Research on E-Science - Serverless Semantic E-Science Framework (SSE-SF)

T. Rubin Jose, P. Sojan Lal

Abstract— :Serverless implementation for a semantic e-science framework (SSe-SF) is all about pushing code to a compute service and networking with third party services and APIs to get the work done. The underlying infrastructure both hardware and software are hidden from the user. Serverless semantic e-science framework (SSe-SF) includes Knowledge Search and Navigation, Identity Management, URI/ Content Negotiation, RDF & RDF Schema annotated information resources, Shared Ontologies, Read-only Onto-Repository & Vocabulary, Onto-learning and merging, Semantic reasoners and Semantic data storage. SSe-SF is applied on computing the Outcome bases education attainment (OBE) calculation.

Keywords: Semantic web, eScience, Ontology, serverless architecture, scientific data representation, ontology engineering, linked data.

I. INTRODUCTION

Knowledge processing capacity of humans are an in-born combination of acquired knowledge and experience, when it is copied for an AI machine with computing speed it becomes far better than humans for huge amounts of data. Humans can recall the facts and can add it with reason, will produce new knowledge for his personal use, when it is represented in any form, will be added with the global knowledge network. However, immense growth of the World Wide Web and its free availability across the space diminishes the comprehensive ability of humans when encountered with the already shared huge volumes of data in WWW [1].

II. SCIENTIFIC DATA PRESENTATION

Scientific data presentation is on the threshold of data and information overload, which may lead the budding scientific community with the huge task of re-representation or re-inventing the wheel. High availability of data storage and computing power should be properly channelized for the appropriate representation of scientific as well as other data for semantic / machine readability.

Knowledge Discovery by all means is to find useful shards of knowledge within the data, with little or no human involvement. The existing information retrieval methods in world wide web mostly in the form of web search engines like Google, Bing or Yahoo are still far from perfect, even with the assistance of automated technologies. They rely on

mere string-based statistics and heuristic ranking [2]; actual meaning of the data is rarely accounted for.

III. E-SCIENCE, ONTOLOGY AND SEMANTIC WEB SERVICES

There is no one scientific laboratory in the world that can support all the resources, tools, raw data or expertise to solve and understand the scientific researchers' problems. It is the distributed global collaboration that can solve it to an extent by means of electronic data sharing. This distributed global dissemination via electronic means is called e-Science [3]. Sharing of results, idea and data in many layers will give rise to new hypothesis as well as people to people relation of whom nobody will ever see face to face. Even those who couldn't complete their ambitions in their life time could be achieved by others in their field of research. But there are many problems in the field of e-Science now, the sharing of data is in heterogeneous form, there is no sustainable and effective e-Science infrastructure now. The standard web technology is finding it very difficult to meet the needs of e-Science. Now the scientists manually search the web for content, which is available in different websites, are processed and interpreted by them to integrate ideas to form human understandable knowledge. Here in the form of ontologies semantic web can make a major impact to the e-Science Infrastructure building, where the scientific societies need to be willing to generate, preserve and share the content. A semantic web is expensive to setup and maintain, and thus is only likely to work for communities where the added value is worthwhile and an 'open source data' philosophy prevails [3].

The aim of the semantic web [2] is that the world's information should be available in a machine-understandable form. Semantic science, the application of semantic know-how and cognitive comprehension under ambiguity to the exercise of science, which requires machine-understandable information of three sorts: ontologies, meta data on observations of the world and prediction theories. Publication of data and theories that can inter-operate by common ontologies should be sole aim of e-Science. The theories can be judged by how well they predict unseen data and can be used for new cases.

An ontology [4] is a formal specification of the meaning of the vocabulary used in an information system. Ontologies

enable information sources to interoperate at a semantic level. Interoperable queries which can expand itself with meaningful link (Semantic Links) will generate machine-understandable knowledge. Already verified information from different sources can generate new interpretation as well as new recommendations in e-Science, if ontologies are made use of. Here we focus on a new framework for e-Science, which can make use of the existing as well as forthcoming recommendations in semantic web.

