CENTRO PER LA RICERCA
SCIENTIFICA E TECNOLOGICA

38050 Povo (Trento), Italy
Tel.: +39 0461 314312
Fax: +39 0461 302040
e−mail: prdoc@itc.it − url: http://www.itc.it

DISTRIBUTED DESCRIPTION LOGICS: DIRECTED DOMAIN CORRESPONDENCES IN FEDERATED INFORMATION SOURCES

Borgida A., Serafini L.

September 2002

Technical Report # 0209−07

Distributed Description Logics:
Directed Domain Correspondences in Federated Information Sources

Alex Borgida[1] and Luciano Serafini[2]

[1] Dept. of Computer Science
Rutgers University
New Brunswick, USA
borgida@cs.rutgers.edu
[2] ITC-IRST
Trento, Italy
serafinit@itc.it

Abstract. A central problem of co-operative information systems is the ability to *integrate* information from multiple sources. Although this problem has been studied for several decades, we argue that there is a need for a more refined approach in those cases where the original sources form a loose federation, with each one wishing to maintain its own independent view of the world. In particular, we motivate with examples the utility of directed non-injective mappings between the individuals in the domains of multiple IS. We then extend the logical formalism of Description Logics, which is used in semantic-web ontolgies and has served well in many other aspects of IS design and integration, to handle such mappings. The result is called Distributed Description Logics, and we consider some of its desirable properties, as well as some theorems concerning its computational aspects.

1 Introduction

A significant problem of modern information management is the integration of information from multiple sources. The standard version presumes a framework where users are accessing through a single interface data from several information sources (local ISs), which can include databases, web data, files, etc. The important goal here is making the users unaware of the original source of the information, thus making the interface uniform. This is usually achieved through a global (conceptual) schema that is queried by users. Local ISs are then related to this by a variety of techniques ("local as view", "global as view"), and query answering consists of identifying relevant ISs, translating the user's query into collections of queries over local ISs, and collating the answers.

A somewhat different, but related, approach is one which preserves the identity of each local IS and its user interface. However, the local system wishes to import information available in other sources, which are related to it directly through bilateral assertions. This approach is more appropriate for so-called *federated* information system, where the local ISs maintain a degree of autonomy. Our work is in this second paradigm.

The traditional view of database integration holds that the schemata of local IS may exhibit miss-matches (often called "conflicts" or "semantic heterogeneities") which need to be resolved in order to allow the information from one source to be properly visible in the other source. Saltor and Rodriguez [11] identify three categories of such heterogeneities between ISs at the semantic level:

- Heterogeneities between object classes.
- Heterogeneities between class structures.
- Heterogeneities between object instances.

The first two items can be thought of as differences between the intensions of predicates (corresponding to classes and their attributes), and have been thoroughly studied in the literature. In the third category, one might find uninteresting conflicts that deal with specific values: two IS may record the capital of China as Beijing and Peking respectively. However, in other situations there may be more interesting, systematic relationships. For example, consider the case when one IS contains personal information (e.g., from credit card purchases), while another one contains census information, which only records information about *households*. The correspondence between households and the people in it is not the identity relation, neither is it a simple functional bijection. Yet it will be important to establish this relationship if the two IS are to be integrated.

In the rest of the paper we further motivate through examples the need for more complex correspondences between the domains of multiple IS in a federated system. In Section 3 we introduce Description Logics and Distributed Description Logics (DDL), provides formal semantics for them, and shows how the examples are handled by the result. In Section 4 we review some desirable properties of distributed Information Systems, and discuss how DDL measure up to them. Finally, we provide a theorem which shows how DDL reasoning can often be translated into reasoning in a single, global but ordinary DL.

The focus on Description Logics in this paper is motivated by past successes in applying them to various problems of Information System development, including IS integration (e.g., [1, 4, 3, 9]), as well as their status as leading contender for languages used in expressing semantic web ontologies (e.g., DAML+OIL [5]) in a peer-to-peer environment like the Internet.

2 Motivating Examples

We will be considering the possible correspondences between individuals in the domains of discourse of multiple IS, IS_1 and IS_2 say.

In the simplest case, during integration concepts in the two IS are marked as being identical. Even if one concept, e.g., the $\mathsf{Employees}$ in IS_1, turns out to match only some subset of a concept in the other IS, e.g., permanent $\mathsf{Employees}$

in IS_2, once the respective individuals in IS_2 are selected, they are essentially related by the identity relation to the corresponding ones in IS_1.

Even early research in schema integration recognized that this need not always be the case. First, there are the obvious possible *scaling conflicts*, such as those induced by the use of different units of measure. The obvious solutions to such heterogeneities involves the use of conversion relation such as Celsius_vs_farenheit(x,y). However, complications may arise even in such situations: Consider the case when the unit of measure is currency; in this situation the conversion function from Euro to Dollars, say, is not the inverse of the function from Dollars to Euro, because banks add a surcharge to all transactions. We must therefore acknowledge the need for **directional mappings** between the domains in ISs. A different complication can arise in the case when the mapping between the domains cannot be described simply by enumeration. For example, suppose that one school assigns grades in the integer range of 1 to 10, by taking a score out of 100, dividing by 10, and rounding the result; another school also starts with a score out of 100, and then assigns A to students with 85 or more, B to scores between 70 and 85, C to scores between 60 and 69, D to scores between 50 and 59, F to students with score below 50. In this case, in describing the correspondence between IS_1 and IS_2, we are limited to saying that an 8 corresponds to *either* a B or an A Note however that this partial information is still important: having a grade of "B or A" is known to be better than having a grade of D!

The following examples explore further intricacies of the relationship between domain elements in different IS.

Example 21. *Suppose* BasicC, IntermC *and* AdvancedC *are 3 increasingly difficult courses on some topic. University* $Univ_1$ *offers* BasicC *and* AdvancedC, *while university* $Univ_2$ *offers* IntermC. *The universities are concerned about what classes a student has completed, in order to check the pre-requisites of other courses they are enrolling in, or to meet the degree requirements. In particular, both universities allow a course x to be substituted for another course y, even as a transfer,* if *x is harder than y,* and *covers most of the material of y (say 80%).* $Univ_1$ *may decide to accept* IntermC *as a substitute for* BasicC*; on the other hand,* $Univ_2$ *may only accept* AdvancedC *as a substitute for* IntermC, *since* BasicC *is weaker than* IntermC. *If we treat course objects as individuals, then the above correspond again to cases where we desire directional mappings: e.g.,* IntermC *corresponds to* BasicC *according to* [illegible] *More interestingly, suppose that courses are modeled as concepts whose extension is the set of students who have completed them. Then we have a situation where we want the instances of* IntermC *to be included among the instances of* BasicC *according to* $Univ_1$, *while* IntermC *should subsume the instances of* AdvancedC *according to* $Univ_2$. *Despite all this,* $Univ_1$ *may not want to view* AdvancedC *as a subclass of* BasicC, *since the courses might disagree on more than 80% of the material.*

Example 22. *Suppose* IS_1 *has information about married couples, while* IS_2 *has information about persons. We therefore need to express correspondences* *between individuals in the two domains, e.g., between* couple23 *in* IS_1*, and each of* Gianni *and* Mary*, in* IS_2*. But there are more general relationships between the information in the two IS: For example, we know that couples are made of two people.*

In this case, IS_1 contains information about individuals that are *abstractions* over individuals in IS_2. Similar examples arise in other situations where the so-called "materialization abstraction" [10] occurs.

