John Adetunji Adebisi, Babajide Samuel Afolabi, Bernard Ijesunor Akhigbe

A Value-based approach to modelling interoperability in Knowledge Organization Systems

Abstract

Existing KOS models are still evidence-based with a focus on how well they are holding up in terms of improvements. However, issues of correctness and reusability based on KOS conformance in terms of requirements re-engineering and metadata restructuring for interoperability gains towards better information access has not be given sufficient attention. This study, therefore, aims to provide insight into how to optimize interoperability in knowledge organization systems (KOSs) using an information system modelling approach. We have designed and implemented a unified value-based model using some identified mechanisms. Enterprise data information was integrated at the conceptual modelling level following the provisions in Diniz et al. (2006) to benefit KOSs in terms of interoperability with the goal of achieving information access. The identified components (or mechanism) required for the value-based model design were identified using appropriate fact-finding techniques, such as active stakeholder participation, interviews, record inspection, direct observation and searches from secondary sources. The design was implemented using extract-transform-load mechanism as a key technique, and the object-based and web-empowered methodology that uses Microsoft technologies such as SQL, ASP.net, web services and Transact SQL codes. Based on potential usage, the various modules that were integrated as a third-party value-based model were done by categorizing all extracted information to showcase seamless information access considering the level of interoperability achieved. This demonstrates the possibility of achieving seamless interoperability with legacy applications like KOS. In future, it will be interesting to implement a prototype of the proposed model, which is ongoing, in order to properly validate our claim.

1. Introduction

With today’s ever-increasing sources of online and offline data, interoperability is becoming more and more indispensable in order not to drown in data while starving for information (Ziegler 2004). In general, the goal of data integration is to combine data from different sources by applying a global data model. This often involves the detection and resolution of conflicts existing in the data schema to achieve homogeneity as well as provide a unified layer. The reason for data integration is two-fold: Firstly, given a set of existing data sources, an integrated view is created to facilitate data access and reuse through a single data access point. Secondly, given a certain information need, data from different complementing sources can be combined to provide a comprehensive basis to satisfy users information needs (Ziegler 2004). There is a remarkable history of research to draw from in terms of data integration for this paper to contribute further in the area of knowledge organisation system (KOS) interoperability. The spectrum of research ranges from early multi-database systems (e.g., Multibase) (Landers and Rosenberg, 1982), over mediator systems (e.g., Garlic) (Brown 2006) to ontology-based integration approaches (e.g., OBSERVER) (Knoblock 1996). What these approaches have in common is the autonomy of data sources that are
to be integrated. This was considered paramount. However, it came with an implication; users were not to be forced to adapt to any standard concerning the structure and semantics of the data they desire, something that was not even previously achievable.

From the literature, one can observe that; integration pertains to the exchange of meaningful information between systems regardless of whether the systems are designed to work together or not (Mousa et al. 2014). While interoperability should have allowed the smooth sharing of information; the interpretation of incoming data and its presentation as received, and the preservation of its original context must also be possible (Bobby 2017). To achieve this, a number of techniques, algorithms and solution(s) have been developed (Lenzerini 2002; McCann et al. 2003 and Chohan et al. 2010). Moreover, data integration can be categorized into two types based on the nature of integration and the techniques. The first type is physical integration using data warehouse (DW) techniques (Kondabolu and Nasina, 2010). The second type is virtual integration that generally adopts data virtualization (DV) as its primary technique (Kondabolu and Nasina 2010; Eden 2007). For meaningful information exchange or integration to happen; providers and consumers need compatible semantics between source and target systems (Rosenthal, 2004). The aim often is to engender better ways of sharing data among systems especially within an enterprise. This means answering the question of what concepts would be used to achieve descriptive metadata for existing information; and, what definition(s) would be adopted to allow newly-built systems to interoperate. The goal will be to create a common platform that will ease the goal of data sharing for which both the new and old systems were built (Rosenthal 2004).

