



HAL
open science

Ontology-based Mediation with Quality Criteria

Muhammad Fahad

► **To cite this version:**

Muhammad Fahad. Ontology-based Mediation with Quality Criteria. 8th International Conference on Business Intelligence (CBI 2023), Jul 2023, Istanbul, Turkey. pp.1-15, 10.1007/978-3-031-37872-0_6 . hal-04106981v2

HAL Id: hal-04106981

<https://hal.science/hal-04106981v2>

Submitted on 25 Jul 2023


HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Ontology-based Mediation with Quality Criteria

Muhammad Fahad ¹


Univ Lyon, Univ Lyon 2, UR ERIC
5 avenue Mendès France, 69676 Bron Cedex, France
f.muhammad@univ-lyon2.fr

Abstract. *Information integration* has a long history since humans started using and collecting information. But, it has been a strong focus of IT research since many recent years. It deals with providing a unified and transparent access to a collection of heterogeneous data sources. In information integration, the formulization of a global schema is a difficult task that manages multiple, autonomous and heterogeneous data sources. This paper presents a semantic system named *OntMed* for an ontology-based data integration of heterogeneous data sources to achieve interoperability between them. Our system is based on the quality criteria (*consistency, completeness and conciseness*) for building the reliable analysis contexts to provide an accurate unified view of data to the end user. The generation of an *error-free global analysis context* with the semantic validation of initial mappings generates accuracy, and provides the means to access and exchange information in semantically sound manner. In addition, data integration in this way becomes more practical for dynamic situations and helps decision makers to work within a more consistent and reliable *virtual data warehouse*.

Keywords: Information Integration, Ontology based Mediation, Ontology, Quality Criteria

1 Introduction

Information integration has a long history since humans started using and collecting information. But, it has been a strong focus of IT research since the 1970s. It deals with providing a unified and transparent access to a collection of heterogeneous data sources. In information integration, the formulization of a global schema is a difficult task that manages multiple, autonomous and heterogeneous data sources. According to the study by Bernstein and Haas [1], large enterprises utilize at least 40% of their budgets on information integration. A forecast about the market for the worldwide data integration and access software estimated to increase from \$2.5 billion to \$3.8 billion in 2007 to 2012 [2]. Although there is a huge amount of research done on this topic, but, it is still a hot issue in IT research. The two major research challenges are Intra-organization and Inter-organization information integration. *Intra-organiza-*

¹  <https://orcid.org/0000-0002-7258-9884>

tional data integration is vital when different components of an organization adopt different systems to maintain. The need for Inter-organizational data integration is required in companies' merger, stock exchanges, etc. Research on this topic evolves with time from a centralized system with single and multiple data stores to managing federated data sources. Then, decentralized systems were designed where distributed integration is done by the application with or without central global schemas. Therefore, in the research literature, there are various approaches for information integration, such as data exchange, mediator-based, P2P data integration and exchange, data warehousing (DWH), etc., [3]. Data exchange aims at materialization of the global view by providing facilities of query answering without accessing the local data sources. There can also be P2P data integration and exchange between several peers. The approach is designed such that it allows queries over one peer, such that each of the peers is equipped with local and external sources.

When a data warehouse is overloaded with data, Chen and Zhao indentify some limitations such as flexibility, efficiency and scalability [25]. Therefore, an approach for data-information integration between *data cubes* is introduced to prevent hurdles in data maintenance and data analysis. There is also a need for data warehouse security techniques, such as data access control, which may not be easy to enforce and can be ineffective. Therefore, DWH also requires a cubic-wise balance method to provide privacy preserving range queries on data cubes in a data warehouse [26]. Along with privacy, quality of information is important as well. This paper is focused on *mediator based data integration* or *virtual warehouse based on quality criteria*. *Virtual* means that the data remains at the local data sources and a mediation layer provides *transparency* in answering querying by managing underlying local heterogeneous data sources.

The rest of the paper is organized as follows. Section 2 presents background terminologies for the reader. Section 3 discusses related work. Section 4 presents our *ontology-mediator based data integration* system (*OntMed*). Section 5 presents the evaluation of our system with the help of Ontology Alignment for Query Answering (OA4QA) track of Ontology Alignment Evaluation Initiative (OAEI). Finally, section 6 concludes our paper and shows future directions.

2 Background

This section presents the background concepts such as categories of DWH and mediation approaches. The first subsection presents three categories of DWH. The second subsection discusses different approaches for ontology-based data integration of heterogeneous data sources.

2.1 Data Warehouse and it's Categories

Inmon provides the first formal definition of a data warehouse as "A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process" [10]. In general, there are three categories of data warehouses [27].

Virtual view approach. First, the virtual view approach deals when the physical data is stored in the local databases (dbs) and the repository of DWH only contains the schema of data. Query pre-processing and query shipping are required to answer queries which were made against the integrated view. This approach results in poor performance.

Materialized view approach. Second, the materialized view approach deals when the data schema and the physical data both reside inside the repository of DWH. This approach leads to data management and maintenance problems due to large volumes of data.

