

Computing and Informatics, Vol. 21, 2002, 383–398

## LINKING HETEROGENEOUS BIODIVERSITY INFORMATION SYSTEMS ON THE GRID: THE GRAB PROTOTYPE

Andrew C. JONES, W. Alex GRAY,  
Jonathan P. GIDDY, Nick J. FIDDIAN

*Department of Computer Science, Cardiff University*  
*PO BOX 916 Cardiff, CF24 3XF, UK*  
*e-mail: andrew, alex, jonathan, nick@cs.cardiff.ac.uk*

Revised manuscript received 11 November 2002

**Abstract.** In the field of biodiversity informatics a wide range of diverse databases and tools already exists. The challenge is to integrate such resources in order to support scientists wishing to explore complex problems of relevance to biodiversity, and to create new resources where necessary. In this paper we outline the relevance of biodiversity informatics requirements to the future development of the GRID, identifying the main issues that need to be addressed in this area. We present GRAB (GRid And Biodiversity), which is a prototype demonstrator illustrating how one particular biodiversity-related task, namely bioclimatic modelling, can be supported in a Globus-based environment. We also describe a much larger-scale GRID application project that is just commencing (BiodiversityWorld) in which a flexible problem-solving environment is to be built for full-scale investigations by scientists working in a number of biodiversity research areas.

**Keywords:** Biodiversity informatics, heterogeneous databases, data integration, collaborative working, Globus

### 1 INTRODUCTION

The GRID [5] promises to offer major benefits to various sectors of e-Science. In this paper we concentrate on one specific sector of e-Science, namely *biodiversity informatics*. Research in this area often requires integration of data of quite different kinds in novel ways. This paper identifies some of these applications and discusses

the relevance of GRID technology in supporting such applications. In particular, we present the GRAB (GRid And Biodiversity) prototype. The background to GRAB is that the UK Department of Trade and Industry provided funding for a number of short GRID demonstrator projects, including GRAB, in order to explore the relevance of GRID concepts in various application domains, using existing GRID software such as Globus [4] and SRB (Storage Resource Broker<sup>1</sup>).

It should be emphasised that GRAB has been implemented as a proof-of-concept, rather than as a full system providing flexible support for biodiversity informatics. Also, although some biodiversity informatics tasks can benefit from High Performance Computing, the current GRAB prototype does not attempt to do this. Speed of retrieval by the data sources used is the limiting factor for performance at present. Our interest has been specifically in how diverse biodiversity resources can be used on the GRID to solve problems of interest to researchers in the field of Biodiversity. But we are currently commencing a new, larger-scale project (BiodiversityWorld) in which we are planning to build a full prototype, sufficiently complete, and harnessing sufficient computing power, for scientific research on a number of key biodiversity topics to be carried out over the GRID.

In the remainder of this paper we first provide a background to biodiversity informatics, describing some of the major issues that are of current interest and identifying some of the tools and data that are needed in this research area; then we focus on the relevance of the GRID to biodiversity informatics. Next, we present our GRAB prototype, describing how it is used and how it has been implemented. We then discuss some of the lessons learned, including the limitations of current GRID technology, and outline our planned future work in the BiodiversityWorld project.

## 2 BACKGROUND

### 2.1 Biodiversity Informatics

*Biodiversity informatics* is concerned with organising and processing knowledge about living things and covers a range of areas including the provision of tools to aid in studying organisms (e.g. collection and analysis of data; classification) and their relationships to each other and to their environment (e.g. organising species of organisms into a taxonomic hierarchy; assessing biodiversity richness). Biodiversity informatics is clearly closely related to bioinformatics, but it is often regarded as a distinct discipline from bioinformatics, rather than as one of its sub-disciplines.

It is sometimes possible to express questions relating to biodiversity quite simply, but answering them is typically a complex process. Four such questions, and some of the relevant techniques, data sources and tools, are as follows.

1. *How should conservation efforts be concentrated?* One major factor influencing decisions regarding conservation is that of *biodiversity richness*. A geographical

---

<sup>1</sup> <http://www.npaci.edu/DICE/SRB/>

region that exhibits high biodiversity richness may be of particular interest for conservation. *Biodiversity richness analysis* involves assessing the biodiversity richness of a region according to some given metrics, which might include *species richness* (number of species present) and *range-size rarity* (a weighted metric favouring rarer species). To carry out this kind of analysis, information about species geography and/or distribution of individual specimens is required, as well as some information regarding boundaries of geographical and political units. An analytical tool that can reason with this data is also required — an example being WorldMap.<sup>2</sup>

