# Reverse Engineering of Data Services (Extended Abstract)

Gianluca Cima[1], Maurizio Lenzerini[1], Antonella Poggi[1,2]

Sapienza Università di Roma
*cima, lenzerini, poggi*@diag.uniroma1.it

**Abstract.** We study the problem of designing reverse engineering techniques for associating semantic descriptions to existing data services. We base our proposal on the Ontology-Based Data Access paradigm, where a domain ontology is used to provide a semantic layer mapped to the data sources of an organization. The basic idea is to perform the reverse engineering of a data service, expressed a query over the data sources, by deriving a query over the ontology that explains the semantics of the data service in terms of the element of the ontology. We illustrate a formal framework for this problem, based on the notion of source-to-ontology rewriting, which comes in three variants, called sound, complete and perfect, respectively. We present a thorough complexity analysis of two computational problems, namely verification (checking whether a query is a source-to-ontology rewriting of a given data service), and computation (computing a source-to-ontology rewriting of a data service).

## Introduction

The architecture of many modern Information Systems is based on data services [13], i.e., services deployed on top of data stores, other services, and/or applications to encapsulate a wide range of data-centric operations. In order to realize the promises of data services, in particular to foster their reuse, it is of vital importance to well document and clearly specify their semantics. While most current techniques manually associate APIs (Application Programming Interface) to data services, and describe their intended meaning with ad-hoc methods, often using natural language or complex metadata [5], we propose a new approach, whose goal is to automatically associate formal semantic descriptions to data services. We base our proposal on the *Ontology-Based Data Access* (OBDA) paradigm [11]. An OBDA specification consists of an ontology expressed in Description Logic (DL) [2], the schema of the data sources forming the information system, and a mapping between the source schema and the ontology. The ontology is a formal representation of the underlying domain, and the mapping specifies the relationship between the data at the sources and the concepts in the ontology. With the OBDA specification at hand, we pursue the idea of expressing the semantics of data services using the elements of the domain ontology, which is assumed to be familiar to the consumer of data services.

---

But how can we automatically produce a semantic characterization of a data service, having an OBDA specification available? The method we propose is to exploit a new reasoning task over the OBDA specification, that works as follows: we express the data service in terms of a query over the sources, and we aim at automatically deriving the query over the ontology that best describes the data service, given the mapping. The following example illustrates this idea.

*Example 1.* Let $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ be as follows:

- $\mathcal{O} = \{\ \mathsf{ErasmusStudent} \sqsubseteq \mathsf{Student},\ \mathsf{MathStudent} \sqsubseteq \mathsf{Student}\ \}$
- $\mathcal{S} = \{\ s_1, s_2, s_3, s_4\ \}$
- $\mathcal{M} = \begin{array}{l} \{(x) \mid s_1(x)\} \rightarrow \{(x) \mid \mathsf{Student}(x)\} \\ \{(x) \mid s_2(x)\} \rightarrow \{(x) \mid \mathsf{Student}(x)\} \\ \{(x) \mid s_1(x), s_3(x,y)\} \rightarrow \{(x) \mid \mathsf{ErasmusStudent}(x)\} \\ \{(x) \mid s_1(x), s_4(x,y)\} \rightarrow \{(x) \mid \mathsf{MathStudent}(x)\} \end{array}$

and consider the data service expressed as the source query $q_{\mathcal{S}}(x) = \{(x) \mid s_1(x) \vee s_3(x)\}$. It is easy to see that the query that best describes $q_{\mathcal{S}}$ in terms of $\mathcal{O}$ is $q_{\mathcal{O}}(x) = \{(x) \mid \mathsf{Student}(x)\}$.

Note that most of (if not all) the literature about managing data sources through an ontology [9,12] deals with user queries expressed over the ontology, and studies the problem of finding an *ontology-to-source rewriting*, i.e., a query over the source schema that, once executed over the data, provides the answers to the original query. Here, the problem is reversed, because we start with a source query and we aim at deriving a corresponding query over the ontology, called a *source-to-ontology rewriting*. Thus, we deal with a sort of reverse engineering problem, which is novel in the investigation of both OBDA and data integration.

