Towards Knowledge Exchange: State-of-the-Art and Open Problems*

Bahar Ghadiri Bashardoost[1[0000−0001−8654−6626]], Kelly Lyons[1[0000−0001−6866−7301]], and Renée J. Miller[2[0000−0002−1484−4787]]

[1] University of Toronto, Toronto, CA
{ghadiri,klyons}@cs.toronto.edu
[2] Northeastern University, Boston, USA
miller@northeastern.edu

Abstract. We discuss our experience in bringing data exchange to knowledge graphs. This experience includes the development of Kensho, a tool for generating mapping rules and performing knowledge exchange between two Knowledge Bases (KBs). We highlight the challenges addressed in Kensho, including managing the rich structural complexity of KBs and the need to handle incomplete correspondences between property paths. We use Kensho to highlight many open problems related to knowledge exchange including how knowledge translation can inform the task of KB integration and population.

Keywords: Knowledge exchange and translation · Data exchange · Mapping generation.

1 Introduction

Knowledge-rich applications can see significant performance improvements by using domain-specific Knowledge bases (KBs). Populating and enriching these KBs has, thus, become an important challenge. Many KB population methods use Information Extraction (IE) techniques in order to harvest facts from unstructured and semi-structured corpora. While very useful, the extracted information can be inaccurate, requiring careful curation to produce high-quality knowledge. In this work, we examine a complementary and powerful approach for KB expansion that is based on knowledge exchange, the process of translating knowledge from one KB to another, even when these KBs use very different concepts, properties, and graphs to represent their knowledge. We consider how to lift knowledge from KBs (such as upper ontologies or hand-curated domain-specific KBs) to expand other KBs. We explore the state-of-the-art in creating KB mappings and using them for knowledge translation. We discuss how this work has implications for using relational or other structured data sources for both KB expansion and for automating mapping creation in OBDA (Ontology-based Data Access).

* This research was funded in part by an NSERC Strategic Partnership Grant

Recent advances have made the discovery of desirable structured sources of knowledge more and more feasible. While these advances help data engineers explore and discover related knowledge to enrich a KB, translating and integrating the newly discovered knowledge into the KB remains challenging. Data exchange [27] is a prominent approach for data translation within and among relational and nested relational databases and thus, it makes sense to investigate how data exchange solutions can be applied to knowledge base exchange. In the data exchange problem, data that is structured under a source schema is transformed into an instance of a target schema. This is accomplished using a set of rules (called mapping rules or a schema mapping) that specify the relationship between the source and target schemas.

Fagin et al. [27] laid the theoretical foundation for data exchange over relational data and identified important data exchange tasks, namely materializing a target instance and answering queries. Target materialization focuses on problems such as determining whether a target instance (i.e., a solution) exists for a given source instance and accompanying set of mapping rules and, if so, how to generate solutions efficiently. Since there can be more than one solution that satisfies the given set of mapping rules, another important problem is to determine whether there exists a preferred solution(s) and if so, whether a preferred solution can be created with reasonable complexity. Having multiple possible solutions raises another challenge. The main goal of data exchange is to allow queries to be answered over a target instance in a way that is consistent with the data stored in the source; however, the query result might differ, depending on the target solution over which the query is evaluated. Research on query answering in data exchange investigates which query answer is the most desirable. In practical applications of data exchange, another important challenge is identifying mapping rules using automated or semi-automated techniques.

To date, data exchange literature has primarily focused on relational or nested relational source and target settings with less work dedicated to data exchange in KBs. In this paper, we present some current work in KB exchange and highlight challenges and open problems for future research.

2 Knowledge Exchange: Where Are we Now

Researchers have begun the systematic investigation of exchanging data among KBs [9,10,11,12]. In recent work, Arenas et al. [10] proposed a formal framework for exchanging knowledge between two KBs that are expressed using $\textit{DL-Lite}_R$. In this work, three main types of solutions were investigated that potentially have desirable characteristics for materialization and query answering.

Another important challenge that needs to be addressed in order to make knowledge exchange possible is identifying a set of mapping rules that correctly describes the relationship between the source and the target. When dealing with KBs, several languages [14,24,55, and others] and frameworks [24,66] have been proposed to facilitate the manual generation of these rules. In addition, there are a few pioneering KB schema mapping generation tools (KMGT), including

Mosto [59,62] and a system by Qin et al. [58] that automatically create mappings given a set of correspondences. In this direction, we have recently proposed Kensho [35] which improves upon the first generation mapping tools by taking into consideration the lessons learned from traditional (nested) relational data exchange and mapping tools (MGT) [16] and extends these, taking into account the unique characteristics of KBs. Kensho can produce mapping rules even in the presence of cycles, incompleteness in the source, and in settings with missing or unknown correspondences between properties or property paths. In addition, Kensho performs knowledge translation using value invention to preserve the proper grouping of data in the target KB.

