



# WPI

WORCESTER POLYTECHNIC INSTITUTE  
INTERACTIVE QUALIFYING PROJECT

## Creating an Open Data Repository for Santa Fe

An Interactive Qualifying Project submitted to the Faculty of  
WORCESTER POLYTECHNIC INSTITUTE  
in partial fulfillment of the requirements  
for the degree of Bachelor of Science

**Gavin Hayes**  
**Arthur Lockman**  
**Nicholas McNary**  
**Nicholas Shannon**

Report Submitted To

Tom Johnson  
Marilyn Ward  
*It's The People's Data*

Scott Jiusto  
David Spanagel  
*Worcester Polytechnic Institute*

April 30, 2016

*This report represents work of WPI undergraduate students submitted to the faculty as evidence of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see <http://www.wpi.edu/Academics/Projects>.*

# Abstract

Like other cities around the world, Santa Fe, New Mexico collects information about its roads, schools, elections, health, and other public resources. It has not been easy for the public to access this data. The Open Data Movement is a global effort seeking to deploy new technologies to bring greater transparency on behalf of the public interest and to enhance government effectiveness. Our project team applied lessons learned from open data initiatives elsewhere to gather and present sample data sets from Santa Fe via a demonstration open data portal and a data dashboard. These examples can be used by the city Information Technology and Telecommunications Department to demonstrate the potential of open data, as it works with city leaders to explore an open data initiative for the city.

# Executive Summary

Around the world, governments and other organizations are collecting information on their constituents. This information can take many forms: maps, tables, graphs, forms, and more. Open Government Data (OGD) specifically refers to the idea that any information collected and maintained by a government agency should be freely and easily accessible to the governed. One of the largest barriers to this data publication is that many of the computer systems that the data is stored on are antiquated and simply cannot interface with modern systems. The second largest barrier is cost. Many computer systems need to be upgraded in order to be able to publish data, and the software that is often used for publication can be costly. These two factors mean that OGD movements require large monetary investments from the supporting governments as well as legislative support to make sure that each government agency actually publishes their information.

It's The People's Data are an organization that focuses on establishing Santa Fe as an example for open government data. They seek to engage citizens in the movement and enlist the taxpayers in using the data that their money goes toward (It's The People's Data, n.d.). ITPD put together a three phase plan to achieve that goal. They have recently entered phase two, which involves reviewing existing methods for implementing open data practices within city environments. The project team worked with them as well as the City of Santa Fe Information Technology and Telecommunications Department to further that goal.

## Project Goal

The goal of this project was to provide the city of Santa Fe with a proof of concept for an open data initiative. By collecting data and creating a prototype system, the team showed how OGD is beneficial to the government as well as the citizens. The system that is created will help the City of Santa Fe and ITPD continue their efforts for government transparency.

# Methodology

While in Santa Fe, the team broke the project down into four primary steps:

1. Gain access to key datasets and determine the data that best suits the interest of the stakeholders
2. Create a repository of information combining the data sources collected as part of Phase 1, providing a resource for building applications using the available information.
3. Create a prototype dashboard for city government workers to view the data, combining new and existing data to bring light to interesting patterns and uses.
4. Determine resources that ITPD can use for future data portal development with the help of the city government.

During project preparation and the early weeks of being on site, the team collected numerous datasets from organizations in Santa Fe and elsewhere in New Mexico. The team identified six main areas of interest for the prototype: elections, education, crime, health, tourism, and roads. These datasets were of greatest interest to ITPD and the data for most of these areas were already being uploaded on various websites. Once the data was gathered, the team created a data repository to hold all the data in one convenient place. Then, using APIs (application programming interface) from the data repository, connections were made to a dashboard creator website. After building the data repository and creating multiple dashboards, the team provided additional resources to both ITPD and the City of Santa Fe to bolster the case for the transparency movement within the city.

## Findings and Outcomes

### Datasets

In talking to ITPD and seeing what the city provided, the team gathered six initial datasets that were useful in a proof of concept open data portal: election, education, crime, health, tourism, and roads. The datasets were scattered in a variety of locations, both physical and digital, which made access difficult at first. The data also came in various forms such as maps, surveys, excel sheets, etc., which could impair interpretation if not given a proper description for users. Every area of interest included some type of useful information but also posed limitations: The same type of data from two sites would be different, the data was not coming from the city, lack of data due to privacy, and data from previous years needed to be used because of the data acquisition rate. In dealing with these limitations, the team had to assess each dataset to find useful information to be used in the platform for dashboard development.

## **Platform**

During the time datasets were being gathered, the team was also looking at several data repository websites that met certain criteria. The team found that CKAN was the most feasible for a proof of concept platform due to it being free and having multiple tools for dashboard development. After setting up CKAN on the City's ITT Department server, all of the datasets were added. The website offered ways to look at the data but no dashboard element. Luckily, CKAN offered APIs for most of the datasets so the information could easily be moved to a dashboard.

## **Dashboard**

Using the APIs from CKAN, the team began creating dashboards from the information the datasets provided. After looking at various websites that offered dashboard creation, Tableau was the most feasible for a proof of concept dashboard. The team created dashboards to showcase school attendance, standardized test results, road conditions, voting locations, and school performance grades.

## **Recommendations**

The team provided some recommendations to the city and ITPD to aid them in creating an open data initiative for the city. ITPD can go to multiple open data organizations in Santa Fe as well as the State to gain support. One key recommendation is that their needs to be an executive order to launch an open data movement within the city. The city council will also need to approve funding for such an initiative, as it will require software purchases and training for employees. By looking at what other cities have done, Santa Fe can focus on what procedures can be implemented to gain support. These resources will be valuable once the team has left Santa Fe so that both ITPD and the Santa Fe City Government can continue their efforts to create a city-wide open data repository.

Additional information and resources for this project can be found on the web at <https://sites.google.com/site/sf16dopen/home>.

# Acknowledgments

## **It's The People's Data**

We would like to thank Tom Johnson and Marilyn Ward for coordinating this project, and communicating on our behalf with stakeholders in Santa Fe to set up meetings and move the project along.

## **The Santa Fe New Mexican**

We would like to thank the Santa Fe New Mexican for supplying us with office space to work on the project. Specifically Henry Lopez, Ray Rivera, and Dan Schwartz.

## **The City of Santa Fe**

We would like to thank Renee Martinez and David Barsanti for working with us and providing us with server hosting resources to make our project a reality.

## **Worcester Polytechnic Institute**

We would like to thank David Spanagel and Scott Jiusto for advising this project, and Fabio Carrera for setting the project up.

# Contents

- Abstract** **i**
  
- Executive Summary** **ii**
  
- Acknowledgments** **v**
  
- 1 Introduction** **1**
  
- 2 Background** **4**
  - 2.1 Open Government Data . . . . . 4
    - 2.1.1 Interest in Open Government Data . . . . . 5
    - 2.1.2 Ideal Formats for Open Data Publication . . . . . 6
  - 2.2 Implications of Data Publications . . . . . 7
    - 2.2.1 Cost of the Open Data Release Model . . . . . 7
    - 2.2.2 Technical Issues . . . . . 8
  - 2.3 Open Data in the United States . . . . . 9
    - 2.3.1 Freedom of Information Act leading to Open Data . . . . . 9
    - 2.3.2 Open Data Action Plan . . . . . 10
  - 2.4 It's The People's Data . . . . . 12
    - 2.4.1 Overview . . . . . 12
    - 2.4.2 It's The People's Data Goals . . . . . 13

2.5	Open Data in Santa Fe, New Mexico . . . . .	14
2.6	Laws for Public Records . . . . .	14
2.6.1	Transparency Portal . . . . .	15
2.6.2	Inspection of Public Records . . . . .	15
2.7	Open Government Data Portal Examples . . . . .	16
2.7.1	Open Government Data in Europe . . . . .	16
2.7.2	Los Angeles GeoHub . . . . .	18
2.7.3	Albuquerque’s Open Data Site . . . . .	18
2.8	Use of Geographic Information Systems in Open Data . . . . .	18
2.8.1	What is a GIS? . . . . .	18
2.8.2	GIS Tools and their Implementations . . . . .	20
<b>3</b>	<b>Methodology</b>	<b>22</b>
3.1	Phase 1: Access and Assess Datasets Available in Santa Fe . . . . .	24
3.1.1	Accessing Data . . . . .	24
3.1.2	Assessing Data . . . . .	24
3.2	Phase 2: Create Centralized Data Repository for Santa Fe . . . . .	25
3.3	Phase 3: Create Prototype Dashboard for Santa Fe . . . . .	27
3.4	Future Recommendations . . . . .	28
<b>4</b>	<b>Findings, Analysis, and Outcomes</b>	<b>29</b>
4.1	Overall Findings on Dataset Acquisition and Utility . . . . .	29
4.1.1	Election Data Findings . . . . .	30
4.1.2	Education Data Findings . . . . .	33
4.1.3	Police Data Findings . . . . .	34
4.1.4	Health Data Findings . . . . .	35



4.1.5	Tourism Data Findings . . . . .	36
4.1.6	Roads Data Findings . . . . .	37
4.2	Portals . . . . .	38
4.2.1	Overview . . . . .	38
4.2.2	CKAN . . . . .	39
4.3	Dashboards . . . . .	40
4.3.1	Tableau . . . . .	40
4.3.2	The Dash . . . . .	44
4.3.3	Power BI . . . . .	45
4.3.4	Jaspersoft . . . . .	46
<b>5</b>	<b>Recommendations for Future Work</b>	<b>48</b>
5.1	Santa Fe Open Data Initiative . . . . .	48
5.2	Legislative Change to Data Management . . . . .	49
5.3	Tools and Technologies . . . . .	49
5.4	Interviewing to advance the Santa Fe Open Data Initiative . . . . .	51
5.5	Santa Fe Civic APPaTHON . . . . .	51
	<b>References</b>	<b>52</b>
	<b>Glossary</b>	<b>55</b>
	<b>Acronyms</b>	<b>56</b>
<b>A</b>	<b>Open Data Stakeholder Discussion Topics</b>	<b>57</b>
A.1	Government Employees . . . . .	57
A.2	Citizens of Santa Fe . . . . .	58
A.3	Local Organizations . . . . .	58

<b>B Tableau Internal Errors</b>	<b>60</b>
<b>C Tableau Dashboards</b>	<b>62</b>
<b>D Dataset Information</b>	<b>68</b>
<b>E Dashboard Software Comparison</b>	<b>72</b>
<b>F CKAN Web Data Connector</b>	<b>75</b>

# List of Figures

2.1	A snapshot of roads data in the Vienna repository. The table column labeled <i>formats</i> <sup>1</sup> shows the different ways that the data can be downloaded. . . . .	17
2.2	Example of World Represented in a GIS (State of Indiana, n.d.). . . . .	19
3.1	Project Flowchart . . . . .	23
4.1	App Finds Locations of Voting Polls and Your Location . . . . .	31
4.2	Click on Closest Polling Location and App Gives You Directions . . . . .	32
4.3	School Grades Map . . . . .	34
4.4	Crimereports.com Socrata Dashboard . . . . .	35
4.5	CKAN Homepage . . . . .	39
4.6	Dashboard Comparison . . . . .	40
4.7	Example Dashboard Using Tableau (published online) . . . . .	41
4.8	Data Source View . . . . .	42
4.9	Visualization Types . . . . .	43
4.10	Worksheet Visualization . . . . .	43
4.11	Dashboard View . . . . .	44
4.12	Example Dashboard Using The Dash . . . . .	45
4.13	Example Dashboard Using Power BI (page 1) . . . . .	46
4.14	Example Dashboard Using Power BI (page 2) . . . . .	46

4.15 Example Dashboard Using Jaspersoft . . . . .	47
4.16 JasperReports Server Main Page . . . . .	47
B.1 Web Data Connector Beginning Execution . . . . .	60
B.2 Web Data Connector Internal Error . . . . .	61
C.1 Santa Fe Streets Dashboard . . . . .	62
C.2 School Attendance Dashboard . . . . .	63
C.3 School Grades Dashboard . . . . .	64
C.4 Santa Fe Voting Locations Dashboard . . . . .	65
C.5 Santa Fe County Crime Reports Dashboard . . . . .	66
C.6 Santa Fe PARCC Results Dashboard . . . . .	67

# List of Tables

3.1	Project Timeline . . . . .	23
3.2	Criteria for selecting a platform for creating the open data repository prototype. . . . .	26
4.1	Comparison of various repository software types. . . . .	38
D.1	Table of datasets. . . . .	71
E.1	Comparison of various dashboard software types. . . . .	74

# Chapter 1

## Introduction

Technology as a whole is constantly changing. Over the last few decades, the human knowledge base has expanded to upgrade technology in almost every industry, especially pertaining to information systems. Cloud storage systems allow easier access to patient records for the healthcare industry. Police can use internet presence as a method for tracking down suspects. As this technology has improved and become more integrated into the everyday functions of society, society as a whole has been moving towards having greater transparency as to what organizations and the government do with taxpayer money and resources. People feel entitled to access information that was previously inaccessible due to the archaic nature of the technology that supported said information. This entitlement was evident as early as the 1760s, when Anders Chydenius helped in the creation of a new law for the Swedish parliament that abolished political censorship and required government documents to be publicly accessible (Mustonen, 2006). For centuries, the concept of open data did not progress substantially. The next big change began in the 1950s when the United States began to look into open information. This endeavor occurred after there were repeated requests from citizens for information about secret government projects, which eventually led to the creation of the Freedom of Information Act (FOIA), which required data to be provided unless it needed security clearance to view (United States Department of Justice, 1996). More recently, the public opinion has been that access to data is a basic right that all people should receive (Carrasco & Sobreperre, 2015). The public desires for governments and organizations to publish data. Thus the concept of Open Government Data was formed, and innovative solutions started to be developed to address the gap between those desiring data and the systems that the data was housed on. One example of these solutions was the creation of the field of Geographic Information Systems (GIS), which are systems that model data spatially using cartography. Through GIS, data are superimposed over maps in unique layers that can be added and removed at will. Dashboards and widgets have become commonplace in displaying government data for public utilization. In this way, our sponsor, It's The People's Data (ITPD), sought to expand the open data initiative in Santa Fe, emulating other cities around the world while also attempting to become a leader within the movement as a whole.

One of the main problems that obstructs data from being integrated with modern

public portals is that the systems in place for storing them are too old to interface with modern computer technology. Obsolete computer systems that store the data perfectly fine within their own areas present issues for cities looking to migrate to open data practices (Berners-Lee, 2009). It can be difficult for cities to create the active links between the old and new systems, which means that little can be done other than simply maintaining the older systems. The city of Santa Fe has room to establish active connections between the old systems that are storing the data and the new systems that will display the data. Data can be exported in batches and then imported into an alternate location to post for public viewing, however this is not an ideal method for staying up to date. As a result, the city of Santa Fe and It's The People's Data were both looking to work on building a central data repository to reorganize the way that data is presented to the public for the city. The goal was to establish digital relationships between the data storage systems and the dashboards that will be used to display the data. These relationships will allow the data to be updated on demand, which in turn will allow the city to have greater efficiency in being transparent.

It's The People's Data's initiative is to utilize government-collected data to identify meaningful trends in information. As a community group from Santa Fe, New Mexico, they seek to engage citizens in the movement and enlist the taxpayers in using the data that their money goes toward (It's The People's Data, n.d.). ITPD put together a three phase plan to achieve that goal. They recently finished phase one, which was to gather the personnel to drive the organization's goals and to begin communication with city departments explaining what ITPD was setting out to do. They have begun moving on to the next phase, which involves reviewing existing methods for implementing open data practices within city environments. The Santa Fe Information Technology and Telecommunications (ITT) department recently purchased extensive licenses for Esri's ArcGIS software in preparation for designing a whole new area for displaying open data to replace their outdated Geocortex platform. ITPD and the Santa Fe IT staff had been looking at Los Angeles' GeoHub as an example for the city's new GIS, as it is also built using ArcGIS. In addition to the progress on the new GIS, Santa Fe made headway on publishing road data collected by previous Worcester Polytechnic Institute (WPI) projects for the city. ITPD was in communication with the GIS staff, members of the city police, as well as various other departments about obtaining data that could be used.

WPI has an extensive history of implementing data driven projects for sponsors in the past, especially at the Venice project site. Most notably, Venipedia, a wiki created by WPI students and faculty to explore data at a fine level of detail, and the Venice Dashboard, a dashboard site that uses widgets to display information in real time to users, have paved the way for adaptations of their infrastructure for Santa Fe. Past WPI projects have created the SantaFedia wiki site and the Santa Fe Dashboard, both based on their Venice counterparts. However, both Santa Fe adaptations are rudimentary, having never been fully implemented to the same extent as the Venice sites. The team used what applicable information from those two sites already existed towards advancing ITPD and the city's goals. The strong technical backgrounds represented among the team lent themselves to the project in that programming was needed to establish the repositories and dashboards (Johnson et al., 2016). Since It's The People's Data was in the early stages of actually gathering data from the city, the WPI team was able to expedite the process of accumulating data from the desired city departments. With the

team's assistance in speeding up this initial process, the city was able to continue creating new views and implementing the data gathered in new ways after the project was completed.

As a result, the ultimate goal of this project was to provide the city of Santa Fe with a proof of concept for an open data initiative. This would entail establishing ways of easily accessing government data and organizing the resources together. The WPI team analyzed the currently available systems for providing this data, gauged the usefulness of specific datasets, and built a prototype system to allow access to some data repositories. The project was broken down in three phases. The first was to gain access to key datasets on behalf of It's The People's Data, and with their assistance, determine the data that was most desirable for inclusion in the repository. The team then constructed a repository for the data, allowing it to be easily accessed and used for applications. Finally, the team created a prototype dashboard to view the data, combining new and existing data to bring light to interesting patterns and uses. This yielded results showing that an open data initiative could work within the city of Santa Fe with proper backing from the government.



# Chapter 2

## Background

This chapter introduces the concept of Open Data, some potential uses for it, examples of its use from around the world, and platforms that can be used for its publication. The chapter begins by covering Open Data at a conceptual and policy level, and moves on to discuss specific implementation examples and methods. These topics provide background for understanding the desire for an Open Data repository in Santa Fe, NM.

### 2.1 Open Government Data

Around the world, governments collect data about their citizens through various means. Whether through census, survey, or tax information, governments maintain large repositories of data about their citizens. Many believe that these data repositories should be publicly accessible, allowing citizens to use this collected data however they see fit (Kucera & Chlapek, 2014; Johnson et al., 2016). This concept is known as OGD. OGD encompasses any data that is collected and maintained by a government or any other public entity. According to The Sunlight Foundation, a nonprofit organization devoted to government transparency, OGD should be:

1. Complete
2. Primary
3. Timely
4. Easily accessible
5. Machine-readable
6. Non-discriminatory
7. Using open standards for publication
8. Available under explicitly stated terms of use (license) which allows its reuse with minimal restrictions

9. Permanent
10. Available to potential users for minimal cost (Sunlight Foundation, 2010).

The concept of OGD is not new. OGD first emerged in the 1760s when legislation was implemented in Sweden that abolished political censorship and required government documents to be publicly available. The theory did not actually catch on at that time, but rather fell aside and was actually suspended from 1772 to 1809 (Mustonen, 2006). Then, the concept made a revival in the 1950s in the form of citizens' requests to the US Federal Government for information about secretive projects that the government was working on. These requests and subsequent actions by citizens eventually led to the creation of the Freedom of Information Act (FOIA). This act mandates that any citizen has a right to gain access to any federal records. All federal agencies are required to release this data unless the information requires some sort of security clearance (United States Department of Justice, 1996). The FOIA is based on the belief that access to government information was "essential to the realization of a civil society, democratization, and a rule of law" (Lassinantti, Bergvall-Kåreborn, & Ståhlbröst, 2014). This belief grounds most of the global OGD campaigns as well. Additionally, the last ten years have seen a shift from the simple desire for data to be available if it is desired to the goal of having governments publish data without prompting. This preemptive approach to releasing data became a focus after a 2007 conference of industry leaders in the open data initiative (Chignard, 2013).

### **2.1.1 Interest in Open Government Data**

Interest in OGD is growing globally, with many different groups working towards the same general goal. Interest in OGD comes from two main places: policy created internally to the government, and citizen groups campaigning for their municipal or federal governments not only to release the data they have collected, but also to do something useful with it.

Within governments, OGD generally has support from legislators and department heads. Requiring this data to be published forces the different government agencies to be accountable, which some argue makes the internal workings of the government smoother. There are also other factors nudging governments along - namely nongovernmental organizations, advocates, and experts who want the government to make smart investments to allow for future innovations (Lassinantti et al., 2014). More specifically, they are looking to open up government data to allow the development of applications and services that put the released data to use; some are interested in using this data to develop applications to make useful information about cities and towns more easily accessible, and others are interested in using the data for commercial gain (Bedini et al., 2014).

Open government data also supports national government goals. Logica Business Consulting in 2012 put together a study showing 6 main motivations for opening up data in the US, UK, Canada, and New Zealand:

1. Increase transparency
2. Stimulate economic growth
3. Improve government services and responsiveness
4. Encourage reuse
5. Improve public relations and attitudes towards government
6. Improve government data and process (Rothenberg, 2012).

