OpenBudgets.eu: Fighting Corruption with Fiscal Transparency

Deliverable 3.2

Visualisations Creation Framework

Dissemination Level	Public
Due Date of Deliverable	Month 16, 31.08.2016
Actual Submission Date	29.09.2016
Work Package	WP3, Budget and Spending Data Visualisation and Exploration
Task	T3.2
Type	Demonstrator
Approval Status	Final
Version	0.2
Number of Pages	14
Filename	Deliverable 3.2 - Visualisations Creation Framework

Abstract: This document documents the selection of the OpenSpending viewer for the basis of the OpenBudgets.eu visualizations Framework. The OS Viewer features are presented and additional development is described.

History

Version	Date	Reason	Revised by
0.1	12.08.2016	First contribution – structuring and introduction	Lazaros Ioannidis
0.3	28.09.2016	Interim review by Sotiris Karabatakis	Sotiris Karabatakis
0.3	28.09.2016	Corrections after review by Paul Walsh	Lazaros Ioannidis, Charalampos Bratsas

Author List

Organisation	Name	Contact Information
OKFGR	Lazaros Ioannidis	larjohn@okfn.gr
OKFGR	Panagiotis-Marios Philippides	filippidis.okfgr@gmail.com
OKFGR	Sotiris Karampatakis	karampatakis@okfn.gr
OKFGR	Charalampos Bratsas	charalampos.bratsas@okfn.org

Executive Summary

The amount of observations in a typical fiscal dataset is prohibitive for a human to process. visualisations provide a more comprehensive and efficient way to understand and analyse large datasets, as stakeholders have already pointed out. The analysis of the data with common operations, as defined in the data cube methodology can result in meaningful visualisations that can be rendered and interacted upon in the users' browsers. Open Spending[1], a sophisticated solution for visualizing fiscal datasets was selected, further developed and integrated into OpenBudgets.eu, as its Viewer component satisfies the stakeholders requirements with various aggregate visualisations. A middleware component has also been developed to provide the Viewer with data from the OpenBudgets.eu triple store.

---

[1] http://next.openspending.org

Abbreviations and Acronyms

OS	OpenSpending
OBEU	OpenBudgets.eu
RDF	Resource Description Framework
FDP	Fiscal Data Package
SVG	Scalable Vector Graphics

Table of Contents

List of Figures

1 Introduction

1.1 Requirements

Fiscal datasets typically contain hundreds or thousands of observations. The more elementary the data, the more useful they can be proven to be, because they allow a wider range of analysis. The analysis operations highlight the value of fiscal data by producing a new set of figures that are the projection of the original data into a more specific field of interpretation. The interpretations vary from predictions over a specific time series, to aggregations, descriptive statistics and others. In either case, the resulting figures are meaningful to humans, by the time they are visually perceptible.

The objective of data visualizations is to make this perception effective and efficient. The perception can be considered effective if the perceiving person is able to analyze and reason about data and evidence. An efficient perception is one that consumes the least possible cognitive effort for the viewer to be effective.

From this point of view, in D3.1 users of fiscal data visualizations have outlined the visualizations that fit their expectations of interpreting the data. As noted, there are five large groups of visualizations that have been identified and can be considered as requirements for the visualizations framework of OpenBudgets.eu:

R1. Showing breakdown of totals
R2. Contextualising big numbers
R3. Showing trends over time
R4. Showing networks of financial flows
R5. Comparing budgeting, commitment and execution

Many of these visualization use cases base their efficiency on interactivity. For instance, in data cube aggregations, the visualization can be more useful, when the user can, in real time, change the drilldown dimension or filter a dimension to focus to specific parts of the observations set. From OpenBudgets.eu, where the DataCube ontology is used to describe the observations, this interaction should be built around data cube operations like slicing and drilling down.

When a visualization represents what its creator desires, it can be used as an efficient means of communication for the rest of the dataset stakeholders. Sharing a fiscal data visualization means that it can be exported outside of its authoring environment and be embedded in other resources. Also, as reverse engineering the visualization to reach back to data is not possible, visualizations might also carry a certification of authenticity and validity against the source data.

Given the above assumptions, and the overall progress of the project till now, the OpenBudgets.eu visualizations framework should have the following properties:

P1. Support the visualization requirements as outlined in D3.1 by offering equivalent visualizations, applicable on fiscal data
P2. Support interactive data exploration, at least per the common data cube specifications
P3. Allow the user to export the visualization and reuse it online
P4. Allow community developers to extend and reuse the implementation

In the following sections, existing approaches are going to be reviewed, based on the aforementioned assumptions. Finally, we present the selected module, on which the OpenBudgets.eu visualization framework will be based - the OpenSpending Viewer.