2. *Where might a given species be expected to occur, under present or predicted climatic conditions?* A task that can contribute to an answer to this question is *bioclimatic modelling*, in which *climate surfaces* are generated so that regions having climates that are similar, judged by appropriate criteria, can be identified. Given the distribution of a particular species, it can then be predicted, for example, whether there are other regions in the world where the climatic conditions are suitable for introduction of that species. For this analysis to be carried out, information on climate is needed in addition to species information, and some analytical tool is again needed — an example in this case is BIOCLIM [2].
3. *Is geography a good predictor of relationship between species lineages?* To answer this kind of question, *phylogenetic analysis* and *biogeography* techniques need to be employed by taxonomists, who specialise in classification of living things. Species are grouped according to similarity in a *phylogenetic tree*: to do this, data on individual species is used, including DNA sequences. Species closely related at a molecular (DNA) level will have some common parent node towards the bottom of such a tree. In order to generate the tree, tools such as MacClade<sup>3</sup> can be used, but the biogeographical analysis needs to be done manually at present.
4. *What useful properties might a particular newly-discovered species have?* Species that are closely related to each other sometimes share useful properties, e.g. for medicinal purposes. Taxonomy has a predictive property, in that once organisms have been classified appropriately, by consideration of known features regarded as important for discrimination, information about a well-studied organism may be used to predict additional properties about a less well-studied organism. Clearly databases containing such information must therefore be available, both for the classification process and for subsequent investigation. In order to support analysis using descriptive data and existing classifications, tools have been or are being developed in projects such as ReTAX [1] and PROMETHEUS [11].

We have thus identified a number of very different kinds of data sources, including *Species Information Systems* (SISs), which might provide information about species

---

<sup>2</sup> <http://www.nhm.ac.uk/science/projects/worldmap/>

<sup>3</sup> <http://phylogeny.arizona.edu/macclade/macclade.html>

geography, specimen distribution or descriptive data; *geographical databases*, providing information about boundaries of geographical and political units, or climate surfaces, etc., and *gene sequence databases*. We have also identified a number of distinct kinds of analytic tools. But central to such activities is the provision of a reliable species catalogue — sometimes referred to as a catalogue of life. The fundamental component of a species catalogue is a list of scientific names, synonyms and associated information for the species it covers — a taxonomic checklist. A stable system of groupings of organisms (normally referred to as *taxa*, of which *species* are an example) and of names for these groupings is essential if scientific questions of relevance to biodiversity are to be answered accurately. For example, resources such as gene sequence databases (e.g. the EMBL Nucleotide Sequence Database<sup>4</sup>) and bioclimatic modelling tools (e.g. BIOCLIM) can only be used effectively to explore phenomena associated with individual species if the data is organised in a taxonomically consistent manner. The biological significance of a checklist is that its contents are subject to expert scrutiny, and each species has an *accepted name*, but also has alternative *synonyms* which allow species to be located under other names, be it in a Species Information System (SIS) providing geographical information about the regions inhabited by individual species, or in other resources like those mentioned above.

The names in a checklist reflect a scientific classification of organisms, and the quality of this checklist is of great importance. In the LITCHI project we developed techniques for analysis of taxonomic checklists for consistency, and for postulating taxonomic reasons for conflicts identified [9, 3]. A comprehensive catalogue is needed if it is to be generally useful, but this is difficult to achieve because of the distributed nature of taxonomic expertise. In the SPICE project [10, 12] we have developed and implemented techniques for co-ordinating a federated catalogue of life, comprising globally distributed, heterogeneous databases covering individual sectors of a taxonomic hierarchy.

So a wide range of data resources and analytic tools already exists and, in particular, techniques that can help provide a comprehensive, consistent catalogue of life have already been developed. Nevertheless, it is currently difficult to bring together these heterogeneous resources in flexible ways to address new problems within the field of biodiversity informatics. It is to the need to be able to do this that we now turn.

## 2.2 The GRID: Relevance to Biodiversity Informatics

The GRID [5, 7] is being developed to improve exploitation of distributed resources — particularly for applications requiring high performance computing, e.g. high energy physics. Within the field of bioinformatics, some applications require high performance computing, and initiatives such as the BioGRID work package

---

<sup>4</sup> <http://www.ebi.ac.uk/embl/index.html>

of Eurogrid<sup>5</sup> aim to support such work. Indeed, the WhyWhere project<sup>6</sup> supports the specific task of global prediction of biodiversity-related phenomena. But the main use of GRID technology within WhyWhere appears to be to co-ordinate high performance computing resources for certain specific computationally intensive tasks. In contrast, many of the biodiversity informatics application areas do not require high performance computing, but support for distributed systems — including high performance networking — and collaborative working, in which tools and data can be brought together to solve new problems, is desirable. To achieve this, resources must be described using appropriate metadata and heterogeneity must be accommodated — including, e.g., heterogeneity of data formats and differences in the schemes according to which knowledge has been organised and named. Appropriate metadata can be used in order to determine how to access and use such resources. Even with this information, it is still necessary to provide lower-level mechanisms for wrapping legacy data and legacy software so they can be used within the GRID. In addition to these facilities we need tools to find appropriate resources and synthesise new knowledge. Another issue that must be addressed is selective access to data: there is frequently good reason to maintain tight control over which parts of a data set a given user has access to. To achieve this, a system that associates particular credentials with individual users is desirable.

