Towards a Semantic Representation for Multi-Scale Finite Element Biosimulation Experiments

Andre Freitas, Margaret Jones, Kartik Asooja, Christos Bellos, Steve Elliott,
Stefan Stenfelt, Panagiotis Hasapis, Christos Georgousopoulos,
Torsten Marquardt, Nenad Filipovic, Stefan Decker, Ratnesh Sahay

Abstract— Biosimulation researchers use a variety of models, tools and languages for capturing and processing different aspects of biological processes. However, current modeling methods do not capture the underlying semantics of the biosimulation models sufficiently to support building, reusing, composing and merging complex biosimulation models originating from diverse experiments. In this paper, we propose an ontology based and multi-layered biosimulation model to facilitate researchers to share, integrate and collaborate their knowledge bases at Web scale. In particular, we investigate the semantic biosimulation model under the context of the multi-scale finite element (FE) modelling of the inner-ear. The proposed ontology-based biosimulation model will provide a homogenized and standardized access to the shared, semantically integrated and harmonized datasets for clinical data (histological data, micro-CT images of the cochlea, pathological data) and inner ear FE simulation models. The work presented in this paper is analyzed and designed as part of the SIFEM EU project.

I. INTRODUCTION

Scientific investigation is evolving in the direction of the creation of large-scale, highly complex, multi-domain and multi-scale scientific models and theories. The interpretation of complex phenomena and the creation of scientific models in these scenarios demand the systematic orchestration of experimental design, experimental execution, observation, data representation, data analysis and knowledge exchange. In this work, we study the problem of multi-scale finite element modelling of the inner-ear. Mathematical modelling and simulation is particularly attractive as a tool in researching the cochlea and its pathology as biopsy, surgical excision and other conventional techniques of pathologic studies are not feasible without further impairing function [1]. However, experiments and modelling done in the study of inner-ear may cross different domains of expertise, may depend on large collaboration networks and may involve large-scale data outputs.

Recent advancement of technologies has given increasing computational power to biomedical researchers for building and managing complex biosimulation systems. However, as their model and simulations grow in complexity, researchers find it more difficult to share, manage, and edit their models. The problem of sharing complex biomedical models becomes severe with the lack of scalable standards for model representation and reuse: as a consequence many modellers cannot readily share and build upon previously coded models. The biosimulation community therefore has a growing need for tools that will help them to more efficiently build, manage and reuse their models.

In this paper, we discuss how Semantic Web technologies can be applied in integrating different types of data associated with inner-ear models (physical, numerical, anatomical, clinical), representing models parameters at different scales, and representing the associated simulation outcome for each case. Semantic Web technologies provide a mechanism for representing and linking heterogeneous data, using Web standards and a flexible and expressive data model.

Semantic Web technologies can provide a principled representation which supports the description of data together with its meaning in a machine interpretable way. The use of a standardized data representation format based on Web standards, and the construction of shared vocabularies and ontologies support semantic interoperability of data across models and systems. While a principled data representation brings the basis for addressing the problems outlined above, it demands the support of a complementary service infrastructure for allowing the capture, accessibility, analysis and support in the interpretation of experimental data. This work concentrates on the discussion of the representation aspect. The work presented in this article is developed as part of the SIFEM EU project[1]. The SIFEM project results in the delivery of an infrastructure, which will semantically integrate finite element simulation data with experimental and clinical data, aiming towards the delivery of a robust

A. Freitas, M. Jones, K. Asooja, S. Decker, R. Sahay are with the DERI National University of Ireland Galway, Ireland (e-mail: andre.freitas, marggie.jones, kartik.asooja, stefan.decker, ratnesh.sahay@deri.org).

C. Bellos is with the Institute of Communications and Computer Systems, Athens, Greece (e-mail: cbellos@biosim.ntua.gr).

S. Elliott is with the Institute of Sound and Vibration Research, University of Southampton, United Kingdom (e-mail: S.J.Elliott@soton.ac.uk).

