Revista de Sistemas de Informação da
FSMA n. 16 (2015) pp. 3 - 15

TRILHA ESTUDANTIL

SISTEMAS DE INFORMAÇÃO
http://www.fsma.edu.br/si/sistemas.html

A Systematic Mapping on Supporting Approaches for Requirements Traceability in the Context of Software Projects

Paulo Robson Campelo Malcher, *Graduate student in Computer Science, PPGCC-UFPA*,
Diogo Adriel Lima Ferreira, *Bachelor in Information Systems, FACOMP- UFPA*,
Sandro Ronaldo Bezerra Oliveira, *Dr. Adjunct Professor IV, PPGCC-UFPA*, e
Alexandre Marcos Lins de Vasconcelos, *PhD. Associate Professor*, CIn-UFPE
{prcmalcher, cetlho}@gmail.com, srbo@ufpa.br, amlv@cin.ufpe.br

***Abstract*— The Requirements Traceability is seen as a quality factor with regard to software development, being present in standards and quality models. In this context, several techniques, models, frameworks and tools have been used to support it. Thus, the purpose of this paper is to present a systematic mapping carried out in order to find in the literature approaches to support the requirements traceability in the context of software projects and make the categorization of the data found in order to demonstrate, by means of a reliable, accurate and auditable method, how this area has developed and what are the main approaches are used to implement it.**

***Index*— Systematic Mapping, Requirements Traceability, Requirements Management.**

I. Introduction

The search for quality hás become a strategic goal for many organizations in order to fulfill all the market demands [1]. Nesse This way, there are many proposed solutions to help improve software quality such as models, patterns and methods for software development. In this context, in order to help the software engineer to better understand his target problem came the requirements engineering.

Requirement engineering includes a set of activities that foster the understanding of the business rules that will be impacted by the software, of the expected return from the software and the final interaction between software and user [2]. The process of the requirement engineering must typically contemplate the following activities [3]:

- Requirement gathering, which corresponds to the initial phase of the requirement engineering process and includes the activities that lead to the discovery of the requirements;
- Requirement analysis, when the gathered requirements are used as a foundation for the system modelling;
- Requirement documentation, when the requirements and models elicited in the previous steps must be described and presented in documents. This is, therefore, a Record and officialization activity for the results of the requirement engineering;
- Requirement verification and validation, in which it is fundamental for the requirements to be carefully evaluated. Hence, the documents created during the previous step must be submitted to verification (assutring that the software is built correctly) and validation (assuting the the correct software is being built);, and
- Requirement management, whose goal is to control the creation and the evoluation of requirements [4].

This paper studies the área of requirements management, because it is the initial area for reaching the maturity in the requirement engineering process implementation in quality models such as CMMI-DEV – *Capability Maturity Model Integration for Development* and MR-MPS-SW – Reference Model for Software MPS. Hence, according to Kotonya and Sommerville, requirement management includes the activities that support control and tracking of requirements, as well as requirement change management at any point of the software life cycle, where a requisite is a condition or a capacity that the system must fulfill [3]. Still in the context of software development quality, there are approaches that help to achieve it, such as the Norm ISO/IEC 12207 [5], and the CMMI [6] and MPS.BR [7] models, where the process of requirements management is deeemed essential to the success of a project, because it must assure that the set of agreed requirements is managed andsupports the project’s planning and execution needs.

This process, besides its goal of controlling the cration and evolution of requirements [8], also includes the following activities: managing the creation of requirements; managing the relationships among requirements; coordinate the requirement validation with the client; and manage the changes in requirements [3][7].

Due to the changes that might happen throughout the life of the systems, the requirements management must contemplate their traceability. Besides, for the coordination and planning of the requisites and their possible changes, there must be a control of the mutual dependencies between the requirements and the artifacts created during the software development process, which brings us directly to Requirements Traceability [4].

Requirement traceability is defined as the ability to describe and follow the life cycle of a requirement in both directions: in the ability to trace an artifact towards its implementation (forward traceability) or in the ability to trace an artifact to its origin (backward traceability), passing throuhgh all the reported specifications [9].

Traceability can also be categorized into horizontal (inter-traceability) and vertical (extra-traceability) [3]. Horizontal traceability is the traceability between different versions or variations of requirements or other artifacts in a specific phase of the life cycle, that is, it allows us to see how the requirements depend on other requirements. On the other hand, vertical traceability is defined between requirements and artifacts created during the development process through the project life cycle, that is, it allows us to see how the requirements relate to the artifacts that are created during the project [10].

In this context, we can see a great concern in having a view of the relationships involving requirements. In fact, the success of the requirements management and, consequently, of the software projects depend directly from how well defined and known are those relationships [11].

A research that focuses on finding out which approaches described in the literature help trace requirements may be worth a lot for future results, because it can bring as results: the discovery of which techniques support its development; which support tools are used to implement them; which support tools are used in its implantation; which models and frameworks were already developed; and lots of information about this research topic. This way, this paper intends to present a view on support approaches for requirement traceability and by approaches we mean processes, models methods, techniques, tools, frameworks and other related items.

Another important point is the method chosen to perform it. In order to meet the goals of this research, we chose the one of the main methods of evidence based software engineering, which is the Systematic Literature Review (SLR). This is a technique classified as a secondary study, given that it depends on the primary studies used to find evidence and build knowledge [12].

It is important to point out that as a research área matures, the number of studies and results grows in a significant way [13]. Hence, the literature separates SLR into two types: convencional systematic literature reviews (SLR) and systematic mapping studies (SMS) [14]. The latter, also known as exploratory study, includes a wide review of primary studies in a specific area, seeking to identify which evidences are available in that area [15]. Performing a mapping provides a general view of a research area and also allows us to know the publication frequency variation with time, the amounts and type of research within it, allowing us to identify trends [13].

