SN Computer Science (2024) 5:335
https://doi.org/10.1007/s42979-024-02652-4

SURVEY ARTICLE

A Gentle Introduction to Controlled Query Evaluation in DL-Lite Ontologies

Gianluca Cima[1] · Domenico Lembo[1] · Lorenzo Marconi[1] · Riccardo Rosati[1] · Domenico Fabio Savo[2]

Received: 12 May 2023 / Accepted: 24 January 2024

Abstract
Controlled query evaluation (CQE) is an approach for confidentiality-preserving query answering where a function called censor alters query answers so that users can never infer data that are protected by a policy given in terms of logic formulae. In this paper, we review some foundational results we have recently found in the context of CQE over Description Logic ontologies. In more detail, we discuss the main characteristics of two notions of censor, CQ censor and GA censor, focusing on the computational complexity of query answering and on the notion of indistinguishability. The latter is a desirable property imposing that a censor always makes a user believe that the underlying data instance might not contain confidential data. As for computational aspects, we characterize the data complexity of answering conjunctive queries for the relevant and practical case of DL-Lite$_{\mathcal{R}}$ ontologies. Since neither CQ censors nor GA censors enjoy both indistinguishability and tractability of query answering in the analyzed setting, we finally recall the notion of IGA censors, a sound approximation of GA censors which instead enjoys both properties, thus paving the way for robust and practical CQE for DL-Lite$_{\mathcal{R}}$ ontologies.

Keywords Description Logics · Information disclosure · Computational complexity

Introduction

Confidential data are data meant to be protected from unauthorized accesses and disclosure. Examples include personal information, e.g., individual phone numbers, street addresses, biometric identifiers, or business data, such as trade secrets, company payment details, customer profiles, to mention a few. Confidentiality preservation is the task of maintaining data protected from breaches, i.e., situations in which confidential data are accessed without permission.

The actions that can be undertaken to prevent data breaches range from general infrastructure security measures, such as the use of firewalls or intrusion detection systems [33], to privileges management for access control, i.e., establishing who can access non-public data, for what purposes, and under which conditions [12], to data perturbation and filtering, such as data anonymization [2], encryption [4], or the use of (virtual) views [38] that show to a user only data that she is authorized to see.

Among the approaches based on some form of data filtering to preserve confidentiality, Controlled Query Evaluation (CQE) has the characteristic to be fully declarative and grounded on logic. In CQE, confidentiality requirements are expressed through a *policy*, i.e., a set of logical assertions that specify the information that must be kept secret, and a function, called *censor*, that is used to modify answers to queries so that confidential data cannot be inferred by the users on the basis of the answers they get and the knowledge they have.

Censors for confidentiality-preserving query answering have been first studied in [34] for complete propositional

This article is part of the topical collection "Advances on Web Information Systems and Technologies" guest-edited by Joaquim Filipe, Francisco José Domínguez Mayo and Massimo Marchiori.

✉ Riccardo Rosati
rosati@diag.uniroma1.it

Gianluca Cima
cima@diag.uniroma1.it

Domenico Lembo
lembo@diag.uniroma1.it

Lorenzo Marconi
marconi@diag.uniroma1.it

Domenico Fabio Savo
domenicofabio.savo@unibg.it

[1] Sapienza University of Rome, Roma, Italy

[2] University of Bergamo, Bergamo, Italy

Published online: 27 March 2024

databases. The framework has then revived years later and extensively investigated for various forms of censors [5–8], even for the case of incomplete databases [9, 13]. More recently, CQE has been tackled in the context of Description Logic (DL) ontologies [14, 22, 23], where data are stored in the ontology component called ABox, whereas a second component, called TBox, captures intensional knowledge.

In this paper, we review some advancements we achieved in the last years in the latter field. More precisely, we recall the general framework for CQE given in [28] and mainly focus on two kinds of censors proposed in that paper, namely *CQ censors* and *GA censors*, recalling their crucial characteristics, previously shown in [17, 18, 20, 28]. In a nutshell, a CQ censor (resp. GA censor) is a function returning a subset of all Boolean conjunctive queries (resp. of all ground atoms, i.e., facts) that are inferred by the ontology and are compliant with the policy. The policy we consider is expressed as a set of denial assertions, each one specifying that the answer to a Boolean conjunctive query is confidential. Thus, a censor c is compliant with the policy if c perturbs query answering so that a user can never infer that a query in the policy is inferred by the ontology even though in fact it is. A censor is *optimal* if it returns a containment-maximal set among the sets returned by all possible censors (either CQ censors or GA censors). Typically, given an ontology and a policy, several optimal censors exist. Then, let $\mathcal{S}$ be the set returned by a censor c, let $\mathcal{T}$ be the TBox of the ontology, and let q_u be a user query: the answer to q_u under censor c is the one inferred by $\mathcal{T} \cup \mathcal{S}$.

Our framework moves from the one proposed in [22, 23], but revises it by defining CQE as a form of skeptical reasoning over all optimal censors. More precisely, we define CQE as the problem of computing, for every user query q_u, the answers to q_u that are in the intersection of the answers computed under all optimal censors. This notion of CQE distinguishes our proposal from previous work, which mainly focuses on the problem of constructing one optimal censor. The latter approach has often to make an arbitrary choice on the censor to use for protecting confidential data, whereas we avoid such a discretionary decision through reasoning over all optimal censors.

In our presentation, we start by CQ censors, which correspond to the confidentiality-preserving censors previously considered in [22, 23]. According to the terminology proposed in [14], such censors allow to realize a Simple Confidentiality Model (SCM), in which the aim is that entailed secrets are filtered out. As noted in [11, 14], an SCM is not always able to protect data from sophisticated attacks, where, e.g., the attacker possesses background knowledge and/or meta-knowledge that may help to reconstruct the confidential part of the ontology. Robustness with respect to these kinds of attacks may be achieved through censors enjoying the so-called *indistinguishability* property. Intuitively, a censor satisfies this property if it modifies answers to queries in such a way that a user can always conceive that the underlying ABox does not contain secrets (this hypothetical ABox is thus indistinguishable from the real one, which possibly contains secrets). As we have shown in [18, 20], CQ censors do not satisfy indistinguishability whereas GA censors do. These last results have strengthened our interest in GA censors, towards the identification of a setting allowing for both robust and practically realizable CQE. In this paper, we thus also review the characteristics of GA censors and discuss whether they allow us to achieve our aim.

Among such characteristics, computational complexity of reasoning turns out to be of paramount importance, because of its inherent connection to practical feasibility. In our studies, we have thus thoroughly investigated the data complexity of CQE [17, 18, 20, 28], i.e., the computational complexity established with respect to the size of the data only [36], which is the complexity measure mainly considered in data-intensive applications. In particular, we recall here the results we devised about the data complexity of answering conjunctive queries according to our revised notion of CQE, under both CQ censors and GA censors, for ontologies specified in the logic DL-Lite$_{\mathcal{R}}$ [15]. We remark that standard conjunctive query answering in DL-Lite$_{\mathcal{R}}$ ontologies is computationally tractable (i.e. solvable in polynomial time), specifically in AC^0 in data complexity [15], that is the same data complexity of evaluating an SQL query over a database. By virtue of this characteristic, this logic is frequently adopted in applications in which large datasets are managed through an ontology [29], and has led to the definition of OWL 2 QL [31], the counterpart of DL-Lite$_{\mathcal{R}}$ within OWL 2, the standard W3C ontology language [32]. As we have shown in [28], CQE for DL-Lite$_{\mathcal{R}}$ under CQ censors is in PTIME in data complexity, whereas it is coNP-hard for GA censors (it is in AC^0 only for queries that are ground atoms, i.e. for the so-called instance checking problem). To re-gain tractability, in [17] we have proposed, in the context of Ontology-based Data Access, a new notion of censor, called *IGA censor*, which is a well-founded sound approximation of GA censors. In this paper we thus recall this additional form of censor, which is particularly interesting because, as shown in [18], it enjoys indistinguishability, and at the same time guarantees tractable query answering, namely in AC^0 in data complexity [17, 18].

