Tractable Extensions of the Description Logic $\mathcal{EL}$ with Numerical Datatypes

Despoina Magka, Yevgeny Kazakov, and Ian Horrocks

Oxford University Computing Laboratory
Wolfson Building, Parks Road, OXFORD, OX1 3QD, UK
{despoina.magka,yevgeny.kazakov,ian.horrocks}@comlab.ox.ac.uk

1 Introduction and Motivation

Description logics (DLs) [1] provide a logical foundation for modern ontology languages such as OWL[1] and OWL 2 [2]. $\mathcal{EL}^{++}$ [3] is a lightweight DL for which reasoning is tractable (i.e., can be performed in time that is polynomial w.r.t. the size of the input), and that offers sufficient expressivity for a number of life-sciences ontologies, such as SNOMED CT [4] or the Gene Ontology [5]. Among other constructors, $\mathcal{EL}^{++}$ supports limited usage of datatypes. In DL, datatypes (also called concrete domains) can be used to define new concepts by referring to particular values, such as strings or integers. For example, the concept $\mathsf{Human} \sqcap \exists \mathsf{hasAge}.(<, \mathsf{18}) \sqcap \exists \mathsf{hasName}.(=, \text{“}\mathsf{Alice}\text{”})$ describes humans, named “Alice”, whose age is less than 18. Datatypes are described first by the domain their values can come from and also by the relations that can be used to constrain possible values. In our example, $(<, \mathsf{18})$ refers to the domain of natural numbers and uses the relation “<” to constrain possible values to those less than 18, while $(=, \text{“}\mathsf{Alice}\text{”})$ refers to the domain of strings and uses the relation “=” to constrain the value to “Alice”.

In order to ensure that reasoning remains polynomial, $\mathcal{EL}^{++}$ allows only for datatypes which satisfy a condition called p-admissibility [3]. In an nutshell, this condition ensures that the satisfiability of datatype constraints can be solved in polynomial time, and that concept disjunction cannot be expressed using datatype concepts. For example, if we were to allow both $\leq$ and $\geq$ for integers, then we could express $A \sqsubseteq B \sqcup C$ by formulating the axioms $\mathsf{A} \sqsubseteq \exists \mathsf{R}.(\leq, \mathsf{5})$, $\exists \mathsf{R}.(\leq, 2) \sqsubseteq \mathsf{B}$ and $\exists \mathsf{R}.(\geq, 2) \sqsubseteq \mathsf{C}$. Similarly, we can show that p-admissibility does not allow for both $<$ (or $>$) and $=$. For this reason, the EL Profile of OWL 2, which is based on $\mathcal{EL}^{++}$, admits only equality (=) in datatype expressions.

In this paper, we demonstrate how these restrictions can be significantly relaxed without loosing tractability. As a motivating example, consider the following two axioms which might be used, e.g., in a pharmacy-related ontology:

$$\mathsf{Panadol} \sqsubseteq \exists \mathsf{contains}.(\mathsf{Paracetamol} \sqcap \exists \mathsf{mgPerTablet}.(=, 500)) \tag{1}$$

$$\begin{aligned}&\mathsf{Patient} \sqcap \exists \mathsf{hasAge}.(<, 6) \sqcap \\ &\quad \exists \mathsf{hasPrescription}.\exists \mathsf{contains}.(\mathsf{Paracetamol} \sqcap \exists \mathsf{mgPerTablet}.(>, 250)) \sqsubseteq \bot\end{aligned} \tag{2}$$

[1] http://www.w3.org/2004/OWL

Axiom (1) states that the drug Panadol contains 500 mg of paracetamol per tablet, while axiom (2) states that a drug that contains more than 250 mg of paracetamol per tablet must not be prescribed to a patient younger than 6 years old. The ontology could be used, for example, to support clinical staff who want to check whether Panadol can be prescribed to a 3-year-old patient. This can easily be achieved by checking whether concept (3) is satisfiable w.r.t. the ontology:

$$\mathsf{Patient} \sqcap \exists \mathsf{hasAge}.(=, 3) \sqcap \exists \mathsf{hasPrescription}.\mathsf{Panadol} \tag{3}$$

Unfortunately, this is not possible using $\mathcal{EL}^{++}$, because axioms (1) and (2) involve both equality (=) and inequalities (<, >), and this violates the *p*-admissibility restriction. In this paper we demonstrate that it is, however, possible to express axioms (1) and (2) and concept (3) in a tractable extension of $\mathcal{EL}$. A polynomial classification procedure can then be used to determine the satisfiability of (3) w.r.t. the ontology by checking if adding an axiom

$$X \sqsubseteq \mathsf{Patient} \sqcap \exists \mathsf{hasAge}.(=, 3) \sqcap \exists \mathsf{hasPrescription}.\mathsf{Panadol}$$

for some new concept name X would entail $X \sqsubseteq \bot$.

Our idea is based on the intuition that equality in (1) and (3) serves a different purpose than inequalities do in (2). Equality in (1) and (3) is used to state a *fact* (the content of a drug and the age of a patient) whereas inequalities in (2) are used to trigger a *rule* (what happens if a certain quantity of drug is prescribed to a patient of a certain age). In other words, equality is used *positively* and inequalities are used *negatively*. It seems reasonable to assume that positive usages of datatypes will typically involve equality since a fact can usually be precisely stated. On the other hand, it seems reasonable to assume that negative occurrences of datatypes will typically involve equality as well as inequalities since a rule usually applies to a range of situations. In this paper, we make a fine-grained study of datatypes in $\mathcal{EL}$ by considering restrictions not only on the kinds of relations included in a datatype, but also on whether the relations can be used positively or negatively. The main contributions of this paper can be summarised as follows:

1. We introduce the notion of a *Numerical Datatype with Restrictions* (*NDR*) that specifies the domain of the datatype, the datatype relations that can be used positively and the datatype relations that can be used negatively.
2. We extend the $\mathcal{EL}$ reasoning algorithm [3] to provide a polynomial reasoning procedure for an extension of $\mathcal{EL}$ with NDRs, where the procedure is sound for any NDR.
3. We introduce the notion of a *safe NDR*, show that every extension of $\mathcal{EL}$ with a safe NDR is tractable and prove that our reasoning procedure is complete for any safe NDR.
4. Finally, we provide a complete classification of safe NDRs for the cases of natural numbers, integers, rationals and reals. Notably, we demonstrate that the numerical datatype restrictions can be significantly relaxed by allowing arbitrary numerical relations to occur negatively—not only equality as currently specified in the OWL 2 EL Profile. As argued earlier, this combination

Table 1. Concept descriptions in $\mathcal{EL}^{\bot}(\mathcal{D})$

NAME	SYNTAX	SEMANTICS
Concept name	C	$C^{\mathcal{I}}$
Top	$\top$	$\Delta^{\mathcal{I}}$
Bottom	$\bot$	$\emptyset$
Conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
Existential restriction	$\exists R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
Datatype restriction	$\exists F.r$	$\{x \in \Delta^{\mathcal{I}} \mid \exists v \in \mathcal{D} : (x,v) \in F^{\mathcal{I}} \wedge r(v)\}$

is of particular interest to ontology engineering, and is thus a strong candidate for the next extension of the EL Profile in OWL 2.

