

# Data Fusion using RDF

Mustafa Jarrar

Birzeit University



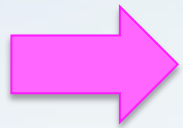
# Watch this lecture and download the slides



Online Courses : <http://www.jarrar.info/courses>

Thanks to Anton Deik for helping me preparing this lecture

# Data Fusion using RDF



**Part 1: Motivation Example (Government Domain)**

Part 2: Fusion of RDF Data

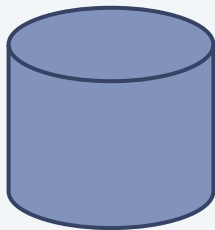
Part 3: Linking resources

Part 4: Practical Session

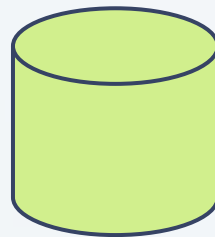
**Keywords:** Domain, databases, RDF, Data integration , linking, queries , dataset, entities , URI, Global Identifier, RDF graphs , Fusion, Practical Session ,relational data, model

# Example from the Government Domain

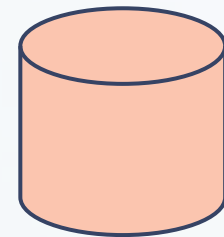
- Consider this simplified example from the Government domain. Consider three governmental agencies that record information about companies.
- In this example, we will integrate the three databases by transforming each one into RDF and then concatenating the resultant RDF tables into one table. After that, we investigate the concatenated data and link the different resources.
- Data integration is simply achieved through concatenation of RDF graphs and linking different resources. It is also achieved when building and executing the queries over the concatenated dataset.



**Companies DB in  
Ministry of  
Justice**



**Companies DB in  
Chamber of  
Commerce**



**Companies DB in  
Ministry of  
Economy**

# Ministry of Justice

Ministry of Justice records some information about companies in addition to the advocates that represent the companies.

## Company

<u>CompanyCode</u>	Name	Advocate_id	Country	City	Website
JK452	Future Tech	Y5263U	Palestine	Bethlehem	www.fttech.ps
YH852	Palestine Antiques	H782YU	Palestine	Jerusalem	www.antiques.ps

## Advocate

<u>Advocate_id</u>	Advocate_name	Country
H782YU	Tony Deik	Palestine
Y5263U	Ismail Mousa	Jordan

# Ministry of Justice: To RDF

## Company

CompanyCode	Name	Advocate_id	Country	City	Website
JK452	Future Tech	Y5263U	Palestine	Bethlehem	www.ftech.ps
YH852	Palestine Antiques	H782YU	Palestine	Jerusalem	www.antiques.ps

## Advocate

Advocate_id	Advocate_name	Country
H782YU	Tony Deik	Palestine
Y5263U	Ismail Mousa	Jordan



Subject	Predicate	Object
:JK452	:Type	:Company
:JK452	:Name	"Future Tech"
:JK452	:Advocate	:Y5263U
:JK452	:Country	"Palestine"
:JK452	:City	"Bethlehem"
:JK452	:Website	"www.ftech.ps"
:YH852	:Type	:Company
:YH852	:Name	"Palestine Antiques"
:YH852	:Advocate	:H782YU
:YH852	:Country	"Palestine"
:YH852	:City	"Jerusalem"
:YH852	:Website	"www.antiques.ps"
:H782YU	:Type	:Advocate
:H782YU	:Name	"Tony Deik"
:H782YU	:Country	"Palestine"
:Y5263U	:Type	:Advocate
:Y5263U	:Name	"Ismail Mousa"
:Y5263U	:Country	"Jordan"

# Chamber of Commerce

Chamber of Commerce records information about companies in addition to information about companies' owners.

