



SC118DI07171

D02.01.3 Specification of the EU Budget Vocabulary

Document Metadata

Property	Value
Date	27/10/2015
Status	For review
Version	1.02
Authors	Nikolaos Loutas – PwC EU Services Brecht Wyns – PwC EU Services Stefanos Kotoglou – PwC EU Services Makx Dekkers – AMI Consult
Reviewed by	Pieter Breyne – PwC EU Services
Approved by	Athanasios Karalopoulos – European Commission, ISA Programme Vassilios Peristeras - European Commission, ISA Programme

This study was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Contents

1.	INTRODUCTION	1	
1.1.	OBJECTIVE		1
1.2.	APPROACH		1
2.	USE CASES AND INFORMATION REQUIREMENTS	3	
2.1.	THE EU BUDGET AS LOD: USE CASES		3
2.1.1.	<i>Increase public understanding</i>		3
2.1.2.	<i>Compare budgets</i>		4
2.1.3.	<i>Follow the money</i>		5
2.1.4.	<i>Data analytics</i>		7
2.1.5.	<i>Data journalism</i>		7
2.2.	INFORMATION REQUIREMENTS		8
3.	ANALYSIS OF RELATED INITIATIVES	10	
3.1.	NATIONAL BUDGETS		10
3.1.1.	<i>State of government budget data publication</i>		10
3.1.2.	<i>EU - Czech Republic</i>		11
3.1.3.	<i>EU - Finland</i>		12
3.1.4.	<i>Brazil</i>		13
3.1.5.	<i>United States of America</i>		17
3.2.	THE WORLD BANK: BOOST INITIATIVE		18
3.3.	OPEN SPENDING BUDGET DATA PACKAGE		22
3.4.	CONCLUSION		22
4.	STRUCTURAL ELEMENTS OF THE EU BUDGET	24	
4.1.	LEVELS OF THE EU BUDGET		24
4.2.	CONCEPTUAL ELEMENTS OF THE EU BUDGET		25
5.	IDENTIFICATION AND AGGREGATION OF CLASSES AND PROPERTIES	34	
5.1.	DESIGN PRINCIPLES		34
5.2.	EXCLUDED ELEMENTS		34
5.2.1.	<i>Unused elements and attributes</i>		34
5.2.2.	<i>Excluded element: Data</i>		35
5.2.3.	<i>Excluded element: HResources</i>		36
5.2.4.	<i>Excluded element: HResource</i>		36
5.2.5.	<i>Excluded element: Amounts</i>		37
5.2.6.	<i>Excluded element: Relations</i>		37
5.3.	CONCEPTUAL DATA MODEL OF THE EU BUDGET		38
5.3.1.	<i>Class: Amount</i>		38
5.3.2.	<i>Class: Monetary Value</i>		42
5.3.3.	<i>Class: Political Category (Catpol)</i>		42
5.3.4.	<i>Class: Corporate Body</i>		44
5.3.5.	<i>Class: EU Programme</i>		45
5.3.6.	<i>Class: Nomenclature</i>		46
5.3.7.	<i>Class: Staff</i>		51
6.	EU BUDGET VOCABULARY IN RDF	53	
6.1.	URI TEMPLATE		53
6.2.	CLASSES		53
6.3.	PROPERTIES PER CLASS		53
6.3.1.	<i>Amount [bud:Amount]</i>		53
6.3.2.	<i>Monetary Value [bud:MonetaryValue]</i>		54
6.3.3.	<i>Political Category [bud:Catpol]</i>		55
6.3.4.	<i>Nomenclature [bud:Nomenclature]</i>		56
6.3.5.	<i>Staff [qb:DataSet]</i>		58
6.4.	METADATA		59
6.5.	CONTROLLED VOCABULARIES		61

7.	EU BUDGET VOCABULARY IN OWL	64
8.	NEXT STEPS	65
	BIBLIOGRAPHY	67
	ANNEXES	70
	ANNEX I: TERMINOLOGY	70
	ANNEX II: URI NAMESPACE REGISTRATION REQUEST	72

List of Figures

Figure 1: drivers of data transparency	3
Figure 2: Personal tax indication per function of government, based on gross salary	4
Figure 3: The BOOST initiative.....	4
Figure 4: Financial Transparency System (FTS).....	5
Figure 5: Components of dBrain [7]	6
Figure 6: Where does my money go?	7
Figure 7: The NY times budget analysis.....	8
Figure 8: Revenue and expense categories of the Finnish budget	13
Figure 9: Example of a code lists of the USA budget.....	17
Figure 10: BOOST Web Interface [15]	21
Figure 11: conceptual data model of the EU budget.....	38

List of Tables

Table 1: State of Budget Publication	10
Table 2: Czech budget – examples of code lists	11
Table 3: Czech budget - balance sheet items	11
Table 4: Example – Expense item of the 2013 budget	14
Table 5: Example - RDF representation	16
Table 6: Polish Budget – BOOST data entities	19
Table 7: Polish Budget – BOOST code lists	20
Table 8: Obligatory attributes of the budget resources in the <i>budget data package</i>	22
Table 9: Recommended attributes of the budget resources in the <i>budget data package</i>	22

Specification of the EU Budget Vocabulary

Table 10: levels of the EU budget.....	24
Table 11: elements of the EU budget.....	25
Table 12: Nomenclature attributes.....	31
Table 13: Justification for excluding the <data> attributes.....	35
Table 14: <i>Amount</i> properties.....	39
Table 15: <i>Amount</i> subclasses.....	40
Table 16: example of <i>Commitment</i>	40
Table 17: example of <i>Commitment</i> with reserve.....	41
Table 18: <i>Monetary Value</i> classes and properties.....	42
Table 19: example of <i>Monetary Value</i>	42
Table 20: <i>Political Category</i> properties.....	42
Table 21: <i>Catpol</i> subclasses.....	43
Table 22: example of <i>Catpol</i>	43
Table 23: example of commitment instances for catpol ceilings.....	44
Table 24: <i>Corporate body</i> classes and properties.....	44
Table 25: example of <i>corporate body</i>	45
Table 26: <i>EU Programme</i> classes and properties.....	45
Table 27: example of <i>EU Programme</i>	46
Table 28: <i>Nomenclature</i> classes and properties.....	47
Table 29: <i>Nomenclature</i> subclasses.....	48
Table 30: example of <i>nomenclature</i> subclasses.....	49
Table 31: <i>Staff</i> classes and properties.....	51
Table 32: <i>Staff</i> subclasses.....	51
Table 33: example of <i>administrator</i>	52
Table 34: Classes of the RDF Data Model.....	53
Table 35: RDF properties of bud:Amount.....	53
Table 36: Sub-classes of bud:Amount.....	54
Table 37: RDF properties of bud:MonetaryValue.....	54
Table 38: RDF properties of bud:Catpol.....	55

Specification of the EU Budget Vocabulary

Table 39: Sub-classes of bud:Catpol	55
Table 40: RDF properties of bud:Nomenclature	56
Table 41: Sub-classes of bud:Catpol	57
Table 42: Staff DataCube - dimensions	59
Table 43: Staff DataCube - observations	59
Table 44: Metadata - dataset properties.....	59
Table 45: Metadata - distribution properties (RDF)	60
Table 46: Named Authority Lists	61
Table 47: Named Authority Lists for staff tables	62
Table 48: Budget-specific controlled vocabularies.....	63

1. INTRODUCTION

This document is prepared in the context of Action 1.1 of the Interoperability Solutions for European Public Administrations (ISA) programme¹ of the European Commission. ISA action 1.1 aims at improving semantic interoperability in European eGovernment systems².

Studies conducted on behalf of the European Commission [1] show that businesses and citizens still face difficulties in finding and reusing public sector information. In its communication on Open Data [2] of 12 December 2011, the European Commission states that the availability of information in a machine-readable format as well as a thin layer of commonly agreed metadata could facilitate data cross-referencing and interoperability. Therefore, the value for reuse of the data would be considerably enhanced.

A type of public sector information that would benefit from increased interoperability and reusability is budgetary data. In this vein, this report presents a dissemination vocabulary for publishing the yearly final adopted EU budget as machine-readable linked open data, henceforth referred to as the "EU budget vocabulary".

1.1. Objective

The EU budget vocabulary is designed to facilitate the exchange, increase the understandability and foster the reusability of budgetary information published by the EU.

The expected benefits from applying the EU budget vocabulary include:

- **Improved transparency towards citizens** by increasing the public understanding of the yearly adopted EU budget;
- **Better decision making** by providing high-quality, fit-for-purpose and well-structured data to support decision making;
- **Increased harmonization in budget publication** across institutions, government levels and countries by providing a common way of structuring public budgets.
- **Increased flexibility for integrating data** by embracing linked data principles; and
- **Greater technology independence** by supporting multiple machine-readable open data formats.

1.2. Approach

The development of the EU budget vocabulary is based on the process and methodology for developing semantic agreements³ of the ISA Programme [3]. The methodology outlines the roles that the different actors in this process play, and the process steps that need to be taken. Besides a consensus building process, which is rather heavyweight for the purpose of this work, the methodology specifies a

¹ The ISA Programme of the EU: http://ec.europa.eu/isa/index_en.htm

² Action 1.1. of the ISA Programme: http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-action_en.htm

³ Methodology for creating semantic agreements: <https://joinup.ec.europa.eu/node/67006>

technical process which describes the stages that must be undertaken in the development of a vocabulary. Hence, our approach consists of the following phases:

- **Chapter 2** | Determine use cases and requirements
- Develop a domain model
 - **Chapter 3** | Identify and analyse other solutions and data models
 - **Chapter 4** | Identify and analyse existing structural elements of the EU budget
 - **Chapter 5** | Specify data entities, attributes and controlled vocabularies
- **Chapter 6** | Specify the EU budget vocabulary in RDF

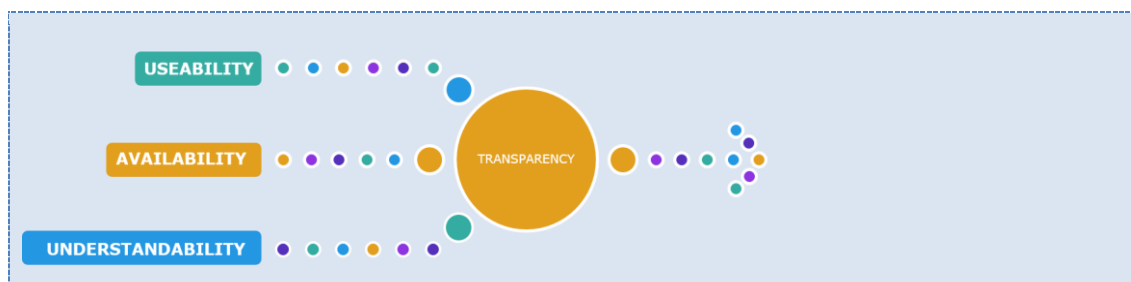
2. USE CASES AND INFORMATION REQUIREMENTS

The first step in the development process of the EU budget vocabulary, which is described in section 1.2, aims at identifying how users of budgetary open data could benefit from the vocabulary. The identified use cases are described in section 2.1. Based on the identified use cases, specific requirements for budget data are elicited and listed in section 2.2.

2.1. The EU budget as LOD: Use cases

In order to increase the transparency of the EU budget, it should be published in a more usable and understandable way. Specifying an RDF dissemination vocabulary for the EU budget by reusing, to the extent possible, terms from existing vocabularies will hence increase its understandability. Moreover, since linked data was identified as a key technology for data integration [4], publishing budget data as linked open data (LOD) will increase its availability and usability.

Figure 1: drivers of data transparency



Based on case studies presented in light of the Global Initiative for Fiscal Transparency⁴ by the cities of New York and Chicago, the governments of Kenya and Mexico, the Open Knowledge Foundation [5] and based on a study conducted by the ISA Programme of the European Commission [6], a number of use cases for publishing EU budget data as LOD using the EU budget vocabulary have been identified. These use cases are elaborated in the following sections.

2.1.1. Increase public understanding

National budgets that are published by governmental institutions often consist of complicated documents and bulky data sets. Moreover, budget data is described using a specific jargon, which is not always understandable by the wide public and data reusers. Readers of the budget aim at gaining a quick understanding of public budgets rather than performing in-depth analyses of published documentation therefore benefit from having budget data available in more understandable ways.

Example | Open Spending

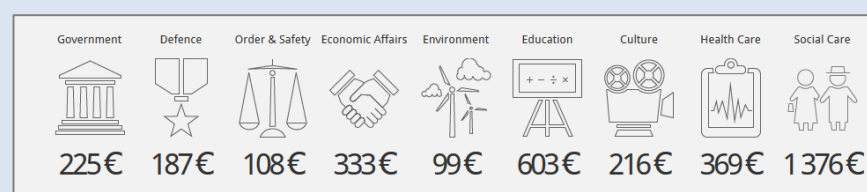
The Open Spending community provides a platform on which public budgets can be published, analysed and visualised. The European multiannual budget (MFF) was already imported into the Open Spending tool by use of a script. The developer of the script could however not guarantee the quality because the data was imported from relatively unstructured and complex sources.

⁴ GIFT: Global Initiative for Fiscal Transparency: <http://fiscaltransparency.net/>

Example | Meie Raha: Estonian state budget

Meie Raha⁵ is a non-profit initiative that aims at publishing the Estonian state budget in an understandable way by structuring and visualising major income and expenditure articles. Via an interactive web tool, users can simulate modifications in expenditures and directly see the tax impact their manipulations. Uploading the budget data into the tool required substantial manual work to gather data from various resources, tables and PDF documents. A machine-readable format of the collected data is made available by the third party organisation⁶. However, the intervention of manual manipulations and the variety in data sources endanger the accuracy and integrity of the provided data.

Figure 2: Personal tax indication per function of government, based on gross salary



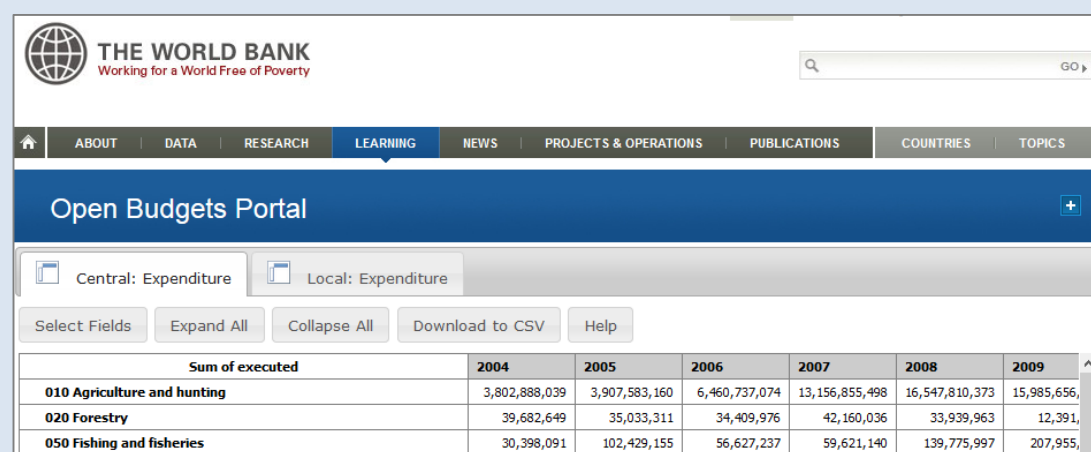
2.1.2. Compare budgets

Publishing budgets in a common way, based on linked data principles, allows users to compare more easily the budgets across years, between countries, public institutions, policy domains, etc. For example, in order to compare how much money European Member States spend on education, expenditure should be categorised in a common way across countries. Applying linked data principles facilitates data integration, as the same vocabularies are used and specific elements of the data can be linked.

Example | The World Bank: BOOST

The BOOST initiative of the world bank provides a common way to describe national budgets. Although BOOST allows budgets to remain country-specific, all boost budgets should have common features. The BOOST initiative is further elaborated in section 3.2.

Figure 3: The BOOST initiative



⁵ Meie Raha – “Our Money”: <http://meieraha.ee/>

⁶ The Mei Raha data sets can be downloaded via <http://meieraha.ee/?page=data>

2.1.3. Follow the money

In order to obtain an exhaustive view of the public spending process, i.e. from tax collection and political decisions over to tendering and then to actual payments, budget data should be integrated with other datasets such as procurement data. The future integration of budget data with other data sets can be done more easily and flexibly by using linked data technologies for the publication of the data sets.