2 Preliminaries

In this section we introduce $\mathcal{EL}^{\bot}(\mathcal{D})$, an extension of $\mathcal{EL}^{\bot}$ [3] with numerical datatypes. In the DL literature datatypes are better known as concrete domains [6]; we call them datatypes to be more consistent with OWL 2 [2]. The syntax of $\mathcal{EL}^{\bot}(\mathcal{D})$ uses a set of *concept names* N_C, a set of *role names* N_R and a set of *feature names* N_F. $\mathcal{EL}^{\bot}(\mathcal{D})$ is parametrised with a *numerical domain* $\mathcal{D} \subseteq \mathbb{R}$ ($\mathbb{R}$ is the set of real numbers). N_C, N_R and N_F are countably infinite sets and, additionally, pairwise disjoint. We call (s, y), where $s \in \{<, \leq, >, \geq, =\}$ and $y \in \mathcal{D}$, a *$\mathcal{D}$-datatype restriction* (or simply a *datatype restriction* if the domain $\mathcal{D}$ is clear from the context). Given a $\mathcal{D}$-datatype restriction $r = (s, y)$ and an $x \in \mathcal{D}$, we say that x satisfies r and we write $r(x)$ iff $(x, y) \in s$, where $s \in \{<, \leq, >, \geq, =\}$ and s is interpreted as the standard relation on real numbers. Table 1 recursively defines concepts in $\mathcal{EL}^{\bot}(\mathcal{D})$, where C and D are concepts, $R \in N_R$, $F \in N_F$ and r is a $\mathcal{D}$-datatype restriction. An *axiom α (in $\mathcal{EL}^{\bot}(\mathcal{D})$)* is an expression of the form $C \sqsubseteq D$, where C and D are concepts. An ($\mathcal{EL}^{\bot}(D)-$)*ontology* $\mathcal{O}$ is a set of axioms. A concept E is said to *positively (negatively) occur* in an axiom $C \sqsubseteq D$ iff it occurs in D (C). An interpretation of $\mathcal{EL}^{\bot}(\mathcal{D})$ is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set, the *domain of the interpretation*, and $\cdot^{\mathcal{I}}$ is the *interpretation function*. The interpretation function maps each $A \in N_C$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, each $R \in N_R$ to a relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and each $F \in N_F$ to a relation $F^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \mathcal{D}$. Note that we do not require the interpretation of features to be functional. In this respect, they correspond to the data properties in OWL 2 [2]. The constructors of $\mathcal{EL}^{\bot}(\mathcal{D})$ are interpreted as indicated in Table 1. *An interpretation $\mathcal{I}$ satisfies an axiom* $\alpha = C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ (written $\mathcal{I} \models \alpha$). If $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{O}$, then $\mathcal{I}$ *is a model of* $\mathcal{O}$ (written $\mathcal{I} \models \mathcal{O}$). If every model $\mathcal{I}$ of $\mathcal{O}$ satisfies the axiom α then we say that $\mathcal{O}$ *entails* α and we write $\mathcal{O} \models \alpha$. We define the *signature of an ontology* $\mathcal{O}$ as the set $sig(\mathcal{O})$ of concept, role and feature names that occur in $\mathcal{O}$. We say that an axiom in $\mathcal{EL}^{\bot}(\mathcal{D})$ is in normal form if it has one of the forms: $A \sqsubseteq B'$ (NF1), $A_1 \sqcap A_2 \sqsubseteq B$ (NF2), $A \sqsubseteq \exists R.B$ (NF3), $\exists R.B \sqsubseteq A$ (NF4), $A \sqsubseteq \exists F.r$ (NF5) or $\exists F.r \sqsubseteq A$ (NF6), where A, A_1, A_2, $B \in N_C^{\top}$, $B' \in N_C^{\top,\bot}$, $R \in N_R$, $F \in N_F$ and r is a $\mathcal{D}$-datatype restriction. The normalization procedure is the same as for the $\mathcal{EL}^{++}$ [3]; more details can be found in the technical report [7]. ($N_C^{\top} = N_C \cup \{\top\}$, $N_C^{\top,\bot} = N_C \cup \{\top, \bot\}$).

3 Numerical Datatypes with Restrictions

In this section we introduce the notion of a Numerical Datatype with Restrictions (NDR) which specifies which datatype relations can be used positively and negatively. We then present a sound polynomial consequence-based classification procedure for $\mathcal{EL}^{\bot}$ extended with NDRs. Finally we prove that the procedure is complete if the NDR satisfies special safety requirements.

Definition 1 (Numerical Datatype with Restrictions). *A* numerical datatype with restrictions (NDR) *is a triple* $(\mathcal{D}, O_+, O_-)$*, where* $\mathcal{D} \subseteq \mathbb{R}$ *is a numerical domain and* $O_+, O_- \subseteq \{<, \leq, >, \geq, =\}$ *is the set of* positive *and, respectively,* negative relations. *An axiom in* $\mathcal{EL}^{\bot}(\mathcal{D})$ *is an axiom in* $\mathcal{EL}^{\bot}(\mathcal{D}, O_+, O_-)$ *if for every positive (negative) occurrence of a concept* $\exists F.(s, y)$ *in the axiom,* $s \in O_+$ *(*$s \in O_-$*). An* $\mathcal{EL}^{\bot}(\mathcal{D}, O_+, O_-)$-ontology *is a set of axioms in* $\mathcal{EL}^{\bot}(\mathcal{D}, O_+, O_-)$*.*

We are going to describe a classification procedure for $\mathcal{EL}^{\bot}(\mathcal{D}, O_+, O_-)$, which is closely related to the procedure for $\mathcal{EL}^{++}$ [3]. In order to formulate inference rules for datatypes we need to introduce notation for satisfiability of a datatype restriction and implication between datatype restrictions. For two $\mathcal{D}$-datatype restrictions r_+ and r_-, we write $r_+ \rightarrow_{\mathcal{D}} r_-$ iff $r_+(x)$ implies $r_-(x)$, $\forall x \in \mathcal{D}$. We write $r_+ \rightarrow_{\mathcal{D}} \bot$ iff there is no $x \in \mathcal{D}$ such that $r_+(x)$ holds. In the opposite cases, we write $r_+ \nrightarrow_{\mathcal{D}} r_-$ and $r_+ \nrightarrow_{\mathcal{D}} \bot$. We assume that deciding whether $r_+ \rightarrow_{\mathcal{D}} r_-$ and $r_+ \rightarrow_{\mathcal{D}} \bot$ can be done in polynomial time. It is easy to see that this is the case when $\mathcal{D}$ is the set of natural numbers, integers, reals or rationals for the set of relations $\{<, \leq, >, \geq, =\}$.

