.

The mentioned SOX and RUL estimators, the battery models and the cloud layer are key elements of the interoperable BMS developed as one of the main outcomes of BIG LEAP. According to the methodology proposed in the project, the first step prior to the development of the algorithms and their integration into the different BMS layers is the definition of their specifications and the collection of data for their development. This is addressed in Task 3.1 of the project.

The development and effectiveness of SOX and RUL algorithms heavily depend on the quality and relevance of the data used for this purpose. Appropriate data collection and standardization are fundamental to create robust models that can reliably monitor and predict battery states. This involves gathering extensive data from various battery types, operating conditions, and usage scenarios. The data must be meticulously processed and analysed to identify key patterns and correlations that inform algorithm development.

A typical approach in battery engineering consists of deploying extensive laboratory tests to generate this mentioned data. However, this is a time-consuming approach that requires several months (or even years) to obtain a sufficiently populated database. Due to this reason, BIG LEAP project proposes the use of open access data and the operational data provided by some of the partners to develop the SOX and RUL algorithms and the battery models. This deliverable explains the process carried out to gather this data, analyse it and select the most appropriate databases for the development of the algorithms and models,

In short, deliverable D3.1 addresses all the activities carried out in Task 3.1 of BIG LEAP project. First of all, Section 1 defines the specifications of the SOX and RUL algorithms to be developed within WP3. This includes the definition of the estimator itself, the methodology in which the algorithm will be based (this will be developed in T3.2 of the project), the characteristics of the data required to develop the algorithm, and the criteria that will be used to evaluate the effectiveness of the estimators. Then, Section 2 addresses the data compilation activity. As previously mentioned, this step is divided into the identification of open access databases, and the provision of operational data from the project partners. This section explains the process carried out to identify the most appropriate data for the development of each of the algorithms. In the next step, Section 3 defines the guidelines for the standardization of this data. Finally, Section 4 includes the methodology that will be used in the following tasks of WP3 to integrate the data into the SUNDIAL platform, which will be a key step for the development of the cloud layer of BIG LEAP project.

1. Specifications of SOX/RUL algorithms

This section defines the specifications of all the SOX (SOC, SOH, SOP and SOS) and RUL algorithms to be developed in BIG LEAP project. For each algorithm, the following information is defined:

- The definition of the estimated variable. For instance, in the case of SOC, what does "State of Charge" mean.
- The methodology that will be used to develop the algorithm. In any case, note that the algorithms will be developed in T3.2 and the detailed methodology and validation will be extensively explained in Deliverables D3.2 and D3.3.
- The characteristics of the data required to estimate the variable. The data requirements will be then considered in Section 2 of this deliverable for the definition of the most appropriate data batches for each algorithm or estimator.
- The evaluation criteria: which key performance indicators (KPIs) will be defined for each algorithm.

1.1. State of Charge (SOC)

1.1.1. SOC definition

The SOC indicates the remaining battery electric charge in a battery at a specific time instant t ($C(t)$) compared to a fully charged battery (C_{full}). In other words, it shows the remaining autonomy of the system until it is completely discharged. As can be seen in the following equation, the SOC shows the ratio between the remaining charge and the capacity of the battery when it is fully charged, both in Ah.

$$SoC(t) = \frac{\mathrm{C}(t)}{C_{full}} \cdot 100$$

Different variables such as battery temperature, charge and discharge current and battery SOH influence battery capacity, so it is necessary to develop estimators that take all these factors into account.

When creating a SOC estimator, it is important to keep in mind that the actual measurement will have noise and error. The SOC algorithm must be able to take this error into account and be as robust as possible in order to obtain the most accurate estimation and be affected as little as possible by this error and noise.

The developed algorithm should be tested at different temperatures, charging, and discharging currents and SOH in order to properly estimate the SOC under different conditions.

1.1.2. Methodology for SOC estimation

The development of the SOC algorithm will be based on an iterative, data-driven approach. Initially, data will be collected (which will include data from public databases as well as real application data provided by the different partners), obtaining measurements of battery voltage, current and temperature under various charging and discharging conditions, as well as collecting historical data of charging and discharging cycles. Once the data is obtained, preprocessing will be performed including cleaning and filtering to remove outliers and noise, as well as normalization to ensure consistency in the measurements.

In the model development stage, appropriate modelling techniques will be selected, including machine learning techniques, such as neural networks specific to deal with timeseries data. The model will be trained using the public datasets and validated by cross-validation to assess its accuracy and robustness.

Subsequently, the model will be optimized using the real application data, adjusting parameters, and retraining the model to improve the accuracy of the SOC for each specific cell type.

1.1.3. Requirements of data for SOC estimation

For the development and testing of the SOC algorithm, the data must meet certain characteristics. First, they must be accurate, with precision voltage and current measurements and temperature measurement accuracy. The temporal resolution of the data is also crucial, so they must be recorded at consistent time intervals, such as every second. In addition, the data must be representative, covering a wide range of operating conditions, including different states of charge, discharge, temperatures, and usage cycles. It is essential to have a sufficient volume of data to cover all these scenarios to ensure the robustness of the algorithm.

Table 1 reviews the data requirements for SOC estimation. It is worth noting that the current resolution is defined with the relative unit C-rate, which refers to the charged/discharged current in relation to the battery capacity. This unit is considered to be more appropriate than Ampere (A), due to the variance in capacity between different batteries.

Table 1. SOC data requirements.

SOC data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.005 V (cell level)	1 Hz	1ºC	0.1 Hz	0.01 C	1 Hz	N/A

1.1.4. Evaluation criteria

To evaluate the performance of the algorithm, different key performance indicators (KPIs) will be used. The accuracy of the SOC will be one of the main indicators, evaluated by the Mean Absolute Percentage Error (MAPE) in the SOC estimation, which should be less than 3% according to the project KPI matrix (see Deliverable D1.3).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

being n the number of evaluation points, i the current evaluation point, Y_i the actual value, and $\hat{Y}_i$ the predicted value.

In order to account for large deviations, the Root Mean Square Error (RMSE) will also be evaluated, defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

Additionally, the maximum error will be also used as an evaluation criterion, providing insight into the worst-case performance of the algorithm:

$$Max.Error = max\left|Y_i - \hat{Y}_i\right|$$

The robustness of the algorithm will also be evaluated, considering its performance under extreme temperature and load conditions, as well as its ability to handle noisy or incomplete data. Computational efficiency is another important KPI, measured by the processing time per SOC estimation and the use of resources such as CPU and memory.

1.2. State of Energy (SOE)

1.2.1. SOE definition

The SOE is a similar state compared to the SOC, but in this case the SOE shows the remaining energy at a given instant ($Energy_{remaining}(t)$) relative to the dischargeable energy when the cell is fully charged ($Energy_{full}$). That can be expressed as follows:

$$SoE(t) = \frac{Energy_{remaining}(t)}{Energy_{full}} \cdot 100$$

The advantage of using this state over the SOC is that it shows the energy state in which the applications are dependent, as the battery voltage varies across the SOC.

Similar to the SOC, the SOE estimations vary also depending on variables such as battery voltage, current or SOH.

1.2.2. Methodology, requirements of data and evaluation criteria for SOE

Both the SOC and SOE algorithms will utilize the same neural network architecture, which will process identical input data to generate two distinct outputs: the SOC and the SOE of the battery. Consequently, the methodology and data requirements for both algorithms are identical and are detailed comprehensively in Section 1.1.2 and Table 2. Similarly, the KPIs used to evaluate the performance of the SOE algorithm will be also the ones specified in Section 1.1.4.

In this context, Table 2 reviews the data requirements for SOE estimation.

Table 2. SOE data requirements.

SOE data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.005 V (cell level)	1 Hz	1ºC	0.1 Hz	0.01 C	1 Hz	N/A

1.3. State of Health (SOH)

1.3.1. SOH Definition

The SOH represents the health condition of the battery as it ages. The most common way to quantify SOH is by comparing the currently usable discharge (diminished) capacity with the nominal capacity of the new battery:

$$SOH = \frac{C_{effective}}{C_{nominal}}$$

Typically, this value is given as a percentage. Another parameter to quantify the SOH is the State of Resistance (SOR) of the battery:

$$SOR = \frac{R_{effective}}{R_{nominal}}$$

As a battery ages, its resistance generally increases, leading to higher I^2R joule losses. The diminished usable capacity and increased losses due to resistance have a compounding negative effect on the energy available from the battery.

Since both discharge capacity and internal resistance vary with operating conditions, it is important to consider values taken under the same reference conditions for accurate SOH (State of Health) and SOR (State of Resistance) definition. For Big Leap, two separate approaches will be developed by two consortium partners, Berner Fachhochschule (BFH) and Fraunhofer (FHG).