Ontological Engineering normally refers to the routine of activities that include the development process, life cycle and the methodologies, tools and languages for building ontologies. The recent developments in the allied areas have given a smooth transition from traditional methods to agile development for ontological engineering also. Semantic web services in the form of microservices which is a descendant of service-oriented architecture has given rise to free access to many research organization's ontologies. The onto-query maintained by many in open public is in the form of small, impartial, fully autonomous microservices built around a specific ontology. The scattered microservices can never be grouped under a single umbrella, since the developed ontology is not in a unique format. This is one of the main difficulty faced by the researchers in semantic web as well as in e-Science. Secondly the microservices can be written in different general-purpose or domain-specific languages, which may force the researcher to use the right language or a specific set of libraries for the particular job for a single microservice.

IV. SERVERLESS SEMANTIC E-SCIENCE FRAMEWORK (SSE-SF)

A typical three-tier application is usually used to support the main architectural premises of the Semantic Web, which is a stack of languages, often drawn in a figure presented by Tim Berners-Lee in his XML 2000 address [2].
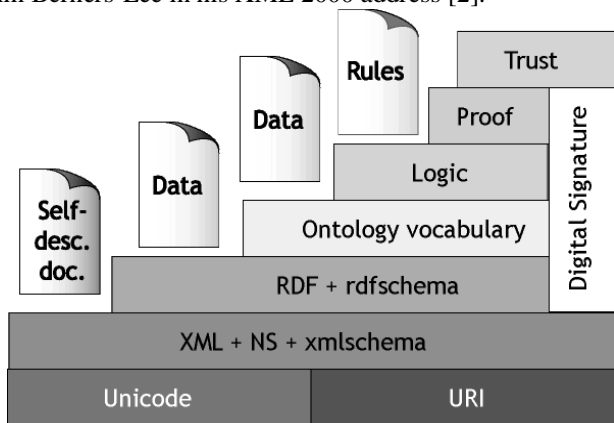

Figure 1 Semantic Web Pyramid proposed by T. Berners-Lee

Software design has evolved tremendously in recent times. Multi- layered and multi-tiered design dealing with individual functionality in each layer or mixed. Even though it looks manageable, too many layers could lead to inefficiencies with poor architectural implementation. A small deviation can often cascade and cause the developer to alter every layer throughout the system, costing substantial time and energy in operation and testing.

The semantic web pyramid (fig .1) has been implemented by many in different platforms. Its tools and frameworks are available in different domains. Protégé - A free, open-source ontology editor and framework for building intelligent systems is one among them developed by Stanford Center for Biomedical Informatics Research.
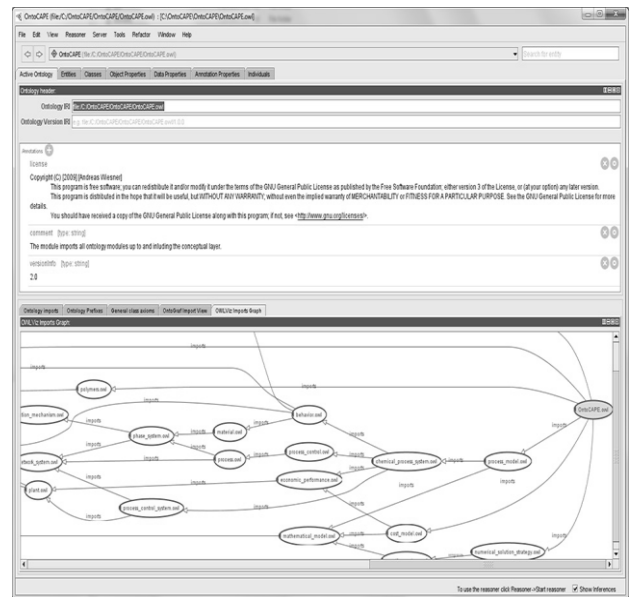

Figure 2 Protégé - ontology editor and framework

Serverless implementation for a semantic e-science framework may be the first of its kind at this time. The serverless technology is still early in its infancy, even though big players like Amazon, Google and IBM already rolled out their flavours of it. It's all about pushing code to a compute service and networking with third party services and APIs to get the work done. The underlying infrastructure both hardware and software are hidden from the user, where the servers still running are maintained by someone else. Server less architecture includes UI design, publish/subscribe infrastructure, workflows/orchestration, active databases, API gateways and management, and data services [5]. A traditional internet delivered application has a client communicating with a long-lived server process that handles most aspects of the application's logic. In a serverless architecture, this applications logic is orchestrated as different functions and handled by third-parties (fig. 3). The five principles of serverless architecture [5] as envisaged by Sabarski are:

1. Use a computer service to execute code on demand (no servers).
2. Write single-purpose stateless functions.
3. Design push-based, event-driven pipelines.
4. Create thicker, more powerful front ends.
5. Embrace third-party services.

