Large Scale Integrating Project

Grant Agreement no.: 257899

D1.2 – Requirement Analysis Report and Specification of the SmartVortex Assessment Framework

SMART VORTEX –WP1-D1.2

Project Number	FP7-ICT-257899
Due Date	September 30th 2011
Actual Date	September 29th 2011
Document Author/s:	• Hans-Ulrich Heidbrink, HHk, ICT • John Lindström, JML, LTU • Magnus Löfstrand, ML, LTU • Matthias Johansson, MJ, Alkit • Stephan Lukosch, SL, TUD • Tore Risch, TR, UU • Dominic Heutelbeck DH, FTK • Jens Grabarske JG, FTK • Arne Byström, AB, HAG • Vahid Kalhori, VK, SAN • Martin Helgosson, MH, SAN • Daniel Wedlund, DW, SAN • Jonas Larsson, JL, VOL • Stefan Radomski, SR, TUD • Mssimo Mecella, MM, ROM • Felix Engel, FE, FUH • Stefan Knoll, SK, TUD • Jordan Janero da Silva, JJS, TUD • Andreas Hundsdörfer, AH, ICT • Alessandro Russo, AR, ROM
Version:	0.6
Dissemination level	Public with confidential Annex D2.1A
Status	Final
Contributing Sub-project and Work package	All partners in the project
Document approved by	HHk, ICT

SEVENTH FRAMEWORK PROGRAMME

Co-funded by the European Union

Document Version Control			
Version	**Date**	**Change Made (and if appropriate reason for change)**	**Initials of Commentator(s) or Author(s)**
0.1	05.6.2011	Initial structural Draft with copy of the use cases and scenarios from D1.1 document and KPI structure	HHk
0.2	18.8.2011	Restructuring regarding recommendations from the Pre-Review Meeting in Luxemburg	HHk
0.3	21.9.2011	KPIs from TUD , Volvo CE, Changes from Hägglunds; Extension of introduction	HHk, SK, JL, AB, JL
0.4	23.9.2011	Additions to the collaboration use cases, KPIs for WP8 and ISP-1, Executive view incomplete draft, rewrite of ISP-1 case	DH, SK, JL, FE, HHk
0.5	25.9.2011	Detailed Hägglunds use case incorporated KPIs improved	AB, JL, HHk, SK
0.6	27.9.2011	Preliminary version for internal review	HHk
1.0	28.9.2011	Final Version for RTDC Review	HHk

Document Change Commentator or Author		
Author Initials	**Name of Author**	**Institution**
HHk	Hans-Ulrich Heidbrink	InConTec GmbH

Document Quality Control			
Version QA	**Date**	**Comments (and if appropriate reason for change)**	**Initials of QA Person**
V 0.6	30/09/2011		RTDC

Catalogue Entry

Title	SmartVortex Requirement Analysis Report and Specification of the SmartVortex Assessment Framework
Creators	Hans-Ulrich Heidbrink
Subject	Deliverable SmartVortex D1.2
Description	
Publisher	
Contributor	
Date	September 30th , 2011
ISBN	
Type	Delivery SmartVortex
Format	
Language	English (US)
Rights	SmartVortex Consortium

Citation Guidelines

EXECUTIVE SUMMARY

Along a McKinsey report on "**Big Data**", analyzing large data sets will become a "key basis for competition, underpinning new waves of productivity growth, innovation and surplus as long as the right policies and enablers are in place". McKinsey sees the power of "innovating new business models, products and services" and "supporting of human decision making" in the computation of "Big Data" [MCK01]. A similar development can be seen regarding data streams.
In industrial applications **high volume data streams** are generated by simulations, sensors or actuators etc. in for instance machining operations or collaboration processes, which can interact with the previously mentioned data streams.

The SmartVortex project handles some aspects of data stream management in industrial environments within three integrated sub-projects (ISP-1/2/3), based on the industrial partners' visions of the innovative **Functional Product** business model. In this document, we describe and elaborate on the findings from the requirement gathering phase of the SmartVortex project.

A focus on impact oriented requirements for solutions to be applicable for different industry segments requires the consortium to look for sustainable results whilst going beyond the state-of-the-art research. The selected use-cases provide a high innovation potential in the areas of the following main topics:

- **Processing of (i.e. search in, analysis of** , and computation on):
    - Large data streams generated by sensors in equipment, simulations, collaboration activities (cross organizational, cross team, and cross domain) etc.
    - Inferred data streams, i.e. resulting from processing other data streams
    - Non-streaming data sets generated in product development or product life-cycle management

- **Visualization** of the computational results like trend, risk and threshold validation in a new graphic interface also supporting querying data streams

The capturing of tractable data should render pertinent information to be used in collaboration and decision support, related to developing, supporting, maintaining, or monitoring Functional Products.

The project is targeting three integration sub-projects (ISPs) with one or more use cases, which are briefly described below:

- ISP-1
    - Use case 1 – (Perspective FE-Design) Usage of simulation data streams generated by FEM simulations which are integrated with design tools in a collaborative engineering environment to reduce the design cycle time and the all overall simulation time for very time extensive simulation procedures with a high volume of visualization results. Trend estimation calculation integrated in the data streaming process will allow to cut simulation time while entering in a collaboration and optimization process which includes designers and changes at the simulation model or stimuli

    - Use-case 2- Hägglunds drives, recording of real-time data streams, transfer using storage media or telecommunication solutions and computation of the stored data streams while reading the stream from the storage media

- ISP-2
    - Use case 1 – Sandvik Coromant, usage of different types of data streams in machining to support collaboration between engineers and customers, monitor equipment, optimize customers' usage, and assure quality in simulation models and optimization methods

- Use case 2 – Hägglunds Drives, usage of data streams to monitor hydraulic motors and related equipment, and to learn collaboratively more about system performance and behavior prior to problems
- Use case 3 – Volvo CE, usage of data streams in construction equipment to monitor usage, performance, service need, and enable collaboration during service/maintenance

- ISP-3
  - Use case 1 – All industrial partners, usage of data streams to support collaboration in a complex cross-organizational scenario. The scenario focuses on a fictive use case where all consortium partners use data streams to collaboratively develop and monitor a cross-organizational product.. This collaborative process needs to provide trust, security, and intellectual property protection for the data streams involved.

Industrial partners' business motivations accompanied by the use cases have generated test-beds for the validation of the research results. The industrial partners have provided requirements and use cases supporting business goals such as:

- Productivity gain during the simulation process
- Higher reliability, availability and risk of damage reduction by improving the preventive maintenance process and offer of service oriented products based on intangible properties (Hägglunds Drives)
- New products and improvement of existing products by data acquisition and validation to enter in new functional engineering methods (Sandvik Coromant)
- Service oriented business models that guarantee higher availability with risk limitation for the provider (Volvo CE)

With openness towards research and implementation, the consortium partners (including the industrial partners) have shared strictly confidential information about many types of measurements, data sources, and data streams involved while envisioning new business models moving towards selling functions. The main drivers for the research activities are the industrial partners' interest and willingness to enter into new business models such as selling a function. These business models, which are based on intangible assets, open up new business opportunities and motivate to start using new functional engineering methods building on data stream processing technology.
The identified scenarios use cases, and data streams involved, are described in the following chapters. In addition, described are also the proof of concept for generic industrial applicability, the validation, and measures of success

TABLE OF CONTENTS

1 INTRODUCTION

1.1 The SmartVortex Vision

In the last years there a trend in the industry emerged which fosters the offer of intangible assets within vendor-customer relation even if tangible assets represent a high volume and a very precious part of the delivery. These underlying business models are called "Service Oriented Models" where the product offer is no longer the equipment itself, but the contractual agreement to specific performances of intangible features like "transportation volume, reliability, physical parameters like torque, power, cutting speed or other measurable parameters" which are taken as contractual basis for between vendor and customer. While leasing and rental models still calculate the financing of the used equipment based on price list items which are seen as goods service oriented models define the goods only as means to an end to reach a specific business goal.

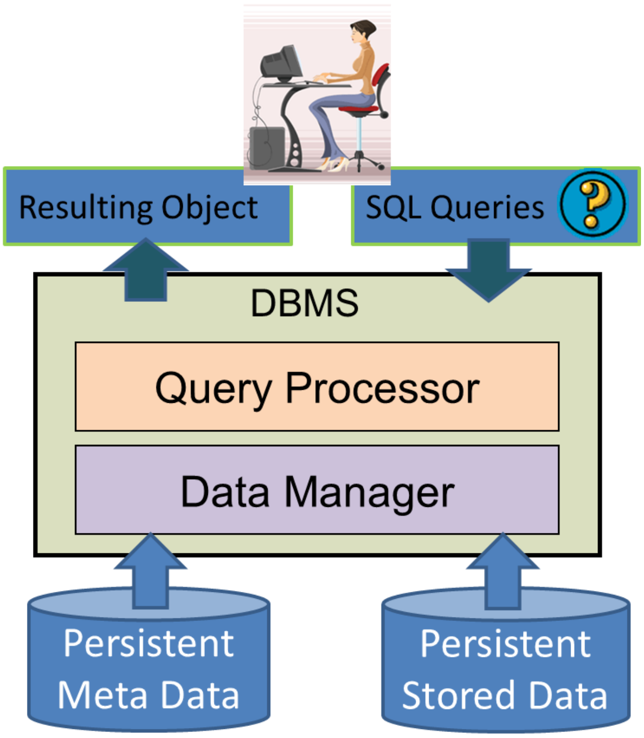
To enter in service oriented business models in the manufacturing industries requires exact knowledge about the behavior of the goods, the equipment or the systems, their usage conditions for early detection of upcoming failure modes or maintenance requirements and all dependencies within the delivered systems. Additionally it requires a new way of engineering and design which is called "Functional Engineering". The design requires much higher efforts in simulation and validation of the goods prior to production to reduce the risk of failing once such business models get applied. One very significant change in the design methodology established functional engineering due to the split of design tasks and input parameters from each other used for the design and construction activities. The use and applicability of the parameters that will be taken for the design requires better knowledge about risk factors, criticalities, boundary conditions and system behavior. Another contributing way to reduce design circle time is to shorten the time spent for simulation and validation without decreasing the quality of the results. Especially simulation cycles where the trend shows quite early the expected results can be shortened and optimized.
This deeper understanding of functional dependencies is combined with measurement of actual data derived from test systems or out of predicted simulation values which are taken to validate the design and operation behavior of the systems and to optimize the used simulation and design models applying the achieved results. To get representative and exact predictions – produced with either high volumes of actual sensor measurement data or with a high volume of data generated from similar systems – the results need to be reviewed and analyzed. One method which reduces time and efforts is based in a technology where data streams produced either by life operating sensors and or derived from computation activities which have produced extremely large data files are validated during the streaming mode to recognize immediately threshold violations, risk situations, uncontrolled usage patterns which require immediate action and care and uncritical values that can be excluded from further investigations.

A core objective of the SmartVortex is to use this data effectively for supporting collaboration to improve the product lifecycle. Sustainable collaboration is a critical skill and competence in organizations. As systems and products become more complex, experts and engineers need to collaborate to manage integration and communication of the subsystems they design and develop. Parallel to this, users, clients and stakeholders are becoming increasingly often included in various phases of a product lifecycle, including design, user feedback, innovation and improvement cycles. Collaboration in the product's lifecycle can be supported by technology; e.g. information systems for planning and controlling the production process as well as for archiving, administrating and providing product related data. In complex design and engineering phases, different tools and techniques are useful at different point in the design. However, groups of engineers often do not have the knowledge to choose or select effective tools and techniques for specific situations. Furthermore, groups might not have the skills or knowledge to effectively appropriate the tools available to them to easy get relevant information out of the data streams. SmartVortex addresses those existing gaps in collaboration and will provide intelligent support for the selection and appropriation of collaboration support tools based on a group's current interaction and process phase and the availability and accessibility of information.

In the domain of SmartVortex, organizations are not isolated, but are members of federations of partners in which data is shared and tasks are worked on collaboratively across-organizational borders. Data exchange between organizations nowadays is quite commonplace and explicit calls are used to place orders, enquire about business deals or send other data which may be important in a B2B scenario. However, on the technical side, these calls are often isolated in nature and lack therefore the support of a full-blown middleware solution.
SmartVortex has the vision of a cross-organizational middleware, meaning, a framework which can automatically connect organizations seamlessly while still protecting the intellectual property of the partners and minding the different company policies. Instead of proposing a monoculture where every partner has to adopt central concepts, every organization is free to use its own terms, definitions and structures which will automatically be translated or mapped for the other partners - or hidden if other members of the federation are not allowed to see the data.
Using this mechanism, SmartVortex allows for a more fluid and agile user experience in collaboration sessions, while still being able to guarantee the highest standards regarding data protection and privacy

To improve the acceptance of application engineers that are not programming specialists easy access to the data which concern the design engineers is required and need to be addressed to speed up the design and validation process. Support given by graphical access methods can solve these deficits and allow the engineers to concentrate on their mainstream activities of design and validation. SmartVortex addresses those existing gaps in the current design methodology and will open the door to novel service oriented business models reducing risk for the industries and minimizing the resource consumption to make the European industries more competitive.

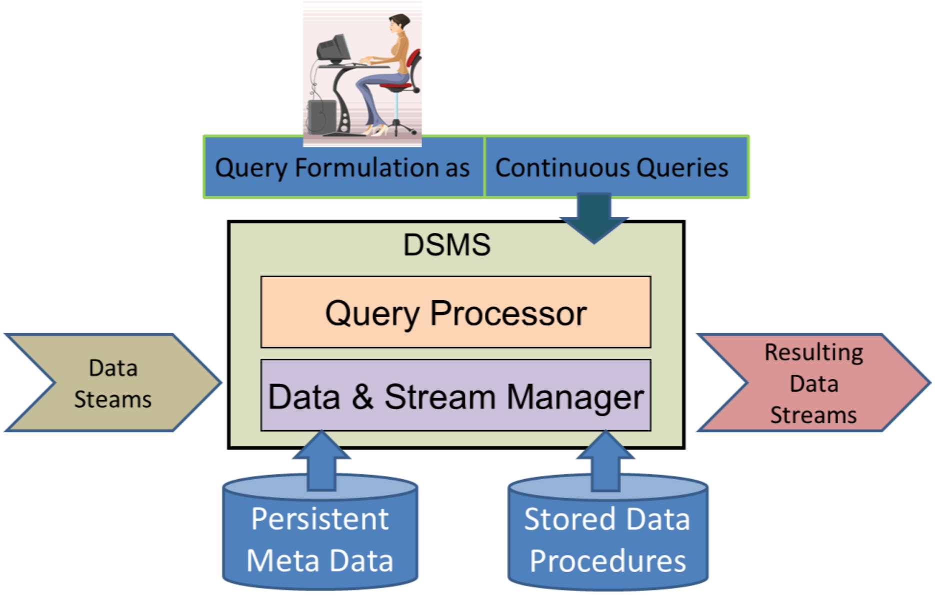
1.2 Data Base Management Systems (DBMS) and Data Stream Management Systems (DSMS)



While a DBMS acts on persistent data or relations were each single query creates a random access on the actual content of the data base, in DSMS systems the data streams are transient and there continuous queries allow sequential access on streaming objects to produce enriched or calculated resulting data streams for visualization or prognostic representations.

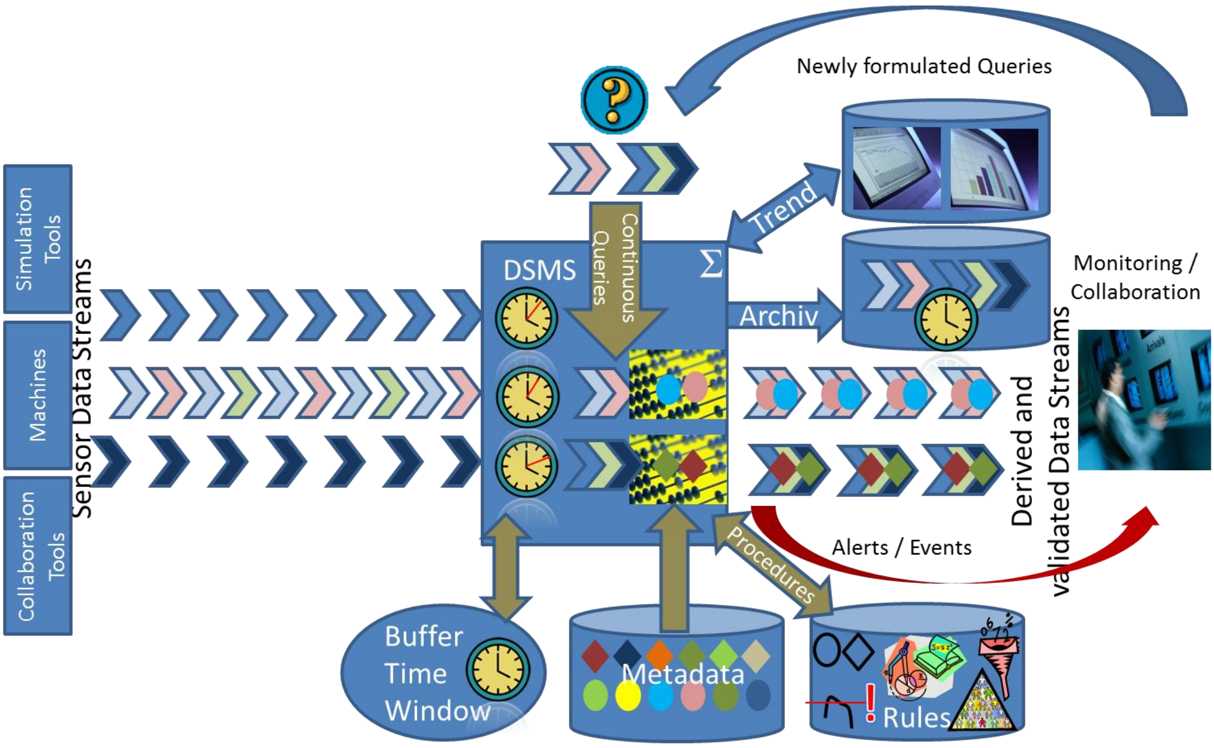
Picture 1: Data Base Management System (DBMS) Principle

While data in DBMS systems get managed in relations (tables) in a DSMS data streams get added as basic data objects. Those data streams can be envisioned as long running continuous series of time and value pairs. Each formulated query is acting continuous on those data streams. In the DSMS value pairs are computed by procedures based on stored rule sets and stored metadata.



Picture 2: Data Stream Management System (DSMS) principle

The entire process applying a DSMS technology in the three different kind of raw data streams for the selected use cases in simulation, real-time sensor monitoring and collaboration or hybrids of those streams is shown in the following picture. This picture simplifies the underlying DSMS technology. In SmartVortex this technology will be built in as a federated system which gives higher flexibility, scalability and enables parallel computing, while providing integration with systems that relay on DBMSs like PLM, ERP or other relational oriented systems. Furthermore, DSMS functions can be installed in a distributed environment. Please see D2.3 and D3.2 for the architecture and the chosen data model.



Picture 3: Data streaming process

1.3 Objectives and Tasks of SmartVortex

The goal of SmartVortex is to provide a technological infrastructure consisting of a comprehensive suite of interoperable tools, services and methods for the intelligent management and the analysis of massive data streams produced in all phases of the product and design life cycle to achieve better collaboration and decision making in large-scale collaborative projects concerning industrial innovation engineering. Data streams will be used for validation, behavior prediction, simulation and simulation model improvement during all phases of the product life cycle and especially for the steady improvement of the offered services, combined with product enhancements, improvements which are used to perform those service offers and collaboration data to make design decisions traceable and improve the communication between the stakeholders.
To evaluate the different application oriented and technology views that the project has been taken we refer to Annex 1: Views on data streams

1.4 DSDM and PLM

Product lifecycle management (PLM) technologies can be defined as a collaborative, integrated and information-driven approach that includes people, procedure and technology The objective of these technologies is to support organizations in planning and controlling their product lifecycles by providing methods and tools for information and process management as well as for the integration of enterprise software. Besides the management of data and processes, PLM provides tools for collaboration among networked participants in product value chains.
Due to the changes in technologies, business and economy, today’s organizations are confronted with new challenges for their product lifecycle (e.g. to design more complex products or to support geographically dispersed design teams). Form the literature, this paper identifies different needs for future PLM solutions. For example, Ming et al. indicates the need for new technology solutions to support collaboration across multi-organizations and virtual teams. Here, intelligent support can be used to manage the collaboration activities that are tailored to the special needs of global and virtual teams. Abramovici [MA07] analyzed the future trends of given PLM solutions along the product lifecycle. He sees main weaknesses of given PLM solutions in the poor support of product lifecycle activities outside the production phase and missing industry standards for PLM meta-data models and for PLM processes.
According to new technologies and sensors that can be used to generate data streams from products in use, SmartVortex identifies a further potential in the use of product data streams to improve PLM in all phases of the product lifecycle. For example, organizations increasingly provide support and maintenance for their products. Product support requires products that are tagged or wired with sensors to monitor their performance. Monitoring and analyzing product data streams could lead to intervention to use insights to improve the design, and to respond quickly to failure. However, often the complexity of products requires a collaborative approach and analysis of different perspectives.
Collaboration support has extensively been studied in various research domains such as groupware, group (decision) support systems, concurrent design tools and group facilitation. As a result, different types of collaboration support technologies are developed with different aims based on different domain perspectives. For instance, multi-criteria decision making tools are developed to compare design alternatives and make design choices, while collaborative modeling tools are developed to support analysis and building shared mental models. In complex design and engineering phases, different tools and techniques are useful at different point in the design. However, designers often do not have the knowledge to choose or select effective tools and techniques for specific situations. Furthermore, groups might not have the skills or knowledge to effectively appropriate the tools available to them. Therefore, support in adopting and using collaboration support tools might benefit collaborative design. In conclusion, SmartVortex indicates several challenges for a PLM technology that supports collaboration by using product lifecycle data stream:

1. *Detect the need for collaboration support*: a PLM technology needs to provide tools and methods to analyze data streams from product lifecycle to detect the need for collaboration;

2. *Finding the right information*: a PLM technology needs to provide tools and methods to filter relevant information from the product lifecycle data streams, information which is needed to support collaboration;

3. *Organization and management of collaboration:* a PLM technology need to provide tools and methods that are efficient but at the same time rigorous to manage the quality of collaboration.

2 DETERMINATION OF THE USER BASE

2.1 General policy analysis survey of users

One of the core elements of the SmartVortex architecture is the technological infrastructure for capturing, communicating and processing measurement data from sensors and actuators in different kinds of technical equipment. In SmartVortex, the technical equipment providing the sensor data are products of the industrial partners Sandvik Coromant, Hägglunds Drives, Volvo CE and Philips (i.e. cutting tools, electro hydraulic drive systems, construction equipment and Finite Element simulation). The specific types of sensors to be studied are described in detail for each of the use cases in the confidential document D1.2A.

2.2 Analysis of basic requirements of the different competence areas related to a common infrastructure

For all sensor data streams, in all use cases, there are a number of basic requirements that need to be fulfilled in order to be able to develop a common technological infrastructure for data capture, communication, processing and analysis. Identifying and analyzing these common requirements will improve the opportunities of developing a technological infrastructure general enough not only to be applied for all of the project's use cases, but also for other applications, making it possible to exploit the results of the project in a wider industrial scope.
The basic requirements of the SmartVortex platform with regard to real time measurement data streams originating from sensors and actuators have been divided into four categories as follows: data capture system requirements, data communication network requirements, backend data processing system requirements and user interface requirements.

3 METHODOLOGY ADOPTED TO IDENTIFY PROJECT REQUIREMENTS

The initial phase of the SmartVortex project has been used to discuss, analyze and document the needs of the industrial consortium members using a top-down approach. Starting from the use scenarios, use cases have been identified and furthermore resulting requirements. This evolutional process allowed checking all upcoming requirements on general utilization ability and market conformity. Only that requirements that could be envisioned as useful in other companies and industry domains have been selected given the commitment of the industrial partners to provide test beds that can show the benefits of the created solutions. While some of the business models, which have been taken from the industry as reason to invest in future new state-of the art technology, need to be kept secret, exact data descriptions for the use cases have been put in confidential documents. Nevertheless the kinds of data which are generating the benefits in the future demonstrators have not been disclosed for publication.
The entire process and the chosen methodology including the recognized use scenarios are described in detail in Annex 2 and 4. The determination of the user base including base requirements are described in Annex 3.

