****

**HCI 598  
Ashley Otero  
M3: The Design**  
**Department of Energy Regional Test Centers Data Project**

# SUMMARY

The Department of Energy (DOE) has been federally funding Sandia National Laboratories to provide system field performance and degradation monitoring data via the Regional Test Center (RTC) data project. The RTCs are field sites that test PV products in a variety of climates and the for photovoltaic performance and reliability. The DOE mandates all data collected by the RTCs to be publically available for consumption by industry, government officials, other research institutions, and the general public (DOE, 2014). As governmental systems have traditionally been locked down behind firewalls and in classified environments, the ability for the public at large to access this data, even if it is not generally useful in everyday life, is a step forward in overall transparency of the final purpose and use of taxpayer dollars (Okamoto, 2017).

The data itself requires a certain amount of baseline knowledge to understand the meaning and function. This introductory knowledge is assumed, as a primer on photovoltaic systems is outside the scope for the system. Meaningful application and visualization of the data allows for greater dissemination of the photovoltaic performance and reliability, which in turn will: 1) potentially engage with external industry partners to allow for a more self-sustaining business model for the RTCs, and 2) validate the necessity and value in the research taking place at the RTCs for the DOE.

There are three main tasks that all users who access the RTC data system will want to accomplish:

* 1. Export data based on a specific set of user-defined criteria.
  2. Find site specific information such as technical specifications and location.
  3. Manipulate a visualization of data based on a specific set of user-defined criteria.

# REQUIREMENTS

The types of users for this system are separated into two groups: unauthenticated general users (which may consist of researchers, industry partners, and the general public), and authenticated industry partner users. All unauthenticated users will see the same interface and have access to the same public data. Authenticated users will be able to see the data that is specifically assigned to their company and marked as proprietary, as well as all publically available data. Proprietary data is flagged within the database and assigned by industry partner or company. Authenticated users are assigned to the specific industry partner or company they belong to and cannot change roles or companies without database administrative changes.

## 2.1 UnauthenticatedUserRequirements

An unauthenticated user must be able to…

1. Find current specifications for any of the RTC sites
   1. Module manufacturer
   2. Number of modules per string
   3. Number of strings per inverter
   4. Number of inverters for the array
   5. Manufacturer and model of inverters
2. Find the location and climate details for any of the RTC sites
   1. Location will include latitude and longitude
   2. Climate details will include:
      1. Link to 25 year weather report
      2. Global horizontal irradiance
      3. Temperature
      4. Direct normal irradiance
      5. Average wind speed
      6. Relative humidity
      7. Diffuse horizontal irradiance
      8. Rainfall per day
3. Select one or more data types and export a CSV file of raw data (see tables)
4. Compare two systems within the same time period within the same RTC site
5. Export visualization data as an image or PDF

## 2.2 Authenticated User Requirements

**In addition to the above**, an authenticated user will require a login and password to access proprietary (non-public) data. These users will be limited to their own specific data and assigned individual usernames and passwords.

An authenticated user must be able to…

1. Successfully use a username and password to login as an authorized user
2. See a dashboard of performance and reliability data for their systems
3. Export a CSV of raw data for their systems
4. Export visualization data for their systems as an image or PDF

