Semantic Integration for Model-based Life Science Applications

Tomasz Gubała - t.gubala@cyfronet.pl (1,4), Katarzyna Prymula (3), Piotr Nowakowski (1) and Marian Bubak (1,2,4)
(1) Distributed Computing Environments Team (DICE), Academic Computer Centre Cyfronet AGH, ul.Nawojki 11, 30-950 Kraków, Poland
(2) AGH University of Science and Technology, Department of Computer Science, Al. Mickiewicza 30, 30-059 Kraków, Poland
(3) Department of Bioinformatics and Telemedicine, Medical College, Jagiellonian University, Św. Łazarza 16, 31-530 Kraków, Poland
(4) Informatics Institute, University of Amsterdam, POBox 94323, 1090 GH Amsterdam, The Netherlands

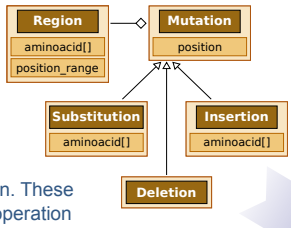
Motivation

The Semantic Integration approach to modeling helps **transcribe complex domain knowledge into well-structured information models based on semantics**. We present how this methodology has been applied to deliver real-life solutions to computational challenges in such areas as bioinformatics, the *Virtual Physiological Human* model and other life science domains.

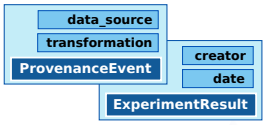
Methodology Overview

Goal: to build a new scientific application based on domain expert knowledge. The diagram shows how Semantic Integration helps in this task.

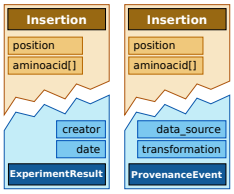
1. We start with domain model **entities** and their **fields** which represent core **concepts** of the application domain. These are defined in cooperation with the domain expert.



2. We add **facets**: behaviors or formats which domain **entities** may exhibit in our system. Facets form the integration model which represents how the application should work. This step may be performed by application developers independently.

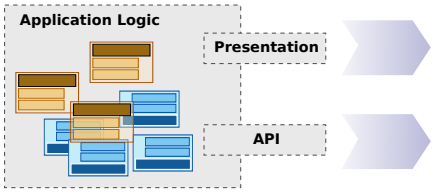


3. Semantic Integration performs automatic juxtaposition of both models (domain and integration) where all **entities** are able to have all **facets** (in essence, a "Cartesian product"). This is where science (domain) meets technology (integration).



4. The resulting complete model serves as the foundation for application logic. We build the logic layer using **facets** to store, embed, present, reformat and transfer domain **entities**.

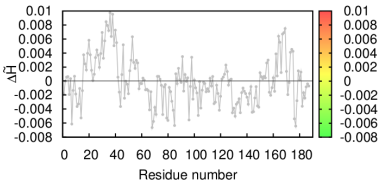
5. Based on the same logic and a common model we now provide a UI presentation layer (e.g. Web-based) for human users and an API (e.g. REST-based) for other, external or internal tools, services etc. In this way we semantically integrate tools for scientists and developers.

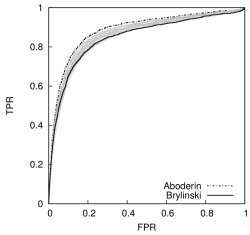


Case Study - Bioinformatics

Objective: to apply the Semantic Integration approach to develop **domain** and **integration** models for storage of *protein hydrophobicity profiles*.

The idea is to record a series of such profiles for each protein under investigation. Simulations are run on demand to compute, store and publish profiles.





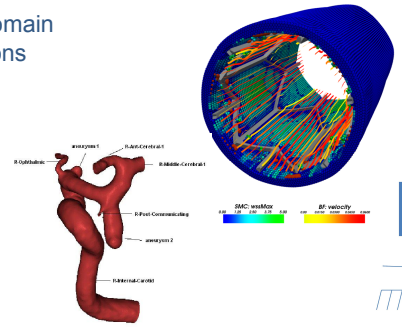
The application has to process and store large volumes of data (1.5 GB), and effectively perform on-demand lookups and presentations of protein profiles (including complex graph analyses such as ROC curves).

Outcome:
- distribution of concern allows us to focus on the core domain model when interacting with scientists
- any iterative change in the domain model is easily propagated to logic and user interface layers
- application designers and scientists may freely alter the *domain* model without affecting the *integration* model
- the framework may be adapted to the requirements of new application domains while retaining overall genericity.

Further Applications of Semantic Integration

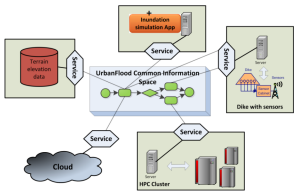
Biology and medicine

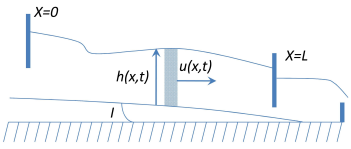
- modeling of scale separation domain knowledge for blood flow simulations (vascular stents)
- support for workflow definition, management and execution of multi-scale brain aneurism simulations
- building a registry and common semantic API for a set of cloud-based *Virtual Physiological Human* simulations.



Crysis management

- providing a common information exchange and storage model for early warning systems running ad-hoc flood simulations, including a dedicated high-throughput store for sensor data received from geographically distributed water dikes





- enabling registration and reuse of workflows of multidimensional, complex irrigation system simulations.

(*) Pictures inside this box are a courtesy of consortium members of projects: MAPPER (EU grant 261507), VPH-Share (EU grant 269978) and UrbanFlood (EU grant 248767).

1. T. Gubała, M. Bubak, P.M.A. Sloot; *Semantic Integration of Collaborative Research Environments*, In: M. Cannataro (Ed.) Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine and Healthcare, chapter XXVI, pages 514-530, 2009, IGI Global.
2. M. Bubak, T. Gubała, M. Kasztelnik, M. Malawski: *Building Collaborative Applications for System-level Science,* In: W. Gentzsch, L. Grandinetti, G. Joubert (eds.) Advances in Parallel Computing, Vol. 18, 2009, ISBN 978-1-60750-073-5, doi: 10.3233/978-1-60750-073-5-299
3. M. Brylinski, K. Prymula, W. Jurkowski, M. Kochanczyk, E. Stawowczyk, L. Konieczny, I. Roterman. *Prediction of functional sites based on the fuzzy oil drop model.* PLoS Comput Biol, 3(5), 2007
4. M. Bubak, T. Szepieniec, K. Wiatr, (editors). *Building a National Distributed e-Infrastructure - PL-Grid - Scientific and Technical Achievements.* Springer 2012.

dice.cyfronet.pl