

QualiMaster

A configurable real-time Data Processing Infrastructure
mastering autonomous Quality Adaptation

Grant Agreement No. 619525

Deliverable D1.2

Work-package	WP1: Use Cases
Deliverable	D1.2: Full Use Cases and Requirements
Deliverable Leader	SPRING
Quality Assessor	Mohammad Alrifai
Estimation of PM spent	7
Dissemination level	Public (PU)
Delivery date in Annex I	31.12.2014
Actual delivery date	31.12.2014
Revisions	5
Status	Final
Keywords:	Use cases, actors, application scenarios, visualization

Disclaimer

This document contains material, which is under copyright of individual or several QualiMaster consortium parties, and no copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the QualiMaster consortium as a whole, nor individual parties of the QualiMaster consortium warrant that the information contained in this document is suitable for use, nor that the use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

© 2014 Participants in the QualiMaster Project

List of Authors

Partner Acronym	Authors
MAX	Steve Hutt, Maciej Kurek
LUH	Mohammad Alrifai, Claudia Niederée
SUH	Holger Eichelberger
SPRING	Holger Arndt, Stefan Burkard, Elke Pelikan
TSI	Apostolos Dollas, Ekaterini Ioannou

Table of Contents

Table of Contents	4
Executive summary.....	7
1 Introduction	8
1.1 Requirements Collection Approach.....	8
1.2 Application Scenarios Overview.....	9
1.3 Components of the QualiMaster Infrastructure	9
1.4 Terminology.....	11
1.5 Changes with respect to D1.1	12
1.6 Structure of the Deliverable.....	13
2 Key Actors of the QualiMaster Infrastructure	13
2.1 Application Users.....	13
2.2 Infrastructure Users	14
2.3 Component Providers	15
3 QualiMaster User Survey	17
3.1 Implementation of survey	17
3.2 Summary of Results and Inputs from Experts	18
3.2.1 Responses to Questionnaire.....	18
3.2.2 Multiple Responses grouped by User Profile	21
3.2.3 Feedback received via telephone calls	23
3.3 Major Findings and Conclusions	24
4 Market Player Financial Volatility.....	24
4.1 Concept of Volatility	24
4.2 Calculation of Volatility.....	25
4.2.1 The VIX Calculation: Step-by-Step.....	26
4.2.2 Existing risk strategies: Portfolio hedging with Volatility-Options	27
4.2.3 Introduction of Seasonality.....	28
4.3 Expected Results in QualiMaster – Description and Mockups.....	29
4.3.1 Volatility as an indicator	30
4.3.2 Future prognosis of Volatility.....	31
4.3.3 Interpretation of Volatility with respect to Systemic Risk.....	31
5 Application Domain & Business Use Cases	32
5.1 QualiMaster Applications	32
5.2 Application Use Cases for Systemic Risk Assessment for Institutional Financial Clients...	33
5.2.1 The Hedge Fund Industry	33
5.2.2 Use Cases for Institutional Financial Clients	34
5.2.3 Application Use Cases for Systemic Risk Assessment for Regulatory Bodies	40
5.3 Visualization.....	44

6	Data, Algorithm and Quality Requirements	49
6.1	Data and Data Stream Requirements	50
6.1.1	Types of Data Sources	50
6.1.2	Real-time vs. Static Sources	50
6.1.3	Data Load	51
6.1.4	Requirements	51
6.2	Requirements for Supported Algorithms	52
6.2.1	Correlation	52
6.2.2	Mutual Information	53
6.2.3	Transfer Entropy	54
6.2.4	Algorithm flexibility	55
6.2.5	Requirements	55
6.3	Platform Quality Requirements	55
6.3.1	Quality Dimensions	55
6.3.2	Requirements	57
7	System Requirements and Use Cases	58
7.1	Pipeline Designer	58
7.1.1	Use Case: Define New Pipeline	58
7.1.2	Use Case: Modify Pipeline Definition	61
7.1.3	Use Case: Delete Pipeline Definition	63
7.2	Adaptation Manager	64
7.2.1	Use Case: Define Quality Characteristics of Processing Elements	65
7.2.2	Use Case: Define Pipeline Quality Characteristics	66
7.2.3	Use Case: Define Reactive Adaptation Rules	68
7.2.4	Use Case: Define Proactive Adaptation Rules	69
7.2.5	Use Case: Monitor Execution of Adaptation Rules	71
7.2.6	Use Case: Change adaptation settings	72
7.3	Platform Administrator	73
7.3.1	Use Case: Define Platform Quality Parameters	74
7.3.2	Use Case: Modify Platform Quality Parameters	75
7.3.3	Use Case: Add Data Processing Algorithm	76
7.3.4	Use Case: Modify Data Processing Algorithm	78
7.3.5	Use Case: Add Hardware-based Data Processing Algorithm	79
7.3.6	Use Case: Modify Hardware-based Data Processing Algorithm	81
7.3.7	Use Case: Configure Pipeline Sources and Sinks	82
7.3.8	Use Case: Start Pipeline	84
7.3.9	Use Case: Stop Pipeline	85
7.3.10	Use Case: Configure QualiMaster Platform for Software-based Execution	87

7.3.11	Use Case: Configure QualiMaster Platform for Hardware-based Execution.....	89
7.3.12	Use Case: Start QualiMaster Platform	91
7.3.13	Use Case: Stop QualiMaster Platform	91
7.3.14	Use Case: Instantiate Platform	92
7.3.15	Use Case: Monitor Execution.....	93
7.4	Component Providers	94
7.4.1	Algorithm Provider	94
7.4.2	Data Provider.....	94
7.4.3	Application Provider	94
8	Application Testing.....	94
8.1	Preparation of Data Sets.....	94
8.1.1	Financial Data Sets.....	94
8.1.2	Social Web Data Sets	95
8.2	Application Test Examples.....	96
9	Conclusions and Outlook	99
	References	100

Executive summary

Collecting requirements about a system scheduled for realization helps defining the terminology, stabilizing the common vision and detailing the functional and quality requirements. This deliverable reports on the results of the requirements collection for the QualiMaster project, and in particular the QualiMaster Applications for systemic risk analysis in the financial domain and the underlying QualiMaster infrastructure. We present and discuss the actors who will interact with the applications and the QualiMaster infrastructure and, in particular, the descriptions of individual use cases, i.e., their specific interactions with the QualiMaster components. Furthermore, we provide requirements for the data streams to be processed by the QualiMaster infrastructure and the algorithms to be applied in a data analysis pipeline for systemic risk calculation. This is a consolidated deliverable which integrates part of the initial requirements collected and described in D1.1 and the refinement of the requirements collected in D1.2. It supersedes D1.1.

1 Introduction

This deliverable summarizes the requirements that have been collected in the QualiMaster project for the two targeted QualiMaster application scenarios and the underlying QualiMaster infrastructure. This document has to be considered as a consolidated deliverable which integrates part of the initial requirements collected and described in D1.1 and the refinement of the requirements collected in D1.2. Thus, it supersedes D1.1.

For collecting the requirements, a use-case based approach has been chosen, which enables an intuitive and user-centered starting point for better understanding and discussing the system functionality. As a first step in describing the use cases, a set of stakeholders or key actors have been identified, which interact with the system in different ways and in different roles.

In addition to describing the use cases of the two QualiMaster applications from the financial domain, we also compiled a set of system use cases for covering the core functionalities and actors of the QualiMaster infrastructure. This addresses the configuration of processing pipelines, the management of the adaptation (the flexible adaptation support is one of the special features of the QualiMaster infrastructure) and the characteristic aspects of the set up and administration of the overall QualiMaster infrastructure.

The description of the stakeholders and the use cases is complemented by the identification of requirements towards the data and the description of relevant quality requirements for the system. They also provide important input for WP2 and WP4, which are both concerned with quality aspects of the processing pipelines.

This collection and documentation process of the core use cases has served as a good trigger for discussions about the functionalities, terminologies, and dependencies within the consortium.

1.1 Requirements Collection Approach

Use cases are a popular means for collecting requirements in a user-oriented way. Starting from a set of actors, i.e., a set of persons or other systems that interact with the system under consideration, use cases describe the flow of interaction of those actors with the system. The advantages of use cases are that they are very intuitive and easy to understand due to their textual form. Furthermore, they do not only support the description of the normal flow of interaction (the so called use case scenario), which helps in the identification of required system functionalities. They also foster the description of exceptional cases, which already gives a broader picture of the expected system functionality.

One of the disadvantages of use cases is that they are restricted to functional requirements. Therefore a separate part has been added to this deliverable, which documents non-functional requirements especially with respect to the data and the quality requirements, which have been identified during the discussion of the use cases.

For the documentation of the use cases tables have been used, which are a simplified form of the table-based templates suggested by Cockburn [5] for this purpose. Each table contains

- the use case name and an unique identifier,
- the involved actors,
- the goal of the use case,
- the preconditions for the use case and the postconditions that are established by successful use case execution,
- the scenario description (interaction steps) for successful execution of the use case (typically the actor and the system alternate in their interaction),
- the description of exceptional cases in the interactions and interaction variants,

- the business constraints for the use case (business rules) and
- the processed data or employed system functionality.

The use case identifier is assigned to the use cases for easing cross-referencing between use cases. In particular, referenced use cases in preconditions are intended to be transitive across all references so that indirectly referenced use cases do not need to be listed explicitly.

1.2 Application Scenarios Overview

The QualiMaster project will validate its results in terms of two application scenarios on systemic risk analysis, one focusing on institutional financial clients and one on regulatory bodies. The use cases for the application scenarios will be discussed in more detail in Section 4.

Systemic risk had been identified as a key factor in the stability of financial markets. In its broadest form, it represents scenarios where financial markets, and the exposure of financial institutions to those markets, become strongly correlated or coupled, potentially leading to industry-wide institutional failure. The need to measure and identify predictive signals of systemic risk is one of the most challenging issues facing institutional users and market regulators today. This is particularly so given the velocity of financial markets and the need to rapidly identify and act on hotspots before contagion sets in. It is well understood that today's financial markets are correlated to a degree significantly greater than historically and not simply within asset classes. Due to the Risk-On-Risk-Off nature of today's trading, broad ranges of asset classes have become strongly correlated, so that diversification of risk is more challenging. The financial data analysis will be complemented with taking into account Social Web data for supporting and stabilizing the prediction of systemic risk, which we expect can be used for identifying additional indicators, and for contextualizing risk predictions.

The second application scenario targets **institutional financial clients**, such as Hedge funds, Banks or Asset Managers. In this scenario, we aim at pre-trading risk analysis and real-time, real money trading risk analysis. The implementing applications will be integrated with the trading applications of SPRING and enhance them by a multi-variant and multi-market risk analysis approach, which is expected to add important insights about systemic risks, and help avoiding fatal losses of capital under management. As a collateral outcome, this will help stabilizing the capital markets at the roots, i.e., within the trading system of the financial industry.

1.3 Components of the QualiMaster Infrastructure

QualiMaster aims at a configurable infrastructure for real-time data stream processing, which adapts itself to the actual needs and the runtime requirements imposed by actual data streams. In this section, we discuss the basic components of the QualiMaster infrastructure in order to introduce the background. Please note that this section does not aim at introducing the overall system architecture of the QualiMaster project, as the architecture is currently under development and is described in Deliverable D5.2.

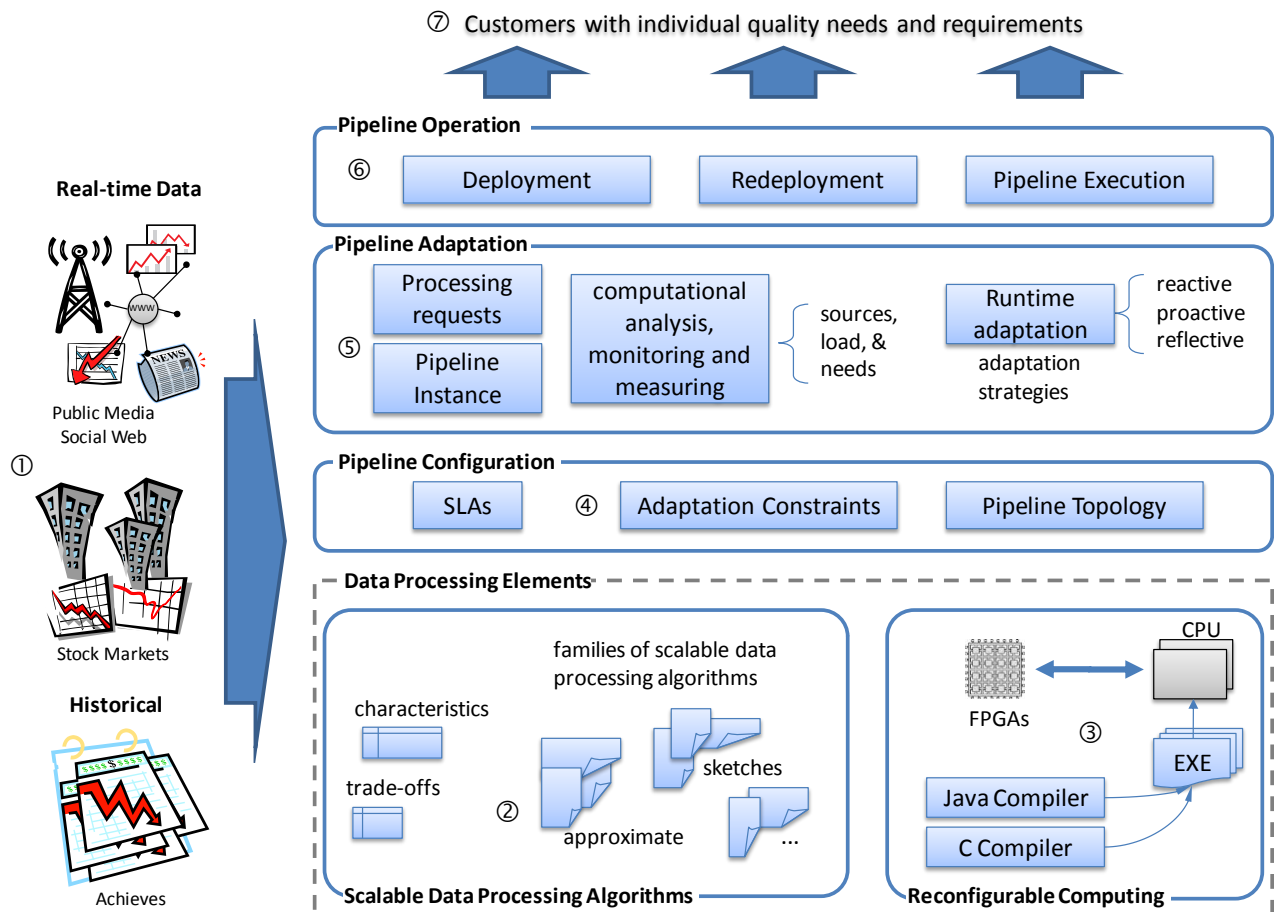


Figure 1: Overview of the QualiMaster Infrastructure

Figure 1 depicts an overview of the QualiMaster infrastructure as also envisioned in the Description of Work (DoW). Basically, this figure was communicated at the Kick-off meeting to all participants and the authors of this document had this figure in mind while collecting and describing the QualiMaster use cases.

The **QualiMaster infrastructure** consists of an environment for adaptively executing data stream pipelines as well as related tools for configuring and managing that environment. We will call the execution environment the **QualiMaster platform**. The term **configuration** refers to the activity of setting up the platform itself and the data analysis to be executed on the platform. The configuration is performed by an expert human being and a proper configuration is a prerequisite for successful and efficient data analysis. Basically, the notion of configuration originates from Software Product Line Engineering (SPLE) [17, 23], a successful approach for systematic software reuse. By applying SPLE techniques, we aim at an efficient and consistent configuration (of a generic “template”) of the QualiMaster platform in order to save time, effort and computational resources. In contrast, **adaptive execution** refers to autonomous activities carried out by the QualiMaster platform in order to maintain the actual quality of the data analysis and the efficiency of the use of the physical computing resources. In turn, the adaptive execution relies on the configuration that implicitly defines the boundaries and the validity of the autonomous activities.

Various **data sources** such as stock market data, public social media as well as collected historical archives will serve as the input for the data analysis ①. The actual data analysis will be performed by **data processing algorithms** ②, such as the identification of causality in multivariate times series. One core idea of making the data analysis in QualiMaster adaptable is the notion of **algorithm families**. An algorithm family is a group of algorithms performing the same analysis step, while the individual algorithms differ in their actual execution quality, i.e., requiring different amount of memory or producing more accurate results. Let us consider three algorithms *A*, *B* and

C for the identification of causality. As the algorithms perform the same task, they belong to the processing family for causality identification. However, algorithm *A* produces highly precise results, but requires vast amounts of memory. In contrast, algorithm *B* is imprecise within acceptable boundaries, while it consumes significantly less memory. Switching between these algorithms at runtime allows reacting on changing conditions in the data streams and the execution environment. However, in the described case, algorithm *B* will impact the quality of the result produced. In particular, the notion of data processing families applies naturally to algorithms, which are realized for specialized **reconfigurable hardware** ③, e.g., data flow computing hardware developed by the partner MAX. For example, algorithm *C* is precise, memory and execution time efficient, but requires a certain amount of specialized hardware, i.e., specifically configured FPGA processors (which of course must be present and available as they are an efficient but expensive and, thus, limited resource).

To perform a certain data analysis, data processing families are combined to a **data processing pipeline**. The data processing pipeline is at the heart of the **configuration** ④ of a QualiMaster platform, also specifying the execution hardware, the algorithm families, quality constraints for Service Level Agreements (SLA) and the adaptive behaviour. A data processing pipeline consists of sources, sinks, data processing elements and data flows [4] connecting data **sources** ① with **processing elements**, processing elements among each other and processing elements with data **sinks** ②. In QualiMaster, a data processing element of a data processing pipeline is realized through a specific algorithm family, thus enabling variants of the data processing pipeline, i.e., adaptive execution of data processing pipelines ⑤. One challenge for the adaptive execution is the selection (and modification) of the most appropriate algorithm within each processing element / family at runtime, i.e., to determine the actual pipeline instance for execution. This is supported by the analysis of the overall (end-to-end) quality of the pipeline and the impact on the data and processing quality introduced by the (combination of) variable data processing algorithms. In our example above, selecting algorithm *B* might save resources, e.g., in high load situations, but may also imply a reduced quality of the analysis results of the subsequent and, thus, entire pipeline. Finally, pipeline execution and adaptation is supported by the (low-level) QualiMaster platform in terms **pipeline operations**, ⑥ such as starting or stopping a pipeline.

As indicated above, adaptivity also needs knowledge about the underlying execution platform. To illustrate, if the data stream comes from Germany and there is a MAX dataflow system in Greece, it is entirely possible that some pre-processing will be done in Germany and the adaptive pipeline will have its next stage in Greece, but it is unlikely that the data can go back and forth a lot due to the communications overhead. However, if both the data and all resources are available in the same location, the platform may include multiple accesses of some resource in the adaptive pipeline.

1.4 Terminology

In this section, we introduce some further terms we will use throughout this deliverable.

- **Platform instantiation** is the process of turning the configuration into an executable and optimized (version of the implementation of the) QualiMaster platform. Akin to the term “configuration”, also the term “instantiation” (also called “product derivation”) originates from SPLE [17, 23]. Basically, the QualiMaster platform will be realized as generic but possibly not (fully) configured piece of software, which may include more functionality than actually required for executing a certain set of pipelines, e.g., measurement and monitoring mechanisms for a wider set of qualities. Based on the configuration, the process of instantiating the platform will turn the generic QualiMaster platform into a specific instance, e.g., adding, disabling or removing unused monitoring mechanisms. Further, it will take care of the appropriate integration of hardware algorithms and the hardware execution, including the choice for different strategies of realizing a dataflow as indicated in the geographically distributed example above.

- A **quality parameter** [20] is a measurable and quantifiable property of a computational element (also other terms are used in literature, e.g., *quality dimension* or *quality attribute* depending on the community [10]). A computational element may be a data processing algorithm, a data flow, a data analysis pipeline or a physical compute resource. Examples for quality parameters are numbers of tuples per time unit (data flow or data pipeline level), execution time or memory usage (compute resource level). We will distinguish between primitive quality parameters directly measured by the infrastructure or the hardware (as the aforementioned examples) and derived quality parameters defined by the platform administrator (such as a domain-specific kind of throughput). Primitive quality parameters will be built into the QualiMaster platform to be measured. However, quality parameters that remain unused in the configuration can be disabled in or removed from the running QualiMaster platform instance during platform instantiation.
- **Quality characteristics** describe how a set of quality parameters behave for a computational element over time in a certain setting, e.g., a financial data processing algorithm under high load. Typically, the quality parameters cannot be considered constant, so that mathematical or statistical means must be applied to capture quality characteristics appropriately. Quality characteristics are defined for individual algorithms, propagate to algorithm families (through the selection of an algorithm) and, ultimately, to the end-to-end quality characteristics of a data processing pipeline.
- A **quality constraint** is a logical expression involving and restricting quality parameters in order to define the validity of the actual execution, e.g., that the throughput of a certain data processing element (implying the underlying selected data processing algorithm) shall not be below a given number. Quality constraints will formalize quality (also known as non-functional) requirements collected, e.g., for a specific pipeline or an organization running the QualiMaster infrastructure. Specific categories of quality constraints deal with the cost of execution, the adaptation or the pipeline itself. On pipeline level, quality constraints will in particular define the SLAs of a data processing pipeline for both, source (pipeline input) and sink (pipeline output) side, i.e., the SLAs negotiated with the customer. Quality constraints bound the adaptation space, i.e., the violation of a quality constraints must be avoided (although they may have to be tolerable for a short period of time until the result of an adaptation can be enacted) and may be used as triggers for reflective adaptation or in order to indicate exceptional situations.

1.5 Changes with respect to D1.1

This document is a consolidated deliverable which integrates part of the initial requirements collected and described in D1.1 and the refinement of the requirements collected in D1.2. It supersedes D1.1. The main changes done with respect to D1.1 are the following:

- A new section (section 3) has been dedicated to the survey done among SPRING's financial customers in order to refine the initial requirements stated in D1.1.
- One of the major conclusions drawn from the survey is that Volatility should be introduced into the QualiMaster platform. Volatility has been considered as a new requirement of the QualiMaster platform and a whole section has been dedicated to its description (section 4).
- From the survey it also emerged that sophisticated visualization need to be introduced in order to facilitate the interpretation of the results. Section 5.3 is dedicated to visualization.
- In D1.1, three separate configuration tools had been conceived, one for each Infrastructure User, in order to describe their individual use cases. In D1.2 the three tools are integrated into one single QualiMaster Infrastructure Configuration Tool (QM-IC tool). This change is reported in section 7.
- The group of roles defined in D1.1 has been extended with another group, namely Component Providers. Component Providers deliver interfaces or algorithms on purpose, and, thus, support the Infrastructure Users with data or implementation related assets. A description on Component Providers has been added in section 2.3.

- Algorithm flexibility has been added as requirement in section 6.2.4.
- The use case “Change adaptation settings” for the Adaptation Manager has been added in section 7.2.6.
- In WP1 there has already been initial activity for the application testing. This activity was not reported in D1.1 and has been added in section 8 of D1.2.

1.6 Structure of the Deliverable

The rest of this deliverable is structured into nine sections. Section 2 introduces stakeholders or key actors interacting with the QualiMaster application and the different parts of the QualiMaster infrastructure. Section 3 reports on the user survey executed for the refinement of the requirements and use cases. Section 4 describes the Volatility, which has been introduced as measurement of the QualiMaster platform based on the results of the user survey. Section 5 is dedicated to the two QualiMaster application scenarios and describes the respective use cases. Section 6 provides an initial collection of non-functional requirements with respect to data, algorithms derived from the application scenarios and the QualiMaster platform. Section 7 contains the system use cases for the QualiMaster infrastructure in terms of three subsections, each focussing on one of the key actors for the infrastructure: the Pipeline Designer, the Adaptation Manager, and the Platform Administrator, respectively. Section 8 is dedicated to the testing from the application perspective. Finally, Section 9 presents some conclusions and outlines the next steps.

2 Key Actors of the QualiMaster Infrastructure

An actor represents a group of users, who interact with a system and who have a similar view on a system. Actors may be persons, but also companies, organizations or even computer systems [5]. Actually, actors are stakeholders of the system, but not all (groups of) stakeholders are required to interact with the system. Further, a real person may take the role of different actors, e.g., depending on the organization structure of a company running the QualiMaster infrastructure. In this section, we will define the key actors of the QualiMaster infrastructure in terms of two distinct groups, namely application users in Section 2.1 and infrastructure users in Section 2.2. In Section 3 and 5, we will describe the use cases according to these groups of actors, respectively.

2.1 Application Users

These are the end users that interact with the financial applications to be built on top of the QualiMaster Infrastructure. They access the system with a task at hand (e.g. risk analysis of a certain market player) and use the QualiMaster applications to perform this task.

Application users do not need to know how the underlying QualiMaster platform is configured or developed. The respective financial application should, however, support some flexibility regarding the data analysis such as the selection of the market players to analyze, the time span to be considered in the analysis, etc. This functionality should also be supported through the graphical user interface of the application.

The most important actors for QualiMaster in the group of the Application users are:

- Hedge Fund Manager
- Investment Company
- Investment Bank
- Regulator

We will detail these actors in the table below:

Application Users	
Actor	Hedge Fund Manager
	A Hedge Fund Manager oversees and makes decisions about the investments in a hedge fund. To be successful, a hedge fund manager must consider how to gain a competitive advantage, a clearly defined investment strategy, adequate capitalization, a marketing and sales plan and a risk management strategy. QualiMaster strives to provide the Hedge Fund Manager with a tool to achieve this competitive advantage mainly in terms of systemic risk management. This can be used for portfolio optimization or risk management and hedging.
Actor	Investment Company
	An Investment Company is a corporation or trust engaged in the business of investing the pooled capital of investors in financial securities. This is most often done either through a closed-end mutual fund or an open-end mutual fund. The open-end fund must be willing to buy back shares from investors every business day. Exchange-traded funds (or "ETFs" for short) are open-end funds or unit investment trusts that trade on an exchange. Open-end funds are most common, but exchange-traded funds have been gaining in popularity. Closed-end funds generally issue shares to the public only once, when they are created through an initial public offering. Their shares are then listed for trading on a stock exchange. Investors who do no longer wish to invest in the fund cannot sell their shares back to the fund (as they can with an open-end fund). Instead, they must sell their shares to another investor in the market.
Actor	Investment Bank
	Management of enterprise wide risk across a wide range of asset types has become a major regulatory requirement in addition to being a prerequisite for effective capital allocation. The need to include a network of potential exposure outside of the investment bank has recently become recognised, in the sense that internal liquidity is no longer a sufficient indicator of financial stress. By providing an Investment Bank with a tool that identifies in real time the co-dependencies and avenues of contagion between major market participants, they will be better able to manage such exposures.
Actor	Regulator
	Historically, Regulators have been retrospective in their analysis of major systemic risk events. Increasingly, Regulators have access to real time exchange data and are looking to leverage this data to provide more timely, and ultimately proactive management of the financial system. The QualiMaster project will provide Regulators with a unique opportunity to view in real time a systemic risk network identify the sources and sinks of risk and view contagion through network topology.