S. Stenfelt is with Linköping University, Linköping, Sweden (e-mail: stefan.stenfelt@liu.se).

P. Hasapis, C.Georgousopoulos, are with the INTRASOFT International S.A., Luxembourg (e-mail: Panagiotis.Hasapis, Christos.Georgousopoulos@intrasoft-intl.com).

T. Marquardt is with University College London - Ear Institute, London, United Kingdom (e-mail: t.marquardt@ucl.ac.uk).

N. Filipovic is with the Bioengineering and Development Centre, BioIRC, Kragujevac, Serbia (e-mail: fica@kg.ac.rs).

[1]http://sifem-project.eu

multi-scale model of the inner-ear. This paper is organized as follows: section II presents the challenges involved in the creation of multi-scale finite element biosimulation models; section III describes existing works on the use of ontologies for finite element analysis and multi-scale integration of biosimulation models; section IV describes the semantic representation and its different conceptual model dimensions for FE biosimulation experiments, which is followed by section V, where we present the conclusions.

II. MULTI-SCALE INNER-EAR MODELLING

The creation of a realistic model of the inner ear function and behavior is a complex scientific task which depends on the understanding of different structures which crosses the boundaries of different domains of expertise. The hearing process involves the orchestration between anatomical structures which have different scales and are governed by different physical principles.

The cochlea consists of a coiled labyrinth, which is about 10mm across and has about 2.5 turns in humans, embedded in the temporal base of the skull. It is filled with fluid and divided into three main fluid chambers: the scala vestibuli, the scala media and the scala tympani. The basilar membrane (BM) is a structure that runs along the coil of the cochlea, which separates the scala media and the scala tympani chambers.

The motion in the cochlea is driven by the middle ear, via a flexible window at the basal end of the upper fluid chamber, and the pressure at the basal end of the lower fluid chamber is released by another flexible window. It is thus the difference in pressure between the upper and lower fluid chambers that drives the BM.

The organ of Corti sits on top of the BM, and contains two types of hair cells (HC): (i) inner hair cell (IHC), which converts the motion of the stereocillia into neural impulses that then pass up the auditory pathway into the brain and (ii) outer hair cells (OHC) within the organ of Corti that play a more active role in the mechanics of the cochlea.

The physical models describing distinct structures within the inner ear follow different physical descriptions which are intrinsically interdependent. The vibration model of the basilar membrane is different from the mechanical model for the organ of Corti and from the mechanical and electrical behavior of the hair cells connected to the OC. The creation of a model for the cochlea mechanics depends on the composition of different physical models, with different supporting clinical data, combining efforts from groups with different domains of expertise. The challenges for multi-scale biosimulation are described below.

Interoperability & Reproducibility: Multi-scale biosimulation modelling depends on the composition of models crossing different domains of expertise. Reuse of the biosimulation results by other researchers typically involves the reproduction of the experiment using information gleaned from the associated research paper. Where data is made available from the original experiment, differences in data representation define a barrier to reproduction, review,

comparison and reuse.

Data Integration & Accessibility: A concrete example of multi-scale modelling in the inner-ear domain is the integration, between simulations for the basilar membrane model, with the mechanical model for the organ of corti developed by other researchers with the model for the hair cells. Typically, the experimental and model data are siloed in different environments, for different research groups, in heterogeneous (possibly poorly structured) formats. In such a scenario, simple tasks such as finding specific data from previous simulations can only be achieved, in general, by reading, analysing and replicating associated publications.

Validation & Data Analysis: As models get more complex, as is the case with multi-scale simulation models, the complexity of the simulation process drastically increases, as the number of possible combinations of simulation parameters increases. Different parts of the simulation model need to be validated and supported by experimental and clinical data, which need to be integrated into the simulation workflow.

The challenges described above can be addressed by the provision of a principled, standards-based representation model which is able to represent biosimulation data unambiguously in both a human and a machine interpretable way. The construction of a representation model for biosimulation experiments are particularly challenging as the representation model needs to cover different model domains.

