**Journal of Kerbala University , Vol. 14 No.1 Scientific . 2016**

# RFID and LOD for distributed data sources Information Integrity

# تحديد التردد الراديوي وتقنية ربط البيانات المفتوحة لتكاملية المعلومات لمصادر البيانات الموزعة

**Dr.Asaad Sabah Hadi**
**Babylon university, Information technology college, software department.**
asaadsabah@yahoo.com

## Abstract

Information integrity is very important factor for multiple information resources. The main problem for data storage is the accuracy and consistency of the data that are stored for a long time. The data integrity is the opposite of data corruption, where the aim of data integrity is to prevent any accidental changes of these data. The main idea about this paper is to design and implement a system for information integrity to the universal ID for Iraqi people. The ID contain many suggested items such that its information is check directly with the main website, that issued that information, therefore each item is checked for a specific period that entered by user in order to check its validity with the source . The main idea about LOD (Linked Opened Data) is publishing the structured data in RDF (Resource Description Framework) in a form of subject, predicate, and object, therefore all the website that contain the updated information for the ID must be connected as LOD by using a center that we named UDC(Universal Database Center). In order to make the ID more accessible we suggest to use RFID (Radio Frequency IDentification) where the information can be checked by multiple RFID readers. The most familiar hurdles with RFID is the security and privacy that we resolve it by using LOD.

Keyword: ***Information, Data, RDFID,LOD, RDF, Information integrity***.

**الخلاصة**

تعتبر تكاملية البيانات من المواضيع المهمة وخاصة بالنسبة للبيانات المسحوبة من اكثر من مصدر. تعتبر دقة و تماسك البيانات المخزونة لفترات طويلة من اهم المشاكل التي تواجه عملية خزن البيانات. ان تكاملية البيانات هي عكس فقدان البيانات حيث ان الهدف من تكاملية البيانات هو محاولة منع فقدان او تغيير البيانات لأي سبب. الهدف الرئيسي من هذا البحث هو تصميم وتطبيق نظام لغرض تكاملية البيانات الموجودة في الهوية الموحدة للمواطن العراقي والتي تتكون من مجموعة من الحقول المقترحة والتي يتم فحصها دوريا وحسب الوقت المقترح من قبل المستخدم وذلك لغرض التأكد من صحتها ومطابقتها مع المصدر. ان الفكرة الرئيسية من ربط البيانات المفتوحة (LOD) هي عملية خزن البيانات والمعلومات بصيغة (RDF) والتي تتكون من فاعل-علاقة-مفعول به وبذلك يجب تحويل جميع مواقع البيانات التي نحتاجها الى هذه الصيغة مما يسهل ربطها والاستفادة منها بواسطة مركز قواعد البيانات المركزي. لغرض تسهيل عملية الوصول الى معلومات الهوية المحدثة بكل سهولة تم اقراح اضافة (RDFID) والتي يتم قراءتها بواسطة قارئ خاص ممكن وضعه بأماكن مختلفة. ان من اهم المعوقات عند استخدام RFID هو السرية والخصوصية وتم حلها بواسطة استخدام LOD.

الكلمات المفتاحية : ***المعلومات، البيانات، تحديد الترددات الراديوية، ربط البيانات المفتوحة، اطار وصف الموارد، تكاملية المعلومات.***

## I. Introduction

Most companies migrating the critical application mission onto open standard RDBMS(Relational Database Management Systems) for many reasons like : Reducing skill base, lack of flexibility to meet business requirements, lack of independent software vendors, total cost of ownership[1]. The primary business challenge is business continuity in which the migration strategy ensure the continuity to a new application[1]. Whenever migration of a specific application the next challenge occur in which we need to decide which business processes will be migrated to the new application. Data is the foundation of the modernization process. If we want to meet the business

requirement then we need to reach clean migration which involve : data that is easy to maintain, data is organized in a usable format[2][1]. The Evolution of applications has led to the need for a method suitable for all uses.

