Periodic Report for Review

Grant Agreement number: **619209**

Project acronym: **AMIDST**

Project title: **Analysis of MassIve Data STreams**

Funding Scheme: **FP7-ICT-2013-11**

Date of latest version of Annex I against which the assessment will be made: October 4th 2013

1st X 2nd ☐

Period covered: **from** January 1 2014 **to** December 31 2014

Project co-ordinator name, title and organisation:

Associate Professor Thomas Dyhre Nielsen

Aalborg University

Tel: +45 9940 8853

E-mail: tdn@cs.aau.dk

Project website address: www.amidst.eu

1. Publishable Summary

1.1 Summary description of project context and objectives

Today, omnipresent sensors are continuously providing streaming data on the environments in which they operate. For instance, a typical monitoring and analysis system may use streaming data generated by sensors to monitor the status of a particular device. Analysis and monitoring systems should be designed to make predictions about the future behaviour of the device, or diagnostically infer the most likely system configuration that has produced the observed data. Sources of streaming data with even a modest updating frequency may produce extremely large volumes of data, thereby making efficient and accurate data analysis and prediction difficult. This calls for scalable data analytics.

The objective of AMIDST is to develop a toolbox providing a scalable framework that facilitates efficient analysis and prediction based on information captured in large volumes of streaming data. The work includes developing and scaling up existing algorithms in order to make the AMIDST toolbox flexible and versatile enough as to cope with the needs and requirements of a wide variety of applications. AMIDST will provide a generic framework for analysis of extremely large volumes of streaming data, thereby adding, creating and increasing the value of existing and new data resources as well as providing a means for more timely and efficient decision making.

Each of the three industrial partners of the AMIDST consortium provides a real-world use case with complex data for test and evaluation of the developed framework. One solution is under development for each use case provider. Each use case represents one domain of commercial exploitation of effective solutions whereas the general framework will be applicable to a wide range of other domains. With the objective of creating a strong positive synergy, AMIDST takes an integrated European approach and joins partners with high interests in probabilistic modelling methods as well as techniques and algorithms for analysis of extremely large data volumes.

The AMIDST consortium has a strong and balanced combination of research and industrial partners. The academic partners ensure a scientific approach to theoretical and methodological aspects of the project. The industrial partners illustrate the importance of the potential developments provided by AMIDST for the EU economy, as they represent four strategic EU areas: software development, the automotive industry, energy, and finance.

1.2 Description of work performed and main results

The work in AMIDST is organised into ten work packages (1) Requirements analysis and Evaluation, (2) The AMIDST framework for representation and reasoning, (3) Exact and approximate inference, (4) Learning AMIDST models from data, (5) Developments in the HUGIN Software Tool, (6) Solution for Manoeuver Recognition in Highway Traffic, (7) Real Time pattern recognition in drilling logs, (8) Efficient and accurate risk prediction for credit operations, (9) Dissemination and exploitation and (10) Management. Work packages 1 to 5, which are focusing on the development of the AMIDST framework and supporting tools, were launched early in the project period. This work has included a thorough analysis and specification of the requirements of the industrial use case providers. The requirements of the use case providers drive the developments of the framework and their feedback will be taken into account throughout the development process. In addition to identifying the requirements of the three use case providers, the process has produced a generic method for requirement engineering in small projects. The development of the

general AMIDST modelling framework has also been initiated. This includes the specification of both static and dynamic modelling components, where the design of these components is driven by the requirements and application scenarios defined by the use case providers.

At month 6, the work on the use case solutions (work packages 6 to 8) were officially launched. Work package 6 started work already in month 3 in order to take advantage of available resources and the availability of a prototype vehicle. The three use case work packages have worked on instantiation of the AMIDST framework, data preparation and identification of necessary software developments to support the AMIDST work.

The work on the requirements engineering, the development of the framework and supporting tools and the solution for the use case, has already produced new methods and algorithms, and papers have been published in the proceedings of scientific conferences and the methods and algorithms have been or are being implemented in the AMIDST toolbox.

So far, the AMIDST project has produced the following results:

- A project fact sheet describing the AMIDST project
- A quality assurance plan for the work performed
- A requirements document for the use cases
- Four conference publications based on the technical work already performed have been published and presented at the targeted conference. One journal paper has been published.
- Media exposure through news articles, press releases, web announcements.

A dedicated project website with the URL www.amidst.eu has been developed. The website is used for both internal and external dissemination of information related to the AMIDST project. The website contains all public information on AMIDST including links to public deliverables, press releases, publications, presentations, etc. A LinkedIn page for AMIDST has also been created and it is actively being used to publish news on the progress of the AMIDST project.

1.3 Expected final results and potential impacts

The AMIDST consortium has a strong and balanced combination of research and industrial partners. The academic partners ensure a scientific approach to theoretical and methodological aspects of the project. The industrial partners, on the other hand, illustrate the importance of the potential developments provided by AMIDST for the EU economy, as they represent four strategic EU areas (software development, automotive industry, energy, and finance).

