

ETRACTIONS & DATA SCIENCE I: LINDA – A LINKED DATA PLATFORM FOR SMEs

Electronic Transactions

by Koussouris S., Lampathaki F., Askounis D.
Demo Presenter: Papaspyros D.



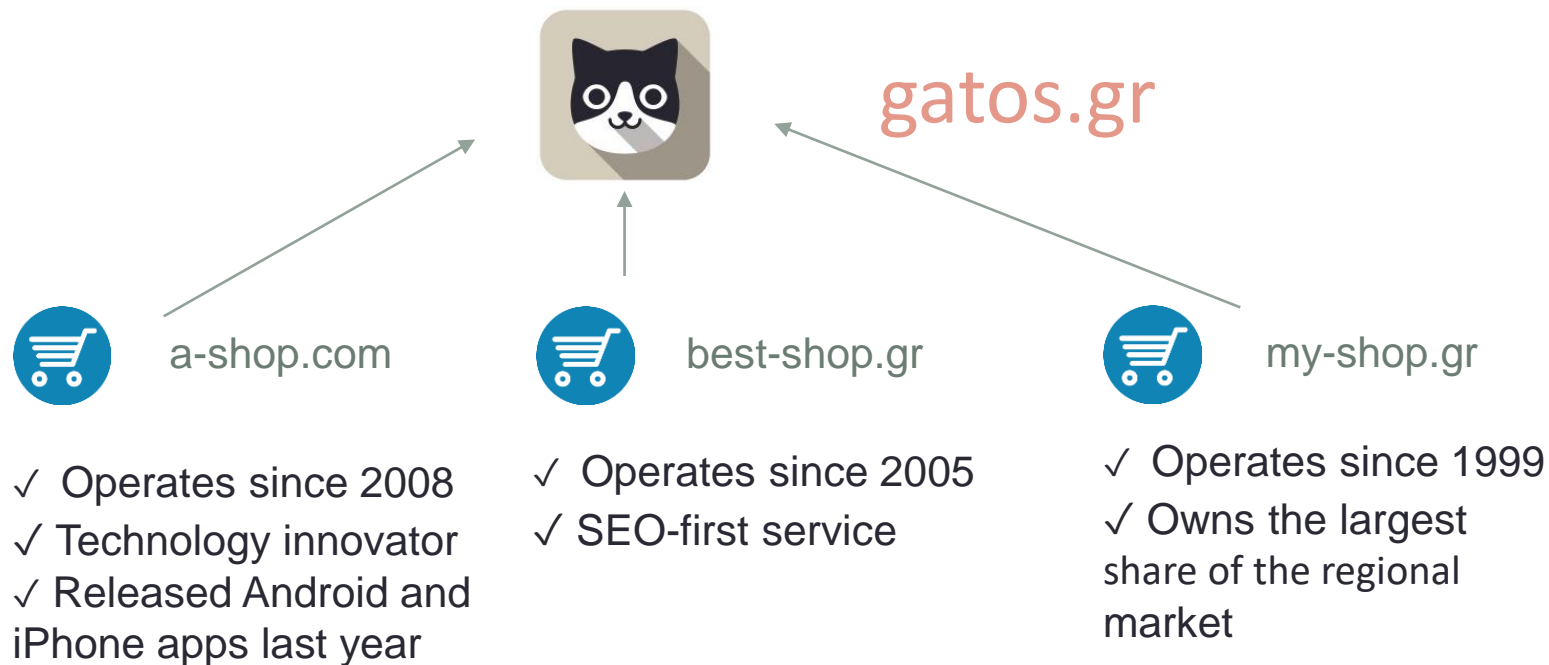
Introduction

- Problem statement
- Software interoperability and Data Semantics
- Linked Data
- LinDA Presentation

[Source: Companies and information: The leaky corporation | The Economist]

Problem Statement

- A new startup (gatos.gr) is developing a platform that will allow customers to compare offers from various other e-shops.
- The startup owners have managed to reach an agreement with three big players in the local market:



Integration efforts (1/3)

- Since a-shop.com developed their mobile applications last year, they have a state-of-the-art REST API service from which gatos.gr can fetch real-time product information.
- Although the format is fairly easy to integrate, the developers of our startup still have to create a mapping between the terms used in the API and their own terminology.

```
"product": {  
  "name": "SONY TV 4K 55\"",  
  "in_stock": 21,  
  "price": 549.99  
}
```

Integration efforts (2/3)

- Best-shop.gr have dedicated a lot of effort to improving their SEO ranking. By using appropriate HTML tags and descriptive class names, they make sure that Google crawlers will make the most out of their pages. This doesn't mean, however, that importing these products will be an easy task for our startup.

```
<article class="product-item">
  <h3 class="product-title">SONY TV </h3>
  <span class="resolution">4K</span>
  <span class="tv-size">55 " </span>
  <div class="price-info">
    <span class="price">205.00</div>
    <span class="tax">49.20</div>
    <span class="total-price">254.20</span>
  </div>
</article>
```

