

OpenBudgets.eu: A Platform for Semantically Representing and Analyzing Open Fiscal Data

Fathoni A. Musyaffa¹, Lavdim Halilaj^{1,2}, Yakun Li², Fabrizio Orlandi^{1,2},
Hajira Jabeen¹, Sören Auer^{3,4}, and Maria-Esther Vidal^{3,5}

¹ University of Bonn, Bonn, Germany

² Fraunhofer IAIS, Sankt Augustin, Germany

³ TIB Leibniz Information Center Science and Technology, Germany

⁴ L3S Research Center, University of Hannover, Germany

⁵ Simón Bolívar University, Venezuela

Abstract. Budget and spending data are among the most published Open Data datasets on the Web and continuously increasing in terms of volume over time. These datasets tend to be published in large tabular files – without predefined standards – and require complex domain and technical expertise to be used in real-world scenarios. Therefore, the potential benefits of having these datasets open and publicly available are hindered by their complexity and heterogeneity. Linked Data principles can facilitate integration, analysis and usage of these datasets. In this paper, we present OpenBudgets.eu (OBEU), a Linked Data -based platform supporting the entire open data life-cycle of budget and spending datasets: from data creation to publishing and exploration. The platform is based on a set of requirements specifically collected by experts in the budget and spending data domain. It follows a micro-services architecture that easily integrates many different software modules and tools for analysis, visualization and transformation of data. Data is represented according to a logical model for open fiscal data which is translated into both RDF data and a tabular data formats. We demonstrate the validity of the implemented OBEU platform with real application scenarios and report on a user study conducted to confirm its usability.

1 Introduction

According to a recent survey done by the Open Knowledge Foundation for the Global Open Data Index (GODI) [15], among 94 observed countries, 88 have made their budget data publicly available. This makes budget data the most popular type of open data published by public administrations on a country level. The Open Data Barometer (ODB) [14] reports that budget and spending datasets are among the most important datasets, along with company registers, contracts, and land ownerships, that are needed to restore public trust.

The significance of publishing open government data, including spending and budget data, are stressed by several motivations, ranging from the improvement of: democratic control and participation in politics [6,19], transparency and compliance [17,19], law enforcement [6], efficiency, and effectiveness [19]. Opening up

data also generally reduces the barrier between government and citizens and enables comparative analysis [6,19]. An interesting case study, demonstrating the advantages of publishing budget data, has been conducted in Brazil and described in [5]. However, we can gain much deeper insights from the published data if the datasets can be compared across different public administrations that have similar characteristics. The linked data paradigm can help here to harmonize and analyze open budget and spending data. A major challenge, however, is to devise a software platform, which facilitates the harmonization of heterogeneous budget and spending data, while facilitating a variety of applications ranging from comparative analysis to participatory budgeting.

In this paper, we present the OpenBudgets.eu (OBEU) platform for linked open budget and spending data analysis. We collect requirements for a linked budget and spending data analytics platform, we illustrate use cases using several actual datasets and provide the platform design for a linked budget and spending data architecture. In particular, we make the following three contributions. *First*, we provide a conceptual open data architecture, based on requirements that have been specifically collected for the open budget and spending data life cycle. *Second*, we realize the concept by extending an existing open data platform for budget and spending according to identified requirements, in particular, by adding support for linked data and integrating additional tools for data analysis and citizen participation. *Third*, we evaluate the platform in terms of usability and applicability on real-world scenarios provided by three different municipalities.

This article is structured as follows: requirements for linked open budget and spending platform are described in Section 2. Then, the software components of the OBEU platform are detailed in Section 3, and their implementation in Section 4, while the evaluation is in Section 5. Finally, Section 6 provides an overview of related work before concluding in Section 7.

2 Requirements

To collect the requirements for a linked open budget and spending data platform, we performed an analysis of several sources, namely open data life cycle [1] as well as several open data publishing guidelines (GODI [15], ODB [14], 5-star data ratings [2], Open Data Policy Guidelines [18]) and collected requirements through the OpenBudgets Community⁶.

There are several open data assessment methodologies as well as open data publishing guidelines. The W3C EGOV interest group provides recommendations and group notes⁷, which includes the Data Cube (DCV) and Data Catalog (DCAT) vocabularies and the Internationalization Tag Set (ITS). The 5-Star data rating provides several suggestions, such as making the data: available online, structured, in a non-proprietary format, provided with URIs, and linked to

