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

D81(3): Dissemination Plan

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

University of Ljubljana (LJU)

Revision: 04

Deliverable D81(3): Dissemination Plan

DATA MINING TOOLS AND SERVICES FOR GRID COMPUTING ENVIRONMENTS

Deliverable D81(3): Dissemination Plan

Responsible author(s):   Jernej Trnkoczy
Co-author(s):   Vlado Stankovski

Revision history

Deliverable administration and summary		
Project acronym: DataMiningGrid		**ID:** IST-2004-004475
Document identifier:	DataMiningGrid-del-D81(3)DisseminationPlan-s	
Leading Partner: LJU		
Report version: 04		
Report preparation date: 5 April 2006		
Classification: Public		
Nature: Report		
Author(s) and contributors: Jernej Trnkoczy, Vlado Stankovski		
Status:		Plan
		Draft
		Working
		Final
	X	**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
30.10.2005	Jernej Trnkoczy	Plan	Setup of pre-filled deliverable document template
03.02.2006	Jernej Trnkoczy	Draft	First draft version
24.02.2006	Vlado Stankovski	Draft	Comments added
05.03.2006	Jernej Trnkoczy	Draft	Web statistics included
18.03.2006	Jernej Trnkoczy	Working	Partners input added
24.03.2006	Werner Dubitzky, Alice McQuillan	Submitted	Final checks and minor consistency and other corrections

Note that other documents may supersede this document. A list of newest public DataMiningGrid deliverables can be found at the http://www.DataMiningGrid.org/dissemination.htm.

Copyright

Citation

Jernej Trnkoczy, Vlado Stankovski (2006). Deliverable D81(3): Dissemination Plan. 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 project IST 2004 004475 DataMiningGrid. DataMiningGrid is a 24-month project that started on September 1st, 2004 and is funded by the European Commission and by the industrial Partners. Their support is appreciated.

The Partners in the project are University of Ljubljana (LJU), University of Ulster (UU), Fraunhofer Institute for Autonomous Intelligent Systems (FHG), DaimlerChrysler (DC) and Israel Institute of Technology (TECH). The content of this document is the result of extensive discussions within the DataMiningGrid© Consortium as a whole. The participants in this working group were: LJU, UU, FHG, DC and TECH. This report is a collaborative effort of all the above organizations.

More information

Public DataMiningGrid reports are available through DataMiningGrid public Web site at www.DataMiningGrid.org.

Executive Summary

The purpose of the Dissemination Plan is to provide a formal planning document to realize the objectives of the dissemination activities. This is the third dissemination document and therefore describes the plan (as of month 18 of the project) for using and disseminating knowledge throughout the project. It aims to go a step further then Annex I [Annex 04] and identifies in detail what precisely needs to be done, when, why and how the Dissemination Plan's objectives and tasks are to be implemented. The Plan seeks to define roles and responsibilities, identify audiences, key messages (current and future), and highlight the methods of communication that should be used, to give details of how the success of dissemination activities are measured and evaluated. The Section 'where we are now' details the methods of communications that are already in place in order to disseminate information about the DataMiningGrid project.

The Plan also explains that, like audiences, new key dissemination messages will also have to be identified and refined as the project progresses. The key messages identified are:

- Details on the project, i.e., the project's purpose and aims;
- The project's potential to change the way scientists work;
- The potential use of project's results in industry and commercial applications;
- Major developments and software releases; and
- New targeted user groups.

The success of the DataMiningGrid project is related to its dissemination. Therefore, all Partners are committed to achieving the highest possible degree of dissemination; both to increase their own scientific visibility and impact, and to support distribution of the DataMiningGrid project, results, and developed technologies among different application sectors, communities and other fora.

The DataMiningGrid dissemination mission can be summarized as The identification of key audiences, active promotion and raising of awareness of the DataMiningGrid project through clear, consistent and timely communications, and publication of research results in relevant journals and conference and workshop proceedings.

Table of Contents

1 Introduction

1.1 Objectives of this document

In order for the project to be successful, it is vital to proactively raise awareness of the project, attract interest and participation from many different sciences and industry disciplines, publishing the research results arising from the project in high-impact journals and at scientific meetings and conferences.

This document is a part of 'Dissemination, Awareness and Exploitation' work package of the DataMiningGrid project. It collects and presents in a structured and coherent way the planned dissemination activities, the activities that were already carried out in order to create the critical mass of interest in the project and its results. The purpose of the Dissemination Plan is to provide a formal planning document for using and disseminating knowledge throughout the project. It is vital for the success of dissemination activities, that all those involved have a clear understanding of the aims of the dissemination activities and what is realistically achievable. This Plan will be revised and updated to reflect the impact of the project.

The Dissemination Plan builds on the Technical Annex [Annex04] and compliments the Exploitation Plan [Exploitation05] to further highlight how the dissemination for the DataMiningGrid project will develop over the two years of its existence. The aim is to provide a clear communication strategy for all Partners and members of the Management Board.

1.2 Document amendment procedure

The Dissemination Plan is the Deliverable D81(3) of WP8. It will be updated at month 24. The produced addendum will be delivered to the Project Technical Board and the Project Manager.

Requests to amend the document must be made to the Dissemination Manager Jernej Trnkoczy (jernej.trnkoczy@fgg.uni-lj.si).

2 Objectives of Dissemination

The aim of the dissemination activities is to create critical mass of interest, which is necessary for the deployment, on the target scale, of the DataMiningGrid technology.

In particular, the dissemination activities have the following objectives:

- To raise awareness of the benefits of the DataMiningGrid to general public, attract potential customers and generate expectation towards the project results, in order to prepare its exploitation;
- To identify and target specific user group audiences and applications (research disciplines, industrial and commercial groups etc.). The promotional activities will especially be aimed at the communities related to data analysis and text mining (i.e., statistics, data/text mining, knowledge discovery in databases, machine learning, pattern recognition, artificial intelligence, knowledge management, information retrieval, and knowledge engineering) and at grid and distributed computing communities. Our approach to targeted user groups will be elaborated by deliverable D82: Description of User Groups;
- To raise awareness of the benefits of the DataMiningGrid project and its results by ensuring that an appropriate message is delivered;
- To ensure that specific user groups become involved in the project so that they can be converted to 'real' users;
- To share the technical results of the project with the scientific community, publishing the project research results in high-impact research publications, international journals and conference proceedings; and
- To improve the knowledge of building data-mining applications in grid computing environments so as to create new opportunities for building quality products and services.

