A Large Scale, Knowledge Intensive Domain Development Methodology

Mayukh Bagchi

IECS Doctoral School, Department of Information Engineering and Computer Science (DISI), University of Trento, Trento, Italy, <mbagchi.research@gmail.com>

Mayukh Bagchi is a PhD scholar—cum—graduate teaching assistant at the Department of Information Engineering and Computer Science (DISI), University of Trento, Italy. His background is in library and information science and mathematics. His current research work revolves around adaptive data and knowledge representation and computational multilingualism. His general academic interests include symbolic and connectionist AI, mathematical epistemology, interdisciplinary KO, knowledge management, cognitive sciences, computational intelligence, HCI, open science ecosystem and altmetrics. Currently he is also exploring the conceptual convergence between KO and different mathematical theories as developed in pure mathematics and applied natural sciences.

Bagchi, Mayukh. 2021. "A Large-Scale, Knowledge-Intensive Domain-Development Methodology." *Knowledge Organization* 48(1): 8-23. 65 references. DOI:10.5771/0943-7444-2021-1-8.

Abstract: Since time immemorial, organization and visualization has emerged as the pre-eminent natural combination through which abstract concepts in a domain can be understood, imbibed and communicated. In the present era of big data and information explosion, domains are becoming increasingly intricate and facetized, often leaving traditional approaches of knowledge organization functionally inefficient in dynamically depicting intellectual landscapes. The paper attempts to present, *ab initio,* a step-by-step conceptual domain development methodology using knowledge graphs, rooted in the rudiments of interdisciplinary knowledge organization and knowledge cartography. It briefly highlights the implementation of the proposed methodology on business domain data, and considers its research ramifications, originality and limitations from multiple perspectives. The paper concludes by summarizing observations on the entire work and particularizing future lines of research.

Received: 4 April 2020; Revised: 9 August 2020, 15 September 2020; Accepted: 22 September 2020

Keywords: knowledge, domain, methodology, knowledge graph

1.0 Introduction

Throughout the history of human intellectual development, the universe of knowledge has been growing with such accentuation and multi-dimensionality, that it has become increasingly difficult to comprehend, analyse and reason over any specific arena of knowledge activity, which we term as a domain. Such an unswerving growth can be primarily attributed to the various modes of formation and growth of subjects (Ranganathan 1967; Ranganathan 1972; Satija et al. 2014). Therefore, there has always been a dire need for systematic development, visualization and evolution of domain knowledge. Several research streams have analyzed the developmental and organizational intricacies of domains from different vistas, including bibliographic knowledge organization systems, semantic knowledge representation paradigms like descriptive ontologies (Giunchiglia et al. 2006), cognition-based knowledge cartography (Okada et al. 2014), domain analysis as in information science (Hjørland and Albrechtsen 1995; Hjørland 2002) and network-theory based information visualization (Spence 2014). Investigating and organizing domains has remained the forte of GLAMs (galleries, libraries, archives and museums) ever since their inception. Especially speaking of libraries and their custodians, organizing knowledge resources and providing access to them through a subtle combination of knowledge organization tools has always been a passion and a challenge. Bibliographic knowledge organization systems (KOSs) mostly work in a mix-and-match fashion—term lists (such as authority files, glossaries, etc.) enforce standardized terminology with their definitions within a domain, classification and categorization schemes impose a hierarchical (broadly, taxonomic) backbone to the key concepts within and between knowledge domains, relationship lists (such as thesauri, classification ontologies, etc.) facilitate the intra and inter semantic linkages between those concepts in knowledge domains and cataloging codes and metadata standards complete the cycle by providing access to the knowledge resources (Joudrey et al. 2015). It is clear that bibliographic KOSs are principally focused on document description, indexing and retrieval. Knowledge representation (KR) formalisms, on the other hand, accords an entity-centric formalization to knowledge domains, in varying specificities (Giunchiglia et al. 2014). Descriptive ontologies (Giunchiglia et al. 2006), in particular, exhibit characteristic features such as explicit enumeration of semantic linkages, shared conceptualization, reusability, real-time computability, inferential and representational competence and logic-based constraint formalization of a domain. Knowledge cartography— the academic discipline and applied study of different realizations of intellectual landscapes in a domain—has been for decades a major moving force in effective depiction of knowledge. It boasts of a host of powerful knowledge mapping and sensemaking techniques like concept mapping, argument and evidence mapping, issue mapping and web mapping (Okada et al. 2014). Its cognitive foundations facilitates in securing concrete representations of abstract conceptualizations in a domain, envisaging novel scenarios of visual planning, reasoning and memory, and theorizing meaningful configurations for mapping knowledge domains. Domain analysis is a transdisciplinary approach of singular importance in information science, wherein knowledge domains are ideated as "thought or discourse communities" (Hjørland and Albrechtsen 1995). It is noteworthy for its cooperative consideration of perspectives such as knowledge organization structures, synergy patterns, linguistic and communication disposition, psychological proclivity, relevance criteria and information needs, in interpreting knowledge domains. Its embodiment of three diverse paradigms—social, functionalist and philosophical-realistic—contributes to its transdisciplinary anatomization of knowledge domains. Lastly, the research arena of information visualization explores knowledge domains from the visual analytic perspective, being rooted in diverse foundations such as digital sociology, mathematical network theory, graph analytics, creative design and human computer interaction (Spence 2014). Contributions to domain visualization from the information science community chiefly involve statistical and visual metrics of the likes of link analysis and prediction, longitudinal mapping and meta-modelling of knowledge. Further, computational techniques like dimensionality scaling and reduction, cluster analysis and self-organizing maps are also valued processes in mapping and analyzing knowledge domains. Computing centrality, similarity, appropriate paths and detecting communities imparts the algorithmic prowess to domain information visualization.

It is true beyond doubt that each of the aforementioned research streams, inclusive of their techniques, metrics and intuition, are well informed to provide interpretation of knowledge domains in their own distinctive manner. But, each of them in singularity, are severely limited in modelling knowledge with diversity and multi-dimensional scalability, i.e., with high volume, velocity, variety, validity, veracity, volatility and value. The present work attempts to address some of these shortcomings by presenting, *ab initio*, a step-by-step conceptual domain development methodology using knowledge graphs, rooted in the rudiments of knowledge organization and knowledge cartography. It has the following contributory objectives:

- the methodology accommodates unbridled scalability (hence, large scale) in terms of diversity of knowledge resources, especially when domains are becoming increasingly intricate and facetized as a result of information explosion and big data
- it is knowledge intensive, i.e., it employs concepts and procedures from research fields like knowledge organization and knowledge representation, both of which aim to inject semantics and contextuality in domain-based knowledge processing systems, thus aiding in complex querying and deep reasoning-inferencing
- to propose a transdisciplinary, step-by-step conceptual methodology for domain development, encompassing foundational cardinals from cognitive science, information science, knowledge representation, database theory and knowledge cartography
- to identify the various hotspots within the broad arenas of knowledge organization and knowledge management, where the proposed methodology can contribute the most
- to briefly describe the appropriateness of the proposed conceptual domain development methodology in exposure to real, domain-focused literature and datasets
- to delineate the various lines of research in knowledge based systems that emerge out of the proposed methodological analysis, and to provide impetus to research in semantic knowledge management systems in general.

The remainder of the paper has been structured as follows: Section 2.0 reviews the state of the art methodologies and models revolving around domain knowledge organization and domain development; Section 3.0 introduces the step-by-step conceptual domain development methodology in details, including motivational analysis of the design decisions made at each step; Section 4.0 includes a brief case study of the implementation of the methodology on real-time business domain data and related literature; Section 5.0 illustrates the research ramifications, originality and the limitation of the work; and, Section 6.0 concludes the paper by summarizing the observations of the entire work and particularizing the future research proposals.

2.0 State of the art

To date, there has been a considerable number of studies, both foundational and applied in nature, examining the various concepts, techniques, methodologies and issues associated with domain knowledge organization systems and domain development, in general.

