34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

An Architecture for Integrating Cancer Model Repositories*

Stelios Sfakianakis, Vangelis Sakkalis, Kostas Marias, Georgios Stamatakos, Steve McKeever, Thomas S. Deisboeck, and Norbert Graf

Abstract— **The TUMOR project aims at developing a European clinically oriented semantic-layered cancer digital model repository from existing EU projects that will be interoperable with the US grid-enabled semantic-layered digital model repository platform at CViT.org (Center for the Development of a Virtual Tumor, Massachusetts General Hospital (MGH), Boston, USA) which is NIH/NCI-caGRID compatible. In this paper we describe the modular and federated architecture of TUMOR that effectively addresses model integration, interoperability, and security related issues.**

I. Introduction

Cancer is a remarkable disease that involves intractable neoplastic growth, invasion of surrounding tissue and metastasis mechanisms. In order to perform in silico experiments there is a need to study the various phases and scales describing different levels of biocomplexity using mathematical modeling and simulation. Such computational multiscale models often bypass the initial tumor genesis stage and focus mainly on the growth phase. In order to better relate various phenomena occurring at different scales, it is necessary on the one hand to account for microscopic processes when trying to predict macroscopic tumor growth, but on the other hand one needs to be able to correlate microscopic variables with a number of clinically meaningful parameters related to macroscopic phenomena often measured in clinical practice.

Microscopic models, more often based on discrete event mapping, are suitable to the description of individual cell dynamics according to some stochastic rules and are mainly applicable at the sub cellular and cellular levels, whereas the macroscopic models are more often based on continuum approaches assuming that the solid tumor behavior can be predicted in terms of its global interaction with the surrounding and underlying tissue properties and a few internal parameters related to the proliferation rate0[2][3]. We may also refer to another intermediate spatial scale known as mesoscopic models that consider the behavior of a group of cells and their interactions as clusters sharing similar physical properties [4]. Hence, models of cancer progression span a range of granularities, from modeling molecular pathways of individual cells, to the geometric-cell level where the behavior of groups of tumor cells are simulated as individual entities.

However in practice, each of the aforementioned models are often developed independently from different specialized research groups around the globe that focus their research on specific temporal and spatial scales. Obviously such a fragmented approach limits the potential of examining cancer under a global prism that consolidates crucial information from different levels of complexity.

So far, significant but highly fragmented efforts have been made on both sides of the Atlantic to develop and use models of pathophysiology in order to better understand human function and promote individualized, patient-specific optimization of disease treatment. The TUMOR project, an EU FP7 funded project, is developing a European clinically oriented semantic-layered cancer digital model repository from existing EU Virtual Physiological Human (VPH) related projects designed to be interoperable with the US grid enabled semantic-layered digital model repository platform at CViT which is NIH/NCI-caGRID compatible. Models and data will drive advances in cancer modeling with the ultimate goal to build an integrated, interoperable transatlantic research environment offering the best available models and tools for clinically oriented cancer modeling and serving as an international validation/ clinical translation platform for predictive, in-silico oncology.

To achieve this ambitious goal, an interoperable, transatlantic environment is needed to offer a range of services to international cancer modelers, bio-researchers and eventually clinicians in fostering both basic cancer research and individualized optimization of cancer treatment.

This paper mainly focuses on the envisaged architecture (Section III) that deals with integration, interoperability, and security related issues. A clinical perspective that highlights the need for implementing workflows capable of interfacing models operating on different scales (molecular-microscopic-macroscopic interactions) is described in section II.

II. The TUMOR clinical perspective

To enable interoperable EU US executable models a generic multilevel simulation execution environment is being developed. The US CViT and EU model repositories host models on macro, micro, and molecular levels that may act in complementary and supportive ways. For example, a "European" mesoscopic model may model the lifecycle of

*This work was supported in part by the European Commission under the TUMOR (FP7-ICT-2009.5.4-247754) project.