All the above issues are relevant to e-Science in general — for example, the Triana<sup>7</sup> system now provides a GRID-enabled problem-solving environment partially tailored to high-energy physics, in which workflows can be constructed specifying how operations are to be sequenced. Moreover, most of these issues are addressed, at least to some extent, in the Globus toolkit. Yet biodiversity data is characterised by a higher degree of structural and semantic complexity than is often the case elsewhere, and complex data integration processes are often needed to synthesise useful results. Accordingly, the GRID offers a useful environment for biodiversity informatics, but biodiversity informatics presents challenges to the current GRID technology such as Globus in terms of the metadata richness and the flexibility of access that are needed.

### 3 THE GRAB PROTOTYPE

#### 3.1 Overview

As explained in Section 1, GRAB was built as a proof-of-concept prototype, rather than as a full prototype of a biodiversity informatics problem-solving environment. In order to constrain the project to a manageable scope, we concentrated specifically on bioclimatic modelling, and only on certain aspects of this task. A fixed sequence

---

<sup>5</sup> <http://www.eurogrid.org/wp1.html>

<sup>6</sup> [http://biodi.sdsc.edu/ww\\_home.html](http://biodi.sdsc.edu/ww_home.html)

<sup>7</sup> <http://www.triana.co.uk/>

of operations is supported, via a Web front-end to a server capable of communicating through Globus with the resources used. These resources are:

- the SPICE catalogue of life;
- two SISs — the International Legume Database & Information Service (ILDIS)<sup>8</sup> and Fishbase<sup>9</sup> — which act as sources for images and geographical information about individual species, and
- a database we have built from public domain climate information obtained from the US National Climate Data Centre (NCDC)<sup>10</sup>.

A limited amount of use is made of metadata in order to enable appropriate choice of SIS, as we shall see in Section 3.3, but the metadata used is not intended to be fully representative of what would be employed in a large-scale system of this sort.

### 3.2 Typical GRAB Session

A GRAB session essentially entails finding climate data pertaining to a species of interest, developing a ‘climate envelope’ from this climate data, and then finding species that are known to exist in regions falling within this climate envelope. A typical GRAB display is shown in Figure 1: it comprises the main GRAB page and a monitor that indicates which resource the GRAB server is accessing at any given time (if any). In more detail, the current fixed sequence of operations is as follows:

1. The user enters a search string and all matching scientific names in the SPICE catalogue are retrieved and displayed. Alongside the names retrieved that are synonyms, the accepted name is also displayed. In Figure 1 the result of searching for ‘*Faba f\**’ is shown. There is only one matching scientific name, *Faba faba*, and it is a synonym of *Vicia faba*. It is this latter name that will be used in subsequent searches.
2. The user selects an accepted name from this list, then full information on this species is retrieved from the appropriate SIS and displayed, including geographical distribution and images. Figure 1 illustrates that immediately after selecting an accepted name, the appropriate SIS is contacted; Figure 2 illustrates the results for *Vicia faba*.
3. The user selects a region from the geographical distribution, then associated climate data is retrieved from the NCDC database and displayed. In this case we have selected ‘Iceland’; Figure 2 illustrates that the climate database has been consulted, while Figure 3 shows the data for climate survey stations in Iceland.

---

<sup>8</sup> <http://www.ildis.org/>

<sup>9</sup> <http://www.fishbase.org/>

<sup>10</sup> <http://lwf.ncdc.noaa.gov/oa/documentlibrary/pdf/climatesoftheworld.pdf>

