Lessons Learnt in the Implementation of CMMI® Maturity Level 5

Isabel Lopes Margarido*, Raul Moreira Vidal
Department of Informatics Engineering
Faculty of Engineering, University of Porto
Porto, Portugal
{isabel.margarido,rmvidal}@fe.up.pt

Marco Vieira
Department of Informatics Engineering
Faculty of Sciences and Technology, University of Coimbra
Coimbra, Portugal
mvieira@dei.uc.pt

***Abstract*— CMMI® has proven benefits in software process improvement. Typically, organisations that achieve a CMMI level rating improve their performance. However, CMMI implementation is not trivial, in particular for high maturity levels, and not all organisations achieve the expected results. Certain CMMI implementation problems may remain undetected by SCAMPI℠ since only a sample of the organisation is analysed during the appraisal and assessing the quality of implementation of some practices may be difficult. In this paper we present the case of three CMMI level 5 organisations. From the lessons learnt and based on an extensive bibliographic research, we identify a set of problems and difficulties that organisations willing to implement CMMI should be aware of and provide a set of recommendations to help avoid them. As future research we will develop a framework to help to evaluate the quality of implementation of CMMI practices.**

Keywords- Capability Maturity Model Integration, high maturity, quality of implementation, Standard CMMI Appraisal Method for Process Improvement.

I. INTRODUCTION

The Capability Maturity Model Integration (CMMI®) [1] is a process improvement maturity model for the development of products and services. CMMI has two representations: the continuous representation which includes four Capability Levels (CL), from CL0 to CL3, this being applied to individual Process Areas (PA); and the staged representation, to which this paper refers, which is composed of five maturity levels (ML), from ML1 to ML5, and being applied across PA. Assuming that an organisation that does not completely fulfil the requirements of the other maturity levels is considered to be at ML1, each ML is achieved by implementing its specific and generic goals and all the preceding ones.

Organisations that implemented CMMI typically improve performance in terms of quality of products and processes, schedule and costs. Consequently, processes become more predictable and customer satisfaction increases [2]. However, this is not always the case. Approaches such as the Team Software Process (TSP℠) allow some organisations to achieve lower defect rates in delivered products than the average of organisations with Capability Maturity Model® level 5 [3]. There is more variance in performance results when using CMMI, as it is a generic model of good practices, not of detailed processes. Thus, when using a prescriptive process like TSP results are understandably more predictable. In practice, the differences in performance from organisation to organisation, using CMMI, depend not only on the context of the business, projects and team but also on the methodologies used in the implementation of CMMI practices.

Implementing CMMI ML 4 and 5, also known as high maturity levels (HML), is particularly challenging. Due to their dependency on metrics HML rely on the quality of implementation of the ML 2 Process Area Measurement and Analysis (MA). When implementing MA organisations define metrics, how they are collected, stored and analysed. Such metrics need, however, to be meaningful to the organisation and be aligned with its business goals. Considering ML 4, to implement Organisational Process Performance (OPP) it is essential to establish the organisation standard processes, build process performance models (PPM) and establish process performance baselines (PPB). In the case of Quantitative Project Management (QPM) it is necessary to ensure that projects establish and monitor quantitative goals based on organisation, client and team goals, and select sub-processes to use and monitor their performance. In ML5, Organisation Performance Measurement (OPM) is implemented to guarantee that the organisation performs towards meeting its goals. If problems, deviations or needs for improvement are detected Causal Analysis and Resolution (CAR) is used to analyse root causes of problems and eliminate them. OPM also includes necessary practices to select and pilot improvements, verify effects in processes performance before deployment, and regulate improvements deployment [1].

The Standard CMMI Appraisal Method for Process Improvement (SCAMPI℠) may be used to benchmark the maturity of a company in the CMMI model [9], by identifying process weaknesses, investigating and determining the degree of satisfaction of CMMI Process Areas Goals, and assigning maturity ratings [4]. As it is

Research partially sponsored by Fundação para a Ciência e a Tecnologia (FCT): Programa Operacional Potencial Humano (POPH) of QREN, and Fundo Social Europeu (FSE).

not cost and effort effective to appraise an entire organisation and all its projects, SCAMPI includes sampling rules, ensuring that the appraised subset of business units and projects is representative of the entire organisation. The selection of evidences necessary from each appraised business unit follows coverage rules that determine the level of affirmations and artefacts that need to be provided [4]. However, these rules do not assure that all organisation projects use the practices of the appraised ML. After all evidences are collected, the appraisal team characterises the implementation of CMMI practices as a weakness or a strength for each model practice and each basic unit or support function. Each practice in each basic unit or support function may be termed Fully Implemented (FI), Largely Implemented (LI), Partially Implemented (PI), Not Implemented (NI) or Not Yet (NY). Aggregation rules at implementation level are used to derive organisational unit-level characterisation. A goal is considered Satisfied if all associated practices at organisational unit level are characterised as LI or FI and the aggregation of weaknesses of the goal do not have negative impact on its achievement [4].

In the past, problems were detected on the implementation of high maturity levels [5], and issues such as lack of capability, poor performance and/or lack of adherence to processes were found in the application of CMMI [6]. The problem is that SCAMPI may miss some implementation issues due to the fact that only a sample of the organisation is appraised and not all business units provide evidences. Furthermore, measuring organisations performance is outside the scope of SCAMPI. The assessment [7] is done by verifying if the techniques applied allow achieving CMMI goals. There are two process areas where performance improvement is explicitly analysed: application of CAR, the effect of implemented actions on process performance should be evaluated; and OPM, the selection and deployment of incremental and innovative improvements should be analysed [7, 8]. In this paper we analyse problems and difficulties occurring in implementation of CMMI, presenting lessons learnt in the implementation of maturity level 5 in three organisations, with a set of recommendations that could help to avoid them. We compile the results in a way that can be used as a checklist by organisations intending to implement CMMI.

This paper comprises: section II with problems and difficulties that may occur in implementing CMMI, section III with problems of three organisations appraised CMMI level 5, section IV presents lessons learnt and recommendations towards preventing such problems, section V with results discussion and section VI with conclusions and future research.

II. Literature Review

There are several common problems in the implementation of CMMI, particularly in HML. Like any structure, the model needs a good foundation to be stable and efficiently evolve to higher maturity levels. The demands of levels 2 and 3 should prepare organisations to adequately use measurement at higher levels, by monitoring appropriate and adequate metrics.

In 2000, a survey [5] to understand what CMM level 4 and 5 companies used in the implementation of the model showed that practices were not clearly institutionalised. Hollenbach and Smith [9] stated that the Software Engineering Institute (SEI) concluded that some companies did not understand the statistical nature of CMMI level 4 and that companies that achieved CMMI HML rating did not enjoy an established consensus on necessary characteristics of level 4. In 2004, the US Department of Defence (DoD) [10] recognised that not all organisations programmes are appraised, resulting in practices not being implemented organisation wide and baselines eroding once a certain ML is achieved. In response to this problem Pyster proposed a set of solutions, one of which is already covered in SCAMPI version 1.3: provide guidance on how to select representative samples and aggregate results from subordinate organisations [11].

Many companies face problems when implementing CMMI HML that arise from complex practices such as measurement and quantitative management or use of effective performance models to predict the future course of controlled processes. In fact, some of the difficulties found in evolving the processes and implementing new Process Areas are related to the need to move to statistical thinking and quantitative management [12]. Statistical Process Control (SPC) is a useful tool in the implementation of metrics programmes, based on the Shewhart Control Charts. Florac *et al.* [13] stated that when using SPC organisations "grow frustrated", as they apply SPC either to complex processes with several sub-processes or to the organisation's entire software process. Their work included an example of the application of SPC in defining PPB for inspections and realising that there were sub-processes present, which were not evident at first. SEI compiled process performance models used by CMMI ML 5 appraised organisations [14].

Kitchenham *et al.* [15] analysed the database of a large CMMI level 5 corporation and found that metrics were at times collected but could not be correlated and they did not have meaning for upper management. The authors noted important concerns to be taken into consideration when storing data and designing databases, such that data analysers and decision makers can actually use them. They also proposed the use of the M^3P (Model, Measure, Manage Paradigm) framework, which extends Goal Question Metric (GQM), by providing links between the collected metrics, the development environment and the business context. Some authors [16] argue that several measurement programmes in organisations fail by defining too many measures not actually implemented and analysed in decision making.

When the MA system is complex, the support of information systems becomes essential to avoid errors and overhead. Johnson, Kou *et al.* [17] recommended the use of telemetry to improve software development management. They proposed a tool whose sensors are connected to all tools used in software development and collect process and product data non-intrusively. In their paper they define which metrics need to be monitored, which data should be collected and how to store and analyse them.

There is also a problem in metrics definition, which at times impedes an understanding of how data are collected and analysed, and finding the common and special causes of variation. As Goulão already mentioned, many metrics exist but the values that allow their calculation or themselves are frequently expressed in natural language [18]. Breuker, Brunekreef *et al.* [19] emphasised the difference between the definition of software metrics in literature available (books and papers), the tools for data collection specifications and those tools actually collecting data. Literature needs to clearly define software metrics, and practitioners should be aware of this when implementing the MA system. Several authors state that metrics programs fail due to the production of metrics which do not allow proper analysis of the performance and capabilities of the organisation processes [20]. There is also a problem of metrics adequacy.

Hamon and Pinette [21] indicated a series of bad practices in the deployment of indicators to control projects, including: *1)* meaningless, useless and non goal-driven indicators; *2)* complicated indicators without triggers for actions; 3*)* inexperienced implementers; *4)* complex solutions, hard to maintain; *5)* same indicators for all situations (ignoring specific needs of different projects); *6)* too many indicators at the beginning; *7)* out of date measurement plans; *8)* return of investment of the metric ignored; *9)* bad quality data and processes that are not followed.

Leeson [22] indicated more problems in CMMI implementation: *1)* senior management not involved in establishing objectives, policies and the need for processes; *2)* sponsor not playing its role and delegating authority; *3)* software Engineering Performance Group not managed; *4)* organisations focused on achieving a maturity level more than improving quality of their products or services.

The later author also described various difficulties faced by organisations when interpreting CMMI: they lack a global view of the model; they do not understand the relationship between measurement and project monitoring, and indeed of several ML 2 generic practices (GP), or the difference between capability and maturity levels. Some organisations misinterpret ML 2 and 3, which in more extreme cases causes the failure of many programmes. When not considering HML practices, organisations fail to understand the ultimate goal to be attained, as they do not adequately see the direction they are taking at lower maturity levels before moving to HML.

All the aforementioned problems are the result of a deficient implementation of CMMI but they become more nefarious when they are not detected in appraisal. We now present some limitations of the SCAMPI method, which is focused on practices implementation. The objective of the appraisal is not to verify how people are actually doing things and the quality of their results [23]. With such orientation malpractices may be missed by the appraisal team. It is assumed that SCAMPI results depend on the quality of the appraisal team and, in fact, they reflect their knowledge, experience and skills [4].

SCAMPI relies on organisation's honesty: it provides evidences and supports the choice of the projects to appraise [23]. Either the lead appraiser is very rigorous in the choice of projects and critique about evidences, or the appraisal outcome may be biased by the organisation.

In the appraisal only a small number of affirmations sustain the practices. SCAMPI V1.3 coverage rules limit the number of affirmations. We consider that a single affirmation from a business unit, safe by coverage rules, can demonstrate it is not following one of the practices. Furthermore, not all programs of the organisations are analysed in the appraisal [10], only a small percentage, so it is easier to have no guarantees that the entire organisation works the same in all projects or programs – that means that practices may not be institutionalised.

Pricope and Horst indicated that SCAMPI is described in natural language and does not provide an activity-oriented graphical description of the appraisal process. They proposed a method to measure SCAMPI, introducing a quality metric for activities to allow determination of a level of weakness or strength of the appraisal elements [24]. The proposed method is useful in quantifying the conducted appraisal; however it does not evaluate how practices are actually done, nor does it evaluate organisation performance.

Sunetnanta *et al.* [25] proposed a model that constitutes a repository for pooling data from all projects, used in organisations working with different offshore units, however we consider it to be applicable to any organisation. The repository allows collection of evidences as the projects are ongoing, as well as analysis of projects and of appraisal results. The quantitative assessment of projects is done by scoring the number of times an activity is executed and by checking if the activity was executed or not. There is a limitation, though: evidences still need to be evaluated and analysed by the appraisal team. An evidence may be generated but if empty it shows that the expected activity was not performed, and even when generated it would be necessary to assess whether people actually did the practice, or whether an artefact that is mandatory was just produced.

Several pilot projects were conducted to implement results-based appraisals [26]. The work was based on measures of results, including the Telecommunication Quality Management System – TL 9000 standard. The measurement repository built in pilot environment had no documented linkage to processes and practices in the standard process or in the CMMI. The objective of the pilots was to identify and validate an appraisal that would assess performance measures. The information collected on SCAMPI should be useful to trigger the appraisal team for further investigation in face of unexpected performance, have results-oriented findings, have records for posterior assessments and result in recommendations related to performance and benchmarking. From the identified challenges the appraisals took longer (5% to 10% more than a regular appraisal) and became more expensive. The industry benchmarks varied in value, raising doubts as to their applicability.

III. Case Studies Results

To verify the problems that a CMMI implementation may present we conducted three case studies in multinational organisations that develop software (CI, CII and CIII) assessed at CMMI for Development ML 5, staged representation. Case studies on CI and CIII were conducted immediately after the appraisal. The purpose of the case studies was to identify real problems and difficulties in the implementation of CMMI and find recommendations to avoid them. We mainly focused on MA and HML, but also analysed the other CMMI PA. The research questions we intended to answer, which we considered when designing the case studies and analysing all data, were the following: *1)* what was the strategy to evolve to the new ML? *2)* what difficulties and problems occurred in the implementation of the new practices? *3)* what is the process definition? *4)* how was the process defined? *5)* how are people using the process? *6)* how are people collecting, analysing and interpreting process data? *7)* what is the impact of the new process on people work?

While considering these case studies we detected some of the problems mentioned in the previous section, along with new ones, which we compile in this section.

A. Case Study I

In CI we interviewed the CMMI programme sponsor, posing not only direct questions but ending with an open-end question to which the interviewee answered by narrating the story of the program. We carried out similar interview with the program responsible. Both interviews allowed us to identify interviewees, projects and other documentation to analyse. We analysed their Quality Management System (QMS), the Information System and the SCAMPI A repository. We also interviewed practices and tools implementers, teams of projects including the appraised ones, in other words, those whose documentation we analysed.

From the analysis of CI we found that the main problem stemmed from rapidly evolving to ML5 without giving enough time to have stable tools, processes, processes performance baselines and people behaviour.

Underestimate time to implement HML – Time to implement CMMI level 5 is often underestimated; i.e., the implementation takes longer than expected as it is considered that there is enough time when there is not. CI had to re-plan the CMMI implementation programme several times until Six Sigma was introduced and gave them a better understanding of the demands of HML.

Understand the statistical nature of level 4 [9, 12] – In CI a move to statistical thinking and quantitative management was the most challenging part of the implementation.

Dissemination problems – After CI was appraised at CMMI ML5 people recognised that communication had improved, however some were still having difficulties in applying the new practices because the dissemination of information regarding processes and tools usage was not totally effective.

Lack of institutionalisation [5, 10] – We verified in CI that not all projects' teams were applying the new practices. This problem was also related to people behaviour and resistance to change.

Meaningless uncorrelated metrics [15] – In CI we found a case of metrics being misinterpreted due to lack of understanding of the context of one business area.

Metrics definition (collect and analyse data) [18, 19, 21] – People in CI were still having difficulties in collecting data in certain contexts and in their interpretation.

Metrics categorisation – In CI the collection of data for high maturity had been occurring for a short period of time so the baselines were not stable enough, meaning that it was not possible to distinguish between different categories of data (for different market, team experience, team size, project size), so the data were compiled in PPB only categorised by technology.

Baselines not applicable to all projects – After CI being rated ML5 baselines were still not completely stable, or adequate for all types of projects. There had not been enough time to collect data from different contexts where they could occur and verify: *1)* if new metrics were needed; *2)* if there were differences in performance and, if so, in which contexts; *3)* if in certain circumstances the procedure to collect the data should be different.

Abusive elimination of outliers – Outliers are data points that occur outside statistical limits of control of a process, and which indicate that it is out of control. They can result from an error committed on collection or the circumstances in which they occurred can be completely characterised and are not expected to reoccur. In CI there was one situation of eliminating an outlier without realisation that it had occurred at least once in some projects. In certain cases these data points are not outliers, they are only indicators that the process improved its performance [Spirula member, 2010 personal communication].

Not all projects are measurable – CI tools were not yet prepared to collect data in certain projects, because the projects data structure was different from the standard ones. Maintenance projects metrics differ from the outsourced projects and the development ones. We found that tools were prepared to collect data from development projects but were not configured to collect data from maintenance projects, projects with several phases and outsource projects. Furthermore, measurements specific to maintenance and outsource projects were not defined.

Tools setup – CI evolved the existing Information Systems to support the implementation of the new practices. After the appraisal people were still detecting problems and requesting improvements. Such requests were the result of using the tools in practice and in different projects contexts.

Overhead – In CI people considered that collecting new metrics introduced overheads even if tools collected part of them. From use of tools and application of new practices people were still identifying new requirements for tools or developing ones that were missing, contributing to overhead.

People behaviour – In CI changing mentality was a challenge. Some people did not see value in new practices or stated that they were not applicable to their projects. It is difficult to convince people to report effort accurately; they normally report contract hours, not real effort. It is also difficult for them to report effort as they are finishing tasks, leaving the reporting until later, worsening the problem.

B. Case Study II

CII is a business unit, located in several countries, that is part of a CMMI level 5 organisation that we name CIIG. In CII we conducted an interview with the responsible for the CMMI programme beginning with directed questions and finishing with descriptive questions regarding the story of the program. Afterwards we analysed CIIG QMS and Information System. In the end we presented and discussed the results and our conclusions with the CMMI program responsible.

In CII we found several problems related to metrics, but most metrics limitations came from the fact that size was not being measured, and time spent on tasks stopped being accurately collected. The resistance to change and a difficulty in presenting, to CIIG, metrics adequate for the business unit CII are the origin of most of the identified problems.

Multicultural environment – CIIG acquired others and imposed its processes on acquired companies. Good practices and certain metrics of acquired companies are often lost, along with good visibility of processes. Another problem is that people from different cultures have different ways of working. In certain cultures orders are taken without question, while in others people need to understand the benefits of working in a certain way, otherwise they will resist change.

Lack of institutionalisation [5, 10] – In CIIG this problem also occurred. Not all projects and business units performed at the same maturity level.

Metrics definition (collect and analyse data) [18, 19, 21] – CII faced this problem because CIIG imposed KLOC (thousand lines of code) as the applicable size metric.

Not all projects are measurable – Due to the nature of CII projects, the use of KLOC was not adequate for most of them. Consequently, many derived measures were not used by this business unit.

Baselines not applicable to all projects – This problem also occurred in CIIG as PPB were centralised but not applicable to all business units' realities and projects, e.g. the phases of development lifecycle had different durations depending on the business. Another problem was that productivity used a business day as unit of time but in some locations it had different duration in number of hours.

People behaviour – People in CII also stopped reporting effort accurately, only reporting contractual hours of work.

Effort estimates – While in CIIG effort estimation is based on their historical data of effort and size, CII estimates were based on expert judgment and did not use any tools or models.

C. Case Study III

In CIII we interviewed a consultant involved in the appraisal of the organisation, i.e. a person who performed an actual observation on the case.

The main difficulty faced by CIII was to move to statistical thinking.

Underestimate time to implement HML – in CIII time to implement HML was also underestimated. Implementation turned out to be more complex than anticipated and the programme took longer than initially planned.

Understand the statistical nature of level 4 [9, 12] – was one of the challenges faced by CIII, because changing mentality to ML5 is a significant shift. The preparation and usage of the quantitative component takes time to mature. This problem may have been one of the causes for those involved to *underestimate time to implement HML*.

Introduction of HML forgetting ML 2 and 3 [22] – this problem occurs often when implementing HML and CIII was no exception.

First data collected were uncorrelated – In CIII the collected data were not correlated from the outset, which implied conducting new data collection cycles and a new search for correlations. This may have been one of the causes of the problem *introducing HML forgetting ML 2 and 3* occurring in CIII.

IV. Lessons Learnt and Recommendations

Some of the problems identified in the previous section reflect a poor implementation of the MA process area, from ML 2, and affect the organisation results. Such problems become more evident when implementing ML 4 because the correlation of values and problems in the collected data affect PPM and PPB. There are several challenges that organisations have to face when implementing CMMI and HML in particular. SCAMPI cannot appraise the entire organisation and does not analyse performance measures – if it did, it would become even more expensive. CMMI rating *per se* is not a guarantee of achieving expected performance results and organisations need to be aware that there are different methods that can be used on its implementation. However, if some recommendations such as the ones we propose in this section are followed, CMMI implementation can be easier, and the problems discussed before can be avoided.

Entry Conditions – When planning a move towards high maturity it is important to have enough time to have mature levels and institutionalised practices, understand

and analyse the needs for HML, find correlations between variables, have stability of metrics, process, tools and work habits, have meaningful performance indicators and have enough stable data points to have statistically meaningful historical data. Organisations need to carefully plan business and process improvement objectives, temporal horizon and resources: time, internal and external human resources, tools, training, etc.

HML only work with a stable base. The introduction of ML 4 and 5 can only occur after ML 2 and 3 are mature and institutionalised [22]. It is important to guarantee the involvement of a quantitative expert (often a statistician) preferably with experience in software and if possible also in CMMI. A permanent statistician with software and CMMI experience can help better understand processes behaviour and correlations between variables, along with providing adequate statistical tools to different contexts. Introducing a Six Sigma initiative in the organisation eases the introduction of the statistical knowledge necessary to the organisation workers. There must be a top down and bottom up revision of the organisation's processes, improvements/innovations, goals and quantitative goals.

Process Definition – The implementation of the model should reflect the culture of the organisation, and not be a copy imposed on personnel. Processes definition should identify current processes (*as is*) and improvements (*to be*) so that these reflect an organisation's culture and people good practices [22]. When defining processes it is important to involve the experts, including those who use the process to do their work: project, technical and quality managers; developers; testers, etc.

In multicultural organisations and when acquiring new companies imposing processes can result in loss of knowledge and resistance to change. Different business units should be allowed to share practices used and lessons learnt. Each business unit will then gradually and naturally adopt the other's practices if they better fulfil needs. This approach allows creating processes without losing good practices in a way that allows the organisation to benefit from cultural differences. There should be goals specific for different business units, departments and projects, which must be related to the organisation business goals. This setting allows having goals monitored at all levels, avoiding losses of visibility by middle management in each level.

Commitment from the entire organisation is essential, including involvement of top management as well as middle management and the people who are actually doing the work [22]. Training needs to be adequate for each role and to include not only the *what to do*, *how to do* and *hands on* components but also the *why shall we do it*, *what will we achieve* and *how do we see it*. Top management needs to set goals, plan gradual institutionalisation, monitor and reward, i.e. recognise the good work of projects and departments. For that it is essential that they understand the processes. To have people commitment it is crucial that they understand the new practices. This can be achieved by coaching projects and people [27], guiding and accompanying them.

Metrics and processes definitions mature when used in practice because it is when problems arise that it becomes more evident how procedures can actually be done. It is necessary to give some time to let processes and metrics mature before producing their final versions.

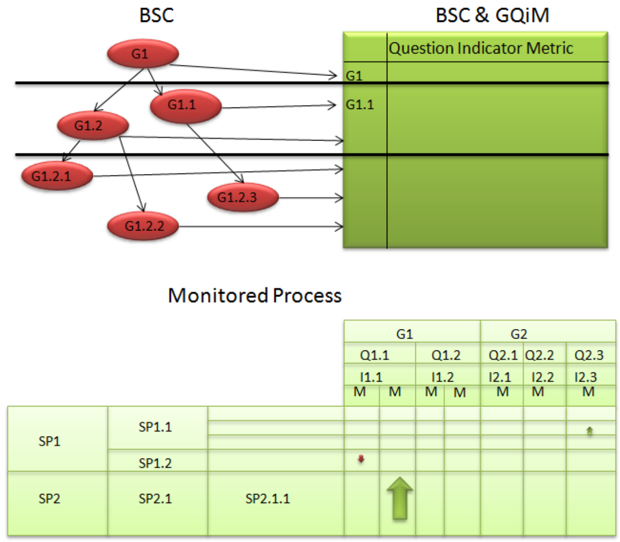

Figure 1. Mapping BSC into GQiM, into processes and sub-processes (SP). G- goal; Q- question; I- indicator; M- metric; SP- Sub-process.

Metrics Definition – To establish business objectives and identify the indicators of the processes performance, organisations can use the goal-driven measurement [28] method, a combined application of the Balanced Score Card (BSC) and Goal Question (Indicator) Metric (GQ(i)M). We represent the method in Figure 1. The business goals metrics are established using BSC and are drilled down from organisation's goals to business units' goals and ultimately projects and individuals goals. The metrics are derived by using the Goal Question indicator Metric (GQiM) and mapped with the organisations different levels of goals. The most relevant goals/sub-goals for the business strategy are elicited and realistic objective values established for those goals indicators. Regarding the monitored processes, data of the metrics used to calculate the indicators are collected in the different sub-processes. If the current processes performance does not allow achieving the quantitative goals (e.g. decrease a metric value, such as number of defects; or increase a metric value, as % of code being reviewed) then process performance improvement projects can be pursued to help achieve them. Monitored processes are being followed.

Understanding metrics is a process that is completed when projects are using them as defined by the final versions of the processes related to their collection or analysis. It is utterly necessary to train the entire organisation; the effort put into such tasks should not be undervalued.

Measures need to be defined with a set of repeatable rules for collecting and unambiguously understanding both the data and what they represent [13]. If different people use them differently, then their definition is inadequate.

The level of detail of metrics needs to be completely defined and understood; the different types of projects' context must be considered, including the technology used. For example, in some technologies there are more files, the time to execute unit tests is negligible, etc. Those factors affect the definition of the metrics. Another example is project type: outsourced, maintenance and development projects, for instance, will have different measurement and control needs.

Basic software processes about which data should be collected should be defined, then concatenated and the data decomposed in different ways to provide adequate information at project and organisation level [15]; if necessary, data should be normalised to make them visible to top management. It is preferable to begin with a sub-process executed often and with a small number of variables in order that results come faster. When the process is stable, this can then be extended to other processes and more complex ones [13].

Metrics databases take time to become stable and allow the construction of relevant PPM and PPB. The data need to be categorised. Florac *et al.* refer to this process as "separating or stratifying data that belong to the same cause system [13]". But to have adequate categorisation it is necessary that the different projects fully cycle to completion. Either the organisation has a significant number of concurrent projects with small lifecycles or the organisation begins to work with first limited baselines that evolve with time. Pilot projects are useful for stabilising processes, procedures and tools. The way people use tools may change the way metrics should be collected. Only after those projects are over and the practices are clearly defined will the organisation be ready for training, the processes/procedures and tools be fully and correctly documented and people be able to learn and apply the practices. Changes may then be deployed in order that processes become institutionalised.

Metrics Usage – Certain outliers can be removed from databases but it is necessary to pay close attention to those instances not immediately understood; they can indicate a process is having a new behaviour (better or worse performance). They may be a common situation and some are indicators of the existence of a different process, with a different behaviour and therefore originate new sub-processes [13]. One way of avoiding the error of abusively eliminating such outliers is to quarantine them, i.e. monitor the process without the outlier in parallel to the process with the outlier. Then decide the most adequate action: *1)* perform CAR; *2)* eliminate the outlier; *3)* establish a new baseline because process performance improved; *4)* create new sub-processes, in case of having sub-processes. Florac *et al.* give an example of how to do it [13].

Regarding effort estimation, expert judgment is more adequate in certain circumstances, in particular when there is absolutely no previous knowledge of the project [29]. Effort estimation does not necessarily need to be based on KLOC for it to be based on historical data; it can be based on other size metrics, phase duration or the time spent on task. When no data are available at all, iterative planning should be carried out so that when data from previous cycle are available they can be used to plan the following.

Tools Setup – It is important to understand that tools need time to be set up, especially when evolving existent ones. In order to avoid overhead in the data collection process, the information system needs to have limited human intervention, e.g. reporting effort and measuring code. Effort spent on different software applications for doing the tasks may be measured and part of the effort automatically labelled; the person only verifies and corrects eventual errors by the end of a block of tasks. This avoids forgetting to report effort or constantly interrupting tasks to manually report. The information system should be composed by automatic storage tools connected to the development environment [17]. Data collection should be automated, as manual data collection is time-consuming and has a great propensity for error [21]. It is utterly necessary that data collection is precise, if it was not so previously, people need to change their mentality and display discipline.

The data collected when correcting those tools defects which have impact on the definition of the metrics and of the process should not be used to build PPB, because the process is not stable.

Table I summarizes problems and recommendations that shall be considered when implementing CMMI, in particular ML 4 and 5, so it can be used as a checklist. We also indicate the PA, Specific Practice (SP), Generic Goal (GG) or Generic Practice (GP), which are possibly affected. Most of the recommendations collected by us were applied in practice by the organisations of the case studies and were helpful in overcoming the detected problems.

V. Discussion

In the following paragraphs we further discuss results.

A. *Validity and Limits to Generalisation*

Due to access limitations, the three case studies had a different design so they cannot be considered multiple-case studies [30]. Only part of the design of CI was repeated on CII, and in CIII we only interviewed a consultant involved in the appraisal. We can classify it has a semi-multiple case study. In CI and CII we used multiple sources of evidence, assuring construct validity. However in CIII we could not assure it. In all cases we had our results reviewed by key informants. To ensure internal validity we did pattern matching by classifying information and aggregating it under each category; built explanations and addressed rival explanations. External validity was partially tested by replicating part of the design used in CI in CII. Nonetheless, for each case study we used theory.

Regarding limits to generalisation, we only analysed three cases but some of the problems that we identified were also found in the literature review, consequently we consider that these problems can be common to other organisations implementing CMMI, measurement programs or doing software process improvements.

TABLE I. PROBLEMS AND RECOMMENDATIONS.

Category	Problem	Recommendations	Refs	PA/GG
Entry Conditions	*P1. Underestimate time to implement HML*	R1: Plan considering activities such as maturing levels, analysing and understanding HML, maturing PPB and PPM, collecting data repeatedly until meaningful performance indicators can be systematically obtained.	CI, CIII	
	P2. Introducce HML forgetting ML 2 and 3.	R2: Before moving to HML guarantee that ML 2 and 3 are mature and institutionalised.	CIII,22	
	P3. Understand the statistical/quantitative nature of level 4: Underestimation of time to change mentality from ML 3 to quantitative thinking, and time to implement ML 5.	R3: Involve a statistician with experience in software and preferably on CMMI. R4: Introduce Six Sigma initiative. R5: Review goals and quantitative goals top down and bottom up when implementing CMMI.	CI,CIII , 9, 12, 22	
Process Definition and Implementation	*P4.* Copy processes from CMMI.	R6: Processes shall reflect the culture of the organisation, not be a copy of the model imposed to the personnel. R7: Involve experts and process users in the definition of processes.	CII, 22	
	P5. Multicultural environments: people dealing differently with change.	R8: Interaction between business units to share processes and lessons learnt to design processes together.	CII	
	P6. Impose processes on acquired organisations with good practices losses.	R8, R9: Have goals specific of different business units, departments and projects, related to the organisation business goals. Have indicators to monitor them at different report levels.	CII, 16	
	P7. Dissemination Problems: Difficulties in applying new practices, in particular in understanding how to collect, analyse and interpret metrics.	R10: Have commitment from the entire organisation: involve top management, middle management and the people who are actually doing the work. Have a sponsor. R11: Train top management on: sponsorship; goal setting; monitoring and rewarding (at different goals levels); on the process (understand it). R12: Have different levels of training. Specialised training for sponsors and top management, process group and all roles that are affected by changes. Shall include what to do, how to do, hands on, benefits and how can benefits be seen. R13: Coaching of projects and people (guiding and accompanying) and monitoring (from top management).	CI, 22, 28	GP2.5 GP2.6
	P8. Lack of Institutionalisation: No all projects used the new practices.	R14: Top management: set goals (when, who, what); include goals for gradual institutionalisation, monitor and reward. R15: Metrics and processes definitions mature when used in practice, need time to define final versions. R13	CI, CII, 5, 10	GG 2 GP2.5
Metrics Definition	*P9. Meaningless Uncorrelated Metrics:* Misinterpretation of metrics due to lack of context information.	R16: Use goal driven measurement to establish quantitative goals. R17: Measures defined with a set of repeatable rules for collecting and unambiguously understand the data and what it represents. R18: Use different size measures according with the workproduct. R19: Identify different context that need to be associated with the metrics in order to adequately interpret them. R20: Do variables normalisation to ensure that metrics are usable in the entire organisation.	CI, 16, 29, 13	MA SP1.4 MA SP2.2
	P10. Metrics definition (collect and analyse data): Not adequate to all contexts, vague, allowing errors in collected data due to different interpretations.		CI, CII, 18, 19, 21 , 13	MA SP1.3 MA SP1.4 MA SP2.1 MA SP2.2
	P11. First collected data were uncorrelated	R21: Conduct all necessary data collection cycles to find correlated metrics.	CIII	OPP
	P12. Metrics Categorisation: Not all contexts data available. Unstable baselines without different categories.	R22: Give time for the metrics databases to become stable and allow the construction of relevant PPM and PPB. Different projects full cycles completed. R23: Categorise data. R24: Aggregate normalised data only for global view.	CI	OPP
	P13. Baselines not applicable to all projects		CI, CII, 17	OPP SP1.3 OPP SP1.4 QPM SP2.2
Metrics Usage	*P14. Abusive elimination of outliers:* exceptional causes of variation occurring once per project or new baseline being established.	R25: Quarantine outliers which cause is not immediately identified. R26: Recognise data points that are not outliers but are unique and recurrent.	CI, 17	MA SP1.4 MA SP2.2
	P15. Not all projects are measurable: Not collecting data from projects with a data structure different from the standard. Not using all derived metrics because of lack of definition of base measures adequate to context.	R14, R27: base measures should be defined for different work and then normalised to allow calculating derived measures.	CI, CII	MA SP1.3 MA SP2.3
	P16. Effort Estimates: without using historical data of effort or size.	R28: Expert judgment is more adequate in certain circumstances. R29: Use any related historical data: size, phase duration, time spent on task. R30: Do iterative planning.	CII, 30	PP SP1.2 PP SP1.4
	P17. People Behaviour: inaccurate personal data reports.	R31: Never use personal data for people evaluation purposes. R12, R13	CI, CII	
Tools Setup	*P18. Tools Setup:* Problems in tools after deployment. Using the tools in practice and in different projects contexts allowed to identify undetected problems and necessary improvements.	R32: Tools are improved when used in practice, save time for their setup. R33: When correcting tools defects that have impact in the metrics definition and the process, do not use the collected data to build PPB.	CI	OPM SP2.2 OPM SP2.3
	P19. Overhead in tools usage (data collection not completely automatic). *P20.* New needs still being identified, new tools still being developed.	R34: Once PPM and PPB are stable only collect data that is needed. R35: Use automatic and unperceived data collection systems, with limited human intervention (start/stop and confirm). R36: Guarantee that data collection is precise (discipline and change people mentality).	CI, 15, 21	

B. Problems Analysis

The number of problems found in each organisation increase with the depth and insight provided by a more complete design of the case study. Nevertheless we found two groups of problems common to two different groups of two organisations (see Figure 2).