⁶ www.openbudgets.eu

⁷ https://www.w3.org/standards/techs/egov#w3c_all

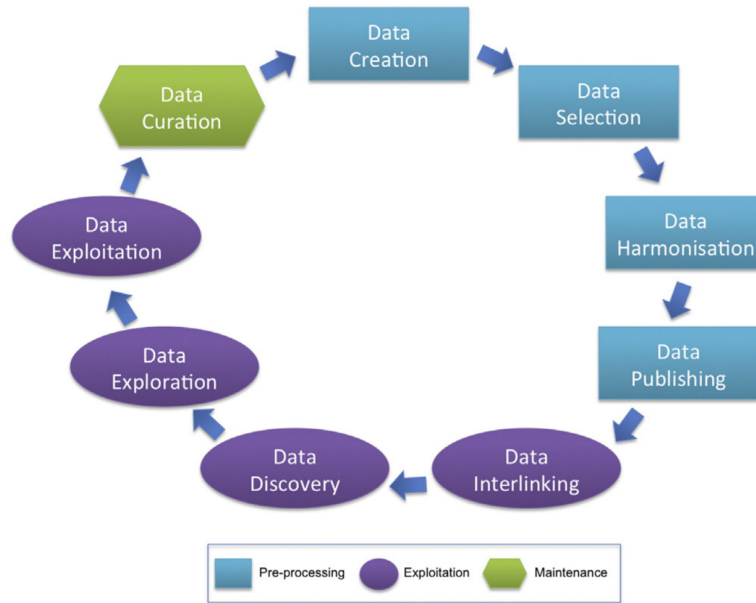


Fig. 1: Open Data Life Cycle as proposed by Attard et. al.[1].

other data [2]. In addition, the Sunlight Foundation provides several comprehensive suggestions [18].

Furthermore, a questionnaire was conducted with the OpenBudgets.eu community. The respondents belong to different interest groups e.g. governance transparency, journalism, active public participation, e-government, technical implementation as well as research. In this gathering, 66 functional, 13 non-functional requirements and 29 data quality indicators for a linked budget and spending platform are collected [4]. The summary of the gathered requirements can be found in an online document in more detail⁸.

Attard et. al. [1] propose an open data life cycle (Figure 1) which consists of three main stages: *pre-processing*, *exploitation*, and *maintenance*. The pre-processing stage consists of data creation, selection, harmonization, and publishing. The exploitation stage consists of data interlinking, discovery, exploration, and exploitation. Finally, the maintenance stage consists of data curation. Based on the open data life cycle, open data assessments and publishing guidelines, we have summarized several key requirements that should be implemented in the linked open budget and spending data platform.

Data Creation. Creating the datasets in public administrations is usually part of daily procedures. The main steps within the data creation are: providing documentation, providing provenance information, and ensuring that the datasets are authoritative.

⁸ <https://github.com/fathoni/icwe-2018-openbudgets-supplement>

Data Selection. Data selection involves the removal of existing private and personal data, as well as identification of conditions for publishing the data. Determining the list of available classifications (i.e., code lists, a list of predefined concepts that is used to group budget and items), checking for missing data, and enlisting available investment alternatives (in the context of participatory budgeting) are part of the requirements.

Data Harmonization. Making the datasets conformant with the open data publication standards is the focus of data harmonization. Steps within data harmonization include: creation of RDF data model that supports budgets, revenues, incomes, transactions, classifications, amount, payer, payee and currency; acquisition of metadata; clarification of data usage license; semantic mapping of CSV data format to RDF; mapping of OpenSpending data model to RDF; association of targeted amount to actual spending; and the linking of data items. Published datasets should ideally be provided as structured data in an *open format* using an *open license*.

Data Publishing. The main data publishing stage consists of different steps, such as data loading from CSV format or an API, providing kiosk mode on the data web page, as well as performing a customizable continuous integration, download option and links to Freedom of Information Act / Access to Documents. Ideally the published datasets should be easily and publicly accessible through an API as well as a bulk download; associated with license, contributors and contact points information. The datasets should also be openly licensed and published in a sustainable manner, i.e., hosted on a government Open Data portal, an official website, or a preservable public platform (e.g., Github).

Data Interlinking. Data interlinking connects datasets and items within the datasets to other resources. The main step for datasets interlinking is a mapping between related classifications from different datasets, for example, mapping a functional classification (e.g., health, education, public infrastructure, etc) from a public administration with another functional classification published by different public administrations, which would enable comparative analysis. Datasets should also be published as RDF and have a dereferenceable URI.

Data Discovery. The existence of open data should be discovered by data consumers. From the requirements perspective, data discovery can be enhanced by the availability of free-text search, the availability of semantic search, providing search result ranking, the availability of explorable processed datasets, availability of metadata, availability of feature to perform different levels of query, and implementation of a user-friendly user interface.

Data Exploration. In order to enable the data exploration, simplified consuming options should be provided. The related steps for this requirement are: explained flow of budgeting process, tracking of budget version, availability of localized or translated data, querying by administrative regions or institutions, availability of search feature, availability of data exploration samples, visual exploration of both RDF and non-RDF data, availability of visualization suggestions, previewing the visualization, availability of geographical visualization, exporting and sharing high quality and indisputable visualization, tracking of user data

processing work flow and cache processing data, budget comparison by using different dimensions (public administrations, time, and function), filtering (by spending or administration type), availability of top-level aggregation, and attachment of participatory budgeting result.

