****

**HCI 598  
Ashley Otero  
M5: User Testing**

**Department of Energy Regional Test Centers Data Project**

# System Summary

## 1.1 System Overview

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 for photovoltaic performance and reliability.

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.

## 1.1 Users

The types of users the system is targeting are researchers and industry representatives. The design was focused primarily on researchers, as I have the most access to them, in terms of ongoing professional relationships. This direct access will allow for ongoing iterations and continuous feedback after the conclusion of this initial project. These researchers are typically highly scientific and technical people, and most have a Ph.D. in their field. While not designers, developers, or programmers, they are intimately familiar with the workings and meaning of data collected by the photovoltaic systems and have demands on what type of visualizations will be most persuasive and valuable for the RTCs. All have basic computing skills and immeasurable knowledge in the varying data types across the five different RTC sites.

## 1.1 Supported Tasks

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.

# Evaluation Methods

## 2.1 Gathering User Data

To begin my data gathering, I interviewed 6 different participants from within Sandia National Laboratories. Due to time and travel constraints, I was unable to get 100% researchers within the Photovoltaic and Distributed Systems organizational unit. I gathered volunteers via email for a brief usability session. Each participant was interviewed briefly before the usability test took place, and then emailed after the test was complete with two surveys to complete. Due to the small sample size, participation was 100%.

## 2.2 Interview Questions (Pre-Test)

Users were asked demographic and background questions before the usability testing session took place.

1. What is your current role or job title?
2. What is your educational background (BA/BS, MS, Ph.D.)?
3. How long have you been with Sandia?
4. What was your previous role before coming to Sandia?
5. Are you familiar with the RTC program?
6. Are you familiar with the solar industry?
7. Do you consider yourself a technical person?

## User Demographics

The user demographics were generated by the pre-test interview.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User | Demographics | Role | Education | Time w/ Sandia | Previous Role | Familiar w/ RTC? | Familiar w/ Solar? | Technical? |
| A | Male, early 30s | Researcher | MS | 4 years | Tech company | Yes | Yes | Yes |
| B | Male, early 50s | Senior Scientist | Ph.D | 13 years | Other lab | Yes | Yes | Yes |
| C | Female, mid 20s | Communications Specialist | BA | 18 months | Private sector | Yes | Yes | Somewhat |
| D | Female, mid 20s | Project Controller | MBA | 2 years | Private sector | Yes | Yes | Somewhat |
| E | Female, late 50s | Principle Investigator | Ph.D | 5 years | Academia | Yes | Yes | Yes |
| F | Male, late 20s | Technical Staff | MS | 8 years | Student | Yes | Yes | Yes |

## 2.4 User Tasks

Prior to the user testing beginning, a short script was recited:

Hello and thank you for taking the time out of your schedule to participate in the Regional Test Center Data Program usability testing. This testing is to explore the usability of the system built to display RTC Data. There is no pass or fail for this test, as it is recording issues that may not be clear to the developers of the system, but which affect the way that users can interact with the system. If you encounter any issues with the system, please understand that is due to poor system design and is not a reflection on your ability to use the system. You will be presented with three tasks to perform using the system. Feel free to expand on your feelings during the course of the test.   
  
The user was then given three tasks:

1. Find number of modules per string for the RTC site you select.
2. Export data as a CSV from an available system based on whatever criteria you select.
3. Manipulate a visualization of data.

## 2.5 Post-Test Questions

Users were asked for their thoughts on the system itself, in addition to comments recorded during the test itself. These questions were open ended to allow for a greater insight into the user’s mind. One user did not have any feedback for either of the two questions presented.

1. What did you find most confusing about the system?
2. Do you have suggestions for improvement?

|  |  |  |
| --- | --- | --- |
| Users | What did you find most confusing about the system? | Do you have suggestions for improvement? |
| A | No clear instruction on how to export data | Add instructions |
| B | Difficult to get to visualizations from site page | Prevent user from submitting without date range |
| C | Excel showed an error when opening CSV | Limit calendar to available dates |
| D | N/A | Onboarding or walkthrough for first time visitors |
| E | N/A | Training prior to use |
| F | N/A | N/A |

## 2.6 Post-Test Questionnaires

Users were emailed two links to questionnaires to complete anonymously and without the proctor present to limit the potential bias. Those questionnaires are available via the following links: <https://goo.gl/forms/AGNmtyfynTw7M5D42> (Usefulness, Satisfaction, and Ease of Use) and <https://goo.gl/forms/cUevyRmTdGzLBwCe2> (System Usability Scale). The questions for both questionnaires originate from standardized formats.

### Usefulness, Satisfaction, and Ease of Use

This questionnaire was built in Google Forms based on Arnold Lund’s original survey including 30 questions rated on a Likert scale of 1-7, with one being “strongly disagree” and 7 being “strongly agree.”

