



WORCESTER POLYTECHNIC INSTITUTE

A MAJOR QUALIFYING PROJECT SUBMITTED TO THE FACULTY OF THE
WORCESTER POLYTECHNIC INSTITUTE IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE ON
MARCH 17, 2016

MATTERS. Data Mining

Bogotov, Dmytro dbogotov@wpi.edu

Hennessy, Jillian jrhennessy@wpi.edu

supervised by

Professor Elke Rudensteiner

Abstract

ABSTRACT HERE!!!111

Executive Summary

IT GOES HERE!!!111

Acknowledgements

HERE!!!111

Contents

| | |
|---------------------------------------|-----------|
| List of Figures | 7 |
| List of Tables | 8 |
| 1 Introduction | 9 |
| 2 Background and Related Work | 12 |
| 2.1 System Architecture | 13 |
| 2.1.1 Model | 13 |
| 2.1.2 Controller | 14 |
| 2.1.3 View | 14 |
| 2.2 Requirements | 15 |
| 2.2.1 API and Documentation | 15 |
| 2.2.2 Metric Builder | 17 |
| 3 Methodology | 20 |
| 3.1 Metric Builder | 20 |
| 3.1.1 Back End | 20 |

| | |
|--|-----------|
| CONTENTS | 5 |
| 3.1.2 Front End | 21 |
| 3.2 API | 24 |
| 3.2.1 Authentication | 24 |
| 3.2.2 Methods | 25 |
| 3.2.3 Query string and output format | 26 |
| 3.2.4 Documentation | 27 |
| 3.2.4.1 Overview | 28 |
| 3.2.4.2 API Key | 28 |
| 3.2.4.3 States | 28 |
| 3.2.4.4 Metrics | 29 |
| 3.2.4.5 Years | 29 |
| 3.2.4.6 Examples | 29 |
| 3.3 Changes to the database | 30 |
| 4 Testing | 31 |
| 4.1 User Testing Procedure | 31 |
| 4.1.1 User Tasks | 32 |
| 4.2 Results | 33 |
| 4.3 Changes made | 36 |
| 5 Conclusion | 40 |
| 5.1 Recommendations | 40 |
| 6 Appendices | 42 |
| 6.1 Appendix 1 | 42 |

| | |
|--|-----------|
| CONTENTS | 6 |
| 6.2 Appendix 2 | 45 |
| 6.2.1 Informed Consent Agreement for Participation in a Research Study . | 45 |
| 6.2.2 Script for Investigators | 47 |
| 6.2.3 Survey | 50 |
| 6.2.4 Results | 54 |
| References | 55 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | System architecture | 14 |
| 3.1 | Data Explorer page | 22 |
| 4.1 | Initial location of the API Documentation link | 35 |
| 4.2 | Footer | 36 |
| 4.3 | Data Explorer sidebar with edit link | 37 |
| 4.4 | Metric Builder | 38 |
| 4.5 | Metric Builder's canvas | 38 |
| 4.6 | "Save" menu | 39 |
| 4.7 | API Documentation page | 39 |

List of Tables

| | | |
|-----|---|----|
| 3.1 | API User database relation | 30 |
| 3.2 | User Metric database relation | 30 |
| 4.1 | Average ratings for the user study on the Metric Builder | 34 |
| 4.2 | Average ratings for the user study on the API Documentation | 34 |

Introduction

With the total amount of data in the world growing steadily at a fast rate, the need to analyze large amounts of data in datasets, or big data, is a major key to productivity growth, and innovation [10]. With this increase in information and the detail of the information available across the web, the Internet will continue to provide a basis for the continued growth of data in the future. Data is a part of every industry and business in today's world, from retail to government. Big data and analysis can help to create value in these industries by making information more usable, showing more accurate performance measures, and providing detailed analytics that lead to better decision making.

In this data-driven time, it is important to be able to make decisions quickly using the data that is available. Big data visualization is an effective way to present the important information amongst the large amount of data used and helps to drive complex analysis [1]. With the use of big data analysis, it can make otherwise too large amounts of data meaningful to those who looking to interpret significance from it.

The Massachusetts High Technology Council, or MHTC, is a group of technological, professional and higher education executives across the state of Massachusetts. They are comprised of higher education CEOs, senior executives from Massachusetts and more. For

over thirty-eight years, MHTC has advocated for various policies and programs to create and keep both a healthy and competitive business climate. Their goal is to establish and keep Massachusetts as a competitive place for successful business and talent building, including a specific focus within the technology sectors. In part because of the MHTC and its members, who are regarded as the region's most venerable technology association, Massachusetts is regarded as one of the top competitive areas for businesses that are high tech [2].

MATTERS, the Massachusetts Technology, Talent, and Economic Reporting System, was developed by MHTC along with Worcester Polytechnic Institute (WPI) and other institutions as a collaborative effort designed to aid users in understanding a variety of public data in an effort to help make Massachusetts the top location for high technology businesses. MATTERS uses this data collected to both measure and evaluate the current state of Massachusetts compared to other states in the country and provides policy makers and advocates with easily accessible and search-able dynamic data to assist in their efforts regarding decisions to help retain and grow business in the state. MATTERS is able to provide a way to centralize a wide range of talent, cost, and economic metrics and state rankings from federal and state government sources, non-profit organizations and media outlets into one location for its users [3].

Previously to our project, WPI IQP, MQP and graduate student teams worked to develop the MATTERS dashboard as it was when we began working. The goal of our Major Qualifying Project (MQP) is to provide additional features to the MATTERS website that will give our users additional ways to interact with the data and metrics found within the site. The idea for the Metric Builder was presented by the MHTC members as a means for the MHTC to be able to visualize their own indicator that they were in the process of

creating. Users will be able to create an account on the MATTERS site and be able to build and display the results of their own unique metric formulas for ranking states. Users will also be able to retrieve any of the data points found throughout the site for their own personal use through the creation of our own API. These new features will provide authorized users an opportunity to use and share the data in even more innovative ways, aiding in the MHTC's goal toward making Massachusetts a leading state for high technology business. These two features create further access to the site's data and additional ways to manipulate and visualize the data to interpret it in new ways. These features also create an expanded user base for MATTERS and different privileges for different types of users, such as the already existing administrative users, and general users who also would like to create their own indicators.

Background and Related Work

The creation of a Metric Builder page and an API for the MATTERS site is a result of two prior project groups and additional efforts by WPI students who worked to establish the initial features and the design framework for MATTERS to exist. The initial development of MATTERS came from a team of students completing their Interactive Qualifying Project (IQP) with the aid of other WPI graduate student teams. This team of IQP students researched and came to the resulting decisions regarding the front-end visualization for the MATTERS dashboard [4]. This includes the inclusion of decisions regarding colors, and visualization types such as charts and tables, that would be used to represent the data within the system in a manner that was both intuitive for the users and a quick way to show analytics. In parallel, the graduate student teams worked to develop the back-end for MATTERS. They made it possible to extract the data wanted from various sites and then parse and clean this retrieved data, before then uploading it to a database. This data would be called upon to be shown in the various visual methods for the users [5].