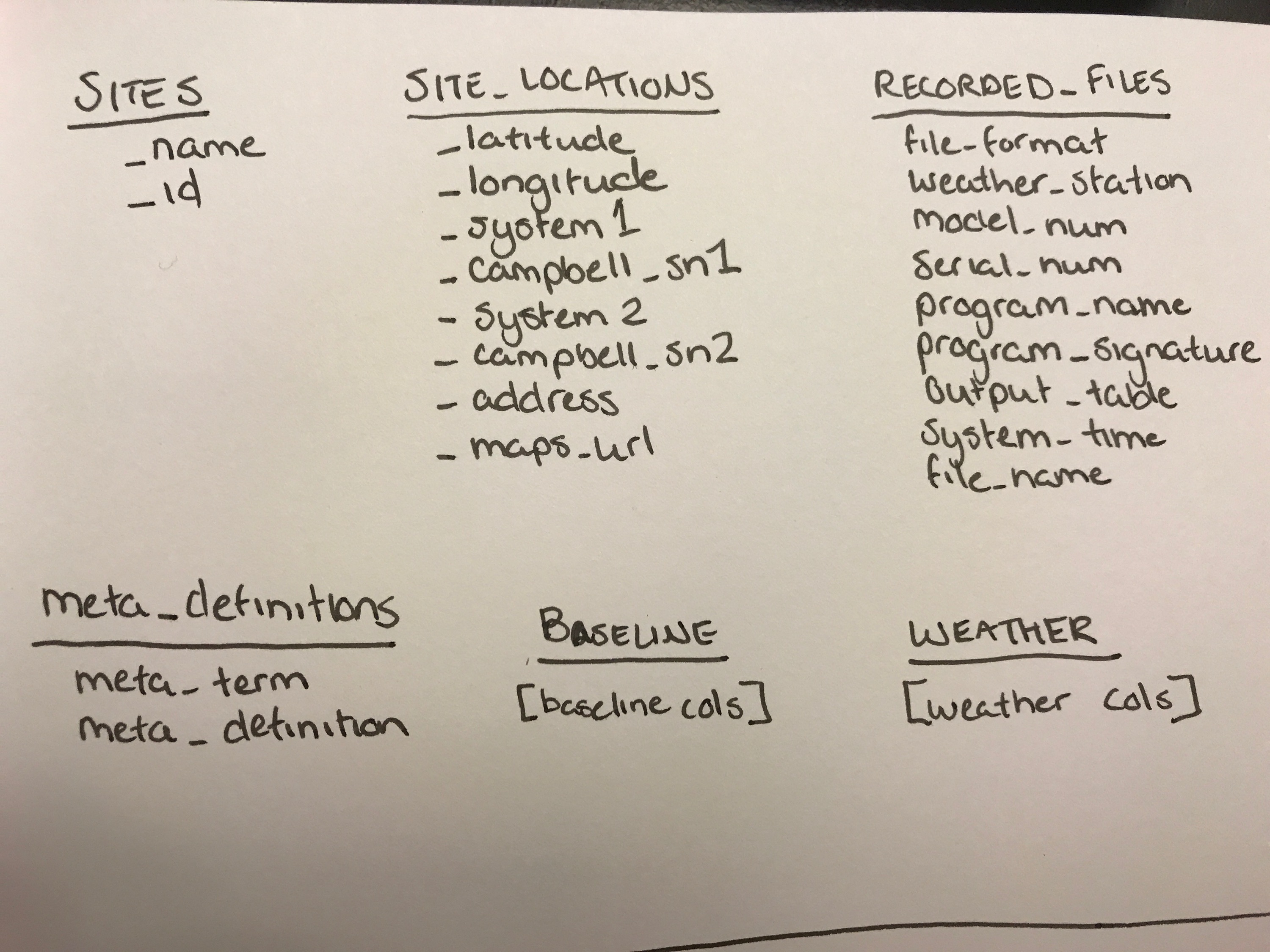
## Table 1: Baseline Data Types

|  |  |
| --- | --- |
| Term | Definition |
| TIMESTAMP | A string of the format yyyy-mm-dd HH:MM:SS. |
| RECORD | An integer indicating the number of entries since the logger was started |
| Sys1Vdc\_Avg | DC voltage of system 1 in volts |
| Sys1Str1Idc\_Avg | DC current of system 1, string 1 in amperes |
| Sys1Str2Idc\_Avg | DC current of system 1, string 2 in amperes |
| Sys1Vac\_Avg | AC RMS voltage of system 1 in volts |
| Sys1Iac\_Avg | AC RMS current of system 1 in amperes |
| Sys1Wac\_Avg | AC active power of system 1 in watts |
| Sys1VARac\_Avg | AC reactive power of system 1 in volt-amps-reactive |
| Sys1PowerFactor\_Avg | AC Power factor of system 1 (unitless) |
| Sys1Frequency\_Avg | AC frequency of system 1 in hertz |
| ModTemp1\_Avg | Module temperature 1 in degrees C |
| ModTemp2\_Avg | Module temperature 2 in degrees C |
| ModTemp3\_Avg | Module temperature 3 in degrees C |
| ModTemp4\_Avg | Module temperature 4 in degrees C |
| ModTemp5\_Avg | Module temperature 5 in degrees C |
| ModTemp6\_Avg | Module temperature 6 in degrees C |
| ModTemp7\_Avg | Module temperature 7 in degrees C |
| ModTemp8\_Avg | Module temperature 8 in degrees C |
| Sys2Vdc\_Avg | DC voltage of system 2 in volts |
| Sys2Str1Idc\_Avg | DC current of system 2, string 1 in amperes |
| Sys2Str2Idc\_Avg | DC current of system 2, string 2 in amperes |
| Sys2Vac\_Avg | AC RMS voltage of system 2 in volts |
| Sys2Iac\_Avg | AC RMS current of system 2 in amperes |
| Sys2Wac\_Avg | AC active power of system 2 in watts |
| Sys2VARac\_Avg | AC reactive power of system 2 in volt-amps-reactive |
