Proposals for supporting model-driven development users: a systematic mapping study
Protocol version: v1.2

Motivation:

We are looking for literature reporting assistants/methods/tools/approaches for supporting model-driven development (MDD) tools' users. We focus on proposals applicable to MDD tools that allow users to establish models and then generate code from them, manually or automatically. Consequently, we propose a systematic mapping regarding this topic. We establish the systematic mapping protocol in order to select a set of primary studies for concluding about a set of research questions.

Research questions:

- **RQ1: Which kind of proposals exists?** With this research question we seek to know the distribution of solutions proposed in the literature. These include tools, methods, frameworks, and intelligent assistants, among others.
- **RQ2: What limitation(s) do the authors establish in their work?** We aimed to know the limitations that the authors of the proposals establish on their own work (if any). This question was obtained from the validation with the target audience. We will use this information to find out the frequency of limitations among the proposals.
- **RQ3: What goal(s) does the proposal aim to achieve?** We aimed to know the goals that the authors of the proposals set for their own work (if any). This question was obtained from the validation with the target audience. We will use this information to know the frequency of goals between proposals.
- **RQ4: Are the proposals empirically evaluated?** We want to know whether the proposals available in the literature have been empirically evaluated. We will use this information to find out the frequency of empirical evaluation of the proposals.
    - **RQ4.1: Which quality attributes have been empirically evaluated?** We want to know what quality attributes have been considered in the empirical evaluations. We will use this information later to classify the attributes in Moody's (2003) framework.
    - **RQ4.2: How is the concept "user" defined on the proposals?** We want to know how authors define users in their proposals. We will find this definition in the profile of the subjects used for the empirical evaluation (if any). We will also obtain this information from the name adopted by the authors to refer to the users of their proposal. We will use this information to list the different user profiles that the authors have adopted in the proposals and, thus, be able to group them.
    - **RQ4.3: Which environment is used to deploy the proposal?** We want to know in which environment the authors deploy their proposals (if any). We refer an environment to the tool (or toolset) in which the proposal is run and not the technology with which it is developed. For example, if a proposal is developed using an Eclipse toolset and is executed in Acceleo (an MDD tool), the environment where the proposal is deployed will be Acceleo. We will use such information in order to obtain more information about the potential users of the proposals by using the environment definition.

Search strategy:

First, we perform a database search by using a search string in six scientific databases: IEEE, ACM, Scopus, ScienceDirect, SpringerLink, and Web of Science. Then, we select a set of primary studies based

on the inclusion/exclusion criteria. Moreover, we choose a sample of such primary studies by using the quality assessment (e.g., the top 25 of quality assessment). Finally, we use such sample in order to perform a snowballing search.

Search strategy – Database search:

We use PICO (Population, Intervention, Comparison and Outcomes) in order to identify keywords and formulate a search string from research questions.

- **Population**: In software engineering, population may refer to specific software engineering role, category of software engineer, an application area or an industry group (taken from Petersen et al. (2015) citing Kitchenham and Charters (2007)). In our context, our population is a software engineering area: model-driven development.
- **Intervention**: In software engineering, intervention refers to a software methodology, tool, technology, or procedure. In our context, we are looking for proposals that support model-driven development users—e.g., methods, tools, frameworks, intelligent assistants, etc.
- **Comparison**: We compare the different features for each proposal. Even though we compare such proposals, our mapping study is not limited to studies with empirical evaluation. We also include proposals without empirical evaluation.
- **Outcomes**: We do not consider measurable outcomes of empirical studies. However, we extract data regarding such empirical studies specially focused on the users and metrics that they define for validating their proposal.

We identify some keywords by using the previous PICO. We group them into three different sets:

- **Set 1:** Terms comprising the population:
    - Model-Driven Development (MDD); Model-Driven Engineering (MDE); Mode-Driven Architecture (MDA); Model Driven Software Engineering (MDSE); Model-Based Software Engineering (MBSE); Low code; Non code.
- **Set 2:** Terms related to the intervention:
    - (Support*; assist*; help*; maintain*; promote*; ease*; facilitate*; simplify*)
      NEAR TO
      (user; developer; tester; software engineer; architect; usability, usage)
- **Set 3:** Terms related to the comparison:
    - Approach; proposal; concept; idea; method; manner; technique; procedure; program

Finally, we establish a search string based on such term sets:

("Model-Driven Engineering" **OR** "MDE" **OR** "Model-Driven Architecture" **OR** "MDA" **OR** "Model Driven Software Engineering" **OR** "MDSE" **OR** "Model-Driven Development" **OR** "MDD" **OR** "Model-Based Software Engineering" **OR** "MBSE" **OR** "Low code" OR "Non code")
AND
("support*" **OR** "assist*" **OR** "help*" **OR** "ease" **OR** "facilitate*" **OR** "simplify*")

