Venetian Open Data

Providing greater access to the Venice Project Center's Data Resources

**An Interactive Qualifying Project Report
submitted to the faculty of
Worcester Polytechnic Institute
in partial fulfillment of the requirements of the
Degree of Bachelor of Science**

Submitted to:
Project Advisor: Fabio Carrera, WPI Associate Teaching Professor
Project Co-Advisor: Michael Aghajanian, WPI Professor

Submitted by:
Ruofan Ding
Jon Paul Gualdarrama
Ryan Horton
Michael Perrone

Date: December 19, 2014
ve14-open@wpi.edu
https://sites.google.com/site/ve14open/home

Authorship

The authors of this report are Ruofan Ding, Jon Paul Gualdarrama, Ryan Sedghi Horton, and Michael Perrone. The specified WPI students were responsible for working on equal parts of this report.

Acknowledgments

First, we would like to thank our project advisors; Professor Fabio Carrera and Professor Michael Aghajanian, both of whom have provided us with more than enough assistance to adequately complete our project in the last two months. Professor Carrera was an invaluable source of information and contact point. Having been a native of Venice, Professor Carrera has many years of experience with dealing with the city and ensuring that our questions and concerns have been addressed with regards to our projects and outreach. Professor Aghajanian was a very useful tool for solidifying our goals and staying organized. With several years of business management, Professor Aghajanian is extremely efficient and professional. He was very beneficial to us and helped guarantee that our goals were feasible and reported in a way that promotes our work for future projects.

We would also like to thank Dr. Federica Bettio. Dr. Bettio was extremely kind to us and went out of her way to make sure that we were granted access to the Venice Open Data website. She also made sure to answer all our questions regarding the publication of data, which allowed us to act with confidence and make sure that our recommendations for future groups are accurate and well researched. While not directly assisting with the creation of our widget or visualizations, Dr. Bettio played an integral role with respect to our project and the Venice Open Data Website.

We would like to extend our gratitude to Benjamin Lichtner, Dylan Kirby, and Kyle Miller for assisting our project team and the entire Venice Project Center with all technical issues revolving around the CK Console and Venipedia. Their assistance helped introduce our teams to the inner workings of the VPC websites and helped to make sure that we would be able to work with the sites without causing accidental damage. They were exceptional sources of information and played critical roles in the usage of these systems by both our team and the rest of the VPC.

Finally, we thank Professor Robert Traver for his excellent teaching style and high degree of professionalism. Professor Traver played a large role in preparing for our arrival in Venice and gave us the tools that we needed to succeed. We offer our thanks and gratitude to the Professor who taught us so much and helped us to create IQP proposals that we could look back on with approval and confidence.

Abstract

The Venice Project Center Open Data team's mission was to provide greater access to the VPC's data resources. This mission had three main objectives: to create a link between the VPC and dati.venezia.it, Venice's Open Data website (VOD), to visualize data from the VOD and add them to the VPC website in order to connect Venice's data to our website, and to combine datasets from both sources to show that datasets, when combined, are more valuable than individually.

To create the link between the two organizations, we updated the dataset on *pili portabandiera* (flagstaff pedestals) and published this data to the VOD. We then used information on the Venice Connected Wi-fi network and the Internet Reporting Information System (IRIS) from dati.venzia.it to create visualizations that were uploaded to the Venice Project Center website in order to complete our second objective. To finalize our project and complete our last objective, we utilized IRIS data and VPC data on fountains to create a map showing IRIS reports pertaining to fountains around Venice.

Executive Summary

The mission of the Venice Project Center (VPC) Open Data Project is to provide greater access to the VPC's data resources. Venice has one main Open Data website: dati.venezia.it, henceforth referred to as the Venice Open Data website or VOD. The VOD contains eighty-five datasets about all aspects of the city. The Venice Project Center also has a large amount of data about the city. Until now, there has never been a link between these two sites, even though the VPC has a longstanding relationship with the Venetian government. We achieved the mission of our project through three objectives:

- Determining the process for uploading VPC data to dati.venezia.it and successfully adding a dataset to it.
- Demonstrating that we can utilize dati.venezia.it and display it in easier to read formats on the VPC.
- Combining our data with dati.venezia.it data as a proof of concept that the pooling of data is more valuable than having two individual data resources.

To determine the method for publishing data to dati.venezia.it, we used the *pili portabandiera* (flagstaff pedestals) dataset as a test case. To make sure the published data is up to date, we were required to recollect flagstaff pedestals information. Using the data from 1997, we were able to find the location of each flagstaff and collect data in various fields while also determining if flagstaffs were added or removed in the last seventeen years. This data was updated on our flagstaff pedestals spreadsheet which was then published on Venice's Open Data website after obtaining the proper permissions. In order to publish this data to the VOD, we had to create a download page for all VPC datasets. This was done because a download link is required to publishing on the VOD. This step was integral in completing our first objective.
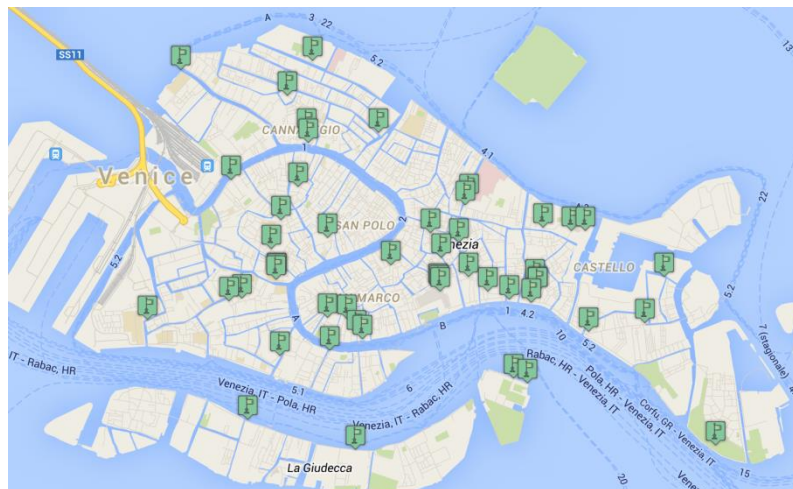

Figure 1: Flagstaff Pedestal Map of Venice

In order to utilize dati.venezia.it data on the VPC website, we created the Citizens'
Complaints Widget for the VPC Dashboard using data from Venice's Internet Reporting
Information System (IRIS). This system is a simple to use website that allows anyone to file a complaint about various issues within the city. Utilizing this data, we created a web scraper which granted us the ability to quickly view changes in complaint statuses. This led to the creation of a live widget located on the Venice Dashboard. It also lead to a number of visualizations that were created to showcase useful data on the rate of complaint resolutions and variety of factors that may play into how long it takes to resolve them.

Figure 2: IRIS Widget

Our final objective was completed by cross-referencing VPC data with IRIS data. To do this, we created a map similar to the widget. However, all points that are a complaint referring to a piece of public art such as fountains are displayed as the appropriate VPC symbol rather than as one of our widget markers. Clicking these markers will allow users to view data on both the collected VPC data and registered IRIS complaint.
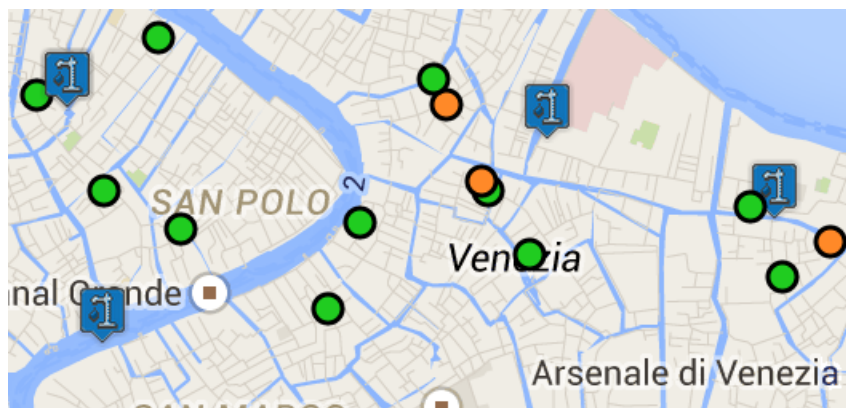

Figure 3: Map of Cross-Referenced Data

Despite the work completed during this project, there is still a long way to go. The foundation for the link between the Venice Open Data site and Venice Project Center has only begun to be laid, and it will take more work to create a strong link between these parallel sites. Future groups can use the information we have provided and criteria created to aid them in fully building the link that will serve as the pipeline between Venice's Open Data website and the Venice Project Center. In addition, our widget and IRIS visualizations are just the tip of the iceberg in terms of what can be accomplished by utilizing this data. Future groups can work to improve Venice's infrastructure maintenance and other issues by using the data analytics we have provided and potentially use similar styles of work to apply our methods to new datasets about the city. By achieving our goals, we hope to have paved the road for many future groups and the Venice Project Center as a whole.

