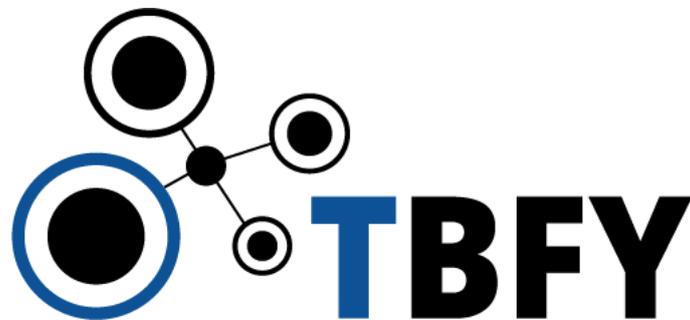


Innovation Action (IA)

ICT-14-2016-2017

H2020-ICT-2017-1

Enabling procurement data value chains for economic development,
demand management, competitive markets and vendor intelligence



Deliverable D5.1

Report on best practices for software engineering, publishing and using procurement data

Date	30/06/2018
Author(s)	Francisco Yedro, Oscar Corcho, Carlos Badenes (UPM)
Dissemination level	PU
Work package	5
Version	1.0

Document metadata

Quality assurers and contributors

Quality assessor(s)	Divna Djordjevic
Contributor(s)	Francisco Yedro, Oscar Corcho, Carlos Badenes

Version history

Date	Version	Description
17/04/2018	0.1	Elaboration of the ToC
17/05/2018	0.2	Review on existing practices for public procurement data delivery
20/05/2018	0.3	Software development methodology added
04/06/2018	0.4	Comments added on the current version of existing practices and public procurement data delivery, and deliverable ready for external QA
30/06/2018	1.0	QA comments received and added to the document

Executive summary

This document presents Deliverable D5.1 “Report on best practices for software engineering, publishing and using procurement data” of the TheyBuyForYou project. This deliverable is developed as part of Work Package 5 (WP5) “Standards, best practices, and integration”, tasks T5.1 “Best practices for the publication and governance of procurement knowledge graphs” and T5.3 “Software engineering methodology”.

The first part of this document reviews the current state-of-the-art on the publication of open procurement data worldwide, with special attention to those initiatives that have focused on the generation and publication of knowledge graphs about public procurement data, since these will be useful to inform the development of the TheyBuyForYou APIs that is being described in Deliverable D5.2 and further versions of that document.

The ontology-based initiatives described in this deliverable include the Public Contract Ontology (PCO), the Public Procurement Ontology (PPROC), and the upcoming eProcurement Ontology (ePO). Additionally other relevant, non ontology-based worldwide initiatives, with a strong influence on the public procurement data governance in the TheyBuyForYou project, have been considered. I.e. the Open Contracting Data Standard (OCDS). Public procurement includes: the procurement process (tendering, selection, notification, etc.) and its associated data and documents, the organisations and individuals that are involved in the process (e.g., buyers, managing entities and suppliers), and other elements that need to be considered in the context of public procurement (e.g. goods and services, public budgets, etc.). Our analysis is mostly focused on these core parts of the public contracting process. The work on the TheyBuyForYou APIs will go further into the direction of considering additional components of public procurement.

Although public procurement not only refers to the procurement process, and its associated data and documents, throughout all its stages (tendering, selection, notification, etc.), but also to the organisations and individuals that are involved in the process (as buyers, managing entities and suppliers), as well as to other elements that need to be considered in the context of public procurement: goods and services, public budgets, etc., our analysis is mostly focused on such core part of the public contracting process. Work on the TheyBuyForYou APIs will go further into the direction of considering also those other components of public procurement.

The second part of this document is focused on describing how we are adapting an agile-based software development methodology, such as SCRUM, for the purpose of driving software engineering tasks in this project. We will describe how we have adapted the basic principles of an agile-based methodology so as to consider the distributed geographical and organisational setting in which an EU project like TheyBuyForYou operates, including how user stories will be initially derived from the work on business cases that is being performed in the context of WP6. This methodology should be seen as an initial proposal on how to optimise the way in which all the TheyBuyForYou organisations will be working together towards producing the software that will be released by the project, and will be updated at regular intervals taking into account the feedback and lessons learned from its application in the context of the project.

