

A Linked Data Application Development Framework (LDADF) *

Yusuf Mashood Abiodun

Faculty of Communication and Information Sciences,
Department of Computer Science, University of Ilorin, Ilorin, Nigeria.
yusufmashoodabiodun@gmail.com

Abstract: The launch of Linked Open Data Project in 2007 has resulted into making the Web a giant global data space hosting millions of linked RDF data triples on various domains in Linked Open Data Cloud. These data are now freely available for reuse. However, while some Web developers have developed notable Linked Data applications consuming and integrating different types of linked open data from the Linked Open Data Cloud to provide valuable services to people, many Web developers are yet to understand Linked Data principles and standards due to difficulties they face in learning semantic and linked data technologies. This research therefore proposes a Linked Data Application Development framework to help Web developers in overcoming challenges associated with designing and implementing linked data applications.

Key Words: linked data applications, Spring framework, development, web developers, semantic web, object-oriented frameworks

1. Problem Description

The launch of Linked Open Data Project [1] in 2007 has resulted into a giant global data space consisting of reusable linked RDF data on various domains.

However, while there are developers building Linked Data applications that are consuming data from Open Linked Data Cloud to provide useful services to people, there are still many web developers that are unable to apply the linked data principles in developing linked data applications. This is evident in the slow pace adoption of linked data among many web application developers. World Wide Web Consortium confirmed the slow pace of adoption by setting up a new Working Group called Data Activity with a mission to support data providers in publishing their data and also providing supports that will ease the development of linked data applications for application developers. In its inaugural minutes [2] of meeting held on 26th February, 2014, members of the group unanimously agreed that there is lack of awareness of Linked Data standards among the web application developers which make them to be put off when they are informed of Linked data.

Some of the contributing factors discouraging developers from adopting linked data technology include lack of necessary competencies, tools, methodological & frameworks supports, difficult in learning linked data standards, lack of guidelines, reusable libraries etc [5,8, 9]. Though some of these problems have been addressed. For instance, W3C has published a lot recommendations and best practices for publication of RDF/Linked Data.

* Research Supervisor: Dr. R. G. Jimoh

Also, there are now various tools and libraries that could be used in developing linked data applications. With the progresses made so far, developers are still having challenges. While developers new to the field of semantic/linked data are facing problem of complexity in learning the technology, the earlier adopters are also having challenges in implementing of linked data applications. The authors of [8] suggested these problems could be addressed through the adoption of conventional software engineering practices with the linked data design approaches such as (i) guidelines, best practices and design patterns; (ii) software libraries; and (iii) software factories to provide a ready-made infrastructure for the developers. However, the authors only briefly described each of the design approach with identification of relevant tools, but failed to develop the desired infrastructure.

It is in view of the aforementioned above that this research seeks to adopt object-oriented software engineering practices in developing a framework for linked data application. The proposed framework will identify and integrate the existing Linked Data/Semantic libraries & APIs such as Jena, Sesame, Silk, RDR, etc. The framework will also be integrated with a popular object-oriented framework already familiar to developers. This will enable the developers to easily adopt linked applications and also reusing the beneficial features of spring framework which simplify development of enterprise applications.

2. Research Questions and Hypotheses

The goal of this research is to develop a framework that will simplify the learning curve of linked data technology for developers and thereby making it easy for them in reusing linked data design models and codes for designing and implementing linked data applications. In order to achieve the stated goal, the research will investigate the following research questions:

- (i) What are the main features of Linked Data applications that distinguish them from conventional web applications?
- (ii) What are the main software components that usually constitute the architectures of linked data applications?
- (iii) How can the use of framework ease the learning curve for new developers coming into the world of semantic web/linked data applications?
- (iv) How can the use of frameworks help developers who are earlier adopters in overcoming challenges associated with the design and implementation of linked data applications?
- (v) Of what benefits will the use of framework be for developing linked data applications?

The research hypotheses are as follows:

- (i) Defining the features of linked data applications will assist developers in deciding when to apply linked data technology for developing web applications.
- (ii) Using of software engineering practice and frameworks will enable more developers to embrace linked data technology.
- (iii) Integrating a linked data application framework with a popular object-oriented framework will simplify the design and implementation of linked data application for the developers.

