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

D73(2)(PC2,PC3): Collaboration Report

Due date of deliverable: M12 (30 August 2006)

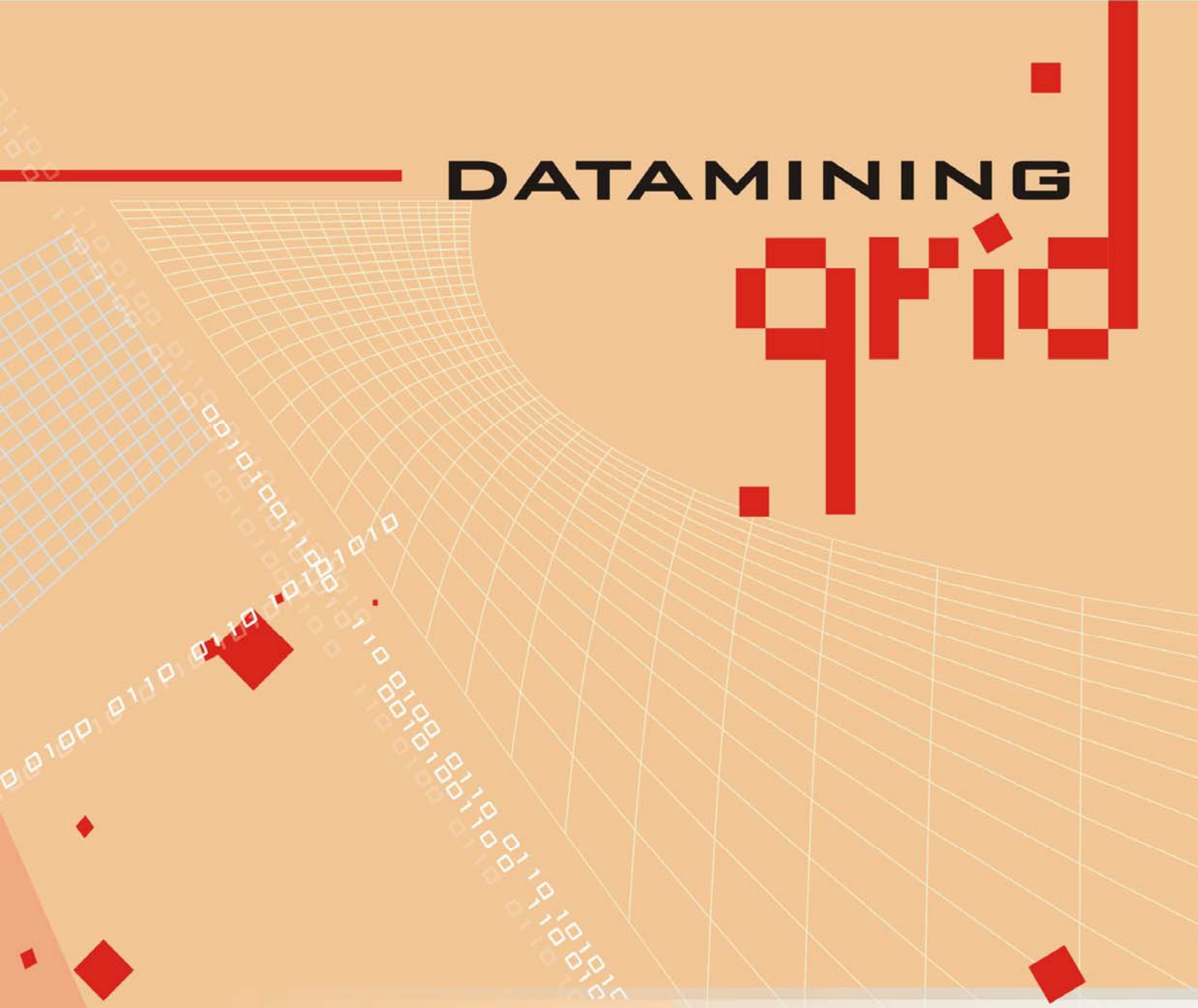
Actual deliverable submission date: 5 November 2006

