Joule

CellPress

Perspective

Principles of the Battery Data Genome

Logan Ward,[1] Susan Babinec,[1,*] Eric J. Dufek,[2,*] David A. Howey,[3,4,*] Venkatasubramanian Viswanathan,[5,*] Muratahan Aykol,[6] David A.C. Beck,[7] Benjamin Blaiszik,[1,8] Bor-Rong Chen,[2] George Crabtree,[1,9] Simon Clark,[10] Valerio De Angelis,[11] Philipp Dechent,[12,13] Matthieu Dubarry,[14] Erica E. Eggleton,[7] Donal P. Finegan,[15] Ian Foster,[1] Chirranjeevi Balaji Gopal,[6] Patrick K. Herring,[6] Victor W. Hu,[7] Noah H. Paulson,[1] Yuliya Preger,[11] Dirk Uwe-Sauer,[12,13] Kandler Smith,[15] Seth W. Snyder,[2] Shashank Sripad,[5] Tanvir R. Tanim,[2] and Linnette Teo[7]

SUMMARY

Batteries are central to modern society. They are no longer just a convenience but a critical enabler of the transition to a resilient, low-carbon economy. Battery development capabilities are provided by communities spanning materials discovery, battery chemistry and electrochemistry, cell and pack design, scale-up, manufacturing, and deployments. Despite their relative maturity, data-science practices among these diverse groups are far behind the state of the art in other fields, which have demonstrated an ability to significantly improve innovation and economic impact. The negative consequences of the present paradigm include incremental improvements but few breakthroughs, significant manufacturing uncertainties, and cascading investment risks that collectively slow deployments. The primary roadblock to a battery-data-science renaissance is the requirement for large amounts of high-quality data, which are not available in the current fragmented ecosystem. Here, we identify gaps and propose principles that enable the solution by building a robust community of data hubs with standardized practices and flexible sharing options that will seed advanced tools spanning innovation to deployment. Precedents are offered that demonstrate that both public good and immense economic gains will arise from sharing valuable battery data. The proposed Battery Data Genome looks to broadly transform innovations and revolutionize their translation from research to societal impact.

CONTEXT & SCALE

Batteries are a key enabling technology in the transition to a low-carbon economy but have yet to enjoy the revolutionary data-science gains enjoyed by other fields. In the proposed Battery Data Genome, we identify gaps hindering this transformation and put forth organizing and operating principles that can drive uniform practices that are the foundation of the solution. Our path forward builds a community of data hubs with standardized practices and, critically, with flexible sharing options. Together, these will enable data-science advances across a reimagined international battery community that will catalyze a new generation of innovation with broader economic impact. As a call to action, the Battery Data Genome is a crucial step to address the global low-carbon challenge. The authors are committed to continue advancing the Battery Data Genome by expanding and linking the existing data hubs and by building the community and organizational structure needed for success.

INTRODUCTION

Batteries are a cornerstone of decarbonization; they enable electrified transportation and significant renewable energy generation on the electrical grid.[1,2] Their impact continues to expand rapidly as performance steadily improves, costs continuously decline, and deployments accelerate.[3,4] The Nobel Committee acknowledged the importance of batteries by awarding the 2019 Chemistry Prize for the development of lithium-ion batteries which "laid the foundation of a wireless, fossil fuel-free society."[5] Batteries are an international enterprise with many countries innovating across the global battery value chain. For example, the US Department of Energy's Energy Storage Grand Challenge,[6] the European Union's Battery 2030+ research initiative[7] and the UK's Faraday Battery Challenge,[8] are all focused on organizing a cohesive battery community as the catalyst for an innovative, robust clean-energy economy that addresses urgent climate issues.

Despite pressing needs, deep commitments, and broad excitement, battery innovation and investment have yet to enjoy the revolutionary data-science gains enjoyed by other fields. Upgrading battery data and software sharing practices, which are decades behind contemporary paradigms in communities such as genomics, therapeutics, protein crystallography, atomistic simulations, and materials science, would transform this important international industry.[9–19] (See supplemental information for further details.) The Human Genome Project (HGP) is one well-known example of broad data-sciences impact: a 2011 analysis of the $3.8B US federal government HGP investment spread over 13 years and ending in 2003, indicates that it created $796B of direct economic impact, $244B of personal income, and 310,000 jobs in the first 8 years after its completion. A 2019 analysis of The Materials Genome Initiative (MGI)[19] concludes that it provides significant impacts in science and technology across diverse industries, such as medicine, catalysis, transportation, etc. These are but a few instances in a rich history demonstrating that an open dataset with common tools can catalyze extraordinary innovations that not only seed public good through knowledge, but also generate enormous economic gains of a magnitude that are otherwise improbable.

In contrast, it has taken more than 30 years from the initial commercialization of Li-ion batteries to achieve today's energy and power density and cycle life.[20,21] Productivity and innovation gaps in today's approach include: the sluggish pace of linking theoretical advances with experimental validation; tedious materials discovery; time-consuming performance and life testing that slows validation; and an inability to fully understand and mitigate device degradation-pathways that compromise system designs. While batteries today are often described in idealized settings using physics-based or even empirical models, these are often insufficient when applied outside the restraints under which they were developed. While some level of success can still be achieved with traditional methods, recent history illustrates that these collectively lead to cascading investment risks that slow market adoption, with negative societal and economic impact.[22] Creating more sophisticated tools, which can include aspects of both data-driven and physics-based approaches,[23] will require more data to capture the complexity of both battery technologies and applications. Establishing coordinated battery-data-science efforts will provide distinct economic and social impacts in an accelerated timeframe.

The primary roadblock is that modern data science requires large quantities of data, and that no single entity can or will provide it to the battery community at the scale needed to generate a suite of breakthrough tools that will catalyze a new innovation paradigm. Academics, national laboratories, and small companies are often willing to open-source data, but their databases are typically too small. Successful multinational battery corporations likely have sufficient data to support modern data-science projects but will not risk their competitive edge by sharing.

While behind other communities, the battery community has a unique opportunity to rapidly spur technology development and deployment due the current global focus and realization of the need for batteries and a distinct interest in evolving how battery researchers are trained. The significance of batteries as a key component of climate-change solutions requires an innovation transformation that can be catalyzed with data sciences. Leveraging the alignment of climate need, emerging workforce, and rapidly evolving tools suggests that this is the time for the battery community to establish its own data genome.

We propose the Battery Data Genome (BDG) as a global initiative to assemble a massive collection of battery databases. We use the genome terminology less to

[1]Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439, USA
[2]Idaho National Laboratory, 1955 N. Fremont Ave., Idaho Falls, ID 83415, USA
[3]Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
[4]The Faraday Institution, Harwell Campus, Didcot OX11 0RA, UK
[5]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[6]Toyota Research Institute, 4440 El Camino Real, Los Altos, CA 94022, USA
[7]Department of Chemical Engineering, University of Washington, 105 Benson Hall, Seattle, WA 98195, USA
[8]University of Chicago, 5730 S Ellis Ave, Chicago, IL 60637, USA
[9]University of Illinois at Chicago, Chicago, IL 60607, USA
[10]SINTEF Industry, New Energy Solutions, Sem Saelands vei 12, Trondheim 7034, Norway
[11]Sandia National Laboratories, 1515 Eubank Blvd SE, Albuquerque, NM 87185, USA
[12]RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany
[13]Helmholtz Institute Münster (HI MS), IEK-12, Forschungszentrum Jülich, Jägerstrasse 17-19, 52066 Aachen, Germany
[14]Hawaii Natural Energy Institute, University of Hawaii, Honolulu, HI 96822, USA
[15]National Renewable Energy Laboratory, 15013 Denver West Pkwy, Golden, CO 80401, USA
*Correspondence: sbabinec@anl.gov (S.B.), eric.dufek@inl.gov (E.J.D.), david.howey@eng.ox.ac.uk (D.A.H.), venkatv@andrew.cmu.edu (V.V.)
https://doi.org/10.1016/j.joule.2022.08.008

link batteries to a biological genome but more to invoke the conceptual analogy of a historically immense, complex, and urgent set of problems faced by the battery community, which has similar lofty goals and similar urgency. This follows approaches of other earlier activities including the MGI.

As a collection of connected data hubs with uniform standards and practices and with open software that serves needs from data collection through life and performance analysis, the structure will enable the BDG to serve as both a data-generation and data-harvesting asset rooted in flexible sharing options to accommodate diverse organizational and individual constraints. The BDG will include contributions from scientifically diverse communities and span all technology readiness levels (TRLs) from theory, experiment, piloting of both materials and devices, mature manufacturing, and authentic field deployments. As the BDG grows it will serve as a key source for guiding data generation by elevating data standardization and collection practices through the development of best practices, community outreach and education efforts, and by establishing key challenge problems. Global scope is a key enabler to assemble a sufficiently abundant database in the face of sharing concerns. The global vision additionally acknowledges that battery contributions to climate-change solutions are international enterprises that require participation of emerging and currently underrepresented economies, despite their limited resources. Thus, the BDG will enable global advances, especially in deployment which is critical to achieving net zero world goals.

