Green Button and SEED Platform: Enhancing the SEED Platform to Support Building Portfolio Investment and Policy Decision Making

Azizan Aziz, Bertrand Lasternas, Vivian Loftness, Haopeng Wang, Yunjeong Mo, Ting Wang, Chenlu Zhang, Shilpi Sharma, Yujie Xu, Kai-Wei Hsu, Amit Nambiar: Center for Building Performance and Diagnostics, Carnegie Mellon University
Elena Alschuler: View Inc. Harry Bergmann: U.S. Department of Energy

ABSTRACT

In 2009, the ARRA stimulus funding for smart grid projects resulted in the tripling of smart meter deployment. To date, more than 60 million smart meters have been deployed nationwide. The availability of finely granulated energy consumption data through Green Button provides an enormous potential for energy data analytics.

In this paper, we will present a building portfolio-level utility data analytics platform, capitalizing on 15-minute-interval Green Button and monthly utility data. The analytics is geared to city executives and large building portfolio owners to support investment decisions, policy development and evaluations. For facility managers the tool also supports building operation optimization. The tool is developed on top of the DOE's Standard Energy Efficiency Data (SEED) platform, an open source software application that manages energy performance data of large groups of buildings. Integrated with the existing PostgreSQL relational database, a parallel time-series database retrieves, stores and analyzes the 15-minute interval Green Button data.

The project improves on the current utility data input, focusing on "real-time"[1] data collection and data quality control. The fully integrated data platform has APIs to enable third parties to access the data for apps development. The analytics from the apps will enable city executives/building owners to evaluate the effectiveness of their energy benchmarking program or climate action plan by tracking the before and after change in energy usage or greenhouse gas emissions. Another feature that identifies operational waste will help building operators target inefficient buildings to realize instant savings.

Introduction

The American Recovery and Reinvestment Act (ARRA) stimulus funding of 2009 for smart grid projects resulted in more than 60 million smart meters being deployed in the U.S (DOE 2014). In 2012, the White House introduced the Green Button initiative to give utility customers secure access to their real-time energy usage and consumption (White House 2015). The availability of granulated energy consumption data provides potential opportunities and real challenges. The sheer volume of time-series utility data from a large number of buildings creates challenges in data collection, storage, quality control, and database management. With the spread of the Internet of Things (IoT) and better network capabilities, streaming data from various IP addressable devices and sensors within buildings with various communication protocols and systems provide further challenge for data collection and management. In addition, the

[1] In this instance, the 15-minute interval Green Button data is collected daily in a single XML file with 96 values.

management of large datasets from various sources is time consuming, potentially costly and often challenging for rigorous and meaningful analyses for energy conservation and improved building occupant comfort.

The U.S. Department of Energy developed the Standard Energy Efficiency Data (SEED) platform, an open source software application to support "data-driven energy efficiency program design and implementation." The SEED platform is part of a suite of toolkits that the U.S Federal Government is developing to help "standardize, systematize and link data so that building owners, contractors, researchers, financiers, and other experts can aggregate and share information about building energy performance." These public tools, and a growing number of private tools, utilize a common set of data definitions, called the Building Energy Data Exchange Specification (BEDES) (DOE 2016). DOE intends for SEED to remain a fully interoperable piece of this system (Alschuler et al. 2014).

Data Collection and Storage

Existing SEED Platform

The SEED platform is an open source web-based application developed to manage energy performance data of large groups of buildings. The software application provides an easy, flexible, and cost-effective method to improve the quality and availability of data, originating from multiple sources, through a single portal. The SEED application is written in Python/Django, with AngularJS, Bootstrap, and other javascript libraries used for the front-end. The back-end database is PostgreSQL, a powerful, open source object-relational database with more than 15 years of active development and a strong reputation for reliability, data integrity, and correctness (PostgreSQL 2016). The SEED application has a browser-based interface for users to upload and manage their building data, and a set of APIs for developers to create apps and data visualization tools to meet any organizations' needs. In addition to providing a flexible and cost-effective platform for energy data management, this software application provides a foundation that can help demonstrate the economic and environmental benefits of energy efficiency, support implementation of energy conservation programs, and target investment activity (GitHub 2015).

The main features within the current platform include data upload from tax assessor data and Energy Star Portfolio Manager (PM) data. Users import tax assessor data into the platform, map the terms into BEDES format and define relationships between buildings and tax lots. To reconcile different data sources (e.g tax assessor and PM), SEED simplifies and automates the data matching process by displaying likely matches to the user. With user approvals, SEED stores this information for future data matching. SEED also supports the manual upload of XML, csv and Excel-based utility data files (Alschuler et al. 2014). Figure 1 below illustrates the SEED user interface.

