Abstract—This work, which is connected to the Future Internet Public Private Partnership (FI-PPP) Integrated Project FI-STAR, presents a validation approach for Future Internet applications based on the use of analytics. In particular, it discusses how to use and combine software use and health statistics for the assessment of user-perceived Quality of Experience, in order to monitor user satisfaction, the risk of user churn, and the status of the corresponding ecosystem.

Keywords—QoE; QoS; response times; usage; analytics; churn; ecosystem

I. INTRODUCTION

Since more than a decade back in time, Quality of Experience (QoE) has become a key issue of concern for operators and providers, as bad QoE implies the risk of user churn [1]. Indeed, when a service or application does not meet its stakeholders expectations, economic loss is an almost unavoidable consequence. In particular, innovative applications are at risk once they do not succeed to satisfy their users.

In many sectors, the concern for quality has led to market entry barriers related to compliance, certification, and access to mission-critical data. In health and care, for example, IEC 80001 compliance, ISO 13485 certification, and access to data such as patient records are considered problematic [2]. Software product lines have been successfully used by companies to capture such domain-specific knowledge and thereby achieve systematic reuse across their product portfolio [3]. Such reuse is achieved in a software product line by engineering design specifications and components that embed commonality and variability across use cases of potential products. The impact is faster development and productization and better quality of applications and services.

The Future Internet Public Private Partnership (FI-PPP), a “European programme for Internet-enabled innovation”, builds on this idea of product lines and attempts to scale it from a single product or services company to a whole industry [4]. FI-PPP aims at establishing an evolving set of common components, called Generic Enablers (GE) that capture solutions to common problems in the building of internet-enabled applications and domain-specific problems such as interoperability with common devices and systems and those outlined above. The hope is that the resulting infrastructure advances the European markets for smart infrastructures, increases the effectiveness of business processes delivered through the Internet, and ultimately stimulates the economy.

In its first phase, a set of GEs have been developed, which aim at providing the basis for innovative applications in virtually any application domain (e.g. e-Health, logistics, energy, etc.) within development cycles that are significantly shorter than those achieved so far. The GEs are offered by potentially competing manufacturers and producers. Application and service developers acquire these GEs for building applications in question.

The GE-based approach is comparable to buying the ingredients for a delicious home-prepared meal in a supermarket. Obviously, both the quality of the ingredients and their skilful preparation determine the quality of the prepared meal. The host can judge the quality of the meal by looking at its look, smell, and taste. The ultimate judgments of that quality, however, is seen in the appraisals of the host’s guests and in the amount that people eat and are willing to return to eat upon the host’s invitation. Translation of this metaphor to the domain of the Future Internet, makes it obvious that (1) the quality of the GEs and (2) the way these GEs are composed make a difference for a developed application as well as the corresponding ecosystem [5]. The impact of these two concerns can be seen from (a) the comments of the users, and (b) the degree of usage.

How hosts, respectively product and service organizations, achieve good-enough quality throughout the whole value chain, from ingredients to the guests’ experience and attitude, is the research underlying this paper. Our approach is based on the idea that the health of applications and their ingredients (such as GEs) needs to be measured, and that its impact on usage needs to be monitored, in order to be able to assure sufficient Quality of Experience.

The FI-PPP Integrated Project FI-STAR [6] will address such validation, and develop and implement the corresponding measurement and analysis tools as follow-up of the ongoing requirement elicitation work. This paper reflects the approach to application and GE validation within FI-STAR and its seven use cases.

The remainder of the paper is structured as follows. Section II introduces an example of a FI-PPP based system and reviews existing work for quality evaluation of such system. Section III describes the analytics-based approach for QoE prediction and assessment. Section IV summarizes and concludes the paper with planned future work.
II. BACKGROUND

Building a new system that meets its quality requirements is inherently difficult. Such requirements are often stated qualitatively like “the system must be fast”, hence are ambiguous and thus difficult to verify [7]. When implementing such requirements the following kinds of problems may be encountered. Developers build a system that delivers less than the stakeholders expect. This results in stakeholder dissatisfaction and might render a system useless. Developers build a system that delivers more than the stakeholders need. This results in a system that is more expensive than necessary.

Quality is particularly important for heterogeneously sourced systems such as FI-PPP-based systems. When engineering such system, developers depend on components, applications, and services provided by third-parties. Developers give such trust only if solution providers keep their promises regarding the service levels that will be achieved. Analytics provide transparency for evaluating such third-party contributions, for predicting the quality of the system, and for monitoring if the running service performs as promised. Analytics also provide the basis for root-cause analysis if quality objectives have not been met.