Example | Financial Transparency System (FTS), European Commission

The Financial Transparency System of the European Commission is an online platform that gives an overview of the beneficiaries of funds awarded by the European Commission. The following information is provided:

Data for any given year is not published until the following year.

Search results include:

- the name of the beneficiary;
- the subject or purpose of the expenditure;
- the location of the beneficiary;
- the amount and type of expenditure (operational vs. administrative);
- the department that awarded the funding, e.g. a directorate-general or agency;
- the relevant part of the EU budget in which the amount is committed;
- the amount was booked in the accounts;
- if applicable, the type of action, such as an EU programme; and
- if applicable, the coordinator, i.e. the beneficiary responsible for redistributing funds in a multi-beneficiary project.

Publishing the EU budget as linked open data would allow to further integrate the budget and spending data of the Financial Transparency System.

Figure 4: Financial Transparency System (FTS)

The screenshot shows the FTS interface with search filters for Year 2013 and France. The search results table contains the following data:

Subject of grant or contract	END-USER DRIVEN DEMO FOR CBRNE
Total amount	24,766,823 €
Year	2013
Responsible department of the Commission	Directorate-General for Enterprise and Industry
Reference (Budget)	CPM.313077-1082601.1 (BGUE)
Budget line and number	Security research (02.04.01.02)
Action Type	Research: Framework programme n°7
Funding Type	Grants
Expense Type	Operational
Action Location	N/A
Geographical Location	N/A

Example | dBrain – Digital Budget and Accounting System

The South Korean Digital Budget and Accounting System (DBAS), also known as *dBrain*, is a financial management information system that manages entire business processes, including

- budget formulation;
- budget execution;
- procurement;
- accounting;
- settlement; and
- performance evaluation [7].

dBrain was developed after an expansion of the size public finance in South Korea. The government needed a powerful system that was able to handle the increasing number of transactions, decisions and data. Before dBrain, many isolated systems were used for public finance management. Today, dBrain consists of 21 unit systems, for example:

- a project management system;
- a budget management system;
- a Revenue and Electronic Bill Presenter payment system; and
- a procurement management system.

The integration of budget data with other data through dBrain allowed the government to

- better use the national budget by avoiding duplicative expenditure and allowing better oversight and control to both policy makers and the public;
- overcome budgeting and accounting difficulties by efficiently exchanging information between government departments;
- ease the review of budgeting and payment information by the congress; and
- make accurate budgeting decisions and forecasts [8]

Figure 5: Components of dBrain [7]

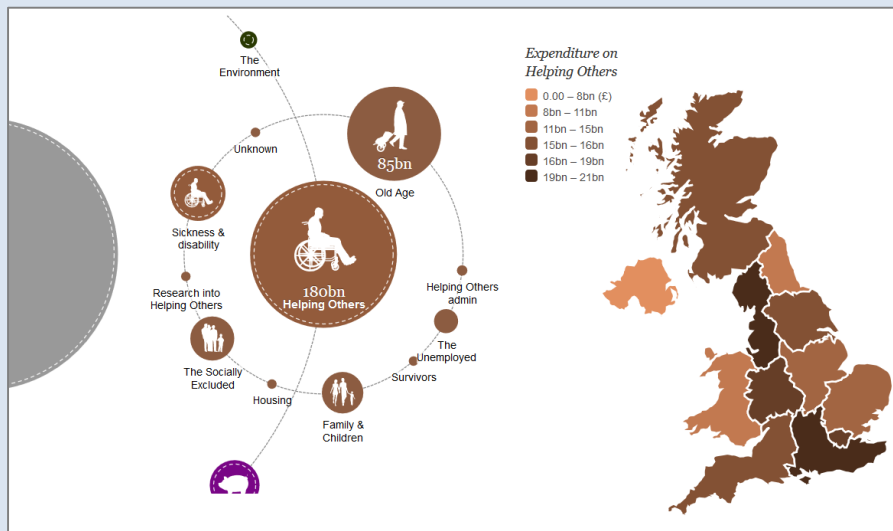


Example | Where does my money go?

*Where does my money go?*⁷ is an initiative that to represent public spending data in the United Kingdom based on regions, functions of government and government departments. Combining national and regional budgets as well as spending data allows the users to follow their tax money from collection to expenditure.

⁷ Where does my money go? is available via <http://wheredoesmymoneygo.org>.

Figure 6: Where does my money go?



2.1.4. Data analytics

Many elements of the Multiannual Financial Framework (MFF) and the yearly EU budgets, such as the allocation of financial and human resources, have a significant impact on the day-to-day operations of EU institutions. Therefore, it is of great importance that the budget is taken into consideration in major business decisions. Making the vast amount of EU budget information available in a well-structured, machine-readable open format would foster the opportunity to conduct accurate and efficient analyses. Such analyses could support business decisions in public administrations as well as in enterprises.

Example | Data value chains

During his presentation at the 2015 SEMIC conference, Dr. Aurer from the University of Bonn explained that linking distributed and heterogeneous data sets can facilitate enterprise development, production and services [9]. In other words, applying linked data technologies brings many opportunities for data analytics, especially when distributed and heterogeneous data sets need to be integrated, fostering more efficient and effective decision-making. The integrated chain of such decentralised datasets, i.e. datasets that are owned by different organisations, in a structured way is called a data value chain.

2.1.5. Data journalism

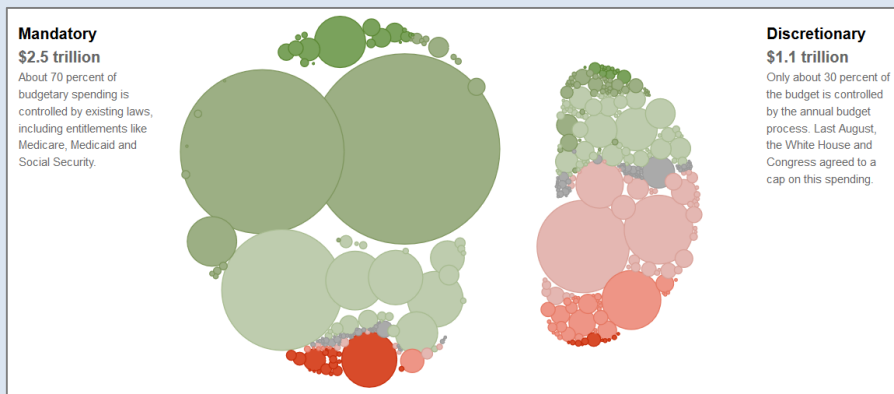
The ever increasing amount of digitized information lead to new ways of producing and disseminating knowledge in society [10]. Data journalism can help journalists in two ways. First, it helps journalists to identify and understand complex information. Second, data journalism can help journalists to tell compelling stories through the use of data visualisation techniques, for example. Publishing information as linked open data could ease the analysis of the big amounts of data that often come from different sources.

Example | The New York Times

The New York Times performed a thorough analysis of the United States 2013 budget proposal. The newspaper created a visualisation which allowed readers to approach the budget data from four different angles: "all spending", "types of spending", "changes" and

“department totals”. In the article⁸, the main focus is drawn towards the data whereas the amount of text is kept to a minimum. In other words: *the data tells the story*.

Figure 7: The NY times budget analysis



2.2. Information requirements

Based on the use cases, the following information requirements were elicited:

1. **Provision of data dictionary.** The meaning and structure of the elements and values of which the budget is made up are often unclear. Therefore, the publication of a data dictionary and structural metadata, such as the conceptual metadata of the EU budget vocabulary and the associated code lists, is required in order to increase the users’ understanding of the budget data.
2. **Use of persistent URIs.** Budget data resources should be identified using unique resource identifiers (URIs) in order to foster the integration with other data sets and to facilitate their reuse.
3. **Support traceability of budget data.** Traceability of the budget would enable users to elicit from the published data all relevant information that is related to the development of the budget. Such information could include a trace of modifications that were applied to the published numbers, amendments, references to laws and decisions, etc.
4. **Support multilingualism.** In the context of the European Union, the budget information should be made available in several languages. Hence, the EU budget vocabulary should support multilingualism.
5. **Visualised information.** In order to foster public understanding of the EU budget, the information should be provided in a format that can be easily integrated with data visualisation tools.
6. **Provide data as a service.** Users of the budget data have different requirements regarding the format in which they want to obtain the data. Therefore, the data should be provided as a service from which multiple machine-readable formats can be retrieved, including CSV, XML, RDF, etc. Moreover, for specific user groups, it might be relevant to only query a specific subset of the data. Therefore, it would be useful to have a search interface

⁸ Four Ways to Slice Obama’s 2013 Budget Proposal:

http://www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html?_r=1&_t=27/10/2015

allows filtering the information and downloading the filtered subset, e.g. for specific political domains, institutions or geographic regions. In order to foster the reuse of the budget data, the data should be accessible via APIs or query interfaces.

3. ANALYSIS OF RELATED INITIATIVES

In order to develop a vocabulary that responds to the needs of its potential users, it is crucial to identify how already existing initiatives deal with the publication of budget data. Section 3.1 gives an overview of how EU Member States and some non-EU countries deal with the publication of the budget data. In section 3.2, the World Bank's BOOST initiative for the publication of government budgets is described.

3.1. National budgets

The maturity regarding the publication of public sector information differs across governments. Therefore, before analysing existing examples of publishing government budgets, an analysis of the state of government budget publication is described.

3.1.1. State of government budget data publication

In order to assess the maturity level of open data sets, Tim Berners-Lee developed a five-star deployment scheme for Open Data⁹, which we applied to the published budgets of EU Member States. Besides the budgets published by EU Member States, several interesting initiatives from outside the European Union can be taken into account for the specification of an EU budget vocabulary. Table 1 gives an overview of the assessment.

Table 1: State of Budget Publication

Rating	Level	Countries	Total
★	Published	Bulgaria Cyprus Denmark Greece Malta Portugal Romania Spain UK	9 Member States
★★	Machine readable	Luxemburg Austria Belgium Croatia Estonia Hungary Ireland Slovakia Sweden	9 Member States
★★★	Open format	Czech Republic Finland France Germany Italy Latvia Lithuania Netherlands Poland Slovenia USA	10 Member States
★★★★	Use of URIs	None	0 Member States
★★★★★	Linked	Brazil	0 Member States

Existing initiatives that publish their government budgets in machine-readable formats and for which vocabularies exist, could provide valuable input for the development of an EU budget vocabulary. Therefore, these initiatives are further analysed in the following sections.

⁹ Deployment scheme for Open Data: <http://5stardata.info/>

3.1.2. EU - Czech Republic

Although the Open Knowledge Foundation only ranks the Czech Republic on place 23 in the *openness of government budgets* ranking [11], the Czech budget is an interesting example to be considered in the scope of the development of an EU budget vocabulary. The budget is published in CSV. Moreover, the structural elements are described in XSD. Table 2 provides an overview of some of the code lists that are published on the website of the Czech ministry of finance¹⁰.

Table 2: Czech budget – examples of code lists

Code List	Description
Classification of economic activities (NACE)	Classification of Economic Activities by the Czech Statistical Office.
Classification of Government Functions (COFOG)	Classification of Government Functions following COFOG.
Type of management	Types of management include <i>centrally</i> or <i>locally</i> managed organization
Number of inhabitants category	Interval distribution of number of inhabitants in municipalities
NUTS	Classification of territorial units for statistics according to the Czech Statistical Office from the district level
Codes institutional sectors and subsectors (ISEKTOR)	Code list of institutional sectors and subsectors according to the Czech Statistical Office.
PVS revenues and PVS expenditures	Structural classification of revenues and expenditures of the state budget, for example: funding, tax revenue, revenue from the EU budget for agriculture, etc.
Specific indicator - expenditure, specific indicator - revenue and cross-sectional indicator	Mandatory indicators of the state budget established by the state budget law for the year and individual chapters. Further subdivided into comprehensive, specific and cross-sectional indicators.
Budget item	Code list representing the hierarchy of budget items, e.g. 1. Tax revenues 1.1. Taxes on income, profits and capital gains 1.1.1. Taxes on personal income

The elements specified in the budget reports, which can be downloaded in CSV, differ per type of report. The balance sheet, for example, contains the fields that are listed in Table 3.

Table 3: Czech budget - balance sheet items

Column	Fieldname	Description
--------	-----------	-------------

¹⁰ The metadata of the Czech budget is available via <http://monitor.statnipokladna.cz/en/2014/data/web-services>

1	ZC_VYKAZ:ZC_VYKAZ	Statement
2	ZC_VTAB:ZC_VTAB	Statement table
3	0FISCPER:0FISCPER	Fiscal year/period
4	ZC_UCJED:ZC_UCJED	Entity number
5	ZC_ICO:ZC_ICO	Company ID
6	0FM_AREA:0FM_AREA	Chapter
7	ZC_KRAJ:ZC_KRAJ	County
8	ZC_NUTS:ZC_NUTS	NUTS entity
9	ZC_POLVYK:ZC_POLVYK	Statement item
10	ZSYN_UCET:ZC_SYNUC	Synthesis account
11	ZU_AOBTTO:ZU_AOBTTO	Gross – current period
12	ZU_AOKORR:ZU_AOKORR	Correction – current period
13	ZU_AONET:ZU_AONET	Net – current period
14	ZU_MONET:ZU_MONET	Net – previous period

3.1.3. EU - Finland

As part of a program to open up public sector information, the Finnish Ministry of Finance has made available the proposed budget in machine-readable formats, including CSV and XML. Since 2014, all budget proposals are produced in machine-readable format within 36 hours after publication of the budget. Besides the machine-readable files, the Finnish government has published structural metadata describing the budget¹¹.

There are 14 classes in the budget data model:

- Base (Juuri)
- Year (Vuosi)
- Work (Teos)
- Position (Kanta)
- Department (Osasto)
- Login number (Tuloluku)
- Login point (Tulomomentti)
- Budget section (Paaluokka)
- Expenditure section (Menoluku)
- Expenditure subsection (Menomomentti)

¹¹ The Finnish budget files and metadata are available via <http://budjetti.vm.fi/indox/opendata/>
27/10/2015

- Comparison numbers (Vertailuluvut)
- Statement of cash (Laskelmaraha)
- Budget statement (Budjettilaskelma)
- Amendment (Muutos)

The Finnish data model is using a code list for categorising revenue and expenditure, as represented in Figure 8.

Figure 8: Revenue and expense categories of the Finnish budget

- [11. Taxes and similar REVENUE](#)
- [12. MISCELLANEOUS REVENUE](#)
- [15. LOANS](#)
- [23. PRIME MINISTER'S OFFICE](#)
- [24. MINISTRY OF FOREIGN AFFAIRS of administration](#)
- [25. JUSTICE Ministry is responsible](#)
- [26. IMMEDIATE The Ministry is responsible](#)
- [27. DEFENCE The Ministry is responsible](#)
- [28. Valtiovarainministeriön hallinnonala](#)
- [29. EDUCATION AND CULTURE The Ministry is responsible](#)
- [30. Agriculture and Forestry Ministry is responsible](#)
- [31. TRANSPORT AND COMMUNICATIONS administrate](#)
- [32. The Ministry's administrative sector](#)
- [33. SOCIAL AND HEALTH The Ministry is responsible](#)
- [35. ENVIRONMENT The Ministry is responsible](#)

3.1.4. Brazil

The Secretary of Federal Budget in order to provide greater transparency to its processes by enabling further participation of Brazilian citizens, developed the "Federal Budget in Open Format", which is the publication of the Federal budget and spending data in period from 2000 to 2013.

The objective of the project "Federal Budget in Open Format" was to create an OWL¹² ontology¹³ of the expenses of the federal budget classification (Modelo ontológico da Classificação das Despesas do Orçamento Federal Brasileiro), covering the categories and concepts included in the budget of the Technical Manual 2013 - MTO¹⁴. OWL is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans.