Citizens are interested in open data for a simple main reason: their local, state, and federal governments have been collecting large amounts of information on their behalf, and they would like to have access to the data that has been collected. Many view access to this data as one of their basic rights (Carrasco & Sobreper, 2015).

Another perceived benefit of OGD is transparency in government and improvement of government process. A focus group conducted with representatives from four different Czech municipalities produced a concise list of the most important benefits they saw in the release of OGD that matched closely with the list above. The study shows that there is enormous perceived benefit of government transparency. If all data is released in a timely manner, then theoretically citizens would be able to easily see if any government agency hasn't completed their stated objective or broken the rules set out by a piece of legislation. The participants in the study also indicate that they perceive OGD to stimulate economic growth and improve government services. This aligns with the desire cited earlier within government agencies to use this data to improve their own services.

### **2.1.2 Ideal Formats for Open Data Publication**

The growth and development of the OGD movement follows a similar progression to the development of the open-source software community - albeit at a different time. The Open Source community has democratized the development of software tools, making it so that anyone can contribute to a project; similarly the open data community hopes that the proliferation of open data will help to make it so that any citizen can meaningfully participate in their government. Activists do recognize that for the majority of people, most government open data portals are not accessible. The means of accessing the data can be non-trivial: the level of knowledge or technology required to access the data is not something that the average user could achieve without substantial effort. Open data activists therefore seek to create intermediary bodies that present the data in an understandable and accessible way to the general public. These intermediaries three governing principles: intermediaries should be data-driven and take large amounts of data and make it available; intermediaries should be open and make the data they pull from available; and intermediaries should seek to actively engage the citizens in public issues (Baack, 2015).

## 2.2 Implications of Data Publications

Some problems could prevent OGD from becoming widely published. The first is an issue of privacy. Certain datasets that governments collect need to have all identifying information removed from them, a process called anonymizing, before being released to prevent giving away the citizens' personal information. This data might also compromise some intellectual property that is otherwise protected by law. Detailed publication of information about city infrastructure could put the city at risk for an attack against the public works. Data publication in government also has a non-zero cost associated with it – not usually in the technology purchase, but simply in terms of paying staff to build and manage the data portals. In 2012, the UK estimated that it could cost between £53,000 and £500,000 for their transparency agenda to be implemented (National Audit Office, 2012). Some municipalities might simply not be able to cover the cost or time requirement to pull together their open data portals.

### 2.2.1 Cost of the Open Data Release Model

Sometimes, Open Data seems like just another special interest group fighting for tax payer money. "Decreased government spending aggravates the problem since it is not a given that lost income from releasing previously licensed data for free will be compensated by the national government or the counsel" (Conradie & Choenni, 2014). Unfortunately, at least to start, releasing data to the public requires additional resources for almost any organization. There must be justification of Open Data in order to expand the budget to accommodate it in public organizations. In order for Open Data to be a lasting practice, it must become established as the main reason for data to be collected at all. Once it is understood that the means of collecting data is supported by the goal of having Open Data, the costs can be better justified.

Paying for the start-up cost is a critical issue in creating an Open Data portal. Software licensing is often a large upfront expenditure that simply may be outside of the organization's capabilities, as mentioned in the introduction to this section. "The participating organizations such as City Works earn income through the sale of their data" (Conradie & Choenni, 2014). Costs could be minimized if there were significant models of successful Open Data efforts to follow. Currently almost every Open Data release is unique and requires advanced decision making for an effective solution. Templates or clearer models could make it more appealing and accessible to city officials. Organizations looking to release data also need technical expertise and a desire to work on an initiative. Over a long period of time, optimization of data release has lowered overhead costs, but gathering the funds to make a change is still a challenge. "Research has shown that the data recovery model results in proportionally low amounts of income and that less restrictive data release can in some cases lead to increases in revenue or, in the case of free data, lower transaction costs" (De Vries et al., 2011). Organizations need to include proper incentives in order to pursue the creation of such a portal.

## 2.2.2 Technical Issues

Not all problems of Open Data are political problems. The ability to release data also depends on the technical environment in a particular area. Fortunately Open Data can scale to the area at hand, but its effectiveness depends on its ability to overcome the following technical issues.

### Keeping Data Up to Date

One of the first rules of data storage is to have Uniform Resource Identifiers (URIs) be the names for things on the web. URIs are used to identify resources on the web; the most common URI is a URL which provides an indicated resource. However, websites tend to violate the first rule of data storage and store data in multiple places (Berners-Lee, 2009). This problem used to be significantly worse in Web 1.0 where everything was static content. Static content must be updated manually by a Webmaster. Data could be shared between websites, but everything had to be done manually. Sometimes the original author would ask for a linkback or to be emailed if an individual wanted to include the author's content on the latter website. Use of another person's content requires a level of trust between the author and the viewer to use the content responsibly and keep it up to date. If an author is receiving linkbacks or emails, they can be aware of where their content is listed in case they want to take action.

Link access provides a solution by keeping data in only one place, which in turn prevents data duplication. In order to build a media rich website and to make use of data, a simple URL is not sufficient. The author of data creates a way to control how the data is shared by providing an API to access the data. Web applications that make use of APIs have immediate instant data to work with and process. APIs data release does not have to implement its own programming language necessarily and can be as simple as SQL queries providing output as CSV (Edoceo, 2014). Yahoo Finance is a website that implements such methods (Greenido, n.d.).

### Data Protection

Open data introduces standards that did not previously exist on managing government data. Accountability becomes more complicated with the allocation of the provider, processor, owner, and maintainer. Government accountability has consistently been an issue of concern, and skeptics can challenge the idea that open data will actually improve government accountability. If government transparency is to be improved, "examination of Open Data at a local level could be conducive to reach this aim" (Conradie & Choenni, 2014).

Releasing the data presents several security concerns, but can also increase security. While open data does not make classified information become publicly available, it gives the public more power in the way they can interact with their government. Instead of individuals deciding what to release to the public, everyone can make informed decisions based on the

data at hand.

In order to avoid upsetting the public, most local governments (Conradie & Choenni, 2014). Privacy issues right now are barriers of entry to Open Data release, but remains important as long as data is being released. "Data can be either useful or perfectly anonymous but never both" (Ohm, 2010).

## **2.3 Open Data in the United States**

This section covers the policies and plans for creating open government data portals in the United States.

### **2.3.1 Freedom of Information Act leading to Open Data**

The idea of OGD is familiar to the US Government. In 1966, the FOIA was signed into law by President Lyndon B. Johnson. It allowed disclosure of information and documents that were previously unreleased to the public. Information that is disclosed to the public includes (United States Department of Justice, 1996):

- Descriptions of Organizations where the Public can obtain information
- Substantive rules and statements of general policy adopted into law or by an agency
- Opinions and dissenting opinions made in the adjudication of cases
- Manuals and instructions to staff that affect the public.

There are many more guidelines and procedures that outline what the government can and cannot show to the public. The FOIA has been amended several times since it was introduced, building the foundation for OGD that is in place today. Many of these amendments revolve around national security. The fear that national security can be disrupted by this act goes back to President Reagan who included language in his executive orders allowing certain government agencies to withhold large amounts of data. These orders were then altered by President Clinton between 1995 and 1999. Even today, President Obama created retroactive classification due to national security concerns. This retroactive classification made it so that the government can still classify information that has been released to the public if it is relevant to national security. While the FOIA has been in place, many presidents have placed restrictions on what can and cannot be publicly available. These restrictions were aimed towards the national level. When it gets down to state and local government, many of these restrictions do not apply and mean that the majority of data needs to be released to the public.

Even though the act was aimed towards federal agencies, the act has been trickling down to the city and state levels. The FOIA has great impacts on the accountability of the

government for their decision making and to help inhibit corruption (Birkinshaw, 2010). This idea that the FOIA will decrease government corruption is partially responsible for sparking the OGD movement. Many cities and states want to get their governments to become a part of the open data movement. This would let citizens play a role in how the government works by being able to understand what is happening in the community. Even though the idea of open data is becoming widely accepted, there are still debates going on throughout the nation (The Economist, 2015). Some argue that not all of the data should be publicized. A good portion of data that is found is considered to be useless, because it does not contain metadata, but gets publicized anyway. Other types of data, for a more national level, is kept away from the public. These arguments depend on what the information is about and how it could be used for or against the people. The strongest point is that open data can be beneficial to the people to help improve the government by empowering the people to be a part of their local government. Participation in government is critical for any democratic governmental body to develop.

### **2.3.2 Open Data Action Plan**

In 2013, the US and other G8 leaders endorsed the Open Data Charter which listed out five main procedures that will be the foundation for the availability of open data (G8, 2013):

1. Open Data by Default
2. Quality and Quantity
3. Usable by All
4. Releasing Data for Improved Governance
5. Releasing Data for Innovation.

Using those five procedures, the G8 leaders outlined what each of the procedures meant and what the goal of each procedure was to be. Those five procedure were to be implemented at the best possible times during each of the nation's time-frames. The G8 leaders signed this charter knowing and understanding the importance of open data for the public. With that understanding and the 5 main principles they determined what would be done to accomplish those tasks.

Each nation recognized the significance of free access to data that will be published openly. As data is acquired, government bodies will take steps to publish the data in a way that the public can easily understand. The government can release a wide range of datasets that can be accessed by the public. However, some data is not available to the public for safety reasons. Each nation understands that even though data needs to be open to the public to help the society and economy, there are legitimate reasons why data sets cannot be shown. To get this data open to the public by default, each nation needed to come up with a plan on how they would carry out acquisition, maintenance, and publishing of the data.

The next procedure addresses the quality and quantity of the data that is released. There are endless amounts of information that each nation must be able to maintain so that they can publish the most important datasets to the citizens. This data must be accurate, written clearly, be fully described, and to be released as early as possible. If data is not actively being published and updated, then there would be little use for it because it would not describe the current conditions. Each data set released can help bring together new information that a person would not be able to see if only looking at one section. This also requires the government bodies to listen to the feedback from the people using the data to improve the data that is being released.

For the data to be usable by all, the data needs to be open in many forms. The data, being available to the public in these various forms let them re-use it as they wish. Also, the data needs to be free of charge to let everyone have access. For this goal to be met, the data that is useful to the public cannot be restricted by the government. This also means that the data can be easily accessed by the use of open formats. CSV files are the best for universal access as they allow users to open the data contained in the file on different types of software.

Open data can strengthen the relationship between government and its citizens. The use of open data means that citizens can be kept up-to-date on governmental policies and express their opinion on them. The shared information is usually how the data was collected and the technical applications the data can provide. The data is shared not only with the citizens but also with other governmental bodies to stimulate innovation. By making data accessible, people can use the information they gather to benefit society and benefit the economy. In order to stimulate this, the countries recognize that they need to work on encouraging people to find the data as well as make it in multiple machine-readable formats for everyone's use.

The U.S., using these five procedures, made the following four commitments (United States Government, 2014).

1. Publish open data in a discoverable, machine-readable, useful way
2. Work with public and civil society organizations to prioritize open data sets for release
3. Support innovators and improve open data based on feedback
4. Continue to release and enhance high-priority data sets

The commitments were part of President Obama's vision to create a plan to manage the nation's information as an asset. This vision reinforced the benefits that open data will provide to the American people.

Putting government data online and making it easy to find and use - while continuing to rigorously protect privacy - can help American families find the right health care provider, identify the college that provides the best value for their money, keep their families safe by knowing which products have been recalled, and much more. (United States Government, 2014).



Since the G8 Summit, the United States has made significant progress in releasing national data for everyone's use. Two important examples of progress have been "Data.gov" and "Project Open Data." Data.gov is the main location for the United States government's data. Managed by the U.S. General Services Administration, Data.gov lets governmental bodies from various locations to add their data to the site. It takes those government's metadata and aggregates them onto their site. The other progression the U.S has made was creating the Project Open Data outline to give other people information on how they acquired the data and how they put the data up online. With that type of information, organizations and other governments can find useful information on how to get their data to the public in an easy way.

## 2.4 It's The People's Data

### 2.4.1 Overview

It's the People's Data (ITPD) is a small organization based in Santa Fe, NM that would like to help identify and retrieve data that the local government has and bring it to the people that deserve access to this information. They are a grass-roots movement with no board or set staff. They have no membership requirements other than attending meetings and taking on some tasks voluntarily to help the organization meet its goals. ITPD was created as a proof of concept project for the Society of Professional Journalists. A proof of concept involves proving the feasibility or practicality of a theory, to ensure that the concept is indeed possible or worthwhile before extensive work is done on it. ITPD are endorsed but *not sponsored* by other organizations in New Mexico. ITPD also has not required any funding other than borrowing meeting space from the local newspaper the Santa Fe New Mexican (Johnson et al., 2016). Their main initiative is to:

...make sense of the changing information environment and enlist citizens in a community effort to identify and retrieve the data they have paid local government to collect, to learn new analytic skills to tease information out of that data for a myriad of purposes and then to communicate the results of that analysis to a variety of audiences in the community (It's The People's Data, n.d.).

With this access, they would like to relay important information the public by analyzing the vast data sets for important information that is useful to a variety of community members. They have long term goals that, with the collective help from the government and the people, can be accomplished and have great impacts on the community. As of right now their activism only covers the city of Santa Fe, however they hope to provide a model for other cities of similar size to use.

## 2.4.2 It's The People's Data Goals

In May 2015, It's The People's Data sent out an initiative plan to the Santa Fe citizens. It outlined what ITPD would be doing in order to deliver open data to the public. Laying out their strategies and time line of events, ITPD created objectives to make the Santa Fe's government data easily accessible, readable, and be the main mission for the citizens and government officials:

1. Have the data collected by the government be easily available to the public in its original form
2. Change the culture of public employees and politicians to recognize the importance of people's data availability and have it be a primary mission at all levels of the government
3. Have journalism institutions help in the open data plan
4. Teach the public and government officials how to analyze data to understand the information it holds
5. Have broad collaboration to maximize the openness and participation
6. Make Santa Fe first small city to gain recognition as championing It's The People's Data
7. Produce a "cookbook" to show the process of It's The People's Data initiative.
8. Build an easily accessible database of the city's databases.

Comparing these objectives to the concept of OGD, ITPD is on target for getting the city of Santa Fe to have their data open for the public. While some of ITPD's objectives are for their own organization, the others meet the tasks as outlined for OGD. One of the main points that It's The People's Data is trying to tell the public is that everyone needs to be involved in the creation and maintenance of Open Data. With support from such a wide range of citizens, everyone's goals can be accomplished and be shown to other small governments as a way to help improve their output of data. The team's work with various local and national groups provides information useful in determining the success an open data can have on small city governments.

To reach their goals, It's The People's Data created a three phase plan from June 1st, 2015 and ending in mid-Summer 2016 (Santa Fe, NM, Johnson, & It's The People's Data, 2015). In the first phase, the organization recruited participants to help announce the initiative as well as making a presentation to show the local organizations about their plan. In the second phase, they are taking a census of other city departments to use as models. These departments are to be used as proofs of concept to help create a plan of action. Finally, phase three improves methods for collecting the information. This involves training officials on how to use the data tools and the publication of a "cookbook" to show the initiative process.

## 2.5 Open Data in Santa Fe, New Mexico

This section covers the current state of OGD in Santa Fe as it pertains to our project.

### Current Data Repositories

The main free online source for data is the City of Santa Fe Internet Mapping Site. This mapping site is an example of a Geographic Information System (GIS). A GIS uses maps that represent data and lays them over each other to get different types of information. Adding layers to maps offers to the viewer a whole new type of data that would not have been found through the use of one single map. Different utilities, city information, and transportation are just a few examples of the layers that may be provided on GIS maps. GIS implementations are further explored in Chapter 2.7. With regards to the Geocortex system currently in place, some users may find it clunky and hard to navigate through. In addition to the software being hard to use as a first time user, the data is also old and is not up to date. The site needs improvement but overall has a great use in publicizing governmental data. Most of the data that can be found on this type of application is directly linked to the citizens way of life. Looking through other cities open data sites, most use a GIS to show different types of information around the city. Using this system to accomplish the main goals of OGD would be easily accomplished but would need people actively maintaining and updating the information that goes into the software.

One other system that uses information from CK Console as well as gathering information from other sites is the Santa Fe Dashboard. This website displays its data that it is gathering in the form of individual widgets. Each widget is specified to one particular type of data and there are multiple types of data that could be shown. One good dashboard for reference is the Venice dashboard that was created in an IQP. The Venice dashboard may not be used by the citizens as much, but can be used to help tourists. It shows weather, incoming and outgoing flights, hotels that are available, and other important information that is pertinent for tourists and citizens. Using the ways that this data is collected, it would be beneficial to the City of Santa Fe which also sees many tourists during different times of the year. However, the Santa Fe Dashboard is very hard to locate if a person does not know exactly what they are trying to find. On top of the difficulty of finding the dashboard, some widgets are not working.

## 2.6 Laws for Public Records

New Mexico is familiar with the idea of having data publicly available. In the New Mexico Statutes Annotated (NMSA), there are several statutes that pertain to having certain documents available to the public. These acts include Chapter 10 Section 16D Sunshine Portal Transparency, Chapter 14 Article 2 Inspection of Public Records, and various others. Most of the other public records laws are focused on elections and what candidates need to share with

the public. These NMSA acts then trickle down to the city level and take on various forms that the city uses to comply with these laws. The City of Santa Fe created the Santa Fe City Code of 1987 (SFCC 1987) that uses the NMSA acts to create ethical guidelines people have to follow. These guidelines can be over each department for the inspection of public records, but the most important guidelines for transparency are focused more for candidate information.

### **2.6.1 Transparency Portal**

In 2011, the state of New Mexico created an act that would start the creation of a Transparency Portal for the State of New Mexico. This would create a free publicly available data site that would host the state's financial information. The site that was created is called <http://sunshineportalnm.com/> and takes data from various financial sources and hosts the data on this site. It also has links to various government bodies, useful citizen information, and tourism links. The site is updated monthly which shows that New Mexico is actively involved with making data publicly available. Even though this is a valuable resource for the state, cities are looking for their own portal to show valuable information to the public that may be hard to locate. This is what ITPD is looking to do with the help of the WPI team in this project. By gathering useful datasets and information that is already currently available, there can be a proof of concept open data portal for the City of Santa Fe. As of right now, there is only a GIS mapping software for the city but that does not provide enough information for the city to be transparent. That is why ITPD is pushing for an open data portal for the city; it would be useful to the citizens as well as the government departments so that information can be shared between both parties without any problems. Currently, there is no information in the Santa Fe City Code or City Charter about implementing a transparency site. This would be very useful in getting a start into development of a transparency site and would need the help of all the city departments.

### **2.6.2 Inspection of Public Records**

In NMSA, the Inspection of Public Records Act (IPRA) is very descriptive in what can and cannot be shown to the public. These exceptions included medical records of mental and physical exams of a person, letters of reference, trade secrets, some instances of protected personal identifiers, and several other pieces of information. Other than those exceptions, everything that goes through the government should be publicly available. To access this data, the IPRA establishes the duties of the data custodian as well as the process a person needs to take to acquire data. To make this easier for the custodian and the person wanting the information, the city has created a simple form people can fill out whenever they would like to have access to public records. There are fees depending on the amount of files the person wants and the type of files. The process is very simple but can take a very long time for data to get into the person's hands.

## 2.7 Open Government Data Portal Examples

There are many examples of OGD initiatives from around the world that can be used as guidance for organizations to create their own open data initiatives. However, there is not one single approach to open data that works for every government. Each organization working to develop an open data initiative must look at what their specific goals and outcomes are in order to build a system that works for them and provides a tangible benefit for the people who are looking to put it to use (Lassinantti et al., 2014). The examples in this section show different initiatives from around the world and their outcomes.

### 2.7.1 Open Government Data in Europe

European states and municipalities have been leading the charge in creating OGD portals. Many cities and towns have on their own developed OGD portals for their citizens to use for a myriad of reasons. In Sweden two different municipalities implemented Open Data, each in their own way (Lassinantti et al., 2014).

The city of Stockholm kicked off their initiative by holding a hackathon called Open Stockholm, which focused on having individual developers and businesses build applications that could take advantage of an open data portal. At the conclusion of their implementation plan, Stockholm had a continuously-updating Open Data portal for their citizens to use. The municipal government in Skellefteå took a different approach. Stockholm largely saw OGD as a means to drive techno-economic growth. The leaders of the initiative in Skellefteå saw things differently, and sought to use an Open Data portal as a means to drive societal growth and communication between citizens (Lassinantti et al., 2014).

Both approaches had specific goals in mind, and as such overlooked a few key elements, as described by Lassinantti. Stockholm's approach was designed very much for the short-term, with their main goal being the development of applications to be used by citizens immediately. They neglected to include a method for citizens to report and correct information that they found to be wrong (Lassinantti et al., 2014). On the other hand Skellefteå's approach significantly slowed down the development process, and moved the top priority to be creating an open government rather than just a set of open data (Lassinantti et al., 2014). In doing so, they did not enable as many applications to be built on the platform for immediate consumption.