## Company

CompanyRegistrationNo	CompanyName	CompanyAbbreviation	Address
8327848	Palestine Antiques	PalAnt	P.O.Box 43792, Jerusalem
5236964	Majdal Solutions	MajSol	Al-Manara, Ramallah

## Owner

OwnerNo	IDNumber	PassportNo	FirstName	LastName	DoB	Province	City
56566u	998568213	236555	Laith	Abdo	15-5-77	Jerusalem	Al-Ram
87858o	986532145	356245	Majd	Maher	17-3-85	Ramallah	Beitunia

## Company\_Owner

CompanyRegistrationNo	OwnerNo	Role	FromDate
8327848	56566u	CEO	20-07-2008
5236964	87858o	Technical Manager	17-06-2009

# Chamber of Commerce: To RDF



CompanyRegistrationNo	CompanyName	CompanyAbbreviation	Address
8327848	Palestine Antiques	PalAnt	P.O.Box 43792, Jerusalem
5236964	Majdal Solutions	MajSol	Al-Manara, Ramallah

OwnerNo	IDNumber	PassportNo	FirstName	LastName	DoB	Province	City
56566u	998568213	236555	Laith	Abdo	15-5-77	Jerusalem	Al-Ram
87858o	986532145	356245	Majd	Maher	17-3-85	Ramallah	Beitunia

CompanyRegistrationNo	OwnerNo	Role	FromDate
8327848	56566u	CEO	20-07-2008
5236964	87858o	Technical Manager	17-06-2009

Subject	Predicate	Object
:8327848	:Type	:Company
:8327848	:Name	"Palestine Antiques"
:8327848	:Abbreviation	"PalAnt"
:8327848	:Address	"P.O.Box 43792, Jerusalem"
:5236964	:Type	:Company
:5236964	:Name	"Majdal Solutions"
:5236964	:Abbreviation	"MajSol"
:5236964	:Address	"Al-Manara, Ramallah"
:56566u	:Type	:Owner
:56566u	:IDNumber	"998568213"
:56566u	:PassportNo	"236555"
:56566u	:FirstName	"Laith"
:56566u	:LastName	"Abdo"
:56566u	:DoB	"15-5-77"
:56566u	:Province	"Jerusalem"
:56566u	:City	"Al-Ram"
:87858o	:Type	:Owner
:87858o	:IDNumber	"986532145"
:87858o	:PassportNo	"356245"
:87858o	:FirstName	"Majd"
:87858o	:LastName	"Maher"
:87858o	:DoB	"17-3-85"
:87858o	:Province	"Ramallah"
:87858o	:City	"Beitunia"

Subject	Predicate	Object
:8327848	:Owner	:56566u
:8327848 Owner_56566u	rdf:type	rdf:Statement
:8327848 Owner_56566u	rdf:subject	:8327848
:8327848 Owner_56566u	rdf:predicate	:Owner
:8327848 Owner_56566u	rdf:object	:56566u
:8327848 Owner_56566u	:role	"CEO"
:8327848 Owner_56566u	:FromDate	"20-07-2008"
:5236964	:Owner	:87858o
:5236964 Owner_87858o	rdf:type	rdf:Statement
:5236964 Owner_87858o	rdf:subject	:5236964
:5236964 Owner_87858o	rdf:predicate	:Owner
:5236964 Owner_87858o	rdf:object	:87858o
:5236964 Owner_87858o	:role	"Technical Manager"
:5236964 Owner_87858o	:FromDate	"17-06-2009"



# Ministry of Economy

Ministry of Economy records information about companies, their owners, and their advocates.

## Company

CompanyNumber	CompanyFullName	Capital	CapitalCurrency	Location	LawyerNo
4354IU	Palestine Antiques	54,000	USD	Jerusalem	185652r
8755UB	Samson Trading	500,000	USD	Bethlehem	199632m

## Owner

OwnerNo	OwnedCompany	OwnerIDNumber	OwnerName	OwnerCity
HGR56u	4354IU	998568213	Laith Abdo	Jerusalem
BA856i	8755UB	858965321	George Musa	Nablus

## Lawyer

LawyerNo	LawyerName	LawyerAffiliation
185652r	Tony Deik	Al-Haq, Palestine
199632m	Sana Ibrahim	Al-Mizan, Palestine

# Ministry of Economy: To RDF

CompanyNumber	CompanyFullName	Capital	CapitalCurrency	Location	LawyerNo
4354JU	Palestine Antiques	54,000	USD	Jerusalem	L85652r
8755UB	Samson Trading	500,000	USD	Bethlehem	L99632m

