A Model for Information Integration Using Service Oriented Architecture

C. Punitha Devi
Department of Computer Science and Engineering, Pondicherry University, Puducherry, INDIA
punitha_c@yahoo.com

V. Prasanna Venkatesan
Department of Banking Technology, Pondicherry University, Puducherry, INDIA
prasanna_v@yahoo.com

S. Diwahar
Department of Computer Science and Engineering, Pondicherry University, Puducherry, INDIA
diwahar.sivaraman@gmail.com

G. Shanmugasundaram
Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, INDIA
sundar_gss2004@yahoo.co.in

Abstract—Business agility remains to be the keyword that drives the business into different directions and enabling a 360 degree shift in the business process. To achieve agility the organization should work on real time information and data. The need to have instant access to information appears to be ever shine requirement of all organizations or enterprise. Access to information does not come directly with a single query but a complex process termed Information integration. Information integration has been in existence for the past two decades and has been progressive up to now. The challenges and issues keep on persisting as information integration problem evolves by itself. This paper addresses the issues in the approaches, techniques and models pertaining to information integration and identifies the problem for a need for a complete model. As SOA is the architectural style that is changing the business patterns today, this paper proposes a service oriented model for information integration. The model mainly focuses on giving a complete structure for information integration that is adaptable to any environment and open in nature. Here information is converted into service and then the information services are integrated through service oriented integration to provide the integrated information also as service.

Index Terms—Information Integration, Service Oriented Architecture, IIaaS, Service Oriented Integration, Information Integration characterization.

I. INTRODUCTION

The very nature of information has changed in the past decade. It has become increasingly digital, networked, and overwhelming. Information also flows free of the containers that were previously controlled as the gatekeepers. Information is needed in the form that directly reflects to the changing growth of business. This change in information has lead for instance access to data or current data, to standardize the business process, decoupling of information /data from applications, enable information integration to provide real time analysis and insight and able to connect partners in real time [1]. This type of information cannot be obtained from a single data source or with a single query to multiple data sources, but with efficient information integration.

As per this need new approaches, techniques, tools and models that would help in obtaining the requested information has to be formed or the present ones have to be changed to adapt to the changed requirements. It's essential to see whether the existing methodologies could be changed rather than to invest efforts and cost on new methodology [2].

This process of information integration evolved from database technologies. Data initially resided in single databases; it grew to multi databases, then to distributed databases and further to federated databases [3]. Complexities also grew along with the data bases, hence data integration was considered to be a major process where many challenges and approaches started to grow. Data integration started to gain enormous attention as data spread across various applications and databases which adhered to heterogeneous types [4].

As technology improved application development leading to evolvement of software products became the trend. Data as well as information was accessed through these applications hence application integration took its strand in the name of EAI. EAI was widely used by all type of business and the access to data sources was inbuilt in the systems [5]. Integration became highly complex due to interoperability and materialization of

data. Data sources became too many and data started to grow in volume and type. Query processing was considered as the core process in such a distributed environment. Hence the concept of Enterprise Information Integration (EII) came into existence [6].

EII is also known as query mediator that translates users' queries into queries on the data sources and integrates the results of those queries so that it appears to have come from single database. EII was only able to retrieve and read data but updates if any were to be done through EAI only [7]. Data federation resulted in data virtualization. Organizations such as Oracle, IBM, Microsoft etc., built products for data virtualization through federation [8].

These techniques concentrated on an enterprise wide solution for information integration but due to the upcoming of internet, globalization and expansion a different look towards integration of information across heterogeneous information systems is given. Service oriented architecture has been identified as an architectural style suitable for integration inside enterprise and across organizations [9][10]. The ideal integration technique would need to solve issues [11]:

- Provide data dynamically – on demand
- Automatically integrate data with respect the domain knowledge
- Incorporation of standards to information

These issues persist in case of all technologies explained above. They work on static data, and data however said to be provided in real time as in EII, but actually not true in reality. Instead it provides near real time data. With these as requirements there exist no integration technique and addressing these requirements cannot be achieved in a single approach. But it involves complex set of tasks, adaptability with an environment, flexible enough to adapt and to provide right information to the right people at the right time. These requirements lead to a need for a model for information integration which has been proposed in this paper, a service oriented information integration model.

This paper brings about the challenges and issues in case of existing traditional approaches to information integration in the related work section followed by the description of the proposed model for information integration with service orientation and its principles. The last section gives the qualitative comparison of these three technical approaches.

II. RELATED WORK

Information integration is defined as combining information from various heterogeneous sources that differ conceptually, contextually and typographical representations i.e., structured or semi structured or unstructured [12][13]. In general, information integration needs [14] to

- Support all information types, structured, unstructured and semi-structured
- Provide context, i.e., where does the information fit in the schema / taxonomy of the receiving repository / application, and what are the relevant behavioral constraints
- Perform seamless integration

These form the vital concept and branches out to various methods and issues. This concept has been live for the past two decades and research has been carried out in various angles pertaining to information integration. This section provides a walkthrough with explanations of the existing approaches, tools and models as milestones to identify the research gap and a destination to formulate the problem statement.

