



HAL
open science

Literature Classification Data for a Systematic Mapping Study on Multi-Paradigm Modeling for Cyber-Physical Systems

Ankica Barisic, Antonio Cicchetti, Ivan Ruchkin, Dominique Blouin

► **To cite this version:**

Ankica Barisic, Antonio Cicchetti, Ivan Ruchkin, Dominique Blouin. Literature Classification Data for a Systematic Mapping Study on Multi-Paradigm Modeling for Cyber-Physical Systems. 2021. hal-03338607

HAL Id: hal-03338607

<https://hal.science/hal-03338607>

Preprint submitted on 8 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Literature Classification Data for a Systematic Mapping Study on Multi-Paradigm Modeling for Cyber-Physical Systems

Authors

Ankica Barišić^{1,2}, Antonio Cicchetti³, Ivan Ruchkin⁴, Dominique Blouin⁵

Affiliations

1. NOVA LINCS, University NOVA of Lisboa, Portugal
2. Université Cote d'Azur, I3S, INRIA, France
3. Mälardalen University, Västerås, Sweden
4. University of Pennsylvania, US
5. LTCI, Telecom Paris, Institut Polytechnique de Paris, France

Corresponding author(s)

Ankica Barišić (a.barisic@campus.fct.unl.pt), Antonio Cicchetti (antonio.cicchetti@mdh.se), Ivan Ruchkin (iruchkin@cis.upenn.edu), Dominique Blouin (dominique.blouin@telecom-paris.fr)

Abstract

This dataset includes the set of papers retrieved to perform a systematic mapping study (SMS) on multi-paradigm modeling (MPM) for cyber-physical systems (CPS). Moreover, it includes several tables that map the literature into several perspectives, notably: used modeling formalisms and processes, part of the CPS addressed by the research, domain of expertise of paper authors, and relevance of the papers at review date. The set of papers is selected over a period ranging from 2006 to 2021, according to publication dates. The selection of the papers and their mapping has been performed by means of a rigorous process based on precise research questions and peer-review. Furthermore, the process has been supported by a web-based survey management application. Both the selection of existing publications and their mappings by means of the included perspectives provide interested readers/researchers with interesting data potentially re-usable for multiple purposes, notably: analyzing the progress of research on modeling of CPS, studying further the papers pertaining to a specific (set of) characteristic(s), performing a follow-up study related to other development technologies.

Keywords

Multi-paradigm modeling, cyber-physical systems, model-driven engineering, modeling formalisms, modeling tools, systematic mapping study

Specifications Table

| | |
|---------------------------------------|---|
| Subject | Computer Science (alt. Software) |
| Specific subject area | Model-Driven Software Engineering of Cyber-Physical Systems |
| Type of data | Tables |
| How data were acquired | Systematic Mapping Study |
| Data format | Raw Analyzed Filtered |
| Parameters for data collection | <p>The set of considered papers (Raw data) was retrieved by means of a well-defined search-string run over the research databases: ("multi-paradigm" OR "multi-formalism" OR "heterogeneous formalism" OR "unified modelling formalism" OR "multi-model language") AND (((("cyber physical" OR CPS OR "cyber-physical" OR cyberphysical OR smart) AND system*) OR CPS)AND ("modelling approach" OR "modeling approach" OR "integrate modelling" OR "integrate modeling" OR "model driven" OR "model-driven") AND ("software engineering" OR "software system"))</p> <p>The second part of Raw data was manually extracted from another SMS reported in [1] for the period 2011-2017.</p> <p>Starting from the initial set of papers, we applied the following criteria: Exclusion Criteria:</p> <ul style="list-style-type: none"> ● E1 - Informal literature (powerpoint slides, conference reviews, informal reports) and secondary/tertiary studies (reviews, editorials, abstracts, keynotes, posters, surveys, books).; ● E2 - Duplicated papers; ● E3 - Papers that did not report a method for MPM approach for CPS; ● E4 - Papers with the same content in different paper versions; |

