

Semantic Web LOGD Project Report

USA Budget Analysis of Different agencies over a period
of 28 years.

SPRING 2016
SECTION: 501

TABLE OF CONTENTS

Group Information	2
Introduction:.....	3
Target Audience.....	3
Description of data sources.....	4
Data Integration	5
Data Product Results.....	6
SUMMARY	10

GROUP INFORMATION

Group Name: Sunish-Mayank Team

Project Name: USA Budget analysis of various agencies.

Project Type: LOGD

Number of Members: 2

Live Website URL for Demo: <http://www.utdallas.edu/~sss140830/>

Details of Team Member

1. Sunish Sheth – sss140830
2. Mayank Agrawal – mxa144530

INTRODUCTION:

Goal: The main goal of our project is to analyze the ratio of the budget assigned to the agencies by the government and the expenditure of those agencies. Additionally, we analyzed the revenue generated by those companies and compared it with the expenditure. We have used three datasets in our project to understand the following concepts about the government agencies:

- 1) Budget vs Expenditure ratio – Analyzed about the efficiency in assigning budget to various agencies
- 2) Budget vs Revenue Ratio – Future we have used this comparison to calculate the profit made by the agency

Resources Used:

- 1) We used three different Datasets: Dataset 401, Dataset 402, and Dataset 403.
- 2) SPARQL Queries
- 3) Google Visualization API
- 4) HTML
- 5) JavaScript
- 6) Bootstrap

About this Report: This report discusses the target audience for this project. Then it explains the data and discusses about the ontologies used to describe the structured data. After that, some focus is laid on the data integration and its value and then some results are displayed. Finally, the report justifies the LOGD project choice and concludes with a brief summary.

TARGET AUDIENCE

This project can be relevant/useful for the following entities:

1. **US Government** – We have the data about how much US government has assigned Budget to the particular agencies in the respective years. We have data about how much an agency has spent in the respective year (Expenditure). And finally we have the data about how much revenue an agency has generated in that respective year. Hence we can understand two different factors by analyzing the result which would be helpful for the government entities to assign budget in the respective years.
 - a) The budget assigned by the government for the particular agencies is sufficient enough for the agencies or not. We could calculate this from the budget and Expenditure ratio.
 - b) We can calculate the revenue vs expenditure. Hence we could deduce that which all agencies are providing the profit to the government.
2. **Agency** – This Data would be useful for the agencies to be well prepared. As they already have the past year expenditure and budget data. Agencies can also check their performance by comparing past year Profit Percentage. If the company has more requirement, then the budget allocated by the government in the past then it could take preventive actions in advance.

3. Users – This will be helpful for the user who are investing in the agencies. They could make good decision based on companies' performance based on their yearly performance.

DESCRIPTION OF DATA SOURCES

We have used the following two different datasets:

Dataset 401: This dataset provides the information about the actual budget assigned to the agencies by the government.

Dataset 402: This dataset provides the information about the actual expenditure by the agencies in the various tasks that they have performed. This Entity could be greater than the value in Dataset 401 because they could spend less than what was actually assigned to them and could spend little bit more because the budget assigned to them was not enough to perform all the tasks.

Dataset 403: This Dataset provides the revenue generated by these agencies. This revenue is based on the expenditure value that took out of budget in Dataset 402.