There are 22 classes in the ontology for the expenditure of the federal budget [12]:

- Economic Category (CategoriaEconomic)
- Classifier (Classificador)
- Expense Element (ElementoDespesa)
- Sector (Esfera)
- Financial Year (Exercicio)
- Source of Funds (FonteRecursos)
- Government Spending (GrupoNatDespesa)

¹² <http://www.w3.org/TR/owl-features/>

¹³ <http://vocab.e.gov.br/2013/09/loa#>

¹⁴ http://www.orcamentofederal.gov.br/informacoes-orcamentarias/manual-tecnico/MTO_2014.pdf

- Expense Item (ItemDespesa)
- Subtitle (Subtitulo)
- Application Mode (ModalidadeAplicacao)
- Program (Programa)
- Action (Acao)
- Project (Projeto)
- Activity (Atividade)
- Special Operation (OperacaoEspecial)
- Budget Plan (PlanoOrcamentario)
- Function (Funcao)
- Subfunction (Subfuncao)
- Budget Unit (UnidadeOrcamentaria)
- Organisation (Orgao)
- Identifier of Use (IdentificadorUso)
- Primary Result (ResultadoPrimario)

The budget documents are available in RDF as N-Triples¹⁵. The tables below give an example of the transformation of an “Expense Item” (ItemDespesa) from the 2013 budget into RDF statements. Table 4 describes the entities of the “Expense Item” extracted from an N-Triple, with their respective attributes.

Table 4: Example – Expense item of the 2013 budget

Classes & attributes	Value/ Description
Sector Esfera	Fiscal budget Orçamento Fiscal (10)
Body Órgão	Ministry of Planning, Budget and Management Ministério do Planejamento, Orçamento e Gestão (47000)
Budget Unit Unidade Orçamentária	Ministry of Planning, Budget and Management Ministério do Planejamento, Orçamento e Gestão (47101)
Function Função	Administration Administração (04)
Subfunction Subfuncao	Information Technology Tecnologia da Informação (126)
Program Programa	Democracy and Improvement of Public Administration Democracia e Aperfeiçoamento da Gestão Pública (2038)
Action Ação	Management and Enhancement of Information Technology Resources and Logistics of the Federal Public Service

¹⁵ <http://www.w3.org/TR/n-triples/>
27/10/2015

Specification of the EU Budget Vocabulary

	Gestão e Aprimoramento dos Recursos de Tecnologia da Informação e de Logística do Serviço Público Federal (20U2)
Subtitle Subtítulo	National Nacional (0001)
Budget plan Plano Orçamentário	Integration of Information Systems of the Federal Government Integração dos Sistemas Informatizados do Governo Federal (0003)
Identifier of use Identificador de Uso	Non-recurrent resources Recursos não destinados à contrapartida (0)
Source of resources Fonte de Recursos	Ordinary resources Recursos Ordinários (100)
Economic Category Categoria Econômica	Current expenditures Despesas Correntes (3)
Group nature of expenditure Grupo de Natureza da Despesa	Other current expenses Outras Despesas Correntes (3)
Enforcement mode Modalidade de Aplicação	Direct investments Aplicações Diretas (90)
Expense element Elemento de Despesa	To be defined A Definir (00)
Primary Result Resultado Primário	Discretionary primary, not PAC, considered in the calculation of RP Primária discricionária, não PAC, considerada no cálculo do RP (2)
PLOA ¹⁶ value (attribute) Valor PLOA	R\$ 0,00
LOA ¹⁷ value (attribute) Valor LOA	R\$ 1.686.356,00
LOA value with credits (attribute) Valor LOA mais Créditos	R\$ 2.337.313,00
Committed value (attribute) Valor Empenhado	R\$ 0,00
Payment value (attribute) Valor Liquidado	R\$ 0,00
Amount payable (attribute) Valor Pago	R\$ 0,00

¹⁶ Projeto de Lei Orçamentária Annual: Draft Annual Budget

¹⁷ Lei Orçamentária Annual Budget, also called initial allocation

Table 5 shows the RDF representation of the "Expense Item". Each line represents a Triple. At the first line the resource of the "Expense Item" is defined. Row 2 to row 17 define the classifiers related to the "Expense Item" and rows 18 to 23 set the financial values of the "Expense Item".

Table 5: Example - RDF representation

Row	Subject	Predicate	Object
1		<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://vocab.e.gov.br/2013/09/loa#ItemDespesa>
2		<http://vocab.e.gov.br/2013/09/loa#temEsfera>	<http://orcamento.dados.gov.br/2013/id/Esfera/10>
3		<http://vocab.e.gov.br/2013/09/loa#temUnidadeOrcamentaria>	<http://orcamento.dados.gov.br/2013/id/UnidadeOrcamentaria/47101>
4		<http://vocab.e.gov.br/2013/09/loa#temFuncao>	<http://orcamento.dados.gov.br/2013/id/Funcao/04>
5		<http://vocab.e.gov.br/2013/09/loa#temSubfuncao>	<http://orcamento.dados.gov.br/2013/id/Subfuncao/126>
6		<http://vocab.e.gov.br/2013/09/loa#temPrograma>	<http://orcamento.dados.gov.br/2013/id/Programa/2038>
7		<http://vocab.e.gov.br/2013/09/loa#temAcao>	<http://orcamento.dados.gov.br/2013/id/Acao/20U2>
8		<http://vocab.e.gov.br/2013/09/loa#temSubtitulo>	<http://orcamento.dados.gov.br/2013/id/Subtitulo/20U20001>
9	_:AmapX3aXItemDespesaX40XX40X7988	<http://vocab.e.gov.br/2013/09/loa#temPlanoOrcamentario>	<http://orcamento.dados.gov.br/2013/id/PlanoOrcamentario/47101203820U20010003>
10		<http://vocab.e.gov.br/2013/09/loa#temIdentificadorUso>	<http://orcamento.dados.gov.br/2013/id/IdentificadorUso/0>
11		<http://vocab.e.gov.br/2013/09/loa#temFonteRecursos>	<http://orcamento.dados.gov.br/2013/id/FonteRecursos/100>
12		<http://vocab.e.gov.br/2013/09/loa#temCategoriaEconomica>	<http://orcamento.dados.gov.br/2013/id/CategoriaEconomica/3>
13		<http://vocab.e.gov.br/2013/09/loa#temGND>	<http://orcamento.dados.gov.br/2013/id/GrupoNatDespesa/3>
14		<http://vocab.e.gov.br/2013/09/loa#temModalidadeAplicacao>	<http://orcamento.dados.gov.br/2013/id/ModalidadeAplicacao/90>
15		<http://vocab.e.gov.br/2013/09/loa#temElementoDespesa>	<http://orcamento.dados.gov.br/2013/id/ElementoDespesa/00>
16		<http://vocab.e.gov.br/2013/09/loa#temResultadoPrimario>	<http://orcamento.dados.gov.br/2013/id/ResultadoPrimario/2>
17		<http://vocab.e.gov.br/2013/09/loa#valorProjetoLei>	"0.00"^^<http://www.w3.org/2001/XMLSchema#decimal>
18		<http://vocab.e.gov.br/2013/09/loa#valorDotacaoInicial>	"1686356.00"^^<http://www.w3.org/2001/XMLSchema#decimal>

19	<http://vocab.e.gov.br/2013/09/loa#valorLeiMaisCredito>	"2337313.00"^^<http://www.w3.org/2001/XMLSchema#decimal>
20	<http://vocab.e.gov.br/2013/09/loa#valorEmpenhado>	"0.00"^^<http://www.w3.org/2001/XMLSchema#decimal>
21	<http://vocab.e.gov.br/2013/09/loa#valorLiquidado>	"0.00"^^<http://www.w3.org/2001/XMLSchema#decimal>
22	<http://vocab.e.gov.br/2013/09/loa#valorPago>	"0.00"^^<http://www.w3.org/2001/XMLSchema#decimal>

3.1.5. United States of America

The Office of Management and Budget (OMB)¹⁸ assists the President in the creation of the President's budget by gathering data from agencies and compiling it into the final plan to be approved by the President.

The totals for the current and upcoming fiscal years are only projected amounts. Annual budget documents provide Congress, State and local governments, and the public with a complete description of the President's budget plans for the coming fiscal year. Documents are available in pdf; however, many of the tables found are also available for separate viewing and downloading as spreadsheets in xls and csv¹⁹. Separate files for receipts, outlays, and budget authority are available in each of these formats.

The OMB is using code lists in some of the classes. For example in the Budget Authority spreadsheet, 5 classes are using code lists (i.e. Agency Code, Bureau Code, Account Code, Treasury Code, Subfunction Code). The Public Budget Database User's Guide provides²⁰ and overview of the classes and the code lists that are used.

Figure 9: Example of a code lists of the USA budget

Agency Code	Agency Name	Bureau Code	Bureau Name	Account Code	Account Name	Treasury Agency Code	Subfunction Code	Subfunction Title	BEA Category	On- or Off-Budget
002	Judicial Bra 25	Courts of A	5101	Registry Administration	10	752	Federal litigati	Mandatory	On-budget	
002	Judicial Bra 25	Courts of A	5114	Judiciary Information Technology Fund	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 25	Courts of A	5114	Judiciary Information Technology Fund	10	752	Federal litigati	Mandatory	On-budget	
002	Judicial Bra 25	Courts of A	5115	National Fine Center	10	752	Federal litigati	Mandatory	On-budget	
002	Judicial Bra 26	Administrat	0927	Salaries and Expenses	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 26	Administrat	0936	Judicial survivors' annuity program	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 26	Administrat	0940	Study of construction of office building	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 30	Federal Juq	0928	Salaries and Expenses	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 30	Federal Juq	8123	Gifts and donations, Federal Judicial Ce	10	752	Federal litigati	Mandatory	On-budget	
002	Judicial Bra 34	Bicentennia	0933	Bicentennial activities	10	808	Other general	Discretionary	On-budget	
002	Judicial Bra 35	Judicial Ret	0941	Payment to Judiciary Trust Funds	10	752	Federal litigati	Mandatory	On-budget	
002	Judicial Bra 35	Judicial Ret	8110	Judicial Survivors' Annuities Fund	10	602	Federal emplc	Mandatory	On-budget	
002	Judicial Bra 35	Judicial Ret	8120	Operation of the Public Defender Servic	10	752	Federal litigati	Discretionary	On-budget	
002	Judicial Bra 35	Judicial Ret	8122	Judicial Officers' Retirement Fund	10	602	Federal emplc	Mandatory	On-budget	

For the budget authorities and the outlay files, the account details are categorised using the following keys:

1. **Agency** codes generally correspond to the Cabinet department or independent agency with primary responsibility for the program.
2. **Bureaus** are major subdivisions within Cabinet departments that correspond to major organizational areas.
3. **Account**. These numbers are assigned by the Treasury Department and are used for budget presentation and financial management.
4. **Subfunction**. Accounts are categorized by the subfunctions. Accounts that include multiple subfunctions are subdivided into subfunction components.

¹⁸ <https://www.whitehouse.gov/omb/>

¹⁹ <https://www.whitehouse.gov/omb/budget/Supplemental>

²⁰ https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/db_guide.pdf

5. **Treasury Agency.** These codes are assigned by the Department of the Treasury and are defined in the Treasury Financial Manual, Supplement to Volume I, Federal Account Symbols and Titles²¹, which contains the titles corresponding to these codes.
6. **Budget Enforcement Act category (BEA).** The Budget Enforcement Act (the "Act") made distinctions between mandatory and discretionary accounts for certain provisions of the Act.
7. **Grant/Non-grant.** Grants to State and local governments are separated from non-grant outlays.
8. **On or Off Budget** identifies if an account is on or off the budget.

Receipt account data are categorised as follows:

1. **Source category.** These are major categories of governmental receipts, such as individual income tax receipts or corporation income tax receipts.
2. **Source subcategory.** Presents further detail within source categories
3. **Agency** codes generally correspond to the Cabinet department or independent agency with primary responsibility for the program.
4. **Bureaus** are major subdivisions within Cabinet departments that correspond to major organizational areas.
5. **Account.** These numbers are assigned by the Treasury Department and are used for budget presentation and financial management.
6. **Treasury Agency.** These codes are assigned by the Department of the Treasury and are defined in the Treasury Financial Manual, Supplement to Volume I, Federal Account Symbols and Titles²², which contains the titles corresponding to these codes
7. **On or Off Budget** identifies if an account is on or off the budget.

3.2. The World Bank: BOOST initiative

In 2009, the World Bank launched an initiative to foster the efficiency and transparency of public spending: the *BOOST* initiative. BOOST provides a tool for describing and publishing government budgets. The tool aims at

- facilitating the analysis of the allocation and use of public expenditure;
- making public spending data on national and sub-national level more open and accessible; and
- supporting policy recommendations by facilitating more practical and insightful analytical work.

BOOST allows budgets to remain country-specific. Nevertheless, all BOOST budgets should have common features. Each data set contains information on the approved budget, on the revised budget and on the actual expenditure amounts. The numbers are broken down in several categories [13]:

- Government level (central or local);

²¹ http://www.fiscal.treasury.gov/fsreports/ref/fastBook/fastbook1_march_2015.pdf

²² http://www.fiscal.treasury.gov/fsreports/ref/fastBook/fastbook1_march_2015.pdf

- Administrative unit (typically a ministry, department, agency, university, towns and villages);
- Economic classification (wages, goods and services, capital expenses, etc.);
- Functional classification (sector and sub-sector);
- Program classification (if the country uses program-based budgeting); and
- Financing source (budget revenue, domestic or foreign borrowing, etc.).

The amount of levels that are specified within these categories as well as the code lists used to express each of the values within these categories, are country-specific. Poland, for example, has applied the BOOST classes as indicated in Table 6.

Table 6: Polish Budget – BOOST data entities

BOOST Elements	Polish Elements	Explanation
Administrative	admin1	Government level: central or local
	admin2	Government agency
	admin3	Government sub-agency
	admin4	Provinces
	admin5	Counties
	admin6	Municipalities
	admin7	Municipality code
	admin8	Type of local government
Functional	func1	Represents 33 functional categories of public spending, including "agriculture and hunting", "information technology", "national defence", etc.
	func2	Represents sub-categories of functions. Within defence (func1) for example, func2 values could include "air troops", "ground troops", "military police", etc.
Economic	econ1	Represents 8 economic categories, including "grants", "benefits to individuals", "capital expenditure", etc.
	econ2	Represents one economic sub-categories, e.g. "grants for cultural institutions", "grants for non-public universities", etc.
Financing source	fin_source	The source of financing for a particular expenditure.
Custom variables	exp_type	Type of expenditures contained in a particular line of data, corresponding to frequently used economic spending categories: "personnel", "non-personnel recurrent", "capital" and "Other".
	transfer	This variable is used to indicate inter-budgetary transfers between the central government and local levels. It can be

Specification of the EU Budget Vocabulary

		used as a filter to avoid double counting when performing analyses on the data.
Amounts	approved	Represents the planned/approved budget amounts.
	adjusted	Represents amended budget amounts.
	executed	Represents the actual expenditure (budget execution).

The World Bank does not impose standard reference data sets, such as code lists, on countries using BOOST to publish budgetary data. The code lists used for budgetary data in the Open Budgets Portal are therefore country-specific. Each country publishes a BOOST user manual²³, which contains the code reference lists. Table 7 gives an overview of the code lists which are defined for the Polish budget [14]. For elements admin6 and admin7 (approximately 2.940 unique values), no code lists were added to the user's manual in the interest of brevity. Not making available the full code lists significantly limits the usability of the Polish BOOST data model.

