

Towards Assessing Open Source Communities' Health using SOC Concepts *

Oscar Franco-Bedoya¹, Marc Oriol¹, Carlos Müller², Jordi Marco¹, Pablo Fernández², Manuel Resinas², Xavier Franch¹, and Antonio Ruiz-Cortés²

¹ Universitat Politècnica de Catalunya

² University of Seville

{ohernan,moriol,jmarco,franch}@essi.upc.edu

{cmuller,pablofm,resinas,aruiz}@us.es

Abstract. Quality of an open source software ecosystem (OSS ecosystem) is key for different ecosystem actors such as contributors or adopters. In fact, the consideration of several quality aspects (e.g., activeness, visibility, interrelatedness, etc.) as a whole may provide a measure of the healthiness of OSS ecosystems. The more health a OSS ecosystem is, the more and better contributors and adopters it will gather. Some research tools have been developed to gather specific quality information from open source community data sources. However, there exist no frameworks available that can be used to evaluate their quality as a whole in order to obtain the health of an OSS ecosystem. To assess the health of these ecosystems, we propose to adopt robust principles and methods from the Service Oriented Computing field.

1 Introduction

Software ecosystems (SECOs) are emerging as an alternative approach for understanding complex software systems context. SECO can be defined as a network of related actors, interacting with a shared market for software and services [2].

SECOs are crucial in the field of Open Source Software. An aspect that makes OSS ecosystems different is the existence of communities: contributors communicate bugs, committers confirm patches, the community produce releases, branches/forks, etc. Community members interact using mailing lists, blogs, forums, etc. Under these circumstances, ensuring the health of SECOs for OSS quality becomes crucial, and these processes and resources need to be considered in a comprehensive quality assessment program.

In our approach, we propose to assess the health of an OSS ecosystem by monitoring a set of Key Health Indicators (KHIs), which are factors or attributes that determine the healthiness in an OSS ecosystem. To define the KHIs, we will use QuESo, a quality model for OSS ecosystem [1].

* This work is partially supported by the EC-FP7 project RISCOSS (agr. 318249), the European Commission FEDER, the Spanish and Andalusian RDI programmes (P12-TIC-1867, TIN2012-32273, TIN2013-44641-P, TIC-5906, IPT-2013-0890-3) and the network of excellence of the Spanish MINECO (TIN2014-53986-REDT)

To monitor these KHIs, we envision that the knowledge, practices and methods in other fields with similar challenges can provide a robust solution to accomplish the aforementioned goals. Particularly, we believe that the current state of the art in Service Oriented Computing (SOC) related to quality assessment has strong similarities that can be ported into OSS ecosystem health analysis.

To this aim, we are developing a new platform, named SALMonADA-OSS, adopting the techniques from SOC to the OSS ecosystem health.

2 SALMonADA-OSS

SALMonADA-OSS is intended to be a monitoring and analysing platform to assess the OSS ecosystem's health. This platform will be based on a previous work named SALMonADA [3]. This is a monitoring and analysing platform that is capable of identifying and explaining if any violation of the Service Level Agreement (SLA) for web services occurs. Here we describe how we align the problem of assessing SLAs from SOC to OSS ecosystem's health, and the enhancements we are conducting to SALMonADA, resulting in the SALMonADA-OSS platform.

2.1 Alignment with SOC

OSS development and maintenance usually requires different software management tools, such as mailing lists, bug tracking systems and version repositories; each one providing an aspect of the OSS ecosystem's health. We envision that these management tools can be offered as services. Options are: (1) The tool already provides a web service interface (e.g. JIRA bug tracking system provides a RESTful service), and (2) The management tool is wrapped into a web service.

Under this environment, the OSS community health monitoring and analysis can be conducted following analogous principles as in SOC. Similarly to SLAs, we propose the ecosystem Health Level Agreement (eHLA), which will establish the KHIs that should be assessed. Unlike SLAs, where the conditions are usually established to technical low-level metrics (e.g. response time <2s), eHLAs establishes conditions on a higher-level of abstraction based on probabilistic methods over defined distributions. For instance, given the distribution of activeness levels: very active, active, not very active and inactive, the user may state that he wants that the OSS community have a 80% probability of being very active or active. ($\text{Activeness_VeryActive} + \text{Activeness_Active} > 80\%$).

The values for such KHIs are computed from lower level metrics, such as number of commits, messages in forums, etc. and combined using probabilistic models instantiated in the form of Bayesian networks.

2.2 Monitoring and analysing process

SALMonADA-OSS will be automatically configured to monitor and analyse the metrics included in an eHLA (see Fig. 1). As depicted, the Composer reads the

eHLA and configures the Monitor to measure the KHIs. The Monitor is a web-service based monitor that observes and computes metrics from the different web services that expose the software management tools (e.g. JIRA, GIT,...). The monitor measures low-level metrics (e.g. number of commits) and aggregates all the gathered data to compute the KHIs using bayesian network methods. These KHIs are then reported to the analyzer to perform the analysis of the eHLA fulfillment. If the conditions stated in the eHLA are not fulfilled, the analyzer will identify the violation, and report it to the interested parties.

Because the monitoring process is based on web services, the architecture and design of the initial SALMonADA can be reused, and enhanced to assess the KHIs in the field of OSS Health. The enhancements and scientific contributions that we are conducting in SALMonADA-OSS are briefly detailed below:

- A new SLA-based language, named eHLA, to express the conditions over KHIs to assess the OSS ecosystem's health.
- Monitoring management tools as web services. In particular, we will reuse and extend the web service monitor of SALMonADA, named SALMon [5].
- Aggregating low-level metrics to KHIs, using bayesian-network methods.
- Analysing the fulfillment of eHLAs, reusing methods from SLA analysers. We will reuse and extend the analyzer of SALMonADA, named ADA [4].

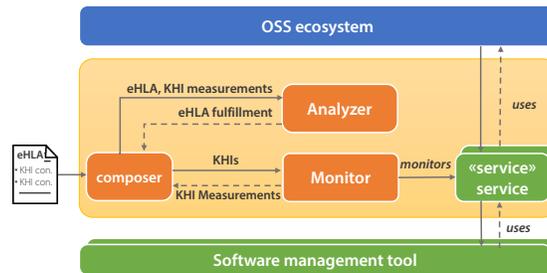


Fig. 1. OSS monitoring and analysing process in SALMonADA-OSS

References

1. O. Franco-Bedoya, D. Ameller, D. Costal, and X. Franch. Queso: a quality model for open source software ecosystems. In *ICSOF*, 2014.
2. S. Jansen, S. Brinkkemper, and A. Finkelstein. Business network management as a survival strategy: A tale of two software ecosystems. In *IWSECO*, 2009.
3. C. Müller, M. Oriol, X. Franch, J. Marco, M. Resinas, A. Ruiz-Cortés, and M. Rodriguez. Comprehensive explanation of sla violations at runtime. *TSC*, 7(2), 2014.
4. C. Müller, M. Resinas, and A. Ruiz-Cortés. Automated Analysis of Conflicts in WS-Agreement Documents. *TSC*, 7(4), 2013.
5. M. Oriol, X. Franch, and J. Marco. Monitoring the service-based system lifecycle with salmon. *ESWA*, In Press, 2015.