Table of contents

1	INTRODUCTION	5
2	STATE OF THE ART IN THE PUBLICATION AND ACCESSIBILITY OF PROCUREMENT DATA AS KNOWLEDGE GRAPHS	7
2.1	ONTOLOGY-BASED PUBLICATION OF PUBLIC PROCUREMENT DATA.....	7
2.2	NON-ONTOLOGY-BASED PUBLICATION OF PUBLIC PROCUREMENT DATA.....	9
3	PUBLICATION AND GOVERNANCE OF PROCUREMENT DATA IN THEYBUYFORYOU	12
3.1	PRINCIPLES AND GUIDELINES FOR THE PUBLICATION OF THE PROCUREMENT DATA KNOWLEDGE GRAPH	12
3.1.1	<i>Technical guidelines</i>	<i>12</i>
3.1.2	<i>Licensing guidelines.....</i>	<i>13</i>
3.2	DATA GOVERNANCE GUIDELINES FOR THE MANAGEMENT, MAINTENANCE AND UPDATING OF THE PROCUREMENT DATA KNOWLEDGE GRAPH	13
4	SOFTWARE ENGINEERING METHODOLOGY.....	14
4.1	WHAT IS SCRUM?.....	14
4.1.1	<i>The Scrum TEAM in TheyBuyForYou.....</i>	<i>14</i>
4.1.2	<i>Scrum Events in TheyBuyForYou.....</i>	<i>15</i>
4.2	HOW THE METHODOLOGY WILL BE APPLIED IN THEYBUYFORYOU.....	16
4.2.1	<i>Organization of the sprints</i>	<i>17</i>
5	REFERENCES	18

1 Introduction

Procurement is a huge business – it affects virtually all sectors and organisations. In the public sector alone, spending on goods and services across the EU is estimated to exceed €2trn per annum¹. In the same time, governments and state-owned enterprises are confronted with massive challenges: they must deliver services with greatly reduced budgets; prevent losses through fraud and corruption; and build healthy, sustainable economies. Managing these competing priorities at a national level and across many disparate agencies is notoriously difficult; coordinating them across Member States seems almost impossible. In times of enhanced transparency and accountability, there is a pressing need for better insight into, and management of government spending. In the absence of data and tools to analyse and oversee this complex process, too little consideration is given to the development of vibrant, competitive economies when buying decisions are made. Too often, the same suppliers are given contracts without considering the total, longer-term benefits to the government’s bottom line or the broader implications of tendering decisions in terms of employment and skills development. At the same time, we struggle to identify fraud in public tendering; the OECD estimates that more than 82% of these cases remain undetected across all OECD countries,² while a recent study by RAND Europe puts the costs of corruption in EU public tendering as high as €990b a year.³

Private sector procurement faces similar challenges. Decision makers are expected to deliver the best possible value and to run efficient and accountable tendering procedures, whether the beneficiaries are company shareholders or taxpayers. While the process tends to be less mechanistic than in the public sector and decisions and framework conditions are not subject to public scrutiny, these decisions do need to be transparent internally, and to demonstrate optimality in terms of costs and value for the company in choosing new suppliers. This becomes increasingly difficult as business becomes more global and the international flow of goods, services, finances, and people reaches levels unimagined a few decades ago.⁴

With so much at stake, there is no doubt that procurement, whether in the public or the private sector, needs to be rethought. The TheyBuyForYou project aims at delivering the tools, alongside enabling technologies and datasets, to make procurement across the EU more efficient, competitive, accountable, and fair. We believe an open approach plays a critical role in this context – in the case of public entities, they are increasingly required to publish information about their procurement processes. By using transparency data on government procurement, enriched and integrated with other open PSI sources, and processed and analysed using cutting-edge data science methods, we will shed light on the opportunities and challenges faced by all direct stakeholders in taking more informed and effective decisions.

This deliverable reports the work that is being performed in the context of WP5, and more specifically, as part of tasks T5.1 “Best practices for the publication and governance of procurement knowledge graphs” and T5.3 “Software engineering methodology”. The aim of this deliverable is hence describing the methodology that will be applied in the context of the project for the development of the procurement data provisioning and publishing mechanisms, as well as all the associated tools that will give support to the business scenarios that are being developed in the context of WP6.

The deliverable is structured into three main sections:

¹ <http://ec.europa.eu/DocsRoom/documents/20679/attachments/1/translations/en/renditions/native>

² <http://www.oecd.org/gov/ethics/48994520.pdf>

³ <http://www.politico.eu/article/corruption-costs-eu-990-billion-year-rand-study-fraud-funding/>

⁴ <http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/global-flows-in-a-digital-age>

- Section 2 gives an overview of existing and ongoing efforts on the publication of open public procurement data. It describes both ontology-based and non-ontology-based approaches that have been used for this purpose.
- Section 3 discusses on the main principles that will be followed by the TheyBuyForYou for the publication of open procurement data, as well as the governance of all the data sources that will be gathered, enriched and interlinked as part of the activities of WP1 and WP2. This part of the deliverable is complimentary to the work that is presented in deliverable D5.2 (and subsequent deliverables) on the TheyBuyForYou API, as well as to the work presented in the Data Management Plan.
- Section 4 describes the software engineering methodology that will be followed in the context of the project for the development of the different software components and APIs. This methodology is inspired by agile-based methodologies and takes into account the usual characteristics of an EU project where software development teams are distributed both geographically and across organisations.