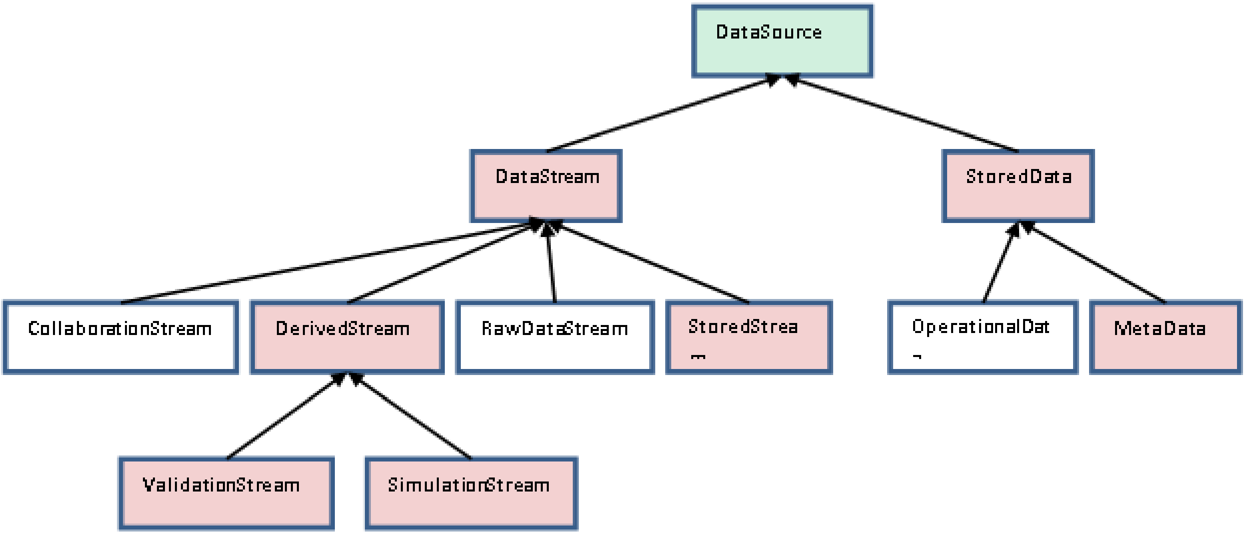
4 USE CASE DESCRIPTIONS FOR THE INTEGRATED SUB-PROJECTS 1-3

4.1 ISP-1 Data streaming computation from stored or generated file sets

There are two use cases which have been identified working either with data streams deriving from stored file sets containing data streams or from simulation tools generating data streams

4.1.1 Use-Case ISP-1.1 Data streams derived from complex simulations in the virtual design process

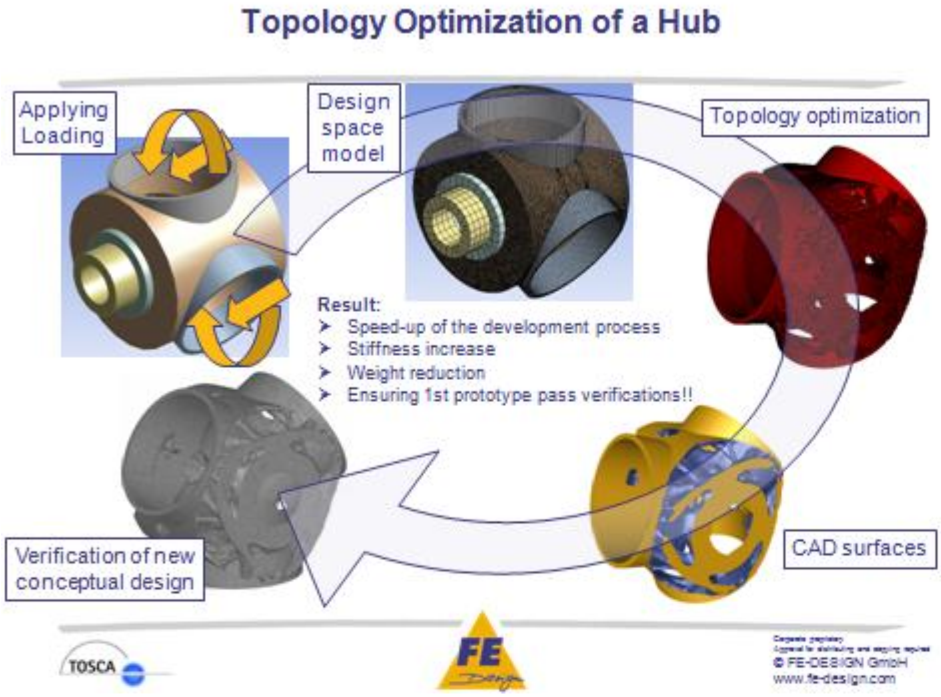
The use case works with the following elements (highlighted) of the DSDM schema:



Picture 4: Data stream schema with highlighted elements participating at use-case ISP-1-1.1

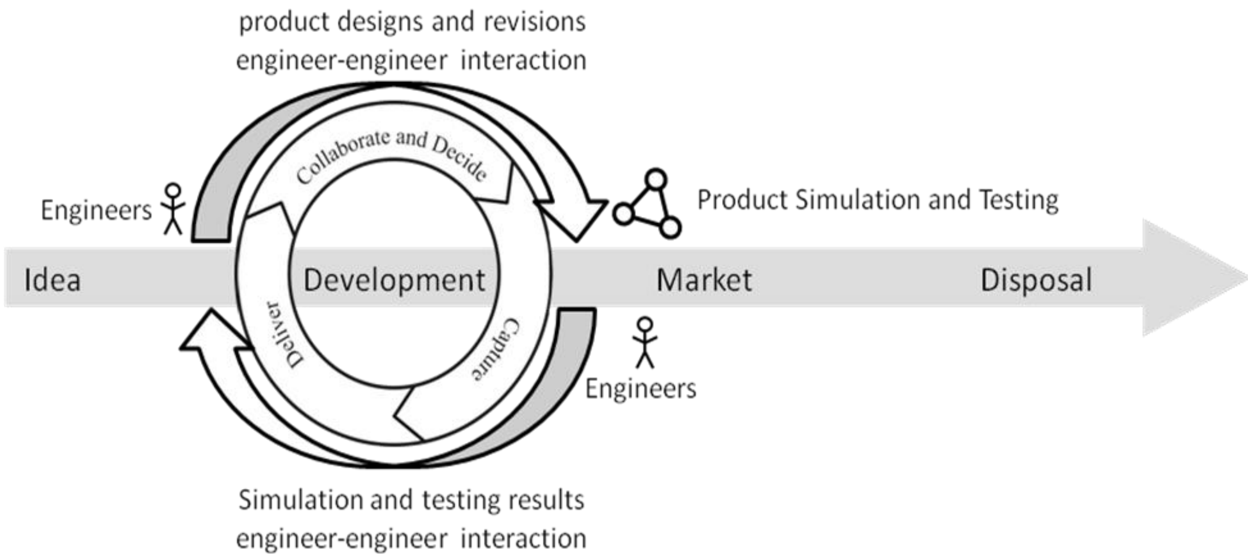
4.1.1.1 Scenario

Today's companies are driving their product development with sophisticated processes and applications, to reduce time-to-market of their products. Computer aided development and production is a given, as often in some industry sectors the time from product idea to the time when it hits the market is even longer than the time on the market itself. Therefore in every phase of the product lifecycle not only software applications are used to design the product, but simulation approaches are established to verify the correctness of physical behavior long before the first prototype of the product is existing physically.



Picture 5: Process of simulation and optimization of a component

Traditionally a design concept will be developed and a prototype will be designed. A more modern approach is to introduce optimization methods into the design process. With these it becomes possible to easily develop prototypes which fulfill all physical requirements. With this the number of design cycles can be reduced and therefore the time-to-market and the development costs.



Picture 6: The producers and consumers of product data streams in ISP-1.

Nowadays the developed products do not only have to fulfill one requirement. In most cases the component has to be effective and optimal for different needs which lead to a number of requirements like minimal weight with a stiffness constraint and also a given dynamic behavior. In most cases the idea of the designer is to fulfill all requirements at the same time but in many cases it is not possible to fulfill all requirements at the same time.

The optimization process is an automated iterative process based on a number of Finite Element simulations and an intelligent optimization approach. The complexity of the optimized components leads to requirement to create simulation models consisting of millions of elements. Due to the complexity of the models and the huge amount of data produced during the process the optimization is normally performed on a compute cluster with a high performance. During the process the optimization system has to decide for each small element (like a "Lego brick") if the element has to remain in the structure or if it has to be removed.
The success of the process is dependent on the possibility to find a valid solution which fulfills all the given requirements. In the ideal situation the constraints are chosen correctly so that the optimizer is able to find a configuration of material which is valid. The existence of a solution cannot be predicted before the iterative optimization process is executed. Within the first iterations it is up to the user to perform some kind of a trend analysis based on the current situation which can be monitored from the simulation results and the current material distribution.
If the optimization fails or if it can be ensured that the optimization process will not be efficient enough then a change of the simulation and optimization model is necessary. For this all involved people from different departments involved into the definition of the requirements of the developed components have to communicate which constraints can be modified (to be less restrictive) to allow the optimizer to find a valid solution. Also the simulation models have to be changed in some cases if e.g. the optimization progress show that some physical boundary conditions are not transferred correctly to the optimization model.
This requires collaboration by people from different domains (e.g. acoustic, mechanical and production engineers) to come up with a set of optimization and simulation settings leading to an optimal design proposal from the simulation and optimization system.
The SmartVortex infrastructure will be used to build solutions that support the monitoring of the process (extraction of data, plots and visualization), the collaboration and the decision making process of engineers.
By employing the DSMS and by using intelligent feature extraction methods it will be possible to control the progress and the potential of the optimization process via and advanced user interface. The overall engineer-engineer interaction will be supported by tools for generating shared understanding between different engineering domains. In addition, by capturing user interaction and activities, such as executing certain simulations or the identification of socially contextualized decision making events, the design rationale of product data and thus important knowledge will be preserved.
As a result of such analysis activities, design changes, or adoptions will take place and the optimization process will start again to enable the optimization system to find a design proposal.

4.1.1.2 Raw data

The raw data produced by most of the FEM simulation tools are ASCII or binary file containing measurement unit, scale, frequency, stimuli, mesh designators and x,y,(z) coordinate pairs representing the resulting stress

4.1.1.3 Metadata

Metadata is data about data, and is needed for the understanding of the analysis of the data collected. The metadata which are used in this use case relate to the specifics of the simulation models, the structure of the simulated object and the applied technology and application oriented rules and standards.

4.1.1.4 Derived data

The derived data stream is defined by the usefulness of the simulation results. The simulation results are pre-scanned and specific patterns, curve shapes and threshold violations and trend assumptions are selected to complement or reduce the derived data stream.

4.1.1.5 Validation

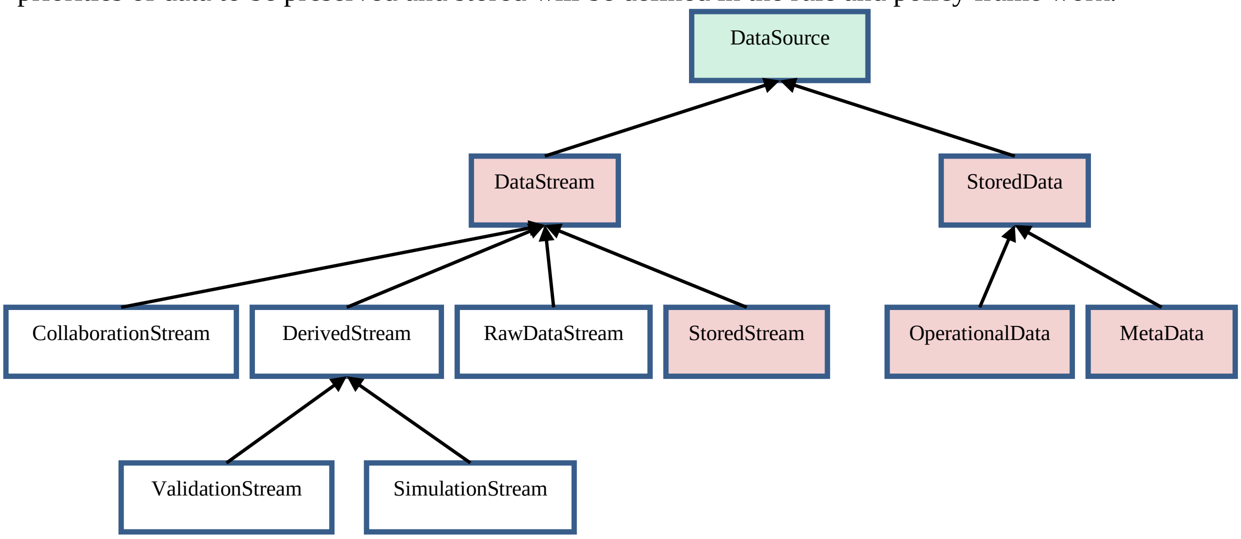
In the area of simulation it especially important to calculate trends and predictions related to given model specific threshold values. Furthermore it is important to recognize, why the violation occurred like what is the difference between expected value and calculated prediction value. This kind of calculation will be performed in this use case.

4.1.1.6 Action

Collected data are used when analyzed to trigger actions or alerts, either to notify design engineers start collaboration processes or to automate actions in the development environment. Actions in the simulation use case are built to create variances in the simulation related to input stimuli, thresholds or selection of exactness to the applied simulation models. Generally a collaboration process gets started including the participating designers to point out the cause of the deviation and perspective solution for changes to improve the design. Furthermore, in What-If scenarios input parameters can be virtually changed to elaborate the impact of changes to actual design constraints.

4.1.2 Use-case ISP-1-1.2 Data stream computation of data streams that have previously been stored in file sets (Use Case provided by Hägglunds Drives and Volvo CE)

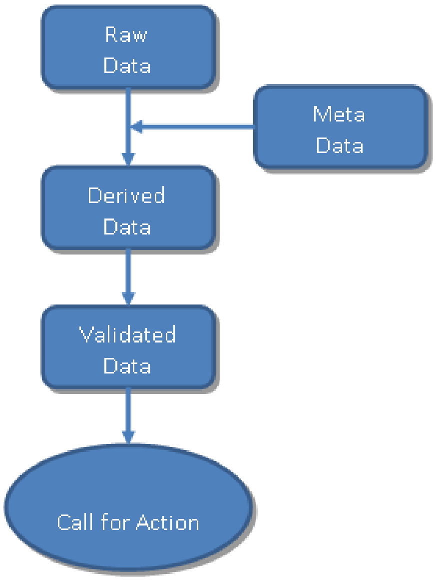
As already mentioned in chapter 3.2 at both companies it can be happen that the transmission situations requires to store the data in either large file repositories or dump them from random access memories. The structure of the data remains identical as described in the following chapters 5.2.1 – 5.2.3. The priorities of data to be preserved and stored will be defined in the rule and policy frame work.



Picture 7: Data stream schema with highlighted elements participating at use-case ISP-1-1.2

4.2 ISP-2 Computation of real-time data streams

Based on the findings in the ISP-1 subproject a general methodology was chosen to describe the requirements of integrated subprojects in the same way to recognize progress in the project and to allow the different consortium members to contribute following standardized schema. Picture 8 shows the basic workflow concept which has been followed establishing this requirement document.



Picture 8: Basic Data stream computation workflow with resulting action event

From a data modelling perspective the workflow schema which shows in application oriented way which workflow elements are combined a cannot be taken to be directly implemented in a data model for the data stream management system. This application oriented schema need to be applied to the technological view o data streams, which is shown in the following picture 9.

Formulated information requirements from the industrial business cases related to real-time data streams show that basic data from sensors, actuators and other types of units are not enough to meet the information needed to be able to offer services and functions to the customer in a new way. Some of this required information must be based on data calculated and derived from the basic data.

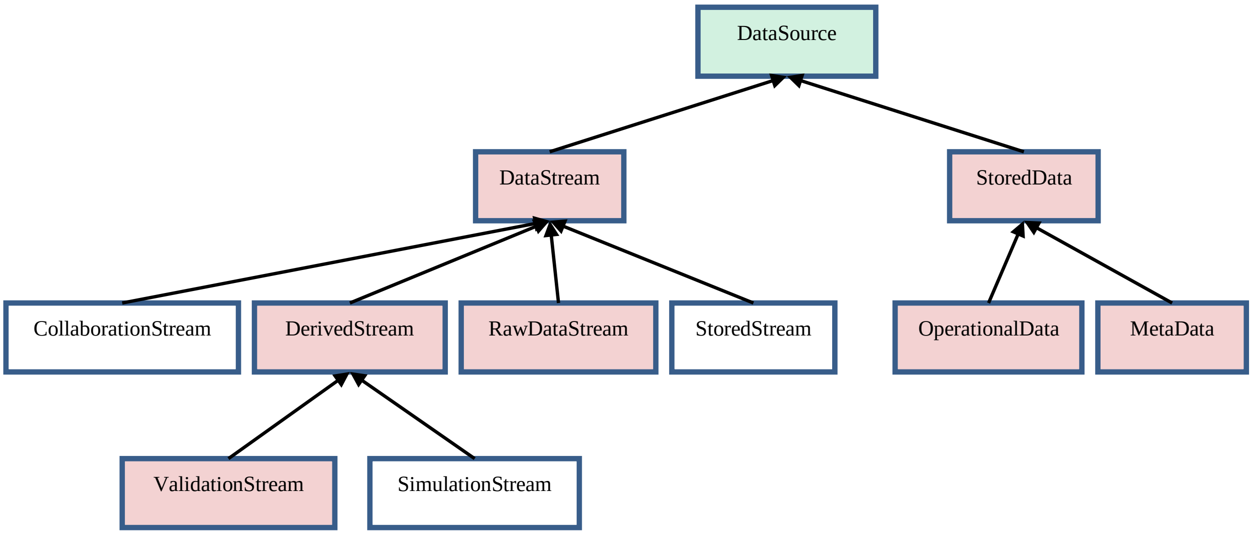
When it comes to large amounts of data the derived data has to be calculated in real-time to be able to deliver the information within time requirements for the user as a designer or analyst.
Some examples of derived data pointed out as important from the industry requirements gathering work are

- vibration levels over a certain time frame in a machine, measured in a specific scale,
- sound levels over a specific time where the operator is working in the cockpit of a machine,
- the build-up of hydraulic pressure in certain time frames, such as when shifting gears in a large machine,
- considerable deviating data for analysis of operator behavior and possible reasons for sudden machine collapse.

These information needs are based on the ability to propose new service and function-oriented business offers to the customers. These derived data streams have to be matched with other data to be able to design machine parts that will fit a set of services for the customer in a satisfactory way.
ISP-2 is also demonstrating the increasing complexity of the ISPs regarding the cross-organizational nature of operation. In ISP-1 the processes mostly reside within a single organization, and the SMART VORTEX technologies are uses in order to improve the local process. In ISP-2, the data streams are originating from a different organization than where they are consumed in order to execute the business model. The participating organizations need to be very careful about which data to share. ISP-3 will illustrate even more complex distributed scenarios.

4.2.1 Use-Case ISP-2-1: Real-time data streaming and computation from actual sensor data in hydraulic systems (Use case provided by Hägglunds Drives)

The most important factors for Hägglunds drive systems are **high availability** with no unplanned stops and best possible **Total Cost of Ownership** (TCO) for the customer. An important factor to improve the TCO is to increase the energy efficiency by optimization of the operating point. However this means higher usage of the system components where Hägglunds Drives needs the online monitoring tools to follow the system behavior to keep the high **reliability** and to improve the **maintenance possibilities** for increased system availability.



Picture 9: Data stream schema with highlighted elements participating at the use cases in ISP-2

Hägglunds Drives wants to improve the knowledge of system performance in customer installations in order to learn about system behavior prior to problems. To get statistical data, also some data from system monitoring needs to be stored for trend analysis. The expected benefits are:

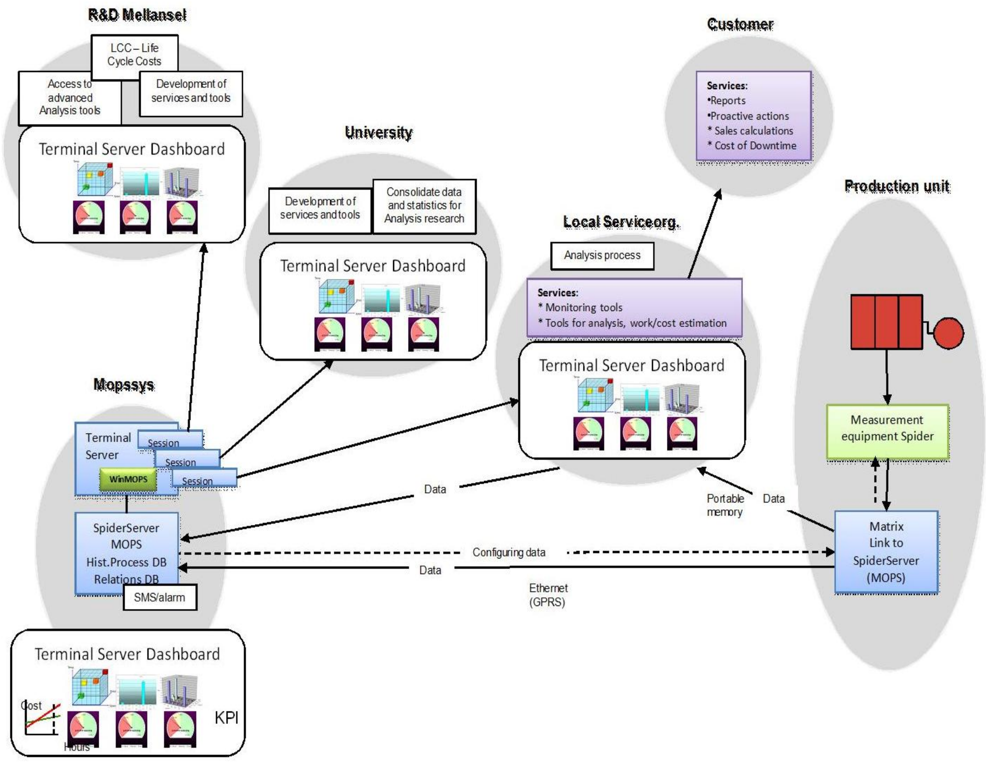
- Central consolidation of process information for analysis
- Collected process data can be used for dimensioning/capacity calculations
- Understanding of system behavior in customer installations
- Status monitoring via Key Performance Indicators (KPIs)
- Addition of analysis capability to support proactive maintenance
- Possibility to integrate process data with data from maintenance system
- Reduce time to fulfill and create reports about service level agreements
- Standard interface available for customers to access their data

The online monitoring tools also gives the additional effects for further application and system knowledge, improved proactive fault monitoring, improved system monitoring, extension of system knowledge data base, and additional management and operational level reports using real-time data.

Hägglunds Drives has a vision to provide functional sales based on hardware, software and services where there is a need to monitor system drive parameters online where data stream management will be crucial.

Picture 10: **Hägglunds Drives pilot environment (first stage)**

Initially, a working concept using a shredder machine at Hägglunds Drives will be developed and verified. The next step is to test the concept at a customer site in a real operational environment. Picture 10 shows the shredder machine and Picture 11 the schematic set up for the information communication with the customer site. The shredder machine uses the same schematic set up.



Picture 11: **Hägglunds Drives information handling environment (second stage)**
(Copyright – Hägglunds Drives)

4.2.1.1 Scenario

4.2.1.1.1 Objectives, Tasks:

Defining DSMS queries for fault identification in Hägglunds Drives (HDAB) Electrohydraulic drive system. The drive is designed to power a shredder, crushing waste wood. The drive cycle of the drive is intermittent and subjected to quite varying and high loads. Finding failure in time will increase the availability of the drive system since unplanned repair or maintenance can be avoided. The objectives are to develop methods for efficient and scalable search of product data and meta-knowledge produced during the entire life-cycle of a product, and further to develop collaboration environments supported by decision making systems.

