OpenBudgets.eu: Fighting Corruption with Fiscal Transparency

Deliverable 4.5
Integrated OpenBudgets.eu Platform – Final

Dissemination Level	Public
Due Date of Deliverable	Month 24, 30.04.2017
Actual Submission Date	Month 28, 10.08.2017
Work Package	WP 4, OpenBudgets.eu Requirements, Platform Architecture Integration and Development
Task	T4.2
Type	Demonstrator
Approval Status	Final
Version	0.4
Number of Pages	1617
Filename	D4.5 - Integrated OpenBudgets.eu Platform – Final

Abstract:

This deliverable presents the OBEU architecture and the high-level description of its components. All software is available at the OBEU GitHub repository https://github.com/openbudgets and details about the documentation are also included in deliverable D4.4.

History

Version	Date	Reason	Revised by
0.1	23.04.2017	Initialization	Lazaros Ioannidis
0.2	26.06.2017	First version	Tiansi Dong, Lazaros Ioannidis
0.3	25.07.2017	Second version	Lazaros Ioannidis, Tiansi Dong
0.4	04.08.2017	Final version	Fabrizio Orlandi

Author List

Organisation	Name	Contact Information
OKGR	Lazaros Ioannidis	larjohn@gmail.com
OKGR	Sotirios Karampatakis	karampatakis@okfn.gr
UBonn	Tiansi Dong	tdong@uni-bonn.de
FhG	Fabrizio Orlandi	orlandi@iai.uni-bonn.de
UEP	Jakub Klímek	klimek@ksi.mff.cuni.cz
UEP	Marek Dudáš	marek.dudas@vse.cz

Executive Summary

This deliverable presents the final stable version of the integrated OBEU platform. In particular, its architecture is described as well as technical details about its components. Thanks to the modular and distributed architecture of the platform, the deployment of the OBEU platform is currently located at four different places: Fraunhofer IAIS, UEP, OKFGR, and OKI. In total, ten components are documented: (1) Indigo, (2) OBEU Viewer, (3) Data-mining tools, (4) LinkedPipes ETL, (5) Microsite, (6) Participatory Budgeting, (7) Virtuoso, (8) Rudolf, (9) RDF Browser, (10) RDF alignment. As this deliverable is of type "Demonstrator" only a high-level description of its components is provided. More details about the tools are provided in the respective GitHub pages and in deliverable D4.4.

These tools cover the whole operations with budget data: from dataset uploading, to data transformation and visualization; from dataset searching to data-mining and user commenting and participating in the budget allocation process.

All software is available at https://github.com/openbudgets. Municipalities can easily install the whole package following the instructions on their server, or embed components/visualizations into their existing web-site (e.g. by using the OBEU Microsite https://github.com/openbudgets/microsite) .

Abbreviations and Acronyms

LOD	Linked Open Data
LP-ETL	LinkedPipes ETL
OS	OpenSpending
OBEU	OpenBudgets.eu
FhG	Fraunhofer IAIS
OKI	Open Knowledge International
OKGR	Open Knowledge Greece
UEP	University of Economics, Prague

Table of Contents

List of Figures

1 Introduction

In this deliverable, we summarize the final platform of the OBEU project, which incorporates components of the OpenSpending project[1]. OBEU platform extends OpenSpending in the following three aspects:

- Data representation
    - High qualified RDF representation of Budget data
    - Associated software tools (both expert-level and dummy-level) to transform Fiscal Data Package (FDP) into the RDF data model
    - User friendly graphic representation of Budget data
- Data Analysis and Mining
    - Promising data analysis and mining tools
    - An easy extendable distributed computing architecture
    - User friendly graphic interface for data analysis and mining
- An independent SAAS module (Microsite) which can be easily plugged into the web-page of municipality to let citizens and stakeholders to view and comment on budget data.
- An independent Participatory Budgeting service, which facilitates citizens to express their opinion in the budget allocation phase.

The architecture of the OBEU platform is illustrated in Figure 1.