Figure 3 shows the relationship between the core elements involved in a multi-scale cochlear mechanics model and their relationship with a *semantic conceptual model* (SCM). From the simulation perspective, the conceptual model supports the communication between different components of the multi-scale model. The left side of the picture shows the communication between the simulation output of the basilar membrane model as an input to the organ of Corti model, which communicates its output to the hair cells model. From this point, the communication from the HC simulation is propagated back to the OC and later to the BM, as an iterative feedback mechanism.

In a loosely coupled simulation model the output of one component is sent as input parameters and boundary conditions to the next level. The semantic model in this case integrates under the same ontology variables and parameters involved in the input and output of the simulation, facilitating the composition of different models.

Experimental and clinical data (pure tone audiogram, otoacoustic emissions, echocg, among others), already integrated into the same semantic conceptual model, are used as input parameters for the simulation model or in the validation of the simulation output. Geometrical data extracted from histological sections and microCT images are also integrated into the same model.

By using a machine processable and structured representation, algorithms can be used to support the automation of the data analysis and model validation, supporting

the experimentation with different parameters, mathematical models in the simulation process (bottom part of Figure 3). Additionally, domain experts can enrich the existing domain knowledge by providing explicit links to information in the literature, or by persisting their analysis into the same integrated model.

Section IV discusses the representation dimensions in details.

III. RELATED WORK

This section focuses on the description of related work in two domains: ontology-based approaches to model finite element simulations and multi-scale biosimulation models integration.

A. Automated finite element analysis with an ontology-based approach

In [2] an ontology-based framework to increase finite element modelling (FEM) reliability is proposed for automated finite element analysis (FEA). The proposed ontology for automated FEA is supported by a hierarchical approach to make transfer formats for data, for information and for knowledge so that human intervention is not required to rebuild data/information/knowledge lost in the transfer. It uses a three-stage automated FEA method used in problem definition (for artifact markup), in problem formulation (for domain knowledge re-use) and in the solution routine (for the automation of FEM analysis). In order to enable data exchange between designers and analysts, two OWL encoded ontologies were developed namely a design domain ontology and a FEA ontology which represents a generic FEA activity.

B. Multi-scale integration

Multi-scale model integration refers to the merging of biosimulation models across several size scales, from cellular to tissue level, or across multiple time scales.

Gennari et al [3] took three different biosimulation models of the heart, at three different scales i.e. a cardiovascular fluid dynamics model, a model of heart regulation and a sub-cellular model of the arteriolar smooth muscle, and built a lightweight ontological framework, called Application Model Ontology (AMO), using small subsets of the reference ontologies, to annotate these biosimulation models semantically and to map between matching concepts. To merge computational models of biological processes into reusable, multi-scale models, two reference ontologies were used, namely the Foundational Model of Anatomy (FMA) [4] and the Ontology of Physics for Biology (OPB) [5] (which are respectively a reference ontology of human anatomy and an ontology for the physics of biological systems). The researchers were then able to address the question of how heart rate and blood pressure depend on calcium uptake into arteriolar smooth muscle cells, a question that could not be answered using one biosimulation model alone.
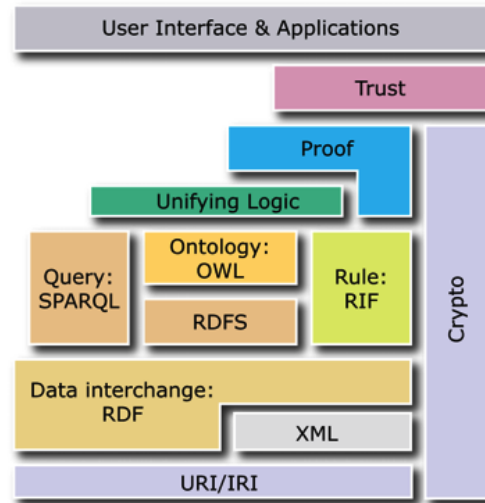

Fig. 1. Semantic Web representation stack.

