

# Grid Enabling Your Data Resources with OGSA-DAI

Mario Antonioletti<sup>1</sup>, Malcolm Atkinson<sup>2</sup>, Neil P. Chue Hong<sup>1</sup>,  
Bartosz Dobrzelecki<sup>1</sup>, Alastair C. Hume<sup>1</sup>, Mike Jackson<sup>1</sup>,  
Kostas Karasavvas<sup>2</sup>, Amy Krause<sup>1</sup>, Jennifer M. Schopf<sup>2,3</sup>,  
Tom Sugden<sup>1</sup>, and Elias Theocharopoulos<sup>2</sup>

<sup>1</sup> EPCC, University of Edinburgh, JCMB, The King's Buildings,  
Mayfield Road, Edinburgh EH9 3JZ, UK

N.ChueHong@epcc.ed.ac.uk

<sup>2</sup> National e-Science Centre, University of Edinburgh & Glasgow,  
Edinburgh EH8 9AA, UK

<sup>3</sup> Distributed System Laboratory, Argonne National Laboratory,  
Argonne, IL, 60439 USA

**Abstract.** OGSA-DAI (Open Grid Services Architecture - Data Access and Integration) provides an extensible software framework allowing data resources, such as files, relational and XML databases, to be exposed through Web services acting within collaborative Grid environments or, more modestly, in stand-alone mode. OGSA-DAI may be deployed to WSRF-based platforms, such as the Globus Toolkit 4, as well as non-WSRF based ones, such as the UK OMII Server or standard versions of Tomcat and axis. Regardless of the platform, the core functionality provided remains the same. OGSA-DAI allows data resources to be accessed and integrated into the main infrastructures presently being used to construct Grids. OGSA-DAI provides a number of optimisations that reduce unnecessary data movement by shifting work to the Web service and encapsulating multiple client-Web service interactions into a single one, and allows for functionality to be added or customised based on the application. OGSA-DAI is widely used and is available from [www.ogsadai.org.uk](http://www.ogsadai.org.uk). It is also bundled with the OMII-UK and Globus Toolkit distributions. This paper gives an overview of what OGSA-DAI is, how it works, presents some usage scenarios, and outlines future enhancements.

**Keywords:** Data, Databases, Grid, OGSA-DAI.

## 1 Introduction

With current advances in technology and the decreasing cost of storage, increasingly large amounts of data are being produced, maintained, kept on-line, and shared within communities. For instance, astronomers are collecting data together, such as surveys of the sky made at different wavelengths and resolutions, and making it collectively available through *Virtual Observatories* [1]; biologist

are gathering DNA and genomic data from different species and making this data available to biologists through data stores, providing a rich source of data to pursue insights into biological systems [2] and in the health sector, digital medical data are being collected and maintained by hospitals allowing experts to collaborate in patient diagnosis and providing case histories that can be used to inform a prognosis for patients suffering from similar maladies [3,4].

The need to access disparate data sources, often spanning multiple institutions, can lead to new insights and discoveries to be made. By combining different wavelength data for the same patch of sky, astronomers have been able to make new discoveries that would not have otherwise been possible from a single survey [5,6]. Biologists now have the capability of performing cross-species comparisons to determine new genes and their function [7]. Doctors can improve the diagnosis of breast cancer by comparing current mammographs with old mammographs in combination with the associated patient histories [8]. The advantage being able to share data and resources in a controlled manner within a collaborative environment is clear. The provision of generic middleware to facilitate this process is the ethos that is currently driving the evolution of the Grid and, in the data area, OGSA-DAI (Open Grid Services Architecture - Data Access and Integration) provides software which makes it easy to publish and share data across organisational boundaries, and develop applications which use both public and personal data resources, through a secure, extensible framework based on web service standards.

OGSA-DAI is not the only solution currently available for data in the Grid space. *Storage Resource Broker* (SRB) [9], developed by the San Diego Supercomputer Center, provides access to collections of data primarily using attributes or logical names rather than using the data's physical names or locations. SRB is primarily file oriented, although it can also work with various other data object types. OGSA-DAI on the other hand takes a database oriented approach to its access mechanisms. *WebSphere Information Integrator* (WSII), a commercial product from IBM, provides data searching capabilities spanning organisational boundaries, provides a means for federating and replicating data, as well as allowing for data transformations and data event publishing to take place [12]. A more detailed comparison between OGSA-DAI and WSII can be found in [10]. *Mobius* [11], developed at Ohio State University, provides a set of tools and services to facilitate the management and sharing of data and metadata in a Grid environment. To expose XML data in Mobius, the data must be described using an XML Schema, which is then shared via their Global Model Exchange. Data can then be accessed by querying the Schema using, for example, XPath. OGSA-DAI, in contrast, does not require an XML Schema to be created for each piece of data; rather, it directly exposes that information (data and metadata/schema) and relies on the resource's intrinsic querying mechanisms to query its data. These three products all provide mechanisms to share data across organisational boundaries, however they complement the functionality provided by OGSA-DAI.