2 Implementation

2.1 Existing Implementations

In OpenBudgets.eu the data visualization components are required to be web based, in order to be openly accessible to all the stakeholders. According to this requirement the data flow will start from accessing the data source. Ideally, this interaction will happen transparently with an HTTP API. The data then can then get transformed at the browser and finally fed into a JavaScript visualization library, client-side.

Currently, the most popular and actively developed approach for visualizing data in the browser is D3.js, an open source JavaScript library usually used to generate SVG graphics. On top of it, there is NVD3, which provides ready-made, but very customizable standard chart components. An alternative SVG library is Raphael.js, while a third drawing library, paper.js uses the HTML Canvas element to render graphics.

There are already various applications that are built on top of these libraries and include GUIs for data analysis. While most of the applications are proprietary, there exists an open source option oriented towards data analysis, Weave, a platform built with Adobe Air. While Weave possesses some of the properties required, it is fairly complex to be used by the end users of OpenBudgets.eu. For this project, OpenSpending is developed as part of the foundation platform with the following qualities:

1. It is completely modular, which means we can replace or improve components without breaking the whole architecture
2. It is already made with fiscal data in mind
3. It supports the common data cube operations
4. It is completely based on open source components (D3.js is the rendering component)
5. It is actively developed and supported by a worldwide community

2.2 OpenSpending and the OpenSpending Viewer

As described in the project's source code repository, a project to make government finances easier to explore and understand. It started out as "Where does my money go", a platform to visualize the United Kingdom's state finance, but has been renamed and restructured to allow arbitrary financial data to be loaded and displayed. The latest version of OpenSpending was designed to be modular, so that third party applications can be easily built and integrated by organizations and users. The default OpenSpending data flow is represented in the following figure.
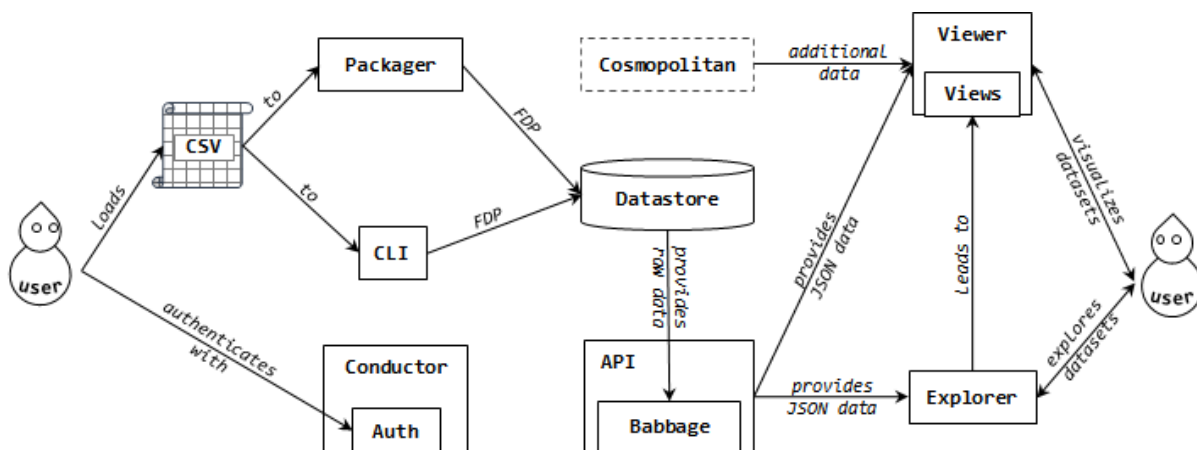

Figure 1: Visualisations Creation Framework

The component of interest for this deliverable is the OpenSpending Viewer. The OS Viewer is a modern, graphical shell around the Babbage UI module of OpenSpending. While the original