Table 7: Polish Budget – BOOST code lists

Budget elements (PL)	Code list	Example values
	Level of government budget	1. Central Government Budget 2. Local Government budget
admin2 admin1	Agency code list	1. Chancellery of the President 2. Chancellery of the Sejm 15. General courts 88. General prosecutor organisational units
admin3	Sub-agency code list	15.01. Ministry of Justice 15.02. The Court of Appeal in Warsaw 15.03. The Court of Appeal in Katowice
admin4	Provinces code list	02. Voivodeship Dolnoslaskie 04. Voivodeship Kujawsko-pomorskie 06. Voivodeship Lubelskie
admin5	Counties code list	201. County boleslawiecki 202. County dzierzoniowski 203. County glogowski
admin8	Local government type code list	1. Voivodeship 2. County 3. City with county status 4. Urban municipality
func1	List of functional domains	010. Agriculture 020. Forestry 050. Fishing and fisheries
func2	List of functional sub-domains	01001. Advisory Centre for Agriculture 01002. Voivodeship advisory centres for Agriculture

²³ Country-specific manuals and code lists are available via <http://wbi.worldbank.org/boost/country>
27/10/2015

		01003. Dissemination of agricultural advisory services
econ1	List of economic categories	2. Grants and subsidies 3. Benefits to individuals 7. Settlements with banks
econ2	List of economic sub-categories	224. Earmarked grants [...] for state-owned cultural institution 226. Grant operators of non-public university budget [...] 227. Grant operators for public higher education [...]
fin_source	Sources of financing	0. Budget revenues 1. Programmes financing from European Union non-repayable funds 3. Financing with foreign loans and credits
exp_type	Types of recurrent expenses	1. Personnel 2. Non-personnel recurrent 3. Capital 4. Other

From a practical perspective, BOOST requires budgets to be specified in a spreadsheet, complying with the categorisation as described above. These spreadsheets are then published in the World Bank’s *Open Budgets Portal*²⁴. Moreover, the web interface of the Open Budgets Portal allows a user to browse and filter the budgetary data in a pivot-like manner, as represented in Figure 10. Moreover, filtered datasets can be exported in CSV.

Figure 10: BOOST Web Interface [15]

The screenshot shows the BOOST web interface. On the left, there is a pivot table configuration panel with the following sections:

- Pivot Fields:** A tabbed interface with 'Pivot Fields' and 'Filters' tabs.
- Available Fields:** A list of fields including YEAR, ADMIN1-6, FUNC1-2, ECON1-2, FIN_SOURCE, EXP_TYPE, and TRANSFER.
- Column Labels:** A list containing 'YEAR' with a close icon.
- Row Labels:** A list containing 'FUNC1' with a close icon.
- Available Values:** A list containing APPROVED, ADJUSTED, and EXECUTED.
- Selected Values:** A list containing 'EXECUTED' with a close icon.
- Buttons: 'Apply', 'Reset', and 'Cancel'.

On the right, there is a data table with the following structure:

	2005	2006
Sum of EXECUTED		
01 General purpose state services	920,556,098	1,039,418,903
02 Foreign relations	256,251,569	235,322,778
03 National defense	156,681,499	216,106,284
04 Justice	126,850,303	188,641,215
05 Enforcement of public order and national security	765,184,461	990,356,272
06 Education	2,739,172,044	3,680,468,011
07 Research and innovations	139,468,246	199,490,664
08 Culture, arts, sports and youth activities	321,593,662	510,129,841
09 Healthcare	2,415,247,842	3,155,472,086
10 Social care and social insurance	5,001,392,765	5,890,342,755
11 Agriculture, forestry, fishery and water service	702,900,187	688,284,381
12 Environment protection and hydrometeorology	54,192,641	108,528,246
13 Industry and construction	20,257,074	40,793,833
14 Transports, roads, communications and IT	218,934,443	594,995,684
15 Utilities and housing	735,050,134	1,145,431,645
16 Fuel and energy	595,424,852	569,903,481
17 State debt servicing	470,555,256	428,592,896
18 Restoring the state reserves	59,144,009	83,703,482

At the top of the data table area, there are buttons: 'Select Fields', 'Expand All', 'Collapse All', 'Download to CSV', and 'Help'.

²⁴ The World Bank Open Budgets Portal: <http://wbi.worldbank.org/boost/>

3.3. Open Spending Budget data package

Open Spending has launched an initiative to develop a light-weight format for budget data that assures understandability and usability: the *budget data package*. The data model of the budget data package prescribes the following attributes for the metadata of budget resources [16].

Table 8: Obligatory attributes of the budget resources in the *budget data package*

Attribute	Type	Description
currency	string	the currency of items in the data; value is an ISO 4217 currency code
dateLastUpdated	date	the date when the dataset was last updated
datePublished	date	the date when the dataset was published
fiscalYear	date	the fiscal year represented by the dataset
granularity	string	the level of disaggregation in the data; value is one of "aggregated" or "transactional"
standard	string	the version of the budget data package specification used by the budget data package
status	string	the stage in the budget cycle represented by the data in the budget data package; value may be "proposed", "approved", "adjusted", or "executed"
type	string	the type of data represented by the resource; value is one of "expenditure" or "revenue"

Additionally, each metadata object should include, where relevant:

Table 9: Recommended attributes of the budget resources in the *budget data package*

Attribute	Type	Description
location	string	the two-letter country code (ISO 3166-1 alpha-2) associated with the budget data package

3.4. Conclusion

The analysis of the state of government budget dissemination shows that most of the countries publish their national budgets in a machine-readable format, usually that being XML and csv. Within the European Union, none of the Member States make use of URIs to identify concepts of the budget, nor do they link their data to other data to provide context. The only country that publishes its budget as linked open data is Brazil.

However, the RDF vocabulary of the Brazilian budget data only describes elements in Portuguese. Therefore, it is not possible to directly reuse elements in the RDF vocabulary of the EU budget.

Even though no reusable elements could be found, the study of related initiatives has identified a number of good practices for disseminating national budgets, including:

1. Provide **different categorisations** for expenses within one budget. For example, classifying expenses by public administration, by policy area, by function of government and by location;
2. Publish a **data dictionary** alongside the budget. Understanding a national budget and its specific jargon is quite challenging. Therefore, publishing a description of how the budgetary data and related concepts should be interpreted significantly contributes to the understandability and usability of the budget data;
3. In order to address different user needs, budgets should be made available in **multiple, fit-for-purpose, formats**. Both machine- as human-readable formats should be supported. Practice has shown that using linked data technologies for data publishing can address this requirement; and
4. As users are often interested in a **subset** of information, it is a good practice to allow them to only **access** the information they need, for example by providing search and querying mechanisms, and by providing the data as a service in the form of linked data.

4. STRUCTURAL ELEMENTS OF THE EU BUDGET

The EU budget is published on Eur-Lex²⁵ in PDF, CSV and XML formats. The budgetary data is structured in a document-oriented fashion using titles, chapters, items, etc. The first step in creating an RDF vocabulary would therefore be to create a resource-oriented data model. In this chapter, the existing elements which are specified in the EU budget are explained. In chapter 4, a conceptual data model is designed, based on which an RDF vocabulary is developed in chapter 6.

4.1. Levels of the EU budget

The EU budget which is currently published on Eur-Lex is structured following six levels of granularity: section, expenditure, title, chapter, article and item. Table 10 provides an example of how these levels are expressed in XML.

Table 10: levels of the EU budget

Levels	Examples
Section	"European Parliament" Sections are not expressed within an XML file. For each section, a separate XML file is published.
Expenditure	
Title	"Persons working with the Institution"
Chapter	"Members of the Institution"
Article	"Salaries and allowances"
Item	"Salaries"
XML	
<pre><?xml version="1.0" encoding="UTF-8"?> <nmc-expenditure id="N10C4B" alias="E"> <bud-heading> <p>EXPENDITURE</p> </bud-heading> <bud-data tot="tot" exprev="exp"/> <nmc-title id="N10C54" alias="1"> <bud-heading> <p lang="en">PERSONS WORKING WITH THE INSTITUTION</p> </bud-heading> <bud-data tot="tot" exprev="exp"/> <nmc-chapter id="N10C5E" alias="1 0"> <bud-heading> <p>MEMBERS OF THE INSTITUTION</p> </bud-heading> <bud-data tot="tot" exprev="exp"/> <nmc-article id="N10C67" alias="1 0 0"> <bud-heading> <p lang="en">Salaries and allowances</p> </bud-heading> <bud-data tot="tot" exprev="exp"/></pre>	

²⁵ The official EU Budget is available via <http://eur-lex.europa.eu/budget/www/index-en.htm>

```

<nmc-item id="N10C71" alias="1 0 0 0">
<bud-heading>
<p lang="en">Salaries</p>
</bud-heading>
<bud-data exprev="exp">
<amounts>
<amount year="n" catpol="5.2.11">
<figure>73 643 709</figure>
</amount>
<amount year="nm1" catpol="5.2.11">
<figure>71 393 074</figure>
<reserve>
<figure>1 216 926</figure>
<alias>10 0</alias>
<total>
<figure>72 610 000</figure>
    
```

4.2. Conceptual elements of the EU budget

The levels that are listed in Table 10, such as titles and chapters, are used to provide a hierarchical structure to the EU budget. These structural elements, however, do not provide sufficient input for creating a data-oriented view to the EU budget, i.e. to define classes, properties and relationships that provide meaning to the budget data. Nevertheless, the elements that give meaning to budget data can be derived based on the published XML files and based on a study on the XML structure of CIBA, the application used by the Publications Office to import, structure and publish the EU budget [17]. In order to validate XML files, CIBA makes use of Data Type Definition (DTD) files. For the purpose of identifying XML elements which are relevant for the creation of a dissemination vocabulary of the EU budget, we based ourselves on the analysis of the *abb.dtd* file, which describes the *Activity Based Budget* structure. The XML elements of the EU budget that were identified based on the analysis of the *abb.dtd* file, their definition and allowed sub-elements are listed in Table 11. Furthermore, the first column of the table indicates for each XML element whether it is used in the published XML files of the EU budget²⁶.

Table 11: elements of the EU budget

✓ Element used in published XML files ✗ Element not used in published XML files

	XML element	Definition	Sub-elements
✗	<abb>	This is root element and defines ABB document type. ABB element does not allow any attributes.	<metadata> <nmc-section> <nmc-sectpart> <nmc-expenditure> <nmc-revenue> <nmc-title> <nmc-chapter> <nmc-article> <nmc-item> <nmc-subitem> <nmc-annex> <nmc-grsec>

²⁶ The analysis was performed on the published XML files of the 2015 budget: <http://eur-lex.europa.eu/budget/www/index-en.htm>

			<bud-data> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-res-cond> <bud-text>
X	<nmc-section>	It is the highest element in the budget structure. Sections correspond to EU institutions. Section 3, for example, describes the budget of the European Commission. Attributes: refer to Table 12	<nmc-expenditure> <nmc-revenue> <nmc-sectpart> <nmc-annex> <nmc-grsec> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference>
X	<nmc-sectpart>	Part of section nomenclature. Attributes: refer to Table 12	<nmc-expenditure> <nmc-revenue> <reuse-link> <nmc-annex> <nmc-grsec> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference>
✓	<nmc-expenditure>	This element includes all the expenditures of particular section. Attributes: refer to Table 12	<nmc-title> <pos-expenditure> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-data> <reuse-link>
X	<nmc-revenue>	This node includes all the revenues of particular section. Attributes: refer to Table 12	<nmc-title> <pos-revenue> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-data> <reuse-link>
✓	<nmc-title>	Nomenclature items are used to structure the budget following titles, chapters, articles, items, etc. as described in section 4.2. Attributes: refer to Table 12	Lower-level <nmc-> elements <pos-item> <bud-heading> <bud-intro>

			<p><bud-remark> <bud-legal> <bud-reference> <bud-data> <reuse-link></p>
✓	<nmc-chapter>		<p><nmc-article> <pos-chapter> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-data> <reuse-link></p>
✓	<nmc-article>		<p><nmc-item> <pos-article> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-data> <reuse-link></p>
✓	<nmc-item>		<p><nmc-subitem> <pos-item> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <bud-res-cond> <bud-data> <reuse-link></p>
✓	<nmc-subitem>		<p><pos-subitem> <bud-heading> <bud-intro> <bud-data> <reuse-link></p>
X	<nmc-annex>		<p><nmc-annex> <nmc-grsec> <nmc-revenue> <nmc-expenditure> <pos-annex> <bud-heading> <bud-intro> <bud-remark> <bud-legal> <bud-reference> <reuse-link></p>

X	<nmc-grsec>	This element stores information used to generate staff tables.	<pos-grsec> <bud-heading> <bud-text> <nmc-grsec>
✓	<bud-data>	This element allows storing content related to amounts committed or paid to particular budget part, human resources or relations between budgetary lines. Attributes: <ul style="list-style-type: none"> • exprev – if these values are revenues or expenditures • tot – total if this attribute is not set for a <bud-data> element, it means that the data element is specified at its lowest level of granularity, i.e. that it does not represent the result of a summation (“total”) of sub-elements. <ul style="list-style-type: none"> • type – horizontal/reserve 	<amounts> <hresources> <relations> <epnote>
✓	<bud-heading>	This element allows setting a heading. The only sub-element, which is allowed, is <p> element (paragraph). It is possible to set translate attribute to specify text language.	<p>
X	<bud-intro>	Introduction of titles.	<p> <reuse-link> <list> <table>
✓	<bud-remark>	This element allows defining remark on budget line.	<p> <reuse-link> <list> <table>
✓	<bud-legal>	This element holds information about legal basis of this budget line.	<p> <reuse-link> <list> <table>
✓	<bud-reference>	This element holds information about references of concrete budget line.	<p> <reuse-link> <list> <table>
✓	<bud-res-cond>	This element allows defining conditions under which reserve can be released.	<p> <reuse-link> <list> <table>

X	<bud-text>	This element allows defining budgetary data in basic ²⁷ structure. Classical budget-line structure is not supported. Typical usage is in "staff" part of section.	<p> <reuse-link> <list> <table>
✓	<amounts>	This element only wraps "amount" elements.	<amount>
✓	<amount>	<p>Element that defines amount of money received/spent for related budget line. Attributes:</p> <ul style="list-style-type: none"> • aele – possible values are true/false. The AELE (European Free Trade Association) is an intergovernmental organisation set up for the promotion of free trade and economic integration to the benefit of its four Member States. • peco – possible values are true/false. PECO (Pays d'Europe Centrale et Orientale) is the acronym for countries of central and eastern Europe in French. • catpol – reference to MFF code. The MFF (Multiannual Financial Framework) lays down maximum amounts ('ceilings') for each broad category of expenditure ('MFF headings') for a clearly determined period of time (several years). • year – specifies year for which amount is located. Everytime, three last years are managed, „n“ stands for current year, „nm1“ is previous year and „nm2“ is 2 years ago. If currently prepared budget is for year 2014, then „n“ stands for 2014, „nm1“ is 2013 and „nm2“ is 2012. • computed – possible values are true/false. Computed value. • comp – possible values are true/false. • delegation – possible values are true/false. 	<figure> <reserve>
✓	<figure>	Figure element holds information about money. Attribute commpay – possible values are comm/pay. It defines that	None

²⁷ Basic structure means support only for paragraph or reuse definitions, or item list, or table definitions.

		<p>this figure is commitment or payment. Depending on parent element, commitment, payment, commitment reserve, payment reserve can be found.</p>	
✓	<reserve>	<p>This amount of money is released only when reserve conditions are met.</p>	<p><figure> <alias> <total></p>
✓	<alias>	<p>This element can be found in reserve element and holds a reference to a budget line of "reserve" type, i.e. the destination line. These destination reserve lines hold the sum of all reserves referenced to a particular destination line. The reserve destination lines are thus repeating information which has already been stated in the <reserve> class.</p>	<p>None</p>
✓	<total>	<p>Total element is used to hold sum of commitment + commitment reserve or payment + payment reserve.</p>	<p><figure></p>
✓	<relations>	<p>The <i>relations</i> element is used within the CIBA system to set relations between budget lines. The element is used to indicate which expense figures need to be stated in another location within the budget. Relations are of several types:</p> <ul style="list-style-type: none"> • The admin relations help generating the tables for Title XX: "Administrative expenditure allocated to policy areas". • The aliases specified in traditional relations are displayed in the summary tables of the chapters. These summary tables are automatically generated by CIBA. • Admin support: not used anymore. • Policy is an information related to the Human resources. • Aggregated this type will be applied to lines that group all the expenditures related to an office (for instance: line "26 01 09 01" groups all the expenditures of the Publications Office). • Detailed: this type will be applied to lines of which the sum will be retrieved in the aggregated ones (for instance: all the expenditures lines of the Annex 2 are the details of the aggregated one "26 01 09 01"). 	

		The field PP ("pour partie") is the amount allocated to the relation (optional). Only figures are allowed in that field. The field Value will be used to insert the alias of the concerned line. Setting a relation between lines produces an update transaction (RelationUpdate).	
✓	<hresources>	This element wraps all human resources.	<bud-data> <old.hresources> <new.hresources>
✓	<hresource>	This element allows defining human resources released for particular part, meaning how many people is allocated. Attributes: <ul style="list-style-type: none"> • type • year 	
✓	<epnote>	Some type of note.	<p>

All nomenclature (nmc) items listed above share a set of common attributes, as listed in Table 12.