NEAR TO

("user" **OR** "developer" **OR** "tester" **OR** "software engineer" **OR** "analyst" **OR** "architect" **OR** "usability" **OR** "usage")
AND
("approach" **OR** "proposal" **OR** "concept" **OR** "idea" **OR** "method" **OR** "manner" **OR** "technique" **OR** "procedure" **OR** "program" **OR** "assistant")

Search strategy – Snowballing:

We propose a forward and backward snowballing in order to discover other proposals not present on the data base search results. We continue the Snowballing procedure until we do not discover any new primary studies. We use the procedure proposed by Wholin (2014):

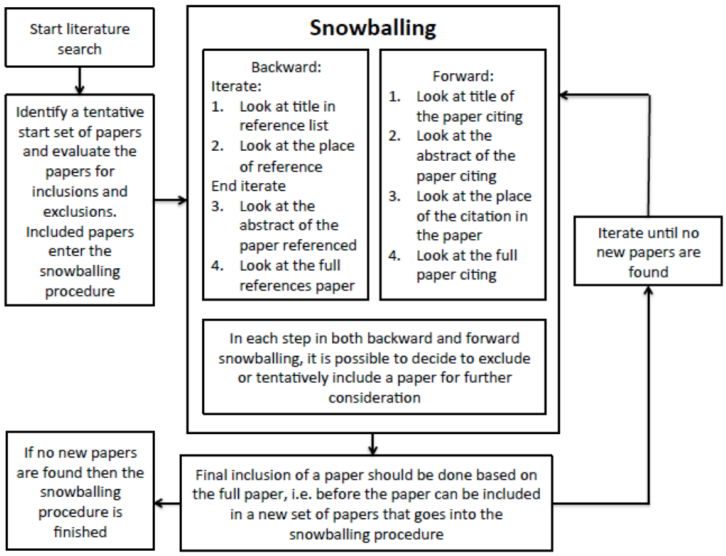


Figure 1. Snowballing procedure.

Inclusion and exclusion criteria:

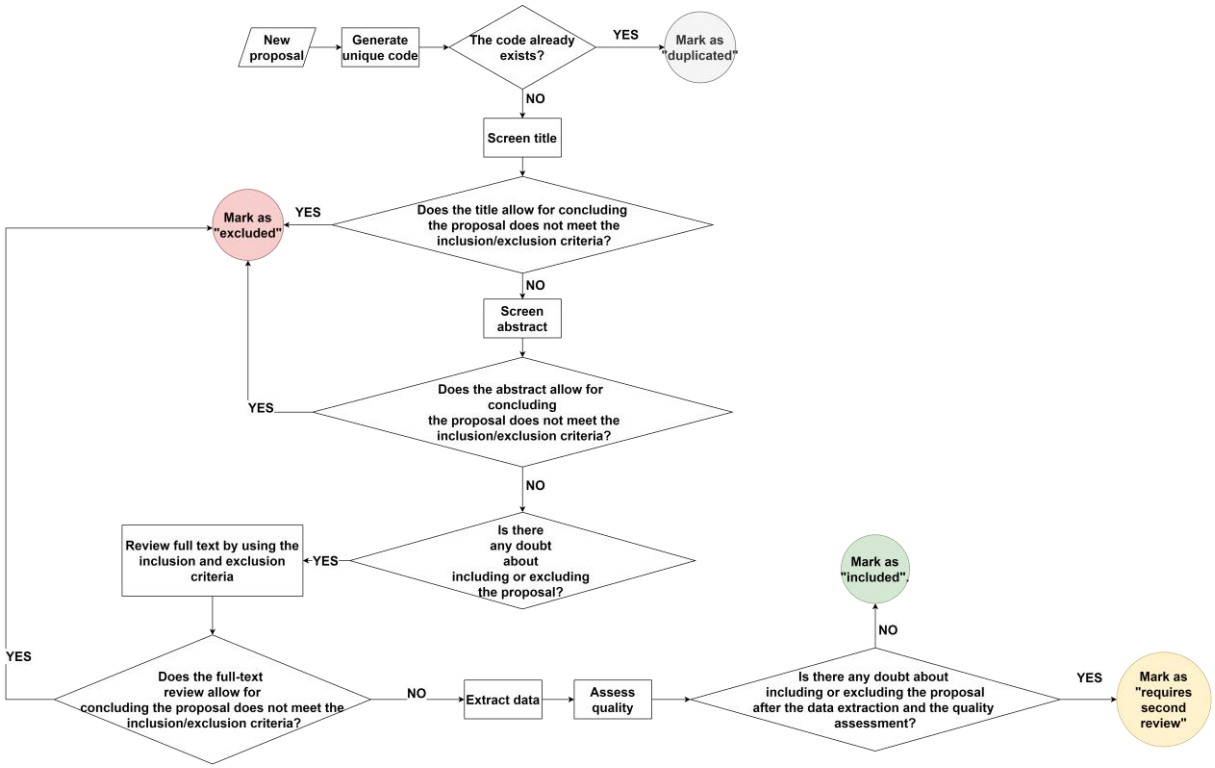
We propose the following procedure in order to review each study result from the different search strategies:

Initially, we need to check whether the study has been previously reviewed. So, we propose to generate a unique code for each article by using the following template: *First and second author last name + two first last name letters of the remaining authors (if present) + year.*
Example:
Title: Model-based tool support for the development of visual editors: a systematic mapping study.
Authors: David Granada, Juan M. Vara, Francisco Pérez Blanco, and Esperanza Marcos.
Year:2017
Code: GranadaVaraPeMa2017

For studies with only one author, the full name of the author followed by the year for generating the unique code.

