Privacy-Preserving Ontology Publishing: The Case of Quantified ABoxes w.r.t. a Static Cycle-Restricted $\mathcal{EL}$ TBox

Franz Baader[1] Francesco Kriegel[1] Patrick Koopmann[1]
Adrian Nuradiansyah[1] Rafael Peñaloza[2]

[1]Technische Universität Dresden & [2]University of Milano-Bicocca

34th International Workshop on Description Logics at Bratislava, Slovakia

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Dataset (qABox) Rules ($\mathcal{EL}$ TBox)

Privacy policy
(a set of $\mathcal{EL}$ concepts)

Quantified ABox: $\exists X.\mathcal{A}$

(*ABox with atomic assertions, individuals, and existentially quantified variables*)

$\exists\{x\}.\{\mathit{relative}(\mathit{BEN}, x), \mathit{Actor}(x), \mathit{spouse}(x, \mathit{JERRY}), \mathit{Comedian}(\mathit{JERRY})\}$

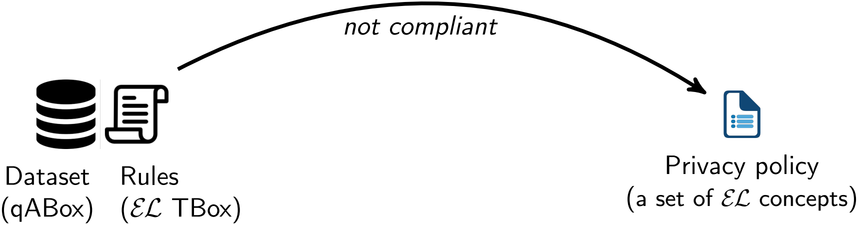
$\mathcal{EL}$ **TBox** $\mathcal{T}$**:**

$\{\mathit{Comedian} \sqsubseteq \mathit{Actor}\}$

Policy $\mathcal{P}$**:**

$\{\exists \mathit{relative}.(\mathit{Actor} \sqcap \exists \mathit{spouse}.\mathit{Actor})\}$

Policy-Compliance w.r.t. Static $\mathcal{EL}$ TBoxes



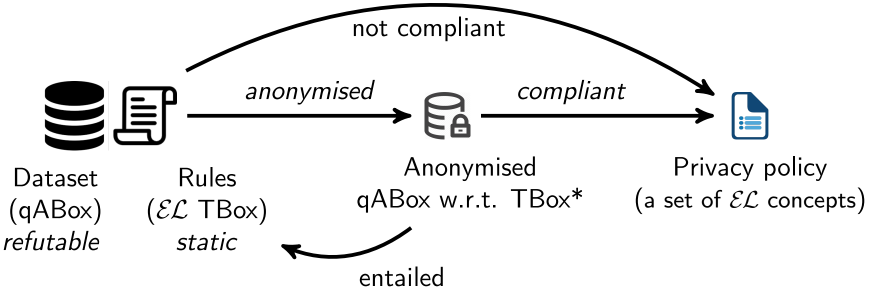
Quantified ABox: $\exists X.\mathcal{A}$
(*ABox with atomic assertions, individuals, and existentially quantified variables*)
$\exists\{x\}.\{\mathit{relative}(\mathit{BEN}, x), \mathit{Actor}(x), \mathit{spouse}(x, \mathit{JERRY}), \mathit{Comedian}(\mathit{JERRY})\}$

$\mathcal{EL}$ **TBox** $\mathcal{T}$**:**
$\{\mathit{Comedian} \sqsubseteq \mathit{Actor}\}$

