

# 9<sup>th</sup> International Conference on Information Society and Technology

# Linking Open Drug Data the Arabic dataset

Guma Lakshen: School of Electrical Engineering

Valentina Janev, Sanja Vraneš: Mihajlo Pupin Institute University of Belgrade



# Overview

## Linking Open Drug Data: the Arabic dataset

---

**Motivation:** Using the **Linked Data approach** in the pharmaceutical and drug industry in the Arabic region

**Methodology:** Design and implementation of **ALDDA**  
**(Arabic Linked Drug Data Application)**

**Results of Analysis:** SPARQL queries for **querying** Arabic data set linked with **DBpedia** and **Drugbank**

**Conclusions and Main Contributions**



# Motivation

## Use Case: Arabic Drugs Data sets

### The Arabic region.....

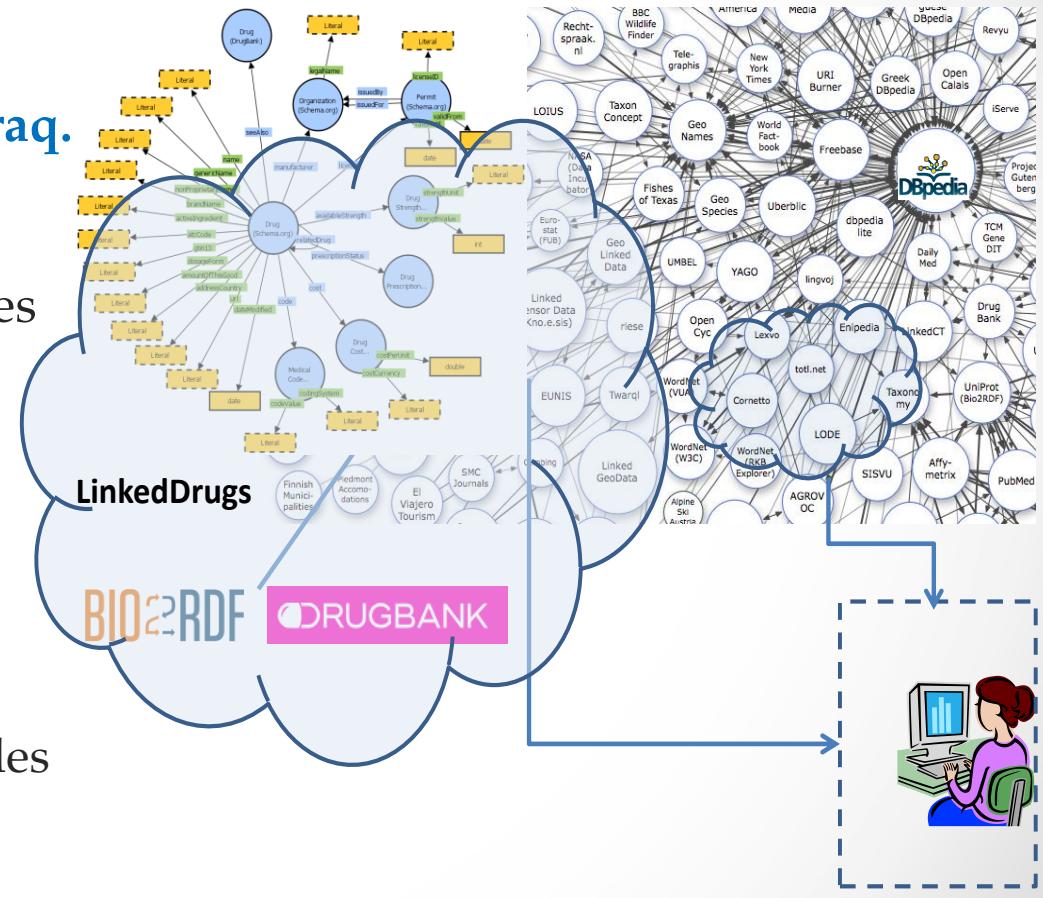
- 23 Countries.
- 422M Population, (2006).
- 13.2 KM<sup>2</sup>
- Located in North Africa and south west Asia.
- Arabic Language is one of 6 official languages in the UN.
- Partially read and understood by more than 1.8 billion Muslims in 56 countries worldwide.



