



IST-2004-004475

DataMiningGrid

Data Mining Tools and Services for Grid Computing Environments

Specific Targeted Research or Innovation Project
2.3.2.8 Grid-based Systems for Complex Problems Solving

D82(2): Description of User Group(s)

Due date of deliverable: M18 (28 February 2006)

Actual deliverable submission date: 5 April 2006

Start date of project: 1 September 2004

Duration: 24 months

Fraunhofer Institute for Autonomous Intelligent Systems (FHG)

Revision: 03

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	



DATAMINING

GRID

**Deliverable D82(2):
Description of User
Group(s)**



DATA MINING TOOLS AND SERVICES FOR GRID COMPUTING ENVIRONMENTS

Deliverable D82(2): Description of User Group(s)

Responsible author(s): Terence Dörflinger
Co-author(s): Werner Dubitzky, Jürgen Franke, Nahum Korda,
Thomas Niessen, Michael May, Gerhard Paaß,
Francois Perrevort, Matthias Röhm, Vlado
Stankovski, Jernej Trnkoczy

Revision history

Deliverable administration and summary		
Project acronym: DataMiningGrid	ID: IST-2004-004475	
Document identifier:	DataMiningGrid-del-D82(2)DescriptionOfUserGroups-s	
Leading Partner: FHG		
Report version: 03		
Report preparation date: 28.2.2006		
Classification: Public		
Nature: Report		
Author(s) and contributors: Terence Dörflinger (FHG) in collaboration with all Partners		
Status:	-	Plan
	-	Draft
	x	Working
	-	Final
	-	Submitted
	-	Approved

The DataMiningGrid © Consortium has addressed all comments received, making changes as necessary. Changes to this document are detailed in the change log table below.

Date	Edited by	Status	Changes made
25.10.04	Jernej Trnkoczy	Plan	A report template will be defined that all reports will follow.
24.03.06	Terence Dörflinger	Working	Working version of 'Description of User Groups'
26.03.06	Werner Dubitzky, Alice McQuillan	Submitted	Final checks and minor corrections.

Notice that other documents may supersede this document. A list of latest public DataMiningGrid deliverables can be found at the DataMiningGrid Web page at www.DataMiningGrid.org/dissemination.

Copyright

This report is © DataMiningGrid Consortium 2006. Its duplication is restricted to the personal use within the Consortium, funding agency and project reviewers.

Citation

Terence Dörflinger (2006), Deliverable D82(2). DataMiningGrid Consortium, c/o University of Ljubljana, www.DataMiningGrid.org

Acknowledgements

The work presented in this document has been conducted in the context of the EU Framework Programme VI project IST 2004 004475 DataMiningGrid. DataMiningGrid is a 24-month project that started on 1 September 2004 and is funded by the European Commission as well as by the industrial Partners. Their support is appreciated.

The Partners in the project are University of Ulster (UU), Fraunhofer Institute for Autonomous Intelligent Systems (FHG), DaimlerChrysler (DC), Israel Institute of Technology (TECH) and University of Ljubljana (LJU). The content of this document is the result of extensive discussions within the DataMiningGrid© Consortium as a whole.

More information

Public DataMiningGrid reports and other information pertaining to the project are available through the DataMiningGrid project's public Web site under www.DataMiningGrid.org.

Executive Summary

A critical success factor for the DataMiningGrid project is to ensure that project outcomes meet the user demand. Deriving from the general estimation of potential user groups for the project results, it is necessary to identify and define targeted user groups, which are intentionally and deliberately addressed by the project outcomes.

Considerable efforts have been made to determine these targeted user groups. Concerning the joint exploitation efforts, this has been done with respect to the DataMiningGrid Software Packages and the DataMiningGrid training activities, concerning the individual exploitation efforts the targeted user groups have been described on a Consortium Partner level. Moreover some controlling efforts have been successfully performed.