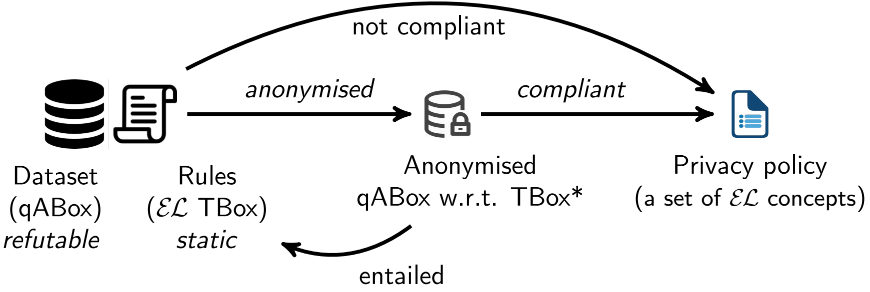
Policy $\mathcal{P}$**:**
$\{\exists \mathit{relative}.(\mathit{Actor} \sqcap \exists \mathit{spouse}.\mathit{Actor})\}$

BEN is an instance of the policy w.r.t. $\exists X.\mathcal{A}$ *and* $\mathcal{T}$ $\Rightarrow$ ***not compliant!***

Optimal Compliant Anonymizations



Optimal Compliant Anonymizations



- *being **optimal**: not strictly entailed by the other compliant anonymizations
- **Cycle-restricted TBoxes** are considered:
  no $C \sqsubseteq_{\mathcal{T}} \exists w.C$ for each concept C and each non-empty word $w \in \Sigma_{\mathsf{R}}*$
- **Canonical compliant anonymizations** $\exists Y.\mathcal{B}$: a class of anonymizations covering all optimal compliant anonymizations

How to Compute A Canonical Compliant Anonymization

1. **Saturate the qABox** $(\exists X.\mathcal{A} \Rightarrow \mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A}))$
   Saturation always terminates for cycle-restricted TBoxes

How to Compute A Canonical Compliant Anonymization

1. **Saturate the qABox** ($\exists X.\mathcal{A} \Rightarrow \mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A})$)
   Saturation always terminates for cycle-restricted TBoxes

2. **Create copies** $y_{u,\mathcal{K}}$ **of each object** u of the $\mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A})$ s.t.
   - each $y_{u,\mathcal{K}}$ is a variable in $\exists Y.\mathcal{B}$
   - $\mathcal{K} \subseteq \mathsf{Atoms}(\mathcal{P}, \mathcal{T})$ is a **repair type** that specifies which instance relationships that want to be removed by $\exists Y.\mathcal{B}$ ($C \in \mathcal{K}$ *implies* $(\exists X.\mathcal{A})^{\mathcal{T}} \models C(u)$)

How to Compute A Canonical Compliant Anonymization

1. **Saturate the qABox** ($\exists X.\mathcal{A} \Rightarrow \mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A})$)
   Saturation always terminates for cycle-restricted TBoxes

2. **Create copies** $y_{u,\mathcal{K}}$ **of each object** u of the $\mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A})$ s.t.
   - each $y_{u,\mathcal{K}}$ is a variable in $\exists Y.\mathcal{B}$
   - $\mathcal{K} \subseteq \mathsf{Atoms}(\mathcal{P}, \mathcal{T})$ is a **repair type** that specifies which instance relationships that want to be removed by $\exists Y.\mathcal{B}$ ($C \in \mathcal{K}$ *implies* $(\exists X.\mathcal{A})^{\mathcal{T}} \models C(u)$)

3. **Define a compliance seed function (csf)** s that assigns each individual to a repair type s.t.
   - for each $P \in \mathcal{P}$ with $\mathsf{sat}^{\mathcal{T}}(\exists X.\mathcal{A}) \models P(a)$, the repair type $s(a)$ contains an atom subsuming P
   - s is further used to **induce** $\exists Y.\mathcal{B}$, e.g., create assertions for $\exists Y.\mathcal{B}$ s.t. $C \in \mathcal{K}$ implies $\exists Y.\mathcal{B} \not\models C(y_{u,\mathcal{K}})$

Complexity of the Computation and Optimizations

Theorem (ISWC '20, CADE '21)

There is an algorithm to compute the set of all optimal compliant anonymizations of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{P}$ and $\mathcal{T}$ that

- is deterministic and runs in exponential time, and
  (*the number of seed functions and variables is exponential*)
- has access to an NP-oracle
  (*remove the non-optimal anonymizations*)

Complexity of the Computation and Optimizations

Theorem (ISWC '20, CADE '21)

There is an algorithm to compute the set of all optimal compliant anonymizations of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{P}$ and $\mathcal{T}$ that

- is deterministic and runs in exponential time, and
- has access to an NP-oracle

Can we improve the complexity?

