Inconsistency-Tolerant Ontology-Based Data Access Revisited: Taking Mappings into Account

Meghyn Bienvenu

LaBRI - CNRS & University of Bordeaux, France
meghyn.bienvenu@labri.fr

Abstract. We give a brief overview of our recent work on inconsistency-tolerant OBDA with mappings, published at IJCAI'18 [4].

Ontology-based data access (OBDA) aims to improve access to data (typically stored in relational databases) by using an ontology to provide a conceptual view of the data that describes the semantic relationships holding between different terms [10]. The focus of the work reported in this abstract is handling data inconsistencies in OBDA. It is widely acknowledged that real-world data suffers from numerous data quality issues, and errors in data are frequent. In the context of OBDA, such errors can lead to logical contradictions, in which case standard OBDA semantics (based upon classical first-order logic) trivializes. Fixing the errors by making changes to the underlying data is typically impossible, as we often do not have permission to modify the data (and even if we do, it may not be clear which modifications should be made). A solution is to adopt inconsistency-tolerant semantics, which allow meaningful answers to be obtained from inconsistent data.

The problem of querying inconsistent data using alternative semantics has been extensively studied by the database community, under the name of consistent query answering [1, 2]. In the database setting, inconsistencies arise from violations of integrity constraints, and a repair is a database that satisfies the constraints and differs minimally from the original database. Various notions of repairs have been considered, among them, subset repairs ($\subseteq$-repairs), which are inclusion-maximal consistent subsets of the database, and symmetric difference repairs ($\oplus$-repairs), which may both add and delete facts and minimize the set of such changes. Consistent query answering then amounts to computing those query answering that hold in every repair.

Recent years have seen a flurry of activity on the topic of inconsistency-tolerant query answering of DL knowledge bases, with proposals of different inconsistency-tolerant semantics [8, 7], complexity analyses of query answering under said semantics [11, 3], and some first implemented systems [6, 9, 12]. We refer readers to the survey [5] for further references. Many of the considered semantics are based upon the notion of an ABox repair, defined as a $\subseteq$-maximal subset of the ABox that is consistent w.r.t. the TBox. These include the AR semantics (which requires a query to hold w.r.t. every repair, as in consistent query answering), brave semantics (the dual of AR, which requires a query to

		AR	IAR	brave
GAV	DL-Lite	coNP-c	in AC^0	in AC^0
	PTime DLs	coNP-c	coNP-c	NP-c
GAV$^{\neg,\neq}$	DL-Lite	Π_2^p-c	Π_2^p-c	Σ_2^p-c
	PTime DLs	Π_2^p-c	Π_2^p-c	Σ_2^p-c

Fig. 1. Data complexity of conjunctive query entailment under AR, IAR, and brave semantics, for GAV and GAV$^{\neg,\neq}$ mappings. The results hold for both $\subseteq$- and $\oplus$-repairs. Lower bounds for PTime DLs hold for all DLs extending $\mathcal{EL}_\bot$.

hold in *some* repair), and IAR semantics (a strengthening of AR semantics, which queries the intersection of all repairs).

Existing works have focused on a simplified version of ontology-based data access (OBDA), in which the data is given as a set of ABox facts over the TBox signature, leaving open the question of how best to adapt repair-based semantics to handle mappings. There are (at least) two natural options: either consider the repairs of the ABox that results from applying the mappings to the data ('map-then-repair' approach), or compute repairs at the database level using the mapping and TBox to determine consistent database instances ('repair-at-source' approach). The latter approach has not been considered before in the OBDA literature, and we argue that it presents two important advantages w.r.t. the map-then-repair approach. First, it avoids the arguably undesirable situation where a repair contains ABox facts that originate from database tuples that are jointly inconsistent w.r.t. the mapping and TBox, and second, it is much more easily adapted to handle database integrity constraints.

In this work, we formalize the repair-at-source approach and investigate its computational properties. We begin by proposing a notion of OBDA repair, which is defined at the level of the database, with the mapping and ontology serving to define consistent instances. As the repairs involve modifications of the underlying (closed-world) database, we in fact consider two notions: $\subseteq$-repairs and $\oplus$-repairs. New variants of the AR, brave, and IAR semantics are then defined based upon these two kinds of OBDA repairs.

We perform a detailed study of the data complexity of OBDA under these semantics, considering both DL-Lite and the general class of 'data-tractable' DLs, which includes DLs of the $\mathcal{EL}$ family and more expressive Horn DLs like Horn-$\mathcal{SHIQ}$. We consider two forms of global-as-view (GAV) mappings, one which only allows positive atoms in mapping bodies and a more expressive variant where mapping bodies may contain negated atoms and inequalities (GAV$^{\neg,\neq}$). Mappings with complex bodies (in particular, negated atoms) are supported by existing OBDA systems and have proven useful in OBDA applications.

Our results (displayed in Figure 1) show that for plain GAV mappings, the complexity is the same as in the simple OBDA setting without mappings; in particular, the tractability results for DL-Lite under IAR and brave semantics are preserved. By contrast, adding negated atoms leads to a jump in complexity, with all problems moving to the second level of the polynomial hierarchy.

References

1. Arenas, M., Bertossi, L.E., Chomicki, J.: Consistent query answers in inconsistent databases. In: Proceedings of PODS (1999)
2. Bertossi, L.E.: Database Repairing and Consistent Query Answering. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2011)
3. Bienvenu, M.: On the complexity of consistent query answering in the presence of simple ontologies. In: Proceedings of AAAI (2012)
4. Bienvenu, M.: Inconsistency-tolerant ontology-based data access revisited: Taking mappings into account. In: Proc. of IJCAI (2018)
5. Bienvenu, M., Bourgaux, C.: Inconsistency-tolerant querying of description logic knowledge bases. In: Reasoning Web, Tutorial Lectures. Lecture Notes in Computer Science, vol. 9885, pp. 156–202. Springer (2016)
6. Bienvenu, M., Bourgaux, C., Goasdoué, F.: Querying inconsistent description logic knowledge bases under preferred repair semantics. In: Proceedings of AAAI (2014)
7. Bienvenu, M., Rosati, R.: Tractable approximations of consistent query answering for robust ontology-based data access. In: Proceedings of IJCAI (2013)
8. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant semantics for description logics. In: Proceedings of RR (2010)
9. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant query answering in ontology-based data access. Journal Web Sem. **33**, 3–29 (2015)
10. Poggi, A., Lembo, D., Calvanese, D., Giacomo, G.D., Lenzerini, M., Rosati, R.: Linking data to ontologies. Journal of Data Semantics **10**, 133–173 (2008)
11. Rosati, R.: On the complexity of dealing with inconsistency in description logic ontologies. In: Proceedings of IJCAI (2011)
12. Tsalapati, E., Stoilos, G., Stamou, G.B., Koletsos, G.: Efficient query answering over expressive inconsistent description logics. In: Proceedings of IJCAI (2016)