Open Data Life Cycle (Attard et al, 2015)				(Gökgöz et al, 2015)	(Musyaffa et al, 2018)
No/Level	**Title**	**Description**	**Significance to Open Fiscal Data Domain (Not at all important / slighlty important / important / highly important / very important)**	**Related D4.2 Requirements**	**Related OFDP Quality Factors**
1	Data Creation	Creating the datasets in public administrations is usually part of daily procedures. Among the requirements within data creation are documentation, provenance, and authoritative.	Very Important/Critical since it is related to the existence of the datasets as well as ensuring a valid source and understandability for the created fiscal datasets.	Proper documentation, Provenance	Authoritative, Data Existence, Documentation
2	Data Selection	Data selection involves the removal of existing private and personal data, as well as identification of conditions for publishing the data. Curating the list of available classifications (i.e.,code lists), checking for missing data, and enlisting available investmets alternatives are part of the requirements.	Highly Important, the availability of privacy concerns hinders the analysis of the data and incomplete code lists prevents the datasets from easily described.	Curation for Code lists, Red Flag upon missing data, List of available investment alternatives	Complete Code List
3	Data Harmonization	Making the datasets conform with open data publication standards is the focus of data harmonization. Several requirements within data harmonization includes: 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; mapping of OpenSpending data model to RDF; association of targeted amount to spending; and the linking of data items. Published datasets should ideally provided as stuctured data in an open format using open license.	Highly Important, properly modeled and well integrated fiscal datasets allow more straightforwarrd analysis that attracts open data / civic / academic / research enthusiasts.	RDF data structure for budgets, RDF data structure for transactions, Mapping OpenSpending Data Package to RDF, Modelling of code lists in RDF, Data structure for modelling revenues/incomes, Ability to model payer, payee, amount, date, currency; Ability to attach concrete targets to spending, Link ability, Semantic Mapping of CSV, Acquisition of metadata, Clear licensing information	Open Format, Open License, Structured Data

4	Data Publishing	The main data publishing stage consists of diferent requirements, such as data loading from CSV format or an API; providing kiosk mode on the data web page, as well as fully customizable continous integration, download button and links to Freedom of information act /Access to Documents. Ideally the published datasets should be easily and publicly accessible online through API as well as bulk download, with license, contributors, contact point and English information provided. The datasets should also ideally be free of charge and published in a sustainable manner.	Very Important, there is no open data without data publishing stage The way data are published determines data consumers engaged in further open data life cycle stage.	Loading of CSV, Loading from an API, Kiosk mode, Fully-customizable CI, Download button all the way, Links to FoI/ATD Tools	API Availability, Available Online, Contact Point, Easily Available, English Info Available, Free of Charge, In Bulk, In Digital Form, Mentioned Contributors, Mentioned License, Public, Sustainable Publication
5	Data Interlinking	Data interlinking connects datasets and items within the datasets to other resouces, which makes the datasets have richer contexts. One of the requirements for datasets interlinking is that there is a mapping between related classifications from different datasets. Datasets should also be published in RDF and hence have a dereferenceable URI.	Slightly important to important. In theory interlinked fiscal data would allow fiscal data enrichment which would provide extra context for comparative analysis acoss heterogeneous fiscal datasets, however most datasets are not interlinked due to technical barriers as well as no such comparative analysis 'kiiller apps'.	Code lists' mappings support	Dereferenceable LD URI , RDF Availability
6	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, processed datasets that can be explored, availability of metadata, feature to perform different levels of query, implementation of a user-friendly user interface.	Highly Important. The ease of discovering fiscal data should help fiscal data enthusiasts to collect all the necessary fiscal data effectively and efficiently.	Free-text search, Semantic search, Relevance ranking, Exploration of processed datasets, Metadata, Different levels of difficulty in queries, User-Friendliness of the UI, Export and share.	Dataset Filtering, Search Mechanism

