



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(1) (PC2,PC3) Collaboration Report

Due date of deliverable: M12 (31 August 2005)

Actual deliverable submission date: 5 October 2005

**Start date of project:** 1 September 2004

**Duration:** 24 months

Technion – Israel Institute of Technology (TECH)

Revision: 6

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
<b>PU</b>	<b>Public</b>	<b>X</b>
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**

**qr10**

**D73(1) (PC2,PC3) Collaboration  
Report**



# DATA MINING TOOLS AND SERVICES FOR GRID COMPUTING ENVIRONMENTS

D73(1) (PC2,PC3) Collaboration Report

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

## Revision history

Administration and summary	
<b>Project acronym:</b> DataMiningGrid	<b>ID:</b> IST-2004-004475
<b>Document identifier:</b>	DataMiningGrid-del-D73(1)(PC2,PC3)CollaborationReport-s
<b>Leading Partner:</b> TECH	
<b>Report version:</b> 06	
<b>Report preparation date:</b> 05.10.2005	
<b>Classification:</b> PUBLIC	
<b>Nature:</b> REPORT	
<b>Author(s) and contributors:</b> Nahum Korda and DataMiningGrid Partners	
<b>Status:</b>	- Plan
	- Draft
	- Working
	- Final
	<b>X Submitted</b>
	- 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
01.09.2005	Nahum Korda	Working	First working version
15.09.2005	Werner Dubitzky	Submitted	Final consistency checks, minor corrections
16.09.2005	Alice McQuillan, Werner Dubitzky	Submitted	English grammar and expression, formatting

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](http://www.DataMiningGrid.org/dissemination).

## Copyright

This report is © DataMiningGrid Consortium 2004. Its duplication is restricted to the personal use within the Consortium and the Commission.

## Citation

Nahum Korda (2005). Deliverable D73(1) (PC2,PC3) Collaboration Report. DataMiningGrid Consortium, c/o University of Ulster, [www.DataMiningGrid.org](http://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 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](http://www.DataMiningGrid.org).

## Executive Summary

In the deliverable D72 (PC1) 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 initialise directly with other EU-funded projects or other research groups.

Regarding concertation activities, we have joined CT5 Roadmap (now in CT1), CT7 Training, and CT2 Dissemination efforts, and became actively engaged in T5 Data Management of CT1 through our representative coordinating this task.

Regarding the independent collaboration activities, we have established a working relationship with CoreGRID's *Knowledge and Data Management* workgroup (their WP2), as well as with the research groups at the Università della Calabria, and Cardiff University. The latter two research groups developed technologies that we examined and are considering their application in the DataMiningGrid project.

Both of these courses of action proved useful during the first year of the project and significantly advanced our efforts. Accordingly, it is our intention to continue pursuing both of these courses of action. We expect that our involvement in the joint dissemination, training and roadmapping activities and concertation workgroups will increase during the second year, following the availability of the project outcomes.

## Table of Contents

Executive Summary .....	6
Table of Contents .....	7
1 Introduction .....	8
2 Concertation Activities .....	9
3 Collaboration Efforts .....	10
4 Conclusions and Plans for the Next Period .....	11

## 1 Introduction

In the deliverable D72 (PC1) 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 initialise 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.

Some conclusions from our collaboration experience, and the collaboration plans for the second year of the project are presented in the Section 4 below.

## 2 Concertation Activities

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

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 which the DataMiningGrid project could save resources by joining forces with other projects, undertaking similar training and dissemination activities. These activities will become more significant during the second year of the project, when the project results will be available.

In addition, we adopted the indicators proposed in CT6 in our collaboration plan, and intend to issue an evaluation report using these indicators at the end of the project.

### 3 Collaboration Efforts

The collaboration efforts were envisioned in the D72 (PC1) 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,
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 that are responsible for the coordination of the CoreGRID's dissemination efforts. As a result, DataMiningGrid representatives were invited to the meeting of the CoreGRID's *Knowledge and Data Management* workgroup held<sup>1</sup> on June 15, 2005 in Barcelona, Spain. At the conference we presented DataMiningGrid's objectives and current results, an invited the participants to utilise the technologies being developed for their projects, 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 this project.

In addition, we also established a working relationship with the research group lead by Ali Shaikh Ali at Cardiff University they have developed technology, which is complementary to our project.

After carefully examining both of these technologies, we decided at present not to integrate their solutions due to their 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.

In the future, as we progress with the implementation of our wrappers, we anticipate difficulties where the experience of the aforementioned groups (especially WSRF Weka) might be useful. Therefore, although we cannot utilise their services at present, we are not rejecting the possibility of future collaboration.

---

<sup>1</sup> The group meeting was held during the *CoreGRID Industrial Advisory Board Meeting*.

## 4 Conclusions and Plans for the Next Period

Grid is an evolving technology, and data mining on grid is in its infancy, just making its initial steps. Accordingly, there are many useful and interesting initiatives and projects that seem closely related to the ambitions and objectives of the DataMiningGrid project. However, on closer examination it becomes evident that fundamental differences in approaches and intentions differ, as well as a variety of underlying technologies and implementations, thus making re-use extremely difficult, and often impossible.

Nevertheless, from our collaboration efforts it has become obvious that the exchange of hands-on experience with the emerging technologies, and brainstorming on fertile research directions are absolutely necessary to further advance our grid research efforts in general. Possibly, the most important aspect of the concertation and collaboration activities is the spontaneous forming of clusters from various European research groups with common research agendas, and their joint planning of future research directions, and collaboration on joint projects.

Both courses of action that we outlined in our Collaboration Plan (i.e. joining the existing concertation activities, and initiating our own collaboration activities) proved useful and significantly advanced our efforts. The representatives of the Consortium members that are directly involved in the various concertation activities expressed their satisfaction with the fact that researchers have an opportunity to meet in person, exchange information on their research efforts and ambitions, and discuss future research directions.

Accordingly, it is our intention to continue pursuing both of these courses of action. We expect that our involvement in the joint dissemination, training and roadmapping activities and concertation workgroups will increase during the second year, following the availability of the project outcomes.