The classification procedure for $\mathcal{EL}^{\bot}(\mathcal{D})$ takes as an input an ontology $\mathcal{O}$ whose axioms are in $\mathcal{EL}^{\bot}(\mathcal{D})$ and in normal form and applies the inference rules in Table 2 to derive new axioms of the form NF1, NF3 and NF5. The rules are applied to already derived axioms and use existence of axioms in $\mathcal{O}$, $r_+ \rightarrow_{\mathcal{D}} \bot$ or $r_+ \rightarrow_{\mathcal{D}} r_-$ as side-conditions. The procedure terminates when no new axiom can be derived. It is easily checked that the procedure runs in polynomial time (there are only polynomially many possible axioms of the form NF1, NF3 and NF5 over $sig(\mathcal{O})$) and that the rules in Table 2 are sound (the conclusions of the rules are logical consequences of their premises).

The completeness proof is based on the canonical model construction similarly as for $\mathcal{EL}^{++}$ [3]. In order to deal with datatypes in the canonical model we introduce a notion of a datatype *constraint*. Intuitively, a constraint specifies which datatype restrictions should hold in a given element of the model and which should not.

Definition 2 (Constraint). *A* constraint *over* $(\mathcal{D}, O_+, O_-)$ *is defined as a pair of sets* (S_+, S_-)*, such that* $S_+ = \{(s_+^1, y_1), \ldots, (s_+^n, y_n)\}$ *with* $s_+^i \in O_+$*,* $S_- = \{(s_-^1, z_1), \ldots, (s_-^m, z_m)\}$ *with* $s_-^j \in O_-$*,* y_i*,* $z_j \in \mathcal{D}$*,* $(s_+^i, y_i) \nrightarrow_{\mathcal{D}} (s_-^j, z_j)$

Table 2. Reasoning rules in $\mathcal{EL}^{\bot}(\mathcal{D})$ ($A, B, C, E \in N_C^{\top}$, $C' \in N_C^{\top,\bot}$, $R \in N_R$, $F \in N_F$)

IR1	$\frac{}{A \sqsubseteq A}$		**IR2**	$\frac{}{A \sqsubseteq \top}$	
			CR1	$\frac{A \sqsubseteq B}{A \sqsubseteq C'}$	$B \sqsubseteq C' \in \mathcal{O}$
CR2	$\frac{A \sqsubseteq B \quad A \sqsubseteq C}{A \sqsubseteq D}$	$B \sqcap C \sqsubseteq D \in \mathcal{O}$	**CR3**	$\frac{A \sqsubseteq B}{A \sqsubseteq \exists R.C}$	$B \sqsubseteq \exists R.C \in \mathcal{O}$
CR4	$\frac{A \sqsubseteq \exists R.B \quad B \sqsubseteq C}{A \sqsubseteq D}$	$\exists R.C \sqsubseteq D \in \mathcal{O}$	**CR5**	$\frac{A \sqsubseteq \exists R.B \quad B \sqsubseteq \bot}{A \sqsubseteq \bot}$	
ID1	$\frac{}{A \sqsubseteq \bot}$	$A \sqsubseteq \exists F.r_+ \in \mathcal{O}$, $r_+ \to_{\mathcal{D}} \bot$	**CD1**	$\frac{A \sqsubseteq B}{A \sqsubseteq \exists F.r_+}$	$B \sqsubseteq \exists F.r_+ \in \mathcal{O}$
CD2	$\frac{A \sqsubseteq \exists F.r_+}{A \sqsubseteq B}$	$\exists F.r_- \sqsubseteq B \in \mathcal{O}$, $r_+ \to_{\mathcal{D}} r_-$			

and $(s_+^i, y_i) \not\to_{\mathcal{D}} \bot$ *for* $1 \leq i \leq n$, $1 \leq j \leq m$ *and* m, $n \geq 0$. *A constraint* (S_+, S_-) *over* $(\mathcal{D}, O_+, O_-)$ *is* satisfiable *iff there exists a* solution of (S_+, S_-) *that is a set* $V \subseteq \mathcal{D}$ *such that every* $r_+ \in S_+$ *is satisfied by at least one* $v \in V$ *but no* $r_- \in S_-$ *is satisfied by any* $v \in V$.

Our model construction procedure works only for the cases where we can ensure that every constraint over a numerical domain is satisfiable. This leads us to a notion of safety for an NDR.

Definition 3 (NDR Safety). *Let* $(\mathcal{D}, O_+, O_-)$ *be an NDR.* $(\mathcal{D}, O_+, O_-)$ *is* safe *iff every constraint over* $(\mathcal{D}, O_+, O_-)$ *is satisfiable.*

Definition 4 (Strong and Weak Convexity). *The NDR* $(\mathcal{D}, O_+, O_-)$ *is* strongly convex *when for every* $r_+^i = (s_+^i, y_i)$ *and* $r_-^j = (s_-^j, z_j)$, *with* $s_+^i \in O_+$, $s_-^j \in O_-$ *and* y_i, $z_j \in \mathcal{D}$ $(1 \leq i \leq n,\ 1 \leq j \leq m)$, *if* $\bigwedge_{i=1}^{n} r_+^i \to_{\mathcal{D}} \bigvee_{j=1}^{m} r_-^j$, *then there exists an* r_-^j $(1 \leq j \leq m)$ *such that* $\bigwedge_{i=1}^{n} r_+^i \to_{\mathcal{D}} r_-^j$. $(\mathcal{D}, O_+, O_-)$ *is* weakly convex *when the implication holds for* $n = 1$.

For example the NDR $(\mathbb{Z}, \{<, >\}, \{=\})$ is weakly convex but not strongly convex. It is weakly convex since the implications $((<, y) \to_{\mathbb{Z}} \bigvee_{j=1}^{m}(=, z_j))$ and $((>, y) \to_{\mathbb{Z}} \bigvee_{j=1}^{m}(=, z_j))$ never hold. However, it is not strongly convex: it is $(>, 2) \wedge (<, 5) \to_{\mathbb{Z}} (=, 3) \vee (=, 4)$, but also $(>, 2) \wedge (<, 5) \not\to_{\mathbb{Z}} (=, 3)$ and $(>, 2) \wedge (<, 5) \not\to_{\mathbb{Z}} (=, 4)$.