The 4 core benefits of Serverless Applications include zero administration, auto-scaling, pay-per-use, increased velocity.
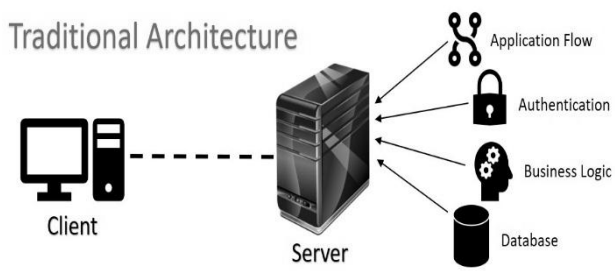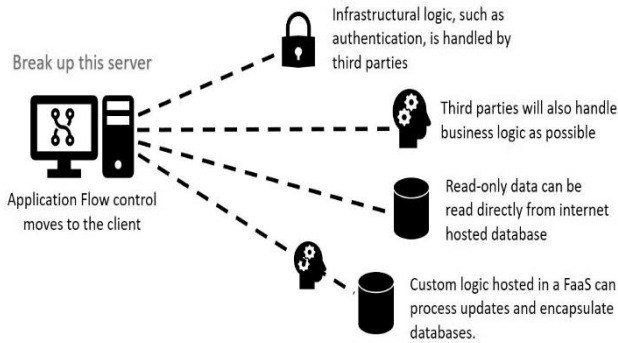
Figure 3 Serverless Architecture

Basically in a serverless architecture the application flow control is moved to the client and an array of third party services are entrusted with infrastructure logic, business logic, databases and custom logic. The server scaling and load balancing are done by the vendor itself in these third parties, thereby the user or the developer is least bothered about horizontal or vertical scaling. Designing a serverless semantic architecture for e-Science based on the above is shown in figure 4. AWS and GCP has an array of services that we can use for building the Serverless Semantic e-Science Framework (SSe-SF)

The Key components in SSe-SF are Knowledge navigation, Identity Management, URL/ Authentication, RDF & RDF Schema annotated information resources, Shared Ontologies, Read-only Onto-Repository & Vocabulary, Onto-learning & merging, Semantic reasoners and Semantic data storage / search.
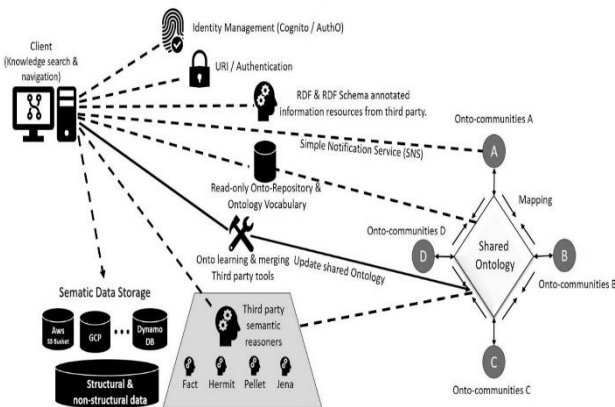

Figure 4 Serverless Semantic e-Science Framework (SSe-SF)

A. Knowledge Search and Navigation

New knowledge is emerged when knowledge workers find the similarities and differences among pieces of information already available, but for that they spend much of the time now. Semantic knowledge management tools deployed on client machine by various vendors can be made use to avoid the delay in this. Knowledge search and navigation shall consist of a combined facility that can explore meta-data annotations as well as traditional free text search [6]. QuizRDF is an example of such tool. These client side tools can be used to fetch a particular node of the ontology and can navigate to its immediate surroundings in order to create appropriate relations from the queries.

B. Identity Management

User identification and management is one of the tedious task in e-Science distributed environment. A third party tool like Amazon Cognito or AuthO, can be used for identity management. Integration with public identity providers such as Google, Twitter, Facebook or importing already populated user directory owned by each research lab to these kinds of third party identity management services providers will ease the difficulties of user identification.

C. URI/ Content Negotiation

Both Web documents and concepts from the real world like organizations, people, topics, things are described using RDF in a computer processable way via URI in semantic web by Web and Ontology developers [7]. 303 URIs and hash URIs, which one to use hinges on the situation, both have advantages and disadvantages while using RDF data and the HTML data. Here the content negotiation is fairly complex, which can be achieved by third party services in SSe-SF.