Datamart approach. Third, the datamart approach extracts data for a special purpose from the primary DWH for a datamart application. In this way, DWH only consists of limited knowledge and unable to manage huge amounts of data.

When data is stored in multi-dimensional form, it is visualized as a cube. When an enterprise implements datamart approach for building DWH, many cubes are formalized and each cube is an independent data aggregation. Each dimension in the cube depicts subsections of data which can be comparable and aggregatable. These independent cubes are isolated bits of information when investigated by analysts. They retrieve information from one single angle and not from a global view. This leads to problems like data duplication, inconsistency and query integrity [28]. For an example consider an example of data cubes illustrated in Fig. 1. The data of Cube Z consist of cube X and Cube Y, which may contain duplication of data and inconsistency issues may arise. Therefore, one must consider quality criteria when handling information integration of data cubes.

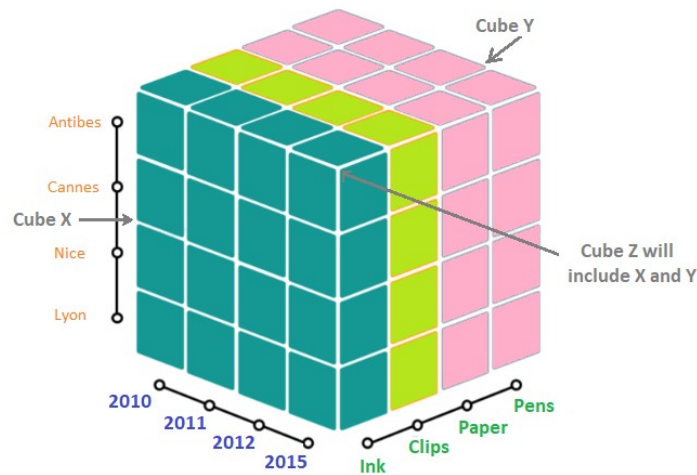


Fig.1. An example of data cubes

2.2 Broad categories of Ontology based information integration

In research literature, there exist three approaches for querying heterogeneous data sources. According to the survey on ontology based information integration by Wache et al. [4], existing approaches can be classified into three broad categories. These different categories for ontology based data integration are represented in Fig. 2 and elaborated below.

Single Ontology Approach. In a single ontology approach, all source schemas are directly related to a shared global ontology. The shared global ontology provides a uniform interface and facilitates querying to the end-user. SIMS is a system that exploits a single ontology approach for querying heterogeneous ontologies [5].

Multiple Ontology Approach. In multiple ontology approach, each local data source is represented by its own local ontology. Local ontologies are mapped to each other on the basis of their similarities, i.e., their inter-ontology mappings are defined. The user query is rewritten for each local source and finally results are merged as per the inter-ontology mappings. OBSERVER is a system that exploits multiple ontology approach for querying heterogeneous ontologies [6].

Hybrid Ontology Approach. Hybrid ontology approach takes the advantages of both the above approaches [7]. For each local data source, there is a representative local ontology. All the local ontologies are mapped to a global shared ontology that provides a unified view. The advantage of this approach is that new data sources can be easily integrated without changing in existing mappings.

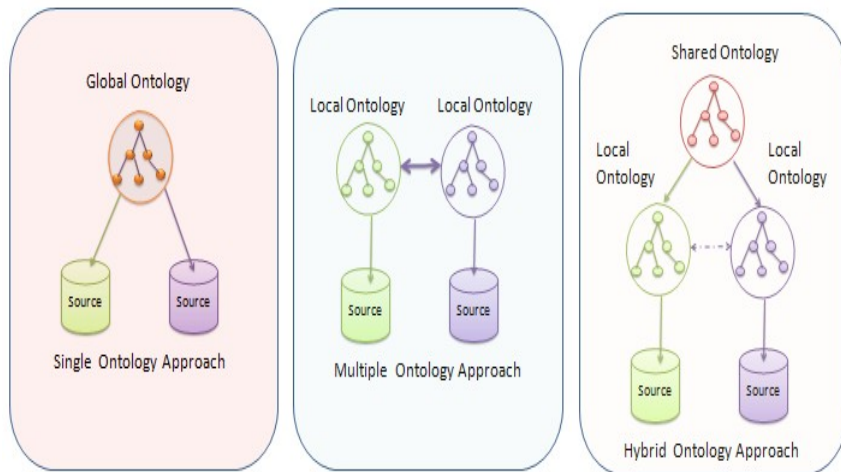


Fig.2. Different approaches for Ontology based Data Integration

On the basis of these approaches, mapping between the local and global ontologies are specified. When the source schemas are defined with respect to the global schema, the approach is known as Local-As-View (LAV). But when the global schema is defined with respect to the local schema, the approach is known as Global-As-View (GAV). A mixed approach can also be defined as GLAV.