| | |
|--|--|
| | <ul style="list-style-type: none"> ● E5 - Papers written in other than the English language; ● E6 - Purely hardware, or electrical engineering perspective papers; <p>Inclusion Criteria:</p> <ul style="list-style-type: none"> ● I1 - Publication date from 1/1/2006 ->; ● I2 - Explicit mentioning of multi-paradigm modeling of cyber-physical systems; ● I3 - Papers that report a methodology, metric or formalism for modeling of CPS; ● I4 - Analysis of relevant application domains for MPM of CPS. <p>For an included set of primary studies, we analyzed data based on the following questions:</p> <ul style="list-style-type: none"> ● Does the paper report a multi-paradigm modeling approach? <ul style="list-style-type: none"> ○ To which extent does the paper present a complete model-driven development approach? ○ Which part of the CPS is modeled? ○ Which formalism(s) is(are) used for modeling the CPS? ○ What is the integration mechanism for the presented models? ● How was the approach presented? <ul style="list-style-type: none"> ○ Does the paper report a model/meta-model? ○ Does the paper report a tool/language? ○ Does the paper report the model of the adopted process? ● Is the approach domain-specific? <ul style="list-style-type: none"> ○ Which application domain is addressed? ● Does the paper report the actors/stakeholders involved in the modeling of the CPS? <ul style="list-style-type: none"> ○ Does the paper report the modeler’s background knowledge? ○ What is(are) the technical background of the authors? <p>We also performed a quality assessment based on the following questions:</p> <ul style="list-style-type: none"> ● What is the ranking of the paper according to the conference/journal where it was published? ● What is the relevance of the paper according to its related citations? ● How clearly is the problem of the study described? ● How clearly is the research context stated? ● How rigorously is the method evaluated? ● How explicitly are the contributions presented? ● How explicitly are the insights and issues for future work stated? |
| <p>Description of data collection</p> | <p>Conducting the SMS involved six phases: (1) Identifying relevant research - We obtained data for this SMS twofold: Using automatic search over digital libraries and manually extracting the</p> |

primary studies from another SMS reported in [1], which had similar search keywords for obtained data and inclusion criteria. See Part 1 archive for the detailed definitions.

(2) **Selecting primary studies** - All primary studies passed the 'screening' phase based on the title, abstract, and keywords and were imported to the ReLis platform [15]. Each paper was analyzed by two different persons that independently evaluated the inclusion/exclusion criteria. In the case of conflicting opinions, a third person was involved. See Part 2 archive for the list of primary studies.

(3) **Performing data analysis** - Obtained primary studies were assigned randomly to different reviewers for analysis. A private shared group was created in Mendeley Library [2]. All bibliographic meta-data for primary studies were imported to this group and were set to have the same citation key (unique identifier) as in the ReLis system. Reviewers were invited to join this group and were asked to upload full papers. Reviewers were provided with a set of instructions and videos about how to use the ReLis system, how to extract data, and when to exclude papers. They were able to monitor the individual and overall classification status.

(4) **Assessing the quality of studies** - Each reviewer reported on the quality of the study based on a set of seven predefined quality assessment questions. Quality questions were defined to observe the impact of the studies. It was mandatory for each reviewer to select the answer to a question based on the predefined scale, except in a case of assessment related to the number of citations, which was performed independently in June 2021. The number was obtained from Google Scholar [3] independently to have a fair number retrieved in a set time frame. To reflect the confidence of reviewers, we used two self-assessment criteria: reviewers' confidence about the content of the study and reviewers confidence about the quality of the study. When a reviewer was not confident about a paper under review, an additional reviewer was assigned to review the paper and the assessment scores were then discussed until consensus was reached.

(5) **Extracting required data** - In the ReLis system, the data extraction form was generated by using the integrated domain-specific language for specifying forms. The form consisted of the questions and descriptions from the data extraction form (defined by protocol), as well as from the quality assessment questions. All the Boolean and multi-choice questions were mandatory to be answered by reviewers before they were able to submit their classification. All questions had at least one text paragraph so reviewer observation could be described freely. After submitting the classification, reviewers were able to edit their answers if necessary. They could follow their progress, as well as the progress of other reviewers and competition status in the ReLis 'Data Extraction' dashboard.

(6) **Synthesizing the data** - Extracted information was cross-reviewed by all experts with the aim of identifying possible clerical errors.

