Supporting the Data Dimension of Multi-* Simulations through Data-Aware Service Choreographies

Michael Hahn, Dimka Karastoyanova, Frank Leymann
Institute of Architecture of Application Systems (IAAS), University of Stuttgart, Germany

Motivation
Efficient data management, placement and exchange are crucial tasks in scientific simulations, especially if multiple simulations are coupled as so-called multi-* simulations. The notion of choreographies [1] enables scientists to model such coupled simulations from a global perspective through the interconnection of different scientific workflows and simulation software without the need to directly provide complex technical details [3].

- Choreography models currently reflect only the control perspective of the interconnection and lack specification of data-related aspects.
- However introducing data on the level of the choreography would enable scientists to model data flow explicitly and as a result to automate the staging, placement and exchange of data during run time in an optimal manner.
- Therefore, corresponding concepts for data-aware choreography modeling, execution and data flow optimization are required in order to improve the overall acceptance of choreographies.

The main goals of the project are to:
- extend the choreography meta model and the modeling and run time environment to support data-aware choreographies,
- investigate data flow optimization concepts for simulation choreographies, i.e. optimization of data placement, exchange, staging and placement
- and to create an approach and realization for a Transparent Data Exchange (TraDE) Middleware that supports the defined concepts.

State of Current Work
Fig. 1 shows our data-aware management life cycle for choreographies which is based on the choreography life cycle described in [4]. The idea is to introduce data already on the level of the choreography and support the scientists throughout the whole life cycle regarding the data perspective of their simulations.

- Scientists model their coupled simulations as a data-aware choreography.
- The choreography model and its data model are transformed to a collection of abstract workflow models.
- The abstract models have to be refined manually or automatically by specifying the simulation control and data flow between the orchestrated simulation services and software.
- The refined models are deployed to one or more Scientific Workflow Management Systems (SWIMS), e.g. using [2]

- The deployed workflow models are instantiated to conduct the coupled simulation. During execution corresponding control and data flow events are emitted.
- The emitted events of the workflows and the data are collected and analyzed during monitoring to ensure that everything is carried out as expected.
- The choreography and workflow models together with the results from monitoring are taken into consideration to find possible optimizations in the analysis phase.
- The new Transparent Data Exchange (TraDE) phase supports data awareness throughout the whole life cycle and potential optimizations regarding the data perspective of choreographies.

The choreography data flow is realized through a network of TraDE Middleware nodes that act as a new intermediate layer between the services, service middleware, physical data sources and users as shown in Fig. 2.

Future Work
- Architecture and realization of TraDE Middleware nodes and networks
- Integration of TraDE Middleware with existing SWIMS
- Identification and realization of data exchange optimization strategies for the different life cycle phases

References