On the semantic engineering of scientific hypotheses as linked data

Bernardo Gonçalves, Fabio Porto, Ana Maria C. Moura

Extreme Data Lab (DEXL Lab)
National Laboratory for Scientific Computing (LNCC),
Av. Getulio Vargas 333, Petrópolis, Brazil
{bgonc,fporto,anamoura}@lncc.br

Abstract. The term 'hypothesis' is part of the Linked Science Core Vocabulary (LSC) as one of the core elements for making scientific assets explicit and linked in the web of data. Hypotheses are generally understood as propositions for explaining observed phenomena, but eliciting and linking hypotheses can be a challenge. In this paper, we elaborate on a semantic view on hypotheses and their linkage, by striving for minimal ontological commitments. We address the engineering of hypotheses as linked data, and build upon LSC by extending it in order to accommodate terms necessary in model-based sciences such as Computational Science. Then we instantiate the extended LSC by eliciting and linking hypotheses from a published research in Computational Hemodynamics.

Keywords: Linked Science Core Vocabulary (LSC), Scientific Hypothesis, Semantic Engineering, Hypothesis Linkage, Computational Science.

1 Introduction

Data-intensive science has large-scale data management as a key technology for enabling the scientific practice. Nevertheless, there is still significant challenges w.r.t. real-world semantics (meaning) for humans as cognitive agents to be able to browse the data deluge [4]. In this sense, Linked Science emerges as a promising program [11]. It has the potential to enable a semantic-sensitive linkage between scientific assets, providing support to both humans and machines.

Scientists need to access data still in order to formulate and evaluate hypotheses [4]. The term 'hypothesis' is part of the Linked Science Core Vocabulary (LSC) as one of the core elements for linking science [11]. Nevertheless, eliciting and linking hypotheses can be a challenge. Scientific hypotheses are falsifiable statements [19], which are proposed to explain a phenomenon.[1] Let us consider a well-known Einstein's hypothesis to refer as an example. It is identified in Wikipedia as (i) the *mass-energy equivalence*, and presented together with (ii) the famous mathematical equation $E = mc^2$. There can be variations in the formulation of this mathematical expression, yet referring to the same hypothesis. Wikipedia's article[2] is introduced with the sentence "In physics, *mass-energy*

[1] http://en.wikipedia.org/wiki/Hypothesis.
[2] http://en.wikipedia.org/wiki/Mass-energy_equivalence. Access on 7/31/2012.

equivalence is the concept that the mass of a body is a measure of its energy content." This hypothesis is strongly supported by experiments, and explains the transfer of energy and mass as a general phenomenon (cf. Wikipedia's article).

That example illustrates a semantic view on scientific hypotheses that draws on their existence apart from a particular statement formulation in some mathematical framework. The mathematical equation is not enough to identify the hypothesis, first because it must be physically interpreted, second because there can be many ways to formulate the same hypothesis. The link to a mathematical expression, however, brings to the reified hypothesis concept (Wikipedia's entry) higher semantic precision. Another link, in addition, to an explicit description of the explained phenomenon (emphasizing its "physical interpretation") can then (reasonably) succeed in bringing forth the intended meaning.

In this paper, we elaborate on a semantic view on scientific hypotheses and their linkage, by striving for minimal ontological commitments [7]. We address the engineering of hypotheses as linked data, and build upon LSC [11] by extending it in order to accommodate terms necessary in model-based sciences such as Computational Science. We focus on this powerful new scientific discipline [21], where scientific hypotheses are assumptions that constrain the interpretation of observed phenomena for computer simulation. Then we instantiate the extended LSC by eliciting and linking hypotheses in a published research in Computational Hemodynamics [1]. This paper points out the important role hypotheses are to play as conceptual entities in Linked Science [11] and theory-driven eScience [2].

The paper is organized as follows. In Section 2 we comment on a related work background, and in Section 3 we introduce hypotheses in Computational Hemodynamics. In Section 4 we present our semantic view on hypotheses and its engineering in Linked Science by extending LSC. This section is fully illustrated with examples from Computational Hemodynamics. In Section 5 we instantiate the extended LSC in a published research on the modeling and simulation of the human cardiovascular system. Finally, in Section 6 we conclude the paper.

2 Related Work

The HyBrow (Hypothesis Browser) conceptual framework [20] addresses hypothesis modeling in Bioinformatics. It aims at providing biologists with a unified eScience infrastructure for hypothesis *formulation* and *evaluation* against observed data. HyBrow is based on an OWL ontology and application-level rules to contradict or validate hypothetical statements. As an upgrade of HyBrow, the HyQue [3] framework adopts linked data technologies and employs Bio2RDF to add to HyBrow semantic interoperability capabilities. HyBrow/HyQue's hypotheses are domain-specific statements that correlate biological processes (seen as events) in First-Order Logic (FOL) with free quantifiers. Hypothesis formulation in HyBrow/HyQue is constrained to a FOL-based model-theoretic semantics in favor of hypothesis evaluation. Our point, nevertheless, is that such requirement could also be met without hardwiring hypothesis *modeling* and *encoding* (cf. [13]). In our work, we strive for eliciting and linking hypotheses as *conceptual* entities and capitalize on the *co-existance* of different formulations (possibly in different languages) of the same hypothesis on the web.

