An Ontology for Clinical Trial Data Integration

Ratnesh Sahay*, Dimitrios Ntalaperas†, Eleni Kamateri‡, Panagiotis Hasapis§, Oya Deniz Beyan*,
Marie-Pierre F. Strippoli¶, Christiana A. Demetriou‖, Thomai Gklarou-Stavropoulou**, Mathias Brochhausen††,
Konstantinos Tarabanis‡, Thanassis Bouras†, David Tian‡‡, Aristos Aristodimou[x],
Athos Antoniades[x], Christos Georgousopoulos§, Manfred Hauswirth*, and Stefan Decker*
*Digital Enterprise Research Institute (DERI), NUI Galway, Ireland
†UBITECH Research, Athens, Greece
‡Centre for Research and Technology Hellas (CERTH), Greece
§INTRASOFT International S.A., Luxembourg
¶Department of Psychiatry, Lausanne University Hospital, Switzerland
‖Departments of EM/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Cyprus
**ZEINCRO HELLAS S.A., Athens, Greece
††University of Arkansas for Medical Sciences, Little Rock, USA
‡‡School of Computer Science, University of Manchester, Manchester, UK
[x]University of Cyprus, Cyprus

Abstract—A set of well-integrated clinical terminologies is at the core of delivering an efficient clinical trial system. The design and outcomes of a clinical trial can be improved significantly through an unambiguous and consistent set of clinical terminologies used in a participating clinical institute. However, due to lack of generalised legal and technical standards, heterogeneity exists between prominent clinical terminologies as well as within and between clinical systems at several levels, *e.g.*, data, schema, and medical codes. This article specifically addresses the problem of integrating local or proprietary clinical terminologies with the globally defined universal concepts or terminologies. To deal with the problem of ambiguous, inconsistent, and overlapping clinical terminologies, domain and knowledge representation specialists have been repeatedly advocated the use of formal ontologies. We address two key challenges in developing an ontology-based clinical terminology (1) an ontology building methodology for clinical terminologies that are separated in global and local layers; and (2) aligning global and local clinical terminologies. We present Semantic Electronic Health Record (SEHR) ontology that covers multiple sub-domains of Healthcare and Life Sciences (HCLS) through specialisation of the upper-level Basic Formal Ontology (BFO). One of the main features of SEHR is layering and adaptation of local clinical terminologies with the upper-level BFO. Our empirical evaluation shows an agreement of clinical experts confirming SEHR's usability in clinical trials.

Keywords—*Clinical Trial; Clinical Terminology; Semantic Interoperability; Ontology Building Methodology; Ontology Alignment; Ontology Evaluation*

I. INTRODUCTION

Ontology building is a set of activities including the ontology development steps, the ontology life cycle, supporting tools, and languages applied coherently for modelling domain knowledge. A large number of ontologies have been developed by different groups, under different approaches, and with different methods and techniques. However, in comparison to the software engineering counterpart, ontology building is still in its infancy. The advancement of technology and significant improvement in availability of structured information, ontology practitioners with the goal of speeding up the ontology development process, are starting to reuse [1] as much as possible

(i) other ontologies such as Gene Ontology (GO) [2], GALEN [3] and ontology modules [4]; (ii) ontology statements and ontology design patterns [5]; and (iii) non-ontological resources [6] such as thesauri, databases, XML schemas, UML models and classification schemas (*e.g.*, DSM-IV[1]) which already have greater degree of consensus. Developers realised that in a distributed setting, an ontology should not be developed entirely from scratch, but by reusing and possibly reengineering other ontologies, databases, XML schemas, thesauri, UML models, classification schemes, and other knowledge resources, as well as by taking into account good practices in the development process. In the field of biomedicine, domain experts have advocated the use of ontologies to deal with the problem of terminological heterogeneity in a formal and consistent way [7], [8]. Unfortunately, clinical terminologies are complex in terms of their structure, granularity, and the scope of use. Consequently, building an ontological representation of a clinical terminology is not a straightforward job. For instance, clinical terminologies could be of a global use or created locally by clinical sites. Integrating global and local terminologies has become a major challenge for the domain and technology experts [9].