At least two reviewers were assigned to perform manual qualitative data analysis for each column in Part 2 archive, in order to cluster the

| | |
|---------------------------------|---|
| | <p>approaches in groups by similarity. Different reviewers described approaches with different terms that were in practice equivalent, and these synonymous terms were merged. Reviewers processed the extracted data to identify unambiguous labels for each question under analysis and a tentative cluster name. Finally, reviewers iteratively refined and merged the data until a small number of well-defined clusters emerged. This process was performed independently for each question. Throughout the iterations, the researchers cross-checked their work and synchronized their definitions of the clusters. See Part 3 archive for the clustered results of data synthesis.</p> |
| Data source location | <p>We extracted the primary studies from the following research databases: ACM Digital Library (ACM) [4], IEEExplore (IEEE) [5], Science Direct (SD) [6], Springer Link (SL) [7], Scopus [8] Also, from [1], to which we applied our inclusion and exclusion criteria</p> |
| Data accessibility | <p>Repository name: Multi-Paradigm Modeling for Cyber-Physical Systems literature - Mendeley Data Repository Data identification number: 10.17632/jy6ww3hmyw Direct URL to data: http://dx.doi.org/10.17632/jy6ww3hmyw</p> |
| Related research article | <p>A. Barišić, I. Ruchkin, D. Savic, M. Abshir Mohamed, R. Al-Ali, L. W. Li, H. Mkaouar, R. Eslampanah, M. Challenger, D. Blouin, O. Nikiforova, and A. Cicchetti: Multi-Paradigm Modeling for Cyber-Physical Systems:A Systematic Mapping Review, In submission to Journal of Systems & Software.</p> |

Value of the Data

- The development of Cyber-Physical Systems (CPS) is a complex endeavor that requires engineers to manage the interplay of many interdisciplinary design methodologies. There have been increasing research efforts and tools targeting different aspects of CPS development, making it difficult for researchers, practitioners, and outsiders to orient themselves in the available solutions and open challenges. This data provides such an orientation.
- Engineers interested in working with CPS can benefit from the proposed multifaceted classifications of the existing literature. They can use this data as a guide towards specific efforts and solutions for their problems (e.g., the design of CPS hardware exploiting a specific modeling formalism). Researchers can also refer to this data to understand and analyze the state-of-the-art for specific research problems (e.g. designing the behaviour of a CPS).

- This data can be used both in part or as a whole to motivate and propose studies or solutions for specific CPS challenges, as well as to perform follow-up literature reviews and mapping studies on modeling of CPS.

Data Description

The provided data archive is split into three parts. Each part is a spreadsheet containing multiple tables. The first table of each spreadsheet is a metadata table, explaining each table and each column in that part. Below we provide a short summary of the data tables in each of the three parts.

Part 1 - Protocol description.xlsx is a comprehensive description of the methodology of this systematic mapping study. Part 1 contains the following data tables:

- SMS Objective: the goal and justification of this study.
- Research Questions: the research questions defining data collection and application of the Q2.3 x Q4: Correlation between processes and rolesPICOC criteria [9].
- Search Keywords and Queries: the keywords that we used to search for primary studies, the exact search query that we executed, and the count of studies returned from each online database.
- Inclusion and Exclusion: the criteria according to which a paper can be included or excluded in the literature review.
- Quality Assessment Strategy: single-choice questions over predefined scored categories for reviewers about the paper's quality; similar questions for reviewers about their confidence in their review; the sources of quality-related data
- Data Extraction Strategy: the data schema extracted from the collaborative reviewing platform; this schema includes the full set of questions for the reviewers, structured around the four research questions. Many questions were single-choice over a fixed set of answers (including yes/no questions); others were questions with free textual responses.
- Protocol Validation: the full set of questions for the reviewers to evaluate and suggest improvements to this SMS, including on the research questions inclusion/exclusion criteria, search strings, and so on. This validation survey was used as an intermediate measure of quality control over the process of this SMS.

Part 2.xlsx (raw primary studies and analyzed annotations) is an archive with 560 primary studies considered in this research. Part 2 contains the following data tables:

- Primary studies: the full dataset of 153 primary studies (raw data) and their annotated information (analyzed data) from the collaborative reviewing platform. These 153 studies

are a result of a combination of 233 manually included studies and 153 studies included from screening, minus 233 excluded during the classification phase.

