Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Temporalized $\mathcal{EL}$ Ontologies for Accessing Temporal Data: Complexity of Atomic Queries

Víctor Gutiérrez-Basulto[1] and **Jean Christoph Jung**[1] and **Roman Kontchakov**[2]
[1] Department of Computer Science, Universität Bremen, Germany
[2] Department of Computer Science and Information Systems, Birkbeck, University of London, UK

Abstract

We study access to temporal data with $\mathcal{TEL}$, a temporal extension of the tractable description logic $\mathcal{EL}$. Our aim is to establish a clear computational complexity landscape for the atomic query answering problem, in terms of both data and combined complexity. Atomic queries in full $\mathcal{TEL}$ turn out to be undecidable even in data complexity. Motivated by the negative result, we identify well-behaved yet expressive fragments of $\mathcal{TEL}$. Our main contributions are a semantic and sufficient syntactic conditions for decidability and three orthogonal tractable fragments, which are based on restricted use of rigid roles, temporal operators, and novel acyclicity conditions on the ontologies.

1 Introduction

In recent years, the use of ontologies to enrich plain data with a semantic layer has become one of the outstanding applications of description logic (DLs) technologies in the Semantic Web. The ontology-based data access (OBDA) setting provides information systems with various advantages, e.g., a friendlier vocabulary for accessing heterogenous data is given by the ontology, and means of querying potentially incomplete data are provided by taking account of the implicit knowledge derived from the data and the ontology. Due to the increasing need to account for the temporal dimension of data available on the Web [Roth and Tan, 2013; Dong and Tan, 2015], the DL community has recently investigated extensions of the OBDA paradigm for temporal data. The initial efforts concentrated on temporal query languages with atemporal ontologies [Gutiérrez-Basulto and Klarman, 2012; Klarman and Meyer, 2014; Baader, Borgwardt, and Lippmann, 2015; Borgwardt, Lippmann, and Thost, 2015; Borgwardt and Thost, 2015]. On the other hand, temporal ontology languages can enhance conceptual modelling with temporal aspects [Artale et al., 2015], which are required, e.g., in applications managing data from sensor networks. In this line, the research has focused on temporal extensions of *DL-Lite* that support rewritability of temporal queries into the monadic second-order logic with order or into two-sorted first-order logic with $<$ and $+$ [Artale et al., 2013b; 2015]. Since standard relational database management systems have such built-in predicates, they can in principle evaluate the FO$(<,+)$-rewritings. However, no temporal extensions of other classical DLs have been investigated yet in the context of OBDA, which is partly because of the intractability and often even undecidability of the standard reasoning tasks (e.g., subsumption) [Artale et al., 2007; Gutiérrez-Basulto, Jung, and Lutz, 2012; Gutiérrez-Basulto, Jung, and Schneider, 2014]. On the other hand, temporal data has also been studied in classical database theory [Chomicki and Toman, 2005]. In their seminal paper, Chomicki and Imielinski [1988] identified $\textsc{Datalog}_{1S}$ as a decidable extension of $\textsc{Datalog}$ with one successor function. Here we make the first (to the best of our knowledge) attempt to link temporal OBDA with temporal deductive databases [Chomicki, 1990; Baudinet, Chomicki, and Wolper, 1993].

In this paper, we study $\mathcal{TEL}$, a temporal extension of $\mathcal{EL}$ [Baader, Brandt, and Lutz, 2005]. The underlying DL component, $\mathcal{EL}$, underpins the OWL 2 EL profile of OWL 2 and the medical ontology SNOMED CT, which provides the vocabulary for electronic health records (EHRs). Indeed, applications managing EHRs must be able to provide information, e.g., on when and for how long some drug has been prescribed to a patient, so that drugs that interact adversely are not prescribed at the same time. Clinical trials [Shankar et al., 2008; O'Connor et al., 2009] also require a unified conceptual model for specifying temporal constraints of protocol entities such as *'a viable participant should have had a vaccination with live virus 5 days ago'* or '*blood tests of a patient should be run every 3 days*'. These statements can be encoded in $\mathcal{TEL}$:

$$\mathsf{Patient} \sqcap \bigcirc_P^5 \exists \mathsf{vaccinated.LiveVirus} \sqsubseteq \mathsf{ViableParticip}, \quad (1)$$

$$\mathsf{Patient} \sqcap \bigcirc_P^3 \mathsf{ReqBloodTest} \sqsubseteq \mathsf{ReqBloodTest}. \quad (2)$$

Our main objective is to establish the limits of decidability and tractability of the query answering problem over $\mathcal{TEL}$ ontologies, in terms of both data and combined complexity. In order to set the foundations, we focus on *temporal atomic queries*. On the one hand, an atomic query $\mathsf{ViableParticip}(x,t)$ with the temporal concept inclusion (1) effectively encodes a tree-shaped temporal conjunctive query. On the other hand, using (1) to extend the vocabulary with a concept $\mathsf{ViableParticip}$ is closer to the spirit of the OBDA paradigm than repeating the same conjunction in all similar user queries. Moreover, a recurrent pattern $\mathsf{ReqBloodTest}$ is expressible as an atomic query

$\mathsf{ReqBloodTest}(x,t)$ with the temporal concept inclusion (2) but not expressible as a query without temporal concept inclusions like (2). As we shall see, even for atomic queries rather surprising (and challenging) results are obtained.

Our main contributions are complexity bounds, algorithms and rewritability into DATALOG$_{1S}$ for atomic query answering in fragments of $\mathcal{TEL}$. Since query answering over unrestricted $\mathcal{TEL}$ turns out to be undecidable (in data complexity), we investigate its fragments to attain decidability and tractability. First, for $\mathcal{TEL}^{\bigcirc}$, which allows only the 'next-' $\bigcirc_F$ and 'previous-time' $\bigcirc_P$ operators, we identify *ultimate periodicity* as a natural semantic condition ensuring decidability, more precisely, PSPACE data complexity (the question of decidability of the full $\mathcal{TEL}^{\bigcirc}$ is left open for future work). Then, we identify a number of fragments with better computational properties. (*i*) For the fragment of $\mathcal{TEL}^{\bigcirc}$ without rigid (not changing over time) roles on the right-hand side of concept inclusions, we construct a polynomial rewriting into DATALOG$_{1S}$, and so, establish PSPACE-completeness for data complexity. This fragment contains all $\mathcal{EL}$ ontologies as well as both (1) and (2). (*ii*) Over *temporally acyclic* $\mathcal{TEL}^{\bigcirc}$-ontologies (with rigid roles and concepts), query answering is PTIME-complete in both data and combined complexity. This tractable fragment contains (1) and fully captures all atemporal $\mathcal{EL}$ ontologies and may prove particularly useful in applications; it, however, does not contain (2). (*iii*) Query answering over *DL-acyclic* $\mathcal{TEL}^{\bigcirc}$ ontologies is NC^1-complete for data complexity (in principle, highly parallelizable). This fragment contains many acyclic $\mathcal{EL}$ ontologies as well as both (1) and (2) (note that large parts of SNOMED CT are in fact acyclic). We remark that our two novel acyclicity conditions (each constraining only one dimension) are inspired by the 'traditional' notion of acyclicity in (temporal extensions of) DLs [Haase and Lutz, 2008; Gutiérrez-Basulto, Jung, and Schneider, 2015]. Finally, (*iv*) we show that the language with only $\Diamond_P$ and $\Diamond_F$ (sometime in the past/future) on the left-hand side of concept inclusions also enjoys PTIME query answering.

Omitted proofs can be found at tinyurl.com/TempEL16.