The concept of interoperability in enterprise applications (EAs) requires the proactive production of useful semantic agreement, and not just document correspondences among existing systems. This will help the enterprise satisfy the need for new requirements, such as the collection of new data or the establishment of new data sharing arrangements. This will also lead to the reduction of costs by cutting down unneeded semantics and diversity of representation (Rosenthal 2004). This should exclude concept definitions and relationships. For example, one may provide data on the aircraft that are operated by a country’s airline, while somebody also may want data on all the aircraft that regularly use the airports in that country. To properly interpret and integrate this type of data, one must understand what portion of information is needed to actually satisfy the foregoing information need. This study considers “interoperability” as it pertains to tolerating organizational “realities”, which is often ignored. This definition is needful to help narrow the practicality of this work to the central authority of the system to be integrated. As a result, we use the requirements of a third-party application to achieve the aim of our research. The objective is thus to design and implement a unified value-based model using some identified mechanisms that will serve as the components of the proposed model. Therefore, as in Diniz et al.
(2006), this paper provides insight on how to optimize interoperability in knowledge organization systems (KOSs) using an information system modelling approach. Third-party application interoperability requires specific activities, which are quite difficult to re-engineer. However, with the techniques employed in this paper, it was possible to simplify their metadata as well as bring more value to them. This has implications for legacy applications like KOS and, based on potential usage, it will be possible to make various modules in KOS interoperable, as with third-party applications. The paper is further structured as follows: section 2 is committed to reviewing the literature; 3 and 4 contain the theoretical foundation and the methodology adopted respectively (in the paper); and the proposed value-based approach to modelling data interoperability; finally, the paper’s conclusion is presented in section 5.

2. Literature

Some authors have attempted to apply theory of value to integration and interoperability. For example, the work of Keizer et al. (2015) used the domain modelling approach to resolve the problem of heterogeneity that often originates from differences in terminologies, domain coverage and root ontology. This challenge comes from the use of application profiling as a way to extend individual domain knowledge. This makes it difficult for a non-universal layer to be used as a common layer to address different application needs. Achimugu et al. (2010) proposed an enhanced model for the software engineering process of EAs. The goal was to contribute to how the critical appraisal of software architecture and methodology as a tool for efficient software engineering processes was carried out. The research provided vital insight as to how to remove the bottleneck that is usually faced during the integration of disparate systems for better interoperability. Considering the impact of software architecture with respect to system components, the study provided a guide to software architecture and a best practice example both of which were adequate to support our data integration model. This was important since, an in-depth look at the architecture of an EA is necessary to carry out a re-engineering exercise.

Both the work of Pokharel (2010) and Fatudimu et al. (2013) considered the place of a framework for the integrated mining of heterogeneous data in decision support systems. These studies highlighted the need to discover knowledge from both structured and unstructured data if completeness and comprehensiveness for data sharing is to be possible, while stressing the need to tailor interoperability towards knowledge sharing. Furthermore, Lindström and Polyakova (2010) considered a customer relationship management tool, which can be studied to make interoperability towards information access work. Based on the results from analysis of the tool, an integrated customer-centric model was developed. Thus, it is absolutely necessary to consider users’ needs to have easy access to information from the context of enterprise databases.
3. The theoretical foundation

The relational database model (RDM), which forms the theoretical basis of EAs, is limited in terms of data interoperability. When third-party applications that require re-use of data from other sources are involved, the RDM literally fails during operation (Lenzerini 2002). Software engineering researchers have proposed different approaches and software vendors themselves have intensified efforts at finding useful tools to help with the task of data integration. However, these efforts remain an art that relies on human labour, due to the absence of a comprehensive philosophical guide or appropriate theoretical foundation. This means that the overall modelling process of data integration still depends heavily on a data integrator's experiences (Brodie and Liu 2010). In other words, existing EA data models are designed to support only a single version of the “truth” within their domain of existence. Therefore, there will naturally be difficulties during data integration when the data collected from diverse sources are to be looked at from different perspectives or from different levels of abstraction (Liu 2012).

Shadow theory (ShT) serves as the theoretical foundation in this paper. The premise for the choice of ShT is because of its conception that whatever can be observed and stored into databases regarding a subject matter are just shadows, that is, mental entities existing only in the viewers’ cognitive structures (Liu 2012). These mental entities can be constrained by the viewers’ internal model of reality. This is particularly true of an implicitly or explicitly chosen perspective(s) of reality that can be formally represented. This option of being able to formally represent reality as postulated by ShT allowed the use of algebraic theory to model data operations based on their meanings and not just their logical structure. Point-free geometry aided representation of data within the semantic space. Since this can be decomposed or aggregated in different ways concurrently; W(hat)-tags were used to draw on shadows for their meanings; and E(quality)-tags were used to recognize which meanings can be treated as the same. Enterprise customer data integration was exemplified to illustrate the data model design and operational principles.

Ideally, data interoperability is contingent upon the issue of reliability and effectiveness criteria. To bring EA integration up to speed in this regard, a combination of ontological, methodological, epistemological and ethical means can be adopted to ensure best practice and see that the foregoing criteria are met. Based on software engineering practice like that highlighted in Lakatos (1978) and Gruner (2011), efforts have concentrated on applying ShT to the re-engineering of EA as an artefact. As a result, the methodological and epistemological side of software development has been pursued.