The Open Data initiatives in Greece highlight a few other aspects of Open Data that are key to understanding the elements of a successful implementation. In 2003, the European Union announced the "Directive on the re-use of public sector information". This directive introduced a common framework for government agencies to make their data public (Alexopoulos, Spiliotopoulou, & Charalabidis, 2013). Local and state governments in Greece use this platform to make their data available with one caveat: they do not seem to follow any consistent metadata standards. Their data is available, but it is largely not machine readable, making

large scale data analysis virtually impossible. Searching through this data manually is incredibly time consuming (Galiotou & Fragkou, 2013), meaning that many of the useful conclusions that could be found in the available data will never be seen.

Vienna, Austria also developed an Open Data portal of their own. The Vienna portal in the end was used as a template for the nation of Austria to build upon when building out their own OGD portal. It is not defined by any legislation officially, but it does have support from many local legislators and is backed by in-place policies. Unlike in the US, Austria does not have a FOIA-like act; the laws instead require that any government agency provide citizens with requested data in a timely manner so long as these requests do not impede the day-to-day operations of the department. An interesting outcome of the OGD portal is that the number of information requests from external stakeholders did not significantly decrease. Even with the portal, groups outside the government were still requesting information not provided by the portal. Department heads in government still viewed OGD as a beneficial endeavor despite this, as public administration itself is in fact seen as the main beneficiary of OGD. OGD in this case led to more information being available to the people for voting, diversity in opinions and debates, and increased civilian representation in their government (Parycek, Höchtl, & Ginner, 2014). The Vienna portal correctly addressed one of the key issues shown in the Grecian portals: availability of data in machine-readable formats. As can be seen in Figure 2.1, data is available for each entry in the "Roads" category in many different machine readable formats (CSV, JSON, KML, KMZ, etc.)

Titel	Veröffentlichende Stelle / Datenverantwortliche Stelle	Veröffentlicht auf data.gv.at am	Letzte Änderung auf data.gv.at	Format	Kosten	Lizenz
<a href="#">Straßengraph Wien</a> Der Straßengraph ist ein topologischer Graph des Abschnittsnetzes der Wiener Straßen. Die Netzlinien...	Stadt Wien / Magistratsabteilung 21 - Stadtteilplanung und Flächennutzung	09.12.2013	29.01.2016	wfs, CSV, gml, JSON, shp, wms, rss+xml, gif, jpeg, KML, kmz, png	kostenlos	cc-by- at-30
<a href="#">Straßenbelagsflächen - Wien</a> Darstellung der verschiedenen Straßenflächen (Fahrbahn, Gehsteig, ...) getrennt nach Belagsarten ( Gußasp...	Stadt Wien / Magistratsabteilung 28 - Straßenverwaltung und Straßenbau	10.12.2013	29.01.2016	wms, gif, jpeg, png	kostenlos	cc-by- at-30
<a href="#">Straßengraph ASFINAG - Wien</a> Der Straßengraph ist ein topologischer Graph des Abschnittsnetzes der Wiener Straßen. Die Netzlinien...	Stadt Wien / ASFINAG	09.12.2013	29.01.2016	wfs, CSV, gml, JSON, shp, wms, rss+xml, gif, jpeg, KML, kmz, png	kostenlos	cc-by- at-30

Figure 2.1: A snapshot of roads data in the Vienna repository. The table column labeled *formats*<sup>1</sup> shows the different ways that the data can be downloaded.

These four examples provide a few guiding principles for creating OGD portals, and what some of the outcomes of an initiative might be. The first conclusion is that initiatives need

<sup>1</sup>A format refers to a way that data can be represented. This portal allows users to download data in many different formats (CSV, JSON, KML, XML, etc.) which have varying degrees of machine-readability (Österreich, n.d.).

to be structured so as to actually provide a long-term benefit to the government developing the portal. Skellefteå and Vienna both showed that there are in fact long-term benefits from portals developed with these benefits in mind. The second is that the data needs to be in a format that is machine-readable and able to be linked back to the source (Lassinantti et al., 2014). Without key metadata, the data is effectively useless as it cannot be analyzed to create useful conclusions. Finally, OGD will not necessarily decrease workload on the government publishing said data in terms of FOIA-like requests, but rather in the end OGD will benefit the government directly by increase ease of access for the employees as well as the public.

## **2.7.2 Los Angeles GeoHub**

The city of Los Angeles, CA recently created a brand new Open Data portal called the LA GeoHub (City of Los Angeles, n.d.). The LA GeoHub is built upon ArcGIS (discussed in Chapter 2.8.2). ArcGIS allows citizens to search through the vast amount of data that the city has collected and build customized maps containing the data that they want to be displayed. The portal also provides an API, allowing for users and businesses to build applications using the data that the city has collected. The GeoHub provides an excellent example for how to implement an open data portal at a city level.

## **2.7.3 Albuquerque's Open Data Site**

As of right now Albuquerque is leading in the amount of information being shared to the public in New Mexico. Albuquerque is the largest city in New Mexico, but Santa Fe is the state capital and lacks open data for the public. Albuquerque gives a clear understanding of what information is most useful to the public and provides useful information for how to publish the data. Using their site, the team can understand the best way to publicize the data sets. This benefits the City of Santa Fe as well as ITPD in that the site provides a reference to creating a favorable website that anyone can use.

# **2.8 Use of Geographic Information Systems in Open Data**

## **2.8.1 What is a GIS?**

As is mentioned briefly in Chapter 2.5, Geographic Information Systems (GIS) are data visualizations involving the mapping of data via a geographic approach. Certain identifying information is used for this mapping process, including but not limited to:

- Latitude/Longitude

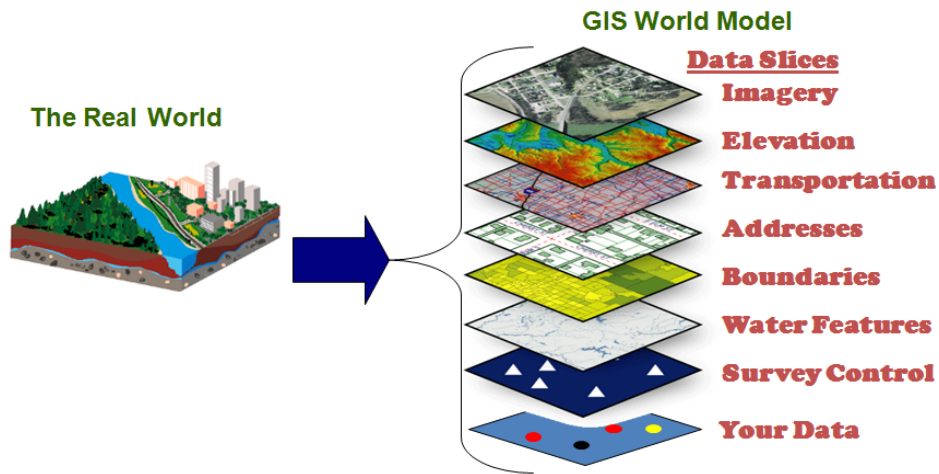


Figure 2.2: Example of World Represented in a GIS (State of Indiana, n.d.).

- Address
- Zip Code
- GPS

Geographic Information Systems (GIS) first became conceptualized in 1962, when Roger Tomlinson worked with the newly formed Canada Land Inventory (CLI). They were at the beginning stages of planning a large scale mapping which would result in about 1,500 new maps. Tomlinson was convinced that computers could be used to automate the map creation and analysis, and so wrote up his proposal for Geographic Information Systems (Esri, 2012). Since then, technological improvements have made them far easier to implement, update, and use to compare relevant data sets in one centralized mapped location with relation to their surrounding geography.

Data sets are uploaded to a central database location, and this communicates with the GIS infrastructure. The data set will tie location data to each object in the set, which the GIS will use to place the objects on a map. The information is shown as layers that can be switched on or off depending on the desired view. This allows for ease of access in comparing desired variables to determine what (if any) relationship they may have. Items that may not be seen as impacting one another have the possibility of displaying surprising trends that can influence city planning and strategy. This world model is exhibited in Figure 2.2, showing the different layers that represent the real world in a GIS.

GIS systems can be implemented internally within an organization, but in the case of cities are more frequently made publicly available. The current trend within technology is to increase transparency, as the public is more likely to support a project that they can visibly see the outcomes of. Making the data available to all will increase the people's trust in the institution that produces the data.

GIS tools provide the means for combining data that might normally be displayed in



different mediums to be shown together for the purpose of drawing additional conclusions. It can also be used to show change over time by examining one or more factors at individual time slices, then updating the viewed time frame to see how it is progressing (*GIS (geographic information system)*, 2011).

## **2.8.2 GIS Tools and their Implementations**

### **MapInfo**

Pitney Bowes' leading GIS software is the MapInfo series, including such products as MapBasic, MapXtreme, MapMarker, and of course MapInfo and its extensions (*Location Intelligence*, n.d.). WPI has been involved with MapInfo in the past when a project team working in the London Borough of Merton was tasked with joining the implementation of the MapInfo GIS software that they had recently purchased. This team first reviewed a strategy guide produced by the Greater London Authority on implementing a GIS, and then moved into the next stages of creating samples for the Borough to learn from. The team trained Merton employees in the use of MapInfo and how it can be optimally used to analytically display data for their area (Bradbury, Fox, & Frank, 2002).

### **Quantum GIS**

An open source desktop GIS software, Quantum GIS (QGIS) has become a solution for the more expense conscious users who are looking for a simpler solution that will not cost them thousands of dollars in licensing fees. Some WPI Project Centers that used to focus on MapInfo have now started moving towards QGIS as a means of reducing costs and licenses required. As MapInfo offers more robust mapping capabilities, some project centers use the MapInfo maps combined with QGIS extensions (*Quantum GIS*, n.d.).

### **Esri ArcGIS**

Esri is one of the leading providers of GIS software with their ArcGIS series. Catering to a broad range of audiences, from individual companies to entire cities, ArcGIS offers extensions for most conceivable GIS usages. They offer a desktop software and a server side software to allow for reaching out to a range of users, and among the multitude of add-on modules are Network Analysis, Spatial Analysis, and 3D Analysis, to name but a few. As mentioned in Chapter 2.7.2, Los Angeles recently unveiled their collaboration with Esri to create the LA GeoHub, a new site for all of their open data to be featured in one place. The new platform that will be getting built out in Santa Fe will be focused around the same principles (Esri, n.d.).

## Geocortex

Latitude Geographics extends upon the Esri ArcGIS infrastructure to create their own product, the Geocortex GIS viewer. A partner company of Esri, Latitude has been expanding upon the mapping capabilities that Esri provides since 1999 (Esri, n.d.). This is the foundation of the current GIS software being implemented by Santa Fe, though as mentioned in Chapter 2.5, the Santa Fe system is significantly outdated.

Santa Fe currently has layers broken down to:

- City Information
- Property Information
- Districts
- Facilities
- Transportation
- Recreational
- Natural Resources – Floodplain – Topography
- Utilities
- Imagery

Depending on the layer, most of the recent data sets range from 2009-2012, with nothing newer dated for comparison. This does lead to a slight gap in the availability of public information. However, the Santa Fe GIS Services team have been working on building a completely updated interface for their public GIS, which will hopefully address most of the older information in the Geocortex system.

# Chapter 3

## Methodology

During the initial background phase, the project was going to be directed towards citizens and how they might view an open data portal. The project changed early during implementation to focus on showing a proof of concept open data portal to the City of Santa Fe government to help progress a movement toward centralized open data within the Santa Fe government. The shift in focus resulted from the need for more data to move to a format conducive to open publication. In order to provide fully open data in an easy to use manner, all the city departments need to be brought on board with the open data initiative. Having full participation from the city departments advances the original goal of providing the citizens a way to access government data to increase ease-of-access to city resources and increase public participation in local government. The project was broken down into the following primary phases:

1. Gain access to key datasets and determine the data that best suits the interest of the stakeholders
2. Create a repository from the data sources collected as part of Phase 1, providing a resource for building online applications.
3. Create a prototype dashboard for viewing the data, combining new and existing data to bring light to interesting patterns and uses.
4. Determine resources that ITPD can use for future data portal development with the help of the city government.

The project's main focus was on creating the data repository as well as the prototype dashboard. The project produced the following major deliverables:

1. Proof of Concept data repository for Santa Fe, NM
2. Multiple prototype dashboards
3. Recommendations for future work

#### 4. Thornburg Grant application to support open data movement in Santa Fe

Figure 3.1 outlines project objectives. Additionally, Table 3.1 shows the projected timeline for the project.

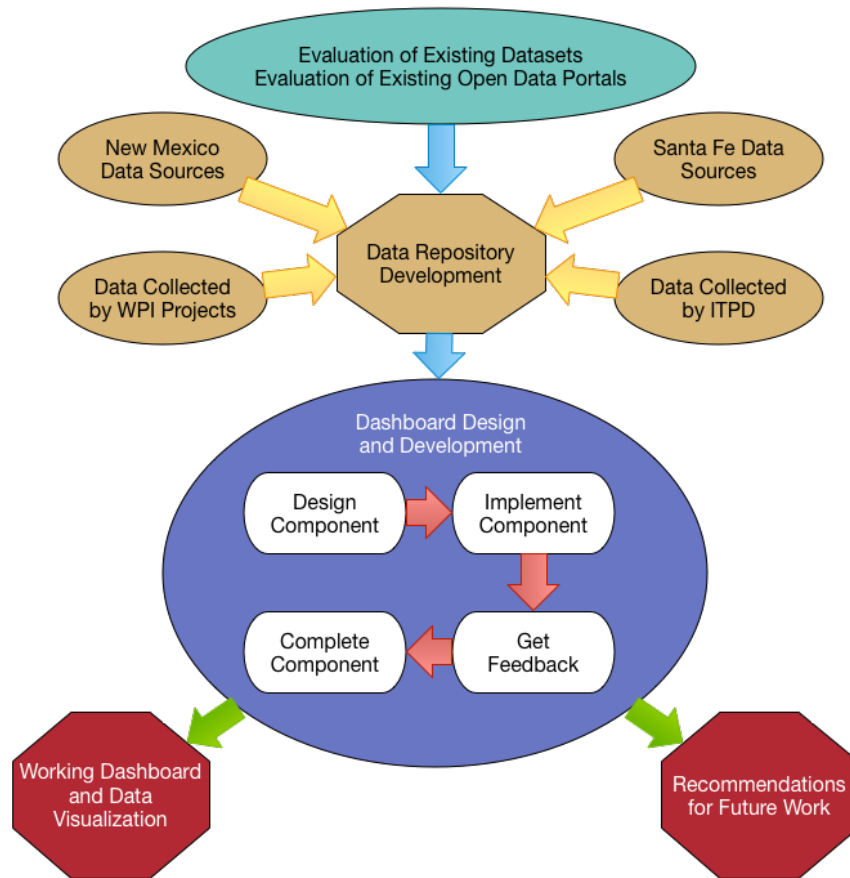


Figure 3.1: Project Flowchart

Project Phase	Task	Week							
		Prep	1	2	3	4	5	6	7
Assess and Access Datasets	Collect list of potential datasets to include	█							
	Determine viability of using each dataset		█	█					
	Select final datasets for initial repository				█				
Construct Data Repository	Assess potential platforms for data repository		█	█	█				
	Add data to selected platform				█	█	█		
	Document platform API						█	█	
Create Prototype Dashboard	Assess potential platforms for dashboard				█	█			
	Construct dashboard widgets						█	█	
	Document dashboard construction and maintenance							█	█

Table 3.1: Project Timeline

Some resources that were useful in completing this project were the Open Knowledge Open Data Handbook (Open Knowledge, 2016), the New York Open Data Handbook (State of

New York, 2015), as well as various resources from Santa Fe and other cities. Each phase is a different part of the project and did not necessarily occur in sequential order. The following sections discuss the methods used and plans made for achieving the stated desired outcomes.

## **3.1 Phase 1: Access and Assess Datasets Available in Santa Fe**

### **3.1.1 Accessing Data**

The team identified six main areas of interest for the prototype: elections, education, crime, health, tourism, and roads. Election information was the primary interest of the Thornburg Foundation, an organization based in the Santa Fe area that has shown interest in good government reforms. Inclusion of the election data allowed the team to apply for the General Operating Support Grant from the Thornburg Foundation, which supported the WPI Santa Fe Project Center in continuing its work in the area. The other five areas, as well as election data, were of greatest interest to It's The People's Data. The key pieces of information acquired were: what data is collected; how that data is collected, stored, and updated; and how the groups with the data can make datasets available to the public. Accessing the datasets was the main priority to complete the ITPD's objectives. Some city agencies were already making data available online which expedited accessing certain datasets. This showed that datasets were being created in the government and that they were also being updated. Once the data had been acquired, the team worked with the City of Santa Fe's Information Technology & Telecommunications (ITT) Department to help incorporate the data into applications.

### **3.1.2 Assessing Data**

Once ITPD and the team had access to the data, the next step was to assess that data to determine which datasets had valuable information. The team used data found from the six areas of interest to show the proof of concept dashboard to the city government. To determine how to best portray the data, the team interviewed five city employees: two members of the ITT Department, the City Clerk, and two members of the Tourism Santa Fe Department. These interviews helped the team gather necessary information in order to create future plans for the city to continue the open data initiative. The employees provided information on how they publicize government data and if their process needs to be changed in any way (Martinez & Barsanti, 2016) (Vigil, 2016). The data were then assessed to determine how to best display them in the repository and dashboard.

## 3.2 Phase 2: Create Centralized Data Repository for Santa Fe

During the second phase of the project, the team combined Phase 1 data into a publicly accessible repository. The data we found in phase one was spread out across multiple sources and services, making standardized access difficult for a user or an application developer. The repository provides three main functions:

1. Data collection from sources both internal and external to Santa Fe
2. Data serving in a variety of machine and human readable formats
3. Direct access to the data, allowing applications to pull updates as soon as they are available

Ideally the city will be able to purchase an off-the-shelf solution to provide these services in the long term. Implementing a large scale repository is a substantial undertaking, and as such did not fall under the scope of this project. The team analyzed six platforms that would be useful in creating such a repository, namely:

1. Socrata (closed source, requires annual fee)
2. Opengov (already in use in the city for publishing financial information)
3. CKAN (partially free and open source, paid plans available)
4. DKAN (partially free and open source, paid plans available)
5. Junar (closed source, requires annual fee)
6. OpenDataSoft (closed source, requires annual fee)

Each of these platforms provided similar services for varying prices. None of them were completely free (if support was desired), and some required large annual fees in addition to hosting fees. To assess which of those platforms would be best for the city, each of the potential platforms were evaluated in accordance with the criteria shown in Table 3.2.

Those criteria provided a reasonable method of judging which solution would work best for the city of Santa Fe. Once a platform was selected, the city and It's The People's Data approved the selection and provide resources to implement it.

<b>Criterion</b>	<b>Explanation</b>
<u>Cost</u>	The solutions need to balance cost and capabilities to not stretch city monetary resources too much: the city of Santa Fe is currently operating at a large budget deficit (Johnson et al., 2016).
<u>API (application program interface) Support</u>	The repository needs to provide an API to allow applications to access the data that are hosted by the repository. This would allow for applications and dashboards to be developed using the data contained within.
<u>Data Source Support</u>	The platform that the portal is built on will need to be able to link to and update the data from antiquated server technology that the city utilizes if good data is to be provided to the people.
<u>External Update Support</u>	The portal needs to have some form of support for adding in more data from sources external to the main datasets to allow for citizens to interact with the portal. A key component of any open data portal is the ability for citizens to submit updates to data if they see anything that is incorrect, or if they wish to add data to the portal themselves.

Table 3.2: Criteria for selecting a platform for creating the open data repository prototype.

## Data Security

Especially when starting from scratch, security is extremely important in building a new digital system, as any device connected to the internet is vulnerable to attacks.

While some attacks may just be vandalism, it is important to employ preventive measures instead of reactive measures because of the nature of an open data platform, which relies on trust. Security is required in order to fulfill the users' and government's expectations. Users expect an open data platform to provide instant access to up to date accurate information. Governments as well expect these platforms to not compromise their own security and to please the people. Some governments even plan on using it for their internal information sharing. If security of an open data website is compromised, it can erode users' trust in the government or government's trust in open data, which in turn can lead to the changes being undone through the process of addressing the security issues.

In order to deliver the data in a secure manner, certain data security controls were utilized, which were described by the Open Data Center Alliance as being key to protecting information in the cloud (Caballero et al., n.d.):

- Access control: Controlling who or what can access which data when, and in what context.
- Information classification: Identifying the sensitivity of the data and the impact of unauthorized access, as well as the organization's need for data integrity and data availability.

- Data encryption: Applying the appropriate encryption techniques to enforce data confidentiality requirements.
- Data masking techniques: Further increasing data security in the cloud through anonymization and tokenization.
- Security information and event management: Tracking and responding to data security triggers, to log unauthorized access to data and send alerts where necessary.
- Backup, archiving, and deletion: Identifying backup requirements and how those relate to secure storage and secure destruction of data when it is no longer needed.

