Analyzing the Ontology Approaches and the Formation of Open Ontology Model: A Step for Organisational Ontology Employment

Jawahir Che Mustapha Yusuf[1]
Institute of Research and Postgraduate Studies (IRPS)
UniKL-Malaysian Institute of Information Technology
Kuala Lumpur, Malaysia
jawahir@miit.unikl.edu.my

Mazliham Mohd Su'ud
Universiti Kuala Lumpur (UniKL)
UniKL-Malaysia France Institute
Kuala Lumpur, Malaysia
mazliham@unikl.edu.my

Patrice Boursier[1]
Department of Artificial Intelligence, FCSIT
University of Malaya
Kuala Lumpur, Malaysia
patrice@um.edu.my

***Abstract*—Three different ontology-based approaches have been used in previous researches to improve the semantic interoperability in an integrated information system. The approaches can be identified as the single, the multiple and the hybrid ontologies. Organisations seeking to improve their information system capability realise the benefits of using semantic technology based on ontology. However, clear guidelines are not available to select the appropriate ontological approach. The selection of the approach should be according to various organisational needs, contexts and management styles. This research is significantly important to provide flexible and adaptable way to start employing ontology, because semantic information systems are still immature in many organisations. In current research the study of different ontology-based approaches is presented. The focus is on the semantic integration challenge based on multi-sources data integration. Viability of all approaches and guides for ontology employment are presented in order to provide options for the organisations to upgrade their current system to new system. There is no specific approach that has been proven to be a successful implementation. Therefore, a new general reference model is proposed in this research work, which is based on the three approaches called Open Ontology Model. The proposed model is designed to work in dual directions which are top-down and bottom-up implementation to make the specification of ontology mappings more flexible and usable. This model would be of interest to novice system developers who plan to use it as a starting point to develop their first semantic information system. Developers might decide any single or combination of approaches based on the nature of their organisation.**

Keywords-ontology-based information system; semantic heterogeneity; data integration.

I. INTRODUCTION

The needs for knowledge sharing and exchange within organisations have become the most significant and prominent cause of data integration. Therefore, information system interoperability is a key to increase cooperation between all data owners to ensure successful data integration. At present information systems are increasingly large-scale, complex and multi-traits. Information sharing and exchange processes are going to be more challenging. Data integration procedures must follow good abstraction principles to solve interoperability problems concerning on the structure, the syntax, the system and the semantic. The focus of this research is on semantic integration which is one of the main issues in multi-sources data integration.

According to [1], semantic integration is the task of grouping, combining or completing data from different sources by considering explicit and precise data semantics. Semantic integration has to ensure that only data related to the same real-world entity is merged. Ontology is a current practice to resolve semantic conflicts in diverse information sources. Ontology itself is an enabling technology (a layer of the enabling infrastructure) to enforce knowledge sharing and manipulation [2]. Any abstract or concealed information can be clearly described according to specific concepts by using ontology.

Researches to employ ontology approaches for integration of multiple data sources are still growing and more demanding as semantic reconciliation can resolve other types of interoperability problems. Three approaches have been used in previous researches that can be identified as single, multiple and hybrid ontology [9][31]. Large number of systems still holds implicit information even though they might have well support on technical data interoperability. Realizing the growing importance of semantic interoperability, organisations are beginning to use ontologies in their system applications. However, common guidelines to find the ontology approaches that are best suited for different organisational needs, contexts and management styles are still unclear. There are organisations that start with complex approach or approach that is not suitable to some types of organisations. In fact, there exists a

[1] Author is also associated with Laboratoire L3i Université de La Rochelle, La Rochelle, France.

 ISBN: 978-1-61208-175-5

much simpler, cost-effective and quick alternative to be exploited with some improvement. Knowing the advantages and disadvantages of different approaches are not enough to help choose the right approach for a given application. More importantly, there should be a mechanism in place to help the organisations decide the necessary information system upgrades on the basis of their management structure and nature. Furthermore, system developers must deliberately choose proper ontological methods at early stages of system development. Otherwise, invalid result from queried information might yield bogus decision due to poor understanding on the knowledge.

This paper discusses different ontology-based approaches for supporting multi-sources data integration. Viability of all approaches and guides for ontology employment are presented in order to provide options for the organisations to upgrade their current system to new system. A new ontology-based model that is called Open Ontology Model (OOM) is also proposed in this research work. It is intended to be used as general reference model to novice system developers who plan to use it as a starting point to develop their first semantic information system. Developers can take advantage of each ontology approach and may build their systems by stages depends on organisation system requirements and the current resources available. Currently, the prototype of this research work based on the OOM is under implementation.

The rest of this paper is structured as follows: Section II elaborates related researches on ontological-based approaches. Meanwhile, Section III presents the viability of ontological approach and guides to ontology employment. The formation of OOM is detailed out in Section IV. Section V briefs the motivation of this research work. Finally, conclusion is added in Section VI.

