B2B Infrastructures in the Process of Drug Discovery and Healthcare

Uwe Radetzki[1], Paul W. Walland, Mike J. Boniface
IT Innovation Centre, University of Southampton
2 Venture Road, Chilworth, Southampton SO16 7NP, UK

Abstract. In this paper we describe a demonstration of an innovative B2B infrastructure which can be used to support collaborations in the pharmaceutical industry to achieve the drug discovery goal. Based on experience gained in a wide range of collaborative projects in the areas of grid technology, semantics and data management we show future work and new topics in B2B infrastructures which arise when considering the use of patient records in the process of drug discovery and in healthcare applications.

Keywords. B2B Infrastructures, Grid, B2B Collaborations, Security, Trust, Data fusion

Introduction

In the last decades, advances in the life science sector have facilitated the rapid acquisition of vast amounts of data in a multitude of projects, which today leads to a huge number of distributed databases and analysis methods [1]. The data is used by the life science community for various in-silico experiments that are by their very nature, hypothesis driven, ad-hoc and highly specialised to the problem with which they are associated. For example, in medicine, sequences provide a basis for the study of susceptibility to disease and the development of new preventative and therapeutic approaches, whereas, in cell biology, the interactions between components of cellular circuitry can be studied.

In the pharmaceutical industry many additional challenges are faced in the development of new drug products. Drug discovery is a complex high risk process with many development phases involving people from various disciplines accessing rich and diverse information sources under the constraints of external factors such as regulations and strong competition. The drug discovery process, in general referred to as a pipeline, includes several distinct phases from the generic research through to clinical trials, project lifecycle management and hopefully the final drug product. Pharmaceutical companies are faced with the challenge of how to adopt new approaches and technologies for the distribution, integration, discovery, visualisation, annotation and validation of information in order to provide decision makers with accurate information and help to reduce companies' costs and risks whilst increasing the probability of successfully designing and developing a new drug.

The EU IST SIMDAT Project [2] is deploying Grid technology to support new business models for carrying out bioinformatics research within the drug discovery

[1] Corresponding author: Uwe Radetzki, tel: +44 23 8076 0834, fax: +44 23 8076 0833, email: ur@it-innovation.soton.ac.uk

pipeline. Innovative solutions to assist this pipeline exist in large pharmaceutical and biotechnology companies as well as in academic institutions. Increasingly, it is the use of B2B infrastructure technology to enable successful partnerships within the pharmaceuticals business that will help achieve the drug discovery goal.
In this paper, we show how an innovative B2B infrastructure drawn from the SIMDAT's Grid solution portfolio can be used to support B2B collaborations in the pharmaceutical industry to achieve the drug discovery goal (section 1). The demonstration will include workflows, distributed data access techniques, semantic web and several Grid infrastructure components. We further discuss new challenges for future B2B infrastructures arising when including patient data in the process of drug discovery and personalised healthcare (section 2).

1. Demonstrating B2B infrastructures for the pharmaceutical industry

The focus of our demonstration is based on a scenario drawn from the SIMDAT project using bioinformatics in the collaborative development of new drugs through Grid-enabled partnerships during the target identification phase of the pipeline. This scenario shows the potential of business-to-business (B2B) and business-to-academic (B2A) collaborations within the pharmaceutical industry. In the specific business case, we will demonstrate how a B2B infrastructure can support drug discovery partnerships between big pharmaceutical (GlaxoSmithKline) and SME biotechnology suppliers (Inpharmatica) in the analysis and annotation of selected sets of proteins. The demonstration scenario comprises the following elements:

1. Inpharmatica publishes a service to provide aggregate and detailed annotation of protein data including similarity based on structure, ligand binding sites and sequence annotation in various predetermined formats.
2. GlaxoSmithKline subscribes to, and requests results using the service.
3. Each request for the Grid services identifies a set of data that requires additional annotation and the required format for the resulting annotation.
4. Inpharmatica delivers the resulting annotation to complete the transaction.

