

# Towards a Multi-Perspective Methodology for Big Data Requirements

Evangelia Kavakli, Rizos Sakellariou, Iliada Eleftheriou  
*University of Manchester, UK*

Julien Etienne Mascolo  
*Centro Ricerche Fiat SCPA, Italy*

**Abstract**—This poster describes work in progress that is concerned with requirements engineering in the context of big data. Drawing experience from a H2020 project, this poster argues for a classification of requirements from different perspectives that can be used to guide the requirements elicitation process and form the basis of a common methodology for requirements engineering in big data applications.

## I. INTRODUCTION

The continuous growth of data along with the availability of cloud resources and the development of data analytics have led to an intensification of activities related to software development projects built around big data. Big data system development rests largely on the “orchestration” of a set of technologies that can meet the data characteristics and enable the effective conversion of data to applicable knowledge. However, rapid technological changes together with the involvement of stakeholders from diverse backgrounds, who often have no expertise in data mining or data analytics, lead to vague requirements or requirements which are difficult to be accurately acquired [1].

Recently, there has been an emerging effort to address Requirements Engineering (RE) in the context of big data. The majority of these approaches stresses the need to properly address the so-called big data V-characteristics (primarily volume, velocity and variety) in the definition, analysis and specification of the system functional and non-functional requirements [2]. Some related research focuses on system requirements (for example, challenges in selecting and integrating different technologies [3], [4]) or user requirements (for example, provision of appropriate analytics tailored to the level of user expertise [5]). Although alignment between business goals and big data user and system requirements is generally considered critical for achieving value through big data, emerging approaches that deal with requirements engineering for big data systems tend either to treat business goals as a secondary issue or not consider them at all [2]. Even less so, they do not recognise the need to identify relevant key performance indicators for measuring and validating the impact of big data technologies on the business goals. This work is an early attempt to cover this gap by proposing a multi-perspective methodology for RE.

This work has been supported by the European Commission under Grant Agreement No. 780787 (I-BiDaaS).

## II. A BIG DATA REQUIREMENTS METHODOLOGY

The proposed methodology draws on work in the area of early RE. Whilst late requirements focus on the target technical system, early requirements address the interplay between business intentions and system functionality. In early requirements elicitation, business intentions are typically conceptualized using goal-based languages, as in the context of Goal-Oriented Requirements Engineering [6]. According to the goal-oriented paradigm, elicitation of requirements can be seen as the systematic transformation of high-level business goals to specific system requirements that operationalise these goals. This analysis takes place on the boundary between the business strategic view (what needs to be accomplished) to the end-users’ operational view (who does it). End-users may range from non-expert business users who introduce new data into the system and/or use the big data analytics services and results, to expert big data application developers and big data infrastructure providers. Understanding different end-users and their needs is a key input for reaching an agreement about the intended system. As a result, requirements elicitation for big data applications encompasses three interrelated perspectives:

- *business requirements* describe specific needs of an organisation that must be addressed through a specific big data analytics activity or project.
- *user requirements* describe the needs of a particular stakeholder or group of stakeholders who will be impacted.
- *system requirements* describe the behaviour that the big data system (or a system component) should expose, or the capabilities it should own in order to realize the intentions of its users.

In the context of the H2020 project I-BiDaaS [7], [8], this initial classification was further elaborated in a top-down manner, based on the review of related research and was revised in a bottom-up manner through the generalization of over 300 specific requirements collected in the context of 9 real use case scenarios within three industrial sectors (telecommunications, banking and manufacturing). This resulted in 24 generic big data requirement categories in the above perspectives. These generic requirements can be used to guide requirements elicitation prompting stakeholders to focus on and express application specific requirements in the relevant categories. In addition, considering big data requirements from different perspectives enables better business/Big Data system alignment and assists the traceability among business and

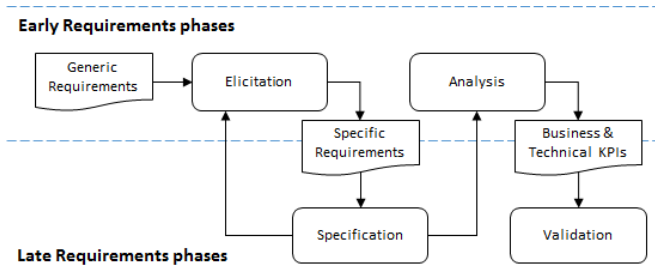


Fig. 1. The proposed way-of-working

system performance. In particular, key performance indicators (KPIs) related to business goals at the business level can lead to the specification of technical indicators that can be used for validating the proposed system solution. This methodology is shown in Fig. 1.

### III. AN INDUSTRIAL USE CASE

This section provides a walk-through of how the proposed way-of-working has been applied to an automobile manufacturer use case, henceforth referred to as CRF, in the context of the I-BiDaaS project. The overall goal for CRF is to use Big Data in order to improve production line efficiency by effectively managing possible interruptions due to process issues such as machine breakdown, and unscheduled maintenance, so as to avoid excessive costs and financial damage.