The magnitude of the BDG's challenge to sharing norms is not underestimated. At the root of the BDG are stakeholders who must balance their individual aspirations with essential sharing requirements of this vision—individual contributors, public institutions, private profit-generating organizations, and a broad customer base. Competition for visibility, research funding, profit, and market share serve as a strong deterrent to cooperation. Public institutions generally align with the broad BDG goals if success metrics, such as publications, are not compromised. Profit-generating organizations will benefit from the use of open community data and software while protecting their trade secrets and propriety information with flexible sharing practices. The BDG's design principles and organizational structure weave the promise of transformative capabilities with accommodations of individual concerns for competition. Ultimately, standardized practices and interoperable tools will enable cooperation and grow a more cohesive community, as they have done in many other fields. We expect that, much like other fields, this will spur additional innovation and accelerate the path from benchtop to deployment. As detailed below, opportunities in discovery, manufacturing, validation, and deployment will emerge across the development cycle, reducing risk.

A subset of the authors has formed a BDG seed committee for a globally organized approach to drive a more cohesive and innovative community from today's silos of diverse technologies. The proposed path draws on previous successful initiatives with considerations that address not only technical barriers but also key soft issues, such as establishing how the benefits of collaboration through data sharing outweigh the resulting competitive disincentives. Metrics to gauge success are also proposed in the supplemental information. Thus, this paper is not only an analysis of issues with a path to resolution, but also a call to action.

DATA-SCIENCE OPPORTUNITIES CREATE BROAD IMPACTS

History shows that innovation in one area drives rapid growth and innovation broadly across multiple segments. For example, the human genome created in the HGP

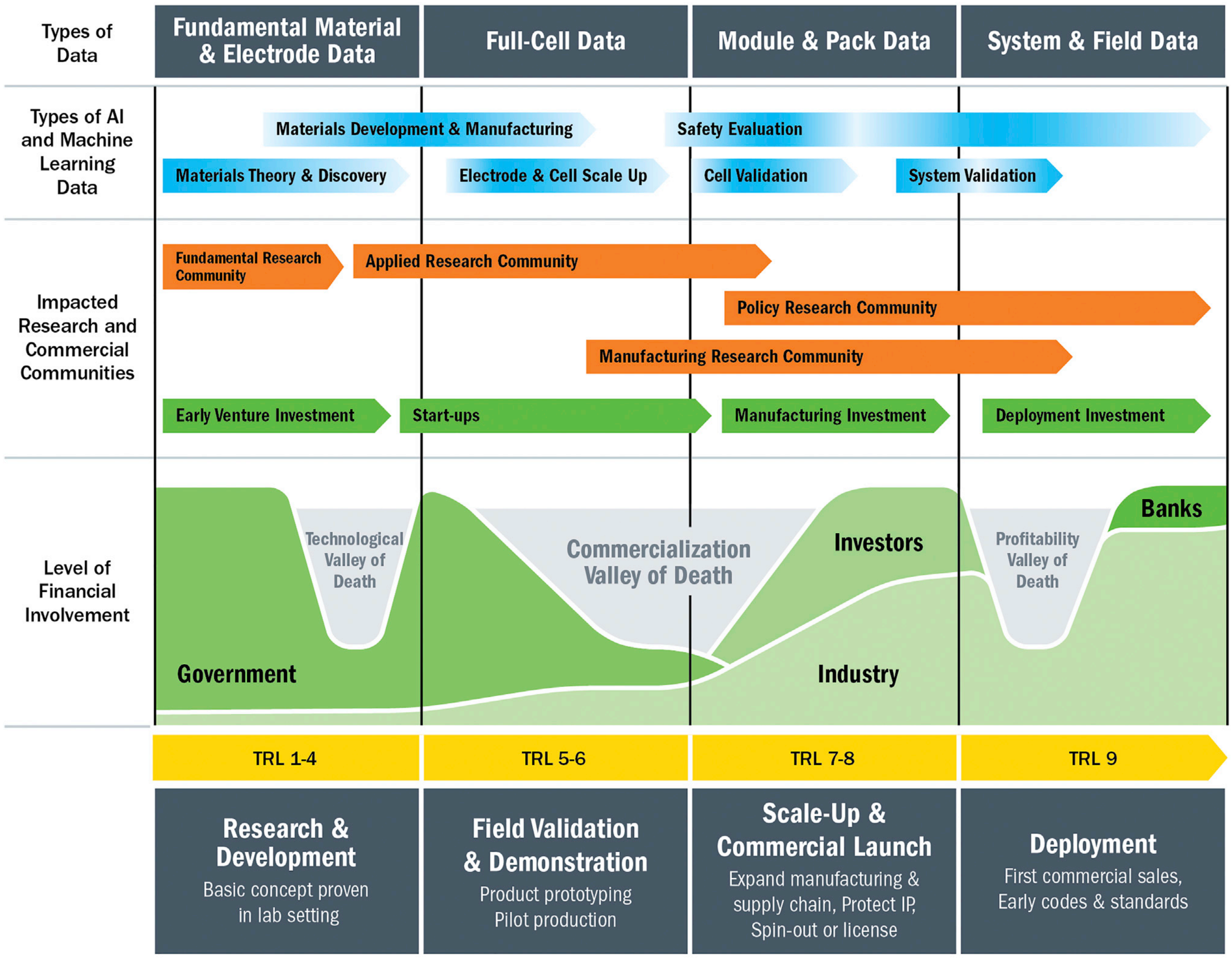

Figure 1. Battery AI/ML innovations of all types span all technology readiness levels (TRLs) and create positive impacts to many stakeholders in both research and commercial communities

For the AI/ML tools in the blue arrows, the intensity depicts the level of impact in advancing battery technologies.

seeded breakthrough innovations in forensic sciences as well as drug discovery. We assert that opportunities for rapid innovations will similarly be enabled by connecting data-science best practices with the battery community's deep scientific, manufacturing, and deployment capabilities and that this will lead to broad scientific, societal, and commercial impact. Figure 1 suggests relationships between BDG-seeded artificial intelligence and machine learning (AI/ML) capabilities for various battery-related data-science objectives and research and commercial impacts as a function of technology maturity expressed as TRLs as commonly defined by many organizations including the National Aeronautics and Space Adminstration (NASA) and other US government agencies.[24]

AI/ML technology innovation spaces

A comprehensive data-sharing ecosystem will seed a suite of innovative AI/ML tools at each TRL as in Figure 1. For example, AI/ML materials theory would initially seed materials discovery, which guides the development of high-yield manufacturing. Next, the materials data inform appropriate benchtop electrode designs, which provide cell-validation data that smooths the transition to pilot-line quality control and then transitions to scale-up followed by mature manufacturing. Providing access to larger and more diverse datasets as part of the BDG will open up additional AI/ML tools. As clearly pointed out by Sendek et al., most current ML work in batteries has used limited tools due to the size of datasets, and increasing the quantity of high-quality data would open the possibility for deep learning and other high data need methods.[25] The ultimate vision is for a unified structure that enables the flow of data across the individual segments—from theory and discovery to deployment and validation.

Impacted communities

Research impacts include theoretical materials discovery, rapid inverse design where tools directly identify materials that will provide desired functionalities, and development of new methods for evaluation protocols in laboratories. Commercial impacts will arise from more abundant, precise, and transferable information and standardized protocols that collectively accelerate the product-development cycle across each step. Shorter time frames with more precise and abundant information reduce risks that slow investments from seed funding, across the various "valleys of death" and finally to large manufacturing capital investments which produce hardware for field deployments.

One example of the above is the ubiquitous, time-intensive battery-life prediction step that has been a natural initial target for AI/ML solutions. Recent studies have used ML methods to predict cycle life using fewer than 100 initial cycles,[26–29] to identify early indicators of Li plating (a major safety concern) by analyzing cycle-by-cycle electrochemical data,[30] and to predict battery end-of-life noninvasively from operational data without taking systems offline,[29] while synthetic datasets provided insights into degradation paths with reduced experimental burdens.[31–33] Similar studies and opportunities for advancement have also been suggested for calendar-life studies and emerging chemistries.[23,34] A study using a more diverse dataset provides tools that predict life across a broad range of cathode chemistries and electrolyte formulations, which accelerate materials research.[28] Another example is the discovery of new battery materials with targeted properties, which has been significantly accelerated by inverse design using ML,[35,36] and fully autonomous materials discovery laboratories are now emerging.[37–40] Despite the impact of these early studies, successes are infrequent due in part to limited availability of shared data.