Example 23. *Consider a situation where there are two IS:* $IS_{Harvard}$ *and* IS_{MIT}*, servicing the needs of college libraries in some town. The libraries have information about* copies of books*, which can be taken by borrowers or are available on the shelf. On the other hand,* $IS_{Student}$ *is a database accessed students, who want to know which library they should go to if they need some book. Notice that the student does not care about which copies of a book are available, so we have once again an abstraction: the student's* Tractatus *corresponds to* TractatusCopy1*,... in* $IS_{Harvard}$*, as well as* TractatusCopy2 *in* IS_{MIT}*. Moreover, the student only wants to know about a material being located at MIT, if there is a copy of it currently on the shelf at the MIT library.*

The above examples establish the need to consider in greater detail the mapping between the domain of objects in the IS being integrated into a federation. First, one needs a *directed general mapping* between the domains of the IS involved, not just a function. Second, there are two aspects of these mappings:

- How are specific individual objects related to each other? (e.g., couple23 in IS_1 and Mary in IS_2).
- What general statements can one make about the mappings of individuals? (e.g., Couple instances in IS_1 correspond to exactly two Person instances in IS_2).

3 Formalization using Description Logics

Logic has proven to be a very powerful tool in understanding Information Systems. In particular, logic appears to be suitable for several tasks:

- Clarifying and making unambiguous the terminology of the data model being introduced. ("What is a bridge rule?")
- Finding errors in the IS designed using a formal data model, by detecting such conditions as classes which cannot have any instances because their specification is incoherent.
- Providing a formal specification of the operations to be supported by the IS (e.g., query answering) so that one prove that the implementation meets the desiderata.

For these reason, we will be using a logic-based data model to investigate the properties of information integration in the presence of complex domain mappings. In particular, we will be using Description Logics, which are object-centered knowledge representation formalisms that have proven to be useful in the design and querying of Information Systems [1].

They are of independent interest, since DL-based systems such as DAML+OIL [5] are the current leading contenders for use in expressing semantic web ontologies, which, given the nature of the Internet, clearly form distributed information sources.

3.1 Description Logics and Information Systems

Description logics view the world as being populated by individuals that can be grouped into classes, called *concepts*, and that can be related to each other by binary relationships, called *roles*. A specific DL provides a specific set of "constructors" for building more complex concepts and roles (much like a programming language type system provides type constructors for building complex types from simpler ones). For example, concept constructors such as conjunction (written as $A \sqcap B$) and value restriction ($\forall r.C$) can be used to describe familiar object classes such as the following necessary conditions on $Students$: $Person \sqcap \forall attends.Univeristy \sqcap \forall age.Integer$.

A typical DL would start with atomic concepts A, as well as constants ANYTHING and NOTHING, denoting the most general concept and the incoherent concept respectively, and then build more complex descriptions according to the syntax in the second column of Figure 1. (An additional concept constructor that we may use, but is not available in $\mathcal{SHIQ}$, is enumeration, which lists the set of instances of the concept: $\{IN_1, \ldots, IN_n\}$.)

Construct name	Syntax	Semantics
primitive concept	A	$A^{\mathcal{I}}$
top concept	ANYTHING x	$\Delta^{\mathcal{I}}$
bottom concept	NOTHING	$\emptyset$
conjunction	$C_1 \sqcap \ldots \sqcap C_n$	$C_1^{\mathcal{I}} \cap \ldots \cap C_n^{\mathcal{I}}$
disjunction	$C_1 \sqcup \ldots \sqcup C_n$	$C_1^{\mathcal{I}} \cup \ldots \cup C_n^{\mathcal{I}}$
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
value restriction	$\forall R.C$	$\{d \in \Delta^{\mathcal{I}} \mid R^{\mathcal{I}}(d) \subseteq C^{\mathcal{I}}\}$
exists restriction	$\exists R.C$	$\{d \in \Delta^{\mathcal{I}} \mid R^{\mathcal{I}}(d) \cap C^{\mathcal{I}} \neq \emptyset\}$
number	$\geqslant n\, R$	$\{d \in \Delta^{\mathcal{I}} \mid \lvert R^{\mathcal{I}}(d)\rvert \geq n\}$
rrestrictions	$\leqslant n\, R$	$\{d \in \Delta^{\mathcal{I}} \mid \lvert R^{\mathcal{I}}(d)\rvert \leq n\}$
qualified number	$\geqslant n\, R.C$	$\{d \in \Delta^{\mathcal{I}} \mid \lvert R^{\mathcal{I}}(d) \cap C^{\mathcal{I}}\rvert \geq n\}$
restrictions	$\leqslant n\, R.C$	$\{d \in \Delta^{\mathcal{I}} \mid \lvert R^{\mathcal{I}}(d) \cap C^{\mathcal{I}}\rvert \leq n\}$
	Role	**Semantics**
primitive role	P	$P^{\mathcal{I}}$
role inverse	R^-	$\{(y, x) \mid (x, y) \in R^{\mathcal{I}}\}$

Fig. 1. Syntax and semantics of the $\mathcal{SHIQ}$ Description Logic [7]

One can then make several kinds of assertions using descriptions. Most familiarly, one can claim that one description, D, **subsumes**/is more general than another one, C, written as $C \sqsubseteq D$. For example, $\mathsf{EMPLOYEE} \sqsubseteq \mathsf{PERSON}$. Subsumption is useful in automatically organizing class descriptions/ontologies into an Is-A hierarchy, and can also be used to detect if a concept C is incoherent, by checking whether $C \sqsubseteq \mathsf{NOTHING}$. Secondly, one can assert the membership of an *individual* in a concept (e.g., $\mathsf{STUDENT} \sqcap \neg\mathsf{MALE}(\mathsf{Anna})$) or the inter-relatedness of two individuals ($\mathsf{attends}(\mathsf{Anna},\mathsf{Harvard})$).

Collections of subsumption assertions specify the terminology used to describe some application domain. Such a collection is called a *T-box*, and resembles the schema of an IS. Collections of assertions about individuals describe some state of world, and form an *A-box*, which resembles a database of facts in an IS. A *(DL) knowledge base* $\mathbf{K}$ will then be a pair $\langle \mathbf{T}, \mathbf{A} \rangle$, where $\mathbf{T}$ is a terminology and $\mathbf{A}$ is an A-box.

Description Logics have been quite successful at capturing the semantics of IS, either directly or as a result of algorithmic mappings from more traditional data models such as the Extended Entity Relationship model. For example [2], starting from a typical ER relation ENROLLMENT, relating STUDENT and COURSE through edges labeled who and what, with a cardinality upper bound of 5 on who, one would get a collection of subsumptions:

1. $\mathsf{ENROLLMENT} \sqsubseteq \forall \mathsf{who}.\mathsf{STUDENT} \sqcap (=1\,\mathsf{who}) \sqcap \forall \mathsf{what}.\mathsf{COURSE} \sqcap (=1\,\mathsf{what})$
(* who *and* what *are functions connecting each enrollment object to a student and a course* *)

2. $\mathsf{STUDENT} \sqsubseteq \forall \mathsf{who}^{-}.\mathsf{ENROLLMENT} \sqcap (\leqslant 5\,\mathsf{who}^{-})$
(* who *only connects enrollments to students, and a student is involved in at most 5 such connections* *)
3. $\mathsf{COURSE} \sqsubseteq \forall \mathsf{what}^{-}.\mathsf{ENROLLMENT}$

Finally, we note that a description logic has a complexity of reasoning (e.g., computing subsumption) that depends on the set of constructors. Much of the effort in DL research has been in identifying different sets of such constructors which have at least decidable inferences, and characterizing their computational complexity.

