Financial Use Case:
Initial Requirements and Scenario Definitions
Work Package 2 Task 2.1 Deliverable 2.1

Authors
Holger Arndt, Stefan Burkard
Spring Techno

Distribution list:

Group:	Others:
WP Leader: Spring Techno Task Leader: Spring Techno	Internal Reviewer: Athena INFORE Management Team INFORE Project Officer

Document history:

Revision	Date	Section	Page	Modification
0.5	19/02/2019	1-6	1-18	Creation
0.6	28/02/2019	Appendix A Appendix B	18-25	Questionnaire and Expert User Responses Added
0.7	13/03/2019	All	All	Internal Reviewer Comments and Edits
0.8	14/03/2019	All	All	Internal Review Comments Incorporated
1.0	26/03/2019	-	-	Final Version

Approvals:

First Author:	Holger Arndt (Spring Techno)	Date:	28/02/2019
Internal Reviewer:	Nikos Giatrakos (Athena)	Date:	14/03/2019
Coordinator:	Antonios Deligiannakis (Athena)	Date:	26/03/2019

Table of contents:

1 Executive Summary 4
2 Introduction 5
2.1 Requirements Collection Approach 5
2.2 Objectives Overview 6
2.3 Structure of the Deliverable 6
3 Key Role Players of the INFORE Financial User Scenario 7
4 Requirements 8
4.1 Data and Data Stream Requisites 8
4.2 Requisites for Supported Algorithms 8
4.2.1 Systemic Risk Use Case 8
4.2.2 Forecast and Decision-Support Use Cases 10
4.3 Platform Quality Requisites 11
4.3.1 Quality Dimensions 11
4.3.2 Initial Requisites 12
5 Financial Area User Scenarios 13
5.1 Analysis of the Expert User Interviews and Questionnaires 13
5.2 Overview of the User Scenarios 13
5.3 Details on the User Scenarios 13
5.3.1 Forecasting Price Swings 13
5.3.2 Systemic Risk Prediction and Early Warning 14
5.3.3 Decision-Support for Investment Opportunities 15
6 Conclusions and Outlook 17
Appendix: Expert User Questionnaires 18
Appendix A: Original questions from the questionnaire 18
Appendix B: Anonymised answers of the expert users on the questions from Appendix A 20

1 Executive Summary

Deliverable D2.1 reports on the results of the requirements' collection for INFORE's Financial use case and INFORE's user scenarios for forecasting price swings, systemic risk prediction and decision support for investment opportunities. In particular, extracting actionable insights from the markets' behavior, spotting profitable investments based on trading instruments' volatility and seasonal price patterns, as well as the in-depth analysis of market risk are the use case scenarios from the INFORE Description of Work (DoW - see WP2), which are studied at a deeper level through the expert user interviews. Derived from the expert interviews, this deliverable details the role of players who interact with the financial applications. Furthermore, individual user scenarios and their specific requirements and interactions with INFORE are summarized. On top of this, the deliverable provides preconditions for the data streams to be processed by INFORE and the algorithms to be applied to price swing prediction, risk calculation, and detection of investment opportunities.

2 Introduction

In this document, user-requirements for the INFORE financial use case scenario are collected and summarized. We provide the basis for the further practical development of the INFORE Financial use case tools and components, considering the interactions of this use case and of the financial data streams it produces with INFORE's algorithms and the overall architecture. An enhanced version of the requirements collected in the project will be documented in Deliverable D2.2, which is due on Month 18 of the project.

For collecting the requirements, a questionnaire and corresponding expert user interviews have been processed and their contents are included in Appendix A and Appendix B, respectively. This course of action enables a practical, user-centered starting point for setting out the development of INFORE's Financial applications. As a first step, a set of user scenarios has been identified and presented.

The main definitions of the initial use cases consist of the identification of user requirements and the structure of the use case workflow.

2.1 Requirements Collection Approach

SPRING is active in software development and analysis for the financial and investment industry and has identified three main finance industry user scenarios that will benefit from the advanced forecasting and risk analysis capabilities envisioned by INFORE. The workflows and requisite details are supported by the interviews based on questionnaires given out to 3 area experts, who gave insights of their work and needs, that lead to concrete user scenarios reflecting the user needs in conjunction with the project objectives. The utilized questionnaire and responses from expert users are included in Appendix A and Appendix B, respectively.

User scenarios are a useful method for collecting requisites in a user-centered way. Starting from a set of role players, i.e., a set of persons or other systems that interact with the system under consideration, user scenarios describe the interaction of those role players with the system. The advantages of user scenarios are that they are intuitive and easy to handle due to their textual form. Besides that, they do not only support the description of the normal flow of interaction, which helps in the identification of required system functionalities, but in an advanced form (which will be used in the upcoming Deliverable D2.2 in Month 18), they also support the description of exceptional cases, which provides a more in-depth view of the expected system specifications.

