Ministero per i Beni e le Attività Culturali

Archivio di Stato di Venezia

# Beni archivistici e big data

## Il progetto GAWS:

## *Garzoni, Apprenticeship, Work, Society*

Andrea Erboso

CONFERENZA GARR 2018
Cagliari, 5 ottobre 2018

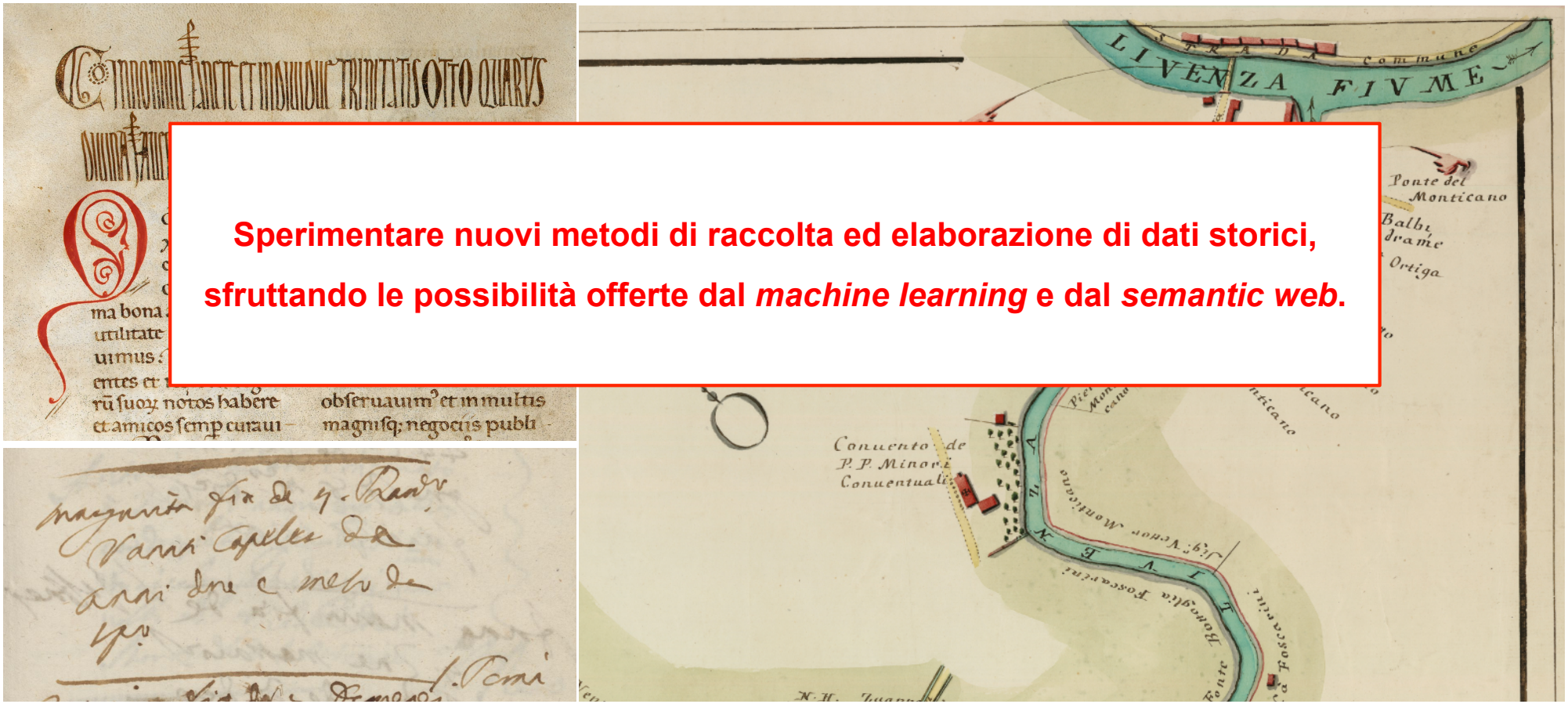
# Archivi: “big data” del passato

**Archivio di Stato di Venezia:**

- 120 km di scaffali di documenti
- Più di 800 fondi archivistici
- Più di 1100 anni di storia documentata

**Un numero incalcolabile di informazioni registrate in forma non strutturata**

# Archivi: “big data” del passato




**Sperimentare nuovi metodi di raccolta ed elaborazione di dati storici, sfruttando le possibilità offerte dal *machine learning* e dal *semantic web*.**