| Sys2PowerFactor\_Avg | AC Power factor of system 2 (unitless) |
| Sys2Frequency\_Avg | AC frequency of system 2 in hertz |
| ModTemp9\_Avg | Module temperature 9 in degrees C |
| ModTemp10\_Avg | Module temperature 10 in degrees C |
| ModTemp11\_Avg | Module temperature 11 in degrees C |
| ModTemp12\_Avg | Module temperature 12 in degrees C |
| ModTemp13\_Avg | Module temperature 13 in degrees C |
| ModTemp14\_Avg | Module temperature 14 in degrees C |
| ModTemp15\_Avg | Module temperature 15 in degrees C |
| ModTemp16\_Avg | Module temperature 16 in degrees C |
| RefCell1Irrad\_Avg | Irradiance of plane of array reference cell 1 in watts per square meter |
| RefCell1Temp\_Avg | Temperature of plane of array reference cell 1 in degrees C |
| RefCell1R\_Avg | Shunt reading of plane of array reference cell 1 in volts |
| RefCell2Irrad\_Avg | Irradiance of plane of array reference cell 2 in watts per square meter |
| RefCell2Temp\_Avg | Temperature of plane of array reference cell 2 in degrees C |
| RefCell2R\_Avg | Shunt reading of plane of array reference cell 2 in volts |
| LocalAmbientTemp\_Avg | Local ambient temperature in degrees C |
| POAIrrad1\_Avg | Irradiance from plane of array pyranometer (CMP-11) in watts per square meter |
| ICP7019\_1CJCTemp\_Avg | Temperature of DAQ module number 1 |
| ICP7019\_2CJCTemp\_Avg | Temperature of DAQ module number 2 |
| ICP7019\_3CJCTemp\_Avg | Temperature of DAQ module number 3 |
| ICP7019\_4CJCTemp\_Avg | Temperature of DAQ module number 4 |
| ICP7019\_5CJCTemp\_Avg | Temperature of DAQ module number 5 |
| ICP7019\_6CJCTemp\_Avg | Temperature of DAQ module number 6 |
| ICP7019\_7CJCTemp\_Avg | Temperature of DAQ module number 7 |

