HAL
open science

OEIRM: An Open Distributed Processing Based Interoperability Reference Model for e-Science

Zhiming Zhao, Paola Grosso, Cees De Laat

To cite this version:

Zhiming Zhao, Paola Grosso, Cees De Laat. OEIRM: An Open Distributed Processing Based Interoperability Reference Model for e-Science. 9th International Conference on Network and Parallel Computing (NPC), Sep 2012, Gwangju, South Korea. pp.437-444, 10.1007/978-3-642-35606-3_52 . hal-01551380

HAL Id: hal-01551380

https://inria.hal.science/hal-01551380v1

Submitted on 30 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

OEIRM: An Open Distributed Processing based Interoperability Reference Model for e-Science

Zhiming Zhao, Paola Grosso, Cees de Laat

System and Network Engineering research group
Informatics Institute, University of Amsterdam
Science Park 904, 1098XH, Amsterdam, the Netherlands
{Z.Zhao|P.Grosso|C.T.A.M.Delaat}@uva.nl

Abstract. E-Science applications are often interdisciplinary and require resources from different infrastructures; the interoperability between heterogeneous infrastructures is an important requirement for constructing large scale experiments. Analyzing interoperability issues by including both application and infrastructure aspects promotes a global view on interoperability of different layers, and we argue it can converge to an optimal reference model to guide the development of service layers in e-Science infrastructures. This paper proposes an Open Distributed Processing based interoperability reference model for e-Science.

Keywords: e-Science, Open Distributed Processing, Interoperability, Cloud, Grid.

1 Introduction

E-Science applications are characterized by big data and interdisciplinary approaches, and require not only large computing and storage capacities, but also effective cooperative computing model to handle the application complexity [1]. Collaborating different infrastructures essentially enable the construction of large scale experiments; however, the technical diversities between infrastructures often make such collaboration difficult, and the service layer interoperability between different infrastructures has been highlighted as a key functional requirement [2].

Historically, e-Science infrastructures are often developed for specific application domain(s); the diverse application requirements result in different focuses on the infrastructure services. Even started with homogenous architecture, the evolution of technologies may still push the infrastructure upgrade towards different directions. Therefore, the interoperability between different infrastructures and the compatibility with the legacy applications from the early infrastructures will remain as a requirement in e-Science for a long term [3]. Many of the existing interoperability solutions focus on specific layers in the context e-Science: between infrastructures [4, 5], between middleware [6], and between workflows [7], and typically via iterative steps: building adapters or connectors between two infrastructures and then deriving new service layer models for standardization via certain community efforts. Such

iterations can continuously promote the evolution of standards for infrastructures and the above service layers, but will not completely solve the interoperability problems when the diversity between infrastructures and the missing links between standards remain [36]. Providing interoperability solutions only at a specific layer without a global view of the entire e-Science context hampers the convergence of service layer evolution. An interoperability reference model is needed to complement the model of the application and infrastructure [8].

In this paper, we propose an interoperability reference model to analyze related issues and review the existing solutions in e-Science. We discuss how the interoperability issues should be handled in the context of the application life-cycle. The research is conducted in the context of EU F7 project ENVRI [9].

2 An Open Distributed Processing based reference model

The Open Distributed Processing (ODP) model captures the design and development issues in complex distributed systems into five viewpoints, namely enterprise, information, computational, engineering and technology [10]. Although it is originally developed in the enterprise IT context, its decomposed view on system design and development contributes valuable model for analyzing the interoperability in e-Science. Following the ODP methodology, we highlight five viewpoints in analysing the interoperability in the e-Science context: scientific, information, computational, middleware and infrastructure, as shown in Fig. 1.