Towards an envisioned knowledge base system where global and local terminologies can co-exist together meaningfully, a key knowledge engineering challenge is to build and align the local ontologies with global ones. Interestingly, the problem of aligning ontologies is a well-studied problem [10] and recently the linked data[2] initiative further improved the alignment situation at the instance level. We argue that building and aligning clinical terminologies in a distributed setting like clinical trials is not a straightforward *equivalent* or *similarity* statement, but needs a mechanism that can support layering of knowledge bases *i.e.*, in local and global spaces, and local adaptation of resources. In our earlier work, we proposed the Plug and Play Electronic Patient Record (PPEPR) [11] Methodology, that ontologises the Health Level Seven

---

[1]http://www.psychiatry.org/practice/dsm/dsm-iv-tr
[2]http://linkeddata.org/

(HL7[3]) standard. A HL7 knowledge base generally separates in global and local parts. The PPEPR Methodology includes development steps for arranging ontologies in layers *i.e.*, global and local spaces, and adapting local ontologies with the global ones. The Semantic Electronic Health Record (SEHR) ontology[4] proposed in this article is built using the PPEPR Methodology. The SEHR ontology is developed as part of the Linked2Safety[5] [12] EU project. One of the main objectives of the Linked2Safety project is to build a semantically-interlinked platform that can enable patient recruitment across the European clinical sites.

The structure of this article is as follows: first we present an integration scenario describing the types of domains and the clinical terminologies used by the Linked2Safety project. The integration scenario highlights three main features of an ontology building in a distributed setting. Second, we briefly present the PPEPR Methodology and a mechanism for layering and adaptation of local and global terminologies. Finally, we present an evaluation of the SEHR ontology that shows a mutual agreement on the correctness and completeness of the clinical terminologies described within SEHR ontology.

II. Integration Scenario

The Linked2Safety consortium includes three clinical partners namely the University Hospital Lausanne (CHUV[6]), the Cyprus Institute of Neurology and Genetics (CING[7]), and ZEINCRO[8]. Table I shows the Linked2Safety clinical partner, domain of study, and the number of clinical terminologies per partner.

Clinical Partner	Domain	Termi nologies
CHUV	Adverse Event, Cardiovascular, Demographic, Medical History, Migraine and Psychiatric Disorder	127
CING	Adverse Event, Breast Cancer, Diabetic, Genetics, and Neurology	134
ZEINCRO	Adverse Event, Concomitant Medication, Demographic, Medical History and Respiratory	764

TABLE I: Linked2Safety Clinical Partners, Domains, and Terminologies

Each clinical partner created locally defined clinical terminologies and therefore, the scope of these terminologies is local. For instance, CHUV uses terminologies covering Adverse Event, Cardiovascular, Demographic, Migraine and Psychiatric Disorder domains. CHUV's psychiatric disorder related terminologies partially correspond with the DSM-IV classification. Similarly, the other two partners partially cover the terminologies for their domain of study. The main challenge here is not a complete coverage of all the domains mentioned (second column of the Table I), but to include these partially covered domains under a consistent, unambiguous, and unifying framework of terminology. It is important to notice that the clinical partners study overlapping domains

---

(*e.g.*, Adverse Event, Demographic, Medical History) and their integration is a main issue in formulating a clinical query across these domains. For example, Listing 1 shows a sample subject selection criteria that asks for a number of males that are suffering from the MetabolicSyndrome with BMI (Body Mass Index) value greater than 25 and showed mental disorders (*e.g.*, PsychoticDisorders) symptom in the past. An answer returned by the query above, *i.e.*, number of subjects, could play a major role in deciding locations, resources, and technologies required for conducting a clinical trial. However, answering such a query requires integrating several terminologies and data originating from clinical sites.

```
Return number of MALE diagnosed with MetabolicSyndrome
and BMI > 25 and showed PsychoticDisorders symptoms
```

Listing 1: Example Subject Selection Criteria for Clinical Trials

Key Requirements: As discussed, ontology can provide a formal, consistent, and unambiguous means to represent the above mentioned clinical terminologies. However, considering the arrangement of several partially covered clinical domains and local terminologies, an ontology building method require three main features: (i) **reusability** of existing non-ontological knowledge sources (*e.g.*, XML schemas, CSV, Excel, and natural text describing classifications); (ii) **layering** of ontologies in global and local spaces; and (iii) **adaptation** of local ontologies with an upper or global ontology.

III. The PPEPR Methodology

The PPEPR methodology is grounded on existing methodologies and domain experiences [11]. Figure 1 presents the PPEPR methodology which consists of five phases: (i) the **scoping** phase establishes the purpose of ontology building and identifies resources that can support the ontology building process; (ii) the **technology support** phase evaluates Semantic Web languages and supporting tools that can fulfil the requirements of the scoping phase; (iii) the **modelling** phase provides detailed guidelines for constructing ontologies; (iv) the **alignment** phase resolves ontological heterogeneity; finally (v) the **testing** phase ensures consistency and correctness of ontologies with respect to the previous phases and requirements. Particular development steps are allocated to each phase, which indicate the order (with optional routes) in which the activities should be performed. The modelling, the alignment, and the testing phases are iterative until the required ontologies and their alignments have been constructed. Considering the space limitation, we discuss the Modelling phase, *i.e.*, the core phase involved in the development of SEHR ontology.

