User Driven Dynamic Grid Services to Integrate New Users and Old Systems in eScience Facilities

Tom KIRKHAM[1], Thomas VARSAMIDIS[1], Julian GALLOP[2],
Laurent LERUSSE[2], Brian MATTHEWS[2]
[1]*School of Informatics, University of Wales Dean Street, Bangor, Gwynedd, UK*
E-mail: thomas@informatics.bangor.ac.uk, tdk1@informatics.bangor.ac.uk,
[2] eScience Centre, STFC Rutherford Appleton Laboratory, Chilton, Oxon
E-mail: j.r.gallop@rl.ac.uk, L.Lerusse@rl.ac.uk, b.m.matthews@rl.ac.uk

Abstract: The paper builds on the initial investigations in the Business Data Integration Framework for the Small-to-Medium Enterprise (BDIFS) project into using Grid Services to build dynamic eBusiness integration environments. The concept of sharing user-driven translation mappings in dynamic grid service environments is introduced. These translations provide the mapping between the source and target system and when combined with dynamic Grid Services in BDIFS, they are presented as remotely hosted services for facilities to use. The translations that the services provide are created via a tool which allows users to create the new translation mappings. This separation from basic service functionality and more specific translation functionality allows a bank of translations to be built up to fit a generic service invocation model. In this paper we demonstrate that for eScience facilities this provides a new approach to integrating temporary users of facilities with the data that the users create.

1. Introduction

The BDIFS project has been investigating the issues surrounding integration of legacy systems in Small-to-Medium Enterprises (SME) in North Wales since 2001 [1]. The issues addressed by the project are the challenges SMEs face in adopting automated eBusiness, linked specifically to the costs and technical challenges of implementing Enterprise Application Integration (EAI) software [2] and pressure to do so from larger customers [3].

This paper advances the BDIFS project by combining the integration issues that SMEs are confronted with in relation to data and user integration issues, with the same issues that are facing large scale science facilities as they adopt more distributed technologies. Large-scale facilities are often provided in central laboratories as an experimental resource, which is shared across the scientific community. Typically, experimental scientists visit the facilities to use its specialised instruments, e.g. analysis of a chemical sample. These analyses generate a large amount of data subsequently processed by the scientist. Using a case study in the Central Laser Facility (CLF), one of a number of large-scale facilities operated by the UK Science & Technology Facilities Council (STFC) [4] as an example, we address the issue of how to present data from legacy applications running on STFC's science facilities to meet the needs of visiting scientists. Such people are increasingly the council's customers, and often require wide and varied forms of data management.

To date, the management of data provided to visiting scientists in STFC has often been done by customized code and design, specifically tailored to the individual environments it was designed for. Therefore, in order to present data to a portal or other user-driven tool, customized programming needs to take place to present it in the appropriate format. The

BDIFS project assists business data integration and simplifies this data extraction process by developing a Service Orientated Architecture to aid data translation by a centralized set of resources. The case study presented applies the approach in an actual project and enhances the user-driven method for creating data translations.

2. Related Work

Grid computing is commonly associated with eScience and historically has mainly been applied in the domain of High Performance Computing. Within the UK, the flagship project is the National Grid Service (NGS) [5]. Funded by JISC, STFC and EPSRC, it forms the core UK Grid, and has been deployed for the production use of computational and data Grid resources in all branches of academic research [6]. A key issue for the NGS in recent years has been making data more accessible to users. To this, the development of portal technology for the NGS has been successful in linking it to the site's eScience facilities [7].

Whilst services like the NGS have been specialized to represent HPC resources to the community, emerging Grid computing technologies have the potential to represent other resources. Large-scale EU projects have looked to the Grid to provide resources to be incorporated into new generation business models, where service providers offer their resources in Grid models. Projects include GRASP [8] which focuses on charging models for application service provision, Trustcom [9] investigating establishing trust relationships and service level agreements between the services provided, and Akogrimo [10] which deals with the provision of mobile services as resources in Grid-based applications.

