SAND2022-14826C

Global Energy Storage Database: Enhancing Features and Validation Procedure

Ujjwol Tamrakar, Alvaro Furlani Bastos, Samuel Roberts-Baca, Sahil Bhalla, Will McNamara, Tu A. Nguyen†
Sandia National Laboratories
Albuquerque, New Mexico, USA
Email: tunguy@sandia.gov†

***Abstract*—Large-scale deployment of energy storage systems is a pivotal step toward achieving the clean energy goals of the future. An accurate and publicly accessible database on energy storage projects can help accelerate deployment by providing valuable information and characteristic data to different stakeholders. The U.S. Department of Energy's Global Energy Storage Database (GESDB) aims at providing high-quality and accurate data on energy storage projects around the globe. This paper first provides an overview of the GESDB, briefly describing its features and overall usage. This is followed by a detailed description of the procedure used to validate the database. In doing so, the paper aims at improving the usability of the website while enhancing its value to the community. Furthermore, the presented validation procedure makes the underlying assumptions transparent to the public so that data misinterpretation can be minimized/avoided.**

***Index Terms*—Energy storage, energy storage policy, global energy storage database, grid storage technology, validation.**

I. Introduction

Energy storage will play a vital role in addressing climate change and minimizing (and ultimately negating) the carbon footprint of the electric grid [1]. To this end, publicly available and accurate information on energy storage projects are critical for the widespread deployment and adoption of energy storage technologies. High-quality data can drive energy policy decisions, accelerate grid integration, and reduce uncertainty for investors. However, having such data available at one convenient location is difficult. Challenges include: 1) Limited availability of publicly accessible data on energy storage projects; 2) Data may not be properly validated; 3) Lack of simple tool(s) to analyze raw data and extract useful information. The U.S. Department of Energy (DOE)'s Global Energy Storage Database (GESDB) aims at addressing these challenges by providing a platform for unbiased, research-grade database of grid-connected energy storage projects worldwide. GESDB provides detailed characteristic information of various energy storage installations throughout the world. It also provides a collection of U.S. state and federal policy information related to energy storage. The database is intended for a variety of stakeholders including academia, researchers, utilities, industry, policymakers, and investors. The primary objective of the GESDB is to serve as a repository for the ever-growing fleet of energy storage installations worldwide. It is expected that maintaining such a repository will assist in the widespread deployment of energy storage and help achieve the goal of de-carbonization.

The objective of this paper is to give an overview of the various features of the GESDB so that users can utilize the database to its full potential. Since accurate validated data is critical, this paper will also outline a robust validation process that has been designed to validate the database. The expectation is that this validation procedure will not only help improve the quality of the database but also provide insights for the user as to how to consume and extrapolate correct information from the database.

II. Features of the Current GESDB

This section describes all the existing features from the current release of the GESDB. The GESDB can currently be accessed at [2]. Fig. 1 shows a screenshot of the main landing page. The GESDB has been developed over the years by various partners. An overview of the history and development of the GESDB to its current state is described in [3], [4]. As of the publication of this paper, the GESDB has a total of 1,694 project entries from around the world. Each of these projects have more than 100 sub-fields which provide a plethora of information that characterizes the energy storage project. This amounts to data on about 190 GW of energy storage projects worldwide.

Fig. 1: A screenshot of the Global Energy Storage Database's main landing page.

The database covers a wide array of energy storage technologies and applications. More specifically, the technology

type of the projects has been carefully curated based on the classification proposed in the recently published DOE's Energy Storage Handbook [5]. Similarly, the applications for which a particular project is used are categorized in a standard manner as well according to DOE's Energy Storage Handbook. The applications are categorized as follows:

1) *Bulk energy services (General energy applications)*: Energy arbitrage, renewable energy time-shift
2) *Ancillary services*: Frequency regulation, operating reserve (spinning, non-spinning, supplementary), frequency response and virtual inertia, voltage support, ramp support, black start
3) *Transmission infrastructure services*: Transmission upgrade deferral, transmission congestion relief, stability damping control
4) *Distribution infrastructure services*: Peak shaving and upgrade deferral, voltage regulation, reliability and resilience
5) *Customer energy management services*: Time-of-use, demand charge, and net-metering management
6) *Other services*: Transportation services, microgrid applications

