Controlled Query Evaluation with Epistemic Dependencies: Algorithms and Experiments (Extended Abstract)

Lorenzo Marconi[1], Flavia Ricci[1] and Riccardo Rosati[1]

[1]Sapienza University of Rome, Italy

Abstract

This work summarizes our paper accepted to the 24th International Semantic Web Conference focusing on Controlled Query Evaluation over Description Logics ontologies. We express the data protection policy using epistemic dependencies (EDs), and use optimal ground atom (GA) censors as tools for exposing the facts entailed by the ontology in a maximal, policy-compliant way. We study the complexity of answering Boolean unions of conjunctive queries with respect to the intersection of all optimal GA censors. We identify a class of EDs for which the examined entailment problem over DL-Lite$_\mathcal{R}$ ontologies is first-order rewritable, and we empirically validate the efficiency of our method.

Keywords

Description Logics, Confidentiality Preservation, Query Answering, First-Order Rewritability

Providing vast amounts of structured and semantically rich information, the use of ontologies poses new challenges in knowledge management and data security. Indeed, while these technologies offer advanced tools for querying and inference, they also raise important concerns about the possible unintentional disclosure of sensitive information: seemingly innocuous queries can, when combined with ontological knowledge, lead to disclosing confidential data.

Controlled Query Evaluation (CQE) [1, 2, 3, 4] is a framework that handles this issue by providing access via queries only to data that conforms to a *data protection policy* expressed in terms of logical formulas. A central notion in CQE is that of *censor*, which represents the part of the (logical consequences of the) ontology that can be safely disclosed to the end user. In particular, we focus on *GA censors* [5], which are sets of ground atoms entailed by the ontology and compliant with the policy (a more formal definition is provided below).

In CQE, policies are typically defined as sets of *denials*, i.e. first-order (FO) sentences of the form $q \rightarrow \bot$, where q is a *Boolean conjunctive query* (BCQ). Such formulas are used to define the information that must be kept confidential: the system is required to guarantee that users cannot infer that the sentence q is entailed by the ontology. The recent work [6], though employing a notion of censor that differs from ours, described how a richer language of rules, called *epistemic dependencies* (EDs) [7], can be used for data protection purposes. EDs are a special case of EQL-Lite(CQ) [8] sentences, and are formally defined as follows.

Definition 1. *An* epistemic dependency (ED) *is a sentence τ of the form*

$$\forall \mathbf{x}_1, \mathbf{x}_2 \, (K q_b(\mathbf{x}_1, \mathbf{x}_2) \rightarrow K q_h(\mathbf{x}_2))$$

where $q_b(\mathbf{x}_1, \mathbf{x}_2)$ is a CQ with free variables $\mathbf{x}_1 \cup \mathbf{x}_2$, $q_h(\mathbf{x}_2)$ is a CQ with free variables $\mathbf{x}_2$, and K is an epistemic operator.

In the same spirit as in the aforementioned work, we use EDs as disclosure rules to govern the publication of data. Intuitively, if σ is any substitution assigning the universal variables of an ED τ to constants, the fact that the ontology entails $\sigma(q_b)$ may only be disclosed if $\sigma(q_h)$ can also be made

DL 2025: 38th International Workshop on Description Logics, September 3–6, 2025, Opole, Poland

marconi@diag.uniroma1.it (L. Marconi); ricci.1883245@studenti.uniroma1.it (F. Ricci); rosati@diag.uniroma1.it (R. Rosati)
0000-0001-9633-8476 (L. Marconi); 0000-0002-7697-4958 (R. Rosati)

public. More formally, we say that an FO theory Φ *satisfies* an ED τ (in symbols $\Phi \models_{\mathsf{EQL}} \tau$) if, for every assignment σ of the free variables of q_b with constants, if $\Phi \models \sigma(q_b)$ then $\Phi \models \sigma(q_h)$. If Φ satisfies all EDs of a policy $\mathcal{P}$, then we say that Φ *satisfies* $\mathcal{P}$ (in symbols $\Phi \models_{\mathsf{EQL}} \mathcal{P}$). Note that EDs can also be used to express denials, as one can have $q_h = \bot$.

To describe our ontology, we rely on Description Logics (DLs) [9].[1] A DL ontology is an FO theory $\mathcal{T} \cup \mathcal{A}$, where $\mathcal{T}$ (the TBox) is a set of intensional axioms and $\mathcal{A}$ (the ABox) is a set of facts. Hereinafter, we call *CQE instance* the triple $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a TBox, $\mathcal{P}$ is a policy and $\mathcal{A}$ is an ABox. Moreover, we define an *optimal GA censor* for $\mathcal{E}$ as any maximal (w.r.t. set inclusion) set of ground atoms $\mathcal{C}$ such that $\mathcal{T} \cup \mathcal{A} \models \mathcal{C}$ and $\mathcal{T} \cup \mathcal{C} \models_{\mathsf{EQL}} \mathcal{P}$. The next example shows the case of a policy consisting of EDs coupled with a DL ontology.