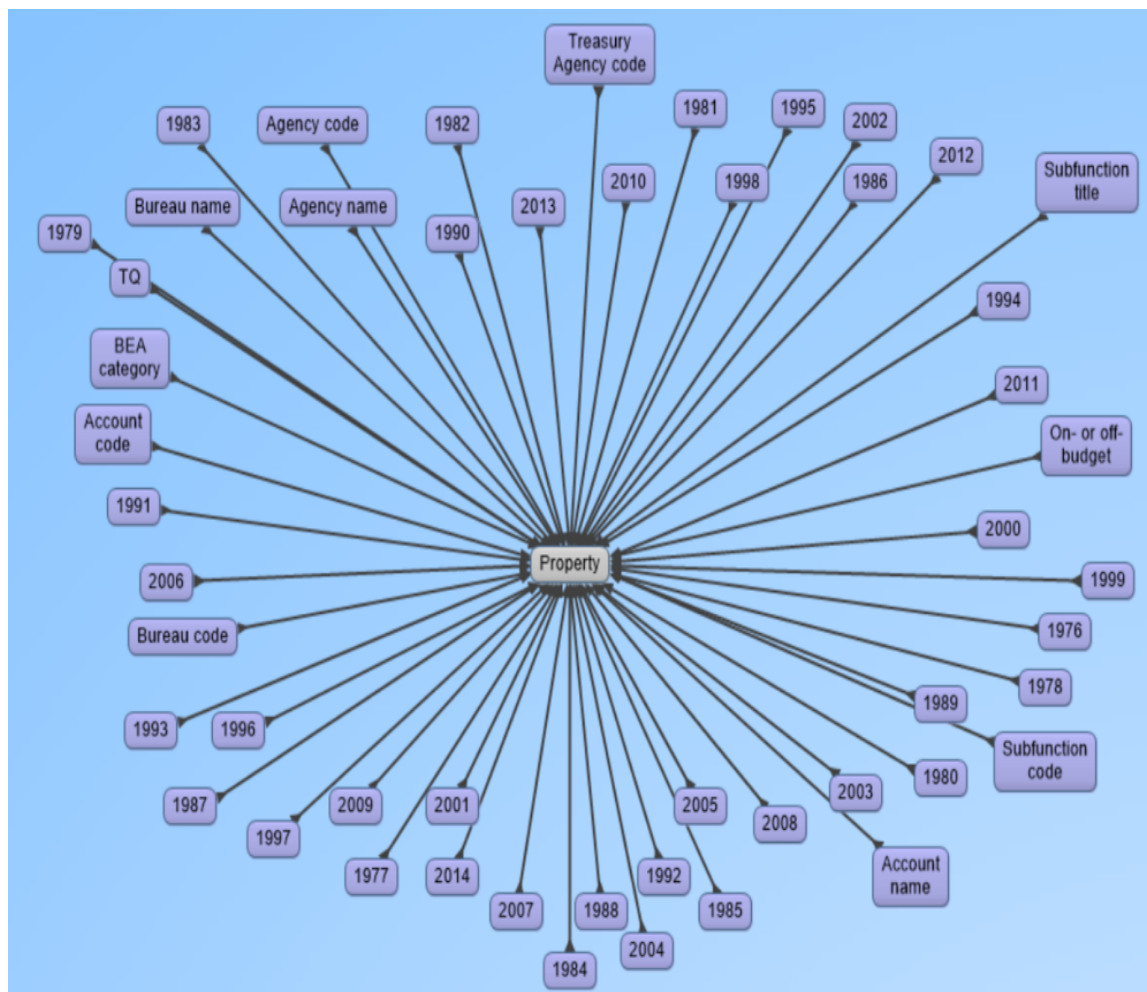


Fig 1: Shows the different predicates or property in our data

The above diagram shows the different properties or the different attributes of our data that we obtained from each dataset.

DATA INTEGRATION

Data is integrated in the following steps:

- 1) Dataset 401, Dataset 402 and Dataset 403 are available publicly online.
- 2) We wrote Sparql query to fetch the data as follows.
 - a) We got the sum of all the budget assigned to any particular agency for the given year – Dataset 401.
 - b) We got the sum of the expenditure (money spent) by any particular agency for the given year – Dataset 402.
 - c) We added the above result to obtain the total revenue generated by the agency in the particular year.
 - d) We got the sum of the revenue (money gained) by any particular agency for the given year – Dataset 403.
 - e) We then took an intersection of the above data to find out the common agencies from the above list.
 - f) We combined all the three dataset on agency name. These agencies are different departments in our government which are provided with budget from our government. For example, Department of Health Care, Department of Agriculture, Judicial Branch. All these departments are funded by the government. Government keeps tracks of their expenses and revenue generated from these agencies.
 - g) Hence, we calculated the profit or loss based on the subtraction of revenue and expenditure from the above data. Thus, this would help people and the government make informed decisions on the next budget for gaining maximum profit.
- 3) Passing the result from the SPARQL query to the Google Visualization API.