The team followed the requirements set by the data providers and the standard policies employed by the City of Santa Fe's ITT Department. There was no need for extraordinary security precautions since no sensitive data was encountered. Modern security programming practices published on credible well-established websites and taught in WPI classes were employed throughout the entire development process in order to guarantee the permanence of the data.

### **3.3 Phase 3: Create Prototype Dashboard for Santa Fe**

For the final step of the project the team created a prototype dashboard to show the effectiveness of open data. Dashboards are user-facing views that compactly display key information that the user desires. This dashboard uses the repository created in phase two as a backbone to publicize the data. The team looked at four dashboards that had abilities to present the data in an ideal format:

- The Dash
- Power BI
- Jaspersoft
- Tableau

In considering dashboards, the ability of each one to present the data in a useful format was weighed. The ideal dashboard would have all of the data displayed in one place that depends on a data repository for the specific information that it needs to show. The data that the team, in collaboration with ITPD and the City of Santa Fe government, acquired could not be interpreted by an outside user without displaying the data in an understandable format.

The Esri ArcGIS platform was going to be looked at because Santa Fe was in the process of implementing this new GIS platform for the city, but the platform was unfortunately



not pushed live while the team was in Santa Fe. This new system was going to be a huge step forward from the old GeoCortex Internet Mapping Site that they had in place previously. The new system, an Esri ArcGIS Server platform, featured top of the line integration with all of the leading methods for displaying data. Some of the data that the team collected could be incorporated into the city's GIS platform. As a result of the platform not being finished during the project's duration, the team had to focus on what other cities have done with the Esri ArcGIS to make any recommendations.

### **3.4 Future Recommendations**

Once the team was finished with the creation of the repository and the dashboards, recommendations for future improvement were given to ITPD and government workers for Santa Fe. These recommendations will be used to continue the open data movement within the city. ITPD can use these resources to continue their movement from outside the government and show how transparency is valuable to the people. The Santa Fe ITT has welcomed these efforts which they see supporting their case for more open data. There are multiple resources from cities to software agencies that can help inform Santa Fe on ways they can advance their transparency. This will be a valuable resource once the team has left Santa Fe so that both ITPD and the Santa Fe City Government, together, can continue their efforts to establish a true city data repository for government transparency.

# Chapter 4

## Findings, Analysis, and Outcomes

This chapter discusses what the team found during the weeks in Santa Fe. It will report on the findings from research on each element in the proof of concept in the following order:

1. Datasets: the data that was acquired and how it was used
2. Portals: the portal platforms that were researched and which one was chosen
3. Dashboards: the dashboard platforms that were researched and which one was chosen

### 4.1 Overall Findings on Dataset Acquisition and Utility

Once arriving at Santa Fe, the team's first goal was to find valuable datasets for the city. In talking to ITPD and seeing what the city provided, the team had six initial datasets that were useful in a proof of concept open data portal: Election, education, crime, health, tourism, and roads. The datasets were scattered in a variety of locations, both physical and digital, which made access difficult at first. The data also came in various forms, maps, surveys, excel sheets, etc., which could impair interpretation if not given a proper description for users. The metadata for each dataset needed to be looked up which was time consuming because not all of the information was available. Something Santa Fe can do in the future is release metadata with their actual datasets, following the example set by the City of Albuquerque, which created a form that departments fill out to inform the public about their metadata. Each form has contact information, data refresh, and dataset dictionary and descriptions for the provided formats. The datasets that the team acquired are:

- Campaign Finance Reports
- Campaign Contributions and Expenditures

- PARCC Standardized Test Results
- School Grades
- School Proficiencies
- Attendance Report
- HeadStart Report
- Crime Report Data
- Healthcare Coverage
- Public Health Indicators
- US Health Status
- Santa Fe Street Conditions

The full table of datasets can be seen in Appendix D in Table D.1.

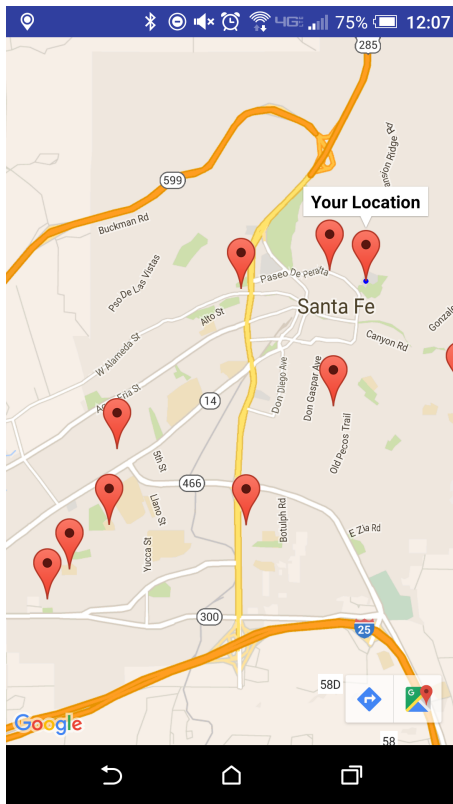
#### **4.1.1 Election Data Findings**

- (a) Two sites are being used for Campaign Finance information but have different monetary values because of different update intervals and how each site organizes financial information.
- (b) The City will be using Voting Information Project (VIP) Tool to inform voters about voting locations. VIP was not working when the team looked into it but will be operational when there are elections in Santa Fe.
- (c) The city of Santa Fe does not recognize electronic signatures as a viable form of notarization.

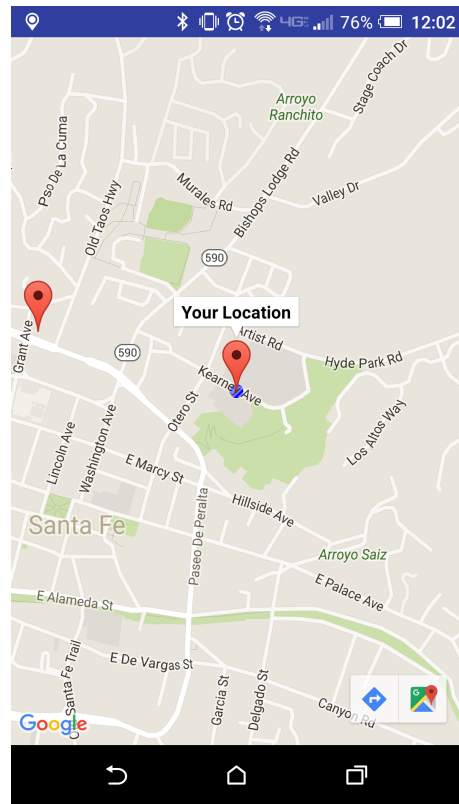
The team collected information from the New Mexico Secretary of State's office about election information. On their website, <https://www.cfis.state.nm.us/media/>, there is a small amount of election information. One reason for the lack of information may have to do with the wording in one Santa Fe law, Campaign Practices 1-19-32 B, where if submitted electronically the document will be uploaded electronically. This could mean that if candidates submit expenditures via paper, those documents may not be uploaded for access on the internet. A work around to this, to access candidate expenditures that were missing, was through a website call <http://www.followthemoney.org>. Going through their site, a person can look at the election overview for a particular year in a state. For New Mexico, the best data that followthemoney.org has is currently 2014. They are actively updating 2016 data as more information is released to them and 2015 had no elections. There are discrepancies between the two sites: The monetary

values are different for the same candidate. After talking to a worker for followthemoney.org, we found out that the site uses a different classifications to create the monetary value for expenditures. We do not know how they classify the financial information that they acquire, but this site has been a reliable resource for candidate expenditures.

The City of Santa Fe's website also provided more information regarding election data. Santa Fe provided the locations of the twelve polling locations that were used for the general election on March 1, 2016. Santa Fe is split into 4 districts and the community can vote at any polling place. With this information, the team created an app that directs the user to a polling site of their choosing. The app uses your location and pinpoints the 12 polling sites around Santa Fe. The user can then click on the closest polling site and then go to Google maps which provides the user with directions. This app came from the idea of the Voting Information Project (VIP) app that was provided to us by our sponsor. After looking over the app, no information was being actively shown and that is why the team created this application. This app was built to demonstrate how translating a specific citizen need for data into a practical application that addresses that need can be very simple. The VIP app will be active again once voting opens in New Mexico.

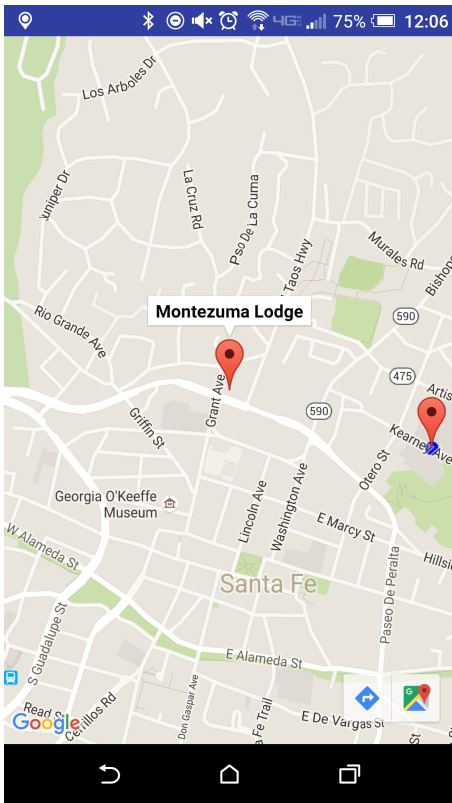


(a) App First Step

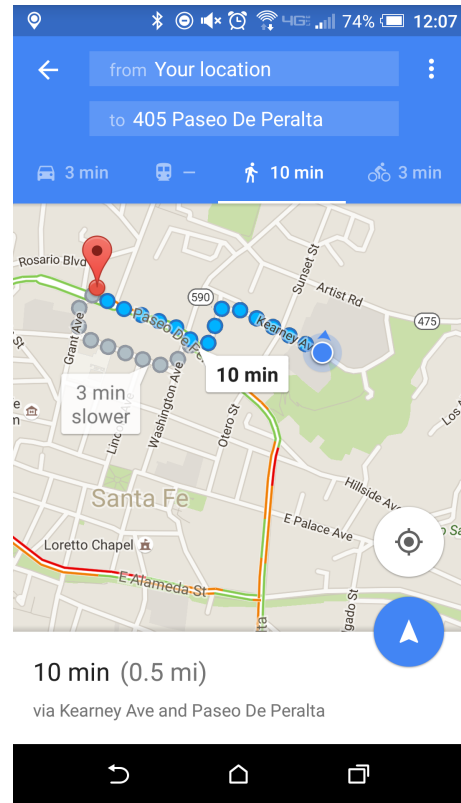


(b) App Second Step

Figure 4.1: App Finds Locations of Voting Polls and Your Location



(a) App Third Step



(b) App Fourth Step

Figure 4.2: Click on Closest Polling Location and App Gives You Directions

In the City of Santa Fe, the City Clerk is the main Custodian of Records and is the contact person for several pieces of data that involve the city. The City Clerk for Santa Fe is Yolanda Vigil, and she has provided information on how she helps manage the election data. Two main documents for election information are the Code of Ethics and Financial Disclosure Statement forms that are filled out by all government workers (Vigil, 2016). On the <http://www.santafecountynm.gov/>, most department workers have easy access to this link. Going into the Santa Fe city site, there is a section that shows the Financial Disclosure Statements but not the Code of Ethics forms. Yolanda sends them a CD with the forms that they fill out and they send those forms back in to Yolanda filled out. However, because the city does not recognize the electronic signatures, candidates must sign them by hand and turn them in to Yolanda so the forms can be scanned, put up onto the city website, and then the hard copy gets filed away. This is a roadblock the team is facing and can not do much about and leads to the question why this is not all done digitally to make uploading, reading, and accessing forms easier. There are ways to read scanned PDFs, but the software needed is expensive and the city would need approval and new practices would need to be implemented. In our weekly meetings with ITPD, information came up that in the past someone purposely deleted electronic election data. This, of course, is a huge problem with digital items and if there are no backups everything can be lost. This shows that even though digital systems can be very reliable, people can have malicious intent and destroy valuable information.

### 4.1.2 Education Data Findings

- (a) The NM Public Education Department uses various resources such as Tableau, PDFs, and maps to display school data. If data was shown in one platform, it would be easier to combine and correlate datasets.
- (b) One important piece of information is the PARCC results for the students showing a decrease in the 'proficiency and above' level in higher grade levels, as these are the standards the state is trying to meet.
- (c) Early childhood development has also been gaining information from various organizations. This data has been put up on <http://www.santafedatahub.org/> in dashboard formats using Tableau. Other dashboards on this website include school, career, and other youth information.

Education is one of the most important features for any community. The level of education the community provides can help the city grow as well as give students a way to reach bigger and better futures. The New Mexico Public Education Department (NMPED) provides a vast amount of data that citizens can use to see how well students are performing and how well each individual school is performing. For instance, NMPED publishes testing information (New Mexico Public Education Department, 2015a) for grades K-11 with the most recent results being from the 2015 school year. Looking at the data, one can learn that there has been a decrease in the number of students reaching the 'proficiency and above' level in different academic areas. In relation to this, there has been a decline in the number of students taking the test in the past years which leads to growing problems for the school district: the percentage of students who are poorly prepared to move to the next level in their academic careers has been increasing.

One specific test is the PARCC, Partnership for Assessment of Readiness for College and Careers, which is New Mexico's standardized test designed to see if students are on track to be successful in school, college, and in their careers. PARCC is split into five levels that show the students' performance level, and a level three score or higher means that they are able to move up. New Mexico has released a dashboard (New Mexico Public Education Department, 2015b) using Tableau that shows the percentage of students that were put into each performance level. Another resource that was brought up during our final presentation were dashboards created for early childhood development and school data located on <http://www.santafedatahub.org/> that also use Tableau. This website provides an opportunity for the city and ITPD to look at how dashboards can be created as well as get help from individuals who have experience working with Tableau. Looking directly at Santa Fe, there is a vast majority of students that fall below level three for performance for many grades. The schools in Santa Fe also get graded based on test scores and other factors that are for promoting learning: Only 4 out of the 29 public schools in the Santa Fe area have an A rating. This presents a very big problem for the city, because poor ratings can lead to less funding for the schools which in turn further erodes then quality of education for the students.

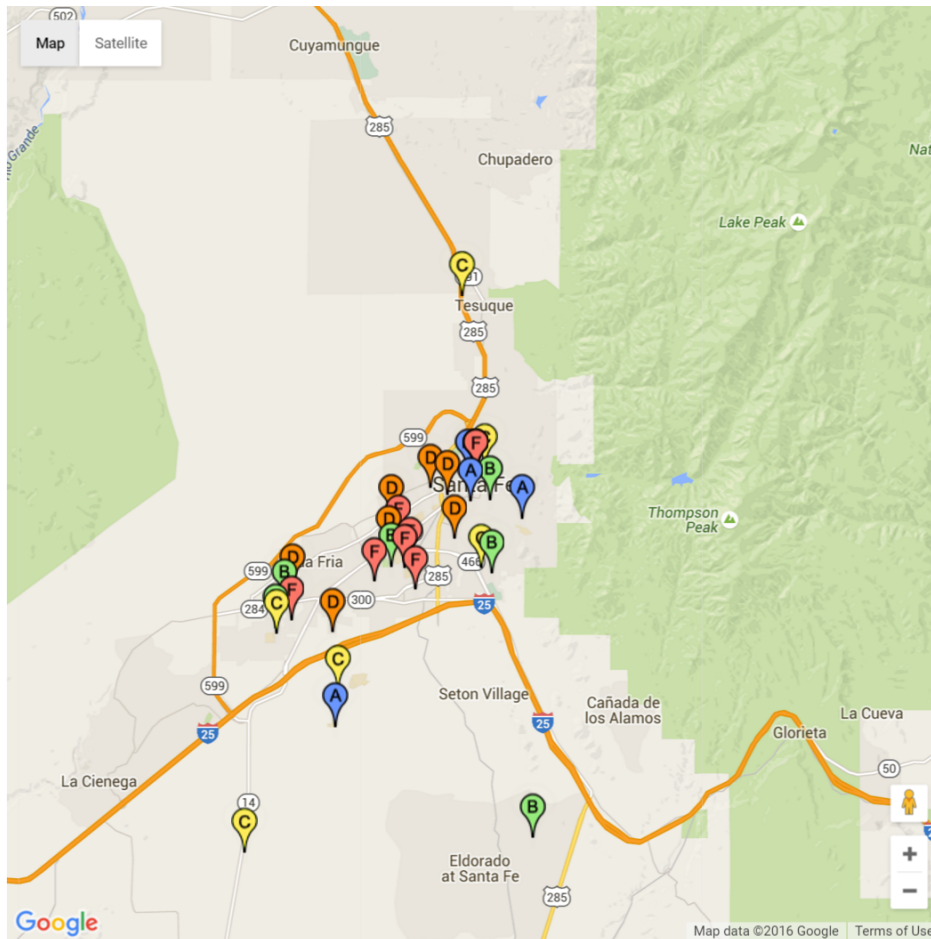


Figure 4.3: School Grades Map

### 4.1.3 Police Data Findings

- (a) The Santa Fe County Sheriff uploads crime information to crimereports.com showing crime information as well as sex offenders in the city.
- (b) The County Sheriff department could give access to the city police so the city can upload their crime information.

Crime data has also been provided to us with the help of ITPD. This information comes from a website called crimereports.com. The user can search any address to find crime information that has happened in that area. Based on the map location, the site gives a list of crimes that have happened. The user can click on the crime to pull up additional information. The information shows the type of crime and the time the crime was committed. The site also puts up information on sex offenders within that area. There are two ways the data can be published externally for a dashboard setting: using the map that is provided, or exporting a list of data that takes information straight from the map. This website is currently updated by

the County Sheriff's office. The Sheriff's Office should give access to the City of Santa Fe Police Department so that they can also update the map with reports they respond to. This would be very useful to the Santa Fe Police department seeing as, based on what we have been told, is that their software is not fully up to date and there is a difficulty gathering any data from their system.

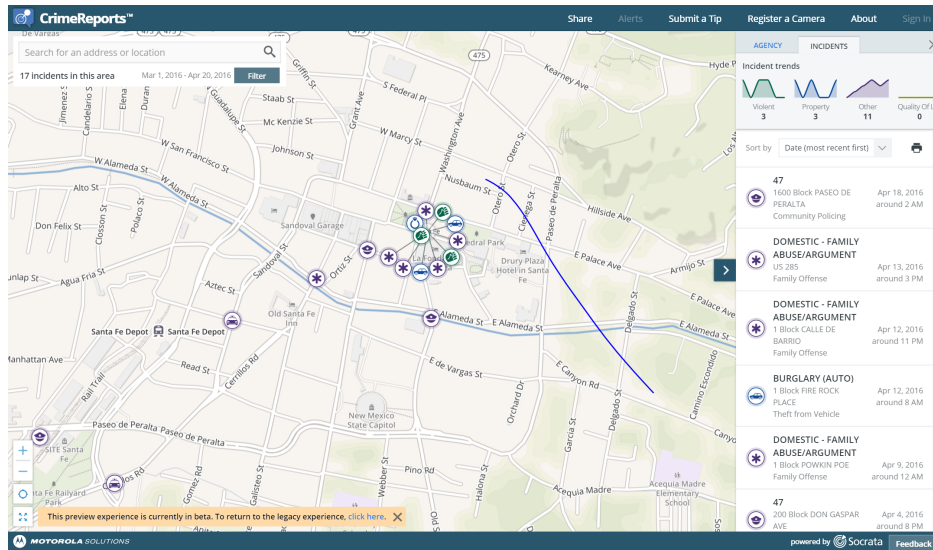


Figure 4.4: Crimereports.com Socrata Dashboard

#### 4.1.4 Health Data Findings

- (a) Health information is difficult to acquire due to HIPAA, which protects "individually identifiable health information." The team's objective was not to share individual information, but HIPAA can stop health data from being accessed.
- (b) New Mexico has two sites that release some health information: NM Indicator Based Information System and NM Health Department. However, both are either hard to navigate or do not give hospital specific data.
- (c) The team looked at the United States Centers for Disease Control and Prevention's (CDC) website to get city specific data from census data that is collected.
- (d) American Hospital Directory and American Hospital Association provide useful hospital information but require a subscription to look at any of said data.

Health information is difficult to acquire. One of the big problems facing the team was the limited types of health related datasets that were publicly available due to information being withheld from the public. One key law that prohibits individual information from being shown to the public is HIPAA, Health Insurance Portability and Accountability Act. HIPAA provides data privacy for "individually identifiable health information:" such as name, address,



birth data, and Social security number so that the private information cannot be used by a third party for corrupt uses. The team's objective was not to share individual information, but with a lot of the information kept under lock and key there is no easy way to acquire specific datasets.