S. Sfakianakis, V. Sakkalis and K. Marias are with the Institute of Computer Science at FORTH, Vassilika Vouton, GR-70013 Heraklion, Crete, Greece (e-mail: {sakkalis; ssfak; kmarias}@ics.forth.gr)

G. Stamatakos is with the Institute of Communication and Computer Systems, Department of Electrical and Computer Engineering, National Technical University of Athens, GR-15780, Athens, Greece (e-mail: gestam@central.ntua.gr)

S. McKeever is with the Department of Computer Science, Wolfson Building, Parks Rad, Oxford OX1 3QD (e-mail: Steve.McKeever@cs.ox.ac.uk)

T. Deisboeck is with the Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA (e-mail: deisboec@helix.mgh.harvard.edu)

N. Graf is with the Saarland University Hospital, Paediatric Haematology and Oncology, D-66421, Homburg, Germany (e-mail: Norbert.Graf@uniklinikum-saarland.de)

groups of cells ('geometric cells') based on various environmental factors (e.g. distance from nutrients, states of immediately adjacent cells, availability of oxygen), probability, and events [6]. This so called "top-down" approach is based on how cells change their state based on factors outside of the cells themselves. For the clinical application, drugs and radiotherapy are regarded as environmental factors. On the other hand a US CViT model of reference may take an agent-based approach where the individual cells have behaviors built into them. More specifically, molecular pathways are basically models of the chemical processes that cause a cell to shift state or result in an action such as cell division, cell death (necrosis/ apoptosis) and even proliferation or migration. So basically in this case the "bottom-up" approach is being able to model within a single cell the molecular and chemical processes, and their resulting actions.

During the past few decades most of the approaches in in-silico oncology have been targeted to the provision of insight into the tumor growth mechanisms (cancer biology modeling) rather than the production of concrete clinically exploitable systems that would be able to support the clinician in the process of selecting the most appropriate treatment scheme and/or schedule for the individual patient [5]. Obviously the former is a prerequisite for the latter; yet the latter dictates an additional philosophy [6] markedly different from the one adopted by conventional cancer biology modeling, which is mainly a bottom-up biocomplexity level approach.

Further information on available biosimulation software is summarized by Ho et al. [7] and Deisboeck et al. [8], which give an excellent overview of 'in silico' cancer modeling by reviewing selected studies on modeling the progression and therapy of highly malignant brain tumors.

To support such complex fusion of models, as well as seamless integration with needed input data a specialized workflow editor needs to include a number of tools dealing with the pre-processing (transformation) of data, anonymization/ pseudonymisation processes, the linking and execution of bottom-up/ top-down models, the visualization and the validation of the results. A clinician doesn't want to deal with the building of such a workflow, instead it is desirable to start a workflow and in an interactive and intuitive dialogue to proceed through the whole workflow until the model is executed (Figure 1). If the result of the model is to be used in the clinical setting, it needs to be automatically validated (validation tool within the workflow) and delivered in due time allowing it's use in the decision process (through a Decision Support System – DSS) for treating the patient.

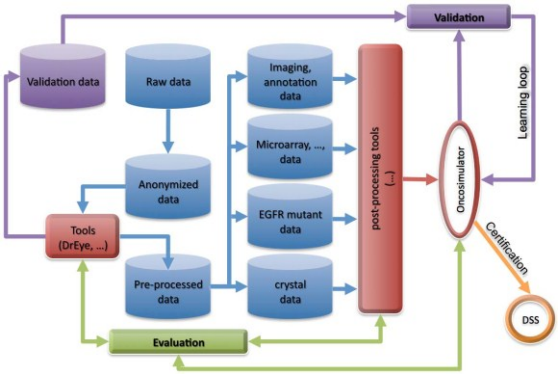

Figure 1 The workflow in more detail

III. ARCHITECTURE

A. Functional Requirements

The integrative TUMOR clinical workflow perspective described in the previous section can be translated to a set of functional requirements for the design of its architecture. These are as follows:

- The users should be able to upload their cancer models and to use appropriate *metadata* in order to efficiently locate them afterwards and maintain their versioning history. Such metadata will consist of publishing information about the author, creation date, etc. They can also include access control information, which will permit or disallow their discovery by other users, licensing information to protect for intellectual property rights, and so on. It is actually the users who decide which model to share and what restrictions should be put on its use.
- The cancer models shared are accompanied by the necessary information that permits their actual execution. This information could include the code (in an executable or source format) and the necessary data used at the model's runtime.
- Data can also be used in more than one cancer model and therefore there's a need for supporting *data repository* in addition to the *model repository* described above. Proper metadata are provided with the uploaded data and provide hints about their purpose, type, lineage, etc.
- Different models with diverse biocomplexity levels and directions (bottom-up, top-down) are to be linked together in order to simulate cancer growth in a more *holistic* way. In particular, computational models are executed by exchanging data that correspond to their output and input parameters, so that a higher level of modeling is achieved. It is important to have an intuitive user interface for the domain experts to build these *hypermodels* dynamically and the paradigm of *scientific workflows* is the one to follow [9].
- After the completion of a new workflow that connects two or more cancer models together, the users should be assisted to run the associated workflows by providing input parameters and data coming from the data repository. The execution should be transparent and leverage the metadata accompanied by each constituent cancer model in order to identify the required parameters, data, and execution environment.
- The "transatlantic" scenarios are implemented through the workflow paradigm described above by retrieving the models both from the EU and the US model repositories.

- Intellectual properties of the users are protected while social networking facilities that are extremely popular these days are also accommodated.
- Privacy and security are built in. User authentication with "single sign on" ensures that the identity of the users is always available and proper access control mechanisms can be applied in every component of the platform.

An additional requirement is that according to the current legal and ethical regulations and restrictions, in both Europe and the US, even the exchange of retrospective data seems not feasible. Data therefore needs to be stored locally in Europe or the US and not exchanged between partners. To overcome this problem tools and models need to be exchanged and shared to run simulations with locally provided data. This solution has significant implications for the type of infrastructure that TUMOR is developing.

B. *Integration Scheme*

TUMOR presents many integration, interoperability, and security related challenges. In designing the architecture we have followed the approach of views and viewpoints, which is standardized by IEEE/ISO [10]. Starting with the requirements and the functionality that the TUMOR platform aims to deliver, we identify the use cases that Figure 2 exhibits:

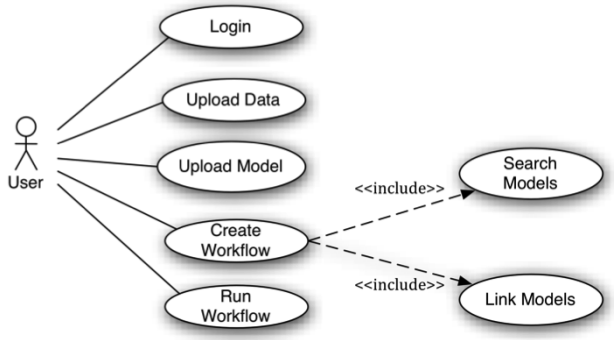

Figure 2 The main TUMOR use cases

The *login* use case represents the user authentication process that is a prerequisite to any use of the platform. Based on the supplied user credentials the user profile information can be retrieved and proper authorization decisions can later be based on. Uploading data and models captures use cases where the users transfer, possibly publish and share their digital assets and artifacts. The model integration and linking is supported by workflow technologies but this requires two additional functionalities: 1) searching the model repositories based on some criteria or just browsing their contents, and 2) linking the selected models by connecting their output and input parameters. The constructed workflow can then be executed and after its successful termination the results can be retrieved.

