An Approach for Secure Semantic Data Integration at Data as a Service (DaaS) Layer

Shoohira Aftab, Hammad Afzal, and Amna Khalid

Abstract—**Data integration provides a uniform view of a set of heterogeneous data sources and facilitates users to query without any knowledge of the underlying heterogeneous data sources. In current era, Service Oriented Architecture and Cloud computing together has enabled users to access services over the Internet at a low cost. Cloud computing model provides a layer which is responsible for providing data to the other layers and services i.e., Data as a service (DaaS) layer. The issue of providing an integrated view of data can be handled using Semantic data; the data stored in a way that is understandable by machines and integratable without human intervention. However, integrating data using semantic web technology without enforcing any access management will raise privacy and confidentiality concerns. Different data owners store data in heterogeneous format based on their requirements. This leads to the data interoperability problem. This research proposes a framework that would allow data from different sources to be integrated using the concept of semantic data, thus resolving the issue of interoperability and also devises an access control system for defining explicit privacy constraints.**

Index Terms—**Access control management, data as a service (DaaS), data integration, interoperability.**

I. Introduction

In real world the data is stored and managed in different forms. Based on variety of requirements, users prefer different methods and format of data storage. Combination of data from these different sources can bring many useful and hidden results that may not be obtained from single data source. This technique to combine data residing in different sources and providing a consolidated view of data to the users is known as "Data Integration [1]. Data integration is of great importance in solving many problems in commercial as well as scientific domains. Commercial domains include when two organizations integrate their data to get results that would help to earn more revenue. Scientific domain includes integrating data to get better results that would use for betterment of human society.

In database domain, various data integration technologies are available, which include combining data from several disparate sources, then storing it using various technologies and provide a unified view of the data. Data warehouse is used to provide unified view of data from two or more enterprises. Ref. [2] the need for data integration increases proportionally with the recent increase in volume of data and need for sharing of data. Integrating information from multiple sources leads to many obvious advantages. It reduces the efforts of data gathering and enables to infer information that would otherwise be impossible. Data collection is done using different methods and techniques. Hence a large amount of information is stored in this way. By applying proper techniques to combine and integrate this data, much useful information can be retrieved. This useful information can help in assisting many tasks that otherwise would be impossible.

Table I explains some scenario of different domains that would explain the importance of data integration from heterogeneous sources. The information that can be extracted in each case cannot be possible when data sources are viewed in isolation [3].

TABLE I: Real Life Scenario

Domain	Scenario of Data Integration from Heterogeneous Source
Law enforcement agency	Law enforcement agencies such as intelligence agencies benefit from accessing database of various police forces to assist in their effort to fight against terrorism, drugs trafficking and other criminal activities.
Insurance company	Insurance companies can identify many possible fraudulent claims using data from external sources i.e., other insurance company and police records.
Medical researchers	Medical researchers and epidemiologists, with access to records across geographical and ethnic boundaries, are in a better position to predict the progression of certain diseases.

Despite the fact that data integration from heterogeneous sources has many advantages, one of the main problems in integrating data is that data stored in one database is not compatible with other. This shortcoming has given rise to a phenomenon called Data interoperability which is defined as the ability to interpret data correctly beyond its original data source. One useful way of providing data interoperability is to use Semantic Web technologies. Semantic interoperability means the ability of a data source to provide data along with the meaning of data. It is concerned with representation of data i.e., Syntax as well as with transferring of meanings i.e., semantic of data. The main goal of Semantic Data is to organize data in such a way that it can be interpreted meaningfully without human intervention. This is achieved by adding data about data which is known as metadata. Each data element is linked with a controlled and shared vocabulary, also known as ontology which is often developed in XML based technologies that make them independent of any particular information system. [4]. As data from heterogeneous sources have conflicts not only in structure but also in context or value, semantic is used to act as middle layer between them. However its effectiveness is closely tied to the consistency and expressivity of ontology used in the

Manuscript received December 2, 2013; revised March 14, 2014.

The authors are with the National University of Science and Technology, Islamabad, Pakistan (e-mail: shoohira@gmail.com).

integration process.