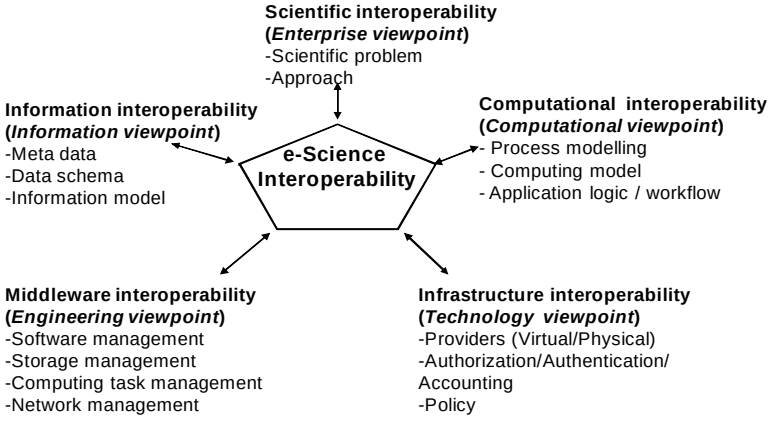

Fig. 1. An Open e-Science Interoperability Reference Model (OEIRM).

In the next section, we will review the available solutions of the interoperability using this model.

3 Interoperability in e-Science: from infrastructure to application

The five viewpoints proposed in Fig. 1 model the interoperability related issues not only in the service layers in infrastructure but also the characteristics of scientific domain and applications.

3.1 The scientific interoperability

From the scientific viewpoint, e-Science activities often require technologies or software components originally developed by different domains. The interoperability between different scientific domains and disciplines has been highlighted as the basic approach in e-Science. For instance, modeling and simulation based computing methods extend traditional laboratory based experimental sciences such as chemistry, physics and finance [11, 12], and semantic web technologies have been promoted as basis for building information models and system integration in large collections fields [13]. Moreover, many simulation models have also exhibited their power in different contexts, for instance finite element simulation model shows its strong power not only in fluid dynamics but also in studying financial problems [14]. In specific fields, the BRIDGE project aims at the distributed optimization using different simulation models and cross both ChinaGrid and EU SIMDAT [15]. These types of interoperability comprises of the basis of the interdisciplinary activities in e-Science.

3.2 The information interoperability

The information interoperability refers to not only data content but also meta metadata, information models and all kinds of descriptions of the services that e-Science applications need. The solutions can be roughly classified into two groups. The first one is to map diverse information models using connectors, for instance the Advanced Resource Connector ARC between EGEE and NorduGrid [16]. Semantic level interoperability enables the data exchange at the meta level, for instance the Linked Open Data principles proposed by the EU environmental sciences is for sharing scientific data [17], semantic web services based mediators [4], and semantic information preprocessing framework [18]. The second one is to synchronize diverse information representation mechanisms to one common model, for instance OGF promotes a solution named Grid Laboratory Uniform Environment (GLUE) [19]. TPT (Taiwan prescription template) is an information level interoperability solution proposed by Kuo to enable the medical exchange among different infrastructure, and is promoted as the standards for improving the patient information [20]. The information interoperability potentially realizes the semantic level information exchange between different e-Science domains.

3.3 The computational interoperability

From the computational viewpoint, 1) the application can be interoperable with the others by sharing and exchanging the intelligence in e-Science applications, i.e. the composition of the application logic 2) the execution engine of the application can be interoperable to execute applications from the other engines. Currently, the scientific workflow management systems from scientific computing context, and the MapReduce related Cloud computing model from large internet companies such as Google are two main forces pushes the evolution of the cooperative computing model. Although aiming at different types of data and application contexts, these two models merging from different perspectives to get benefits from each other, for instance MapReduce is included as an execution model in Kepler [20], or DAG is include in MapReduce to extend its flow control [21]. Workflow bus was proposed to realize engine level interoperability between different systems [22]. The computational interoperability focuses on the runtime issues of the e-Science applications.