LSC provides core terms for making scientific assets explicit and linked in the web of data.[3] In [11, p. 12], Kauppinen et al. instantiate LSC to a research in Environmental Conservation that investigates "the *notion* that hazards to Amazonian forests have declined over the last decade" [16] (the *emphasis* is ours). In fact, by dealing with that hypothesis as a conceptual entity, the scientists make it possible to change its statement formulation or even to assert a semantic mapping to another incarnation of the hypothesis in case someone else reformulates it. The preservation of the hypothesis conceptual identity is particularly interesting in that case, since it can then be tracked in public affairs.

Brodaric et al.'s Science Knowledge Infrastructure ontology (SKIo) [2] is a foundational work to leverage data-driven eScience to a theory-driven paradigm. SKIo extends the top-level ontology DOLCE [17], following a top-down approach to characterize concepts of the scientific method. SKIo aims at providing ontological distinctions of terms like 'theory', 'law', 'problem', which can appear in different contexts with subtle different meanings. This qualifies SKIo as a *reference ontology* (in the sense of Guarino [8, p. 5], also called *foundational* [17, p. 3]) for scientific knowledge representation. On the one hand, a well-founded, fine-grained ontology such as SKIo can be used to support meaning negotiation in science, enabling the semantic interoperability of scientific assets. On the other hand, SKIo's top-down ontology engineering approach requires from scientists (as independent knowledge engineers) to subscribe to abstract ontological commitments [10]. This justifies the need for a *lightweight ontology* [17, p. 2], a second kind of ontology, like LSC, to serve as a shareable, minimally laden vocabulary that fits the needs of a community (cf. [7]). One can take benefits of both artifacts by instantiating the lightweight ontology, yet by referring to the reference one as an interlingua. This is sought for in our work, which considers SKIo as a reference ontology and LSC as a lightweight ontology for Linked Science. In this paper we concentrate on extending and instantiating LSC for realizing Linked Science in a research. An alignment of the extended LSC to SKIo can be addressed in future work for the semantic interoperability of scientific assets in different researches. Then the technique for hypothesis linkage introduced here (see Section 4.3) shall be extended to map hypotheses in different researches, in support of the (decentralized) growth of scientific knowledge on the web.

Next section introduces hypotheses in Computational Hemodynamics. In Computational Science, state-of-the-art models are the vehicle of several entangled hypotheses about a studied phenomenon [21]. The terms 'hypothesis' and (modeling) 'assumption' are used interchangeably in that field. We then stick to LSC's minimal commitment in the definition of lsc:Hypothesis as "any kind of hypothesis" [11], and do not distinguish assumptions from (say) laws, empirical regularities, (under-)theories—for such distinctions, refer to SKIo [2].

3 Hypotheses in Computational Hemodynamics

Among plenty of natural phenomena that are addressed by research groups from our institution, we have chosen to work in this paper with Hemodynamics. The

[3] Version 11/29/2011. Available at `http://linkedscience.org/lsc/ns-20111129`.

reason is that the sheer complexity of the human cardiovascular system (CVS) stresses the nature and role of hypotheses in the formulation of a complex mathematical model. The computational modeling of blood flow in vascular vessels can support investigations about the development of pathologies [5]. A model used to simulate such a phenomenon has to be simple enough to allow for a numerical treatment at reasonable computational costs. Yet, it has to provide all the information that is essential for their comprehension. A relevant feature of the phenomena of in CVS is their "multiscale" nature both with respect to time and space variables. Blood vessels in different regions of CVS vary significantly in terms of their diameter, wall thickness, elasticity, etc. As of time scales, the long-term formation of atherosclerotic plaques can be a response of the vascular tissue to specific stresses induced by the blood during heart beats ($\approx 0.8\,s$). A plaque developed, e.g., in the carotid artery, could affect the blood flow rate to the brain and change the overall circulation in CVS.

The challenge of CVS modeling with multiscale techniques is addressed in the literature by several groups, one of which leading the HemoLab project at our institution.[4] In their published research papers, an implicit complex hypothesis is typically formulated as a sophisticated mathematical model, with the implicit meaning that the model is fit in simulating the phenomenon of interest. That final, synthesized hypothesis formulated as an effective model can be used to make predictions. For this reason we shall refer to it henceforth as a hypothetico-deductive (H–D) system [9] whenever is relevant to distinguish it as such.