Usually, the first step in data sharing tasks is alignment (a.k.a., matching). The output of this step is a set of correspondences each expressing a relationship among a (set of) resource(s) in the source and a (set of) resource(s) in the target. For instance, in Figure 1, the output of the first step (the alignment task) is the red and blue lines, that indicate the concept $Person$ in the source corresponds to concept $Person$ in the target, and similarly, $workAddress$ in the source corresponds to $address$ in the target. Note that if one uses only the information provided through correspondences for translating knowledge from source to target, some of the information in the source might not be transferred to the target. For instance, in this example, while people and their addresses may be copied over to target, people will not be associated with their own work address in the target. In addition, looking more closely at the structure of the target, one might notice that two people in the target are associated using a relationship called *related*. Thus a desirable solution for the exchange might be a target instance in which the *related* property is also populated using some/all of the possible relationships between people in the source. For instance, depending on the semantics of the target,if two people work on a project in the source, it may be desirable to model those people as *related* in the target. Alternatively, if one person is related to another through *hasSupervisor* in the source, it may be desirable to model these two people as *related* in the target. In fact, both of these relationships may be desirable to model in the target. An important goal of schema mapping creation is to ensure that associations like these between resources are modelled and preserved by the mapping rules.

In traditional data exchange, a large number of tools that automatically generate the mapping rules between a source and a target do so by taking as input a set of correspondences between the source and target (i.e., the alignment), and enriching them with information obtained from the structural characteristics of the source and target. Except for MostoDex [63,64] (see Section 3 for more details), all KMGTs, including Kensho, follow the same strategy. Kensho, generates the mapping rules in two main steps: 1) semantic association discovery; and 2) correspondence interpretation. The goal of the first step is to capture all semantic associations between aligned resources of each KB. One key feature of Kensho is that in this phase, it captures all the ways that resources are associated with each other, of course, up to a certain depth (when there are cycles). For instance, two people in the source are associated if they work on a project,

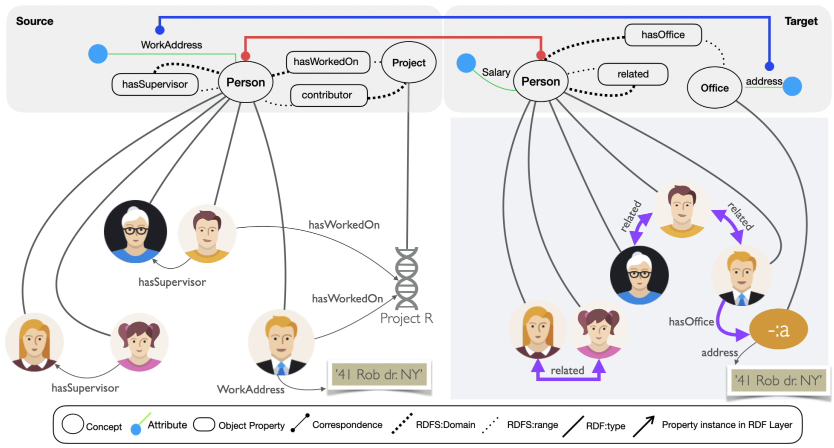

Fig. 1. RDFS and RDF layers of two KBs and correspondences between them. Target KB is populated using source instance.

or if they have the same supervisor, or if they are in a supervision chain of a certain length, etc. The goal of the second step is to interpret the set of given correspondences *collectively* using the discovered associations among resources of each KB. To see why this step is necessary, note that in the example of Figure 1, we will fail to transfer the relationship between people and their work addresses if we consider each correspondence separately and not collectively. Note that the associations among resources that participate in the correspondences provide cues on how to weave these correspondences together to find mapping rules that not only transfer resources that participate in correspondences, but also their relationships with each other.

Most of the advantages of Kensho over previous work stems from the fact that Kensho captures all possible ways that resources of KBs are associated with each other. For instance, in the example of Figure 1, Kensho will suggest a rule that interprets the two correspondences together since it captures the fact that people and their work addresses are associated in the source and in the target and this provides a cue that this association should be preserved by the mapping. The associations also provide means for proper value invention. For instance, to preserve the association between a person and their work address, for each individual of type *Person* and for each of their work addresses, a new blank node of type *Office* will be created which has the value of the source's *workAddress* as the value of its *address* attribute and participates in a *hasOffice* relation with the transferred individual of the type *Person*. One problem with capturing all associations among resources is that a large number of possible associations may be identified that subsequently have to be considered in the interpretation phase. To address this problem we have used a set of validity and pruning measures to reduce the possible choices that need to be considered as a valid mapping rule [35]. In addition, we have proposed a set of ranking heuristics to help reduce the burden of selecting the best set of mapping rules.

3 Challenges and Opportunities

In this section, we present a vision for some of the most important open problems related to knowledge translation and exchange. Some of the challenges mentioned in this section are already studied in depth in traditional (non-KB) data exchange literature. One goal of this section is to highlight important challenges already identified in the parent field. We establish a call to use lessons learned from the solutions proposed for these challenges to further broaden the research in the KB exchange area. We also talk about some challenges which are more unique to KBs and thus have been studied less in previous work. We consider the evaluation of mapping generation tools which is not a trivial task. Unfortunately most approaches used to evaluate traditional mapping generation tools can not be readily adapted to KBs. We present new evaluation challenges and discuss opportunities for leveraging the Ontology Alignment Evaluation Initiative (OAEI) which provides a series of correspondence discovery tasks [8].

