BattINFO Converter: An Automated Tool for Semantic Annotation of Battery Cell Metadata

Nukorn Plainpan, Simon Clark, and Corsin Battaglia*

Driven by the global demand for batteries with improved performance, cost efficiency, and sustainability, battery research has expanded rapidly during recent years, generating a surge in data which offers the potential to accelerate the discovery of new battery materials, optimize battery cell manufacturing processes, and enable predictive modeling of battery aging. However, the lack of adoption of a community-wide battery data reporting standard results in significant challenges in integrating battery research data across different studies, limiting the potential for automated analysis and large-scale comparative studies. To foster the widespread adoption of a standardized battery data reporting framework, BattINFO converter, an open-source, Python-based web application is developed that facilitates the conversion of battery cell data and metadata from an Excel table into a structured, semantically annotated, linked data format compliant with the BattINFO ontology. The BattINFO ontology defines clear relationships between data and metadata items, thereby enhancing machine readability, interoperability, and discoverability. By facilitating the adoption of a standardized approach to data annotation, BattINFO converter supports more efficient data sharing, analysis, and integration within the battery research community.

1. Introduction

Batteries represent a key enabling technology for realizing net-zero carbon emission.[1–4] Despite recent advancements in battery research, the quest to develop even more efficient and longer lasting, yet cheaper, and more sustainable batteries is still ongoing.[5–8] Publicly funded academic research in the field of batteries has been significantly supplemented by industrial research and development, largely financed by private and corporate venture capital investments totaling 2 billion USD for early-stage and >20 billion USD for growth-stage battery manufacturing and recycling from 2020 to 2023.[2] These investments have catalyzed the development and the commercialization of a number of cutting-edge innovative battery technologies for electric vehicles and stationary energy storage systems.

Intense research and development efforts in academia and industry during recent years have also led to the generation of enormous volumes of data, ranging from battery materials properties to battery cell manufacturing parameters to battery pack performance metrics and battery materials recycling efficiencies. While such data holds immense potential for promoting technological breakthroughs, it remains largely inaccessible, partially due to proprietary reasons, but mainly due to the lack of standardization. The lack of standardization of data structure and format, the fragmentation of data across various stakeholders, and the absence of advanced data analytics prevent the effective use of this surge in data. As a result, valuable insights that could accelerate battery development remain untapped, while harnessing this data effectively presents a transformative opportunity for accelerating battery innovation.[9–13]

The creation of the BattINFO ontology was driven by the need to address these challenges and unlock the vast potential of the data generated by the battery research community.[14–16] The BattINFO ontology provides a framework for human-readable and machine-interpretable semantic data annotation, enabling the generation of linked data connecting diverse datasets thereby facilitating complex queries and automated rapid retrieval of relevant information.[17] Annotating battery data according to the BattINFO ontology ensures that data are findable, accessible, interoperable, and reusable aligning with the FAIR data principles, promoting long-term data sustainability and usability. Properly structured and annotated data also facilitate leveraging the potential of modern artificial intelligence methods, leading to more accurate models, better predictions, and deeper insights.

Despite the clear benefits of structuring data in compliance with the BattINFO ontology, several potential barriers to its widespread adoption exist.[18] The financial and time resources required to transition from the currently established data management practices or current proprietary data structures to comply with the BattINFO ontology may be seen as prohibitive. Also, despite a clear benefit

N. Plainpan, C. Battaglia
Empa - Swiss Federal Laboratories for Materials Science and Technology
Ueberlandstrasse 129, 8600 Dübendorf, Switzerland
E-mail: corsin.battaglia@empa.ch

N. Plainpan, C. Battaglia
Department of Information Technology and Electrical Engineering
ETH Zurich
Gloriastrasse 35, 8092 Zurich, Switzerland

N. Plainpan, C. Battaglia
School of Engineering
Institute of Materials
EPFL
Station 15, 1015 Lausanne, Switzerland

S. Clark
New Energy Solutions
SINTEF Industry
Sem Sælands vei 12, 7034 Trondheim, Norway

Batteries & Supercaps

Research Article
doi.org/10.1002/batt.202500151

Chemistry
Europe
European Chemical
Societies Publishing

to the community as a whole, the short-term benefit to individual researchers appears low. However, often it is rather a lack of technical expertise in data management and ontology-based data management that hinders the implementation of an effective use of the BattINFO ontology within an academic or industrial battery lab. Attempting to semantically annotate battery data and metadata manually, although human readable, remains a daunting error-prone task even for an expert and thus requires automation.

