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

Results

Project Information

TheyBuyForYou
Grant agreement ID: 780247

Closed project

Funded under H2020-EU.2.1.1.

Overall budget € 3 279 240

EU contribution € 2 925 693

Coordinated by SINTEF AS Norway

Start date 1 January 2018
End date 31 December 2020

Deliverables

Other (21)

Cross-lingual document similarity service
Implementation of the document similarity computation service deployed as a web-service used through a RESTful API.

Common spending templates framework v2
Final version of the common spending templates framework, including clustering common spending patterns across different languages and abstracting their
Ingested data v1

The data that has been ingested will be made available to WP2 so that it can be enriched and published. Where possible, the graph will be released openly so that everyone can make use of it. This data will be continuously updated throughout the project and beyond, and can be considered as an intermediate point in the process of data publication. This specific deliverable will consist in a snapshot of such data and documentation on the underlying schema.

Internal communication tools

This deliverable will provide an initial version of a fact sheet to inform interested users about the project’s approach and goals. Furthermore, this deliverable will account for the established infrastructure for securing a tight collaboration and internal communication in the project.

Data curation, linking, reconciliation, and categorisation components

We will provide the open source code for the data curation, linking and categorisation components that will be configured and developed based on the partners’ background technology components, including the unitary tests to ensure consistency of the datasets created, as well as the code necessary to integrate with the reconciliation service.

Data comparison components

We will provide the source code and the service deployment for the components to compare data inside and across organisations. This deliverable will include also an update to the data curation, linking, reconciliation and categorisation components (2nd version of open source code).

Data source catalogue and audit

This deliverable will consist in a catalogue of identified data sources, together with the evaluation of their quality, and the provision of their corresponding metadata. The catalogue will be continuously updated during the project and beyond. Therefore, this specific deliverable will be setting the basic infrastructure for the catalogue deployment and maintenance, as well as the initial data sources that will be already identified early in the project.

Knowledge graph publication

The enriched data will be published according to established standards and best practice for the publication of data on the Web. The configurations for the services to use in order to do this publication will be made available early in the project and the data will be continuously updated from there on.
Public spending real-time monitoring and analytics demo v1
Initial implementation of the public spending monitoring framework implemented as a Web application, including clustering of related documents across languages and visualizations of cluster data.

Procurement APIs and platform release v1
First version of TheyBuyForYou’s procurement APIs and integrated software offerings

Ingested data v2
Second snapshot of the continuously updated data.

Public spending real-time monitoring and analytics demo v2
Final implementation of the public spending monitoring framework, including analysis of trends in public spending and demand for it.

Visualization and interaction components v1
Components based on existing technology to allow users to easily combine local and remote data sources into new visualizations, and to storify data into summaries and infographics.

Data gathering, extraction, pre-processing and normalisation components v3
Third version of the code for the ingestion of data.

Project manual and quality assurance plan
Detailed report on project management and quality assurance procedures.

Visualization and interaction components v3
Basic prototype JavaScript library will be developed and made available that can put the instructions in the output file together into a customizable story. The JavaScript library will be open-source and documented in a public code repository. The library will also be used to create the basic HTML Web pages made available directly from the editor.

Data gathering, extraction, pre-processing and normalisation components v1
Here we will provide the open source code, which will be maintained on GitHub, for the different types of scripts and software code that will be used to ingest public procurement data from different organisations. This software will be continuously evolving throughout the rest of the project, and will be submitted in two additional versions.

Data gathering, extraction, pre-processing and normalisation components v2
Second version of the code for the ingestion of data.
Visualization and interaction components v2
A core component of the proposed tool will be the open file format. The format will serve as an open specification for storing data, instructions for visualization and display. The release will cover configurable infographics software including automatic storytelling techniques from Task T4.3.

Procurement APIs and platform release v2
Second version based on new methods and data from WPs 1, 2, 3, and 4.

Common spending templates framework v1
An initial version of the framework, including discovery of typical groups or sequences of public spending orders and tools for their visualization.

Documents, reports (14)

Report on best practices for software engineering, publishing and using procurement data
This will provide a global review of best practices in the areas of the project’s agile development and implementation approach of software engineering, open data and procurement knowledge graphs.

Outreach strategy for communication and community engagement
The outreach strategy will lay out the project’s approach and measures to build and engage with target audiences and users. This deliverable also includes the provision of the project website and a supplementary fact sheet to inform interested users about the project’s approach and progress.

Business case requirements analysis and specifications
This deliverable will provide a report that will contain a set of tables summarising the initial requirements, a detailed roadmap and a basis for the products to be developed in each business case. It will consist of an initial testing of the available technology by key business case stakeholders, in order to assess the overall approach and the specification of each business case.

Innovation and Exploitation plan
The exploitation plan will specify TheyBuyForYou’s approach towards utilising and transferring the project’s tools and results.

Data provider engagement report
A report on the data provider engagement activities that have been run throughout the project, including a plan for further engagement activities beyond the end of the project and lessons learned.
Interim impact report
This deliverable will provide a provisional report on the impact of community dissemination and engagement activities conducted at this stage.

White paper on the publication and governance of open procurement data
The white paper will define requirements and evidence-based standards for the publication, governance and collaborative curation of open procurement data.

Open-domain, cross-lingual automatic storytelling methods
Here we will add to D4.3 to handle heterogeneous content in multiple domains and languages. The actual software will be delivered also at M24, but part of D4.6.