2.2 Infrastructure Users

In contrast to application users, infrastructure users directly interact with the platform in order to define data analysis pipelines, the adaptation space of individual pipelines or to administer a platform (including its initial setup). To perform their tasks, Infrastructure Users utilize specific tools and, thus, have specific requirements towards the QualiMaster infrastructure.

The following types of infrastructure users have so far been identified in the QualiMaster project:

- Pipeline Designer
- Adaptation Manager
- Platform Administrator

We will detail these actors in the table below:

Infrastructure Users	
Actor	Pipeline Designer
	The Pipeline Designer defines the structure of data processing pipelines for performing specific analysis tasks. In particular, a Pipeline Designer identifies the data sources, the data sinks, the data processing elements (families) to be used in a pipeline and the data flow among the processing elements. The task of the pipeline designer may also include the selection of adequate visualizations for the pipeline processing results.
Actor	Adaptation Manager
	The Adaptation Manager defines and specifies the adaptive behaviour of the system. This includes defining the quality characteristics of the different data processing elements, the methods for measuring them and defining methods for estimating the end-to-end quality of pipelines. In addition, the Adaptation Manager also has to define a set of rules on the pipeline level for reactive and proactive adaptation as well as prediction mechanisms for quality parameter for proactive adaptation. Furthermore, the Adaptation Manager monitors and analyzes the execution of adaptation rules and reflects the results of the analysis by adjusting these rules when required to further optimize the adaptations (reflective adaptation).
Actor	Platform Administrator
	The Platform Administrator sets up, installs and maintains the QualiMaster infrastructure. This includes the administration of the physical computing resources, the algorithm and algorithm families pool, the reconfigurable hardware units (such as Data Flow Engine boards) as well as the storage of data. In addition, the Platform Administrator is also in charge of monitoring the pipeline operation and of starting and stopping pipelines and, thus, taking the responsibility of the physical compute resources.

2.3 Component Providers

Due to clarification of the configuration and integration approach in QualiMaster, we extend the actual groups of roles with another group, namely Component Providers. Component Providers deliver interfaces or algorithms on demand, and, thus, support the Infrastructure Users with data or implementation related assets. Actually, Infrastructure Users may also take the role of Providers if appropriate. In particular, Component Providers can be external to the organization of the Infrastructure Users, e.g., being employed by third parties.

The most important actors for the Providers group are:

- Algorithm Provider
- Data Provider

We detail these actors in the table below.

Component Providers	
Actor	Algorithm Provider
	<p>The Algorithm Provider delivers algorithms for QualiMaster pipelines. Algorithm Provider and Platform Administrator communicate in terms of (family) interfaces, i.e., the Algorithm Provider delivers algorithms complying with the communicated (family) interfaces so that the Platform Administrator can easily configure them for inclusion into his/her QualiMaster infrastructure. Although not an algorithm, an Algorithm Provider may also deliver the implementation of data source access compound needed to translate the data of a data source so that the QualiMaster infrastructure can handle it.</p>
Actor	Data provider
	<p>The Data Provider owns (to some extend) a data set and agrees to deliver it to a running QualiMaster infrastructure, i.e., to the (organization of) the Infrastructure Users. Therefore, the Data Provider announces a data access interface, an access address and, if required, negotiates the licensing and the pricing of the data.</p>
Actor	Application Provider
	<p>The Algorithm Provider delivers data analysis applications based on a QualiMaster infrastructure to Application Users. Therefore, the Application Provider may discuss the design of new pipelines with the Application Designer. In particular, the Application Provider designs and develops new user applications on top of the QualiMaster platform, i.e., based on technical data sink interface (e.g., realized as a communication protocol), the Application Provider develops Application in any programming language, visualizes the resulting data provided by the QualiMaster platform and enables the user, if appropriate, to issue user triggers via an external event interface that may lead to an adaptation in the QualiMaster platform. Furthermore, the Application Provider may provide to the Application User insights into the calculation performed by the respective pipelines as well (as far as supported) their actual quality parameters.</p>

3 QualiMaster User Survey

In order to refine the QualiMaster use cases and requirements, a survey has been prepared in the scope of WP1. The aim of the survey has been to collect inputs from the potential industrial users of the QualiMaster platform. In fact, the questionnaire has been prepared by the partner SPRING who works with customers in the financial online trading industry. Section 3.1 describes the implementation of the survey, listing the questions and possible answers. Section 3.2 summarizes the results and main inputs from experts. Section 3.3 summarizes the main findings and conclusions.

3.1 Implementation of survey

The questionnaire has been prepared in two iterations. In the first iteration, the first version of the questionnaire has been prepared. It consisted of 12 questions. After a discussion with the other partners of the consortium, it has been decided to shorten the questionnaire in order to higher the chances of getting valuable impact from the questionnaire recipients. Therefore, the questionnaire has gone through a second iteration of preparation and the second and final version has been prepared. It consists of the following six questions:

1. **Which answer describes you best?** This question has been inserted at the beginning of the survey in order to know the profile of the professional responding to the questionnaire. The possible answers are the following:
 - I manage institutional funds
 - I am an institutional analyst without direct trading access
 - I am a financial infrastructure provider (hardware or software)
 - I am a private trader
 - I am a scientist
 - Other (please specify)
2. **In which timeframe do you usually trade?** The aim of this question was to identify the timeframe used by the professional during his trading activities. The possible answers are the following:
 - Monthly
 - Weekly
 - Daily
 - Intraday
 - High Frequency
3. **Which risk analysis software do you use?** This is an open question and aims to get inputs on the state-of-the-art platforms used by the responders for their trading activities.
4. **What are the main functionalities that a risk analysis platform should have?** This question aims to get inputs on the requirements of the overall QualiMaster platform. The main functionalities selected by the potential customers will be considered as priority during the development of the platform. The possible answers to this question are:
 - Give early warning of expected strong market downturns
 - Give early warning of expected strong market volatility (up or downturns)
 - Show the correlation between the different assets in my portfolio
 - Show the correlation between my portfolio and the overall market
 - Give concrete buy/sell signals.
5. **How important are the following visualization features of a risk platform?** This aims to collect user requirements on the visualization system which will be integrated in the QualiMaster platform. The possible answers are:

- 2D charts
- 3D charts
- Risk metrics as quotes
- Risk metrics as complex visual analytics charts
- Detailed insight to mathematical calculations
- Printable fact sheet
- Other (please specify)

6. The QualiMaster project needs your help. This question aims to create a contact base around the QualiMaster consortium and possibly to recruit potential customers interested to test the QualiMaster prototype. The possible answers are:

- Please keep me informed
- I would like to help the project with my scientific and financial industry related input
- I would be interested to test the QualiMaster prototype as a beta tester

The questionnaire has been sent out to by email to several contacts among Spring Techno's customers and contact base. In addition, specific customers have been contacted via telephone in order to have a more direct and concrete feedback.

3.2 Summary of Results and Inputs from Experts

The questionnaire has been sent to 250 people, which were recruited by the business customers of spring, the partners of the business contacts of spring and via target group-specific online-discussion platforms. 29 people have responded by written answers. More than the half of them announced further interest in the QualiMaster development and also in active involvement. Based on these responses, feedback has also been collected by means of interviews made via telephone calls. This section summarizes the received responses to the questionnaire (section 3.2.1) and the feedback received by means of telephone conversations (section 3.2.3).

3.2.1 Responses to Questionnaire

Question 1: Which answer describes you best?

The table below summarizes the response to question 1.

Answer	Responses
I manage institutional funds	3.7%
I am an institutional analyst without direct trading access	7.41%
I am a financial infrastructure provider (hardware or software)	33.3%
I am a private trader	33.3%
I am a scientist	22.2%

The following three comments have been received as response to the open question "Other (please specify)":

- I am setting up an algorithmic hedge fund
- Contract Risk Analyst

Question 2: In which timeframe do you usually trade?

The table below summarizes the response to question 2.

Answer	Responses
Monthly	22.7%
Weekly	22.7%
Daily	27.2%
Intraday	22.7%
High frequency	18.1%

Question 3: Which risk analysis software do you use?

The responses to the open question 3 are summarized below.

- My own custom risk analysis software
- Openquant
- ABW
- Custom made software
- P&L
- None
- FastTrack.net
- Custom made. Realtime risk management with alerts etc
- Excel
- riskPro FirE Sophis
- proprietary (own design)
- Proprietary software
- None
- Matlab
- Prop
- Excel application I created with visual basic
- I generally use my own approach, which is extremely strict.
- Risk Manager

Question 4: What are the main functionalities that a risk analysis platform should have?

The table below summarizes the response to question 4.

	Non important	Less important	Important	Very important
Give early warning of expected strong market downturns.	3.6%	14.3%	28.6%	53.6%
Give early warning of expected strong market volatility (up- or downturns).	3.6%	14.3%	28.6%	53.6%
Show the correlation between the different assets in my portfolio.	7.1%	14.3%	28.6%	50%

Show the correlation between my portfolio and the overall market.	3.6%	7.1%	57.1%	32.1%
Give concrete buy/sell signals.	21.4%	32.1%	17.9%	28.6%

The comments to the open question were the following:

- Show exposure of my portfolio to risk factors show detailed sources of systematic risk show detailed sources of idiosyncratic risk show contribution to risk of positions analyze and propose portfolio hedges attribute past performance and volatility to systematic and idiosyncratic sources forecast future portfolio volatility forecast future portfolio exposure to systematic risk factors
- Valuation using different models stress testing
- I think the main issue missed by risk platforms is the risk of unknown unknowns. Anything is possible. So you've got to be careful that your platform is only used as a tool, not a solution. Underlying data should be able to be analyzed in maximum detail by the user. You need maximum flexibility, especially when responding to emerging risks that had not been previously considered.

Question 5: How important are the following visualization features of a risk platform?

The table below summarizes the response to question 5.

	Non important	Less important	Important	Very important
2D charts	3.8%	11.5%	46.1%	38.5%
3D charts	18.5%	33.3%	37%	11.1%
Risk metrics as quotes	4%	12%	36%	48%
Risk metrics as complex visual analytics charts	7.7%	7.7%	50%	34.6%
Detailed insight to mathematical calculations	12%	32%	40%	16%
Printable fact sheet	14.8%	33.3%	22.2%	29.6%

The comments to the open question "Other (please specify)" are the following:

- Neural relationships between causal factors
- Audit trail precise mapping of real life contracts to contract types in the system wide range of risk factors, consistently managed
- Single-glance dashboard summarizing exposures, performance, other single factor metrics
- The eye processes information more effectively in a graphical format, rather than just numbers
- Correlation between up/down with the released of new information to the Market

3.2.2 Multiple Responses grouped by User Profile

The main questions which have been considered as input to the QualiMaster requirements are questions 4 (What are the main functionalities that a risk analysis platform should have?) and 5 (How important are the following visualization features of a risk platform?). They have been designed as question with multiple choice answer, question 4 having five possible answers while question 5 having six possible answers. In order to understand how the different profiles have addressed these two questions, we have grouped the profiles of the interviewed people into two main groups. The first group includes the following profiles:

- Institutional fund manager
- Institutional analyst without direct trading access
- Financial infrastructure provider
- Private trader

This group has been indicated as “Professional”. The second group includes all the scientists and has therefore been indicated as “Scientists”.

Figure 2 and Figure 3 below indicate the results of the answers of the two groups to the questions 4 and 5. The horizontal axis reports the multiple choices of the two questions while the vertical axis reports the number of the answers. For the sake of simplicity, the answers have been grouped into “Important”, which includes the answers “Important” and “Very important”, and “Non important”, which includes the answers “Less important” and “Non important”.

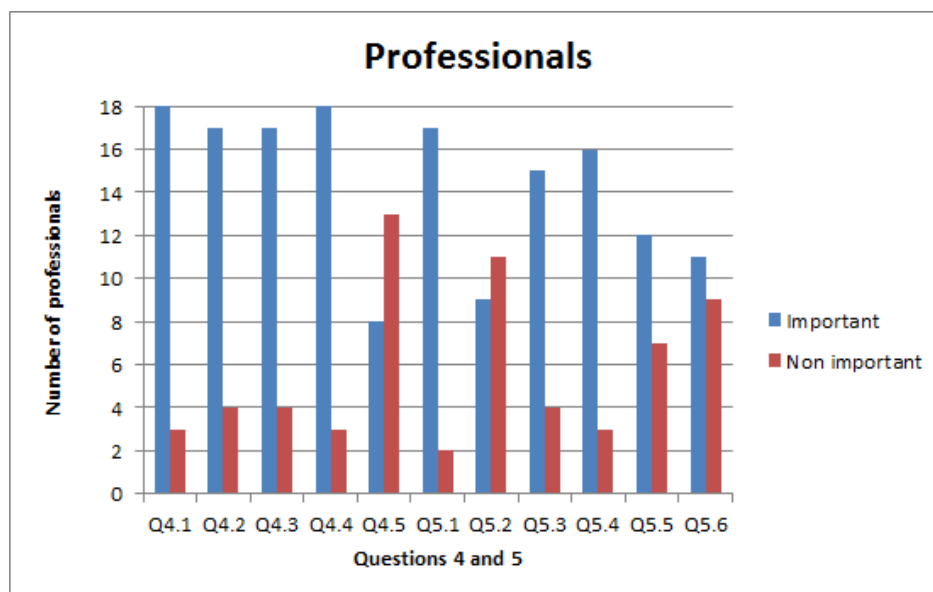


Figure 2: Responses to multiple-choice questions 4 and 5 for the group of “Professionals”.

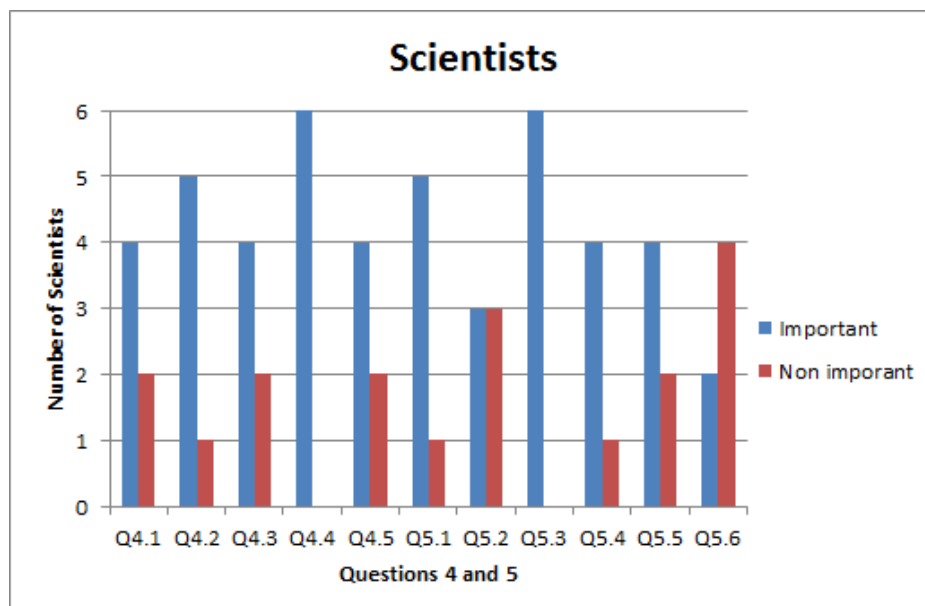


Figure 3: Responses to multiple questions 4 and 5 for the group of “Scientists”.

The answers in

Figure 2 and Figure 3 show that financial industry professionals regard volatility measurements and correlations measurements as important (Questions 4.1 to 4.4). Both suggested types of volatility indication and prognosis, namely giving early warning of strong market downturns and giving early warning of strong overall volatility – including strong market up-turns – are regarded almost equally important by the professionals (Questions 4.1 and 4.2). The same is true for both suggested correlation indications: The correlation between the different assets in the portfolio and the correlation between the portfolio and the overall market.

The majority of professionals does not see concrete buy and sell signals as an important feature of a systemic risk platform (Question 4.5).

The group of scientists – in comparison – also regards volatility and correlation measurements as important features. However, with one third of their votes in the “no important” field, they do not give an equally strong statement for “strong volatility downturns” (Q4.1) and “show correlation between the different assets in my portfolio” (Q4.3) as the industry professionals.

In contrast to the professionals, the group of scientists mostly would like to see concrete buy and sell signals (Q4.5).

Coming to the visualization features (Question 5), both groups find 2D charts important (Q5.1), while 3D charts (Q5.2) are less important (among the scientists group) or mostly not important (among the professionals). “Risk metrics as quotes” (Q5.3) are regarded as important by both groups, while the scientists group gives a very strong vote for this feature.

“Risk metrics as complex visual analytics charts” (Q5.4) is an important feature for both groups, while “Detailed insight to mathematical calculations” (Q5.5) is also regarded as important, but with a more controversial note, as about one third of the votes goes to “non important” with this feature.

Finally a “printable factsheet” (Q5.6) seems to be a nice-to-have feature for the professionals group and not important for the scientists.

In terms of the evaluation of answers, in general, the project will focus more on the statements of the professionals group, as it is the main target group. However, we will carefully listen to the group of scientists to get additional input and suggestions.

3.2.3 Feedback received via telephone calls

Several financial industry professionals have been interviewed by means of telephone calls. One of the interviewed professionals has provided significant feedback on two critical functionalities which the QualiMaster platform needs to address: “Strategy” and “Time Window”. Regarding the “Strategy”, the interviewed professional says that “for evaluating Hedge Funds and their demand for data and risk analysis, it is important to know their Strategy”. The suggested strategies are reported in Figure 4 below.

Market Neutral	Event Driven	Opportunistic
<ul style="list-style-type: none"> ▪ Equity Market Neutral ▪ Fixed Income Arbitrage ▪ Convertible Arbitrage 	<ul style="list-style-type: none"> ▪ Merger Arbitrage ▪ Distressed Securities 	<ul style="list-style-type: none"> ▪ Global Macro ▪ Long-Short Equities ▪ Short Selling ▪ Emerging Markets

Figure 4: Strategy suggested by industry professional via telephone call.

Regarding the “Time Window”, the interviewed professional says the following: “A further important differentiator is the time window. There are Funds that make decisions on monthly basis. The other extreme are High Frequency Trading Funds”. The time windows that he suggests are: monthly, weekly, daily, intraday, high frequency trading.

Another financial industry professional has been asked to give his opinion on how the QualiMaster approaches systemic risk analysis. He says that the project is trying to quantify some MARKET risk measures based on some non-linear measures of dependence, such as the copula functions and trying to relate these measures with the flow of information provided by the news. He suggests that, also based on academic literature, systemic risk should depend on the shock that the observer would like to discuss. Depending on how this shock is defined, there might have a "domino" effect or a sort of "tsunami" that damages the entire financial system/economy. The professional believes that it does not matter how the shock is defined. An institution might need to hold a capital buffer for being able to repay its own liabilities, in case one of these shocks get generated.

In addition to contacting financial industry professionals, the consortium also contacted the leading persons in charge of other projects addressing similar topics to the ones addressed in QualiMaster. The aim of contacting these people was to gain insights and to evaluate possible areas of exchange or cooperation. The first person to be contacted has been a professor from the University of Osnabrück, Germany. The professor has managed a project titled “Social media sentiment and stock market prediction”. The project has already ended, but the contacted person has sent us a document summarizing the main results. One of the partners of this project is Lenz & Partner (www.lp-software.de) a renown financial data provider with industry contacts. This partner has expressed the intention to collaborate and exchange knowledge with the QualiMaster project.

Another project that has been contacted is SSIX, an EU project which has just been approved and which starts in March 2015. The project deals with the creation of sentiment analysis indices out of social networks. The person contacted is Laurentiu Vasiliu, from PERACTON LTD.

3.3 Major Findings and Conclusions

The major conclusions that can be drawn from the survey and that have been considered as inputs to the QualiMaster requirement engineering are:

- Volatility should be introduced in the platform as measurement. This can be appreciated in the response to question 4 in section 3.2.1 and also in section 3.2.2. A discussion on Volatility is reported in section 4.
- Sophisticated visualization need to be introduced in order to facilitate the interpretation of the results. This is reported in the responses to question 5 in section 3.2.1. A discussion on visualization is included in section 5.3.
- Prognosis/forecast should be used in order to take the systemic risk one step further (see response to question 4 in section 3.2.1). A discussion on prognosis is provided in section 4.

4 Market Player Financial Volatility

Following the results of the survey described in Section 3, Volatility will be introduced into the QualiMaster platform as a measurement. In addition to measuring general market Volatility as well as single stock- and single financial instruments Volatility as an indicator, in a further step, a future Volatility forecast and an interpretation of Volatility with respect to Systemic Risk will be performed.

4.1 Concept of Volatility

Among traders and fund managers, Volatility is an important sentiment measure, which reacts to market movements. There are several existing volatility indicators, published by stock exchanges, with the CBOE Volatility Index (VIX) being the most popular one.

The existing volatility indices measure the implied volatility for a basket of put and call options related to a specific index or Exchange Traded Funds (ETFs). The VIX measures the implied volatility for a basket of S&P 500 call and put options. (Call options give their owner the right to buy an underlying – in this case the S&P 500 – for a specified price at a specified future date. Put options certify the right to sell for a specified price at a specified future date.) The VIX is designed as an index, which tends to return to a median level over time. It is specifically designed to measure the expected 30-day volatility for the S&P 500. See chapter “Calculation of Volatility” for more details.

The VIX is not by itself a predictive indicator, but it can identify sentiment extremes. In general, the indicator declines when the stock market gradually rises - and it advances when stocks decline. Quick and sharp stock market declines often produce spikes in the VIX. Spikes above specific VIX levels suggest an extremely bearish sentiment¹.

Volatility as a tradable asset: VIX Futures & Options

The Chicago Board of Options exchange introduced exchange-traded VIX futures and options between the years 2004 and 2006, making the VIX a regulated tradable asset. The success and demand in these exchange products can be seen in their daily volume, as combined trading activity in VIX options and futures has grown to over 800,000 contracts per day.

Different existing Volatility indices

¹ compare: http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:volatility_index

In addition to the VIX Index, CBOE calculates several other indexes, including the volatility of certain commodities, single stocks and foreign currencies. Each Index groups the financial volatility of the major market players of a specific market segment. Most indices are self-explaining through their name, of example the “CBOE U.S. Energy Sector ETF Volatility Index “ is a group of Exchange traded funds, which focus on the energy sector.

Below is a list of example volatility indices, issued by the CBOT²:

- CBOE U.S. Energy Sector ETF Volatility Index (VXXLESM)
- CBOE Emerging Markets ETF Volatility Index (VXEEMSM)
- CBOE EFA ETF Volatility Index (VXEFSM)
- CBOE Gold Miners ETF Volatility Index (VXGDXSM)
- CBOE Silver ETF Volatility Index (VXSLVSM)
- CBOE Brazil ETF Volatility Index (VXEWZSM)
- CBOE China ETF Volatility Index (VXFXISM)
- CBOE Equity VIX® on Apple (VXAPLSM)
- CBOE Equity VIX® on Amazon (VXAZNSM)
- CBOE Equity VIX® on Goldman Sachs (VXGSSM)
- CBOE Equity VIX® on Google (VXGOGSM)
- CBOE Equity VIX® on IBM (VXIBMSM)

4.2 Calculation of Volatility

For the calculation of the VIX, the CBOE uses near-term and next-term put and call options. (Near-term options have at least 1 week left until expiration. Next-term options usually have 1-2 months until expiration.) Each option price carries an implied volatility or [Standard Deviation](#). The CBOE calculates a weighted average of implied volatility to find the expected 30-day (calendar days) volatility for the S&P 500.



Figure 5: Example VIX chart, illustrating the mean reversion tendency of the index. After peaks, the volatility level tends to return to lower levels over time.

² source: CBOE: White Paper – The CBOE Volatility Index VIX, 2014

In short terms, there are four steps involved in the calculation:

- Select the near-term and next-term put and call options.
- Calculate the implied volatility for each option.
- Calculate a weighted average of implied volatility for these options.
- Multiply this weighted average by 100.

As a result of this calculation, the VIX tells us the weighted 30-day standard deviation of annual movement in the S&P 500. Example: A value of 20 would expect a 20% move, up or down, in the next 12 months. This annualized number can be broken down to monthly or daily values. The table below shows VIX levels with the expected volatility in the S&P 500 on a monthly or daily basis³.

CBOE Volatility Index (\$VIX)	Expected Monthly Volatility in S&P 500	Expected Daily Volatility in S&P 500
10%	2.89%	0.63%
12%	3.46%	0.76%
14%	4.04%	0.88%
16%	4.62%	1.01%
18%	5.20%	1.13%
20%	5.77%	1.26%
22%	6.35%	1.39%
24%	6.93%	1.51%
26%	7.51%	1.64%
28%	8.08%	1.76%
30%	8.66%	1.89%
32%	9.24%	2.02%
34%	9.81%	2.14%
36%	10.39%	2.27%
38%	10.97%	2.39%
40%	11.55%	2.52%
42%	12.12%	2.65%
35%	10.10%	2.20%
40%	11.55%	2.52%
45%	12.99%	2.83%
50%	14.43%	3.15%
55%	15.88%	3.46%
60%	17.32%	3.78%
65%	18.76%	4.09%
70%	20.21%	4.41%
75%	21.65%	4.72%
80%	23.09%	5.04%
85%	24.54%	5.35%
90%	25.98%	5.67%
95%	27.42%	5.98%
100%	28.87%	6.30%

4.2.1 The VIX Calculation: Step-by-Step

“The VIX Index is a volatility index comprised of options rather than stocks, with the price of each option reflecting the market’s expectation of future volatility. Like conventional indexes, the VIX calculation employs rules for selecting component options and a formula to calculate index values” [31].

³ Source: http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:volatility_index

The generalized formula used in the VIX calculation is [31]:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2 \quad (1)$$

WHERE...

σ is	$VIX/100 \Rightarrow VIX = \sigma \times 100$
T	Time to expiration
F	Forward index level desired from index option prices
K_o	First strike below the forward index level, F
K_i	Strike price of the ith out-of-the-money option; a call if $K_i > K_o$; and a put if $K_i < K_o$; both put and call if $K_i = K_o$.
ΔK_i	Interval between strike prices - half the difference between the strike on either side of K_i :

$$\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$$

(Note: ΔK for the lowest strike is simply the difference between the lowest strike and the next higher strike. Likewise, ΔK for the highest strike is the difference between the highest strike and the next lower strike.)

