M. Sonntag[1] S. Hotta[1] D. Karastoyanova[1] D. Molnar[2] S. Schmauder[2]

Workflow-Based Distributed Environment For Legacy Simulation Applications

Stuttgart, May 2011

[1] **Institute of Architecture of Application Systems (IAAS)**
University of Stuttgart, Universitaetsstrasse 38, 70569 Stuttgart, Germany
{sonntag, hotta, karastoyanova}@iaas.uni-stuttgart.de
http://www.iaas.uni-stuttgart.de

[2] **Institute of Materials Testing, Materials Science and Strength of Materials (IMWF)**
University of Stuttgart, Pfaffenwaldring 32, 70569 Stuttgart, Germany
{david.molnar, siegfried.schmauder}@imwf.uni-stuttgart.de
http://www.imwf.uni-stuttgart.de

Abstract Computer simulations play an increasingly important role to explain or predict phenomena of the real world. We recognized during our work with scientific institutes that many simulation programs can be considered legacy applications with low software ergonomics, usability, and hardware support. Often there is no GUI and tedious manual tasks have to be conducted. We are convinced that the information technology and software engineering concepts can help to improve this situation to a great extent. In this poster presentation we therefore propose a concept of a simulation environment for legacy scientific applications. Core of the concept are simulation workflows that enable a distributed execution of former monolithic programs and a resource manager that steers server work load and handles data. As proof of concept we implemented a Monte-Carlo simulation of precipitations in copper-alloyed iron and tested it with real data.

Keywords Simulation workflows, distributed simulations, BPEL, Web services, Monte-Carlo.

Reference Sonntag, M., Hotta, S., Karastoyanova, D., Molnar, D. and Schmauder, S. (2011) "Workflow-Based Distributed Environment For Legacy Simulation Applications". In: Proceedings of the 6th International Conference on Software and Data Technologies (ICSOFT 2011), Poster Paper, SciTePress Digital Library.

Stuttgart Research Centre for Simulation Technology (SRC SimTech)

SimTech – Cluster of Excellence
Pfaffenwaldring 7a
70569 Stuttgart
publications@simtech.uni-stuttgart.de
www.simtech.uni-stuttgart.de

Workflow-based Distributed Environment for Legacy Simulation Applications

Mirko Sonntag[1], Sven Hotta[1], Dimka Karastoyanova[1], David Molnar[2], Siegfried Schmauder[2]
[1] Institute of Architecture of Application Systems (IAAS)
[2] Institute of Materials Testing, Materials Science and Strength of Materials (IMWF)

Motivation

Many simulation applications are based on legacy software:

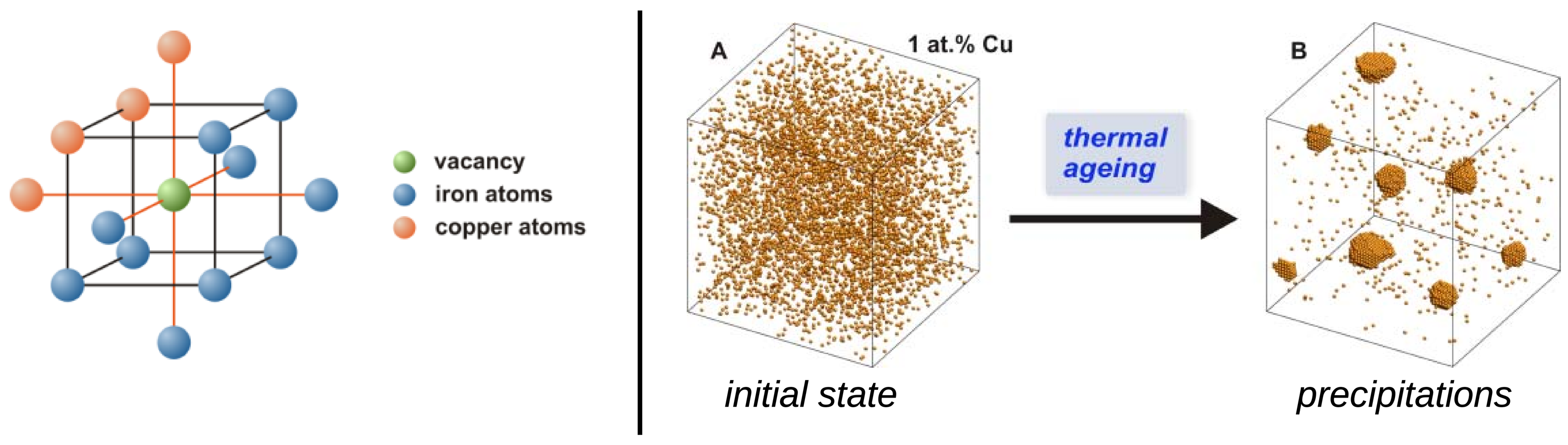
- The Software was developed over years by many different authors
- There is no time, money or knowledge to re-implement the tools in a modern programming language
- The software often does not benefit from multi-core CPUs and distributed computing
- The programs often cannot deal with parallel invocations
- Scientists have to carry out manual tasks (e.g. copy files, start tools)