2 State of the art in the publication and accessibility of procurement data as knowledge graphs

Public procurement data can be considered as a specific type of public sector information (PSI) data, and consequently, PSI directives (Directive 2003/98/EC⁵, which entered into force on 31 December 2003, and its revision on Directive 2013/37/EU⁶, which entered into force on 17 July 2013) and their corresponding national legislations are applicable to this type of data. The latest directive has undergone in 2017 a public online consultation, and as a result, in April 2018 a proposal for a revision has been adopted by the European Commission⁷, which is now being discussed by the European Parliament and Council.

Furthermore, the EU directive for public procurement (Directive 2014/24/EU⁸) and the associated national legislations, many of which have entered into application in 2018 (e.g., the Spanish Law 9/2017⁹), are also applicable. Such legislation usually refers to the corresponding PSI directives or applicable national laws (e.g., in the Spanish case, this happens in the “Disposición Adicional Decimotercera” , which modifies law 37/2007 on PSI reuse (see footnote above).

Although the previous discussion is only focused on the EU context, which is the main focus of the TheyBuyForYou project, similar contexts are applicable worldwide. In any case, our discussion will be mostly focused on the EU context from now on.

As a result of all this legal framework, national public procurement portals have been created, which live together with regional, local as well as EU-wide public procurement portals. However, there is no common agreement across the EU (not even, in many cases, inside the same country) on the data formats to be used for exposing such data sources and on the data models that need to be used for exposing such data, what leads to a large heterogeneity in the data that is being exposed. A simple analysis on the EU open data portal, after querying for datasets related to “procurement”¹⁰ (which may not even provide all the datasets that are related to public procurement, as well as include datasets that are not relevant) shows around 1000 datasets in the area, with a large variety of formats and many different types of data models.

Several proposals have been made in order to try to homogenise the data models and formats used for the publication of procurement data. In the following sections we describe the most relevant efforts where ontologies have been proposed (section 2.1) as well as other worldwide efforts that are worth analysing, since they will also be considered in the context of our project as one of the main data sources (section 2.2).

2.1 Ontology-based publication of public procurement data

Ontologies provide a commonly agreed and shared vocabulary for a specific domain. They are being used extensively in the publication of open data worldwide, especially in the context of Linked Data publication (which will be described in section 3, as well as in deliverables D5.2 and onwards).

In the context of public procurement, a public procurement ontology or ontology network will represent concepts such as those of public contracts, associated documents, public administrations, buyers, suppliers, budgets, etc., together with their corresponding attributes (either compulsory or optional). An in-depth recent review of existing ontologies in the area of public contracting can be found at [1], a paper that was co-authored by one of the

⁵ <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32003L0098>

⁶ <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32013L0037>

⁷ <https://ec.europa.eu/digital-single-market/en/proposal-revision-psi-directive>