- The raw data has each study's unique key, title, publication type (journal, conference, workshop), citation count, publication year, authors, publication venue, source (online database), and search type (manual or automatic).
- The analyzed data has the answers to the data extraction form (see Part 1 - Protocol).
- Excluded - screening: the list of 174 studies (raw data) excluded in the screening stage and their reason for exclusion (analyzed data).
 - The raw data has the unique key, title, publication venue, publication year, and the authors of the excluded study.
 - The analyzed data has the reason for exclusion.
- Excluded - classification: the list of 233 studies (raw data) excluded in the classification stage and their reason for exclusion (analyzed data).
 - The raw data has the unique key, title, publication venue, publication year, and the authors of the excluded study.
 - The analyzed data has the reason for exclusion.

Part 3.xlsx (filtered data per RQ) is an archive with the primary studies clustered by their characteristics after our analysis. Each table is a filtering based on data from a question in the data extraction form (see the last table of Part 1). Part 3 contains the following data tables:

- Q1.1: Development phases supported by MPM studies
 - Based on Q11 in data extraction form
 - Studies filtered by whether they support the editing, analysis, integration, simulation, generation, transformation and verification activities with models
- Q1.2: CPS components modeled by MPM studies
 - Based on Q1.2 in data extraction form
 - MPM studies filtered by whether they model hardware, software, network, environment, or stakeholders
- Q1.3: Formalisms reported by MPM studies
 - Based on Q1.3 in data extraction form
 - MPM studies filtered by reported formalism to analyze the CPS
- Q1.3 x Q1.2: Formalisms mapped to CPS components
 - Based on sheets Q1.2 and Q1.3
 - Mapping of which formalism(s) is (are) used to address CPS components
- Q1.4: Integration mechanisms reported by MPM studies
 - Based on Q1.4 in data extraction form MPM studies filtered by mechanism used to integrate different CPS models
- Q2a: CPS parts modeled in all studies
 - Based on Q1 in data extraction form
 - MPM studies filtered by the part of CPS modeled by the approach

- Q2b: system qualities supported in all studies
 - Based on Q1 in data extraction form
 - MPM studies filtered by the system quality supported by CPS modeling approach
- Q2.1: models and metamodels in all studies
 - Based on Q2.1 in data extraction form
 - MPM studies filtered by whether a study reports an instance model and/or a metamodel for the approach and/or use case
- Q2.2a: novel languages/tools reported by all studies
 - Based on Q2.2 in data extraction form
 - A list of novel languages and tools reported in each MPM study
- Q2.2b: languages used by all studies
 - Based on Q2.2 in data extraction form
 - A list of languages used in each MPM study
 - A list of papers clustered by the use of multiple languages
- Q2.2c: tools used by all studies
 - Based on Q2.2 in data extraction form
 - A list of tools used in each study
 - A list of papers clustered by the use of multiple tools
- Q1.3 x Q2.2b x Q2.2.c: cross-analysis or relationships between formalisms/languages/tools
 - Based on Q1.3, Q2.2b, and Q2.2.c in data extraction form
 - A list of formalisms clustering papers as mentioning just the formalism, the formalism and the language, and/or the formalism and the tool
- Q2.3: types of process descriptions in all studies
 - Based on Q2.3 in data extraction form
 - MPM studies filtered by whether the process description is textual, diagrammatic, or model-based
- Q3: domains in single/multi-paradigm studies
 - Based on Q3 and Q2 in data extraction form
 - Study counts filtered based on whether the study is domain-specific and/or MPM
- Q3.1: domains and subdomains in all studies
 - Based on Q3.1 in data extraction form
 - Studies filtered by domain and subdomain of application
- Q3.1a: multi-domain studies
 - Based on Q3.1 in data extraction form
 - Multi-domain studies filtered by combination of domains
- Q3xQ1.3: Correlation between application domains and formalisms
 - Based on Q3 and Q1.3 in data extraction form
 - Quantitative relations between the application domains and reported formalisms
- Q4: roles of actors/stakeholders in all studies
 - Based on Q4 in data extraction form
 - Studies filtered by the type of actor/stakeholder considered in them
- Q4.1: technical expertise of modelers in all studies