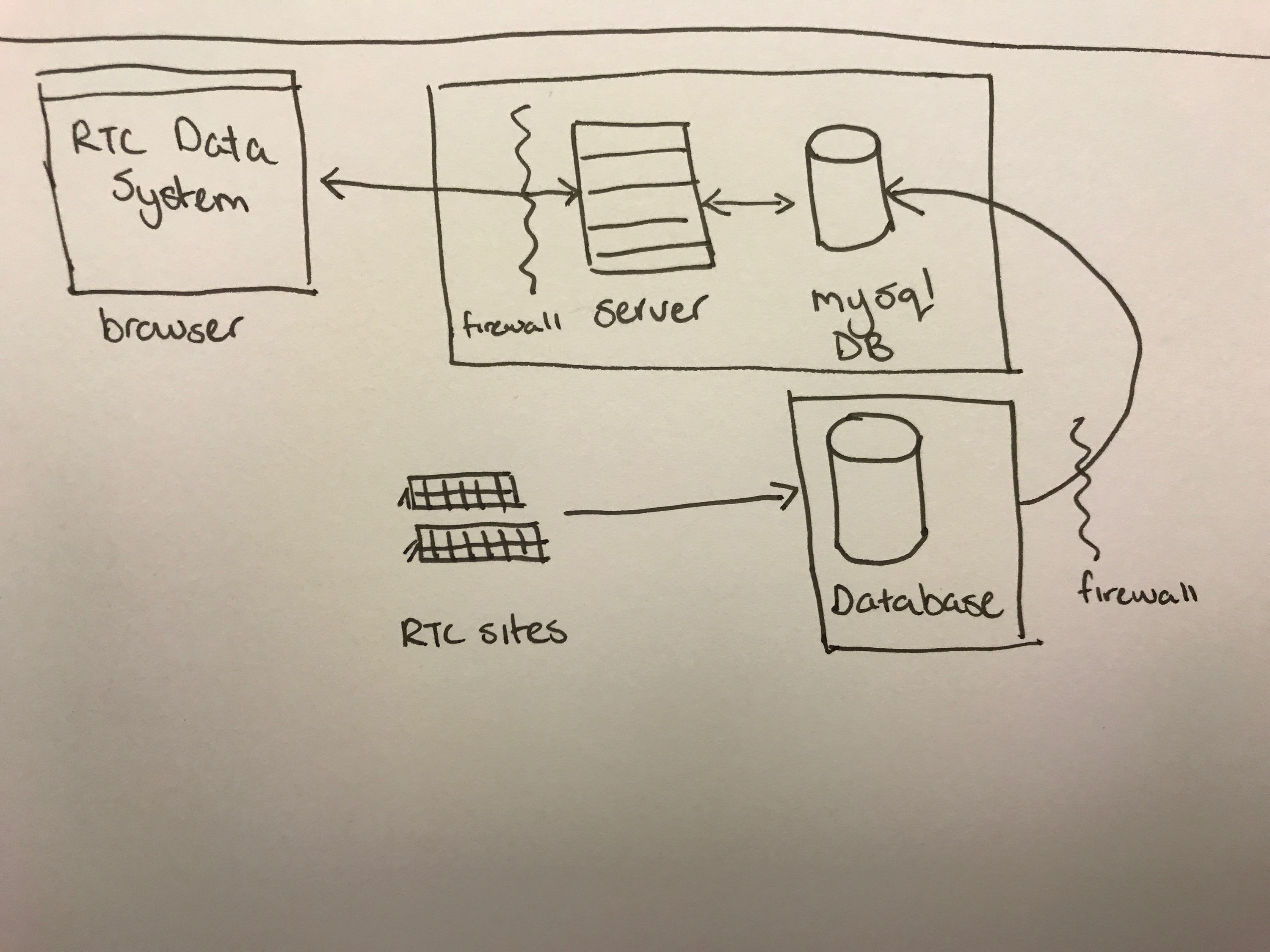
## Table 2: Weather Data Types

|  |  |
| --- | --- |
| Term | Definition |
| TIMESTAMP | A string of the format yyyy-mm-dd HH:MM:SS. |
| RECORD | An integer indicating the number of entries since the logger was started |
| Global\_Wm2\_Avg | Global Horizontal Irradiance in Watts per square meter |
| Direct\_Wm2\_Avg | Direct Normal Irradiance in watts per square meter |
| Diffuse\_Wm2\_Avg | Diffuse Horixontal irradiance in watts per square meter |
| Pressure\_mBar\_Avg | Atmospheric pressure in millibars |
| WS\_ms\_Mean | Mean wind speed in meters per second |
| Wdir\_Mean | Mean wind direction in degrees east of north |
| Wdir\_Std | Wind direction standard deviation in degrees |
| WS\_ms\_Std | Wind speed standard deviation in meters per second |
| WS\_ms\_3sec\_Max | Maximum observed wind speed sustained over 3 seconds, in meters per second |
| WD\_deg\_SMM | Wind direction at which the maximum wind speed was observed |
| Temp\_C\_Avg | Ambient dry-bulb air temperature in degrees C |
| RH\_pct\_Avg | Relative humidity in percent |
| Panel\_Temp\_C\_Avg | Temperature of the data logger in degrees C |
| Battery\_V\_Avg | Battery voltage of the data logger in volts |
| ProcessTime\_S\_Max | Maximum processing time in microseconds |
| Global\_mV\_Avg | Millivolt reading of the global horizontal instrument |
| Rain\_mm\_Tot | Millimeters of rainfall since the logger was started |
| Rain\_mm\_Daily | Millimeters of rainfall since midnight |
| Temp\_CMP22\_C\_Avg | Temperature of the global horizontal irradiance pyranometer |