A. Approaches

The actual handling of the data in pursuit of ideal information integration is explained in the approaches. These approaches form the basic design aspect of the existing information integration models and frameworks. The approaches originated from data integration strategy itself it went on to evolve as per the changing requirements of business. It has been found that most of researchers [15][16][17][18][19][20] have considered the general approaches to be

- Materialization
- Federation

Data materialization consolidates the data to form a common model and makes querying and data access easy for the requested data. This is also named as the data warehousing approach were the data needed are extracted transformed and loaded to get the require information. This approach has been widely adopted in EAI techniques. Data warehouse was an instant hit among many enterprises but started to lose its shine as data sources increased [21][22].

In contrast data federation approach gained attention where the data are retrieved directly from the data sources by extensive querying over the views of the multiple data sources. Mapping of data sources were achieved through Global-as-view and Local-as-view methods. Global-as-view forms a global schema with reference to the underlying sources and data are viewed through this joint view. Local- as-view structures a global schema irrespective of the data sources and each data source is related by a local view over the global schema. [17][19]

Apart from these notable approaches the other approaches that could be named are

- Ontology based integration [16] [23]
- Navigation based integration [16] [23]

Ontology based approach has its origin as semantics. This approach focuses completely on the problem of

semantic integration which is based on the multiple ontologies. Semantic web and web services is a new lead in this type of integration. Data were realized through web services. The major problem of this approach is the need for a valid methodology to adopt the approach. Navigation based approach works completely on the links provided through the web pages. As the number of web sources increases the page links increases and a model has to be applied to store the interconnection, paths, along with input parameters. This approach was significant in case of information integration among deep web sources. This approach lost its potential with rapid changes and increase in web sources. It is evident that these approaches to information integration still keep maturing and has open issue and challenges.

B. Models and Frameworks

Information integration consists of a complex process where approaches form only a part of it. The models and frameworks help to structure out the various tasks of the process and help in turn to adopt the appropriate approach. Some of the prominent models and frameworks of information integration are specified in concise below.

The literature study has been done on several research contributions based on several models and frameworks that have been implemented for information integration. To consolidate the review findings the various functionalities predominantly addressed by the models and frameworks are:

- Query plan and processing [24][25][26][27][28]
- Query optimization [20][24][25][26][29]
- Integration approach LAV or GAV [20][25][30]
- Data integration [15][20][26][28][29][31][32]
- Semantics and Ontology [20][33][34][35]
- Web information sources [24][25]
- User interface [12]

It is evident that these functionalities purely pertain to the mapping layer where it concentrates on retrieval of data that corresponds to the needed information. It could also be noted that each of these functionalities was not completely considered by a framework each of them had their own limitations. For example Information Manifold framework [29] allows only declarative queries to be addressed on the predefined set of sources. Any improvements on the query processing could not be done. The models concentrated much on the integration aspects and in handling heterogeneity of data sources.

Likewise each model or frameworks have exhibited their functionalities through various technologies like

- Database management [20][29] [32]
- Artificial intelligence [29]
- Machine learning [25]
- Software agents and Multi agents [36]
- Computational logic [3]
- Wrappers and Mediators [25][27] [28] [37]

- Web Services and Service Oriented Architectures [33][35][37][38][39][40]

The technologies pertain to the need of the system and as the complexities of information integration increased due to different types and high volume of data, highly adaptable and flexible technologies were adopted. To name a few, artificial intelligence and data base technologies were used to build the three components in the information manifold [29] framework. When more intelligence and self reasoning were the need of a system agents are adopted. Multiple agents such as user agent, broker agent, communicator agent, ontology agent, and resource agent were employed for various tasks of Infoslueth framework [41]. Multi domain information extraction and integration of spatial and temporal data were first used in Infomosaic framework [26]. Most of the models and frameworks have used wrappers and mediators [25][27][28][37].

The need for dynamic integration or on the fly integration [17][24] arises due to business agility. Real time data and information and ability for data sources to be discovered and data sources had to be added as integration proceeded. Reuse of data, information or functionality exposed by the application or process was also in need. Service oriented architecture and webs services were considered a solution for the integration challenges such business agility, reuse, standardization and visibility. Information integration systems built on SOA or Web services have centered on

- Semantic based frameworks [35] [38][39] [42]
- Workflow based models [43]
- Application specific models and frameworks [21][44][45][46][47]
- Information as service [48][49]

Frameworks for eg., semantic based frameworks in specific deal completely on the semantic layer and how ontologies and semantics could be improved to enhance information integration. Likewise each models and frameworks are biased either towards one specific layers or applications. Information as a service is one concept originated through IBM's recent Information on demand Strategy [50]. It was easily adopted by many organizations like Informatica, Composite Softwares, Denodo Technologies and lot more as a step towards achieving the complete benefits of SOA at the data and information level [51].

