Chapter 2

Using Public Data From Different Sources

Yair Cohen
Farmingdale State University of New York, USA

ABSTRACT

The United States federal government agencies as well as states agencies are liberating their data through web portals. Web portals like data.gov, census.gov, healthdata.gov, ed.gov and many others on the state level provide great opportunity for researchers of all fields. This chapter shows the challenges and the opportunities that lie by merging data from different public sources. The researcher collected and merged data from the following datasets: NYSED school report card, NYSED Fiscal Profile Reporting System, Civil Rights Data Collection, and Census 2010 School District Demographics System. The challenges include data validation, data cleaning, flattening data for easy reporting, and merging datasets based on text fields.

INTRODUCTION

On January 21, 2009 on his first day in his office, President Obama signed two executive orders: The Memorandum on Transparency and Open Government and The Memorandum on the Freedom of Information Act (Cuillier & Piotrowski, 2009). These memorandums stated that the federal administration is committed to creating an unprecedented level of openness in Government, and that transparency will enhance the public trust. It encouraged government transparency as an important step toward government accountability and democracy. Directing the federal agencies to take affirmative steps to make information public and use modern technology to inform citizens about what is known and done by their Government. The executive orders

DOI: 10.4018/978-1-5225-3616-1.ch002
Using Public Data From Different Sources

also called for the Chief Technology Officer (CTO) to present recommendations for Open Government Directive (Obama, 2009a). The recommendation of the federal CTO became the base for President Obama published Open Government Directive (OGD) presented in December 2009. The OGD was based on transparency, participation, and collaboration that were described as the cornerstone of an open government. The OGD has led to the development of data.gov website which provides direct access to enormous amounts of unrefined government data with the hope that researchers, private entities, and other government agencies will be able analyze and develop new uses for the data that will enhance government activities (Bertot & Jaeger, 2010). In 2016 data.gov was elected to be one of the Best Free Business Websites, the explanation stated that data.gov is the “home of the US government’s open data,” it is an online government data aggregator based on its OGD to embrace a new era of open and accountable government. In an age when data are seemingly abundant in our daily lives and with production expanding at an astonishing pace, Data.gov is in its eighth year of bringing data to the American public in an effort to foster government transparency, and is a recognized tool of the open data movement. The platform has spawned a host of data.gov sites in other countries (e.g., Canada, the UK, Australia, France, Singapore), with 106 countries now bringing the open data movement forward, according to a 2016 United Nations global survey. States also followed the OGD and developed websites on the state level. New York State launched in 2013 open.ny.gov a comprehensive state data website that provides user-friendly access to data from New York State agencies, localities, and the federal government. The state of California developed data.ca.gov in order “bring government closer to citizens and start a new shared conversation for growth and progress in our great state”. The state of Texas developed data.tx.gov, each state has developed its own data sharing site following the footsteps of the federal initiative (Best of the best business websites, 2016). As of April 2017, data.gov hosts over 192,000 datasets across government offices. The data catalog includes data from the following departments: agriculture, climate, consumer, ecosystems, education, energy, finance, health, local government, manufacturing, maritime, ocean, public safety, and science & research.

However, not all government agencies connected to data.gov. For example, the US Census Bureau stored its massive data under census.gov, the site hold valuable information for social science researchers on topics like age, sex, migration, household, poverty, education, income, and more.

All the departments that connected to data.gov have created their own site in order to make it easy for the end users to find the dataset they are looking for. The US Department of Health and Human Services (HHS) issued their own website healthdata.gov that host data from the various department agencies (e.g. CCD, FDA, CMS, and others), the department also focused on making data easier for developers
Comparative Performance Evaluation of Effects of Modifier in Asphaltic Concrete Mix