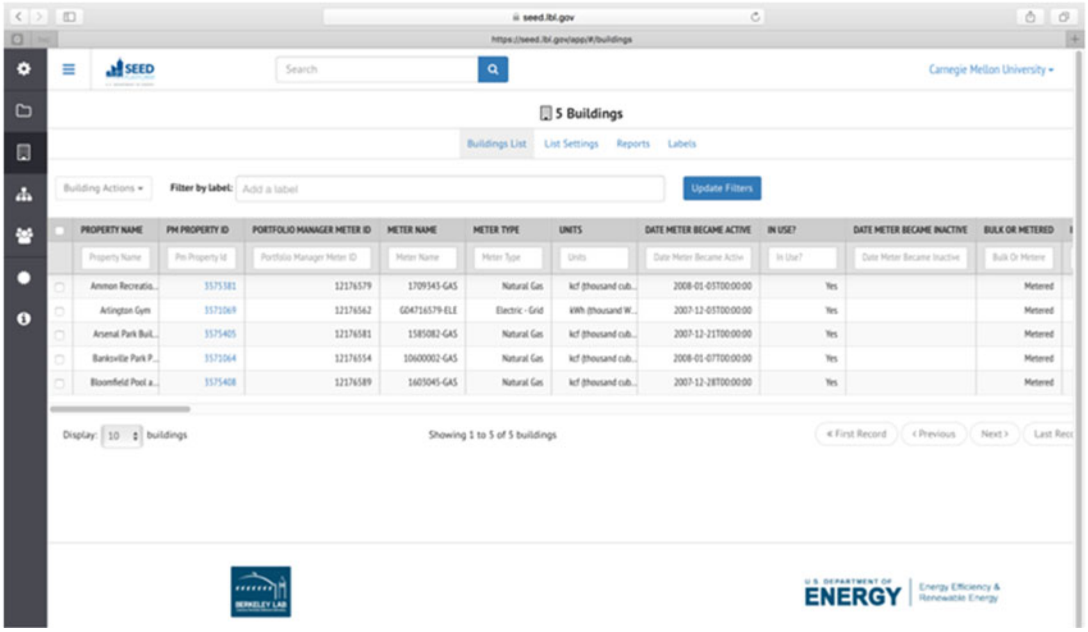
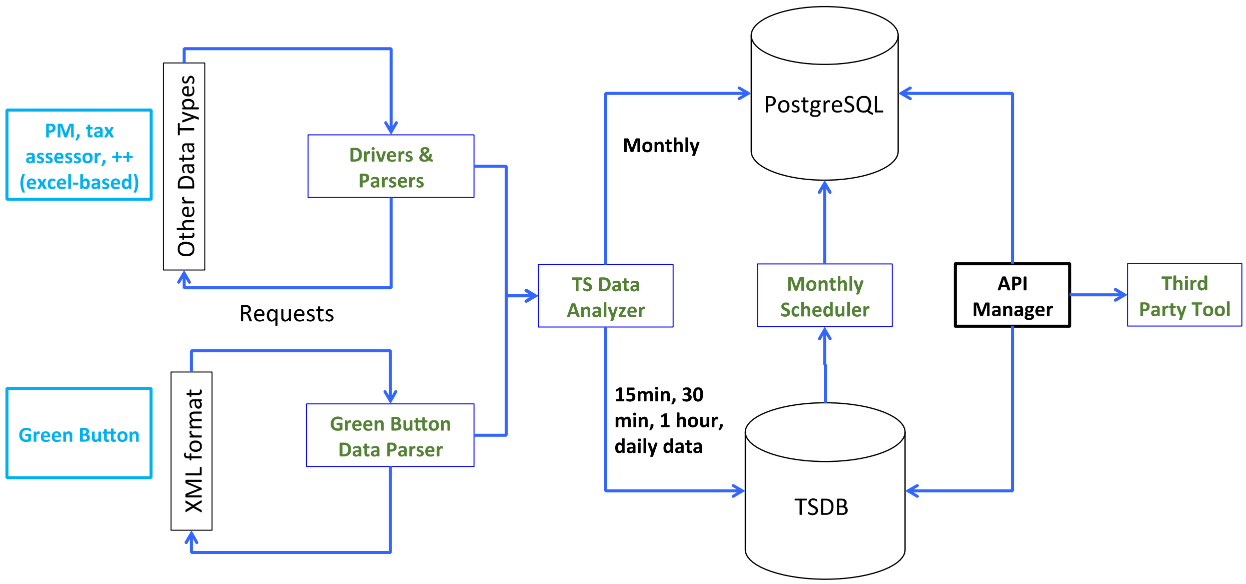