Firstly, let us consider some of the knowledge organization systems available in the health domain. SNOMED-CT (http://www.snomed.org/) is an all-inclusive controlled terminology reference relating to clinical data. International Classification of Diseases (ICD) (https://icd.who.int/en) is a foundational global standard for diagnostic health information management. The Diagnostic and Statistical Manual of Mental Disorders (DSM) (https://www.psychiatry.org/psychiatrists/practice/dsm) is an internationally recognized standard for structuring knowledge related to mental health and psychiatry. The Medical Subject Heading (MeSH) (https://www.nlm.nih.gov/mesh/meshhome.html) thesaurus is another such controlled vocabulary system focused on biomedical and health knowledge management. In astronomy, the Three Kingdom System (https://www.isko.org/cyclo/3ks) attempts a coherent organization of the key conceptual entities existing in the vast expanse of our universe. The Mathematics Subject Classification (MSC), maintained by the American Mathematical Society, structures all knowledge within the domain of mathematics through an alphanumeric ordering scheme (https://zbmath.org/classification/). Similar knowledge organization schemes exist for the physics domain (PhySH-Physics Subject Headings; https://physh.aps.org), museum domain (Nomenclature for Museum Cataloging; https://www.isko.org/cyclo/nomenclature) and for economic and business activities (ISIC-International Standard Industrial Classification). Similarly, there are organizational schemes for other domains that cannot be discussed in detail, due to restriction of the scope of discussion.

Ontologies and ontology development methodologies have also been key architectures in conceptualizing, formalizing and developing knowledge domains for semantics embedded information systems. IDEF5 ontology description and capture method (Benjamin et al. 1994) remains one of the first ontology development methodologies, wherein domain objects are uniquely identified by their distinct properties, and complex relationships between them are captured in alignment with human conceptualizations. Uschold and King (1995) proposed a step-wise ontology development approach having the following stages: purpose identification, ontology building (including capture, codification and integration of existing ontologies), evaluation and finally documentation of ontologies. The TOVE enterprise modelling approach proposed by Grüninger and Fox (1995) improvised upon its predecessors in domain development by incorporating motivational scenarios, informal and formal competency questions, and first-order logic based axiom specification within its framework. METHONTOLOGY by Fernández- López, Gómez-Pérez and Juristo (1997) metamorphoses the art of building ontologies from scratch into a well-ordered, systematic engineering process, grounded in ontological life-cycle principles. Noy and McGuinness (2001) proposed Ontology Development 101 as a practical knowledge engineering methodology especially geared for development of domain ontologies. Prieto- Díaz (2003) proposed an ontology development model based on reusing faceted classification principles and the domain analytic paradigm from the field of library and information science. More recently, YAMO by Dutta, Chatterjee and Madalli (2015) developed a methodology for constructing large scale faceted ontologies inspired by canonical principles of S. R. Ranganathan's analytico-synthetic approach in library classification.

Knowledge cartography as a research area involved in mapping domains from different perspectives is less heard of but reserves its own importance. Okada, Shum and Sherborne (2014) interpret knowledge cartography as a specific engineered form of information visualization, collaboratively and collectively spatializing an arena of knowledge activity- domain(s)—which otherwise have no intrinsic spatial expression. They further catalogue the different genres of knowledge cartography such as mind mapping, concept mapping, argument and evidence mapping, issue mapping, web mapping and ontologies. Concept mapping, inclusive of its theoretical foundations and software implementations, has been discussed at length in Cañas and Novak (2007). Zeiliger and Esnault (2009) elaborate on the networked and gradational construction of formalized knowledge, using web mapping and validates it with an appropriate case study. Selvin (2014) sheds light on some of the special skills, concepts and frameworks required in construction of knowledge maps and for further collaborative analysis. The integrated mapping of data, metadata, interpretations and intellectual landscapes in a specific domain can cohesively act as a supporting tool for concept analysis, qualitative research and collaborative knowledge modelling strategy (Sierhuis and Shum 2014; Vasconcellos 2014; Basque et al. 2014). Skupin (2004) asserted that cartographic representation can provide unique insights on knowledge domain visualization. Plantié and Riccio (2010) investigated the convergence of knowledge cartography and social network representation. Concept maps have also been applied as visual interfaces to digital libraries (Shen, Richardson and Fox 2003). Even within the information retrieval community, in the domain visualization arena, knowledge maps have been generated and navigated through to pursue required information needs (Shiffin and Börner 2004; Chen 2003).

Knowledge graphs is the go-to solution for populating, reasoning and visualizing knowledge domains in recent semantic information systems. Ehrlinger and Wöß (2016) explored the various definitions of knowledge graphs, consolidating its various interpretations and common attributes. The architecture of knowledge graphs has also been progressively analysed through a DIKW hierarchy (DIKW: data-information-knowledge-wisdom; Duan et al. 2018). Qingjie et al. (2014) focus on powerful domain knowledge graphs generated by multi-source information fusion and multidimensional relationships. Nickel et al. (2015) review relational machine learning in the context of knowledge graphs and examine how statistical models can be trained on knowledge graphs to predict new facts about the representative domain. Property graph data models have also been developed to streamline and facilitate the process of knowledge management (Zhang 2017). Melo and Paulheim (2017) propose an algorithmic approach to refine noisy and incomplete knowledge graphs. There have also been proposals of synchronous knowledge graphs, generated using unification of heterogeneous IoT (internet of things) data sources using linked data (Le-Phuoc et al. 2016). Grainger et al. (2016) propose the semantic knowledge graph, capable of traversing and ranking any relationship within a domain, primarily with the help of an inverted index and a complementary un-inverted index. Singhal (2012), introducing the very essence of the Google knowledge graph, adds that due to its semantic enhancement functionality, it is able to understand and search for things, not strings. Chah (2018) attempts to understand the underlying ontology behind Google's knowledge graph. Algorithmic frameworks and techniques to mine meaningful insights from large scale graphs has been elaborately discussed in Aridhi and Nguifo (2016). Deshpande et al. (2013) discusses, in detail, the building up, maintenance and usage of knowledge bases. Bagchi (2019) conceptualized a knowledge architecture grounded in the emerging technological paradigm of knowledge graphs. Finally, Krötzsch et al. (2018), in their landmark paper on attributed description logic, enriched description logic with a new kind of attribute-value pair termed "annotations" that allowed to express constraints on them—for achieving increased expressivity while reasoning with knowledge graphs.

Domain analysis as a transdisciplinary paradigm to analyze knowledge domains and elicit insights (intuitive or empirical) from such analysis has emerged as a key approach in information science with a strong research backbone. Hjørland and Albrechtsen (1995) introduce domain analysis as a novel approach in information science, situating the study of "knowledge domains as thought or discourse communities." It brings to fore the foundational perspective of information science as a discipline, which is as social as it is intellectual. It continues describing in-depth the various transdisciplinary facets in theorizing and understanding knowledge domains (Hjørland 2017). Hjørland (2002) introduces eleven landmark approaches to domain analysis, ranging from special classifications and discourse studies to artificial intelligence. Smiraglia (2015a, 2015b) ideates domain analysis as a methodological conglomeration towards "identifying a specified knowledge base," interprets it in terms of ontology extraction and carries out an informetric analysis to characterize its research vibrancy within knowledge organization and associated research streams.

3.0 Methodology: description and motivational analysis

Figure 1 (see below) diagrammatically presents the proposed conceptual domain development methodology using knowledge graphs, rooted in the rudiments of knowledge organization and knowledge cartography. The steps are listed as follows:

1. Knowledge space
2. Knowledge analysis (comprising mental models and facet analysis)
3. Knowledge manoeuvre (i.e., knowledge base formation)
4. Ontology development
5. Knowledge graph population
6. Knowledge validation
7. Reasoning-inferencing
8. Knowledge visualization (founded on knowledge cartography cardinals)
9. Knowledge evolution

In what follows, each step of the conceptual methodology is described in great detail, including motivational analysis and illustrative examples as and when needed.