The approach that we are using is to annotate data stored in multiple sources using common controlled and shared vocabularies (i.e. ontology) [5]. Despite usage of semantic inter-operability to address issue of heterogenous data sources, data owners have been reluctant to adopt the approach, especially in various critical situations when the domains may involve confidential data (health data, national security), or significant business or scientific information [6]. It is important to develop a solution that enables widespread integration and sharing of data, along with easy and effective privacy control of data owners and ensuring data security.

Data as a service enables data access on demand to the data users using internet as communication medium. Data services enable multiple users to access data simultaneously, allow easy access and always provide latest version of data. However, reliability to store and manage data securely is always required to gain customer trust to use data service. Based on several studies of DaaS, security issues that are raised are due to following reasons. First, there is no clear distinction between the service provider and data provider that causes problem for data provider to enforce any privacy requirements. Second, in traditional web service security models, the main focus is on service provider and service consumer, and not on the data provider. Third, handling user permissions and obligations, with the fact that data as a service can deal any type of data, has not been considered [7].

The characteristics of DaaS layer raises the issue that roles of data provider and service provider are not very clearly defined which leads to the issue of access rights [7]. Secondly, while composing the data from different resources; there is not clear solution to handle heterogeneous data. Both factors impact the security and interoperability of data that is used.

A. Example Scenario

In order to understand the problem fully let's take a hypothetical example. Suppose we have data from an education department and a medical scientist wants to get some statistics about which percentage of students from which region are more prone to diseases. To do the analysis, they may request to use the data from education department. The education department allows the researcher to use the data, provided that personal data remains confidential. Now let us consider another scenario where an emergency control system requires having latest contact information including residential details of every student. Education department agrees to share this personal information of students to emergency control system. So based on the above explained scenario the education department can share different types of data with different organizations. In other words, Domain X is data provider and the sharing policy of domain X states that domain X can share data A with domain Y, data B with domain Z and data A and B with domain W.

B. Contributions

This research devises a framework that will help in solving the issues of security and interoperability of Data at DaaS layer. We have called this framework as Secure Semantic Data Interagtor (SSDI). The framework will be providing following features.

- A framework for data integration from heterogeneous resources.
- Access control management from the perspective of data providers and data consumer,
- A Data as a service that would be accessible using internet.
- A solution for handling data interoperability issues.

Rest of the paper is organized as follows. Section II is related work, which mainly focuses on the recent and related research that has been done. Section III is about the proposed framework architecture. Section IV discusses the implementation details and Section V gives concluding remarks.

II. Related Work

This section focuses on the latest research that has been carried out in the field of data integration. The section is organized to covers security and interoperability in data integration, semantic data security and DaaS security issues. The solutions that are already in use and the proposed methods from researchers are discussed. The novelty of our proposed approach against the existing solutions is also discussed.

In [8] privacy has been taken into account for DaaS layer. They have given data annotation technique for applying security policies. Their security policies have considered the difference between data provider and data user. However they haven't considered the data integration perspective for data security. [9] Their work is based on the integration of data by converting it to semantic format. The data is integrated using Jena API [10], and have kept source information in separate knowledge-base and for tracking source information about source is attached to the end nodes.

Another related research [11] involves data mashup's security. Mashup is a special type of web service in which data, computation and user interface elements are combined together to form new application. Data mashup is the type of mashup applications that combine data from different resources, that is available through web services known as Data as a Service. They have proposed to define data privacy need declaratively and then by using of query rewriting approach to ensure that only authorized user have access to data. Their main focus is on the special data class i.e. data mashup, however we have also considered to use this approach in a modified way. First the data owners will define privacy policy and then by using query rewriting security will be ensured. Ref. [7] also provided a security of mashup data. This technique proposes two phase process to provide security solution for web mashup; three way handshake to introduce user to server and select information for session. It provides cryptographic mechanism for providing User authentication, data confidentiality and Data integrity. However we will consider that basic security framework for web-services would be sufficient to authenticate users and provides data integrity as well.

In the domain of semantic data security, there are different approaches available to provide security to semantic data. Ref. [2] proposed an approach to encrypt the nodes in RDF-graph. Their approach consists of transforming the

graph into smaller sub-graph and then encryption of data in that graph but as our solution is specific for data integration, encrypting RDF graph will impose a major overhead. [5] Provided authorization and privacy of semantic web services. They proposed the idea of security policy and privacy in semantic web services using service parameter of OWL-S. Their work focuses on mainly two kinds of security policies; Privacy and Authorization. Privacy policy deals with data confidentiality and authorization policy deals with to accept requests from certain clients only.