7	Data Exploration	In the data exploration stage, simple ways to consume the data are performed. The related requirements for this stages are: exporting and sharing high quality and indisputable visualization, previewing the visualization, availability of geographical visualization, explained flow of budgeting process, visual exploration of both RDF and non-RDF data, search result relevance ranking, availability of data exploration samples, availability of visualization suggestions, tracking of user data processing workflow and cache processing data, tracking of budget version, budget comparison with using different dimensions (public administrations, time, and function), filtering (by spending or administration type), availability of top-level aggregation, availability of localized or translated data, querying by administrative regions or institutions, and attach participatory budgeting result.	Highly Important. Data exploration allows the datasets to be explored in a generic way for common use cases. This stage results an easy to understand fiscal data analytics which should be suitable for the dissemination of fiscal data analytics for most people.	Good quality visualizations, Indisputable visualizations, Provide geo visualizations, Preview, Budget process model, Visual exploration (RDF), Non Semantic exploration,Provide samples, Suggest first, Do not repeat, Version tracking of budgets, Entity comparison, Temporal comparison, Functional comparison, Filter by spend, Filter by administration type, Get top-level aggregates, Localized data, Query by institution, administrative regions	Visualization
8	Data Exploitation	The next level of data cycle is exploting the data, which is a more advanced step in consuming the data and allows users to provide analsis, mashup or some other innovations by using, reusing or distributing the data. The requirements involved in the data explotiation stage include building custom visualization, performing exploit analysis, filtering commensurable objects, detecting outliers, extrapolating the data, aggregating the data (by time interval, temporal trend of planned vs actual spent money, normalizing by key metrics, differentiatiing 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.	Important. The advanced level of data exploitation would typically provide unique actionable insights that can be taken for interested parties, however, the trade off between technical barrier and the possible collected insgiht may not seem to be paid off.	Build custom visualizations, Exploit analysis, Filtering commensurable objects, Outlier detection, Extrapolations on data, Aggregation by time interval, Temporal trend of the difference between planned and actual spending, Normalize by key metrics, Real vs. Nominal, Contextual information, Break Down functionality, Displaying results, Attach Targets to spending	

9	Data Curation	Data curation is important to ensure data sustainability. The requirements within data curation include pointing missing data, indexing both tabular and RDF graph data structure, and gatheing budget votes in terms of participatory budgeting. Datasets should ideally published with metadata, updated requlary, and timely. A version tracking for datasets would also be desirable.	Very Important. This stage determines whether the fiscal datasets will always be available, authoritative, and sustainable. This stage will make the fiscal data enthusiasts to be more confidence that their efforts in analyzing some particular data will always be reusable for furter fiscal data publication.	Point missing data, Indexing data w.r.t. tabular vs. graph structures, Gathering votes	Metadata, Regular Update, Up to date, Version Tracking

Note:

1. Categorization may not always binary.
2. One requirement / quality factors may belong to several levels. In this case, the requirement is mapped to the level with strongest association sense.

OpenBudgets.eu Platform Functional Requirements (Gökgöz et al, 2015)				Tools, Satisfiability, ODLC Level		
No	**Features**	**Functionality**	**Description**	**Utilized/Integrated Tools**	**Satisfied (Yes / No / Partial)?**	**ODLC Level**
F001	**Data Model**	RDF data structure for budgets	Budgeting information representation by utilizing Data Cube Vocabulary DCV.	OBEU Ontology	Yes	3
F002		RDF data structure for transactions	Transaction information representation by using DCV.	OBEU Ontology	Yes	3
F003		Mapping OpenSpending Data Package to RDF	Providing mapping from OpenSpending Fiscal Data Model to OpenBudgets Data Model	FDP2RDF Pipeline	Yes	3
F004		Curation for Code lists	System for managing code lists.	GitHub (Not Integrated) + RDF Browser + Virtuoso (and SPARQL Endpoint)	Yes	2
F005		Modelling of code lists in RDF	Utilizing/extending available RDF vocabulary to model the code lists.	OBEU Ontology	Yes	3
F006		Data structure for modelling revenues/incomes	Providing data structure definition for modeling revenue/income data.	OBEU Ontology	Yes	3
F007		Budget process model	Modeling the flow of budget from the very beginning (planned budget) until the very end (transaction and achieved result).	OBEU Ontology (through it's budget-phase property)	Yes	7
F008		Code lists' mappings support	Provide localization and external mappings.	UI Alignment	Yes	5
F009		Ability to model payer, payee, amount, date, currency	Provide support for modelling transactions and its properties.	OBEU Ontology	Yes	3
F010		Ability to attach concrete targets to spending	Provide a way for community to associate a specific budget with concrete targets, and the community could evaluate the resulting target in the end of the budgeted project.	OBEU Ontology	Yes	3
F011		Link ability	Every data items must have a link (URI).	OBEU Ontology. Also depending on the datasets, whether the datasets provide both approved/drafted and executed budget phase.	Yes	3
F012	**Data acquisition, loading semantic lifting**	Loading of CSV	Provide support for loading CSV data, as well as selecting specific column in the CSV data.	OS Packager, LInkedPipes ETL	Yes	4
F013		Semantic Mapping of CSV	Provide support for mapping CSV columns to corresponding RDF properties.	LinkedPipes ETL	Yes	3