4.2.1.1.2 Requirements to describe and analyse:

Better monitoring of critical systems and parameters enables to in advance be able to detect and predict failures avoiding interruptions and damages. Parameters to monitor are for instance effect distributions of system parameters over time (pressure, temperature, rpm, leakage flow, vibrations etc.). After identification of critical parameters, the system will be analyzed and modeled to improve system availability.
As an example, an increase in oil temperature in the motor casing or an increase in the loop temperature at the low pressure side port versus high pressure port will indicate an imminent failure.
The temperature difference between inlet and outlet oil are at each timestamp initially calculated and evaluated against a predetermined set-point.
Whenever the set-point value is exceeded the DSMS will by a sliding window evaluate the slope of the failure sequence and determine if the slope is acceptable or not. First, the slope are determined and set through engineering judgment, once the algorithm within the DSMS is trained with previous monitored fault data and online data the algorithm will improve.
Suggested sensors for failure identification can be viewed in the confidential document D1.2A, recommended placements of sensors in the simplified hydraulic scheme in the following picture 12.

4.2.1.2 Raw data

The raw data is planned to be collected from various sensors and data sources at adequate frequencies (i.e. with such granularity that the data collected is useful and reveals variations or abnormalities). The data relate to the Electro-hydraulic drive system condition and operating behavior.
L= ABS(sensor NR22-sensorNR23) %ABS = absolute value%
D=sensorNR20-(sensorNR22 or NR23)

4.2.1.3 Metadata

Metadata is data about data, and is needed for the understanding of the analysis of the data collected. The metadata to be collected and compiled describe the Electro-hydraulic drive and pump system and the usage of the motors including environmental data. Specific business model related parameters are part of this data set which reflect the usage history and the part identification for elements within the Electro-hydraulic drive system. Meta data in this use case are e.g.
X1= Setpoint1 for L (L is defined below)
X2 = Setpoint2 for D (D is defined below)

4.2.1.4 Derived data

Derived data is data that can be computed/compiled/extracted from one or more raw data elements, i.e. the data is derived from underlying data items. The derived data which will be extracted combine meta data with actual measured data and generate a data stream about the system condition and the evolving risk while operating the system.

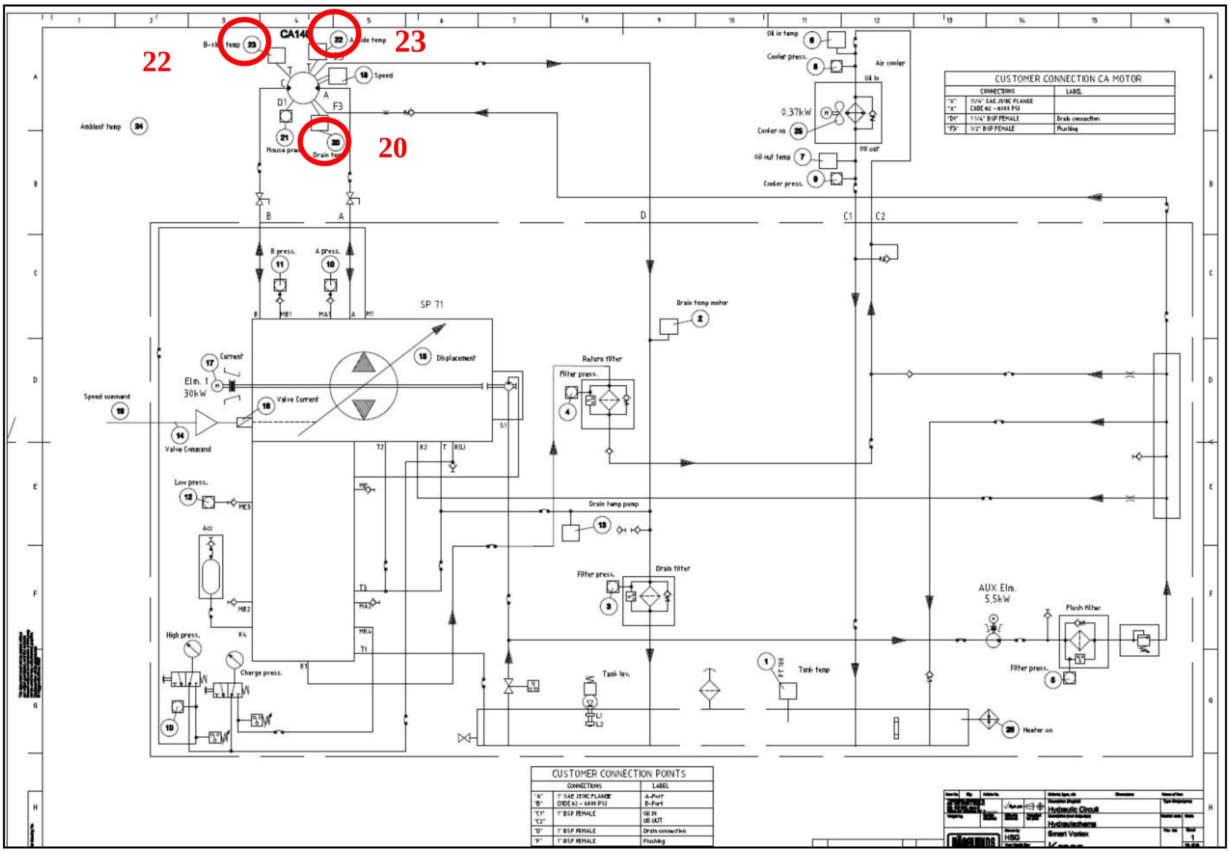
4.2.1.5 Validation

Validation means to compare independent measured values of parameter data with calculated values from other raw data and metadata. The calculation should render an “ideal” value. If the values are the

same (or within a tolerance), the parameter values are validated. Else, they are falsified and one has thus identified losses or other errors in the system.
Alert limits SensorNR22 and Sensor23 < temp limit
As mentioned in the text above, IF L > X1 or D >X2, then the slope of DL/dt and DD/dt are evaluated during a time interval Δt . Here set-point of slope are determined through engineering judgment.

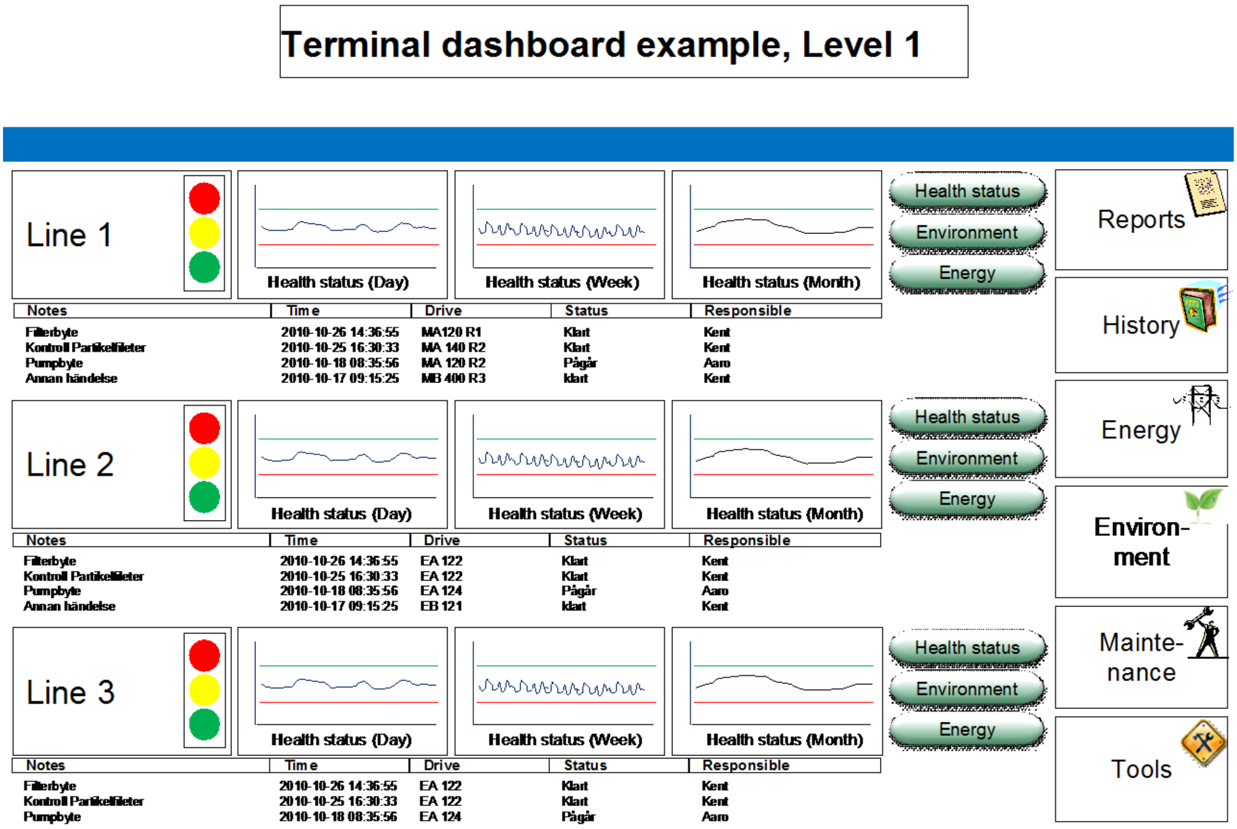
4.2.1.6 Actions

Collected data can be used when analyzed (in real-time) to trigger actions (either to notify operators or automatic actions) in the manufacturing or operations environment. Already identified actions that can be possible to invoke the system operators in a collaborative process and propose specific system or maintenance actions. In the current use case the system need to get stopped to avoid damage to the Electro-Hydraulic Motor systems.



Picture 12: Logic diagram with marked sensors used in the use case

The outcome of the continuous analyses will be available in a “cockpit chart” at an operator dashboard similar to the figure example in picture 13 below:

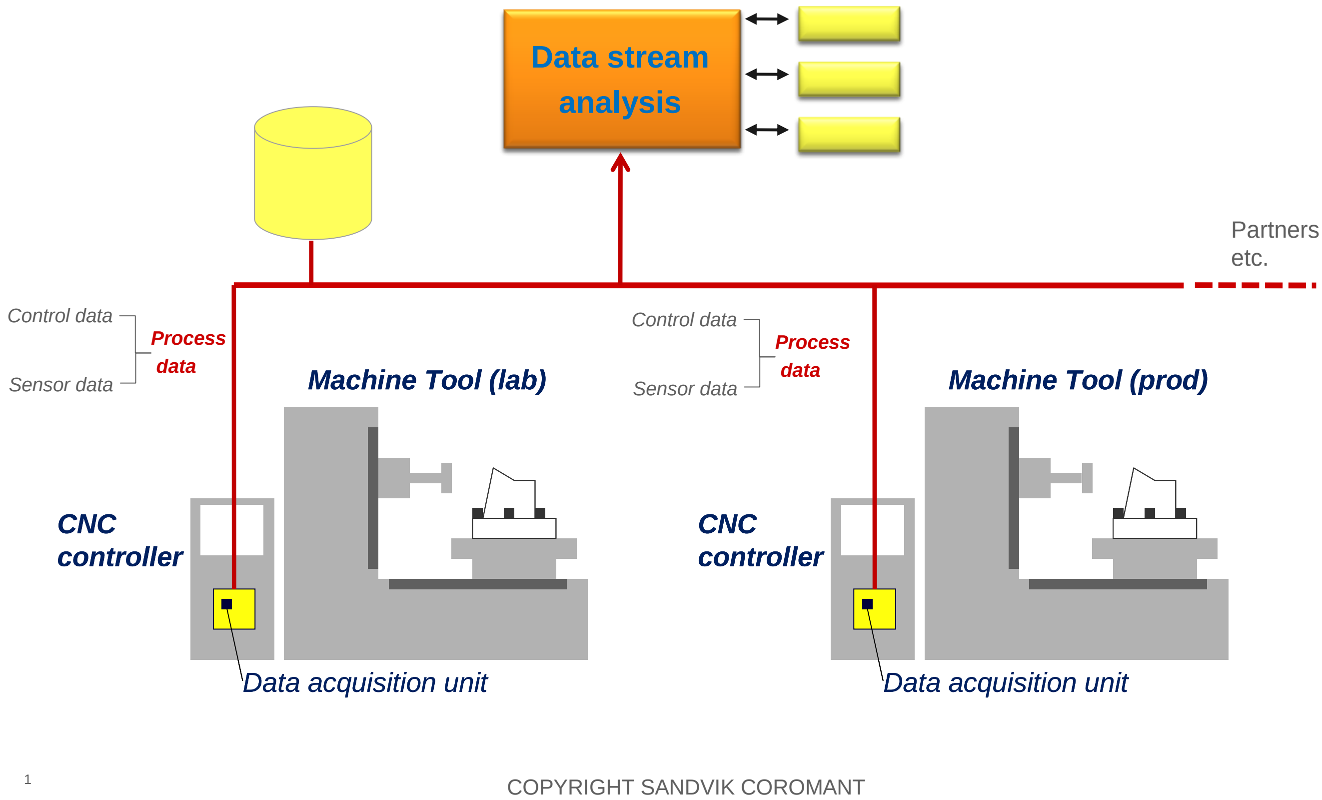


Picture 13: **Terminal dash board example**
(Copyright – Hägglunds Drives)

4.2.2 Use-Case ISP-2-2: Real-time data streaming and computation from actual sensor at milling tools (Use case provided by Sandvik Coromant)

There are many factors and dependencies in machining that must be considered when developing efficient sustainable machining solutions. In order to develop, validate and provide functional products within this area, all parts in the machining system (including cutting tools, machine tools, part to be machined, materials, controller, machining strategy, cutting data etc.) need to be considered in a comprehensive way. Consequently, data parameters from all these different but interlinked parts need to be captured, combined, and analysed. The complexity is high due to very large amounts of data (static and dynamic), and that these data originate from many different sources such as:

- PLM systems (product data of components to be machined)
- Manufacturing resource management systems (data about cutting tools, machine tools etc.)
- CAM, process planning systems (machining strategies, cutting data, clamping methods etc.)
- Material databases (material data)
- Sensors (in-process monitoring data of the machining process)
- Controllers (in-process data such as axis positions, current power values etc. parameters for productivity analyses etc.)
- Metrology and quality evaluation (critical dimensions, tool wear parameters etc.).



Picture 14: Sandvik Coromant pilot set-up (Copyright – Sandvik Coromant)

A challenge is to be able to capture and interlink all wanted data from a large number of machine tools and machining operations, and to be able to use these large data sets for comprehensive analyses (also including identification of meta-knowledge).
Efficient methods and tools are necessary in order to be able to implement, verify and validate functional products and related models, processes and software tools. In addition, needed is to support identification of the important customer values, i.e. both spoken and unspoken customer needs.
Monitoring and efficient usage of large data streams from many different sources, both internal and external ones, enables the support for sustainable machining solutions.
The data parameters to monitor and collect can be grouped into the three main categories, i.e. process, resources (equipment/machines) and product data. A detailed categorization and classification of data and data sources, as well as data quality issues are needed to be dealt with during the project. The data collected during the project originates from e.g. the machine shown in picture 15.

In essence, the purpose is to capture relevant information that can be related to the machining of a component and interlink these with parameters related to the resources and the product data. The process-related parameters include data before, during and after the machining process. By capturing process data and combine these with data from resources and product data a comprehensive view of the machining situation will be available and represent knowledge. This will improve collaboration in the PLM and facilitate development of new solutions, process improvements, and trouble- shooting as well as to efficiently validate and verify models, software tools and procedures used in the functional product itself. Further, both a proactive and reactive view can be applied.

Picture 15: Test Milling Machine used for the SmartVortex ISP-2 scenario

4.2.2.1 Scenario

4.2.2.1.1 Objectives:

Demonstrate operational data acquisition from a machining process, subsequent data processing and evaluation at Sandvik Coromant. A concept for large data stream management, quality assurance of data and models will be developed.

Develop collaboration environments supported by decision making systems

Case study #2 will in the first steps be carried out within Sandvik Coromant and Sandvik Tooling Supply. When the technology and methods are feasible to be scaled up, Hägglunds Drives will be jointly involved as an external collaboration partner.

4.2.2.1.2 Problems / Challenges:

The following challenges appear and need to be addressed:

- Collaboration between involved business partners. The following steps need to be addressed:
  - stakeholder identification
  - identify data and information needs related to stakeholders
  - identification of data to be extracted from large data streams
  - selection and prioritization
- Processing massive data streams, the following steps need to be addressed:
  - quality assurance of data and models
  - identification of meta-knowledge
  - specification and implementation of software modules for data acquisition and processing
  - implementation of a method for data management when scaling up from one machine tool to a large number of machine tools

4.2.2.1.3 Requirements:

How to find unmet customer needs, optimize utilization of products/services

The re-link of the product data stream need to create feedback from the market/customers with company ideation and development processes to:

- identify consumer needs
- monitor the customer's key performance indexes regarding the level of capability to utilize the introduced services and product solutions
- suggest possible improvements for the customer's process and facilitate improvement of the customer's process
- capture new needs for product upgrading or totally new innovations
- quality assure simulation models and optimization methods used in product and service development (needs to be quantified)

4.2.2.2 Raw data

The raw data is planned to be collected from various sensors and data sources at adequate frequencies (i.e. with such granularity that the data collected is useful and reveals variations or abnormalities). The raw data are generated from either sensors at the tool or at the machining elements or are transferred from the NC-control system.

4.2.2.3 Metadata

The used meta data belong to the tooling process and define the process itself, i.e. the used tool, the used machine and contain information about the work piece.

4.2.2.4 Derived data

Computation activities which combine machine, tool, process and work piece meta data with the actual measured data are generating data stream objects that cannot be measured directly but are required to identify and prepare specific tool and process parameters for the final validation process

4.2.2.5 Validation

For the validation, data objects resulting from the derived data streams are taken for risk and quality calculations and compared with the previously generated and used simulation models.

4.2.2.6 Action

Collected data could be used to trigger actions either to notify operators or automatic actions in the machining environment. Actions can mainly be divided into two categories:
1. Real–time process deviation notification or adaption,
2. Off-process actions e.g. model refinement, lessons learnt, recommendations etc. (Further extensions will be defined during the project)

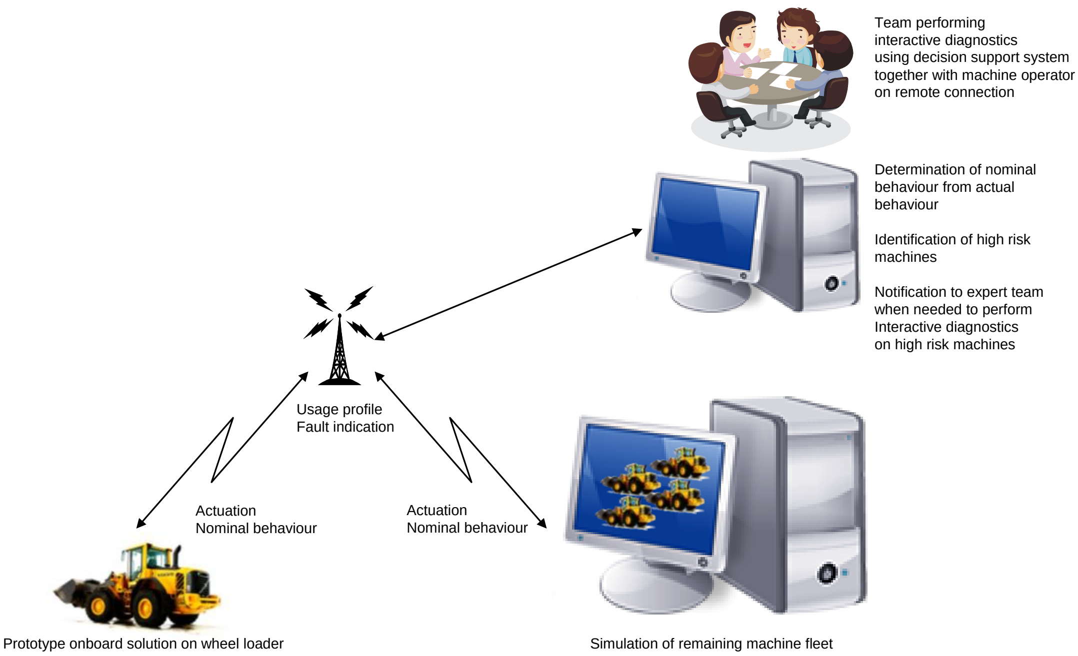
To collect the raw data, metadata, derive data, and produce validation requests continuous queries executed by the DSMS are needed. Already identified queries are related to quality parameters and machine or tool damage situations. (Further queries need to be developed):

4.2.3 Use-Case ISP-2-3: Real-time data streaming and computation from actual sensor in construction equipment (Use case provided by Volvo Construction Equipment)

Volvo CE has a vision that customer support sells services to customers where the machines are monitored in more detail. In this project, the information needed by such a support is realized for one or more machines. Upon alarms, immediate and corrective actions are then taken to avoid failures or updates maintenance schedules if there is a perceived increase regarding for instance the level of failure risk. The streaming data from the machines can then be compared to find machines behaving differently to the rest of the complete Volvo CE customer base machine fleet. This is an approach that reduces the need for diagnoses based on detailed knowledge of the machine behavior, which in practice often cannot be realized due to resource needs and the complexity in modeling material imperfections etc. since not only electronics fail but also mechanics. The alternatives to this so-called model based diagnosis are rule-based and case-based reasoning. Rule-based reasoning uses a set of fixed rules set up to find every possible error based on expert knowledge. Case based reasoning mimics the human mind where similar historical cases are compared with to find the one with the closest match. Apart from simple comparison of trends and absolute values with compensation of operator and application characteristics, case-based reasoning could be employed in this project to compare events among the various machines using data from previous failures among the machines. If the data is real-time, corrective actions can be undertaken using short term trends early and fast enough to avoid also some the random failures that cannot be predicted from long term trends.

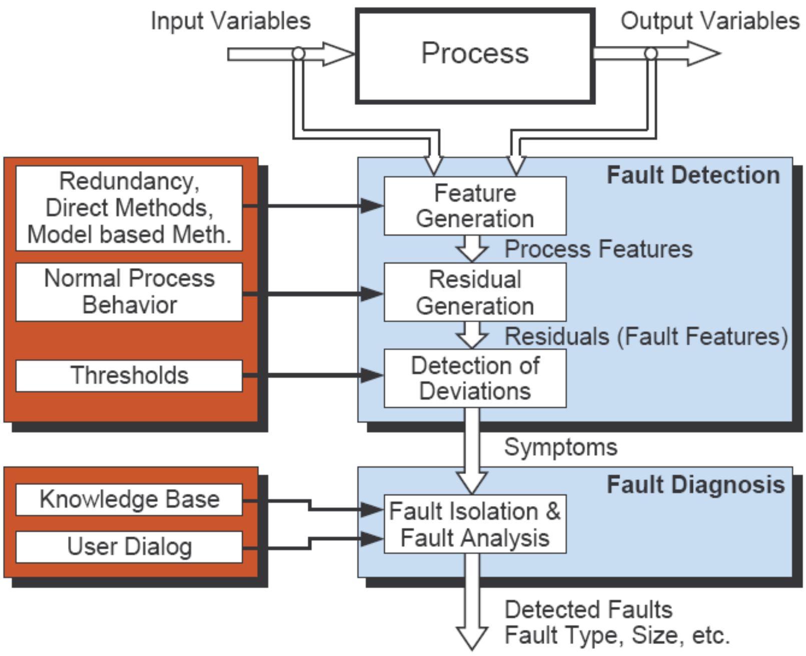
Customer support would in this vision have a central monitoring function with a number of operators surveying the connected machines, whose volume could exceed 10 000. Similar requirements can be imagined in nuclear power or large chemical plant surveillance. The number of operators should of course be kept low which means that a large amount of information needs to be condensed into decision information to be presented to the operators. If an operator finds a machine of particular interest, the operator could be given the option to get more detailed information. The latter mechanism and the remains of this paragraph are implemented in this project. The operator could then demand a different type of data stream from that machine with more internal signals and higher time resolution to evaluate the severity of the problem but also the cause of male function initiated by the operator, the environment or the machine itself. Another option is to automatically provide more details to the operator if severe problems appear that require calls for urgent action or in other cases put it in a task list for investigation.

In the Smart Vortex project, Volvo CE will realize data driven fault detection and remote fault diagnosis for a wheel loader demonstrator. Machine usage characterisation and fault detection is primarily performed using on-board solutions including the Smart Vortex DSMS. Information is sent to a test set-up of a Volvo CE centre upon request to tap any information needed for comparison with other machines in the fleet to find deviations in general (data driven approach) and for remote fault diagnosis purposes. The goal in Smart Vortex is to prove (through analysis) an improved uptime of the machine by 10%. Another contribution from Smart Vortex will result from less travels needed when using remote diagnostics (compared to the on-site assessments required today).