In addition to initial AI/ML life prediction tools, new data-sharing hubs are also emerging. Often nation-centered, early independent battery-data activities include those of Battery Archive, BIGMap, Batteries Europe, and the Faraday Institution. Their diversity of locations and formats underscores the critical need for a singular approach to improve uniformity.[41]

BATTERY DATA GENOME ORGANIZING AND OPERATING PRINCIPLES

The BDG faces two significant data challenges: (1) heterogeneity—establishing the data and metadata conventions that will make heterogeneous battery data useful and enable interoperability; and (2) scale—rapid, large-scale capture of data from many sources and contributors. The scale challenge is a data-engineering issue that has been solved in other genomes; the heterogeneity challenge, however, is more complex and specific to battery data. Heterogeneity here refers to the broad spectrum of the most critical information; it arises from multiple phenomena covering a range of length scales, from molecular to pack dimensions, and time domains from seconds to decades (see Figure 2). Not only are the data naturally complicated, it also is rapidly evolving as new experimental and developmental

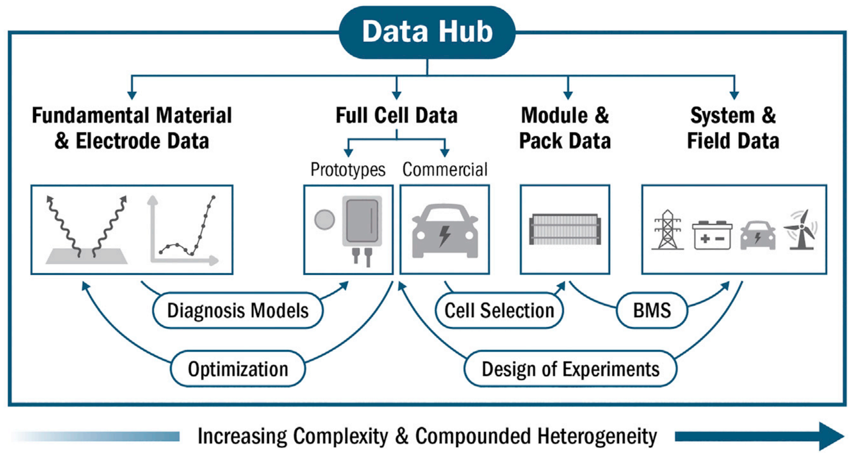

Figure 2. The four primary segments of the Battery Data Genome

capabilities allow researchers to generate even more diverse data across these spectrums at unprecedented scales and precision. The heterogeneity challenge is further compounded by bench-scale reproducibility, cell-to-cell manufacturing uniformity, and thermal control. Early efforts at establishing a more unified ontology to describe battery data are underway.[42] The BDG organizing and operating principles both expand on the early work and address the foundational challenges from multiple angles.

Organizing principles

The heterogeneity and scale challenges can both be addressed by organizing the data into four segments based on the functionality of the component or system under study—an approach that can naturally accommodate variability, while driving interoperability. The four segments, which are meant to facilitate more uniform data harvesting and generation, are illustrated in Figure 2 and described below:

(1) Fundamental material and electrode data from studies that probe bulk and interfacial processes at the atomic and molecular scale,[18,43–46] for both materials discovery and commercial electrodes[47–49]; this serves as the core set to support theory development and validate *ab initio* property predictions from materials theory. Materials characterization can include electrochemical data and spectroscopy, X-ray characterization, etc. Electrode characterizations can include optical, acoustic, post-test tear down, and synchrotron-based studies.
(2) Full-cell evaluation data reflect the interplay among materials, electrodes, cell designs, and their performance for prototypes[50,51] and commercial single cells.[27,52–54] This segment should also include nonelectrochemical data, such as temperature, pressure, and manufacturing process variations.
(3) Module/pack characterization data reflect not only system performance[55] but the ability of the pack and battery management system (BMS) design to manage cell-to-cell and environmental heterogeneities.
(4) System and field data are generated during authentic "real-life" use, which is usually nonuniform and in contrast to research cycling protocols. Field data provide insights into the ultimate intended goals and is required to understand gaps between lab and field situations. These data are often not openly available but are critical for the transition to global decarbonization. The information is generated at the system level, and so approaches will be needed

to separate data of the entire system from its many cells and its BMS. The value of field data as a complement to laboratory data—giving insight into real world battery usage and performance—has been demonstrated in early examples and is seen as a key to providing better validated models, new techniques, and improved visibility of important data-generation activities which may be missing.[29,56]

This approach is designed to support today's technologies as well as future materials, chemistries, and device structures. For example, while Li-ion technology is currently a very active field, the organizing principles will translate to emerging Li-metal technologies, new low-cost aqueous chemistries, flow cells, and the many evolving approaches to long-duration storage. Early battery data hubs already use these organizing principles for some of their specific data types: (1) the Battery Archive, which provides data for battery degradation studies[54,57]; (2) the Battery Evaluation and Early Prediction (BEEP) tools, which focus on optimization of fast-charging protocols for batteries[58]; and (3) Galvanalyser, which aims to unify the gathering and querying of data from different types of battery testers. Furthermore, several emerging databases on fundamental material and electrode segments link battery chemistry to material properties[59] and electrode microstructure to performance.[43,60–63] There are also databases linking battery information to other phenomena, e.g., various electric vehicle (EV) battery-lifetime tools[64] leveraging open meteorological data[65] and drive cycles.[66] These specialized databanks can serve as models and guides to establishing the structure and metadata requirements for the BDG.

Theoretical data

All four data segments can be bolstered by simulations or generation of "synthetic data," to complement experimental datasets. This "close the design loop" approach is a critical strategy to increase innovation pace and efficiency by extending experimental results and analyzing multi-fidelity datasets. Synthetic data examples include first principles atomic calculations, cell-level synthetic data[31–33,67] and full-system digital twins. Benefits of digital twins include the use of physically rooted electrochemical performance and aging models to generate virtual experimental data (hence "digital twin") which is otherwise prohibitively costly and time-consuming. Another example is the digitization of the battery manufacturing process to speed protocols and procedures for both material development and cell design.[68] Synthetic data also provides the opportunity to benchmark the predictive performance of ML algorithms. The completely new function the BDG could offer is to seed partially autonomous laboratories by coordinating experimental and synthetic data access.

Operating principles

The BDG is designed around six key operating principles:

The first operating principle, the standards principle, holds that uniform standards and protocols guide how experiments should be performed, how existing data can be adapted, what types of data should be collected, and ultimately how we train researchers. This uniformity is the foundation for interoperability, which is critical for breakthrough innovations arising from cross-fertilization, and is immediately needed for both theory and experiments.[34,69–77] Success here requires community consensus both on how data is organized and then on how to implement in both data generation and acquisition.

The second operating principle, the metadata principle, is that metadata are as important as performance data. Metadata should include not just detailed electrochemical information but also the chemical, structural, and physical characteristics of the cells and materials involved in an experiment. Metadata are complex, exceptionally heterogeneous, and require detailed reporting protocols to insure adequacy and consistency. Full-cell and electrode metadata could include all test conditions, references and protocols, additional characterizations (such as safety), and experimental control details. Additional details on metadata are in the supplemental information.

Standardizing metadata while maximizing value, flexibility, and impact is a significant challenge. Establishing what information is needed and its format will require not only extensive community discussion, but also innovations to enable broadest possible participation. The BDG will manage trade-offs by organizing into layers of detail: primary, secondary, and tertiary:

- Primary metadata are the minimal mandatory information that must be reported to cover the basic concepts will be structured as a table of contents. It will include data owners and generators, a unique identifier for each cell, experimental objective, the nature of the device/s that were tested including chemistry, and specific tests. Examples are shown in Table S1.
- Secondary metadata are more comprehensive, with *significant* experimental details including cell and electrode design, commercial-cell identification codes, the manner of attachments to test equipment, etc. The journals *Joule* and *Advanced Energy Materials* already have battery experimental reporting requirements that provide initial concepts.[78]
- Tertiary metadata refers to highly specialized, nonroutine information that likely vary substantially between subfields and should be defined by stakeholder's consensus—for example, material synthesis protocols and simulation model parameters.

To facilitate the ease of metadata entry some of the data hubs have started to develop metadata entry tables and forms which can be more directly used to populate the identified primary, secondary, and tertiary metadata. While these are in the early stages of development it can be readily envisioned that future iterations may evolve to the development of more advanced metadata capture and upload possibilities using tools such as electronic laboratory notebooks (ELNs). Developing ELNs will be aided once each of the classes of metadata are more directly agreed upon across the BDG community.

The third operating principle, the quality principle, is that all performance and metadata must be cleaned, curated, and quality controlled. before submission to a data hub. This principle is the foundation of ensuring scientific robustness of the data. Advancing quality will include developing test protocols from bespoke procedures used in early research up to rigorous commercial evaluation standards.[79,80] It will also involve developing tools for cleaning data. Examples of widespread challenges include: (1) significant clean-up of time-series current, voltage, time, and temperature data is already a well-known issue. Open-sourced tools from academic and national labs (e.g., Battery Data Toolkit) are presently available, but more tools will be required. Future efforts, including the ranking or grading of datasets based on the metadata, the quality of the data, and the plausibility (or verification) of repeatability may also evolve as the BDG advances. (2) The location of and funding for efficient storage and back-up for a large volume of data is an issue that may be partially solved by support of large organizations, such as funding agencies and philanthropic and industry groups.