C. Research Gap

Information integration is a vibrant field powered not only by engineering innovation but also by evolution of the problem itself [23]. In its progress the issues and challenges grew and different approaches, models frameworks also evolved, which has been brought out in the related work section. The review has clearly brought that information integration has not addressed all its requirements in complete [17][38][52][53][54][55]. The

visible aspects that has been identified as research issues which may lead to addressing all the requirements of information integration are

- Firstly the frameworks and models that are addressed are not open in nature, they do not provide space for existing systems, if do not provide definite solution to suit any type of application or system.
- Secondly they do not adopt themselves to the environment for which information integration model or framework is formed or to any other environment where the model could be used. They have been talked about as a separate entity. But in real, information integration is one field that has to happen simultaneously as per the need of the system. This is not considered as such presently.
- Third, to specify, there exists no proper flow pertaining to the information integration systems. It lacks appropriate chaining of process and tools.

To consolidate information integration should be ubiquitous to provide right information to the right people at the right time and act according to any change in the business scenario.

III. PROPOSED WORK

A. Motivation

Information integration is moving towards a goal to achieve a new pinnacle in accessing heterogeneous information residing in geographically distributed sources. The ever changing business climate and requirements are shaping a new breed of information integration approach. The scope of data as well information management are expanding significantly beyond the traditional Data warehousing, Business Intelligence, content market, and Data virtualization, instead, it has to create value for the users by integrating information [56][57]. These business requirements and the research gap that has been identified as a result of an extensive literature survey have been considered as the influencing factors for the research problem.

One architectural style that proves to meet the changing business requirements is SOA. Where the business functions are broken down into various tasks and tasks are made to access as services [58]. Here the requirement to unlock data from its applications and completely throw away information islands could be solved through adoption of SOA and its principles while creating information services. Information to be made a reusable resource that allows usage along with constrains and standards with respect to the stakeholders.

B. Proposed Problem

A comprehensive and complete characterization of information integration has been stated by [59]. It gives a complete definition and states the requirement for such information integration to be achieved. SOA deployment has reached maturity in being adopted at the process level but some researchers say that SOA benefits could be completely realized when adopted at the data and information level [48][49].

The problem statement to put in a line is

"To build a layered model for Information integration adopting SOA and deliver integrated information as a service".

C. Proposed Solution

Effective decisions and planning depends highly on information availability. This information has to be consistent and could be achieved through loose coupling of data and applications, which has been brought out by Service Oriented Architecture. SOA represents a new way of thinking about everything in a company's IT structure, including how one thinks about data. It begins with the goal of achieving consistency between data sources. In order to achieve data consistency, it begins by separating data from its tight dependency on the business applications that created it and update it. One critical step for making this happen is to ensure that each component of data can be used independently from its current implementation. With service oriented architecture, information should be thought of as reusable resources.

The research gap that has been identified precisely indicates the need for a model that could very well suit all types of systems irrespective of domain, any business scenario, and adapt to any suitable environment and provide the flow of execution of the tools if any used. Hence a model that is open, scalable and flexible, and completely structured is the basic qualities of the proposed paper.

A layered approach has been taken into account, as it appropriately helps to support strong separation of concerns, abstraction, loose coupling and that in turn supports flexibility, scalability and modularity. This structure consolidates the dynamic interoperability between components because components can be replaced and upgraded without affecting other layers [44].

Here SOII builds a service oriented system where information is discovered, integrated, modeled and delivered as service. These information services are integrated to deliver the requested information as a service. SOA benefits are realized completely by including it to the data level [48]. The model is shown in Fig. 1.
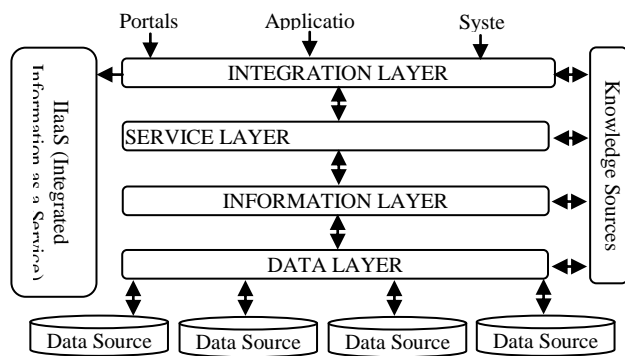
Fig. 1. Layer Diagram

This model is initiated or activated through any system, or portals or applications or machines or devices that send a request for some information that could not be obtained from a single data source.

The Integration Layer

This layer acts as an application or system interface layer for this framework, in addition to the integration functionality exposed.

