

# Auf dem Weg zu einer Cyber-Infrastruktur für die Simulationstechnik



Simulationstechnik ist seit Jahrzehnten unersetzlich bei der Erforschung komplexer, naturwissenschaftlicher und ingenieurwissenschaftlicher Phänomene, aber auch in den Lebenswissenschaften und den Geisteswissenschaften. Der Fortschritt in den Techniken für Modellierung und Simulation und auch im Bereich des Hardware- und Software-Komponentenentwurfs hat dazu beigetragen, neuartige Synergien zwischen Wissenschaftlern und den Informationstechnologien zu identifizieren. Im Rahmen des Exzellenzclusters „Simulation Technology“ (SimTech) der Universität Stuttgart liegt ein Fokus auf der Entwicklung von Methoden und Techniken für die IT-Unterstützung. Die bis dato isolierten Simulationsansätze und IT-Infrastrukturen sollen zusammengebracht, also integriert, werden, um so eine ganzheitliche Systemwissenschaft zu entwickeln.

## 1. Einführung

Ein Hauptziel von SimTech ist es, mit einer neuartigen und mächtigen IT-Infrastruktur die IT-Unterstützung für Simulationen zu verbessern. Diese soll den Wissenschaftlern helfen, sich auf die eigentlichen Aufgaben ihrer Forschung zu konzentrieren. Eine solche Infrastruktur, auch Cyber-Infrastruktur genannt, umfasst ein Netz von Hochleistungsrechenplattformen, Software-Programmen und -Werkzeugen, Sensornetze, Speicher, Rechenressourcen- und Datenverwaltungssysteme sowie Visualisierungsumgebungen. Auf dem Weg zur Cyber-Infrastruktur wurden bereits in Zusammenarbeit mit Naturwissenschaftlern IT-Ansätze für die Datenverwaltung und Visualisierung entwickelt, mit denen die Effizienz der Modellierung und die Ausführung von Simulationen, sowie die interaktive Visualisierung der Simulationen und der Simulationsergebnisse unterstützt werden. Die Cyber-Infrastruktur soll die Anforderungen heutiger Simulationstechnik erfüllen wie beispielsweise Automatisierung von Simulationsverfahren, Benutzerfreundlichkeit, Wiederverwendbarkeit der Simulations-Software und Ressourcen, effizienter Umgang mit großen Datenmengen und Datenbereitstellung, Langläufigkeit der Simulationen, interaktiver Umgang zwischen Simulationen und Wissenschaftlern, Bereitstellung von Echtzeitdaten aus Sensornetzen und mobilen Geräten für die Simulationen, interaktive Visualisierung und Analyse von großen Datenmengen, die auch Unsicherheiten aufweisen können. Die Cyber-Infrastruktur soll zudem hochskalierbar sein und eine dynamische Ressourcen- und Datenverwaltung bereitstellen.

In diesem Beitrag geben wir eine Übersicht über unsere Forschungserfolge und stellen unsere Ideen für die zukünftige Forschung vor.

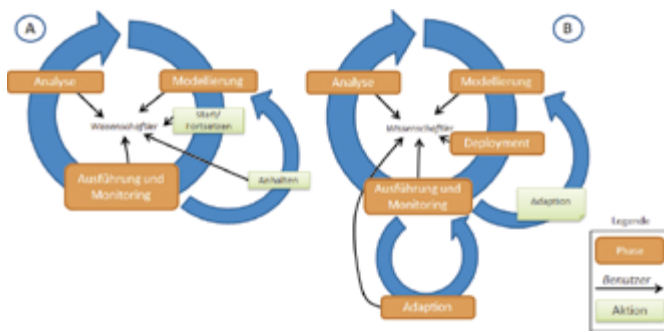
## 2. Infrastruktur für interaktive Simulationen

### 2.1. Konzepte für die Modellierung und Ausführung von Simulationsworkflows

Ein wichtiges Ziel der bisherigen Forschung von SimTech war die Konzeption und Realisierung einer IT-Infrastruktur für

Simulationen, die an die Bedürfnisse der Wissenschaftler während ihrer Forschung angepasst ist. Haupteigenschaft einer solchen Infrastruktur muss die Benutzerfreundlichkeit sein, also eine einfache Nutzung sowohl bei der Modellierung von Experimenten aus unterschiedlichen naturwissenschaftlichen und ingenieurwissenschaftlichen Bereichen als auch bei der Ausführung und Anpassung des Verlaufs eines Experiments und ebenso bei der Visualisierung der Ergebnisse.

Um diese Ziele zu erreichen, wurden bewährte Konzepte, Techniken und Technologien aus der Informatik und Software-Technik angewandt, die hohe Akzeptanz im Geschäftsbereich genießen, wie etwa die Workflow-Technologie. Unter Verwendung des Workflow-Konzepts wurde die *Modellierung* (siehe **(01A)**) einer Simulation einfacher gestaltet, da die Experimente mit Hilfe eines grafischen Modellierungswerkzeugs schnell, effizient und einfach zu erstellen sind, ohne jegliches Wissen über die Software, die diese Schritte genau umsetzt (also implementiert). Ergebnis der Modellierung ist ein Workflow-Modell, das eine Simulation darstellt. In einem Workflow-Modell werden sowohl die Schritte (auch Aktivitäten genannt) einer Simulation definiert als auch die richtige Reihenfolge, in der diese Schritte ausgeführt werden müssen. Die bedingte Ausführung von Aktivitäten ist einfach zu modellieren, ebenso die sequenzielle als auch parallele Ausführung. Die Aktivitäten in einer Simulation können dabei für Datenbearbeitungsprogramme oder Berechnungsfunktionen stehen, welche in naturwissenschaftlichen Simulationen gebraucht werden. Das erstellte Modell kann parametrisiert werden, sodass es mit unterschiedlichen Parameterwerten ausgeführt werden kann. Diese Parameter werden aber erst bei der Ausführung angegeben, somit wird die Wiederverwendbarkeit der Modelle ermöglicht. Die Wiederverwendbarkeit der Simulationsmodelle wird zusätzlich gesteigert durch die Entkopplung des Modells von den Software-Programmen, mit denen die einzelnen Schritte realisiert werden. Das führt dazu, dass das gleiche Simulationsmodell, also der gleiche Simulationsworkflow, mehrmals verwendet werden kann um unterschiedliche Software für die Ausführung der Simulationsschritte zu kombinieren. Die Verbindung zwischen einer Akti-



01

Lebenszyklus von Simulationen (A)  
und Workflows (B)

