Approaching OBDA Evolution through Mapping Repairs[⋆]

Domenico Lembo[1], Riccardo Rosati[1], Valerio Santarelli[1],
Domenico Fabio Savo[1], Evgenij Thorstensen[2]

[1] Sapienza Università di Roma
`lastname@dis.uniroma1.it`

[2] University of Oslo
`evgenit@ifi.uio.no`

Ontology-based data access (OBDA) is a new paradigm for accessing source databases through mediation of a conceptual domain view, given in terms of an ontology [9]. A major issue in OBDA is the design of an OBDA specification and the management of its evolution. An *OBDA specification* is constituted by an ontology, usually a Description Logic (DL) TBox, a schema of the source databases, and a declarative mapping specifying the semantic relationship between the data at the sources and the elements of the ontology. In the following we denote it by $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$, where $\mathcal{T}$ is the TBox (a set of DL axioms), $\mathcal{S}$ the source schema, i.e., a relational signature possibly with integrity constraints (ICs), and the mapping $\mathcal{M}$ is a set of mapping assertions of the form $\phi(\boldsymbol{x}) \rightsquigarrow \psi(\boldsymbol{x})$, where $\phi(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ are queries over $\mathcal{S}$ and $\mathcal{T}$, respectively, both with free variables $\boldsymbol{x}$ (such an assertion is called a GLAV mapping assertion if both $\phi(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ are conjunctive queries (CQs), while it is a GAV mapping assertion if it is GLAV and $\psi(\boldsymbol{x})$ is an atom without occurrences of non-free variables).

The design of the specification is normally conducted in an iterative fashion, and changes are continuously implemented to its various components. Also, the entire specification is often a lively artifact, continuously modified due to, e.g., changes in the requirements. Due to these characteristics, OBDA design and maintenance must be supported by adequate tools and methodologies. The mapping is certainly the component of the specification which has received so far less attention, and thus consolidated tools supporting its design are currently not available. Mapping design is a time-consuming and complex operation, which typically (and especially in complex scenarios) has to be conducted manually [1]. Of course, modifying the mapping due to changes in the other components of the specification is tedious and time-consuming as well.

In this paper we study the evolution of OBDA specifications. We start our investigation by observing that many approaches exist for both *ontology evolution* [12] and *database schema evolution* [11]. However, to the best of our knowledge, no previous study has analyzed evolution in the presence of mappings connecting an ontology to a database schema. In this sense, a problem that is close to OBDA is *ontology matching and alignment*, which is based on the use of a notion of *mapping* to integrate different ontologies. Several works have studied the problem of repairing inconsistent mappings in this context (e.g., [3, 7, 8, 10]). However, the framework of ontology matching, and in particular the notion of mapping, is very different from OBDA.

We adopt a *mapping-centered* notion of OBDA evolution: given an OBDA specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$, we want to repair the mapping $\mathcal{M}$ given a modification of the

---

[⋆] The present paper is an extended abstract of [6].

TBox $\mathcal{T}$ and/or of the source schema $\mathcal{S}$. We think that, at least for a first analysis of evolution in OBDA, this is a natural assumption: indeed, the mapping is an information that depends on both the TBox and the source schema, while the TBox and the schema are (at least in principle) semantically independent entities.

Following the classical approaches to belief revision, we look for a notion of repair of a mapping that is based on two general principles: (i) preserving *consistency* of the OBDA specification; (ii) expressing *minimal change* with respect to the initial OBDA specification. With respect to consistency preservation, we adopt a non-classical notion of inconsistency for an OBDA specification, called *global mapping inconsistency*, recently introduced in [4, 5]. According to this notion, a mapping $\mathcal{M}$ is inconsistent with respect to a TBox $\mathcal{T}$ and a source schema $\mathcal{S}$ if there exists no instance D for $\mathcal{S}$ such that D is consistent with $\mathcal{J}$ and $\mathcal{M}$ *is active* on D, i.e., every query over the source schema appearing in $\mathcal{M}$ has a non-empty answer in D.