R	Risk-free interest rate to expiration
$Q(K_i)$	The midpoint of the bid-ask spread for each option with strike K_i .

In the QualiMaster platform, volatility will be calculated and forecasted for a wide range of market players. Due to these requirements, that are wider than those of the existing volatility indices, the QualiMaster platform will use a set of transparent proprietary formulas, which are applicable to the different asset classes.

4.2.2 Existing risk strategies: Portfolio hedging with Volatility-Options

The negative correlation of volatility to stock market returns makes volatility products a reasonable instrument for the diversification and hedging of a portfolio. Introducing volatility contracts into a trading strategy represents an alternative to exiting a part of the long positions, when the anticipated risk level rises. The negative correlation of the highly volatile VIX to the S&P 500 index makes it possible to use VIX futures or options as a hedge to protect a portfolio against a market crash. This strategy is especially recommended, when the level of the VIX is still low, while a rise in volatility is anticipated.

There are several documented strategies of hedging with VIX options and futures, while simultaneously keeping the costs and possible negative side effects of this kind of portfolio insurance low⁴.

4.2.3 Introduction of Seasonality

Seasonality is a factor, which many traders and fund managers take into account with their asset allocation decisions. The term “seasonality” means that in a defined historic timeframe, certain markets tend to have recurring patterns of up- and downswings during the year. (Yearly seasonality is the most common and most important kind of pattern. More specialized research of seasonality includes e.g. the “four year president cycle”, the “decade cycle” or recurring patterns during the trading day.)

The QualiMaster platform will take the seasonality of volatility into account for its calculations.

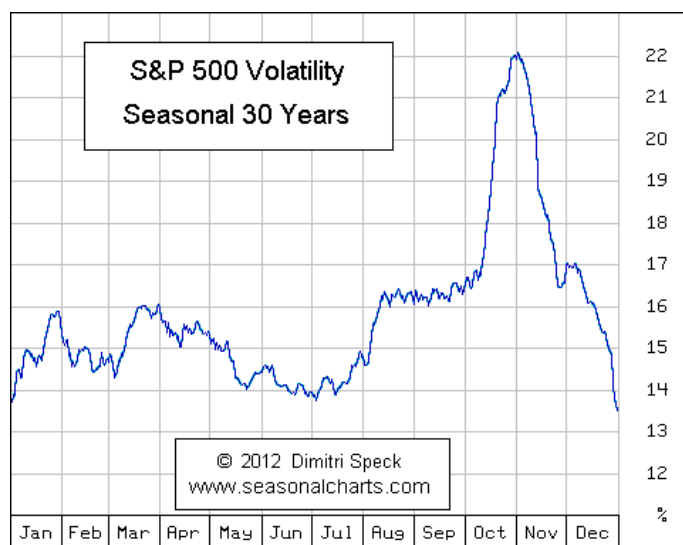


Figure 6: 20-day volatility of S&P 500 Index - Date: 19820406 – 20120314 - Source: <http://www.seasonalcharts.com>

⁴ For hedging examples, compare: <http://www.theoptionsguide.com/portfolio-hedging-using-vix-calls.aspx>



Figure 7: 20-day volatility of DAX (Perf.), until 1988 Dafox Blue Chips - Date: 19820406 – 20120314 - Source: <http://www.seasonalcharts.com>



Figure 8: Volatility spike in October 2014 - marked area. (In this example: Dec14 SP500 VIX Future contract.)

The seasonal volatility charts of the S&P500 (Figure 5) and the DAX (Figure 6) both show peaks between the month October and November. The most recent trading activity before this deliverable confirmed this seasonal assumption: In October 2014 there has been a significant volatility peek in stock indices (Figure 7).

4.3 Expected Results in QualiMaster – Description and Mockups

In the QualiMaster platform, volatility measurements will be shown as a cascade of three main categories of results:

1. Volatility as an indicator
2. Future prognosis of Volatility

3. Interpretation of Volatility with respect to Systemic Risk

From category one to three of the above, the QualiMaster platform gives more and more interpretation of the current risk situation:

While (1) "Volatility as an indicator" adds value by enhancing the concept of volatility to a much bigger selection of market players than it is currently usual in volatility measurement, it leaves a lot of room for the user to interpret the volatility measurements and chartings.

"Future Prognosis of Volatility" (2) – as the name suggests - applies a prognostic approach to evaluating systemic risk through the analysis of volatility. It gives the user a guideline of a projected future volatility for each market player, while still leaving a wide room for own conclusions.

Finally "Interpretation of Volatility with respect to Systemic Risk" describes a feature of the QualiMaster platform, which gives back suggestions and evaluations on how the currently measured and projected volatility could be seen in the light of systemic risk.

4.3.1 Volatility as an indicator

Further than the concept of volatility as an existing index the QualiMaster platform will add value by calculating and charting volatility for a variety of different market players. This will give the user the opportunity to not only evaluate risk on the basis of general market volatility, but also on the basis of volatility of single stocks of assets, that are part of his portfolio or that are considered for buying or for giving a detailed indication.

Volatility as an indicator can be considered as a first level of interpretation. In it, the user needs to draw most conclusions himself. With the concept of future prognosis of volatility, which is explained in the following paragraph, the QualiMaster platform will gain a high level of interpretation and potential guidance of the user. Both concepts may run in parallel on the platform and can be selected on demand by the stakeholder.



Figure 9: Volatility as an indicator (mockup graph): The QualiMaster platform will calculate Volatility for a number of market players, selected by the user. The horizontal blue lines mark an –

adjustable - area, in which volatility is considered as normal und thereby would not indicate an elevated risk level. (source: SPRING)

4.3.2 Future prognosis of Volatility

Based on custom development and proprietary algorithm and software products by SPRING (compare www.vectorbull.com) , there will be a graphical and numerical future prognosis of volatility. This prognosis can be created by the user for a list of stocks of other financial assets.

The algorithms are mainly based on an evaluation of historical data in combination with real-time data, using neural networks, machine learning techniques and fuzzy logic.

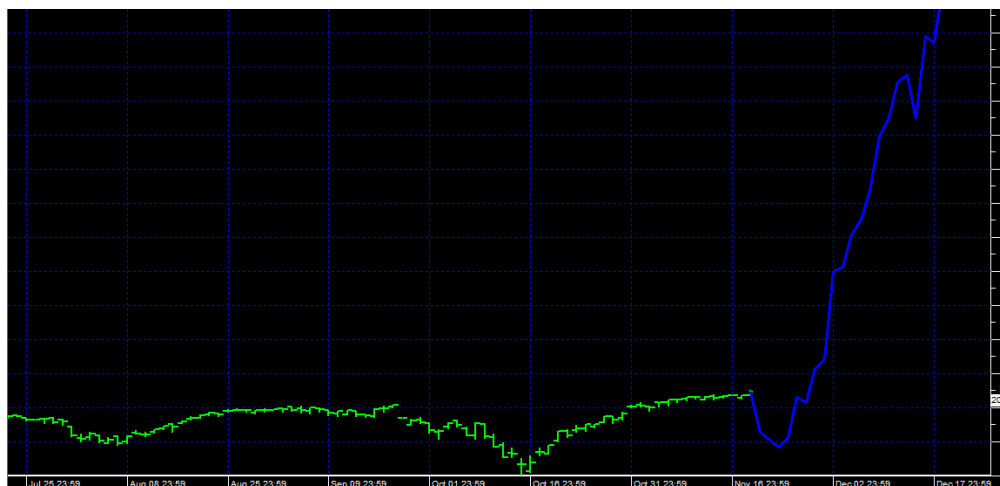


Figure 10: Green bars: daily volatility index bars; blue line: prognosis of future index movement. x-axis: time; y-axis: Volatility index value;

In the above example, the daily volatility fluctuates around a relatively low index range for several weeks (green bars).

In custom settings, the last bar shown in the visualization – and used for calculation – is the last completed bar. This would be “yesterday” in the case of daily bars and e.g. the bar for the last completed half hour in the case of 30-minute bars. The calculation can however go up to the level of milliseconds.

The blue line represents the prognosis of future development of volatility for an index, stock of other financial instrument, which was selected by the user. In the above example, the predicted volatility would first go slightly down and would then sharply increase in value, indicating a volatility peak ahead.

4.3.3 Interpretation of Volatility with respect to Systemic Risk

A step further than the prognosis of volatility goes the direct interpretation of volatility with respect to systemic risk.



Figure 11: Mockup screen of the volatility measurement of a selected market player.

The above graph shows a mockup screen of the volatility measurement of a selected market player. On demand, a user may select several market players to be shown at the same time.

The green bars represent volatility as an indicator (see paragraph above). The dark-blue line represents a future prediction of volatility (see paragraph above). Finally, the colored areas show the interpretation of the current and forecasted future volatility by the QualiMaster platform.

In the above example, a green background stands for a volatility level within the range, which is set as “normal”. Once the prognosis breaks above this range (marked by the upper light-blue line), the background color changes to orange, indicating an elevated volatility- and thus risk level. The background color once more changes to red, when the current price actually breaks above the upper threshold (light-blue line), indicating a high risk level. The background color changes back to orange and then green, once the price - and then the prognosis level - changes back into the “normal” volatility range.

In addition to the interpretation of volatility through color codes, it will be considered to give direct suggestions in a text window such as e.g.: “Based on the current development of volatility, it is suggested to hedge your portfolio.” In addition, it will be considered to suggest different types of hedges to reduce the exposure to risk and systemic risk.

5 Application Domain & Business Use Cases

In this section, we discuss the use cases of the QualiMaster applications. In Section 3.1 we provide an overview on the applications. In Section 3.2, we detail the application use cases for systemic risk assessment for institutional financial clients and in Section 3.3 the application use cases for regulatory bodies. These use cases and the envisioned data analysis pipeline resulted from intensive discussions between the partners MAX and SPRING.

5.1 QualiMaster Applications

Five specific application domains have been identified, three of them belonging to the business domain “Risk assessment for institutional financial clients” and two of them belonging to the domain “Systemic Risk Analysis for Regulatory Bodies“. The five application use cases are illustrated in Figure 2.

Basically, financial data, news and social media data ① will be considered as input streams (details will be given in Section 4). The instantiated QualiMaster infrastructure ② with specific data processing pipelines for the financial applications will process the input streams. The output of the processing will be prepared for the actual application by data analytics and visualization techniques ③. Finally, the five applications ④ will present the analysis results to the financial end users, i.e., the application user actors introduced in Section 2. We will describe the use cases of these five applications in the remainder of this section. As indicated above, the applications can be assigned to the two financial business domains ⑤.

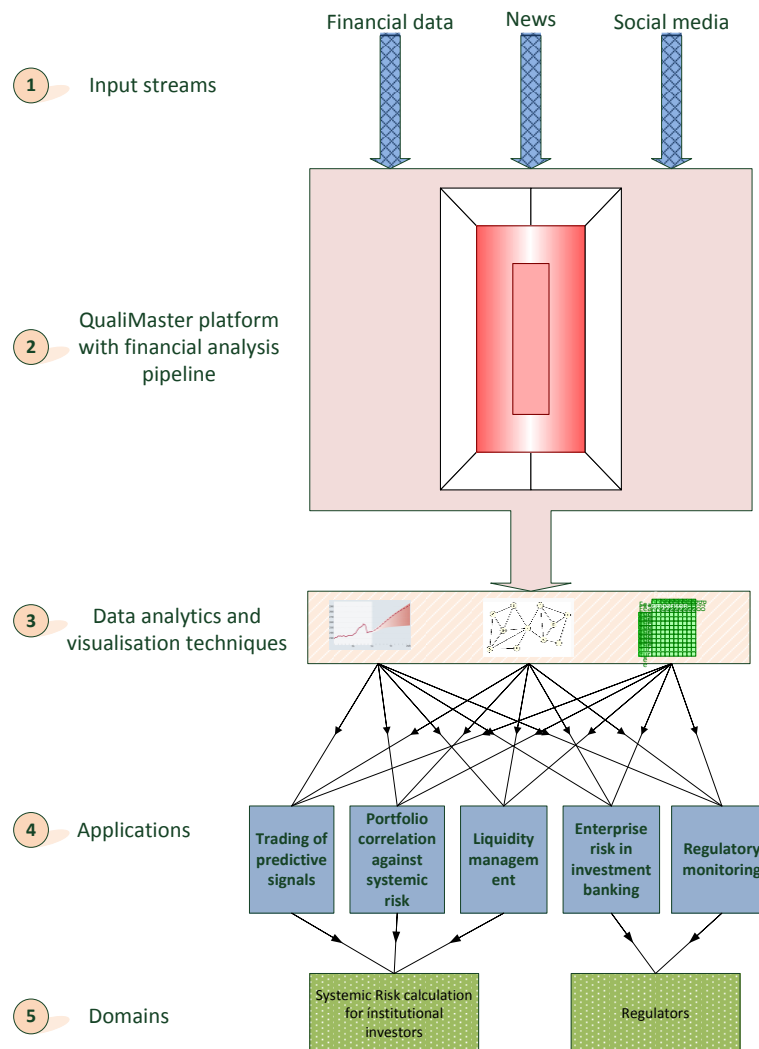


Figure 12: Information flow between QualiMaster infrastructure and applications

5.2 Application Use Cases for Systemic Risk Assessment for Institutional Financial Clients

SPRING is very active in the Hedge Fund industry, and has identified three main application use cases that will benefit from QualiMaster support and advanced risk analysis. In the following paragraphs, we first provide some background on the underlying business domain (Section 5.2.1), namely on the hedge fund industry and describe then the use cases for the three applications in the business domain of systemic risk assessment for institutional financial clients (Section 5.2.3).

5.2.1 The Hedge Fund Industry

A hedge fund is a pooled investment vehicle administered by a professional management firm, and often structured as a limited partnership, limited liability company, or similar vehicle. Many hedge

fund investment strategies aim to achieve a positive return on investment regardless of whether markets are rising or falling ("absolute return"). Some hedge funds have several billion dollars of Assets Under Management (AUM). As of June 2013, the estimated size of the global hedge fund industry was US\$ 2.4 trillion. As of February 2011, 61% of worldwide investment in hedge funds comes from institutional sources. In June 2011, the hedge funds with the greatest AUM was Bridgewater Associates (US\$ 58.9 billion), Man Group (US\$ 39.2 billion), Paulson & Co. (US\$ 35.1 billion), Brevan Howard (US \$31 billion), and Och-Ziff (US\$ 29.4 billion). Bridgewater Associates, had \$70 billion under management as of 1 March 2012.

Hedge funds employ a wide range of trading strategies but classifying them is difficult due to the rapidity with which they change and evolve. However, hedge fund strategies are generally said to fall into four main categories: global macro, directional, event-driven, and relative value (arbitrage). These four categories are distinguished by investment style and each have their own risk and return characteristics.

Because investments in hedge funds can add diversification to investment portfolios, investors may use them as a tool to reduce their overall portfolio risk exposures. Managers of hedge funds use particular trading strategies and instruments with the specific aim of reducing market risks to produce risk-adjusted returns, which are consistent with investors' desired level of risk. Hedge funds ideally produce returns relatively uncorrelated with market indices.

The total capital invested globally in hedge funds increased to a record level for the fourth consecutive quarter in Q2 2013, according to the latest HFR Global Hedge Fund Industry Report [30]. The total hedge fund capital increased by a net total of US\$ 40 bn in 2Q13 to a record US\$ 2.41 trn. The total number of hedge funds increased to over 10,000 funds for the first time since 2006. Positive capital inflows occurred across all fund sizes, with firms below US\$ 500 m in AUM experiencing combined inflows of approximately US\$ 2.4 bn. The industry's largest firms, those in excess of US\$ 5 bn in AUM, experienced net inflows of US\$ 6.1 bn, while firms between US\$ 1 bn and US\$ 5 bn experienced inflows of US\$ 5.8 bn [22].

5.2.2 Use Cases for Institutional Financial Clients

In this section, we describe the use cases for institutional financial clients, namely for

- UC-TOPS1, UC-TOPS2: Application Trading of Predictive Signals
- UC-PCASR1: Application Portfolio Correlation against Systemic Risk

5.2.2.1 UC-LM1, UC-LM2: *Application liquidity managementTrading of Predictive Signals*

In this setting, the actor (Hedge Fund Manager, Investment Company) is looking for new investment opportunities. From his own market analysis, he/she identifies one or more market players that seem to have good trading opportunities. We will call this specific application *Trading Of Predictive Signals Application* (TOPS App). In the first use case (UC-TOPS1), the actor checks co-dependencies against existing portfolio members. In the second use case (UC-TOPS2), the actor checks co-dependencies against all markets.

Use Case Identifier	UC-TOPS1
Use Case Name	Application Trading of predictive signals against existing portfolio
Actor	Hedge Fund Manager, Investment Company
Goal	Assist investment decision makers in selecting market players for new investments, taking into account the systemic risk
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The actor is able to make decisions for modifying his portfolio members

Scenario Sequence	<ol style="list-style-type: none"> 1. Actor starts the QualiMaster TOPS App, enters his login credentials and becomes an application user. 2. Depending on the role of the user, the TOPS App allows the user to select more or less detailed market segments and output visualizations. 3. The user selects 'New market player' and 'existing portfolio member', chooses analysis mode 'Compare players' visualization mode 'Dependency table'. 4. The TOPS App through the QualiMaster infrastructure with running data analysis pipeline tells the user whether the new market player(s) have strong dependencies to the market players in which they have already invested in. This is based on correlation analysis of time series including real-time and historical data. 5. Based on this information, the user decides to invest in the new market player or not.
Extensions	<ol style="list-style-type: none"> 1a Login is not permitted due to invalid credentials and the actor is informed by the TOPS App 3a Based on his role (senior or junior investment manager), the user is able to access/select more intermediate steps of analytics and can select specific pipeline paths for analysis. He can see on which input streams the application Trading of predictive signals is mainly based on. 3b Depth and complexity of data selection, filtering, data representation, and visualisation can be increased and decreased. For example, one additional factor in measuring risk, using predictive signals could be a statistical filtering process to find social media inputs that have proven a good track record in predicting correlation. 3c The user is able to select/de-select each input stream. This may be done by a user-triggered adaptation of the data analysis pipeline. For example, he deselects the impact of social media on the application trading of predictive signals.
Business Rules	
Data/Functions	<ol style="list-style-type: none"> a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations, e.g. comparison and table visualization
Use Case Identifier	UC-TOPS2
Use Case Name	Application trading of predictive signals against all markets
Actor	Hedge Fund Manager, Investment Company
Goal	Assist investment decision makers in selecting market players for new investments, taking into account the systemic risk
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).

Postcondition	The actor is able to make decisions for modifying his portfolio members
Scenario Sequence	<ol style="list-style-type: none"> 1. Actor starts the TOPS App, enters his login credentials and becomes an application user. 2. Depending on the role of the user, the TOPS App allows the user to select more or less detailed market segments and output visualizations. 3. The user selects 'New market player' and 'all markets', chooses analysis mode 'Compare players' visualization mode 'Dependency table'. 4. The TOPS App through the QualiMaster infrastructure with running data analysis pipeline tells the user which dependencies the new market player(s) has against all market segments. The user can now check whether there are dependencies to market segments in which the investor would not like to invest in (for example emerging markets). This is based on correlation analysis of time series including real-time and historical data. 5. Based on this information, the user decides to invest in the new market player or not.
Extensions	<ol style="list-style-type: none"> 1a Login is not permitted due to invalid credentials and the actor is informed by the TOPS App 3a Based on his role (senior or junior investment manager), the user is able to access/select more intermediate steps of analytics and can select specific pipeline paths for analysis. He can see on which input streams the application Trading of predictive signals is mainly based on. 3b Depth and complexity of data selection, filtering, data representation, and visualisation can be increased and decreased. For example, one additional factor in measuring risk, using predictive signals could be a statistical filtering process to find social media inputs that have proven a good track record in predicting correlation. 3c The user is able to select/de-select each input stream. This may be done by a user-triggered adaptation of the data analysis pipeline. For example, he deselects the impact of social media on the application trading of predictive signals.
Business Rules	
Data/Functions	<ol style="list-style-type: none"> a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations, in particular comparison and table visualization

5.2.2.2 Portfolio Correlation against Systemic Risk

The actor (Hedge Fund Manager, Investment Company) wants to check if his already existing portfolio is diversified in terms of systemic risk or not. The given (successful) use case scenario

sequence assumes that the actor wants his portfolio to be as diversified as possible. We will call the application *Portfolio Correlation Against Systemic Risk Application* (PCASR App).

Use Case Identifier	UC-PCASR1
Use Case Name	Application portfolio correlation against systemic risk
Actor	Hedge Fund Manager, Investment Company
Goal	Assist investment decision makers in enhancing the diversification (against systemic risk) of their already existing investments
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The actor is able to make decisions for modifying his portfolio members
Scenario Sequence	<ol style="list-style-type: none"> 1. Actor starts the QualiMaster PCASR App, enters his login credentials and becomes an application user. 2. Depending on the role of the user, the PCASR App enables the user to select more or less detailed market segments and output visualizations. 3. The user selects all market players he/she has in his portfolio. In case a market player of his portfolio is not provided by the underlying QualiMaster data analysis pipeline, he/she selects a market index that represents his market player best. For analysis mode, he selects 'Compare members'. For visualization, he selects 'Cluster representation'. 4. The PCASR App provides a Cluster visualization and a corresponding numerical table. The user can now see how strong the co-dependency from each of his portfolios' market players against each other is. This is based on correlation analysis of time series including real-time and historical data. 5. In case of strong clustering and/or recognized dependency loops, the user can reduce position sizes or even close positions.
Extensions	<ol style="list-style-type: none"> 1a Login is not permitted due to invalid credentials and the actor is informed by the PCASR App 3a Based on his role (senior or junior investment manager), the user is able to access/select more intermediate steps of analytics and can select specific pipeline paths for analysis. He can see on which input streams the portfolio correlation against systemic risk is mainly based on. 3b Depth and complexity of data selection, filtering, data representation, and visualisation can be increased and decreased. For example, one additional factor in measuring portfolio correlation might be a co-dependency module that measures and visualizes market players that are often mentioned together in social media. 3c The user is able to select/de-select each input stream. This may be done by a user-triggered adaptation of the data analysis pipeline. For example, he deselects the impact of social media on the portfolio correlation against systemic risk analysis.

	4a The PCASR App does not report significant co-dependencies. 5a Subsequent to 4a the user does not need to act.
Business Rules	
Data/Functions	a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations, in particular cluster and table visualizations

5.2.2.3 Liquidity Management

The investor (Hedge Fund Manager, Investment Company) wants to check the systemic risk of the whole market and market segments of interest to modify his liquidity management. For example, in the case of high overall systemic risk, the investor will reduce market player position sizes and hold more liquidity. We will call the related application the *Liquidity Management Application* (LM App). In the first use case (UC-LM1), the actor checks general systemic risk on markets while in the second use case (UC-LM2), the actor checks systemic risk on market players he has invested in.

Use Case Identifier	UC-LM1
Use Case Name	General Systemic Risk Assessment for Markets
Actor	Hedge Fund Manager, Investment Company
Goal	Assist investment decision makers in modifying the liquidity of the investment pool with respect to the systemic risk
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The system provides risk information so that the actor is able to make decisions for modifying his portfolio liquidity.
Scenario Sequence	<ol style="list-style-type: none"> 1. Actor starts the QualiMaster LM App, enters his login credentials and becomes a user of the application 2. Depending on the role of the user, the LM App presents selection options for more or less detailed market segments and output visualizations 3. The user selects 'common major markets', chooses analysis mode 'Compare against portfolio' and chooses 'Current systemic risk', 'Tendency' and 'Prognosis' for output. Tendency and prognosis are based on the user selection of the time horizon, e.g., this moment, the last five minutes, the last hour, the last day, the last week, etc. 4. The LM App tells the user through the results of the data analysis pipeline running on the QualiMaster infrastructure, that the current systemic risk (based on real-time co-dependency data) is low, but the previous tendency was raising, also the prognosis says, there is a relevant chance, that risk will raise more. 5. Based on this information, the user decides to reduce the size of investment in general, raising the amount of available liquidity from 20% to 30%.

Extensions	<p>1a Login is not permitted due to invalid credentials and the actor is informed by the LM App</p> <p>3a Based on the role of the user (Senior or junior investment manager), the user is able to access/select more intermediate steps of analytics and can select specific pipeline paths for analysis. For example, he can see, on which input streams the risk analysis is mainly based on. The risk analysis of German blue chips is based on 80% price information, 15% social media streams and 5% News streams.</p> <p>3b The user is able to select/de-select each input stream. This may be done by a user-triggered adaptation of the data analysis pipeline. For example, he/she deselects the impact of social media on the risk analysis.</p> <p>4a The LM App does not report significant risks.</p> <p>5a Subsequent to 4a the user does not need to take new decisions.</p>
Business Rules	
Data/Functions	<ul style="list-style-type: none"> a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations

Use Case Identifier	UC-LM2
Use Case Name	Specific Systemic Risk Assessment for Individual Market Players
Actor	Hedge Fund Manager, Investment Company
Goal	Assist investment decision makers in modifying the liquidity of the investment pool with respect to the systemic risk
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The system provides risk information so that the actor is able to make decisions for modifying his portfolio liquidity.
Scenario Sequence	<ol style="list-style-type: none"> 1. Actor starts the QualiMaster LM App, enters his login credentials and becomes a user of the application 2. Depending on the role of the user, the LM App presents selection options for more or less detailed market segments and output visualizations 3. The user selects market segment 'German Blue Chips', selects Market players 'BMW', 'Volkswagen' and Currency 'EURUSD' (The user has invested in the German automobile industry and he knows that the profits of those market players are strongly dependent of exports to North America. So he includes the currency rate in the risk analysis). 4. The LM App tells the user through the results of the data analysis pipeline running on the QualiMaster infrastructure that the systemic risk in the German main market is stable (using

	<p>real-time co-dependency calculations). However, the risk for the automobile market player slightly raised based on a comparison of the current risk and historical risk data. The system also shows that there is a prognosis of raising currency rate risk.</p> <p>5. As this may affect the automotive market players, the user decides to reduce the position size for those market players.</p>
Extensions	<p>1a Login is not permitted due to invalid credentials and the actor is informed by the LM App</p> <p>3a Based on the role of the user (Senior or junior investment manager), the user is able to access/select more intermediate steps of analytics and can select specific pipeline paths for analysis.</p> <p>3b The user is able to select/de-select each input stream. This may be done by a user-triggered adaptation of the data analysis pipeline. For example, he deselects the impact of social media on the risk analysis.</p> <p>4a The LM App does not report significant risks.</p> <p>5a Subsequent to 4a the user does not need to take new decisions.</p>
Business Rules	
Data/Functions	<p>a. User information, login and authentication functions</p> <p>b. Market segment information</p> <p>c. User triggers into the QualiMaster infrastructure</p> <p>d. Output visualizations</p>

5.2.3 Application Use Cases for Systemic Risk Assessment for Regulatory Bodies

Since the credit crisis the environment for risk management within the Investment Banking (IB) sector, especially with regard to the regulatory response, has changed dramatically towards greater regulatory oversight together with significantly increased internal changes for improving risk management. Post Dodd-Frank in the US and European Market Infrastructure Regulation (EMIR) in the EU, IBs have been asked to provide the results of extensive stress tests designed to allow the regulators to assess the financial health of individual IBs. These stress tests can take many forms, from simple what-if scenarios (simple in definition but generally resource intensive to produce) to more complex tail risk analytics such as Value at Risk (VaR) or Comprehensive Risk Measure (CRM). Regulators use these stress tests to understand and specify capital requirements for IBs, which may apply bank wide or to specific trading activities.