SLR and SMS became two important tools to aggregate and build knowledge in software engineering, presenting the following adavantages [15]: the well defined method makes it less probable that the results found in literature are biased, even though it does not protect against bias in the primary studies; they can provide information on the effects of a specific phenomenon through a wide array of configurations and empirical methods. If the studies offer consistent results, the systematic review offers evidence that the phenomeno is robust and repeatable. On the other hand, if they are inconsistent, the variation sources can be studied and determined. In the case of quantitative studies, it is possible to combine data using meta-analysis techniques. This increases the probability of detecting real effects that small individual studies are not able to find.

Given that, this work represents a systematic literature review whose goal is to identify the research related to support approaches for requirements traceability and answer questions on the research on this field. The results found with this method will be used to create a catalog of traceability support techniques in order to help achieve success in software projects.

Besides this introductory section, this work is organized as follows. Section 2 describes the systematic mapping performed. Section 3 presents the results and their analysis. Section 4 discusses the threats for the validation of this research. Section 5 presents work related to this research and finally, Section 6 presents our final considerations.

II. Research Method

This research performed the mapping with the goal of finding and analyzing the largest possible number os relevant and recongnized primary research on requirement traceability in order to answer the research questions.

The systematic reviews and systematic mapping studies follow a specific method that begin with the definition of the review protocol and divides a systematic review into three steps: planning, conduction and presentation [15].

The planning phase is the first step of a systematic mapping and focuses on the need for a detailed protocol that describes the process and the methods that will be used. The protocol is the instrument that consolidates all definitions in this phase. Identifying the research questions translates as the most important point in the planning phase of the mapping, because the entire research will be limited b y the scope of the question that needs to be answered [12].

The protocol defined must address some points: the review goal, the research question, the scope and restrictions applied to the research; the sources that were mined for the review in search of primary studies; the identification of key-words; the generation of search strings; the inclusion and exclusion criteria and how they will be applied for the selection of the studies; the evaluation of the primary studies; the data extraction process and how they will be summarized. Besides, the review protocol must be evaluated to assure that the planning is viable [15]. For that, many researchers suggest that specidalist be consulted in order to review the protocol and/or test the protocol execution.

The execution phase of the systematic mapping comprises the primary studies selection, data extraction and evaluation phases. The last step of the systematic mapping is the writing of the mapping report according to the findings of the data analysis and synthesis.

Hence, this section presents the main sections of our protocol, followed by the systematic mapping conducted in this study. In order to guide the construction of this research protocol, we used the instructions and guidelines presented in [15], where the authors specifiy that systematic reviews and systematic mapping studies have the following issues in common:

1. Identification of the need to perform a systematic review;
2. Definition of the research question;
3. An encompassing and exhaustive search for primary studies;
4. An evaluation of the quality of the included studies;
5. An identification of the data that is necessary to answer the research question;
6. Data extraction;
7. Summary and synthesis of the results in the studies (meta-analysis);
8. Interpretation of the results in order to determine their applicability
9. Report writing.

The systematic mapping protocol can be found in its entirety at the web address given by *http://spider.ufpa.br/projetos/ spider_rastreabilidade/SPIDER_ProtocoloRevisao.pdf*.

A. Goals of this mapping

The goal of this systematic mapping study was to identify approaches that support the activity of Requirement Traceability, including processes, models, methods, techniques, tools and others. Hence, we defined the following structure for the goal, as proposed in [18]:

- **Analize:** experience reports and scientific publications throuh a study based on a systematic mapping;
- **With the goal of:** identifying apporaches that support Requirements Traceability activities, whose importance is to find the origin of each element that belongs to the development process, as well as the requirements that are related to them and how they will be affected by any change, offering a possible impact analysis of those changes.
- **In relation to:** the definition and usage of processes, frameworks, tools and other instruments used for the implantation and execution of Requirement Traceability activities in software development organizations;
- **From the point of view of:** researchers and software development and support organizations;
- **In the context:** both academic and industrial.

B. Research Question

The research questions in a systematic mapping study are much more general (and usually have an exploratory characteristic), in contrast with those that are defined in systematic literature reviews, which need to be more precise. Hence, the research question that guided this mapping was the following:

(Q1) Which approaches exist to support the activities of Requirement Traceability?

Additionally, the following secondary research questions were used to guide the research and define the profile of the existing publications in the specialized literature:

1. What is the distribution of studies per publication year?
2. What is the distribution of studies per author?
3. What is the distribution of studies per institution?

4. What is the distribution of studies per publication type?

The structure of the main question was defined according to the structure *Population, Intervention, Context, Outcomes, Comparison* (PICOC), recommended in [15]. Nevertheless, only the items *Population*, *Intervention* e *Outcomes* (PIO), were considered relevant to the research. Hence, we defined the followign structure to the main research question:

- **Population (P):** Software organizations and software projects;
- **Intervention (I):** Patterns to support the activities of Requirements Traceability;
- **Outcomes (O):** Process models, techniques, methods, tools and frameworks for Requirements Traceability.

C. *Research Scope and Restrictions*

We defined a research scope in order to assure its viability. It can be defined through the definition of the source selection criteria and some restrictions.

For the selection of research sources, we defined the following criteria:

- Availability for web search;
- Availability for article search from the UFPA domain;
- Availability of full text articles from the UFPA domain or from the Google and/or Google Scholar search engine;
- Availability of the full text either in English or in Portuguese;
- Source relevance.

We have the following restrictions to the research:

- The research cannot have a finantial impact due to the fact that there are no funds allocated to this Project. Hence, we only selected sources that could be accessed for free (we also considered sources that could be accessed for free from with the uFPA domain);
- We only considered studies obtained from the selected sources that are conformant to the inclusion and exclusion criteria;
- The research was restricted to the papers published between January, 1st, 2003 until December, 2013, comprising a period of 11 years, due to the need to identify the most current approaches to support the activities for Requirements Traceability.

