EU FISCAL DATA

AS LINKED OPEN DATA

ISA² ACTION 2016.18:
"DEVELOPMENT OF AN OPEN DATA SERVICE, SUPPORT AND TRAINING PACKAGE IN THE AREA OF LINKED OPEN DATA, DATA VISUALISATION AND PERSISTENT IDENTIFICATION"

7 July 2017

TOPICS OF TODAY

1. **Fiscal transparency**
   Why linked open data?
2. **The journey so far**
   EU Budget Vocabulary & data publication
   FTS Vocabulary
   e-procurement ontology
3. **The next steps**

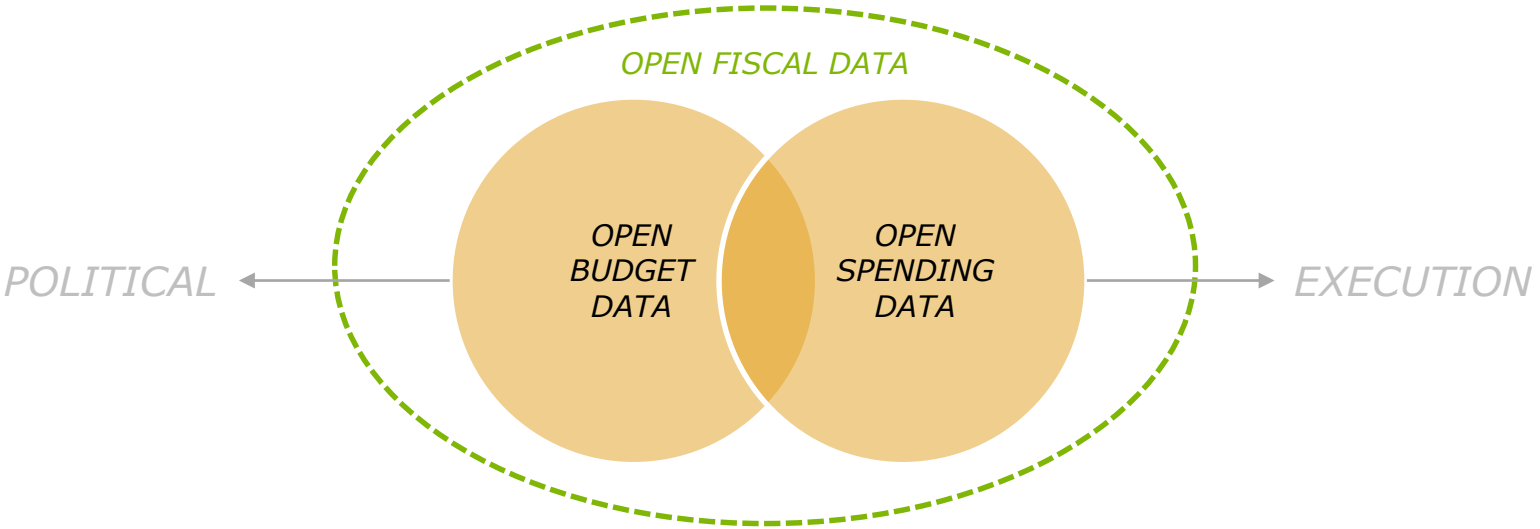
1

Fiscal transparency

Why linked open data?

OBJECTIVE: OPEN FISCAL DATA

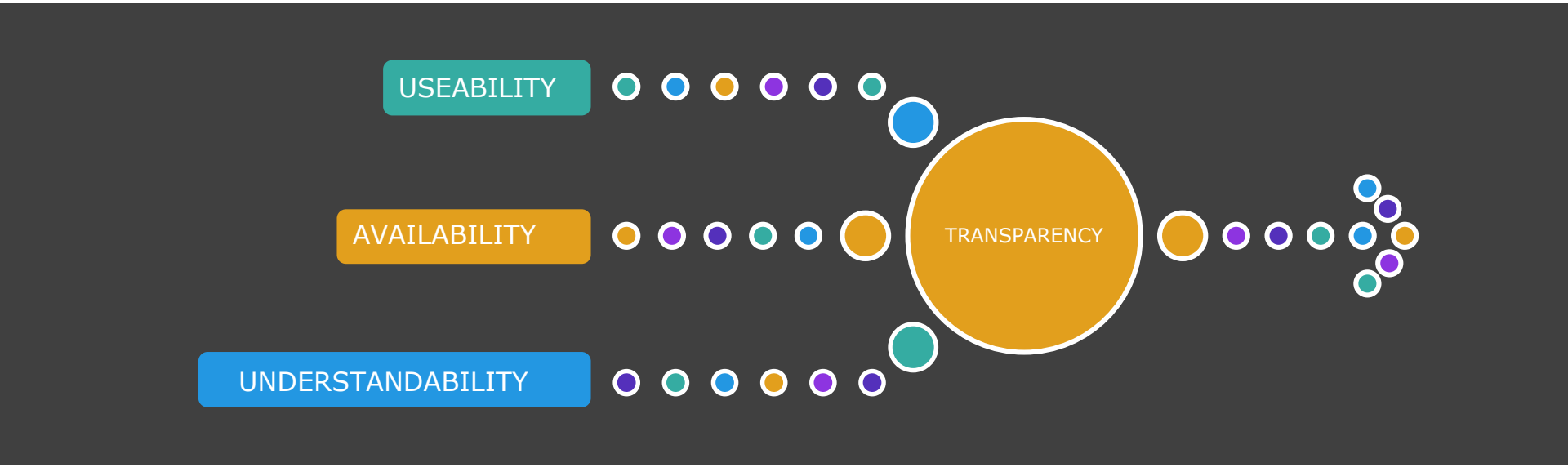
> "Open budget data is defined as the government budget data that are made accessible to the public (**online**) an editable (**machine-readable**) and reusable format, without any restriction (**free/legally open**)."
>
> - *World Bank* [source, p.3]



Source: the term "open fiscal data" as introduced in http://community.openspending.org/resources/handbook/

FISCAL TRANSPARENCY DRIVERS

To improve the **transparency** of EU fiscal data,
it should be **published** in a more **usable** and **understandable** manner.



FISCAL TRANSPARENCY: USERS & REQUIREMENTS

Who uses the data?

- Reporters
- Media
- Researchers
- Civil organisations
  Budget watchdogs, unions, issue advocates
- Private Firms
- Citizens
- Public Administrations

What do they want to do?

- **Understand** the data
- **Communicate** and build stories
- **Analyse** the data
- Develop **applications** and **services**
- Take informed **decisions**:
  corporate strategies, policy making, etc.

Source: GIFT (New York City & Chicago), TED study

FISCAL TRANSPARENCY: USERS & REQUIREMENTS

What do the users need?