For the documentation of the user scenarios, tables have been used. Each table contains:

- the user scenario name and a unique identifier,
- the involved role players,
- the goal of the user scenario,
- the user-specific requirements, derived from the expert interviews on the execution of the user scenario,
- the scenario description (interaction steps) for successful execution of the user scenario (typically, the role player and the system alternate in their interaction).

2.2 Objectives Overview

The INFORE project will validate its results with respect to the Financial domain in terms of three objectives: Forecasting price swings, Systemic risk prediction with early warning and Decision support for investment opportunities. The user scenarios for the financial domain will be discussed in more detail in Section 5.

Forecasting price swings is essential for almost all categories of trading. A swing forecasting the rising price of an asset would initiate a setup to buy it. A swing forecasting falling prices would lead to sell the asset, or open a short position on it (a 'bet' on falling prices). In both cases, forecasting price changes may indicate an opportunity, and lets investors act accordingly. Forecasting in the financial markets is as old as the markets themselves and several forecasting approaches have been developed, mainly based on two different, concurrent analysis aspects:

- The technical analysis follows the approach of building forecasts based on the past market data, assuming trends and patterns exist in the time series of asset prices.
- The fundamental analysis sets its focus on economic data, earnings, announcements and other market driving aspects.

The user scenario challenges revolve around the development of novel methods to extract market patterns. INFORE's distributed machine learning algorithms are to extract the market patterns, while INFORE's complex event forecasting component will then foresee their apparition. These components together with the processing power of a High Performance Computing (HPC) infrastructure and/or Big Data platforms would enable the analysis of wide sectors of assets in real time, thus expecting to achieve a competitive advantage compared to classical methods.

Systemic risk is identified as one of the major key factors for the stability of financial markets. It represents the situation where financial markets, and the exposure of financial institutions to those markets, become strongly dependent to each other, potentially leading to industry-wide market failure. The need to measure and identify systemic risk is one of the most challenging issues facing institutional users today. This is so because of the given velocity in financial markets and it has forced the need to rapidly identify and act on hotspots before contagion sets in. It is well known that nowadays financial markets are correlated to a degree significantly higher than in the past, and not simply within asset classes. Due to the risk-oriented nature of actual trading, wide ranges of asset classes have become strongly correlated, so that diversification of risk is more challenging.

The use case scenario targets institutional financial clients, such as Banks, Hedge Funds or Asset Managers. In this scenario, the challenge is to perform pre-trading risk analysis and real-time, real money trading risk analysis. The implemented algorithms will be integrated with the trading applications of SPRING and enhance them by a multivariate and multi-market risk analysis approach, which is expected to add important insights about systemic risks, and help avoiding fatal losses of the managed capital. To enable an early warning in case of rising systemic risk, INFORE needs to analyze systemic risk continuously and with small delay (near real time).

Finally, INFORE aims at providing investors with **support for their investment decisions**. This support is based on identifying new investment opportunities and analyzing risk on already existing investment ideas.

2.3 Structure of the Deliverable

The rest of this deliverable is structured into four sections. Section 3 introduces users or key role players interacting with the INFORE financial user scenario and the different parts of INFORE's architecture. Section 4 describes the non-functional requirements focusing on data and data streams, analysis algorithms and platform quality. Section 5 describes the financial user scenarios. Section 6 presents some conclusions from the requirements collection process and outlines the next steps in WP2.

3 Key Role Players of the INFORE Financial User Scenario

The application users are the role players that interact with the visual analytic tools to be built on top of INFORE's architecture. They access the system with a task at hand (e.g., risk analysis of a certain asset) and use the INFORE' tools to perform this task.

Users do not need to know how the underlying INFORE platform is configured or developed. The respective visualization tool 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 role players for INFORE in the group of the financial users are:

- Hedge Funds
- Investment Companies
- Investment Banks

We detail these role players in the table below:

Financial user scenario role players	
Role player	**Hedge Funds**
Taking investment decisions over the hedge fund portfolio is a key task of a fund manager. For a successful hedge fund, it is required to achieve positive returns that are largely uncorrelated with the overall markets. This so called "alpha" is mostly generated by using niche strategies, which gain an edge in the markets. Furthermore, the strategy of risk management is of foremost importance. INFORE will be designed to provide the Hedge Funds with a complete set of tools to achieve these advantages by forecasting price swings, identifying investment opportunities and managing systemic and portfolio risk. Hedge funds can use it for portfolio optimization or risk management and hedging.	
Role player	**Investment Company**
An Investment Company usually sets up its own exchange traded investment vehicles like e.g. Exchange-traded funds ("ETFs") and invests the pooled capital of its investors there. From a trading related point of view, Investment Companies have similar requisites as Hedge Funds. However, they often have even stricter rules regarding trading style and risk management. With the trading vehicles of Investment Companies, there is a differentiation between closed-end mutual funds and open-end mutual funds. While open-end funds buy back shares from investors every business day, closed-end funds usually offer shares to the public only once, when created. Public trading is later possible on a stock exchange.	
Role player	**Investment Bank**
Investment Banks manage enterprise-wide risk across a wide range of asset types. This has become a major regulatory requisite 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 recognized, 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, it will be better able to manage such exposures.	