- (iv) Use of framework for implementing linked data applications will enable developers to deliver application within a short period through extensive reuse of analysis model, design model and codes.
- (v) Integrating a linked data application with an existing object-oriented framework will enrich the features of linked data applications.

3. Relevancy:

Open Data Movements globally have been leading campaigns calling for the adoption of open data by governments of the world in order to make them more transparent and accountable to their people. Governments such as UK, US and some others have responded by launching data portals hosting thousands of datasets on various public sectors with the aim of promoting transparency and also improving their economies. Through the Open Linked Data project launched in 2007, many datasets from government data portals have been transformed and republished as RDF data in line with the principles of Linked Data. The efforts have resulted into a huge cloud of linked datasets. Web developers are expected to reuse the published datasets from the governments' portals and Linked Open Data Cloud in building Linked Data applications that will provide valuable services to people as envisaged by open data movements. However, while some developers have used these opportunities to build linked data applications, many developers are yet to adopt and apply linked data principles and standards due to earlier stated reasons above. We hope this proposed framework will help more web developers in adopting Linked Data principles and standards to build more valuable applications for the betterment of people. In addition, the research will also contribute to the on-going efforts of truly making the Web a Web of documents and data.

4. Related Work:

A group of researchers [8] did a good work to design a conceptual architecture for linked data application and also identified all the interacting components that make up the architecture. They also advocated for the use of software engineering and design approaches i.e (1) guidelines, best practices and design patterns; (2) software libraries; and (3) software factories to provide ready-made solutions for developing linked data applications. They only briefly described each of the design approach with identification of relevant tools, but did not build a ready-made solution as envisaged.

A related work to this research is the development of a flexible integration framework for semantic 2.0 applications [9]. The research was successful in developing a framework and also integrated it with an object-oriented framework i.e Rubby on Rail framework. However, the framework is not suitable for linked data applications because it was developed as at the time the linked data technology was just evolving.

In the Linked Data book [6], the authors laid the foundation guide for the development of linked data applications by providing architectural patterns and components for implementing linked data applications. They went further to describe the main tasks and techniques needed for developing linked data applications. However, descriptions provided are difficult to comprehend and apply by web application developers that are just coming to the world of semantic web and linked data communities.

Also following the same step of Linked Data book was EUCLID project [7], a two years project funded under the EU Seventh Framework Programme for the purpose of providing comprehensive educational curriculum to the real needs of data practitioners. One of its lecture deliverables was a guide on building linked data applications. Also, the architecture of linked data application based on patterns, layers and components was described. The guide only captured information on the Linked Data application development frameworks

Another research [3] performed an empirical survey of 98 Semantic Web applications, in order to identify the most common shared components and the challenges in implementing these components. In their findings, the authors observed that though not explicitly stated, most of the applications surveyed applied principles of Component-based software engineering (CBSE) in implementing their applications. The authors only recommended the use of CBSE in building semantic applications.

There are also various life cycles that have been developed for publishing linked data on the web of data. These life cycles only focus on publishing linked data; neglecting processes to be carried out by developers to develop linked data application that can consume and manipulate data. Developers therefore found it difficult to comprehend and apply the life cycles within the context of software engineering.

5. Proposed Approach & Preliminary Results:

The research will adopt process for developing Object-Oriented Frameworks. In addition, refactoring method will also be used to capture aggregation and reusable components for linked data application development framework. The proposed approach will include the following steps:

1. **Definition of characteristics usually possessed by semantic web/linked data application:** In order to define the characteristics, we are adopting the NeOn methodological [5] process which set questionnaires for deriving semantic application characteristics based on the following three dimensions.
 - i. Questionnaire about Ontologies: help the application developers to determine the characteristics of the ontologies that the application will make use of.
 - ii. Questionnaire about Data: help the application developers to determine the characteristics of the data that the application will consume or manipulate and its relation with the ontologies or data schemas which data may conform to.
 - iii. Questionnaire about Reasoning: help the application developers to determine the characteristics of the reasoning that the application will apply to the ontologies and data.

The questionnaire about the data dimension will be extended to capture more requirements based on Linked data principles because they are not fully captured in NeOn.