1.3.1. Methodology for SOH Estimation (BFH)

The implementation of this algorithm will be based on a Machine Learning (ML) algorithm applied to the BFH Statistical Data Frame (SDF). In this approach, the historical usage data of the battery is stored as histograms as opposed to a time series. The advantage is that the size of the data frame doesn't change throughout the lifetime of the battery. The histograms capture how much time the battery spent in certain operating conditions, for example how much time the battery spent in a certain temperature, C-rate, or SOC range. The Histograms can also capture multiple dimensions at once, for example how much time the battery spent in a certain C-rate range while in a certain SOC range.

The ageing of a battery is heavily dependent on the conditions it was operated in during its lifetime. The ML algorithm will be trained on SDFs generated from open data sets that reflect these various conditions. BFH approach involves Random Forest regression methods as it is robust to new data and to outliers and more importantly, it allows for the estimation of features importance.

1.3.2. Data Requirements for SOH Estimation (BFH)

The Data for SOH estimation and its validation must be precise and have a high enough sampling rate to capture dynamic processes, such as 1Hz for the voltage and current, and 0.1Hz for the cell temperature. The development data must be representative of commonly used cell chemistries and cover a variety of operating conditions and ageing patterns.

Table 3. SOH Data Requirements

SOH data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.001 V (cell level)	1 Hz	0.1ºC	0.1 Hz	0.1 C	1 Hz	N/A

1.3.3. Methodology for SOH Estimation (FHG)

The general approach for the AI-based SOH estimation algorithm is to combine the neural network (NN) architectures incorporating the properties of the long short-term memory (LSTM) and convolutional neural networks (CNN). Depending on the performance and accuracy attention techniques may be used to improve the estimation further. To create a generalized SOH model transfer learning (TF) strategies will be combined with a smart hyperparameter tuning approach. This will enable a generalized, accurate SOH estimation for different cell chemistries and for different battery cell applications (entirely usage and load profiles) different. This approach should enable to widely eliminate the influence of the measurement noise on the SOH estimation accuracy.

1.3.4. Data Requirements for SOH Estimation (FHG)

Current, cell voltage, and temperature are required as data input for the SOH estimation. As higher currents, even if they are only shortly applied, have a large impact on the SOH directly, and indirectly due to the temperature rise, the current needs to be sampled at a relatively high frequency. The temperature as the other main driver of aging must sampled accurately, but due to the large mass of the cell and therefore the slow thermal response a lower sampling rate is sufficient. The data requirements for cell voltage are state-of-the-art.

Table 4. SOH Data Requirements

SOH data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.001 V (cell level)	10 Hz	0.1ºC	0.1 Hz	0.01	50 Hz	N/A

1.3.5. Evaluation Criteria

The primary key performance indicator (KPI) of the SOH algorithm, is that it should achieve a Mean Absolute Percentage Error (MAPE) of less than 3%.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|$$

As in the SOC and SOE cases, the RMSE will be also used as evaluation criteria, in order to account for large deviations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2}$$

Finally, the algorithm's performance should be assessed based on its estimation efficiency, meaning it should quickly and effectively process data to produce health evaluations with minimal computational overhead.

1.4. State of Power (SOP)

1.4.1. SOP Definition

The SOP represents the power limits of a battery at any given time and operating condition. This algorithm must provide an estimation of the maximum charge and discharge power the battery can sustain over a certain period without causing premature degradation. There are many different approaches to calculating and describing power limits, but they all commonly share the fundamental principles of

1. Ensuring the terminal voltages of all cells in the battery system remain within safe operating bounds.
2. Maintaining system design limits on power and current
3. Optimizing the inherent trade-off between battery pack performance and degradation.

To avoid rapidly changing system power limits, the charge and discharge power limits $p_{max,ch}$ and $p_{max,dch}$ should be predicted up to a future time horizon of ΔT seconds. Within this predicted timespan, the battery must be able to sustain the estimated power limits without exceeding system design limits. These power limits are continuously re-estimated before the time horizon ΔT is reached, creating a moving estimation window that is updated with a frequency $f_{update} > \frac{1}{\Delta T}$.

1.4.2. Methodology for SOP Estimation

The SOP will be determined using an Equivalent Circuit Model (ECM), whose parameters will be determined by an on-board impedance analysis developed in task 2.2. Batteries are not ideal voltage sources, as their OCV is dependent on State of Charge (SOC), and they have an internal resistance.

The ECM that will be used for the SOP algorithm is the polarization model. This model features parallel RC-elements that model time-dependent polarization effects (activation and diffusion) in the battery.

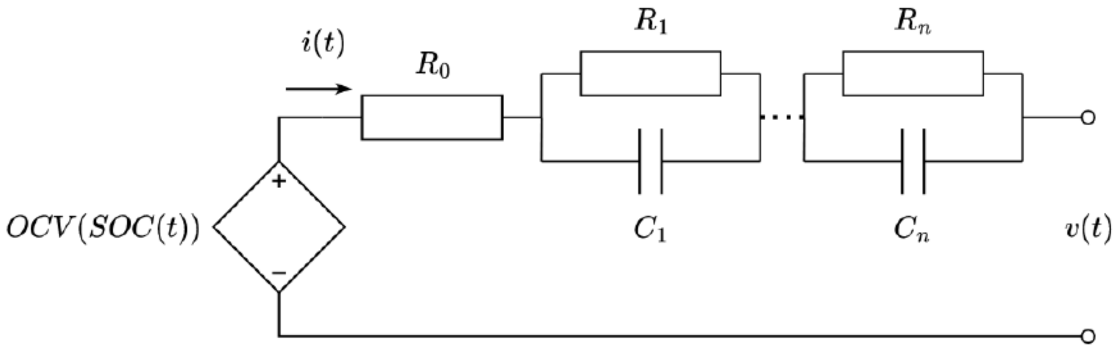

Figure 1. n^{th} Order ECM

The number of parallel RC-elements used in the model depends on the fidelity of the impedance measurement developed in task 2.2 and thus remains to be determined.

The power limits $p_{max,ch}$ and $p_{max,dch}$ for a time horizon ΔT will be determined by calculating the instantaneous Δv due to the R_0 series resistance as well as the time-dependant Δv due to the polarization effects based on the ECM. The sum of Δv after ΔT must not lower the cell terminal voltage below v_{min} or raise it above v_{max}.

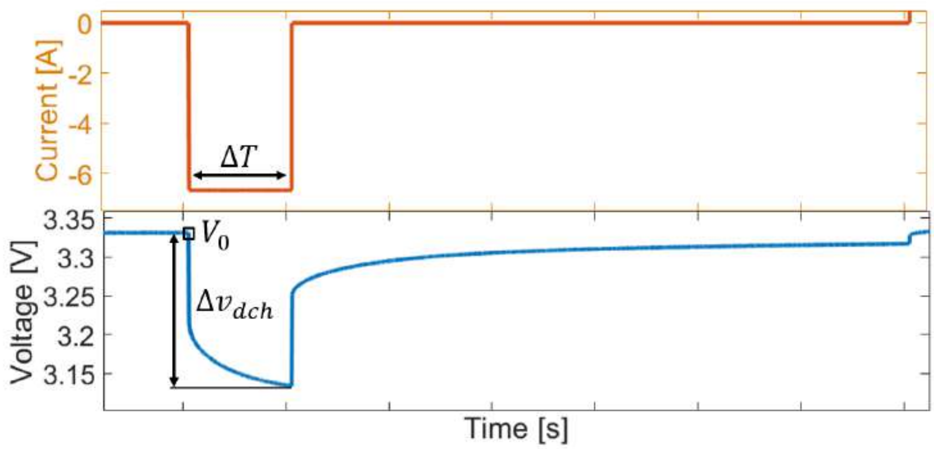

Figure 2. Voltage Response to a Current Pulse

1.4.3. Data Requirements for SOP Estimation

The data for the SOP estimation must be precise and have a very high sampling rate such as 10Hz in order capture highly dynamic processes. It is crucial to be able to correlate fast variations in cell current to the cell's voltage response. The development data for the algorithm must include pulse tests as well as Electrochemical Impedance Spectroscopy (EIS) tests to compare the results.

Table 5. SOP Data Requirements

SOP data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.001 V (cell level)	10 Hz	0.1ºC	0.1 Hz	0.01 C	10 Hz	N/A

1.4.4. Evaluation Criteria

The primary key performance indicator (KPI) of the SOP algorithm, is that it should achieve a Mean Absolute Percentage Error (MAPE) of less than 3%.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|$$

The algorithm must effectively provide an accurate estimate of the available system power for a given estimation window in near real-time, while maintaining low computational overhead.