In this document, these objectives are planned in a clear and measurable way. For this purpose, we introduce the widest possible promotion of the DataMiningGrid technology through the Internet and traditional media (press, participation and events etc.). The measures for success of dissemination activities include:

- Web site statistics;
- Publication of research results in international journals, conferences, symposia, and workshops; and
- Success of the initiatives for the wide promotion of the DataMiningGrid (media coverage, the number of information sheets written, events where DataMiningGrid is promoted, etc.).

In order to run consistent and successful dissemination, activities key dissemination *messages* have been identified. New key dissemination messages will also have to be identified and refined as the project progresses. The key messages identified in the campaign so far include:

- What the project is about;
- The project's aims and objectives;
- The project's potential to improve the way scientists work;
- The project's potential to contribute improvements of grid and distributed data mining technology;
- Who is involved in the project;
- Major developments in grid and data-mining technology;
- DataMiningGrid software releases; and
- DataMiningGrid technology exploitation and user communities.

It is also important to note here that not all key messages will be relevant to all targeted users. Therefore, the Consortium must ensure that appropriate messages are tailored to the right audiences.

3 Dissemination Plan

This section describes the planning, set-up, monitoring and refining of a framework for the dissemination of the project results. It is worth making the following preliminary considerations:

- The DataMiningGrid results are much more attractive in the context of a worldwide effort in the area of grid technologies;
- The strategy for an effective dissemination to the industry community should be different from the strategy for the dissemination to the scientific community; and
- The concept of *data mining in grid computing environments* is relatively novel and therefore largely unknown in the industrial community. A more effective way to attract industrial interest is a concrete demonstration of applicability of the new DataMiningGrid technologies to real-world complex problem-solving problems.

It is important that the experience that the scientific community makes with this technology will be converted to attract the interest of industry. The DataMiningGrid will use its commercial Partners to address the potential users of the project results and to raise awareness and create interest in technologies that we are developing.

Dissemination activities of the DataMiningGrid project will include maintenance of an up to dated project and state-of-the-art Web site, editorial cover of the new DataMiningGrid technology and products, publications in scientific journals and conference proceedings, etc. A digital library has been set up for the project, and is used to archive and make available the project results in a form of documents. The results of DataMiningGrid will be presented during all these activities. This will contribute to greater awareness of the DataMiningGrid project, its results and participants, and thus, of the European Union.

In order to exploit the project's results, the DataMiningGrid Consortium promised to ensure an outreach and training effort, which can proactively market DataMiningGrid services to new research communities in academia and industry, capture new requirements for the middleware and service activities, and provide the necessary education to enable new users to benefit from the DataMiningGrid technology. Training activities are described in deliverable D83 Exploitation Plan and D82 Description of User Group(s).

In the following sections a detailed Plan of current/future dissemination activities is described.

3.1 Dissemination manager

The dissemination of the DataMiningGrid project is a challenging task that requires significant human resources. We invested part of our initial efforts to create a group of people dedicated to the dissemination of the DataMiningGrid project. At the time of writing, the following persons are fully or partially staffed

in the DataMiningGrid Dissemination Office at LJU, which is leading and coordinating the dissemination activities:

- Jernej Trnkoczy, Dissemination Manager;
- Mateja Šmid, responsible for marketing;
- Vlado Stankovski, dealing with technical aspects;
- Robert Klinc, graphics designer.

The dissemination activities are supervised by professor Žiga Turk.

Other Partners have dedicated the following human resources to dissemination activities:

- Werner Dubitzky, University of Ulster;
- Francois Perrevort, Fraunhofer Institute;
- Jürgen Franke, DaimlerChrysler; and
- Assaf Schuster, Israel Institute of Technology.

3.2 Corporate image and style

In order for the DataMiningGrid to be easily recognizable, we have built a strong corporate image, brand and style. The DataMiningGrid logotype, templates for slide presentations, information sheets, posters, Web portal, etc. were all designed on the basis of an agreed style. It is crucial to ensure that all participants use the DataMiningGrid styles and the corresponding templates that have been developed.

3.3 Media relations

Media relations (interviews, news releases, briefings etc.) are one of the dissemination methods for all DataMiningGrid Partners. All Partners are expected to maintain contacts with the relevant media, which will help with placement of DataMiningGrid stories, press releases and announcements in the future. DataMiningGrid partners will make proposals and will be involved in editing of special issues of journals in the domain of grid computing and data mining.

3.4 Contacts

All Partners are responsible to establish a set of collaborations and alliances with institutes, departments and organizations.

In order to facilitate the establishment of contacts with parties interested in the project, the national points of contact were established. Each Partner dedicated a person who will act on a national scale and will represent point of contact for local people who are interested in DataMiningGrid project.

- UU: Werner Dubitzky   w.dubitzky@ulster.ac.uk
- FHG: Terence Dörflinger terence.doerflinger@ais.fraunhofer.de

- DC: Juergen Franke   juergen.franke@daimlerchrysler.com
- TECH: Assaf Schuster   assaf@cs.technion.ac.il
- LJU: Jernej Trnkoczy   jernej.trnkoczy@fgg.uni-lj.si

3.5 Joint dissemination and information exchange activities

Joint dissemination and information exchange activities aim to align the project's dissemination activities with those of other EU-funded grid-related projects under the GridCoord project. The DataMiningGrid project aims to establish inter-project collaboration, especially with other European (under GridCoord) grid development projects/activities. The collaboration with European projects is covered in the project's work package, WP7: Concertation, under task T72: Joint fora for the exchange and dissemination.

In addition, we can consider the information exchange with the scientific community (mainly researchers with experiences in the grid and data mining technologies) to be the dissemination to users formally interested in contribute to the technical development of the DataMiningGrid.

3.6 Dissemination materials

For dissemination purposes, a variety of promotion and dissemination materials were/will be produced and disseminated. These include:

- Technical/scientific papers reporting the results of the project in national and international journals and at conferences;
- Slide presentations and posters be presented at conferences, workshops, trade exhibitions and similar events;
- A *basic fact sheet* describing general facts about project and individual *information sheets* describing the project, its objectives and results;
- A folder complete with information sheets;
- A flyer with basic facts about project and it's objectives;
- A CD-ROM containing the project's publications, demos and presentations (subject to IPR constraints); and
- Small promotional gifts.