The notions introduced in this paper are relevant in a plethora of scenarios. For the sake of brevity, we mention only two of them. Following the ideas in [6,7], it can be shown that our notions of source-to-ontology rewriting can be used to provide the semantics of open datasets and open APIs published by organizations, which is a crucial aspect for unchaining all the potentials of open data. In [10], the concept of realization of source queries, corresponding to one of the notions studied here, is used for checking whether the mapping provides the right coverage for expressing the relevant data services at the ontology level.

The contributions provided by this paper can be summarized as follows. We propose a formal framework for the problem of semantically characterizing a data service through an ontology (Section 3). We introduce the notions of *perfect*, *sound*, and *complete* source-to-ontology rewritings, and we define two basic reasoning tasks, namely *verification* and *computation*. The former checks whether a given query is a source-to-ontology rewriting of a data service, whereas the latter computes one such rewriting. We show that, although the ideal notion is the one of perfect source-to-ontology rewriting, there are cases where, with the given mapping, no query over the ontology can precisely characterize the data service at hand. Thus, we introduce *maximally sound* and *minimally complete* source-to-ontology rewritings, which intuitively aim at approximating the perfect rewriting of a data service at best, with the goal of either precision (sound rewriting), or recall (complete rewriting).

We study the verification and the computation problem for complete (Section 4) and sound (Section 5) source-to-ontology rewritings in one of the most popular OBDA setting considered in the literature, namely where the ontology language is *DL-Lite*$_{\mathcal{R}}$ [4,3], each mapping assertion maps a conjunctive query (CQ) over the source to a CQ over the ontology, and both the data service and the source-to-ontology rewriting are expressed as unions of CQs. For complete source-to-ontology rewritings we present algorithms for verification and computation, and characterize the complexity of both tasks. For the case of sound rewritings, we do the same for verification, and we precisely determine the cases where a maximally sound rewriting is not guaranteed to exist. For the lack of space, we do not tackle here the case of perfect rewritings, but we point out that results for complete and sound rewritings can be combined to study this case.

To the best of our knowledge, the problem studied in this paper has been (partially) addressed only in [6,10]. The former provides upper bound complexity results for complete rewritings, and the latter focuses on both *DL-Lite*$_{\mathcal{R}}$ and the $\mathcal{EL}$ family of ontology languages, and studies perfect rewritings only, under a slightly different semantics with respect to the one proposed here.

This paper is an extended abstract of [8]. Hence, while we assume basic knowledge about databases [1] and Description Logics (DL) [2], for specific concepts and notations, we refer to the Preliminaries section of [8].

## Framework

We implicitly refer to an OBDA specification $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$. Intuitively, given a data service expressed as a query $q_{\mathcal{S}}$ over $\mathcal{S}$, we aim at finding the query over $\mathcal{O}$ that precisely characterizes $q_{\mathcal{S}}$ w.r.t. $\Sigma$. Since the evaluation of queries over $\mathcal{O}$ is based on certain answers, this means that we aim at finding a query over $\mathcal{O}$ whose certain answers w.r.t. $\Sigma$ and $D$ exactly capture the answers of $q_{\mathcal{S}}$ w.r.t. $D$ for every $\mathcal{S}$-database $D$. So, we are naturally led to the notion of perfect source-to-ontology rewriting. In what follows, $q_{\mathcal{S}}$ refers to a query over $\mathcal{S}$, and $q_{\mathcal{O}}$ to a query over $\mathcal{O}$ of the same arity.

**Definition 1.** *$q_{\mathcal{O}}$ is a* perfect $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting *of $q_{\mathcal{S}}$ if for every $\mathcal{S}$-database $D$, $Mod_D(\Sigma) \neq \emptyset$ implies $q_{\mathcal{S}}^D = cert_{q_{\mathcal{O}},\Sigma}^D$.*

As noted in [6,10] and illustrated in the next example, a perfect source-to-ontology rewriting of $q_{\mathcal{S}}$ may not exist.

*Example 2.* Refer to Example 1, and consider the data service exprresed as the source query $q_{\mathcal{S}}(x) = \{(x) \mid s_1(x)\}$. By inspecting the mappings, one can see that, since the certain answers of Student include the values stored both in $s_1$ and in $s_3$, such concept is too general for exactly characterizing $q_{\mathcal{S}}$. On the other hand, it can also seen that both ErasmusStudent and MathStudent are too specific, and therefore we can conclude that no perfect $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ exists.

