Integrating services for integrated medicine
Building virtual physiomes with semantics services

David Lambert

Knowledge Media Institute, Open University, Milton Keynes, United Kingdom
d.j.lambert@open.ac.uk

Abstract. The anticipated advent of personalised healthcare is predicated on sophisticated computer modelling of individuals' physiomes, but the integration of the many diverse computational services, operating across scales and domains, remains a major hurdle. We present the Living Human Digital Library, a project in the biomechanics domain where semantic web services will help manage and integrate the services which provide both technical and community support functionality. The semantic services approach is intended to satisfy both short-term project requirements, and develop the technical underpinnings for multi-domain integration with other physiome projects.

1 Introduction

The practice of medicine will soon be radically altered by the cheap sequencing of individual patients' genomes and computer modelling of the interaction of that genome the environment and medical interventions [1]. While the genome hogs the limelight, for the clinician and patient it is the outcome for the whole organism—the *physiome*—that is the ultimate concern. Physiology is the study of the physical and mechanical as well as the bio-chemical behaviour of organisms. Consider that mice subjected to mechanical vibrations alter their bone density positively in response [2], a result with implications for astronauts and the Earth-bound suffering from bone loss. Although the cause and effect occur at a macro scale, they are mediated by processes involving proteins and genes. To fully understand these processes, multiple data sources and algorithms representing diverse scales and subsystems must be integrated, a task which has been identified as an informatics 'Grand Challenge'.

These 'virtual physiological humans' (VPH) will be built for specific purposes from components which will be developed and maintained by separate communities focused on particular physiological subsystems. This paper outlines the programme of the Living Human Digital Library (LHDL) project [3], which is collecting data and developing infrastructure to model the human musculoskeletal system (section 2). As well as providing ongoing support to the biomechanics research community, LHDL will investigate the techniques for constructing the virtual physiome: we believe semantic web services (section 3) are best placed to handle the interoperation of services within and between communities. LHDL presents many challenges for semantic services research (section 4), and differs significantly from other projects in bioinformatics (section 5).

2 The Living Human Digital Library

LHDL will establish a distributed fabric of community tools, data repositories and algorithms for medical researchers and clinical workers in biomechanics: one of the partners is the Istituti Ortopedici Rizzoli, a centre of excellence in orthopaedics research and treatment. The immediate objective of LHDL is to ease the sharing of data and computational facilities within the biomechanics community by using web and Grid technologies, supplemented by semantics. The longer-term ambition is to extend this approach to the integration of other communities and their systems. Such integration must work across scales and between disciplines (e.g. chemistry, biomechanics, clinical) and sub-systems (e.g. renal, cardiac, orthopaedic). VPH is about coordination: the intention is not to create a single federation of services that define a single VPH, but rather a framework to enable the integration of services to suit particular requirements—even to the point of modelling individual humans for clinical purposes.

The principle desktop tool is LhpBuilder [4], an application which enables a user to create, store, and manipulate Virtual Medical Entities (VMEs). VMEs are collections of data such as MRI images, gait analysis data or finite element analysis results. LhpBuilder can perform operations such as extracting two-dimensional slices from volume data, virtual palpations, or combining motion-capture data with bone images. LHDL is a community of experts as much as a group of technologies, so we provide community services such as forums, mailing lists and file storage: these are accessed at Biomed Town (`www.biomedtown.org`) via a web browser.

The LHDL project has identified semantic web services as the key to achieving its short and long term aims. The storage facilities of Biomed Town will be made available as web services, and LhpBuilder will be made network aware to take advantage of the storage grid. LhpBuilder's own computational functions will similarly be exported as services on a computational grid. The community services are to be semantically annotated, too: processes such as user registration and authentication will be semantically enabled, making easier single sign-on between services, for example, or compliance with security and privacy policies.

3 Semantic web services

Semantic web services [5] extend the benefits of semantic web technologies to the provision of services. By marking-up services semantically, software can reason about them, and automate many of the complex, time consuming and error-prone tasks that otherwise must be done by the scientist or clinician: workflow planning, service selection [6], mediation between services and data formats, process monitoring, and ensuring permissions for access to data are met [7].