Figure 1.1 SEED platform user interface currently imports building

Enhanced SEED Platform

To increase the functionality of the SEED platform, our team from Carnegie Mellon University is developing and refining a utility data upload, storage and retrieval feature. This feature will enable the manual utility data upload from Excel-based files and support automatic data upload of XML Green Button files from utility providers. We investigated various strategies for the integration of the utility data module into the current SEED platform. In general, the structure of the framework is as follows; a software driver periodically requests interval usage data from utility company (service provider) or the user uploads the excel-based file manually. Once the XML or excel-based file is uploaded, it will go through a parser that formats the data to a SEED time-series data. The data will then be stored in databases to be accessed by third-party apps. These apps can be developed third party vendors or owners of the data itself. This framework is designed for scalability and to minimize modifications to the existing platform. Figure 2 illustrates the overall framework with the various modules to integrate an additional database to manage the time-series utility data. The five modules and associated tasks are listed below:

1. “Data Driver and Parser” modules import the XML and Excel-based files into the SEED.
2. “Time Series (TS) Data Analyzer” module determines where and how to store the imported data.
3. A secondary database (TSDB) manages time series interval data of higher granularity (15 minute to daily intervals).
4. “Monthly Scheduler” module aggregates and pushes monthly data to the existing PostgreSQL database.
5. API Management Platform enables queries from third-party app developers.



The following sections describe in detail the various modules in the database.

Green Button Data Parser

Green Button data is obtained via one of two methods, Green Button Connect My Data (CMD), and Green Button Download My Data (DMD). The data from either source is in the standard Green Button XML format. However Green Button CMD can be accessed automatically through RESTful web services. Due to the standardized Green Button format, the Green Button data parser is applicable to both Green Button CMD and Green Button DMD. The new Green Button data parser (Figure 3) incorporates the following new features to the SEED user interface:

- An interface for users to specify a Green Button web service (URL) to retrieve Green Button data. The URL will be specified and provided by the utility (service provider).
- A “start” and “stop” function to allow user control of the Data Import Scheduler.
- A backfilling function to retrieve all available legacy data.
- Display of the data retrieved from the Green Button server. The displayed data originates from the TSDB through the REST service with a read query.