- **Raw data**
  Some users need raw data, in order to perform their own analyses.
- **Data as a service:** query datasets, APIs
  Especially in big data sets, users are interested to retrieve only the information they need when they need it. APIs allow users to build applications consuming the data
- **Visualised information**
  Gain quick understanding
- **Accurate, structured data**
  Data should be provided at the right level of granularity and in the desired format. Aggregated data often doesn't allow accurate analytics.
- **Understandable information**
  Provide sufficient contextual information for users to understand the data.

Source: TED study

CHALLENGES CAN BE OVERCOME WITH LINKED OPEN DATA

CHALLENGE	SOLUTION: LOD COMES WITH...
Understandability of the data: terminology & structure are difficult to understand	**Dissemination vocabularies** & **data dictionaries** published online
Information is **scattered** across different data sets: EU Budget, MFF, FTS, etc.	**Linkable** datasets, integrating data from different sources
Low **usability** & **interoperability:** data not interpreted the same way, difficult to link and integrate.	**Common data formats**, developed and adopted to avoid misinterpretation.
Difficult to meet **different** user needs: Format, focus, granularity,...	Data as a **service**, which allows users to query & export in any format needed
Machine readability and **reasoning** are required for advanced analytics	**Data** and **ontologies** published in **machine-readable** formats

SOLUTION

MAKE EU FISCAL DATA AVAILABLE AS 5-STAR LINKED OPEN DATA

Publish data on the web ★

as structured data ★★

in a non-proprietary format ★★★

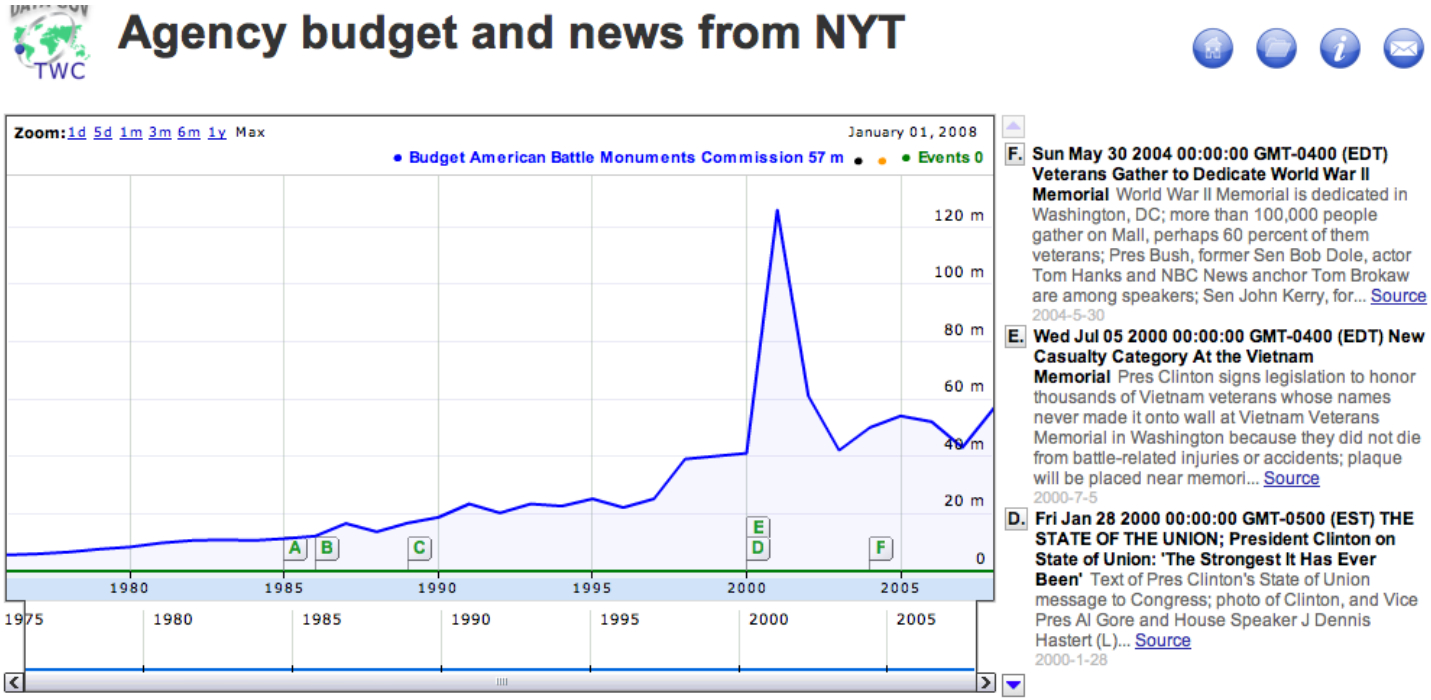
using URIs to denote things ★★★★

linking your data to other data ★★★★★

USE CASE: LINKING BUDGET DATA WITH OTHER DATA

Example: US Budget & related news items [source]

- Linking data from different sources: budget & NY Times
- Increase understanding by linking facts with insights
- Visualisation

