

# Gridification and virtualization: enabling e-Science in the life sciences

Dietmar Sommerfeld  
Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen  
Am Fassberg, 37077 Göttingen, Germany  
dsommer@gwdg.de

Thomas Lingner  
Department of Bioinformatics  
University of Goettingen  
Goldschmidtstr. 1, 37077 Goettingen, Germany  
thomas@gobics.de

Janko Heilgeist  
Computing Center of the Max-Planck-Society  
Boltzmannstr. 2, 85748 Garching, Germany  
jkh@rzg.mpg.de

Harald Richter  
Department of Computer Science  
Clausthal University of Technology  
Arnold-Sommerfeld-Str. 1,  
38678 Clausthal-Zellerfeld, Germany  
hri@tu-clausthal.de

## ABSTRACT

The MediGRID [1] project, which is part of the German e-Science initiative D-Grid, aims to provide a community Grid for researchers in the fields of bioinformatics, medical image processing, biomedical ontology, and clinical research applications. Users in the life science domain usually do not have a strong computer science background. Therefore, it is crucial to make the Grid solution as user friendly as possible. To achieve this goal, MediGRID is building up an infrastructure that follows two principles: gridification and virtualization.

The basic Grid middleware used in MediGRID is the Globus Toolkit 4, which provides services for security, execution and data management, and resource discovery. On top of Globus, MediGRID employs further specialized middlewares. The first is the Grid Workflow Execution Service (GWES) [2] that orchestrates the distributed execution of workflows on Grid resources. Application workflows consist of several successive program executions and intermediate data transfers. Second, the Storage Resource Broker (SRB) [3] Data Grid middleware is utilized to handle large amounts of distributed data and corresponding metadata. We use SRB to create a global logical namespace in which file replica can be accessed by unique file names, independent of the physical locations. At the top level, a portal framework is employed to provide a convenient graphical user interface.

The middleware layers allow to virtualize the MediGRID resources. This means, end users do not interact with specific clusters and storage elements but deal with a high-level abstraction of the Grid. For this, extensive gridification work is necessary to set up the Grid infrastructure and adapt the applications to Grid computing. We will demonstrate the whole process with the gene-finding application AUGUSTUS [4].

## REFERENCES

- [1] S. Kottha, et al., Medical image processing in MediGRID. In Proceedings of the German e-Science Conference, Baden-Baden, Germany, 2007.
- [2] A. Hoheisel, Grid Workflow Execution Service - Dynamic and interactive execution and visualization of distributed workflows. In Proceedings of the Cracow Grid Workshop 2006, Cracow, 2006.
- [3] C. Baru, R. Moore, A. Rajasekar, M. Wan, The SDSC storage resource broker. In CASCON '98: Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research, IBM Press, 1998, p. 5.
- [4] M. Stanke, O. Schoffmann, B. Morgenstern, and S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources, BMC Bioinformatics, vol. 7, p. 62, 2006.