3.1 Step 1: knowledge space

This step of the methodology in part, and the next step "knowledge analysis" are intuitively inspired by the concept of idea plane as enunciated by S. R. Ranganathan in his theoretical basis of Colon Classification (Ranganathan 1967; Ranganathan 1987), the analytico-synthetic classification scheme he devised. The idea plane mandates rigorous thinking and fine-tuning in terms of decision making and design choices. The essence remains that the quality of analysis done at this plane determines the quality of the rest of the design architecture. Following such a notion, the step prescribes finalizing two crucial determinants, which together constitute

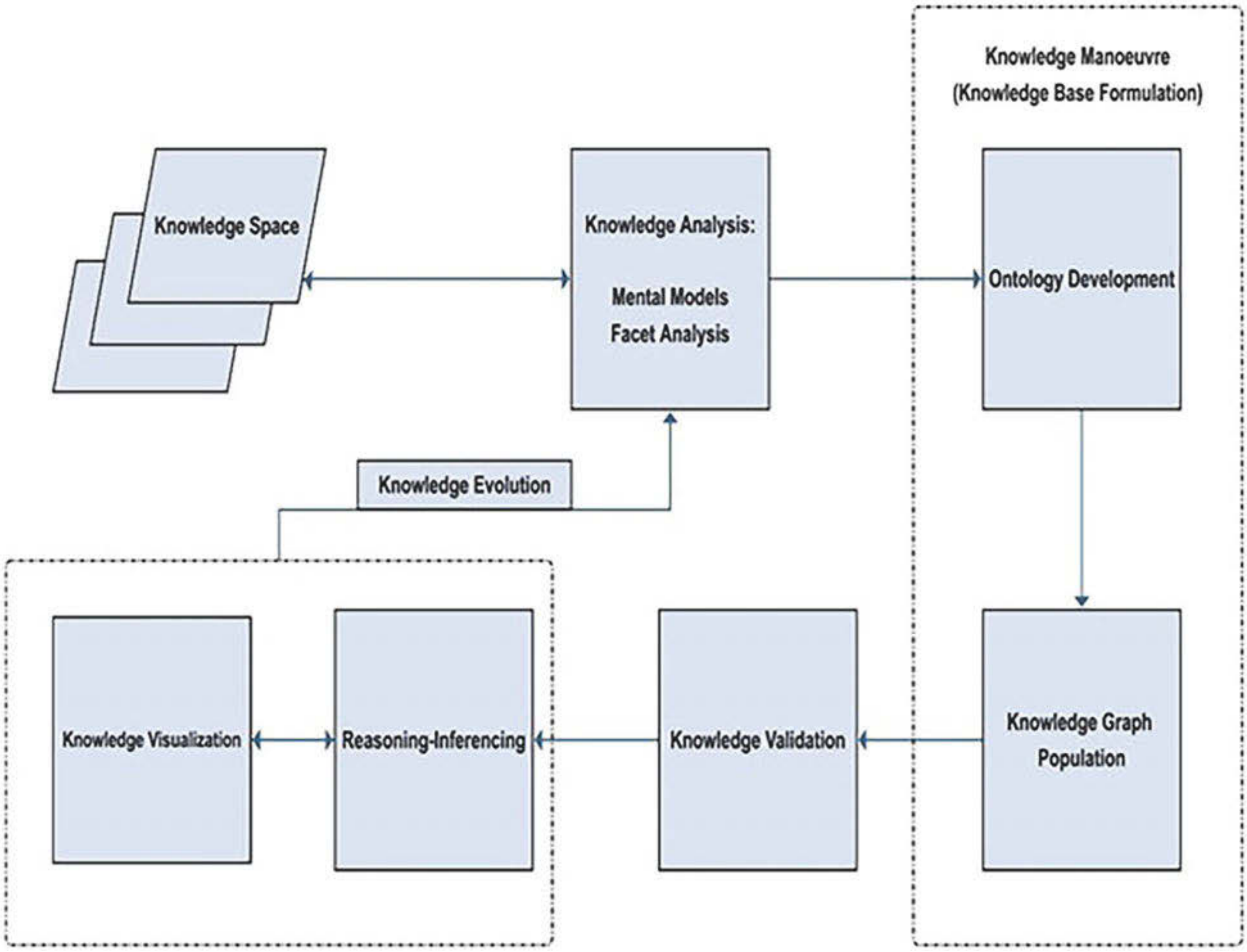

Figure 1. Domain development methodology.

the knowledge space: fixing the interpretation of what is meant by a domain and recommending the domain knowledge resources needed to be collected for future analysis.

Knowledge space, in the context of the methodology, is an arena of knowledge activity that is also known as domain in the parlance of formal sciences. The interpretation of the domain has been kept limited to that of an application arena (like healthcare records management, consumer complaint management within the larger business domain etc.) or those relating to communities (such as pandemic containment, earthquake response and recovery, etc.). The steps of the methodology have been found unsuitable for very broad interpretation of domains (such as considering a discipline like chemical sciences or physical sciences to be a domain), and such limitations will be addressed appropriately with reasons.

The domain knowledge resources to be aggregated for the succeeding steps are apportioned into two different streams, which work in sync and inform each other: namely, domain specific literature and domain specific datasets.

– Domain literature should compulsorily include encyclopedias and dictionaries since they provide prefatory knowledge focused on the domain. They should also refer to trend reports collated by institutions having authority in the domain in order to gain insight about the issues or topics trending in recent times.

– Domain specific datasets engineered out of real-life scenarios and projects are another form of valued domain knowledge resources. The instances with values of their characteristics (rows) and the attributes (columns), in conjunction with domain literature, is carefully observed so as to conceive an intuitive understanding of the domain.

3.2 Step 2: Knowledge analysis

As previously mentioned, the very conceptualization of this step is based on the notion of idea plane, which fine-tunes the overall quality of any methodology or system design.

Knowledge analysis deals with an exhaustive, theoretical identification and codification of the core conceptual underpinnings of the domain as innately captured within the domain knowledge resources in the previous step. The step, in the context of the overall methodological framework, can be construed as a sphere of discourse or a universe of study deliberating over a set of entities, variables, attributes and relationships, reined in by a set of common minimum constraints and scopic manoeuvres. Ingetraut Dahlberg, the pioneering German information scientist and philosopher, perceived knowledge as a shared commodity with a purpose, which is neither raw nor unattached to human perception (Dahlberg 2006). She further explicated the very essence of the conceptual knowledge analysis in the following four ways: knowledge units (essentially the concepts), knowledge elements (reflective of the various characteristics of each concept), larger knowledge units (aggregations of hierarchically lower concepts) and knowledge systems (formed by planned and systematic cohesion of knowledge units). Patrick Wilson, noted librarian and information scientist, had a more broad conceptualization of knowledge analysis as cohesive of two concept spaces: descriptive, dealing with knowledge representation and organization, and, exploitative, pertaining to knowledge reasoning and discovery (Wilson 1968).

The uniqueness of this step lies in the fact that it is aided by the novel dual of facet analysis (Broughton 2006; Vickery 1968) and mental models (Johnson-Laird 1983)—the two respected paradigms from conceptual analysis and cognitive mapping working in intuitive coherence The step also aids in making a well informed distinction between essential and non-essential attributes (specific to datasets in the knowledge space) to be considered for developing the upcoming ontological schema. But before localizing their overarching significance in our methodology, it is worthy to perform another ancillary step, which can reveal a preliminary structural understanding of the knowledge space. A lightweight exploratory data analysis (EDA) (Chatfield 1995) should be carried out on the domain datasets in order to get a hold of the dimensions they encompass, thus contributing towards the much validated data-driven knowledge modelling. It is also important to note that in case of very big datasets, the methodology also recommends conducting a data wrangling exercise using standard procedures.

3.2.1 Mental models

The cognition inspired theory of mental models (Johnson-Laird 1983) is being embraced for the first time ever by any domain development methodology, or to speak more broadly, by any semantic knowledge management architecture. Being a key component of human cognitive apparatus, mental models are defined as the cognitive ability of forming mental representational networks from observing interactions in a domain of discourse. They greatly aid in learning, developing and conceiving explanatory and predictive knowledge by intuitive perception of a knowledge domain, ultimately accentuating conceptual development and cognitive reasoning-inferencing (Vosniadou 1992; Vosniadou 2002). Scientific research in cognitive psychology and cognitive neuroscience have shown encouraging endorsement for practical employment of mental models (Kosslyn 1994). The following characteristics attempts at a fuller understanding of the intricate theory of mental models:

- the theory provides ontological, epistemological and metaconceptual awareness of the salient concept components of the knowledge domain, while simultaneously mediating revisions and constructions of changing conceptualizations
- mental models performs the role of perceptual vehicles through which implicit knowledge of a domain becomes explicit, reigned in by prior beliefs and assumptions
- mental models exhibit autonomy and dynamism; contribute in interpretation of information in perceptual information processing and fixing incoherency in the cognitive knowledge base
- mental models uphold the commonalities amongst several possible models of comprehension, and is rooted in mental logic constructs such as counterexamples and abduction (ability to formulate explanations)

The proposed methodology, in particular, utilizes the conceptual tool of generative questions from the theory of mental models (Vosniadou 2002). It adds to the theory, especially in the context of its application to domain development, the innovation termed as whiteboard conceptual model.