2 Preliminaries

We begin by introducing $\mathcal{TEL}$, a temporal extension of the classical DL $\mathcal{EL}$. Let $\mathsf{N_C}$, $\mathsf{N_R}$, $\mathsf{N_I}$ be countably infinite sets of *concept*, *role* and *individual names*, respectively. We assume that $\mathsf{N_R}$ is partitioned into two infinite sets, $\mathsf{N_R^{rig}}$ and $\mathsf{N_R^{loc}}$, of *rigid* and *local role names*, respectively. $\mathcal{TEL}$ *concepts* are defined by the following grammar:

$$C, D \ ::= \ A \ \mid \ C \sqcap D \ \mid \ \exists r.C \ \mid \ \bigcirc_* C \ \mid \ \Diamond_* C,$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $* \in \{F, P\}$. A $\mathcal{TEL}$*-TBox* (*ontology*) $\mathcal{T}$ is a finite set of *concept inclusions* (*CIs*) $C \sqsubseteq D$ and *concept definitions* (*CDs*) $C \equiv D$ for $\mathcal{TEL}$ concepts C, D. Data is given in terms of *temporal ABoxes* $\mathcal{A}$, which are finite sets of assertions of the form $A(a,n)$ and $r(a,b,n)$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, $a, b \in \mathsf{N_I}$, and $n \in \mathbb{Z}$. We denote by $\mathsf{ind}(\mathcal{A})$ the sets of individual names occurring in $\mathcal{A}$, and by $\mathsf{tem}(\mathcal{A})$ the set $\{n \in \mathbb{Z} \mid \min \mathcal{A} \leq n \leq \max \mathcal{A}\}$, where $\min \mathcal{A}$ and $\max \mathcal{A}$ are, respectively, the minimal and maximal time points in $\mathcal{A}$. The size, $|\mathcal{T}|$ and $|\mathcal{A}|$, of $\mathcal{T}$ and $\mathcal{A}$ is the number of symbols required to write $\mathcal{T}$ and $\mathcal{A}$, respectively, with time points $n \in \mathbb{Z}$ encoded in *unary*. A *temporal knowledge base* (*KB*) $\mathcal{K}$ is a pair $(\mathcal{T}, \mathcal{A})$.

An *interpretation* $\mathfrak{I}$ is a structure $(\Delta^{\mathfrak{I}}, (\mathcal{I}_n)_{n \in \mathbb{Z}})$, where each $\mathcal{I}_n$ is a classical DL interpretation with domain $\Delta^{\mathfrak{I}}$: we have $A^{\mathcal{I}_n} \subseteq \Delta^{\mathfrak{I}}$ and $r^{\mathcal{I}_n} \subseteq \Delta^{\mathfrak{I}} \times \Delta^{\mathfrak{I}}$. Rigid roles $r \in \mathsf{N_R^{rig}}$ do not change their interpretation in time: $r^{\mathcal{I}_n} = r^{\mathcal{I}_0}$ for all $n \in \mathbb{Z}$. We usually write $A^{\mathfrak{I},n}$ and $r^{\mathfrak{I},n}$ instead of $A^{\mathcal{I}_n}$ and $r^{\mathcal{I}_n}$, respectively, and the mapping $\cdot^{\mathfrak{I},n}$ is extended to complex $\mathcal{TEL}$-concepts as follows:

$$(C \sqcap D)^{\mathfrak{I},n} = C^{\mathfrak{I},n} \cap D^{\mathfrak{I},n},$$
$$(\exists r.C)^{\mathfrak{I},n} = \big\{ d \mid \text{ there is } e \in C^{\mathfrak{I},n} \text{ with } (d,e) \in r^{\mathfrak{I},n} \big\},$$
$$(\bigcirc_* C)^{\mathfrak{I},n} = C^{\mathfrak{I}, n \,\mathsf{op}_*\, 1},$$
$$(\Diamond_* C)^{\mathfrak{I},n} = \big\{ d \mid d \in C^{\mathfrak{I}, n \,\mathsf{op}_*\, k} \text{ for some } k > 0 \big\},$$

where op_* stands for $-$ if $* = P$ and for $+$ if $* = F$. We use strict $\Diamond_*$ ($k > 0$) but our results do not depend on the choice.

An interpretation $\mathfrak{I}$ is said to be a *model* of $C \sqsubseteq D$, written $\mathfrak{I} \models C \sqsubseteq D$, if $C^{\mathfrak{I},n} \subseteq D^{\mathfrak{I},n}$, for all $n \in \mathbb{Z}$; and a model of $C \equiv D$ if $C^{\mathfrak{I},n} = D^{\mathfrak{I},n}$, for all $n \in \mathbb{Z}$. We call $\mathfrak{I}$ a *model of a TBox* $\mathcal{T}$, written $\mathfrak{I} \models \mathcal{T}$, if $\mathfrak{I} \models \alpha$ for all $\alpha \in \mathcal{T}$. Note that TBoxes are interpreted *globally* in the sense that all CIs and CDs must be satisfied at every time point. A concept D *subsumes* a concept C with respect to $\mathcal{T}$, written $\mathcal{T} \models C \sqsubseteq D$, if $\mathfrak{I} \models C \sqsubseteq D$ for all models $\mathfrak{I}$ of $\mathcal{T}$.

For ABoxes $\mathcal{A}$ we adopt the *standard name assumption*: $a^{\mathfrak{I},n} = a$ for all $a \in \mathsf{ind}(\mathcal{A})$, $n \in \mathbb{Z}$ (and thus $\mathsf{ind}(\mathcal{A}) \subseteq \Delta^{\mathfrak{I}}$). The relation $\models$ is extended to ABoxes by taking $\mathfrak{I} \models A(a,n)$ iff $a \in A^{\mathfrak{I},n}$ and $\mathfrak{I} \models r(a,b,n)$ iff $(a,b) \in r^{\mathfrak{I},n}$; $\mathfrak{I}$ is a model of $\mathcal{A}$ iff $\mathfrak{I} \models \alpha$ for all $\alpha \in \mathcal{A}$. An interpretation $\mathfrak{I}$ is a *model* of a KB $(\mathcal{T}, \mathcal{A})$, written $\mathfrak{I} \models (\mathcal{T}, \mathcal{A})$, iff $\mathfrak{I} \models \mathcal{T}$ and $\mathfrak{I} \models \mathcal{A}$. Finally, $\mathcal{K} \models A(a,n)$ if $\mathfrak{I} \models A(a,n)$ for every model $\mathfrak{I}$ of $\mathcal{K}$.

As the query language, we consider *temporal atomic queries* (*TAQs*) of the form $A(x,t)$ with $A \in \mathsf{N_C}$, x an *individual variable* and t a *temporal variable*. Given $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a *certain answer to* $A(x,t)$ *over* $\mathcal{K}$ is a pair $(a,n) \in \mathsf{ind}(\mathcal{A}) \times \mathsf{tem}(\mathcal{A})$ with $\mathcal{K} \models A(a,n)$. We study the complexity of the query answering problem over temporal knowledge bases:

TAQ answering	
Input:	TBox $\mathcal{T}$, ABox $\mathcal{A}$, TAQ $A(x,t)$, a pair (a,n).
Question:	Is (a,n) a certain answer to $A(x,t)$ over $(\mathcal{T}, \mathcal{A})$?