1.5. State of Safety (SOS)

1.5.1. SOS Definition

SOS represents how safe a battery is. While the general concept of safety is intuitive, numerically quantifying it is an imprecise task often subject to interpretation. Currently, there is no unified standard for defining and classifying SOS, leading to multiple ways of indicating a battery's safety.

One approach is to define a safety operation area (SOA), within which the battery can operate with very low risk. The SOA specifies that the battery's voltage and temperature must remain within certain values, which differ for each battery and must be provided by the manufacturer. Figure 3 shows a common diagram of the SOA for a battery.

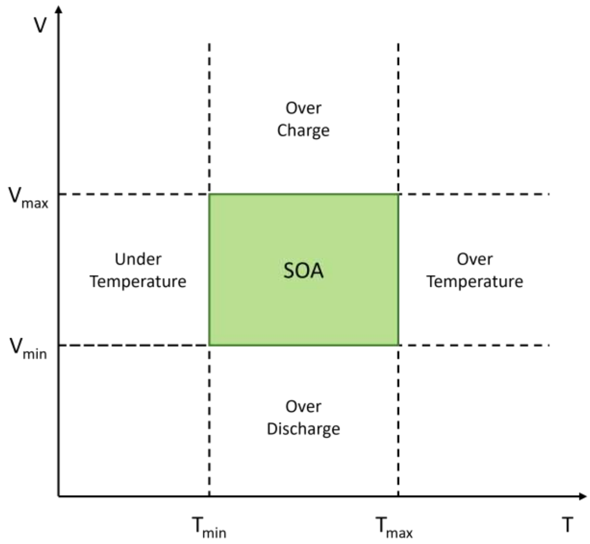

Figure 3. SOA for a battery that depicts different states during the charging procedure [1].

The European Council for Automotive R&D (EUCAR) uses a hazard level description to qualitatively categorize battery hazard states into eight levels, from 0 to 7, with increasing severity [2].

A common quantitative method to define hazard levels is based on hazard risk, where the lower the risk, the safer the battery. Hazard risk (H_R) is defined as the product of hazard severity (H_S) and the likelihood of occurrence (H_L) [3]:

$$\mathrm{H_R} = \mathrm{H_S} \cdot \mathrm{H_L}$$

Another way to define SOS is as the reciprocal of the probability of abuse [4]:

$$\mathrm{f_{SOS}(x)} = \frac{1}{\mathrm{f_{abuse}(x)}}$$

where $f_{SOS}(x)$and $f_{abuse}(x)$ are the safety and abuse functions respectively, and x represents all types of state and control variables that describe the behavior of the battery, for example voltage, temperature, current, etc.

Based on these interpretations, SOS can be defined as a dynamic and quantitative measure that integrates multiple factors affecting battery safety to provide real-time assessments. It evaluates the probability of safe operation under varying conditions by considering factors such as voltage, temperature, current, mechanical deformation, SOC, SOH, internal impedance, and other abuse conditions.

1.5.2. Methodology for SOS Estimation

To estimate the SOS algorithm, a generic theoretical framework must be established. Assuming SOS is the reciprocal of the abuse function, the final equation is derived from this basis. Here is a summary, with further details available in [4].

To keep $\boldsymbol{SOS}$ within a workable range, it is set between 0 and 1, like SOC:

$$\boldsymbol{f_{sos}(x) = \frac{1}{g(x) + 1}}$$

where $\boldsymbol{g(x)}$ is the abuse function for values ≥ 0. Preferably, $\boldsymbol{g(x)}$ is quadratic, defined by parameters $\boldsymbol{m}$ and $\boldsymbol{d}$:

$$\boldsymbol{f_{sos}(x) = \frac{1}{m[h(x) + d]^2 + 1}}$$

being $\boldsymbol{h(x)}$ the adapted abuse function.

The values of $\boldsymbol{m}$ and $\boldsymbol{d}$ can be derived by defining two points. The resulting equation is:

$$\boldsymbol{f_{sos}(x) = \frac{1}{0.25[\frac{x - x_{100}}{x_{\zeta} - x_{100}}]^2 + 1}}$$

This equation allows the derivation of a safety expression for any variable x by defining the safety and compromised safety states.

As a probability function, $\boldsymbol{SOS}$ is the product of all safety subfunctions:

$$SOS = f_1(x_1) \cdot f_2(x_2) \cdot \cdots \cdot f_n(x_n)$$

Each subfunction has a minimum value ζ to ensure safe behavior and a maximum of 1, with key values being:

- $SOS = 1\ Completely\ safe\ (all\ functions\ are\ 1)$
- $SOS = \zeta\ warning\ (one\ function\ may\ be\ at\ \zeta)$
- $SOS = \zeta^n\ \ minimum\ (all\ functions\ are\ \zeta)$
- $SOS < \ \zeta^n\ unsafe\ (all\ functions\ are\ below\ \zeta)$

The SOS algorithm will be developed using a data-driven approach. During the model development phase, suitable modeling techniques, including machine learning methods, will be chosen. The model will be trained with the public datasets mentioned in Section 2 and validated to evaluate its accuracy and robustness. Following this, the model will be optimized with real application data, involving parameter adjustments and retraining to enhance the algorithm's accuracy.

1.5.3. Data Requirements for SOS Estimation

The following table reviews the data requirements for the SOS algorithm.

Table 6. SOS Data Requirements

SOS data requirements						
Voltage Resolution	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
0.01 V (cell level)	10 Hz	1ºC	0.1 Hz	0.1 C	10 Hz	N/A

1.5.4. Evaluation Criteria

To validate the baseline SOS algorithm, data from safety tests obtained from open sources, as detailed in Section 2, will be used. These tests are considered the best framework for validating the SOS algorithm because they reflect conditions outside the recommended bounds, precisely the situations the SOS algorithm is designed to detect. The safety tests include three different methods: nail penetration, thermal abuse, and short circuit test.

Unlike other algorithms, the SOS algorithm cannot be compared to a close-to-real SOS value because there is no physical measure for "safety." Therefore, the validation focuses on the algorithm's ability to prevent catastrophic events, such as thermal runaway. The evaluation of the SOS algorithm will be conducted through both physical and virtual tests. SOX values will be monitored throughout the duration of both virtual and physical tests.

1.6. Remaining Useful Life (RUL)

1.6.1. RUL definition

Remaining Useful Life (RUL) is defined as the time remaining until the battery requires replacement or significant maintenance, i.e., until it reaches its End of Life (EOL). BIG LEAP project approaches the transition of batteries from its first life to their second life. Therefore, RUL is divided into two separate concepts: Remaining Useful First Life (RU1L) and Remaining Useful Second Life (RU2L). RU1L is defined as the remaining time until a battery reaches its EOL during its first life (EO1L), and RU2L is defined as the remaining time until a battery reaches its EOL during the second time (EO2L).

Table 7 reviews the criteria that will be used in BIG LEAP project for the definition of EO1L and EO2L, which directly affect into the definition of RU1L and RU2L.

Note that many batteries used in second life do not directly start at their 80% SOH, as many of them are discarded before ending the typical EO1L criteria due to other reasons. This is the reason why the SOH threshold difference between EO1L and EO2L is reduced.

Table 7. EO1L and EO2L definition criteria.

EO1L and EO2L definition criteria	
EOL	**SOH threshold**
End of 1st Life	80%
End of 2nd Life	70%

1.6.2. Methodology for RUL estimation

On the one hand, RU1L estimation algorithm will be based on a data-driven approach, in which the relation between the typical degradation factors (time in rest, SOC during rest, temperature during rest, depth of discharge in operation, middle SOC in operation, temperature in operation, and charge and discharge currents in operation) and the associated degradation will be modelled. The algorithm will be trained with the data specified in Section 2, and will allow evaluating the impact of different future scenarios into the RU1L.

Similarly, the RU2L prognostic algorithm is also a data-driven approach. The proposed scheme models the comparatively fast second-life aging dynamics of the energy storage system, derived from factors such as trends in capacity degradation and the evolution of the cells' series resistance. The impact of the other cell parameters, including temperature, depth of discharge, C-rate, and, most importantly, the first life uses history, play a key role in the proposed RU2L scheme. Machine learning-based approaches will be used in training the RU2L algorithm, considering the data listed in section 2. The aim of the algorithm is to estimate the reaming cycle life of the energy storage system for the present use cases. The effectiveness of the proposed scheme will be evaluated considering the performance indices defined in sub-section 1.6.4.

1.6.3. Requirements of data for RUL estimation

The specifications of the data required for the development and training of RU1L and RU2L algorithms is defined in Table 8. As it can be seen, the requirements are smoother compared to the previous algorithms, as the dynamics affecting the long-term degradation of batteries are lower.

