

Contents

1	Introduction	1
1.1	The Fourth Research Paradigm and Scientific Workflows . . .	1
1.2	Business Workflow Technologies for Scientific Workflows . .	3
1.3	Research Questions and Approach	5
1.4	Contribution	7
1.5	Structure of the Thesis	9
I	Foundations	13
2	Business Workflows	15
2.1	Introduction and Basic Terminology	15
2.2	Business Process Model and Notation (BPMN)	20
2.3	Web Services Business Process Execution Language (BPEL) .	25
2.4	BPMN to BPEL Mapping Strategies	29
3	Scientific Workflows	37
3.1	Introduction and Basic Terminology	37
3.2	Business Workflows vs. Scientific Workflows	39
4	Model-driven Software Development (MDSD)	43
4.1	Introduction and Basic Terminology	43
4.2	Model Transformations	46
4.3	Domain-specific Languages (DSLs)	48
4.4	Xtend	49
4.5	Xtext	53
5	Grid Computing	57
5.1	Introduction and Basic Terminology	57
5.2	Grid Middleware	60

5.2.1	Globus Toolkit 4 (GT4)	61
5.2.2	UNICORE 6	63
5.3	Grid Security Infrastructure (GSI)	64
5.4	BIS-Grid Workflow Engine	66
 II Model-Driven Scientific Workflow Engineering with MoDFlow		71
6	Introduction to MoDFlow	73
6.1	Basic Assumptions	73
6.2	Concept and Components of MoDFlow	74
7	Scientific Workflow Requirements on MoDFlow	81
7.1	Common Requirements on Scientific Workflows and SWfMS	81
7.2	Definition of Requirements on MoDFlow	85
8	Scientific Workflow Model Representation with MoDFlow.BPMN	91
8.1	Basic Design Considerations	91
8.2	Representation of Workflow Activities	95
8.3	BPMN Metamodel Subset	95
8.4	BPMN Metamodel Extensions	108
9	Scientific Workflow Model Mapping with MoDFlow.BPMN2BPEL	121
9.1	Basic Design Considerations	121
9.2	IWM2EWM Mapping Extensions	124
9.3	Step 1: BPMN Process Expansion	125
9.4	Step 2: BPMN Mapping	129
9.5	Step 3: Workflow Engine Adaptation	136
10	Utilization and Extension of MoDFlow	139
10.1	Realization of a DWM and a DWM2IWM mapping	139
10.1.1	Creation of DSLs	140
10.1.2	Adoption of Existing Scientific Workflow Languages .	142
10.1.3	Mapping of Data Flow-centric to Control Flow-centric Workflow Languages	143

10.2	Extension of MoDFlow.BPMN and MoDFlow.BPMN2BPEL	146
10.2.1	General Extension of MoDFlow.BPMN	146
10.2.2	General Extension of MoDFlow.BPMN2BPEL	146
10.2.3	Definition of Workflow Activities	147
10.2.4	Integration of Third Party Software	147
10.2.5	Adoption of other BPEL Process Engines	148
10.2.6	Adoption of other Executable Workflow Languages	148
11	Requirements on the MoDFlow Framework	149
11.1	Classification of BPMN-to-BPEL mapping	150
11.1.1	Classification Scheme for Model Transformation Problems	150
11.1.2	Classification Scheme for Model Transformation Languages	154
11.2	Requirements for a Transformation Framework	158
11.3	Requirements for a BPMN-to-BPEL Transformation Chain	159
12	Design of the MoDFlow Framework	161
12.1	Transformation Framework	161
12.2	BPMN-to-BPEL Transformation Chain	164
III	Evaluation	169
13	Implementation of the MoDFlow Framework	171
13.1	Transformation Framework	171
13.1.1	Implementation Decisions	172
13.1.2	Implementation	176
13.2	BPMN-to-BPEL Transformation Chain	185
13.2.1	Implementation Decisions	186
13.2.2	Implementation	193
14	Application Scenarios	207
14.1	Validation in Software Engineering	207
14.2	Preparations for Scientific Workflow Execution in Grid Environments	208

Contents

14.2.1	Workflow Interface for Scientific Workflows	208
14.2.2	Support for BIS-Grid Workflow Engine	210
14.2.3	Support of Globus Toolkit 4 Delegation Service	210
14.2.4	Workflow Activity for GT4 Job Submissions	213
14.2.5	Scientific Workflow Transformation Chain	215
14.3	Scenario I: Optimization of 3D-Images	216
14.4	Scenario II: Fishery Simulation	219
14.5	Scenario III: Publication Workflows in PubFlow	222
14.6	Discussion and Threats to Validity	233
15	Related Work	235
15.1	Utilization of Business Workflow Technologies for Grid and Scientific Workflows	235
15.2	Mapping of BPMN to BPEL	238
15.3	Transformation Chaining	243
IV	Conclusion and Future Work	247
16	Summary and Conclusion	249
17	Future Work	255
V	Appendix	261
A	WSDL Definition for Scientific Workflows	263
B	BPMN Workflow for Application Szenario I	267
C	BPMN Workflow for Application Szenario II	273
D	Xtext grammar PubFlow.DSL for Application Szenario III	277
E	PubFlow.DSL Workflow for Application Szenario III	281
	Bibliography	285