D. *Source Selection*

Based on the selection criteria and the research restrictions, we selected the following databases as sources for the research. In them we performed the searches for primary studies according the search availability and the relevance for the área of Software Quality:

- IEEEXplore Digital Library;
- El Compendex;
- ISI Web of Knowledge;
- ACM;
- Scopus;
- Annals of WAMPS – Annual Workshop of MPS;
- Annals of WER – Requirement Engineering Workshop;
- Annals of SBQS – Brazilian Symposium on Software Quality.

E. *Identification of Keywords and Synonims*

Keywords were identified based on the research questions and according to the structure Population, Intervention and Results. According to the research restrictions, they were defined in English and Portuguese. Here, we present the keywords in English for the main research question, according to the analysis performed on the topic of Requirement Traceability as defined in the software quality models (CMMI-DEV and MR-MPS-SW).

- **POPULATION:** *Software, Project, Development, Organization, Enterprise, Company, Industry, Institute, Research Group, Technology Center*;
- **INTERVENTION:** *Requirements Traceability, Traceability, Requirements Tracing*;
- **RESULTS:** *Model, Process, Framework, Method, Technique, Methodology, Knowledge, Activity, Task, Tool, Software, Program, System, Application, Environment, Workbench*;

F. *Generation of Search Strings*

The search string groups the keywords using the OR and AND operators. The OR operator is used to group keywords and their synonyms, per element (Population, Intervention and Results), while the AND operator is used to group the set of keywords defined for all elements, according to the PICO structure (or PIO, in this case), as follows [16]:

P <and> I <and> C <and> O

It is worth to point out that the element Comparison (or Control), the C in the expression above, is not in the context of this work. Hence the set of keywords for this element is empty.

- ("Software" AND ("Project" OR "Development" OR "Organization" OR "Enterprise" OR "Company" OR "Industry" OR "Institute" OR "Research Group" OR "Technology Center")) AND ("Requirements Traceability" OR “Traceability” OR "Requirements Tracing") AND ("Model" OR "Process" OR "Framework" OR

"Method" OR "Technique" OR "Methodology" OR "Knowledge" OR "Activity" OR "Task" OR "Tool" OR "Program" OR "System" OR "Application" OR "Environment" OR "Workbench")

Another important thing to point out is that we need to make some changes in the search string according to the syntax of the search string for each research source.

G. Selection of Primary Studies

The selection of primary studies can be divided into the following steps: definition of inclusion and exclusion criteria for primary studies and definition of their selection process.

The inclusion and exclusion criteria for primary studies will guide the researchers in the selection of the studies that were gathered from the research sources and also define the research rigour and prevent the researchers from having any bias during the selection. The exclusion criteria in this work were the following:

- Papers that are not freely available for download (full text) in the searched databases or through a manual search in Google or Google Scholar (for papers that are not offered with full text)*;*
- Papers that clearly do not answer the research questions;
- Repeated papers (in more than one searched database) were considered only once;
- Duplicated papers were restricted to the newer or more complete version, except in the case where there were complementary information;
- Studies that were classified as abstracts, keynote speeches, courses, tutorials, workshops and similar items;
- Papers that do not mention the keywords of the search string in their tile, abstract or keywords, except in the case of papers that discuss improvements to the software process in which we can observe the possibility of dealing with Requirements Traceability;
- Exclusion in the case of non insertion of the study in the context of Software Projects, Software Industry or Software Engineering;
- Exclusion if the paper is not presented in onde of the accepted languagens (English or Portuguese).

The inclusion criteria for the papers were based on the following principles:

- Studies that present either primary or secondary approaches (patterns and CASEs) of support to the activities of Requirement Traceability;
- Studies that present reports of indsuty experience, or either experimental or theoretical researches, as long as they present application examples, experiment descriptions or real use case of their approaches (patterns or CASEs) to support the activites of Requirement Traceability.

We used the following resources to perform the primary selection phase of this study:

- Two researchers (one masters and one undergratuate student);
- Access to the research sources through the Federal University of Pará domain;
- Validations on documents and procedures of the systematic review made during meetings with the Project SPIDER coordinator and Project advisor.

During the conduction of the systematic mapping, the primary studies were identified according to the following process:

- At first, we performed some pilot tests on the digital databases with the goal of adapting the search string for each repository. Ths procedure allowed for the execution of an automatic search to be performed the same way in all search engines, with minor adaptations for each tool. This strategy allows us to improve the future replication of this study;
- After that, we performed the searches in all selected sources using the search strings. Studies that were clearly irrelevant to the research were discarded. Articles were cataloged using the Mendeley Desktop tool, establishing a spreadsheet with the list with possible primary studies for each researcher;
- Based on the reading of the title and the abstract, the papers were evaluated according to the exclusion and inclusion criteria anf the result was recorded;
- Both researchers responsible for the article selection came to a consensus, when necessary. This happened when they did nto initially agree on the inclusion of an article (when one decided for the inclusion and the other for the exclusion);
- In the consensus phase, if there was a disagreement on the inclusion of an article, it was included in the research;
- Afterwards, the agreement índex was measured through the calculation of the Kappa value used in [33], for future reference;
- The primary studies identified were afterwards fully read and then a quality evaluation and a data extration strategy were applied, as explained in the next subsections.

H. Evaluation of the Primary Studies

The quality of an article can be measured by the relevance and the scientific value of its content. In the quality evaluation of the primary study we considered also an exclusion criterion to be applied during the research. Even though there is no universal of research quality, most checklists include issues that have the goal to evaluate the extent to which the bias is minimized and internal and external validations are maximized [15].

Therefore, during the analysis of primary studies and result gathering, we applied the quality criteria, allowing for an additional step of study validation, in order to identify possible papers that still needed to be excluded from our research and to observe individually the degree of importance of the studies for any comparisons we might deem necessary at the data synthesis phase [15].

Additionally, the quality evaluation might serve as study recommendation for future research, offering information on the information quality for each evaluated study [15].

Studies that might have been excluded for not fitting in the aforementioned quality criteria were set aside with the description of the reasons for their exclusion. After this step, the papers included in the research passed through the data extraction phase.