Table of Contents

Executive Summary	6
Table of Contents.....	7
1 Introduction	8
2 Objectives of the Description of User Groups.....	9
3 Description of User Groups	10
3.1 Joint Exploitation	10
3.1.1 DataMiningGrid Software Packages	10
3.1.2 Specific data mining algorithms.....	12
3.1.3 DataMiningGrid Training Services	13
3.2 Individual Exploitation	14
3.2.1 University of Ulster.....	14
3.2.2 DaimlerChrysler	18
3.2.3 Israel Institute of Technology	18
3.2.4 University of Ljubljana	19
Conclusions.....	21
4 References.....	22
5 Abbreviations	22

1 Introduction

For the successful exploitation of the various DataMiningGrid results, it is essential to meet the demand of the targeted user groups. To ensure this, it is necessary to predict the character of future users as accurate as possible. Having a clear understanding of future users enables the project Partners to customise products and services according to user's requirements on a joint as well as on an individual exploitation basis.

The D82(2) Description of User Group(s) deliverable is part of WP8 Dissemination, Awareness and Exploitation. It builds on the Annex I to the Contract and compliments the Dissemination Plan and the Exploitation Plan. It further highlights how user groups change in the light of the evolving project and its results. The aim is to provide a clear strategy for all Partners and the Consortium.

2 Objectives of the Description of User Groups

Having detailed knowledge about potential user groups for the DataMiningGrid project outcomes is necessary for a variety of reasons. A clear picture is important to ensure that project outcomes meet the users' demands. Moreover, it affects other important project aspects such as:

- Specification of requirements;
- Effective dissemination; and
- Effective exploitation.

To ensure that user demands will be met, the following actions have to be carried out:

- Identifying potential user groups for project results;
- Identifying targeted user groups for project results; and
- Controlling whether targeted user groups get addressed.

The identification of potential user groups focuses on the general overview of possible users. It provides valuable information for the definition of markets that are subject of penetration in the exploitation. The targeted user groups consist of the part of potential user groups that have been decided to particularly address by the specific project results. For the success of the DataMiningGrid project it is necessary that the identified targeted user groups will actually benefit from the developed results. Hence, controlling efforts have been performed to evaluate whether the demands of the targeted user groups correspond to the anticipated results of the project.

The version of Deliverable D82(1) Description of User Group(s) covered potential as well as specifically targeted users. This final version of the document will strictly focus on the targeted users.

3 Description of User Groups

A user group is a group of individuals with common interests in a technology or an application. In the scope of the DataMiningGrid project, user groups will be referred to as group of individuals that have interest or need for one or more specific results of the project. One can distinguish between groups of targeted users, potential users and actual users:

- **Potential user groups:** All user groups that potentially have a demand for or interest in the project's outcomes;
- **Targeted user groups:** Specific user groups that are intentionally addressed by the project and its results;
- **Actual user groups:** User groups that will actually use the project results when they are available.

While it is fairly easy to determine a diffuse class of potential users and to define a group of targeted users, it is virtually impossible to know in advance who exactly the actual users will be. Therefore, the goal must be to estimate as accurately as possible the actual users in advance. This estimation should then match with the definition of targeted user groups in order to avoid acceptance problems.

The deliverable version D82(1) Description of User Group(s) covered potential as well as targeted users. This final version of the document will exclusively focus on targeted users.

The rationale of Deliverable D82(2) is aligned with that of D83(3) Exploitation Plan, meaning that it distinguishes between joint and individual exploitation.

3.1 Joint Exploitation

A business solution based on DataMiningGrid results can be characterised by the equation

$$\begin{aligned}
 & \textit{DataMiningGrid-based business solution} \\
 & = \\
 & \textit{DataMiningGrid Software Packages} \\
 & + \\
 & \textit{Specific data mining algorithms} \\
 & + \\
 & \textit{DataMiningGrid software package configuration}
 \end{aligned}$$

For the three different categories of results, different user groups can be identified.