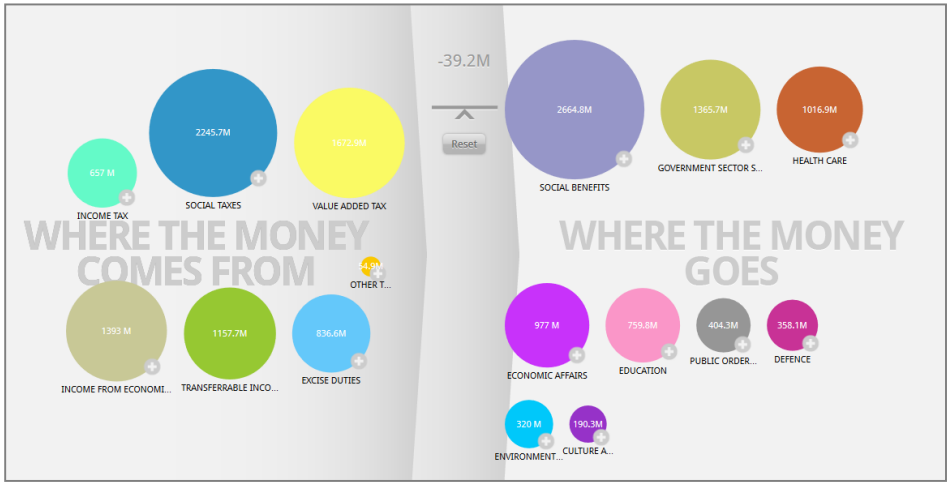


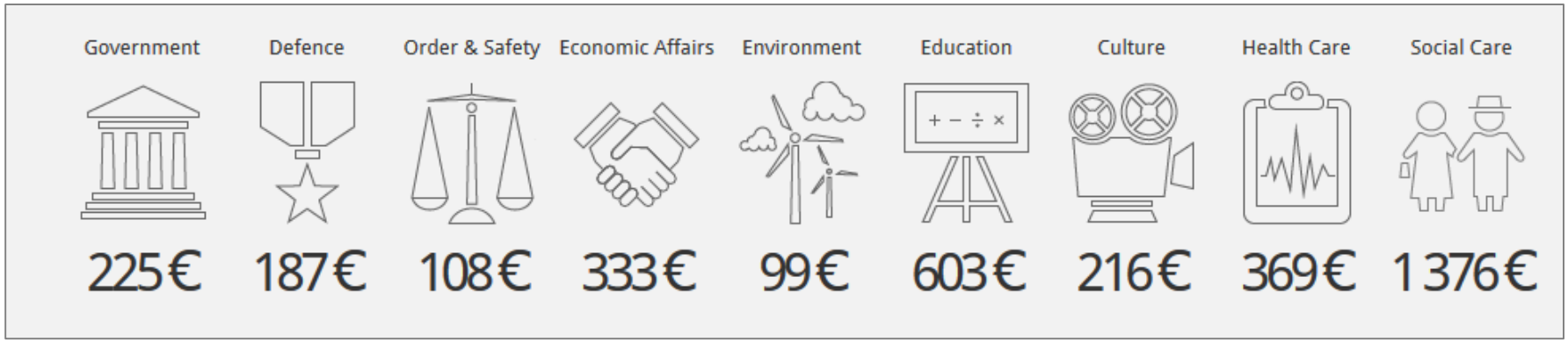
USE CASE: FOLLOW THE MONEY

Example: Estonian budget website

http://meieraha.ee/

- Visualise the income & expenditure
- Help citizens understand what happens with their tax money
- Underlying data unavailable

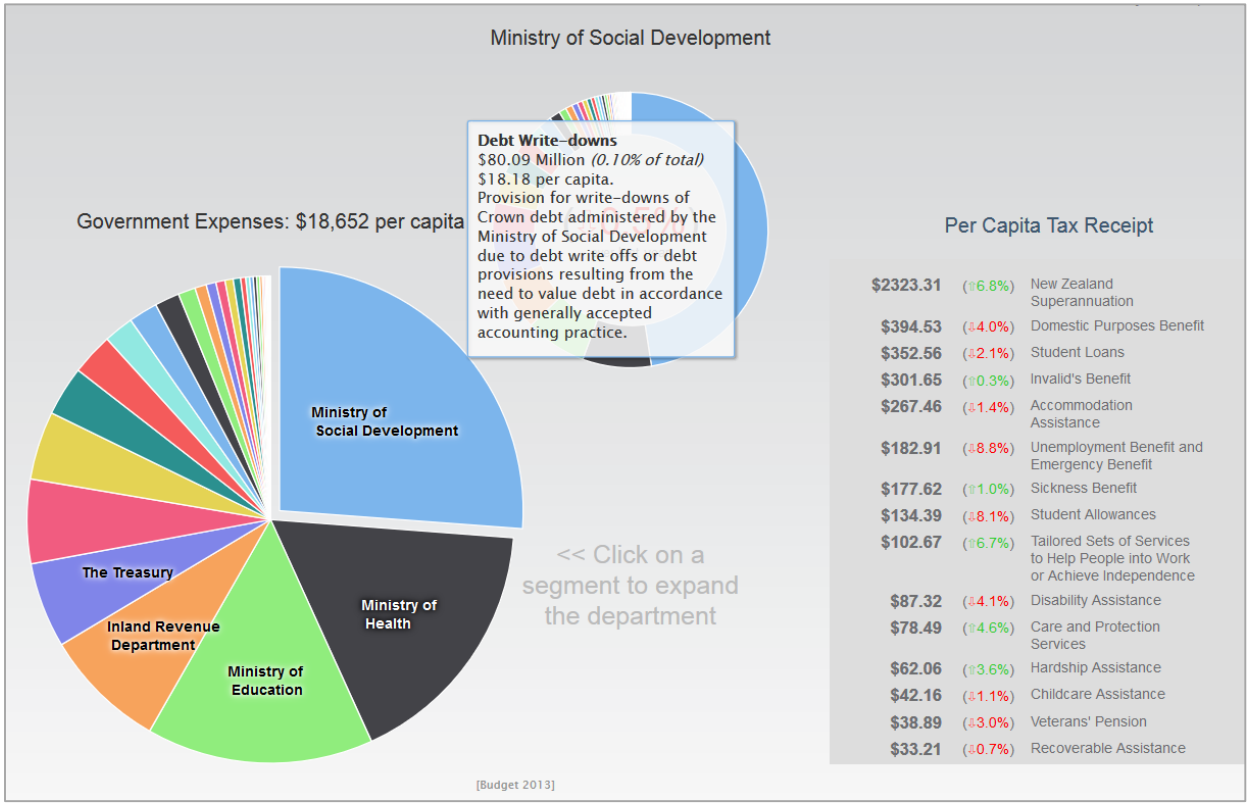




USE CASE: BUDGET ANALYTICS

Example:
New Zealand – where are my taxes?
http://wheresmytaxes.co.nz/

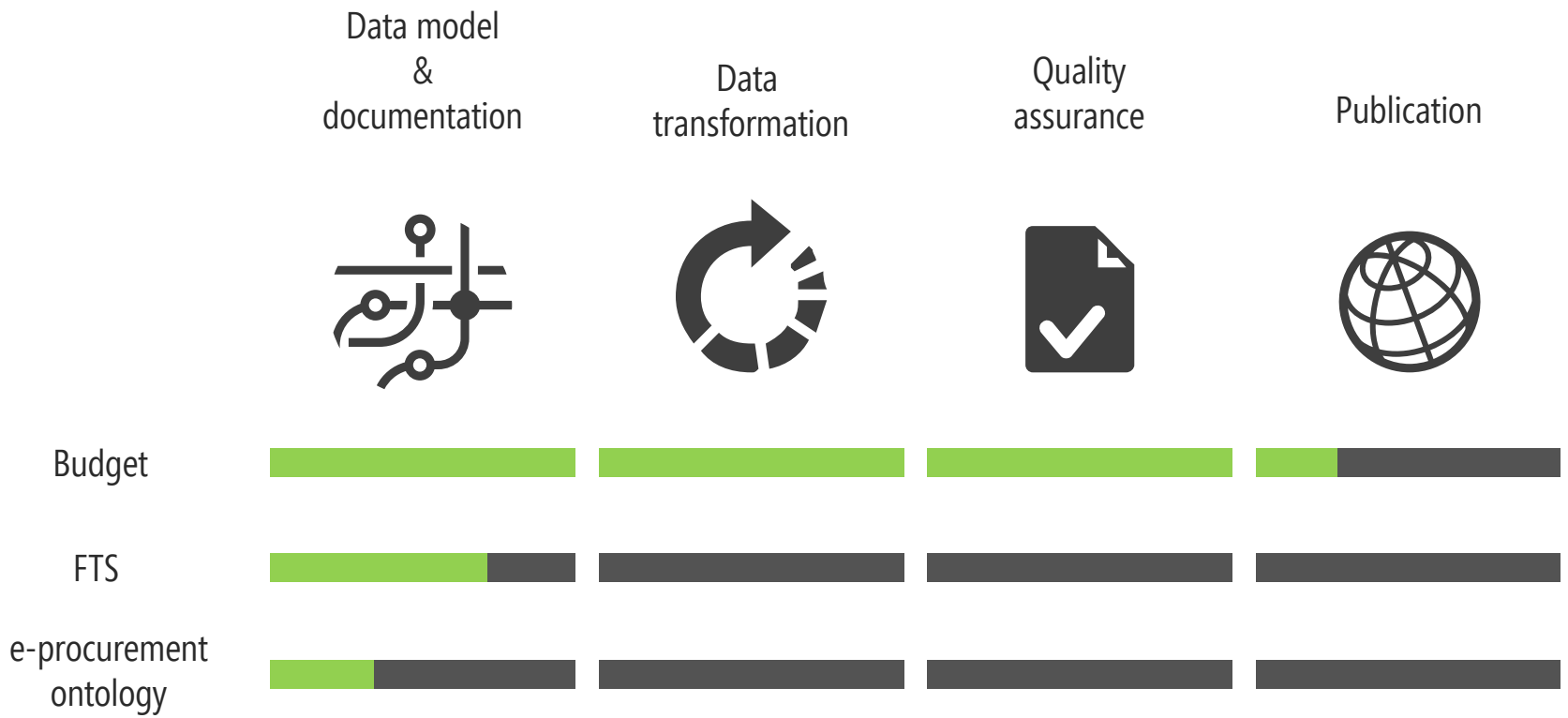
- Visualise the budget
- Categorise per functional domain
- Show changes compared to previous years



