

Guido Scherp: A Framework for Model-Driven Scientific Workflow Engineering

1. **Gutachter:** Prof. Dr. Wilhelm Hasselbring (Universität Kiel)

2. **Gutachter:** Prof. Dr. Odej Kao (Technische Universität Berlin)

Datum der Prüfung: 25.06.2013

Zusammenfassung:

So genannte *Scientific Workflows* werden zunehmend im Kontext daten-intensiver Wissenschaften eingesetzt, um komplexe Verarbeitungen von Forschungsdaten effizient und zuverlässig in verteilten Infrastrukturen wie Grids auszuführen. *Scientific Workflow Management Systeme* (SWfMS) unterstützen Wissenschaftler in der Modellierung und Ausführung von Scientific Workflows, wobei oft zwischen der Modellierung durch einen Wissenschaftler auf der *domänenspezifischen Ebene* und der automatisierten Ausführung auf der *technischen Ebene* unterschieden wird. Erste SWfMS wurden von Grund auf neu entwickelt inklusive entsprechender Workflow-Technologien und -Sprachen. Bereits existierende und etablierte *Business Workflow*-Technologien aus dem betrieblichen Bereich wurden ursprünglich nicht genutzt, beispielsweise weil Scientific und Business Workflows unterschiedliche Lebenszyklen abbilden und auf Grund inkompatibler Schnittstellen und Kommunikationsprotokolle der jeweiligen Infrastrukturen.

Im Zuge der Etablierung von Service-orientierten Architekturen (SOAs) in betrieblichen IT-Infrastrukturen wurden zahlreiche Web Service-Standards und entsprechende Technologien entwickelt. Die *Web Services Business Process Execution Language* (BPEL) ist beispielsweise ein Standard für die Implementierung und Ausführung von Business Workflows in einer SOA. So genannte *Service Grids* haben das SOA-Architekturmuster für wissenschaftliche IT-Infrastrukturen übernommen und nutzen dabei die bereits existierenden Standards und Technologien. Somit ist BPEL generell auch für die Ausführung von Scientific Workflows auf der technischen Ebene geeignet, was bereits in zahlreichen Projekten und Publikationen gezeigt wurde. Allerdings ist BPEL eine Workflow-Sprache für IT-Experten und kann in der Form nicht zur Modellierung eines Scientific Workflows durch einen Wissenschaftler auf der domänenspezifischen Ebene genutzt werden. Es fehlen eine geeignete Abstraktion für BPEL, die speziell für den Einsatz auf der domänenspezifischen Ebene von Scientific Workflows zugeschnitten ist, sowie eine passende Abbildung auf

die technische Ebene.

Diese Herausforderungen der Abstraktion und der Abbildung werden in der Dissertation mit Hilfe der *Business Process Model and Notation* (BPMN) und Techniken aus der *modellgetriebenen Softwareentwicklung* adressiert. Dazu wird mit *MoDFlow* ein modellgetriebener Ansatz vorgestellt, um domänenspezifische Scientific Workflow-Modelle über eine auf BPMN basierende Zwischenschicht in eine technisch ausführbare Form zu überführen. Die Zwischenschicht basiert auf MoDFlow.BPMN, was eine Untermenge von BPMN definiert mit eigenen Erweiterungen für die wissenschaftliche Domäne. MoDFlow.BPMN2BPEL beschreibt drei aufeinanderfolgende Transformationsschritte zur Abbildung von MoDFlow.BPMN nach BPEL auf der technischen Ebene. Zudem werden in MoDFlow mehrere Möglichkeiten beschrieben, um MoDFlow.BPMN und MoDFlow.BPMN2BPEL zu nutzen und zu erweitern. Ein Schwerpunkt liegt dabei auf der Erstellung von so genannten domänenspezifischen Sprachen (DSLs) zur Modellierung von Scientific Workflows auf der domänenspezifischen Ebene. Mit dem *MoDFlow-Framework* wird eine Implementierung des Ansatzes bereitgestellt, die auf dem Eclipse Modeling Framework (EMF) aufsetzt.

Das MoDFlow-Framework wird in drei Anwendungsszenarien evaluiert, wobei unterschiedliche Nutzungsmöglichkeiten angewendet werden. Die ersten beiden Szenarien definieren Parameterraumexplorationen in Scientific Workflows und führen diese in einer Grid-Infrastruktur aus. Damit wird die technische Realisierbarkeit des Ansatzes gezeigt. Das dritte Szenario ist eine Kooperation mit dem Projekt PubFlow, in dem eine Infrastruktur zur Erstellung und Ausführung von Scientific Workflows für Datenpublikationen aufgebaut wird. Mit Hilfe des Frameworks Xtext wird eine textuelle DSL erstellt, die zunächst Entwickler beim Umgang mit Workflow-Technologien unterstützt. Dies Szenario zeigt die praktische Nutzbarkeit des MoDFlow-Frameworks. PubFlow plant im nächsten Schritt eine grafische DSL basierend auf der Notation von BPMN sowie einen entsprechenden Editor speziell für Wissenschaftler zu erstellen.

Veröffentlicht als:

Guido Scherp. A Framework for Model-Driven Scientific Workflow Engineering, Kiel Computer Science Series, 2013.

Online unter <http://eprints.uni-kiel.de/21932/>