A. The Modelling Phase

The Modelling phase starts with the task of lifting clinical terminologies and ends with the local adaptation of ontologies. The overall goal of this phase is to build the SEHR ontology.

1) Lifting Resources: The PPEPR Methodology proposed seven lifting rules that automatically transform structured resources such as a XML schema to a corresponding ontology [11]. However, the clinical terminologies available from the clinical partners are primarily semi-structured (*i.e.*, in excel, text formats). Hence, the lifting of clinical terminologies is
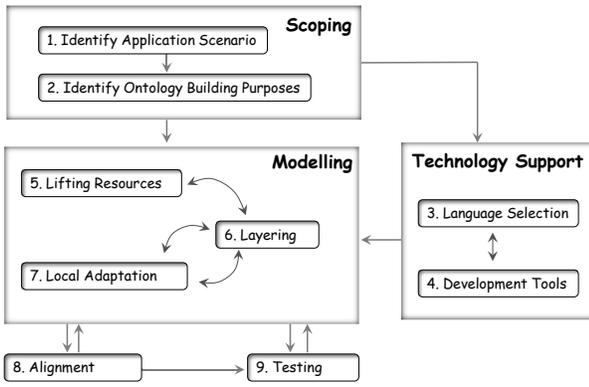
Fig. 1: PPEPR Ontology Building Methodology

mainly a manual process. For instance, to identify a "SubClassOf" relation, we analysed all sets of terminologies under a common heading/title/column (*e.g.*, Myocardial Infarction and Stoke described under the Cardiac Disorders heading), type of terminologies (*e.g.*, quantitative, qualitative), and publishers annotations. The lifting task involves creating correspondences between semi-structured resources and ontological constructs (Class, SubClassOf, Property). Similarly, the type of ontology property, *i.e.*, ObjectProperty or DataProperty is decided on a value space used by a clinical terminology, *e.g.*, storing medical code, specialised data type like mg/dl.

```
Class: chuv:MetabolicSyndrome
SubClassOf: acgt:Syndrome
Class: chuv:Schizophrenia
SubClassOf: chuv:PsychoticDisorders
```

Listing 2: Lifting Clinical Terminologies

Listing 2 shows a snippet of the SEHR ontology. The lifting correspondences, *i.e.*, "SubClassOf" inheritance relations have been identified first and later implemented using OWL. Ontology examples presented in this article use a non-logician syntax, called the OWL Manchester Syntax [13], which produces a less verbose syntax. In Listing 2 the local clinical terminology, "chuv:MetabolicSyndrome", creates a "SubClassOf" relation with Advancing Clinico-Genomic Trials on Cancer (ACGT) ontology (prefix "acgt") which is global in the scope of use. However, a "SubClassOf" relation between "Schizophrenia" and "PsychoticDisorders" is straightforward as both terms are local to the CHUV's clinical system. The ACGT Master Ontology (ACGT MO) was developed in the ACGT project, which focused breast cancer and nephroblastoma (Wilms' Tumor) [14]. One aim of ACGT was to create an ontology-driven clinical trial management system. The ACGT MO is an extension of Basic Formal Ontology (BFO) and the SEHR ontology is an extension of ACGT MO. Most importantly, the ACGT MO has the significant number of similar and/or matching clinical terminologies used in the Linked2Safety environment as compared to other clinical trial ontologies [15], [16]. The basic development strategies of ACGT MO were to provide ontological representation of data captured by case report forms (CRF) and subsume relevant classes on the BFO. BFO is an upper ontology that is based on sound theoretical and logical principles, which has proven to be highly applicable to the biomedical domain. There exists

an OWL implementation of BFO[9] Version 1.1. Version 2 of BFO is under development[10]. The next section describes the arrangement of the global and local ontologies resulting from local terminologies and upper layer ontology.