---

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

Figure 1. Architecture of the OBEU platform
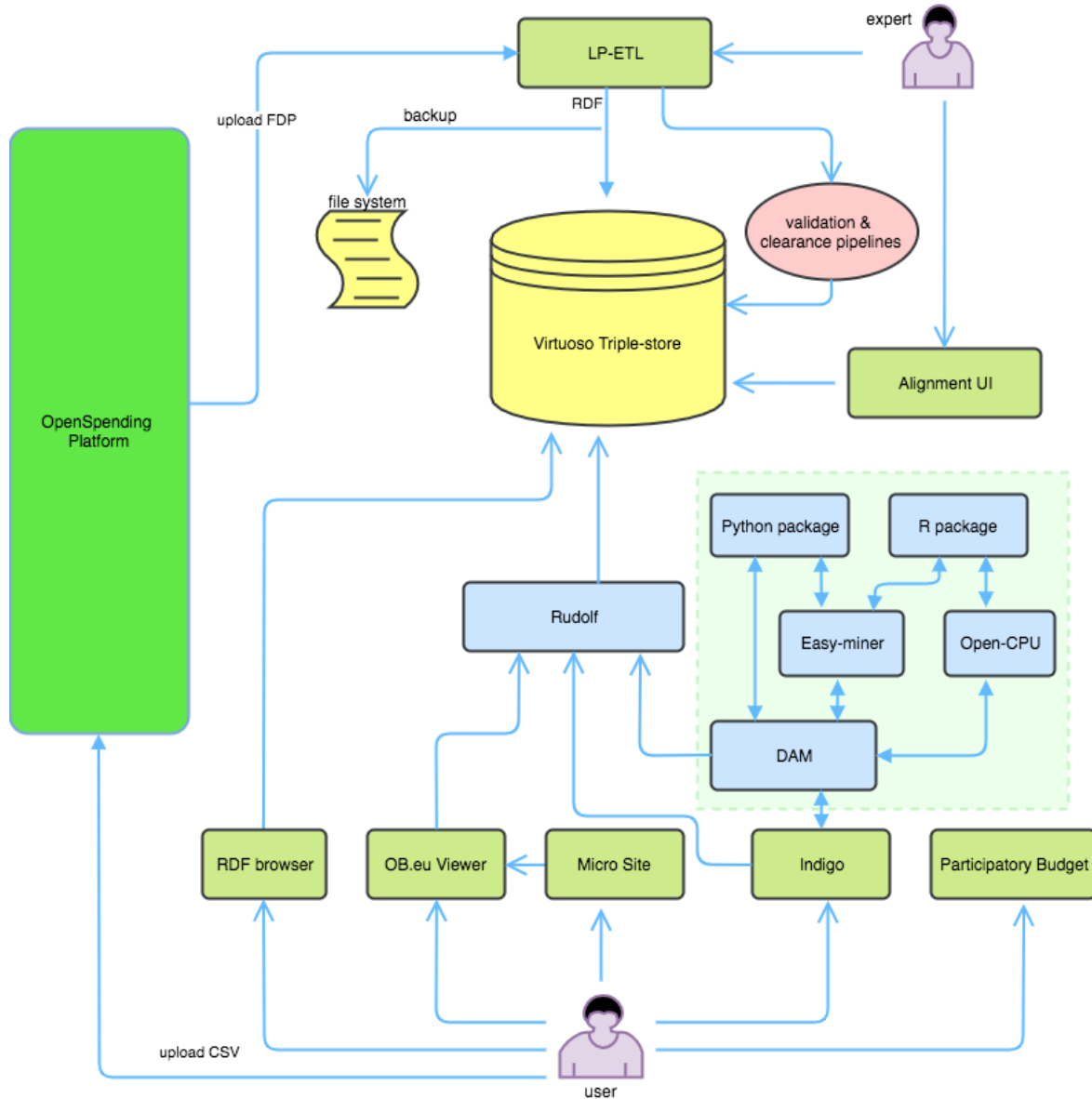
As illustrated in Figure 1, regular users can access: (1) OpenSpending platform, (2) RDF bowser, (3) OBEU viewer, (4) Microsite - an easy SAAS tool for viewing and commenting budget data - (5) Indigo, for search datasets, and analysing data, (6) Participatory Budgeting interface, for participating in the budget-related decision processes. Expert users can use additional tools, such as (7) LP-ETL for transforming budget data into RDF format, (8) RDF alignment tool.