II. Revision on the Ontology Approaches

The use of ontologies for data integration is applicable to various numbers of applications. This part describes top-down and bottom-up ontology development. Then, the three ontology approaches based on previous researches contribution in [8][9][10][31] are revised. More recent researches are added to show some earlier approaches still relevant in particular domain background. Indeed, the formation of the Open Ontology Model (OOM) is rooted from the three approaches. The advantages and disadvantages of each approach are not to be emphasized. The concern is more with the numerous types of organisational environments which need to decide the most suitable ontology approaches for their information system upgrade.

A. A Glance on Top-down and Bottom-up Ontology

In computer science perspective, ontology is important for data integration in order to facilitate shared and exchanged information. Generally, two popular trends exist in the development of ontology approaches; top-down and bottom-up designs. In the top-down design, each term in source ontologies is created from the primitive term in the set of top-level ontology. The set of top-level ontology is provided first. Secondly, source ontologies that contain more specific terms are extended from the set of top-level ontology. Since source ontologies only use the vocabulary of a top-level ontology, therefore terms are comparable easily. In the top-level ontology, only common terms are described at a very abstract level. Therefore, adding up existing ontologies should not become a problem as many upper-ontologies (or upper-domain ontologies) are developed under consideration it can be easily reused. The knowledge-base CYC [39], SUMO [38], Sowa’s upper ontology [41], WordNet [42], DOLCE [40] and UMBEL [43] are the examples of top-level ontology.

On the other hand, the bottom-up ontology design is aimed to build shared, global ontology by extracting data from source ontologies. Firstly, source ontologies that contain specific terms are constructed from data source schema (or catalogues, labels etc) to describe the meaning of the information. Secondly, source ontologies of all disparate data sources are mapped to construct primitive terms or abstract concepts of the top-level ontology (common shared vocabulary). This way, the related terms between low-level and top-level ontologies are still comparable.

B. Ontology Approaches Revisiting

In *single ontology approach*, a global ontology is derived by data interpretation from all connected data sources as depicted in Fig. 1a. One common shared vocabulary is provided to denote the semantics between data sources. Global ontology development efforts primarily focus on the formation of general knowledge used in multi-purpose applications. A few former systems based on the single approach can be located in the Carnot system [12] that utilises the global CYC ontology [11], an ontology modularization technique in ONTOLINGUA [13], TAMBIS for connecting biological data sources [14], and SIMS [15] as the tightly-coupled system that is tested in the domains of transportation planning and medical trauma. This approach is still utilised in recent years with some improvements such as for spatial data integration in SPIRIT [5][16], a geo-ontology construction for web spatial data query system, three-level ontology architecture for geo-information services discovery in [17] and OCHRE [36] core ontology for combining cultural heritage information from diverse local schemas.

In most real-time implementation, it is not easy to completely achieve mutual agreement within data owners to use one common vocabulary. Thus, m*ultiple ontology approach* is aimed for data integration by mapping different ontologies without using global schema. Each data source is described by its own disparate ontology (Fig. 1b). Inter-ontology mapping technique must be used to enable association between ontologies. Mapping provide a common layer from which several ontologies could be accessed, and hence could exchange information in semantically sound manners [18]. This approach is presented in earlier systems such as OBSERVER system [19] for domain of bibliographic references, combination of two different geographic ontologies using bi-directional integration in [21], MAFRA system [20] and SEWASIE [6]

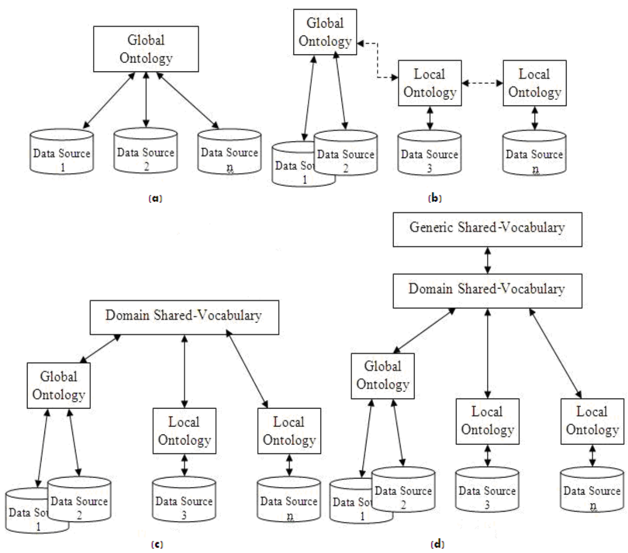

Figure 1. Different ontology approaches: (a) single ontology approach, (b) multiple ontology approach, (c) and (d) two types of hybrid ontology approach that are simulated from [9]

system that use multiple ontologies to provide access to heterogeneous web data and the ontology translation (*bridging axioms)* to merge two related ontologies in OntoMerge XML-based system [4]. More recent work on the approach can be found in [34], where YAGO ontology [33] was automatically derived from Wikipedia and WordNet, further work in [32] to combine high-level axioms from the SUMO and YAGO, and MEMO [35] an automatic merging of two source ontologies, which uses clustering techniques in order to help the identification of the most similar ontologies.