To eliminate these adoption barriers, we introduce here BattINFO converter (https://battinfoconverter.streamlit.app/), an open-source, Python-based web application designed to automate the conversion of battery data and metadata from an Excel file into a structured, semantically annotated, JavaScript Object Notation for Linked Data (JSON-LD) format compliant with the BattINFO ontology. We choose the battery coin cell format as example due to its widespread use in academic and industrial research labs,[19–23] providing a filled example Excel file and its corresponding linked data file generated by the BattINFO converter. By employing BattINFO converter, battery researchers without data science and programming background can seamlessly convert their battery cell data into an interoperable, semantically annotated, linked data format that can be included as supplementary information in scientific publications or be shared on open-access data repositories, thereby supporting greater transparency, reproducibility, data sharing, and knowledge integration in the battery research community.

2. BattINFO Converter Implementation

BattINFO converter (**Figure 1**A) is a Python-based web application that automates the semantic annotation of battery cell metadata

without the need for technical knowledge in semantic annotation or ontologies. The user can simply fill out the metadata template (the Excel template – Figure 1B, this will be discussed in more detail in the 'Battery Metadata Template in Excel Format' section) and upload this to the web application (the user can simply drag and drop the filled Excel template file), and the web application will convert the Excel template into a corresponding JSON-LD file (Figure 1C).

To avoid unnecessary adoption barriers, the user is not required to compile or install any additional software. BattINFO converter is a web application. Hence, the user only needs a web browser and an internet connection. The source code of this application is open-source (the link to the source code is provided in the 'Code Availability' section). Therefore, the user can verify how the application works and can be assured that the application will not store nor send out any information the user has uploaded through the web application to anywhere. The concerned user can also choose to take the source code and run BattINFO converter locally to ensure data privacy.