F014		Acquisition of metadata	Provide support for capturing metadata (public administrations, year, data uploader) for each loaded data package.	OS Packager, LinkedPipes ETL	Yes	3
F015		Loading from an API	Provide support for acquiring datasets from API, if the datasets are not available as bulk download.	NA	No	4
F016	**Exploration, search**	Visual exploration (RDF)	Provide support for graph browsing, to find relationship between data items.	RDF Browser	Yes	7
F017		Non Semantic exploration	Provide support for datasets faceted browsing or tabulated view.	Indigo, OS Viewer	Yes	7
F018		Free-text search	Search trough a keyword for: datasets, attributes, field names.	Indigo, Virtuoso (through SPARQL endpoints - only for experts)	Partial	6
F019		Semantic search	Provide SPARQL endpoints for advanced users, provide pre-stored script for common queries, and user friendly interface for users without SPARQL expertise.	Indigo, Virtuoso (through SPARQL endpoints - only for experts)	Yes	6
F020		Exploration of processed datasets	Provide an aggregate API for searching analysis results previously done.	OS Viewer and Babbage/Rudolf API by caching in the backend.	Yes	6
F021		Metadata	Provide a way to search additional metadata within the datasets.	Virtuoso (through SPARQL Endpoint), Indigo - Data Search	Yes	6
F022		Different levels of difficulty in queries	Provide different levels of query difficulties, for example: 1) lowest level e.g., drop down menu; basic questions; off-the-shelf visuals. 2) intermdiary level e.g., access / filter data, pivot tables. 3) SPARQL queries.	OS Viewer through visualization browsing, Virtuoso through SPARQL queries.	Yes	6
F023		User-Friendliness of the UI	Provide user friendly interface.	Indigo	Yes	6
F024		Relevance ranking	Provide search relevance ranking.	No	No	6
F025	**Visualization**	Build custom visualizations	Provide visualization customization, by column / relation selection in one or more datasets.	OS Viewer through visualization browsing, KPI	Yes	8
F026		Exploit analysis	Provide an interface for using features offered in the Analysis Tools, allowing users to flexibly combine data and hide complex queries.	OS Viewer, Indigo, KPI, and Babbage / Rudolf API	Yes	8
F027		Provide samples	Give beginners some templates to start playing, without requiring them to read any documentation. Give helpful hints and maybe a tutorial., e.g., ideal sample would ideally be real, successful news stories.	OS Viewer	Yes	7
F028		Suggest first	Provide suggestions regarding visualizable datasets, instead of overwhelmingly provide all visualzations.	OS Viewer, Indigo, KPI	Yes	7

F029		Do not repeat	Provide tracking of users' movement towards user interface and allow the users to repeat the interaction with updated datea.	NA	No	7
F030		Export and share	Let the information shareable in other platform, as well as printing.	OS Viewer, iframe export is supported but not for print.	Partial	7
F031		Good quality visualizations	Provide good quality visualizations for printing.	OS Viewer provide an intuitive visualization, but it is not for print	Partial	7
F032		Indisputable visualizations	Provide a way to verify the source of the data that is used for the visualization.	OS Viewer (Uploader name and partial email is provided, but without uploader's email domain)	Partial	7
F033		Provide geo visualizations	Provide maps that shows datasets geographical availability - if the datasets have geographical information.	NA	No	7
F034		Point missing data	Based on the patterns of previuosly available datasets, provide a way to point if a dataset is missing.	NA	No	9
F035		Preview	Provide a way to preview the datasets.	NA	No	6
F036	**Analytics**	Filtering commensurable objects	Provide agregate analytics that can be used for commensurable objects (objects with the comparable size in any terms).	KPI	Partial	8
F037		Version tracking of budgets	Provide budget evolution analysis for tracking budgets over its preparation phase.	NA/Available specifically for Bonn (proof of concept) but not for other public administrators as it is very data-structure specific.	No	7
F038		Indexing data w.r.t. tabular vs. graph structures	Optimize indexing for both tabular and graph data structures.	Data Analytics and Mining component	Yes	9
F039		Outlier detection	Provide a way to find disproportionately used categories based on outlier detection algorithms.	Data Analytics and Mining component	Yes	8
F040		Extrapolations on data	Outline trends for predicting / recommending future budget allocations.	Time Series	Yes	8
F041		Aggregation by time interval	Provide aggregation feature for e.g., sum or average of budget / spending amounts over a defined period of time (e.g., quarter, year).	NA, very dependent on datasets	No	8
F042		Temporal trend of the difference between planned and actual spending	Analyze the differences between planned and actual expenditure during the course of a time, then get insight with regards to the differences.	NA	No	8
F043	**API**	Entity comparison	Provide a way to query datasets from multiple entities.	NA	No	7