3.2 Distributed Description Logics

Suppose I is a nonempty set of indexes, and we have a collection of information systems IS_i, for $i \in I$, each described by some (potentially different) description logic $\mathcal{DL}_i$. (The IS_i could be full DL knowledge bases $\mathbf{K}_i$, or just T-boxes $\mathbf{T}_i$.) Let us now try to express connections between them, to form a co-operative IS.

The pioneering work of Catarci and Lenzerini [4] proposed the continued use of description logics. In particular, subsumption-like assertions could relate descriptions in different knowledge bases: $GradStudent_2 \sqsubseteq_{int} Student_1$ would indicate that every graduate student in the part of the world described by IS_2

was also a student in the overlapping part of the world described by IS_1. However, the semantics in [4] implies that inter-schema assertions only have an effect on those *individuals that are shared* between the respective IS domains – i.e., the correspondence between the domain elements is identity. The same assumption seems to underlie other work on federated and heterogeneous databases which relies on Description Logics (e.g., [8]). The reason for this is, in part, the inability of current Description Logics to describe concepts that consist of *new* objects: a new concept definition can only retrieve a subset of the current set of individuals.

In order to deal with our more complex examples, we turn for inspiration to the work of Ghidini and Serafini [6] on Distributed First Order Logic. The idea is to introduce in the semantics binary relations $\mathbf{r}_{ij}$ describing the correspondences between the domains of each IS_i and IS_j, and to use so-called **bridge rules** to constrain these relationships.

In order to support directionality, the bridge rules in a set $\mathfrak{B}_{jk}$ will be viewed as describing "flow of information" from IS_j to IS_k *from the point of view of* IS_k (i.e., IS_k "importing" information from IS_j), and hence $\mathfrak{B}_{jk}$ may be different from $\mathfrak{B}_{kj}$.

Based on studies in [6], here are some types of constraints on the correspondence relationships that one might like to express using bridge rules:

1. Every A-object in IS_1 corresponds only to G-objects in IS_2.
2. All H-objects in IS_2 have a corresponding A-object in IS_1.
3. Each A-object has at least/at most n corresponding object in IS_2.
4. The correspondence relation from IS_1 to IS_2 is the identity relationship.
5. The correspondence relations between IS_1 and IS_2 are converses.

In this paper we will study the first two kinds of bridge rules.

To formalize things, we begin by labeling each description E in $\mathcal{DL}_i$ with its index i (written as $i\!:\!E$). This will help disambiguate descriptions in different $\mathcal{DL}_i$. However, when talking about subsumption within a single IS_i, we will use the more readable $i\!:\!A \sqsubseteq B$, instead of the more formal $i\!:\!A \sqsubseteq i\!:\!B$.

Definition 31. *Given concepts C and E of $\mathcal{DL}_i$ and $\mathcal{DL}_j$ respectively, a* bridge rule *from i to j is an expression of the following two forms:*

$$i\!:\!C \xrightarrow{\sqsubseteq} j\!:\!E \quad \textit{called an } \text{into rule}$$
$$i\!:\!C \xrightarrow{\sqsupseteq} j\!:\!E \quad \textit{called an } \text{onto rule}$$

An into-rule specifies that C-objects in IS_i correspond only to E-objects in IS_j, while an onto-rule states that the only things that correspond to E-objects in IS_2, are C-objects in IS_1.

In Example 2.2,one would have the bridge rule $1\!:\!\mathsf{COUPLE} \xrightarrow{\sqsubseteq} 2\!:\!\mathsf{PERSON}$ to indicate that every couple has corresponding persons, but would not normally include the bridge rule $1\!:\!\mathsf{COUPLE} \xrightarrow{\sqsupseteq} 2\!:\!\mathsf{PERSON}$, because there may be unmarried persons, who may therefore note correspond to any couples.

In order to deal with specific correspondences between individuals, we can follow two approaches. First, if the description logic is sufficiently expressive, we can state such correspondences by using bridge rules. For example, the correspondence of couple23 to Gianni and Mary in Example 2.2 can be expressed by bridge rules $1 : \{\mathsf{couple23}\} \xrightarrow{\sqsubseteq} 2 : \{\mathsf{Gianni,\ Mary}\}$ and $1 : \{\mathsf{couple23}\} \xrightarrow{\sqsupseteq} 2 : \{\mathsf{Gianni, Mary}\}$, if the description logics support concepts formed by enumeration. Otherwise, we need to introduce the individual-level equivalent of bridge-rules.

Definition 32. *If x is an individual in $\mathcal{DL}_i$, while $y, y_1, \ldots$ are individuals of $\mathcal{DL}_j$, then a* (partial) individual correspondence *is an expression $i : x \mapsto j : y$, while a* complete individual correspondence *is an expression $i : x \stackrel{=}{\mapsto} j : \{y_1, y_2, \ldots\}$*

The former indicates that y is a possible translation in the domain j of x in the domain i. For instance $1 : \mathsf{couple23} \mapsto 2 : \mathsf{Gianni}$ and $1 : \mathsf{couple23} \mapsto 2 : \mathsf{Mary}$ state that couple23 in IS_1 can be translated into Gianni and Mary in IS_2. Notice that this assertions do not capture fully the relationship between couple23, Gianni and Mary, because additional objects may still be in correspondence with couple23. Hence the need for complete correspondences

$$1 : \mathsf{couple23} \stackrel{=}{\mapsto} 2 : \{\mathsf{Gianni, Mary}\}$$

We are now ready to define distributed description logics.

Definition 33. *Start with a collection of description logics $\{\mathcal{DL}_i\}_{i \in I}$. A* distributed T-box *(DTB), $\mathfrak{T} = \langle \{\mathbf{T}_i\}_{i \in I}, \mathfrak{B} \rangle$ consists of a set of T-boxes $\{\mathbf{T}_i\}_{i \in I}$, and a set $\mathfrak{B} = \{\mathfrak{B}_{ij}\}$ of bridge rules from i to j for every $i \neq j \in I$. For every $k \in I$, all descriptions in $\mathbf{T}_k$ must be in the corresponding language $\mathcal{DL}_k$, and for every bridge rule $i : A \xrightarrow{\sqsubseteq} j : B$ or $i : A \xrightarrow{\sqsupseteq} j : B$ in $\mathfrak{B}_{ij}$, the concepts A and B must be in the languages $\mathcal{DL}_i$ and $\mathcal{DL}_j$ respectively.*

Definition 34. *A* distributed A-box *(DAB) $\mathfrak{A} = \langle \{\mathbf{A}_i\}_{i \in I}, \mathfrak{C} \rangle$ consists of a set of A-boxes $\{\mathbf{A}_i\}_{i \in I}$, and a set $\mathfrak{C} = \{\mathfrak{C}_{ij}\}$ of partial and complete individual correspondences from i to j for every $i \neq j \in I$. For every $k \in I$, all descriptions in $\mathbf{A}_k$ must be in the corresponding language $\mathcal{DL}_k$, and for every correspondence rule $i : x \mapsto j : y$ or $i : x \stackrel{=}{\mapsto} \{y_1, y_2, \ldots\}$ in $\mathfrak{C}_{ij}$, the individual name x must be in $\mathcal{DL}_i$, and $y, y_1, \ldots$ must be in $\mathcal{DL}_j$.*

Definition 35. *A* distributed DL knowledge base *is then a pair $\langle \mathfrak{T}, \mathfrak{A} \rangle$, consisting of a distributed T-box and a distributed A-box*

Note that when speaking of complete knowledge bases, it might have been more reasonable to have the usual collection of component ISs $\{\mathbf{K}_i\}$, plus information $\langle \mathfrak{B}, \mathfrak{C} \rangle$ about the mapping between concepts and individuals of different *IS*s, in the form of bridge rules and individual correspondences. We have chosen the current definition in order to facilitate factoring out distributed T-boxes, which are the focus of this paper.