Following this initial front-end and back-end design and implementation of the MATTERS dashboard, a team of students working on their Major Qualifying Project (MQP) for WPI then worked to improve the administration center and the data integration for the

dashboard. This involved the data integration pipeline manager as well as the Administration Center's development to allow for MHTC and other administrators who had access to be able to easily upload their desired data and metrics, as well as an easy way to view this data on the administrative end. This MQP team was able to integrate 20 new data sources for the MATTERS dashboard, make the Administration Center usable for non-technical users, and improve upon the look and feel of the MATTERS dashboard as a whole [5].

2.1 System Architecture

The MATTERS Website consists of two parts - *admin panel* and *dashboard*. While the *admin panel* is used by administrators to work with data (basic CRUD operations), in this project we focused on the *dashboard* subsystem. This subsystem follows MVC (Model-View-Controller) architecture. You can see its visual structure on the Figure 2.1

2.1.1 Model

System's data layer is built with the PostgreSQL database. The main tables are

Statistics holds metrics' data in the form of a year, state, metric and value tuple;

States holds states' metadata like name, abbreviation and if it is a peer state;

Metrics holds metrics' metadata like description, trend and category;

The access to these tables is provided via stored procedures.

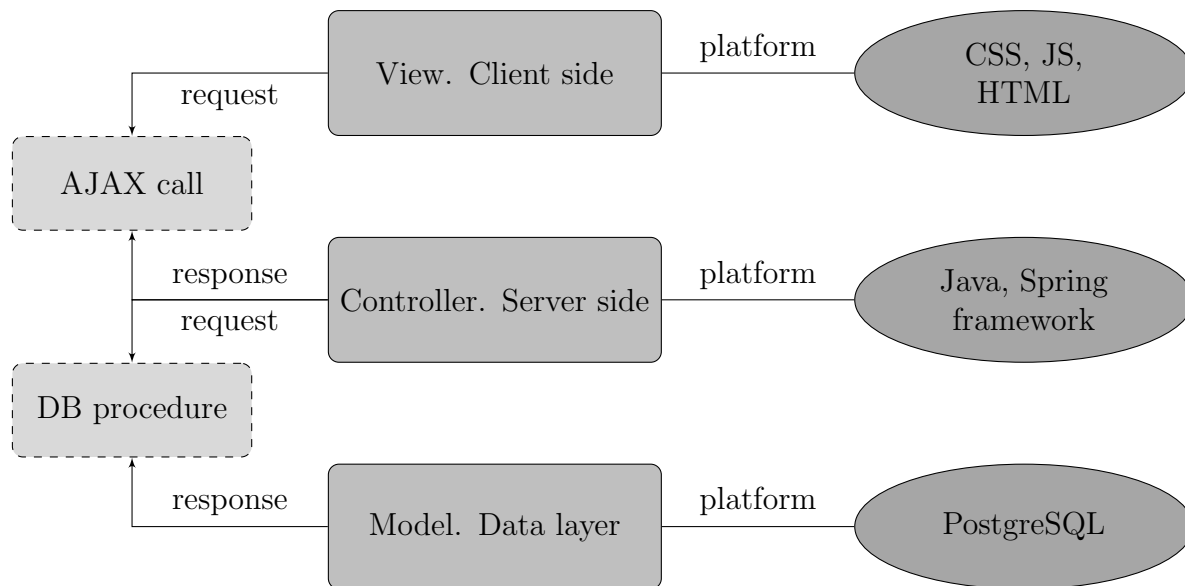


Figure 2.1: System architecture

2.1.2 Controller

The business logic of the system resides on a Java-based server. Spring framework is used as a server architecture solution. Front end request that comes to the system goes through controllers which, in turn, invoke services. Services query the data layer and return data stored in models. These data models get pushed back to the front end.

2.1.3 View

The presentation layer is represented by the HTML based website. Client side generates queries to the server via JavaScript AJAX calls, then uses returned data to generate the content. The website has a significant amount of logic on the client side as it needs to build tables, plots and charts. The two main JavaScript libraries used on the client side are jQuery

and D3. jQuery makes it easy to manipulate DOM while D3 is used for generating vector based plots and charts.

Visual representation is based on the responsive Twitter Bootstrap 3 framework, which is a collection of CSS classes and small JavaScript snippets.

2.2 Requirements

The main goal of creating an API for MATTERS as well as creating the Metric Builder page for the site and what it would entail, required looking into other similar features of different sites and making decisions regarding implementation, design and security features for these aspects of the MATTERS site.

2.2.1 API and Documentation

An API is an Application Programming Interface. APIs are tools that allow other programs to interact with the software program that the API belongs to without having to give someone complete access to your code base. In a sense they expose some of the internal workings of software to the public but in a way that is limited by what the API chooses to provide. APIs aid in providing a way to share a potentially large amount of data in an efficient way. According to the chief data officer of Philadelphia, Mark Headd, APIs "allow a specific audience to use data more quickly, easily, and efficiently when they are looking to do something specific with the information." APIs are beneficial for not only keeping certain aspects of code private, but also for saving users time. Even in cases of open source programs where all of the code for a program is visible, the number of code can be so large that it is inefficient

to search through the whole code base for specific data [6].

APIs will be beneficial from both a business aspect and for programs that aim share information and research with many people. An advantage of an API being an automated tool is that it can process a large number requests without any added cost and work from the developers. APIs also present the data requested in a manner that is useful and easily manipulated for the users' purposes. For APIs whose goals are to share certain data across systems there is a great benefit efficiency wise, as many manual procedures are cut out by the APIs automatic generation. When these tasks are done manually, they waste time and are both laborious and repetitive, as well as being more prone to error; all of these things being costly on the developer end. While it is more time consuming and costly to develop an API initially, the overall benefit once it is developed is very clear [6].

Additionally, APIs not only encourage innovation and the ability to manipulate and extrapolate data by external collaborators, they provide a means to do so that is more secure for the group creating the API. They reduce the risk of how and what data is obtains and encourage good practice for how to properly manage the data available [7].

One important step to gaining the full benefits from an API is making sure that the API is secure. This can be achieved through requiring authentication and authorization of users before they gain access to the API's features. It is good practice to ensure that the users are who they say they are and that they are given permission before using data or parts of a developer's program. This also allows for a record of who exactly is accessing the data, in the event that there is an issue so that a specific user can be tracked down [8].

In addition to creating the API for a site, it is necessary to create good documentation to accompany the API, to make the API as user friendly and simplistic as possible. API

Documentation provides developers with everything that they need to know in a concise manner in order to use the APIs features with other applications. The API Documentation needs to contain the classes, functions and other important aspects of the API [9]. The documentation guide allows developers to easily interact with a site's API and their own code, as many developers style their code in their own unique way.