3. Battery Metadata Template in Excel Format

A schema is a formal description of a database's structure that specifies the data types, relationships, and constraints governing the data (in this case coin cell metadata). One can think of a schema as a template to fill in the metadata. Several file types such as .xml or .json can be used as a template for this purpose[24] These file types share many advantages for use as a template, such as being non-proprietary, human-readable, and also designed for machine-to-machine communication. Nevertheless, these file types are text-based, which are more prone to errors
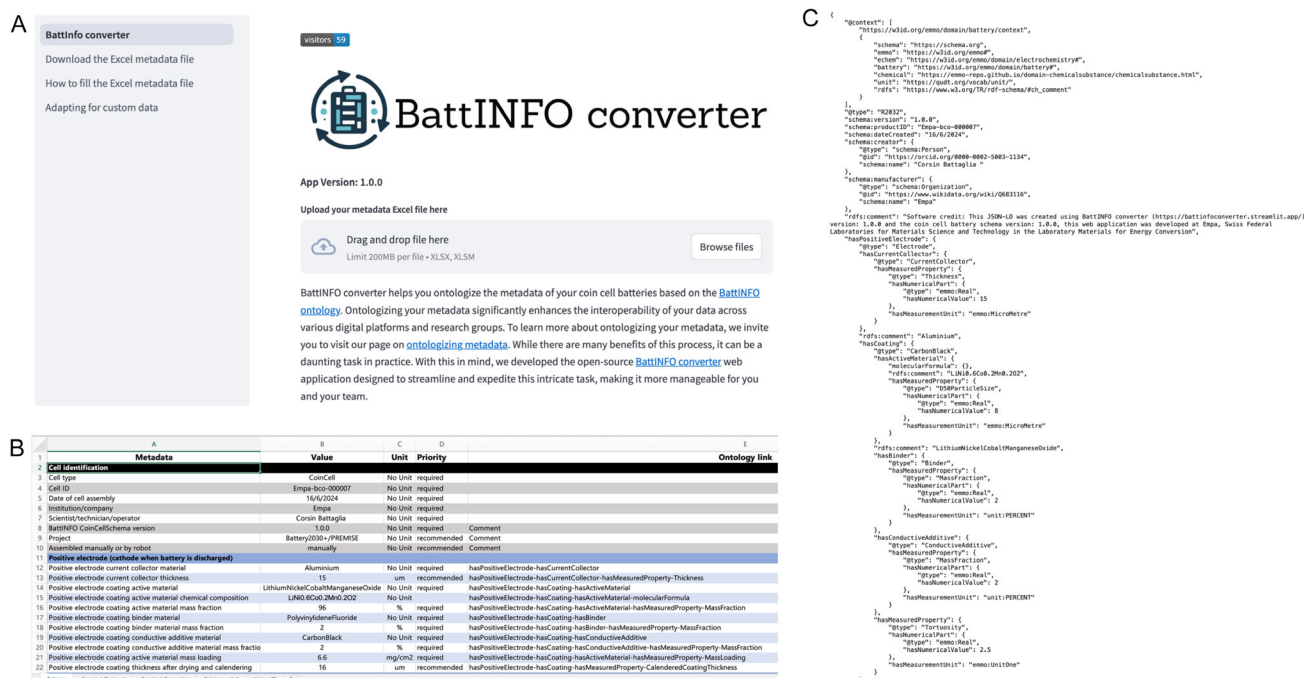

Figure 1. A screenshot of A) the BattINFO converter web application, B) the Excel template version 1.0.0, and C) the corresponding JSON-LD file.

Batteries & Supercaps

Research Article
doi.org/10.1002/batt.202500151

Chemistry Europe
European Chemical
Societies Publishing

and are more tedious in nature for the non-technical users to fill in. Thus, we look for a tabulated file format to make schema editing more intuitive. The ideal file type would be a file type that is familiar to most non-technical users. With these criteria in mind, we chose .xlsx (tabulated file from Microsoft Excel) as the file type of choice. Despite being proprietary, there are open-source solutions to interact with Excel files, such as LibreOffice.[25] Excel is already widely used by the majority of users, which helps lower the learning curve when using .xlsx as a file type for a metadata template. Excel also offers many functionalities that allow users to have both controllability and customizability of the template through Excel. We acknowledge that the .xlsx format comes with certain disadvantages, including OS-dependent encoding. Nevertheless, we believe that its advantages and widespread availability far outweigh these disadvantages, which is why we chose the .xlsx format for our metadata template. These advantages make .xlsx a compelling file type of choice.

The example of the Excel template can be found on the BattINFO converter web application page. We have included both the blank Excel template and an example template that is filled with one of the coin cells produced in our lab as an example. The full documentation of the structure of the Excel template and how to fill or make adjustments to it is described on the BattINFO converter web page (https://battinfoconverter.streamlit.app/). In brief, the Excel template consists of five tabs, but the tab that will be most relevant to the general user is the 'Schema' tab. The other four tabs are used to store settings and information required for our web application to generate a semantically annotated JSON-LD file. The user will only need to make adjustments to the information in these four tabs if one aims to create a custom Excel template for their custom cells.

The Schema tab contains a predefined table of metadata items. In this current work, we present the predefined metadata form of a coin cell battery. The list contains 133 metadata items that we believe are relevant and important for each fabricated coin cell battery. The metadata items are divided into six main categories: cell identification, positive electrode, negative electrode, electrolyte, separator, and cell assembly. We carefully select each metadata item by discussing with the users in our lab to ensure that each item is relevant.