Ref. [12] proposed architecture for privacy preserving in Data integration, although they haven't considered issues involved in data interoperability their architecture provides the basis for SSDI framework. This framework have considered two conflicting challenging issues: sharing of critical information for greater good while minimizing privacy backlash. The architecture consists of three major components: the privacy policy formulation framework, the privacy preserving query processing framework, and the privacy preserving mediation engine. The architecture for this approach is basis for SSDI framework as well.

As our model is based on the idea that all the data will be converted into Semantic data hence related research work are also included. [13] Proposed W-graph, a bridge to automatically retrieve ontology from relational databases. They have adopted a medium model method i.e., to convert relational database in to a middle model and then apply ontology to this middle model. Their main achievement is to dynamically adopt to change in ontology or relational database. Ref. [14] Proposed a semi-automated method to convert data present in relational database to semantic web ontology. Their work focus on finding hidden pattern from the data that is stored in relational database and combined with input from domain expert to get more accurate results. The patterns found from data are then used in mapping rules to convert database to ontological language. Ref. [15] developed "OntoGrate", a database 2) Knowledge base 3) World Wide Web 4) semantic data. First, ontology is developed from database schemas. Second, matching are developed in ontology from name, structure and relationships. Third, mappings are developed with the help of input from domain expert about relationship in data. Fourth, data mining techniques are used to find candidate mapping. The interaction between domain expert and integration system is by User Interface. Inference Engine use mapping to execute query as well as, refine rules and perform consistency and redundancy checks.

Many researches are based on the conversion techniques to transform data into semantic format. Ref. [16] RDF123 defines mapping to transform data from spreadsheet to RDF graph. Ref. [17] designed a mapping model so that new RDF documents can easily be designed and updated. Their work focuses on developing flexible and simple programming interface. Ref. [18] D2RQ provides a declarative language for processing and accessing non-RDF database to RDF. Ref. [19] creates mapping from database by first considering particular cases for mapping and then using that mapping to convert ontology query to SQL to retrieve data. Ref. [20] extended the work of DB2OWL, combining data and services provide following benefits Analysis of data, query data from different domain Research on data from different domains.

III. Architecture

Architecture of our proposed Secure Semantic Data Integrator (SSDI) Framework is explained in this section. As already explained in the previous section, main focus of this research is to find a solution for issues of interoperability and security while integrating data from heterogeneous sources.

A. SSDI Architecture

The main emphasis of this research is on data integration technique. SSDI uses the strength of semantic data to address the issue in data integration technique. Along with providing a solution that will ensure the security of data will increase the confidence of reluctant data owners to share their data.

Data sharing is a complex phenomenon that involves sharing with different users on different terms. Our proposed SSDI framework enhances the DaaS layer to combine data from different sources and handle issue of data privacy by deploying an access control management system.

The architecture of the SSDI is explained as follows. Data will be provided by data providers which are also the owner of the data. Data will be converted into semantic format using the ontologies provided by the data owner. Data policies are provided by data owners and stored in the access control management system. After conversion of data into semantic format i.e., RDF the data is annotated by the policies provided by the data owner. All the data will be integrated after and will be stored in the RDF store.

To access the data, end users will query the data which will be processed (parsed and rewritten) by SSDI's access control management to enforce security policies. Retrieved results will be provided to the users.

B. Modules

SSDI framework is divided into the following modules based on their functional properties.

1) Data integration

This module is responsible for integration of data from different sources. Data sources are responsible to provide data along its conceptual model manifested in ontology. Data is then converted into semantic Data. RDF is used for the representation of semantic data. Semantic data will resolve the issue of interoperability. Regardless of the data original format, after conversion to semantic data; data will be compatible with the rest of the data. This involves resolving the compatibility issues along with integrating data from heterogeneous sources. Data providers will also provide privacy policies that will be used to provide access control on the data.

