Modelling Data-Driven Digital Twins of EV Batteries for Predictive Analytics

Afroditi Fouka
Information Management Unit (IMU)
Institute of Communication and Computer Systems (ICCS)
National Technical University of Athens (NTUA)
Athens, Greece
afroditifouka@mail.ntua.gr

Katerina Lepenioti
Information Management Unit (IMU)
Institute of Communication and Computer Systems (ICCS)
National Technical University of Athens (NTUA)
Athens, Greece
klepenioti@mail.ntua.gr

Alexandros Bousdekis
Information Management Unit (IMU)
Institute of Communication and Computer Systems (ICCS)
National Technical University of Athens (NTUA)
Athens, Greece
albous@mail.ntua.gr

Gregoris Mentzas
Information Management Unit (IMU)
Institute of Communication and Computer Systems (ICCS)
National Technical University of Athens (NTUA)
Athens, Greece
gmentzas@mail.ntua.gr

***Abstract*—As one of the key components of electric vehicles, the Li-ion Battery Management System (BMS) is crucial to the industrialization and marketization of electric vehicles. Developing advanced and intelligent BMSs has been gathering the research interest. However, the internal states of the battery are affected by several factors, thus making the application of predictive analytics algorithms a challenging task. With the recent advances in modelling tools and diagnostics, there is an opportunity to fuse this knowledge with emerging ML techniques towards creating a battery digital twin. In this paper, we propose a data-driven digital twin of EV batteries in order to support the implementation of predictive analytics algorithms. The architecture has been modelled according to the RAMI 4.0 principles in order to provide a systematic way of modelling and development data-driven digital twins for supporting predictive analytics of battery states.**

Keywords—digital twin, Li-ion battery, electric vehicle, RAMI 4.0, data analytics, machine learning.

I. Introduction

Lithium-Ion (Li-Ion) batteries have been widely applied as energy storage systems, such as electric vehicles (EVs) and Hybrid Electric Vehicles (HEVs) [1]. "Batteries are a key enabler for European competitiveness and decarbonization" as stated in the strategic agenda of the European Battery Partnership and will be one necessary tool to make Europe "fit for 55 within 2030" [2].

As one of the key components of electric vehicles, the Li-ion Battery Management System (BMS) is crucial to the industrialization and marketization of electric vehicles. Developing advanced and intelligent BMSs for the Li-Ion battery packs has been gathering the research interest. However, the internal states of the battery are affected by several factors, thus making the application of predictive analytics algorithms a challenging task [3]. Battery modeling and state estimation are key functions of the advanced BMS. Accurate modeling and state estimation can ensure reliable operation, optimize the battery system and provide a basis for safety management [4].

Battery management is critical to enhancing the safety, reliability, and performance of the battery systems [5][6]. Effective management of lithium-ion batteries is a key enabler for a low carbon future, with applications including electric vehicles and grid scale energy storage [7]. The lifetime of these devices depends greatly on the materials used, the system design and the operating conditions. To this end, there is an increasing research interest on Machine Learning (ML) models and algorithms dealing with lifetime prognostics [8][9]. These research works develop and test ML algorithms using a variety of input parameters in order to achieve various objectives, such as: End Of Life (EOL) prediction, Remaining Useful Life (RUL) estimation, State Of Health (SOH) estimation, State Of Charge (SOC) estimation, etc. With the recent advances in understanding battery degradation, modelling tools and diagnostics, there is an opportunity to fuse this knowledge with emerging ML techniques towards creating a battery digital twin [7].

Modelling the digital twin needs to follow the principles of the Reference Architectural Model Industrie 4.0 (RAMI 4.0). Overall, in the literature, there are only a few case studies that follow the RAMI 4.0 model, and even fewer not requiring much effort to reach the level of practical implementation [10]. However, the key issue of any design and system development in the context of Industry 4.0 is the proper implementation of RAMI 4.0 in various operations [10]. To this end, there is the need for architectural frameworks that will enable the systematic design and development of digital twins so that they tackle the big data-rich, complex, and uncertain environments in a holistic way.

In this paper, we propose a data-driven digital twin of EV batteries in order to support the implementation of predictive analytics algorithms aiming at addressing the aforementioned objectives. The proposed architecture aims at supporting several and dynamic predictive analytics processes, employing data from the heterogeneous data sources. To this end, the various objectives can be addressed dynamically and bootstrapped into a single software instance following the specific use case requirements and the available data. The architecture has been modelled according to the RAMI 4.0 principles and guidelines in order to provide a systematic way of modelling and development data-driven digital twins for supporting predictive analytics of battery states.

The rest of the paper is organized as follows: Section II presents the related works on Li-Ion battery digital twins. Section III outlines the RAMI 4.0 guidelines guiding the design and development of digital twins and then, models the proposed data-driven digital twin for EV batteries. Section IV demonstrates the implementation and applicability of the proposed digital twin to a real-life use case. Section V concludes the paper and presents our plans for future work.

II. Related Work

Battery management is critical to enhancing the safety, reliability, and performance of the battery systems [5][6]. Effective management of lithium-ion batteries is a key enabler for a low carbon future, with applications including electric vehicles and grid scale energy storage [7]. The lifetime of these devices depends greatly on the materials used, the system design and the operating conditions. With the recent advances in understanding battery degradation, modelling tools and diagnostics, there is an opportunity to fuse this knowledge with emerging machine learning techniques towards creating a battery digital twin [7].