OwnerNo	OwnedCompany	OwnerIDNumber	OwnerName	OwnerCity
HG856u	4354JU	998568213	Laith Abdo	Jerusalem
BA856i	8755UB	858965321	George Musa	Nablus

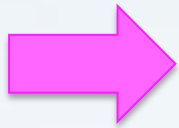
LawyerNo	LawyerName	LawyerAffiliation
L85652r	Tony Deik	Al-Haq, Palestine
L99632m	Sana Ibrahim	Al-Mizan, Palestine



Subject	Predicate	Object
4354JU	Type	"Company"
4354JU	Name	"Palestine Antiques"
4354JU	Capital	"54,000"
4354JU	CapitalCurrency	"USD"
4354JU	Location	"Jerusalem"
4354JU	Lawyer	L85652r
8755UB	Type	"Company"
8755UB	Name	"Samson Trading"
8755UB	Capital	"500,000"
8755UB	CapitalCurrency	"USD"
8755UB	Location	"Bethlehem"
8755UB	Lawyer	L99632m
HG856u	Type	"Owner"
HG856u	Owns	4354JU
HG856u	IDNumber	"998568213"
HG856u	Name	"Laith Abdo"
HG856u	City	"Jerusalem"
BA856i	Type	"Owner"
BA856i	Owns	"8755UB"
BA856i	IDNumber	"858965321"
BA856i	Name	"George Musa"
BA856i	City	"Nablus"
L85652r	Type	"Lawyer"
L85652r	Name	"Tony Deik"
L85652r	Affiliation	"Al-Haq, Palestine"
L99632m	Type	"Lawyer"
L99632m	Name	"Sana Ibrahim"
L99632m	Affiliation	"Al-Mizan, Palestine"

# Data Fusion using RDF

Part 1: Motivation Example (Government Domain)