We remark that the present paper is intended to be an introduction to the problem of CQE in DL ontologies, and in particular in DL-Lite$_{\mathcal{R}}$, built through basic definitions, examples, and foundational results. To this aim, we gather together into this single article some material already appeared in slightly different shapes in various different papers [17, 18, 20, 28]. To complete our treatment, we also provide here some new results on the so-called *ICQ censors*, which are censors obtained by intersecting all possible

optimal CQ censors, as IGA censors are obtained through the intersection of all optimal GA censors, and, as IGA censors, have the characteristic of being unique, given an ontology TBox and a policy.

We finally note that confidentiality-preservation in DLs and ontologies have been studied also through approaches different from CQE. For example, in [25] the authors propose a technique to import an ontology $\mathcal{K}_h$ into another ontology $\mathcal{K}_v$ and reason on $\mathcal{K}_h \cup \mathcal{K}_v$ using the axioms in $\mathcal{K}_v$ but just asking queries expressed over a subset of the signature of $\mathcal{K}_h$, being the rest of the signature of $\mathcal{K}_h$ non-accessible to users for confidentiality reasons. Authorization views in DLs are instead studied in [16, 35], whereas a probabilistic logic-based framework is considered in [21], and anonymization for Linked Data is addressed in [24]. Information disclosure in the context of Ontology-based Data Access is instead studied in [3].

The rest of the paper is organized as follows. In the next section, we provide some preliminaries. In the subsequent section, we deal with CQ and ICQ censors, followed by which, we attack the case of GA and IGA censors, in both sections providing complexity results on CQE for DL-Lite$_\mathcal{R}$ ontologies and discussing the indistinguishability property for all these kinds of censor. Finally, we provide some final discussions and conclude the paper.

Preliminaries

We use standard notions of function-free first-order (FO) logic, and in particular, we consider Description Logics (DLs), which are fragments of FO logic that can be used to represent the domain of interest through *individuals* (also known as constants), representing real-world objects, *atomic concepts*, i.e., unary predicates representing sets of objects, and *atomic roles*, which denote binary relations between objects [1]. We assume to have the pairwise disjoint countably infinite sets $\Sigma_\mathcal{C}$, $\Sigma_\mathcal{R}$, $\Sigma_\mathcal{I}$, and $\Sigma_\mathcal{V}$ for atomic concepts, atomic roles, individuals, and *variables*, respectively.

Complex expressions of concepts and roles are specified by means of suitable operators applied to atomic concepts and atomic roles. Different DLs allow for different operators in the construction of complex expressions.

For a DL language $\mathcal{L}_\mathcal{T}$, an $\mathcal{L}_\mathcal{T}$ TBox $\mathcal{T}$ is a finite set of assertions allowed in the language $\mathcal{L}_\mathcal{T}$, adopting symbols from $\Sigma_\mathcal{C} \cup \Sigma_\mathcal{R}$ as predicates and symbols from $\Sigma_\mathcal{I} \cup \Sigma_\mathcal{V}$ as terms. The set of atomic concepts, roles, and individuals mentioned in the assertions of $\mathcal{T}$ constitutes the *signature* of $\mathcal{T}$. Given a TBox $\mathcal{T}$, an ABox $\mathcal{A}$ for $\mathcal{T}$ is a finite set of *ground atoms* (which we also refer to as *facts*) of the form $A(a)$ and $P(a, b)$, where A and P are an atomic concept and an atomic role, respectively, occurring in the signature of $\mathcal{T}$, whereas a and b belong to $\Sigma_\mathcal{I}$. In what follows, when a TBox $\mathcal{T}$ is given, whenever we refer to an ABox $\mathcal{A}$, we implicitly assume that $\mathcal{A}$ is for $\mathcal{T}$.

For a DL language $\mathcal{L}_\mathcal{T}$, an $\mathcal{L}_\mathcal{T}$ ontology is a finite theory $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ constituted by an $\mathcal{L}_\mathcal{T}$ TBox $\mathcal{T}$ and by an ABox $\mathcal{A}$. The TBox $\mathcal{T}$ specifies the intensional knowledge of a modeled domain, whereas the ABox $\mathcal{A}$ specifies the extensional knowledge.

The semantics of an ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ is given in terms of *FO interpretations* [1] $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$, where $\Delta^\mathcal{I}$ is the interpretation domain (i.e. a non-empty set of objects), and $\cdot^\mathcal{I}$ is the interpretation function, which assigns to each constant c occurring either in $\mathcal{T}$ or in $\mathcal{A}$ a domain object $c^\mathcal{I} \in \Delta^\mathcal{I}$, to each unary predicate A in the signature of $\mathcal{T}$ a subset $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, and to each binary predicate P in the signature of $\mathcal{T}$ a subset $P^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. An interpretation $\mathcal{I}$ is a *model* of an ontology $\mathcal{O}$ if $\mathcal{I}$ *satisfies* all the assertions occurring in $\mathcal{O}$. An ontology $\mathcal{O}$ is said to be *consistent* if it has at least one model, *inconsistent* otherwise. Given an ontology $\mathcal{O}$ and an FO sentence ϕ, we say that $\mathcal{O}$ *entails* ϕ, denoted by $\mathcal{O} \models \phi$, if ϕ is true with respect to every model of $\mathcal{O}$. If this is not the case, then we say that $\mathcal{O}$ does not entail ϕ, denoted by $\mathcal{O} \not\models \phi$.

In this paper, we are particularly interested in DL-Lite$_\mathcal{R}$ ontologies, where DL-Lite$_\mathcal{R}$ is the member of the DL-Lite family [15] underpinning OWL 2 QL [30], i.e., the OWL 2 profile specifically designed for efficient query answering.

Concepts and roles expressions in DL-Lite$_\mathcal{R}$ are formed according to the following syntax:

$$B \longrightarrow A \mid \exists R$$
$$R \longrightarrow P \mid P^-,$$

where B is a *basic concept*, i.e., an expression of the form A, with $A \in \Sigma_\mathcal{C}$, $\exists P$, with $P \in \Sigma_\mathcal{R}$, or $\exists P^-$. The expressions $\exists P$ and $\exists P^-$ are called unqualified existential restrictions, which denote the set of objects occurring as first or second argument of P, respectively. R denotes a *basic role*, i.e., an expression of the form P or P^- (called the inverse of the atomic role P).

A DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ consists of a finite set of assertions of the following form[1]:

$$B_1 \sqsubseteq B_2 \quad B_1 \sqsubseteq \neg B_2$$
$$R_1 \sqsubseteq R_2 \quad R_1 \sqsubseteq \neg R_2.$$

Assertions of the left-hand side are called *positive inclusion assertions* (or, simply, positive inclusions) specified, from the top to the bottom, between concepts and roles, respectively. Assertions of the right-hand side are called *negative inclusion assertions* (or, simply, negative inclusions), also

[1] For DL-Lite$_\mathcal{R}$ assertions we adopt the well-known variable-free DL syntax [1].

called disjointnesses, specified, from the top to the bottom, between concepts and roles respectively.