Nonetheless, it is worth highlighting that a H–D system is only the scientist's finished work [9] and can hardly be grasped (neither is it described) at once. Instead, it is worked out in modeling steps as the scientist goes back and forth by assuming and revising simpler hypotheses and assembling them together. This is something important to be considered for the sake of reproducibility. We shall refer from now on to hypotheses in Computational Hemodynamics throughout this paper to illustrate our semantic engineering of hypotheses as linked data.

4 Semantic Engineering of Scientific Hypotheses

A scientific hypothesis is a falsifiable statement [19]. That is, it must be prone to be either supported or refuted by observation and experimentation. Another point to note is that in model-based sciences, a formal language with some notation constitutes the technical manner to express hypotheses as *models*, while non-formal expressions like image sketches or natural language itself are used as more flexible alternatives to convey meaning in papers, books, conversations.

4.1 A Semantic View on Hypotheses

In a careful examination on what a scientific hypothesis is, we note that (i) its falsifiability grounds it in the *observable* world, while (ii) its statement *formulation* allows to be assigned for it truth values. An additional feature we should add still is that (iii) it comprises the scientist's *interpretation* of the observed phenomenon [9], and this third feature brings forth the hypothesis' conceptual

[4] `http://macc.lncc.br`.

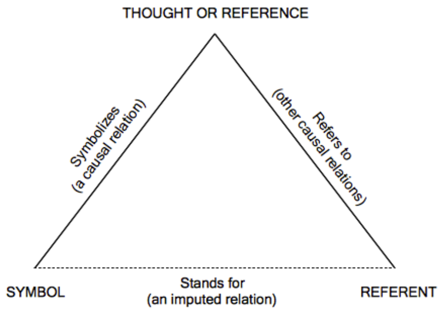

Fig. 1. Ogden and Richards' meaning (or semiotic) triangle (from [12] apud. [18]).

nature which is important for semantic interoperability. This hypothesis three-fold notion can be compared to Ogden and Richards' meaning (or semiotic) triangle (see Fig. 1) [18], if we consider the *hypothesis* conceptual identity (thought or reference), its statement formulation as a *model* that can be evaluated (symbol), and its reference to a *phenomenon* that can be observed or measured (referent).

Ogden and Richards' meaning triangle has been adopted by Kuhn in his characterization of *concept* for the purpose of semantic engineering [12], which fits well to this semantic view on scientific hypotheses. In Linked Science, as we commented on the Einstein's hypothesis example, all of those three corners of the triangle are relevant to be made explicit. In next section we elaborate on the design of this semantic view on hypotheses in the framework of Linked Data.

4.2 Hypotheses as Linked Data

We design our semantic view on scientific hypotheses as a Model–Hypothesis–Phenomenon triad and call it the hypothesis triangle (see Fig. 2). Data and phenomenon-related entities (viz., Region, Time, and Observable) ground the hypothesis triangle in the observable world. All of those entities are RDF resources by design.

With the hypothesis triangle, scientists are able to express themselves in multiple co-existing forms on each of its three corners. This is captured by assigning to them RDF properties. For a prompt example, let us consider a hypothesis in Computational Hemodynamics (see Fig. 3). In the current state-of-the-art [1, 5], the blood flow in the microvascularity (say) of the hands is assumed to behave analogously to an electrical circuit: a resistor-capacitor connection in parallel (standing for the flow in the arterioles), in series with another resistor (standing for the flow in the capillaries). The rationale is that the blood flows like an electrical current. The arterioles' wall tissue absorbs ("dissipates") it, while still stretching itself (locally accumulating blood) in response to a blood pressure gradient ("voltage"). The capillaries in turn have a very small diameter, for which deformation is neglectable w.r.t. the resistance to the flow. In this illustration (Fig. 3), we are using known terms such as rdfs:label, rdf:value, dc:description and foaf:depiction as RDF properties. These, once arranged together, can all be worth as expressions of hypotheses, models, and phenomena. The convention of

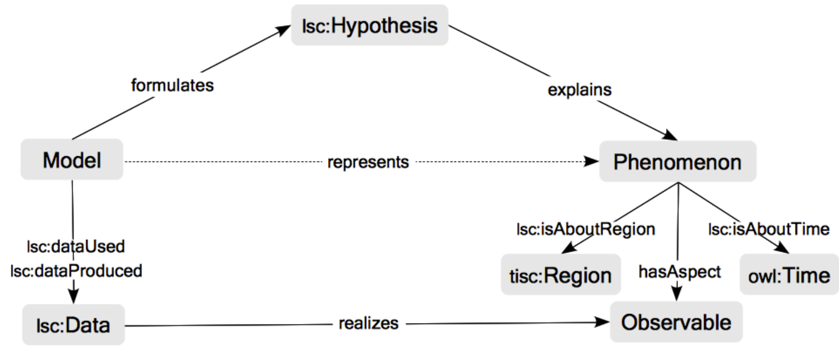

Fig. 2. A triadic notion of lsc:Hypothesis (thought or reference), which explains a Phenomenon (referent) and is formulated as a Model (symbol). Terms with a namespace as prefix come from LSC (version 11/29/2011) or from its imported vocabularies. The others are proposed here to account for hypotheses in model-based sciences with no loss in generality.

proper RDF properties for the RDF resource Model can benefit further from ontologies for representing mathematics on the semantic web [15].