3.3 Formal Semantics of Description Logics

We present next a formal semantics for the preceding definitions, which forms the basis of the logic that we desired.

We start with an interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, which assigns subsets of the domain $\Delta^{\mathcal{I}}$ to atomic concepts, and subsets of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ to atomic roles, as well as distinct values of $\Delta^{\mathcal{I}}$ to different named individuals. The interpretation then proceeds recursively, driven by the syntax of complex concept and role constructors, as shown in column three of Figure 1.

The following notation will be used to describe satisfaction and entailment in description logic T-boxes:

$\mathcal{I} \models C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.
$\mathcal{I} \models \mathbf{T}$ iff $\mathcal{I} \models A_i \sqsubseteq B_i$, for all $\mathrm{A}_i \sqsubseteq \mathrm{B}_i$ in $\mathbf{T}$.
$\mathrm{C} \sqsubseteq \mathrm{D}$ iff $\mathcal{I} \models C \sqsubseteq D$ for all possible interpretations $\mathcal{I}$.
$\mathbf{T} \models C \sqsubseteq D$ iff $\mathcal{I} \models C \sqsubseteq D$ for all interpretations $\mathcal{I}$ such that $\mathcal{I} \models \mathbf{T}$.

These definitions are extended to A-boxes according to the rules

$\mathcal{I} \models C(a)$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$
$\mathcal{I} \models r(a, b)$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in p^{\mathcal{I}}$.
$\mathcal{I} \models \mathbf{A}$ iff $\mathcal{I} \models \pi$ for every assertion $\pi = C(a), p(a, b)$ in $\mathbf{A}$.
$\mathbf{K} \models C(a)$ iff $\mathcal{I} \models C(a)$ for all interpretations $\mathcal{I}$ such that $\mathcal{I} \models \mathbf{K}$. Similarly for $p(a, b)$.

We provide semantics for distributed description logics by using local interpretations for the individual information systems, and connecting their domains using relations $\mathbf{r}_{ij}$.

Definition 36. *A* distributed interpretation $\mathfrak{I} = \langle \{\mathcal{I}_i\}_{i \in I}, \mathbf{r} \rangle$ *of* $\mathfrak{T}$ *consists of interpretations* $\mathcal{I}_i$ *for* $\mathcal{DL}_i$ *over domain* $\Delta^{\mathcal{I}_i}$, *and a function* $\mathbf{r}$ *associating to each* $i, j \in I$ *a binary relation* $\mathbf{r}_{ij} \subseteq \Delta^{\mathcal{I}_i} \times \Delta^{\mathcal{I}_j}$. *We use* $\mathbf{r}_{ij}(d)$ *to denote* $\{d' \in \Delta^{\mathcal{I}_j} \mid \langle d, d' \rangle \in \mathbf{r}_{ij}\}$, *and for any* $D \subseteq \Delta^{\mathcal{I}_i}$, *we use* $\mathbf{r}_{ij}(D)$ *to denote* $\bigcup_{d \in D} \mathbf{r}_{ij}(d)$.

Definition 37. *A distributed interpretation* $\mathfrak{I}$ d-satisfies *(written* $\mathfrak{I} \models_d$*) the elements of a DTB* $\mathfrak{T} = \langle \{\mathbf{T}_i\}_{i \in I}, \{\mathfrak{B}_{ij}\} \rangle$ *according to the following clauses: For every* $i, j \in I$

1. $\mathfrak{I} \models_d i : B \xrightarrow{\sqsubseteq} j : G$, *if* $\mathbf{r}_{ij}(B^{\mathcal{I}_i}) \subseteq G^{\mathcal{I}_j}$ *(** Satisfy into-bridge rules. **)*
2. $\mathfrak{I} \models_d i : A \xrightarrow{\sqsupseteq} j : H$, *if* $\mathbf{r}_{ij}(A^{\mathcal{I}_i}) \supseteq H^{\mathcal{I}_j}$ *(** Satisfy onto-bridge rules. **)*
3. $\mathfrak{I} \models_d i : A \sqsubseteq B$, *if* $\mathcal{I}_i \models A \sqsubseteq B$ *(** Satisfy local subsumptions. **)*
4. $\mathfrak{I} \models_d \mathbf{T}_i$ *if* $\mathcal{I}_i \models A \sqsubseteq B$ *for all* $A \sqsubseteq B$ *in* $\mathbf{T}_i$
5. $\mathfrak{I} \models_d \mathfrak{T}$ *if, for every* $i \in I$, $\mathfrak{I} \models_d \mathbf{T}_i$, *and* $\mathfrak{I}$ *d-satisfies every bridge rule in* $\bigcup \mathfrak{B}_{ij}$.

Finally, $\mathfrak{T} \models_d i : C \sqsubseteq D$ *if, for every distributed interpretation* $\mathfrak{I}$, $\mathfrak{I} \models_d \mathfrak{T}$ *implies* $\mathfrak{I} \models_d i : C \sqsubseteq D$.

Concerning individuals, we have the following

Definition 38. *A distributed interpretation* $\mathfrak{I}$ d-satisfies *the elements of a DAB* $\mathfrak{A} = \langle \{\mathbf{A}_i\}_{i \in I}, \{\mathfrak{C}_{ij}\} \rangle$ *according to the following clauses: For every* $i, j \in I$

1. $\mathfrak{I} \models_d i : x \mapsto j : y$, *if* $y^{\mathcal{I}_j} \in \mathbf{r}_{ij}(x^{\mathcal{I}_i})$ *(** Satisfy individual correspondences **)*

2. $\mathfrak{I} \models_d i : x \stackrel{=}{\mapsto} \{y_1, y_2, ...\}$ *if* $\mathbf{r}_{ij}(x^{\mathcal{I}_i}) = \{{y_1}^{\mathcal{I}_j}, {y_2}^{\mathcal{I}_j}, ...\}$ *(** Satisfy complete correspondences **)*
3. $\mathfrak{I} \models_d i : C(a)$, *if* $\mathcal{I}_i \models A(a)$ *(** Satisfy local assertions. **)*
4. $\mathfrak{I} \models_d i : p(a, b)$, *if* $\mathcal{I}_i \models p(a, b)$
5. $\mathfrak{I} \models_d \mathbf{A}_i$ *iff* $\mathfrak{I} \models_d \pi$ *for every assertion* $\pi = C(a), p(a, b)$ *in* $\mathbf{A}_i$.
6. $\mathfrak{I} \models_d \mathfrak{A}$ *if, for every* $i \in I$, $\mathfrak{I} \models_d \mathbf{A}_i$, *and* $\mathfrak{I}$ *d-satisfies every individual correspondence in* $\bigcup \mathfrak{C}_{ij}$.

