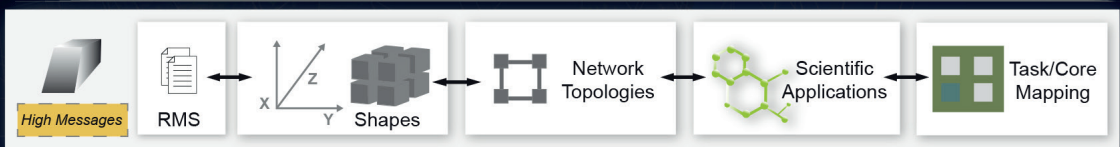
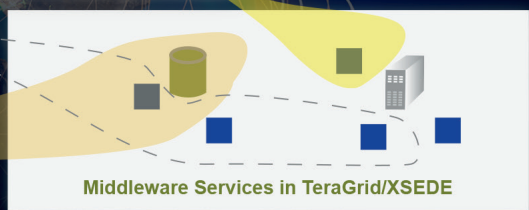


```

<jSDL:JobSpecification>
  <jSDL:JobIdentification> ...
  ...
  <jSDL:Resources>
    ...
    <glue2:NetworkInfo>
      ...
      <glue2:NetworkInfo>
        ...
        <Shape>
          <DimX> 1 </glue2dw:DimX>
          <DimY> 1 </glue2dw:DimY>
          <DimZ> 2 </glue2dw:DimZ>
        ...
      </glue2:NetworkInfo>
    ...
  </jSDL:Resources>
</jSDL:JobSpecification>
  <jSDL:DataStaging>
    ...
  </jSDL:DataStaging>
</jSDL:JobSpecification>

```



Design and Applications of an Interoperability Reference Model for Production e-Science Infrastructures

Morris Riedel

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)

Design and Applications of an Interoperability Reference Model for Production e-Science Infrastructures

Morris Riedel

Schriften des Forschungszentrums Jülich

IAS Series

Volume 16

ISSN 1868-8489

ISBN 978-3-89336-861-7

Contents

1	Introduction	1
1.1	Terminologies	3
1.2	Thesis Objectives and Contributions	4
1.3	Related Topics Out of Thesis Scope	6
1.4	Selected Publications and Demonstrations	8
1.5	Thesis Structure	10
2	State-of-the-art e-Science Infrastructures	13
2.1	Grid and e-Science Infrastructure Concepts	14
2.1.1	e-Science Infrastructure Fundamentals	14
2.1.2	Resource Sharing in e-Science Infrastructures	16
2.1.3	Classification of e-Science Applications	19
2.2	Key Technologies and Standards for e-Science Infrastructures	23
2.2.1	Resource Management Systems and Grid Middleware	23
2.2.2	Service-Oriented Technologies	24
2.2.3	Common Open Standards in the e-Science Domain	25
2.3	e-Science Infrastructure Interoperability Challenges	28
2.3.1	Classification of e-Science Infrastructure Types	28
2.3.2	Tightly Coupled Middleware Clusters in e-Science Infrastructures	30
2.3.3	Benefits of a Network of Interoperable Services	32
2.4	Conclusion	35
3	Related Work	37
3.1	Identification of Relevant Approaches and Factors	38
3.1.1	Reference Model Foundations and Factors	38
3.1.2	Open Grid Services Architecture Analysis	42
3.1.3	Component-based Approach Review	49
3.2	Survey of Related Reference Models	53
3.2.1	Enterprise Grid Alliance Reference Model	54
3.2.2	OASIS Service Component Architecture	56
3.2.3	Reference Model for Open Distributed Processing	58
3.2.4	Common Component Architecture	59
3.2.5	Coloured Petri Nets Reference Model	61
3.3	Classification of Component-based Approaches	64
3.3.1	Additional Layer Concepts	65
3.3.2	Neutral Bridge Concept	68
3.3.3	Gateway Approach	70
3.3.4	Mediator Approach	71
3.3.5	Adapter Approach	72