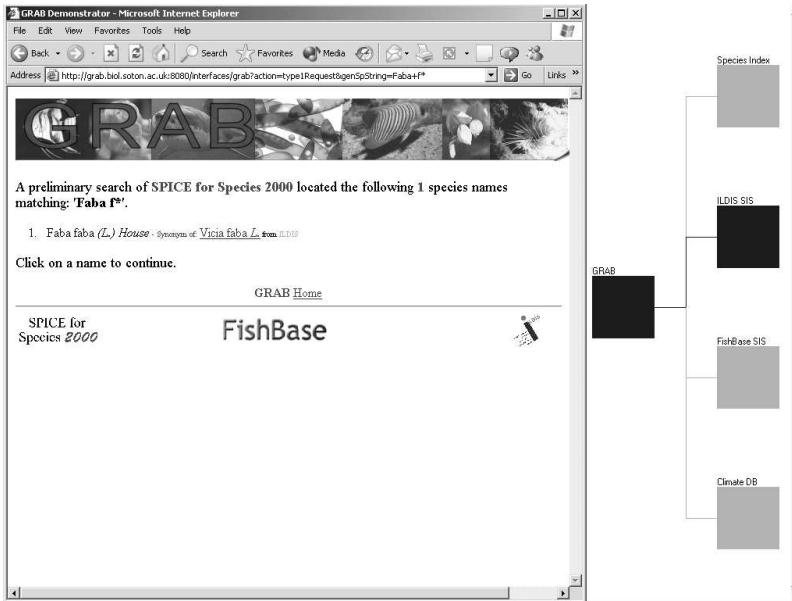


Fig. 1. Catalogue of life synonym display

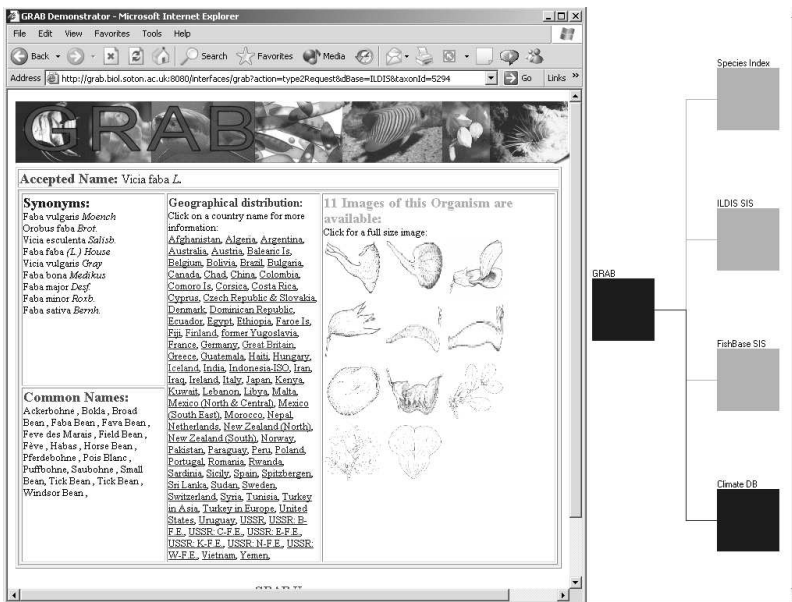


Fig. 2. Species information for *Vicia faba*

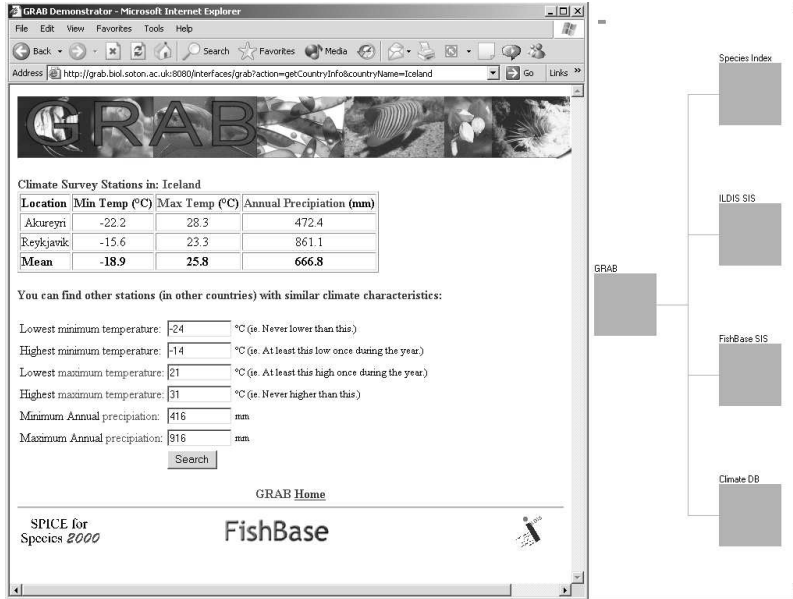


Fig. 3. Data retrieved pertaining to climate survey stations in Iceland

4. Also illustrated in Figure 3 is the climate envelope generated by GRAB from analysis of the survey station data. At this point the user can make use of the climate envelope obtained, or modify it to find other regions with similar climatic characteristics. In this case the highest minimum temperature has been raised to 0°C prior to obtaining a list of stations whose climate variation falls within this revised envelope, as illustrated in Figure 4.
5. The user selects a region from the new list of regions displayed, and species native to that region are retrieved from the SISs. Figure 5 illustrates the result of retrieving species native to the UK.
6. The sequence can be repeated from step 3, if desired, by selecting one of the native species displayed.