Table 8. RUL data requirements.

RUL data requirements							
RUL	**Voltage Resolution**	**Voltage Sample Rate**	**Temp. Resolution**	**Temp. Sample Rate**	**Current Resolution**	**Current Sample Rate**	**Other Signals**
RU1L	0.05 V (cell level)	1 Hz	1ºC	0.1 Hz	0.1 C	0.2 Hz	N/A
RU2L	0.05V (Cell level)	1 Hz	1ºC	0.1 Hz	0.1 C	0.2 Hz	N/A

1.6.4. Evaluation criteria

RU1L algorithm will be evaluated according to the following KPIs:

- On the one hand, the RU1L definition is expected to be within an error threshold of the 10%:

$$RU1L_{error} = abs\left(\frac{RU1L_{estimation} - RU1L_{measured}}{RU1L_{measured}}\right) \cdot 100 < 10\%$$

- On the other hand, as the RU1L definition will be based on the estimation of the capacity loss, the MAE of the capacity estimations will be also evaluated. An error below the 5% is expected for this measure:

$$MAE_Q = \sum_{i=1}^{N} \frac{SOH_{i-est} - SOH_{i-measured}}{SOH_{i-measured}} \cdot 100 < 5\%$$

Similarly, the RU2L algorithm's performance will be assessed using Absolute Percentage Error (APE) and Mean Absolute Error (MAE). Performance indices should be limited to within 10% and 5% and can be defined as:

$$RU2L_{APE} = \left|\frac{RU2L_{estimation} - RU2L_{measured}}{RU2L_{measured}}\right| \cdot 100 < 10\%$$

$$RU2L_{MAE} = \left|\frac{1}{N}\sum_{i=1}^{N} \frac{SOH(i)_{estimation} - SOH(i)_{measured}}{SOH(i)_{measured}}\right| \cdot 100 < 5\%$$

2. Data Gathering for SOX/RUL algorithms

This section presents the battery operation data collected during the development of task T3.1. This data has been collected aiming its use in the successive tasks of WP3, that is to say, in order to be useful in the development and parametrization of the battery models, SOX and RUL estimators.

Section 2.1 focuses on the identification of open access databases. These databases may include battery data obtained at laboratory environments, where characterization or degradation experiments are held at static and dynamic conditions; or battery data obtained from real operation. In this section, a series of open-access databases are identified, and for each algorithm (SOC, SOE, SOH, SOP, SOS and RUL) the most appropriate ones are defined, based on the characteristics of their data.

On the other hand, section 2.2 focuses on the battery operational data provided by the battery OEMs participating in BIG LEAP project (OCTAVE, SIRO, SOLITEK and CORVUS). This data complements the data collected from the open access databases and represents data from real operation.

2.1. Open Access Databases

The first step for the selection of the most appropriate data for each algorithm has been the identification of a series of databases containing battery operation data. These databases have been identified based on the previous work made by different agents and publicly available in sites such as Battery Archive [2], Battronics [3], or in different publications [4] [5] [6] [7]. Moreover, some databases have been identified searching in the web key words such as "battery operational database". Table 9 below shows the list of the identified databases. The table assigns an ID to each database, and displays the data source (laboratory, field or synthetic data), the sample size, the cathode chemistry and the reference of the batch.

Table 9. Identified Open Access Databases

Open Access Battery Databases					
ID	**Name**	**Data Source[1]**	**Sample Size**	**Cathode Chemistry**	**Ref.**
1	CALCE CS2	Lab	15	LCO	[5]
2	CALCE CX2	Lab	12	LCO	[6]
3	CALCE PL	Lab	16	LCO	[7]
4	CALCE Storage and Test	Lab	144	LCO	[8]
5	CALCE Accelerated Cycle Life	Lab	192	LCO	[9]

6	HNEI Relaxation Benchmark	Lab	12	NMC LFP	[10]
7	HNEI Synthetic V vs. Q Training Dataset	Syn	N/A	LFP	[11]
8	HNEI Synthetic Training Diagnosis Dataset	Syn	N/A	LFP	[12]
9	HNEI Synthetic Training Prognosis Dataset	Syn	N/A	LFP	[13]
10	ORNL/Sandia Mechanically Induced Thermal Runaway	Lab	105	NMC LMO LFP LCO	[14]
11	SANDIA 2020 cycling data	Lab	30 (LFP) 24 (NCA) 32 (NMC)	LFP NCA NMC	[15]
12	Sandia Cell Cycle Testing Data	Lab	24	LCO LFP NCA NMC	[16]
13	Sandia Cell Abuse Testing Data	Lab	12	LCO LFP NCA NMC	[17]
14	Oxford Battery Degradation Dataset 1	Lab	8	LCO	[18]
15	Oxford energy trading battery degradation dataset	Lab	6	NMC	[19]
16	Oxford Path Dependent Battery Degradation Part 1	Lab	12	NCA	[20]
17	Oxford Path Dependent Battery Degradation Part 2	Lab	12	NCA	[21]
18	Oxford Path Dependent Battery Degradation Part 3	Lab	12	NCA	[22]
19	UofM Cyclic Aging Dataset	Lab	21	NMC	[23]
20	NASA Randomized Battery Usage Data Set	Lab	28	LCO	[24]
21	NASA Accelerated Battery Life Testing Data Set	Lab	14	NCA	[25]
22	Stanford Aging Based on Real Driving Profiles	Lab	10	NMC	[26]
23	Toyota Fast Charge Cycle Life Prediction	Lab	124	LFP	[27]

24	Toyota-MIT-Stanford Fast Charging	Lab	192	LFP	[27]
25	XJTU battery dataset	Lab	55	NMC	[28]
26	AESA BIT-MIT Battery Degradation Dataset	Lab	77	LCO-NMC	[29]
27	EVBattery Real-World Charging Dataset	Field	464	-	[30]
28	Battery Failure Databank	Lab	-	-	[31]
29	Experimental cycling data	Lab	130	LCO NCA NMC	[32]
30	Experimental galvanostatic discharge tests at different C-rates	Lab	18	NCA LFP NMC	[33]
31	EIS data of li-ion battery for SOH estimation using ML	Lab	-	NMC	[34]

[1] Data source: Laboratory (Lab), Field (Field), Synthetic (Syn)

In the following sub-sections, an analysis of the databases identified in Table 9 is carried out, focusing on the specific data characteristics required by each algorithm. As a result of this analysis, the most appropriate databases for the development of each algorithm (SOC, SOE, SOH, SOP, SOS, RU1L and RU2L) are defined.

2.1.1. Databases for SOC and SOE algorithms

As outlined in Section 1, the development of the SOC and SOE algorithms requires comprehensive data including measurements of battery voltage, current, and temperature under various operating conditions. These data are essential for the neural network to learn to estimate both SOC and SOE. To link specific conditions to SOC and SOE values, it is important that the data cover a diverse set of circumstances.

Analysis of the database is intended to determine whether they contain sufficient variability under these conditions to support robust development of SOC and SOE algorithms. Given the need for these algorithms to be adaptable to different battery chemistries, the process of identifying suitable databases is categorized by chemistry types, including NMC (Nickel-Manganese-Cobalt), LFP (Lithium Iron Phosphate), LCO (Lithium Cobalt Oxide), and NCA (Nickel Cobalt-Aluminium Oxide).

To ensure a robust development process, the selected databases will be evaluated on several criteria: the diversity of conditions they represent, the accuracy and consistency of the recorded measurements and the extent to which they cover the full range of expected operating conditions for each battery chemistry.

These databases will serve as the foundation for training and validating the neural network models, aiming to create accurate and reliable SOC and SOE estimators.

2.1.1.1. Databases for NMC chemistry

First, Table 10 shows the selected database for NMC chemistry. In this case, it is concluded that database #12 is the most appropriate, as it includes a sufficient variety of conditions, with different charge and discharge rates and temperatures. However, in this case only the batteries were cycled under static charging/discharging rates, they have not been cycled under dynamic profiles. Despite this, the database is considered to offer sufficient variability to model the SOC and SOE algorithms. Database #25 is also an option to consider, so it will be kept as a backup in case they are needed.

Table 10. Selected databases for NMC chemistry for SOC/SOE estimation.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• It includes variety charge and discharge rates combinations. • Data is correctly labelled. • Different temperatures have been considered. • As disadvantage, no dynamic cycling has been performed.
25	XJTU battery dataset	• Cyclic aging in 6 different charge and discharge strategies until 80% • Discharge and charge under different conditions. • Only one temperature is considered.