DATA PRODUCT RESULTS

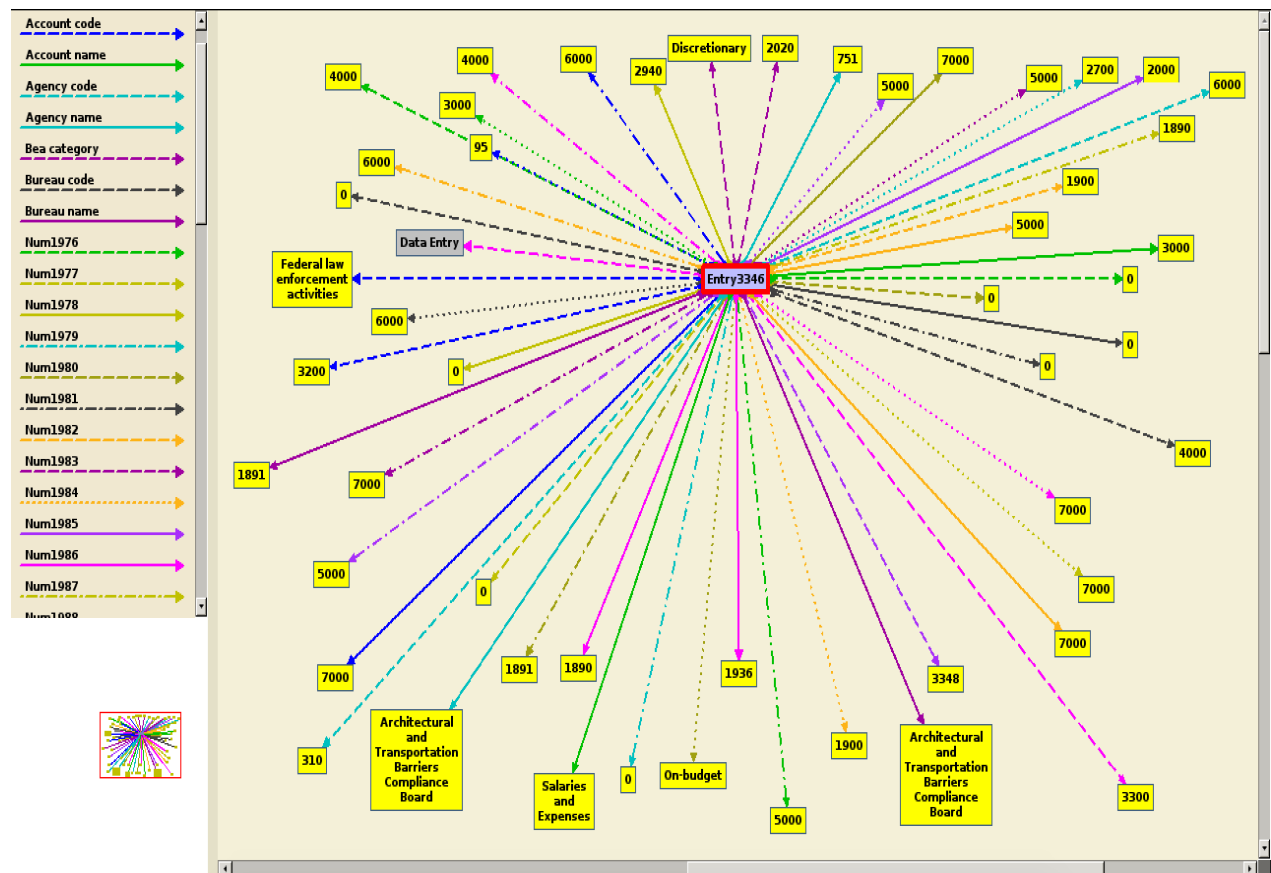


Fig 2. Shows the actual data from the dataset of an Entry3346

Suppose if we will consider **Entry3346**. Every Entry in the dataset has data updated over many years (1976-2014). Every Entry corresponds to one Agency, its corresponding Agency Name, Account name, Bureau Name and their respective codes. Our intension is to find the data corresponding to each Agency across all the Accounts for each year. So we need to add the data for the same Agency name spread across multiple tuples for the Account name. Thus we will find the data for that particular Agency, for a particular year across all the Accounts.

We find the same data from each dataset (401 - budget allocated, 402 - outlays/Expenditure, 403 - Receipts/Revenue) for a particular agency in a particular year to compare results.

This will help us deduce Budget to Expenditure ratio and also take a look at the Revenue model and the Profit or Loss generated from each agency over the period and thus making wise decision on the amount of money allocated to each agency.

SPARQLer Query Results

year	callret-1	callret-2	callret-3
"http://data-gov.tw.rpi.edu/vocab/p/401/num1976"	"345651"	"325188"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1977"	"431149"	"393091"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1978"	"459168"	"436667"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1979"	"521374"	"481357"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1980"	"608924"	"566677"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1981"	"655993"	"640755"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1982"	"733473"	"709648"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1983"	"822519"	"786872"	"0"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1984"	"904067"	"866355"	"12349"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1985"	"1054691"	"966226"	"56603"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1986"	"1045239"	"1071016"	"2199"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1987"	"1273133"	"1180500"	"2851"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1988"	"1343581"	"1342371"	"2615"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1989"	"1480851"	"1498507"	"25651"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1990"	"1752108"	"1646115"	"42403"
"http://data-gov.tw.rpi.edu/vocab/p/401/num1991"	"2118278"	"1996763"	"57495"

Fig 3. Sample SPARQLer Query Result for Judicial Branch from all the 3 datasets. Callret-1 corresponds to Budget, Callret-1 corresponds to Expenditure, Callret-1 corresponds to Revenue.