# Motivation

## Use Case: Interlinking Arabic Drugs Data sets

- Sample Drug Datasets:  
Lebanon, Saudi Arabia, Egypt, Iraq
  - Datasets for interlinking:
  - DrugBank - 766,000 RDF triples  
for 5,818 drugs.
  - Dbpedia - 38.3 million things,  
23.8 million localized, 20 different  
Chapters.
  - LinkedDrugs - 248,000 drug  
products, over 99,000,000 RDF triples  
and over 278,000 links to generic  
drugs from the LOD Cloud



# Motivation

## Use Case: Linking Arabic Drugs Data sets

Answering user questions such as:

**Query1:** Retrieve relative information for a drug in Arabic language (if exists) from other identified datasets, such as DrugBank and DBpedia.

**Query2:** Retrieve equivalent drugs; and compare active ingredients, contradictions, and prices;;

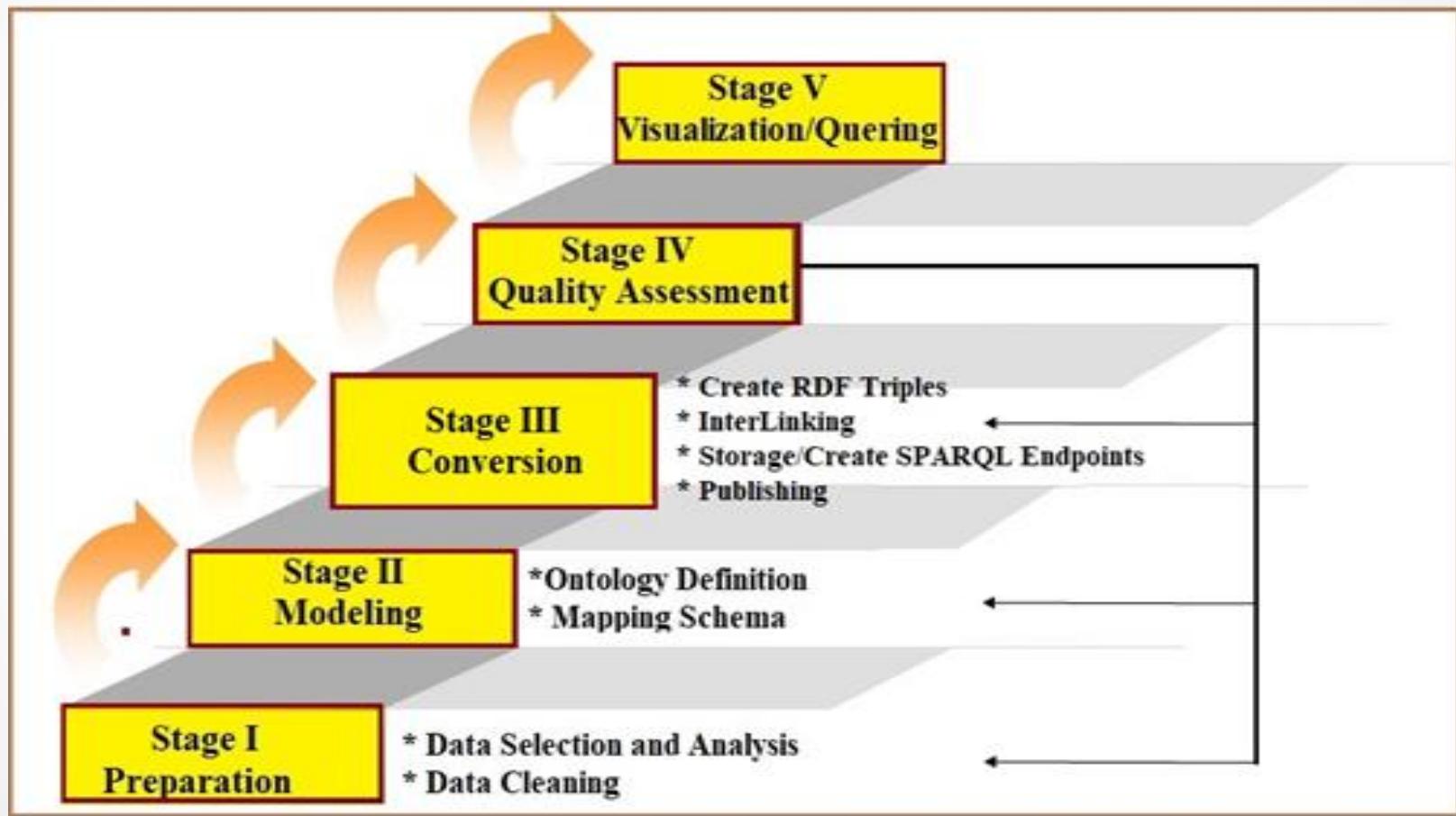
**Query3:** Retrieve valuable information about equivalent other drugs with different brand name, manufacturer, strength, form, price, etc.;