2.1.1.2. Databases for LFP chemistry

Table 11 shows the selected database for LFP chemistry. In this case, database #24 is the most appropriate choice. Although there is no variability in discharge rate and temperature, it is considered to be the most comprehensive database available. Databases #23 and #30 will be left as backup, as they show less variability than #24.

Table 11. Selected databases for LFP chemistry for SOC/SOE estimation.

Selected databases for LFP chemistry		
ID	**Name**	**Database Analysis**
24	Toyota-MIT-Stanford Fast Charging	• It includes cyclic aging in 5 different six-step charging protocols for 100 cycles. • As disadvantage, only one discharge rate and one temperature are used.

23	Toyota Fast Charge Cycle Life Prediction	• Cyclic aging until failure under high C-Rates • As disadvantage, only one discharge rate and one temperature are used.
30	Experimental galvanostatic discharge tests at different C-rates	• Different temperatures, C-rates and DoDs are tested. • No dynamic discharge or charge profile is used. • Only few cells are tested.

2.1.1.3. Databases for LCO chemistry

After analysing the different options that exist for LCO chemistry cells, it has been considered, as depicted in Table 12, that #12 is the most appropriate, since it takes into account charges and discharges under different currents and also considers different temperatures. The only drawback is that there are no cycles under dynamic profiles. Even so, it is considered a good option to have as a reserve the dataset #29, which also contains cycles under different conditions.

Table 12. Selected databases for LCO chemistry for SOC/SOE estimation.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• It includes variety charge and discharge rates combinations. • Data is correctly labelled. • Different temperatures have been considered. • As disadvantage, no dynamic cycling has been performed.
29	Experimental cycling data	• Different temperatures, C-rates and DoDs are tested. • Good number of cells tested. • No dynamic discharge or charge profile is used

2.1.1.4. Databases for NCA chemistry

The last chemical analysed is the NCA. For the same reasons cited in 2.1.1.1.1 and 2.1.1.3, and as shown in Table 13, it is believed that the best option for this chemistry is #12. However, in case #12 is not sufficient, databases #29 and #30 are also considered as options to be considered.

Table 13. Selected databases for NCA chemistry for SOC/SOE estimation.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• It includes variety charge and discharge rates combinations. • Data is correctly labelled. • Different temperatures have been considered. • As disadvantage, no dynamic cycling has been performed.

29	Experimental cycling data	• Different temperatures, C-rates and DoDs are tested. • Good number of cells tested. • No dynamic discharge or charge profile is used
30	Experimental galvanostatic discharge tests at different C-rates	• Different temperatures, C-rates and DoDs are tested. • No dynamic discharge or charge profile is used. • Only few cells are tested.

Therefore, the databases best suited to the needs of the SOC/SOE estimation algorithms are believed to be #12 and #24, although databases #11, #23, #25, #29 and #30 will be kept as backups,

2.1.2. Databases for SOH algorithm

2.1.2.1. BFH Algorithm

There are many open battery ageing datasets available to train the SOH algorithm on. Many of them vary multiple factors such as temperature and C-Rate, which is advantageous for the training of the algorithm. All chosen datasets were measured on high-precision laboratory equipment which meet the data requirements. Database #22 will be used to test and validate the algorithm's performance with a dataset containing a drive profile that closely resembles the real-world operation.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
19	UofM Cyclic Aging Dataset	• Different Temperatures, C-Rates and DODs • Dataset goes to 70% SOH and beyond
22	Stanford Aging Based on Real Driving Profiles	• Dynamic drive profiles, but only one temperature
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for LFP chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
23	Toyota Fast Charge Cycle Life Prediction	• High C-Rates (fast charging) but only one temperature
24	Toyota-MIT-Stanford Fast Charging	• Ageing with multiple different high C-Rate protocols

		• Only one temperature
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for LCO chemistry		
ID	**Name**	**Database Analysis**
1	CALCE CS2	• Different C-Rates, no dynamic tests
2	CALCE CX2	• Different C-Rates, no dynamic tests
3	CALCE PL	• Different C-Rates and DoDs, no dynamic tests
4	CALCE Storage and Test	• Useful for calendar degradation analysis at different temperatures
5	CALCE Accelerated Cycle Life	• Different Temperatures, C-Rates, and cutoff C-rates, no dynamic tests
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
14	Oxford Battery Degradation Dataset 1	• Dynamic drive cycles, but only one temperature
20	NASA Randomized Battery Usage Data Set	• Different Temperatures, C-Rates, DODs • Older data set
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for NCA chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

2.1.2.2. FHG Algorithm

The first key component for accurate SOH estimation is the accuracy and data rate of the current measurement, and secondly the same for the temperature measurement. Therefore, data sets with different C-rates, with and without dynamics, sampled at different rates and accuracy are crucial for training and evaluation of the SOH algorithm.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
19	UofM Cyclic Aging Dataset	• Different Temperatures, C-Rates and DODs • Dataset goes to 70% SOH and beyond
22	Stanford Aging Based on Real Driving Profiles	• Dynamic drive profiles, but only one temperature
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for LFP chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
23	Toyota Fast Charge Cycle Life Prediction	• High C-Rates (fast charging) but only one temperature
24	Toyota-MIT-Stanford Fast Charging	• Ageing with multiple different high C-Rate protocols • Only one temperature
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for LCO chemistry		
ID	**Name**	**Database Analysis**
1	CALCE CS2	• Different C-Rates, no dynamic tests
2	CALCE CX2	• Different C-Rates, no dynamic tests
3	CALCE PL	• Different C-Rates and DoDs, no dynamic tests

4	CALCE Storage and Test	• Useful for calendar degradation analysis at different temperatures
5	CALCE Accelerated Cycle Life	• Different Temperatures, C-Rates, and cutoff C-rates, no dynamic tests
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
14	Oxford Battery Degradation Dataset 1	• Dynamic drive cycles, but only one temperature
20	NASA Randomized Battery Usage Data Set	• Different Temperatures, C-Rates, DODs • Older data set
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests

Selected databases for NCA chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• Different Temperatures and C-Rates, no dynamic tests
29	Experimental cycling data	• Different Temperatures, C-Rates and DODs, no dynamic tests
30	Experimental galvanostatic discharge tests at different C-rates	• Different Temperatures, C-Rates and DODs, no dynamic tests

2.1.3. Databases for SOP algorithm

Data sets for the SOP algorithm development must contain at least a pulse characterisation test, and ideally also an electrochemical impedance spectroscopy (EIS) test. All the chosen datasets were measured using high-precision laboratory equipment, ensuring that the data requirements outlined in 1.4.3. are met.

The public datasets may not be sufficient for the novel impedance analysis to be developed in WP2, in which case they will be supplemented by additional laboratory tests and validation.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• EIS data available, but no pulse characterisation tests
19	UofM Cyclic Aging Dataset	• Pulse characterisation and EIS data available
22	Stanford Aging Based on Real Driving Profiles	• Pulse characterisation and EIS data available

Selected databases for LFP chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• EIS data available, but no pulse characterisation tests
23	Toyota Fast Charge Cycle Life Prediction	• Pulse characterisation tests, but no EIS
24	Toyota-MIT-Stanford Fast Charging	• Pulse characterisation and first life EIS data available

Selected databases for LCO chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• EIS data available, but no pulse characterisation tests

Selected databases for NCA chemistry		
ID	**Name**	**Database Analysis**
12	Sandia Cell Cycle Testing Data	• EIS data available, but no pulse characterisation tests
16	Oxford Path Dependent Battery Degradation Part 1	• Pulse characterisation and EIS data available
17	Oxford Path Dependent Battery Degradation Part 2	• Pulse characterisation and EIS data available
18	Oxford Path Dependent Battery Degradation Part 3	• Pulse characterisation and EIS data available

2.1.4. Databases for SOS algorithm

As mentioned in Section 1.5.4, the SOS algorithm will require data from abuse tests for validation and training. From the open-source databases collected, two primary sources containing abuse test data have been identified, as shown in the table.

Selected databases for NCA chemistry		
ID	**Name**	**Database Analysis**
13	Sandia Cell Abuse Testing Data	• Cyclic testing beyond manufacturer specs for LCO, LFP, NCA, NMC chemistries.

28	Battery Failure Databank	• Contains data of heat output, heat breakdown, cell mass before and after thermal runaway, and mass ejected from various abuse tests, such as nail penetration and thermal abuse. Identified chemistries are NMC, NCA

2.1.5. Databases for RUL1 algorithm

As specified in section 1, in order to develop an appropriate RUL1 algorithm, data including an evolution of the SOH (i.e., battery degradation) and the associated circumstances (at least temperature, voltage and current data) is required. In order to properly associate specific circumstances to specific degradations, it is preferred that the data includes a variety of circumstances. For instance, for degradation tests held at laboratory environments, a variety of circumstances is typically obtained by deploying tests at different temperatures, SOC variations, and charging and discharging C-rates (including zero SOC variations and C-rates for calendar degradation). Therefore, the analysis of the data is focused on identifying if the databases include enough variations in the mentioned circumstances in order to properly develop the RU1L algorithm.