Given an interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ of the form described above, the interpretation function $\cdot^{\mathcal{I}}$ extends to DL-Lite$_{\mathcal{R}}$ basic concepts and roles as follows.

$$(\exists P)^{\mathcal{I}} = \{o \mid \exists o'.\ (o, o') \in P^{\mathcal{I}}\}$$
$$(P^{-})^{\mathcal{I}} = \{(o, o') \mid (o', o) \in P^{\mathcal{I}}\}$$

To complete the definition of the semantics of a DL-Lite$_{\mathcal{R}}$ ontology, we define when an interpretation $\mathcal{I}$ satisfies TBox and ABox assertions:

- $\mathcal{I}$ satisfies a TBox positive concept inclusion assertion $B_1 \sqsubseteq B_2$ if $B_1^{\mathcal{I}} \subseteq B_2^{\mathcal{I}}$;
- $\mathcal{I}$ satisfies a TBox positive role inclusion assertion $R_1 \sqsubseteq R_2$ if $R_1^{\mathcal{I}} \subseteq R_2^{\mathcal{I}}$;
- $\mathcal{I}$ satisfies a TBox negative concept inclusion assertion $B_1 \sqsubseteq \neg B_2$ if $B_1^{\mathcal{I}} \cap B_2^{\mathcal{I}} = \emptyset$;
- $\mathcal{I}$ satisfies a TBox negative role inclusion assertion $R_1 \sqsubseteq \neg R_2$ if $R_1^{\mathcal{I}} \cap R_2^{\mathcal{I}} = \emptyset$;
- $\mathcal{I}$ satisfies an ABox assertion $A(a)$ if $a^{\mathcal{I}} \in A^{\mathcal{I}}$;
- $\mathcal{I}$ satisfies an ABox assertion $P(a, b)$ if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$.

As usual in query answering over DL ontologies, we focus on the language of conjunctive queries. A *Boolean conjunctive query (BCQ)* q is an FO sentence of the form $\exists \vec{x}.\ \phi(\vec{x})$, where $\vec{x}$ are variables in $\Sigma_{\mathcal{V}}$, and $\phi(\vec{x})$ is a finite, non-empty conjunction of atoms of the form $\alpha(\mathbf{t})$, where $\alpha \in \Sigma_{\mathcal{C}} \cup \Sigma_{\mathcal{R}}$, and each term in $\mathbf{t}$ is either a constant in $\Sigma_{\mathcal{I}}$ or a variable in $\vec{x}$. A union Q of BCQs is a Boolean FO sentence of the form $\bigvee_{i=1}^{n} q_i$, where q_i is a BCQ for each $i = 1, \ldots, n$.

We recall that, for every DL-Lite$_{\mathcal{R}}$ TBox $\mathcal{T}$ and union Q of BCQs, it is possible to effectively compute an FO query q_r, called the *perfect reformulation of Q with respect to $\mathcal{T}$*, such that, for each ABox $\mathcal{A}$, we have $\mathcal{T} \cup \mathcal{A} \models Q$ if and only if q_r evaluates to true over the ABox $\mathcal{A}$ seen as an interpretation [15]. This yields the well-known result that answering unions of BCQs over DL-Lite$_{\mathcal{R}}$ ontologies is *FO-rewritable*, and therefore the underlying decision problem is in AC^0 in the size of the ABox, i.e., in the so-called *data complexity* [36]. All the complexity results in this paper concern with data complexity.

Given a TBox $\mathcal{T}$, a *denial assertion* (or simply *denial*) over $\mathcal{T}$ is an FO sentence of the form $\forall \vec{x}.\phi(\vec{x}) \rightarrow \bot$ over the signature of $\mathcal{T}$, such that $\exists \vec{x}.\phi(\vec{x})$ is a BCQ. Given a DL-Lite$_{\mathcal{R}}$ ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ and a set of denials assertions $\mathcal{P}$ over $\mathcal{T}$, we note that the theory $\mathcal{O} \cup \mathcal{P}$ is consistent if and only if $\mathcal{O} \not\models \exists \vec{x}.\phi(\vec{x})$ holds for each denial assertions $\forall \vec{x}.\phi(\vec{x}) \rightarrow \bot$ occurring in $\mathcal{P}$ [27].

In the following, with **CQ** and **GA** we denote the languages of BCQs and ground atoms, respectively, all specified over the alphabets $\Sigma_{\mathcal{C}}, \Sigma_{\mathcal{R}}, \Sigma_{\mathcal{I}}$, and $\Sigma_{\mathcal{V}}$. Note that **GA** $\subseteq$ **CQ**.

Given an ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ and a language $\mathcal{L} \subseteq$ **CQ**, with $\mathcal{L}(\mathcal{O})$ we refer to the subset of $\mathcal{L}$ containing all those sentences constructible using the atomic concepts and atomic roles in the signature of $\mathcal{T}$ as predicates, and the constants occurring in $\mathcal{T} \cup \mathcal{A}$ and the variables in $\Sigma_{\mathcal{V}}$ as terms. Given a language $\mathcal{L} \subseteq$ **CQ**, a TBox $\mathcal{T}$, and an ABox $\mathcal{A}$, we denote by $\mathsf{cl}_{\mathcal{L}}^{\mathcal{T}}(\mathcal{A})$ the set of those sentences in $\mathcal{L}(\mathcal{T} \cup \mathcal{A})$ that are entailed by $\mathcal{T} \cup \mathcal{A}$, i.e., $\mathsf{cl}_{\mathcal{L}}^{\mathcal{T}}(\mathcal{A}) = \{\phi \mid \phi \in \mathcal{L}(\mathcal{T} \cup \mathcal{A})$ and $\mathcal{T} \cup \mathcal{A} \models \phi\}$.

CQ Censors

A *CQE specification* is a pair $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, where $\mathcal{T}$ is a DL TBox and $\mathcal{P}$ is a policy over $\mathcal{T}$, i.e., a finite set of denial assertions over $\mathcal{T}$. A *CQE instance* is a pair $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$, where $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ is a CQE specification and $\mathcal{A}$ is an ABox for $\mathcal{T}$. We sometime say that $\mathcal{A}$ is an ABox for $\mathcal{J}$. In the following, when a TBox $\mathcal{T}$ is given, we always assume that the coupled policy is specified over $\mathcal{T}$, that each considered ABox $\mathcal{A}$ is over the signature of $\mathcal{T}$, and that, unless otherwise specified, $\mathcal{T} \cup \mathcal{A}$ and $\mathcal{T} \cup \mathcal{P}$ are consistent.

Example 1 Consider the CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, where:

$$\mathcal{T} = \{\ \mathsf{Oncol} \sqsubseteq \exists \mathsf{WorksIn},\ \exists \mathsf{Cures} \sqsubseteq \exists \mathsf{WorksIn},\ \exists \mathsf{Cures}^{-} \sqsubseteq \mathsf{Patient}\ \}$$
$$\mathcal{P} = \{\ \mathsf{Oncol}(x), \mathsf{Cures}(x, y) \rightarrow \bot\ \}$$

In words, the DL-Lite$_{\mathcal{R}}$ TBox $\mathcal{T}$ sanctions that (i) every oncologist (Oncol) works in (WorksIn) some place, and (ii) if somebody cures (Cures) someone else, the first individual must work somewhere and the latter is a patient (Patient). The data protection policy specified by $\mathcal{P}$ hides the existence of patients cured by oncologists. □

Definition 1 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. A *conjunctive query censor* (*CQ censor*) $\mathsf{cens}(\cdot)$ for $\mathcal{J}$ is a function that, for each ABox $\mathcal{A}$, returns a possibly infinite set $\mathsf{cens}(\mathcal{A})$ such that (i) $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cl}_{\mathbf{CQ}}^{\mathcal{T}}(\mathcal{A})$ and (ii) $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cens}(\mathcal{A})$ is consistent.

Example 2 Consider the CQE specification $\mathcal{J}$ of Example 1. The following functions are CQ censors for $\mathcal{J}$.