- Whiteboard conceptual modelling involves informal, graphical domain modelling wherein key conceptual entities in a domain are drawn as labelled nodes, interconnected through semantic relationships, and with both nodes and relationships having attributes. The whiteboard conceptual model follows as a direct consequence of the mental model of the domain as conceptualized by the domain expert, and requires as a prerequisite considerable understanding of the knowledge space and its associated preliminary EDA. Free and open source lexico-semantic resources (http://compling.hss.ntu.edu.sg/omw/summx.html) can also be consulted at this sub-step towards enriching the domain model.
- Generative questions (Vosniadou 2002) are utilized at this stage of the methodology for its informing approach towards improving the whiteboard conceptual model. These questions are such that they are quite difficult to

answer on the basis of known or stored background domain knowledge, but require genuine innovative cognition and solution approaches. In other terms, generative questions are dependent on the dynamism of mental models and cannot elicit readymade solutions. The generative questions formulated at this stage challenges the existent mental model of the domain and improves it. It is also important to note that generative questions are for improving the conceptual development of the knowledge domain and not focused on evaluating the competency of the final knowledge artifact.

3.2.2 Facet analysis

Facet analysis is construed as one of the overarching theoretical approaches in knowledge organization (Broughton 2006; Vickery 1960). Broadly speaking, it has its basis in the principle of logical division (https://www.isko.org/cyclo/facet_analysis) aided by a unique train of characteristic features, which groups together domain isolates (smallest possible unit of domain knowledge) into domain facets (which are conceptual generalizations of domain isolates) each of which further belongs to one amongst an agreed upon set of domain categories (fixing the order of these categories under which the domain facets are assigned should depend on domains and the level of formalisation like lightweight or highly formal). In our methodology, after the whiteboard conceptual model is agreed upon after enrichment through cycles of generative questions (the number of cycles depends on domain appropriateness and necessity), a facet analytic study is recommended. Depending on the domain specific conceptual analysis done till now, a certain set of domain facets (classes of entities) are framed and their ordered sequence is determined (if needed). Any freely available lexico-semantic resource (like, WordNet) can be consulted at this stage for a better visualization of facets, hierarchies and for augmenting their standardization and shareability. The Whiteboard Conceptual Model, which is a completely intuitive semantic model, is standardized through the means of these domain facets and re-drawn, thus giving an intuitive network an informal character (informal, in the sense of machine processable formalization). It is to be noted that this sub-step of facet analysis is guided throughout by the canonical principles of knowledge organization as postulated by S. R. Ranganathan (it is to be noted that these canonical principles are general in the context of knowledge organization, and can be applied beyond the analytico-synthetic paradigm).

3.3 Step 3: knowledge manoeuvre (knowledge base formation)

The need for formalisation of the informal analysis of the previous step into a formal knowledge base informs the present step of the methodology—knowledge manoeuvre—comprising two sub-steps, namely ontology development and knowledge graph population. It is the first step in the proposed domain development methodology that is dependent on software components, and a non-negotiable condition in this step is the use of free and open source software tools. The recommended software platforms for engineering the entire knowledge manoeuvre stage includes: either open-source ontology editors or open-source NoSQL graph databases, and the choice of the open source software platforms must be grounded on the premise of information interoperability (which implies that either of them can be utilized towards creating the knowledge artifact required by the end of this step, and it should be interoperable with the other software platform). Leading open source ontology editors like Protégé (https://protege.stanford.edu/) and open source graph databases like Neo4j Community Edition (https://neo4j.com/licensing/) fulfill this requirement, besides many other similar open source software platforms.

Ontology development: A formal, descriptive domain ontology (Guarino et al. 2009) is developed at this sub-step of knowledge manoeuvre out of the robust, informal domain model finalized during knowledge analysis. An open source integrated ontology development environment (local or cloud-based) or an open source graph database is highly preferred for developing the ontological schema. Following the enriched whiteboard conceptual model informed by generative questions and facet analysis, class hierarchies, relationships, data-attributes and axiomatic constraints are created and formalized into an ontological model of the domain. The choice of the level of formalization of the ontological schema (whether lightweight or highly formal) is left up to the expertise of the knowledge engineer, which further depends on the target objectives (like computational efficiency, rich visualization etc.) of the entire domain development process.

Knowledge graph population: This stage of the knowledge manoeuvre step consolidates the formation of the knowledge base through population of the ontological schema developed in the previous sub-step with representative entries (meaning, instances representative of the classes formalized). The requirement of software components is, again, paramount in this sub-step. In case of the usage of ontology editors, there are two options of populating the schema: i) population of the ontological schema manually, one by one; and, ii) in case of big datasets with lot of instances, plugins for ontology editors (Cellfie Plugin [https://github.com/protegeproject/cellfie-plugin] for importing spreadsheet data, in case of Protégé) are available for doing the task. In case of open source graph databases, their query-cum-modelling language (like CYPHER for Neo4j; https://www.opencypher.org/) is sufficient for populating the ontological schema (graph databases can handle big data with ease; for example, Neo4j can easily handle trillions of nodes in a semantic relationship graph, and up to a quadrillion, theoretically).

It is also important to explain the reason behind anchoring the design choices made in the knowledge manoeuvre step on information interoperability cardinals. The open source software platforms should be chosen in such a way that if required, a knowledge graph produced by the ontology editor can be easily imported into the graph database for knowledge processing, without any structural or functional disorientation, and vice-versa. Neo4j Community Edition has an exclusively developed plugin—NSMNTX (https://neo4j.com/docs/labs/nsmntx/current/)—that facilitates usage of RDF data in Neo4j graph databases for purposes as diverse as RDF import-export, model mapping and reasoning-inferencing.

3.4 Step 4: knowledge validation

To ascertain the overall veracity of the knowledge artifact as formalized in the knowledge manoeuvre step as well as to provide another chance at improving the knowledge model, the logical consistency of its TBox-ABox-RBox is checked, and the resulting populated graph is validated following state-of-the-art domain validation techniques and services. The outcome of this stage is a semantically enriched knowledge base, ready for querying, reasoning-inferencing and visualization. It is achieved through two successive validation layers, one based on philosophical notions and the other requiring computational prowess.

– Firstly, an ontological analysis of the schema is recommended to be carried out by using the OntoClean methodology (Guarino and Welty 2002). The methodology based on its subtle analysis of meta-properties is primarily drawn from philosophical ontology notions of validating ontological commitments. A developed ontological schema is analyzed against six meta-property criteria: identity, unity, rigidity, dependence, permanence and actuality (the last two being later research additions). The stress here is to appraise the ontology's commitment to formal ontological engineering principles and not to make it very heavyweight in terms of formalization.
– From the perspective of software based validation tools, there exists various dimensions or perspectives from which an already constructed knowledge graph can be refined. It can be refined to predict and add new knowledge to the knowledge graph, i.e., with an intention to make it even more complete. It can also be refined to detect outliers, inconsistent patterns and missing nodes within the knowledge graph, with an aim to make it as much as error free as possible (Paulheim 2016). Relational machine learning (Nickel et al. 2016) algorithms combined with standard data cleaning techniques play a pivotal role in such graph refinement. Further, there are several free, web-based ontology validation tools (such as OOPS!—http://oops.linkeddata.es/), which can be utilized for model enrichment against diverse criterias such as logical consistency, reality representation and issues with knowledge modelling.

3.5 Step 5: reasoning-inferencing:

The methodology also proposes two parallel reasoning-inferencing facilities (semantic reasoning-inferencing engines and graph computing algorithms based analytics), and associated querying and visualization techniques. In fact, this step is informed by and is often simultaneous with the next step of knowledge visualization. Querying the knowledge base is primarily dependent on the query languages, query service plug-ins and protocols provided by the software platform committed during the knowledge manoeuvre step of the methodology (like CYPHER query-cum-modelling language in Neo4j or SPARQL protocol in case of RDF data).

– Semantic reasoners of varying inferential complexity provided as plug-ins with standard ontology editors forms the first category of reasoning-inferencing facility within the methodology. They are mostly based on logic languages (first order logic being the most popular language basis; description logics being a decidable fragment of first order logic). The motivation is to infer new domain knowledge by reasoning existing knowledge that has been explicitly formalized. Ontology editors (like Protégé) provide many such semantic reasoning engines of differing computational efficiency (examples being Pellet, HermiT, Fact ++, etc.). More recently, the limitation of conventional description logics based reasoning with respect to knowledge graphs have given rise to the potential future applicability of attributed description logics, which is still in its infancy and is undergoing serious developmental research (Krötzsch et al. 2018). It supports development of multi-attribute value annotation pairs representative of annotations, thus capturing in the annotated nodes and relationships. Attributed description logics provide a layer of concept names, role names and individual names over and above plain description logics (DLs) to transform it into the type of ontology-based formal semantics aptly suited for reasoning over knowledge graphs.