Definition 1 (Global mapping inconsistency [5]). *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification. We say that $\mathcal{M}$* is globally inconsistent for *$\langle \mathcal{T}, \mathcal{S} \rangle$ if there does not exist a source instance D satisfying the ICs in $\mathcal{S}$ such that (i) $\mathcal{M}$ is active on D; and (ii) there exists an interpretation that agrees with D on the predicates of $\mathcal{S}$ and satisfies $\mathcal{J}$.*

Global mapping inconsistency provides a more meaningful notion of inconsistency than the classical one in the context of OBDA: for instance, in all the cases when the source schema is a relational database schema with standard integrity constraints, the OBDA specification is inconsistent according to the classical semantics if and only if its TBox is inconsistent (which in turn implies that this notion is trivial for many DLs).

With respect to minimal change, we propose two different notions of repair. The first one is called *deletion-based mapping repair* and reflects the simple idea of repairing a mapping through a (subset-)minimal deletion of assertions from the initial mapping.

Definition 2 (Deletion-based mapping repair). *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification such that $\mathcal{M}$ is globally consistent for $\langle \mathcal{T}, \mathcal{S} \rangle$, $\mathcal{T}'$ a consistent TBox, $\mathcal{S}'$ a consistent source schema, and $\mathcal{M}'$ a mapping such that $\mathcal{M}' \subseteq \mathcal{M}$. We say that $\mathcal{M}'$ is a* deletion-based mapping repair (DMR) for $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S}' \rangle$ *if: (i) $\mathcal{M}'$ is globally consistent for $\langle \mathcal{T}', \mathcal{S}' \rangle$, and (ii) there exists no mapping $\mathcal{M}''$ such that $\mathcal{M}''$ is globally consistent for $\langle \mathcal{T}', \mathcal{S}' \rangle$, and $\mathcal{M}' \subset \mathcal{M}'' \subseteq \mathcal{M}$.*

The second notion of repair, called *entailment-based mapping repair*, relies on the notion of *mapping entailment set (MES)*: the mapping entailment set of an OBDA specification $\mathcal{J}$ for a mapping language $\mathcal{L}$ is the set of mapping assertions in $\mathcal{L}$ that are logical consequences of $\mathcal{J}$. Then, the repairs are the globally consistent subsets of the MES that are selected according to a preference criterion (the notion of *fewer changes* [2]) that formalizes the intuitive principle of preferring insertions over deletions. In practice, a mapping $\mathcal{M}_1$ has fewer deletions (resp., fewer insertions) than a mapping $\mathcal{M}_2$ with respect to a mapping $\mathcal{M}$ if $\mathcal{M} \setminus \mathcal{M}_1 \subset \mathcal{M} \setminus \mathcal{M}_2$ (resp., $\mathcal{M}_1 \setminus \mathcal{M} \subset \mathcal{M}_2 \setminus \mathcal{M}$), and $\mathcal{M}_1$ has fewer changes than $\mathcal{M}_2$ with respect to $\mathcal{M}$ if either $\mathcal{M}_1$ has fewer deletions than $\mathcal{M}_2$ with respect to $\mathcal{M}$, or $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same deletions with respect to $\mathcal{M}$, and $\mathcal{M}_1$ has fewer insertions than $\mathcal{M}_2$ with respect to $\mathcal{M}$.

Definition 3 (Entailment-based mapping repair). *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification such that $\mathcal{M}$ is globally consistent for $\langle \mathcal{T}, \mathcal{S} \rangle$, $\mathcal{T}'$ a consistent TBox, $\mathcal{S}'$ a* *consistent source schema, $\mathcal{L}$ a mapping language, and $\mathcal{M}'$ an $\mathcal{L}$-mapping. We say that $\mathcal{M}'$ is an* entailment-based $\mathcal{L}$-mapping repair *($\mathcal{L}$-EMR) for $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S}' \rangle$ if: (i) $\mathcal{M}'$ is globally consistent for $\langle \mathcal{T}', \mathcal{S}' \rangle$; and (ii) there exists no $\mathcal{L}$-mapping $\mathcal{M}''$ such that $\mathcal{M}''$ is globally consistent for $\langle \mathcal{T}', \mathcal{S}' \rangle$ and $MES_{\mathcal{L}}(\langle \mathcal{T}', \mathcal{S}', \mathcal{M}'' \rangle)$ has fewer changes than $MES_{\mathcal{L}}(\langle \mathcal{T}', \mathcal{S}', \mathcal{M}' \rangle)$ with respect to $MES_{\mathcal{L}}(\mathcal{J})$.*