3.4 The middleware interoperability

The middleware interoperability refers to cross invocation of services among different service layers, and mainly deals with the diversity between middleware in invoking services, in managing data involved by the service invocation. Typical approach is to provide a gateway node to map the invocation interface of both sides. Wang proposed a gateway between the submission tool in gLite and the CNGrid GOS environment, and a scheduler decides the jobs from the gateway queue to middleware of which side [23]. Another way to solve such interoperability issues is to adopt a standardized service interface for job submission, storage and network management; OGF promoted a standards named HPC basic profile (HPCBP) towards this direction [24]. Web services interoperability (WS-I) is another initiative to promote the interoperability for the web service stacks [25], although it faces several challenges to commit is final mission. An important problem that in the middleware interoperability is how to efficiently utilize the resources from different domains after the middleware interoperability has been enabled. Saga deals with the load balance issue in this context by proposing a HPCBP gateway, so that the jobs from other infrastructures will be checked against the allocation table of the local infrastructure scheduler with certain load balance constraints [26]. The Grid meta scheduler (GMS) can schedule computing tasks of workflows among several middleware gLite, Unicore and Globus environments, but mainly batch and job farming based tasks [27]. GridARS is another example but focuses on reserving the resources among different infrastructures [28].

3.5 The infrastructure interoperability

From the infrastructure viewpoint, distributed resources have been aggregated and virtualized as different stack of services; from Clusters, Grids and Clouds. There are

several standardization initiatives for Cloud models, for instance Open Cloud Computing Interface(OCCI) [29], Open Virtualization Format [30], and Cloud data management interface (CDMI) [31]. Moreover, whole organization related issues such as management policies for sharing data and resources, including authorization, authentication and accounting issues are included, for instance security level interoperability [32,33], using model checking approach to verify the policies and rules [34].

Table. 1 summarizes the existing work in one table.

Table 1. Interoperability in e-Science.

	Involved issues	Approach and examples
Scientific interoperability	-Scientific problems - Scientific methods	-Problem to large system scale, and involve different domains. Interdisciplinary optimization, Bridge project [15]. -Computational sciences in many fields [11,12] Simulation models extended to different fields [14]
Information interoperability	-Scientific data -Resource description, meta data, information model	- Community based ontology framework, Linked Open Data [7] - Ontology mapping , Semantic based information preprocessing[18], ARC[16], Semantic Web Services mediators[4], - Standardized information model, GLUE [18], TPT[19],
Computational interoperability	-Joint computing model -Runtime embedding -cross invocation	- MapReduce in Scientific workflow [20], flow control in MapReduce [21] - Runtime engine: Workflow bus[22],
Middleware interoperability	-Cross data access -Cross job submission -Cross infrastructure load balance	- Gateway based protocol mapping [23] - HPCBP [24], WS-I [25] - GridARS [28], GMS [27],
Infrastructure interoperability	-Authentication/Authorization/ Accounting/Security/Policy -Infrastructure model	- Authentication/Authrization/Accounting: SAML, X.509 [32,33], XACML[34] -OCCI [29], OVF[30], CDMI[31]

4. Related work

Interoperability has been recognized as a key issue in different contexts, and several reference models have been proposed. Riedel proposed an Infrastructure Interoperability Reference Model (IIRF) [35], which is now in the OFG community effort. The main idea of the IIRF is to fill the missing links between the existing earlier interoperability related middleware or standards, in particular the interoperability between infrastructure for High performance computing and high throughput computing. Open Grid Forum set up a working group called Grid Interoperability Now! (GIN) [36]. The IIRF focuses on the different standards in the existing Grid middleware; however, its basic concepts are compatible with model presented in this paper.

5. Summary

The research presented in this paper is conducted in the context of the EU ENVRI project; we aim at an interoperability model for promoting the resource sharing

among infrastructures from different environmental science domains [9]. Using the Open Distributed Processing (ODP) model, we proposed an interoperability reference model named OEIRM, and analyzed the interoperability issues in the life cycle of e-Science applications using five different viewpoints of the model.

