



Linked Data for content analytics in Celi

Semantics 2014 - Leipzig

Alessio Bosca



Speech
Applications



Semantic
Search

Opinion Mining



Text
Analytics



Social Media
Intelligence



Agenda

- ✓ **Presentation of Celi**
- ✓ **Technologies (and what we do with them)**
- ✓ **Focus on LOD for content analytics in Celi**
- ✓ **... what we'd like to do**



Adaptive Language
Technology

Enabling



Speech
Applications



Semantic
Search



Text
Analytics

Opinion Mining



Social Media
Intelligence



1999
CELI srl was born

2002
Speech Technology

2006
BlogMeter

2010
Milan, Rome,
Trento

2011
Cross Library

2013
Korean Market

1999

2005

2010



blogmeter





Speech
Applications



Semantic
Search



Text
Analytics



Opinion Mining



Social Media
Intelligence

4

Torino
Milano
Trento
Roma

Seats

6

Italy
Belgium
France
Spain
Corea
Poland

Markets

15

Years of experience

4

NLP components
Speech technology
Social Media Intelligence
Digital Humanities

Business branches

50

Employees +
Collaborators

>100

Active clients



Speech
Applications



Semantic
Search

Opinion Mining



Text
Analytics



Social Media
Intelligence



6

Scuola Normale Superiore
Università di Torino
Università di Pisa
Università di Trento
Fondazione Bruno Kessler
Politecnico di Milano

Agreements with research centers

15

Research projects

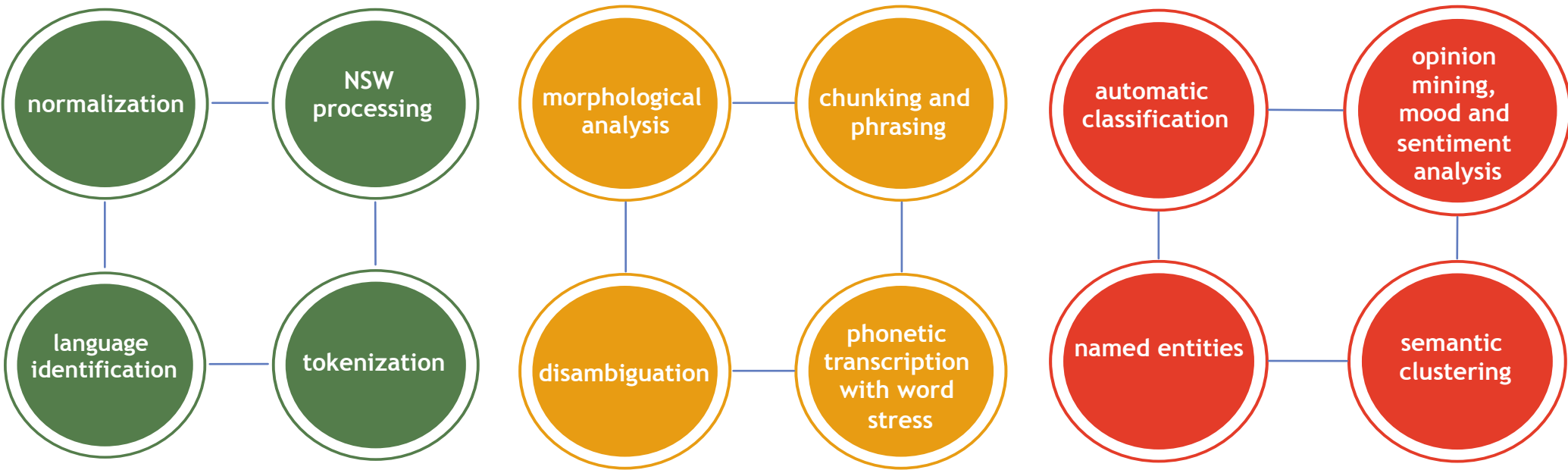
>50

Published papers

Relationships with the scientific community



Core technology



Techs



Kestrel



**Virtuoso
OpenSource**



Guava



Clients

Semantic Solutions

Speech Technology

Social Media Monitoring



"InfoCamere"

