

# Automated analysis with event log enrichment of the European public procurement processes

2nd International Workshop on **Digital Justice, Digital Law, and Conceptual Modeling (JUSMOD)**

*6 November 2023*

Roberto Nai<sup>1</sup>, Emilio Sulis<sup>1</sup>, Laura Genga<sup>2</sup>

<sup>1</sup>Computer Science Department - University of Turin

<sup>2</sup>Industrial Engineering and Innovation Sciences - Eindhoven University of Technology (TU/e), The Netherlands

# Agenda

- Introduction
- Case study
- Methodology
- Results
- Conclusion and future work



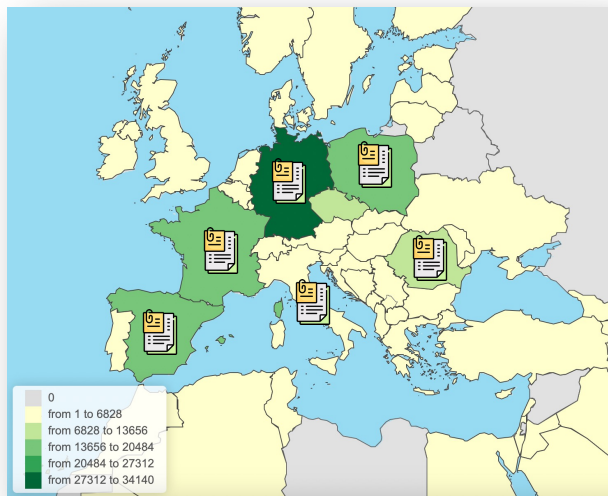
# Introduction

- Public tender processes are often lengthy, which has an impact on the legal and political system as a whole.



# Introduction

- In addition, special attention is paid to the **European landscape**, where there is difficulty in standardizing the legislative framework because of the different characteristics and practical applications peculiar to each member state.



# Introduction

- Starting from *timed activities recorded in public administration information systems* we propose the adoption of automated **Process Mining (PM)** technique to:

**RQ1** grasp relevant insights on public tenders;

**RQ2** comparing the procedures of public tenders in different European member states;

**RQ3** enrich the *legal event log* by extracting features from texts of tender notices, using Natural Language Processing (NLP) techniques.

# Case study

- Legal data of tenders and processes related can be found in the **Tenders Electronic Daily (TED)** website, the European community collection system.

The screenshot shows the TED website interface. At the top, there's a navigation bar with links like 'Site map', 'About TED', 'Help', 'Legal notice', 'Cookies', 'Contact', and 'English (en)'. Below this is a search bar with the text 'Search TED for public tenders'. The main content area displays 'Search result' for 'EU solidarity with Ukraine'. It includes a table with columns: Document number, Description, Country, Publication date, and Deadline. The first result is '611994-2023' for 'Italy-Rome: Database services'.

| Document number | Description   | Country | Publication date | Deadline |
|-----------------|---|---------|------------------|----------|
| 611994-2023     | Italy-Rome: Database services<br>Type of buyer: Other<br>Notice type: Contract or concession award notice - standard regime<br>Type of procedure: Negotiated without prior call for competition<br>Type of contract: Services | ITA     | 10/10/2023       |          |



735 thousand notices a year



~670 billion euro



3 sector (works, services, supplies)