Another mode of multiple ontology integration is done via one shared-vocabulary to make these ontologies simply comparable to each other. This most adopted approach is known as the *hybrid ontology*. Generic ontology and domain ontology are the type of shared-vocabulary. Domain shared-vocabulary can be specified from or without generic shared-vocabulary (Fig. 1c and Fig. 1d). Generic shared-vocabulary usually contains very basic terms in a universe of discourse while domain shared-vocabulary models more specific concept of the world. In some hybrid ontology approach, domain shared-vocabulary is split up into top-domain and domain ontologies as described in [22]. Particularly, hybrid ontology approach set a global top-level ontology to appear as a common reference framework (foundational ontologies) for multi-application and/or multi-domain. The aim is to encourage ontology reuse to facilitate semantic interoperation between applications [10]. At the low-level, all source ontologies that are involved in the integration will use the terms specified in the shared-vocabulary. Simultaneously, each source ontology does not need to be concerned with the context of other source ontologies.

Wache et al. [9] described concisely on the implementation of the hybrid ontology approach in former systems such as COIN, MECOTA and BUSTER. The same approach is used by Elmore et al. [23] to solve a problem of losing data when one global ontology is used. They proposed computer agents over shared-vocabulary to merge only relevant ontologies within participating data sources (USA national lab system). In [3], the authors extended a hybrid ontology approach by defining the XML schema for each data source. The XML schema was then used to create local ontologies before abstracting the equivalent concepts in global ontology. In order to relate between global and local ontology, a mapping rule was applied using path-to-path approach with XQuery language for global query. Bellatreche et al. [24] attempted to achieve a fully automated technique for heterogeneous sources integration of electronic catalogues within engineering databases. Their technique preserves the autonomy of various data sources in which all data sources reference a shared-ontology, and possibly extend it by adding their own concept specializations. In GeoNis [7], semantic mediator was used to solve semantic heterogeneity of geographic data sources. GeoNis provides an ontology mapping between local and top-level ontology, and software support for semantic mismatches. Another related work, GeoMergeP system [25] also created for geographic data sources to focus on the improvement of semantic matching techniques (semantic enrichment and merging).

III. THE VIABILITY OF ONTOLOGY APPROACHES AND GUIDELINES FOR ONTOLOGY EMPLOYMENT

This section justifies the viability of all ontology approaches for different types of organisational environment. Basic guidelines for selecting the appropriate approach in multi-source data integration are also presented.

A. Viability of the Single, the Multiple and the Hybrid Ontology Approaches

In the early generation of ontology-based information systems, data integration adopted the single ontology approach. All data sources should abide with the same agreement to grant a very similar view on the domain. This means all data owners are required to retain and use a single, common ontology definition as well as at it local schema. Single ontology environment depicts that the newly added data source is modelled using terms from general, shared domain model only. Furthermore, a global ontology is also possible to be extended if the new data source goes beyond what is modelled in the current global ontology. Any changes such as alteration and deletion in data source will also imply the changes in global ontology. However, all the tasks are bounded by the size of the required data sources.

The integrated system based on the single ontology approach is applicable to certain environments which comply with specific principles. The single ontology mechanism is fine if data sources schema have no pre-existing ontologies and at once agreeable to use a global vocabulary. Data integration could be done if all data sources are able to share similar view on a domain of interest. The former mechanism

 ISBN: 978-1-61208-175-5

(i.e., SIMS), if changes occur in any data sources, will affect current global ontology and their mappings with other data sources. In order to resolve this issue, the creation of a mapping rule such as in [3] between a global ontology and local schema could be applied. Therefore, new sources can easily be added without the need to use a global ontology modification but only the mapping rule. Integration method in [37] is also feasible because the authors created user ontology that was independent of databases and similarity functions to compare related entities and instances in the system. User ontology allows users to express queries in their own terms according to their own conceptualizations without having to know the underlying modeling and representation of data in heterogeneous databases. Any updates in both the user ontology and the databases will not affect the system. Another issue of using this approach is the possibility to lost a valuable concepts of information could happen as described in [23]. If two or more data sources do not have a common view on some prospective information, it will not be appended in global ontology. This issue can still be resolved if some uncommon concepts are critically decided upon to be a sharable concepts in global ontology.

In other perspective, this approach is hard to support due to the complexities involved in integrating the ontologies and maintaining consistency across concepts from different ontologies with only a single shared-vocabulary [19]. On top of that, data sources should have full autonomy to sustain its own datasets. Thus, this approach is possible to be applied in less distributed environments where only fewer data sources exist and this situation enables simple ontology mapping process to be done. In a less heterogeneous organisational model such as in intra-government agencies, this approach can also be considered. Additionally, the frequency of future changes also should be nominal to avoid complexities while maintaining the integrated system. Overall, when the principles in single ontology approaches are difficult to be attained an alternative ontology approaches could be considered.