2) Data store

Data is stored in an RDF store after conversion into RDF format. Update or retrieval of data can be made through web services based interface.

3) Access control management

Privacy and access control management service is applied on semantic data. This involves incorporating the data

polices as provided by Data providers and a Query Rewriting module to enforce those policies. Data will be annotated with the provided policies. Query rewriting will be performed to ensure that only access is allowed to only authorize person.

4) Interface

SSDI framework has a main role of interfaces. SSDI interfaces are categorized on user interaction with the system and system internal component interactions. All interfaces are service oriented. SSDI will be using Data as a Service to make it accessible to the users by internet.

IV. Core Techniques

This section describes the design of core techniques that are utilized in designing SSDI.

A. Data Integrator

Data integrator has prime and fundamental importance in SSDI framework. All the data will be converted in the format of semantic data and then united. Data from data owners will transformed and combined together to provide results.

Responsibilities of the module are following:

- Receive data from heterogeneous data sources along with its conceptual model (ontology).
- Gather security requirements from each data owner
- Capture and maintain security policies. Users will define their security requirement by allowing domains to access data.
- Transformation of data to semantic format i.e., RDF
- Interpret security policies by annotation of semantic data, allowed domains information will be attached with the nodes.
- Binding all data together to form an integrated data repository. This repository will be used to

Data integrator is divided into further three modules, as shown in Fig. 1; each sub module is responsible for a set of functionalities. These submodules are named as *Collect*, *Convert* and *Combine* are the sub-modules.

1) Collect

This sub module is responsible to collect data and security policies from data owners. Interface between data owners and this module is via web services. Data owners will interact with web services to provide data and security policies.

Input parameters include Data files (they can be in variety of formats such as excel, csv etc), security requirements, and ontologies. Data files are mandatory as an input, while security requirement and ontology can be left blank. In case security requirement and ontologies are left blank maximum security i.e., accessible to only data owner and global ontologies files will be considered.

Output of this component will be comprised of data files, security requirements with respect to user i.e., data owner information. Ontologies will also be an output parameter either provided by data owner or SSDI ontologies will be used.

2) Convert

Convert is the Second module in data integrator. Its main responsibility is to provide conversion facility. Data, ontologies, security policy and data owner information are input to this sub-module. Data will be converted to semantic format; ontologies will aid the process of conversion. Data annotation will be added afterward based on the security policy and data owner's information.

Output of *Collect* will serve as input in *Convert*. i.e. Data files and transformed ontologies Different types of convertors are taken as off-the-shelf product and are used to transform data files in the RDF format. At this point individual data is converted into the semantic format along with their security policies.