2

The journey so far

STATUS IN JULY 2017



THE EU BUDGET PUBLICATION IN 2015

PDF
2.000+ pages
Not machine readable

CSV
3-star LOD
Incomplete
Complicated structure

XML
3-star LOD
Unclear terminology
For formatting, not data-oriented

- No data model disclosed
- No use of standardised reference data, such as the Name Authority Lists
- Different file per institution (Parliament, Commission, etc.), difficult to create an integrated view
- Terminology unclear: aele, peco, nmc, catpol, etc.
- Isolated from other fiscal data sets such as FTS and e-procurement
- Data not provided as a service, no flexibility

➔ Limited reusability of the data

Published on http://eur-lex.europa.eu/budget/www/index-en.htm

THE EU BUDGET PUBLICATION IN 2015

Before we started the LOD projects, the following formats were available online:

- **PDF**:
  - Over 22.000 pages & not machine-readable.
  - Very difficult to use this format, unless the users know really well what they are looking for.
  - Impossible to do analytics & build applications
- **CSV**
  - Better, it's machine readable and an open format!
  - However: the data is not complete. It only contains the headings and the numbers. Numbers are also aggregated, so it is difficult to know what can be added up and what is already a result of a computation.
- **XML**
  - the data is complete, but the purpose of the XML is to build the PDF
  - The XML is thus not focused on enriching the data, but simply on formatting the PDF.
  - It is not really fit-for-purpose for re-use by third parties.
  - The terminology used in the XML is difficult to understand comprehensible, e.g. peco, nm, aele,... .
  - The numbers are listed at different levels of granularity
  - Even though a useful format is used, the content is not suited to do any of the activities listed before: analytics, data visualisation, etc.

LIMITATIONS OF THE 2015 PUBLICATION FORMATS

Title Chapter Article Item	Heading	FF	Appropriations 2015
02 01 05	***Support expenditure for research and innovation programmes in the 'Enterprise and industry' policy area***		
02 01 05 01	Expenditure related to officials and temporary staff implementing research and innovation programmes — Horizon 2020	1.1	10 866 475

What is the total amount spent on Horizon 2020 in the past years?

Who received this money?
Which companies? Which universities?

Under which conditions can this money be spent?

When are procured services for this expense due?

How is this spent?
Grant? Contract? Prize?

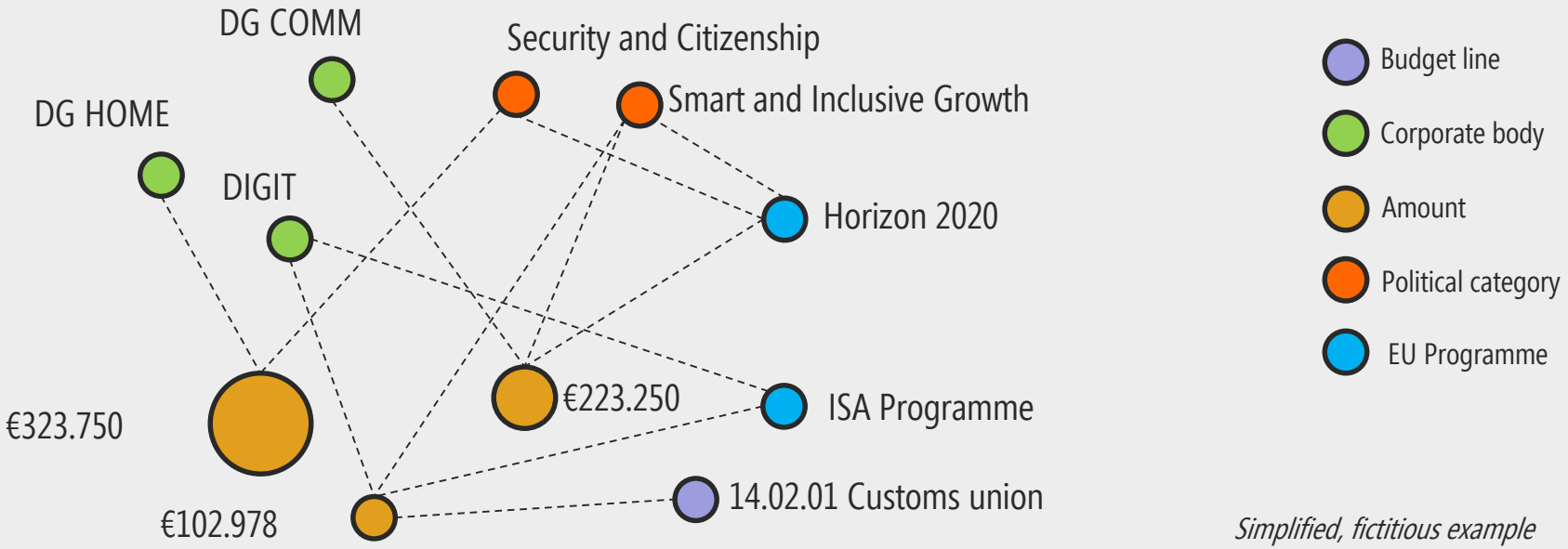
Which other expenses are allocated to this MFF heading? What is its ceiling?

How does this heading translate into Bulgarian?

In which countries are recipients located?

