A Formal Investigation of Semantic Interoperability of HCLS Systems

Ratnesh Sahay, Antoine Zimmermann, Ronan Fox,
Axel Polleres, Manfred Hauswirth

Digital Enterprise Research Institute
National University of Ireland, Galway, Ireland
`firstname.lastname@deri.org`

Abstract. Semantic interoperability facilitates Health Care and Life Sciences (HCLS) systems in connecting stakeholders (e.g., patient, physician, pharmacy) at various levels as well as ensure seamless use of healthcare resources (e.g., data, schema, applications). Their scope ranges from local (within, e.g., hospitals or hospital networks) to regional, national and cross-border. The use of semantics in delivering interoperable solution for HCLS systems is weakened by fact that an Ontology Based Information System (OBIS) has restrictions in modeling, aggregating, and interpreting global knowledge (e.g., terminologies for disease, drug, clinical event) in conjunction with local information (e.g., policy, profiles). This chapter presents an example-scenario that shows such limitations and recognizes that enabling two key features, namely the *type* and *scope* of knowledge, within a knowledge base could enhance the overall effectiveness of an OBIS. We provide the idea of separating knowledge bases in *types* (e.g., general or constraint knowledge) with scope (e.g., global or local) of applicability. Then, we propose two concrete solutions on this general notion. Finally, we describe open research issues that may be of interest to knowledge system developers and broader research community.

1   Introduction

Nowadays, Health Care and Life Sciences (HCLS) systems are facing challenges to bring healthcare stakeholders together, such that healthcare resources (e.g., data, schema, and applications) are seamlessly accessed across all related domains. Knowledge and HCLS specialists have advocated the use of semantics to (1) create interlinked networks of HCLS resources and (2) overall management and integration of HCLS resources. Ontology Based Information Systems (OBISs) have recently received attention for their flexibility and automation in maintaining, updating, exchanging schemas and underlying data. This strength is due to the fact that schema and data are loosely-coupled (as compared to cohesively-weak schema and data in traditional databases) and both represented and integrated in a logical fashion. However, OBIS in its current form has limitations in dealing with the requirements imposed by heterogeneous information

systems. One major limitation is its inability to distinguish between various types of knowledge used or exchanged between information systems. Additionally, knowledge expressed within any typical information system are generally tied to a context (*e.g.*, place, event, time) and their appropriate interpretation is scoped or limited to that context.

Semantics and the reasoning mechanism behind ontological knowledge bases are centralized in the way data and schemas are accumulated and processed. Therefore, when an OBIS experience the use-cases where data, schema, and applications are heterogenous and distributed then that impedes the expected results. This limitation has roots in formalism and corresponding reasoning mechanism underlying ontology-based knowledge bases. Ontology as the symbolic layer is closest to concepts in the real world. An ontology may be defined as the specification of a representational vocabulary for a shared domain of discourse which may include definitions of classes, relations, functions and other objects (Gruber, 1993). Ontologies are good in describing general invariant concepts and mappings or relation among those concepts. When ontologies are applied for information systems, then they describe information which is attached to multiple parameters, for example, information that is local and specific to some domain, time-dependent, constraints applicable to certain domain of discourse.

This chapter discusses how to formally represent information in use in electronic patient records (EPR) and related knowledge bases, where the data are distributed, heterogeneous and multi-contextual. We especially explore how existing formalisms are able to deal with the difficult issues provoked by heterogeneity in a globalized information system. To do this, we present in a plausible use case scenario where two hospitals in different countries are involved, as well as labs and clinics. This serves to identify essential issues arising in such environment. We then show that Semantic Web technologies can help solving these issues and consolidate interoperability. Yet, these technologies fail at several levels in this multi-scoped situation. Therefore, we investigate formal approaches that have been proposed on top of Semantic Web technologies to deal with these crucial aspects of world-wide knowledge base systems. As a result of this investigation, we classify the approaches according to five essential features that are meaningful to dealing with our example scenario. We conclude that no approach fully solve the issues but some can be combined to improve the overall formalism. Especially, we notice that those issues eventually amount to delimiting the *scope* and *type* of a knowledge base or its subparts. Subsequently, we detail how to define an extension of existing work to treat more appropriately the identified features. Finally, we discuss the remaining foundational problems that are still not addressed by the presented approaches but are critical to the interoperability of these systems. This way we hope to offer a roadmap and directions for future research in semantic-enabled HCLS system at Web-scale.

We start the chapter with background information about semantics in HCLS systems (Section 2). We then describe our use-case scenario (Section 3). We show how to apply various formalisms to this scenario in four sections overviewing the state of the art: first, we present two general theories of reasoning with context

(Section 4); second, we detail some instanciation of one of the model of context (Section 5); third, we present more concrete formalisms for the description of context on the Semantic Web (Section 6); fourth, we provide other formal approaches built on top of semantic technologies that deals with the identified problems of our scenario (Section 7). After this extensive state of the art, we provide a summary and analysis of the studied approaches (Section 8). Then, we present our proposal for combining existing approaches to better deal with scopes and types of knowledge (Section 9). Finally, we discuss the remaining open research issues that we deem crucial to enable interoperability (Section 10).

2 Semantics for HCLS

An overwhelming amount of HCLS knowledge is represented in natural language, information models, clinical repositories (databases), ontologies for terminologies, vocabularies, etc. Additionally, the involvement of various stakeholders, such as hospitals, healthcare standards, pharmacy, patients, multiplies the integration complexity of this domain. Intelligent processing, logical aggregation of information, synthesis and analysis, and the development of knowledge systems that can serve purposeful ends are needed. HCLS has been one of the primary field of application for knowledge representation and reasoning systems. In the past researchers have tried to formalize and integrate the knowledge bases in HCLS systems and many of the successful systems in earlier times were centralized and limited to sub domain or particular application of a HCLS domain (Szolovits, 1982; Kashyap & Sheth, 1996). Current HCLS systems are much more open and available to global society where stakeholders mobility and seamless use of the overall system is of prime concern.

As discuss above, ontology based information systems (OBISs) offer greater flexibility and automation for the management and integration of very complex intertwined HCLS data and schemas. HCLS specialists increasingly argue in favor of Semantic Web technologies for representing medical and clinical knowledge (Rector, Qamar, & Marley, 2006) in a well formalized way. However, current Semantic Web technologies alone are still too limited to provide a unified framework for all the varieties of applications and sub-domains of life sciences. Also, they show their limits when integrating and exchanging data between different systems. The W3C HCLS Interest Group[1] and various research projects have taken initiatives for the ontological representation of healthcare information models and their integration with HCLS terminologies and vocabularies (Bicer, Laleci, Dogac, & Kabak, 2005; Rector et al., 2006; Sahay, Akhtar, & Fox, 2008; Fox, Sahay, & Hauswirth, 2008).

This integration is crucial to effectively achieve a unified-view of electronic health records (EHR). However, this approach is still facing core integration issues such as ontological heterogeneity, ambiguous separation between global and local healthcare knowledge (Jahnke, Bychkov, Dahlem, & Kawasme, 2005). For

---

[1] http://www.w3.org/2001/sw/hcls/

example, high level ontologies that are shared by all systems cannot describe all possible sub-domains (Berners-Lee & Kagal, 2008) that may be needed by local clinics. Therefore, they must extend common knowledge with domain ontologies, as well as define internal policies, rules and vocabularies that are specific to their context.

In conclusion, HCLS is a complex domain and any data integration system which connects healthcare institutes must facilitate heterogeneous systems at two levels (1) information model specific data, and (2) domain and/or institute specific terminologies or vocabularies. These two levels must interoperate to aggregate and exchange medical records from disparate healthcare systems. In this section we describe these two levels and explain how regional and/or local clinical practices influence the modeling of clinical data.

2.1 HCLS Information Models

An information model allows modeling of domain and/or institute specific message requirements. Health Level Seven (HL7[2]) standard (version 3) develops information model specific data standards for storing and exchanging information in the healthcare industry. HL7 is the most widely used healthcare standard and shares many semantic equivalences with other influential standards such as openEHR[3] and CEN13606[4].

The HL7 (version 3) information modeling process recognizes three interrelated types of information models: Reference Information Model (RIM), Domain Message Information Model (DMIM), and Refined Message Information Model (RMIM). The RIM is a unified model which covers all domains in healthcare and defines data structures, data types and vocabularies, as well as the relationships between them. DMIM is a refined subset of the RIM and is used to create messages for a particular domain (*e.g.*, Lab, Hospital Admin). RMIM is a subset of a DMIM and is used to express the information content for a message or set of messages (*e.g.*, Lab-Test Order Message). All three interrelated models use the same notation and have the same basic structure but differ from each other in their information content, scope, and intended use.

In the example scenario presented in Section 3, Galway University Hospital (GUH) uses the ontological representation of RIM and creates local ontology using DMIM and RMIM models.

2.2 HCLS Terminologies and Vocabularies

HCLS terminologies and vocabularies (*e.g.*, *SNOMED* (Spackman, 2008), *LOINC*[5]) describe medical concepts. When these concepts are placed in a clinical record they become associated with an observation (*e.g.*, lab test), time (*e.g.*, effective

---

[2] http://www.hl7.org/

[3] http://www.openehr.org/

[4] http://www.cen.eu/

[5] http://loinc.org/

time), policy and profile (*e.g.*, drug policy, patient profile), and relationships with other records. These associations influence the interpretation of the concepts (Rector & Brandt, 2008). For example, (1) a clinical *observation* concept (*e.g.*, blood sugar test) has an effective time during which it is valid, and (2) a diabetic concept placed in a "family history" segment of a patient record does not imply that the individual patient is a diabetic patient.

Standard compliant patient records are constructed from the combination of messages derived from an information model and several terminologies or vocabularies referring to message attributes. On the other hand, many healthcare institutes do not use the standard specific information model, rather their messages are constructed using general models. The presence of different healthcare standards, large scale applicability, and limitations of syntactic integration solutions, motivated the healthcare specialists to apply Semantic Web technologies to resolve heterogeneity in a formal and consistent way.

HCLS information models, terminologies and vocabularies can be expressed as a set of RDF, RDFS and OWL propositions. RDF is best for expressing medical or patient data and OWL allows more expressive propositions to be expressed, like those that represent general knowledge rather than specific patient data elements. The relationships between heterogeneous healthcare data and knowledge can be formally expressed in OWL constructs. The reasoner underlying an expressive Web language like OWL can be used to entail consistent sets of inferred knowledge about the healthcare Web resources.
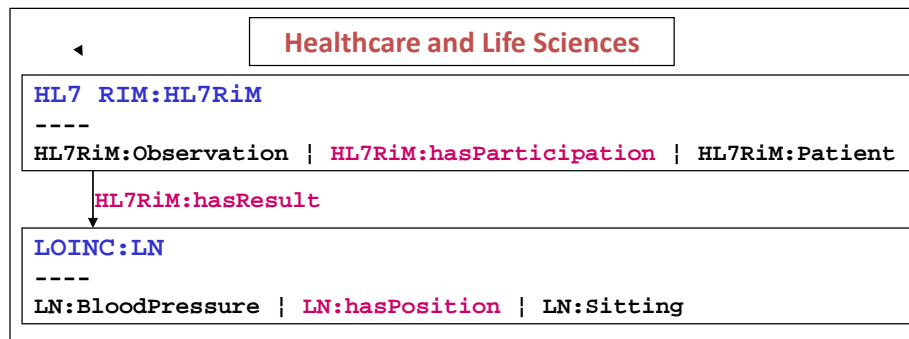

Fig. 1. Ontologies for Information Model and Terminologies

Figure 1 shows the predicates (*i.e.*, RDF triples) describing relationship between information model (HL7 RIM) specific data and general *LOINC* vocabularies. The upper (HL7 RIM) RDF triple describes the lab observation (*e.g.*, blood test) for a participating patient, and lower triple (*LOINC*) is used to describe the blood pressure (BP) and physical position (*e.g.*, sitting, standing) of the patient while measuring the BP.

3 Use-Case Scenario: Lab-Test Order

The scenario describes clinical events in course of primary medical care provided to a patient suffering from acute chronic diabetes. Clinical events are recorded as a part of patient medical history where two healthcare providers, namely, hospitals, use different healthcare standards to model the electronic patient records (EPRs). The scenario highlights the interoperability issues between HCLS systems due to (1) heterogeneity of data, information models, terminologies and vocabularies, and (2) inability to model and execute local, *i.e.*, context-specific, policy and profiles on a more general framework like the Semantic Web.