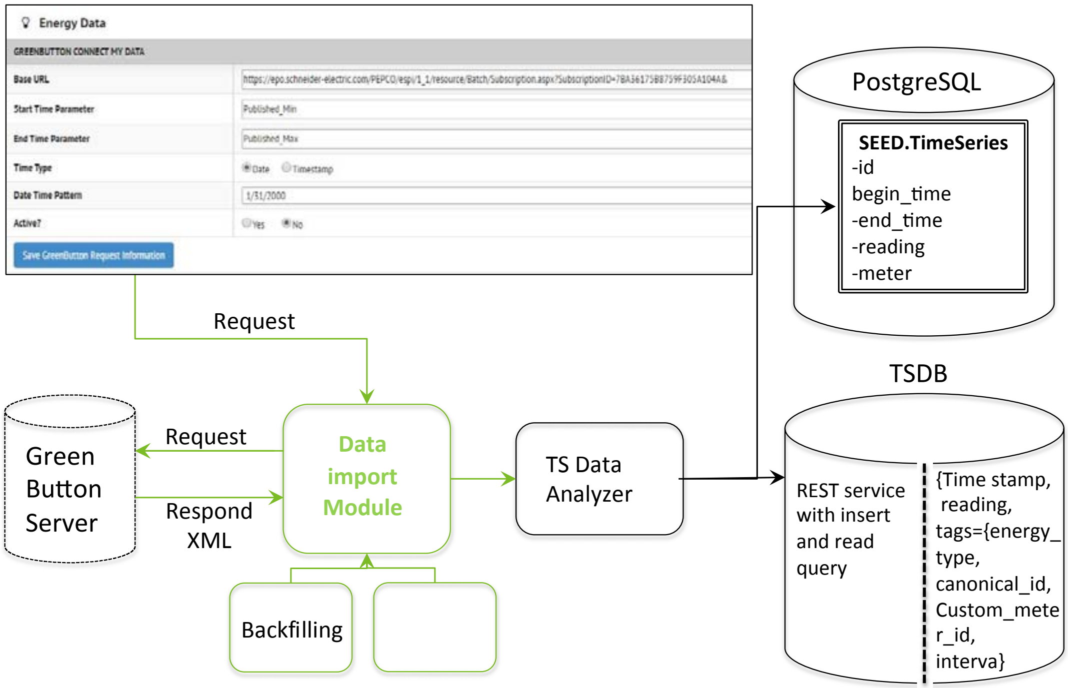


Figure 3. Green Button data parser diagram

Time Series (TS) Data Analyzer

The TS Data Analyzer detects the data time interval and utility type (electricity, gas, water, etc.), provides the meter_ID and other pertinent information (unit, utility type, etc.) from the PostgreSQL database, and inserts the utility data into the time-series database (TSDB) or PostgreSQL based on their time interval. The TS Data Analyzer module will directly push utility data from buildings with monthly interval into the existing PostgreSQL. For other time intervals (15min, 30min, hourly and daily), the TS Data Analyzer will direct the data to the TSDB (Figure 4).

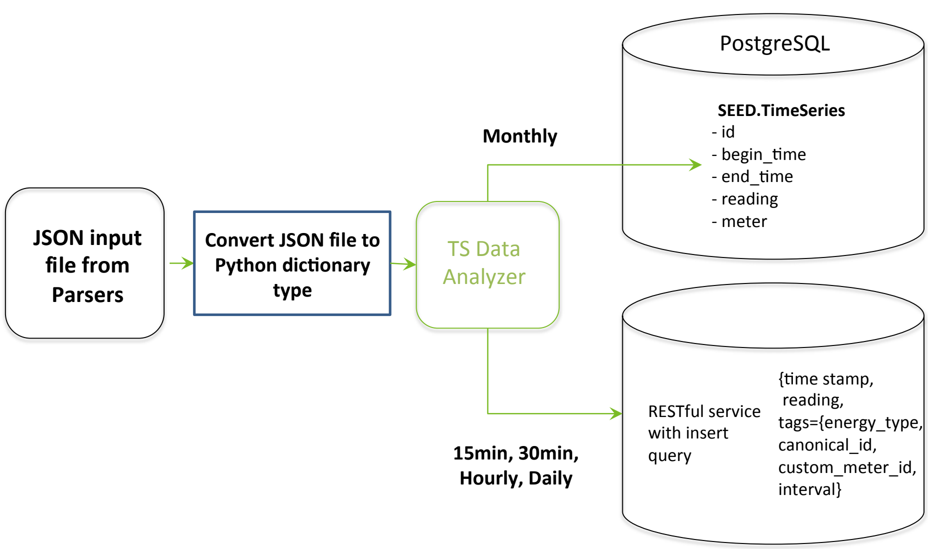


Figure 4. Schema for the TS Data Analyzer

Monthly Scheduler

The Monthly Scheduler module retrieves and aggregates data from the TSDB and saves the monthly aggregated data into the existing PostgreSQL. The Scheduler runs on the first day of every month, queries data from the TSDB through RESTful web services (provided by the TSDB), aggregates the interval data into one month, and pushes the aggregated data into the time series table in PostgreSQL. The aggregator is also triggered each time data is imported into the TSDB.

Time Series Database

Due to limitations of the PostgreSQL relational database for time-series data, we are implementing the use of a secondary time-series database, parallel to the existing PostgreSQL. The TSDB is a specialized database optimized for time-series data storage and processing. The native support is highly optimized for time range query and aggregation calculation. The TSDB is the centerpiece of the whole system, potentially hosting billions of data points per day. For this reason, performance, scalability, aggregation functions and community support are crucial aspects when choosing a TSDB. We opted for KairosDB, an open source time-series database the meets the above-mentioned requirements. In addition, KairosDB supports RESTful API, which makes it compatible with our system design and the existing SEED platform architecture. Due to the modularity of the systems design, the SEED platform should also be able to accommodate other time series databases.

Application Programming Interface (API)

In addition to a set of APIs for internal use to enable the multiple modules to interact, the enhanced platform will have an API program manager to enable interaction with third-party systems and applications outside of the SEED platform. We are currently considering an API platform with our collaborators at Lawrence Berkeley National Lab (LBNL) as a hub to connect with multiple third-party applications, services and systems.

Data-driven Apps for Investment Decisions and Operation Optimization

The following section describes the development and types of user interfaces to support actionable intelligence towards better investment decisions, policy evaluations and optimum building operations and maintenance. The user interfaces were developed based on findings from multiple usability studies with participants from stakeholders of large portfolio owners ranging from city officials (City of Pittsburgh and Washington DC), private sector building portfolio executives and the facility management executives from a university campus. The first usability study solicited the types of information needed and the second round of study was focused on getting feedback on the refined interfaces.

