



Prevalence, common causes and effects of technical debt: Results from a family of surveys with the IT industry[☆]

Robert Ramač^a, Vladimir Mandić^{a,*}, Nebojša Taušan^b, Nicolli Rios^c, Sávio Freire^d, Boris Pérez^e, Camilo Castellanos^f, Darío Correal^f, Alexia Pacheco^g, Gustavo Lopez^g, Clemente Izurieta^h, Carolyn Seamanⁱ, Rodrigo Spinola^j

^a University of Novi Sad, Serbia

^b INFORA Research Group, Serbia

^c Federal University of Rio de Janeiro, Brazil

^d Federal University of Bahia and Federal Institute of Ceará, Brazil

^e Universidad de Los Andes and Francisco de Paula Stder. Univ, Colombia

^f Universidad de Los Andes, Colombia

^g University of Costa Rica, Costa Rica

^h Montana State University and Idaho National Laboratories, United States

ⁱ University of Maryland Baltimore County, United States

^j Salvador University and State University of Bahia, Brazil

ARTICLE INFO

Article history:

Received 10 January 2021

Received in revised form 15 June 2021

Accepted 28 September 2021

Available online 13 October 2021

Keywords:

Technical debt

Survey

InsighTD

Causes of technical debt

Effects of technical debt

ABSTRACT

Context: The technical debt (TD) metaphor describes actions made during various stages of software development that lead to a more costly future regarding system maintenance and evolution. According to recent studies, on average 25% of development effort is spent, i.e. wasted, on TD caused issues in software development organizations. However, further research is needed to investigate the relations between various software development activities and TD.

Objective: The objective of this study is twofold. First, to get empirical insight on the understanding and the use of the TD concept in the IT industry. Second, to contribute towards precise conceptualization of the TD concept through analysis of causes and effects.

Method: In order to address the research objective a *family of surveys* was designed as a part of an international initiative that congregates researchers from 12 countries—InsighTD. At country level, national teams ran survey replications with industry practitioners from the respective countries.

Results: In total 653 valid responses were collected from 6 countries. Regarding the prevalence of the TD concept 22% of practitioners have only theoretical knowledge about it, and 47% have some practical experiences with TD identification or management. Further analysis indicated that senior practitioners who work in larger organizations, larger teams, and on larger systems are more likely to be experienced with TD management. Time pressure or *deadline* was the single most cited cause of TD. Regarding the effects of TD: *delivery delay*, *low maintainability*, and *rework* were the most cited.

Conclusion: InsighTD is the first family of surveys on technical debt in software engineering. It provided a methodological framework that allowed multiple replication teams to conduct research activities and to contribute to a single dataset. Future work will focus on more specific aspects of TD management.

© 2021 Elsevier Inc. All rights reserved.

[☆] Editor: Raffaella Mirandola.

* Corresponding author.

E-mail addresses: ramac.robert@uns.ac.rs (R. Ramač), vladman@uns.ac.rs (V. Mandić), nebojsa.tausan@infora.rs (N. Taušan), nicolli@cos.ufrj.br (N. Rios), savio.freire@ifce.edu.br (S. Freire), borisperezg@ufps.edu.co (B. Pérez), cc.castellanos87@uniandes.edu.co (C. Castellanos), dcorreal@uniandes.edu.co (D. Correal), alexia.pacheco@ucr.ac.cr (A. Pacheco), gustavo.lopez_h@ucr.ac.cr (G. Lopez), clemente.izurieta@montana.edu (C. Izurieta), cseaman@umbc.edu (C. Seaman), rodrigo.spinola@unifacs.br (R. Spinola).

<https://doi.org/10.1016/j.jss.2021.111114>

0164-1212/© 2021 Elsevier Inc. All rights reserved.

1. Introduction

Software companies are often in situations that require them to make trade-offs between software quality and preset deadlines and goals. Making such trade-offs may result in short-term wins, but long-term it may introduce friction in software development and maintenance processes. This is aptly described with a metaphor of *technical debt* (Avgeriou et al., 2016; Kruchten et al., 2012; Shull et al., 2013; Rios et al., 2018a). The concept

of technical debt (TD) describes a phenomenon that impacts software projects and makes them difficult to manage (Rios et al., 2018a). For example, lack of coding standards can result with bad code, i.e. code debt and code smells (Alves et al., 2016; Rios et al., 2018a; Zazworka et al., 2014). It seems that TD issues are prevalent in organizations and teams of all sizes. Some recent studies indicated that on average 25% of development effort is spent—wasted—on TD caused issues (Besker et al., 2019, 2018a).

The technical debt metaphor was coined by Ward Cunningham in the 90s with an explicit intention to better communicate important aspects of internal software quality—software attributes that are not directly visible to the users (Buschmann, 2011; Rios et al., 2018a; Kruchten et al., 2012). Interestingly in the early 2000s, software engineers were more occupied with the concept of value—software attributes that are directly visible to the users—and with approaches to deliver that value to the market more efficiently, e.g. Agile approaches, lean software development, Value-based Software Engineering (Beck et al., 2001; Mandić et al., 2010; Boehm, 2003). Just in recent years, the concept of TD is gaining more attention from the research community as well as from the industry (Rios et al., 2018a).

Essentially, TD represents *shortcuts* made during various stages of software development that result with a more costly future regarding system maintenance and evolution. A more elaborate definition was formulated by McConnell (2008) as: *technical debt contextualizes the problem of outstanding software development tasks as a kind of debt that brings a short-term benefit to the project that may have to be paid later in the development process with interest*. For example, a poorly designed class tends to be more difficult and costly to maintain than if it had been implemented using good object-oriented practices. Furthermore, McConnell distinguishes two kinds of technical debt: *intentional*—purposefully injected by acknowledging benefits of doing so—and *unintentional*—result of poor engineering practices (McConnell, 2008). Beside this definition, the literature recognizes many other definitions of TD (Avergiou et al., 2016; Kruchten et al., 2012; Shull et al., 2013; Rios et al., 2018a).

By definition, TD is a trade-off between short-term benefits and long-term effects, where taking short-term benefits causes some long-term issues, i.e. *negative* consequences. However, there are some indications that in some situations injecting TD can have overall *positive* effects, e.g. start-up companies (Besker et al., 2018b).

In past years, a significant effort was made by researchers to better understand and identify technical debt from the analysis of source code, i.e. code debt and code smells (Alves et al., 2016; Rios et al., 2018a; Zazworka et al., 2014). However, this is just one perspective. Software development can be seen as a chain of various decisions (business, design, architecture, etc.), which starts well before coding, and all those decisions impact the creation of a product (Mandić et al., 2010), and consequently the creation of technical debt. One of the conclusions of the Dagstuhl Seminar was that all software development artifacts can carry technical debt (Avergiou et al., 2016). Therefore, we need to better understand the connections between non-coding activities and TD, such a cognition will allow us to balance a delivered value of software with accompanying technical debt.

The nature of the TD phenomenon itself calls for empirical approaches with sufficiently large and representative datasets. Such empirical studies are still rare. However, there are empirical studies focused on understanding TD causes, effects, and practices to manage the accumulated debt (Martini and Bosch, 2017; Martini et al., 2014; Yli-Huumo et al., 2016; Yli-Huumo et al., 2014; Ernst et al., 2015). For example, Yli-Huumo et al. conducted an interpretive case study with two mid-sized organizations in Finland, and they reported 7 scenarios of why workarounds are

made in projects (Yli-Huumo et al., 2016; Yli-Huumo et al., 2014). Martini et al. investigated relations between software architecture and TD items (Martini et al., 2014). In their case study, they involved 7 study units with 25 participants (Martini et al., 2014). In 2015, Ernst et al. conducted a study—so far, the largest empirical investigation of TD—with 536 participants from 3 large organizations (Ernst et al., 2015). The results of their study confirmed the usefulness of the metaphor, but at the same time they accentuated the lack of existing tools for managing TD. The existing empirical studies are lacking the heterogeneity of the data sample, e.g. studies with large number of participants are conducted with a few organizations or in one country. Therefore, we need further empirical investigations that aim at increasing the representativeness of the sample and collected dataset.

The objective of this research is to collect representative empirical data about the TD concept, foremost about its prevalence within the IT industry alongside with the identification of TD causes and effects. This objective is further elaborated with the following research questions:

- RQ1: To what extent are software professionals familiar with the concept of TD?
- RQ2: What causes lead software development teams to incur TD?
- RQ3: What effects does TD have on software projects?