The requirements for stress testing were certainly present, though in a reduced form, prior to the credit crisis. The difference today is not just the severity of the tests and their application, but more importantly the way in which cross-sector stress data is analysed. Prior to the credit crisis the focus was almost exclusively on the health of individual institutions. Little attempt was made to understand the nature of systemic risk transfer amongst IBs or to apply such understanding to the capital requirements of individual IBs. With the identification of Systemically Important Financial Institutions, regulators have acknowledged the importance of systemic risk and the need to reflect this in capital requirements.

Stress testing is an important component of identifying systemic risk in that it identifies the most vulnerable institutions. However, since stress tests generally take many months to complete, the information can rapidly become outdated as market events overtake prior results. Furthermore the

nature of stress contagion or alternatively, the topology of systemic risk networks, cannot be inferred without further information. Therefore it is important both for regulators, and for IBs themselves, to have access to the current state of the financial network in order to guide decision making during periods of stress.

The use cases below identify how QualiMaster will enable both regulators and IBs to proactively manage and rapidly respond to stress events in the financial markets, providing a valuable tool for reducing both impact of events as they occur and insight into the nature of systemic risk to inform policy making.

5.2.3.1 Enterprise Risk in Investment Banking

The actor (Investment Bank) wishes to understand better the interdependencies in the market and protect against contagion. We will call the realizing application the Enterprise Risk in Investment Banking Application (ERIB App). In the first use case (UC-ERIB1), the actor checks co-dependencies against existing portfolio members. In the second use case (UC-ERIB2), the actor checks co-dependencies against all markets.

Use Case Identifier	UC-ERIB1
Use Case Name	Checking co-dependencies against existing portfolio members
Actor	Investment Bank
Goal	Assist in capital allocation and systemic risk management
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The actor is able to make decisions for capital allocation and take measures to minimize systemic risk in the bank.
Scenario Sequence	<ol style="list-style-type: none"> 1. The actor starts the QualiMaster ERIB App, enters his login credentials and becomes an application user. 2. Depending on the role of the user, the ERIB App enables the user to select more or less detailed market segments and output visualizations. 3. The user selects 'New market player' and 'existing portfolio member', chooses analysis mode 'Compare players' visualization mode 'Dependency table'. 4. The ERIB App through the QualiMaster infrastructure with running data analysis pipeline tells the user whether the new market player(s) have strong dependencies to the MPs in which they have already invested in. 5. Based on this information, the user may modify assets/liabilities and take protection.
Extensions	<ol style="list-style-type: none"> 1a Login is not permitted due to invalid credentials and the actor is informed by the ERIB App 3a The user may choose to focus on a specific market participant in order to understand their dependency network. Having understood the degree to which the market participant is exposed to contagion may decide to reduce exposure. 3b The user may monitor the systemic risk network topology via, for example, centrality metrics, to understand in real time the health of the financial network. 3c The user may choose to take the systemic risk network and

	overlay the investment banks exposures by market participant in order to derive an investment bank specific systemic risk network.
Business Rules	
Data/Functions	<ul style="list-style-type: none"> a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations, in particular comparison and table visualization

Use Case Identifier	UC-ERIB2
Use Case Name	Check co-dependencies against all markets
Actor	Investment Bank
Goal	Assist in capital allocation and systemic risk management
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	The actor is able to make decisions for capital allocation and take measures to minimize systemic risk in the bank.
Scenario Sequence	<ol style="list-style-type: none"> 1. The actor starts the QualiMaster ERIB App, enters his login credentials and becomes an application user. 2. Depending on the role of the user, the ERIB App enables the user to select more or less detailed market segments and output visualizations. 3. The user selects 'New market player' and 'all markets', chooses analysis mode 'Compare players' visualization mode 'Dependency table'. 4. The ERIB App through the QualiMaster infrastructure with running data analysis pipeline tells the user which dependencies the new market player(s) has against all market segments. The user can now check whether there are dependencies to market segments in which the investor would not like to invest in (for example emerging markets). 5. Based on this information, the user decides to invest in the new market player or not.
Extensions	<ul style="list-style-type: none"> 1a Login is not permitted due to invalid credentials and the actor is informed by the ERIB App 3a The user may choose to focus on a specific market participant in order to understand their dependency network. Having understood the degree to which the market participant is exposed to contagion may decide to reduce exposure. 3b The user may monitor the systemic risk network topology via, for example, centrality metrics, to understand in real time the health of the financial network. 3c The user may choose to take the systemic risk network and overlay the investment banks exposures by market participant

	in order to derive an investment bank specific systemic risk network.
Business Rules	
Data/Functions	<ul style="list-style-type: none"> a. User information, login and authentication functions b. Market segment information c. User triggers into the QualiMaster infrastructure d. Output visualizations, in particular comparison and table visualization

5.2.3.2 Regulatory Monitoring

Regulators have access to real time market data from sources such as exchanges. Due to the volume of data, they currently do not have a comprehensive view of the state of the financial market. QualiMaster will provide such a view in real time that will enable them to identify key drivers of contagion and focus any preemptive response. The related application allows the actor to monitor the systemic risk network for possible signs of increasing dependency or contagion. We will call this application the *Regulatory Monitoring Application* (RM App).

Use Case Identifier	UC-RM
Use Case Name	Regulatory Monitoring
Actor	Regulator
Goal	Provide regulator with real time comprehensive monitoring of the state of the financial network.
Precondition	QualiMaster platform is running (UC-PA12) and the QualiMaster data analysis pipeline is started (UC-PA8).
Postcondition	Actor is able better regulate the financial markets.
Scenario Sequence	<ol style="list-style-type: none"> 1. The actor starts the QualiMaster RM App, enters his login credentials and becomes a user of the application. 2. Depending on the role of the user, the RM App allows the user to select more or less detailed market segments and output visualizations. 3. The user selects 'market participant' (node) to obtain detailed breakdown of information flow. 4. The RM App through the results of the configured data processing pipeline running on the QualiMaster infrastructure tells the user the market dependencies of the selected market participant. 5. Based on this information, the user decides to investigate the liquidity position of the market participant (via external sources). 6. The user monitors the centrality metric of systemic risk network. 7. The RM App displays the real time centrality metrics through the QualiMaster infrastructure. 8. The user responds to severe changes in centrality metrics as a signal for heightened monitoring.
Extensions	1a Login is not permitted due to invalid credentials and the actor is informed by the RM App

	<p>3a The user specifies pre-dependency analysis filters to input data in order to focus on specific asset types or regional segments. To perform efficient processing, this leads to a user-triggered adaptation of the data analysis pipeline.</p> <p>3b The user utilizes social data to enhance market sentiment insight and capture breaking news. This may be done by a user-triggered adaptation of the data analysis pipeline.</p>
Business Rules	
Data/Functions	<p>a. User information, login and authentication functions</p> <p>b. Market segment information</p> <p>c. User triggers into the QualiMaster infrastructure</p> <p>d. Output visualizations, in particular centrality metric of systemic risk</p>

Additional remarks

Some use cases indicate that the user provides feedback back to the QualiMaster infrastructure, which leads to a **user-triggered adaptation** of the data processing. However, this affects the costs of data processing. So two user models should be considered:

- 1) The user utilizes the QualiMaster infrastructure and the running pipeline 'as it is', without the possibility of pipeline modification. This would enable lower usage costs.
- 2) The user can send triggers to adapt the pipeline processing at runtime. This will result in higher and dynamically changing costs for using the system. Please note, that such an adaptation enables changes to a running data processing pipeline within boundaries given by the pipeline design in terms of structure and quality constraints. Such triggers do not imply the ability to modify or specify a new pipeline. This is a task of the Pipeline Designer.

5.3 Visualization

The results of the user survey (see section 3) also gave inputs to understand the requirements of the QualiMaster visualization system. In this paragraph we mention some examples of how the visualization will be addressed in the design and development of the QualiMaster platform. We provide some examples for the previously mentioned use cases.

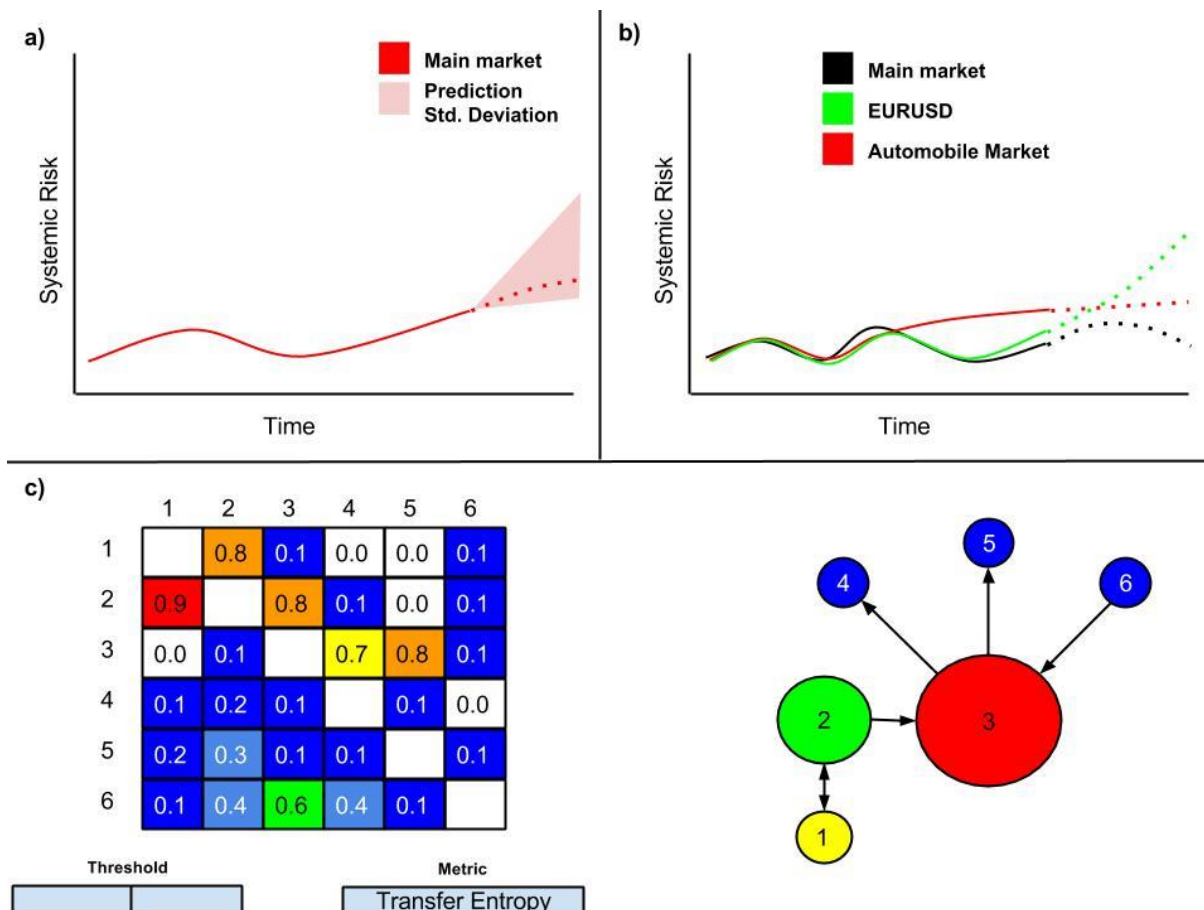


Figure 13: Example of visualisation for the use cases AppLM, PCASR and TOPS.

AppLM-1

The LM App tells the user through the results of the data analysis pipeline running on the QualiMaster infrastructure, that the current systemic risk is low, but the previous tendency was raising, also the prognosis says, there is a relevant chance, that risk will raise more. The corresponding visualization is presented in Figure 13 a). A time plot with historical systemic risk is shown using solid line. The dotted line is the prediction for systemic risk. Uncertainty, for example ± 2 std. deviations, is plotted as pink area on the diagram. Selecting standard deviation range should be possible (for example ± 2 std. deviations)

AppLM-2

The user selects market segment 'German Blue Chips', selects Market players 'BMW', 'Volkswagen' and Currency 'EURUSD' (The stakeholder has invested in the German automobile industry and he knows that the profits of those market players are strongly dependent of exports to North America. So he/she includes the currency rate in the risk analysis). The LM App tells the user through the results of the data analysis pipeline running on the QualiMaster infrastructure that the systemic risk in the German main market is stable. However, the risk for the automobile market player slightly raised. The system also shows that there is a prognosis of raising currency rate risk. A visualization of this scenario is presented in Figure 13 b). A time plot with historical systemic risk and predictions for the three selected entities is shown as solid lines. The forecasts are presented as dotted lines. Although omitted, uncertainty can also be plotted as in Figure 13 a).

PCASR-1

The user selects all market players in his portfolio. In case a market player in the portfolio is not provided by the underlying QualiMaster data analysis pipeline, the user selects a market index that better represents his market player. For analysis mode, the user selects 'Compare members'. For visualization, he selects 'Cluster representation'. The PCASR App provides a Cluster-like visualization and a corresponding numerical table as seen in Figure 13 c). A table is displayed. Color hue indicates the strength of dependency between two portfolio market players. Different hue settings are enabled. In the example, market player 2 is strongly dependent on market player 1. At the same time, market player 4 has no dependency on market player 1.

A graph corresponding to the table is also displayed. Nodes are sized/colored according to their degree. The user should have a choice on coloring/sizing configuration of graph display.

Only links of strength of 0.5 and above are displayed on the graph. A choice of threshold values is available. If a link is below certain threshold, it is not displayed in the graph. It is also blanked in the table according to the hue setting. This aids in visual assessment of market player importance.

A choice of dependency measure is available. Symmetrical and asymmetrical measures are available. In the first case the corresponding graph is directed, in the latter it is undirected.

TOPS-1 and TOPS-2

The use cases TOPS-1 and TOPS-2 can also be represented by the example in Figure 13 c). In this case, the graph nodes consist of the existing portfolio and of the new market players (for TOPS-1) or they represent the new market players and all market segments (for TOPS-2). The user can contrast graphs and/or tables with and without the new market players.

ERIB

The user selects 'New market player' and 'existing portfolio member', chooses analysis mode 'Compare players' visualization mode 'Dependency table'. The ERIB use cases through the QualiMaster infrastructure with running data analysis pipeline tells the user whether the new market player(s) have strong dependencies to the MPs in which they have already invested in. Depending on whether a symmetrical or asymmetrical dependency metric is used, different visualizations techniques are available. Some examples of visualization techniques for the EIRB use cases are illustrated in Figure 14 and in Figure 15.

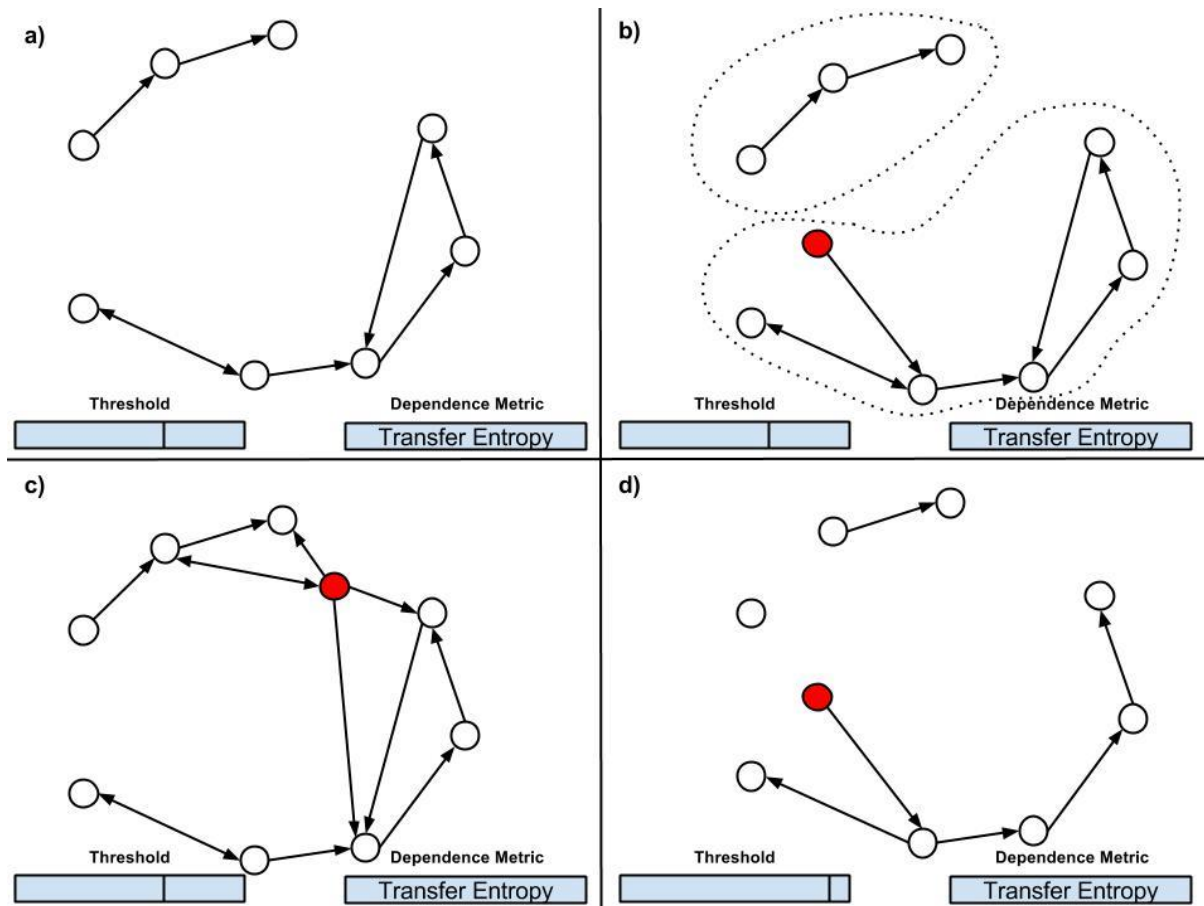


Figure 14: Example of visualisation for the EIRB use cases.

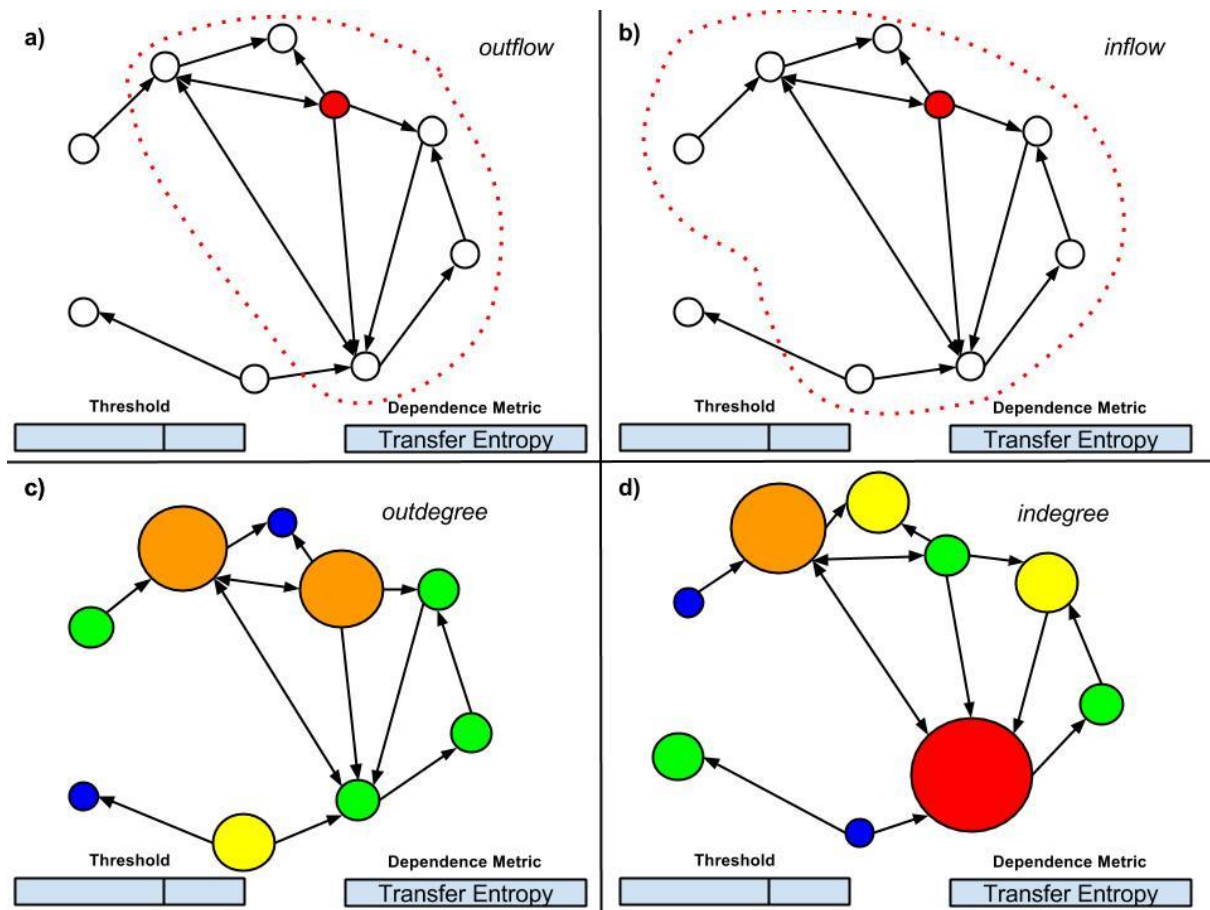


Figure 15: Example of visualisation for the EIRB use cases.

The user can see his existing portfolio dependency graph, as seen in Figure 14 a) or one with a new nodes added, as seen in Figure 14 b) and Figure 15 c). The red dot is the new market player. Depending on which market player is added, different connections will appear. The visualization can be in the form of a table and/or of a graph. The metrics used to determine dependence can be symmetrical or asymmetrical. For example the correlation or transfer entropy. As previously illustrated, either directed or undirected graphs are going to be created. A threshold setting is present. This determines which connections should be highlighted in the table, and which connections should be included in graphs. In Figure 14 b) the graph components are indicated. The user can indicate whether components should be highlighted. In Figure 14 d) the threshold is increased and some of the connections are removed.

The graph can be further customized. In Figure 15 a) the range of reachable outflow nodes for the new market player are indicated. The new market player has either a direct or indirect impact on the selected nodes. In Figure 15 b) the range of inflow nodes are indicated. Those are the nodes which have either a direct or indirect impact on the new market player. This type of analysis allows the user to see indirect dependencies between the new market player and current portfolio members. The user can customize the graph visualization. In Figure 15 c) and d) the user selected coloring and sizing of nodes according to their inflowing or outflowing connections. This can aid in quickly assessing important nodes. Multiple settings should be available.

RM

The centrality metric can be visualized in two ways. Either as a plot or as a graph. In a plot, the user selects some dependency metric and centrality metric. A time plot is created for the historical

data using those settings. This is similar to previously presented visualizations. A graph visualization of centrality is presented in Figure 16.

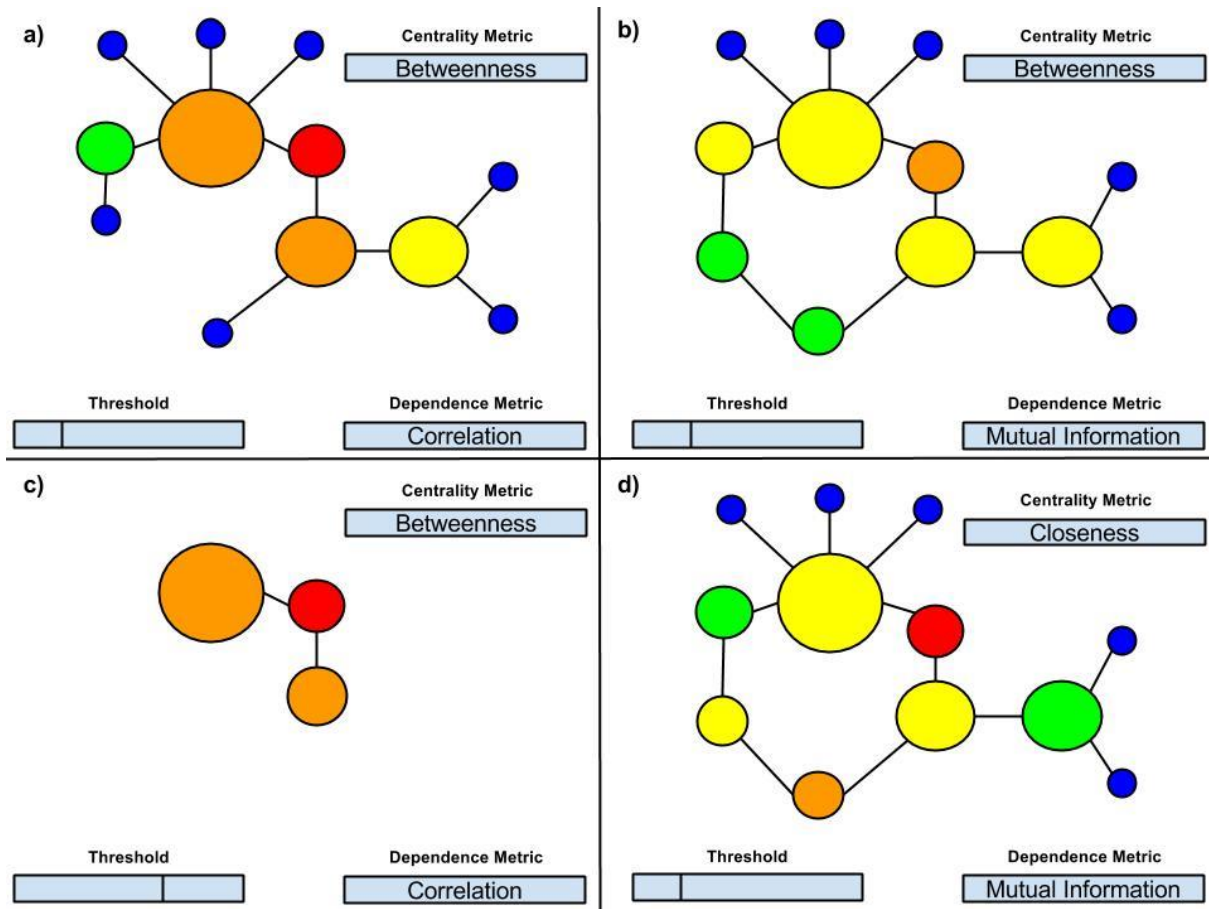


Figure 16: Example of graph visualization of centrality for the RM use case.