During the focus group meetings, the participants were presented with thirteen different types of data graphics containing various information on utility usage and cost at the portfolio level all the way to building level information (EUI, monthly/yearly usage and spending, ranking of buildings based on usage and cost, etc.) and general overview information about the portfolio (building area, distribution of buildings of different typology, building geographical data, etc.).

Questions regarding the attractiveness of the graphic, understandability of the data representation and usefulness of the analytics were posed for each of the data graphics. The participants were also given questions to gauge their ability to interpret the data graphics accurately to assess the usefulness of the data representation.

The participants rated detailed building level data analyses (graphic 13: "Heat Map," graphic 11: "Potential Savings", and graphic 12: "LEAN Analysis") with the highest average scores in all categories of accuracy, easy-to-understand, usefulness, and attractiveness of the graphics. It demonstrates that even policy makers and executives need detailed and specific building level data for actionable decision-making.

The "heat map" graphic (Figure 5) scored the highest in attractiveness. This graphic utilizes interval data (15min, 30min or hourly) to represent energy usage intensities, color coded, by the hour (x-axis) for every day (y-axis). We chose to represent higher intensities in the red color family and lower intensities are represented in green. This data graphic assists the users to understand building operation schedule and identify unnecessary usage, denoted in red rectangles, during unoccupied hours (weekends and nights). For a building portfolio, users can compare buildings that are operated well and ones that are not optimized. Users can use this information to lower energy usage by changing operation schedules, adding more building controls, and developing strategies to engage building occupants to reduce energy consumption.

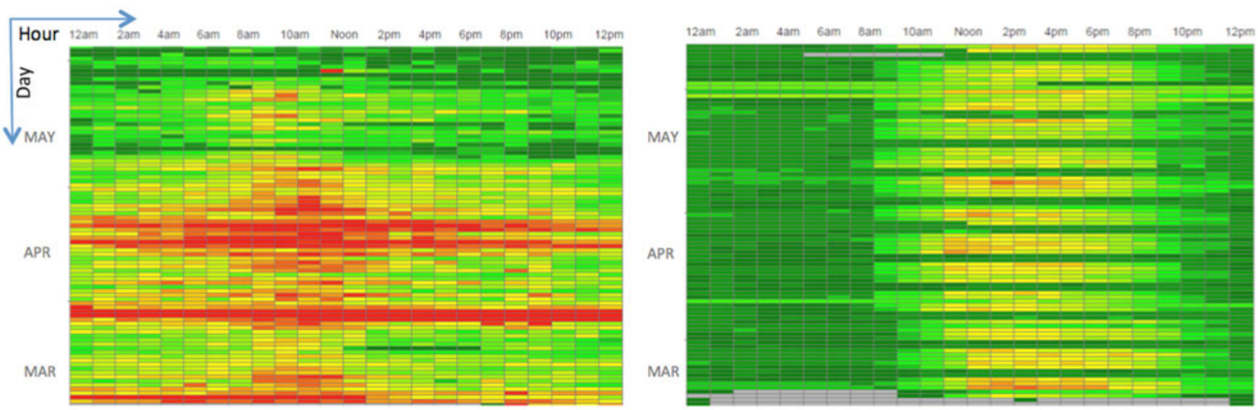


Figure 5. Heat Map data graphic comparing a well-operated building (right) and a building that is using the same amount of energy during unoccupied hours compared to occupied hours (left).

Rated second, the "potential savings" graphic (Figure 6), rank-ordered buildings based on their potential savings. The potential savings are calculated based on the difference of the building's EUI against the median EUI within that category. A building with bigger EUI differential potentially has bigger savings. This information allows building owners to plan their budget allocations for capital improvements/retrofits, building audits and commissioning and leasing terms.

The 25 buildings are selected from the **highest average energy consumption (kBTU) for 2011-2013 per sq. ft.**

	Building Name	Energy Use Intensity (kBTU/sf)	Avg. Yearly Expenditure
1	East Brunswick - 710 Turnpike, NJ	214	$110,583
2	Naperville South Office, IL	194	$112,822
3	Dupont - DC, DC	174	$100,082
4	Cincinnati - 925 Annex, OH	162	$72,155
5	Greensburg South Main Street, PA	142	$41,338

Figure 6. "Potential savings" table

Finally, the "LEAN analysis" graphic (Figure 7) provides information on a building's heating, cooling and base (lighting, plug-load and domestic hot water) loads. The graphic enables the users to identify buildings that consumes significant amount of energy during heating/cooling periods and high base loads year-round. This information allows building owners and facility managers to identify building inefficiencies (poor building enclosure, inefficient equipment) and target retrofits, such as boiler or furnace replacements or building re-commissioning based on the load curves. The "LEAN analysis" graphic is a powerful data analytics tool to enable entire portfolio level analyses to identify the worst performing facilities, a pre-cursor to manual site inspections.