In order to address the research objective a *family of surveys* was designed as a part of an international initiative that congregates researchers from 12 countries—InsighTD.¹ This is the first aggregated data analysis of the *InsighTD* survey. The empirical dataset consists of 653 valid responses from industry practitioners in 6 countries. By publishing this report, we finalized the analysis of data for basic TD questions about the prevalence, causes, and effects. As a part of InsighTD research roadmap other aspects of TD management will be addressed in the future.²

Findings presented in this paper are of interest for industry practitioners as well as for researchers in the field of TD, who can benefit from the insights about the most common causes and effects of TD. The contributions of the paper are:

- (1) Regarding the familiarity of the TD concept within the industry, 31% of participants never heard about it, while 27% of those who are familiar with it, have actively managed TD on projects. Furthermore, the only demographic variable that does not have significant effect on the TD familiarity is *system age* ($\chi^2 = 15.192, p = 0.2311$), while the other variables have a significant effect.
- (2) Our findings show that various TD types are present in the industry. There is no single TD type that significantly stands out. We identified the top 5 TD types by occurrence as: *design debt* (21.99%), *test debt* (19.94%), *code debt* (14.66%), *architecture debt* (10.7%), and *documentation debt* (9.09%).
- (3) Conceptualization of TD causes with a mind map, i.e. visualization of categories of causes and their frequencies. Identification of major TD causes. The single most cited TD cause is *deadline*, i.e. time pressure caused by short deadline.
- (4) Conceptualization of TD effects with a mind map, i.e. visualization of categories of effects and their frequencies. Identification of major TD effects. The top three effects of TD are *delivery delay*, *low maintainability*, and *rework*.

¹ <http://www.td-survey.com>.

² InsighTD research roadmap and publication map can be found on the link: <http://www.td-survey.com/publication-map/>.

The rest of the paper is organized as follows. After discussing the related work and our previous work in Section 2, we describe the research approach in Section 3. Section 4 presents the results of the survey. In Section 5 we discuss findings and observations. Trustworthiness of the study is discussed in Section 6, followed by our final remarks in Section 7.

2. Related work

This section starts with the discussion about survey studies on TD (Section 2.1). After this, Section 2.2 describes the related work on causes and effects of TD. Our literature review process started with an initial dataset of references. The initial dataset was collected during our previous meta-study of systematic literature reviews on TD (Rios et al., 2018a). From the initial dataset we selected references that were focusing on empirical studies of TD causes and effects. The selected references were used as starting references for a form of the snowballing search. Two main inclusion criteria were: (1) survey-based research, i.e. empirical studies that used a survey as the main instrument, and (2) empirical studies focusing on TD causes and TD effects.

Finally, in Section 2.3 we discuss our previously disseminated results on the topic and how this work complements them.

2.1. Survey research on TD

Technical debt and related topics became a subject of empirical investigations after 2010. The early empirical investigations used qualitative, interview based, approaches (Taksande, 2011), because little was known about the phenomena. Later, this was followed by observational studies, experience reports, and case studies (Siebra et al., 2012; Morgenthaler et al., 2012; Tom et al., 2013; Zazworka et al., 2013). This early accumulation of knowledge about the TD phenomena paved a road for more structured empirical investigations using surveys. In this section, the extent and representativeness of the survey-based research on TD is presented.

Table 1 presents surveys on TD excluding surveys and replications of the InsignTD project. Also, surveys that were used in case studies as pre-study or follow-up questionnaires are excluded. Since 2014, there were 11 conducted surveys. Regarding the sample size, it varies from 40 in (Silva et al., 2019) up to 536 in (Ernst et al., 2015). Research focus of the surveys mainly revolves around general TD topics, including, TD causes and effects, monitoring and management practices, productivity issues, and software quality (Table 1).

The Ernst et al. (2015) survey is the largest with 536 valid answers, however the survey was conducted in 3 large organizations thus the results of the survey can be biased towards large organizations. On the other hand, Codabux et al. (2017) managed to collect responses from 67 participants in 33 organizations resulting with a good diversity of company sizes. Furthermore, the study of Codabux et al. has international character, involving participants from 14 countries.

Overall, 3 studies are national (Ernst et al., 2015; Rocha et al., 2017; Apa et al., 2020), 4 studies are international (Holvitie et al., 2014; Codabux et al., 2017; Holvitie et al., 2018; Arvanitou et al., 2019), and for the remaining 4 studies the country was not reported (Table 1).

In summary, although the survey as a research instrument is convenient for collecting responses from large samples, evidently survey research needs to deal with challenges of low response rates and consequently lower representativeness of the population. This was one of the main motivations to choose an international family of surveys for the research approach of this study.

2.2. Research on causes and effects of TD

Avgeriou et al. (2016) indicated that the causes of TD can be a process, a decision, an action (or lack thereof), or an event that triggers the existence of a TD item, such as schedule pressure, unavailability of a key person, or lack of information about a technical feature. Also, they reported that TD can affect the value of the system, the costs of future changes, the schedule, and system quality. Those perceptions on causes and effects of TD reflect the consolidation of the opinion of a couple of researchers and practitioners during the Dagstuhl Seminar 16162. Nevertheless, there are other works on the topic, before and after the mentioned seminar, that support and, also, add new information to the TD causes and effects body of knowledge.

Codabux and Williams (2013) conducted an industrial case study to understand how TD affects the adoption of Agile software development. The study used participant observations, semi-structured interviews, and a questionnaire for collecting data. The authors analyzed the data using a coding scheme, where each question was related to a theme. For example, causes and impacts were associated with motivations behind incurring TD and consequences of TD, respectively. In total, 28 practitioners (product owners, scrum master, developers, and testers) participated in the case study. Regarding the causes of TD, the practitioners indicated that managerial decisions and resource constraints (a limited timeframe and the unavailability of developers to continue working on the feature) are the main causes for incurring TD. Starting the project from scratch was the main effect felt by practitioners.

Yli-Huumo et al. (2014) performed a case study of two product lines in a midsize Finnish software company for understanding the relationship between TD causes, effects, and management. The data was collected from interviews, and qualitative data analysis techniques were also applied. From 12 interviews, the authors found that the lack of time given for development, pressure put on the development team, complexity of the source code, business decisions (lack of technical knowledge and communication challenges), lack of coding standards and guides, junior coders, lack of knowledge about future changes, and lack of documentation are the causes of TD. On the other side, the effects recognized in the case study were time-to-market benefit, increased customer satisfaction, extra working hours, errors and bugs, customer dissatisfaction, and complexity of the source code.

In another study of the area, Ernst et al. (2015) performed an industrial survey to identify how software practitioners understand the TD concept and whether they use practices and tools for managing TD items. The study also considered a follow-up step with one-hour interviews for collecting data. In the survey, the authors asked that the participants order a predefined list of 14 causes of TD using as criterion the amount of debt that each cause could bring to the project. The survey received 536 responses, whereas the interviews were conducted with seven respondents. As a result, the authors identified that immature choices done in architecture are a key cause of TD.

Yli-Huumo et al. (2015) conducted a case study in two software organizations for identifying the benefits and consequences of taking workarounds. The data collection was performed by using semi-structured interviews with 17 practitioners. Open coding was used for analyzing the data and identifying categories related to the workarounds. The authors found that time pressure, complicated code base, no time for changing the selected software components, and prioritization of features based on their business value were the causes of workarounds. Also, the effects identified were decreased code maintainability, extra working hours, extra costs, major refactoring, lack of motivation to work with the code base, increased time for newcomers to start, outdated software components, lack of new features available in

Table 1
Survey-based research on TD.

No.	Reference	Research Focus	Year	Representativeness		
				Part.	Org.	Count.
1.	(Holvitie et al., 2014)	General TD in Agile software development.	2014.	80	–	3
2.	(Ernst et al., 2015)	General TD causes.	2015.	536	3	1
3.	(Rocha et al., 2017)	TD management, focusing on the code quality and code debt.	2017.	74	–	1
4.	(Codabux et al., 2017)	General TD and TD risk assessment.	2017.	67	33	14
5.	(Besker et al., 2017)	Estimation of TD costs.	2017.	258	–	–
6.	(Holvitie et al., 2018)	General TD in Agile software development.	2018.	184	–	3
7.	(Besker et al., 2018a)	TD and software developer productivity.	2018.	43	–	–
8.	(Martini et al., 2018)	TD monitoring and tracking.	2018.	226	15	–
9.	(Arvanitou et al., 2019)	TD monitoring.	2019.	60	11	9
10.	(Silva et al., 2019)	TD perception and TD management.	2019.	40	12	–
11.	(Apa et al., 2020)	General TD and TD management.	2020.	259	–	1

newer versions of components, and decreased code maintainability when scaling the feature.