Based on these scenarios and the requirements of the project, we have identified the following software components and their responsibilities:

- The European Model and Data Repository: This is the "main" model repository, located in Europe. In addition to storing the cancer models of the European users and their anonymized data, this repository also maintains the users profile information.
- The US Model Repository: This is the American model repository, located in the US and operated by CViT. This is where US-CViT users store their models and data. It can be accessed from the European side but only the models can be transferred, due to the legal and ethical requirements.
- The Workflow Editing and Enactment environment, which is the web based application that allows the construction of simulation experiments through the linking of the available cancer models [11]. In order to do so, the Workflow Environment accesses the EU and US model repositories and selectively retrieves models from their contents. It is hosted inside the EU and therefore it has access to the data stored in the EU repository. Nevertheless since it is a web application, it has to make authorization decisions based on the users profile in order to restrict the data access mechanisms only to the European users. The execution of the workflows is taken care of by a cluster of processing machines physically collocated with the workflow environment's server side.
- The Common Access Point (CAP, for short): This is the main "entrance" to the platform. It is a web portal for interacting with the majority of the TUMOR services. Behind this portal there will be the EU Model and Data repositories and also the users profile database.

The deployment architecture of TUMOR is shown in Figure 3.

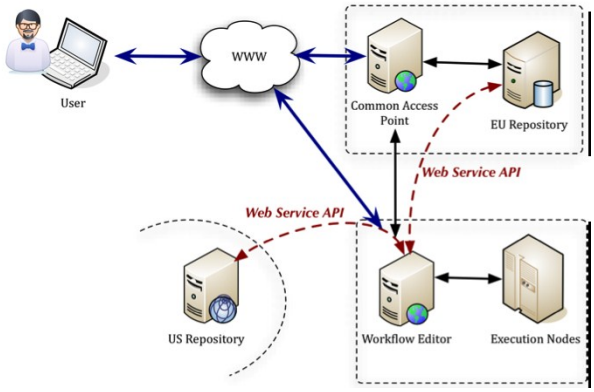

Figure 3 The deployment of the TUMOR platform

C. *Information View*

The models themselves are described using TumorML [12], a new markup language (ML) for describing cancer models. The development of TumorML contributes to enabling some of the key aims within the TUMOR project. Firstly, by annotating cancer models with appropriate document metadata, digital curation is facilitated in order to make publishing, search, and retrieval of cancer models easier for researchers and clinicians using the TUMOR

digital repository. Second, markup will be used to describe abstract interfaces to published implementations allowing execution frameworks to run simulations using published models. Finally, TumorML markup facilitates the composition of compound models, regardless of scale and source, enabling multiscale models to be developed in a modular fashion, and models from the US CViT to be integrated with EU models in the TUMOR transatlantic scenarios.

D. *Engineering View*

The TUMOR environment is built as an online platform where its services are accessible over the World Wide Web. The architecture therefore is designed with "service orientation" in mind [13].

In essence there is a programmatic interface for the "cross database" search and transmission of the models, so that no patient data is transmitted outside the European Union due to the lack of a legal framework and the implicated ethical and security issues. Therefore the main components that exhibit such an application programmatic interface (API) are the model repositories. Based on these APIs the Workflow Environment can browse, search, and retrieve cancer models and related data sets. There are two basic extensions to the baseline of SOAP/WSDL type of Web Services offered by the model repositories: Some data sets can be pretty large so the XML encoding imposed by the standard Web Services introduces a major performance tax. In these cases a more lightweight approach based on Representational State Transfer (REST) [14] is followed, i.e. the datasets are retrieved via simple HTTP(S) URLs. Secondly, in order to support the workflow based model integration facilities, the models related metadata need to be semantics based. Therefore Semantic Web technologies [15] are employed in various places. The TumorML descriptions of the models are RDF compliant and therefore can be searched and retrieved using SPARQL, while the linking with more specialized domain ontologies for the model descriptions, or with the data that are required for the model execution, are also greatly facilitated.

On the security front, there is the need for authenticating the users with the minimal possible distraction (Single Sign On) and also supporting authorization and access control. To address both of these concerns, TUMOR uses the OAuth 2.0 (Open Authorization, version 2.0 - http://oauth.net/2/) protocol that is also supported by Google, Microsoft, and Facebook in their web applications. Using OAuth the Workflow Environment can access the model repositories on the users' behalf without knowing their passwords or other authenticating information.