In order to cope with the situations illustrated in the example, we introduce the notions of sound and complete source-to-ontology rewritings, which, intuitively, provide sound and complete approximations of perfect rewritings, respectively.

**Definition 2.** *$q_{\mathcal{O}}$ is a* sound *(respectively,* complete*) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ if for every $\mathcal{S}$-database $D$, $Mod_D(\Sigma) \neq \emptyset$ implies $cert^D_{q_{\mathcal{O}},\Sigma} \subseteq q^D_{\mathcal{S}}$ (resp., $q^D_{\mathcal{S}} \subseteq cert^D_{q_{\mathcal{O}},\Sigma}$).*

*Example 3.* We refer to Example 2, and observe that $\{(x) \mid \mathsf{ErasmusStudent}(x) \wedge \mathsf{MathStudent}(x)\}$ is a sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, whereas $\{(x) \mid \mathsf{Student}(x)\ \}$ is a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$.

Obviously, $q_{\mathcal{O}}$ is a perfect $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ if and only if $q_{\mathcal{O}}$ is both a sound, and a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$. There are also interesting relationships between the notions of $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewritings introduced here and the usual notions of rewritings studied in OBDA.

**Proposition 1.** *$q_{\mathcal{O}}$ is a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ if and only if $q_{\mathcal{S}}$ is an $\mathcal{O}$-to-$\mathcal{S}$ $\Sigma$-rewriting of $q_{\mathcal{O}}$. If $q_{\mathcal{S}}$ is a perfect $\mathcal{O}$-to-$\mathcal{S}$ $\Sigma$-rewriting of $q_{\mathcal{O}}$, then $q_{\mathcal{O}}$ is a perfect $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$.*

It is easy to see that different sound or complete source-to-ontology rewritings of $q_{\mathcal{S}}$ may exist, and therefore it is reasonable to look for the "best" approximations of $q_{\mathcal{S}}$, at least relative to a certain class of queries.

**Definition 3.** *$q_{\mathcal{O}} \in \mathcal{L}$ is* an $\mathcal{L}$-maximally sound *(respectively,* $\mathcal{L}$-minimally complete*) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ if $q_{\mathcal{O}}$ is a sound (resp. complete) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ and no $q' \in \mathcal{L}$ exists such that $(i)$ $q'$ is a sound (resp., complete) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, $(ii)$ $cert_{q_{\mathcal{O}},\Sigma} \sqsubseteq cert_{q',\Sigma}$ (resp., $cert_{q',\Sigma} \sqsubseteq cert_{q_{\mathcal{O}},\Sigma}$), and $(iii)$ there exists an $\mathcal{S}$-database $D$ s.t. $cert^D_{q_{\mathcal{O}},\Sigma} \subset cert^D_{q',\Sigma}$ (resp., $cert^D_{q',\Sigma} \subset cert^D_{q_{\mathcal{O}},\Sigma}$).*

*Example 4.* We refer again to Example 2, and observe that while $\{(x) \mid \mathsf{Student}(x)\ \}$ is the minimally complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ in the class of UCQs, both $\{(x) \mid \mathsf{ErasmusStudent}(x)\}$, and $\{(x) \mid \mathsf{MathStudent}(x)\}$ are maximally sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewritings of $q_{\mathcal{S}}$ in the class of CQs, while $\{(x) \mid \mathsf{ErasmusStudent}(x) \vee \mathsf{MathStudent}(x)\}$ is so in the class of UCQs.

Given the general framework presented so far, it is natural to consider the following two basic computational problems, for classes $\mathcal{L}_{\mathcal{S}}$ and $\mathcal{L}_{\mathcal{O}}$ of queries:

- *Verification*: given $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$, $q_{\mathcal{S}} \in \mathcal{L}_{\mathcal{S}}$ over $\mathcal{S}$ and $q_{\mathcal{O}} \in \mathcal{L}_{\mathcal{O}}$ over $\mathcal{O}$ of the same arity as $q_{\mathcal{S}}$, verify whether $q_{\mathcal{O}}$ is a sound (resp., complete, perfect) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewritings of $q_{\mathcal{S}}$.
- *Computation*: given $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$, and $q_{\mathcal{S}} \in \mathcal{L}_{\mathcal{S}}$ over $\mathcal{S}$ compute any $\mathcal{L}_{\mathcal{O}}$-maximally sound (resp., $\mathcal{L}_{\mathcal{O}}$-minimally complete, perfect) $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, if it exists.