Table 12: Nomenclature attributes

	Attribute	Description
✓	id	identifier of individual budget node
✓	alias	short node locator within the structure (e.g.: SEC3)
✓	programs-link	Concrete nomenclature can be associated with programs. Then reuse which counts over all lines related to given programs can be defined. Multiple programs codes can be defined, they are separated by " " (pipe). The program codes are listed in the "Program" code list, which is published as part of the metadata of the EU budget XML files on Eur-Lex.
✓	agency-code	Separated entities within the European Union with a precise mission. A number of EU agencies have been established, with two main types: <ul style="list-style-type: none"> • executive agencies: organisations established in accordance with Council Regulation (EC) No 58/2003 (OJ L 11, 16.1.2003) with a view to being entrusted with certain tasks relating to the management of one or more Community programmes. These agencies are set up for a fixed period. Their location has to be at

		<p>the seat of the European Commission (Brussels or Luxembourg).</p> <ul style="list-style-type: none"> • decentralised agencies: body governed by European public law; it is distinct from the Community Institutions (Council, Parliament, Commission, etc.) and has its own legal personality. It is set up in order to accomplish a very specific technical, scientific or managerial task. <p>Agencies have their own budget (managed within the European Commission budget - Volume 3) and budget documents. Concrete nomenclature can be associated with each agency. The agency codes are listed in the "Agency" code list, which is published as part of the metadata of the EU budget XML files on Eur-Lex.</p>
X	office-code	A code attached to establishment plan (staff table), nmc-grseq and budget line.
✓	type	<ul style="list-style-type: none"> • aibl (Activity Including a Budget Line): lines of this type are final (without children) with associated figures • awbl (Activity Without a Budget Line): lines of this type are final (without children) without figures in a budget structure
X	nodeId	node identifier
X	isRestoreFigure	true if restore figures
X	isRestoreText	true if restore text
X	conc_tech_status	see Conciliation statuses
X	conc_poli_status	see Conciliation statuses
X	proposalBy	see Conciliation statuses
X	firstApprovalBy	see Conciliation statuses
X	secondApprovalBy	see Conciliation statuses
✓	isAdminSupport	Support for administrative stuff budget line (i.e. not an operational line), if and only if the adminSupport is true, or more relations of type adminSupport can be specified.
✓	pp-status	pilot project status to which the budget line is linked
✓	pp-stage	pilot project stage
✓	pa-status	preparatory action status to which the budget line is linked

✓	pa-stage	preparatory action stage
---	----------	--------------------------

5. IDENTIFICATION AND AGGREGATION OF CLASSES AND PROPERTIES

The elements of the EU budget XML files as described in chapter 4 consist of both classes and properties. The paragraphs below describe

- which elements should be considered as classes and which should be considered as properties;
- how similar elements can be grouped;
- which elements could be excluded from the dissemination vocabulary; and
- which code lists could be developed and maintained.

5.1. Design principles

In order to identify which elements should be incorporated in the dissemination vocabulary of the EU budget, the following design principles will be applied:

- **Represent numbers at their lowest level of granularity.** The RDF dissemination of the budget should only state data at its lowest level of granularity.
- **No duplication of information.** A specific item of expenditure will only be stated in one location of the published budget. Data will not be repeated, i.e. no summary tables will be included.
- **Accuracy, validity and completeness.** The published budget should only state data of which the validity, accuracy and completeness is guaranteed.
- **Focus on data.** The RDF budget should separate the data from its visual representation.

5.2. Excluded elements

Following the design principles for the RDF dissemination vocabulary of the EU budget, many of the XML elements identified in chapter 3 should be excluded from the data model. These elements and the justification for excluding them from the dissemination vocabulary are described in the following sections.

5.2.1. Unused elements and attributes

As indicated in Table 11, some of the XML elements that are defined in the CIBA system are not used in the dissemination of the EU budget. Stakeholders from the Publications Office of the EU confirmed that the following unused elements can be excluded:

- <abb>
- <nmc-sectpart>
- <nmc-revenue>
- <nmc-annex>
- <bud-intro>
- <bud-text>

Besides those elements which should be excluded, the following attributes were identified as obsolete:

- Within the <amount> element: *computed* and *comp* are not used;
- Within the nomenclature elements, the following attributes can be excluded: *nodeId*, *isRestoreFigure*, *isRestoreText*, *conc_tech_status*, *conc_poli_status*, *proposalBy*, *firstApprovalBy*, *secondApprovalBy* and *office-code*. Moreover, the *type* attribute is not needed because the difference between AWBL and AIBL does not need to be specified as activities without a business line (AWBL) are no longer included in the published budget.

Two of the elements that were identified as *unused* in Table 11 are not excluded from the dissemination vocabulary. In order to publish the whole budget in one file, the <nmc-section> elements should be used to group information per EU institution. Second, the staff tables are currently not published as part of the XML files. If the staff numbers need to be included in the RDF version of the budget, the <nmc-grsec> element should be included in the data model.

5.2.2. Excluded element: Data

The Data element stores content related to amounts committed or paid to a particular budget part, human resources or relations between budgetary lines. A design principle for the dissemination vocabulary is to only specify expenses at their lowest level of granularity. As a consequence,

- each expense will only appear once in the budget; and
- values resulting from calculations with amounts will not be included in the published RDF budget.

This design principle allows users of the data to unambiguously interpret the budget information and avoids the risk that expense items are counted multiple times when analysing the data.

Following the design principle, the data elements and its attribute are irrelevant for the EU budget dissemination vocabulary, as justified in Table 13.

Table 13: Justification for excluding the <data> attributes

Properties	Justification for exclusion
Exprev	The <i>exprev</i> attribute indicates whether the stated figures describe expenditures or revenues. As the published budget only describes expenditures, the <i>exprev</i> attribute is superfluous.
Tot	The <i>tot</i> element is used to hold sum of commitment + commitment reserve or payment + payment reserve. As figures are only defined at the lowest level of granularity, the <i>tot</i> attribute will never appear in the RDF dissemination of the EU budget.
Type	The data <i>type</i> attribute distinguishes "horizontal" data elements from "reserve" data items. Reserve figures are always stated twice in the budget: 1. within the appropriate article, chapter and title and once: the <reserve> element applies; and

2. within the destination line, i.e. copying the numbers that have already been stated in <reserve> elements: the <data> element type "reserve" is used.

In the RDF dissemination of the EU budget, only the first reserve statement will be included.

The "horizontal" type defines whether a data element is already stated elsewhere in the budget. As each expenditure will only be included once in the RDF dissemination of the budget, there is no need to identify horizontal data types.

In order to avoid double statement of figures in the RDF dissemination of the EU budget, data elements with the attributes *tot="tot"*, *type="reserve"* and *type="horizontal"* will be excluded from the budget. The exclusion of these elements makes the <data> element obsolete.

5.2.3. Excluded element: HResources

This class wraps up all human resources, i.e. it aggregates all instances of "HResource". The HResources element will be excluded from the dissemination vocabulary for two reasons:

1. The *HResources* item only serves as a container for *HResource* elements. Therefore, the *HResources* element has no purpose in the RDF vocabulary; and
2. The *HResource* element will be excluded from the RDF vocabulary, therefore the *HResources* element has no purpose.

5.2.4. Excluded element: HResource

Information on Human Resources can be stated on two locations in the EU budget:

1. Within the budget nomenclature:
The <HResource> element is used to specify the amount of human resources attributed to a certain expenditure item or article.
2. In staff tables:
The <grsec> element is used to describe the amount of human resources in staff tables. These staff tables are not coupled to nomenclature elements.

In light of creating a conceptual data model for a dissemination vocabulary, the first method of describing human resources would provide the most valuable and linkable information. Stakeholders from the Publications Office of the EU stated that the accuracy of these numbers cannot be guaranteed. Moreover, the <hresource> numbers as described in the published budget XML files are not included in the PDF and CSV files of the EU budget. The staff numbers as described in the <grsec> elements are updated before publishing the budget. Therefore, the <grsec> staff information is more reliable than <hresource> staff information.

Since guaranteeing data accuracy and data validity is a key design principle for the dissemination vocabulary, the information stated in the <hresource> element will be excluded from the data model.

5.2.5. Excluded element: Amounts

The *amounts* element only serves as a container for *amount* elements. Therefore, the *amounts* element has no purpose in the RDF vocabulary.

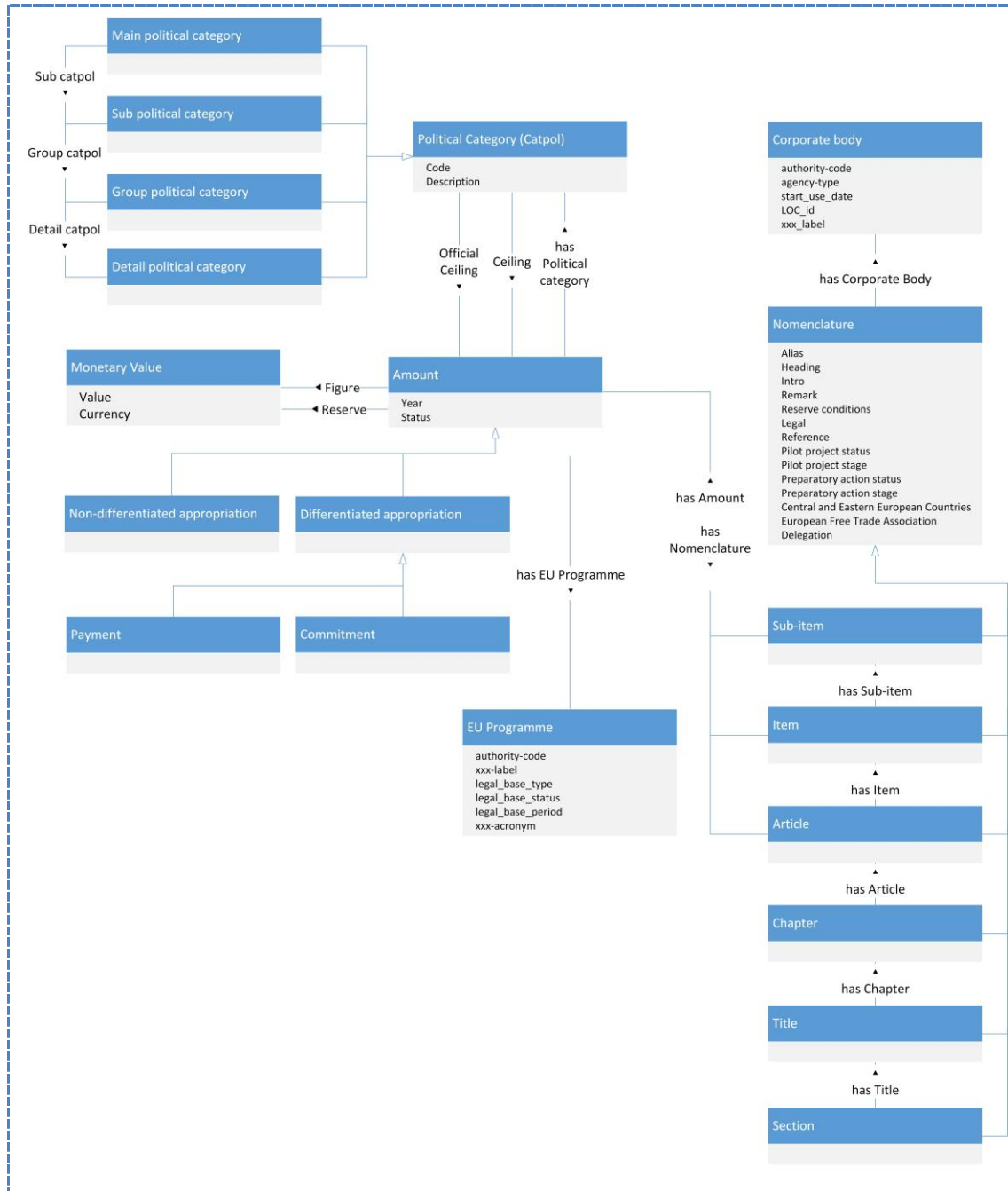
5.2.6. Excluded element: Relations

The *relations* element is used within the CIBA system to set relations between budget lines. The element is used to indicate which expense figures need to be stated in another location within the budget. For example, the tables of chapter XX are populated by figures that are originally stated elsewhere in the budget. Since the RDF dissemination of the budget will state each figure only once, the *relations* element is not needed. As a consequence, chapter XX will not be described in the RDF dissemination of the budget.

5.3. Conceptual data model of the EU budget

Figure 11 represents the conceptual data model of the EU budget as a UML diagram. The classes and properties are explained in the following sections.

Figure 11: conceptual data model of the EU budget



5.3.1. Class: Amount

The *amount* class represents amounts of money. The EU budget follows the principle of annuality, which means that amounts always relate to a given financial year. Nevertheless, many of the European Union’s operations run over multiple years, in

which case the notion of differentiation is used. Differentiated appropriations are split into commitment appropriations and payment appropriations [18]:

- Commitments are legal pledges to provide finance, provided that certain conditions are fulfilled [19]. “Commitments” are described as real budgetary decisions, whereas “payments” refer to the money that the EU actually expects to pay out in a given year [20]; and
- Payments are budgeted cash or bank transfers to the beneficiaries.

Moreover, if amounts are only released when specific conditions are met, they are called *reserves*. As shown in Table 11, a separate XML element is defined for describing the reserve conditions: *bud-res-cond*. Since reserve conditions are specified at the level of nomenclature, the *reserve conditions* element is modelled as a property of the *nomenclature* class.

Table 14: Amount properties

Class: Amount		
Properties	Definition / values	Data type
Year	Specifies the year for which underlying figures are accounted.	Date
Status	Specifies the status of the amount which is accounted for: <ul style="list-style-type: none"> • Adopted: final adopted budget. Amounts from the current year usually represent the adopted budget; • Amended: final adopted budget, including post-adoption amendments. Amounts from the previous year usually represent the amended budget; and • Outturn: amounts from two years ago usually represent the actual outturn of the budget. 	Values from a NAL ²⁸ : <ul style="list-style-type: none"> - Adopted - Amended - Outturn
Has Nomenclature	Property specifying lowest level of nomenclature from the nomenclature taxonomy (to be created). The nomenclature taxonomy can change from year to year.	Values from a controlled vocabulary.
Has Political Category	Reference to the applicable political category from the MFF. Reference should be made to the lowest suitable level in the catpol taxonomy.	Values from a controlled vocabulary
Has Programme EU	Recorded amounts can be associated with programs. The property can be repeated, multiple programs codes can be defined per amount.	Values from the EU Programmes NAL
Figure	<i>Figures</i> hold information about monetary values that are recorded on amounts.	Monetary Value

²⁸ Named Authority List: <http://publications.europa.eu/mdr/authority/>
27/10/2015

Reserve	<i>Reserves</i> hold information about monetary values that are only released when reserve conditions are met.	Monetary Value
---------	--	----------------

Table 15 gives an overview of the distinct subclasses under the *amount* class. These classes inherit the properties of the *amount* class.

Table 15: Amount subclasses

Class: Amount		
Sub-class		Definition / values
Differentiated appropriation		Many of the European Union’s operations run over multiple years, in which case the notion of differentiation is used. Differentiated appropriations are split into commitment appropriations and payment appropriations [18].
Non-differentiated appropriation		Appropriations that are not differentiated, i.e. these are not split into payments and commitments.
Commitment, subclass of <i>differentiated appropriation</i> .	a of	Commitments are legal pledges to provide finance, provided that certain conditions are fulfilled [19]. “Commitments” are described as real budgetary decisions, whereas “payments” refers to the money that the EU actually expects to pay out in a given year [20].
Payment, subclass of <i>differentiated appropriation</i> .	a of	Payments are budgeted cash or bank transfers to the beneficiaries.

The table below gives an example a *commitment* class, based on the European Commission budget of 2015.

Table 16: example of Commitment

Example: Commitment instance	
Id ²⁹	COM_2.5.11_15_Comm
Year	2015
Status	Adopted
Figure ³⁰	COM_FC_15_020511
Reserve	N/A
Has Corporate body	COM

²⁹ In the examples, identifiers are fictitious. In practice, classes will be identified using Unique Resource Identifiers (URIs).

