Preliminary findings from a survey on the MD state of the practice

Preliminary findings from a Survey on the MD* State of the Practice

Marco Torchiano, Federico Tomassetti
Department of Control and Computer Engineering
Politecnico di Torino, Italy
Email: marco.torchiano|federico.tomassetti@polito.it

Filippo Ricca, Alessandro Tiso, Gianna Reggio
DISI
Università di Genova, Italy
filippo.ricca|alessandro.tiso|gianna.reggio@disi.unige.it

***Abstract*—In the context of an Italian research project, this paper reports on an on-line survey performed with 155 software professionals, with the aim of investigating about their opinions and experiences in modelling during software development and Model-driven engineering usage. The survey focused also on used modelling languages, processes and tools.**

A preliminary analysis of the results confirmed that Model driven engineering, and more in general software modelling, are very relevant phenomena. Approximately 68% of the sample uses models during software development; Among them, 44% generate code starting from models and 16% execute them directly. The preferred language for modelling is UML but DSLs are used as well.

Keywords: Industrial survey, Model driven engineering, State of the practice

I. Introduction

Models can be used in different ways in software development. They can be only used as a tool to support design and discuss it with stakeholders, or they can be used as primary artefact to produce the final system.

Model driven engineering (MD*[1]) is a software development methodology defined in the last decade that is based on the concept of model. Model-driven engineering raises the abstraction level from the code to the models: in practice, the code is generated from the models by means of (semi) automatic transformations or the models are directly executed. The primary advantages claimed by MD* are improvements in productivity, portability, maintainability and interoperability [4]. Despite the many potential claimed benefits, there are few empirical works attempting to evaluate in a systematic way its adoption and to assess its effectiveness in an industrial context. For example, Mohagheghi and Dehlen claim that there is a need for more empirical studies evaluating MD* to prove the benefits of its adoption [5].

In this paper we try to fill this gap. Here we present some initial results from an on-line survey, performed with 155 Italian software professionals, about modelling during software development and MD*.

A recent related work [3] conducted a similar survey in the British industry. Differently from us, their study is not a survey penetration of MD* in industry; in addition the population in our study is wider, encompassing all kind of developers and not only MD* practitioners. Moreover, the research questions are different: they are more interested to identify the social, technical and organizational factors affecting the MD* success or failure and understanding how MD* is applied in the industry, while we are more interested in other aspects.

The paper is structured as follows. Sect. II presents the study definition and design of the conducted survey. Sect. III summarizes execution, while Sect. IV sketches some selected results. Sect. V discusses the preliminary results and, finally, Sect. VI concludes the paper.

II. Study definition and design

The survey presented in this report was conducted with the aim of investigating to what extent and how Italian companies deal with software models and MD*. The survey follows the same schema/framework of [6].

The goals of the survey are the following: (1) to understand the real penetration of software modelling in the Italian industry, (2) to understand and document if and how MD* is currently being applied in industry, (3) to understand which motivations led to or prevented modelling and MD* adoption and (4) to understand which processes, modelling languages and tools are used in this context.

The main effect studied in this work (i.e., the quality focus) is the maturity in the adoption of software models and MD* in software development. The perspective is of software engineering researchers, interested to understand what are the software models, notations and MD* techniques relevant in industry, and which of them should be better supported by proper methods and tools. The context consists of a sample of 155 Italian software professionals (e.g., project managers, architects, developers) belonging to several companies having a different size (e.g., small, medium and large) and a different application domain, mostly consultancy, IT companies, and companies having an IT department.

A. Research questions

Given the above goals, we aim at addressing three research questions:

RQ1: *What is the practical relevance of software modelling and MD* in the Italian industry?* We are interested

[1] we use the term MD* to indicate MDE/MDD/MDA

to know and identify how many, how and why professionals use software models and apply MD* techniques in industrial software development.

RQ2: *Which modelling languages and notations are used in the modelling phase and for MD*?* We are interested to know if professionals adopt general purpose modelling languages (e.g., UML) or domain specific languages (DSLs) and whether they prefer graphical or textual languages.

RQ3: *What kind of processes and tools does Italian industry adopt to support modelling and MD*?* This research question aims at investigating the current tool support and adopted processes during the modelling phase.

B. Areas of interest

We identified four areas of interest:

Information about the organization. In particular, we collected business domain, organization size, respondent's group/business unit size, kind of projects conducted and their average duration, average experience and skills of unit members.

Software models usage. We collected information about the usage of models during the software development process, motivation about their adoption or non-adoption, expected results and benefits of modelling (reached or not) and mean size/complexity of models. We also collected information about the application context of models, i.e., in which phases of the development they are used (e.g., during analysis and design) and for which reason (e.g., for modelling data logic only).

MD* adoption. We collected information about the adoption of MD* such as: years use of MD*, experience in MD* and percentage of projects in which models are used with respect to all the projects of an organization. Moreover, we also collected information about code generation (e.g., degree of code generation with respect to the final product), execution of models by means of specialized interpreters and usage of automatic transformations (model-to-model and model-to-text).

Used modelling languages and tools. We collected information concerning modelling languages, notations and tools used by professionals. In particular, we are interested to understand which type of models are used, i.e., graphical or textual and general purpose or domain specific. About the tools, we are mainly interested to understand how much is in percentage their adoption and which kinds of tools are the more used in the modelling context (e.g., graphical editors for producing models).

C. Identification of the target population and of the sample

The target population of our survey consisted of professionals and decision makers in software organizations in Italy. The sampling was performed in different ways: (i) using the Commerce Chamber database and randomly selecting some contacts; (ii) as a convenience sampling relying on the network contacts of the two research units involved (Torino and Genova); (iii) sending invitation messages on mailing lists concerning programming and software engineering; (iv) publishing a note on an on-line magazine for developers (programmazione.it); and (v) advertising it on a large Italian conference for developers (CodeMotion 2011). In total, we obtained 181 responses: 155 complete questionnaires and 26 incomplete (discarded before the analysis phase).

ID	Question
SUB03	In which domain does your company operate?
SUB04	How many employees does your company count, including both part-time and full-time staff?
DEV08	During the software development process are developed and used some models in our organization? For model we mean both diagrams (e.g., UML) and text according to any DSL
LAN25	Does you company use UML?
LAN26	Does your company use UML profiles?
MOD13	Which kind of IT personnel (e.g., developer) does typically write the models?

Table I
SAMPLE OF QUESTIONS IN THE QUESTIONNAIRE

D. Questionnaire Design

The questionnaire has been developed following a standard schema [2]. To receive more answers we decided that the questionnaire should take no longer than approximately 10-15 minutes to complete and it was designed accordingly[2]. It contains 31 open and multiple-choice questions but the total number of questions really answered by the respondents depend on their level of adoption in MD* and modelling (e.g, respondents non-adopting modelling in their software process answered to only 8 questions).

Corresponding to the areas of interest (Sect. II-B), the questionnaire consists of four sections. Section 1, common to all respondents, characterizes with 6 questions their organization. Section 2 collects, with 8 questions, information about the usage of models during the development process. In this section the main question is the DEV8 question (see Table I). Sections 3 and 4 — completed by subjects answered yes at question DEV8 only — collect, with 17 questions, information about MD* adoption and MD* tools/processes. Clearly, section 4 is completed only by subjects adopting MD*. Other examples of questions are given in Table I. The complete questionnaire (in Italian) is available at: *http://softeng.polito.it/tomassetti/MDQuestionnaire.pdf.*

III. SURVEY PREPARATION AND EXECUTION

The survey was put on-line since the 1st of February 2011 until the 15th of April 2011 (two and a half months). The procedure followed to prepare, administer, and collect the

[2]The time for completing the questionnaire was on average less than 6 minutes

questionnaire data is made up of the following five main steps:

Preparation and design of the questionnaire. To design the questionnaire we adopted the Goal-Question-Metric approach [1]. From the goals presented in Sect. II we derived the questions and the metrics necessary to answer them. Three different pilot questionnaires were conducted with software professionals before putting on-line the final questionnaire.

On-line deployment. Once ready, the questionnaire was uploaded to the LimeSurvey survey tool[3] to permit the automatic collection of data. Some effort was required to implement data validation by means of Javascript (the language used by LimeSurvey) for the questions needed a constraint (e.g., "this answer must be an integer between 5 and 10").

Invitation to participate. Organizations were sampled as detailed in Section II-C. Once the contact persons were identified, we invited people, via email, to register to the survey and to complete the on-line questionnaire. We also broadcast invitation on selected mailing lists and on-line magazines/conferences including in the message a link to a registration form where the participants could register themselves and compile the questionnaire.

Monitoring. During the data capture phase, the research unit of Torino monitored the progress of the questionnaire submission. This allowed us to send selective reminders to contacts who did not respond or did not completed the questionnaire. Some people that reported some difficulties about the questions, because of internal policies of the company or because involved in very different projects with different companies at the same time, asked to us some suggestions on how answering.

Data analysis. After questionnaires have been collected, analyses were performed with the aim of answering the research questions formulated in Sect. II-A.

IV. Analysis

For space constraints, in this section only a subset of the questions of the questionnaire will be considered.

A. Demographics

The most of the companies where the respondents work are in the IT domain (104), then come services (15) and telecommunications (11). The distribution of the companies size where the respondents work is presented in Figure 1.

B. RQ1: Relevance

Among the 155 complete questionnaires, we have 20 companies (13%) always using modelling, 85 (55%) using it sometimes, and 50 (32%) never using modelling.

Such proportion varies significantly as the size of company varies, as illustrated in Figure 2. We observe that the use of modelling (i.e., always + sometimes) is positively correlated with the size (i.e., it is more frequent in large companies) with two exceptions: individual professionals and micro-sized companies. Instead, the systematic use of modelling (i.e., always) is more frequent in micro and large companies; medium-sized companies and individual professionals adopt it occasionally.

In general, we observe that the most cited motivations for not using modelling are that the respondent believe models require *too much effort* (50%) and could turn out to be *not useful enough* (47%). Moreover, 10 respondents (7%) indicated the lack of time as one of the reason for not using models.

In terms of relevance we should also note that 30% of the surveyed developers – that is 44% of those using modelling – generate some amount of code from them (an average of 41% of modelled components). As far as run-time model interpretation is concerned, 11% of developers produced models that were interpreted at run time (16% of developers doing modelling).

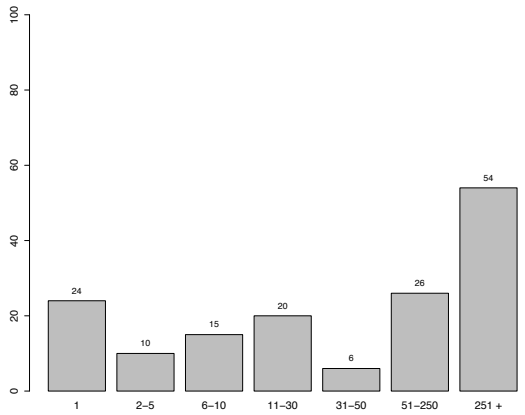

Figure 1. Size of respondents' companies

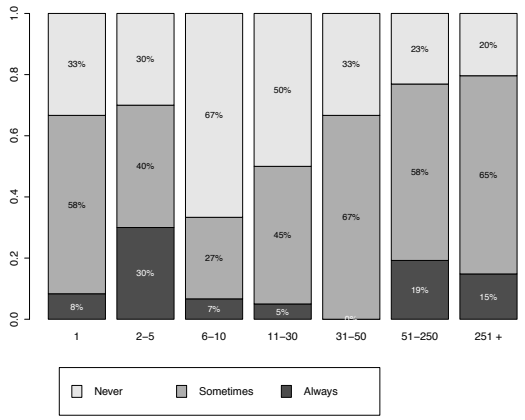

Figure 2. Proportion of modelling usage per company size

[3]http://www.limesurvey.org/

C. RQ2: Languages and notations

We found that 76% of professionals (80 out of 105) creating models use UML; among them 11% use also UML profiles, 51% do no use them, and the remaining 38% state to not know if them are used in their organization.

In our sample, only 21% of professionals using models appear interested in Domain Specific Languages (DSLs). Among them 50% use a purely textual notation, 23% a purely graphical one, and 27% a mix of textual and graphical notations.

D. RQ3: Processes and tools

As far as processes are concerned, we investigated which role typically performs the modelling. Usually modelling is performed by multiple roles at the same time. For this reason, to the corresponding question in the questionnaire several answers were permitted (e.g., developer and project manager). Results show that architects or project managers perform modelling in 76% of the cases, developers write models in 72% of the cases, while domain experts are involved in just 11% of the cases.

As far as model manipulations, it appears the only 10% of modelers perform automatic transformations between models. While 16% of the modelers developed editors or other support tools for models. Since models can evolve, 53% of the modelers adopt versioning of models.

V. Discussion

The methods and technologies related to modeling are extremely relevant in the Italian industry involved in software development: 68% of respondents use them. Typically modeling is widely applied in large companies while medium and small companies adoption is not systematic and rates are relatively limited. Micro-sized companies (at least in our sample) represent an exception: they exhibit the larger rate of systematic adoption of modeling techniques.

The two most frequent causes that prevent adoption of modeling are the perceived limited effectiveness (lack of usefulness) and the fear of large investments (too much effort). An important issue, especially among the companies that never used modeling, is the lack of specific skills.

General purpose techniques appear dominant, with DSLs concerning just 21% of the cases. UML is largely used among modeling adopters, although at a basic level, in fact just 11% of UML users take advantage of advanced features such as profiles. Most modeling initiatives are targeted at technical users, while domain experts have been involved in just 11% of the cases.

In terms of advanced use of modeling that qualifies as MD*, 44% of the companies doing modeling, generate a significant portion of code from models: 41% of the modeled features on average. Run-time execution of models is adopted by a smaller portion of companies using modeling: just 16%.

A possible threat to the validity of our study may derive from the auto-exclusion of developers not interested in modeling and MD*: they could have avoided answering since the topics of the survey are not relevant for them, as a result we could have an overestimated relevance of modeling. Nevertheless we are confident that the survey provides a faithful representation of the companies that perform some kind of modeling.

VI. Conclusions and future work

In this paper we presented preliminary results from a survey performed to analyze how Italian companies use and perceive modeling with different goals and levels of expertise. We found that modeling is adopted by a relevant portion of the respondents and among them a large number use MD* techniques, i.e., code generation, model transformation, and run-time model interpretation.

As future work we will complete analysis of data available to examine how the different kinds of issues in adopting modeling and MD* are related to the characteristics of the companies. We also plan to compare the state of adoption in Italian companies to the situation in other countries both replicating this study in other nations or using available reports (e.g., [3]).

Acknowledgements

We would like to thank the people who took the time to take part in our survey and the organizers of CodeMotion for the help provided to diffuse our survey.

References

[1] V. Basili, G. Caldiera, and D. H. Rombach. *The Goal Question Metric Paradigm, Encyclopedia of Software Engineering*. John Wiley and Sons, 1994.

[2] M. Ciolkowski, O. Laitenberger, and S. Biffl. Software reviews: The state of the practice. *IEEE Software*, 20(6):46–51, November/December 2003.

[3] J. Hutchinson, J. Whittle, M. Rouncefield, and S. Kristoffersen. Empirical assessment of MDE in industry. In *Proc. of the 33rd International Conference on Software engineering*, ICSE '11, pages 471–480, New York, NY, USA, 2011. ACM.

[4] A. G. Kleppe, J. Warmer, and et al. *MDA Explained: The Model Driven Architecture: Practice and Promise*. Addison-Wesley Longman Publishing Co., Inc, 2003.

[5] P. Mohagheghi and V. Dehlen. Where is the proof? - a review of experiences from applying mde in industry. In *Proc. of the 4th European conference on Model Driven Architecture: Foundations and Applications*, ECMDA-FA '08, pages 432–443, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] M. Torchiano, M. Di Penta, F. Ricca, A. De Lucia, and F. Lanubile. Migration of information systems in the italian industry: A state of the practice survey. *Inf. Softw. Technol.*, 53:71–86, January 2011.