In the multiple ontology approach, the tasks such as insertion, exclusion or alteration of data sources are easily supported. Each data source has its own autonomy without being dependent on a global schema. The correlation between pre-existing multiple ontologies is easier than creating a global ontology because a smaller community is involved in the mapping process [20]. SEWASIE [6] developer also claimed that at the local level, things may be done more richly than at a wider level. In contrast, to compare different ontology sources are more challenging without common vocabulary. Furthermore, inter-ontology mapping is also prone to the complexities in query process. Although the use of inter-ontology mapping in [20] and [6] are rational, but system developers must also be concerned with the integration of large different ontologies. We might involve more complicated tasks of creating multiple mapping processes if existing mapping rules cannot be applied directly on new local ontology. Otherwise, this approach is simple and feasible.

Inter-ontology mapping is actually quite challenging to define in the environment when more than two information sources exist in the domain of interest. Mapping tasks become more complex as system developers might discover more semantic heterogeneity problems to correlate the ontologies between all the multiple sources. Many other mapping techniques are not clearly defined [26] and still remain as a research attention over recent years. Some discussions upon mapping for multiple ontology approaches can be referred at [26][27]. In other point of fact, the integration of a particular type of information within geographic and non-geographic data encompasses excellent implementation when using this approach, for instance in the domain of disaster management, forestry, land planning, and agriculture just to name a few. These kinds of information are typically distinct and independent in nature, and also in its description. They usually contain at least one common concept that could be related to strengthen the meanings of information. Thus, promising for data integration to facilitate effective information sharing under specific domain.

Data sources autonomy is partially vanished in former systems, which were based on the hybrid ontology approach. The existing ontologies cannot easily be reused and need to be redeveloped from scratch [9] by referring to the shared-vocabulary. Path to path approach and abstraction method as used in [3], and Ontology-based Database (OBDB) approach introduced in [24] could resolve the problem because the newly added data source is still able to maintain the autonomy by using its own local concepts. The hybrid ontology is a well-known approach that allows new data sources to be added easily in the ontology-based system. If new data source contains concepts that are not described with ontologies, local ontologies will be created for it by referring to the general terms established in shared-vocabulary. The sharable terms which are not specified in shared-vocabulary will be added directly in shared-vocabulary as general terms. Then, the mapping process of new terms is created to relate between local and shared-vocabulary. If new data sources come with pre-existing ontologies, system developer should investigate whether shared-vocabulary (upper to very upper level) is present or not. With the existence of shared-vocabulary, the different source ontologies should refer to the upper ontology with liberty to preserve its own concepts. The source ontologies may extend the upper ontology as much as required. Without shared-vocabulary, the different source ontologies could be connected using bottom-up direction to produce it common terms. The global ontology as in the single and the hybrid ontology approaches are actually transfers the burden of information correlation and filtering on the query processing system [19]. With global shared-vocabulary, the integration of pre-existing ontologies using global-local mapping rules will lessen the complexities in creating the query process compared with inter-ontology mapping.

B. A Proposed Guidelines for Ontology Employmet

Ontology-based information system for organisations (public and corporate sector) is still an immature field. Readiness for change to apply a formal ontological approach is a key factor to successful modern application integration solution. The selection of appropriate ontology approach is

 ISBN: 978-1-61208-175-5

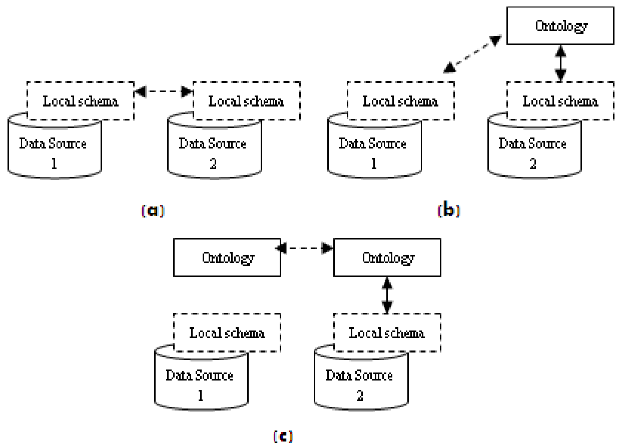

Figure 2. Integration of various system's structure: (a) Less explicit systems, (b) less explicit system and ontology-based system, (c) multiple ontology-based systems

solely depended on the organisational environment. Although the hybrid ontology perform well in most situations, the single and the multiple ontologies are also practical. Both approaches offer fast, economical and can possibly to be extended to the hybrid ontology. Once the data owners agree to use the ontology, they must properly decide on the ontology approach which is suitable for their organisation. So, organisational nature, scopes, information needs and resources are important in selecting the practical approach for ontology-based multi-source data integration.