Data Exploitation. The next level of data cycle is exploiting the data, which is a more advanced step in consuming the data and allows users to provide analysis, mash-up or some other innovations by using, reusing or distributing the data. The requirements involved in the data exploitation stage include building custom visualization, performing exploit analysis, filtering commensurable objects, detecting outliers, extrapolating the data, aggregating the data by time interval, availability comparison between planned vs. spent money, normalizing by key metrics, differentiating between real vs. nominal value (e.g., inflation adjustments), providing contextual information, breaking down the budget and spending items, and attaching spending to participatory budgeting result.

Data Curation. Data curation is important to ensure data sustainability. Steps within data curation include pointing missing data, indexing both tabular and RDF graph data structures, as well as gathering budget votes for participatory budgeting. Datasets should ideally be published with detailed metadata and updated regularly and in a defined time interval. If possible, a version tracking for datasets should be provided.

3 Architecture

The high-level overview of the OBEU platform is provided in [Figure 2](#). A more detailed data interaction between different components is provided in [Figure 3](#), with most tools are developed during the duration of OBEU project by OBEU partners, except tools within OpenSpending platform (which has been developed earlier and is being continuously developed) and Virtuoso Triple Store. As can be seen in [Figure 2](#), there are five layers that build up the OBEU platform: data storage layer, data transformation layer, API layer, platform layer and application layer. These layers are described in the following sections.

3.1 Data Storage Layer

The Virtuoso triple store is used to host all the graphs resulting from the data transformation pipelines. Non-RDF datasets coming from the OpenSpending (OS) packager interface are hosted in Amazon S3 cloud storage. This layer partially satisfies the requirements of Data Publishing and Curation in [Section 2](#).

3.2 Data Transformation Layer

Unifying heterogeneous budget and spending datasets is a challenging task due to their heterogeneity in terms of schema/structure, syntax, and format. Representing different open budget and spending datasets adhering to a unified and integrated data representation formalism significantly eases data analysis.

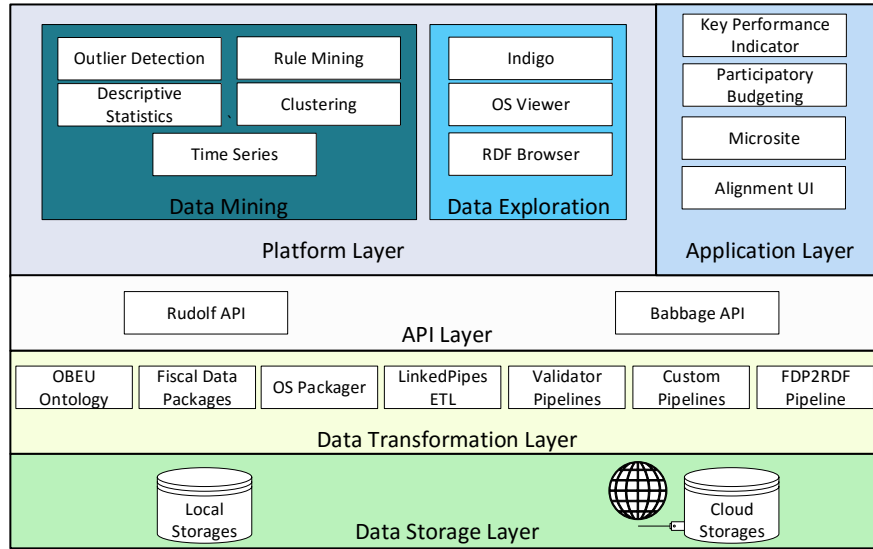


Fig. 2: Logical Overview of the OBEU platform.

Currently, there are two major data models for representing open budget and spending datasets: the OpenBudgets.eu (OBEU) data model [13] and the Fiscal Data Package (FDP) data model [20].

The *OBEU data model* and ontology represent spending and budget datasets in RDF employing the Data Cube Vocabulary (DCV) specification. The OBEU Ontology has been designed by [3], after surveying fourteen different data models for budget and spending. The OBEU data model allows multidimensional On-line Analytical Processing (OLAP), which is suitable to process multidimensional data in budget and spending data. Representation of budget and spending datasets as RDF allows linking the data with other datasets and knowledge bases e.g. on the Linked Open Data (LOD) cloud.