IV. SEMANTIC REPRESENTATION

The provision of a machine-interpretable semantic representation for multi-scale FE biosimulation experiments depends on two main dimensions: (i) the construction of a comprehensive *conceptual model* which describes the entitites, attributes and relationships of the FE biosimulation experiments domain including all associated dimensions (physical, mathematical, anatomical, among others) and (ii) the use of a *standardized data model* which supports a semantic representation of the conceptual model and associated instance simulation data. The next subsections describe these two dimensions:

A. Standardized Semantic Data Model

Data interoperability can be achieved by defining the semantics of the content consistently across the heterogeneous data sources. The aim of SIFEM project is to semantify the clinical inner-ear and biosimulation data and models using Semantic Web technologies and standards, making the information machine understandable and processable.

Figure 1 shows the Semantic Web stack[2], which is a hierarchy of languages and technologies, where each layer exploits and uses capabilities of the layers below. The bottom layer, Uniform Resource Identifier (URI) represents a naming protocol for uniquely identifying Web resources. The Extensible Markup Language (XML) layer with the XML namespace and XML Schema definitions ensures a common syntax is used in the Semantic Web.

Generally, biomedical data resources on the Web are distributed. Consequently, the description of resources should be encoded in a way that facilitates integration from a large number of sources. On the top of the URI referencing mechanism and XML document exchange mechanism, the Resource Description Framework (RDF) layer provides a graph-structured data format to encode descriptions about

[2]http://www.w3.org/2007/03/layerCake.png

Web resources. RDF is the data model for Semantic Web data. Having interlinked RDF data sets, mechanisms for querying the RDF data are necessary. SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language) is a declarative query language, which allows for specifying queries against data in RDF. SPARQL is the standard query language for semantic data. The RDF graph-based encoding of Web resources covers only parts of the meaning of the data. RDF(S) constructs allows modelling class or property hierarchies. Data may require logic-based representation constructs. The Web Ontology Language (OWL) is a logic-based language which suport inferences, i.e., drawing conclusions based on existing facts. The Rules Interchange Format (RIF) allows for encoding and exchanging logical rules. Logical consequences can be inferred from a set of axioms with the help of logical reasoners.

The Semantic Web technologies can improve the knowledge representation and integration of the clinical data by providing standards to represent the data at Web scale. Moreover, such semantic representation supports intelligent knowledge processing such as reasoning, semantic approximation and alignments.

B. Multi-scale Conceptual Model

The domain knowledge can be represented as ontologies using the layered structure representation Figure 1). The representation of the cochlear mechanics biosimulation experiments is dependent on different domains. This section describes each domain dimension and existing ontologies for its representation.

1) Physics Model: The physics model layer allows a principled description of physical quantities, their relationships and associated units involved in a FE biosimulation experiment. The *Ontology of Physics for Biology (OPB)* [5] is an ontology focused on the description of physics for biosimulation from a foundational perspective. The Semantic Web for Earth and Environmental Terminology (SWEET) [6] [7] is a group of ontologies which describe physical quantities and units, among other concept domains. Examples of SWEET classes are: *Pressure* and *Density* ; OPB class examples include: *Physical entity*, *Physical property* and *Physical state*.

2) Algorithmic & Finite Element Model: This layer provides the description of the algorithms which are used within a given simulation setup, including the algorithms input parameters, a categorization for the algorithms and the interrelationships between different algorithms. The *Kinetic Simulation Algorithm Ontology (KiSAO)* [8] categorizes and describes the most common biosimulation algorithms. This layer also descries the main concepts in the FE domain (e.g. Mesh, boundary conditions). On the FE domain, Sun et al. [2] developed a Finite Element Analysis ontology which contains the basic concepts of finite element models. The finite element modelling markup language (femML) [9] is an attempt to address the problem of data representation and exchange in the finite element modelling (FEM) domain.
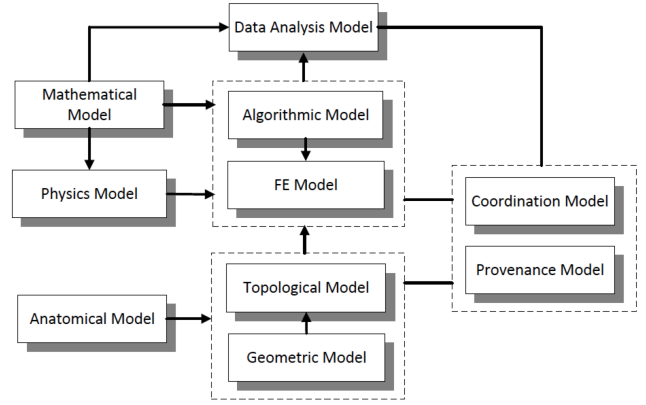