Table of Contents

Table of Figures

Table of Tables

1.0   Introduction

Data plays a very important role in modern society. Everything from producing a song in a music studio to running a city requires extensive usage and analysis of data. The city of Venice is no different, and collecting meaningful information about the city is the mission of Worcester Polytechnic Institute's (WPI) Venice Project Center (VPC). Over the last 25 years, VPC students have been gathering, analyzing, and publishing data on Venice. This year, we have completed the Venetian Open Data Project to create a link between the VPC and dati.venezia.it, Venice's Open Data site (VOD), to allow for easier public access to the VPC's rich database and show how useful data from both organizations can be to the city of Venice.

Figure 4: Overhead View of Venice

While the Venetian government and the VPC have a long history of information sharing, up until this point, dati.venezia.it and the VPC have been working independently to make data about the city available to the public. VOD is the Venetian government's main channel for data publication. Venice also provides access to websites through which citizens can provide information about the city. For example, the Internet Reporting Information System (IRIS), allows citizens to log complaints about urban maintenance issues they find around the city. In parallel, the Venice Project Center has used the Venipedia wiki site since 2007 to allow public access to the VPC's proprietary data.

Keeping this valuable information on two different websites without any link between the two makes research difficult and increases the probability of a user not finding the data they need if they are unaware of the existence of one of these sites. Additionally, separating two significant datasets such as these can potentially lead to reduced efficiency for both organizations. For example, volunteers working in Venice for the *Protezione Civile* (Emergency Management Agency) are currently collecting data on public art, most of which is publicly accessible on the VPC website, as can be seen in Figure 5, on the following page. Obtaining this data from the VPC would be mutually beneficial to both the *Protezione Civile* and the VPC, at the minimum

improving the efficiency of the volunteer's data collection efforts. Thus, creating a link between the data of the Venetian government and that of the VPC would make it easier for everyone from a casual user to a government organization to find useful data from both datasets.
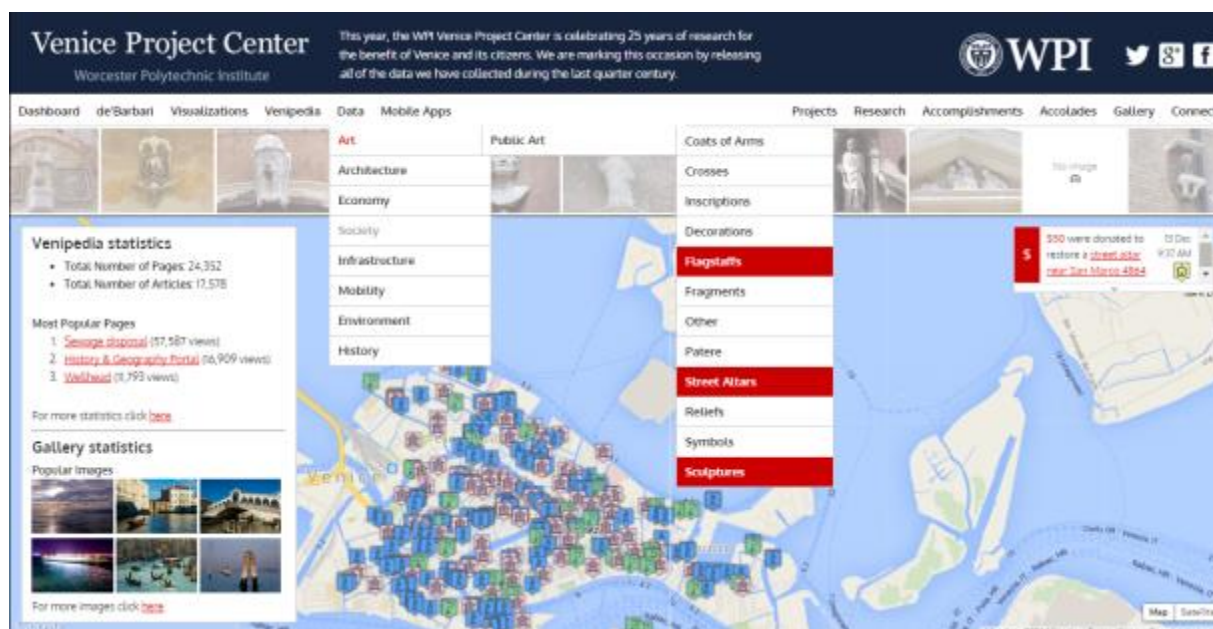
Figure 5: Main Page of the Venice Project Center Website

The mission of this project consists of three objectives: firstly, to increase the availability of VPC data by linking data to the VOD database, secondly, to do the same for VOD data and create visualizations that we can add to the Venice Project Center site, and thirdly, to join datasets data from the VPC and dati.venezia.it in a useful visualization that is more valuable than the individual datasets. In working to meet the three objectives, we considered the tasks necessary for each objective separately. To complete our first objective, we evaluated and updated the VPC dataset on *pili portabandiera* (flagstaff pedestals) and then published it to Venice's Open Data website. To assist future students with the publication process, we also created a list of criteria that can be used to prioritize and select other WPI Venice Project Center datasets for publishing. To complete our second objective, we created a number of visualizations that made use of VOD data and added them to VPC websites. Our third objective required us to combine a VPC dataset with one from Venice's website.

The completion of these tasks resulted in a number of deliverables. These consist of the published flagstaff pedestal data, a widget for the Dashboard page on the VPC website, a download page for VPC datasets, and visualizations of data from the VOD site. These efforts will make VPC data more accessible via the newly created connection to the Venice Open Data

website (VOD), and provide a platform upon which future projects can be built.

2.0 Background

The term "Open Data" was introduced in 1995; however, the principles that characterize "Open Data" were not formulated until 2007. Today, Open Data is defined as "data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike." (Open Knowledge Foundation, 2011) Three fundamental concepts of "open" data as defined by the Open Knowledge Foundation in 2011 are:

- **Availability and Access** – The data must be available in its entirety in a convenient and modifiable form for no more than a reasonable reproduction cost.
- **Reuse and Redistribution** – The data must be allowed to be reused and redistributed as well as mixed with other datasets.
- **Universal Participation** – Data publishers may not discriminate against any user. Restrictions (e.g. corporate use only) are not allowed.

By applying these concepts, entities can ensure that their datasets are accessible to all while encouraging interoperability.

2.1 Economic and Social Benefits of Open Data

The benefits that come from employing these principles are numerous. In particular, Venice can reap both economic and social benefits from opening data to the public. A breakdown of the economic benefits as analyzed by McKinsey Global Institute can be seen in the graph below.

Potential value in open data, $ billion

Education — 890–1,180
Transportation — 720–920
Consumer products — 520–1,470
Electricity — 340–580
Oil and gas — 240–510
Health care[1] — 300–450
Consumer finance — 210–280

The values here are drawn from examples of open-data potential and are not a comprehensive sizing of potential value across the two sectors.

Total — 3,220–5,390

[1]Includes US values only.

Source: McKinsey Global Institute analysis

Figure 6: Estimated Economic Gains in Billions. Solid blue bar represents the minimum potential gain. Dashed blue line represents range of potential gain.

Governments create commercial possibilities for businesses and consumers when they Open Data to the public. (Open Knowledge, 2013) An estimated three trillion dollars[1] of potential economic benefit generated by Open Data initiatives and projects was noted in a recent McKinsey report. (U.S. Open Data Action Plan, 2014) These values were determined through a financial investigation into seven domains. Contributors from the McKinsey Global Institute, McKinsey's Global Public Sector Practice, and McKinsey's Business Technology Office (BTO) chose what they determined to be seven sectors that provide the broadest range possible in both private and public sectors. The seven chosen domains were: Education, Transportation, Consumer Products, Electricity, Oil and Gas, Healthcare[2], and Consumer Finance. The study further examined how much each domain would economically benefit from Open Data through means such as smarter investments, access to information that help view topics from alternate perspectives, and strategies useful for various fields which can help improve the quality of service. For example, within the domain of Education, it was investigated how much the domain would benefit from the improvement of five "levers:" Improved Instruction, Matching Students to Programs, Matching Students to Employment, Transparent Education Financing, and

---

[1] Roughly divided between the United States ($1.1 Trillion), Europe ($900 Billion), and the rest of the world ($1.7 trillion)

[2]Healthcare is only examined within the United States' possible economic benefits

Efficient System Administration.

The full McKinsey Report continues to individually break down each domain to find the root economic gains in each lever of each domain. The result is Figure 6, which can be seen at the beginning of this section. (Dobbs, Manyika, Woetzel, Farrell, & Kuiken, 2013) Open Data applications can create major possibilities for app developers, urban planners, and businessmen to take full advantage of consumer needs in their area. (Marijn Janssen, 2012) In particular, some of the financial and geographical data released through Open Data initiatives have helped smaller sized enterprises find much needed loans.

Open Data also allows individuals and governments to reap social benefits. The increased availability of information and overall increased ease of access create a more equal opportunity for success. For example, some datasets allow both residents and business owners to determine which neighborhoods are safest. Other datasets enable residents to easily find the services they need and enable businesses to quickly find potential customers. Tourists can use Open Data to manage their trips in advance and utilize small applications called widgets that provide real-time information, such as traffic levels and bus schedules. Cities have even hosted competitions for developers to make use of newly released datasets. (Gurin, 2014) These benefits exemplify the opportunity of Open Data.