Obviously, the majority of the current system holds less explicit information system integration (Fig. 2a). Modern information system is encouraged to embed more semantics in their systems to allow better information integration and this could be achieved by using ontology. Based on Section III-A, the single ontology approach is recommended if the data owners and their system conform to the following states:

1. Each data source contains at least one common concept and some uncommon concepts are declared sharable in global ontology to avoid data loss.
2. Each data owners participating in the integration process agree to use similar definition of global ontology.

Small-scale enterprise and intra-agencies usually possess common datasets that are maintained in distributed location. The single ontology will be practical for them in order to achieve low-cost, low-risk and fast deployment of semantic-based integrated system. The multiple ontology approach works very well if only two data sources are involved in the integration. Otherwise, hybrid ontology approach is more convenient as mapping process beneath global ontology simplify the complexities in inter-ontology mapping. In order to develop their first ontologies, data heterogeneities will be the first problem faced by the developers. Many research such as in [3][4][7][17][23][24] gave solutions to reconcile the heterogeneities.

In another situation, a possible integration could occur between less explicit data source with an ontology-based system (Fig. 2b). The first problem is to match local schema with pre-existing ontology. There is a possibility to reuse existing ontology as a global ontology (single ontology approach) if each data sources is able to share similar concepts. Otherwise, new ontology for non-ontology-based data source could be developed to enable peer-to-peer or hybrid ontology integration.

More challenges would be face by the system developers to integrate multiple ontologies (Fig. 2c). The problem here is the ontology heterogeneity. Even if each data source has its own ontology, the heterogeneity problems will still not resolved. Ontology merging is a common approach to combine existing ontology into common vocabulary that incorporates possible aspects of participating ontologies [27]. Another way to integrate multiple ontologies is thru ontology matching in order to define equivalent relation between different ontologies. The system developers should be able to resolve the inter-ontology integration complexities and maintaining consistency across different concepts. Euzenat and Shvaiko [28] described in detail how the matching technique should work for multiple ontologies. Even though having few complexities along with high cost and long-time implementation, the hybrid ontology approach could work well with pre-existing ontologies.

With regards to the selection of ontology approaches single ontology approaches will never suit with sustained and entrenched organisational models due to its costly transformation and maintenance process. Multiple ontology approaches is feasible if the developer is able to maintain all ontologies. They might create inter-ontology mapping (traversing semantic relationship) via terminological relationship. Less complexity in inter-ontology mapping can be achieved if ontologies which are to be integrated are nominal. Thus, this approach is not recommended for huge number of different specific ontologies as it becomes a great effort to traverse and understand all the semantic relationship. As such, the hybrid ontology approach that is supported with broad mapping techniques can almost fit all environments.

A notion that could add little add-ons to the organization ontology modelling theory is presented: Even though ontology is to describe the explicit meaning of knowledge, there is no explicit or better approaches for ontology employment since it really depends on the organisational structure and its management style, in accord with their scopes, the type of external information needs, and also the available resources such as personnel, financial, physical and their internal information itself.

IV. THE FORMATION OF OPEN ONTOLOGY MODEL

OOM (Fig. 3) is a general reference model for organisations data integration at semantic level. This model is meant for various domains of application (i.e., E-Government, Crisis Management etc.), to interconnect multi-sources data particularly on database components. Ontology building is expected to work in dual directions; top-down and bottom-up implementation. The model is aimed to be a flexible model for ontology employment by the organisations. The ontology-based model should in principle adopt a general to specific approach. Thus, the model is

 ISBN: 978-1-61208-175-5

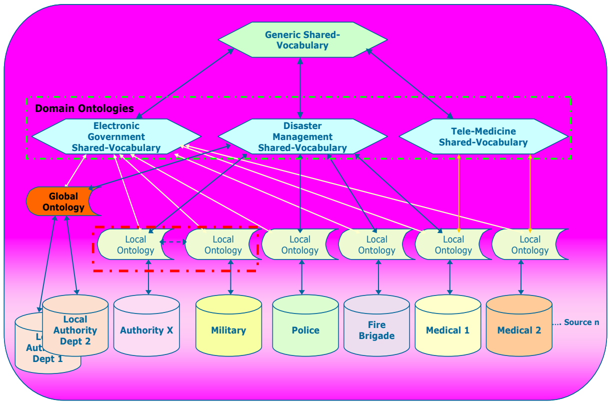

Figure 3. The Open Ontology Model

adequately expansive for explicit semantic data integration to avoid potential problems of under-specification. Afterward, the organisations can legitimately simplify the model according to their management needs. Obviously, the OOM is designed as the combination of available ontology approaches that feasible in most organizations environment.

In this model, each classes and property is assigned with *primary identifier* as in PLIB ontology [29] to map between concepts. The model approach works with or without existing source ontologies. It is assumed that generic or domain shared-vocabulary exists to be referred by low-level ontology (top-down to bottom-up). But it doesn't mean that explicit mapping correlation must be made to refer to the upper ontology. This happens when the participating organizations decide to use the single ontology approach. The single ontology is constructed with consideration on the existence of the upper ontology, so that the single ontology will be constantly ready for upgrading into hybrid ontology for connecting multiple data sources. That is also similar with the organisations who decide to use the multiple ontology approach. Two participating data sources shall contain its own ontology that is created in advance with respect that there exist a generic or domain shared-vocabulary. In future, mapping rules to connect between two ontologies may be used to adapt with hybrid ontology environment.