Trustcom and Akogrimo use emerging WS-Resource standards found in frameworks such as the Web Services Resource Framework (WSRF) [11]. Their aim is to dynamically group resources to present business applications [12]. To them, the Grid model includes the current state of resources which can be taken into account when selection is taking place. This use of states enables applications which use this Grid model to pick and choose between services providers offering the same service, but with different metrics of quality. In a similar fashion, BDIFS presents a platform where service providers can supply specific elements of the integration process, thus allowing businesses to integrate more readily. The difference in the BDIFS project compared to larger scale Grid Service projects is that it focuses on the specific problem of SMEs have when they need to integrate with larger partners [13]. This comparison is elaborated in Table 1. The approach to BDIFS has enabled the design and application to grow to the extent where SMEs can now begin to use it, and deployment models have been presented [14].

Table 1: Comparing VO Needs

	Traditional Grid (e.g. UK National Grid Service)	**Next Generation Grid (e.g. Akogrimo, Trustcom etc)**	**BDIFS**
Motivation	To provide computational resources otherwise unavailable to an organisation.	To provide a next generation of business applications that consist of workflows composed of multiple services.	To provide and develop access to complex ebusiness integration solutions otherwise out of reach of SMEs.
Service Creation	Services are permanently linked to VO.	Services are created dynamically from Service Provider domains	Services are created dynamically from central service that links to user populated service logic.
VO Management	Main VO is managed by a single organisation, with services have to abide by rules.	Main VO is managed by single organisation, with dynamic VOs managed by this domain but include services managed by 3rd parties. These services are monitored according to VO rules.	Main VO is managed by single organisation that also manages service creation. The users are issued with clients to access the VO and support is given in this process.

The challenge presented by the science facilities is similar to the issue of business integration, as data from legacy systems must be translated. The main difference is the closed nature of the facilities and the specialized nature of the data management removes the need of traditional businesses models to encourage service use and development. Instead, service development and use needs to be user-driven. This enables service users to represent specific data management translations and can move the integration of data from these facilities away from non-standard forms of integration to a model where centralized translations sets can be stored. Centralized storage & management of translation/integration mechanisms is an emerging area of research [15] and is yet to be applied to the Grid. The BDIFS project applies the concepts of storing and harvesting data translations in the CLF facility, and binds these translations to dynamic Grid Services for use.

3. BDFIS for eScience

STFC facilities are supplied to external scientific users from universities or industry. Users create data and wish to access it in varied ways. As the number of visiting scientists increases so do these creation and access patterns. Similarly, the way users' home systems and data produced are integrated varies widely. The approach follows a common business model where software linking facilities to applications (e.g. visualization software) is done differently from one laboratory to another. This creates problems for integrating a steadily changing stream of new partners, which often want to use data in different ways.

When analysed, the process of data usage from science facilities consists of three elements: the production of the data from the facility, its translation to the required format, and the integration into the target application the user wishes to use (see Figure 1).

Figure 1: The Data Integration Scenario

Whilst it has often been commonplace to interpret this process from facility to facility in isolation, or in some cases on a user-by-user basis, there are important advantages in pursuing a single, standardized method to achieve this integration pattern.

Within the business community the problem has often been addressed by using Enterprise Application Integration (EAI) servers. However, the BDIFS project has already illustrated that for small businesses this approach is too complex and costly. For example, using EAI for the eScience facilities would involve the development of a complex server facility to aid integration. This process would satisfy the operational requirements within the originating facilities, but transfer to other facilities may not be possible; the process could soon be left working in isolation and become a black box.

The solution that BDIFS presented to SMEs is to break up the process in Figure 1 into a Service-Oriented Architecture (SOA). The instant appeal for SMEs lies in that the complexity of the translations in the integration could be hosted remotely and supplied by specialists on a pay-per-use basis. For the eScience facility the SOA approach is attractive as it can be used as an easy to design, integrate with and interoperable method for collecting resources to aid user integration. Individual integration steps will be achieved using specific services that can be documented and stored in accessible ways, thus allowing the expertise built to be easily adopted and used by other facilities. This service composition centres on the formation of a computing environment consisting of the services the user needs. In Grid computing terms this is a Virtual Organisation (VO), which is formed to perform the translations as illustrated in Figure 2.