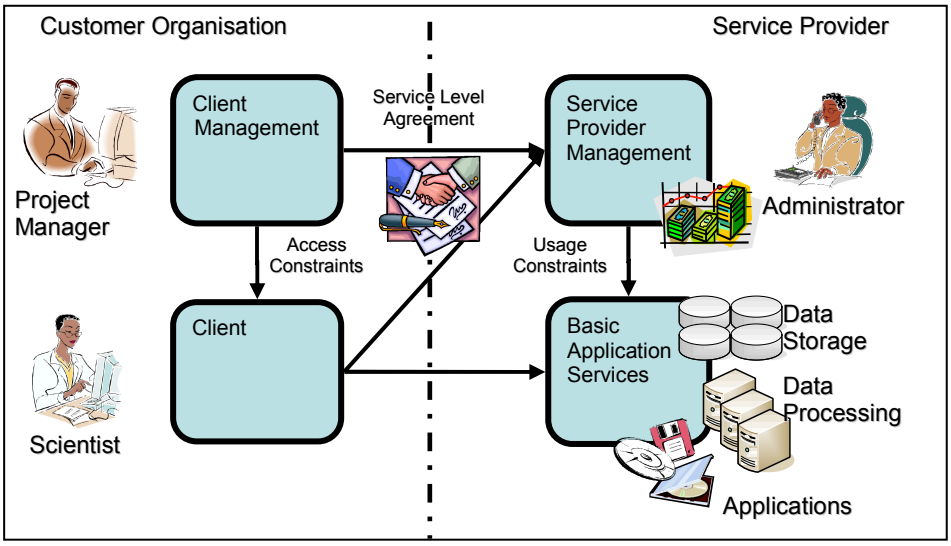

Fig. 1: Managing B2B collaborations

While this scenario is typical for collaborations in the pharmaceutical sector, it also unveils requirements for the underlying B2B infrastructure supporting cost effective dynamic collaborations. This includes business relationship management, minimal authorisation and administration efforts, explicit trust decision points, and industrial operational security policies (ISO/IEC 17799:2005). Dynamic collaborations allow organizations to participate in distributed and managed scientific processes whilst still enforcing commercial level security policies for the protection of intellectual property rights and business objectives. Figure 1 illustrates the workflow of establishing and managing B2B collaborations. These collaborations are underpinned by the GRIA B2B infrastructure [3], a secure and interoperable Open-Source Grid middleware. We will demonstrate how the infrastructure supports the following more general requirements for B2B collaboration.

Dynamic Trust and Security

Trust and Security are major issues in establishing business relationships between companies in the drug discovery pipeline. In GRIA we have developed security mechanisms based on Service-Oriented Architecture (SOA). GRIA has dynamic policy management features to define and enforce policies for accessing service and consumer management functions, and for each individual data item and computation. This dynamic policy feature enables users to orchestrate interactions between services operated by totally independent providers, and to share results with other users as required. The GRIA 5 consumer management package also supports integration with 'home site' user authentication, and service policies can specify which site(s) are trusted to authenticate users against each individual policy clause. This makes it very easy to set up business relationships and start providing services to new customers within a few minutes.

On top of these security mechanisms we demonstrate an End-to-End (E2E) Security framework developed by NEC [6]. E2E Security supports the provisioning of security

services, including encryption, integrity protection, and authentication, from end-to-end even in the presence of intermediate systems like proxies or message queuing systems.

Service Level Agreements and Quality of Service

In B2B collaborations, service providers and customers trade resources (applications, data, processing, storage) under the terms of bilateral Service Level Agreements (SLAs). An SLA describes quality of service (QoS) and are used as contracts between service providers and service consumers (customers), working out risks (in terms of costs) to providers and working out value (i.e. price) to customers. An SLA gives a promise to provide services, for instance to initiate a specific processing step on protein data. We will demonstrate how to create and negotiate SLAs to regulate the use of resources in our business scenario.