3.1 Mapping Generation: Input Evidence

Metadata Evidence Mapping generation is carried out using initial evidence about how KBs correspond. In Kensho and other KMGTs [35,58,59,62], this evidence is a set of correspondences. In the ontology alignment literature, correspondences are distinguished as being 1:1, aka simple correspondences (e.g., connecting a source concept with a target concept) or n:m, aka complex correspondences (e.g., connecting a source path/property – and therefore the endpoints of the path – with a target path/property) [25]. Kensho was the first KMGT to consider complex correspondences. An important innovation in Kensho is that it does *not* assume that the set of correspondences is complete. From the given correspondences and KB structure, Kensho can suggest additional possible correspondences that would map more of the source data. One important type of correspondence that, to the best of our knowledge, is not considered in KMGTs is metadata-data (MAD) correspondences [40,51]. Such correspondences articulate the need for a mapping that transforms data values to metadata (e.g., a property name) or vice versa. Arguably, such transformations may be more prevalent in KBs than in relational or semi-structured models. Generating proper mappings from MAD correspondences can be challenging.

Another type of important correspondence which needs further investigation is complex correspondences which indicate that a function can be used to transform a (set of) source value(s) into a (set of) target value(s), or that a filter should be imposed on source values that are being transferred. It is straight forward to extend algorithms in current KMGTs to cover most of these correspondences. However, one advantage that these types of correspondences might bring to the table is in helping to impose some constraints on invented values, which can in turn help improve the query answering capabilities of the materialized target. For instance, imagine that in the example of Figure 1, there was a correspondence that expresses that if a person supervises another person in the source, their *salary* is greater than \$10k. In this case, although the salary of individuals cannot be materialized, a constraint can be added to the target to reflect this fact. Using this added constraint can help in answering queries, for instance, if someone asks for people who earn more than \$10k, we know that at least those who were supervisors in the source must be returned as an answer. Of course, studying such correspondences also requires theoretical advances, for example, understanding changes needed in exchange settings so that such constraints can be translated and understanding the effects of doing so on the complexity of the important tasks of materialization and query answering (Afrati et al. [2] discuss some relevant theoretical underpinnings for such tasks in relational data). One other interesting example of such correspondences are those that require aggregation over values in the source. Note that aggregating values is not a trivial task specially in KBs since KBs adhere to an open-world data model [18].

Data Evidence An alternative to using correspondences to drive mapping generation (sometimes called a schema-driven approach) is to use data. In traditional data exchange, *data-driven* or *example-based* MGTs use source and target data examples. In some approaches, the data is a single full instance of the source schema and the corresponding target instance. In other approaches, smaller sets of example data are used (e.g., a set of source tuples and their corresponding target tuples).

The seminal work of Gottlob and Senellart [37,68] proposed a theoretical framework for the problem of deriving a set of mapping rules from a single ground data example, that is, an instance of the source and an instance of the target schema which do not include nulls. One of the most important contributions of this work is that it casts the problem as an optimization problem by introducing a cost function that expresses how well a rule helps in translating the given source example. Given a *set* of universal examples (pairs of source and target instances where the target instance is a universal solution of the source instance), EIRENE [4] checks whether there exist a set of rules that can fit all the examples, and if so, automatically generates the most general set of rules. In addition, given a set of universal data examples and a membership oracle, ten Cate et al. [19,20] show that a set of mapping rules can be learned in some specific traditional settings. These approaches require very high quality input including examples with certain characteristics (e.g., universal examples). Some of the work in this area can produce mapping rules even when the examples have lower quality (and may even be incorrect) making the approaches more robust if non-experts are providing examples [15,57]. In the future, it is important to understand how data driven approaches can be adopted in KMGTs particularly due to the web scale nature of many KBs making them ideal settings for using data examples.

Mixed Evidence For KBs, MostoDex [63] uses both an example and a set of source-to-target correspondences. The single example provided to MostoDex must be complete and correct, which puts a large burden on the domain expert. In contrast, for relational exchange, Kimmig et al. [43,44] propose an approach using both data examples and metadata evidence (source-to-target correspondences and source constraints). Importantly in this work, both the data and metadata evidence may be incomplete or incorrect (including schema evidence like correspondences or mined joined paths). They propose Collective Mapping Discovery (CMD) to reason over this possibly inconsistent evidence to produce a schema mapping (a set of mapping rules) that best explain the evidence. An important research direction for KGMTs is mapping generation that can use incomplete or incorrect correspondences and data examples, perhaps leveraging automatically generated examples from the Linked Open Data (LOD) cloud.

Combined Correspondence and Mapping Generation In the spirit of fully combining logical and statistical reasoning, a KGMT that collectively learns correspondences, examples, and mappings would be an important milestone. Some work including ILIADS [72], PARIS [71], and CODI [41] do this for ontology merging (learning same-as, isa, and equivalent-class axioms to merge two ontologies within a domain), but do not handle deep structural and semantic heterogeneity that cannot be resolved with simple axioms.

3.2 Mapping Refinement

A major challenge in mapping generation is in resolving ambiguity in the relationship between the source and target. Often the evidence provided is not sufficient to select a single definitive semantic interpretation (e.g., does the target *related* property represent a supervision relationship or something else?). Thus, usually mapping generation involves a refinement phase which is often a user-in-the-loop process of eliciting more information to refine the mapping.