Reference [11] provided an overview of the opportunities provided by the digital twin technology for the EVs. They classified the applications of digital twin in the following categories: intelligent driver assistance, autonomous navigation, converters and inverters, consumer-centered development, digital design and manufacturing, health monitoring, BMSs. Reference [6] presented a cloud battery management system to improve the computational power and data storage capability. The application of equivalent circuit models in the digital twin for battery systems is explored with the development of cloud-suited state-of-charge and state-of-health estimation approaches. Furthermore, a state-of-health estimation algorithm with particle swarm optimization is innovatively exploited to monitor both capacity fade and power fade of the battery during aging.

Reference [5] proposed a digital twin architecture for automotive battery systems, on which digital services for various stakeholders along the manufacturing and product life cycle of a battery system can be established. The results feature an UML meta model as a first step toward implementing of a digital twin for battery systems. Reference [7] presented the state-of-the-art in battery modelling, in-vehicle diagnostic tools, data driven modelling approaches, and how these elements can be combined in a framework for creating a battery digital twin.

Reference [12] presented the development history, basic concepts and key technologies of the digital twin, and summarized current research methods and challenges in battery modeling, state estimation, remaining useful life prediction, battery safety and control. Furthermore, based on digital twin, they described the solutions for battery digital modeling, real-time state estimation, dynamic charging control, dynamic thermal management, and dynamic equalization control in the intelligent battery management system. [13] set up a data pipeline and digital battery twin to track the battery state, including State of charge (SOC) and State of Health (SOH). Pushing this data into the cloud twin system using IoT-technology, they fit battery models to the data and infer for example, cell individual internal resistance from them.

III. Modelling Data-Driven Digital Twins of EV Batteries

In this Section, we outline the RAMI 4.0 background (Section III.A) and then, we describe the proposed data-driven digital twin of EV batteries (Section III.B) which follows the Industry 4.0 principles. Section III.C presents the data model that structures the related data and information and represents the connection between the different entities that take place in the battery data analytics.

A. Background on RAMI 4.0 and Digital Twins

The German Federal Ministry of Education and Research defines Industry 4.0 as “the flexibility that exists in value-creating networks by the application of Cyber Physical Systems (CPS)” [14]. In this context, RAMI 4.0 is based on a three-dimensional model consisting of the Architecture Layers, Life Cycle and Value Stream, and Hierarchy Levels dimensions. RAMI 4.0 considers any technical asset as an entity that can be represented in the digital world to conform an I4.0 component. The main scope of each dimension is described below.

Architecture Layers: The Architecture Layers enable the development of Industry 4.0 software solutions in a consistent way so that different operations are interconnected, taking into account the physical and the digital world. There are six (6) architecture layers, as shown in Fig. 1:

Asset Layer: It represents the reality, i.e. the physical assets and the users.

Integration Layer: It provides information related to the assets in the appropriate format by connecting elements and people with information systems.

Communication Layer: It provides standardization of communication by means of uniform data format and deals with the physical support of information processing.

Information Layer: It provides pre-processing of events and execution of event-related rules by enabling their formal description for the interpretation of the information.

Functional Layer: It enables the formal description of functions and creates the platform for horizontal integration of various functions.

Business Layer: It ensures the integrity of functions in the value stream and enables mapping business models and the outcomes of the overall process.

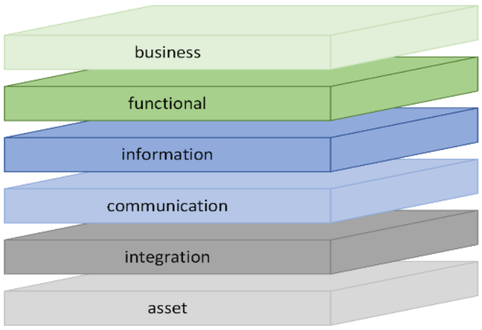

Fig. 1. The Architecture Layers of RAMI 4.0.

Life Cycle and Value Stream: The second axis in RAMI 4.0 represents the lifecycle of products and systems and is taken from the IEC 62890 standard [15]. The product lifecycle model introduces a differentiation between product type and product instance.

Hierarchy Levels: The third axis of RAMI 4.0 is the hierarchical representation of the different functional levels of the factory, based on the IEC 62264 [16] and IEC 61512 standards. These hierarchy levels are: Connected World, Enterprise, Site, Area, Work Centers, Work Units or Station, Control Device, Field Device, and Product.

In this context, a digital twin is the container for integrating information, executing operations, and producing data describing its activity which can be in different formats, from different software tools, and not necessarily deployed in one central repository [17]. Both the physical and the digital twins are equipped with networking devices to guarantee a seamless connection and a continuous data exchange between a generic physical system (or process) and its respective Digital Twin [14], while they are able to support predictive analytics [18][19]. The digital twin is implemented by the Asset Administration Shell (AAS). The AAS consists of a number of sub-models in which all the information and functionalities of a given asset - including its features, characteristics, properties, status, parameters, measurement data, and capabilities - are described [20]. To facilitate the design of these sub-models, there is the need of applying the RAMI 4.0 Architecture Layers in alignment to the AAS architecture, as depicted in Fig. 2. The Asset and the Integration Layer deal with the physical asset ("Thing"), while the rest of the Layers deal with the digital world ("Administration Shell").

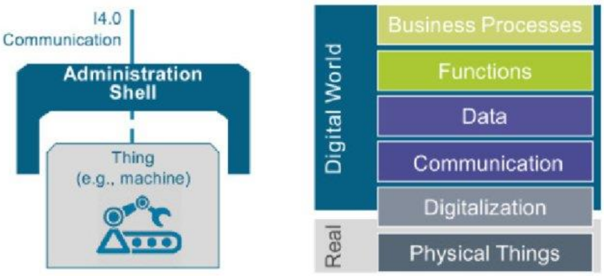

Fig. 2. The physical architecture of a battery pack of an EV (Source: [14]).