3.1 Background

In Galway, Ireland community clinic, Dr. Paul Smith is a primary care physician. Galway University Hospital (GUH) has a pathological laboratory, which is directed by Dr. John Colleen who is responsible for approving and releasing all test results. The information system of the GUH laboratory is standard, *i.e.*, HL7 Version 3, compliant and able to receive orders and return result reports electronically. Dr. Gustav Roth is an internal medicine physician in the City hospital, Göttingen, Germany (GOET). City hospital laboratory is directed by Dr. David Hahn and the only pathological laboratory available in Göttingen. GOET laboratory information systems use general vocabularies and terminologies for information modeling such as *GALEN* [6], *FOAF* [7], and able to send and receive lab test orders electronically.

3.2 Patient Case History 1

Dr. Paul Smith has a patient, Sean Murphy, who is examined on the 16th of July 2007 because of a poor wound healing, weight loss, and increase in urine production. Dr. Paul draws a blood sample, labels it with a bar code, and sends it to Galway University Hospital laboratory, GUH. Sean Murphy has been identified based on his Irish PPS number: 678970W. The blood sample is identified as 7854698. Dr. Paul fills out the electronic order form in his office system for a Glycosylated hemoglobin (HbA1c) test and sends it to Galway University Hospital Laboratory (GUH). Table 1 shows the summary of the lab order data.

GUH received the electronic order and returned a message accepting the order with intent to fulfill it on a routine basis. GUH Lab performs the requested HbA1c test on the 18th of July 2007, after receiving the sample. The result of the HbA1c test is 9%. The LOINC code for the hæmoglobin count is 4548-4. Dr. John reviewed the results of the HbA1c test and noticed that the blood sugar is abnormally high (normal reference range 4%-5.9%).

Dr. John authorized the results of the HbA1c lab tests with an indication that Sean's blood sugar level is abnormal high. The order was complete so a

---

[6] http://www.co-ode.org/galen/
[7] http://www.foaf-project.org/

notification is sent to Dr. Paul Smith. It is important to mention that, even though GUH Lab system is HL7 version 3 compliant and can electronically exchange information, it requires manual effort and paper work to integrate the patient observations.

Sean has been diagnosed as an acute chronic diabetic (Type 2) and that requires him to frequently test his blood sugar level. GUH has special drug-policy for diabetics (Type 2) where patients are treated with either Insulin or Avandia (an insulin equivalent drug), but not both. Sean receives the Insulin therapy and any future abnormal blood sugar level will require immediate and appropriate insulin treatment for him. His nature of job demands frequent travels within Ireland and his travel constraints bring complexities whenever he is in an emergency situation. Every emergency situation requires him to visit a nearby clinic or hospital, and each clinical visit requires repetitive lab tests that may have been performed earlier. Sean original medical records are with GUH lab information system and immediate availability of Sean's medical records is crucial in emergency situation.

3.3 Patient Case History 2

On Christmas holidays, Sean went to Göttingen, Germany for a week to meet his friends. On his way back to home he had a major car accident and doctors need to operate him as soon as possible. Sean is identified based on his international driving licence number 345678IE. He had a year of medical history as an acute chronic diabetic patient and his current medical records are in Galway University Hospital (GUH) Lab and other Irish hospitals where Sean visited frequently. Coincidentally, Galway University Hospital (GUH) Lab and City hospital, Göttingen, collaborate as a part of common EU healthcare framework and they can share information. In emergency situations like this, where time is critical, the automated integration of patient data can really improve the overall quality of primary healthcare services. In this scenario, GUH and GOET lab systems are able to send and receive patient's records electronically but require manual intervention, paper work, and phone calls to make sensible understanding of medical records and Sean was running out of time. City hospital, Göttingen has decided to perform all kinds of test locally without depending on his medical history. Sean was able to speak and provided an informal description of his medical history. Dr. Gustav Roth has ordered a similar test dated back on 16/07/2007 to examine his current blood sugar level before going into any surgical treatment. GOET drug policy on type 2 diabetics does not suggest any restrictions for Avandia and Insulin treatments, and thus Dr. Roth decides to prescribe Avandia in complement of Sean's insulin treatment. Table 1 show the summary of the lab order data.

In this scenario, GUH and GOET lab systems are Semantic-Web-enabled and able to send and receive patient's records electronically. But, due to differences in their domain-specific models (local models), terminology schemes, local policy, etc. information between Galway and Göttingen hospitals cannot be integrated in automated way. Figure 2 shows a snippet of the ontologies used in GUH (left)

Items	Lab Value (GUH)	Lab Value (GOET)
Lab order requisition number	6560-89	7779-78
Specimen identifier	7854698	89756
The ordering physician	Dr. Paul Smith	Dr. Gustav Roth
Physician health identifier	68374532	6837-4532
Patient	Sean Murphy	Sean Murphy
Patient health identifier	678970W:IrishPPSId	345678IE:Drivinglicence
Patient birth date	15/07/1974	15-07-1974
Date of Lab test Order	16/07/2007	22-12-2008
Lab Test ordered	Glycosylated hemoglobin	hæmoglobin A1c
LOINC/SNOMED Lab Test Code	4548-4 (LC)	43396009 (SD)
Priority of Lab Test	Routine	Urgent

Table 1. Sean's lab test orders dated 16/07/2007 (GUH) and 22/12/2008 (GOET)

and GOET (right). Figure 3 shows the table 1 equivalent data described semantically, *i.e.*, RDF triples. GUH is using ontologies that are strongly influenced by HCLS current standards (*e.g.*, RIMOWL) while GOET is using more Semantic Web oriented ontologies (*e.g.*, *GALEN*, *FOAF*) which better ensures interoperability with other Linked Data (*e.g.*, other administrative systems may take advantage of *FOAF* data). They model medical, clinical and patient knowledge using OWL, reusing well established ontologies such as *SNOMED* and *Galen*. They extend them to represent local domain knowledge, internal policies, etc. The data, *e.g.*, the patient records, are represented in RDF. This presents a realistic situation since GUH has already been involved in a project where Semantic Web technologies were used (Sahay, Fox, & Hauswirth, 2009).

3.4 Context and Constraint for OBIS

As a mechanism for representing information, RDF (Hayes, 2004), RDFS (Brickley & Guha, 2004) and OWL (Group, 2009) are very general and primitive (Klyne, 2001). As discussed above, ontologies are monotonic and make an Open World Assumption (OWA). Therefore, they are good at describing general high-level invariant concepts and their mappings. In ontology-based information systems (OBIS) the knowledge engineers often encounters a situation when describing detailed relationships between physical-world objects and their mappings becomes impossible or tedious task. In traditional ontology modeling and integration approaches, it is assumed that everything is global and the local perspective of the domain is largely ignored. This shows the limitation of ontology-based information systems (OBIS) where it is necessary to distinguish between the *scope* and `type` of knowledge for both the information models and ontologies. For example, in Figure 1 it is noticeable that the information-model-specific triple describes a blood pressure (BP) observation within a particular hospital for a specific patient unlike the LOINC triple that provide a very general description of a blood pressure (BP). When both types of knowledge, general and constraint,
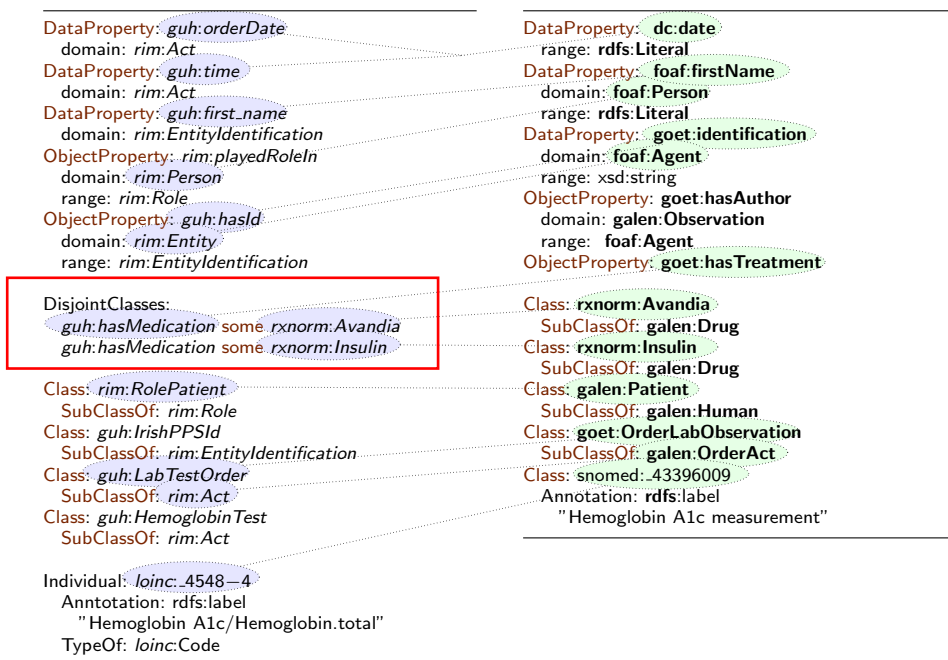
DataProperty: *guh:orderDate*
  domain: *rim:Act*
DataProperty: *guh:time*
  domain: *rim:Act*
DataProperty: *guh:first_name*
  domain: *rim:EntityIdentification*
ObjectProperty: *rim:playedRoleIn*
  domain: *rim:Person*
  range: *rim:Role*
ObjectProperty: *guh:hasId*
  domain: *rim:Entity*
  range: *rim:EntityIdentification*

DisjointClasses:
  guh:hasMedication some rxnorm:Avandia
  guh:hasMedication some rxnorm:Insulin

Class: *rim:RolePatient*
  SubClassOf: *rim:Role*
Class: *guh:IrishPPSId*
  SubClassOf: *rim:EntityIdentification*
Class: *guh:LabTestOrder*
  SubClassOf: *rim:Act*
Class: *guh:HemoglobinTest*
  SubClassOf: *rim:Act*

Individual: *loinc:_4548−4*
  Anntotation: rdfs:label
    "Hemoglobin A1c/Hemoglobin.total"
  TypeOf: *loinc*:Code

DataProperty: **dc:date**
  range: **rdfs:Literal**
DataProperty: **foaf:firstName**
  domain: **foaf:Person**
  range: **rdfs:Literal**
DataProperty: **goet:identification**
  domain: **foaf:Agent**
  range: xsd:string
ObjectProperty: **goet:hasAuthor**
  domain: **galen:Observation**
  range: **foaf:Agent**
ObjectProperty: **goet:hasTreatment**

Class: **rxnorm:Avandia**
  SubClassOf: **galen:Drug**
Class: **rxnorm:Insulin**
  SubClassOf: **galen:Drug**
Class: **galen:Patient**
  SubClassOf: **galen:Human**
Class: **goet:OrderLabObservation**
  SubClassOf: **galen:OrderAct**
Class: snomed:_43396009
  Annotation: **rdfs**:label
    "Hemoglobin A1c measurement"

Fig. 2. Extract of GUH (left) and GOET (Right) ontologies. Correspondences are informally represented by dotted lines. The rectangle shows the local policy axiom.

receive a similar treatment for their interpretation, aggregation, information retrieval, then conflicts and inconsistency arises. From the information modeling perspective the major issues in OBIS are:

1. The ability to distinguish local and global information;
2. The ability to deal with informing from different contexts;
3. The ability to model time-dependent information;
4. The ability to model constraints (*e.g.*, policy, profiles) and validation.