The ingestion of datasets can be performed through a step-by-step wizard using the *OS Packager*⁹. The usage of the OS Packager is recommended for users that do not have a strong technical background in transforming the data, i.e., those who are not familiar with linked data and SPARQL queries. With the OS Packager, the user annotates the datasets based on the schema of their data. These annotations are then saved as a JSON file. This JSON file, along with the original CSV format, make up the *Fiscal Data Package* (FDP) data model. It should be noted that not all structures of the budget and spending data are supported by FDP. In such case, a manual data transformation should be done

⁹ <https://github.com/openspending/os-packager>

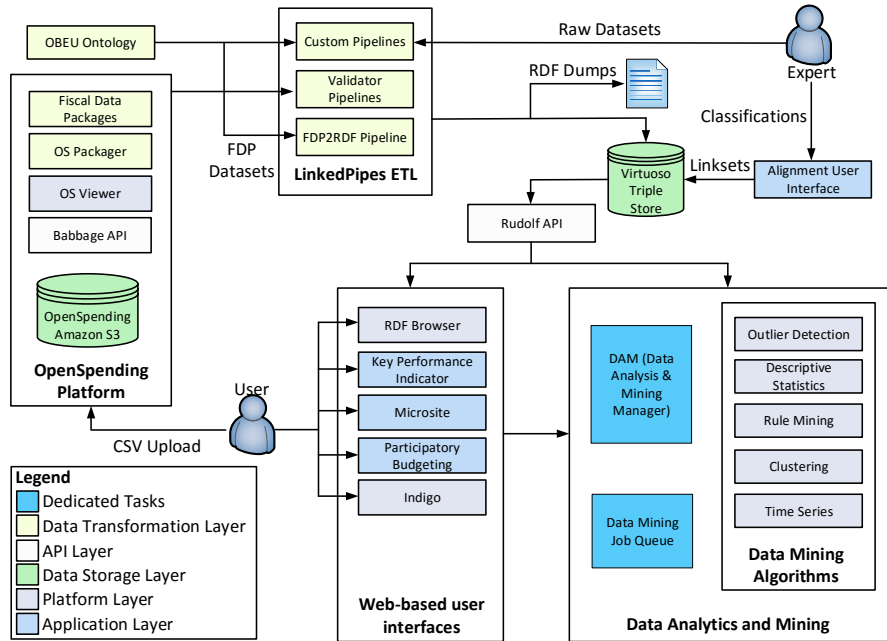


Fig. 3: The Data Flow within the OpenBudgets.eu platform.

using an Extract-Transform-Load (ETL) tool. A comparison between supported features and limitation of OBEU data model and FDP is given in [11].

LinkedPipes ETL [9] is used for the ETL process. Using *LinkedPipes* requires some understanding of RDF concepts and constructing SPARQL queries. However, the users can flexibly arrange the components in such a way that fits the structure of the datasets to build up a *custom pipeline* for their own dataset transformation, so that all necessary information in their datasets can be represented in the OBEU data model. Datasets in the FDP format can also be transformed into the OBEU data model using a reusable *FDP2RDF* [12] pipeline, provided that the datasets have been correctly structured and annotated in FDP. Since building a transformation pipeline is an error-prone process, a reusable *validation pipeline* template [12] is provided to check the constraints imposed by the DCV and OBEU data model. Transformation pipelines from *LinkedPipes* can be exported into the JSON-LD format, which can then be imported in any *LinkedPipes* instance.

On a side note, providing datasets with metadata significantly improves the accessibility of the datasets. FDP provides metadata when the user annotates and uploads the datasets using the OS Packager wizard. In the OBEU data model, users specify the metadata using several components, such as DCAT-AP distribution (for e.g., datasets access URL, format, license) and DCAT-AP

datasets (for e.g., datasets title, description, IRI and contact point). The components in this layer facilitate Data Creation, Harmonization and Interlinking.

3.3 API Layer

The *Rudolf API* [8] provides an API that fetches fiscal datasets from the OBEU RDF triple store. The Rudolf API derives data from the OS data store and serves the data for further tasks, such as data analytics and mining, as well as data visualization. This layer facilitates the Data Publishing and Discovery aspect.

3.4 Platform Layer

Indigo is the main dashboard that lets users choose available datasets to be explored (Figure 5). Users can then navigate through several other features, such as *Data Analytics and Mining* (DAM) and visualization. DAM provides a playground for scientists to experiment with budget and spending datasets. Within the DAM component several algorithms are implemented, including several types of *outlier detection*, as well as *descriptive statistics*, *rule mining*, *clustering* and *time series* algorithms.

Two types of visualization are provided within the platform: standard and customized visualizations. A standard visualization is provided directly by the OS Viewer¹⁰. Customized visualizations can be easily integrated as well (e.g. as in the case of the city of Bonn¹¹).

*RDF browser*¹² is designed to enable exploration of specific dataset entities using the particular URIs. By using the RDF Browser, users can inspect the relationship of items, in the form of URIs, within the datasets. The tools in this layer address the requirements Data Discovery, Exploration and Exploitation.

3.5 Application Layer

Alignment UI enables mapping between related concepts of classifications that are published by different public administrations. The interlinking of related concepts across different classifications enables comparative analysis across different datasets that have related concepts.

Microsite simplifies fiscal data Web site creation and embedding on public administration websites. Users are provided with configurable administrator dashboard to set which localization, data types and visualization are embedded. Web page visitors can then comment on the showcased datasets.