Lebenszyklus von Simulationen (A) und Workflows (B). Die Simulation festgelegt. Damit wird es möglich, die gleiche Simulation mehrfach bereit zu stellen, sodass die gleiche Aktivität in den unterschiedlichen Bereitstellungen von unterschiedlicher Software durchgeführt werden kann. Das ist vor allem wichtig, wenn sich z.B. die Qualität der Ergebnisse, die die Software produziert, oder die Dauer der Berechnungen unterscheidet. Die Wahl einer passenden Bereitstellung ist vor allem dann von Bedeutung, wenn die Simulationsergebnisse sehr präzise sein sollen oder die Simulation oder Teile davon sehr schnell ablaufen sollen. Mit Hilfe der Workflow-Technologie konnten wir die *Automatisierung* der wissenschaftlichen Experimente/Simulationen steigern, da sich viele manuelle Schritte in einem Simulationsmodell einfach automatisieren lassen. Darüber hinaus lassen sich die modellierten Workflow-Schritte automatisch von einem so genannten Workflow Management System (WfMS) ausführen. Die Simulationsworkflows werden erst dann als ausführbar betrachtet, wenn diese auf dem Workflow Management System bereitgestellt werden. Nach der Bereitstellung des Simulationsmodells kann die Simulation mehrmals parallel ausgeführt und damit die Zeit der Durchführung aller benötigten Simulationsdurchläufe verringert werden. Die *Ausführung* der Simulationsdurchläufe (auch Workflow-Instanzen genannt) passiert automatisch und wird von der Navigator Komponente des WfMS verwaltet. Der Navigator ist dafür verantwortlich der bedingten Ausführungsreihenfolge des Simulationsmodells richtig zu folgen (durch zu navigieren) und alle auszuführenden Aktivitäten in der korrekten Reihenfolge mit den korrekten Eingangsdaten anzustoßen. Das Anstoßen jeder Aktivität wird in Zusammenarbeit mit einem Middleware-System

realisiert. Der Navigator übergibt dieser Middleware die Eingangsdaten für die Ausführung der Aktivität und Informationen darüber, welches Software-Programm oder welcher Typ von Software-Programm die Aktivität ausführen soll. Zusätzlich können auch Anforderungen an die Qualität der Software, wie z.B. Antwortzeit, Datenqualität oder Sicherheit angegeben werden. Die Middleware agiert als Mittler, der das Programm anhand dieser Anforderungen auffindet, ihm die Eingangsdaten übergibt und somit die Ausführung der Simulationsaktivität startet. Nach Beendigung dieser Ausführung werden die Ergebnisdaten der Middleware übergeben, welche diese Daten dann an den Navigator weiterleitet. Anschließend kennzeichnet der Navigator die entsprechende Aktivität in der entsprechenden Workflow-Instanz als ausgeführt und setzt die Navigation über das Simulationsmodell fort. Diese Aktionen wiederholen sich für jede Aktivität in jeder Workflow-Instanz.

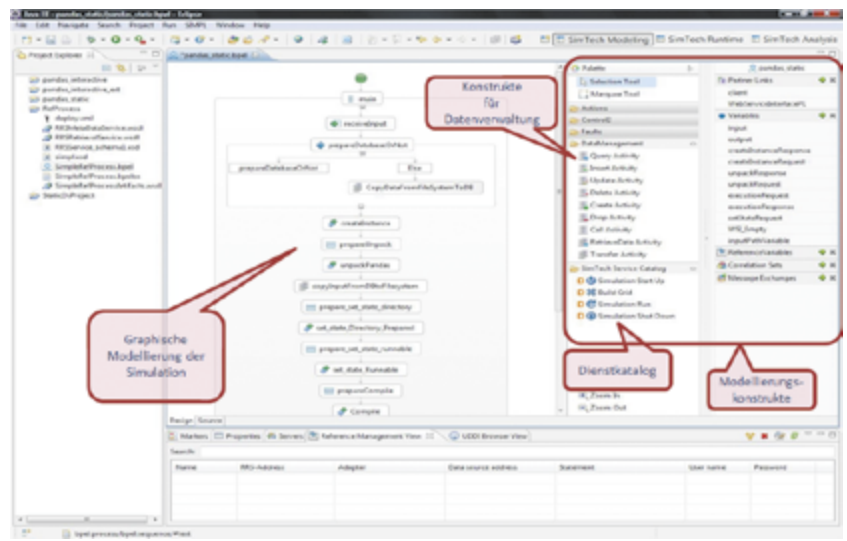
Zusätzlich zu der vereinfachten Modellierung, automatischen Ausführung und verbesserten Effizienz können wir aus der Workflow-Technologie weitere Vorteile wie beispielsweise Skalierbarkeit, Zuverlässigkeit und Fehlertoleranz übernehmen [1]. Da die Software, die im naturwissenschaftlichen und ingenieurwissenschaftlichen Bereich erstellt und benutzt wird, sehr heterogen ist, d.h. es werden diverse Programmiersprachen, Plattformen und Middleware verwendet, besteht oft das Problem, dass diese nicht unmittelbar mit einander kommunizieren können. Diese Problematik kann mit unterschiedlichen Ansätzen der *Software-Integration* angegangen werden. Wir haben vor allem den Ansatz der *Service-Orientierung* (Dienstorientierung) verfolgt. Service-orientierte Systeme stellen Software als Dienste (Services) zur Verfügung. Dienste sind ein Ansatz, um existierende Software zur Benutzung bereit zu stellen; d.h. Dienste sind kein Paradigma zur Programmierung von Software. Die Software wird über stabile Schnittstellen bereitgestellt, im Fall der Web-Service-Technologie werden diese Schnittstellen in einer einheitlichen Beschreibungssprache (WSDL – Web Service Description Language) beschrieben. Wir haben die Web-Service-Technologie verwendet, um existierende Simulations-Software als Dienste in Simulations-Workflows einheit-

lich verwenden zu können. Die Dienstorientierung konnte die Heterogenität der Simulations-Software beseitigen und hat die Integration für die Wissenschaftler erleichtert. Konkret haben wir gängige Service-Middleware um generische Adapter für unterschiedliche Simulations-Software erweitert wie z.B. Opal [3], Dune und PAN-DAS.

Diese Dienste wurden in einem grafischen Modellierungs-Werkzeug zur Verfügung gestellt, so dass sie von den Wissenschaftlern direkt in einem Simulationsmodell eingebunden werden können (siehe (02)). Damit wird die Komplexität der Software-Integration vor den Wissenschaftlern verborgen und die Benutzerfreundlichkeit der gesamten Infrastruktur verbessert.

Viele Schritte während einer Simulation werden von Wissenschaftlern durchgeführt, wie z.B. die Bewertung von Zwischenergebnissen oder die Parametereingabe. Die besondere Berücksichtigung von menschlichen Benutzern in dienstorientierten Simulations-Workflows war eine der Herausforderungen in der Forschungsarbeit. Um für die Wissenschaftlern angepasste Möglichkeiten zu schaffen, über gängige Kommunikationsmittel mit den Simulationsworkflows zu interagieren, wurde die Architektur des WfMS um akzeptierte Interaktionskanäle wie beispielsweise Benachrichtigungen und Rückmeldungen per E-Mail oder Skype erweitert [2]. Mit diesen Techniken ist es zum Beispiel möglich, dass ein Wissenschaftler eine lang laufende Simulation startet und dann auf eine Konferenzreise geht. Treten während der Ausführung der Simulation Fehler auf, so wird der Wissenschaftler per Skype oder, falls er zu diesem Zeitpunkt nicht online ist, per E-Mail benachrichtigt. Er kann dann, wiederum per Skype oder E-Mail, auf diese Meldung reagieren und beispielsweise Parameter so verändern, dass die Simulation weiter ausgeführt werden kann.