3.1.1 DataMiningGrid Software Packages

The products arising from the project are the DataMiningGrid Software Packages. These packages can be applied within OGSA-DAI and Globus Toolkit 4

environments respectively, enabling users to build their own (data mining) applications on top of the extended middleware. Each package has been carefully chosen so that it is largely independent from all other packages. This modularity facilitates reuse in other projects, in which only part of the functionality of the full DataMiningGrid components is required. The following packages will be provided at the end of the project:

DataMiningGrid Data Services package

The Data Services package contains all server-side services and client-side components related to data transport with OGSA-DAI and GridFTP. They especially contain all OGSA-DAI activities developed by the Consortium, such as the 'DB2ARFF' activity for transforming the data contained in a database table into Weka's ARFF format. Provisions are made for other format transformations to be added.

DataMiningGrid Information Integrator package

The Information Integrator package contains all required services and components for issuing user-defined queries against the application repository for finding a data mining algorithm appropriate for the task at hand. The client-side components, amongst others, are capable of processing these metadata descriptions in order to provide a dynamically compiled graphical user interface for defining parameter values and inputs. Furthermore, this package also contains a Web-based mechanism for integrating new algorithms into the system and specifying their respective descriptions in the process.

DataMiningGrid Execution System package

This package contains the service-based resource broker, which is responsible for finding machines appropriate for the job to execute. The package also provides services for monitoring each job and dealing with errors, which may occur during remote execution.

DataMiningGrid Full System package

The Full System package comprises the whole suit of software components developed by the Consortium. It consists of all packages mentioned above plus the required units for using some components from within the Triana workflow editor. These units are lightweight wrappers around the respective client classes, but do not contain any system logic themselves. As they rely on the client components contained in the other packages mentioned above, the units do not form a package on their own (such as, e.g., the Data Services package).

3.1.1.1 Targeted users

Since the DataMiningGrid Software Packages will be open source, they are potentially available for anyone interested. Besides the internal usage of the DataMiningGrid Software Packages within the DataMiningGrid project, four different kinds of user groups can be identified.

The first one consists of the end users (e.g., industrial companies with large amounts and/or distributed data in need of faster and/or better quality data

mining technology) that will develop their own grid-enabled data mining solutions using the DataMiningGrid Software Packages.

Another user group are software service providers that will develop customised applications for their customers using the DataMiningGrid Software Packages.

Furthermore, the DataMiningGrid Software Packages could be used by software companies such as data mining tool vendors or data analysis service providers (like IBM, Oracle, SAS, SPSS, OLAP database producers) that can integrate the DataMiningGrid technology to improve their products.

Finally, the DataMiningGrid Software Packages are intended for use by research institutes, universities as well as the grid and data mining community for educational purposes as well as to support the enhancement of grid and data mining middleware and software in general and the DataMiningGrid technology in particular (e.g. in other EU-funded projects).

The DataMiningGrid project encourages the use and adoption of the DataMiningGrid Data Services package by the grid community and EU grid research projects as pointed out in the Collaboration Report (Deliverable D73(1)).

3.1.1.2 Controlling efforts

The DataMiningGrid Software Packages contain the results – especially grid enhancements and data mining interfaces - in a user-friendly bundle that makes them easy to adapt and deploy in grid environments. In the Consortium, the industry is represented by DaimlerChrysler, who will use the DataMiningGrid technology in-house. This shows that the DataMiningGrid Software Packages can be useful for large companies that want to build their own grid-enabled data mining applications.

Meetings and discussions with companies like Purple Insight and DataSynapse showed that there is strong interest in DataMiningGrid technology by with software companies and software service providers. In these talks the need for expert training on DataMiningGrid Software Packages became obvious, which is now an additional joint exploitation activity of the DataMiningGrid Consortium. The DataMiningGrid Training Services represent one result of the consultations.

Moreover, the DataMiningGrid outcomes were presented to the grid community and other EU-funded and nationally funded (e.g. UK e-Science) projects within the TG5 Data Management workshops. These workshops are part of the EU-IST grid concertation initiative.