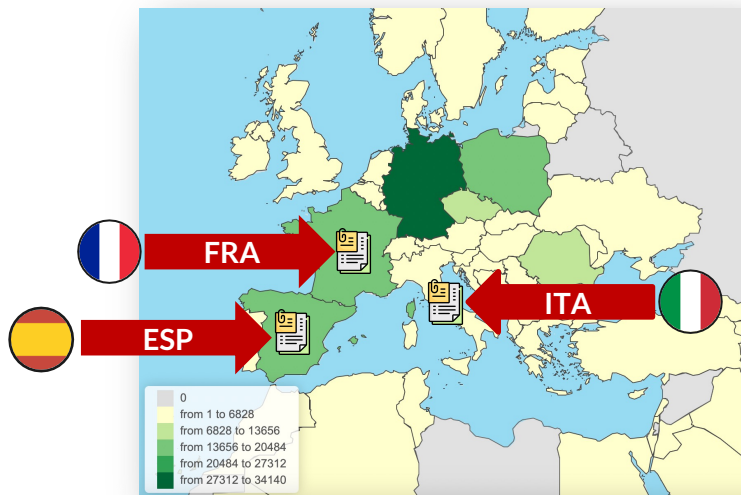
24 languages



<https://ted.europa.eu/TED/browse/browseByMap.do>

# Case study

- Following the suggestion of domain experts, we focused on regional authorities of the **French** (FRA), **Italian** (ITA), and **Spanish** (ESP) cases.
  - According to the domain experts, the three member states are similar in the legal review system.



# Case study

- Tender **metadata** entry available on TED (in **CSV** format):
  - document number (*key value*), NUTS (country / region / province), sector, amount, activities and dates.

Metadata



| Document number ▾  | Description   | Country ▾  | Publication date ▾ | Deadline ▾ |
|--------------------|---|------------|--------------------|------------|
| <u>673351-2023</u> | Italy - Energy-efficiency consultancy services - SERVIZIO DI OTTENIMENTO E GESTIONE DEI TITOLI DI EFFICIENZA ENERGETICA PER IL SUPERCOMPUTER EUROFUSION PRESSO IL DATA CENTER CINECA DI CASELECCHIO DI RENO E ALTRE FUTURE INSTALLAZIONI<br>Type of buyer: Body governed by public law<br>Notice type: Contract or concession notice - standard regime<br>Type of procedure: <u>Open</u><br>Type of contract: <u>Services</u> | <u>ITA</u> | 06/11/2023         | 04/12/2023 |



# Case study

- Tender data entry available on TED (in PDF format):
  - unstructured text containing tender's information divided into paragraphs.

Data



ITALY - SEWER CLEANING SERVICES - SERVIZIO DI CONTROLLO, PULIZIA, SPURGO, DISINTASAMENTO E VIDEOISPEZIONE DELLE RETI E DEGLI IMPIANTI DEL SERVIZIO IDRICO INTEGRATO DELL'EX ATO SAVONESE

196/2023

Contract or concession award notice – standard regime

## 1. Buyer

### 1.1 Buyer

*Official name:* [CONSORZIO PER LA DEPURAZIONE DELLE ACQUE DI SCARICO DEL SAVONESE S.P.A.] —

*Legal type of the buyer:* Public undertaking, controlled by a local authority

*Activity of the contracting authority:* Environmental protection

## 2. Procedure

### 2.1 Procedure

*Title:* SERVIZIO DI CONTROLLO, PULIZIA, SPURGO, DISINTASAMENTO E VIDEOISPEZIONE DELLE RETI E DEGLI IMPIANTI DEL SERVIZIO IDRICO INTEGRATO DELL'EX ATO SAVONESE

*Description:* LOTTO 1: "APPALTO PER IL SERVIZIO DI CONTROLLO, PULIZIA, SPURGO, DISINTASAMENTO E VIDEOISPEZIONE DELLE RETI E DEGLI IMPIANTI DEL SERVIZIO IDRICO INTEGRATO DELL'EX ATO SAVONESE - CENTRO OVEST 1 - (SETTORE DI LEVANTE)" LOTTO 2: "APPALTO PER IL SERVIZIO DI CONTROLLO, PULIZIA, SPURGO, DISINTASAMENTO E VIDEOISPEZIONE DELLE RETI E DEGLI IMPIANTI DEL SERVIZIO IDRICO INTEGRATO DELL'EX ATO SAVONESE - CENTRO OVEST 1 - (SETTORE DI PONENTE)"

*Procedure identifier:* 3cd026d6-91b3-4c1a-ac04-3e975b6b0761

*Type of procedure:* Open

### 2.1.1 Purpose

*Nature of the contract:* Services

*Main classification (cpv):* 90470000 Sewer cleaning services

### 2.1.4 General information

*Additional information:* Il responsabile del procedimento in fase di affidamento dott. Mirco Conterno