IV. CONCLUSION

The TUMOR project is a transatlantic effort with the goal of becoming the starting point for an international validation environment supporting joint modeling applications development as well as verification and validation of the clinical relevance of cancer models. To ensure the clinical relevance of this joint effort, the development of the project is based upon specific clinical scenarios that will be implemented within the proposed integrated EU-US workflow environment prototype. As an end result, a specific, predictive oncology workflow involving both EU and CViT models will be demonstrated, which will clearly highlight the need for and added value of interoperability.

The presented architecture takes into consideration all the technical issues and user requirements to realize this goal. At the same time a number of challenges still need to be addressed, mainly concerning the transatlantic sharing of the data and digital models from the legal and the security point of view.

ACKNOWLEDGMENT

The authors wish to thank all project partners as well as the EC reviewers for their constructive comments.

REFERENCES

[1] A. Roniotis, V. Sakkalis, K. Marias, I. Karatzanis and M. Zervakis, "In-depth analysis and evaluation of diffusive glioma models", IEEE Trans. Inf. Tech., Vol. 16, No. 3, pp. 299-307, 2012.

[2] V. Sakkalis et al., "Evaluation framework for the multilevel macroscopic models of solid tumor growth in the glioma case", Conf Proc IEEE Eng Med Biol Soc. 2010, pp. 6809-12, 2010.

[3] A. Roniotis, G. Manikis, V. Sakkalis, M. Zervakis, I. Karatzanis, and K. Marias "High grade glioma diffusive modeling using statistical tissue information and diffusion tensors extracted from atlases", IEEE Trans. Inf. Tech., Vol. 16, No. 2, pp. 255-263, 2011.

[4] P.P. Delsanto et al., "A multilevel approach to cancer growth modeling," J. Theor. Biol., Vol. 250, No 1, pp. 16-24, 2007.

[5] T. S. Deisboeck and G. S. Stamatakos Eds, Multiscale Cancer Modelling. Chapman & Hall/CRC, Boca Raton, Florida, USA.

[6] DD Dionysiou, GS Stamatakos, NK Uzunoglou, et al., "A four-dimensional simulation model of tumour response to radiotherapy in vivo: parametric validation considering radiosensitivity, genetic profile and fractionation," J Theor Biol Vol. 230, pp. 1–20, 2004.

[7] RL Ho and LT Bartsell, "Biosimulation software is changing research," Biotechnol Annu Rev, Vol. 10, pp. 297-302, 2004.

[8] TS Deisboeck et al. "In silico cancer modeling: is it ready for prime time?," Nat Clin Pract Oncol., Vol. 6, pp.34-42, 2009.

[9] A. Belloum et al., "Scientific workflows," Scientific Programming, Vol. 14, No 171, pp.3-4, 2006.

[10] ISO/IEC 42010:2007. Systems and software engineering - Recommended practice for architectural description of software-intensive systems. 2000-10-09 http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45991

[11] S. Sfakianakis et al., "Web-Based Authoring and Secure Enactment of Bioinformatics Workflows," Grid and Pervasive Computing Conference (GPC), 2009 pp.88-95, 4-8 May, 2009.

[12] D. Johnson, J. Cooper, and S. McKeever, "TumorML: Concept and Requirements of an In Silico Cancer Modelling Markup Language," Conf Proc IEEE Eng Med Biol Soc. 2011, pp.441-4, 2011.

[13] F. Curbera, M. Duftler, et al., "Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI," Internet Computing, IEEE, Vol. 6, No 2, pp. 86–93, 2002.

[14] RT Fielding and RN Taylor, "Principled Design of the Modern Web Architecture," ACM Trans on Internet Technology (TOIT), Vol. 2, No 2, pp. 115–150, 2002

[15] N. Shadbolt et al., The semantic web revisited, Intelligent Systems, IEEE, Vol. 21, No 3, pp. 96–101, 2006.