There are two sites that the state of New Mexico has provided some information: [nmhealth.org](http://nmhealth.org) and [ibis.health.state.nm.us](http://ibis.health.state.nm.us). The first site goes to the New Mexico Department of Health website that has some general health statistics for the state but does not go into detail for specific counties or hospitals. County and hospital data would be very useful for the general public to see where the best type of doctor is or what the best hospital is for a certain procedure. Certain hospital data can be acquired through the American Hospital Directory (AHD) or the American Hospital Association (AHA). These require fees but seem to provide very useful information about hospitals. The other site is New Mexico's Indicator Based Information System which is New Mexico's public health data resource. The site contains numerous datasets on various health topics such as risk and resiliency factors, health care services, and health outcomes. Some of the datasets provided are in graph form and are easy to understand while others are based on a query the user creates. Using the site is very difficult for a first time user and does not have a helpful guide on how to create the best type of query with the data that is provided.

Another useful resource for some health pertinent information was from the US Census factfinder site and the CDC. This factfinder site provided health insurance data based on several different identifying subjects such as race, age, education, and jobs. This data can be narrowed down to a specific county and provides a percentage of people who have and who do not have health insurance. The CDC also has good information for the state based on a survey that was done over the phone. The CDC used the BRFSS, Behavioral Risk Factor Surveillance System, survey to get a better understanding for the health across the nation. The data provided is within an excel sheet that has a general question that can lead to a series of breakout questions. In New Mexico, there were 50 topics that were asked ranging from age to drug use to general health information and even to active military duty. This survey has some useful information about the health of New Mexico, the team just needs to filter out the unnecessary data.

#### **4.1.5 Tourism Data Findings**

- (a) The team first looked at how the Venice Dashboard worked with using code to scrape hotel sites to figure out hotel booking data. Unfortunately, the scraper did not work for Santa Fe.
- (b) The team contacted Santa Fe Tourism to get other types of data. They are in the process of acquiring a market analysis system for hotels that can eventually predict future booking information, but right now they have nothing digital that can be shared.

Seeing as Santa Fe is a very busy tourist destination, the team looked to see if tourism

data could be acquired using techniques from past projects. After talking to our sponsor and the ITT department, the team concluded that they would like to have a way to know hotel booking information. The Venice Dashboard that was made for the Venice Project Site has a widget that tells you how many rooms are available. After talking to Tomaso on how the widget works, he provided code he uses for the Venice Dashboard. The code is a script that goes on to booking.com and pulls information from the site. This would be useful, but the code currently does not work due to booking.com upgrading their website which has the script malfunction when trying to find data.

With the help of ITPD, the team contacted Tourism Santa Fe for more information about the hotel booking feature their site offers (Delgado & Randall, 2016). From our meeting, Santa Fe Tourism does not have a lot of statistical data about hotel bookings. They are in the process of getting an outside market intelligence system to gather lodging data. Hotels who participate will get data that, once enough is collected, can be used to make future predictions. Along with that info, the team is getting access to the Rocky Mountain Lodging Report that has monthly statistics for hotel data. The Rocky Mountain Reports are monthly: therefore, the reports are do not represent real time lodging information. Another service is the STAR benchmark which reports a hotel's performance. This is a subscription based benchmark that comes in monthly, weekly, or daily reports. We need access to both reports to see what useful data can be obtained. For other information such as tourist destinations, websites often link to their city tourism page because there are not many datasets for attractions.

#### **4.1.6 Roads Data Findings**

- (a) The team was given excel sheet describing the roads conditions in Santa Fe.
- (b) The Public Works Department could add geographic identifiers in order to incorporate the data on a map for a better understanding.

With the help of our sponsor, we acquired information for the roads throughout Santa Fe. The information provided gave us ratings for each road, district location, the estimated cost to fix the road, and more information all located in an excel sheet. This can be valuable information for drivers so that they know what roads are bad to drive on. This information can also be expanded on during road maintenance. One example of road's data being updated for the public is shown in the Los Angeles Geohub with their web app called Street Wize. Street Wize uses road data to show when there is road maintenance going on to let drivers know ahead of time that a road may be closed. The application LA Geohub has can be used by anyone at any time and actively updates when the maintenance time is known. This is a practical way the excel sheet can be used in Santa Fe, it would just need to have additional resources to let users know when certain areas are being worked on.

## 4.2 Portals

### 4.2.1 Overview

From the list provided in Section 3.2 of the Methodology Chapter, the team looked at the feasibility of six portals for hosting the proof of concept data repository. Table 4.1 briefly describes each considered platform.

Software Name	Website	Price	API	Data Sources Supported	Public Updates
CKAN	<a href="http://www.ckan.org">http://www.ckan.org</a>	\$0 for software, hosting fees vary	Yes	All	Yes
DKAN	<a href="http://www.nucivic.com/dkan/">http://www.nucivic.com/dkan/</a>	\$0 for software, hosting fees vary	Yes	All	Yes
Socrata	<a href="https://www.socrata.com/">https://www.socrata.com/</a>	Call for Price (too expensive)	Yes (requires paid key)	All	No
Junar	<a href="http://junar.com/">http://junar.com/</a>	\$1000/month	Yes	All	No
OpenGov	<a href="http://opengov.com/">http://opengov.com/</a>	Unknown (city already purchased)	Yes	Finance only	No
DKAN	<a href="https://www.opendatasoft.com/">https://www.opendatasoft.com/</a>	Call for Price	Yes	All	No

Table 4.1: Comparison of various repository software types.

Out of the table, and what the team researched, the best platform on the market is Socrata. The user interface is helpful to first time users and a number of datasets can be put onto the software for use. The downside to this platform is that it is very expensive and the team would need excess funds to work with the platform. Even though Socrata was the best on the market, the price range was too high for the team to use as a proof of concept. We then looked at the two platforms which are free to use (aside from hosting costs): DKAN and CKAN.

CKAN was the most comprehensive of the free solutions, but left room for improvement. The largest issue with CKAN is that it has no support, meaning that if we encountered an error there was no one to call for help. CKAN itself does provide error documentation, but doesn't provide instructions for fixing issues that might occur, leaving the user to find answers on support forums. DKAN provided a more satisfactory and simplified setup experience

than CKAN, however DKAN gave the team issues immediately when importing the first dataset. DKAN also seemed to struggle with the size of the datasets that were being imported.

## 4.2.2 CKAN

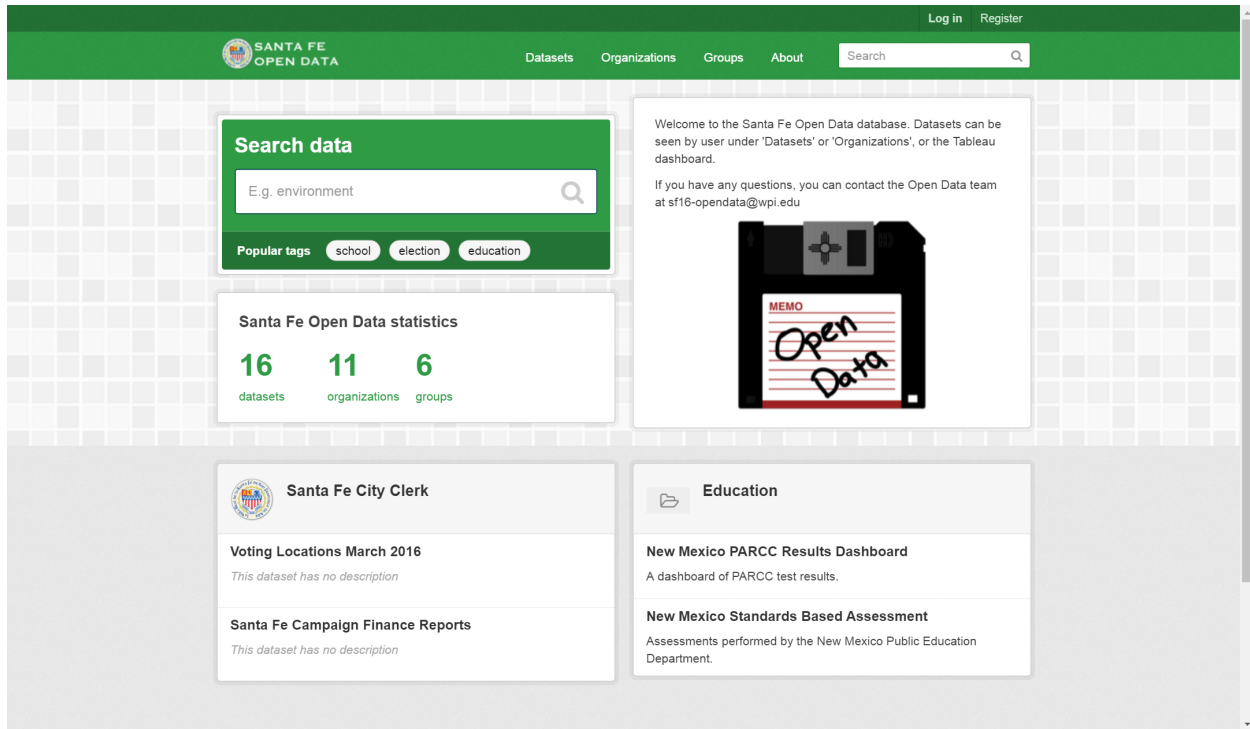


Figure 4.5: CKAN Homepage


Both CKAN and DKAN provide excellent services that the team could use for a proof of concept platform. In the end CKAN was chosen for our platform due to its relative ease of use and compatibility with the file formats that the data was provided in.

CKAN organizes data with three different identifiers: organizations, groups, and tags. CKAN organizations are groups in the real world that provide data, like the Santa Fe City Clerk's office or the New Mexico Department of Transportation. Datasets are owned by organizations. Once entered, datasets can be further categorized with tags. A dataset can have any number of tags applied to it. Tags behave effectively as search keywords, meaning that if a user searches for the word *education* they'll see all datasets tagged with *education*. Finally, datasets can be categorized by groups. Groups are created by the user, and are used to collect datasets from different organizations into one place.

The repository itself can be found at <http://santafeopendata.itsthepeoplesdata.org>. This web page is its temporary home, until the city rolls our repository into something more permanent.

### 4.3 Dashboards

Several dashboard services were researched to determine the ideal method of displaying the compiled data. The following four dashboards were identified as possibilities, each featuring slightly different options that would impact how the finished product was presented. Taking into account factors such as ease of use, data sources available, and variety of tools for manipulating data, the team analyzed each of the dashboards and concluded that Tableau Desktop would be the best for balancing user-friendliness with extensive capabilities for linking and manipulating data.

DASHBOARD COMPARISON 





CRITERION	 TABLEAU	 THE DASH	 POWER BI	 JASPERSOFT
Price	\$2,000 / user + \$500 / user / year	\$795 / user / year	\$120 / user / year	\$4,100 - \$154,000 / year
Database Support	●	●	●	●
Web Source Support	●	●	●	●
Social Media Support	●	●	●	●
Connects to CKAN	●	●	●	●
No Storage Limits	●	●	●	●
Customizable Data Refresh	●	●	●	●

Figure 4.6: Dashboard Comparison

See Appendix E for a more detailed table comparing the dashboard options. The following section outlines in greater detail each of the dashboards that were tested, starting with Tableau and how it ended up being used.

#### 4.3.1 Tableau

##### Overview

Tableau’s visualization platform is a business intelligence tool aimed at connecting a user to a data source and helping create views for displaying data in meaningful ways. It can connect to databases, take in local files, and perform a variety of other server operations. The data are then displayed in sheets within Tableau, with each data source receiving its own workspace where

different visualizations can be made. The data source itself can be modified within Tableau to add calculated columns, and it offers both live connections and the ability to extract the data. The benefit of Tableau is that dashboards can then be created combining information from a few or many data sources in the workbook as desired. These dashboards can also feature web connections embedded directly into them, so that a previously made page can simply be added to the dashboard with no additional work required.



Figure 4.7: Example Dashboard Using Tableau (published online)

## Establishing Connections

Since Tableau was the chosen dashboard software, the team created a connector that would pull data from CKAN and import it into Tableau. CKAN stores data in a datastore that is unique to its platform, which meant that a custom solution was required to establish a connection between Tableau and the repository. This was one of the benefits of Tableau, which, while also featuring the ability to connect to most databases, also allowed for connections to customized "Web Data Connectors." This is the term that Tableau uses for web based connectors that are built by the user. CKAN provides an API which allows access to the datas-

tore, so an html page was created. The webpage makes the necessary calls to CKAN to retrieve the data and then imports it into Tableau, establishing a connection that can be updated when the user desires. The code for the CKAN Data Connector can be found in Appendix F and also at the website <https://s3-us-west-2.amazonaws.com/sf-opendata/CKANDataConnector.html>. Some errors were received during the data import stage, though these did not prevent continuation and were short term. More information about the errors can be found in Appendix B

## Manipulating Data and Creating a Dashboard

Once data has been successfully imported to Tableau, it appears in a view in the bottom left labeled as the "Data Source":

The screenshot shows the Tableau interface with a data source view. The connection is set to 'Extract' and requires an extract. The table below represents the data source view.

Abc	#	#	#	#	Abc	#	#	#	#	#
webdata-direct.11fb2190whp... PARCC Assessment	webdata-direct.11fb219... District Code	webdata-direct.11fb219... Percent Level 4	webdata-direct.11fb219... School Code	webdata-direct.11fb2190wh... Count of Students	webdata-direct.11fb2190whp... District Name	webdata-direct.11fb219... Percent Level 5	webdata-direct.11fb219... Percent Level 2	webdata-direct.11fb219... Percent Level 3	webdata-direct.11fb219... Percent Level 1	webdat... _Id
Algebra 1	1	21.4000	0	6,732	Albuquerque Public Sc...	0.9000	33.2000	26.5000	18.0000	
Algebra 1	1	12.2000	6	41	Albuquerque Public Sc...	0.0000	39.0000	36.6000	12.2000	
Algebra 1	1	3.8000	7	26	Albuquerque Public Sc...	0.0000	65.4000	15.4000	15.4000	
Algebra 1	1	2.0000	16	50	Albuquerque Public Sc...	0.0000	52.0000	16.0000	30.0000	
Algebra 1	1	0.0000	17	20	Albuquerque Public Sc...	0.0000	45.0000	10.0000	45.0000	
Algebra 1	1	33.3000	28	21	Albuquerque Public Sc...	9.5000	19.0000	38.1000	0.0000	
Algebra 1	1	<i>null</i>	30	1	Albuquerque Public Sc...	0.0000	<i>null</i>	0.0000	0.0000	
Algebra 1	1	0.0000	39	28	Albuquerque Public Sc...	0.0000	50.0000	14.3000	35.7000	
Algebra 1	1	17.2000	47	64	Albuquerque Public Sc...	0.0000	39.1000	26.6000	17.2000	
Algebra 1	1	0.0000	51	23	Albuquerque Public Sc...	0.0000	39.1000	21.7000	39.1000	
Algebra 1	1	0.0000	61	23	Albuquerque Public Sc...	0.0000	39.1000	13.0000	47.8000	
Algebra 1	1	11.1000	63	63	Albuquerque Public Sc...	0.0000	41.3000	30.2000	17.5000	
Algebra 1	1	16.0000	69	25	Albuquerque Public Sc...	0.0000	36.0000	32.0000	16.0000	

Figure 4.8: Data Source View

Once connected, the data can then be manipulated in views called Worksheets. On a worksheet, there are a variety of ways to arrange and view data. Tableau has built in visualizations that can be created using "Dimensions" and "Measures", which are Tableau's words for variables depending on their context. The standard visualizations that can be made are:

1. Bar Chart
2. Line chart
3. Pie chart
4. Map
5. Scatter Plot
6. Gantt chart
7. Bubble chart
8. Histogram chart
9. Bullet chart
10. Heat map
11. Highlight table
12. Treemap
13. Box-and-whisker plot

Figure 4.9: Visualization Types

Variables can be placed in certain fields such as Columns, Rows, Pages, Filters, and Marks, which Tableau then associates with possible visualizations. From those different combinations of fields a worksheet can be filled with a view of the data, as seen here:

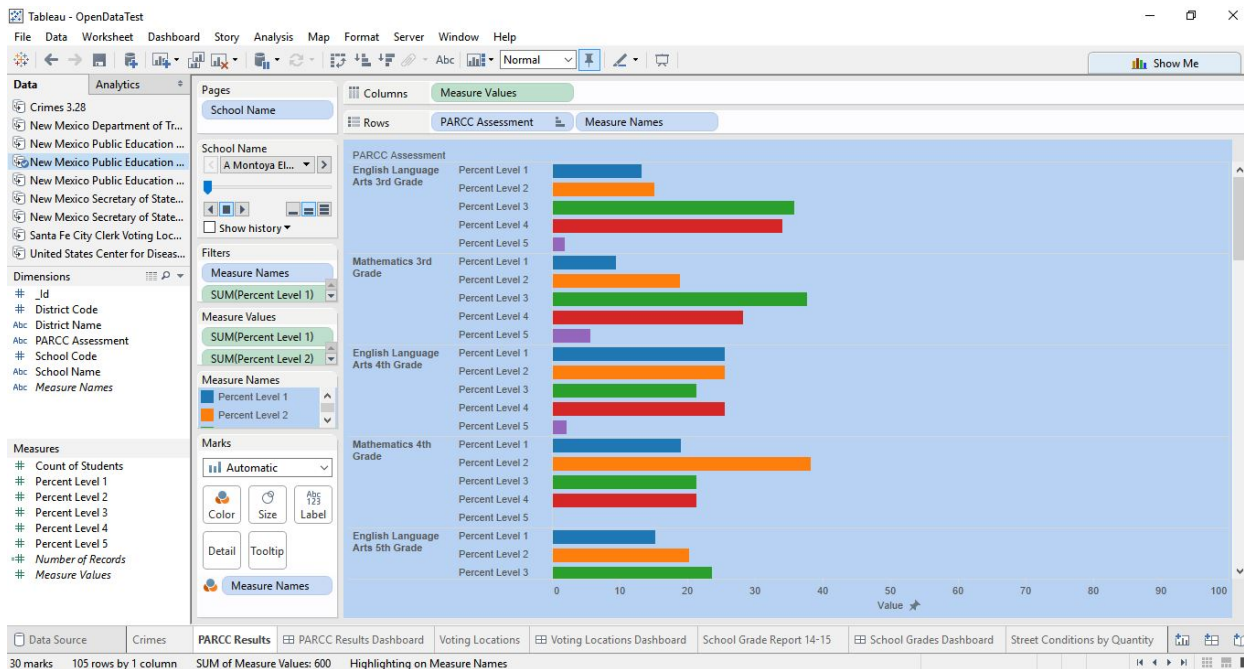


Figure 4.10: Worksheet Visualization



Once a worksheet has been completed, it can then be added to a Dashboard, either on its own or with multiple worksheets being displayed together. Here is an example of the above worksheet displayed in dashboard format:

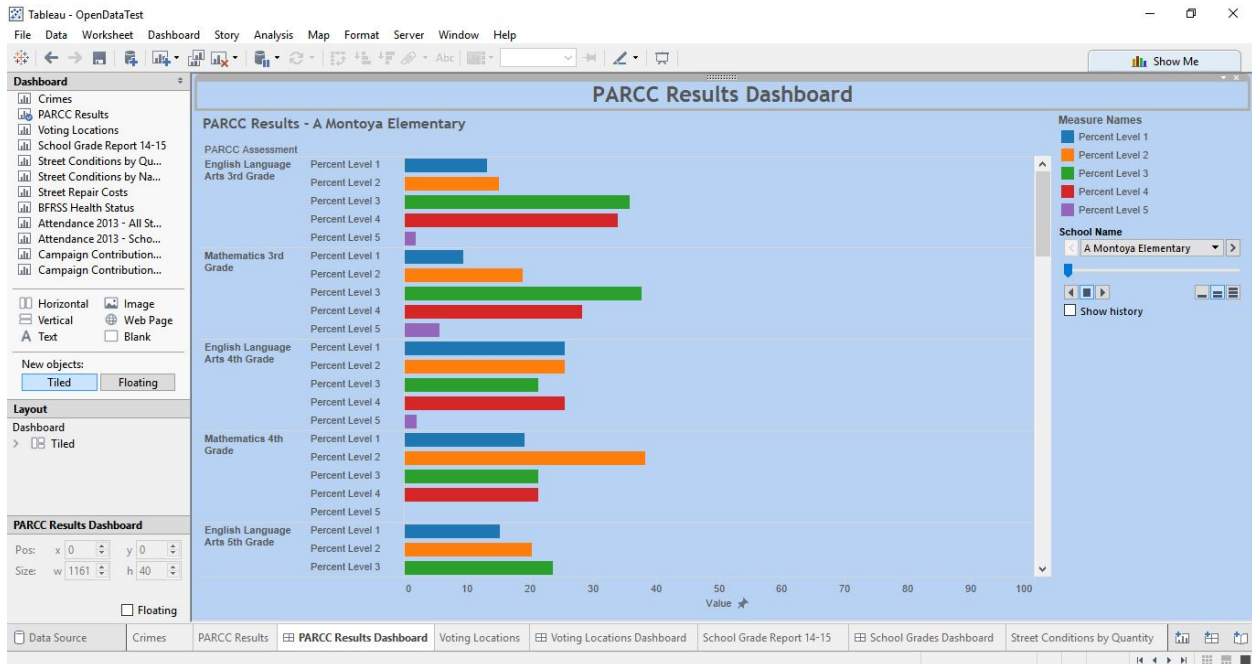


Figure 4.11: Dashboard View

All of the dashboards that were created can be viewed online at <https://public.tableau.com/profile/nicholas.shannon#!/> or in Appendix C.