Picture 16: Volvo CE pilot set-up (Copyright – Volvo CE)

The use case includes research on a framework/implementation of a wheel loader demonstrator, and simulation. Further, the use case addresses running machines in which faults regarding the transmission can be detected (through comparison with variables from other machines being used in similar settings) and where interactive remote diagnostics can be performed effectively (resulting in a reduced need for travels and thus improved uptime for the machines).



Picture 17: Process for monitoring, fault detection and analysis at Volvo CE (Copyright – Volvo CE)

The framework will be further described during the research, but in general the vision is that the availability of live data collected from a large population of machines gives the possibility to introduce data driven fault detection and condition prognosis. Only machines being used in similar ways in terms of operator characteristics, external loads etc are compared when searching for anomalies in the signals of interest. One signal could be the transmission clutch pressure build-up for a particular gear switch. The distinction between where to apply a data driven approaches such as general clustering and case based reasoning in comparison to model based, needs to be clarified during the project.

Below is a picture (Picture 18) showing the machine (or construction equipment) that will be used as data source:

Picture 18: Volvo CE L70 wheel loader equipped with sensors (Copyright – Volvo CE)

Machine usage characterization

The fault detection, fault diagnosis, and condition prognosis are based on trends, deviations among machines etc., but also on information on how the machine is used including operating cycle time, and average velocity as well as other factors. This information is immediately useful for other purposes such as product planning and product development.

In Smart Vortex, this type of information is stored as sequences in time of different types of operating cycles that in turn are made of phases. The operating cycles can be:

- Short cycle loading
- Load and carry
- Transport

Whereas the phases can be:

- Approaching pile
- Bucket fill
- Reverse from pile
- Towards load receiver
- Bucket emptying
- Reverse from load receiver

The algorithms needed for finding the start- and end time points of each phase are developed in a parallel project outside of SmartVortex. Within the parallel project, research is performed for deriving the operating cycle types and applications described later in this document.

For each phase, at least the information below is stored:

- Duration in time
- Load duty [rev*Nm^3]. This is the accumulated damage to a component. This can for one revolution of a shaft be the average torque raised to some exponential between 3 and 7 depending on damage type of interest. For a complete time trajectory, the damage is the sum of one revolution's damages.

- Average value, value at start and end point for:
  - Load in bucket [kg]
  - Engine torque, speed and momentary fuel consumption
  - Machine position and velocity from GPS
  - Road slope
- Stored at switching points together with time stamp:
  - Lever positions: lift +/-/0, tilt +/-/0 and/or steering +/-/0
  - Gear N,R1,R2,R3,R4, F1, F2, F3, F4
- Sampled at 10 Hz. The sampled data can be retrieved through the telematics system. Otherwise, it is analysed to produce information on application etc. and not stored:
  - Angles of lift, tilt and steer
  - Machine position and velocity
  - Road slope
  - Transmission output torque and speed
  - Hydraulic cylinder forces and speeds. Bucket lift, tilt and steering cylinders.
  - Transmission clutch pressures

For each operating cycle, the following information is derived and stored on top of the information above stored for each phase:

- Cycle type: short loading cycle, load and carry, transport
- Application: Bucket, pallet or timber
- Altitude (via engine inlet pressure)
- Weather (given by ambient temperature and windshield wiper)
- Material handling (light or heavy)
- Loaded weight [kg]
- Cycle time
- Distribution in time of:
  - Cylinder forces and speeds
  - Transmission input and output torque and speed
  - Road slope
  - Fuel consumption
  - Engine torque and speed
  - Gears
  - Brake and throttle pedal position

The operation of deriving the data streams above for each operating cycle is dependent on meta-data on:

- Sensor positions
- Sampling frequencies
- Sensor data types (temperature, vibration etc.)
- Sensor ID (for comparison with manufacturer data sheets that are also stored)
- Software versions
- Operator ID (gives information on training and work hours)
- Average behaviour of other alike machines. This is updated regularly from the Volvo CE centre. Per application, material handling, cycle type and phase:
  - Load duty and its increase slope
  - Fuel consumption
  - Distribution in time of … (see operating cycle above)
  - Time trajectories of transmission clutch pressure build up per gear switch
- Machine ID (gives machine model and configuration)
- Customer:
  - Agreed operating cycles and maximum loadings and velocities

- Service history
- Cost of down time
- Longest possible stop time
- Position
- Weather
- Environment

The information above is sent to the Volvo CE centre upon request from the centre or when on-board memory is full. The information is compressed at the centre if needed, ensuring that only time distributions per operating cycle type and phases for each operating cycle type remain with the chosen sampling rate (e.g. once a day), and not time trajectories of measurement data. Using this stored data, it would be possible to find changes from one day to another in for example the time distribution of transmission output from torque during bucket fill in short cycle loading.

4.2.3.1 Scenario

4.2.3.1.1 Objectives:

Provide possibility for adequate product feedback to stakeholders such as research and development organizations, service organization and customer. (Analytic tools will however be developed outside of the project).

Develop collaboration environments supported by decision making systems.

4.2.3.1.2 Problems / Challenges:

The main challenges are to acquire additional product knowledge and enable increased availability of the construction equipment.

4.2.3.1.3 Requirements:

In the project the Use CareTrack and Matris systems is used for demonstrative purposes holding data of every heavy production machine. The data monitored are: performance data, position, service information, and real-time use data from on-board sensors via telematics systems. (Further development of Volvo CE's back office systems, to be able to handle all information contributed from the project, will be managed outside of the project. The back office systems will help interpret the results and offer support decision making based on the gathered information).

4.2.3.1.4 Machine classification

It is valuable to be able to classify machines into groups with same machine models and similar usage as defined in earlier. Any deviations in the operation of the machine components among machines in the same group will be less due to external factors, and more due to manufacturing tolerances and component condition. This narrows down the scope when performing fault detection or fault diagnosis. Below is an example of a question that could be given at a Volvo CE centre level with access to the complete set of machines produced by Volvo CE with the Smart Vortex tools installed

- Show 3 machine groups (given certain machine model and machine configuration and operator experience in hours) based on clustering techniques where the classification is performed with the following as input (moving average for last week)
  - Time distribution of
    - Applications
    - Operating cycle types

- Per operating cycle type and application combination
    - Average handling
    - Average bucket load
    - Average machine velocity
    - Average ambient temperature

4.2.3.1.5 Detecting deviations and faults within context groups

For every machine group from above, from now on called context groups, we now can state a number of questions to find deviations in general that need further attention. The common context will make it easier in the end to understand the causes of the faults or deviations that have occurred. A few examples of relevant questions are given below:

- For each context group, show the 5 machines which produced most critical error codes last week. Any ghost error codes are not included (induced by other errors).
- For each context group, show the 5 machines which had the steepest slope for load duty and/or fuel consumption last week.

4.2.3.1.6 Detecting deviations and faults when comparing with nominal behaviour

To detect deviations in machine components, where the external environment and applications contribute less than in the previous example, Volvo CE would like to compare the operation variables with an average or nominal behaviour gained from the complete machine fleet of the same model.

The nominal behaviour is analysed and following sent out to the machines in the form of thresholds etc. depending on usage profile. This help to calibrate the fault detection routines on each machine.

In the Smart Vortex project, an example where the pressure build up in transmission clutches will be studied. If the pressure build up is not correct, the results will be slippage which must be avoided. Any deviations from the nominal behaviour can help to indicate that slippage might occur in the future. The following data stream is needed to study the gear shift from forward first gear to forward second gear:

- Show the average of the 10 latest gear shifts F1-F2 for the time data of control valve downstream pressure starting at gear switch and ending 1 s after clutch speed difference is zero.

Given this type of data stream for all possible gear switch combinations:

- Show the group of 10% machines of a given model that deviates the “most” from the others in terms of all the pressure build up data streams

A similar case is the deviation of gear shift comfort among machines. This demands on-board Fast Fourier Transformation (FFT) filtering:

- Show the average of the 10 latest gear shifts F1-F2 for the frequency range of 1-3 Hz for FFT filtered time data of longitudinal accelerometer starting at gear switch and ending 1 s after clutch speed difference is zero.
- Show the group of 10% machines of a given model that deviates the “most” from the others in terms of all the FFT filtered longitudinal acceleration data streams

4.2.3.1.7 Detecting deviations and faults when comparing with nominal behaviour

The deviating groups of machines could even be hierarchic, if the deviating groups in the last section were retrieved from each context group. In this way, the deviating machines can be studied in more detail using interactive fault diagnosis:

- Given a deviating machine with respect to transmission clutch pressure build up data stream, initiate a remote interactive fault diagnosis session. Operator secures safety.

- Show in real-time in a flow diagram the pressure of all transmission clutches with sampling frequency 50 Hz. At the same time display the reference pressure from transmission Electronic Control Unit (ECU)
- The mechanic demands actuation of valves according to given scheme and/or manually. Show deviation of valve position to reference signal.
- Play sound of mobile microphone with upper frequency limit from database. Machine operator places microphone at suitable position.

4.2.3.2 Raw data

The raw data is planned to be collected from various sensors and data sources at adequate frequencies (i.e. with such granularity that the data collected is useful and reveals variations or abnormalities). The data describing the wheel loader system including the operational user behavior, the hydraulic system and the environmental and working conditions the are taken from either direct mounted sensors, the CAN-Bus, the central control unit or the telemetric system of the wheel loaders. Video and audio capabilities are enriching the data streams in specific cases.

4.2.3.3 Metadata

The metadata describe beside the wheel loader and the used hydraulic all sensor systems and the operation modes which are contractual agreed and belong to the warranted characteristics. Maintenance history data and business model specific data extend the meta data to allow risk and threshold calculations.

4.2.3.4 Derived data

In this use case especially trends, summaries of events or errors or time related processing of handling data contribute to the derived data streams.

4.2.3.5 Validation

The validation is identical to the validation mentioned in the other ISP-2 use cases.
To collect the raw data, metadata, derive data, and produce validation requests, queries to pose to the DSMS are needed. Queries will be developed during the project.

4.2.4 Use-case Background ISP-2-3: Customer usage modelling. (Background provided By Volvo CE)

This use case is not a real part of project, but must be made visible for setting requirements of future usage of the tools and methods of the project. To mark this, this section has been written in "Italic" letters. The findings of this section are not seen as deliverable, but give background knowledge for future market oriented activities of the consortium partners.

In product planning and product development, Volvo CE needs to have a clear view of how the machines are used in different regions and applications and the distribution among applications. This information is used for planning product features and for sizing components and designing systems to achieve high enough fatigue life and good fuel economy. On a higher level, Volvo CE needs to be able to estimate the total cost of ownership for a large set of our customers for a new model range. How much could operator support affect the fuel consumption and how could a new gear shift strategy affect transmission fatigue life, both given how the new machine would be used by our customers. To answer these questions, Volvo CE really needs to measure the machine independent variables, such as road slope, curvature etc. Volvo CE work in parallel projects on how to analyze the signals measured *on machines to derive the machine independent variables and how to measure in a cost efficient manner as the measurements will take place on a large part of the machines produced.*

The data quickly grows large and thus Volvo CE works in parallel projects on also how to compress the information using pattern recognition to divide the data into typical machine operation cycles, where each cycle is characterized with a few attributes and saved rather than saving the sampled time data. Some processing thus needs to be performed locally. Storage and higher level analysis however is performed on a Volvo CE central server.

There is no real-time demand in this use case, rather there is a need to empty the measurement buffer regularly on the machines due to limited local storage. In total this will generate a data stream to the server.

By and large Volvo CE wants to measure and validate

- *operator characteristics (to validate operator models) and operators' influence on performance. Preferred velocity for given outer conditions.*
- *environment (temperature, humidity etc.)*
- *ground surface profile / curvature / stiffness / slope …*
- *position using GPS*
- *properties of handled material*
- *type of attachment currently used*
- *time spent at each torque and speed combination for engine and in analogue for other components. Torque and speed in this case are divided into a finite number of ranges, such as 100-200 Nm and 100-200 rpm. The latter combination will have figure related stating the number of seconds spent in that operating range.*
- *accumulated load duty [Nm*rev] for various components*
- *Emissions for governments*

Volvo CE would also need, in the central server, relate the machine's position to GIS information that gives us information of ground material properties, country (Which can be used to draw conclusions on operator characteristics), surrounding terrain profile etc.

The data is stored and directly updates a number of Volvo CE customer and application profiles primarily used by product planning. Product planning prepares business opportunity descriptions based on this data in which it is shown how new products will improve customer satisfaction in the end for a chosen part of the customers and their applications.

The data also updates a number of machine usage profiles that product development use for sizing components and for developing control strategies that will improve operability and fuel efficiency for the largest part of the applications.

Typical questions:

- *What is the top 5% torque range the transmission output shaft is subjected to during periods of time > 5 s for the application re-handling and operating cycle "short cycle" and phase "bucket loading"*
- *For which regions 3 do the operators have the largest influence on fuel economy in comparison to environment and applications?*

SmartVortex will lead to flexibility in how questions can be asked to the usage data and the possibility to combine factors in a more complex fashion. This will lead to less need for (expensive) questionnaires used by product planning and marketing and sales by 20%. This will be shown analytically but based on test cases where product planning employees try the SmartVortex system for *a large number of artificial data to cover information on which country the machine is in and similar operation or environment related use paradigms.*

At the same time, fewer measurements need to be performed on existing machines, so called CUP's (customer usage profiling). A reduction of 20% of tests here will be shown analytically compared to the machine monitoring of today. The analysis is based on evaluation of the Smart Vortex system by product platform engineers using actual measurement data where machine independent variables are measured as explained above while making them available in an intuitive fashion using visualization means for the queries.

4.3 ISP-3 Product data stream sharing and collaboration in complex cross-organizational scenarios

In this case study, we integrate the data streams and collaboration across the boundaries of the individual industrial partners. The semantic social collaboration framework is deployed for each partner.

The vocabularies are adapted to the individual requirements and structures present within these environments.

The rules and policy framework is used to specify constraints for a cross-organizational mechatronic engineering scenario, where partners only share parts of their product data streams within collaboration and decision making tools.

In order to be able to test and verify the tools and methods for cross-organizational sharing of industrial data that is developed in the project, a virtual collaboration scenario is used, wherein a fictive but realistic business relationship exists between the companies taking part in the case study.

The scenario is designed to validate and demonstrate all aspects of the rules and policy framework.

It develops and highlights the use of the collaboration tools for data sharing in cross-organizational settings. The data streams to be shared will be real data streams in the sense that the data is representative for the collaborative work scenario, but the data is not sensitive.

Thus, the sensitivity of the data is simulated.

Similarly, the business relationship between the industrial partner companies participating in the case study is fictive but realistic. In this case we envision a mobile garbage crusher scenario, where a garbage crusher is composed of different components of the individual industry partners into a single functional product.

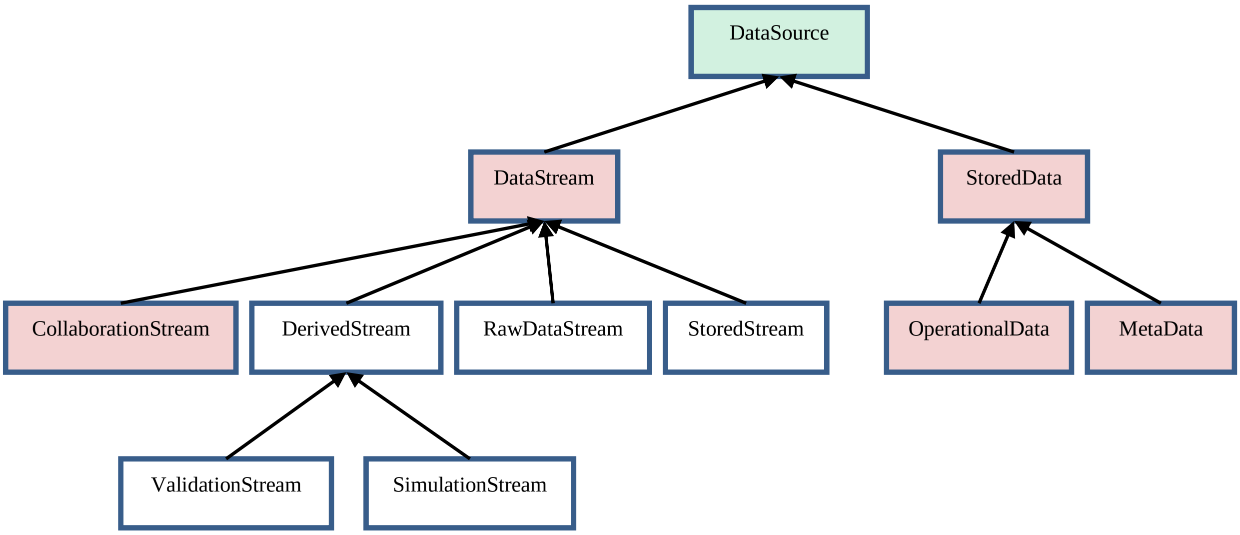
The scenario demonstrates how the SMART VORTEX framework can support policy based data sharing in a cross-organizational business collaboration situation where a subset of the data being shared is subject to access constraints.

Mechanisms for de-sensitizing data streams are also included e.g. downgrading data in a way that removes sensitive contextual metadata from the data in a controlled way.

This coordination of collaboration in cross-organizational settings is acting a level above the systems as described in ISP-1 and 2. This work requires the implementation of a new customizable social graph, rules, policies and collaboration environment.

Some collaboration tools will be built from scratch, while others extend upon existing domain specific tools requiring extension and integration.

Finding experts within a company for any kind of problem is a complex and time consuming task. Few isolated applications for expert finding exist, however the demand for effective solutions keep on being a non-trivial task [cf. M06, PC08, S09, ZT07]. Numerous factors are involved in the indication process that designates the right person as expert for being a part of a problem solution. The bases for such a designation process are all accessible and relevant information about the expertise of a possible candidate. Such information could be the persons profile, data from collaborative tasks, role in a specific environment or the involvement in production processes (e.g. design, development, production or testing), but are not limited to those. On top of an ontology based representation of such data, Entity/Relationship search approaches could be applied in order to support expert finding and hence maintenance tasks or decision making process.



Picture 19: Data stream schema with highlighted elements participating at use-case ISP-3

4.3.1 Use-case ISP-3: Cross-Partner Collaboration

This is a fictive use case where Sandvik, Hägglunds and Volvo CE would collaborate on condition monitoring of our machine elements using acceleration information, temperature and so forth to estimate remaining life of the machine element in question. For Hägglunds it could be sliding bearings, for Volvo CE gears and for Sandvik cutting tools. The partners are interested to support this use case.

We would have a common knowledge base that can be shared that is basic research on condition monitoring and wear and fatigue while the application knowledge would be hidden from the partners. Any claims on the common results should be handled by the information system such that the dividing of results is unambiguous. It shall be possible to search the system for background, foreground and accompanying common findings, to be able to judge whether we are exposing too much or if another partner's foreground could be of interest.

Typical questions:

- Is this result coming from partner background, foreground or common established knowledge?
- Should we keep this result hidden as it might be of interest to patent by the other partners?
- What are the use rights for this result for all partners?

With the SmartVortex solutions for a number of project leaders (not legal experts) a measure should be established what percentage of misunderstandings exists related to the written agreements. This should be rectified by 30%. The time needed to answer question should be shortened by 50%.

The base components of the SmartVortex suite to support operationalization of the solving of these problems are the social middleware originating from WP8 and the rules and policy framework from WP 11 and WP12, where the first WP captures the distributed network of entities and their relationships, the second WP implements rules and policies on top of this knowledge. Together these enable the data stream management and collaboration tools to reason about these questions.

4.3.1.1 Scenario

The scenario combines results of ISP-1 and -2 into a cross-organizational environment. In the ISP-1 Delft workshop two processes were determined to improve decision making in product lifecycle:

engineer to engineer collaboration and interaction; problem solving for maintenance. Both processes are for instance possible part of a Product Service System (PSS). PSS is a focus shift from traditional production offers towards a combination of production and service offers. The intrinsic characteristic of PSS it is that manufacturer and costumer are cooperating tight coupled (cf. [BWB10]). In that PSS is from the manufacturer perspective a gainful advancement, it means improvement of knowledge about costumer product usage as well as the further costumer requirements respectively product development. Hence, such knowledge is a useful flow back of information for product optimization. On the other hand manufacturers are faced with the challenge to offer product relevant services of diverse escalation stages, as for instance service hotlines, remote or on-site assistance of experts. Such a service offer implementation is even more complex to handle in case that parts of a product are produced outsourced by contract partners. In this case expert knowledge about products is not completely in-house available.

Different service escalation stage produces different costs and efforts and hence down time of manufacturers products. Hence, it is important to establish cross organizational structures along with the provision of relevant information to offer engineer to engineer collaboration and interaction and problem solving for maintenance.

The ISP-3 scenario will focus on these processes from a cross-organizational perspective. The major challenges in both processes are related to collaborative design decision support problems and require the participation of various designers and engineers from different organizations. Identifying the best possible participants for the collaborative process will be one of the major tasks for the Semantic Social Collaboration Suite (SSCS). The following sections focus on the process support itself.

4.3.1.2 Objectives:

Set up of a cross-organizational IPR protected and self-configurable collaboration environment. This will guide to two main tasks which address the following demands:

4.3.1.2.1 Intelligent self-configurable collaboration environment to support maintenance

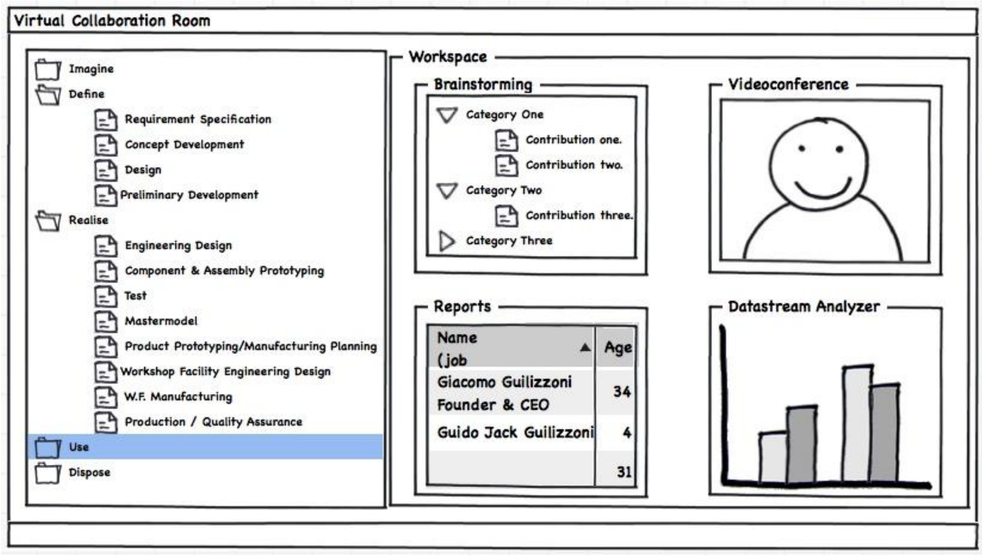
The construction industry is currently putting a lot of efforts to develop new mechanisms to collect and monitor data generated from construction equipment to obtain a better feedback of the machinery use [I03].. The feedback data is valuable to these companies because they can better understand current limitations of their equipment and allow engineers to work together to constantly improve equipment's availability and reliability.

However, besides the efforts of collecting large amounts of data generated by equipment, one limitation present in these companies nowadays is the lack of a suitable environment to allow engineers to collaborate with each other processing such amount of data and making decisions about equipment improvement or even equipment maintenance.

With the natural evolution of diverse types of technology, companies are migrating their maintenance equipment policies from the "Fail And Fix" (FAF) methodology to the "Predict And Prevent" (PAP) maintenance methodology [IL09]. Maintenance addresses the fundamental needs of predictive and intelligent tools to monitor the degradation of an equipment usage in order to allow interventions to be taken before a unscheduled downtime or unexpected breakdown.

For example, if an excavator starts to present possible failure signs, identified by the telemetric data, e.g. that its engine is operating over a certain limit, experts then need to be brought together into a collaborative environment to analyze the problem and take decisions to avoid an abrupt breakdown. Besides solving an emergent situation, such collaborative environment should also support experts to improve future generations of the equipment. The near breaking down state can be interpreted as feedback data that allows experts identifying the design or usage causes that led to such extreme situations and improve future generations of the equipment.

