Real-world Assessment of Policy-Protected OBDA (Extended Abstract)

Divya Baura[1], Diego Calvanese[2]

[1]Umeå Universitet, Umeå, Sweden
[2]Free University of Bozen-Bolzano, Bolzano, Italy

Abstract

Within the Ontology Based Data Access (OBDA) framework, users can query relational data sources using an ontology to which the source is linked via declarative mappings. In a world where data sharing is widespread, ensuring privacy while managing data poses a significant challenge. Controlled Query Evaluation (CQE) is a privacy preserving query answering framework in the presence of ontologies, where policies representing confidential information are used to devise suitable censors that enforce data protection. The integration of CQE within OBDA was recently proposed through the *Policy-Protected OBDA* (PPOBDA) framework, which is based on embedding policies into mappings. Such framework is essentially theoretical, and the effectiveness with which PPOBDA policies are able to capture real-world privacy requirements has not been assessed so far. In this work, we carry out such an evaluation, utilizing the well-known MIMIC-III hospital dataset, which recently has been mapped, by adopting the OBDA framework, to the *Fast Healthcare Interoperability Resources* (FHIR) ontology. We identify relevant privacy requirements by analyzing the legal regulations on data sharing expressed in HIPAA of US Federal Law and GDPR of the EU, show how they can be expressed via PPOBDA policies, and analyze the impact of these policies on the answers to a set of representative queries. Our analysis exposes both strengths and weaknesses of the PPOBA framework in relation to these practically relevant privacy regulations. Furthermore, we perform a performance evaluation of the OBDA framework implemented over the MIMIC-III dataset via the FHIR ontology, assessing the overhead introduced by the PPOBDA policies and its implications on such real-world use case.

1. Introduction

Ontology-Based Data Access (OBDA) [1, 2] provides a powerful framework for querying relational data sources using ontologies. It supports user-friendly access to data by allowing queries over a conceptual vocabulary of ontologies while relying on mappings to translate these queries to the underlying data. Our work focuses on OBDA systems that use lightweight description logics, particularly OWL 2 QL. While OBDA offers efficient query answering, it also raises privacy concerns, as sensitive information can be derived from the data retrieved through the mappings, combined with the inferences via the ontology axioms.

To address such concerns, Controlled Query Evaluation (CQE) has emerged as a privacy-preserving mechanism within OBDA [3, 4, 5]. CQE enforces privacy through policies that specify which information must be protected, using a censor to filter query answers accordingly. Building on this, the Policy-Protected OBDA (PPOBDA) framework introduces policies that are first order denial assertions encoded in mappings, aiming for stronger integration of privacy and data access [6].

In this work, we explore the application of PPOBDA in the healthcare domain, where privacy is paramount due to legal and ethical considerations. We reference major regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) [7] and the General Data Protection Regulation (GDPR) [8] to identify key privacy requirements [9] and then use these requirements to test the relevance of policies expressed in the PPOBDA framework. Our case study uses the MIMIC-III hospital dataset [10] structured via the OMOP Common Data Model [11] and mapped to the FHIR

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

divya.baura@umu.se (D. Baura); diego.calvanese@unibz.it (D. Calvanese)
https://www.umu.se/personal/divya-baura/ (D. Baura); https://www.inf.unibz.it/~calvanese/ (D. Calvanese)
0000-0002-5237-9927 (D. Baura); 0000-0001-5174-9693 (D. Calvanese)

ontology [12, 13, 14]. Due to the size and complexity of FHIR, we apply ontology modularization techniques to manage the execution process.

We present representative PPOBDA policies, analyze their effectiveness in expressing real-world privacy constraints, and assess limitation particularly in addressing practices like data de-identification. Our experimental evaluation investigates how privacy policies impact query results and performance. These findings contribute to understanding how policy-driven privacy mechanisms can be practically applied in OBDA systems.

Code and resources are available at https://github.com/divyabaura/PPOBDA-policies.

2. Methodology

In this section, we outline the steps undertaken to construct and validate a privacy-aware Ontology-Based Data Access (OBDA) framework applied to a real-world healthcare dataset.