Finally, $\mathfrak{A} \models_d i : C(a)$ *if, for every distributed interpretation* $\mathfrak{I}$, $\mathfrak{I} \models_d \mathfrak{A}$ *implies* $\mathfrak{I} \models_d i : C(a)$. *Similarly for* $p(a, b)$ *replacing* $C(a)$.

3.4 Some Examples Revisited

We will now recast some of the earlier informal examples into the notation of DDLs, and briefly explore the logical consequences of the resulting theories.

To begin with, the course correspondences in Example 2.1 yield two bridge rules:

$$1 : \mathsf{AdvancedC} \stackrel{\sqsubseteq}{\longrightarrow} 2 : \mathsf{IntermC}$$
$$2 : \mathsf{IntermC} \stackrel{\sqsubseteq}{\longrightarrow} 1 : \mathsf{BasicC}$$

These allow each IS to import appropriate information from the other, but does not entail $1 : \mathsf{AdvancedC} \sqsubseteq \mathsf{BasicC}$, because, among others, the two bridge rules concern mappings in opposite directions.

Consider next Example 2.3, involving libraries. We have three T-boxes $\mathbf{T}_h$, $\mathbf{T}_m$ and $\mathbf{T}_s$. $\mathbf{T}_h$ and $\mathbf{T}_m$ describe the information systems of the libraries. They both have concepts BOOK, corresponding to copies of books that can be loaned, and concepts PERSON, to model the borrowers. The role taken_by is meant to record who has borrowed a book, so that the concept BOOK_ON_SHELF, in $\mathrm{IS}_{Harvard}$, can be defined as follows

$$\mathsf{BOOK\_ON\_SHELF} \equiv \mathsf{BOOK} \sqcap \neg \exists \mathsf{taken\_by}.\mathsf{PERSON}$$

The T-box $\mathbf{T}_s$, modeling the students' information system, also includes a concept called BOOK, but its meaning is different, since, as we mentioned before, these are abstractions over the libraries' copies of the book. The following bridge rules are intended to capture the fact that students see books based on copies from the respective libraries

$$h : \mathsf{BOOK} \stackrel{\sqsubseteq}{\longrightarrow} s : \mathsf{BOOK} \tag{1}$$
$$m : \mathsf{BOOK} \stackrel{\sqsubseteq}{\longrightarrow} s : \mathsf{BOOK} \tag{2}$$

Note that this does not imply that *all* books at the Harvard library can be seen in the students database (the mapping $\mathbf{r}_{hs}$ may be partial). Nor does it imply that the same book copy cannot be mapped to different books in $\mathrm{IS}_{Student}$ – this would require a different kind of bridge rule, one expressing that the correspondence has a certain cardinality.

In addition, $\mathbf{T}_s$ has a role $\mathsf{located\_at}$ to capture the name of the library where the student should go to get the material in question, assuming it is available there. For convenience, $\mathbf{T}_s$ contains a concept, $\mathsf{AVAILABLE\_BOOK}$, that lets students tell quickly if some book is available:

$$\mathsf{AVAILABLE\_BOOK} \equiv \mathsf{BOOK} \sqcap \exists \mathsf{located\_at}$$

Since students only want to hear about material located at a library if there are some copies of it available there, we use onto bridge rules as follows[3]:

$$h : \mathsf{BOOK} \sqcap \neg\exists \mathsf{taken\_by.PERSON} \xrightarrow{\sqsupseteq} s : \exists \mathsf{located\_at}.\{\text{"Harvard"}\} \tag{3}$$

$$m : \mathsf{BOOK} \sqcap \neg\exists \mathsf{taken\_by.PERSON} \xrightarrow{\sqsupseteq} s : \exists \mathsf{located\_at}.\{\text{"Mit"}\} \tag{4}$$

Note that in the absence of additional bridge rules, no information should leave $\mathbf{T}_s$, since the bridge rules given so far act in one direction only.

Ignoring one of the libraries, $\mathbf{T}_m$ say, in order to simplify matters, we can then define the distributed T-box $\mathfrak{T}_{lib} = \langle \mathbf{T}_h, \mathbf{T}_s, \mathfrak{B}_{hs} \rangle$, where $\mathfrak{B}_{hs}$ contains bridge rules 1 and 3. Figure 2 provides an example distributed interpretation $\mathfrak{I}_{lib}$ for $\mathfrak{T}_{lib}$

$$
\begin{aligned}
\Delta^{\mathcal{I}_h} &= \{\mathsf{Tractatus(1)}, \mathsf{Tractatus(2)}, \mathsf{DB\_Pples}, \mathsf{Mario}\} \\
\mathsf{BOOK}^{\mathcal{I}_h} &= \{\mathsf{Tractatus(1)}, \mathsf{Tractatus(2)}, \mathsf{DB\_Pples}\} \\
\mathsf{PERSON}^{\mathcal{I}_h} &= \{\mathsf{Mario}\} \\
\mathsf{taken\_by}^{\mathcal{I}_h} &= \{\langle \mathsf{Tractatus(1)}, \mathsf{Mario} \rangle\} \\
\Delta^{\mathcal{I}_s} &= \{\mathsf{Tractatus}, \mathsf{Philosophical\_Investigations}, \text{"Harvard"}, \text{"Mit"}\} \\
\mathsf{BOOK}^{\mathcal{I}_s} &= \{\mathsf{Tractatus}, \mathsf{Philosophical\_Investigations}\} \\
\mathsf{located\_at}^{\mathcal{I}_s} &= \left\{ \begin{array}{c} \langle \mathsf{Tractatus}, \text{"Harvard"} \rangle \\ \langle \mathsf{Philosophical\_Investigations}, \text{"Mit"} \rangle \end{array} \right\} \\
\mathbf{r}_{hs} &= \left\{ \begin{array}{c} \langle \mathsf{Tractatus(1)}, \mathsf{Tractatus} \rangle \\ \langle \mathsf{Tractatus(2)}, \mathsf{Tractatus} \rangle \end{array} \right\}
\end{aligned}
$$

Fig. 2. Example of distributed interpretation for $\mathfrak{T}_{lib}$

$\mathfrak{I}_{lib}$ satisfies bridge rule (1); indeed, $\mathbf{r}_{hs}(\mathsf{BOOK}^{\mathcal{I}_h}) = \{\mathsf{Tractatus}\} \subseteq \mathsf{BOOK}^{\mathcal{I}_s} = \{\mathsf{Tractatus}, \mathsf{Philosophical\_Investigations}\}$;

$\mathfrak{I}_{lib}$ also satisfies bridge rule (3); indeed, $\mathbf{r}_{hs}(\mathsf{BOOK\_ON\_SHELF}^{\mathcal{I}_h}) = \mathbf{r}_{hs}(\{\mathsf{Tractatus(2)}\}) = \{\mathsf{Tractatus}\}$, which is a superset of $(\exists \mathsf{located\_at}.\{\text{"Harvard"}\})^{\mathcal{I}_s} = \{\mathsf{Tractatus}\}$.