As the RU1L algorithm developed in BIG LEAP project may be adaptable for a variety of battery chemistries, the identification of databases is split into different chemistries. In other words, for each chemistry (NMC, LFP, LCO and NCA) the most appropriate databases are identified.

First of all, Table 14 below shows the database selected for NMC chemistry. In this case, it is concluded that database #19 is the most appropriate, as it includes enough variety of circumstances, with different temperatures, charge and discharge C-rates and SOC variations. However, only two different delta SOCs are evaluated, and the database does not include information about calendar degradation. In any case, it is considered to be enough variability to model a RUL model for first life. Databases #11, #25 and #29 also include a variety of circumstances, but not as wide as database #19. Therefore, they will be maintained as back-up in case database #19 shows any limitation.

Table 14. Selected databases for NMC chemistry.

Selected databases for NMC chemistry		
ID	**Name**	**Database Analysis**
19	UofM Cyclic Aging Dataset	• It includes variety of temperatures, charge and discharge rates combinations and delta SOCs. Different pressures are also evaluated. • Data is correctly labelled with capacity evolution. • As disadvantage, only two different delta SOCs are included. No calendar degradation is analysed.

Table 15 shows the database selected for LFP chemistry. In this case, it is found that database #11 is the most appropriate option. Compared to the database selected for NMC, in this case there is no variability in the charging rate, but it is still the most comprehensive database for LFP. Databases #23, #24 and #30 are left as back-up, as they show even less variability in the circumstances compared to database #11.

Table 15. Selected databases for LFP chemistry.

Selected databases for LFP chemistry		
ID	**Name**	**Database Analysis**
11	SANDIA 2020 cycling data	• It includes variety of temperatures, discharge rates and delta SOCs. In the case of NMC, it includes data of 32 experiments. • Data is correctly labelled with capacity evolution, in all cases until 80% SOH. • As disadvantage, all tests are held at 0.5C charging rate, and the same middle SOC (50%) is used for all the delta SOCs. No calendar degradation is analysed.

Besides, Table 16 shows the selected databases for LCO chemistry. The three databases are provided by the University of Maryland, and they individually analyse the effects of some of the typical degradation factors: Database #4 focuses on the calendar degradation (effect of temperature and SOC), database #3 focuses on the effect of discharge rate and SOC variation, and database #5 focuses on the effect of temperature and discharge rate. Even if the different databases do not include data of the same exact cell, they all make use of a prismatic LCO cell. Therefore, during the development of the RUL algorithm (task T3.2), it will be evaluated if merging the data of the three databases of Table 16 is a feasible solution. In case it is found that merging the databases is not a feasible option, databases #12 and #20 are also left as back-up options.

Table 16. Selected databases for LCO chemistry.

Selected databases for LCO chemistry		
ID	**Name**	**Database Analysis**
4	CALCE Storage and Test	• It analyses the calendar degradation: different combinations of temperature and SOC values are analysed. • Data is correctly labelled.
3	CALCE PL	• It includes variety of discharge C-rates and delta SOCs. • Data is correctly labelled. • Effect of temperature and charge C-rate is not evaluated.
5	CALCE Accelerated Cycle Life	• It includes a variety of temperatures and discharge C-rates. Different charge protocols are analysed, but only focusing on varying the CV phase ending point. • Data is correctly labelled.

Finally, Table 17 displays the selected databases for NCA chemistry. In this case, there is neither a preferred database, as it is concluded that the compromise between the identified advantages and disadvantages is similar in all the cases. Database #11 does not include the effect of the charging rate or calendar degradation; Databases #16-18 (treated as a single database) do not include the effect of the temperature or the delta SOC; and Database #21 does not analyse the effect of calendar degradation and it may require a pre-processing to correctly label the degradation associated to each set of circumstances. Database #29 is left as back-up option for NCA chemistry.

Table 17. Selected databases for NCA chemistry.

<table>
<tr><th colspan="3">Selected databases for LCO chemistry</th></tr>
<tr><th>ID</th><th>Name</th><th>Database Analysis</th></tr>
<tr><td>11</td><td>SANDIA 2020 cycling data</td><td>• It includes variety of temperatures, discharge rates and delta SOCs. In the case of NCA, it includes data of 24 experiments.
• Data is correctly labelled with capacity evolution, in all cases until 80% SOH.
• As disadvantage, all tests are held at 0.5C charging rate, and the same middle SOC (50%) is used for all the delta SOCs. No calendar degradation is analysed.</td></tr>
<tr><td>16</td><td>Oxford Path Dependent Battery Degradation Part 1</td><td rowspan="3">• Databases #16, #17 and #18 are treated as a single database, as the same battery cell is used.
• Different charge and discharge rates, and different calendar conditions are analysed.
• Path dependence is analysed.
• As disadvantage, influence of temperature and delta SOC is not analysed</td></tr>
<tr><td>17</td><td>Oxford Path Dependent Battery Degradation Part 2</td></tr>
<tr><td>18</td><td>Oxford Path Dependent Battery Degradation Part 3</td></tr>
<tr><td>21</td><td>NASA Accelerated Battery Life Testing Data Set</td><td>• It includes a variety of temperatures, charge and discharge rates, and delta SOCs.
• Not calendar aging is analysed.
• Data is not labelled with periodical check-ups.</td></tr>
</table>

In short, Databases #3, #4, #5, #11, #16-#18, #19 and #21 are found to be the most appropriate options to develop the RU1L algorithm for the different battery chemistries. In addition, Databases #12, #20, #23, #24, #25, #29, and #30 are left as potential back-up options in case the previously mentioned databases show any limitations during the algorithms development phases.

2.1.6. Databases for RU2L algorithm

To accurately estimate the remaining useful second-life (RU2L) of the first-life retried battery energy storage system (SL-BESS), it is crucial to access a comprehensive database containing both historical and current data on battery performance, state of health degradation, and usage.

The database should provide valuable insights for RUL estimation, including the battery's initial health state (SOH), aging factors, failure modes, and end-of-life (EOL) indicators. Consequently, the data analysis is focused on determining whether the databases encompass enough variations in the mentioned circumstances to develop the RU2L algorithm properly. Like the RU1L algorithm, the RU2L prognostic scheme developed in the BIG LEAP project needs to be adaptable to different battery chemistries. However, obtaining open-access RU2L data based on various battery chemistries is challenging because second-life battery applications are not as common as first-life applications. At the time of preparing this document, an LCO cathode chemistry-based second-life battery usage data-based has been identified, as detailed in Table 18.

Table 18. RU2L Algorithm Database for LCO Battery Chemistry

Selected databases for LCO chemistry		
ID	**Name**	**Database Analysis**
2	CALCE CX2	• Capacity drop results up to ~26% (>20%) for multiple cells considering the test conditions- • at different discharge C-rates and cut-off voltages, • 100% depth-of-discharge.
5	CALCE Accelerated Cycle Life	• Accelerated cyclic aging at different temperatures and C-rates for different charging protocols.

In addition, one of the BIG LEAP project partners, OCTAVE, provided *second-use* battery data for the development of the RU2L prediction scheme (see Section 2.2). Second-use batteries do not necessarily reach the end of their first life (i.e., ≥20% capacity drop); however, they retire from their typical high-power first-life applications, such as the transportation sector. Certainly, the reason is the state of health degradation in terms of the power delivery capacity of the battery. So, typically, the low-power demands stationary applications are the more economical and potential alternatives for these first-use retired batteries with significant remaining useful life - namely second-use. In view of the lack of second-life battery data, in this study, the second-use battery data are being considered for developing the RU2L prognostic algorithm. OCTAVE provided the second-use battery data for NMC battery chemistry, as mentioned in Table 19. Section 2.2 discusses the OCTAVE's in-house second-use battery data in more detail.

Table 19. Second-use NMC Battery Data from OCTAVE

Selected second-use databases for NMC chemistry		
ID	**Name**	**Database Analysis**
#	**OCTAVE 2nd-use**	• Includes voltage, current, time, and temperature measurements of cell level and • SOC and SOH data using developed in-house algorithms.

2.2. Battery Operational Data

Apart from the open-access data previously presented in Section 2.2, during Task 3.1 of Big Leap project industrial partners (SIRO, SOLITEK, CORVUS and OCTAVE) have shared operational data of their batteries. This data corresponds to the operation of their batteries, and include at least current, voltage and temperature timeseries information. Depending on the provider, SOC and SOH information is also provided.