In terms of GESDB's website architecture, there are three main web pages which are described in the subsections below:

A. Projects Page

The *Projects* page is the primary page for the GESDB where the users can navigate, search, and download a particular project or a set of projects. Fig. 2 shows a screenshot of the projects page along with the main components of the page. The page consists of an interactive global map from which the users can navigate to individual projects at a specific location. The page also provides an advanced search option where the users can search for desired projects by name, location, technology type, rated power, duration, and status. The information on the selected project(s) can then be downloaded in either CSV or JSON file formats.

B. Policies Page

As the energy storage policy landscape in the U.S. continues to evolve, both at the federal and state levels, the GESDB aims to keep track of this changing scenario. Relevant policy-making activities specific to energy storage at the federal and state levels are published and made available to the public through the *Policies* page in the GESDB. A screenshot of the policies page is shown in Fig. 3. This page provides summaries of energy storage policies, legalisation, and regulatory rules. Along with this, data regarding energy storage project deployments and short, analytical papers on significant policy topics (referred to as *Issue Briefs* on the website) are also available. The page also provides links to various Federal Energy Regulatory Commission (FERC) orders related to energy storage in one central location.

Fig. 2: A screenshot of the Projects page. The main navigation components are marked in red.

Fig. 3: A screenshot of the Policies page. The states for which detailed reports are available are highlighted in dark blue.

C. Statistics and Visualization

The *Statistics* page is the latest addition to the GESDB. The objective of this page is to provide interactive statistics and visualizations of the data maintained within GESDB. This page provides a set of simple tools for users to quickly process and analyze raw data available in GESDB. Users can interact with various filters and tools provided within each of the visualizations to obtain desired information in the plots. These plots can then be downloaded either in PDF format or as images. Fig. 4 shows a screenshot from the statistics page. The plot of energy storage installations by year is shown as an example. In addition to this, the following plots are available:

1) Cumulative Sum of Energy Storage Installations by Year
2) Pie-chart Breakdown of Energy Storage Installations by Technology Type
3) Global Map of Energy Storage Installations

4) Rated Power of Energy Storage Installations by Application
5) Rated Capacity of Energy Storage Installations by Application

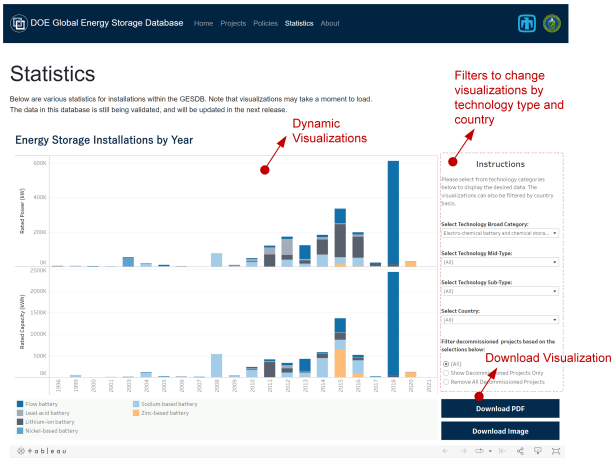

Fig. 4: A screenshot of the Statistics page. The main navigation components of the visualization page are pointed out in red.

III. Validation Methodology

To improve the quality, accuracy, and reliability of GESDB, a robust validation methodology has been developed. This methodology can validate the existing database and any new projects that will be added into the database in the future. In this section, a detailed overview of the developed validation procedure is provided. The objective of describing the validation procedure in this paper is to make the validation procedure transparent so that users can gain confidence in extracting information from the database. It is expected that through this detailed description of the procedure, misinterpretation of the data can be minimized/avoided. Fig. 5 shows a schematic diagram of the procedure developed for validating the GESDB. The process consists of three major steps:

A. Step 1: Identification

The first step involves checking the data entries for different kinds of errors. The flowchart providing further details of the identification process is illustrated in Fig. 6. The identification process takes in a GESDB schema file and the actual database as the inputs. The GESDB schema is a separate file that defines the validation rules for each field in the database. Based on these carefully defined rules, automated Python scripts check for various errors in the database. The database also inputs all validation status results from previous validation steps. Various checks are then performed to identify possibly erroneous data. Some of the checks are listed below:

1) Check the required data type for each field and whether the field entry is required or optional.
2) Validate any applicable lower or upper bound on the numerical values.
3) For fields that require a specific value such as *Applications*, *Status of the Project*, *Technology Type*, etc., check whether the entry is within the valid set of values. As mentioned earlier, the values for these fields must be in alignment with the definitions provided in DOE's energy storage handbook.
4) Check to validate if the address fields (such as *City*, *State/Province/Territory*, and *Country*) have valid entries and the project's latitude and longitude values match with the given address information.
5) Check to validate if the dates (such as *announced*, *constructed*, *commissioned*, and *de-commissioned*) are in the correct format (mm-dd-YYYY) and chronological order.
6) Check for possibly inaccurate data. Examples include – checks for whether the ratio of storage devices' capacity to rated power is within a threshold value.

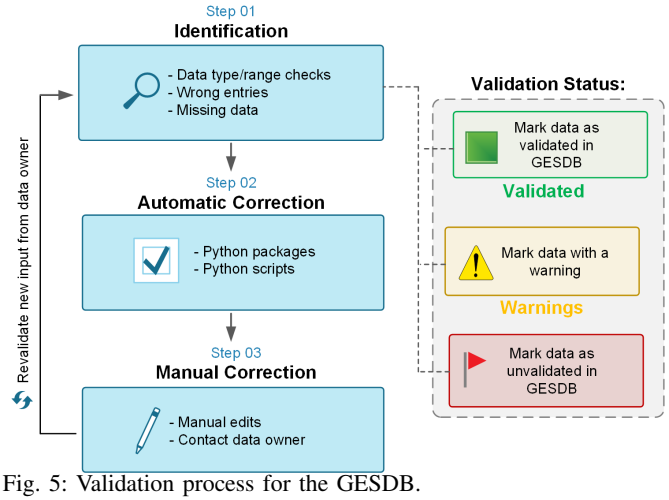

Fig. 5: Validation process for the GESDB.

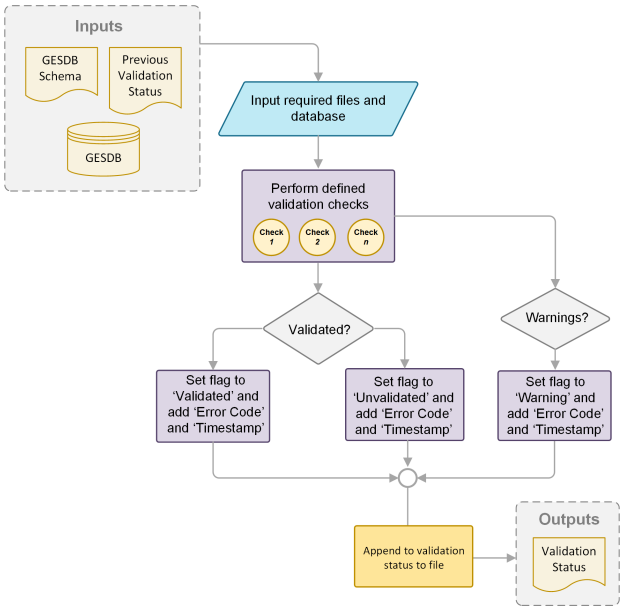

Fig. 6: Flowchart of the identification process.

The checks mentioned above are all automated through Python scripts/modules. These Python scripts go through each

field within each database entry and check the validity of the entry based on the corresponding rules defined in the GESDB schema. For each kind of check, a separate Python function is implemented. These functions take in the data from each field and each database entry as the input and outputs a validation status (*Validated*, *Warnings*, or *Unvalidated*), along with a description of the errors, corresponding error codes, and a timestamp to mark when the validation was last performed. These statuses have the following meanings:

- *Validated*: This status means that the data field entry is correct based on all rules defined in the schema file.
- *Warnings*: Data is marked with a warning status if there are *possible* errors. Warnings need manual inspection and correction of data in most cases. For instance, data fields are marked with a *Warning* status if the ratio of storage devices' capacity to rated power is not within a threshold value or if there are issues with date formatting and chronological order.
- *Unvalidated*: This status means that there are obvious errors in the data field. For instance, wrong data type, data lower and/or upper bound being violated.