- Based on Q4.1 in data extraction form
- Studies filtered by the technical background of the CPS modeller assumed or required in the approach
- Q4.2: research field of the authors of all studies
 - Based on Q4.2 in data extraction form
 - Studies filtered by the descriptions of the research field of the authors
- Q2.3 x Q4: correlation between processes and actors
 - Based on Q2.3 and Q4 in data extraction form
 - Quantitative relations between the processes and the actor roles
- Q2.1 x Q4.1
 - Based on Q2.1 and Q4.1 in data extraction form
 - Quantitative relations models/metamodels and the required modeler expertise
- Q1.3 x Q4.2
 - Based on Q1.3 and Q4.2 in data extraction form
 - Quantitative relations between formalisms and author fields
- Q1.4 x Q4.2
 - Based on Q1.4 and Q4.2 in data extraction form
 - Quantitative relations between integration mechanisms and author fields
- Q3 x Q4.2
 - Based on Q3 and Q4.2 in data extraction form
 - Quantitative relations between application domains and author fields
- QAcitations: citations retrieved for classified studies in June 2021
 - We removed inputs from QA2 from Quality Assessment form
 - Median was calculated based on total number of citations per each year
- QA1: list of venues and venue types
 - the primary studies are merged by publication venue and venue type
- QA.C: analysis of the countries from the participating authors of primary study
 - Studies are filtered by the contributing country and combination of multiple countries involved in study is analysed
- QA.D: analysis of contributing institutions based on authors list in primary study
- QA & SA: analysis of quality assessment questions
 - scores per each question were calculated
- All paper IDs: list of all paper identifiers
 - Used in the cross-question sheets
- Validation: includes the validation formulas which were used to ensure that all studies were classified and that there is no duplication of inputs

Experimental Design, Materials and Methods

To create a comprehensive overview of a given research area, Evidence-Based Software Engineering (EBSE) provides two core tools for evidence-based studies: Systematic Literature Reviews (SLRs), focusing on identifying the best practices on a given topic based on empirical evidence, and Systematic Mapping Studies (SMSs), aiming at creating a comprehensive overview of a given research area [7,10,11].

In order to achieve our goals, we chose to design SMS based on the process suggested by Kitchenham [12]. In such reviews, the conducting researchers should follow a rigorous and reproducible method to obtain, evaluate, and interpret information related to selected research questions. According to the method, the scientific papers with the keywords corresponding to Multi-Paradigm Modeling (MPM) of Cyber-Physical Systems (CPS) have been collected for the period from 2006 till 2017 from digital libraries. The papers were filtered, eliminating redundant research results and those not corresponding to the field. Then, the papers were analyzed according to several research questions, which gave reviewers the ability to discuss different aspects of MPM approaches, application for CPS, as well as the people involved in the area. The method has three phases: planning, conducting, and reporting. Outcomes were published in terms of an intermediate technical report of MPM4CPS COST action [13] and summarized in [14].

In the planning phase, we specify research questions and perform a PICOC (Population, Intervention, Comparison, Outcome, Context) analysis. Then, the review protocol (i.e. Part 1 archive) was established by defining the search strategy, selection criteria, quality assessment checklist, and data extraction strategy. We evaluated our research protocol threefold:

(1) Query Testing

To clearly define keywords, we analyzed all of our research questions separately and divided the keywords into three main groups: (i) cyber-physical systems, (ii) modeling approaches for cyber-physical systems, (iii) combination of modeling approaches for cyber-physical systems.

For each group, we created a search substring and then integrated the sub-strings into the final search string.

(i) The first group consists of terms referring to cyber-physical systems such as cyber-physical system, embedded system, real-time system, hybrid system, sensor networks system, smart system, critical system and we created the first part of our research string as a combination of the main item.

(ii) The second group included the keywords that referred to modelling and simulation terms (modelling and simulation, model-driven, modelling approach, model-approach) which we obtained the second part from our research string.

(iii) Finally, the third group referred to formalism and paradigm like multi-paradigm, multi-formalism, heterogeneous formalism, unified modelling formalism, multi-model language.

In order to find the final, most appropriate combination of these sub-strings that presents our final search query, we defined several versions of query strings and conducted a search process in all the targeted databases.

The three versions of search queries that we tested are:

Search string 1 - ("cyber physical" OR "cyber-physical" OR CPS OR cyberphysical OR smart OR critical) AND (("integrate model*" OR "composable model*") OR ("integrate simulation" OR "composable simulation")) AND ("model driven" OR model-driven OR "model based" OR model-based)".

Search string 2 - ("multi-paradigm" OR "multi-formalism" OR "heterogeneous formalism" OR "unified modelling formalism" OR "multi-model language") AND (("cyber physical" OR "cyber-physical" OR cyberphysical OR smart) AND system*) AND ("modelling approach" OR "modeling approach" OR "integrate modelling" OR "integrate modeling" OR "model driven" OR "model-driven") AND ("software engineering" OR "software system").

