



2016 RIVA DEL GARDA



OpenTED Browser: Insights into European Public Spending

Yann-Aël Le Borgne ¹, Adriana Homolova ², Gianluca Bontempi ¹

¹ Machine Learning Group,
Université Libre de Bruxelles, Brussels, Belgium

² Independent Data Journalist, Amsterdam, Netherland



Public procurement

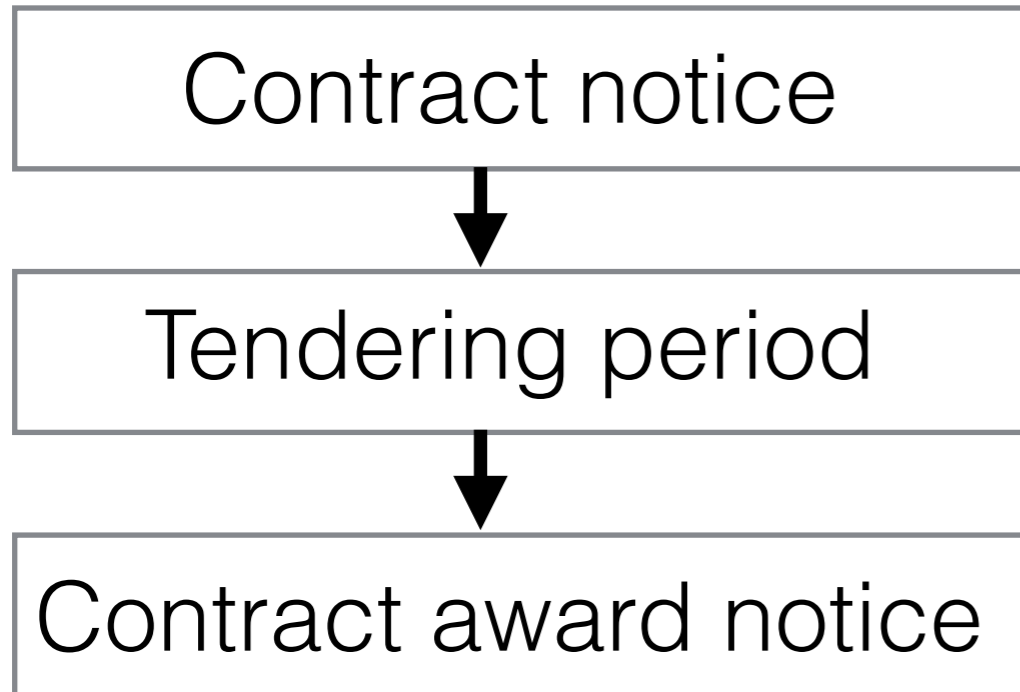
“Public procurement refers to the process by which public authorities, such as government departments or local authorities, purchase work, goods or services from companies.”

- DG GROW, Internal Market, Industry, Entrepreneurship and SMEs.

Every year, over 250 000 public authorities in the EU spend around **400 billions euros** through public procurements.

One third of OECD total government expenditure.

Procurement process



Publication on
Tenders Electronic Daily
(**'TED'**)

ted.europa.eu/TED

- The procedure is intended to be fully transparent
- For projects above about €100K, a contract notice must be published in Supplement S of the Official Journal of the European Union

OpenTED

<http://ted.openspending.org>

- Initiative started in 2013 as part of the European Investigative Journalism Conference, to lower the barrier to the access and analysis of TED data.

OpenTED Insight into European Tenders

Welcome

Welcome to OpenTED, an initiative that aims to make information from the joint European procurement system more easily accessible to journalists and researchers. The project aims to answer such questions as: Who are the biggest suppliers to a local government? What goods and services are the public bodies in a particular country purchasing? Are anti-corruption procedures being followed correctly?

As of 2016, the European Commission has published machine-readable CSV bulk extracts of the TED data on its open data portal.

This makes the original objective of the OpenTED project — providing such CSV data — unnecessary. Hooray!