Lemma 1. $(\mathcal{D}, O_+, O_-)$ *is safe iff it is weakly convex.*

Proof. We assume that $(\mathcal{D}, O_+, O_-)$ is not weakly convex and we prove that it is non-safe. Since it is not weakly convex we have that for some $r_+ \to_{\mathcal{D}} \bigvee_{j=1}^{m} r_-^j$

there exists no r_-^j such that $r_+ \rightarrow_{\mathcal{D}} r_-^j$. We define (S_+, S_-), with $S_+ = \{r_+\}$ and $S_- = \{r_-^j\}_{j=1}^m$ and we prove that (S_+, S_-) is not satisfiable. (S_+, S_-) is indeed a constraint because $r_+ \not\rightarrow_{\mathcal{D}} \bot$ (otherwise $r_+ \rightarrow_{\mathcal{D}} r_-^j$ is true for every r_-^j) and for every r_-^j, $r_+ \not\rightarrow_{\mathcal{D}} r_-^j$ (otherwise $r_+ \rightarrow_{\mathcal{D}} r_-^j$ is true for at least one r_-^j). Additionally, it is not satisfiable, because from $r_+ \rightarrow_{\mathcal{D}} \bigvee_{j=1}^m r_-^j$ there can be found no x such that $r_+(x)$ and $\bigwedge_{j=1}^m \neg r_-^j(x)$.

We prove that if $(\mathcal{D}, O_+, O_-)$ is not safe, then it is not weakly convex. Since it is not safe then there exists a non-satisfiable constraint (S_+, S_-), where $S_+ = \{r_+^i\}_{i=1}^n$ and $S_- = \{r_-^j\}_{j=1}^m$. We have $S_+, S_- \neq \emptyset$ because otherwise a solution for (S_+, S_-) exists. Since (S_+, S_-) is not satisfiable there exists no x for $1 \leq i \leq n$ such that $r_+^i(x)$ and $\bigwedge_{j=1}^m \neg r_-^j(x)$, or otherwise written, $r_+^i \rightarrow_{\mathcal{D}} \bigvee_{j=1}^m r_-^j$. From this and $r_+^i \not\rightarrow_{\mathcal{D}} r_-^j$ (from the constraint definition), $(\mathcal{D}, O_+, O_-)$ is not weakly convex. □

Theorem 1 (Completeness). *Let $(\mathcal{D}, O_+, O_-)$ be a safe NDR, let $\mathcal{O}$ be an $\mathcal{EL}^\bot(\mathcal{D}, O_+, O_-)$-ontology containing axioms in normal form and let $\mathcal{O}'$ be the saturation of $\mathcal{O}$ under the rules of Table 2. For every A, $B \in (N_C^\top \cap sig(\mathcal{O}))$, if $\mathcal{O} \models A \sqsubseteq B$, then $A \sqsubseteq B \in \mathcal{O}'$ or $A \sqsubseteq \bot \in \mathcal{O}'$.*

Proof. The proof is analogous to the completeness proof for the $\mathcal{EL}^{++}$ language [3]; we build a canonical model $\mathcal{I}$ for $\mathcal{O}$ using $\mathcal{O}'$ and show that if $A \not\sqsubseteq B \in \mathcal{O}'$ and $A \not\sqsubseteq \bot \in \mathcal{O}'$ then $\mathcal{I} \not\models A \sqsubseteq B$.
For every $A \in N_C$, $F \in N_F$, define $S_+(A, F)$ and $S_-(A, F)$, as follows:

$$S_+(A, F) = \{r_+ \mid A \sqsubseteq \exists F.r_+ \in \mathcal{O}', A \sqsubseteq \bot \notin \mathcal{O}'\} \quad (3)$$
$$S_-(A, F) = \{r_- \mid \exists F.r_- \sqsubseteq B \in \mathcal{O}, A \sqsubseteq B \notin \mathcal{O}'\} \quad (4)$$

We now show that $(S_+(A, F), S_-(A, F))$ is a constraint over $(\mathcal{D}, O_+, O_-)$. First we prove that $r_+ \not\rightarrow_{\mathcal{D}} \bot$, $\forall r_+ \in S_+(A, F)$, which is true because otherwise due to rule **ID1** it would be $A \sqsubseteq \bot \in \mathcal{O}'$, in contradiction to the definition of $S_+(A, F)$. Additionally, there is no $r_+ \in S_+(A, F)$ and $r_- \in S_-(A, F)$ such that $r_+ \rightarrow_{\mathcal{D}} r_-$, otherwise from $A \sqsubseteq \exists F.r_+ \in \mathcal{O}'$, $\exists F.r_- \sqsubseteq B \in \mathcal{O}$ and **CD2** it would be $A \sqsubseteq B \in \mathcal{O}'$ which contradicts the definition of $S_-(A, F)$. Since $(S_+(A, F), S_-(A, F))$ is a constraint over $(\mathcal{D}, O_+, O_-)$ and $(\mathcal{D}, O_+, O_-)$ is safe, there exists a solution $V(A, F) \subseteq \mathcal{D}$ of $(S_+(A, F), S_-(A, F))$. We now construct the canonical model $\mathcal{I}$:

$$\Delta^\mathcal{I} = \{x_A \mid A \in (N_C^\top \cap sig(\mathcal{O})), A \sqsubseteq \bot \notin \mathcal{O}'\} \quad (5)$$
$$B^\mathcal{I} = \{x_A \mid x_A \in \Delta^\mathcal{I}, A \sqsubseteq B \in \mathcal{O}'\} \quad (6)$$
$$R^\mathcal{I} = \{(x_A, x_B) \mid A \sqsubseteq \exists R.B \in \mathcal{O}', x_A, x_B \in \Delta^\mathcal{I}\} \quad (7)$$
$$F^\mathcal{I} = \{(x_A, v) \mid v \in V(A, F)\} \quad (8)$$

We prove that $\mathcal{I} \models \mathcal{O}$ by showing that $\mathcal{I} \models \alpha$, when α takes one of the NF1-NF6.

NF1 $A \sqsubseteq B$: We need to prove $A^\mathcal{I} \subseteq B^\mathcal{I}$. Take an $x \in A^\mathcal{I}$. By (6), $x = x_C$ such that $C \sqsubseteq A \in \mathcal{O}'$. From $A \sqsubseteq B \in \mathcal{O}$ and since $\mathcal{O}'$ is closed under **CR1**, we have $C \sqsubseteq B \in \mathcal{O}'$. Hence $x = x_C \in B^\mathcal{I}$ by (6).

If $B = \bot$, then we need to show that $A^{\mathcal{I}} = \emptyset$. If there exists $x \in A^{\mathcal{I}}$, then by (6) $x = x_C$ such that $C \sqsubseteq A \in \mathcal{O}'$. Since $\mathcal{O}'$ is closed under **CR1** and $A \sqsubseteq \bot \in \mathcal{O}'$, we have $C \sqsubseteq \bot \in \mathcal{O}'$. Thus, $x = x_C \notin \Delta^{\mathcal{I}}$ by (5), which contradicts our assumption that $x \in A^{\mathcal{I}}$.

We examine separately the case when $A = \top$. We have that $x_A \in \Delta^{\mathcal{I}}$ and we need to show that $x_A \in B^{\mathcal{I}}$. From rule **IR2**, we have that $A \sqsubseteq \top \in \mathcal{O}'$. From rule **CR1**, $A \sqsubseteq B \in \mathcal{O}'$; since $x_A \in \Delta^{\mathcal{I}}$ and $A \sqsubseteq B \in \mathcal{O}'$ we get $x_A \in B^{\mathcal{I}}$ by (6).