³⁰ In practice, figures will not get unique identifiers assigned. In the RDF data set, identifiers for figures are unnecessary.

Has Political category	1.1.DAG
Has EU Programme	N/A
Article	2.5.11

The table below gives an example of two *commitment* instances to which reserves are attributed based on the European Commission budget of 2015.

Table 17: example of *Commitment* with reserve

Properties	Values
Id ³¹	CJ_15_1 0 0 0_Comm
Year	2015
Status	Adopted
Item	1.0.0.0
Has Corporate body	CJ
Has Political category	5.2.41
Has EU Programme	N/A
Figure ³²	CJ_FC_15_020511 The <i>monetary value</i> instance linked to this amount specifies a value of 23.916.500 Euro.
Reserve	CJ_RC_15_020511 The <i>monetary value</i> instance linked to this amount specifies a value of 2.000.000 Euro.

The example of *reserve* statements as described above does not relate to the earlier example of *commitment* in Table 16, since the example had no reserve linked to it. The example above is a reserve statement categorised under nomenclature item 1.0.0.0 of the Court of Justice: "Remunerations and other entitlements". The *reserve conditions* property of that *nomenclature* instance describes the conditions under which the committed value of 2.000.000 will be released: "*Appropriations entered in the reserve may be released when the European Parliament and the Council have reached an agreement on the basis of the decision to amend the Protocol on the Statute of the Court of Justice of the European Union by increasing the number of judges at the General Court by nine*".

³¹ In the examples, identifiers are fictitious. In practice, classes will be identified using Unique Resource Identifiers (URIs).

³² In practice, figures will not get unique identifiers assigned. In the RDF data set, identifiers for figures are unnecessary.

5.3.2. Class: Monetary Value

The *monetary value* class expresses the values of figures in the budget. The class is used to assign a value to *amount* figures and reserves, as explained in section 5.3.1, and to spending ceilings for *political categories* as explained in section 5.3.3.

Table 18: Monetary Value classes and properties

Class: Figure		
Properties	Definition / values	Data type
Currency	The currency in which the monetary values are expressed.	Values from the Currencies NAL
Value	Number expressing the monetary value.	Number

The table below gives an example of four *monetary value* instances that are linked to the *commitment* instance given in Table 16.

Table 19: example of Monetary Value

Properties	Values
Id ³³	COM_FC_15_020511
Currency	http://publications.europa.eu/resource/authority/currency/EUR
Value	26 791 000

5.3.3. Class: Political Category (Catpol)

The MFF (Multiannual Financial Framework³⁴) lays down maximum amounts, *ceilings*, for each broad category of expenditure for a clearly determined period of time. These categories of expenditure are referred to as "MFF headings". Each MFF heading refers to a political category and has a specific code: the *catpol* code. All amounts in the EU budget refer to their lowest level *catpol* code.

Table 20: Political Category properties

Class: Political Category (Catpol)		
Properties	Definition	Data type
Code	The MFF <i>catpol</i> code.	Identifier
Description	Title of the political category. This property has multilingual labels.	String

³³ In practice, figures will not get unique identifiers assigned. In the RDF data set, identifiers for figures are unnecessary.

³⁴ More information can be found on

http://ec.europa.eu/budget/explained/budg_system/fin_fwk0713/fin_fwk0713_en.cfm. (19.4.2013)

Ceiling	Ceiling for a specific year as defined in the MFF.	Amount
Official ceiling	Official ceiling for a specific year as defined in the MFF.	Amount
Sub-category	Property specifying the sub-categories of a political category (hasMainCatpol, hasSubCatpol, hasGroupCatpol, hasDetailCATpol).	Political category

Table 21 gives an overview of the distinct subclasses under the *catpol* class. These classes inherit the properties of the *catpol* class.

Table 21: *Catpol* subclasses

Class: Catpol	
Sub-class	Definition / values
Main political category	The sub-classes of the <i>catpol</i> class represent the aggregate structure of <i>catpol</i> types.
Sub political category	
Group political category	
Detail political category	

The table below gives an example of three *catpol* sub-classes, based on the metadata file that is published alongside the 2015 EU budget.

Table 22: example of *Catpol*

Example: <i>catpol</i> subclasses			
Properties	Class: main political category	Class: sub political category	Class: detail political category
Code	1	1.1	1.1.DAG
Description	Smart and inclusive growth	Competitiveness for growth and jobs	Decentralised agencies
Ceiling	The <i>ceiling</i> is expressed by instances of the <i>amount</i> class, as explained below.	The ceilings are expressed by instances of the <i>amount</i> class. These amount instances are not further described in this example.	
Official ceiling	The official ceiling is expressed by instances of the <i>amount</i> class, as explained below.	The official ceilings are expressed by instances of the <i>amount</i> class. These amount instances are not further described in this example.	
Sub-category	1.1	1.1.1	N/A
Sub-category	1.2	1.1.DAG	N/A
Sub-category	N/A

The *commitment* instances that express the ceiling for *catpol* instance 1, are described in the table below. For describing *catpol* ceilings, the following properties are not used: *catpol*, *nomenclature*, *type*, *corporate body*, *reserve* and *EU programme*. Payment ceilings, as opposed to commitment ceilings, are not included in the example, as the payment ceilings are not defined at the level of each heading. They are only defined at the overall level, covering all MFF headings. Commitment ceilings are defined at the level of each heading.

Table 23: example of commitment instances for catpol ceilings

Example: <i>commitment</i> instances for <i>catpol</i> ceilings		
Properties	Instance 1	Instance 2
Year	2015	2014
Figure	The <i>monetary value</i> instance linked to this amount specifies a value of 66 813 000 000 Euro.	The <i>monetary value</i> instance linked to this amount specifies a value of 63 973 000 000 Euro.

The official ceiling values for this example are the same as the ceiling values listed in the table above.

5.3.4. Class: *Corporate Body*

The *corporate body* class holds information on the EU bodies to which appropriated amounts are allocated.

In the budget, many references are made to EU institutions and agencies. Therefore, a class *corporate body* is included in the data model. The *corporate body* class models the following elements:

- The `<nmc-section>` element because each section refers to an EU institution, which are considered instances of the *corporate body* class.
- The *agency-code* property of *nomenclature*, as agencies are instances of the *corporate body* class.

Table 24: *Corporate body* classes and properties

Class: <i>Corporate body</i>		
Properties	Definition / values	Data type
authority-code	Acronym that uniquely identifies the corporate body, e.g. AGRI, DEVCO, etc.	Identifier
agency-type	"Decentralised", "executive", "joint undertaking" or "other".	String
Start_use_date	Creation date of the organisation	Date
LOC_id	City in which the organisation is seated. Refers to the Places NAL.	Identifier

xxx_label	Title of the corporate body. "xxx" stands for a three-character language code.	String
-----------	--	--------

The table below gives an example of two *corporate body* instances that are linked to the *commitment* instances given in Table 16.

Table 25: example of *corporate body*

Example: corporate body instances		
Properties	Instance 1	Instance 2
authority-code	COM	GSA
agency-type	N/A	Decentralised
Start_use_date	N/A	12/07/2004
LOC_id	N/A	Brussels (Belgium)
LOC_id	N/A	Prague (Czech Republic)
ENG_label	European Commission	European GNSS Agency (GSA)

5.3.5. Class: *EU Programme*

"EU policies are implemented through a wide range of programmes and funds which provide financial support to hundreds of thousands of beneficiaries such as farmers, students, scientists, NGOs, businesses, towns, regions and many others" [21]. The programmes of the 2014-2020 multiannual financial framework, the amounts allocated to each of them and their legal bases, when adopted, are represented by the *EU Programme* class.

Table 26: *EU Programme* classes and properties

Class: <i>EU Programme</i>		
Properties	Definition / values	Data type
authority-code	Acronym that uniquely identifies the EU program.	Identifier
xxx_label	Full title of the EU program. "xxx" stands for a three-character language code.	String
legal_base_type	Type of legal base for the program. Values: - decision - codecision	Literal
legal_base_status	Status of the legal base. Values - proposed - adopted	Literal

legal_base_period	Period for which the program runs, as defined in the legal base. Usually compliant to the period to which an MFF applies, e.g. 2009-2013 or 2014-2020.	Period of time
xxx_acronym	Short description of the program, usually in a few keywords or an acronym. "xxx" stands for a three-character language code.	String

The table below gives an example of two *EU Programme* instances as described in the metadata that is published alongside the EU budget of 2015.

Table 27: example of *EU Programme*

Example: EU programme instances		
Properties	Instance 1	Instance 2
authority-code	COPERNICUS	CFSP2020
ENG_label	European Earth Observation Programme (Copernicus)	Common foreign and security policy (CFSP)
legal_base_type	Codecision	Decision
legal_base_status	Adopted	Adopted
legal_base_period	2014-2020	2014-2020
ENG_acronym	COPERNICUS	CFSP

5.3.6. Class: *Nomenclature*

The *nomenclature* class provides contextual information about *amount* instances, such as legal information, introductory texts, references, etc.

The nomenclature elements of the EU budget as listed in Table 11 share the same sub-elements. Therefore, the nomenclature elements are grouped into one class. The nomenclature elements which share similar sub-elements and attributes include:

- nmc-section;
- nmc-title;
- nmc-chapter;
- nmc-article;
- nmc-item; and
- nmc-subitem.

The nomenclature element *nmc-grsec* – which specifies staff tables – is excluded from the list above, as fewer similarities with the other nomenclature elements could be identified. The *nmc-grsec* is modelled through the *staff* class.

The following sub-elements of nomenclature elements are considered properties: *bud-heading*, *bud-intro*, *bud-remark*, *bud-legal* and *bud-reference*. Table 28 gives an overview of the properties for the *nomenclature* class.

Table 28: Nomenclature class properties

Class: Nomenclature		
Properties	Definition	Data type
Alias	The identifier of a budget line	Identifier
Heading	This element allows setting a heading, i.e. the title of the nomenclature instance.	String
Intro	Introduction to the nomenclature instance.	String
Remark	This element allows defining remarks on a budget line.	String
Reserve conditions	The conditions that need to be met before releasing reserve figures that are recorded within the nomenclature instance. Reserve conditions can be of various natures. Therefore, they are described as free text.	String
Legal	This element holds information about legal basis of this budget line.	String
Reference	This element holds information about references to other documents, such as articles in the Official Journal of the EU and communications of the European Commission.	String
Pilot project status	Pilot project status to which the budget line is linked.	Values from a NAL: <ul style="list-style-type: none"> • New • Continuation
Pilot project stage	Pilot project stage. Pilot projects can run for two years. Pilots in their second year of implementation are supposed to turn into legal basis or preparatory actions [22].	Values from a NAL: <ul style="list-style-type: none"> • PP1: 1st year of implementation • PP2: 2nd year of implementation • CPP: completion phase of PP
Preparatory action status	Preparatory action status to which the budget line is linked.	Values from a NAL: <ul style="list-style-type: none"> • New • Continuation
Preparatory action stage	Preparatory action stage. Preparatory actions can run for three years. Preparatory actions in their third year are supposed to turn into legal basis [22].	Values from a NAL: <ul style="list-style-type: none"> • PA1: 1st year of implementation • PA2: 2nd year of implementation • PA3: 3rd year of implementation • CPA: completion phase of PA

Central and Eastern European Countries	This property indicates whether the recorded expenditure relates to a PECO country. PECO (Pays d'Europe Centrale et Orientale) is a French acronym for countries of central and eastern Europe.	Boolean
Free trade association	This property indicates whether the expense relates to contributions from the EFTA States pursuant to the Agreement on the European Economic Area.	Boolean
Delegation	The delegation property indicates whether an expenditure is related to a delegation.	Boolean
Has Corporate body	Reference to corporate bodies, such as Agencies or EU Institutions, for to which the amounts within a nomenclature instance relate. The property can be repeated, i.e. multiple corporate bodies can be linked to one nomenclature instance.	Values from the Corporate Bodies NAL

The structure and hierarchy of the budget can differ per section, i.e. there is not one nomenclature structure that applies to the whole budget. Moreover, the nomenclature can change from one year to another. Nomenclature changes are described in the general introduction of the draft general budget³⁵. Table 31 below gives an example of two *nomenclature* instances that are linked to the *commitment* instance given in Table 16.

Table 29 gives an overview of the distinct subclasses under the *nomenclature* class. These classes inherit the properties of the *catpol* class. Only the *article*, *item* and *sub-item* classes can have *amounts* linked to them as shown in Table 30: *Nomenclature* subclass specific propertyTable 30: *Nomenclature* subclass specific property. Moreover, following the design principle that numbers are only represented at their lowest level of granularity, as explained in section 5.1, *nomenclature* subclasses to which an *amount* is linked cannot have child instances. For example, if an *article* instance is linked to *amount* instance, the *article* instance cannot contain *item* instances.

Table 29: Nomenclature subclasses

Class: Nomenclature	
Sub-class	Definition / values
Section	The sub-classes of the <i>nomenclature</i> class represent the aggregate structure of <i>nomenclature</i> types.
Title	
Chapter	
Article	

³⁵ The “general introduction” of the draft general budget of 2015 is available via <http://eur-lex.europa.eu/budget/data/DB/2015/en/SEC00.pdf> as from page 358.

Item	
Sub-item	

Table 30: Nomenclature subclass specific property

Class: Nomenclature			
Sub-class	Property	Definition	Data Type
Article	Has Amount	Nomenclature is used to categorise amounts, as explained in section 5.3.1. The <i>amount</i> property is used to identify which amounts are recorded within a nomenclature instance.	Amount
Item			
Sub-item			

Table 31: example of nomenclature subclasses

Example: EU programme instances		
Properties	Class: Chapter	Class: Article
Alias	2.5	2.5.11
Heading	European satellite navigation programmes (EGNOS and Galileo)	European GNSS Agency
Intro	N/A	N/A
Remark	N/A	<p>The Union contribution to the European GNSS programmes is granted with a view to financing activities relating to:</p> <ul style="list-style-type: none"> the completion of the deployment phase of the Galileo programme, consisting of the construction, establishment, protection of the space and ground infrastructure, as well as preparatory activities for the exploitation phase including activities relating to the preparation of service provision, the exploitation phase of the Galileo programme, consisting of the management, maintenance, continuous improvement, evolution and protection of the space and ground infrastructure, the development of future generations of the system and the evolution of the services provided by the system,

Specification of the EU Budget Vocabulary

		certification and standardisation operations, provision and marketing of the services provided by the system and all other activities needed to ensure that the programme runs smoothly.
Reserve conditions	N/A	N/A
Legal	N/A	Regulation (EU) No 1285/2013 of the European Parliament and of the Council of 11 December 2013 on the implementation and exploitation of European satellite navigation systems and repealing Council Regulation (EC) No 876/2002 and Regulation (EC) No 683/2008 of the European Parliament and of the Council (OJ L 347, 20.12.2013, p. 965), and in particular Article 1(4) thereof.
Reference	N/A	N/A
Pilot project status	N/A	N/A
Pilot project stage	N/A	N/A
Preparatory act status	N/A	N/A
Preparatory act stage	N/A	N/A
Central and Eastern European Countries	N/A	N/A
Delegation	N/A	N/A
Corporate Body	COM	COM
Has Corporate body	N/A	GSA
Free trade association	True	True
Top-level	2	2.5
Sub-level	2.5.1	N/A

Amount	N/A – a chapter cannot link to an amount.	COM_2.5.11_15_Comm
Amount	N/A	COM_2.5.11_15_Pay
Amount	N/A	COM_2.5.11_14_Comm
Amount	N/A	COM_2.5.11_14_Pay
Amount	N/A	COM_2.5.11_13_Comm
Amount	N/A	COM_2.5.11_13_Pay

5.3.7. Class: Staff

Besides amounts, the EU budget also records allocated staff numbers linked to corporate bodies. Although the staff data is out of scope for the dissemination vocabulary of the EU Budget, the following tables give an indication of how they could be modelled with currently available data. Note that this class does not represent an individual staff member.