Semantic interoperability problems

- Even though some of the terms between the various business partners may be similar or even identical, this does not mean that they represent the same real-world concept.
- In our example, “price” for a-shop.com refers to total price, but “price” for best-shop.gr refers to product price before taxes.
- No automated system can distinguish between these two uses of the word “price” without any context information.

```
"product": {  
  "name": "SONY TV 4K 55\"",  
  "in_stock": 21,  
  "price": 549.99  
}
```

```
<article class="product-item">  
  <h3 class="product-title">SONY TV </h3>  
  <span class="resolution">4K</span>  
  <span class="tv-size">55" </span>  
  <div class="price-info">  
    <span class="price">205.00</div>  
    <span class="tax">49.20</div>  
    <span class="total-price">254.20</span>  
  </div>  
</article>
```

Ontologies

- In order to achieve true semantic interoperability between separate software systems, use of tools such as ontologies has to be introduced.
- An **Ontology** or **Vocabulary** is a formal definition of entities, their properties and their relationships within a specific domain.
- A widely adopted ontology, “*Friend of a Friend*”, is a great example of an ontology that defines:
 - Classes (e.g foaf:Person, foaf:Group) that define entities commonly used in social networks,
 - Datatype Properties (e.g foaf:name, foaf:homepage) that define attributes of these classes, and
 - Functional Properties (e.g foaf:member), which specify relationships between instances of the different entities.
- You can read more about Friend of a Friend and find examples of how the ontology could be used here:

<http://linda.epu.ntua.gr/vocabulary/2277/friend-of-a-friend/>

GoodRelations Ontology for Semantic Web-based E-Commerce

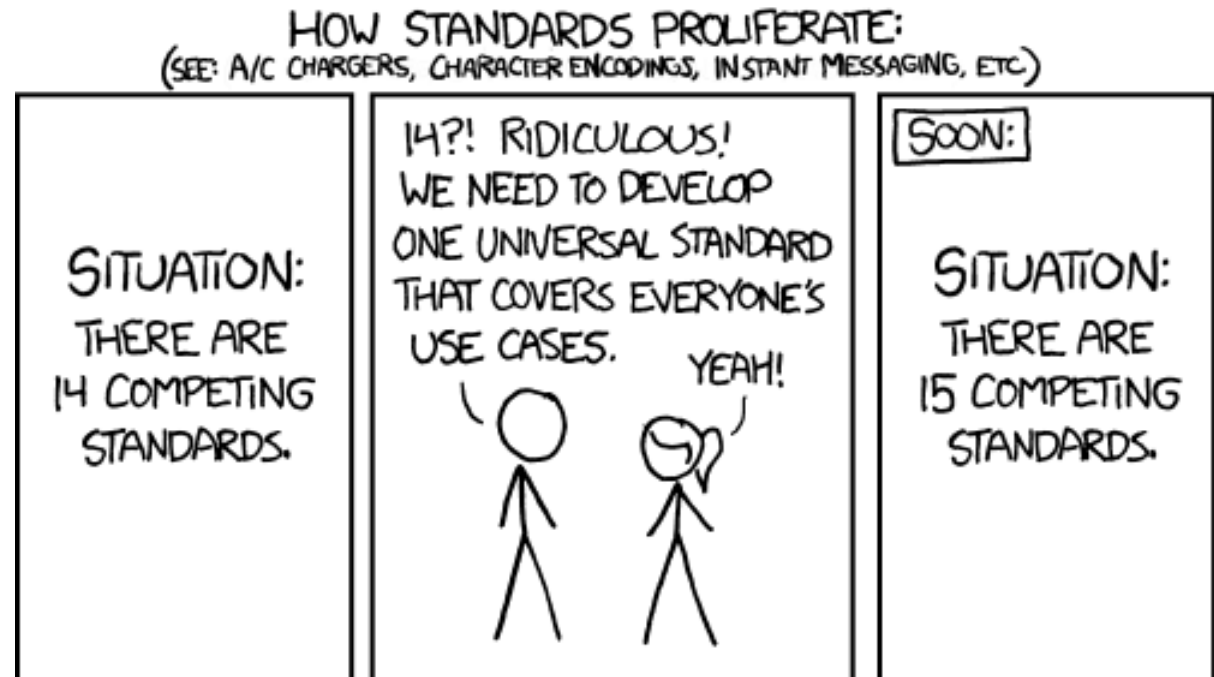
- GoodRelations Ontology for Semantic Web-based e-Commerce is an ontology that focuses into the e-commerce domain.
- Like most ontologies, GoodRelations reuses elements from other ontologies (such as Friend of a Friend), usually by extending existing classes (sub-classing) or defining new properties.
- Ontology outline:
 - Basic classes: Brand, Product, BusinessEntity, Offering, License etc.
 - Basic properties: BusinessEntity:legalName, Offering:eligibleCustomerTypes, Offering:hasPriceSpecification etc.
 - Find out more at <http://linda.epu.ntua.gr/vocabulary/1963/the-goodrelations-ontology-for-semantic-web-based-e-commerce/>

• Example

```
products:sonyTV
  a gr:ProductOrServiceModel;
  gr:name `Sony LED TV`;
  gr:hasBrand brands:sony ;
  seo:SupportedResolution `4K` .
```


Developing ontologies – the danger

- The main purpose of ontologies is to support software interoperability.
- This means that ontologies should only cover a clearly defined domain, and reuse concepts from other ontologies whenever possible.
- Otherwise...