Kensho is part of an *integration by example paradigm* [52] in which a domain expert's feedback is used iteratively to refine the mapping rules. A body of traditional data exchange literature is dedicated to investigating the best set of examples to show to domain experts (to resolve ambiguity or choices in mapping creation), ways to automatically generate these examples, show them to the domain experts, and incorporate domain experts' feedback which might be contradictory or wrong [5,23,77, and others]). On the theoretical front, the complexity of the problem of whether a set of mapping rules can be uniquely identified by a set of finite examples has been investigated for various traditional settings with mapping rules in different languages such as GAV and LAV rules [3,6,21]. We are currently working on approaches for interacting with domain experts, by finding examples which are *sufficient illustrations* [77] of rules generated by Kensho and incorporating the feedback to improve and refine the final set of mapping rules which can be used to exchange data between two KBs.

One of the important features of Kensho is that it does not rely on the existence of any constraints other that domain, range, and subsumption. However, if other constructs are contained in the source or target, they can be used in refinement to eliminate (or lower the confidence of) some ambiguous mapping options. For instance, cardinality constraints such as `owl:minCardinality`, `owl:maxCardinality`, `owl:FunctionalProperty` and `owl:InverseFunctionalProperty` or disjointness constraints can sometimes help select a property path that best represents the relationship between two concepts. Note that if the KB is already populated with some instances, even if the cardinality constraints are not present, it might be the case that these constraints can still be automatically inferred (e.g., Bühmann and Lehmann [17] provide some examples of this). In general, we feel that KB profiling and mining may be helpful in improving mapping quality, in a similar way that database profiling and dependency mining [1] has been used to mine for join paths or other constraints that can guide mapping generation.

Depending on the problem at hand, instance based methods can be used to help refine possible mapping alternatives. For instance, if we are trying to transfer as much data as possible from source to target, heuristics can be introduced to give higher priority to mapping rules that translate more source data. In the presence of a set of positive (and/or negative) examples, rule mining approaches [31,50,54, and others] can help identify important paths (or rules) that best fit the given examples. In addition, rule mining approaches can help enrich the set of constraints that are used in order to further refine mapping rules. It is interesting to see whether enriching the source/target ontologies with additional induced constraints produced by running a rule mining algorithm can help generate better mapping rules.

3.3 Knowledge Exchange

Of course an important reason to create mappings is to perform knowledge exchange. The goal in MGTs is to produce a *universal* solution [26], as universal solutions are the only solutions that represent the whole space of solutions in a precise sense [27]. It is important to note that early systems did this heuristically because the theory of data exchange had not yet been developed [52,53,56]. The tools motivated the development of the theory so we could formally reason about the solutions the tools produced – and formally prove they were "good" solutions. Later, as the theory progressed, the theory fed back into the tools. Fagin et al. [28] defined *core* solutions as the smallest universal solutions for relational settings and Mecca et al. [49] used this idea to produce mappings that are guaranteed to produce core solutions. Later on Chirkova et al. [22] showed that when the exchange setting consists of nested relational DTDs, materialization of the solution can be reduced to the materialization of a solution in a relational setting and provided an algorithm to do so. One advantage of the Chirkova et al. work is to enable applications to take advantage of theoretical results and efficient algorithms already proposed in relational data exchange even when they are dealing with nested relational settings (++SPICY [48] is an example).

The study of what are the best solutions for KB exchange has been investigated by Arenas et al. [10]. They define a notion of universal solution (for KBs represented in DL-Lite$_R$), but also two related notions of universal UCQ-solutions (a relaxation of the notion of universal solutions) and UCQ-representations. However, their proposed setting is somewhat different than what is considered in current KMGTs that consider target constraints. An important research question is how to use this formal work to guide KMGTs and if we can use the theory to create tools that produce mappings with certain formal properties.

Most MGTs and Kensho assume that the target is not populated with any data. But this might not be the case in real world application. For relational exchange, this problem has been addressed using a *solution-aware chase*, a procedure that adds and merges source data into an existing target instance [30]. An interesting open problem is to understand how to do this in KBs where inference is more complex. In addition, practical solutions will need to use deduplication and record linkage to create links such as `same-as` and resolve conflicts in values.

3.4 Scalability and Optimization

One of the most important challenges that KMGTs face is the scale of the KBs. Kensho uses a simple method for slicing the KBs in order to be able to deal with larger scale source and target [34]. However, more sophisticated methods such as modularization [36,45,70] or partitioning [38,67,69] can be adopted to help in efficient generation of mapping rules when dealing with large scale KBs.

In Kensho, mapping rules are expressed using SPARQL. However, there are declarative languages specifically designed for expressing the mapping rules proposed in literature. One such language is R2R [14] is a powerful language for representing mapping rules for large KBs. An interesting open question is to study the use of declarative languages like R2R that can perhaps be optimized and then "compiled" into SPARQL or other execution languages.