2 Components

2.1. Indigo for exploring & visualizing datasets

2.1.1. Introduction

Indigo http://apps.openbudgets.eu/indigo/#indigo/ is a user-interface for exploring and analysing datasets. It is an open-source software https://github.com/okgreece/indigo developed by OKFGR as part of the project.

Indigo can apply different data processing functions to fiscal data sets to gain deeper insights. Indigo is designed to work with OpenSpending out of the box, so that you can use it to analyze datasets in the public Open Spending repository or your own datasets, uploaded through the Open Spending packager. More algorithms will be developed and included into Indigo along with the ability to create and reuse custom fiscal indicators by combining aggregated data from various sources.

2.1.2. Main function

Indigo has three main functions:

    (1) explore and search budget datasets
    (2) provide graphical user-interface for data mining
    (3) visualizing data-mining results

2.1.3. Relation to other components

Indigo retrieves datasets from Rudolf API, it sends data-mining requests to (and receives data-mining results from) the Data Analysis and Mining (DAM) module.

2.2. OBEU Viewer

2.2.1. Introduction

OBEU Viewer[2] is the main front-end for visualization of the OpenBudgets.eu datasets, which is a customisation of the OpenSpending Viewer[3]. It is an open-source software to visualize budget datasets, developed by OKI. A tutorial of Openspending Viewer is available at http://docs.openspending.org/en/latest/developers/viewer/ . The main customisations of the OBEU Viewer over the OpenSpending version are on the capability to work with OBEU RDF data and different styling/layout.

---

[2] http://eis-openbudgets.iais.fraunhofer.de/viewer
[3] https://github.com/openspending/os-viewer

2.2.2.   Main Functions

OBEU Viewer has four main functions:

(1) preview data package resources.

(2) read metadata for a data package.

(3) provide a selection of views to visualise the budget data.

(4) perform other core actions for single data packages, such as downloading the raw data sources.

2.2.3.   Relation to other components

OBEU Viewer retrieves dataset from both OpenSpending datasets, and the OBEU RDF triplestore.

2.3.   Data-mining tools

2.3.1.   Introduction

With the new OBEU platform workflow, users send data analysis and mining request to the Data Analysis and Mining (DAM) component through Indigo. DAM is the central control module for data mining. It retrieves concrete datasets from Rudolf, perform data-preprocessing, and send concrete data-mining task to the right data-mining server. DAM is also responsible for accepting the data-mining results from other data-mining server, and sending the results to Indigo for visualization.

Data analysis and mining is realised by the following tools:

(1) Data-mining pre-processing tool
https://github.com/openbudgets/preprocessing_dm,

(2) central interface for data mining https://github.com/openbudgets/DAM,

(3) specific data mining packages https://github.com/openbudgets/outlier_dm,

(4) data mining server at UEP https://github.com/kizi/easyminer,

(5) data mining server at OKFGR http://okfnrg.math.auth.gr/ocpu/test/,

(6) interfaces with UEP and OKFGR servers:
https://github.com/openbudgets/uep_dm and

https://github.com/openbudgets/okfgr_dm

More details about the data mining tools are presented in D2.4 and D2.5.

2.3.2.   Main function

The data analysis and mining provides several functions:

(1) Descriptive statistics
(2) Time series analysis and prediction
(3) Clustering
(4) Rule mining
(5) Outlier detection based frequency or local density

The details are presented in D2.4 and D2.5

2.3.3.   Relation to other components

Data-mining package is connected with Indigo (front side) to select and visualise data, and Rudolf API to query and fetch OBEU RDF datasets.