Workflows possess great potential to improve the tool support for many scientists. There is a need to:

- Automate and optimize simulation processes
- Exploit recent development in hard- and software
- Improve user-friendliness and flexibility of simulation applications
- Integrate different scientific applications

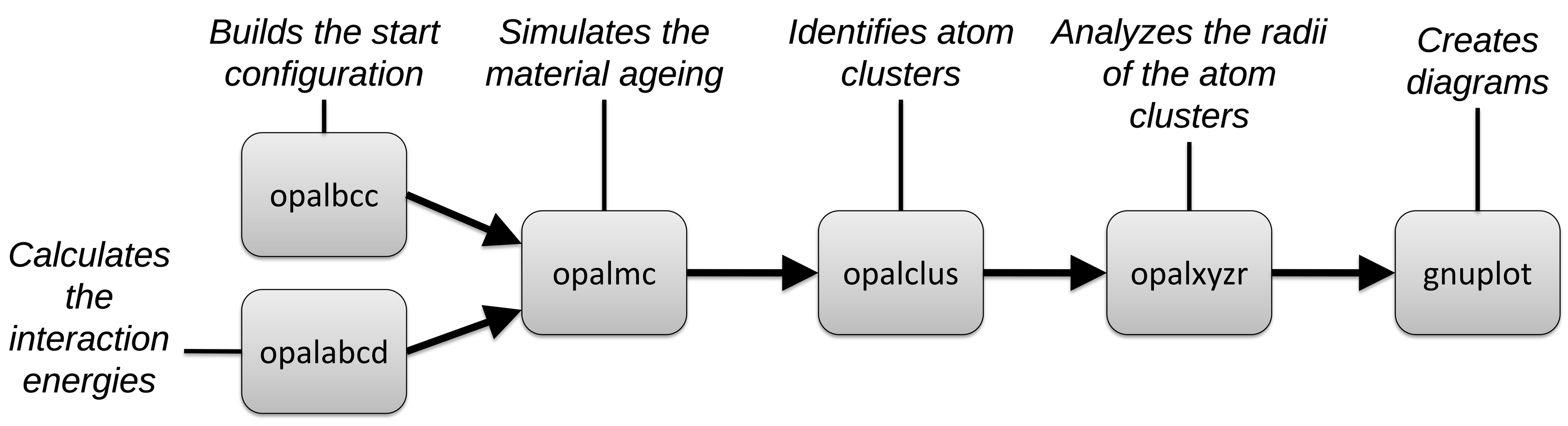
Use Case: MC-Simulation for Solid Bodies

The example simulation investigates the thermal ageing of copper-alloyed steel on an atomistic scale. The steel is modeled as body-centered cubic (bcc) crystal lattice with different atom types and a single vacancy. Over time, the atoms change their positions by jumping into the vacancy. Atom clusters (a.k.a. precipitations) develop which decrease the material strength [1].



Current simulation tool and disadvantages

The simulation is carried out with a legacy tool called opal [2] that consists of five monolithic Fortran 77 programs and gnuplot. The tools have no GUI and some parameters are hard-coded. The programs are started manually by the scientists. This is tedious and time-consuming because files have to be copied and commands have to be typed in a shell. The simulation runs only sequentially because opal allocates a complete CPU and Fortran 77 does not benefit from multi-core processors. That means the intermediary files produced by opalmc can be processed not until opalmc finishes.