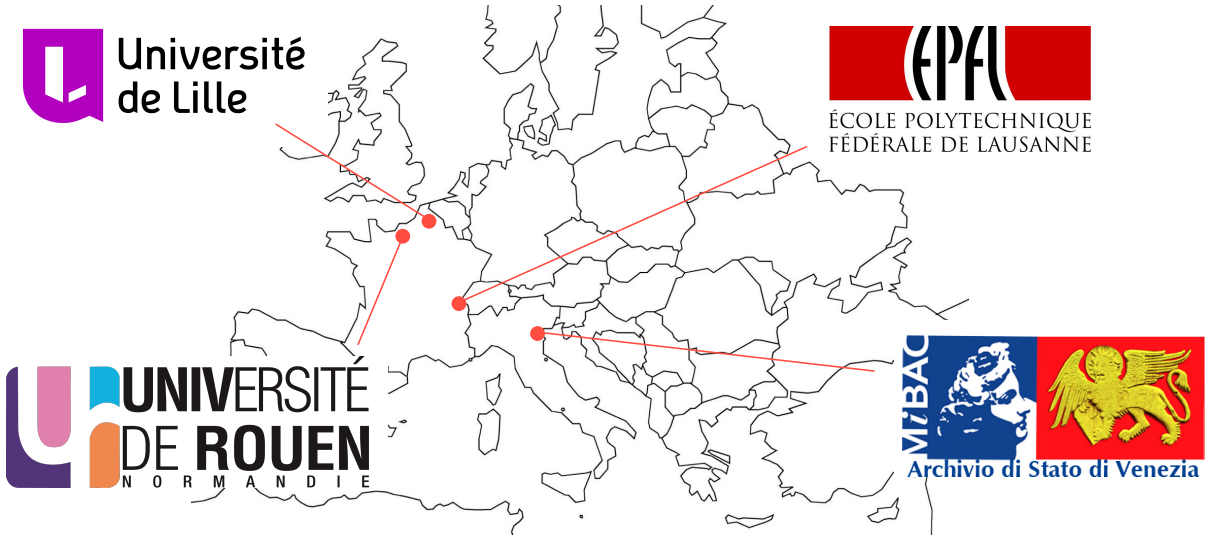
# Il progetto GAWS

## GAWS: *Garzoni, Apprenticeship, Work and Society in Early Modern Venice*

Progetto ANR/FNS 2014-2018

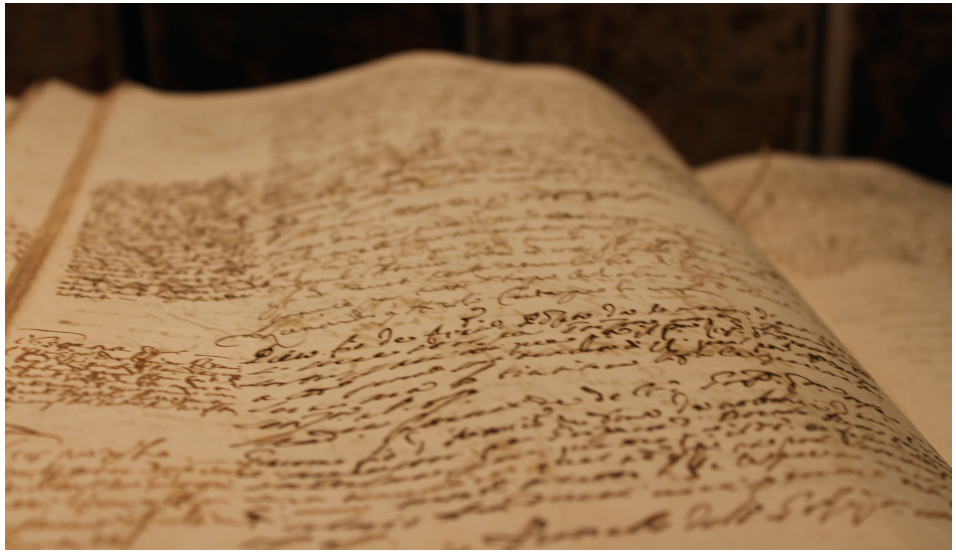
Principal Investigators: Anna Bellavitis, Frédéric Kaplan, Valentina Sapienza

https://garzoni.hypotheses.org/



# La fonte: gli Accordi dei Garzoni della Giustizia Vecchia

- 53.890 contratti di apprendistato tra GARZONE e MAESTRO
- 1300 professioni
- 32 registri per un totale di 14.236 carte
- Registrazioni effettuate dallo stesso ufficio pubblico
- Continuità cronologica (1575-1772)
- Serialità delle informazioni contenute