Table 32: Staff classes and properties

Class: Staff		
Properties	Definition / values	Data type
Year	The EU budget staff tables specify staff numbers for the current and the past year. The <i>year</i> property encodes the year to which a staff record relates.	Date
Staff count	The staff count property carries the actual number of staff within a certain function, type and year.	Integer
Has Corporate body	Staff records are created for specific corporate bodies. This property allows specifying to which bodies the staff records refer.	Values from the Corporate Bodies NAL
Post type	The post type indicates whether the staff count relates to permanent or temporary staff.	Values from a NAL: - Permanent - Temporary - Contractual

Table 21 gives an overview of the distinct subclasses under the *staff* class. These classes inherit the properties of the *staff* class.

Table 33: Staff subclasses

Class: Staff	
Sub-class	Definition / values
Administrator	<i>"Administrators are typically engaged in drafting policies and implementing EU law, analysing and advising. An administrator</i>

	<p><i>may find him/herself playing a key role in the EU's legislative and budgetary processes, coordinating the broad economic and other policies of the Member States, taking part in trade negotiations with non-EU countries, or representing the Institutions in international forums. Others might be inspecting the fishing fleets in the Member States, developing or managing a specific scientific research programme, or legal research and analysis for the European Court of Justice or drafting a decision of the European Ombudsman. Translators and interpreters are also recruited as administrators. We offer a very wide range of career opportunities for university graduates including administration, law, finance, economics, communication and science to name but a few.</i> [23]</p> <p>Property: AD-Grade indicates the grade of administrators. Administrator careers cover grades AD5 to AD16</p>
Assistant	<p><i>"Assistants are generally employed in an executive and technical role (administrative, financial, communication, research, policy development and implementation etc). They play an important role in the internal management of the Institutions, notably in budgetary and financial affairs, personnel work, computing, and document management and scientific laboratory work."</i> [23]</p> <p>Property: AST-Grade indicates the grade of assistants. Assistant careers cover grades AST1 to AST11.</p>

The table below gives an example of two *staff* instances as published in the European Commission budget of 2015.

Table 34: example of administrator

Example: Administrator instances		
Id	Instance 1	Instance 2
AD-Grade	AD12	AD12
Post type	Permanent	Temporary
Year	2015	2015
Staff count	20	14
Corporate body	COM	COM
Corporate Body	OLAF	OLAF

6. EU BUDGET VOCABULARY IN RDF

6.1. URI template

The RDF classes and properties minted in the RDF distribution of the EU budget vocabulary will be defined in the CELLAR namespace. Hence, the URI pattern implemented by the CELLAR will be followed.

6.2. Classes

Table 35: Classes of the RDF Data Model

Class name	Identifier
Amount	bud:Amount
Monetary Value	bud:MonetaryValue
Nomenclature	bud:Nomenclature
Political category	bud:Catpol

The staff concept, which is presented as a class in the conceptual data model, will be modelled following the RDF Data Cube Vocabulary³⁶. Therefore, *staff* will be presented as a DataCube DataSet: qb:DataSet.

6.3. Properties per Class

6.3.1. Amount [*bud:Amount*]

Table 36: RDF properties of bud:Amount

Label	RDF property	Range	Usage note	Card.
Year	dct:date	rdfs:Literal	Formatted as YYYY following W3CDTF.	1..1
Status	adms:status	skos:Concept	Indicates the type of amount which is accounted for, from a NAL (adopted, amended or outturn).	0..1
Has Nomenclature	bud:hasNomenclature, a subproperty of dct:subject	skos:Concept	Property referring to the lowest level of the nomenclature taxonomy that categorises the amount.	1..1

³⁶ RDF Data Cube Vocabulary: <http://www.w3.org/TR/vocab-data-cube/>
27/10/2015

Has Corporate body	bud:hasCorporateBody	skos:Concept	A corporate body to which the amount is allocated, taking values from the <i>Corporate Bodies</i> NAL.	1..n
Has Political category	bud:hasPoliticalCategory, a subproperty of dct:subject	bud:Catpol	Property referring to the lowest level of applicable political category from the MFF.	0..1
Has EU programme	bud:hasProgramme	skos:Concept	A programme to which the amount is allocated. Takes values from the <i>EU Programmes</i> NAL.	0..n
Figure	bud:figure	bud:MonetaryValue	A figure holds information about monetary values of amounts.	1..1
Reserve	bud:reserve	bud:MonetaryValue	A reserve holds information about monetary values of amounts that are only released under certain conditions.	0..1

All sub-classes of the *Amount* class are RDF classes:

Table 37: Sub-classes of bud:Amount

Sub-class	RDF Identifier
Differentiated Appropriation	bud:DifferentiatedAmount
Non-Differentiated Appropriation	bud:NonDifferentiatedAmount
Commitment, a subproperty of bud:DifferentiatedAmount	bud:Commitment
Payment, a subproperty of bud:DifferentiatedAmount	bud:Payment

6.3.2. Monetary Value [*bud:MonetaryValue*]

Table 38: RDF properties of bud:MonetaryValue

Label	RDF Property	Range	Usage note	Card.
Currency	cv:currency	skos:Concept	The currency in which the monetary value is expressed,	1..1

			taking values from the <i>Currencies</i> NAL.	
Value	cv:monetaryValue	rdfs:Literal (datatype xs:decimal)	Number expressing the monetary value.	1..1

6.3.3. Political Category [*bud:Catpol*]

Table 39: RDF properties of *bud:Catpol*

Label	RDF property	Range	Usage note	Card.
Code	adms:identifier	adms:Identifier	A unique identifier for a political category: the MFF catpol code.	1..1
Description	dct:description	rdfs:Literal	Title of the political category. This property has multilingual labels.	1..n
Ceiling	bud:ceiling	bud:Amount	A figure specifying the expenditure ceiling for a specific year, as specified in the MFF.	1..1
Official ceiling	bud:officialCeiling	bud:Amount	A figure specifying the expenditure's official ceiling for a specific year, as specified in the MFF.	1..1

All sub-classes of the *Political Category* class are RDF classes:

Table 40: Sub-classes of *bud:Catpol*

Label & RDF Identifier	RDF Properties
Main political category <i>bud:MainCatpol</i>	<p><i>bud:hasSubCatpol</i> indicates which <i>sub political category</i> instances relate to this main political category.</p> <ul style="list-style-type: none"> ▪ Domain: <i>bud:MainCatpol</i> ▪ Range: <i>bud:SubCatpol</i>
Sub political category <i>bud:SubCatpol</i>	<p><i>bud:hasGroupCatpol</i> indicates which <i>group political category</i> instances relate to this sub political category.</p> <ul style="list-style-type: none"> ▪ Domain: <i>bud:SubCatpol</i> ▪ Range: <i>bud:GroupCatpol</i>

Group political category bud:GroupCatpol	<p>bud:hasDetailCatpol indicates which <i>detail political category</i> instances relate to this group political category.</p> <ul style="list-style-type: none"> ▪ Domain: bud:GroupCatpol ▪ Range: bud:DetailCatpol
Detail political category bud:DetailCatpol	None.

6.3.4. Nomenclature [*bud:Nomenclature*]

Table 41: RDF properties of bud:Nomenclature

Label	RDF property	Range	Usage note	Card.
Alias	dct:identifier	rdfs:Literal	A budget line identifier.	1..1
Heading	dct:title	rdfs:Literal	A name assigned to the nomenclature. This property can have multilingual labels.	1..n
Introduction	bud:intro, subproperty of dct:description	rdfs:Literal	A budgetary introduction related to the nomenclature.	0..n
Remark	bud:remark, subproperty of dct:description	rdfs:Literal	A remark on budgetary line related to the nomenclature.	0..n
Reserve condition	bud:reserveCondition, subproperty of dct:description	rdfs:Literal	A condition on reserve amounts within the nomenclature.	0..n
Legal	dct:conformsTo	dct:Standard	A legal document related to the nomenclature.	0..n
Reference	dct:references	foaf:Document	A reference to other documents, such as articles in the Official Journal of the EU and communications of the European Commission.	0..n

Pilot project status	bud:ppStatus, subproperty adms:status	a of	skos:Concept	Pilot project status to which the budget line is linked, taking values from a NAL.	0..1
Pilot project stage	bud:ppStage, subproperty adms:status	a of	skos:Concept	Pilot project stage, taking values from a NAL.	0..1
Preparatory action status	bud:paStatus, subproperty adms:status	a of	skos:Concept	Preparatory action status to which the budget line is linked, taking values from a NAL.	0..1
Preparatory action stage	bud:paStage, subproperty adms:status	a of	skos:Concept	Preparatory action stage, taking values from a NAL.	0..1
Central and Eastern European Countries	bud:peco		rdfs:Literal (datatype: xs:boolean)	This property indicates whether the expense relates to a PECO country.	0..1
European Free Trade Association	bud:aele		rdfs:Literal (datatype: xs:boolean)	This property indicates whether the expense relates to contributions from the EFTA States pursuant to the Agreement on the European Economic Area.	0..1
Delegation	bud:delegation		rdfs:Literal (datatype: xs:boolean)	Indicates whether an expenditure is related to a delegation.	0..1

All sub-classes of the *Nomenclature* class are RDF classes:

Table 42: Sub-classes of bud:Catpol

Label & RDF identifier	RDF Properties
Section bud:Section	<ul style="list-style-type: none"> ▪ bud:hasCorporateBody indicates which corporate body is associated with the section. <ul style="list-style-type: none"> ○ Range: skos:Concept ▪ bud:hasTitle indicates which <i>title</i> instances relate to this section.

	<ul style="list-style-type: none"> ○ Domain: bud:Section ○ Range: bud>Title
Title bud>Title	<ul style="list-style-type: none"> ▪ bud:hasChapter indicates which chapter instances relate to this title. <ul style="list-style-type: none"> ○ Domain: bud>Title ○ Range: bud:Chapter
Chapter bud:Chapter	<ul style="list-style-type: none"> ▪ bud:hasArticle indicates which article instances relate to this chapter. <ul style="list-style-type: none"> ○ Domain: bud:Chapter ○ Range: bud:Article
Article bud:Article	<ul style="list-style-type: none"> ▪ bud:hasAmount indicates which amount is linked to an article, item or sub-item. <ul style="list-style-type: none"> ○ Domain: bud:Article, bud:Item, bud:SubItem ○ Range: bud:Amount ▪ bud:hasItem indicates which item instances relate to this article. <ul style="list-style-type: none"> ○ Domain: bud:Article ○ Range: bud:Item
Item bud:Item	<ul style="list-style-type: none"> ▪ bud:hasAmount indicates which amount is linked to an article, item or sub-item. <ul style="list-style-type: none"> ○ Domain: bud:Article, bud:Item, bud:SubItem ○ Range: bud:Amount ▪ bud:hasSubItem indicates which sub-item instances relate to this item. <ul style="list-style-type: none"> ○ Domain: bud:Item ○ Range: bud:SubItem
Sub-item bud:SubItem	<ul style="list-style-type: none"> ▪ bud:hasAmount indicates which amount is linked to an article, item or sub-item. <ul style="list-style-type: none"> ○ Domain: bud:Article, bud:Item, bud:SubItem ○ Range: bud:Amount

6.3.5. Staff [qb:DataSet]

The staff tables are not in scope of the EU Budget Vocabulary. However, if it is decided to include the staff tables in the future, they could be modelled following the RDF Data Cube Vocabulary³⁷. Within RDF Data Cube, actual staff count numbers would be called observations. The key components of RDF Data Cube are dimensions, attributes and measures:

- “The **dimension** components serve to identify the observations. A set of values for all the dimension components is sufficient to identify a single observation;
- The **measure** components represent the phenomenon being observed;

³⁷ RDF Data Cube Vocabulary: <http://www.w3.org/TR/vocab-data-cube/>
27/10/2015

- The **attribute** components allow us to qualify and interpret the observed value(s). They enable specification of the units of measure, any scaling factors and metadata such as the status of the observation (e.g. estimated, provisional).” [24]

6.3.5.1.Dimensions – qb:dimension

The following dimensions are used to identify staff count observations:

Table 43: Staff DataCube - dimensions

Dimension	Range	Usage note
Function	skos:Concept qb:codeList	The function and grade to which an observation relates, taking values from a NAL (AD5 to AD12 and AST1 to AST11)
Post type	skos:Concept qb:codeList	The type of post, taking values from a NAL (permanent, temporary and contractual).
Year	rdfs:Literal	Formatted as YYYY following W3CDTF.
Has Corporate body	skos:Concept qb:codeList	Values from NAL corporate bodies.

6.3.5.2.Observation – qb:Observation

Table 44: Staff DataCube - observations

Observation	Range & data type	Usage note
Staff count	Range: rdfs:Literal Datatype: xs:nonNegativeInteger	A number indicating how much staff is budgeted for a given set of dimension values.

6.4. Metadata

The metadata of the EU budget dataset, which will be published on the European Open Data Portal, will be described according to the DCAT Application Profile (DCAT-AP) for Data Portals in Europe³⁸. DCAT-AP prescribes five obligatory classes. As indicated in the first column of the tables below, each class has mandatory (M), recommended (R) and optional (O) properties.

Table 45: Metadata - dataset properties

M: Mandatory property – R: recommended property – O: optional property

Dataset		
Properties	URI	Values

³⁸ The DCAT Application Profile for Open Data Portals in Europe: https://joinup.ec.europa.eu/asset/dcat_application_profile/
27/10/2015

M	description	dct:description	This dataset comprises the data of the official published version of the general budget of the European Union for the financial year 2015
M	title	dct:title	General budget of the European Union for the financial year 2015
R	contact point	dcat:contactPoint	http://.../publ . This URI identifies a resource of type vcard:Organization
R	dataset distribution	dcat:distribution	http://.../RDF
R	dataset distribution	dcat:distribution	http://.../CSV
R	dataset distribution	dcat:distribution	http://.../PDF
R	dataset distribution	dcat:distribution	http://.../XML
R	keyword/ tag	dcat:keyword	Budget_2015
R	keyword/ tag	dcat:keyword	Expenditure
R	publisher	dct:publisher	http://publications.europa.eu/resource/authority/corporate-body/PUBL
R	theme/ category	dcat:theme, subproperty dct:subject	of http://eurovoc.europa.eu/100205
O	language	dct:language	http://publications.europa.eu/resource/authority/language/ENG
O	related resource	dct:relation	http://.../PDF
O	release date	dct:issued	2015-06-24
O	spatial/ geographical coverage	dct:spatial	http://publications.europa.eu/resource/authority/country/EUR
O	temporal coverage	dct:temporal	schema:startDate=2013 schema:endDate=2015

Table 46: Metadata - distribution properties (RDF)

M: Mandatory property – R: recommended property – O: optional property

Distribution (RDF)			
	Properties	URI	Values
M	access URL	dcat:accessURL	http://.../RDF/generalbudget2015.rdf
R	description	dct:description	This distribution contains the general EU budget of 2015, the amended EU budget

			of 2014 and the outturn budget of 2013 in RDF format.
R	format	dct:format	http://publications.europa.eu/resource/authority/file-type/RDF
R	licence	dct:license	http://ec.europa.eu/geninfo/legal_notices_en.htm
O	release date	dct:issued	2014-06-24
O	title	dct:title	General budget of the EU for 2015

6.5. Controlled vocabularies

Table 47 gives an overview of existing or proposed Named Authority Lists to which the EU Budget Vocabulary will refer.

Table 47: Named Authority Lists

RDF property	Part Class	of	Vocabulary name	Comment
bud:hasCorporatebody	bud:Amount bud:Nomenclature qb:Dimension		Corporate Bodies NAL	http://publications.europa.eu/mdr/resource/authority/corporate-body/ The NAL will include at least the following items: - authority-code - agency-type - start_use_date - LOC_id - xxx_label "xxx" represents a language code, such as "ENG".
dct:format	dcat:Distribution		File Types NAL	http://publications.europa.eu/mdr/authority/file-type/
cv:currency	bud:MonetaryValue		Currencies NAL	http://publications.europa.eu/mdr/authority/currency/
bud:hasProgramme	bud:Nomenclature		EU Programmes NAL	NAL is under development. CIBA metadata can be used as input for the development of the metadata. The table will include at least the following items: - authority-code - xxx_label - xxx_acronym - legal_base_type - legal_base_status - legal_base_period "xxx" represents a language code, such as "ENG".

adms:status	bud:Amount	Amount statuses NAL	Values: - Adopted - Amended - Outturn
bud:ppStatus bud:paStatus	bud:Nomenclature	Pilot project and preparatory action statuses NAL	Values: - New - Continuation
bud:ppStage	bud:Nomenclature	Pilot project stages NAL	Values: - PP1 - PP2 - CPP
bud:paStage ³⁹	bud:Nomenclature	Preparatory action stages NAL.	Values: - PA1 - PA2 - PA3 - CPA

Table 48 gives an overview of Named Authority Lists which could be created for staff data, which are currently out of scope for the dissemination vocabulary of the EU budget.