Babbage UI module is the one that actually implements the visualizations, the OS Viewer provides a responsive interface that facilitates the common data cube operations via a user-friendly interface. The user interface of the OpenSpending Viewer can be seen in the annotated image below.
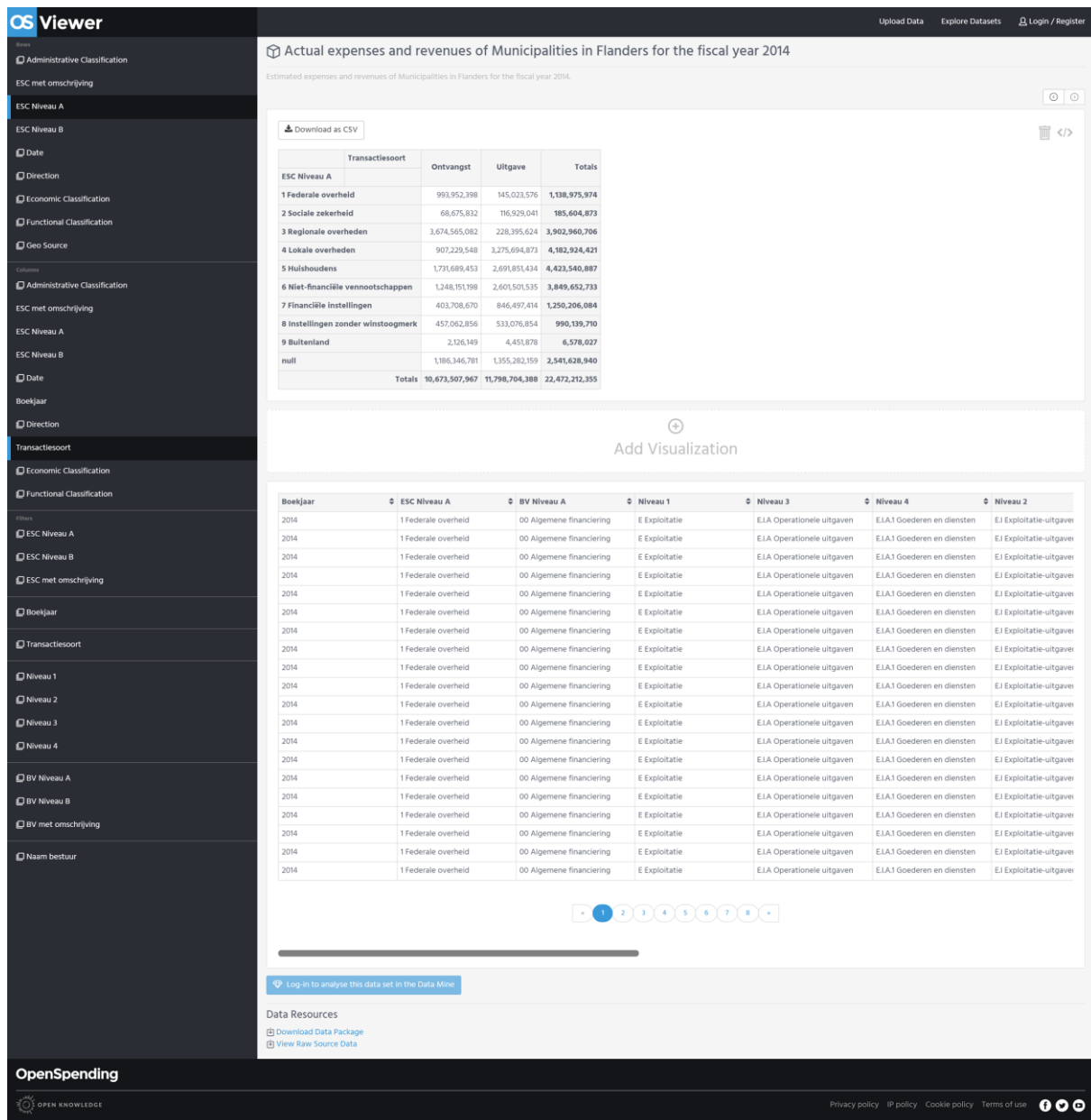

Figure 2: The OpenSpending Viewer

In the left part of the interface, the dimensions of the data set are displayed to allow the user to select those that will be used to aggregate the data subset. The subset is also defined by filtering the dataset observations, by providing specific values to the responsive dimensions. Depending on the already built aggregate, the OpenSpending viewer adapts the interface to restrict the available options and assist the users create their visualizations. OpenSpending currently supports the visualization type described below.