Wissenschaftler folgen sowohl bei der Erstellung von Simulationen als auch bei der Durchführung von Experimenten fast immer der Trial-and-Error-Methode, also dem beständigen Wechsel zwischen der Modellierung und der Ausführung einer Simulation, ein Kennzeichen der interaktiven Natur der naturwissenschaftlichen Experimente. Dieser Hauptunterschied sowie weitere Unterscheidungen zu den herkömmlichen Workflows werden in (01)



Modellierungswerkzeug  
für Simulations-Workflows

dargestellt. Um die Trial-and-Error-Methode weiter zu unterstützen, wurden die Workflow-Technologie und die entsprechende Infrastruktur, also Modellierungswerkzeug, Ausführungsumgebung und Monitoring erweitert. Dieser Ansatz wird *Model-as-You-Go* genannt und hilft den Wissenschaftlern, einen unvollständigen Simulations-Workflow während der Ausführung zu vervollständigen, also weiter zu modellieren. Dies war in den existierenden Workflow-basierten Ansätzen nicht möglich, weil die Vollständigkeit des Workflow-Modells für eine Ausführung vorausgesetzt war. Der *Model-as-You-Go-Ansatz* gibt den Wissenschaftlern die Möglichkeit, die laufenden Simulationen zu verändern und dabei mit ihnen wie bei der Erstellung eines Simulationsmodells zu interagieren. Dies wurde durch die Integration des Modellierungswerkzeugs mit dem WfMS ermöglicht. Dabei werden die Änderungen im Modell an das WfMS und die Informationen über den Verlauf der Simulation an das Modellierungswerkzeug übertragen. Um den Verlauf der Simulationen zu beobachten, wurden die Monitoring-Mechanismen aus der herkömmlichen Workflow-Technologie übernommen und erweitert, um die oben genannte bidirektionale Interaktion zu ermöglichen [4]. Beispielsweise kann ein Wissenschaftler, der die Strukturänderung von Knochen simulieren möchte, einen oder mehrere Schritte im Simulationsworkflow zurückgehen, wenn er erkennt, dass die bisher angenommenen Parameter nicht zu dem gewünschten Ergebnis führen. Dabei

bleiben die Ergebnisse bisheriger Workflow-Schritte erhalten, was insbesondere bei lang laufenden Simulationen wünschenswert ist.

Weitere Herausforderungen beim Einsatz von Simulationsworkflows bestehen in der Verwaltung der Datenmengen, die eine Simulation benutzt oder produziert, und in der Unterstützung von Echtzeiteigenschaften sowie der Visualisierung der Ergebnisse der Simulationen. Die neuartigen Konzepte zur Unterstützung von Simulationsworkflows, die oben beschrieben wurden, lassen sich mit generischen Konzepten und Mechanismen, die diese Herausforderungen bewältigen, erweitern. Diese Aspekte werden in den nächsten Kapiteln im Detail diskutiert.

## 2.2. Unterstützung von verzahnten Simulationen

In Zukunft sollen die Konzepte und die Infrastruktur so erweitert werden, dass es möglich wird, auch komplexe Interaktionen zwischen einzelnen Simulationen zu modellieren und auszuführen. Ziel ist die Erstellung von Multiskalen-, Multi-Physik-Simulationen auf Basis von einzelnen Simulationsworkflows, die in unterschiedlichen naturwissenschaftlichen Bereichen bereits vorhanden sind. Simulationen, die sich zu einer komplexeren Simulation – wie z.B. der ganzheitlichen Simulation des menschlichen Körpers (also von der Molekularebene, Zellebene bis zur Skelett-Simulation) und Simulationen von Festkörpern (auf Atomebene, Molekularebene und Makroebene) – zusammensetzen lassen, sind verzahnt und tauschen ständig Daten in einer bestimmten Interaktionsreihenfolge aus. Die Durchführung solcher komplexen Simulationen und die Integration der oft verteilten Simulationen soll von einer *Cyber-Infrastruktur* übernommen werden. Durch dieses Zusammenspiel wird auch die Kooperation von Wissenschaftlern aus unterschiedlichen Organisationen und Bereichen wesentlich vereinfacht.

Die verzahnten Simulationen werden mit Hilfe von so genannten *Choreographien* modelliert. Choreographien definieren die korrekte Interaktionsreihenfolge und den Datenaustausch von Workflows und sind bereits ein erprobtes Konzept bei Geschäftsanwendungen. Die dort bekannten Modellierungskonzepte und Korrelations-

mechanismen sollen auch für die Modellierung von verzahnten Simulationen übernommen und angepasst werden, um die komplexe Kopplung und die Kontextabhängigkeiten zwischen den Simulationen zu ermöglichen. Ein Beispiel ist die Kopplung einer Festkörpersimulation, die das Verhalten von Kupferatomen im Eisenatomgitter im Lauf der Zeit simuliert, mit einer Molekulardynamiksimulation, die die Scherkräfte, d.h. in unterschiedliche Richtungen ziehende Kräfte auf ein Eisengitter simuliert. Die Ergebnisse dieser so genannten Kinetischen Monte-Carlo-Simulation sollen vollautomatisch in die Molekulardynamiksimulation fließen. Bisher sind diese Simulationen nicht automatisch gekoppelt und der Datentransfer erfolgt durch manuelles Kopieren. Eine Modellierung mit Hilfe von Choreographien soll hier eine automatisierte Verzahnung der beiden Simulationen ermöglichen. Dadurch kann die Durchführung umfangreicher und komplexer Parameterstudien deutlich beschleunigt und damit erst sinnvoll anwendbar gemacht werden.

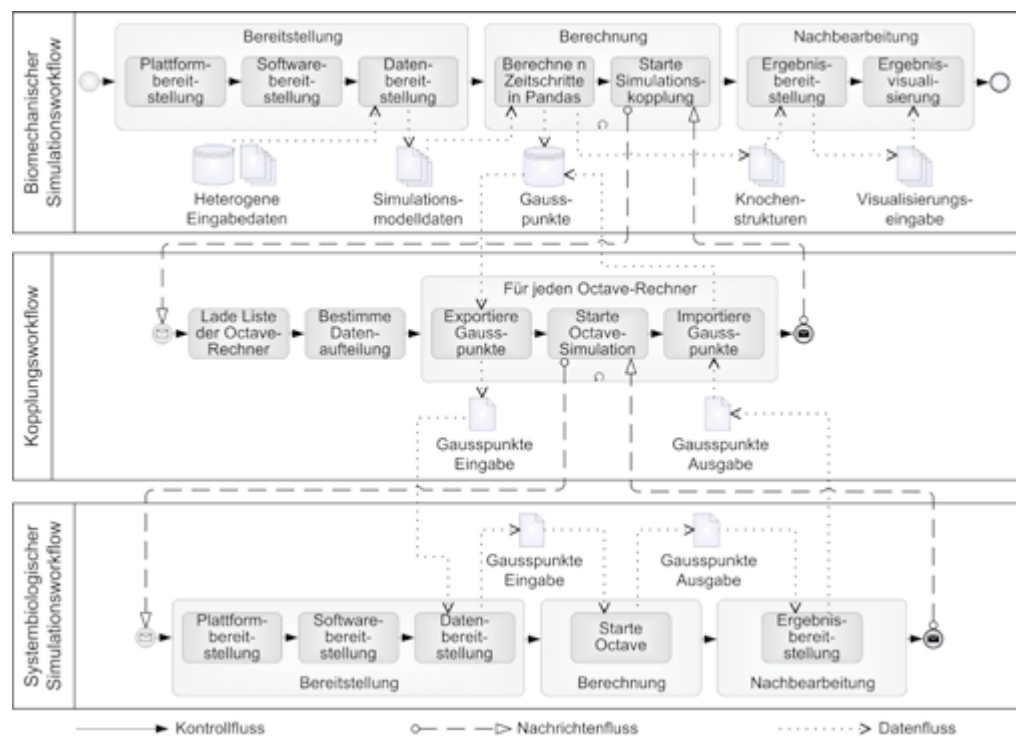
Um die Ausführung verzahnter Simulationen effizienter zu gestalten, soll vor allem das Potenzial existierender Cloud-Infrastrukturen genutzt werden. Cloud-Infrastrukturen stellen Zugriff auf Rechen- und Speicherressourcen nach dem Pay-Per-Use-Modell zur Verfügung. Es gehört zu den Zielen der aktuellen Forschung, das SimTech WfMS selbst in der Cloud bereitzustellen, um diese Ressourcenkapazität auszunutzen. Auch die verzahnten Simulationen sollen einmal in der Cloud ausgeführt werden können [5].