Data Source and Standardization. We utilize the MIMIC-III clinical dataset[10], a large, de-identified dataset containing health-related information from ICU patients. To harmonize the data with existing standards, we employ the open-source `mimic-omop` ETL tool [15] to transform MIMIC-III into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), specifically version 5.4. This transformation facilitates the use of standardized vocabularies from the OHDSI initiative [16].

Ontology Selection and Module Extraction. To align the data with a semantic model, we adopt the FHIR Ontology [13], which provides an OWL-based formalization of FHIR resources. Due to its size (over 1,450 classes), we extract a relevant module using the Syntactic Locality Module Extractor[17] with the STAR method. The seed signature includes essential clinical concepts such as `Patient`, `Address`, and `CodeableConcept`. This ensures that the extracted module retains semantic integrity while being computationally tractable.

Metadata Extraction and Integration. We leverage Ontop's CLI [18] to extract metadata from the transformed OMOP database. This metadata is represented as a JSON file and includes schema-level information such as table names, column data types, and foreign key constraints. It forms the structural bridge between the database and the ontology, enabling OBDA reasoning.

Mapping Specification. We incorporate existing R2RML mappings from Xiao et al. [14], linking OMOP CDM elements to FHIR RDF representations. These mappings span over 100 data elements across 11 OMOP tables and 11 FHIR classes, and they allow SPARQL queries to be translated into SQL over the relational schema while respecting the semantics of FHIR.

System Setup and Execution. We have deployed the OBDA system using Ontop and connected it to a PostgreSQL instance containing the OMOP-transformed MIMIC-III data. The system supports semantic querying via SPARQL and enforces policy-aware access through the integration of the PPOBDA framework, which will be detailed in subsequent sections.

3. Legal Analysis and Results

We combine our analysis of how PPOBDA addresses key legal privacy requirements of HIPAA and GDPR with experimental results on PPOBDA policy enforcement over the MIMIC-III dataset. To provide an intuitive understanding of how in PPOBDA, policies affect the instances of a class C, assume that all denials in the set $\mathcal{P}$ of policies containing an atom that unifies with $C(x)$ are $\forall x, \vec{y_i}. (C(x) \wedge \alpha_i(x, \vec{y_i}) \rightarrow \bot)$, for $i \in \{1, \ldots, k\}$. Then, the PPOBDA mapping reformulation algorithm replaces the atom $C(x)$

with $C(x) \wedge \bigwedge_{1 \leq i \leq k} \neg \exists \vec{y_i}.\, \alpha_i(x, \vec{y_i})$ [19], and such expression is rewritten w.r.t. the TBox and unfolded w.r.t. the original mappings and incorporated in the source part of the new mapping assertions, thus expressing the PPOBDA policies.

Protected Health Information (PHI). Under HIPAA, PHI comprises 18 identifiers—demographic, geographic, and medical data that require strict protection [20]. We have defined within PPOBDA several policies to safeguard PHI by ensuring that identifiable attributes are concealed. We list here a few meaningful examples, and refer to the GitHub repo for the full PPOBDA specification:

$$\begin{aligned}
&p_1 : \forall x.\forall y.\forall z.\, \mathit{Patient.gender}(x, y) \wedge \mathit{Patient.address}(x, z) \to \bot \\
&p_2 : \forall x.\forall y.\forall z.\, \mathit{MedicationStatement.subject}(x, y) \wedge \mathit{link}(y, z) \to \bot \\
&p_3 : \forall x.\forall y.\, \mathit{Patient.id}(x) \wedge \mathit{Patient.generalPractitioner}(x, y) \to \bot
\end{aligned}$$

Policy p_1 blocks any query combining gender and address; p_2 prohibits linking medication statements to patient identity; p_3 hides the pairing of patient ID and practitioner assigned to specific patient. A Limited Data Set (LDS) consists of health information with certain identifiers removed, reducing the chances of identifying an individual.Embedding p_1 into mapping M_1 satisfies a LDS by only returning patient gender when `location_id IS NULL`, ensuring address is never exposed alongside gender.

De-identified Data. HIPAA's de-identification methods (Expert Determination, Safe Harbor) rely on transformations (e.g., truncating ZIP codes) that remove the risk of re-identification [21]. Instead, the only effect of the additional negated atoms introduced in mappings through PPOBDA policies is to filter out entire tuples from the result (when they violate a policy), but these atoms are not able to induce any transformation on the result, in particular to apply any of the available de-identification functions (e.g., obfuscation, truncation, anonymization, or generalization). Therefore, denials are not suited to perform de-identification in the PPOBDA framework.