This description provided the starting point to discover requirements by documenting the high-level business goals and associated KPIs as terms of reference. In collaboration with CRF stakeholders and technology providers, the requirements elicitation followed a mostly top-down approach whereby business requirements were further refined in order to identify specific user requirements, whose analysis resulted in the definition of system requirements. This process was facilitated by the use of a questionnaire.

Fig. 2 depicts a summary of the CRF requirements. Starting from the business requirements, the strategic business goal R1 was refined into more operational business goals (R2 and R3), and associated KPIs (R4 - R6). At the user requirements level, requirements were described in terms of the characteristics of different data sources that are planned to be used (requirements R7 and R8); the analytics capability of the proposed solution envisaged (R9); and the different interface requirements of the end-users that will consume the analytics results (R10 - R12). Further refinement of the above user requirements, resulted in the generation of the system requirements, both functional (R13) and non-functional (R14 and R15).

Furthermore, the KPIs defined at the business level have been mapped onto specific indicators at the big data system level. For example, the “Product/service quality improvement” KPI at the business level has been mapped onto “Data quality”, which relates to the accuracy of the analytics models at the system level. Such indicators have been used to determine whether the proposed Big Data solution meets CRF needs, during validation.

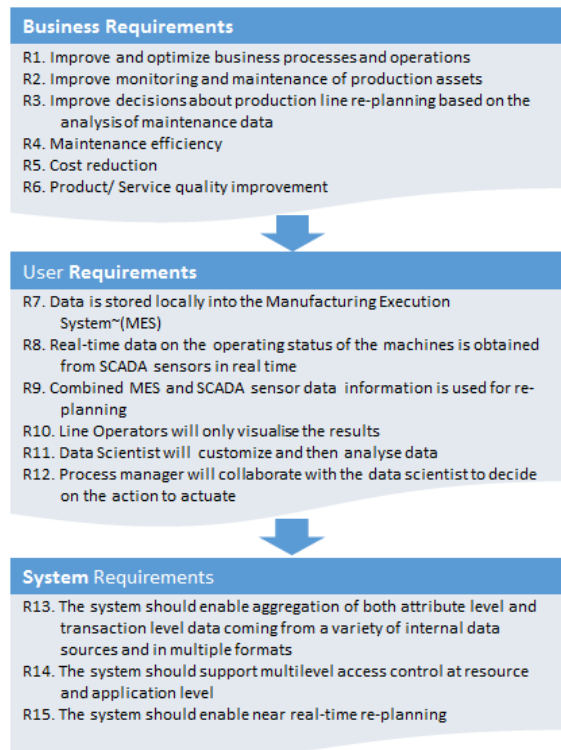


Fig. 2. From high level to specific CRF requirements

This example provides only an excerpt of the elicited requirements, however, it demonstrates how the proposed methodology guides the identification of different big data requirements. A detailed list of all requirements elicited and further information is provided in [7]. Current efforts focus on the empirical validation of the proposed methodology through application to use cases from different domains.

### REFERENCES

- [1] P. Wang *et al.*, “Eliciting big data requirement from big data itself: A task-directed approach,” in *2017 6th International Workshop on Software Mining (SoftwareMining)*, Nov 2017, pp. 17–23.
- [2] D. Arruda, “Requirements engineering in the context of big data applications,” *SIGSOFT Softw. Eng. Notes*, vol. 43, no. 1, pp. 1–6, Mar. 2018.
- [3] O. Hummel *et al.*, “A collection of software engineering challenges for big data system development,” in *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Aug 2018, pp. 362–369.
- [4] M. A. Serhani *et al.*, “Quality profile-based cloud service selection for fulfilling big data processing requirements,” in *2017 IEEE 7th International Symposium on Cloud and Service Computing (SC2)*, Nov 2017, pp. 149–156.
- [5] G. Camara *et al.*, “Big earth observation data analytics: Matching requirements to system architectures,” in *Proceedings of the 5th ACM SIGSPATIAL*. New York, NY, USA: ACM, 2016, pp. 1–6.
- [6] J. Horkoff *et al.*, “Goal-oriented requirements engineering: an extended systematic mapping study,” *Requirements Engineering*, vol. 24, no. 2, pp. 133–160, 2019.
- [7] I-BiDaaS Consortium, “D1.3: Positioning of I-BiDaaS,” 2018. [Online]. Available: <http://www.ibidaas.eu/sites/default/files/docs/Ibidaas-d1.3.pdf>
- [8] I. Arapakis *et al.*, “Towards specification of a software architecture for cross-sectoral big data applications,” in *2019 IEEE World Congress on Services (SERVICES)*, July 2019, pp. 394–395.