Variables used to conduct the LEAN analysis include monthly electricity and gas consumption data, building size, cooling degree days (CDD) and heating degree days (HDD) (Kissock et al. 2004). Basically, heating load is decided based on the linear relationship between gas consumption and heating degree days, and cooling load is decided based on the linear relationship between electricity consumption and cooling degree days. Lighting and plug-load base load is estimated by the average value of the electricity consumption during the lowest CDD months. Similarly, domestic hot water base load is estimated by the average value of the gas consumption during the lowest HDD months (Donelley et al. 2013). This method needs to be varied to accommodate for the specific energy sources used for heating or cooling a building (e.g. an all electric building uses electricity to cool *and* heat the spaces).

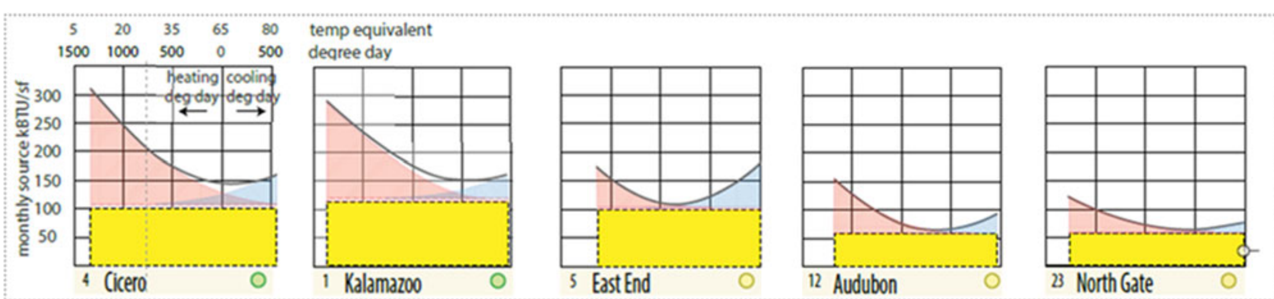


Figure 7. "LEAN analysis" data graphic

Based on the findings from the usability study, we developed new graphical interfaces and refined the top three data graphics for prototype development. The interfaces will help inform building owners, executives and policy makers in decision making for investments and operations using a visually oriented and data driven narrative. Three major sections have been identified from the user testing to create this narrative: utility consumption, facility efficiency and energy saving solutions. The “consumption” section presents the utility cost, usage and greenhouse gas emissions to highlight the total impact of energy consumption on the portfolio. The “efficiency” section identifies specific buildings and the potential savings by evaluating the energy usage of each building compared to national medians for the same building typology. Lastly, the “solutions” section recommends potential retrofits or operational adjustments for the target buildings. A portfolio “overview” section presents the overall information of the building portfolio (Figure 8). In this section, the user will get information on the number of buildings for each building type and the total gross floor area for that building type. This general overview will set the context of the entire portfolio to enable the users to make decisions on future investments. Different stakeholders will find different interfaces/pages within the app to be more applicable for their task. For example, the interface illustrating the GHG emission will be useful for policy makers/building owners to evaluate the environmental impact of their portfolio’s energy consumption. Facility managers will benefit from the “heat map” interface which identifies wasteful energy usage outside of regular working hours.

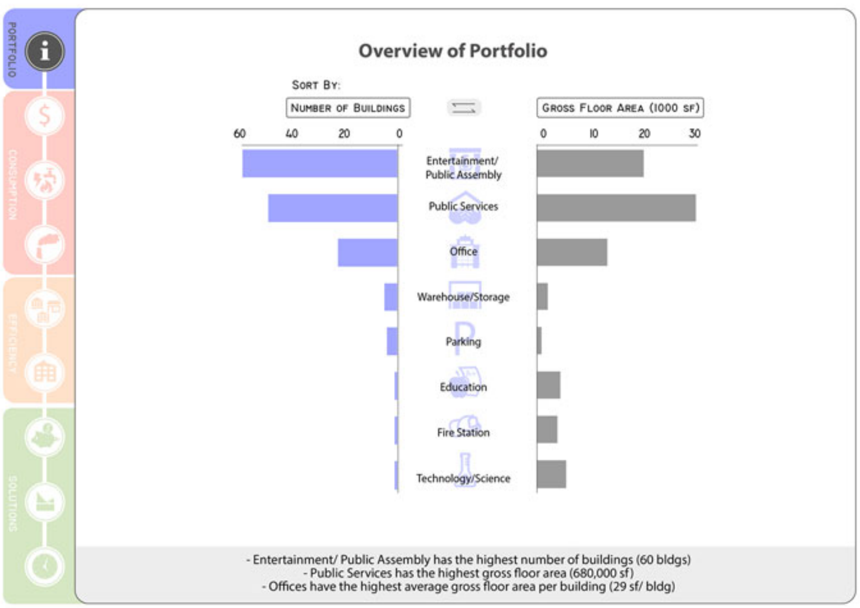


Figure 8. This data graphic provides an overview of the building portfolio’s number of building and total gross floor area for each building type.