F044		Temporal comparison	Provide comparison over time, e.g., budgeted / spent over time.	NA/Partial, available specifically for Bonn datasets (as a proof of concept) but not for other public administrators as it is very data-structure specific.	No	7
F045		Functional comparison	Provide a comparison over functional classification (e.g.: education), also along different entities and time.	Virtuoso, OS Viewer (manual navigation through classification), Babbage/Rudolf API	Partial	7
F046		Filter by spending amount	Provide a way to filter spending amount by queries on a desired public entity.	Virtuoso, Indigo	Yes	7
F047		Filter by administration type	Filter by administrative classifications (managing offices).	Virtuoso	Yes	7
F048		Get top-level aggregates	Provide a way to query for total allocated/spent amount across all (e.g.: countries) in (e.g.: years 2010 - 2018).	OS Viewer (partial, since multiyears aggregation is not supported)	Partial	7
F049		Normalize by key metrics	Provide normalization by population (also breakdown population by gender, age, etc.)	KPI	Yes	8
F050		Real vs. Nominal	Provide necessary adjustments, e.g., inflation adjustments.	LinkedPipes ETL Pipeline Fragment	Yes	8
F051		Localized data	Provide a feature for data localization and translation (e.g.: entities titles, budget lines).	OBEU Ontology, LinkedPipes ETL (also data dependent)	Yes	7
F052	**SAAS Interface**	Kiosk mode	Provide activity report and software management.	Microsite	Yes	4
F053		Fully-customizable CI	Provide continuous, customizable integration for ready-to-deploy working copy.	Microsite	Yes	4
F054	**Use Journalism**	Download button all the way	Provide a way for journalist to download and store the data at every step of the analysis.	OS Packager, Virtuoso	Yes	4
F055		Contextual information	Provide additional contextual information (e.g., the budget-holder responsible for, on and-off budget items, data with regards to population, relation with Eurostat data).	NA	No	8
F056		Proper documentation	Provide information on methodology, data sources, and how the mapping has been done. The information should be done on a dataset level.	OS Packager, LinkedPipes ETL	Yes	1
F057		Provenance	Provide provenance information.	OS Packager, LinkedPipes ETL, OS Viewer	Yes	1
F058		Red Flag	Provide 'red flag' feature to report missing data.	NA	No	2

F059	**Use Transparency**	Links to FoI/ATD Tools	Provide link to ask the AsktheEU.org/fragdenstaat/Freedom of information act/Access to Documents	NA	No	4
F060		Break Down functionality	provide a way to break down the budget items into major categories, institutions, etc.	OS Viewer	Yes	8
F061		Clear licensing information	Provide information licensing information to encourage reuse of visualizations or data.	LinkedPipes ETL	Yes	3
F062		Query by institution, administrative regions	Provide a way for filtering per dataset as well as aggregates of all data that refers to the institution.	OS Viewer, depending on the availability of administrative classification on the datasets	Yes	7
F063	**Use Participation**	List of available investment alternatives	Provide a way for municipalities to create and update a list of potential investment options for users to pick.	Participatory Budgeting Tool	Yes	2
F064		Gathering votes	Store votes from citizens to the proposed investment alternatives and potentially provide a way to comments/feedback on these items.	Participatory Budgeting Tool	Yes	9
F065		Displaying results	Count votes according to the agreed process and then display the vote to other users and municipality.	Participatory Budgeting Tool	Yes	7
F066		Attach Targets to spending	Provide a way to show the spending result, e.g., hospitals built or disease index reduced. Provide a way to attach spending to concrete results so that the spending can be later scrutinized.	NA	No	8