Right to Erasure (RTE). The *Right to Erasure* (RTE), defined in Article 17 of the GDPR [22], allows individuals to request deletion of their personal data under certain conditions. This is essential for protecting user privacy, especially in data access systems like OBDA. We consider two approaches to supporting RTE in the OBDA setting:

(1) RTE in PPOBDA. While OBDA systems typically lack control over the underlying data sources, PPOBDA can simulate erasure by ensuring requested data is excluded from query results, even if not physically deleted. Upon a user's RTE request, appropriate denial policies can be added to censor access to the relevant data. For instance, the following policies can hide sensitive patient data:

$$\begin{aligned}
&p_4 : \forall x.\forall y.\forall z.\, \mathit{Encounter.location}(x, y) \wedge \mathit{Location.name}(y, z) \to \bot \\
&p_5 : \forall x.\forall y.\forall z.\, \mathit{Procedure.code}(x, y) \wedge \mathit{Procedure.performedDateTime}(x, z) \to \bot \\
&p_6 : \forall x.\forall y.\, \mathit{Patient.id}(x) \wedge \mathit{Observation.subject}(x, y) \to \bot
\end{aligned}$$

Policies p_4–p_6 respectively block results revealing encounter locations, procedure timestamps, or observation subjects for erased patients. Though the underlying data sources records remain intact, this method ensures data is functionally inaccessible.

(2) RTE via Ontology-Based Updates. For full RTE compliance, ontology-based updates can be used to translate high-level deletion requests into source-level deletions [23]. In this approach, deletions specified at the ontology level are compiled into the minimal necessary changes to the source database. However, such updates may introduce side-effects. For example, deleting a patient's marital status might also remove their name and patient status if both are stored in the same row of a source table, due to shared mappings. This highlights the complexity of achieving RTE via physical deletion.

Right to Rectification (RTR). RTR, as defined in Article 16 of the GDPR, grants individuals the right to request corrections to their personal data if it is inaccurate or incomplete [24]. This right ensures that data subjects can maintain the accuracy and integrity of their information, preventing incorrect or outdated records from being used. Similar to RTE, PPOBDA cannot modify source values or adjust query answers to reflect corrections. Instead, ontology-based updates [23] can translate rectification requests into source updates, but may produce side-effects (e.g., changing eligibility facts tied to corrected birth dates). Hence, in general, RTR compliance is managed similarly to RTE compliance.

Experimental Results. We evaluated PPOBDA on MIMIC-III data (via OMOP CDM and FHIR) using the six policies p_1–p_6 (see above) and 17 queries. For each query and policy, we measured average execution time (over three runs) and result counts. Our key observations are: *(i)* Each policy successfully suppresses results for at least one query, demonstrating precise enforcement of privacy constraints. *(ii)* Queries unaffected by a given policy return the same result count and exhibit comparable execution times to the baseline. *(iii)* Embedding privacy policies introduces negligible overhead, indicating that policy-induced filtering does not impede performance.

4. Conclusions and Future Work

We showed that PPOBDA can enforce many HIPAA/GDPR requirements, such as, hiding PHI combinations and blocking "forgotten" data from query results while remaining efficient. At the same time, denial assertions as policies cannot perform true de-identification (obfuscation/truncation) or delete/rectify data at the source. To bridge these gaps, we are extending PPOBDA so that mappings can apply simple transformations (e.g., value obfuscation) and invoking ontology-based updates to propagate erasure/rectification requests down to the database. These enhancements aim to achieve full regulatory compliance without sacrificing query performance.

Acknowledgments

This research has been partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, by the HEU project CyclOps (GA n. 101135513), by the Province of Bolzano and FWF through project OnTeGra (DOI 10.55776/PIN8884924). by the Province of Bolzano and EU through projects ERDF-FESR 1078 CRIMA, and ERDF-FESR 1047 AI-Lab, by MUR through the PRIN project 2022XERWK9 S-PIC4CHU, and by the EU and MUR through the PNRR project PE0000013-FAIR.