**OBIETTIVO:**

**Raccogliere le informazioni strutturandole affinché siano esplorabili**

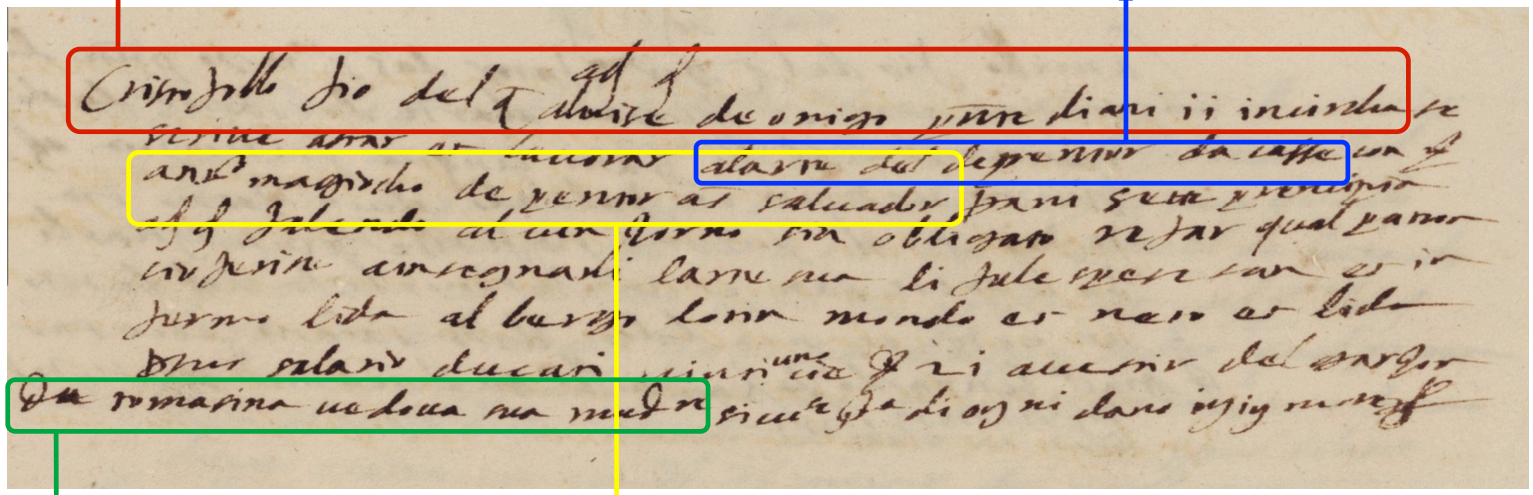
# Il metodo di lavoro

1. Digitalizzazione del materiale
2. Studio della fonte e modellizzazione delle informazioni
3. Sviluppo dell'ontologia e dell'architettura informatica
4. Sviluppo della maschera di inserimento e test
5. Inserimento dei dati / disambiguazione
6. Bonifica dei dati
7. Esplorazione dei dati

# Modellizzazione dei dati

**Garzone:** Cristofollo *quondam* Alvise de Onigo, di 11 anni

**Professione:** Arte del *depentor da casse*



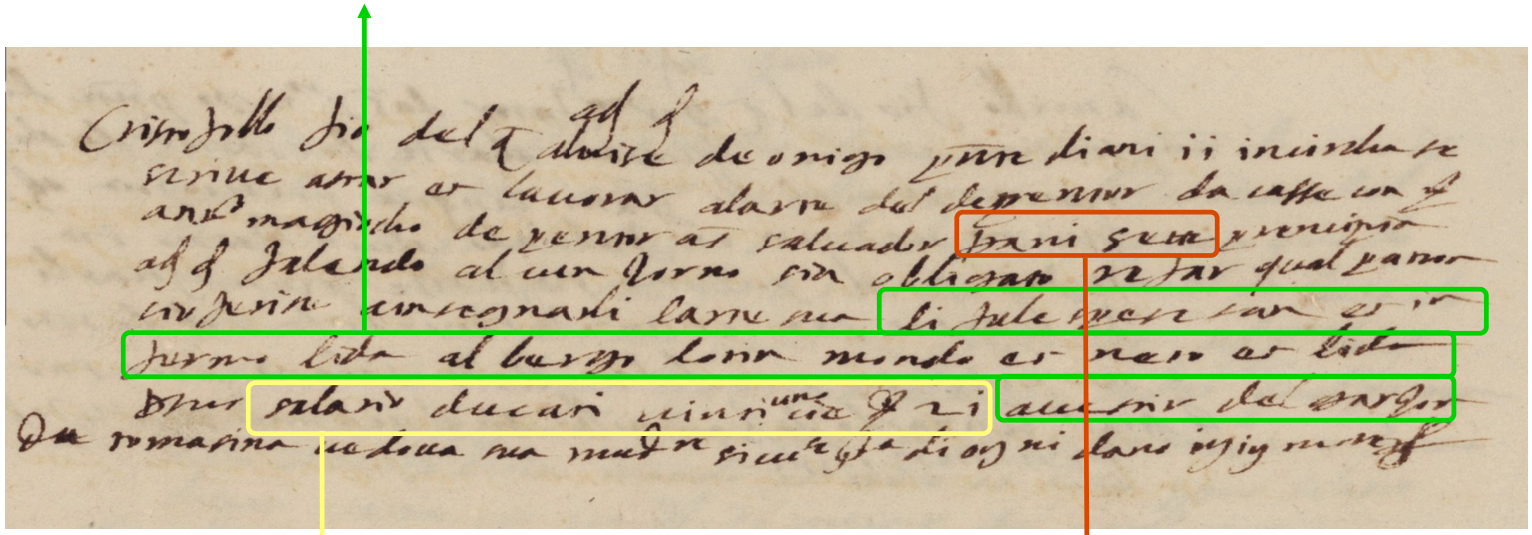
**Maestro:** Antonio Magiocho *depentor*, con bottega a San Salvador