2. **Domain Analysis for Linked Data application domain:** in order to identify and characterize the problem domain, the research will adopt the following steps as stipulated in [10]:
 - i. Outline the situation and the problem: the problem domain area is semantic web/linked data applications.
 - ii. Examine existing solutions: this involves selection of semantic web/linked data applications using the defined characteristics in order to identify and extract the general functionalities common to all of them..
 - iii. Identify key abstractions: applying component based software engineering process to obtain software components in line with the identified functionalities.
 - iv. Identify what parts of the process the framework will perform
 - v. Ask for input from clients and refine the approach

The above steps will result into obtaining domain analysis model as presented in figure 1 below:

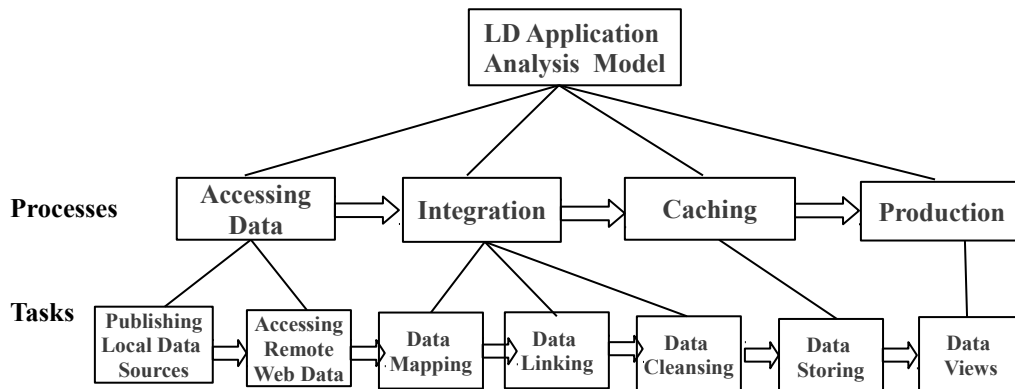


Figure 1: Analysis Model for Linked Data Application Framework

3. **Framework design:** this will be carried out following the steps illustrated in [11]. A framework design is a software design that, when implemented, provides the general and abstract functionality identified in analysis model [11].

The two main sub-processes are:

- i. **Architectural Design:** the objective is to identify the objects/components that constitute the system and how they collaborate using the analysis model as input. Activities involved are:
 - (a) **Identify Abstractions:** refining the analysis model tasks to obtain high-abstracted classes from which the clients instantiate. The derived three main classes' Names from the analysis model which client instantiate are:



Figure 2: The Derived Main High-Level Abstracted Classes.

For effective design and implementation of the proposed framework, the identified three above High-abstracted Classes will be classified as sub-frameworks.

- (b) **Identify Generic Design Solutions:** studying the sub-frameworks and reuse design patterns to provide solutions. For instance, Strategy Design Pattern through use of Composition [12] will be reused to provide solution for DataAccessing Class/Sub-framework as shown in figure 3 below:

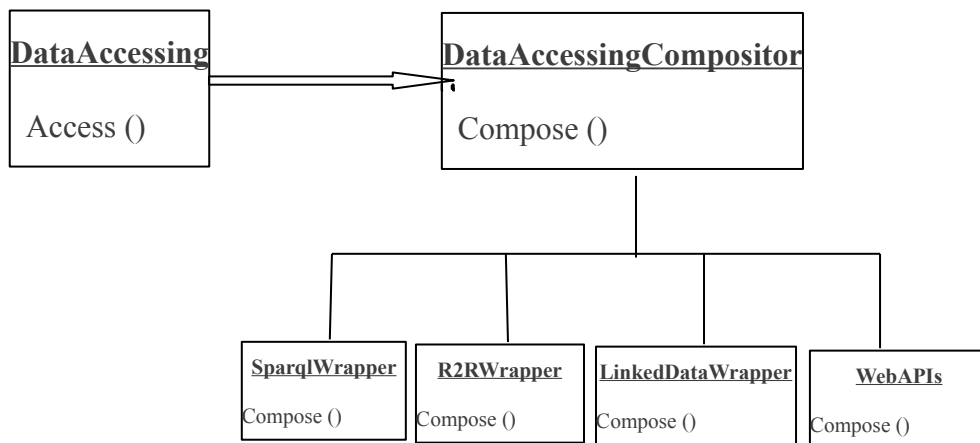


Figure 3: Use of Strategy Design Pattern for DataAccessing Sub-framework

- (c) **Assigning of Classes'/Objects' Responsibilities:** these under construction
- ii. **Detailed Design:** involves detailed definition of collaborations among the objects of the architectural design. Under construction.