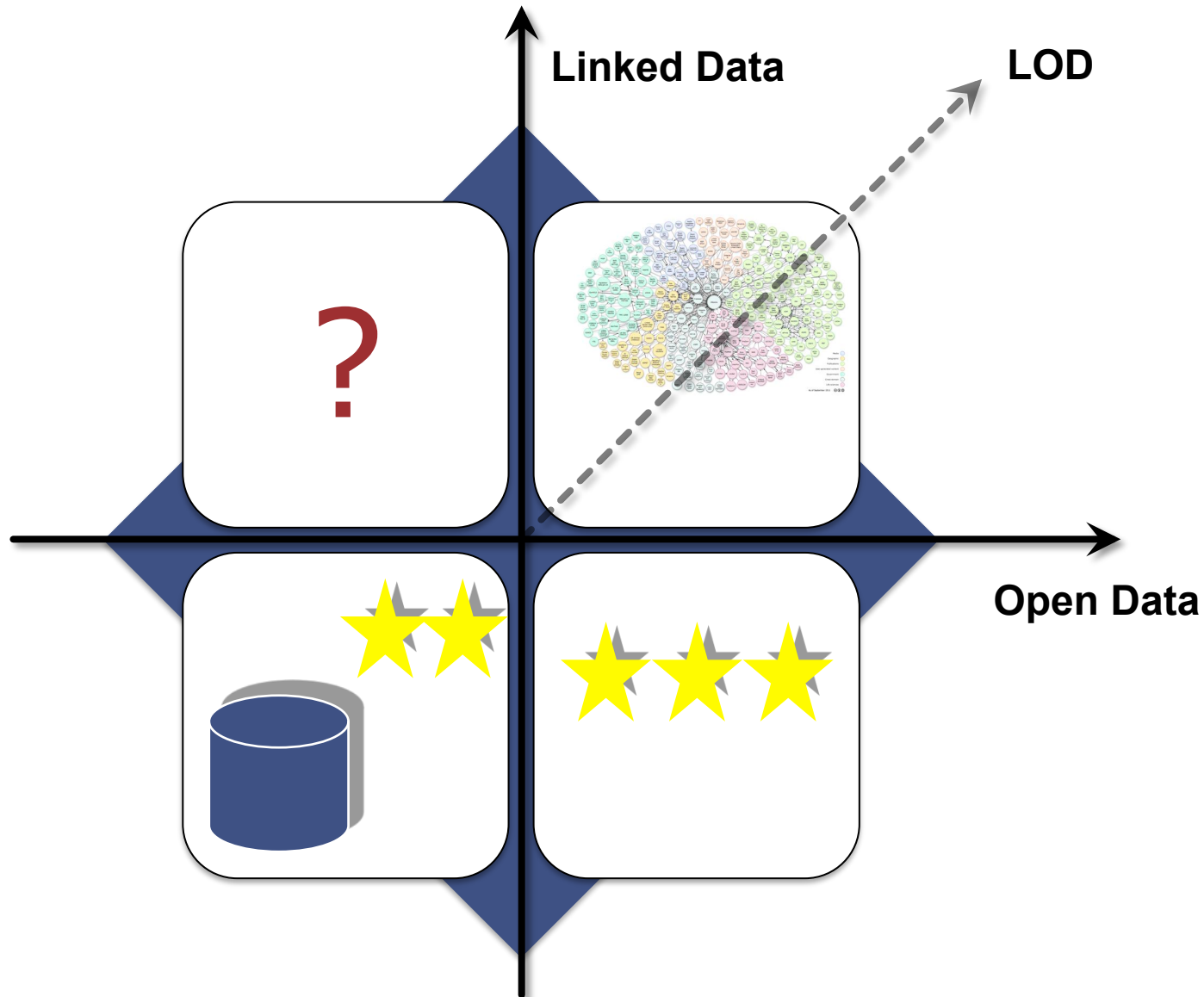
LA STAMPA



NUANCE



Linked (and/or Open) Data



Private Sector: how Celi exploits L(O)D

- **as user**

LODs as linguistic resources for NER, content enrichment, machine linking, discovery search...