*Legal basis:*

Directive 2014/24/EU

## 5. Lot

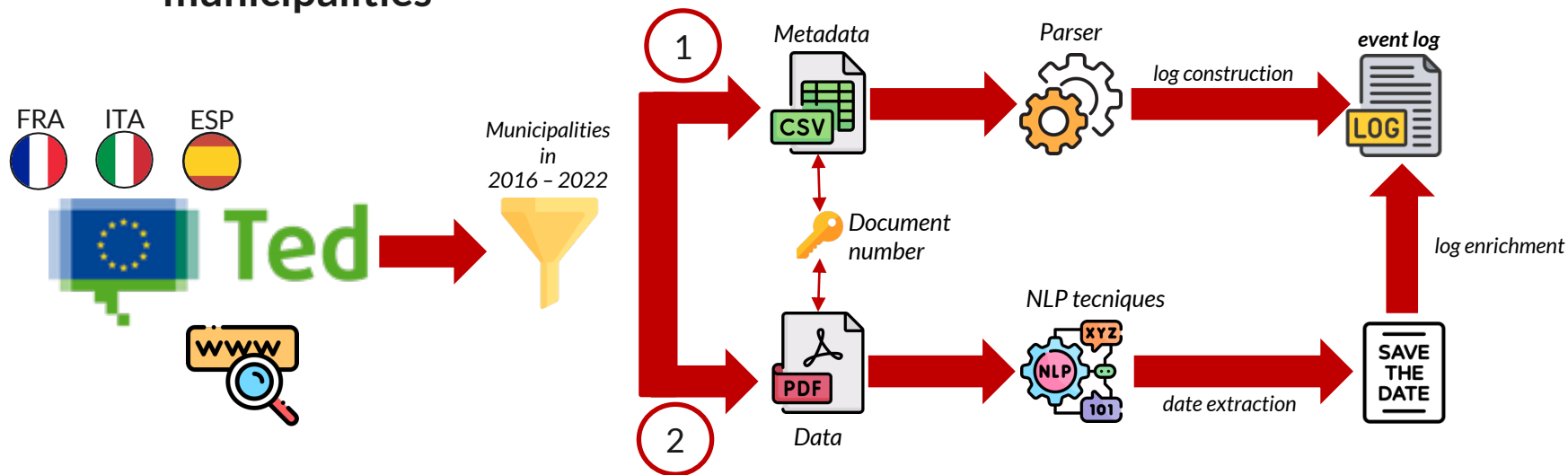
### 5.1 Lot: LOT-0001

*Title:* CIG: 9908883CEB

*Description:* LOTTO 1 - SETTORE DI LEVANTE: Comuni di Albisola Superiore, Albisola Marina, Bergeggi, Celle Ligure, Quiliano, Savona, Stella, Vado Ligure e Varazze LOTTO 1: € 689.000,00 oltre IVA di cui: € 260.000,00 inclusi costi per la sicurezza per il primo anno solare; € 260.000,00 inclusi costi per la sicurezza per eventuale rinnovo di max 1 anno; € 65.000,00 per eventuale proroga, stimata in massimo 90 (novanta) giorni necessaria alla conclusione delle procedure per l'individuazione di un nuovo contraente; € 104.000,00 per eventuali modifiche ai sensi dell'art. 106

# Methodology

- The event log was created from the collection of **annual tender** and **award** notices available on the TED portal.
  - Our final collection: **FRA**, **ITA**, and **ESP** cases from **2016** to **2022** for **municipalities**



# Methodology

- An **event log** is a set of **traces**, that includes at least *three* basic features:
  - 1) the *identifier* of the trace (**Case ID**);
  - 2) the name of the **Activity** which generated the event;
  - 3) the execution **Timestamp** of an activity.