Start date of project: 1 September 2004 **Duration**: 27 months

Technion Institute of Technology (TECH)

Revision: 3

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)	-



D73(2)(PC2,PC3): Collaboration Report

DATA MINING TOOLS AND SERVICES FOR GRID COMPUTING ENVIRONMENTS

D73(2)(PC2,PC3): Collaboration Report

Responsible author(s): Nahum Korda
Co-author(s): DataMiningGrid Partners

Project funded by the European Community under the 'Information Society Technology' Programme Contract IST-2004-004475

Revision history

Administration and summary		
Project acronym: DataMiningGrid		**ID:** IST-2004-004475
Document identifier:	DataMiningGrid-del-D73(2)(PC2,PC3)CollaborationReport-s	
Leading Partner: TECH		
Report version: 03		
Report preparation date: 5.11.2006		
Classification: CONFIDENTIAL (Consortium, Commission)		
Nature: REPORT		
Author(s) and contributors: Nahum Korda and DataMiningGrid Partners		
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
01.09.2006	Nahum Korda	Working	First working version
25.10.2006	Werner Dubitzky	Submitted	Final consistency checks, minor corrections
05.11.2006	Werner Dubitzky	Submitted	Finalized for submission.

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

Copyright

Citation

Nahum Korda (2006). Deliverable D73(2)(PC2,PC3): Collaboration Report. DataMiningGrid Consortium, c/o University of Ulster, 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 27-month project that commenced 1st of September 2004 and is funded by the European Commission and by 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 DataMiningGrid public web site under www.DataMiningGrid.org.

Executive Summary

In the Deliverable D72: Final Collaboration Plan we have established two parallel courses of action: (1) contribution to the concertation activities initiated by the EC, and (2) independent collaboration activities that we want to initialize directly with other EU funded projects or other research groups. Both of these directions were actively pursued during the project's lifetime. This report focuses on the concertation and collaboration activities of M13-M27.

In addition, we succeeded in establishing collaboration ties with two industrial partners (Intel, Israel and Purple Insight, UK), leading to the initial commercial application of the DataMiningGrid platform.

At the end of the project, we are issuing the Impact Assessment Report included in this document, providing an overall evaluation of the collaboration efforts in the project:

- DataMiningGrid successfully utilized concertation efforts in order to obtain all technologies that were critical for the accomplishment of the project objectives, and that were necessary for the implementation of the DataMiningGrid platform.
- DataMiningGrid successfully utilized the concertation efforts in order to minimize dissemination costs while maximizing dissemination impact.
- DataMiningGrid participated successfully in the concertation and standardization efforts, and contributed to them probably more than proportionally to its overall budget and size.
- DataMiningGrid succeeded in collaborating with industrial partners in developing initial commercial applications of the DataMiningGrid platform.

Overall, given its small size (five partners, initially, two year duration) the DataMiningGrid partners have contributed considerably to concertation and collaboration efforts. We believe that a great deal of the effort of the DataMiningGrid project will have its effects in the wider grid and data mining community, especially in Europe.

Table of Contents

1 Introduction

In the Deliverable D72: Final Collaboration Plan we have established two parallel courses of action: (1) contribution to the concertation activities initiated by the Commission, and (2) independent collaboration activities that we want to initialize directly with other EU funded projects or other research groups.

Our concertation efforts were driven by the desire to closely follow the developments and research activities in all related areas, and to avoid working solely within the limited scope of our project in a kind of research "vacuum". These activities are presented in the section 2 below.

The independent collaboration activities that we initiated were lead by our desire to re use as much of the available technologies as possible, and thus minimize the research and development efforts within the project, allowing us to focus on the essential functionalities that we want to develop. These activities are presented in the Section 3 below.

In addition, in the Deliverable D72: Final Collaboration Plan we committed ourselves to issue an overall estimation of our collaboration efforts in a Project Impact Assessment Report, following the impact assessment indicators proposed in the Concertation Task 6. This report is presented in the Section 4 below.