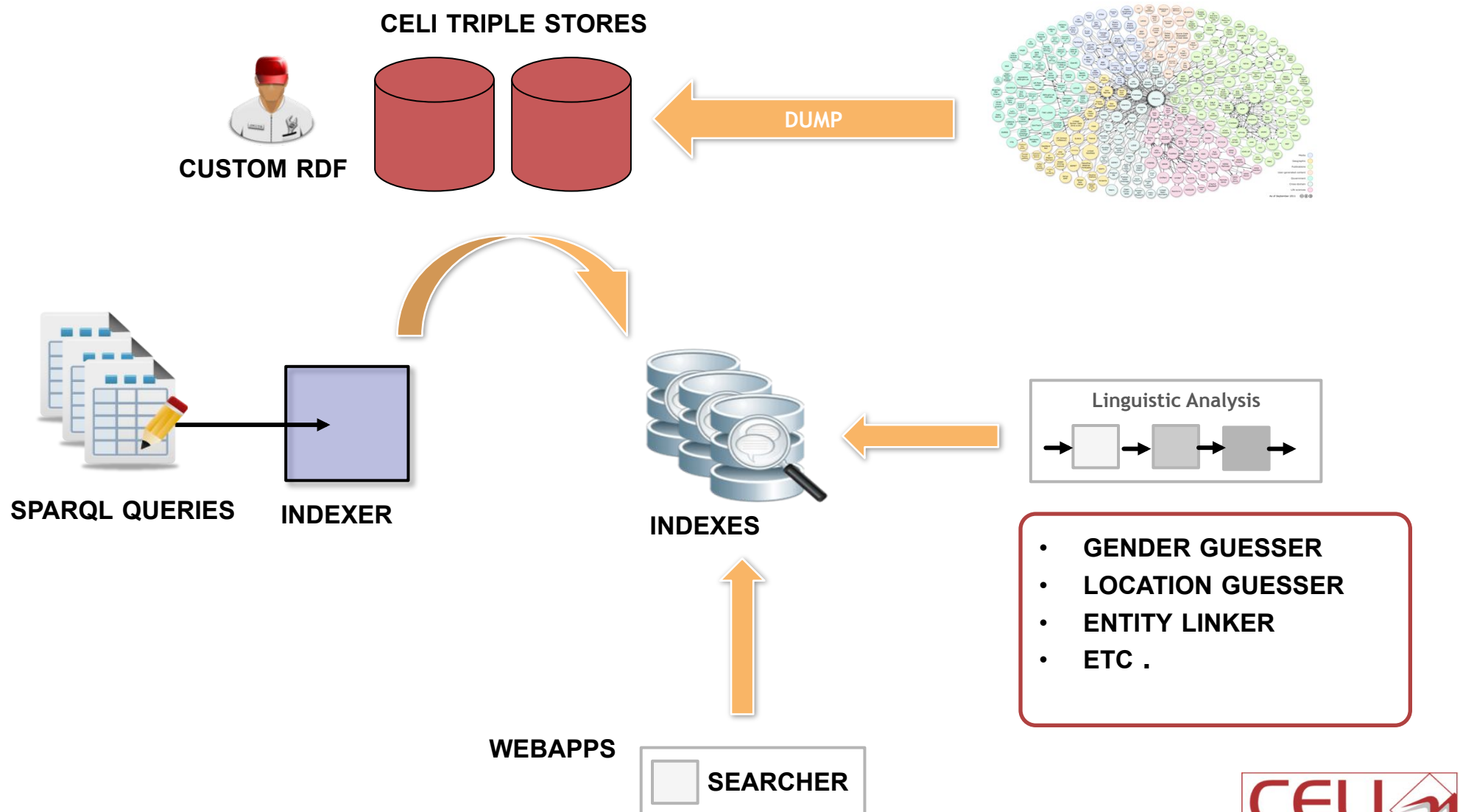
- **as provider** for the PA publishing, data integration

- **internal use** (e.g. assets management)

