1. Publishable Summary

The goal of the FERARI project is to provide the means for building large scale distributed systems that

- provide support for large scale services,
- provide Complex Event Processing technology for business users in Big Data architectures,
- integrates machine learning tasks in the architecture,
- provide support for flexible and adaptive analytics workflows and
- exemplifies the potential of the new architecture in the telecommunications and cloud domain.

In the FERARI project, we will develop a reference architecture that supports in situ processing (right where the data is generated) as a first choice for scalable processing of large scale services. Furthermore, Complex Event Processing (CEP) will be integrated in the architecture bringing stream processing much closer to the business world providing a seamless model that applies CEP as part of Big Data applications in a way easily consumable by business users. Furthermore, distributed machine learning is integrated into the architecture to allow learning sophisticated models from the data streams. The architecture all these functionalities are based on is adaptive and event-driven, i.e., it is able to adapt workflows within the architecture to changing data distributions. This architecture

- will be implemented and made available as an open source framework and
- will be the basis for the solutions FERARI provides for the telecommunications and cloud domain.

The official start date of the project was 2st of February 2014. Overall, all project objectives have been met. All deliverables have been produced and submitted in time. Phase 1 – the reporting period - has as its major objective to define a first version of the general architecture and theory as well as to define the use-cases in order to identify requirements on the architecture as well as new scientific requirements. A further goal is to identify privacy guidelines and regulations that are to be followed. During the first year, the major results have been:

1. analyzing the state-of-the-art in robust stream monitoring, flexible event processing and in-situ methods;
2. defining a first version of the general architecture in an agile, use-case driven fashion based on the STORM framework (this version has already been made available as open source framework on https://bitbucket.org/sbothe-iais/ferari);

3. identifying privacy regulations that apply to the telecommunications as well as the cloud use-case and to anonymize the corresponding data accordingly;
4. specifying the use-cases together with the end user, selecting the necessary data that has then been provided by the data provider;
5. implementing the Complex Event Processing engine PROTON from the partner IBM on STORM in order to integrate it into the FERARI architecture.

The major practical result of the first year has been the definition of the architecture and the implementation of a new method for anonymization of data according to the new EU opinion 5/2014 on anonymization techniques. The description of the architecture is provided in deliverable D2.1. The deliverable contains a thorough evaluation of Big Data Streaming Platforms that lead the consortium to deciding for STORM as the platform for the FERARI architecture. The deliverable than gives an overview on the general architecture and describes the building blocks of the FERARI framework that will be implemented. A second major practical result was porting the IBM Proton CEP engine to STORM. A first implementation of a basic FERARI framework can be reached from the FERARI open source repository (https://bitbucket.org/sbothe-iais/ferari).

In order to define the architecture and framework in such a way that it fits the requirements of the application scenarios, we investigated both the mobile fraud mining scenario and the cloud health monitoring scenario together with domain experts from HT on-site. A description of the use-cases and the subsequent requirements can be found in deliverable D1.1. Since the data for these use-cases is privacy-sensitive, the consortium investigated the requirements on anonymization techniques together with domain experts and the legal department of HT. A description of the privacy aspects and the anonymization techniques used in FERARI can also be found in deliverable D1.1.

The major scientific contributions in year involve new techniques for robust stream monitoring, distributed machine learning and in-situ methods. They can be found, together with a comprehensive state-of-the-art analysis on robust stream monitoring, in-situ methods and flexible event processing in the deliverables D3.1, D4.1 and D5.1.

In summary, there has been great progress on defining and implementing a FERARI framework which led to new and exciting ideas which resulted in a high number of high quality publications in top journals and conferences already in the first year. Making a first implementation of the framework available open source raised interest from the industry and led to many invited talks at major companies such as Siemens and Deutsche Telekom. The consortium is collaborating intensively and is working in an atmosphere of partnership and open exchange of ideas.

Further information about FERARI can be found under www.ferari-project.eu.