Hybrid ontology approach is anytime viable to associate less or more data sources. If the participating data sources in the integration process have no pre-existing ontologies, each local ontology will be created with reference to shared-vocabulary. The local ontology possibly will extend its body to have more specific entities and properties. In the pre-existence of ontology, this source still has the autonomy to maintain its name concepts. The *primary identifier* is used to indicate the similarity or different concepts between participating data sources and it upper-vocabularies. Fig. 4 depicts the top-down to bottom-up mapping implementation with the use of *primary identifier*.

Local ontology is defined based on the schema of the local database. Data owners will decide their own definition of local ontology concepts. Concepts that are rational to be disclosed will be pulled out to domain-shared list. Concealed concepts (shaded in Fig. 4) will not be shared but can be

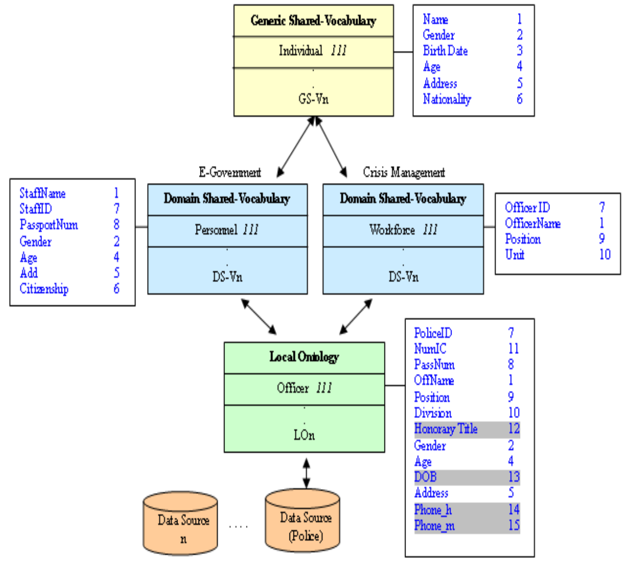

Figure 4. Top-down to bottom-up mapping

accessed locally or may be shared (right away or later) in different domain. Generic and domain shared-vocabulary are the list of shared concepts for all participating data sources. In our approach, the design of shared-vocabulary begins with inspirational approach [30]. For instance, 'National Security Division' as the principal initiates the specification of generic and domain shared-vocabulary that is substantially potential to be shared with the group of the data owners. Concerned with the importance of information sharing, the data owners may collaboratively [30] use the existing shared-vocabulary as the anchor and supportively extend it if necessary. However, the data source owners will not be attentive to each other's data. This is important for most of the intelligence systems that are confidentiality-related. Some ontology standards (ISOs, ANSI etc.) and/or other common top-level ontologies (WordNet, OpenCyc etc.) may be reused during the ontology design time.

V. MOTIVATION OF THE RESEARCH WORK

The prototype of research work based on the OOM is currently under implementation. The attention is given to perform an ontology-based integrated system beneath the crisis management domain within the Malaysian public agencies, particularly amongst local authorities, police, fire brigades and medical agencies. An example of study is drawn from digitized, multi-format documents that are collected before and after disasters. The data sets are typically stored in heterogeneous GIS-based (raster images or vector) proprietary or open formats such as Shapefile, MapInfo TAB, GML, KML JPEG2000, DEM, GeoTIFF, etc. Besides, some photographic images, text-documents, video and audio clips which are collected aftermath of a disasters allows the decision makers to see the big picture of the disaster events. Even though they are maintained and distributed by different information systems, formats, organizations and locations, but their contents might carry

 ISBN: 978-1-61208-175-5

one and the same calamity story, situation, related and supporting each other. Access to all of this valuable data needs high performance of information retrieval and integration mechanism that is effective at gathering, analyzing and outputting the required information.

Malaysia has good mechanism in managing disasters and the committee was established at three different levels (Federal, State and District under National Security Division Secretariat) to coordinate all the activities related to disaster. Various agencies perform their own daily work routine and maintain their own information either manually or in digitized form (flat files, databases and etc.). During disaster events, huge amount of information are acquired to be disseminated amongst them. However, the required datasets are not only difficult to obtain from system network but lack of automated data coordination at operational level such as during counter-disaster, rescue and relief activities. In addition, if information system is utilised, each agency may use different terminology to refer to similar data, and different document format to store spatially and semantically related information. Ontology usage in information system is still at infant stage amongst the Malaysian public agencies. Furthermore, ontology in this domain is not yet exists in the context of Malaysian disaster management. This research opens up significant opportunities to achieve more flexible and adaptable way to start employing ontology within many organisations.