1. The system helps me be more effective.
2. The system helps me be more productive.
3. The system is useful.
4. The system gives me more control over the activities in my life.
5. The system makes the things I want to accomplish easier to get done.
6. The system saves me time when I use it.
7. The system meets my needs.
8. The system does everything I would expect it to do.
9. The system is easy to use.
10. The system is simple to use.
11. The system is user friendly.
12. The system requires the fewest steps possible to accomplish what I want to do with it.
13. The system is flexible.
14. Using the system is effortless.
15. I can use the system without written instructions.
16. I don't notice any inconsistencies as I use the system.
17. Both occasional and regular users would like the system.
18. I can recover from mistakes quickly and easily in the system.
19. I can use the system successfully every time.
20. I learned to use the system quickly.
21. I easily remember how to use the system.
22. It is easy to learn to use the system.
23. I quickly became skillful with the system.
24. I am satisfied with the system.
25. I would recommend the system to a friend.
26. The system is fun to use.
27. The system works the way I want it to work.
28. The system is wonderful.
29. I feel I need to have the system.
30. The system is pleasant to use.

### System Usability Scale

This questionnaire was built in Google Forms based on John Brooke’s original survey in 1986, containing 10 questions rated on a Likert scale of 1-5, with one being “strongly disagree” and 5 being “strongly agree.”

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in the system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going in the system.

## 2.7 Testing Environment

Each session took place in the same laboratory with the same computer setup and internet/network access. The computer was a 13” MacBook Pro (Retina, Early 2015) with 3.1 GHz Intel Core i7 procesor, 16 GB 1867 MHz DDR3 of RAM, and running macOS Sierra (version 10.12.6). The computer was hard-wired to the Sandia network via Ethernet. Conditions were typical of an office setting, with overhead lighting and ergonomical desk and chair setup.

# Results

## 3.1 Usability Testing Results

The results of the usability testing generated the following errors. For reference, the type codes are included below.

1. Usability disaster: must fix
2. Major problem
3. Minor problem
4. Cosmetic problem and/or related to prototype
5. New feature request

|  |  |  |
| --- | --- | --- |
| Users | Issue | Type |
| ABCDF | Didn't see start/end date | 1 |
| E | Issue loading site from map | 1 |
| ABE | Didn't see top navigation on home page | 3 |
| BCF | Didn't see data available dates | 2 |
| CDF | Was unsure of system definitions purpose | 4 |
| ALL | Didn't realize Build Query would export CSV | 1 |
| A | Asked for system definitions in a tooltip | 5 |
| ALL | Was confused by error thrown by Excel on CSV open | 4 |
| AD | Thought time/temp/etc on visualization page was clickable | 3 |
| B | Found specs redundant on visualization | 3 |
| F | Thought colors between two systems were too hard to distinguish | 3 |

## 3.2 Usefulness, Satisfaction, and Ease of Use Results

The Usefulness, Satisfaction, and Ease of Use questionnaire does not contain a scoring formula like the System Usability Scale (below). Scores from each user were tallied and averaged. The user number column does not correspond to a specific user, as the results were anonymous and could not be attributed to a specific participant in the usability studies.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Q13 | Q14 | Q15 |
| 1 | 6 | 6 | 7 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 |
| 2 | 4 | 4 | 5 | 4 | 6 | 6 | 4 | 5 | 6 | 7 | 6 | 6 | 5 | 6 | 7 |
| 3 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 6 |
| 4 | 2 | 2 | 4 | 2 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 6 | 5 | 5 | 5 |
| 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 3 | 3 | 4 | 4 | 5 | 3 | 4 |
| 6 | 5 | 6 | 5 | 4 | 5 | 6 | 5 | 6 | 5 | 3 | 2 | 3 | 2 | 4 | 3 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User | Q16 | Q17 | Q18 | Q19 | Q20 | Q21 | Q22 | Q23 | Q24 | Q25 | Q26 | Q27 | Q28 | Q29 | Q30 |
| 1 | 7 | 6 | 6 | 7 | 7 | 7 | 7 | 6 | 7 | 7 | 6 | 6 | 7 | 6 | 6 |
| 2 | 7 | 7 | 7 | 7 | 6 | 6 | 6 | 6 | 6 | 7 | 5 | 4 | 4 | 7 | 6 |
| 3 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 4 | 5 | 6 | 5 | 5 | 5 | 6 | 6 | 5 | 6 | 5 | 5 | 4 | 4 | 3 | 4 |
| 5 | 6 | 5 | 6 | 5 | 5 | 6 | 5 | 5 | 5 | 6 | 4 | 5 | 6 | 5 | 5 |
| 6 | 6 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