2.2.1 Pivot table & Facts Table

The Facts table is the simplest form of visualization in the OS Viewer. It is a paginated tabular view of the observations. One step further is the pivot table which displays summarized data based on an aggregation. The observations are first grouped by one or more dimensions and their measures are aggregated using the sum function. The users can select the aggregation

dimensions, filter observations by one or more dimensions, and define sorting and pagination of the aggregated results. This same generalized aggregation method is used in all the following visualizations.

admin2	approved	adjusted	executed
289 National Social Insurance Company	95,669,941,944	97,723,755,276	96,639,396,103.05
322 National Health Insurance Company	47,822,780,140	47,393,840,372	46,345,733,744
200 General actions	37,940,295,934	37,890,814,943.87	37,655,817,669.860
001 Council of the municipality of Chisinau	20,420,115,377.99	23,150,265,864.360	21,610,698,546.780
129 Ministry of Education	12,787,662,120	13,764,292,063.54	12,935,482,955.540
133 Ministry of Internal Affairs	9,945,057,180	10,826,536,949.97	10,658,835,683.17
128 Ministry of Health	8,046,841,884	9,105,786,969.06	8,855,352,769.11
125 Ministry of Agriculture and Food Industry	7,392,266,204	8,994,031,594.99	8,783,593,021.870
264 Ministry of Transports and Road Infrastructure	9,025,987,704	9,182,846,216.04	8,675,358,773.63
131 Ministry of Justice	4,871,891,655	5,081,240,088.99	4,850,558,983.900
132 Ministry of Defense	4,279,516,976	4,598,410,010.48	4,535,867,952.920
440 Fund for financial support to administrative-territorial units	3,524,600,000	4,479,650,704	4,479,032,204
122 Ministry of Finance	4,568,062,023	4,620,535,541.14	4,449,914,756.430

Figure 3: Pivot Table

2.2.2 Line & Bar charts

The line chart displays information as a series of data points connected by straight line segments. The observations are grouped by a single dimension. An ideal use case is when the user wants to visualize a trend in data over intervals of time – a time series – thus the line is drawn chronologically.

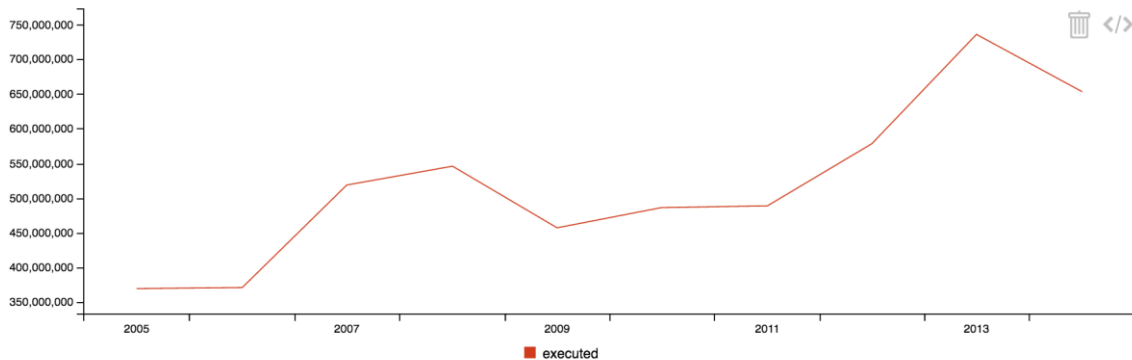

Figure 4: Line Chart

The bar chart presents the grouped data with rectangular bars with lengths proportional to the values that they represent. Each bar represents a value of the categorical variable that represents the drilldown dimension.
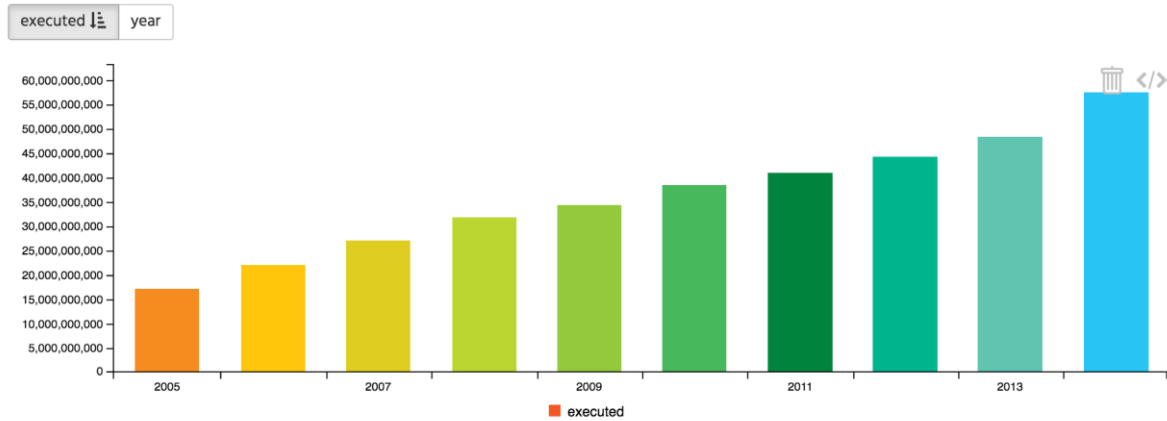
Figure 5: Bar chart