2.2.2 Metric Builder

The main feature that is to be developed is the Metric Builder page which will be available for the registered users on the MATTERS website only. A user will be able to select any number of the metrics already available in the MATTERS database and combine their data to create their own personal formula or indicator. The user can then use their created formula to display the states' rankings based on this indicator. This will allow users to see and evaluate states based on what they think is most important as a whole or combination of various factors, and interpret the data in their own way, simply. By allowing users to create their own data set from all the metrics available and then provide their own weights to each of these data values, we are allowing users to see the value that each state has based on how important they find a large number of metrics.

It is important to ensure that many different data types are able to be combined to create one final value. It is also important to consider the effects that certain data types will have on the final value received from the user's indicator. For example, the MATTERS database contains rankings, percentages, nominal values, and other data types. Certain data types could end up dominating in a user created indicator when combined with significantly smaller value data types. There are also complexities involved with certain data types that

are represented in the reverse of the majority of the data types. For example, both rankings and some other data types use numerical values in which a lower value indicates a better performance by the state, versus the majority of other metrics in which the larger the value is, the better a state is performing in that area.

Finally, there are many metrics that do not contain a value for every single year that other metrics contain. Situations where one metric being used in the user's Metric Builder equation does not contain values for years that another metric being used in the same equation need to be considered. There needs to be a uniform solution for what to do in these instances in a way that does not deceive the users of the indicator, as this would affect the calculations and following visualizations across years where these issues exist.

These factors need to be taken into consideration in order to provide accurate and understandable visualizations and results when users create their own indicator in the Metric Builder. It is for this reason good to consider special handling or removal of certain data types, as well as potential ways to normalize the final numerical values received from the formulas the users have created in the Metric Builder.

We looked at various economic indicators to come up with an idea for how to allow our users to combine sets of data to use for comparisons. One common approach in economic indicators is to use weighted averages [10]. For example, one of the most popular economic indicators, CPI, or the Consumer Price Index, uses weighted averages of data from hundreds of different consumer goods and services to measure inflation and deflation. Each of the goods considered by the CPI is weighted based on its importance [11].

Weighted averages work by taking each number in a given data set, and multiplying the value by its given weight. This will give you a new value reflecting the product of the data

and its weight. Add all of these products together from the data set to get the total value. Then, add up the total of all the weights. Finally divide the total value by the total weight in order to receive the weighted average of your data set. The weights are used to show the importance that each specific item in a data set has in order to calculate an overall value of all the data together [12].

Methodology

3.1 Metric Builder

3.1.1 Back End

Implementing a Metric Builder the core question was how to store and compute a user metric. We have come to these two options

1. Store a user metric data in the database
 - Pros
 - Fast to retrieve; no computation needed
 - Cons
 - Waste of space in the database
 - User metric data remains static; update in the underlying data does not cause an update in the user metric
2. Store a user metric as a metadata only and compute values on-the-fly (on-demand)
 - Pros

- Efficient use of the database space
- User metric data is always in sync with underlying data
- Cons
 - Might be time-consuming to compute

Looking at the cons of each option, we attempted to approximate how bad they are. Having n users each creating m user metrics each of which has data for k years we have $m \cdot n \cdot k \cdot 50$ entries in the database. This data would be a snapshot. If any data point in the metrics changes, user metrics remain out of sync.

For the second option, it was difficult to approximate time it takes to compute a user metric as it depends on the server load and the number of metrics involved in the user metric. We implemented the second option and benchmarked it with 3 users trying to compute largest possible user metric (all metrics with all states selected). It took around 1.5 seconds to load all 3 user metrics. This drove our design decision towards the second option - computing metrics on-the-fly.

3.1.2 Front End

The main goals for the front end of the Metric Builder feature was to design a webpage that was easy to use, easily accessible, intuitive, and consistent with the design of other pages on the MATTERS site. The Metric Builder feature needed to be a tool that allowed users with any level of technological background to easily build their metric formula from preexisting metrics, assign weights to their chosen metrics, and be able to make changes to their metric formula at any time.

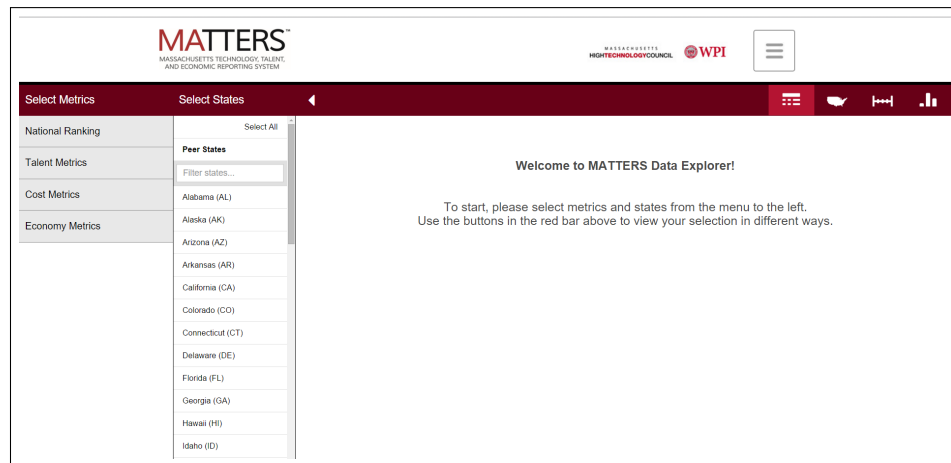


Figure 3.1: Data Explorer page

The front end for the Metric Builder page was created through the use of HTML, CSS and JavaScript. To make the page consistent with the rest of the site, code was reused from the Data Explorer page. The Data Explorer page, as seen in 3.1, within MATTERS allows users to select preexisting metrics to visualize the data on the site in multiple different ways, including a bar chart, line graph, heat map, or table. The framework for the Data Explorer was the page most similar to how we wanted the Metric Builder to look, and we were able to reuse and modify code from that page to create our new feature.

The Data Explorer page allows users to select metrics and states that they would like to view through the use of the sidebar. Because the intended users for the Metric Builder feature are meant to be authorized users with experience using the MATTERS Data Explorer, we used the sidebar from the Data Explorer as the basis for how the users would select their metrics on the Metric Builder page as well. Since the Metric Builder does not need a user to select states, that part of the sidebar was removed. We also removed rankings from the metric selection column of the sidebar since rankings are manipulations of various other

metrics already found in the database.

Once a user selects the metrics they wish to use in order to create their own metric formula, they have the ability to assign each of them different weights to decide how much of a factor each of the metric's data will be in the formula. To adjust the weight of a metric, users can use the sliders that appear under each metric that they have selected. The sliders work horizontally from left to right and allow a weight from zero (which would be the same as not including the metric) to one hundred. Users can click on a section of the slider to move to that weight, click and drag to a specific weight or use the arrow keys to move the slider's value up or down. The idea for the sliders came from a New York Times site which allows users to interactively calculate their rent and other values using visual sliders ???. The sliders are visually appealing and simple to use and understand.