## 3.3 System Usability Scale Results

The SUS scores were calculated by adding the scores of each item (0-4). Odd questions were modified by -1, even items were modified by -5. The sum of the scores was multiplied by 2.5 to obtain the overall score. SUS scores have a range from 0 to 100, with the industry average of 68. The user number column does not correspond to a specific user, as the results were anonymous and could not be attributed to a specific participant in the usability studies.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | SU Score |
| 1 | 4 | 2 | 4 | 1 | 4 | 1 | 4 | 1 | 5 | 1 | **87.5** |
| 2 | 4 | 1 | 4 | 2 | 4 | 1 | 5 | 1 | 4 | 1 | **87.5** |
| 3 | 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | **100.0** |
| 4 | 2 | 2 | 4 | 2 | 4 | 2 | 4 | 2 | 3 | 2 | **67.5** |
| 5 | 4 | 2 | 4 | 2 | 4 | 2 | 3 | 2 | 5 | 3 | **72.5** |
| 6 | 3 | 1 | 3 | 2 | 4 | 1 | 4 | 2 | 3 | 2 | **72.5** |
|  |  |  |  |  |  |  |  |  |  | **AVERAGE** | **81.25** |

## 3.4 Google Analytics Results

This information was gathered from a private Google Analytics account for the dates of October 1, 2017 through November 9, 2017. No sessions or users were excluded from the view.



# Discussion

## 4.1 Conclusions

There were significant usability issues that popped up at the very beginning of testing. The developer within me desperately wanted to make the changes, since they were so simple once users stumbled over them, but to maintain the integrity of the results, I abstained. For instance, the upper-right-hand navigational bar was unclear to half of users (A, B, E). Increasing the prominence on the page by changing the color, making the text bold, or center-aligning the navigation could have improved visibility. Another glaring usability issue was in the date selector. Most users (A, B, C, D, and F) didn’t see the start and end date fields, as they were not styled as a typical form input, but as a single bottom border hortizontal line with labels above the input field.

In addition to completely missing the start/end dates, the system did not allow a graceful resolution if the dates were not selected. It simply gave an unhelpful system style error message and didn’t give the user options to return to the previous page or restart the query process. The lack of built in form validation proved to be a huge roadblock to ALL users at one point or another in the process, and definitely added to their frustration levels. Simply adding this validation could have built trust with the users as well as increasing their self-confidence in using the system.

There were repeated instances of what I assumed to be clear and obvious interactions for the user proving to be not quite so. My naming convention of “Build Query” on the forms for exporting data was a mismatch from the task instruction of “export CSV,” which proved to be a stumbling point for all six users tested. Of the usability issues recorded during testing, 3 were considered a usability disaster (highest priority), 1 was a major problem, 4 were minor problems, 2 were cosmetic problems, and 1 was a new feature request.

## 4.2 Implications for Design

The results of the Usefulness, Satisfaction, and Ease of Use questionnaire didn’t give me any tangible action items, as the scores were all over the chart. The averages for users were 6.3, 5.7, 6.9, 4.3, 5.1, and 4.3 for an overall average score of 5.4. The highest average scores for each individual question were “I don't notice any inconsistencies as I use the system” and “I easily remember how to use the system” which were both scored at an average of 6.33.

The System Usability Scale overall score was alarmingly high, 81.25. Industry average is 68. I removed myself from the environment to try to remove the potential for my presence to invite positive-leaning bias. Unfortunately, the data still appeared to be overwhelmingly positive. Under normal conditions, I would be thrilled to see a high SUS score, but given the amount of huge usability issues that arose during the test, I would disagree in the effectiveness of the current design. This leads me to believe that the results of the SUS should be discarded for any use other than historical comparison between iterations, as they are not accurate nor do they give any meaningful weight to the efficacy of the current design.

Based on the peer review feedback from Milestone 3, I did include Google Analytics on the site itself. The numbers were interesting to me, as the number of unique users and sessions were much greater than expected (17 and 31, respectively). This tells me that users returned to the RTC Data project application after the usability test was over, as the site is not currently indexed in search engine results, nor has it been advertised by staff within Sandia.

## 4.3 Reflections

I would have liked to have significantly more users to test the system, but was unable to recruit more than 6 due to travel and project schedules. This meant I ended up with a few users who did not have a highly technical background or years of familiarity with solar in general, much less with the RTC project. In the future, I would plan the usability testing far ahead to make sure I had more than enough people recruited to participate. Also, as a part of ongoing iterations on the design and development of the system, I would have a standing usability testing meeting during each sprint to continue to reinforce the user centered design principles I wish to follow.

As the design iterates and improves, I would like to continue with this type of quick testing to provide feedback and give me (as a developer) clarity on my own blind spots in designing the system. Additionally, creating additional types of testing (timed task completion, eye tracking/heatmap, and overall cognitive load) will tease out the important yet difficult to pinpoint usability issues that may exist. I would also add another proctor for the testing, so I could be more involved in note taking and observing, rather than trying to also run the test at the same time.

While I do realize that the SUS is overused in the industry and did not allow for deeper, more thoughtful analysis, I felt it was important to get a baseline number to work from for future iterations. I also felt it was important to allow the users time to reflect and review on their experience with the system, rather than forcing them to complete to surveys back to back with me serving as a proctor.

# References

Lund, A.M. (2001) Measuring Usability with the USE Questionnaire. STC Usability SIG Newsletter, 8:2.

Brooke, J. (2013) SUS: a retrospective. J. Usability Studies 8, 2, 29-40.