⁸ <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32014L0024>

⁹ <https://www.boe.es/buscar/doc.php?id=BOE-A-2017-12902>

¹⁰ <https://www.europeandataportal.eu/data/en/dataset?q=procurement>

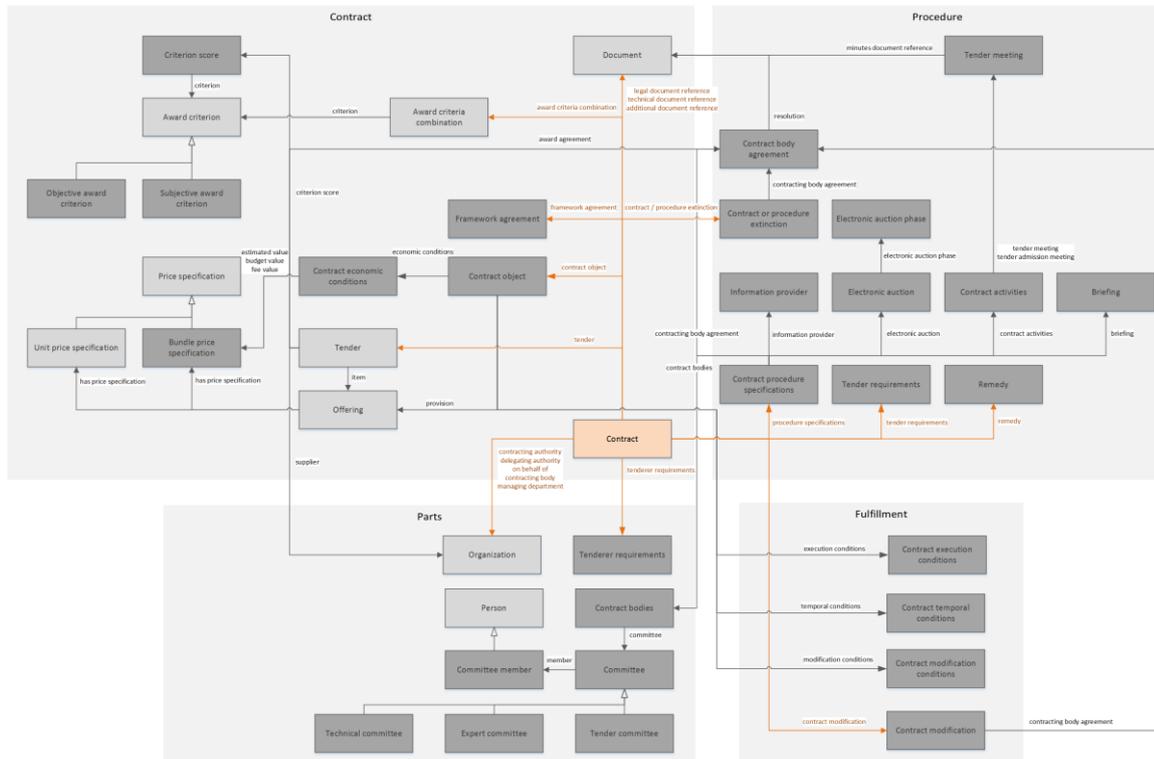


Figure 2. A diagram of the main classes in the PPROC Ontology

eProcurement Ontology¹⁴ (ePO). This ontology, which is expected to be available by the end of 2018, will be the result of the ongoing work funded by the EU Publications Office, which is being recorded in the GitHub repository that is referenced above. This ontology aims at providing a commonly agreed vocabulary about public procurement according to the EU legislation, and focuses on covering all the steps that public contracts need to go through (notification, tendering to awarding, ordering, invoicing and payment). Several scenarios obtained from a range of stakeholders have set the basis for the development of this ontology, and ongoing work is mostly focused on the definition and revision of the terms that will be part of the glossary of terms.

2.2 Non-ontology-based publication of public procurement data

As discussed in the introduction, different data providers (e.g., public administrations) provide their public procurement data according to different models and formats, and there is no common agreement EU-wide (not to say worldwide) to provide such data in a unified manner. In Europe, contracting portals like TED may be seen as a way to homogenise the data that is being provided, but unfortunately this portal is only used for those contracts that are larger than a predefined budget threshold, and hence this does not cover the whole richness of types of public contracts (e.g., those that are really small) nor does it force the usage of this format for those contracts that do not need to be published there.

National portals, especially with the application of the latest EU directive on public contracting, are also forcing public administrations in every country to publish their contracts in a centralised portal, so as to ensure better access for potential suppliers to public procurement, as well as to increase transparency. However, some of these

¹⁴ <https://github.com/eprocurementontology/eprocurementontology/>

portals are only starting to appear and be used for this purpose (e.g., the Spanish one¹⁵) and every country uses its own data format (e.g., in Spain, the CODICE¹⁶ XML-based format is being used).

Outside this context, the only relevant data model that is getting some important traction worldwide is the **Open Contracting Data Standard**¹⁷ (OCDS), initially developed for the Open Contracting Partnership¹⁸ by the World Wide Web Foundation¹⁹ and now being managed by Open Data Services Co-operative²⁰. This data model has been mostly developed with a focus on transparency in the public contracting procedure worldwide, so as to allow analyses of contracting data inside and across different institutions. Furthermore, the data model (which is outlined in Figure 3) provides guidelines for global identifiers to be used to identify contracts worldwide, as well as for the different stages through which contract procedures go, so that it may be used as well to provide some support for contracting processes and buyer profiles. The implementation is based on the usage of the JSON data format, and several conversion and data management and exploitation tools exist, as well as validators, many of which are open sourced.

Deliverable D5.2 contains several references to this data model, given the fact that one of the TheyBuyForYou partners (OpenOpps) is currently maintaining and publishing a large database of public contracts using OCDS as the underlying model. Therefore, its REST API, which is briefly described in deliverable D5.1, is strongly based on the structure of this data model.