[3] Technical note: although we use enumerated concepts, such as {"Harvard"}, which are not in $\mathcal{SHIQ}$, their elements are string constants, which do not have properties of their own. It is known that such constructs can be eliminated by using mutually exclusive primitive concepts

Note also that the bridge rule $h:\mathsf{BOOK} \xrightarrow{\sqsubseteq} s:\mathsf{BOOK}$ is satisfied even though $\mathsf{BOOK}^{\mathcal{I}_h}$ is not contained in $\mathsf{BOOK}^{\mathcal{I}_s}$. Finally, one of the logical consequences (d-entailments) of $\mathfrak{T}_{lib}$ is

$$s:\exists \mathsf{located\_at}.\{\text{"Harvard"}\} \sqsubseteq \mathsf{AVAILABLE\_BOOK}$$

This follows because the bridge rules allow us to infer in IS_s that anything that is located at the Harvard library must be a book, and hence an instance of the concept AVAILABLE_BOOK defined earlier.

The above example exhibits in fact a common pattern of inference in our DDL:

starting from
- A subsumes B in IS_1
- A is mapped **into** G by a bridge rule
- B is mapped **onto** H by a bridge rule

conclude that
- G subsumes H in IS_2

The following is another example of this inference: suppose $\mathbf{T}_1$ contains the subsumption assertion

$$\mathsf{Villa} \sqsubseteq \mathsf{SecondResidence}$$

and there are bridge rules

$$1:\mathsf{SecondResidence} \xrightarrow{\sqsubseteq} 2:\mathsf{Dwelling}$$
$$1:\mathsf{Villa} \xrightarrow{\sqsupseteq} 2:\mathsf{Cottage}$$

One can then conclude that $2:\mathsf{Cottage} \sqsubseteq \mathsf{Dwelling}$. The intuition behind this inference is depicted in Figure 3.4

4 Some properties of DDL

In this section, we will restrict our attention to the simplest kinds of DDL, namely distributed T-boxes involving only two IS, and a single set of bridge rules between them: $\mathfrak{T}_{12} = \langle \mathbf{T}_1, \mathbf{T}_2, \mathfrak{B}_{12} \rangle$.

The following is a list of intuitively desirable properties for such a system, based on our motivations:

1. When deducing things at IS_i in the distributed system, all local information should be available. (Technically, if $\mathbf{T}_i \models i:X \sqsubseteq Y$, then $\mathfrak{T}_{12} \models_d i:X \sqsubseteq Y$, for $i = 1, 2$.)
2. In the absence of bridge rules, no information should pass between the component systems. (In other words: $\langle \mathbf{T}_1, \mathbf{T}_2, \emptyset \rangle \models_d i:X \sqsubseteq Y$ if and only if $\mathbf{T}_i \models i:X \sqsubseteq Y$, for $i = 1, 2$.)

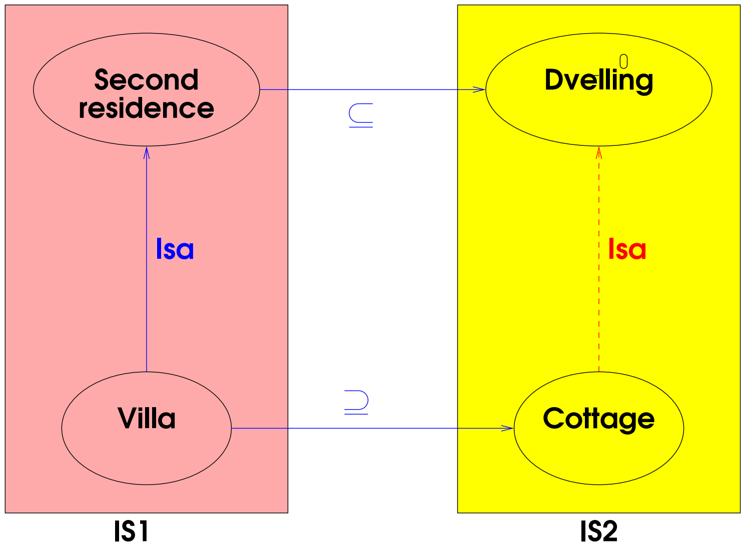

Fig. 3. Following the solid lines (containment statements) we can infer the dashed line which is a subsumption statement in the second Tbox

3. A DDL should exhibit "directionality": we have said that $\mathfrak{B}_{12}$ contains bridge rules that are set up to provide information flow from IS_1 to IS_2. Therefore, such a set up should not affect, by itself, reasoning in IS_1. Formally, we would like to have $\mathfrak{T}_{12} \models_d 1 : A \sqsubseteq B$ iff $\mathbf{T}_1 \models A \sqsubseteq B$. This would also allow for more effective reasoning because there would be no need for a feedback loop between new inferences in IS_2 and those in IS_1.
4. A distributed system should not allow local inconsistencies to "pollute" the entire system, in the sense that if the information at IS_i is not satisfiable, then deductions at other sites should not be affected. Formally, this would mean that if $\mathbf{T}_i$ is inconsistent then $\mathfrak{T}_{12} \models_d j : X \sqsubseteq Y$ iff $\mathbf{T}_j \models X \sqsubseteq Y$ for $j = 1, 2$.

It can be easily seen that our definition of DDL does have the first two properties, if the component T-boxes are consistent.

As far as "backflow" is concerned, it is unfortunately possible to use onto bridge rules to require certain properties of descriptions in IS_1. In particular, a rule such as $1 : A \xrightarrow{\sqsupseteq} 2 : \mathsf{ANYTHING}$ requires every distributed interpretation to have the property $A^{\mathcal{I}_1} \neq \emptyset$, because the extension of the ANYTHING concept is never empty, and hence there must always be individuals in $A^{\mathcal{I}_1}$ that correspond to it.

To express this informal reasoning using subsumption statements, let us introduce a new role, $\mathsf{UNIVERSAL}_{\mathsf{ROLE}}$. It relates every possible pair of objects ($\mathsf{UNIVERSAL}_{\mathsf{ROLE}}{}^{\mathcal{I}_1} = \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_1}$). Then the DTB $\mathfrak{T}$', with empty T-boxes and one bridge rule $\{1 : A \xrightarrow{\sqsupseteq} 2 : \mathsf{ANYTHING}\}$, d-entails the the formula

$1{:}\mathsf{ANYTHING} \sqsubseteq \exists\mathsf{UNIVERSAL_{ROLE}}.A$ because, if A's interpretation is never empty, than every object must be related by $\mathsf{UNIVERSAL_{ROLE}}$ to some object in A.

Interestingly, for some practical DLs such problems do not actually arise, as stated in the following result:

Proposition 41 (no "back-flow" for $\mathcal{SHIQ}$). *Let $\mathfrak{T}_{12}$ be a consistent DTB $\langle \mathbf{T}_1, \mathbf{T}_2, \mathfrak{B}_{12}\rangle$, where $\mathbf{T}_1$ and $\mathbf{T}_2$ contain only subsumptions from the DL $\mathcal{SHIQ}$. Then $\mathfrak{T}_{12} \models_d 1{:}A \sqsubseteq B$ iff $\mathbf{T}_1 \models A \sqsubseteq B$.*

As far as inconsistency propagation, unfortunately if one of the component T-boxes is unsatisfiable then there is no interpretation satisfying the entire DTB, and therefore every possible assertion is automatically d-entailed by an unsatisfiable DTB. We must however point out that this kind of "pollution propagation" is not peculiar to DDL – it is a feature of most formal models of distributed IS expressed in DLs, and, in fact, of most logic-based approaches.