3.7 Collection of dissemination materials and information

In order to attract potential users of the dissemination, a complete and up-to-date database of information and related news within the world of grid and data mining technologies will be created and maintained for the project lifetime. The aim is the search, collect, select and publish all relevant documents and other media material. Research all material for the project members and for people interested in grid technologies. Results of the research will be maintained in an electronic form in the digital library accessed by the Web portal. Some results on

projects will be published. The following list describes the kind of material in course of collection:

- Grid and data mining related technical and scientific papers;
- Grid and data mining related slide presentations, tutorials and posters; and
- Important grid and data mining related events (conferences, workshops, meetings, exhibitions and demonstrations) calendar.

DataMiningGrid project will also publish own project results (papers, public deliverables, software etc.) and dissemination material. These materials will be published on Web portal and includes:

- DataMiningGrid public deliverables;
- DataMiningGrid open source software components;
- DataMiningGrid slide presentations;
- DataMiningGrid technical/scientific papers;
- DataMiningGrid press releases; and
- DataMiningGrid fact and information sheets.

3.8 DataMiningGrid Web portal

Dissemination of the project through DataMiningGrid Web portal is one of the key dissemination elements. The Web portal will primarily be aimed at attracting new users, at every level. The users are expected to be non-technical and technical users, students and occasional visitors. The portal will be enriched with new contents during the projects lifetime. The maintenance and creation of the Web portal and its physical location is the responsibility of University of Ljubljana.

3.9 DataMiningGrid digital library

A digital library has been set up for the project, using the technologies developed in the projects IST-2001-33127 SciX (LJU) and is used to archive and make available the project results and other important project related materials. Advanced searching, classification, commenting, annotating facilities is provided. The digital library is publicly available through DataMiningGrid Web portal. The digital library will be enriched with new content during the project lifetime.

3.10 Mailing lists and direct mail campaigns

The lists will be used to communicate among different groups in the project and also to notify/inform highly interested external parties about the project progress. Three different e-mail lists were set up:

- dmgridmm@lists.ulst.ac.uk: The Management Board List's purpose is to discuss and communicate project-level aspects of the project;

- dmgridtc@lists.ulst.ac.uk: The Technical List is reserved for discussions/issues regarding technical development and management of technical issues and meetings; and

- dmgrid@lists.ulst.ac.uk: The General List is reserved for all discussions, issues, etc that do not strictly belong to topics covered by above two lists and are/or of interest to all involved.

In addition to mailing lists we are planning to conduct direct mail campaigns. Emails describing important project progress and results will be sent to people from our database of contacts.

3.11 Participation/organization of conferences and exhibitions

The DataMiningGrid Consortium will organize several workshops/conferences. The DataMiningGrid project will also be presented and promoted at several conferences and exhibitions, in order to reach a wide audience. These will be posted on the Web portal.

4 Where are we now?

During the third reporting period of the project the dissemination activity has been focusing on:

- The Web portal promotion and maintenance;
- The establishment of a base of DataMiningGrid related information and news that can be attractive for the potential users;
- The creation of contacts and relationships with potential users, both from industry and research;
- The development of promotional material to support the dissemination activity;
- The promotion of the project through conferences and different media; and
- The organization of workshops in spring 2006.

In the following sections, the dissemination actions that were already taken will be explained in detail. In Section 5 the overall dissemination efforts will be evaluated.

4.1 News releases

To date there have been the following activities regarding news releases:

- An article at the IST Results page entitled Digging deep to unlock the Grid (see Appendix for some more detail);
- The release that highlighted the launch of the DataMiningGrid project and was distributed at the European Grid Technology Days 2004, IST-FP6 Grid Projects Launch and Concertation in Brussels;
- The Cutter Consortium published an article about DataMiningGrid project online in March 2005 (see Appendix A);
- DataMiningGrid project was one of the two projects described in the Invest Northern Ireland newsletter in June 2005 (see Appendix A);
- DataMiningGrid Consortium (Dubitzky W., Huang, C.-H., Lanza V., Rajasekaran S) has been involved in the co-editing of a special issue on Health Grid: Toward Collaborative and On-Demand Healthcare in the Journal of Clinical Monitoring and Computing, Springer, Volume 19, Numbers 4-5, October 2005;
- DataMiningGrid Consortium is involved in co-editing a special issue on Data Mining in Grid and Web Services Computing Environments: Challenges and Applications in Future Generation Computer Systems in The International Journal of Grid Computing: Theory, Methods and Applications, Elsevier (to appear in 2006);
- DataMiningGrid Consortium (Dubitzky W., Mann R., Rana O) is involved in co-editing a special issue on Computational Analysis and Exploration of Distributed Data in the journal *Concurrency and Computation: Practice and Experience, Wiley Publisher* (to appear in 2006);

- DataMiningGrid Consortium (Dubitzky W.) is involved in co-editing a special issue on Critical Review of Computational Methodologies for Systems Biology in the journal Briefings in Bioinformatics, Oxford University Press (to appear in December 2006).

As a result of scientific research carried out in the project the following **scientific/technical articles** have been published:

- Assaf Schuster and Ran Wolff; *Association Rule Mining in Peer-to-Peer Systems*; Special Issue on Distributed and Mobile Data Mining, IEEE Transactions on System, Man, Cybernetics, Part B;
- Stankovski V., May M., Franke J., Schuster A, McCourt D., Dubitzky W.; *A Service-Centric Perspective for Data Mining in Complex Problem Solving Environments*; in Proc of Int'l Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'04), Vol II, pp780-787, 2004; June 2004;
- Tzachi Birk, Liran Liss, Assaf Schuster, and Ran Wolff; *A Local Algorithm for Ad Hoc Majority Voting via Charge Fusion*; 18th International Symposium on Distributed Computing (DISC). Amsterdam, October 2004;
- A. Schuster and R. Wolff; *Communication Efficient Distributed Mining of Association Rules*; Data Mining and Knowledge Discovery, 8, 171--196, 2004;
- Amir Bar-Or, Danny Keren, Assaf Schuster, and Ran Wolff; *Hierarchical Decision Tree Induction in Distributed Genomic Databases*; 1st workshop on Grid Data Mining (in conjunction with ICDM). Brighton, November 2004;
- Denis Krivitski, Assaf Schuster, and Ran Wolff; *Local Hill Climbing in Sensor Networks*; Workshop on Data Mining in Sensor Networks. Newport Beach, April 2005;
- A. Bar-Or, D. Keren, A. Schuster and R. Wolff; *Decision Tree Induction in High Dimensional, Hierarchically Distributed Databases*; SIAM Conference on Data Mining (SDM). Newport Beach, CA, April 2005;
- Swain M., Hunniford T., Mandel J., Palfreyman N., Dubitzky W.; *Modeling Gene-Regulatory Networks using Evolutionary Algorithms and Distributed Computing*; in Proc. of the 5th IEEE/ACM International Symposium on Cluster Computing and the Grid, BioGrid 2005, May 2005;
- Stahl F., Berrar D., Goncalves Silva C.S., Rodrigues J.R., Brito R.M.M., Dubitzky W.; *Grid Warehousing of Molecular Dynamics Protein Unfolding Data*, in Proc. of the 5th IEEE/ACM International Symposium on Cluster Computing and the Grid, BioGrid 2005, May 2005;
- Mark Silberstein, Anna Tzemach, Nikolay Dovgolevsky, Maayan Fishelson, Assaf Schuster, Dan Geiger; *Superlink online: distributed system for linkage analysis of large inbred pedigrees*; Israel Bioinformatics Symposium, Tel Aviv, May 2005;
- Denis Krivitski, Assaf Schuster, and Ran Wolff; *A Local Facility Location Algorithm for Sensor Networks*; International Conference on Distributed Computing in Sensor Systems (DCOSS), Marina del Rey, June 2005;