The example of the columns in the 'Schema' tab is shown in **Table 1**. Here, we are showing three examples out of 133 metadata items. There are four columns: 'Metadata', 'Value', 'Unit', and 'Ontology link'. The 'Metadata', 'Value', and 'Unit' columns store the name, the corresponding value, and the corresponding unit (if any) of the given metadata item. From our example, the

positive electrode current collector material is aluminum, the positive electrode current collector thickness is 15 micrometers and the tortuosity of the separator is 2.5. The 'Priority' column is our recommendation on the importance of each of the metadata items. There are three categories for this: 'required', 'recommended', and 'optional'. While the more items filled, the better, BattINFO converter will perform the conversion even if some fields are empty (the value field is left empty for the particular metadata item). The fields that are truly necessary are 'Cell type', 'Cell ID', 'Date of cell assembly', 'Institution/company', and 'Scientist/technician/operator' field. If one of these fields is left empty, the app will not perform the conversion. A special requirement is also needed for the 'Institution/company' and 'Scientist/technician/operator' fields. For these two fields, a unique online identifier is also needed. For the 'Institution/company' field, we recommend using a link from www.wikidata.org, and for the 'Scientist/technician/operator' field, we recommend using an Open Research and Contributor ID (ORCID). The 'Ontology link' column holds the information on how to describe each metadata item using the BattINFO ontology framework.[17] In general, the way an item is described using an ontology framework is usually hierarchical and nested in nature. The nested and intertwined structure can be seen by visualizing the resulting JSON-LD as a graph (**Figure 2**). We have developed a method that allows users to represent the nested relationship of the ontologized object in linear text, facilitating nontechnical users to ontologize custom new metadata items.

Despite an overwhelming number of metadata items in the Excel template, the user can choose to fill only parts of the whole template. In addition to this, only a few metadata items differ between cells in practice. Many items are kept identical among the produced cells, and only a few key metadata items are selectively altered among fabricated battery cells. This reflects the nature of how one would perform an experiment: only one or a few variables are changed while many are kept identical to assess the effect of a certain variable of interest. Hence, the user can create a template file where most of the metadata items usually kept identical among cells are pre-filled. The user can then duplicate the template file and fill in only a few items that differ for a specific cell, reducing the time to record the metadata of the cell while still maintaining all cell metadata. As automated, robot-assisted battery assembly[26,27] and cycling platforms become more common, most metadata is already available in digital form. A simple Python script can easily extract this metadata. Libraries such as openpyxl allow for automatic population of Excel templates, eliminating the need for manual input.

Table 1. Example of the metadata items in the Schema tab of the Excel template file.