**Part 2: Fusion of RDF Data**

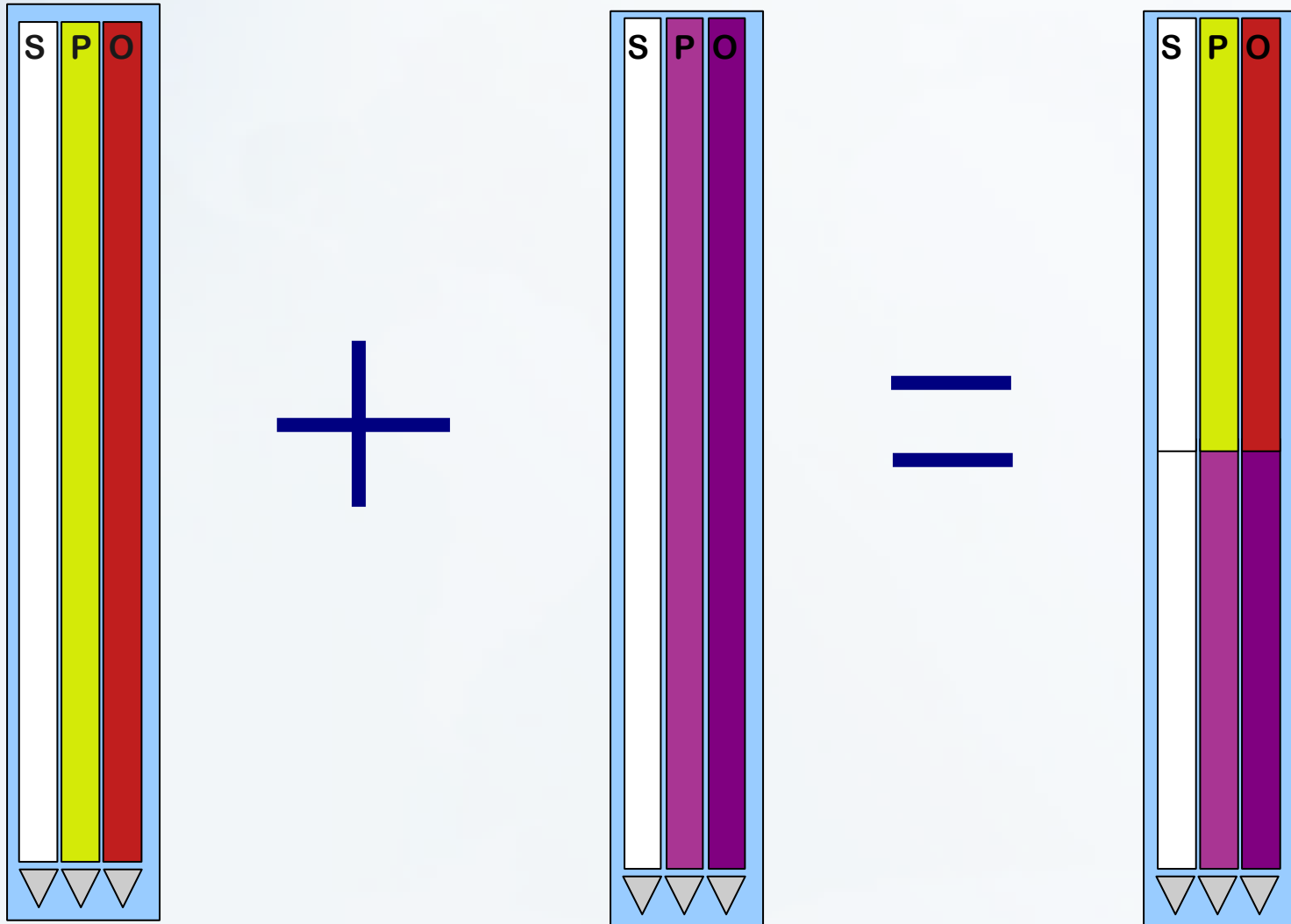
Part 3: Linking resources

Part 4: Practical Session

**Keywords:** Domain, databases, RDF, Data integration , linking, queries , dataset, entities , URI, Global Identifier, RDF graphs , Fusion, Practical Session ,relational data, model

# Fusion of RDF Data

As simple as ...



# In our example

- Appending the three graphs into one RDF.
- Query the RDF as one graph, though not necessarily connected.

Subject	Predicate	Object
:JK452	:Type	:Company
:JK452	:Name	"Future Tech"
:JK452	:Advocate	:Y5263U
:JK452	:Country	"Palestine"
:JK452	:City	"Bethlehem"
:JK452	:Website	"www.tech.ps"
:YH852	:Type	:Company
:YH852	:Name	"Palestine Antiques"
:YH852	:Advocate	:H782YU
:YH852	:Country	"Palestine"
:YH852	:City	"Jerusalem"
:YH852	:Website	"www.antiques.ps"
:H782YU	:Type	:Advocate
:H782YU	:Name	"Tony Deik"
:H782YU	:Country	"Palestine"
:Y5263U	:Type	:Advocate
:Y5263U	:Name	"Ismail Mousa"
:Y5263U	:Country	"Jordan"

Subject	Predicate	Object
:A154U	:Type	"Company"
:A154U	:Name	"Palestine Antiques"
:A154U	:Capital	"\$4,000"
:A154U	:CapitalCurrency	"USD"
:A154U	:Location	"Jerusalem"
:A154U	:Lawyer	:B755U8
:B755U8	:Type	"Company"
:B755U8	:Name	"Jamoon Trading"
:B755U8	:Capital	"\$00,000"
:B755U8	:CapitalCurrency	"USD"
:B755U8	:Location	"Bethlehem"
:B755U8	:Lawyer	:B9612m
:H6856u	:Type	:Owner
:H6856u	:Owner	:A354U
:H6856u	:IDNumber	998568213
:H6856u	:Name	"Lath Abdo"
:H6856u	:City	"Jerusalem"
:BA856i	:Type	:Owner
:BA856i	:Owner	"B751U8"