The performance of the transformation queries or programs generated by MGTs has been an important research problem [42] and will remain so for KMGTs. To handle the open-world nature of KBs, Kensho generates SPARQL queries with many `OPTIONAL` clauses. Generally, `OPTIONAL` clauses are processed as left outer joins in most query engines and thus running nested `OPTIONAL` clauses is expensive. Recently, there has been research on methods for efficient handling of SPARQL `OPTIONAL` clauses [76], since similar to our approach, `OPTIONAL` clauses are required for many data integration tasks when dealing with KBs. We believe our approach (as well as many other KB integration approaches) can immensely benefit from research on how to optimize the execution of queries that use SPARQL `OPTIONAL` clauses.

3.5 Application to other Integration Problems

We believe the most important application of Kensho is populating an existing domain-specific ontology using other currently available structured data sources. The source of data might not be a KB as long as it can be automatically converted to one. In our evaluation [35], we showed that Kensho can effectively populate a worker expertise KB using open data published by the US Patent and Trademark Office (USPTO). To use the USPTO corpus as our source KB, we started with a subset of the USPTO's patent XML corpus[3] and automatically created a linked data corpus from it using Xcurator [39], and further enriched it using Vizcurator [32].

The ontology based data access initiative (OBDA) [75] aims to facilitate the integrated access of heterogeneous relational data sources using a target KB. OBDA does not require target materialization and instead, uses the KB as a virtual view over the relational source(s). To enable query answering, they use mapping rules to re-write KB queries in order to enable users to access data in the source by querying the target. Thus in these approaches mapping rules need to be created such that they facilitate efficient query re-writing and deal with source and target models with different expressive power. Xiao et al. [75] state that"mapping creation and management is probably the most complicated OBDA design-time task". Additional research is needed on KGMTs to support the unique requirements of OBDA [46].

3.6 Evaluation

Evaluating the effectiveness of mapping rules is not a trivial task. Computing precision and recall is the standard in the literature for comparing different set of correspondences. However, mapping rules are not simple sets. Two sets of different mapping rules may be equivalent, so it is not sufficient to check if a specific syntactic mapping is produced by a KMGT. In addition, testing equivalence of queries and mappings is undecidable in general, and many of the mapping languages used in practice have this property. Hence, over time, tools and benchmarks have been developed to help evaluate the effectiveness of mapping generation tools. STBenchmark [7] was one of the first for MGTs. One of the contributions of this work is a suite of mapping micro scenarios (or mapping patterns) that represents a minimum set of transformations that should be readily supported by mapping tools. This idea was generalized by the meta-data generator iBench [13] that permits the efficient creation of benchmarks with large and complex schemas and data exchange scenarios that require value invention. These tools are designed to evaluate settings that involve relational sources and targets (or in the case of STBenchmark nested relational). DTSBenchmark [60] was the first to devise a set of scenarios when the source and target are both KBs. LODIB [65] refined these scenarios based on patterns that occur often on examples from the LOD Cloud. Later work [61] proposed a framework called MostoBM, which contains the set of scenarios from DTSBenchmark [60] as well as a metadata generator that allows the complexity of the scenarios to be tuned. MostoBM uses a few parameters that can be changed to systematically generate settings with different degrees of complexity, including three schema-level parameters (namely, *depth* of the class relationships, *breadth* of the class relationships, and *the number of attributes*). However, these parameters are not as

[3] https://www.google.com/googlebooks/uspto-patents-grants-text.html

extensive as the iBench parameters and do not control important factors such as the use of value invention and the number of alternative interpretations for a correspondence. To begin to address this, we recently proposed a new set of scenarios for KB exchange that covers settings that require value invention, incomplete source and target KBs, incomplete correspondences, and KBs containing cycles [33], though notably our work does not go as far as providing the KB equivalent of an iBench metadata (exchange scenario) generator.

Real world scenarios play an important role in knowledge exchange and especially in the evaluation of ranking measures in mapping generation tools. Note that all of the mapping rules generated by Kensho are consistent with the correspondences and source/target KBs but, depending on the context, some mappings may be more desirable than others. Returning to our example in Figure 1, a user must decide if they want the target *related* property left empty, populated with the source *hasSupervisor* property, populated with the $hasWorkedOn/contributor$ path, or some combination (union) of these. And of course given cycles, one could also consider the source $hasSupervisor/hasSupervisor$ path (to get a person's second-line supervisor) and so on. None of these mappings are wrong based on the structure of the KBs and the given correspondences. The semantics of what *related* means in the target is missing and must be determined by a human, perhaps with the aid of data examples. Situations like these make it important for tools to rank *most-likely* mappings well, so the user is not overwhelmed by options. Such ambiguity is inherent in integration and it poses important evaluation challenges. Currently there are not enough open source real world scenarios that can be used for settings which involve KBs. One resource we used in the evaluation of Kensho was the OAEI corpora which provides a large set of source and target ontologies together with a curated set of *ground-truth* correspondences. We were able to re-purpose the data provided in this initiative to better evaluate our knowledge translation tool in a real world setting. We believe knowledge translation and exchange research can greatly benefit from a community effort which identifies corpora that can be shared to produce comparable evaluations of KGMT. This research can also benefit from initiatives that push for open source settings which are more focused on the challenges specific to this area.