## 2.3 Proposed System Database Schema



## 2.4 Proposed Infrastructure



# 3. DESIGN SPACE

## 3.1 Design Tradeoffs

### 3.1.1 Value for the Multiple Audiences

Users belonging to the researcher user type were hesitant to consider the use one system to adequately serve multiple audiences, specifically raising concerns that the general public would be unable to understand the use and value of the system without significant instruction or context. This led to pushback on a simplified version of the design, as the definition for data types would be required for non-researcher and non-industry user types. This also led to discussion about adding a general information or introduction to solar and PV reliability page separate from the public data query builder. Additional research and user interviews will be required to determine if this addition will be valuable or clutter the interface.

### Limitations on Comparison Data

Most of the comparisons of systems are not apples to apples type comparisons. For instance, the baseline and weather systems collect completely different datasets and would not be able to be compared side by side effectively. For sites with multiple identical arrays setups by different module manufacturers, it is easier to normalize the variances and accurately compare the visualization data produced by each site. Additionally, the limitations of having every database column as a potential filter for selection would make the interface not only difficult to interpret, but increase the cognitive load even for advanced users, and could also potentially create a strain on the computing resources of the end user’s browser client.

### Responsive Design

Typically, most web applications are designed and developed with the assumption that it must be accessible on a variety of devices, including mobile phones and tablets. Due to the sheer amount of data types present to build the queries for the public data portion of the system, it is not feasible to optimize the design for devices smaller than a landscape tablet. Users interviewed were not concerned with the lack of mobile support, as they expected to mostly interface with the tool in a desktop or laptop setting. Additionally, since files (XLS, PDF, JPG, etc.) are expected to be downloaded, these users felt it would be uncomfortable to store such information on their phone or tablet devices.

## 3.2 Difficult Requirements to Support

### 3.2.1 Securing Proprietary Data

Securing proprietary partner data is outside the scope of this project, but was repeatedly raised as a potential issue for implementation.

### 3.2.2 Risk Mitigation of Infrastructure

The inherent risks of using third party infrastructure services, such as a hosting and domain provider quickly proved to be more effort than was valuable to meet the requirements for this system. An on-premise virtual machine infrastructure consisting of a Linux, Apache, MySQL, and PHP (LAMP) technology stack that resides within the Sandia Open Network and behind the Sandia network firewall will be used (Riley, 2016). This reduces the overhead of additional resources to continually scan for vulnerabilities and apply patches to an outside infrastructure, as these services are included as part of any software system application developed on Sandia servers. Since the existing infrastructure will meet or exceed the documented requirements, there is no need to pursue outside resources.

### Two Factor Authentication

Securing authentication was another customer requirement, and the suggestion of using two-factor authentication to reduce the number of unauthorized attempts to retrieve data was supplied as a potential mitigation tactic. Implementation of this authentication security is outside of the scope of this project.

# 4. THE DESIGN

The design for the RTC Data system is a web-based application. Based on the increased scope of the system from the initial request, the system will only include the public data portions, but will show dummy data for the authenticated dashboard screens.

## 4.1 Initial Display (Home Page)

The initial screen that displays in a browser is a map of the United States with specific markers for each of the five Regional Test Center sites: Albuquerque, New Mexico; Golden, Colorado; Cocoa, Florida; Henderson, Nevada; and Williston, Vermont. Additionally, the main Regional Test Center logo (as approved and also required by the Department of Energy) will be present in the upper left hand corner of the screens to maintain consistency with the main RTC website, as well as conform to design patterns. A primary navigation (which serves as secondary on the initial screen, due to the map) will be present in horizontal, right-aligned navigation bar within the header.

