PhD Topic 1: Platform-as-a-Service for Environmental Informatics

First Supervisor	Prof. Dr. Schahram Dustdar
Second Supervisor	Prof. Dr. Günther Raidl
Primary Field	Large-scale data- and compute-intensive systems
Secondary Field	Resource Management
Interdisciplinary References	D2, D3, D4

Environmental domains, such as resource management, green business and smart solutions, will, on the one hand, bring different types of data that are integrated into data services, and, on the other hand, introduce several stakeholders that do not rely on a single way to program, analyze, and simulate domain-specific problems. For example, data analytics will be implemented in different languages, such as Java, Matlab and R. To enable efficient management and execution of diverse simulations, data analytics and decision support is required. Fostering the development and optimization of algorithms, processes and applications for such simulations, analytics and decision support, requires scalable and flexible platforms to deal with the scale and the complexity of data services, algorithms, and programs for multiple stakeholders. Especially, environmental simulations, analytics, and decision supports are data-intensive and compute-intensive and will involve different stakeholders tackling multi-objective optimization. To this end, it is crucial to support programmable platform-as-a-service for scientists, developers, community users and other types of stakeholders.

The research topic on platform-as-a-service for environmental informatics (PaaSEI) will aim at developing a novel cloud-based platform-as-a-service integrating multiple types of data services and computational services for visual data analytics, near-real time monitoring and analysis, and simulations of environmental problems. The first challenge that this topic has to deal with is to support different models of analytics and simulation programs – batch sequential, parallel, and workflow-based programs as well as streaming ones – written in different languages to be executed in the same platform-as-a-service. To this end, scalable and extensible execution environments must be developed by utilizing cloud computing technologies/models and parallel and distributed processing techniques. The second challenge is that it needs to address and manage the dependencies among data services, analytics services, simulation services, governance policies and stakeholders that enable the development and optimization of novel simulations, data analytics processes, and decision supports. To this end, novel techniques for capturing and managing information about

stakeholders, their views and constraints on services for environmental informatics must be developed. Furthermore, domain-specific languages can be investigated for environmental domain experts to program analytics and simulations atop PaaSEI, enabling them to focus on solving domain problems without having extensive IT knowledge required for dealing with complex data and computational services.

References:

[1] Hong-Linh Truong, Schahram Dustdar: A Survey on Cloud-based Sustainability Governance Systems. International Journal of Web Information Systems, March, 2012. To appear.

[2] Hong Linh Truong, Schahram Dustdar: Cloud computing for small research groups in computational science and engineering: current status and outlook. Computing 91(1): 75-91 (2011)

[3] Michael Reiter, Uwe Breitenbucher, Schahram Dustdar, Dimka Karastoyanova, Frank Leymann, Hong Linh Truong: A Novel Framework for Monitoring and Analyzing Quality of Data in Simulation Workflows. eScience 2011: 105-112

[4] Matija König, Hong-Linh Truong, Schahram Dustdar, and Vlado Stankovski. 2011. Information modelling for sustainable buildings. In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS '11). ACM, New York, NY, USA, 507-510. DOI=10.1145/2095536.2095641 http://doi.acm.org/10.1145/2095536.2095641