

Clinical Disease Registry Ontology for Blood Transfusion

Hisham Ammar Gadi, Riyadh Al-Shammari, and Muhammad Imran Razzak

Department of Health Informatics, College of Public Health and Health Informatics

King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

hishamgadi_13@live.com, {alshammari, razzakmu}@ngha.med.sa

Abstract: *This paper discusses the subject of clinical disease registry based ontology. Disease registry is used to aid blood bank staff to utilize their resources effectively to improve service outcomes. In this paper, we presented ontology to incorporate elements important to patient blood transfusion; blood group, disease and date of next transfusion. The ontology was built using Protégé program that utilizes the semantic web approach in designing ontologies. Using SPARQL queries, selected identification elements were selected and shown clearly and properly.*

Keywords: *SPARQL, blood transfusion, clinical ontology, Protégé*

1. Introduction:

Disease registry is a system that collect patient data (demographic and disease related) in order to improve the provision of care or to help healthcare providers to plan properly for future. It plays very important role in post marketing surveillance of pharmaceuticals. Disease registry can provide health providers with reminder to check certain tests in order to reach at certain quality. Patient disease registry can be classified into: Product Registries, Disease or Condition Registries and Health Services Registries [1]. Product registries are concerned with the type of medicine or medical device used by patients. A disease or a condition registry is concerned to identify and collect data form a particular disease, injury or medical condition that might be on a national or international level. The health service registry is mainly concerned with collecting data to measure the outcome of the service or care that is provided for the patient. The purpose of this paper is to design an ontology that might be helpful in aiding health service registry, or clinical registries to provide a better care and service while at the same time improving the efficiency of the healthcare facility. According to Orphanet report in 2014, there are currently 641 rare disease registries in Europe that covers the regional national level [2]. We designed an ontology that will target specific group of blood diseases that require regular transfusion for the patient. In busy and large blood banks, shortage of some blood products is always an issue. To be able to identify specific patients who are requiring multiple transfusions and to be able to prepare or to forecast the needs of the blood bank inventory will be extremely helpful for the operational need of the blood bank activity.

2. Related Work

A patient registry can be defined as software system that uses observational study methods to collect uniform data to evaluate specified outcomes for a population defined by a particular disease, condition or exposure and which serves one or more predetermined scientific, clinical, or policy purposes [3]. It consists of a database to read from, the registry application and the set of rules to apply to the database. In fact, diseases registries are an extremely valuable data source, whether to collect data on diagnosed patients or to provide relevant data that are meaningful to healthcare providers and researchers for patients or people who are not involved in clinical trials. Furthermore, registries can provide a helpful source to aid healthcare providers and

researchers to develop the current diagnostic, treatment and data collection methods as well as aiding them to adapt to the continuous changing in the practice of medicine. Good design and use of registries, however, requires strong understanding of both the potential for bias that threatens all observational studies and the methodological and operational tools that can be used to minimize the influence of such biases as stated by Gliklich [4].

The main and obvious purpose of patient registries is to document the natural history of the disease, with or without treatment. The natural history may be variable from one population group to another and usually changes over time, especially if a new therapy becomes available as stated by Clarke [5]. Rare diseases especially in the context of increased life expectancy and ageing. In general, the benefits of patient disease registries can be summarized in the followings by Gonzalez [6].

- A. Obtain 'real world' therapeutic effectiveness and safety data.
- B. Large patient numbers can detect rare adverse events.
- C. Heterogeneity among numerous investigative sites.
- D. Research collaboration with interactive communication & data reporting to investigators.
- E. It will make better use of regular diagnostic procedures than research or hypothetical procedures.
- F. Can be conducted in any phase of product development.
- G. Flexible, multi-faceted, e.g., sub-studies.
- H. Study subjects are heterogeneous such as
 - a. Various treatments.
 - b. Concomitant meds.
 - c. Co-morbidities.
- I. Hypothesis generation when an a priori hypothesis is difficult to define.
- J. Supportive data for label extensions.
- K. Evidence-based medicine for outcomes & reimbursement.
- L. Cost effective on a per patient basis.

However, data collected from patient disease registry doesn't always provide all the needed information. Because patient diseases registry data has many limitations such as [3]:

- A. Non-randomized data cannot be used for promotional or competitive claims.
- B. Data generally not 100% verified.
- C. Variability in data definitions, interpretation, abstraction & collection intervals.
- D. Selection bias due to non-sequential patients.
- E. Inability to perform desired analyses due to limitations of data captured.
- F. Capture of irrelevant data that is not, or cannot be reported.
- G. Analysis of observational data requires experienced biometrics personnel.
- H. Perceived diminished value of research evidence than controlled trials.
- I. Journal reviewers may be less accepting of observational data.