3 Related Work

Data integration is essential since the analysis context is built using data from different heterogeneous data sources and providing an end-user with the unified view of underlying data [8]. For the integration of heterogeneous data and on-line analytical processing, data warehousing has long been adopted that aims at building a centralized database which contains all (or selected) data originating from various heterogeneous data sources designed in the multidimensional manner [9]. Two well-known techniques for building warehouses are provided by *Inmon* [10] and *Kimball* [11]. In addition, on-line analytical processing (OLAP) systems may stock terabytes of data and handle queries from millions of end-users. Since in the data warehouse, heterogeneous data is warehoused that facilitates the decision processes, this approach is recognized by its performance due to its query response time. With the time when there is a change in data, updates are required by the decisional tools for the accurate decision making. The update process needs much additional cost for the refreshment of data. Data warehouse by mediation is thus increasingly seen as a solution to this problem that aims at building a virtual data warehouse to build analysis context on-the-fly.

Kharlamov et al. develop Ontology Based Data Access (OBDA) system which contributes a conceptual view of data and the ontology performs as a mediator between the user and the data [32]. It enables querying a database which is represented with ontology by abstracting away from the technical schema-level details of the local data sources. Later on, Can and Unalir reuse the Ontology Based Access Control (OBAC) model within the scope of OBDA which supports a privacy framework [33]. The aim of their work is ontology-based secure data access so that information integration can be performed in a privacy-aware manner.

Besides these, there are some works for information integration of data cubes. Kaur and Kaur implement cost-effective parallelization techniques for cluster based algorithm to compute data cubes [24]. Their approach exploits existing sequential proposals and is intended to target load balance and formulates the communication effectively and efficiently. Similarly to facilitate information integration, Chen and Zhao propose rule-based cube exploration by introducing inter-relationships that hold among data cubes to avoid loss of data semantics [25]. They mix OLAP technology with association rules to tackle the problem of navigation of large data so that decision makers can benefit by using low capacity devices such as mobile devices. Liu et al. also present a cubic-wise balance approach for the privacy preservation and better range query accuracy in data cubes. Their approach provides a precise estimation of the answer for range-sum queries, enabling preservation of the confidential informa-

tion of individual data cell in a data cube [26]. Huang et al. propose a semantic cube model approach for data warehouse enhancement to reduce the duplication of data and to improve the performance of query integrity [28]. An end-user designs the generalization relationship between different cubes. To reap the targeted benefits, they exploit connectivity by identifying the relationship between data cubes.

Ontology-based information integration is being used in multiple domains. Sobral et al. propose an ontology-based framework to support integration and visualization of data from intelligent transportation systems [29]. They develop Visualization-oriented Urban Mobility Ontology as a semantic pillar for knowledge-assisted visualization tools. Their domain ontology consists of three facets which interrelate the characteristics of spatio-temporal mobility data, visualization mechanisms and expert knowledge. Belitz-Hellwich implements information exchange between a logistics simulation model and a finite element model through ontologies [30]. They connect source models of both domains, build interfaces between the models and the ontologies, and design an application that executes the exchange of information between them.

Now-a-days, the use of ontologies for describing semantics of data in data warehouse design has revolutionized the performance in terms of accuracy and provides means to exchange information in semantically sound manners in heterogeneous environments. Although there is already a lot of research in this area, there are still many issues that need to be resolved, especially concerning the quality of results. Based on the previous research [12], we analyzed various situations when local ontologies are merged together that need inconsistency, incompleteness and inconciseness checking. Therefore, we build a quality criteria based on *consistency, completeness and conciseness* parameters for building the reliable analysis context in the data warehouse design based on a mediation approach. We further investigated our approach that exploits the semantic validation of initial mappings from the information present in the local ontologies that form global analysis context on-the-fly. Global ontology as an analysis context provides a conceptual unified view to the end-user about the underlying data. The user performs a query over the data. As this query is expressed in a global schema terminology, it must be reformulated in terms of local data sources such that it can be executed. Otherwise, query results are incomplete or null. Once the results are retrieved from the global schema, they need to be merged together and presented to the user in terms of a global terminology. In this whole process, we believe that the generation of error free global analysis contexts with semantic validation of initial mappings would generate more accuracy, provide means to exchange information in semantically sound manners, data integration by this way becomes more practical for dynamic situations and helps decision maker to work within more consistent and reliable virtual data warehouse. There are many other problems under this topic, such as data extraction, cleaning, reconciliation, and optimization of query answers, but these are out of scope of this paper. Our work is limited to identification of mappings between local data sources, validation of mappings based on quality criteria, generation of global schema, answering queries formulated with the vocabulary of the global schema and aggregation of final results.

4 Ontology-based Data Integration

Traditional data warehouse systems based on *ETL* mechanisms *extract*, *transform*, and *load* data from several heterogeneous data sources into a single queryable data. This approach of data integration is tightly coupled because the data is located in the same repository during the execution of query, therefore provides fast query response. As of modern research (almost from 2009 onward), the trend in data integration has favored loosening the coupling between the data. This requires mediation based uniform query-interface, where an end-user can write a query, and the mediation layer transforms it into specialized queries over the original databases, and retrieves the results. One example of a data warehouse based on mediation approach is presented by [13].