Thus we have been able to find species native to some region having similar climate to at least one of the regions the originally-submitted species was native to (step 5), and to retrieve information relating to the originally-submitted species (step 2) and the additional species found (step 6). By this means it is possible for a biologist to consider, for example, whether a particular species of plant could be introduced successfully and beneficially into a new geographical region.



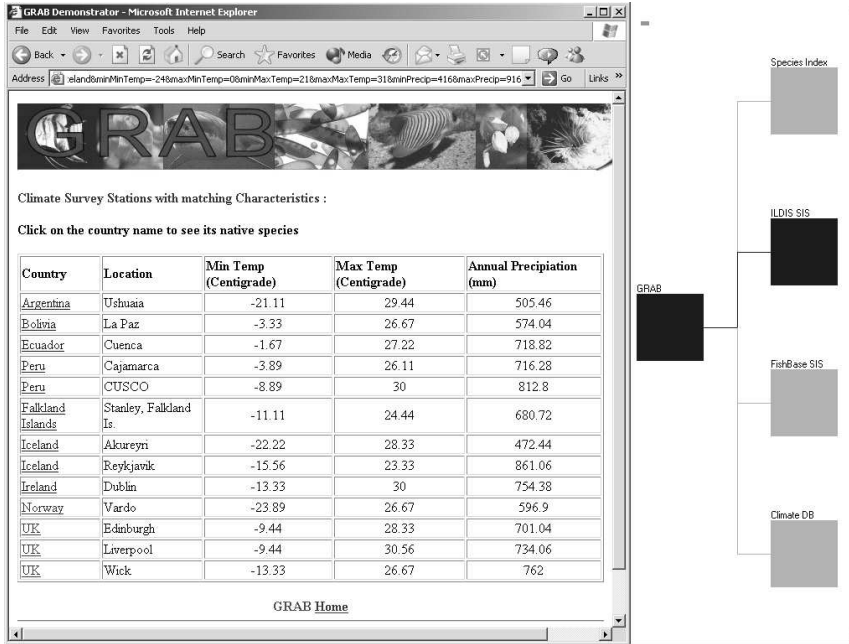


Fig. 4. Climate survey stations falling within specified climate envelope

### 3.3 Implementation of GRAB

The strategy adopted in the implementation of GRAB was initially to build a system in which the GRAB server communicated with the GRAB resources using CGI requests, receiving results back as XML data, and subsequently to integrate these components into a Globus environment. The reasons for adopting this divide-and-conquer approach were that (a) all the data resources used needed some preparation so that they could supply data in XML format, and (b) the programmer employed for the main part of the project had no prior Globus knowledge, and so it seemed sensible for him to seek to acquire this knowledge as a background task while working on other aspects of the project. Creation of the climate database was a lengthy undertaking: the database had to be populated and the CGI/XML interface had to be built. The other components of the system already existed, and so the main task was simply to build a suitable front-end to each of them.

The pre-Globus version of GRAB was indistinguishable in perceived behaviour from the Globus-enabled version, except that the monitor had not been implemented at that point. The architecture of the new Globus-enabled version is illustrated in Figure 6. Implementing the Globus version of GRAB entailed using the Globus Resource Allocation & Management (GRAM) service to invoke tasks on the machines hosting the GRAB databases, passing parameters to these tasks that were previously

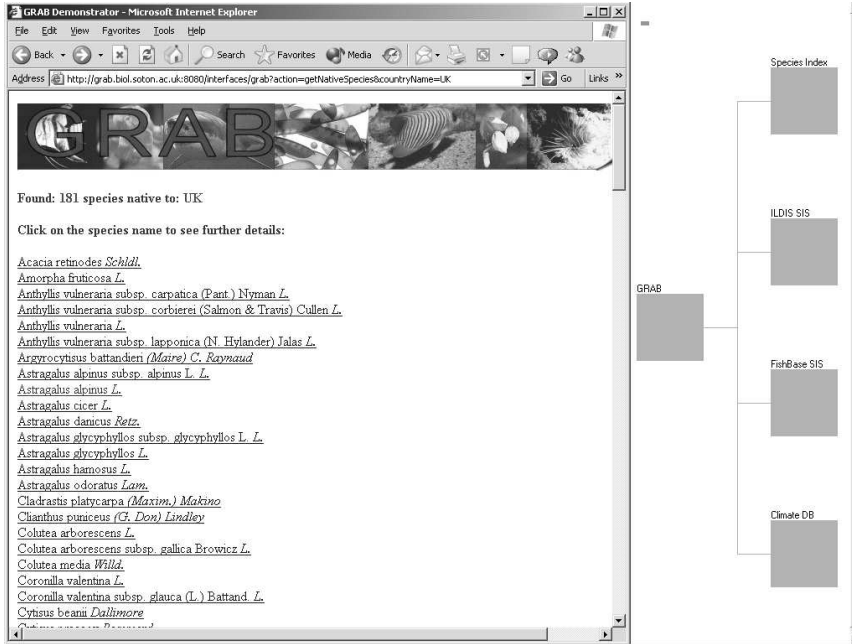


Fig. 5. Species native to UK

passed to the CGI routines. The Globus Access to Secondary Storage (GASS) facility is used to communicate (temporary) result files, which contain the required data formatted as XML documents. For the fixed sequence currently implemented, a fairly small number of operations are needed for execution as remote jobs:

- Retrieve from Catalogue the taxa (species) matching the search string passed as a parameter;
- Retrieve from Catalogue the ‘standard data’ (accepted name, synonyms, etc.) for a named species;
- Retrieve from SIS the image URLs and geography data for a named species;
- Retrieve from climate database the survey station data for a specified country, and
- Retrieve from climate database the survey station data for stations whose climate parameters lie within a specified ‘envelope’.

We have completed migration of the SISs and climate databases to Globus-based access; migration of the Catalogue of life is currently underway: at present the Catalogue operations listed above are still invoked using CGI. The XML response format will remain unchanged, however.

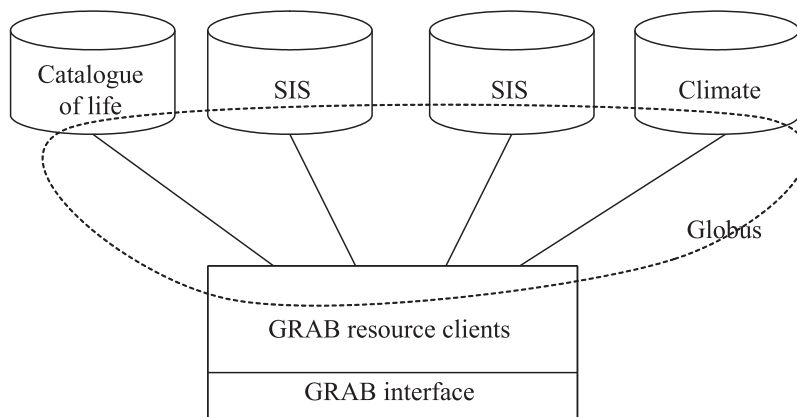


Fig. 6. GRAB architecture

We exploit two additional features of Globus, namely the Globus Security Infrastructure (GSI) and the Metacomputing Directory Service (MDS), but this is not perceived by the end-user. Nevertheless, these two features play roles that are potentially very important. GSI allows single sign-on to all the resources used across the environment, employing certificates for authentication. At the moment it is the System Manager who sets up this authentication: once this has been done, anyone can access GRAB via the Web front end. (Clearly an important development would be for each individual user to have to be authenticated, but this is a task we have yet to do.) MDS is used for locating and accessing a suitable SIS when further species details are requested (step 2 in the sequence of operations given in Section 3.2). Extracts from the MDS information for the two SISs are given in Figure 7. Using metadata in this way in the present prototype is perhaps somewhat contrived, because we know the two SISs that are to be used, but nevertheless it illustrates the process of resource discovery in a simple way: we can search for a database of class `GrabTaxonDatabase`, and specify the name of the database (e.g. 'ILDIS') or the name of the higher taxon that it contains information about (e.g. 'Leguminosae').

The monitor is a simple Java applet that monitors GRAB activity; it does not depend on Globus. Nevertheless, we have found this monitor to be useful when demonstrating GRAB, as it illustrates clearly what is happening within the GRAB system, while this is not evident from the main Web interface window.

## 4 DISCUSSION AND FUTURE WORK

### 4.1 Lessons Learned

We have successfully implemented a prototype demonstrator illustrating how data from various sources can be brought together using Globus and, in particular, illus-

**ILDIS:**

```

...
objectClass: GrabTaxonDatabase
Grab-Taxon-name: ILDIS
Grab-Taxon-higherTaxon: Leguminosae
Grab-Taxon-contact: grab.biol.soton.ac.uk
Grab-Taxon-executable:
    /home/globus/mybin/ILDISImageInterfaceServer

```

**Fishbase:**

```

...
objectClass: GrabTaxonDatabase
Grab-Taxon-name: FishBase
Grab-Taxon-higherTaxon: Pisces
Grab-Taxon-contact: grab.biol.soton.ac.uk
Grab-Taxon-executable:
    /home/globus/mybin/FishBaseImageInterfaceServer
...

```

Fig. 7. Partial listing of MDS data for ILDIS and Fishbase

trating the role that a catalogue of life can play in locating species-related information. We shall now mention a number of lessons learned from this project.

#### 4.1.1 Installing and Understanding Globus

The GRAB project was initiated when the UK GRID Support Centre was just being established. We found that the Globus software was difficult to install and manage with the limited support that was initially available. At the end of the project an experienced Globus developer joined the GRAB team, and the remaining problems we had were quickly rectified. Support of experienced Globus developers is thus important to the success of Globus-based projects.