EU BUDGET AS LINKED OPEN DATA

- To answer these questions, we need the budget and related data as linked open data
- Linked data allows to approach the data from different **angles**:
  amounts, programmes, political category, Directorate-General, and many more
- Can be **linked** to by other datasets, e.g. FTS
- The data is now defined at the **granularity** at the **data entity level** (data oriented), while before it was defined at the file level (defining titles, chapters, articles).



EXAMPLES OF EU FISCAL DATA AS LINKED OPEN DATA

What is the total amount spent on Horizon 2020 in the past years?

Who received this money?
Which companies? Which universities?

Under which conditions can this money be spent?

When are procured services for this expense due?

How is this spent?
Grant? Contract? Prize?

Which other expenses are allocated to this MFF heading? What is its ceiling?

In which countries are recipients located?

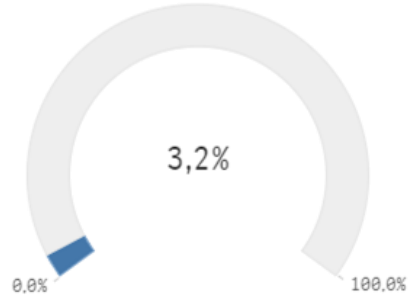
How does this heading translate into Bulgarian?

REAL EXAMPLE: DYNAMIC VISUALISATION (H2020)

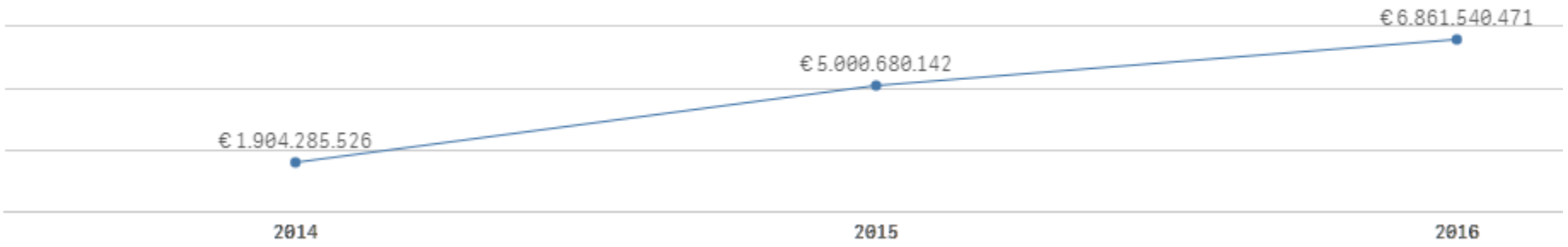
EU Programme

- H2020 ✓
- AFIS2020
- AMF
- CEF
- CF
- CF_DET

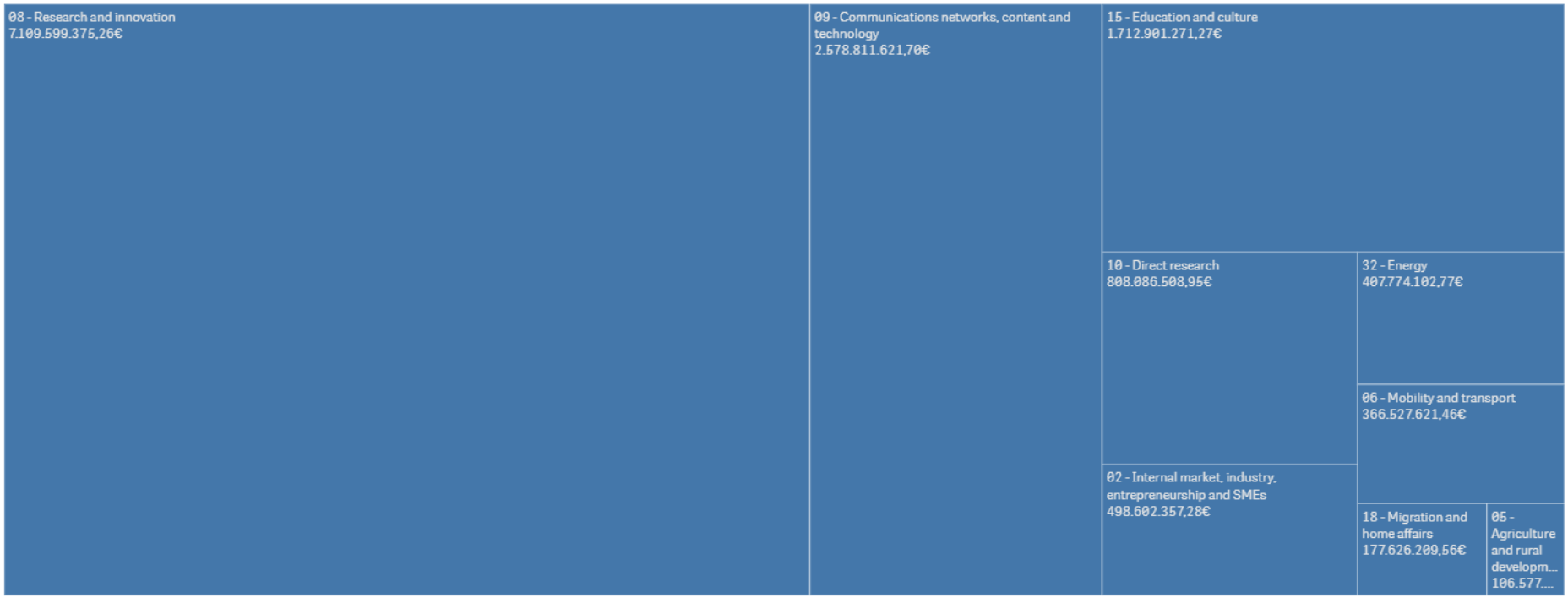
Year	Change %
2014	-
2015	163%
2016	37%



Change over time



REAL EXAMPLE: H2020 DYNAMIC VISUALISATION



REAL EXAMPLE: "SEARCH" QUERY H2020 & EURATOM

On the Open Data Portal of the EU, under the "Linked Data" tab:

SPARQL

You can search for the metadata stored in the EU Open Data Portal triple store by using the SPARQL endpoint query editor below.

Namespaces *

```
PREFIX odp:  <http://data.europa.eu/euodp/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX bud: <http://data.europa.eu/bud#>
```

SPARQL Query *

```
SELECT

' Amounts spent on Horizon 2020 or Euratom. ' AS ?ResultOf

sum(xsd:decimal(REPLACE(REPLACE(str(?MV_Value),' ',''),',','.'))) AS ?
Amount

WHERE {

 ?Title bud:heading ?TiHeading.
 FILTER (lang (?TiHeading) = 'en').
 ?Title bud:alias ?TiAlias.
 ?Title bud:hasChapter ?Chapter.
 ?Chapter bud:hasArticle ?Article.
 ?Chapter bud:hasArticle ?Article.
```

Format *

HTML

ResultOf	Amount
Amounts spent on Horizon 2020 or Euratom.	15719717668

EXAMPLES OF EU FISCAL DATA AS LINKED OPEN DATA

What is the total amount spent on Horizon 2020 in the past years?