In the rest of this paper, if not otherwise stated, we refer to the most common setting studied in OBDA, i.e., where $(i)$ the ontology is expressed in *DL-Lite*$_{\mathcal{R}}$, $(ii)$ $\mathcal{S}$ is a relational database schema without integrity constraints, and $(iii)$ both $\mathcal{L}_{\mathcal{O}}$ and $\mathcal{L}_{\mathcal{S}}$ denote the class of UCQs. Interestingly, in this case, we have the following.

**Proposition 2.** *If $q_1$ and $q_2$ are two UCQ-minimally complete, or UCQ-maximally sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewritings of $q_{\mathcal{S}}$, then they are equivalent w.r.t. $\Sigma$.*

## Complete source-to-ontology rewritings

In this section, we study both the verification and the computation problem for complete source-to-ontology rewritings.

**Verification.** Suppose we want to check whether $q_{\mathcal{O}}$ is a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$. Obviously, if $q_{\mathcal{S}}$ is contained in $\mathsf{PerfRef}_{q_{\mathcal{O}},\Sigma}$, then for every $\mathcal{S}$-database $D$ consistent with $\Sigma$, we have that $q_{\mathcal{S}}^{D} \subseteq cert_{q_{\mathcal{O}},\Sigma}^{D}$ and therefore the answer is positive. However, if $q_{\mathcal{S}}$ is not contained in $\mathsf{PerfRef}_{q_{\mathcal{O}},\Sigma}$, it might be that $q_{\mathcal{O}}$ is still a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, in the case where the non-emptiness of $q_{\mathcal{S}}$ in $D$ reveals the presence of inconsistencies. From this observation, we derive the following characterization.

**Proposition 3.** *$q_{\mathcal{O}}$ is a complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}(\boldsymbol{t})$ if and only if $q_{\mathcal{S}} \sqsubseteq \mathit{PerfRef}_{q_{\mathcal{O}},\Sigma} \vee \mathit{PerfRef}_{\mathcal{V}_{\mathcal{O}}^{\boldsymbol{t}},\Sigma}$.*

The following theorem characterizes the complexity of verification for complete source-to-ontology rewritings.

**Theorem 1.** *Verification for complete source-to-ontology rewritings is NP-complete.*

**Computation.** Our algorithm for the computation of minimally complete source-to-ontology rewritings is below.

```
Algorithm 1
Input: $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$, $q_{\mathcal{S}}(\boldsymbol{t}) = q_{\mathcal{S}}^{1}(\boldsymbol{t}) \vee \ldots \vee q_{\mathcal{S}}^{n}(\boldsymbol{t})$ over $\mathcal{S}$
Output: $q_{\mathcal{O}}(\boldsymbol{t})$ over $\mathcal{O}$
begin return $q_{\mathcal{O}} = \{\boldsymbol{t} \mid \mathcal{M}(q_{\mathcal{S}}^{1}) \wedge \top(\boldsymbol{t}) \vee \ldots \vee \mathcal{M}(q_{\mathcal{S}}^{n}) \wedge \top(\boldsymbol{t})\}$ end
```

Intuitively, the algorithm computes the output query as union of CQs obtained by simply applying the mapping $\mathcal{M}$ to each CQ $q_{\mathcal{S}}^{i}$ in $q_{\mathcal{S}}$, using $\top$ to bind the variables that are not involved in the application of $\mathcal{M}$ to $q_{\mathcal{S}}^{i}$.