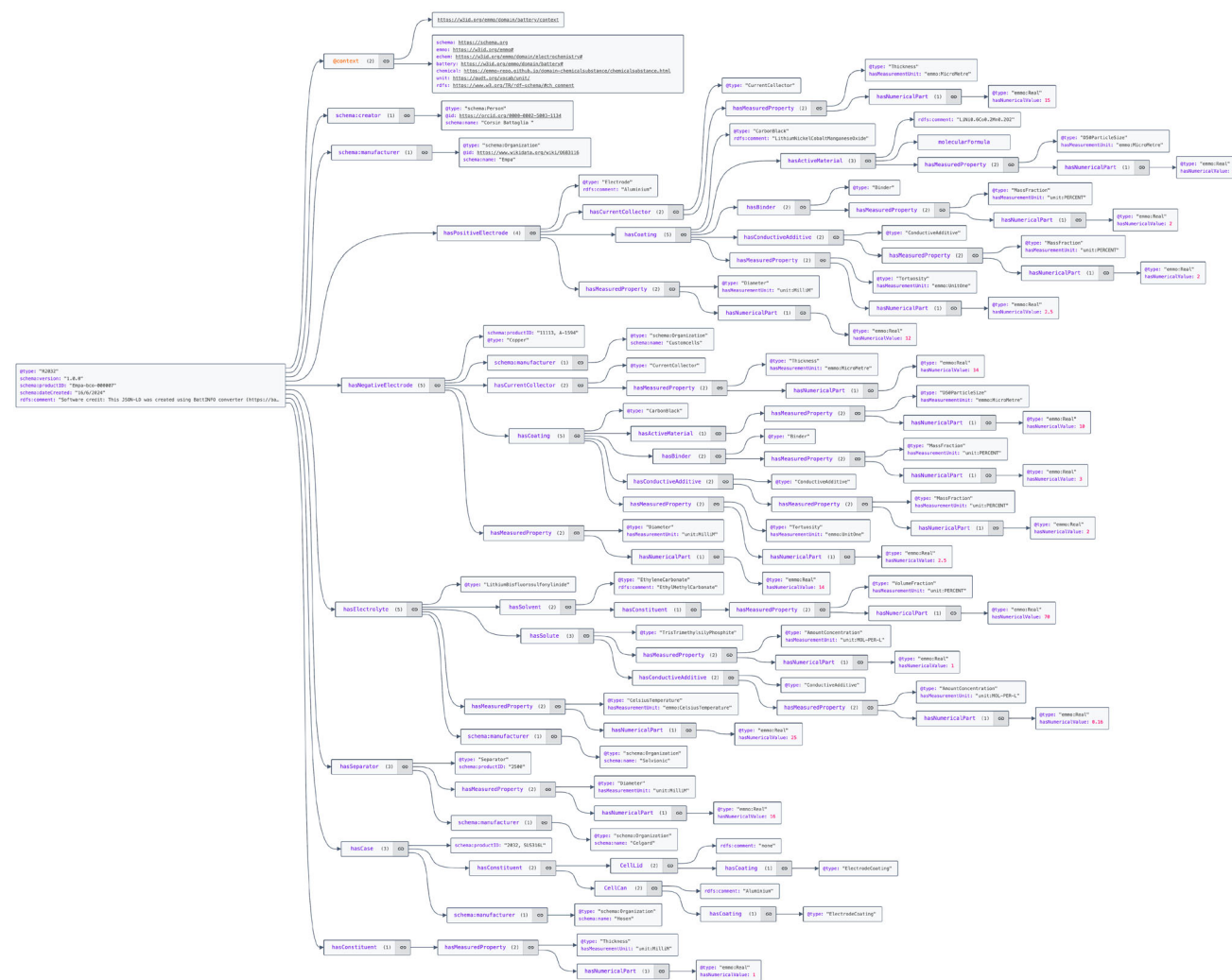| Metadata	Value	Unit	Priority	Ontology link
Positive electrode current collector material	Aluminum	No unit	Required	hasPositiveElectrode-hasCurrentCollector
Positive electrode current collector thickness	15	um	Recommended	hasPositiveElectrode-hasCurrentCollector-hasMeasuredProperty-Thickness
Separator tortuosity	2.5	No unit	Optional	hasSeparator-hasMeasuredProperty-Tortuosity

Figure 2. Illustration of an example coin cell battery metadata JSON-LD as a graph.

4. Control and Customization of the Excel Template

While we have provided the Excel template for a coin cell battery, we recognize that adopting the BattINFO converter application in other laboratories may require adjustments to the template. Certain controllability may be needed to ensure a standard of the produced metadata. In contrast, some users may require customization to ensure the recorded metadata is relevant to their specific research. With this in mind, we have designed the BattINFO converter application and the Excel template to accommodate users in both dimensions.

Controllability ensures the standard of the recorded metadata and the resulting JSON-LD. Controllability can come in many forms. A good example of this is the right to access and edit certain fields. The lab manager can lock most of the cells and only allow users to fill out the 'Value' column in the 'Schema' tab. This way, the lab manager can ensure that users do not accidentally (or intentionally) change the Excel template, preserving the metadata structure. Another example includes data validation, where only certain input types, such as numbers

only, are allowed in specific cells. Controlled vocabulary via a drop-down menu is another interesting feature. This could be beneficial in situations where only a few predefined inputs are appropriate for a given metadata item. For instance, the separator provider for the battery cell in a particular lab should only come from a few providers. The use of controlled vocabulary ensures that users do not input unexpected values or make typographical errors. These are some examples of how one could have controllability over the Excel template. These forms of controllability can be achieved directly using built-in functions in Excel without any adjustment to the BattINFO converter application itself. This is another strong advantage of using .xlsx as the file type of choice for the metadata template.