The quality criteria we applied to the primary studies were adapted from [12], given that it described criteria that were extensive enough to include the scope of studies under consideration, with some changes to fit the goals and research questions of this systematic review. The quality criteria were the following:

- Introduction/Planning
  - The goals or research questions of the study are clearly defined (including the reasons to perform the sudy)?
  - The type of study is clearly defined?
- Development
  - Is there a clear description of the context in which the research was perfomed?
  - The paper is either well or adequately refenced (does it present related or similar works and is based in models and theories well spread in the literature)?
- Conclusion
  - Does the study report clearly and non ambiguosly the results?
  - Are the goals or research questions of the study reached?
- Criteria for the investigation question
  - Does the study list primarily or secondarily hte models, processes, methods, techniques, methodologies, etc. used to support the activities of Requirement Traceability?
  - Does the study present support tools to the activities of Requirement Traceability?
- Specific criterion for experimental studies
  - Is there a method or set of methods described for the study execution?
- Specific criterion for theoretical studies
  - Is there a non biased process for study selection?
- Specific criterion for systematic reviews
  - Is there a rigorous protocol that was described nad followed?
- Specific criterion for reporto f industry experience
  - Is there a description of the organization(s) where the study was conducted?

The execution of the process of the primary studies evaluation step occurred according to the following process:

- The primary studies selected were fully read and evaluated according to the quality criteria. In order to evaluate their adequacy to the quality criteria, we adopted the strategy proposed in [12] that used the Likert-5 scale, allowing for graded answers from 0 (totally disagree) to 4 (totally agree), as presented in Table I. In order to help the evaluation following the Likert-5 scale, we defined scales specific for each quality criterion.

Table I. Likert-5 scale

Scale	Description
Totally agree (4)	Must be granted in the case where the work presents in the text all the criteria and fully complies with the topic.
Partially agree (3)	Must be granted when the paper complies partially with the criteria of the issue.
Neutral (2)	Must be granted when it is not clear whether the paper complies or not with the criteria.
Partially disagree (1)	Must be granted when mosto f the criteria are not met by the evaluated paper.
Totally disagree (0)	Must be granted when there is no part of the paper that meets the criteria of the issue at hand.

- Then we must use the sum of the points and fit the paper in one of the quality levels presented in Table II.

Table II. Quality Levels

Grade level	Evaluation
Excellent	>86%
Very Good	66%-85%
Good	46%-65%
Average	26%-45%
Bad	< 26%

III. Results and Analysis

This section described our main contribution, the application of the systematic literature mapping techniques. Its application during the conduction phase is presented according to the protocol described in the previous section. In this work, the conduction phase was divided into: primary search, primary studies selection and data extraction.

A. Primary Search

According to the definitions of the review protocol on the issue of search method used on the selected sources, we performed a search in each primary source using a specific search string. From this, we found a total of 1509 studies, 433 of which in IEEE, 437 at EI Compedex, 124 at Scopus, 316 at ACM, 175 at ISI Web Knowledge; 6 at SBQS; 9 at WAMPS; and, finally, 9 at the WER. The distribution found is shown at Figure 1.

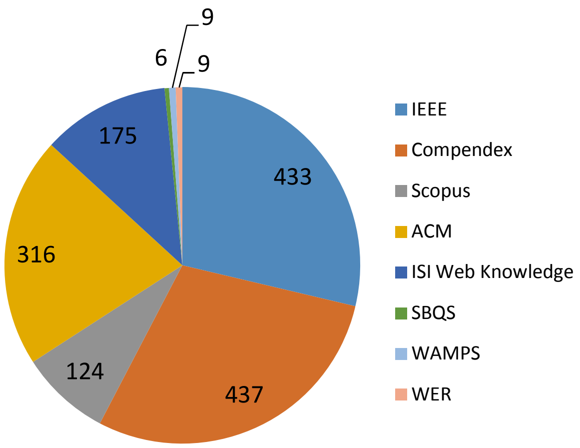

Fig.1. Distribution of the studies found at each research source.

B. Selection of Primary Studies

Based on the selection process defined in the review protocol, were selected the primary studies after reading the title, keywords, abastract and conclusion for each study found and afterwards applied the inclusion and exclusion criteria.

After the selection process the number of studies decreased a lot, falling to 411. The evolution of the numbers in the primary studies selection process is presented in Table III. It is important to point out that the full list of selected papers for each inclusion and exclusion criterion is presented in *http://cin.ufpe.br/~srbo/UFPA_TCC_DiogoAdriel.pdf.*

Table III. Overview of the Primary Studies Selection

Primary Studies Selection				
Sources	**Returned papers**	**Excluded papers**		**Included papers**
		Duplicated	**Others**	
IEEE	433	0	269	164
Scopus	124	95	13	16
ACM	316	61	182	73
ISI Web Knowledge	175	110	41	24
EI Compedex	437	228	99	110
SBQS	6	0	0	6
WAMPS	9	0	0	9
WER	9	0	0	9
Partial total		494	604	
Total	**1509**	**1098**		**411**

We can see that 1098 of the studies found were excluded by the researchers. The main reasons for the exclusions were: does no present primarily or secondary Requierments Traceability support approaches, were repeated, that is, were available in more than one of the sources used, did not mention the search keywords at the title, abstract or paper keywords, were not freely available for download.

After the primary studies selection process, we can verify that even though ISI Web of Knowledge was responsible for a considerable number of the studies initially found (close to 12% of the initial results), this source had a small representation in the final seelction when compared to the others (with the exception of Scoupus), given that this source was responsible for less than 6% of the studies. This can be explained by the order in which the selection process was performed, given that many of the studies of those sources were included in the repeated studies exclusion criterion (they were also present in other sources). Another aspect on the study duplicity can be seen when we observe that the EI Compendex was the source with the largest number of studies found, but more than 228 of those were excluded because they were already present in another source database.