**Theorem 2.** *Algorithm 1 computes the UCQ-minimally complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$.*

The algorithm shows that the UCQ-minimally complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ always exists. Moreover, if $q_{\mathcal{S}}$ is a CQ, then it is a CQ. Finally, we observe that the complexity of Algorithm 1 does not depend on $\mathcal{O}$, is in ExpTime in $\sigma(\mathcal{M})$, and in PTime in $\sigma(q_{\mathcal{S}})$, where we use $\sigma(x)$ to denotes the size of $x$. It can be shown that a PTime algorithm for computing the UCQ-minimally complete source-to-ontology rewriting would imply a PTime algorithm for CQ containment. So, assuming PTime $\neq$ NP, computation cannot be solved in PTime.

## Sound source-to-ontology rewritings

We now turn to verification and computation of sound source-to-ontology rewritings.

**Verification.** Notice that, since for an $\mathcal{S}$-database $D$ consistent with $\Sigma$, $\mathsf{PerfRef}_{q_{\mathcal{O}},\Sigma}^{D}$ computes exactly $cert_{q_{\mathcal{O}},\Sigma}^{D}$, checking whether $q_{\mathcal{O}}$ is a sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ means checking whether for all $\mathcal{S}$-databases $D$, either $Mod_{D}(\Sigma) = \emptyset$ or $\mathsf{PerfRef}_{q_{\mathcal{O}},\Sigma}^{D} \subseteq q_{\mathcal{S}}^{D}$. This observation leads to the following characterization.

**Proposition 4.** *$q_{\mathcal{O}}(\boldsymbol{t})$ is a sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ if and only if $\mathsf{PerfRef}_{q_{\mathcal{O}},\Sigma} \sqsubseteq q_{\mathcal{S}} \vee \mathsf{PerfRef}_{\mathcal{V}^{t}_{\mathcal{O}},\Sigma}$.*

The following theorem characterizes the complexity of the verification problem for sound source-to-ontology rewritings.

**Theorem 3.** *Verification for sound source-to-ontology rewritings is $\Pi^p_2$-complete.*

**Computation.** We address the problem of computing UCQ-maximally sound source-to-ontology rewritings. Our main result is that there are many cases where a UCQ-maximally sound source-to-ontology rewriting of a query is not guaranteed to exist. To illustrate the result, we introduce a specific setting, that we call *restricted*, obtained from the general one by: $(i)$ limiting the ontology language to *DL-Lite*$_{\mathsf{RDFS}}$, $(ii)$ limiting the mapping to pure GAV, and $(iii)$ limiting $q_{\mathcal{S}}$ to UCQJFEs. The following shows that, surprisingly, as soon as we try to extend such setting, we lose the guarantee of the existence of source-to-ontology rewritings that are maximally sound.

**Theorem 4.** *UCQ-maximally sound source-to-ontology rewritings of a query $q_{\mathcal{S}}$ may not exist if we extend the restricted setting with each of the following features: (*i*) disjointness axioms in the ontology; (*ii*) inclusion axioms with $\exists R$ as right-hand side in the ontology; (*iii*) LAV mapping assertions, even without joins involving existential variables in the right-hand side; (*iv*) non-pure GAV mapping assertions; (*v*) the fragment of UCQ for $q_{\mathcal{S}}$ beyond UCQJFE.*

*Proof sketch.* We present the proof for case 5. Consider the OBDA specification $\Sigma = \langle \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$, where $\mathcal{O}$ has no axiom, and $\mathcal{M}$ is the following pure GAV mapping:

$$\begin{array}{ll} \{(x,y) \mid s_1(y,y) \wedge s_3(x)\} & \rightarrow \{(x,y) \mid R(x,y)\} \\ \{(x,y) \mid s_1(x,y)\} & \rightarrow \{(x,y) \mid R(x,y)\} \end{array}$$

and let $q_{\mathcal{S}}$ be the query $\{() \mid s_1(x,y) \wedge s_1(y,z)\}$. Observe that $q'_{\mathcal{O}} = \{() \mid R(x,y) \wedge R(y,z)\}$ is the complete $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, but is not sound, because the query $q'_s = \{() \mid s_1(x,y) \wedge s_1(z,z) \wedge s_3(y)\}$ is a disjunct of $\mathsf{PerfRef}_{q'_{\mathcal{O}},\Sigma}$ such that $q'_{\mathcal{S}} \not\sqsubseteq q_{\mathcal{S}}$. Conversely, one can verify that each of the following queries is a sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$:

- $q_0 = \{() \mid R(x,y) \wedge R(y,y)\}$,
- $q_1 = \{() \mid R(x,y) \wedge R(y,z_1) \wedge R(z_1,y)\}$,
- $q_2 = \{() \mid R(x,y) \wedge R(y,z_1) \wedge R(z_1,z_2) \wedge R(z_2,y)\}$,
- ...