Request:

- The request is categorized into one or more information request with the help of domain knowledge.
- Information about single or multiple information needed are analyzed and maintained.
- The request is passed as information service request to the service layer.

Response:

- The information service request is obtained as information service after complex processing.
- Depending upon the service type and strategy the services are integrated.
- The integrated service is now the integrated information as a service.
- Depending on the initiators request a relevant response is built and exposed as service.

The Service Layer

This layer handles the functionalities of the services that are created. The request and response are handled here with respect to the initiator of the model. This layer acts between integration layer and information layer.

Request:

- The initiators request is treated as information service request is processed with the help of domain ontology.
- Request is analyzed for it to be single or multiple services and split into various requests for services.
- This information services request is passed down for the need of associated information as information request.

Response:

- The services layer accepts the requested information as a response from the information layer
- The candidate service is converted into a service with respect to the template selected and stored in the service repository and registered into the local registry as well.
- The service created is passed to the integration layer

The Information Layer

This layer communicates with the service layer to get the information request, and with data layer to accept the data response from.

Request:

- The request that is handled by this layer is the Information request.
- The information request is processed for the identification of relevant data with the help of semantics and ontology.

Response:

- The data requested for is received as response from the data layer.
- The data is converted to information and modeled with business rules.
- The information which is to be made available as service, is sent to the service modeler.
- The modeled information is brought out as candidate service with the help of service rules.

The Data Layer

This layer as the name implies deals directly with the distributed heterogeneous data sources in order to retrieve of right data. This layer communicates with the integration layer and the data sources.

Request:

- The data request is handled here for obtaining the relevant data
- The data request is processed based on the domain ontology and continued to retrieve the resultant data set.
- The queries are formulated for the resultant data set.
- Simultaneous retrieval of data are facilitated

Response:

- The data are obtained through data federation approach towards the data sources
- The data are aligned with the request and passed to the information layer.

Data and Knowledge Sources

The data sources are assumed to be of heterogeneous data that are distributed geographically. The data sources are not tied to the application or organization, the data silos are just operated upon without altering their location and privacy. The information transparency is the major

advantage here, the access are recorded as to have accessed the data.

Knowledge sources pertain to the domain ontology along with the business rules and policies with respect to the data sources. These form the major aspect as the request and response of this model depends on this interpretation. The domain ontology is assumed to be probably a large and a detailed one as it is to give minute classifications and relations in order to retrieve the appropriate data from the sources.

IV. DISCUSSION

A model depicts the design of the concept under which it's laid on. This model follows the SOA style where the various information needed are converted as services and integration happen at the service level. The capabilities of this model could be well explained when compared with the other information integration strategies such as EAI and EII. This model has been compared for the qualities that are desirable for efficient information integration. In order to bring out the uniqueness and the benefits of Service oriented information integration the comparison is done on the basis of certain features grouped under 3 essential aspects of information integration. They are identified to be Business, Technology and Quality aspects. The features chosen are in accordance to the challenges of information integration point of view [48][53]. The features are briefed here

Business Aspects

The essential features that could be relevant to business requirements are explained as business aspect. Change is the word that rules the business world along with a need to access current data and information. Hence three most required aspects of current business scenario have been projected to show the model's benefits.

Business Agility: This feature indicates the dynamism of the system i.e., how far it adapts to changes in real time. Current business world keeps changing, from requirements to governance. The system which is not affected by changes is able to meet business agility.

Inter Enterprise: This reveals the need to use distributed data both within an enterprise and out of enterprise such as internet, cloud data sources

Real Time Information: This indicates the need for the current information in processing and delivering the required information which would lead to better business insights.

Technical Aspects

Technology refers to the executable element of the business proposals or requirements. Here for information integration the technical aspects intends in bringing out a completely automated and able to incorporate legacy sources. With respect to these 3 most significant basic technical aspects are considered to exhibit the models technical benefits.

Leveraging Existing application: Business is in need to utilize the time and amount spent on existing applications. The benefits of existing applications could be extracted and made use of.

Fully Automated: This principal normally refers to the extent of work done my machines / computers rather than human. If the needed information is processed completely without human intervention then we say that this feature is fully automated.

Read and Write: This indicates the ability of the data sources and system to provide rights to read and write.

Quality Aspects

Since the proposed model is service oriented in nature four most predominant non functional aspects which also falls in line with the information integration requirements have taken into consideration to bring out the uniqueness and efficiency in the model.

Loosely Coupled: This is term that is been the real need of business today. Application consumers do not need to have knowledge beforehand about the system functionality other than where to find it. Application functionality and the programs that invoke them can be changed independently of each other, instead of requiring a redesign to the components.

Reusability: The feature that is the most essential in case of any system is reusability. It states the need for the information and data access logic to be reused.

Correctness, Completeness and Consistency: This feature indirectly indicates the reliability of the system. This refers to the performance of the system on execution.