Through a systematic mapping study, Li et al. (2015) recognized that the following quality attributes are compromised due to the presence of debt: maintainability, reliability, security, portability, and performance efficiency.

Concerning architectural TD, Martini et al. (2014) performed a multiple case study including seven Scandinavian sites in five large software development organizations. The goal of the study was to investigate the causes for the accumulation of architectural TD and to investigate the trends in the accumulation and recovery of architectural TD. Through interactive workshops, the authors collected the data and analyzed it using open and axial coding, and a deductive approach. The authors identified that causes of TD are related to business factors, design and architecture documentation (lack of specification/emphasis on critical architectural requirements), reuse of legacy/third party/open source, parallel development, effects uncertainty, non-completed refactoring, technology evolution, and the human factor. In 2017, Martini and Bosch (2017) included two other Scandinavian sites and one large organization in that study to understand which items of architectural debt are the most dangerous for accumulated interest. The authors performed interviews and workshops, and identified a set of effects related to architectural TD, organizing those effects in a model.

In another case study, Martini and Bosch (2015) investigated the factors that should be considered during TD prioritization. They used questionnaires and interviews for collecting data. The authors reported that the effects of architectural TD are big deliveries, many code changes, “double” the effort, number/complement of tests, quality issues (bugs), wrong effort estimation, contagious architectural technical debt, probable hidden architectural technical debt, and developers idling.

In summary, the current evidence on causes and effects of TD is limited to the point of view of a small set of practitioners not providing a complete view on causes that lead to the occurrence of TD and effects felt by practitioners due to the presence of TD.

Technical debt can be injected into almost any artifact and during any stage of software development resulting with different types of technical debt—TD types—e.g., code, architecture, design, documentation, test, process, or people debt (Rios et al., 2018a). Elaborate definitions of TD types are given in Section 4.1.3 alongside with survey results.

2.3. Previous work

This work is part of the InsignTD project, which intends to investigate the causes, effects, and management of TD (Rios et al., 2018b). Several papers have already been disseminated in the technical literature reporting the results of the project. In this subsection, we discuss our previous work on causes and effects

of TD, and show how this work composes the big picture we are drawing with InsignTD.

Rios et al. (2018b) briefly presented the design of the project, along with the discussion of the top 10 causes and effects of TD recognized in the Brazilian software industry. After, the whole set of causes and effects from the Brazilian replication of the study was organized into probabilistic diagrams for supporting the analysis of TD causes and effects in software projects (Rios et al., 2019). Based on this work, Rios et al. (2020) reported a detailed discussion about the design of the InsignTD family of surveys and described the results of new analyses performed over the first execution of InsignTD in Brazil.

Replications of the survey have been performed in several countries and the initial analysis on causes and effects of TD have been incrementally reported. Pérez et al. (2020) identified the causes of TD in the Chilean software industry and Ramač et al. (2020) recognized the causes and effects of TD in the Serbian software industry. Both studies compare their results to the ones reported by Rios et al. (2019). So far, considering only the results of the isolated replications' results, we have investigated the following research questions: (1) *Are software professionals familiar with the concept of TD?* (2) *What causes lead software development teams to incur TD?* and (3) *What effects does TD have on software projects?*

However, a consolidated view of causes and effects of TD across the available³ replications of InsignTD is still missing. Moreover, until recently, the InsignTD dataset has not been large enough to make specific data analysis on causes and effects of TD. This paper represents the first opportunity to explore such specific analyses.

For more information on the big puzzle we are working on within the InsignTD project, including other research variables, an interested reader can visit <http://www.td-survey.com/publication-map/>.

3. Research approach

The research approach encompasses the selection of the research strategy, organization, and execution of the methodological tasks. Consequently, the presentation of the research approach consists of: (a) selected research strategy and the argumentation in favor of the selection, (b) InsignTD project as a managing organization that guides the study, (c) design of the survey, which is the main research instrument, (d) survey family as a mode of survey execution, (e) data collection, with emphasis on sampling, (f) data analysis, presenting both quantitative and qualitative methods, and (g) the aggregation of the collected survey family data.

³ Currently, we already have the data from Brazil, Chile, Colombia, Costa Rica, Serbia, and the United States.

Table 2
Argumentation for Survey Research Strategy in InsignTD.

No	Survey Research Strategy	InsignTD project
1	Survey can be used to ask question about what has happened in the past (Kitchenham and Pfleeger, 2003; Ghazi et al., 2019).	Existing experience with TD, collected from IT practitioner, is in the focus of the analysis in InsignTD.
2	Surveys can easily collect large number of variables (Wohlin et al., 2006).	Each RQ in InsignTD project is related with a subset of questions in the survey.
3	Survey allows to form an assertions (or claims) about target population (Babbie and Seacat, 1973)	Answer to RQ1 represent an assertion about the familiarity with the TD concept.
4	Survey allows to form an explanation about target population (Babbie and Seacat, 1973)	RQ2 and RQ3 represent explanations about the causes and effect of TD.
5	Survey research allows the utilization of both qualitative and quantitative methods for analysis (Wohlin et al., 2006).	InsignTD relies on both qualitative and quantitative methods for analysis.
6	Surveys are cost effective, time effective, easy to distribute especially if the survey is implemented as web-based (or online) survey (Lethbridge et al., 2005).	In InsignTD, survey was implemented and distributed as online survey. Cost and time effectiveness, and ease of distribution are useful properties.

3.1. Research strategy

The InsignTD project adopted the *survey research strategy* (Kitchenham and Pfleeger, 2003; Molléri et al., 2016) as an overall methodological guidance for designing a family of surveys. In short, the survey research strategy consists of methodological steps that center on a survey or questionnaire⁴ to collect the data from the sample of the population. Based on the analysis of that data, survey research aims to draw conclusions about the entire population (Babbie and Seacat, 1973), or to collect empirical data, which is a more attainable aim for survey research in the software engineering field. The decision to rely on this research strategy was made since it is seen as the most suitable for achieving the research objective. To argument this decision, Table 2 presents how some of the main properties of the survey strategy are related with InsignTD.

The implementation of the survey research strategy in InsignTD followed the recommendations for survey research in software engineering proposed by Molléri et al. (2016). The following sections present the adoption of these recommendations in InsignTD. To facilitate the presentation, these sections refer to Fig. 1 which relates the organization of InsignTD with the main research activities.

3.2. The InsignTD project and research questions

The InsignTD project is an international initiative that focuses on the TD phenomenon and its manifestation in the software development process. The project organization consists of the core team and national replication teams. The core team is an international group of senior scientists who led the project preparation and made all relevant decisions regarding the project organization and research. The core team also coordinates the efforts of national replication teams. Replication teams are groups of scientists from each country that participates in InsignTD. At the time of writing this article, 34 researchers organized in national teams from 12 countries joined the initiative.⁵ The main responsibility of the replication teams is to execute the InsignTD tasks in their own countries but also to facilitate the work of the core team, or to contribute members to the core team. In Fig. 1, the core and replication teams are presented with rectangles, whereas the replication teams are additionally emphasized as rounded rectangles.

The core team has set the foundation of the InsignTD project by defining the goal, research questions and by selecting the

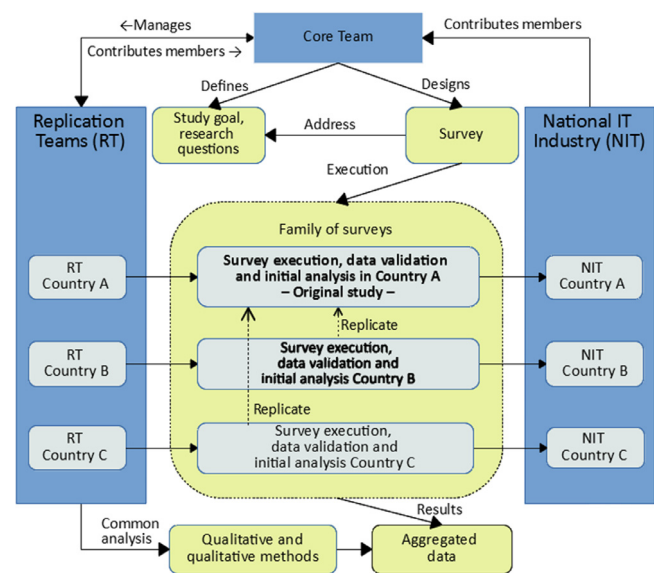


Fig. 1. InsignTD project organization.

overall research strategy. The goal of InsignTD is to produce generalizable knowledge on the TD phenomenon and its manifestation in the software development process. To achieve this goal, the project focuses on several aspects of TD, three of these aspects were in the focus of this study and articulated in the following research questions:

RQ1: To what extent are software professionals familiar with the concept of TD?