Participatory Budgeting component allows public administrations to announce their budget plan and then let their citizen vote on their preferred budget allocation. The application within participatory budgeting components allows citizens to be more proactive on budget allocation decision making.

Key Performance Indicator (KPI)¹³ provides an analysis of fiscal performance

¹⁰ <https://github.com/openspending/os-viewer>

¹¹ <https://github.com/shurkhovetsky/obeu-vis>

¹² <https://github.com/okgreece/RDFBrowser>

¹³ <https://github.com/okgreece/KPIs>

from a specific dataset and organization. In KPI, users are also provided with configurable administrator panel. The indicators examined within KPI includes employment cost index to expenditure, total revenue to population, expenses per citizen, among many other indicators. The tools in this Application layer partially satisfy the requirements in several open data life cycles, including Data Selection, Publishing, Interlinking, Exploitation and Curation.

4 Implementation

The Docker light-weight virtualization technology is used to integrate the components in different layers of the OBEU platform which is shown in Section 3. The components are running within different Docker containers, and the access to different components is controlled by a Nginx web server which is also running within a Docker container. The internal communication between some components also goes through Nginx.

The management of the different docker containers is done by Dockerfile and Docker Compose. Dockerfile is used to build a Docker image and run Docker container, the configurations of Docker containers are done by Docker Compose. By using this management schema, the OBEU platform can be updated easily if there are some updates in any components, and the platform is also easily portable. Further documentation regarding integration and how to instantiate a new OBEU platform is accessible online¹⁴.

5 Evaluation

The OBEU platform is evaluated using (i) three large-scale trials conducted with municipalities, (ii) a survey on UI usability, and (iii) performance measurements. In addition, an online documentation¹⁵ for the evaluation is also provided to accompany this paper, which states whether each requirement has been satisfied/partially satisfied/unsatisfied by developing and integrating related tools.

5.1 Use cases

To test the tools developed in the OBEU platform, large-scale trials were conducted in three municipalities: Bonn, Paris and Thessaloníki. Seven different testing scenarios have been developed for each trial: (1) data ingestion with OS Packager, (2) automated data transformation to RDF from OS Packager, (3) ETL pipelines for RDF semantic lifting fiscal data using LinkedPipes, (4) Visualizations, (5) Microsite, (6) Data Analytics and Mining and finally (7) Participatory Budgeting. Each municipality had to perform the same seven testing scenarios and was then asked afterward for comments and feedback. The main outcome of these testings is detailed in the project trials deliverable [16].

¹⁴ <https://github.com/openbudgets/integration>

¹⁵ <https://github.com/fathon/iawe-2018-openbudgets-supplement>

The first large-scale trial is implemented with the city of Bonn in Germany. Datasets from the city of Bonn are rich and complex, involving positive and negative values to indicate the expenditure and income, dimensions that do not uniquely define a budgeting item, as well as complex, nested classifications. The data structure available for the Bonn datasets was not supported by the current common structure in the OS Packager. Therefore, the dataset upload was performed in two scenarios: First by utilizing the custom pipeline to accommodate their data structure complexity, and second, by simplifying the initial datasets to adapt the supported structure by the OS Packager tool. Both scenarios successfully transformed the datasets to OBEU data model, with the consideration that the first scenario needs more technical expertise. Embeddable visualization for the Bonn datasets was generated and tailored using the Microsite. The outlier detection algorithms, using either Local Outlier Factor (LOF) or frequency-based algorithms, have successfully detected unusual budgeting trends in the case of Beethoven's 250th birthday celebration in the budget year 2020. The City of Bonn has found that the Participatory Budgeting tool was easy to configure and use. However, implementing such a Participatory Budgeting tool for the citizen requires a large amount of political and bureaucratic work and consequently the Participatory Budgeting tool is not used openly for now. The main feedback received from the city of Bonn was, that using the OBEU platform simplifies data ingestion (which would take a long time to comprehend), and once the datasets are properly ingested, the subsequent requirements in data life cycle are adequately fulfilled.

The municipality of Paris is the second trial participant for the OBEU platform. Paris has already provided their datasets openly in a clean CSV format, and their datasets can be directly transformed to FDP using OS Packager. A custom pipeline was not necessary for Paris, since the CSV and FDP format can be transformed to the OBEU RDF format using a reusable FDP2RDF pipeline. Visualizations have been tested and a Microsite generated. According to this trial, the Microsite revealed to be the most visible component of the platform. Suggestions for improvement include making personalized data visualization available for other users, allowing for tabbed visualizations (default, expert and showcases visualization), and allowing regulation of the Microsite by the communication/publication department. Data mining has not been used in the Paris case, due to the lack of personnel with data mining expertise. Since Paris has also implemented their own participatory budgeting tool, the OBEU Participatory Budgeting tool has not been used in practice.