2.2.3 Tree Map

The tree map displays data grouped by one dimension hierarchically (tree-structured) as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension of the data.
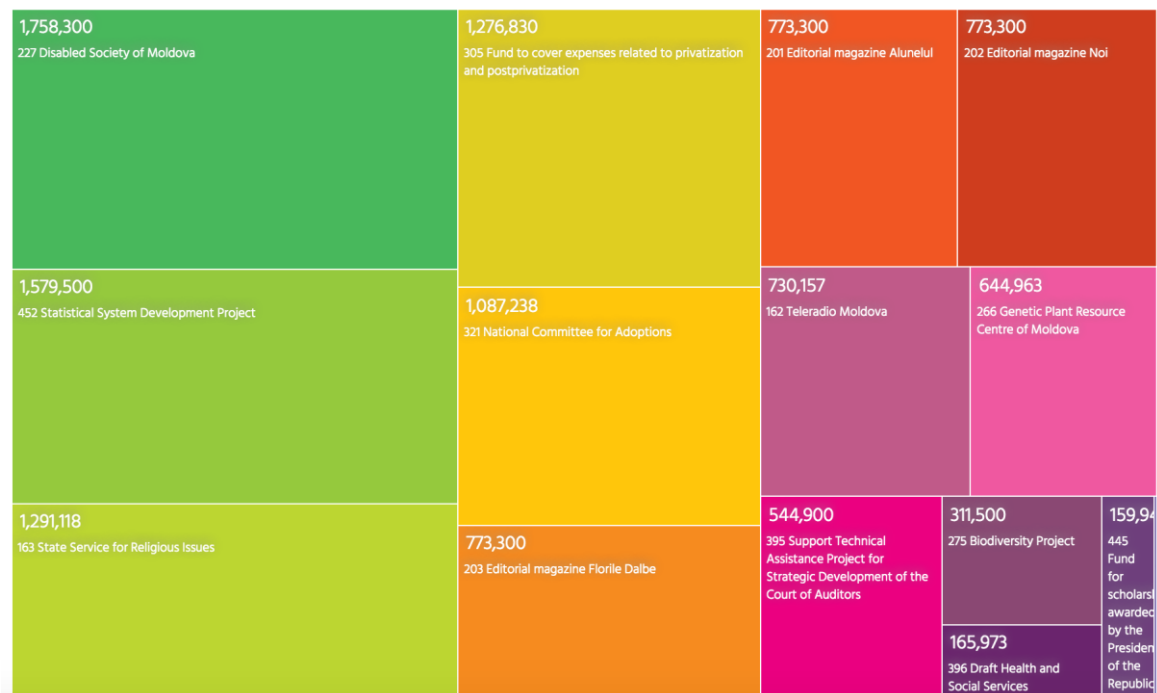

Figure 6: Tree map

2.2.4 Bubble tree

Bubble tree provides a similar visualization as the tree map. The rectangular areas are replaced by circles that orbit around the total amount.
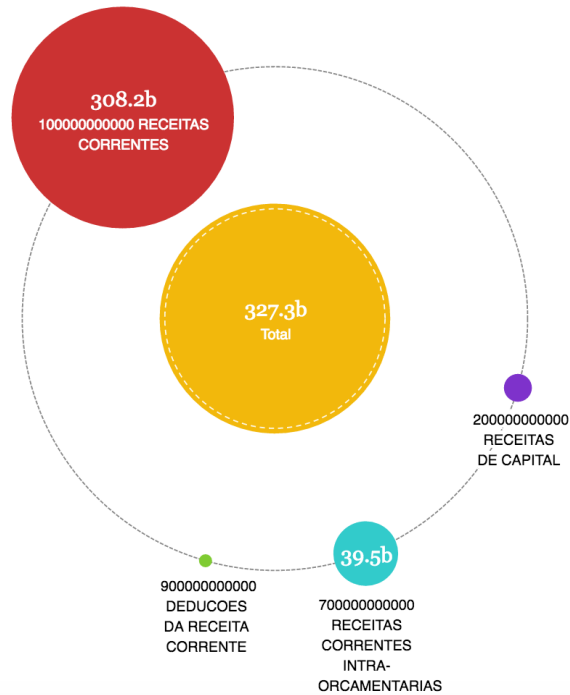
Figure 7: Bubble Chart

2.2.5 Map

If location data is provided in the budget data, then the OS Viewer is using it to provide a map. The region colors represent the aggregated amounts of the grouped observations.