The outputs from all of these checks are then aggregated by a main Python script, which checks for validation statuses from all the modules and evaluates whether there are any errors. The aggregator also combines all the error descriptions and error codes generated by individual checks into a single location. The outputs are then written into a validation status file. This approach provides extensive modularity and flexibility to the validation procedure. The individual validation checks can easily be refactored in the future without impacting the other checks or the aggregation process. Furthermore, if additional checks/rules are defined for a specific field in the future, a new Python module can quickly be created and integrated into the main aggregator script with minimal effort.

B. Step 2: Automatic Correction

Once the erroneous data entries are identified, the next step in the validation procedure is *Automatic Correction* as shown in Fig. 5. In this step, existing Python packages and custom Python scripts are leveraged to automate the correction of erroneous data.

An example of this step is the correction of the projects' locations. Some of the issues with the fields related to the projects' locations in the entries of the current GESDB include the following:

- Typos or incorrect values in the address fields (such as *City*, *State/Province/Territory*, and *Country*).
- Missing latitude and/or longitude values. Some database entries had these missing fields set to 0, which is a valid, but incorrect, value.
- Rounded-off latitude and/or longitude values, which would cause some projects to be displayed on the sea.

These issues are fixed through the *geopy* library in Python, which allows to correct the typos in the address fields, as well as estimate the correct latitude and longitude values based on the address information.

C. Step 3: Manual Correction

Not all of the erroneous data entries can be corrected through automated scripts. Hence, the final step in the validation process is to correct any erroneous data through manual edits. However, the baseline validation procedures carried out in Steps 1 and 2 provide an easy way to locate erroneous data for manual correction. The previous steps allow not only to locate the errors, but also contain detailed information on why those values are incorrect (error description, error codes), so that manual editing is simplified to some level. This step involves extensive research on the project to acquire the correct information on the project and may involve contacting the data owner or the contact point of the project to get the most up-to-date and accurate information. Once, the corrections in Step 2 and Step 3 are completed the identification process of Step 1 is repeated again for a final validation check. (Note: The GESDB team is continuously working to correct these erroneous data with each new release of the GESDB. Readers should look into the website for the latest information regarding the validation status.)

IV. Results from Validation Procedure

In this section, the results from the validation procedure performed in the current GESDB are presented. This provides some quantitative measures to assess the quality and accuracy of the current database.

A. Error/Warning Message Count in Current GESDB

Table I shows the results obtained from running the developed validation procedure in the current version of the GESDB (as of June 28, 2022). Different errors and the total count of these errors are listed. Note that error counts include data fields that are marked as unvalidated and data fields with warnings. There are a total of 10701 data fields where data that is defined as required in the schema are missing. Even though this number seems large, it should be kept in mind that the rules defined in the schema file are quite stringent. Hence, even though some of the fields of a particular project may have warnings/unvalidated status, the project still has usefeul information for the user. Additional information on each project, that can add more value to the GESDB are the ones that are missing and the team is continuously working towards adding this information. Various other types of errors/warnings are also reported in Table I but the total count of these errors is comparatively low. This initial analysis shows that the GESDB currently has the basic information available on most projects but there is room for improvement. The errors/warnings need to be corrected initially through some automatic procedures and finally, manual verification is still needed to completely validate the database.

B. Examples of Wrong Location Detection and Correction Method

The following list provides some examples of the issues in the address fields that have been corrected through Python scripts according to the procedure presented in Section III-B:

TABLE I: Results from Validation Procedure

Required Data Missing	Data Range Violations	Invalid Data Entries	Capacity to Rated Power Ratio Exceeded	Date Error Warning
10701	1795	126	70	90

- Project #1: *State/Province/Territory* is *Trentino* rather than *Tretino*;
- Project #166: *City* is *Miryang* rather than *Milyang*;
- Project #323: *Country* is *United States* rather than *Unites States*;
- Project #1188: *Country* is *Spain* rather than *United States*;
- Project #1385: *Country* is *India* rather than *United States*.