NF2 $A_1 \sqcap A_2 \sqsubseteq B$: We prove $(A_1 \sqcap A_2)^{\mathcal{I}} \subseteq B^{\mathcal{I}}$. Take an $x \in (A_1 \sqcap A_2)^{\mathcal{I}}$; then, $x \in A_1^{\mathcal{I}}$, $x \in A_2^{\mathcal{I}}$ and by (6) $x = x_A$ for some concept name A such that $A \sqsubseteq A_1 \in \mathcal{O}'$ and $A \sqsubseteq A_2 \in \mathcal{O}'$. Since $A \sqsubseteq A_1 \in \mathcal{O}'$, $A \sqsubseteq A_2 \in \mathcal{O}'$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{O}$, closure under rule **CR2** gives $A \sqsubseteq B \in \mathcal{O}'$ or $x \in B^{\mathcal{I}}$, by (6).

NF3 $A \sqsubseteq \exists R.B$: We show $A^{\mathcal{I}} \subseteq (\exists R.B)^{\mathcal{I}}$; take an $x \in A^{\mathcal{I}}$. By (6), $x = x_C$ where $C \sqsubseteq A \in \mathcal{O}'$. Since $A \sqsubseteq \exists R.B \in \mathcal{O}$ and $\mathcal{O}'$ is closed under **CR3**, we have $C \sqsubseteq \exists R.B \in \mathcal{O}'$. Since $x_C \in \Delta^{\mathcal{I}}$, we have $C \sqsubseteq \bot \notin \mathcal{O}'$ and, hence, $B \sqsubseteq \bot \notin \mathcal{O}'$ by **CR5**. Thus, $x_B \in \Delta^{\mathcal{I}}$ and $(x_C, x_B) \in R^{\mathcal{I}}$ by (7). Since $B \sqsubseteq B \in \mathcal{O}'$ by **IR1**, we have $x_B \in B^{\mathcal{I}}$ by (6). Thus, $x = x_C \in (\exists R.B)^{\mathcal{I}}$.

NF4 $\exists R.B \sqsubseteq A$: We prove $(\exists R.B)^{\mathcal{I}} \subseteq A^{\mathcal{I}}$; take an $x \in (\exists R.B)^{\mathcal{I}}$. Then, there exists $y \in \Delta^{\mathcal{I}}$ such that $(x, y) \in R^{\mathcal{I}}$ and $y \in B^{\mathcal{I}}$. By (7) and (6) $x = x_C$ and $y = x_D$ such that $C \sqsubseteq \exists R.D \in \mathcal{O}'$ and $D \sqsubseteq B \in \mathcal{O}'$ respectively. Since $\exists R.B \sqsubseteq A \in \mathcal{O}$ and $\mathcal{O}'$ is closed under **CR4**, $C \sqsubseteq A \in \mathcal{O}'$. By (6), $x = x_C \in A^{\mathcal{I}}$.

NF5 $A \sqsubseteq \exists F.r_+$: We show that $A^{\mathcal{I}} \subseteq (\exists F.r_+)^{\mathcal{I}}$; take an $x \in A^{\mathcal{I}}$. By (6), there exists a concept name C such that $x = x_C$ and $C \sqsubseteq A \in \mathcal{O}'$. Since $A \sqsubseteq \exists F.r_+ \in \mathcal{O}$ and $\mathcal{O}'$ is closed under **CD1**, we have $C \sqsubseteq \exists F.r_+ \in \mathcal{O}'$. We use (3) and (4) to build $(S_+(C, F), S_-(C, F))$; we have $r_+ \in S_+(C, F)$. By (8) we have $(x_C, v) \in F^{\mathcal{I}}$ for every $v \in V(C, F)$. Since $r_+ \in S_+(C, F)$, there exists $v \in V(C, F)$ such that v satisfies r_+ and, hence, $x = x_C \in (\exists F.r_+)^{\mathcal{I}}$.

NF6 $\exists F.r_- \sqsubseteq B$: We prove that $(\exists F.r_-)^{\mathcal{I}} \subseteq B^{\mathcal{I}}$; take an $x \in (\exists F.r_-)^{\mathcal{I}}$. By (5), there exists $C \in (N_C^{\top} \cap sig(\mathcal{O}))$ such that $x = x_C$. By (3) and (4) we construct $(S_+(C, F), S_-(C, F))$. Since $x_C \in (\exists F.r_-)^{\mathcal{I}}$, by (8), there exists $v \in V(C, F)$, such that $r_-(v)$ and $V(C, F)$ is a solution for $(S_+(C, F), S_-(C, F))$. Hence, $r_- \notin S_-(C, F)$, and so, $C \sqsubseteq B \in \mathcal{O}'$ by (4). By $C \sqsubseteq B \in \mathcal{O}'$ and (6), $x_C \in B^{\mathcal{I}}$.

We now show that if $A \sqsubseteq B \notin \mathcal{O}'$ and $A \sqsubseteq \bot \notin \mathcal{O}'$, then $\mathcal{O} \not\models A \sqsubseteq B$ by proving $\mathcal{I} \not\models A \sqsubseteq B$ (since $\mathcal{I} \models \mathcal{O}$). $A^{\mathcal{I}} \not\subseteq B^{\mathcal{I}}$ holds, because $x_A \in \Delta^{\mathcal{I}}$ (from $A \sqsubseteq \bot \notin \mathcal{O}'$ and (5)), $x_A \in A^{\mathcal{I}}$ (from $A \sqsubseteq A \in \mathcal{O}'$ using rule **IR1** and by (6)) and $x_A \notin B^{\mathcal{I}}$ (from $A \sqsubseteq B \notin \mathcal{O}'$ and (6)). □

4 Maximal Safe NDRs for ℕ, ℤ, ℝ and ℚ

In this section we present a full classification of safe NDRs for natural numbers ($0 \in \mathbb{N}$), integers, reals and rationals. Table 3 lists all maximal safe NDRs for ℕ, ℤ, ℝ and ℚ. Due to space constraints we present proofs only for the maximal NDRs of natural numbers, that is NDR_1, NDR_2, NDR_9 and NDR_{10}. For these we show that: (i) they are safe (ii) extending any of them leads to non-safety and (iii) every safe NDR w.r.t. ℕ is contained in one of the NDR_1, NDR_2, NDR_9 or NDR_{10}. Table 4 presents some basic transformations between (satisfiable) constraints.