2.4.   ETL Pipelines for generating RDF datasets

2.4.1.   Introduction

LinkedPipes ETL (LP-ETL) http://etl.linkedpipes.com/ is a generic Linked Data based ETL tool for data transformation and publishing. It manages pipelines, which represent individual data processing tasks. Each pipeline consists of interconnected components, where each component represents a configurable data transformation, and the interconnection represents the data flow among the components. A tutorial is available at https://github.com/openbudgets/pipeline-fragments/blob/master/CSV2RDF/documentation/CSV2OBEURDFPipelineDocumentation.pdf

2.4.2.   Main functions

Within the scope of the OpenBudgets.eu platform, it has two main functions.

Firstly, there is a single FDP2RDF pipeline, transforming fiscal data in the Fiscal Data Package format used by OpenSpending to its RDF representation using the OpenBudgets.eu data model, which is based on the RDF Data Cube Vocabulary.

Secondly, when the data is not available as a Fiscal Data Package, LP-ETL allows users to define their custom pipelines, which transform their data to the OpenBudgets.eu data model. The data produced by LP-ETL in both ways is then stored in the OpenBudgets.eu triplestore, in which the other tools can access it.

2.4.3.   Relation to other components

LinkedPipes ETL uploads datasets into the OBEU Virtuoso RDF/triple store.

2.5.   Microsite

2.5.1.   Introduction

The Microsite is developed to provide citizens with a way to visualize, interact, and discuss about Budget and Spending of their towns, cities and countries. Moreover, it provides a simple way for municipalities to offer these services embedded into their own websites without the need to host OBEU data or OBEU software. A demo is installed at Fraunhofer IAIS http://microsite-obeu.iais.fraunhofer.de/vizmanager/1/

2.5.2.   Main functions

The main functions of Microsite is to let users select and view datasets, and write their feedback, comments on budget data. Microsite is an open-source software. Its source and its usage is located at https://github.com/openbudgets/microsite

2.5.3.   Relation to other OBEU components

The Microsite reuses the OBEU Viewer to visualize datasets.

2.6.   Participatory Budgeting Tool

2.6.1.   Introduction

The Participatory tool is developed to enable citizens to influence budget allocation of a city. With a voting process provided by this tool, citizens can vote for their preferred proposals. Each participant is assigned a number which represents the public budgets that she/he can allocate to the preferred project. Educational resources for citizens gathered in D.7.4, will provide online materials to understand i.e. how to use the tool, the budget cycle, terms used or how to influence and monitor the budget. This is an open source software located at https://github.com/openbudgets/participatory-budgeting A demo is available at http://openbudgets-monitoring-sample.herokuapp.com/

2.6.2.   Main function

The main function of this tool is to let users to

(1) express their budget allocation priorities during the budget approval process and along the lines and within the process defined by each administration concerned;

(2) monitor budget transactions;

(3) audit budget compromised;

(4) give feedback to the administrations.

2.6.3. Relation to other components

This is an independent tool developed by CIVIO.

2.7. Triplestore Virtuoso

RDF triples are stored in Virtuoso https://virtuoso.openlinksw.com/rdf/, which is an existing open source software. Its main function is to store and query knowledge in the RDF format. A SPARQL endpoint is provided for expert-users to query/maintain the triple-store.

2.8. RDF Browser

2.8.1. Introduction

RDF triples have unfortunately long forms, which are not friendly for people to read. RDFBrowser is a customizable, easy to deploy, crossdomain and feature rich Linked Data publishing platform, developed by OBEU partner OKGR. RDFBrowser is a PHP web application, able to be deployed in most environments out of the box with minimum effort, lowering the barrier for publishing Linked Data on the Web. The feature rich HTML generator offers multilingual support, language detection, picture galleries, geo-information features and content filtering. All aspects of the application can be configured through an administrator panel on the fly, reducing administrative effort. https://github.com/okgreece/RDFBrowser. The HTML generator can be easily themed using a modern template engine. The backend of RDF Browser is triplestore Virtuoso. More details are presented at D4.4.