3.7 Evolution

Source and target KBs or schemas may not be static and can change through time. When this happens, mappings may become invalid or inconsistent. Starting the mapping generation from scratch will waste work and can be expensive especially in cases where significant domain expert intervention is required [29]. In addition there is no guarantee that the regenerated mapping rules will reflect the original semantics of the previous mapping and have the same capabilities in query answering [29,74]. Hence, researchers have studied how to *identify inconsistencies* and *adapt mappings* to conform to the new source and target. This has been done in MGTs such as ToMAS [73,74]. ToMAS considers relational and nested-relational schemas and adapts mappings in the face of schema changes (additions or deletions in the schema) as well as semantic and structural changes (such as changes in the schema constraints or nesting structure).

In addition, in traditional data exchange schema mapping evolution has been proposed [29] based on mapping composition and inversion. For OBDA mappings, Lembo et al. [47] introduce two different notions of repair. The first notion, called deletion-based mapping repair, reflects the idea of repairing through a minimal number of deletion of assertions from the original mapping. The second notion, called entailment based mapping repair, aims to preserve the assertions which are implied by the original mapping, as much as possible. Unfortunately, to the best of our knowledge, there is not yet any work in this area for settings which involve source and target KBs. Indeed some of the motivating scenarios used to evaluate KMGTs come from KB evolution (specifically different versions of DBpedia) [62]. This is an issue since often the source KB resides in a dynamic environment such as web with no centralized authority, and thus its structure might change often and without any prior notice. It is important to see how various repair notions proposed in the literature can be used in the context of knowledge translation and to facilitate the mapping adaptation.

4 Conclusions

We have laid out an extensive research agenda for Knowledge Exchange. While we have focused on the development and evaluation of tools that perform mapping discovery and knowledge translation, we believe that as with data exchange, further development of the theory and foundations of knowledge exchange is critical to informing and fueling the development of better, more robust and accurate, tools and algorithms. We also believe the necessity of being able to exchange knowledge between the heterogeneous and ever growing web of knowledge bases will be a catalyst for the development of new mathematical tools and principles for understanding the foundations of knowledge exchange.

References

1. Abedjan, Z., Golab, L., Naumann, F., Papenbrock, T.: Data Profiling. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2018)
2. Afrati, F., Li, C., Pavlaki, V.: Data exchange in the presence of arithmetic comparisons. In: EDBT. p. 487498 (2008)
3. Alexe, B., ten Cate, B., Kolaitis, P.G., Tan, W.C.: Characterizing schema mappings via data examples. TODS **36**(4), 1–48 (2011)
4. Alexe, B., ten Cate, B., Kolaitis, P.G., Tan, W.C.: Eirene: Interactive design and refinement of schema mappings via data examples. PVLDB **4**(12), 1414–1417 (2011)
5. Alexe, B., Chiticariu, L., Miller, R.J., Tan, W.C.: Muse: Mapping understanding and design by example. In: ICDE. pp. 10–19 (2008)
6. Alexe, B., Kolaitis, P.G., Tan, W.C.: Characterizing schema mappings via data examples. In: PODS. pp. 261–272 (2010)
7. Alexe, B., Tan, W.C., Velegrakis, Y.: STBenchmark: towards a benchmark for mapping systems. PVLDB **1**(1), 230–244 (2008)

