Data Complexity of Querying Description Logic Knowledge Bases under Cost-Based Semantics (Extended Abstract)

Meghyn Bienvenu[1] , **Quentin Manière**[2]
[1]Université de Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Talence, France
[2]LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France
meghyn.bienvenu@labri.fr, quentin.maniere@inria.fr

Abstract

This extended abstract summarizes our recent AAAI 2026 paper on the data complexity of querying description logic knowledge bases under cost-based semantics (Bienvenu and Manière 2026). Our complexity analysis notably pinpoints the precise complexity of optimal-cost certain answer semantics (for which no non-trivial upper bound was known) and exhibits the first tractability results for cost-based semantics, for the case of DL-Lite$^{\mathcal{H}}_{\text{bool}}$ ontologies with a fixed cost bound.

Context & Motivation Ontology-mediated query answering (OMQA) has been extensively studied within the KR and database communities as a means of improving data access by exploiting semantic information provided by an ontology (Poggi et al. 2008; Bienvenu and Ortiz 2015; Xiao et al. 2018). Ontologies are typically formulated in decidable fragments of first-order logic (FO), with description logics (DLs) being a popular choice (Baader et al. 2017). Given an ontology (or TBox in DL parlance) $\mathcal{T}$, a dataset (or ABox) $\mathcal{A}$, and a query $q(\vec{x})$, the OMQA task boils down to finding the certain answers, i.e. tuples of constants $\vec{a}$ for which the instantiated query $q(\vec{a})$ is entailed from the knowledge base (KB) $(\mathcal{T}, \mathcal{A})$. Observe that under this classical semantics, when the input KB is inconsistent, every answer tuple is trivially a certain answer, so OMQA trivializes.

A prominent approach to tackling this issue is to adopt alternative inconsistency-tolerant semantics in order to be able to extract meaningful information from inconsistent KBs, cf. (Lembo et al. 2010) and survey (Bienvenu and Bourgaux 2016). Many of these semantics are based upon repairs, defined as inclusion-maximal consistent subsets of the ABox. For example, the AR semantics considers the query answers that hold in all repairs, while the brave semantics returns those answers holding in at least one repair. Note that the line of work on repair-based semantics targets scenarios in which the TBox axioms are deemed fully reliable, so inconsistencies derive solely from errors in the ABox. However, in practice, it can be useful to allow for TBox axioms which typically hold but may admit rare exceptions. Such 'soft' ontology axioms can be addressed qualitatively, using generalized notions of repair that have been proposed for existential rule ontologies (Eiter, Lukasiewicz, and Predoiu 2016), or employing non-monotonic extensions of DLs that support defeasible axioms cf. (Bonatti, Lutz, and Wolter 2009; Giordano et al. 2013; Britz et al. 2021). Another option is to adopt a quantitative approach, using the recently proposed cost-based semantics for DL KBs (Bienvenu, Bourgaux, and Jean 2024), henceforth abbreviated to (BBJ 2024).

Cost-Based Semantics In a nutshell, the idea is to use a *weight function* $\omega : \mathcal{T} \cup \mathcal{A} \mapsto \mathbb{N}_{>0} \cup \{\infty\}$ to annotate TBox axioms and ABox assertions in a KB $(\mathcal{T}, \mathcal{A})$ with (possibly infinite) weights. These weights are then used to assign a cost, $\omega(\mathcal{I})$, to each interpretation $\mathcal{I}$, by summing up the weights of the axioms and assertions that are violated in $\mathcal{I}$ (with each distinct violation of a TBox axiom incurring a separate penalty). The *optimal cost* of a *weighted knowledge base* (WKB) $\mathcal{K}_\omega$ is the minimal cost over all interpretations: $optc(\mathcal{K}_\omega) = \mathsf{min}_{\mathcal{I}}(\omega(\mathcal{I}))$. To query a WKB, we may either consider the set of interpretations achieving the optimal cost, or we may fix a cost bound k and consider all interpretations having cost at most k. We can then define the sets of certain and possible answers as those answers that hold respectively in all or some interpretation(s) whose cost does not exceed the optimal cost / given cost bound k, yielding four different entailment relations ($\models^{opt}_c, \models^{opt}_p, \models^k_c, \models^k_p$).

As noted in (BBJ 2024), the optimal-cost certain answer semantics ($\models^{opt}_c$) generalizes both the classical certain answer semantics and the AR semantics based upon weighted ABox repairs. Increasing the cost bound k beyond the optimal cost allows one to identify answers that are robust in the sense that they hold not only for the optimal-cost interpretations. Optimal- and bounded-cost possible answers generalize query satisfiability and can serve to compare candidate answers based upon their incompatibility with the KB.

Existing Results & Open Questions The computational complexity of querying inconsistent weighted KBs under cost-based semantics was investigated in (BBJ 2024). The complexity analysis was fairly comprehensive, considering both combined and data complexity, conjunctive (CQs) and instance queries (IQs), and DLs ranging from the lightweight DL $\mathcal{EL}_\bot$ to the expressive DL $\mathcal{ALCO}$. One important question that was left open, however, was the data complexity of optimal-cost certain semantics (arguably the most useful of the semantics), for which no non-trivial upper bound was provided. Moreover, as the considered DLs allow neither inverse roles nor role inclusions, they do not yield any results for DLs of the DL-Lite family (Calvanese

	BCS^k, IQA_p^k, CQA_p^k	IQA_c^k	CQA_c^k	BCS, IQA_p, CQA_p	IQA_c, CQA_c	$\mathrm{IQA}_{p,c}^{opt}$, $\mathrm{CQA}_{p,c}^{opt}$
$\mathcal{EL}_\bot$ / $\mathcal{ALCHIO}$	NP ‡	coNP ‡	coNP ‡	NP †	coNP †	Δ_2^p †
DL-Lite$_{\mathrm{core}}$ / DL-Lite$_{\mathrm{bool}}^{\mathcal{H}}$	in AC^0	in AC^0	coNP	NP	coNP	Δ_2^p

Table 1: All results are completeness results, unless stated otherwise. †: lower bound from (BBJ 2024). ‡: lower bound for $k \geq 3$ from (BBJ 2024), improved to $k \geq 1$ in the reported work. For CQA, lower bounds already hold for connected acyclic Boolean CQs. Results hold both for binary and unary encoding of weights, except for Δ_2^p-hardness (only for binary encoding).

et al. 2007), which are the most commonly utilized in the context of OMQA.