- Kfir Karmon, Liran Liss and Assaf Schuster; *GWiQ-P: An Efficient Decentralized Grid-Wide Quota Enforcement Protocol.* The 14th IEEE Intl Symposium on High Performance Distributed Computing (HPDC). Research Triangle Park, NC, July 2005;

- Swain M., Hunniford T., Mandel J., Palfreyman N., Dubitzky W., *Reverse-Engineering Gene-Regulatory Networks using Evolutionary Algorithms and Grid Computing*, in Journal of Clinical Monitoring and Computing, Springer, Volume 19, Numbers 4-5, pp329-337, October 2005;

- Huang C-H., Lanza V., Rajasekaran S., Dubitzky W., Journal of Clinical Monitoring and Computing special issue on HealthGrid – *Bridging Life Science and Information Technology*, Springer, Volume 19, Numbers 4-5, pp259-262, October 2005;

- Berrar D., Stahl F., Goncalves Silva C.S., Rodrigues J.R., Brito R.M.M., Dubitzky W., Towards Data Warehousing and Mining of Protein Unfolding Simulation Data, in Journal of Clinical Monitoring and Computing, Springer, Volume 19, Numbers 4-5, pp307-317, October 2005;

- Trnkoczy J., Kindermann J., Turk Ž., Stankovski V.; *Data Mining Requirements for Emerging Grid Environments: Digital Libraries Use Case*, Proc. of the eChallenges 2005 conference, October 2005;

- Dolenc M., Stankovski V., Turk Ž.: *Grid Technology in Civil Engineering.* Innovation in Civil and Structural Engineering Computing, (Saxe-Coburg publications on computational engineering). Stirling: Saxe-Coburg, 2005, pp. 75-96;

- Stankovski V., Kravtsov V., Niessen T., Ackermann M. and Dubitzky W.; *Towards Knowledge Discovery in Databases in Future Grid Computing Environments: Analysis and Design*; to appear in Knowledge Engineering Review, Elsevier (submitted);

- A. Bar-Or, D. Keren, A. Schuster, and R. Wolff; *Hierarchical Decision Tree Induction in Distributed Genomic Databases*; Transactions on Knowledge Discovery and Engineering. (to appear);

- Assaf Schuster, Dan Trock and Ran Wolff; *A High-Performance Distributed Algorithm for Mining Association Rules*; to appear in Knowledge and Information Systems;

- Silva C.G., Ostropytskyy V., Loureiro-Ferreira N., Berrar D., Dubitzky W., and Brito R.M.M., *ProFOUS: A Protein Folding and Unfolding Simulation Data Repository* (submitted);

- Dubitzky W., Granzow M., Berrar D. (Editors), Fundamentals of Data Mining in Genomics and Proteomics, Springer Verlag (to appear in 2nd half of 2006);

- Kravtsov V., Niessen T., Stankovski V., Schuster A.: *Service-based Resource Brokering for Grid-Based Data Mining;*GCA'06 – The 2006 International Conference on Grid Computing and Applications" conference (submitted).

4.2 Publicity material

Various publicity materials have been produced so far. The following material has been designed:

- DataMiningGrid folder;
- Basic fact sheet;
- Flyer;
- Pens printed with DataMiningGrid logotype and URL address of Web portal;
- Cups printed with DataMiningGrid logotype and URL address of Web portal.

Some of the designed material has already been produced and distributed at meetings, conferences and workshops. To date we have distributed 5 000 flyers, 800 pens and 50 cups. In the third reporting period additional 25 T-Shirts and 40 cups with DataMiningGrid logotypes were produced and distributed.

4.3 Events organized by DataMiningGrid

The following actions were undertaken:

- 12-13 May, 2004, Organized International Workshop on Grid-based Systems for Complex Problem Solving, Ljubljana, Slovenia;
- 21-24 June 2004: The 2004 International Multi Conference in Computer Science & Computer Engineering (PDPTA '04), Las Vegas, Nevada, USA. The Consortium organized two concurrent sessions: GRID COMPUTING & TECHNOLOGY – I and II [PDPTA04];
- 1-4 November, 2004: Organized Workshop on Data Mining and the Grid (DM-Grid 2004) held in conjunction with the 4th IEEE International Conference on Data Mining (ICDM'04), Brighton, UK [DMGrid04];
- 9-12 May, 2005: Membership in International Program Committee of the Third International Workshop on Biomedical Computations on the Grid (BioGrid 2005), Cardiff, UK, [BioGrid05];
- 19 October, 2005: Organized mini-conference Grid Technologies: Infrastructure and Applications, Ljubljana; Slovenia;
- 19-21 October, 2005: Organized the Workshop: Grids for Business - Challenges at the Turning Point, in conjunction with the eChallenges-2005 conference, Ljubljana, Slovenia;
- 27-30 November, 2005: Co-organized the Workshop on Data Mining and the Grid (DM-Grid 2005), in conjunction with the 5th IEEE International Conference on Data Mining (ICDM'05), New Orleans, Louisiana, USA;
- 9-12 April, 2006: Co-organizing the International Workshop on Knowledge Discovery in Life Science Literature (KDLL 2006) to be held in conjunction with The 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006), Singapore;
- 10-13 September 2006: Co-organizing and Proposal for ECCB'06 Workshop on Distributed, High-Performance and Grid Computing in Computational Biology (GCCB 2006) to be held as part of 5th European Conference on Computational Biology (ECCB 2006).