Table 1: Description of role players for the financial user scenarios

4 Requirements

4.1 Data and Data Stream Requisites

The INFORE system must support streaming data sources, as well as downloadable historical financial data. The requisites originating from these data sources are discussed in the following paragraphs.

The INFORE applications will be built on two types of data sources. Real-time and near real-time data from the financial area is the main source. With respect to this, INFORE 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 INFORE are international interest rates. The financial data is provided by SPRING through a specific API to the project partners.

The streaming data (each asset's data comes as a separate stream) is available in the following format:
"Date, Time, Price, Volume" for each data tick.

In addition to the real-time data, the INFORE system will incorporate static data sources. This is historical financial data, collected for up to 50 years, depending on the availability of market players (available through SPRING). The historical data of INFORE has a similar format and characteristics as the real-time data and covers most market players. INFORE needs to be able to batch process the past data to provide historical time series of predictions.

This historical data is available in different formats:

- Stored as "Tick by Tick" data, in the order that the data arrived from the data streams in the past. The format of this data is "Date, Time, Price, Volume".
- As time-based-compressed data, where ticks are condensed to time frames of 1 minute, 30 minute, 60 minute, daily and so on. The format of this data is "Date, Time, Open, High, Low, Close, Volume". "Date" and "Time" stands for the end of a specified time period, e.g. 5 minutes. "Open" represents the first price occurred in this time frame, "High" stands for the highest price, "Low" stands for the lowest price and "Close" for the last prices of the specified time frame. "Volume" represents the summarized asset volume for the specified time period.

Initial requisites: From the data stream background introduced by the sections above, we summarize the following requisites in the style of the controlled natural language (possible with additional information in natural language). INFORE Financial application must support:

- multiple, concurrent, real-time data sources. In particular, this includes structured financial stock market data.
- filtering of data according to certain criteria. Remark: Filtering is defined through a configuration, and may be influenced by user triggers.
- the integration of historical and streaming data processing. Remark: This also includes analytics queries over purely historical data.
- at least 500 stock market messages per second per market player under normal load.
- growth rates up to factor 20 over normal load for stock market streams.

4.2 Requisites for Supported Algorithms

4.2.1 Systemic Risk Use Case

INFORE 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. 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 co-dependence 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 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, INFORE will adapt such algorithms to allow incremental update of metrics suitable for asynchronous data.

A systemic risk network 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) 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 dimensionality to a subset of interest prior to the causality analysis.

We now give an overview of the main co-dependence and causality algorithms to be properly extended (e.g., for distributed processing) and incorporated in INFORE.

Pearson's Correlation

Pearson's correlation is a measure of the linear dependence between 2 variables. In the case of 2 statistical variables X_t and Y_t, it is a measure of the extent at which a movement of X happens at the same time with a proportional movement of Y. The correlation lies between -1 and 1. If it is 1 (respectively -1), then a movement of X in one direction happens at a similar time as a proportional movement of Y within the same (respectively, opposite) direction. At the opposite extreme, a correlation of zero means there is no linear relation between the movements of X and Y.

Note the importance of the word linear here since freelance variables can have correlations of zero, however, the reverse is not true.

The classical approach to applied math correlation estimation assumes that X_t and Y_t update synchronously. This is often not the case for normal stock market data. Therefore, any candidate technique to extracting correlation within the financial markets needs to synchronize these statistic variables. The selection of the synchronization strategy is vital, since a poor choice will cause biased results depending on the frequency of data points within the stream. Totally different approaches to synchronization are tested together with interpolation and Fourier correlation.

Pros

Correlations are often computed quickly compared to alternative strategies. Computing correlations is a business customary technique that is widely notable and well accepted. Further, correlations are used with efficiency in additional advanced studies like Correlation Networks.

Cons

Pearson's correlation solely exposes linear relations between statistics and should fail to properly represent non-linear co-dependencies in time series. Particularly, Pearson's correlation is not invariant beneath monotonic transformations of the marginal distributions. Alternatives to Pearson's correlation, like Spearman's rank correlation coefficient or the more general category of Rank Correlation measures somewhat mitigate this restriction. A lot of well-known correlation metrics do not infer directivity or cause-and-effect relationships, which could be a key feature of a general risk network.