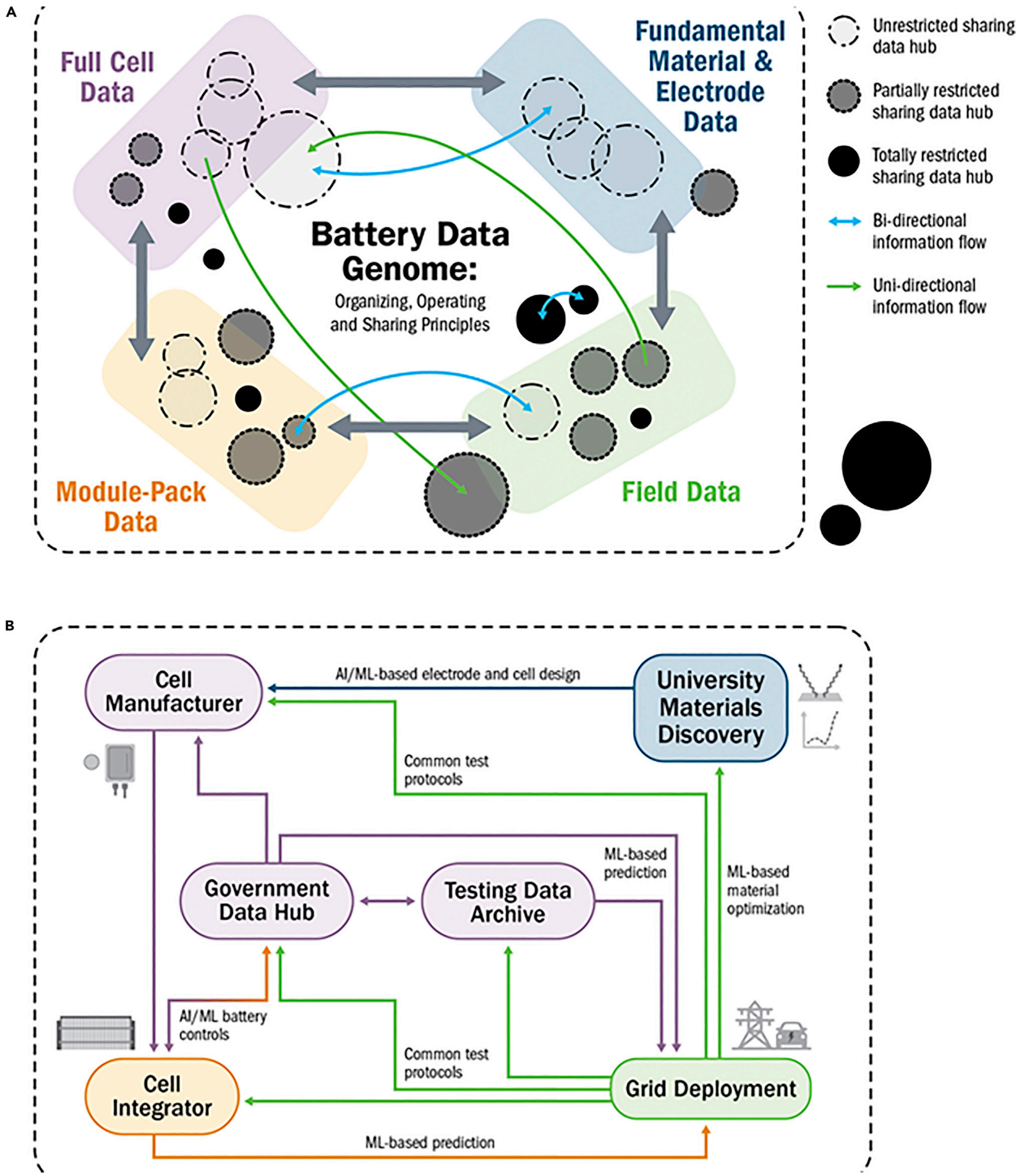

Figure 3. An example of coordinated efforts

(A) Visual representation of the BDG (dashed line) where data hubs (represented as circles) are grouped based on data segments. Data flow freely between segments (thick arrows) and can also flow between data hubs, although not all hubs are completely open. Some data hubs within the BDG may

Figure 3. ***Continued***
have severe sharing restrictions but can still participate with other data hubs if access restrictions are met as shown by the two black circles connected with bidirectional data sharing. Other data hubs not part of the BDG (lower right) are expected to still exist.
(B) An example of data flow across different data hubs as described in the text. The arrows show data flow, while the text denotes tools and protocols made possible by the BDG.

The fourth operating principle, the sharing principle, states that BDG success does not require all data to be shared openly, but only that all participants use agreed-upon standards. To drive participation, there will be a push to align BDG principles with data reporting to journals and across key government funded activities where data sharing is reasonable. Additionally, in the early stages of BDG planning, techniques for anonymization of data will be explored. The sharing principle does not compromise the BDG foundational need for significant amounts of data, because success in any given project requires that only a fraction of the data which is already stockpiled. Similarly, it does not diminish—in fact, it enhances—impact by encouraging broad use of standards without committing to the risk of data exposure in advance. Figure 3A illustrates sharing principles as several venues for communicating between data sources with different levels of openness. As shown in Figure 3A, multiple data hubs are loosely grouped based on four data segments described in the text. For fully open data hubs, information freely flows both between similar hubs and different data segments of the BDG. It is anticipated that lower TRL information will tend to be more open. In other data hubs where there is some level of access restriction, information can still be exchanged but not all users will have full access.

Figure 3B depicts an example of sharing principles as information flow across segments and data hubs, both forward and backward in TRL. In this example, a group of university researchers performing materials discovery develops new material information that is shared openly with a cell manufacturer. The manufacturer develops AI/ML-based electrode- and cell-optimization approaches using the university data plus data from an open-testing data archive and a partially open government data hub. To contribute back to the BDG, the cell manufacturer shares some of their cell performance data back to the testing data archive. Manufacturing data is also shared with a cell integrator. The integrator again uses data from the government and open data sources and a grid installation to develop ML-informed advanced controls. Ultimately, the grid developer uses the integrated system with advanced controls to decarbonize power generation. As a participating member of the BDG, the developer shares information, which is then used to further optimize future materials and develop both new, more targeted (and shorter duration) test protocols and more refined cell and management designs.

Through the coordination provided by BDG operating and organizing principles, the data sharing is seamless and enables AI/ML tools and common software to be independently developed. It is anticipated that much of the higher TRL data will be anonymized, especially field and manufacturing information. Other examples related to BMS development, fast-charge-protocol development, or recycling and second life can follow similar processes where data flow across data hubs. When coupled with AI/ML as part of the BDG dramatic reduction in the time from discovery to deployment is envisioned.

The fifth principle, the software principle, is that at least one complete suite of interoperable software must be open-sourced. It would be a missed opportunity to not reimagine the role of software in a comprehensive vision since fragmented software development and inaccessibility have hampered battery-data-science advances.

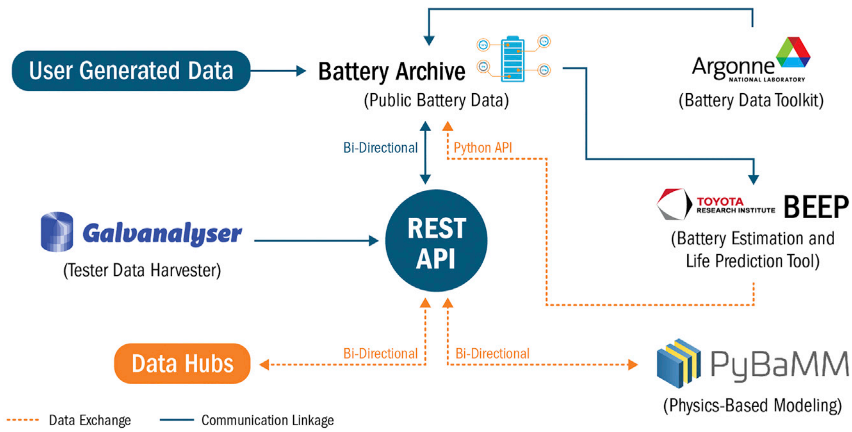

Figure 4. Several open-source packages and data hubs are emerging, and groups are starting to integrate them to exchange battery data and metadata

Open and interoperable software tools that enable data management, testing, and analysis are broadly needed. Interoperability catalyzes the BDG vision by enabling coordination across (standardized) data hubs and comparison of data against evolving testing standards; it enables streamlining and uniformity of test procedures and hardware configurations across all stages of development and types of analysis. It is critical that the software be compatible with common tools for plotting, analysis, and prediction, and other functions.