## II. Information Integrity

Data integration is a very important challenge for multiple autonomous and heterogeneous data sources[3][2]. One of the major obstacle for setting up a data integration is writing the semantic mapping between the mediated schema and the source[2][4]. The context for the operation of the data integration system is dynamic where the optimizer has very limited information also there are two scenarios, the first is that the plan that look good in optimization time may behaves poor if the sources does not respond as expected, whereas the second show that the optimizer may not have enough information in order to decide on a good plan[2]. The XML (eXtensible Markup Language) offer a syntactic format for sharing data among data sources but it lack in addressing the semantic integration issues, because the source may continue to share XML files that are completely meaningless[2][4].

The Data migration is moving the data to the application that requested it, therefore there must be a cooperation between the application and the sources in order to take the correct data. In order to reach a clean data migration we must make many conditions, the first is organizing that data is a usable format, the second is facilitating the data maintenance[3][1].

The main challenge in any information integration is that the heterogeneous data format is non-relational hierarchical format. There are three integration strategies for data : Consolidated, Shared, Federated. The Consolidated data migration have only one database and there is a central location to manage it. The Shared data integration move the data and all events from multiple database sources to a queue or affiliated resources that are created to serve one or more applications. The Federated data integration have an individual data source which is self-maintaining and updating[1].

## III. Data Base[5]

The data is a raw facts whereas the information is the result of making some processing on the raw data. The data is the building block of the information. The metadata (data about data) describe the characteristics of data and a set of relationships between the data within the database. The DBMS(Data Base Management System) manage the database structure and direct access to the data that are stored in the database. The DBMS is the intermediary between the database and the user. The database structure can be stored as fixed files or a graph of separated representation that are collected according to some criteria[6]. The advantages of DBMS are :[5]

- Minimize the data inconsistency
- Data integration
- Improve the data security

Location can be used to classify the database into two main types : Centralized database and Distributed database. In centralized database, the database that support the data are located in a single site whereas in a distributed database, the database is distributed through multiple website[5]. Database can also be classified according to data representation into : Unstructured data, Structured data and Semi structured data. Semi-structured and Unstructured data storage are addressed into new database generation known as XML (Extensible Markup Language) database.

