



Methods for Federating and Transferring Data to eScience Applications  
*S. Koulouzis*

# Methods for Federating and Transferring Data to eScience Applications

Spyridon Koulouzis

The main objective of this research is to investigate efficient, scalable, flexible and transparent methods for moving large volumes of data between workflow tasks and distributed, heterogeneous and independent storage resources.

We address the challenges rising from the exchange of large data volumes between web services in scientific applications by proposing a data pipeline model to reduce workflow execution times and demand for storage and network resources [1, 2].

We also propose a data resource federation architecture which is technology agnostic and enables a unified view of the data resources providing an abstraction layer under which independent data storage resources are coordinated [3].

To address scalability and performance issues we have extended our initial data management architecture [4] by introducing modules that can be deployed on multiple and heterogeneous infrastructures. Next, we investigated how programmable networks can reduce the execution time of data and I/O intensive workflows [5].

To demonstrate the usage of the proposed methods and tools, we have applied them to real world applications. The performance and usability of our data pipeline model for web services is evaluated with two workflows [6, 2, 7]. Next, we applied our storage federation approach to a well known data-intensive workflow based on Montage [5]. Finally, we analyze usage data of our storage federation approach coming from the VPH-Share project infrastructure which is used for executing medical applications.

## References

- [1] S. Koulouzis, E. Meij, M. Marshall, and A. Belloum, “Enabling Data Transport between Web Services through alternative protocols and Streaming,” in *eScience, 2008. eScience '08. IEEE Fourth International Conference on*, pp. 400–401, December 2008.
- [2] S. Koulouzis, R. Cushing, K. Karasavvas, A. Belloum, and M. Bubak, “Enabling Web Services to Consume and Produce Large Datasets,” *Internet Computing, IEEE*, vol. 16, pp. 52–60, January 2012.
- [3] S. Koulouzis, D. Vasyunin, R. Cushing, A. Belloum, and M. Bubak, “Cloud Data Federation for Scientific Applications,” in *Euro-Par 2013: Parallel Processing Workshops*, vol. 8374 of *Lecture Notes in Computer Science*, pp. 13–22, Springer Berlin Heidelberg, 2014.

- [4] S. Koulouzis, A. Belloum, M. Bubak, P. Lamata, D. Nolte, D. Vasyunin, and C. de Laat, “Distributed Data Management Service for VPH Applications,” *IEEE Internet Computing*, vol. 20, pp. 34–41, March 2016.
- [5] S. Koulouzis, A. S. Belloum, M. T. Bubak, Z. Zhao, M. Živković, and C. T. de Laat, “SDN-Aware Federation of Distributed Data,” *Future Gener. Comput. Syst*, vol. 56, pp. 64–76, March 2016.
- [6] S. Koulouzis, E. Meij, and A. Belloum, “Enabling Large Data Transfers Between Web Services,” in *5th EGEE User Forum 2010/Book of Contributions* (E. Floros, ed.), (Uppsala), pp. 153–154, 2010.
- [7] S. Koulouzis, E. Zudilova-Seinstra, and A. Belloum, “Data transport between visualization web services for medical image analysis,” *Procedia Computer Science*, vol. 1, no. 1, pp. 1727 – 1736, 2010. ICCS 2010.