2.8.2. Main Function

Since OBEU platform exploits the Linked Data Principles for the purposes of structured data description and publishing, RDFBrowser acts as the resource Content Negotiator and Description Generator for the Platform of OBEU. It is also an entry point for users

to explore datasets or other related resources stored on the triplestore, in a follow-your-nose manner.

2.8.3. Relation to other components

RDFBrowser interacts with the triplestore to retrieve information about requested resources. It also interacts with other Linked Data enabled applications such as Alignment tool to provide description about resources upon request.

2.9. RDF alignment tool

2.9.1. Introduction

Ontology matching is a task to find out which ontologies refer to the same concept or existence. This is by no mean trivial, and ontology matching also exists in the OBEU project, for example, whether two IRIs refer to the same economical classification?

An open-source RDF alignment tool is developed by OBEU OKFGR partner to do collaborative, system aided, user driven ontology matching application. This alignment tool offers a simple GUI environment for matching two ontologies / vocabularies with aid of configurable similarity algorithms. We undertake research for the evaluation and validation of the default settings, taking into account expert users feedback.

Multiple users can work on the same project simultaneously. The application offers social features, as users can vote, providing feedback, on the produced linksets. The linksets are available through a SPARQL endpoint and an API.

Alignment is the outcome of the experience working with heterogeneous public budget data, and has been used to align SKOS Vocabularies describing budget data across diverse level of administrations of the EU and its member states.

More detail is presented in D4.4 and on the application repo http://github/okgreece/RDFBrowser.

2.9.2. Main Function

Alignment is used to create, validate and maintain high quality linksets between different ontologies or resource Collections as Concept Schemes. These kind of collections was used in the scope of the project as dimension values.

2.9.3. Relation to other components

Alignment interacts with the triplestore to store and retrieve mappings between different resources.

2.10. Rudolf

2.10.1. Introduction

Rudolf is an open source software developed by OKFGR partners https://github.com/openbudgets/rudolf. It is an HTTP API that wraps a SPARQL endpoint to expose OpenBudgets.eu RDF-schema - compatible datasets, according to OpenSpending's babbage API specification.

Rudolf is the HTTP data backend for many applications in OpenBudgets.eu. It needs not to be directly used by end users, rather it is contacted by other applications or developers. The main role of Rudolf is to provide an easy-to-use abstraction over the OpenBudgets.eu triple store. As getting data from a triple store requires the execution of SPARQL queries, this generates the requirement to build SPARQL strings in each and every application that wants to use data from the triple store.

This practical difficulty can be an obstacle for making the data available and used which is against the main scope of the OpenBudgets.eu project. For this purpose, the Rudolf API builds the SPARQL query itself, given the actions and parameters that come as HTTP query parameters. The parameter scheme exposed by Rudolf closely replicates the successful OpenSpending Babbage API. The out-of-the-box support for applications like the OpenSpending Viewer comes as a bonus. Another advantage is that applications that depend on the Rudolf API (like Indigo) can be reused with any OpenSpending deployment, requiring little or zero modification.

2.10.2. Main function

The main function of this tool is to let users to query Virtuoso triple-store, according to OpenSpending's babbage API specification.

Rudolf also caches the model of the datasets to save execution time. In a preloaded triple store it is recommended that one runs the cache building commands, as outlined in the README file of the source code.

2.10.3. Relation to other components

This tool communicates with Virtuoso RDF triple-store.

3. Conclusions

In this document we introduce all the ten different components of the OBEU platform at a high level. All these components, as well as the OBEU platform are open-source software, and located at Github for public access https://github.com/openbudgets. More details documenting the usage of the tools are described in D4.4. Future software updates will be pushed to the OBEU Github repository above.