Figure 1: Patient Data Sharing Solution shows such a heterogeneously sourced system, a simplified and anonymized version of a FI-STAR use case scenario (www.fi-star.eu). The system allows patients and clinicians to collect and exchange biometric and other patient data. The system creates value by empowering the patient with rapid feedback about his condition and by providing treatment decision-support to the clinician.

According to the system architecture specification, the system consisted components, applications, devices, and services sourced from multiple parties. Patients would access the system with their personal mobile phone. The patient data collector, sharing proxy, and analysis applications would be developed by a software product company active in the healthcare domain. The sensors would be procured from a device manufacturer. User authentication services would be provided by the relevant national authority. The electronic health record would be managed by the hospital for which the solution was designed. The hospital-internal private cloud services, accessed by the clinician with one of the common web-browsers, would be provided by a local service provider. Components for connectivity and interoperability, finally, would be provided as GEs by FI-WARE platform providers.

A potentially wide variety of quality characteristics need to be fulfilled for given components, applications, and services to become useful. Such quality characteristics include functional suitability, performance, compatibility, usability, reliability, security, maintainability, and portability [8]. The quality levels achieved by a component or application is specified in the release requirements of that component or application. Warranties are used to guarantee that a product performs as promised in the specification. Usually, such a warranty is agreed between the supplier and the customer in a licensing contract [9]. Correspondingly, if a supplier provides a service...
for a customer, they agree on the quality of the service in a service level agreement (SLA). An SLA again specifies the quality levels, which the supplier gives warranty for. Norms, standards, and certificates are used to specify minimal quality levels to be achieved by products in a given industry [2].

Once developed, integrated, and deployed, the quality of the system affects the quality of the user’s experience [10]. Quality can be so good that it allows the supplier to compete with alternative solutions [11]. If quality falls below the utility breakpoint, however, users will turn away and discard the solution [12].

One approach to manage quality proactively is the use of software analytics [13]. With analytics, attributes of software entities are measured, the measurements analysed and transformed into indicators that are useful for decision-making [14]. Such measurements give transparency, thus allows developers and management to decide about the course of actions for evolving the software system [15].

A wide variety of analytics are used to manage the quality of the software engineering process, the quality of the resulting software products, and software systems that are in operation. Developer dashboards improve awareness of a project’s situation to support planning and coordination [16]. Such dashboards include information about the organization, plans and tasks, source code and builds, and quality assurance [17]. Prior to release, analytics allow analysing performance and reliability of software and services [18]. Similar analytics and geo-location are used to monitor and improve performance of the service in a real-world context with the intended users [19]. Voting buttons were proposed for measuring quality of experience. In comparison to laboratory testing, such late-stage analytics give diverse and representative results because they come from real use. Learning organizations use them to validate and improve testing assumptions.

Even though analytics are effective for managing quality of software, their use is difficult to plan. In particular, it is unclear what an effective analytics approach is for managing quality when a heterogeneously sourced system such as the one outlined above is being developed. Too many variables could be measured, and trade-offs need to be made between ease of data collection and value of the analysis [20]. In addition, the composition of a system with multiple heterogeneous parts by one player and the use of the same part by different players makes standardization of a small set of broadly useful measurements important.

III. APPROACH

Our approach of predicting quality of experience (QoE) is based on three models: a measurement model, a composition model, and a lifecycle model.

The measurement model defines how quality attributes are measured and used to assert about properties of software or of users. It closely follows ISO/IEC 15939 for analytics measurement and ISO/IEC 25010 for quality attributes.

The software composition model defines how quality propagates as a result of composing software into real-world solutions. The approach follows the ideas of soft goal networks that allow deriving high-level global quality properties from low-level measurements [21].

The software lifecycle model determines when measurements are made and quality assessed or predicted. It follows the principles of product management [22], where a release of a software product is prepared, made available for customers, and integrated by such customers into larger solutions.

A. Measurement Model

The measurement model describes how data is collected to make assertions about quality of service and of experience. In our cooking metaphor, such data collection corresponds to the host that probes the ingredients or meal and interviews the guests. Probes include looking, taking a smell, and tasting the food and asking guests whether they like its appearance and taste. The host uses this data to understand whether the food meet the desired quality standards and to understand the guest’s experience with it. Some of these properties can be derived from the corresponding measurement. For example, bad smell can be an indicator for bad food. Other properties can be inferred from indirect measurements. For example, whether the food was good can be inferred by asking the guests about their opinion. Similarly, experienced cooks are able to accurately predict the guest’s experience based on the just tasting the food. The assessment of the ultimate success is different, though. As hosts, we would define it as whether the guests are eating or not. This can be assessed by observing whether the guests are eating or not.