**Garante:** Domina Tomasina vedova, madre del garzone

ASVe, *Giustizia Vecchia*, *Accordi dei Garzoni*, b. 112, reg. 152, c. 116v (29 novembre 1582)

# Modellizzazione dei dati

**Condizioni materiali dell'apprendistato:** spese, alloggio, cure personali...



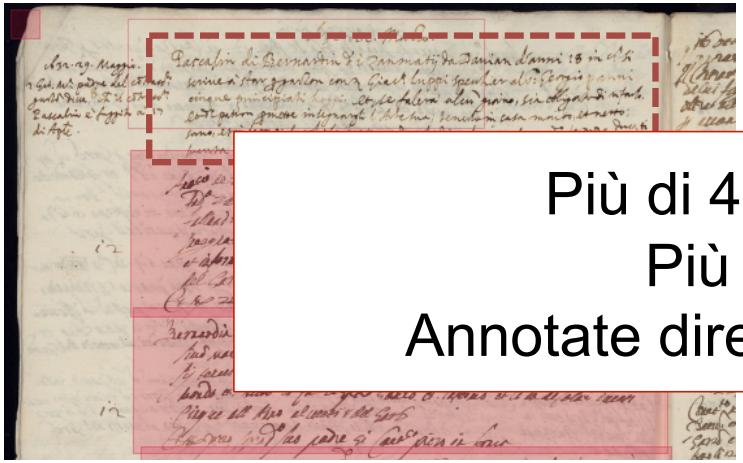
**Durata del contratto:** 7 anni

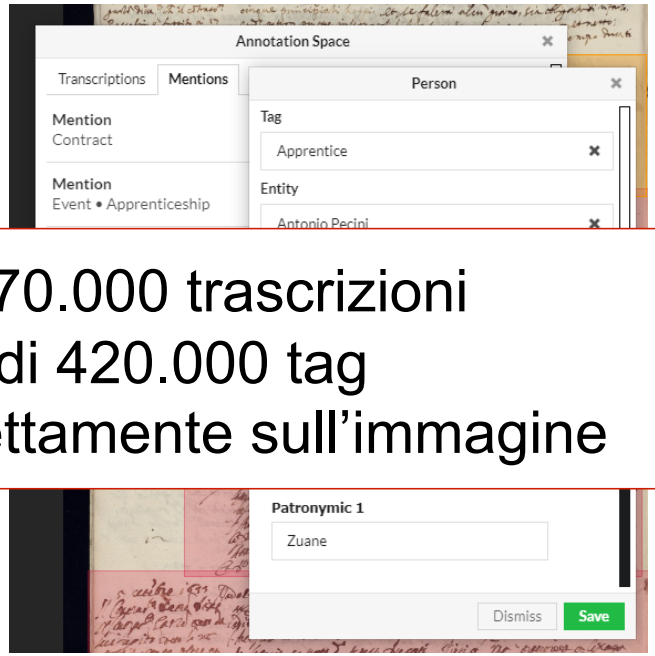
**Salario :** 21 ducati corrisposti in questo caso dal maestro al garzone