Linked Data – the idea

- The main strength of the world wide web lies in the ability to **link** between different web pages.
 - This way, a webpage may provide its customer a link to another website in order to retrieve additional information about a topic.
- Could we apply the same principle on **data**?

Linked Data – Definition

- Linked Data is a data representation paradigm that is based on using Unique Resource Identifiers (URIs) to identify entities and the HTTP protocol to provide access to information about these entities.
- This practically means that linked data, much like websites, can live on different places, be maintained by different organizations, and still be used as a single system from the user's perspective.

Linked Data – our example



www.gatos.gr/latest-products

Product	Cost (€)	Availability	Characteristics
Sony LED TV 55"	549.99	18	www.sony.com/sony-led-tv-4k
Coffee maker Elekom EK-662	39.99	4	www.elecom.com/ek662



www.sony.com/sony-led-tv-4k

Frequency	Resolution	Processor	Weight (kg)
800 Hz	3840 x 2160	X-Reality PRO	37.60



www.elecom.com/ek662

Power (Watt)	Coffee type	Pressure (Bar)
800	Brewed	4.0

Linked Data – Characteristics

- Linked Data allow us to **easily reference** the same entity in different datasets.
- Using linked data we can refer to and extend data **external** to our organization.
- Linked data usage is ideal for **data exchange** between different systems, especially when each one of the system only maintains **part of** the overall **information** regarding each entity.
- Linked data usage **reduces** the cost of data exchange and maintenance, while **increasing** the cost of data generation and usage.
- Benefits of linked data greatly depend on **correct usage** of the paradigm and **well-designed datasets**.

interoperability

Small-and-medium
Enterprises

data maintenance

public open data



LinDA

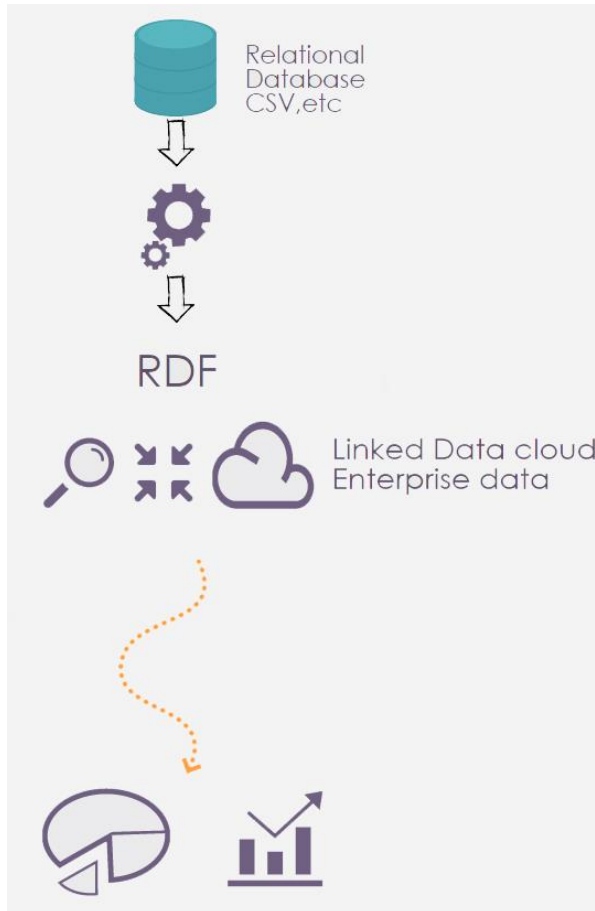
data curation

semantic enrichment

What is LinDA?

- LinDA is a toolbox for small and medium enterprises (SMEs) that facilitates the consumption, managements, reusability and analysis of corporate as well as public linked data.

Data lifecycle in LinDA



- Data from an SQL database or an Excel or CSV file are uploaded to LinDA.
- LinDA semi-automatically generates a transformation to enrich data with semantic information.
- Public open linked datasets are made accessible to the end user.
- User can use both public and private datasets to create queries, run analysis and generate charts.

LinDA Tools (1/2)

- **Vocabulary Repository**
A database of commonly used ontologies to facilitate semantic enrichment of datasets.
- **Transformation Engine**
Enables the transformation of existing relational datasets to linked data.
- **Query Designer**
Allowed the generation of SPARQL queries against one or more datasets through a GUI.

LinDA Tools (2/2)

- **Analytics Toolbox**

A toolbox of commonly used analytic algorithms applied on linked data.

- **Visualization Engine**

Enables users to generate linked data visualizations such as charts or maps.

You can find an open installation of LinDA at

<http://linda.epu.ntua.gr>

QUESTIONS

etransactions@epu.ntua.gr

Dr. Koussouris S. - skous@mail.ntua.gr

Dr. Lampathaki F. - flamp@epu.ntua.gr

Papaspyros D. - dpap@epu.ntua.gr