2.2 Barriers to Open Data

Despite all the potential benefits of government Open Data, there are many barriers that can hinder its success. The most significant of these are low data quality, nonexistent data standards, and a lack of user feedback. There is very little quality assurance that surrounds public data. As a result, many datasets are incomplete, inaccurate, or out of date. Additionally, the lack of Open Data standards results in ill-defined formats, non-centralized distribution, inconsistent metadata, and difficult dataset searching. All these inconsistencies make it more difficult for consumers to use. Finally, there is rarely any feedback from the users. Often data publishers measure success by the number of datasets they publish and not by the way their data is utilized. This leads to a less successful product for publishers and a less productive experience for data consumers. (Marijn Janssen, 2012)

Privacy is a major non-technical barrier to the success of Open Data. Concerns about privacy are common to data publishers, individuals, and governments alike. Potential data publishers fear that they could expose the individuals who are represented in their data to

invasions of privacy. For example, Venice's Open Data Website has datasets that record the gas usage of approximately 130,000 residential and government properties within the city. In spite of the fact that this data could be valuable for studies and efficiency improvement, citizens, businesses, and governmental agencies, might not want their gas usage available for public review. This may be due to the fear of reputational damage. When the data regarding usage (i.e. gas usage) is released, it has the possibility to cause consumers to worry that their data will be tied to them and therefore harm their economic or social standing.

Hypothetically, through this same data release, credit card companies can raise interest rates on households now that they can make a much more detailed decision on how much a house is wasting with regards to electricity and gas. In terms of businesses, Open Data can expose poor environmental or labor practices by companies which would certainly damage the companies' reputation and revenue. Companies can also damage themselves by using Open Data about their consumers to create advertisements that show that the company knows too much about their consumer. (Dobbs, Manyika, Woetzel, Farrell, & Kuiken, 2013) When a government releases data, it must consider the privacy of its citizens as well as its own security.

2.3 Examples of Model Open Data Websites

Despite the barriers to their success, the socioeconomic benefits of Open Data have led cities and nations around the world to start their own Open Data websites. While a number of these sites feature very basic capabilities, other sites are ahead of the pack in their functionality and interactivity with their users. Two examples of such sites are Boston's Open Data site data.cityofboston.gov, and the European Union's site, open-data.europa.eu.

Boston has over 500 datasets and views in the form of maps, charts, calendars, and raw data. One can sort them by popularity and filter based on search terms. Datasets can be sorted by date of publication, date of update, and number of times accessed. It is very easy for all types of users to access the data. Users can view the data straight in the browser, and if the data cannot be viewed directly, a link to a page or an application where the data can be viewed is provided. Many datasets can be exported in a variety of formats depending on the layout of the data in the source file.

The European Union Open Data Portal website provides a good example from which

Venice's site could learn. It is very easy for technical users to access the data.[3] Even for non-technical users, it is easy to access data via applications and visualizations on the website. (European Union Open Data Portal, 2013) The homepage features lists of the most recently updated datasets, most popular datasets, top publishers, and most popular search terms. The datasets can be searched with keywords or by exact titles. In addition, the metadata can be queried to obtain general information, such as the number of datasets per publisher. Users can also interact with the administration of the site by requesting datasets that do not exist.

It is clear that with their variety of features, both of these websites exemplify the three principles of Open Data. Boston offers downloads in multiple data formats, and the European Union even offers multilingual datasets. Both sites allow users to use and redistribute their data freely, and neither imposes any restriction on who can download their data. In these ways, Boston and the EU's Open Data websites fulfill the guiding principles of Open Data, making them excellent examples from which the VOD can learn.

2.4 Venice Open Data

Venice's Open Data is grounded in its website: dati.venezia.it. This site serves as the main connection point between data publishers and parties interested in using the published data. The Internet Reporting Information System (IRIS) website serves as a tool for citizens to communicate with their government, providing the city with feedback on urban maintenance and cleanliness issues. That data is exposed on dati.venezia.it. These two sites together form the basis for the current state of Venice's Open Data Project.

2.4.1 Venice Open Data (dati.venezia.it)

Venice's Open Data Website, seen in Figure 7 below, serves as the vehicle through which the Venetian government can share data about the city with the world. The design of the site centers on the ability to find and download a dataset. One can directly view the list of datasets or search through it for a particular dataset. Each dataset also has a dedicated page with its description and a list of pertinent information, discussed in §**3.1.2**. Additionally, each dataset's copyright must meet the three principles of Open Data, particularly the third principle requiring universal accessibility to the data.

---

[3] The site has RESTful APIs, as well as a query engine in SPARQL to access the data in RDF. It also provides SDKs for a number of languages.

Figure 7: Dati.venezia.it Homepage

Despite the seemingly simple design of Venice's Open Data Website, it has a number of issues, including the aforementioned Italian-only limitation and the difficulty of accessing data. Interested parties that do not speak Italian have to consider the time and resources necessary to translate datasets into their language before utilizing them. Furthermore, much of the information is scattered and hard to access. Many datasets on Venice's Open Data Website lack explanations for the column headers. In addition, when compared to Open Data sites such as those of Boston and the European Union, Venice's Open Data website lacks many of the features which make those sites successful and useful to those interested in utilizing its data. Nevertheless, what these characteristics reveal is that Venice's Open Data Website, though basic, strives to be a true Open Data source.

2.4.2 Internet Reporting Information System

Since the IRIS website's creation in May 2008, 78% of the nearly 22,000 submitted reports have been resolved.[4] It is evident that many Venetian citizens make use of the site to voice their complaints about problems they see around the city. These complaints vary from issues with public lighting to animal waste. Additionally, the high rate of resolution of reports shows that the city takes the reports seriously and actively monitors the site. Thus, the information contained on the site is valuable to Venice's citizens and its government, making it an ideal resource for those interested in analyzing data about Venice.

Venice's IRIS website serves as a rapid method for reporting damages and other various complaints to the city. The website, which can be seen below, has a much more user friendly

---

[4] As of November 17, 2014

interface compared to the Venice Open Data website seen above. Immediately upon navigating to the website, users have three options, "Fai una segnalazione," "Vedi le segnalazioni," and "Verifica una segnalazione," or "Make a report," "See reports," and "Verify a report." These options are displayed using unique icons, colors, and large text to make them stand out from the rest of the page and simple for users to decide which service they require.
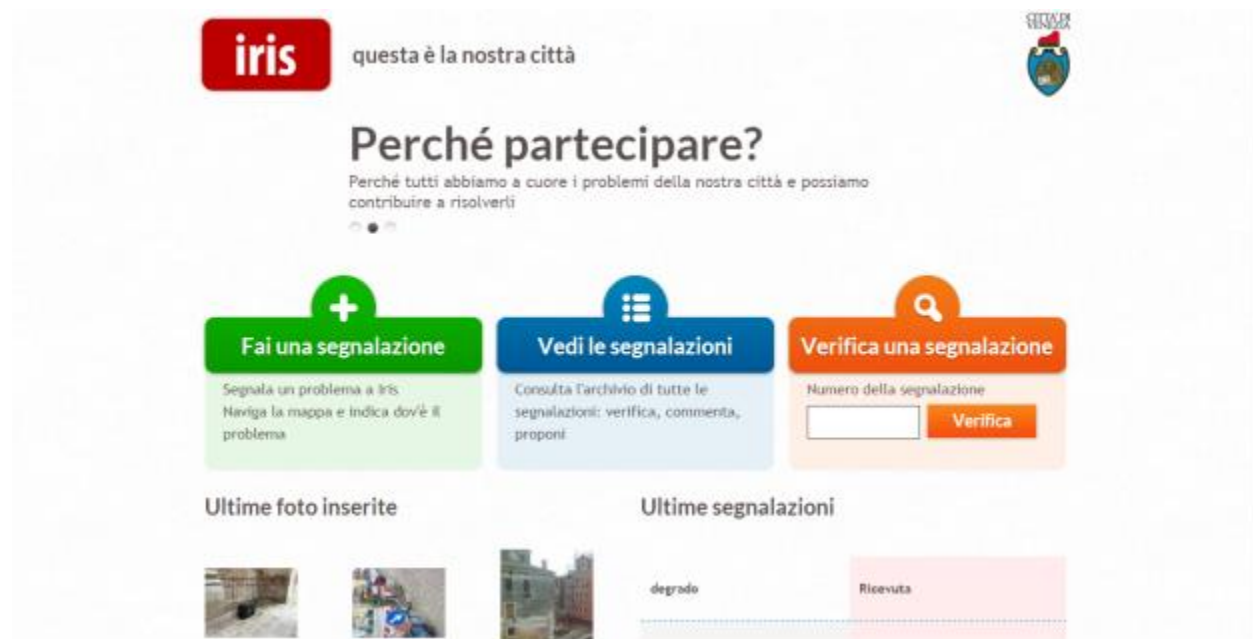
The "Make a report" option redirect users to easy-to-use pages where they can input information about a complaint they would like to report or add information to an existing report. Similarly, clicking "Verify a report" brings users to a page containing information about a report including the location reported on, the current status of the report, and other pertinent information.