$\mathsf{cens}_1(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}_{\mathbf{CQ}}^{\mathcal{T}}(\mathcal{A})$ all the CQs containing: (i) at least one conjunction $\mathsf{Oncol}(x) \wedge \mathsf{Cures}(x, y)$ or (ii) at least one atom $\mathsf{Oncol}(a)$, with $a \in \Sigma_{\mathcal{I}}$, s.t. $\mathcal{T} \cup \mathcal{A} \models \exists x.\mathsf{Cures}(a, x)$.

$\mathsf{cens}_2(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ all the CQs containing: (*i*) at least one conjunction $\mathsf{Oncol}(x) \wedge \mathsf{Cures}(x, y)$ or (*ii*) at least one atom $\mathsf{Cures}(a, x)$, with $(a, x) \in \Sigma_{\mathcal{I}} \times \Sigma_{\mathcal{V}}$, s.t. $\mathcal{T} \cup \mathcal{A} \models \mathsf{Oncol}(a)$.

$\mathsf{cens}_3(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ all the CQs containing at least one atom $\mathsf{Cures}(t_1, t_2)$, with $t_1, t_2 \in \Sigma_{\mathcal{I}} \cup \Sigma_{\mathcal{V}}$. □

Given two CQ censors $\mathsf{cens}_1(\cdot)$ and $\mathsf{cens}_2(\cdot)$ for a CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, we say that $\mathsf{cens}_2(\cdot)$ is *more informative than* $\mathsf{cens}_1(\cdot)$ if (*i*) $\mathsf{cens}_1(\mathcal{A}) \subseteq \mathsf{cens}_2(\mathcal{A})$ holds for each ABox $\mathcal{A}$ and (*ii*) $\mathsf{cens}_1(\mathcal{A}) \subset \mathsf{cens}_2(\mathcal{A})$ holds for at least an ABox $\mathcal{A}$.

Definition 2 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification and $\mathsf{cens}(\cdot)$ be a CQ censor for $\mathcal{J}$. We say that $\mathsf{cens}(\cdot)$ is an *optimal CQ censor* for $\mathcal{J}$ if there does not exist a CQ censor $\mathsf{cens}'(\cdot)$ for $\mathcal{J}$ that is more informative than $\mathsf{cens}(\cdot)$.

Example 3 Consider the CQE specification $\mathcal{J}$ of Example 1 and recall the CQ censors of Example 2. We have that $\mathsf{cens}_1(\cdot)$ and $\mathsf{cens}_2(\cdot)$ are optimal CQ censor for $\mathcal{J}$ while $\mathsf{cens}_3(\cdot)$ is not, since $\mathsf{cens}_2(\cdot)$ is more informative than $\mathsf{cens}_3(\cdot)$. Indeed, consider the ABox $\mathcal{A} = \{\mathsf{Oncol}(ann), \mathsf{Cures}(ann, bob), \mathsf{Cures}(sam, tom)\}$. One can verify that $\mathsf{cens}_3(\mathcal{A}) \subset \mathsf{cens}_2(\mathcal{A})$. Moreover, from the definition of such functions, it is possible to see that $\mathsf{cens}_3(\mathcal{A}') \subseteq \mathsf{cens}_2(\mathcal{A}')$ for any ABox $\mathcal{A}'$. Intuitively, $\mathsf{cens}_3(\cdot)$ hides also all the BCQs containing facts over the role Cures which could be safely disclosed, like $\mathsf{Cures}(sam, tom)$. ▽

As shown by Example 3, there may exist more than one optimal CQ censor for a given CQE specification $\mathcal{J}$. We denote by $\mathsf{optCQCens}(\mathcal{J})$ the set of optimal CQ censors for $\mathcal{J}$.

An important property of optimal CQ censors is that they do not hide information implied by the pair ABox and TBox except in cases where the hiding is functional to not disclose policy-protected data. More formally, optimal CQ censors enjoy what we call *knowledge preservation property*, i.e. if $\mathsf{cens}(\cdot)$ is an optimal CQ censor for a CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, then $\mathsf{cens}(\mathcal{A}) = \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ holds for each ABox $\mathcal{A}$ such that $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$ is consistent. It is easy to see that, in general, the same property is not satisfied by non-optimal CQ censors.

While in other works such as [22, 23] the study was focused on verifying the existence of an optimal censor and its computation, we follow a different approach, and define CQE as a form of skeptical reasoning over all the optimal censors of the underlying CQE specification. To this aim, we provide the definition of entailment under CQ censors.

Definition 3 Let $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ be a CQE instance, where $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, and ϕ be an FO sentence. We say that ϕ is *CQ-Cens entailed by* $\mathcal{E}$, denoted by $\mathcal{E} \models^{\mathsf{cqe}}_{\mathrm{CQ}} \phi$, if $\mathcal{T} \cup \mathsf{cens}(\mathcal{A}) \models \phi$ holds for every $\mathsf{cens}(\cdot) \in \mathsf{optCQCens}(\mathcal{J})$.

Example 4 Consider the CQE instance $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$, where $\mathcal{J}$ and $\mathcal{A}$ are as in Example 3. The following three queries are given:

$q_1 = \exists x.\mathsf{WorksIn}(ann, x)$
$q_2 = \exists x.\mathsf{Cures}(x, bob)$
$q_3 = \mathsf{Cures}(ann, bob)$
$q_4 = \mathsf{Cures}(sam, tom)$.

One can verify that queries q_1, q_2, and q_4 are CQ-Cens entailed by $\mathcal{E}$, while q_3 is not. □

We now provide the data complexity of the problem of entailment of BCQs in CQE with respect to CQ censors, focusing on DL-Lite$_{\mathcal{R}}$ as ontology language.

Theorem 1 [[28]] *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification such that $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox, let $\mathcal{A}$ be an ABox, and let q be a BCQ. The problem of deciding whether $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\mathsf{cqe}}_{\mathrm{CQ}} q$ is in PTIME in data complexity.*

Inspired by the work on consistent query answering [26, 27], and towards the identification of a notion of censor that allow us to have a single optimal censor, we introduce below the definition of intersection-based CQ censor.

Definition 4 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. An *intersection-based CQ censor* (*ICQ censor*) $\mathsf{cens}_{\cap}(\cdot)$ for $\mathcal{J}$ is a function that, given an ABox $\mathcal{A}$, returns a possibly infinite set $\mathsf{cens}_{\cap}(\mathcal{A}) = \bigcap_{\mathsf{cens}(\cdot) \in \mathsf{optCQCens}(\mathcal{J})} \mathsf{cens}(\mathcal{A})$.

With the notion of ICQ censor in place, we can provide the following notion of entailment.

Definition 5 Let $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ be a CQE instance and ϕ be an FO sentence. We say that ϕ is *ICQ-Cens entailed by* $\mathcal{E}$, denoted by $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\mathsf{cqe}}_{\mathrm{ICQ}} \phi$, if $\mathcal{T} \cup \mathsf{cens}_{\cap}(\mathcal{A}) \models \phi$ holds, where $\mathsf{cens}_{\cap}(\cdot)$ the ICQ censor for $\mathcal{J}$.

We now examine the data complexity of the problem of entailment of BCQs in CQE under ICQ censors, again by considering DL-Lite$_{\mathcal{R}}$ as ontology language. First, we provide the following straightforward lemma, which is crucial to study the computational complexity of this problem.

Lemma 1 *Let $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ be a CQE instance and q be a BCQ. We have that $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{CQ}} q$ if and only if $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{ICQ}} q$.*

Proof The fact that $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{ICQ}} q$ implies $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{CQ}} q$ is trivial.