Looking at the issue more closely in the context of DTBs, if $\mathbf{T}_2$ is unsatisfiable, then anything can be derived concerning IS_1, which is particularly disturbing given our stance on "no backflow". Even in the case when only $\mathbf{T}_1$ is unsatisfiable, there is still some desire for this inconsistency not to "infect" the reasoning of IS_2, at least not to the point that *all* conclusions of the form $2{:}X \sqsubseteq Y$, for every X and Y are d-entailed by $\mathfrak{T}_{12}$. Note that there other possible reasons for the unsatisfiability of a DTB: (i) some bridge rules are not satisfiable (e.g., $1{:}\mathsf{NOTHING} \xrightarrow{\sqsupseteq} 2{:}\mathsf{ANYTHING}$); (ii) though each component on its own is satisfiable, it is the combination that is not: $\left\langle \{A \sqsubseteq \mathsf{NOTHING}\}, \{\mathsf{ANYTHING} \sqsubseteq B\}, 1{:}A \xrightarrow{\sqsupseteq} 2{:}B \right\rangle$.

One can try to modify this aspect of a logical symbolism by using a modified semantics. The original work on Distributed First Order Logic [6] used sets of ordinary interpretations to achieve this effect. We have investigated our own simpler version of such a semantics: introducing a new, special interpretation for ordinary DL, $\mathcal{I}^\delta = \left\langle \Delta^{\mathcal{I}^\delta}, .^{\mathcal{I}^\delta} \right\rangle$. $\Delta^{\mathcal{I}^\delta}$ is any non empty set, and $.^{\mathcal{I}^\delta}$ makes the denotation of every description be the whole domain $\Delta^{\mathcal{I}^\delta}$. Intuitively $\mathcal{I}^\delta$ offers an interpretation to locally inconsistent T-boxes; indeed, in $\mathcal{I}^\delta$, every subsumption $A \sqsubseteq B$ is satisfied, including $\mathsf{ANYTHING} \sqsubseteq \mathsf{NOTHING}$. Redefining d-entailment to allow §0 as one of the $\mathcal{I}_i$, produces a DDL which does limit the effect of inconsistent T-boxes in the desired way.

5 Relating DDL and ordinary DL

The following results provide in certain cases a connection between DDL and ordinary DLs, supplying, among others, algorithms for reasoning in the DDL.

Definition 51. *Given a family of description logics $\{\mathcal{DL}_i\}_{i\in I}$ the global description logic $\mathcal{GDL}$ is defined as follows:*

1. *The primitive concepts of GL consist of $i{:}A$ for all primitive concepts or constant concepts (such as* ANYTHING*) A of $\mathcal{DL}_i$, for every $i \in I$.*

2. *the primitive roles of GL include $i\!:\!p$ for all primitive or constant roles p of $\mathcal{DL}_i$, $i \in I$.*
3. *furthermore, for any $i, j \in I$ and $i \neq j$, R_{ij} is a new role of GL (intuitively, representing correspondences between domains).*
4. *There are new top and bottom concepts, $\mathsf{ANYTHING}_g$ and $\mathsf{NOTHING}_g$ respectively.*

Now define a mapping $\#()$ from concepts/roles of $\mathcal{DL}_i$ to $\mathcal{GDL}$, which starts from the primitive concepts, roles and individuals M of the corresponding IS_i: $\#(i\!:\!M) = i\!:\!M$. For complex concepts and roles, the function is defined by structural recursion: if ρ is a concept constructor taking k arguments, then $\#(i\!:\!\rho(M_1, \ldots, M_k)) = i\!:\!\mathsf{ANYTHING} \sqcap \rho(\#(M_1), \ldots, \#(M_k))$. (The intersection with $i\!:\!\mathsf{ANYTHING}$ is needed in order to limit the effect of complement.) For example, the concept $\forall p.C$ has constructor $\forall$, and two arguments, the role and the restriction; therefore $\#(\, i\!:\!\forall p.C\,) = i\!:\!\mathsf{ANYTHING} \sqcap \forall(i\!:\!p).(i\!:\!\mathsf{ANYTHING} \sqcap i\!:\!C)$.

Applying $\#()$ to a DTB $\mathfrak{T} = \langle \{\mathbf{T}_i\}_{i \in I}, \mathfrak{B}\rangle$, yields a T-box $\#(\mathfrak{T})$ in the language $\mathcal{GDL}$, consisting of the following:

1. (* *Copies of axioms from local T-boxes.* *)
   $\#(i\!:\!A) \sqsubseteq \#(i\!:\!B)$ for all $i\!:\!A \sqsubseteq B \ \in \mathbf{T}_i$;
2. (* *Translations of into bridge rules as value restrictions on* R_{ij}. *)
   $\#(i\!:\!A) \sqsubseteq \forall \mathrm{R}_{ij}.\#(j\!:\!G)$ for every into bridge rule $i\!:\!A \xrightarrow{\sqsubseteq} j\!:\!G \in \mathfrak{B}$;
3. (* *Translations of onto bridge rules as existential restrictions on the inverse of* R_{ij}. *)
   $\#(j\!:\!H) \sqsubseteq \exists \mathrm{R}_{ij}^{-}.\#(i\!:\!A)$ for every onto bridge rule $i\!:\!A \xrightarrow{\sqsupseteq} j\!:\!G \in \mathfrak{B}$;
4. (* *Restrictions on role* R_{ij} *to ensure that it connects only objects in* $\mathbf{T}_i$ *and* $\mathbf{T}_j$. *)
   $\mathsf{ANYTHING}_g \sqsubseteq \forall \mathrm{R}_{ij}.j\!:\!\mathsf{ANYTHING}$ (* *the range of* R_{ij} *is* $\Delta^{\mathcal{I}_j}$ *)
   $\neg(i\!:\!\mathsf{ANYTHING}) \sqsubseteq \forall \mathrm{R}_{ij}.\mathsf{NOTHING}_g$ (* R_{ij} *is undefined outside* $\Delta^{\mathcal{I}_i}$ *)
5. (* *Specifying that all* $i\!:\!\mathsf{NOTHING}$ *are the incoherent concept.* *)
   $i\!:\!\mathsf{NOTHING} \sqsubseteq \mathsf{NOTHING}_g$
6. (* *Making sure that the* $i\!:\!\mathsf{ANYTHING}$ *acts as the proper local top of IS-A hierarchies* *)
   $i\!:\!A \sqsubseteq i\!:\!\mathsf{ANYTHING}$, for every atomic concept A of $\mathcal{DL}_i$;
7. (* *Making sure that every i-role has as domain and range* $i\!:\!\mathsf{ANYTHING}$ *)
   $i\!:\!\mathsf{ANYTHING} \sqsubseteq \forall(i\!:\!s).(i\!:\!\mathsf{ANYTHING})$ for every role s of $\mathcal{DL}_i$ (* *the range of* $i\!:\!s$ *is in* $\Delta^{\mathcal{I}_i}$ *);
   $\neg(i\!:\!\mathsf{ANYTHING}) \sqsubseteq \forall i\!:\!s.\mathsf{NOTHING}_g$ (* $i\!:\!s$ *is undefined outside* $\Delta^{\mathcal{I}_i}$ *)

The following theorem states that d-entailment can often be reduced to ordinary DL-reasoning through the use of the above translation:

Theorem 52. *Suppose $\mathfrak{T}$ is a DTB, where none of the $\mathcal{DL}_i$ use role constants or role constructors. Then $\#(\mathfrak{T}) \models \#(i\!:\!X) \sqsubseteq \#(i\!:\!Y)$ if and only if $\mathfrak{T} \models_d i\!:\!X \sqsubseteq Y$.*

The main obstacle in generalizing the above theorem for the case when some $\mathcal{DL}_i$ has role constants or constructors is ensuring that any interpretation $\mathcal{I}*$

of every composite role in $\#(\mathfrak{T})$ has as domain and range the interpretation of i:ANYTHING. This works by induction for role constructors such as conjunction, disjunction and inverse. However, it is no longer the case for role complement, or constants such as universal roles. In this case, the description logic in which $\#(\mathfrak{T})$ is expressed may need to have additional, possibly more powerful, constructors such as roles defined as the product of domain and range concepts.