Although patient diseases registry is considered an important tool in hospitals, many studies such as Gonzalez [6] have suggested that patient diseases registry is still facing many operational challenges such as:

- a. Research naïve investigators & sites
- b. Site may not have a trained Study Coordinator
- c. Enrolling & training large numbers of sites
- d. Capture & cleaning of large volumes of data
- e. Site & patient retention
- f. Determining the appropriate balance of on-site/escalated monitoring vs. remote site management.

An example of an effective implementation of disease registry is Chronic Disease Management System to identify and manage diabetes mellitus patients in Singapore [8]. The System links between healthcare facilities database to make a proper tool for monitoring the outcome and the prognosis of diabetes patient. Another example is the Niemann-Pick disease type C registry [9] which is a rare neurological disease that requires more knowledge and information to help healthcare providers to have a better diagnosis and treatment procedures to deal with it. The registry was collaboration between healthcare providers in 17 countries that till 2012 was able to collect data from 163 identified patients.

3. Disease Registry Ontology

The ontology was build using Protégé® 4.3.0. there were no similar ontologies to be found that deals with topic and purpose in regarding of blood bank related blood transfusions, however, because the overall concept have some similarities with standard disease registries, the ontology was designed and build to focus on patients Number of Blood Transfusions , Blood Group and Specific Diseases. During the search for a related ontology to this work, a related ontology was found that has some similarities in the concept as deals with hospital specific activity. Buranarach designed ontology to create a reminder system of the purpose of visits for chronic patients [10]. The ontology proposed in this paper deals with frequency of transfusion for hematological diseases , so the main concept of both ontologies is to aid the hospital in improving the service and care provided, however, the structure, purpose, data and goals intended of both ontologies are different.

The designed ontology is defined by the following classes:

1. Patient
2. Blood Group (that include the subclasses of A,B,AB and O blood groups)
3. Disease (That include the subclasses Sickle Cell Anemia, Thalasemia And Thrombocytopenia)
4. No Of Transfusion
5. User (That include the subclasses Physician and Blood Bank Technician)

Data properties include; Admission Date, Blood Group, Disease, Medical Record Number (MRN) , No Of Transfusions, Next transfusion Date Patient Name and User ID. Object properties include has blood group, has disease and has No of tans fusion. Relations between the individuals were created using the data properties.

After entering the data for the patients , the following is required to be able to locate and identify the expected results based on the queries entered in the SPARQL queries (Note: if the relations were wrong or invalid, the queries won't run and show no result). The first query tested was to identify patients with B blood group; (**SELECT ?PatientName ?BloodGroup WHERE { ?x CR:PatientName ?PatientName . ?x CR:Bloodgroup ?BloodGroup FILTER (?BloodGroup = "B") }**) and the result can be seen in Figure 1.

```

SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/XMLSchema#>
PREFIX http: <http://www.w3.org/2005/01/rdf-schema#>
PREFIX CR: <http://www.semanticweb.org/ivens/ontologies/2014/4/LitMed-ontology-948#>

SELECT ?PatientName ?Disease WHERE { ?x CR:PatientName ?PatientName . ?x CR:Bloodgroup ?BloodGroup FILTER ( ?BloodGroup = "B" ) }

```

PatientName	Disease
"Steve"^^<http://www.w3.org/2001/XMLSchema#Name>	"Thrombocytopenia"^^<http://www.w3.org/2001/XMLSchema#string>
"Omar"^^<http://www.w3.org/2001/XMLSchema#Name>	"Thrombocytopenia"^^<http://www.w3.org/2001/XMLSchema#string>

Fig. 1: SPARQL query showing B blood group patients

The query was able to locate the patients Carlos and John who both share a B blood group. The second query entered was to locate patients who are diagnosed with Thrombocytopenia; (**SELECT ?PatientName ?Disease WHERE { ?x CR:PatientName ?PatientName . ?x CR:Disease ?Disease FILTER (?Disease = "Thrombocytopenia") }**) and the results can be seen in **Figure.2**. As seen in the result from the query, it was able to locate the patients **Steve** and **Omar** who both are diagnosed with Thrombocytopenia. The third query is to identify patients who have multiple transfusions (for the sake of to show the success of the query the No Of Transfusion of 3 was selected); **SELECT ?PatientName ?NoOfTransfusion WHERE { ?x CR:PatientName ?PatientName . ?x CR:NoOfTransfusion ?NoOfTransfusion FILTER (?NoOfTransfusion = 3) }**. The results can be seen at **Figure. 3**. The third query located the patients Omar and Carlos who both received 3 blood transfusions.