Fig. 2.   Representation dimensions for FE biosimulation models.
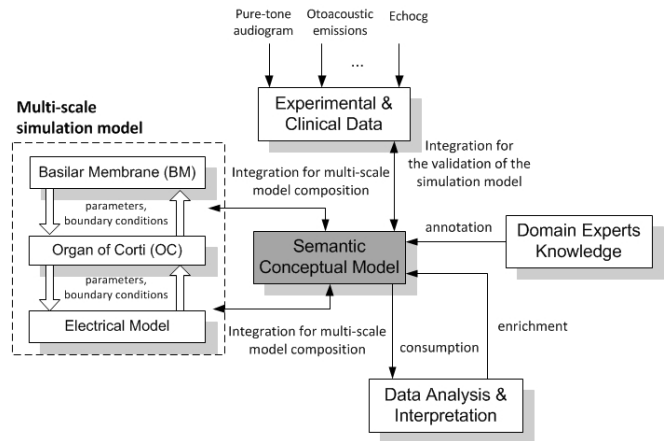

Fig. 3.   SIFEM Multi-scale cochlear model and semantic integration between the different scale components.

3) Material Properties Model: The Materials Ontology [10] serves as a base to integrate heterogeneous materials information resources (such as databases, computational simulations, documents and equation solvers) mapping concepts between these resources in order to enable more flexible search and interoperable representation for materials information. The design of this ontology is discussed in [10] followed by an example of its application to exchange data from three well known thermal property databases.

4) Data Analysis Model: This dimension of the conceptual model provides a symbolic description of the simulation output, where the symbolic features extracted from the numerical results are classified under a data analysis ontology. Value ranges, maxima, minima, derivative, function zeroes are examples of values which can be used to describe the numerical output of an experiment. This allows the description of an observed behavior of a simulation experiment in a machine-readable manner so that researchers can answer questions such as "Which model displays a behavior matching my experimental data?" The *terminology for the description of dynamics (Teddy)* is a data analysis ontology which provides a set of function analysis features [8].

5) Geometrical & Topological Models: FE biosimulation experiments use geometrical data associated with the anatomical structures The connection between the symbolic-level and the geometrical-level structures is still an area of research where there is a lack of standardized solutions. 3D models are digital representations of objects that can be processed by computer applications [11] which can be used to represent objects or phenomena such as electromagnetic fields. Current 3D modelling systems handle the geometric representation of digital shapes (using shape analysis and segmentation) but do not provide a symbolic (semantic) description. In [12] the development of a Semantic Web environment, Java-based, prototype for 3D annotation called be_SMART (Beyond Shape Modelling for Understanding Real World Representations) is described. More specifically, this system performs the geometric-topological inspection and semantic structuring and annotation of 3D shapes using metadata from the Common Shape Ontology.

6) Coordination & Provenance Model: Provenance is the information about artefacts, processes, and agents involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. PROV-O[3] is a specication, promoted by the World Wide Web consortium, for recording the provenance of web resources. The PROV-O ontology defines a model to enable the interoperable interchange of provenance information in heterogeneous environments such as the Web. Prospective provenance descriptors are used on the scope of this project to describe biosimulation experiments, while retrospective provenance descriptors describe the actual execution flow of the experiments.

7) Mathematical Model: Mathematical objects are a fundamental element in the description of biosimulation experiments. Explicitly representing mathematical knowledge can enable the comparison between the mathematical models behind different experiments, facilitating their reuse. There are two main initiatives for standardizing the representation of mathematical objects: MathML and OpenMath. *MathML* focuses on the syntactical representation of mathematical objects and its presentation. MathML allows the OpenMath terminology to be embedded inside a MathML structure.
OpenMath aims at encoding the semantics of mathematical objects by maintaining an extensible Content Dictionary mechanism.