The conceptualization shown in Fig. 2 extends LSC by striving for minimal ontological commitments.[5] Three terms are new, namely, Phenomenon, Observable and Model. They come with five additional relational terms: explains, formulates, represents, hasAspect and realizes; all of them are potentially $n \times n$. The hypothesis triangle relations explains, formulates, represents turn out to be functional in the scientist's final decision in adopting a particular model m_1 to formulate a hypothesis h_1, which is meant to explain phenomenon p_1. To anticipate next section, all that lies within the scope of a research, where such instances are to be made semantically explicit. The represents link is dashed to point out that its instances do not have to be asserted, since they can be inferred by a rule, namely, for all $\langle m, h, p \rangle \in M \times H \times P$, if formulates($m$, h) and explains(h, p), then represents(m, p); where M, H and P are sets of models, hypotheses and phenomena, respectively. In model-based sciences, it is such a triple $\langle m, h, p \rangle$ that can afford to convey a scientific hypothesis unambiguously.

A Phenomenon, as originally defined by Kant, is "any observable occurrence," which is distinguished from 'noumenon' (thing-in-itself, not directly accessible to observation).[6] A Phenomenon isAboutRegion and isAboutTime, and it hasAspect Observable. We adopt the term Observable, differing to (say) 'physical quantity', in order to refer to the quantifiable observable world but still cover non-quantities such as genes and astronomical objects.[7] The term Data appears as a core element in LSC. We then add the link realizes to Observable in order to tie up the

[5] In particular, we have strived to make it possible for computational scientists to promptly recognize and instantiate this extended LSC in their research.

[6] http://en.wikipedia.org/wiki/Phenomenon.

[7] Although in this paper we focus on Computational Science, we have strived not to restrict this conceptualization to that discipline.

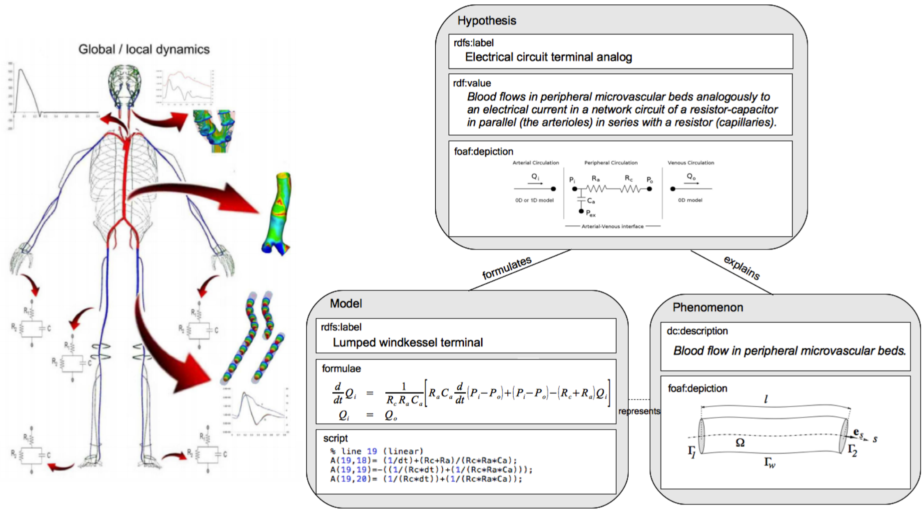

Fig. 3. On the left, an illustration of the CVS emphasizing the multiscale modeling for different regions of the system. On the right, a hypothesis explaining the blood flow in the microvascularity of peripheral beds as an instance of the hypothesis triangle.

hypothesis triangle according to its grounding in the measurements of observables. Refutation can be considered a function $\rho : H \rightarrow [0, 1]$, as a measure of the distance between data produced (model output, or, conceptually, hypothesis predictions) and data used (model input, or, conceptually, phenomenon observations). This metrics can be designed as an RDF datatype property refutation to be assigned by the scientist users in order to explicitly assess the quality of their hypotheses (models) in explaining (representing) their observed phenomena.