4.4 Other

Within the third reporting period Werner Dubitzky (UU) has joined the editorial board of two journals related to data mining where he represents, among other things, grid activities in connection with data mining and bioinformatics.

- Int. J. of Data Mining and Bioinformatics (IJDMB), InderScience Publishers (2005 – present): http://www.inderscience.com/;
- BMC Bioinformatics, BioMed Central, The Open Access Publisher (2005 – present) http://www.biomedcentral.com/bmcbioinformatics/ .

4.5 Events where DataMiningGrid was presented

The DataMiningGrid Partners held presentations of their project at the following conferences (the list does not include the events organized by the project):

- 16-17 September 2004: European Grid Technology Days 2004, IST-FP6 Grid Projects Launch and Concertation, Brussels, Belgium;
- 1-2 June 2005: European Grid Technology Days 2005, Concertation Meeting of IST Grid Projects, Brussels, Belgium. The Consortium was actively involved in TG5 Data Management with Michael May (FHG) as TG5 Chair and Martin Swain (UU) as presenter;
- 16 June 2005: an overview of the DataMiningGrid project was presented at a Core Grid WP2 meeting; and
- 12 January 2006: Workshop TG5: Data Management - Best practice solutions for Grid data management, Edinburgh, United Kingdom. The Consortium was actively involved in this workshop with Michael May (FHG) as TG5 Chair and Martin Swain (UU) as presenter.

4.6 Events where DataMiningGrid was promoted

The events where DataMiningGrid has been promoted are those where printed publicity material was available to delegates (usually in the form of flyers, posters, information sheets or small promotional gifts). There are obviously events where DataMiningGrid was both presented and promoted, but these are listed only in the section 'Events where DataMiningGrid was presented', to avoid too much repetition.

DataMiningGrid project was promoted at the following events:

- 30 November to 1 December 2004: e-science and data mining workshop, Edinburgh, UK;
- 21-22. February 2005: Life Science Grid Conference, Zürich, Switzerland.
- 27 June to 5. July 2005: Advanced Course on Knowledge Discovery (ACAI 05), Ljubljana, Slovenia;
- 6-8 July 2005: First SEKT Summer School on Semantic Web, Ljubljana, Slovenia;
- 8 July, 2005: Kick Off Meeting IHK Innovation forum, Bonn, Germany;

- 19-21 July 2005: 22nd CIB-W78 Conference on Information Technology in Construction, Dresden, Germany;
- 7-11. August 2005: International Conference on Machine Learning (ICML 2005) Bonn, Germany;
- 11-13. August 2005: 15th International Conference on Inductive Logic Programming, Bonn, Germany;
- 21-24 August 2005: The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Chicago, USA;
- 15-17 November 2005: European Banking and Insurance Fair, Frankfurt, Germany.

One of the DataMiningGrid Partners, Fraunhofer AIS, is the coordinator of KDNet (the Knowledge Discovery Network of Excellence). They have used the KDNet communication and dissemination channels to promote DataMiningGrid to the European and worldwide KD Community. During all KDNet events, AIS has promoted DataMiningGrid on plenary talks on grid computing or has dedicated sessions of the events to grid computing.

4.7 Contacts

The DataMiningGrid Partners collected contact names at events, meetings and via Web from the beginning of the project. These contacts will form a list of people to target with updates about DataMiningGrid and help identify the number of interested DataMiningGrid users in the future. These contacts will be used for direct e-mail campaigns and newsletters. To date a list of 107 people are e-mailed with important news on the project.

4.8 Web portal

This section objectively describes the current shape and structure of the DataMiningGrid Web portal (set up at www.DataMiningGrid.org), designed and located at University of Ljubljana.

The DataMiningGrid Web portal is aimed at presenting the widely accessible dissemination and exploitation of the project results. Visitors of the Web portal are expected to be technical and non-technical users, students and occasional visitors. The Web portal was designed to:

- Allow and find basic information about the project, technology used and Partners involved;
- Allow downloading of all public deliverables, developed software, and relevant papers/presentations about the project;
- Support the information exchange, thanks to the adoption of digital library technology.

We consider the DataMiningGrid Web portal as a powerful dissemination platform for reaching project's target audiences and communicating project achievements. The Web portal is designed according to best practices for disseminating

information and results over the WWW. We took all the requirements and remarks (especially those coming from project officer) into account to assure:

- Clear visibility of (a) the project's achievements, (b) its contributions and impact beyond the Consortium, (c) the added value at a European level, (d) the fact that the project is co-financed by the Information Society Technologies part of the EU RTD Framework Programme;
- Provision and maintenance of a frequently updated section indicating the current status of the project;
- Provision and maintenance of a section with the press releases relating to the major project achievements;
- Presentation on the role and contribution of all Partners involved.
- Ability for the public to view / download 'Public' deliverables listed in the 'Description of Work';
- Policy of updating the Web site when it comes to news, latest achievements, events, future developments etc;
- Ease of the navigation, accessibility and good degree of the cross-browser compatibility.

All Partners are responsible for the advertisement of the DataMiningGrid Web portal. The following actions were taken:

- Registration to the main internet search engines;
- Web sites of all Partners have reciprocal links to the DataMiningGrid Web portal;
- Agreements with related projects in Europe and US to have reciprocal links to the DataMiningGrid Web portal;
- Distribution of flyers, fact sheets, pens and cups with URL address of the Web portal.

The DataMiningGrid Web portal has and will be continuously updated.

Since the beginning of the project, the DataMiningGrid Web statistics have been collected. The Web portal statistics is one of the key indicators of project dissemination success. For information on Web statistics refer to Section 5.

5 Evaluation of Dissemination Activities

Evaluation and reporting on the dissemination results is essential to tune up the dissemination activities. In this section, we define methods to measure the effectiveness of the dissemination activities and provide actual measured figures on these indicators.

5.1 General indicators of dissemination success

Table 1 includes figures for the first three reporting periods of the project. It shows at a glance what the Partners need to measure and report. Some of the measured dimensions include:

- Evaluation of media coverage in terms of number of articles, interviews etc.;
- A number of information sheets prepared;
- Number of events where DataMiningGrid is promoted and/or presented;
- Number of contacts made;
- The statistics of Web portal hits.