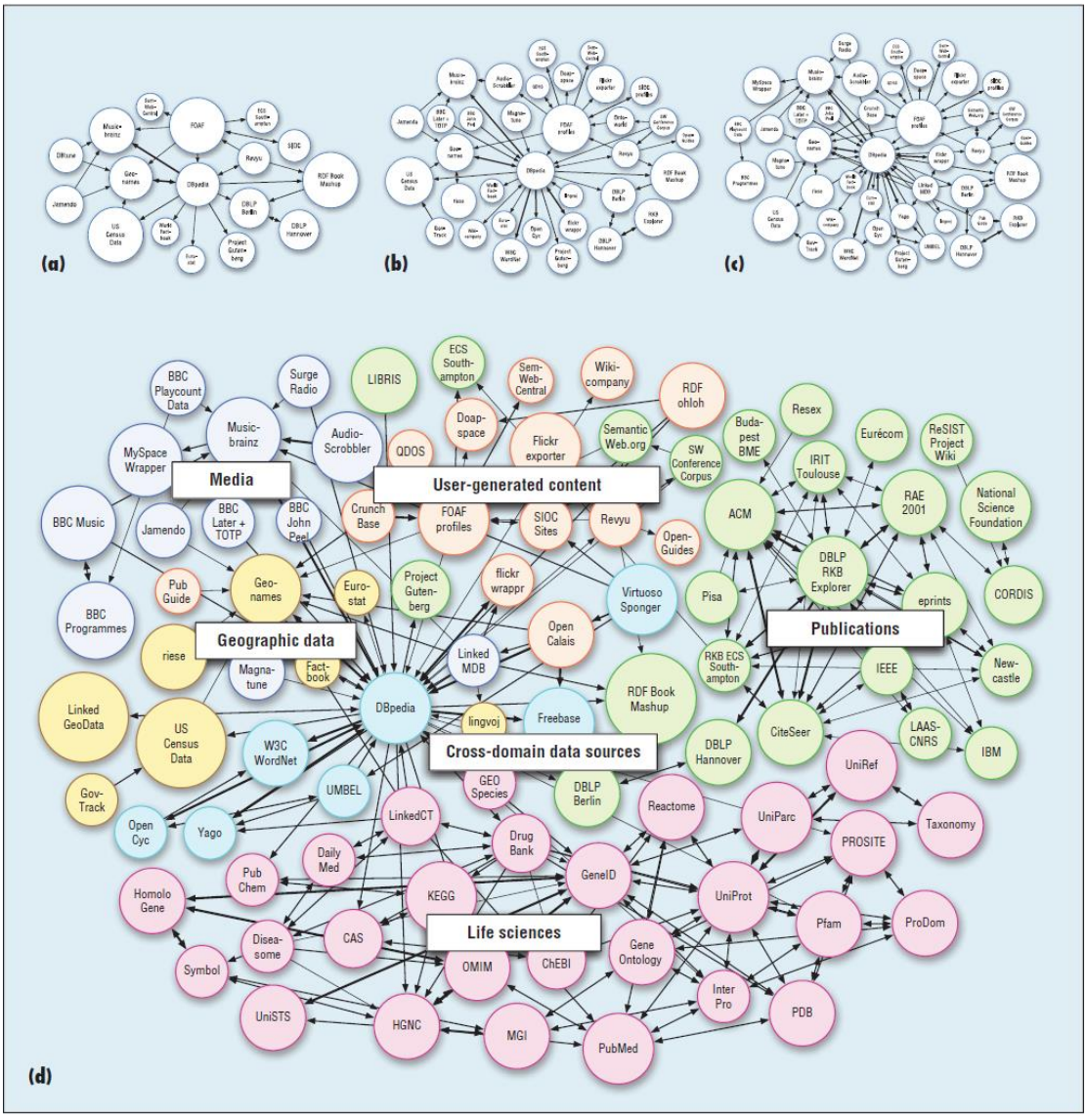
## IV. Linked Opened Data

The main idea about Linked Open Data is using HTTP URIs (Hypertext Transfer Protocol Uniform Resource Identifiers) to characterize arbitrary real-world entities and analyze Web Documents[7][8][9]. The entities data is represented using RDF (Resource Description Framework) and these description may contain some links to other entities which is described by other data resource. The form of the links is represented by Triples (Subject-predicate-object).

There are many properties for the Linked Data Web [6][7][8]:

- The Linked Data Web is opened in which the sources for the new data can discovered online by following the links.
- The application can resolve the URIs that represent some vocabulary if these vocabulary is unfamiliar.
- Anyone can broadcast his data to the Linked Data Web.
- The links connects many entities as a Graph and that enable to discover new data sources.

In the last years there are a fast growth of the Linked Opened Data . Figure (1) Shows the growth from 2007 to 2009 and now it became very huge links[8][9].

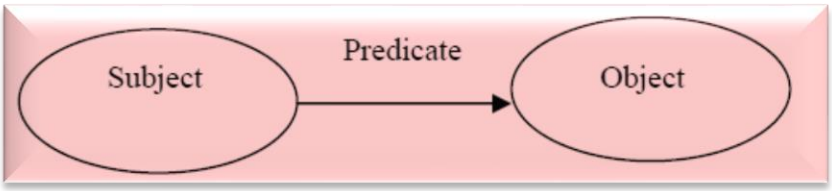



**(Figure-1) The Linking Open Data Cloud growth : (a) 2007 (b) April 2008 (c) Sept. 2008 (d) 2009**

representing the data in RDF which improve the machine to read the structure of the data, also solve the proprietary data format that decrease the software to reach many types of data where the LOD makes all the data available to all[6][10][11][8].

## V. RDF

The RDF is a general language that is used to represent Web metadata [12]. The main objective of RDF is representing the meaning of the web metadata. The meaning here is mean "Semantic meaning" , because the main problem of the web data is understanding its semantic meaning [6][10][8]. The RDF is a very simple data model that accumulate the semantic with better inference[11][13]. The model allow anyone to inquire about any resources. Anything that have a unique URI can be stored in RDF format that consist of a set of triples of the form (s, p ,o), where s is the subject and p is the predicate and o is the object[12]. Figure (2) show the RDF triple[13].