2) Layering: The Layering step takes as input global/upper and local ontologies; and produces as output layered ontologies. The layering task arranges ontologies into global and local spaces.
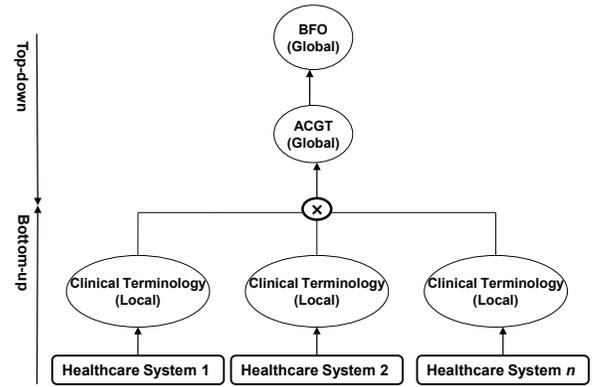

Fig. 2: Layering of global and local ontologies

Figure 2 shows that global ontologies are arranged in a top-down fashion where ACGT extends the BFO ontology. Local ontologies are created from the local clinical terminologies and later aligned with the ACGT ontology. In Figure 2 the circled cross symbol ($\otimes$) means alignment of ontologies. Two conceptual ambiguities exist between local ontologies: (i) semantically similar concepts are named differently; (ii) corresponding concepts are represented at different structural levels. To deal with both issues (*i.e.*, different naming schemes and structural differences) one option is to provide directed alignments between local ontologies. However, directed set of local alignments (*i.e.*, bi-directional alignments) would result in a quadratic size $n \times (n - 1)$ alignment for n number of clinical systems. Therefore, the SEHR ontology employs a hybrid approach of aligning the majority of local terminologies with the global ontologies, thus allowing only minimal set of bi-directional local alignments. This way, we reduced the bilateral correspondences between clinical applications by delegating the majority of alignments with the upper-level (global) ontologies. Further, the meeting points of the "bottom-up" and "top-down" arrows in Figure 2 (left hand side), require extensions within local and global ontologies to find suitable correspondences between them. It is important to note that in addition to the equivalent constructs (*i.e.*, equivalentClass, equivalentProperty), we also consider a "SubClassOf" relation as a mean to align two separate ontologies. The modeling and alignment phases are iterative (see Figure 1) until a consistent set of alignments/correspondences are established. Regardless of resources being (semi-)structured or unstructured, domain and ontology experts are required to sit together to find and verify alignments/correspondences between clinic terminologies and ontological constructs. The next section

---

[9]http://www.ifomis.org/bfo
[10]http://code.google.com/p/bfo/

describes mechanisms and choices for extending both types of ontologies.

3) Local Adaptation: The final Local adaptation step within the Modelling phase takes as input global and local ontologies and has as output extended global and local ontologies meeting local requirements. The notion of local adaptation was first proposed by the DILIGENT methodology [17]. DILIGENT local users adapt the global ontology by introducing local changes, and a central board controls the contents of the global and local ontologies. DILIGENT local ontologies are reengineered versions of global ontologies. In that sense, DILIGENT does not consider different sources that may provide input separately to global and local ontologies. On a similar line, Linked2Safety local ontologies originate from various sources or clinical partners. The local adaptation step is motivated by normal practices where local clinical applications diverge from the standard (or commonly agreed) guidelines by introducing local terminologies.

```
Class: acgt:Syndrome
SubClassOf: acgt:NonInfectiousDiease
Class: NonInfectiousDiease
SubClassOf: acgt:Disease
Class: acgt:Disease
SubClassOf: snap:Disposition
```

Listing 3: Global Ontologies: BFO & ACGT

The local adaptation phase ensures that: (i) local ontologies are generalised enough to resemble the concepts defined in global ontology, and (ii) global ontologies are specialised enough to resemble the concepts defined in the local ontologies. The PPEPR Methodology [11] deals with adaptation of local concepts, *i.e.*, how locally defined concepts fit together with global ontologies. For SEHR ontology, the local adaptation step aligns and extends the local ontologies originating from the three clinical partners.

```
Class: chuv:MetabolicSyndrome
SubClassOf: acgt:Syndrome
```

Listing 4: CHUV Local Terminologies

```
Class: cingBreastCA:BRCA1, cingBreastCA:BRCA2
SubClassOf: acgt:GeneExpression
```

Listing 5: CING Local Terminologies

```
Class: zeincroConcmed:ConcomitantMedication
SubClassOf: acgt:TherapeuticProcess
```

Listing 6: ZEINCRO Local Terminologies

Listing 3 shows a snippet of the ACGT ontology, which describes the top level concepts of local ontologies shown in Listings 4–6. For example, the "Syndrome" class represents all types of syndromes. Further, classes such as "NonInfectiousDisease" and "Disease" are super classes of the "Syndrome" class. We notice that the "SubClassOf" relation between the "acgt:Disease" and "snap:Disposition" classes is an alignment between ACGT and BFO (prefix "snap") ontologies. Similarly, alignments are required between the SEHR and ACGT ontology. For example, a local concept such as "MetabolicSyndrome" can be inherited from any of the classes, *e.g.*, "Syndrome", "NonInfectiousDisease", or simply from the

"Disease" class. To deal with a situation where creating an appropriate alignment between global and local ontologies could be a complex task, the PPEPR Methodology suggests three approaches for the adaptation of the local concept.