Search string 3 - ("multi-paradigm" OR "multi-formalism" OR "heterogeneous formalism") AND ("Modeling and Simulation" OR "Integrate modeling") AND ("cyber-physical system" OR "hybrid system" OR "embedded system" OR "real-time system" OR "smart system").

The best candidate was selected in the Validation survey.

(2) Validation Survey with MPM4CPS COST Action Members

We performed the internal validation of protocol and announced the survey MPM4CPS COST Action members during one of the meetings, in which we introduced the participants of COST network to the SMS objective. For this purpose, we created a Google sheets survey [\footnote{https://docs.google.com/spreadsheets/d/1X0B7G1RqNWvsNYILY6UPRPq34V8btbk-EVts-pVltgM}](https://docs.google.com/spreadsheets/d/1X0B7G1RqNWvsNYILY6UPRPq34V8btbk-EVts-pVltgM) that describes the key aspects. For each of these aspects we created a separate Google sheet, similar to one reported in Part 1 archive, which contained information about SMS objective, SMS requirements, SMS team and list of defined activities and users who are responsible for these activities. The participants were invited to evaluate Research Questions, Search Sources, inclusion and exclusion criteria, quality assessment strategy and data extraction strategy, respectively. Each question in the survey was linked with a particular Google sheet. A survey also collected general information about the participants and their expertise concerning SMS studies.

A total of 10 participants participated in this survey. Five participants were knowledgeable in systematic reviews, though not self-identified as experts, three of them identified themselves as experts and the remaining two were not experts. The participants chose Search string 2 to be most appropriate for executing the SMS, and gave their feedback and ranking regarding all the aspects, upon which the protocol was refined.

(3) The pilot session with reviewers

After the survey was concluded, we refined the protocol in accordance with the obtained feedback. However, we found it was necessary to provide additional descriptions regarding answering each question related to data extraction. In order to provide comprehensive descriptions, we organized a special work session dedicated to this task during MPM4CPS cost action meeting. All meeting participants helped in providing the appropriate descriptions, regarding each question. As a follow up to the workshop meeting, we organized the pilot session with three reviewers. Each reviewer was assigned to read and classify two primary studies. The complete protocol, with a special answer sheet, was provided to reviewers in the form of a Google Sheet document. After completing their assignments

independently, reviewers had a live session with the SMS project lead. During this session, each of the questions was re-evaluated, and reviewers expressed their doubts and eventual problems which they encountered while providing their answers. The completeness of the answers given in text boxes was also evaluated and discussed.

Several necessary clarifications were identified as a result of this session. For instance, it was not trivial for the reviewers to understand that it was necessary to provide the description for the different types of instance models and meta-models which they found in primary studies, as some were provided for a case study which was used to illustrate the approach, while others were provided for approach itself.

Further, it was not clear to reviewers that text box descriptions were mandatory to be filled in the case that the Boolean answer to the question is marked with 'Yes'. Finally, one of the reviewers found it confusing which information should be extracted from what source when they were collecting data on the technical background of the authors. Once all the concerns were discussed, we updated the descriptions for each affected question in Data Form.

To conduct the review, we used the ReLiS platform [15], primarily because it supports multi-user data extraction. It features a domain-specific modeling editor tailored for researchers who perform systematic reviews and an architecture that enables live installation and deployment of multiple concurrently running projects. ReLiS is a framework to automatically install systematic review projects on the cloud. Its generic and dynamic architecture allows users to install projects during the process without manual intervention [16]. In addition, ReLiS allows researchers to conduct reviews collaboratively.

Conduction phase followed six phases, reported in Description of Data Collection, which were synthesized into two main steps for reviewers which the ReLiS platform distinguishes:

- (1) Screening - where the studies were examined to fit inclusion criteria based on abstract, title and keywords
- (2) Classification - where the studies were examined in full, and data was extracted in accordance to Data Form (see Part 2 archive)

Ethics Statement

The collection of this data did not include any experimentation with human subjects or animals.

All the raw, analyzed, and filtered data presented in this dataset are faithfully reported. This includes also all the corresponding sources. This dataset has not been published elsewhere, and all the authors approve its submission. We made all the necessary efforts to collect, analyze, filter, and report the state-of-the-art literature contained in the dataset in a faithful manner, by following an adequate research protocol. Nonetheless, should these authors or a third party discover errors or inaccuracies in the dataset, we are ready to promptly notify and cooperate to adopt the necessary countermeasures.

Acknowledgments

This work was supported by MPM4CPS IC1404 COST action and partially supported by NOVA LINCS with the support of FCT Portuguese national funds (UID/CEC/04516/2019) and the Knowledge Foundation in Sweden through the MINEStrA and SACSys projects.

We would like to thank Prof. Paulo Carreira, Prof. Geylani Kardas, Prof. Ivan Lukovic, and Prof. Mauro Iacono for their feedback during the construction of the protocol. We thank Prof. Vasco Amaral, Prof. Miguel Goulao, Prof. Bedir Tekinerdogan, Dr. Claudio Gomes, Dr. Ferhat Erata, Prof. Dušan Savić, Mustafa Abshir Mohamed, Dr. Hana Mkaouar, Dr. Raheleh Eslampanah, Prof. Moharram Challenger, Dr. Dominique Blouin, Prof. Oksana Nikiforova for their contribution to the classification of the primary studies. Finally, we thank Prof. Eugene Syriani and Brice Bigendako for their constant support with the ReLis platform.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

References

- [1] A. Barišić, J. Cunha, Sustainability in Modelling of Cyber-Physical Systems: A Systematic Literature Review - Intermediate Technical Report (4 2017).doi:10.5281/ZENODO.1481946.
- [2] Primary Studies for Systematic Mapping Study on MPM4CPS. <https://www.mendeley.com/community/smr-mpm4cps-public/>, (accessed 14 June 2021).
- [3] Google Scholar. <https://scholar.google.com/>, (accessed 14 June 2021).
- [4] ACM Digital Library(ACM). <https://dl.acm.org/>, (accessed 14 June 2021).
- [5] IEEEExplore (IEEE). <https://ieeexplore.ieee.org/Xplore/home.jsp>, (accessed 14 June 2021).
- [6] Science Direct (SD). <https://www.sciencedirect.com/>, (accessed 14 June 2021).
- [7] Springer Link (SL). <https://link.springer.com/>, (accessed 14 June 2021).
- [8] Scopus. <https://www.scopus.com/home.uri>, (accessed 14 June 2021).
- [9] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering—a systematic literature review, *Information and software technology* 51 (1) (2009) 7–15.

- [10] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. "Lessons from applying the systematic literature review process within the software engineering domain." In: *Journal of systems and software* 80.4 (2007), pp. 571–583.
- [11] K. Petersen, S. Vakkalanka, and L. Kuzniarz. "Guidelines for conducting systematic mapping studies in software engineering: An update." In: *Information and Software Technology* 64 (2015), pp. 1–18.
- [12] B. Kitchenham, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report (7 2007).doi:10.1016/j.scico.2013.03.017.
- [13] Ankica Barišić, Dušan Savić, Rima Al-Ali, Ivan Ruchkin, Dominique Blouin, Antonio Cicchetti, Raheleh Eslampanah, Oksana Nikiforova, Mustafa Abshir, Moharram Challenger, Claudio Gomes, Ferhat Erata, Bedir Tekinerdogan, Vasco Amaral, Miguel Goulão: *Systematic Literature Review on Multi-Paradigm Modelling for Cyber-Physical Systems*, In: *COST IC1404, Deliverable WG4*, DOI: 10.5281/zenodo.2528953, 2018.
- [14] A. Barišić, I. Ruchkin, D. Savic, M. Abshir Mohamed, R. Al-Ali, L. W. Li, H. Mkaouar, R. Eslampanah, M. Challenger, D. Blouin, O. Nikiforova, and A. Cicchetti: *Multi-Paradigm Modeling for Cyber-Physical Systems: A Systematic Mapping Review*, In submission to *Special Issue on Advanced Techniques and Emerging Trends in Smart Cyber-Physical Systems, Future Generation Computer Systems*.
- [15] ReLiS : a tool for conducting Systematic Review. <http://relis.iro.umontreal.ca/auth.html>, (accessed 14 June 2020).
- [16] Brice M. Bigendako and Eugene Syriani. "Automatically Installing and Deploying Tools for Conducting Systematic Reviews in ReLiS", *CoRR* (2017), vol. abs/1705.01978