Consumption

As a general overview, the graphs in this section describe the total utility cost, usage and greenhouse gas emissions of the entire portfolio over a period of time (Figure 9). The user can sort the data by utility type or building type to help identify the highest consumers. Consumption trends may also show the results of retrofits and operational changes or areas of improvement for future consideration. With this knowledge, the user can evaluate the effectiveness of their decisions on the overall portfolio.

Figure 9. This consumption data graphic provides an overview of the building portfolio utility cost (left), usage, and CO_2 emissions (right).

Efficiency

The efficiency section helps executives/building owners select buildings to retrofit or operate in a more efficient manner. By comparing the building's energy consumption per square foot to the U.S. national median of the equivalent building type (Figure 10, left), the executive/owner can identify the building types and specific buildings to improve. A second graph shows the comparison in terms of savings (Figure 10, right); if the building performed at the national median, the building owner could save a certain amount of money.

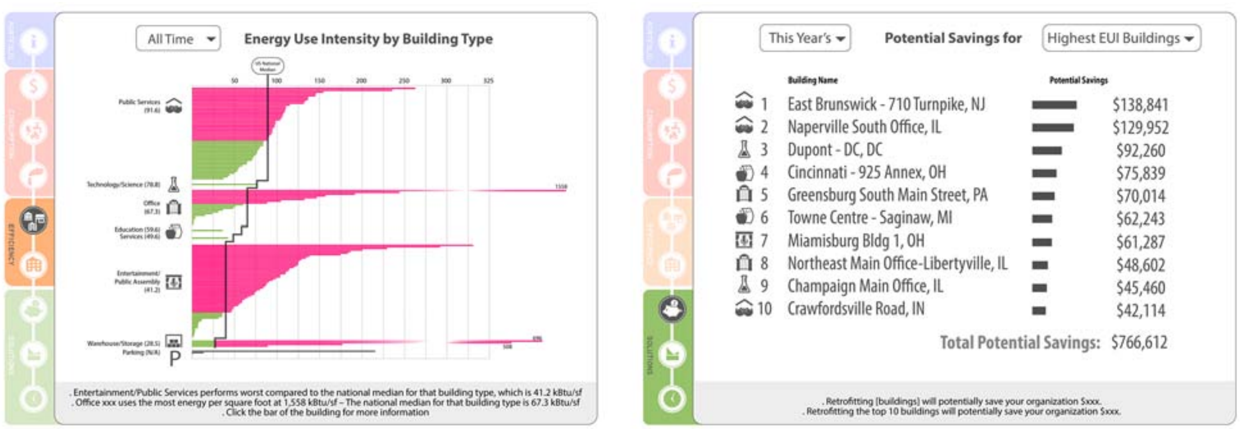


Figure 10. This efficiency data graphic identifies specific buildings and the potential savings by evaluating the energy usage of each building compared to national medians for the same building typology

Solutions

The solutions section helps the executive make decisions about the type of retrofits to invest in as well as identify waste in building operations. A heat map shows the usage per hour of a certain building (Figure 11, left). High consumption is shown in red and low consumption is shown in green. The owner can immediately recognize and eliminate the extra utility usage during unoccupied hours. Using LEAN regression analysis, the buildings' plug loads, heating loads, cooling loads and hot water loads can be expressed. The executive can use this data to identify the type of retrofits that are needed (Figure 11, right).