This battery operational data is considered critical for the battery providers, and therefore it is not described in the current deliverable. The data sharing among the project partners has been granted by the Consortium Agreement (CA) signed by all the partners.

3. Data Standardization

Data standardisation is a crucial process in any collaborative project, especially where multiple partners may come to work together using the same databases. Standardisation ensures that the data collected and used is consistent, regardless of its origin. This is particularly important when developing advanced algorithms for lithium-ion battery condition monitoring and management, as variability in data quality and format can significantly affect the results and effectiveness of the algorithms.

The diversity of data sources, which may come from different project partners or from different existing public databases, requires a common data structure to facilitate analysis and integration. A standardised file format, such as a .csv file, allows for easy manipulation and processing of data, ensuring that all participants can contribute and access a unified and consistent dataset.

Therefore, to meet the requirements of the project and to facilitate the development of the battery status algorithms, it is believed that the use of a .csv file format with the columns (using ';' to divide the different columns) in the following order may be advantageous:

1) Label Data Origin:
    a. Description: indicates the origin of the data, either the name of the project partner or the name of the database from which the data originates.
    b. Example: Partner1, DatabaseXYZ.
2) Label Chemistry:
    a. Description: Specifies the chemistry of the lithium-ion battery. This can include different chemical compositions such as LFP (Lithium Iron Phosphate), NMC (Nickel Manganese Cobalt), etc.
    b. Example: LFP, NMC.

3) Timestamp:
    a. Description: timestamp of the data record, converted to milliseconds. This ensures an accurate and uniform time reference.
    b. Example: 15684.
4) Average Voltage:
    a. Description: Average voltage of the different cells that make up the module or system.
    b. Example: 3.7.
5) Average Temperature:
    a. Description: Average temperature recorded from the different temperature sensors.
    b. Example: 25.3.
6) Current:
    a. Description: Current measured in amperes (A).
    b. Example: 1.5.
7) State of Charge (SOC):
    a. Description: State of charge of the battery, expressed as a percentage. If this information is not directly available, it can be obtained by the Coulomb Counting method.
    b. Example: 80.
8) Voltage vector or multiple columns of voltage:
    a. Description: Set of voltage values measured on different cells or modules of the battery.
    b. Example: 3.65, 3.66, 3.64.
9) Temperature vector or multiple columns of temperature:
    a. Description: Set of temperature values measured at different points of the battery if more than one temperature is available.
    b. Example: 25.0, 25.1, 24.9.
10) Last recorded SOH value:
    a. Description: Last recorded value of the battery's state of health (SOH), if available.
    b. Example: 95.

Therefore, Table 20 shows an example of how the data in the .csv would be split.

Table 20. Example of the .csv after the data standardization.

Label Data Origin	Label Chemistry	Timestamp [ms]	Average Voltage [V]	Average Temperature [ºC]	Current [A]	SOC [%]	Voltage Vector [V]	Temperature Vector [ºC]	Last Recorded SOH Value [%]
Partner1	LFP	15684	3.7	25.3	1.5	80	3.65,3.66,3.64	25.0,25.1,24.9	95
Partner1	LFP	15884	3.8	26.1	1.7	85	3.75,3.74,3.76	26.0,26.2,25.9	90

4. Methodology for secure integration of data into SUNDIAL platform

This chapter outlines the security methodologies implemented to guarantee secure access to the SUNDIAL platform. The architecture leverages VPNs, firewalls, user authentication, and role-based access control to provide a robust security framework, ensuring data integrity and protection against unauthorized access. These strategies can be adapted for the project needs and other methodologies can be implemented.

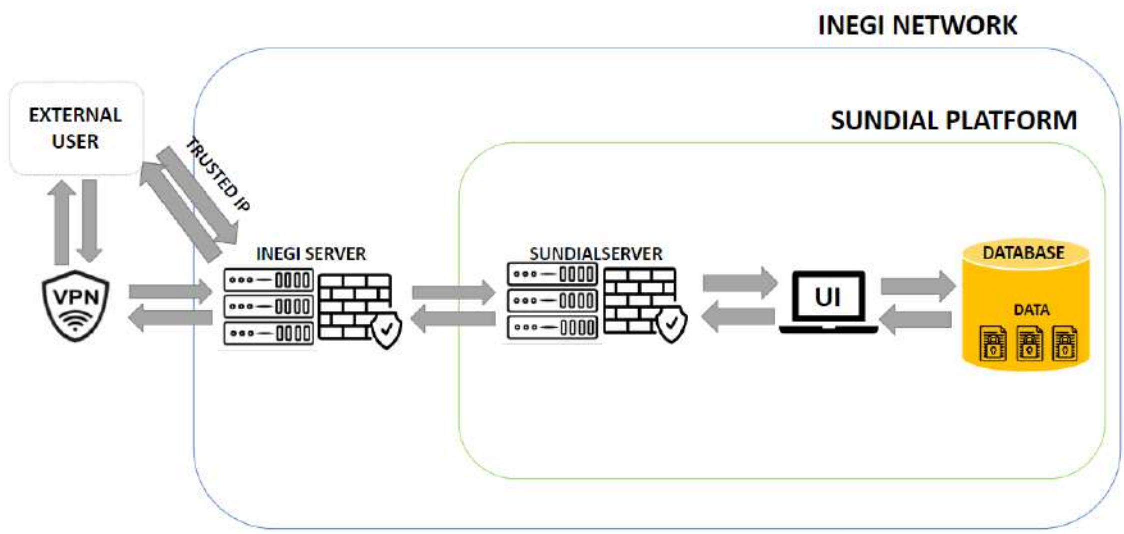

Figure 4. Secure integration of data into SUNDIAL platform.

4.0. VPN

This security layer is the first one and is the most important because it allows partners to find and securely access INEGI network. The Virtual Private Network (VPN) setup ensures that external users can connect to the INEGI network over an encrypted communication channel, providing a secure entry point into the system.

For partners who are unable to use VPN, it is crucial to establish secure and controlled access to the INEGI network, in that cases will be open connection to the server for that partner's network.

4.1. Firewall

To ensure the secure operation of the SUNDIAL platform, two dedicated firewalls need to be configured. These firewalls provide a layered security approach, controlling access to INEGI network and SUNDIAL platform.

By implementing and maintaining two firewalls, the INEGI network ensures robust security for the SUNDIAL platform. The first firewall protects the INEGI network from external attacks and controls who can access the SUNDIAL platform. The second firewall ensures that only users within INEGI network can access the platform. Together, these firewalls provide a defence in depth strategy, enhancing the overall security of the SUNDIAL platform.

4.2. User authentication and Data Encryption

Ensuring secure access to the SUNDIAL platform and protecting the integrity and confidentiality of data is extremely important. By implementing robust user authentication and data encryption measures, the INEGI network ensures the secure access and protection of the SUNDIAL platform data. User authentication mechanisms verify the identity of users and restrict access to authorized individuals only, while data encryption protects the confidentiality and integrity of data. These measures collectively enhance the overall security of the SUNDIAL platform, safeguarding it against unauthorized access and potential threats.

5. Conclusions

In this deliverable, the specifications for the SOX (SOC, SOH, SOP, SOE, SOS) and RUL algorithms have been defined first. The key points of each algorithm include the definition of the estimated variable, the development methodology, the data requirements and the evaluation criteria.

In the case of SOC and SOE, algorithms will be developed to estimate the state of charge and battery energy using machine learning techniques. These algorithms will be trained and validated with accurate and representative data, and their performance will be measured using key indicators such as MAPE.

On the other hand, the SOH will use a statistical data framework and a neural network framework, which will be used to estimate battery health status. Both methods require accurate, high-frequency data to capture ageing dynamics.

The SOP algorithm will be done using an Equivalent Circuit Model, which will estimate the charge and discharge power limits. The algorithm must be able to provide accurate estimates in real time, with data captured at high frequency.

The SOS algorithm shall evaluate through an abuse function that considers multiple factors affecting battery safety. This algorithm will be validated through physical and virtual testing to ensure its ability to prevent catastrophic events.

Finally, the RUL will be split into two algorithms, one for the first and one for the second battery life, using methods based on historical data and machine learning techniques. The objective is to estimate the time remaining until the battery needs replacement or significant maintenance.

To develop and validate the SOX (SOC, SOE, SOH, SOP, SOS) and RUL estimation algorithms, open databases have been selected. The databases have been chosen after a thorough analysis of their content and suitability to the specific requirements of each algorithm.