Figure 2: The SOA Integration Approach

The approach in Figure 2 includes a step where the appropriate translation is chosen from a bank of existing data translations. This enables the integration process to be broken down. A laboratory providing several facilities may invest in developing translation services to a specific service model and store them as central resources for use by all facilities, leveraging the experience of developing previous integration solutions for users.

4. Implementation

The BDIFS design is based on Grid Service WSRF standards implemented using the Globus Toolkit Version 4 [16]. The Virtual Organisations are controlled by a central BDIFS *BaseVO* using a Dynamic VO manager to create temporary VO instances, and populating them with the required services to match the visiting scientist's needs. This approach to creating VOs is used in the existing BDIFS implementation for business and has been tuned for hosting by the Akogrimo project where VOs are formed to manage groups of actors.

The translation mapping and service repository is largely created using PHP, MySQL and JavaScript. These are combined with Globus Toolkit version 4 to implement the Grid services, to enable the deployment of new translations as services. The process of the deployment of a translation as a service instantiation can be seen in Figure 3.

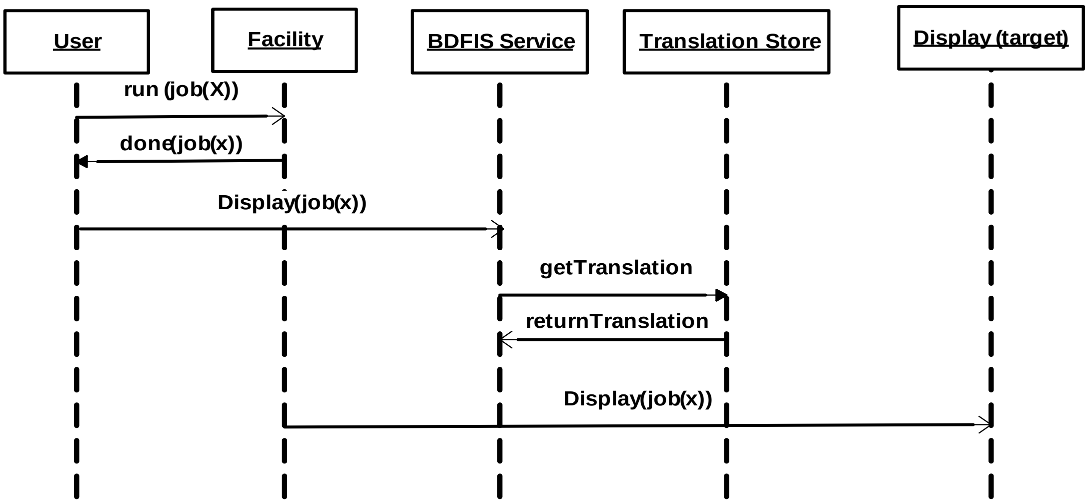

Figure 3: Deploying a Translation as a Service

The data in the Translation Store is a new innovation for BDIFS and is directly linked to the application of BDFIS at CLF in 2007. This is the Translation Mapping Service (TMS), and is a tool to aid users to create the translations between the data formats their services use to perform the data integration of Figure 3. At the centre of the TMS is a graphical user interface which presents users with a form where they can map legacy data onto a desired target data format. For example, in an environment supporting a SME, this can be used to map legacy purchase order data to an ebXML [17] format purchase order. Once this mapping is complete, the saved translation is indexed within the Translation Store. An example mapping that was used in the CLF translation is illustrated in Figures 4, 5 and 6. This takes data from a raw format as generated by the instrument to a standard exchange format known as NEXUS, which is widely used in scientific facilities.

In the example of Figures 4-6, the three data formats are input into the GUI by the user. When submitted, the BDIFS framework calculates the exact points of each document where marked variables occur, by detecting the point at which the first variable occurs and counting back to capture the text before it. With this entry point now available, BDIFS can determine where the first value starts in the transformation. The end points are calculated in the same way, and a series of stable reference points are built up in the document.

```
FNAME:20061109GL0000003N_PUMP_RIG.dat;SECTION:LA3\N_PUMP\RIGOWSKI
DATE:20061109;TIME:13:59:17.0
TIMES:50357
SHOTNUM:3
DIM:1
ARRAY:1000
DATASIZE:SNG;FORMAT:1D traces
CTSTIME:OFF
PARENT:
CHANS:5;CNAMES:RIG_3
RIG_6
RIG_9
RIG_12
RIG_15
DATASTATE:0;
EOH]
```