Open software enables challenge activities to drive further software development—indeed a virtuous cycle. For example, early community-driven software code efforts have already contributed to solving some battery-data processing and analysis[58,81–86] challenges, including simulation frameworks (e.g., Python Battery Mathematical Modelling, PyBAMM[81]). Yet, significant opportunities remain: (1) battery-cycling protocols that are not standardized or easy to translate between different experimental equipment; (2) workflows for data management—parsing, validation, and sharing—are idiosyncratic and data formats are highly nonuniform; and (3) AI/ML code for analysis and automation is a new area, and embryonic code is *ad hoc*, lacking modularization and interoperability. Further details on these gaps are provided in the supplemental information.

Existing software projects already have linkages that establish an interconnected research community with interoperable data tools. Figure 4 offers our proof-of-concept integration that we developed between complimentary capabilities in the public-facing data hub Battery Archive,[57,87] the private-test-lab database software Galvanalyser, the data analysis tool BEEP[58] and the modeling software PyBAMM[81] (described in Github[88]). Battery Archive provides long-term archival storage and interactive visualization of cycling data and benefits from data parsers available in BEEP, and PyBaMM offers a modeling framework for simulating the cycling protocols *in silico*. Groups at Argonne National Laboratory have developed complementary algorithms to prepare the data for ML and visualization in the Battery Data Tool Kit.

The proof of concept shown in Figure 4 suggests a variety of possible workflows. For example, an individual lab uses Galvanalyser to store their battery test data privately in a local database; they select a subset of these data to be public and transfer it to Battery Archive and then compare their results against similar tests by other groups with data also stored on Battery Archive. Another lab subsequently transfers the data from Battery Archive to BEEP to explore charging algorithms and lifetime, and a third group transfer the data to PyBAMM to develop a new continuum model for battery degradation. Crucially, we also developed a representational state transfer application programming interface (REST API)—a standardized set of rules to enable communication between different pieces of software—which is already in use to exchange information between Battery Archive and Galvanalyser and will also be used to exchange information with PyBaMM.

These tools are a start, but must be expanded; e.g., BEEP added an importer to process Battery Archive time-series files and Battery Archive added an importer for the HDF5 files generated by the Battery Data Tool Kit. As integration progresses, so does our understanding of what constitutes primary metadata, and what are the needs for further development. For example, Battery Archive is now developing tools to make BEEP directly available through Jupyter Notebook on the BA servers, facilitating easier use of ML tools. A key outcome from early discussions is that the REST API is now available as new data hubs appear, as discussed in the operating principles above.

Challenge problems driving community engagement and addressing broad issues

While the organizing and operating principles are rooted in the standardization and data unification needed for harvesting data from across the globe, we realize that just harvesting data and establishing guidelines is unlikely to be sufficient to drive success for the BDG and attain the ultimate goal of a reduced carbon future. Still, it is difficult to determine where data gaps lie. One approach to ensure dataset sufficiency is the use of challenge problems, which will be established to use both harvested and generated data for the purpose of accelerating tool development and setting priorities for new data generation. These challenge activities also uniquely reveal issues and provide a detailed understanding of data gaps.

Until suitable datasets are available for the significantly difficult challenge problems, we suggest a two-tiered approach that balances the need for rapid start driving community engagement and an expanding scope with greater complexity. Tier 1 challenges will use currently available data with a key interest in addressing cell life as shown in the middle of Figure 1. More complex Tier 2 challenges will emerge as appropriate open datasets become available and will be focused across the entire development cycle from materials discovery to system safety and performance. Each challenge problem is carefully designed to focus on topics that are difficult to address using data derived from a single institution, which therefore would require the BDG. The problems are also specifically focused on key challenges across the TRL scale shown in Figure 1. As more explicitly detailed in the supplemental information, to accelerate Tier 2 challenges, the authors are in the planning stage of generating a key dataset that will include opportunity for both experimental and synthetic data across a range of aging and use conditions. This tiered approach is seen as a way to both drive early engagement and further expand the data available as part of the BDG.

The present Tier 1 and Tier 2 challenges are dominated by existing Li-ion chemistries as these offer significant data, but there is a critical need for future Tier 2 challenges to include emerging battery chemistries and designs, especially those for long-duration storage. Possible areas for future releases include flow-battery electrolyte systems, materials discovery, other alkali-metal-based chemistries, and multivalent systems. An example is for an electrode satisfying multiple performance criteria, such as high or low operating voltage, high working ion capacity and mobility, low expansion on intercalation, and minimal side reactions. The challenge is to satisfy all the requirements simultaneously while negotiating the trade-offs.

The first consideration in challenge problems is the topic areas. As one example we focus here on cell-level challenges, which can be categorized into performance, state of health, life, safety, and manufacturing:

(1) Performance: What metadata are most critical for successfully predicting device performance over a decade of cycling? Are algorithms for one cell chemistry extensible to others?
(2) State of health: Can capacity at any point in cell life cycle be estimated from *brief* current/voltage excursions? What type of data and models best represent performance evolution during aging? What limits an algorithm's ability to predict future performance in any scenario and at any state of health?
(3) Life: What is the minimum number of cycles needed to predict cycle life (for example, to 80% of initial capacity) with different accuracies? How do cycle-life predictions for one use-scenario translate to another; for example, can fast-charge data predict life for energy arbitrage on the grid? How well do predictions transfer from one chemistry to another; for example, can Li-ion NMC-cell data help predict iron phosphate behaviors?
(4) Safety: Is there a signal that indicates an impending thermal runaway? Can the thermal response of a cell during thermal runaway, as well as cell-to-cell variation, be predicted based on the properties of a cell (chemistry, format, and energy and power density)? Can the combination of engineering models, performance data, and ML provide insights into fault-tolerant materials, components, and cell designs?
(5) Materials to manufacturing: What are the root causes of cell variability versus performance metrics such as cycle life in relatively uncontrolled field deployments and temperature resistance? Can manufacturers solve inverse problems to optimize performance by improving cell design (e.g., electrode geometry, tab placement, etc.)? Can quantification of these uncertainties be translated to commercially significant risk factors in order to catalyze deployment investments most directly?

Table S2 lists 18 different challenge categories for the community to expand upon and prioritize. Safety[89] and other categories will require broader community discussions to reach consensus on priority directions and need for supporting benchmark datasets.

BATTERY DATA GENOME: THE PATH FORWARD

The BDG launch is designed as a phased approach facilitated by two global committees focusing on either technical or commercial impacts which address and so sustain stakeholder interests (Figure 5). The "scientific council" (green arrow) will develop technical standards and protocols and drive broad data-science advancements via engagements focused on their community; the "sustainability council" (blue arrow) focuses on economic and commercial issues which complement the purely technical and which drive sustainability, and it will have a broader stakeholder mix. Committee memberships should rotate, as do boards of professional societies. Coordination through community workshops with broad outreach will be vital to establishing this framework.

Phase 1 focuses on creation of a comprehensive data-science structure via a series of workshops. More details on the workshops are in the supplemental information, including those focused on setting standards, a strategy for interoperability, and establishing a roadmap of increasingly demanding targets to achieve the BDG goals.

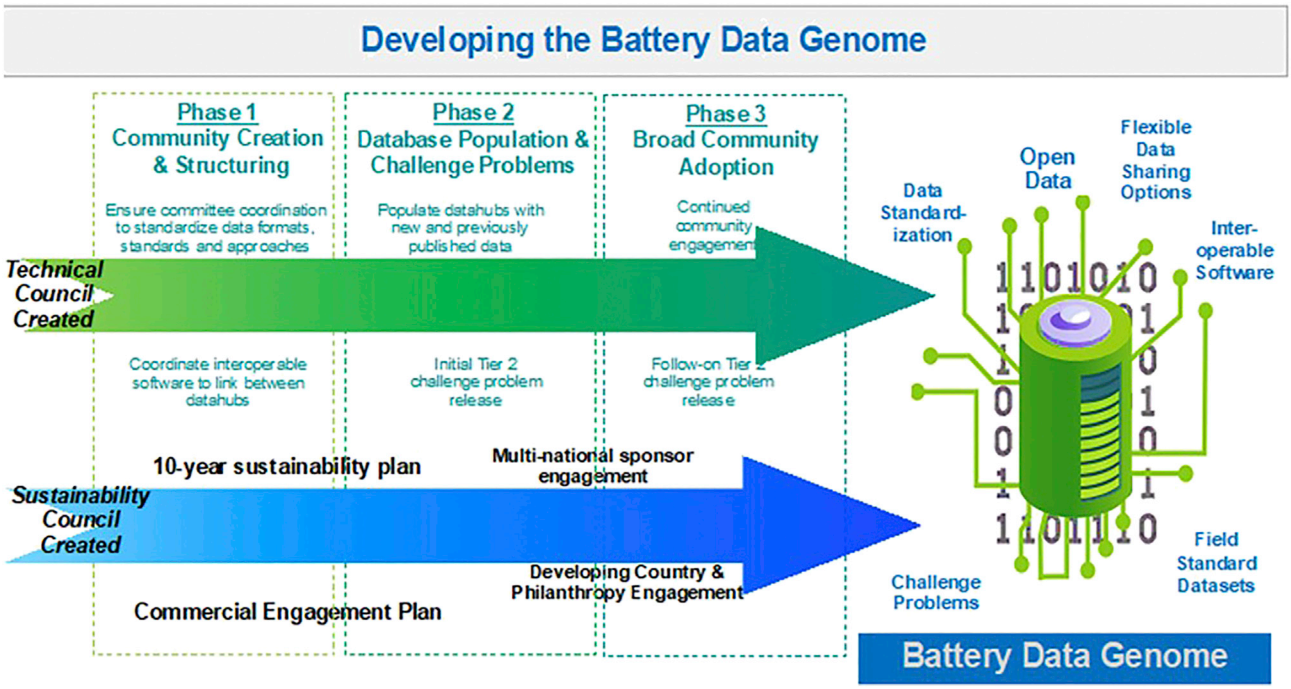

Figure 5. Implementation roadmap for the Battery Data Genome—a multiphase approach with science and sustainability-focused councils