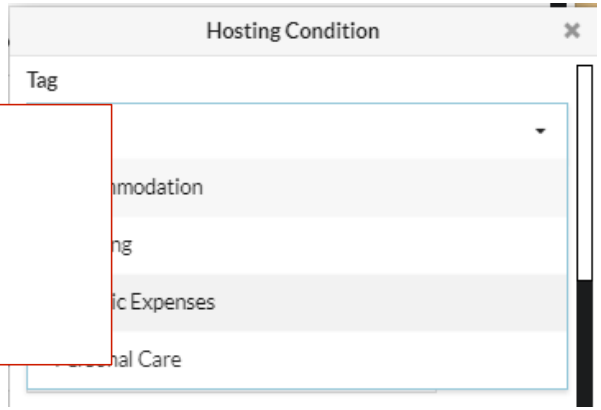
ASVe, *Giustizia Vecchia*, *Accordi dei Garzoni*, b. 112, reg. 152, c. 116v (29 novembre 1582)

# Inserimento dei dati

SEGMENTAZIONE DELLE IMMAGINI → TRASCRIZIONE DELLE INFORMAZIONI → INSERIMENTO DELLE TAG



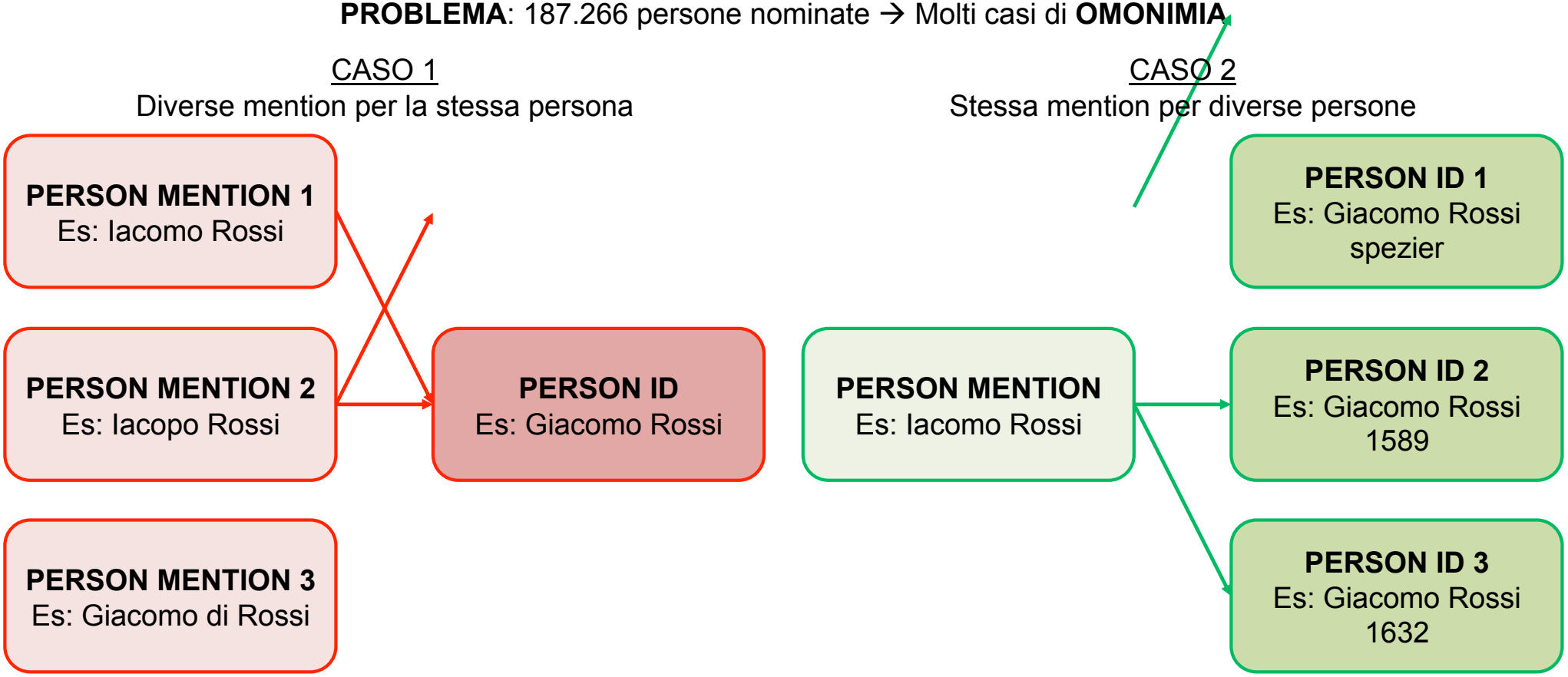



Più di 470.000 trascrizioni
Più di 420.000 tag
Annotate direttamente sull'immagine