Figure 8: Geo-map

2.2.6 Sankey

The Sankey diagram is a specific type of flow diagram, in which the width of the arrows is shown proportionally to the flow quantity. This diagram can be used to visualize the relationship between two dimensions, by which the observations are grouped.

Figure 9: Sankey

2.2.7 Radar Chart

The radar chart is a chart and/or plot that consists of a sequence of equal-angle spokes, with each spoke representing one of the first dimension values. The data length of a spoke is proportional to the magnitude of the measure for the data point relative to 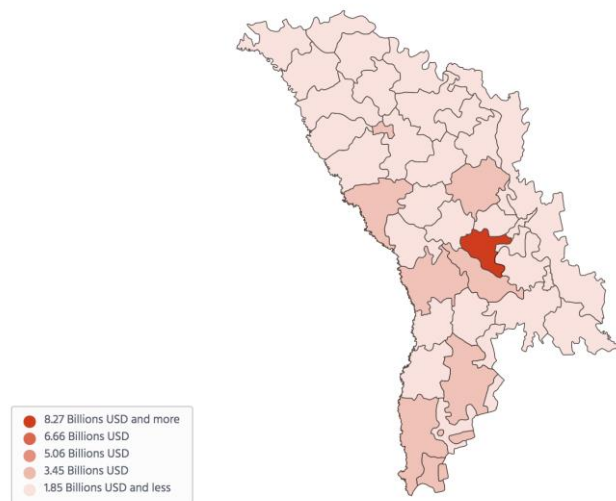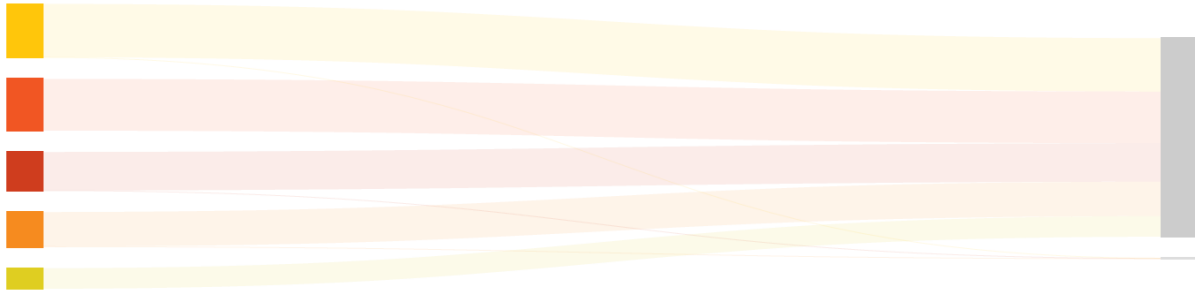the maximum magnitude of the second dimension across all data points. A line is drawn connecting the data values for each spoke.
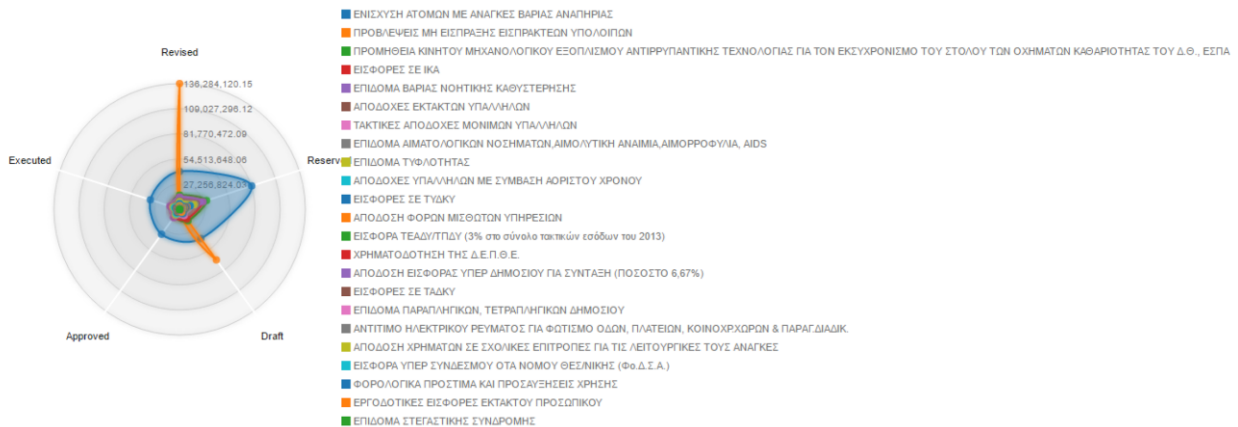
ΕΝΙΣΧΥΣΗ ΑΤΟΜΩΝ ΜΕ ΑΝΑΓΚΕΣ ΒΑΡΙΑΣ ΑΝΑΠΗΡΙΑΣ
ΠΡΟΒΛΕΨΕΙΣ ΜΗ ΕΙΣΠΡΑΞΗΣ ΕΙΣΠΡΑΚΤΕΩΝ ΥΠΟΛΟΙΠΩΝ