D. RDF & RDF Schema annotated information resources

The resource description framework(RDF) designed for standardization, which is built on object-attribute-value triple, is used for web-based properties based on meta-data descriptions. RDF Schema with richer representation formalism can be used for basic ontological modelling, where classes, subclasses, subproperties, domain and range properties [8] and so forth can be included by the third party.

E. Shared Ontologies

Solution for Semantic heterogeneity is a shared ontology. Available ontology building methods and tools reveals that ontologies can be built from scratch, by reusing and re-engineering other ontologies, by a process of merging or aligning or by using an ontology learning approach [9]. Five of the prominent methods are Cyc, Uschold & King, Gruninger & Fox, METHONTOLOGY and On-To-Knowledge methods [10]. Most of the approaches are fixated on development activities, especially on the ontology implementation, and they do not pay too much attention to other important aspects related to the management, evolution and evaluation of ontologies. Ontology development tools like Ontolingua, WebOnto, oilEd (Language-dependent tools) and Protégé, WebODE, OntoEdit

(language-independent tools) have been created by organizations to aid in the ontology process. Many of these tools are open source and added functionalities can be enabled through third party plugins.

F. Read-only Onto-Repository & Vocabulary

A public read-only Onto-repository hosted by a third party can reduce the process load on the shared ontology by the onto communities. Shared vocabularies can eliminate ambiguities which may occur by the terms used in the different data sets or in a discovered new relationship. It may also support to organize knowledge using standard formalisms to achieve the real power of linked data. To describe and define different forms of vocabularies in a standard format W3C offers RDF and RDF Schemas, Simple Knowledge Organization System (SKOS), Web Ontology Language (OWL), and the Rule Interchange Format (RIF) [11]. The complexity required by a specific application decides the selection of specific standard.

G. Onto-learning and merging

There are a few ontology learning tools which can assist, to detect new relations, to find and set up new concepts and to build a taxonomy or enrich an existing one. Some of them are ASIUM, LTG Text Processing Workbench or OntoLT, TERMINAE, OntoLearn, KAON and Text-2-Onto [12]. In order to support the shared ontology, third party hosted aligning and merging tools can establish different kinds of mappings either by preserving the original ontologies or can generate a new unique ontology from the original ontologies. FCA-Merge, GLUE and the PROMPT plug-in tools can achieve this for the user.

H. Semantic reasoners

A semantic reasoner or reasoning engine infer logical consequences from a set of asserted facts or axioms, which enable the semantic web to function. Some of the reasoning tools are: DLP, FaCT, RACER, Pellet, Jena, CEL, Cerebra Engine, QuOnto, KAON2, HermiT [13]. Many of them differ in their inference procedures, reasoning and implementation. It's API services will greatly support in the SSe-SF.

I. Semantic data storage for SSe-SF

Structural and non-structural data storage support for the SSe-SF can be achieved by the cloud storage services which supports the REST API. Distributed storage of shared ontologies also can be achieved by cloud services already available.

V. RESULTS BASED EDUCATION ATTAINMENT THROUGH SSE-SF

Outcome Based Education focuses on clearly defined outcomes which is to be attained by the candidate over a period of time. OBE assessments, direct and indirect from various stakeholders and its attainment calculation is cyclic process, where computing power and knowledge population is varied based on each course. Applying the SSe-SF to OBE calculations reduced cost and saved the time.
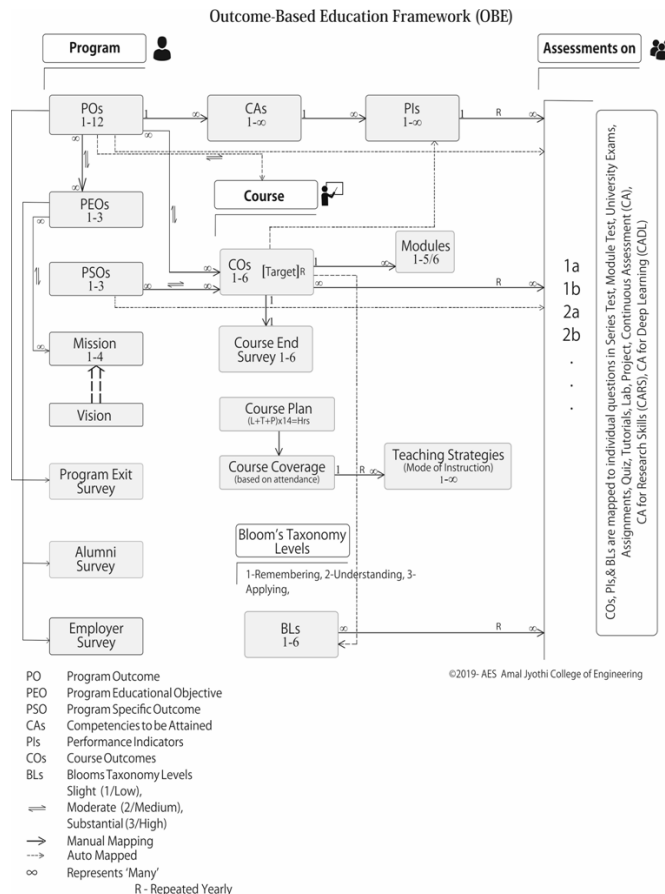