- Welcome
- What is TED?
- Get the data
- Get Involved
- Related work
- About us

TED open data



European Union Open Data Portal

EUROPA > Open Data Portal > Data > Publisher > Directorate-General for Internal... >

Data Applications Linked Data Developers' corner About

Tenders Electronic Daily (TED) (csv subset) - procurement notices

» Publisher

Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs

Resources		
DOWNLOAD	TED - Contract award notices 2006	CSV
DOWNLOAD	TED - Contract award notices 2007	CSV
DOWNLOAD	TED - Contract award notices 2008	CSV
DOWNLOAD	TED - Contract award notices 2009	CSV
DOWNLOAD	TED - Contract award notices 2009-2015	CSV
DOWNLOAD	TED - Contract award notices 2010	CSV
DOWNLOAD	TED - Contract award notices 2011	CSV
DOWNLOAD	TED - Contract award notices 2012	CSV
DOWNLOAD	TED - Contract award notices 2013	CSV
DOWNLOAD	TED - Contract award notices 2014	CSV
DOWNLOAD	TED - Contract award notices 2015	CSV
DOWNLOAD	TED - Contract notices 2006	CSV
DOWNLOAD	TED - Contract notices 2007	CSV
DOWNLOAD	TED - Contract notices 2008	CSV
DOWNLOAD	TED - Contract notices 2009	CSV
DOWNLOAD	TED - Contract notices 2009-2015	CSV
DOWNLOAD	TED - Contract notices 2010	CSV
DOWNLOAD	TED - Contract notices 2011	CSV
DOWNLOAD	TED - Contract notices 2012	CSV
DOWNLOAD	TED - Contract notices 2013	CSV
DOWNLOAD	TED - Contract notices 2014	CSV
DOWNLOAD	TED - Contract notices 2015	CSV

- Since 2015: Available as CSV files on the EU Open Data portal thanks to DG-GROW.
- Provides both contract notices and contract award notices.
- About 4GB of data

TED open data

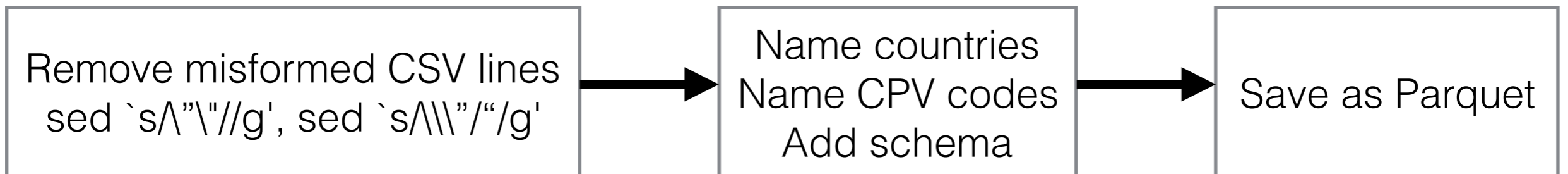
- Cumbersome to deal with GBs of data. Excel/R will show their limits. Most non-IT users (entrepreneurs, journalists, citizen, ...) struggle to browse the data.
- Example of query: “Find the biggest supplier for pharmaceutical products in Portugal in 2006-2015”.



- Aim of OpenTED browser: To make it easy to filter and visualise partnerships in procurement data