RQ2: What causes lead software development teams to incur TD?

RQ3: What effects does TD have on software projects?

InsignTD seeks to investigate the causes of TD, without differentiating whether the debt is incurred intentionally or unintentionally, and the effects without differentiating whether they positively or negatively impact the development process. For this reason, RQ2 and RQ3 remained agnostic with respect to these aspects, so were the survey participants since they were not instructed to consider them either.

Number of InsignTD studies that focus on one or more of the presented RQs have already been published within the InsignTD project⁶. These studies, however, relied on partial datasets, e.g.

⁴ Both terms are used to denote the list of questions that participants are asked to answer.

⁵ Project statistics on 28.Nov.2020.

⁶ <http://www.td-survey.com/publication-map/>.

data collected in a single country, or data collected in companies using agile approach. This study, is the first study whose results are based on an aggregated dataset, and thus brings us closer to generalizable knowledge on TD which is the goal of the InsignTD project.

3.3. Motivation for, and design of InsignTD family of surveys

Software engineering recognizes two major approaches for analyzing aggregations of empirical studies (Shull et al., 2008; Santos et al., 2019): (1) systematic literature reviews (SLR), i.e. meta-studies of published research, and (2) replications of empirical studies, i.e. a *family of empirical studies*. The advantages of replications over SLRs are:

- (1) Replications share the same research goal, and if they share the same research questions then we call them exact replications (Santos et al., 2018, 2019).
- (2) Replications are conducted by independent researchers or research teams.
- (3) Research design (research protocol and instruments) is shared by all replications, which allows for the collection of data in a uniform and consistent way.
- (4) Often, data analysis protocols are shared by replications.
- (5) All researchers have access to raw data, unlike SLRs, and therefore an aggregated data analysis is possible.
- (6) Joint conclusions are not affected by publication bias (Santos et al., 2019).

The replications, that is *families of empirical studies*, are therefore perceived to have higher internal validity over SLRs. Families of surveys in the SE field are still rare and one example is from the requirements engineering domain (Wagner et al., 2019). To the best of our knowledge, InsignTD is the first family of surveys in the TD domain.

In survey research, the questionnaire (or the survey) is a central research instrument and it is used to collect data from the target population (Babbie and Seacat, 1973). The survey used in InsignTD was designed by the core team, and the details about this design are already presented in scientific publications (Rios et al., 2018b, 2020). The following sections present the most relevant information regarding the survey and its design.

The InsignTD survey⁷ consists of 28 questions and these questions are organized into five parts. The first part are the demographics questions that are used to characterize the respondents, while the subsequent three parts aim to elicit the data for answering the three stated RQs respectively. Fifth part targets TD management aspects that are not part of this study. Demographic questions are all closed-type questions, while the other parts use both closed and open-ended type questions. A subset of survey questions related to RQs addressed in this study are presented in Table 3. To design a survey, the core team relied on the existing knowledge from scientific sources and on insights from senior researchers and industry experts, as presented in Rios et al. (2018b).

Once the survey was designed, replication teams were free to execute the survey as is, or they could introduce minor adaptations. For example, if needed, replication teams translated the questions to the national language and allowed answers to open-ended questions to be both in English and in the national language. The reasons for these adaptations were the insights from the pilot study which suggested that they can improve the quality of the responses (Mandić et al., 2020). This ensured that the language barrier was neutralized and that the survey can now be

conducted in any country that joins the InsignTD initiative and that the aggregated analysis could now be complemented with data from various countries.

Following the methodological steps that are set in the InsignTD project, the designed survey was executed as a *family of surveys*. In InsignTD, a family of surveys was realized by independent execution of surveys and analysis by replication teams. Each replication team executed the survey in their country, that is in their national IT industry and followed the identical procedure for the analysis. In the central part of Fig. 1, the survey family is presented with a dotted-lined rounded rectangle which contains blocks for each survey executions of each replication team within their national IT industry. Further more, Fig. 1 also shows that replicated surveys addressed the same study goal and RQs, and that the data is aggregated in a common data repository and then analyzed.

The first survey in this survey family was executed by the Brazilian team (Rios et al., 2018b), thus this study is referred as the *original study*. All subsequent survey executions, executed by other replication teams, replicate the original study and therefore are referred as the replication studies. In Fig. 1, the block with the original study is emphasized with bold letters. The replication studies were presented below and related with the original study with the dotted arrows. The survey was implemented as an online survey.

3.4. Data collection

TD represents a concept specific to software engineering practice, hence the target population for this survey consists of IT practitioners working in the software development industry. The sampling from this target population was done following the *convenience sampling* approach (Wohlin et al., 2012; Baltes and Ralph, 2020). In short, convenience sampling is a non-probability sampling approach where the sample is drawn based on convenience, or availability of the subjects. This sampling approach was seen as suitable since any IT practitioner in any role working in the IT industry was seen as a suitable candidate for the survey. Consequently, each replication team sent survey participation invitations using all or some of the following channels: (a) invitation using their personal contacts who work in the IT industry, (b) asking their contacts to further disseminate the invitation to their colleagues, (c) in some countries, IT cluster organizations were contacted and asked to distribute the invitation among their members, and lastly (d) IT practitioners were invited by posting the invitation to specialized forums using internet-based platforms and, such as LinkedIn⁸ or Facebook.⁹

Upon survey execution, each replication team was responsible for validating the data. The validation of open-ended questions was done by checking (i) the completeness of the answers and (ii) did the respondents provide a valid TD example. The completeness check was done by making sure that the answers properly address the questions. The evaluation of the TD example focused on the answer to Q13 where the respondents were asked to describe an example of TD. Depending on whether the provided example could be related with one of the TD types, the answers were used or discarded during latter analysis. This evaluation revealed whether the respondents properly understand the TD phenomenon and therefore increased the credibility of their answers. Each answer was read by at least three scientists, and to discard the answer that was seen as invalid, a consensus of at least two scientists was required. Finally, the validation of closed questions was ensured by constraints set by the online survey, and it was publicly accessible since mid 2018 until the end of 2019. As the online survey tool *Google forms* were used.

⁷ Survey description and the final version of the survey is available on following link: <https://bit.ly/2SCOM9S>.

⁸ www.linkedin.com.

⁹ www.facebook.com.

Table 3

Subset of the InsignTD Survey's Questions for demographics, RQ1, RQ2, and RQ3. Adapted from Rios et al. (2018b).

RQ	Qn	Question	Type
Demographics	Q1	What is the size of your company?	Closed
Demographics	Q2	In which country you are currently working?	Closed
Demographics	Q3	What is the size of the system being developed in that project? (LOC)	Closed
Demographics	Q4	What is the total number of people of this project?	Closed
Demographics	Q5	What is the age of this system up to now or to when your involvement ended?	Closed
Demographics	Q6	To which project role are you assigned in this project?	Closed
Demographics	Q7	How do you rate your experience in this role?	Closed
Demographics	Q8	Which of the following most closely describes the development process model you follow on this project?	Closed
RQ1	Q9	How familiar you are with the concept of Technical Debt?	Closed
RQ1	Q10	In your words, how would you define Technical Debt?	Open
RQ1	Q11	How close to the above TD definition is your understanding about TD?	Closed
RQ1	Q12	Are there any parts of the definition above from McConnell that you disagree with? Are there some things that you think are TD that are not included in this definition?	Open
RQ1	Q13	Please give an example of TD that had a significant impact on the project that you have chosen to tell us about:	Open
RQ1	Q14	Why did you select this example?	Open
RQ1	Q15	About this example, how representative it is?	Closed
RQ2	Q16	What was the immediate, or precipitating, cause of the example of TD you just described?	Open
RQ2	Q17	What other cause or factor contributed to the immediate cause you described above?	Open
RQ2	Q18	What other motives or reasons or causes contributed either directly or indirectly to the occurrence of the TD example?	Open
RQ2	Q19	Considering all the cases of TD you've encountered in different projects, and the causes of those TD cases, which causes would you say are the most likely to lead to TD (ordered by likelihood of causing TD)? Please list up to 5 causes.	Open
RQ3	Q20	Considering the TD item you described in question 13, what were the impacts felt in the project?	Open
RQ3	Q21	Considering all the cases of TD you've encountered in different projects and the effects of that TD that you have personally experienced, which 5 effects would you classify as the effects that have a bigger impact (ordered by their level of impact).	Open