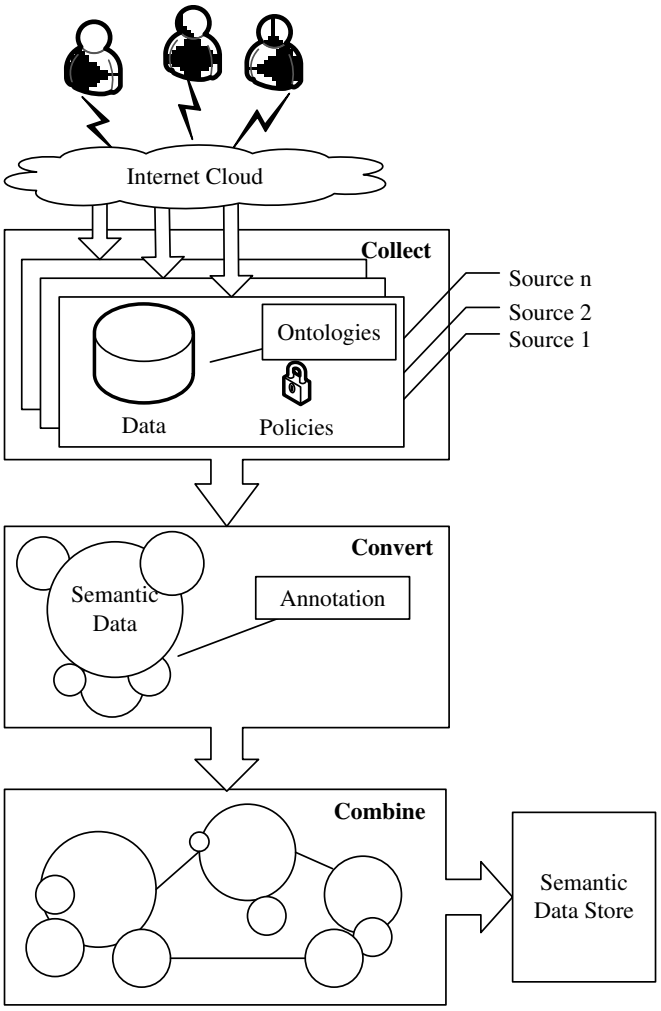

Fig. 1. Data integrator.

3) Combine

Combine is the third and last Sub-module in the Data Integrator. When data is converted successfully based on the security needs of the data users, it will be combined together. All the data will be transformed into an RDF graph.

Output of Convert will be served as input to combine. Annotated semantic data is taken as an input and combined; however, it retains the information about individual ownership of the data.

B. Data Store

Data is stored in the form of semantic data that uses RDF for representation. Update or retrieval of data will be made through web services based interface. In Fig. 2 representation of layers of data is shown, detail description of layers are mentioned in Table II.

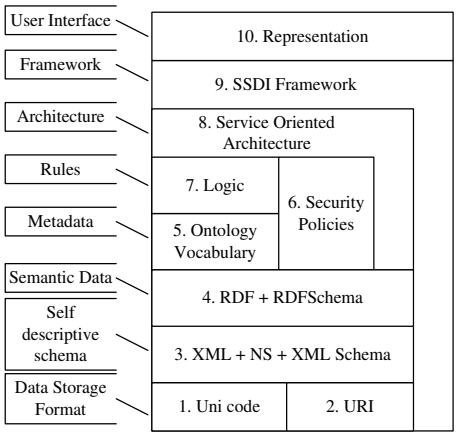

Fig. 2. Data layer.

SSDI have large portion of processing based on data. Thus Data is most important element for the concept of this framework. To understand data present in this framework it is divided logically into ten parts. These ten parts form eight layers

TABLE II: DATA LAYER

Layer	*Description*
Layer 1	UniCode is the first layer in SSDI frame work is of Uni code. Unicode is used as a standard for universal coding of text.
Layer 1	URI is parallel to the Unicode is layer of URI I.e. universal Resource Identifier. It identifies the any object uniquely over the internet.
Self Descriptive Schema	XML+NS+XML Schema is based on UNI and Unicode we have XML+ NS+ XML Schema. XML is designed to store and transport data, Namespace to avoid conflicts in data elements from different xml documents and schema to structure a XML document.
Semantic Data Schema	RDF+ RDF schema: Resource Description framework is a general purpose language designed to describe knowledge in the web based environment. Schema provides the structure for storage and processing purposes. This is the bases for Semantic data
Metadata	Ontology is description of knowledge, data about data, and specification of conceptualization. Ontology vocabulary provide aids to the RDF for processing, storing, connecting data
Rules + Data	Security policies for access control management are stored along with the data. Security policies overlap ontology vocabulary and logic layer, as it has data for security requirement as well as the logic for implementing it.
Rules	Logic are the rules about the Data, Data ownership, accessibility etc
Architecture	Service oriented Architecture: Service oriented Architecture provides services for accessing the data, and thus implementing any logic on it.
Framework	Secure semantic Data Integrator (SSDI) framework layer is the highest layer from data perspective, it combines all the below layer and aids the functionality by providing a solution to data interoperability and data security problems.
User Interface	Representation is user interface layer, it provides all type of users access to provide and retrieve data.

There are many stores available for rdf. Jena triple store will be used by SSDI [21].

The data to be used in this system can be categorized based on the access of the data, into two parts. Table III contain detail description of data types that are being used in SSDI.

TABLE III: DATA TYPES

Data Type	Description
Public Data	Public data that has no security constrain on it. Either available publically over the World Wide Web or the Data provider hasn't applied any security constrains on it and declared it as public data.
Private Data	Private Data, that has security constrains on it. This data and security policies are defined by data provider/ Domain Controller. Private data is further dynamically divided based on the domain it belongs to.