A data warehouse based on a mediator approach consists of three main elements; *data source schemas* as local ontologies, *a global ontology* and *correspondence rules* between local-global ontologies for the query answering. Building a global ontology is essential as it allows the execution of decisional queries on underlying sources, query conversion from the global schema vocabulary to the data sources vocabulary, and building the data cube on-the-fly from the obtained results from the different data sources. But, the mediation layer has to face a number of problems; since local ontologies may represent the same knowledge in different ways producing various mismatches while the construction of a global schema. In addition, it is very hard to perform their manual integration beyond a certain degree of size (number of concepts, properties and relations), complexity of axioms, and number of ontologies. Therefore, it needs fully automatic techniques for supporting interoperability in dynamic environments such as building virtual data warehouse where analysis context is made on-the-fly. This dilemma requires the need of a reliable ontology merging approach that should be capable enough to find semantic heterogeneity in source ontologies and resolve it automatically to produce accurate, consistent and coherent global merged ontology on-the-fly based on a number of local source ontologies. Since it is in the analysis and decision domain, validation of initial mappings that form global schema is more than a challenging task for decision making and sound manipulation of data. Therefore, we took this initiative to integrate quality criteria inside the mediation layer to build a reliable analysis context to achieve correctness of desired query results.

4.1 Representative Schema for Local Data Sources

According to Haw et al. (2017) the mapping between a database and an ontology is regarded as a case of data integration [31]. Concepts of the ontology are mapped on the relevant entities in the db. We believe each of the data sources has some representative schema, i.e., local source ontologies already exist. If not, use some tool to generate individual ontologies [14, 15].

4.2 Generation of Global Schema

The second step of ontology-based data integration is to formulate the global schema. Our approach is flavored by building a mediator based data integration that facilitates to physically reside data in original sources. It minimizes the burdens for end-users from locating sources significant to a query, interacting with each one in isolation, and manually integrating data from multiple data sources. For our ontology-based data integration system, we use our previously designed ontology merging system DKP-AOM that deals with the problem of providing a unified view (in terms of global ontology) of local ontologies which represent distributed and heterogeneous data sources. Further information about our DKP-AOM system can be found in [16]. The user creates a query by using vocabulary of the global schema and the mediator exploits some techniques to carry out the query to achieve meaningful answers. For this, it needs to translate a query expressed with the mediated schema terminology into one that refers directly to the local schema.

4.3 A Novel Framework for Data Integration with Reliable Mediation

We design a novel framework for the data integration with reliable mediation named *OntMed* based on a hybrid approach; where local ontologies represent local data sources, and our DKP ontology merging framework generates a global ontology automatically from the local ontologies. Finally a query engine executes query and retrieves result by query rewriting, obtains local answers, and constructs global result. Fig. 3 depicts the data integration approach adapted by our system *OntMed*.

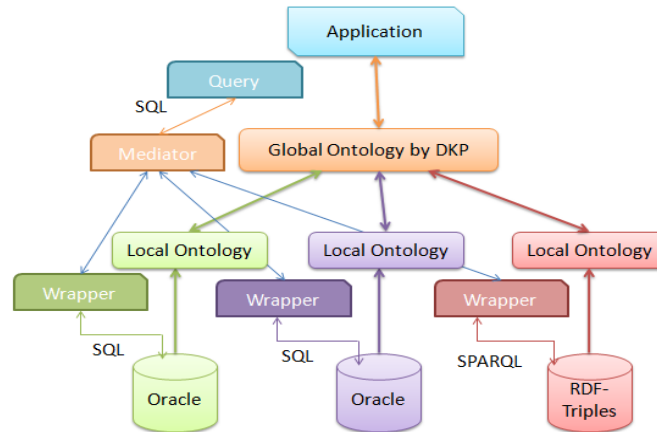


Fig.3. Data Integration approach adapted by OntMed

It is well argued in the research area that often the generated results of data integration suffer from inconsistency, incompleteness and redundancy. Therefore, we implemented the quality criteria based on the Consistency, Completeness and Redundancy parameters [17, 18] in a mediation layer to build an accurate analysis context so that

execution of simple queries on the global ontology reveals correct results. Fig. 4 depicts our quality criteria for achieving reliable mediation and also summaries various errors. The important thing which should be considered is that global ontology should be free from these errors as it serves an analysis context on which user queries are performed and final result is aggregated. The details about these criteria are as follows.

Inconsistency in the global ontology. Inconsistency in the global ontology means that there exists some disagreement within the ontology. It also means there exists some conflicts or sort of contradictory knowledge inferred from the concepts, definitions and instances within the ontology. Inconsistency in the global ontology produces ambiguities, contrary axioms, contradictions in final results and compromises precision. There are mainly three types of errors that can cause inconsistency and ambiguity in the ontology. These are Circulatory errors, Partition errors and Semantic inconsistency errors.