3.1.2 Specific data mining algorithms

In addition to the DataMiningGrid Software Packages – for a functional business solution – specific data mining algorithms are needed. Several demonstrators, which are developed within the project, add these missing pieces in a domain specific way, thus demonstrating the potential of the DataMiningGrid Software

Packages for solving research and business problems. These vertically integrated solutions can be seen as products in their own right, e.g., the text mining solution for quality management or the bioinformatics applications.

Some or all of these demonstrators will potentially be turned into products that can be subsequently exploited by the individual Consortium Partners. Nevertheless, it should be noted that the demonstrators (by their very nature as demonstrators) are not immediately marketable products but will need significant additional investments for product development. The products are unlikely to be open source, since they incorporate on loosely coupled, proprietary data mining algorithms.

3.1.2.1 Targeted users

As the specific data mining algorithms will be individually offered on a Partner level basis, they will be described in the individual exploitation section later in this document.

3.1.3 DataMiningGrid Training Services

From an end users perspective, the DataMiningGrid client and server packages combined with the specific data mining algorithms do not provide a full solution for a business problem, since domain specific additions as well as customisations to the software are not included.

To help users creating their own individual grid-enabled data mining solutions using the DataMiningGrid Software Packages, the DataMiningGrid Consortium plans to jointly offer training services.

These services include the transfer of know-how for the

- Installation, operation and maintenance of a (GT4) grid environment;
- Installation, integration, operation and maintenance of a (Condor) workload management system;
- Installation, integration, customisation, operation and maintenance of the DataMiningGrid Software Packages; and
- Installation, integration, customisation, operation and maintenance of data mining algorithms.

This know-how resides with different Partners to a different extent (in line with the Partners' roles in the project). It can be used to enter further research projects as a grid expert and can help to spread knowledge about the grid to the research community, especially to consumers of grid middleware that are not themselves interested in grid research.

Partners active in industrial exploitation such as DC can use it to communicate their knowledge about the grid within their organisation. This could potentially lead to large-scale take-ups of grid technology in the automotive industry.

Partners cooperating with external companies such as FHG can use it to communicate knowledge about the grid in other companies.

3.1.3.1 Targeted users

The DataMiningGrid Training Services address software developers, especially grid experts who want to learn about the

- Installation, operation and maintenance of a (GT4) grid environment;
- Installation, integration, operation and maintenance of a (Condor) workload management system;
- Installation, integration, customisation, operation and maintenance of the DataMiningGrid Software Packages; and
- Installation, integration, customisation, operation and maintenance of data mining algorithms.

These Services will be most interesting for third party software service providers that wish to offer new customised solutions to the industry based on the DataMiningGrid Software Packages. They are also aiming at software departments of large industrial companies that wish to develop customised grid-enabled data mining application in-house.

3.1.3.2 Controlling efforts

The DataMiningGrid Training Services represent a direct result of the controlling efforts carried out by some partners for the DataMiningGrid Service packages. Discussions and meetings with software companies and software service providers showed the demand for expert training activities. This way the content and scope of training activities had been adjusted directly according to the expectations of software companies.

3.2 Individual Exploitation

The Consortium Partners have individual aims concerning exploitation of DataMiningGrid technology. In the following, the targeted user groups for the specific, Partner-individually exploited results will be described.

3.2.1 University of Ulster

The expected result of the University of Ulster's (UU) contribution towards the DataMiningGrid project consists of the following items:

- A specification for the access of data services for the other components in DataMiningGrid project (making efficient use of existing technologies and standards);
- To implement and distribute whatever resources are required to provide access and pre-processing methods to support the data mining activities of the project; and

- To implement two demonstrator applications showing how the DataMiningGrid project can be used to support research activities in bioinformatics.

3.2.1.1 Targeted Users

The University of Ulster’s (UU) contribution to the DataMiningGrid project is initially intended for use by the other project Partners. They will implement the data access components of their software according to the specifications provided by the UU. Ultimately, the target user groups are researchers, developers and other users wishing to develop and use data mining techniques within grid-computing environments. The demonstrator components developed at UU will be of interest to wider bioinformatics and biology community, including pharmaceutical industry which is increasingly relying on data mining and bioinformatics technologies for their businesses. User groups associated with UU’s requirements and applications are depicted in Table 1.