Example 1. Consider the OBDA specification $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ where: $\mathcal{T} = \{C \sqsubseteq F,\ \ C \sqsubseteq A,\ \ \exists R \sqsubseteq B,\ \ A \sqsubseteq B\}$, $\mathcal{S} = \{T_1/2\}$, $\mathcal{M} = \{T_1(x,y) \rightsquigarrow R(x,y),\ \ T_1(x,y) \rightsquigarrow C(x)\}$, and consider the TBox $\mathcal{T}' = \mathcal{T} \cup \{A \sqsubseteq \neg\exists R\}$. It is easy to see that $\mathcal{M}$ is not globally consistent for $\langle \mathcal{T}', \mathcal{S} \rangle$, since for every database that activates $\mathcal{M}$ the mapping produces two facts of the form $R(x,y)$ and $C(x)$, which violate the TBox assertion $C \sqsubseteq \neg\exists R$ inferred by $\mathcal{T}'$. Then, the DMRs of $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S} \rangle$ are $\mathcal{M}'_1 = \{T_1(x,y) \rightsquigarrow R(x,y)\}$ and $\mathcal{M}'_2 = \{T_1(x,y) \rightsquigarrow C(x)\}$. On the other hand, one can easily verify that the following are GAV-EMRs of $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S} \rangle$: $\mathcal{M}''_1 = \{T_1(x,y) \rightsquigarrow R(x,y),\ T_1(x,y) \rightsquigarrow F(x)\}$ and $\mathcal{M}''_2 = \{T_1(x,y) \rightsquigarrow C(x)\}$. Such repairs are contained in all GAV-EMRs of $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S} \rangle$. $\square$

Then, we define query entailment under mapping repairs, which corresponds to a form of skeptical reasoning over all the mapping repairs.

Definition 4 (Query Entailment). *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification such that $\mathcal{M}$ is globally consistent for $\langle \mathcal{T}, \mathcal{S} \rangle$, $\mathcal{T}'$ a consistent TBox, $\mathcal{S}'$ a consistent source schema, D a source instance satisfying the ICs in $\mathcal{S}'$, and q a Boolean CQ over over the signature of $\mathcal{T}'$. We say that q* is entailed under DMR (resp. $\mathcal{L}$-EMR where $\mathcal{L}$ is a mapping language) by $\mathcal{J}$, $\mathcal{T}'$, $\mathcal{S}'$, and D *if $(\langle \mathcal{T}', \mathcal{S}', \mathcal{M}' \rangle, D) \models q$ for every DMR (respectively, for every $\mathcal{L}$-EMR) for $\mathcal{J}$ under update $\langle \mathcal{T}', \mathcal{S}' \rangle$.*

Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ and $\mathcal{T}'$ be as in Example 1, and let $D = \{T_1(a,b)\}$. Given the DMRs $\mathcal{M}'_1, \mathcal{M}'_2$ and the GAV-EMRs $\mathcal{M}''_1, \mathcal{M}''_2$ above described, it follows that the query $B(a)$ is entailed under DMR by $\mathcal{J}$, $\mathcal{T}'$, $\mathcal{S}'$, and D; moreover, the query $\exists x. F(x)$ is not entailed under DMR by $\mathcal{J}$, $\mathcal{T}'$, $\mathcal{S}'$, and D, while it is entailed under GAV-EMR.

We finally provide initial results on the complexity of query entailment, focusing on *DL-Lite*$_R$ TBoxes, *simple* source schemas (i.e., without ICs), and on GAV and GLAV mappings. We first establish an exact bound for query entailment under deletion-based mapping repairs, which holds for both GAV and GLAV mappings.