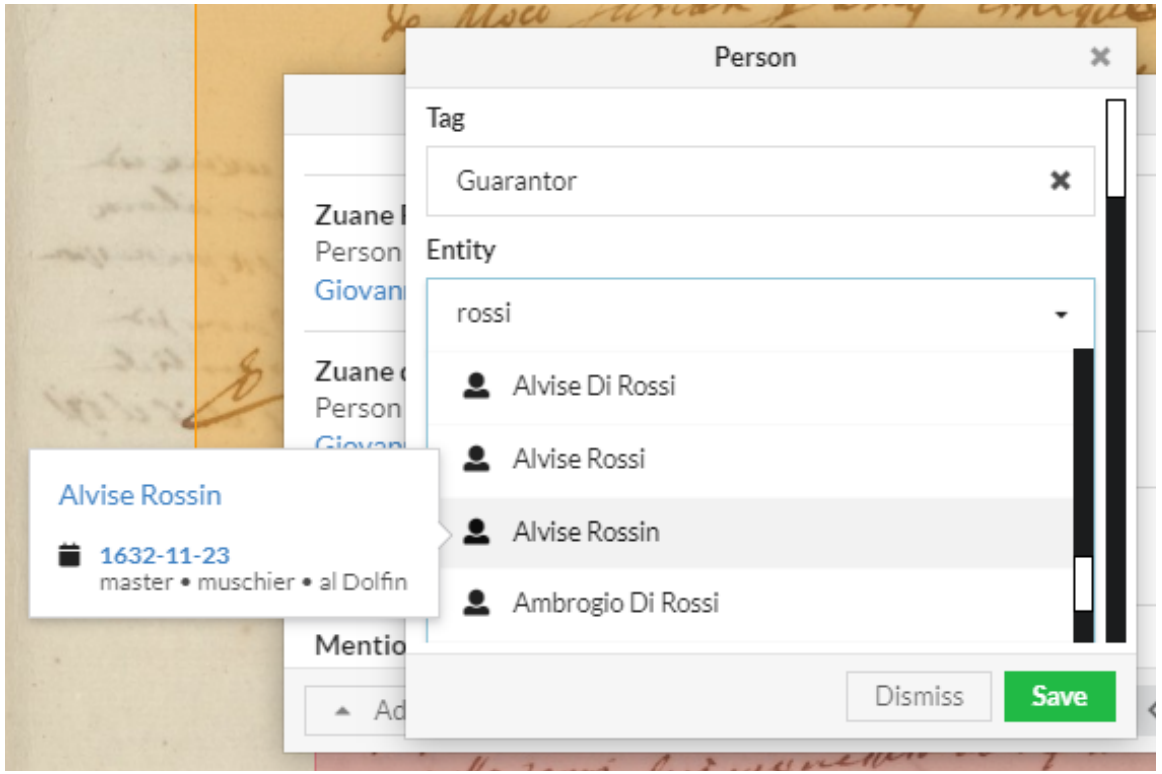
# Disambiguazione

**PROBLEMA**: 187.266 persone nominate → Molti casi di **OMONIMIA**

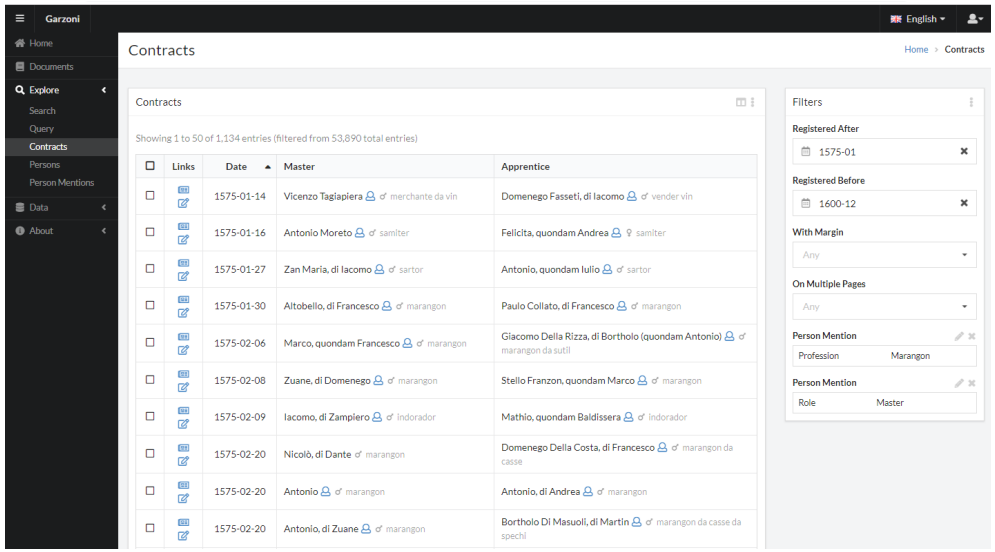


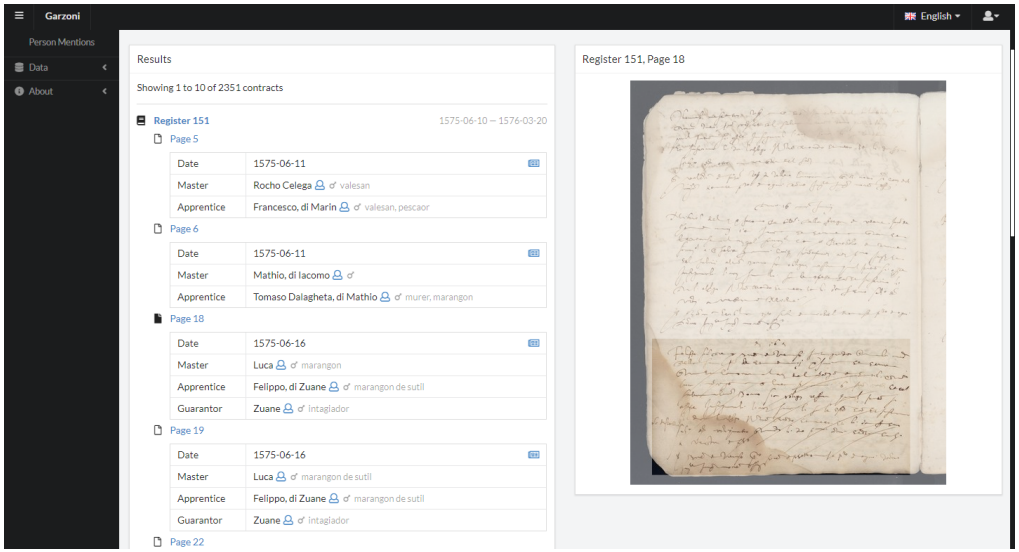
# Disambiguazione

**SOLUZIONE → MACHINE LEARNING**



# Esplorazione dei dati




- Ricerca testo libero
- Filtri
- Query SPARQL
- Segnatura archivistica / navigazione per immagini

- Risultati collegati all'immagine
- Consultazione online o estrazione dati (xlsx, ods, json)

# Conclusioni e Sviluppi futuri

## CONCLUSIONI

- Valorizzazione di una fonte archivistica
- Sviluppo di nuovi strumenti di esplorazione
- Verificabilità dei dati inseriti
- Team multidisciplinare

## SVILUPPI FUTURI

- Implementazione di nuovi dati da nuove serie archivistiche
- Geolocalizzazione dei dati topografici → rappresentazione grafica delle info
- Sistemi OCR per la lettura assistita

# Credits

## Main collaborators:

Anna Bellavitis – GRHis, Rouen University, France.
Riccardo Cella – GRHis, Rouen University, France.
Giovanni Colavizza – DHLAB, Lausanne, Switzerland.
Lucia Corazza – Venice, Italy.
Jacopo Cossu – Venice, Italy.
Davide Drago – Venice, Italy.
Maud Ehrmann – DHLAB, Lausanne, Switzerland.