## 3. Der Umgang mit Daten für Simulationsworkflows

Simulationsworkflows müssen oftmals große Datenmengen verarbeiten, die in einer Vielzahl unterschiedlicher Formate vorliegen. Damit diese Daten von den im Workflow eingebundenen Programmen verarbeitet werden können, müssen sie entsprechend bereitgestellt und dabei in die passenden Formate transformiert werden. Der in (03) gezeigte Kopplungsworkflow muss z.B. Daten von einem Datenbank-basierten Format der biomechanischen Simulation in ein spezifisches Dateiformat der systembiologischen Simulation transformieren. Außerdem benötigt die systembiologische Simulation nur einen

Teil der Daten der biomechanischen Simulation, so dass dieser Teil der Daten entsprechend aus der Datenbank ausgewählt und extrahiert werden muss. Dies erhöht die Komplexität der Workflowmodellierung. Die Workflows werden dabei i.d.R. von den an den Simulationsergebnissen interessierten Wissenschaftlern selbst modelliert. Da Wissenschaftler aber selten vertiefte Kenntnisse im Bereich der Workflowmodellierung und der Datenbereitstellung besitzen, bedeutet dies, dass sie sich weniger auf ihre Kernaufgaben konzentrieren können, d.h. auf die Entwicklung von mathematischen Simulationsmodellen, die Durchführung der Simulationen und die Interpretation der Ergebnisse.

Zur Behebung dieses Problems wurde ein Rahmenwerk entwickelt (*SIMPL – SimTech, Information Management, Processes, and Languages*), mit dem die Aktivitäten zur Datenbereitstellung in Simulationsabläufen auf einfache Weise modelliert werden können [6, 7]. Wissenschaftler sollen keine Implementierungsdetails spezifizieren müssen, sondern lediglich die Kernaspekte der Datenbereitstellung in Form von Patterns (Muster) beschreiben, welche typische Datenbereitstellungsschritte in Simulationsworkflows darstellen. Jedes Pattern fasst mehrere feingranulare Workflow-Schritte zusammen, was bereits die Anzahl der für Wissenschaftler sichtbaren Schritte reduziert. Weiterhin müssen Wissenschaftler nur wenige abstrakte Parameterwerte für die Patterns spezifizieren. Wenn für eine Simulation beispielsweise Daten von einem Computer auf einen anderen transferiert werden müssen, so nutzen Wissenschaftler hierfür ein entsprechendes Datentransferpattern. Als Parametrisierung dieses Patterns werden lediglich der Speicherort der Eingabedatei sowie das Programm, für welches die Datei bereitgestellt werden soll, festgelegt. Um komplizierte Details, wie z.B. die oben beschriebene Transformation der Daten zwischen verschiedenen Formaten, müssen sich die



03

Wissenschaftler dabei nicht kümmern. Über diese Abstraktionsunterstützung hinaus wurden Maßnahmen zur Effizienzsteigerung der Datenverarbeitung in Simulationsworkflows untersucht. Dabei wurde insbesondere ein Ansatz entwickelt, welcher die Verarbeitung von Daten im lokalen Speicher eines Workflows optimiert [8].

Auf Basis des SIMPL-Rahmenwerks wurde das so genannte „Comprehensive Data Management Layer“ (CDML) entwickelt [9]. Das CDML wird speziell im Maschinenbau eingesetzt und löst die Heterogenität von Programmen und Datenquellen aus verschiedenen Projektphasen oder Organisationseinheiten eines Unternehmens auf. Es ermöglicht insbesondere, computerbasierte Simulationen mit echten physischen Tests, z.B. Crashtests von Autos, zu kombinieren und damit die Kosten von Produkttests zu reduzieren.

In Zukunft soll das beschriebene SIMPL-Rahmenwerk für die Datenbereitstellung in Simulationsworkflows zu einer vollständig datengetriebenen Simulationsinfrastruktur weiterentwickelt werden. Dabei wird insbesondere die Reproduzierbarkeit und Nachvollziehbarkeit von Simulationen und ihren Ergebnissen sowie die Bewertung von Simulationsmodellen unterstützt. Hierzu ist die umfassende

*Workflows für eine Simulation des Knochenwachstums, welche z.B. bei Osteoporose-Patienten eingesetzt wird. Die Simulation koppelt zwei mathematische Simulationsmodelle. Das biomechanische Modell beschreibt den Knochen auf einer makroskopischen Gewebeebene. Das systembiologische Modell arbeitet etwas genauer, indem es die Interaktion von Zellen innerhalb des Knochengewebes betrachtet. Jedes Modell wird durch einen eigenen Simulationsworkflow realisiert, während der zusätzliche Kopplungsworkflow den Datenaustausch zwischen diesen beiden Modellen durchführt.*

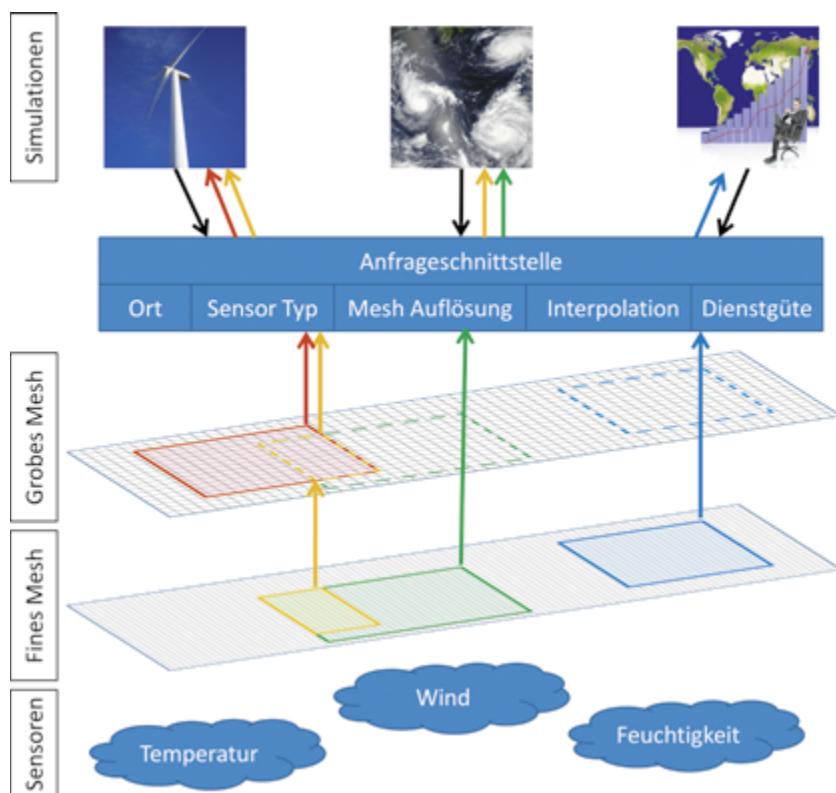
Sammlung und Verwaltung so genannter Provenance-Informationen erforderlich. Dies sind Metriken und Beschreibungsdaten zu den Simulationsabläufen und den Simulationsergebnissen. Diese Provenance-Informationen beschreiben u.a., wann welcher Workflow-Schritt ausgeführt wurde, welches Programm mit welchen Eingabedaten dabei verwendet wurde und wie lange diese Ausführung gedauert hat. Das Erfassen von Provenance-Informationen ist aber nur der erste Schritt. Hat man diese erst einmal erfasst, dann können sie die Grundlage für vielfältige Analysen bilden, wobei fortgeschrittene Analysemechanismen hierfür erst noch entwickelt werden müssen. Die Provenance-Analyse wird Wissenschaftlern maßgeblich helfen, ihre Simulationsergebnisse und Simulationsmodelle besser verstehen