B. Modelling EV Battery Digital Twin according to RAMI 4.0

In this Section, we present the modelling of the EV battery data-driven digital twin in the frame of RAMI 4.0. Fig. 3 depicts the various components of the BMS and the respective software components of the digital world. The following sub-sections describe each Architecture Layer out of those depicted in Fig. 1 for the Li-Ion batteries aiming at supporting predictive analytics.

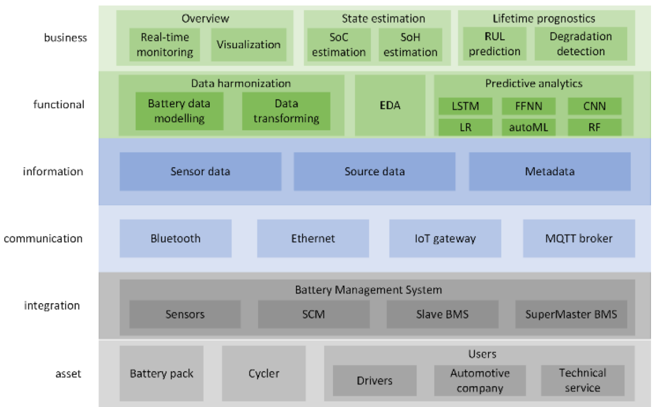

Fig. 3. The EV battery physical and software components in the context of the RAMI 4.0 Architecture Layers.

1) *Asset Layer:* In the EV, the power for its operations is provided by a high voltage Battery Pack (BP) that is composed of different components, organized in levels. The first level consists of the cells, which are the smallest, packaged components in a BP. The cells are organized in modules, that constitute the second level. Accordingly, the modules compose the pack, the third level of a BP. The units of each level are connected in a parallel or serial configuration. The number of components that constitute its level is configurable and depends on the application that it is used for. In the Fig. *3*, the elements of the physical architecture of the BP are categorized in the Asset Layer.

In the same layer, the proposed approach includes the battery cycler. A battery cycler is an experimental instrument that can analyse battery function through charge and discharge cycles and measure the cells response over time. Cyclers are utilized in cycling ageing tests on cells. A cycler contains many channels that are attached to cells. Moreover, the cycler is wrapped in an environmental chamber in order to perform the cycling tests in a stable environmental temperature. The cyclers support the configuration of various specifications that can guide different test procedures and thus provide a controlled environment for testing [9]. During the defined tests, a cycler can measure several parameters such as capacity, efficiency and self-discharge. This layer also includes the users, i.e. the drivers, the automotive company and the technical service.

2) *Integration Layer:* One of the key components of electric vehicles, the Li-ion Battery Management System (BMS) is crucial to the industrialization and marketization of electric vehicles [21]. BMS is a system designed to monitor and optimize battery behavior. It is also capable of handling different cell types and chemical compositions and managing various settings, including multiple packs. The main functions of the BMS include battery data acquisition, modeling and state estimations, charge and discharge control, fault diagnosis and alarm, thermal management, balance control, and communication [21]. Additionally, BMS can calculate the SoX and therefore increase the overall safety and performance of the system [9]. BMS software is designed to be accessed and updated remotely in order to provide remote maintenance functionality and anticipate faulty situations beyond the testing phase [9]. To this end, the Battery Management System represents the Integration Layer, as depicted in Fig. *3*. This layer consists of the elements that control and monitor the components in each level of BP.

Each level of BMS's components is equipped with sensors that can record measurements. The integration of sensors in the interconnection at cell level can enable additional advanced functionalities of the system [IISA 2022]. Beyond the data that arise from sensors, additional data can be obtained from each level of the BP. As a result, data from sensors, parameters that describe the function of the component and aggregations of both categories of data can be available. Starting from the first layer of the BP, each cell is connected to a Smart Cell Manager (SCM) that can record its measurements and transmit data to a Slave BMS. Respectively, the packs of the third level of a BP transmit data to a super Master BMS.

Although the sensors integration advances the data availability and the data analysis capabilities, the installation of multiple sensors increases the wiring complexity in BPs and thus increases the manufacturing cost. BPs include special systems to damp oscillations and absorb vibrations providing high reliability for all physical connections among sensors, transmitters and receivers, but also having a strong impact on