This initial loading screen, while adhering to many standard web patterns, may also seem overly simplistic due to the lack of text or other context available. The intent was to focus the user on selecting a site to query and narrowing the potential pathways to additional sites. While simple in design, this may feel empty or incomplete to users who stumble across the system, as opposed to those who have a researcher or industry partner background and understand the value and purpose of the system.

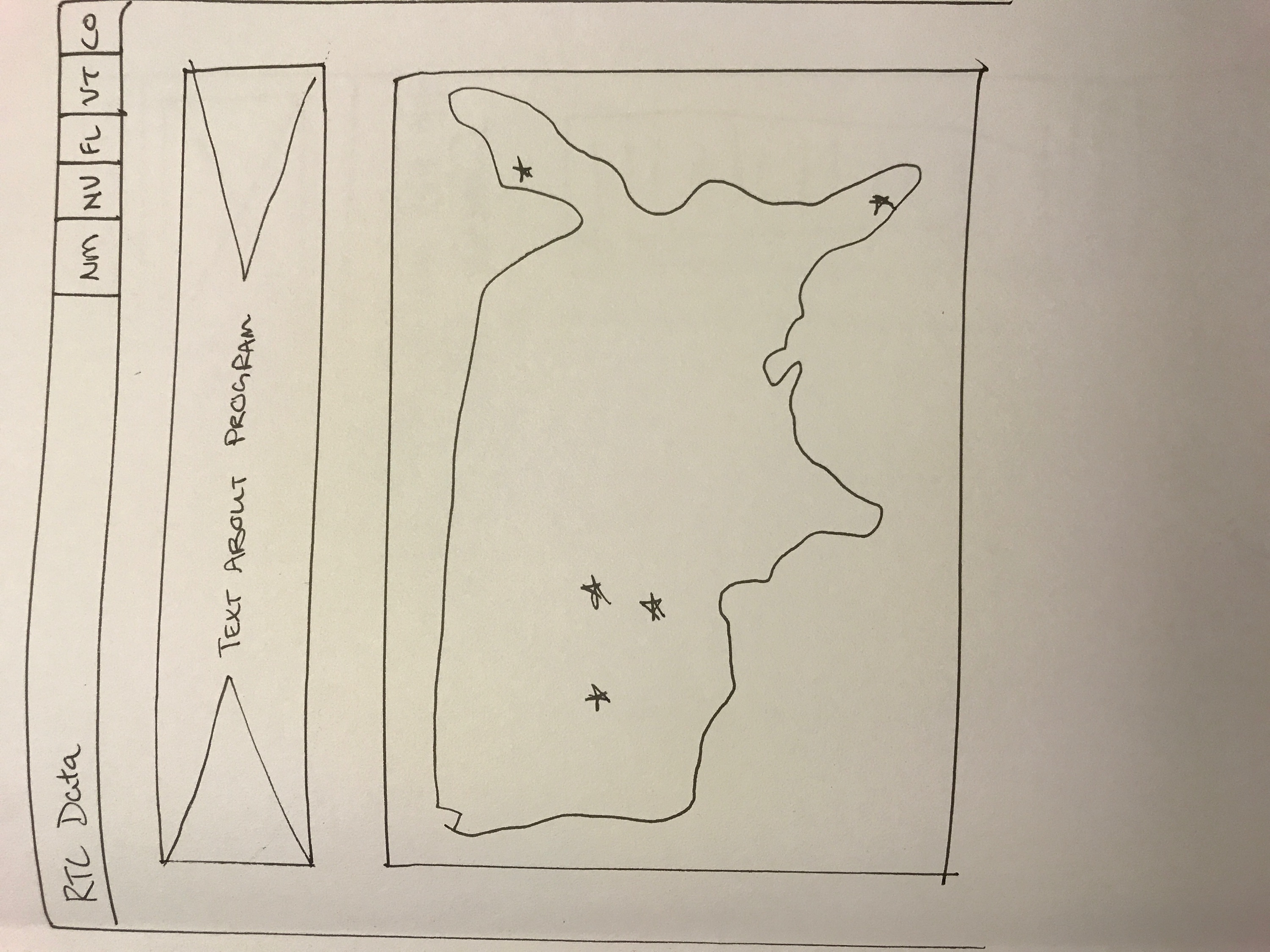


Figure 4.1: RTC Homepage Sketch

## 4.2 RTC Site Page

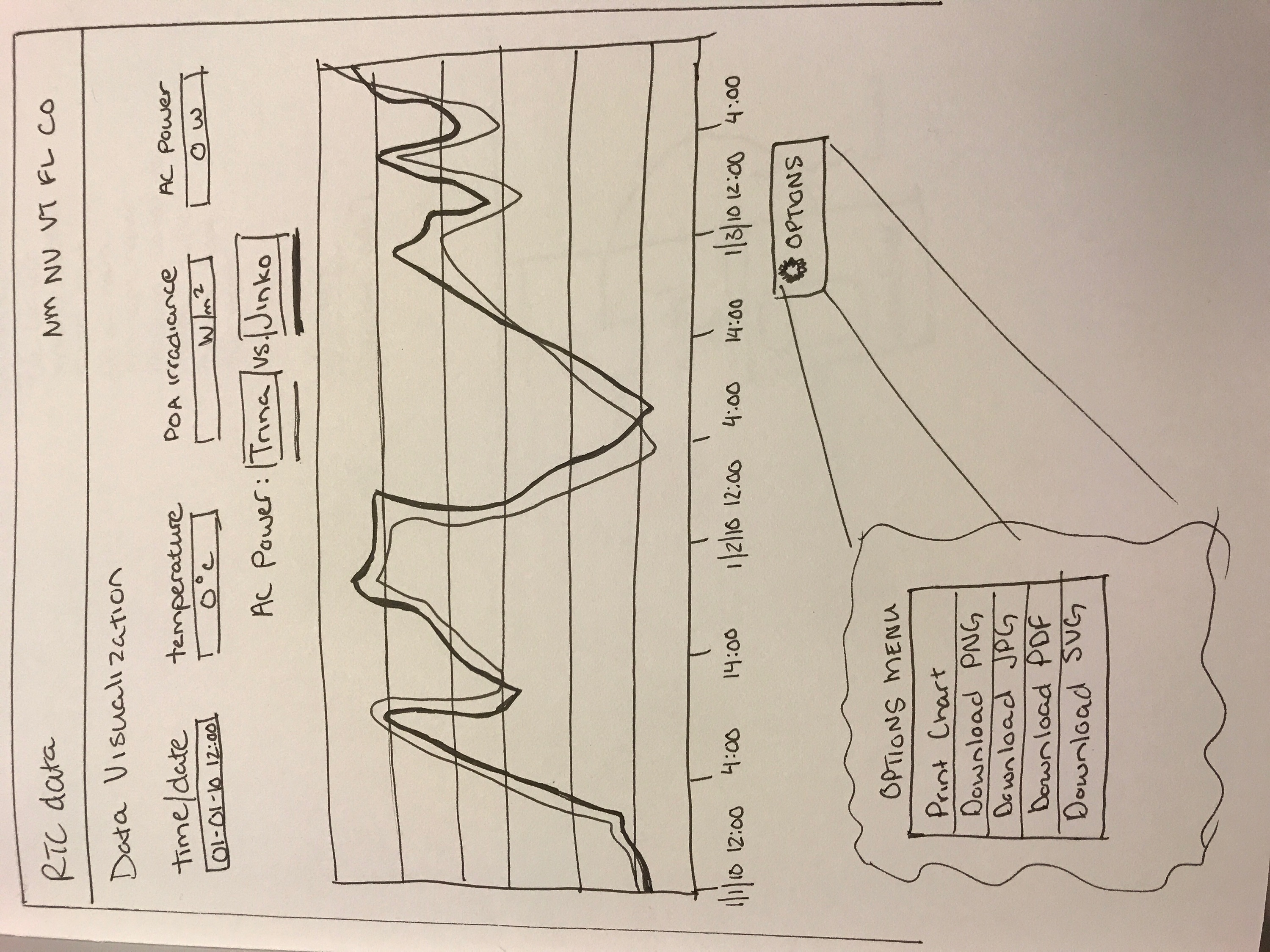
After a user selects which site they wish to query, a detailed screen with the location, technical specifications, photos, and climate data for the site will load. This interface will also contain all the available systems and data types for querying, in addition to the definitions of the data types for greater user context. At this point, the user can select a system, date range, and one or more data types to build a query. Once they have selected their desired inputs, the user will click “Get Raw Data.” Upon successful completion of the query, a CSV file will be downloaded via the user’s web browser. When the user opens the CSV, the file will reflect the data types as column headers, and the rows will correspond to each day within the specified date range.