Top-Down: Extend the global ontology with more specialised concepts that resemble the concepts defined in local ontologies. However, any extension in the global ontology needs careful analysis on (i) granularity: considering the real purpose of an upper-level (or global) ontology an extension must use domain neutral concepts; and (ii) constraint: classes at the same level in an upper-level ontology (*e.g.*, BFO) are generally described as mutually disjoint. An instance of a class at the lower level might attempt to instantiate sibling classes and cause inconsistencies. For example, in terms of granularity, ACGT represents the breast cancer domain and each of the concepts ("Syndrome" or "NonInfectiousDisease") from the ACGT ontology could be further extended with a lab specific concept like "MetabolicSyndrome" (see Listing 4) without trigging any inconsistencies (*i.e.*, disjointness is between "InfectiousDisease" and "NonInfectiousDisease" concepts).

Bottom-Up: Extend the local ontology with more generalised concepts that resemble the concepts defined in the global ontology. Measuring *connectedness* of two concepts to create a common higher concept is an open research issue [18]. In our case, we measured this through informal means and contexts around them, *i.e.*, all sets of resources (*e.g.*, annotations, data-fields, value space) describing a clinical concept. For example, "BRCA1" and "BRCA2" (see Listing 5) terminologies from CING are described under the genetic category and used for a breast cancer status reporting. Considering the contexts of both the terminologies we used the ACGT class "acgt:GeneExpression" as a super class of 'BRCA1" and "BRCA2".

Middle-Out: Instead of specialising or generalising global or local concepts, another approach could be to add a specialised class as a sibling of similar type of concepts. For example, concepts like "DoseModulation", "Pharmacotherapy", etc. in the ACGT ontology describe a single medical therapeutic process. However, the "ConcomitantMedication" terminology from ZEINCRO describes a therapeutic process that involves two or more therapeutic processes given during the same time period. For example, chemoradiotherapy is the concomitant (combining) of chemotherapy and radiation therapy. Therefore, we aligned the "ConcomitantMedication" class as a sibling of other therapeutic processes (*e.g.*, DoseModulation, Pharmacotherapy, GeneTherapy) describing a therapy that involves multiple treatment methods.

We applied these three approaches in creating bridges (or alignments) with the ACGT and BFO ontologies. These three approaches could be applied independently or in combination depending on requirements from different clinical scenarios. Considering the heterogeneities of different clinical scenarios, we intentionally avoid any fit-for-all suggestion. The next section describes an evaluation strategy and an overall opinion of the clinical experts in terms of SEHR's usability for clinical trail data integration.

IV. EVALUATION

Various approaches have been considered for the evaluation of ontologies [19], [20] depending on the kind and purpose of an evaluation. There is no fit-for-all method for evaluating different kinds of ontologies. SEHR ontology is in its early stage of development, therefore, we decided to use the "Human assessment and conformity with requirements" method of evaluation [20], which is carried out by the clinical experts who seek to verify the adherence of SEHR ontology to certain criteria and patterns. The evaluation process started with detecting any syntactic inconsistencies. We used the OOPS! tool [21], a Web-based tool intended to detect potential syntactic errors in formal ontologies. The initial evaluation concluded with minor pitfalls (*e.g.*, missing annotations, missing domain and/or range in properties). OOPS! helped us to prepare a syntactically clean version of the SEHR ontology.

Further, in order to simplify the evaluation process three different questionnaires (per clinical partner) have been circulated covering two analysis parts (i) correctness and completeness: the questions were designed to detect inconsistencies and weaknesses allowing the clinical experts to express any disagreement, and propose corrections; (ii) usability and simplicity: domain experts were asked to answer a tailored version of the System Usability Scale (SUS) [22] in order to evaluate the understanding and agreement felt about the SEHR ontology. It contains 7 Likert scale questions (stating the degree of agreement or disagreement) [23]. Considering the space limitation, we present the second part of the evaluation, *i.e.*, usability and simplicity of the SEHR ontology depicted in Table II.