Transfer Entropy

Information entropy, or just entropy for brief, was introduced by Shannon and is the measure of the uncertainty of a random number or equivalently the typical range of bits delineated by a random variable. The higher the entropy, the more uncertain a random variable is, or the more data is obtained on average by sampling from the random variable.

Formally, given a random number X, its entropy is the average range of bits necessary to represent the outcome of X. This is the formula:

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

In the above formula, H_X denotes the entropy of X and p(x) is the probability of a specific outcome x of X. Given two data series X_t and Y_t, one will raise the question: "How much data is encoded in X_t if we assume that we have already grasped past values X_s and Y_s for s<t?". Asking this question in a different way: "How many bits are on average 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 question is zero might be if X_t is usually up to Y_{t-1}. In general cases, this value will not be zero.

Assume that each, X and Y, are Markov processes, then the transfer entropy from Y to X is outlined by:

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

The latter formula can be extended to account for a group of past values of X, not simply at X_{t-1}. Similarly, one may increase the amount of past values for Y. The computation of transfer entropy needs estimating the joint distribution of (X_t, X_{t-1}, Y_{t-1}) which can be achieved through binning or kernel estimation, among other ways.

Pros
Information entropy is not cruciate, i.e., $T_{Y \to X}$ should not essentially be equal to $T_{X \to Y}$. It makes a distinction between driving and responding parts. For this purpose, transfer entropy is helpful on directed networks of market players where we seek the market players that drive market movements.

Cons
Transfer entropy's calculation is much more computationally intensive, compared to correlation estimation.

4.2.2 Forecast and Decision-Support Use Cases

INFORE will implement algorithms designed to predict price swings based on multivariate time series. Price swing forecasts aim at giving a probability of a certain price movement ahead of time. Typically, this is done by giving an expected price movement within a defined timeframe, along with a probability of this expected move, which is based on the analytics of existing time series from the past. There are huge numbers of different approaches to prediction, where some focus on machine learning (restricted to a certain algorithm) and some others focus on the price data of the asset itself. We will focus on a broad range of asset data (almost data, that is not supposed to affect the analyzed asset, for example assets from other countries) along with a group of different machine learning algorithms. A prerequisite for all applied algorithms and strategies is that they provide an advantage over commonly used approaches. In the following, the main sets of strategies to be applied in the INFORE Financial use case are described. We will focus on Machine Learning Strategies.

Machine Learning Methods
Machine learning is a technique that allows deep levels of analysis without the need of hardcoded algorithms. It has already been used in proprietary software of SPRING to support pattern recognition for financial price swing forecasts. Practically, historical data time series are taken to perform supervised learning. Throughout the primary stages of the INFORE project, an evaluation of different machine learning techniques is going to be provided to the user to look for the ones which bring the best results, related to performance measures and user needs. Finally, the results are going to be used as a basis for the methods of the INFORE financial use case. As a benchmark for the analysis of methodologies, the requirements from the expert interviews and from the INFORE DoW are going to be applied.

Initial Requisites
From the algorithmic background introduced at the previous sections, we summarize the following requisites in the style of controlled natural language (possible with additional information in natural language).

- The INFORE system for systemic risk analysis must support the calculation of correlation for synchronized time-series data streams.
- The INFORE system for systemic risk analysis must support the calculation of correlation networks based on time-series data streams.
- The INFORE system for systemic risk analysis must support the calculation of transfer entropy for time series data streams. Remark: Techniques for improving the performance will be applied.
- The INFORE system for price swing prediction must support time series analysis based on machine learning algorithms for synchronized time series data streams.

Alternative ways of implementing the algorithms will form the basis for respective algorithm families. Furthermore, the algorithms will be considered for implementation in WP6.

4.3 Platform Quality Requisites

Based on the previous sections, we now present general quality requisites for the INFORE platform. In Section 4.3.1, we start with a discussion of quality dimensions. In Section 4.3.2, we summarize initial requisites drawn from the discussion of the quality dimensions.

4.3.1 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 INFORE system.

Timeliness
As represented in Section 4.1, INFORE should enable the processing of real-time data streams to provide up-to-date analysis results in addition to historical information. This requires the capability to support real-time calculations over high rate data streams, like the quote tick data from the stock markets. By combining real-time and historical information, INFORE will construct trend analysis and predictions of the future evolution in the stocks market. Real-time processing of enormous data volumes arriving at high rates is usually computationally expensive (in terms of CPU and storage consumption). Therefore, means must be provided to express preferences in terms of timeliness versus the computational cost.

Coverage
The coverage of the information sources to be processed in INFORE shall be maximized to provide a comprehensive market analysis. However, there is a trade-off between the amount of data being processed and the computational cost. Processing more information consumes more computing resources. On the other hand, there is a trade-off between the amount of the data to be processed and efficiency, with a given set of resources consumption constraints (more information takes longer to be processed). Here, user triggers from the applications might enable the user to influence the calculation and to express precise preferences in terms of coverage versus performance.