- Reasoning services employing graph computing algorithms constitute the more advanced and complex reasoning-inferencing paradigm within the methodology (the layer that adds the analytical flavour to reasoning-inferencing and ultimately provides novel insights about domain development). Neo4j in particular, and other open source graph data management software platforms in general, offer numerous graph-computing algorithms for performing powerful graph analytics (Needham and Hodler 2019), reasoning and visualization over the representative knowledge space (domain) it is dealing with. Our methodology efficiently harnesses this unique algorithmic power and provides a strong reasoning-inferencing layer over the consolidated knowledge base. Some of the broad categories of algorithms and its various constituent types heavily utilized by graph data management platforms for analytical reasoning purposes (Needham and Hodler 2019) are as follows:
  - Centrality algorithms, dealing with overall connectivity, ease of information spread and relative importance of distinct nodes in facilitating flow of information in a graph-connected network. Examples include pageRank, betweenness, closeness and harmonic centrality algorithms
  - Community detection algorithms, dealing with the formation and detection of communities/clusters in a graph-network and their tendency to stay together or break apart. Examples include Louvain, label propagation, connected components and clustering coefficient algorithms
  - Path finding algorithms, dealing with making an informed decision regarding choosing the most optimized path in a specific traversal, alongside judging the quality and availability of paths in a graph-connected network. Examples include minimum weight spanning tree, shortest path, A*, random walks and numerous other traversal algorithms
  - Similarity algorithms, ascertaining the degree of similarity of nodes in a graph-based network. Examples include Jaccard similarity, cosine similarity and Euclidean distance based similarity algorithms.

3.6 Step 6: knowledge visualization

For visualizing the knowledge graph, the methodology proposes to leverage semantic-information visualization platforms inspired by the cardinals of knowledge cartography (Okada et al. 2014). The foundational cardinals are briefly explained below as the underlying motivation for this step:

- visualization of knowledge domains is particularly essential in detecting saliency of domains, i.e., distinguishing the knowledge enriched hotspots within the landscape of the specific domain, which have potential to contribute to advanced research in domain specific information seeking and understanding
- visualization simplifies the communication of umpteen semantic intricacies and dependencies in a knowledge domain
- domain visualization, in combination with graph analytics, offers a great opportunity to elicit various patterns in a knowledge domain, depending on the parameters chosen
- knowledge cartography empowers visual analytics such as inferring parametric dimensionality of a domain or communicating knowledge diversity within domains
- it facilitates the cognition to perception cycle, wherein we visually sense our cognitive conceptualization of a domain (of course, after standardization)

Software platforms for engineering this step are widely available as standalone software tools or as plugins interoperable with graph-based knowledge structuring tools. The methodology is flexible in its outlook in this case, and recommends adoption of any of the following knowledge visualization options as per the visualization needs of the specific domain context.

- Ontology editors (like Protégé) have several plugins available to facilitate the visual navigation of the developed knowledge graph (examples: OWLViz, OntoViz, OntoGraf, VOWL, SOVA, etc.)
- Graph databases also offer several options of visualizing domain entities (nodes), relationships, attributes and properties. They usually provide embeddable tools (like Neovis.js in case of Neo4j), embeddable libraries (like the very popular D3.js) and third party tool plugins (example: yFiles etc.) for rich graph visualization (https://neo4j.com/developer/tools-graph-visualization/)
- Graph databases specialized in RDF data manipulation and visualization can also be used for the purpose (like GraphDB Free; www.ontotext.com/products/graphdb/)
- Open source, third-party information visualization platforms (such as Gephi; gephi.org/) also offer support for visual navigation and inferencing of ontologies

3.7 Step 7: knowledge evolution

The final step—knowledge evolution—is inspired by the theme of continuous conceptual recurrence and evolution as enshrined in S. R. Ranganathan's "Spiral of Scientific Method" (Ranganathan 1957). It injects dynamism into the methodology by envisioning the successive steps of top-level ontology alignment, ontology mapping and/or knowledge inclusion and ontology modularization and merging, resulting in an evolved manifestation of the knowledge analysis stage. This step of the methodology is overarching in its conceptual significance, and also practically feasible if there is noteworthy evolution of domain knowledge. In such a scenario:

- the ontological schema developed in the knowledge manoeuvre step should be aligned to a top-level ontology (like DOLCE, IEEE SUMO, UFO etc; the choice of the top-level ontology depends on the expertise of the knowledge engineer), which should remain same for possible cycles of evolution in that specific instance of the domain development methodology
- after ontology alignment, there can be two possible sub-scenarios: i) enrichment of the developed ontological schema using other overlapping ontologies in the same domain through ontology mapping; and/or, ii) inclusion of new domain knowledge entities (like concepts, relationships, etc.) in the face of evolution of knowledge domains
- it also accommodates the idea of knowledge modularity and merging, wherein domain knowledge models can be modularized into perspective-aware patterns (possibly through ontology modularization techniques based on algebraic logic) and an appropriate set of such domain knowledge patterns be algebraically merged to form a more wholesome domain knowledge model (the use of ontology modularization and merging techniques in this case remains paramount). For example, in case of a community information service focused domain such as pandemic information management, an already developed domain knowledge model using the proposed domain development methodology can be further enriched, if newer perspectives, in the form of perspective-specific ontology fragments from similar ontologies be integrated.

4.0 Brief highlights of a case study

It is needless to mention that the proposed methodology is conceptually rigorous and robust, and has been practically implemented on business domain data (consumer complaints data, to be specific), which attests its evaluatory effectiveness. The principal focus of the paper remains the presentation of the synergy between the steps of the conceptual domain development methodology, its motivational analysis and its seamless integration with latest available semantic technology-components. Towards that objective, some highlights of the practical implementation are briefly mentioned below, and a fuller discussion with illustration of the practical implementation remains a subject for a future technical paper.

In the case study on consumer complaints domain, datasets representative of complaints about consumer-specific business entities and interactions was consolidated from the online, open access data catalog of the data.gov platform. For preliminary, prefatory knowledge about the domain, encyclopedias like the *Encyclopedia of Consumer Culture* (Southerton 2011) were consulted. Initially, in general, the datasets can be in any standard file format. As per the technical compatibilities of the chosen software platform for the case study—Neo4j Graph Platform Community Edition—it had to be converted into the CSV (comma separated-value) format, for the conversion of which numerous software libraries made available in different standard programming languages (like Python, R, etc.) to be utilized. Further, the CSV files should be made to undergo the process of data cleaning, which screens, diagnoses and edits outliers, inconsistent patterns and missing data. The informal knowledge analysis comprising the complementary sub-stages of mental models and facet analysis were carried out as conceptually explicated in the description of the methodology above. The knowledge manoeuvre and the knowledge visualization steps of the methodology specific to the implementation are briefly illustrated below, whereas an implementational discussion of the knowledge validation and the reasoning-inferencing steps is not aligned with the focus of the brief highlight.

The required information from the standardised CSV file was loaded along with the headers, which annotates the loaded information with context. It was read and loaded line-wise at a very fast speed, and, as needed, the output of the contextual loading can be limited to a specific number of lines (for example 100) for validating the correctness of the steps performed. From the loaded information actualized with the “LOAD CSV” command, Neo4j allows one to create ontological schema and knowledge graphs out of it, leveraging the intrinsic powers of the CYPHER query language, and allows one to semantically relate them together. A very simple illustration-cum-explanation of an use case in the consumer complaint domain using CYPHER scripting is given below:

```
“USING PERIODIC COMMIT
LOAD CSV WITH HEADERS FROM
‘file:///CC.csv' AS line WITH line
CREATE CONSTRAINT ON (c:Complaint)
   ASSERT c.id IS UNIQUE;
CREATE CONSTRAINT ON (c:Company)
   ASSERT c.name IS UNIQUE;
CREATE CONSTRAINT ON (r:Response) ASSERT
   r.name IS UNIQUE;
CREATE (complaint:Complaint { id: TOINT ( line.
   ‘Complaint ID`)})
MERGE (company:Company {name: UPPER(line.
   Company)})
```