lyse helfen, indem sie Zusammenhänge zwischen solchen Parameteränderungen und den Simulationsergebnissen untersucht und z.B. Empfehlungen gibt, inwieweit der CO<sub>2</sub>-Ausstoß verringert werden sollte. In Kombination mit dem bereits entwickelten SIMPL-Rahmenwerk wird diese Analyse von Provenance-Informationen die Produktivität der Wissenschaftler weiter deutlich erhöhen.

#### 4. Simulationen in Echtzeit – Das Global Sensor Grid

Echtzeitsimulationen werden immer wichtiger, um Entscheidungen anhand der Entwicklungen eines physischen Phänomens zu treffen. Beispielsweise ermöglichen Kenntnisse über die zukünftige Entwicklung eines Windfeldes zahlreiche Anwendungen, z.B. die Optimierung der Stromerzeugung aus Windkraft bis hin zur Analyse der Ausbreitung von Schadstoffen. Um die Qualität der Simulationsergebnisse zu verbessern, nutzen Simulationen zur Parametrisierung verstärkt Echtzeitdaten aus Sensoren. Dabei können Simulationen auf eine ständig wachsende Infrastruktur aus vielen weltweit verteilten Sensoren und Sensornetzen zurückgreifen. So wurden allein im Rahmen des Earthscope-Projektes [10] tausende von seismischen Sensoren primär an der Westküste der USA ausgebracht, um zukünftig besser auf Erdbeben reagieren zu können. Andere Projekte wie das CENSE-Projekt von HP [11] gehen gar von Billionen von Sensoren in absehbarer Zukunft aus, auf welche jeweils eine Vielzahl verschiedener Echtzeitsimulationen gleichzeitig zugreifen können.

Die enormen Datenvolumina und die Heterogenität, mit der Daten von Simulationen bereitgestellt werden müssen, stellen IT-Infrastrukturen vor enorme Herausforderungen hinsichtlich der skalierbaren Integration dieser Daten. Dazu wurde das Global Sensor Grid (GSG) entwickelt, welches ein breites Spektrum von Methoden bietet, um eine skalierbare Integration von Sensordaten in Simulationen sicherzustellen. Das GSG wird durch eine Menge global verteilter Rechner realisiert, die die Vermittlung von Sensordaten zwischen verschiedenen Sensornetzen und Konsumenten, also den angebundenen Simulationen, sicherstellen. Diese Rechner bilden ein Netzwerk von Vermittlungsknoten,



und validieren zu können. Weiterhin wird die Analyse Empfehlungen für Optimierungen von sowohl Simulationsworkflows als auch Simulationsmodellen geben. In der Klimaforschung möchten Wissenschaftler z.B. mithilfe von Simulationen die Auswirkungen von Änderungen des CO<sub>2</sub>-Ausstoßes auf das Weltklima erforschen. Hierbei kann die Provenance-Ana-

die an strategisch wichtigen Punkten sitzen, beispielsweise nahe an den jeweiligen Sensoren oder den Konsumenten sowie an zentralen Umschlagpunkten im Netzwerk, um bei der Übertragung und Vermittlung von Daten sowohl eine möglichst geringe Verzögerung als auch eine effiziente Nutzung der Bandbreite im vermittelnden Netz zu garantieren.

Die Anbindung eines Sensornetzes an das GSG erfolgt durch die Bereitstellung eines oder mehrerer als Gateway bezeichneter Vermittlungsknoten. Zu diesen werden sensornetzspezifische Daten übertragen und vorverarbeitet, so dass diese den Simulationen in einem globalen Koordinatensystem zugänglich gemacht werden können. Konsumenten können an beliebige Vermittlungsknoten angebunden werden und über eine einfache Anfrageschnittstelle (siehe (04)) definieren, welche Sensordaten gefunden und zugänglich gemacht werden sollen. Dies beinhaltet insbesondere die Art der Sensordaten, die geografische Ausdehnung sowie die Häufigkeit mit der Sensordaten gemessen und gesendet werden sollen. Beispielsweise kann eine Windfeld-Simulation sämtliche verfügbaren Winddaten in der Umgebung von Stuttgart mit einer Aktualisierungsrate von zehn Minuten fordern. Die Vermittlungsknoten ermitteln dann kooperativ Gateways und optimale Pfade im Netz zwischen Gateways und Konsumenten. Zur Minimierung der Bandbreite werden insbesondere Überlappungen in den Datenströmen zwischen Gateways und Konsumenten genutzt [10], um gemeinsame Vermittlungsstrecken im Vermittlungsnetz zu nutzen. Sollte also eine weitere Simulation sich für Windfelddaten interessieren, werden diese möglichst gemeinsam im Netz übertragen und somit redundante Übertragungen vermieden. Ferner ermöglicht der Einsatz von Prädiktionstechniken [13] eine signifikante Reduktion der Verzögerungen und der genutzten Bandbreite. Falls also keine Änderung der gemessenen Winddaten erfolgt oder sich das Windfeld gemäß des Modells eines Prädiktors verändert, müssen diese nicht über das Netz übertragen werden, sondern stehen der Simulation direkt durch das Prädiktormodell zur Verfügung. Die für das GSG entwickelten Methoden zur autonomen Anpassung ermöglichen eine hohe Anpassungsfähigkeit an geänderte Anfragen der Konsumenten. Gerade bei Anfra-

gen, die sich auf ein bewegliches physisches Phänomen konzentrieren, kann somit der Anfragebereich dynamisch und effizient angepasst werden. Gleichzeitig können neue Vermittlungsknoten und Sensornetze einfach und effizient integriert werden. Mit dem rasanten Fortschritt bei der Verbreitung mobiler Geräte wird die Fragestellung der Integration dieser Geräte im Kontext von Echtzeitsimulationen immer bedeutender. Die Forschung an dem GSG richtet sich daher in Zukunft insbesondere auf zwei wichtige Anwendungsfälle. Einerseits sollen Echtzeitsimulationen verstärkt von der Nutzung der auf mobilen Geräten befindlichen Sensoren profitieren können. Dies erfordert insbesondere Konzepte zur Integration der mit Unsicherheiten behafteten Daten in eine Echtzeitsimulation. Andererseits soll eine effiziente Integration der Sensordaten in Simulationen erfolgen, so dass sogar eine Ausführung von Echtzeitsimulationen auf eher ressourcenschwachen mobilen Geräten erfolgen kann. Zur Unterstützung der Ausführung von Simulationen soll dazu die Anbindung an moderne IT-Infrastrukturen untersucht werden, auf denen die Berechnungen der Simulationen teilweise aus der Ferne ausgeführt werden können. Weiterhin soll durch eine Reduktion der Modelle einer Simulation [14] der Ressourcenbedarf einer Echtzeitsimulation reduziert werden, so dass letztlich sogar eine vollständige Ausführung auf einem ressourcenschwachen Gerät erfolgen kann. Da die Reduktion der Simulationsmodelle typischerweise auch eine Abnahme der Qualität der Simulation erzeugt, werden im Rahmen des GSG Methoden untersucht, um den erforderlichen Ressourcenbedarf und den Energieverbrauch von Echtzeitsimulationen zu minimieren und gleichzeitig eine für den Nutzer genügende Qualität der Echtzeitsimulation bei reduzierten Modellen und unsicheren Daten zu garantieren.