### 4.3.2 The Dash

The Dash is an online dashboard service with both free and paid variations. While The Dash is very good at creating links to data sources, whether by embedding sites directly in the dashboard or by finding information stored in tables in the data source, it offers very little in the way of manipulating data beyond their preconfigured charts and tables. Unlike the other dashboards, The Dash doesn't separate out the data from the view. It performs manipulation of data within a single widget when the data is linked, so this restricts ease of creating views that will analyze data as desired. It also limits the filtering of data within a widget on the dashboard.

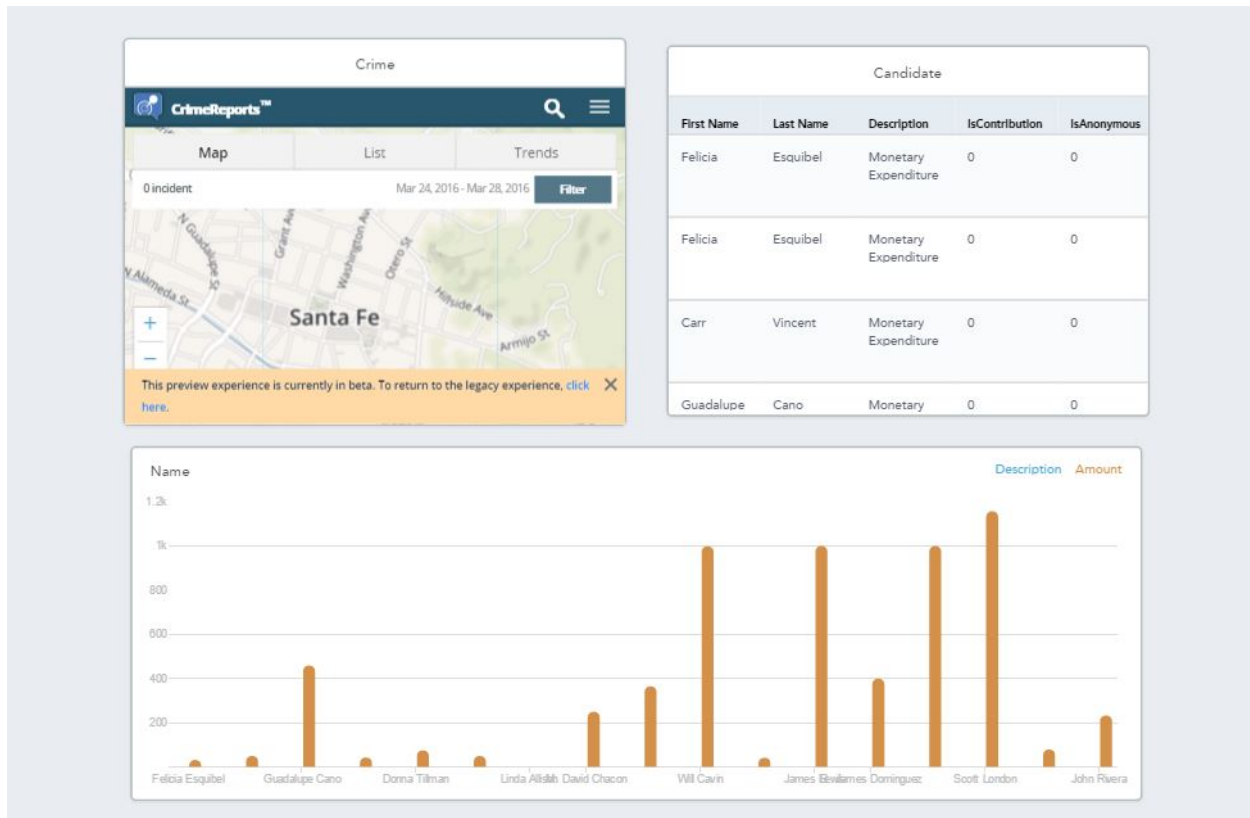


Figure 4.12: Example Dashboard Using The Dash

### 4.3.3 Power BI

Microsoft's Power BI is a business intelligence tool for connecting data sources and creating dashboards and other ways to analyze and view data. It features the ability to connect to a multitude of sources, including standard database types, local files, and web addresses, to name a few. A benefit of Power BI is that it first establishes a connection with the data source, then either creates a link or imports the data locally, depending on the type of source. These data can then be manipulated within Power BI and used to produce a multitude of charts, graphs, and tables for displaying the information that is desired. Another key feature that is supplied is for the views within a file to interact with each other through page and workbook level filters. With certain parameters set up, selecting one specific set of data on a certain view can filter another view to only show data related to the first selection. One drawback of Power BI is that its web integration, while able to get at web data sources, cannot actually embed a website within the dashboard. Since some sources of data are already established and only need to be embedded, this limits the ability to include all sources in Power BI.

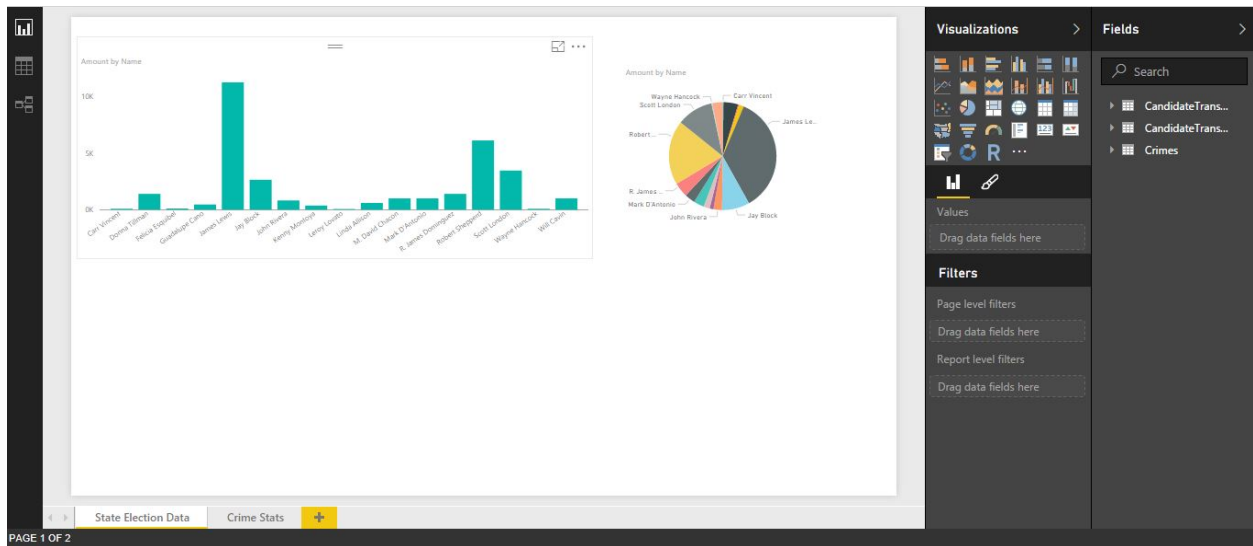


Figure 4.13: Example Dashboard Using Power BI (page 1)

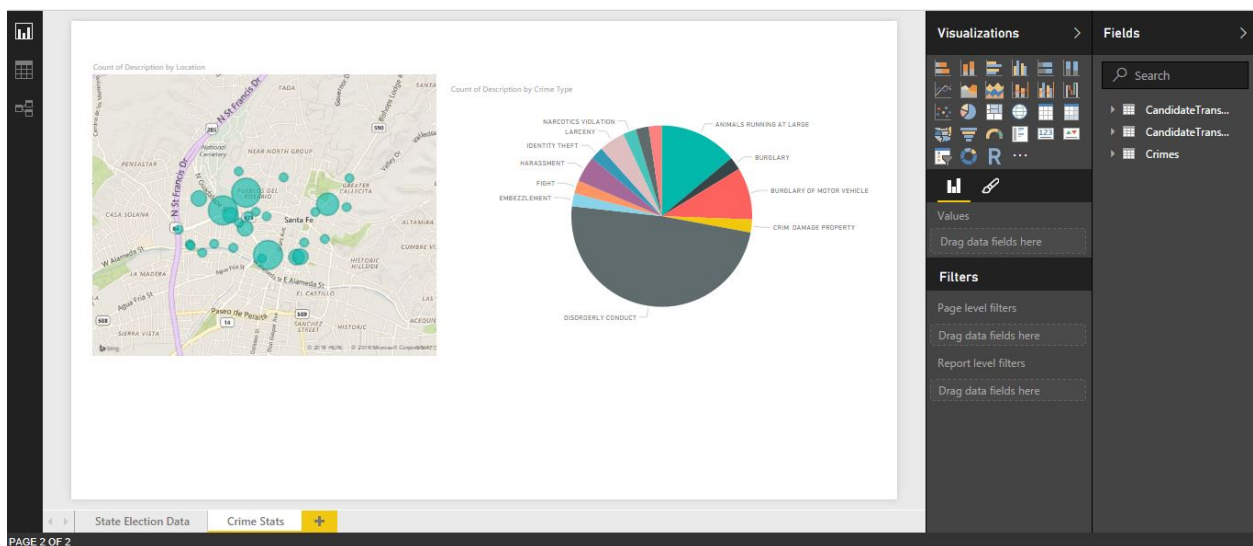


Figure 4.14: Example Dashboard Using Power BI (page 2)

### 4.3.4 Jaspersoft

Recently acquired by TIBCO, Jaspersoft is a reporting platform that specializes in extracting datasets and visualizing them in reports. Jaspersoft is best known for its advanced reporting and analytic capabilities, which are all housed within the JasperReports Server that Jaspersoft provides. As a paid option, Jaspersoft's products are among the more expensive. This is in part because of the amount of features that are included within the software suite. Extensive support is also available to those who purchase the product, while an active community also provides a free outlet for questions, although there is no guarantee one will receive a useful

answer on that route. From a user perspective, Jaspersoft is aimed towards advanced users, people who have programming and reporting experience. Users without that background will have trouble establishing the necessary connections, and may quickly get bogged down in the user interface, which presents a plethora of advanced options. In terms of data connections available, Jaspersoft focuses on server connections, featuring extensive options for MySQL and PostgreSQL specifically, among others. The JasperReports platform does offer Ad Hoc View and Dashboard creation, but both are again geared towards users with a substantial background in the subject area.



Figure 4.15: Example Dashboard Using Jaspersoft

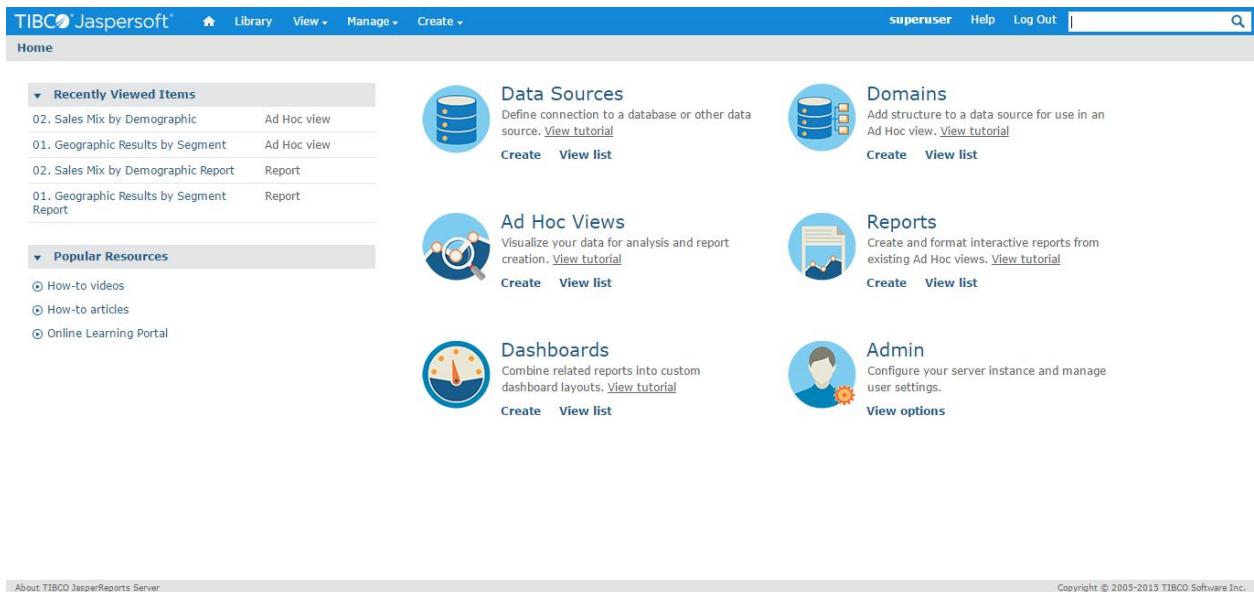


Figure 4.16: JasperReports Server Main Page

# Chapter 5

## Recommendations for Future Work

This chapter discusses the team's recommendations for future work that the city of Santa Fe can complete, especially focusing on how a citywide open data initiative can be developed. The team's recommendations are:

1. If the city wants to continue their movement towards a more transparent government, they would benefit from creating an open data initiative.
2. A legislative change would be most effective in driving substantial change from the existing data management and publication methods practiced by city employees.
3. The city would benefit from investing in a professional supported platform such as Socrata for publishing data, and using Esri ArcGIS and Tableau to create public views of city information.
4. If the city is to create a city-wide open data policy, organizations and individuals both inside and outside the city government should be interviewed to determine which data would be best for initial inclusion in the portal and dashboards. This would aid in directing the initiative's starting point.
5. The city should utilize the data that the team collected in the repository in the upcoming Santa Fe Civic APPaTHON that they are holding in summer 2016.

### 5.1 Santa Fe Open Data Initiative

**Recommendation:** If the city wishes to continue their movement towards a more transparent government, they would benefit from starting an open data initiative.

As mentioned in Section 3.1.1 of the Methodology Chapter, the team filled out a General Operating Support Grant from the Thornburg Foundation. The Thornburg Founda-

tion would provide a grant for the Santa Fe Project Site if certain election data was compiled, specifically campaign finance reports and code of ethics forms. In searching for this information, it was determined that the finance forms were only available as PDFs and the code of ethics forms were not digitized at all (Vigil, 2016). This problem of certain data either not being stored in useful formats or not being stored digitally at all became more apparent as further datasets were investigated.

An open data initiative for the city as a whole would help to update the city in how it stores, publishes, and maintains documents. Open data practices would assist employees in becoming more familiar with data formats that are useful to both government employees and citizens. The initiative would also help the city to become more modern in the manner that they treat data and its publication. The team recommends a central data repository as a means of beginning to organize data within the government, which aids both employees and citizens that are seeking the data.

## **5.2 Legislative Change to Data Management**

**Recommendation:** A legislated change would be most effective in driving substantial change from the existing data management and publication methods practiced by city employees.

In the City of Santa Fe, not every department releases their data. When comparing the information released by Santa Fe to information released by Albuquerque, there is much wider data diversity present in the latter. With the help of all departments within a city, data can be released in a timely manner. Albuquerque can be a valuable resource to Santa Fe due to its proximity and can provide departments information on how to enhance transparency in government data sharing.

Other resources include the open data implementations of other cities across the US. One key city is Philadelphia, which has been advancing their open data movement since 2012. In 2012, the City of Philadelphia created an open data policy that was signed as an executive order. This immediately started the open data movement within the city government with specific guidelines on how to get public participation as well as government participation. This executive order is a model that the City of Santa Fe can use for seeing how to go about creating and implementing their own plan. In order to bring about meaningful change in the city's open data practices, the city government would be best served creating a city-wide mandate which describes the methods that the city departments will use to publish their data.

## **5.3 Tools and Technologies**

**Recommendation:** The team recommends an investment in a professional and supported hosting platform such as Socrata for publishing data to. For the purpose of visualizing data, the

team recommends using the city's existing Esri ArcGIS platform for geographic data and investing in Tableau Desktop for creating public dashboards.

Even though CKAN was the platform used to show the city government open data usefulness, it is not the best platform out on the market for long-term use. For CKAN, hosting and support payments would be necessary for long-term usage. Even though these additional resources can be used, CKAN is not the most user-friendly platform available for both the public and for government workers. A better long-term solution for the city would be to implement Socrata, which is more advanced and offers a more user friendly environment. Pricing is not shown on their website and the city would need to set up a meeting with a representative to talk about pricing and training. Socrata might be substantially expensive, but the meeting would provide further information about the feasibility of it from both a fiscal and a usability standpoint so that an educated decision can be made. For mapping data, the city's Esri ArcGIS system can be used. ArcGIS is ideal for creating geographic visualizations from the information provided. The city is already in the process of implementing an ArcGIS system which means that they can incorporate some of the data once the city finishes their upgrade (Martinez & Barsanti, 2016). As a result of the current upgrade to Esri ArcGIS, the team would encourage the city to generate more location-based data to create specialized applications for the city. Finally, Tableau Desktop is an ideal visualization platform for creating different ways to view data and then publish a dashboard for the public to view. While there is a moderate cost associated with it of \$2000/user for the initial software, with a \$500 cost/user/year for getting the latest updates and support, the price has been deemed worthwhile by the team in exchange for the vast possibilities that Tableau provides. Most of the datasets that the team worked with did not contain geographic variables, which ruled out ArcGIS as a visualization method. Tableau fulfilled that need, as well as providing a way to create and publish dashboards for public viewing. Using free Academic licenses for the software, the team created the prototype dashboards found in Appendix C, which are examples that the team recommends the city follow in publishing their data in easy to view formats.

By moving over to a digital repository, it would help the city to move their entire document storage system to digital platform. This would make uploading and accessing data easier for the city and the public. This includes having electronic signatures being a proper form of notarization. Electronic signatures would allow for easy upload by the city departments without needing to print out the form, have the person sign the form, scan it, and then upload it as a PDF without any way of easily taking the data from the form. Having online documents also does not hinder any previous practices government employees perform. The city currently has a storage facility where they keep all the printed copies of documents and reports. This can still be used as a backup if desired, but the primary storage would need to be digital storage. Additionally, most online storage platforms provide methods for easily backing up the system at regular intervals, which prevents loss of information. Online documents can easily be reviewed and changed if needed. This will greatly help the city once they commit to having government transparency.

## 5.4 Interviewing to advance the Santa Fe Open Data Initiative

Recommendation: The team recommends that if a policy driven change is to be made, the general public and local organizations should be interviewed to garner support and ascertain which information the public would make greatest use of to begin with, which would help direct the initiative's starting point.

Further interviews of government employees could be conducted by the city with support from ITPD if desired to gather more information about how specific resources can be released to the public. When the city begins their open data movement, with the help of ITPD, they can start surveying the public or start city council discussions about the topic of open data to get the public's perspective on the topic. Local organizations can be valuable as well due to the fact that they represent the public. Example questions for these groups can be found in Appendix A. These questions cover a broad range of topics related to the project: the questions seek insight into the participant's feelings on open data, what kinds of data they would like to see, what kinds of data would be useful to them, where they would go to get data, and what they feel the most convenient way of accessing this data would be.

## 5.5 Santa Fe Civic APPaTHON

Recommendation: The team recommends that the City of Santa Fe utilize the team's data repository in the hackathon that is planned for this summer.

One component of the project involved preparing for the Santa Fe Civic APPaTHON which was to be held near the end of the project time period. Members of the team were going to participate as instructors and helpers throughout the hackathon (Johnson et al., 2016). Additionally, the team was going to be kicking off the hackathon with a presentation of the work completed over the course of the project. On the second week of work in Santa Fe, the hackathon was moved to the end of June. Therefore, the team's involvement had to be cut back. Despite this, the data collected are still able to be used in the hackathon if so desired and the dashboards created can also be used as examples. The work that the team completed should be utilized in the hackathon to begin involving the public in designing and using Santa Fe's open data system.