GUH and GOET model the same domain of knowledge in a common format (RDF and OWL). These systems define medical knowledge according to two different contexts in a heterogeneous way. There are two levels of contextual heterogeneity:

– *intra-contextual heterogeneity*, which occurs within a context. As seen in the example, several vocabularies are used for describing different types of information (*e.g.*, *RxNorm* [8] for drugs, Loinc for types of tests). Multiple terminologies must be integrated in a modular way. Local systems also have

[8] http://www.nlm.nih.gov/research/umls/rxnorm/

```
:sean a rim:Person ;                                :345678IE a galen:Patient ;
   rim:playedRoleIn [a rim:RolePatient] ;               goet:identification "Irish Driving licence" ;
   guh:hasID :678970W .                                 foaf:family_name "Murphy" ;
:678970W a guh:IrishPPSId ;                             foaf:firstName "Sean" ;
   rdfs:label "Sean Murphy's Irish PPS number" ;        goet:hasTreatment [a rxnorm:Avandia] .
   guh:first_name "Sean" ;                          :6837−4532 a galen:Doctor ;
   guh:last_name "Murphy" ;                             goet:identification "Professional Id" ;
   guh:hasMedication [a rxnorm:Insulin] .               foaf:family_name "Roth" ;
:paul a rim:Person ;                                    foaf:firstName "Gustav" .
   rim:playedRoleIn [a rim:RolePhysician] ;         :7779−78 a goet:OrderLabObservation ;
   guh:hasID :68374532 .                                goet:hasAuthor :6837−4532 ;
:68374532 a guh:PID ;                                   goet:hasPatient :345678IE ;
   rdfs:label "Paul Smith's professional Id" ;          galen:orders :Test777767 .
:6560−89 a guh:LabTestOrder ;                       :Test777767 a galen:BloodSugarTest ;
   guh:orderedBy :paul ;                                dc:date "2008−12−22T00:30:00" ;
   guh:hasPatient :sean ;                               goet:hasMeasurement [a snomed:_408254005] ;
   guh:orderDate "2007−07−16" ;                         goet:hasSpecimen :89756 .
   guh:time "13:07:07" ;                            :s89756 a galen:Specimen .
   guh:orders :Test743890 .
:Test743890 a guh:HemoglobinTest ;
   rim:effectiveTime "M0717173034" ;
   guh:specimen :7854698 ;
   guh:hasCode loinc:_4548−4 .
```

Fig. 3. Extract of Lab-Test orders at GUH (left) and GOET (right).

to deal with knowledge of differing natures, *e.g.*, axioms about a domain of discourse (*e.g.*, an hemoglobine test is a kind of act) and internal policy (a patient cannot be treated with both Insulin and Avandia).

− *inter-contextual heterogeneity* occurs between contexts. If the terminologies are different, the systems cannot interoperate unless some relations are asserted between the domain ontologies or a translation mechanism exists. Correspondences between local ontologies are informally presented in Figure 2 using dotted lines. Besides, corresponding concepts of distinct ontologies can be modeled in very different ways. For instance, GUH here uses an object property for patient's identification, while GOET is using a datatype. Thus, systems should be able to tolerate such heterogeneity. Finally, we can see that identifying context is crucial, notably to see which policy belongs to which context.

Outline of the state of the art. The following sections detail the existing formal approaches that were devised to deal with the problem of heterogeneity in multi-contextual knowledge systems. In Section 4, we described foundational work on modeling context, especially focusing on McCarthy's and Guha's approach on the one hand, and Giunchiglia *et al.*on the other hand. The later approach have been more successful in its adoption and led to several instantiation of the approach, that we discuss in Section 5. While these formalisms essentially focus on reasoning accross contexts or modularity, they offer a limited infrastructure for "reifying" the notion of context, describing it and constraining it. These aspects are better covered by development of Guha *et al.*'s work that we present in Section 6. Finally, we show that other essential concerns of interoperable

HCLS systems, such as modeling constraints, policies, rules, are also tackled by proposed extensions of semantic formalisms that are quite independent of the handling of heterogeneity and context.

4   State of the Art - 1: Context Formalisms

4.1   Logic for Context

In artificial intelligence, the notion of context generally concerns the representation and use of information. The notion is used to account for phenomena such as the context of validity of information (Kleer, 1986) and the efficiency of reasoning in narrower contexts (Guha, 1991). An influential but sketchy attempt to formalize context dependency was made by John McCarthy (McCarthy, 1987). The idea was to mark the dependency of propositions on context and track this dependency through changes of context and stages of reasoning. The basic step was to move from a (simple) proposition to the (meta) proposition that the proposition in question is true in a context. The syntax for such meta-propositions was `ist(c,p)` for proposition `p` and context `c`. Contexts thus entered the theory as objects, enabling the theory to express changes of context and the effects of such changes on propositions. Changes of context typically involve making or dropping assumptions, so the syntax also allows compound, functional terms for contexts such as `assuming(p, c)` (the context obtained from `c` by assuming `p`). With such terms, the theory can cover logically interesting consequences of context change, expressed as `ist(c, p) ⇒ ist(d, q)` and similar formulas, called lifting axioms. McCarthy's ideas were later developed by others, and found their way into a working AI system (Cyc, in the form of micro-theories applied for medical domain (Lenat, 1995)). McCarthy's work made contexts as first class objects in the domain that can be quantified, be the range of functions, and so on. All this allows one to write very expressive formulae that lift axioms from one context to another. While this is very convenient, it also makes it quite difficult to provide an adequate model theory and extremely difficult to compute with (Akman & Surav, 1996).

4.2   Contextual Reasoning - Local Model Semantics

In the early nineties, a different line of thought (compared to McCarthy's and Guha's works on context) was proposed—Local Model Semantics (LMS)—by Fausto Giunchiglia (Giunchiglia, 1993; Ghidini & Giunchiglia, 2001), who argued that people do not use all of the knowledge in an attempt to solve a problem instead they construct a "local theory" where each independent theory (or context) is related to some particular domain knowledge. Giunchiglia's formalization of context is motivated by the problem of locality where the reasoning process uses only a subset of the world knowledge. Therefore, in LMS, context is a theory of the world that encodes an agent's perspective of it and that subset is used during a given reasoning process. While reasoning, the user

can switch from one context to another in case the original context is not adequate to solve the problem. Giunchiglia's approach is more towards formalizing contextual reasoning than McCarthy's work of formalizing context as a first class object.

Serafini *et al.* then proposed refinements of LMS to concretely realize a distributed reasoning framework, in Section 5.2 we discuss those refinements, namely Distributed Description Logics (DDL) (Borgida & Serafini, 2003) and a contextualized form of OWL, called C-OWL (Bouquet, Giunchiglia, van Harmelen, Serafini, & Stuckenschmidt, 2003). For instance, DDL is an underlying formalism behind the Contextualized Web Ontology Language (C-OWL) that allows separate ontological models for a domain and correspondences between entities of different models are formalized via domain relations which are in turn used to interpret the bridge rules. Reasoning will then be carried out locally with respect to a single context and is shared only via explicitly specified connections, following the principle of locality and compatibility. Similar authors (Bouquet, Giunchiglia, van Harmelen, Serafini, & Stuckenschmidt, 2004) have attempted to align and map medical ontologies (e.g., *GALEN*, *UMLS* [9], *TAMBIS* [10]) using three proposed features of contextual ontologies, *i.e.*, directionality of information, local domain, and context based ontology mappings.

5 State of the Art - 2: Context and Constraint for OBIS

5.1 Standard Approach-OWL

The representation and reasoning of contextual knowledge is outside the scope of OWL semantics (Motik, Patel-Schneider, & Cuenca-Grau, 2009; Schneider, 2009). When reasoning is performed across different ontologies, then these ontologies share a single and global interpretation domain. The mappings between two ontologies become part of the overall model which is interpreted globally. In OWL, the ability to combine ontologies is restricted to the import of the complete ontology and to the use of the imported elements by direct reference. Thus, a set of local ontologies is globalized in a unique shared model, via the import mechanism. OWL also supports mappings constructs that implicitly exist in terms of the mutual use of statements across ontologies (*e.g.*, *subClass*, *sameAs*, *equivalentClass*, *equivalentProperty*, *differentFrom*, and *AllDifferent* axioms).

Although we have already raised some doubts in the informal discussion above on the suitability of current Semantic Web technologies with respect to heterogeneity and context, we can explore what OWL can offer to overcome these problems. This language partially addresses modularity and expressing correspondences between various ontologies.

OWL Solution. In the example scenario, several well identified terminologies are reused. OWL provides an import feature thanks to the property `owl:imports`

---

[9] http://www.nlm.nih.gov/research/umls/
[10] http://www.cs.manchester.ac.uk/ stevensr/tambis/

which helps to modularly integrate several domain ontologies. By stating that the GUH ontology imports RIM and LOINC, the axioms of these ontologies are automatically made part of the GUH ontology.

In the example scenario, concepts and properties of the two hospitals are modeled differently, but correspondences can be identified and expressed in OWL. In Listing 1.1, we present the axioms that should be added to make the two systems interoperate[11]. Notice that these mappings relating terms in two different contexts cannot be distinguished from mappings between terms of imported ontologies within one context.

```
Class: ( foaf:Person ) SubClassOf: ( rim:Entity )
Class: ( galen:OrderAct ) SubClassOf: ( rim:Act )
Class: ( rim:playedRoleIn some rim:RolePatient ) EquivalentTo: ( galen:Patient )
Class: ( guh:LabTestOrder ) EquivalentTo: ( goet:OrderLabObservation )
Class: ( guh:HemoglobinTest and (rim:measures some loinc:_4548−4) )
    EquivalentTo: ( galen:BloodSugarTest and (goet:hasCode some {snomed:
        _43396009}) ) )


EquivalentProperties: ( guh:first_name ) ( foaf:firstName )
EquivalentProperties: ( guh:hasMedication ) ( goet:hasTreatment )

Instance: ( guh:sean ) sameAs: ( goet:345678IE )
```

Listing 1.1. Extract of OWL supported mapping definitions

This approach is the only possible way of dealing with heterogeneity which fully complies with established Semantic Web standards. It has been argued that it improves interoperability of HCLS systems (Rector et al., 2006), when compared to previous standards in this field, such as HL7. However, these standards are clearly not enough to solve the important issues presented in Section 3.

OWL Limitations. First, while a form of modularity is offered by OWL, its import statement can only handle the reuse of full ontologies without being able to specify subparts of them. This is particularly problematic with large ontologies like *SNOMED*, which have to be fully integrated, even when only a small subdomain is needed.

Second, not all relevant mappings can be expressed in OWL. For example, (1) the *ObjectProperty* guh:hasId and the *DatatypeProperty* goet:identification are designed for a similar purpose (identify the person) but OWL semantics does not allow to map between *ObjectProperty* and *DatatypeProperty*; (2) OWL does not support operations on attributes, *e.g.*, the concatenation of two *Datatype-Properties* (*e.g.*, guh:orderDate, guh:time) into a single *DatatypeProperty* (*e.g.*, dc:date). Other examples include unit or currency conversion.

Third, OWL does not include any feature for distinguishing between universal facts (*e.g.*, a patient is a person) and local policy or profile (*e.g.*, people should

---

[11] As in Figure 2, we are using the OWL Manchester syntax (Horridge & Patel-Schneider, 2009).

not be treated with both Insulin and Avandia). Additionally, OWL does not permit identifying the context of an axiom or term. The implication of these two limitations is that policies have to be represented as DL axioms, and these axioms are affecting all contexts identically. In our scenario, according to GUH, Sean is treated with Insulin. When he goes to GOET, the record indicates that he has been given Avandia. Thanks to the aforementioned mappings, the terms hasMedication and hasTreatment can be interpreted interchangeably, so that GOET can understand GUH record automatically. But it leads to a contradiction with the GUH policy because Sean has now both treatments. Yet, it should not be the case because GOET does not have this policy, and therefore should not detect an inconsistency. Note that undesired interactions can be reduced by using subsumption instead of equivalence in mappings, but the problem remains.

Fourth, OWL is not tolerant of diverging modeling of a knowledge domain. Different points of view can equally well describe a domain of interest, while being partially incompatible. Interpreting all axioms and assertions as equally true, in all contexts, may easily lead to inconsistency or nonsensical entailments.

5.2 Distributed Description Logics (DDL)

Distributed Description Logics (DDL) (Borgida & Serafini, 2003) is a formalism which was developped to formalize contextual reasoning with Description Logic ontologies. Indices $i \in I$ are used to determine from which context an ontology or an axiom comes from. Given, for instance, an axiom $C \sqsubseteq D$ from an ontology O_i, DDL uses the prefixed notation $i\!:\!C \sqsubseteq D$ to highlight the context of the axiom. Moreover, cross-context formulas can be defined to relate different terminologies. These particular formulas are called bridge rules and written either $i\!:\!C \xrightarrow{\sqsubseteq} j\!:\!D$ or $i\!:\!C \xrightarrow{\sqsupseteq} j\!:\!D$ where i and j are two different contexts, and C and D are terms from the contextual ontologies O_i and O_j respectively. A bridge rule $i\!:\!C \xrightarrow{\sqsubseteq} j\!:\!D$ (resp. $i\!:\!C \xrightarrow{\sqsupseteq} j\!:\!D$) should be understood as follows: from the point of view of O_j (*i.e.*, in the context j), C is a subclass (resp. superclass) of D.