¹⁵ <https://contrataciondelestado.es/>

¹⁶ <https://contrataciondelestado.es/wps/portal/codice>

¹⁷ <http://standard.open-contracting.org/latest/en/>

¹⁸ <http://www.open-contracting.org/>

¹⁹ <http://www.webfoundation.org/>

²⁰ <http://www.opendataservices.coop/>

Schema for an Open Contracting Release

Each release provides data about a single contracting process at a particular point in time. Releases can be used to notify users of new tenders, awards, contracts and other updates. Releases may repeat or update information provided previously in this contracting process. One contracting process may have many releases. A 'record' of a contracting process follows the same structure as a release, but combines information from multiple points in time into a single summary.

ocid `string [1..]`
A globally unique identifier for this Open Contracting Process. Composed of a publisher prefix and an identifier for the contracting process. For more information see the [Open Contracting Identifier guidance](#)

id `string [1..]`
An identifier for this particular release of information. A release identifier must be unique within the scope of its related contracting process (defined by a common ocid), and unique within any release package it appears in. A release identifier must not contain the # character.

date `string (format: date-time)` The date this information was first released, or published.

tag `array[1..] enum string` `planning` `planningUpdate` `tender` `tenderAmendment` `tenderUpdate` `tenderCancellation` `award` `awardUpdate` `awardCancellation` `contract` `contractUpdate` `contractAmendment` `implementation` `implementationUpdate` `contractTermination` `compiled`
One or more values from the [releaseTag codelist](#). Tags may be used to filter release and to understand the kind of information that a release might contain.

initiationType `constant string` `tender`
String specifying the type of initiation process used for this contract, taken from the [initiationType codelist](#). Currently only tender is supported.

parties `array` `Organization`
Information on the parties (organizations, economic operators and other participants) who are involved in the contracting process and their roles, e.g. buyer, procuring entity, supplier etc. Organization references elsewhere in the schema are used to refer back to this entries in this list.

buyer `Organization reference`
A buyer is an entity whose budget will be used to pay for goods, works or services related to a contract. This may be different from the procuring entity who may be specified in the tender data.

planning `Planning`
Information from the planning phase of the contracting process. This includes information related to the process of deciding what to contract, when and how.

tender `Tender` The activities undertaken in order to enter into a contract.

awards `array` `Award`
Information from the award phase of the contracting process. There may be more than one award per contracting process e.g. because the contract is split among different providers, or because it is a standing offer.

contracts `array` `Contract`
Information from the contract creation phase of the procurement process.

language `string,null` default `en`
Specifies the default language of the data using either two-letter [ISO639-1](#), or extended [BCP47 language tags](#). The use of lowercase two-letter codes from [ISO639-1](#) is strongly recommended.

relatedProcesses `array` `Related Process`
If this process follows on from one or more prior process, represented under a separate open contracting identifier (ocid) then details of the related process can be provided here. This is commonly used to relate mini-competitions to their parent frameworks, full tenders to a pre-qualification phase, or individual tenders to a broad planning process.

Figure 3. Overview of the OCDS data model used to describe a contract (extracted from the OCDS website).

3 Publication and governance of procurement data in TheyBuyForYou

Once we have reviewed the main (ontology-based and non-ontology based) data models and formats that are being used in the state of the art for the publication of procurement data (mostly public), in this section we will describe the main principles and guidelines that will be used in the context of TheyBuyForYou for the management, governance and publication of our knowledge graph. We expect these guidelines and principles to remain for the whole duration of the project, although some aspects may need to be revisited especially taking into account developments around public contracting data publication in Europe and worldwide. Whilst the actual implementation of these guidelines in the form of a specific data model and ontology network, as well as in the form of the TheyBuyForYou API, will be continuously updated during the project.

3.1 Principles and guidelines for the publication of the procurement data knowledge graph

3.1.1 Technical guidelines

As discussed in deliverable D5.2, the TheyBuyForYou knowledge graph will be generated from a wide range of data sources (mostly the buyer profiles and transparency portals of public administrations EU-wide), many of which will be collaborating with TheyBuyForYou in the provision of the relevant data (ideally by producing their data according to the TheyBuyForYou ontology), while in other cases there will be a need to obtain that data by processing (e.g., Web scraping) the data sources, what may lead to some additional noise in the data that is maintained in the knowledge graph. Deliverable D5.2 provides a figure with the current overview of the architecture for data acquisition, processing and enrichment, prior to its storage in the TheyBuyForYou knowledge graph.