The Metric Builder displays options to name the user's custom metric, save their metric and also go back and edit or delete the metric they have created. There is a simple to use button that says "Remove" next to each metric chosen, making it easy for users to change the metrics included in their formula. The input box to name the custom metric and save the changes can be found below all of the chosen metrics and sliders. This was done to help users and encourage them to properly weigh all the selected metrics before leaving saving their custom metric formula to use, as users typically look at a webpage from top to bottom.

Finally, above all of the metrics selected is a box displaying the basics of their created metric formula. As the users update metric selections and their various weights, the formula will update and show the basics of the math that needs to be done on the back end. This includes showing the users that they are multiplying each data point by the chosen weight before adding each metric's data values together, and that sum is being divided by the total

weight as a weighted average approach. This is shown in a simple way to make sure that users can follow along and are aware of exactly what they changing when they adjust their formulas. Once content, each time a user saves and updates their formula they will be prompted and subsequently directed to the Data Explorer page where they can begin using their custom metrics and displaying the data with the various visualization options.

3.2 API

Implementing an API, there are 2 problems to consider

- Do we implement a closed API or open API? If closed one, how do we implement authentication.
- Which methods do we put in the API?
- How our query string will look like? What is the output format?

3.2.1 Authentication

Although MATTERS system is a public resource, we wanted to at least track who uses our API, how and how often. This requirement made us chose a closed API option. Thinking of authentication we had a few options

- Bind to user
- Simple *API key* (token) authentication
- Complex *handshake* mechanism

- *Handshake* mechanism with random value

Binding to API user t regular user would allow any registered user to access API. Although this kind of authentication is the easiest one to implement we wanted a separate API users.

Simple *API key* authentication is the one when API users are given unique token (long random string) which they need to provide with each request. Based on this token system is able to accept or deny connection and track individual users.

Handshake mechanism with random value involves the following steps

- Client sends in username (not password)
- Server responds by sending back unique random number
- Client encrypts using user's password plus number as key
- Client sends hashed or encrypted value back to server
- Server encrypts using the user's same password plus number as key
- Server compares two hashed or encrypted values, if same then grant access

Provided that the main purpose of the API being closed is tracking an activity for statistical purposes, *API key* authentication was implemented.

3.2.2 Methods

Main method implemented in the API is `api/data?`. This method returns all data points for given metrics, states and years.

Since `api/data?` needs specific IDs as arguments, there is a need for supporting methods which show information about metrics and states. So we also implemented `api/metrics?` and `api/states?`.

3.2.3 Query string and output format

There are two options for constructing a query string.

- `api/method?argument1=value1&argument2=value2&...`
- `api/method/value1/value2/...`

To make our API user friendly we implemented the first option since it clearly shows which value corresponds to which arguments and these argument-value pairs may be in any order.

The final query strings are given in Listing. 3.1

Listing 3.1: API methods examples

```
1  /* method */
2  /api/data?metric=&state=&year=&apiKey=
3  /* example */
4  /api/data?metric={16,32}&state=MA&year=*&apiKey=key
5
6  /* method */
7  /api/metrics?apiKey=
8  /* example */
9  /api/metrics?apiKey=key
```

```
10
11 /* method */
12 /api/states?apiKey=
13 /* example */
14 /api/states?apiKey=key
```

The output format JSON (JavaScript Object Notation) was chosen as it both compact and human and computer readable.

3.2.4 Documentation

In order for the developers to know how to make proper API calls, it was necessary to create the API Documentation page for MATTERS. The documentation provides users with full information on the MATTERS API and how to use it, as well as examples of valid API calls and the output that they would return. It is important to ensure that the API Documentation is free of both natural language and code sample errors, as this tool is what developers unfamiliar with using the API rely on to understand how to properly access and extract data [13].

In order to make the documentation easy to follow, the documentation was divided into sections. The main sections of the API Documentation were: an overview, the methods, the parameters, and examples. The layout of the MATTERS API Documentation was based on the API Documentation of Tradestation, which had been very simple and easy to follow [14]. The Tradestation documentation provided a summary, path, parameters, and examples of both an API call and what it would return. An example of their documentation

can be found at: <http://tradestation.github.io/webapi-docs/en/users/accounts/>. For the MATTERS API, it was necessary to include information on where to access a user's API key, as well as the proper format to get the data desired for different, years, states, and metrics. The API Documentation sections are explained in the following sections. The full API Documentation can be found in Appendix 1.

3.2.4.1 Overview

The API Documentation begins with a summary about what the MATTERS API is and what data points the API tool gives users access to. This section also includes an interactive table of contents which allows users to quickly navigate to a specific section of the documentation to look for information. This allows users to easily find what they are looking for without having to scroll through the entire documentation page and provides a brief overview of everything that follows.

3.2.4.2 API Key

Developers without authorization do not have the ability to use the MATTERS API until they have been given an API key. These users are directed to the Contact Us page in order to send an email request to become an authorized user with their own key. This section also explains that the API key is used in all API calls in order to validate the user's request.

3.2.4.3 States

Users are instructed on the proper format needed to retrieve a single state, list of states, or all states. If a user does not know any other information regarding the states in the

MATTERS system, they are shown where to find any other information needed via an API request.

3.2.4.4 Metrics

Users are given the proper format in order to request data for a single metric, list of metrics or all metrics in the MATTERS database. The documentation explains that in order to access data from metrics, the metric parameter values must be the proper metric's ID number. The user is also given the path to all information regarding the metrics, which will give the user the metric names, ID numbers and details.

3.2.4.5 Years

Users are given the proper format to request data from specific years. Users may request data for a single year, range of years or all years in the database.

3.2.4.6 Examples

Following the parameters and methods of the MATTERS API, the documentation provides the users with two examples of valid API requests and the data which they would return. The example requests include the proper format of the parameter values and the various different types of calls the users can make to request specific data. Under the path name showing an API request the option to display the return value of that specific example. Users can view this to see the return data as well as the JSON format it would be given in.

| Name | <i>Type</i> | Comment | Reference |
|---------------|---------------------|---|-----------|
| Id | <i>INT</i> | Unique auto-incremented identifier | none |
| Name | <i>VARCHAR</i> | API user's name | none |
| ApiKey | <i>VARCHAR(160)</i> | Unique random string; used for authentication | none |

Table 3.1: **API User** database relation

| Name | <i>Type</i> | Comment | Reference |
|---------------|----------------|--|----------------------------|
| Id | <i>INT</i> | Unique auto-incremented identifier | none |
| Name | <i>VARCHAR</i> | User metric's title | none |
| Value | <i>TEXT</i> | JSON formatted list of metric id - coefficient | none |
| UserID | <i>INT</i> | Id of a user - author of the user metric | Id in User relation |

Table 3.2: **User Metric** database relation

3.3 Changes to the database

We had to add two tables to the database - **APIUser** and **User Metric**. See Table. 3.1 and Table. 3.2.