2 Concertation Activities

Regarding our contribution to the concertation activities, we have realized that as a STREP with limited resources we need to focus our attention to the activities that either have a direct relevance to the project's efforts, or from which the project could directly benefit.

The former group of the concertation activities includes our contribution to the technological focus T5 Data Management of the Concertation Task (CT) 1. We have exploited the fact that the T5 task leader (Michael May of Fraunhofer AIS)

is the representative of one of the DataMiningGrid's consortium partners to become more involved in the activities of this workgroup.

The latter group of the concertation activities includes CT5 Roadmap (now in CT1), CT7 Training, and CT2 Dissemination. These are the activities through DataMiningGrid could save resources by joining forces with other projects undertaking similar training and dissemination activities. Accordingly, DataMiningGrid joint forces with the InteliGrid project to organize a major dissemination event eChallenges 2005, in October 21, 2005, at the University of Ljubljana.

3 Collaboration Efforts

The collaboration efforts were envisioned in the D72 Final Collaboration Plan as evolving in five consecutive steps:

1. Scouting and intelligence gathering regarding relevant technologies, primarily those developed by other EU funded projects,
2. Decision regarding which of the parallel European projects are relevant for collaboration,
3. Establishment of a joint work plan for collaboration,
4. Joint evaluation of the collaboration results, and
5. Planning of the future collaboration activities that outlive the actual projects' lifetimes.

Following this plan, we established contacts with the CoreGrid project through CETIC who are responsible for the coordination of the CoreGrid's dissemination efforts. As the result, DataMiningGrid representative was invited to the meeting of the CoreGrid's Knowledge and Data Management workgroup held on June 15, 2005 in Barcelona, Spain. At the conference we presented DataMiningGrid's objectives and current results, invited the participants to utilize the technologies being developed in the project to their purposes, and extended our willingness to collaborate on joint data mining applications.

The collaboration with the CoreGrid's Knowledge and Data Management workgroup is of particular interest to the DataMiningGrid, since the workgroup's coordinator, Domenico Talia of the Universita della Calabria, leads a research group that developed WS wrappers that we considered using for the project.

In addition, we also established working relation with the research group lead by Ali Sheik Ali of the Cardiff University that developed similar technology.

After carefully examining both of these technologies, we decided at the present not to integrate their solutions due to the different approach we have regarding the execution of the actual algorithms. Both of these implementations execute algorithms as services installed on predefined execution machines, while we are

interested in providing such algorithms rather as an executable code that does not require any pre installation.

We successfully collaborated with WSRF Weka team leading to the implementation of the Grid-enabled WEKA services.

During period P2 (M13-M27) of the project, we established collaboration ties with two industrial partners, which lead to the initial commercial implementation of the DataMiningGrid platform.

Some Partners of the consortium (notably FHG, TECH, and UU) have also been successful in bidding for new grid projects in FP6 IST Call 5. The new projects where these partners are involved included Chemomentum (www.Chemomentum.org), QosCosGrid (www.QosCosGrid.eu) and ACGT (http://test-acgt.healthgrid.org/).

4 Project Impact Assessment Report

The Concertation Task 6 proposed 11 indicators for the overall impact assessment of the projects developing Grid technologies. These indicators are grouped into 7 categories:

1 The cooperation among initiatives in the member states and EC initiatives has increased. (a) Number of steps taken and actions initiated, and (b) ERA Coordination.

2 The influence of members of the ERA in the global grid community has increased. (a) Degree of involvement in Grid Standards Community (GGF and OASIS), (b) Downloads of European Grid Middleware.

3 Interaction between academic and industrial partners has increased. (a) Participation of commercial representatives in European grid events, (b) Number of commercial products including grid technology from European Projects has increased.

4 Degree of grid penetration in Complex Problem Solving and new application areas. (a) Percentage and absolute number of applications using grid technology, (b) Number of grid enabled computing resources, which are made available by vendors and academia.

5 Grid-enabled collaboration within business communities (like distributed supply chains) has increased. (a) Number of grid enabled analysis methods, problem solving environments and workflow tools in selected application areas (engineering and pharma), which are relevant for industrial usage.