The diversity of the scientific domains drives the interdisciplinary research in e-Science from various angles; solving the diversity by providing interoperability support only promotes the evolution of scientific domains and all kinds of support technologies, but will not eventually remove such diversity. From this point of view, interoperability will be a continuous problem exists in the e-Science context. Different interoperability technologies as we see from the above analysis play a central role in enabling the collaborations in different contexts, but they also raises a challenge problem for the e-Science researchers: how we should deal with the interoperability issues in an evolutionary vision of the infrastructure and e-Science applications? As the next step, we will develop the reference model to guide the development of future e-Science application and service layers of infrastructures.

Acknowledgments. We would like to thank the FP7 EU funded Integrated project ENVRI project (project number 283465) and the Dutch national and education network SURFnet, through the GigaPort Research on Network (RoN) project for sponsoring this research..

References

1. Francis, J., Alexander, A., Alexander., S.: Big data. Computing in Science and Engineering, 13:10–13 (2011)
2. Geraci, A.: IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries. IEEE Press, Piscataway, NJ, USA (1991)
3. Charalabidis, Y., Janssen, M. Glassey, O.: Introduction to cloud infrastructures and interoperability minitrack. Hawaii International Conference on System Sciences, 0:2177 (2012)
4. Ngan, L., Feng, Y., Rho, S., Kanagasabai, R.: Enabling interoperability across heterogeneous semantic web services with owl-s based mediation. Asia-Pacific Conference on Services Computing. IEEE, 0:471–476 (2011)
5. Arcieri, F., Fioravanti, F., Nardelli, E., Talamo, M.: A layered it infrastructure for secure interoperability in personal data registry digital government services. Research Issues in Data Engineering, International Workshop on, 0:95–102 (2004)
6. Blair, G., Grace, P.: Emergent middleware: Tackling the interoperability problem. IEEE Internet Computing, 16:78–82 (2012)
7. Zhao, A., Belloum, A., Laat, C., Hertzberger, B., Using jade agent framework to prototype an e-science workflow bus. In 7th IEEE International Symposium on Cluster Computing and the Grid, pages 655–660 (2007)
8. White, L., Wilde, N., Reichherzer, T., El-Sheikh, M., Goehring, G., Baskin, A., Hartmann, B., Manea, M.: Understanding interoperable systems: Challenges for

the maintenance of soa applications. Hawaii International Conference on System Sciences, 0:2199–2206 (2012)
9. EU 7th framework programme. Common operations of environmental research infrastructures, grant agreement no 283465, www.envri.eu.
10. Linington, P., Milosevic, Z., Tanaka, A., Vallecillo, A.: Building enterprise systems with ODP. CRC press, Taylor & Francis group (2012)
11. Cassel, L.: Interdisciplinary computing is the answer: now, what was the question? ACM Inroads, 2(1):4–6 (2011)
12. Smarkusky, D. and Toman, S.: An interdisciplinary approach in applying fundamental concepts. In Proceedings of the 10th ACM conference on SIG-information technology education, SIGITE ’09, pages 224–228 (2009)
13. Vattan, S., Goel, A.: Semantically annotating research articles for interdisciplinary design. In Proceedings of the sixth international conference on Knowledge capture, K-CAP ’11, pages 165–166 (2011)
14. Qiu, G., Kandhai B., Sloot, P.: Understanding the complex dynamics of stock markets through cellular automata. Phys Rev E, 75:046116 (2007)
15. Wang, Y., Roberto, D., Boniface, M., Qian, D., Cui, D., Jiang, J.: Cross-domain middlewares interoperability for distributed aircraft design optimization. eScience, IEEE International Conference on, 0:485–492 (2008)
16. Ellerta, M., Grnagerb, M., Konstantinovc, A., et. al.: Advanced resource connector middleware for lightweight computational grids. Future Generation Computer Systems, page 219240 (2007)
17. Hausenblas, M.: Utilising linked open data in applications. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, pages 7:1–7:4, (2011)
18. Zhao, Z., Taal, A., Grosso, P., Laat, C.: Resource discovery in large scale network infrastructure. In IEEE Int’l conf on Networking architecture and storage, (2011)
19. Field, L., Andreozzi, S., Konya, B.: Grid information system interoperability: The need for a common information model. IEEE International Conference on e-Science 0:501–507 (2008)
20. Kuo, C., Li, Y., Lee, P., Wu, Y.: An interoperability infrastructure with portable prescription for improving patient safety – the framework of a national standard in taiwan. Computer Science and Information Engineering, World Congress on, 1:293–297 (2009)
21. Budiu, M., Delling, D., Werneck., R.: Dryadopt: Branch-and-bound on distributed data-parallel execution engines. Parallel and Distributed Processing Symposium, International, 0:1278–1289 (2011)
22. Zhao, Z., Booms, S., Belloum, A., Laat, C., Hertzberger, B.: Vle-wfbus: a scientific workflow bus for multi e-science domains. In Proceedings of the 2nd IEEE International conference on e-Science and Grid computing, pages 11–19 (2006)
23. Wang, Y., Scardaci, D., Yan, B., Huang, Y.: Interconnect egee and cngrid e-infrastructures through interoperability between glite and gos middlewares. e-Science and Grid Computing, International Conference on, 0:553–560 (2007)