We utilized the second option, "See reports." This option redirects users to a page where users can perform an advanced search for up to two hundred and forty datasets at a time. It also includes a link to a statistics page where the number of received and resolved complaints and a breakdown of complaints into categories can be found.

The IRIS website also features a create map function that allows users to view a map of up to two hundred and forty complaints at a time. While functional, the function does not work for complaint categories that have more than 240 reports. This means that 45% of the complaint

categories cannot be displayed in their entirety. Therefore, while the IRIS site has many features that make it easy to understand and use, it there are opportunities for improvement with regards to visualizing the data.

3.0 Methodology

The Open Data project had three main objectives:

1. Add VPC data to Venice's Open Data website (VOD) to create a permanent link between the two sites.
2. Produce visualizations of data from dati.venezia.it for use on the VPC website
3. Merge data from the VPC and the VOD to show that datasets together have more value than individually

Objectives One and Two can be visualized as a cycle in which data from the VPC is added to Venice's site and information from the VOD is utilized to create visualizations on the Venice Project Center site. This cycle can be seen in Figure 9 below.
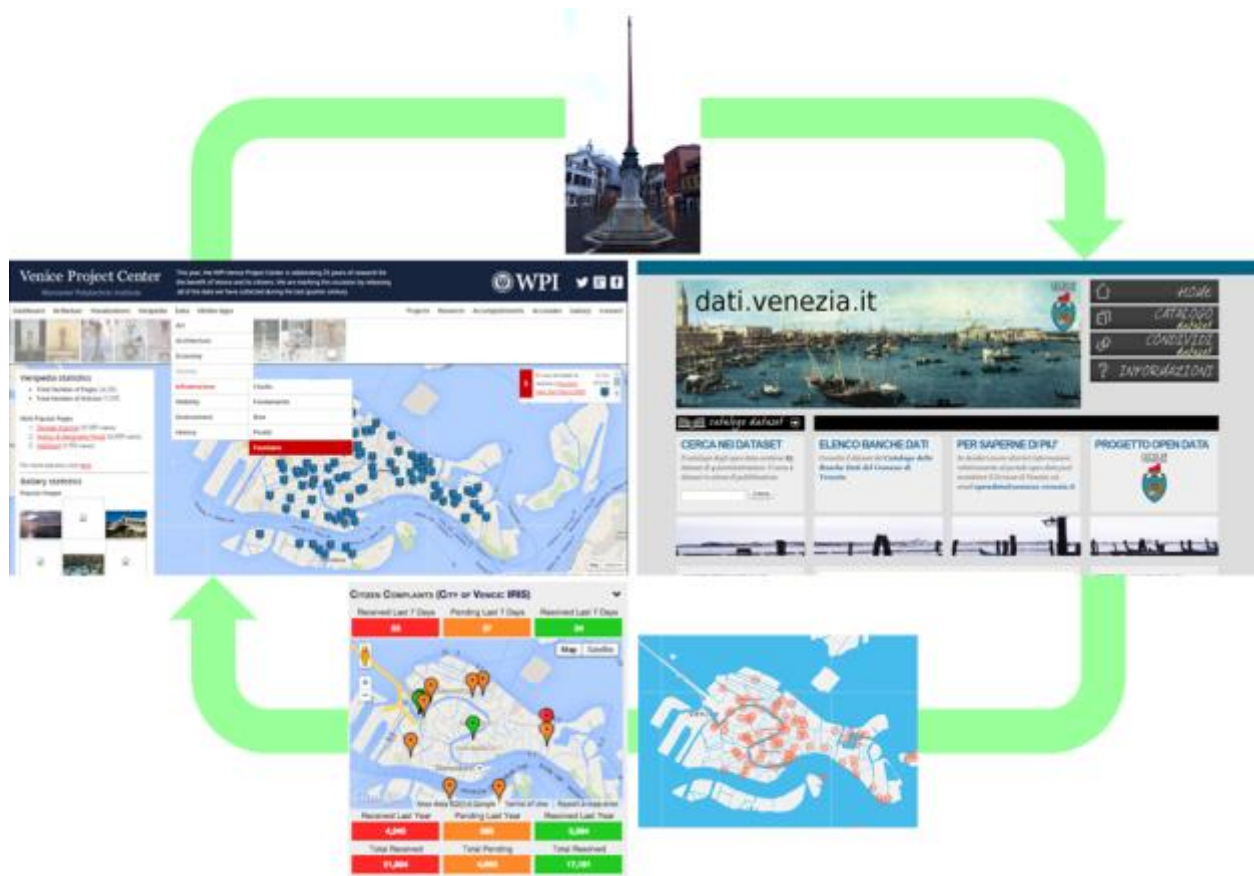
Figure 9: Open Data Project Cycle

3.1 Adding Venice Project Center Data to dati.venezia.it

Completing our first objective began with choosing a dataset to add to the VOD. Since the VPC is the first non-governmental agency to attempt to publish data to Venice's Open Data site, we were focused on building the connection between these two sites and decided to choose a dataset that would be relatively simple to update and therefore could be quickly updated to the

VOD once the connection had been established. For this reason, we chose to update the VPC dataset on *pili portabandiera* (flagstaff pedestals). Having chosen the dataset to work on, three main tasks were developed to complete the objective of adding Venice Project Center data to the VOD:

1. Update flagstaff pedestal data
2. Create a download page for Venice Project Center data
3. Publish the dataset to Venice's Open Data website, providing a URL to the corresponding download page


Figure 10: Contributing to the Venice Open Data Initiative

3.1.1 Update Flagstaff Pedestal Data

To update the flagstaff pedestal dataset, we analyzed and reviewed the existing flagstaff data in the VPC database to assess its level of completeness. The completeness criterion included both information on which flagstaff pedestals were not present in the dataset and which details were missing. Additionally, because the original flagstaff pedestal data was collected before the online VPC database was created, we compared the dataset found in the VPC database with the original files from the 1997 Interactive Qualifying Project (IQP), "A Computerized Catalog of Flagstaff Pedestals in Venice," to ensure that the transfer from the data file to the VPC database was without error. We determined that the flagstaff pedestal data in the VPC dataset was accurate and matched the original data collection efforts.

Before updating the data, we chose fields in the dataset that we would update for each flagstaff. Many of the existing fields, such as pedestal inscription and historical notes, were
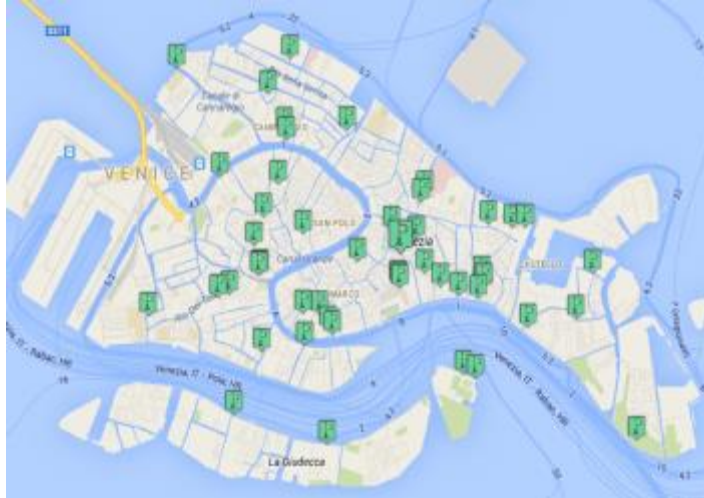
Figure 11:Flagstaff Pedestal Map of Venice

immutable and complete and therefore could be overlooked for updating, but some of these immutable fields, such as latitude and longitude, were not complete, making them prime candidates for data collection. Mutable fields, such as flagstaff color, the presence of a flagstaff on the pedestal, and whether or not the flagstaff was in use, could be easily updated. We selected a subset of mutable and immutable but incomplete fields to update that would give an overall picture of the current condition of the flagstaff pedestal; the list of selected fields can be found in Appendix B.

Using these fields, we went around Venice to collect the data. Generally, we could easily locate flagstaff pedestals based on their latitude and longitude or, in the case of pedestals without coordinates, based on their address. At each pedestal, we collected the data to fill in the selected fields, correcting any outdated information. After the dataset had been updated, it was published to Venice Project Center's CK Console and Venipedia.

3.1.2 VPC Data Download Page

Although there are many useful datasets on the VPC website, previously the only way to retrieve the raw data was through the CK Console. The CK Console is a power saw, and for this job, an everyday user just needs scissors. Furthermore, access to the CK Console is limited to VPC developers. To address this issue we developed a download page for VPC data, eliminating the need to go to the CK Console to retrieve individual datasets. The page can be used internally by the VPC or by outside organizations, making it an invaluable resource.