Frederic Kaplan – DHLAB, Lausanne, Switzerland.
Claudio Lorenzini – University of Udine, Italy.
Valentina Sapienza – IRHiS, Lille, France.
Francesca Stopper – Venice, Italy.
Orlin Topalov – DHLAB, Lausanne, Switzerland.
Gloria Zuccarello – Venice, Italy.
Francesca Zugno – Venice, Italy.

## Associate collaborators:

Sandra Bazin – University of Paris 4, Paris, France.
Michela Dal Borgo – Archivio di Stato, Venice, Italy.
Bert De Munk – University of Anvers, Netherland.
Andrea Erboso – Archivio di Stato, Venice, Italy.
Martina Frank – University Ca'Foscari, Venice, Italy.
Giovanna Giubbini – Archivio di Stato, Venice, Italy.
Nikita Grishin – EPFL, Lausanne, Switzerland.
Michel Hochmann – Ecole Pratique des Hautes Etudes, Paris.
Matej Klemenčič – University of Ljubljana, Slovenia.
Nina Kudis – University of Rijeka, Croatia.
Corine Maitte – University of Marne-la-Vallée, France.
Camille Perez – Musée de Roanne, France.
Luciano Pezzolo – University Ca' Foscari, Venice.
Giorgio Tagliaferro – University of Warwick, UK.
Francesco Trentini – University Ca' Foscari, Venice, Italy.
Damir Tulic – University of Rijeka, Croatia.
Patrick Wallis – London School of Economics, London, UK.
Beatrice Zucca – Historienne.