Depending on the metric setting, either a directed or undirected graph is visualized. The links are included based on the selected threshold value. As seen in Figure 16 a) and b), different metrics impact the graph structure. Higher threshold setting can remove some links from the graph as seen in Figure 16 c). The centrality metric also impacts the graph as seen in Figure 16 d). The size is used to indicate the degree of the node, and the color is used to indicate its centrality metric. A plot represents the centrality measure for the graph. It can be configured for a different centrality metrics and dependency metrics. The plot shows trend predictions, and follows the format presented in Figure 16 a).

6 Data, Algorithm and Quality Requirements

In this section, we discuss the requirements that arise from the data and the algorithms that will be used by the QualiMaster applications as introduced in Section 5.1. First, in Section 6.1 we discuss the requirements collected so far for the data (sources). Then, in Section 6.2 we present the requirements for the algorithms to be applied. Finally, in Section 6.3 we will discuss initial quality tradeoffs and derive quality requirements for the QualiMaster platform as a basis for the adaptive execution of pipelines. All sections start with a discussion on the background and conclude with initial requirements. The requirements will be marked with a unique identifier for further and future reference and will be given in a simplified form of controlled natural language, which is frequently used in requirements engineering to avoid ambiguities (e.g. [2, 6, 11]).

6.1 Data and Data Stream Requirements

The QualiMaster infrastructure must support various data sources with different characters and specifications. The requirements originating from these data sources are discussed in the following paragraphs.

6.1.1 Types of Data Sources

As introduced in Section 5.1, the QualiMaster applications will be built on two types of data sources. Data from the **financial domain** is the main source. With respect to this, QualiMaster must handle data from foreign exchange rates (currently around 140 currency pairs), futures on indices and commodities, bond markets, stocks from worldwide exchanges and market indices. Additional sources for financial data that might be also incorporated in QualiMaster are international interest rates. The financial data is provided by SPRING through a specific API to the project partners.

The second QualiMaster data source is **Web data**. This is a collection of contributions from individual sources on the Web (e.g., posts about products, stocks, companies, real state, jobs) as well as experts from the financial domain (e.g., posts about recent expert analysis, studies, job reports, predictions) in micro-blogging systems such as Twitter. In particular, LUH is collecting Twitter data from the public sample stream via its REST API [28]. The API returns a random sample of 1% of all public tweets. LUH is using several parallel streams to increase the amount of collected tweets. In order to increase the amount of data related to the financial domain, additional focused streams will be collected using the public filter stream API [28]. For example:

- Streams filtered by a static set of financial terms using relevant and general terms such as “financial market”, “stocks”, “banks”, “Dow Jones” etc, that are time independent. The filter terms will be defined in the configuration and may need manual reconfiguration over time or adjustments by user triggers.
- Streams filtered by a dynamic set of financial terms using a set of terms that are related to current events. The terms shall be extracted from current news, and updated continuously.
- Streams of news agencies and users that are identified as “experts” in the financial domain.

While the financial data is typically composed by numbers, the Web data is composed by more complex data types. For instance, a tweet from Twitter contains the actual text and a lot of metadata [29]. The metadata contains fields related to the tweet such as creation time, hashtags used, URLs included in the tweet and whether tweet is a retweet or not and the geo coordinates or location of the user when sending the tweet. In addition, the metadata includes fields related to the user, such as number of followers, friends, tweets, location, time zone and many more.

Furthermore, QualiMaster will collect and analyze information from online news that can have an impact on the financial market, such as political news (e.g., new elections, armed conflicts), science and technology (e.g., trends in technology that may affect some industrial sectors), and news on natural disasters (e.g., major earth quakes, tsunamis etc).

While the financial data is typically structured and composed by numbers, the Web data is less structured, heterogeneous and composed by more complex data types. In addition to the more complex format, Web data can have noise, incorrect values, may be biased and the meaning or the actual interpretation is typically uncertain (since the messages are expressed in evolving natural languages). QualiMaster must be able to handle the different data formats of the incorporated data sources.

6.1.2 Real-time vs. Static Sources

In addition to the real-time data, the QualiMaster infrastructure will incorporate static data sources. This is historical financial data, collected for up to 20 years, depending on the availability of market players (available through SPRING). The historical data of QualiMaster has the same format and characteristics as the real-time data and covers most market players. QualiMaster needs to be

able to batch process the past data to provide historical time series of the systemic risk. The collection of Twitter data from the public sample stream by LUH started already in February 2013. There have been 3.7 billion tweets (~2.22 TByte) collected till end of March 2014, which can be used as historical data. In order to increase the amount of data related to the financial domain, additional focused streams shall be collected. This includes the Web data stream examples described in Section 6.1.1. Data Stream Characteristics

We will detail the characteristics of the data streams in terms of three dimensions, the data load, the message rate and the volume.

6.1.3 Data Load

6.1.3.1 Number of streams

QualiMaster must be able to handle various data streams arriving at the same time. For instance, there will be more than 200 pairs of foreign exchange rates, around 500 virtual streams for the futures, indices and bonds from one source. Information coming from international stock exchanges will provide some thousand stocks, which implies the same number of virtual streams.

6.1.3.2 Rate of Messages (i.e., messages per second arriving at QualiMaster)

In average market situations, an individual stock causes around 400 messages per second (see, for example, the Microsoft stock at NASDAQ with estimated 500.000 trade counts per day). According to the experience of the industrial partners in QualiMaster, the number of messages may grow by a factor of around 20 in times of larger market movements, news or other impacts.

In addition to the messages coming from the financial data, QualiMaster must be able to handle the messages coming from the Web data, including News articles and tweets. The News articles are around 400 per day. The tweets collected by LUH via Twitter's public sample stream (1% of all tweets) have an average rate of 300 per second. Obviously, this can increase the total rate of tweets which, as listed in the official statement of 2011, has an average rate of 4600 per second.

6.1.3.3 Volume of Messages (i.e., messages per second arriving at QualiMaster)

A larger exchange of about 2000-3000 stocks (e.g., NASDAQ) produces about 4 million messages per second. This results in a volume of about 10 GByte data per day, which QualiMaster must be able to handle. In addition, there will be around 5 GByte tweets that are collected and provided to QualiMaster by LUH via Twitter's public sample [28]. This corresponds to 1-3% of all tweets (this is the union of all sampled tweets that are collected from three parallel streams, each receiving 1% of all public tweets).

6.1.4 Requirements

From the data stream background introduced by the sections above, we summarize the following requirements in the style of controlled natural language (possible with additional information in natural language).

- REQ-DS1: The QualiMaster infrastructure must support multiple real-time data sources, possible with each source having a different type of data. *Information:* In particular, this includes structured financial stock market data and social Web data.
- REQ-DS2: The QualiMaster infrastructure must support filtering of data according to criteria. *Information:* Filtering is at least defined through the configuration and may be influenced by user triggers.
- REQ-DS3: The QualiMaster infrastructure must support the integration of historical data sources and data processing. *Information:* This also includes queries over historical data.
- REQ-DS4: The QualiMaster infrastructure must support at least 400 stock market messages per second per market player under normal load. *Information:* We plan for 1.500 market players, i.e., 600.000 stock market messages per second.

- REQ-DS5: The QualiMaster infrastructure must support growth rates up to factor 20 over normal load for stock market streams.
- REQ-DS6: The QualiMaster infrastructure must support processing at least 1% of all public tweets from Twitter.
- REQ-DS7: The QualiMaster infrastructure must support growth rates up to 10% of all public tweets from Twitter.
- REQ-DS8: The QualiMaster must support processing online public news and financial blogs.

6.2 Requirements for Supported Algorithms

QualiMaster will implement algorithms designed to identify co-dependency and causality in multivariate time series. Co-dependency describes the degree to which time series tend to move together, typically captured by correlation. Co-dependency measures are symmetric. Causality describes the degree to which a time series is influenced by the prior behavior of another time series [14]. There are a number of different approaches to Causality, most of which utilize the concept of Information or Entropy. Since we are interested in contagion effects we will focus primarily on causality, although codependence will be useful as a way to quickly identify possible causal relationships.

The co-dependency or causality relationships between major market participants will be inferred from streaming market and social data, suitably filtered and normalized. Note that data of different market participants will arrive asynchronously. Classical time series analysis assumes that a multivariate vector of values may be sampled at regular time points. Where possible, QualiMaster will adapt such algorithms to allow incremental update of metrics suitable for asynchronous data.

A systemic risk network [8] quantifies the linkage between major market participants and provides a framework for identifying instability in financial markets. Major changes in the topology of the network (e.g., a sudden increase in the risk of contagion [1]) can signal ongoing financial stress. For co-dependency measures undirected, weighted networks will describe the degree of risk clustering. For causal measures directed, weighted networks will describe the major sources and flow of information.

The computation of information flow is in general significantly more expensive than correlation. Therefore, an initial correlation analysis will be done for a broad range of market participants to reduce dimension to subset of interest prior to the causality analysis.

We now give an overview of the main codependence and causality algorithms to be incorporated in QualiMaster.

6.2.1 Correlation

Pearson's correlation is a measure of the linear dependence between two variables. In the case of two time series X_t and Y_t , it is a measure of the extent to which a movement of X happens simultaneously with a proportional movement of Y . The correlation is a value between -1 and 1. If this value is 1, respectively -1, then a movement of X in one direction happens at the same time as a proportional movement of Y in the same, respectively the opposite direction. At the other extreme, a correlation of 0 means that there is no linear relation between the movements of X and Y . Note the importance of the word linear here since independent variables will have correlations of 0, but the reverse is not true.

The classical approach to statistical correlation estimation assumes that the time series X_t and Y_t update synchronously. This is not the case in general for financial time series so that a feasible method to correlation in the financial domain needs synchronise the time series. The choice of the method is important since a poor choice can lead to biased results depending on the relative frequency of data points in the time series. Different approaches to synchronisation will be tested including interpolation (e.g., based on the last value) and Fourier correlation (see [21] for a survey).

An alternative to the classical correlation statistics first introduced in [13] will also be implemented. Here, an asynchronous estimator for the correlation of times series is defined based on the degree of inter-event overlap. The algorithm admits an inline version based on a state machine representation. The algorithm can be used to measure lead/lag relationships.

Pros

Correlations can be computed quickly compared to other methods. Computing correlations is an industry standard technique that is widely known and well accepted. Further, correlations have been used efficiently in more advanced studies such as Correlation Networks (see [19] for an example, or [3] for a survey).

Cons

Pearson's correlation only exposes linear relations between time series and may fail to properly represent non-linear co-dependencies in time series data. In particular, Pearson's correlation is not invariant under monotonic transformations of the marginal distributions. Alternatives to Pearson's correlation, such as Spearson's or the more general class of Rank Correlations mitigate this somewhat. More importantly, correlation metrics do not infer directionality or cause-and-effect relationship, which is a key feature of a systemic risk network.

6.2.2 Mutual Information

In the initial requirements collected in D1.1, Granger Causality had been chosen for the estimation of causality between time series. However, during the refinement of the requirements in D1.2, it has been decided to replace Granger Causality with Mutual Information.

The reason for replacing Granger Causality with Mutual Information is that Mutual Information and Transfer Entropy are similar measures, both relying on an information metric for the time series. However they differ in that the former is symmetric while the latter is asymmetric. Thus MI is cheaper to compute but less informative and TE is more expensive to compute but more informative. The idea is that the sequence of algos captures increasingly detailed but related statistics on the time series so that we can employ adaptivity techniques. If the algos were all completely different there would be no concept of "zooming in" on the data and we would lose an application of adaptivity.

Information entropy, or simply entropy for short, was defined by Shannon [27] and is a measure of the uncertainty of a random variable or equivalently the average number of bits described by the random variable. The higher the entropy, the more uncertain it is or the more information is obtained on average by sampling the random variable. More formally, given a random variable X , its entropy is the average number of bits necessary to represent an outcome of X . The precise formula is

$$H_X = - \sum_x p(x) \log_2 p(x) \quad (2)$$

In formula 2, H_X denotes the entropy of X and $p(x)$ is the probability of a particular outcome x of X .

Given two time series X_i and Y_i , one can ask the question "How much information is encoded in X_i if we assume that we already know Y_i ?" Another way of asking this question is "How many bits on average are necessary to encode an outcome of X_i if we assume that we have already encoded Y_i ". One example where the answer to the above question is 0 could be if X_i is always equal to Y_i . In general, this value will not be 0.

Assume that both, X and Y , are Markov processes, then the mutual information [37] between X and Y is defined by:

$$I(X; Y) = \sum_{x,y} p(x_t, y_t) \log_2 \left(\frac{p(x_t, y_t)}{p(x_t)p(y_t)} \right) \quad (3)$$

The computation of mutual information requires estimation of the joint distribution of (x_i, y_i) which may be achieved through binning or kernel estimation, amongst other methods [38-40].

Pros

Mutual information is sensitive to non-linear signal properties. It is known that financial data often follows heavy tailed distributions, to capture those relationships the dependency measure needs to take into account higher order statistics [34-36].

Cons

Mutual Information is more computationally intensive than the Hayashi-Yoshida linear correlation coefficient estimator [33]. It also requires substantially more data to provide statistically significant results [38-40]

6.2.3 Transfer Entropy

Information entropy, or simply entropy for short, was defined by Shannon [27] and is a measure of the uncertainty of a random variable or equivalently the average number of bits described by the random variable. The higher the entropy, the more uncertain it is or the more information is obtained on average by sampling the random variable. More formally, given a random variable X , its entropy is the average number of bits necessary to represent an outcome of X . The precise formula is

$$H_X = - \sum_x p(x) \log_2 p(x) \quad (2)$$

In formula 2, H_X denotes the entropy of X and $p(x)$ is the probability of a particular outcome x of X .

Given two time series X_t and Y_t , one can ask the question “How much information is encoded in X_t if we assume that we already know past values X_s and Y_s for $s < t$?” Another way of asking this question is “How many bits on average are necessary to encode an outcome of X_t if we assume that we have already encoded past values of X_t and Y_t ”. One example where the answer to the above question is 0 could be if X_t is always equal to Y_{t-1} . In general, this value will not be 0.

Assume that both, X and Y , are Markov processes, then the transfer entropy [15] from Y to X is defined by:

$$T_{Y \rightarrow X} = \sum_{x,y} p(x_t, x_{t-1}, y_{t-1}) \log_2 \left(\frac{p(x_t | x_{t-1}, y_{t-1})}{p(x_t | x_{t-1})} \right) \quad (3)$$

Formula 3 can be extended to look at a set of past values of X , not just at x_{t-1} . One can also increase the number of past values for Y . The computation of transfer entropy requires estimating the joint distribution of (x_t, x_{t-1}, y_{t-1}) which may be achieved through binning or kernel estimation, amongst other methods [14].

Pros

Information entropy is not symmetric, i.e., $T_{Y \rightarrow X}$ must not necessarily be equal to $T_{X \rightarrow Y}$. It can be used to distinguish between driving and responding elements [25]. For that purpose, transfer entropy is useful when creating directed networks or market players and seeking for the market players that are driving market movements. In that respect, transfer entropy is similar to G-causality. Its advantage over G-causality is that it is sensitive to non-linear signal properties [26].

Cons

Transfer entropy is more computationally intensive than correlation. It also requires substantially more data than G-causality to provide statistically significant results. Different techniques have been suggested to improve performance [14].

6.2.4 Algorithm flexibility

Algorithm flexibility requires that the set of available algorithms in the QualiMaster infrastructure, in particular an instantiated version for a certain domain such as finance can be adjusted as required. Actually, also the algorithm families grouping similar algorithms with different quality tradeoffs and the pipelines using the algorithm families shall be flexible. This is targeted by the core concepts of the QualiMaster project, in particular in terms of configuring the platform and instantiating it according to the configuration. In more details, algorithm flexibility is covered by the use cases

- UC-PA3 (add data processing algorithm)
- UC-PA4 (modify data processing algorithm)
- UC-PA5 (add hardware-based algorithm)
- UC-PA6 (modify hardware-based algorithm).

These use cases also implicitly handle the management of the algorithm families. Combining algorithm families to pipelines and, thus, achieving flexibility there is covered by UC-PD1 (define new pipeline), UC-PD2 (modify pipeline definition) and UC-PD3 (delete pipeline definition).

6.2.5 Requirements

From the algorithmic background introduced by the sections above, we summarize the following requirements in the style of controlled natural language (possible with additional information in natural language).

- REQ-ALG1: The instantiated QualiMaster infrastructure for systemic risk analysis must support the calculation correlation for synchronized time series data streams. *Information:* At least interpolation and Fourier correlation will be considered.
- REQ-ALG2: The instantiated QualiMaster infrastructure for systemic risk analysis must support the calculation of correlation networks based on time series data streams.
- REQ-ALG3: The instantiated QualiMaster infrastructure for systemic risk analysis must support the calculation of Granger causalities on point processes of time series data streams.
- REQ-ALG4: The instantiated QualiMaster infrastructure for systemic risk analysis must support the calculation of transfer entropy for time series data streams. *Information:* Techniques for improving the performance will be applied.

Alternative ways of implementing the algorithms, e.g., those mentioned in REQ-ALG1, will form the basis for respective algorithm families. Furthermore, the algorithms will be considered for implementation in reconfigurable hardware in workpackage 3, i.e., depending on the approach for translating algorithms to hardware, appropriate algorithms will be chosen for demonstration and experimentation.

6.3 Platform Quality Requirements

Based on the sections before, we present now general quality requirements for the QualiMaster platform. In Section 6.3.1, we start with a discussion of quality dimensions. In Section 6.3.2, we summarize the requirements drawn from the discussion of the quality dimensions.

6.3.1 Quality Dimensions

In this section, we discuss basic quality dimensions such as timeliness, coverage, accuracy, efficiency (performance) and resource consumption in the context of the applications and the QualiMaster infrastructure.

In order to clarify the relevant Quality Dimensions, WP4 conducted a quality survey in order to determine the relative importance of individual dimensions and quality properties. This survey covered pipeline- and data-processing related aspects, infrastructure aspects as well as the expected user triggers and high-level infrastructure settings for the adaptation. We report now on

the understanding of the quality dimensions of the consortium from a requirements point of view. For more details on the quality survey and the derived quality taxonomy, please refer to D4.1.

Timeliness

As described in Section 6.1, QualiMaster must enable the processing of real-time data streams to produce up-to-date analysis results in addition to historical data (REQ-DS6). This requires the capability to deal with high velocity data streams such as the financial tick data from the stock markets (REQ-DS1). By combining both, real-time and historical data, QualiMaster will be able to produce trend analysis and predictions of future development in the financial market. Real-time processing of large volume data with high velocity is typically computationally expensive (in terms of CPU and memory consumption). Therefore, means must be provided to express preferences in terms of timeliness vs. computational cost.

Coverage

The coverage of the data sources to be processed in QualiMaster shall be maximized to produce a comprehensive market analysis. However, there is a trade-off between the volume of data being processed and the computational cost. Processing more data will probably consume more computational resources. On the other hand, there is a trade-off between the volume of the data to be processed and efficiency, within a given set of resources consumption constraints (more data require more time to be processed). Here, user triggers from the applications may enable the user to influence the calculation and to express preferences in terms of coverage vs. performance.

Accuracy

The accuracy of prediction models vary depending on several factors such as the characteristics of the data and the underlying statistical and machine-learning models. Using more historical data and more accurate models typically is associated with some additional computational cost in terms of resources consumption and computation time. While increasing the accuracy of the applied model is desired, it is expected that the timeliness and performance of the analysis can be negatively affected. Therefore, the QualiMaster infrastructure must allow for balancing between these quality parameters.

Computational Performance

To achieve timeliness but also to satisfy the user, it is required to minimize the overall computation time of the analysis tasks. This may be achieved by parallelizing algorithmic tasks as well as in software-based algorithms as well as on reconfigurable computing (e.g., using multiple DFE boards in parallel). Further, work package 3 will analyze the required algorithms for translating them into hardware and to eliminate potential bottlenecks of software-based execution (as this was done by MAX in several application settings in the past). Generally, there is a trade-off between the efficiency and the timeliness, coverage and accuracy properties, which should be taken into account when applying adaptation tasks.

Resources Consumption

To enable further pipelines, value added computing or elastic resources, it is required to minimize the overall consumption of computational resources as much as possible. Given a set of constraints on the computational resources (CPU and memory) the system should be able to maximize the quality of the produced analysis and/or satisfy domain or user defined quality requirements (in terms of accuracy, coverage and timeliness) as much as possible.

Content quality

The data processing in the QualiMaster use cases rely on financial as well as social web data. In particular in web data analysis source data may be noisy, imprecise or not exact. Thus, it is important for the processing to characterize the degree of relevance, i.e., whether information is relevant for the actual processing as well as the credibility of the data / the produced results. A decreasing content quality may negatively affect the accuracy of the combined / overall results.

Scalability

The QualiMaster infrastructure aims at (real-time) processing of large data amounts in the field of Big Data. Big Data is typically characterized in terms of the 3 V's, i.e., Volume, Velocity and Variety⁵. It is required that the QualiMaster infrastructure is designed and optimized for processing of Big Data in these dimensions and takes these dimensions also in account for runtime optimization, i.e., to enable and to characterize the scalability of the QualiMaster approach. Please note that these dimensions are typically used to characterize the data to be processed, e.g., the amount of data in Bytes. Instead, related runtime quality measures such as volume per time unit (for Volume) or number of data sources (for Variety) must be taken into account to characterize and optimize for scalability. Scalability is correlated to timeliness, computational performance and coverage, but affected by the maximum available resources.

6.3.2 Requirements

In this section, we summarize the discussion above in terms of initial (dedicated) quality requirements, which give a first indication for the tradeoffs to be specified and handled by the adaptivity in QualiMaster. Preferences among the individual quality dimensions will be specified in terms of pipeline or adaptation constraints as well as adaptation rules (see use cases in Section 7).

- REQ-Q1: The QualiMaster infrastructure must support timeliness in the processing of real-time data streams in order to produce up-to-date analysis results.
- REQ-Q2: The QualiMaster infrastructure must support means to customize the coverage of the data sources to produce a comprehensive market analysis. *Information:* Preferences of coverage vs. performance may be given in terms of user triggers (see also REQ-DS2).
- REQ-Q3: The QualiMaster infrastructure must support means to specify the accuracy of the performed calculation, for example in terms of the confidence of the processing, the confidence of the results or the overall error rate due to the composition of various algorithms.
- REQ-Q4: The QualiMaster infrastructure must support dynamic means to exploit mechanisms to maximize computational performance. *Information:* The flexible integration of hardware-based computing will enable the adaptivity to dynamically exploit the benefits of reconfigurable computing where applicable.
- REQ-Q5: The QualiMaster infrastructure must provide means to measure and optimize its resource usage.
- REQ-Q6: The QualiMaster infrastructure must provide means to characterize and react on the content quality of the data processing.
- REQ-Q7: The QualiMaster infrastructure must be designed and optimized during realization and at runtime for scalability in order to cope with Big Data challenges. In particular, this includes the number and diversity of data sources, covering the number of stocks to be processed, as well as the market depth.

⁵ http://en.wikipedia.org/wiki/Big_data

7 System Requirements and Use Cases

In this section, we describe the identified system use cases from the point of view of an infrastructure use. Those use cases are structured along the three types of infrastructure users, which have been identified as actors in the context of the QualiMaster Infrastructure. Further, we will use the term repository for mechanisms storing and retrieving configuration data for such as the Pipeline Repository for all information directly related to pipelines, e.g., the data (processing) flow.

Initially in D1.1, three separate configuration tools had been conceived, one for each Infrastructure User, in order to describe their individual use cases. There, we called the tools “Pipeline Configuration Tool”, “Adaptation Manager Tool” and “Platform Administration Tool”. According to the initial use cases, an actor implicitly had access to “his/her” tool(s) and to none of the other tool(s) so that user management aspects such as logging into a tool were not detailed. However, an implementation may realize these conceptually separated tools also in terms of one integrated configuration tool, which then needs some form of user management. An integrated tool supports the various Infrastructure Users to obtain an overview on the configuration aspects maintained by the other roles. For example, the Infrastructure Administrator may look into a configured pipeline in order to better understand the expected resource needs of the pipeline when approving changes to existing pipelines or starting new pipelines. Thus, we decided to integrate the three separate tools as views into one QualiMaster Infrastructure Configuration Tool (QM-IC tool) and to discriminate the roles by a login mechanism, e.g., through the authentication against the repositories mentioned above. Upon login, an Infrastructure User is identified and its actor role is determined, so that the actual user gains write access according to the use cases and, in case of missing write permissions, read-only access to view the other configuration settings.

7.1 Pipeline Designer

As defined in Section 2.2, the Pipeline Designer creates a data processing pipeline by combining data processing elements. For this purpose, the following three use cases have been identified for the Pipeline Designer:

- UC-PD1: Define new pipeline
- UC-PD2: Modify Pipeline definition
- UC-PD3: Delete Pipeline definition

Some elements of the pipeline may be implemented in terms of dedicated hardware, e.g., a FPGA-based processor, such as a Maxeler dataflow supercomputer. These elements will be possible to be included as needed in the processing but need configuration by the Platform Administrator (see UC-PA5, UC-PA6, UC-PA11).

7.1.1 Use Case: Define New Pipeline

This use case enables the Pipeline Designer to define a new pipeline based on underlying technical configuration parts using the QM-IC tool. The definition of a pipeline includes the validation of syntax, semantics and feasibility of the pipeline and finally storing the new pipeline configuration.

Use Case Identifier	UC-PD1
Use Case Name	Define new pipeline
Actor	Pipeline Designer
Goal	Define a data stream analysis pipeline based on underlying configuration information, existing data processing elements and static quality validation.