Using Theorem 5.2, we can obtain reasoners for a variety of DDL.

Proposition 53. *A DDL such that all $\mathcal{DL}_i$ are contained in some decidable description logic $\mathcal{DL}_0$ which supports (i) qualified existential restriction, and (ii) arbitrary subsumption assertions in T-boxes, can use the decision procedure of $\mathcal{DL}_0$ to decide unsatisfiability and d-entailment.*

This result applies even when the $\mathcal{DL}$ involved have role constructors such as conjunction, disjunction, inverse, composition, role hierarchies, and transitive roles. We get as corollaries that DDLs with $\mathcal{DL}_i$ that are in $\mathcal{SHIQ}$ [7] can use their reasoners for determining $\models_d$.

We can extend the above translation to deal with A-boxes and individual correspondences.

1. (* *Copies of local assertions* *)
   $\#(i:C)(i:a)$ for $C(a)$ in $\mathbf{A}_i$
   $\#(i:p)(i:a, i:b)$ for $p(a,b)$ in $\mathbf{A}_i$
2. (* *Translation of correspondences* *)
   $\mathrm{R}_{ij}(i:x, j:y)$ for every correspondence $i:x \mapsto j:y$ in $\mathfrak{C}$.
3. (* *Translation for complete correspondences* *)
   $\mathrm{R}_{ij}(i:x, j:y_1), \ldots,\ \mathrm{R}_{ij}(i:x, j:y_n),\ (\leqslant n\,\mathrm{R}_{ij})(i:x)$ for every complete correspondence $i:x \overset{=}{\mapsto} j:\{y_1, \ldots, y_n\}$ in $\mathfrak{C}$.

The key novelty is the translation of $\overset{=}{\mapsto}$, which places an upper bound on the number of values that can correspond to an individual. As a result, the DL into which translation occurs must also support number restrictions. Fortunately, $\mathcal{SHIQ}$ does have number restrictions.

We remark that the full expressive power of reasoners such as $\mathcal{SHIQ}$ may not necessarily be required in order to accommodate such translations (there are no nested existential quantifiers) so that DDL with very simple $\mathcal{DL}_i$ may have reasoners with lower complexity. Also, note that although Theorem 5.2 makes it seem like all $\mathbf{T}_i$ should come from the same $\mathcal{DL}$, there may be alternate ways to proceed, which make it interesting to study the use of different kinds of Description Logics, of different expressive power, in the different component T-boxes of the integrated IS.

The above results can also help focus our search for additional kinds of bridge rules to add to DDL, by trying to preserve Theorem 5.2. This means that the new bridge rules should be expressible as DL descriptions using the special roles R_{ij}. For example, each of the following kinds of bridge rules has the corresponding translation shown beside it:

- Every A-object has at least n corresponding G-object: $A \sqsubseteq \geqslant n\, \mathrm{R}_{12}.G$.
- The correspondence relation from IS_1 to IS_2 is the identity: $\mathrm{R}_{12} = \mathsf{IDENT}_{\mathsf{ROLE}}$.
- The correspondence relation is "injective" in the sense that different objects map to distinct images: $\mathsf{ANYTHING}_i \sqsubseteq \forall \mathrm{R}_{12}.(\leqslant 1\, \mathrm{R}_{12}^{-})$.
- The correspondence relations between IS_1 and IS_2 are converses, when directionality is not involved (e.g., scaling functions): $\mathrm{R}_{12} = \mathrm{R}_{21}^{-}$.

6 Summary

The seminal work of Catarci and Lenzerini [4] on integrating ISs described by DLs, made an implicit assumption that the local ISs have the same notion of what individual objects are, and that there was only one set of (subsumption) assertions relating IS_1 and IS_2. We have argued that in cases such as federated IS, when there is no single global view, these conditions need to be relaxed, by allowing general correspondence relationships between objects in the local domains, and by having "directed" import assertions. These intuitions were formalized in DDL using the notion of bridge rules and individual correspondences. We note that directed integration rules are also provided in [3], but these are general Horn logic clauses, for which no reasoning is supported.

We have identified other desirable properties of federated DL, such as "no feedback", and localizaling the effect of inconsistencies so that one IS does not "infect" the reasoning of the entire system.

Among the interesting results obtained are a translation of DDL reasoning to DL reasoning, which requires only qualified existential restrictions and general theories, thus providing algorithms for reasoning in DDLs where all local ISs have $\mathcal{ALCNR}$ or $\mathcal{SHIQ}$ T-boxes.

References

1. Alexander Borgida. Description logics in data management. *IEEE Trans. on Knowledge and Data Engineering*, 7(5):671–682, 1995.
2. D. Calvanese, M. Lenzerini, and D. Nardi. Unifying class-based representation formalisms. *J. of Artificial Intelligence Research*, 11:199–240, 1999.
3. D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. A principled approach to data integration and reconciliation in data warehousing. In *Proc. Intl. Workshop on Design and Management of Data Warehouses (DMDW'99), Heidelberg, Germany*, page 16, 1999.
4. T. Catarci and M. Lenzerini. Representing and using interschema knowledge in cooperative information systems. *International Journal of Intelligent and Cooperative Information Systems*, 2(4):375–398, 1993.
5. DAML+OIL Web Ontology Language, W3C Submission, September 2001, `http://www.w3.org/Submission/2001/12/`
6. C. Ghidini and L. Serafini. Distributed First Order Logics. In D. Gabbay and M. de Rijke, editors, *Frontiers Of Combining Systems 2*, Studies in Logic and Computation, pages 121–140. Research Studies Press, 1998.

7. I. Horrocks and U. Sattler. A description logic with transitive and inverse roles and role hierarchies. *J. Logic and Computation*, 9(3):385–410, 1999.
8. V. Kashyap and A. Sheth. Semantic and Schematic Similarities Between Database Objects: A Context-Based Approach. VLDB Journal 5(4): 276-304, 1996.
9. E. Mena, V. Kashyap, A. Illarramendi and A. Sheth. Imprecise Answers on Highly Open and Distributed Environments: An Approach based on Information Loss for Multi-Ontology Based Query Processing. International Journal of Cooperative Information Systems (IJCIS), special issue "Intelligent Integration of Information", 9(4):403-425, December 2000.
10. A. Pirotte, E. Zimanyi, D. Massart, and T. Yakusheva. Materialization: A Powerful and Ubiquitous Abstraction Pattern. In *Proc. Conf. on Very Large Database (VLDB'94)*, pages pp. 630–641, Morgan Kauffman, 1994.
11. F. Saltor and E. Rodríguez. On Intelligent Access to Heterogeneous Information. In F. Baader, M.A. Jeusfeld & W. Nutt (eds), *Proceedings of the 4th Workshop KRDB-97*, Athens, Greece, August 30, 1997, pp. 15:1-15:7