Example 1. *A company establishes that all the salaries of its employees must be kept undisclosed, except for those who hold a managerial position; moreover, it requires that consensual personal relationships between managers and their team members not be publicly revealed. This policy can be formally represented through the following set of EDs:*

$$\mathcal{P} = \{\, \forall x, y\, (K\,\mathit{salary}(x, y) \rightarrow K\,\mathit{manager}(x)),\\ K\exists x, y\, (\mathit{managerOf}(x, y) \wedge \mathit{consRel}(x, y)) \rightarrow K\bot \,\}$$

where manager is a unary predicate indicating that an individual is a manager, and salary, consRel and managerOf are binary predicates modelling, respectively, the salary level of a person, the consensual relationship between two individuals and the relationship where one individual manages another. By employing the existential quantifier, the second ED asserts that, for any manager (or employee), the fact that a consensual relationship exists with one of their employees (or managers) must itself be concealed—not just the two parties' identities.

In addition, our knowledge about the company is defined by the following ontology:

- *A TBox* $\mathcal{T} = \{\exists \mathit{managerOf} \sqsubseteq \mathit{manager}, \mathit{manager} \sqsubseteq \exists \mathit{respDept}\}$*, meaning that* (i) *if one person manages another, then he or she is a manager; and* (ii) *every manager is responsible for at least one department.*
- *An ABox* $\mathcal{A} = \{\mathit{managerOf}(\mathit{lucy}, \mathit{tom}), \mathit{consRel}(\mathit{lucy}, \mathit{tom}), \mathit{salary}(\mathit{lucy}, \mathit{150k}), \mathit{salary}(\mathit{tom}, \mathit{50k})\}$*, which describes a situation in which Lucy manages Tom, they have a consensual relationship, and they receive a salary of $150,000 and $50,000, respectively.*

Intuitively, a GA censor certainly does not contain either $\mathit{managerOf}(\mathit{lucy}, \mathit{tom})$ *or* $\mathit{consRel}(\mathit{lucy}, \mathit{tom})$ *(in order not to violate the denial) and, in any case, it does not contain the fact* $\mathit{salary}(\mathit{tom}, \mathit{50k})$ *(as Tom is not a manager). Moreover, every optimal GA censor includes both* $\mathit{manager}(\mathit{lucy})$ *and* $\mathit{salary}(\mathit{lucy}, \mathit{150k})$*, because the fact that Lucy is a manager can be logically deduced from the ontology and, by revealing this information, knowing also her salary does not violate the policy.* ∎

In this setting, our objective is to answer *Boolean unions of conjunctive queries* (BUCQs) based on a formal entailment semantics that balances information disclosure with policy compliance. In particular, we investigate the problem of checking whether a BUCQ is entailed by the TBox and the intersection of all the optimal GA censors. This task, known as IGA-entailment, consists in checking whether $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models q$, where $\mathcal{C}_{\mathsf{IGA}}$ is the intersection of all the optimal GA censors of a CQE instance $\mathcal{E}$ (in this case, we write $\mathcal{E} \models_{\mathsf{IGA}} q$).

The work [12] showed that IGA-entailment is *FO-rewritable* in the case of DL-Lite$_{\mathcal{R}}$ ontologies and policy consisting of denials. Formally, this means that the IGA-entailment of a BUCQ q can be checked through an algorithm that first rewrites q into an FO query q_r that does not depend on the ABox and, in a second moment, evaluates q_r over the ABox. From a theoretical perspective, such a property ensures a nice computational behavior, as its direct implication is that the problem of determining the IGA-entailment of a BUCQ with DL-Lite$_{\mathcal{R}}$ ontologies and denials is in AC^0 in data complexity [13].

[1] An up-to-date overview of CQE in the context of DLs can be found in [10, 11].

On the practical side, experiments under these conditions were carried out in [14], within the OBDA framework.