VI. Conclusion and Future Works

Various ontology-driven information system approaches for multi-sources data integration is presented to provide direction for ontology employment among different organizations. Based on this study, the organizations should not adhere to employ directly specific model approach but are given as much autonomy as possible with respect to their nature along with their resource allocation and acquisition. Both the single and the multiple ontologies have high level of implementation feasibility because the approaches provide a quick way to develop quick, low risk and low-cost system application. Furthermore, the approaches may be extended to hybrid ontology when greater integration of heterogeneous data sources is required. A hybrid ontology approach can almost fit all environments but the challenge of having more ontology heterogeneity could delay the development. Besides a flexible OOM that is feasible in most organizations environment is also proposed. The ontology is designed to follow inspirational and collaborative approach with the top-down to bottom-up implementation. The OOM could be replicated in developing the semantic-based application for various domains of interest.

The presented model approach to design an ontology provides the basis for developing and implementing the ontology-based system. The system is aimed to improve multi-source, multi-format document query and integration particularly for disaster management domain. Further research is focusing to make better the ontology building, along with testing and evaluating the concepts in domain and application ontology. The ontology matchmaking is primarily come into focus to help achieve the goal of automatic data search and integration to response a specific query.

Acknowledgment

We would like to express our sincere thanks to IRPS-UniKL and Majlis Amanah Rakyat (MARA) for the research funding and administrative support, Laboratoire L3i Université de La Rochelle for research and administrative support, and we are also thankful to Mr. Muhammad Alam and Ms. Kim de Silva for advising us on this paper.

References

[1] P. Ziegler and K. R. Dittrich, "Three Decades of Data Integration - All Problems Solved?," in 18th IFIP World Computer Congress (WCC 2004), Building the Information Society, vol. 12, Aug 2004, pp. 3-12.

[2] T. R. Gruber, "Ontology of Folksonomy: a Mash-up of Apples and Oranges," International Journal on Semantic Web and Information Systems, vol. 3(1), 2007, pp. 1-11.

[3] L. Zhang, Y. Ma, and G. Wang, "An Extended Hybrid Ontology Approach to Data Integration," in Proceedings of 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09), Tianjin, China, 2009, pp. 1-4.

[4] D. Dou, D. McDermott, and P. Qi, "Ontology Translation on the Semantic Web," On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, 2003. pp. 952-969.

[5] C. B. Jones, A. I. Abdelmoty, D. Finch, G. Fu and S. Vaid, "The Spirit Spatial Search Engine: architecture, ontologies and spatial indexing," Geographic Information Science, 2004, pp. 125-139.

[6] S. Bergamaschi, F. Guerra, and M. Vincini, "A peer-to-peer information system for the Semantic Web," Agents and Peer-to-Peer Computing, 2005, pp. 161-182.

[7] L. Stoimenov, A. Stanimirovi, and S. Djordjevic-Kajan, "Semantic Interoperability using multiple ontologies," Proceedings printed as book, Eds. Fred Toppen, Marco Painho, AGILE, 2005, pp. 26-35.

[8] U. Visser, H. Stuckenschmidt, and C. Schlieder, "Interoperability in GIS-enabling technologies," in Proceedings of the 5th AGILE Conference on Geographic Information Science, Palma de Mallorca, Spain, 2002, pp. 25-27.

[9] H. Wache, et al., "Ontology-based integration of information - a survey of existing approaches," in Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, 2001, pp. 108-117.

[10] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," ACM Sigmod Record, vol. 33, 2004, pp. 65-70.

[11] D. Lenat and R. V. Guha, Building Large Knowledge-based Systems: Representation and Inference in The Cyc Project: Addison-Wesley, 1990.

[12] C. Collet, M. N. Huhns, and W. M. Shen, "Resource integration using a large knowledge base in Carnot," IEEE Computer, 1991, pp. 55-62.

[13] T. R. Gruber, "Ontolingua: a mechanism to support portable ontologies," Knowledge Systems Laboratory, Stanford University, California, Tech. Rep. Version 3, 1992.

[14] R. Stevens, et al., "TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources," Bioinformatics, Oxford Univ. Press, England, vol. 16(2), 2000, pp. 184-185.

[15] Y. Arens, C. A. Knoblock, and W. M. Shen, "Query reformulation for dynamic information integration," Journal of Intelligent Information Systems, vol. 6, 1996, pp. 99-130.

[16] G. Fu, C. B. Jones, and A. I. Abdelmoty, "Building a geographical ontology for intelligent spatial search on the web," in Proceedings of International Conference on Databases and Applications, 2005, pp. 167-172.

[17] F. Probst and M. Lutz, "Giving meaning to GI web service descriptions," in 2nd International Workshop on Web Services: Modeling, Architecture and Infrastructure (WSMAI), Porto, Portugal, INSTICC Press, 2004.

[18] Y. Kalfoglou and M. Schorlemmer, "Ontology mapping: the state of the art," The Knowledge Engineering Review, vol. 18, 2003, pp. 1-31.