C. Data Extraction

When analyzing the number of studies published each year, we can demonstrate and confirm what other authors state on the growth of the number of researches on Requirement Traceablity in the context of software projects in the last decade.

According to the results found in the search of the sources of selected research, in the year of 2003 only 6 studies on the research topic were published, but from this year on there was a large growth. Even though from 2011 on there was a decrease in this amount, the value found in 2013 is still superior to the one found in 2003, a fact that might have been motivated by the increasing recognition of the importance of this area. The graphic in Figure 2 illustrates the distribution through the years of primary studies identified by the selection process.

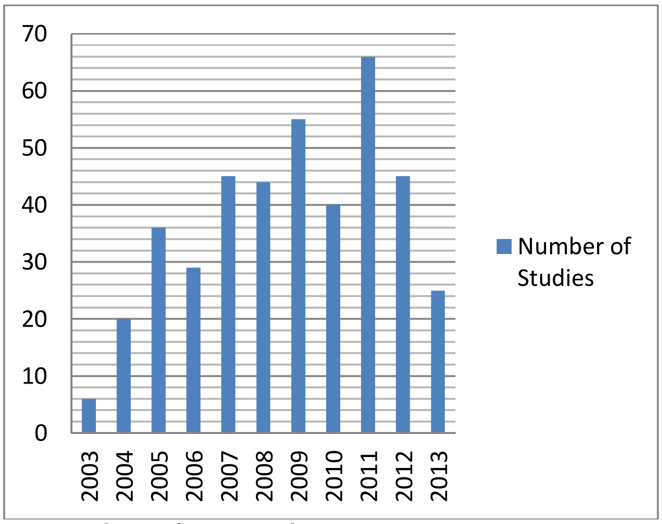

Fig.2. Distribution of Primary Studies per Year.

In the graphic at Figure 3 we can observe the number of publications per authors that were included in this research based on their research within the specified time frame. It is valid to inform that other authors were omitted because of the small number of papers published when compared to those that appear in the graphic. Due to presentation issues, we included only the tem most prolific authors.

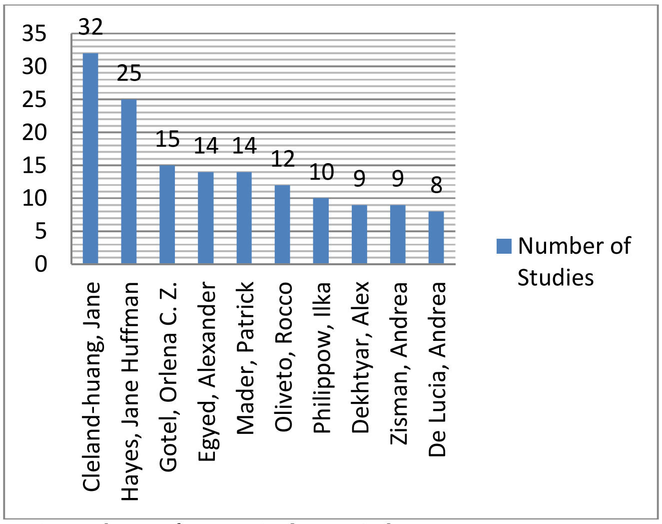

Fig.3. Distribution of Primary Studies per Author

The distribution of papers per affiliation institution can be seen in the graphic at Figure 4, where the number of institutions with publications is higher than visualized in the graphic (some where omitted). The graphic shows only the main institutions that publish papers on the issue of Requirement Traceability.

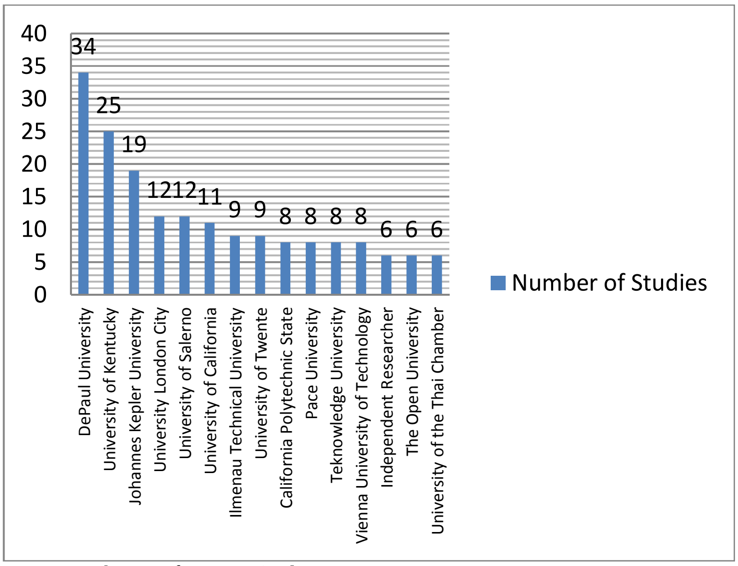

Fig.4. Distribution of Primary Studies per Institution

The distribution of papers per publication type is depicted in the graphic of Figure 5. It shows that most primary studies (79% or 328 studies) were published in events and another large part (20%, or 83 studies) were published in journals and finally, less than 2% of the results (2 studies) are technical reports. We can see that most studies included were published in events, which may be explained by the fact that computer science is a relatively new science. Hence, there are not that many journals specialized in Software Engineering. Also, some researchers tend to prioritize the publication at events due to the speed of publication at this venue. The list of events that published the selected studies is available at the document stored at the address *http://cin.ufpe.br/~srbo/UFPA_TCC_DiogoAdriel.pdf.*