8. Algergawy, A., Faria, D., Ferrara, A., Fundulaki, I., Harrow, I., Hertling, S., Jiménez-Ruiz, E., Karam, N., Khiat, A., Lambrix, P., Li, H., Montanelli, S., Paulheim, H., Pesquita, C., Saveta, T., Shvaiko, P., Splendiani, A., Thiéblin, É., Trojahn, C., Vatascinová, J., Zamazal, O., Zhou, L.: Results of the ontology alignment evaluation initiative 2019. In: OM. pp. 46–85 (2019)
9. Arenas, M., Botoeva, E., Calvanese, D.: Knowledge base exchange. In: DL. vol. 4 (2011)
10. Arenas, M., Botoeva, E., Calvanese, D., Ryzhikov, V.: Knowledge base exchange: The case of OWL 2 QL. Artificial Intelligence **238**, 11–62 (2016)
11. Arenas, M., Botoeva, E., Calvanese, D., Ryzhikov, V., Sherkhonov, E.: Exchanging description logic knowledge bases. In: KR (2012)
12. Arenas, M., Pérez, J., Reutter, J.: Data exchange beyond complete data. JACM **60**(4), 28 (2013)
13. Arocena, P.C., Glavic, B., Ciucanu, R., Miller, R.J.: The iBench integration metadata generator. PVLDB **9**(3), 108–119 (2015)
14. Bizer, C., Schultz, A.: The R2R framework: Publishing and discovering mappings on the web. In: COLD. pp. 97–108 (2010)
15. Bonifati, A., Comignani, U., Coquery, E., Thion, R.: Interactive mapping specification with exemplar tuples. In: SIGMOD. pp. 667–682 (2017)
16. Bonifati, A., Mecca, G., Papotti, P., Velegrakis, Y.: Discovery and correctness of schema mapping transformations. In: Schema matching and mapping, pp. 111–147 (2011)
17. Bühmann, L., Lehmann, J.: Universal OWL axiom enrichment for large knowledge bases. In: EKAW. pp. 57–71 (2012)
18. Calvanese, D., Kharlamov, E., Nutt, W., Thorne, C.: Aggregate queries over ontologies. In: ONISW. pp. 97–104 (2008)
19. ten Cate, B., Dalmau, V., Kolaitis, P.G.: Learning schema mappings. In: ICDT. pp. 182–195 (2012)
20. ten Cate, B., Kolaitis, P.G., Qian, K., Tan, W.C.: Active learning of GAV schema mappings. In: PODS. pp. 355–368 (2018)
21. ten Cate, B., Kolaitis, P.G., Tan, W.C.: Database constraints and homomorphism dualities. In: PC. pp. 475–490 (2010)
22. Chirkova, R., Libkin, L., Reutter, J.L.: Tractable xml data exchange via relations. In: CIKM. pp. 1629–1638 (2011)
23. Chiticariu, L., Tan, W.C.: Debugging schema mappings with routes. In: VLDB. vol. 6, pp. 79–90 (2006)
24. Dou, D., McDermott, D.V., Qi, P.: Ontology translation on the semantic web. J. Data Semantics **2**, 35–57 (2005)
25. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, 2nd edn. (2013)
26. Fagin, R., Haas, L., Hernández, M., Miller, R., Popa, L., Velegrakis, Y.: Clio: Schema mapping creation and data exchange. Conceptual modeling: foundations and applications pp. 198–236 (2009)
27. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: semantics and query answering. Theoretical Computer Science **336**(1), 89–124 (2005)
28. Fagin, R., Kolaitis, P.G., Popa, L.: Data exchange: Getting to the core. ACM Trans. Database Syst. **30**(1), 174–210 (2005)
29. Fagin, R., Kolaitis, P.G., Popa, L., Tan, W.C.: Schema mapping evolution through composition and inversion. In: Schema matching and mapping, pp. 191–222. Springer (2011)
30. Fuxman, A., Kolaitis, P.G., Miller, R.J., Tan, W.C.: Peer data exchange. ACM Trans. Database Syst. **31**(4), 1454–1498 (2006)

31. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with AMIE. VLDB J. **24**(6), 707–730 (2015)
32. Ghadiri Bashardoost, B., Christodoulakis, C., Hassas Yeganeh, S., Hassanzadeh, O., Miller, R.J., Lyons, K.: Vizcurator: A visual tool for curating open data. In: WWW Companion. pp. 195–198 (2015)
33. Ghadiri Bashardoost, B., Miller, R.J., Lyons, K.: Towards a benchmark for knowledge base exchange. In: DI2KG (2019)
34. Ghadiri Bashardoost, B., Miller, R.J., Lyons, K., Nargesian, F.: Knowledge translation: Extended technical report. Tech. rep., University of Toronto (2019), available at `https://tny.sh/KBTranslation`
35. Ghadiri Bashardoost, B., Miller, R.J., Lyons, K., Nargesian, F.: Knowledge translation. PVLDB **13**(11), 2018–2032 (2020)
36. Ghazvinian, A., Noy, N.F., Musen, M.A.: From mappings to modules: Using mappings to identify domain-specific modules in large ontologies. In: K-CAP. pp. 33–40 (2011)
37. Gottlob, G., Senellart, P.: Schema mapping discovery from data instances. J. ACM **57**(2), 6:1–6:37 (2010)
38. Grau, B.C., Parsia, B., Sirin, E., Kalyanpur, A.: Automatic partitioning of OWL ontologies using E-Connections. DL **147** (2005)
39. Hassas Yeganeh, S., Hassanzadeh, O., Miller, R.J.: Linking semistructured data on the web. In: WebDB (2011)
40. Hernández, M.A., Papotti, P., Tan, W.C.: Data exchange with data-metadata translations. PVLDB **1**(1), 260–273 (2008)
41. Huber, J., Sztyler, T., Noessner, J., Meilicke, C.: Codi: Combinatorial optimization for data integration - results for OAEI 2011. In: OM. pp. 134–141 (2011)
42. Jiang, H., Ho, H., Popa, L., Han, W.: Mapping-driven XML transformation. In: WWW. pp. 1063–1072 (2007)
43. Kimmig, A., Memory, A., Miller, R.J., Getoor, L.: A collective, probabilistic approach to schema mapping. In: ICDE. pp. 921–932 (2017)
44. Kimmig, A., Memory, A., Miller, R.J., Getoor, L.: A collective, probabilistic approach to schema mapping using diverse noisy evidence. TKDE (2018)
45. Konev, B., Lutz, C., Walther, D., Wolter, F.: Model-theoretic inseparability and modularity of description logic ontologies. Artificial Intelligence **203**, 66–103 (2013)
46. Lembo, D., Mora, J., Rosati, R., Savo, D.F., Thorstensen, E.: Mapping analysis in ontology-based data access: Algorithms and complexity. In: ISWC (2015)
47. Lembo, D., Rosati, R., Santarelli, V., Savo, D.F., Thorstensen, E.: Mapping repair in ontology-based data access evolving systems. In: IJCAI. pp. 1160–1166 (2017)
48. Marnette, B., Mecca, G., Papotti, P., Raunich, S., Santoro, D., et al.: ++Spicy: an open-source tool for second-generation schema mapping and data exchange. PVLDB **19**(12), 1438–1441 (2011)
49. Mecca, G., Papotti, P., Raunich, S.: Core schema mappings. In: SIGMOD. pp. 655–668 (2009)
50. Meng, C., Cheng, R., Maniu, S., Senellart, P., Zhang, W.: Discovering meta-paths in large heterogeneous information networks. In: WWW. pp. 754–764 (2015)
51. Miller, R.J.: Using schematically heterogeneous structures. In: SIGMOD. pp. 189–200 (1998)
52. Miller, R.J., Haas, L.M., Hernández, M.A.: Schema mapping as query discovery. In: VLDB. pp. 77–88 (2000)
53. Miller, R.J., Hernández, M.A., Haas, L.M., Yan, L.L., Ho, C.H., Fagin, R., Popa, L.: The Clio project: managing heterogeneity. SIGMOD Record **30**(1), 78–83 (2001)