In terms of model-theoretic semantics, this is formalized by assigning a distinct Description Logic interpretation $\mathcal{I}_i$ to each contextual ontology O_i, instead of having one single global interpretation. Thus, there is as many domains of interpretation as there are contexts. Additionally, cross-context relations are made explicit by so-called domain relations, that is set-theoretic binary relations between each pairs of contexts (formally, $r_{ij} \subseteq \Delta^{\mathcal{I}_i} \times \Delta^{\mathcal{I}_j}$). Two contextual interpretations $\mathcal{I}_i$ and $\mathcal{I}_j$ satisfy a bridge rule $i\!:\!C \xrightarrow{\sqsubseteq} j\!:\!D$ (resp. $i\!:\!C \xrightarrow{\sqsupseteq} j\!:\!D$) iff $r_{ij}(C^{\mathcal{I}_i}) \subseteq D^{\mathcal{I}_j}$ (resp. $r_{ij}(C^{\mathcal{I}_i}) \supseteq D^{\mathcal{I}_j}$).[12]

The advantage of this approach is the identification of context, a better robustness with respect to heterogeneity, improved modularity. However, it still misses some of the requirements that were identified in the example-scenario.

---

[12] For a set S, $r_{ij}(S) = \{x \in \Delta^{\mathcal{I}_j} \mid \exists y \in S, \langle x, y \rangle \in r_{ij}\}$.

Solution in DDL. In the scenario, ontologies would be related thanks to C-OWL (Bouquet et al., 2003) bridge rules, which instantiates DDL for the Description Logic of OWL. A P2P reasoning system called Drago (Serafini & Tamilin, 2005) implements a fragment of C-OWL and could be used in each hospital. Each peer manages its own context by reasoning with its internal ontology and "incoming" bridge rules. Messages are sent to neighbour peers according to a distributed algorithm involving bridge rules in order to take advantage of knowledge from other contexts.

In our healthcare use case, GUH and GOET may implement a Drago reasoner. GOET expresses the correspondences by way of bridge rules, as shown with a few examples in Listing 1.2.

---

guh:(*rxnorm*:*Insulin*) $\overset{\equiv}{\longrightarrow}$ **goet**:(*rxnorm*:*Insulin*)
guh:(*rxnorm*:*Avandia*) $\overset{\equiv}{\longrightarrow}$ **goet**:(*rxnorm*:*Avandia*)
guh:(*rim*:*playedRoleIn* some *rim*:*RolePatient*) $\overset{\equiv}{\longrightarrow}$ **goet**:(**galen**:**Patient**)
guh:(*guh*:*hasMedication*) $\overset{\equiv}{\longrightarrow}$ **goet**:(**goet**:**hasTreatment**)
guh:(*guh*:sean) $\overset{=}{\longmapsto}$ **goet**:(**goet**:345678IE)

Listing 1.2. Extract of DDL bridge rules

Because of the semantics of bridge rules, no inconsistency can be derived in this case. So DDL reduces the problem of diverging policies. In fact, DDL decreases interactions between different ontologies, which in turn decrease the chance of inconsistency.

Limitations. Bridge rules are not able to represent mappings between object and datatype properties, nor can they express operations on datatypes. Besides, C-OWL uses the same import mechanism as OWL. Additionally, the non-standard semantics of DDL may be counter intuitive, sometimes. Neither disjointness nor cardinality constraints are "transferred" from an ontology to the other via bridge rules. That is, if Insulin and Avandia are disjoint in GUH, and there are the bridge rules above, it cannot be inferred that Insulin and Avandia are disjoint in GOET. However, a variant of DDL has been defined to treat this specific problem (Homola, 2007). Finally, the problem of policy is not completely solved. By adding the bridge rules of Listing 1.3, the GOET system can infer that a patient must not be treated with both Avandia and Insulin, which is what we tried to avoid.

---

guh:(*guh*:*hasMedication* some *rxnorm*:*Insulin*) $\overset{\equiv}{\longrightarrow}$ **goet**:(**goet**:**hasTreatment** some
      rxnorm:*Insulin*)
guh:(not *guh*:*hasMedication* some *rxnorm*:*Avandia*) $\overset{\equiv}{\longrightarrow}$ **goet**:(not **goet**:
      hasTreatment some *rxnorm*:*Avandia*)

Listing 1.3. Other possible bridge rules

While this example may seem a bit artificial, it shows that some restriction would have to be made on bridge rules to avoid undesired inferences. These restrictions are not, by themselves, supported by the formalism.

5.3 Other contextual reasoning formalisms

Contextual reasoning formalisms are characterized by a non-standard semantics where several ontologies are assigned distinct interpretations. Apart from DDL, this family of formalisms includes $\mathcal{E}$-connections, Package-based Description Logics and Integrated Distributed Description Logics.

Package-based Description Logics. In package-based Description Logics (P-DL (Bao, Caragea, & Honavar, 2006)), each ontological axiom is associated with an identifier of the ontology, similarly to DDL. Moreover, as in other contextual formalisms, a distinct interpretation is assigned to each ontology in a network of ontologies. However, cross-ontology knowledge can only take the form of *semantic imports* of ontological terms. The reason behind this is that this formalism was essentially designed to compensate the drawbacks of the OWL import mechanism and improve modularity of Web ontologies.

As an example, the GOET ontology only uses a few terms from the *GALEN* ontology. Since the *GALEN* ontology is extremely big and highly expressive, its complete import would result in a very complex, hard to manage ontology. However, while this helps building local ontologies in a modular way, it does not very much help expressing cross-context knowledge such as the correspondences that needed to bridge GUH and GOET. As an ontology alignment formalism, semantic imports have a very limited expressiveness. To express a complex correspondence such as, for instance, the one of Listing 1.4, one would have to first import the terms `guh:HemoglobinTest`, `rim:measures` and `loinc:_4548-4`, then add a *local* axiom in standard DL using the same approach as in Section 5.1.

```
Class: ( guh:HemoglobinTest and (rim:measures some loinc:_4548−4) )
    EquivalentTo: ( galen:BloodSugarTest and (goet:hasCode some {snomed:
        _43396009}) )
```

Listing 1.4. An example of correspondence that cannot be represented with semantic imports.

Integrated Distributed Description Logics. Integrated Distributed Description Logics (IDDL (Zimmermann, 2007)) is a formalism that address similar issues as DDL but take a different paradigm than other contextual frameworks. Usually, cross-ontology assertions (*e.g.*, bridge rules in DDL, links in $\mathcal{E}$-connections, semantic imports in P-DL) define knowledge from the point of view of one ontology. That is to say that the correspondences are expressing the relations "as witnessed" by a local ontology. On the contrary, IDDL asserts correspondences from a "third party"'s point of view which encompasses both the ontologies in relation. One consequence of this approach is that correspondences can be manipulated and reasoned with independently of the ontologies, allowing operations like inversing or composing ontology alignments, as first class objects.

In terms of model theory, this is represented by using an additional domain of interpretation to the whole network of ontologies, as if it was a single ontology.

The local domains of interpretation, assigned to all ontologies, are then related to the global domain by way of the so-called *equalizing functions* (ε_i). These functions map the elements of local domains to elements of the global domain. Formally, a correspondence $i : C \xleftrightarrow{\sqsubseteq} j : D$ from a concept C of ontology O_i to concept D of ontology O_j is satisfied whenever $\varepsilon_i(C^{\mathcal{I}_i}) \subseteq \varepsilon_j(D^{\mathcal{I}_j})$.

Furthermore, a reasoning procedure for this formalism has been defined (Zimmermann & Duc, 2008), where a central system detaining the correspondences can determine global consistency of a network of ontologies, by communicating with local reasoners of arbitrary complexity. This formalism is useful for federated reasoning systems, while the interactions between local ontologies are rather weak. By separating local reasoning and global reasoning, it better prevents interactions between contexts, thus being quite robust to heterogeneity.

In our example, let us assume that the correspondences of Listing 1.5 are defined to make the two systems interoperate.

```
guh:( rxnorm:Insulin ) ⟷≡ goet:( rxnorm:Insulin )
guh:( rxnorm:Avandia ) ⟷≡ goet:( rxnorm:Avandia )
guh:( rxnorm:Avandia ) ⟷⊥ goet:( rxnorm:Insuline )
guh:( rim:playedRoleIn some rim:RolePatient ) ⟷≡ goet:( galen:Patient )
guh:( guh:hasMedication ) ⟷≡ goet:( goet:hasTreatment )
guh:( guh:sean ) ⟷= goet:( goet:345678IE )
guh:( guh:hasMedication some rxnorm:Insulin ) ⟶≡ goet:( goet:hasTreatment some
      rxnorm:Insulin )
guh:( not guh:hasMedication some rxnorm:Avandia ) ⟶≡ goet:( not goet:
      hasTreatment some rxnorm:Avandia)
```

Listing 1.5. Extract of IDDL correspondences

policy of the form $C \sqsubseteq \neg D$ would only influence another ontology if a disjointness is asserted at the alignment level, *e.g.*, One can see the similarity of these correspondences and bridge rules. Yet the resulting inferences differ. No inconsistency will arise from these correspondences. However, thanks to the third correspondence, the system would detect an inconsistency of a medicine is asserted to be Avandia and Insuline at the time. While this formalism decreases undesired interaction of knowledge, especially with respect to policies, its drawback is the possible missing inferences at the local level. Moreover, correspondences are not more expressive than in DDL.

$\mathcal{E}$-connections. $\mathcal{E}$-connections is another formalism for reasoning with heterogeneous ontologies (Kutz, Lutz, Wolter, & Zakharyaschev, 2004). Again, different ontologies are interpreted distinctly but formally related using particular assertions. Instead of expressing correspondences of ontological terms, an ontology can connect to another by using special terms (called *links*) which can be combined in conjunction with terms from another ontology. The semantics of links is very similar to the semantics of roles in Description Logics, except that instead of relating elements from the same domain of interpretation, they relate two

different domains. In principle, $\mathcal{E}$-connections serve to relate ontologies about very different domains of interest. For instance, an ontology of laboratories in GUH could be connected to an ontology of medical staff used in GOET. To do this, one can define the link ⟨hasDirector⟩ and use it in GUH ontology as in Listing 1.6.

```
guh:Laboratory ⊑ ∃⟨ hasDirector ⟩goet:StaffMember
```

Listing 1.6. An axiom of an ontology using a *link* in the $\mathcal{E}$-connections formalism.

Thus, $\mathcal{E}$-connections are particularly useful for ontology design by modularly reusing and connecting existing blocks. However, one of the main focus of this chapter is on relating existing ontology systems on overlapping domains. So, although $\mathcal{E}$-connections is a relevant formalism for the management of heterogeneity, its applicability to the type of scenario we are interested in is weak.

6   State of the Art: 3 - Context in Semantic Web technologies

6.1   Models of provenance

RDF model-theory (Hayes, 2004) provides reification as a mechanism for making statements about statements. There are significant differences between reification and contexts both in what they are intended for and in their structure. Reification is intended to enable statements about potential statements (which may or may not be true). They can be useful for making statements about provenance (Watkins & Nicole, 2006; Stoermer, Bouquet, Palmisano, & Redavid, 2007). Named graphs are also used for making statement about provenance (Carroll, Bizer, Hayes, & Stickler, 2007). In named graphs triples become quadruples where the fourth element is ID(URI), *i.e.*, origin of the graph. In addition to the reification mechanism, named graphs, and quadruples in RDF, the system/language CWM/N3[13] supports a construct called contexts. This notion of context is not substantially different from reification. Since these approaches to include additional information about the graph have no coupling with the truth of the triple that has been reified, they cannot be used to relate the truth of a triple in one graph to its truth in another graph. Consequently, it is hard to see how reification, named graphs, quadruples can be used to mediate data aggregation as far as "truth of a triple" is concerned.

6.2   Contextual RDF(S)

The author of (Guha, McCool, & Fikes, 2004) proposed an extension of RDF(S)—Context Mechanism—to incorporate contextual knowledge within RDF model theory. A simpler version of OWL is assumed to be interoperable with the

---

[13] http://www.w3.org/2000/10/swap/doc/

proposed context mechanism. The most basic change in RDFS model-theory introduced—by the addition of contexts—is that the denotation of a resource is not just a function of the term and the interpretation (or structure), but also of the context in which that term occurs. Most importantly, the proposed context mechanism allows RDF statements to be true only in their context. The goal of this RDFS extension is to aggregate triples that are true in the graphs being aggregated, and because of the close coupling between truth and contexts, they cannot be a posteriori introduced at the RDF Vocabulary level (Lutz & Sattler, 2000; Donini, Nardi, & Rosati, 2002; Wagner, 2003; Analyti, Antoniou, Damásio, & Wagner, 2004). They appear in the internals of the model theory, in the definition of an interpretation and satisfaction.