C. Access Control Management

Access Control Management is the backbone of SSDI. It controls the access to data and limits it to only authorized users. Users will request for data and after verification for that user's access to the data, result data set will be sent back to users [22]. For enforcement of security, this module works with the aid from Data Integrator Module.

Requirement for the privacy and access are handled in Data Integrator Module. As already explained Data Integrator module is responsible for gathering security details about data and applying necessary modifications on data. Security requirements are defined by Data Owner and it is received along with data. Data will be annotated as per security requirements of data. Once data is in its final shape, data will be accessible to users.

1) Users

SSDI have divided the users into different type based on their authorization to access different types of data and services. Table IV contain SSDI user roles along with their description.

TABLE IV: USER ROLES

User	Role Descriptions
System Administrator	System administrator is responsible for controlling domain. This includes domain creation and assignment of domain controller to a domain.
Data Provider/ Domain Controller	Owner of data is responsible for providing data to the SSDI, controls and create the policies on data, Uses data based on domain access/data policies. It also responsible for the authorization of Data users' assignment to their domain.
Data User	A member of some domain uses Data based on domain access/data polices.
Guest User	Uses data that has no restriction on its use i.e., public data

2) Data access cycle

There is complete data access cycle that is explained in this section. Complete cycle includes user for retrieving data results will be using Web services [23]. User interface is used as an access point to the data. User request to access data is sent to the "Set Query" and results are delivered through "Send Results". Fig. 3 shows the interaction of users to the system for retrieval of data. It also elaborated the path that will be followed by data, in order to fulfil users' data retrieval request. Main steps involved in the cycle along with their responsibility are as follows:

a) Set query

User will enter a valid SPARQL query using user Interface. This query along with unique identifies for the user will be transferred to SSDI by this web service. The unique identifier associated with every user will be used for the retrieval of user's domain information.

b) Query rewriting

SSDI Query rewriting service will process the query; it will enforce the security requirements applied by the data owner. Constrains will be added in the query to enforce the privacy. In data annotation information about the domains for which the data is accessible is added to the data. Using the User's domain information constraints will be added to check if the data is allowed by its data owner to be accessible to user.

c) Get results

Once the query processing is complete as per privacy requirement, query will be executed on the integrated data available in the data store to retrieve the result data set. Results will be in the RDF Format.

d) Send results

The data received from the query execution will be sent to the user. For simplicity the data will be available to user in the RDF format; however functionality for converting RDF in other format can be added later.

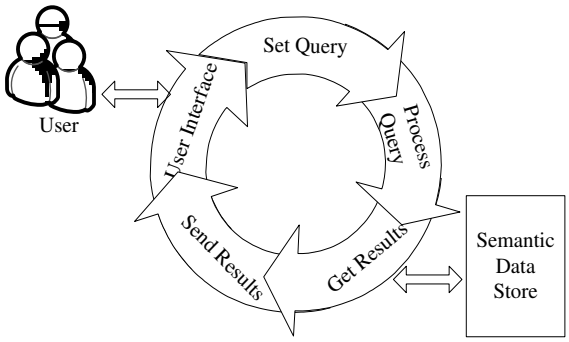

Fig. 3. Data access cycle.

D. Interface

Along with all the other modules Interfaces are also very important aspect of SSDI. Usability of the Application is as important as the other aspects. Interfaces in SSDI are not only the communication point between Human User and the System but also provide communication between different modules of the SSDI. For clarity SSDI interfaces are divided into two categories. First is User interface that provide the communication between the human users and the system. Second is SSDI Interface which includes the communication between user interface and SSDI other component and, communication in SSDI internal modules.

Following are the details for the architecture of Interfaces.

1) User interface

SSDI User Interface is designed to ensure the system is accessible from everywhere. User Interface is browser based and all the services are available through internet. This feature increases the system usability. Table V have detail description of User interface breakdown that is used in SSDI. It contains explanation of different screens and their purpose in the system.