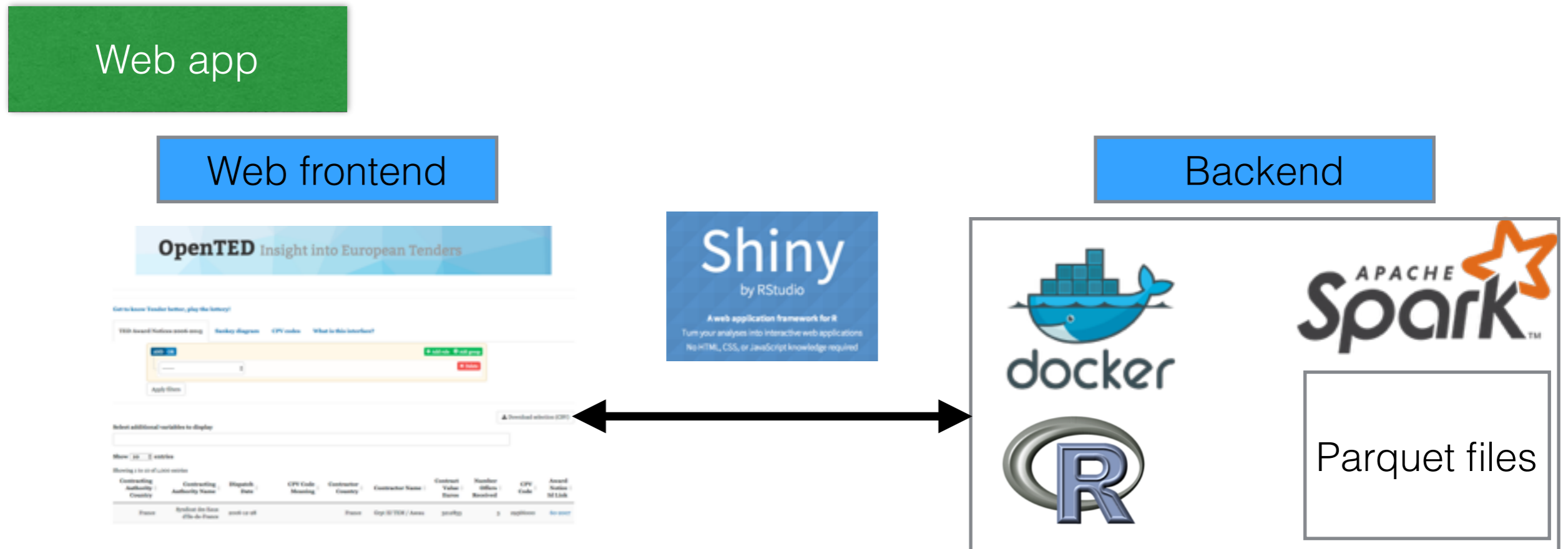
OpenTED browser

Preprocessing



- Apache Parquet - Benefits of databases without DB infrastructure:
 - Allow SQL querying and distributed processing.
 - Compressed: Reduce data size by factor 7 (From 2GB to 300MB).
 - Columnar data format.

OpenTED browser



- Shiny: High-level, reactive Web application framework for R. Code is less than 400 lines (frontend: ~80 lines, backend: ~250 lines).
- Docker: App is self-contained, reproducible.
- All code on Github - <https://github.com/Yannael/OpenTED>

OpenTED browser - Demo

OpenTED Insight into European Tenders

Get to know Tender better, play the lottery!

TED Award Notices 2006-2015 Sanky diagram CPV codes What is this interface?