**(Figure-2) RDF Triple**

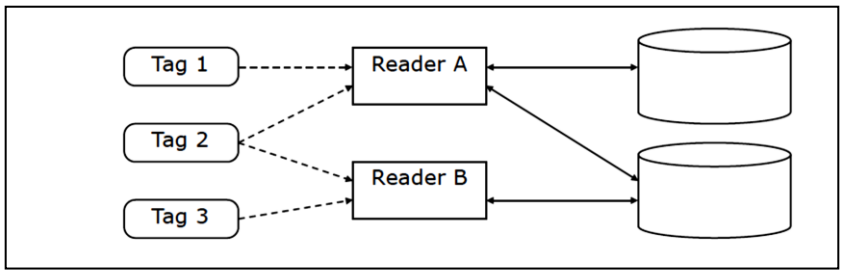
The SPARQL (Simple Protocol And RDF Query Language) is a language ,which recommended by W3c,that are used to fetch data from RDF files . It is similar to SQL constructs[7][12].

## VI. RFID [14]

The RFID (Radio Frequency Identification) is a well-known Automated Data Collection technology that uses the radio-frequency waves to transfer the data between a movable item and a reader for categorizing, identification, tracking purpose.

This technology refer to the system that merge all the electronic, digital, physical elements which enable the RFID Tags and readers for collecting, using, storing the required data. The elements include tags , servers, readers, and the RFID Software. The RFID tag is a small piece of material contain three components: small antenna, wireless transduces which link to microchip unit that contain memory storage, and an encapsulating material which can attached or inserted into new material.

There are many different kinds of RFID system which are used in many different applications. They have a different operating frequencies, power sources and functionalities. Figure (3) show the RFID System Interaction[15].




**(Figure-3) Illustration of RFID System Interaction**

The RDFID databases associate tag-identification data with arbitrary records, the records may contain product information, tracking logs, ..,etc. Independent databases can be built using unrelated users or may be integrated in a federated or centralized database system[16].

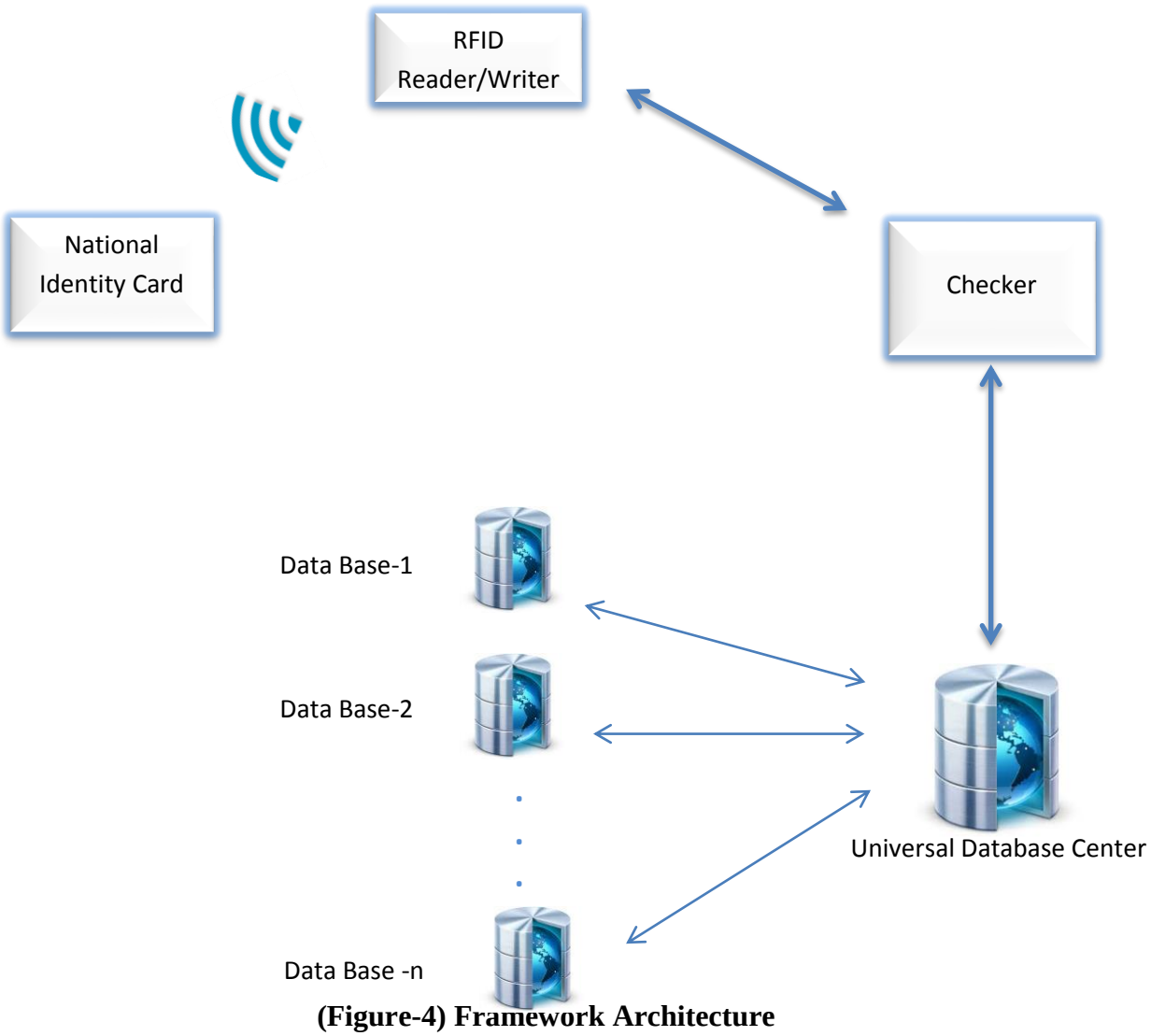
Different RFID systems operate at a variety of radio frequencies. Each range of frequencies offer its own power requirement, performance , and operating range[14].