Information Hiding: This refers to the level of abstraction and transparency exhibited by the data sources. As business operates with all type of data this feature helps in providing a controlled and authenticated view of the information.

The traditional EAI and EII are integration strategies that have been dominating the enterprises for a long period of time, even though they had their own limitations. Enterprise information integration (EII) is the ability to support a unified view of data and information for an entire organization.[6][7] Enterprise application integration is an integration framework composed of a collection of technologies and services which form a middleware to enable integration of systems and applications across the enterprise.[5][7] With reference to the definition and the related tools and structure cited, EAI and EII could be well understood as to be compared with the proposed work.

Table 1 shows the level of aspects in the respective category.

Table 1. Qualitative Comparison

Attributes	Enterprise Application Integration	Enterprise Information Integration	Proposed Service Oriented Information Integration
Business Aspects			
Business Agility	Low	Medium	High
Inter Enterprise	Low	Low	High
Real time Information	Low	Medium	High
Technology Aspects			
Leveraging Existing Systems	Low	Medium	High
Automation	Medium	Low	High
Read and Write	Medium	Low	High
Quality Aspects			
Loose Coupling	Low	Medium	High
Correctness, Completeness, Consistency	Medium	Medium	High
Reusability	Low	Low	High
Information Hiding	Low	Medium	High

The table values of specifying high, low, or medium is stated with respect to its functionality, and explained below.

Business Aspects

Business Agility: A system or model is said to be completely agile when it remains open and withstands any changes in the business, the proposed model is of such a kind and hence graded high. A system can also adhere to changes posed by its own business environment for example within an enterprise, and EII strategy is graded medium as it falls within the enterprise level. Likewise it could be flexible enough to changes at the application or technology level and EAI systems or models withstands only to this level of changes hence graded low against business agility.

Inter Enterprise: When it is an open system, and also allows incorporating standards in data or information retrieval within and outside the enterprise is stated high as provided by the proposed system. When the access to distributed data is limited to an enterprise as in EII it could be graded medium and when the access is done with manual interventions as done in case of EAI, where access is restricted to applications.

Real Time Information: Information that seems to be current and it goes out with the need to get the right information, at the right time. The proposed system has been aimed at it and as services are used it works on current data, so it is graded high. In some cases federated access to distribute data is through mediated schema and however provides real time data with the existing data sources and does not allow data sources to be dynamically discoverable and queryable. EII which

provides this variety of data is graded medium. Whereas data warehousing approach adopted in EAI uses static and historical data at an interval [EII success and challenges]

Technology Aspects

Leveraging Existing Systems: Existing applications and data sources need not be converted or made changes but can be used as such. The only requirement in case of SOA an interface has to be created and implementation is differed from its interfaces. Hence usage of legacy systems or data becomes possible so graded high. In case of EII the data sources are mediated and the changes are made to query process so as to use legacy data. Since this does not avail the usage completely, it could be graded medium. As in this case of EAI it cannot be availed as such or made available with modifications, hence low.

Automation: This feature as the meaning conveys, if a system works completely for the request without human intervening, then it graded high or said fully automated, The systems in which the systems produces the results and manually they are integrated i.e executed partially then it could be said that it is medium and when more human intervention is included it is graded low, as in EII. *Read and Write:* The basic nature of any data is to be consistent. Most of the systems consider it to be complex problem and leave it simple by allowing it to be only read on access, such systems are graded low as in EII. Like EAI both read and write are enabled on access but does not reflect on all places they have been used, hence low consistency and graded medium. One system that allows access to read and write options to the data sources as in the proposed model is graded high.

Quality aspects

The proposed model has been intended to provide a complete model addressing all the requirements of information integration. The basic performance factors are looked for in the model in relation to this discussion section.

Loose Coupling: A system that provides independence to its components and decouples application logic from the implementation of Information services thereby future changes would impact the implementation of information services and not to the data models. In EII loose coupling refers to decoupling of application logic and data models, so graded medium. EAI is tightly coupled and graded low.

Reusability: The ability of the model to provide information as reusable resources proves to be graded high in reusability as in the proposed system which achieves it through information services. EII and EAI provide low reusability as each information retrieved is accessed for the specific purpose and cannot be reused for another such scenario.

Correctness, Completeness, Consistency: The three c's provide a basic quality of the system. The information retrieved is exact and is directly used by the model, the information obtained with respect business rules and standards and if changes are reflected throughout the data

sources then the model could be graded high as in proposed model. Both EII and EAI lack in specifying all the three aspects, either they satisfy correctness or consistency, but not completeness so they are graded medium.

Information Hiding: This aspect refers to the level of abstraction and graded high in case of proposed system as it provides information hiding at service level. In EII the views form the focal point, and provide limited information hiding. But in EAI the complete database are visible and graded low.