AGENCE NATIONALE DE LA RECHERCHE ANR

FNSNF FONDS NATIONAL SUISSE SCHWEIZERISCHER NATIONALFONDS FONDO NAZIONALE SVIZZERO SWISS NATIONAL SCIENCE FOUNDATION

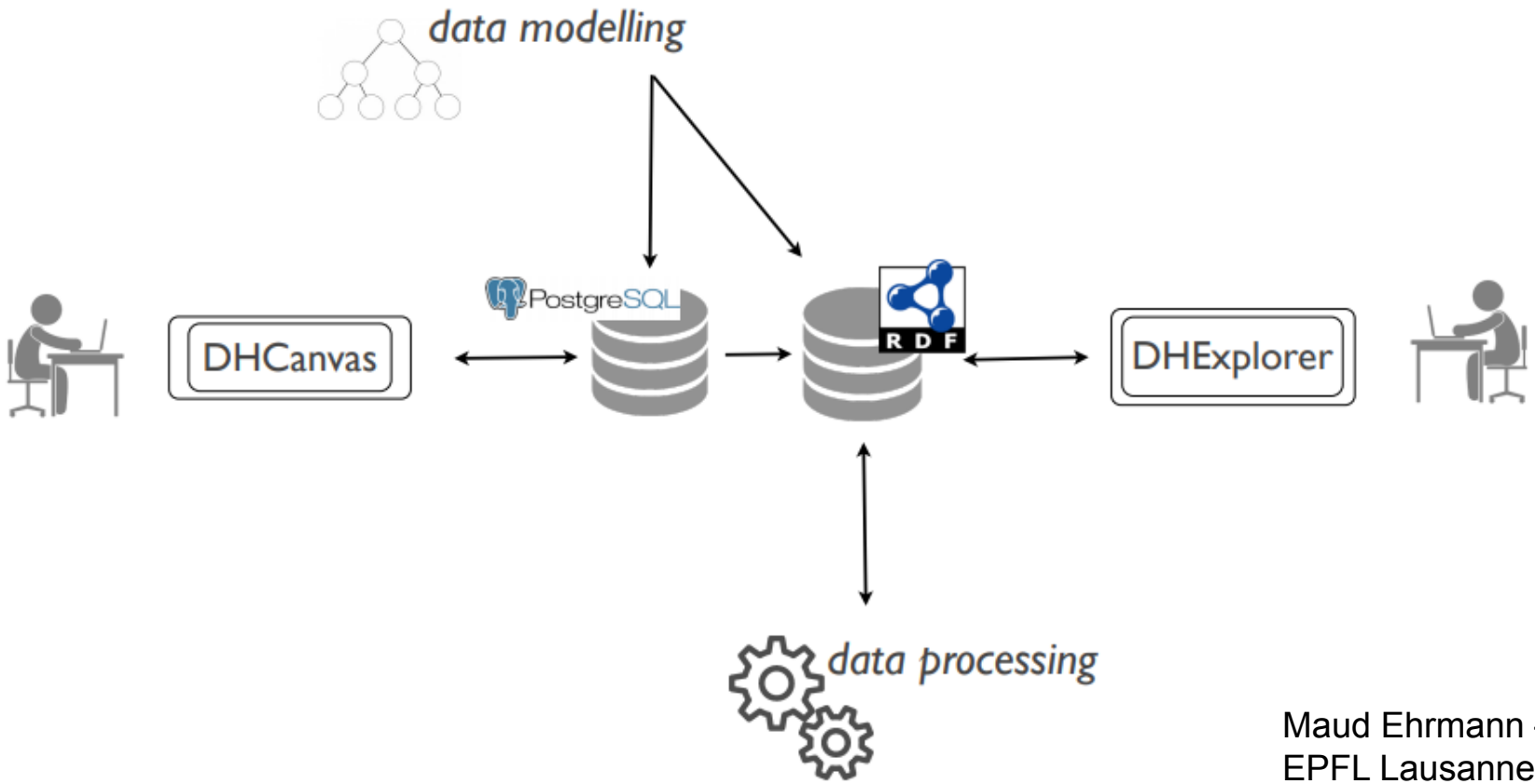
MiBAC Archivio di Stato di Venezia

Université de Lille

UNIVERSITÉ DE ROUEN NORMANDIE

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Architettura informatica




Maud Ehrmann – DHLAB
EPFL Lausanne