#### 4.1.2 Usefulness of Current GRID Software

As mentioned in Section 1, our brief was to use existing GRID software such as Globus and SRB. On close examination, SRB did not appear to provide facilities relevant to the software architecture adopted: it is primarily a platform-independent, distributed file access system. For this reason we decided for our prototype to use Globus only. Yet we have found that Globus provides facilities primarily at the computation/data level, and in a form not well suited to database applications. As with the CGI-based approach initially taken, we have had to develop a mechanism for supporting specific 'canned queries' to the component databases. This is acceptable

for the present GRAB system, but would not be suitable for a full-scale biodiversity informatics problem-solving environment in which data from various sources is to be combined flexibly in ways determined by the user of the system. We are aware that the UK Database Task Force<sup>11</sup> is working on this problem, and when a solution is delivered it will increase the range of tasks to which Globus can sensibly be applied.

With regard to GRID facilities at higher levels, there is an increasing interest in viewing the GRID as having three layers: *Computation/Data*, *Information* and *Knowledge* [8]. But the Information and Knowledge layers have yet to be fully built: currently, for example, the metadata used in GRAB is *of necessity* fairly primitive, since MDS is LDAP (Lightweight Directory Access Protocol)-based and hence essentially hierarchical with no explicit ontology. Sophisticated GRID tools and middleware, for such tasks as finding resources, using metadata to interpret data, and scientific visualisation, are needed in biodiversity informatics — as in other disciplines — if the potential of the GRID is to be realised. Again, progress is being made in this area: the Open Grid Services Architecture [6] is an important development, since it provides a framework not only for locating appropriate services but also for making explicit how they should be used.

#### 4.1.3 Data Quality and Semantics

In Section 3.2 it was assumed that the accepted name obtained from the catalogue of life was the appropriate term to use in subsequent queries to SISs. This is not necessarily the case, since taxonomists may differ in their opinions. Some of the problems in a given SIS might be that (a) the species is referred to by one of the synonyms held in the catalogue, not the accepted name, or (b) a different species is referred to by the same name. In order to support accurate retrieval, tools that can detect such conflicts, such as LITCHI (see Section 2.1), need further development so that they not only create consistent checklists but also maintain relationships between conflicting ones. Again, it will be noted that *Vicia faba* (the broad bean) is recorded as native to Iceland in the ILDIS system. In fact, it is cultivated there but is not native to that country. This is an example of where semantic problems can lead to inaccurate results.

#### 4.2 BiodiversityWorld

BiodiversityWorld is a three-year project which, at the time of writing, has just commenced. It is biology-led, with the intention that biological research in three of the four areas mentioned in Section 2.1 should be supported — biodiversity richness analysis & conservation evaluation; bioclimatic modelling & climate change, and phylogenetic analysis & biogeography. The essential aim is to build a GRID-based system that will support collaborative research in these three areas, providing flexible access to a representative range of biodiversity data and analytic tools. Unlike

---

<sup>11</sup> <http://www.cs.man.ac.uk/grid-db/>

in the GRAB project, this will entail the development of a full problem-solving environment supporting the creation of workflows in which quite heterogeneous resources can be brought together to support the biologists' research. Part of the computer science research required will be to track developments in GRID support for databases, resource discovery and metadata and to contribute to these developments where appropriate. Another computer science issue that will need to be investigated is how collaborative research among biologists can best be supported by a problem-solving environment.

## 5 CONCLUSIONS

We have outlined some of the main areas of current research interest within the field of biodiversity informatics, and discussed the benefits that the GRID might offer to this research discipline. The GRAB prototype illustrates the feasibility of data integration using current Globus technology, but also highlights some of the areas in which further development is necessary, particularly the areas of database support, resource discovery and metadata.

## Acknowledgements

The GRAB project was funded by a grant from the UK Department of Trade and Industry. We are grateful to ILDIS, Fishbase and Species 2000 for access to their systems and data. The BiodiversityWorld project is funded by a grant from the UK Biotechnology and Biological Sciences Research Council to the Universities of Reading, Cardiff and Southampton, and The Natural History Museum. We are grateful to John S. Robinson for the substantial contribution he made to the implementation of the GRAB prototype.