- **crafting** of RDF artifacts for custom projects and applications

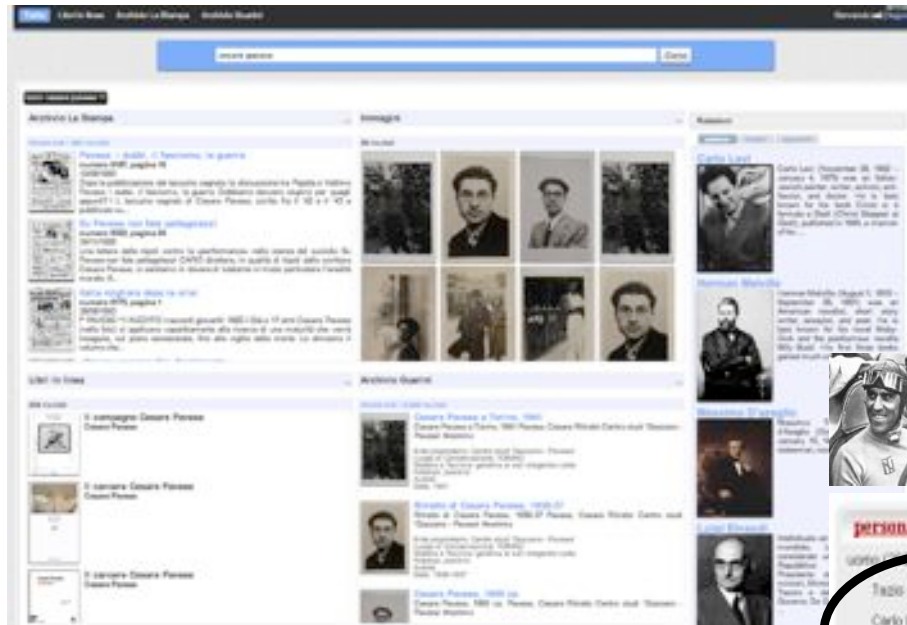


LOD for NER

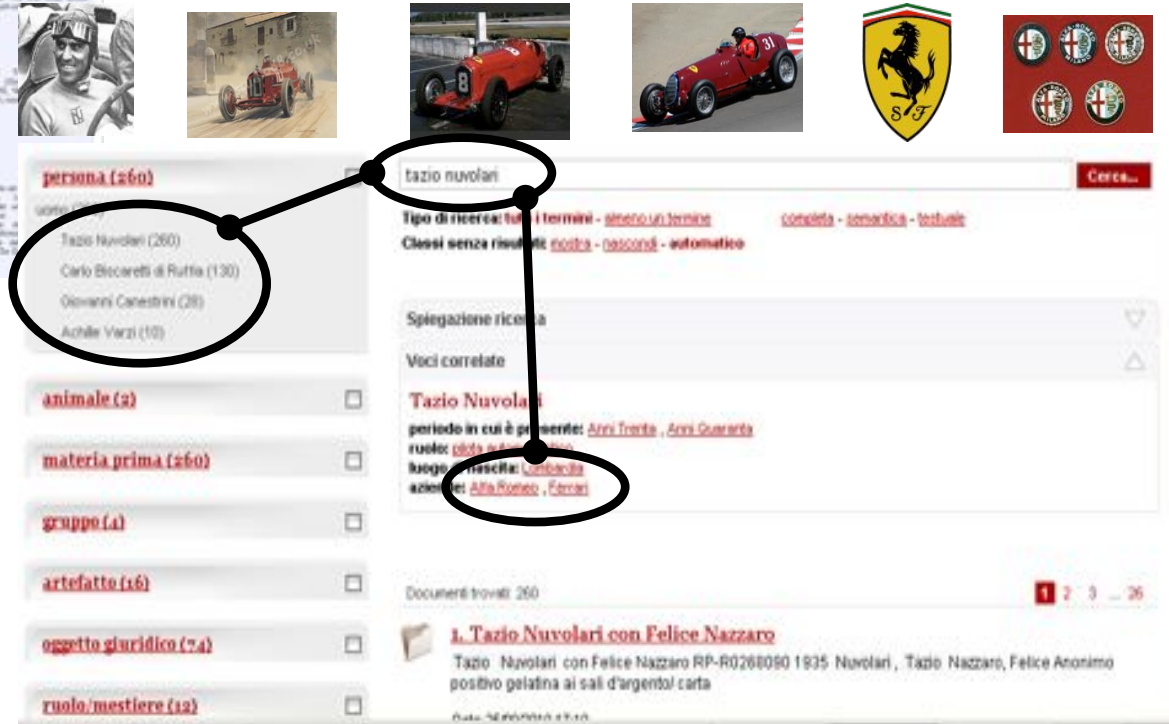


Faceted Semantic Search

Relations between Facets



Browse through
documents
and contents



LOD for CLIR

The screenshot shows the Organic.Edunet website interface. At the top, there is a logo for Organic.Edunet and a search bar containing the word 'картофель'. Below the search bar are links for Home, About, and Feedback. On the left side, there is a 'Apply Filters' section with various categories like Educational Context, Language, Technical Format, Collection, Educational Role, and Type. The main content area displays search results for 'картофель'. The first result is 'Potato', which includes a description, author information, typical age range, rating, keywords, and a language selector. The second result is 'Diseases of potatoes', which also includes a description, typical age range, rating, keywords, and a language selector. The interface is clean and organized, with a focus on providing educational resources.