Precondition	QualiMaster platform is configured (UC-PA10, optionally UC-PA11), data sources and sinks are configured (UC-PA7), quality characteristics are defined (UC-AM1)
Postcondition	New pipeline is validated and successfully stored
Scenario Sequence	<ol style="list-style-type: none"> 1. The Pipeline Designer starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Pipeline Designer provides his / her credentials and selects the definition of a new pipeline. 4. The QM-IC tool grants the access, updates the Configuration from the repository and shows an editor to enter the new pipeline. Available parts of a pipeline in particular sources, sinks and progressing elements (algorithm families) are shown to simplify the creation of a pipeline. 5. The Pipeline Designer enters the pipeline by selecting processing elements, defining the data flow between sources and data processing elements, among data processing elements, and, finally, from data processing elements to sinks. 6. The QM-IC tool checks the syntactic and semantic validity of the pipeline (as far as possible interleaved with step 5). This includes, whether successors of processing elements can be linked, whether sources are connected and paths to sinks are present. 7. The Pipeline Designer adds quality constraints, such as SLAs for sources and sinks, constraints on the output quality of the processing elements or constraints on the connecting data flows. 8. The QM-IC tool checks the syntactic and semantic validity of the quality constraints (as far as possible interleaved with step 6). 9. The Pipeline Designer initiates a static analysis of the feasibility of the pipeline, i.e., whether quality constraints can be met and whether the infrastructure is basically feasible for executing the pipeline (e.g., based on the actual or the maximum available resources). 10. The QM-IC tool performs the static analysis of the end-to-end pipeline quality and whether the underlying pipeline infrastructure is basically capable of executing the configured pipeline. 11. The Pipeline Designer stores the configured pipeline (using a symbolic name for the pipeline design) into the pipeline repository of the infrastructure. 12. The QM-IC tool acknowledges the successfully stored pipeline. 13. The Pipeline Designer logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Pipeline Designer enters wrong credentials and the QM-IC tool refuses the access and queries the credentials again. 4b The Pipeline Designer may access the configuration of other pipelines in order to reuse existing parts.

	<p>6a In case of syntactic or semantic errors in the constraint syntax, the QM-IC tool displays appropriate messages in human readable form and highlights the involved elements. The use case continues at step 3.</p> <p>8a In case of syntactic or semantic errors in the constraint syntax, the QM-IC tool displays appropriate messages in human readable form and highlights the involved elements. The use case continues at step 5.</p> <p>10a In case of an infeasible infrastructure, the QM-IC tool indicates missing resources and suggests the increase of resources. The use case continues at step 3, 5 or 7.</p> <p>10b In case that overall quality constraints cannot be fulfilled, the QM-IC tool highlights critical parts or critical data flows. The use case continues at step 3, 5 or 7.</p> <p>12a In case of a syntactically, semantically or not validated pipeline, the QM-IC tool informs the Pipeline Designer about the actual status and stores the draft pipeline for further configuration.</p> <p>12b In case of a physical storage error, the QM-IC tool informs the Pipeline Designer about the failed pipeline repository action.</p>
Business Rules	<ul style="list-style-type: none"> Invalid pipelines cannot be executed on the QualiMaster platform. Pipeline configurations which exceed the actual resources of the platform and the underlying hardware cannot be executed.
Data/Functions	<ul style="list-style-type: none"> a. Access to sources and sinks b. Access to quality parameter c. Syntactic pipeline analysis d. Semantic pipeline analysis e. Static pipeline quality analysis f. Access the pipeline repository g. Structural display (or visualization) of pipelines including editor

Additional remarks:

We consider the technical information about the pipeline in the responsibility of the pipeline administrator (at the moment the technical information about sources and sinks such as IP addresses, credentials, the adapter etc) in order to separate concerns between pipeline design and its technical realization.

Actually, the QM-IC Tool may provide a graphical or a textual editor for the pipelines. Actually, the QualiMaster consortium prefers a graphical editor representing the analysis data flow in order to ease the adoption by real life Infrastructure Users, also as nowadays composition of analysis flows is typically done in programming languages rather than using a high-level graphical interface. In this case, syntactic as well as semantic pipeline analysis will in particular be based on the Configuration Meta Model, i.e., its translation into IVML, the INDENICA variability modeling language [7, 24] used to specify the Configuration Meta Model as well as semantic Configuration constraints defined by the QualiMaster consortium in IVML. The resulting IVML model then used for instantiating the QualiMaster platform / the pipeline utilizing the SPLE tooling provided by SUH. For more details on IVML and the Configuration Meta Model please refer to D4.1, for details regarding the instantiation to D5.2.

In practice, steps 5-8 of the use case scenario may happen in an interleaved and incremental fashion. In the use case scenario, we provided these in a conceptually separated form. In fact, closing the QM-IC tool without storing the pipeline will lead to a warning.

7.1.2 Use Case: Modify Pipeline Definition

This use case enables the pipeline designer to modify an existing pipeline definition based on underlying technical configuration parts using the QM-IC tool. Akin to the definition of a new pipeline (UC-PD1), the modification of a pipeline definition requires the validation of syntax, semantics and feasibility of the pipeline and, finally, storing the new pipeline configuration.

Use Case Identifier	UC-PD2
Use Case Name	Modify pipeline definition
Actor	Pipeline Designer
Goal	Modify the definition of an existing data stream analysis pipeline based on underlying configuration information, existing data processing elements and static quality validation.
Precondition	QualiMaster platform is configured (UC-PA10, optionally UC-PA11), data sources and sinks are configured (UC-PA7), quality characteristics are defined (UC-AM1)
Postcondition	Existing pipeline definition is modified, validated and successfully stored
Scenario Sequence	<ol style="list-style-type: none"> 1. The Pipeline Designer starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Pipeline Designer provides his / her credentials. 4. The QM-IC tool grants access, updates the Configuration from the repository and shows the existing pipelines. 5. The Pipeline Designer selects the modification of an existing pipeline and specifies which of the pipelines stored in the pipeline repository shall be modified. 6. The QM-IC tool shows an editor displaying the pipeline selected in Step 5. Available parts of a pipeline in particular sources, sinks and progressing elements (algorithm families) are shown to simplify the modification of the pipeline. 7. The Pipeline Designer modifies the pipeline adding, changing or removing processing elements and by (re)defining the data flow between sources, sinks and data processing elements. 8. The QM-IC tool checks the syntactic and semantic validity of the pipeline (as far as possible interleaved with step 7). This includes, whether successors of processing elements can be linked, whether sources are connected and paths to sinks are present. 9. The Pipeline Designer adds, removes or changes quality constraints such as SLAs for sources and sinks, constraints on the output quality of the processing elements or constraints on the connecting data flows. 10. The QM-IC tool checks the syntactic and semantic validity of the quality constraints (as far as possible interleaved with step 9). 11. The Pipeline Designer initiates a static analysis of the feasibility

	<p>of the pipeline, i.e., whether quality constraints can be met and whether the infrastructure is basically feasible for executing the pipeline (e.g., based on the actual or the maximum available resources).</p> <p>12. The QM-IC tool performs the static analysis of the end-to-end pipeline quality and whether the underlying pipeline infrastructure is basically capable of executing the configured pipeline.</p> <p>13. The Pipeline Designer stores the modified pipeline into the pipeline repository of the infrastructure.</p> <p>14. The QM-IC tool asks the Pipeline designer whether the existing pipeline definition shall be overwritten.</p> <p>15. The Pipeline Designer acknowledges that the existing pipeline shall be overwritten.</p> <p>16. The QM-IC tool acknowledges the successfully stored pipeline.</p> <p>17. The Pipeline Designer logs out from the QM-IC tool.</p>
Extensions	<p>4a The Pipeline Designer enters wrong credentials and the QM-IC tool refuses the access and queries the credentials again.</p> <p>5a No pipeline definitions are available for modification, i.e., the QM-IC tool will not show pipelines for selection and the scenario ends.</p> <p>5b The Pipeline Designer may access the configuration of other pipelines in order to reuse existing parts.</p> <p>8a In case of syntactic or semantic errors in the constraint syntax, the QM-IC tool displays appropriate messages in human readable form and highlights the involved elements. The use case continues at step 7.</p> <p>10a In case of syntactic or semantic errors in the constraint syntax, the QM-IC tool displays appropriate messages in human readable form and highlights the involved elements. The use case continues at step 9.</p> <p>12a In case of an infeasible infrastructure, the QM-IC tool indicates missing resources and suggests the increase of resources. The use case continues at step 7, 9 or 11.</p> <p>12b In case that overall quality constraints cannot be fulfilled, the QM-IC tool highlights critical parts or critical data flows. The use case continues at step 7, 9 or 11.</p> <p>15a The Pipeline Designer does not acknowledge that the existing pipeline shall be overwritten. In this case, the scenario may continue at step 7, 9, 11 or 13.</p> <p>16a In case of a syntactically, semantically or not validated pipeline, the QM-IC tool informs the Pipeline Designer about the actual status and stores the pipeline for further configuration.</p> <p>16b In case of a physical storage error, the QM-IC tool informs the Pipeline Designer about the failed pipeline repository action.</p>
Business Rules	<ul style="list-style-type: none"> Invalid pipelines cannot be executed on the QualiMaster

	platform. <ul style="list-style-type: none"> Pipeline configurations which exceed the actual resources of the platform and the underlying hardware cannot be executed.
Data/Functions	<ol style="list-style-type: none"> Access to sources and sinks Access to quality parameter Syntactic pipeline analysis Semantic pipeline analysis Static pipeline quality analysis Read and write access to the pipeline repository Access to the pipeline repository Structural display (or visualization) of pipelines including editor

Additional remarks:

In practice, steps 7-10 of the use case scenario may happen interleaved and in an incremental fashion. In the use case scenario, we provided these in a conceptually separated form. In fact, closing the QM-IC tool without storing the pipeline will lead to a warning. Further, step 14 may also be used for storing the modified pipeline as a new pipeline. This corresponds to the definition of a new pipeline (UC-PD1).

7.1.3 Use Case: Delete Pipeline Definition

This use case enables the pipeline designer to delete an existing pipeline. Please note that this step just deletes the definition of the pipeline so that it cannot be (re)instantiated and schedules the instantiated artefacts of the pipeline for (eventual) physical deletion by the Platform Administrator. Further, stopping a running pipeline is a task of the Pipeline Administrator.

Use Case Identifier	UC-PD3
Use Case Name	Delete pipeline definition
Actor	Pipeline Designer
Goal	Delete an existing data stream analysis pipeline.
Precondition	QualiMaster platform is configured (UC-PA10, optionally UC-PA11), data sources and sinks are configured (UC-PA7), quality characteristics are defined (UC-AM1)
Postcondition	Existing pipeline definition is deleted from the pipeline repository
Scenario Sequence	<ol style="list-style-type: none"> The Pipeline Designer starts the QM-IC tool. The QM-IC tool asks for the login credentials. The Pipeline Designer provides his / her credentials. The QM-IC tool grants access, updates the configuration from the repository and shows the existing pipelines. The Pipeline Designer selects an option for deleting an existing pipeline definition and specifies which of the pipelines stored in the pipeline repository shall be deleted. The QM-IC asks the Pipeline Designer whether the pipeline selected in Step 5 shall be deleted. The Pipeline Designer acknowledges the deletion of the pipeline

	<p>selected in Step 5.</p> <p>8. The QM-IC tool deletes the successful deletion of the pipeline selected in Step 5.</p> <p>9. The Pipeline Designer logs out from the QM-IC tool.</p>
Extensions	<p>4a The Pipeline Designer enters wrong credentials and the QM-IC tool refuses the access and queries the credentials again.</p> <p>5a. No pipeline definitions are available for deletion, i.e., the QM-IC tool does not show pipelines for selection and the scenario ends at step 5.</p> <p>6a The Pipeline Designer does not acknowledge that the existing pipeline shall be deleted. In this case, the scenario continues at the pipeline selection of step 5.</p> <p>8a. In case of an access error, the QM-IC tool informs the Pipeline Designer about the failed deletion.</p>
Business Rules	<ul style="list-style-type: none"> Actual pipeline execution requires responsibility about physical compute resources and is, thus, a responsibility of the Platform Administrator.
Data/Functions	<p>a. Access to the pipeline repository</p> <p>b. Structural display (or visualization) of pipelines</p>

7.2 Adaptation Manager

As introduced in Section 2.2, the role of the Adaptation Manager is to define and specify the adaptive behaviour of the QualiMaster infrastructure. In particular, the Adaptation Manager prepares the running QualiMaster platform to act upon the adaptation needs of the system, by small- and large-scale changes, e.g., the tuning a threshold to reduce or increase sensitivity (small scale adaptation), or change of the pipeline altogether or the number of nodes for the processing (large scale adaptation). Given that the real-time response of the QualiMaster platform might need to be smaller than the amount of time needed for large scale pipeline reconfiguration in the presence of special-purpose hardware (e.g., full reconfiguration of the MAX system might take several seconds), the Adaptation Manager will need to assess (in conjunction with the Platform Administrator) the desirability of the adaptation process vis-a-vis the real-time system requirements, as well as optimising the scheduling of workloads onto the reconfigurable hardware to minimise the amount of time spent reconfiguring.

The tasks of the Adaptation Manager are detailed through the following use cases:

- UC-AM1: Define quality characteristics of processing elements
- UC-AM2: Define pipeline quality parameters
- UC-AM3: Define reactive adaptation rules
- UC-AM4: Define proactive adaptation rules
- UC-AM5: Monitor execution of adaptation rules (i.e., reflective adaptation)
- UC-AM6: Change adaptation settings

Some use cases may overlap with the Pipeline Designer (from a domain perspective) or the Platform Administrator (from a resource perspective), in particular UC-AM3 and UC-AM4. However, depending on the actual organization structure, the role of the Adaptation Manager may

also jointly be filled by a Pipeline Designer or a Platform Administrator. As described above, all Infrastructure Users utilize the QM-IC tool.

7.2.1 Use Case: Define Quality Characteristics of Processing Elements

This use case describes how quality parameters and quality characteristics for processing elements are defined. Thereby, the quality parameters to be measured for a processing element (and how the parameters can be measured) as well as relating the quality characteristics (behavior over time determined by analysis or lab measurements of the individual algorithms) to the processing elements are specified. This is the basis for deriving quality characteristics for entire pipelines. We will detail the notion of quality characteristics in future deliverables (in particular D1.2 and D2.1).

Use Case Identifier	UC-AM1
Use Case Name	Define quality parameters of processing elements
Actor	Adaptation Manager
Goal	Define quality and adaptation parameters of data processing elements and methods for measuring them
Precondition	Data processing algorithms are added to the infrastructure (UC-PA3 or UC-PA4) and platform quality parameters are defined (UC-PA1 or UC-PA2).
Postcondition	The description of data processing elements is augmented with quality characteristics that can be taken into account in the quality-driven adaptation process (and the definition of adaptation rules).
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool grants access, updates the configuration from the repository and displays the list of existing data processing elements. 5. The Adaptation Manager selects one of the data processing elements. 6. The QM-IC tool views the description and properties of the selected data processing element. 7. The Adaptation Manager defines/modifies the quality characteristics of the selected processing element providing their description and metrics, e.g., based on measuring the processing element in certain settings. 8. The QM-IC tool validates the provided characteristics, in particular, whether the underlying quality parameters can actually be determined by the QualiMaster infrastructure at runtime. The QM-IC tool saves the changed/defined quality characteristics. 9. The Adaptation Manager defines how individual quality parameters can be measured in terms of low-level monitoring functionalities provided by the QualiMaster platform. Further, derived (calculated) quality parameter calculated can be specified based on already defined quality parameters, in particular those determined by low-level measurements.

	<p>10. The QM-IC tool validates the provided measurement methods, e.g., whether actual implementations of the methods are provided by the QualiMaster platform.</p> <p>11. The Adaptation Manager requests saving the definitions.</p> <p>12. The QM-IC tool saves the changes and definitions.</p> <p>13. The Adaptation Manager logs out from the tool.</p>
Extensions	<p>4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>4b Data processing elements displayed as groups of families/clusters that share the same functionality.</p> <p>7a Quality characteristics are defined for a whole family of processing elements and apply for all members of the family.</p> <p>8a If quality characteristics for already known quality parameters are missing, the QM-IC issues a warning that adaptation may not be able to take these parameters into account.</p> <p>8b If quality characteristics for unknown quality parameters are specified, the QM-IC issues a warning that the underlying quality parameters must be specified or measured.</p> <p>9a If no measurement methods can be defined, the Adaptation Manager may specify constant values, such as quality levels for each individual member of the family.</p> <p>10a Akin to 8a and 8b</p> <p>12a If saving the information fails, the QM-IC informs the Adaptation Manager by an error message.</p>
Business Rules	<ul style="list-style-type: none"> Quality parameters must either be measured at runtime or defined by the Adaptation manager If measurement of quality parameters is not defined, the adaptation may ignore the related quality parameters. If quality characteristics for quality parameters are not defined, the adaptation may ignore these quality parameters.
Data/Functions	<p>a. Access to the data processing elements repository (including metadata).</p> <p>b. Formatting of data processing element information such as listing or grouping.</p> <p>c. Access to the description of derived quality measurement methods.</p> <p>d. Access to the configuration of the QualiMaster platform.</p>

7.2.2 Use Case: Define Pipeline Quality Characteristics

This use case complements UC-AM1 (Section 7.2.1) by defining how to derive the quality characteristics of entire data analysis pipelines based on the characteristics of the constituting processing elements. Basically, we aim at end-to-end quality characteristics, but also quality characteristics for pipeline parts might be needed. Thus, we refer in this use case to both options using the term “(end-to-end) quality characteristics”, but we will detail this use case based on further discussions in future deliverables (in particular D1.2 and D4.1).

Use Case Identifier	UC-AM2
Use Case Name	Define pipeline quality characteristics
Actor	Adaptation Manager
Goal	Define methods for measuring/estimating the (end-to-end) quality characteristics (including costs) of a pipeline by propagating quality, error and cost of individual data processing elements through the pipeline.
Precondition	Pipelines are defined (UC-PD1, UC-PD2) and the quality characteristics of the individual data processing elements are available (UC-AM1)
Postcondition	The end-to-end quality and cost of the pipeline can be measured/estimated
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool displays the list of existing pipelines 5. The Adaptation Manager selects one pipeline. 6. The QM-IC tool displays the structure of the processing elements of the selected pipeline. 7. The Adaptation Manager defines or modifies the methods for measuring/estimating the (end-to-end) quality characteristics (including costs) of the pipeline depending on the quality characteristics of the individual processing elements. 8. The QM-IC tool validates the methods, e.g., whether the required quality characteristics of the individual data processing elements are available to calculate the (end-to-end) pipeline characteristics. 9. The Adaptation Manager requests saving the definitions. 10. The QM-IC tool saves the changes 11. The Adaptation Manager logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 4b The QM-IC tool highlights those processing elements without complete specification of quality characteristics. 8a The QM-IC tool detects inconsistencies, e.g., missing individual quality characteristics of the processing elements and emits a warning. The use case continues at step 7. 8b The QM-IC tool identifies missing measurement methods for quality parameters and emits a warning. 10a The QM-IC tool cannot store the related elements and informs the Adaptation Manager about the error. The use case continues at step 7.
Business Rules	<ul style="list-style-type: none"> • Only consistently specified (end-to-end) quality characteristics

	can be propagated and determined.
Data/Functions	<ul style="list-style-type: none"> a. Access to the Pipeline Repository. b. Access to the data Processing Elements Repository. c. Structural display (or visualization) of pipelines. d. Access to the software artefact repository to store/modify implemented quality measurement methods so that the infrastructure instantiation process can integrate them.

7.2.3 Use Case: Define Reactive Adaptation Rules

In this use case, the specification of the actual behavior of the reactive adaptation is described. Reactive adaptation focuses on the detection of certain triggers and to quickly perform adaptation changes without extensive planning or predictions of quality characteristics. The specification of predictive adaptation rules will be described as an extended use case in Section 7.2.4.

The specification of the reactive adaptation is based on the quality parameters and characteristics introduced by UC-AM1 and UC-AM2. As mentioned above, the Adaptation Manager role may also be filled by the Platform Designer or the Platform Administrator, each from his/her specific (domain vs. resource) view. The aim of this use case is to specify the adaptive behaviour and the boundaries of the adaptation space rather than implementing it in a fixed way. This enables adjusting, modifying or evolving the adaptation behaviour without digging into its actual implementation (which will be derived as one step during the platform instantiation, see UC-PA14). Please note that we use the generic term “adaptation rule” to denote the elements of a (declarative) adaptivity specification. We will detail this use case as well as the notion of adaptation rules in future deliverables (in particular D1.2 and D4.1)

Use Case Identifier	UC-AM3
Use Case Name	Define reactive adaptation rules
Actor	Adaptation Manager
Goal	Define quality-driven adaptation rules to be implemented by the adaptation module at run-time on the pipeline level
Precondition	Processing pipelines exist (UC-PD1, UC-PD2) and methods for measuring/estimating their (end-to-end) quality characteristics are defined (UC-AM2)
Postcondition	Define quality-driven adaptation rules to be considered and executed by the adaptation module at run-time on the pipeline level.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool grants access and displays the existing pipelines. 5. The Adaptation Manager selects the pipeline for the definition of adaptation rules. 6. The QM-IC tool displays the actual adaptation rules for the selected pipeline. 7. The Adaptation Manager defines (or modifies) a set of quality parameters (such as data load, velocity or resources

	<p>consumption, quality impact on the pipeline) to be monitored.</p> <p>8. The QM-IC tool validates the parameters, thresholds and tradeoffs against the available measurement methods for quality parameters.</p> <p>9. The Adaptation Manager defines (or modifies) the set of reactive adaptation rules (including thresholds and tradeoffs for the quality parameters) to be executed by the adaptation module.</p> <p>10. The QM-IC tool validates the rules against the parameters specified in step 7 as well as the rules for potential inconsistencies.</p> <p>11. The Adaptation Manager requests saving the definitions.</p> <p>12. The QM-IC tool saves the changes.</p> <p>13. The Adaptation Manager logs out from the QM-IC tool.</p>
Extensions	<p>4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>4b The QM-IC tool highlights those processing elements without complete specification of quality characteristics.</p> <p>8a Validation fails as underspecified quality parameters shall be used. An error message is displayed to the Adaptation Manager and the use case continues at step 7.</p> <p>10a Validation fails as underspecified quality parameters shall be used. An error message is displayed to the Adaptation Manager and the use case continues at step 7.</p> <p>10b Validation fails as inconsistencies or cyclic dependencies have been specified in the reactive rules and the use case continues at step 5.</p> <p>12a Saving the changes fails for some reasons so that the QM-IC informs the Adaptation Manager in terms of an error message. The use case continues at step 7.</p>
Business Rules	<ul style="list-style-type: none"> Inconsistent or invalid adaptation rules shall not be turned into an implementation or considered at runtime.
Data/Functions	<p>a. Access to the Pipeline Repository</p> <p>b. Access to the Processing Elements Repository</p> <p>c. Adaptivity rule validation</p>

7.2.4 Use Case: Define Proactive Adaptation Rules

This use case is actually an extension to the definition of proactive adaptation rules (UC-AM3), as in addition to the adaptation rules also mechanisms to predict quality parameters and characteristics must be specified. In order to keep the use cases readable, we did not describe these extensions within UC-AM3, but provide an extended description in this section.

Use Case Identifier	UC-AM4
Use Case Name	Define proactive adaptation rules
Actor	Adaptation Manager

Goal	Define quality-driven adaptation rules to be implemented by the adaptation module at run-time on the pipeline level
Precondition	Processing pipelines exist (UC-PD1, UC-PD2) and methods for measuring/estimating their (end-to-end) quality characteristics are defined (UC-AM2)
Postcondition	Define quality-driven adaptation rules to be considered and executed by the adaptation module at run-time on the pipeline level.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool grants access and displays the existing pipelines. 5. The Adaptation Manager selects the pipeline for the definition of adaptation rules. 6. The QM-IC tool displays the actual adaptation rules for the selected pipeline. 7. The Adaptation Manager defines (or modifies) a set of quality parameters (such as data load, velocity or resources consumption, quality impact on the pipeline) to be monitored and their thresholds and tradeoffs that should trigger a reactive adaptation. 8. The QM-IC tool validates the parameters, thresholds and tradeoffs against the available measurement methods for quality parameters. 9. The Adaptation Manager defines methods for predicting individual quality parameters and characteristics, e.g., in terms of software components. 10. The QM-IC tool validates that the data provided by the prediction methods fits to the related quality parameter or characteristic. 11. The Adaptation Manager defines (or modifies) the set of proactive adaptation rules to be executed by the adaptation module in response to the triggers (such as adjusting the filtering and sampling of data or switching to an alternative execution path of the pipeline, etc.) 12. The QM-IC tool validates the rules against the parameters specified in step 7 as well as the rules for potential inconsistencies. 13. The Adaptation Manager requests saving the definitions. 14. The QM-IC tool saves the changes. 15. The Adaptation Manager logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 8a Validation fails as underspecified quality parameters shall be used. An error message is displayed to the Adaptation Manager. The use case continues at step 7.

	<p>10a Validation fails as for example the data types of the prediction mechanism and the quality parameters do not match or cannot be converted. The use case continues at step 9.</p> <p>12a Validation fails as underspecified quality parameters shall be used. An error message is displayed to the Adaptation Manager. The use case continues at step 9.</p> <p>12b Validation fails as inconsistencies or cyclic dependencies have been specified in the predictive rules or the plan derivation. The use case continues at step 9.</p> <p>14a Saving the changes fails for some reasons so that the QM-IC tool informs the Adaptation Manager in terms of an error message. The use case continues at step 9.</p>
Business Rules	<ul style="list-style-type: none"> • Inconsistent or invalid adaptation rules or specifications for plan derivation shall not be turned into an implementation or considered at runtime. • Inconsistent prediction mechanisms shall not lead to runtime failures.
Data/Functions	<ol style="list-style-type: none"> a. Access to the Pipeline Repository b. Access to the Processing Elements Repository c. Adaptivity rule validation d. Prediction component validation e. Predictive plan validation f. Access to the software artefact repository to store/modify prediction components so that the infrastructure instantiation process can integrate them.

7.2.5 Use Case: Monitor Execution of Adaptation Rules

This use case aims at the identification of improvement potential or problems caused by the adaptation of the pipeline execution (reflective adaptation with human-in-the-loop). Therefore, the QualiMaster infrastructure will provide detailed logs on the executed adaptation actions and the QM-IC tool supports the Adaptation Manager in reviewing and analyzing the log results.