3.5. Data analysis

The data is analyzed using both quantitative and qualitative research methods. Descriptive and median statistics, alongside with different charting and diagramming techniques, were the most frequently used quantitative methods in replication studies. In this study, descriptive statistics with diagrams and charts were used to analyze the closed-type questions. The respondents answered these questions by selecting one of the choices that were offered in form of categorical, ordinal or interval scale and these choices were counted and analyzed. Additionally, the Pearson χ^2 test with contingency tables was used to analyze the relations between the pairs of observed variables. Relations that were seen as interesting, i.e. statistically significant, were presented with mosaic plots. Mosaic plots with shading (Brown and Preece, 1992) were chosen as a representation tool due to their expressiveness which allows them to show both the relation between two categorical variables, and the strength of that relation. To calculate the χ^2 statistics and generate the mosaic plots, the R statistical tool¹⁰ was used.

The answers to open-ended questions were analyzed using the *qualitative coding technique*, which is one of the qualitative methods (Miles and Huberman, 1994). In short, coding denotes the process of labeling pieces of text which are seen as interesting with a meaningful name, i.e. the *code*. This analysis consisted of three main steps. The first step was the thorough familiarization with the survey answers. The second was the *coding* of the answers. This step is also refereed to as *open coding*, since everything that is seen as interesting was coded. The third step is the grouping of codes to *categories*. During this step, the open codes are re-read and related based on their similarities to a higher-order code, that is to its category. This step is also refereed as axial coding. Each national team coded the dataset collected in their national context. To derive the codes and categories, a consensus of at least two scientists was required.¹¹ Resulting codes, categories and their relations were described and interpreted.

The execution of the third step however differed in the original study and in replications. The third step in the original study consisted of axial coding, while in replications, besides axial coding, replication teams had to map their codes to the ones derived in the original study. To map the codes the replication teams relied on code descriptions from the original study and on the consensus among replication team members to confirm the mapping. This way, the list of codes and categories evolved with each replication by confirming existing and by adding new codes and categories to the list from the original study. The consolidated lists of codes and categories for TD causes and effects are given in Sections 4.3.1 and 4.3.4.

3.6. Aggregation of data and results

The data collected in different countries is aggregated into a single common dataset, however the aggregation differed for answers on closed and open questions. The answers to closed questions were merged into a common dataset as is, since these questions offer predefined answers to choose from. Only in cases when the survey is translated to the national language, the chosen answers were first translated to English, and then merged into a common set. The answers to open-ended questions however, are in free form text. This text was provided in national languages or in English. Therefore, instead of merging the raw data, each replication team completed the analysis of the data, and then contributed the analysis results into a common, consolidated dataset. This was possible, since each replication relied on the initial coding scheme that was derived in the original study, thus the results of the replication studies confirmed the existing codes and possibly further enriched them with new ones.

For example, to aggregate the answers to open-ended questions relevant for answering RQs stated in this study, replications' data from Brazil, Chile, Colombia, Costa Rica, United States and Serbia were used. Each replication team analyzed the relevant answers, identified, and described codes and categories. The provided descriptions made the meaning of codes and categories understandable to each replication team, which in turn facilitated the merger of codes and categories into a consolidated dataset. This consolidated set of data was used to produce the results that are presented in the following section.

¹⁰ <https://www.r-project.org/>.

¹¹ Analysis procedure that is used for RQ1 is available online: <https://bit.ly/3fpfY57>. Readers can familiarize with the level of details, and explanations in analysis procedures.

Table 4
Participants per country.

Replication	Responses	Valid Resp.	Valid Resp. (%)
Brazil	112	107	16.39%
Chile	104	89	13.63%
Colombia	159	134	20.52%
Costa Rica	156	145	22.21%
Serbia	91	79	12.10%
United States	123	99	15.16%
Total	745	653	100%

4. Results

The data collection resulted with 745 responses from practitioners that came from Brazil, Chile, Colombia, Costa Rica, Serbia and the United States. Table 4 shows the number of participants per country. The column *responses* represents the number of practitioners that answered the survey in the respected country, while the column *valid responses* represents the number of participants that remained after the data was validated. In total, 12.35% of the dataset was discarded, leaving 653 valid responses.

The study results are based on an aggregated analysis that took into account valid responses from participants from each country. The countries had the following share of the total number of valid responses: Brazil 16.39%, Chile 13.63%, Colombia 20.52%, Costa Rica 22.21%, Serbia 12.10% and the United States 15.16% (Table 4).

The demographics section presents the exemplar project in Section 4.1.1, participants in Section 4.1.2 and types of TD in Section 4.1.3. Section 4.2 reveals the identified relations between various demographic variables and the familiarity with the TD concept. Finally, the identified causes and effects were related with the TD types and this relation is presented in Sections 4.3. The presentation of results also includes the authors' interpretation of the results in the form of enumerated findings.

4.1. Demographics data

To collect the demographic data, survey questions Q1–Q8 were used, and these questions were all of closed type. Before the start of the survey, however, participants were also instructed to think of an exemplar project. The exemplar project would then serve as a context or as a reference point to which other survey questions would refer to.

4.1.1. Exemplar project

The exemplar project is a project in which the participant was involved and which he or she deems as representative from the TD point of view. Participants were instructed that having a clear idea about the exemplar project is an important part of the survey, since the majority of answers had to be provided from the perspective of this project.¹² For example, in Q13, participants were asked to describe the example of TD that they encountered in the exemplar project. Later during the analysis, the TD example was mapped to one of the TD types. Depending on whether the TD example could, or could not be mapped to a TD type, the answer was accepted or discarded from further analysis.

The exemplar project is characterized with data regarding the development approach, system age, system and team size. This data is illustrated in Fig. 2. The most dominant development approaches used were the hybrid approach (44.72%) that consisted of combining agile approaches with more traditional

development methodologies (e.g. waterfall), and the agile approach (41.96%). The remaining 13.32% of the participants stated that some traditional methodology was used in the development of the exemplar project.

More than half of the specified systems were between 1 and 5 years of age (56.51%), while others were scattered across less than 1 year old (17.15%), between 5 and 10 years old (15.31%) and older than 10 years of age (11.03%).

Regarding system size, most of the systems were between 10KLOC and 1MLOC (65.24%). Systems with less than 10KLOC took up 13.78% of all systems while systems between 1MLOC and 10MLOC took up 13.63%. Systems larger than 10MLOC accounted for only 7.35% of all specified systems in the survey.

Teams that worked on the exemplar project ranged from teams that counted less than 5 people to large teams that had more than 30 people. The majority of teams counted between 1 and 20 people (76.42%) while the remaining 23.58% was distributed across teams that counted from 21 to 30 people (7.5%) and more than 30 people (16.08%).

4.1.2. Participants

During the survey, the participants were asked to choose the role they had in the exemplar project from the provided list, or to specify the role if it was not offered in the list. The provided list of roles included for example: developer, tester and software architect roles. The complete list of proposed roles is enumerated in Table 7. Afterwards, the participants were asked to specify the level of experience they had while working in the specified role. The experience scale included: Novice (minimal or textbook knowledge without connecting it to practice), Beginner (working knowledge of key aspects of practice), Competent (good working and background knowledge of area of practice), Proficient (depth of understanding of discipline and area of practice), and Expert (authoritative knowledge of discipline and deep tacit understanding across area of practice).

The results show that in total there were (note that the percentages were rounded): 2% novice, 12% beginner, 34% competent, 32% proficient, and 21% expert participants. It can be concluded that only 14% of the participants can be classified as juniors (novice and beginner) that have little experience, while 87% can be classified as experienced in their roles. Table 5 shows the distribution of different experience levels over various roles. The roles marked with an asterisk (*) are roles that were not predefined in the InsignTD questionnaire, thus represent the roles defined by participants. These roles were bundled into the *other* category. The most dominant role is the developer¹³ role that accounts for 50% of all respondents. The majority of respondents for this role can be classified as experienced (84.71%).