3.1.2.1 Implementing the Download Web Page

The VPC's raw data is stored in Firebase, a system which provides a real-time database

and backend. All data is stored as JSON objects in a tree structure. The database tree has nodes called "groups," "data," etc. The "groups" node contains the basic information for each dataset and has an ID for each tuple, or row, of the dataset. Tuple IDs can be used as a key to retrieve a certain tuple from the dataset contained in the "data" node of the tree.

We created a single download page for users to download datasets in comma separated value (CSV) format. This web page lists all the datasets available to the public. Administrators of the website can decide which datasets can be downloaded by changing the metadata file. The metadata file is in JSON format, and contains an array of information for the datasets to be published. The information includes CK Console group name, dataset title, and a brief introduction to the dataset. When a user clicks the download button, the web app retrieves all the tuple IDs based on the dataset name from "groups." Then, it uses the tuple IDs to retrieve the tuples, and join all of the tuples into a single CSV file. The browser then downloads the CSV file.

In addition, we also noticed that CK Console does not show download progress when a user downloads a dataset. Because downloading some CSV files can take more than one minute, without seeing the download progress, a user may consider that the CK Console has failed to download the file while it is actually still retrieving the data. In this case, the user is likely to leave our website. To avoid that, we added a progress bar feature to our download page. When a user clicks the download button, a progress bar will pop over and show the percentage of data has been retrieved, which will help users estimate the remaining time for downloading the file.

To summarize, the download page can be used by the public to download raw data from VPC, bypassing the CK Console. It also serves as an interface between the Venice Project Center and Venice's Open Data Website (VOD). The download link provided to the VOD for VPC data leads users to our download page.

3.1.2.2 VPC Dataset Download

On the Venice Project Center Website users can see different kinds of datasets under the "Data" drop-down menu and view the geographic information of each dataset on the interactive map. However, prior to this project, there was not an easy way for users to download the raw data for each dataset from VPC website. Therefore, we added a download icon next to each dataset. When a user clicks the download icon, the corresponding download page for that dataset opens.

3.2 Visualizing dati.venezia.it Data

Our second objective required the selection of datasets from Venice's website for which we could create effective visualizations. We chose to select one dataset which would lend itself to a simple visualization and one which required more data processing. Searching the VOD's database, we selected the dataset on Venice Connected Wi-fi Hotspots and the dataset from the Internet Reporting Information System (IRIS). Based on these two datasets, we established two tasks:

1. Produce a map-based visualization of the Venice Connected Wi-fi Hotspot Data
2. Develop visualizations for IRIS data

3.2.1 Map Visualization of Wi-fi Hotspots

The Venice Connected Wi-fi Network map was created with Google Maps API. We overlaid transparent circles over the latitudes and longitudes provided by Venice's Open Data Website. This allowed us to create a map that shows where users should go in order to be able to connect to Venice's city-wide Wi-fi network.

3.2.2 Citizens' Complaints Widget

This widget parses data from IRIS and displays it in an easy to read format from which users can extract useful information. It contains statistics on reports received and completed and displays a map of complaints received, pending, and closed in the last week.

In the top right of the widget, there is an arrow. This arrow points to a google doc where information about using the widget can be found. The blue text stating "City of Venice: IRIS" serves as a link to the IRIS statistics website. The widget displays statistics regarding how many complaints were received (Total, Last Week, and Last Year), as well as how many complaints are still pending and how many have been resolved. This provides a live status report of how many new reports are being made, how many old reports are being solved, and where they are in the city.

This portion of the project was completed in two steps:

1. **Gathering report data**: IRIS exposes a CSV file of all the reports that have been made. Unfortunately, it does not provide any kind of historical data about the report. The one piece of historical data that is most important is the date that a report is closed. IRIS does, however, show this data on the individual pages for each report. In order to get this data, we built a web crawler and scraper that emulates a browser to collect that data out of the HTML page. It then joins that data with the data from the CSV, and stores it in Firebase. The scraper now runs

twice a day to fetch the updated data.

2. **Creation of the widget**: Finally, the widget itself needed to be created in two parts. The first part was to design the layout of the widget. This was modelled off of other dashboard widgets. The second part was to code the widget by hooking into the data gathered from step 1. It uses Firebase to access the data and Google Maps API to create the map.

This widget was integrated into the VPC Dashboard. Refer to §4.3.1, Figure 20 to see the final results of this.

3.2.2.1 Visualizing IRIS Data

Using primarily the JavaScript library d3.js, we were also able to create many graphics to help display the IRIS and VOD data in an easily understandable format. Two examples of this are shown in Figures 12 and 13, The plot on the left shows the time that has passed since each report was opened, while the plot on the right shows the time to close reports by region of Venice. Further details and analysis of both graphs can be found in §4.3.2.
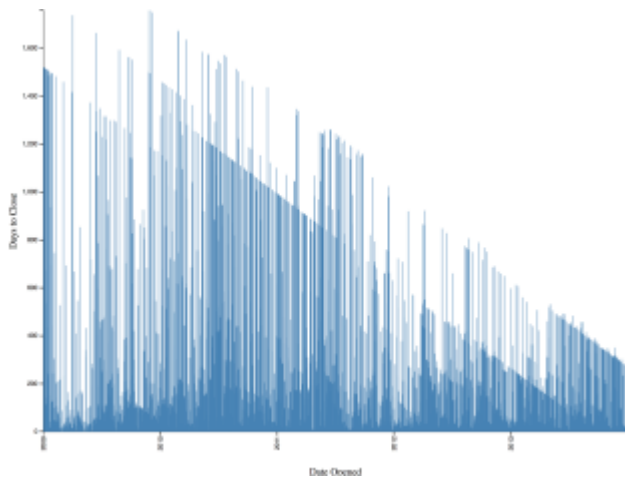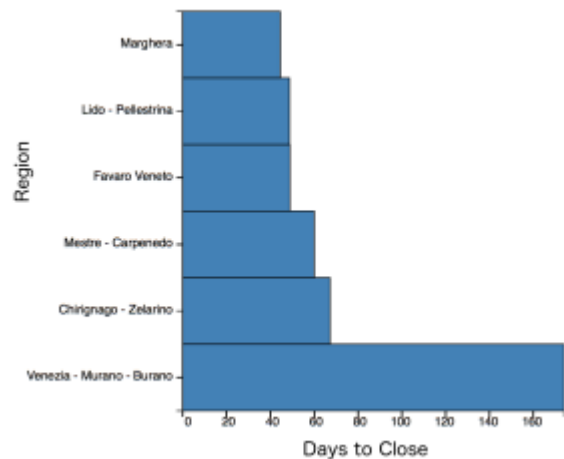

Figure 12: Days to Close Report vs Year

Figure 13: Region vs Days to Close Report

3.3 Combining IRIS and VPC Data

We combined IRIS data with VPC data to produce something better than either independently in order to execute our final objective. Specifically, we combined VPC fountain data with IRIS complaint data. If there were any complaints that had the word *fontana* (fountain) in the description and also had a location within 10 meters of a fountain in the VPC data set, we marked that complaint, the assumption being that the complaint refers to that fountain. We then graphed these complaints again and denoted the marked complaints with the fountain icon. The final result of this can be seen in §4.4, Figure #23. The combination of these datasets created something more valuable than any one of them independently could provide.

4.0 Results & Analysis

Through our correspondence with our city liaison, Dr. Federica Bettio, we were able to develop a plan for continuing to expand the WPI Venice Project Center (VPC) and integrate it with the Venice Open Data website (VOD). This chapter discusses the results of our research and our analysis of the work completed. Our results consisted of two halves of work. The first half was to upload our flagstaff pedestal data to dati.venezia.it in order to prove the possibility of publishing VPC data on official city websites. The second half of our results were to create visualizations and widgets on VPC websites based off data taken from the City of Venice.

Figure 13: Open Data Project Cycle

4.1 Analysis of Collected Flagstaff Data

Through the collecting and updating of data of Venetian flagstaffs, we were able to publish VPC data on dati.venezia.it. This is significant not only because it marks an expansion of VPC data onto other websites, but it is also the first dataset to be published by an entity external to the Venetian government on the VOD. As can be seen in Figure 15, on the following page, we successfully published the flagstaff pedestal dataset to dati.venezia.it, establishing a link between their site and the VPC.

4.1.1 Results of Publishing Data to dati.venezia.it

While finding each flagstaff and the city and recording new data regarding each site posed no major challenge, publishing the data to the VOD had many administrative obstacles. This created many issues due to the fact that it lessened our ability to publish multiple datasets and could very likely discourage others from trying to publish data on dati.venezia.it. We found that if the City of Venice can streamline the process for third parties to apply for data publishing rights, it would allow for much more data to be released about the city and create a large repository of valuable information on the VOD site. With all this said, it can be noted that future VPC teams will not have to go through the same difficulties since we have completed this process already.