Figure 5 Outcome Based Education Framework

VI. CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

Knowledge Management is a crucial factor and a strategically important capability for success for any enterprise today. It is an interdisciplinary process where humans, organizations, cultural and language considerations play a role along with highly practical information technology support. We have presented SSe-SF with a view to support this knowledge integration. Serverless architecture for the employment of the full potential of semantic web was our aim. Further research areas of SSe-SF will be in the knowledge grid of eScience applications in genomics, astrophysics, bioengineering or biochemistry where enormous data sources are available with no semantic structure. Since life sciences grow very rapidly so does the terminology of the field, where a shared world model that can be supported by many shall be applied

VII. REFERENCES

1. R. Cailliau and J. Gillies, How the Web Was Born:The Story of the World Wide Web., Oxford University Press., 2000.
2. T. Berners-Lee, J. Hendler and O. Lassila, "The semantic web," Scientifc American, p. 5, 2001.
3. C. Goble, "Using the Semantic Web for e-Science: Inspiration, Icubation, Irritation.," in ISWC, LNCS, Springer-Verlag, Berlin, 2005.
4. S. B, "Ontology," in Luciano Floridi (ed.), Blackwell

Guide to the Philosophy of Computing and Information, Oxford: Blackwell, 2003, pp. 155-166.

5. P. Sbarski, Serverless Architectures on AWS, New York, USA: Manning, 2017.
6. Hebeler, M. Fisher, R. Blace and A. Perez-Lopez, Semantic Web Programming, Indianapolis: Wiley, 2009.
7. Davies, R. Studer and P. Warren, Semantic Web Technologies: Trends and research in ontology-based systems, New Delhi: Wiley India, 2012.
8. T. Segaran, C. Evans and J. Taylor, Programming the Semantic Web, USA: O'Reilly, 2011.
9. D. Fensel, J. Hendler and H. Lieberman, Spinning the Semantic Web, London: The MIT Press, England, 2005.
10. G. Perez, a. Lopez and O. Corcho, Ontological Engineering, London: Springer, 2011.
11. W3C, " Vocabularies, W3C World Wide Web Consortium Recommendation," [Online]. Available: https://www.w3.org/standards/semanticweb/ontology. [Accessed 06 September 2018].
12. J. Davis, D. Fensel and F. Van, Towards The Semantic Web: Ontology-Driven Knowledge Management, London: John Wiley & Sons Ltd., 2004.
13. J. Cardoso and A. P. Sheth, Semantic Web Services, Processes and Applications, New Delhi: Springer (India), 2011.

VIII.   AUTHORS PROFILE

Rubin Thottupurathu Jose, MCA, M Tech. Scholar in the area of Semantic query optimization and Ontology Engineering, already published 4 Journal Papers and 8 Conference publications in the area. Attended short course in the Protégé tool in Stanford University, USA. He has got a total of 15 years of experience in the academic field of Computer Science and Engineering.

Dr P. Sojan Lal has more than 30 years of experiences, with international companies in Middle East and educational institutions in India. He has 6 books published under his name. He is an approved research supervisor in many of the Indian universities.

Dr. Sojan Lal been listed in "Marquis Who's Who in the World", representing the world's most accomplished individuals. He authored 65 Journal papers and guided 4 PhD.

1. Qualifications: BE (Mechanical Engineering, 1985), M.Tech(Computer Science, 1993), PhD (Faculty of Technology, 2002), MBA(2011, UK), DBA(2018, USA)