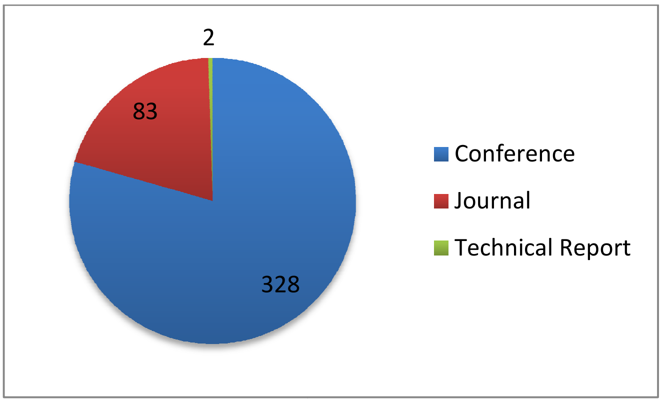

Fig.5. Distribution of Primary Studies per Publication Type

We can divide the support approaches to traceability found in the literature in the following classes: techniques, models, frameworks and tools. In this paper we define those approaches as:

- Techniques are the procedure or set of procedures whose goal in to find a specific result, be it in the field of science, technology, arts of any other activity;

- Model is the representation or simplified interpretation of reality using artifacts;
- Framework or conceptual structure is the set of concepts used to solve a problem from a specific domain;
- Tools is a utensil, device or mechanism used to perform a task.

The graphic in Figure 6 shows the division of these approaches by the number of articles that show them.

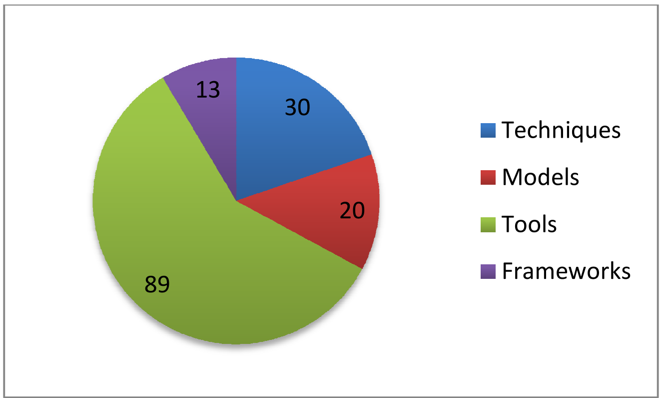

Fig.6. Distribution of Approaches

On the topic of techniques, we identified close to 30 techniques for traceability support in the literature. It is possible to divide them into traceability visualization and traceability link generation techniques. The visualization techniques are those that allow us to see the results of the traceability, such as traceability matrices, which was one of the techniques most referenced in the literature found [19][20][21][22] and several other papers, and also the graph representation discussed in papers such as [23][24]. |Most of the techniques for generation and maintanace of traceability links belong to the Information Retrieval area. Examples of these techniques are the Vector Space Model (VSM) [25][26], Latent Semantic Information (LIS) [27][28]. Some benefits derived from the use of those in the scenario of an organizational improvement program are the following:

- As to the neeed to define a mechanism that allows to track the dependency between requirements and work products, we can better use any link generation techniques that use information recovery and as to the traceability visualization, the traceability matrix is still the most satisfactory tool for a wider view of traceability
- In the context of consistence maintance between requirements and work products, based on the evaluation and the fact that the identified products were corrected, we can verify that based on the reviews on traceability performed and on their links, we can use link visualization to identify inconsistencies and offer corrections. A technique that offers a better visualization of this kind is the one that involves graph implementation;
- Finally, in order for us to have the possibility of adding additional requirements to the project, removing or changing existing requirements, the updates must be made in the established traceability structure and what will differ is the way to generate the traceability links: if it is manual, the change must also be made manually and in the case of using either semi-automatic or automatics changes, the impact usually is identified automatically.

From the models found in literature, we can highlight mainly the Traceability Information Model (TIM), that is a basic model on traceability much referenced by the literature because in offers a conceptual view of na implementation of a traceability strategy and was mentioned in [29][30]. As to the tools, several proposals were found and we can mention some of the most relevant, such as RETRO [31] and DOORS [32].

Some other approaches found in the systematic mapping can be seen and detailed in [33].

IV. Threats to Validity

In this section we present improvements for future instances of studies similar to this one. The first point we mention is the number of researchers involved. As established in the mapping protocol, only two researchers performed the process of search and selection of primary studies. This number may be increased in order to eliminate biases during the research. Nevertheless, the results found were validated with the researcher-advisor,

It is also important to point out as an improvent, the care that should be taken to build the search string to perform the primary searches in the automatic sources, because initially we created a single search string. After its usage, we verified the need to refine it and change its structure to better adequate it to the search engine of each source.

As to the amount of studies, we can highlight that the studies found were enough for the research, that is, presented enough subsidies for the composition of a catalog of approaches.In spite of that, in order to improve the generalization of the research, other databases could have been added to the research. Nevertheless, to comply with the source selection criteria (namely, to be a source that did not require and expenses from the researchers), some of the sources that could futher help improve our results were left out of the research.

V. Related Work

We performed a search at Google Scholar and several other databases relevant for our topic, such as IEEE and ACM with the goal of finding important information on the topic of interest of this paper. We found two works that are referenced as follows:

- "*Requirements traceability state-of-the-art: A systematic review and industry case study* [16];
- "*A Review of Traceability Research at the Requirements Engineering Conference[RE@21]*" [17].

The first paper is a systematic review and case study in the industry on the state of the art of Requirement Traceability between 1997 and 2007. The research questions for the review can be summarized as the search for the definition of Requirement Traceability, its main challenges, tools and techniques that might help in its application. The research was performed in five sources (IEEE, ACM, Springer Link, Inspec and Compendex). Based on the inclusion and exclusion criteria, the authors selected 52 primary studies that were used to answer the research questions of the review.

In this review we could observe that even though it was published in 2012, the last year selected was 2007, that is, eight years ago, and because of that, newer approaches may be available in the literature.