Under which conditions can this money be spent?

When are procured services for this expense due?

Who received this money?
Which companies? Which universities?

How is this spent?
Grant? Contract? Prize?

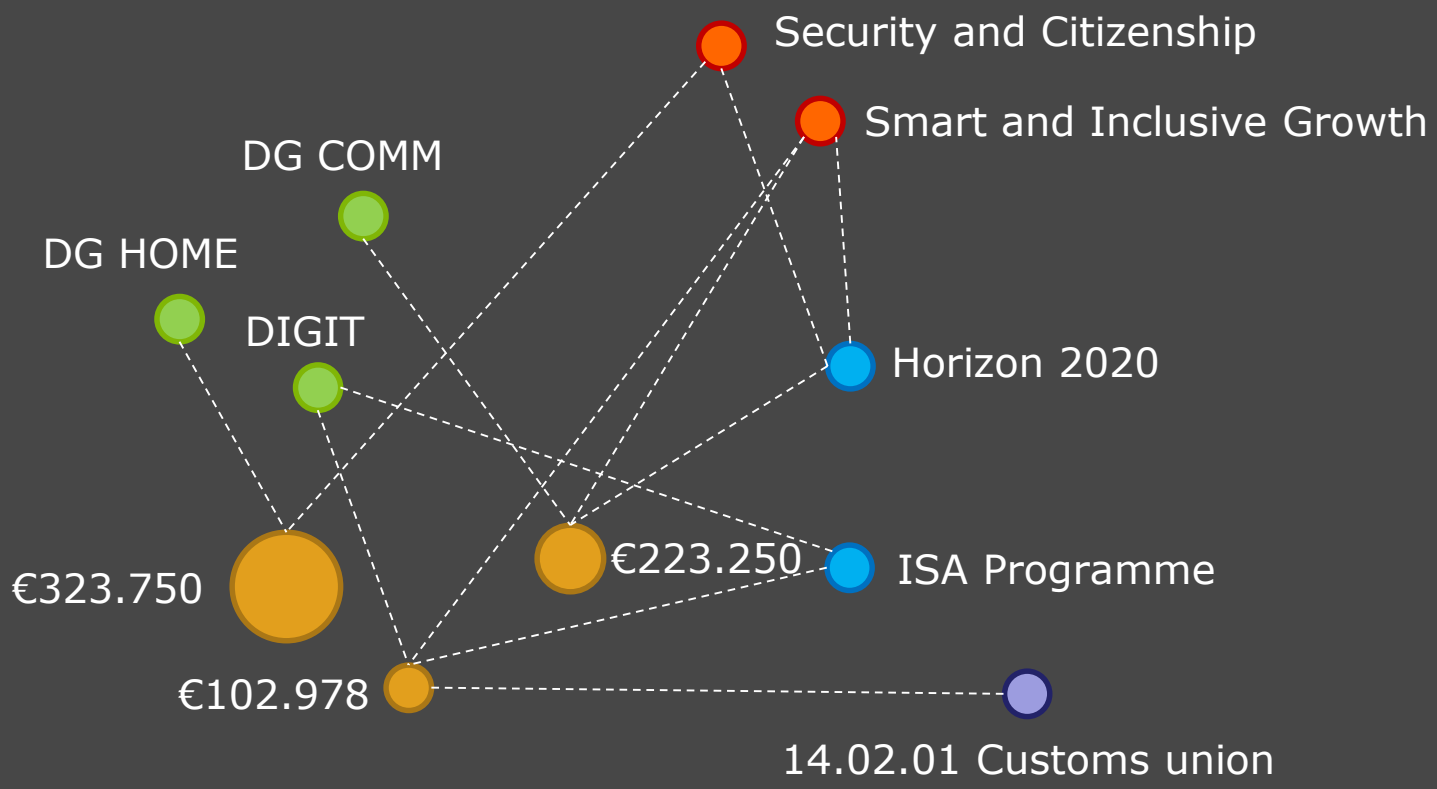
Which other expenses are allocated to this MFF heading? What is its ceiling?

In which countries are recipients located?

How does this heading translate into Bulgarian?

EU BUDGET AS LOD

Who received the money? Where are they located?



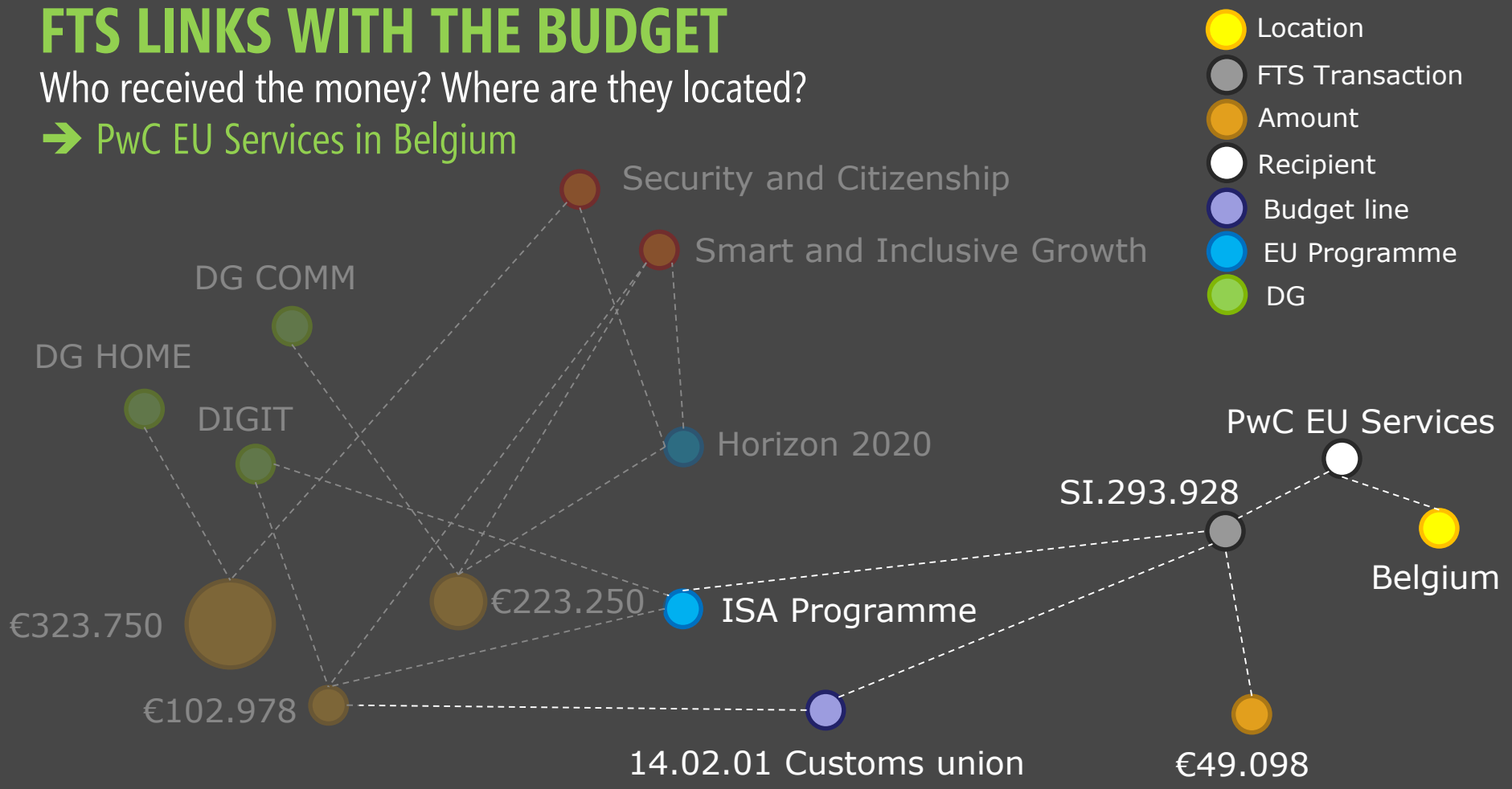
- Location
- FTS Transaction
- Amount
- Recipient
- Budget line
- EU Programme
- DG