The Web Services Modelling Ontology (WSMO) [8] is a leading semantic web services framework, similar to OWL-S [9]. WSMO has four key concepts: domain ontologies, goals, web services, and mediators. Domain ontologies capture the users' conceptualisation of the objects and relations of interest. Goals represent user intentions, while services are available software systems which can be invoked to satisfy goals. Goals have four principle components: *preconditions* and *postconditions*, predicates which must hold on the inputs and outputs respectively; and *assumptions* and *effects*, which are conditions on the world which should hold true before and after respectively. WSMO insists on a clear distinction between user goals and their realisation by web services, thus enabling capability-based invocation. Mediators intervene in several places where otherwise heterogeneity would cause incompatibility: between ontologies, *OO-mediators* perform ontology mapping wherever necessary; *WW-mediators* allow web services to interact correctly, primarily addressing choreography mismatches; user goals are mapped to web services by *WG-mediators*; and *GG-mediators* allow the creation of new goals by composing others.

The Internet Reasoning Service III (IRS) [10], in its current implementation, adopts and extends the epistemological commitments of WSMO. The IRS has been used in several domains including business process management, e-learning, and e-government. It can call web services exposed via SOAP, XML-RPC, or HTTP GET, publish legacy code written Java and Common Lisp as web services, and can itself be invoked via SOAP or HTTP GET requests (thus supporting the REST philosophy). A process of 'elevation' deals with mapping the XML messages of services to internal ontological representations expressed in OCML [11].

4 Challenges

The development of LHDL is in its early stages, but several issues are already apparent:

Quantity of data Biomechanics not only deals with large volumes of data, it deals with large quanta, too. A single VME may be hundreds or thousands of megabytes in size. This is sometimes solved by third-party transfers, where the controller initiates a data transfer from a source to its final destination but does not itself participate, but then semantic mediation cannot happen. Conversely, funnelling such information flows through centralised brokers, which will often be required to massage the data formats, may not scale. We will need to examine mechanisms for truly distributed semantic middleware (in the sense of service registries, ontological shim, message buffers, process monitoring and others).

Interaction modalities The nature of the applications is often different to the traditional workflow orientation: interactive manipulation of models, for instance. This will require sophisticated choreography and state management.

Laws and ethics Genetics research typically operates on publicly accessible databases which do not identify anyone. In the LHDL domain, biomechanics, much of the data is recovered from cadavers whose identity must be protected, and many of the applications will be clinical and thus concern individual patients.

Metadata maintenance VMEs might be very long lived, and contain important patient data. They are also associated with metadata that must be kept in synchrony with current ontology and software versions, either through modification or by some process of mediation.

Integration across scale and domain The heterogeneity here is immense, and the mediation involved will be important. The multiple domains bring their own ontologies; different scales of space and time complicate orchestration and choreography; and the range of services required will make it difficult for any single human to be sufficiently expert to select the most appropriate services. Finally, the research communities themselves must be knit together to make the user experience acceptable.

5 Related work

The notion of integrated modelling of the human physiome is first proposed in [12]. The development of a database-centric model for the physiome idea is developed in [13], but it underestimates the importance of computational services and has been obsoleted by the web services [14] approach. In the service-oriented era, there are three large, informatics-centric life sciences programmes which demand mention: ^{my}Grid and BioMOBY and GEMSS. Grid Enabled Medical Simulation Services (GEMMS) [15] explored the feasibility of providing computationally intensive medical applications using the Grid, primarily for practising clinicians. These services included bone structure capture for facial reconstruction surgery, optimising radiation propagation paths for radiosurgery, and computation of drug inhalation patterns. Semantic issues were not addressed.

^{my}Grid [16] is a large programme providing geneticists with workflow tools for web services. While ^{my}Grid has focused on retrofitting semantics to existing web services and tools, BioMOBY [17] set out to create a unified ontology, with services strictly adhering to the standard terminology and XML message structures. Despite the ontology itself being developed collaboratively, in an 'open source' way, this approach precludes incorporation of legacy services and third-party annotation. In both projects, familiarity of the practitioners was considered more important than the expressiveness of the ontology languages. ^{my}Grid and BioMOBY target a community which had long used distributed data stores and computation: in essence, they automated a manual task, incorporating ad-hoc service semantics which suit their users' needs. LHDL has a commitment to applying a comprehensive semantic web services framework to a field which has not historically much benefitted from web or grid services, with the intention that the technologies developed there will scale to support the cross-domain, multi-scale simulations of the future.