Regarding company size, the participants are somewhat equally distributed into the following company sizes: 32.16% in small companies that count up to 50 employees, 38.44% in medium sized companies that have 51 to 1000 employees and the remaining 29.4% work in large companies that count more than 1000 employees.

4.1.3. Technical debt types

Types of TD were derived based on the exemplar project during which the survey participants recognized or experienced situations that led to the injection of TD. The analysis of valid responses resulted with the identification and extraction of the TD Types from the described situations. Fig. 3 depicts the distribution of TD Types.

¹² Survey description and the final version of the survey is available on following link: <https://bit.ly/2SCOM9S>.

¹³ the developer role often spans over the entire software development process, e.g. it often includes system design and architecture activities, testing and so on.

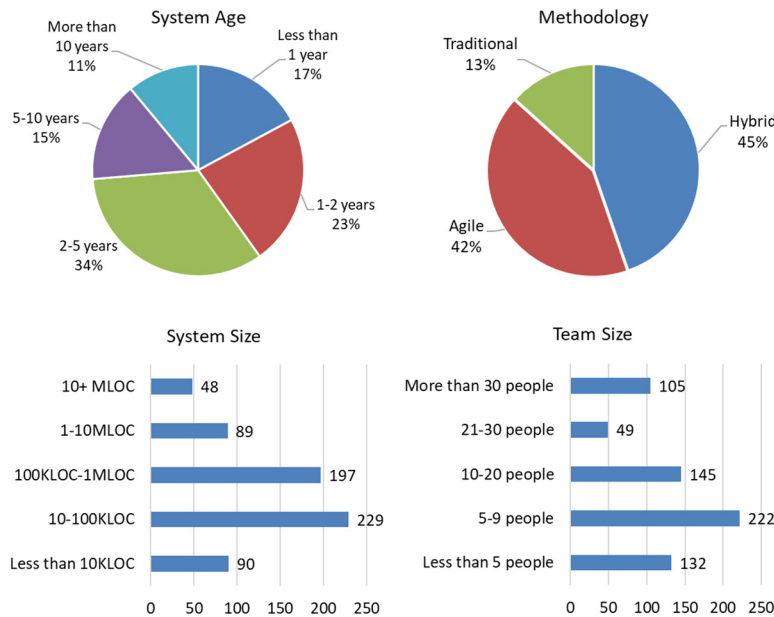
Fig. 2. Demographics data for the exemplar project ($N = 653$).

Table 5

Participants' roles and experience levels per role. R1—Developer; R2—Project leader/manager; R3—Software architect; R4—Tester; R5—Requirements analyst; R6—Process analyst; R7—Database administrator/analyst; R8—Business analyst; R9*—Performs multiple functions; R10*—Quality analyst; R11*—Configuration manager; R12*—Infrastructure analyst; R13*—Infrastructure manager; R14*—Quality analyst manager.

Role	N (%)	Junior (%)		Experienced (%)		
		Novice	Beginner	Competent	Proficient	Expert
R1	327 (50.08)	1.83	13.46	37.00	30.28	17.43
R2	108 (16.54)	0.93	8.33	29.63	35.19	25.93
R3	87 (13.32)	1.15	8.05	24.14	40.23	26.44
R4	46 (7.04)	0.00	6.52	34.78	34.78	23.91
R5	24 (3.68)	0.00	20.83	45.83	25.00	8.33
R6	19 (2.91)	0.00	31.58	36.84	15.79	15.79
R7	14 (2.14)	0.00	21.43	28.57	28.57	21.43
R8	13 (1.99)	7.69	15.38	30.77	15.38	30.77
R9*	5 (0.77)	0.00	20.00	40.00	20.00	20.00
R10*	3 (0.46)	0.00	0.00	33.33	33.33	33.33
R11*	2 (0.31)	0.00	0.00	0.00	0.00	100.00
R12*	2 (0.31)	50.00	0.00	50.00	0.00	0.00
R13*	2 (0.31)	0.00	0.00	0.00	50.00	50.00
R14*	1 (0.15)	0.00	0.00	0.00	0.00	100.00

Design debt (21.99%) refers to violations of good design principles and practices. This type of debt can be discovered by analyzing the source code. For example, using code review techniques or static code analysis tools (Li et al., 2015).

Test debt (19.94%) refers to any issues in regard with testing activities (Guo and Seaman, 2011). For example, planned testing that was not conducted or low-test coverage.

Code debt (14.66%) refers to problems found in the source code that negatively affects readability and maintainability of the code. For example, too complex methods or non-existing coding standards. In order to remove code debt, code refactoring is needed (Rios et al., 2018a).

Architecture debt (10.70%) refers to issues and problems with the architecture of the system that affect the architectural requirements (e.g., performance, maintainability, robustness, etc.) (Kruchten et al., 2012). For example, violation of the modularity principle or poor separation of concerns leads to significant problems with further evolution, maintenance and extensions of the system.

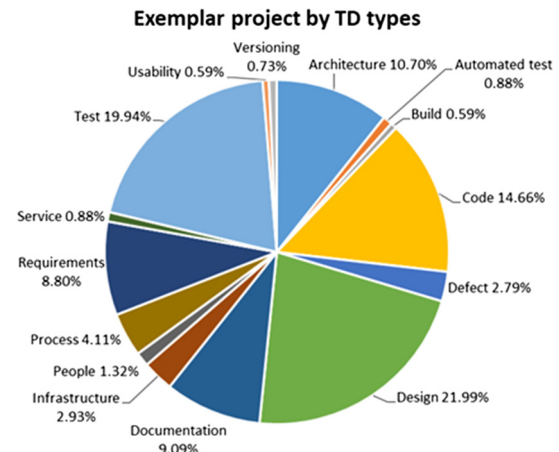


Fig. 3. Exemplar TD types.

Documentation debt (9.09%) refers to any issues and problems discovered in the documentation of the software project, including missing documentation. It is discovered by looking for inconsistent, missing, inadequate or incomplete documentation (Guo and Seaman, 2011). For example, technical documentation is outdated and not reflecting the actual state of the system.

Requirements debt (8.80%) refers to the trade off between agreed requirements and requirements that are actually implemented in a software product. For example, this can be manifested as partially implemented requirements or requirements that are implemented for some cases only (Kruchten et al., 2012).

Process debt (4.11%) refers to inefficient processes. For example, a process changed over time and the existing practices are not efficient any longer (Alves et al., 2016).

Infrastructure debt (2.93%) refers to issues that if present in an organization significantly reduce the teams' abilities to deliver quality products (Alves et al., 2016). For example, using outdated tools or postponing regular software or system updates.

Defect debt (2.79%) refers to known defects that are discovered by testing activities, however the corrective actions are

postponed or never implemented. Decisions to postpone the corrective actions can accumulate a significant amount of TD in the system (Snipes et al., 2012).

People debt (1.32%) refers to any issues related with people. For example, lack of specific skills and knowledge that would require additional training or hiring (Alves et al., 2016). People related issues are much more complex, they involve socio-organizational factors that directly affect productivity and people satisfaction (Rios et al., 2018a; Mellblom et al., 2019).

Finding 1. There is no single TD type that stands out. The top-5 debt types by its occurrence are: design, test, code, architecture, and documentation debts.

Other debt types had less than 1% occurrences each: *service* (0.88%), *automated test* (0.88%), *versioning* (0.73%), *usability* (0.59%), and *build* (0.59%). Precise definitions for these debt types can be found in (Mandić et al., 2020).

4.2. The familiarity with the TD concept (RQ1)

To assess the extent of familiarity with the TD concept, participants were asked to self-assess their own familiarity with TD. Afterwards they were given the McConnell's definition and asked to specify how close is their understanding of TD to the definition. Fig. 4.a shows that 69% of participants were familiar with the TD concept and that 31% of the participants have never heard of it. Out of 69% of the participants that were familiar with TD, 22% had only a theoretical knowledge (they read about the TD concept in books or articles), whereas 47% had more practical experience and were either on projects where they actively managed TD or TD was recognized but not managed.

Fig. 4.b shows that a large percentage of the participants specified that the McConnell's definition was either close or very close to their understanding of TD (78%), while 4% of the participants seem to disagree with the definition. From Fig. 4.b it can also be seen that the percentage of participants that had no prior knowledge of TD decreased to 18%, compared to 31% of respondents that answered that they have never heard of TD before the definition was revealed. This indicates that after presenting the McConnell's definition, the TD concept resonated with some of participant's past experiences that prior to the definition revelation answered that they have never heard of TD before.