The questionnaires were answered by nine clinical experts. The majority of the clinical experts (55.56% indifferent, 44.44% agreement) declared that they could contribute to the SEHR ontology (the question 1 of Table II). Regarding the question 2 of Table II, the answers vary (22.22% disagreement, 33.33% indifferent, 44.44% agreement), but most of the clinical experts found the ontology fairly easy to understand. This is also validated by the fact that several improvements have been suggested in the first part of the evaluation by the clinical experts. The same conclusion derives also from the question 3 where most of the clinical experts needed further theoretical support to be able to understand the SEHR ontology (44.44% indifferent, 55.56% agreement). Similarly, the clinical experts (question 5) agreed that some additional support is required for the clinical community in order to understand the SEHR ontology (22.22% disagreement, 55.56% indifferent, 22.22% agreement). Moreover, most of the clinical experts understand the conceptualisation (the question 6) with 33.33% indifferent and 44.44% agreement. Regarding integration of clinical terminologies (the question 4), the clinical experts found the SEHR ontology well-integrated (44.44% indifferent, 44.44% agreement), however, in few cases suggested corrections in the alignments that fit their local requirements (11.11% disagreement). Similarly, for the completeness (the question 7), the clinical experts agree that SEHR ontology covers the needs of the clinical trial domain (44.44% indifferent, 44.44% agreement) but suggested conceptualisation fixes (*i.e.*, hierarchy and/or names of classes and properties) that suit their local requirements (11.11% disagreement).

Initially, the clinical experts had difficulty understanding the SEHR ontology due to lack of technical background. However, with improvements in definitions (*i.e.*, class name, annotations, labels), it has become clear that with guidance from the developers' side clinical experts can understand the ontology as well as provide clear instructions for its effective development. The evaluation process will continue to provide feedback and in the near future we are aiming to extend the size of the evaluation panel (*i.e.*, 18-20 clinical experts).

V. RELATED WORKS

The work presented in this article can be compared along two dimensions (i) recently proposed clinical trial ontologies; and (ii) ontology building methodologies. There have been a few initiatives to build a clinical trial ontology [15], [24], [25], [16], however, they focused mainly on the globally defined concepts and ignoring the local aspects. Similarly, several ontology building methodologies have been proposed in the last two decades. A series of methods and methodologies for developing ontologies from scratch have been reported in [26] and these can be summarised as follows: In 1990, Lenat and Guha published the general steps [27] and some interesting points about the Cyc development. Some years later, in 1995, on the basis of the experience gathered in developing the Enterprise Ontology, the first ontology building guidelines were proposed in [28]. The methodology METHONTOLOGY [29] appeared at the same time. Some years later, the On-To-Knowledge methodology appeared as a result of the project with the same name [30]. However, all these methods and methodologies do not consider distributed and layered construction of ontologies. In this respect, in 2004, the DILIGENT methodology [17] was proposed. This new methodology was intended to support domain experts in a distributed setting when they need to engineer and evolve ontologies.

All the approaches mentioned above focus on ontologising upper conceptual models of the biomedical and/or clinical trial domain. None of them consider local applications and related issues. The PPEPR Methodology has two development steps that deal with the problem of layered ontologies and how local ontologies could be adapted with an upper ontology. Above all, the works mentioned above lack detailed methodology for ontologising the clinical trial domain where resources are separated in global and local spaces.

VI. CONCLUSION

Integration of clinical terminologies is a difficult task. Domain experts have advocated the use of ontologies in resolving the heterogeneity of clinical terminologies. The various types of ontologies emerged recently differ in their scope and applicability causing a conceptual gap between upper-level, domain-specific, and application-specific ontologies. This conceptual gap is obvious as each domain and/or application ontology is defined at different levels of granularity. We argue that resolving the conceptual gap between ontologies in a distributed setting such the clinical trial domain is not a straightforward equivalent or similarity statement, but needs a mechanism that can support layering and local adaptation of ontologies. We developed SEHR ontology – a clinical trial ontology that is built on a mechanism to incorporate layering and local adaptation of clinical terminologies. SEHR ontology is evaluated by the clinical experts from various sub-domains of the Healthcare

No.	Question	High disagree	Dis-agreement	In-different	Agreement	High agree
1	I think that I could contribute to the SEHR ontology	0.00%	0.00%	55.56%	44.44%	0.00%
2	I find the SEHR ontology easy to understand	0.00%	22.22%	33.33%	44.44%	0.00%
3	I think that I would need further theoretical support to be able to understand the SEHR ontology	0.00%	00.00%	44.44%	55.56%	0.00%
4	I found that the various concepts in the SEHR ontology were well integrated	0.00%	11.11%	44.44%	44.44%	0.00%
5	I would imagine that most clinical experts would understand the SEHR ontology very quickly	0.00%	22.22%	55.56%	22.22%	0.00%
6	I am confident that I understand the conceptualisation of the SEHR ontology	0.00%	22.22%	33.33%	44.44%	0.00%
7	The concepts/properties of the SEHR cover the needs of the clinical trial domain	0.00%	11.11%	44.44%	44.44%	0.00%