Also a fourth query was entered to show up all the patients with their names, admission date , MRN number and No Of transfusions for each patient; **SELECT ?Patient_MRN ?Patient_Name ?Admission_Date ?NoOfTransfusion WHERE { ?x CR:MRN ?Patient_MRN . ?x CR:PatientName ?Patient_Name . ?x CR:AdmissionDate ?Admission_Date . ?x CR:NoOfTransfusion ?NoOfTransfusion }**. The result can be seen in **Figure.4**. the result showed all patients with the requested informations along with it.

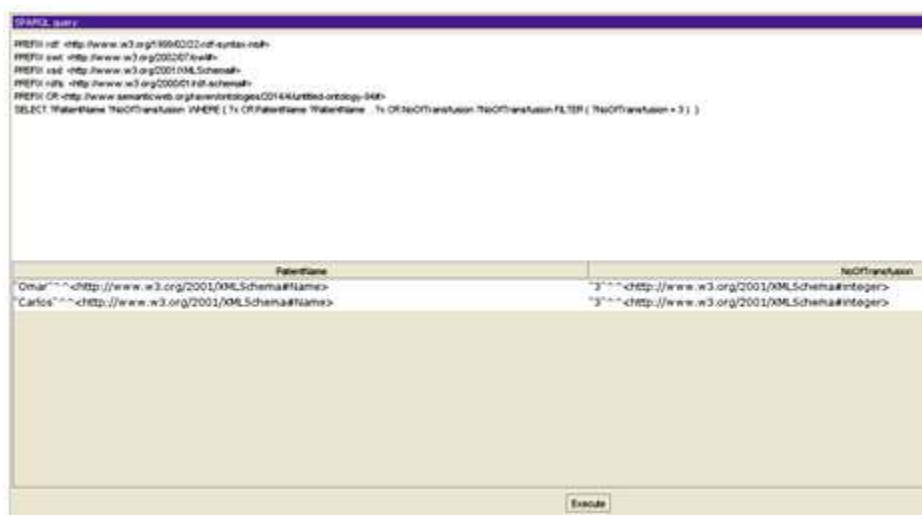


Fig. 2: SPARQL query showing Thrombocytopenia patients



Fig. 3: SPARQL query showing Patients who have similar number of transfusions

4. Result and Discussion:

As seen previously in the results, the search queries managed to locate each request and its relation to the patient. For the blood bank to know how many patients have similar blood group will help them to anticipate the shortage of such blood group. Also, knowing the number of transfusions of each patient and the designated blood group will assist the blood bank in planning for what type of blood group they need to stock up. Also, blood bank staff can locate when the next transfusion date with the patient MRN, Name is and blood group as seen in Figure.5 which will leave them to prepare in advanced for the blood groups intended for the next transfusion.

All of the previous result showed extreme usefulness in locating the desired data and required information to provide the best care for patients that require constant transfusion. Some of the issues that blood bank staff face on regular basis is that the face shortage of some blood groups prior to next date of transfusion or that they don't get notified of the next transfusion date in advance. This ontology if incorporated with the hospital database which contain most of this information as it needs it to identify similar properties entered in the ontology, it will provide a great opportunity for the blood bank to improve the efficiency and overall productivity of the service it provides while at the same time reducing the waiting time for patients and manage their appointed transfusion effectively and in a safe manner. Also, it will help the hospital management to identify what improvement they need to implement for the blood bank (more space, more staff, increasing blood components storage are just a few examples). It can help to identify the patients with the selected disease in order to prepare better diagnosis and treatment procedures for such patients. In order for all of that to succeed, the link between the ontology must be proper and successful because without proper information and data, no ontology can be successful.

The screenshot shows a SPARQL query interface with the following query text:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX OR: <http://www.semanticweb.org/ontology/2014/4/ontology-2014#>
SELECT ?Patient_MRN ?Patient_Name ?Admission_Date ?NoOfTransfusion
WHERE {
  ?Patient_MRN ?Patient_Name
  ?Patient_MRN ?Admission_Date ?Admission_Date
  ?Patient_MRN ?NoOfTransfusion ?NoOfTransfusion
}

```

Patient_MRN	Patient_Name	Admission_Date	NoOfTransfusion
222992	Omar	11-11-2013	2
756700	Omar	09-01-2014	3
647599	Zhor	27-10-2013	3
487443	Carlos	22-12-2013	3
005843	Steve	25-12-2013	3

Fig. 4: SPARQL query showing Patient Name, MRN, Admission Date and No of received Transfusions

The screenshot shows a SPARQL query interface with the following query text:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX OR: <http://www.semanticweb.org/ontology/2014/4/ontology-2014#>
SELECT ?Patient_MRN ?Patient_Name ?BloodGroup ?NextDate
WHERE {
  ?Patient_MRN ?Patient_Name
  ?Patient_MRN ?BloodGroup
  ?Patient_MRN ?NextDate
}