Our results concern both the combined and data complexity of the problem: for data complexity, the TBox is fixed. As usual, for a complexity class $\mathcal{C}$ and a class $\mathcal{X}$ of TBoxes, we say that *TAQ answering over* $\mathcal{X}$ *is* $\mathcal{C}$*-hard in data complexity* if there is some $\mathcal{T} \in \mathcal{X}$ such that answering TAQs over $\mathcal{T}$ is $\mathcal{C}$-hard. Conversely, *TAQ answering over* $\mathcal{X}$ *is in* $\mathcal{C}$ *in data complexity* if, for all $\mathcal{T} \in \mathcal{X}$, answering TAQs over $\mathcal{T}$ is in $\mathcal{C}$.

As classes $\mathcal{X}$, we will in particular look at full $\mathcal{TEL}$ and its fragments $\mathcal{TEL}^{\Diamond}$ and $\mathcal{TEL}^{\bigcirc}$ in which, respectively, only the temporal operators $\Diamond_*$ and $\bigcirc_*$ are allowed. Note that $\Diamond_*$ on the left-hand side and $\Box_*$ (with the usual semantics) on the right-hand side of CIs can be expressed in $\mathcal{TEL}^{\bigcirc}$, e.g., instead of $\Diamond_P A \sqsubseteq X$ or, equivalently, $A \sqsubseteq \Box_F X$, take $A \sqsubseteq A'$ and $\bigcirc_P A' \sqsubseteq A' \sqcap X$, for a fresh A'. Thus, *rigid concepts*, which do not change their interpretation in time, can be expressed in these two fragments using $\Diamond_P \Diamond_F$ on the left-hand side of CIs.

3 Query Answering in $\mathcal{TEL}$: Undecidability

We first pinpoint different sources of complexity for the query answering problem in $\mathcal{TEL}$ in order to identify computationally well-behaved fragments in the sequel.

We begin by showing that TAQ answering over $\mathcal{TEL}^{\Diamond}$ is undecidable. The known undecidability of subsumption in $\mathcal{TEL}^{\Diamond}$ [Artale et al., 2007] translates only into the combined complexity of TAQ answering. We strengthen the result to obtain undecidability in data complexity by reducing the halting problem for the universal Turing machine. We exploit the crucial observation that disjunction, although not in the syntax, can be simulated with $\Diamond_*$ [Artale et al., 2007].

Theorem 1 *TAQ answering over $\mathcal{TEL}^{\Diamond}$ is undecidable in data complexity.*

The proof can also be adapted to the *non-strict* semantics of $\Diamond_*$ using the *chessboard technique* [Gabbay et al., 2003].

$\mathcal{TEL}^{\bigcirc}$, unlike $\mathcal{TEL}^{\Diamond}$, is not capable of expressing disjunction. Still, its expressiveness makes answering TAQs hard:

Theorem 2 *TAQ answering over $\mathcal{TEL}^{\bigcirc}$ is non-elementary in combined complexity and* PSPACE*-hard in data complexity.*

The proof of PSPACE-hardness is close in spirit to that for DATALOG$_{1S}$ [Chomicki and Imielinski, 1988]; note that the lower bound holds even in the restriction of $\mathcal{TEL}^{\bigcirc}$ without $\exists r.C$ on the right-hand side of CIs. For the non-elementary lower bound, we take inspiration in the construction for the product modal logic LTL$\times$**K** [Gabbay et al., 2003, Theorem 6.34]. Our proof requires a careful implementation of the yardstick technique [Stockmeyer, 1974] with only Horn formulas.

Decidability of TAQ answering in full $\mathcal{TEL}^{\bigcirc}$ is left open as interesting and challenging future work; more insights on the difficulty of the problem are given in Sec. 4. Nevertheless, we show that extending $\mathcal{TEL}^{\bigcirc}$ with certain DL constructs that are harmless for data complexity of atemporal query answering [Krisnadhi and Lutz, 2007] immediately leads to undecidability. Let $\mathcal{TELI}^{\bigcirc}$ and $\mathcal{TELF}^{\bigcirc}$ be the extensions of $\mathcal{TEL}^{\bigcirc}$ with *inverse roles* r^- and *functionality* axioms *func*(r), respectively.[1] For both languages, we reduce the halting problem for the universal Turing machine to prove:

Theorem 3 *TAQ answering over $\mathcal{TELI}^{\bigcirc}$ and $\mathcal{TELF}^{\bigcirc}$ is undecidable in data complexity.*

In the rest of the paper, we study decidability and complexity of TAQ answering in various fragments of $\mathcal{TEL}^{\bigcirc}$ and $\mathcal{TEL}^{\Diamond}$.

4 Foundations of Query Answering in $\mathcal{TEL}^{\bigcirc}$

In this section, we lay the groundwork for the development of algorithms for query answering in fragments of $\mathcal{TEL}^{\bigcirc}$ by introducing canonical quasimodels, which are succinct abstract representations of the *universal model* of the KB, see also [Artale et al., 2013b; 2015]. They can also be viewed as a generalization of the canonical structures used for query answering in pure $\mathcal{EL}$ [Lutz, Toman, and Wolter, 2009].

In the sequel, we assume that $\mathcal{TEL}^{\bigcirc}$-TBoxes are in *normal form*, that is, they consist of CIs of the form

$$A \sqcap A' \sqsubseteq B, \qquad A \sqsubseteq \exists r.B, \qquad X \sqsubseteq A,$$

where A, A' and B are concept names and X is a *basic concept* of the form A, $\bigcirc_* A$, or $\exists r.A$, for a concept name A. Observe that, without loss of generality, $\bigcirc_*$ is restricted to the left-hand side of CIs: for instance, $A \sqsubseteq \bigcirc_F B$ is equivalent to $\bigcirc_P A \sqsubseteq B$. It is routine to show that every $\mathcal{TEL}^{\bigcirc}$-TBox can be transformed into the normal form by introducing fresh concept names; see, e.g., [Baader, Brandt, and Lutz, 2005].

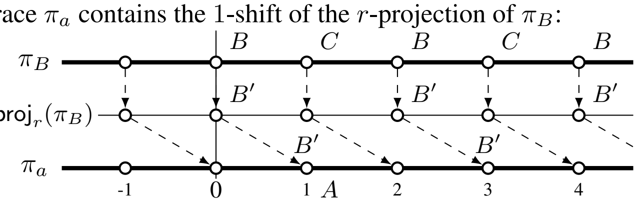
Fix now a KB $(\mathcal{T}, \mathcal{A})$ with a $\mathcal{TEL}^{\bigcirc}$-TBox $\mathcal{T}$ in normal form and let CN be the set of concept names in $(\mathcal{T}, \mathcal{A})$. A map $\pi\colon \mathbb{Z} \to 2^{\mathsf{CN}}$ is a *trace for* $\mathcal{T}$ if it satisfies the following:

(t1) if $A \sqcap A' \sqsubseteq B \in \mathcal{T}$ and $A, A' \in \pi(n)$, then $B \in \pi(n)$;

(t2) if $\bigcirc_* A \sqsubseteq B \in \mathcal{T}$ and $A \in \pi(n)$, then $B \in \pi(n \mathsf{\ op}_* 1)$.

Traces are the building blocks of quasimodels: they represent the temporal evolution of individual domain elements. For example, for $\mathcal{T} = \{\bigcirc_P C \sqsubseteq B, \bigcirc_P B \sqsubseteq C\}$, the map π such that $\pi(i)$ is $\{B\}$ for odd i and $\{C\}$ for even i is a trace for $\mathcal{T}$.