4.3 Linking Hypotheses in a Local Research

To make only a final hypothesis (a H–D system [9]) explicit may not provide much *conceptual traceability* about a research (ibid.). This issue can be addressed, however, by eliciting simpler hypotheses and linking them properly in the derivation of more complex ones. While *complex* hypotheses are formed by combining two or more hypotheses already assumed, *atomic* hypotheses constitute for the scientist a single unit of thought either because it has been borrowed from another research or because it has been assumed at once from scratch.

Nevertheless, hypotheses (e.g., in Mathematical Modeling) are entangled in such a way that the primitive can no longer be identified in the resulting one. Therefore, we do not attempt to prescribe any logical structure for hypothesis combination. Rather, we borrow prov:wasDerivedFrom[8] as a semantically lightweight relation under a notion of provenance and consider that a complex hypothesis is a blend of others. We use prov:wasDerivedFrom as an ordering relation to make up a data structure for hypothesis linkage as follows.

[8] From the PROV Ontology, available at `http://www.w3.org/TR/prov-o/`.

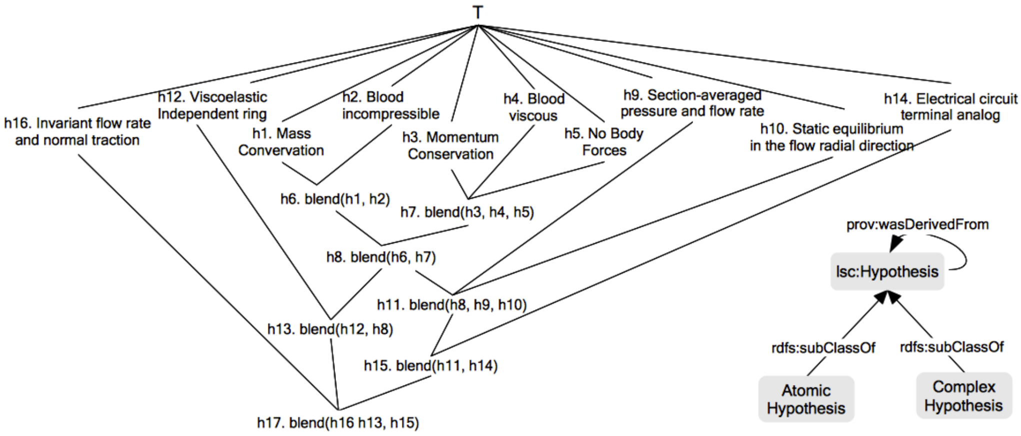

Fig. 4. On the lower right, the atomic/complex hypothesis distinction, and prov:wasDerivedFrom for hypothesis linkage. A hypothesis lattice is formed by considering a set of hypotheses equipped with prov:wasDerivedFrom as a strict order $<$ (from the bottom to the top). Hypotheses directly derived from exactly one hypothesis are atomic, while those directly derived from at least two hypotheses are complex.

Def. 1 *Let H be the set of hypotheses in a local research, and $<$ be a strict order (asymmetric and transitive). For all $h_1, h_2 \in H$, we write $h_1 < h_2$ if h_1* ***was derived from*** *h_2. More specifically, if $h_1 < h_2$ and for no $h \in H$, $h_1 < h < h_2$, then we write $h_1 \prec h_2$ and say that h_1 was directly derived from h_2.*

Def. 2 *We call $h \in H$ an* ***atomic hypothesis*** *if there exists exactly one $h_1 \in H$ such that $h \prec h_1$. Otherwise, we call $h \in H$ a* ***complex hypothesis*** *if there exists at least two hypotheses $h_1, h_2 \in H$ such that $h \prec h_1$ and $h \prec h_2$.*

Def. 3 *There is a special hypothesis $h_0 \in H$, such that for all $h \in H \setminus \{h_0\}$, $h < h_0$. We call h_0 H's* ***top hypothesis****. (trivially, h_0 assumes nothing).*

Our design approach for hypothesis linkage turns out to form a lattice data structure [22], but both its formalization and the semantics of hypothesis blending fall outside the scope of this paper. Fig. 4 shows the hypothesis lattice which comprises the H–D system (h17) of our application case in Computational Hemodynamics (cf. Section 5). Symbol '$\top$' is used as a label for the top hypothesis h0. The hypothesis shown in Fig. 3 is h14 in this hypothesis lattice.