```

Patient_MRN	Patient_Name	BloodGroup	NextDate
222992	Omar	A	08-08-2014
756700	Omar	B	19-06-2014
647599	Zhor	A	20-08-2014
487443	Carlos	B	19-06-2014
005843	Steve	O	20-08-2014

Fig. 5: SPARQL query showing Patient MRN, Name, Blood Group and Next transfusion date

5. Conclusion

We built the ontology for blood transfusion by using Protégé® 4.3.0 and utilized the semantic web approach in designing ontologies. Using SPARQL queries, selected identification elements were selected and shown clearly and properly. As demonstrated and discussed above, the success and failure of the proposed ontology will rely on the data set it will connect and read to. The design of the ontology is not permanent as adjustments and additional criteria might be included according to the need and desires of the organization. Such ontology will be beneficial for researches studying the epidemiology as an example. Sickle Cell Anemia patients and the frequency of their need of blood transfusion as it will provide valuable research data.

6. References

- [1] Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. (Prepared by Outcome DECIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHS290200500351 T O1.) AHRQ Publication No. 07-EHC001-1. Rockville, MD: Agency for Healthcare Research and Quality. April 2007
- [2] Orphanet Report Series, Rare Disease Registries in Europe January 2014 <http://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf> retrieved on 18th of May 2014.
- [3] Newton, John and Garner, Sarah: Disease Registers in England, A report commissioned by the Department of Health Policy Research Program in support of the White Paper entitled Saving Lives: Our Healthier Nation, Published by the Institute of Health Sciences, University of Oxford, Old Road, Oxford, OX3 7LF, ISBN 1 8407 50286, February 2002.
- [4] Gliklich, Richard, Dreyer, Nancy A., Leavy, Michelle B., Velentgas, Priscilla and Khurana, Laura: 2012, Standards in the Conduct of Registry Studies for Patient-Centered Outcomes Research, Guidance Document for the Patient-Centered Outcomes Research Institute, Outcome, A Quintiles Company, 201 Broadway, Cambridge, MA 02139
- [5] Clarke JT, Leavy MB. Case Example 29: Data collection challenges in rare disease registries. In: Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 2nd ed. (Prepared by Outcome DECIDE Center [Outcome Sciences, Inc. d/b/a Outcome] under Contract No. HHS290200500351 TO3.) AHRQ Publication No. 10-EHC049. Rockville, MD: Agency for Healthcare Research and Quality. September 2010. 247-248.
- [6] Gonzalez, Ferreira I, Marsal JR, Mitjavila F, Parada A, Ribera A, Cascant P, 2009. Patient registries of acute coronary syndrome: assessing or biasing the clinical real world data? *Circ Cardiovasc Qual Outcomes*. 2009 Nov;2(6):540-7.
- [7] Hernberg-Ståhl, Elizabeth, 2011, Making the most of patient registries, Late Phase Solutions Europe AB, Rare Disease Research Forum- Challenges and Solutions Stockholm KI, February 21, 2011.
- [8] Matthias Toh, Helen SS Leong And Beng Kuan Lim, 2009, Development of a Diabetes Registry to Improve Quality of Care in the National Healthcare Group in Singapore, *Ann Acad Med Singapore* 2009;38:546-51
- [9] Patterson Marc C, Mengel Eugen, Wijburg Frits A, Muller Audrey, Schwierin Barbara, Drevon Harir, Vanier Marie T and Pineda Mercé, 2013, Disease and patient characteristics in NP-C patients: findings from an international disease registry, *Orphanet Journal of Rare Diseases* 2013, 8:12 doi:10.1186/1750-1172-8-12 <http://dx.doi.org/10.1186/1750-1172-8-12>
- [10] Buranarach Marut, Chalortham Nopphadol, Thein Ye Myat And Supnithi Thepchai, 2010, Design and Implementation of an Ontology-based Clinical Reminder System to Support Chronic Disease Healthcare, *IEICE Transactions* 01/2011; 94-D:432-439. DOI:10.1587/transinf.E94.D.432, 2010 <http://dx.doi.org/10.1587/transinf.E94.D.432>