In the ISP-3 a new collaboration environment will be developed that brings together the right experts from different organizations, to make decisions based on the relevant related information from the data streams in order to solve certain maintenance and improvement problems. The following Picture 20 introduces the Virtual Collaboration Room, a mockup representing a possible collaboration environment that can be used (Picture 20). The virtual collaboration room will offer access to necessary collaboration tools, information and bring together people to take decisions about equipment maintenance.



Picture 20: Virtual Collaboration Room, an intelligent environment to support experts in maintenance processes.

The scenario uses the results of ISP-1 and 2 to process data streams of a certain equipment started to give unconventional signs that it is about to breakdown or stop working, after some unsuccessful automatically software agent interventions.

The collaboration environment realizes this situation and starts its intelligent configuration process to bring together, based on an automatically pre-diagnosis of the problem, professionals from different organizations that might understand current causes of the problems, or professional who already experienced similar situations. The environment also configures tools and provides information from the data streams that professionals should use to diagnose the problem and to communicate with each other. Some example of tools and its configuration can be seen in the Picture 20, as videoconferencing tools where professional involved in the diagnosis process are automatically brought together, graphs presenting current values of equipment variables, reports containing previous similar stories and their solutions, and brainstorming tools to capture creative ideas on how to diagnose and solve the current problem. Compared to current collaboration environments, we will investigate a self-configuration of the environment depending on the situation being handled and in order to support an effective collaboration process.

4.3.1.2.2 Decision support model for decision based design and collaborative decision support framework

The conceptual product design is characterized as a collaborative task between different organizations with several decision-points that involve the participation of various designer and engineers for product definition and optimization. During this process, a design team collects and monitors data streams generated from parameter studies and simulation as a foundation for its decision making process. Streams are handled in two forms of data: verbal-conceptual and visual-graphic.

In the Smart Vortex project we will develop new collaboration environment that bring together the right experts from different organizations to make decisions in a collaborative environment. During this process, the system should provide relevant information from the data streams which is needed to make a decision as well as relevant information about possible experts profile, which should be part of the problem solution. Data stream of the product lifecycle will be analyzed for possible decision points in the definition and realization stage. The environment detects these points and starts its intelligent configuration process to bring together engineers and designers from different domains, such as mechanical and electronic design, which provide knowledge or skills that will be needed for decision making. The environment also configures collaborative tools that the design team can use to analyze and interpret data as well as to support the communication of a distributed team. According to the massive data stream of the product lifecycle, the system needs to filter relevant data for the design decision-making process.

The system should further capture previous decisions and their influence on the product lifecycle. This data can be used for future decision making as best practice or solutions that did not solve a problem but provide information for its fail.

4.3.1.2.3 Problems / Challenges:

One goal of SmartVortex is to provide a technological infrastructure to achieve better cross-organizational collaboration and decision making. As a result the following challenges appear and need to be addressed:

- to support group formation and collaboration in a cross-organizational environment;
- to manage data exchange in a cross-organizational environment;
- to support all needed collaboration activities that are necessary for collaboration and decision making;
- to adapt collaboration tools and collaborative process to new situations in the product lifecycle management, which results out of a constantly changing data stream.

4.3.1.2.4 Requirements:

To achieve better collaboration and decision making in SmartVortex, we need

- to identify and analyze collaboration tools the industry partners generally use in the product lifecycle;
- to identify and analyze collaboration activities the industry partners generally use in the product lifecycle;
- to identify and analyze data and information the industry partners need for collaboration and decision making;
- to identify and analyze rule and policies the industry partners generally use for the exchange of data in a cross-organizational environment;
- to identify and analyze rules and policies the industry partners generally use for group formation.

As a result out of this analysis we need to define a formal description for a collaboration that can be used to design and adapt a collaboration process;

- to define a formal description for relevant information about the employees that can be used for group formation;
- to define rules for the use of data out of the data streams for a collaboration process;
- to define rules for the adaptation of a collaboration process;
- to define rules and policies for group formation;
- to define rules and policies for the for the exchange of data in a cross-organizational environment;

The collaboration data stream consisting of data generated in collaboration and decision making processes. There are different kinds of data that can be generated out of such interaction, such as: the design decisions of a product that a group takes, documents generated about design decisions and records from group meetings. As collaboration data streams can easily become complex due to a high

amount of data around a product, we adapt the data stream computation schema to process the collaboration data stream and to filter relevant information that is presented to stakeholders.
The framework is not constraint to process only data based on documents, models of decisions, generated by the collaboration among stakeholders. Another goal is to detect the behavior of stakeholders, whenever a collaborative session takes place. This information is particularly important for intelligent applications that are supposed to guide stakeholders during collaborative sessions for achieving effective results. This way, the applications may use the collaboration data stream to analyzing collaboration and project performance according to metrics, such as, commitment, quality, impact, shared understanding, progress and quality of intermittent results.

4.3.1.3 Raw data

The Raw Data entity represents the data that is generated by the collaboration tools and by the interaction of its user. Such kind of data can be generally interpreted as the logging data of the system. For example, two participants of a chatting session can exchange messages to schedule a meeting. In this case, we intend to store the information that is being exchanged by the participants.

4.3.1.4 Metadata

The information transmitted through the data streams just represents the amount of data generated by the collaboration tools. As our major goal is to support different kinds of adaptation within our tools, we need to gather as much data as possible besides the outcomes of the raw data. At this point we are interested in collecting information about entities that are related to the outcomes of the raw data layer, in this case: the tools that are being used, the users that employ those tools and the processes that are carried out from the users. We classify those three variables as being part of our Meta Data layer.

It has to keep in mind that within SmartVortex, the data is used in a cross-organizational manner. This implies that the data to be collected may be subject to different terminologies and interpretations. Mapping mechanisms will be used to translate these concepts seamlessly during a collaboration session to make sure that the metadata gives a consistent picture for all parties involved.

For this to work, a common infrastructure is used. This infrastructure uses rules and definitions to map foreign concepts to local representations using a common language which consists of terms defined in a so-called general social ontology. These mechanisms can be applied to both conventional data sets and data streams and they can be used not only for a better understanding between cooperating parties, but also to enforce privacy and data protection policies by restricting the data a foreign party may receive.

4.3.1.5 Derived data

As the collaboration data streams contain high amounts of information (coming from the raw data) it would very complex that a human analyst processes all of the raw data without any filtering mechanism that can transform the high amounts of data into processed information, reducing its volume. That is why the next layer in the framework represents the Derived Data layer. The goal of this layer is to automatically process incoming data from the raw data layer using the meta-data information, transforming data into information. For example, the system is able to calculate from the data streams the current collaboration situation like the number of contribution for the categories of a brainstorming session or information about the behavior of a group.

4.3.1.6 Validation

The Derived Data represents a proposal of transformed data from data streams with the support of meta-data, making some system variables updated. However, for the decision-making process, a continuous update might not be suitable because the users might not have enough time to check and react to the variation of these data. Therefore, we intend to create a Validate Data layer that constantly analyzes the derived data using predefined rules. For example, the number of contribution for a category can be compared to a predefined scale that indicates if the number of contributions is sufficient as input for a decision-making. A system can further use this information to analyze group dynamics and to decide when to intervene.

4.3.1.7 Action

The transformation steps performed through the Raw Data, Derived Data and Validated Data prepare the data coming from the collaboration data streams to be used by the stakeholders for decision-making or by the system for interventions. We defined this process within the last layer of our framework; the Call for Actions layer. For example, a system will monitor the collaborative brainstorming process and detect small or insufficient quality contributions. The system make an intervention by changing the focus of the participants to these contributions and provide prepared information like criteria or background information to support the generation of better contributions.

4.4 Verification of requirements for industrial appliance

Related to the value prepositions mentioned in chapter 4 the consortium has verified if the dedicated research areas can be applied in general to the industry or similar acting business models. To really create impact based on the findings and results of the SmartVortex project it is required to verify the applicability and portability of the achievements in other companies or industry segments. We have no doubt that the results that are envisioned by the current industry partners can also be achieved in following commercial implementations in the European industries.

4.4.1 ISP-1 Computation of data streams deriving from stored simulation result file sets

Simulation is consumes a high portion of the design process. The experts are very rare and require high investments in salaries, environment, education and experience building. Reducing their workload will allow to generate a higher throughput and increase therefore their productivity. This can be applied to all kind of simulation processes where high data volumes are generated. Typical examples are noise simulations to validate Wi-Fi networks in medical applications and equipment, thermal analysis, cooling flow analysis, all kind of electromagnetic analyses, and many more.

4.4.2 ISP-2 Computation of real time data streams deriving from sensors and actuators

Volvo CE expects to get about **10% increased uptime** regarding construction equipment (machines) when the results from the SmartVortex projects are implemented and integrated with the organization and information systems used to manage the construction equipment in the field. Further, Volvo CE foresee a **decrease in the travel time and travelling required for maintenance,** and in particular the first diagnosing trip, as more monitoring information is available to analyze and base planning upon.

Hägglunds Drives expects to get **increased system availability** in the electro hydraulic drive systems as the main result. Additional potential effects are also further knowledge and tools enabling energy optimization, improved proactive fault monitoring, improved system monitoring, extension of system knowledge data base, and additional management and operational level reports using real-time data.

Sandvik Coromant expects to get **increased availability, reliability and quality** of delivered products and services as well as increased capability to implement **green sustainable machining**, Further on, Sandvik Coromant sees a potential effect also in an increased capability to follow up products and services used and identification of new needs.

All these factors and earnings are applying for the automotive, machinery test, test facility, electronic and logistic companies, transportation and manufacturing industries. We see a very high potential of re-use capabilities for many different applications. Being successful, some are intended to be productized by a number of participating consortium members.

4.4.3 ISP-3 Data streams within collaborative environments

To increase efficiency in information exchange in collaborative environments, the involved organizations need to be able to do or provide the following:

- **use a shared vocabulary** (GSO – general social ontology)
- **agree upon a social communication protocol**
- **establish trust levels**
- **specify rules and policies for access control and IPR management**
- **collaborate**

Potential applications using the above are for instance foreseen at Hägglunds Drives for pre-sales decisions and sales activities, system dimensioning work, work related to system reconfiguration when the system has been deployed at a customer site, as well as for support for maintenance and trouble shooting.

Sandvik Coromant sees potential that the points above will make information exchange more reliable and efficient when collaborating with business partners.

Volvo CE sees a high of potential using the above results in applications where on the one hand user onsite and on the other hand central located user groups communicate and collaborate.

In general it can be stated that due to globalization efforts in the entire industry the importance to support collaboration processes is steadily growing. Where ever data streams are involved suited collaboration solutions need to be existing. SmartVortex can close this gap and provide with its research results solutions that can be used in all data stream oriented collaboration efforts, regardless of the industry section or domain.

5 IDENTIFICATION OF CONCEPTUAL REQUIREMENTS FOR THE DATA STREAM MANAGEMENT SYSTEM (DSMS) BASED ON THE DEFINED USE CASES

5.1 Scope of the SmartVortex usage cases related to data stream management

To cover the use cases the DSMS must be extremely flexible in terms of which kinds of data can be processed, what kind of algorithms can be implemented to process the data in the desired way, how the computations over the data streams can be scaled to many CPUs, which queries can be formulated over the streams, and how the results of the querying and processing is delivered to the user interface components for presentation. The performance of the DSMS must be good enough to give the desired answers to queries in real time, even when the system is scaled up to hundreds of thousands of simultaneous data sources.

5.2 Conclusions from the SmartVortex requirement analysis for the Data Stream Data Management (DSDM) architecture

For each of the use cases, a data processing application needs to be developed that performs the desired computations over the data streams and delivers the results of the queries in the desired way, with high enough performance. Developing these data processing applications using a DSMS is expected to be easier and less error prone, compared to developing the applications from scratch in a lower level programming language (e.g. C++). The DSMS components for processing data streams (e.g. searching, filtering, splitting, joining) are reused for all the use cases (and other applications), which improves maintainability and stability of the systems.

6 SMARTVORTEX ASSESSMENT FRAMEWORK FOR THE VALIDATION OF THE PROJECT OBJECTIVE

6.1 Test beds provided to validate the SmartVortex integrated subproject use cases

To verify that the SmartVortex technology works, test beds will be used to generate large volumes of data streams with some of the streams containing anomalies or values that should be noted, as well as used for aggregation and validation purposes.

6.1.1 Test bed use case ISP-1-1.1: FEM Simulation of mechanical design data

The test bed is defined by the data. Every computer can be used and no additional equipment is needed. The simulator the delivers the data must be defined here with referenced to the file formats produced by the simulator output. This use case will be done using the FEM simulation results generated during the computation of a stimulated example part with respect to noise and vibration. The used simulation system is called TOSCA and based on an exchangeable kernel. This kernel can be either, ABAQUS, ANSYS or NASTRAN, which is widely used in the industry an can be seen as standards for FEM simulation

6.1.2 Test beds use-case ISP-1-1.2 and ISP-2-2.1: Hägglunds Drives: Real time data streaming and stored data streams

The test bed has two options depending on possibility to get live data streams available over the Internet. Option 1 includes the use of live data streams and simulated data streams, whereas option 2 includes only simulated data streams. The test bed should comprise many data streams and within those one or more data streams should contain anomalies etc. that should be detected.

If option 1 is possible, the shredder at Hägglunds Drives will be used and run in provoked modes as well as for instance low oil pressure or similar conditions can be induced to produce sensor values that should be detected.

6.1.3 Test beds use case ISP-2-2.1: Sandvik Coromant: Real time data streaming

The test bed has three options depending on possibility to get data streams over the Internet combined with optional post process data.

1) includes the use of on-line streams
2) includes the use of off-line data streams (stored e.g. on file)
3) includes on-line and/or off-line data streams combined with optional post-process data such as surface roughness and tool wear

The test bed should comprise many data streams and within those one or more data streams should contain anomalies etc. that should be detected.

Sandvik Coromant aims to be able to capture and interlink all wanted data from a large number of machine tools and machining operations, and to be able to use these large data sets for comprehensive analyses, including identification of meta-knowledge.

6.1.4 Test beds use-case ISP-1-1.2 and ISP-2-2.3 Volvo CE: Real time data streaming and stream storage

The test bed has two options depending on possibility to get live data streams available over the Internet. Option 1 includes the use of live data streams and simulated data streams, whereas option 2 includes only simulated data streams. The test bed should comprise many data streams and within those one or more data streams should contain anomalies etc. that should be detected.

If option 1 is possible, the excavator in the pilot set-up at Volvo CE will be used and run in provoked modes as well as for instance low maintenance conditions can be induced to produce sensor values that should be detected.

Furthermore Volvo CE will equip two of their wheel loaders with all requested sensors and transmission capabilities to transfer the generated data stream. Even signals that are only giver by internal operations on the CAN bus or in the motor control unit will be made available but under specific confidentiality requirements.

6.1.5 Test bed for collaboration, cooperation and decision making

The test bed definition is still in progress but since there are no additional signals required we can state that all stream related information from actuators and sensors is available in the ISP-1 and ISP-2 test beds. The Meta data are under discussion since we have to deal here with person oriented information that require special confidentiality and care. Due to different legal backgrounds in the different European countries we need to respect laws like in Germany the “Betriebsverfassungsgesetz” and agreements with the local worker units. Nevertheless we believe there will be the possibility to use neutralized information that don’t allow conclusions on the real person mapped with its skill set, experience and additional person related background, that we consider as social graph.

6.1.6 Test bed for the graphical user interface in data stream querying

In order to arrive at measurable success criteria for the evaluation of the user interface regarding DSMS query expression and consumption, we will follow a three step approach:

1. Identify important, prototypical use-cases of the SmartVortex user interface regarding expression and consumption of DSMS queries.
2. Conduct user-studies in a controlled environment; if real data streams are missing or classified, we will make use of mockup data.
3. The findings of the controlled user studies are to be confirmed with on-site user studies.

The test bed specification mostly depends on requirements and test bed specifications for the identified use cases and scenarios.

The evaluation and assessment of the visual and multi-modal environment requires to identify, for each use case and scenario, the context of use, i.e., involved users, tasks to be performed, required equipment (hardware, software, machines and other materials), and the physical and social environments.

The first step consists of identifying primary end-users and characterizes them in terms of:

- organizational experience, role and qualifications
- system skill/knowledge
- task experience
- level of training
- technical capabilities and I/O device skills

To evaluate the overall usability and functionalities, it is required to clearly define tasks and goals that users are supposed to perform and achieve. Tasks and goals should be decomposed into sub-tasks and sub-goals, linked to each other (i.e., for each sub-task specify the corresponding sub-goals). For each

(sub-) goal relevant satisfaction criteria to be measured should be specified, and characteristics of (sub-)tasks which may influence usability should be described.
A set of key tasks will be selected to be executed in a controlled environment, in order to cover all relevant usability and functional aspects to be assessed.
In addition relevant characteristics of the required equipment need to be described. Visual and multi-modal interactions may require setting up a hardware infrastructure consisting of digital video cameras, microphones, loudspeakers, smartboards, multitouch devices and surfaces. If required, smartphones, Tablet PCs and other mobile devices should be considered as part of the test bed. At this early stage, attributes or performance characteristics of the hardware to be used cannot be clearly determined and should be considered in subsequent phases.

6.1.7 Test bed for the application of graphical query languages

To test the graphical query language the same test bed as in 6.1.6 defined will be used.

6.2 Success measures

For the success measures the KPI (Key Performance Indicator) method will be applied.
Key Performance Indicators define a set of values that are used to measure against. These raw sets of values, when fed to systems to summarize information against, are called indicators. Indicators identified as possible candidates for SmartVortex KPIs can be grouped into the following sub-categories:
• Quantitative indicators, which can be presented as a number.
• Practical indicators that interface with existing processes.
• Directional indicators, specifying whether the project is getting better or not.
• Actionable indicators, sufficiently within the project's control to effect change.
KPIs can be related to a variety of different aspects (technical, commercial, organizational, financial, social etc.). The key performance indicators received from SmartVortex partners will be analyzed, processed and linked to the evaluation criteria whenever appropriate.

The key performance indicators are established for each work package and its deliverables.
In the research and data modeling oriented work packages are mainly documents or algorithm defined as results that are contributing to the performance of the integrated sub-projects. These results are seen as fundaments for the establishment of the prototypes and demonstrators. For those deliverables mainly time and completion oriented measures or KPIs can be defined. The more important KPIs lay in the features that make the prototype useful, valuable and applicable, and result in expected reactions of the representation layers or the visualization or monitoring equipment. Therefor the KPIs are shown for both, ISPs and work packages in Annex 5.

7 CONCLUSIONS

The selected use cases are very generic for the applicability in other industrial domains. Even if they are adapted to specific evaluated problems of the consortium partners, they can easily be adapted to other scenarios and other dedicated problems related to streaming data. Changes are required in the content of the rule, archival and meta data DBMSs. This is obvious while working with different sensor sources, rules defining e.g. threshold, risks or trends and the combined required meta data descriptions. The visualization method is in general configurable and the used query language allows entering new or different queries for the selection of steam objects and the calculation of the data streams. The supported visualization devices representing the current state of the art display technologies spanning from handheld phones, over tablets to complex monitoring systems and are based on established visualization standards.

8 INDEX OF TABLES

9 INDEX OF PICTURES

10 EXPLANATION OF TERMS USED IN THIS DOCUMENT

Terms and their explanations used in the document:

API	Application Programming Interface
CAD	Computer Aided design
CAE	Computer Aided Engineering
CAM	Computer Aided Manufacturing
CAN-Bus	Controller Area Network bus (in Automobile network applications)
CPU	Central Processor Unit
DBMS	Data Base Management System
DSMS	Data Stream Management System
EMC	Electro Magnetic Compatibility
EMI	Electro Magnetic Interference
FAF	Fail And Fix
FEM	Finite Element Method
GPS	Global Positioning System
GSO	General Social Ontology
I/O	Input and Output
IC	Integrated Circuit
IPv4/6	Internet Protocol
IPR	Intellectual Property Rights
IT	Information Technology
KPI	Key Performance Indicator
L-Rating	Benchmark in DSMS measuring the traffic on Autobahn lanes
MAC	Media Access Control (internet address)
Matris	Machine performance measurement application at Volvo CE
MOPSsys	Application of s service provider at Hägglunds Drives
QoS	Quality of Service
PAP	Predict And Prevent
PC	Personal Computer
PCB	Printed Circuit Board
PDA	Personal Digital Assistant
PDM	Product Data Management
PLM	Product Lifecycle Management
RAM	Random Access Memory
SI	Signal Integrity
SSCS	Semantic Social Collaboration Suite
TBD	To Be Defined
TCO	Total Cost of Ownership
Wi-Fi	Data Transfer using Wireless Network Connection Technology

11 REFERENCES AND LITERATURE

Papers:

[M06] Maybury, M. T. (2006). **Expert Finding Systems**. *Intelligence*, (September).

[PC08] Petkova, D., & Croft, W. B. (2008). **Hierarchical Language Models for Expert Finding in Enterprise Corpora**. *International Journal on Artificial Intelligence Tools*, *17*(01), 5. doi: 10.1142/S0218213008003741.

[S09] Serdyukov, P. (2009). Search for expertise.

[ZT07] Zhang, J., Tang, J., & Li, J. (2007). LNCS 4443 **- Expert Finding in a Social Network**. *Advances in Databases: Concepts, Systems and Applications*, 1066-1069. doi: 10.1007/978-3-540-71703-4_106.

[MCK11] Report McKinsey & Company: **Big Data: The next frontier for innovation, competition, and productivity**. (May 2011)

[I03] Iung, B. (2003) **From remote maintenance to MAS-based e-maintenance of an industrial process**. Journal of Intelligent Manufacturing, 14, pp. 59-82.

[IL09] Iung, B., Levrat, E., Marquez, A.C., & Erbe, H. (2009) **Conceptual framework for e-Maintenance: Illustration by e-Maintenance technologies and platforms**. Annual Reviews in Control, 33, pp. 220-229.

[MA07] M. Abramovici, **Future Trends in Product Lifecycle Management (PLM)**, in *The Future of Product Development*, F.-L. Krause, Ed. Springer Berlin Heidelberg, 2007, pp. 665-674.

[MI08] X. G. Ming et al., **Collaborative process planning and manufacturing in product lifecycle management,** *Computers in Industry*, vol. 59, no. 2-3, pp. 154-166, Mar. 2008.

[BWB10] J. Brunsmann , W. Wilkes, H. Brocks (2010**). Exploiting Product and Service Lifecycle Data. International Conference on Product Lifecycle Management.**

Standards:

ISO9241-11: User interface standard please see:
http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf

SmartVortex Deliverables:

Delivery D2.1: **SmartVortex Information model**
Delivery D2.2: **An RDF-based model describing streams and other data from products, social networking systems, collaboration and products in use for scenarios**
Delivery D2.3: **SmartVortex Suite architecture and technical specification**
Delivery D3.1: **Semantic models of data streams**
Delivery D3.2: **Query language survey and selection criteria**
Delivery D16.2: **Software development quality assurance plan**

ANNEXES 1-5

CONTENT OF ANNEXES

Index of Pictures

11.1

1 ANNEX 1: VIEWS ON DATA STREAMS

1.1 Data streams used in the different applications

From an application oriented view the project takes five different kinds of data streams with different tasks in account:

a) **Streams stream which are produced from simulation programs get stored in large file repositories that are representing raw and unweighted simulation results.** These data sets are generated solving complex simulation equations for the electro-magnetic or thermal validation of e.g. electronic printed circuit boards, where the electrical behavior of the components in combination with the physical layout of a board need to be tested about thermal robustness or electro-magnetic conformity. Due to simulators using Monte-Carlo or other stochastic methods which are testing for instance radiation strength in very broad frequency spectra for each net on the board today extreme large data sets are produced which require reduction of data to allow recognizing the critical nets in the design. SmartVortex will support the engineers to focus faster and more dedicated on the real problem areas of the electrical designs.