More precisely, if we define $q_n$ to be $\{() \mid R(x,y) \wedge R(y,z_1) \wedge R(z_1,z_2) \wedge \ldots \wedge R(z_{n-1},z_n) \wedge R(z_n,y)\}$, for $n \geq 2$,, then it can be shown that every $q_n$ is a sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$, and for no pair $(i,j)$, with $i \neq j$, $i,j \geq 0$, $cert_{q_i,\Sigma} \sqsubseteq cert_{q_j,\Sigma}$. It follows that the infinite union $q$ of $q_0$, $q_1$, and all $q_n$'s can be shown to be the maximally sound $\mathcal{S}$-to-$\mathcal{O}$ $\Sigma$-rewriting of $q_{\mathcal{S}}$ in the class of positive queries. ☐

We observe that, despite its limitations, the expressive power of the restricted setting is sufficient for several meaningful applications. Indeed, several popular ontologies are expressible in *DL-Lite*$_{\mathsf{RDFS}}$, and the form of pure GAV mapping is exactly the one originally defined in the literature of data integration. Moreover, UCQJFEs capture data services expressible in the famous SPJ (Select, Project, Join) fragment of Relational Algebra, with the only limitation of forbidding projection outside join, which makes them suitable for all tasks related to source profiling. Interestingly, it can be shown that in the restricted setting maximally sound source-to-ontology rewritings always exist.

## Conclusion

We have presented a framework for semantically characterizing data services through ontologies, and carried out a comprehensive analysis for the most common OBDA setting. We plan to continue this work along several directions. For example, in the unrestricted setting, we aim at (*i*) studying the problem of checking for the existence of a UCQ-maximally sound source-to-ontology rewriting of a query, and computing it in case it exists, and (*ii*) singling out the minimal class $\mathcal{L}_O$ of queries that guarantees the existence of an $\mathcal{L}_O$-maximally sound source-to-ontology rewritings. Furthermore, we will extend our analysis to OBDA settings based on other DLs as ontology languages.

**Acknowlegments.** Work supported by MIUR under the SIR project "MODEUS" – grant n. RBSI14TQHQ, and by Sapienza research project "PRE-O-PRE".

## References

1. S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison Wesley Publ. Co., 1995.
2. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.
3. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
4. D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, and G. Vetere. *DL-Lite*: Practical reasoning for rich DLs. In *Proc. of DL 2004*, volume 104 of *CEUR,* `ceur-ws.org`, 2004.
5. M. J. Carey, N. Onose, and M. Petropoulos. Data services. *Comm. of the ACM*, 55(6):86–97, 2012.
6. G. Cima. Preliminary results on ontology-based open data publishing. In *Proc. of DL 2017*, volume 1879 of *CEUR,* `ceur-ws.org`, 2017.
7. G. Cima, M. Lenzerini, and A. Poggi. Semantic technology for open data publishing. In *Proc. of WIMS 2017*, 2017.
8. G. Cima, M. Lenzerini, and A. Poggi. Semantic characterization of data services through ontologies. In *Proc. of IJCAI 2019*, 2019.
9. M. Lenzerini. Managing data through the lens of an ontology. *AI Magazine*, 39(2):65–74, 2018.
10. C. Lutz, J. Marti, and L. Sabellek. Query expressibility and verification in ontology-based data access. In *Proc. of KR 2018*, pages 389–398, 2018.
11. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.

12. G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, and M. Zakharyaschev. Ontology-based data access: A survey. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 5511–5519, 2018.
13. Z. Zheng, J. Zhu, and M. R. Lyu. Service-generated big data and big data-as-a-service: An overview. In *Proc. of the 2013 IEEE Int. Conf. on Big Data*, pages 403–410, 2013.