IV. CONCLUSION

Information integration is realized as ever challenging field due to enormous growth of information in volume and type. This paper has taken it as a focal point and has given an overview of current state of art which has directed to identify the research gap. The research gap addresses the three aspects openness, flexibility and completeness as the lag in case of existing models and real need of them in a model which is taken primarily into consideration. The proposed problem is stated based on the aspects identified and SOA. SOA is the building block that could unveil these problems by delivering the relevant information as a service. A model that follows a layered service oriented structure has been proposed, and is also compared with the other information integration techniques and is found to be effective as the framework adheres to SOA principles. The integrated information is delivered as a service. In order to bring out the real benefits of the framework the respective number of components has to be analyzed and designed further.

REFERENCES

[1] Chuck Ballard, Fabio Hasegawa, Gord Owens, Soren Ravn, Pedersen, Klaus Subtil, *Moving Forward with the On Demand Real-time Enterprise* , SG24-7104-00, IBM Redbook, September 2006.

[2] Hal Records, Nancy M. Records, Robert Behling, "Managing Information Integration In Today's Business", *Issues in Information Systems,* Volume VII, No. 2, 2006, pp 359-364.

[3] Yannis Dimopoulos and Antonis Kakas, "Information integration and computational logic", *Special Issue on the Future Technological Roadmap of Compulog-net*, (URL: http://www.compulog.org/net/Forum/Supportdocs.html.)

[4] A.P Sheth, J. A. Layon, "Federated Database system for managing distributed, heterogeneous and autonomous databases", *ACM Computing Surveys*, Vol 22,no.3,1990, pp 153-236.

[5] David S. Linthicum, *Enterprise Application Integration,* Addison Wesley, First Edition November 05, 1999 ISBN: 0-201-61583-5.

[6] *What Is Enterprise Information Integration (Eii)?*, The Gilbane Report, vol. 12, No.6, 2004.

[7] Alon Halevy, Anand Rajaraman, Joann Ordille, "Data Integration in Teenage Years", *VLDB `06,* ACM 1 59593 385 9/06/09.

[8] Noel Yuhanna and Mike Gilpin, *The Forrester Wave™: Data Virtualization, Q1 2012*, Jan 2012.

[9] Daniel Liebhart, Guido Schmutz, Peter Welkenbach, *Service-Oriented Architecture—An Integration Blueprint,* Packt Publishing enterprise.

[10] *"A Revolution in Agility: Business Integration Through Service-Oriented Architecture"*, Oracle White Paper, 2008.

[11] Corinne Baragoin, Joachim Dirker, Carolyn Elkins, Ian Harvey, Fred Lo, *Getting Started on Integrating Your Information*, IBM Redbook, Feb 2003.

[12] N. W. Paton, C. A. Goble, S. Bechhofer, " Knowledge based information integration systems", *Elsevier Information and Software Technology, Vol* 42 (2000) pp. 299–312.

[13] M. A. Roth, D. C. Wolfson,J. C. Kleewein,C. J. Nelin, Information integration:A new generation of information technology, IBM Systems Journal, Vol 41, No 4, 2002 0018-8670 pp 563-577.

[14] A. D. Jhingran, N. Mattosm,H. Pirahesh, "Information integration: A research agenda", IBM systems Journal, Vol 41, no 4, 2002.

[15] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, Riccardo Rosati, " Information Integration:Conceptual Modeling and Reasoning Support", 3rd IFCIS, International Conference on Cooperative Information Systems, Pgs 280-289, Aug 1998.

[16] Aditya Telang and Sharma Chakravarthy and Yan Huang," Information Integration Across Heterogeneous Sources: Where Do We Stand and How to Proceed?", International Conference on Management of Data COMAD 2008, Mumbai, India, CSI, 2008.

[17] Mohammad Golam Saklayen, "Autonomous Information Integration: A Theoretical Perspective", Available: http://www.cs.wayne.edu/~jamil/Dr.JamilCD/my-pub-papers/PDFFILES/jamil-saklayen-2005.pdf.

[18] Mukesh Mohania and Manish Bhide,"New trends in information integration", IBM research Lab, 2007.

[19] Feng beiming, Ma Manfu, Gou Heping, " An Architecture to integrate distributed information integration, ACM, Eighth International Conference Grid and Cooperative Computing, 45 – 49, 2009.

[20] Jaime A Reinoso Castillo, Adrian Silvescu, Doina Caragea, Jyotishman Pathak, Vasant G Honavar, " Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Information Sources - A Federated Ontology-Driven Query-Centric Approach", *IEEE* , International Conference on Information Reuse and Integration, 2003.

[21] Kai Yang, Robert Steele, "A Semantic Integration Solution for Online Accommodation Information Integration", 978-1-4244-8756-1*, IEEE*, 2011.

[22] Alon Y. Halevy, Naveen Ashishy, Dina Bittonz, Michael Careyx, Denise Draper, Jeff Pollockk, Arnon Rosenthal, Vishal Sikkayy , "Enterprise Information Integration: Successes, Challenges and Controversies", *ACM SIGMOD*, June 2005.