## REFERENCES

- [1] ALBERDI, E.—SLEEMAN, D.: ReTAX: A Step in the Automation of Taxonomic Revision. *Artificial Intelligence*, Vol. 91, 1997, No. 2, pp. 257–279.
- [2] BUSBY, J.: BIOCLIM — a Bioclimatic Analysis and Prediction System. In *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*, C. Margules and M. Austin, Eds. CSIRO, 1991, pp. 64–68.
- [3] EMBURY, S. M.—BRANDT, S. M.—ROBINSON, J. S.—SUTHERLAND, I.—BISBY, F. A.—GRAY, W. A.—JONES, A. C.—WHITE, R. J.: Adapting Integrity Enforcement Techniques for Data Reconciliation. *Information Systems*, Vol. 26, 2001, No. 8, pp. 657–689.
- [4] FOSTER, I.—KESSELMAN, C.: The Globus Toolkit. In *The Grid: Blueprint for a New Computing Infrastructure*, Chap. 11, I. Foster and C. Kesselman, Eds. Morgan Kaufmann, San Francisco, CA, 1999, pp. 259–278.

- [5] FOSTER, I.—KESSELMAN, C. eds. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco, CA, 1999.
- [6] FOSTER, I.—KESSELMAN, C.—NICK, J.—TUECKE, S.: *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. <http://www.globus.org/research/papers/ogsa.pdf>.
- [7] FOSTER, I.—KESSELMAN, C.—TUECKE, S.: *The Anatomy of the Grid: Enabling Scalable Virtual Organization*. *The International Journal of High Performance Computing Applications*, Vol. 15, 2001, No. 3, pp. 200–222.
- [8] JEFFERY, K. G.: *GRIDs in ERCIM*. *ERCIM News*, April 2001.
- [9] JONES, A. C.—SUTHERLAND, I.—EMBURY, S. M.—GRAY, W. A.—WHITE, R. J.—ROBINSON, J. S.—BISBY, F. A.—BRANDT, S. M.: *Techniques for Effective Integration, Maintenance and Evolution of Species Databases*. In *12th International Conference on Scientific and Statistical Databases (2000)*, O. Günther and H.-J. Lenz, Eds., IEEE Computer Society Press, pp. 3–13.
- [10] JONES, A. C.—XU, X.—PITTAS, N.—GRAY, W. A.—FIDDIAN, N. J.—WHITE, R. J.—ROBINSON, J. S.—BISBY, F. A.—BRANDT, S. M.: *SPICE: A Flexible Architecture for Integrating Autonomous Databases to Comprise a Distributed Catalogue of Life*. In *11th International Conference on Database and Expert Systems Applications (LNCS 1873) (2000)*, M. Ibrahim, J. Küng, and N. Revell, Eds., Springer Verlag, pp. 981–992.
- [11] PULLAN, M. R.—WATSON, M. F.—KENNEDY, J. B.—RAGUENAUD, C.—HYAM, R.: *The Prometheus Taxonomic Model: a Practical Approach to Representing Multiple Taxonomies*. *Taxon*, Vol. 49, 2000, No. 1, pp. 55–75.
- [12] XU, X.—JONES, A. C.—PITTAS, N.—GRAY, W. A.—FIDDIAN, N. J.—WHITE, R. J.—ROBINSON, J. S.—BISBY, F. A.—BRANDT, S. M.: *Experiences with a Hybrid Implementation of a Globally Distributed Federated Database System*. In *2nd International Conference on Web-Age Information Management (LNCS 2118) (2001)*, X. S. Wang, G. Yu, and H. Lu, Eds., Springer Verlag, pp. 212–222.



**Andrew C. Jones** is a lecturer in the Department of Computer Science at Cardiff University. He received his PhD in 1992, working in the area of object-oriented systems, and subsequently worked on active object-oriented databases. More recently he has specialised in biodiversity informatics, including data quality issues and federated heterogeneous biological databases, as a partner in the LITCHI and SPICE projects. At present he is researching in the area of GRID-based biodiversity problem-solving environments: he is a partner in the BiodiversityWorld project and was previously a partner in the GRAB project.



**W. Alex GRAY** is Professor of Advanced Information Systems in the Department of Computer Science at Cardiff University. A specialist in interoperable distributed information systems, he has 25 years experience and over 150 scientific articles. In the field of biodiversity informatics, he was a partner in the LITCHI, SPICE and GRAB projects, and is currently a partner in the BiodiversityWorld project. He has directed 7 major research projects as a Principal Investigator.



**Jonathan P. GIDDY** is the Grid Technologies Co-ordinator for the Welsh e-Science Centre at Cardiff University. His interests include distributed computing architectures, autonomic computing and market-based scheduling. In his previous position as a research scientist for the Distributed Systems Technology Centre in Australia, he was a key developer of the Nimrod software for running parametric simulations on the Grid. He received the Bachelor of Computer Science with Honours from the University of Wollongong, Australia.



**Nick J. FIDDIAN** is Professor and Head of Computer Science at Cardiff University. He has 25 years research experience and 100 research publications. He has worked closely with Professor Gray on several large projects. In the field of biodiversity informatics these projects include SPICE and BiodiversityWorld, and also the smaller GRAB project.