Finding 2. The large majority of practitioners understands or agrees with the McConnell's definition of TD (78%). Whereas, only 47% had some practical experiences with TD identification or management. Furthermore, the definition resonated well with practitioners' experiences who had no prior knowledge about the TD concept.

In order to gain better insights on how demographics and TD familiarity interact additional analyses were conducted. For every demographics variable a two step analysis was done. First, a χ^2 test was conducted to see if there is a statistically significant association (dependency) between demographic variables on one side, and TD familiarity and the apprehension of McConnell's definition on the other. This was possible because the variable under study is categorical and the dataset is relatively large. To conduct this test the R language was used with the help of the *stats* package,¹⁴ all in combination with the R Studio development environment. Second, if the test showed the existence of statistically significant associations then mosaic plots with shading were used to visualize data associations.

Table 6 shows that no significant association (p – value > 0.05, red font) was found between system age and TD familiarity, as well as between company size, team size, system age and methodology and the apprehension of McConnell's definition. The following text present the in depth analysis of the significant associations (p -values in black font) between the demographics variables and the familiarity with the TD concept. The associations between the demographic variables and the apprehension of the McConnell's definition were not analyzed regardless of the statistical significance. In depth analysis result of the apprehension of the TD definition may contribute to areas such as the IT education or similar, however it is the authors' opinion that this analysis offers no contribution to answers for RQ1 and it was therefore omitted from the study.

Country versus TD familiarity. Fig. 5.a depicts a mosaic plot of prevalence of the TD concept in industry, with respect to country. Analysis of the figure shows that practitioners that come from Serbia are more likely to have never heard of the TD concept. If practitioners come from the United States or Colombia it is of low probability that they never heard of the TD concept, and regarding United States practitioners they are more likely to have had some practical experience in managing TD. On the other hand, practitioners from Colombia are more likely to have practical experience in TD identification. Brazilian practitioners are more likely to have only theoretical knowledge of the TD concept and less likely to have had some practical experience in managing TD.

Finding 3. Practitioners coming from different countries may have different levels of familiarity with the TD concept. For example, practitioners from USA or Colombia are more likely to be familiar with TD, whereas practitioners from Serbia are not.

Company size versus TD familiarity. Fig. 5.b depicts the prevalence of the TD concept in industry, with respect to company size and shows that practitioners who come from small companies are less likely to have practical experience with TD management.

System size versus TD familiarity. Fig. 6.a depicts a mosaic plot of prevalence of the TD concept in industry, with respect to system size. The analysis of the plot in the figure shows that practitioners that work on systems that have less than 10KLOC are more likely to have only theoretical knowledge of TD and are less likely to have practical experience with managing TD. On the other hand, practitioners that work on systems that have between 1 and 10MLOC are more likely to have experience in TD management.

Team size versus TD familiarity. Looking into teams size versus TD familiarity (Fig. 6.b) it can be observed that practitioners who work in large teams, more than 30 people, are the most likely to have practical experience with managing TD, and are less likely to only have theoretical knowledge of TD. On the other side, practitioners that work in small teams, less than 5 people, are more likely to have never heard of the TD concept and less likely to have practical experience in managing TD.

Experience versus TD familiarity. Fig. 7.a depicts the prevalence of the TD concept in industry, with respect to practitioner experience. Practitioners that are beginners are more likely to have never heard of the TD concept and also less likely to have practical experience in TD management. Practitioners that classify as competent are more likely to have never heard of the TD concept, and also equally likely to have theoretical knowledge on TD. Also, like beginners, they are less likely to have practical experience in managing TD. Expert practitioners are less likely to have never heard of TD and most likely to have practical experience in TD management.

¹⁴ <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/chisq.test>.

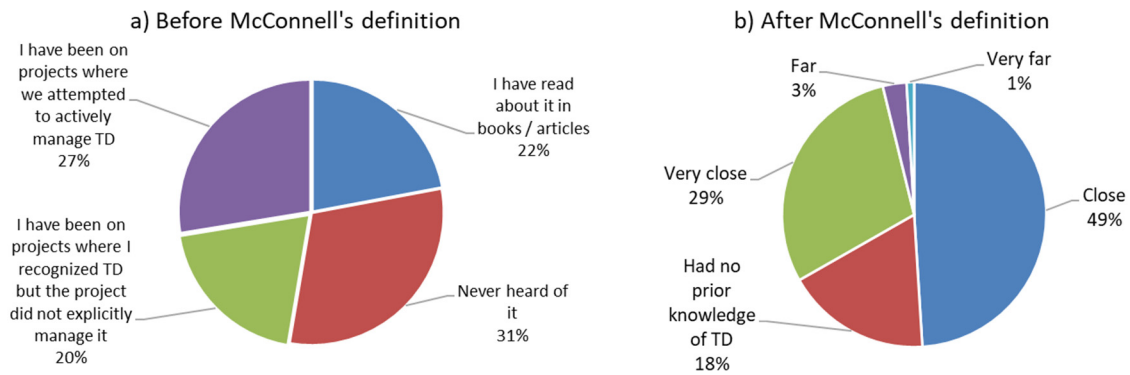


Fig. 4. Participants' familiarity with the TD concept: (a) before TD definition, (b) after TD definition ($N = 653$).

Table 6

Results of a χ^2 test between demographics data and TD familiarity (Q9) and TD definition compliance (Q11). χ^2 —chi square test results; df—degrees of freedom; p-value—threshold $\alpha = 0.05$.

Demographic variable	Q9: Familiarity			Q11: Apprehension		
	χ^2	df	p-value	χ^2	df	p-value
Country	87.912	15	2.428e−12	35.658	20	0.01687
Company size	13.865	6	0.03118	6.6222	8	0.5779
System size	37.586	12	0.0001794	28.379	16	0.02848
Team size	47.547	12	3.746e−06	20.757	16	0.1881
System age	15.192	12	0.2311	17.802	16	0.3356
Role	96.746	39	8.229e−07	85.1	52	0.002567
Experience	70.369	12	2.732e−10	71.783	16	4.851e−09
Methodology	39.688	6	5.246e−07	11.791	8	0.1608

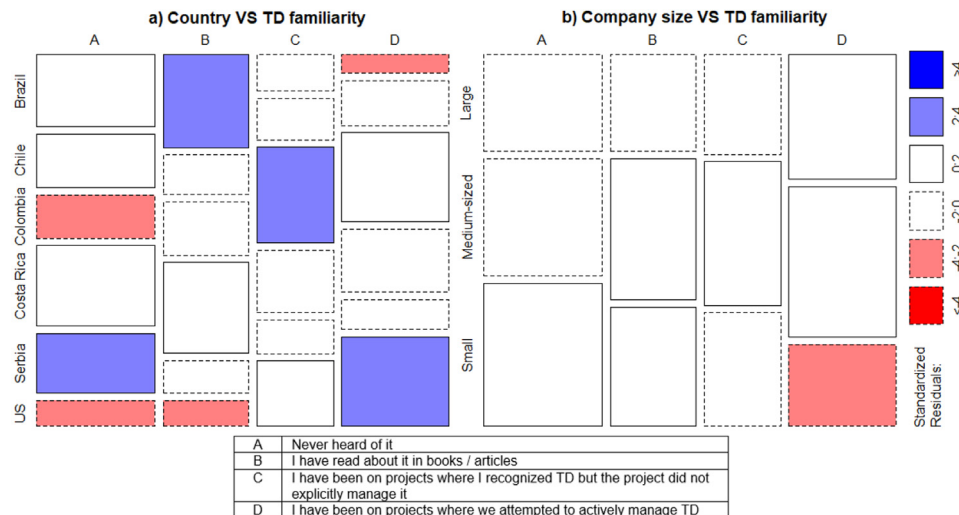


Fig. 5. The mosaic plot for country and company size versus TD familiarity (Q9).

Finding 4. As the system size or company size or team membership count or level of expertise increases, the likelihood that IT practitioners have experience with TD management increases as well.