Another important point to highlight concering this paper is that it does not present an evaluation of the quality of the primary studies selected, as used in [14] – it just presents in the review protocol the quality criteria defined, but does not report in its application. Hence, the work performed in this research differentiated from that previous one in the goal and research question, given that we do not only wish to find the characterization of requirement traceabilitym the main challenges for its implementation, the main techniques and tools available in the literature, but also in a wider way, any approached referring to requirement traceability, such as processes, frameworks, models and meta-models, among others, as well as finding which are the main application contexts for the approaches found. This difference is even clearer when we consider the difference between the amount of papers selected for each research.

The second study is a review on traceability in the context of the "*Requirements Engineering ConferenceRE@21*" performed in a period of 20 yeards. Nevertheless, the work does not present clearly the start and end dates of the period under study. Given the distribution of the results, we can assume that the period of time for article selection was between 1993 and 2013. The main goal of that study was to verify how the traceability research in that conference contributed to the area. In this way, we can summarize the research issues for that review as the discovery of topics on traceability, challenges, tools, artifacts, methods and, finally, the main authors and institutions that work on this are found in the conference literature.

It is important to point out that this last study does nor define its review as a systematic literature review, but it presents some definitions that are used in these studies. Based on the search strategy defined in the paper, 76 papers were foun, but since six were excluded, they took into consideration 70 primary studies.

The authors did not present objective criteria for the inclusion or exclusion of the studies, what gives the review a more subjective character, something that is not well perceived in a systematic review, which should strive for the elimination of biases from their authors. Another point to highlight is that it was not performed an evaluation of the quality of the secondary studies. Besides, the review concentrates in a single conference, which can be seen as a limitation of this work.

Besides, our work has the additional differential against the studies [16] and [17] the higher number of papers selected (411) based on our inclusion, exclusion and quality criteria. We also observed that most of the studies analyzed in those papers were also selected in ours (due to our selection period – 2003 to 2013, some of their papers were not able to be selected).

As a smilar point, we can highlight the dfistribution of papers selected by type of publication, countries and researchers. Our work complements the above mentioned works by the catalog of the approaches presented in [33].

Also, given that we extend the paper [33], it is worth mentioning that the main difference between that work and the one presented in [32] is the fact that the former presents details of the application of the Systematic Mapping method, highlighting all the information generated on the planning, execution and presentation phases, while [32] presents only one of the products generated by a Systematic Mapping, that is the detailing of the Catalog of Approaches for Requirement Traceability.

VI. Conclusions

In this paper we presented the process of a systematic mapping study, from its planning until the data extraction. Due to the reduced space of this paper, some information on the mapping were not presented.

The goal of this mapping was mainly to characterize the search phenomenon through data categorization and the presentation of possible approaches that might help implement requirement traceability in the context of software projects.

As a future goal, we should first highlight the further detailing of the techniques, frameworks, models and tools found as a result of this work. This activity was partly performed and published in [33], with indications of the approaches found to support the implementation of the MPS.Br quality model [8].

After that, we intend to elaborate a catalog of the techniques that support traceability in order to reach one of the main challenges proposed in the First Workshop of Challenges to Traceability (GCW'06), whose principle is that traceability is a critical factor for software projects success. Nevertheless, there is no consensus on the best techniques and methods to implement it. Hence, we expect to be able to create a database of knowledge on best practices, terminology, approaches and case studies on traceability. Besides, it is important to perform periodically a new systematic mapping on the subject at hand to analyze trens and new approaches in the line of researche on Requirement Traceability, contemplating the studies performed after 2013. This could also possibly include new keywords in the search string, as well as all the several types of traceability mentioned in Section I.

Another important future work could be the analysis of the classification of the works in the are of Requirement Engineering proposed in [35] and the exploration of how the selected studies fit into that classification.

AKNOWLEDGEMENT

This paper received funding from CAPES based on the concession of an institutional masters scholarship. This work is part of the project projeto SPIDER – *Software Process Improvement: DEvelopment and Research*, institutionalized at – Federal University of Pará (*www.spider.ufpa.br*).

REFERENCES

[1] Ferreira, P. J. A. V. Rastreabilidade de ponto de função. Dissertação de Mestrado. PPGI/UNIRIO. Rio de Janeiro, 2010.

[2] Pressman, R. S. Engenharia de software. Tradução Rosângela Velloso Penteado. São Paulo: McGraw-Hill, 2006.

[3] Kotonya; Sommerville, I. Requirements Engineering – Processes and Techniques. John Wiley & Sons, 1998.

[4] Spence, I.; Probasco, L. Traceability Strategies for Managing Requirements with Use Cases. White Paper, Rational Software Corporation. 2000.

[5] ABNT - Associação Brasileira De Normas Técnicas. NBR ISO/IEC 12207:2009 - Engenharia de Sistemas de Software - Processos de Ciclo de Vida de Software. 2009.

[6] SEI. Capability Maturity Model Integration. CMMI for Development. Version 1.3. Carnegie Mellon, USA. 2010.

[7] Softex. Melhoria do Processo de Software Brasileiro (MPS.BR) - Guia Geral 2012. Disponível em: http://www.softex.br/wp-content/uploads/2013/07/MPS.BR_Guia_Geral_ Software_20121.pdf. 2012.

[8] Thayer, R. H., Dorfman, M. Software Requirements Engineering, 2nd edition, IEEE Computer Society Press. 1997.

[9] Gotel, O. C. Z.; Finkelstein, A. C. W. An Analysis of the Requirements Traceability Problem, In: Proceedings of the International Conference on Requirements Engineering, pp. 94-101, Colorado Springs, CO, EUA, Abr. 1994.

[10] Genvigir, E. C. Um Modelo para Rastreabilidade de Requisitos de Software Baseado em Generalização de Elos e Atributos. Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada. INPE. 2009. Diponível em http://mtc-m18.sid.inpe.br/col/sid.inpe.br /mtc-m18@80/2009/03.02.14.17/doc/publicacao.pdf?languagebutton=pt -BR. Último acesso em Julho/2015.

[11] Leal, M. D. Uma abordagem semiautomática de geração de informações de rastreabilidade de software. Dissertação de Mestrado. Universidade Federal do Pará. Belém, 2011.