This page has a significant cognitive load for users, just due to the sheer amount of information displayed. In future user testing, an alternate layout that separates the technical specifications and overall site data from the query builder portion of the page will be tested. In addition, the required linear process of the query building process will be highlighted to keep users on task and prevent errors.

   
Figure 4.2: Site Specific Sketch

## 4.3 RTC Site Visualization Page

After a query of raw data has been run and completed, a user will be redirected to a page with the visualization of available data within their requested date range. This page will allow users to interact with the visualization of data, including adjusting the date range/time, zooming in or out to increase or decrease the focus on a particular set of data, and filter by toggling the specific modules that can be compared side by side. In addition, users can download a snapshot of the visualization in various different formats, including PNG, JPG, PDF, or SVG.

   
Figure 4.3: Visualization Sketch, including Options menu choices

# FUTURE TECHNOLOGIES & SOCIAL IMPLICATIONS

## 5.1 Emerging Technologies

Data visualization will continue to be an important factor in how large datasets are disseminated to various audiences. In addition, the visualization of this data will allow even non-technical users to begin to make conclusions based on the data (McCauley, 2017), rather than staring at a CSV of raw data. This data will help to inform decision makers at all levels of the sustainability of photovoltaic infrastructure over time at different climates.

Though the current design is optimized, and arguably built for a desktop experience, any future designs will need to be created with the mobile and tablet market in mind. Currently, the visitors of the main Regional Test Centers website are mostly on desktop or laptop devices, using current browsers, such as Chrome, Firefox, and Internet Explorer 11/Edge. If there is a significant shift in analytics that proves the visitors are beginning to use the site on a smaller resolution, such as a tablet or mobile device, it would make a compelling argument to reevaluate the purpose of the system and address the new needs of users.

## 5.2 Ethical Issues and Social Implications

Having open data available to the general public is a step forward in government transparency of the final purpose and use of taxpayer dollars (Okamoto, 2017). The Department of Energy has required this data to be open and available to the public, but future administrations may rescind or abandon this mandate. The ongoing support of the United States government is currently critical for the Regional Test Centers to maintain their existence past fiscal year 2018. If the funding was dramatically reduced, the potential harm for the RTC program would be drastic. A lack of governmental funding could force the RTCs to begin to approach industry for a more direct partnership involving research for payment, rather than the current arrangement of one between industry and research institution as two non-competing but equally important roles.

The inherent risk of storing proprietary data within an externally-facing web server is another consideration that requires ongoing mitigation planning. The potential loss of trust between industry partners and the RTC program, as well as Sandia National Laboratories in general must not be understated. In addition, a breach of this private data may cause financial harm to the industry partners or reveal some sort of trade secrets, directly impacting their bottom line.

# References

1. Riley, D., & Stein, J. (2016). *The Regional Test Center Data Transfer System* (SAND2016-9461) (United States, Department of Energy, Sandia National Laboratories). Albuquerque, New Mexico: Sandia National Laboratories.
2. Okamoto, K. (2017). Introducing Open Government Data. *Reference Librarian*, *58*(2), 111-123. doi:10.1080/02763877.2016.1199005
3. McCauley, R., & Graves, B. (2017, January 30). How Data Visualization is the Future of Information Sharing. http://www.govtech.com/fs/How-Data-Visualization-is-the-Future-of-Information-Sharing.html