6 The capability and functionality of next generation grid toolkits and middleware has been increased. (a) Number of new products resulting

from the GRID IST projects.

7 New grid generic toolkits and middleware makes the application of grid technology possible and easier. (a) Evaluation reports by external users of the new products, tools and environments.

We decided to utilize the same indicators, in order to evaluate the concertation efforts undertaken in the DataMiningGrid project. The results are presented in the

Table 1. Project Impact Assessment

Indicator		Assessment	Comment
1	Number of steps taken and actions initiated.	Following the dissemination and concertation plans issued at the beginning of the project, we organized a major joint event with the IntelliGrid project.	DataMiningGrid and InteliGrid joint event: eChallenges 2005, October 21, 2005 University of Ljubljana
2	ERA Coordination.	No active participation.	
3	Degree of involvement in Grid Standards Community (GGF and OASIS).	Project's representative actively participated in ETSI (GRID Work Group)	
4	Downloads of European Grid Middleware.	Open source DataMiningGrid services and tools will be available under Apache License V2 also at SourceForge.net.	
5	Participation of commercial representatives in European grid events.	One relevant event organized with 2 commercial representatives.	DataMiningGrid and InteliGrid joint event: eChallenges 2005, October 21, 2005 University of Ljubljana
6	Number of commercial products including grid technology from European Projects has increased.	The project factually provided two commercial applications in partnership wth industrial partners.	Intel, Israel and Technion - Israel Institute of Technology (Grid monitoring application), PurpleInsight, UK and University of Ljubljana (DataMiningGrid tools and WSRF-compliant services)
7	Percentage and absolute number of applications using grid technology.	10+	WEKA, R, Spin!, Lagramge 2, DC ontology learning (DaimlerChrysler), Text mining (Fraunhofer AiS), Gene regulatory networks, any data mining algorithm has the potential to be grid enabled in a very short process which takes 20 minutes of the DataMiningGrid application developers' time, DataMiningGrid Application Enabler (in conjunction with the WSRF-compliant services), DataMiningGrid client side software

			components, DataMiningGrid Web-based client
8	Number of grid enabled computing resources, which are made available by vendors and academia.	4 Globus Toolkit 4 installations with underlying Condor pools (20-64 machines in each pool).	The testbed will be open for participation of external organizations after the termination of the project.
9	Number of grid enabled analysis methods, problem solving environments and workflow tools in selected application areas (engineering and pharma), which are relevant for industrial usage.	4	Grid-enabled WEKA, Grid-Enabled R, DaimlerChrysler text-mining and ontology learning applications, Lagramge 2 (partial differential equation discovery software). In addition, any data mining algorithm has the potential to be grid enabled in a very short process which takes 20 minutes of the DataMiningGrid application developers' time.
10	Number of new products resulting from the GRID IST projects.	The project provided two commercial applications during the project lifetime, although many more commercial applications could be built using open source DataMiningGrid services and tools.	
11	Evaluation reports by external users of the new products, tools and environments.	Evaluation by both commercial partners using the first DataMiningGrid applications.	Intel, Israel and PurpleInsight, UK.

Summarizing the project impact indicators, the following could be concluded:

- DataMiningGrid successfully utilized concertation efforts in order to obtain all technologies that were critical for the accomplishment of the project objectives, and that were necessary for the implementation of the DataMiningGrid platform.
- DataMiningGrid successfully utilized the concertation efforts in order to minimize dissemination costs while maximizing dissemination impact.
- DataMiningGrid participated successfully in the concertation and standardization efforts, and contributed to them probably more than proportionally to its overall budget and size.
- DataMiningGrid succeeded in collaborating with industrial partners in developing initial commercial applications of the DataMiningGrid platform.

5 Conclusions

Overall, given its small size (five partners, initially, two year duration) the DataMiningGrid partners have contributed considerably to concertation and collaboration efforts. We believe that a great deal of the effort of the DataMiningGrid project will have its effects in the wider grid and data mining community, especially in Europe.