WEBSITE ACCESS:

The entire project is live on the website. URL: <http://www.utdallas.edu/~sss140830/>

USA Government Revenue System
all values in dollars (USD)

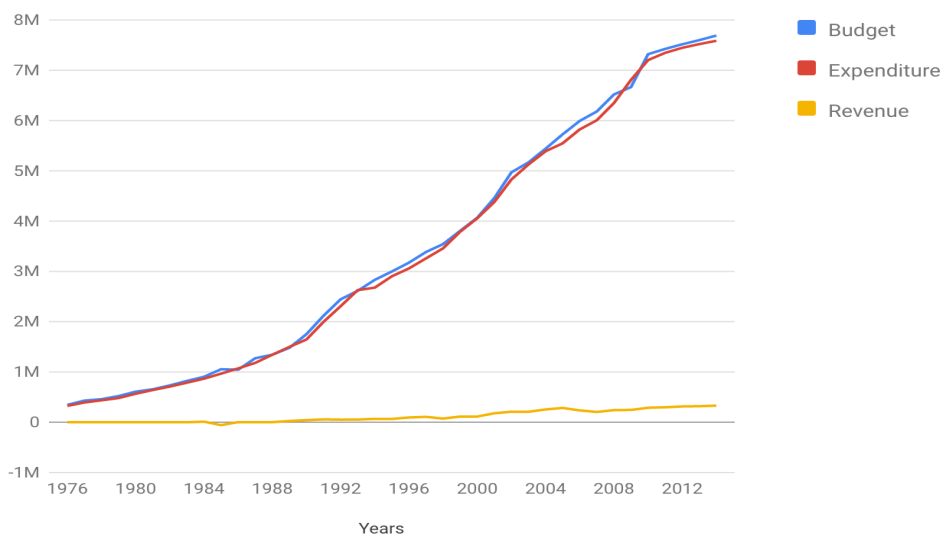


Fig 4: This figure shows the aggregated graph of the budget v/s Expenditure v/s Revenue for Judicial Branch Agency.

USA Government Revenue System

all values in percentage

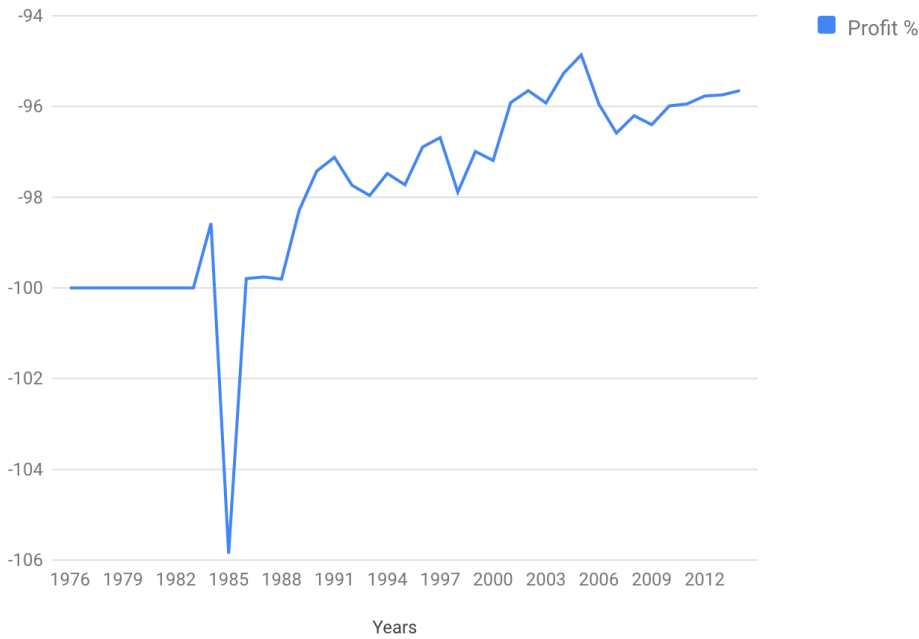
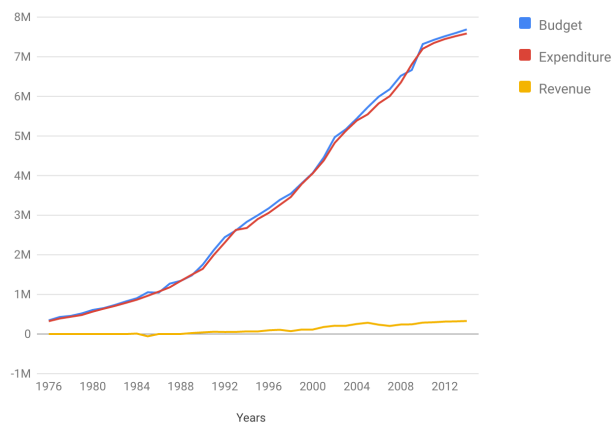


Fig 5: This figure shows the Profit/Loss percentage graph for Judicial Branch Agency.