Accuracy
The accuracy of prediction models varies depending on many factors like the characteristics of the data and also the underlying applied math and machine-learning models. Using a lot of historical information and more correct models is usually related to some extra cost in terms of resource consumption and computation time. Whereas increasing the accuracy of the applied model is desired, it is expected that the timeliness and performance of the analysis are often negatively affected. Therefore, INFORE should allow reconciliation between these quality parameters.

Computational Performance
To achieve timeliness necessitates reducing the computation time of the analysis tasks. This could be achieved by parallelizing algorithmic tasks and/or better configure the use of HPC and/or Big Data platforms infrastructure. Generally, there is a tradeoff, which ought to be taken into consideration, between the quality and also the timeliness, coverage and accuracy properties.

Resource Consumption
Given a group of constraints on the machine resources (CPU, memory or other) INFORE ought to be ready to maximize the standard of the processed analysis and/or satisfied field or user outlined quality requisites (in terms of accuracy, coverage and timeliness). This may involve elastic resource allocation so that the financial application receives at each point in time the amount of resources specified by quality requirements as analyzed before.

4.3.2 Initial Requisites

In this section, we summarize the discussion above in terms of initial (dedicated) quality requisites, which give a first indication for the tradeoffs to be specified and handled in INFORE. Preferences among the individual quality dimensions will be specified in terms of pipeline or adaptation constraints as well as adaptation rules (see user scenarios in Section 5).

- INFORE must support timeliness in the processing of real-time data streams in order to produce up-to-date analysis results.
- INFORE must support means to customize the coverage of the data sources to produce a comprehensive market analysis. Remark: Preferences of coverage versus performance may be given in terms of user triggers.
- INFORE must support means to specify the accuracy of the performed calculation.
- INFORE must provide means to measure and optimize its resource usage.

5 Financial Area User Scenarios

In this section, we discuss the specific financial domain user scenarios of INFORE. In Section 5.1 we detail the process of gathering expert knowledge. In Section 5.2 we provide an overview on the user scenarios. In Section 5.3 we detail the user scenarios for institutional financial clients.

5.1 Analysis of the Expert User Interviews and Questionnaires

We identified 3 expert users from the financial area, whose experience (see Appendix B) covers several or all of the tasks of INFORE financial user scenario, which are:

- **Predictions** for the financial markets, based on hidden market patterns.
- **Early warnings for systemic risk** to keep a balanced and risk-adjusted portfolio of assets or goods.
- **Decision support on relative strength** and weakness of under-valued or over-valued assets.
- Discovery of **trends and correlations** over financial time series.
- Identification and monitoring of **clusters of stocks and markets**.
- Forecast/suggestions on **winning opportunities**

From the interviews and questionnaires, we extracted common workflows and categorized the needs to build concrete user scenarios.

Three specific application user scenarios have been identified. The instantiated INFORE prototype with specific data processing pipelines for the financial applications will process the input streams. The output of the processing will be post-processed for the financial application and visualization tool presentation. The three user scenarios will present the analysis results to the financial end users, i.e., the application role players introduced in Section 3. We will describe the three user scenarios in the remainder of this section.

5.2 Overview of the User Scenarios

All user scenarios follow the overall goal of decision support for investments, which covers all the aspects mentioned in Section 5.1. Depending on the user scenario, either the risk reduction in existing investments or investment opportunities in new asset will be supported.

5.3 Details on the User Scenarios

In this section, we describe the user scenarios for institutional financial clients, namely for

- FPS: Forecasting Price Swings,
- SRP: Systemic risk prediction and early warning,
- DSI: Decision support for investment opportunities.

5.3.1 Forecasting Price Swings

In this user scenario, role players (Hedge Funds, Investment Company) are looking for new investment opportunities. They use INFORE to identify upcoming price swings, that either seem to have good trading opportunities (Buying assets) or that predict falling prices of assets, that are already in the investor's portfolio (Selling assets). So INFORE will provide decision support for new investments, as well as for already investments that have already been performed.