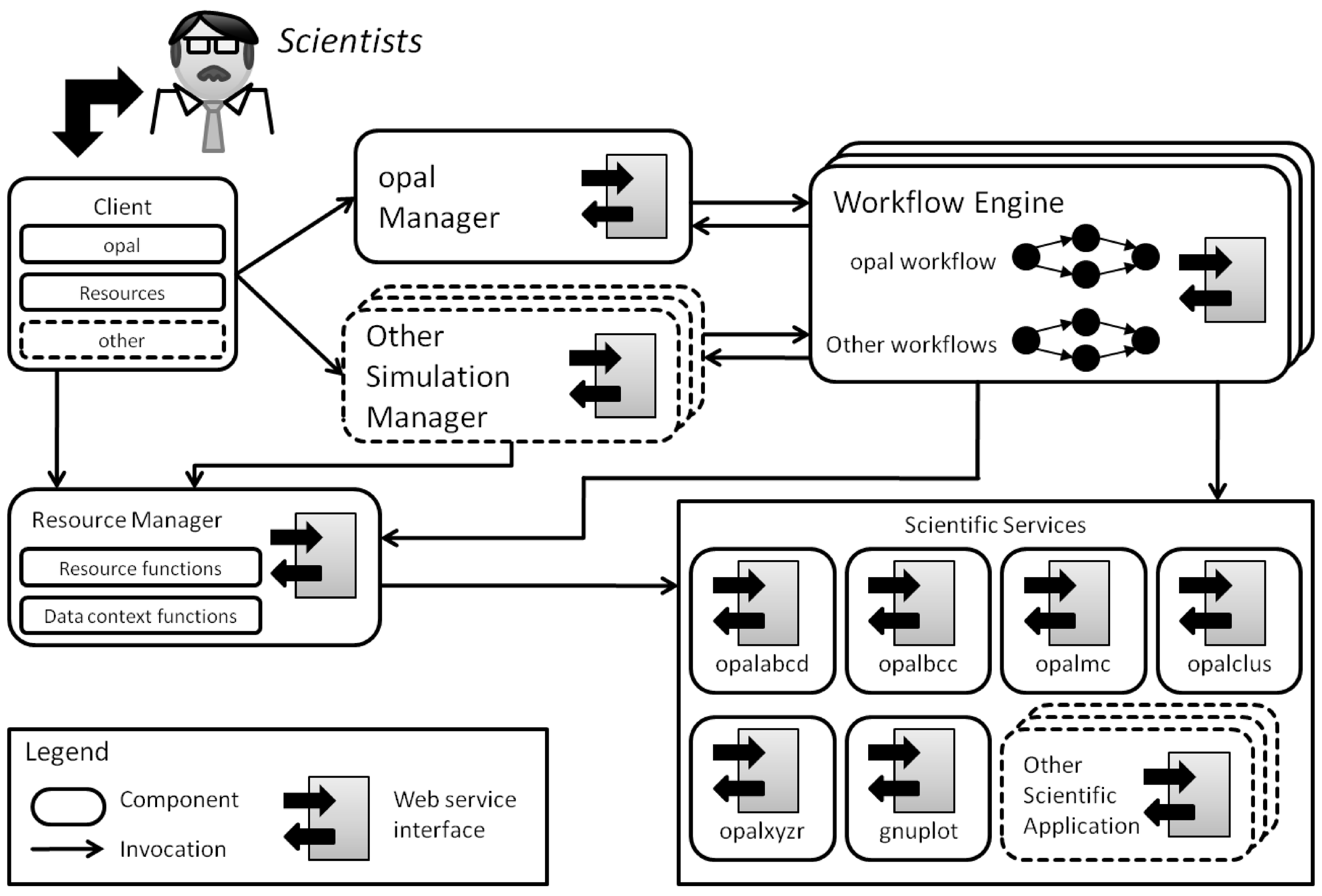


Workflow-based Simulation Environment

The situation is improved when the legacy software is re-engineered based on workflows and Web Services (WS). Workflows enable the automation of manual steps, the integration of different pieces of software as well as the distributed and parallel execution of software. We have developed an extendible architecture for the distributed execution of legacy simulation applications with workflows as centerpiece.