Figure 4: Original Data Format

```
FNAME:<VALUE1>;SECTION:<VALUE2>
DATE:<VALUE3>;TIME:<VALUE4>
TIMES:<VALUE5>
SHOTNUM:<VALUE6>
DIM:<VALUE7>
ARRAY:<VALUE8>
DATASIZE:<VALUE9>;FORMAT:<VALUE10>
CTSTIME:<VALUE11>
PARENT:<VALUE12>
CHANS:<VALUE13>;CNAMES:<VALUE14>
DATASTATE: <VALUE15>
E0H]<VALUE16>
```

Figure 5: Original Data Format with Values Marked in Regular Format

```
<NXroot NeXus_version="3.9.0" XML_version="mxml"
file_name="performance-xrnl.nxs" file_time="">
 <NXentry name="gemini">
  <CNXdaily name="daily">
   <fname>value1/</fname>
   <section>value2/</section>
   <date>value3/;TIME:value4/</date>
   <time>value5/</time>
   <shotnum>value6/</shotnum>
   <dim>value7/</dim>
   <datasize>value9/</datasize>
   <format>value10/</format>
   <cstime>value11/</cstime>
   <chans>value13/</chans>
   <datastate>value15/</data state>
   <NXdata name=value1/>
    <NXdata>
    value16/
    </NXdata>
```

Figure 6: Final Data Mapping, based on Desired Format with Appropriate Values Included

Collections of similar information make up the mapping for that specific translation and are stored within the Translation Store. The method allows users to add to the store and increase its technical content without having to write any programming code. When a translation request requiring a translation mapping to be selected from the repository is made, the Grid Service Framework provides the generic services within the Dynamic VO that is formed (see Figure 3). Furthermore, users coming to BDIFS that may not be the original creators of the translation, still have the same service presented to them for use. This approach allows the efforts of the service designers in both the eScience and eBusiness communities to build up a knowledge base that can be shared with wider organisations.

5. Business Case

The method of service creation describe above is ideal for temporary integration in businesses, such as when a business acquires another company and needs short-term data integration between systems. This model is also use to SMEs which have to integrate with multiple larger partners. Research in North Wales has illustrated that the quality of partner and contract for low tech outsourced jobs such as packing can be increased if integration can be achieved [13]. This is important as often the contracts last during seasonal periods of 2-3 months and then a new partner is needed.

Within eScience, the dynamic Virtual Organisation creation has the potential to allow visiting scientists to experiment and create various output formats of data from science facilities for analysis on- or off-site in short time periods. This is a radical improvement on traditional manual based ways of giving scientists piecemeal access and the non-scalable design of specific integration environments when a new set of experiments are conducted.

When combined with dynamic workflow compositions, the mix of translation services can be presented in complex, rule-based translation scenarios that are conducted depending on the values presented in the data. A GUI to allow users to dictate rules and create workflows in the same simple manner of the services is currently being implemented for BDIFS. This type of composition would allow users to practically dictate the design and behaviour of the Service-Oriented Architecture implemented using Grid services.

The above demonstrate a dual business case for the BDIFS project. The first side of that refers to the original eBusiness integration case for SMEs. There, the project's aims at creating an eBusiness integration solution that SMEs with little IT knowledge can use and deploy with minimal external help and support. A second aim exists for creating a set of eScience VOs for temporary scientists in the same fashion as the stable National Grid Service type VOs are established in the science community.

6. Future Investigation

The concept of temporary VOs for visiting scientists requires more investigation. Work on BDIFS so far has only looked at the use of such VOs for assisting service sharing for data integration. The combination of these VOs with other services and more application-specific ones as in the example of visualization services will further strengthen the output. Also, security in the system and use of federation with respect to identity management needs formal implementation and testing.

The GUI for creating the translation services is currently web-based and also needs more security control mechanisms, especially for user management. The indexing and referencing of translations needs to be defined in more detail and within larger scale tests, and this would ensure that application saleability is enhanced. The workflow composition needs extensive testing and linking to a mechanism related to defined VO policies.