Testing

After the back end code and the initial front end of the Metric Builder, API and API Documentation features were complete, our team decided to receive user feedback to see where we could improve our design. We created a two part user study to first test the usability, intuitiveness, and simplicity of the Metric Builder's user interface and finally to test the clarity and completeness of API Documentation and to make sure there are no errors in the code samples or the natural language text within the documentation.

4.1 User Testing Procedure

All of the users were students from WPI and were asked to complete the user tasks on a laptop provided by our team. Even though the users in our study are not the typical expected users of the MATTERS site, we believe that their input is valuable. If the features of the site are easy to use for users inexperienced with the MATTERS site, then we should expect the features to be easy for the authorized users in the future as they are more experienced with the MATTERS site or with developing APIs. After completing the tasks for each part of the user study, the users filled out a survey asking them to rate the easiness of different aspects of the tasks on a scale of one to five. While the users were completing their tasks, our team

was also there observing their actions in order to get better insight on their clicking behavior and thought process while navigating the site. We also recorded the operating system and browser the users were doing the study on to note any important differences this may bring about.

4.1.1 User Tasks

The user tasks given to the users in the following order and were printed on a sheet of paper so that the users could reference the task details at any time:

User Study Tasks: Part 1, Metric Builder feature

1. *Navigate to the Metric Builder feature*
2. *Please create your own metric formula, which includes Unemployment Rate, Corporate Income Tax Rate and Median Household Income, with weights 10, 5, and 22 respectively.*
3. *Name your new metric "My Metric" and save it.*
4. *Please find and select your created metric on the Data Explorer page.*
5. *Edit your metric formula so that Unemployment Rate is no longer a part of the metric and Median Household Income has a weight of 15.*
6. *Save your changes to My Metric.*
7. *Please delete My Metric*
8. *You will now receive a short survey about your experience using the Metric Builder.*

User Study Tasks: Part 2, API feature and Documentation

- 1. Navigate to the API Documentation page*
- 2. Your API Key is "secret"*
- 3. Please get the data for the metric Unemployment Rate for Massachusetts and Florida for the year 2012 using the API tool and the documentation as a guide*
- 4. You will now receive a short survey to complete about your experience using the API and API Documentation.*

The survey that the users completed at the end of each part of the user study can be found in Appendix 2.

4.2 Results

Thirteen users completed the user study and filled out the subsequent surveys. Most users performed the tasks on Google Chrome or Safari and all users performed the tasks on either a Mac Operating System or Windows Operating System. All of the data collected from the user studies can be found in Appendix 2. Table 4.1 shows the average of the users' ratings on a scale of one to five for each question on the survey regarding the Metric Builder feature.

Table 4.2 shows the average of the users' ratings on a scale of one to five for each question on the survey regarding the API Documentation.

Our team considered an average rating of above four to be a good score, meaning that the specific task was easy to perform with our user interface. The only tasks that did not score above our desired rating was for How easy was it to edit 'My Metric' to make the appropriate changes? and How easy was it to find the API Documentation?. User feedback on the open ended survey responses, as well as observations made by our team during the

| Survey Question | Average User Rating |
|--|---------------------|
| How easy was it to find the Metric Builder? | 4.385 |
| How easy was it to select the metrics "Unemployment Rate", "Corporate Income Tax Rate", and "Median Household Income" in the Metric Builder? | 4.308 |
| How easy was it to weigh the metrics accordingly? | 4.615 |
| How easy was it to name and save your metric? | 5.000 |
| How easy was it to find My Metric on the Data Explorer page? | 4.231 |
| How easy was it to edit My Metric to make the appropriate changes? | 3.846 |
| How easy was it to remove My Metric? | 5 |

Table 4.1: Average ratings for the user study on the Metric Builder

| Survey Question | Average User Rating |
|--|---------------------|
| How easy was it to find the API Documentation? | 3.250 |
| How useful is the API Documentation? | 4.167 |

Table 4.2: Average ratings for the user study on the API Documentation

user studies were able to confirm that these two tasks were the most difficult or confusing and provided insight as to how to improve it.

For the Metric Builder, users noted that they wanted an option to allow them to edit their metric from the Data Explorer page. Our observations confirmed that while on the Data Explorer page, users were actively searching for an edit button while they were viewing their custom made metric. Many times the users would find the "Create new metric link quickly, which if clicked would bring the user to the Metric Builder page where they could also edit their metric, but were hesitant to click the link because they were not interested in starting over with a new metric.

For the API Documentation, many users commented on their difficulty finding the link

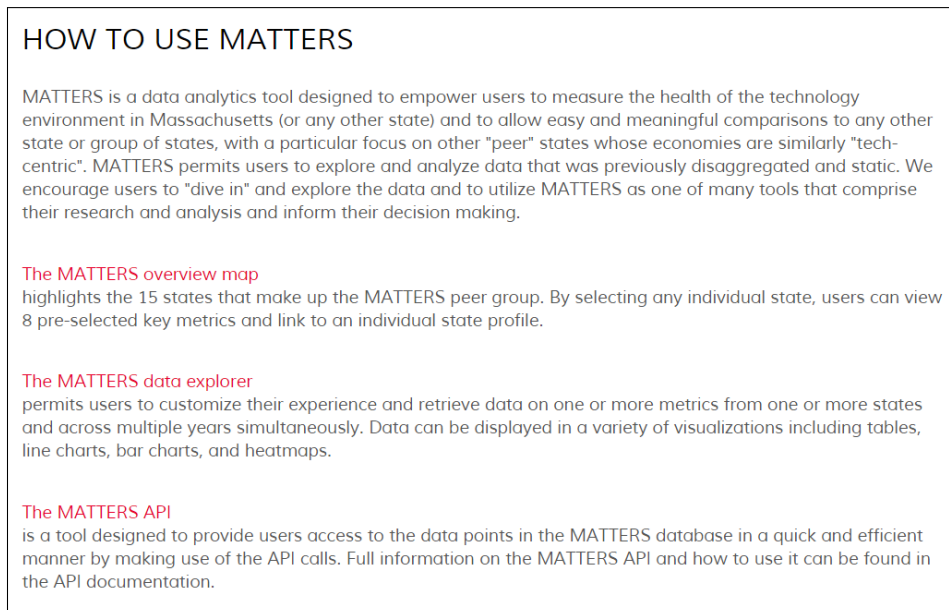


Figure 4.1: Initial location of the API Documentation link

for the API Documentation page. Ultimately, twelve out of the thirteen users did not have much of an issue eventually navigating to the API Documentation link found on the How to use MATTERS page as seen in Figure 4.1; however, it was more time consuming than it should have been and many users clicked on multiple other links before navigating to the correct page. The users who stated that they did not know what an API was mainly looked for the API Documentation link in the main drop-down menu of the site before navigating to the About page, followed by clicking on the link to How to use MATTERS and finally finding the API section which led them to the documentation. Most of the users who were aware of what an API was, ultimately navigated to the API Documentation page the same way; however, many of them scrolled to the bottom of the home page of the site looking for a link similar to the Developers links found at the bottom of many webpages.