**Query4:** Retrieve drug reference information to highlight possible contradiction e.g. drug/drug, drug/allergy, drug/special cases (e.g. Pregnancy), etc.;

**Query5:** For an active ingredient retrieve advanced clinical information i.e. pharmacological action, pharmacokinetics, etc.;

**Query6:** Compare prices for a particular; drug, showing drug, cost, manufacturer, and country.

# Methodology



# Methodology – Step2: Data Mapping

## 1:Iraq (Excel Data file )

Original Attribute	Mapped Attribute
Scientific name	genericName
Trade name	brandName
Packaging&dosage form	dosageForm
Authorization holder (manufacturer)	manufacturer1
No. & date of registration	licenceValidFrom

## 2. Syria (Excel Data file)

Original Attribute	Mapped Attribute
Scientific name of the preparation	genericName
The commercial name of the product	brandName
Name	Manufacturer1
Caliber	Amount
Package	dosageForm
Price for the public	CostPerUnit

## 3. Saudi Arabia (Web database)

Original Attribute	Mapped Attribute
Generic Name	genericName
Trade Name	brandName
Strength Value	strengthValue1
DosageForm	dosageForm
Manufacturer Name	manufacturer1
Price	costPerUnit
Registration No	licenceValidFrom
Volume	Amount

## 4. Lebanon (Web database)

Original Attribute	Mapped Attribute
ATC	atcCode
Ingredients	activeSubstance1/ activeSubstance2/ activeSubstance3/ activeSubstance4/ activeSubstance5/ strengthValue1/ strengthValue2/ strengthUnit1/ strengthUnit2
Name	brandname
Dosage	dosageForm
Laboratory	manufacturer1
Price	costPerUnit
Registration No	licenceValidFrom
Exch_date	licenceValidUntil

## Methodology – Step 3: Data Interlinking

For Example: DBpedia Reconciliation service based on atcCode

```
PREFIX drugbank: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT *
WHERE
{
    ?s dbo:atcPrefix ?atcPrefix .
    OPTIONAL { ?s dbo:atcSuffix ?atcSuffix . }
    BIND (concat(?atcPrefix, ?atcSuffix) AS ?atcCode)
    FILTER regex(?atcCode, '<drugAtcCode>')
}
```

Similar procedure is done for brand Name, Chemical Substance, and generic Name in Drug synonyms.



# Results and findings:

- 31906 distinct drugs.
- 23971 interlinked drugs.
- >75% of the drugs are interlinked with Dbpedia in **order to enrich** the datasets with open data.
- For example running the SPARQL query:

```

prefix dbo:<http://dbpedia.org/ontology/>
prefix drugbank:<http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>
SELECT * WHERE {
?drug a <http://schema.org/Drug> .
?drug drugbank:genericName ?genericName .
?drug rdfs:seeAlso ?seeAlso .
{ SERVICE<http://dbpedia.org/sparql>
{ ?seeAlso dbo:abstract ?abstract } }
FILTER (?genericName = 'paclitaxel')
FILTER (langMatches(lang(?abstract), "ar")) }
```

Which extracts abstract info from Dbpedia in Arabic language for the 'taxol' which is an Organic composite similar to the 'paclitaxel' drug. Gives output.

"في 1988 توصل الباحثون في جامعة جونز هوبكنز إلى أن تاكسول taxol ، وهو مركب محضر من لحاء شجر الطقسوس بالمحيط الهادئ ، يمكن أن يفيد النساء المصابات بسرطان حاد في المبيض. كما اقترح الباحثون سنة 1991 في مركز أندرسون للسرطان في هيوستن أن مادة تاكسول يمكن أن تفيد السيدات المصابات بسرطان الثدي أيضاً. في دراسات تمت على 25 سيدة مصابة بسرطان متقدم في الثدي ولمتمكن من الاستجابة للعلاج الكيميائي ، شعر غالبية السيدات بانكماش الورم بعد تسع شهور من العلاج التجريبي "@ar".