**Case ID;Activity;Timestamp;Sector;Amount;Nuts;Country**

Trace 1

2017106814;PUBLICATION;2017-03-17;S;1035000.0;ITC13;IT  
2017106814;PARTICIPATION;2017-05-09;S;1035000.0;ITC13;IT  
2017106814;AWARD;2017-06-07;S;1035000.0;ITC13;IT  
2017106814;CONTRACT-START;2017-09-01;S;1035000.0;ITC13;IT  
2017106814;CONTRACT-END;2022-07-31;S;1035000.0;ITC13;IT

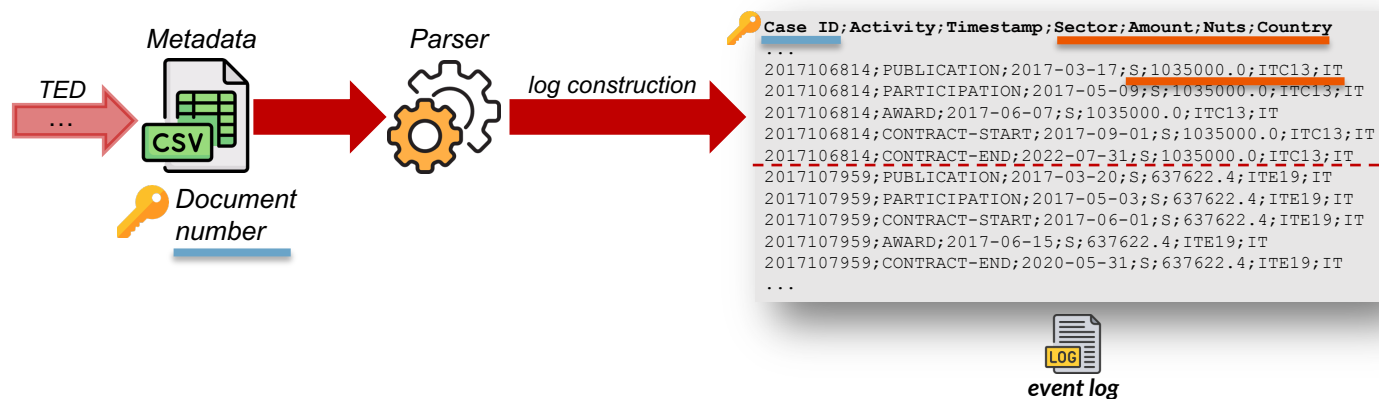
Trace 2

2017107959;PUBLICATION;2017-03-20;S;637622.4;ITE19;IT  
2017107959;PARTICIPATION;2017-05-03;S;637622.4;ITE19;IT  
2017107959;CONTRACT-START;2017-06-01;S;637622.4;ITE19;IT  
2017107959;AWARD;2017-06-15;S;637622.4;ITE19;IT  
2017107959;CONTRACT-END;2020-05-31;S;637622.4;ITE19;IT

...

# Methodology

- In our case study, tender document-number is the Case ID.
- In addition, we add in the event log file as attributes the information on the corresponding tender's sector, NUTS, country, amount.



# Methodology

- Starting from the **textual data** of the tenders, we found an event that was *not* available in the metadata, i.e. the **bids opening**.
- To enrich the event log with the bid opening, we dive into date extraction.

**Spark NLP**  
DateMatcher annotator

VS

^[a-z]\$\_

RegEx

ITALY - SEWER CLEANING SERVICES - SERVIZIO DI CONTROLLO, PULIZIA, SPIRUGO, DIRAMMENTAMENTO E VIDEOSPESSIONE DELLE RETI E DEGLI IMPIANTI DEL SERVIZIO IDRICO INTEGRATO DELL'EXATO SAVONESE

196/2023  
Contract or concession award notice - standard regime

1. Buyer  
Official name: [CONSORZIO PER LA DEPURAZIONE DELLE ACQUE DI SCARICO DEL SAVONESE S.P.A.] ---  
Legal type of the buyer: Public undertaking, controlled by a local authority  
Activity of the contracting authority: Environmental protection

---

**IV.3.8) Modalità di apertura delle offerte**  
Data: 18.1.2019 - 09:30  
Luogo:  
 Sede comunale degli uffici pubblica istruzione siti in Busto Arsizio, via Alberto da Giussano 12, 21052 Busto Arsizio.  
 Persone ammesse ad assistere all'apertura delle offerte: sì  
 Informazioni complementari sulle persone ammesse e la procedura di apertura: Legali rappresentanti/ procuratori delle imprese interessate oppure persone munite di specifica delega. In assenza di tali titoli, la partecipazione è ammessa come semplice uditore.