Figure 4.2: Footer

4.3 Changes made

As a result of the user ratings, comments and our observations made while completing the user study tasks, we made a few front end changes to make our features easier to use and easier to find. This included adding a link to the API Documentation in the footer of the site in addition to the link on the "How to use MATTERS" page, as seen in Figure 4.2, and as shown in Figure 4.3, changing the Create new metric link on the Data Explorer page to mention that the link would also allow users to edit their metric.

These changes led to the final design of the front end of the Metric Builder and API Documentation which can be seen in the figures below. Figures 4.4 and 4.5 show the Metric Builder page when a user first visits the page and after a user has selected and assigned weights to some of the metrics from the Metric Selection column of the sidebar. Once the user has decided to save their metric formula, if they choose to go back and edit any part of it, the "Save" button will be replaced by an "Update" button and options are added to "Cancel" any changes or remove the custom metric completely. These additional options can all be seen in Figure 4.6. Finally, there were no changes that needed to be made to the API Documentation itself as a result of the user testing. Figure 4.7, shows the API Documentation

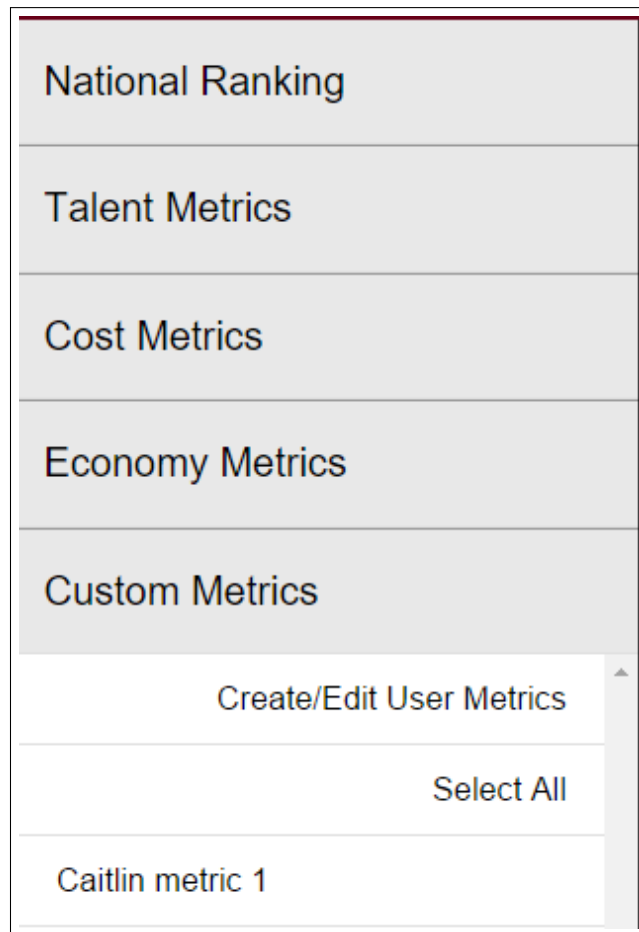


Figure 4.3: Data Explorer sidebar with edit link

page as it appears when a user first navigate to the page from either the footer or the "How To Use MATTERS" page.

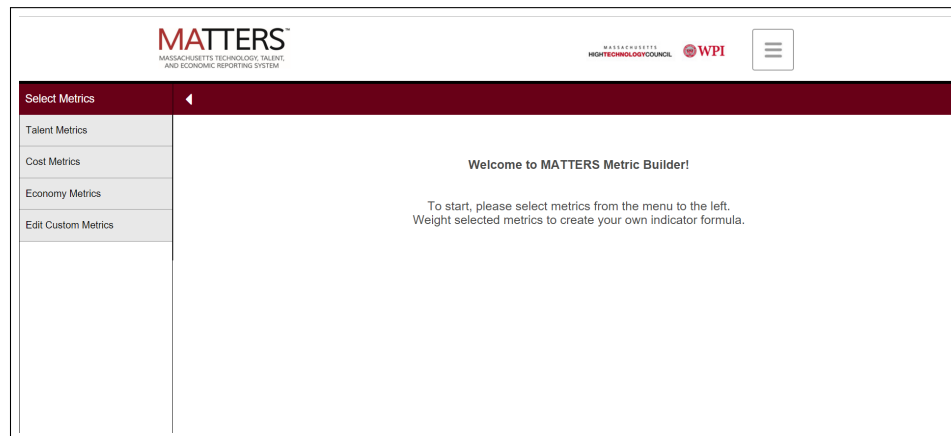


Figure 4.4: Metric Builder

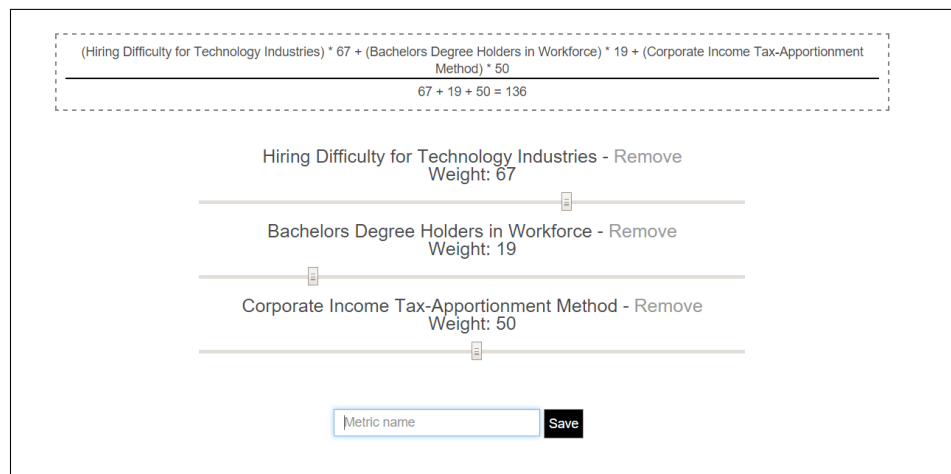
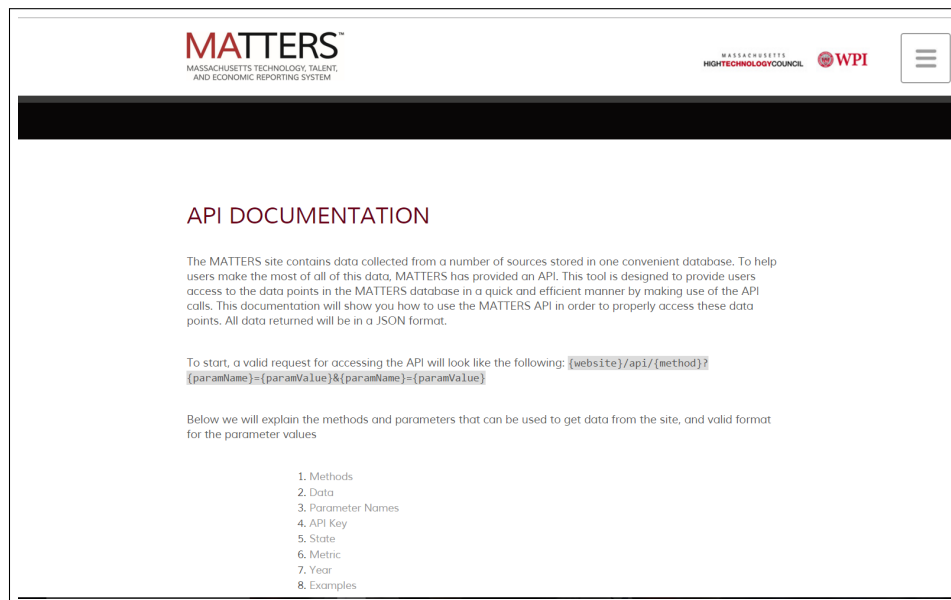