An important final result of AMIDST is the AMIDST toolbox. It will provide a scalable framework that facilitates efficient analysis and prediction based on information captured in streaming data. The framework is based on probabilistic graphical models. The work will include scaling up existing and new algorithms for probabilistic graphical models to cope with large data sets. The toolbox will have functionality for learning probabilistic graphical models from historical data as well as for performing inference in probabilistic graphical models.

The establishment of an open source community to support the future development and usage of the AMIDST toolbox is an important objective of the project. Also, an AMIDST handbook is produced describing procedures and guidelines on how to analyse extremely large volumes of streaming data using the AMIDST toolbox.

The AMIDST research project includes three use cases provided by the industrial partners of the project as described below. The three use case solutions are used to rigorously test the framework on real and complex data. Each use case represents one domain of commercial exploitation of effective solutions whereas the general framework will be applicable to a wide range of other domains. The use cases defined by three of the industrial partners represent an important contribution to their application domains, and correspond to manoeuver recognition in traffic situations, risk prediction and profiling in financial institutions and pattern recognition for oil drilling logs.

1.1.1 Manoeuver recognition in highway traffic

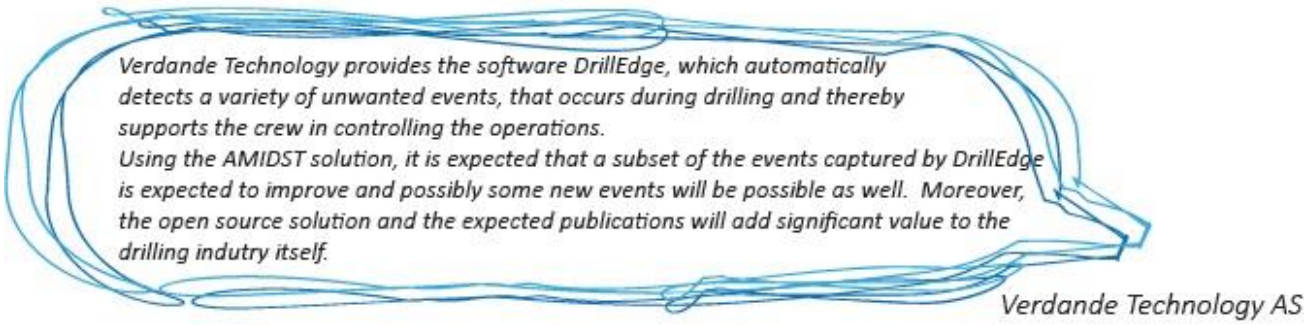
Daimler AG, a world leading car manufacturer, provides this use case and its data.

Improved manoeuver recognition in traffic situations has a high benefit to society in terms of reduction of traffic accidents, less injuries and fatalities, reduction of congestions and thus less environment pollution and better traffic flow leading to a positive society impact.

1.1.2 Real time pattern recognition in drilling logs

Verdande Technology AS, a Norwegian software development company, provides this use case and its data.

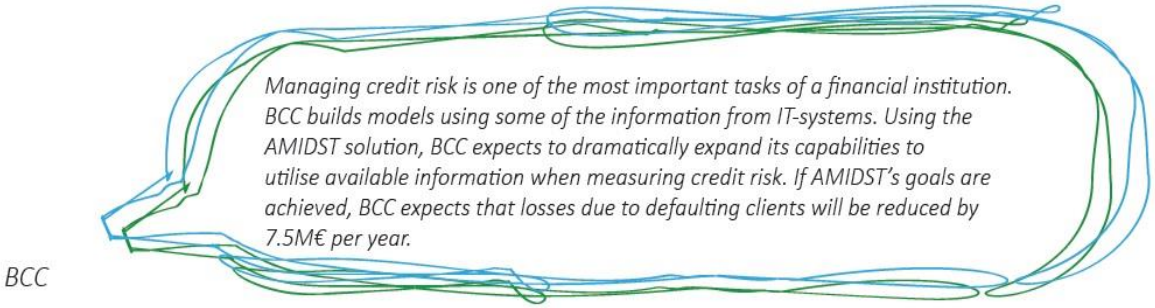
A real time pattern recognition for oil drilling logs will reduce the fraction of drilling time that is non-productive. It will also reduce the dependence on manual interpretation of drilling data to avoid drilling problems.



1.1.3 **Efficient and accurate risk prediction for credit operations**

BANCO DE CRÉDITO SOCIAL COOPERATIVO S.A. (BCC) A Spanish bank, provides this use case

An efficient and effective solution for the risk prediction problem in banks can be crucial to reduce losses due to inefficient business procedures. The risk prediction solution can be used as a tool for monitoring the evolution of customers, in terms of credit operations risk, as well as to design marketing campaigns based on risk profiling. Solvency of institutions will be increased by applying this solution.



BCC