The basic function of RFID is identification. The tags return identifiers that used to retrieve other data record. The underlying technologies and principles for various types of tags are related to RFID tags. The RFID-style tags can be splited into five broad classes shown in table (1)[17].

**(Table -1) Tag Functionality Classes**

| Class | Name | Memory | Power Source | Features |
|---|---|---|---|---|
| A | EAS | None | Passive | Article Surveillance |
| B | Read-only EPC | Read-Only | Passive | Identification Only |
| C | EPC | Read/Write | Passive | Data Logging |
| D | Sensor Tags | Read/Write | Semi-Passive | Environmental Sensors |
| E | Motes | Read/Write | Active | Ad Hoc Networking |

## VII. Framework Architecture




**(Figure-4) Framework Architecture**

Let us explain the main components of our proposed system in Figure(4):

- ***National Identity Card*** : It is RFID card that contain written information and saved it in its small memory.
- ***RFID Reader/Writer*** : It capture the RFID card and read the main information from it then contact with the checker to check its correctness and update all the information so that we reach the information integrity. There are many type of RFID reader/writer but we use (Scanfob N330B Bluetooth NFC MIFARE Reader/Writer) which allow the data transmission over Bluetooth or USB.
- ***Checker*** : It check the ID code in the RFID card and check the information integrity by conforming the current information with the information stored in the Universal Database Center. The checker convert the information in the RFID to RDF in order to make the compatibility with the Universal Database Center which is stored all the information in RDF format.
- ***Universal Database Center :*** It is a central database that store the last update of the whole information in the RFID by linking with all the LOD website which is designed to publish the information that are required in the RFID. The universal Database Center has the same features of DBpedia in the LOD[11].
- ***Database*** : The number of database may change according to the stored information in the RFID. For example, for the passport no. the Universal Database Center will link with the Passport General Department website for updating the passport no. whereas linking with the General Directorate of Nationality website for updating the Nationality ID information and so on. We must mentioned that the database type for the universal database center and all the databases must be put in RDF format in order to give the simplicity for gathering all the information semantically. Here we have two scenarios :

1. All databases will be built in RDF format :

In this scenario all databases is already built as we want then our system will be fetched the information directly to the universal database center.

2. All databases build in relational database :

In this scenario , the most common, all the databases information is put in tables therefore our system will convert these information into RDF format .