ΠΡΟΜΗΘΕΙΑ ΚΙΝΗΤΟΥ ΜΗΧΑΝΟΛΟΓΙΚΟΥ ΕΞΟΠΛΙΣΜΟΥ ΑΝΤΙΡΡΥΠΑΝΤΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΓΙΑ ΤΟΝ ΕΚΣΥΧΡΟΝΙΣΜΟ ΤΟΥ ΣΤΟΛΟΥ ΤΩΝ ΟΧΗΜΑΤΩΝ ΚΑΘΑΡΙΟΤΗΤΑΣ ΤΟΥ Δ.Θ., ΕΣΠΑ
ΕΙΣΦΟΡΕΣ ΣΕ ΙΚΑ
ΕΠΙΔΟΜΑ ΒΑΡΙΑΣ ΝΟΗΤΙΚΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ
ΑΠΟΔΟΧΕΣ ΕΚΤΑΚΤΩΝ ΥΠΑΛΛΗΛΩΝ
ΤΑΚΤΙΚΕΣ ΑΠΟΔΟΧΕΣ ΜΟΝΙΜΩΝ ΥΠΑΛΛΗΛΩΝ
ΕΠΙΔΟΜΑ ΑΙΜΑΤΟΛΟΓΙΚΩΝ ΝΟΣΗΜΑΤΩΝ,ΑΙΜΟΛΥΤΙΚΗ ΑΝΑΙΜΙΑ,ΑΙΜΟΡΡΟΦΥΛΙΑ, AIDS
ΕΠΙΔΟΜΑ ΤΥΦΛΟΤΗΤΑΣ
ΑΠΟΔΟΧΕΣ ΥΠΑΛΛΗΛΩΝ ΜΕ ΣΥΜΒΑΣΗ ΑΟΡΙΣΤΟΥ ΧΡΟΝΟΥ
ΕΙΣΦΟΡΕΣ ΣΕ ΤΥΔΚΥ
ΑΠΟΔΟΣΗ ΦΟΡΩΝ ΜΙΣΘΩΤΩΝ ΥΠΗΡΕΣΙΩΝ
ΕΙΣΦΟΡΑ ΤΕΑΔΥ/ΤΠΔΥ (3% στο σύνολο τακτικών εσόδων του 2013)
ΧΡΗΜΑΤΟΔΟΤΗΣΗ ΤΗΣ Δ.Ε.Π.Θ.Ε.
ΑΠΟΔΟΣΗ ΕΙΣΦΟΡΑΣ ΥΠΕΡ ΔΗΜΟΣΙΟΥ ΓΙΑ ΣΥΝΤΑΞΗ (ΠΟΣΟΣΤΟ 6,67%)
ΕΙΣΦΟΡΕΣ ΣΕ ΤΑΔΚΥ
ΕΠΙΔΟΜΑ ΠΑΡΑΠΛΗΓΙΚΩΝ, ΤΕΤΡΑΠΛΗΓΙΚΩΝ ΔΗΜΟΣΙΟΥ
ΑΝΤΙΤΙΜΟ ΗΛΕΚΤΡΙΚΟΥ ΡΕΥΜΑΤΟΣ ΓΙΑ ΦΩΤΙΣΜΟ ΟΔΩΝ, ΠΛΑΤΕΙΩΝ, ΚΟΙΝΟΧΡΩΡΩΝ & ΠΑΡΑΓΔΙΑΔΙΚ.
ΑΠΟΔΟΣΗ ΧΡΗΜΑΤΩΝ ΣΕ ΣΧΟΛΙΚΕΣ ΕΠΙΤΡΟΠΕΣ ΓΙΑ ΤΙΣ ΛΕΙΤΟΥΡΓΙΚΕΣ ΤΟΥΣ ΑΝΑΓΚΕΣ
ΕΙΣΦΟΡΑ ΥΠΕΡ ΣΥΝΔΕΣΜΟΥ ΟΤΑ ΝΟΜΟΥ ΘΕΣ/ΝΙΚΗΣ (Φο.Δ.Σ.Α.)
ΦΟΡΟΛΟΓΙΚΑ ΠΡΟΣΤΙΜΑ ΚΑΙ ΠΡΟΣΑΥΞΗΣΕΙΣ ΧΡΗΣΗΣ
ΕΡΓΟΔΟΤΙΚΕΣ ΕΙΣΦΟΡΕΣ ΕΚΤΑΚΤΟΥ ΠΡΟΣΩΠΙΚΟΥ
ΕΠΙΔΟΜΑ ΣΤΕΓΑΣΤΙΚΗΣ ΣΥΝΔΡΟΜΗΣ