Development & evaluation report of tools & business cases v1
This deliverable will provide the results of evaluating the business cases, together with an updated roadmap for the sustainability of the deployed TheyBuyForYou technology, and provide a feedback to the technology providers in WPs 1 to 4.

Innovation and Exploitation report
The report will specify how the project’s tools and results have been exploited through the course of the project.

Evaluation of the infographic and storytelling generation technology
This deliverable describes the methods used during and results gained from the ideation process, documenting the developed concepts and an analysis of the prototypes regarding their expected strengths and weaknesses. Furthermore, it describes the results gathered during the evaluation activities conducted as part of T3.1. It specifically documents the performance of the different visualization concepts in the different tests related to the three scenarios and existing procurement datasets.

Data-shape aware automatic storytelling methods
Here we will document the methods and models used to identify commonly used patterns in structured datasets (e.g., CSV) and suggest templates for stories. The actual software will be delivered also at M12, but part of D4.6.

Knowledge graph publication and consumer engagement report
Here we will provide a report on the data consumer engagement activities that have been run with potential users. We will report on the usage of the procurement knowledge graph and provide a plan for continued engagement and usage facilitation activities beyond the end of the project. The final snapshot of the data will be provided here.

Final impact report
This deliverable will provide the final report on the impact of community dissemination and engagement activities conducted during the project. Furthermore, this deliverable will also include an economic impact assessment of the project.

Demonstrators, pilots, prototypes (1)

Procurement APIs and platform release v3
Final version based on new methods and data from WPs 1, 2, 3, and 4.

Open Research Data Pilot (1)

Data management plan
Report on how data will be handled in TheyBuyForYou during the project's lifetime and after its completion, in accordance with the guidelines on FAIR Data Management in H2020.

Publications

Conference proceedings (5)

An Overview of the TBFY Knowledge Graph for Public Procurement
Author(s): Ahmet Soylu, Brian Elvesæter, Philip Turk, Dumitru Roman, Oscar Corcho, Elena Simperl, Ian Makgill, Chris Taggart, Marko Grobelnik and Till C. Lech
Published in: Proceedings of Poster and Demos of International Semantic Web Conference (ISWC) 2019, 2019

Attacking Diophantus - Solving a Special Case of Bag Containment
Author(s): George Konstantinidis, Fabio Mogavero
Published in: Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems - PODS '19, 2019, Page(s) 399-413
DOI: 10.1145/3294052.3319689

Towards Integrating Public Procurement Data into a Semantic Knowledge Graph
Predictive Data Transformation Suggestions in Grafterizer using Machine Learning

Author(s): Ahmet Soylu, Óscar Corcho, Elena Simperl, Dumitru Roman, Francisco Yedro Martínez, Chris Taggart, Ian Makgill, Brian Elvesæter, Ben Symonds, Helen McNally, George Konstantinidis, Yuchen Zhao, Till C. Lech
Published in: Proceedings of the EKAW 2018 Posters and Demonstrations Session co-located with 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW), November 2018, Nancy, France, 2018

Scalable Cross-lingual Document Similarity through Language-specific Concept Hierarchies

Author(s): Carlos Badenes-Olmedo, José Luis Redondo-García, Oscar Corcho
Published in: Proceedings of the 10th International Conference on Knowledge Capture - K-CAP '19, 2019, Page(s) 147-153
DOI: 10.1145/3360901.3364444

Book chapters (2)

Towards an Ontology for Public Procurement Based on the Open Contracting Data Standard

Author(s): Ahmet Soylu, Brian Elvesæter, Philip Turk, Dumitru Roman, Oscar Corcho, Elena Simperl, George Konstantinidis, Till Christopher Lech
DOI: 10.1007/978-3-030-29374-1_19

Towards a Knowledge Graph Based Platform for Public Procurement

Author(s): Elena Simperl, Oscar Corcho, Marko Grobelnik, Dumitru Roman, Ahmet Soylu, María Jesús Fernández Ruíz, Stefano Gatti, Chris Taggart, Urška Skok Klima, Annie Ferrari Uliana, Ian Makgill, Till Christopher Lech
Published in: Metadata and Semantic Research - 12th International Conference, MTSR 2018, Limassol, Cyprus, October 23-26, 2018, Revised Selected Papers, Issue 846, 2019, Page(s) 317-323
DOI: 10.1007/978-3-030-14401-2_29
Datasets

Datasets via OpenAIRE (6)

TheyBuyForYou Knowledge Graph data dump
Author(s): TBFY
Published in: Zenodo

TheyBuyForYou Knowledge Graph data dump
Author(s): TBFY
Published in: Zenodo

TheyBuyForYou Knowledge Graph data dump
Author(s): TBFY
Published in: Zenodo

TheyBuyForYou Knowledge Graph data dump
Author(s): TBFY
Published in: Zenodo

Crowdsourcing Document Similarity Judgements
Author(s): Amaral, Gabriel Maia Rocha
Published in: Zenodo

JSON data from the OpenOpps and OpenCorporates API
Author(s): TBFY
Published in: Zenodo

Software

Software via OpenAIRE (6)

librairy/jrc-en-model
Author(s): Badenes-Olmedo, Carlos
DOI: 10.5281/zenodo.3543446
Publisher: Zenodo

librairy/jrc-es-model