Table 1. Indicators for dissemination success

Indicator	M1-M6	M7-M12	M18	M24
Number of conference publications	2 + 2*	7+1*	1+2*	--
Number of journal publications	1	2+5*	4+4*	--
News releases (including online)	1	2	1	--
Web portal visits	Refer to section 5.2	Refer to section 5.2	Refer to section 5.2	--
Number of information sheets	1	0	0	--
Number of events where DataMiningGrid has been promoted	5	10	1	--
Number of events where DataMiningGrid has been presented	2	3	2	
Number of events (co)organized by DataMiningGrid Consortium	3	1+2*	3+2*	--

* In preparation

5.2 Web portal statistics

One of the most relevant indicators of success of dissemination activities are the Web site statistics. Detailed statistics report can be found online on www.DataMiningGrid.org/statistics. The following statistics have been derived from the official DataMiningGrid Web portal.

5.2.1 Web site hits on well-known search engines:

Table 2 shows six well-known search engines (left column) and six DataMiningGrid keywords (top horizontal). The numbers relate to where the DataMiningGrid Web site appeared when searches were executed, for example 27 means that www.DataminingGrid.org appeared 27th item on the list. The

numbers in round '()' brackets refer to the first (M1-M6) and the ones in squared '[]' brackets to the second [M6-M12] reporting period of the project. Those numbers are included for comparison.

Table 2: DataMiningGrid Web portal statistics based on different search engines

	DataMiningGrid	Grid	Data mining	Grid problem solving environments	FP6 grid-related project
Google	(1) [1] **1**	(>200) [>200] **197**	(>200) [188] **105**	(>200) [149] **71**	(15) [19] **25**
Lycos.com	(>200) [3] **1**	(>200) [130] **>200**	(>200) [>200] **>200**	(>200) [52] **54**	(>200) [3] **10**
Altavista	(2) [1] **1**	(>200) [39] **>200**	(>200) [11] **55**	(>200) [2] **9**	(27) [6] **5**
Yahoo	(2) [1] **1**	(>200) [>200] **155**	(>200) [185] **126**	(42) [16] **9**	(6) [11] **4**
MSN Search	(2) [1] **1**	(>200) [>200] **>200**	(>200) [142] **>200**	(>200) [5] **7**	(179) [25] **10**

5.2.2 Log file statistics

The log file analysis for advanced Web statistics allows us to monitor all the Web visits and visitors. After the first 18 months of the project the following figures in were obtained (the figures in M1-M6 and M6-M12 columns are for the first two reporting periods of the project and are included for comparison):

Table 3: Web statistics for each reporting period

	M1-M6	M7-M12	M12-M18	M18-M24
Successful requests	5631	36209	24033	--
Average successful requests per day	25	202	132	--
Successful requests for pages	2769	15625	8233	--
Average successful requests for pages per day	12	87	45	--
Distinct files requested	90	622	1006	--
Distinct hosts served	500	2540	2714	--
Data transferred	46.87 megabytes	885.75 megabytes	665.77 megabytes	--
Average data transferred per day	218.89 kilobytes	4.97 megabytes	3.68 megabytes	--

The following figures show the countries of the computers, which requested files, and their share of traffic. Figure 1 shows portal statistics for first reporting

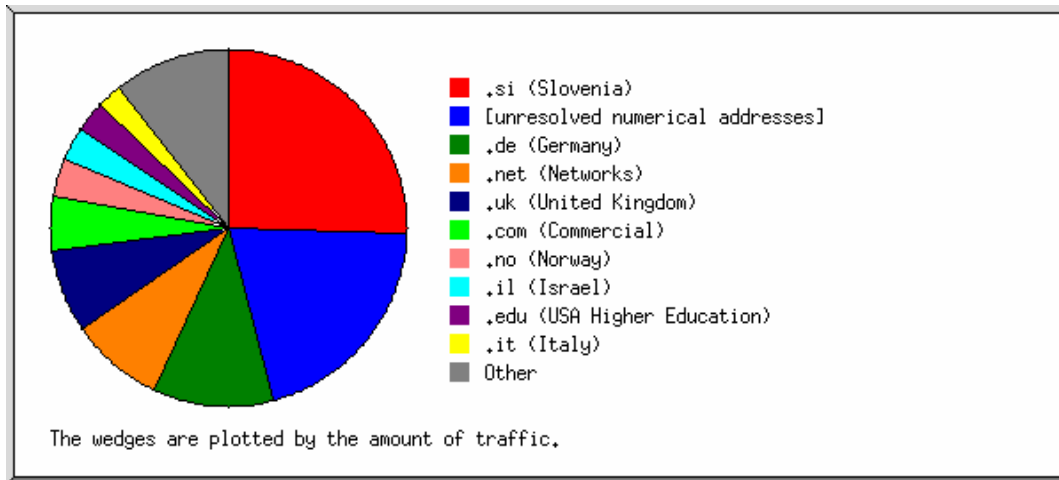period, Figure 2 for second reporting period and Figure 3 for the current reporting period (M12-M18).

Figure 1. Portal statistics by country for M1-M6 period of the project.
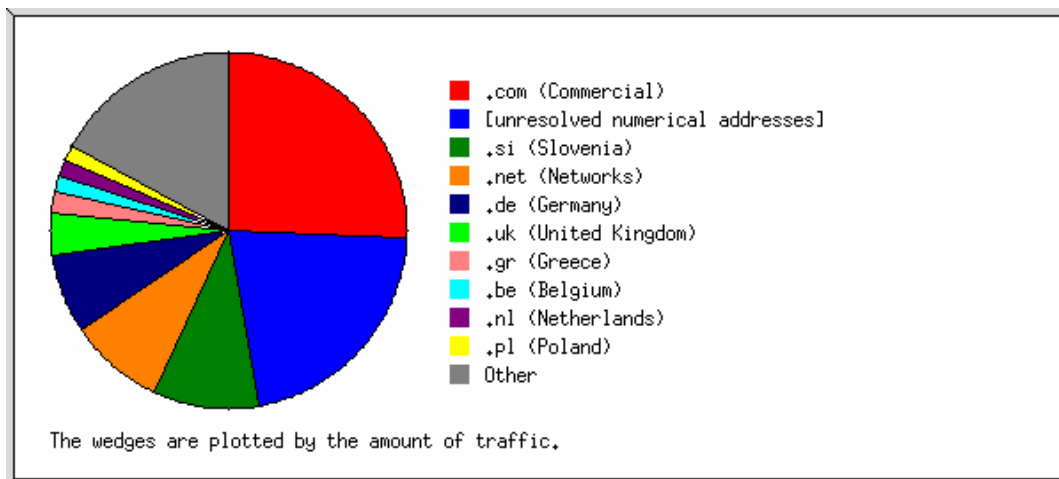
Figure 2: Portal statistics by country for M7-M12 period of the project.