TABLE V: USER INTERFACE STRUCTURE

Page Name	Description
Welcome	This is the first page and entry point of the system. This page will have a welcome message and option to either login or signup for the system. For login system Users have to provide their unique ID, Password and the account type. Username and the account type should be unique.
Sign Up	For new users a form based page will take the necessary input. There are two options for the type: data owner and the Data user. Data Owner or the Domain Owner will have the right to allow Data User account. Data Owner account will be authorized by Administrator only.
Data User Home	Home page of data user will have information about his domain and domain's owner info. This page is for submission of SPARQL query. "Set Query" Service will be called as an effect of this submitting SPARQL query. Results for the query are also provided on this page.
Data Owner Home	Data owner home page will have functionality for manage data users accounts and Data management option. This page shows the list of already uploaded data and their associated security requirements.
Data Add & Security requirement	This page is to Upload new data in the system. File name path and its associated type is compulsory to be given. Ontology definition can be defined by user or alternatively global ontology will be used in the system. Security requirement will be added in the form of domain list that are authorized to access the data. Unique id of the data owner will be associated with the data. This ID will be used to update the data in future.
Data Update & security Requirement	Data owner can select the already update data and either update data or associated security requirements. Data can also be deleted.
Administrator	Administrator is responsible for activation of data owner accounts and for configurations of the system.
Configurations	Setup for global ontology in the setup
Account approval	Requests for the approval of account creation from data owner are approved or denied.

2) SSDI interface

SSDI inter-module Communication is based on the services. Table VI elaborated description of services, along with input and output parameters, which are being used in different interfaces for inter-module communication.

a) User addition

For addition of user account in the system this service is called. There are multiple types of Users in the system. This service is called when user requests for the account.

b) Data addition

This interface will be used by Data owner only. Data will converted by calling conversion services. This will activate the "collect" part of the data integrator.

c) Update data

This interface will be used by Data owner only. This interface includes Update or Deletion of Data.

d) Retrieve data

This interface is used to retrieve or search the data.

TABLE VI: INTERFACE STRUCTURE

Service	Input	Output
User Addition	User information	Creation of a valid User Account
Data Addition	Data & security Requirement	Data and security requirement stored in the system
Update Data	Updated data OR security requirement	Updated Data or Updated security requirement associated.
Retrieve Data (Search Data)	SPARQL Query	RDF Result Set

V. CONCLUSION AND DISCUSSION

Main focus of this work is the design of our proposed framework, Secure Semantic Data Integrator (SSDI). This system utilizes full potentials of semantic data to make secure data integration possible. SSDI enables the user to share the data originating from heterogeneous sources, with security setting of their own. This feature maintains the data ownership status which makes data owner and data provider classification distinct. We have discussed the architecture of the SSDI that elaborates the high level structure of the system. It future elaborates the detailed design of the system along with the core implementation techniques.

VI. FUTURE WORK

The approach presented here is in its preliminary phases, major improvements can be done in refining the techniques and concept. Currently the focus of this approach is to provide a framework that can not only integrate data but also provide a secure environment. However improving the framework to incorporate the advance ontology processing would increase the usability and applicability of the system.

REFERENCES

[1] M. Lenzerini, "Data integration: a theoretical perspective. In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems," in *Proc. PODS '02*, pp. 233–246, New York, NY, USA, 2002.

[2] Dataintegration. [Online]. Available: http://www.dataintegration.info/data-integration

[3] J.-H. Zeng, "Research and Practical Experiences in the Use of Multiple Data Sources for Enterprise Level Planning and Decision Making: A Literature Review," Spring 1999.

[4] Semagix. [Online]. Available: http://www.semagix.com/what-is-semantic-data.htm

[5] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug *et al*., "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat Biotechnol*., vol. 25, pp. 1251–1255, 2007.

[6] C. Clifton, "Murat Kantarcio AnHai Doan, Gunther Schadow, Jaideep Vaidya, Ahmed Elmagarmid, and Dan Suciu. 2004. Privacy-preserving data integration and sharing," in *Proc. the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery (DMKD '04)*, ACM, New York, NY, USA.

[7] A, Rezgui, M, Ouzzani, A, Bouguettaya, and B, Medjahed, "Preserving privacy in web services," in *Proc. WIDM '02 Proceedings of the 4th international workshop on Web information and data management*, pp. 56–62, ACM New York, NY, USA.