In a standard RDF(S) entailment a vocabulary is interpreted as a tuple $V = \{\mathcal{U}, \mathcal{PL}, \mathcal{TL}\}$ that consists of a set $\mathcal{U}$ of URI references, a set $\mathcal{PL}$ of plain literals, and a set $\mathcal{TL}$ of typed literals where $\mathcal{U}, \mathcal{PL}, \mathcal{TL}$ are mutually disjoint. An interpretation I of vocabulary (V) is a tuple $\mathcal{I} = \{\mathcal{IR}, \mathcal{IP}, \mathcal{LV}, \mathcal{IS}, \mathcal{IL}, \mathcal{IEXT}\}$, where $\mathcal{IR}$ is a non-empty and finite set, called the domain or universe of $\mathcal{I}$, $\mathcal{IP}$ is a finite set of properties, $\mathcal{LV} \subseteq \mathcal{IR}$ is a finite set of literal values with $\mathcal{PL} \subseteq \mathcal{LV}$, $\mathcal{IS}$ is a mapping $\mathcal{IS} : \mathcal{U} \longmapsto \mathcal{IR} \cup \mathcal{IP}$, $\mathcal{IL}$ is a mapping $\mathcal{IL} : \mathcal{TL} \longmapsto \mathcal{IR}$ and $\mathcal{IEXT}$ is an extension function $\mathcal{IEXT} : \mathcal{IP} \longmapsto 2^{(\mathcal{IR} \times \mathcal{IR})}$. In the proposed context mechanism, minor changes in the definitions of standard RDF(S) interpretation (Hayes, 2004) are as follows:

1. vocabulary $V = \{\mathcal{U}, \mathcal{CU}, \mathcal{PL}, \mathcal{TL}\}$ contains set $\mathcal{CU}$ of contextual URIs;
2. a set $\mathcal{C} \subseteq \mathcal{IR}$, is introduced to denote context ($\mathcal{C}$);
3. a mapping $\mathcal{IS}$ from power set $2^{(\mathcal{U} \times \mathcal{CU})}$ in V into $\mathcal{IR} \cup \mathcal{IP}$, the power set $2^{(\mathcal{U} \times \mathcal{CU})}$ corresponds to resource-context (*i.e.*, $\langle \mathcal{U}, \mathcal{CU} \rangle$) pairs.

The context mechanism updates the standard RDF(S) satisfaction by allowing set of context-dependent graphs instead of a single graph, for example, if $\mathcal{E}$ is a ground triple[14] $\langle s, p, o \rangle$ in the context c then $\mathcal{I}(E, c) = \texttt{true}$ if c, s, p and o are in V, $\mathcal{IS}(p, c)$ is in $\mathcal{IP}$ and $\langle \mathcal{IS}(s, c), \mathcal{IS}(o, c) \rangle$ is in $\mathcal{IEXT}(\mathcal{IS}(p, c))$. Otherwise $\mathcal{I}(E, c) = \texttt{false}$. Considering the updates within standard RDF(S) interpretation and satisfaction, now graphs will be merged with regard to where they occur. It means that, the definition of entailment is updated so that a ground graph G_1 in a context C_1 is entailed by a set of graph-context pairs $\langle G_i, C_i \rangle$ if $\langle G_1, C_1 \rangle$ is true under every interpretation under which ($\langle G_i, C_i \rangle$) is true. The proposed context mechanism may lead to non-monotonic aggregation—depending on the expressivity of *lifting rules*[15] suggested by the authors of (Guha et al., 2004)—in the following sense. A graph G_1 might imply φ but the aggregation of this graph (including lifting rules) with other graphs might not imply φ.

In our example-scenario, if the contents at URLs `galway.hospital/guh.rdf` and `Gottingen.hospital/goet.rdf` are available as RDF then we can have a context corresponding to these URLs and the contents of an URL is said to

---

[14] subject (s), predicate (p), object (o) without blank node

[15] lifting rules are basically normal *imports* that brings contents of one context to another

be true in that context. Furthermore, lifting rules can be defined to import all or part of the contents from data sources. For example, if we assume that an extended version of RDFS could express the *disjoint* axiom in GUH drug policy as below, then the truth of this axiom can be scoped to the context `galway.hospital/guh.rdf`.

$$\exists \mathrm{guh:hasMedication.rxnorm:Avandia} \sqsubseteq \neg \mathrm{guh:hasMedication.rxnorm:Insulin}$$

Similarly, subclass axioms (as below) of goet drug policy can be scoped to `Gottingen.hospital/goet.rdf`.

```
Class: rxnorm:Avandia
    SubClassOf: galen:Drug

Class: rxnorm:Insulin
    SubClassOf: galen:Drug
```

In GOET, Sean medical treatment record showed that he has been given Insulin therapy at GUH and Avandia treatment is forbidden in conjunction with Insulin therapy. The truth of GUH drug policy is applicable only within GUH context and when GUH and GOET records will be aggregated then GUH drug policy could be easily ignored (thus avoiding inconsistency) by using appropriate lifting rule. Similarly, when Sean is back to Galway, GOET drug policy would not influence his further treatment in Galway or any other places.

7 State of the Art - 4: Other formal handling of heterogeneity

7.1 Database-Style Integrity Constraints for OWL

This approach is motivated by data-centric problems in DL/OWL based applications. The authors of (Motik, Horrocks, & Sattler, 2007) have established the relationship between the role of Integrity Constraints (IC) in databases, *i.e.*, (1) data reasoning (*e.g.*, in checking the integrity of a database) and schema reasoning (*e.g.*, in computing query subsumption), and (2) DL/OWL knowledge bases (*e.g.*, schema (*TBox*) reasoning and data (*ABox*) reasoning). In this approach an additional *TBox* is introduced to model constraint axioms, which results in the knowledge base containing two *TBoxes* and an *ABox*. In *TBox* reasoning, constraints behave like normal *TBox* axioms, and for *ABox* reasoning they are interpreted as constraints in relational databases. This approach is very relevant in solving profile and policy issues of our example scenario. For example, to avoid inconsistency due to hospital specific drug policy, axiom:

$$\exists \mathrm{guh:hasMedication.rxnorm:Avandia} \sqsubseteq \neg \mathrm{guh:hasMedication.rxnorm:Insulin}$$

can be placed in *TBox* for constraints and when *TBox* reasoning is performed only standard axioms could be taken into account. In case of *ABox* reasoning,

constraints axioms can act as Integrity Constraints. To some extent, it helps formalizing policies but since it does not identify the context of these constraints, their utility for this purpose is limited. Moreover, as a standard OWL, robustness to heterogeneity is poor.

7.2   Modular Web Rule Bases

Although this approach is not based on current Semantic Web standards, it is relevant to this survey. The framework proposed in (Analyti, Antoniou, & Damásio, 2008) makes the distinction between global knowledge, local knowledge and internal knowledge. The framework is based on a rule-based language rather than Description Logics and provides an approach to express and reason with modularity on top of Semantic Web. In this framework each predicate in a rule base is constrained with "uses" and "scope", which in turn determine the reasoning process. The framework also treats different forms of negation (weak or strong) to include Open-World Assumption (OWA) as well as Closed-World Assumption (CWA) (Analyti, Antoniou, Damásio, & Wagner, 2008). This rule-based framework provides a model-theoretic compatible semantics and allow certain predicates to be monotonic and reasoning is possible with inconsistent knowledge bases. This framework addresses a few issues of our example scenario because rules can express some DL axioms and can be exchanged with certain restrictions (private, global or local). For example, the drug policy rule of our example scenario:

$\mathcal{F} \leftarrow$ *hasMedication*(?x,?y), *Avandia*(?y), *hasMedication*(?x,?z),*Insulin*(?z)

can be expressed and treated appropriately. However, the major problem we observe is how DL-based ontologies (as majority of HCLS ontologies are DL ontologies) and rules can work together. The integration of DL with rules is still an open research problem (Eiter, Ianni, Krennwallner, & Polleres, 2008). Moreover, this framework is not concerned about the heterogeneity of the knowledge model, and does not provide an expressive way of relating contextual ontologies.

7.3   Query-Based Data Translation

The query-based approach translates data from one knowledge source to another, and is close to the problem of expressing complex correspondences of the example-scenario presented in this chapter. In this approach mappings between ontologies are first expressed in expressive alignment language (Euzenat, Scharffe, & Zimmermann, 2007) and then grounded and executed to a combined query language and engine, SPARQL++ and PSPARQL, called PSPARQL++ (Euzenat, Polleres, & Scharffe, 2008). Listing 1.7 show how (a) two ontology entities (guh:orderDate, guh:time) could be concatenated to a single entity (dc:date) and (b) a conversion is possible between an object property and a datatype property by using proposed cast-function that converts *xsd:string* to RDF resource. Expressive correspondences between ontology instances can be constructed using "SPARQL CONSTRUCT" to create additional dataset and query upon them.

```
(a) CONSTRUCT { ?X dc:date fn:concat(?Date,"T",?Time). }
      WHERE { ?X guh:orderDate ?Date . ?X guh:time ?Time . }

(b) CONSTRUCT { ?X guh:hasId rdf:Resource(fn:encode−for−uri(?Id)) . }
      WHERE { ?X goet:identification ?Id . }
```

Listing 1.7. Mappings expressed as a SPARQL CONSTRUCT query

This approach allows one to express complex correspondences like concatenating attributes or even datatype to object properties and one can avoid some undesired interactions between knowledge of various sources. However, one major limitation is that the query result depends on how the correspondences are written and the knowledge in the domain ontologies are largely unexploited. Similarly, complex correspondences can be expressed in Rule Interchange Format (RIF), which offers a rich set of built-in functions (*e.g.*, *string* manipulations), as well as a formal semantics[16] for interoperating with RDF and OWL knowledge bases. However, RIF is yet to be approved as W3C's standard recommendation and so far is still a work in progress, which is why we do not further focus on this approach here.

7.4 Reasoning with Inconsistencies

Robustness to heterogeneity is an important aspect in healthcare integration scenarios. One of the most problematic consequences of heterogeneity is the occurrence of undesired inconsistencies. Therefore, we believe it useful to investigate formal approaches for handling inconsistencies. There are two main ways to deal with inconsistent ontologies. One is to simply accept the inconsistency and to apply a non-standard reasoning method to obtain meaningful answers in the presence of inconsistencies. An alternative approach is to resolve the error, that is, to repair the ontology, whenever an inconsistency is encountered.

Repairing or revising inconsistent ontology is, in principle, a possible solution for handling inconsistency. However, one major pragmatic issue we observe is that healthcare institutes may not expose and/or allow repair of their knowledge bases due to various legal constraints. Also, in a typical Semantic Web setting, importing ontologies from other sources makes it impossible to repair them, and if the scale of the combined ontologies is too large as in the case of HCLS ontologies then repair might appear ineffective. Other work focus on revising mappings only (Meilicke, Stuckenschmidt, & Tamilin, 2008), but they are meant to be used at alignment discovery time, which we are not discussing in this chapter.

Reasoning with inconsistencies is also possible without revision of the ontology. One effective way of tolerating inconsistencies consist of using paraconsistent logics (Béziau, Carnielli, & Gabbay, 2007). Paraconsistent logics use a "weaker" inference system that entails less formulas than in classical logics. This way, reasoning can be done in the presence of inconsistency. A paraconsistent extension

---

[16] http://www.w3.org/TR/rif-rdf-owl/

of OWL was proposed in (Huang, van Harmelen, & ten Teije, 2005). Alternatively, defeasible argumentation (Chesñevar, Maguitman, & Loui, 2000) and its implementation Defeasible Logic Programs (DeLP (García & Simari, 2004)) have been introduced to reason and resolve inconsistencies. In this case, the TBox is separated into 2 subsets, one being *strict*, which means that it must always be used in reasoning, the other being *defeatable*, which means that an argumentation process may defeat them and nullify them for a particular reasoning task.

While we want to tolerate inconsistency when reasoning with an ontology defined in another context, it is not desirable to tolerate local inconsistencies in a HCLS system. The system should have a strict logical framework when it only treats local data, that are existing in a unique and well understood context. Unfortunately, the approaches mentioned here are not able to distinguish local knowledge and external knowledge. They do not allow specification of the types of mappings we need, and are not capable of treating policies.

