

Software + Services for Data Storage, Management, Discovery, and Re-Use

Alex Wade¹, & Lee Dirks²

¹Microsoft Research Cambridge, 7 JJ Thomson Ave., Cambridge, United Kingdom,
awade@microsoft.com

²Microsoft Research, One Microsoft Way, Redmond, WA, United States,
ldirks@microsoft.com

With a growing demand for transparency and openness around scientific research and an emphasis on the sharing of scientific workflows and datasets, there is a similarly increasing number in the variety of client and web-based tools required to manage each stage in the lifecycle of individual datasets. Datasets are produced from a variety of instruments and computations; are analyzed and manipulated; are stored and referenced within the context of a research project; and, ideally, are archived, stored, and shared with the scientific community and the rest of the world. Each of these efforts, however, requires a number of user actions involving a growing number of systems and interfaces. In an effort to preserve the flexibility and autonomy of the researchers, but also to minimize the logistical effort involved, we present an approach to this problem through the integration of workflow execution, project collaboration, project-based dataset management and versioning, and long-term archiving and dissemination. This example demonstrates the orchestration of a number of existing Microsoft Research projects; however, the interaction between each uses existing web interoperability protocols and can easily support the replacement of individual architectural components with related services.