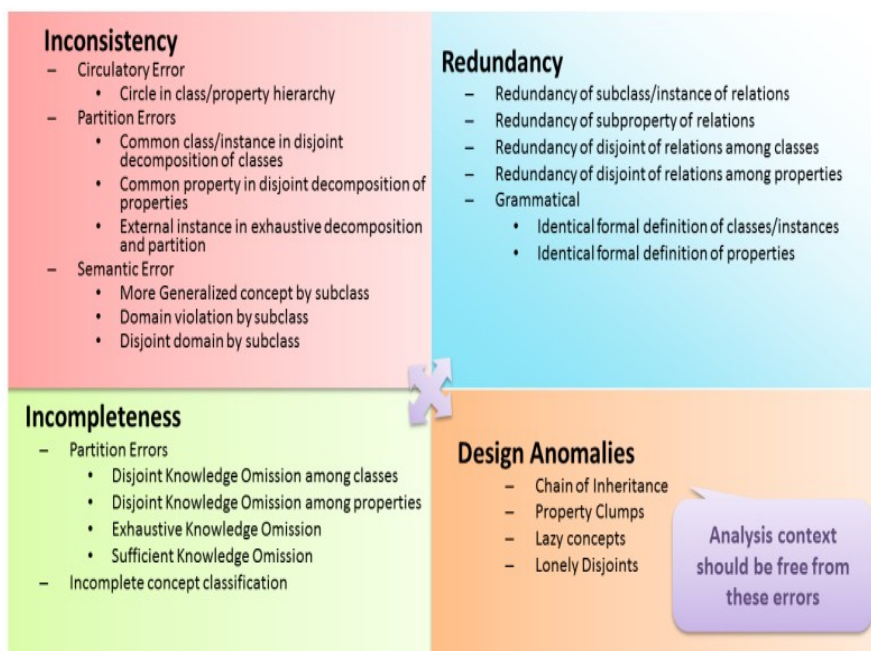


Fig.4. Quality criteria for Reliable Mediation

Incompleteness in the global ontology. Incompleteness occurs when the global ontology contains various types of domain knowledge in the form of concepts, properties and definitions, but overlooks some of the important information about the domain. The incompleteness of the domain knowledge often creates ambiguity, and lacks reasoning and inference mechanisms. The incompleteness errors are due to in-

complete concept specification and partitions errors due to disjointness and exhaustive knowledge omission between concepts.

Redundancy in the global ontology. Another important aspect is to make global ontology concise, so that it stores only necessary and sufficient knowledge about the concepts. Inconciseness or repetition or redundancies in global ontology not only compromise usability and convenience, but also create problems for the maintenance, preservation and manageability of concepts and properties within ontologies. Generally, there are redundancy errors when particular information is inferred more than once from the relations, classes and instances found in the global ontology.

Design Anomalies in the global ontology. Besides taxonomical errors, Baumeister and Seipel pointed out some anomalies based on bad designs. These anomalies restrict simplicity, coherence and maintainability of taxonomic structures within the ontology [19]. These anomalies do not generate incorrect reasoning from the axioms, but highlight controversial, problematic and poorly designed areas in ontology that cause issues of maintainability. Therefore it is advisable to detect and repair or remove these anomalies. Ontologies free from these anomalies improve the usability, clarity and provide better maintainability over the semantic web.

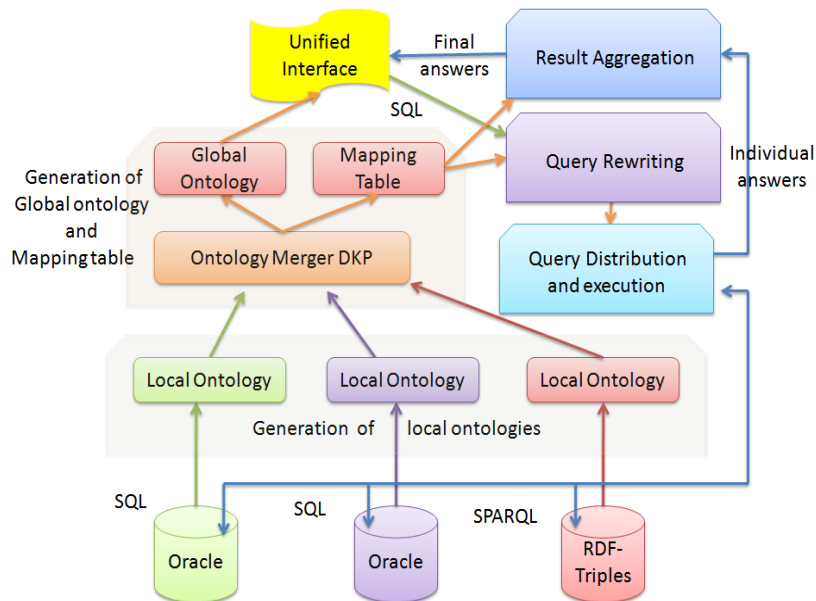


Fig.5. Execution of different steps by OntMed

Once the global ontology is generated by the DKP ontology merging system, this serves as an analysis context for our data integration system. We have participated in the conference and OA4QA track of OAEI 2015 and results of our system are com-