The last large-scale trial is the municipality of Thessaloníki. Data Ingestion with OpenSpending could be done easily since the municipality of Thessaloníki has already published Open Data which can be exported into different formats, including CSV that is already structurally compatible with OS Packager. A minor issue was found during the testing of the OpenSpending packager, which requires year/date columns in the datasets. Custom ETL pipelines were also created as a template so that other municipalities from Greece can reconfigure the pipeline and reuse it. The developed KPI visualizations offer rich financial per-

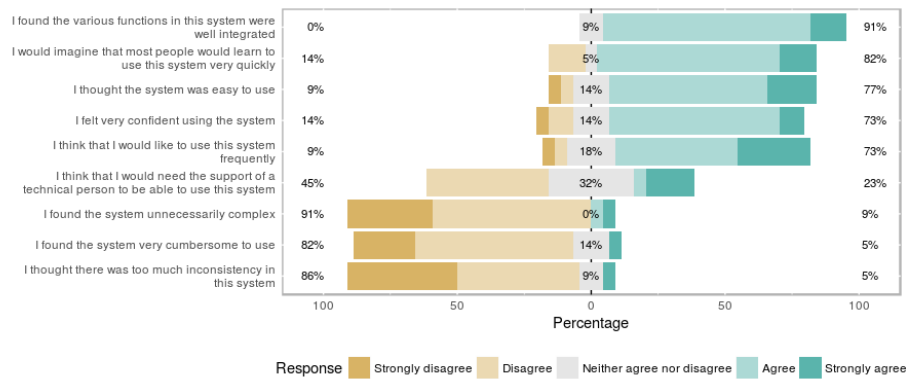


Fig. 4: An Aggregated UI evaluation result from several OBEU tools.

formance indicators for Thessaloniki¹⁶. As with other municipalities, some of the data mining tools require domain expertise. However, under expert supervision, insights and predictions over the data were found useful. An implementation of Participatory Budgeting was tested and currently planned for publication to the citizens of Thessaloniki.

5.2 Usability

Using Likert-scale-based questionnaire, we evaluated several tools deployed within OBEU, namely OS Packager, OS Viewer, Microsite, KPI admin panel, KPI, and LinkedPipes. The aggregated questionnaire result is summarized in Figure 4. The detailed, non-aggregated usability test result is available online¹⁷.

5.3 Performance

The OBEU platform is deployed on a server with CPU Intel®Xeon®CPU E5-2660 v3 @ 2.60GHz, 35 GB of RAM, 1 TB of disk and 4 GB Swap Memory. The Virtuoso triple store manages 12.6 million triples at the time of writing. There are 253 distinct datasets, 305 distinct classifications, 240 distinct data structure graphs, totaling 798 of distinct graphs.

The performance evaluation focuses on the data search and query performance. The entrance point of the platform for a normal user is Indigo, which sends a search request to the Rudolf API to load a certain size of datasets. Afterwards a user can search, as illustrated in Figure 5, e.g. using "bonn" as a keyword. We use a script to initiate API calls, and measure the runtime difference between a different number of datasets in the search request to the Rudolf API.

¹⁶ <http://kpi.okfn.gr/>

¹⁷ <https://goo.gl/Kqkbc6>

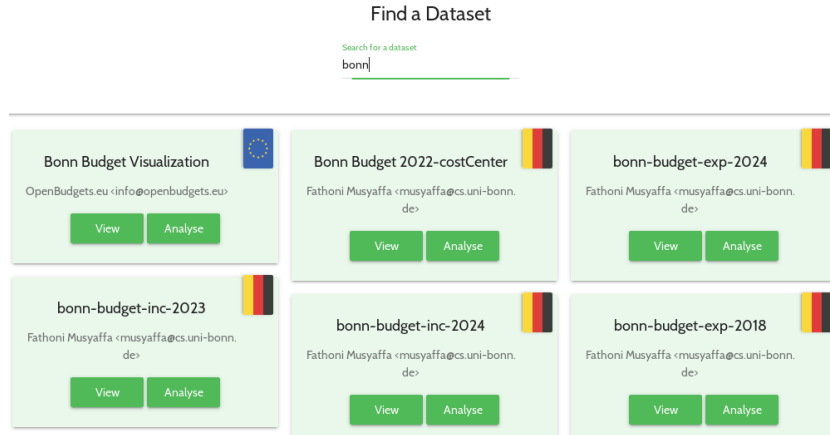


Fig. 5: Search with a keyword in Indigo.

The runtime of searching through the Rudolf API with or without using a keyword is shown in Figure 6. We initiate 30 API calls for each data items size and plot the averaged runtime for each number of datasets. It can be seen that with increasing number of the datasets, the runtime deteriorates not much.