Declaration on Generative AI

During the preparation of this work, the authors used Chat-GPT-4o in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, J. on Data Semantics 10 (2008) 133–173.

[2] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, in: Proc. of the 27th Int. Joint Conf. on Artificial Intelligence (IJCAI), IJCAI Org., 2018, pp. 5511–5519. doi:10.24963/ijcai.2018/777.

[3] P. A. Bonatti, L. Sauro, A confidentiality model for ontologies, in: Proc. of the 12th Int. Semantic Web Conf. (ISWC), volume 8218 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 17–32. doi:10.1007/978-3-642-41335-3_2.

[4] B. C. Grau, E. Kharlamov, E. V. Kostylev, D. Zheleznyakov, Controlled query evaluation for Datalog and OWL 2 Profile ontologies, in: Proc. of the 24th Int. Joint Conf. on Artificial Intelligence (IJCAI), AAAI Press, 2015, pp. 2883–2889.

[5] D. Lembo, R. Rosati, D. F. Savo, Revisiting controlled query evaluation in description logics, in: Proc. of the 28th Int. Joint Conf. on Artificial Intelligence (IJCAI), ijcai.org, 2019, pp. 1786–1792.

[6] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, 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. doi:10.1007/978-3-030-62419-4_8.

[7] W. Moore, S. Frye, Review of HIPAA, Part 1: History, protected health information, and privacy and security rules, J. of Nuclear Medicine Technology 47 (2019) 269–272.

[8] P. Voigt, A. von dem Bussche, The EU General Data Protection Regulation (GDPR) – A Practical Guide, Springer, 2017.

[9] The HIPAA Privacy Rule, 2025. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/.

[10] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark, MIMIC-III, a freely accessible critical care database, Scientific Data 3 (2016) 1–9.

[11] OMOP Common Data Model, 2025. URL: https://ohdsi.github.io/CommonDataModel.

[12] HL7 FHIR Release 5, 2023. URL: http://www.hl7.org/fhir/structuredefinition.html.

[13] FHIR Ontology, 2025. URL: http://build.fhir.org/fhir.ttl.

[14] G. Xiao, E. R. Pfaff, E. Prud'hommeaux, D. Booth, D. K. Sharma, N. Huo, Y. Yu, N. Zong, K. J. Ruddy, C. G. Chute, G. Jiang, FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common Data Model, J. of Biomedical Informatics 134 (2022) 104201.

[15] Mapping the MIMIC-III database to the OMOP schema, 2018. URL: https://github.com/MIT-LCP/mimic-omop.

[16] I. Reinecke, M. Zoch, C. Reich, M. Sedlmayr, F. Bathelt, The usage of OHDSI OMOP – A scoping review, in: German Medical Data Sciences 2021: Digital Medicine: Recognize – Understand – Heal, IOS Press, 2021, pp. 95–103. doi:10.3233/SHTI210546.

[17] Syntactic Locality Module Extractor, 2020. URL: https://owlcs.github.io/owlapi/apidocs_5/uk/ac/manchester/cs/owlapi/modularity/SyntacticLocalityModuleExtractor.html.

[18] Ontop, 2025. URL: https://github.com/ontop/ontop/releases.

[19] D. Baura, D. Calvanese, L. Marconi, Implementing controlled query evaluation in OBDA, in: Proc. of the Joint Ontology Workshops Episode 10: The Tukker Zomer of Ontology (JOWO), volume 3882 of *CEUR-WS.org*, CEUR Workshop Proceedings, 2024. URL: https://ceur-ws.org/Vol-3882/st4dm-1.pdf.

[20] Other requirements relating to uses and disclosures of protected health information, 2013. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514.

[21] De-identification of Protected Health Information, 2025. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/.

[22] Art. 17 GDPR Right to erasure, 2025. URL: https://gdpr-info.eu/art-17-gdpr/.

[23] R. E. Wandji, D. Calvanese, Ontology-based update in virtual knowledge graphs via schema mapping recovery, in: Proc. of the 8th Int. Joint Conf. on Rules and Reasoning (RuleML+RR), volume 15183 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 59–74.

[24] Art. 16 GDPR Right to rectification, 2025. URL: https://gdpr-info.eu/art-16-gdpr/.