In order to describe interactions of domain elements, we require more notation. Let π be a trace for $\mathcal{T}$. For a rigid role $r \in \mathsf{N}_\mathsf{R}^{\mathsf{rig}}$, the *$r$-projection of* π is a map $\mathsf{proj}_r(\pi)\colon \mathbb{Z} \to 2^{\mathsf{CN}}$ that sends each $i \in \mathbb{Z}$ to $\{A \mid \exists r.B \sqsubseteq A \in \mathcal{T}, B \in \pi(i)\}$; for a local role $r \in \mathsf{N}_\mathsf{R}^{\mathsf{loc}}$, $\mathsf{proj}_r(\pi)$ is defined in the same way on 0 but is $\emptyset$ for all other $i \in \mathbb{Z}$. Given a map $\varrho\colon \mathbb{Z} \to 2^{\mathsf{CN}}$ and $n \in \mathbb{Z}$, we say that π *contains the n-shift of ϱ* and write $\varrho \subseteq^n \pi$ if $\varrho(i - n) \subseteq \pi(i)$, for all $i \in \mathbb{Z}$. For example, let $\mathcal{T} = \{\exists r.B \sqsubseteq B'\}$ with rigid role r. In the picture below, the trace π_a contains the 1-shift of the r-projection of π_B:



If r is local then π_a has to contain B' only at 1 (but not at 3, etc.). We are now fully equipped to define quasimodels.

Let $D = \mathsf{ind}(\mathcal{A}) \cup \mathsf{CN}$ henceforth. A *quasimodel* $\mathfrak{Q}$ *for* $(\mathcal{T}, \mathcal{A})$ is a set of traces π_d, $d \in D$, for $\mathcal{T}$ such that

(q1) $A \in \pi_a(n)$, for all $A(a, n) \in \mathcal{A}$;

(q2) $B \in \pi_B(0)$, for all $B \in \mathsf{CN}$;

(q3) $\mathsf{proj}_r(\pi_b) \subseteq^0 \pi_a$, for all $r(a, b, n) \in \mathcal{A}$;

(q4) if $A \in \pi_d(n)$ then $\mathsf{proj}_r(\pi_B) \subseteq^n \pi_d$, for all $d \in D$, $n \in \mathbb{Z}$ and $A \sqsubseteq \exists r.B$ in $\mathcal{T}$.

Intuitively, quasimodels represent models of $(\mathcal{T}, \mathcal{A})$: each π_a stands for the ABox individual a; each π_B, on the other hand, represents *all* individuals that witness B for CIs $A \sqsubseteq \exists r.B$ in $\mathcal{T}$. The latter is, in fact, the crucial abstraction underlying quasimodels. Note that traces π_B are normalized: B occurs at time point 0, which is compensated by the shift operation in **(q4)**. For example, in the picture above, if $A \sqsubseteq \exists r.B \in \mathcal{T}$ then, in any model, a has an r-successor that belongs to B at moment 1. Such a successor can be obtained as a 'copy' of trace π_B shifted by 1 so that its origin, 0, matches moment 1 for a. Then, by **(q4)**, a belongs to B' at all odd moments.

For the purposes of query answering we need to identify *canonical (minimal)* quasimodels. We define the canonical quasimodel as the limit of the following saturation (chase-like) procedure. Start with initially empty maps π_d, for $d \in D$, and apply **(t1)**–**(t2)**, **(q1)**–**(q4)** as rules: **(q3)**, for example, says 'if

[1] with the usual semantics: $(r^-)^{\mathfrak{I},n} = \{(e, d) \mid (d, e) \in r^{\mathfrak{I},n}\}$; $\mathfrak{I} \models$ *func*(r) iff $e_1 = e_2$, for all $(d, e_1), (d, e_2) \in r^{\mathfrak{I},n}$ and $n \in \mathbb{Z}$.

$r(a,b,n) \in \mathcal{A}$ and $A \in \mathsf{proj}_r(\pi_b)(i)$, then add A to $\pi_a(i)$.' Then we have the following characterization:

Theorem 4 *Let $\mathfrak{Q} = \{\pi_d \mid d \in D\}$ be the canonical quasimodel of $(\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ a $\mathcal{TEL}^{\circ}$-TBox. Then, for any $A \in \mathsf{CN}$,*

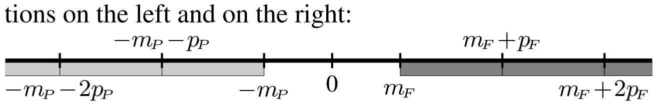
$$(\mathcal{T}, \mathcal{A}) \models A(a,i) \quad \textit{iff} \quad A \in \pi_a(i), \quad \textit{for } a \in \mathsf{ind}(\mathcal{A}),\ i \in \mathbb{Z}.$$

The procedure for constructing the canonical quasimodel deals with infinite data structures (traces) and is generally not terminating. So, although Theorem 4 provides a criterion for certain answers, it does not immediately yield a decision algorithm for full $\mathcal{TEL}^{\circ}$. We remark that known techniques for dealing with such infinite structures cannot be easily applied: for example, MSO (over $\mathbb{Z}$), a standard tool for decidability proofs in temporal DLs [Gabbay et al., 2003], is not sufficient to encode the canonical quasimodel directly because **(q4)** requires $+$. In fact, the key to showing decidability for (fragments of) $\mathcal{TEL}^{\circ}$ is finding a *finite* representation of traces.

The starting point of the rest of the paper is a semantic condition on the canonical quasimodel, *ultimate periodicity*, which ensures decidability in data complexity. Let $\mathcal{T}$ be a $\mathcal{TEL}^{\circ}$-TBox and $\mathfrak{Q}$ the canonical quasimodel for $(\mathcal{T}, \emptyset)$. We say that $\mathcal{T}$ is *ultimately periodic*, if there is $p \in \mathbb{N}$ such that all π_B, $B \in \mathsf{CN}$, in $\mathfrak{Q}$ are *ultimately p-periodic*, that is, for each $B \in \mathsf{CN}$, there are positive integers $m_P, p_P, m_F, p_F \leq p$ satisfying the following conditions:

$$\pi_B(n - p_P) = \pi_B(n), \text{ for all } n \leq -m_P,$$
$$\pi_B(n + p_F) = \pi_B(n), \text{ for all } n \geq m_F.$$

Intuitively, an ultimately p-periodic trace has repeating sections on the left and on the right:



The condition of ultimate periodicity is rather natural. On the practical side, it is motivated by applications with recurrent patterns such as health care support [Shankar et al., 2008], see CIs (1) and (2) in Section 1. From the theoretical point of view, any satisfiable LTL formula has an ultimately periodic model [Manna and Wolper, 1984].

We next show that ultimate periodicity is indeed sufficient for decidability in data complexity.

Theorem 5 *TAQ answering over ultimately periodic $\mathcal{TEL}^{\circ}$-TBoxes is* PSPACE*-complete in data complexity.*

PSPACE-hardness follows from (the proof of) Theorem 2. We prove the matching upper bound by *rewriting* an ultimately periodic $\mathcal{TEL}^{\circ}$-TBox $\mathcal{T}$ into DATALOG$_{1S}$ [Chomicki and Imielinski, 1988]. First, we take temporal rules

$$r(x,y,t \pm 1) \leftarrow r(x,y,t), \qquad \text{for } r \in \mathsf{N}_\mathsf{R}^{\mathsf{rig}} \text{ in } \mathcal{T}, \quad (3)$$
$$B(x,t) \leftarrow A(x,t), A'(x,t), \qquad \text{for } A \sqcap A' \sqsubseteq B \text{ in } \mathcal{T}, \quad (4)$$
$$B(x,t) \leftarrow r(x,y,t), A(y,t), \qquad \text{for } \exists r.A \sqsubseteq B \text{ in } \mathcal{T}, \quad (5)$$

which reflect rigidity of roles and standard $\mathcal{EL}$ concept inclusions on ABox individuals. Second, we observe that, for any trace π_d in the canonical quasimodel $\mathfrak{Q}$ of any $(\mathcal{T}, \mathcal{A})$, if $A \in \pi_d(n)$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$ then, by **(q4)**, π_d contains the n-shift not only of $\mathsf{proj}_r(\pi_B)$ but also of π_A. Since $\mathcal{T}$ is ultimately periodic, for each trace π_B, we fix integers m_P, p_P, m_F, p_F and take the following rules with a fresh predicate F_B:

$$A(x,t+i) \leftarrow B(x,t), \quad \text{for } 0 \leq i < m_F, A \in \pi_B(i),$$
$$A(x,t+i) \leftarrow F_B(x,t), \text{ for } 0 \leq i < p_F, A \in \pi_B(m_F + i),$$
$$F_B(x, t+m_F) \leftarrow B(x,t), \qquad F_B(x,t+p_F) \leftarrow F_B(x,t),$$

and symmetric rules with m_P, p_P and fresh P_B. Intuitively, the rules in the first line replicate the (irregular) part of π_B from 0 to m_F. The two rules in the last line add recurring markers F_B at the start of each period while the rules in the second line replicate the period of π_B starting from each marker F_B.