24. Ruiz-Alvarez, A., Smith, C., Humphrey, M.: Bes++: Hpc profile open source c implementation. Grid Computing, IEEE/ACM International Workshop on, 0:41–48 (2008)
25. Schoenberger, A., Schwalb, J., Wirtz, G.: Has ws-i's work resulted in ws-* interoperability? Web Services, IEEE International Conference on, 0:171–178 (2011)
26. Kazushige S., Aida, K., Miura., K.: Mutual job submission architecture that considered workload balance among computing resources in the grid interoperation. Grid Computing, IEEE/ACM International Workshop on, 0:19–25 (2011)
27. Mirto, M., Passante, M., Aloisio, G.: A grid meta scheduler for a distributed interoperable workflow management system. Computer-Based Medical Systems, IEEE Symposium on, 0:138–143 (2010)
28. Takefusa, A., Nakada, H., Takano, R., Kudoh, T., Tanaka, Y.: Gridars: a grid advanced resource management system framework for intercloud. In Proceedings of the conference on Cloud computing technology and science, pages 705–710 (2011)
29. Open Grid Forum. Open cloud computing interface infrastructure. http://forge.ogf.org/sf/wiki/do/viewPage/proejcts.occi-wg/wiki/Infrastructure
30. Distributed management task force. Open virtualization format specification. In dsP0243 version 1.1.0 (2010)
31. Storage networking industry association. Cloud data management interface. In Storage networking industry association, tech report, (2010)
32. Cantor, J., Philpott, P., Maler, E.: Assertions and protocols for the oasis security assertion markup language. In OASIS Standard (2005)
33. Moses, T., et. al. extensible access control markup language. In OASIS Standard (2005)
34. Maarabani, M., Cavalli, A., Hwang, I., Zaidi, F.: Verification of interoperability security policies by model checking. High-Assurance Systems Engineering, IEEE International Symposium on, 0:376–381 (2011)
35. Riedel, M., Memon, M., Memon, A., Mallmann, D., Lippert, T., Kranzlmuller, D.: e-science infrastructure integration invariants to enable htc and hpc interoperability applications. IEEE International Symposium on Parallel and Distributed Processing Workshops 0:922–931 (2011)
36. Riedel, M., Laure, E., Soddemann, T., Field, L, et al.: Interoperation of world-wide production e-science infrastructures. Concurr. Comput. : Pract. Exper., 21(8):961–990 (2009)