Once the data is stored in the TheyBuyForYou knowledge graph, there are several possibilities for exposing it on the Web. It may be exposed as a data dump (in a single or multiple files) in a data format such as JSON, it may be exposed in a Web-oriented presentation language such as HTML, and/or it may be exposed following Linked Data principles²¹:

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful RDF information.
- Include RDF statements that link to other URIs so that they can discover related things.

In our project, we will focus on the latter, although this will not prevent us from providing data dumps for those willing to make analyses on the whole set of datasets (e.g., massive processing of all the tender notices) as well as providing HTML views on all the items in the knowledge graph. Linked Data principles will ensure that the data from our knowledge graph is accessible through the HTTP (or HTTPS) protocol, as discussed in D5.2, in different formats, using content negotiation. We will aim at providing data, at least, in RDF Turtle (and possibly N3) format, in JSON-LD, in JSON, and in CSV, using our REST API, as well as in HTML for human consumption.

The API will be authenticated (although there will be also, at least initially, a possibility to use it without any authentication, for occasional use), so that we will be able to record API usage as well as to apply authorisation mechanisms for the part of the data that will not be publicly accessible.

²¹ <https://www.w3.org/DesignIssues/LinkedData.html>

3.1.2 Licensing guidelines

Licensing is relevant in order to let TheyBuyForYou internal users (e.g., those responsible for business cases) as well as third party reusers of our contracting data, know under what conditions our data can be used and reused.

Different types of data will be made available in the TheyBuyForYou knowledge graph. In general, the following principles will be applicable:

- Open licenses with attribution requirements (e.g., CC-BY 4.0) for internal TheyBuyForYou users during the duration of the project. That, is users will be free to copy and redistribute the work and make derivative works and remixes based on it, in any medium of format, only if they give the author or licensor the attribution for any purpose, even commercially, in the manner specified by these
- Dual licenses (open and non-open) for data that has been created in the context of the project with tools that were available by any of the project partners before the project started. This will be the case for the OCDS-based contracting data provided by OpenOpps and for the data about organisations provided by OpenCorporates. The dual license will be as follows:
 - Open license in the case that such data is being used for the generation of applications that are available openly, with non-commercial purposes (e.g., CC-BY-NC)
 - Closed license for data being used for the generation of applications that are being commercialised.
- Open licenses with attribution requirements (e.g., CC-BY 4.0) for the data from those data providers that agree on providing data according to the TheyBuyForYou ontology (e.g., Zaragoza, MJU, and those that may decide during the course of the project to adapt to this model, so that no further processing or enrichment is necessary).

The TheyBuyForYou Data Management Plan provides additional details on the licensing mechanisms to be used for the data sources, and the specific licenses and provisions for different types of data may also evolve over the course of the project.

3.2 Data governance guidelines for the management, maintenance and updating of the procurement data knowledge graph

In TheyBuyForYou we will aim at making sure that data providers make their procurement data available for third party reusers (including our own TheyBuyForYou business cases) in their own domains/sites, according to the TheyBuyForYou ontology network. This means that we will actually ensure that a pure Linked Data set of principles is applied for the purpose of data publication.

Since this will not be possible during the course of the project with every single public data provider in the EU, we will also run processes of ingestion, processing and enrichment of the data that will be made available in other forms by public administrations across the EU. This will mean that we will also maintain a centralised data repository for the TheyBuyForYou knowledge graph, which will combine a replication of the data that is being provided by third party data providers, plus the data that will be ingested using other means (e.g., Web scrapers). Every single data provider will be treated differently, and update frequencies for the data that the TheyBuyForYou knowledge graph will take into account will be specified for each data source, so that data reusers for our knowledge graph will always have an up-to-date view on how the data is being maintained in our knowledge graph.

4 Software engineering methodology

This section discusses on the software engineering methodology that will be applied for the implementation of the software components that will be required to provide support to all the functionalities provided by the project, both in terms of the provision of data in the knowledge graph as well as in the exploitation of such data sources.

Our project will use a software engineering methodology based on agile development principles and based on SCRUM. Section 4.1 provides details of the general Scrum-based methodology, based on the general description provided at [3].

4.1 What is Scrum?

Scrum is a framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value. Scrum itself is a simple framework for effective team collaboration on complex products.

Scrum is not a process, technique, or definitive method. Rather, it is a framework within which one can employ various processes and techniques. Scrum makes clear the relative efficacy of product management and work techniques so that one can continuously improve the product, the team, and the working environment.

4.1.1 The Scrum TEAM in TheyBuyForYou

The Scrum Team consists of a Product Owner, the Development Team, and a Scrum Master:

- **Product Owner:** The Product Owner is responsible for maximizing the value of the product resulting from work of the Development Team. How this is done varies widely across organizations, Scrum Teams, and individuals. In the case of TheyBuyForYou, the product owner will be Oscar Corcho, who is the principal investigator of UPM at the consortium.