Suppose now that $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{CQ}} q$. By definition, we have that $\mathcal{T} \cup \mathsf{cens}(\mathcal{A}) \models q$ holds for every $\mathsf{cens}(\cdot) \in \mathsf{optCQCens}(\mathcal{J})$. It immediately follows that $q \in \mathsf{cens}(\mathcal{A})$ (otherwise $\mathsf{cens}(\cdot)$ would not be optimal) for every $\mathsf{cens}(\cdot) \in \mathsf{optCQCens}(\mathcal{J})$, and thus $q \in \mathsf{cens}_{\cap}(\mathcal{A})$, where $\mathsf{cens}_{\cap}$ is the ICQ censor for $\mathcal{J}$. We therefore conclude that $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{ICQ}} q$. □

From the above result and Theorem 1, we immediately have the following result.

Theorem 2 *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification such that $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox, $\mathcal{A}$ be an ABox, and q be a BCQ. The problem of deciding whether $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{ICQ}} q$ is in PTIME in data complexity.*

It is interesting to note that, for CQE instances with a DL-Lite$_{\mathcal{R}}$ TBox, both CQ censors and the ICQ censor return sets for which a representation as a finite set of BCQs does not always exist. This behavior implies that in general it is not possible to materialize a set of BCQs to be used for deciding query entailment under these kinds of censors.

Theorem 3 *There exists a CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, where $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox and* $\mathsf{optCQCens}(\mathcal{J}) = \{\mathsf{cens}(\cdot)\}$*, and an ABox $\mathcal{A}$ such that there does not exist any finite subset $\mathcal{S}$ of $\mathsf{cens}(\mathcal{A})$ satisfying the following: $\mathcal{T} \cup \mathcal{S} \models q$ if and only if $\mathcal{T} \cup \mathsf{cens}(\mathcal{A}) \models q$, for each BCQ q.*

Proof Consider the following DL-Lite$_{\mathcal{R}}$ CQE specification $\mathcal{J} = \langle \mathcal{T}, P \rangle$, where:

$$\mathcal{T} = \{R \sqsubseteq P, \exists R^- \sqsubseteq \exists R\}$$
$$\mathcal{P} = \{\forall x, y.R(x, y) \rightarrow \bot\}.$$

Is it easy to verify that $\mathsf{optCQCens}(\mathcal{J})$ contains only the following optimal CQ censor:

$\mathsf{cens}(\cdot)$: given an ABox $\mathcal{A}$, $\mathsf{cens}_1(\mathcal{A})$ returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ all the CQs containing at least one atom over the predicate R.

Now, consider the following ABox $\mathcal{A} = \{R(a, b)\}$. One can verify that $\mathsf{cens}(\mathcal{A})$ is semantically equivalent to the following infinite set of queries:

$\{ P(a, b),$
$\exists x, y.P(a, x) \wedge P(x, y),$
$\exists x, y.P(x, b) \wedge P(b, y),$
$\exists x, y, z.P(a, x) \wedge P(x, y) \wedge P(y, z),$
$\exists x, y, z.P(x, b) \wedge P(b, y) \wedge P(y, z),$
$\exists x, y, z, w.P(a, x) \wedge P(x, y) \wedge P(y, z) \wedge P(z, w),$
$\exists x, y, z, w.P(x, b) \wedge P(b, y) \wedge P(y, z) \wedge P(z, w),$
$\ldots \}.$

It is easy to see that, for each natural number $k \geq 1$, if we let $\mathcal{S}_k$ be the set of all queries of length at most k (i.e., with at most k atoms in their body), then there exists a query $q' \in \mathsf{cens}(\mathcal{A})$ of length $k + 1$ such that $\mathcal{S}_k \not\models q'$. By construction, it easily follows that no finite subset $\mathcal{S}$ of $\mathsf{cens}(\mathcal{A})$ exits such that $\mathcal{T} \cup \mathcal{S} \models q$ if and only if $\mathcal{T} \cup \mathsf{cens}(\mathcal{A}) \models q$, for each BCQ q. □

We further observe that, in the above proof, $\mathsf{cens}(\cdot)$ is the only optimal CQ censor for $\mathcal{J}$. Then it coincides with the ICQ censor for $\mathcal{J}$. Thus, the result holds also for the ICQ censor case.

Another important property of censors that is often addressed in the CQE literature is the so-called indistinguishability. The idea behind this property is that to preserve the confidentiality of a CQE instance $\langle \mathcal{J}, \mathcal{A} \rangle$, there must exist an ABox $\mathcal{A}'$ that contains no sensitive information and is indistinguishable from $\mathcal{A}$ in the eyes of the user. The formal definition is given below.

Definition 6 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. A CQ censor $\mathsf{cens}(\cdot)$ for $\mathcal{J}$ satisfies the *indistinguishability* property if, for each ABox $\mathcal{A}$, there exists an ABox $\mathcal{A}'$ such that (*i*) $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is consistent and (*ii*) $\mathsf{cens}(\mathcal{A}) = \mathsf{cens}(\mathcal{A}')$.

Note that, according to the Definition 1, an ICQ censor is in fact a CQ censor (even though not necessarily optimal), and thus the above definition can also be applied to ICQ censors.

Unfortunately, as already observed in [20], in general, CQ censors do not enjoy the indistinguishability property. This is illustrated by the following example.

Example 5 Consider the CQE specification $\mathcal{J} = \langle \mathcal{T}, P \rangle$, where:

$$\mathcal{T} = \emptyset$$
$$\mathcal{P} = \{\forall x.R(x, b) \rightarrow \bot\}.$$

Note that we have only one censor in $\mathsf{optCQCens}(\mathcal{J})$, which is:

$\mathsf{cens}(\cdot)$: given an ABox $\mathcal{A}$, $\mathsf{cens}(\mathcal{A})$ returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ all the CQs

containing at least one atom over the predicate R having in the second position the constant b.

Consider the ABox $\mathcal{A} = \{R(a,b)\}$. We have that $\mathsf{cens}(\mathcal{A})$ is semantically equivalent to $\{\exists x.R(a,x)\}$. According to Definition 6, there must exists an ABox $\mathcal{A}'$ such that (i) $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is consistent and (ii) $\mathsf{cens}(\mathcal{A}) = \mathsf{cens}(\mathcal{A}')$. Since the TBox is empty, the only way to entail the query $\exists x.R(a,x)$ is to have in $\mathcal{A}'$ an atom of the form $R(a,\alpha)$, where α is a constant. If $\alpha = b$, then $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is not consistent, while for every $\alpha \neq b$ we have that $R(a,\alpha) \in \mathsf{cens}(\mathcal{A}')$ but $R(a,\alpha) \notin \mathsf{cens}(\mathcal{A}')$. ▽

Moreover, since in the above example $\mathsf{optCQCens}(\mathcal{J}) = \{\mathsf{cens}(\cdot)\}$, we have that $\mathsf{cens}(\cdot)$ is the ICQ censor for $\mathcal{J}$. Thus, the above example also shows that ICQ censors are not guaranteed to enjoy the indistinguishability property.

GA Censors

As shown in the previous section, CQ censors do not enjoy some notable properties, such as having a finite representation or that of indistinguishability. In this section, we provide a new notion of censor which instead enjoys both of these properties.

Definition 7 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. A *ground atom censor* (*GA censor*) $\mathsf{cens}(\cdot)$ for $\mathcal{J}$ is a function that, for each ABox $\mathcal{A}$, returns a set $\mathsf{cens}(\mathcal{A})$ such that (i) $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ and (ii) $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cens}(\mathcal{A})$ is consistent.

Example 6 Consider the CQE specification $\mathcal{J}$ of Example 1. The following functions are GA censors for $\mathcal{J}$.

$\mathsf{cens}_4(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ all the facts of the form $\mathsf{Cures}(a,b)$.

$\mathsf{cens}_5(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ all the facts of the form $\mathsf{Cures}(a,b)$ such that $\mathcal{T} \cup \mathcal{A} \models \mathsf{Oncol}(a)$.