Phase 1 involves expanded participation and outreach to increase community breadth and awareness. It is anticipated that, initially, large portions of the BDG data will come from the authors of this manuscript. Indeed, using BatteryArchive as an example, many of the datasets were generated by co-authors of this manuscript. As other data hubs are developed for different purposes and linked together, we expect that a similar early trend will occur and that, over time, broad community engagement will advance. Starting now and advancing to that point in the near future, methods and protocols to grade and qualify data will be a key focus of the BDG.

In Phase 2, data hubs are populated with newly published data using software tools that build in interoperability; several challenge problems as described above and in the supplemental information, will be run, key benchmark datasets will be developed (Tables S3 and S4) and roadmaps will be refined. Phase 2 thus becomes a joint data expansion and core participant sustainment activity. The focus of Phase 3 is tuning the work structure via feedback—ideally from an even broader community. Each of the phases will require organizational growth and participation. Early efforts (including those described in the supplemental information) have and will continue to focus on technical publications and professional meeting workshops and discussions. As the BDG moves from its current early stages in Phase 1 to Phase 3 activities to engage trade organizations and investment entities, the development of resources for education and building data-reporting practices with key journals to enhance data submission to the BDG will all become more necessary. The authors will also lead by example by not only setting the vision and implantation of the BDG but by actively expanding the data content which resides in BDG data hubs.

The BDG vision and its principles do not inherently lead to sustained activity, even with initial successes. The sixth principle, the sustainability principle, is that financial and organizational activities must begin in concert with technical activities to ensure longevity. The two perspectives must be interwoven to synergize as both are of paramount importance to an enthusiastic and productive community. The complexity of this task is high—a viable structure is difficult to predict but exciting to anticipate.

The sustainability committee (blue arrow) will include mature industry, start-ups, nonprofits, investors, and government representatives—several authors of this paper volunteer for the first phase. Activities could include:

- Setting an initial 10-year financial timeline with: (1) a quick start via seed funding from government and other agencies including nonprofits and philanthropy; (2) development of medium-term support via financial devices such as data-sharing fees (open access fees are currently employed by many journals); and (3) setting a year 5 milestone to create long-term support.
- Establishing a commercial engagement framework to build long-term funding that supports continued data collection and curation, which are woven into the BDG design.

An example project with early commercial engagement that can create impact and coalesce the group might be establishing DERs (distributed energy resources) in developing economies. Climate change is a global issue but resources to meet the challenge are nonuniform, with especially large gaps in developing economies that often depend on philanthropic support. Datasets and open software from the BDG can help build economically optimized and versatile off-grid power systems through detailed data that provide an understanding of battery performance and lifetime.

CONCLUDING REMARKS

We propose the launch of a global battery-data-science initiative, the BDG, for the transformation of battery innovation to contemporary data-science practices in research, development, and deployment. The fundamental gap of large open data resources is solved by creation of data standards and the building a robust community of interconnected data hubs, with flexible sharing practices and an open suite of interoperable software. Battery ML and AI advancements will arise from access to large open datasets and tools. These will drive an unprecedented pace of innovations in research, large-scale manufacturing, and commercial investment that will catalyze the wide-scale battery deployments needed for deep decarbonization, including those in developing economies. If even a portion of the gains seen in similar efforts are realized, such as those of the HGP, the BDG will directly impact the global population.

There is a precedent, and we offer a plan. Early indicators suggest that initial activities will not naturally coalesce and that a hesitancy will remain. From this perspective, we reason that—while the goal of coordinating the global battery community and launching the BDG may seem grandiose and improbable today—ultimately it will be realized as a natural step driven by the desire to innovate vigorously, to combat climate change, and to provide energy security with broad economic gains. The risk is not in moving forward, but in staying as we are now.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.joule.2022.08.008.

ACKNOWLEDGMENTS

The authors would like to recognize the contributions from each of the reviewers. The recommendations and items brought forth for consideration aided considerably in the preparation and ultimate impact of this manuscript. S.B., N.H.P., and L.W. acknowledge financial support from Laboratory Directed Research and Development (LDRD), funding from Argonne National Laboratory, provided by the Director, Office of Science, of the US Department of Energy (DOE) under contract no. DE-AC02-06CH11357. M.D. acknowledges support from the Office of Naval Research #N00014-18-1-2127 and from the State of Hawaii. Y.P. and V.D.A. acknowledge support from the US DOE Office of Electricity, Energy Storage Program, and Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US DOE's National Nuclear Security Administration under contract DE-NA-0003525. K.S., E.J.D., T.R.T., B.-R.C., D.P.F., and S.S acknowledge support from the US DOE Vehicle Technologies Office. S.C. acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 957189 (BIG-MAP). The National Renewable Energy Laboratory is operated by Alliance for Sustainable Energy, LLC for the US DOE under contract no. DE-AC36-08GO28308. Idaho National Laboratory is operated by Battelle Energy Alliance under contract no. DE-AC07-05ID14517 for the US DOE. The views expressed in the article do not necessarily represent the views of the DOE or the US Government. P.D. and D.U.-S. gratefully acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF) for funding within the research cluster greenBattNutzung (03XP0302C) and the German Council of Science and Humanities for funding of the Center for Ageing, Reliability and Lifetime Prediction of Electrochemical and Power Electronic Systems(CARL). E.E.E, V.W.H., and L.T. acknowledge Professor Daniel T. Schwartz at the University of Washington and support provided by Army Research Office contracts #W911NF1710550 and #W911NF1810171 for developing electrochemical data-science activities in partnership with the Electrochemical Society. D.A.H. acknowledges funding from Research England's Connecting Capability Fund ("Pitch-In" project, CCF18-7157) and from the Faraday Institution (EP/S003053/1, grant no. FIRG003).

AUTHOR CONTRIBUTIONS

The ideas described in this paper were refined over the course of 2 years of video conferences. L.W. organized the meetings and wrote the initial draft of this perspective article, with guidance from V.V. and S.B. S.B. and E.D. led the writing and editing of the final draft, with significant input from S.S., D.H., V.V., G.C., and I.F. The technical groups were led by E.E. (organizational and operating principles), C.G. (software principles), V.V. and N.P. (challenge problems), and E.D. (urgent need and path forward). All authors contributed to the concept development and writing of this manuscript.

DECLARATION OF INTERESTS

D.A.H. is co-founder of Brill Power Ltd. and is a technical advisor at Habitat Energy Ltd. V.V. is a Technical Consultant at QuantumScape Corporation and Chief Scientist at Aionics Inc.

REFERENCES

1. Arbabzadeh, M., Sioshansi, R., Johnson, J.X., and Keoleian, G.A. (2019). The role of energy storage in deep decarbonization of electricity production. Nat. Commun. *10*, 3413.

2. Davies, D.M., Verde, M.G., Mnyshenko, O., Chen, Y.R., Rajeev, R., Meng, Y.S., and Elliott, G. (2019). Combined economic and technological evaluation of battery energy storage for grid applications. Nat. Energy *4*, 42–50.

3. Nykvist, B., and Nilsson, M. (2015). Rapidly falling costs of battery packs for electric vehicles. Nat. Clim. Change *5*, 329–332.

4. Nykvist, B., Sprei, F., and Nilsson, M. (2019). Assessing the progress toward lower priced long range battery electric vehicles. Energy Policy *124*, 144–155.

5. Nobel Committee for Chemistry (2019). Press release: The Nobel Prize in Chemistry 2019. https://www.nobelprize.org/prizes/chemistry/2019/press-release/.

6. U.S. Department of Energy (2020). Energy Storage Grand Challenge Draft Roadmap. https://www.energy.gov/sites/prod/files/2020/07/f76/ESGC%20Draft%20Roadmap_2.pdf.