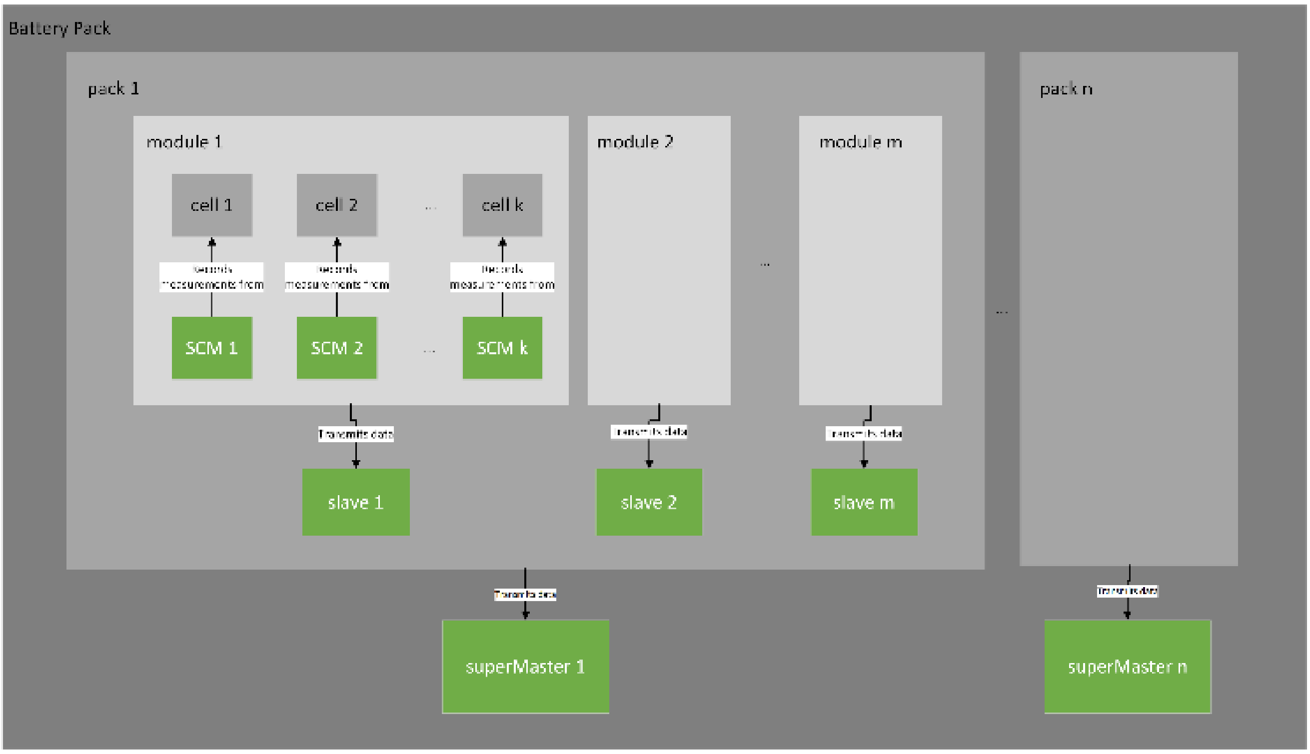

Fig. 4. The physical architecture of a battery pack of an EV.

the final production cost. The approach of the Smart Cell Manager (SCM) consists of the implementation of a system that can be embedded directly into each individual cell.

The slave BMS receives data from the module and sends them to an IoT gateway that can initiate the process of storing, processing and analyzing the data.. The captured data usually refer to sensor data, such as voltage, current and temperature but also calculated aggregations of these parameters, and parameters that describe the function of the module such as the total discharge and calculate additional features such as State of Charge (SoC). Moreover, the available aggregations depend on either the number components that compose the module or the time. For example, a slave can calculate the average value of the voltage of the cells that compose the module or the average value of a parameter during the aggregation time, which has been configured from the gateway. Additionally, a slave BMS transmits to the gateway the data that are available from the SCMs.

Following a similar approach, the super Master BMS receives data from the pack and sends them to the IoT gateway. The Super Master BMS supports different functionalities from the slave BMS. In contrast to the slave that handles data that characterize the state of the module and the respective cells, the Super Master BMS records data that describe the overall state of the pack. Consequently, the parameters that can be transmitted from a Super Master BMS are either aggregations of related parameters or variables that describe the overall state of the pack.

3) Communication Layer: In order to provide the advanced data analytics capabilities of the Battery Digital Twin, data move from the physical elements of the BP to the higher layers through the Communication Layer. This layer consists of the technologies and enablers that facilitate the transmission of the data. Specifically, Bluetooth, Ethernet and Controller Area Network (CAN) bus are the technologies considered that among others enable the transmission of the cells' data from several SCMs to a slave BMS and the transmission of the data to the gateway. The gateway receives the data, converts it into MQTT frames and sends them over the vehicle and public network to a MQTT broker. The latter is hosted on a server for storing data and performing data analysis. Message Queuing Telemetry Transport (MQTT) is a lightweight protocol for Internet of Things (IoT) applications to transfer data to and from a cloud following a publish-subscribe-methodology (see [3]). Communication is performed when the vehicle is online, e.g., while driving. The gateway establishes the connection to the MQTT broker, publishes data periodically and subscribes to results and firmware update topics.

4) Information Layer: The data are communicated through the Communication Layer to the higher levels of the Digital Twin. Following RAMI 4.0, the Information Layer holds the total information upon which the proposed Digital Twin can be built. This information is encapsulated in the data that are available in the considered setup, the data generated by sensors, the additional source data that are generated by the EV or the cycler and any additional metadata related to these. In order to provide a short overview of these data, Fig. 5 and Fig. 6 present the parameters that can be measured and recorded from a slave and a Super Master BMS. These parameters refer to either sensor data or source data. As depicted in Fig. 5, the data that are recorded from a slave BMS can be sensor data such as voltage and temperature and they are measured by the SCM in the level of cells. Moreover, through the slave BMS additional features such as SoC and total discharge capacity of the module, which is an indicator for the State of Health (SoH) are calculated. Furthermore, additional aggregations of these parameters such as the minimum and the maximum value of voltage or the average value of SoC between the cells during the time elapsed between the calculations can be calculated. Finally, in this level the information provided can include parameters such as the balance number, which indicates the number of cells that take part in the procedure of balancing. All the above parameters refer to source data.

In a similar way, sensor and source data are recorded from a Super Master BMS as shown in Fig. 6. The difference from the data of a slave BMS, is that the first ones describe the overall condition and functionality of the BMS while the latter, the data describe in more detail its condition. Another difference is the fact that from the BP additional metadata can be obtained, such as the firmware needed for the communication between the different elements of the BP.

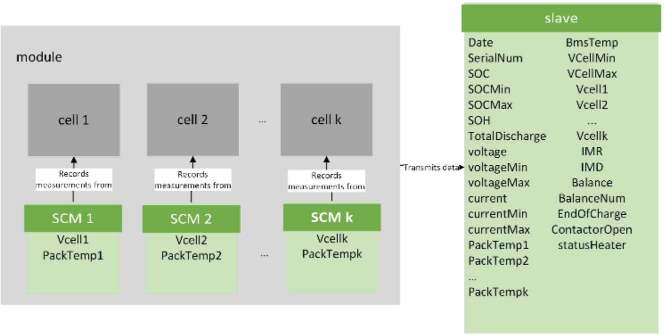

Fig. 5. Data produced by a module of a pack.

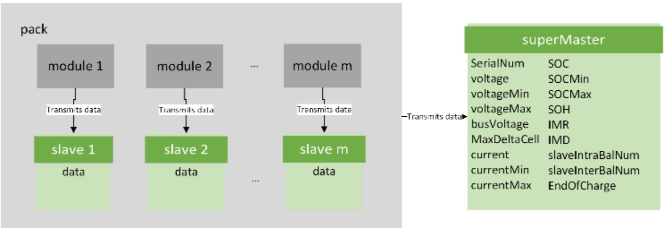

Fig. 6. Data produced by a pack of a battery pack.

In a similar manner, a plethora of data arises from the procedure of cycling through cyclers, with the difference that the data which arise from cycling refer only to cell level. Primarily, cycling can provide sensor data such as cell's voltage, current, temperature, charge and discharge capacity, energy and power. Furthermore, cyclers provide source data. Additional features that can be obtained from these conditions are the internal resistance, the inclination of the voltage curve dV/dt and the function's status, which describes if the cell operates through charging, discharging or resting. By the same token, further parameters can be calculated such as the cycle number, which participates in the analysis of the cell's health. Moreover, metadata that refer to scalar values such as the temperature of the chamber and the charging protocol which has been applied to the test and the cycle life of the cell can be collected. According to the cycler's functionality, it is feasible to predefine some parameters of the testing procedure and in this way, it is possible to approach more realistic driving conditions resulting in more representative data. In detail, the charging protocol can be defined by the value of the C-rate that is going to be applied and the number of different C-rates if it is desirable. Moreover, the temperature of the chamber can be designated in order to simulate the driving conditions in different locations.

5) Functional Layer: The Information Layer represented by the three types of data mentioned above includes various datasets from several data sources that may follow different data models. Consequently, the procedure of data harmonization is crucial for any data analysis,. Harmonization refers to the action of gathering data with different types, from different sources, organizing and unifying them in a suitable way for the analysis. The main goal of this procedure is the unification of the data structure that will enable the processing of multiple datasets in a unified way. Additionally, the development of a harmonization process disassociates the analytics process from the specific dataset and its characteristics.

In the same context, the inclusion of the minimum required variables for the implementation of the analysis can be ensured and the additional required features can be constructed independently from the format of the initial dataset. An important aspect of this procedure is the fact that the variables under examination should be based on the same hypothesis. For example, if the discharging phase of a battery is represented as a negative number in the majority of the available datasets, through the harmonization process, this constraint can be applied to all the datasets in a unified way, by implementing the corresponding transformations. Other aspects that harmonization can help refer to more technical challenges that are presented in the data processing steps. As an example, consider the datasets that contain useful data in an unreadable of unsuitable for further analysis format, i.e. objects or strings, or even some time-relevant information that is recorded in a different format per dataset. Finally, another critical aspect that harmonization can solve is the different sampling rates that the different datasets follow.

Therefore, there are different data sources for battery data analytics and the datasets contain both, similar and different parameters, depending on the component from which they are collected. Under these circumstances, the need for a unified and suitable battery data model defining the entities that are able to describe most of the elements related to the battery data analytics in a common way, is clear. To this end, this present work proposes the battery data model described in Section C, that can facilitate the harmonization process and optimize the data fusion that will enable better data analysis results.

The procedures of battery data modelling and data transforming play a key role to homogenizing the data. Consequently, the battery data predictive analytics can be independent of the characteristics of each dataset. In other words, the analytics can be flexible, dynamic and generalized by using and combining knowledge from different data sources. Another crucial procedure for the examination of the data is the Exploratory Data Analysis (EDA) of each dataset. Through this method the behavior of each variable and the relation among them can be mapped and defined. Furthermore, the correlations among the variables existed in a dataset can also be defined, while some fundamental statistical features can be calculated. Finally, additional useful characteristics of a dataset can be calculated, such as the lifecycle or the time duration of the battery that it describes.

These methods are the interim steps between the data and predictive analytics. In a similar way, the analytics take place in a dynamic way using data from the different data sources and containing several algorithms and models that can be used for the desirable approaches. A variety of models can be employed for battery data analytics. Indicative examples that can be applied are Long-Short Term Memory neural networks

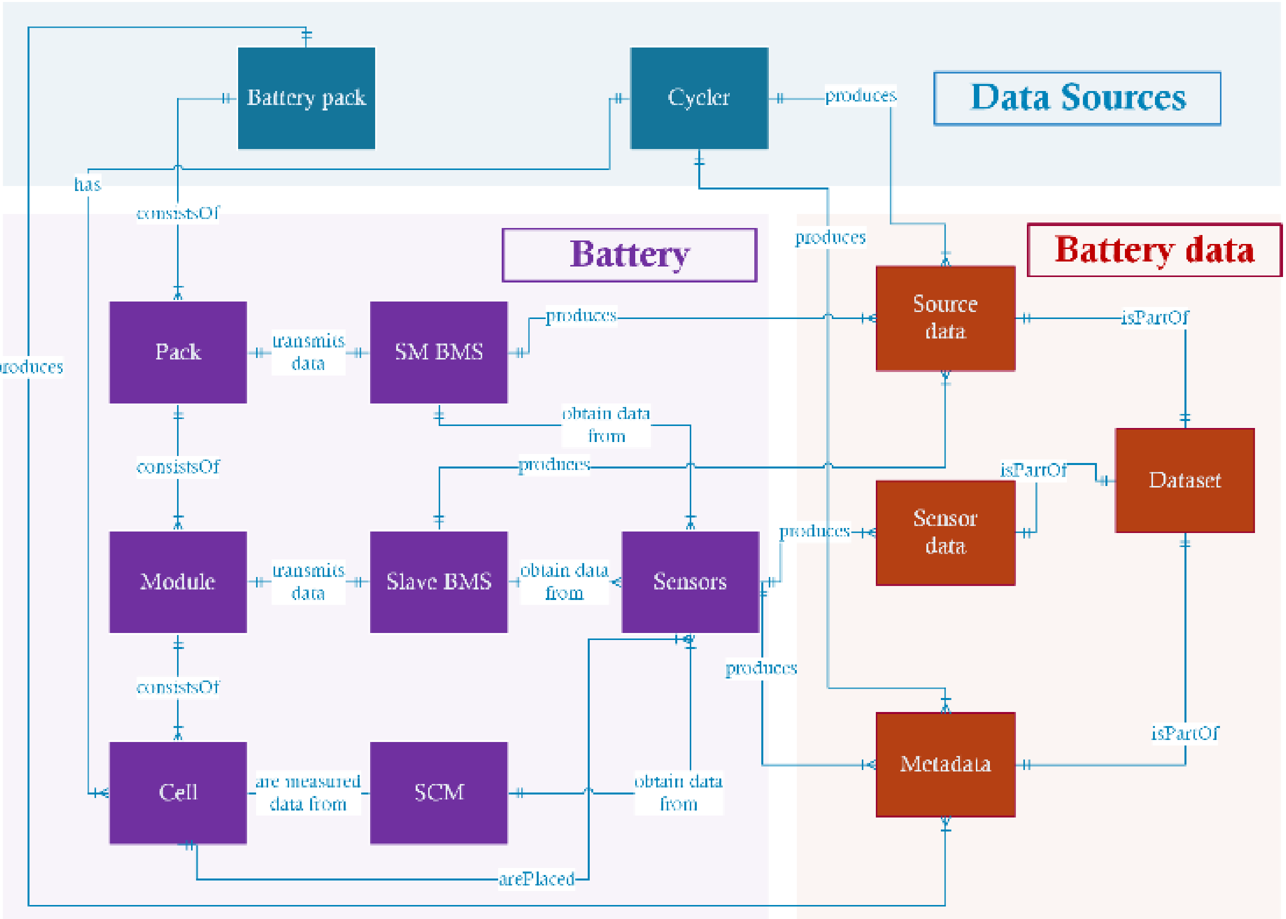

Fig. 7. The battery data model.

(LSTM), Feed Forward Neural Networks (FFNN), Convolutional Neural Networks (CNN), Linear Regression models (LR), Random Forest models (RF). Furthermore, the automated Machine Learning (autoML) methods can be applied to provide optimal models following a more structured experimental process.

6) Business Layer: The results of the predictive analytics on battery data answer several objectives that are of interest of the end users of the digital twin. Some of these are depicted in the Business Layer of the Fig. 3. To begin with, the recipient of the analytics' results can observe an overview of the variables in real time by monitoring their behavior through time. Moreover, the visualization of features and aggregations, as well as the advanced analytics results that have been calculated for a specific dataset can be available. An important focus of the battery data analytics is the state estimation, since it can describe in almost real time the condition of battery's health. A common estimation for batteries in EVs is the State-of-Charge (SoC), which is critical to their safe and reliable operation since this quantity reflects a vehicle's remaining driving range [22] as well as the remaining energy inside a battery during operation [23]. By definition, SoC refers to the capacity of the battery in its current state as a percentage of the capacity in its fully charged state [24]. As a consequence of this definition and due to the battery's degradation, the SoC does not present a linear behavior through his entire lifetime. In a similar way, State-of-Health (SoH) is an essential indication that determines the battery aging and has a value ranging between

0 and 100% [25]. It describes the capacity of the battery in its fully charged state compared to the nominal capacity [24], as it is defined by the constructor of the battery. It is a non-linear quantity and it is highly dependent on the volatility of the loading profiles, ambient temperature, depth of discharge and the way of discharge [25]. The long-term predictions for the batteries are presented by the lifetime prognostics. Specifically, the Remaining Useful Life (RUL) of the battery refers to the remaining cycles until its End Of Life (EOL). This feature is defined usually as the cycle in which the battery will reach the 80% of its nominal capacity. Finally, the detection of the degradation refers to the cycle that the battery will present a smaller amplitude of its nominal capacity. Both of these features do not present the same behavior in all the batteries since they depend on diverse aging mechanisms and their behavior, as well as on the initial condition of the battery.

C. Battery Data Modelling

Fig. 7 shows the aforementioned data model and represents the connection as also the relationship between the different entities that take place in the battery data analytics. This entity relationship diagram is designed to be read from left to right and in this way is indicated the relationship between two entities. The main components of this model are the Data Sources, the Battery, the Battery Data and the Analytics. Each component consists of several entities and each entity is described by several attributes. Every entity has a unique attribute about its identity (Id) and several others that describe it.

As analysed in the physical architecture of the battery in a EV, a BMS consists of several battery packs, a battery pack consists of several modules and each module consists of several cells. Each of these components transmits data or are measured data from it by superMaster BMS, slave BMS and SCM, respectively. A superMaster BMS produces source data but also obtain data from sensors, so it produces and sensor data. Correspondingly, a slave BMS produce both source and sensor data. On the contrary, from SCM it can be obtained only sensor data. A cycler has cells in which are placed sensor and so can be produced sensor data. Moreover, a cycler produces source data and metadata. All the types of data are part of a dataset which is connected to the analytics component and used by it so to be performed the battery data analytics. Finally, a cycler has a cycler model that describes its specifications and the circumstances under which the cycling test has been performed.

IV. Implementation and Deployment of the EV Battery Data-Driven Digital Twin

The digital twin was implemented and deployed in an EV with a BP which was composed from two packs, each pack from sixteen modules and each module from six cells. In correspondence with the physical architecture of this BP, data were recorded from two superMaster BMSs, sixteen slave BMSs and six SCMs. The same cells were placed in two different cyclers for cycling.

Through the communication elements were obtained sensor data, source data and metadata from both BP and cyclers. The data from the cyclers had different structure and formulation since they came from different types of cyclers. The cycling tests were conducted with different charging protocols, so the tests would be more representative of real driving conditions. Moreover, the datasets that produced from these tests, contained both same and different variables. An important note here is that even the same variables were recorded in a different way. For example, the capacity of the cells in one case was recorded through two variables, the charge and the discharge capacity when in the second one, the capacity was recorded from one variable, the overall capacity. These conditions were taken into account in the data harmonization process and confirm the need for a common and sustainable way of data processing and exploitation. Finally, some variables were built during data processing since they were not contained in all the datasets, such as the cycle number.

We used additional cycling data from open datasets in order to enhance the diversity of input data. These data also had to be processed by data harmonization procedure since they had different structure and content and they conducted in different test conditions. Likewise, the data that were obtained from each level of BP were in need of data harmonization since they referred to different levels of BP and contained aggregations or additional calculations.

The data harmonization contains several sub-procedures. First, the baseline for these procedures is the battery data model, which points out the relation between the different components of battery data analytics and how they are described. Second, in the context of data harmonization and in conjunction with the data model, we implemented different functions of data transformation on each dataset. In more detail, the variable names were transformed into common ones and their structure was converted to readable dataframe, in which the first line contained the columns' name and the identification number of measurements was defined to be the index. Some variables were constructed such as the cycle number or harmonized such as the capacity. Additionally, the datetime that was recorded, remodeled in a common way of presentation in order to be comparable and understandable in different plots of data. Furthermore, several datasets contained data in the form of objects and so they were not processable. For this reason, they converted into a numeric or character form depending on the variable that they present. Finally, the sampling rate in several datasets was not stable, so it had to be converted to a steady one or between datasets the rate was different so it had to be unified.

We also conducted an EDA for each dataset in contemplation of visualizing and understanding the behavior of several variables through raw data as well as statistical features analysis. For the predictive analytics, we used several algorithms and models in a flexible way for the overview component, state estimation and lifetime prognostics.

Fig. 8 presents an overview of the datasets that were used in battery data analytics and came from the cycling tests. It contains the number of analyses that have conducted as well as the number of datasets that have been used for the analytics. The following graphs describe the main characteristics of the datasets. The first graph refers to the lifetime of the dataset and depicts if the dataset contains data from the entire lifecycle of the battery, from a part of it or from a single cycle. Moreover, important parameters, which are expressed as metadata, are: the temperature in which is conducted the data collection, the C-rate that was implemented, and the duration of the experimental procedure. These parameters are presented in the overview graph.

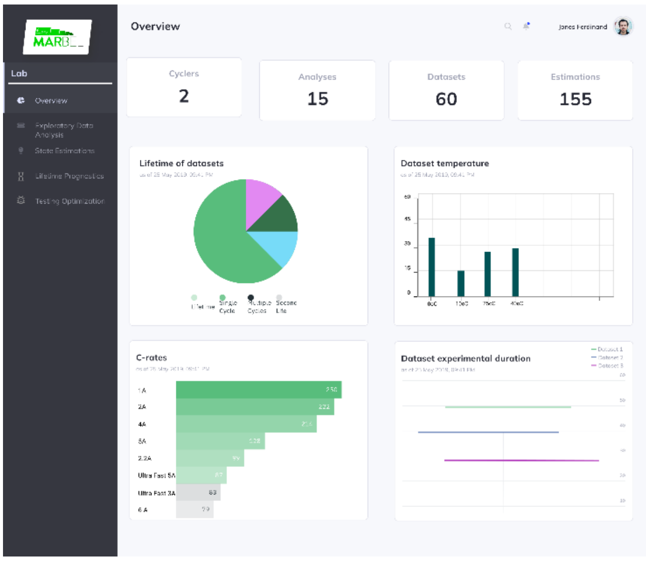

Fig. 8. Overview of battery data used in analytics.

V. Conclusions and Future Work

Li-Ion batteries have been widely applied as energy storage systems, such as EVs. There is an increasing research interest on ML models and algorithms dealing with predictive analytics for SoC, SoH, and RUL predictions. With the recent advances in understanding battery degradation, modelling tools and diagnostics, there is an opportunity to fuse this knowledge with emerging ML techniques towards modelling a battery digital twin, following the principles of RAMI 4.0. In this paper, we modelled, designed, and developed an EV battery data-driven digital twin for supporting predictive analytics in order to tackle the big data-rich, complex, and uncertain Li-Ion batteries behaviour in a holistic way.

During the last years, Transfer Learning (TL) has evolved into an efficient and powerful data-driven tool for smarter battery management [26]. In our future work, we will extend the herein presented digital twin in order to develop TL approaches in order to take advantage of the multitude of data derived from cyclers so that they are used for training ML models that can provide predictions on BMSs, where typically the data availability is limited.