The required DATALOG$_{1S}$-program $\Pi_\mathcal{T}$ contains all the rules above (note that CIs of the form $\bigcirc_* A \sqsubseteq B$ are also covered by the rules for traces π_B). Using the canonical quasimodel and Theorem 4, it is readily seen that $\Pi_\mathcal{T}$ is equivalent to $\mathcal{T}$: for every temporal ABox $\mathcal{A}$, the answers to $\Pi_\mathcal{T}$ over $\mathcal{A}$ coincide with the certain answers to $(\mathcal{T}, \mathcal{A})$. Theorem 5 follows from PSPACE data complexity in DATALOG$_{1S}$ [Chomicki and Imielinski, 1988] and independence of $\Pi_\mathcal{T}$ from $\mathcal{A}$.

Observe that Theorem 5 does not imply decidability of full $\mathcal{TEL}^{\circ}$ since it is open whether every $\mathcal{TEL}^{\circ}$-TBox is ultimately periodic. We thus turn our attention to *sufficient syntactic conditions* for ultimate periodicity and obtain tight complexity bounds for both data and combined complexity for the resulting fragments. We consider two types of conditions: restricted use of rigid roles and acyclicity of concept inclusions.

5 Restricted Use of Rigid Roles

We first consider $\mathcal{TEL}^{\circ}_{\mathsf{loc}}$, the restriction of $\mathcal{TEL}^{\circ}$ in which *only* local roles are allowed. Due to the reduced interaction between the temporal and DL component, we obtain data tractability.

Theorem 6 *TAQ answering over $\mathcal{TEL}^{\circ}_{\mathsf{loc}}$ is* PSPACE*-complete in combined and* PTIME*-complete in data complexity.*

PSPACE-hardness follows from the proof of PSPACE-hardness for entailment in Horn-LTL [Chen and Lin, 1993] and PTIME-hardness from atomic query answering in $\mathcal{EL}$. For the upper bounds, let $(\mathcal{T}, \mathcal{A})$ be a KB with $\mathcal{T}$ a $\mathcal{TEL}^{\circ}_{\mathsf{loc}}$-TBox and $\mathfrak{Q}$ its canonical quasimodel. We take a propositional variable $P_{A,d}$ for each $A \in \mathsf{CN}$ and $d \in D$ and construct a Horn-LTL formula $\varphi_{\mathcal{T},\mathcal{A}}$ whose minimal model is isomorphic to $\mathfrak{Q}$: variable $P_{A,d}$ is true in the model at moment n iff $A \in \pi_d(n)$. We take the conjunction of the following formulas, for $d \in D$:

$$\Box(P_{A,d} \wedge P_{A',d} \rightarrow P_{B,d}), \quad \text{for } A \sqcap A' \sqsubseteq B \in \mathcal{T},$$
$$\Box(\bigcirc_* P_{A,d} \rightarrow P_{B,d}), \quad \text{for } \bigcirc_* A \sqsubseteq B \in \mathcal{T},$$
$$\bigcirc^n P_{A,a}, \quad \text{for } A(a,n) \in \mathcal{A},$$
$$P_{B,B}, \quad \text{for } B \in \mathsf{CN},$$
$$\bigcirc^n P_{B,b} \rightarrow \bigcirc^n P_{A,a}, \quad \text{for } r(a,b,n) \in \mathcal{A}, \exists r.B \sqsubseteq A \in \mathcal{T},$$
$$P_{B',B} \rightarrow \Box(P_{A,d} \rightarrow P_{A',d}), \quad \text{for } A \sqsubseteq \exists r.B, \exists r.B' \sqsubseteq A' \in \mathcal{T},$$

where $\bigcirc^n$ is $\bigcirc_F^n$ if $n \geq 0$ and $\bigcirc_P^{-n}$ if $n < 0$ and $\Box$ is the 'globally' operator. It is readily verified that $\varphi_{\mathcal{T},\mathcal{A}}$ is as required. Crucially, **(q4)** for local roles boils down to the last formula above. Since entailment in LTL is in PSPACE [Sistla and Clarke, 1985] and $\varphi_{\mathcal{T},\mathcal{A}}$ is polynomial in the size of $(\mathcal{T}, \mathcal{A})$, we obtain membership in PSPACE for combined complexity.

To obtain the PTIME data complexity, observe that traces π_B, $B \in \mathsf{CN}$, are ultimately $2^{|\mathcal{T}|}$-periodic because they are traces of the canonical quasimodel for $(\mathcal{T}, \emptyset)$; so, they can be maintained in constant space. Next, traces π_a, $a \in \mathsf{ind}(\mathcal{A})$, are ultimately $2^{|\mathcal{T}|+|\mathcal{A}|}$-periodic, but a closer inspection reveals that the middle irregular section, $m_P + m_F$, is bounded by $|\mathcal{A}| + 2^{|\mathcal{T}|}$, while both periods, p_P and p_F, by $2^{|\mathcal{T}|}$; see, e.g., Lemma 3 [Artale et al., 2013a]. Thus, $\mathfrak{Q}$ can be stored in space bounded by a polynomial in $|\mathcal{A}|$. Since each rule application extends the traces, the saturation procedure for constructing $\mathfrak{Q}$ terminates in polynomial time in the size of $\mathcal{A}$.

Since ontologies without rigid roles at all may be too restrictive for applications, we consider $\mathcal{TEL}^{\circ}_{\mathsf{l\text{-}rig}}$-TBoxes where rigid roles are allowed only in CIs of the form $\exists r.B \sqsubseteq A$.

Theorem 7 *TAQ answering over $\mathcal{TEL}^{\circ}_{\mathsf{l\text{-}rig}}$ is* PSPACE*-complete in data complexity and in* EXPTIME *in combined complexity.*

PSPACE-hardness in data complexity follows from the proof of Theorem 2. For the upper bounds, we construct rewritings into DATALOG_{1S}, similarly to $\Pi_{\mathcal{T}}$ in Section 4 (Theorem 5).

6 Acyclicity Conditions

It is known that acyclicity conditions may lead to better complexity. In particular, acyclic TBoxes are a way of obtaining CTL-based temporal extensions of $\mathcal{EL}$ that have rigid roles and enjoy PTIME subsumption [Gutiérrez-Basulto, Jung, and Schneider, 2015]. In DATALOG_{1S}, a restriction on recursion has also been used to attain tractability [Chomicki, 1990]. From the application point of view, large parts of SNOMED CT and GO [Gene Ontology Cons., 2000] are indeed acyclic. So, we believe that the fragments we consider below are well-suited for temporal extensions of such ontologies.