Usage Control and Accounting

Service providers deploy application services appropriate to their business operation. These services generate usage reports using their own QoS criteria which may be qualitative (e.g. error conditions) or quantitative (e.g. processing time, data transferred). The metrics are flexible and depending upon the application could be CPU time, number of active application licenses, data storage, number of submitted jobs, etc. GRIA uses these reports to monitor customer usage and the level of commitments from existing agreements compared with available capacity. A specific management service records and constrains the customer usage against the service provider capacity and the terms of agreedSLAs.

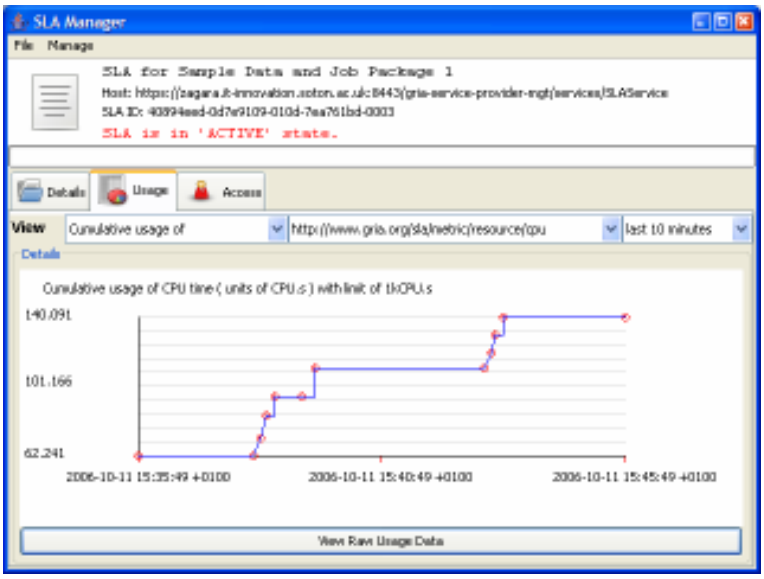

Fig 2: Usage report showing number of CPU seconds

Adaptive Workflows

Defining and enacting distributed workflows is one of the key drivers for grid computing and Service-Oriented Architectures. In our demonstration we use Taverna [4] or InforSense KDE [5] to construct and enact the workflow required for our business case. We will also show how novel service discovery and service registries can be used to adapt workflows on the fly to fulfil business requirements. An adaptive workflow is not fixed to specific service instances but rather discovers and selects services based on business constraints, which could be described for instance by SLAs and current usage..

2. From bioinformatics to healthcare: Issues for future B2B infrastructures

A great deal of experience has been gained from work in GRID, SOA and data management projects, both in the health-care sector and others. This experience has highlighted many of the issues that arise when dealing with the management of health-care records and the handling of personal data. A range of techniques, such as those of data fusion (SANY), media and knowledge handling (ARTEMIS, E-CHASE), secure middleware (GEMSS, GRIA, SIMDAT, NEXTGRID), creation of ontologies (GEMSS, ARTEMIS) and handling of heterogeneous data sources (E-CHASE, SANY) are common across many disciplines. As an integrated bioinformatics and healthcare environment develops, so these areas of expertise will need to be brought to bear on the solution of fundamental challenges which arise.

The following brief discussion gives some specific examples of the new challenges being faced in the developing healthcare environment:

- Data Fusion

  Data is being created from a wide variety of sources and sensors, and as citizens start to live in a world increasingly populated with ubiquitous sensors monitoring their lifestyles and biometrics, so there will be an increasing need to fuse heterogeneous data, interpolate and predict data, use and apply models to extract knowledge about the health of individuals and populations, and to predict, for example, epidemiological developments.

- Heterogeneous data sources

  The collection of data is from a huge variety of sources, which can be historical data held in databases through to live measurements such as blood-pressure and temperature or media based records of x-ray or scanner results. Additionally, this data may be stored or delivered using one of a number of protocols or formats, with different structures or ontologies to support them. Future healthcare systems will need to be able to handle and integrate this wide variety of data.