6 Conclusions

LHDL is an application of semantic web services driven by researchers in biomechanics who have believe semantic technologies will provide them with tangible benefits. LHDL is concerned with biomechanics, leading to a different set of concerns from the usual bioinformatics fare of DNA or protein analysis. Interactivity and visualisation become more important, as do ethical issues relating to the handling of medical records of identifiable individuals. We intend to apply proven general-purpose semantic web technologies in this domain, and extend them to deal with the peculiar difficulties of the field.

References

1. Weston, A.D., Hood, L.: Systems Biology, Proteomics, and the Future of Health Care: Toward Predictive, Preventative, and Personalized Medicine. J Proteome Res **3**(2) (2004) 179–196
2. Judex, S., Monaghan, M., Rubin, C., Dhundale, R.: Altered mechanical demand of the skeleton broadly changes transcriptional activity. Journal of Biomechanics (39(Sl)) (2006) S22
3. Viceconti, M., Taddei, F., Van Sint Jan, S., Leardini, A., Clapworthy, G., Galizia, S., Quadrani, P.: Towards the multiscale modelling of musculoskeletal system. In: Bioengineering modeling and computer simulation. (2007)
4. Viceconti, M., Taddei, F., Montanari, L., Testi, D., Leardini, A., Clapworthy, G., Jan, S.V.S.: Multimod data manager: A tool for data fusion. Comput. Methods Prog. Biomed. **87**(2) (2007) 148–159
5. McIlraith, S., Son, T., Zeng, H.: Semantic web services. IEEE Intelligent Systems **16** (2001) 46–53
6. Paulucci, M., Kawamura, T., Payne, T.R., Sycara, K.: Semantic Matching of Web Services Capabilities. In: The Semantic Web — ISWC 2002: Proceedings. (2002)
7. Galizia, S.: WSTO: A Classification-Based Ontology for Managing Trust in Semantic Web Services. In Sure, Y., Domingue, J., eds.: Proceedings of the 3rd European Semantic web Conference. (2006)
8. Fensel, D., Lausen, H., Polleres, A., de Bruijn, J., Stollberg, M., Roman, D., Domingue, J.: Enabling Semantic Web Services. Springer (2006)
9. Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N., Sycara, K.: OWL-S: Semantic markup for web services (2004)
10. Cabral, L., Domingue, J., Galizia, S., Gugliotta, A., Norton, B., Tanasescu, V., Pedrinaci, C.: IRS-III: A Broker for Semantic Web Services based Applications. In: Proceedings of the 5th International Semantic Web Conference (ISWC2006), Athens, Georgia, USA (2006)
11. Motta, E.: An Overview of the OCML Modelling Language. In: 8 th Workshop on Knowledge Engineering: Methods & Languages KEML 98. (1998)
12. Bassingthwaighte, J.: Design and strategy for the cardionome project. Adv Exp Med Biol (430) (1997) 325–39
13. Dao, N., McCormick, P.J., Dewey, Jr, C.F.: The Human Physiome as an Information Environment. Annals of Biomedical Engineering **28**(8) (August 2000) 1032–1042
14. Gottschalk, K., Graham, S., Kreger, H., Snell, J.: Introduction to web services architecture. IBM Systems Journal **41**(2) (2002)
15. Benkner, S., Engelbrecht, G., Backfrieder, W., Berti, G., Fingberg, J., Kohring, G., Schmidt, J.G., Middleton, S.E., Jones, D., Fenner, J. In: Numerical Simulation for eHealth: Grid-enabled Medical Simulation Services. Elsevier (2004)
16. Stevens, R.D., Robinson, A.J., Goble, C.A.: myGrid: personalised bioinformatics on the information grid. Bioinformatics **19 Suppl. 1** (2003) i302–i304
17. Wilkinson, M.D., Links, M.: BioMOBY: An open source biological web services proposal. Briefings in bioinformatics **3**(4) (2002) 331–341