$\mathsf{cens}_6(\cdot)$: the function that, given an ABox $\mathcal{A}$, returns the set computed by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ all the facts of the form $\mathsf{Oncol}(a)$ such that $\mathcal{T} \cup \mathcal{A} \models \exists y.\mathsf{Cures}(a,y)$. □

As done for CQ censors, we define when a GA censor is more informative than another and when it is optimal.

Given two GA censors $\mathsf{cens}_1(\cdot)$ and $\mathsf{cens}_2(\cdot)$ for a CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, we say that $\mathsf{cens}_2(\cdot)$ is *more informative than* $\mathsf{cens}_1(\cdot)$ if (i) $\mathsf{cens}_1(\mathcal{A}) \subseteq \mathsf{cens}_2(\mathcal{A})$ holds for each ABox $\mathcal{A}$ and (ii) $\mathsf{cens}_1(\mathcal{A}) \subset \mathsf{cens}_2(\mathcal{A})$ holds for at least an ABox $\mathcal{A}$. Furthermore, given a CQE specification $\mathcal{J}$ and a GA censor $\mathsf{cens}(\cdot)$ for $\mathcal{J}$, we say that $\mathsf{cens}(\cdot)$ is an *optimal GA censor* for $\mathcal{J}$ if there does not exist a GA censor $\mathsf{cens}'(\cdot)$ for $\mathcal{J}$ that is more informative than $\mathsf{cens}(\cdot)$. As for CQ censors, given a CQE specification $\mathcal{J}$, there may exist more than one optimal GA censor for $\mathcal{J}$. We denote by $\mathsf{optGACens}(\mathcal{J})$ the set of all optimal GA censors for $\mathcal{J}$.

Example 7 Consider the CQE instance $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ of Example 4 and the GA censors of Example 6. While $\mathsf{cens}_5(\cdot)$ and $\mathsf{cens}_6(\cdot)$ are optimal GA censors for $\mathcal{J}$, it is easy to see that $\mathsf{cens}_4(\cdot)$ is not. Indeed, by applying such censors to $\mathcal{A}$ one would get the following sets of ground atoms:

$$\mathsf{cens}_4(\mathcal{A}) = \{\mathsf{Oncol}(ann), \mathsf{Patient}(bob), \mathsf{Patient}(tom)\}$$
$$\mathsf{cens}_5(\mathcal{A}) = \{\mathsf{Oncol}(ann), \mathsf{Patient}(bob), \mathsf{Cures}(sam, tom), \mathsf{Patient}(tom)\}$$
$$\mathsf{cens}_6(\mathcal{A}) = \{\mathsf{Cures}(ann, bob), \mathsf{Patient}(bob), \mathsf{Cures}(sam, tom), \mathsf{Patient}(tom)\}$$

Clearly, we have that $\mathsf{cens}_4(\mathcal{A}) \subset \mathsf{cens}_5(\mathcal{A})$. Moreover, from the definition of such functions, one can verify that $\mathsf{cens}_4(\mathcal{A}') \subseteq \mathsf{cens}_5(\mathcal{A}')$ for any ABox $\mathcal{A}'$. □

As in the case of optimal CQ censors, also optimal GA censors enjoy the knowledge preservation property, i.e.: if $\mathsf{cens}(\cdot)$ is an optimal GA censor for a CQE specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$, then $\mathsf{cens}(\mathcal{A}) = \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ holds for each ABox $\mathcal{A}$ such that $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$ is consistent. It is easy to see that, in general, the same property is not satisfied by non-optimal GA censors.

We also observe that since $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ is a finite set of ground atoms, it follows from Definition 7 that so is the set returned by a GA censor. As such, it can be represented as an ABox. This means that, unlike CQ censors, there is always a finite representation of the application of a GA censor to an ABox. As shown in [20], such materialization can be computed in polynomial time with respect to the size of the ABox in input.

We now turn our attention to the indistinguishability property and we show that GA censors enjoy it. The definition of the indistinguishability property must first be slightly modified to account for GA censors.

Definition 8 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. A GA censor $\mathsf{cens}(\cdot)$ for $\mathcal{J}$ satisfies the *indistinguishability* property if, for each ABox $\mathcal{A}$, there exists an ABox $\mathcal{A}'$ such that (i) $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ and (ii) $\mathsf{cens}(\mathcal{A}) = \mathsf{cens}(\mathcal{A}')$.

We are now ready to provide the following result.

Proposition 1 [[18]] *Every optimal GA censor for a CQE specification $\mathcal{J}$ satisfies the indistinguishability property.*

Below we define the entailment in CQE under GA censors.

Definition 9 Let $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ be a CQE instance and ϕ be an FO sentence. We say that ϕ is *GA-Cens entailed by* $\mathcal{E}$, denoted by $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{GA}} \phi$, if $\mathcal{T} \cup \mathsf{cens}(\mathcal{A}) \models \phi$ holds for every $\mathsf{cens}(\cdot) \in \mathsf{optGACens}(\mathcal{J})$.

Example 8 Consider the same CQE instance $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ and the same queries q_1, q_2, q_3, and q_4 of Example 4. One can verify that both q_1 and q_4 are GA-Cens entailed by $\mathcal{E}$, while both q_2 and q_3 are not. Note that, as shown in Example 4, the query q_2 is instead CQ-Cens entailed by $\mathcal{E}$. □

Since **CQ** is a more expressive language than **GA**, one might wonder whether GA-Cens entailment is actually a sound approximation of CQ-Cens entailment. The following example shows that, in fact, the two entailment semantics are incomparable with each other.

Example 9 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification, with:

$$\mathcal{T} = \{\ A \sqsubseteq \exists P\ \}$$
$$\mathcal{P} = \{\ \forall x, y.B(x) \wedge P(x, y) \rightarrow \bot,\ \forall x.A(x) \rightarrow \bot\}$$

Now, consider the ABox $\mathcal{A} = \{A(c), B(c)\}$. All optimal GA censors for $\mathcal{J}$, when applied to $\mathcal{A}$, will return the set $\mathcal{C}_1 = \{B(c)\}$. On the other hand, the optimal CQ censors will return one of the two following sets of BCQs:

$$\mathcal{C}_2 = \{\ B(c), \exists x, y.P(x, y),\ \text{ and all the queries in } \mathbf{CQ}(\mathcal{T} \cup \mathcal{P}) \text{ inferred by them }\}$$
$$\mathcal{C}_3 = \{\ \exists x.B(x), \exists y.P(c, y),\ \text{ and all the queries in } \mathbf{CQ}(\mathcal{T} \cup \mathcal{P}) \text{ inferred by them }\}$$

Now, given the two BCQs $q_1 = B(c)$ and $q_2 = \exists x, y.P(x, y)$, we have that: $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{GA}} q_1$ but $\langle \mathcal{J}, \mathcal{A} \rangle \not\models^{\text{cqe}}_{\text{CQ}} q_1$, while $\langle \mathcal{J}, \mathcal{A} \rangle \not\models^{\text{cqe}}_{\text{GA}} q_2$ but $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{CQ}} q_2$. □

This incomparability is also evidenced by the complexity results of the entailment problem of BCQs in CQE under GA censors showing that, if DL-Lite$_{\mathcal{R}}$ is considered as ontology language, the problem is intractable in the case of GA censors.

Theorem 4 [[28]] *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification such that $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox, let $\mathcal{A}$ be an ABox, and let q be a BCQ. The problem of deciding whether $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{GA}} q$ is coNP-complete in data complexity.*

On the other hand, the following result shows that instance checking is tractable and, actually, in AC ^{0} in data complexity.

Theorem 5 [[28]] *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification such that $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox, let $\mathcal{A}$ be an ABox, and let g be a ground atom. The problem of deciding whether $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{GA}} g$ is in AC ^{0} in data complexity.*

As done in the previous section for CQ censors, we now consider an intersection-based censor also for the case of GA censors.