# References

- Alexopoulos, C., Spiliotopoulou, L., & Charalabidis, Y. (2013). Open data movement in greece: A case study on open government data sources. *Proceedings of the 17th Panhellenic Conference on Informatics - PCI '13*, 279. Retrieved from <http://dl.acm.org/citation.cfm?doid=2491845.2491876> doi: 10.1145/2491845.2491876
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 1–11. Retrieved from <http://bds.sagepub.com/lookup/doi/10.1177/2053951715594634> doi: 10.1177/2053951715594634
- Beal, V. (n.d.). *What is application programming interface (api)?* Retrieved from <http://www.webopedia.com/TERM/A/API.html>
- Bedini, I., Farazi, F., Leoni, D., Pane, J., Tankoyeu, I., & Leucci, S. (2014). Open government data: Fostering innovation. *JeDEM - eJournal of eDemocracy and Open Government*, 6(1), 69–79. Retrieved from <http://www.jedem.org/article/view/329>
- Berners-Lee, T. (2009). *Linked data - design issues*. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>
- Birkinshaw, P. (2010). *Freedom of information: The law, the practice and the ideal*. Cambridge University Press.
- Bradbury, N., Fox, T., & Frank, S. (2002). *Merton open space study : Implementing a gis* (Tech. Rep.).
- Caballero, A., Shvartz, A., Li, B., Gevaudan, C., Lamont, I., Souza, J., ... Hirschmann, T. (n.d.). Open data center alliance usage : Data security framework rev 1 . 0 . , 1–21.
- Carrasco, C., & Sobrepere, X. (2015). Open government data: An assessment of the spanish municipal situation. *Social Science Computer Review*, 33(5), 631–644. Retrieved from <http://ssc.sagepub.com/cgi/doi/10.1177/0894439314560678> doi: 10.1177/0894439314560678
- Chignard, S. (2013). A brief history of Open Data. , 1–5. Retrieved from <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>
- City of Los Angeles. (n.d.). *Home | los angeles geohub*. Retrieved from <http://geohub.lacity.org>
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31(SUPPL.1), S10-S17. Retrieved from <http://dx.doi.org/10.1016/j.giq.2014.01.003> doi: 10.1016/j.giq.2014.01.003
- Delgado, C., & Randall, R. (2016, March). Personal communication.
- De Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., ... Whitehouse,

- D. (2011). Popsis-pricing of public sector information study. *European Commission Information Society and Media Directorate-General*.
- Edoceo, I. (2014). *Comma separated values (csv) standard file format*. Retrieved from <http://edoceo.com/utilitas/csv-file-format>
- Esri. (n.d.). *Arcgis for server*. Retrieved from <http://www.esri.com/software/arcgis/arcgisserver>
- Esri. (2012). *The 50th anniversary of gis*. Retrieved from <http://www.esri.com/news/arcnews/fall12articles/the-fiftieth-anniversary-of-gis.html>
- G8. (2013). G8 open data charter. *G8 Lough Erne 2013*(June), 1–10. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/207772/Open\\_Data\\_Charter.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf)
- Galiotou, E., & Fragkou, P. (2013). Applying linked data technologies to greek open government data: A case study. *Procedia - Social and Behavioral Sciences*, 73(December 2015), 479–486. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S187704281300373X> doi: 10.1016/j.sbspro.2013.02.080
- Gis (geographic information system). (2011). Retrieved from <http://education.nationalgeographic.org/encyclopedia/geographic-information-system-gis/>
- Greenido. (n.d.). *Yahoo finance*. Retrieved from <https://au.finance.yahoo.com/>
- It's The People's Data. (n.d.). *It's the people's data*. Retrieved from [https://www.facebook.com/Its-The-Peoples-Data-1599854626919671/info/?tab=page\\_info](https://www.facebook.com/Its-The-Peoples-Data-1599854626919671/info/?tab=page_info)
- Johnson, T., Ward, M., Lockman, A., Hayes, G., McNary, N., & Shannon, N. (2016, Feb). *It's the people's data sponsor interview*.
- Kucera, J., & Chlapek, D. (2014). Benefits and risks of open government data. *Journal of Systems Integration*, 5(1), 30–41.
- Lassinantti, J., Bergvall-Kåreborn, B., & Ståhlbröst, A. (2014). Shaping local open data initiatives: Politics and implications. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 17–33. doi: 10.4067/S0718-18762014000200003
- Location intelligence. (n.d.). Retrieved from <http://www.pitneybowes.com/us/location-intelligence.html>
- Martinez, R., & Barsanti, D. (2016, March). Personal communication.
- Meriam-Webster. (2016). *Metadata*.
- Mustonen, J. (2006). *The World's First Freedom of Information Act*. Anders Chydenius Foundation. Retrieved from [http://www.chydenius.net/pdf/worlds\\_first\\_foia.pdf](http://www.chydenius.net/pdf/worlds_first_foia.pdf)
- National Audit Office. (2012, Apr). *Implementing transparency - national audit office (nao)*. Retrieved from <https://www.nao.org.uk/report/implementing-transparency/>
- Nations, D. (2015). *What is a linkback?* Retrieved from <http://webtrends.about.com/od/glossary/g/linkback.htm>
- New Mexico Public Education Department. (2015a). *Proficiencies state, district and school, by grade 2015*. Retrieved from <http://ped.state.nm.us/assessmentaccountability/AcademicGrowth/NMSBA.html>
- New Mexico Public Education Department. (2015b). *Spring 2015 parcc results*. Retrieved from <https://public.tableau.com/views/NewMexicoPARCCResultsDashboard/>

- Dashboard1?:embed=y&:display\_count=yes&:showTabs=y&:toolbar=no&:showVizHome=no
- Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57(6), 1701.
- Open Knowledge. (2016). *The open data handbook*. Retrieved from <http://opendatahandbook.org/guide/en/>
- Oracle. (2015). *What is a url?* Retrieved from <https://docs.oracle.com/javase/tutorial/networking/urls/definition.html>
- Österreich, C. O. (n.d.). *Katalog | data.gv.at*. Retrieved from <https://www.data.gv.at>
- Parycek, P, Höchtl, J., & Ginner, M. (2014). Open government data implementation evaluation. *Journal of theoretical and applied electronic commerce research*, 9(2), 13–14. Retrieved from [http://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-18762014000200007&lng=en&nrm=iso&tlng=en](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000200007&lng=en&nrm=iso&tlng=en) doi: 10.4067/S0718-18762014000200007
- Quantum gis. (n.d.). Retrieved from <http://www.qgis.org/en/site/index.html>
- Rothenberg, J. (2012). Case study international benchmark. (April), 1–48.
- Santa Fe, NM, Johnson, T., & It's The People's Data. (2015). *It's The People's Data Initiative Position Paper Draft* (Tech. Rep.).
- State of Indiana. (n.d.). *What is gis?* Retrieved from <http://www.in.gov/gis/gis101.htm>
- State of New York. (2015, Feb). *Open new york*. Retrieved from <http://ny.github.io/open-data-handbook/index.html?1455514232222>
- Sunlight Foundation. (2010). Ten Principles for Opening Up Government Information. *Sunlight Foundation*(October 2007), 3. Retrieved from <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>
- Technopedia. (n.d.). *What is web 1.0? definition from technopedia*. Retrieved from <https://www.techopedia.com/definition/27960/web-10>
- The Economist. (2015, Nov). *Out of the box: The open-data revolution has not lived up to expectations. but it is only getting started*. Retrieved from <http://www.economist.com/news/international/21678833-open-data-revolution-has-not-lived-up-expectations-it-only-getting>
- United States Department of Justice. (1996). *Foia update: The freedom of information act, 5 u.s.c. sect. 552, as amended by public law no. 104-231, 110 stat. 3048*. Retrieved from <http://www.justice.gov/oip/blog/foia-update-freedom-information-act-5-usc-sect-552-amended-public-law-no-104-231-110-stat>
- United States Government. (2014). U.s. open data action plan. , 1–21.
- Vigil, Y. (2016, March). Personal communication.

# Glossary

## **API**

an abbreviation of Application Programming Interface; a set of routines, protocols, and tools for building software applications (Beal, n.d.). 8, 18

## **CSV**

an abbreviation of Comma Spliced Values; a standard file format used for data exchange from a variety of applications. 8

## **linkback**

a term that's commonly used to refer to the mention of a website or blog on another website or blog, along with adding a hyperlink to its home page or a specific page so that users can click on it to visit it directly (Nations, 2015). 8

## **metadata**

data that provides information about other data (Meriam-Webster, 2016). 16, 18

## **proof of concept**

A Proof of concept is a realization of a certain method or idea to demonstrate its feasibility, or a demonstration in principle, whose purpose is to verify that some concept or theory has the potential of being used. A proof of concept is usually small and may or may not be complete. ii–iv, 3, 12, 13, 15, 22, 24, 29, 38, 39

## **SQL**

an abbreviation of Structured Query Language; a programming language used for managing data in a relational database management system. 8

## **URL**

an acronym for Uniform Resource Locator and is a reference (an address) to a resource on the internet (Oracle, 2015). 8

## **Web 1.0**

the first stage in the World Wide Web, which was entirely made up of Web pages connected by hyperlinks. Refers to the Web when it was a set of static websites that were not yet providing interactive content (Technopedia, n.d.). 8

# Acronyms

**FOIA**

Freedom of Information Act. 5, 9, 10, 17, 18

**GIS**

Geographic Information Systems. 19, 20

**OGD**

Open Government Data. ii, 4–7, 9, 10, 13, 14, 16–18

# Appendix A

## Open Data Stakeholder Discussion Topics

A key component in developing a product or service for any group of people is gauging the interest and need of that group to ensure that the product fulfills the need. It follows then that in order to develop an open data repository and prototype dashboard the public desires need to be addressed. As part of the project, a field survey must be conducted in order to determine the public interest, and to determine which data and features should be included so as to be most useful for the target audience.

As part of the field survey, three key groups of people would be surveyed: Santa Fe government employees, and the citizens of Santa Fe. Government employees have been identified as a key beneficiary of an open data portal, since having ready access to data from other departments can often make their jobs easier. The citizens are the second key group, as they provide opinions on what would be useful to people outside of the government. The third key group are local organizations that represent the citizens. The following sections provide a sample schedule of questions that would be asked of field survey candidates.

### A.1 Government Employees

The government employees serve to provide data on how each department views open data. This will provide the team with necessary information on whether or not each department is actively updating their information. With any survey, and especially government, there needs to be strict anonymity. Even with the anonymous questions, we still need to know which departments the results come back from to get an accurate representation of the governments viewpoint. One way of getting these results is by using Google Forums and sending it out to the employees. Google forums does not collect personal information unless asked. The only information we would need is what department the person that was surveyed is from.

1. What department do you work for in Santa Fe?

2. Do you know about Open Government Data?
3. What are your feelings on Open Data?
4. What would you use Open Data for?
5. Has your department posted any data online for public use?
6. Do you currently know if your department is actively releasing information to the public?
7. Has your department used any data published by other departments in your work?
8. Would you or your department be interested in collaborating on feedback or development of the platform we are creating as part of our project?

## **A.2 Citizens of Santa Fe**

The citizens serve to provide useful data on how to build a useful platform for publishing open data. This will provide the team with necessary information on if the citizens understand the importance of open data and if they are using the resources currently available. One way of getting these results is to perform a field survey. The field survey will have strict anonymity among the citizens and will only look for whether or not the citizens know about open government data.

1. Do you know about Open Government Data?
2. What kinds of data would you be interested in seeing?
3. What would you use Open Data for?
4. Where would you go to look for government data?
5. What would be the most convenient way for you to access data?

## **A.3 Local Organizations**

Local organizations represent the opportunity for the community to get involved in both using and contributing to open data. Similar to the citizens, organizations will be able to provide the team with feedback on which methods of presenting the data would be ideal. Some other information that can be ascertained would be with regards to what data would be most useful, where they would look for it, and if there is any data that can be sourced from those local organizations and be added to the repositories. With organizational consent, certain quotes may be useful in the final project report. However, to maintain anonymity if organizations so

desire, Google Forms may be utilized in this situation as well. A possible way to categorize responses would be to use market vertical to assign classifications.

1. Do you know about Open Government Data?
2. What are your feelings on Open Data?
3. What kinds of data would you be interested in seeing?
4. What would you use Open Data for?
5. Where would you go to look for government data?
6. What would be the most convenient way for you to access data?
7. Would you or your organization be interested in collaborating on feedback of the platform we are creating as part of our project?



# Appendix B

## Tableau Internal Errors

One problem encountered with Tableau was an unusual tendency for it to experience internal errors for no apparent reason. The errors would occur when the Web Data Connector was run and a selection was made from the list of importable datasets. Tableau would then begin attempting to establish a connection, as pictured here:

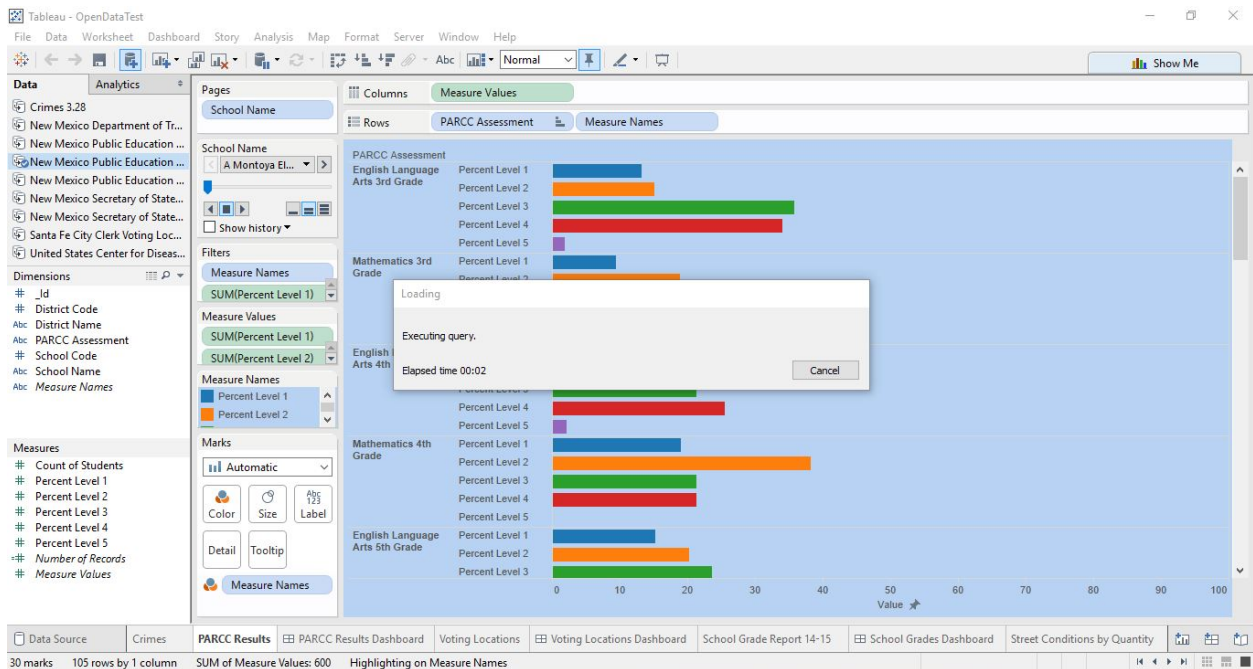


Figure B.1: Web Data Connector Beginning Execution

Then, it would randomly throw an internal error, as pictured here:

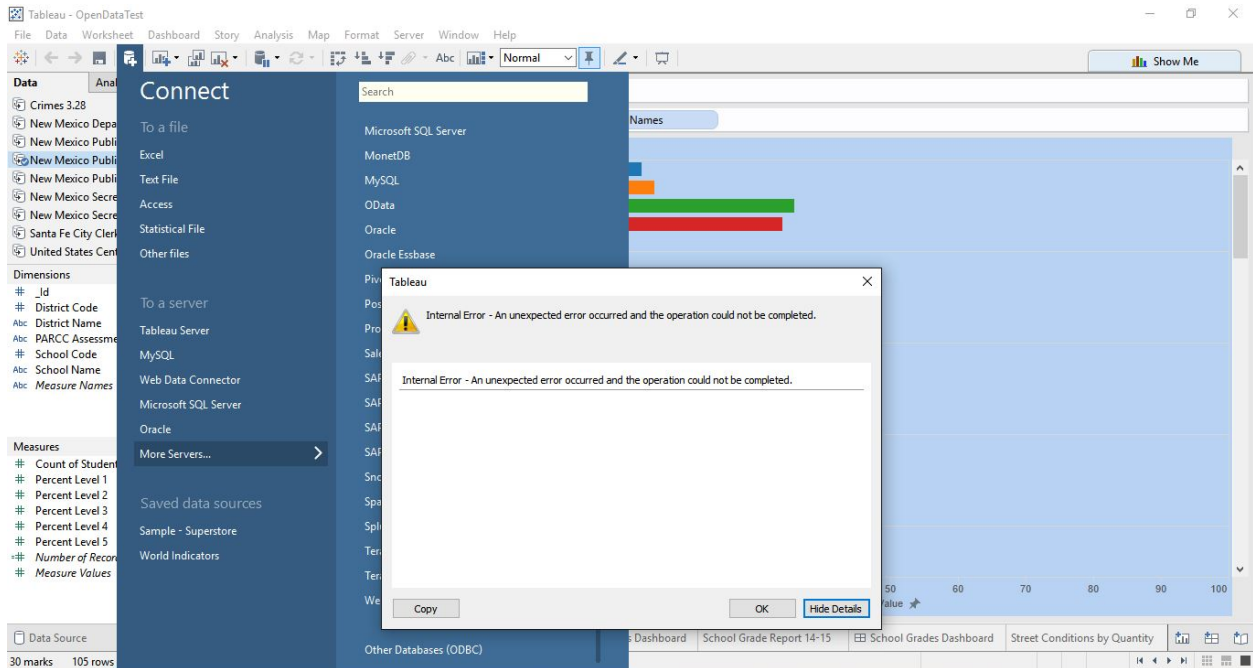


Figure B.2: Web Data Connector Internal Error

This error was not readily explicable, and only occurred on one team member's computer. As a result, the team could not confirm whether the error was an anomaly or whether this was a common Tableau issue. The Tableau support forums did not yield results about similar problems occurring for other users either, which indicates that an assumption that the error was an outlier would not be wholly wrong. Either way, the error would not always occur and therefore did not severely interfere with the ability to use Tableau. Once the datasets were imported successfully once, the data was able to be manipulated error free.

# Appendix C

## Tableau Dashboards

The following pages contain images taken of the tableau dashboards that were created as part of this project.