7. Battery, 2030+. (2021). https://battery2030.eu/.

8. The Faraday Institution. Faraday battery challenge. https://faraday.ac.uk/.

9. wwPDB consortium (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic Acids Res. *47*, D520–D528.

10. Hill, J., Mulholland, G., Persson, K., Seshadri, R., Wolverton, C., and Meredig, B. (2016). Materials science with large-scale data and informatics: unlocking new opportunities. MRS Bull. *41*, 399–409.

11. Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., et al. (2019). Current status of Landsat program, science, and applications. Remote Sens. Environ. *225*, 127–147.

12. Adams, P.D., Afonine, P.V., Baskaran, K., Berman, H.M., Berrisford, J., Bricogne, G., Brown, D.G., Burley, S.K., Chen, M., Feng, Z., et al. (2019). Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB). Acta Crystallogr. D Struct. Biol. *75*, 451–454.

13. Hall, S.R., Allen, F.H., and Brown, I.D. (1991). The crystallographic information file (CIF): a new standard archive file for crystallography. Acta Crystallogr. A Found. Crystallogr. *47*, 655–685.

14. Ward, L., Aykol, M., Blaiszik, B., Foster, I., Meredig, B., Saal, J., and Suram, S. (2018). Strategies for accelerating the adoption of materials informatics. MRS Bull. *43*, 683–689.

15. Khorshidi, A., and Peterson, A.A. (2016). Amp: a modular approach to machine learning in atomistic simulations. Comput. Phys. Commun. *207*, 310–324.

16. Schütt, K.T., Kessel, P., Gastegger, M., Nicoli, K.A., Tkatchenko, A., and Müller, K.-R. (2018). SchNetPack: a deep learning toolbox for atomistic systems. J. Chem. Theory Comput. *15*, 448–455.

17. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555.

18. Usseglio-Viretta, F.L.E., Finegan, D.P., Colclasure, A., Heenan, T.M.M., Abraham, D., Shearing, P., and Smith, K. (2020). Quantitative relationships Between pore tortuosity, pore topology, and solid particle morphology using a novel discrete particle size algorithm. J. Electrochem. Soc. *167*, 100513.

19. de Pablo, J.J., Jackson, N.E., Webb, M.A., Chen, L.-Q., Moore, J.E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D.G., Toberer, E.S., et al. (2019). New frontiers for the materials genome initiative. npj Comp. Mater. *5*, 41.

20. Liu, J., Bao, Z., Cui, Y., Dufek, E.J., Goodenough, J.B., Khalifah, P., Li, Q., Liaw, B.Y., Liu, P., Manthiram, A., et al. (2019). Pathways for practical high-energy long-cycling lithium metal batteries. Nat. Energy *4*, 180–186.

21. Harlow, J.E., Ma, X., Li, J., Logan, E., Liu, Y., Zhang, N., Ma, L., Glazier, S.L., Cormier, M.M.E., Genovese, M., et al. (2019). A wide range of testing results on an excellent lithium-ion cell chemistry to be used as benchmarks for new battery technologies. J. Electrochem. Soc. *166*, A3031–A3044.

22. Isidore, C., and Valdes-Dapena, P. (2021). Hyundai's recall of 82,000 electric cars is one of the most expensive in history. Battery Fires, News Articles (CNN Business). https://www.cnn.com/2021/02/25/tech/hyundai-ev-recall/index.html.

23. Gasper, P., Gering, K., Dufek, E., and Smith, K. (2021). Challenging practices of algebraic battery life models through statistical validation and model identification via achine-learning. J. Electrochem. Soc. *168*, 020502.

24. Persons, T.M., and Mackin, M. (2020). Technology readiness best practices for evaluating the readiness of technology for use in acquisition programs and projects. (U.S. Government Accountability Office Washington United States). https://apps.dtic.mil/sti/pdfs/AD1105846.pdf.

25. Sendek, A.D., Ransom, B., Cubuk, E.D., Pellouchoud, L.A., Nanda, J., and Reed, E.J. (2022). Machine learning modeling for accelerated battery materials design in the small data regime. Adv. Energy Mater. *12*, 2200553.

26. Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.-H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature *578*, 397–402.

27. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391.

28. Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. J. Power Sources *527*, 231127.

29. Aitio, A., and Howey, D.A. (2021). Predicting battery end of life from solar off-grid system field data using machine learning. Joule *5*, 3204–3220.

30. Chen, B.-R., Kunz, M.R., Tanim, T.R., and Dufek, E.J. (2021). A machine learning framework for early detection of lithium plating combining multiple physics-based electrochemical signatures. Cell Rep. Phys. Sci. *2*, 100352.

31. Dubarry, M., and Beck, D. (2020). Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis. J. Power Sources *479*, 228806.

32. Kim, S., Yi, Z., Chen, B.-R., Tanim, T.R., and Dufek, E.J. (2022). Rapid failure mode classification and quantification in batteries: a deep learning modeling framework. Energy Storage Mater. *45*, 1002–1011.

33. Dubarry, M., and Beck, D. (2021). Analysis of synthetic voltage vs. capacity datasets for big data Li-ion diagnosis and prognosis. Energies *14*, 2371.

34. Dufek, E.J., Tanim, T.R., Chen, B.-R., and Sangwook, K. (2022). Battery calendar aging and machine learning. Joule *6*, 1363–1367.

35. Sendek, A.D., Cubuk, E.D., Antoniuk, E.R., Cheon, G., Cui, Y., and Reed, E.J. (2019). Machine learning-assisted discovery of solid Li-ion conducting materials. Chem. Mater. *31*, 342–352.

36. Barrett, D.H., and Haruna, A. (2020). Artificial intelligence and machine learning for targeted energy storage solutions. Curr. Opin. Electrochem. *21*, 160–166.

37. Tabor, D.P., Roch, L.M., Saikin, S.K., Kreisbeck, C., Sheberla, D., Montoya, J.H., Dwaraknath, S., Aykol, M., Ortiz, C., Tribukait, H., et al. (2018). Accelerating the discovery of materials for clean energy in the era of smart automation. Nat. Rev. Mater. *3*, 5–20.

38. Flores-Leonar, M.M., Mejía-Mendoza, L.M., Aguilar-Granda, A., Sanchez-Lengeling, B., Tribukait, H., Amador-Bedolla, C., and Aspuru-Guzik, A. (2020). Materials acceleration platforms: on the way to autonomous experimentation. Curr. Opin. Green Sustain. Chem. *25*, 100370.

39. Dave, A., Mitchell, J., Kandasamy, K., Wang, H., Burke, S., Paria, B., Póczos, B., Whitacre, J., and Viswanathan, V. (2020). Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. Cell Rep. Phys. Sci. *1*, 100264.

40. Crabtree, G. (2020). Self-driving laboratories coming of age. Joule *4*, 2538–2541.

41. dos Reis, G., Strange, C., Yadav, M., and Li, S. (2021). Lithium-ion battery data and where to find it. Energy Ai *5*, 100081.

42. Clark, S., Bleken, F.L., Stier, S., Flores, E., Andersen, C.W., Marcinek, M., Szczesna-Chrzan, A., Gaberscek, M., Palacin, M.R., Uhrin, M., and Friis, J. (2022). Toward a unified description of battery data. Adv. Energy Mater. *12*, 2102702.

43. NREL (2020). Battery Microstructures Library. https://www.nrel.gov/transportation/microstructure.html.

44. Quinn, A., Moutinho, H., Usseglio-Viretta, F., Verma, A., Smith, K., Keyser, M., and Finegan, D.P. (2020). The application of electron backscatter diffraction for investigating intra-particle grain architectures and boundaries in lithium ion electrodes. Cell Rep. Phys. Sci. *1*, 100137.

45. McShane, E.J., Colclasure, A.M., Brown, D.E., Konz, Z.M., Smith, K., and McCloskey, B.D. (2020). Quantification of inactive lithium and solid{\textendash}electrolyte interphase species on graphite electrodes after fast charging. ACS Energy Lett. *5*, 2045–2051.

46. Verma, A., Smith, K., Santhanagopalan, S., Abraham, D., Yao, K.P., and Mukherjee, P.P. (2017). Galvanostatic intermittent titration and performance based analysis of $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$ cathode. J. Electrochem. Soc. *164*, A3380–A3392.

47. Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. J. Power Sources *341*, 373–386.

48. Ecker, M., Tran, T.K.D., Dechent, P., Käbitz, S., Warnecke, A., and Sauer, D.U. (2015). Parameterization of a physico-chemical model of a lithium-ion battery. J. Electrochem. Soc. *162*, A1836–A1848.

49. Rynne, O., Dubarry, M., Molson, C., Nicolas, E., Lepage, D., Prébé, A., Aymé-Perrot, D., Rochefort, D., and Dollé, M. (2020). Exploiting materials to their full potential, a Li-ion battery electrode formulation optimization study. ACS Appl. Energy Mater. *3*, 2935–2948.