---

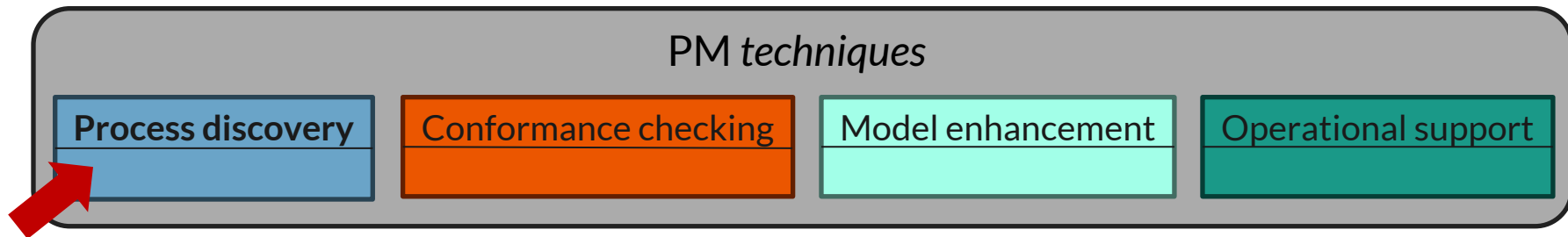
5. Lot  
5.1 Lot LOT-0001  
Title: CIG-969883CCEB  
Description: LOTTO 1 - SETTORE DI LEVANTE: Comuni di Abbiate Superiori, Abbiate Marina, Beregnago, Cella Ligure, Gallarate, Saronno, Stalla, Vado Ligure e Varesina LOTTO 1: € 689.000,00 oltre IVA di cui € 290.000,00 inclusi costi per la sicurezza per il primo anno solare; € 290.000,00 inclusi costi per la sicurezza per eventuale rinnovo di max 1 anno; € 65.000,00 per eventuale proroga, stimata in massimo 90 (inventari) giorni necessaria alla conclusione delle procedure per l'individuazione di un nuovo contraente; € 104.000,00 per eventuali modifiche ai sensi dell'art. 106

Page 1/5

BID-OPENING section

# Methodology

- PM *starts* from event log to answer a variety of process-related questions.



- Most of the commercial PM tools use **Directly-Follows Graph (DFG)** as a first means to *discover* the event log.
  - A DFG is a graph with *nodes* that correspond to *activities* and *directed edges (arcs)* describe the relationship between the various activities.



# Results

- As result, there are **six** activities in the process *discovered* via DFG:



- 1 - initial event is the **PUBLICATION** of the tender;
- 2 - the **PARTICIPATION** of individual entities proposing a bid;
- 3 - the adjudication of the tender (**AWARD**);
- 4 - the contract is issued (**CONTRACT- START**)
- 5 - the contract ending (**CONTRACT-END**);



- 6 - the **BID-OPENING** which takes place *before* AWARD.

# Results



FRA

Start

PUBLICATION  
instant

35 d

PARTICIPATION  
instant

54 d

AWARD  
instant

21 d

CONTRACT-START  
instant

12 mths

CONTRACT-END  
instant

End



Slow decision process



ESP

Start

PUBLICATION  
instant

34 d

PARTICIPATION  
instant

46 d

CONTRACT-START  
instant

17 d

AWARD  
instant

51.6 wks

CONTRACT-END  
instant

End



Slow decision process



ITA

Start

PUBLICATION  
instant

39 d

PARTICIPATION  
instant

36 d

AWARD  
instant

53 d

CONTRACT-START  
instant

29.2 mths

CONTRACT-END  
instant

End



Slow decision process



# Results

- Comparison of three countries' event logs by the number of cases, events, mean and median case duration (*months*), standard deviation (SD) for *services* sector.