TABLE II: Usability Evaluation for the SEHR Ontology

and Life Sciences (HCLS) domain. The majority of clinical experts agree on the uses of SEHR ontology in the clinical trial domain. In the near future, we plan to extend the SEHR ontology with a set of categorical terminologies, *e.g.*, "ExtrapyramidalSideEffects" can be categorised along: 0 = Absent, 1 = Slight, 2 = Mild, 3 = Severe; or "SystolicBloodPressure" can be categorised along:85 = "HighSystolicBloodPressure", 75 = "LowSystolicBloodPressure" depending on patient's age. Our future work also is to extend the SEHR ontology by reusing a set of matching terminologies from the Human Phenotype Ontology [31] and the Human Disease Ontology [11].

REFERENCES

[1] E. Simperl, "Reusing ontologies on the Semantic Web: A feasibility study," *Data Knowledge Engineering*, vol. 68, pp. 905–925, October 2009.

[2] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nature genetics*, vol. 25, no. 1, pp. 25–29, May 2000.

[3] A. L. Rector, J. E. Rogers, and P. A. Pole, "The GALEN High Level Ontology." IOS Press, Jan. 1996, pp. 174–178.

[4] B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler, "Just the right amount: extracting modules from ontologies," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW07. New York, NY, USA: ACM, 2007, pp. 717–726.

[5] V. Presutti and A. Gangemi, "Content Ontology Design Patterns as Practical Building Blocks for Web Ontologies," in *Proceedings of the 27th International Conference on Conceptual Modeling*, ser. ER '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 128–141.

[6] A. J. Yepes, E. J. Ruiz, R. B. Llavori, and D. R. Schuhmann, "Reuse of terminological resources for efficient ontological engineering in Life Sciences," *BMC Bioinformatics*, vol. 10, no. 10, pp. S4+, 2009.

[7] A. L. Rector, R. Qamar, and T. Marley, "Binding ontologies and coding systems to electronic health records and messages," *Applied Ontologies*, vol. 4, no. 1, pp. 51–69, 2009.

[8] P. Lewalle, J. M. Rodrigues, P. E. Zanstra, B. Ustun, D. Kalra, G. Surján, A. L. Rector, V. N. Stroetmann, and M. Virtanen, "A Deployment and Research Roadmap for Semantic Interoperability: the EU SemanticHEALTH project," in *MIE*, ser. Studies in Health Technology and Informatics, S. K. Andersen, G. O. Klein, S. Schulz, and J. Aarts, Eds., vol. 136. IOS Press, 2008, pp. 635–640.

[9] C. Patel, J. J. Cimino, J. Dolby, A. Fokoue, A. Kalyanpur, A. Kershenbaum, L. Ma, E. Schonberg, and K. Srinivas, "Matching Patient Records to Clinical Trials Using Ontologies," in *Proceedings of the 6th International Semantic Web Conference (ISWC)*. Springer, 2007, pp. 816–829.

[10] J. Euzenat and P. Shvaiko, *Ontology Matching*. Heidelberg (DE): Springer-Verlag, 2007.

[11] R. Sahay, R. Fox, A. Zimmermann, A. Polleres, and M. Hauswirth, "A Methodological Approach for Ontologising and Aligning Health Level Seven (HL7) Applications," in *ARES 2011 - Proceedings of the Availability, Reliability and Security for Business, Enterprise and Health Information Systems, Vienna, Austria, August 22-26*. Springer LNCS Series, 2011, pp. 102–117. [Online]. Available: http://axel.deri.ie/publications/saha-etal-2011SALUS.pdf

[12] A. Antoniades, C. Georgousopoulos, N. Forgó, A. Aristodimou, F. Tozzi, P. Hasapis, K. Perakis, T. Bouras, D. Alexandrou, E. Kamateri, E. Panopoulou, K. A. Tarabanis, and C. S. Pattichis, "Linked2Safety: A secure linked data medical information space for semantically-interconnecting EHRs advancing patients' safety in medical research," in *BIBE*. IEEE Computer Society, 2012, pp. 517–522.