In the case of SOC and SOE, databases containing voltage, current and temperature measurements under various operating conditions have been identified. For the SOH algorithm, databases with accurate temperature and current measurements were chosen. The databases chosen for the SOP include pulse characterisation tests and electrochemical impedance spectroscopy (EIS). Databases containing abuse test data, essential to validate the SOS algorithm, have been selected. Finally, to estimate the RUL, databases including variability in degradation conditions were selected.

On the other hand, the need to standardise the data in a common format, such as .csv files, to facilitate consistent manipulation and analysis between the different project partners has been identified and emphasised.

The need for secure data management has also been identified and robust security methodologies, including VPNs, firewalls, user authentication and data encryption, will be implemented to ensure secure access to the SUNDIAL platform and data integrity.

This comprehensive approach to database selection, standardisation and security ensures that the algorithms developed in the BIG LEAP project are accurate, robust and secure, improving battery management and efficiency in various applications.

References

[1] J. Jose Saucedo-Dorantes, D. Alejandro Elvira-Ortiz, C. Gustavo Manriquez-Padilla, A. Yosimar Jaen-Cuellar, and A. Perez-Cruz, "New Trends and Challenges in Condition Monitoring Strategies for Assessing the State-of-charge in Batteries," 2023. doi: 10.5772/intechopen.109062.

[2] "Battery requirements for future automotive applications - EUCAR." Accessed: Jul. 11, 2024. [Online]. Available: https://www.eucar.be/battery-requirements-for-future-automotive-applications/

[3] C. Ashtiani, "Analysis of Battery Safety and Hazards' Risk Mitigation," *ECS Trans*, vol. 11, no. 19, pp. 1–11, Mar. 2008, doi: 10.1149/1.2897967.

[4] E. Cabrera-Castillo, F. Niedermeier, and A. Jossen, "Calculation of the state of safety (SOS) for lithium ion batteries," *J Power Sources*, vol. 324, pp. 509–520, Aug. 2016, doi: 10.1016/j.jpowsour.2016.05.068.

[5] "Battery Data | Center for Advanced Life Cycle Engineering." Accessed: Jul. 11, 2024. [Online]. Available: https://calce.umd.edu/battery-data#CS2

[6] "Battery Data | Center for Advanced Life Cycle Engineering." Accessed: Jul. 11, 2024. [Online]. Available: https://calce.umd.edu/battery-data#CX2

[7] "Battery Data | Center for Advanced Life Cycle Engineering." Accessed: Jul. 11, 2024. [Online]. Available: https://calce.umd.edu/battery-data#PL

[8] "Battery Data | Center for Advanced Life Cycle Engineering." Accessed: Jul. 11, 2024. [Online]. Available: https://calce.umd.edu/battery-data#Storage

[9] "Battery Accelerated Cycle Life Testing Data | Center for Advanced Life Cycle Engineering." Accessed: Jul. 11, 2024. [Online]. Available: https://calce.umd.edu/battery-accelerated-cycle-life-testing-data

[10] A. Fernando, M. Kuipers, G. Angenendt, K.-P. Kairies, and M. Dubarry, "Relaxation data for commercial NMC and LFP cells," vol. 1, 2023, doi: 10.17632/Y8NSTXMDRG.1.

[11] M. Dubarry, "Graphite//LFP synthetic V vs. Q dataset (>700,000 unique curves)," vol. 3, 2021, doi: 10.17632/BS2J56PN7Y.3.

[12] M. Dubarry, "Graphite//LFP synthetic training diagnosis dataset," vol. 1, 2020, doi: 10.17632/BS2J56PN7Y.1.

[13] M. Dubarry, "Graphite//LFP synthetic training prognosis dataset," vol. 1, 2020, doi: 10.17632/6S6PH9N8ZG.1.

[14] L. Lin, H. Wang, J. Li, L. Torres-Castro, Y. Preger, and V. De Angelis, "Mechanically Induced Thermal Runaway for Li-ion Batteries," vol. 1, 2023, doi: 10.17632/SN2KV34R4H.1.

[15] Y. Preger *et al.*, "Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions," *J Electrochem Soc*, vol. 167, no. 12, p. 120532, Sep. 2020, doi: 10.1149/1945-7111/ABAE37.

[16] R. Wittman *et al.*, "Characterization of Cycle-Aged Commercial NMC and NCA Lithium-ion Cells: I. Temperature-Dependent Degradation," *J Electrochem Soc*, vol. 170, no. 12, p. 120538, Dec. 2023, doi: 10.1149/1945-7111/AD1450.

[17] "R&D Data Repository – DOE Office of Electricity Energy Storage Program." Accessed: Jul. 11, 2024. [Online]. Available: https://www.sandia.gov/ess/tools-resources/rd-data-repository

[18] C. Birkl, "Oxford Battery Degradation Dataset 1," 2017, doi: 10.5287/BODLEIAN:KO2KDMYGG.

[19] J. M. Reniers, G. Mulder, and D. A. Howey, "Oxford energy trading battery degradation dataset," 2020, doi: 10.5287/BODLEIAN:GJPDDZVP4.

[20] T. Raj, A. A. Wang, C. W. Monroe, and D. A. Howey, "Investigation of Path-Dependent Degradation in Lithium-Ion Batteries**," *Batter Supercaps*, vol. 3, no. 12, pp. 1377–1385, Dec. 2020, doi: 10.1002/BATT.202000160.

[21] "Path Dependent Battery Degradation Dataset Part 2 - ORA - Oxford University Research Archive." Accessed: Jul. 11, 2024. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:be3d304e-51fd-4b37-a818-b6fa1ac2ba9d

[22] "Path Dependent Battery Degradation Dataset Part 3 - ORA - Oxford University Research Archive." Accessed: Jul. 11, 2024. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:78f66fa8-deb9-468a-86f3-63983a7391a9

[23] "Data Set | UofM Pouch Cell Voltage and Expansion Cyclic Aging Dataset | ID: 5d86p0488 | Deep Blue Data." Accessed: Jul. 11, 2024. [Online]. Available: https://deepblue.lib.umich.edu/data/concern/data_sets/5d86p0488

[24] K. Fricke, R. G. Nascimento, M. Corbetta, C. S. Kulkarni, and F. A. C. Viana, "Prognosis of Li-ion batteries under large load variations using hybrid physics-informed neural networks," *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, vol. 15, no. 1, 2023, doi: 10.36001/PHMCONF.2023.V15I1.3463.

[25] "Prognostics Center of Excellence Data Set Repository - NASA." Accessed: Jul. 11, 2024. [Online]. Available: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/

[26] "OSF | Lithium-ion battery aging dataset based on electric vehicle real-driving profiles." Accessed: Jul. 11, 2024. [Online]. Available: https://osf.io/qsabn/?view_only=2a03b6c78ef14922a3e244f3d549de78

[27] "Experimental Data Platform." Accessed: Jul. 11, 2024. [Online]. Available: https://data.matr.io/1/projects/5c48dd2bc625d700019f3204

[28] P. M. Attia *et al.*, "Closed-loop optimization of fast-charging protocols for batteries with machine learning," *Nature*, vol. 578, no. 7795, pp. 397–402, Feb. 2020, doi: 10.1038/S41586-020-1994-5.

[29] F. Wang, Z. Zhai, B. Liu, S. Zheng, Z. Zhao, and X. Chen, "Open access dataset, code library and benchmarking deep learning approaches for state-of-health estimation of lithium-ion batteries," *J Energy Storage*, vol. 77, Jan. 2024, doi: 10.1016/j.est.2023.109884.

[30] J. Lu *et al.*, "Battery Degradation Dataset (Fixed Current Profiles＆Arbitrary Uses Profiles)," vol. 3, 2022, doi: 10.17632/KW34HHW7XG.3.

[31] "EVBattery - OneDrive." Accessed: Jul. 11, 2024. [Online]. Available: https://onedrive.live.com/?authkey=%21AO%5FaW1XR4%5FX4CA8&id=A583DCDEF1053C71%21477&cid=A583DCDEF1053C71

[32] "Battery Failure Databank | Transportation and Mobility Research | NREL." Accessed: Jul. 11, 2024. [Online]. Available: https://www.nrel.gov/transportation/battery-failure.html

[33] "Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation", doi: 10.5281/ZENODO.6405084.

[34] E. Catenaro and S. Onori, "Experimental data of lithium-ion batteries under galvanostatic discharge tests at different rates and temperatures of operation," *Data Brief*, vol. 35, p. 106894, Apr. 2021, doi: 10.1016/J.DIB.2021.106894.

[35] M. Rashid, M. Faraji-Niri, J. Sansom, M. Sheikh, D. Widanage, and J. Marco, "Dataset for rapid state of health estimation of lithium batteries using EIS and machine learning: Training and validation," *Data Brief*, vol. 48, Jun. 2023, doi: 10.1016/J.DIB.2023.109157.