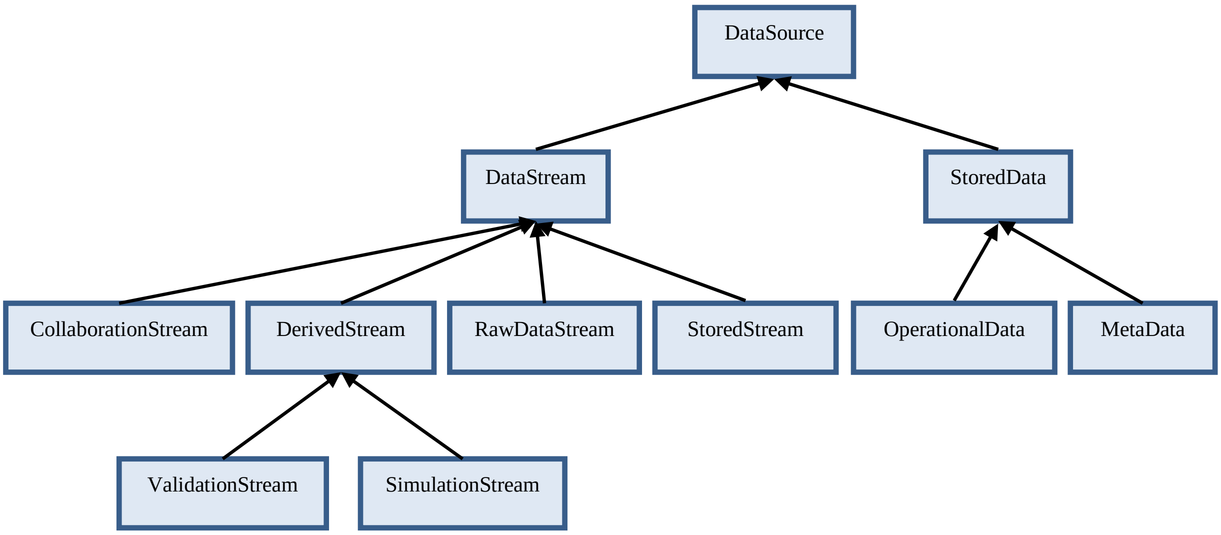
b) **Streams from real-time measurement, generated from actuator and sensors of the products which are in a test or operational mode.** This data can either originate from systems with very high measurement frequency or from a extreme volume of different products which are used under similar usage conditions. SmartVortex addresses both situations and must ensure that values which are in contradiction to each other get corrected, validated regarding thresholds and combined risks or give early warnings to the observers about fault prediction, maintenance needs and usage conditions which violate the contractual basis agreed for the usage. Furthermore the product or system boundary conditions as well as the optimal usage paradigm can be experienced from the designing engineers without damaging the systems or products itself. In parallel the current models used to simulate the future behavior can be adapted and improved. This has a direct impact on the financial health of the company reducing risk and preventing financial damages.

c) **Stream of design objects produced from many engineers during the design phase and the entire life cycle of a product.** In complex systems one single design object might be used as element or reference by many engineers working in different global acting design teams, various domains and performing sketching, authoring, simulation, validation and quality control with very different mostly non compatible or not integrated design tools. Even if a design object represents only one single item is used in many design tools and requires attention of the other engineers if the owner changes this object which might obsolete intended functions of another designer. SmartVortex will investigate methods that allow single design objects hidden in large data streams to notify the designers if given constraints of other design elements get violated or damaged.

d) **Streams of multi-media sourced collaboration objects.** Collaboration activities performed with collaboration and domain specific tools generate during the design and prototyping process data streams which include decision points, meeting notes, access logs, social interactions, chat logs, video conferencing transmissions and desktop sharing elements which are domain specific and contribute to the business process execution in the product life cycle. Traceable and focused information elements out of these massive streams are essential for the progress of the engineering activities. Since those streams cannot be processed manually SmartVortex can extract and tag the important information and decisions made during the collaboration session and will provide a platform to offer pregnant condensed information objects suited to the needs of the participants.

e) **Streams of higher level inferred events.** While analyzing any kind of raw data streams higher level events appear, during male functions of sensors, violation of thresholds, unusual behavior, unexpected or planned status changes or important changes in the design data used as base for the functional engineering design. These data streams in combination with the underlying data models for the reference function are producing higher level inferred data streams the need to start notification processes to inform immediately the consumer about trends, early recognition of starting male functions or identifying needs for soon appearing maintenance activities or human decisions to change the machinery behavior. SmartVortex can analyze this event oriented derived data stream and support predefined actions like specific notification, calls for interventions and other possible human or machinery interactions.

The application oriented view shows that within those different types of data streams, which are produced during all phases of the product life cycle, the recognition of pertinent information and its use to achieve better collaboration and decision making is the highest goal of the SmartVortex research.
During the requirement analysis it has been obvious that this application oriented view need to be interpreted in a different way for the implementation in a data model and for the scientific realization within this project. Out of a simplified view a more sophisticated streaming model has been generated that is shown in picture 21.



Picture 21: Data and data streams in the SmartVortex project

In the Integrated Subprojects ISP-1/2/3 the research concentrates on streams produced from sensors and actuators, simulation data repositories and collaboration activities. For those data streams the industrial partners are providing test data and test beds and will evaluate, test, validate and assess the achieved research results and combined solutions.

1.1.1 Application oriented view

The survey within the SmartVortex project is a preparation step within the requirement analysis to encounter if the envisioned and planned use cases are representative to solve problems which are common in the European industries and which solutions of these problems open up better conditions for functional engineering methods which are the key for service-oriented business models. The SmartVortex team has focused on the industrial partners to verify if requirements dedicated to data steam computing and management show similarities and can be unified to be applied with a single focused solution to each participant's industrial domain and service offer intention.

During the workshops which are documented in Annex 2 requirements to perform the following generic goals have been discussed:

a) Reduction of work iterations and reviewing time while evaluating very large data sets of raw data produced by multiple simulation runs of high complex simulation tools already when reading the data repositories.
b) Cost reduction at maintenance obligations performed with optimized planning of preventive maintenance activities and enlarged service intervals
c) Higher competitiveness reached by preparation and optimization of basis and entrance design parameters for the dimensioning of complex machinery systems
d) Decreased warranty and repair costs and higher availability of systems enabled with early fault trend recognition and prevention of damages in the machinery
e) Higher work quality due to improved user education, training and collaborative communication based on traceable interaction between humans and machinery elaborated with measurement of operational parameters
f) Deeper understanding of physical interrelations during the operational process of complex machinery and dependencies between tools, materials and design strategies and used machinery parameters.
g) Elimination of financial risk and liability entitlements while entering in business models and customer quotes based on service-oriented offers which are based on intangible assets like physical parameters (torque, speed, time to, ..) or commercial parameters (reliability, availability, down times,….). In this business models these parameters need to be contractually agreed as warranted characteristics under specific defined conditions.

Based on the goal to produce impact-oriented results which are applicable not only for the participating industrial project members the consortium has entered in a technical, scientific and impact-oriented survey to get an deeper understanding current practices, tools and techniques used for collaboration and decision making, and challenges involved by the use of data streams that and can be solved with technologies and methods based on data stream computing, data stream management and expected easy to learn and adopt access methodologies related to data streams as described in Annex 3 .

To recognize communalities in all reviewed processes and intended business models a common application oriented documentation sheet was generated and used during the workshops which shows the requirements in a concise manner documenting the life cycle of data stream objects from generation, over consumption to resulting actions. To enable the review process and allow value oriented results the actions as now defined, provide a starting point including identifiers of the required actions that can be used while implementing the prototype applications. This excludes a detailed modeling of all actions that later-on will be required in a productive environment.

Raw data (Streams)	Meta data (stored)	Derived data streams	Validation	Action

Picture 22 : Workshop data evaluation schema

Annex 2 explains the chosen methodology to evaluate the user requirements and needs and investigations in possible use case scenarios using the above mentioned process need analysis and its documentation in more detail.

1.1.2 Technology oriented view on data stream management

As already mentioned the application oriented view would not allow to implement a generic data model (deliverable D2.2) for the required implementation. The simplified application oriented view on data streams requires a better definition of data streams that need to be handled in a data stream management system. Not leaving the application oriented view a generic technological view has been defined which will be taken for the implementation of the data model. This can now easily been mapped to the streaming activities within the use cases that have been discussed with the industrial partners. There also five basic steaming elements exit.

The project takes these five different kinds of technology oriented data streams with different resulting tasks in account:

a) **Simulation streams** are produced by some computational system. For example, the desired limits for temperature readings over time may be continuously computed as a stream by a simulator. In other cases simulation streams are stored in files generated off-line by a simulation program solving complex simulation equations for the electro-magnetic or thermal validation of e.g. electronic printed circuit boards, where the electrical behavior of the components in combination with the physical layout of a board need to be tested about thermal robustness or electro-magnetic conformity. Due to simulators using Monte-Carlo or other stochastic methods which are testing for instance radiation strength in very broad frequency spectra for each net on the board, extreme large data sets are produced which require reduction of data to allow recognizing the critical nets in the design. SmartVortex will support the engineers to focus faster and more dedicated on the real problem areas of the electrical designs.
b) **Raw data streams** from real-time measurements, generated from actuator and sensors of the products which are in a test or operational mode. This data can either originate from systems with very high measurement frequency or from an extreme volume of different products which are used under similar usage conditions. SmartVortex addresses both situations and must ensure that values which are in contradiction to each other get corrected, validated regarding thresholds and combined risks or give early warnings to the observers about fault prediction, maintenance needs and usage conditions which violate the contractual basis agreed for the usage. Furthermore the product or system boundary conditions as well as the optimal usage paradigm can be experienced from the designing engineers without damaging the systems or products itself. In parallel the current models used to simulate the future behavior can be adapted and improved. This has a direct impact on the financial health of the company reducing risk and preventing financial damages.
c) **Validation data streams** are data streams indicating unexpected deviations in data from other streams. Values of validation streams are often continuous, e.g. indicating to what degree a measured value differs from a simulated computed value. In other cases they are discrete, e.g., while analyzing any kind of data streams events are inferred indicating male functions of sensors, violation of thresholds, unusual behavior, unexpected or planned status changes or important changes in the design data used as base for the functional engineering design. Such **notification events** indicating non-expected behavior or significant deviations cause the need to start notification processes to inform immediately the consumer about trends, early recognition of starting male functions or identifying needs for soon appearing maintenance activities or human decisions to change the machinery behavior. SmartVortex use validation streams to support predefined actions like specific notification, calls for interventions, and other possible human or machinery interactions.
d) **Design object streams** produced from many engineers during the design phase and the entire life cycle of a product**.** In complex systems one single design object might be used as element or reference by many engineers working in different global acting design teams, various domains and performing sketching, authoring, simulation, validation and quality control with very different mostly non compatible or not integrated design tools. Even if a design object

represents only one single item is used in many design tools and requires attention of the other engineers if the owner changes this object which might obsolete intended functions of another designer. SmartVortex will investigate methods that allow single design objects hidden in large data streams to notify the designers if given constraints of other design elements get violated or damaged.

e) **Collaboration streams** of multi-media sourced collaboration objects**.** Collaboration activities performed with collaboration and domain specific tools generate during the design and prototyping process data streams which include decision points, meeting notes, access logs, social interactions, chat logs, video conferencing transmissions and desktop sharing elements which are domain specific and contribute to the business process execution in the product life cycle. Traceable and focused information elements out of these massive streams are essential for the progress of the engineering activities. Since those streams cannot be processed manually SmartVortex can extract and tag the important information and decisions made during the collaboration session and will provide a platform to offer pregnant condensed information objects suited to the needs of the participants.

Picture 1 classifies the different kinds of **data sources** processed by Smart Vortex. Both data stored in **repositories** and produced in real-time as **data streams** need to be processed. A particular kind of repository data is **meta-data** describing properties about artifacts (machinery, collaborations, simulations, etc.). **Operational data** from regular enterprise databases are used by SmartVortex for example for accessing customer, pricing, materials, or marketing data. If desired and possible, streams can be saved in a repository as **stored streams** by a program (e.g. a simulator) and replayed later.

Data streams can either be **raw data streams** produced by some artifact or human, or **derived data streams** whose data is continuously computed based other data streams in combination with product meta-data and data stored in repositories. Derived streams can contain both continuous and discrete values. An example of a continuous value is the power consumption of some equipment in use estimated by a mathematical model over other data streams and product meta-data. A discrete value can be, e.g., an action event generated when non-expected behavior of equipment is inferred by a validator.

We notice that the five Smart Vortex kinds of streams listed above can be either raw data streams, derived streams, or stored streams depending on how they are produced.

2 ANNEX 2: METHODOLOGY ADAPTED

2.1 The process of capturing user requirements for ISP-1, ISP-2, ISP-3

While establishing the DOW for the SmartVortex project possible user scenarios have already been discussed in detail. Especially the expected impact provided with the vision of anticipated outcomes were subject of the project proposal. Due to this the consortium was able to enter directly into talks and discussions in workshops and extended brainstorming sessions to drill down the initial ideas to concrete requirements and especially to recognize added values that this project will generate. The workshops were hosted by the industrial partners, where the specialists of all affected domains and departments and the dedicated research members of the assigned work packages have taken part. In several iterations, done either with additional workshops, conference calls or personal on-to-one meetings a common brainstorming session hosted by the Technical University in Delft (TUD) has been prepared. In this session the consortium identified additional research and clarification potentials which have been in following sessions got more deeply specified hosted by LTU and UU.

To elicit requirements for ISP-1 and the collaboration and decision support tools envisioned in the Smart Vortex project a common workshop and brainstorming session with all partners involved was given the following goals:

1. Get an overview of key success factors for collaboration and decision-making
2. Get an overview of processes that are in the focus of decision and collaboration support
3. Identify activities in these processes, decisions and data streams/documents & support used for these decisions
4. Get insight in current practices, tools and techniques used for collaboration and decision-making, and challenges involved.

To support the workshop a group support system was used and in addition SMARTboards for collaborative modeling of the requirements. The group support system helped the team to capture the requirements that were discussed in the meeting. Furthermore, working with the group support system and SMARTboards for collaborative modeling gave the participants insights in different innovative tools and opportunities for collaboration support.
The group first brainstormed on the success factors for collaboration and decision-making both from an academic and from a business perspective. This helped the team to see differences in perspectives on successful collaboration and decision-making within the product life cycle.
A key success factor from an academic perspective is shared understanding among engineers to improve rational decision making, and process and tool support to improve effectiveness of collaboration. From an industry perspective ease of use and accessibility and structuration of information were critical success factors. This shows an interesting difference in focus on tacit versus explicit forms of information sharing, interaction and consensus.
Next, an overview of activities and decisions was created, that could benefit from collaboration support and the needs for collaboration and decision support in these activities. Also an overview of existing collaboration and decision tools was established to avoid duplicate work by re-inventing the wheel.
Based on this brainstorm, four focus scenarios for collaboration and decision support needs were established:

1. Acting on (live) data streams, supporting decision making about these data streams or simulations
2. General engineer to engineer collaboration support and interaction support
3. Problem solving or crisis management, acting on 'alarms' an also acting proactive, seeing trends, forecasting

4. User involvement, capturing knowledge, using user feedback in design and decision making, learning from unspecified use, adoptions, new uses

Based on these four scenarios, heterogeneous teams of academics and business participants created conceptual models of the envisioned collaboration and decision support systems. The models were then presented to the group for discussion and feedback and finally accepted.

2.2 Workshop performed with the Industrial Partners

In preparation to this general common brainstorming and planning team event there have been several workshops, meetings and conference calls performed to cover the project intentions, needs and requirements of the industrial participants. These workshops were set up with the dedicated industrial partner and the work package contributors involved in the different ISPs.

2.2.1 Workshops related to ISP-1

Requirement workshops related to mainly ISP-1 have been conducted according to listing below, to understand the case environments, limitations and gather the requirements:

- 09-Feb at Philips CE in Brugge, Belgium
- 23-Feb at Technical University Delft in Delft, Netherlands
- 06-May at InConTec in Burghaslach, Germany
- 23-May at Philips CE in Eindhoven, Netherland

2.2.2 Workshops related to ISP-2

Requirement workshops related to mainly ISP-2 have been conducted according to the following listing, to understand the case environments, limitations and gather the requirements:

- 24-Jan at Sandvik Coromant in Sandviken, Sweden.
- 7/8-Feb at Volvo CE in Eskilstuna, Sweden.
- 14-Feb at Sandvik Coromant in Sandviken, Sweden.
- 15-Feb at Hägglunds Drives in Örnsköldsvik, Sweden.
- 16-Feb at Hägglunds Drives in Örnsköldsvik, Sweden.
- 17-Feb at Volvo CE in Eskilstuna, Sweden.
- 23-Mar at Hägglunds Drives in Örnsköldsvik, Sweden

2.2.3 Workshops related to ISP-3

Requirement workshops related to mainly ISP-3 have been conducted according to listing below, to understand the case environments, limitations and gather the requirements:

- 23-Feb at Technical University Delft in Delft, Netherlands.
- 02-Mar at Uppsala University in Uppsala, Sweden.

2.3 Identification of user base, legacy systems and service providers

Mirroring the essential finding in the workshops with the industry partners it has been recognized as significant that the discussed possible use scenarios can be treated as very generic and work performed in one of the scenarios can definitely be used supporting the other scenarios. This gave the consortium the assurance that the planned outcome of the project is not only suitable for the participating industry partners but also applicable to the rest of the European or global acting industries which intend to start with service oriented product offers that are based on intangible assets.

No real legacy systems, which could be compared with the aim of the SmartVortex project were identified. Similar approaches but without the planned performance, power, elegance and user friendliness could be found in “Analog Recording Technologies” or “Hybrid Computing Equipment”. But this technology is outdated today and not able to cover the filter and computation process of systems that require fast and immediate access to succinct case describing information and is very limited in the threshold or boundary condition management and therefore not able to be used in today’s digital design and verification world of functional engineering.

Other systems like data logging systems can only cover the specific predefined situations, but are not able to perform complex computing tasks with large and fast changing data streams. One of the important technologies which can be used to differentiate the system status and behavior is pattern recognition which cannot be done based on the above mentioned technologies.

Real service providers which are using comparable technologies are not really visible or they are part of the SmartVortex research team like

- Alkit, which has done measurement of test drives in the automotive and tire industries establishing wireless and broadband technology to transmit the data from the vehicle to the engineers in the laboratories or
- InConTec where the founders been for many years been involved in simulation of very complex electromagnetic behaviors of printed and integrated circuits which are producing extremely large data volume simulating radiation emissions.

Both companies which are active in this market are keen to get technologies as proposed in the SmartVortex project to use data stream sources and management accretions to solve problems easier and better in their market segment. The review of the possible competition has not brought evident results which can be compared with the SmartVortex approach.

3 ANNEX 3: DETERMINATION OF THE USER BASE

3.1 Common requirements to build up a data stream management infrastructure

3.1.1 Data capture requirements

The data capture system must be able to sample a sensor at a sample rate that is sufficient for the particular analysis to be performed. The exact value of the sample rate depends on what kind of analysis is to be performed, which is application specific.
If the sensor is analog, the data capture system must be able to digitize the sensor signal at a precision that is good enough for the analysis, which again is application specific.
The latency of the measurement data capture (including sampling and digitization) must be low enough to be able to keep up with the sample rate required by the analysis.
The data capture system must be installable inside or in the absolute vicinity of the product containing the sensor(s) to be sampled. For mobile systems (e.g. vehicles) this means that the data capture system must move along with the system. As a consequence, the physical dimensions and power supply requirements must be compatible with the limitations imposed by the application and operating environment.
The data capture system must have sufficient processing power to perform at least a simple filtering mechanism based on comparing each sample value with a binary mask and filter value defining which bits of the sample must match the filter value.
The data capture system must have a sufficient amount of onboard memory to perform at least the data capture mentioned above and simple filtering mechanism and to communicate the data over a network connection with the properties described below.
The data capture system must be programmable in a high-level programming language (e.g. C), and a software development kit must be available so that the data capture and primary filtering mechanisms required for the application can be developed by the SmartVortex partners.
The data capture system must have an IP stack implementation and a network programming API (e.g. UNIX sockets).
The data capture system must have a network interface providing IP (IPv4/6) network access with a bandwidth that is sufficient to communicate the captured sensor data in real time. This may be accomplished by using data compression techniques, if the data capture system has sufficient processing power for this. The layer 2 (MAC) technology of the network connection can be any technology that complies with the bandwidth and latency requirements of the application.
It must be possible to produce such a data capture system at a price level that makes it possible to scale up the number of systems required by a potential customer to a number that is dependent on the application. The precise price level is also application-specific and should be estimated at a time when the exploitation of the SmartVortex results is expected to start. This start point is estimated to be given with the begin of the in the SmartVortex exploitation activities.

3.1.2 Data communication network requirements

The communication network must provide IPv4 (and optionally IPv6) connectivity for all data capture systems and for the backend data processing system.
The bandwidth, latency and jitter (variations in latency) of the network must, in normal operation of the system, be compliant with the latency and data rate requirements of the application. Strict guaranties regarding latency and bandwidth (i.e. QoS mechanisms) are not a general requirement, but may apply so for certain applications.
If the data capture system is installed in a mobile system (e.g. a vehicle), the network access technology of the data capture systems must be wireless and provide roaming capabilities and coverage sufficient for the application at the geographical locations where the system is to be used.

The network connection of the backend data processing system should have high enough capacity regarding its bandwidth to support the aggregation of many simultaneously connected data capture systems communicating at their operational data stream rate. The exact number of simultaneously supported communicating data sources is application-specific.

3.1.3 Backend data processing system requirements

The data processing system should be based on commercially available computer hardware running a commercially available (or open source) operating system such as Windows or Linux.
The data processing system should be able to get extended to a computer cluster configuration – i.e., the operating system should have support for this.
A large number of simultaneous data streams analyses should be supported by the backend data processing system. The exact number is application-specific.
The CPU (or CPUs for multi-core or cluster configurations) should be powerful enough for the (application-specific) aggregation of data stream analyses depending on the given requirements.
The (RAM) memory of the backend data processing should have a big enough capacity for the (application-specific) aggregation of data stream analyses.
The secondary storage (disk) of the backend data processing should be big enough for the (application specific) aggregate data stream analysis requirements.
The backend data processing system should provide scalable search and processing of data streams by means of a data stream management system (DSMS)
The DSMS should provide a query language for formulating general queries over the data streams.
The DSMS should provide a plug-in framework for applying application-specific computations over the data streams.
The data processing system should provide an access mechanism (e.g. a web Portal) giving the end users of the system access to the data and the search and processing mechanisms provided by the DSMS.

3.1.4 Conceptual frame work for data stream processing

We expect that the tools we mentioned in the ISP1-3 use cases generate large quantity of data by the interaction of users during their cooperation/collaboration sessions. However, the use of such data is not viable in decision-making processes, due to its high amount. Generally, such processes require already transformed data that aggregates a certain value and therefore can support decision makers performing their tasks.

In this section we intend to introduce a conceptual framework to transform raw data into processed information that can be presented to the user while taking decisions or to provide as input to software agents that are able to decide autonomously. This framework is detailed in picture 1.

The framework is composed of 4 layers: raw data, derived data, validated data and action. Additionally, the framework also contains the description of the raw data, represented by the Meta Data entity. The characteristics of each entity will be detailed in the next paragraphs.

The **Raw Data** represents the data that is generated by the collaboration tools and by the interaction of its user. Such kind of data can be generally interpreted as the logging data of the system. For example, two participants of a chatting session – Jordan and Stephan – can exchange some messages to schedule a meeting. In this case, we intend to store the information that is being exchanged by the participants.

The information transmitted through the data streams just represents the amount of data generated by the collaboration tools. As our major goal is to support different kinds of adaptation within our tools, we need to gather as much data as possible besides the outcomes of the raw data. At this point we are interest in collecting information about entities that are related to the outcomes of the raw data layer, in this case: the tools that are being used, the users that employ those tools and the processes that are carried out from the users. We classify those three variables that we just mentioned as being part of our **Meta Data** layer.