Table 1. UU user groups

Results	Targeted User Group
Data service access specification	<ul style="list-style-type: none"> • DataMiningGrid project Partners • Researchers and developers wishing to use data mining techniques within grid computing environments
Implementation and distribution of resources	<ul style="list-style-type: none"> • DataMiningGrid project Partners
UU demonstrator tasks	<ul style="list-style-type: none"> • Researchers and developers wishing to use data mining and data management techniques within grid computing environments • Bioinformatics and systems biology communities with grid-based data mining and data management requirements

3.2.1.2 Controlling efforts

UU has already ‘tapped’ into different research and IT communities by exploiting their know-how gained in the DataMiningGrid project on grid-enabled data management techniques. UU has participated in two FP6 IST Call 5 proposals on Advanced Grid Technologies, Systems & Services projects – both proposals (see below) are currently under negotiation with the EC:

- Quasi-Opportunistic Supercomputing for Complex Systems in Grid Environments (QosCosGrid) (11 partners, 2 of which are from industry; UU is coordinator); and
- Grid Services Based Environment to Enable Innovative Research (Chemomentum) (10 partners, 2 of which are from industry).

In both projects UU will be in charge of the grid-based data management aspects of the project. In addition, UU has been leading a proposal (8 partners) under the FP6 NEST programme in which UU is both coordinator and in charge of data

management, part of which within grid environments. The involved partners in above projects/proposal constitute already a large community in areas such as biomedical research and development, IT and other areas. Increasingly, these communities require grid-enabled technologies, including data mining and data management. This success points to bright future of UU in terms of collaborating (research) with these communities on grid-based data mining/management and also further into the future on possible commercial usages of this know-how. We also intend to use UU-developed DataMiningGrid components for data handling within the new FP6 IST Call 5 projects.

UU's demonstrators involve two research collaborators (Weihestephan University of Applied Sciences, Freising, Germany and University of Coimbra, Coimbra, Portugal). Both collaborators are strongly linked to industry in their respective field of biotechnology and biochemistry. We are currently in the process of evaluating the protein unfolding simulation data repository (P-Found¹, partly developed under DataMiningGrid) with the respective biochemical and bioinformatics communities. This could open a considerable channel for future user groups.

3.2.1.3 Fraunhofer Institute for Autonomous Intelligent Systems

The Fraunhofer Institute for Autonomous Intelligent Systems (FHG), as a data-mining research and consulting centre, will use the technology to offer specific products and services. FHG will provide:

- A grid-enabled version of the Weka [Weka04] toolkit that will provide several advantages compared to the standalone application such as execution on any suitable machine in the grid, without the need of installing Weka on all these machines.
- A collection of text-mining modules, as well as the associated knowledge how to develop solutions with these modules. The text mining modules can be combined into complex workflows using the common workflow editor. In this way, various pre-processing, clustering, and classification steps may be chained and distributed over the grid.
- Fraunhofer AIS will use the DataMiningGrid Software Packages to build tailor-made data mining applications for individual customers with the demand for highly specific solutions. These applications range from the fast grid-enabled calculation of driving zones for specific spatial data mining tasks to the privacy preserving parallel mining of large text repositories. FHG will offer services which will cover the entire knowledge discovery process, starting with the creation of statistical models, the pre-processing and enrichment of (distributed) data, development, implementation and deployment of (parallel) data mining algorithms

¹ Silva C.G., Ostropytskyy V., Loureiro-Ferreira N., Berrar D., Dubitzky W., and Brito R.M.M., P-Found: A Protein Folding and Unfolding Simulation Data Repository (submitted to ISMB 2006)

(making use of the DataMiningGrid grid-enabled Weka services) to the interpretation of data mining results.