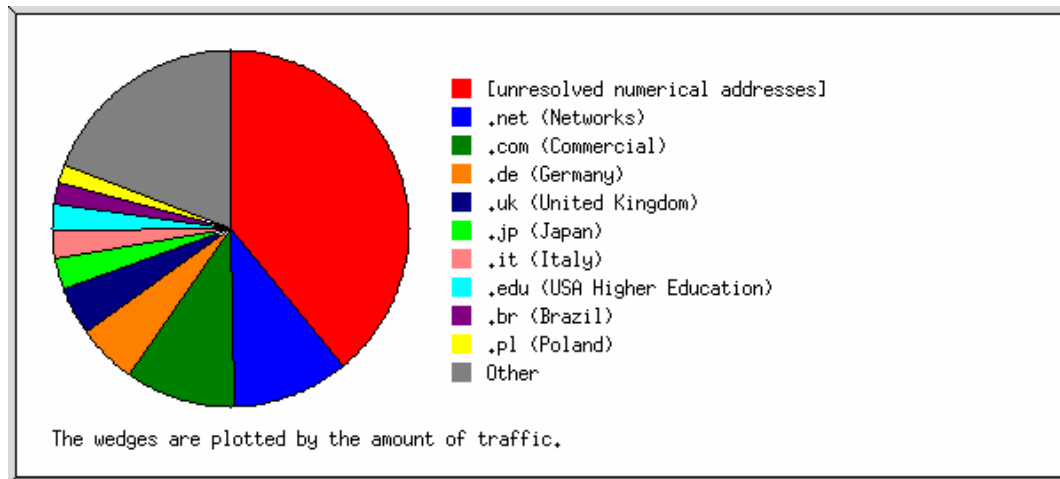The wedges are plotted by the amount of traffic.

Figure 3: Portal statistics by country for M12-M18 period of the project.

The following tables show the countries of the computers, which requested files, and their share of traffic. Table 4 contains data for the first reporting period, Table 5 for second and Table 6 for the last (third) reporting period (M12-M18).

Table 4. Quantified portal statistics by country for the first reporting period.

Requests	Domain
2158	.si (Slovenia)
790	Unresolved numerical addresses
629	.de (Germany)
382	.net (Networks)
427	.uk (United Kingdom)
362	.com (Commercial)
31	.no (Norway)
125	.il (Israel)
93	.edu (USA higher education)
83	.it (Italy)
66	.fr (France)
45	.int (International Treaty Organization)
41	.fi (Finland)
26	.pl (Poland)
25	.nl (Netherlands)
40	.hr (Croatia)
11	.bg (Bulgaria)
17	.es (Spain)
84	.se (Sweden)
11	.br (Brazil)
14	.at (Austria)
15	.pe (Peru)
12	.ee (Estonia)
14	.gr (Greece)
8	.hu (Hungary)
10	.lt (Lithuania)
17	Domain not given
10	.be (Belgium)
8	.gov (USA Government)
6	.ie (Ireland)
4	.jp (Japan)
6	.ro (Romania)
3	.kz (Kazakhstan)
6	.ru (Russia)
5	.org (non profit organisations)
16	.ca (Canada)
4	.in (India)
4	.cy (Cyprus)
2	.tw (Taiwan)
	.mil (USA Military)

Table 5: Quantified portal statistics by country for the second reporting period.

Requests	Domain
7947	.com (Commercial)
7804	[unresolved numerical addresses]
4643	.si (Slovenia)
3392	.net (Networks)
2250	.de (Germany)
1705	.uk (United Kingdom)
290	.gr (Greece)
317	.be (Belgium)
521	.nl (Netherlands)
285	.pl (Poland)
528	.it (Italy)
479	.edu (USA Higher Education)
452	.at (Austria)
383	.es (Spain)
256	.jp (Japan)
337	.ru (Russia)
377	.br (Brazil)
148	.fr (France)
244	.il (Israel)
114	.pt (Portugal)
205	.in (India)
235	.ca (Canada)
102	.pk (Pakistan)
118	.tw (Taiwan)
125	.ro (Romania)
768	[domain not given]
247	.ch (Switzerland)
70	.vn (Vietnam)
236	[unknown domain]
119	.ee (Estonia)
117	.int (International Treaty Organisations)
63	.us (United States)
53	.org (Non Profit Making Organisations)
73	.au (Australia)
86	.sg (Singapore)
35	.ie (Ireland)
41	.se (Sweden)
98	.hu (Hungary)
36	.tr (Turkey)
85	.cz (Czech Republic)

Table 6: Quantified portal statistics by country for the third reporting period.

Requests	Domain
10638	[unresolved numerical addresses]
2328	.net (Networks)
2652	.com (Commercial)
1111	.de (Germany)
836	.uk (United Kingdom)
185	.jp (Japan)
477	.it (Italy)
475	.edu (USA Higher Education)
270	.br (Brazil)
181	.pl (Poland)
257	[unknown domain]
368	.in (India)
242	.fr (France)
247	.at (Austria)
270	.si (Slovenia)
145	.il (Israel)
84	.gr (Greece)
224	.nl (Netherlands)
737	[domain not given]
154	.kr (South Korea)
28	.lv (Latvia)
177	.es (Spain)
78	.pt (Portugal)
65	.th (Thailand)
148	.ca (Canada)
77	.vn (Vietnam)
29	.co (Colombia)
49	.org (Non Profit Making Organisations)
57	.mx (Mexico)
36	.my (Malaysia)
101	.ie (Ireland)
4	.hk (Hong Kong)
70	.tw (Taiwan)
96	.au (Australia)
55	.cn (China)
70	.be (Belgium)
42	.id (Indonesia)
52	.fi (Finland)
173	.info (Informational)
90	.ch (Switzerland)

The visitor domains/countries statistics are very important feedback. The commercial visitors (with the code 'com' and 'net') are on the highest positions, which confirm the commercial's interest in DataMiningGrid project results.