Smaller/Optimized Compliant Anonymizations

- The number of variables in canonical anonymizations is always exponential
- Start with a csf, and then **only introduce necessary variables** stepwise

Complexity of the Computation and Optimizations

Theorem (ISWC '20, CADE '21)

There is an algorithm to compute the set of all optimal compliant anonymizations of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{P}$ and $\mathcal{T}$ that

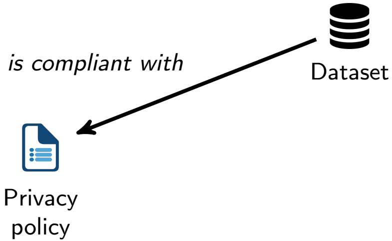
- is deterministic and runs in exponential time, and
- has access to an NP-oracle

Theorem (CADE '21)

Each optimized compliant anonymization induced by a csf s is **equivalent** to the corresponding canonical compliant anonymization induced by s

Implementation: `https://github.com/de-tu-dresden-inf-lat/abox-repairs-wrt-static-tbox`.

Safety for Singleton Policies and Without TBoxes

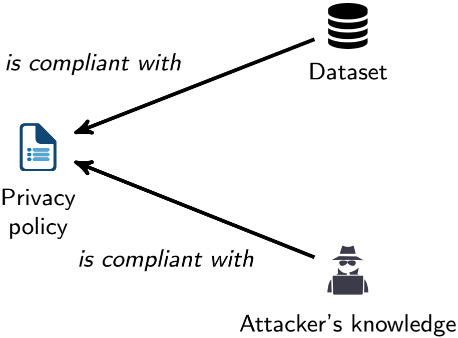


Dataset $\exists X.\mathcal{A}$**:**
$\exists \emptyset.\{father(BEN, JERRY), Comedian(JERRY)\}$

Policy $\mathcal{P}$**:**
$\{Comedian \sqcap \exists father.Comedian\}$

No instance of the policy concept w.r.t. the dataset

Safety for Singleton Policies and Without TBoxes



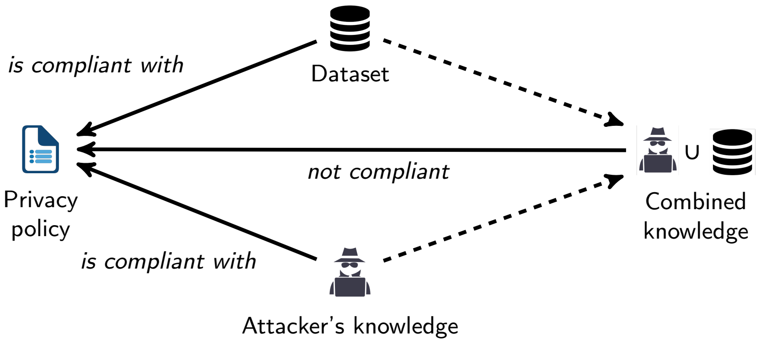
Dataset $\exists X.\mathcal{A}$**:**
$\exists \emptyset.\{father(BEN, JERRY), Comedian(JERRY)\}$

Policy $\mathcal{P}$**:**
$\{Comedian \sqcap \exists father.Comedian\}$

Attacker $\exists Y.\mathcal{B}$ **knows:**
$\exists \emptyset.\{Comedian(BEN)\}$

No instance of the policy concept w.r.t. the attacker's knowledge

Safety for Singleton Policies and Without TBoxes



Dataset $\exists X.\mathcal{A}$:
$\exists\emptyset.\{father(BEN, JERRY), Comedian(JERRY)\}$