If the proposal has already been reviewed, it should be marked as "duplicate". Then, the reviewer should screen the proposal title, checking if it is enough to discard the proposal based on the inclusion/exclusion criteria. If it is not possible to discard the proposal based on the proposal title, the reviewer should screen the abstract. The reviewer should be sure if it is possible to discard the proposal by using the inclusion/exclusion criteria. If the proposal was not discarded after reviewing its title and abstract, the reviewer wonder if there is any doubt for including or excluding the study. If there is any doubt, the reviewer performs a full-text review, always based on the inclusion/exclusion criteria. If the study was not discarded in the previous filters and meets the inclusion criteria, then the reviewer should extract the data and assess the study quality. Finally, if there is any doubt after extracting the data and assessing the study quality, other reviewer should review the study in order to come to an agreement. If there are no doubts about including the study after the data extraction and quality assessment tasks, the reviewer includes the study in the primary study set.

We show such procedure in the following flow diagram:



We propose the following inclusion criteria:

- Does the paper define a specific proposal for supporting MDD users? (I1) Not a compilation of proposals like a literature review, mapping, or systematic literature review.
- Is the proposal designed to support MDD users? (I2) Specifically, we are looking for proposals that allow for supporting/easing/facilitating the execution of MDD tools tasks by MDD users. This includes---but is not limited to---modeling, traceability, debugging, model repair, model verification, and model validation, among others. Proposals that focus on supporting/easing/facilitating the DEVELOPMENT of MDD tools rather than their USE do not meet this inclusion criteria---e.g., proposals that focus on the development of model-to-model transformations to generate new modeling tools do not meet this inclusion criterion.

We propose the following exclusion criteria:

- The proposal does not focus on proposing or evaluating an approach for supporting MDD users (E1). This exclusion criteria aims for differentiate proposals to support the use of MDD tools from the MDD tools themselves. Proposals that show a new MDD tool or MDD applied to a new context should be excluded if their focus is not on supporting the users of an existent MDD tool.
- The proposal is not related to software engineering (E2)
- The proposal is not written in English (E3)
- The proposal is not a peer-reviewed publication (i.e., the proposals is not reported on a journal article, book section, chapters in research books, or conference proceeding) (E4)
- The proposal's full text is not available (E5)

Notice that each exclusion and inclusion criteria have a code following the *I#* and *E#* template. When the reviewer discards a proposal, he/she mark the reason for excluding it by using the corresponding code. Thus, we track the reasons for excluding the proposals during the inclusion and exclusion process.

Data extraction:

We extract the data in order to answer the proposed research questions. Consequently, we propose the following template for extracting the study data:

https://zhaw-my.sharepoint.com/:x:/r/personal/mosq_zhaw_ch/_layouts/15/Doc.aspx?sourcedoc=%7BD894F59F-6D13-43E0-B5D6-CD9018AB946C%7D&file=DataExtraction.xlsx&action=default&mobileredirect=true

Quality assessment:

We propose the following Likert-Scale Questionnaire based on the Ormeño and Panach (2013) systematic mapping protocol:

Subjective question (Yes = 1; Partially = 0; No= -1)

1. Is the proposal clear?
2. Are the limitation(s) clear?
3. Are the goal(s) clear?
4. Are the publications tools downloadable?
5. Is there a clear case study or example illustrating the proposal?
6. Is the whole proposal empirically validated?
7. Are the users clearly described?
8. Are the results clearly explained?

Objective Questions

9. Has the publication been published in journal or conference proceedings?
   Very important = 1; Important = 0; Not important = -1
10. Has the publication been cited by other authors?
    More than 4 = 1; Between 2 and 4 = 0; Less than 2 = -1

Search evaluation:

We propose to evaluate the search by using the Fleis Kappa (K) statistic. We randomly sample a set of papers, then each reviewer reviews it and, finally, they discuss their decisions of inclusion or exclusion in order to come to an agreement. We calculate the K statistic based on the number of inclusions and exclusions after the initial screening decision and before reconciliation. Such statistic allows us for assessing the inclusion/exclusion reliability. Also, we should have at least 3 reviewers in order to improve the inclusion/exclusion reliability at the search evaluation. We interpretate the K statistic result based on Belur et al. (2021) citing Landis and Koch (1977):

Table 4. Interpretation of κ Statistic.

κ Score	Interpretation
<0	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Near perfect

Source: Landis and Koch (1977).