Composition of Linked Data Application Framework with Spring Object-oriented Framework:

The proposed linked data application framework will be integrated with the Spring Object-oriented framework in order to make the linked data application framework accessible to many developers who are already familiar to Spring Framework. The integration will enable the linked data application to draw on the strengths and benefits of Spring Framework which include :

- i. Spring framework is java based platform and also most of the existing RDF libraries are implemented using java programming language.
- ii. Spring Core Container Infrastructure for managing the life cycle of application objects.
- iii. Spring MVC Framework for achieving the separation of Concerns.
- iv. Spring Aspect Programming Infrastructure

4. Framework Implementation Phase: the objective of the implementation phase is to implement the objects, the relationships and the collaborations identified in the design phase [11].

Core activities to be carried out according to [10], are:

- i. Implement the core classes
- ii. Test framework
- iii. Ask client to test the framework
- iv. Iterate to refine the design and add features

7. Evaluation Plan:

In order to evaluate the research hypotheses, the developed linked data application framework will be tested by the researcher using it to develop a linked data application for Nigerian National Petroleum Corporation. The agency currently publishes on its website the Oil & Gas Statistical Data (Monthly, Quarterly & Annually) on various activities ranging from upstream, midstream and downstream in PDF format. This data will be converted to RDF data and also integrated with other relevant data from the web to derive new valuable linked data for the organization. The domain expert in the industry will be involved in the development process. In addition, the framework together with its documentation will be made publicly available to developers to access and use it in developing linked data applications. Thereafter, questionnaires will be designed for developers to answer to measure research hypotheses such as shortening of development period with the use of framework, ease of learning linked data technology through the use of the framework etc.

8. Reflection:

Developing the proposed linked data application framework following the conventional software engineering and existing object-oriented framework development process will enable the successful development of the proposed framework. In addition, the integration of the framework with a popular object-oriented framework, Spring framework will make linked data application technology accessible to more developers.

Acknowledgements

I sincerely wish to express my deep gratitude to my Supervisor, Dr. R. G. Jimoh for his constructive and supportive critique of my work. His mentoring role is enabling me to build my research skills.

References

1. *State of the Open Linked Data Cloud*. Retrieved April 18, 2014 from <http://lod-cloud.net/state/>
2. World Wide Web Consortium: *Data Activity Working Group Inaugural minute of meeting*. Retrieved April 18, 2014 from <http://www.w3.org/2014/02/26-dacg-minutes.html>.
3. Benjamin H, Sheila K, Conor H, & Stefan D.: Implementing Semantic Web applications. In: Proceedings of the 5th International Workshop on Semantic Web Enabled Software Engineering (SWESE 2009).
4. Leigh D., & Ian D. (2012). *Linked Data*, Retrieved on May 23, 2014 from <http://patterns.dataincubator.org/book/>
5. *NeOn Methodology for Building Semantic Applications (2009)*. Retrieved February 9, 2014 from : <http://www.neon-project.org/nw/Deliverables>
6. Heath T., & Bizer C. (2011). *Linked Data Book: Evolving the Web into a Global Data Space*. Retrieved February 12, 2014 from <http://linkeddatabook.com/editions/1.0/>
7. *Educational Curriculum for the usage of Linked Data*. Retrieved February 12, 2014 from <http://euclid-project.eu/>
8. Benjamin H, Richard C, Conor H, & Stefan D: An An empirically-grounded conceptual architecture for applications on the Web of Data . Published in Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions (2012).
9. Eyal O, Armin H, Manfred H, Benjamin H, & Stefan D,: A Flexible Integration Framework for Semantic Web 2.0 Applications. *ieeexplore.ieee.org*
10. [Tal94a] Building object-oriented frameworks, Taligent, Inc., 1994
11. Niklas L. & Axel N,: Development of Object-Oriented Frameworks Authors http://www.researchgate.net/publication/245912789_Development_of_Object-Oriented_Frameworks
12. [Gam94] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides. Design Patterns - Elements of Reusable Object-Oriented Software, Addison-Wesley, Reading, MA, 1994.