8) Anatomical Model: The Foundational Model of Anatomy (FMA) is a reference ontology for anatomy, which satisfies fundamental requirements for ontological representation of human anatomy [4], representing anatomical objects and their relationships. An example of an FMA class is Basilar membrane of the cochlea.

V. Conclusion

In this article, we have identified the challenges involved in integrating and automating multi-scale biosimulation models. In order to address these challenges a principled multi-layered semantic model was discussed. This model is grounded on existing ontologies.

The ontologies described in this paper provide a way of representing and annotating the elements of the SIFEM project conceptual model in a machine-processable, unambiguous, standardized representation format which supports the reuse, access to, and the semantic interoperability of, heterogeneous data across models and systems. The principled semantic representation of the biosimulation and experimental data enables the reproducibility, integration and reusability of different biosimulation experiments and the automation of experimental analysis.

Future work include extending and refining the existing conceptual model and the development of the associated knowledge services, which will use the integrated data under the semantic conceptual model to support the multi-scale FE simulations.

VI. Acknowledgements

References

[1] S. N. Merchant, M. J. McKenna, J. C. Adams, J. B. Nadol Jr, J. Fayad, R. Gellibolian, F. H. Linthicum Jr, A. Ishiyama, I. Lopez, G. Ishiyama *et al.*, "Human temporal bone consortium for research resource enhancement," *JARO-Journal of the Association for Research in Otolaryngology*, vol. 9, no. 1, pp. 1–4, 2008.

[2] W. Sun, Q. Ma, and S. Chen, "A framework for automated finite element analysis with an ontology-based approach," *Journal of mechanical science and technology*, vol. 23, no. 12, pp. 3209–3220, 2009.

[3] J. H. Gennari, M. L. Neal, B. E. Carlson, and D. L. Cook, "Integration of multi-scale biosimulation models via light-weight semantics," in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. NIH Public Access, 2008, p. 414.

[4] C. Rosse and J. L. Mejino Jr, "The foundational model of anatomy ontology," in *Anatomy Ontologies for Bioinformatics*. Springer, 2008, pp. 59–117.

[5] D. L. Cook, J. L. Mejino Jr, M. L. Neal, and J. H. Gennari, "Bridging biological ontologies and biosimulation: the ontology of physics for biology," in *AMIA Annual Symposium Proceedings*, vol. 2008. American Medical Informatics Association, 2008, p. 136.

[6] R. Raskin, "Guide to sweet ontologies," *NASA/Jet Propulsion Lab, Pasadena, CA, USA, Available at: http://sweet. jpl. nasa. gov/guide. doc (last accessed: May 2011)*, 2006.

[7] A. Tripathi and H. A. Babaie, "Developing a modular hydrogeology ontology by extending the sweet upper-level ontologies," *Computers & Geosciences*, vol. 34, no. 9, pp. 1022–1033, 2008.

[8] M. Courtot, N. Juty, C. Knüpfer, D. Waltemath, A. Zhukova, A. Dräger, M. Dumontier, A. Finney, M. Golebiewski, J. Hastings *et al.*, "Controlled vocabularies and semantics in systems biology," *Molecular systems biology*, vol. 7, no. 1, 2011.

[9] J. Michopoulos, "Development of the finite element modeling markup language," in *ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 572. Citeseer, 2002.

[10] T. Ashino, "Materials ontology: An infrastructure for exchanging materials information and knowledge," *Data Science Journal*, no. 0, p. 1006250233, 2010.

[11] M. Spagnuolo and B. Falcidieno, "3d media and the semantic web," *Intelligent Systems, IEEE*, vol. 24, no. 2, pp. 90–96, 2009.

[12] L. De Floriani, A. Hui, L. Papaleo, M. Huang, and J. Hendler, "A semantic web environment for digital shapes understanding," in *Semantic Multimedia*. Springer, 2007, pp. 226–239.

---

[3]http://www.w3.org/TR/prov-o/