Simplified, fictitious example

FTS LINKS WITH THE BUDGET

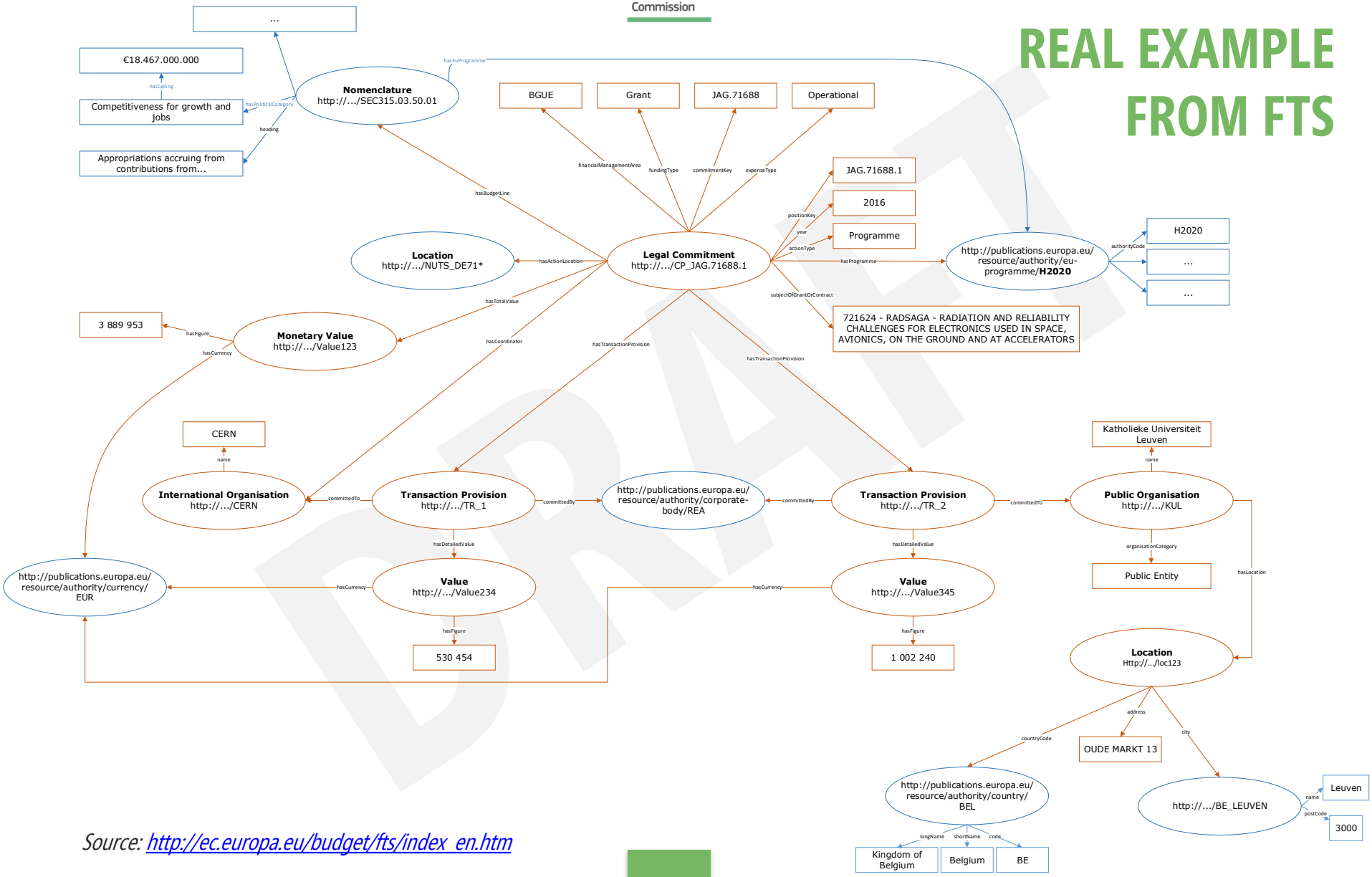
Who received the money? Where are they located?

