

Introduction to OGSA-DAI Services

Konstantinos Karasavvas¹, Mario Antonioletti², Malcolm Atkinson¹,
Neil Chue Hong², Tom Sugden², Alastair Hume², Mike Jackson²,
Amrey Krause², and Charaka Palansuriya²

¹ National e-Science Centre, University of Edinburgh, Edinburgh EH8 9AA, UK
{kostas, mpa}@nesc.ac.uk

² EPCC, University of Edinburgh, JCMB, The King's Buildings,
Mayfield Road, Edinburgh EH9 3JZ, UK
{mario, neilc, tom, ally, michaelj, amrey, charaka}@epcc.ed.ac.uk

Abstract. In today's large collaborative environments, potentially composed of multiple distinct organisations, uniform controlled access to data has become a key requirement if these organisations are to work together as *Virtual Organisations*. We refer to such an integrated set of *data resources*¹ as a *virtual data warehouse*. The *Open Grid Services Architecture - Data Access and Integration* (OGSA-DAI) project was established to produce a common middleware solution, aligned with the Global Grid Forum's (GGF) OGSA vision [OGSA] to allow uniform access to data resources using a service based architecture. In this paper the service infrastructure provided by OGSA-DAI is presented providing a snapshot of its current state, in an evolutionary process, which is attempting to build infrastructure to allow easy integration and access to distributed data using grids or web services. More information about OGSA-DAI is available from the project web site: www.ogsadai.org.

Keywords: Data, Databases, Grid, DAIS, OGSA-DAI, Open Grid Services Architecture, Web Services.

1 Introduction

Access to and the sharing of data across organisational boundaries is an important requirement for a large number of UK national and international collaborative projects. Instead of requiring each of these projects to individually solve the same data access problem, the OGSA-DAI project was established to produce a common middleware solution that allows uniform access to data resources using a service-based architecture. The initial objectives of the project have concentrated on developing the base data access platform and now, gradually, to

¹ A data resource here is taken to mean any entity that can act as a source and/or sink of data together with its associated management framework. Although the framework being developed at the moment works mainly with databases the scope is more general and could encompass file systems and streams.

focus on more sophisticated functionality that offers data integration capabilities, such as *distributed query processing* provided by the OGSA-DQP project [OGSA-DQP] using OGSA-DAI services. Up to the advent of OGSA-DAI, provision for uniform data access through service-based interfaces was absent from the then available Grid toolkits, such as the Globus Toolkit 3.0² (GT3) and Unicore.

In moving towards these ends the development of OGSA-DAI has been guided by a set of key design principles, mainly:

- *Avoid unnecessary communication between a service and its clients.* In order to minimise the number of message exchanges between a service and its clients, multiple interactions are abstracted into a set of *activities* which are then contained in a single document, referred to as a *perform document*³ requiring a single message exchange.
- *Avoid unnecessary data movement.* Wherever possible move the computation to the data. Capabilities already available to a data resource, together with additional capabilities implemented at the service layer such as data transformations or third party data delivery, are exposed through the *activities* which can be linked as a series of pipelined tasks through which data flows, all within the same perform document. These activities are then executed within the scope of a single service interaction near or at the data source.
- *Provide an extensible activity framework.* It is unlikely that OGSA-DAI will provide all the base functionality, implemented as activities, that a given project might require. Thus, from the outset, perform documents and the activity *engine* have been designed to be extensible. New functionality can be implemented as activities and incorporated to work within the existing framework.
- *OGSA-DAI does not provide a complete virtualisation of the underlying data resource.* It is still necessary to know what the underlying data resource is and target suitable queries for that type of data resource. The infrastructure does not automatically do this; OGSA-DAI is not attempting to create a new universal query language suitable for all types of data resource.
- *Provide an extensible architecture.* Allow the OGSA-DAI framework to be customised or extended, e.g to add a stronger security model.
- *Build the middleware using existing standards and, where these do not exist, try to produce standards to fill in the gaps.* Up to release 5, the OGSA-DAI middleware has been based on the GGF *Open Grid Services Infrastructure* recommendation [OGSI] and its dependencies. At the time there were no

² From GT3.2 OGSA-DAI became a contributed component to the Globus Toolkit; an endorsement that such functionality was required. It is also still distributed independently of Globus through the OGSA-DAI project web site, from which there have been over 3300 downloads by approximately a 1000 registered users.

³ An additional advantage of using a document based interface is that only one operation is required at the interface, *perform*. Functionality can easily be added without requiring a change to the interface, as this can be done inside the document.