4. Methodology

The various components of the value-based model proposed in this work and used
to demonstrate the intended value-based approach to modelling interoperability in KOs were identified through fact-finding techniques. The techniques included active stakeholders’ participation, personal interviews, record inspection, direct observation and searches from secondary sources. The information necessary to implement each identified component were extracted using the extract-transform-load mechanism. This mechanism is a database re-engineering technique that work with webservices such as JSON-RPC, JSON-WSP, Web Services Description Language (WSDL), XML Interface for Network Services (XINS), Web Services Conversation Language (WSCL), Web Services Flow Language (WSFL), Simple Object Access protocol (SOAP) and Representational state transfer (REST). This study used REST and Active Server Methods (ASMX) of .Net for the information exchange layer, although WSDL and JSON were also supported as they may be required during a third-party application connection. Third-party application of EA modules was adopted; interoperability was achieved using third-party consumption and decision-making procedures. These were specified by categorising extracted information based on potential usage.

The research design used use-case of Unified Modelling Language (UML) as a design tool. The value-based model was implemented using the Rational Unified Process methodology. By this methodology the Microsoft SQL Management studio and ASP.NET web services were leveraged for implementation purposes. The data virtualisation layer resulting from the interoperability model was specified as comprising of modules that were further partitioned into sub-modules.

4.1. Value-based model

The architecture in Figure 1 exemplifies the value-based model. The architecture is translatable to benefit KOSs by optimizing interoperability. KOS components such as libraries can be organized as presented in the architecture in Figure 1. Keizer et al. (2015) also discussed a distributed database approach using the example as specified in the architecture; each third-party application, internal and external business application and others can be integrated into the enterprise application. This may require a new database designed for the purpose. Unfortunately, this can be complex and error prone to achieve using the conventional application today.

However, at the interoperability layer and the unified information exchange layer, web services are introduced seamlessly with little or no human interaction. The model combined existing knowledge with the knowledge gained from field work and case studies to provide stakeholders with a solution for effective decision making (in the form of associated business rules). For this study, the businesses are the protocols defined to ensure interoperability. For KOS, there will be a need for such protocols to fully allow the intended interoperability and they should be deployed at the value-based interoperability layer. The aim of the protocol is to ensure seamlessness Buyya, et al. (2009). Based on this, access to enterprise application data will be seamless since the
bottleneck of interoperability will have been removed.

Figure 1: A Value-based Integration Model for Enterprise Application showing the Interoperability layer deployed for third-party applications

4.2. Data interoperability

The proposed model was also developed following both the logical and physical foundation of a typical data warehouse. At this stage, the components that already exist in an organization (since we are considering EAs) were identified. The missing components were introduced. The relational and multidimensional nature of the databases used for data warehouses and their data marts was leveraged for ease of localization and interconnection of access tools. First, the logical and practical mode of architecture was defined as shown in Figure 1. This became the configuration for collecting the required data. In all, the system involved a central repository storing the data for an entire organization, an optional operational data store, data marts, and metadata repositories.

5. Value-based system architecture

The architecture presented in Keizer et al. (2015) continues to have a profound impact. It still serves as one of the de-facto standards for data interoperability. This study adopted and adapted it to show the novelty of the proposed value-based model. The value-based interoperability model, with its intended working parts as components, is presented as a systematic whole in Figure 1. The architecture shows how the model is deployed to work, especially in enterprise knowledge organisations. The model could also be deployed for business collaborations. This should be the same in all applications that require specific information for interoperability purposes. This is meant to ensure
that interoperability is managed at every point in time. As a result, at the integration layer, the model data will be the user data. The model provided proper infrastructure for data and application structure. It included the server, network, hardware and software components for connecting and communicating with third-party infrastructures. The technical architecture responded to the requirements of scalability, performance, availability, stability and security. The proposed model is thus robust, reliable, flexible, extensible and parallel.

6.0. Conclusion

We should highlight the fact that sections 3.0 and 4.0 were discussed in an instructive narrative and indeed the descriptive nature of this paper seeks to exemplify how KOS can benefit from interoperability when achieved in EAs. Our aim has been to hopefully whet the appetite of the knowledge organization research community concerning how to implement interoperability in KOS. This paper is therefore not prescriptive, which is one of its limitations. However, it proposes a value-based model capable of providing seamless integration. Clearly, since third-party applications require integration with tightly coupled EAs, they will seek to gain control of enterprise data and the schema structure will not be compromised. As an ongoing work, the value-based model is singled out as contributing to the development of a resourceful approach to knowledge sharing in the context of interoperability based on an integrated collaborative situation. In future, it will be necessary to implement a prototype of the proposed model. This we believe could lead to the value-based approach being considered capable of serving the functional requirements of legacy systems. Thus, the particular advantage of a value-based approach when compared to evidence-based methodology is the fact that it can be harnessed towards providing better interoperability for easy information access.

References