# Results and findings: Find extra information

**Another Example:** To find extra information about **Fentanyl** drug from Dbpedia.

```

prefix dbo:<http://dbpedia.org/ontology/>
prefix drugbank:<http://www4.wiwiss.fu-
berlin.de/drugbank/resource/drugbank/>
prefix dbp:<http://dbpedia.org/ontology/>
SELECT * WHERE {
?drug a <http://schema.org/Drug> .
?drug drugbank:genericName ?genericName .
?drug rdfs:seeAlso ?seeAlso .
{ SERVICE <http://dbpedia.org/sparql>
{
?seeAlso dbo:abstract ?abstract .
?seeAlso dbo:wikiPageRevisionID ?wikiPageRevisionID .

OPTIONAL { ?seeAlso dbp:atcPrefix ?atcPrefix .}
OPTIONAL { ?seeAlso dbp:atcSuffix ?atcSuffix}
OPTIONAL { ?seeAlso owl:sameAs ?sameAs}
OPTIONAL { ?seeAlso dbp:synonyms ?synonyms}}
FILTER (?genericName = 'Fentanyl')
FILTER (langMatches(lang(?sameAs), "ar"))}
```

## Partial Result

"فينتаниل (بالإنجليزية) (Fentanyl): المعروف أيضا باسم fentanyl، Durogesic، Actiq، Sublimaze، Abstral، Instanyl، Onsolis، Fentora، Duragesic وغيرها) هو من مسكنات المخدرات الاصطناعية الفعالة مع بداية سريعة ومدة قصيرة من العمل. وهو ناهض قوي على مستقبلات - μ الأفيونية. وتاريخياً، قد تم استخدامه لعلاج الألم المزمن ويستخدم عادة في مرحلة ما قبل الإجراءات الجراحية بمثابة مسكن للألم وكمدر في توليفة مع البنزوديازيبين. يعتبر الفينتаниل أقوى بـ 80 إلى 100 مرة من المورفين وبشكل تقريبي هو أقوى بـ 40 إلى 50 مرة من الهيروين المستخدم بشكل طبي (النقي 100%) صنع فينتаниل أول مرة من قبل باول جانسين في عام 1960. بعد الاكتشاف الطبي للبيثيدين في السبعينيات، طورت جانسين الفينتаниل عن طريق معايرة نظائر للدواء بيثيدين ذي البنية الكيميائية القريبة للفينتانيل بحثاً عن الفاعلية الأفيونية. الاستخدام الواسع للفينتانيل أدى إلى إنتاج الفينتانيل سيترات (ملح يشكل دمج سيتريك أسيد مع الفينتانيل بنسبة 1:1)



# Results and findings: Find equivalent drugs

Drugs with different brand name comparison

	Drug1	Drug2
<b>BrandName</b>	EBETREXAT	METOJECT
<b>GenericName</b>	methotrexate	methotrexate
<b>ManufacturerLegalName</b>	Codiphia	Alfamed S.A.L.
<b>ActiveIngredient</b>	methotrexate	methotrexate
<b>DosageForm</b>	7.5mg/0.75ml	15mg/0.3ml
<b>CostFull</b>	32984.0 L.L	51182.0 L.L
<b>AddressCountry</b>	LB	LB

	Drug1	Drug2
<b>Drug Number</b>	aldda.b1.finki.ukim.mk/old/data/drugs#35704	aldda.b1.finki.ukim.mk/lod/data/drugs#36482
<b>GenericName</b>	glimepiride	metformin and sulfonamides
<b>ManufacturerLegalName</b>	Sadco	Benta Trading Co s.a.l.
<b>ActiveIngredient</b>	Glimepiride	Metformin HCl
<b>CostFull</b>	12415.0 L.L	28800.0 L.L
<b>AddressCountry</b>	LB	LB

## Conclusions

- There exist a few websites in the Arab region (in English with little information in Arabic) dealing with drugs such as WebTeb, altibbi, and dwaprice.
- Currently only few Arabic drug data exists and they are 2-star format i.e. Excel or PDF format.
- Only 4 countries started an initiatives in Linked data and semantic web: UAE, Egypt, SA, and Lebanon.
- Only a few studies exists in Arabic Language that emphasize on the importance of linked data issue.



## Main Contributions

- Analysis showed that existing Arabic drug data even in 2-star format has serious data quality problems.
- Our methodology can answer a verity of questions based upon user needs, and obtain information and comparisons from DBpedia and Drugbank.
- Due to different lingual background and knowledge of different parts of the Arab region, different lingual results can be obtained especially from DBpedia which enriches knowledge of different users.

Thank you for your attention