=

Subject	Predicate	Object
:B127848	:Type	:Company
:B127848	:Name	"Palestine Antiques"
:B127848	:Abbreviation	"PalAnt"
:B127848	:Address	"P.O. Box 43792, Jerusalem"
:S136964	:Type	:Company
:S136964	:Name	"Majdal Solutions"
:S136964	:Abbreviation	"MajSol"
:S136964	:Address	"Al-Manara, Ramallah"
:S6566u	:Type	:Owner
:S6566u	:IDNumber	"998568213"
:S6566u	:PassportNo	"216553"
:S6566u	:FirstName	"Lath"
:S6566u	:LastName	"Abdo"
:S6566u	:DOB	"15-5-77"
:S6566u	:Province	"Jerusalem"
:S6566u	:City	"Al-Ram"
:B7858o	:Type	:Owner
:B7858o	:IDNumber	"986531243"
:B7858o	:PassportNo	"356245"

Subject	Predicate	Object
:JK452	:Type	:Company
:JK452	:Name	"Future Tech"
:JK452	:Advocate	:Y5263U
:JK452	:Country	"Palestine"
:JK452	:City	"Bethlehem"
:JK452	:Website	"www.tech.ps"
:YH852	:Type	:Company
:YH852	:Name	"Palestine Antiques"
:YH852	:Advocate	:H782YU
:YH852	:Country	"Palestine"
:YH852	:City	"Jerusalem"
:YH852	:Website	"www.antiques.ps"
:H782YU	:Type	:Advocate
:H782YU	:Name	"Tony Deik"
:H782YU	:Country	"Palestine"
:Y5263U	:Type	:Advocate
:Y5263U	:Name	"Ismail Mousa"
:Y5263U	:Country	"Jordan"
:B127848	:Type	:Company
:B127848	:Name	"Palestine Antiques"
:B127848	:Abbreviation	"PalAnt"
:B127848	:Address	"P.O. Box 43792, Jerusalem"
:S136964	:Type	:Company
:S136964	:Name	"Majdal Solutions"
:S136964	:Abbreviation	"MajSol"
:S136964	:Address	"Al-Manara, Ramallah"
:S6566u	:Type	:Owner
:S6566u	:IDNumber	"998568213"
:S6566u	:PassportNo	"216553"
:S6566u	:FirstName	"Lath"
:S6566u	:LastName	"Abdo"
:S6566u	:DOB	"15-5-77"
:S6566u	:Province	"Jerusalem"
:S6566u	:City	"Al-Ram"
:B7858o	:Type	:Owner
:B7858o	:IDNumber	"986531243"
:B7858o	:PassportNo	"356245"
:B7858o	:FirstName	"Lath"
:B7858o	:LastName	"Abdo"
:B7858o	:DOB	"15-5-77"
:B7858o	:Province	"Ramallah"
:B7858o	:City	"Bethlehem"

# Data Fusion using RDF

Part 1: Motivation Example (Government Domain)

Part 2: Fusion of RDF Data

 **Part 3: Linking resources**

Part 4: Practical Session

**Keywords:** Domain, databases, RDF, Data integration , linking, queries , dataset, entities , URI, Global Identifier, RDF graphs , Fusion, Practical Session ,relational data, model

# Linking Resources

How are same entities described in different datasets linked?

By linking the Global Identifier, that is, the URI\*\*!

Let's have a look:

```
:YH852 owl:sameAs :8327848
```

```
:YH852 owl:sameAs :4354JU
```

- Links the company called “Palestine Antiques” in the three databases.
- This is called **entity resolution**/disambiguation.

```
:H782YU owl:sameAs :L85652r
```

- Links the lawyer called “Tony Deik” recorded in the ministry of Justice and the ministry of national economy.
- This is called entity resolution/disambiguation.