The evaluation of query performance is done by executing SPARQL queries against the SPARQL endpoint. To compare the performance of dataset listing using the Rudolf API with a pure SPARQL query, we also measure the average time needed to fetch the list of datasets. The SPARQL query used to list the whole datasets is provided in Listing 1.1. The complete execution of this query takes 67 milliseconds on average. Compared with the Rudolf API execution time (see Figure 6), SPARQL querying is faster. Having an interface in Indigo which calls the Rudolf API is providing an easy-to-use interface for users, with some performance tradeoff. However, the expense of loading time is justified when we are considering the UI usability improvement as evaluated in Figure 4.

Listing 1.1: SPARQL query that retrieves all OBEU datasets.

```
SELECT DISTINCT ?s WHERE {?s a <http://purl.org/linked-data/cube#DataSet>}
```

6 Related Work

A conceptual architecture for open data architecture is proposed by *DIGO* [10]. *DIGO* presents a semantic open data architecture based on five layers: knowledge base layer, syntactic data layer, semantic data layer, fusion data layer, and information layer. *DIGO* elaborates a high-level overview of open data architecture in general domain of open data. In this work, we materialize the conceptual open data architecture specifically for analyzing open budget and spending data on a platform. We propose a more fine-grained platform of linked budget and

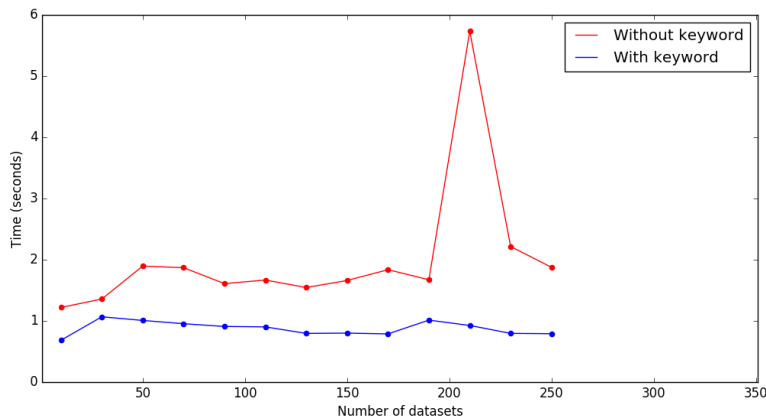


Fig. 6: Runtime of searching through Rudolf API.

spending data analytics platform (compared to DIGO’s work) by considering requirements systematically collected from different sources.

*OpenSpending*¹⁸ (OS) is a platform to analyze open budget and spending datasets. The users can upload and annotate their CSV datasets in the OS platform. The whole platform consists of mainly a data store, API, platform utilities (e.g. conductor, status and incident notifications, command line interface, authorization client, monitoring tool), data packager, data viewer, data explorer, as well as Where-Does-My-Money-Go (an app for analyzing and visualizing tax allocation per taxpayer)¹⁹. There is no linked data support in the OS platform and data representation and analytics are very limited. However, our work extends and integrates well with OS.

LinkedSpending [7] transforms the datasets available in the OS platform into the semantic format by following DCV specification²⁰. There are several components collected, developed and integrated by the LinkedSpending platform, such as ontologies, datasets transformation application, datastore, error handler, web-based datasets browser and LinkedSpending-OS data synchronization tool. After conversion, the datasets can be browsed using faceted search, visualized using CubeViz²¹ or queried using SPARQL. While LinkedSpending has provided the semantic layer and added necessary synchronizer to fetch and convert budget and spending datasets from OS, some other requirements for a semantic budgets and spending platform have not been met, such as unsupported-structure datasets import.

¹⁸ <https://openspending.org/>

¹⁹ <https://docs.openspending.org/en/latest/developers>

²⁰ <https://www.w3.org/TR/vocab-data-cube/>

²¹ <http://cubeviz.aksw.org/>

*Socrata*²² provides an Open Data Portal platform intended specifically for the government on different levels (city, country, state, and federal state organizations) which includes several services: DataSpace (data storage, indexing and retrieval), Data Publishing, Data Discovery and Visualization, and Open Data API. Data supported in Socrata platform ranges from digital content (e.g., video), operational, geospatial, financial, and performance data. *Junar*²³ is an SaaS platform to publish Open Data in general Open Data domain. Junar offers Open Data collection, enhancement (through tables, charts, and maps), publishing (including API), sharing and analysis. *OpenGov*²⁴ offers an open data solution for public administrations, consisting of cloud-based open data publishing, visualization, financial tracking, and collaborative budget builder. *WikiBudgets*²⁵ is an interactive visualization tool for open budget data. In contrast with OBEU platform, we are working specifically for public budget and spending platform with a strong emphasis on linked data. The implementation for the OBEU platform can be done independently if the public administrators have technical resources, since most of the tools utilized in the OBEU platform are documented and freely available, including the source code.