Acyclic TBoxes are finite sets of CDs $A \equiv C$, $A \in \mathsf{N_C}$, such that no two CDs have the same left-hand side, and there are no CDs $A_1 \equiv C_1, \ldots, A_k \equiv C_k$ in $\mathcal{T}$ such that A_{i+1} occurs in C_i, for all $1 \leq i \leq k$ (where $A_{k+1} := A_1$). We say A is *defined in* $\mathcal{T}$ if $A \equiv C \in \mathcal{T}$ and *primitive* otherwise.

Theorem 8 *TAQ answering over acyclic $\mathcal{TEL}^{\circ}$ is in* LOGTIME*-uniform AC^0 in data complexity and in* PTIME *in combined complexity.*

The LOGTIME-uniform AC^0 upper bound is established by *rewriting into FO(+)*: for a given TAQ $A(x, t)$ and TBox $\mathcal{T}$, we construct a two-sorted first-order formula $\varphi_{\mathcal{T},A}(x, t)$ with functions $+1$ and -1 on temporal terms such that $(\mathcal{T}, \mathcal{A}) \models A(a, i)$ iff $\mathcal{A}$ (viewed as an interpretation) is a model of $\varphi_{\mathcal{T},A}(a, i)$, for all ABoxes $\mathcal{A}$, $a \in \mathsf{ind}(\mathcal{A})$, $i \in \mathbb{Z}$. We adapt the technique developed for atemporal $\mathcal{EL}$ [Bienvenu, Lutz, and Wolter, 2012]:

$$\begin{aligned}
\varphi_{\mathcal{T},A}(x,t) &= \mathsf{S}_A(x,t), && \text{if } A \text{ is primitive,}\\
\varphi_{\mathcal{T},A}(x,t) &= \mathsf{S}_A(x,t) \vee \varphi_{\mathcal{T},C}(x,t), && \text{if } A \equiv C \in \mathcal{T},\\
\varphi_{\mathcal{T},B_1 \sqcap B_2}(x,t) &= \varphi_{\mathcal{T},B_1}(x,t) \wedge \varphi_{\mathcal{T},B_2}(x,t), &&\\
\varphi_{\mathcal{T},\exists r.B}(x,t) &= \exists y \big(\mathsf{R}_r(x,y,t) \wedge \varphi_{\mathcal{T},B}(y,t)\big), &&\\
\varphi_{\mathcal{T},\bigcirc_* B}(x,t) &= \varphi_{\mathcal{T},B}(x, t \,\mathsf{op}_*\, 1), &&
\end{aligned}$$

where $\mathsf{S}_A(x, t)$ is a disjunction of all $B(x, t)$ for concept names B with $\mathcal{T} \models B \sqsubseteq A$, and $\mathsf{R}_r(x, y, t)$ is $r(x, y, t)$ for $r \in \mathsf{N_R^{loc}}$ and $\exists t'\, r(x, y, t')$ for $r \in \mathsf{N_R^{rig}}$.

Note that $\varphi_{\mathcal{T},A}$ is an $\text{FO}^{\mathbb{Z}}$-rewriting in the terminology of Artale et al. [2013b; 2015] because the temporal terms range over $\mathbb{Z}$. However, the infinite interpretation of $\mathcal{A}$ is empty after at most $|\mathcal{T}|$ steps from the ABox and so, $\varphi_{\mathcal{T},A}$ can be converted into an FO-rewriting whose temporal terms range over $\mathsf{tem}(\mathcal{A})$ only; see [Artale et al., 2015].

We next introduce novel notions of acyclicity that restrict only one dimension, DL or temporal.

DL Acyclicity

First, we introduce DL-acyclic $\mathcal{TEL}^{\circ}$-TBoxes, which are well-suited as temporal extensions of, say, biomedical ontologies that may require recurrent patterns but have an acyclic DL component. A $\mathcal{TEL}^{\circ}$-TBox $\mathcal{T}$ with concept names CN is called *DL-acyclic* if there is a mapping $\ell_{\mathsf{DL}} \colon \mathsf{CN} \to \mathbb{N}$ such that:

(i) $A \sqsubseteq \exists r.B$ or $\exists r.B \sqsubseteq A \in \mathcal{T}$ implies $\ell_{\mathsf{DL}}(A) > \ell_{\mathsf{DL}}(B)$;
(ii) $\bigcirc_* A \sqsubseteq B$ implies $\ell_{\mathsf{DL}}(A) = \ell_{\mathsf{DL}}(B)$;
(iii) $A \sqcap A' \sqsubseteq B \in \mathcal{T}$ implies $\ell_{\mathsf{DL}}(A) = \ell_{\mathsf{DL}}(A') = \ell_{\mathsf{DL}}(B)$.

A DL-acyclic TBox is of *depth k* if k is minimal such that a witnessing mapping ℓ_{DL} satisfies $\ell_{\mathsf{DL}}(B) \leq k$ for all $B \in \mathsf{CN}$.

Theorem 9 *TAQ answering over DL-acyclic $\mathcal{TEL}^{\circ}$-TBoxes of depth $k \geq 1$ is k-*EXPSPACE*-complete in combined complexity and* NC^1*-complete in data complexity.*

A closer inspection of the non-elementary lower bound proof in Theorem 2 reveals that the TBox used is DL-acyclic and TAQ answering over TBoxes of depth k is k-EXPSPACE-hard. NC^1-hardness in data complexity follows [Artale et al., 2015] by reduction of the word problem of NFAs to TAQ answering, even without the DL dimension.

For the matching upper bounds, fix $(\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ of depth k. We devise a completion procedure, which is based on special LTL-formulas and implies ultimate periodicity of all traces in the canonical quasimodel of $(\mathcal{T}, \mathcal{A})$; cf. Section 5. Given any $\mathcal{A}$, let the slice $\mathcal{A}_i$ consist of all $A(a, i) \in \mathcal{A}$, all $r(a, b, i) \in \mathcal{A}$ and all $r(a, b, i)$ with $r(a, b, j) \in \mathcal{A}$, for some j, and $r \in \mathsf{N_R^{rig}}$. The algorithm separates consequences of the role structure of $\mathcal{A}$ and local temporal consequences of $\mathcal{T}$. In particular, it exhaustively extends $\mathcal{A}$ by all $A(a, i)$ with either

$$(\mathcal{T}, \mathcal{A}_i) \models A(a, i) \;\text{ or }\; B(a, i \,\mathsf{op}_*\, 1) \in \mathcal{A}, \bigcirc_* B \sqsubseteq A \in \mathcal{T}. \quad (6)$$

It turns out that $\mathcal{A}_i$ in (6) can be replaced by its suitably defined *quotient* $\mathcal{B}_i$. Intuitively, the logic can only distinguish distinct trees of depth k, whose number depends on $|\mathcal{T}|$ only; so, the size of $\mathcal{B}_i$ is independent of $|\mathcal{A}|$. By induction on depth k, we define LTL-formulas $\varphi_{a,i}$ of k-fold-exponential size characterizing all $A \in \mathsf{CN}$ with $(\mathcal{T}, \mathcal{B}_i) \models A(a, i)$: we begin from formulas as in Theorem 6; the induction step takes account of the structure of $\mathcal{B}_i$ and incurs an exponential blowup.

Now, for combined complexity, observe that each of the polynomially many $\varphi_{a,i}$ can be analyzed in k-EXPSPACE. For data complexity, observe that checking $(\mathcal{T}, \mathcal{B}_i) \models A(a, i)$ can be done in *constant* time. The second option in (6), however, cannot be implemented directly as the number of steps depends on $|\mathcal{A}|$. Instead, we construct a Büchi automaton that accepts precisely the traces for $\mathcal{T}$ and cast the second option in (6) as the question of whether all traces extending $\mathcal{A}$ have A at position i, which is a regular property and so, is in NC^1.