The Product Owner is the sole person responsible for managing the Product Backlog. Product Backlog management includes:

- Clearly expressing Product Backlog items;
 - Ordering the items in the Product Backlog to best achieve goals and missions;
 - Optimizing the value of the work the Development Team performs;
 - Ensuring that the Product Backlog is visible, transparent, and clear to all, and shows what the Scrum Team will work on next; and,
 - Ensuring the Development Team understands items in the Product Backlog to the level needed.
- **The Development Team:** The Development Team consists of those doing the work of delivering a potentially releasable Increment of “Done” product at the end of each Sprint. A “Done” increment is required at the Sprint Review. Only members of the Development Team create the Increment.

In TheyBuyForYou, the Development Teams, which will be composed of people from different organisations in the consortium, will have the following characteristics:

- They are self-organizing. No one (not even the Scrum Master) tells the Development Team how to turn Product Backlog into Increments of potentially releasable functionality;
- Development Teams are cross-functional (and cross-organisational), with all the skills as a team necessary to create a product Increment;

- Scrum recognizes no titles for Development Team members, regardless of the work being performed by the person;
- Scrum recognizes no sub-teams in the Development Team, regardless of domains that need to be addressed like testing, architecture, operations, or business analysis; and,
- Individual Development Team members may have specialized skills and areas of focus, but accountability belongs to the Development Team as a whole.
- **The Scrum Master:** The Scrum Master is responsible for promoting and supporting Scrum, helping everyone understand Scrum theory, practices, rules, and values. The Scrum Master will be Francisco Yedro, from UPM.

4.1.2 Scrum Events in TheyBuyForYou

Prescribed events are used in Scrum to create regularity and to minimize the need for meetings not defined in Scrum. All events are time-boxed. Once a Sprint begins, its duration is fixed and cannot be shortened or lengthened. The remaining events may end whenever the purpose of the event is achieved, ensuring an appropriate amount of time is spent without allowing waste in the process. The Scrum Events are:

- **Sprint:** The heart of Scrum is a Sprint, a time-box that in the context of TheyBuyForYou will last for one month, during which a “Done”, useable, and potentially releasable product Increment is created. Sprints have consistent durations throughout a development effort. A new Sprint starts immediately after the conclusion of the previous Sprint.

During the Sprint:

- No changes are made that would endanger the Sprint Goal;
- Quality goals do not decrease;
- Scope may be clarified and re-negotiated between the Product Owner and Development Team as more is learned.
- **Sprint Planning:** The work to be performed in the Sprint is planned at the Sprint Planning. This plan is created by the collaborative work of the entire Scrum Team.

Sprint Planning is time-boxed to a maximum of two hours for a one-month Sprint (via audioconference, where discussion will be held on the steps to be followed for the following sprint). The Scrum Master ensures that the event takes place and that attendants understand its purpose. The Scrum Master teaches the Scrum Team to keep it within the time-box.

- **Daily Scrum:** The Daily Scrum is a 15-minute time-boxed event for the Development Team, done via a dedicated Skype channel. The Daily Scrum is held every day of the Sprint. At it, the Development Team plans work for the next 24 hours. This optimizes team collaboration and performance by inspecting the work since the last Daily Scrum and forecasting upcoming Sprint work. The Daily Scrum is held at the same time and place each day to reduce complexity.

The Development Team uses the Daily Scrum to inspect progress toward the Sprint Goal and to inspect how progress is trending toward completing the work in the Sprint Backlog. The Daily Scrum optimizes the probability that the Development Team will meet the Sprint Goal. Every day, the Development Team should understand how it intends to work together as a self-organizing team to accomplish the Sprint Goal and create the anticipated Increment by the end of the Sprint.

- **Sprint Review:** A Sprint Review is held at the end of the Sprint to inspect the Increment and adapt the Product Backlog if needed. During the Sprint Review, the Scrum Team and stakeholders collaborate

about what was done in the Sprint. Based on that and any changes to the Product Backlog during the Sprint, attendees collaborate on the next things that could be done to optimize value. This is an informal meeting, not a status meeting, and the presentation of the Increment is intended to elicit feedback and foster collaboration.

- **Sprint Retrospective:** The Sprint Retrospective is an opportunity for the Scrum Team to inspect itself and create a plan for improvements to be enacted during the next Sprint.