Acknowledgment

This work is partly funded by the European Union's Horizon 2020 project MARBEL "Manufacturing and Assembly of Modular and Reusable EV Battery for Environment-friendly and Lightweight Mobility" (Grant agreement No 963540). The work presented here reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] X. Hu, C. Zou, C. Zhang and Y. Li, "Technological developments in batteries: a survey of principal roles, types, and management needs," IEEE Power and Energy Magazine, vol. 15(5), pp. 20-31, 2017.

[2] M. Fichtner, K. Edström, E. Ayerbe, M. Berecibar, A. Bhowmik, I.E. Castelli, ... and M. Weil, "Rechargeable Batteries of the Future—The State of the Art from a BATTERY 2030+ Perspective," Advanced Energy Materials, pp. 2102904, 2021.

[3] Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li, and Z. Chen, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," Renewable and Sustainable Energy Reviews, vol. 131, pp. 110015, 2020.

[4] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, ... and H. E. Hoster, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," Renewable and sustainable energy reviews, vol. 113, pp. 109254, 2019.

[5] L. Merkle, A. S. Segura, J.T. Grummel and M. Lienkamp, "Architecture of a digital twin for enabling digital services for battery systems," In 2019 IEEE international conference on industrial cyber physical systems (ICPS), pp. 155-160, May 2019.

[6] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte and D. U. Sauer, "Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation," Journal of energy storage, vol. 30, pp. 101557, 2020.

[7] B. Wu, W. D. Widanage, S. Yang and X. Liu, "Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems," Energy and AI, vol. 1, pp. 100016, 2020.

[8] C. Vidal, P. Malysz, P. Kollmeyer and A. Emadi, "Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art," IEEE Access, vol. 8, pp. 52796-52814, 2020.

[9] A. Fouka, K. Lepenioti, A. Bousdekis and G. Mentzas, "A Machine Learning Framework for Li-Ion Battery Lifetime Prognostics," In 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA), pp. 1-8, July 2022.

[10] A. Bousdekis and G. Mentzas, "Enterprise Integration and Interoperability for Big Data-driven Processes in the frame of Industry 4.0. Frontiers in big Data", vol. 4, pp. 644651, 2021.

[11] G. Bhatti, H. Mohan and R. R. Singh, "Towards the future of smart electric vehicles: Digital twin technology," Renewable and Sustainable Energy Reviews, vol. 141, pp. 110801, 2021.

[12] W. Wang, J. Wang, J. Tian, J. Lu and R. Xiong, "Application of digital twin in smart battery management systems," Chinese Journal of Mechanical Engineering, vol. 34(1), pp. 1-19, 2021.

[13] L. Merkle, M. Pöthig and F. Schmid, "Estimate e-Golf battery state using diagnostic data and a digital twin," Batteries, vol. 7(1), pp. 15, 2021.

[14] Platform Industrie 4.0. Details of the Asset Administration Shell Part 1 - The Exchange of Information Between Partners in the Value Chain of Industrie 4.0 (Version 2.0). (2019). Available online at: https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/Details-of-the-Asset-AdministrationShell-Part1.html (accessed May 08, 2023).

[15] International Electrotechnical Commission. Life-Cycle Management for Systems and Products Used in Industrial-Process Measurement, Control and Automation. Geneva: IEC (2017).

[16] International Electrotechnical Commission. Enterprise-Control System Integration-Part 3. Activity Models of Manufacturing Operations Management. Geneva: IEC (2016).

[17] T. Catarci, D. Firmani, F. Leotta, F. Mandreoli, M. Mecella and F. Sapio, "A conceptual architecture and model for smart manufacturing relying on service-based digital twins," in 2019 IEEE International Conference on Web Services (ICWS), pp. 229–236, 2019.

[18] S. Zillner, J. A. Gomez, A. G. Robles, E. Curry, C. Södergård, … and N. Boujemaa, "Data-Driven Artificial Intelligence for European Economic Competitiveness and Societal Progress: BDVA Position Statement," Brussels: BDVA, November 2018.

[19] B. R. Barricelli, E. Casiraghi and D. Fogli, "A survey on digital twin: definitions, characteristics, applications, and design implications," IEEE Access, pp. 7-2953499, 2019.

[20] H. Bedenbender, M. Billmann, U. Epple, T. Hadlich, M. Hankel, … and R. Heidel, "Examples of the Asset Administration Shell for Industrie 4.0 Components–Basic Part," ZVEI White Paper. Frankfurt: ZVEI, 2017.

[21] M. Naguib, P. Kollmeyer, and A. Emadi, "Lithium-ion battery pack robust state of charge estimation, cell inconsistency, and balancing," IEEE Access, vol. 9, pp. 50570-50582, 2021.

[22] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries," IEEE Transactions on Industrial Electronics, vol. 65(8), pp. 6730-6739, 2017.

[23] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," Renewable and Sustainable Energy Reviews, vol. 114, pp. 109334, 2019

[24] M. F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," Nature Machine Intelligence, vol. 2(3), pp. 161-170, 2020.

[25] E. Chemali, P. J. Kollmeyer, M. Preindl, Y. Fahmy, and A. Emadi, "A Convolutional Neural Network Approach for Estimation of Li-Ion Battery State of Health from Charge Profiles," Energies, vol. 15(3), pp. 1185, 2022

[26] K. Liu, Q. Peng, Y. Che, Y. Zheng, K. Li, R. Teodorescu, ... and A. Barai, "Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects," Advances in Applied Energy, pp. 100117, 2022.