Figure 2 illustrates the application of these measurements to software that is used by a human user. The human user corresponds to the guest, the software to the meal, and the host to the software provider. Software analytics are applied at the software, and empirical inquiries performed with users. Both of them allow collecting data for assessing quality of the software, quality of the user experience. Also, either of them also allows assessing the ultimate success of the software: whether it is used or not.

A substantial amount of work exists to understand how to assess software quality with analytics. Many address a selection of the software quality characteristics outlined by ISO/IEC 25010. The most common analytics are time and error-based.
The most common empirical inquiry determines a score of user opinion. Table 1 gives an overview of existing work on how measurement are used to assess software quality. It excludes software qualities that affect stakeholders other than users. Table 2 illustrates the same idea: how measurements will be used to assess the impact of software on the user.

<table>
<thead>
<tr>
<th>Table 1: Measurement of Software Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional suitability</strong></td>
</tr>
<tr>
<td>Performance</td>
</tr>
<tr>
<td>Reliability</td>
</tr>
<tr>
<td>Security</td>
</tr>
<tr>
<td>Usability</td>
</tr>
</tbody>
</table>

Evaluation of functional suitability of a software is usually performed by functional testing. However the result of functional suitability is reflected in terms of functional acceptability from user’s perception. It can be reflected even in usage analytics [25]. As example, during a software use, unnecessary functions will be understood from click a stream that is an implication of functional inappropriateness.

Other aspects of software quality, usually called Quality of Service (QoS), are performance [31] and reliability [32]. QoS usually refers to system components and network delivery capacity. It concerns time behavior, resource utilization, and capacity aspects, in addition to availability, frequency of failures, fault tolerance rate, and recoverability time. Attributes such as throughput, loss ratio, jitter, packet error rate, response time, delay and availability time are vital for measuring in the network layer, and the transport layer between two machines [28, 33]. Servers are measured by essential attributes of load rate, error rate, response time, peak response time, server up time, resource (i.e. CPU, memory, and disk) utilization, and threads [34]. In the application layer, statistics about page errors, frame rate, call success rate and the quality of outputs such as audio, video, and files are identified to measure QoS [35]. Finally, security of an application/component affects the solution health [36, 37]. The Attacks attribute is used to combat security issues such as DOS or malware attacks [38].

A time dependent attribute has the largest coverage for an end-to-end software health assessment. User perceived quality is dominated by response time and waiting time [39, 40]. The perception of quality on the user is typically measured by the Mean-Opinion-Score (MOS)[41]. Availability of the software solution is measured by infinite response time. The response time of an intrusion tolerant system with the steady-state availability is monitored for the security assessment [30]. Therefore response time can be a suitable candidate that simulates waiting time, availability as well as security. Error attribute provides further support for the assessment of software health in security, availability and fault tolerance.

<table>
<thead>
<tr>
<th>Table 2: QoE Measurements mapping to Quality in Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effectiveness</strong></td>
</tr>
<tr>
<td>Efficiency</td>
</tr>
<tr>
<td>Satisfaction</td>
</tr>
<tr>
<td>Freedom from Risk</td>
</tr>
</tbody>
</table>

B. Composition Model

The composition model describes how data is collected to combine assertions about quality of service and of experience. In our cooking metaphor, such composition corresponds to the host that combines and cooks the ingredient into a meal that is served to the guests. The host uses heating and combination to process the ingredients into a result of value higher for the guests than the inputs that were used. The quality of the inputs and the host’s own work affect the quality of the results. The results are at most as good as the worst of the inputs that was used. Skillful preparation of the meal and presentation of it to the guests, however, can increase the value of the whole meal well beyond the sum of the inputs.

Figure 3 shows a software composition model that allows describing the solution shown in Figure 1. Nodes such as the private cloud contained in the secure zone, which again is contained in the hospital correspond to instances of the infrastructure. Patient Data Analysis and Electronic Health Record are two instances of software that run on the private cloud infrastructure. Not shown in Figure 1 are the generic enablers that the Patient Data Analysis contains. The clinician is a user that uses a browser, which communicates with the Patient Data Analysis and the Electronic Health Record software.