8   Comparison of formal approaches

While the previous sections described the state of the art horizontally by merely describing approaches one by one, the present section analyzes the features of formal approaches vertically, that is, each characteristic is compared separately throughout the extent of formal approaches.

Table 2 shows a brief, synthetic summary of what is detailed here. The columns can be seen either as a feature of a formal approach or as an issue related to heterogeneity. Consequently, the content of a cell can be read as "this formal approach possess this feature" (if a $+$ is present) or not (if a $-$ is present) or it can be read as "this issue is addressed by this formal approach" $(+)$ or not $(-)$. In the case of the last two columns, the issue defines a continuum between "this issue is not addressed at all" and "this issue is fully addressed", but all approaches are only partially addressing them. Thereby, we order the extent to which they address the issue in the following way:

Very limited < Limited < Medium < Good < Very good < Excellent

The first column represents *context awareness* (C.A. in the table) which is the ability or possibility to identify the context in which some information or knowledge is described. The second column shows the possibility of *modularizing* ontologies (M. in the table). This is considered fully addressed if subparts of ontologies can be reused rather than reuse of complete ontologies. The third column (P.&P.M. in the table) shows whether the formalism can be used to *manage profiles and policies*. The fourth column gives an indication of *correspondence expressiveness* (C.E. in the table), relatively to the other formalisms. We use here a loose notion of "expressiveness" and our classification is partly based on informal arguments rather than an authentic logical proof. Finally, the fifth column shows the *robustness with respect to heterogeneity* (R.H. in the table). This corresponds to the capability to exploit knowledge from independently designed sources while keeping coherence and relevance of inferences.

Context awareness can be enabled by clearly separating the axioms and facts asserted in distinct ontologies or from distinct provenance. Such formalisms as DDL, P-DL, $\mathcal{E}$-connections and IDDL assign an identifier to each of the ontologies such that an axiom $i\,{:}\,C \sqsubseteq D$ can be associated with an ontology O_i. This also ensures the distinction between local axioms and cross-ontology correspondences, such as bridge rules $i : C \stackrel{\sqsubseteq}{\rightarrow} j : D$. However, these formalisms do not allow one to describe or give a type to contexts. The contextual RDF framework of (Guha et al., 2004) improves on this by treating context identifiers as any other RDF resource identifiers: they are URIs that can be dealt with as first class objects. The case of query-based transformations is a bit particular because the transformation must say from which and to what datasets the translation occurs, but the transformation could actually be used independently of the context. This explains the $+/-$ sign for this formal approach. Modular rule bases uses a mechanism similar to the one of DDL etc but the underlying formalism is not based on Semantic Web languages.

Modularity is almost always associated with context awareness. This is because in order to reuse modules, one must be able to identify them, that is, have a mean of identifying the provenance of knowledge. Moreover, formalisms such as DDL, $\mathcal{E}$-connections, P-DL, IDDL are designed to enable the reuse of external knowledge through the use of cross-context assertions like imports, bridge rules, links, correspondences. Therefore, what identifies a context can be thought of as a module. To a lesser extent, OWL allows some modularisation but is limited the full import of complete ontologies which results—from a logical point of view—in the same ontology as a complete merge of all imported knowledge. OWL/IC has the same behavior as OWL in terms of modularity. Finally, the contextual RDF approach allows to identify contexts and relate them, it is harder to separate knowledge into modular blocks since contexts can overlap and, as first class objects, contexts themselves can be part of a context. These intertwined elements make modularization in contextual RDF(S) more difficult.

Policies and profiles has not been often addressed as a knowledge representation issue. Most of the solutions to this problem are using ad hoc implementations or algorithm that are not explicitly tackled by the underlying formalism. Of course, it is possible to use existing languages—such as RDF and OWL as in (Kolovski, Parsia, Katz, & Hendler, 2005)—to represent policies but OWL cannot formally distinguish between ontological axioms and policy axioms. This is true for most of the formalisms presented here. Notable exceptions are Defeasible Logic Programs for which there is a distinction between "strict" knowledge, which is always true, and "defeatable" knowledge which can be canceled by way of an automatic argumentation process. Policies are typically defeatable by stronger policies (*e.g.*, from GUH's point of view, GOET's policy axioms would be defeated by contradicting local policies). As an alternative, OWL/IC offers a way to separate the usual ontological knowledge from database-style constraints. Unfortunately, both approaches are unable to distinguish the context in which appear the axioms or policies. Thereby, they can only separate the totality of the ontological axioms from the totality of the policy constraints. Finally, Web

Rule bases has a similar aptitude to distinguish different types of knowledge by assigning different reasoning scheme to different sets of rules. Additionally, it separates knowledge from different sources in a similar way as contextual formalisms do. However, it is not complying to Semantic Web technologies and it would hadly be possible to integrate it with description logics systems.

Correspondence expressiveness varies a lot depending on the formalism being used. Most of the correspondences given in the example scenario are directly expressible in OWL due to its high expressiveness. However, OWL constructs still have some limitations which appear in practical cases. In formalisms that allow delimiting contextual knowledge, the expressiveness of correspondences is defined by the types of relations that can be asserted between contexts. The case of the query-based approach to transforming data is notable. This approach allows for defining very fine grained correspondences in the form of a data transformation. Basing the approach on queries decreases the exploitation of reasoning but enables all kinds of structural and functional transformations.

Robustness to heterogeneity is the ability to make consistent and useful inferences in spite of the variations of points of view, modeling approaches and contexts. There are two extreme cases: one is the classical logic approach, the other is the context separation approach. In classical logic, all statements are treated equally in a theory and can equally interfere with any other statement. Therefore, incompatible views are very likely to produce inconstencies or nonsense. This results in a formalism very vulnerable to heterogeneous knowledge. On the opposite side, it is possible to simply separate all conclusions drawn from a context from any other context. This is very robust to heterogeneity since, granted that each context is self consistent, it cannot produce incompatible inferences. However, this way, no context can take advantage of information coming from a different context. Context aware formalisms can be ordered in their ability to tolerate heterogeneity: P-DL < modular rule bases < DDL revisited < DDL < IDDL < $\mathcal{E}$-connection < Query-based. The reason why the query-based approach is so robust is that it does not actually reason with a source ontology to produce new data according to the target knowledge base. Moreover, the transformation can be designed in such a way that it conform to the destination ontology.

Conclusion of the state of the art. Although the presented formalisms are significantly different from each others, there are common aspects to them and some of them are not incompatible. First, in all these approaches, we can recognize the need to delimit parts of the knowledge to assign them either provenance information, a type (policy, ontological knowledge) or a distinct reasoning scheme (OWA, CWA). Thereby, a better solution to the problem of heterogeneity should possess this ability to identify subparts of the knowledge and apply different status to them. Second, context-aware formalisms are usually adequate approaches to modularization thanks to their ability to relate distinct ontologies. Additionally,

---

[17] C.A.: Context-awareness; M.: Modularity; P.& P.M.: Profile and policy management; C.E.: Correspondence expressiveness; R.H.: Robustness to heterogeneity.

	C.A.[17]	M.[17]	P.& P.M.[17]	C.E.[17]	R.H.[17]
DL/OWL	-	+/-	-	Good	Very limited
DDL/C-OWL	+	+	-	Very good	Good
P-DL	+	+	-	Very limited	Limited
DDL Revisited	+	+	-	Very good	Medium
$\mathcal{E}$-connections	+	+	-	Medium	Excellent
IDDL	+	+	-	Good	Very good
RDFS-C	+	+/-	-	Good	Good
DeLP/Paraconsistent	-	-	+/-	Good	Good
Query-based	+/-	-	-	Very good	Very good
Modular Rule bases	+	+	+/-	Limited	Limited
OWL/IC	-	+/-	+/-	Good	Limited

Table 2. Formal Approaches towards Heterogeneity and Context

the separation of knowledge makes them more tolerant to heterogeneous modeling. However, these approaches have very little addressed issues of policies and profiles. Yet, handling constraint-like information differently from ontological knowledge is critical to the management of electronic patient records. Improving the so-called *modular ontology languages* with policies, or making policy-aware formalisms more context sensitive is an important aspect to achieve interoperability of HCLS systems. For this reason, we will present two possible directions for handling policies within context-aware formalisms in Section 9. Since this is by no mean a complete solution to the vast problem of heterogeneity, we thereafter discuss more briefly the other critical issues that should be covered by a semantic formalism to represent the knowledge involved in this type of systems.

9 Towards a framework for handling context and heterogeneity

In this section, we propose a general framework for adding constraint-like axioms to a context-aware formalism such as DDL, P-DL, IDDL. While a semantics must be chosen to allow context aware reasoning, they can be extended with other non-monotonic approaches. There are many possible combinations, too many to compare them all, so we only show possible paths that we think are best suited for the scenarios we consider.

9.1 Adding constraints to context-aware formalisms-1: DL-based

We propose to combine the formalisms that are identifying the context of axioms with a mechanism for handling policies that do not lead to undesired interactions. This approach is partly inspired by (Motik et al., 2007), where axioms are separated into two T-Boxes, one for ontological axioms, the other for integrity constraints. Figure 4 depicts the general idea behind our framework. This way, we define a local T-Box as a pair $\langle D, P \rangle$, where D describes a domain ontology,

and P represents the internal policies. If several local ontologies and policies exist, the overall knowledge is a distributed system $(\langle D_i, P_i \rangle)$. To ensure interoperability, ontology alignments (A_{ij}) are added to the system to bridge the gaps between different terminologies. Note that these alignments could be simple DL axioms, DDL bridge rules, P-DL semantic imports, or IDDL correspondences. The resulting system is a pair of tuples $\langle (\langle D_i, P_i \rangle), (A_{ij}) \rangle$.

To simplify the presentation, consider a simple pair of OBIS as in our example scenario, that is, the system in consideration is $\Omega = \langle (\langle D_{\mathrm{guh}}, P_{\mathrm{guh}} \rangle, \langle D_{\mathrm{goet}}, P_{\mathrm{goet}} \rangle), A_{\mathrm{guh,goet}} \rangle$. The semantics of such a system depends on the type of alignments in place, which defines the type of context-aware formalism used. Let us write $\models_d$ the distributed entailment relation of the considered formalism. Thereby, $\langle (O_i), (A_{ij}) \rangle \models_d k :$ $C \sqsubseteq D$ means that the system of ontologies (O_i) and alignment (A_{ij}) distributively entails the axiom $C \sqsubseteq D$ in ontology O_k.

A policy-enabled distributed entailment $\models_P$ is defined over distributed ontologies Ω as follows. For a given local OWL axiom $\alpha = \mathrm{guh} : C \sqsubseteq D$ in the terminology of the ontology of GUH, $\Omega \models_P \alpha$ if and only if α is distributively entailed by the system composed of $D_{\mathrm{guh}} \cup P_{\mathrm{guh}}, \{D_{\mathrm{goet}}\}$ and $A_{\mathrm{guh,goet}}$ (and vice versa if the axiom belong to the ontology of GOET).
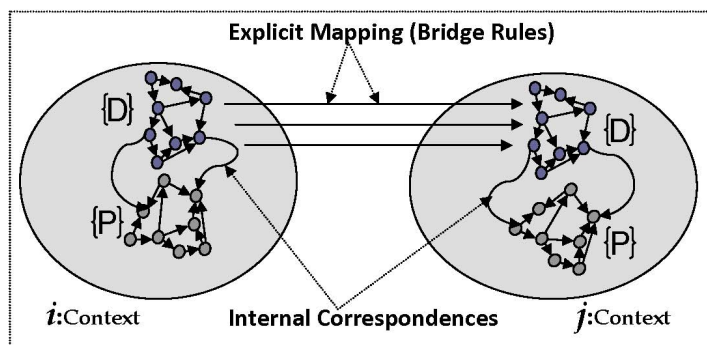

Fig. 4. Enabling Context Specific Policies-DL Based

In other words, only the policy axioms of the ontology which is asking for an entailment is used. In our scenario, it means that if GUH is reasoning, it will take its drug policy into account but not the one of GOET, while GOET would not consider the Avandia-Insulin counter indication of GUH. The very same approach can be easily adapted to DDL, P-DL or IDDL.