7 Conclusion

Through OpenBudgets.eu, a conceptual architecture and implementation of a budget and spending data platform are provided with the goal of simplifying the life cycle of such data. This implementation addresses the challenges and satisfies requirements related to the demanding open data life cycle. It integrates available relevant components and platforms in a micro-services architecture and extends them with extra tools providing additional linked data capabilities. The platform has been evaluated with real application scenarios, usability and performance tests. It is currently ready as an open source product that can be exploited by municipalities and easily re-deployed in various contexts.

Acknowledgments

The work presented is co-funded by the EU H2020 for the project OpenBudgets.eu (GA no. 645833), Fathoni Musyaffa is supported by a DAAD scholarship.

References

1. Judie Attard, Fabrizio Orlandi, Simon Scerri, and Sören Auer. A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4):399–418, 2015.

²² <https://socrata.com/>

²³ <http://www.junar.com/index9ed2.html?lang=en>

²⁴ <https://opengov.com/>

²⁵ <https://www.wikibudgets.org/>

2. Tim Berners-Lee. Linked data - design issues, 2010. <https://www.w3.org/DesignIssues/LinkedData.html>.
3. Marek Dudáš, Jakub Klímek, Jan Kučera, Jindřich Mynarz, Lucie Sedmihradská, Jaroslav Zbranek, and Bela Seeger. The openbudgets data model and the surrounding landscape, 2016. <http://openbudgets.eu/assets/deliverables/D7.7.pdf>.
4. Fahrettin Gökgöz, Sören Auer, and Jaana Takis. D4.2: Analysis of the required functionality of openbudgets.eu. <http://openbudgets.eu/assets/deliverables/D4.2.pdf>.
5. Auraloce Graft, Stefaan Verhulst, and Andrew Young. Open Data's Impact - Brazil's Open Budget Transparency Portal : Making Public How Public Money Is Spent. Technical report, Govlab, January 2016.
6. Noor Huijboom and Tijs Van den Broek. Open data: an international comparison of strategies. *European journal of ePractice*, 12(1):4–16, 2011.
7. Konrad Höffner, Michael Martin, and Jens Lehmann. Linkedspending: Openspending becomes linked open data. *Semantic Web*, 7(1):95–104, 2016.
8. Lazaros Ioannidis, Charalampos Bratsas, Sotiris Karabatakis, Panagiotis Filipidis, and Panagiotis Bamidis. Rudolf: An http api for exposing semantically represented fiscal olap cubes. In *SMAP*, pages 177–182. IEEE, 2016.
9. Jakub Klímek, Petr Škoda, and Martin Necaský. Linkedpipes etl: Evolved linked data preparation. In *ESWC (Satellite Events)*, volume 9989 of *LNCS*, 2016.
10. Alexandre Lopes Machado and José Maria Parente de Oliveira. Digo: An open data architecture for e-government. In *EDOCW*, pages 448–456. IEEE Computer Society, 2011.
11. Fathoni A. Musyaffa, , Fabrizio Orlandi, Hajira Jabeen, and Maria-Esther Vidal. Classifying data heterogeneity within budget and spending open data. In *Proceedings of the 11th ICEGOV*. ACM, 2018.
12. Jindřich Mynarz, Jakub Klímek, Marek Dudáš, Petr Škoda, Christiane Engels, Fathoni A Musyaffa, and Vojtech Svátek. Reusable transformations of data cube vocabulary datasets from the fiscal domain. In *Proceedings of the 4th International Workshop on Semantic Statistics co-located with 15th International Semantic Web Conference*, 2016.
13. Jindřich Mynarz, Vojtech Svátek, Sotirios Karampatakis, Jakub Klímek, and Charalampos Bratsas. Modeling fiscal data with the data cube vocabulary. In *SEMANTiCS (Posters, Demos, SuCCESS)*, volume 1695 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.
14. Open Data Barometer. Global report: Open data barometer, 2017.
15. Open Knowledge Barometer. Place overview: Global open data index, 2016.
16. Fabrizio Orlandi, Tiansi Dong, Sotirios Karampatakis, Piero Hernandez, Fathoni Musyaffa, and Hantong Liu. D7.7 large-scale trial report including best practices. <http://openbudgets.eu/assets/deliverables/D7.7.pdf>.
17. Nigel Shadbolt, Kieron O'Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, Wendy Hall, and m. c. schraefel. Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, 27(3), 2012.
18. Sunlight Foundation. Open data policy guidelines, 2014.
19. Alan Freihof Tygel, Judie Attard, Fabrizio Orlandi, Maria Luiza Machado Campos, and Sören Auer. How much? is not enough: an analysis of open budget initiatives. In *Proceedings of the 9th ICEGOV*, pages 276–286. ACM, 2016.
20. Paul Walsh, Rufus Pollock, Tryggvi Björgvinsson, Steve Bennett, Adam Kariv, and Dan Fowler. *Fiscal Data Package*. Open Knowledge International, 0.3.0 edition. <https://frictionlessdata.io/specs/fiscal-data-package/>.