## 5. Visualisierung für die Simulationstechnik

Die Visualisierung ist ein wichtiges Werkzeug bei der Analyse von Simulationsergebnissen. Häufig können die komplexen Daten, die bei Simulationen entstehen, nur durch eine geeignete visuelle Darstellung verstanden werden. Zum Beispiel kann die Bewegung von Teilchen in Luft-



strömungen viel leichter untersucht werden, wenn Bahnlinien visuell dargestellt werden. Daher spielt die Visualisierung im heutigen Arbeitsablauf der Simulationstechnik eine entscheidende Rolle und ist als Analysewerkzeug nicht mehr wegzudenken. Gerade bei komplexen Simulationen, die durch Simulationsworkflows gestaltet werden, ist die Visualisierung unverzichtbar.

Neben der Entwicklung vollständig neuer Visualisierungsmethoden spielt die Beschleunigung bestehender Verfahren eine wichtige Rolle, da die Visualisierung häufig interaktiv eingesetzt wird und deshalb schnell genug auf Benutzereingaben reagieren muss. Ein typisches Beispiel hierfür ist die Visualisierung von Strömungssimulationen: Hierfür müssen typischerweise sehr viele Bahnlinien oder ähnliche Kurven berechnet und dargestellt werden, was bei großen Datensätzen leicht mehrere Minuten in Anspruch nehmen kann. Ein neuartiges Berechnungsverfahren [15] erlaubt seit kurzem – zusammen mit der Ausnutzung der Leistungsfähigkeit moderner Grafik-Hardware (engl. Graphics Processing Units: GPUs) – die hoch-effiziente und parallele Berechnung dieser Kurven.

Wenn die Berechnungen, die der Visualisierung zugrunde liegen, schnell genug für die Benutzerinteraktion sind, bietet sich die interaktive Erkundung der Simulationsdaten an. Beispielsweise können Parameter für eine optimale

Darstellung angepasst werden; oder der Benutzer vergrößert bestimmte Bereiche oder lässt sich zusätzliche Informationen an relevanten Punkten der Simulation anzeigen, z.B. die Bahnkurven von dort startenden Partikeln. Bei den komplexen Daten aus heutigen Simulationsrechnungen ergeben sich jedoch sehr feine Strukturen in der visuellen Darstellung. Deshalb ist es wichtig, eine exakte Navigation und Positionierung durch den Benutzer zu gewährleisten.

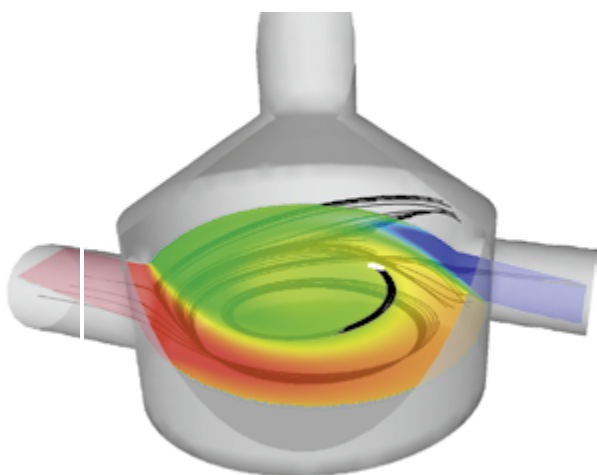
Wir haben eine entsprechende Interaktionstechnik entwickelt, die die Positionierungsgenauigkeit bei Strömungsvisualisierungen automatisch an die Daten

anpasst und so die Analyse selbst von feinsten Strukturen erlaubt [16]. So kann der Benutzer auch in komplexen 3D-Strömungsdaten mit seiner gewohnten Mausegeschwindigkeit die Bahnlinien sehr genau positionieren (siehe (05)).

Eine weitere Herausforderung ist die Visualisierung zeitabhängiger Daten. Die einzelnen Zeitschritte einer Simulation können nacheinander visualisiert und wie in einem Film abgespielt werden. Mit diesem naheliegenden Ansatz ist er aber schwierig, verschiedene Resultate zu vergleichen oder die zeitliche Entwicklung und die Zusammenhänge richtig abzuschätzen. Diese Beschränkung ergibt sich aus den Grenzen der menschlichen Kognition: Es ist für den Benutzer sehr schwierig, komplexe Sachverhalte in Animationen zu erkennen und sich diese zu merken. Eine gute Visualisierung ist deshalb an die kognitiven Fähigkeiten der Nutzer angepasst. Beispielsweise haben wir eine neue Methode für die Strömungsvisualisierung entwickelt, welche die zeitabhängigen Daten in einem einzelnen Bild zusammenfasst und dabei ohne Animation auskommt [17]. So können etwa verschiedene Simulationsdurchläufe sehr einfach miteinander verglichen werden. Eine Anwendung hierfür ist z.B. die Simulation von verunreinigtem Grundwasser. Häufig ist hierbei die genaue Beschaffenheit des Untergrunds nicht bekannt, weswegen viele Simulationsdurchläufe mit veränderten Parametern durchgeführt werden.

Aber auch vermeintlich simple Visualisierungsmethoden wie Balkendiagramme können noch verbessert werden. Ein Problem von Balkendiagrammen ist die Darstellung von Daten, die einen großen Wertebereich abdecken, wie z.B. die Dauer des radioaktiven Zerfalls verschiedener chemischer Elemente. Eine direkte Darstellung solcher Daten würde darin resultieren, dass nur die großen Werte sichtbar wären; die Balken der kleinen Werte wären einfach zu klein. Unsere Methode nutzt nun mehrere Skalen, um solche Daten darzustellen [18]. Dabei bleiben auch sehr kleine Werte sichtbar, und alle Daten können visuell miteinander verglichen werden.

Die zukünftige Forschung wird sich auf die visuelle Analyse sehr großer und komplexer Daten konzentrieren, die mit bisherigen Visualisierungstechniken oder anderen Analyseverfahren nicht gehandhabt



Vom Benutzer platzierte Bahnlinien in einem Mischer, in dem kalte (blau) und heiße Luft (rot) vermischt wird. Durch eine präzise Platzierung der Linien kann das Mischverhalten sehr genau analysiert werden.

werden können. Eine umfassende Untersuchung solcher Daten durch eine direkte Visualisierung ist sehr zeitaufwändig, und es besteht die Gefahr, dass wichtige Details übersehen werden. Dies kann z.B. der Fall sein, wenn der Wissenschaftler viele hunderte oder gar tausende Simulationen durchläufe miteinander vergleichen möchte. Daher sollen die Nutzer der Visualisierung durch Verfahren aus dem Bereich des Data-Mining und des maschinellen Lernens unterstützt werden. So kann der Computer beispielweise die Datenmenge vorab auf relevante Informationen einschränken oder den Anwender auf potentiell interessante Bereiche hinweisen. Diese Art der Datenreduktion beeinflusst auch die Entwicklung neuer Visualisierungsmethoden, indem diese so gestaltet werden, dass sie nur die relevanten Strukturen und Informationen darstellen, z.B. nur die Grenzschichten zwischen verschiedenen Bereichen einer Strömung. Zudem sollen die Visualisierungsmethoden hocheffizient sein, um eine schnelle Darstellung ständig wachsender Datenmengen zu gewährleisten. Hierfür sollen Ansätze aus dem Höchstleistungsrechnen verfolgt werden.