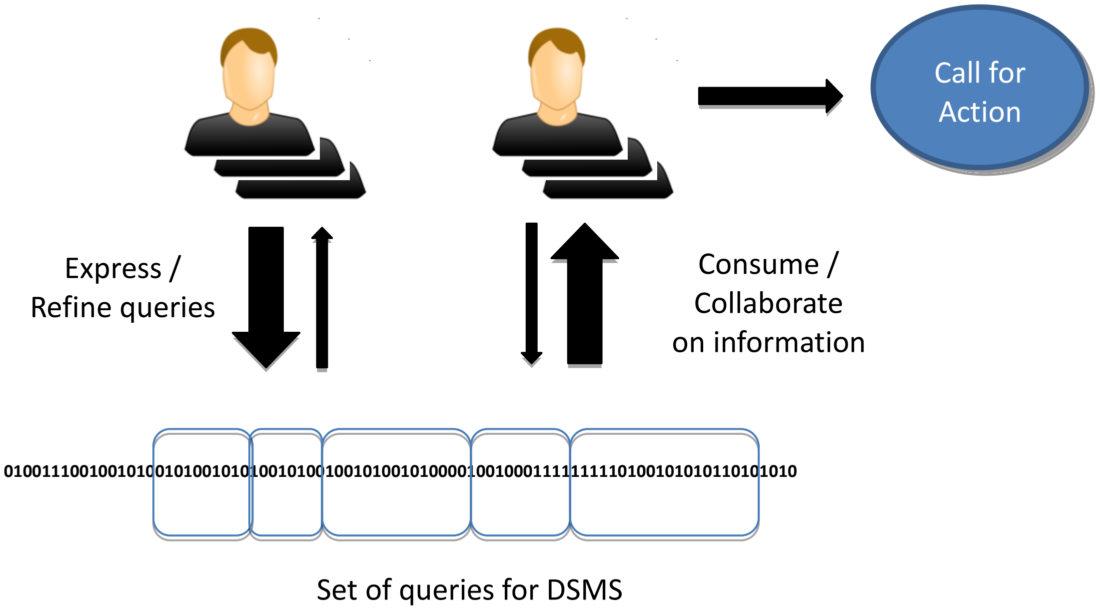
One of the goals of this framework is to facilitate the use of data to be processed by an analyst to ease the decision-making process. As the collaboration data streams contain high amounts of information (coming from the raw data) it would be very complex for a human analyst to process all of the raw data without any filtering mechanism that can transform the high amounts of data into processed information, reducing its volume. That is why the next layer in the framework represents the **Derived Data** layer. The goal of this layer is to automatically process incoming data from the raw data layer using the Meta-data information, transforming data into information. For example, the system is able to calculate from the data streams the fuel's lifecycle usage of an excavator not just its statistics for hours or days.

The derived data represents a proposal of transformed data from data streams with the support of Meta-data, making some system variables updated. However, for the decision-making process, a continuous update might not be suitable because the users might not have enough time to check and react to the variation of these data. Therefore, we intend to create a layer of rules (**Validate Data**) that constantly analyze the data to let the user know about information preprogrammed by a rule.

The transformation steps performed through the Raw Data, Derived Data and Validated Data prepare the data coming from the collaboration data streams to be processed and presented to decision-making stakeholders to that they can collaboratively agree upon a decision to take. We defined this process within the last layer of our framework; the **Actions** layer. For example, whenever a equipment is about to reach a downtime interval for bad usage, experts should be brought together into a collaborative environment, containing telemetric data of the equipment for a possible diagnosis, to take an action as fast as possible to avoid such time and may be also to provide possible interventions for collaboration.

3.1.5 User interface requirements

The user interface of the system should provide the end-users with the means to collaborate on, consume and reason about the raw and derived data streams as well as to manipulate and create new data stream computation instances. Using graphical and multi-modal query languages, the complexities of the underlying, textual query language are hidden from the end-users for a more intuitive access to the information contained within the data streams.



Picture 23: Concepts of Multimodal Interaction with Data Streams

Suitable interaction conceptualizations are to be researched, taking into account the different use cases and scenarios in which the system is utilized, as well as the challenges arising from the volatile and real-time nature of the underlying data streams. The interaction conceptualizations are to be integrated into a collaboration tool, for multiple end-users with their roles to reason about the information and coordinate and agree upon and take actions. Furthermore part of the user interface will use appropriate modalities to notify end-users proactively about noteworthy occurrences regarding the system's state with respect to the users' observable activities.

The collaborative visual environment should enable two different types of interactions, providing support to:

1. synchronous collaboration between co-located users
2. asynchronous distributed collaboration between remote users

To support these tasks and enable collaboration among different workers, the user interface has to identify and provide accessible, flexible and editable artifacts for both individual-based and group-based exploration of data streams. This requires the definition of a group interaction space for collaborative data analysis, where users can interactively access data streams, as well as create, layout, arrange and share artifacts connected with data streams. Each artifact should be enriched with contextual information (such as labels, scales, legends, etc.) that allows users to correctly interpret the information being displayed.

Users collaborating in a shared group interaction space should be able to jointly interact with a single shared view of data streams, or to create different views of the same data streams in a personalized workspace that can optionally be shared with other participants and mapped to or fuse with other views. By enabling the sharing and coordination of multiple views of data streams in a common workspace, the visual environment allows users with different backgrounds and experience to reason on complementary data stream representations.

Additional requirements, particularly targeted to asynchronous distributed collaboration, include the ability to:

- allow users to trigger conversations and remote social interactions via the provided collaboration tools
- use notifications of actions performed by other users (edited views, updates in artifacts, requests or comments issued by other users, etc.)
- use metaphors to represent the timing of performed actions and to access and display the actions' history for the artifacts

- provide representations of user profiles, including their roles, backgrounds and skills (as provided by social networks)

Moreover, the interface should provide users having different skills, roles and capabilities with the possibility to explore and analyze complex data streams with little or no assistance from technical practitioners. To effectively support inexperienced or non-technical users, display and interaction techniques should partly or totally automate the selection of metaphors and interaction modalities, in order to constrain the configuration parameters that users have to define. On the other side, experienced users and technical people should be able to fine tune the appearance and behavior of their displays via the user interface.
As collaboration practices often include mobility and remote interaction, the user interface layer should also provide representations and interaction models for pervasive computing scenarios, where multiple devices with different capabilities (e.g., in terms display size, screen resolution, available devices and computational resources) and demanding little or no technical knowledge (e.g., PDAs and Tablet PCs) are used.

Apart from all technical features mentioned above, the user interface should support also the main marketing related user criteria which are:

- Easy to learn with as much as possible intuitive handling elements
- Fast reaction capabilities to risk situations
- Easy adaptability if extensions are required
- Multi-language capabilities (Unicode) with external managed accessible language files

3.2 Computation of simulation data streams deriving from stored simulation results

In general simulators are computer programs simulating the behavior of different artifacts based on parameters characterizing the simulated artifact. Traditional simulators are programs running some mathematical models to produce results stored in regular files. The results can also be stored in regular databases after proper reformatting. In the context of data streams it can be noticed that simulation methods also can be applied on live streams, rather than being run in batch over stored files. Such a simulator produces a **simulation stream** by implementing a simulation model. As a simple example, one could simulate the allowed temperature fluctuations in a building over time by a mathematical model knowing properties of the building. Often data produced by the simulation program will be stored in a repository and converted to a stream by reading from the repository. For example, the simulated temperature fluctuations over the year may be stored in a disk file and replayed as a live stream.

A **stream validator** can compare a derived stream computed by a simulation program with the actual raw data (e.g. temperature) readings. The stream validator produces a **validation stream**. A validation stream can be a derived stream where the degree of deviations from the expected simulated result is computed. A validation stream can also be a stream of notification events signaling the need for action by some agent.

When significant deviations are detected by the stream validator it may indicate either faulty behavior in the monitored artifact or that the simulation model is invalid. The streamed validation program with some certainty verifies that the simulation model is correct if no deviations between observed and simulated behavior are detected

In the user survey the following scenarios related to stored data streams have been identified:

3.2.1 Scenario regarding simulation data stream validation, extraction and reduction

In any simulation process the focus is given to data that are produced in very high volumes by computational processes like typically FEM, Vibration Sound or EMI simulators are performing. That will be performed taking FEM simulators as used in the mechanical design validation process to reduce the time spent for completing the FEM simulations results in the area of vibration, and sound generation. Recognizing the trend of a simulation result and predicting the final result will help to restart simulation cycles with new stimuli much earlier and allow getting better analytical results that with the currently used Monte-Carlo Method only.

3.2.2 Scenario at Hägglund and Volvo CE regarding stored data streams generated from real-time measurement activities

In some situations where data streams originating from real-time measurements cannot be transferred to the computing centre of the partners due to missing communication infrastructure or extreme high cost when using satellite or phone connections it is intended to store the data streams in a random access way and either to dump and transmit only a specific timeframe that is required to identify situations that have been generating specific trend or to allow the service people to take a copy of the data streams on portable devices. At Hägglunds such situations appear due to the wide distributed installed base which is spanned over many global locations

3.3 Computation of real-time measured data streams deriving from sensors and actuators

While in the computation of stored data streams the data streams originate from program execution in a computing system, in the **Integrated Sub-Project 2** (**ISP-2**) the data streams are generated real-time from sensors actuators or control-units which are initiated by the machinery operation process. The operation condition of the machinery is providing the input for the sensors in real-time manner. This creates different aspects to data stream management, transmission and validation, which will be determined in the use cases which are concentrated in ISP-2

In this this subproject the data streams are generated from actual sensor measurements with either a high frequency of measures or a high volume of sensors which in parallel are transferred to one center which interprets those volume data as combined stream.
The validation process is similar to the validation used in ISP-1 but the algorithm will be used under real-time conditions with varying by volume of data or by fast changing sampling frequencies.

3.3.1 Sandvik Coromant scenario related to real-time data streams

In the Sandvik Coromant scenario some sensors provide high frequencies and other sensors are acting in combination. Sandvik Coromant envisions scenarios where high volumes of tooling machines are delivering in parallel measurements that also can be taken to get better view on the communalities of the usage scenarios at their customers.

3.3.2 Hägglunds Drives scenario related to real-time data streams

In the Hägglunds Drive Systems use scenario the data stream are generated from sensors of Electro-hydraulic drive systems with high frequency sample rates. The installed motors are located around the globe and in some cases located in parts of the world where it is easy to transport the logged data online over a network. In these cases, parts the data stream examination can be done locally for a faster system fault action a high amount of data will be used centrally to give a wider picture of the system usage.

3.3.3 Volvo CE scenario related to real-time data streams

In the Volvo CE scenarios the data streams are generated either from the sensors itself or taken out of the motor control unit of their construction machines. This will start during the project with two special equipped wheel loaders but Volvo intend to spread this observation and controlling measurements over complete fleets of construction machines located around the globe. In addition to the previous described use cases the moving location of their machines plays an important role.

3.4 Computation of collaborative data streams resulting out of cooperation and decision making processes

Decision-making processes generally require the support of various kinds of tools and the interaction of participants – cooperation processes. At this section we intend to present current tools used in the industry by employees to support them to cooperate with each other and to support them in making decisions.

We collected a list of tools used mainly by the industry partners of the Smart Vortex project - Sandvik, Hägglunds and Volvo – during the ISP-1 workshop in Delft. We performed a brainstorming session, conducted by a professional facilitator, to support the process of gathering such information. After the session we compiled a list of tools and created categories to cluster similar types of tools.

Category	Example Tools
Instant Messaging	**MSN, Skype, Google Talk and ICQ**
SVN Repositories	**Subclipse, JIRA, Notepad++ SVN Plugin**
Brainstorming	**Thinktank**
Mark-up/Tagging Tools	**Google SideWiki.**
Shared Calendar	**Google Calendar, Outlook and Lotus Notes**
E-mail	**Gmail, Outlook, Hotmail**
Event Scheduling	**Doodle, Outlook, Lotus Notes**
Screen Sharing	**GoToMeeting, Jing, Skype, VNC and Microsoft Remote Connection**
Screen Recording	**GoToMeeting and Jing, Mikogo, WebEx, Adobe Connect**
Document Co-Edition	**spreadsheet, drawing, document, presentation (Google Docs)**
Document Sharing	**Drop box and rapid share**
Conversation	**Skype, conventional telephone, cell phone**
Collaborative Drawing	**GoToMeeting, Skype**
Bug Tracking	**BugAware, BugTrack and Clearquest.**

Table 1: Categorization of cooperation and decision-making tools.

The above compilation of tools supports us in defining the information that can be generated within collaboration data streams. Based on those categories we can define general abstract objects that can be manipulated and abstract actions that can be performed as part of a whole class of tools and not just of one specific tool. For example, the Instant Messaging class of tools might have an abstract object called message and actions to manipulate the message as send or receive.

These kind of abstract objects and actions represent the basis of the model (further detailed in section 4.3 of D1.2 main document) that describes the collaboration data streams resulting from the cooperation among the participants that are taking decisions.

3.5 Graphical and multi modal query language to access information within data streams

There are two parts to the query language as it is offered to the end-user: the creation of data processing instances as queries over the DSMS, and the consumption of the information delivered by these queries.
Hiding the complexity of the textual query language potentially limits the expressiveness for queries to be expressed. The available semantic subset of the textual query language is a function of the employed modality. To select meaningful subsets requires the identification of suitable trade-offs between expressiveness and ease of use for the different modalities and user's expertise.
Similar considerations apply to the consumption of the information from the expressed queries. The different modalities and employed interaction techniques determine the dimensionality and frequency of information from the DSMS to be perceived and processed by the user with acceptable cognitive load.
The visual and multi-modal query language should support the access data streams (and related metadata) for both local and remote users in various formats. It should allow users to define queries over data streams, parameterize them and create and position resulting views in the personal or shared workspace.
Query building and configuration should be based on conceptualizations with a formal syntax and semantics to be matched with the syntax and semantics of the underlying textual query language. The query building environment has to provide visual and multi-modal tools that allow unskilled end-users to perform data stream queries navigate the results and mine other hidden relevant information. Graphical and multi-modal primitives, as well as basic visual and multi-modal queries, should be made available via a user query library extendable by the user by adding and storing composite queries.
The environment should enable multiple combinations and compositions of conceptualizations to reflect advanced query constructs and allow users to aggregate/disaggregate, filter and browse data in order to focus on relevant information. Depending on the textual query language capabilities, visual and multi-modal query formulation tools should allow both intensional and extensional queries operating on the data stream model or their representation.

4 ANNEX 4: IDENTIFICATION OF SMARTVORTEX USAGE SCENARIOS AND RELATED USECASES PROVIDED FOR THE THREE INTEGRATED SUB-PROJECTS ISP-1, ISP-2, ISP-3

4.1 Business-oriented scope of the SmartVortex usage scenarios

The use cases which have been identified with all industrial partners must create sustainable business value to have impact for the entire industry. The value potential has been identified in the user base and has manifested the selection of the intended or possible use cases.

4.1.1 Data and time reduction handling complex simulation processes in design processes

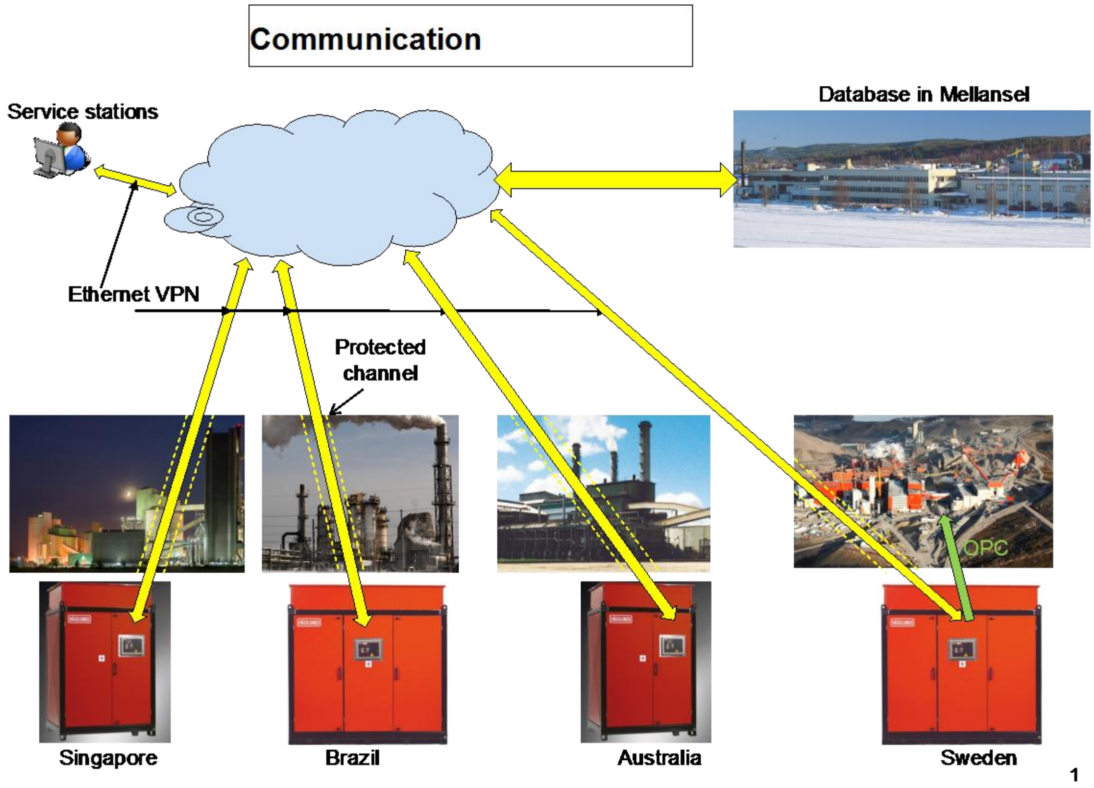
A usage scenario has been identified in the area of FEM simulation, especially Noise and vibration simulation. All FEM simulation driven design activities operating with different frequencies provide similar cases , where extremely large simulation resulting data streams need to be reduced to those which can give the simulation engineer a direct hint to occurring problems identified by the simulation run. This dedicated offer allows the engineers who are rare and cost intensive specialists in its area to concentrate on the real problems and to speed up the steps that can resolve the recognized problems with design related changes or recommendations.

4.1.2 Reduced transmission cost of large data streams enabled with random access technologies, transmission prioritization and local storage capabilities (Hägglunds drives and Volvo CE)

At Hägglunds and at Volvo CE their products, motors and construction machines, are located in areas which often miss the required fast network connection for central analysis, since this is combined with additional high infrastructure cost. This makes it sometimes impossible to receive the data online. Here it is important to be able to examine the data stream locally to reduce the stored information, limit of the data amount and to keep down the data transportation costs. In these cases the validation results need to discover dangerous trends and local required emergency actions. Furthermore system need to set priorities what date can be stored and transferred on request or a subject of long-term machine usage investigation which can be transmitted as block or stochastic result in a compressed and meaningful manner on request, cyclic or with a dump from the repository initiated by a service person at a customer visit.

4.1.3 Improved planning of preventive maintenance and dimensioning of electro hydraulic drive systems with establishment of novel physical parameter based sales and service models. (Hägglunds Drives)

Hägglunds Drives we recognized that providing high reliability and availability will become more and more important in industrial settings. New business model can only emerge if functions are provided and not only products get combined with services. This requires control of many aspects, and for instance real-time monitoring of critical systems allows for detection and prevention of failures so that measures can be applied when indications of a failure arise instead of when they have occurred. To be able to monitor critical systems at customer sites requires not only adequate IT-infrastructures but also the ability for efficient large scale management and analysis of the multiple product data streams generated at the customer sites during the product life-cycles. In addition, meta-knowledge about the data streams needs to be developed to be able to understand them.

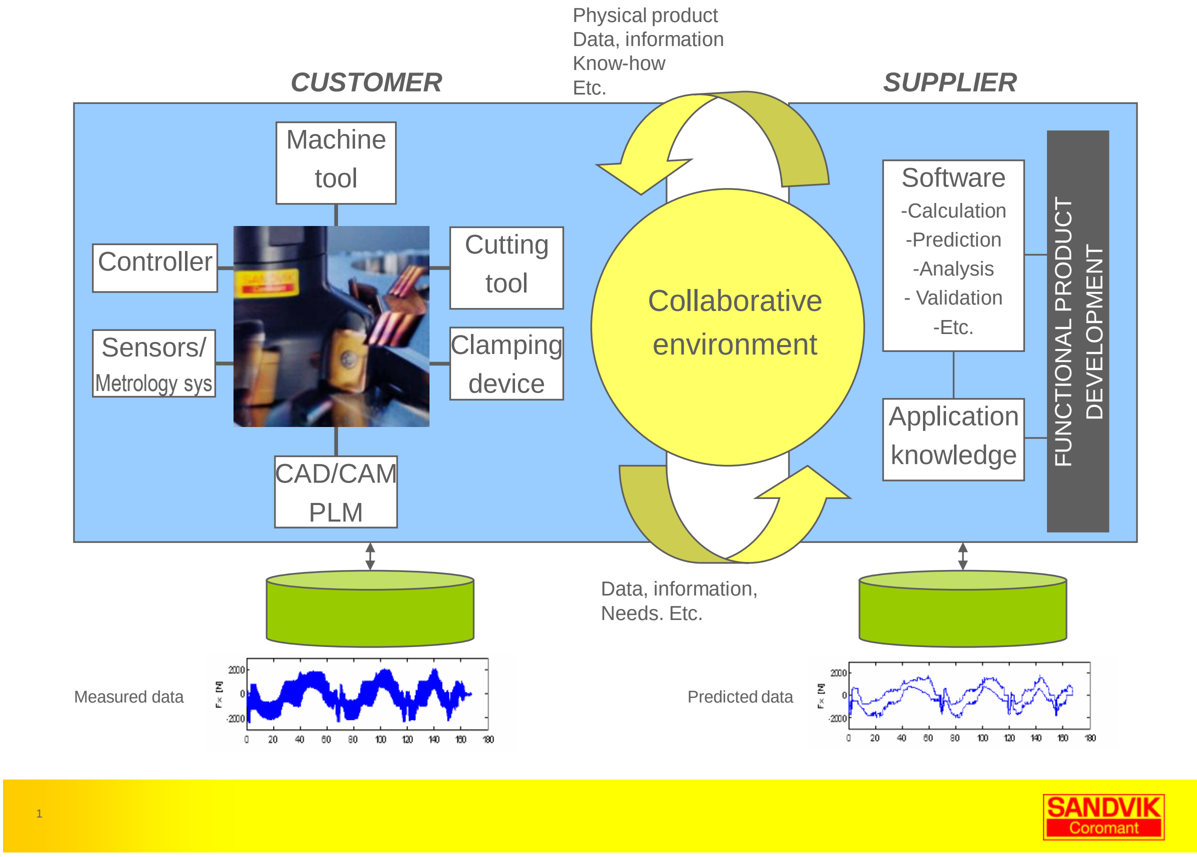


Picture 24: **Hägglunds Drives' vision of future functional product architecture (Copyright - Hägglunds Drives)**

As Hägglunds Drives is interested to learn more and develop capabilities within: availability, reliability, decision making, functional products, optimized lifecycle cost, customer value, data stream management etc., the strategy is to focus on how to optimize the drive system and its life-cycle cost. Thus, the technology and service development will be tighter tied to the business activities as more software and services are required when Hägglunds Drives approaches selling a function with an agreed upon level of availability.

4.1.4 Improvement of tool quality, reliability and machine availability due to function oriented design methodology based on real-time measurement feedback and heuristic model improvement methods (Sandvik Coromant)

Sandvik Coromant has a vision to provide efficient functional product offer through a short term loop and a long term loop. The long term loop enables a closer cooperation between product/service developers and customers, capturing unmet customer needs and requirements for future product upgrades. The short term loop enables possibility to help the customers to optimize the usage of products, and to find the mix of products and services that satisfies the customer needs and increases the capability to implement green sustainable machining. It has been identified that suitable combination of cutting tools, process parameters, machine tools and cutting strategies will support efficient manufacturing of parts leading to less power consumption and a green sustainable machining operation. Further on, Sandvik Coromant sees a potential effect also in an increased capability to follow up products and services used and identification of new user needs.