Architecture

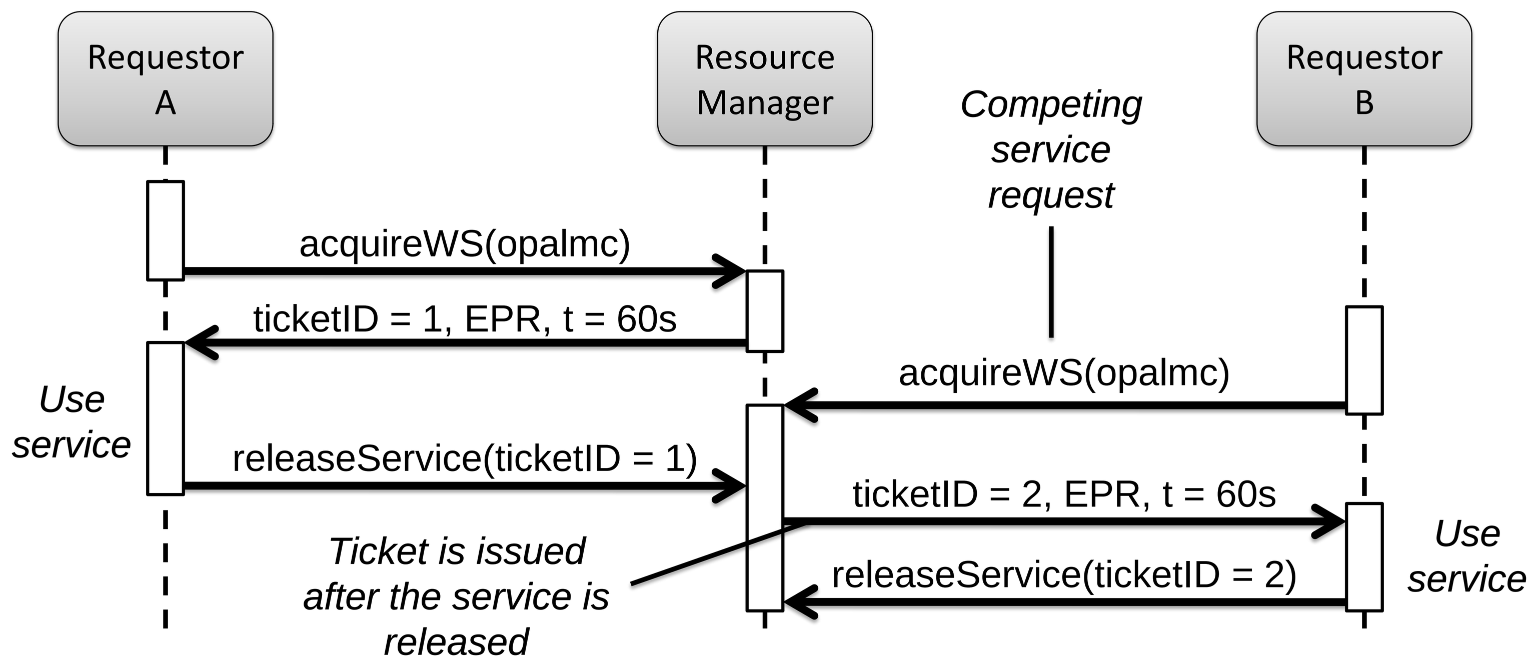
- **Scientific Services:** Thin WS wrappers make the legacy programs accessible in a network. Some scientific programs have to be previously adapted, some can be taken unchanged. This strongly depends on their interfaces and has to be decided on a per-case basis. E.g., we adapted the opal programs so that parameters can be passed by call instead of being hard-coded.



- **Scientific Workflows:** Each simulation requires tailored workflows. We implemented the opal use case two BPEL workflows. Further, we derived generic, re-usable workflow fragments, e.g. to allocate/de-allocate scientific services with the resource manager. These can be used to easily implement other use cases.
- **Simulation Manager:** A service on top of each simulation use case provides domain-specific functionality, e.g. to select and create parameter files, to start and control simulations. It hides the underlying infrastructure from the client.

BUT: Workflows do not solve GUI and data handling issues out of the box. We need to introduce separate components and mechanisms.

- **GUI:** The GUI improves the software ergonomics. It provides general and domain-specific functions to the scientists. The *opal* part contains a wizard to parameterize and start opal simulations and to visualize intermediary results. The *resources* part allows to register participating servers, to monitor the work load and to access produced files.
- **Resource Manager:** The resource manager supervises the registered servers in the network, their properties and installed services. It controls utilization of the services by issuing usage tickets. Moreover, it works as data storage for simulation data. Hence, data can be passed by reference within the system



References

[1] Molnar, D., Binkele, P., Hocker, S., Schmauder, S.: Multiscale Modelling of Nano Tensile Tests for Different Cu-precipitation States in α-Fe. Proc. of the 5th Int. Conf. on Multiscale Materials Modelling, pp. 235-239, Fraunhofer Verlag (2010)
[2] Binkele, P., Schmauder, S.: An atomistic Monte Carlo simulation for precipitation in a binary system. International Journal for Materials Research, 94, pp. 1-6 (2003)