Table 48: Named Authority Lists for staff tables

RDF property	Part of Class	Vocabulary name	Comment
Post type qb:dimension	qb:DataSet	Staff post types NAL	Values: - Permanent - Temporary - Contractual
Function qb:dimension	qb:DataSet	Staff grades NAL	Values: - non-category - AD16 - AD15 - AD14 - AD13 - AD12 - AD11 - AD10 - AD9 - AD8 - AD7 - AD6 - AD5 - AST11 - AST10 - AST9 - AST8

³⁹ The ppStage and paStage NALs could be merged into one NAL.

			<ul style="list-style-type: none"> - AST7 - AST6 - AST5 - AST4 - AST3 - AST2 - AST1 - AST/SC 6 - AST/SC 5 - AST/SC 4 - AST/SC 3 - AST/SC 2 - AST/SC 1 - contractual IV - contractual III - contractual II - contractual I - seconded national expert
--	--	--	--

Table 49 gives an overview of controlled vocabularies which are specific to the EU Budget Data and thus cannot be managed as Named Authority Lists. The controlled vocabularies listed below will be published alongside the data of the EU budget.

Table 49: Budget-specific controlled vocabularies

RDF property	Part of Class	Vocabulary name	Comment
Nomenclature bud:hasNomenclature	bud:Amount	Nomenclature taxonomy	To be created based on the EU budget metadata.
Political Category dct:subject	bud:Amount	Political categories	To be created base on the EU budget metadata

7. EU BUDGET VOCABULARY IN OWL

In order to implement the EU Budget Vocabulary into the CELLAR system of the Publications Office, the data model should be expressed in OWL DL. However, in order to support the use cases and requirements as defined in section 2, the RDF format is the most suitable. Therefore, it is recommended to use the RDF vocabulary for dissemination purposes while the OWL vocabulary can be applied internally. The conceptual data model as described in section 5.3 serves as the common link between the RDF and the OWL vocabulary.

In order to transform the EU Budget Vocabulary into an OWL vocabulary, the following transformations need to be executed:

- All **classes** should be expressed as OWL classes [owl:Class]
- All **properties** should be expressed as OWL data type properties [owl:DatatypeProperty]
- All **relations** should be expressed as OWL object properties [owl:ObjectProperty]

Sub-classes of RDFS classes should be defined as distinct classes.

Naming conventions and design rules of the OWL vocabulary should be followed when expressing the EU Budget Vocabulary in OWL. The RDF representation of the RDF dissemination vocabulary is already compliant with the OWL DL design rules.

8. NEXT STEPS

The objective of this work was to create the specifications for an RDF dissemination vocabulary for the EU budget in order to achieve the following expected benefits:

- improved transparency towards citizens;
- better decision making;
- increased harmonization in budget publication;
- increased flexibility for integrating budgetary data with other datasets; and
- greater technological independence.

The primary use of the RDFS dissemination vocabulary for the EU budget is to enable the publication of budget data as linked open data. The analysis of use cases and existing initiatives for publishing budgets as LOD has identified a number of next steps which could further exploit the above-mentioned benefits.

Extending the CDM of the CELLAR with the new classes and properties of the EU budget vocabulary. In order to implement the specifications developed in this work, the classes and properties of the EU budget vocabulary will be defined under the CELLAR namespace, as an extension of the CDM. This is pre-requisite in order to allow the publishing of the EU budget data as linked data.

Publishing the EU budget as LOD. Two scenarios can be supported. First the EU budget can be made available as an RDF dataset, in the form of an RDF manifestation of the official budget published on EurLex. At a later stage, the EU budget data can be made available through the CELLAR as linked open data.

Making available the code lists linked to the EU budget vocabulary as linked open data. Besides integrating the budget with the FTS, transparency could be significantly improved by also publishing the MFF as Linked Open Data. The EU budget vocabulary takes a first step towards integrating the annual budget data with the MFF, as it provides a link to the MFF headings, i.e. "catpol codes", and spending ceilings. In order to further exploit the benefits such as increased transparency and easier data integration, the MFF could as well be published as an RDF data set, for example as a SKOS taxonomy.

The MFF taxonomy would however not be the only controlled vocabulary that needs to be created. The development of the controlled vocabularies identified in chapter 6, such as the nomenclature taxonomy or the EU Programmes Named Authority List, is crucial for the successful implementation of the EU budget vocabulary.

Linking budget and spending data for monitoring money flows. The transparency towards citizens could be significantly improved by further integrating the published EU budget with spending data, such as this published by the Financial Transparency System (FTS)⁴⁰. The FTS publishes on a yearly basis the overview of all beneficiaries of budgetary commitments. Further integrating both datasets would allow users to get a better view on the spending process from budget publication to the actual transfer of funds through public procurement, for example.

Supporting the development of pilots to demonstrate the value of budget data as LOD. Ideas for such pilots can come from the use cases identified in this work, such as the consumption of EU budget LOD in order to increase the public understanding of the budget, compare budgets, follow the money, perform data analytics and support data journalism.

⁴⁰ FTS: <http://ec.europa.eu/budget/fts/>

Harmonising the dissemination of budget data across the EU. An important next step would be to create a Core Budget Vocabulary for EU Member States, based on the EU budget vocabulary. Such a Core Vocabulary could be used by national governments to disseminate their budgets in a common way, thus fostering the harmonization of budget publication among EU Member States.

BIBLIOGRAPHY

- [1] G. Vickery, "Review of recent studies on PSI re-use and related market developments," Information Economics, Paris, 2011.
- [2] European Commission, "Communication from the Commission to the European Parliament, the Council, the European economic and Social Committee and the Committee of the Regions. Open data. An engine for innovation, growth and transparent governance.," European Commission, Brussels, 2011.
- [3] ISA Programme - European Commission, "Process and methodology for developing semantic agreements," European Commission, Brussels, 2013.
- [4] ISA Programme - European Commission, "Study on Business Models for Linked Open Government Data (BM4LOGD)," European Commission, Brussels, 2013.
- [5] J. Gray, "Open Budget Data: A Landscape Analysis," Open Knowledge Foundation, 2015.
- [6] European Commission - ISA Programme, "Operational pilot providing open data with the use of core vocabularies," European Commission, Brussels, 2014.
- [7] South Korean Ministry of Strategy and Finance, "About dBrain," 8 June 2015. [Online]. Available: [http://eng.openfiscaldata.go.kr/portal/dbrain/english/introduction.do?code=dBrain&leftCd=25&subCd=16&depCd=.](http://eng.openfiscaldata.go.kr/portal/dbrain/english/introduction.do?code=dBrain&leftCd=25&subCd=16&depCd=)
- [8] Open Knowledge Foundation, "D-Brain in Korea: Digital budgeting and accounting system," 8 June 2015. [Online]. Available: [http://community.openspending.org/research/gift/publishing/d-brain/.](http://community.openspending.org/research/gift/publishing/d-brain/)
- [9] S. Auer, *Presentation: "Creating Data Value Chains by Linking Enterprise Data"*, Riga: Fraunhofer, 2015.
- [1] J. Gray, L. Boungru and L. Chambers, "The Data Journalism Handbook," O'Reilly Media, Sebastopol, 2012.
- [1] Open Knowledge Foundation, "Government Budget," 24 April 2015. [Online]. Available: [http://index.okfn.org/dataset/budget/.](http://index.okfn.org/dataset/budget/)
- [1] Brazilian Ministry of the Federal Budget, "Modelo ontológico da Classificação das Despesas do Orçamento Federal Brasileiro," 9 June 2015. [Online]. Available: [http://vocab.e.gov.br/2013/09/loa#.](http://vocab.e.gov.br/2013/09/loa#)

- [1 I. Kheyfets, M. Mastruzzi, D. Merotto and L. Sondergaard, “A New Data Tool to
3] BOOST Public Spending Efficiency - Knowledge Brief,” The World Bank,
Washington, DC, 2011.
- [1 The World Bank, “Poland Government Expenditure Database. User's manual.,” The
4] World Bank, Washington, DC, 2013.
- [1 The World Bank, “Boost Data - Moldova,” 24 March 2015. [Online]. Available:
5] http://obtables.worldbank.org/boost_moldova/.
- [1 Open Spending, “Open Data Package | Specification,” 12 June 2015. [Online].
6] Available: [https://github.com/openspending/budget-data-
package/blob/master/specification.md](https://github.com/openspending/budget-data-package/blob/master/specification.md).
- [1 Publications Office of the European Union, “CIBA - XML structure analysis,”
7] Publications Office of the EU, Luxemburg, 2013.
- [1 Eur-Lex, “The Former Financial Regulation,” 10 07 2015. [Online]. Available:
8] <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:134014>.
- [1 European Commission, “The Budget Explained,” 8 April 2015. [Online]. Available:
9] http://ec.europa.eu/budget/explained/index_en.cfm.
- [2 European Parliament, “Background on the difference between commitments and
0] payments,” 27 10 2010. [Online]. Available:
[http://www.europarl.europa.eu/news/en/news-
room/content/20101027BKG90476/html/Background-on-the-difference-between-
commitments-and-payments](http://www.europarl.europa.eu/news/en/news-room/content/20101027BKG90476/html/Background-on-the-difference-between-commitments-and-payments).
- [2 European Commission, “Programmes,” 10 June 2015. [Online]. Available:
1] http://ec.europa.eu/budget/mff/programmes/index_en.cfm.
- [2 Committee on Budgets - European Parliaments, “Working document on pilot
2] projects and preparatory actions in budget 2012,” European Parliament, Brussels,
2011.
- [2 EPSO, “Careers with the European Union,” 4 September 2015. [Online]. Available:
3] http://europa.eu/epso/discover/types_employment/index_en.htm.
- [2 W3C, “The RDF Data Cube Vocabulary,” 16 January 2014. [Online]. Available:
4] <http://www.w3.org/TR/vocab-data-cube/#cubes-model>.
- [2 European Commission - ISA Programme, “Process and Methodology for Core
5] Vocabularies,” European Commission, Brussels, 2012.
- [2 R. Follet, How to keep score in business?, New Jersey: FT Press, 2012.
6]

ANNEXES

Annex I: Terminology

Term	Definition	XML reference
Commitment	<i>Legal pledges to provide finance, provided that certain conditions are fulfilled [19].</i>	<code><figure commpay="comm" ></code>
Payment	<i>Cash or bank transfers to the beneficiaries.</i>	<code><figure commpay="pay"></code>
Differentiated appropriations	<i>Appropriations for commitments and payments often differ because multiannual programmes and projects are usually committed in the year they are decided and are paid over the years as the implementation of the programme and project progresses. Thus, if the EU budget increases, due for example to enlargement, commitments will increase before payments do. Not all projects and programmes are concluded, and appropriations for payments are therefore lower than for commitments [19].</i>	<code><diff="true"></code>
Non-differentiated appropriations	<i>Apply for administrative expenditure, for agricultural market support and direct payments. Furthermore, not all projects and programmes are concluded, and appropriations for payments are therefore lower than for commitments [19].</i>	Not represented in the XML structure
Ceiling	<i>Limits of expenditure or revenue fixed by law or by agreement, such as in the own resources decision or in the multiannual financial framework. The latter defines an annual ceiling for each expenditure heading in commitment appropriations and an annual global ceiling for payment appropriations [19].</i>	In metadata: <code><property name="ceiling-nm1" value="123"/></code>
Multiannual Financial Framework (MFF)	<i>Multiannual spending plan that translates into financial terms the Union's policy priorities. It sets</i> <ul style="list-style-type: none"> - <i>limits on European Union expenditure over a fixed period and thus imposes budgetary discipline</i> - <i>annual maximum amounts (ceilings) of commitments for the main categories of</i> 	See "CATPOL"

	<i>expenditure (called headings) and an overall payments ceiling [19].</i>	
CATPOL FP MFF Headings	Categorie Politique: the heading of the multiannual financial framework to which the expense item belongs. Also referred to as financial programming (FP) or MFF heading.	XML: <catpol="X.Y.Z."> Metadata: <nomenc name="fp">

Annex II: URI namespace registration request

ID of the request	<TBC>
Responsible unit	EC.OP.C1 and EC.OP.C4 Publications Office - Dissemination and Reuse EC.OP.B1 Publications Office - Official Journals and Case Law
Local register	budget
Contact person	Lotte Joergensen (EC.OP.B1) Norbert Hohn (EC.OP.C1) Willem Van Gemert (EC.OP.C4)
Date of request	16/07/2015
Namespace type (opaque / mnemonic)	Opaque
Already known by third parties (evidence)	<p>“Budget” is a generally accepted term to describe expected income and expenses for a given period. A budget domain is often used by third parties such as national governments to group their budget information. Examples of such domains include:</p> <ul style="list-style-type: none"> ▪ The Australian government budget http://budget.gov.au/ ▪ The Finnish government budget http://budjetti.vm.fi/ ▪ The USA government budget https://www.whitehouse.gov/omb/budget
Requested namespace	The request is for an opaque namespace, which would initially host URIs for EU Budget Vocabulary resources, but might in the future host related resources, such as resources for spending data.
Description of resources in the local register	The namespace will initially be used to host resources related to the EU budget Metadata: classes and properties of the EU budget vocabulary, an RDF data model for the dissemination of the annual EU budget. Data: RDF distribution of the annual EU budget
Similar registered resources	There are no other EU institutions which registered resources within the domain of the EU budget.
Authoritative source (evidence)	The Council Regulation No 1605/2002 of 25 June 2002 on the Financial Regulation applicable to the general budget of the European Communities ⁴¹ states that the EU budget should be published in the Official Journal of the European Communities.

⁴¹ Decision No 1605/2002/EC, Euratom: <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002R1605&qid=1436870720332&from=EN>

ID of the request	<TBC>
	<p>Publishing the Official Journal is a key task of the Publications Office⁴². The Publications Office publishes the Official Journal, as well as other data, on Eur-Lex. The Eur-Lex section of the budget is available via http://eur-lex.europa.eu/budget/www/index-en.htm. Since the Publications Office is responsible for the publication of the budget, it should as well be responsible for the URIs of the minted classes and properties of the budget data model.</p>
<p>Commitment of persistence (evidence)</p>	
<p>Inter-organisational character (evidence)</p>	<p>The EU budget covers all EU institutions, agencies and programmes. Therefore, the budget vocabulary has a strong inter-organisational character. Moreover, one of the purposes of the budget vocabulary is to improve the accessibility and understandability of the budget by national administrations and citizens.</p>
<p>Machine-readable information (evidence)</p>	<p>The purpose of the EU budget vocabulary is to make the budget available as machine-readable information by means of an RDF datasets. Therefore, the resources identified by the URI set will provide data in RDF format.</p> <p>Moreover, the study that was conducted in light of the creation of the EU budget vocabulary pointed out that there is a strong demand to make budgetary data available in a machine-readable format.</p>
<p>Redirection rules</p>	<p>The redirection rules are to be discussed with the CELLAR team. Redirection rules should be created for the following URIs:</p> <p>http://data.europa.eu/1234/</p> <p>The URI for the collection itself will redirect to a landing page of the RDF vocabulary with an overview of all classes and properties.</p> <p>http://data.europa.eu/1234/*</p> <p>These persistent HTTP URIs for the terms in the budget vocabulary will redirect to a dynamic website containing definitions of each term in the conceptual data model of the budget vocabulary, and later potentially related terms. This website will apply content negotiation locally, i.e. so after the redirection. Temporarily, this could be replaced by a static HTML page and some anchor tags (#URLs).</p> <p>http://data.europa.eu/1234/{year}/*</p> <p>The persistent HTTP URIs for a data resource of an annual budget will redirect to an RDF expression of a particular instance in the budget.</p>

⁴² Decision No 2009/496/EC, Euroatom: <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002R1605&qid=1436870720332&from=EN>