The Sprint Retrospective occurs after the Sprint Review and prior to the next Sprint Planning. This is at most a three-hour meeting for one-month Sprints. For shorter Sprints, the event is usually shorter. The Scrum Master ensures that the event takes place and that attendants understand its purpose.

The Scrum Master ensures that the meeting is positive and productive. The Scrum Master teaches all to keep it within the time-box. The Scrum Master participates as a peer team member in the meeting from the accountability over the Scrum process.

4.2 How the methodology will be applied in TheyBuyForYou

This section explains how we will organise the SCRUM-based methodology in the context of the software development to be done in this project. Depending on the task to be done in each sprint, the work to be done will be included in one of the repositories created for the TBFY project (storytelling²², crosslinguality²³, knowledge-graph²⁴ and general²⁵). Each of these repositories is divided into projects, which correspond to each of the product backlog tasks to be performed, and these projects are composed of issues. These issues will be assigned to one or more users at the sprint planning session, and can be in one of the following states:

- **To do:** Containing all the identified issues in the project but still not started.
- **In progress:** Containing all the issues which you have started to work with.
- **Done:** Containing all the finished issues.

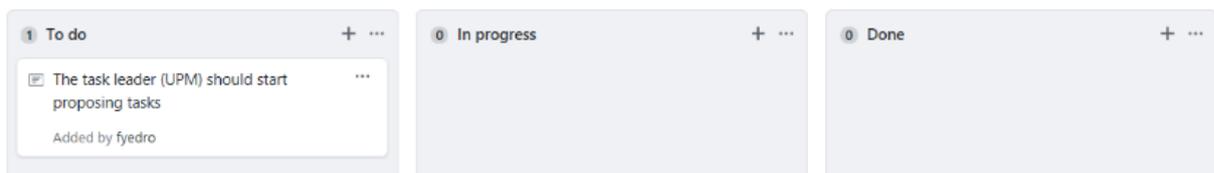


Figure 4. Kanban dashboard for each of the projects to be dealt with in a sprint session

The task leader will be free to propose new tasks as issues in each project assigning them to the corresponding team. These issues will be placed initially in the "To do" column and, depending on the state in which the task is, they will be moved to the "In progress" column or "Done" column.

Our list of issues is open to contributions and requests from any TBFY users/developers creating new ones. They will be able to post a 'REQUESTED' issue/task/story²⁶ at any point in time. This should contain details about the task/issue/user_story that one wants to include, including:

- The description of the task/issue/user_story.

²² <https://github.com/TBFY/storytelling>

²³ <https://github.com/TBFY/crosslinguality>

²⁴ <https://github.com/TBFY/knowledge-graph>

²⁵ <https://github.com/TBFY/general>

²⁶ http://github.com/TBFY/general/issues/new?title=NEW_TASK-ISSUE-USER_STORY

- The project to which it belongs, from the list of projects at <https://github.com/orgs/TBFY/projects> (or on any of the more specific repositories). If one is not sure about the project, this can be left blank. If a new project needs to be created, it can be also proposed there.
- The tasks (from the DoA) that it affects or relates to.
- Some indications about the importance of addressing it urgently or not.

Requests from TheyBuyForYou members that are explicitly discussed in the DoA will be prioritised during the sprint planning session, but in any case, the product backlog will be open to contributions from any other member of the community.

4.2.1 Organization of the sprints

We will have project-wide sprints. Sprints in our project will have a default duration of one month, so as to allow for an easier assignment and control of the tasks that are being performed by our distributed set of partners.

Here there is a brief description of how the sprints are organised:

- Week 1: decisions on epics to be handled and setting up the scene for the sprint
- Week 2-3: development
- Week 4: wrapping up and demo generation

We will have multiple projects/epics running in parallel where sprint leaders become project/epic leaders. Each activity in a project/epic must be a Github issue. Steps to follow creating sprints:

- Define the goal of each sprint & sprint planning
- Assemble their team
- Issue management (open, close and update issues)
- Coordinate their sprint team (daily stand-up etc.)

5 References

- [1] José Félix Muñoz Soro, Guillermo Esteban, Oscar Corcho, Francisco Serón. PPROC, an ontology for transparency in public procurement. *Semantic Web*, vol. 7, no. 3, pp. 295-309, 2016.
- [2] Svátek V., Mynarz J., Węcel K., Klímek J., Knap T., Nečaský M. (2014) Linked Open Data for Public Procurement. In: Auer S., Bryl V., Tramp S. (eds) *Linked Open Data -- Creating Knowledge Out of Interlinked Data*. Lecture Notes in Computer Science, vol 8661. Springer
- [3] Ken Schwaber and Jeff Sutherland, *The Scrum Guide*, 2017, <http://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf>