# DOING THINGS RIGHT OR DOING THE RIGHT THINGS?
## Proposing a Documentation Scheme for Small to Medium Enterprises

Josephine Antoniou, Panagiotis Germanakos and Andreas S. Andreou

*Computer Science Department, University of Cyprus, CY-1678 Nicosia, Cyprus*

Abstract:    Coping with the initial and finest systems' functionality and performance is indeed one of the major problems nowadays, due to the rapid increase and continuous change of customer demands. Hence, it is crucial to move on with a research analysis in an attempt to identify whether documentation, the most reliable source for preserving a software system's quality over the years, is properly created, updated and used in Small to Medium Enterprises (SME) operating in small EU markets, focusing both on the development process and the maintenance activities. Henceforth, the main objective of this paper is to propose a minimum documentation set required to fulfil both the Software Engineering principles and the SME practical needs by comparing literature suggestions with empirical findings. In further support of our documentation set suggestion, we present and discuss the results of a small survey conducted in nine IT-oriented SME in Cyprus and Greece.

## 1 INTRODUCTION

Software documentation may be conceived as the cornerstone of future maintenance activities. As such, the importance of a complete, updated and accurate documentation set is appraised by the Software Engineering community (e.g. (Sommerville, 2007; Pfleeger, 2001; Schach, 2005; Pressman, 2005). This paper aims to uncover the importance of documentation design and maintenance in Small to Medium Enterprises (SME) in small but growing European markets.

The methodology used is based on the construction of a questionnaire (Georgiou & Germanakos, 1999) and its further application on selected SME. in the area of software development and support. The SME that contributed to this work are located in Greece and Cyprus; nevertheless, the findings may be considered applicable to other developing European regions (e-MINDER). The results from the survey are analyzed and juxtaposed with an ideal documentation model resulting from a literature review, to closely examine any procedural documentation areas that need to be improved. This was not an easy task as the process should devise questions that would result in the gathering of the necessary information for inference purposes as regards to current and future documentation plans. The latter are associated with the profile and status of a SME as well as the comparison between existing documentation practices and those suggested by the literature. Specifically, documents created during system development, as well as ones actually used during system maintenance were targeted. Furthermore, the questionnaire aspired to examine the percentage of time spent on documentation during the different phases of the system development cycle to conclude on effects of possible time limitations for documentation development and usage (Baldassarre et al., 2006).

An extensive analysis of the data collected via the questionnaires is included in an attempt to identify the attitudes, reactions, suggestions and perceptions of both the management and rest employees. This work aspires to provide a way for assessing the SME's efficiency in "doing things right" and effectiveness in "doing the right things" within the given market space, which is crucial if a competitive advantage is to be achieved.

The rest of the paper is organized as follows: Section 2 provides a short related literature overview. The complete methodology of our investigation, including the analysis of collected results, is described in Section 3. Consequently, in

Section 4, a proposal for an "optimal" documentation set is presented based on conclusions drawn from the findings of our survey. Finally, we conclude in Section 5 and also provide some future research steps.

## 2 BASIC LITERATURE REVIEW

A brief literature review and an outline of various life-cycle models is presented in this section. Several researchers have focused on documentation issues such as accuracy, timeliness, flexibility and low cost (Arisholm et al., 2005). The authors of "System Documentation as Software" (Dibb et al., 1994) observe that due to new software technology, along with the greater productivity of higher level (4G) programming languages, there is a longer response time for documenting software. As a result, the dilemma of accuracy versus timeliness arises. Similar issues for a specific case study are looked at in (McGregor, 1982), a work concentrating on the important factor of identifying the recipients of the documentation set. Sickler (Sickler, 1982) encourages empirical research in order for creators of documentation to use the best available technology. This is because it has been observed that there is a lack of commitment on behalf of the programming personnel to prepare accurate documentation (Pfleeger & Kitchenham, 2001).

The *Waterfall* model was first put forward by Royce (Royce, 1970), involving a straightforward, trickledown approach to design and development. The resulting documentation provides a solid foundation of information for the writer as well as to track the progress of the project. It comprises a complete system documentation set and provides a theoretical model to be used as a basis of comparison with the documentation sets actually produced by the SME presented later in this paper.

The *rapid prototype* is a working model functionally equivalent to a subset of the product. With the *incremental* model, software is constructed step by step, in the same way that a building is constructed. The *extreme programming* model (Beck, 1999) is a new approach to software development based on the incremental model. The *synchronize-and-stabilize* model refers to a version of the incremental model (Cusumano & Selby, 1997). *The spiral* model is as a waterfall model with each phase preceded by risk analysis. Eventually, *object-oriented* models explicitly reflect the need for iteration, e.g the fountain model (Henderson-Sellers & Edwards, 1990).

The only documentation-driven model is the waterfall model which requires deliverable documents in order for each step of the system life cycle to be regarded as complete. Therefore, special emphasis on documentation is given compared to other models and that is why it perfectly satisfies the objectives of this paper. It should be noted at this point that we target companies belonging to one of the first three levels of the Capability Maturity Model, namely the Initial level (ad-hoc process), the Repeatable level (basic project management) and the Defined level (process definition), as we concentrate on small software providers with the typical characteristics of SME striving to survive in the software industry (Otoya & Cerpa, 1999).

## 3 A METHODOLOGY FOR DOCUMENTATION ANALYSIS AND ASSESSMENT

In order to analyze and assess the documentation set used in the targeted SME, nine companies situated in Greece and Cyprus were approached. A number of employees from each company (3 to 5 persons varying from IT developers, IT Consultants, Business Solutions Consultants, and Information & Media) were asked to fill in a specially prepared questionnaire, which comprises two parts: Part A aims at constructing the profile of an SME by identifying the type and size of each company thus forming the necessary background to record and assess managers' and employees' opinions and perceptions as regards system documentation. Part B aims to evaluate whether system documentation is equally treated and stressed emphasis upon during the development phases. The methodology used in the paper for information gathering combines both primary and secondary research methods (Glass V. G., 1976).

It was considered vital that the directors of the SME consented to the proposed research methods. All the relevant information is based around the activities and general functioning of a company in the scope of developing, assessing and maintaining systems' documentation delivering, eventually, coherently and cohesively the expected output to their customers. The authors' prime task was to accumulate realistic and spontaneous feedback with reference to the answers and attitudes by the respondents.

On the other hand, importance was given to the analysis and creation of a documentation model that

would bridge potential gaps and handle misleading issues of existing models used by the SME at the time of the survey.

The work reported in section 3 is summarized in the following steps: (i) constructing and disseminating the questionnaire, (ii) analyzing the results (iii) proposing an optimal documentation set, and, (iv) providing initial validation of the proposal.

## 3.1 Constructing the Questionnaire

A questionnaire may be defined as:

"*An instrument used for eliciting and recording responses in many, but not all, research projects employing the questioning approach*" (Dibb et al., 1994).

Although it may seem that designing questionnaires is a simple process, experienced researches are quick to point out that nothing is further from the truth (Dibb et al., 1994).

Our major concerns when designing our questionnaire with respect to the targeted SME revolved around the following aims:

(a) To identify the type and size of the participating companies using a system documentation process.

(b) To gain deeper understanding of managers and employees' opinions and perceptions for the development of system documentation.

(c) To evaluate whether documentation is part of every development phase and is equally treated and stressed emphasis upon.

(d) To identify the gaps and weaknesses of existing system documentation models and methodologies currently followed.

(e) To propose a minimum set of documents adequate and efficient for SME to create, maintain and use.

It is vital to clarify at this point that a "sloppy" questionnaire could lead to a great deal of distortion in the communication, from the researchers to the respondents and vice versa. Bearing in mind this, for each question included in the study the following attributes had to be considered:

- Conciseness
- Clarity
- Well structured
- Easy to respond
- Providing clear direction
- On target

In order for the above attributes to be met, a set of potential questions were formulated at first, targeting both the managers and the employees. Next, certain critical checks of the draft questions set were performed to decide the most appropriate form of each question (structured or non-structured, open-end, close-end, etc.), whether each question was relevant and properly worded to obtain meaningful and valid responses, whether the sequencing of the questions was likely to introduce any bias, and if the layout and appearance of the questionnaire was conducive to accurately and easily collect the data.

The questionnaire was divided into two parts: Part A included a set of questions related to the role of the respondents, their experience, their employment history in the SME and also to the size, type and number of projects undertaken, along with some additional development information (e.g. operating system platforms, programming languages etc.). Part B involved a full documentation scheme reported in various classic textbooks, the separate documents of which the respondents should mark to indicate production and usage during original development or/and maintenance activities.

## 3.2 Experimental Analysis – The Proposed Optimal Documentation Set

As previously mentioned, the questionnaire was distributed among the management and employees of each company, focusing on their culture and opinions, with main emphasis stressed upon perceptions and attitudes towards the system documentation design and maintenance. The next step after collecting the responses was to analyze and assess the information gathered.

In order to organize the results we targeted the two parts of the questionnaire separately. Part A provided results that helped us characterize the survey sample and identify the profile of each company in terms of personnel and nature of the projects undertaken. Part B provided the core information for the type of documentation actually used during development and maintenance activities.

Nine companies provided answers to the questionnaire with a total of 31 people, 11 of which are employed as management staff (Directors, Managers, IT Managers, and Project Managers) and 20 are employed as technical staff (Analysts, Programmers, and Team leaders). It is, therefore, straightforward to observe that the survey sample is well balanced, taking into account both the management and the technical perspective, something that plays its own role in deciding upon a globally accepted documentation set as will be described later on.

The employment period in the specific companies as well as the overall experience in the IT field for both groups of employees ranges from less than 1 year to more than 5 years, with most of the management staff having more than five years experience. As one can notice here, the experience is fair to good for technical personnel and good for managers. Both these factors (history, experience) play a decisive role in the objectiveness, credibility and significance of the responses. For example, staff with rich experience and long employment history is more appropriate to convey the necessary documentation information as, in general, they know the procedures and practices followed in the company better than newly recruited personnel and have been involved with an adequately large number of projects thus forming a more accurate picture of specific development aspects.

Overall, the percentage of time spent on the seven SDLC phases is uniformly distributed among company employees, i.e. the employees divide their time almost equally between the different SDLC stages. Therefore, one may not anticipate dramatic reduction in the time spent for documentation issues.
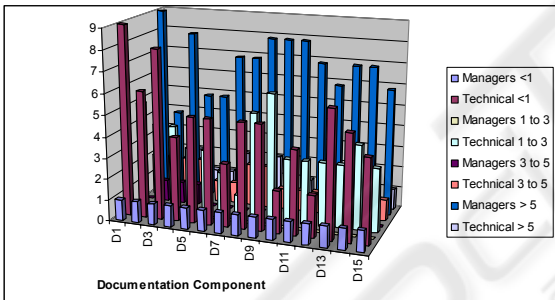


Figure 1: Documentation versus employment period for managers and technical employees in the specific company.
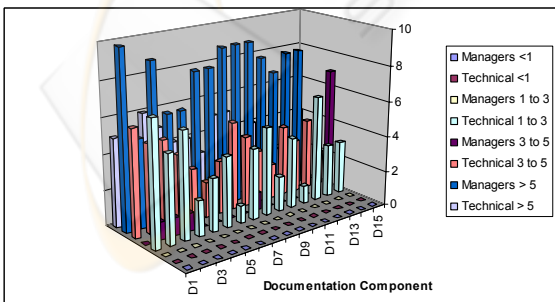


Figure 2: Documentation versus overall experience period for managers and technical employees in the IT field.

Concentrating on the role of the respondents and their employment history on one hand and their experience on the other in contrast with the creation and usage of the documents listed on the questionnaire, we mapped the associated survey results on figures 1 and 2. Further analysis of these results aimed at identifying how the usage of the proposed documentation set varies for different roles (i.e. management and technical staff) and experience both during system development (we will call this phase A) and system maintenance (we will call this phase B).

There is an apparent difference in the above two graphs. While during system development it seems that the managerial staff participates in the documentation development, this participation is considerably decreased in the usage of the documentation set during maintenance. This is quite expected as technical staff does the actual maintenance.

We decided to take into consideration the staff differences in the subsequent analysis by assigning different weights to the suggestions of the managerial and technical staff based on the employment and overall experience periods of the two categories of employees. Higher significance was given to the management during system development, while the answers of the technical staff were rated more important during the maintenance phase. Our goal was to conclude on the preference percentages for each of the documentation components by taking into account the type of personnel that is more involved with and has the highest stake in the particular phase. More specifically, the following weighting scheme in the scale [1, 5] (1 is the least significant) was applied:

Phase A – Development

*Management*
- $w_i$=5 for experience and employment over 5 years
- $w_i$=3 for experience between 3 and 5 years, and employment between 1 and 5 years

*Technical*
- $w_i$=4 for experience and employment over 5 years
- $w_i$=2 for experience between 3 and 5 years, and employment between 1 and 5 years

*Rest Management/Technical:* $w_i$=1

Phase B – Maintenance

*Management*
- $w_i$=3 for experience and employment over 5 years
- $w_i$=2 for experience between 3 and 5 years, and employment between 1 and 5 years

*Technical*

- $w_i$=5 for experience and employment over 5 years
- $w_i$=3 for experience between 3 and 5 years, and employment between 1 and 5 years
- *Rest Management/Technical:* $w_i$=1

After rating a certain respondent the assigned weight determined the times a document selected was inserted in the pool of the survey. For example, if document D1 was marked during maintenance and the responded was rated important with a weight equal to 5 then D1 was inserted 5 times in the general pool of selected documents. This way our sample takes into account the key concepts of role, experience and employment history.
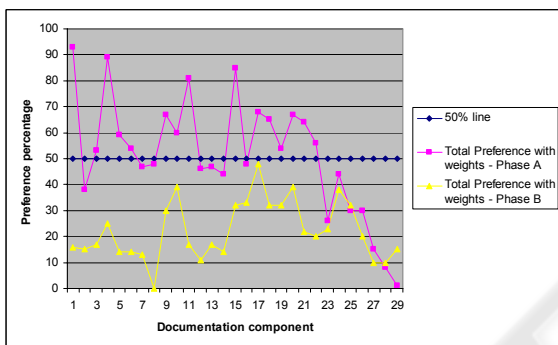


Figure 3: Total percentage of documentation components used for phase A (development) and B (maintenance).

The results regarding the preferences of the survey sample using the above weights are depicted in Figure 3.

Table 1: Documents gathering most responses with weighted scheme.

| System Life Cycle Phase or Activity | Proposed Document Components |
|---|---|
| *Requirements Definition* | System Overview |
| *Requirements Analysis* | Operations Overview |
| | Specification Analysis |
| | System Constraints |
| | Development Assumptions |
| | Data Flow or Object oriented Diagrams |
| | Data Dictionary |
| *Preliminary Design* | Design Overview |
| *Detailed Design* | Design Overview |
| | Design Description |
| | Data Interfaces |
| *Implementation* | Test Description |
| *System Testing* | Presentation/Discussion of Results |
| *Acceptance Testing* | System Overview |
| | System description |

The results shown in Table 1 indicate which of the documents listed in the questionnaire mostly created and used (i.e. received the vote of above 50% of the respondents) during the development and maintenance phase. The set comprises 15 documentation components covering all 7 phases of the SDLC, that combine high usage only during system development since none of the documents used for system maintenance received more than 50% of the votes.

It is apparent that during maintenance there is a reduced usage of the developed documentation set. Additionally, in order to detect potential documentation gaps, we first identified the documentation components missing from the list of preferred deliverables. This provided an indication of what the participating SME do not consider important and can skip in the documentation process, in order to effectively manage their time and resources.

To arrive to the proposed documentation set we have isolated the documents that the technical staff has indicated to mostly use during maintenance and checked whether they are covered by the deliverables gathering above 50% preference during system development. Since we have used weights to reflect overall experience and employment period, the preferences in Figure 3 reflect familiarity of respondents with the specific company's policies as well as experience in the field, achieving a more complete picture of both development and maintenance activities and capturing the business targets and goals of the SME. Since none of the documentation components used in phase B reached 50% of the employee votes, we have selected to check which of the documents reach and surpass the 30% of the votes. Lowering the preference threshold does not bias the results since it became evident that, as in many SME around the world, our sample of companies seems to be more production than maintenance-oriented. Thus, those employees that give emphasis on maintenance issues, although a minority, may pave the way for improving the processes followed within the SME by suggesting documentation usage for maintenance purposes.

A total of 10 documentation components appear to gather the 30% preference during the maintenance phase. However, there is an inconsistency between the selected set presented in Table 1 and these 10 deliverables. Out of these 10 deliverables, 3 are not included in the preferred set with the 50% threshold. These are: (a) Operations Overview of the Detailed Design phase (b) Detailed description of input and output of the Acceptance testing phase and (c)

Internal Storage Requirements of the Acceptance Testing Phase. Thus, to propose a complete documentation set covering all indicated needs of the participating SME we decided to add these 3 documentation components in the list presented in Table 1. Table 2 lists the complete proposed documentation set which comprises now 18 documentation components.

Table 2: Proposed ("optimal") Documentation Set.

| System Life Cycle Phase or Activity | Proposed Document Components |
|---|---|
| *Requirements Definition* | System Overview |
| *Requirements Analysis* | Operations Overview |
| | Specification Analysis |
| | System Constraints |
| | Development Assumptions |
| | Data Flow or Object oriented Diagrams |
| | Data Dictionary |
| *Preliminary Design* | Design Overview |
| *Detailed Design* | Design Overview |
| | Operations Overview |
| | Design Description |
| | Data Interfaces |
| *Implementation* | Test Description |
| *System Testing* | Presentation/Discussion of Results |
| *Acceptance Testing* | System Overview |
| | System description |
| | Detailed Description of input and output |
| | Internal Storage Requirements |

Our final task was to compare the final set with that suggested by literature and identify critical documents that might be missing from the proposed list. As one may observe, Table 2 includes indeed a more or less complete documentation basis with Requirements, Specification and Design fully covered. Implementation and Testing also seem complete although more documents may be added here (e.g. implementation strategy). It should be noted, however, that we did not include an exhaustive list of documents in the questionnaire as we were addressing SME with limited resources (time, budget, staff) and therefore we expected a bearing minimum in documentation management.

## 3.3 Experimental Validation

The only way to validate the proposed documentation scheme was to consult the SME participated in our survey. Therefore, we went back

to the nine companies and discussed mostly with employees (both management and technical personnel) that marked less documents than those in the "optimal" list, attempting to trace the reasons for not creating or using certain types of documents. It was interesting to notice that the majority of the discussion outcomes agreed on the following:

i. Time is very critical and often at the expense of proper documentation

ii. There is absence of a disciplined process to guide and monitor documentation related issues

Additionally, some of the respondents raised a significant issue that is worthy of mentioning here and was related to what type of documents is considered necessary when producing software and why. This issue roots back to the lack of fundamental knowledge and experience as regards the true value of documentation which is usually lost somewhere between the struggle to win new software contracts and the race to catch up with deadlines. In this context, insufficient and inefficient project management results in less control of documentation produced and used, something which was strongly confirmed by the survey participants.

As a final step we asked the respondents to study the proposed documentation scheme and comment on completeness, value and potential difficulties for practical application. The general feeling was that indeed the proposed scheme is critical and extremely helpful especially at the maintenance stage. Nevertheless, the price to be paid in time overhead was estimated to range between 5-15% of the total development time currently spent so as to add the proposed missing documents. Additionally, in 5 out of 9 companies surveyed at least one case was identified where the problems faced during enhancements or improvements of the delivered software product would have been avoided if the proposed documentation scheme was adopted and followed.

## 4 CONCLUSIONS

The initial question posed by this paper was whether things should be done "right" or instead doing the right things is better. Doing things right in our case requires for a complete documentation set according to the theoretical model that follows the Waterfall SDLC. However, as has been indicated several times in this work, SME are usually pressured to deliver a complete system with documentation within a certain time frame and with limited resources. It is, therefore, not possible to develop a complete

documentation set. This is seen from our survey results, as the companies do not develop all of the documentation components indicated by our theoretical model.

This work's ultimate goal was to present an empirical investigation of documentation issues and propose a minimum set of documents required to satisfy both the theory and the companies' needs. In the proposed set we included the documents mostly developed and used by the companies enquired and excluded those less developed and used. For the components that are sometimes used we draw a decision based on the importance of those documents in the theoretical model.

For the research conducted, nine SME have been approached and with the use of a comprehensive questionnaire, the authors were in a position to gather all the necessary information for the analysis process. These results, in combination with a balanced literature review, as well as the use of the Waterfall system development life cycle model as a guide, enabled the identification of the apt minimum set of documents to be used for the specific type of enterprises.

Future work will concentrate on extending the survey sample including SME with different profile characteristics. Our goal will be to link better certain SME organizational and procedural factors with the size and quality of documentation produced and utilized, as well as to trace the sources better and investigating in depth the origin of neglecting such a significant set of activities. Further, additional validation is planned where the obtained results will be cross-checked by new companies that will adopt them.

## REFERENCES

Arisholm E., Briand L.C., Hove S.E., and Labiche Y., 2005. The Impact of UML Documentation on Software Maintenance: An Experimental Evaluation, *Proceedings of the International Conference on Software Engineering Research and Practice (SERP'05)*, Monte Carlo Resort, Las Vegas, Nevada, USA, June 27-30, Technical Report SCE-05-11, Version 2.

Baldassarre M.T., Caivano D., and Visaggio G., 2006. Software Product Lines in Value Based Software Engineering. *Proceedings of the 10th International Conference on Evaluation and Assessment in Software Engineering (EASE 2006)*, April 10-11 (to appear).

Beck K., 1999. Embracing Change with Extreme Programming, *IEEE Computer*, pp. 70-77.

Boehm W. B., 1988. A Spiral Model of Software Development and Enhancement, *IEEE Software*, pp. 61-72.

Cusumano A. M., Selby W. R., 1997. How Microsoft Builds Software, *Communications of the ACM*, pp. 53-61.

Dibb, Simkin, Pride, Ferell, 1994. Marketing: Concepts and Strategies,.5th European edition, Houghton Mifflin.

e-MINDER: Electronic Commerce, Leveraging Network for Developing European Regions, http://www.eminderproject.com

Georgiou S., Germanakos P., 1999. Assessing the Services Quality and Relationship Marketing Parameters of *KAI Computer Services Ltd.* In the U.K. and the Factors that need to be Considered when Venturing into the European Market via the Internet, M.Sc. final year thesis, Leeds University Business School, Leeds, U.K.

Glass V. G., 1976. Primary, Secondary, and Meta-Analysis of Research . *Educational Researcher*, Vol. 5, No. 10, pp. 3-8 doi:10.2307/1174772.

Henderson-Sellers B., Edwards M. J., 1990. The Object-Oriented Systems Life Cycle, *Communications of the ACM*, Vol.33:9, pp. 142-159.

McGregor L. S., 1982. Improving Documentation.

Otoya S., Cerpa N., 1999. An Experience: A Small Software Company Attempting to Improve its Process, *IEEE Computer Society Washington, DC, USA*, pp. 153-160.

Pfleeger L.S., Kitchenham A.B., 2001. Principles of Survey Research. Part1: Turning Lemons into Lemonade. *ACM SÌGSOFT, Software Engineering Notes*, vol. 26 no 6, pp. 16-18.

Pfleeger S.L. 2001. Software Engineering, Theory and Practice (2nd edition). Prentice Hall: New Jersey.

Pressman RS. 2005. Software engineering: A practitioner's approach, 6th Ed. McGraw-Hill: London.

Royce W., 1998. Managing the Development of Large Software Systems: Concepts and Techniques, *1970 WESCON Technical Papers*, Western Electronic Show and Convention, Los Angeles, August 1970, pp. A/1-1 – A/1-9. Reprinted in *Proceedings of the 11th International Conference on Software Engineering*, Pittsburgh, May 1989, pp. 328-38.

Schach SR. 2005. Classical and Object-Oriented Software Engineering, 6th Ed. McGraw-Hill: London.

Sickler J.G., 1982. System Documentation as Software, *Proceedings of the 1st annual international conference on Systems documentation*, ACM Special Interest Group for Design of Communication, ACM Press, NY,USA, pp. 139-142.

Sommerville I. 2007. Software Engineering, 8th Ed., Addison-Wesley.