[12] Costa, C. S. Uma abordagem baseada em evidências para o gerenciamento de projetos no desenvolvimento distribuído de software. Dissertação de Mestrado – Programa de Pós-Graduação em Ciência da Computação - Universidade Federal de Pernambuco, Recife, PE, Brasil. 2010.

[13] Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic mapping studies in software engineering. 12th International Conference on Evaluation and Assessment in Software Engineering, pages 71–80, 2008.

[14] Petticrew, M.; Roberts, H. Systematic Reviews in the Social Sciences: A Practical Guide. Wiley-Blackwell, 2005.

[15] Kitchenham, B.; Charters, S. Guidelines for performing Systematic Literature Reviews. In Software Engineering, Technical Report EBSE-2007-01, Departament of Computer Science Keele University, Keele. 2007.

[16] Torkar, R *et al*. Requirements traceability state-of-the-art: A systematic review and industry case study. International Journal of Software Engineering and Knowledge Engineering, World Scientific Publishing, 22(03), pp. 385-433. 2012.

[17] Nair, S.; De La Vara, J. L.; SEN, S. A review of traceability research at the requirements engineering conference re@ 21. In: Requirements Engineering Conference (RE), 2013 21st IEEE International. IEEE. p. 222-229. 2013.

[18] Santos, G. (2010) "Revisão Sistemática, Mini-Curso". Simpósio Brasileiro de Qualidade de Software – SBQS. Belém – PA. 2010.

[19] Naslavsky, L.; Ziv, H.; Richardson, D. J. Towards traceability of model-based testing artifacts, Proceedings of the 3rd international workshop on Advances in model-based testing, p.105-114. London, United Kingdom. July 09-12, 2007.

[20] Dekhtyar, A.; Hayes, J. H.; Smith, M. Towards a model of analyst effort for traceability research. In Proceedings of the 6th International Workshop on Traceability in Emerging Forms of Software Engineering (TEFSE '11). ACM, New York, NY, USA, 58-62. 2011

[21] Schwarz, H.; Ebert, J.; Winter, A. Graph-based traceability: a comprehensive approach. Softw. Syst. Model. 9, 4 (September 2010), 473-492. 2010

[22] guo, Y.; Yang, M.; Wang, J.; Yang, P.; Li, F. An Ontology Based Improved Software Requirement Traceability Matrix. In Proceedings of the 2009 Second International Symposium on Knowledge Acquisition and. Modeling - Volume 01 (KAM '09), Vol. 1. IEEE Computer Society, Washington, DC, USA, 160-163. 2009.

[23] Omoronyia, I.; Sindre, G.; Stålhane, T. Exploring a Bayesian and linear approach to requirements traceability, Information and Software Technology, v.53 n.8, p.851-871, August, 2011.

[24] Hussain, T.; Eschbach, R. Automated fault tree generation and risk-based testing of networked automation systems. Emerging Technologies and Factory Automation (ETFA), 2010 IEEE Conference on, vol., no., pp.1,8, 13-16 Sept. 2010.

[25] Jiang, H.; Nguyen, T. N.; Chen, I.; Jaygarl, H.; Chang, C. K. Incremental Latent Semantic Indexing for Automatic Traceability Link Evolution Management. Automated Software Engineering, 2008. ASE 2008. 23rd IEEE/ACM International Conference on, vol., no., pp.59,68, 15-19 Sept. 2008.

[26] De Lucia, A.; Oliveto, R.; Tortora, G. Adams re-trace: traceability link recovery via latent semantic indexing. In Proceedings of the 30th international conference on Software engineering (ICSE '08). ACM, New York, NY, USA, 839-842. 2008.

[27] Sultanov, H.; Hayes, J. H. Application of reinforcement learning to requirements engineering: requirements tracing. Requirements

Engineering Conference (RE), 2013 21st IEEE International , vol., no., pp.52,61, 15-19 July 2013.

[28] Eyal-Salman, H.; Seriai, A.-D.; Dony, C. Feature-to-code traceability in a collection of software variants: Combining formal concept analysis and information retrieval. Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on, vol., no. pp.209,216, 14-16 Aug. 2013.

[29] Mirakhorli, M.; Cleland-Huang, J. Transforming trace information in architectural documents into re-usable and effective traceability links. In Proceedings of the 6th International Workshop on SHAring and Reusing Architectural Knowledge (SHARK '11). ACM, New York, NY, USA, 45-52. 2011.

[30] Mirakhorli, M.; Cleland-Huang, J. Tracing architectural concerns in high assurance systems: (NIER track). Software Engineering (ICSE), 2011 33rd International Conference on, vol., no. pp.908,911, 21-28 May 2011.

[31] Pandanaboyana, S.; Sridharan, S.; Yannelli, J.; Hayes, J. H. REquirements TRacing On target (RETRO) enhanced with an automated thesaurus builder: An empirical study. Traceability in Emerging Forms of Software Engineering (TEFSE), 2013 International Workshop on, vol., no. pp.61,67, 19-19 May 2013.

[32] Winkler, S. On Usability in Requirements Trace Visualizations. In Proceedings of the 2008 Requirements Engineering Visualization (REV '08). IEEE Computer Society, Washington, DC, USA, 56-60. 2008.

[33] Malcher, P. R. C.; Oliveira, S. R. B. Indicações de Abordagens para Rastreabilidade de Requisitos no contexto do MR-MPS-SW por meio de uma Revisão Sistemática da Literatura. X Workshop Anual do MPS. Campinas-SP. 2014. Disponível em http://www.softex.br/wp-content/uploads/2014/12/Anais-WAMPS-2014.pdf. Último acesso em Abril/2015.

[34] Bjørnson, F. O.; Dingsøyr, T. Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used. Information and Software Technology. 1055 – 1068. 2008.

[35] Wieringa, R.; Maiden, N.; Mead, N.; Rolland, C. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Eng. 11. 102 – 107, 2006.