RFID Information

For every written information in the RFID we save a check bit that take two value :

✓ A bit (1) if the information is correct and checked with the original website.
✓ A bit (0) if the information is wrong or it is not checked.

All the checking bits must be "1" in a normal situation , because we can't use a card that have wrong information.

We proposed many information field like :

❖ ID No. : It save universal unique number.
❖ Name : It contain First, Middle and last Name.
❖ Position : It save the position of ID owner.
❖ Country : It save the country for the Id owner.
❖ DOB (Date OF Birth) : It contain the date of birth in day, month and year .
❖ Address : It save the address.
❖ Title : it save the title of the owner ( Mr., Mrs., …).
❖ Photo : it save the site that contain the owner photo.
❖ Email: it save the email.
❖ Pobox :it save the Pobox-number.
❖ Pcode :It save the postcode .
❖ Other : It save the Expiry date
❖ Checksum: In order to ensure the validity of these information, we make an AND gate for all bit and save the result in this field . The value of checksum is also stored in the header of the

owner picture , therefore in any checkpoint we can checked the validity of the RDFID information by comparing the values of the check bits with the values of the checksum. The main advantage of saving the value of the checksum in the header of the owner phone is give double security : ***The first*** is the security of the whole information, because the checker can ensure the integrity of these information by comparing this value with the checksum value. ***The Second***, is give more security to the owner photo, because the original photo must contain the correct checksum value.

## VIII. Implementation

We test our proposed system with multiple cases, but for simplicity and drawing purposes here we take a sample of Five RFID information as shown in table(2) below, these information are stored in RDF file in order to simplify the information acquisitions. Figure (5) show the RDF that represent these information whereas figure (6) show a portion of the graph for that RDF.

**(Table-2) RFID information**

| | Noor | Ali | Karrar | Hashim | Mohammed |
|---|---|---|---|---|---|
| Full Name | Noor Kareem Hadi | Ali Abbas Khadum | Karrar Jasim Mahdi | Hashim Hillal Maki | Mohammed Aziz Hatem |
| Key | 001 | 002 | 003 | 004 | 005 |
| Country | Iraq | Iraq | Iraq | Iraq | Iraq |
| Birth day | 1-1-1990 | 1-1-1995 | 1-1-1996 | 1-1-1975 | 1-1-1985 |
| Address | Iraq-Babylon | Iraq-Babylon | Iraq-Babylon | Iraq-Babylon | Iraq-Babylon |
| Title | Mrs. | Mr. | Mr. | Mr. | Mr. |
| Photo | Website-1 | Website-2 | Website-3 | Website-4 | Website-5 |
| Email | No@yahoo.com | al@yahoo.com | ka@yahoo.com | ha@yahoo.com | mo@yahoo.com |
| Tel. | 07801111111 | 07802222222 | 07803333333 | 07804444444 | 07805555555 |
| Postbox | 111 | 222 | 333 | 444 | 555 |
| Post Code | IR01 | IR02 | IR03 | IR04 | IR05 |
| Other(Expiry date) | 1-1-2016 | 1-1-2016 | 1-1-2016 | 1-1-2016 | 1-1-2016 |

For example, if "Karrar Jasim" change his address from "Iraq-Babylon" to "Iraq-Baghdad" then the check-bit value is replaced by "0". The value of the checksum is "0" as shown in Table(3). The Checker will update the wrong value with new one using "Universal Database Center", therefore the check-bit will be updated to "1" to ensure the correctness of the information.

The main idea about building the information in RDF format is to link the symmetric information, for example if many persons have the same value like : the country is Iraq, then we can cluster them later according to some features, and so on. We can find a new information according to the existing one like finding the friend of someone by knowing some identical information between them or by following the links between them.