Picture 25: Sandvik Coromant's vision of future customer-supplier relationship model (Copyright - Sandvik Coromant)

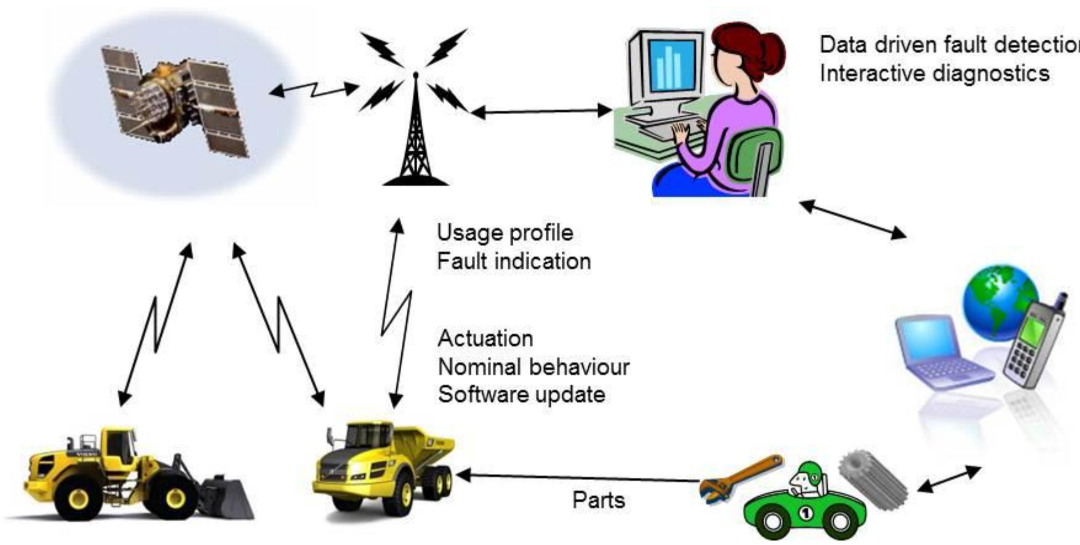
To help Sandvik Coromant understanding the whole picture of customer requirements, a large scale analysis approach needs to be applied as Sandvik Coromant has customers using the products in various ways in many countries all over the world. In order to be able to do large scale analyses, a tool for analysis of product data streams generated by products used at customer sites is needed. Further, identification of meta-knowledge is needed to be able to understand the data streams. By doing this, Sandvik Coromant will enable re-linkage of product data and knowledge generated by customers to product management and the development teams.
The use case which can be generated out of these findings is generic, since especially the machinery and automotive industry is moving towards functional engineering capabilities.

4.1.5 Increase of construction machine fleet availability combined with cost reduction by real-time validation of maintenance requirements, conformity of usage conditions, energy consumption and abrasive wear to allow entering in service oriented business models. (Volvo CE)

Volvo CE has a vision where the construction machinery in operation at customer sites are monitored in more detail than today, giving the possibility to perform improved support for product planning, design, customer support agreement validation, maintenance prediction, data driven fault detection, automated fault diagnosis calibration, interactive remote fault diagnosis and optimization of site logistics and operator driving behavior.
Commonly to all use cases that have been identified at Volvo is the global distribution of locations where their construction machine fleets are used. Connecting places with low mobile phone and satellite coverage represents an impressive high cost factor to transfer the data streams. This necessitates an analysis of how much computing power and storage should be assigned for the

machines and how this should be traded towards the communication cost and availability of communication means. Volvo CE will work on a global view of this problem to address how markets such as the African continent shall be dealt with. At the African continent Volvo CE will develop low costs solutions and have low or infrequent access to the machines. This is a lot harder in comparison to markets where more expensive machines are used and where there is better access and lower communication costs to send large amounts of data.

Especially use cases that take this situation in account create a challenge for the consortium and help to generate solutions that can be applied in similar difficult scenarios.



Picture 26: Volvo CE's vision of future customer-user machine monitoring solutions (Copyright – Volvo CE)

As a first step in realizing this, Volvo CE has recently begun to equip the Volvo CareTrack and Volvo Matris information systems together with every heavy production machine. These information systems enable monitoring by use of onboard telematics equipment. These systems are connected to the vehicle via the CAN-bus, and are able to acquire data from all sensors onboard the vehicle. In a previous set up used by Volvo CE, some data sets were transmitted from the vehicle using the telematics system while other data sets were stored on-board the vehicle electronic control units. The stored data sets were downloaded to the service database when the vehicle was serviced. Due to current limitations in bandwidth and storage capacity, the major part of the information (i.e. the data sets) available is not recorded today.

4.1.6 Higher work quality due to improved user education, training and collaborative communication based on traceable interaction between humans and machinery elaborated with measurement of operational parameters

The goal of SmartVortex is to provide a technological infrastructure consisting of a comprehensive suite of interoperable tools, services and methods for the intelligent management and the analysis of massive data streams produced in all phases of the product and design life cycle to achieve better collaboration and decision making in large-scale collaborative projects concerning industrial innovation engineering. This technological infrastructure can create different kind of business values for the entire industry. To identify the sustainable business values of SmartVortex for the entire industry, different usage scenarios have been analyzed with the industrial partners during the sub-projects ISP-1, ISP-2 and ISP-3. The identified scenarios have been focused on the use of data streams in all phases of the product and design life cycle as well as on different kind of collaboration; such as cross-organization, engineer-to-engineer or engineer-to-customer collaboration. Together with the industrial partners, a set of possible usage scenarios has been selected that present for the industrial partners a high value potential. The next sub-section will describe these usage scenarios with regard to the industrial partners, who will use these scenarios to verify the resulting technological infrastructure of SmartVortex.

5 ANNEX 5: SUCCESS MEASURES AND KPI

5.1 Success measures to be applied for the work package tasks

5.1.1 WP1 Key performance indicators

Title of KPI	Survey of users in Europe and SMARTVORTEX Usage Scenarios and Cooperation Support
Defined	Workshops with partnering user groups evaluation user scenarios, benefits, values and related usage scenarios
Measured	Min. one workshop with each partner separate and min. 1 common workshop with all partners
Target	All workshops performed prior D1.1 delivery

Title of KPI	Industrial Survey about data stream applications
Defined	Cross-check about generic applicability of the selected data stream use cases
Measured	Percentage (%)
Target	100 %

Title of KPI	SmartVortex Requirements Analysis Report and Specification of the SmartVortex Assessment Framework
Defined	Definition of the use case scenarios in ISP1-3
Measured	30.September 2011
Target	On time

Title of KPI	Requirements for real-time data streaming and computation from actual sensor at Electro-hydraulic drive systems (Hägglunds Drives)
Defined	Existence of A) use case description B) definition and description of test data, meta data, validation and resulting actions C) Test data access D) Test bed definition E) Test bed commitment by Hägglunds Drives
Measured	Percentage (%)
Target	100% completion of definition of all sub-points

Title of KPI	Requirements for real-time data streaming and computation from actual sensor at milling tools (Sandvik Coromant)
Defined	Existence of F) use case description G) definition and description of test data, meta data, validation and resulting actions H) Test data access I) Test bed definition J) Test bed commitment by Sandvik Coromant
Measured	Percentage (%)
Target	100% completion of definition of all sub-points

Title of KPI	Requirements for real-time data streaming and computation from actual sensor in construction equipment (Volvo Construction Equipment)
Defined	Existence of

	A) use case description B) definition and description of test data, meta data, validation and resulting actions C) Test data access D) Test bed definition E) Test bed commitment by Volvo CE
Measured	Percentage (%)
Target	100% completion of definition of all sub-points

5.1.2 WP2 Suite Modelling and data & system architecture

Title of KPI	Architecture
Defined	Architectural diagram established and agreed by consortium partners
Measured	Approval by RTDC
Target	Done

Title of KPI	Suite modeling
Defined	Suite modeling and architecture stable
Measured	Scientific or application oriented paper
Target	Paper about the architecture issued until Q2 2011

5.1.3 WP3 Semantic data stream models and access language

Title of KPI	Query language
Defined	Survey over the possible query languages and determination of those
Measured	Number of determined Query languages
Target	4

Title of KPI	Functionality of query language
Defined	Missing query functions for the SmartVortex use-cases identified
Measured	Paper or report about additional functionality
Target	Paper issued during 2011

5.1.4 WP4 Enriched Product Data Processing

Title of KPI	Semantic query handling implementation description
Defined	Scientific description about implementation of RDF-based infrastructure, scalable processing of queries (searching and combining), use of other data formats and relationship tracing
Measured	Papers issued and published until Q1 2012
Target	On time

Title of KPI	Semantic query handling implementation
Defined	Implementation of an RDF-based infrastructure, scalable processing of queries (searching and combining), use of other data formats and relationship tracing
Measured	Prototype test in labor environment
Target	Ready for use with ISP-1 implementation

5.1.5 WP5 Scalable Product Data Stream Management

Title of KPI	Design of DSMS tools
Defined	Design of DSMS tools enabling scalable search in and analysis of high volume data streams
Measured	Paper
Target	Paper in 2011 published

Title of KPI	Use of cluster and cloud resources
Defined	How efficiently can cluster and cloud based resources be used for scalable search in and analysis of high volume data streams
Measured	Published papers
Target	2 until Q1 2012

Title of KPI	Scalable methods
Defined	Develop scalable methods for classification/identification of behavior and on-line monitoring (regarding deviations from expected behavior) for products-in-use
Measured	Papers and code availability for the ISP demonstrators
Target	Min 2 papers and code availability along work plan SmartVortex

5.1.6 WP6 Navigation Visualization and consumption

Title of KPI	Effectiveness of usage and visualization
Defined	Accuracy and completeness with which users achieve specified goals 1. percentage of goals achieved (e.g., number of queries successfully built/intended number of queries); 2. percentage of tasks completed successfully on first attempt (e.g., number of queries successfully built on first attempt); 3. percentage of users successfully completing a given task (e.g., define a set of queries); 4. average accuracy of completed tasks (e.g., evaluating produced queries wrt intended queries); 5. percentage of errors performed by users or reported by the system (e.g., mean number of mistakes/problems for building a visual query); 6. mean number of assistance operations required per task
Measured	% Judgment of users and observers documented with rating list
Target	Point 1-4 maximize; Point 5 and 6 minimize

Title of KPI	Effectiveness related to resource consumption
Defined	Resources expended in relation to the accuracy and completeness with which users achieve goals; it can be measured in terms of: • time to complete a task (e.g., mean time to build a visual query); • tasks completed per time unit (e.g., number of queries built per time unit); • relative time spent by users compared with time spent by an expert user to perform a task (e.g., to build a query);
Measured	Duration (time)
Target	minimize

Title of KPI	Satisfaction of users
Defined	Freedom from discomfort, and positive attitudes towards the use of the visual and multi-modal environment measured through questionnaires, including:

	• rating scale for satisfaction of provided features and interaction modalities • rating scale for ease of use and comprehensibility of the environment • rating scale for ease of learning (learnability) • rating scale for provided support facilities • rating scale for attractiveness of the graphical environment
Measured	% Rating by users using a questionnaire
Target	maximize

Title of KPI	Conformity to existent standards
Defined	ISO9241-11 http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf
Measured	%
Target	100%

While these metrics and associated measurements can give a fairly complete assessment of the performance of the overall user experience, we will only employ subsets suitable for the selected use-cases and user-base.

5.1.7 WP7 Advanced user Information interactions

Title of KPI	Multi-modal user interface quality and usability
Defined	• Responsiveness: ability of the environment to perform (near)real-time interaction modalities recognition • User adaptability and feedback: ability of the environment to adapt the interaction modality to users' needs and requirements, as well as to provide feedback about the correctness of the provided input (voice, gestures, etc.) • Learnability: learning rate and ability of the user to perform and remember available actions provided by the environment, to be weighted wrt complexity of the task, user experience and user cognitive and technical skills • Accuracy: ability of the environment to detect, track and recognize user inputs, actions and commands • Intuitiveness: ability of the environment to provide a clear cognitive association between available commands and the functions or actions they correspond to • Cognitive and physical load: mental and physical load required to the user to perform intended tasks (wrt to user abilities and skills) • Re-configurability: effort and time required to train the environment to support different types of users and interaction modalities
Measured	Time spent
Target	Minimize

5.1.8 WP8 Semantic social collaboration suite

Title of KPI	Collaboration Group Size (CGS)
Defined	Measuring the number of participants that are needed to execute a collaboration process successfully
Measured	
Unit:	Number
Target	Minimize

Title of KPI	Expert Search (ES)
Defined	Support of expert search for maintenance
Measured	
Unit:	Number
Target	Minimize

Title of KPI	Expert Group Formation (EGF)
Defined	Support of expert group formation within the Social Collaboration Suite
Measured	
Unit:	Number
Target	Minimize

5.1.9 WP9 Collaboration and decision making methods & tools

Title of KPI	Mean Time to Incident Solution (MTIS)
Defined	Average amount of time between the occurrence of an incident (a failure / intended failure) and its resolution.
Measured	Average (Time between registration and resolution of an incident)
Unit:	Time
Target	Minimize

Title of KPI	Mean Time to Analyse Causes (MTAC)
Defined	Average amount of time between the registration of an incident and the discovery of the causes.
Measured	Average (Time between registration of an incident and the detection of the causes)
Unit:	Time
Target	Minimize

Title of KPI	Mean Time to Group Composition (MTGC)
Defined	Average amount of time between the registration of an incident and the formation of a collaboration group.
Measured	Average (Time between registration of an incident and the formation of a group \ login of all members)
Unit:	Time
Target	Minimize

Title of KPI	Collaboration Mean Time to Analyse Causes (CMTAC)
Defined	Average amount of time between the start of a collaboration process for analysis an incident and the discovery of causes for an incident.
Measured	Average (Time between starting a collaboration process and the detection of the causes)
Unit:	Time
Target	Minimize

Title of KPI	Collaboration Mean Time to Decision (CMTD)
Defined	Average amount of time between the start of a decision process and the generation of decisions (decision for action / decision for requirements).
Measured	Average (Time between starting a decision process and making decisions)
Unit:	Time
Target	Minimize

Title of KPI	Collaboration Participant Satisfaction (CPS)
Defined	Measuring the satisfaction value of a participant for a given collaboration process
Measured	Survey
Unit:	Percentage
Target	Maximize

Title of KPI	Mean Time to Access (MTTA)
Defined	Average amount of time between the registration of a need for data and the access to this data.
Measured	Average (Time between registration of a need for data and the access to this data)
Unit:	Time
Target	Minimize

Title of KPI	Collaboration Participant Commitment (CPS)
Defined	Measuring the commitment of a participant for a collaboration process result
Measured	Survey
Unit:	Percentage
Target	Maximize

Title of KPI	Technology Support Number (TSN)
Defined	Measuring the operation systems that support the collaboration tool
Measured	
Unit:	Number
Target	Maximize

Title of KPI	Generated Idea Number (GIN)
Defined	Measuring the number of ideas that are generated during a workshop
Measured	
Unit:	Number
Target	Maximize

Title of KPI	Feasible Idea Quote (FIQ)
Defined	Measuring the number of ideas that are selected during a reduce stage of the workshop
Measured	# Feasible Ideas / # Selected Ideas
Unit:	Number
Target	Maximize

Title of KPI	Outcome Clarify Number (OCN)
Defined	Measuring the understanding of the outcome of a collaboration process
Measured	Survey
Unit:	Number
Target	Maximize

Title of KPI	Pattern of Collaboration Number (PCN)
Defined	Measuring the number of patterns of collaboration that can be identified out of the collaboration data stream
Measured	Expert Analysis
Unit:	Number
Target	Maximize

5.1.10 WP10 Semantic collaboration analysis and intelligent process intervention

Title of KPI	Pattern of Collaboration Number (PCN)
Defined	Measuring the number of patterns of collaboration that can be identified out of the collaboration data stream
Measured	Expert Analysis
Unit:	Number
Target	Maximize

Title of KPI	Mean Time to Access (MTTA)
Defined	Average amount of time between the registration of a need for data and the access to this data.
Measured	Average (Time between registration of a need for data and the access to this data)
Unit:	Time
Target	Minimize

5.1.11 WP11 Rules and policy framework

Title of KPI	Management of access / Usage
Defined	Management of access, usage rules, and policies to be able to correctly share data streams and artifacts in terms of information security and intellectual property rights (IPR)
Measured	% of fulfillment described by master thesis (defended during 2011) and two issued papers
Target	100%

Title of KPI	Establishment of best practices for building secure cross-organizational collaboration environments
Defined	Create clear understanding on various levels and non-disclosure/disclosure at different stages on collaborating partners documented in papers. Certified by the industrial partners. The process description determines the key factors for IPR protection and how typical IPR problems can be solved and implemented by the authorized users
Measured	% Written acceptance by all (4) industrial partners
Target	100%

5.1.12 WP12 Cross-organizational IPR management

Title of KPI	IPR and access protection during collaboration processes within either different legal entities or different cost/responsibility centers within one company
Defined	Entering the given access rules either from the paper issued in WP11 and rules or from the participating members of the collaboration.
Measured	Demonstration how to enter or change min 2 different rules. Check of authorization
Target	100% performed

5.1.13 WP13 Smart data stream enabled decision making and cooperation support in complex design and engineering environments

The described use case in ISP-1 use case the KPIs requires acceptance of the DOW change and addition of Fe-Design by the project officer. Due to the short time available until the committed delivery date of this document and the identification of the perspective partner deeper technical discussions apart of the use-case scenario and content have not been possible and are set up for the first week of October 2011.The KPIs will be delivered immediately after those meetings given the fact of a preliminary (not legally binding) positive message about the ability to include Fe-Design GmbH as consortium partner. The KPIs are planned in the direction of performance increase and overall time reduction for the FEM simulation process. This might need also to adapt the KPIs in ISP-1.

5.1.14 WP14 Efficient information lifecycle and data stream management of industrial data for Functional Products business models

Title of KPI	Set-up defined and prepared
Defined	Successful set-up of ISP-2 use cases 1-3 (i.e. Sandvik Coromant, Hägglunds Drives and Volvo CE)
Measured	D14.1, D14.2 test beds accessible and functional
Target	Prepared to start at M13

Title of KPI	Ready for execution
Defined	Successful execution of use-cases defined in ISP-2
Measured	Prototype and demonstrator existent and working in two stages M36 and M46
Target	M36 limited functionality / M46 full functionality

Title of KPI	Prototype functionality and applicability in ISP-2
Defined	Publishing of application oriented paper for each use case I ISP-2
Measured	Number of papers that have been issued
Target	1 each use case

5.1.15 WP15 Product data stream sharing and collaboration in complex cross-organizational scenarios

Title of KPI	Set-up defined and prepared
Defined	Successful set-up of ISP-3 use case concerning cross-domain and cross-organizational collaboration
Measured	D15.1, D15.2 test beds accessible and functional
Target	Prepared to start at M13

Title of KPI	Ready for execution
Defined	Successful execution of use-case defined in ISP-3
Measured	Prototype and demonstrator existent and working
Target	M36 limited functionality / M46 full functionality

Title of KPI	Prototype functionality and applicability in ISP-3
Defined	Publishing of application oriented paper for use case ISP-3
Measured	Number of papers that have been issued
Target	2

5.2 Success measures to be applied in the integrated subprojects

Success measures in the integrated sub-projects can be shown during the demonstration of the prototyping software. In the related use cases situation can be generated that show the reaction in form of a defined action or a visualization alert during the prototype demonstration. For those success criteria oriented KPIs where the demonstration effort during a review does not justify the real time presentation, movies or recordings of the specific measure result will be presented. The value profiting industrial partner will certify that those features have been successfully implemented and will support that with experience reports if necessary.

5.2.1 ISP-1 Simulation of design data

Title of KPI	Calculation of trends in simulation results related to stress, vibration or sound
Defined	Frequency oriented trend calculation in complex stress, vibration or sound simulations can consume and operate real-time related to the simulator output (computing power adapted)
Measured	Comparison output and trend generation
Target	No Difference

Title of KPI	Early recognition of result trends in simulation results related to stress
Defined	Frequency oriented trend calculation in complex stress, sound or vibration simulations should be reduced by trend calculation
Measured	Time spent to recognize stabile vibration or sound trends frequency by frequency
Target	Minimize

Title of KPI	Prediction of simulation results related to stress, sound or vibration
Defined	Prediction calculation out of trend analysis on relation to expected overshoot of given thresholds in stress, sound or vibration, stop of current simulation run; adaptive visualization of prediction and restart of the simulation with new parameters if overshoot is predicted.
Measured	Trend and prediction calculation
Target	Real-time calculation with direct visualization equal to simulation process progress

Title of KPI	Improvement of the complete simulation process
Defined	Time comparison between traditional simulation process and new methodology
Measured	%
Target	20 % maximize

Title of KPI	Demonstrators for entire project

Defined	• stakeholder identification
	• identify data and information needs related to stakeholders
	• identification of data to be extracted from large data streams
	• selection and prioritization of the data
	• Processing of massive data streams
	  o quality assurance of data and models
	  o identification of meta-knowledge
	  o specification and implementation of software modules for data acquisition and processing
	  o implementation of a method for data management when scaling up from one machine tool to a large number of machine tools

Measured	Percentage (%)
Target	100% completion of definition for all sub-points

5.2.2 ISP-2.1 Real time data streaming in Electro-hydraulic drive systems applications

Title of KPI	Data collection tool
Defined	Hardware and software equipment for local collection, storage and transfer of measured drive data from installations in test beds
Measured	75%

Target	Tool for collection and transfer of measured data

Title of KPI	Data base
Defined	Setup of a data base where collected data from the data collection tool can be stored.
Measured	70%
Target	Data base for storage of collected test bed data

Title of KPI	User data visualization environment
Defined	Find a graphical environment for visualization of data
Measured	5%
Target	A user-friendly interface where users can access selected data in a graphical format

Title of KPI	Report function
Defined	Tool for creation of status reports for internal use and for information to end users
Measured	10%
Target	Reports that can be scheduled for automatic creation and sent via mail to specified users

Title of KPI	Failure relations
Defined	Find important signal relations and verify in test bed
Measured	15%
Target	Find important relations between measured signals for failure modes

Title of KPI	Statistics of reported problems
Defined	Collect and calculate statistic values from failure reports. Will be used to identify critical components.
Measured	25%
Target	Statistics from failures

5.2.3 ISP-2.2 Real time data streaming in milling machine operations related to the used tools

Title of KPI	Temperature alert
Defined	Tool is used with illegal feed motion
Measured	Temperature threshold violated
Target	Temperature alert is created

Title of KPI	Customer satisfaction (Sandvik Coromant)
Defined	Capability to • monitor customer's machining process • give recommendations with respect to improvements of the customer's machining process
Measured	% Survey at Sandvik R&D departments and ranking in low, medium, high
Target	> 60% high

Title of KPI	Scalability
Defined	Upscale to multiple machines connected to the DSDM should show identical performance
Measured	Time delay on operation execution
Target	No delays

5.2.4 ISP-2.3 Real time data streaming in construction equipment fleets

Title of KPI	Overload detection
Defined	Wheel loader bucked carries to high load
Measured	Comparison with wheel loader usage allowance
Target	Monitoring system creates alert and show current load

Title of KPI	Gear shift average
Defined	Show the average of the 10 latest gear shifts F1-F2 for the time data of control valve downstream pressure starting at gear switch and ending 1 s after clutch speed difference is zero
Measured	%
Target	100%

Title of KPI	Gear shift average with FFT filtered time data
Defined	Show the average of the 10 latest gear shifts F1-F2 for the frequency range of 1-3 Hz for FFT filtered time data of longitudinal accelerometer starting at gear switch and ending 1 s after clutch speed difference is zero
Measured	%
Target	100%

Title of KPI	Highest deviation
Defined	Show the group of 10% machines of a given model that deviates the “most” from the others in terms of all the FFT filtered longitudinal acceleration data streams
Measured	%
Target	Depending on the available machines to control (will be defined prior related review date)

Title of KPI	Quality of the remote diagnostic set-up
Defined	Jury members shall regard the remote diagnostics setup as good enough in comparison to the real situation of traveling to the machine that has failed in some sense in terms of judging how to repair the machine.
Measured	Judgments of current service personal
Target	7 out of 10 judgments

Title of KPI	Uptime improvement
Defined	The uptime improvement estimation is computed using the saved nr of travel days as well as reduced time delay in ordering and shipping components needed for repair for the most common failures and summed over the chosen machine fleet configuration
Measured	%
Target	Uptime should reach > xx% (confidential figure will be told to the reviewers by Volvo)

Title of KPI	Prevention of failures
Defined	Another contribution to improved uptime is the prevention of failures through finding machines that are in some sense in a risk zone and then performing remote diagnosis to find corrective actions. The risk zone could be defined as the machine with a critical signal deviating the most from the machines being used in similar fashion, for example with the same type of operating cycles. This is difficult to measure and is thus only added to the improvement estimation if the project finds distinct cases where it can be shown that a failure can be avoided using the tools of the Smart Vortex project
Measured	Number of cases

Target	> 1

5.2.5 ISP-3 Data Streams within collaborative environments

Title of KPI	Shared Vocabulary
Defined	A general social ontology(GSO) is selected and extended with terms required for engineering oriented collaboration
Measured	Delivery date
Target	Selection done

Title of KPI	Trust levels
Defined	The rule engine allows to access collaboration data in a different way depending on the trust level
Measured	Trust levels can be defined and selected during ISP3 demo
Target	3 Trust levels are implemented in prototype