Theorem 1. *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification, with $\mathcal{T}$ a DL-Lite$_R$ TBox, $\mathcal{S}$ a simple source schema, and $\mathcal{M}$ a GLAV mapping. Let $\mathcal{T}'$ be DL-Lite$_R$ TBox, $\mathcal{S}'$ a simple source schema, D an instance for $\mathcal{S}$, and q a Boolean CQ over the signature of $\mathcal{T}'$. Deciding whether q is entailed under DMR by $\mathcal{J}$, $\mathcal{T}'$, $\mathcal{S}'$, and D is Π_2^p-complete.*

Then, we study the same problem under entailment-based mapping repairs and GAV mappings, and prove that the Π_2^p exact bound holds also in this case.

Theorem 2. *Let $\mathcal{J} = \langle \mathcal{T}, \mathcal{S}, \mathcal{M} \rangle$ be an OBDA specification, where $\mathcal{T}$ is a DL-Lite$_R$ TBox, $\mathcal{S}$ is a simple source schema, and $\mathcal{M}$ is a GAV mapping. Let $\mathcal{T}'$ be DL-Lite$_R$ TBox, $\mathcal{S}'$ a simple source schema, D an instance for $\mathcal{S}'$, and q a Boolean CQ over the signature of $\mathcal{T}'$. Deciding whether q is entailed under GAV-EMR by $\mathcal{J}$, $\mathcal{T}'$, $\mathcal{S}'$, and D is Π_2^p-complete.*

References

1. N. Antonioli, F. Castanò, S. Coletta, S. Grossi, D. Lembo, M. Lenzerini, A. Poggi, E. Virardi, and P. Castracane. Ontology-based data management for the italian public debt. In *Proc. of the 8th Int. Conf. on Formal Ontology in Information Systems (FOIS)*, pages 372–385, 2014.
2. R. Fagin, G. M. Kuper, J. D. Ullman, and M. Y. Vardi. Updating logical databases. *Advances in Computing Research*, 3:1–18, 1986.
3. E. Jiménez-Ruiz, C. Meilicke, B. C. Grau, and I. Horrocks. Evaluating mapping repair systems with large biomedical ontologies. In *Proc. of the 26th Int. Workshop on Description Logic (DL)*, CEUR Electronic Workshop Proceedings, `http://ceur-ws.org/`, pages 246–257, 2013.
4. D. Lembo, J. Mora, R. Rosati, D. F. Savo, and E. Thorstensen. Towards mapping analysis in ontology-based data access. In *Proc. of the 8th Int. Conf. on Web Reasoning and Rule Systems (RR)*, volume 8741 of *Lecture Notes in Computer Science*, pages 108–123. Springer, 2014.
5. D. Lembo, J. Mora, R. Rosati, D. F. Savo, and E. Thorstensen. Mapping analysis in ontology-based data access: Algorithms and complexity. In *Proc. of the 14th Int. Semantic Web Conf. (ISWC)*, pages 217–234, 2015.
6. D. Lembo, R. Rosati, V. Santarelli, D. F. Savo, and E. Thorstensen. Approaching OBDA evolution through mapping repairs. Manuscript, 2016.
7. C. Meilicke, H. Stuckenschmidt, and A. Tamilin. Reasoning support for mapping revision. *J. Log. Comput.*, 19(5):807–829, 2009.
8. Ö. L. Özçep. Towards principles for ontology integration. In *Proc. of the 5th Int. Conf. on Formal Ontology in Information Systems (FOIS)*, pages 137–150, 2008.
9. 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.
10. G. Qi, Q. Ji, and P. Haase. A conflict-based operator for mapping revision. In *Proc. of the 22nd Int. Workshop on Description Logic (DL)*, CEUR Electronic Workshop Proceedings, `http://ceur-ws.org/`, 2009.
11. E. Rahm and P. A. Bernstein. An online bibliography on schema evolution. *SIGMOD Record*, 35(4):30–31, 2006.
12. F. Zablith, G. Antoniou, M. d'Aquin, G. Flouris, H. Kondylakis, E. Motta, D. Plexousakis, and M. Sabou. Ontology evolution: a process-centric survey. *Knowledge Eng. Review*, 30(1):45–75, 2015.