We aim to extend this scenario to accommodate policies defined using EDs while preserving the *FO-rewritability* property. In particular, we focus on the class of *full EDs*, i.e. EDs whose head contains no existential variable. This class enjoys a desirable property related to security: given a CQE instance $\mathcal{E}$ whose policy is made of full EDs, the intersection of all the optimal GA censors for $\mathcal{E}$ is still a GA censor for $\mathcal{E}$. We also show that, in general, this property does not hold. We exclude, however, the FO-rewritability of IGA-entailment for this class of dependencies, by providing the following complexity result (which holds even in the case the TBox is empty):

Theorem 1. *IGA-entailment is coNP-hard in data complexity in the case of full EDs.*

We thus identify a sufficient condition for full EDs for which IGA-entailment in the case of DL-Lite$_{\mathcal{R}}$ ontologies remains FO-rewritable. Specifically, we require the policy $\mathcal{P}$ to be such that the set Σ of TGDs derived (in the natural way) from $\mathcal{P}$ and from (the inclusion assertions of) $\mathcal{T}$ is *UCQ-rewritable* i.e., given any CQ $q(\mathbf{x})$, there exists a UCQ q' such that, for every set $\mathcal{F}$ of facts and for every ground substitution σ of the free variables of q, $\Sigma \cup \mathcal{F} \models \sigma(q)$ iff $\mathcal{F} \models \sigma(q_r)$ for some $q_r(\mathbf{x}) \in q'$. In this case, we say that $\mathcal{P}$ is *expandable* w.r.t. $\mathcal{T}$.

Theorem 2. *IGA-entailment is FO-rewritable, and thus in* AC^0 *in data complexity, for DL-Lite*$_{\mathcal{R}}$ *TBoxes and policies that are full and expandable w.r.t. the coupled TBox.*

As a final step, focusing on two categories of EDs satisfying this requirement—namely the *acyclic* full (where acyclicity condition is defined as in [6]) and the *linear* full EDs—we carried out experiments to assess the practical viability of our rewriting procedure. We developed a tool that transforms a SPARQL BUCQ into an FO query q_r using only the provided TBox and policy, and then executes q_r over an SQL database storing the ABox. As the above theorem pertains to DL-Lite$_{\mathcal{R}}$, we employed the OWL 2 QL ontology and the 10 queries for OWL 2 QL provided by the OWL2Bench benchmark [15]. The outcome of our experiments, conducted on a standard laptop with an Intel i7 @1.8 GHz processor and 16GB of RAM, is summarized in Table 1. For every test case and every query, we report the rewriting time (t_r) and the evaluation time (t_e) expressed in milliseconds, other than the number of returned tuples (#). In the table, $\mathcal{P}_\emptyset$, $\mathcal{P}_\mathrm{a}$, $\mathcal{P}_\mathrm{b}$, $\mathcal{P}_\mathrm{a}^-$, and $\mathcal{P}_\mathrm{b}^-$ refer to the empty policy, a full acyclic policy, a full linear policy, and, respectively, their "reduced" versions. Moreover, o2b$_i$ with $i \in \{5, 10\}$ refers to the ontology included in the benchmark containing axioms and ground data about i fictitious universities.

We observe that (i) in most cases the evaluation time t_e is acceptable (of the order of seconds), although the seventh query takes several minutes; (ii) the rewriting time t_r—which is nearly identical for o2b$_5$ and o2b$_{10}$ as it does not depend on the ABox—is often negligible and never exceeds three seconds; (iii) for both acyclic and binary policies, t_r values for smaller and larger policies are of comparable magnitude; and (iv) binary policies tend to remove more tuples than acyclic ones, likely because EDs with fewer atoms in their body are more easily "activated".

All the results with full proofs are reported in the extended version of the paper [16].

Acknowledgments

This work was partially supported by: projects FAIR (PE0000013) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU; the MUR PRIN 2022LA8XBH project Polar (POLicy specificAtion and enfoRcement for privacy-enhanced data management); by the EU under the HORIZON.2.1.5 project dAIbetes (grant id. 101136305); and by projects SEED PNR 2021 and SEED PNR 2022 funded by Sapienza Università di Roma.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 in order to: Grammar and spelling check. After using this service, the authors reviewed and edited the content as needed and take full

Table 1
The results of our experiments. For every test case and every query, we report the rewriting time (t_r) and evaluation time (t_e) expressed in milliseconds, plus the number of returned tuples (#).