Data Complexity Study The preceding considerations motivated us to embark on a more detailed data complexity analysis of cost-based semantics, considering various DL-Lite dialects as well as expressive DLs up to $\mathcal{ALCHIO}$. Our study considers the following decision problems. *Bounded cost satisfiability* (BCS) takes as input a WKB $\mathcal{K}_\omega$ and an integer k and decides whether there exists $\mathcal{I}$ with $\omega(\mathcal{I}) \leq k$. BCS^k is similarly defined but k is fixed and not part of the input. *Bounded-cost certain (resp. possible) BCQ entailment* (CQA_c / CQA_p) takes as input a WKB $\mathcal{K}_\omega$, Boolean CQ q and integer k and decides whether $\mathcal{K}_\omega \models_c^k q$ (resp. $\mathcal{K}_\omega \models_p^k q$). CQA_c^k and CQA_p^k are similarly defined, but k is fixed. Finally, *optimal-cost certain (resp. possible) BCQ entailment* (CQA_c^{opt} / CQA_p^{opt}) takes as input a WKB $\mathcal{K}_\omega$ and a BCQ q and decides if $\mathcal{K}_\omega \models_c^{opt} q$ (resp. $\mathcal{K}_\omega \models_p^{opt} q$). We also consider the restrictions of the query entailment problems to the case of instance queries, denoted by IQA_c, IQA_p, IQA_c^k, IQA_p^k, IQA_c^{opt} and IQA_p^{opt} respectively. Note that to better understand the dependency of the complexity on the cost bound k, we distinguish problems based upon whether k is treated as a fixed constant or as part of the input (BCS^k vs. BCS). Thus, we measure complexity with regards to $|\mathcal{A}|$ and, for bounded-cost tasks, also the cost bound k.

Main Results Our results summarized in Table 1. A first major contribution of our work is to establish a Δ_2^p upper bound for the optimal-cost certain and possible semantics, matching an existing lower bound for $\mathcal{EL}_\bot$ and a new lower bound we show for DL-Lite$_{\mathrm{core}}$. This result was obtained by using an intricate quotient construction to establish a small interpretation property, which crucially does not depend on the considered cost. We also strengthen a number of existing lower bounds for the bounded-cost semantics by showing that they hold even for cost bound $k = 1$, as well as providing some new lower bounds for DL-Lite$_{\mathrm{core}}$. Finally, our most challenging and surprising technical result is to show that if we consider DL-Lite$_{\mathrm{bool}}^{\mathcal{H}}$ ontologies and a fixed cost bound, then certain answers for IQs and possible answers for CQs can be computed using first-order rewriting and thus enjoy the lowest possible data complexity (AC^0).

Perspectives We expect that our upper bounds can be adapted to also handle negative role inclusions (to cover also DL-Lite$_{\mathcal{R}}$). For DLs with functionality or number restrictions, it does not suffice to work with finite interpretations, so wholly different methods are required. Developing a practical implementation of the FO-rewritings for the identified tractable cases is another interesting direction.

References

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Bienvenu, M., and Bourgaux, C. 2016. Inconsistency-tolerant querying of description logic knowledge bases. In *Reasoning Web Tutorial Lectures*.

Bienvenu, M., and Manière, Q. 2026. Data complexity of querying description logic knowledge bases under cost-based semantics. In *Proceedings of AAAI*.

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web Tutorial Lectures*.

Bienvenu, M.; Bourgaux, C.; and Jean, R. 2024. Cost-based semantics for querying inconsistent weighted knowledge bases. In *Proceedings of KR*.

Bonatti, P. A.; Lutz, C.; and Wolter, F. 2009. The complexity of circumscription in DLs. *J. Artif. Intell. Res.* 35:717–773.

Britz, K.; Casini, G.; Meyer, T.; Moodley, K.; Sattler, U.; and Varzinczak, I. 2021. Principles of KLM-style defeasible description logics. *ACM Trans. Comput. Log.* 22(1):1:1–1:46.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning (JAR)*.

Eiter, T.; Lukasiewicz, T.; and Predoiu, L. 2016. Generalized consistent query answering under existential rules. In *Proceedings of KR*.

Giordano, L.; Gliozzi, V.; Olivetti, N.; and Pozzato, G. L. 2013. A non-monotonic description logic for reasoning about typicality. *Artif. Intell.* 195:165–202.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D. F. 2010. Inconsistency-tolerant semantics for description logics. In *Proceedings of RR*.

Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; and Rosati, R. 2008. Linking data to ontologies. *Journal of Data Semantics* 10:133–173.

Xiao, G.; Calvanese, D.; Kontchakov, R.; Lembo, D.; Poggi, A.; Rosati, R.; and Zakharyaschev, M. 2018. Ontology-based data access: A survey. In *Proceedings of IJCAI*.