Table 3. Maximal safe NDRs for $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{Q}$ where $\mathcal{D}$ is the domain and O_+ ,O_- is the set of positive and, respectively, negative relations

NDR	$\mathcal{D}$	O_+	O_-
NDR_1	$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$	$\{=\}$	$\{<,\leq,>,\geq,=\}$
NDR_2	$\mathbb{N}$, $\mathbb{Z}$	$\{>,\geq,=\}$	$\{<,\leq,=\}$
NDR_3	$\mathbb{Z}$	$\{<,\leq,=\}$	$\{>,\geq,=\}$
NDR_4	$\mathbb{R}$, $\mathbb{Q}$	$\{<,>,\geq,=\}$	$\{<,\leq,=\}$
NDR_5	$\mathbb{R}$, $\mathbb{Q}$	$\{<,\leq,>,=\}$	$\{>,\geq,=\}$
NDR_6	$\mathbb{Z}$	$\{<,\leq,>,\geq,=\}$	$\{=\}$
NDR_7	$\mathbb{R}$, $\mathbb{Q}$	$\{<,\leq,>,\geq,=\}$	$\{\leq,=\}$
NDR_8	$\mathbb{R}$, $\mathbb{Q}$	$\{<,\leq,>,\geq,=\}$	$\{\geq,=\}$
NDR_9	$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$	$\{<,\leq,>,\geq,=\}$	$\{<,\leq\}$
NDR_{10}	$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$	$\{<,\leq,>,\geq,=\}$	$\{>,\geq\}$

Table 4. Transformations $C_1 \Rightarrow C_2$ preserving constraints and their satisfiability for $\mathbb{N}$, where S_-, S_+ and S are sets of datatype restrictions and $y_1 \leq y_2$, $z_1 \leq z_2$

$C_1 = (S \cup S_+^1, S_-)$, $C_2 = (S \cup S_+^2, S_-)$		$C_1 = (S_+, S \cup S_-^1)$, $C_2 = (S_+, S \cup S_-^2)$	
S_+^1	S_+^2	S_-^1	S_-^2
$\{(<,y)\}$	$\{(\leq,y-1)\}$	$\{(<,z)\}$	$\{(\leq,z-1)\}$
$\{(>,y)\}$	$\{(\geq,y+1)\}$	$\{(>,z)\}$	$\{(\geq,z+1)\}$
$\{(\leq,y_1),(\leq,y_2)\}$	$\{(\leq,y_1)\}$	$\{(\leq,z_1),(\leq,z_2)\}$	$\{(\leq,z_2)\}$
$\{(\geq,y_1),(\geq,y_2)\}$	$\{(\geq,y_2)\}$	$\{(\geq,z_1),(\geq,z_2)\}$	$\{(\geq,z_1)\}$
$\{(=,y_1),(\leq,y_2)\}$	$\{(=,y_1)\}$	$\{(=,z_1),(\leq,z_2)\}$	$\{(\leq,z_2)\}$
$\{(\geq,y_1),(=,y_2)\}$	$\{(=,y_2)\}$	$\{(\geq,z_1),(=,z_2)\}$	$\{(\geq,z_1)\}$
		$\{(<,0)\}$	$\emptyset$

Lemma 2. *Let C_1 and C_2 be as defined in Table 4 and $(\mathbb{N}, O_+, O_-)$ be an NDR. Then (i) C_1 is a constraint over $(\mathbb{N}, O_+, O_-)$ iff C_2 is a constraint over $(\mathbb{N}, O_+, O_-)$ and (ii) if C_1 and C_2 are both constraints over $(\mathbb{N}, O_+, O_-)$, then C_1 is satisfiable iff C_2 is satisfiable.*

Corollary 1. *For $\mathbb{N}$, let NDR_i with $i = 1, 2, 9, 10$. For every $C_1 = (S_+^1, S_-^1)$ over NDR_i there exists a constraint $C_2 = (S_+^2, S_-^2)$ over NDR_i, $y_1, \ldots, y_n \in \mathbb{N}$ and $z_1, \ldots, z_m \in \mathbb{N}$ with m, $n \geq 0$ such that:*

$$S_+^2 \subseteq \{(\leq, y_1), (=, y_2), \ldots, (=, y_{n-1}), (\geq, y_n)\}$$
$$S_-^2 \subseteq \{(\leq, z_1), (=, z_2), \ldots, (=, z_{m-1}), (\geq, z_m)\}$$

where $z_1 < y_1 < \ldots < y_n < z_m$, $z_1 < \ldots < z_m$, $y_i \neq z_j$ ($2 \leq i \leq n-1$, $2 \leq j \leq m-1$) and C_1 over NDR_i is satisfiable iff C_2 over NDR_i is satisfiable.

Lemma 3. NDR_1, NDR_2, NDR_9 *and* NDR_{10} *(all for $\mathbb{N}$) are safe.*

Proof. We prove safety by building a solution V for every (S_+, S_-) over the NDRs; by Corollary 1 we can assume w.l.o.g. the following restrictions:

$\underline{\mathsf{NDR}_1}$: For S_+ we have that $S_+ \subseteq \{(=, y_1), \ldots, (=, y_n)\}$ and for S_- that $S_- \subseteq \{(\leq, z_1), (=, z_2), \ldots, (=, z_{m-1}), (\geq, z_m)\}$ with $z_1 < y_1 < \ldots < y_n < z_m$, $z_1 < \ldots < z_m$ and $y_i \neq z_j$ $(1 \leq i \leq n,\ 2 \leq j \leq m-1)$. $V = \{y_1, \ldots, y_n\}$.

$\underline{\mathsf{NDR}_2}$: For S_+ we have that $S_+ \subseteq \{(=, y_1), \ldots, (=, y_{n-1}), (\geq, y_n)\}$ and for S_- that $S_- \subseteq \{(\leq, z_1), (=, z_2), \ldots, (=, z_m)\}$ with $z_1 < y_1 < \ldots < y_n$, $z_1 < \ldots < z_m$ and $y_i \neq z_j$ $(1 \leq i \leq n-1,\ 2 \leq j \leq m)$. $V = \{y_1, \ldots, y_{n-1}, y'_n\}$, where $y'_n = \max(y_n, z_m) + 1$.

$\underline{\mathsf{NDR}_9}$: For S_+ we have that $S_+ \subseteq \{(\leq, y_1), (=, y_2), \ldots, (=, y_{n-1}), (\geq, y_n)\}$ and for S_- that $S_- \subseteq \{(\leq, z_1)\}$ with $z_1 < y_1 < \ldots < y_n$. $V = \{y_1, \ldots, y_n\}$.

$\underline{\mathsf{NDR}_{10}}$: For S_+ we have that $S_+ \subseteq \{(\leq, y_1), (=, y_2), \ldots, (=, y_{n-1}), (\geq, y_n)\}$ and for S_- that $S_- \subseteq \{(\geq, z_1)\}$ with $y_1 < \ldots < y_n < z_1$. $V = \{y_1, \ldots, y_n\}$. □

Lemma 4. *Let* $\mathsf{NDR} = (\mathbb{N}, O_+, O_-)$. *If (a), (b) or (c), then* NDR *is non-safe.*

(a) $O_+ \cap \{<, \leq, >, \geq\} \neq \emptyset$, $O_- \cap \{<, \leq\} \neq \emptyset$ *and* $O_- \cap \{>, \geq\} \neq \emptyset$.
(b) $O_+ \cap \{>, \geq\} \neq \emptyset$, $O_- \cap \{>, \geq\} \neq \emptyset$ *and* $\{=\} \subseteq O_-$.
(c) $O_+ \cap \{<, \leq\} \neq \emptyset$ *and* $\{=\} \subseteq O_-$.