3.3.6	Middleware Co-existence	73
3.4	Conclusion	75
4	Requirements	77
4.1	Reference Model and Associated Elements Requirements	78
4.1.1	Reference Model Blueprint and Entity Requirements	78
4.1.2	Reference Model Entity Relationships	81
4.1.3	General Technical Reference Architecture Requirements	84
4.2	Functional Requirements	88
4.2.1	Reference Architecture Core Building Blocks	88
4.2.2	Improved e-Science Applications Executions	92
4.2.3	Improved Processing and Data-staging Capabilities	94
4.3	Non-functional Requirements	97
4.3.1	e-Science Production Infrastructure Integration Constraints	97
4.3.2	Requirements for Interoperable Infrastructure Usage Model	100
4.3.3	Process Requirements for Sustained Infrastructure Interoperability	102
4.4	Conclusion	105
5	Architectural Design	107
5.1	Reference Model Design and Associated Architecture Work	108
5.1.1	The Infrastructure Interoperability Reference Model Design	108
5.1.2	Associated Reference Architecture General Design	112
5.1.3	Detailed Associated Reference Architecture Core Building Blocks	116
5.1.4	Associated Reference Architecture Infrastructure Integration Constraints	122
5.1.5	Overall Run-time Pattern for Associated Architecture Work	128
5.1.6	Security Pattern for Associated Architecture Work	132
5.2	Design Layout and Essential Functionality	139
5.2.1	e-Science Application Concepts	139
5.2.2	Application Execution Adjacencies Concepts	144
5.2.3	High Performance Computing Extensions	148
5.2.4	Sequence Support for Computational Jobs	154
5.2.5	Manual Data-staging Concepts	157
5.2.6	Enhanced Accounting and Data Management Concepts	162
5.3	Seven Segment-based Process for Infrastructure Interoperability	167
5.3.1	Segment 1: Open Standards-based Reference Model and Architecture	169
5.3.2	Segment 2: Collaboration with the Right Set of Technology Providers	171
5.3.3	Segment 3: Reference Architecture Implementations	173
5.3.4	Segment 4: Standardisation Feedback Ecosystem	174
5.3.5	Segment 5: Aligned Future Strategies and Roadmaps	176
5.3.6	Segment 6: Harmonised Operation Policies	178
5.3.7	Segment 7: Funding Sources and Cross-Project Coordination	179
5.4	Conclusion	182
6	Impact and e-Science Applications	183
6.1	Seven Segment-based Process Implementation and Impact	184
6.1.1	Segment 1: IIRM and Standards-based Reference Architecture	185
6.1.2	Segment 2: Collaboration of Infrastructures with Technology Providers	188
6.1.3	Segment 3: IIRM Reference Architecture Implementations	189
6.1.4	Segment 4: The GIN and PGI Standardization Feedback Ecosystem	190

6.1.5	Segment 5: Aligning Middleware Roadmaps with EMI and XSEDE	193
6.1.6	Segment 6: Harmonized Security Setups and Operation Policies	195
6.1.7	Segment 7: Funding and Cross-Project Collaborations	196
6.2	Concrete Architectures of Production e-Science Infrastructures	198
6.2.1	Reference Model and Architecture Adoptions	199
6.2.2	European e-Science Infrastructures Setup	201
6.2.3	US and other e-Science Infrastructures Setups	207
6.2.4	Related Models for e-Science Applications	211
6.3	Architecture Implementation for the WISDOM Applications	214
6.3.1	Basic Framework of the WISDOM Initiative	215
6.3.2	Scientific Applications of the Bio-informatics Domain	217
6.3.3	Academic Analysis and Production Infrastructure Setup Experience	218
6.3.4	Reference Model Impact and Applicability	220
6.4	Architecture Implementation for the VPH Applications	225
6.4.1	The STEP Roadmap and the Basic VPH Framework	226
6.4.2	Scientific Applications of the e-Health Domain	228
6.4.3	Academic Analysis and Production Infrastructure Setup Experience	229
6.4.4	Reference Model Impact and Applicability	230
6.5	Architecture Implementation for the EUFORIA Applications	236
6.5.1	The EUFORIA Framework	237
6.5.2	Scientific Applications of the Fusion Domain	239
6.5.3	Academic Analysis and Production Infrastructure Setup Experience	240
6.5.4	Reference Model Impact and Applicability	242
6.6	Architecture Implementations for ESFRI and other Applications	246
6.6.1	Basic Framework for ESFRI Projects	247
6.7	Conclusion	249
7	Conclusion	251

Computational simulations and thus scientific computing is the third pillar alongside theory and experiment in today's science. The term e-Science evolved as a new research field that focuses on collaboration in key areas of science using next generation data and computing infrastructures (i.e. e-Science infrastructures) to extend the potential of scientific computing. During the past decade, significant international and broader interdisciplinary research is increasingly carried out by global collaborations that often share resources within a single production e-Science infrastructure. More recently, increasing complexity of e-Science applications that embrace multiple physical models (i.e. multi-physics) and consider a larger range of scales (i.e. multi-scale) is creating a steadily growing demand for world-wide interoperable infrastructures that allow for new innovative types of e-Science by jointly using different kinds of e-Science infrastructures. But interoperable e-Science infrastructures are still not seamlessly provided today and this thesis argues that this is due to the absence of a production-oriented e-Science infrastructure reference model. The goal of this thesis is thus to present an infrastructure interoperability reference model (IIRM) design tailored to production needs and that represents a trimmed down version of the Open Grid Service Architecture (OGSA) in terms of functionality and complexity, while on the other hand being more specifically useful for production and thus easier to implement. This reference model is underpinned with lessons learned and numerous experiences gained from production e-Science application needs through accompanying academic case studies of the bio-informatics, e-Health, and fusion domain that all seek to achieve research advances by using interoperable e-Science infrastructures on a daily basis. Complementary to this model, a seven segment-based process towards sustained infrastructure interoperability addresses important related issues like harmonized operations, cooperation, standardization as well as common policies and joint development roadmaps.

This publication was edited at the Jülich Supercomputing Centre (JSC) which is an integral part of the Institute for Advanced Simulation (IAS). The IAS combines the Jülich simulation sciences and the supercomputer facility in one organizational unit. It includes those parts of the scientific institutes at Forschungszentrum Jülich which use simulation on supercomputers as their main research methodology.