4.4 Links of the Hypothesis Triangle as Morphisms

From the design solutions presented in the two previous sections, we obtain an interesting framework for linking scientific assets in a research. For example, Fig. 5 shows a complex hypothesis h8 for explaining a general phenomenon of fluid behavior (p8) which is present in our application case (where the fluid is human blood). This hypothesis, as a H–D system for explaining p8, is the bottom element of the hypothesis lattice shown on the top center in Fig. 5. The

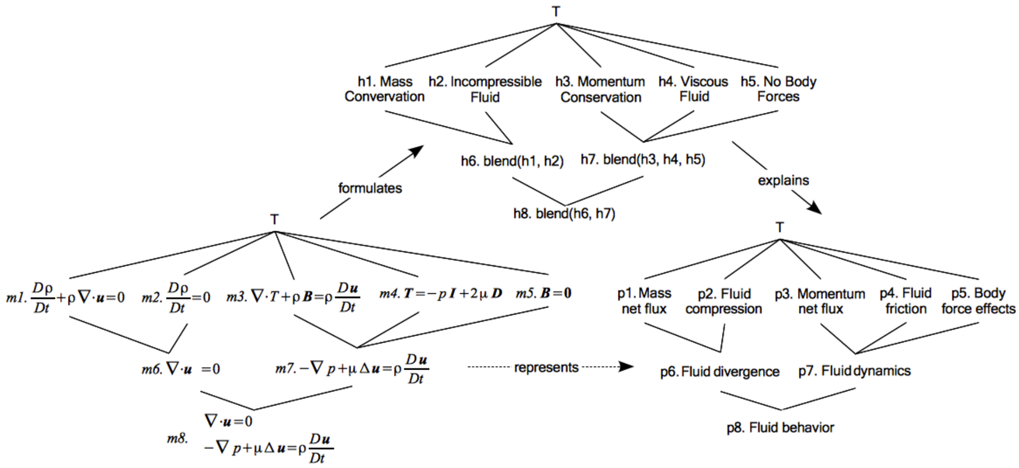

Fig. 5. Hypothesis lattice unfolded into model and phenomenon isomorphic lattices. Model m1 formulates hypothesis h1, which explains phenomenon p1. Similarly, m2 formulates h2, which explains p2, and so on.

hypothesis lattice is unfolded into model and phenomena isomorphic lattices according to the hypothesis triangle (Fig. 2).[9] The lattices are isomorphic if one takes subsets of M, H and P such that formulates, explains and represents are both one-to-one and onto mappings (i.e., bijections), seen as structure-preserving mappings (morphisms). This turns out to be the case in published research papers where scientists propose exactly one hypothesis to explain exactly one phenomenon of interest, and formulate the former in exactly one way.

We are eliciting in Fig. 5, for example, the hypotheses underlying the so-called continuity equation (m1) and the Navier-Stokes equations (m7). These are standard models in the study of fluid mechanics [14] and they are applied in our case in Computational Hemodynamics. The hypothesis lattice shown in Fig. 5 (on the top center) is a sublattice of the hypothesis lattice shown in Fig. 4. Deductions from h8 as a H–D system can be too coarse for predicting conditions of blood flow in vascular vessels, for which predictions from h17 can be adequate.

5 Instantiation of the Extended LSC

In this section we present a published research in Computational Hemodynamics as an instantiation of the extended LSC proposed here. The research we instantiate is reported in the article "*On the potentialities of 3D-1D coupled models in hemodynamics simulations*" by Blanco et al. [1] from our institution. The article elaborates on the potential of such coupled models (introduced preliminarily by Formaggia et al. [5]) for predicting two hemodynamics conditions: (i) the sensitivity of local blood flow in the carotid artery to the heart inflow condition, and (ii) the sensitivity of the cardiac pulse to the presence of an aneurysm. A

[9] We are using symbol '⊤' to denote a top hypothesis, and abusing this notation slightly to mean the same (isomorphically) for the model and phenomenon lattices.

representative account of the extended LSC instantiation on that research is described in Table 1. The scientific assets presented in Table 1 are linked according to the conceptualization shown in Fig. 2.

In particular, Table 1 includes the final complex hypothesis h17 which is Blanco et al.'s H–D system as shown in Fig. 4. With the extended LSC, we can provide scientists with interesting querying functionalities on the web. SPARQL queries can select, e.g., all the atomic hypotheses built into h17, or all the model-hypothesis-phenomenon triples in a research. We present below a SPARQL query Q1 selecting a particular triple $\langle m, h, p \rangle \in M \times H \times P$, namely, the one shown in Fig. 3. It exemplifies a scientist interested in Blanco et al.'s research.