Use Case Identifier	UC-AM5
Use Case Name	Monitor execution of adaptation rules
Actor	Adaptation Manager
Goal	The adaptation manager monitors the execution of the adaptation rules and their impact to identify any needs for adjustments.
Precondition	Adaptation rules are defined (UC-AM3, UC-AM4) and logs of execution history of adaptations are provided by the QualiMaster infrastructure.
Postcondition	Needs for adaptation adjustments or improvements are identified
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool grants access and displays the existing

	<p>pipelines.</p> <ol style="list-style-type: none"> 5. The Adaptation Manager selects the pipeline to monitor the execution of the adaptation for. 6. The QM-IC tool retrieves the adaptation execution logs from the QualiMaster infrastructure and performs reflective analysis on the actual log, historical log information and past analyses. 7. The Adaptation Manager reviews the logs and the analysis results and identifies needs for modifications/extensions of the reactive or proactive adaptation rules and marks them for modification (continuing at UC-AM3 step 2 or UC-AM4 step 2, respectively). 8. The Adaptation Manager logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 6a Retrieving the execution log fails for some reason. The QM-IC tool informs the Adaptation manager accordingly. The use case continues at step 4. 6b The reflective analysis does not identify any issues or improvement potential. Then the Adaptation Manager may continue with a manual analysis. 7a The Adaptation Manager does not identify any needs for changes. Then the use case ends at step 7.
Business Rules	
Data/Functions	<ol style="list-style-type: none"> a. Access to Adaptation Execution Logs (including a sufficient level of details for quality parameters, characteristics, predictions, plans etc) through the QualiMaster infrastructure. b. Reflective adaptation analysis

7.2.6 Use Case: Change adaptation settings

As part of discussions with stakeholders, SPRING identified that enabling the Adaptation Manager to change the “adaptation rules” at full level of detail may not be appropriate in every application settings. Thus, we extend the initial set of use cases by the opportunity to change high-level adaptation settings rather than modifying deep “adaptation rules”. Experienced Adaptation Managers may request deeper level access to the “adaptation rules”, e.g., through view settings. The original use cases UC-AM3 and UC-AM4 may then enable access to various levels of “rules”, e.g., ranging from high-level rules to the most detailed view.

Use Case Identifier	UC-AM6
Use Case Name	Change adaptation settings
Actor	Adaptation Manager
Goal	Influence the adaptation on high level, i.e., without explicit knowledge about the “adaptation rules”.

Precondition	Methods for measuring/estimating (end-to-end) quality characteristics are defined (UC-AM2)
Postcondition	High-level adaptation settings are changed.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Adaptation Manager starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Adaptation Manager provides his / her credentials. 4. The QM-IC tool grants access and displays among others the option to change high-level adaptation settings. 5. The Adaptation Manager selects the option to change high-level adaptation settings. 6. The QM-IC tool displays the actual high-level adaptation settings. 7. The Adaptation Manager changes individual settings such as the relative weighting of quality parameters. 8. The QM-IC tool validates the input. 9. The Adaptation Manager requests saving the definitions. 10. The QM-IC tool saves the changes. 11. The Adaptation Manager logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Adaptation Manager enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 8a Validation fails as invalid weights have been entered. An error message is displayed to the Adaptation Manager and the use case continues at step 6. 10a Saving the changes (and potentially enacting them) fails for some reasons so that the QM-IC tool informs the Adaptation Manager in terms of an error message. The use case continues at step 6.
Business Rules	Inconsistent or invalid high-level adaptation settings shall not be turned into an implementation or considered at runtime.
Data/Functions	<ol style="list-style-type: none"> a. Access to the Pipeline Repository c. Adaptivity rule validation

7.3 Platform Administrator

The Platform Administrator, as introduced in Section 2.2, will setup, install and maintain the QualiMaster infrastructure. In particular, the platform administrator is responsible for the configuration, and (if needed) reconfiguration of hardware used for the execution of the data processing pipelines, including the configuration of special-purpose hardware. This task entails knowledge of the availability of resources as well as quantitative performance aspects (in collaboration with the Adaptation Manager) in order to properly configure the QualiMaster platform. The Administrator also uses the QM-IC tool. In the remainder of this section, the related use cases are described:

- UC-PA1: Define platform quality parameter (as basis for quality constraints)

- UC-PA2: Modify platform quality parameter
- UC-PA3: Add data processing algorithm
- UC-PA4: Modify data processing algorithm
- UC-PA5: Add hardware-based algorithm
- UC-PA6: Modify hardware-based algorithm
- UC-PA7: Configure data sources and sinks
- UC-PA8: Start pipeline
- UC-PA9: Stop pipeline
- UC-PA10: Configure QualiMaster platform for software-based execution
- UC-PA11: Configure QualiMaster platform for hardware-based execution
- UC-PA12: Start platform
- UC-PA13: Stop platform
- UC-PA14: Instantiate platform
- UC-PA15: Monitor the execution

As the tasks of the Platform Administrator may affect running pipelines, in most cases an explicit approval of the ultimate execution is required.

7.3.1 Use Case: Define Platform Quality Parameters

The QualiMaster platform will be able to monitor low-level quality parameters, such as data stream related measurements (e.g., tuples/throughput per second, etc.) or resource consumption (e.g., execution time or memory consumption). This use case is about defining the quality parameters (considering that implementing them in some cases may require manual modifications of the QualiMaster platform). However, measuring certain quality parameters may influence the performance of the execution, in particular, if unnecessary measurements are performed. Therefore, the infrastructure instantiation process (UC-PA14) will take care of those quality parameters actually used in the definition of processing families (UC-PD1, UC-PD-2) or pipelines (UC-PD1, UC-PD-2) and disable or even eliminate unused measurements.

Use Case Identifier	UC-PA1
Use Case Name	Define platform quality parameters
Actor	Platform Administrator
Goal	Define the (low-level) quality parameters for the QualiMaster infrastructure and the methods to measure them in the platform.
Precondition	QualiMaster platform is configured (UC-PA10, optionally UC-PA11)
Postcondition	The quality parameters are defined and can be used for the specification of quality constraints by the Adaptation Manager.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and displays the configuration options. 5. The Platform Administrator selects the list of existing quality parameters.

	6. The QM-IC tool displays the existing quality parameters. 7. The Platform Administrator defines a new quality parameter in terms of its metadata and its implementation (either as a component or by modification of the QualiMaster infrastructure). 8. The QM-IC tool validates the input and saves the changes 9. The Platform Administrator logs out from the QM-IC tool.
Extensions	4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 6a Quality parameters are categorized in groups/clusters that share the same functionality. 7a No new quality parameter is required. The use case stops here. 8a Validation fails and the platform administrator is informed about the related reason. The use case continues at step 3.
Business Rules	
Data/Functions	a. Access to the (metadata) of the platform quality parameters. b. Display of quality parameters as list or as groups c. Access to the software artefact repository to store/modify monitoring components so that the infrastructure instantiation process can integrate and the infrastructure can execute them.

7.3.2 Use Case: Modify Platform Quality Parameters

In addition to the definition of (new) platform quality parameters (UC-PA1), platform parameters may need to be modified, e.g., disabled or a more recent implementation shall be configured. However, due to possible references from the pipeline families, the pipelines and the adaptation specification, the use case does not support the deletion of quality parameters rather than disabling them. Changes to the quality parameters may require a review of the referring elements and, thus, explicit approval.

Use Case Identifier	UC-PA2
Use Case Name	Modify platform quality parameters
Actor	Platform Administrator
Goal	Define the (low-level) quality parameters for the QualiMaster infrastructure and the methods to measure them in the platform.
Precondition	Platform quality parameters are defined (UC-PA1)
Postcondition	The quality parameters are defined and can be used for the specification of quality constraints.
Scenario Sequence	1. The Platform Administrator starts the QM-IC tool 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and lists the configuration options. 5. The Platform Administrator selects the list of existing quality

	<p>parameters.</p> <ol style="list-style-type: none"> 6. The QM-IC tool displays the existing quality parameters. 7. The Platform Administrator changes the metadata of a quality parameter, in particular in terms of enabling / disabling individual quality parameters or changing the underlying implementation. Disabling a quality parameter requires review of the quality characteristics (UC-AM1) or the processing pipeline definitions (UC-PD2). 8. The QM-IC tool validates the input and saves the changes. 9. The Platform Administrator approves the new algorithm(s). 10. The QM-IC tool acknowledges the approval. 11. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 6a Quality parameters are categorized in groups/clusters that share the same functionality. 7a No new quality parameter is required. The use case stops here. 8a Validation fails and the Platform Administrator is informed about the related reason. The use case continues at step 7. 10a No approval happens so that the new parameters will not be considered for actual execution until explicit approval. Then the use case ends here.
Business Rules	<ul style="list-style-type: none"> • Quality parameters shall not be deleted, just disabled and thus shall not be available to the upper level layers. • Modifications of existing quality parameters require explicit approval as running pipelines may be affected.
Data/Functions	<ol style="list-style-type: none"> a. Access to the (metadata) of the platform quality parameters. b. Display of quality parameters as a list or as groups. c. Access to the software artefact repository to store/modify monitoring components so that the infrastructure instantiation process can integrate and the infrastructure can execute them. d. Approval mechanism.

7.3.3 Use Case: Add Data Processing Algorithm

As a basis for defining quality characteristics (UC-AM1), data processing algorithms must be known to the QualiMaster platform. In this use case, the Platform Administrator is enabled to incorporate the algorithms in terms of their implementation. Currently, the specification of the quality characteristics (UC-AM1) is conceptually separated from the introduction of a data processing algorithm (this use case). However, further work may lead to an integrated approach. One example is to specify (default) quality characteristics with the algorithm, e.g., in terms of source code annotation or a manifest file. This may for example simplify the specification and maintenance of quality characteristics in UC-AM1 and even of adding new data processing algorithms.

Use Case Identifier	UC-PA3
Use Case Name	Add data processing algorithm
Actor	Platform Administrator
Goal	Include a new data processing algorithm in the QualiMaster infrastructure.
Precondition	The QualiMaster platform is configured (UC-PA10, optionally UC-PA11).
Postcondition	The new data processing algorithm is incorporated and can be used as part of a data processing family.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the data processing algorithm view. 6. The QM-IC tool displays the already specified data processing algorithms. 7. The Platform Administrator selects that a new algorithm shall be added. 8. The QM-IC tool enables the Platform Administrator to define the metadata for the algorithm. 9. The Platform Administrator provides the metadata, such as the containing algorithm family, the name of the algorithm or its inputs and outputs (or “unstructured”) and specifies the actual implementation (e.g., JAR) of the specific algorithm. 10. The QM-IC tool validates the input. 11. The Platform Administrator requests storing the new algorithm. 12. The QM-IC tool saves the metadata, places the implementation into the processing algorithm repository and request for approving the changes. 13. The Platform Administrator approves the new algorithm(s). 14. The QM-IC tool acknowledges the approval. 15. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 9a The QM-IC tool takes over the metadata from source code annotations or a manifest. 10a Validation fails, e.g., as the implementation is not accessible, not compliant to the algorithm family or metadata is missing. Then the Platform Administrator is informed and the use case continues at step 7. 12a Saving the metadata or the implementation fails. Then the use case continues at step 7. 12b No QualiMaster platform is running so that no approval is needed. The use case stops here. 14a No approval happens so that the new algorithms will not be considered for actual execution until explicit approval. Then the

	use case ends here.
Business Rules	<ul style="list-style-type: none"> Only valid data processing algorithms shall be made available to the upper level layers of the QualiMaster infrastructure. New data processing algorithms become only available to the platform if actually needed and due to explicit approval of the Platform Administrator to ensure consistency of the execution.
Data/Functions	<ol style="list-style-type: none"> Access to the Data Processing Algorithm Repository. This implies access to the software artefact repository. Algorithm approval mechanism.

7.3.4 Use Case: Modify Data Processing Algorithm

In addition to the definition of (new) data processing algorithms (UC-PA3 in Section 7.3.3), processing algorithms may need to be modified, e.g., disabled or a more recent implementation shall be used. However, due to possible references from the pipeline families, the use case does not support the deletion of processing algorithms. Instead, it enables disabling a processing algorithm so that it is not available anymore, but requires a review of the referring elements as not an algorithm family may be empty.

Use Case Identifier	UC-PA4
Use Case Name	Modify data processing algorithm
Actor	Platform Administrator
Goal	Modify an existing data processing algorithm in the QualiMaster infrastructure.
Precondition	The QualiMaster platform is configured (UC-PD10, optionally UP-PD11) and data processing algorithms are specified (UC-PA3).
Postcondition	The modified data processing algorithm is incorporated and can be used as part of a data processing family.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the data processing algorithm view. 6. The QM-IC tool displays the already specified data processing algorithms. 7. The Platform Administrator selects the algorithm that shall be modified. 8. The QM-IC tool enables the Platform Administrator to define the metadata for the algorithm. 9. The Platform Administrator provides the changed metadata such as whether the algorithm is disabled, its name or its updated implementation (e.g., JAR). 10. The QM-IC tool validates the input.

	<p>11. The Platform Administrator requests storing the changed algorithm.</p> <p>12. The QM-IC tool saves the metadata and places the implementation into the processing algorithm repository and requests for explicit approval.</p> <p>13. The Platform Administrator approves the changed algorithm(s).</p> <p>14. The QM-IC tool acknowledges the approval.</p> <p>15. The Platform Administrator logs out from the QM-IC tool.</p>
Extensions	<p>4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>8a The QM-IC tool takes over the metadata from source code annotations or a manifest.</p> <p>10a Validation fails, e.g., as the implementation is not accessible or metadata is missing. Then the Platform Administrator is informed accordingly and the use case continues at step 7.</p> <p>12a Saving the metadata or the implementation fails. Then the use case continues at step 7.</p> <p>12b No QualiMaster platform is running so that no approval is needed. The use case stops here.</p> <p>13a No approval happens so that the changed algorithms will not be considered for actual execution until explicit approval. Then the use case ends here.</p>
Business Rules	<ul style="list-style-type: none"> Only valid data processing algorithms shall be made available to the upper level layers of the QualiMaster infrastructure. New data processing algorithms become only available to the platform if actually needed and due to explicit approval of the Platform Administrator to ensure consistency of the execution.
Data/Functions	<p>a. Access to the Data Processing Algorithm Repository. This implies access to the Software Artefact Repository.</p> <p>b. Algorithm approval mechanism.</p>

7.3.5 Use Case: Add Hardware-based Data Processing Algorithm

Hardware-based data processing algorithms can transparently be used in algorithm families. However, in order to avoid unnecessary overhead, arbitrary switches between software-based and hardware-based algorithms shall be avoided. Further, the platform instantiation process (UC-PA14) may assemble multiple (alternative) hardware-based algorithms for the execution on one reconfigurable hardware unit, e.g., a data flow engine. Please note, that in contrast to software-based algorithms, hardware-based algorithms have a clearly determined performance behavior if once laid out for a specific reconfigurable hardware. Also here, a integration with the quality characteristics (UC-AM1), e.g., in terms of source code annotations or manifest may be considered in future. Please note that the instantiation of the composition of hardware-based algorithms and their actual layout towards a specific hardware unit is considered during the platform instantiation process.

Use Case Identifier	UC-PA5
Use Case Name	Add hardware-based data processing algorithm

Actor	Platform Administrator
Goal	Include a new hardware-based data processing algorithm in the QualiMaster infrastructure.
Precondition	The QualiMaster platform is configured (UC-PD10, optionally UC-PD11).
Postcondition	The new data processing algorithm is incorporated and can be used as part of a data processing family.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and displays the configuration options. 5. The Platform Administrator selects the data processing algorithm view. 6. The QM-IC tool displays the already specified data processing algorithms. 7. The Platform Administrator selects that a new hardware-based algorithm shall be added. 8. The QM-IC tool enables the Platform Administrator to define the metadata for the algorithm. 9. The Platform Administrator provides the metadata such as its name, the containing family or its inputs and outputs (or “unstructured”) and specifies the actual implementation (e.g., JAR) of the specific algorithm. 10. The QM-IC tool validates the input. 11. The Platform Administrator requests storing the new algorithm. 12. The QM-IC tool saves the metadata, places the implementation into the processing algorithm repository and requests for explicit approval. 13. The Platform Administrator approves the changed algorithm(s). 14. The QM-IC tool acknowledges the approval. 15. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 9a The QM-IC tool takes over the metadata from source code annotations or a manifest. 10a Validation fails, e.g., as the implementation is not accessible, it does not comply with the family or (hardware-based) metadata is missing. Then the use case continues at step 8. 12a Saving the metadata or the implementation fails. Then the Platform Administrator is informed and the use case continues at step 8. 12b No QualiMaster platform is running so that no approval is needed. The use case stops here.

	13a No approval happens so that the changed algorithms will not be considered for actual execution until explicit approval. Then the use case ends here.
Business Rules	<ul style="list-style-type: none"> Only valid data processing algorithms shall be made available to the upper level layers of the QualiMaster infrastructure. New data processing algorithms become only available to the platform if actually needed and due to explicit approval of the Platform Administrator to ensure consistency of the execution. Once a hardware-based algorithm is laid out for a certain reconfigurable hardware, it can be executed only on that hardware. In turn, the quality performance parameters are then known for that specific hardware.
Data/Functions	<ol style="list-style-type: none"> Access to the Data Processing Algorithm Repository. This implies access to the Software Artefact Repository. Algorithm approval mechanism.

7.3.6 Use Case: Modify Hardware-based Data Processing Algorithm

Akin to UC-PA4, also hardware-based processing algorithms may be modified or disabled (but actually not deleted). Please note that the instantiation of the composition of hardware-based algorithms and their actual layout towards a specific hardware unit is considered by the platform instantiation process (UC-PA14).

Use Case Identifier	UC-PA6
Use Case Name	Modify hardware-based data processing algorithm
Actor	Platform Administrator
Goal	Modify an existing hardware-based data processing algorithm in the QualiMaster infrastructure.
Precondition	The QualiMaster platform is configured (UC-PD10, optionally UC-PD11) and hardware-based data processing algorithms are specified (UC-PA5).
Postcondition	The modified hardware-based data processing algorithm is incorporated and can be used as part of a data processing family.
Scenario Sequence	<ol style="list-style-type: none"> The Platform Administrator starts the QM-IC tool. The QM-IC tool asks for the login credentials. The Platform Administrator provides his / her credentials. The QM-IC tool grants access and shows the configuration options. The Platform Administrator selects the data processing algorithm view. The QM-IC tool displays the already specified data processing algorithms. The Platform Administrator selects the hardware-based algorithm that shall be modified. The QM-IC tool enables the Platform Administrator to define the metadata for the hardware-based algorithm.

	<p>9. The Platform Administrator provides the changed (hardware-specific) metadata such as whether the algorithm is disabled, its name or its updated implementation (e.g., JAR).</p> <p>10. The QM-IC tool validates the input.</p> <p>11. The Platform Administrator requests storing the changed algorithm.</p> <p>12. The QM-IC tool saves the metadata, places the implementation into the processing algorithm repository and asks for explicit approval.</p> <p>13. The Platform Administrator approves the changed algorithm(s).</p> <p>14. The QM-IC tool acknowledges the approval.</p> <p>15. The Platform Administrator logs out from the QM-IC tool.</p>
Extensions	<p>4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>8a The QM-IC tool takes over the metadata from source code annotations of the provided implementation.</p> <p>10a Validation fails, e.g., as the implementation is not accessible or metadata is missing. Then the use case continues at step 8.</p> <p>12a Saving the metadata or the implementation fails. Then the use case continues at step 8.</p> <p>12b No QualiMaster platform is running so that no approval is needed. The use case stops here.</p> <p>13a No approval happens so that the changed algorithms will not be considered for actual execution until explicit approval. Then the use case ends here.</p>
Business Rules	<ul style="list-style-type: none"> Only valid data processing algorithms shall be made available to the upper level layers of the QualiMaster infrastructure. New data processing algorithms become only available to the platform if actually needed and due to explicit approval of the Platform Administrator to ensure consistency of the execution. Once a hardware-based algorithm is laid out for a certain reconfigurable hardware, it can be executed only on that hardware. In turn, the quality performance parameters are then known for that specific hardware.
Data/Functions	<p>a. Access to the Data Processing Algorithm Repository. This implies access to the Software Artefact Repository.</p> <p>b. Algorithm approval mechanism.</p>

7.3.7 Use Case: Configure Pipeline Sources and Sinks

Conceptually, we separate the definition of data analysis pipelines (on domain level) and the technical specification of pipeline sources and sinks (on platform / resource level). In particular, sources and sinks may need credentials, imply network access restrictions or require an adapter implementation to bring the data from an arbitrary source into the QualiMaster data stream processing platform. Technically, sources and sinks may be considered as an (abstract or draft)

pipeline specification, which is then refined by the Pipeline Designer use cases (UC-PD1 or UC-PD2).

Use Case Identifier	UC-PA7
Use Case Name	Configure pipeline sources and sinks
Actor	Platform Administrator
Goal	Define the technical information about data sources and sinks such as IP addresses, credentials, adapters, etc.
Precondition	
Postcondition	Draft pipeline source and sink definition has been created and successfully stored.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the pipeline source and sink definitions. 6. The QM-IC tool displays the configured sources and sinks. 7. The Platform Administrator selects that a new source shall be entered. 8. The QM-IC tool displays the input for the metadata. 9. The Platform Administrator now enters the metadata, e.g., the internet address, the credentials, the adaptor implementation realizing the integration of the physical source/sink with the QualiMaster infrastructure (e.g., in terms of a JAR file) or the data structure of the associated data stream (or “unstructured”). 10. The QM-IC tool validates the provided information. 11. The Platform Administrator requires saving the information entered above. 12. The QM-IC tool acknowledges that the source / sink has been successfully stored and requests for explicit approval. 13. The Platform Administrator approves the changed source(s) or sink(s). 14. The QM-IC tool acknowledges the approval. 15. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 5a The Platform Administrator selects that a new sink shall be entered. 5b The Platform Administrator selects the existing source that shall be modified. 5c The Platform Administrator selects the existing sink that shall be modified.

	<p>6a If step 5b or 5c was executed before, the already configured metadata for the selected data source or sink is displayed for editing.</p> <p>7a In case of sinks (step 5a or 5c) the Platform Administrator may enter the network addresses for which access shall be granted (or permitted).</p> <p>10a Validation fails, e.g., due to missing required information or as the source or sink cannot be accessed through the network. Then the Platform Administrator is notified and the use case continues at step 9.</p> <p>12a Saving the information fails for some reason and the Platform Administrator is notified by the QM-IC tool accordingly. Then the use case continues at step 9.</p> <p>12b No QualiMaster platform is running so that no approval is needed. The use case stops here.</p> <p>13a No approval happens so that the changed algorithms will not be considered for actual execution until explicit approval. Then the use case ends here.</p>
Business Rules	<ul style="list-style-type: none"> Access limitations to data sources or sinks may apply, e.g., in case of licensed (financial) data or to separate customer groups. Invalid source or sink information shall not be made available to a running QualiMaster platform.
Data/Functions	<p>a. Access to the Pipeline Repository.</p> <p>b. Source and sink validation including network access.</p> <p>c. Access to the software artefact repository to store/modify adapters so that the infrastructure instantiation process can integrate them.</p>

7.3.8 Use Case: Start Pipeline

When a pipeline is completely specified ranging from required quality characteristics over algorithms, algorithm families, adaptation up to the pipeline data flow and the infrastructure instantiation process (UC-PA14) has derived the related software artifacts, the pipeline is ready for execution on the QualiMaster infrastructure. Akin to the acknowledgement of changes, this is currently considered as an explicit task, as compute resources will be allocated and existing pipelines may (potentially) be affected.

Use Case Identifier	UC-PA8
Use Case Name	Start pipeline
Actor	Platform Administrator
Goal	Start a pipeline that has been fully configured, stored and instantiated for execution with the QualiMaster platform.
Precondition	The platform is running (UC-PA12) and pipelines are configured (UC-PD1 or UC-PD2).
Postcondition	The pipeline is deployed, started and being executed adaptively.
Scenario Sequence	1. The Platform Administrator starts the QM-IC tool.

	<ol style="list-style-type: none"> 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and displays the configuration options. 5. The Platform Administrator selects the pipelines. 6. The QM-IC tool displays the configured pipelines. 7. The Platform Administrator selects the pipeline he/she wants to start. 8. The QM-IC tool verifies that it can start the selected pipeline (using the quality impact analysis also accessible to the Platform Designer as described in Section 7.1) and asks the Platform Administrator whether the selected platform shall actually be started now. 9. The Platform Administrator approves the start of the pipeline. 10. The QM-IC tool starts the pipeline on the QualiMaster platform and acknowledges that the pipeline was successfully started. 11. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 8a The QM-IC tool cannot verify that the selected pipeline can be started, e.g., as the execution quality of running pipelines would be affected. The Platform Administrator is informed and the use case continues at step 6. 9a The Platform Administrator does not approve the action and so the use case stops here. 10a The deployment or physical execution fails for some reason. Then the Platform Administrator is notified accordingly.
Business Rules	<ul style="list-style-type: none"> • Only valid, instantiated and executable pipelines can be started. • The start of pipelines shall not affect the execution of already running pipelines.
Data/Functions	<ol style="list-style-type: none"> a. Pipeline quality analysis b. Access to the Pipeline Repository c. Access to the lower level QualiMaster infrastructure, e.g., deployment and start of pipelines

Additional remarks:

In an implementation of the QualiMaster infrastructure, separate tasks for deploying and undeploying a pipeline in addition to starting and stopping might be needed. However, this actually depends on the capabilities of the quality impact analysis, which could determine the most appropriate subset of machines to run a pipeline on and, thus, would make an explicit (un)deployment superfluous.

7.3.9 Use Case: Stop Pipeline

This is the counterpart use case of UC-PA8, i.e., to explicitly stop a running pipeline.

Use Case Identifier	UC-PA9
----------------------------	---------------

Use Case Name	Stop pipeline
Actor	Platform Administrator
Goal	Stop a pipeline that has been already started in QualiMaster platform.
Precondition	QualiMaster platform is running (UC-PD12) and pipelines are configured (UC-PD1 or UC-PD2)
Postcondition	The selected pipeline is stopped.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the pipelines. 6. The QM-IC tool displays the configured pipelines. 7. The Platform Administrator selects the pipeline he/she wants to stop. 8. The QM-IC tool verifies that the pipeline can be stopped without affecting other pipelines and asks the Platform Administrator for final approval. 9. The Platform Administrator approves the action. 10. The QM-IC tool stops the selected pipeline through the QualiMaster platform and acknowledges the Platform Administrator about successfully stopping the selected pipeline. 11. The Platform Administrator logs out from the QM-IC tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 8a The QM-IC tool cannot verify the selected pipeline can be stopped, e.g., as the execution quality of running pipelines would be affected. Then the Platform Administrator is informed and the use case continues at step 6. 9a The Platform Administrator does not approve the action and so the use case stops here. 10a Stopping physical execution or undeployment fails for some reason. Then the Platform Administrator is notified accordingly.
Business Rules	<ul style="list-style-type: none"> • Only running pipelines may be stopped, but stopping a pipeline shall not affect other running pipelines.
Data/Functions	<ol style="list-style-type: none"> a. Pipeline quality analysis b. Access to the Pipeline Repository c. Access to the lower level QualiMaster infrastructure, e.g., stopping and undeploying pipelines

Additional remarks:

In an implementation of the QualiMaster infrastructure, separate tasks for deploying and undeploying a pipeline in addition to starting and stopping might be needed. However, this actually depends on the capabilities of the quality impact analysis, which could determine the most appropriate subset of machines to run a pipeline on and, thus, would make an explicit (un)deployment superfluous.