Die umfassende Analyse komplexer Daten erfolgt typischerweise in Teamarbeit. Daher soll die Visualisierung die kollaborative Arbeit mehrerer Experten unterstützen. Gemeinsames Arbeiten kann beispielsweise an großen Display-Wänden oder aber auch an verteilten Arbeitsplätzen erfolgen. Dabei ist es unter anderem wichtig, dass die Anwender über den gegenseitigen Fortschritt informiert sind und sich über interessante Datenbereiche und Informationen problemlos austauschen können. Die visuellen Schnittstellen und Interaktionstechniken müssen gerade auch für Nicht-Informatik-Experten zugänglich gemacht werden – was beim Umfang der Simulationsdaten und der Komplexität der Analyseaufgaben eine Herausforderung für die Visualisierungsforschung ist.

## Schluss

Im Rahmen des Exzellenz Clusters „Simulation Technology“ (SimTech) wird der Fokus auf die Entwicklung einer ganzheitlichen Systemwissenschaft gelegt, die den Wissenschaftlern bei der Erforschung und Simulation von Naturphänomenen hilft. In diesem Beitrag haben wir unsere For-

## ZUSAMMENFASSUNG

Dieser Beitrag gibt einen Überblick der bisherigen Forschungsergebnisse auf dem Gebiet integrierte Datenverwaltung, Workflows und Visualisierung für integrative Systemwissenschaft und unsere Zukunftspläne und Forschungsschwerpunkte auf dem Weg zu einer Cyber-Infrastruktur für die Simulationstechnik. Unter anderem wird ein Workflow Management System vorgestellt, welches speziell auf die Anforderungen und Bedürfnisse der Wissenschaftler zugeschnitten ist und aus unserer Forschung auf dem Gebiet Simulationstechnik entstanden ist. Des Weiteren wurde ein Rahmenwerk entworfen und umgesetzt, mit dem Aktivitäten zur Datenbereitstellung in Simulationsabläufen benutzerfreundlich modelliert und ausgeführt werden können. Für die skalierbare Integration von Sensordaten in Echtzeitsimulationen wurde das Global Sensor Grid entwickelt. Für die Analyse von Simulationsergebnissen wurden sowohl komplett neue Visualisierungstechniken erarbeitet als auch bestehende Techniken beschleunigt. Auf Basis unserer Forschungsergebnisse soll in Zukunft eine Simulations-Cyber-Infrastruktur zur Unterstützung von verteilten und gekoppelten Simulationen entstehen. Dabei wird die Reproduzierbarkeit und Nachvollziehbarkeit von Simulationen und ihren Ergebnissen, im Allgemeinen bezeichnet als Provenance, eine bedeutende Rolle spielen. Für Echtzeitsimulationen steht die Integration von mobilen Geräten im Vordergrund. Bei der Visualisierung wird der Fokus auf der Analyse sehr großer und komplexer Daten liegen, die bisher nicht gehandhabt werden können.

schungsergebnisse kurz zusammengefasst und unsere zukünftigen Forschungsziele auf dem Weg zu einer Cyber-Infrastruktur für die Simulationstechnologie umrissen. •

Dimka Karastoyanova, Karolina Vukojevic-Haupt,  
Andreas Weiß, Frank Leymann,  
Peter Reimann, Bernhard Mitschang,  
Holger Schwarz, Boris Koldehofe, Kurt Rothermel,  
Marcel Hlawatsch, Daniel Weiskopf

## Literatur

1. Görlach, Katharina; Sonntag, Mirko; Karastoyanova, Dimka; Leymann, Frank; Reiter, Michael: *Conventional Workflow Technology for Scientific Simulation*. In: Yang, Xiaoyu (Hrsg.); Wang, Lizhe (Hrsg.); Jie, Wei (Hrsg.): *Guide to e-Science*, Springer-Verlag, 2011
2. Karastoyanova, Dimka; Dentsas, Dimitrios; Schumm, David; Sonntag, Mirko; Sun, Lina; Vukojevic, Karolina: *Service-based Integration of Human Users in Workflow-driven Scientific Experiments*. In: *Proceedings of the 8th IEEE International Conference on eScience (eScience 2012)*
3. Sonntag, Mirko; Hotta, Sven; Karastoyanova, Dimka; Molnar, David; Schmauder, Siegfried: *Using Services and Service Compositions to Enable the Distributed Execution of Legacy Simulation Applications*. In: Abramowicz, W. (Hrsg.); Llorente, I.M. (Hrsg.); Surridge, M. (Hrsg.); Zisman, A. (Hrsg.); Vayssière, J. (Hrsg.): *Towards a Service-Based Internet, Proceedings of the 4th European Conference ServiceWave 2011, Poznan, Poland, 2011*

4. Sonntag, Mirko; Hahn, Michael; Karastoyanova, Dimka: Mayflower – Explorative Modeling of Scientific Workflows with BPEL. In: *Proceedings of the Demo Track of the 10th International Conference on Business Process Management (BPM 2012)*, CEUR Workshop Proceedings, 2012
5. Vukojevic-Haupt, Karolina; Karastoyanova, Dimka; Leymann, Frank: On-demand Provisioning of Infrastructure, Middleware and Services for Simulation Workflows. In *Proceedings of SOCA 2013*.
6. P. Reimann, H. Schwarz: Datenmanagementpatterns in Simulationsworkflows. In: *Datenbanksysteme für Business, Technologie und Web*, Magdeburg, Germany, 2013.
7. P. Reimann, M. Reiter, H. Schwarz, D. Karastoyanova, F. Leymann: SIMPL – A Framework for Accessing External Data in Simulation Workflows. In: *Datenbanksysteme für Business, Technologie und Web*, Kaiserslautern, Germany, 2011
8. P. Reimann, H. Schwarz, B. Mitschang: Design, Implementation, and Evaluation of a Tight Integration of Database and Workflow Engines. *Journal of Information and Data Management* 2(3), SBC – Brazilian Computer Society, 2011.
9. P. Janowski, B. Mitschang, A. Gollmann: Issues and Characteristics of Testing as Part of the Design Process in Mechanical Engineering. In: *Tagungsband der 15. International Conference on Computer Supported Cooperative Work in Design*, Lausanne, Switzerland, 2011.
10. M. Williams, K. Fischer, J. Freymueller, B. Tikoff, A. Tréhu et al., Unlocking the secrets of the north american continent: An earthscope science plan for 2010–2020, EarthScope, Tech. Rep., February 2010.
11. P. Hartwell, Cense: A central nervous system for the earth, in *Technologies Beyond 2020 (TTM)*, 2011 IEEE Technology Time Machine Symposium on, 2011, pp. 1–1.
12. A. Benzing, B. Koldehofe, and K. Rothermel. Efficient Support for Multi-Resolution Queries in Global Sensor Networks. In *Proc. of the 5th International Conference on COMMunication System softWAre and middleware (COMSWARE 2011)*, pp.1–12, ACM.
13. A. Benzing, B. Koldehofe, and K. Rothermel. Multilevel Predictions for the Aggregation of Data in Global Sensor Networks. In *Proc. of the 14th International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2010)*, pp. 169–178, IEEE.
14. B. Haasdonk, M. Dihlmann, M. Ohlberger: A Training Set and Multiple Bases Generation. Approach for Parametrized Model Reduction Based on Adaptive Grids in Parameter Space. *Math. Comp. Model. Dyn. Vol.17(4)*: 423–442, 2011, Taylor & Francis.
15. M. Hlawatsch, F. Sadlo, D. Weiskopf: Hierarchical Line Integration. *IEEE Transactions on Visualization and Computer Graphics* 17(8):1148–1163, 2011
16. M. Hlawatsch, F. Sadlo, D. Weiskopf: Predictability-Based Adaptive Mouse Interaction and Zooming for Visual Flow Exploration. *International Journal for Uncertainty Quantification* 3(3):225–240, 2013
17. M. Hlawatsch, P. Leube, W. Nowak, D. Weiskopf: Flow Radar Glyphs – Static Visualization of Unsteady Flow with Uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 17(12):1949–1958, 2011
18. M. Hlawatsch, F. Sadlo, M. Burch, D. Weiskopf: Scale-Stack Bar Charts. *Computer Graphics Forum* 32(3):181–190, 2013