[19] E. Mena, V. Kashyap, A. Sheth and A. Illarramendi, "OBSERVER: an approach for query processing in global information systems based on interoperation across pre-existing ontologies," Distributed and Parallel Databases, vol. 8(2), 2000, pp. 223-271.

[20] A. Maedche, B. Motik, N. Silva and R. Volz, "MAFRA - A MApping FRAmework for distributed ontologies," Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, 2002, pp. 69-75.

[21] F. T. Fonseca and M. J. Egenhofer, "Ontology-driven Geographic Information Systems," in 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, November 1999, pp. 14-19.

[22] H. Stenzhorn, E. Beisswanger, and S. Schulz, "Towards a top-domain ontology for linking biomedical ontologies," Studies in Health Technology and Informatics, vol. 129(2), 2007, p. 1225-1229.

[23] M. Elmore, T. Potok, and F. Sheldon, "Dynamic Data Fusion using an Ontology-based Software Agent System," Proceedings of the IIIS Agent Based Computing, Orlando, vol. 7, 2003.

[24] L. Bellatreche, N. X. Dung, G. Pierra and D. Hondjack, "Contribution of Ontology-based Data Modeling to Automatic Integration of Electronic Catalogues within Engineering Databases," Computers in Industry, vol. 57, 2006, pp. 711-724.

[25] A. Buccella, et al., "GeoMergeP: Geographic information integration through enriched ontology matching," New Generation Computing, vol. 28, 2010, pp. 41-71.

[26] S. M. Falconer, N. F. Noy, and M. A. Storey, "Ontology mapping - A user survey," in Proceedings of The Second International Workshop on Ontology Matching, 2007.

[27] J. de Bruijn, et al., "Ontology mediation, merging and aligning," in Semantic Web Technologies: Trends Research and Ontology-based Systems, J. Davies, R. Studer and P. Warren, Eds. Chichester, UK: John Wiley and Son's, 2006, pp. 95-113.

[28] J. Euzenat and P. Shvaiko, "Ontology matching," Springer-Verlag, New York Inc., 2007.

[29] G. Pierra, "The PLIB ontology-based approach to data integration," in Proceedings of the 18th IFIP World Computer Congress (WCC'2004), Toulouse France, 2004, pp. 13-18.

[30] C. W. Holsapple and K. D. Joshi, "A collaborative approach to ontology design," Communications of the ACM, vol. 45, 2002, pp. 42-47.

[31] U. Visser, Intelligent Information Integration for the Semantic Web. Springer-Verlag, New York Inc, 2004.

[32] G. De Melo, F. Suchanek, and A. Pease, "Integrating YAGO into the Suggested Upper Merged Ontology," in Proceedings of 20th IEEE International Conference on Tools with Artificial Intelligence, IEEE Press, 2008, pp. 190-193.

[33] F. M. Suchanek, G. Kasneci, and G.Weikum. "Yago: A core of Semantic Knowledge". in Proceedings of World Wide Web, New York, USA, ACM Press, 2007, pp. 697-706.

[34] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A large ontology from Wikipedia and Wordnet," Journal of Web Semantics: Science, Services and Agents on the World Wide Web, vol 6(3), 2008, pp. 203-217.

[35] F. F. de Araújo, F. L. R. Lopes, and B. F. Lóscio, "MeMO: A Clustering-based approach for merging multiple ontologies," Workshops on Database and Expert Systems Applications, Bilbao, Spain, IEEE Press, 2010, pp. 176-180.

[36] Online Cultural Research Heritage Envirenment. (2011, May 8). Core Ontology. [Online]. Available: http://ochre.lib.uchicago.edu/index_files/Page845.htm

[37] M. Gutiérrez and A. Rodríguez, "Querying heterogeneous spatial databases: Combining an ontology with similarity functions," in Proceedings of the ER Workshop on Conceptual Modeling of GIS, 2004, pp. 160–171.

[38] A. Pease, I. Niles, and J. Li, "The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications," in Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada, 2002.

[39] C. Matuszek, J. Cabral, M. Witbrock, and J. DeOliveira, "An introduction to the syntax and content of Cyc," in Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and its Applications to Knowledge Representation and Question Answering, 2006, pp. 44-49.

[40] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE, Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, 2002, pp. 223-233.

[41] J. F. Sowa, Knowledge Representation - Logical, Philosophical, and Computational Foundations, Pasific Grove, California, USA, Brooks Cole Pulishing Co., 2000.

[42] Princeton University. (2011, June 21), WordNet A Lexical Database for English. [Online]. Available: http://wordnet.princeton.edu/wordnet/

[43] M. K. Bergman, and F. Giasson. (2008, July), Umbel Ontology-Subject Concepts and Named Entities Instantiation, Technical Report, vol. 2. [Online]. Available: http://fgiasson.com/articles/UMBELOntology_vA2.pdf

 ISBN: 978-1-61208-175-5