3.2.1.4 Targeted users

Targeted users for the grid-enabled version of the Weka toolkit are researchers as well as data mining experts that have a demand for parallel, distributed data mining.

Targeted users are people that administrate large text repositories, e.g. document collections, service records, customer contact records or administrative records, which require efficient text mining systems. In addition, the ability to have the raw text databases, method repositories and compute servers at different locations is a possible advantage for users, who want to control the location of their data.

The targeted customers for application building services are industrial companies (e.g., insurance companies, car manufacturers and telecommunication providers) that encounter problems with traditional data mining because of computational inefficiency or insufficient prediction accuracy which can be solved by applying Grid technology. Moreover the developed DataMiningGrid text mining applications can be useful to media companies with large news databases, industrial enterprises with many customers, public administrations with contacts to many citizens, which are all organisations that generate massive amounts of text documents with partially free content.

Additionally, FHG wants to transmit both the acquired technology and the knowledge about setup and use in the EU-funded SIMDAT project, where FHG plays a central role in data mining. FHG and DC will use their strong position in the data mining community, e.g., as coordinator and management board member of the European Coordination Action KD-Ubiq, to push the topic of grid computing and to promote the project’s results in the data-mining community. User groups associated with FHG’s requirements and applications are depicted in Table 2.

Table 2. FHG user groups

Results	Targeted User Group
Grid-enabled version of Weka	<ul style="list-style-type: none"> • Researchers with demand for parallel distributed data mining • Data mining experts that have a demand for parallel distributed data mining • Present and future grid related EU-projects
Collection of text mining modules	<ul style="list-style-type: none"> • Administrators of large text repositories, who require efficient text mining systems
Application building services	<ul style="list-style-type: none"> • Industry

3.2.1.5 Controlling efforts

Fraunhofer AIS will use the Grid-enabled Weka algorithms internally to enhance its data mining services. Discussion with Fraunhofer AIS and University of Ljubljana experts largely influenced the design and the Grid-integration of the algorithms. The Grid-enabled Weka services will in future be deployed on the Fraunhofer AIS Grid, thus enabling scientists to more efficiently calculate data mining results (for commercial as well as for educational purposes).

Concerning the text mining modules, Fraunhofer developed them in close collaboration with DaimlerChrysler, who will use them to analysis diagnostic computer logs for early stage car fault detection. Moreover the development has been influenced by the Fraunhofer AIS 'SeIAg' project with Deutsche Welle, where new stories from large text archive repositories need to be categorised. Hence it can be concluded that the Text Mining algorithms, which are currently developed within the DataMiningGrid project, address the demand of the industry user groups.

Moreover, Fraunhofer AIS offers currently application building services to the Axel Springer publishing company. Within this project it is necessary to determine catchments areas of newspaper shops. To have high quality results within an appropriate time frame, Fraunhofer AIS will use DataMiningGrid technology for efficient calculation. This project proofs that the targeted users do have demand for individual grid-enabled knowledge discovery services.

3.2.2 DaimlerChrysler

DaimlerChrysler (DC) has a direct need to use the applications that are developed in the scope of the DataMiningGrid project in its in-house processes related to quality management and customer relationship management, especially to compute large amounts and distributed, mostly confidential data.

3.2.2.1 Targeted Users

At DC the targeted users are internal departments working on quality and security enhancement. Additional target user groups are all developers in the company, which should be supported by better abilities of finding related textual information.

3.2.2.2 Controlling efforts

The intermediate project results were discussed with the targeted user groups. The users suggestions were taken up and integrated into the further development. To meet the requirements these steps were repeated iteratively.

3.2.3 Israel Institute of Technology

Israel Institute of Technology (TECH) intends to develop generic monitoring capabilities, which may then be deployed as part of Net-batch, Condor, DataGrid [DataGrid04], and other grid systems. TECH's project results consist of:

- Mechanism for rapid deployment of data-mining applications on the grid that could significantly cut down time-to-market of complex data-mining applications; and
- Log mining tool designed to detect various failure points and allow improved configuration and administration of the grid infrastructure.