7.3.10 Use Case: Configure QualiMaster Platform for Software-based Execution

This step is required for bootstrapping the QualiMaster Platform for a certain execution environment or when new hardware becomes available. This includes the configuration of the underlying standard hardware, e.g., their physical resources or the numbers of threads to be used for pipeline execution. Please note that we describe the configuration of hardware-based execution in UC-PA10.

Use Case Identifier	UC-PA10
Use Case Name	Configure QualiMaster Platform for Software-based Execution
Actor	Platform Administrator
Goal	Configure the standard hardware for software-based execution of data flow pipelines.
Precondition	Appropriate hardware is available
Postcondition	The hardware is configured and the QualiMaster platform is configured for software-based execution.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the hardware view. 6. The QM-IC tool displays the configured hardware. 7. The Platform Administrator selects to add new servers. 8. The QM-IC tool allows entering information about the hardware. 9. The Platform Administrator enters the information, e.g., its physical resources, its network identification or the number of threads to be used for data processing pipeline execution. 10. The QM-IC tool validates the input. 11. The Platform Administrator requires storing the information. 12. The QM-IC tool acknowledges the successful storage, starts the platform instantiation process and finally asks for approval. 13. The Platform Administrator approves the changes. 14. The QM-IC tool acknowledges the enactment of the changes. 15. The Platform Administrator logs out from the tool.
Extensions	<ol style="list-style-type: none"> 4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 7a The Platform Administrator selects to change or delete information on an individual server. 8a In case of 7a, the QM-IC tool displays the already configured information. 10a The validation fails due to missing or inconsistent data. Then the Platform Administrator is informed and the use case continues at step 7. 12a Storage of the data fails and the administrator is notified accordingly. The use case continues at step 7. 12b Platform instantiation process fails and the administrator is notified accordingly. The use case continues at step 7. 12c Actually no instance of the QualiMaster platform is running (bootstrapping) so that an approval is not needed, but the

	<p>administrator is informed on how to install and start the platform (UC-PD12).</p> <p>13a The Platform administrator does not approve the changes so that they do not become effective immediately. Then the use case stops here.</p> <p>13b In case of 12c, 13b is not executed.</p>
Business Rules	<ul style="list-style-type: none"> Only consistently configured instances of the QualiMaster platform may be used for pipeline execution. In case of modifying hardware for a running QualiMaster platform an explicit approval is required, e.g., to ensure that the new hardware is actually switched on.
Data/Functions	<ol style="list-style-type: none"> Access to hardware information repository Access to already running QualiMaster platforms Validation of hardware information

7.3.11 Use Case: Configure QualiMaster Platform for Hardware-based Execution

This task is required for bootstrapping the QualiMaster Platform for a certain execution environment or when new reconfigurable hardware becomes available. This includes the configuration of the type and the amount of reconfigurable hardware units (e.g., MAX Data Flow Engines) and how to access them (e.g., through a host computer). While software-based execution is required, UC-PA10 is optional depending on whether reconfigurable hardware is available.

Use Case Identifier	UC-PA11
Use Case Name	Configure QualiMaster Platform for Hardware-based Execution
Actor	Platform Administrator
Goal	Configure the standard hardware for software-based execution of data flow pipelines.
Precondition	Supported reconfigurable hardware is available
Postcondition	The reconfigurable hardware is specified and the QualiMaster Platform is configured accordingly.
Scenario Sequence	<ol style="list-style-type: none"> The Platform Administrator starts the QM-IC tool. The QM-IC tool asks for the login credentials. The Platform Administrator provides his / her credentials. The QM-IC tool grants access and shows the configuration options. The Platform Administrator selects the hardware view. The QM-IC tool displays the configured hardware. The Platform Administrator selects to add new reconfigurable hardware. The QM-IC tool allows entering information about the reconfigurable hardware. The Platform Administrator enters the information, e.g., it's the number of types of Data Flow Engine boards as well as the host computer used for accessing and controlling the Data Flow

	<p>Engines.</p> <p>10. The QM-IC tool validates the input.</p> <p>11. The Platform Administrator requires storing the information.</p> <p>12. The QM-IC tool acknowledges the successful storage, starts the platform instantiation process and finally asks for approval.</p> <p>13. The Platform Administrator approves the changes.</p> <p>14. The QM-IC tool acknowledges the enactment of the changes.</p> <p>15. The Platform Administrator logs out from the tool.</p>
Extensions	<p>4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>7a The Platform Administrator selects to change or delete information on a specific Data Flow Engine board or an entire engine cluster through its host computer.</p> <p>8a In case of 7a, the QM-IC tool displays the already configured information.</p> <p>10a The validation fails due to missing or inconsistent data. Then the Platform Administrator is informed and the use case continues at step 9.</p> <p>12a Storage of the data fails and the administrator is notified accordingly. The use case continues at step 9.</p> <p>12b Platform instantiation process fails and the administrator is notified accordingly. The use case continues at step 9.</p> <p>12c Actually no instance of the QualiMaster platform is running (bootstrapping) so that an approval is not needed, but the administrator is informed on how to install and start the platform (UC-PA12).</p> <p>13a The Platform administrator does not approve the changes so that they do not become effective immediately. Then the use case stops here.</p> <p>13b In case of 12c, 13b is not executed.</p>
Business Rules	<ul style="list-style-type: none"> • Hardware-based execution requires a configuration for software-based execution (UC-PD10) • Only consistently configured instances of the QualiMaster platform may be used for pipeline execution. • In case of modifying reconfigurable hardware for a running QualiMaster platform an explicit approval is required, e.g., to ensure that the new hardware is actually switched on.
Data/Functions	<p>a. Access to hardware information repository</p> <p>b. Access to already running QualiMaster platforms</p> <p>c. Validation of reconfigurable hardware information</p>

7.3.12 Use Case: Start QualiMaster Platform

This is the low-level task to start up a configured QualiMaster Platform during bootstrapping or after maintenance. Currently, we expect that therefore a low-level (shell) command must be issued on one of the servers.

Use Case Identifier	UC-PA12
Use Case Name	Stop QualiMaster Platform
Actor	Platform Administrator
Goal	Start a QualiMaster Platform.
Precondition	The QualiMaster Platform is configured (UC-PA9, optionally UC-PA10) and installed.
Postcondition	The QualiMaster Platform is started
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator accesses the coordinator server and issues the QualiMaster platform startup command. 2. The startup command displays that the QualiMaster platform has successfully been started.
Extensions	2a Errors prevent startup and the Platform Administrator is notified accordingly.
Business Rules	<ul style="list-style-type: none"> • Only consistently configured instances of the QualiMaster platform can be started.
Data/Functions	a. Startup script

7.3.13 Use Case: Stop QualiMaster Platform

This is the low-level task to stop a configured QualiMaster Platform, e.g., during a maintenance interval. Currently, we expect that this task can be performed through the QM-IC tool.

Use Case Identifier	UC-PA13
Use Case Name	Stop QualiMaster Platform
Actor	Platform Administrator
Goal	Stop the QualiMaster Platform.
Precondition	The QualiMaster Platform is running (UC-PA12).
Postcondition	The QualiMaster Platform is stopped
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and shows the configuration options. 5. The Platform Administrator selects the platforms view. 6. The QM-IC tool displays the running platforms. 7. The Platform Administrator selects the platform to be stopped and issues the stop command. 8. The QM-IC tool asks for explicit approval.

	<p>9. The Platform Administrator approves the stop command.</p> <p>10. The QM-IC tool stops the selected platform and the Platform Administrator is acknowledged about successfully stopping the selected platform.</p> <p>11. The Platform Administrator logs out from the tool.</p>
Extensions	<p>4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.</p> <p>8a At least one running pipeline prevents stopping the platform. The Platform Administrator is informed and the use case continues at step 6.</p> <p>9a The Platform Administrator does not approve the stop command. In this case, the use case stops here.</p> <p>10a The QualiMaster platform cannot be stopped due to technical reasons. In the extreme case, the Platform Administrator must enter the control server and stop the platform using a low-level command (akin to UC-PA10).</p>
Business Rules	<ul style="list-style-type: none"> Stopping the platform requires that no pipelines are running, i.e., stopping a QualiMaster platform may lead to the forced stop of pipelines only in very extreme cases.
Data/Functions	<p>a. Access to the platform functionality</p>

7.3.14 Use Case: Instantiate Platform

This use case aims at turning the generic QualiMaster platform into an instantiated one. This includes the instantiation of missing or changed artifacts based on the configuration. The Platform Administrator is responsible for this task as in particular the instantiation of hardware layouts may consume compute resources (on specific build servers).

Use Case Identifier	UC-PA14
Use Case Name	Instantiate Platform
Actor	Platform Administrator
Goal	Turn the generic QualiMaster platform into an instantiated one based on the configuration (models)
Precondition	The required configuration information is provided, at least the hardware information (UC-PA10, UC-PA11). Further artefacts can only be created if configuration information is available through the appropriate infrastructure use cases.
Postcondition	The artefacts described by the configuration are instantiated. If a complete configuration is provided, the platform and the configured pipelines are ready for installation / deployment / execution.
Scenario Sequence	<ol style="list-style-type: none"> 1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and displays the configuration options. 5. The Platform Administrator selects the instantiation view.

	6. The QM-IC tool displays the status of the configuration. 7. The Platform Administrator selects 'instantiate platform'. 8. The QM-IC tool instantiates the platform. 9. The Platform Administrator logs out from the QM-IC tool.
Extensions	4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4. 6a Important configuration information is missing so that the platform cannot be instantiated, i.e., the configuration is inconsistent The Platform Administrator is informed and the use case stops. 7a The Platform instantiation fails for technical reasons. Information about the failure is displayed and the use case stops here.
Business Rules	<ul style="list-style-type: none"> Only consistently configured platforms shall be executed.
Data/Functions	a. Access to all repositories containing configuration information or implementation components. b. Access to the software artefact repository to store the generic QualiMaster platform as a source and to produce the instantiated platform.

7.3.15 Use Case: Monitor Execution

This use case aims at providing the Platform Administrator with actual information of the running platform, such as the number of actual Storm utilized worker nodes or DFEs. The QM-IC tool displays this information as a kind of internal administration cockpit. In addition, the Platform Administrator may also directly utilize the high-level web interfaces of the individual Execution Systems (if applicable).

Use Case Identifier	UC-PA15
Use Case Name	Monitor Execution
Actor	Platform Administrator
Goal	Provide runtime information about the running platform(s)
Precondition	At least one platform is configured and running (UC-PA12).
Postcondition	
Scenario Sequence	1. The Platform Administrator starts the QM-IC tool. 2. The QM-IC tool asks for the login credentials. 3. The Platform Administrator provides his / her credentials. 4. The QM-IC tool grants access and displays the configuration options. 5. The Platform Administrator selects the runtime view as well as the platform to be displayed. 6. The QM-IC tool connects to the platform and displays the actual execution status, e.g., in terms of time series charts. 7. The Platform Administrator logs out from the QM-IC tool.

Extensions	4a The Platform Administrator enters wrong credentials and the QM-IC tool refuses the access. Continue at step 4.
Business Rules	<ul style="list-style-type: none"> Only consistently configured platforms shall be executed.
Data/Functions	<ul style="list-style-type: none"> a. Access to all repositories containing configuration information or implementation components. b. Access to the software artefact repository to store the generic QualiMaster platform as a source and to produce the instantiated platform.

7.4 Component Providers

7.4.1 Algorithm Provider

An Algorithm Provider communicates with the Infrastructure Users in order to deliver or integrate new algorithms. Thus, an Algorithm Provider does not require additional use cases.

7.4.2 Data Provider

A Data Provider communicates with the Infrastructure Users in order to deliver data. Thus, a Data Provider does not require additional use cases. In fact, the discussion with the Data Provider about terms and licenses typically may lead to Service Level Agreements. If important, e.g., to justify and communicate external failures to the Application Users, the Infrastructure Administrator may represent quantifiable SLAs as terms of data source constraints in the QualiMaster configuration.

7.4.3 Application Provider

An Application Provider communicates with the Infrastructure Users in order to develop new Applications. Thus, the role of the Algorithm Provider does not lead to new use cases, neither on Application User side nor on Infrastructure User side.

8 Application Testing

In WP1 there has already been initial activity for the application testing. We define application testing as the testing of the QualiMaster platform from the user point of view. This section provides an early discussion and examples on the application testing activities already done. It will be considered as input for WP6 which is responsible for the entire application testing process. Section 8.1 describes the preparation of the data sets while section 8.2 provides some examples of application testing.

8.1 Preparation of Data Sets

8.1.1 Financial Data Sets

SPRING provides financial data sets. Via the SPRING's real time API, data sets for more than 1200 market players are provided. These data sets have 2 components:

- Streaming real time quotes and delayed quotes from the financial markets over world major exchanges.
- Streaming continuous real time market depth data for each market player. This data is a very realistic simulation of real market depth data. As providing real market depth data for more than 1200 market players is very expensive, simulated data is provided.

For final tests of the QualiMaster system, the market depth data will be switched to real market data for a limited time frame.

This data is fully representative for the different use cases (which need sub sets of this data). As this data is provided during the full project runtime, also different market conditions (stress in the markets, high volatility) are covered.

The selection of the market player took into account a wide geological range as well as a wide range of market player type. The set of market players contains the heaviest traded market players worldwide to take into account high data load for QualiMaster.

The market players in detail are:

- 150 currency pairs
- Major stocks from Europe, USA, Canada, China, Hong Kong
- Exchange traded funds
- World Indices
- Futures on Indices
- Futures on Bonds

The approach on providing data sets for QualiMaster is flexibility on request. If there are additional requirements during project lifetime or requests from stakeholders, data sets can easily be modified and extended.

8.1.2 Social Web Data Sets

Several social media and online news streams are considered in QualiMaster. In addition to RSS feeds of popular news agencies, QualiMaster will collect and evaluate streams of popular and domain-independent micro-blogging platforms such as Twitter as well as domain-specific blogging and micro-blogging platforms such as SeekingAlpha and StockTwits, respectively. In addition, news related social media such as Yahoo News and News Channels on YouTube will be considered and evaluated as a source for detecting and monitoring social interactions and reactions to news events. For a more detailed overview of relevant social web data sources, refer to deliverable D2.1 of WP2.

Already since early 2013, LUH started collecting and archiving the sample Twitter stream that is published via Twitter's public Streaming API. During the first year of the project, additional "crawling" threads have been started using Twitter's filtered Streaming API, in order to retrieve a more focused collection of data that is relevant for the financial domain. Using the filtered Streaming API, only tweets that contain certain keywords (such as stock names) are collected. This ensures that the recall of relevant tweets increases in comparison to the recall of the (un-filtered) Streaming API. In addition, micro-blogs from the social platform StockTwits, which focuses on discussions on stock trading, have been collected.

For the purpose of building a test collection of social web datasets, to be used in combination with the financial datasets, collaboration between WP1 and WP2 started to select interesting financial events that occurred in 2014 and had significant impact on the financial market and perform a focused crawling of Twitter datasets around these events. The goal of this activity is to use these datasets to test the effectiveness of processing and analysis methods that will be developed in WP2 in estimating, detecting, or predicting the impact of these events on the financial markets.

8.2 Application Test Examples

The tests of the application designer and the applications will have 2 phases.

- The first phase is processed to ensure that the software works bug free and that provides good performance. Use and handling need to be easy and self-explainable. These tests will be performed by the technical partners of QualiMaster. Each tester follows predefined test cases. The test cases are built with focus on functionality and processor loading.
- The second phase is processed by possible stakeholders. From the survey (section 3), several participants indicated interest on collaboration and testing of the QualiMaster platform. The testers will be asked to first follow predefined storylines and then follow their own, intuitive storylines. They will have an evaluation sheet, where they can rank their impressions.

Test Phase 1: Technical tests, following story lines

- a. Application Designer
 - i. Testing Scenarios
 - 1. User management (Login/ Logout/ rights management)
 - 2. Application design handling
 - a. Create new app
 - b. Modify existing app
 - c. Delete app
 - ii. Metrics
 - 1. Functionality (Benchmark: bug free)
 - 2. Performance (Benchmark: no overload)
 - 3. Usability (Score of 4 from 5 points)
- b. Applications
 - i. Testing scenarios
 - 1. User management (Login/ Logout/ rights management)
 - 2. Use case dependent storylines (see tables below)
 - ii. Metrics
 - 1. Functionality (Benchmark: bug free)
 - 2. Performance (Benchmark: no overload)
 - 3. Usability (Score of 4 from 5 points)

Test Phase 2: Target user Acceptance test, customer story line

- A. Application designer and applications
 - a. Testing scenarios
 - i. Assisted use following predefined storylines
 - ii. Free and intuitive use
 - b. Metrics
 - i. Ease of use (Score of evaluation sheet)

- ii. Intuitive workflow (Score of evaluation sheet)
- iii. Arrangement/ visualisation of results (Score of evaluation sheet)
- iv. Comprehensibility of results (Score of evaluation sheet)

The tables below identify test cases to target potential performance / usability bottlenecks arising in the use cases. The tests are defined per components which are used by the applications.

Use case identifier	Performance-Related Testing
Targets	Dependency Metric Techniques
Data-set	Average and worst case scenario data sets per each metric.
Goal	Identify performance of dependency metric calculation in the average and worst case scenarios.
Example	A data set to calculate a 5000x5000 Hayoshi-Yoshida correlation matrix across a 10 minutes window.

Use case identifier	Performance-Related Testing
Targets	Centrality Metric Techniques
Data-set	Average and worst case scenario data sets per each metric.
Goal	Identify performance of dependency metric calculation in the average and worst case scenarios.
Example	A data set to calculate a 5000x5000 Hayoshi-Yoshida correlation matrix across a 10 minutes window.

Use case identifier	Performance-Related Testing
Targets	Graph analysis Techniques
Data-set	Average and worst case scenario data sets per each metric.
Goal	Identify performance of centrality metric calculation in the average

	and worst case scenarios.
Example	A data set to calculate a betweenness centrality for 10,000 node directed graph.

Use case identifier	Performance-Related Testing
Targets	3.2.3 functionalities
Data-set	Average and worst case scenario data sets per each functionality. As functionalities can be combined, identify the longest most computationally intensive chain. This is going to be simultaneous calculation of multiple statistics on a transformed domain along with ongoing regression.
Goal	Identify performance of functionality calculation in the average and worst case scenarios.
Example	A data set to calculate a Gaussian process regression on 100,000 points transformed using wavelets. Simultaneous Gaussian process regression with a computationally expensive kernel.

Use case identifier	Usability Testing / Performance Testing
Targets	Graph Visualizations
Data-set	Identify critical scenarios and prepare data sets required to recreate them.
Goal	Identify graph rendering and responsiveness of the GUI.
Example	<p>A 10,000 node directed graph with graph coloring according to its centrality measure. Graph subcomponents are indicated.</p> <ol style="list-style-type: none"> 1. Responsiveness to drastic threshold change 2. Responsiveness to dependency metric change 3. Responsiveness to centrality metric change 4. Responsiveness to time progression. How does the visualization behave as new data arrives.

Use case identifier	Usability Testing / Performance Testing
Targets	Plot Visualizations
Data-set	Identify critical scenarios and prepare data sets required to recreate them.
Goal	Identify plot rendering and responsiveness of the GUI.
Example	<p>A 100,000 point plot transformed using wavelets with gaussian process regression.</p> <ol style="list-style-type: none"> 5. Responsiveness to regression change 6. Responsiveness to time-frequency domain change 7. Responsiveness to statistics calculation 8. Responsiveness to time progression. How does the plot visualization behave as new data arrives.

9 Conclusions and Outlook

In this deliverable, we described the consolidated requirements collection for the QualiMaster applications and the QualiMaster infrastructure. The document integrates the initial requirements and use cases collected in D1.1 and the ones collected during the refinement activity. We have determined and detailed the actors that will interact with QualiMaster. In particular, we identified two groups of actors, namely application users and infrastructure users, which are actually a subset of the stakeholders interested in the QualiMaster project. We then detailed the user-centric view on the QualiMaster application infrastructure for the financial domain in terms of use cases for institutional financial clients and regulators (detailing the business domains / application scenarios in the DoW). Subsequently, we discussed the requirements for the data streams to be processed and the algorithm families to be provided by an instantiated QualiMaster platform for the financial domain. Finally, we described the use cases for the three infrastructure users, namely, the Pipeline Designer, the Adaptation Manager and the Platform Administrator. The deliverable also provided an initial discussion on the application testing.

The requirements described in D1.2 will be taken into account during the whole design and development phase of the QualiMaster platform. It is also important to underline that part of the activities reported in D1.2 will be taken over in WP6, in particular:

- **Visualization:** D1.2 reported on the requirements of the QualiMaster visualization system based on the QualiMaster user survey. The deliverable also illustrated some example of possible user interfaces for the various QualiMaster applications. The activity on visualization will be taken over and completed in Task 6.3 “User interface and visualization”.
- **Application testing:** D1.2 reported on the preliminary activity done for application testing, namely testing the QualiMaster platform from the user point of view. The deliverable mentioned the work related to the preparation of the financial and social web data sets and it reported some examples of application testing. This activity will be taken over and completed in Task 6.1 “Preparation of Evaluation and Data Sets” and in Tsk 6.4 “Automated and Expert Evaluation”.

References

- [1] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Systemic risk and stability in financial networks. *NBER Working Paper*, (18727), January 2013.
- [2] D. M. Berry, E. Kamsties, and M. M. Krieger. From contract drafting to software specification: Linguistic sources of ambiguity—a handbook. Technical report, 2003.
- [3] J. Birch. A short history of correlation networks research, 2012. <http://www.fna.fi/blog/2012/07/06/a-short-history-of-correlation-networks-research/>.
- [4] S. Chakravarthy and Q. Jiang. *Stream Data Processing: A Quality of Service Perspective Modeling, Scheduling, Load Shedding, and Complex Event Processing*. Springer Publishing Company, Incorporated, 1st edition, 2009.
- [5] A. Cockburn. *Writing Effective Use Cases*. Addison Wesley, 2000.
- [6] D. de Almeida Ferreira and A. R. da Silva. A controlled natural language approach for integrating requirements and model-driven engineering. In *International Conference on Software Engineering Advances (ICSEA '09)*, pages 518–523, Sept 2009.
- [7] H. Eichelberger and K. Schmid. A Systematic Analysis of Textual Variability Modeling Languages. In *International Software Product Line Conference (SPLC '13)*, pages 12–21, 2013.
- [8] European Central Bank. Recent advances in modeling systemic risk using network analysis. Technical report, January 2010.
- [9] W. A. Freiwald, P. Valdes, J. Bosch, R. Biscay, J. C. Jimenez, L. M. Rodriguez, V. Rodriguez, A. K. Kreiter, and W. Singer. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *J Neurosci Methods*, 94:105–119, 1999.
- [10] S. Frølung and J. Koistinen. QML: A Language for Quality of Service Specification. Technical Report HPL98-10, HP Labs, HP Software Technologies Laboratory, February 1998.
- [11] V. Gervasi and D. Zowghi. Reasoning about inconsistencies in natural language requirements. *ACM Trans. Softw. Eng. Methodol.*, 14(3):277–330, July 2005.
- [12] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
- [13] T. Hayashi and N. Yoshida. Estimating correlations with nonsynchronous observations in continuous diffusion models. Technical report, 2006. Working paper.
- [14] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches to time series analysis. *Physics Reports*, 441:1–46, 2007.
- [15] A. Kaiser and T. Schreiber. Information transfer in continuous processes. 110:43–62, 2002.
- [16] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown. A Granger Causality Measure for Point Process Models of Ensemble Neural Spiking Activity. *PCBI*, March 2011. DOI: 10.1371/journal.pcbi.1001110.
- [17] F. van der Linden, K. Schmid, and E. Rommes. *Software Product Lines in Action - The Best Industrial Practice in Product Line Engineering*. Springer, 2007.
- [18] Y. Liu and M. T. Bahadori. A survey on granger causality: A computational view, October 2012. <http://www-bcf.usc.edu/liu32/granger.pdf>.
- [19] R.N. Mantegna. Hierarchical structure in financial markets. *Eur. Phys. J. B*, 11:193–197, 1999.

- [20] C. Marchetti, B. Pernici, and P. Plebani. A Quality Model for Multichannel Adaptive Information Systems. In *International World Wide Web Conference (WWW '04)*, pages 48–54.
- [21] V. Mattiussi, M. Tumminello, G. Iori, and R. N. Mantegna. Comparing correlation matrix estimators via Kullback-Leibler divergence. Technical report. <http://ssrn.com/abstract=1966714>.
- [22] Joseph G. Nicholas. HFR Global Hedge Fund Industry Report, 2013. <https://www.hedgefundresearch.com/?fuse=products-irglo>.
- [23] K. Pohl, G. Böckle, and F. van der Linden. *Software Product Line Engineering: Foundations, Principles, and Techniques*. Springer, 2005.
- [24] INDENICA project consortium. Open Variability Modeling Approach for Service Ecosystems, 2012. <http://www.indenica.eu> [validated: February 2014].
- [25] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85, 2000.
- [26] A. Seth. Granger causality, 2007. http://www.scholarpedia.org/article/Granger_causality.
- [27] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379—423, 1948.
- [28] Twitter. Public apis, 2012. <https://dev.twitter.com/docs/streaming-apis/streams/public>.
- [29] Twitter. Tweets - field guide, 2013. <https://dev.twitter.com/docs/platform-objects/tweets>.
- [30] D. A. Vaughan. Selected definitions of "hedge fund", May 2003. Comments for the U.S. Securities and Exchange Commission Roundtable on Hedge Funds, <https://www.sec.gov/spotlight/hedgefunds/hedge-vaughn.htm>.
- [31] CBOE: White Paper – The CBOE Volatility Index VIX, 2014
- [32] N.Huth, and F. Abergel, *High Frequency Lead/lag Relationships Empirical facts*. January 2012, available online <http://xxx.tau.ac.il/pdf/1111.7103.pdf>.
- [33] http://www.ms.u-tokyo.ac.jp/~nakahiro/mypapers_for_personal_use/hayyos03.pdf
- [34] <https://hal.archives-ouvertes.fr/hal-00851429/document>
- [35] <https://fp7.portals.mbs.ac.uk/Portals/59/docs/KNPapers/S.%20STRAETMANS.pdf>
- [36] <http://math.bu.edu/people/murad/pub/hist11-posted.pdf>
- [37] <https://web.cse.msu.edu/~cse842/Papers/CoverThomas-Ch2.pdf>
- [38] Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Phys. Rev. E* 71, 066208
- [39] <http://bechtel.colorado.edu/~balajir/my-papers/physical-review-E.pdf>
- [40] “Estimation of Mutual Information: A Survey”, Janett Walters-Williams and Yan Li