- Management of data sources (patient information, images and media)

  Health data is increasingly coming from a wide variety of sources, and this range will increase in the future. Image data (such as scans, digitised X-rays, in-vitro video monitoring) are created, stored and delivered in different formats and with markedly different levels of descriptive metadata. Future healthcare data management will need to take account of these different formats, and manage the creation of metadata so as to improve the ability of data systems to search and retrieve relevant information. Global ontologies will be essential to allow healthcare professionals from a range of disciplines to access all the data relevant to their needs.

- Integrated middleware

  Building on the experience gained by implementers in academic, scientific and business applications of distributed computing and grid-based systems, the particular needs of future healthcare systems can be addressed in a holistic fashion to ensure that interoperability, security, managed exchange of data and integration of dynamic acquired information (such as personal health monitoring systems) is built-in from the ground up. Service oriented architectures (SOA) and infrastructures (SOI) will need to be developed in which workflows connect all the functional components seamlessly, whilst maintaining data integrity and ease of use for the healthcare professional.

- Personal monitoring and reporting

Future healthcare systems will encompass preventative medicine based on personal monitoring systems. Intelligent systems are needed which will correlate data and patient history, and create alerts and warnings based on data fusion and modelling techniques.

- Security and protected access to data

  There are a number of levels of security and trust raised in the healthcare scenario, which are similar to those approached in other situations. Privacy issues arise in the use, sharing and access of health records, some of which are currently controlled by legislation which is often different in different countries. Health records maintained by different agencies or local bodies will often be held on incompatible systems, and may well be subject to differing local, national and international legislation, as well as personal preferences expressed by the person to whom the data relates. The means to manage these many levels of security and access need to be incorporated into future integrated healthcare systems, and cannot be implemented as an add-on function.

- Virtual Physiological Human

  Creation of the Virtual Physiological Human (VPH) is an opportunity to integrate what are currently localised models of physiological functions into an increasingly complex model. As this model becomes more complete, so it will provide the opportunity for achieving an ever-wider range of functions:-

- from nano to macro
  investigation of behaviour at the molecular level, the cellular level, the organism level and the level of populations
- in-silico drug efficacy
  integrated studies of drug function and biological impact
- training
  virtual training of physicians and other healthcare workers on accurate simulations of the physiology of patients
- integrated physiological models
  comprehensive modelling of the interactions between physiological processes

3. Conclusion

In this paper we describe a demonstration based upon a real situation that has been investigated in the European SIMDAT project. The role of users in the pharmaceutical sector is crucial in identifying the needs of the business-to-business (B2B) and business-to-academic (B2A) communities. The demonstration shows the potential of distributed workflows in grid computing and service oriented architectures in fulfilling those needs. Vital elements, such as metrics and Service Level Agreements are shown operating as an integral part of that workflow. The dynamic collaborations which are formed are supported by minimal administration requirements, yet provided with industrial-strength security policies, allowing high levels of trust to be maintained

between collaborating entities. It is also shown that SLA based business partnerships can allow quality of service to be managed effectively.
We conclude with a discussion of the future requirements of distributed workflows in the healthcare sector, taking into account the future needs of ubiquitous sensors, integrated healthcare provision and the need for maintaining security and privacy. The future needs of implementing the ambitions Virtual Physiological Human project are also briefly discussed. Based on work performed in related grid-based projects it is shown that there is much original work still to be done, but that it can be seen in the context of a continuation of knowledge and expertise gained through the application of grid architecture and SOA's to practical, user driven scenarios.

References

[1] M. Y. Galperin: The Molecular Biology Database Collection: 2006 update, Nucleic Acids Research, Vol.34, Database issue (2006).
[2] EU IST SIMDAT Project, http://www.simdat.org
[3] Grid Resources for Industrial Applications (GRIA), http://www.gria.org/
[4] Taverna workbench, http://taverna.sourceforge.net/
[5] InforSense KDE (Knowledge Discovery Environment), http://www.inforsense.com/
[6] NEC, End-to-End Security