| Country | Cases  | Events | Mean↓<br>(months) | Median<br>(months) | SD<br>(months) |
|---------|--------|--------|-------------------|--------------------|----------------|
| ITA     | 1,063  | 5,156  | 34.8              | 37.1               | 14.3           |
| FRA     | 22,102 | 4,601  | 25.7              | 17.1               | 16.9           |
| ESP     | 838    | 4,063  | 21                | 17.6               | 12.6           |

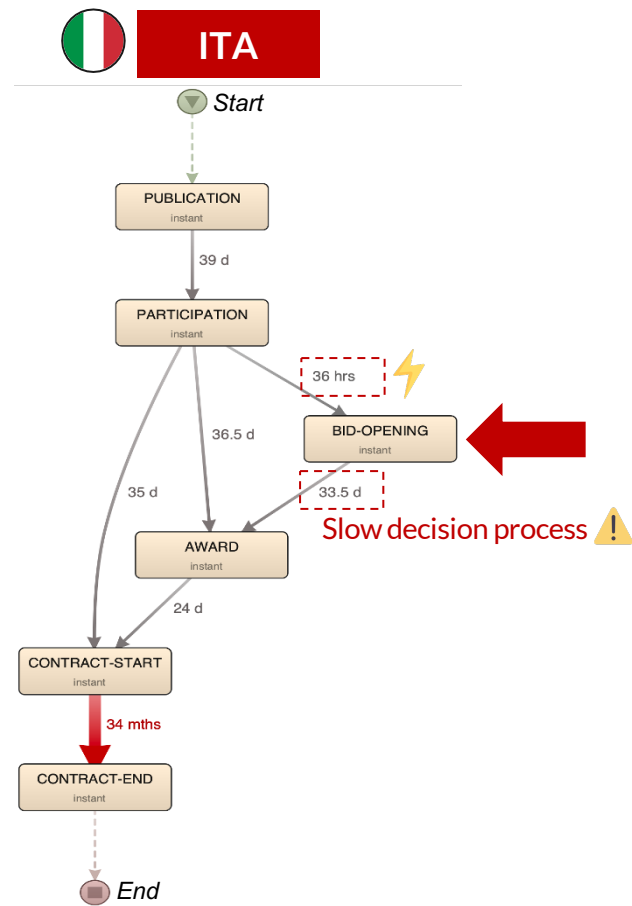
# Results

- Starting from the 1,063 total cases for the ITA country, in the text of tenders the presence of the bid opening section in 270 cases.
- In the remaining 270 cases, the NLP techniques (DateMatcher and RegEx) were applied and compared.

| Method      | Dates found↓ | Timing (min.sec) |
|-------------|--------------|------------------|
| RegEx       | 270 (100%)   | ~ 0.6            |
| DateMatcher | 223 (83%)    | ~ 1.1            |

# Results

- The new event BID-OPENING extracted from the tender notices shows how the opening of received bids **takes place in a short time** (transition from PARTICIPATION to BID-OPENING).



# Results

- Validation of the results was carried out with the domain experts through interviews.
- Domain experts confirmed the validity of the results:
  - to understand the quantity and timing of process activities;
  - to compare processes and timing from different countries;
  - to obtain other useful data from the texts automatically.

# Conclusion and future work




- **Applied PM techniques** to a legal process to understand its execution in reality.
- **Compared the country-based processing** in terms of time performance for municipalities.
- **Applied NLP tasks** to the tender texts to explore some information extraction techniques as well as to obtain a new event to add to the italian cases.



- Adopt **automated variant** analysis techniques using ProM *process comparator* module.
- Explore other NLP techniques to **extract more data from the texts** and enrich the process event log even more, **also for other countries**.
- Explore the **application of prediction algorithms** to predict in advance:
  - outcomes (e.g., is the process going to a compliant or not);
  - processes' duration;
  - next event (e.g. is the process going to subcontract or variant).

# End of presentation

*Thanks*

 roberto.nai@unito.it