The opposite end of controllability, customizability, is equally important. Customization of the Excel template by removing or adding certain fields is possible. It is important to note that the main benefit of semantic annotation is not restricting what metadata must be included, but rather the standardization of how each metadata item is described. When adopting BattINFO converter to one's research, it is very likely that the user may find certain fields in the metadata irrelevant and hence should

Batteries & Supercaps

Research Article
doi.org/10.1002/batt.202500151

Chemistry
Europe
European Chemical
Societies Publishing

exclude the irrelevant metadata items from the user's version of the Excel template. Alternatively, the user may identify metadata items they deem relevant to their system of study and wish to add such items to the Excel template. While removing certain fields is simple, the added metadata items must follow our standards to ensure seamless operation of the application. Since we plan to continuously update this application in the future, it is advisable to review the most recent guidelines on adding new metadata items directly from the application. We provide a separate section on how to achieve this.

We envision that when BattINFO converter is adopted by other laboratories, the laboratory manager would first inspect and customize our Excel metadata template to suit the laboratory's needs, removing nonrelevant items and adding relevant ones, then adding additional layers of controls (examples include features we mentioned in the controllability section). This results in a customized Excel template containing the battery cell metadata with the adequate amount of controls to ensure the integrity of the metadata. The laboratory manager could even pre-fill the metadata template with certain items that remain identical for all cells (or a particular project) to facilitate the final metadata filling by the researcher who fabricates the battery cells.

5. Future Directions

We aim to continue developing the BattINFO converter app. While the current coin cell metadata template is chemistry agnostic, our development road map aims at expanding the Excel template to other cell formats and expanding the semantic annotation to battery performance data. For the first aspect, although users can already customize the Excel template to accommodate their own set of metadata and create an Excel template for other types of battery chemistries, we recognize that creating a new Excel template for other types of battery formats requires substantial amount effort. Therefore, we are working on creating Excel templates for other battery formats, focusing first on the battery formats we are currently researching in our group, which are solid-state batteries, molten-salt batteries, and redox-flow batteries. We will first create the Excel template and test the feasibility of the new Excel template with our lab members. Multiple iterations are required to ensure the newly created metadata template captures relevant and important information for the user. Once the metadata template version that satisfies the users in our lab has been made, we will share the version with our collaborators from partner institutions for further feedback. Once the final version is achieved, we will post them on the BattINFO converter web application page.

As for expanding the semantic annotation to the data, the metadata of a given battery, although important information, does not hold the whole picture of battery performance evaluation. The cycling data are also valuable information that must also be standardized and semantically annotated. This is to ensure a level playing field when comparing battery performance evaluation data generated in different labs.

Various electrochemical characterization protocols are used in assessing the performance of a battery. Galvanostatic charge/discharge cycling, cyclic and open-circuit voltametry, and electrochemical impedance spectroscopy are examples of commonly employed protocols.[28] Despite differences among these protocols, they are electrochemical measurement techniques, meaning that the data from these protocols provide the relationship between voltage (V) and current (I) (either being applied or measured). Therefore, these protocols can be represented as voltage and current as a function of time. In other words, the data from the electrochemical protocols are time series data. We are currently working on the semantic annotation of time series data to extend the standardization and semantic annotation to all electrochemical measurement data. Beyond electrochemical techniques, many other techniques are also employed to assess certain aspects of battery performance. These non-electrochemical techniques range from microscopy techniques such as scanning electron microscopy and transmission electron microscopy to spectroscopy techniques such as Raman spectroscopy and X-ray photoelectron spectroscopy to diffraction techniques such as X-ray diffraction.[29] We are also working on standardizing and semantically annotating these non-electrochemical techniques used in battery characterization. Ultimately, we aim to semantically annotate all information related to the battery—from assembly and cycling protocols to time-series results and any additional characterizations performed on the battery cell. The ideal format should allow easy validation of input information, support versioning, and be machine-readable. We are currently evaluating the best approach in terms of format.