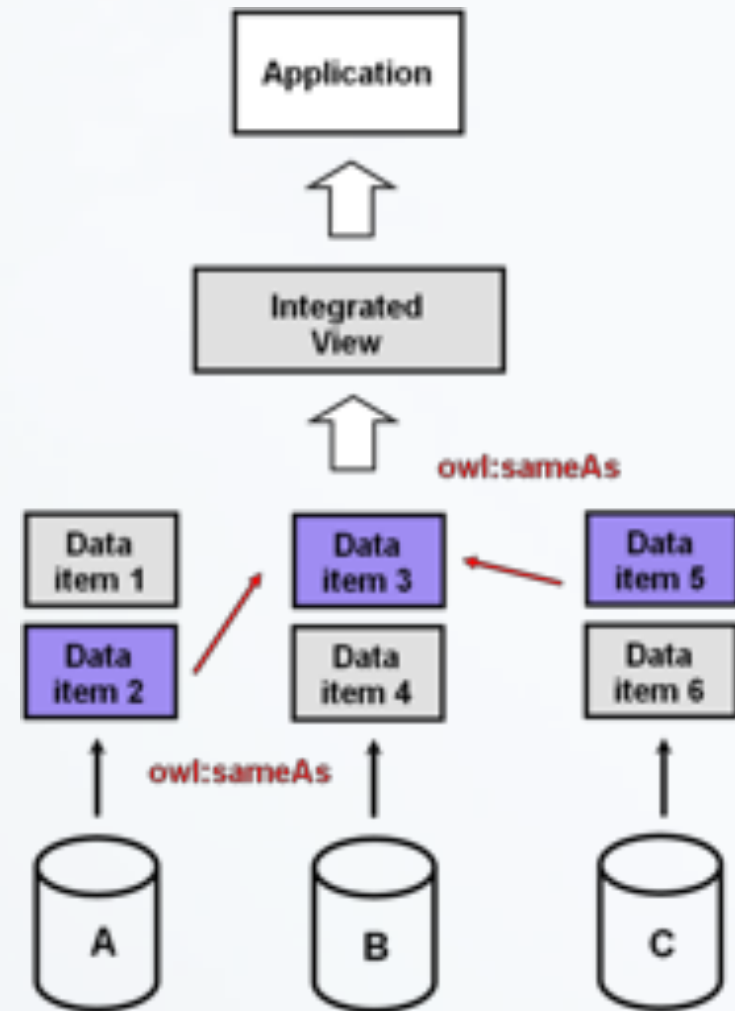
\*\*

Note that in our example we used colons to distinguish URIs. For example :JK452, :H782YU, :Country, and :Name are all URIs.

For example: “:H782YU” might actually be something like: <http://www.palgov.ps//H782YU>

# Data Integration and Fusion

Concatenating RDF graphs and linking entities in different datasets forms an integrated view where applications see all datasets as one integrated database.



Source: Christian Bizer



# Data Fusion using RDF

Part 1: Motivation Example (Government Domain)

Part 2: Fusion of RDF Data

Part 3: Linking resources



**Part 4: Practical Session**

**Keywords:** Domain, databases, RDF, Data integration , linking, queries , dataset, entities , URI, Global Identifier, RDF graphs , Fusion, Practical Session ,relational data, model

# Practical Session

## Description:

From previous practical sessions: “The central management of students’ profiles by the ministry of education is becoming an urgent need in the last years. Many students in Palestine move from one university to another, and they need to transfer their academic records. Also, the ministry of higher education needs to certify the diplomas and mark sheets of students. Moreover, there is a need to centrally manage/monitor students financial aids. Therefore, the ministry of higher education has decided to build a national student registry, such that, each semester every university has to send the academic record of every student to the ministry of education. The ministry will then update and integrate the academic records according to the data combined from all universities into the national student registry.”

The ministry wants to use RDF to integrate this data. Thus, each university must map its relational data (or data in any other model) into RDF, and at the ministry this data is integrated and fused. Map the universities’ relational data into RDF and integrate and fuse it.

# Practical Session

- Each two students form a **group**. Each group must be composed of students from different universities (in their first level degrees).
- Students are expected to use three different mark sheets from different universities to construct 3 different hypothetical relational data schemes of students records.
- Students must populate the three databases (pertaining to the 3 different data schemes) with sample data.
- Students must integrate and fuse all data using RDF.
- Students are highly recommended to use the ontologies developed in previous practical sessions when mapping and integrating RDF data.
- Students must write at least three SPARQL queries on the integrated RDF data that involves data from all 3 sources
- Students must work this practical session using Oracle Semantic Technologies.
- After finalizing their work, each group will be asked to **present their work** to all students, so to **collect comments and feedback**.
- The final delivery include: **(i)** Snapshots of the three hypothetical databases and schemes taken from Oracle DB. **(ii)** The RDF mapping of each database (SPO tables). **(iii)** The integrated final RDF showing how entities were disambiguated. **(iv)** The executed SPARQL queries and their results. Note that this final delivery should have the form of a report where discussion of the various steps are expected to be clear.