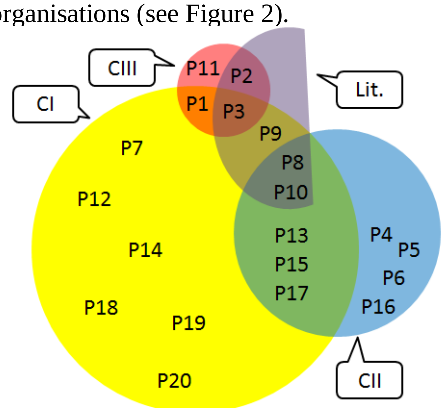

Figure 2. Venn Diagram of Problems: part of CI problems is common to CII, another part is common to CIII and 5 of the problems found in literature (Lit.) are common to the ones found in CI, CII and CIII.

Several problems found in CI were also detected in CII, four of them are related with metrics definition and usage and the other two are related with institutionalisation and people behaviour, respectively. Another two problems found in CI also occurred in CIII, all of them related with assuring entry conditions. CII was just of a business unit of CIIG, who was rated ML5 for a long time, so we cannot verify if they faced similar entry conditions problems. However, we realised that the metrics problems found could be due to CII's lack of understanding of the requirements for HML and statistical nature of ML4. We cannot even conclude that CIII did not face the metrics problems because we did not analyse their PPM, PPB, metrics definitions and usage in person. Five of the problems found in the literature review were also detected in CI, CI and CIII, CI and CII, and CIII, respectively.

C. Recommendations Analysis

R1 was used by CIII. The CMMI implementation plan had a long duration but all activities needed to be conducted on the estimated time. The thought that if an activity had overrun the schedule that time could be recovered by shortening others was abandoned once CIII concluded they could not shorten effort of posterior tasks. R2 was followed by CI and CIII. CI did a GAP analysis to address problems in lower maturity levels, however those processes were affected by changes due to HML implementation and there should have been a new cycle for them to mature. In CIII the move from ML3 to ML5 was uninterrupted, so ML3 matured and did not erode in the meantime. R3 to R5 were also followed by CI and CIII, as they were part of the CMMI implementation process. In CI, Six Sigma helped to gain insight of information needs to achieve quantitative goals, solve problems and design PPB and PPM.

R6 and R7 were followed by CI and CIII by first understanding the existent process, identifying GAPS and involving internal experts and users in the definition of improvements and new processes. CII applied R8 in a single direction, i.e., analysed other business units metrics in order to adopt the ones that could be applicable to their projects lifecycle. R9 was helpful to maintain the visibility of processes and projects at different organisation levels in CI and was part of the CMMI implementation process in CIII.

R10 to R13 were used by both CI and CIII. Regarding the training on benefits and how can they be seen we can not be sure if it was effective. R14 was used by CI and CIII, the dissemination of processes was gradual, as they were ready to be deployed directly from pilot projects to the entire organisation. However, when organisations are large they should consider even more gradual dissemination, spreading practices in a small group of projects and gradually involving new ones, which can be done also profiting from team members mobility. R15 was used by CIII.

R16 was used by both CI and CIII so there was a clear view of which metrics were used to monitor different levels of goals and what was their definition. R17 was also followed but in CI definitions needed to mature to ensure unambiguous collection and interpretation. In CIII it was necessary to define new metrics for ML5 to have the desired confidence, because the integrity of existent data from ML3 could not be assured. With time the definition of metrics was improved to tune the process models. R21 was part of the CMMI implementation process of CIII. R29 was followed by CI. R18 and R27 were followed after the case study, when they defined their specific metrics and implemented the estimation tool.

Even following R31 in CI it was hard to convince people to accurately report effort that is why we suspect that showing the benefits on training may not have been effective. Regarding R32, on both CI and CIII the tools initially used were more rudimentary. As processes, metrics and performance models and baselines were defined more complex tools were adopted or implemented. Regarding R34 the experience in CIII was that initially it was necessary to collect data of all variables they felt could be important to create models and establish baselines. In time the non-used metrics were abandoned, leaving only the necessary ones. In both CI and CIII R35 was followed but it is always difficulty to totally eliminate human intervention in effort report, especially when people have other tasks than just developing code, for example.

VI. Conclusion

Regarding CMMI problems, the DoD stated that it is necessary to "Develop meaningful measures of process capability based not on a maturity level, e.g. Level 3, but on process performance" [10]. CMMI V1.3 is more focused on performance of organisations but SCAMPI is becoming more efficient [31], as it reduced the number of necessary evidences (which may eventually increase the probability of leaving problems undetected) but it does not measure performance.

The difficulties in implementing CMMI, in particular HML, are common to the problems found on metrics programmes and software process improvements in general. In particular Software Engineering metrics are still ambiguous [18, 19], impeding an implementation common to all organisations. With the objective of understanding CMMI problems better, we did the research presented on this paper and compiled them. We also assembled a set of recommendations useful for software development organisations.

There is a wide variety of methods that can be used in the implementation of CMMI practices. As the model is just a guide which tells you what to do, but not how to do it, room is left for various implementations, and these may not always lead to the desired performance results. Furthermore, SCAMPI's objectives do not include appraising performance. Consequently problems and difficulties can occur when implementing CMMI, some of which can persist after appraisal. In future research we will provide a method for self-assessing the quality of implementation of CMMI practices and effects of improvements, based on compliance, efficiency and effectiveness [32]. We believe our framework will aid in prevention of implementation problems and allow better control of organisation performance.

REFERENCES

[1] M. B. Chrissis, M. Konrad, S. Shrum, CMMI for Development®: Guidelines for Process Integration and Product Improvement, 3 ed. Massachusetts: Addison-Wesley, 2011.

[2] D. R. Goldenson, D. L. Gibson, R. L. Ferguson, "Why Make the Switch? Evidence about the Benefits of CMMI," presented at the SEPG, 2004.