## DIE AUTOREN

**Kontakt**

IAAS – Institut für Architektur von Anwendungssystemen  
 Universität Stuttgart  
 Universitätsstraße 38  
 D–70569 Stuttgart  
 Jun.-Prof. Dr. Dimka Karastoyanova  
 Tel. +49 (0) 711/685-88476  
 E-Mail: dimka.karastoyanova@iaas.stuttgart.de  
 Web: <http://www.iaas.uni-stuttgart.de/>

## ■ ABBILDUNG 1

**JUN.-PROF. DR. DIMKA KARASTOYANOVA** (v.r.) ist Juniorprofessorin für Simulationsworkflows am IAAS und im Rahmen des Exzellenzclusters SimTech. Ihre Forschung umfasst Themen wie flexible Simulationsworkflows und verzahnte Workflows für Simulationen und Geschäftsanwendungen.

**PROF. DR. FRANK LEYMANN** (h.l.) ist geschäftsführender Direktor des Instituts für Architektur von Anwendungssystemen (IAAS) an der Universität Stuttgart. Er forscht auf den Gebieten Cloud Computing, Service-orientierte Architekturen, Anwendungsarchitekturen und Integration für Geschäftsprozesse und Simulationstechnik.

**KAROLINA VUKOJEVIC-HAUPT**

(v.l.) ist Doktorandin am IAAS und im Rahmen von SimTech. Ihre Forschung konzentriert sich auf die Konzipierung und Entwicklung von Ansätzen zur Bereitstellung von Simulations-Workflow-Management-Systemen in der Cloud.

**ANDREAS WEISS** (h.r.) ist Doktorand am

IAAS und im Rahmen von SimTech. In seiner Forschung beschäftigt er sich damit Konzepte und Systeme zur Unterstützung von flexiblen, verzahnten Simulationsworkflows für Cloud-basierte, Multi-Skalen- und Multi-Physik-Simulationen zu entwickeln.

## DIE AUTOREN

## ■ ABBILDUNG 2

**DR. BORIS KOLDEHOFE** (l.) ist Forscher und Dozent an der Universität Stuttgart. Seine Forschungsinteressen umfassen skalierbare Kommunikationssysteme, insbesondere Internet-basierte Sensornetze, Publish/subscribe Systeme und Ereignisverarbeitungssysteme.

**ANDREAS BENZING** (m.) ist Doktorand an der Graduiertenschule in SimTech. In seinem Dissertationsvorhaben erforscht er Stromverarbeitungssysteme zur Unterstützung von wissenschaftlichen Simulationen.

**PROF. DR. KURT ROTHERMEL** (r.) leitet den Lehrstuhl Verteilte Systeme an der Universität Stuttgart. Seine Forschungsinteressen liegen in den Bereichen Verteilte Systeme, Rechnernetze, Mobile Systeme und Sensornetze. In SimTech leitet er die Forschung am GlobalSensorGrid Projekt.

## ■ ABBILDUNG 3

**PROF. DR. BERNHARD MITTSCHANG** (r.) leitet die Abteilung Anwendersoftware am IPVS.

**PETER REIMANN** (m.) ist Doktorand am IPVS und an der Graduierten Schule SimTech.

**PD DR. RER. NAT. HOLGER SCHWARZ** (l.) ist Dozent und Forscher in der Abteilung Anwendersoftware am IPVS.

Diese Autoren beschäftigen sich im Rahmen des Exzellenzclusters Simulation Technology primär mit der Datenbereitstellung und Datenverwaltung in Simulationen. Hauptsächliches Augenmerk liegt dabei darauf, die Komplexität der für Wissenschaftler sichtbaren Datenverwaltungsaufgaben soweit zu reduzieren, dass sich Wissenschaftler verstärkt auf ihre Kernproblematik konzentrieren können, nämlich auf die eigentliche Simulation.

## ■ ABBILDUNG 4

**PROF. DR. DANIEL WEISKOPF** (r.) ist am Institut für Visualisierung und Interaktive Systeme (VIS) und am Visualisierungsinstitut der Universität Stuttgart (VISUS) tätig. Das Forschungsgebiet von Prof. Daniel Weiskopf umfasst die Visualisierung, visuelle Analytik und Computergrafik. Im Rahmen von SimTech konzentriert er sich auf die Entwicklung und Untersuchung von Methoden für die wissenschaftliche Visualisierung innerhalb von Simulationsworkflows.

**MARCEL HLAWATSCH** (l.) ist Doktorand am Visualisierungsinstitut der Universität Stuttgart (VISUS) und erforscht im Rahmen seiner Arbeit für SimTech neue Methoden für zeitabhängige Graphen und Workflows, sowie für Skalar-, Vektor- und Tensorfelder. Die von ihm entwickelten Methoden können u.a. für die Strömungsvisualisierung eingesetzt werden.



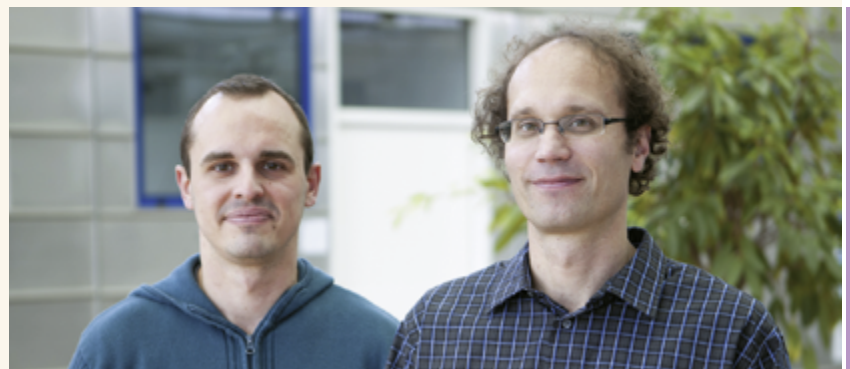
1



2



3



4