➔ PwC EU Services in Belgium



Simplified, fictitious example

REAL EXAMPLE FROM FTS

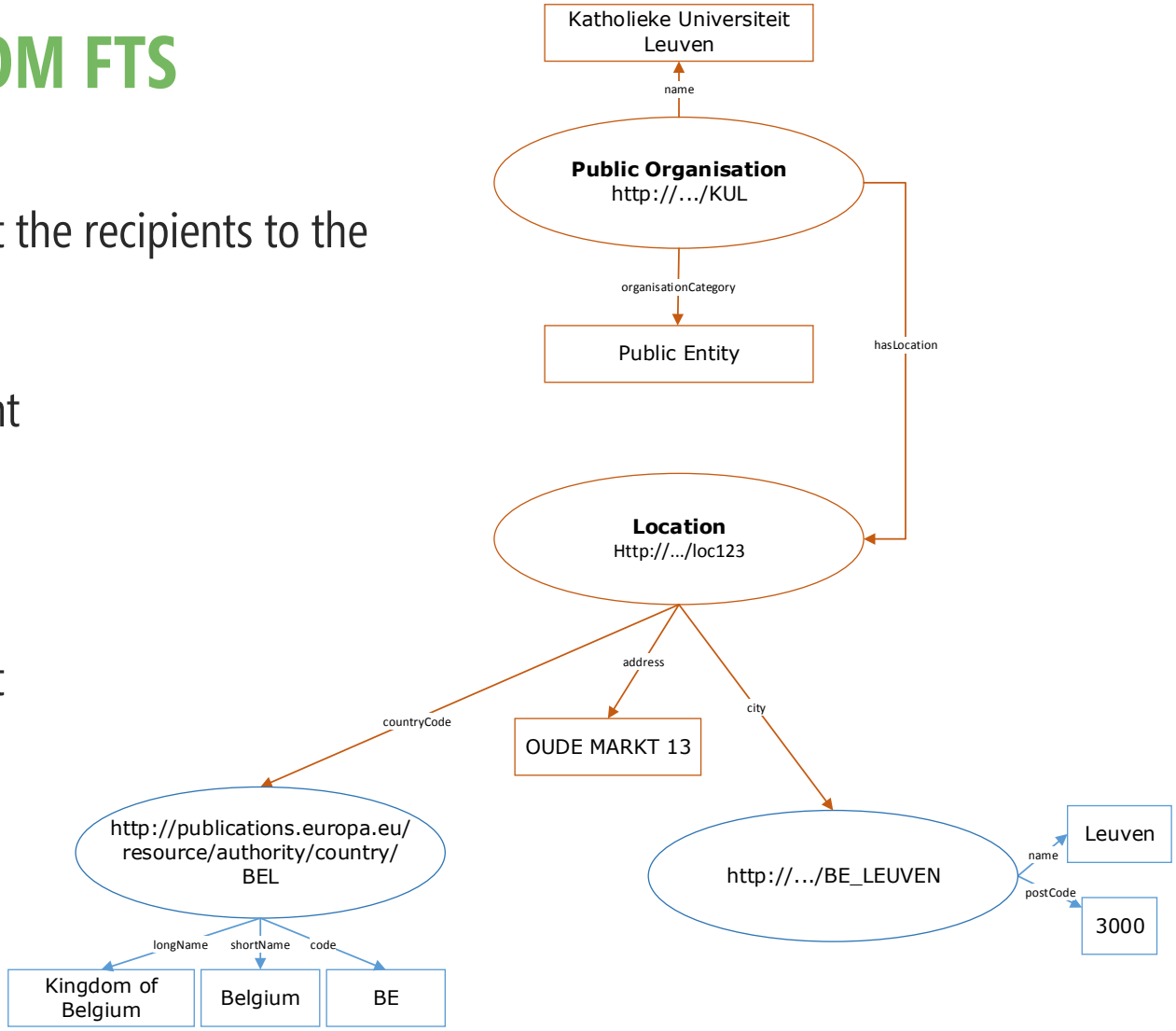


Source: http://ec.europa.eu/budget/fts/index_en.htm

REAL EXAMPLE FROM FTS

FTS links information about the recipients to the budget data, such as:

- Name & type of recipient
- Amount received by each recipient
- Location of the recipient
- Type of expenditure (grant, prize,...)
- ...

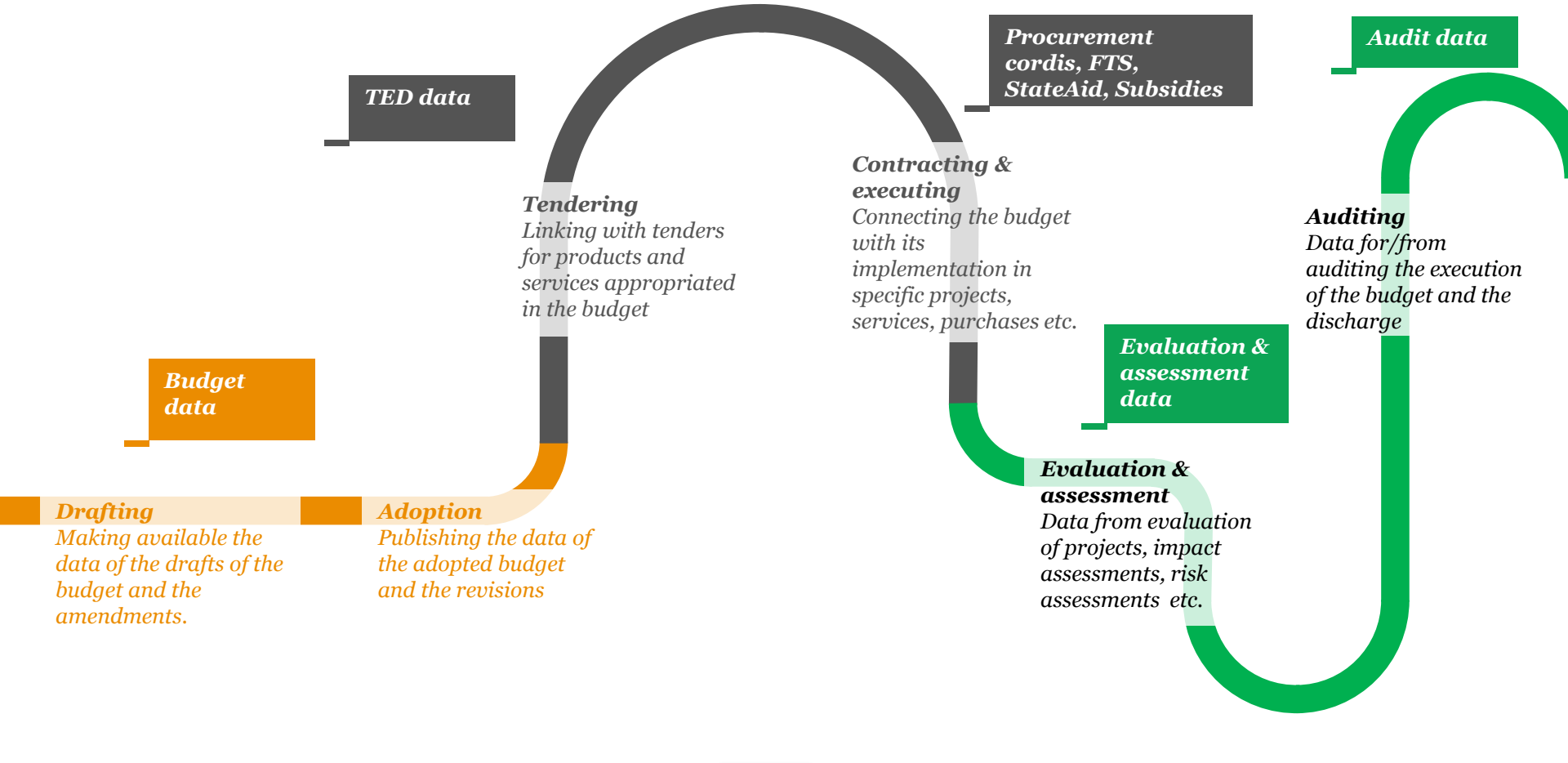


Source: http://ec.europa.eu/budget/fts/index_en.htm

3

Next steps

CLOSING THE CIRCLE



DOCUMENT METADATA

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.

Version: 1.0

Publication date: 7 July 2017

Created by: Nikolaos Loutas, Daniel Brulé, Brecht Wyns [PwC EU Services]
Reviewed by: Agnieszka Zajac [Publications Office of the EU]