Accessing Web Databases using OGSA-DAI in BDWorld

Shirley Crompton[1], Brian Matthews[2],
Alex Gray[3], Andrew Jones[3], Richard White[3]

[1]CCLRC, Daresbury Laboratory, Warrington WA4 4AD, United Kingdom
s.y.crompton@dl.ac.uk
[2]CCLRC, Rutherford-Appleton Laboratory, Didcot, Oxfordshire OX11 0AX, United Kingdom
b.m.matthews@rl.ac.uk
[3]Cardiff School of Computer Science, Cardiff University, Cardiff CF24 3AA, United Kingdom
{w.a.gray, Andrew.C.Jones, r.j.white}@cs.cardiff.ac.uk

Summary. Interoperation of heterogeneous and autonomous data resources is a common data management issue for GRID applications. The OGSA-DAI [1] middleware addresses this need by providing a service-oriented framework aligned with OGSA [2] to faciliate the access and integration of database and semi-structured data resources in the GRID environment. In the BioDA [4] project, we are evaluating the benefits of using OGSA-DAI in bioinformatics GRIDs by establishing communication between OGSA-DAI and GRID project developers as well as through practical case studies involving current projects. In this paper, we describe our experience in applying OGSA-DAI R5 to one of these projects, BiodiversityWorld (BDWorld [5]) - a GRID-based problem solving environment that specialises in the exploration and analysis of patterns in global biodiversity.

BDWorld's database handling is characterised by the diverse types of database used and the heterogeneity of the data, with respect to both the data structures and standards. Many of these databases are also autonomous internet information resources that a client queries by keyword searches rather than SQL. In contrast to the diversity of data resources used, BDWorld currently only requires a limited range of operations on these resources. One such operation is to create a study data set by aggregating data from iterative searches of remote data collections using the same biological taxon object as the search parameter. This particular scenario requires accessing and integrating heterogeneous data from distributed collections and, we feel, should make an ideal test case for OGSA-DAI.

There are two main ways in which we could introduce OGSA-DAI into BDWorld. One possibility is to leverage myGRID's [6] distributed query processing tool, OGSA-DQP [7] and use it to assist in planning the execution and distribution of data-oriented parts of a BDWorld workflow. But this would require a drastic revision to the BDWorld protocols [8]. Moreover, OGSA-DQP currently only supports popular relational database products and most BDWorld data resources are not exposed as relational databases. Another option is to layer a virtual OGSA-DAI Grid Data Service over the existing BDWorld database wrappers. This design (Figure 1) preserves the existing BDWorld universal resource invocation mechanism (see 8) and allows the existing database wrappers to be reused with minor modifications. This basic exemplar illustrates how OGSA-DAI could be modified to access web databases but leaving the data integration task to the BDWorld workflow system.

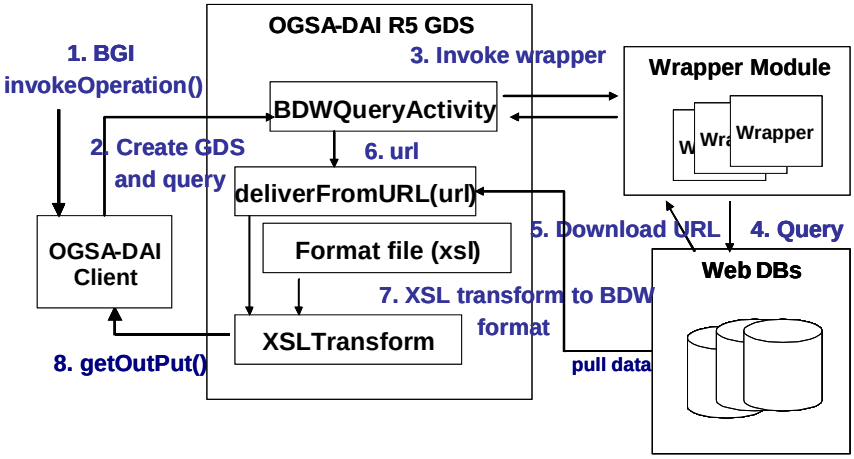

Figure 1. Basic Exemplar

Our experience shows that OGSA-DAI represents a ready-made solution to GRID-enable data resources compliant with popular GRID and web services specifications [9]. It comes with useful functionalities such as data translation and third-part delivery to fast-track the development of data-centric GRID applications. In the case of BDWorld, its application is constrained by the fact that BDWorld has a specialised architecture and communication protocols. Its predominant use of web databases, a type of data resource not supported by OGSA-DAI, also means that we cannot leverage the database and DQP functionalities. Also, our exemplar still requires the use of bespoke wrappers to query each web databases, which are now invoke via a custom OGSA-DAI activity rather than directly by the BDWorld Grid Interface. In conclusion, we feel that BDWorld will not get maximum benefits from the use of OGSA-DAI without having generic query access to its remote data.

References

1. Open Grid Services Architecture – Data Access and Integration, http://www.ogsadai.org.uk/
2. Foster, I., Kishimoto, H., Savva, A., Berry, D., Djaoui, A., Grimshaw, A., Horn, B., Maciel, F., Siebenlist, F., Subramaniam, R., Treadwell, J., Von Reich, J.: The Open Grid Services Architecture, Version 1.0 (2005)
3. GGF Data Access and Integration Services Working Group, https://forge.gridforum.org/projects/dais-wg/
4. Bioinformatics and DAIT, http://isegserv.itd.rl.ac.uk/BioDA/pages/default.htm
5. BiodiversityWorld, http://www.bdworld.org/
6. myGRID, http://www.mygrid.org.uk/
7. Open Grid Services Architecture – Distributed Query Processing, http://www.ogsadai.org.uk/dqp/
8. Jones, A.C., White, R.J., Gray, W.A., Bisby, F.A., Caithness, N., Pittas, N., Xu, X., Sutton, S., Fiddian, N.J., Culham, A., Scoble, M., Williams, P., Bromley, O., Brewer, P., Yesson., C. and Bhagwat, S.: Building a Biodiversity GRID. In: Konagaya, A. and Satou, K. (eds.): Grid Computing in Life Science, (LNCS 3370), Springer-Verlag (2005) 140-151
9. Atkinson, M., Karasavvas, K., Antonioletti, M., Baxter, R., Borley, A., Chue Hong, N., Hume, A., Jackson, M., Krause, A., Laws, S., Paton, N., Schopf, J., Sudgen, T., Tourlas, K. and Watson, P. A New Architecture for OGSA-DAI. In: Proc. 4th e-Science All-Hands Meeting, Nottingham (2005) *in press*