54. Ortona, S., Meduri, V.V., Papotti, P.: Robust discovery of positive and negative rules in knowledge bases. In: ICDE. pp. 1168–1179 (2018)
55. Polleres, A., Scharffe, F., Schindlauer, R.: SPARQL++ for mapping between RDF vocabularies. In: OTM. pp. 878–896 (2007)
56. Popa, L., Velegrakis, Y., Hernández, M.A., Miller, R.J., Fagin, R.: Translating web data. In: VLDB. pp. 598–609 (2002)
57. Qian, L., Cafarella, M.J., Jagadish, H.V.: Sample-driven schema mapping. In: SIGMOD. pp. 73–84 (2012)
58. Qin, H., Dou, D., LePendu, P.: Discovering executable semantic mappings between ontologies. OTM pp. 832–849 (2007)
59. Rivero, C.R., Hernández, I., Ruiz, D., Corchuelo, R.: Generating SPARQL executable mappings to integrate ontologies. In: ER. pp. 118–131 (2011)
60. Rivero, C.R., Hernández, I., Ruiz, D., Corchuelo, R.: On benchmarking data translation systems for semantic-web ontologies. In: CIKM. pp. 1613–1618 (2011)
61. Rivero, C.R., Hernandez, I., Ruiz, D., Corchuelo, R.: Benchmarking data exchange among semantic-web ontologies. TKDE **25**(9), 1997–2009 (2012)
62. Rivero, C.R., Hernández, I., Ruiz, D., Corchuelo, R.: Exchanging data amongst linked data applications. Knowledge and Information Systems **37**(3), 693–729 (2013)
63. Rivero, C.R., Hernández, I., Ruiz, D., Corchuelo, R.: MostoDEx: A tool to exchange rdf data using exchange samples. Journal of Systems and Software **100**, 67–79 (2015)
64. Rivero, C.R., Hernández, I., Ruiz, D., Corchuelo, R.: Mapping RDF knowledge bases using exchange samples. Knowledge-Based Systems **93**, 47–66 (2016)
65. Rivero, C.R., Schultz, A., Bizer, C., Ruiz Cortés, D.: Benchmarking the performance of linked data translation systems. In: LDOW (2012)
66. Schultz, A., Matteini, A., Isele, R., Bizer, C., Becker, C.: Ldif-linked data integration framework. In: COLD. pp. 125–130 (2011)
67. Seidenberg, J., Rector, A.: Web ontology segmentation: Analysis, classification and use. In: WWW. pp. 13–22 (2006)
68. Senellart, P., Gottlob, G.: On the complexity of deriving schema mappings from database instances. In: PODS. pp. 23–32 (2008)
69. Stuckenschmidt, H., Klein, M.: Structure-based partitioning of large concept hierarchies. In: ISWC. pp. 289–303 (2004)
70. Stuckenschmidt, H., Parent, C., Spaccapietra, S.: Modular ontologies: concepts, theories and techniques for knowledge modularization, vol. 5445. Springer (2009)
71. Suchanek, F.M., Abiteboul, S., Senellart, P.: Paris: Probabilistic alignment of relations, instances, and schema. PVLDB **5**(3), 157–168 (2011)
72. Udrea, O., Getoor, L., Miller, R.J.: Leveraging data and structure in ontology integration. In: SIGMOD. pp. 449–460. ACM (2007)
73. Velegrakis, Y., Miller, J., Popa, L.: Preserving mapping consistency under schema changes. The VLDB Journal **13**(3), 274–293 (Sep 2004)
74. Velegrakis, Y., Miller, R.J., Popa, L., Mylopoulos, J.: Tomas: A system for adapting mappings while schemas evolve. In: ICDE. p. 862 (2004)
75. Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R., Zakharyaschev, M.: Ontology-based data access: A survey. In: IJCAI. pp. 5511–5519 (2018)
76. Xiao, G., Kontchakov, R., Cogrel, B., Calvanese, D., Botoeva, E.: Efficient Handling of SPARQL optional for OBDA. In: ISWC. pp. 354–373 (2018)
77. Yan, L.L., Miller, R.J., Haas, L.M., Fagin, R.: Data-driven understanding and refinement of schema mappings. In: SIGMOD. pp. 485–496 (2001)