Q1. Find in Blanco et al.'s research a hypothesis (if any) explaining phenomena of blood flow in microvascular vessels and show which model formulates it.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX lsc: <http://linkedscience.org/lsc/ns#>
SELECT ?hypothesis_name ?model_name
WHERE {
?h rdfs:label ?hypothesis_name . ?m rdfs:label ?model_name .
?h a lsc:Hypothesis . ?p a lsc:Phenomenon . ?m a lsc:Model .
?h lsc:explains ?p . ?m lsc:formulates ?h .
?p dc:description ?d .
FILTER regex(?d, "blood flow", "i") . FILTER regex(?d, "microvascular", "i")
}
-------------------------------------------------------------------------------------------
| hypothesis_name | model_name |
===========================================================================================
| "Electrical circuit terminal analog" | "Lumped windkessel terminal" |-----------------
```

It is worth now to draw attention to our initial motivation w.r.t. the semantic engineering of hypotheses in the context of data-intensive science. The datasets in Table 1 are linked to model m17, which is in turn linked to hypothesis h17. The latter can be an interpretation key to the research, and to those datasets in particular. This can be an interesting line of thought to be investigated further by developing querying patterns that bind hypotheses to data.

Recall that hypothesis evaluation is not addressed in this paper. But as we discuss elsewhere [6], state-of-the-art scientific workflow systems can be extended to manage hypotheses and models. The numerical methods which are necessary for computing model m17 find their place under the term lsc:Method. To cope with them, however, is another challenge and it falls out the scope of this paper. The Linked Science program provides a proper framework for overcoming such a limitation under the partial knowledge view of Linked Data. It allows the level of detail of a published research to be improved in a stepwise manner.

6 Conclusions

In this paper we have elaborated on a semantic view on scientific hypotheses and their linkage, by striving for minimal ontological commitments. We have addressed the engineering of hypotheses as linked data by extending LSC and instantiating it in a research in Computational Hemodynamics.[10]

[10] This extension has been proposed to LSC's authors and is under consideration to be incorporated into a next version of it (the current one dates to 11/29/2011) to be available at `http://linkedscience.org/lsc/ns/`.

Table 1. Representative set of scientific assets of a research in Computational Hemodynamics as an instantiation of the extended LSC.

rdfs:Class	rdf:Resource $\rightarrow$	rdf:Literal
lsc:Researcher	authors1 $\xrightarrow{\text{rdf:value}}$	"P.J. Blanco, M.R. Pivello, S.A. Urquiza, and R.A. Feijóo."
lsc:Research	research1 $\xrightarrow{\text{dc:description}}$	"Simulation of hemodynamic conditions in the carotid artery."
lsc:Publication	pub1 $\xrightarrow{\text{dc:title}}$	"On the potentialities of 3D–1D coupled models in hemodynamics simulations."
lsc:Data	dataset1 $\xrightarrow{\text{dc:description}}$	"Flow rate of 5.0 l/min as an inflow boundary condition at the aortic root, in observation of Avolio (1980) and others."
lsc:Data	dataset2 $\xrightarrow{\text{dc:description}}$	"1D mechanical and geometric data from Avolio (1980)."
lsc:Data	dataset3 $\xrightarrow{\text{dc:description}}$	"MRI images processed for reconstructing the 3D geometry of both the left femoral and the carotid arteries."
Phenomenon	p17 $\xrightarrow{\text{dc:description}}$	"Blood flow in the carotid artery."
tisc:Region	region1 $\xrightarrow{\text{dc:description}}$	"The carotid artery, a part of the human CVS."
owl:IntervalEvent	beat1 $\xrightarrow{\text{dc:description}}$	"A heart beat with period $T = 0.8\,s$."
Observable	ob1 $\xrightarrow{\text{dc:description}}$	"Blood flow rate."
Observable	ob2 $\xrightarrow{\text{dc:description}}$	"Blood pressure."
lsc:Hypothesis	h17 $\xrightarrow{\text{rdfs:label}}$	"blend(h13, h15, h16)"
Model	m17 $\xrightarrow{\text{dc:description}}$	"3D-1D coupled model with lumped windkessel terminals."
lsc:Data	dataset4 $\xrightarrow{\text{dc:description}}$	"Plots of hemodynamic observables in the left femoral artery produced to validate the hypothesis."
lsc:Data	dataset5 $\xrightarrow{\text{dc:description}}$	"Plots of hemodynamic observables in the carotid artery."
lsc:Data	dataset6 $\xrightarrow{\text{dc:description}}$	"Scientific visualization of hemodynamic observables in the left femoral artery produced to validate the hypothesis."
lsc:Data	dataset7 $\xrightarrow{\text{dc:description}}$	"Scientific visualization of hemodynamic observables in the carotid artery both with and without aneurism."
lsc:Prediction	predict1 $\xrightarrow{\text{rdf:value}}$	"Sensitivity of local blood flow in the carotid artery to the heart aortic inflow condition."
lsc:Prediction	predict2 $\xrightarrow{\text{rdf:value}}$	"Sensitivity of the cardiac pulse to the presence of an aneurysm in the carotid."