Proof. For every of the cases (a)-(c) we provide a counterexample that violates the weak convexity condition and, thus by Lemma 1, safety:
(a): $(<, 3) \rightarrow_{\mathbb{N}} (<, 1) \vee (\geq, 1)$ but $(<, 3) \not\rightarrow_{\mathbb{N}} (<, 1)$ and $(<, 3) \not\rightarrow_{\mathbb{N}} (\geq, 1)$. The same counterexample applies when $O_+ \cap \{<, \leq\} \neq \emptyset$, $\{\leq, >\} \subseteq O_-$ and when $O_+ \cap \{<, \leq\} \neq \emptyset$, $\{\leq, \geq\} \subseteq O_-$. For $O_+ \cap \{<, \leq\} \neq \emptyset$, $\{<, >\} \subseteq O_-$ it is $(<, 3) \rightarrow_{\mathbb{N}} (<, 2) \vee (>, 1)$ but $(<, 3) \not\rightarrow_{\mathbb{N}} (<, 2)$ and $(<, 3) \not\rightarrow_{\mathbb{N}} (>, 1)$. A similar example can be given for the the cases when $O_+ \cap \{>, \geq\} \neq \emptyset$.
(b): $(>, 1) \rightarrow_{\mathbb{N}} (=, 2) \vee (\geq, 3)$ but $(>, 1) \not\rightarrow_{\mathbb{N}} (=, 2)$ and $(>, 1) \not\rightarrow_{\mathbb{N}} (\geq, 3)$
$(>, 1) \rightarrow_{\mathbb{N}} (=, 2) \vee (>, 2)$ but $(>, 1) \not\rightarrow_{\mathbb{N}} (=, 2)$ and $(>, 1) \not\rightarrow_{\mathbb{N}} (>, 2)$
$(\geq, 1) \rightarrow_{\mathbb{N}} (=, 1) \vee (\geq, 2)$ but $(\geq, 1) \not\rightarrow_{\mathbb{N}} (=, 1)$ and $(\geq, 1) \not\rightarrow_{\mathbb{N}} (\geq, 2)$
$(\geq, 1) \rightarrow_{\mathbb{N}} (=, 1) \vee (>, 1)$ but $(\geq, 1) \not\rightarrow_{\mathbb{N}} (=, 1)$ and $(\geq, 1) \not\rightarrow_{\mathbb{N}} (>, 1)$
(c): $(<, 3) \rightarrow_{\mathbb{N}} (=, 1) \vee (=, 2)$ but $(<, 3) \not\rightarrow_{\mathbb{N}} (=, 1)$ and $(<, 3) \not\rightarrow_{\mathbb{N}} (= 2)$
$(\leq, 2) \rightarrow_{\mathbb{N}} (=, 1) \vee (=, 2)$ but $(\leq, 2) \not\rightarrow_{\mathbb{N}} (=, 1)$ and $(\leq, 2) \not\rightarrow_{\mathbb{N}} (= 2)$ □

Lemma 5. NDR_1, NDR_2, NDR_9 *and* NDR_{10} *(all for* $\mathbb{N}$*) are maximal safe, that is if any relation is added to* O_+ *or* O_- *they become non-safe.*

Proof. We examine all cases of adding a new relation:
$\underline{\mathsf{NDR}_1}$: If any of the $<$, $\leq$, $>$, $\geq$ is added to O_+, then NDR_1 becomes non-safe due to Lemma 4(a).
$\underline{\mathsf{NDR}_2}$: If $>$ or $\geq$ is added to O_-, then non-safety is due to Lemma 4(b). For adding $<$ or $\leq$ to O_+, non-safety is due to Lemma 4(c).
$\underline{\mathsf{NDR}_9}$: If $>$ or $\geq$ is added to O_-, then non-safety is due to Lemma 4(a). When $=$ is added to O_- then NDR_9 becomes non-safe due to Lemma 4(c).
$\underline{\mathsf{NDR}_{10}}$: If $<$ or $\leq$ is added to O_-, then non-safety is due to Lemma 4(a). When $=$ is added to O_- then NDR_{10} becomes non-safe due to Lemma 4(c). □

It remains to demonstrate that every safe NDR for $\mathbb{N}$ is contained in one of the NDR_1, NDR_2, NDR_9 or NDR_{10}. In the following, we assume that O^i_+ and O^i_- are defined such that $\mathsf{NDR}_i = (\mathbb{N}, O^i_+, O^i_-)$ with $i = 1, 2, 9, 10$.

Lemma 6. *If* $(\mathbb{N}, O_+, O_-)$ *is a safe NDR, then* $O_+ \subseteq O_+^i$ *and* $O_- \subseteq O_-^i$ *for* $i = 1, 2, 9$ *or* 10.

Proof. The proof is by case analysis of possible relations in O_+ and O_-.
Case 1: $O_+ \cap \{<, \leq, >, \geq\} = \emptyset$. In this case, $O_+ \subseteq O_+^1$ and $O_- \subseteq O_-^1$.
Case 2: $O_+ \cap \{<, \leq, >, \geq\} \neq \emptyset$. If $O_- \cap \{<, \leq\} \neq \emptyset$ and $O_- \cap \{>, \geq\} \neq \emptyset$ at the same time, then from Lemma 4(a), the NDR is non-safe. Therefore, we examine two cases: either $O_- \subseteq \{>, \geq, =\}$ or $O_- \subseteq \{<, \leq, =\}$.
Case 2.1: $O_- \subseteq \{>, \geq, =\}$. We distinguish either $O_- \subseteq \{>, \geq\}$ or $\{=\} \subseteq O_-$.
Case 2.1.1: $O_- \subseteq \{>, \geq\} = O_-^{10}$ and $O_+ \subseteq O_+^{10}$.
Case 2.1.2: $\{=\} \subseteq O_-$. By Lemma 4(c) it should be $O_+ \subseteq \{>, \geq, =\} = O_+^2$ otherwise the NDR is non-safe. If $O_- \cap \{>, \geq\} \neq \emptyset$ then the NDR is non-safe by Lemma 4(b); otherwise $O_- = \{=\} \subseteq O_-^2$.
Case 2.2: $O_- \subseteq \{<, \leq, =\} = O_-^2$. If $O_+ \subseteq \{>, \geq, =\}$, then $O_+ \subseteq O_+^2$. Otherwise, $O_+ \cap \{<, \leq\} \neq \emptyset$ and we distinguish whether $O_- \subseteq \{<, \leq\}$ or $\{=\} \in O_-$.
Case 2.2.1: $O_- \subseteq \{<, \leq\} = O_-^9$ and $O_+ \subseteq O_+^9$.
Case 2.2.2: $\{=\} \in O_-$. In this case, the NDR is non-safe by Lemma 4(c). □