Policy $\mathcal{P}$:
$\{Comedian \sqcap \exists father.Comedian\}$

Attacker $\exists Y.\mathcal{B}$ knows:
$\exists\emptyset.\{Comedian(BEN)\}$

*BEN is an instance of the policy concept w.r.t. the dataset and the attacker's knowledge ⇒ the dataset is **compliant with**, but **not safe** for the policy !*

Characterization of Safety for Singleton Policies

Characterization of Safety (SAC '21)

$\exists X.\mathcal{A}$ is safe for $\{P\}$ iff for each individual a,

- if $A \in \text{Atoms}(\{P\})$, then $A(a) \notin \mathcal{A}$
- if $r(a, u) \in \mathcal{A}$ and $\exists r.D \in \text{Atoms}(\{P\})$, then there is no **partial homomorphism** from D to $\exists X.\mathcal{A}$ at u.

Characterization of Safety for Singleton Policies

Characterization of Safety (SAC '21)

$\exists X.\mathcal{A}$ is safe for $\{P\}$ iff for each individual a,

- if $A \in \text{Atoms}(\{P\})$, then $A(a) \notin \mathcal{A}$
- if $r(a, u) \in \mathcal{A}$ and $\exists r.D \in \text{Atoms}(\{P\})$, then there is no **partial homomorphism** from D to $\exists X.\mathcal{A}$ at u.

Partial Homomorphism

It **is like a homomorphism,** but the mapping only maps nodes of the syntax tree of D that are between **the root and a "cut".**

Characterization of Safety for Singleton Policies

Characterization of Safety (SAC '21)

$\exists X.\mathcal{A}$ is safe for $\{P\}$ iff for each individual a,

- if $A \in \text{Atoms}(\{P\})$, then $A(a) \notin \mathcal{A}$
- if $r(a, u) \in \mathcal{A}$ and $\exists r.D \in \text{Atoms}(\{P\})$, then there is no **partial homomorphism** from D to $\exists X.\mathcal{A}$ at u.

Partial Homomorphism

It **is like a homomorphism,** but the mapping only maps nodes of the syntax tree of D that are between **the root and a "cut".**

Complexity of Safety for Singleton Policies (SAC'21)

Safety of a qABox for singleton $\mathcal{EL}$ policies is in P

Computing Canonical Safe Anonymizations

Canonical safe anonymizations $\exists Z.\mathcal{C}$ of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ covers each $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$.

Computing Canonical Safe Anonymizations

Canonical safe anonymizations $\exists Z.\mathcal{C}$ of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ covers each $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$.

Analogous to the computation of canonical compliant anonymizations, but:

- no saturation, no seed function
- each variable is of the form $y_{u,\mathcal{K}}$, but $\mathcal{K}$ is not a repair type, it's just a subset of $\mathcal{EL}$ atoms.
- there is an additional mechanism to avoid partial homomorphisms

Computing Canonical Safe Anonymizations

Canonical safe anonymizations $\exists Z.\mathcal{C}$ of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ covers each $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$.

Analogous to the computation of canonical compliant anonymizations, but:

- no saturation, no seed function
- each variable is of the form $y_{u,\mathcal{K}}$, but $\mathcal{K}$ is not a repair type, it's just a subset of $\mathcal{EL}$ atoms.
- there is an additional mechanism to avoid partial homomorphisms

Results of the Computation (SAC '21)

There is **only one optimal safe anonymization** of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ and computing this can be done in **exponential time**

Smaller Optimal Safe Anonymizations

Using a similar idea as the computation of optimized compliant anonymizations

Theorem (NEW!)