We decided to release our software as an open-source project with the aim to allow every interested researcher to use our software freely and to welcome those who share the same vision of standardization and semantic annotation to contribute to our web application.

6. Conclusion

We have developed BattINFO converter, a Python-based web application for automating semantic annotation of coin cell battery data using the BattINFO ontology. The web application allows seamless conversion of coin cell battery metadata into a semantically annotated metadata file in JSON-LD format, facilitating the process of semantic annotation of the coin cell battery metadata without the need for data science and programming skills. The web application relies on a battery metadata template in Excel format, which balances controllability and customizability. We are currently working on expanding this work by providing templates for different battery formats and extending the work to semantically annotate not only battery cell metadata but also the battery cell performance data. We believe BattINFO converter will accelerate the adoption of semantic annotation of battery data, setting a new standard in the battery field with the hope of achieving more standardized and FAIR research data in the battery community. Our goal for this tool is to serve as a catalyst, empowering battery researchers to share their semantically annotated data, ultimately fostering collaboration and accelerating scientific progress in the near future.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in BattInfoConverter at https://doi.org/10.5281/zenodo.14945413, reference number 14945413.

Code Availability

BattINFO converter web application, along with the coin cell battery schema developed in this work can be accessed at https://battinfoconverter.streamlit.app/. The source code of the app is open-source and hosted at https://github.com/EmpaEconversion/BattInfoConverter. The specific release version for this publication can be found at https://doi.org/10.5281/zenodo.14945413and is also included as Supporting Information.

Keywords: batteries · coin cell battery · electrochemistry · energy storage · metadata · ontology · semantic annotation

[1] United Nations, Tackling Climate Change, https://www.un.org/sustainabledevelopment/climate-action/, (accessed: January 2025).

[2] International Energy Agency, Batteries and Secure Energy Transitions, https://www.iea.org/reports/batteries-and-secure-energy-transitions, (accessed: February 2025).

[3] M. Hannan, A. Q. Al-Shetwi, R. Begum, P. J. Ker, S. Rahman, M. Mansor, M. Mia, K. Muttaqi, Z. Dong, J. Energy Storage 2021, 42, 103040.

[4] B. Tarroja, J. M. Schoenung, O. Ogunseitan, A. Kendall, Y. Qiu, T. Malloy, J. Peters, J. M. Cha, D. Mulvaney, O. Heidrich, M. Baumann, Iscience 2024, 27, 1.

[5] M. Arbabzadeh, R. Sioshansi, J. X. Johnson, G. A. Keoleian, Nat. Commun. 2019, 10, 3413.

[6] P. Ahlgren, T. Jeppsson, E. Stenberg, E. Berg, K. Edström, ChemSusChem 2023, 16, e202300333.

[7] S. Chen, J. Xiong, Y. Qiu, Y. Zhao, S. Chen, J. Energy Storage 2023, 63, 107109.

[8] C. Bauer, S. Burkhardt, N. P. Dasgupta, L. A.-W. Ellingsen, L. L. Gaines, H. Hao, R. Hischier, L. Hu, Y. Huang, J. Janek, C. Liang, H. Li, J. Li, Y. Li, Y.-C. Lu, W. Luo, L. F. Nazar, E. A. Olivetti, J. F. Peters, J. L. M. Rupp, M. Weil, J. F. Whitacre, S. Xu, Nat. Sustain. 2022, 5, 176.

[9] A. K. Stephan, Joule 2021, 5, 1.