For the cases of integers, reals and rationals the proofs are analogous to the case of natural numbers. The interested reader can find details in the technical report [7]. In the following, we provide a brief explanation for the results. We notice two new maximal safe NDRs w.r.t. $\mathbb{Z}$, namely NDR_3 and NDR_6. The reason is that integers do not have a minimal element such as 0 in the case of naturals. In particular positive occurrences of $<$ (or $\leq$) and negative occurrence of $=$ are no longer dangerous (e.g. $(\leq, 1) \not\twoheadrightarrow_{\mathbb{Z}} (=, 1) \vee (=, 0)$ does not hold anymore). Reals and rationals are examples of dense domains: between every two different numbers there always exists a third one. This property is responsible for new safe NDRs. Specifically, O_+ of NDR_2 and NDR_3 can be extended with $<$ and $>$ respectively because the weak convexity property which did not apply for $\mathbb{Z}$ now applies for $\mathbb{R}$ (e.g. $(<, 5) \not\twoheadrightarrow_{\mathbb{R}} (=, 4) \vee (\leq, 3)$). For the same reason, either $\leq$ or $\geq$ can be added to O_- of NDR_6 (e.g. $(\leq, 5) \not\twoheadrightarrow_{\mathbb{R}} (=, 5) \vee (\leq, 4)$).

5 Related Work and Conclusions

Datatypes have been extensively studied in the context of DLs [3, 6, 8]. Extensions of expressive DLs with datatypes have been examined in depth [6] with the main focus on decidability. Baader, Brandt and Lutz [3] formulated tractable extensions of $\mathcal{EL}$ with datatypes using a p-admissibility restriction for datatypes. A datatype $\mathcal{D}$ is p-admissible if (i) satisfiability and implication of conjunctions of datatype restrictions can be decided in polynomial time, and (ii) $\mathcal{D}$ is convex: if a conjunction of datatype restrictions implies a disjunction of datatype restrictions then it also implies one of its disjuncts [3]. In our case instead of condition (i) we require that implication and satisfiability of just datatype restrictions (not conjunctions) is decidable in polynomial time since we do not consider functional features. Condition (ii) is relaxed to the requirement of safety for NDRs since we take into account not only the domain of the datatypes and the types of restrictions but also the polarity of their occurrences. The relaxed restrictions allow for more expressive usage of datatypes in tractable languages, as demonstrated by the example given in the introduction. Furthermore, Baader, Brandt and Lutz did not provide a classification of datatypes that are p-admissible; in our case we provide such a classification for natural numbers, integers, rationals and reals. The EL Profile of OWL 2 [2] is inspired by $\mathcal{EL}^{++}$ and restricts all OWL 2 datatypes to satisfy p-admissibility in such a way that only equality can be used. Our result can allow for a significant extension of datatypes in the OWL 2 EL Profile, where in addition inequalities can be used negatively.

Our work is not the only one where the convexity property is relaxed without losing tractability. It has been shown [8] that the convexity requirement is not necessary provided that (i) the ontology contains only concept definitions of the form $A \equiv C$, where A is a concept name, and (ii) every concept name occurs at most once in the left-hand side of the definition. In some applications this requirement can be too restrictive since it disallows the usage of general concept inclusion axioms (GCIs), such as the axiom (2) given in the introduction, which do not cause any problem in our case.

In this work we made a fine-grained analysis of extensions of $\mathcal{EL}$ with numerical datatypes, focusing not only on the types of relations but also on the polarities of their occurrences in axioms. We made a full classification of cases where these restrictions result in a tractable extension for natural numbers, integers, rationals and reals. One practically relevant case for these datatypes is when positive occurrences of datatype expressions can only use equality and negative occurrences can use any of the numerical relations considered. This case was motivated by an example of a pharmacy-related ontology and can be proposed as a candidate for a future extension of the OWL 2 EL Profile. For the cases where the extension is tractable, we provided a polynomial sound and complete consequence-based reasoning procedure, which can be seen as an extension of the completion-based procedure for $\mathcal{EL}$. We think that the procedure can be straightforwardly extended to accommodate other constructors in $\mathcal{EL}^{++}$ such as (complex) role inclusions, nominals, domain and range restrictions and assertions since these constructors do not interact with datatypes [9]. We hope to investigate these extensions in future works.

In future work we also plan to consider other OWL datatypes, such as strings, binary data or date and time, functional features, and to try to extend the consequence-based procedure for Horn $\mathcal{SHIQ}$ [10] with our rules for datatypes. For example, to extend the procedure with functional features, we probably need a notion of "functional safety" for an NDR that corresponds to the strong convexity property (see Definition 4). In order to achieve even higher expressivity for datatypes we shall study how to combine different restrictions on the datatypes occurring in an ontology so that tractability is preserved. For example, using two safe NDRs in a single ontology may result in intractability, as is the case for NDR_1 and NDR_6 for integers (see Table 3). One possible solution to this problem is to specify explicitly which features can be used with which NDRs in order to separate their usage in ontologies.

References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F.: The Description Logic Handbook: Theory, Implementation, and Applications. 2nd edn. Cambridge University Press (2007)
2. Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P.F., Sattler, U.: OWL 2: The next step for OWL. J. Web Sem. **6**(4) (2008) 309–322
3. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ Envelope. In: IJCAI, Professional Book Center (2005) 364–369
4. Cote, R., Rothwell, D., Palotay, J., Beckett, R., , Brochu, L.: The systematized nomenclature of human and veterinary medicine. Technical report, Northfield, IL: College of American Pathologists (1993)
5. The Gene Ontology Consortium: Gene Ontology: Tool for the unification of biology. Nature Genetics **25** (2000) 25–29
6. Lutz, C.: Description Logics with Concrete Domains-A Survey. In: Advances in Modal Logic, King's College Publications (2002) 265–296
7. Magka, D., Kazakov, Y., Horrocks, I.: Tractable Extensions of the Description Logic $\mathcal{EL}$ with Numerical Datatypes. Technical report, Oxford University Computing Laboratory (2010) posted on `http://web.comlab.ox.ac.uk/isg/people/despoina.magka/publications/reports/NDRTechnicalReport.pdf`.
8. Haase, C., Lutz, C.: Complexity of Subsumption in the $\mathcal{EL}$ Family of Description Logics: Acyclic and Cyclic tboxes. In: ECAI. Volume 178., IOS Press (2008) 25–29
9. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope further. In: In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions. (2008)
10. Kazakov, Y.: Consequence-driven Reasoning for Horn $\mathcal{SHIQ}$ Ontologies. In: IJCAI. (2009) 2040–2045