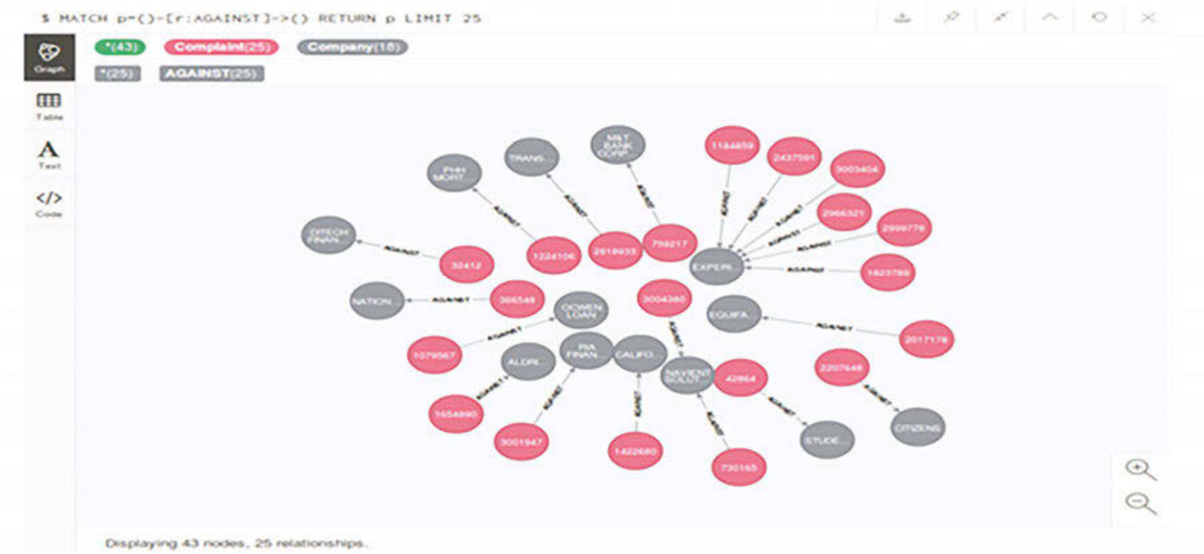

Figure 2. Knowledge graph generated for the complaint-AGAINST-company relationship.

MERGE (response:Response {name: UPPER(line.
  'Company response to consumer')})
CREATE (complaint)-[:AGAINST]->(company)
CREATE (response)-[r:TO]->(complaint)
SET r.timely = CASE line. 'Timely response' WHEN
    'Yes' THEN true ELSE false END,
    r.disputed = CASE line. 'Consumer disputed'
    WHEN 'Yes' THEN true ELSE false END;"

(The above CYPHER script is a small portion of a body of much larger script running into several pages in length. It is just an indicative glimpse of the process of implementation and not the focus of the paper. Further, the specific terms within the script should not be given precedence over the conceptual methodology).

At the outset, the contextually annotated CSV file (here, CC.csv) is loaded line-wise in the software platform. The "USING PERIODIC COMMIT" clause directs the loading mechanism in Neo4j to perform a commit after a specified number of rows, in case the file contains millions of records (i.e., big data), facilitating high scalability and powerful performance. Then constraints are being created on nodes (reflective of entities in a domain; here complaint, company, response) asserting that the particular node is unique and can act as an identifier in required operational circumstances. Since Neo4j by default loads fielded-data from CSV as strings, it is imperative in this scenario to convert some fields into their natural data-types ("Complaint ID" as integer). "MERGE" is one of the unique functionalities offered by the CYPHER query language. It is a conceptual aggregation of the functionalities offered by "CREATE" and "MATCH" command. It searches for a given pattern or a node in the knowledge graph. If it already exists, then merging of the incumbent occurs with the already existing node/pattern and an output is returned. If it does not exist, then a new node/pattern is created with appropriate labelling and returned as an output (in this case, instances of companies and responses are either being created anew or merged with the ones already existing). Then relationships are built using the "CREATE" command specifying the type of the relationship and its semantic direction. In this example code, two relationships are being created—one is the complaint against a company, and the other is the response accorded to that complaint (see Figure-2 and Figure-3). Further, a condition is established using the "SET"" command to determine which response can be classified as a "timely response" and which is a "consumer disputed" one. The visualization follows subsequently.

The above illustration is a very simple use case within the business domain of modelling knowledge about companies, complaints and responses. It can be further connected to numerous other knowledge graphs within the same knowledge space to form a semantically rich knowledge structure capable of complex inferencing. Neo4j is efficient in modelling domains that are much more complex and vast in nature.

5.0 Research implications

The proposed conceptual domain development methodology is expected to have numerous research implications and practical ramifications in the broad arena of domain specific knowledge organization (KO) and knowledge management (KM) systems. It shall contribute as a significant methodological, technological and structural augmentation in research fields like information resource discovery sys-

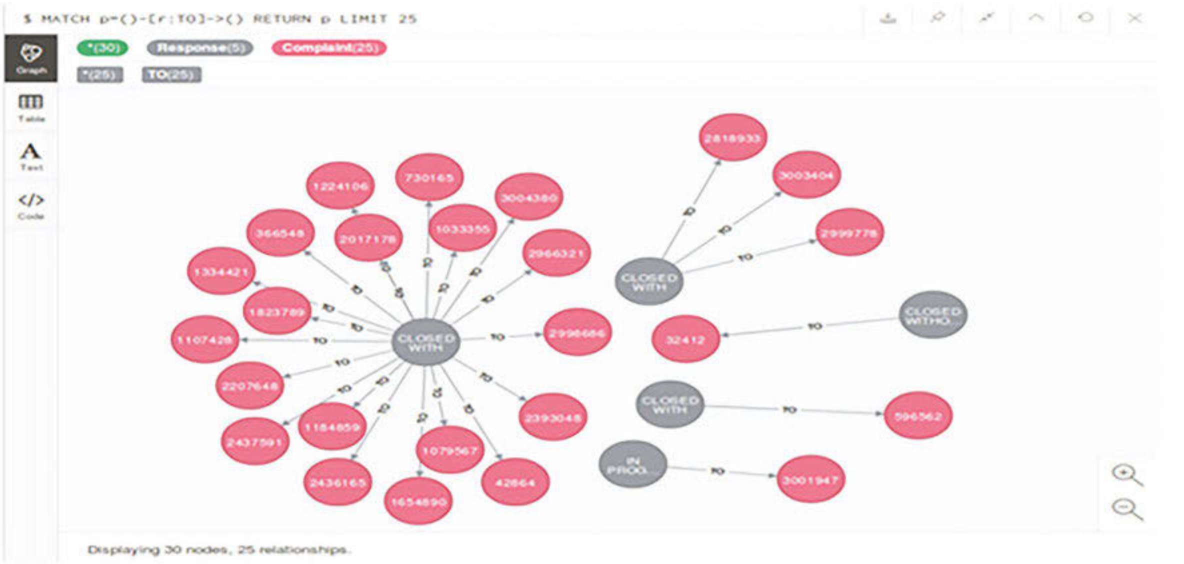

Figure 3. Knowledge graph generated for the response-TO-complaint relationship.

tems, digital repositories, question-answering services and enterprise knowledge management systems, to name a few. Some of the possible research implications are highlighted as follows-

5.1 Digital repositories

Digital repositories are online repositories of digital information objects that provide facilities to organize, annotate, search and retrieve the content included in its collection (Prasad and Madalli, 2008). The content included in the collection of a digital library can further be born digital or digitized, and, accordingly different strategies are adopted to make them organized, sustainable and interoperable. Some of the common characteristic features of a digital library include multiple user-access, easy retrieval of information, user-friendly interfaces, relatively large storage space and value-added content. Knowledge graphs, if incorporated within the technical architecture of next-generation digital repositories, can provide technical competencies like semantically enriched information annotation and organization, visually appealing UI, super-fast search and retrieval, and richer recommendations due to its inherent graph-structured linked data.

5.2 Information resource discovery services

An information resource discovery service, as its name reflects, is an integrated online searching interface for locating any and every resource a particular library has access or subscription to physical library resources, online library-subscribed resources and publisher metadata, open access resources and even licensed resources—made possible through a combination of pre-assembled centralised indexing, unified searching interface and consolidated results (which is different from the federated search architecture). The primary aim of developing such a discovery service is to aggregate and present all the varieties of resources a particular library possesses using a unified search functionality, thus satisfying the clientele's interdisciplinary needs and exposing rich local resources. Marshall Breeding in a landmark paper on "The Future of Library Resource Discovery" (Breeding 2015) elucidated the technical requirements a futuristic discovery system should characterise. Some of the ways in which the proposed knowledge graph-based domain development methodology can help in developing future library resource discovery services are as follows-

- The technical functionalities offered by the semantic technological stack (vast storage, domain visualization, highly scalable indexing, rich semantic annotations, etc.) can be utilized in developing a futuristic large-scale open access global information resource discovery platform (such as a far more efficient open science cloud platform)
- Connected data is at the core of this knowledge graph-based methodology, and, they can naturally provide coverage and harvesting facilities to open-linked library data
- Relational machine learning, especially tailored for knowledge graphs, can aid in conducting gap analysis of the discovery functionalities
- Identity and access management (IAM) and data privacy implementations by the knowledge-intensive semantic platforms can be leveraged to provide a secure and robust data infrastructure to resource discovery platforms
- Graph analytic techniques, offered within the proposed methodological framework, can be utilized to quantita-

tively picturize the performance of a discovery system and can also be used for developing alternative metrics (altmetrics) based assessment for library resources.