petitive, see [20]. Fig. 5 illustrates the execution of different steps of our system for the data integration. We have tested our approach of data integration on a case study related to Shopping Mall data sources which are modeled in two instances of Oracle. These data sources are heterogeneous and separately developed. In addition, we blend our case study with the RDF triple data source, to add more heterogeneous environment in the case study. However, we have incorporated very limited information about the RDF data source in this case study, and mainly it circulates over Oracle data sources. In the near future, we will present this case study.

5 Validation of OntMed Methodology

We implement OntMed (ontology-based data integration via mediation) module as a part of an Ontology Merging System (DKP-AOM). For the evaluation of our system, we choose the Ontology-Based Data Access (OBDA) scenario (see Fig. 6). This scenario is considered as a part of Ontology Alignment for Query Answering (OA4QA) track of Ontology Alignment Evaluation Initiative (OAEI) 2015 [21]. Although it is old, it serves its purpose to validate the methodology. There are three rules which should be taken into account to reduce the number of potentially unintended consequences [22]. The following are these rules:

- (i) *Consistency Rule*. It means that the alignments should not bring unsatisfiable classes in the integrated merged ontology. Only consistent alignments should be detected between source ontologies which lead to consistent integrated ontology.
- (ii) *Locality Rule*. It means that the alignments should link concepts that have similar neighborhoods. Concepts in the integrated ontology must have similar neighboring concepts that are present in the source ontologies.
- (iii) *Conservativity Rule*. It means that the alignments should not propose changing in the classification of the source ontologies. New classification of concepts in the integrated ontology must be avoided and preservation of the same classification of concepts is maintained in the integrated ontology.

There are two ontologies (i.e., QF-Ontology and DB-Ontology) in the OBDA scenario. QF-Ontology offers the vocabulary to formulate the queries. DB-Ontology is connected to the data and it is hidden to the users. The integration based on proposed mappings is achieved between these two ontologies since the data is linked only with the vocabulary of the DB-Ontology. When a query and a pair of source ontologies are provided as input, then the model result set will be calculated using the correspondences for the ontology pair.

The accuracy of resultant model answer sets are measured based on the Precision and Recall [23]. The evaluation of ontology alignment based on a set of reference alignments are taken into account investigating this aspect of data integration. Precision and recall values are computed by initiative campaign based on the capacity of

the generated mappings to answer a set of queries in an OBDA scenario in presence of multiple ontologies. The track is intended to study the practical consequences of logical violations disturbing the detected mappings. In addition, they compare the different repair techniques adopted by the participant ontology matching systems.

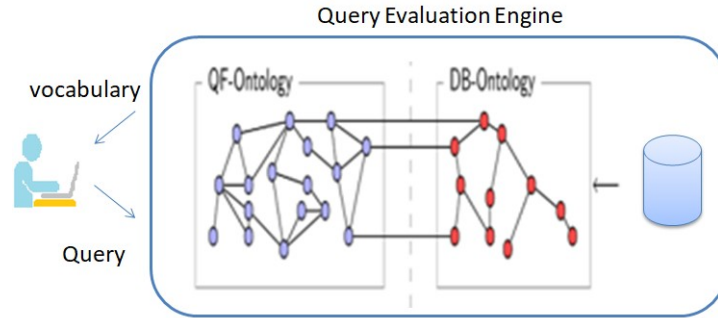


Fig.6. Ontology-based data access scenario

The well-known conference data set is used to promote understanding of the dataset and the queries. However, it is extended with synthetic ABoxes. The results of participated systems are taken into account by the SEALS platform. The evaluation results are generated on available reference alignments of Conference track (RA1) and its repaired version (RAR1). Table 1 illustrates the results for the whole set of queries showing the average precision, recall and f-measure. All the participant Ontology matchers are tested on 18 queries, for which the sum of expected answers is 1724. There is only one answer for some queries. But some queries have as many as 196 answers. The details of participant ontology matching systems, evaluation criteria and evaluation results can be found in their individual research publications. However, the following points highlight the evaluation results showing the salient features of our tool.

- The alignments generated only by the systems (AML, DKPAOM, LogMap, LogMap-C and XMap) allowed answering *all the queries* of the evaluation.
- According to jury of campaign, the best global results have been achieved for violations queries. Ontology Matching tools (AML, DKPAOM, Log-Map, COMMAND, LogMapC and XMAP) generate excellent correct results, in decreasing order of f-measure w.r.t. RA1.
- Our system DKPAOM obtained a remarkable f-measure of 0.999 w.r.t. RAR1 (see evaluation results aggregated by query family²). It means logical violations are successfully handled by our tool.