3.2.3.1 Targeted Users

Concerning the rapid deployment mechanism, the typical users are IT personnel and executives utilising data mining.

Typical users for the log mining tool are grid administrators. User groups associated with TECH’s requirements and applications are depicted in Table 3.

Table 3. TECH user groups

Result	Targeted User Group
Mechanism for rapid deployment of data-mining applications	<ul style="list-style-type: none"> • IT personnel and executives utilizing data mining
Log mining tool	<ul style="list-style-type: none"> • Grid administrators

3.2.4 University of Ljubljana

The University of Ljubljana’s (LJU) is a research and education institution. In this sense, the immediate result that LJU expects is knowledge that shall be exploited to improve teaching activities. Moreover, LJU expects to contribute to the delivery of specific services in three distinct areas:

- Services that could be used in connection to standard digital libraries, which could improve the ways in which digital libraries are used. This specifically refers to certain text-mining and ontology-learning services that we intend to implement;
- Services that could be used for ecosystem modelling that take into account collected data as well as previous knowledge to generate new knowledge; and
- Services that could be used in the medical field, when conducting research studies on greater geographic areas. In this case, the developed data services shall take care about the security, privacy, and governance of the investigated data, therefore enabling on the fly medical analysis of much larger and distributed databases than it has been possible in the past.

3.2.4.1 Targeted Users

Typical users for standard digital libraries are researchers that need to access these around the world and want to discover useful information in these libraries. The technology shall help towards automation of the whole process. It shall provide easier semantic search possibilities for the researchers.

Typical users fro ecosystem modelling are researchers and other users that need to combine mathematical models of natural phenomena together with

experimental data. This approach would not be possible or at least very difficult without the DataMiningGrid technology. The computational grid resources are used to generate quality equations (i.e., models) of the natural phenomena. In the DataMiningGrid project, this shall be demonstrated in a use case of modelling of the Lake Bled ecosystem.

Typical users for medical database processing are researchers in medical centres and health monitoring authorities are the main users of the technology. The technology will enable them to access geographically distributed medical databases and conduct 'on-the-fly' studies. In the DataMiningGrid project, the demonstration is based on a use case dealing with iodine deficiency of children entering high school where data are collected in several regional databases. User groups associated with LJU's requirements and applications are depicted in Table 4.

Table 4. LJU user groups

Identifier	Targeted User Group
Standard Digital Libraries	<ul style="list-style-type: none">• Researchers that need to access digital libraries
Ecosystem modelling	<ul style="list-style-type: none">• Researchers and other users that need to combine mathematical models of natural phenomena together with experimental data
Process medical data in distributed databases	<ul style="list-style-type: none">• Researchers in medical centres and health monitoring authorities

Conclusions

A critical success factor for the DataMiningGrid project is to ensure that project outcomes meet the user demand. Deriving from the general estimation of potential user groups for the project results, it is necessary to identify and define targeted user groups, which are intentionally and deliberately addressed by the project outcomes.

Considerable efforts have been made to determine these targeted user groups. Concerning the joint exploitation efforts, this has been done with respect to the DataMiningGrid Software Packages and the DataMiningGrid Training Activities, concerning the individual exploitation efforts the targeted user groups have been described on a Consortium Partner level. Moreover controlling efforts have been successfully performed.

Hence, it can be concluded that the DataMiningGrid project outcomes are highly likely to meet the users' demands.

4 References

References cited in this document and throughout the project are listed in the DataMiningGrid Project Manual, which accompanies the deliverables and reports. Some of the references can also be searched on the project's Web site's digital library. The Project Manual is available to the EC Services and the reviewers on the project's BSCW document server.

5 Abbreviations

Acronyms and abbreviations used in this document and throughout the project are listed and described more comprehensively in the DataMiningGrid Project Manual. The Project Manual is available to the EC Services and the reviewers on the project's BSCW document server.