Judicial Branch ▾

USA Government Revenue System

all values in dollars (USD)



USA Government Revenue System

all values in percentage

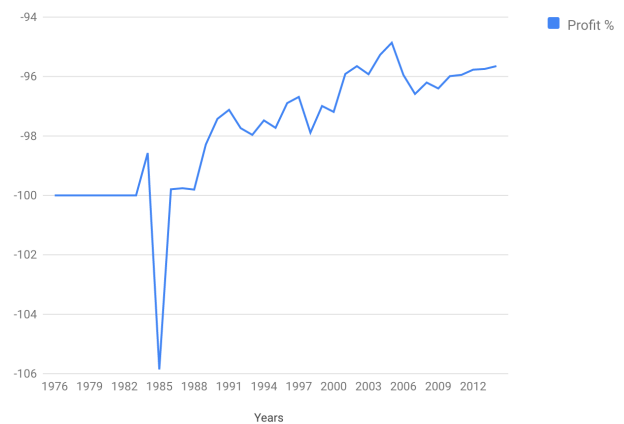


Fig 6: This figure shows the layout of the website and portraying graphs of Judicial Branch Agency.



Fig 7: The user can change the agency to see its respective agency results by selecting agencies from a list displayed in the top left corner.

Description	Timeline for an agency's aggregated Budget/ Expenditure /Revenue from 1976 to 2014. Finally calculating the Profit Percent per agency
Team Name	Sunish - Mayank Team
Creator	<ul style="list-style-type: none"> • Sunish Sheth • Mayank Agarwal
Created	2016/04/07
Datasets	<ul style="list-style-type: none"> • Dataset 401 (Public Budget Database - Budget Authority and offsetting receipts 1976-2014, Executive Office of the President) • Dataset 402 (Public Budget Database - Outlays and offsetting receipts 1962-Current, Executive Office of the President) • Dataset 403 (Public Budget Database - Governmental receipts 1976-2014, Executive Office of the President)
SPARQL queries	<ul style="list-style-type: none"> • http://www.utdallas.edu/~sss140830/JudicialBranch.sparql • http://www.utdallas.edu/~sss140830/DepartmentofAgriculture.sparql • http://www.utdallas.edu/~sss140830/AffordableHousingProgram.sparql • http://www.utdallas.edu/~sss140830/CorpsOfEngineers-CivilWorks.sparql • http://www.utdallas.edu/~sss140830/DepartmentofCommerce.sparql • http://www.utdallas.edu/~sss140830/DepartmentofHealthandHumanServices.sparql
SPARQL endpoint	http://data-gov.tw.rpi.edu/sparql

Fig 8: This figure shows the description of the project and the team working on it.

Any user can access the website mentioned above. One can check the SPARQL queries going to the link mentioned in the website. The above website has a simple and a clean User Interface so it is very easy to access. One can click on the list on the top right corner to change the agencies, the user wants to see the data for. The graphs will change dynamically when the agency is changed. Every time the agency is changed, the SPARQL query is executed and the result obtained is feed to the Google Visualization API and the graphs is updated on the website.

SUMMARY

With the help of this project we have great understanding about how to fetch relevant data and present it to the user. The Critical section of this project is to pick and understand the dataset and then analyze them to get some meaningful output. This project has provided us the opportunity to work and coordinate in a team. This project helped us understand the Google Visualization API, work on large datasets.

Lessons Learned: The most important thing to consider here is thinking in a group. While thinking in a group you need to come to some common conclusion. We had enormous amount of the data. Each and every data provides a separate information. We need to gather some data out of those and get the common understanding. Second major step was to analyze the data and deduce some meaningful expression. At a particular point we were not getting the desired result we were looking for. We then decided to add a new Dataset (Dataset 402), which would help get us our desired result and we would justify our results in a better manner.

Things that could have been done differently: Currently we are writing different queries for each agency. We could instead have written one query and dynamically changed the value of the agency and obtained our results. Thus providing better efficiency and modularization.