Figure 4.5: Metric Builder's canvas



A screenshot of a web form. At the top, there is a text input field containing the text "Caitlin metric 1". To the right of the input field is a black button with the word "Update" in white. Below the input field and the "Update" button are two more black buttons: "Remove" on the left and "Cancel" on the right.

Figure 4.6: "Save" menu



The screenshot shows the MATTERS API documentation page. The header includes the MATTERS logo (MASSACHUSETTS TECHNOLOGY, TALENT, AND ECONOMIC REPORTING SYSTEM), the MASSACHUSETTS HIGHTECHNOLOGY COUNCIL logo, and the WPI logo. A navigation menu is visible on the right. The main content area is titled "API DOCUMENTATION" and contains the following text:

The MATTERS site contains data collected from a number of sources stored in one convenient database. To help users make the most of all of this data, MATTERS has provided an API. This tool is designed to provide users access to the data points in the MATTERS database in a quick and efficient manner by making use of the API calls. This documentation will show you how to use the MATTERS API in order to properly access these data points. All data returned will be in a JSON format.

To start, a valid request for accessing the API will look like the following: `{website}/api/{method}?{paramName}={paramValue}&{paramName}={paramValue}`

Below we will explain the methods and parameters that can be used to get data from the site, and valid format for the parameter values

1. Methods
2. Data
3. Parameter Names
4. API Key
5. State
6. Metric
7. Year
8. Examples

Figure 4.7: API Documentation page

Conclusion

In this project we have built two services for the MATTERS site. The API provides a way for other computer-based systems to work with the MATTERS system. Specifically, it is now possible to query MATTERS data directly, bypassing the presentation layer. The other feature is a Metric Builder. This subsystem allows users to create their own unique metrics based on the existing metrics. Users are able to work with their custom metrics seamlessly in the Data Explorer as if they were just regular ones.

5.1 Recommendations

Recommendations for other teams, either MQP or other teams, working on improving the MATTERS system in the future are as follows:

1. Refactor the client side and server side code. Since the system has been developed by a number of teams with different design patterns and approaches, there is a significant technical debt. It might be worth spending some time rewriting some parts of the system according to the latest coding and technical standards.
2. Documentation. Right now there is a steep learning curve for the new developers who

start working on the project. Good documentation may dramatically decrease the time it takes for developers to start working on the project.

3. Role management. Right now the system supports regular users and API users. Each user metric must have one and only one author. It might be a good idea to merge these two user entities and introduce shared user metrics that are not binded to specific users.

Appendices

6.1 Appendix 1

API Documentation:

API DOCUMENTATION

The MATTERS site contains data collected from a number of sources stored in one convenient database. To help users make the most of all of this data, MATTERS has provided an API. This tool is designed to provide users access to the data points in the MATTERS database in a quick and efficient manner by making use of the API calls. This documentation will show you how to use the MATTERS API in order to properly access these data points. All data returned will be in a JSON format.

To start, a valid request for accessing the API will look like the following: `{website}/api/{method}?{paramName}={paramValue}&{paramName}={paramValue}`

Below we will explain the methods and parameters that can be used to get data from the site, and valid format for the parameter values

1. Methods
2. Data
3. Parameter Names
4. API Key
5. State
6. Metric
7. Year
8. Examples

- URL Path: `matters.mhtc.org/api/data`

This method can be used to access any of the data points from the site

The following are the potential parameter names that can be used along with the data method to access data for certain states, years, and metrics.

The API key is the key associated with the person trying to access the data. You can get your personal API key by contacting us. You must be a registered user with your email address in the system in order to receive an API key. We will respond to the email with your specific API key that you can use to access the data points.

- URL Path: `matters.mhtc.org/api/data?apiKey=yourKey`
- Validation: the API key is used to check if a user is allowed access to the data he wishes to return. For this reason, you must always include the API key parameter followed by your unique key.

State is used to choose which state(s) you would like to access the data of. A list of states and their information can be found by going to: matters.mhtc.org/api/states?apiKey=yourKey

- URL Path: matters.mhtc.org/api/data?state={paramValue}

2015. Worcester Polytechnic Institute. All Rights Reserved.
Sponsored by Mass High Technology Council

<http://localhost:8080/mhtc/api>

1/2

3/3/2016

MATTERS

- Valid {paramValues}:
 - * : this parameter is to be used to represent all states. This is the default setting if no state is given.
 - **peer** : this parameter will select data for all peer states
 - {**state abbreviation**}: giving a single state's two letter abbreviation (ex: {MA} for Massachusetts) will return data for that one state
 - {**state abbreviation, state abbreviation, ...**}: giving a list of proper state abbreviations (ex: {MA, FL}) will return the data for all states listed

Metric is used to choose which metric(s) for the states chosen that you would like to access the data of. A list of metrics and their information can be found by going to: matters.mhtc.org/api/metrics?apiKey=yourKey

- URL Path: matters.mhtc.org/api/data?metric={paramValues}
- Valid {paramValues}:
 - * : this parameter is to be used to represent all metrics. This is the default setting if no metric is given.
 - {**metric ID**}: giving a valid metric ID number (ex: {19}) will return the data for the metric associated with that ID number
 - {**metric ID, metric ID, ...**}: giving a list of metric IDs (ex: {19, 20}) will return the data for all metrics listed

Year is used to choose which year(s) for the states and metrics chosen that you would like to access the data of

- URL Path: `matters.mhtc.org/api/data?year={paramValue}`
- Valid {paramValues}:
 - * : this parameter is to be used to represent all years. This is the default setting if no year is given.
 - year: giving a valid year (ex: 2014) will return the data for the specific year chosen
 - {year-year}: giving a range of years (ex: {2010-2014}) will return the data for all years in the range

EXAMPLE 1. `http://matters.mhtc.org/api/data?state={AL}&metric={18}&year=2012&apiKey=yourKey`

This would return the data points for Alabama for metric 18 for the year 2012, provided that the user has permission to access this data, as a JSON. Click to show:

EXAMPLE 2. `http://matters.mhtc.org/api/data?apiKey=yourKey&state=peer&metric={19}&year=2013`

This would return the data points for all peer states for metric 19 for the year 2013, provided that the user has permission to access this data, as a JSON. Click to show:

6.2 Appendix 2

Full User Study documents.