9.2 Adding constraints to context-aware formalisms-2: Rule-based

As discussed above, DL-based solutions are adhering to the Open World Assumption (OWA), that is, conclusions which cannot be derived from an ontology are treated agnostically. The approach presented in Section 9.1 enables separate

context spaces allocated for the special treatment of the context-specific policies. However the intrinsic nature of a DL-based solution may expose limitations when one expects certain rational conclusions, which are reasonable to infer even under incomplete knowledge. To overcome this limitation and to deal with the conservative stance of the DL-based solution, we propose a rule-based approach that will allow the appropriate treatment of context-specific policies by treating them with Closed World Assumption (CWA). Use cases that require data integration from heterogenous information systems are suitable candidates for combining both the formalisms, that is, OWA and CWA (Eiter et al., 2008). In this approach, the knowledge base is split into two parts similar to the DL-based solution, but a separate treatment is provided for each knowledge type (OWA for normal axioms and CWA for policy axioms, see Figure 5).

This rule-based approach is inspired by (Guha et al., 2004) and (Horst, 2005b, 2005a) where authors in (Guha et al., 2004) proposed that the denotation of a resource is not just a function of the term and the interpretation, but also of the context in which that term occurs, that is, ontology axioms are interpreted within certain contexts (see Section 6.2). The authors suggested an RDFS extension by introducing context within RDFS-Model theory. The extension enables the "Context-Awareness" but the constraints (drug policy) in our example scenario demands expressive semantics (higher expressivity than RDFS), such as *disjointness*.

DL−Axiom:
$\exists\,\mathrm{guh\!:\!hasMedication.rxnorm\!:\!Avandia} \sqsubseteq \neg\mathrm{guh\!:\!hasMedication.rxnorm\!:\!Insulin})$

Rule Equivalent:
$\mathcal{F} \leftarrow$ *hasMedication*(?x,?y), *Avandia*(?y), *hasMedication*(?x,?z),*Insulin*(?z)

Listing 1.8. GUH drug policy.

To overcome this issue, Horst (Horst, 2005b, 2005a) proposed a sound and complete rule-based extension of RDFS (called R-entailment) that involves *datatypes* (useful for modeling HCLS information systems) and a subset of the OWL-vocabulary that includes the property-related vocabulary (*e.g.*, *FunctionalProperty*), the comparisons (*e.g.*, *sameAs, differentFrom, disjointWith*) and the value restrictions (*e.g.*, *allValuesFrom*). These semantic extensions are in line with the `if-semantics` of RDFS.

We suggest to combine both the approaches (Horst, 2005a; Guha et al., 2004), calling it RDFS-CR, and providing further extension to this combination by (1) including *Type* and/or *Status* of the knowledge base (normal axioms, constraint axioms), and (2) defining the *scope* of the knowledge bases (global or local). For example, domain ontology (D) (normal axioms and global scope) and context-dependent policies (P) (constraint axioms and local scope). This approach will allow the reuse of existing DL-based HCLS ontologies as normal axioms and define context specific constraints.

Figure 5 defines the knowledge base as a pair $\langle D, P \rangle$, where D describes a domain ontology and P represents the internal policy. Domain ontologies (D)

will be treated normally by standard RDFS-entailment whereas context-specific policies (P) will have RDFS-CR entailment. Policy axioms within a context will share domain ontologies. For example, RDFS-CR entailment will be applied on drug policy (α).

$\alpha = \exists\,\mathrm{guh\!:\!hasMedication.rxnorm\!:\!Avandia} \sqsubseteq \neg\,\mathrm{guh\!:\!hasMedication.rxnorm\!:\!Insulin}$
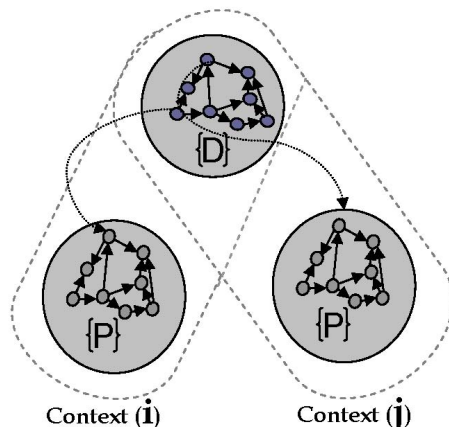
Fig. 5. Enabling Context Specific Policies-Rule Based

One major advantage of this approach is its closeness to the standard semantics. Light-weight extensions of RDFS allows both (1) context-awareness and (2) the ability to model information constraints (IC). Combining integration approaches imposes additional constraints on knowledge-base systems (Eiter et al., 2008), *i.e.*, highly expressive knowledge bases are hard to integrate with less expressive knowledge bases. For example, translating DL-axioms to rules (vice-versa) and faithful preservation of semantics while translation is a known problem. Therefore, we kept the expressiveness of knowledge bases that ensure reliable translation and execution (Horst, 2005a).

10   Towards fully interoperable HCLS systems

This section presents critical issues of semantic-enabled HCLS systems that have not been addressed by this paper. Since the breadth of this domain is so wide, this section can only discuss those problems succinctly, but we hope to provide directions for future research on the key obstacles to better interoperability.

10.1   Temporal aspects of HCLS systems

Temporal information is everywhere in patient records. For example, Lab observations (*e.g.*, blood test, bloop pressure (BP)) have a duration of validity,

vaccines must be renewed after some time, medicines have an effect on the body for a certain time (thus, counter-indications are valid beyond the duration of the treatment). One aspect of implicit, contextual information is its temporal component. The ability to identify, represent and reason about time-dependent information is important for various applications, especially for HCLS systems. With respect to temporal information one is often faced with the problem of implicit time (Moldovan, Clark, & Harabagiu, 2005). Sites often publish a piece of data that is true at the time of publication, with the temporal qualification left implicit. Equally often, this data does not get updated when it no longer holds. Even worse, such implicitly temporally qualified data is often mixed with data that is not temporally qualified. In this chapter, we have not directly addressed the time-dependency issue but we assume that framework that enables the two key features (Type and Scope) within the knowledge bases can act as a ground work for enabling and processing such a temporal information in any OBIS.

10.2   Efficiency of reasoning with non-standard logics

Relying on information systems may be great but in many occasions, practitioners must make quick decisions in order to act as soon as possible, especially in emergency situation such as our case study at Göttingen. Thereby, the efficiency of information systems is crucial in many HCLS environment. We did not take this into account in our survey. We simply looked at the possibility to logical represent and models various aspects of common HCLS systems, especially with respect to EPRs. It may be needed to have different approach depending on the type of system or department. Emergency hospitals would sacrifice expressiveness for the sake of faster feedback, while a clinic which only operates when scheduled a long time before would be more favourable to a systems that takes the most precise, most thorough description of all parameters to avoid any potential mistake. Beyond efficiency, simply having an algorithm may be difficult. Undecidability easily occur when throwing in together rules, standard logic axioms, context, policies, time and so on.

10.3   Healthcare and Social Networks

There is great perceived value in being able to integrate and use EPR and social network data across clinical and social domains. Healthcare stakeholders belongs to various social ecosystems (or networks), and the vision of patient-centric healthcare seems deemed unrealistic without providing a linkage between a patient and surrounding social networks. The idea is very simple and basic, for example, a semantically described patient data can also have a semantic description of his/her social existence (e.g., friends, family, and location). In situations where both of the networks (i.e., healthcare and social) have mutual correspondences, patient primary care (i.e., treatment of patient) and other secondary care (e.g., home-care, remote monitoring, emergency assistance, etc.) could be easily deployed on top such a linked patient related networks. In listing 1.9 (snippet of

listing 1.9) we can observe that an HL7 and *FOAF* artifacts are mapped, such correspondences are link between two networks.

```
Class: ( foaf:Person ) SubClassOf: ( rim:Entity )
EquivalentProperties: ( guh:first_name ) ( foaf:firstName )
```

Listing 1.9. HL7 RIM and *FOAF* Correspondences

Although correspondences between networks can be easily designed but it triggers various challenges with respect to protection, access, policy, etc. of patient and social data. We discussed in the example scenario that each healthcare related stakeholders are open to define their constraints and it is important to interpret and process these context-dependent constraints in a sensible way. The two solutions sketched in Section 9 can act as an enabler for such interlinked and heterogeneous networks.

10.4 Practicality of formal approaches

This chapter have presented theoretical or abstract solutions to semantic interoperability, detailing how they are, in theory, able to represent relevant information. However, it must not be forgotten that these approach have to be implemented in practical applications. The implementation itself would lead to even further research issues that are at least equally important and challenging than what we focused on.

First, the more expressive and powerful the formalism is, the more difficult it is to use and to model knowledge with. This can be overcome by building intuitive interfaces that help the user in these tasks. The problem occurs on both the side of the knowledge engineer building ontologies and for the practitioner who needs to understand what the system is telling. Ontology engineering is in itself a whole research field which gained momentum with the development of the Semantic Web. Yet, extending the ontology formalisms would also require updating the engineering methods that are mostly developed specifically for OWL (and,). For practitioners, explaining inference results is required.

Second, it can be hypothesised that not all systems will ever use a unique data model. It is thus important to rely on data conversion. For instance, the transformation of syntactic to semantic data format (and vice-versa) is the first basic step in exploring implicit semantics within syntactic data. Initiatives like RDB2RDF[18], GRDDL[19], are improving this transformation and help taking advantage of legacy-database within the framework of the Semantic Web.

Third, the actual distribution of systems leads to many practical consideration such as latency, connectivity, routing and parallelising the algorithms in an efficient way. While hospital computers are all interconnected, they cannot provide any amount of information, or answer any amount of queries from the outside. Yet, to ensure distributed reasoning, the computing power must be

---

[18] http://www.w3.org/2001/sw/rdb2rdf/

[19] http://www.w3.org/TR/grddl-primer/

shared over the network. These problems relate to other topics such as the Grid and Cloud computing.

11   Conclusion

In this chapter, we presented how HCLS systems can benefit from semantic technologies to improve interoperability. We focused on formal aspects of knowledge representation formalisms, assessing their ability to effectively model the information commonly needed in electronic patient records (EPR). While the discussion was essentially on theoretical aspects, we showed what are the concrete consequences of applying these formalisms to a real-world scenario. The strength and limitation identified for each formalism became the motivation for proposing a framework that identifies the *type* and *scope* of a knowledge base. We argued that an approach to the problem of heterogeneity—or in general interoperability of HCLS systems—relies on the combination of several integration models. However, complexities of different magnitude may arise from this combination. By this approach, we theoretically address an important problem of our example scenario. Nonetheless, the diversity of the field is so vast that we did not cover some relevant issues that we only sketched to provide a roadmap for further research. We believe that these open research issues also requires attention of a broader research community.

12   Acknowledgement

Bibliography

Akman, V., & Surav, M. (1996). Steps Toward Formalizing Context. *AI Magazine*, *17*(3), 55–72.

Analyti, A., Antoniou, G., & Damásio, C. V. (2008). A Principled Framework for Modular Web Rule Bases and Its Semantics. In Brewka, G., & Lang, J. (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pp. 390–400. AAAI Press.

Analyti, A., Antoniou, G., Damásio, C. V., & Wagner, G. (2004). Negation and negative information in the W3C resource description framework. *Annals of Mathematics, Computing and Teleinformatics*, *2*, 25–34.

Analyti, A., Antoniou, G., Damásio, C. V., & Wagner, G. (2008). Extended RDF as a Semantic Foundation of Rule Markup Languages. *Journal of Artificial Intelligence Research*, *32*, 37–94.

Bao, J., Caragea, D., & Honavar, V. G. (2006). On the Semantics of Linking and Importing in Modular Ontologies. In Cruz, I. F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., & Aroyo, L. (Eds.), *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, Vol. 4273 of *Lecture Notes in Computer Science*, pp. 72–86. Springer-Verlag.

Berners-Lee, T., & Kagal, L. (2008). The Fractal Nature of the Semantic Web. *AI Magazine*, *29*(3), 29–35.

Béziau, J.-Y., Carnielli, W., & Gabbay, D. M. (2007). *Handbook of Paraconsistency*. College Publications.

Bicer, V., Laleci, G. B., Dogac, A., & Kabak, Y. (2005). Artemis Message Exchange Framework: Semantic Interoperability of Exchanged Messages in the Healthcare Domain. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, *34*(3), 71–76.

Borgida, A., & Serafini, L. (2003). Distributed Description Logics: Assimilating Information from Peer Sources. *Journal on Data Semantics*, *1*, 153–184.