THE AGROVOC
THESAURUS HAS
BEEN USED IN
THE
ORGANIC.LINGU
A PROJECT FOR
ONTOLOGY-
BASED CLIR

Sem-web techs for internal models

Information in the CRUNCHED BOOK is represented using combinations of RDF and GRAPH DBS

Places

Verona

- A tomb belonging to the Capulets
- Capulet's house
- A hall in Capulet's house
- A lane by the wall of capulet's orchard
- A room in Capulet's house
- Capulet's orchard
- Juliet's chamber
- Friar Laurence's cell

Romeo

Where he/she appears (Acts)

I II III V

Co-speakers

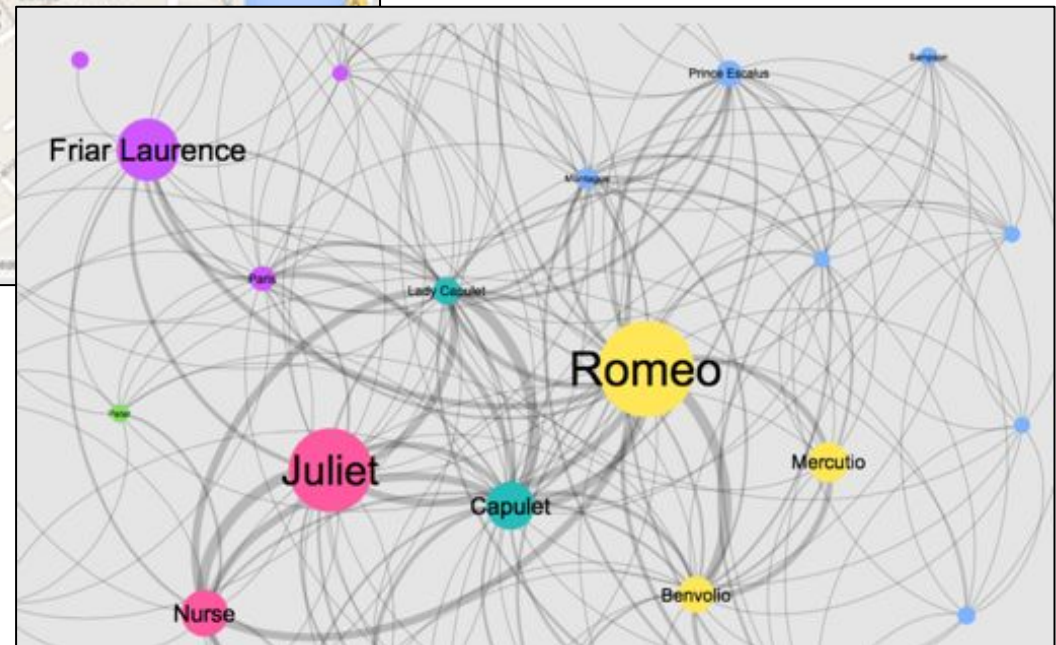
Abraham (1) Apothecary (1) Balthasar (2) Benvolio (7) Capulet (5) First citizen (2) First servant (1) First watchman (1) Friar Laurence (4) Gregory (1) Juliet (5) Lady Capulet (4) Lady Montague (1) Mercutio (4) Montague (3) Nurse (5) Page (1) Paris (2) Peter (1) Prince Escalus (3) Sampson (1) Second Capulet (1) Second servant (1) Second watchman (1) Servant (2) Third servant (1) Third watchman (1) Tybalt (3)

Contributions

Document [Romeo](#) Document [Romeo & Juliet by Zeffirelli trailer](#)
Image [Romeo & Juliet by Zeffirelli](#)
Image [Romeo+Juliet by Baz Luhrmann](#)

Places

A tomb belonging to the Capulets Friar Laurence's cell
A lane by the wall of capulet's orchard Capulet's orchard
A hall in Capulet's house



Public Sector: clear process ...



**Celi for the public sector
(CSI Piemonte):
the Homer project**



(Public sector contd.) ... but ...



**LACK
OF MONEY**

... hard
problems

USE OF “STANDARDS”



OPAQUE DATASETS

**POOR RDF/SPARQL
SUPPORT**



**LACK
OF WILLINGNESS**

Why companies' RDF is not published

- It reflects **customers' needs**



- It reflects **internal data models**

HENCE → OVERFITTING:

Provocation

It would not be **interesting** nor **usable**

WAY OUTS:

having more standard models for particular micro-domains could permit their direct (re)use by the private company (and hence the publication of enhanced versions)

Receipts

Public Sector: use “true” LOD technologies (RDF dumps and SPARQL endpoints)

Private companies: use standard data models, **internally** and for their artifacts

OpenData Community: please stress the *linked* in LOD!

The success of LOD is bound to the use of Linked Data
(as a technology)

The use of LD in the **Private Sector** will positively feedback on the diffusion of the necessary expertise and sensibility in the Public Sector too



Thank You!

Language and
Information Technology

Via San Quintino 31 - 10121 Torino
Tel. +39 011.562.71.15
Fax +39 011.506.40.86
info@celi.it - www.celi.it