Figure 3: Composition Model

The composition model allows propagation of quality properties. Such propagation can be expressed in rules that are evaluated with an instance of the composition model (Figure 1 is such an instance). They determine how a property of one entity, for example a failure of an infrastructure, affects the rest of the software system. A set of availability-related rules would state that failure of infrastructure implies that any dependent software and user will experience the failure. Software that runs on reliable infrastructure, however, would not be affected by the failure. Similarly, a set of performance-related rules would state that the total roundtrip time for a user interaction corresponds to the aggregated time behavior of software, run on the respective infrastructure, and communication channels. Depending on criticality of quality of service and experience, the set of rules can be completed and refined.

C. Lifecycle Model

The lifecycle model describes how software and service infrastructure come into existence and evolve. The evolution stages then give raise to possible quality assurance actions. In our cooking metaphor, the cook would perform quality assurance actions based on the evolution stages of ingredients.
and the meal. He would look for ingredients that are made available to him on the market. Preferably, he would turn to ingredients with trusted quality, for example as indicated by certification labels awarded to some ingredients. In addition, he would touch and take a smell of some of them to assess their quality. Once in the kitchen, he would process and combine the ingredients into a meal. The meal undergoes quality assurance in the kitchen before it is made available to the guests. Once these guests have received the meal, they look at it and take a smell (presumably with delight) before they decide to eat it.

Figure 4 shows a lifecycle model that allows explaining how software is developed, delivered, integrated, and made available as a solution for the healthcare environment shown in Figure 1. Each supplier, indicated by the letter in parentheses in Figure 1, has developed, tested, and released software or infrastructure. The integrator then has performed acceptance testing of the sourced software and infrastructure in his own premises and integrated them into the solution that Figure 1 describes. Again, the integrator tested and released the software solution, before performing site acceptance testing and initiating its usage.

![Software Lifecycle Model](image)

Figure 4: Software Lifecycle Model. FAT = factory acceptance test. SAT = site acceptance test.

The lifecycle model describes quality assurance actions that is performed at each respective lifecycle stage. Factory acceptance testing includes testing of the software in the supplier’s laboratory environment. Software release is accompanied with certification. Such certification is standard practice of application stores such as Google Play and iTunes [42]. Site acceptance testing is performed by the consumer of the released software in a laboratory environment that is as close to the real-world environment as possible. In the healthcare environment, site acceptance testing of software systems is accompanied IEC 80001 and ISO/IEC 27000 [43]. Systems that have passed all these quality assurance hurdles are put into use, where they continue to be monitored [19].

Each quality assurance action involves collection of analytics and possibly empirical data as described by the measurement model. The collected data updates earlier predictions made with the help of the composition model. Such updating allows validation of the prediction and increases confidence in whether the final solution actually meets its quality objectives or not.

The combination of the measurement and composition models enable early prediction. The lifecycle model allows planning for step-wise improvement of these prediction, hence reducing the risks of the final test of where a solution is being used in a real-world environment and the achieved quality of experience level determines success or failure of the system.

IV. SUMMARY AND CONCLUSION

While important for any software, quality assurance is particularly critical for acceptance and successful use of heterogeneously sourced systems. Such systems integrator components from parties that the system integrator has little control over. As a consequence, the risk and the corresponding need for trust much higher than when a single-source software is developed.

This paper introduces a holistic approach for quality assurance of heterogeneously sourced systems. It is based on three models that together allow quality of experience prediction and step-wise validation of these predictions with real-world measurements. The measurement model describes how analytics and empirical data is collected and used for assertion of quality of service and experience. The composition model describes how measurements are propagated through the composed system to estimate overall quality of service and experience. The lifecycle model describes quality assurance actions that are used for validation of system quality.

The paper represents an important step towards unifying the so far separated disciplines of software engineering and performance evaluation in telecommunication systems. It contributes with a QoS and QoE measurement-based approach to managing quality while a software system is constructed. The paper explains the approach in depth with the metaphor of a host that prepares a delicious meal to guests. An exemplar taken from the FI-STAR project is taken to describe how the approach is transferred into a real-world environment.

Future work includes validation of the approach. Analysis of software architectures will be used for refining the composition model. A literature review will be performed for constructing a rule base for QoS and QoE assessment and prediction. Empirical inquiries about engineering process will be used to evaluate the composition model and refine the description of quality assurance practices. A particular focus will be given to the healthcare environment, where quality assurance is particularly important as it may decide on death or life.

V. ACKNOWLEDGMENT

This work has been co-sponsored by the European Commission, DG CNECT, through the FI-PPP Integrated Project FI-STAR.

VI. REFERENCES