Revised

136,284,120.15
109,027,296.12
81,770,472.09
54,513,648.06
27,256,824.03

Executed   Reserv

Approved   Draft

Figure 10: Radar Chart

2.3 Integration

In the original implementation of OpenSpending, the OS Viewer interacts with Babbage API, which consumes data from a relational database, described according to FDP, a simple, open technical specification for government budget and spending data. To support the integration of the OpenSpending Viewer into OpenBudgets.eu, the OS Viewer should be able to access data described with the OBEU RDF fiscal data model [1]. There are the following options to implement this support:

1. Transform the data into FDP
2. Replace the data module of OpenSpending (Babbage API)

In OpenBudgets.eu, both options are viable due to the quality of the RDF-described data. For this deliverable, we developed the second solution by mimicking the functionality of the Babbage API, replacing at the same time the relational database of Babbage with the triple store. The middleware developed, Rudolf [2], is also going to bring in new features, like the global cube, which will make comparing data from different datasets more meaningful.

Like Babbage API, Rudolf exposes 4 endpoints:

1. Cubes: provides a list with all the dataset available at the triple store
2. Model: provides an FDP-like representation of the structure of a specific dataset
3. Facts: provides paginated, filtered and ordered observations of a specific dataset
4. Members: provides an enumeration of all the possible values of a dimension in a specific dataset

5. Aggregates: provides a filtered, paginated, ordered result after aggregating observations with an arbitrary number of drilldown dimensions

Behind the scenes, Rudolf translates the requests to each endpoint into a SPARQL query and submits it to the configured triple store. As the OBEU RDF schema is not completely aligned with FDP, some assumptions have to be made in order to simulate the full functionality of Babbage in Rudolf. For instance, selecting the key and the value attribute for dimensions is a heuristic process, looking for properties like SKOS notation and preferred label inside the data observations. The results of the query are transformed and returned according to the OpenSpending specification. Code lists will be taken account in following versions, in order to aggregate similar items together [3], [4].

OpenSpending will be continuously developed as the core of the OpenBudgets.eu with new applications, features, improvements and bug fixing. The OpenSpending Viewer, together with Rudolf will be part of OpenBudgets.eu first prototype, as specified in the description of deliverable D4.3.

3 References

[1] L. Ioannidis, C. Bratsas, P. Filippidis, S. Karabatakis and P. Bamidis, "Rudolf: an HTTP API for exposing semantically represented fiscal OLAP cubes," in *11th IEEE International Workshop on Semantic and Social Media Adaptation and Personalization, Special Session on "Data Impact: Big, Open, Linked Data innovations" Proceedings*, Thessaloniki, 2016.

[2] J. Mynarz, V. Svátek, S. Karabatakis, J. Klímek ja C. Bratsas, „Modeling fiscal data with the Data Cube Vocabulary," *SEMANTiCS conference, CEUR Workshop Proceedings*, 2016.

[3] P. Filippidis, S. Karabatakis, L. Ioannidis, J. Mynarz, V. Svátek ja C. Bratsas, „Towards Budget Comparative Analysis: the need for Fiscal Codelists as Linked Data," *SEMANTiCS conference, CEUR Workshop Proceedings*, 2016.

[4] P. Filippidis, S. Karabatakis, K. Koupidis, L. Ioannidis ja C. Bratsas, „The Code Lists Case: Identifying and Linking the Key Parts of Fiscal Datasets," *11th IEE International Workshop on Semantic and Social Media Adaptation and Personalization Proceedings - Special session on "Data impact: Big, Open, Linked Data innovations"*, 2016.