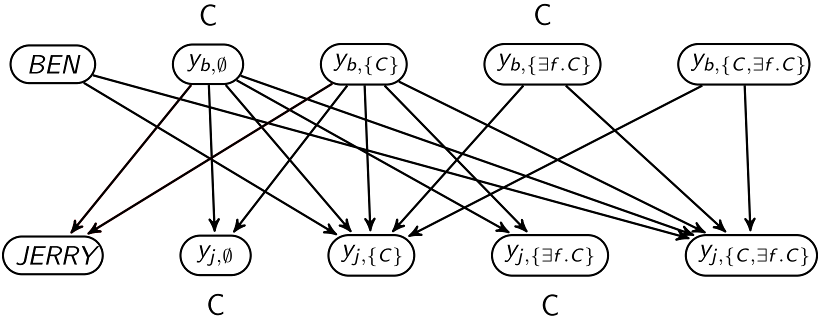
The optimized safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ is **equivalent** to the canonical safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$

Smaller Optimal Safe Anonymizations

Using a similar idea as the computation of optimized compliant anonymizations

Theorem (NEW!)

The optimized safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ is **equivalent** to the canonical safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$

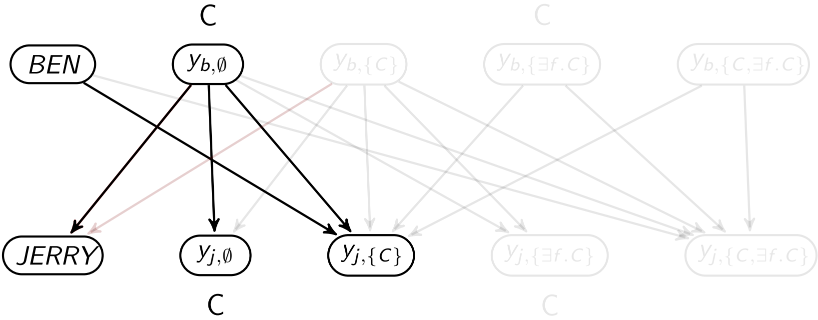


Smaller Optimal Safe Anonymizations

Using a similar idea as the computation of optimized compliant anonymizations

Theorem (NEW!)

The optimized safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$ is **equivalent** to the canonical safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P\}$



What If TBoxes are Taken into Account

Complexity of the Problem

Expressing general policies by singleton policies using TBoxes

- Safety problem for singleton policies is **at least as hard as** safety for general policies when TBoxes are considered
- The safety problem for general policies w.r.t. static $\mathcal{EL}$ TBoxes is in **coNP**.

Remarks and Future Work

Our work **reviewed results** from

- Baader, Kriegel, Nuradiansyah, Peñaloza, *Computing Compliant Anonymisations of Quantified ABoxes w.r.t. $\mathcal{EL}$ Policies*, ISWC '20
- Baader, Kriegel, Nuradiansyah, Peñaloza, *Safety of Quantified ABoxes w.r.t. Singleton $\mathcal{EL}$ Policies*, SAC '21
- Baader, Koopmann, Kriegel, Nuradiansyah, *Computing Optimal Repairs of Quantified ABoxes w.r.t. Static $\mathcal{EL}$ TBoxes*, CADE '21

and **presented new results** in the topic of safety with and without TBoxes.

Remarks and Future Work

Our work **reviewed results** from

- Baader, Kriegel, Nuradiansyah, Peñaloza, *Computing Compliant Anonymisations of Quantified ABoxes w.r.t. $\mathcal{EL}$ Policies*, ISWC '20
- Baader, Kriegel, Nuradiansyah, Peñaloza, *Safety of Quantified ABoxes w.r.t. Singleton $\mathcal{EL}$ Policies*, SAC '21
- Baader, Koopmann, Kriegel, Nuradiansyah, *Computing Optimal Repairs of Quantified ABoxes w.r.t. Static $\mathcal{EL}$ TBoxes*, CADE '21

and **presented new results** in the topic of safety with and without TBoxes.

Possible Future Work:

- Safety w.r.t. general policies and/or (cycle-restricted) TBoxes
- Safety w.r.t. a finite set of concept assertions $\{P_1(a_1), \ldots, P_n(a_n)\}$