# References

- [1] Mustafa Jarrar, Anton Deik: The Graph Signature: A Scalable Query Optimization Index for RDF Graph Databases Using Bisimulation and Trace Equivalence Summarization. International Journal on Semantic Web and Information Systems, 11(2), 36-65, DOI: 10.4018/IJSWIS.2015040102. April-June 2015
- [2] Mustafa Jarrar, Anton Deik, Bilal Faraj: Ontology-Based Data And Process Governance Framework -The Case Of E-Government Interoperability In Palestine . In pre-proceedings of the IFIP International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA'11). Pages(83-98). 2011.
- [3] Mustafa Jarrar and Marios D. Dikaiakos: A Query Formulation Language for the Data Web. The IEEE Transactions on Knowledge and Data Engineering. IEEE Computer Society. Pages(783-798). Volume 24, Number 4, April 2012
- [4] Paolo Ceravolo, Chengfei Liu, Mustafa Jarrar, Kai-Uwe Sattler: Special Issue on Querying the Data Web -Novel techniques for querying structured data on the web. The World Wide Web Journal. Volume(14), Issue (5-6). Springer. August 2011. ISSN:1573-1413.
- [5] Anton Deik, Bilal Faraj, Ala Hawash, Mustafa Jarrar: Towards Query Optimization for the Data Web - Two Disk-Based algorithms: Trace Equivalence and Bisimilarity. Proceedings of the 3rd Palestinian International Conference on Computer and Information Technology (PICCIT 2010). Hebron, Palestine. March 2010.
- [6] Mustafa Jarrar, Marios D. Dikaiakos: Querying the Data Web: the MashQL Approach. IEEE Internet Computing. Volume 14, No. 3. Pages (58-670). IEEE Computer Society, ISSN 1089-7801. May 2010.
- [7] Mustafa Jarrar, Marios D. Dikaiakos: Querying the Data Web: the MashQL Approach. IEEE Internet Computing. Volume 14, No. 3. Pages (58-670). IEEE Computer Society, ISSN 1089-7801. May 2010. Mustafa Jarrar and Marios D. Dikaiakos: A Data Mashup Language for the Data Web . Proceedings of LDOW, WWW'09. ACM. ISSN 1613-0073. (2009).
- [8] Mustafa Jarrar and Marios D. Dikaiakos: MashQL: a query-by-diagram topping SPARQL -Towards Semantic Data Mashups. Proceedings of ONISW'08, part of the ACM CiKM conference. ACM. pages (89-96) ISBN 9781605582559.(2008).
- [9] Mustafa Jarrar: Towards methodological principles for ontology engineering. PhD Thesis. Vrije Universiteit Brussel. (May 2005)
- [10] Mustafa Jarrar, Luk Vervenne, Diana Maynard: HR-Semantics Roadmap- The Semantic challenges and opportunities in the Human Resources domain . Technical Report. The Ontology Outreach Advisory, Belgium. (OOA-HR/2007-08-20/v025). August 2007
- [11] Lyndon Nixon, Malgorzata Mochol, Mustafa Jarrar, Stamatia Dasiopoulou, Vasileios Papastathis, and Yiannis Kompatsiaris: Prototypical business use cases. Deliverable D1.1.2 (WP1.1), The Knowledge Web Network of Excellence (NoE) IST-2004-507482, Luxemburg. January 2005.
- [12] Peter Spyns, Daniel Oberle, Raphael Volz, Jijuan Zheng, Mustafa Jarrar, York Sure, Rudi Studer, and Robert Meersman: OntoWeb- a Semantic Web Community Portal. Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management (PAKM 2002). Pages (189-200). LNCS 2569, Springer. ISBN: 3540003142. December 2002.