Query		o2b$_5$					o2b$_{10}$				
		$\mathcal{P}_\emptyset$	$\mathcal{P}_\mathsf{a}^-$	$\mathcal{P}_\mathsf{a}$	$\mathcal{P}_\mathsf{b}^-$	$\mathcal{P}_\mathsf{b}$	$\mathcal{P}_\emptyset$	$\mathcal{P}_\mathsf{a}^-$	$\mathcal{P}_\mathsf{a}$	$\mathcal{P}_\mathsf{b}^-$	$\mathcal{P}_\mathsf{b}$
q_1	t_r	19	19	20	53	63	15	20	28	51	63
	t_e	513	526	547	683	958	822	560	964	1394	1085
	#	9228	9228	9228	1367	334	19782	19782	19782	2948	730
q_2	t_r	50	249	336	476	511	65	243	386	873	553
	t_e	81	8688	7335	174	206	269	39426	51181	876	402
	#	18872	18736	14829	5957	5957	44190	43889	33193	13009	13009
q_3	t_r	25	41	42	40	109	38	32	43	62	112
	t_e	4	8	6	2	4	8	5	9	5	6
	#	34	34	34	34	28	75	75	75	75	64
q_4	t_r	21	40	35	52	133	33	33	49	68	97
	t_e	6	7	2	4	3	7	5	7	5	3
	#	0	0	0	0	0	0	0	0	0	0
q_5	t_r	21	554	473	250	319	29	482	917	451	334
	t_e	17	30699	29679	1823	1003	44	110763	124305	5056	1283
	#	3574	2020	2020	952	264	6564	3676	3676	1696	394
q_6	t_r	18	114	115	161	116	25	148	109	263	193
	t_e	59	60661	28564	81	230	235	283020	141173	333	282
	#	16236	15834	7811	0	0	35889	35075	17481	0	0
q_7	t_r	75	2029	1976	1378	1847	88	2205	2244	2222	1801
	t_e	66	198641	187908	403	496	334	993701	1119538	1343	812
	#	5489	5489	5091	5489	3292	11969	11969	10971	11969	7241
q_8	t_r	22	52	48	257	263	21	49	44	399	272
	t_e	62	75	63	615	647	186	112	143	1548	963
	#	17904	17904	17904	14668	14668	39278	39278	39278	32350	32350
q_9	t_r	135	1711	2302	1765	3430	195	1778	2290	2877	3144
	t_e	31	1412	1313	137	129	93	10230	10944	598	332
	#	1698	1539	1539	0	0	3434	3196	3196	0	0
q_{10}	t_r	22	1243	1220	184	337	32	1351	1433	380	477
	t_e	78	676	3	110	159	297	4753	3	478	270
	#	642	122	0	642	144	1413	258	0	1413	335

responsibility for the publication's content.

References

[1] J. Biskup, For unknown secrecies refusal is better than lying, Data and Knowledge Engineering 33 (2000) 1–23.

[2] J. Biskup, P. A. Bonatti, Controlled query evaluation for enforcing confidentiality in complete information systems, Int. J. Inf. Sec. 3 (2004) 14–27.

[3] P. A. Bonatti, L. Sauro, A confidentiality model for ontologies, in: Proc. of the 12th Int. Semantic Web Conf. (ISWC), volume 8218 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 17–32.

[4] B. Cuenca Grau, E. Kharlamov, E. V. Kostylev, D. Zheleznyakov, Controlled query evaluation over OWL 2 RL ontologies, in: Proc. of the 12th Int. Semantic Web Conf. (ISWC), 2013, pp. 49–65.

[5] G. Cima, D. Lembo, R. Rosati, D. F. Savo, Controlled query evaluation in description logics through consistent query answering, Artificial Intelligence 334 (2024) 104176.

[6] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, Enhancing controlled query evaluation through epistemic policies, in: Proc. of the 33th Int. Joint Conf. on Artificial Intelligence (IJCAI), Int. Joint Conf. on Artificial Intelligence Organization, 2024, pp. 3307–3314.

[7] M. Console, M. Lenzerini, Epistemic integrity constraints for ontology-based data management, in: Proc. of the 34th AAAI Conf. on Artificial Intelligence (AAAI), AAAI Press, 2020, pp. 2790–2797.
[8] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, EQL-Lite: Effective first-order query processing in description logics, in: Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI), Morgan Kaufmann Publishers Inc., 2007, pp. 274–279.
[9] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation and Applications, 2nd ed., Cambridge University Press, 2007.
[10] P. A. Bonatti, A false sense of security, Artificial Intelligence 310 (2022).
[11] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, A gentle introduction to controlled query evaluation in DL-Lite ontologies, Springer Nature Computer Science 5 (2024) 335.
[12] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, Indistinguishability in controlled query evaluation over prioritized description logic ontologies, J. of Web Semantics 84 (2025) 100841.
[13] S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, Addison Wesley Publ. Co., 1995.
[14] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, Controlled query evaluation in ontology-based data access, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I, volume 12506 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 128–146.
[15] G. Singh, S. Bhatia, R. Mutharaju, OWL2Bench: A benchmark for OWL 2 reasoners, in: Proc. of the 19th Int. Semantic Web Conf. (ISWC), volume 12507 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 81–96.
[16] L. Marconi, F. Ricci, R. Rosati, CQE under epistemic dependencies: Algorithms and experiments (extended version), 2025. URL: https://arxiv.org/abs/2507.17487.