5.3 Enterprise intellectual infrastructure

Enterprise intellectual infrastructure can be considered as the entire gamut of processes, technologies and resources that an institution or an organization or a system utilises for producing, accomplishing and maintaining intellect, knowledge and expertise. It is of utmost practical importance in the arena of future knowledge-intensive enterprise administration, and, at a bigger scale, in effectively managing large, futuristic enterprise networks. Knowledge-graph based domain knowledge organization and management systems, though technical in nature, have numerous general ramifications that can positively catalyse the development of future enterprise administration frameworks. Hiring, training and mentoring policies will change for enterprises as professionals with different skill-sets (both technical and non-technical) will be required for effectively managing these futuristic enterprise KM systems. The conventional managerial policies regarding knowledge management as a whole and performance feedback will also have to be reframed. Further, the tools and methodologies that are currently in popular use in enterprises have to be overhauled, and the data has to be effectively migrated and linked as per the developed standards.

It is also important to note the research value of the proposed domain development methodology in order to foresee its possible research implications in the implementational areas identified above. The following points, very briefly, attest to the original value of the conceptual methodology:

- the conceptual interplay between facet analysis and mental models enriching the knowledge analysis step
- the platform-flexible and platform-interoperable character of the proposed methodology
- a parallel layered reasoning-inferencing mechanism (mathematical logic based semantic reasoners and graph algorithm powered reasoners)
- visualization grounded in the foundational cardinals of knowledge cartography
- the entire methodology, including the knowledge evolution step, conceptually grounded in knowledge organization principles

A limitation of the present study lies in the fact that a semi-automatic interpretation and implementation of the proposed domain development methodology has not yet been attempted.

6.0 Conclusion and future work

The work, in alignment with its stated contributory objectives, proposed a conceptual domain development methodology that has characteristic features like scalability, knowledge- intensivity, context-dependency, methodological transdisciplinarity and scopic domain specificity. Before concluding, it would be relevant to note that though knowledge-graph based domain development systems require more technical soundness and research, they have boundless powers to induce a tectonic shift in the way domains are conceptualised, visualized, analysed and reasoned upon. It also has the ability to effect a positive change in the existing technological culture and practices across institutions, enterprises and societies. Libraries, which are considered as the intellectual heart of a society, will also immensely benefit from incorporating such systems within their technical architecture. Future work in this arena can revolve around:

- development of a faceted representation framework of knowledge graphs as per the established tenets of facet analysis and generalizing it towards a unified theory of faceted knowledge organization based on knowledge hypergraphs, both situated within the broad framework of the methodology
- application of the domain development methodology to diverse application domains and studying the associated issues and intricacies
- generalizing the scope of the methodology towards development of a more high-level, semantics intensive knowledge ecology framework
- analyzing the methodology from the prism of information behaviour and information experience principles.

References

Aridhi, Sabeur and Engelbert Mephu Nguifo. 2016. "Big Graph Mining: Frameworks and Techniques." *Big Data Research* 6: 1-10. doi:10.1016/j.bdr.2016.07.002

Bagchi, Mayukh. 2019. "A Knowledge Architecture using Knowledge Graphs." MS Dissertation, Indian Statistical Institute.

Basque, Josianne, Gilbert Paquette, Béatrice Pudelko and Michel Leonard. 2014. "Collaborative Knowledge Modelling with a Graphical Knowledge Representation Tool: A Strategy to Support the Transfer of Expertise in Organisations." In *Knowledge Cartography: Advanced Information and Knowledge Processing,* eds. Alexandra Okada, Simon Buckingham Shum and Tony Sherborne. London: Springer, 491-517. doi:10.1007/978-1-4471-6470-8_22

Benjamin, P. C., Menzel, C. P., Mayer, R. J., Fillion, F., Futrell, M. T., deWitte, P. S., and Lingineni, M. 1994.

"IDEF5 Method Report." *Knowledge Based Systems, Inc.*

Breeding, Marshall. 2015. "Future of Library Discovery Systems." *Information Standards Quarterly* 27: 24-30. doi:10.3789/isqv27no1.2015.04

Broughton, Vanda. 2006. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." *Aslib proceedings* 58: 49-72.

Chah, Niel. 2018. "OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google's Knowledge Graph." *arXiv preprint.* arXiv:1805.03885.

Chatfield, Christopher. 1995. *Problem Solving: A Statistician's Guide*. London: Chapman and Hall.

Chen, Chaomei. 2017. *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. London: Springer.

Dahlberg, Ingetraut. 2006. "Knowledge Organization: A New Science?" *Knowledge Organization* 33: 11-19.

Deshpande, Omkar, Digvijay S. Lamba, Michel Tourn, Sanjib Das, Sri Subramaniam, Anand Rajaraman, Venky Harinarayan, and AnHai Doan. 2013. "Building, Maintaining, and Using Knowledge Bases: A Report from the Trenches." In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD '13)*. New York, USA: ACM, 1209–1220. doi:10.1145/2463676.2465297

Duan, Yucong, Lixu Shao and Gongzhu Hu. 2018. "Specifying Knowledge Graph with Data Graph, Information Graph, Knowledge Graph, and Wisdom Graph." *International Journal of Software Innovation* 6: 10-25. doi:10.4018/ijsi.2018040102

Dutta, Biswanath, Devika P. Madalli and Usashi Chatterjee. 2015. "YAMO: Yet Another Methodology for Large-Scale Faceted Ontology Construction." *Journal of Knowledge Management* 19: 6-24. doi:10.1108/JKM10-2014-0439

Ehrlinger, Lisa and Wolfram Wöß. 2016. "Towards a Definition of Knowledge Graphs." In *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMAN TiCS2016 and the 1st International Work- shop on Semantic Change & Evolving Semantics (SuCCESS'16)*, eds. Michael Martin,

Martí Cuquet, and Erwin Folmer, Vol. 1695. Leipzig, Germany: CEUR-WS. http://ceur-ws.org/Vol-1695/paper4.pdf

Esnault, Liliane, Amaury Daele, Romain Zeiliger and Bernadette Charlier. 2009. "Creating an Innovative Palette of Services for Communities of Practice with Participatory Design." In *Learning in the Synergy of Multiple Disciplines,* ed. Cress U., Dimitrova V. and Specht M. Lecture Notes in Computer Science 5794. Heidelberg: Springer, 304-309. doi:10.1007/978-3-642-04636-0_29

Fernández-López, Mariano, Asunción Gomez-Perez, and Natalia Juristo. 1997. "METHONTOLOGY; From Ontological Art Towards Ontological Engineering." In *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*. Stanford: Stanford University, EEUU, 33-40. http://oa.upm.es/ 5484/1/METHONTOLOGY_.pdf

Giunchiglia, Fausto, Biswanath Dutta and Vincenzo Maltese. 2014. "From Knowledge Organization to Knowledge Representation." *Knowledge Organization* 41: 44-56.

Giunchiglia, Fausto, Marchese, Maurizio and Zaihrayeu, Ilya. 2006. "Encoding Classifications into Lightweight Ontologies." *Journal of data semantics* 8: 57-81.

Grainger, Trey, Khalifeh AlJadda, Mohammed Korayem, and Andries Smith. 2016. "The Semantic Knowledge Graph: A Compact, Auto-Generated Model for Real-Time Traversal and Ranking of any Relationship within a Domain." In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Montreal: IEEE, 420-429. doi:10.1109/dsaa.2016.51

Grüninger, Michael and Mark S. Fox. 1995. "Methodology for the Design and Evaluation of Ontologies." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI, Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Canada: Morgan Kaufmann, 1-10.

Guarino, Nicola, Daniel Oberle and Steffen Staab. 2009. "What Is an Ontology?" In *Handbook on Ontologies*, eds. S. Staab and R. Studer. Springer, Berlin, 1-17.

