Service-Oriented Science:
Scaling the Application and Impact of eResearch
(Abstract)

Ian Foster
Math & Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, U.S.A.
Department of Computer Science, University of Chicago, Chicago, IL 60637, U.S.A.
foster@mcs.anl.gov

The importance of service-oriented architecture for science is widely recognized. Increasingly, scientific communities are making information tools accessible as services that clients can access over the network, without knowledge of their internal workings. In this way, tools formerly accessible only to the specialist can be made available to all. Equally importantly, new value-added services can be constructed that integrate other services to automate useful tasks. The value of such *service-oriented science* has been demonstrated in disciplines as diverse as astronomy [9], biology [6, 8], and fusion science [7].

I am concerned here with the question of how we can scale the application and impact of such approaches. I argue that a key to progress is achieving a separation of concerns between the construction and hosting of science services, so that scientists can focus on constructing content rather than on the minutiae of operating scalable and robust services. I suggest that if appropriate mechanisms are identified, we can enable distinct content provider, service provider, and resource provider roles, with associated economies of scale [3].

The mechanisms required to achieve these goals are provided, in part, by Grid infrastructure [5]. I review the mechanisms that have been developed to date for Grid infrastructure and experience gained implementing these mechanisms, for example within the open source Globus Toolkit version 4 [2]. I present a range of dynamic service deployment scenarios, in which for example the TeraGrid [1] and Open Science Grid [4] are used to host services for science communities. I discuss how these scenarios demonstrate the potential for scaling service-oriented science

References

1. Catlett, C. The TeraGrid: A Primer, 2002. www.teragrid.org.
2. Foster, I., Globus Toolkit Version 4: Software for Service-Oriented Systems. *IFIP International Conference on Network and Parallel Computing*, 2005, Springer-Verlag LNCS 3779, 2-13.
3. Foster, I. Service-Oriented Science. *Science*, *308*. 814-817. 2005.
4. Foster, I. and others, The Grid2003 Production Grid: Principles and Practice. *IEEE International Symposium on High Performance Distributed Computing*, 2004, IEEE Computer Science Press.
5. Foster, I. and Tuecke, S. Describing the Elephant: The Different Faces of IT as Service. *ACM Queue*, *3* (6). 26-29. 2005.
6. Goble, C., Pettifer, S. and Stevens, R. Knowledge Integration: In silico Experiments in Bioinformatics. *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 2004.
7. Keahey, K., Fredian, T., Peng, Q., Schissel, D.P., Thompson, M., Foster, I., Greenwald, M. and McCune, D. Computational Grids in Action: the National Fusion Collaboratory. *Future Generation Computing Systems*, *18* (8). 1005-1015. October 2002.
8. Stein, L. Creating a Bioinformatics Nation. *Nature*, *317*. 119-120. 2002.
9. Szalay, A. and Gray, J. Scientific Data Federation: The World Wide Telescope. *The Grid: Blueprint for a New Computing Infrastructure (2nd Edition)*, Morgan Kaufmann, 2004.