**(Table-3) Checksum value for RFID information**

| *ID No* | *Name* | *Position* | *Country* | *DoB* | *Addr.* | *Title* | *Photo* | *Email* | *PoBox* | *poBox* | *Pcode* | *other* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Checksum value = 0**

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0# ">
  <rdf:Description rdf:about=http://somewhere/Mohammed>
   <vcard:Other>1-1-2016</vcard:Other>
<vcard:PHOTO>website-5</vcard:PHOTO>
   <vcard:BDAY>1-1-1985</vcard:BDAY>
   <vcard:Extadd>Babylon-Hilla</vcard:Extadd>
   <vcard:FN>Mohammed Aziz Hatam</vcard:FN>
   < vcard:Country>Iraq</vcard:Country>
   <vcard:Pobox>555</vcard:Pobox>
   <vcard:KEY>0005</vcard:KEY>
   <vcard:EMAIL>mo@yahoo.com</vcard:EMAIL>
   <vcard:TEL>07805555555</vcard:TEL>
   <vcard:Pcode>IR05</vcard:Pcode>
<vcard:TITLE>Mr</vcard:TITLE>
  </rdf:Description>
```

**(Figure-5) portion for the RDF/XML for the information in table(2)**


**(Figure-6) sub-graph for RDF in Figure(5)**

## IX. Conclusion

The proposed system gave a high information integrity, for our samples, by using Linked Opened Data Principles which distribute the information in many links for making the information retrieval very quickly and accurately. Using RFID card in our system make the checking procedures very quickly and can serve many cards simultaneously because all the updated information are stored in a universal database center.

**References**

[1] Tom Laszewski, et al , " Oracle Information Integration, Migration, and Consolidation ", book, 2011.
[2] Alon Halevy,et. all, " Data Integration : The teenage Years", Proceedings of the 32Nd International Conference on Very Large Data Bases, pp. 9-16, ACM 1-59593-385-9/ 06/09 , 2006.
[3] Richard Sheffield, "The Web Content Strategist's Bible ", book, CLUEfox Publishing,2009.
[4] Patricia Rodríguez-Gianolli and John Mylopoulos, " A Semantic Approach to XML-based Data Integration", Springer –Verlag , LNCS 2224, pp. 117-132, 2001.
[5] Carlos Coronel , Steven Morris , Peter Rob, " Database System Design, Implementation, and management", Book, 10th edition, Course Technology,2012.
[6] Michael Hausenblas, Marcel Karnstedt, " Understanding Linked Open Data as a Web-Scale Database", IEEE Computer society, 2nd international conference , pp. 56-61,2010.
[7]André Freitas, Edward Curry, João Gabriel Oliveira and Seán O'Riain, " Querying Heterogeneous Datasets on the Linked Data Web Challenges, Approaches, and Trends", IEEE Computer Society, volume 16, number 1, pp. 24-33, 2012.
[8] Christian Bizer, "The Emerging Web of Linked Data ", IEEE Intelligent system, Vol. 24, pp. 87-92, 2009.
[9] Sonya Abbas and Adegboyega Ojo, " Towards a Linked Geospatial Data Infrastructure", Springer, LNCS 8061, pp. 196-210, 2013.
[10] Michael Hausenblas, "Exploiting Linked Data to Build Web Applications", IEEE Internet Computing, 2009.
[11] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer,Christian Becker, Richard Cyganiak, Sebastian Hellmann, " DBpedia - A crystallization point for theWeb of Data", Web Semantics: Science, Services and Agents on theWorldWideWeb 7, pp. 154–165, 2009.
[12] M. Scott Marshall, et al, " Emerging practices for mapping and linking life sciences data using RDF — A case series", Web Semantics: Science, Services and Agents on the World Wide Web 14 ,pp. 2–13, 2012.
[13] Rupal Gupta, Sanjay Kumar Malik, " SPARQL Semantics And Execution Analysis In Semantic Web Using Various Tools", IEEE Computer Society ,2011.
[14] Elisabeth ILIE-ZUDOR, "The RFID Technology And Its Current Applications", MITIP, pp.29-36, 2006.
[15] Monica Chauhan,et al, " A Survey on RFID Technology ", International Journal of Research (IJR) Vol-1, Issue-10 , 2014.
[16] Zachary G. Ives, et. all , " An Adaptive Query Execution System for Data Integration" , ACM, pp. 299-310, 1999.
[17] Klaus Finkernzeller, " RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification", Book, john wiley&Sons, 2003.