User Scenario Description	
Identifier	FPS
Name	Forecasting Price Swings
Role player	Hedge Funds, Investment Company
Goal	Assist investment decision makers in forecasting and evaluating price swings of selected assets. Giving an "edge" in the markets, while taking into account user-specific risk/reward ratios.
Requisites	- Forecasting price swings in several specified timeframes, e.g. 30 minute, 60 minute, daily, etc. - Emphasis on forecasting and/or reacting on sudden, unexpected and large price swings. - Price swings that are not commonly identified and, therefore, provide an "edge" in the market. Examples: Individual seasonal patterns, uncommon price swing patterns that are found by using Machine Learning and similar techniques. - Main markets: Liquid stocks, Main Forex pairs, Index Futures, Commodity Futures.
Data input streams	- Historical price data of the analysed asset along with the same data for identified correlation assets - Real time streams of the analysed asset along with the identified correlating assets (to update the analysis after each selected time frame
Data and Functions	a. User information, login and authentication functions b. User creates or opens a specific workspace with markets to be analyzed. c. User selects the markets to be analyzed and specific analysis factors like timeframe and risk/reward measures. d. Output visualizations, e.g. selected opportunities as chart and numeric. Table visualization as overview of selected forecasts. Summaries of hypothetical historical results of selected forecasts. e. Optional adjustments and optimizations by the user (restart at point b.)

Table 2: Description of user scenario FPS

5.3.2 Systemic Risk Prediction and Early Warning

The role player (Hedge Funds, Investment Company, Investment Bank) wants to check if his already existing portfolio is diversified in terms of risk, or not. The given user scenario workflow assumes that the role player wants his portfolio to be as diversified as possible. INFORE will identify dependencies within the portfolio, which allows the investor to sell/reduce position sizes of assets that have strong dependencies among each other.

User Scenario Description	
Identifier	SRP
Name	Systemic risk prediction and early warning
Role player	Hedge Funds, Investment Company, Investment Bank

Goal	Early warning of systemic risk factors in existing and planned portfolios. Suggestion of risk-reduced portfolios and portfolio members, taking into account the specific requisites of the user.
Requisites	Early detection of potentially risky constellations of portfolio members by evaluating the diversification.
Data input streams	- Real time streams of the analysed portfolio assets to continuesly analyse the portfolio internal dependencies
Data and Functions	a. User information, login and authentication functions. b. User creates or opens a specific workspace with portfolio to be analyzed. c. Output visualizations, in particular cluster and table visualizations. d. Optional adjustments and optimizations by the user (restart at point b).

Table 3: Description of user scenario SRP

5.3.3 Decision-Support for Investment Opportunities

The investor (Investment Company or Investment Bank) checks the risk of market segments of interest or the overlying market to modify their management of liquidity. Rising risk will lead the investor to reduce asset position sizes and hold more liquidity. The opposite scenario is that the investor may reduce liquidity and invest in more assets (based for example on Predictive Price Swings), when the risk in the market is relatively small.

User Scenario Description	
Identifier	DSI
Name	Decision support for investment opportunities
Role player	Investment Company, Investment Bank
Goal	Identification of selected investment opportunities over a broader range of markets. Automated “cherry picking” of best opportunities from a large basket of markets and market players.
Requisites	Measurement of under-valued or over-valued assets. Identification of opportunities, which provide an above average risk to reward factor and medium to long term (12 to 36 and more months) positive return. Opportunities, that are or not strongly correlated with traditional benchmark markets (e.g., the S&P500 index) are preferred.
Data input streams	- Historical price data of the analysed asset along with the same data for identified correlation assets - Real time streams of the analysed asset along with the identified correlating assets to update the analysis after each selected time frame - Real time streams of the analysed portfolio assets to continuously analyse the portfolio internal dependencies
Data and Functions	a. User information, login and authentication functions. b. Broader market segment information, preview of selected potential investment opportunities.

	c. User selects markets to be analyzed and specific analysis factors like timeframe and risk/reward measures. d. Output visualizations. e. Optional adjustments and optimizations by the user (restart at point c).

Table 4: Description of user scenario DSI

6 Conclusions and Outlook

In this deliverable we described the results of the early requirements collection for the INFORE financial user scenario. We detailed the user-centric view of INFORE for the financial domain in terms of user scenarios for institutional financial clients. Subsequently, we discussed the requisites for the data streams to be processed and the algorithms to be provided by an instantiated INFORE platform for the financial domain. Finally, we described the three user scenarios, namely, “Forecasting Price Swings” (FPS), “Systemic risk prediction and early warning” (SRP) and “Decision support for investment opportunities” (DSI).

The collection of the user scenarios and requisites as well as the documentation process has already served as a good trigger for discussions about the functionalities, terminologies, and dependencies within the consortium. As scheduled in the DoW, this initial set of user scenarios and requisites will be further studied in the following months, which will lead to an extended and revised version of this deliverable (D2.2 on Month 18). In particular, D2.2 will detail the actual analysis algorithms to be used for risk management and prediction. The preparation of Deliverable D2.2 will incorporate the advancements on INFORE architecture and the algorithms described in Deliverables of Work Packages 5 and 6. This collection of the user scenarios and requisites will also guide the work in INFORE until the completion of the project, while actual results in the form of INFORE’s prototypes will undergo further expert user evaluation according to our workplan.