[23] Philip A. Bernstein, Laura Haas, "Information Integration in the Enterprise", *Communications of the ACM*, Vol 51 No. 9, September 2008.

[24] K. C.-C. Chang, B. He, Z. Zhang "Toward Large Scale Integration: Building a MetaQuerier over Databases on the Web". In Proceedings of the Second Conference on Innovative Data Systems Research (CIDR 2005), California, January 2005.

[25] Craig a. Knoblock, Steven minton, Jose luis ambite, Naveen ashish, Ion muslea, Andrew g. Philpot, Sheila tejada, "The aridane approach to web-based information integration", *International Journal of*

Cooperative Information Systems, World Scientific Publishing Company, 2001.

[26] Aditya Telang, Sharma Chakravarthy "Information Integration across heterogeneous domains: current scenario, challenges, the InfoMosaic Approach", *Technical Report CSE-2007*, University of Texas.

[27] Ali Kiani, Nematoallaah Shiri "A Genaralised Model for Mediator Based Information Integration", 11th International Database Engineering and Applications Symposium (IDEAS 2007), 0-7695-2947-X, 2007.

[28] Oliver M. Duschka, Michael R. Genesereth "Infomaster- An Information Integration Tool", International Workshop on Intelligent Integration, Friedurg, Germany, 1997.

[29] Alon Y. Levy, "The Information Manifold Approach to Data Integration", *IEEE Intelligent Systems*, 1998.

[30] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y.Sagiv, J. Ullman, V. Vassalos, J. Widom, "The TSIMMIS approach to mediation: data models and languages", *Journal of Intelligent Information Systems* Vol 8 No 2, (1997) pp.117–132.

[31] Nicola Leone, Thomas Eiter, Riccardo Rosati, Edyta Kalka, "The INFOMIX System for Advanced Integration of Incomplete and Inconsistent Data", *ACM,* SIGMOD '05, Maryland USA, 1-59593-060-4, 2005.

[32] Maurizio Lenzerini, "Data Integration: A Theoretical Perspective" *ACM SIGMOD,* Symposium on principles of database systems, 2002.

[33] Bostjan Grasic, Vili Podgorelec, "Automating Ontology Based Information Integration Using Service orientation", *Wseas Transactions On Computers*, Vol 9, No 6, 1109- 2750 547 June 2010.

[34] David f. Barrero, M aria d. R-moreno, "Information Integration In Searchy: An Ontology and Web Services Based Approach", *International Journal of Computer Science and Applications*, Vol. 7 No. 2, pp. 14 - 29, 2010.

[35] Wang Zenglei, Zhang Kaixing, "An Enterprise Information Integration Architecture Based on Semantic and Grid", *IEEE*, 978-1-4244-5540-9, 2010.

[36] Ya.Li, Hairui.Wang Lin.Jin, Yuping.Liu, "A Multi Agent Framework for Enterprise Information Integration", *IEEE*, International Conference on Measuring Technology and Mechatronics Automation, 978-07695-3962-1, 2010.

[37] X.Y. Li, Y. Qian, "A Web service Based Enterprise Information Integration Model", *IEEE*, Proceedings of 4th International Conference on Computer Science and Education, 978-1-4244-3251-0, 2009.

[38] Aditya Telang and Sharma Chakravarthy and Yan Huang, "Information Integration Across Heterogeneous Sources: Where Do We Stand and How to Proceed?", *CSI* ,International Conference on Management of Data COMAD 2008, India, , 2008.

[39] Song Xiao, Zhang Lin, Yu Peng-fei, "Semantic SOA-Based Enterprise Information System Integration Technology", International Conference on Smart Manufacturing Application, April. 9-11, Korea, 2008.

[40] Xiangyu Li, "An Integration Framework for Information System based on Web Service", *IEEE*, 978-1-4244- 5265-1, 2010.

[41] R.H. Bayardo et al., "InfoSleuth: Agent-based semantic integration of information in open and dynamic environments", Proceedings of the ACM SIGMOD, 1997, pp. 195–206.

[42] Jingtao Zhou, Haicheng Yang, Mingwei Wang, Rongxia Zhang, Tao Yue, Shusheng Zhang, Rong Mo, "A Survey of Semantic Enterprise Information Integration", *IEEE,* Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, 978-1-4244-7386-1, 2010.

[43] Patrick C.K. Hung, Dickson K.V.Chiu, "Devolping Workflow-based Information Integration with Exception Support in a Web Services Environment", *IEEE,* Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.

[44] Bahram Amini, Roliana Ibrahim, Mohd Shahizan Othman, A Framework for Personalized Information Integration in Higher Education Institutes, *International Journal of Computer Applications*, Vol 2, No.4, 0975 – 8887, June 2011.