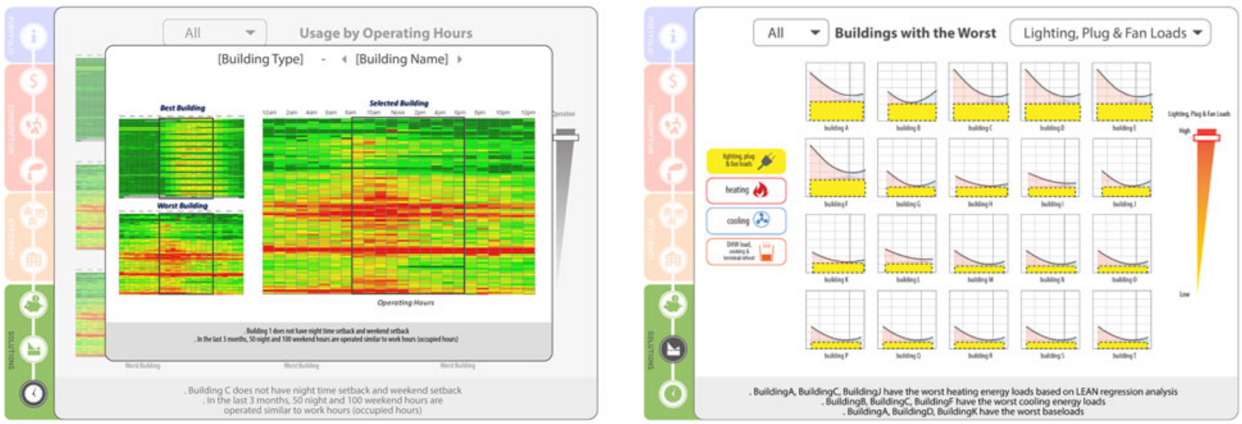


Figure 11. This “heat map” on the left provides an hourly overview of the building energy usage and the LEAN regression data graphic on the right isolates heating, cooling and baseload energy consumptions

Conclusion

The SEED platform’s new features of importing, storing, and managing data from Excel-based and Green Button utility data will enable city executives/policy makers and building owners to make data-driven investment decisions, policy development and evaluation and help building managers optimize facility operations. Building owners can potentially share and compare their portfolio’s building performance indicators against other building portfolios, facilitating sharing and collaboration. Through data import automation and data quality checks, the enhanced platform will streamline energy benchmarking and disclosure law initiatives. The next logical step in the platform’s enhancement is to enable automated data import from gas and water utilities to provide a holistic picture of resource consumption and efficiency for the entire building portfolio. Other enhancements could include the integration of building automation systems and indoor/outdoor environmental quality data. These datasets will be necessary to correlate the impact of energy consumption on indoor environmental quality and occupant comfort and satisfaction (Aziz et al. 2015). This expanded platform will also support the engagement of other building stakeholders such as building occupants, in addition to building owners, facility managers and the general public towards energy conservation (Yun et al. 2013).

Acknowledgement

The work is funded by the U.S. Department of Energy through the Consortium for Building Energy Innovation (CBEI).

References

Alschuler E., Antonoff J., Brown R. and Cheifets M. 2014. "Planting SEEDs: Implementation of a Common Platform for Building Performance Disclosure Program Data Management." In *Proceedings of the ACEEE 2014 Summer Study on Energy Efficiency in Buildings.* 25-35. Pacific Grove, CA: ACEEE.

Aziz A., Lasternas B., Loftness V., Mo Y. and Zhang C. 2015. "Building Data Analytics: Energy and Indoor Environmental Quality." In *Proceedings of the Engineering Sustainability 2015.* Pittsburgh, PA: University of Pittsburgh. 2015.

DOE (Department of Energy). 2014. *Smart Grid System Report: Report to Congress.* Washington, D.C.: U.S. Department of Energy.

Donnelly M., Kummer J. and Drees K. "LEAN Energy Analysis: Using Rgression Analysis to Assess Building Energy Performance." Washington, DC: Johnson Control, 2013.

———. 2016. *Building Energy Data Exchange Specification (BEDES).* Washington, D.C.: U.S. Department of Energy. http://energy.gov/eere/buildings/building-energy-data-exchange-specification-bedes.

GitHub. 2015. *Standard Energy Efficiency Data (SEED) Platform*. 2016. Berkeley, CA: Lawrence Berkeley National Laboratory. https://github.com/SEED-platform/seed

Kissock K. and Seryak J. "LEAN Energy Analysis: Identifying, Discovering and Tracking Energy Savings Potential." *Society of Manufacturing Engineers: Advanced Energy and Fuel Cell Technologies Conference.* Linovia, MI, 2004. 11-13.

PostgreSQL. 2016. *PostgreSQL*. 2016. Global Development Group. http://www.postgresql.org/

White House. 2015. *Green Button Initiative Makes Headway with Electric Industry and Consumers*. July 22. Washington, D.C.: White House. https://www.whitehouse.gov/blog/2015/07/22/green-button-initiative-makes-headway-electric-industry-and-consumers.

Yun R., Scupelli P., Aziz A., and Loftness V. 2013. "Sustainability in the Workplace: Nine Intervention Techniques for Behavior Change." In Proceedings of the *8th international conference on Persuasive Technology*. 253-265. Sydney: Springer-Verlag Berlin Heidelberg.