² <https://oaei.ontologymatching.org/2015/oa4qa/results.html>

Table 1. Global Evaluation Results published by OA4QA track
 **taken from the original paper [21]

Matcher	Answered Queries	ra1			rar1		
		Prec.	Rec.	F-m.	Prec.	Rec.	F-m.
AML	18/18	0.778	0.750	0.759	0.757	0.750	0.746
COMMAND	14/18	0.557	0.611	0.575	0.536	0.611	0.562
CroMatcher	12/18	0.464	0.479	0.434	0.407	0.458	0.404
DKPAOM	18/18	0.667	0.618	0.635	0.666	0.639	0.648
GMap	9/18	0.324	0.389	0.343	0.303	0.389	0.330
JarvisOM	17/18	0.409	0.430	0.386	0.407	0.430	0.385
Lily	11/18	0.389	0.368	0.376	0.389	0.389	0.389
LogMap	18/18	0.750	0.750	0.741	0.729	0.750	0.728
LogMapC	18/18	0.722	0.694	0.704	0.722	0.694	0.703
LogMapLite	11/18	0.409	0.423	0.379	0.351	0.402	0.348
Mamba	14/18	0.556	0.528	0.537	0.556	0.528	0.537
RSDLWB	6/18	0.131	0.180	0.127	0.129	0.180	0.126
ServOMBI	6/18	0.222	0.222	0.222	0.222	0.222	0.222
XMAP	18/18	0.778	0.675	0.702	0.720	0.654	0.671

6 Conclusion

Data integration is crucial since the analysis context is built using data from different heterogeneous sources. With the grown usage of data warehouse, the question about the usage of a mediation approach for building analysis context on-the-fly has become even more important in today's dynamic world. Although, there has been already a lot of research in this area, there are still many issues that need to be resolved especially concerning with the quality of results. There are various points about the ontology merging and conceptual schema merging. In general, an ontology is a broad concept than conceptual schema and a conceptual schema is regarded a sub-concept of ontology. Therefore, merging conceptual schema requires less effort than ontology merging. Integration of databases has been the focus of years of research. There are many approaches in the research literature for the data integration with/without the support of ontologies. Ontologies conceptualize concepts with their generalization and formulate properties and axioms to equip the semantics. The relational model on the other hand provides limited semantic description about the data. For example, there is no or very limited generalization support, no axiomatic definitions, etc. An absence of the rich semantics also poses a difficulty in recognition of concepts during their mappings and hence their merging raises the level of difficulty.

This paper contributes an ontology-based data integration framework. Integration of data sources also depends on the representation and quality of its representative schema or ontology. The more representation model is good, the more integration is

easily done with quality. The presented approach based on the quality criteria (consistency, completeness and conciseness) for building reliable analysis contexts in data warehouse design suits well and provides accurate unified view of data to the end user. In addition, our approach builds the virtual data integration environment with the least human intervention. Our automatic ontology merger generates the global ontology from the local ontologies. Our approach exploits semantic validation of initial mappings from the knowledge that is present inside the local ontologies that form global analysis context on-the-fly. The generation of an error-free global analysis context with the semantic validation of initial mappings generates more accuracy, and provides means to exchange information in semantically sound manners. In addition, data integration by this way becomes more practical for dynamic situations and helps decision maker to work within more consistent and reliable virtual data warehouse.

One of the future directions is to present a case-study on ontology based information integration where we demonstrate how various errors in the mapping phase create hurdles achieving correctness and accuracy in the integration process and aggregation of final results. Other is to evaluate our framework on a recent dataset and report its performance.

Acknowledgement

The research depicted in this paper is funded by the French National Research Agency (ANR), project ANR-19-CE23-0005 BI4people (Business Intelligence for the people).

References

1. Bernstein, P.A., Haas, L.M.: Information integration in the enterprise. *Communications of the ACM (CACM)*, vol. 51, no. 1, pp. 72-79, (2008).
2. Ziegler, P., and Dittrich, K.R.: Three Decades of Data Integration - All Problems Solved?, *WCC*, pp. 3–12, (2004).
3. IDC, *Worldwide Data Integration and Access Software, 2008-2012, Forecast*. Doc No. 211636, Apr. (2008).
4. Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Häubner, S.: *Ontology-Based Integration of Information - A Survey of Existing Approaches*. In *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing*, (2001).
5. Arens, Y., Knoblock, C.A., Hsu, C.: *Query Processing in the SIMS Information Mediator*. In *The AAAI Press*, May, (1996).
6. Mena, E., Kashyap, V., Sheth, A.P., Illarramendi, A.: *OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*. In *Proceedings of the 1st IFCIS International Conference on Cooperative Information Systems (CoopIS 1996)*, pages 14-25, (1996).