Definition 10 Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification. An *intersection-based GA censor* (*IGA censor*) $\mathsf{cens}_{\cap}(\cdot)$ for $\mathcal{J}$ is a function that, given an ABox $\mathcal{A}$, returns the set $\mathsf{cens}_{\cap}(\mathcal{A}) = \bigcap_{\mathsf{cens}(\cdot) \in \mathsf{optGACens}(\mathcal{J})} \mathsf{cens}(\mathcal{A})$.

Also in this case, for each CQE specification, there exists by definition only one IGA censor. We note that, given a CQE specification $\mathcal{J}$, the IGA censor for $\mathcal{J}$ is, according to Definition 7, also a GA censor for $\mathcal{J}$. Moreover, as stated in [18], IGA censors enjoy the indistinguishability property.

We now examine the data complexity of the problem of entailment of BCQs in CQE under IGA censors, again by considering DL-Lite$_{\mathcal{R}}$ as ontology language. We first give the definition of entailment, which, as done in analogous definitions of entailment given before, is given with respect to a generic ontology language for the CQE specification.

Definition 11 Let $\mathcal{E} = \langle \mathcal{J}, \mathcal{A} \rangle$ be a CQE instance and ϕ be an FO sentence. We say that ϕ is *IGA-Cens entailed by* $\mathcal{E}$, denoted by $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{IGA}} \phi$, if $\mathsf{cens}_{\cap}(\mathcal{A}) \models \phi$ holds, where $\mathsf{cens}_{\cap}(\cdot)$ the IGA censor for $\mathcal{J}$.

It is straightforward to verify that IGA-Cens entailment is a sound approximation of the GA-Cens entailment.

Proposition 2 [[17]] *Let $\mathcal{J}$ be a CQE specification, $\mathcal{A}$ be an ABox, and ϕ be a FOL sentence. If $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{IGA}} \phi$, then $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\text{cqe}}_{\text{GA}} \phi$.*

However, differently from the case of CQ censors, entailment under GA censors does not coincide in general with entailment under the IGA censor. The following example shows, indeed, that the converse of Proposition 2 does not necessarily hold.

Example 10 Let $\mathcal{J}$ and $\mathcal{A}$ be as in Example 7. The IGA censor $\mathsf{cens}_{\cap}(\cdot)$ for $\mathcal{J}$ is such that $\mathsf{cens}_{\cap}(\mathcal{A}) = \{\ \mathsf{Patient}(bob), \mathsf{Cures}(sam, tom), \mathsf{Patient}(tom)\ \}$.

Consider now the queries of Example 8. While both q_1 and q_4 are GA-Cens entailed by $\langle \mathcal{J}, \mathcal{A} \rangle$, it is immediate to see that q_4 is the only query IGA-Entailed by $\langle \mathcal{J}, \mathcal{A} \rangle$. □

We conclude this section by providing the data complexity of IGA-Cens entailment of BCQs.

Theorem 6 [[17, 18]] *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{P} \rangle$ be a CQE specification such that $\mathcal{T}$ is a* DL-Lite$_{\mathcal{R}}$ *TBox, let $\mathcal{A}$ be an ABox, and let q be a BCQ. The problem of deciding whether $\langle \mathcal{J}, \mathcal{A} \rangle \models^{\mathsf{cqe}}_{\mathrm{IGA}} q$ is in AC^0 in data complexity.*

As shown by Theorem 6, with respect to GA-Cens entailment, the adoption of the IGA censor allows for a significant improvement in data complexity of query answering, still preserving the fact that we are adopting a kind of censor that enjoys the indistinguishability property.

Conclusions

In this paper, we provided an introduction to Controlled Query Evaluation in the context of Description Logics ontologies, by presenting a series of results from our previous investigations. We also introduced the new notion of the ICQ censor and studied its relationship with CQ censors. We have therefore considered four different CQE semantics, each of which is based on a different notion of censor, also highlighting some of their fundamental properties, namely, knowledge preservation, indistinguishability, and the possibility of materializing a censor. We also recalled the data complexity results of evaluating conjunctive queries on CQE instances by considering each of such semantics in the case where the TBox is expressed in DL-Lite$_{\mathcal{R}}$, i.e., the logical underpinning of the OWL 2 profile OWL 2 QL.

We remark that our approach is both fully declarative and semantically neat, which are its distinguishing features: the policy is expressed through logic formulas and confidentiality is enforced by automated reasoning on the CQE specification, according to the formal semantics adopted for censors. In other terms, to ensure data confidentiality it is only necessary to encode the privacy requirements into formulas, thus abstracting away from any underlying implementation mechanism.

Two main limitations of the techniques that we have described in this paper are the limited expressiveness of both the TBox and the policy language and the impossibility of expressing preferences for operating a further selection among optimal censors. To overcome the above limitations, in [18, 19] we extended both the ontology language and the policy language, showing that, for the case of the IGA censor, the computational complexity of BCQ entailment does not increase. More specifically, in the first work we considered DL-Lite$_{\mathcal{A}}$ as ontology language and denials with inequalities (under some safeness conditions) for expressing the policy and, in the latter, we adopted DL-Lite$^{\mathcal{H}}_{horn}$ for expressing the TBox and allowing denials with number restrictions in the policy. Moreover, in [18], we proposed an intersection-based notion of censor that takes into account preferences between ontology predicates for inducing a choice among the set of all the optimal censors and, consequently, obtaining a less severe censor (in terms of disclosable information). Finally, the CQE semantics presented in [10] is based on the so-called longest honeymoon approach, where, given a sequence of user queries, a (progressively refined) subset of optimal GA censors is selected for answering them, with the aim of delaying as much as possible the necessity of censoring the answer. This property is called maximal cooperativeness, and BCQ entailment under this semantics has been shown to still be in AC^0 in data complexity.

We believe that all these results open the way towards practical implementations of CQE engines for DL ontologies and Ontology-based Data Access (OBDA) [29, 37]. Some interesting results in this direction have already been presented in [17], where we have experimented IGA-Cens entailment over an OBDA benchmark. Such an experimental evaluation shows that controlled query evaluation can be realized in the practice using off-the-shelf OBDA engines. Analogous results have been obtained for the extended framework of [18].

The study of the CQE problem can still be extended in several directions. First, the PTIME upper bound for CQ-Cens entailment of BCQs over DL-Lite$_{\mathcal{R}}$ CQE instances should be refined. In addition, it may be of interest to cover the entire family of OWL 2 profiles, thus deepening the analysis of CQE for ontologies expressed in OWL 2 EL and OWL 2 RL, which was already started in [22, 23, 28].

Finally, the investigation of different forms of policy could be still pursued, to improve the abilities of the CQE framework in the enforcement of confidentiality.

The main contribution of this work is the collection in a single treatment of a series of our fundamental results on CQE in DL ontologies, and in particular in DL-Lite, previously appeared in various papers.

introduction to the problem, built through basic definitions, examples, and pointers to the original works, to which we refer the reader for further technical aspects.

advancements we achieved in the last years in this latter field. More precisely, we recall the general framework for CQE given in [28], focus on two kinds of censors proposed in that paper, namely CQ censors and GA censors, and report their most crucial characteristics, previously shown in

In this paper, we have studied the approach to CQE based on instance indistinguishability and identified a semantically well-founded notion of CQE that enjoys first-order rewritability in the case of DL-Lite$_{\mathcal{R}}$ ontologies. We believe that this result opens the way towards practical implementations of CQE engines for DL ontologies and Ontology-based Data Access. We are currently working to achieve this goal. Another important future direction is a deeper study of the user model. Our framework inherits from its predecessors a relatively simple model, which assumes that the user knows (at most) the TBox and all the query answers returned by the system, and considers only the deductive abilities of the user over such knowledge. This user model might need to be enriched to capture more realistic data protection scenarios.