7. Conclusions

This paper makes the case for user driven service-oriented Grid approach using dynamic virtual organisations for eScience applications, a case which to date has had few practical examples supporting it. BDIFS successfully realises this approach and is currently being tested in real-word case studies for further refinements. Future applications of BDIFS are planned within SMEs in the North Wales area and funding is being sought to establish the University as a permanent host for the SMEs BDIFS infrastructure. The small scale implantation at STFC is currently under investigation with respect to combining it with existing data management infrastructures to make the prototype wider ranging. If the BDIFS solution is not adopted at STFC on a larger scale, the SOA approach as outlined in this paper and the existing work at CLF will provide the best use of resources and a clear way forward for future data integration projects in the laboratory.

References

[1] Business Data Integration Framework for the SME homepage: http://www.bdifs.org *Last accessed 22.04.07*

[2] I. Westarp. F. Weitzel. T. Buxmann. P. König, "The Status Quo and the Future of EDI" *Proceedings of the 1999 European Conference on Information Systems (ECIS'99). Copenhagen, Denmark*

[3] L. Schubert, B. Koller, N. Bieber. "The Requirement for an Interoperability Service: Power-Distribution in an EBusiness World." In *Exploiting the Knowledge Economy: Issues, Applications, Case Studies*, Paul Cunningham and Miriam Cunningham (Eds), IOS Press, 2006 Amsterdam, ISBN: 1-58603-682-3 *Proceedings of eChallenges e-2006, Barcelona.*

[4] Science and Technology Facilities Council homepage: http://www.scitech.ac.uk/Home.aspx *Last accessed 22.04.2007*

[5] UK National Grid Service homepage: http://www.grid-support.ac.uk *Last accessed 22.04.2007*

[6] N. Geddes. "The National Grid Service of the UK." *Second IEEE International Conference on eScience and Grid Computing (eScience'06) Amsterdam.*

[7] A. Akram, D. Chohan, X.D. Wang, D. Meredith, R. Allan, "*STFC Portal Infrastructure to Support Research Facilities*." *Global Grid Forum 14, Chicago, USA, 2005*

[8] EU GRASP project homepage: http://www.eu-grasp.net. *Last accessed 22.04.07*

[9] A. Svirskas, M.D. Wilson, A.E. Arenas, E.C Lupu, P. Giambiagi, N. Tuptuk, D. Chadwick, T. Dimitrakos B.Roberts. "Aspects of Trusted and Secure Business-Oriented VO Management in Service Oriented Architectures." *1st IEEE Workshop on Service Oriented Solutions for Cooperative Organizations Munich*

[10] S. Wesner, T. Dimitrakos, K. Jeffrey. "Akogrimo - The Grid goes Mobile" *ERCIM news no59 2004*

[11] OASIS technical committee page for WSRF: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrf *last accessed 22.04.07*

[12] I. Foster,(eds), J, Frey (eds), S. Graham, (eds), S. Tuecke "Modelling Stateful Resources with Web Services v. 1.1." *Globus Alliance, 2004.*

[13] T.Kirkham, T. Varsamidis "Supporting EBusiness integration in North Wales" In *Exploiting the Knowledge Economy: Issues, Applications, Case Studies*, Paul Cunningham and Miriam Cunningham (Eds), IOS Press, 2006 Amsterdam, ISBN: 1-58603-682-3 *Proceedings of eChallenges e-2006 Barcelona*

[14] T. Kirkham, T. Varsamidis, J. Gallop, B. Matthews, M. Wilson "A Framework to Provide EBusiness Support for African SMEs" In *IST Africa 2007 Conference Proceedings*, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information Management Corporation, 2007, ISBN: 1-905824-04-1

[15] T. Dohzen, M. Pamuk, S.-W. Seong, J. Hammer, M. Stonebraker, "Data integration through transform reuse in the Morpheus project," *ACM SIGMOD International Conference on Management of Data (Demo Track). Chicago*

[16] Globus Toolkit Version 4 homepage: www.globus.org/toolkit/docs/4.0/ *last accessed 22.04.07*

[17] Official OASIS site dedicated to the ebXML standards, includes links and information on standard specifications: www.ebXML.org *last accessed*