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., & Stuckenschmidt, H. (2003). C-OWL: Contextualizing Ontologies. In Fensel, D., Sycara, K. P., & Mylopoulos, J. (Eds.), *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, Vol. 2870 of *Lecture Notes in Computer Science*, pp. 164–179. Springer-Verlag.

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., & Stuckenschmidt, H. (2004). Contextualizing ontologies. *Journal of Web Semantics*, *1*(4), 325–343.

Brickley, D., & Guha, R. V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004. Tech. rep., World Wide Web Consortium (W3C).

Calvanese, D., Franconi, E., Haarslev, V., Lembo, D., Motik, B., Tessaris, S., & Turhan, A.-Y. (Eds.). (2007). *Proceedings of the 20th International Workshop on Description Logics DL'07, June 8 - 10, 2007, Brixen/Bressanone, Italy*. Bolzano University Press.

Carroll, J. J., Bizer, C., Hayes, P. J., & Stickler, P. (2007). Named graphs, provenance and trust. In Ellis, A., & Hagino, T. (Eds.), *Proceedings of the 14th International Conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*, pp. 613–622. ACM Press.

Chesñevar, C. I., Maguitman, A. G., & Loui, R. P. (2000). Logical Models of Argument. *ACM Computing Survey*, *32*(4), 337–383.

Cristani, M., & Cuel, R. (2005). A Survey on Ontology Creation Methodologies. *International Journal of Semantic Web and Information Systems*, *1*(2), 49–69.

Donini, F. M., Nardi, D., & Rosati, R. (2002). Description logics of minimal knowledge and negation as failure. *ACM Transaction on Computational Logics*, *3*(2), 177–225.

Eiter, T., Ianni, G., Krennwallner, T., & Polleres, A. (2008). Rules and Ontologies for the Semantic Web. In Baroglio, C., Bonatti, P. A., Maluszynski, J., Marchiori, M., Polleres, A., & Schaffert, S. (Eds.), *Reasoning Web, 4th International Summer School 2008, Venice, Italy, September 7-11, 2008, Tutorial Lectures*, Vol. 5224 of *Lecture Notes in Computer Science*, pp. 1–53. Springer-Verlag.

Euzenat, J., Polleres, A., & Scharffe, F. (2008). SPARQL Extensions for Processing Alignments. *IEEE Intelligent Systems*, *23*(6), 82–84.

Euzenat, J., Scharffe, F., & Zimmermann, A. (2007). Expressive alignment language and implementation. Deliverable D2.2.10, Knowledge Web NoE.

Fox, R., Sahay, R., & Hauswirth, M. (2008). PPEPR for Enterprise Healthcare Integration. In Weerasinghe, D. (Ed.), *Electronic Healthcare, First International Conference, eHealth 2008, London, UK, September 8-9, 2008. Revised Selected Papers*, Vol. 1 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 130–137. Springer-Verlag.

García, A. J., & Simari, G. R. (2004). Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, *4*(1–2), 95–138.

Ghidini, C., & Giunchiglia, F. (2001). Local Models Semantics, or contextual reasoning=Locality+Compatibility. *Artificial Intelligence*, *127*(2), 221–259.

Gil, Y., Motta, E., Benjamins, V. R., & Musen, M. A. (Eds.). (2005). *The Semantic Web - ISWC 2005: 4th International Semantic Web Conference, Galway, Ireland, November 6-10, 2005, Proceedings*, Vol. 3729 of *Lecture Notes in Computer Science*. Springer-Verlag.

Giunchiglia, F. (1993). Contextual Reasoning. *Epistemologica*, *16*, 345–364.

Group, W. O. W. (2009). OWL 2 Web Ontology Language Document Overview W3C Recommendation 27 October 2009. Tech. rep., World Wide Web Consortium (W3C).

Gruber, T. R. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In and (Ed.), *Formal Ontology in Conceptual Analysis and Knowledge Representation.*

Guha, R. V. (1991). *Contexts: a Formalization and Some Applications.* Ph.D. thesis, Stanford University, Stanford, CA (USA). Revised version at `http://www-formal.stanford.edu/guha/guha-thesis.ps`.

Guha, R. V., McCool, R., & Fikes, R. (2004). Contexts for the Semantic Web. In van Harmelen, F., McIlraith, S., & Plexousakis, D. (Eds.), *The Semantic Web - ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004. Proceedings*, Vol. 3298 of *Lecture Notes in Computer Science*, pp. 32–46. Springer-Verlag.

Hayes, P. J. (2004). RDF Semantics, W3C Recommendation 10 February 2004. Tech. rep., World Wide Web Consortium (W3C).

Homola, M. (2007). Distributed Description Logics Revisited.. In Calvanese et al. (Calvanese, Franconi, Haarslev, Lembo, Motik, Tessaris, & Turhan, 2007).

Horridge, M., & Patel-Schneider, P. F. (2009). OWL 2 Web Ontology Language Manchester Syntax, W3C Working Group Note 27 October 2009. Tech. rep., World Wide Web Consortium (W3C).

Horst, H. J. T. (2005a). Combining RDF and Part of OWL with Rules: Semantics, Decidability, Complexity.. In Gil et al. (Gil, Motta, Benjamins, & Musen, 2005), pp. 668–684.

Horst, H. J. T. (2005b). Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, *3*(2–3), 79–115.

Huang, Z., van Harmelen, F., & ten Teije, A. (2005). Reasoning with Inconsistent Ontologies.. In Kaelbling, & Saffiotti (Kaelbling & Saffiotti, 2005), pp. 454–459.

Jahnke, J. H., Bychkov, Y., Dahlem, D., & Kawasme, L. (2005). Context-Aware Information Services for Health Care. *Revue d'Intelligence Artificielle*, *19*(3), 459–478.

Kaelbling, L. P., & Saffiotti, A. (Eds.). (2005). *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005.* Professional Book Center.

Kashyap, V., & Sheth, A. (1996). Semantic and schematic similarities between database objects: a context-based approach. *The Very Large Data Base Journal*, *5*, 276–304.

Kleer, J. D. (1986). An Assumption-Based TMS. *Artificial Intelligence*, *28*(2), 127–162.

Klyne, G. (2001). Information modelling using rdf, constructs for modular description of complex systems. web published.

Kolovski, V., Parsia, B., Katz, Y., & Hendler, J. A. (2005). Representing Web Service Policies in OWL-DL.. In Gil et al. (Gil et al., 2005), pp. 461–475.

Kutz, O., Lutz, C., Wolter, F., & Zakharyaschev, M. (2004). $\mathcal{E}$-connections of abstract description systems. *Artificial Intelligence*, *156*(1), 1–73.

Lenat, D. B. (1995). CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, *38*(11), 33–38.

Lutz, C., & Sattler, U. (2000). Mary Likes all Cats. In Baader, F., & Sattler, U. (Eds.), *Proceedings of the 2000 International Workshop on Description Logics (DL2000), Aachen, Germany, August 17-19, 2000*, Vol. 33 of *CEUR Workshop Proceedings*, pp. 213–226. Sun SITE Central Europe (CEUR).

McCarthy, J. L. (1987). Generality in Artificial Intelligence. *Communications of the ACM*, *30*(12), 1029–1035.

Meilicke, C., Stuckenschmidt, H., & Tamilin, A. (2008). Supporting Manual Mapping Revision using Logical Reasoning. In Fox, D., & Gomes, C. P. (Eds.), *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pp. 1213–1218. AAAI Press.

Moldovan, D. I., Clark, C., & Harabagiu, S. M. (2005). Temporal Context Representation and Reasoning.. In Kaelbling, & Saffiotti (Kaelbling & Saffiotti, 2005), pp. 1099–1104.

Motik, B., Horrocks, I., & Sattler, U. (2007). Adding Integrity Constraints to OWL. In Golbreich, C., Kalyanpur, A., & Parsia, B. (Eds.), *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, June 6-7, 2007*, Vol. 258 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

Motik, B., Patel-Schneider, P. F., & Cuenca-Grau, B. (2009). OWL 2 Web Ontology Language Direct Semantics W3C Recommendation 27 October 2009. Tech. rep., World Wide Web Consortium (W3C).

Rector, A. L., & Brandt, S. (2008). Why do it the hard way? The Case for an Expressive Description Logic for SNOMED. *Journal of the American Medical Informatics Association*, *15*(6), 744.

Rector, A. L., Qamar, R., & Marley, T. (2006). Binding Ontologies & Coding Systems to Electronic Health Records and Messages. In Bodenreider, O. (Ed.), *KR-MED 2006, Formal Biomedical Knowledge Representation, Proceedings of the Second International Workshop on Formal Biomedical Knowledge Representation: "Biomedical Ontology in Action" (KR-MED 2006), Collocated with the 4th International Conference on Formal Ontology in Information Systems (FOIS-2006), Baltimore, Maryland, USA, November 8, 2006*, Vol. 222 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

Sahay, R., Akhtar, W., & Fox, R. (2008). PPEPR: Plug and Play Electronic Patient Records. In Wainwright, R. L., & Haddad, H. (Eds.), *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Fortaleza, Ceara, Brazil, March 16-20, 2008*, pp. 2298–2304. ACM Press.

Sahay, R., Fox, R., & Hauswirth, M. (2009). Semantic Service-oriented Design and Development Methodology for Enterprise Healthcare Integration. In Filipe, J., & Cordeiro, J. (Eds.), *WEBIST 2009 - Proceedings of the Fifth International Conference on Web Information Systems and Technologies, Lisbon, Portugal, March 23-26, 2009*, pp. 159–166. INSTICC Press.

Schneider, M. (2009). OWL 2 Web Ontology Language RDF-Based Semantics W3C Recommendation 27 October 2009. Tech. rep., World Wide Web Consortium (W3C).

Serafini, L., & Tamilin, A. (2005). DRAGO: Distributed Reasoning Architecture for the Semantic Web. In Gomez-Perez, A., & Euzenat, J. (Eds.), *The Semantic Web: Research and Applications, Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29 - June 1, 2005, Proceedings*, Vol. 3532 of *Lecture Notes in Computer Science*, pp. 361–376. Springer-Verlag.

Spackman, K. A. (2008). SNOMED CT Style Guide: Situations with Explicit Context, Style Guide 14 January 2008, Version 1.03. Tech. rep., SNOMED CT.

Stoermer, H., Bouquet, P., Palmisano, I., & Redavid, D. (2007). A Context-Based Architecture for RDF Knowledge Bases: Approach, Implementation and Preliminary Results. In Marchiori, M., Pan, J. Z., & de Sainte Marie, C. (Eds.), *Web Reasoning and Rule Systems, First International Conference, RR 2007, Innsbruck , Austria, June 7-8, 2007, Proceedings*, Vol. 4524 of *Lecture Notes in Computer Science*, pp. 209–218. Springer-Verlag.

Szolovits, P. (1982). *Artificial Intelligence in Medicine*. Westview Press, Inc., Boulder, CO, USA.

Wagner, G. (2003). Web Rules Need Two Kinds of Negation. In Bry, F., Henze, N., & Maluszynski, J. (Eds.), *Principles and Practice of Semantic Web Reasoning, International Workshop, PPSWR 2003, Mumbai, India, December 8, 2003, Proceedings*, Vol. 2901 of *Lecture Notes in Computer Science*, pp. 33–50. Springer-Verlag.

Watkins, E. R., & Nicole, D. A. (2006). Named graphs as a mechanism for reasoning about provenance. In Zhou, X., Li, J., Shen, H. T., Kitsuregawa, M., & Zhang, Y. (Eds.), *Frontiers of WWW Research and Development - APWeb 2006, 8th Asia-Pacific Web Conference, Harbin, China, January 16-18, 2006, Proceedings*, Vol. 3841 of *Lecture Notes in Computer Science*, pp. 943–948. Springer-Verlag.

Zimmermann, A. (2007). Integrated Distributed Description Logics.. In Calvanese et al. (Calvanese et al., 2007), pp. 507–514.

Zimmermann, A., & Duc, C. L. (2008). Reasoning on a Network of Aligned Ontologies. In Calvanese, D., & Lausen, H. (Eds.), *Web Reasoning and Rule Systems, Second International Conference, RR 2008, Karlsruhe, Germany, October/November 2008, Proceedings*, Vol. 5341 of *Lecture Notes in Computer Science*, pp. 43–57. Springer-Verlag.

Zimmermann, A., Sahay, R., Fox, R., & Polleres, A. (2009). Heterogeneity and Context in Semantic-Web-Enabled HCLS Systems.. Vol. 5871 of *Lecture Notes in Computer Science*, pp. 1165–1182. Springer-Verlag.