Further, some of the energy storage projects in South Korea would have the *Country* field set to *South Korea* while others would be *Korea, South*. This inconsistency would result in *South Korea* and *Korea, South* being considered different countries, which affects statistics based on projects' locations. For consistency, the value *South Korea* has been adopted throughout the database. Finally, Table II reports the number of projects that had their latitude and/or longitude values corrected based on the address information (out of the 1,694 current projects in the database).

TABLE II: Results from Validation Procedure – Location Fields

	Latitude	Longitude	Both
Missing	–	–	2
Set to 0	1	220	6
Rounded off	–	19	398

V. Limitations and Future Plans

Although the developed validation procedure will greatly improve the quality of the current database there are however some limitations. There are instances where the identification/corrections steps of the validation procedure will not be able to identify problematic data without some manual intervention. For instance, if a database entry is incorrectly labeled to be a 1MW/1MWh Li-ion system, it would be virtually impossible for the validation procedure to classify these values as incorrect, as they are logically valid and fall within typical ranges. The only approach for detecting this error would be to manually go over each field for each project (which may be impractical) or if an user reports this error.

The GESDB is in continuous development and a number of features and updates are planned to improve its usability and value. A separate page will be created for a record of projects that were funded by the U.S. DOE's Energy Storage Program. The next major feature that is planned to be released in the next version of GESDB is a dedicated page for new project submissions. Currently, users can only provide information through email. The new submission page will provide a convenient method to provide new information for the GESDB. At the back-end, the submission page will run the validation procedure outlined in Section III to validate the data as it is being entered. To enrich the database further a yearly request for information (RFI) will be sent out to get information on the latest energy storage projects. From the database management aspect, it is the team's objective to transition the current database to a more formal database structure (SQL, MongoDB) instead of using plain CSV/JSON file format. This will provide more flexibility as the database grows in the future and will provide better usability for advanced users. Finally, an API is also under planning for efficiently accessing/querying the database.

VI. Conclusions

An overview of the features and usage of a database for grid-connected energy storage installations around the globe called the Global Energy Storage Database (GESDB) was presented in this paper. The database can serve as a useful resource for various stakeholders including academia, researchers, industry, investors, and policymakers to accelerate technological advancement and deployment of energy storage systems on the power grid. The paper also presented a validation procedure to validate and improve the accuracy of the current database. It is expected that this makes the content and the validation procedure transparent and publicly accessible so as to minimize misinterpretation of data.

Acknowledgment

The authors would like to thank Atri Bera of Sandia National Laboratories for the internal review of this paper.

This work is supported by the US DOE Energy Storage Program managed by Dr. Imre Gyuk of the DOE Office of Electricity Delivery and Energy Reliability. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly-owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy National Nuclear Security Administration under contract DE-NA-0003525.

References

[1] R. Masiello, R. Fioravanti, B. Chalamala, and H. Passell, "Electrification, decarbonization, and the future carbon-free grid: the role of energy storage in the electric grid infrastructure [point of view]," *Proceedings of the IEEE*, vol. 110, no. 3, pp. 324–333, 2022.

[2] "U.S. Department of Energy's Global Energy Storage Database (GESDB)," https://gesdb.sandia.gov.

[3] J. Hernández, C. Christensen, and I. Gyuk, "Developing a flexible platform to improve the DOE global energy storage database." Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2016.

[4] J. Hernández, I. Gyuk, and C. Christensen, "DOE global energy storage database—a platform for large scale data analytics and system performance metrics," in *IEEE International Conference on Power System Technology (POWERCON)*. IEEE, 2016, pp. 1–6.

[5] T. Nguyen, R. Byrne, R. Trevizan, A. Bastos, F. Wilches-Bernal, R. Concepcion, R. Tonkoski, and A. Headley, "Applications and grid services," Tech. Rep., 2020.