Appendix: Expert User Questionnaires

Appendix A: Original questions from the questionnaire

INFORE- User requirements questionnaire

1. User background information

- Company/organisation
- Professional position [and years of experience]
- Domain of expertise
- Background studies (university degree major, etc.)
- What are your main job tasks?
- To whom are you responsible for performing these tasks?

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
- Provide examples of use case studies
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
    - Why is this a problem?
    - How do you currently solve the problem?
    - How would you ideally like to solve the problem?
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
- Are you able to program/set up a new/custom data processing workflow?
- How long does it take to program/set up a new data processing workflow?
- Are you capable of optimising your data analysis operations?
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
- B. Which new information would you like to extract from these (old and new) data sources?

- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required
  - Ability to change algorithm parameters graphically
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast events of interest
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required
  - Ability to change algorithm parameters graphically
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast (defined or currently unknown) events of interest
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- F. What are your expectations regarding the system usability?
- G. What is the expected added value from INFORE for you and your corporation?

Appendix B: Anonymised answers of the expert users on the questions from Appendix A

1. User background information

- **Company/organisation**
  Expert User 1: Investment Company for private equity investment
  Expert User 2: Professional trading consultant and advisor
  Expert User 3: Trading coach and Hedge Fund Manager
- **Professional position [and years of experience]**
  Expert User 1: +20 Years
  Expert User 2: +20 Years
  Expert User 3: CEO, Technical Analyst and Financial Consultant

- **Domain of expertise**
  Expert User 1: Option strategies trading
  Expert User 2: Futures spread trading, seasonality trading, stocks, Forex
  Expert User 3: Technical Analysis, Prediction of stocks, currencies, Trading Psychology

- **Background studies (university degree major, etc.)**
  Expert User 1: Professional Training and Trainee as banker
  Expert User 2: No answer
  Expert User 3: Business Engineering

- **What are your main job tasks?**
  Expert User 1: Creating continous performance with minimum drawdown
  Expert User 1: Realtime Trading
  Expert User 2: “Alpha” and constant 10%+ p.a. performance with minimum risk
  Expert User 3: Analysis and Prediction of plausibilites with historical data (based on quantitativ finance)
  Expert User 3: Support for trading systems
  Expert User 3: Coaching Financial Behaviour

- **To whom are you responsible for performing these tasks?**
  Expert User 1: My company & our clients
  Expert User 2: Private coach and advisor, responsible for educating and advising my clients and for delivering trading performance
  Expert User 3: No answer

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
Realtime stock/ Options data	Tick data stream	+3000 streams	Positive performance with low drawdown	Excel (VBA), Tradestation (Easy Language)	Manual & Semi automatic	Both combined	1/s
Realtime Data Historical Data	Stream/ text files	80-100 Currencies, Commodities	Quantitative Analysis	Seasonal Advisor Trading view ATAS Volume	Automatic & Semi automatic	Historical & realtime	daily
Realtime Futures, stocks, Forex	Tick data stream	+2000 streams		Trade-station (Easy Language), Metatrader 4, proprietary analysis tools	Manual & Semi automatic	Both combined	1/s

- **Which is the aim of the analysis you perform (what kind of insights do you try to find)?**
  Expert User 1: Prediction of rising Volatility for single market players
  Expert User 2: Identifying trading opportunities with high reward/risk probability; identifying futures spread trading and seasonal pattern trading opportunities
  Expert User 3: Try to find profitable chances in the market
  Expert User 3: A positive chance risk reward over 3:1
  Expert User 3: Prognosis about the direction
  Expert User 3: Suitable Time for entry and exit

- **What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?**
  Expert User 1: Doing massive parallel analytics continuously (+K)
  Expert User 2: High data load regarding computational power for analysis, when identifying spread and seasonal opportunities
  Expert User 3: The right timing (Exit and Entry)
  Expert User 3: Finding volatil markets
  Expert User 3: Sensitive fundamental data and market influence

Expert User 3: Changing Correlations
Expert User 3: Different market cycles

- **Provide examples of use case studies**
  Expert User 1: Monitor stocks
  Expert User 1: Focus on low volatility
  Expert User 1: Predict rising volatility
  Expert User 1: Check availability of fitting Call/ Put- options
  Expert User 1: Open positions
  Expert User 1: Manage trades
  Expert User 2: Identifying and monitoring seasonal patterns across different asset classes
  Expert User 2: Identifying opportunities in Futures calendar spreads (same commodity and different Futures trading months)
  Expert User 2: Identifying opportunities in Futures intermarket spreads (making use of winning opportunities, trading different commodities against each other)
  Expert User 2: Monitoring official trading publications like the "Commitment of Traders" Report and drawing trading conclusions from it
  Expert User 3: 20 Currencies which aren't correlating