AND OR + Add rule + Add group

Contracting_Authority_Country equal Portugal ✕ Delete

CPV_Code begins with 336 ✕ Delete

Dispatch_Date greater or equal 2012-01-01 ✕ Delete

Dispatch_Date less or equal 2013-12-31 ✕ Delete

Apply filters

Show 20 entries

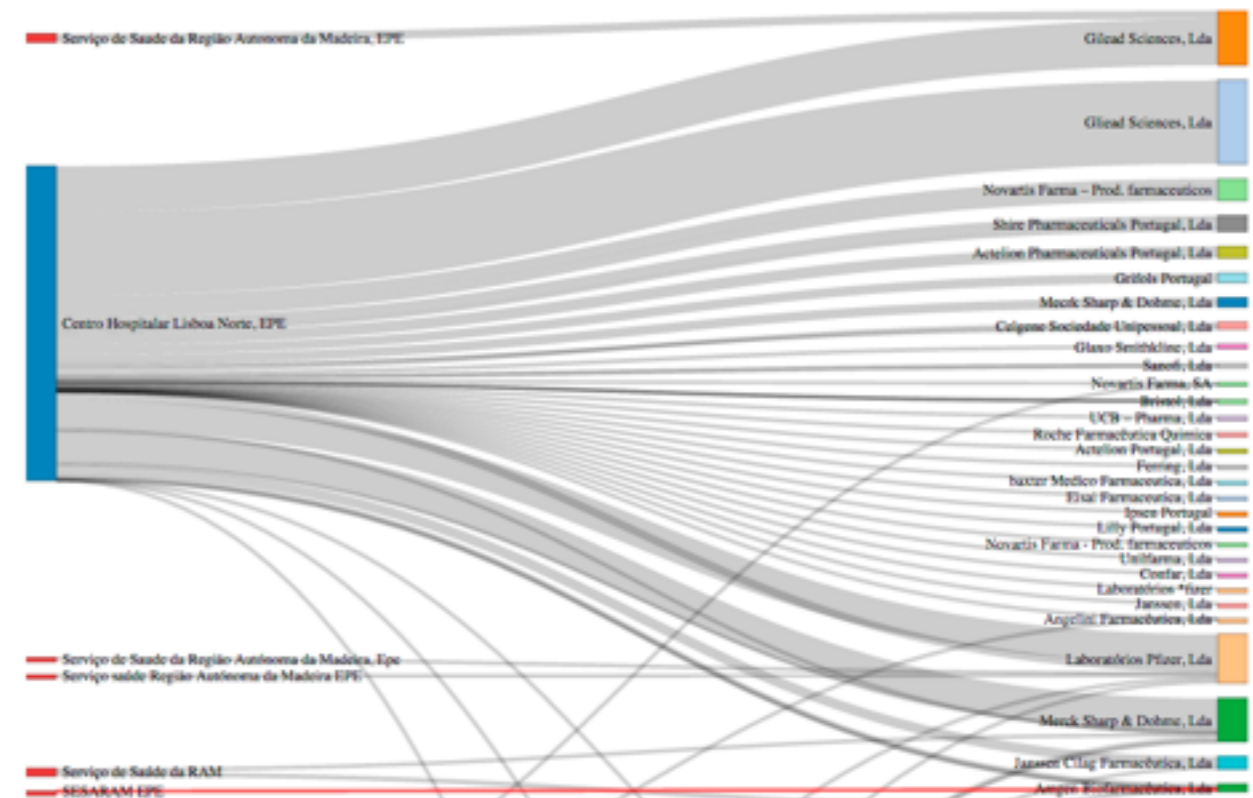
Showing 1 to 10 of 302 entries

Contracting Authority Country	Contracting Authority Name	Dispatch Date	CPV Code Meaning	Contractor Country	Contractor Name	Contract Value Euros	Number Offers Received	CPV Code	Award Notice Id Link
Portugal	Serviço Saúde Região Autónoma da Madeira, EPE	2012-07-27	Various medicinal products	Portugal	Roche Farmacêutica Quimica, Lda	668037	1	33690000	231998-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Glucose solutions					33692700	235648-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Medicinal products for the blood and blood-forming organs					33621000	235652-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Antineoplastic agents					33652100	235653-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Pharmaceutical products					33600000	235654-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Pharmaceutical products					33600000	235656-2012
Portugal	Serviços Partilhados do Ministério da Saúde, EPE	2012-08-08	Contraceptives					33641400	235657-2012

TED Award Notices 2006-2015 Sanky diagram CPV codes What is this interface?

Summary

Number of contracting authority:	66
Number of contractors:	268
Total number of contracts:	302
Number of contracts (values=0):	224
Total value of contracts (€):	84795435



Open issues and future work

- (Further) Data curation:
 - Disambiguation: Siemens, Seimens, Siemens S.A., Siemens SA, ...
 - Missing/incorrect values

- Analytics:
 - Indicators of corruption

Fazekas, Mihály, István János Tóth, and Lawrence Peter King. "An Objective Corruption Risk Index Using Public Procurement Data." *European Journal on Criminal Policy and Research* (2016): 1-29.

<http://www.redflags.eu/>

<http://digiwhist.eu>



- Graph-based anomaly detection

Akoglu, Leman, Hanghang Tong, and Danai Koutra. "Graph based anomaly detection and description: a survey." *Data Mining and Knowledge Discovery* 29.3 (2015): 626-688.

Thanks for your
attention!

Questions?

<http://mlg.ulb.ac.be>

<http://yleborgne.net/opented>