Temporal Acyclicity

We next relax acyclicity by admitting recursion in the DL dimension (but not in temporal); thus, temporally acyclic TBoxes include general $\mathcal{EL}$-TBoxes. A $\mathcal{TEL}^{\bigcirc}$-TBox $\mathcal{T}$ with concept names CN is *temporally acyclic* if there is $\ell_{\bigcirc}\colon \mathsf{CN} \to \mathbb{N}$ such that:

(i) $\bigcirc_P A \sqsubseteq B$ or $\bigcirc_F B \sqsubseteq A \in \mathcal{T}$ implies $\ell_{\bigcirc}(B) = \ell_{\bigcirc}(A)+1$;

(ii) $\exists r.B \sqsubseteq A$ or $A \sqsubseteq \exists r.B \in \mathcal{T}$ implies $\ell_{\bigcirc}(A) = \ell_{\bigcirc}(B)$;

(iii) $A \sqcap A' \sqsubseteq B \in \mathcal{T}$ implies $\ell_{\bigcirc}(A) = \ell_{\bigcirc}(A') = \ell_{\bigcirc}(B)$.

Temporally acyclic TBoxes cannot, unlike DL acyclic ones, express rigid concepts. Still, we can partition concept names $\mathsf{N_C}$ into local $\mathsf{N_C^{loc}}$ and rigid $\mathsf{N_C^{rig}}$ and obtain the following:

Theorem 10 *TAQ answering over temporally acyclic $\mathcal{TEL}^{\bigcirc}$ (with rigid concepts) is* PTIME*-complete in data and combined complexity.*

The lower bounds are from $\mathcal{EL}$. For the upper bounds, we show a *small quasimodel property*: traces of the canonical quasimodel of any $(\mathcal{T}, \mathcal{A})$ with such a TBox $\mathcal{T}$ satisfy

$$\pi_a(j) = \pi_a(j'), \qquad \text{if } j, j' > u + |\mathcal{T}| \quad \text{or} \quad j, j' < l - |\mathcal{T}|,$$
$$\pi_B(j) = \pi_B(j'), \qquad \text{if } j, j' > |\mathcal{T}| \quad \text{or} \quad j, j' < -|\mathcal{T}|,$$

where $u = \max \mathcal{A}$ and $l = \min \mathcal{A}$. Intuitively, the canonical quasimodel has a restricted temporal extension that stretches only $|\mathcal{T}|$ time points beyond $\mathcal{A}$. By the small quasimodel property, the procedure for constructing the canonical quasimodel can be implemented in polynomial time: traces π_d require only polynomial space, and rules **(q1)–(q4)** extend the traces.

Inflationary $\mathcal{TEL}^{\Diamond}$

Next, we follow an approach suggested by Artale et al. [2013b] (in the context of temporal *DL-Lite*) and restrict $\mathcal{TEL}^{\Diamond}$ by allowing $\Diamond_*$ only on the *left-hand side* of CIs. This fragment is denoted by $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$, for *inflationary* $\mathcal{TEL}$ (which is related to inflationary DATALOG$_{1S}$ [Chomicki, 1990]). Note that $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ extends general $\mathcal{EL}$-TBoxes. Yet, the complexity is the same:

Theorem 11 *TAQ answering over $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ is* PTIME*-complete in both data and combined complexity.*

We need to show only the upper bounds. Observe that $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ can still be viewed as a fragment of $\mathcal{TEL}^{\bigcirc}$; see Section 2. In fact, one can show an analogue of Theorem 4 with the following replacement of **(t2)**:

(t2′) if $\Diamond_* A \sqsubseteq B \in \mathcal{T}$ and $A \in \pi_d(n)$, then $B \in \pi_d(n')$ for all $n' > n$ if $* = P$ and for all $n' < n$ if $* = F$.

We establish a special shape of the traces in the canonical model of any $(\mathcal{T}, \mathcal{A})$. Let $\varrho\colon \mathbb{Z} \to 2^{\mathsf{CN}}$ be a map and let $l, u \in \mathbb{Z}$ with $l \leq u$. We say that ϱ is an $[l, u]$-*bow tie* if

– for all $i > u$, we have $\varrho(i+1) \supseteq \varrho(i)$, and if $\varrho(i+1) = \varrho(i)$ then all $\varrho(i')$, for $n' \geq i$, coincide;

– symmetrically, for all $i < l$, we have $\varrho(i-1) \supseteq \varrho(i)$, and if $\varrho(i-1) = \varrho(i)$ then all $\varrho(i')$, for $i' \leq i$, coincide.

These properties mean that ϱ grows monotonically to the right of u and to the left of l; in other words, ϱ has *inflationary* behaviour. We prove that the traces π_d in the canonical quasimodel $\mathfrak{Q}$ of $(\mathcal{T}, \mathcal{A})$, for any $\mathcal{A}$, enjoy the following properties:

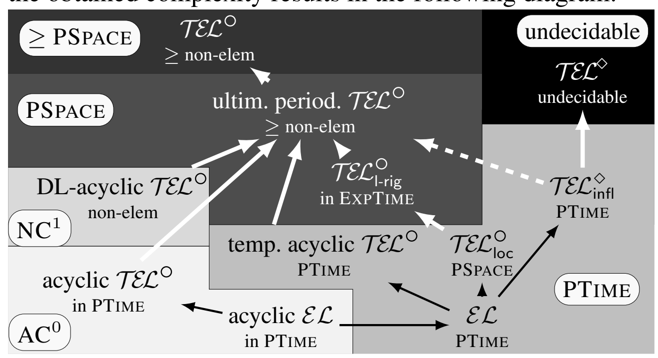
– π_a is a $[\min \mathcal{A}, \max \mathcal{A}]$-bow tie, for each $a \in \mathsf{ind}(\mathcal{A})$;

– π_B is a $[0, 0]$-bow tie, for each $B \in \mathsf{CN}$.

Thus, the traces in $\mathfrak{Q}$ can be represented in polynomial space because only the middle section and at most $|\mathsf{CN}|$ steps at both ends need to be stored. Since the traces are extended with every rule application, the procedure terminates after polynomially many steps; Theorem 11 follows.

7 Discussion and Future Work

We summarize the fragments of $\mathcal{TEL}$, their relationships and the obtained complexity results in the following diagram:



where the solid lines are inclusions between DLs, the dashed line is a reduction that preserves answers to all queries (model conservative extension). The data complexity is indicated by shading and the combined complexity is below the language.

Our data-tractability results show theoretical adequacy of the identified fragments of $\mathcal{TEL}$ for data-intensive applications. Our two novel forms of acyclicity, DL- and temporal, are somewhat close in spirit to *multi-separability* [Chomicki, 1990]: the latter, however, puts a weaker restriction on recursion but a stricter one on the interaction between the temporal and data component. DL-acyclic $\mathcal{TEL}^{\bigcirc}$ is the first (to the best of our knowledge) DL shown to have NC^1-complete query answering (the large gap between data and combined complexity is also remarkable). On the practical side, there is evidence that such data-tractable fragments should be sufficient for many biomedical applications. Following the principles of OBDA, our framework provides a means of defining temporal concepts in the ontology for these applications: temporal concepts capture both (restricted) tree-shaped temporal conjunctive queries (CQs) and recurring temporal patterns.