[3] N. Davis and J. Mullaney, "The Team Software ProcessSM (TSPSM) in Practice: A Summary of Recent Results," CMU/SEI CMU/SEI-2003-TR-014, ESC-TR-2003-014, 2003.

[4] SCAMPI Upgrade Team, "Standard CMMI® Appraisal Method for Process Improvement (SCAMPI SM) A, Version 1.3: Method Definition Document," CMU/SEI, CMU/SEI-2011-HB-001, 2011.

[5] R. Radice, "Statistical Process Control in Level 4 and Level 5 Software Organizations Worldwide," presented at the Software Technology Conference, 2000.

[6] R. Charette, L. M. Dwinnel, J. McGarry, "Understanding the Roots of Process Performance Failure," CROSSTALK The Journal of Defense Software Engineering, vol. 17, pp. 18-22, 2004.

[7] S. Masters, S. Behrens, J. Mogilensky, C. Ryan, "SCAMPI Lead Appraiser SM Body of Knowledge (SLA BOK)," CMU/SEI CMU/SEI-2007-TR-019, ESC-TR-2007-019, 2007.

[8] CMU/SEI. (2008, 21-06-2012). Criteria for Audits of High Maturity Appraisals. Available: http://www.sei.cmu.edu/cmmi/solutions/appraisals/himataudits.cfm

[9] C. Hollenbach and D. Smith. (2002) A portrait of a CMMI SM level 4 effort Systems Engineering. 52-61.

[10] M. Schaeffer, "DoD Systems Engineering and CMMI," presented at the CMMI Technology Conference and User Group, 2004.

[11] A. Pyster, "What Beyond CMMI Is Needed to Help Assure Program and Project Success? Unifying the Software Process Spectrum." vol. 3840, M. Li, *et al.*, Eds., ed: Springer Berlin / Heidelberg, 2006, pp. 75-82.

[12] A. Takara, A. X. Bettin, C. M. T. Toledo, "Problems and Pitfalls in a CMMI level 3 to level 4 Migration Process," presented at the Sixth International Conference on the Quality of Information and Communications Technology, 2007.

[13] W. A. Florac, A. D. Carleton, J. R. Barnard, "Statistical Process Control: Analyzing a Space Shuttle Onboard Software Process," IEEE Software, vol. 17, pp. 97-106, 2000.

[14] R. W. Stoddard II and D. R. Goldenson, "Approaches to Process Performance Modeling: A Summary from the SEI Series of Workshops on CMMI High Maturity Measurement and Analysis," CMU/SEI CMU/SEI-2009-TR-021, ESC-TR-2009-021, 2009.

[15] B. Kitchenham, C. Kutay, R. Jeffrey, C. Connaughton, "Lessons Learnt from the Analysis of Large-scale Corporate Databases," presented at the International Conference on Software Engineering, Shanghai, 2006.

[16] L. F. S. Monteiro and K. M. d. Oliveira, "Defining a catalog of indicators to support process performance analysis," Journal of Software Maintenance and Evolution: Research and Practice, vol. 23, pp. 395-422, 2011.

[17] P. M. Johnson, H. Kou, M. Paulding, Q. Zahng, A. Kagawa, T. Yamashita, "Improving Software Development Management through Software Project Telemetry," IEEE Software, vol. 22, pp. 76-85, 2005.

[18] M. C. P. A. Goulão, "Component-Based Software Engineering: a Quantitative Approach," Doctoral Thesis, Departamento de Informática, Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia, Lisboa, 2008.

[19] D. Breuker, J. Brunekreef, J. Derriks, A. N. Aicha, "Reliability of software metrics tools," presented at the International Conference on Software Process and Product Measurement, Amsterdam, 2009.

[20] M. P. Barcellos, "Uma Estratégia para Medição de Software e Avaliação de Bases de Medidas para Controlo Estatístico de Processos de Software em Organizações de Alta Maturidade," Doctoral Thesis, Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

[21] P. Hamon and O. Pinette, "Les indicateurs Mesure & Analyse," Spirula22-06-2010 2010. Available: http://www.spirula.fr/commun/BonnesPratiques/Bonnes_Pratiques_Indicateurs.pdf

[22] P. Leeson, "Why the CMMI® does not work," presented at the SEPG Europe, Prague, Czech Republic, 2009.

[23] J. Armstrong, R. Barbour, R. Hefner, D. H. Kitson, "Standard CMMI SM Appraisal Method for Process Improvement (SCAMPI SM): Improvements and Integration," Systems Engineering, vol. 5, pp. 19-26, 2002.

[24] S. Pricope and L. Horst, "Towards a Systematic Metric Based Approach to Evaluate SCAMPI Appraisals," presented at the 10th International Conference Product-Focused Software Process Improvement, PROFES 2009, Oulu, Finland, 2009.

[25] T. Sunetnanta, N.-O. Nobprapai, O. Gotel, "Quantitative CMMI Assessment for Offshoring through the Analysis of Project Management Repositories " presented at the Third International Conference Software Engineering Approaches for Offshore and Outsourced Development, SEAFOOD 2009, Zurich, Switzerland, 2009.

[26] L. McCarthy, "Piloting Results-Based Appraisals," presented at the 9th Annual CMMI Technology Conference & User Group, Denver, Colorado, 2009.

[27] W. S. Humphrey, TSP SM: Coaching Development Teams: Addison-Wesley, 2006.

[28] R. E. Park, W. B. Goethert, W. A. Florac, "Goal-Driven Software Measurement - A Guidebook", CMU/SEI, CMU/SEI-96-HB-002, 1996.

[29] S. Grimstad and M. Jorgensen, "A framework for the analysis of software cost estimation accuracy," presented at the Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, Rio de Janeiro, Brazil, 2006.

[30] R. K. Yin, Case Study Research Design and Methods, Fourth ed.: SAGE, 2009.

[31] M. Philips, "CMMI V1.3 Planned Improvements," presented at the SEPG Europe 2010, Porto, Portugal, 2010.

[32] I. Lopes Margarido, J. P. Faria, R. M. Vidal, M. Vieira, "Towards a Framework to Evaluate and Improve the Quality of Implementation of CMMI® Practices," in PROFES, Madrid, 2012.