4.1.2 Analysis of the Flagstaff Pedestal Data

The flagstaff pedestal dataset contains information on 56 pedestals around Venice proper. From our data collection, we were able to determine that almost 70% of pedestals are in a state of disuse and do not have flagstaffs, as shown on the following page in Figure 16. Additionally, of the thirty-nine that do have flagstaffs, only 15%, a total of six flagstaffs, are currently in use. Since the data was originally collected almost two decades ago in 1997, two pedestals have lost the flagstaffs they once had. At the same time, however, five flagstaffs have been installed on previously unused pedestals. This information has been tabulated in Table 2 below.
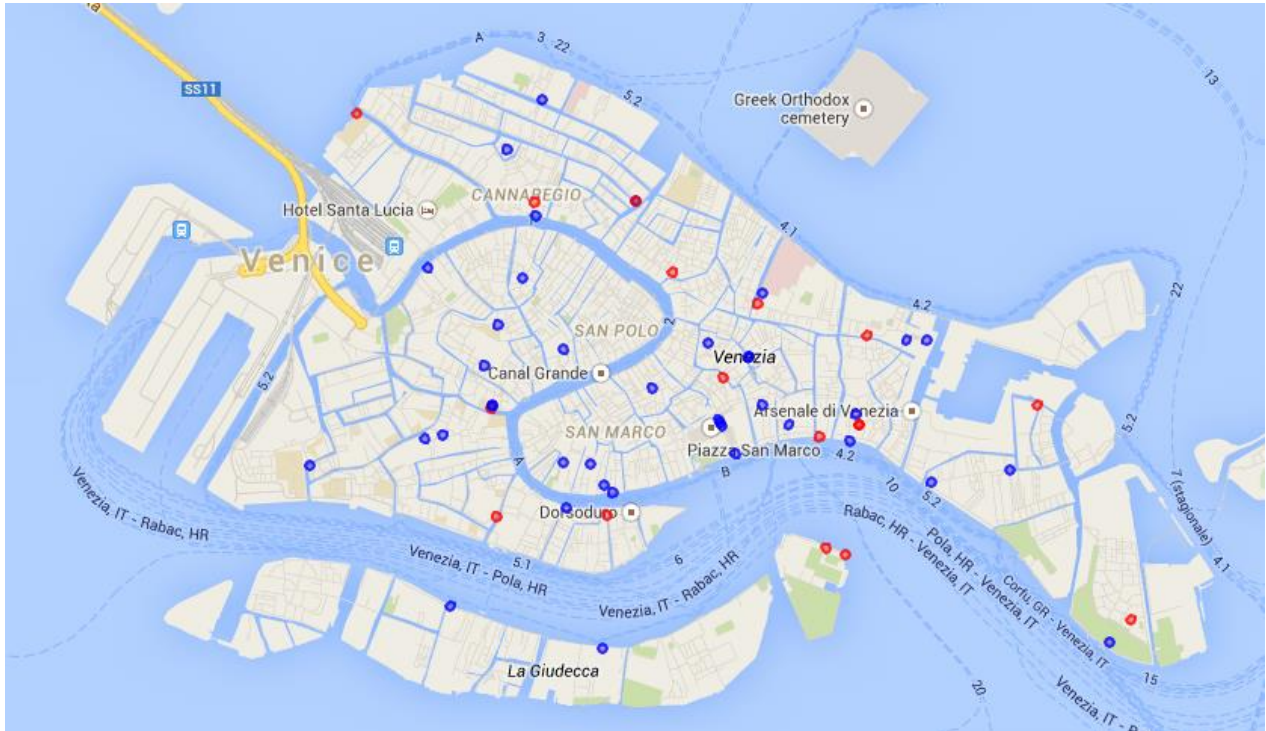
Figure 15: Flagstaff Pedestal Map of Venice Created by the Open Data Team

Table 1: Flagstaff Pedestal Data Analysis

Total	56
Have flagstaffs	39
Have lost flagstaffs	2
Have gained flagstaffs	5
In use	6

4.2 Venice Project Center Download Page

The publishing of our updated flagstaff data not only led to the creation of the VPC download page in terms of the webpage itself, but also in the additional feature for the Venice Project Center homepage. We successfully created this user-friendly download page. It possesses

all the functionalities as we expected, including posting the datasets for download, allowing users to download a dataset in CSV format, search functionality, and analytics. As an interface of Venice Project Center Data, the download page makes VPC raw dataset available to the world, under the Creative Common Attribution-ShareAlike 4.0 International License. Based on the principles of Open Data, published VPC datasets can be used for research purpose as well as commercial purpose.

Figure 16: VPC Download Page

Figure 17: Download Feature on VPC Homepage

4.3 Visualizations of dati.venezia.it Data

      The Venice Connected Wi-fi Hotspot data was utilized to create a map of all areas on Venice proper where users can access wireless internet. Because the data was readily accessible through Venice's website, creating the map was relatively simple. Previous knowledge about the Google Maps API facilitated the process of creating the map. The result from creating this map was that it proved that it was possible to take data from the city and display it in a meaningful and easy-to-read manner. This proof is what allowed us to move forward and create the IRIS widget.
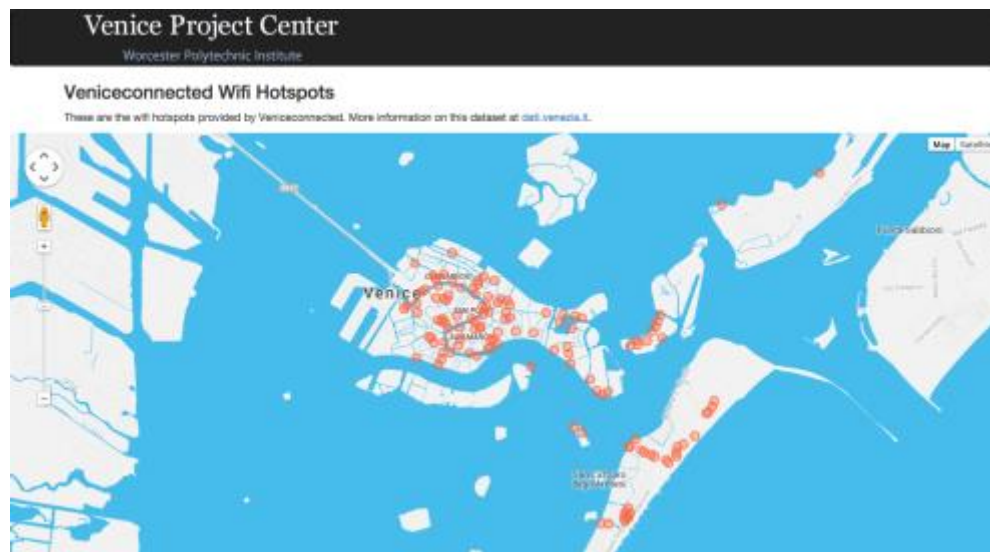

Figure 18: Generated Venice Connected Hotspot Map

4.3.1 Citizens' Complaints Widget

      As another task for completing one of our three objectives, we successfully created multiple visualizations to show VOD data in a more user-friendly way. We found many issues with the way that data is stored on dati.venezia.it. First of all, the data is stored in codified files which are difficult to read. Besides this, some datasets are missing critical information needed for a full data analysis. An example of this is Venice's Internet Reporting Information System (IRIS) which lacks the date upon which the complaints were resolved. This makes determining the time to resolve a complaint much more difficult than the work it would have taken to include that portion of information in the overall dataset. Regardless of this issue, we jumped through several hoops, and did get our hands on that data. Finally, we worked to prove the concept that data can be taken from other city sources and displayed in simple manners on the VPC website or Dashboard, seen below.
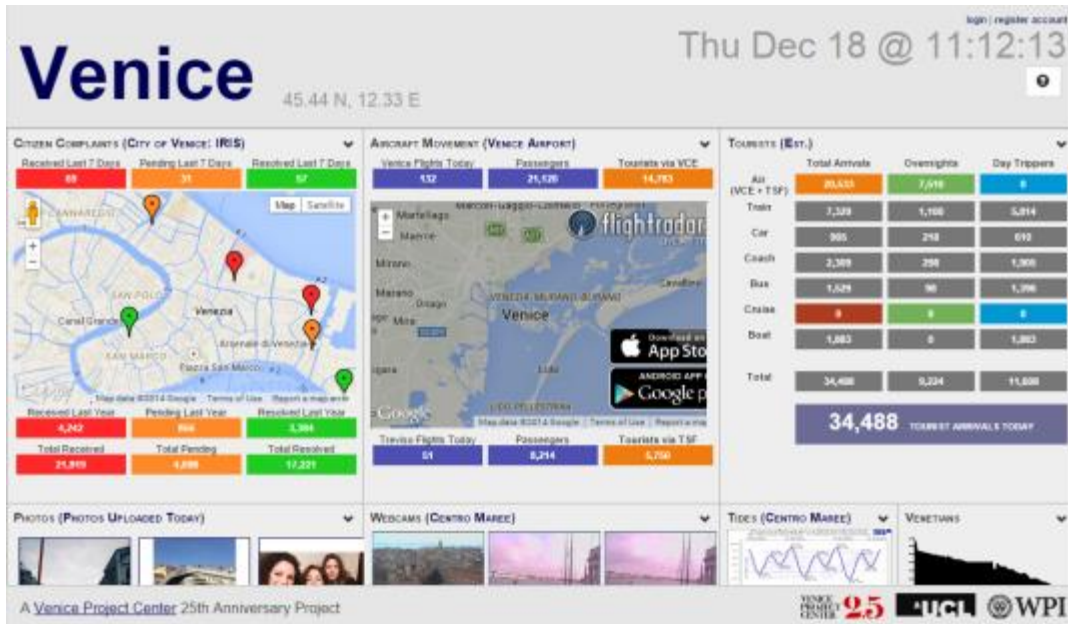
Figure 19: IRIS Widget on the VPC Dashboard

An image of the widget created can be seen, in more detail, below. The widget takes information from the IRIS website, including information that has to be calculated through alternate means and displays it on the widget.