As our immediate future work, we will address decidability of (full) $\mathcal{TEL}^{\bigcirc}$ and then consider CQs with the $+$ operation on temporal terms. We expect that our positive results can be lifted to CQs using the *combined approach* [Lutz, Toman, and Wolter, 2009], which utilizes a structure similar to our canonical quasimodel. We will also study succinct and expressive representations of temporal data. For example, the only known algorithm for DATALOG$_{1S}$ with *binary* encoding of timestamps in the data runs in EXPTIME in the size of the data [Chomicki and Imielinski, 1988]. We, however, presume that careful materialization should be sufficient to deal with the issue. We will also consider *interval* encoding of temporal ABoxes, e.g., $A(a, [n_1, n_2])$, and settings capturing *infinite* temporal periodic data as introduced by Kabanza, Stévenne, and Wolper [1990] and Chomicki and Imielinski [1993].

References

[Artale et al., 2007] Artale, A.; Kontchakov, R.; Lutz, C.; Wolter, F.; and Zakharyaschev, M. 2007. Temporalising tractable description logics. In *Proc. TIME-07*, 11–22.

[Artale et al., 2013a] Artale, A.; Kontchakov, R.; Ryzhikov, V.; and Zakharyaschev, M. 2013a. The complexity of clausal fragments of LTL. In *Proc. LPAR-19, 2013*, 35–52.

[Artale et al., 2013b] Artale, A.; Kontchakov, R.; Wolter, F.; and Zakharyaschev, M. 2013b. Temporal description logic for ontology-based data access. In *Proc. IJCAI-13*.

[Artale et al., 2015] Artale, A.; Kontchakov, R.; Kovtunova, A.; Ryzhikov, V.; Wolter, F.; and Zakharyaschev, M. 2015. First-order rewritability of temporal ontology-mediated queries. In *Proc. IJCAI-15*, 2706–2712.

[Baader, Borgwardt, and Lippmann, 2015] Baader, F.; Borgwardt, S.; and Lippmann, M. 2015. Temporal query entailment in the description logic SHQ. *J. Web Sem.* 33:71–93.

[Baader, Brandt, and Lutz, 2005] Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proc. IJCAI-05*.

[Baudinet, Chomicki, and Wolper, 1993] Baudinet, M.; Chomicki, J.; and Wolper, P. 1993. Temporal deductive databases. In *Temporal Databases*. Benjamin/Cummings.

[Bienvenu, Lutz, and Wolter, 2012] Bienvenu, M.; Lutz, C.; and Wolter, F. 2012. Deciding FO-rewritability in $\mathcal{EL}$. In *Proc. DL-12*.

[Borgwardt and Thost, 2015] Borgwardt, S., and Thost, V. 2015. Temporal query answering in the description logic EL. In *Proc. IJCAI-15*, 2819–2825.

[Borgwardt, Lippmann, and Thost, 2015] Borgwardt, S.; Lippmann, M.; and Thost, V. 2015. Temporalizing rewritable query languages over knowledge bases. *J. Web Sem.* 33:50–70.

[Chen and Lin, 1993] Chen, C.-C., and Lin, I.-P. 1993. The computational complexity of satisfiability of temporal Horn formulas in propositional linear-time temporal logic. *Inf. Process. Lett.* 45(3):131–136.

[Chomicki and Imielinski, 1988] Chomicki, J., and Imielinski, T. 1988. Temporal deductive databases and infinite objects. In *Proc. PODS-88*, 61–73.

[Chomicki and Imielinski, 1993] Chomicki, J., and Imielinski, T. 1993. Finite representation of infinite query answers. *ACM Trans. Database Syst.* 18(2):181–223.

[Chomicki and Toman, 2005] Chomicki, J., and Toman, D. 2005. Temporal Databases. In *Handbook of Temporal Reasoning in Artificial Intelligence*, 429–468. Elsevier.

[Chomicki, 1990] Chomicki, J. 1990. Polynomial time query processing in temporal deductive databases. In*Proc. PODS*.

[Dong and Tan, 2015] Dong, X. L., and Tan, W. 2015. A time machine for information: Looking back to look forward. *PVLDB* 8(12):2044–2055.

[Gabbay et al., 2003] Gabbay, D.; Kurucz, A.; Wolter, F.; and Zakharyaschev, M. 2003. *Many-dimensional modal logics: theory and applications*. Elsevier.

[Gene Ontology Cons., 2000] Gene Ontology Cons. 2000. Gene ontology: Tool for the unification of biology. *Nature Genetics* 25:25–29.

[Gutiérrez-Basulto and Klarman, 2012] Gutiérrez-Basulto, V., and Klarman, S. 2012. Towards a unifying approach to representing and querying temporal data in description logics. In *Proc. RR-12*, 90–105.

[Gutiérrez-Basulto, Jung, and Lutz, 2012] Gutiérrez-Basulto, V.; Jung, J. C.; and Lutz, C. 2012. Complexity of branching temporal description logics. In *Proc. ECAI-12*.

[Gutiérrez-Basulto, Jung, and Schneider, 2014] Gutiérrez-Basulto, V.; Jung, J.; and Schneider, T. 2014. Lightweight description logics and branching time: a troublesome marriage. In *Proc. KR-14*.

[Gutiérrez-Basulto, Jung, and Schneider, 2015] Gutiérrez-Basulto, V.; Jung, J.; and Schneider, T. 2015. Lightweight temporal description logics with rigid roles and restricted TBoxes. In *Proc. IJCAI-15*, 3015–3021.

[Haase and Lutz, 2008] Haase, C., and Lutz, C. 2008. Complexity of subsumption in the $\mathcal{EL}$ family of description logics: Acyclic and cyclic TBoxes. In *Proc. of ECAI-08*.

[Kabanza, Stévenne, and Wolper, 1990] Kabanza, F.; Stévenne, J.; and Wolper, P. 1990. Handling infinite temporal data. In *Proc. PODS-90*, 392–403.

[Klarman and Meyer, 2014] Klarman, S., and Meyer, T. 2014. Querying temporal databases via OWL 2 QL. In *Proc. RR-14*, 92–107.

[Krisnadhi and Lutz, 2007] Krisnadhi, A., and Lutz, C. 2007. Data complexity in the $\mathcal{EL}$ family of description logics. In *Proc. LPAR-07*, 333–347.

[Lutz, Toman, and Wolter, 2009] Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. IJCAI-09*.

[Manna and Wolper, 1984] Manna, Z., and Wolper, P. 1984. Synthesis of communicating processes from temporal logic specifications. *ACM Trans.Program.Lang.Syst.* 6(1):68–93.

[O'Connor et al., 2009] O'Connor, M. J.; Shankar, R. D.; Parrish, D. B.; and Das, A. K. 2009. Knowledge-data integration for temporal reasoning in a clinical trial system. *Int. J. Med. Inform.* 78:S77–S85.

[Roth and Tan, 2013] Roth, M., and Tan, W. 2013. Data integration and data exchange: It's really about time. In *Proc. CIDR-13*.

[Shankar et al., 2008] Shankar, R. D.; Martins, S. B.; O'Connor, M. J.; Parrish, D. B.; and Das, A. K. 2008. Representing and reasoning with temporal constraints in clinical trials using semantic technologies. In *BIOSTEC-08, Revised Selected Papers*, 520–530.

[Sistla and Clarke, 1985] Sistla, A. P., and Clarke, E. M. 1985. The complexity of propositional linear temporal logics. *J. ACM* 32(3):733–749.

[Stockmeyer, 1974] Stockmeyer, L. J. 1974. *The Complexity of Decision Problems in Automata Theory and Logic*. Ph.D. Dissertation, MIT, Cambridge, Massachusetts, USA.