[45] Shuwang Ke, Jitao Fang, XiangqianDing, Shusong Yu, Dan Li, " The Design and Development of Shop-level Information Integration System" , *IEEE*, International Conference on Electrical and Control Engineering, 978-0-7695-4031-3, 2010.

[46] Yong-Gang Gong, Xin Chen, "Healthcare Information Integration and Shared Platform Based on Service-Oriented Architectures", *IEEE,* 2nd International Conference on Signal Processing Systems (ICSPS), 978-1-4244-6893, 2010.

[47] Haiyan Chen, Shigang Qin, Jianxun Liu, Jian Cao, "SOA-Enabled Health Information Integration Platform (HIIP): A Case Study", *IEEE*, Fifth International Conference on Semantics, Knowledge, Grid, 978-0-7695-3810-5, 2009.

[48] Asit Dan, Robert Johnson, and Ali Arsanjani, "Information as a Service: Modeling and Realization", *IEEE* International Workshop on Systems Development in SOA Environments (SDSOA'07), 2007.

[49] Matt Estes,Fania, Jason, Li-Yi Lin, Ivy Lin, Shawn Ramse, Bala Rasaratnam, Mick Symonds, *Open Data Center Alliance Master Usage Model: Information as a Service Rev. 1.0,* 2013.

[50] "Information as a service: Achieve new levels of flexibility, consistency and availability," IBM Corporation, ibm.com/software/data/ondemandbusiness, 2005.

[51] Noel Yuhanna and Mike Gilpin, *The Forrester Wave: Information-As-A-Service Q1 2010,* Report 2010.

[52] Laura Haas "Beauty and the Beast: The Theory and Practice of Information Integration" ICDT 2007, Vol. 4353 (2006), pp. 28-43.

[53] Laura M. Haas, Aya Soffer, "New Challenges in Information Integration", DaWaK 2009, *Springer-verlag LNCS* 5691, pp.1-8, 2009.

[54] Haresh Luthria, Fethi A. Rabhi, "Service–Oriented Architectures: Myth or Reality?" *IEEE Software*, 0740-7459, 2012.

[55] C.Punitha Devi, V. Prasanna Venkatesan , G. Shanmugasundaram, "Research opportunities in Information Integration", *AICTE* National conference on Internet and Web Service Computing (NCIWSC 2012), Pondicherry Engineering College,India, 2012

[56] Noel Yuhanna and Mike Gilpin, *Information Fabric 3.0,* Forrester Report, 2013.

[57] Brian Hopkins, Alex Cullen, Mike Gilpin, Boris Evelson, Gene Leganza, and Mackenzie Cahill, *Data Virtualization Reaches Critical Mass*, Forrester Report, 2011.

[58] Kathryn B. Laskey, Kenneth Laskey, "Service oriented architecture", Vol1, John Wiley & Sons, Inc., 2009.

[59] C.Punitha Devi, V. Prasanna Venkatesan , G. Shanmugasundaram, "Information Integration Redefined", *ARPN Journal of Systems and Software*, Vol 2, No:4, 2222-9833, 2012.

Authors' Profiles

C. Punitha Devi obtained her B.Sc in Computer Science (1996) and M.C.A (1999) from Bharathidasan University and M.Tech in Computer Science & Engineering (2007) from Pondicherry University. Currently she is pursuing PhD in Department of Computer Science and Engineering, Pondicherry University. She is having 10 years of teaching experience. She has published one book and papers in national and international journals/ conferences. Her research area includes Service Oriented Architecture and Software Engineering.

Dr. V. Prasanna Venkatesan is currently an Associate Professor, Department of Banking Technology, Pondicherry University. He earned his B.Sc in Physics (1986) from Arignar Anna Arts College, Karaikal. He received his M.C.A (1989) from Pondicherry Engineering College, M.Tech in Computer Science & Engineering (1995) from Pondicherry University and Ph.D in Computer Science & Engineering (2008) from Pondicherry University. He is having more than 20 years of teaching experience. He has published 4 books and papers in national and international journals/ conferences. His research area includes Software Architecture, Banking Technology, Object Oriented Modelling and Design, Smart Banking.

Diwahar. S received his B.E Degree in Electronics and Communication Engineering from DMI College of Engineering, Chennai under ANNA University. Currently pursuing M.Tech in Network and Internet Engineering from Pondicherry University, INDIA.

Dr. G. Shanmugasundaram obtained his B.Tech in Information Technology (2005) from BCET, Pondicherry university. He received his M.Tech in Computer Science and Engineering (2008) from SMVEC, Pondicherry University. He was awarded Doctorate from Pondicherry University (2013). Currently he is working at SMVEC, Pondicherry. He is having 2 years of teaching experience and 1 year in software development. He has published several papers in national and international journals/ conferences. His research area includes Service Oriented Architecture, Service Discoverability, Software Metrics and Web Technologies.