[10] S. Puls, E. Nazmutdinova, F. Kalyk, H. M. Woolley, J. F. Thomsen, Z. Cheng, A. Fauchier-Magnan, A. Gautam, M. Gockeln, S.-Y. Ham, M. T. Hasan, M.-G. Jeong, D. Hiraoka, J. S. Kim, T. Kutsch, B. Lelotte, P. Minnmann, V. Miß, K. Motohashi, D. L. Nelson, F. Ooms, F. Piccolo, C. Plank, M. Rosner, S. E. Sandoval, E. Schlautmann, R. Schuster, D. Spencer-Jolly, Y. Sun, B. S. Vishnugopi, et al., Nat. Energy 2024, 9, 932.

[11] Y.-K. Sun, ACS Energy Lett. 2021, 6, 2187.

[12] J. Li, C. Arbizzani, S. Kjelstrup, J. Xiao, Y.-y. Xia, Y. Yu, Y. Yang, I. Belharouak, T. Zawodzinski, S.-T. Myung, R. Raccichini, J. Power Sources 2020, 452, 1.

[13] L. Ward, S. Babinec, E. J. Dufek, D. A. Howey, V. Viswanathan, M. Aykol, D. A. Beck, B. Blaiszik, B.-R. Chen, G. Crabtree, S. Clark, Joule 2022, 6, 2253.

[14] S. Clark, F. L. Bleken, S. Stier, E. Flores, C. W. Andersen, M. Marcinek, A. Szczesna-Chrzan, M. Gaberscek, M. R. Palacin, M. Uhrin, J. Friis, Adv. Energy Mater. 2022, 12, 2102702.

[15] S. P. Stier, X. Xu, L. Gold, M. Möckel, Energy Technol. 2024, 12, 2301305.

[16] S. Clark, C. Battaglia, I. E. Castelli, E. Flores, L. Gold, C. Punckt, S. Stier, P. Veit, ChemSusChem 2025, https://doi.org/10.1002/cssc.202500458.

[17] The Battery Interface Ontology, https://big-map.github.io/BattINFO/index.html, (accessed: January 2025).

[18] M. Vigo, S. Bail, C. Jay, R. Stevens, Int. J. Hum. Comput. Stud. 2014, 72, 835.

[19] B. R. Long, S. G. Rinaldo, K. G. Gallagher, D. W. Dees, S. E. Trask, B. J. Polzin, A. N. Jansen, D. P. Abraham, I. Bloom, J. Bareño, J. Electrochem. Soc. 2016, 163, A2999.

[20] V. Murray, D. S. Hall, J. Dahn, J. Electrochem. Soc. 2019, 166, A329.

[21] A. U. Schmid, M. Kurka, K. P. Birke, J. Energy Storage 2019, 24, 100732.

[22] F. Dai, M. Cai, Commun. Mater. 2022, 3, 64.

[23] A. Nam, H. Song, J. Koo, G. Choi, Y. Kim, H. Kim, C. Shin, Y. Kim, J. Nah, Y.-J. Kim, P. Yoo, Commun. Mater. 2024, 5, 29.

[24] R. Elmasri, Fundamentals Of Database Systems, 7th Ed., Pearson, Boston 2021.

[25] LibreOffice, https://www.libreoffice.org/, (accessed: January 2025).

[26] J. T. Yik, L. Zhang, J. Sjölund, X. Hou, P. H. Svensson, K. Edström, E. J. Berg, Dig. Disc. 2023, 2, 799.

[27] E. Svaluto-Ferro, G. Kimbell, Y. Kim, N. Plainpan, B. Kunz, L. Scholz, R. Läubli, M. Becker, D. Reber, P. Kraus, R.-S. Kühnel, C. Battaglia, Batteries & Supercaps 2025, https://doi.org/10.1002/batt.202500155.

[28] W. Zuo, R. Liu, J. Cai, Y. Hu, M. Almazrouei, X. Liu, T. Cui, X. Jia, E. Apodaca, J. Alami, Z. Chen, Chem. Rev. 2025, 125, 369.

[29] C. Gervillié-Mouravieff, W. Bao, D. A. Steingart, Y. S. Meng, Nat. Rev. Electr. Eng. 2024, 1, 547.