50. Murray, V., Hall, D.S., and Dahn, J.R. (2019). A guide to full coin cell making for academic researchers. J. Electrochem. Soc. *166*, A329–A333.

51. Tanim, T.R., Dufek, E.J., Evans, M., Dickerson, C., Jansen, A.N., Polzin, B.J., Dunlop, A.R., Trask, S.E., Jackman, R., Bloom, I., et al. (2019). Extreme fast charge challenges for lithium-ion battery: variability and positive electrode issues. J. Electrochem. Soc. *166*, A1926–A1938.

52. Ecker, M., Gerschler, J.B., Vogel, J., Käbitz, S., Hust, F., Dechent, P., and Sauer, D.U. (2012). Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data. J. Power Sources *215*, 248–257.

53. Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D.U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. J. Power Sources *247*, 332–338.

54. Preger, Y., Barkholtz, H.M., Fresquez, A., Campbell, D.L., Juba, B.W., Romàn-Kustas, J., Ferreira, S.R., and Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. J. Electrochem. Soc. *167*, 120532.

55. Tanim, T.R., Shirk, M.G., Bewley, R.L., Dufek, E.J., and Liaw, B.Y. (2018). Fast charge implications: pack and cell analysis and comparison. J. Power Sources *381*, 56–65.

56. Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., and Howey, D.A. (2021). The challenge and opportunity of battery lifetime prediction from field data. Joule *5*, 1934–1955.

57. Battery Archive. https://www.batteryarchive.org/.

58. Herring, P., Balaji Gopal, C., Aykol, M., Montoya, J.H., Anapolsky, A., Attia, P.M., Gent, W., Hummelshøj, J.S., Hung, L., Kwon, H.-K., et al. (2020). BEEP: a Python library for battery evaluation and early prediction. SoftwareX *11*, 100506.

59. Doan, H.A., Agarwal, G., Qian, H., Counihan, M.J., Rodríguez-López, J., Moore, J.S., and Assary, R.S. (2020). Quantum chemistry-informed active learning to accelerate the design and discovery of sustainable energy storage materials. Chem. Mater. *32*, 6338–6346.

60. Martin Ebner, V.W. (2014). Tool for tortuosity estimation in lithium ion battery porous electrodes. J. Electrochem. Soc. *162*, A3064–A3070.

61. Cooper, S.J., Bertei, A., Shearing, P.R., Kilner, J.A., and Brandon, N.P. (2016). Taufactor: an open-source application for calculating tortuosity factors from tomographic data. SoftwareX *5*, 203–210.

62. NREL (2020). MATBOX: Microstructure Analysis Toolbox. https://www.nrel.gov/transportation/matbox.html.

63. Betz, J., Bieker, G., Meister, P., Placke, T., Winter, M., and Schmuch, R. (2018). Theoretical versus practical energy: a plea for more transparency in the energy calculation of different rechargeable battery systems. Adv. Energy Mater. *9*, 1803170.

64. Neubauer, J., and Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. J. Power Sources *257*, 12–20.

65. NREL (2020). National Solar Radiation Database: typical meteorological year. https://nsrdb.nrel.gov/data-sets/tmy.

66. NREL (2020). Transportation Secure Data Center. https://www.nrel.gov/transportation/secure-transportation-data/.

67. Mayilvahanan, K.S., Takeuchi, K.J., Takeuchi, E.S., Marschilok, A.C., and West, A.C. (2022). Supervised learning of synthetic big data for Li-ion battery degradation diagnosis. Batter. Supercaps *5*, e20210016.

68. Ayerbe, E., Berecibar, M., Clark, S., Franco, A.A., and Ruhland, J. (2022). Digitalization of battery manufacturing: current status, challenges, and opportunities. Adv. Energy Mater. *12*, 2102696.

69. Barai, A., Uddin, K., Dubarry, M., Somerville, L., McGordon, A., Jennings, P., and Bloom, I. (2019). A comparison of methodologies for the non-invasive characterisation of commercial Li-ion cells. Prog. Energy Combust. Sci. *72*, 1–31.

70. Chen, S., Niu, C., Lee, H., Li, Q., Yu, L., Xu, W., Zhang, J.-G., Dufek, E.J., Whittingham, M.S., Meng, S., et al. (2019). Critical parameters for evaluating coin cells and pouch cells of rechargeable Li-metal batteries. Joule *3*, 1094–1105.

71. Nagpure, S.C., Tanim, T.R., Dufek, E.J., Viswanathan, V.V., Crawford, A.J., Wood, S.M., Xiao, J., Dickerson, C.C., and Liaw, B. (2018). Impacts of lean electrolyte on cycle life for rechargeable Li metal batteries. J. Power Sources *407*, 53–62.

72. Ue, M., Sakaushi, K., and Uosaki, K. (2020). Basic knowledge in battery research bridging the gap between academia and industry. Mater. Horiz. *7*, 1937–1954.

73. Long, B.R., Rinaldo, S.G., Gallagher, K.G., Dees, D.W., Trask, S.E., Polzin, B.J., Jansen, A.N., Abraham, D.P., Bloom, I., Bareño, J., and Croy, J.R. (2016). Enabling high-energy, high-voltage lithium-ion cells: standardization of coin-cell assembly, electrochemical testing, and evaluation of full cells. J. Electrochem. Soc. *163*, A2999–A3009.

74. Naumann, M., Spingler, F.B., and Jossen, A. (2020). Analysis and modeling of cycle aging of a commercial $LiFePO_4$/graphite cell. J. Power Sources *451*, 227666.

75. Genovese, M., Louli, A.J., Weber, R., Hames, S., and Dahn, J.R. (2018). Measuring the coulombic efficiency of lithium metal cycling in anode-free lithium metal batteries. J. Electrochem. Soc. *165*, A3321–A3325.

76. Rahe, C., Kelly, S.T., Rad, M.N., Sauer, D.U., Mayer, J., and Figgemeier, E. (2019). Nanoscale X-ray imaging of ageing in automotive lithium ion battery cells. J. Power Sources *433*, 126631.

77. Dubarry, M., Devie, A., Stein, K., Tun, M., Matsuura, M., and Rocheleau, R. (2017). Battery energy storage system battery durability and reliability under electric utility grid operations: analysis of 3 years of real usage. J. Power Sources *338*, 65–73.

78. Stephan, A.K. (2021). Standardized battery reporting guidelines. Joule *5*, 1–2.

79. Ruiz, V., Pfrang, A., Kriston, A., Omar, N., den Bossche Van, P., and Boon-Brett, L. (2018). A review of international abuse testing standards and regulations for lithium ion batteries in electric and hybrid electric vehicles. Renew. Sustain. Energy Rev. *81*, 1427–1452.

80. Joint Research Centre (European Commission) (2018). Standards for the performance and durability assessment of electric vehicle batteries: possible performance criteria for an ecodesign regulation. (Publications Office). https://data.europa.eu/doi/10.2760/24743.

81. Sulzer, V., Marquis, S.G., Timms, R., Robinson, M., and Chapman, S.J. (2020). Python Battery Mathematical Modelling (PyBaMM). J. Open Res. Softw. *9*, 14.

82. Torchio, M., Magni, L., Gopaluni, R.B., Braatz, R.D., and Raimondo, D.M. (2016).

LIONSIMBA: a Matlab framework based on a finite volume model suitable for Li-ion battery design, simulation, and control. J Electrochem Soc. *163*, A1192–A1205.

83. Smith, R.B., and Bazant, M.Z. (2017). Multiphase porous electrode theory. J. Electrochem. Soc. *164*, E3291–E3310.

84. Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. J. Electrochem. Soc. *166*, A3189–A3200.

85. Lewis-Douglas, A., Pitt, L., and Howey, D.A. (2020). Galvanalyser: a battery test database. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.14959.

86. Feng, X., Merla, Y., Weng, C., Ouyang, M., He, X., Liaw, B.Y., Santhanagopalan, S., Li, X., Liu, P., Lu, L., et al. (2020). A reliable approach of differentiating discrete sampled-data for battery diagnosis. eTransportation *3*, 100051.

87. de Angelis, V., Preger, Y., and Chalamala, B.R. (2021). Battery lifecycle framework: a flexible repository and visualization tool for battery datafrom materials development to field implementation. Preprint at ECSarXiv. https://doi.org/10.1149/osf.io/h7c24.

88. Herring, P., Balaji Gopal, C., Aykol, M., Montoya, J.H., Anapolsky, A., Attia, P.M., Gent, W., Hummelshøj, J.S., Hung, L., Kwon, H.-K., et al. (2020). BEEP GitHub. SoftwareX *11*, 100506. https://doi.org/10.1016/j.softx.2020.100506.

89. Finegan, D.P., Zhu, J., Feng, X., Keyser, M., Ulmefors, M., Li, W., Bazant, M.Z., and Cooper, S.J. (2021). The application of data-driven methods and physics-based learning for improving battery safety. Joule *5*, 316–329.