[8] M. Mrissa, S.-E. Tbahriti, and H.-L. Truong, "Privacy Model and Annotation for DaaS," in *Proc. 2010 IEEE 8th European Conference on Web Services (ECOWS)*, p. 3, 10, 1-3 Dec. 2010.

[9] H. Alani, W. Hall *et al*., "Building a Pragmatic Semantic Web," *Intelligent Systems*, IEEE, vol. 23, no. 3, p. 61, 68, May-June 2008.

[10] Jena. [Online]. Available: http://jena.apache.org/

[11] M. Barhamgi *et al*., "Privacy-Preserving Data Mashup," in *Proc. 2011 IEEE International Conference on Advanced Information Networking and Applications (AINA)*, p. 467, 474, 22-25 March 2011.

[12] *Private-Ive: Framework for privacy preserving Data Integration.*

[13] S.-H. Yang and J.-Z. Wu, "Mapping Relational Databases into Ontologies through a Graph-based Formal Model," in *Proc. 2010 Sixth International Conference on Semantics Knowledge and Grid (SKG)*, p. 219, 226, 1-3 Nov. 2010.

[14] A. M. Iqbal, A. Moh'd, and Z. Khan, "A semi-automated approach to transforming database schemas into ontology language," in *Proc.* 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE), p.000930, 000933.

[15] D.-J. Dou *et al*., "Integrating Databases into the Semantic Web through an Ontology-Based Framework," in *Proc. 22nd International Conference on Data Engineering Workshops*, p. 54, 2006.

[16] RDF123: from Spreadsheet to RDF.

[17] M. Farouk and M. Ishizuka, in *Proc. AIKED'2012 Proceedings of the 11th WSEAS international conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, pp. 195-200.

[18] C. Bizer, "D2RQ - treating non-RDF databases as virtual RDF graphs," in *Proc. the 3rd International Semantic Web Conference (ISWC2004)*, pp. 100-109.

[19] C. Bizer, "D2RQ - treating non-RDF databases as virtual RDF graphs," in *Proc. the 3rd International Semantic Web Conference (ISWC2004)*, pp. 345-400.

[20] R. Ghawi and N. Cullot, "Database-to-Ontology Mapping Generation for Semantic Interoperability," *In Third International Workshop on Database Interoperability (InterDB 2007)*, 2007.

[21] S. Gerbracht, "Possibilities to encrypt an RDF-Graph Published in:Information and Communication Technologies: From Theory to Applications," *ICTTA*, 2008.

[22] L. Kagal, T. Finin, M. Paolucci *et al*., "Authorization and privacy for semantic Web services," *Intelligent Systems*, IEEE, vol. 19, no. 4, pp. 50,56, Jul.-Aug. 2004

[23] S. Ali, S. Khusro, and A. Rauf, "A cryptography-based approach to web mashup security," in *Proc. 2011 International Conference on Computer Networks and Information Technology (ICCNIT)*, p. 53, 57, 11-13 July 2011.

Shoohira Aftab was born in 1988 in Pakistan. She received her B.E degree in software engineering from National University of Science and Technology, Islamabad Pakistan in 2010. She is currently a scholar of masters' degree program in computer science from National University of Science and Technology, Islamabad, Pakistan. Her current research interest includes semantic data, data integration technologies, cloud computing, information security in modern data integration and semantic data technologies, and access control management systems.

Hammad Afzal is working at National University of Sciences and Technology, Islamabad, Pakistan as an assistant professor since May, 2010. He has previously done his doctoral studies, working under the supervision of Dr. Goran Nenadic and Dr. Robert Stevens, based in School of Computer Science, University of Manchester UK. Dr. Hammad Afzal has also worked as a research fellow at Digital Enterprize Research Institute (DERI), National University of Ireland, where he was working with Dr Paul Buitelaar, as part of the Unit of Natural Language Processing. He is currently involved in a teaching and a number of UG and PG projects, mainly in the fields of natural language processing, semantic web, linked data and smart phone applications (UG).

Amna Khalid from Pakistan has received her B.E degree in software engineering from National University of Science and Technology, Islamabad, Pakistan in 2010. She is currently a scholar of masters' degree program in computer science from National University of Science and Technology, Islamabad, Pakistan. Her current research interest includes data analytics, semantic data integration technologies, cloud computing, information security, and computer science education.