[13] M. Horridge and P. F. Patel-Schneider, "OWL 2 Web Ontology Language Manchester Syntax, W3C Working Group Note 27 October 2009," World Wide Web Consortium (W3C), Tech. Rep., Oct. 27 2009. [Online]. Available: http://www.w3.org/TR/2009/WD-owl2-manchester-syntax-20090421/

[14] M. Brochhausen, A. D. Spear, C. Cocos, G. Weiler, L. Martín, A. Anguita, H. Stenzhorn, E. Daskalaki, F. Schera, U. Schwarz, S. Sfakianakis, S. Kiefer, M. Doerr, N. M. Graf, and M. Tsiknakis, "The ACGT Master Ontology and its applications - Towards an ontology-driven cancer research and management system," *Journal of Biomedical Informatics*, vol. 44, no. 1, pp. 8–25, 2011.

[15] S. W. Tu, S. Carini, A. Rector, P. Maccallum, I. Toujilov, S. Harris, and I. Sim, "OCRe: An Ontology of Clinical Research," in *11th International Protg Conference*, Amsterdam, Netherlands, June 2009.

[16] R. Brinkman, M. Courtot, D. Derom, J. Fostel, Y. He, P. Lord, J. Malone, H. Parkinson, B. Peters, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, L. Soldatova, C. Stoeckert, J. Turner, J. Zheng, and the OBI consortium, "Modeling biomedical experimental processes with OBI," *Journal of Biomedical Semantics*, vol. 1, no. Suppl 1, p. S7, 2010.

[17] H. S. Pinto, S. Staab, and C. Tempich, "DILIGENT: Towards a fine-

[11]http://bioportal.bioontology.org/ontologies/1009

grained methodology for Distributed, Loosely-controlled and Evolving Engineering of Ontologies," in *16th European Conference on AI (ECAI)*. IOS Press, 2004, pp. 393–397.

[18] A. C. Varzi, "Spatial Reasoning and Ontology: Parts, Wholes, and Locations," in *Handbook of Spatial Logics*, M. Aiello, I. Pratt-Hartmann, and J. van Benthem, Eds.   Springer, 2007, pp. 945–1038.

[19] M. B. Almeida, "A proposal to evaluate ontology content," *Applied Ontology*, vol. 4, no. 3-4, pp. 245–265, 2009.

[20] J. Brank, M. Grobelnik, and D. Mladenic, "A survey of ontology evaluation techniques," in *In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*, 2005.

[21] M. Poveda-Villalón, M. del Carmen Suárez-Figueroa, and A. Gómez-Pérez, "Validating ontologies with oops!" in *EKAW*, ser. Lecture Notes in Computer Science, A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Aquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez, Eds., vol. 7603.   Springer, 2012, pp. 267–281.

[22] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability evaluation in industry*, P. W. Jordan, B. Weerdmeester, A. Thomas, and I. L. Mclelland, Eds.   London: Taylor and Francis, 1996.

[23] N. Casellas, "Ontology Evaluation through Usability Measures," in *OTM Workshops*, ser. Lecture Notes in Computer Science, R. Meersman, P. Herrero, and T. S. Dillon, Eds., vol. 5872.   Springer, 2009, pp. 594–603.

[24] Y. Megan Kong, C. Dahlke, Q. Xiang, Y. Qian, D. Karp, and R. H. Scheuermann, "Toward an ontology-based framework for clinical research databases," *Journal of Biomedical Informatics*, vol. 44, no. 1, pp. 48–58, Feb. 2011.

[25] R. D. Shankar, S. B. Martins, M. O'Connor, D. B. Parrish, and A. K. Das, "An ontology-based architecture for integration of clinical trials management applications." *American Medical Informatics Association (AMIA) Annual Symposium Proceedings 2007*, pp. 661–665, 2007.

[26] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "Methodologies, tools and languages for building ontologies: where is their meeting point?" *Data Knowledge Engineering*, vol. 46, pp. 41–64, July 2003.

[27] D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project*, 1st ed.   Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1990.

[28] M. Uschold, "Building Ontologies: Towards a Unified Methodology," in *16th Annual Conference of the British Computer Society Specialist Group on Expert Systems*, 1996, pp. 16–18.

[29] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "METHONTOLOGY: from Ontological Art towards Ontological Engineering," in *Proceeding Symposium on Ontological Engineering of AAAI*, Stanford, USA, Salvatore T. March 1997, pp. 33–40.

[30] S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure, "Knowledge Processes and Ontologies," *IEEE Intelligent Systems*, vol. 16, pp. 26–34, January 2001.

[31] P. N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos, "The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease," *American Journal of Human Genetics*, vol. 83, no. 5, pp. 610–5, Nov. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18950739