7. Cruz, I.F., Xiao, H.: Using a Layered Approach for Interoperability on the Semantic Web. In Proceedings of the 4th International Conference on Web Information Systems Engineering (WISE), pages 221-232, Rome, Italy, December (2003).
8. Lenzerini, M.: Data Integration: A Theoretical Perspective. PODS, pp. 233–246. (2002).
9. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology, ACM SIGMOD Record, v.26 n.1, p.65-74, March (1997).
10. Inmon, W.H.: Building the Data Warehouse. John Wiley & Sons, Inc., New York, USA, (1992).
11. Kimball, R.: The operational data warehouse, DBMS, 11(1), 14-16, (1998).
12. Fahad, M., and Qadir, M.A.: A Framework for Ontology Evaluation, 16th ICCS Supplement Proceeding, vol. 354, pp.149-158, France (2008).
13. Maiz, N., Fahad, M., Boussaid, O., Tayab, F.B.: Automatic Ontology Merging by Hierarchical Clustering and Inference Mechanisms, In proceedings of 10th International Conference on knowledge Management and Knowledge Technologies (I-Know'10), Sept 1-3, Messe Congress Graz, Austria, (2010).
14. Upadhyaya, S.R., and Kumar, P.S.: ERONTO: A Tool for Extracting Ontologies from Extended E/R Diagrams, ACM Symposium on Applied Computing (2005).
15. Xu, Z., Zhang, S., Dong, Y.: Mapping between Relational Database Schema and OWL ontology for Deep Annotation. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), IEEE, (2006).
16. Fahad, M., Moalla, N., Bouras, A.: Detection and Resolution of Semantic Inconsistency and Redundancy in an Automatic Ontology Merging System, Journal of Intelligent Information System (JIIS), Vol. 39(2) pp. 535-557, (2012).
17. Gómez-Pérez, A.: Evaluating ontologies: Cases of Study. IEEE Intelligent Systems and their Applications, vol. 16(3): 391–409, (2001).
18. Gomez-Perez, A., Fernández-López, M., Corcho, O.: Ontological engineering: With examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web. Springer, London, (2004).
19. Baumeister, J., Seipel, D.S.: Smelly Owls—Design Anomalies in Ontologies, 18th Intl. Florida AI Research Society Conference, pp. 251–220. AAAI Press. (2005).
20. Fahad M.. Initial results for ontology matching workshop 2015, DKP-AOM: Results for OAEI 2015. In CEUR Workshop Proceedings 1766, pp. 82–96. 5 (2015). <http://oaei.ontologymatching.org/2015/conference/index.html>
21. Cheatham, M., Dragisic, Z., Euzenat, J., Faria, D., Ferrara, A., et al.: Results of the Ontology Alignment Evaluation Initiative 2015. 10th ISWC workshop on ontology matching (OM), Oct, Bethlehem, United States. pp. 60-115. (2015).
22. Solimando, A., Jimenez-Ruiz, E., and Guerrini G.: Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings. International Semantic Web Conference (ISWC), (2014).
23. Solimando, A., Jimenez-Ruiz, E., and Pinkel, C.: Evaluating Ontology Alignment Systems in Query Answering Tasks. Poster paper at International Semantic Web Conference (ISWC), (2014).
24. Kaur, P. and Kaur P.: New Approach of Computing Data Cubes in Data Warehousing, International Journal of Information & Computation Technology. Volume 4, Number 14 pp. 1411-1417, (2014).
25. Chen Z. and Zhao T.: A cube model approach for Data warehouse, Proceedings of the 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering, pp. 846-849. (2015).

26. Liu, Y., Sung, S.Y., Xiong, H.: A cubic-wise balance approach for privacy preservation in data cubes. *Information Sciences*, vol. 176, pp.1215–1240. (2006).
27. Alejandro, G.T., Marotta, A.: *An Overview of Data Warehouse Design Approaches and Techniques*. (2001).
28. Huang, S., Chou, T., Seng, J.: Data warehouse enhancement: A semantic cube model approach. *Information Sciences*, Volume 177, Issue 11, pp. 2238-2254. (2007).
29. Sobral, T., Galvão, T., Borges, J.: An Ontology-based approach to Knowledge-assisted Integration and Visualization of Urban Mobility Data, *Expert Systems with Applications*, Vol. 150, (2020).
30. Belitz-Hellwich, W.: *An Ontology-Based Platform for Information Integration; Supporting Sustainable Smart Transportation Infrastructure*, Feb 17, (2023).
<https://www.diva-portal.org/smash/get/diva2:1737543/FULLTEXT01.pdf>
31. Haw, S.-C., May, J.-W., and Subramaniam, S.: Mapping relational databases to ontology representation: A review. In *ICDTE'17*, pages 54–55. (2017).
32. Kharlamov, E., Hovland, D., Jimenez-Rui, E., et al.: Ontology based data access in statoil. *Journal of Web Semantics*, 44:3–36. (2017).
33. Can, O. and Unalir, M.: Revisiting Ontology Based Access Control: The Case for Ontology Based Data Access, *Proceedings of the 8th International Conference on Information Systems Security and Privacy (ICISSP)*, pp. 515-518. (2022).