Guarino, Nicola and Welty, Christopher. 2002. "Evaluating Ontological Decisions with OntoClean." *Communications of the ACM* 45: 61-65.

Hjørland, Birger. 2002. "Domain analysis in Information Science. Eleven approaches – Traditional as well as Innovative." *Journal of Documentation* 58: 422-462.

Hjørland, Birger and Hanne Albrechtsen. 1995. "Toward a New Horizon in Information Science: Domain analysis." *Journal of the American Society for Information Science* 46: 400-425.

Hjørland, Birger. 2017. "Domain Analysis." *Knowledge Organization* 44: 436-64.

Johnson-Laird, P. N. 1983. *Mental models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.

Joudrey, D.N., A.G. Taylor and D.P. Miller. 2015. *Introduction to Cataloging and Classification.* Santa Barbara, CA: Libraries Unlimited.

Kosslyn, S.M. 1994. *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.

Krötzsch, Markus, Maximilian Marx, Ana Ozaki and Veronika Thost. 2018. "Attributed Description Logics: Reasoning on Knowledge Graphs." In *Proceedings of 27th International Joint Conference on Artificial Intelligence (IJCAI 18)*, ed. Lang, J. Sweden: AAAI, 5309–5313. doi:10.24963/ijcai.2018/743

Le-Phuoc, Danh, Hoan Nguyen Mau Quoc, Hung Ngo Quoc, Tuan Tran Nhat and Manfred Hauswirth. 2016. "The Graph of Things: A Step towards the Live Knowledge Graph of Connected Things." *Journal of Web Semantics* 37: 25-35.

Melo, André, & Heiko Paulheim. 2017. "An Approach to Correction of Erroneous Links in Knowledge Graphs." In *K-CAP Workshops and Tutorials Proceedings,* New York: ACM, 54-57.

National Academy of Sciences. 2004. *Mapping Knowledge Domains*. Washington, DC: The National Academies Press. doi:10.17226/11048

Needham, Mark and Hodler, Amy. 2019. *Graph Algorithms: Practical Examples in Apache Spark and Neo4j.* USA: O'Reilly Media, Inc.

Nickel, M., K. Murphy, V. Tresp and E. Gabrilovich. 2016. "A Review of Relational Machine Learning for Knowledge Graphs." *Proceedings of the IEEE* 104: 11--33.

Novak, J. D. and A. J. Cañas. 2008. "The Theory Underlying Concept Maps and How to Construct and Use Them." Technical report, Florida Institute for Human and Machine Cognition. Available at http://cmap.ihmc.us/docs/pdf/TheoryUnderlyingConceptMaps.pdf

Noy, Natalya F. and Deborah L. McGuinness. 2001. "Ontology development 101: A Guide to Creating Your First Ontology." *Technical Report SMI-2001-0880*, Stanford: Stanford Knowledge Systems Laboratory, Stanford University. https://protege.stanford.edu/publications/ontologydevelopment/ontology101.pdf

Okada, Alexandra, S. Buckingham Shum, and Tony Sherborne. 2014. *Knowledge Cartography: Software Tools and Mapping Techniques*. London: Springer-Verlag.

Paulheim, Heiko. 2017. "Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods," *Semantic Web* 8: 489–508.

Plantié, M. and P.M. Riccio. 2010. "Knowledge Cartography and Social Network Representation: Application to Collaborative Platforms in the Scientific Area" In *Sixth International Conference on Signal-Image Technology and Internet Based Systems*, Kuala Lumpur, Malaysia. doi:10.1109/SITIS.2010.45

Prasad, A. R. D. and D.P. Madalli. 2008. "Faceted Infrastructure for Semantic Digital Libraries." *Library review* 57: 225-234.

Prieto-Díaz, Rubén. 2003. "A Faceted Approach to Building Ontologies." In *Proceedings Fifth IEEE Workshop on Mobile Computing Systems and Applications*, Washington, DC: IEEE, 458-65. doi:10.1109/IRI.2003.1251451

Qingjie, Lv, Xu Lingyu, Yu Jie, Wang Lei, Xun Yunlan, Shi Suixiang and Liu Yang. 2014. "Research on Domain Knowledge Graph based on the Large Scale Online Knowledge Fragment", In *IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*, Ottawa, Canada: IEEE, 1-6. doi:10.1109/WARTIA.2014.6976259.

Ranganathan, S. R. 1957. "Library Science and Scientific Method." *Annals of Library and Information Studies* 4: 19–32.

Ranganathan, S. R. 1967. *Prolegomena to Library Classification*. London: Asia Publication House.

Ranganathan, S. R. 1972. "Impact of Growth in the Universe of Subjects on Classification." In *Ranganathan memorial issue,* eds. S. R. Ranganathan and Bernard Ira Palmer, Copenhagen: Danish Centre for Documentation.

Ranganathan, S. R. 1987. *Colon Classification. 7th ed., rev. and ed. by M. A. Gopinath*. Bangalore: Sarada Ranganathan Endowment.

Satija, M.P., Madalli, Devika P. and Dutta, Biswanath. 2014. "Modes of Growth of Subjects". *Knowledge Organization* 41: 195-204. doi:10.5771/0943-7444-2014-3-195.

Selvin, A. M. 2014. "Performing Knowledge Art: Understanding Collaborative Cartography." In *Knowledge Cartography: Advanced Information and Knowledge Processing,* eds. Alexandra Okada, Simon Buckingham Shum and Tony Sherborne. London: Springer, 233-259. doi:10.1007/978-1-4471-6470-8_11

Sierhuis, M. and S. J. Shum. 2014. "Human-Agent Knowledge Cartography for e-Science: NASA Field Trials at the Mars Desert Research Station." In *Knowledge Cartography: Advanced Information and Knowledge Processing,* eds. Alexandra Okada, Simon Buckingham Shum and Tony Sherborne. London: Springer, 381-399. doi:10.1007/978-1-4471-6470-8_17

Singhal, Amit. 2012. "Introducing the Knowledge Graph, Things, not Strings." Blog post at http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html.

Skupin, A. 2004. "The World of Geography: Visualizing a Knowledge Domain with Cartographic Means." In *Proceedings of the National Academy of Sciences, 101(Supplement 1)*, 5274-5278. doi:10.1073/pnas.0307654100

Smiraglia, Richard P. 2015a. "Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster." *Knowledge Organization* 42: 602-611.

Smiraglia, Richard P. 2015b. *Domain Analysis for Knowledge Organization: Tools for Ontology Extraction*. Chandos Information Professional Series. Amsterdam: Chandos.

Southerton, D. 2011. *Encyclopedia of Consumer Culture*. USA: Sage Publications.

Spence, R. 2001. *Information Visualization (Vol. 1)*. New York: Addison-Wesley.

Tergan, S. 2005. "Digital Concept Maps for Managing Knowledge and Information." In *Knowledge and Infor-*

mation Visualization, eds. Sigmar-Olaf Tergan and Tanja Keller. Heidelberg: Springer Verlag, 185-204. 185-204. doi:10.1007/11510154_10

Uschold, M. and M. King. 1995. "Towards a Methodology for Building Ontologies." In *Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95*, Montreal, Canada.

Vasconcellos, M. 2014. "An Experience of the Use of the Cognitive Mapping Method in Qualitative Research." In *Knowledge Cartography: Advanced Information and Knowledge Processing,* eds. Alexandra Okada, Simon Buckingham Shum and Tony Sherborne. London: Springer, 471-490. doi:10.1007/978-1-4471-6470-8_21.

Vickery, Brian Campbell. 1960. *Faceted Classification: A Guide to the Construction and Use of Special Schemes*. London: ASLIB.

Vosniadou, S. and W. F. Brewer. 1992. "Mental Models of the Earth: A Study of Conceptual Change in Childhood." *Cognitive Psychology* 24: 535–585.

Vosniadou, S. 2002. "Mental Models in Conceptual Development." In *Model-based reasoning. Science, technology, values*, eds. L. Magnani and Nancy Nersessian. New York: Kluwer Academic Publishers, 353-368.

Wilson, Patrick. 1969. "Two Kinds of Power: An Essay on Bibliographical Control." *The Library Quarterly* 39: 112-114. doi:10.1086/619728

Zhang, Yuyu, Hanjun Dai, Zornista Kozareva, Alexander J. Smola and Le Song. 2018. "Variational Reasoning for Question Answering with Knowledge Graph." In *Thirty-Second AAAI Conference on Artificial Intelligence*. North America: AAAI, 6069-6076.