6.2.1 Informed Consent Agreement for Participation in a Research Study

Investigator: Jillian Hennessy or Dmytro Bogatov

Contact Information: Jillian (jrhennessy@wpi.edu) or Dmytro (dbogatov@wpi.edu)

Title of Research Study: Metric Builder and API Documentation User Testing

Introduction You are being asked to participate in a research study. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your

participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

Purpose of the study: The purpose of this study is to test and receive user feedback on our Metric Builder and API features of the MATTERS site. We are looking to see how easy it is use these features and if there are any changes to the design that need to be made.

Procedures to be followed: You will be asked to use the Metric Builder and API and complete a set of tasks given to you. During this time, you will be observed by the instructor to see how you navigate the site/ how long these tasks take to complete. After completing the tasks, you will be given a short survey about your experience with the Metric Builder and API features and their design.

Risks to study participants: There are no risks to you if you choose to take part in this user study. You may choose to stop this study at any time and your responses/ observation will not be used.

Benefits to research participants and others: There are no direct benefits to completing this study. You will be helping our MQP team to ensure our features are ready to be deployed and used by their intended targets as well as find any potential problems with the features that we may want to fix.

Record keeping and confidentiality: We will not use any of your personal information as part of the results of the study. Our observations will be linked to the ID of the survey however your name will not be used. Any responses given in the survey will remain anonymous. Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or it's designee and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI

IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you.

Compensation or treatment in the event of injury: You do not give up any of your legal rights by signing this statement.

For more information about this research or about the rights of research participants, or in case of research-related injury, contact: Please see the contact information at the beginning of this document. In addition, you may also contact the IRB Chair (Professor Kent Rissmiller, Tel. 508-831-5019, Email: kjr@wpi.edu) and the University Compliance Officer (Jon Bartelson, Tel. 508-831-5725, Email: jonb@wpi.edu).

Your participation in this research is voluntary. Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

By signing below, you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

6.2.2 Script for Investigators

Have them look over and sign the consent form

You will also sign the consent form

Have them complete part one of the survey

User Study Tasks: Part 1, Metric Builder feature

log them into the MATTERS site so they have access to the Metric Builder feature

we will observe as they complete the tasks, and will record the time for how long it takes to find the metric builder page, and their metric in the data explorer

if they cannot find the link in a reasonable time, bring them to the Metric Builder page and make a note

Navigate to the Metric Builder feature Please create your own metric formula, which includes Unemployment Rate, Corporate Income Tax Rate and Median Household Income, with weights 10, 5, and 22 respectively.

Name your new metric My Metric and save it.

Please find and select your created metric on the data explorer page.

Edit your metric formula so that Unemployment Rate is no longer a part of the metric and Median Household Income has a weight of 15.

Save your changes to My Metric.

Please delete My Metric

You will now receive a short survey about your experience using the Metric Builder.

Have them complete part 2 of the survey on their own

User Study Tasks: Part 2, API feature and Documentation

we will observe as they complete the tasks, and will record the time for how long it takes to find the API Docs page

if they cannot find the link in a reasonable time, bring them to the API Docs page and make a note

Navigate to the API Documentation page

Your API Key is secret

Please get the data for the metric Unemployment Rate for Massachusetts and Florida for the year 2012 using the API tool and the documentation as a guide

You will now receive a short survey to complete about your experience using the API and API Documentation.

Have them complete part three of the survey

Upon completion make sure the survey has been submitted, ask them if they have any other questions or feedback, make any notes, and thank them for participating

6.2.3 Survey

MQP User Study

General Questions

* Required

Choose your OS *

☐ Mac OS X

☐ Windows

☐ Linux

☐ iOS

☐ Android

☐ Other : _____

Choose your browser *

☐ Safari (Desktop)

☐ Safari (iOS)

☐ Google Chrome

☐ Mozilla Firefox

☐ Opera

☐ Internet Explorer

☐ Edge

☐ Android Browser

☐ Other : _____

NEXT

33% complete

* Required

Metric Builder User Testing

Metric Builder User Testing

How easy was it to find the Metric Builder? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to select the metrics "Unemployment Rate", "Corporate Income Tax Rate", and "Median Household Income" in the metric builder? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to weigh the metrics accordingly? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to name and save your metric? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to find "My Metric" on the data explorer page? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to edit “My Metric” to make the appropriate changes? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How easy was it to remove “My Metric”? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

Do you have any comments regarding Metric Builder feature?


Your answer

Do you know what an API is? *

- ☐ Yes
- ☐ No

BACK

NEXT

 66% complete

Never submit passwords through Google Forms.

MQP User Study

* Required

API Documentation User Testing

API Documentation User Testing

How easy was it to find the API Documentation? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------|
| Hard | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Easy |

How useful is the API Documentation? *

| | 1 | 2 | 3 | 4 | 5 | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------|
| Poor | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Excellent |

Do you have any comments regarding the API Documentation?

Your answer

BACK

SUBMIT



100%: You made it.

6.2.4 Results

Full user survey results and comments can be found in the Excel file labeled User Results

References

- [1] Big data: The next frontier for innovation, competition, and productivity.
- [2] About mhtc.
- [3] About matters.
- [4] Alex Fortier, Long Hoang Nguyen Duc, Kevin Mee, and Westley Russell. Advancing massachusetts technology, talent and economy reporting system (matters). http://www.wpi.edu/Pubs/E-project/Available/E-project-042915-094249/unrestricted/Advancing_MATTERS.pdf.
- [5] Nilesch Patel, Stefan Gvozdenovic, and Ted Meyer. State cost competitiveness dashboard. http://www.wpi.edu/Pubs/E-project/Available/E-project-050814-222635/unrestricted/Final_report_5.6.2014_FINAL-revisions.pdf.
- [6] Greg Sleter. What's an api and why do you need one?
- [7] Brian Proffitt. What apis are and why they're important.
- [8] Andy McGregor, David Flanders, and Malcolm Ramsey. The advantage of apis.

- [9] Gregory Koberger. The importance of good api documentation: Creating meaningful partnerships for businesses.
- [10] Weighted average.
- [11] Consumer price index - cpi.
- [12] The top 10 economic indicators: What to watch and why.
- [13] Hao Zhong and Zhendong Su. Detecting api documentation errors.
- [14] Get user accounts.