- **What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?**
  Expert User 1: Main problem is the correct prediction of rising volatility. The success rate needs to be higher. There is currently no workaround.
  Expert User 2: Managing risk, when a spread trade or seasonal pattern trade does not work out as planned.
  Expert User 2: High realtime dataload, when calculating and simulating winning opportunities in this trading logic
  Expert User 3: Quality of data
  Expert User 3: Timing
  Expert User 3: Fundamental News („unpredictable" Events like 2015 „Swiss Volatility") to process and interpret data correctly
  Expert User 3: Detect and adjust changes in market cycles (Like Brexit Event or elections)

  - **Why is this a problem?**
    Expert User 1: The failing quote can produce losses in investment
    Expert User 2: Risk management is key to professional investment – for most investors even more important than high performance. Real time risk management, considering a varity of factors, can be improved
    Expert User 3: It lowers the quality of the analysis

  - **How do you currently solve the problem?**
    Expert User 1: Filter for best predictions
    Expert User 2: Work with available risk and computational performance facilities
    Expert User 3: No current solution

  - **How would you ideally like to solve the problem?**
    Expert User 1: Having higher success rate in terms of better predict volatility
    Expert User 2: Take into account more risk factors, learning algorithms that calculate risk; higher

performance when identifying seasonal and spread trading opportunities
Expert User 3: Enhancing event prediction

- **Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?**
  Expert User 1: No. But an advanced export of results would be fine for internal automation.
  Expert User 2: No
  Expert User 3: Metatrader
  Expert User 3: Realtime Data Feed
  Expert User 3: Direkt Market Access Broker

- **Are you able to program/set up a new/custom data processing workflow?**
  Expert User 1: Yes, in a limited way by using Excel.
  Expert User 2: No
  Expert User 3: Yes, with parts of the tools mentioned above

- **How long does it take to program/set up a new data processing workflow?**
  Expert User 1: Several days
  Expert User 2: Days
  Expert User 3: Few Month because of backtesting

- **Are you capable of optimising your data analysis operations?**
  Expert User 1: No
  Expert User 2: Partially, using Tradestation “Easy Language”
  Expert User 3: Partially

- **On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)**
  Expert User 1: Laptop, Server
  Expert User 2: Trading Computer with several screens, Server
  Expert User 3: VPS Server
  Expert User 3: Personal Multiscreen workstation

3. Expected benefits from using INFORE

- **A. Please mention more data sources that would you like to use and why.**
  Expert User 1: No new
  Expert User 2: No answer
  Expert User 3: Fundamental Data Realtime
  Expert User 3: Correlation
  Expert User 3: Divergence between different markets
  Expert User 3: Diversification
  Expert User 3: Pattern Plausibility

- **B. Which new information would you like to extract from these (old and new) data sources?**
  Expert User 1: No new, but better information/ analysis
  Expert User 2: No answer

Expert User 3: Connections of different patterns
Expert User 3: Fundamental aspects
Expert User 3: Market cycles

- **C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?**
  Expert User 1: Market drops
  Expert User 2: Prediction of sudden risks of decline in value of the current portfolio
  Expert User 3: Volatility
  Expert User 3: Market direction

- **D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?**
  Expert User 1: Yes
  Expert User 2: Yes
  Expert User 3: A speed up of data analysis caused in inaccurate data and results

- **E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)**
  - Ability to design data processing workflows with no code required [2] [3] [2]
  - Ability to change algorithm parameters graphically [2] [4] [4]
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries [4] [4] [3]
  - Ability to interactively explore the data in order to detect patterns/features of interest [3] [5] [4]
  - Ability to accurately forecast events of interest [5] [5] [5]
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). [3] [3] [3]

- **E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5:** Very useful)
  - Ability to design data processing workflows with no code required [2] [3] [3]
  - Ability to change algorithm parameters graphically [2] [4] [5]
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries [4] [4] [3]
  - Ability to interactively explore the data in order to detect patterns/features of interest [3] [5] [4]
  - Ability to accurately forecast (defined or currently unknown) events of interest [5] [5] [4]
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). [3] [3] [3]

- **F. What are your expectations regarding the system usability?**
  Expert User 1: Fast learning curve
  Expert User 2: Easy to use and customized to my current trading style and focus
  Expert User 3: Easy to understand
  Expert User 3: Customizable
  Expert User 3: Transmissible to multiple markets
  Expert User 3: Quick results by using

- **G. What is the expected added value from INFORE for you and your corporation?**
  Expert User 1: Stabilized performance curve
  Expert User 2: Better performance with seasonal and spread opportunities
  Expert User 2: More accurate and faster risk analysis of the existing portfolio and positions
  Expert User 3: Improving our historical and realtime data
  Expert User 3: Trading Strategy update
  Expert User 3: Better probability predictions