Acknowledgements This work was partially supported by: projects FAIR (PE0000013) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union—NextGenerationEU; Glaciation project funded from the European Union's HE research and innovation programme (Grant agreement No. 101070141); ANTHEM project funded by the National Plan for NRRP Complementary Investments (CUP: B53C22006700001).

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors state that there is no conflict of interest.

References

1. Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider PF, editors. The description logic handbook: theory, implementation and applications. 2nd ed. Cambridge University Press; 2007.
2. Bayardo RJ, Agrawal R. Data privacy through optimal k-anonymization. In: Aberer K, Franklin MJ, Nishio S (eds) Proc. of the 31th IEEE Int. Conf. on Data Engineering (ICDE). IEEE Computer Society Press; 2005. pp. 217–228
3. Benedikt M, Cuenca Grau B, Kostylev EV. Logical foundations of information disclosure in ontology-based data integration. Artif Intell. 2018;262:52–95.
4. Bhanot R, Hans R. A review and comparative analysis of various encryption algorithms. Int J Secur Appl. 2015;9:289–306.
5. Biskup J. For unknown secrecies refusal is better than lying. Data Knowl Eng. 2000;33(1):1–23.
6. Biskup J, Bonatti PA. Lying versus refusal for known potential secrets. Data Knowl Eng. 2001;38(2):199–222.
7. Biskup J, Bonatti PA. Controlled query evaluation for enforcing confidentiality in complete information systems. Int J Inf Secur. 2004;3(1):14–27.
8. Biskup J, Bonatti PA. Controlled query evaluation for known policies by combining lying and refusal. Ann Math Artif Intell. 2004;40(1–2):37–62.
9. Biskup J, Weibert T. Keeping secrets in incomplete databases. Int J Inf Secur. 2008;7(3):199–217.
10. Bonatti P, Cima G, Lembo D, Marconi L, Rosati R, Sauro L, Savo DF. Controlled query evaluation in OWL 2 QL: A "longest honeymoon" approach. In: Proc. of the 21st Int. Semantic Web Conf. (ISWC), Lecture Notes in Computer Science. Springer, 2022.
11. Bonatti PA. A false sense of security. Artif Intell. 2022;310:103741.
12. Bonatti PA, de Capitani di Vimercati S, Samarati P. An algebra for composing access control policies. Int J Inf Secur. 2002;5(1):1–35.
13. Bonatti PA, Kraus S, Subrahmanian VS. Foundations of secure deductive databases. IEEE Trans Knowl Data Eng. 1995;7(3):406–422.
14. Bonatti PA, Sauro L. 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.
15. Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R. Tractable reasoning and efficient query answering in description logics: the DL-Lite family. J Autom Reason. 2007;39(3):385–429.
16. Calvanese D, De Giacomo G, Lenzerini M, Rosati R. View-based query answering in description logics: semantics and complexity. J Comput Syst Sci. 2012;78(1):26–46.
17. Cima G, Lembo D, Marconi L, Rosati R, Savo DF. Controlled query evaluation in ontology-based data access. In: Proc. of the 19th Int. Semantic Web Conf. (ISWC), volume 12506 of Lecture Notes in Computer Science. Springer; 2020. pp. 128–146.
18. Cima G, Lembo D, Marconi L, Rosati R, Savo DF. Controlled query evaluation over prioritized ontologies with expressive data protection policies. In: Proc. of the 20th Int. Semantic Web Conf. (ISWC), volume 12922 of Lecture Notes in Computer Science. Springer, 2021. pp. 374–391.
19. Cima G, Lembo D, Marconi L, Rosati R, Savo DF, Sinibaldi D. Controlled query evaluation over ontologies through policies with numerical restrictions. In: Proc. of the 4th IEEE Int. Conf. on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE; 2021. pp. 33–36.
20. Cima G, Lembo D, Rosati R, Savo DF. Controlled query evaluation in description logics through instance indistinguishability. In: Proc. of the 29th Int. Joint Conf. on Artificial Intelligence (IJCAI). 2020. pp. 1791–1797.
21. Cuenca Grau B, Horrocks I. Privacy-preserving query answering in logic-based information systems. In: Proc. of the 18th Eur. Conf. on Artificial Intelligence (ECAI). 2008. pp. 40–44.
22. Cuenca Grau B, Kharlamov E, Kostylev EV, Zheleznyakov D. Controlled query evaluation over OWL 2 RL ontologies. In: Proc. of the 12th Int. Semantic Web Conf. (ISWC), volume 8218 of Lecture Notes in Computer Science. Springer; 2013. pp. 49–65.
23. Cuenca Grau B, Kharlamov E, Kostylev EV, Zheleznyakov D. Controlled query evaluation for datalog and OWL 2 profile ontologies. In: Proc. of the 24th Int. Joint Conf. on Artificial Intelligence (IJCAI). 2015. pp. 2883–2889.
24. Cuenca Grau B, Kostylev EV. Logical foundations of linked data anonymisation. J Artif Intell Res. 2019;64:253–314.
25. Cuenca Grau B, Motik B. Reasoning over ontologies with hidden content: the import-by-query approach. J of Artif Intell Res. 2012;45:197–255.
26. Lembo D, Lenzerini M, Rosati R, Ruzzi M, Savo DF. Inconsistency-tolerant semantics for description logics. In: Proc. of the 4th

Int. Conf. on Web Reasoning and Rule Systems (RR). 2010. pp. 103–117.
27. Lembo D, Lenzerini M, Rosati R, Ruzzi M, Savo DF. Inconsistency-tolerant query answering in ontology-based data access. J Web Semant. 2015;33:3–29.
28. Lembo D, Rosati R, Savo DF. Revisiting controlled query evaluation in description logics. In: Proc. of the 28th Int. Joint Conf. on Artificial Intelligence (IJCAI). 2019. pp. 1786–1792.
29. Lenzerini M. Managing data through the lens of an ontology. AI Mag. 2018;39(2):65–74.
30. Motik B, Cuenca Grau B, Horrocks I, Wu Z, Fokoue A, Lutz C. OWL 2 Web Ontology Language profiles (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. Available at http://www.w3.org/TR/owl2-profiles/.
31. Motik B, Fokoue A, Horrocks I, Wu Z, Lutz C, Cuenca Grau B. OWL Web Ontology Language profiles. W3C Recommendation, World Wide Web Consortium, October 2009. Available at http://www.w3.org/TR/owl-profiles/.
32. W3C OWL Working Group. OWL 2 Web Ontology Language: Document Overview. W3C Recommendation, 27 October 2009. Available at http://www.w3.org/TR/owl2-overview/.
33. Shostack A. Threat modeling: designing for security. Wiley; 2014.
34. Sicherman GL, de Jonge W, van de Riet RP. Answering queries without revealing secrets. ACM Trans Database Syst. 1983;8(1):41–59.
35. Stouppa P, Studer T. Data privacy for *ALC* knowledge bases. In: Proc. of the 2nd Int. Symp. on Logical Foundations of Computer Science (LFCS), 2009. pp. 409–421.
36. Vardi MY. The complexity of relational query languages. In Proc. of the 14th ACM SIGACT Symp. on Theory of Computing (STOC). 1982. pp. 137–146.
37. Xiao G, Calvanese D, Kontchakov R, Lembo D, Poggi A, Rosati R, Zakharyaschev M. Ontology-based data access: A survey. In: Proc. of the 27th Int. Joint Conf. on Artificial Intelligence (IJCAI), 2018. pp. 5511–5519.
38. Zhang Z, Mendelzon A. Authorization views and conditional query containment. In: Proc. of the 10th Int. Conf. on Database Theory (ICDT), volume 3363 of Lecture Notes in Computer Science. Springer; 2005. pp. 259–273.