Figure 20: IRIS Widget

4.3.2 Other Visualizations of IRIS Data

The two plots below in Figures 21 and 22 are enlarged versions of the examples shown in §3.2.2.1 in Figures #12 and 13. The plot in Figure #21 shows the time that has passed since each report was opened, grouping reports by the date on which they were opened. The most noticeable feature of the plot is that it trends downward. At first glance one might say that the time to close reports is decreasing over time. This is not the case. If we consider reports that were opened one week ago, the average time to close these reports must be less than one week, as anything taking longer wouldn't be closed yet.

In Figure #22, the plot shows the time to close reports by areas of the city of Venice. Notice that it takes longer to close the complaints on the island of Venice than in Marghera on the mainland. One reason this is the case might be that it is more difficult to get the proper service team to the location on the islands because there are no cars. It is difficult to know if this is correct or not. Further analysis on the IRIS data was not performed because the focus of our project was to link VPC and VOD data.
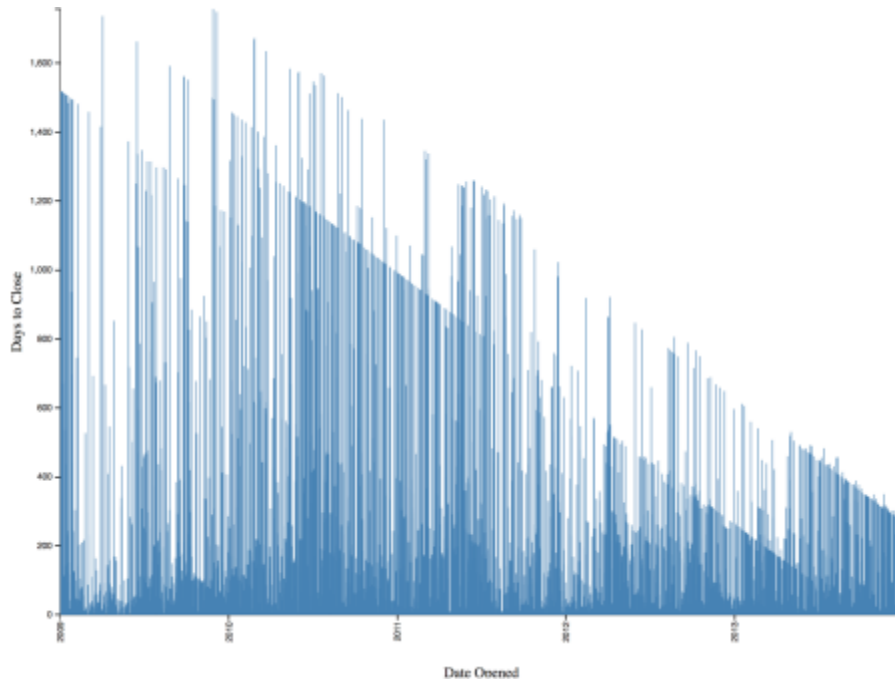

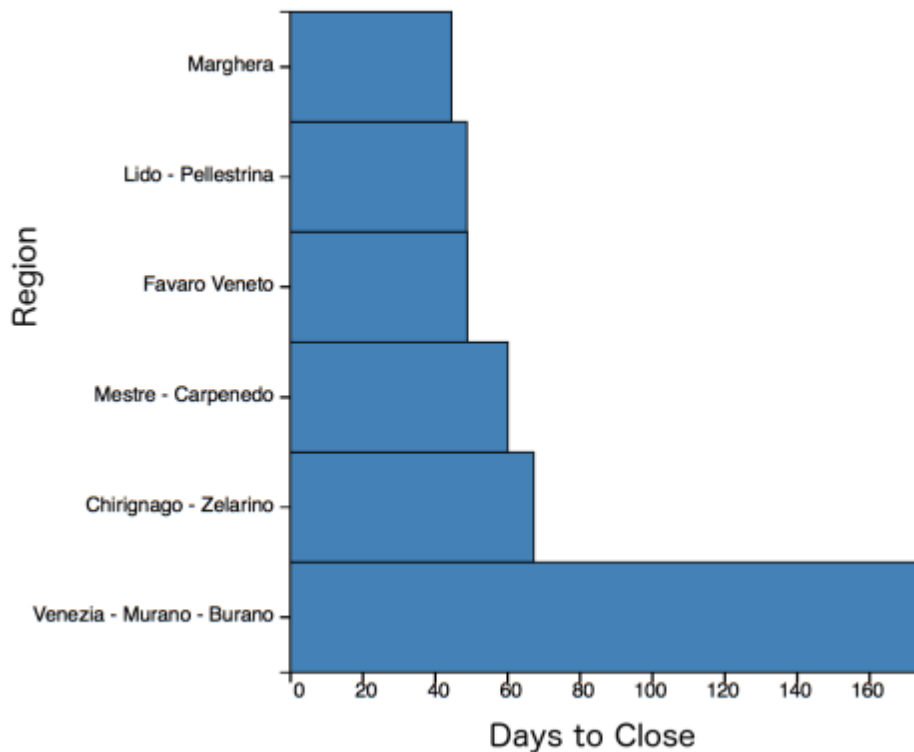**Figure 21: Days to Close Report vs Year**

4.4 Cross-reference of VPC and VOD Data

We found it important that as a final addition to our project we should cross-reference VPC data with IRIS data. To do this, we created a map similar to the widget which displays complaints. In addition, the map displays the appropriate VPC symbol rather than as one of our widget markers if it is found that a complaint is regarding an item that the VPC has data on. Clicking these markers will allow users to view data on both the collected VPC data and registered IRIS complaint.

This map, shown in Figure 23, is just one example of what can be done when combining data from different sources. This combination of VPC and IRIS data is a proof of concept that information from these two sources together is more valuable than either dataset individually. Further groups can look for other datasets that could be cross-referenced or combined with VPC data to create other visualizations or useful tools for future VPC students and outside organizations alike.
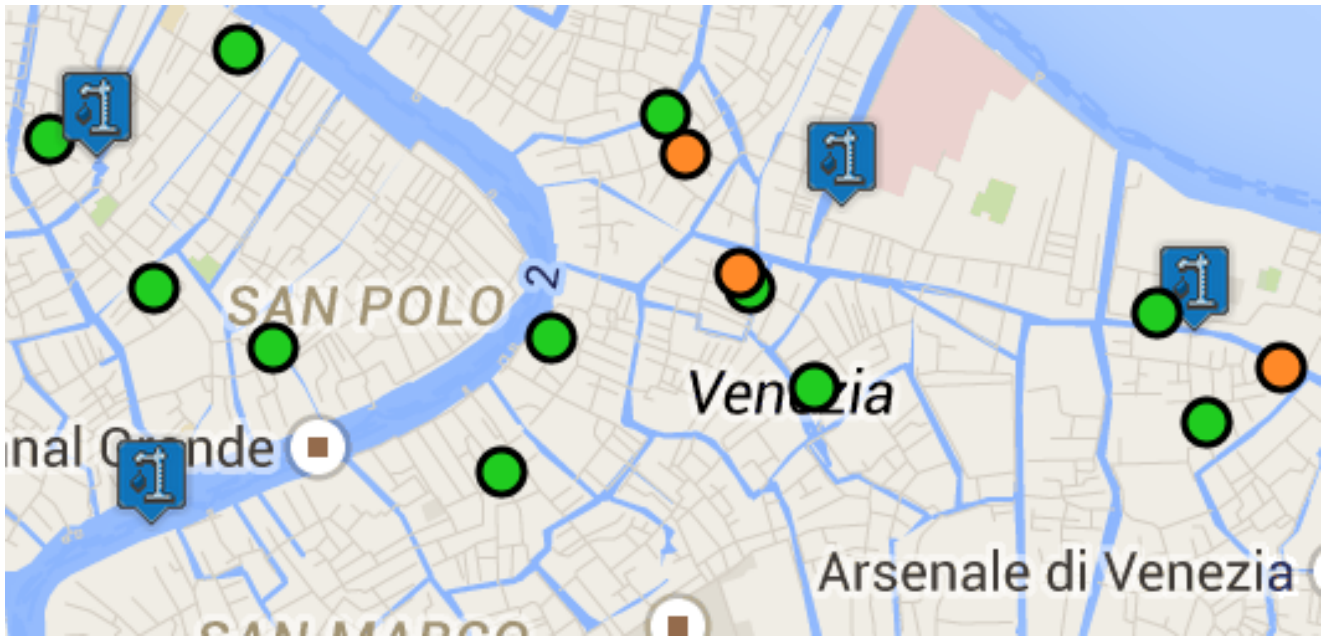
Figure 23: Map of Cross-Referenced Data

5.0 Recommendations

The work we have completed has led us to recognize some recommendations for the VOD and for future VPC work. These recommendations are intended to suggest ways in which both agencies can continue to increase the accessibility and usability of their respective websites, as well as the openness of their data. In addition, we have concluded that an explanation of the procedures involved in using the VOD, working with the CK Console, and utilizing our deliverables should be provided for future groups to facilitate their work. This section will cover the usage of Venice's Open Data website, the process for data publication, and the criteria that should be used in order to determine the optimal choices for data updating and publication.