lsc:Conclusion	conclusion1 $\xrightarrow{\text{rdf:value}}$	"3D-1D coupled models allow to perform quantitative and qualitative studies about how local and global phenomena are related, which is relevant in hemodynamics."

In our work we have taken a direction tailored not to reduce hypotheses to a rigid logical structure, but to seek for them proper forms of expression as linked data. In this way, our approach allows for the co-existance of hypotheses and their formulations over multiple scientific domains and formalisms. The very problem of hemodynamics multiscale modeling is an astonishing example of hypotheses co-existance across multiple scales and the boundaries of disciplines.

We have shown that an effort in eliciting and linking of hypotheses can be rewarded with interesting functionalities in terms of conceptual traceability. The hypothesis lattice (see Fig. 4) is a data structure meant for the management of hypothesis evolution, as the scientist user operates over it by reflecting her cognitive operations on the scientific problem at hand. We aim at providing scientists with such a tool. We are developing an algebraic specification of abstract data types such as model, hypothesis and phenomenon for scientists to operate over on the web—e.g., by assuming, borrowing and revising hypotheses.

Significant effort still has to be carried on until we have sophisticated computational models reproducible online. This work is a step towards conceptual traceability, which might open some seaways for sailing on the big data [4].

Acknowledgments.
We thank three anonymous reviewers for providing valuable comments on a previous version of this manuscript, and LSC's authors Tomi Kauppinen, Alkyoni Baglatzi and Carsten Kessler for their openness and feedback w.r.t. our extension proposal to LSC. The first author would like to express his gratitude to José Karam Filho for delivering insightful lectures on the principles of mathematical modeling that contributed to ideas of this paper. This research received financial support from the National Research Council (CNPq) under grants no. 141838/2011-6, 309502/2009-8 and 382489/2009-8.

References

1. P. Blanco, M. Pivello, S. Urquiza, and R. Feijóo. On the potentialities of 3D–1D coupled models in hemodynamics simulations. *J. Biomech.*, 42(7):919–30, 2009.
2. B. Brodaric, F. Reitsma, and Y. Qiang. SKIing with DOLCE: Toward an e-Science knowledge infrastructure. In *Proc. of FOIS'08*, pages 208–19, 2008.
3. A. Callahan, M. Dumontier, and N. H. Shah. HyQue: Evaluating hypotheses using semantic web technologies. *Journal of Biomedical Semantics*, 2(Suppl 2):S3, 2011.
4. J. P. Collins. Sailing on an ocean of 0s and 1s. *Science*, 327(5972):1455–6, 2010.
5. L. Formaggia et al. Multiscale modelling of the circulatory system: A preliminary analysis. *Comput. Vis. Sci.*, 2(2-3):75–83, 1999.
6. B. Gonçalves, F. Porto, and A. Moura. Extending scientific workflows for managing hypotheses and models. In *Proc. of the 6th Brazilian eScience Workshop*, 2012.
7. T. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum-Comput St*, 43(5-6):907–28, 1995.
8. N. Guarino. Formal Ontology and Information Systems. In *Proc. of FOIS'98*, pages 3–15, 1998.
9. N. R. Hanson. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science.* Cambridge University Press, 1958.
10. K. Janowicz. Observation-Driven Geo-Ontology Engineering. *Transactions in GIS*, 16(3):351–74, 2012.
11. T. Kauppinen, A. Baglatzi, and C. Kessler. *Data Intensive Science*, chapter Linked Science: Interconnecting Scientific Assets. CRC Press, 2012.
12. W. Kuhn. *Research Trends in Geographic Information Science, part 1*, chapter Semantic Engineering, pages 63–76. LNG&C. Springer, 2009.
13. W. Kuhn. Modeling vs Encoding for the Semantic Web. *Semantic Web*, 1(1-2):11–15, 2010.
14. W. M. Lai, D. Rubin, and E. Krempl. *Introduction to Continuum Mechanics.* Elsevier, 4th edition, 2009.
15. C. Lange. Ontologies and languages for representing mathematical knowledge on the semantic web. *Semantic Web (to appear)*, 2012.
16. W. F. Laurance, A. K. M. Albernaz, and C. D. Costa. Is deforestation accelerating in the Brazilian Amazon? *Environmental Conservation*, 28(4):305–11, 2001.
17. C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Ontology Library: WonderWeb Deliverable D18. Technical report, ISTC-CNR, 2003.
18. C. Ogden and I. Richards. *The meaning of meaning.* Harcourt, 8th edition, 1948.
19. K. Popper. *The logic of scientific discovery.* Routledge, 2nd edition, 2002.
20. S. Racunas, N. Shah, I. Albert, and N. Fedoroff. Hybrow: a prototype system for computer-aided hypothesis evaluation. *Bioinformatics*, 20(1):257–64, 2004.
21. P. M. A. Sloot. The cross-disciplinary road to true computational science. *Journal of Computational Science*, 1(3):131, 2010.
22. J. Sowa. *Knowledge Representation.* Brooks / Cole, 1st edition, 1999.