Figure C.1: Santa Fe Streets Dashboard

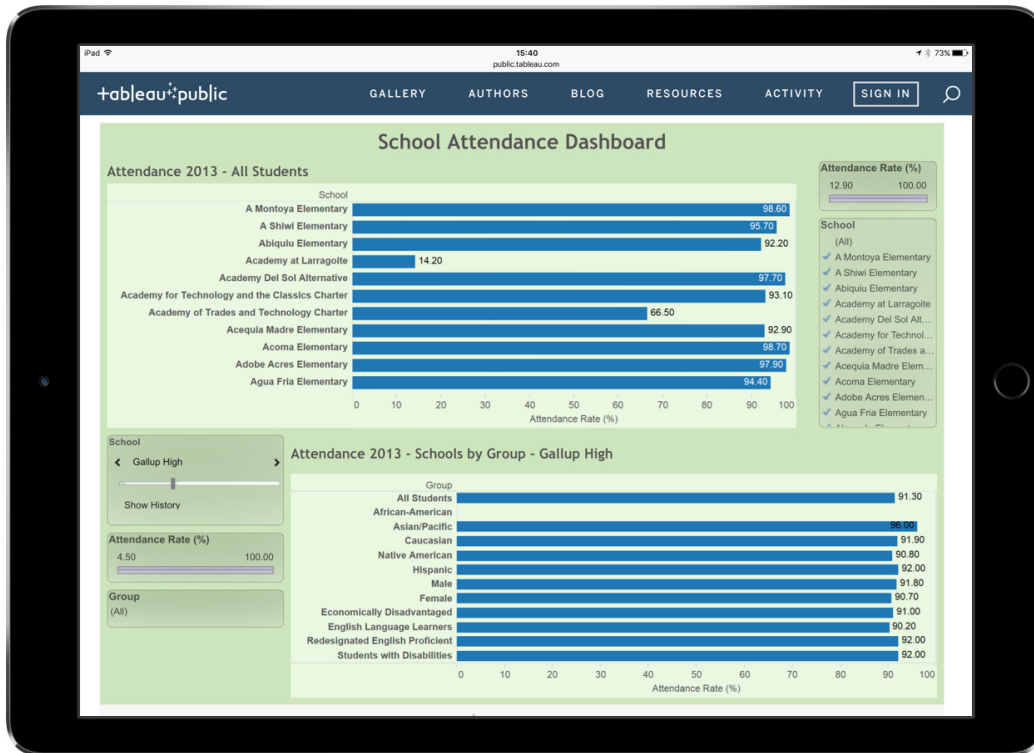


Figure C.2: School Attendance Dashboard

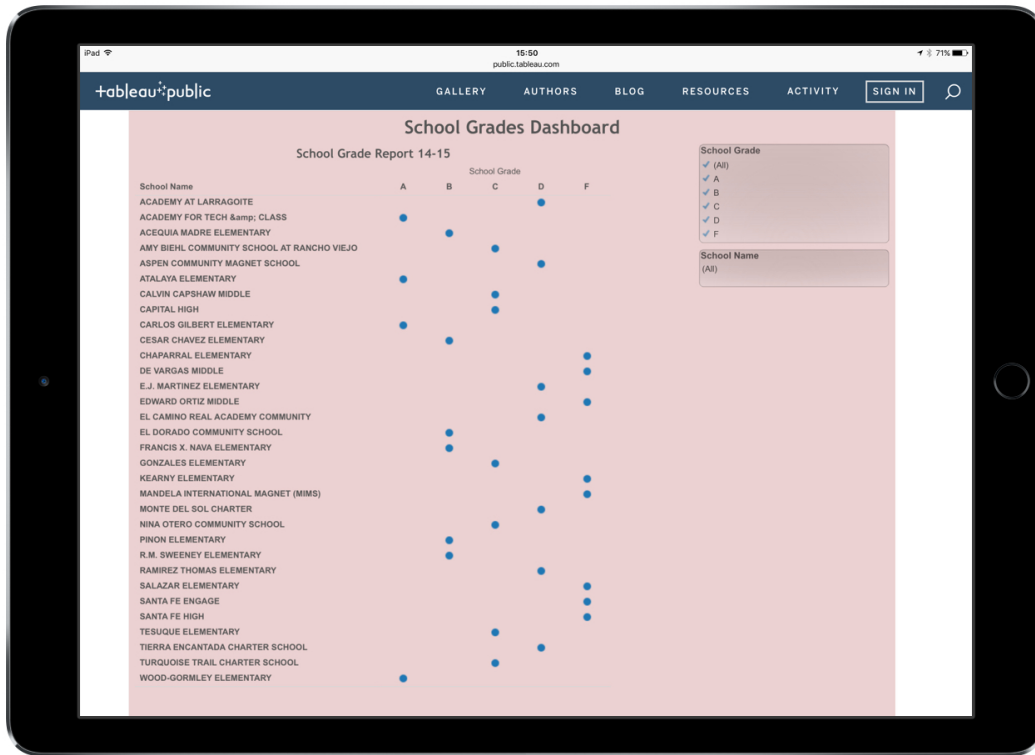


Figure C.3: School Grades Dashboard

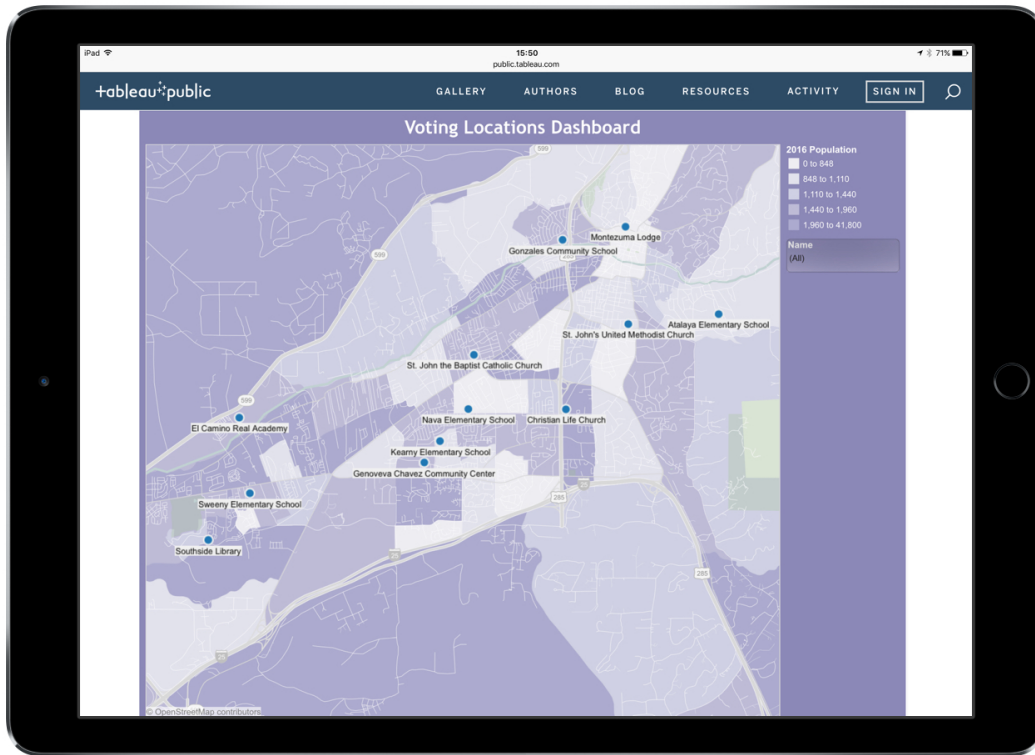


Figure C.4: Santa Fe Voting Locations Dashboard

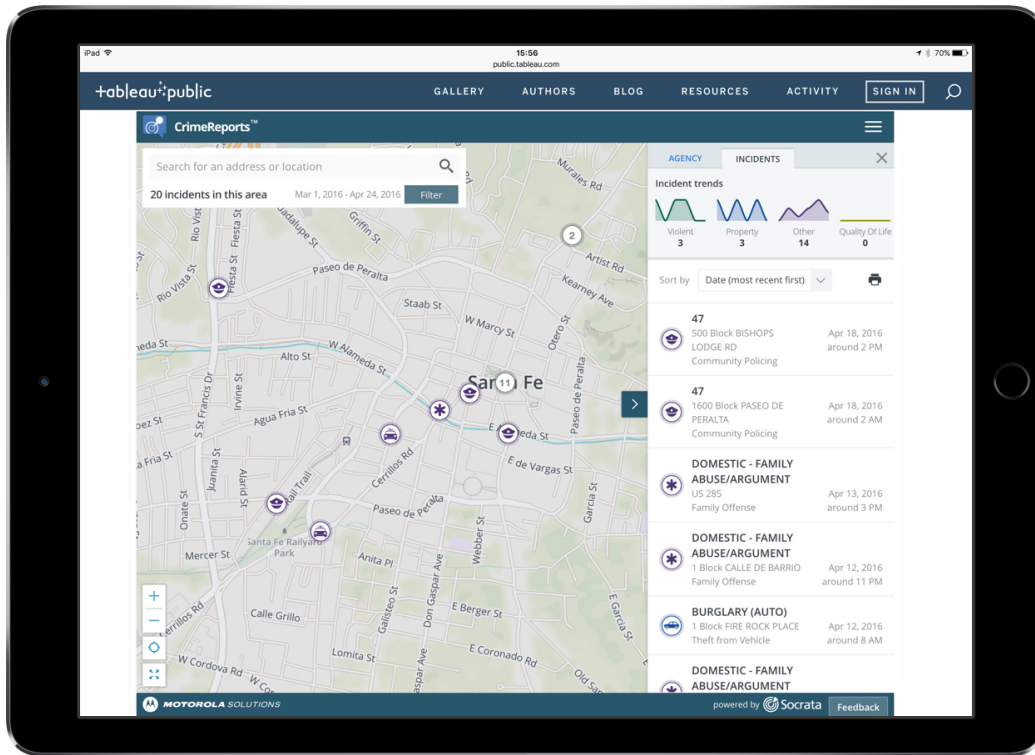


Figure C.5: Santa Fe County Crime Reports Dashboard



Figure C.6: Santa Fe PARCC Results Dashboard



# Appendix D

## Dataset Information

89

The table on the next few pages describes the details of each of the datasets that were included in the open data portal.

Category	Data Type	Source	Contact Person	URL	Format	Update Frequency
Election	Campaign finance reports	Followthemoney		<a href="http://www.followthemoney.org/election-overview?s=NM&amp;y=2014">http://www.followthemoney.org/election-overview?s=NM&amp;y=2014</a>	Spreadsheet	
Election	Campaign contributions and expenditures	New Mexico Campaign Finance System	Brad Winter (Secretary of State)	<a href="https://www.cfis.state.nm.us/media/">https://www.cfis.state.nm.us/media/</a>	CSV	After Major Elections

Election	Campaign finance reports	City of Santa Fe, NM	Yolanda Vigil (City Clerk)	<a href="http://www.santafenm.gov/campaign_finance_statements_and_participating_candidate_expendit_1">http://www.santafenm.gov/campaign_finance_statements_and_participating_candidate_expendit_1</a>	PDF	Seems to be Monthly
Election	Campaign expenditures and finance reports	Federal Election Commission		<a href="http://www.fec.gov/data/DataCatalog.do?cf=downloadable">http://www.fec.gov/data/DataCatalog.do?cf=downloadable</a>	CSV, XML, RSS	
Education	PARCC Standardized test results	New Mexico Public Education Department	Beverly Friedman (NMPED Custodian of Record)	<a href="https://public.tableau.com/views/NewMexicoPARCCResultsDashboard/Dashboard1?:embed=y&amp;:display_count=yes&amp;:showTabs=y&amp;:toolbar=no&amp;:showVizHome=no">https://public.tableau.com/views/NewMexicoPARCCResultsDashboard/Dashboard1?:embed=y&amp;:display_count=yes&amp;:showTabs=y&amp;:toolbar=no&amp;:showVizHome=no</a>	Dashboard, Graphics	Yearly
Education	School grades	New Mexico Public Education Department	Beverly Friedman (NMPED Custodian of Record)	<a href="http://webapp2.ped.state.nm.us/SchoolData/SchoolGrading.aspx">http://webapp2.ped.state.nm.us/SchoolData/SchoolGrading.aspx</a>	Map, Excel, CSV	Yearly

Education	School proficiencies, attendance reports	New Mexico Public Education Department	Beverly Friedman (NMPED Custodian of Record)	<a href="http://ped.state.nm.us/AssessmentAccountability/AcademicGrowth/NMSBA.html">http://ped.state.nm.us/AssessmentAccountability/AcademicGrowth/NMSBA.html</a>	PDF	Yearly
Education	Head start preschool reports	US Department of Health and Human Services		<a href="http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2014-hs-program-factsheet.html">http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2014-hs-program-factsheet.html</a>	PDF Reports	
Crime	Crime report data	Santa Fe County Police Department	Jessica Rodarte (Custodian of Records)	<a href="https://preview.crimereports.com">https://preview.crimereports.com</a>	Map, Table	Crimes go up after 72 hours.
Health	Healthcare coverage	US Census Bureau		<a href="http://factfinder.census.gov/bkmk/table/1.0/en/ACS/14_5YR/S2701/0500000US35049">http://factfinder.census.gov/bkmk/table/1.0/en/ACS/14_5YR/S2701/0500000US35049</a>	CSV, PDF, RTF, XLS	
Health	Public health indicators	New Mexico Department of Health	Daniel Jacobs (Records Custodian)	<a href="https://ibis.health.state.nm.us/">https://ibis.health.state.nm.us/</a>	Maps, Data downloads	Depends on what info
Health	All health information available in the state	New Mexico Department of Health	Daniel Jacobs (Records Custodian)	<a href="http://nmhealth.org/data/all/">http://nmhealth.org/data/all/</a>	Maps, PDF Reports	Depends on what info

Health	US Health Status	Center for Disease Control		<a href="https://chronicdata.cdc.gov/Behavioral-Risk-Factors/BRFSS-Table-of-Health-Status/y94n-rgzt">https://chronicdata.cdc.gov/Behavioral-Risk-Factors/BRFSS-Table-of-Health-Status/y94n-rgzt</a>	CSV, JSON, PDF, RSS, XLSX, XML, XLS	
Education	PARCC Standardized Test Proficiencies	New Mexico Public Education Department		<a href="http://ped.state.nm.us/ped/NMPARCCindex.html">http://ped.state.nm.us/ped/NMPARCCindex.html</a>	XLS	Yearly
Public Works	Santa Fe Street Conditions	New Mexico Department of Public Works		<a href="http://www.santafenm.gov/public_works">http://www.santafenm.gov/public_works</a>	XLSX, CSV	

Table D.1: Table of datasets.

# Appendix E

## Dashboard Software Comparison

The table on the next few pages compares the different dashboard software solutions considered for use as part of the project.

72

Software Name	Website	Price	Data Sources Supported	Data Capacity Limit	Data Refresh Rate
Microsoft Power BI	<a href="https://powerbi.microsoft.com/en-us/">https://powerbi.microsoft.com/en-us/</a>	Free, or \$120 per year per user	Excel, CSV, XML, Text, JSON, Folder, SQL Server DB, Access DB, SQL Server Analysis Services DB, Oracle DB, IBM DB2 DB, MySQL DB, PostgreSQL DB, Sybase DB, Teradata DB, Microsoft Azure, Web, SharePoint List, OData Feed, Hadoop File, Active Directory, Microsoft Exchange, Dynamics CRM Online, Facebook, Google Analytics, SAP Hana DB, Salesforce, ODBC, R Script, Smartsheet	1GB (free), 10GB (paid)	Daily (free), 8x per day (paid)

The Dash	<a href="http://thedash.com">http:// thedash.com</a>	Free, or \$795 per year for 5 users + \$159 per year for each additional user	Text & Images, Clock, Weather, Nest, App Store, appFigures, Github, Google Analytics, Chartbeat, Pingdom, Twitter, Instragram, Fitbit, Withings, Strava, Gmail, Google News, BBC, CNN, ESPN, Bleacher Report, Hacker News, Reddit, RSS Feeds, Google Sheets, URL	None	Upon Page Refresh
Tableau	<a href="http://www.tableau.com/">http://www .tableau.com/</a>	Free for students and educators, \$1999 per user one time, plus \$500 per year for maintance (support, product updates)	Action Vectorwise, Amazon Redshift, Amazon Elastic MapReduce, Cloudera Hadoop Hive and Impala, DataStax, EXASOL, Firebird, Google Analytics, Google BigQuery, Hortonworks Hadoop Hive, HP Vertica, IBM DB2, IBM Netezza, MapR, Microsoft Access, Microsoft Excel, Microsoft PowerPivot, Microsoft SQL Server, Microsoft SQL Server Analysis Services, Microsoft SQL Server PDW, Microsoft Windows Azure Marketplace DataMarket, MySQL, OData, Oracle Databases, Oracle Hyperion Essbase, ParAccel Analytics Database, Pivotal Greenplum, PostgreSQL, Progress OpenEdge, Salesforce.com, SAP HANA, SAP NetWeaver Business Warehouse, SAP Sybase IQ, Splunk Enterprise, Tableau Data Extract, Teradata V2, Teradata Aster Data nCluster, Teradata OLAP Connector, CSV, Web	None	User Determined

Jaspersoft	<a href="http://jaspersoft.com/">http://jaspersoft.com/</a>	\$4,100 per year - \$154,200 per year depending on EC2 Infrastructure	JDBC 2.1, Bean, Jaspersoft OLAP, XML/A, MySQL, Oracle RDBMS, PostgreSQL, IBM DB2, Microsoft SQL Server, Infobright, Vertica, JBoss Metamatrix, JBoss Teiid, Greenplum DB, Ingres Vectorwise, Netezza, Sybase ASE, Teradata, Mongo DB 2.4x, Hive 1/2, Impala, Cassandra, Salesforce, Spark 1.1, ElasticSearch, Custom Data Sources	Variable depending on product plan	User Determined
------------	---	---	---	------------------------------------	-----------------

Table E.1: Comparison of various dashboard software types.

# Appendix F

## CKAN Web Data Connector

75

The source code on the following pages is the source for the CKAN to Tableau web data connector that was written to enable the Tableau Dashboards to connect to the CKAN portal.

```
1 <html>
2 <meta http-equiv="Cache-Control" content="no-store" />
3 <head>
4   <title> Santa Fe Open Data Web Connector </title>
5   <script src="https://connectors.tableau.com/libs/tableauwdc-1.1.1.js" type="text/javascript"></script>
6   <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.1/jquery.min.js" type="text/javascript"></
   script>
7
8   <script type="text/javascript">
9     resource_id = '';
10    dataToReturn = [];
11    last_record_token = '';
12    var accum = 0;
13    var total = 0;
14    function getCKANData(response)
15    {
16      dataToReturn.push.apply(dataToReturn, response.result.records);
17      var data = {
```



```
18     resource_id: tableau.connectionData, // the resource id
19 };
20 if (accum < total)
21 {
22     var countToFetch = 100;
23     var limit = '';
24     if (total - accum < 100)
25     {
26         countToFetch = total - accum;
27         limit = "&limit=" + countToFetch;
28     }
29     var t = $.ajax({
30         url: 'http://santafeopendata.itsthepeoplesdata.org' + response.result._links.next + limit,
31         dataType: 'jsonp',
32         cache: true,
33         success: function(data) {
34             accum += countToFetch;
35             getCKANData(data);
36         }
37     });
38 }
39 else
40 {
41     console.log('Finished!');
42     console.log('Fetched ' + accum + ' of ' + total + ' records.');
```

```
43     for (i in dataToReturn)
44     {
45         for (var key in dataToReturn[i])
46         {
47             if (dataToReturn[i].hasOwnProperty(key))
48             {
49                 if (/^\$(((1-9)\d{0,2}(,\d{3})*|((1-9)\d*)?\d))(\.\d\d)?/.test(dataToReturn[i][key])
50                     || /\d+(,\d)+./test(dataToReturn[i][key]))
51                 {
52                     dataToReturn[i][key] = dataToReturn[i][key].replace('$', '').replace(',', ', ');
53                 }
54             }
55         }
56     }
```

```

57     tableau.dataCallback(dataToReturn, last_record_token, false);
58 }
59 }
60
61 function makeRadioButton(name, value, text) {
62     var label = document.createElement("label");
63     var radio = document.createElement("input");
64     radio.type = "radio";
65     radio.name = name;
66     radio.value = value;
67     radio.onclick = function ()
68     {
69         resource_id = radio.value;
70     }
71
72     label.appendChild(radio);
73
74     label.appendChild(document.createTextNode(text));
75     return label;
76 }
77
78 (function () {
79
80     var myConnector = tableau.makeConnector();
81
82     myConnector.getColumnHeaders = function () {
83         tableau.log('Fetching headers');
84         tableau.log(tableau.connectionData);
85         var data = {
86             resource_id: tableau.connectionData, // the resource id
87             limit: 1,
88         };
89         $.ajax({
90             url: 'http://santafeopendata.itsthepeoplesdata.org/api/action/datastore_search',
91             data: data,
92             dataType: 'jsonp',
93             cache: true,
94             success: function(data) {
95                 console.log('Got data for headers');

```

```
96     var firstLine = data.result.records[0];
97     var fieldNames = Object.keys(firstLine);
98     var fieldTypes = [];
99     for (i in Object.keys(firstLine))
100     {
101         var key = Object.keys(firstLine)[i];
102         var fieldType = 'string';
103         if (parseInt(firstLine[key]).toString() === firstLine[key]) fieldType = 'int';
104         else if (parseFloat(firstLine[key]).toString() === firstLine[key]) fieldType = 'float';
105         else if (/^\$([1-9]\d{0,2}(,\d{3})*)|([1-9]\d*)?\d)(\.\d\d)?/.test(firstLine[key])
106             || /\d+(\,\d)+\./test(firstLine[key])) fieldType = 'float'; //modified
107         else if (key.toLowerCase().indexOf('datetime') !== -1) fieldType = 'datetime';
108         else if (key.toLowerCase().indexOf('date') !== -1) fieldType = 'date';
109         fieldTypes.push(fieldType);
110     }
111     tableau.log(fieldNames);
112     tableau.log(fieldTypes);
113     tableau.headersCallback(fieldNames, fieldTypes);
114 }
115 });
116 }
117
118 myConnector.getTableData = function(lastRecordToken){
119     var dataToReturn = [];
120     var hasMoreData = false;
121     last_record_token = lastRecordToken;
122
123     var data = {
124         resource_id: tableau.connectionData, // the resource id
125     };
126     var t = $.ajax({
127         url: 'http://santafeopendata.itsthepeoplesdata.org/api/action/datastore_search',
128         data: data,
129         dataType: 'jsonp',
130         cache: true,
131         success: function(data) {
132             total = data.result.total;
133             getCKANData(data);
134         }
135     });
```

```
135     });
136   }
137
138   tableau.registerConnector(myConnector);
139 })();
140
141 window.onload = function () {
142   $.ajax({
143     url: 'http://santafeopendata.itsthepeoplesdata.org/api/action/package_list',
144     dataType: 'jsonp',
145     cache: true,
146     success: function(data){
147       for (i in data.result)
148       {
149         var url = 'http://santafeopendata.itsthepeoplesdata.org/api/action/package_show?id=' + data.result[i];
150         $.ajax({
151           url: url,
152           dataType: 'jsonp',
153           cache: true,
154           success: function(data1) {
155             for (j in data1.result.resources)
156             {
157               if (data1.result.resources[j].datastore_active)
158               {
159                 var radio_home = document.getElementById("radioButtonLocation");
160                 var b = makeRadioButton("dataSourceName", data1.result.resources[j].id, (data1.result.
organization.title + ': '
161                                     + data1.result.resources[j].name));
162                 radio_home.appendChild(b);
163                 radio_home.appendChild(document.createElement('br'));
164               }
165             }
166           });
167         });
168       }
169     }
170   });
171 }
172 $(document).ready(function () {
```

```

173     $("#submitButton").click(function(){
174         var checkedBoxes = $('input[name=dataSourceName]:checked');
175         if (!checkedBoxes || checkedBoxes.length == 0)
176         {
177             return;
178         }
179         tableau.connectionName = checkedBoxes[0].nextSibling.data;
180         tableau.connectionData = resource_id;
181         tableau.submit();
182     });
183 });
184 </script>
185 </head>
186 <body>
187     <p>Choose a dataset, then press "Get the Data" to retrieve data from CKAN.</p>
188     <p><div id="radioButtonLocation"></div></p>
189     <p><button type="button" id="submitButton">Get the Data</button></p>
190 </body>
191 </html>

```

88

Listing F.1: CKAN to Tableau web data connector.