5.1 Venice's Open Data System

Originally, publishing information to the VOD was a difficult and long process. This is undesirable and can lead to much less information being published than the amount that would otherwise be published. A reason for this is that the VOD uses a platform similar to a blog that is not simple to go through. By changing to another system such as Socrata, the VOD can fix both issues. Socrata is the platform used by websites such as Boston's and the European Union's Open Data Websites and allows for a much user-friendly experience. Users can just create a username and password and begin their work with the website.

5.1.1 Recommendations to Improve dati.venezia.it

Venice can look to Boston and the EU as good sources of information on how to further improve their own site. Boston, like the European Union, uses the platform Socrata to run their website. Socrata is talked about above and is, by far, one of the best aspects of both these sites. It allows for a much easier user interface and better layout than the blog-style of the VOD. In addition, Socrata makes it easier to search and query the data in order for users to find exactly what they want.

As for the European Union's website, they offer services such as the "Developer's Corner." This allows users to assist in building on the data and therefore improving the quality of the site. The EU also offers a variety of applications that can assist in better viewing their data which makes it more desirable to users. Additionally, the EU site offers datasets in a number of languages. In contrast, the VOD's data is available only in Italian. The resources required to translate into multiple languages is significant, but if the VOD published their data in a more

36

widely spoken language, such as English,[5] it would be much easier for users worldwide to make use of the data.

It is possible for the VOD to become better than Boston and the EU; however, we then have to take into account the feasibility of such an effort. Naturally, as Boston is seen as one of the best-in-class sites, it is difficult to pick out what specifically would make it better. Therefore it can be said that the most obvious improvement would be further streamlining of site processes, correction of website errors, and the continuous uploading and updating of data. For the VOD to make a transition from their current platform to Socrata and then increase the number and quality of datasets to match those of Boston and the EU, it would take a tremendous effort. Understandably, this would take a lot of resources with regards to financing and programming.

5.2 Future WPI Projects

Much of the work that we did prior to beginning our project work in Venice was centered around building the relationship between the VPC and VOD and developing a methodology for publishing VPC datasets to the VOD. Because we focused on publishing only the flagstaff pedestal dataset, much of this work was not utilized during our time in Venice. However, VPC students in future years can make use of the work we accomplished to facilitate further work with the VOD, and for this reason, we have detailed below the process involved in obtaining access to the VOD and preparing data for publication, followed by the selection criteria we developed for VPC datasets and how to use them.

5.2.1 Preparing to Publish Data to the VOD

Publishing the data consists of obtaining approval for those who will be working with the VOD, updating the datasets to meet metadata standards, and submitting it to the Venice website administration for publication. The approving agency is the VOD. One of their requirements is that we complete the "Operating Agreement for the Publication of 'Open Data.'" This process can be completed by working with the site advisor as well as members of the Venice website administration that they are referred to.

Upon getting approval, the team will receive information on their publication standards and can begin working to meet these standards by completing dataset modifications and writing the necessary metadata. Necessary modifications will include translation of the dataset and its

---

[5] This is examined in terms of countries that utilize English rather than the just those where English is the native language.

pertinent information into Italian, as well as dataset specific actions, such as linking images to the file. The metadata needs to contain a number of fields about the data including:

- Detailed description of the dataset
- Applicable licensing
- Link to the data
- Format
- Granularity (level of data processing)
- Dates uploaded and last updated
- Managing authority and link to their site
- Administrative level and organizational unit
- Contact persons and their email addresses
- Geographic origin of data
- Subject area and search tags

5.2.2 Dataset Selection Criteria

This set of criteria should be used to systematically determine a prioritized list of the datasets to be published by future teams. These have been developed through discussion with our advisors and through responses obtained from Dr. Federica Bettio, a member of the Venice website administration.

The criteria that will be used are:

- **Level of Completeness**: Similar to the analysis on the flagstaff pedestal data, the dataset in question will be evaluated to determine if all examples that exist of that data are recorded and if all fields of the existing data are filled. The score for this criterion will be directly proportional to the relative completeness of the dataset.

- **Age of the Data**: Data from the VPC is only updated when the data is used in a project. For this reason, some of the data has gone unedited since it was collected and has become outdated. Data that has not been updated since its creation would require more time and resources than more recently edited and more relevant datasets. Therefore, the length of time since a dataset was last updated will be inversely proportional to the score in this field.

- **Shelf Life**: This term describes the length of relevance of the data to interested parties; data such as canal depth or flagstaff height is consistent and would have a longer "shelf life" than data about this morning's high tide. Accordingly, datasets with longer shelf lives will be given higher scores in this area.

- **Usefulness for Tourists**: This will examine how tourists to Venice could use the dataset in question. A dataset's ranking in this category will be weighed against the usefulness to residents and researchers depending on to whom the data is more relevant.

- **Usefulness for Residents**: Analogous to the previous criterion, this will examine a dataset's usefulness to residents of the city. Weighting will again change based on which users view the data.

- **Usefulness for Researchers**:[6] This will be scored based on the possible relevance to researchers looking to make use of the data.

- **Difficulty of Translation**: Some publishing platforms require the metadata to be in a particular language. For example, Venice website requires Italian. In addition, some datasets have complicated metadata and column headers that require thorough explanations. Datasets that are less complicated and therefore take less effort to translate will receive a higher score here.

- **Data Structure Conversion Effort**: Depending on the format in which it is saved, some datasets could require additional processing in order to meet standards or increase the usability and diffusibility of the data. Because of this, datasets that require more processing will receive a lower score here.

5.2.3 Weighing Criteria

A Likert scale should also be used to assess the eight criteria. Each set can receive a score of one (1) to five (5) in each area, and each area can receive an individual weighting. The weighted scores for each criterion should then be totaled to determine the final score for each dataset. The weights range between a multiplier of one (1) to ten (10) and can be seen in Table 1 on page 13. These weightings reflect our judgment and the requirements of the administration of the Venice website. The analysis done with these selection criteria will enable the publication of datasets to the Venice website.

---

[6] This includes researchers in government agencies, non-governmental organizations, and independent entities.

Table 2: Selection Criteria and Weights

Level of Completeness	10
Age of the Data	5
Shelf Life	4
Usefulness to Tourists	2
Usefulness to Residents	2
Usefulness to Researchers	2
Difficulty of Translation	1
Data Structure Conversion Effort	1

Appendices

Appendix A: Bibliography

Creative Commons. (n.d.). *Attribution 3.0 Unported*. Retrieved from Creative Commons: http://creativecommons.org/licenses/by/3.0/legalcode
European Union Open Data Portal. (2013). *European Open Data Portal*. Retrieved from European Open Data Portal: https://open-data.europa.eu/

Gurin, J. (2014, September 12). *How Open Data Is Transforming City Life*. Retrieved from Forbes: http://www.forbes.com/sites/techonomy/2014/09/12/how-open-data-is-transforming-city-life/

Italian Open Data License v2.0. (2012, March). Retrieved from dati.gov.it: http://www.dati.gov.it/iodl/2.0/

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. (2013, October 1). Open Data: Unlocking Innovation and Performance with Liquid Information. Retrieved November 14, 2014, from http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovatio n_and_performance_with_liquid_information

Marijn Janssen, Y. C. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*.

Open Knowledge. (2013, February 4). *What Is Open?* Retrieved from Open Knowledge: https://okfn.org/opendata/

Open Knowledge Foundation. (2011, February 2). *What Is Open Data?* Retrieved from Open Data Handbook: http://opendatahandbook.org/en/what-is-open-data/

U.S. Open Data Action Plan. (2014, May 9). Retrieved from White House website: http://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan. pdf

Appendix B: List of Updated Fields from Flagstaff Pedestal Dataset

flagstaff attachment side
flagstaff being used
flagstaff bracket material
flagstaff cap present
flagstaff cap type
flagstaff color
flagstaff diameter cm
flagstaff height m
flagstaff material
flagstaff number of attachment brackets
flagstaff number of cracked brackets
flagstaff number of missing brackets
flagstaff painted
flagstaff position
flagstaff present
flagstaff usable
pedestal base average side height cm
pedestal base average side length cm
pedestal base number of sides
photo filename
image url
image-urls