The log file statistics also allows us to monitor the number of downloads of online documents. In the third reporting period the following figures were obtained:

Table 7: Requests by file.

Requested file	Number of requests
Call for papers: Future Generation Computer Systems	311
Call for papers KDLL	127
Article: Weka4WS: a WSRF-enabled Weka Toolkit for Distributed Data Mining on Grids	123
Deliverable D11(1)	91
Deliverable D11(2)	63
Focus on 6 Newsletter	60
Deliverable D71(1)	53
Deliverable D81(2)	47
Deliverable D81(1)	43

6 Critical Success Factors

In order for the Dissemination Strategy to achieve its aim and contribute to the success of the DataMiningGrid project, a number of factors need to be addressed:

- **Focus**: Everyone involved with dissemination should concentrate his or her efforts on what and who really matters. Bigger is not necessarily better;

- **Communication within the project**: It is vital that the Partners communicate with each other. Dissemination will be discussed during Management Meetings. Other means of communication (voice-conferencing, e-mail and telephone) will be used to help close the communication gap;

- **Guidance, support and assistance**: University of Ljubljana is leading the dissemination activities and as such, we are there to support, guide and assist when required;

- **Resources**: The task is huge, and the resources are small. This means that it is not always feasible to do everything we would like to do (or indeed are asked to do by those involved in other activities). We should, however, ensure that we produce quality information; and

- **Success of other parts of the project**: Of course, we can only disseminate information that is relevant, timely and progressive so we are at the mercy of other activities to make progress in their own areas.

7 Conclusions and Future Work

The dissemination activities have been running at full strength at the lead Partner site (LJU) since day one of the project. Indeed, much preparatory work was undertaken in advance of the start of the project.

The main dissemination activities undertaken in this reporting period include promoting the DataMiningGrid technology at international events, organization of a DM-Grid workshop and preparations for organization of KDLL 2006 workshop. Also, several research papers have been published at international conferences and seven journal papers have been published and submitted. One news release has also been published.

In addition to this, the corporate image was strengthened with the help of our Web portal and different promotional materials. Presentation templates are available and everyone may have a clear understanding about the look and feel of DataMiningGrid as a brand.

The main dissemination evaluation method is without doubt the Web portal statistics. The registration of our Web portal to main search engines and promotion of the Web portal through publicity material has obviously been very successful. Our Web portal is high ranking in the search engines ranked results lists (Table 2). The Web portal statistics shows a very high interest in DataMiningGrid project results. Another important thing is that large portions of these visitors are coming from commercial domains.

All Partners will be requested to continue contributing, not only statistics for reports, but also actual content and ideas. The Partners should not wait for material to be generated centrally but rather are encouraged to generate their own.

The figures supplied so far indicate a very positive outlook of dissemination activities for this reporting period. From the measurement data collected for this report it appears that the areas that specifically need improvement are number of issued news releases. We also need to increase the DataMiningGrid visibility in industrial/commercial domain. This is more likely to happen in the last reporting period, when the Consortium will have final products (software components and services) to offer.

There is a number of new dissemination activities planned for the next reporting periods, including a DataMiningGrid CD-ROM and Newsletter, which will be issued in the latter stages of the project. We are also planning to design a separate information sheet for each demonstration that will be demonstrated at the end of the project. By doing so, we will hopefully attract more industrial/commercial interest in the project. This, in conjunction with focusing on the current activities should result in an improvement in dissemination activities. The challenge for the future is to increase activity in all areas, ensuring more audiences are identified and targeted with the appropriate methods of communication.

8 References

References cited in this document are listed in the DataMiningGrid Project Manual. Most of these references can also be obtained using the digital library set up at our project Web site at www.DataMiningGrid.org under downloads | digital library.

Appendix A

DataMiningGrid on IST Results Web page (Digging deep to unlock the Grid):

 [Digging deep to unlock the Grid](#)

Unlocking the true power of the Grid for data mining is a long-cherished aim of computer scientists and researchers are making important strides to achieve that goal by developing the necessary tools and techniques.

28 Mar 2006 - [Read more](#) (below)

Unlocking the true power of the Grid for data mining is a long-cherished aim of computer scientists and researchers are making important strides to achieve that goal by developing the necessary tools and techniques.

The real power of Grid computing lies in sharing resources across a network. These can be CPU cycles, storage, peripherals, network bandwidth, data and software. Ultimately, this will lead to the grand goal envisioned by Grid researchers in which Grid users will be able to seamlessly access and harness geographically-widely distributed computing resources as if they were using a local system.

However, "trust, security, data privacy and reliability [or quality of service] in Grid computing is still a largely unresolved problem," says Dr Werner Dubitzky, Professor of Bioinformatics at the School of Biomedical Sciences at the University of Ulster and co-coordinator of the IST-funded [DataMiningGrid](#) project. "These issues are particularly important when commercial computing jobs are distributed across sites not belonging to the company that issued the jobs."

DataMiningGrid is investigating some of these problems for the specific field of, predictably, data mining. This is important for two reasons. Data mining is a technology that has been developed to analyse and interpret large quantities of data. It is one of the most powerful technologies used in astronomy, finance, and biological sciences.

One of the key objectives of the project is to build technologies that facilitate the Grid-enabling of data mining technology, ranging from data pre-processing, analysis and post-processing techniques, even if these intrinsically reside in widely dispersed locations. It is hoped that this technology will eventually help to improve the effectiveness and performance of data mining applications and provide a much wider access to data mining technology.

By using a series of mature or near mature tools to manage issues like scheduling, workflow management, and data access and integration, DataMiningGrid does not reinvent the wheel and can focus on the core problem: extracting relevant information from vast data sets across a Grid.

"We're a research project, so we're not going to be producing a commercial product. We're putting together some demonstrators to show how the tools we develop can effectively mine data across a Grid," says Dr Dubitzky. "Having said that, we are of course highly interested to bring this technology to the market using suitable exploitation channels."

The project faces several critical challenges. First, the requirements for data mining applications vary widely across different domains and sectors. To bring them all under a unified systems architecture is difficult. Second, in many data mining problems the data must remain at its source, because of the volume of data, for privacy or other reasons. In this case analysis must be executed close to where the data resides. In addition to this, one logical data set may be physically distributed across different locations. These requirements and constraints pose a significant challenge.

Dr Dubitzky says the project is on target to produce a selected set of demonstrator applications by the summer 2006, including a demonstrator for text mining. The DataMiningGrid is an important effort in realising the true potential of Grid computing.