Role versus TD familiarity. Fig. 8 depicts a mosaic plot of prevalence of the TD concept in industry, with respect to practitioner roles. For this analysis practitioner roles were categorized in order to generate a better illustration of the association between roles and TD familiarity. The categories are presented in Table 7. After categorizing participant roles the χ^2 test was run again and resulted in the following: $\chi^2 = 73.845$, $df = 18$, $p - value = 9.997e - 09$. As the results prior to role categorization these results indicated that there indeed is a statistically significant association in regards to TD familiarity.

Table 7

Categories of participant roles.

Category	Roles assigned to category
DEV	Developer
MNGT	Project leader/project manager, Quality analyst manager
ARC	Software architect
QA	Tester, Quality analyst
ANALYST	Process analyst, Requirement analyst, Database administrator/analyst, Business analyst
ALL	Performs multiple functions
SYS-CM	Configuration manager, Infrastructure analyst, Infrastructure manager

Analysis of Fig. 8 shows that if practitioners fall into the Analyst category they are more likely to have only theoretical

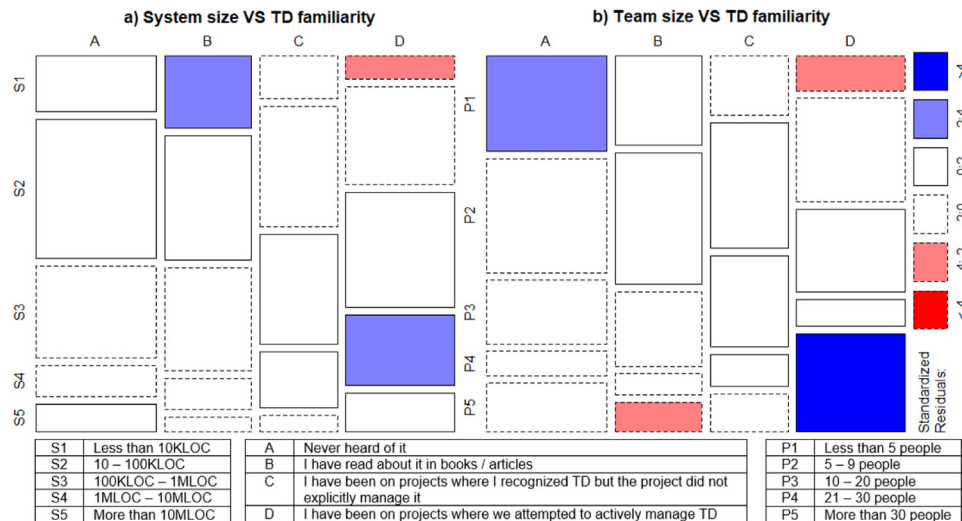


Fig. 6. The mosaic plot for system and team size versus TD familiarity (Q9).

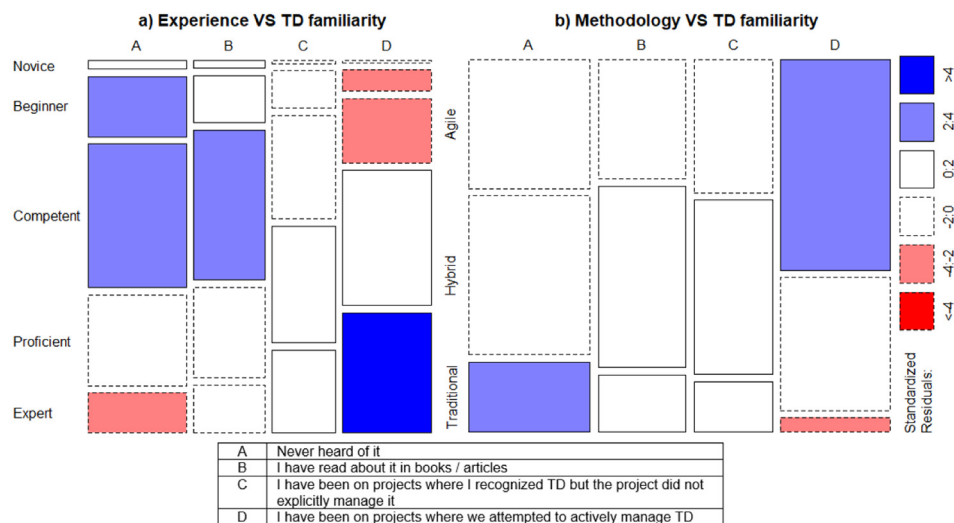


Fig. 7. The mosaic plot for methodology and practitioner experience versus TD familiarity (Q9).

knowledge of TD but are less likely to have practical experience in TD identification or management. Practitioners that are software architects are less likely to have never heard of the TD concept and more likely to actually have some practical experience in TD management. Practitioners that are developers seem to be equally likely to have never heard of the TD concept, have only theoretical knowledge about TD or have some practical experience in TD identification and management. The same can be said for practitioners that fall into the *MGMT* and *QA* categories. The *ALL* and *SYS-CM* seem to lean this way also, but the sample is insufficiently large to make any assumptions.

Finding 5. Practitioners working as software architects are more likely to have practical experience in TD management, whereas analysts, despite their theoretical knowledge about TD, are more likely to have never participated in a project where TD was recognized or managed.

Methodology versus TD familiarity. Fig. 7.b depicts the prevalence of the TD concept in industry, with respect to the used development approach. The data shows that traditional development

approaches are more likely to have practitioners working with them that have never heard of the TD concept and less likely to have practical experience in TD management. On the other hand, practitioners using agile approaches are more likely to have practical experience in TD management.

Finding 6. Practitioners using non-agile like approaches, i.e. traditional approaches, are more likely to have never heard of the TD concept and less likely to have any experiences with TD management.

4.3. Technical debt causes and effects

The participants gave their insight into the causes and effects of TD, in the context of the exemplar project that they specified. The analysis of causes and effects of TD consisted of the: (a) conceptualization of TD causes, (b) identification of the most common causes and effects of TD, and (c) identification of associations between the TD types (Section 4.3.3) on one side, and causes and effects on the other. Most common causes and effects were identified using coding and categorization and the

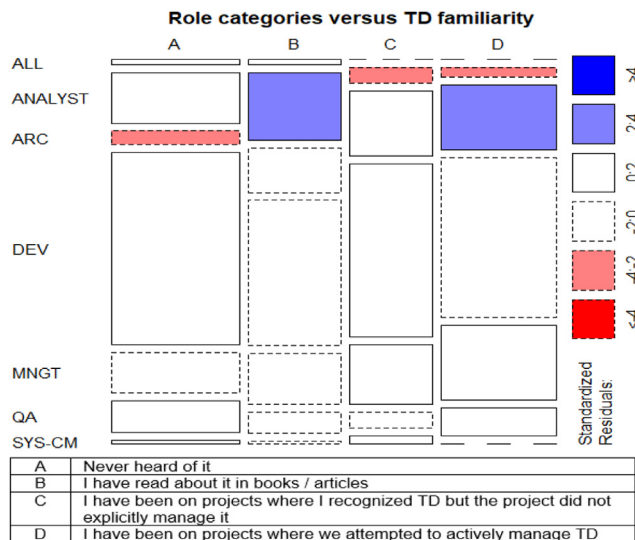


Fig. 8. The mosaic plot for role categories versus TD familiarity (Q9).

results were counted and charted. The associations between the TD types, and causes and effects were identified using the χ^2 test and contingency tables, while for the presentation of these associations, the mosaic plots were used. The analysis results contribute to answers to RQ2 and RQ3.

4.3.1. Conceptualization of TD causes

The TD causes were extracted using the participants answers to survey questions Q16–Q19. Further analysis of these causes led to the derivation of eight categories. These categories, ordered by occurrence frequency and with exemplary causes are:

1. *Planning and management* (725): Multitasking different clients and projects, misconduct, poor leadership.
2. *Development issues* (473): Lack of quality, legacy system, legacy artifacts.
3. *Methodology* (389): Tests not performed, poor task distribution.
4. *Lack of knowledge* (311): Lack of experience, information.
5. *People* (208): Overconfidence, Non-sharing of knowledge, lack of team communication.
6. *Organizational* (133): Lack of training, lack of priority for the project.
7. *External factors* (119): Payment dynamics, discontinued component.
8. *Infrastructure* (42): Inadequate choice of technology, tool, platform, required infrastructure unavailable.

The identified causes, derived categories, and their relationship, as well as their frequencies, represents the conceptualization of TD causes. This conceptualization is presented in a form of mind map on Fig. 9. The mind map reveals that the two dominant categories or group of causes are the *Planning and management* and *Development issues*, and the causes that most frequently occurred in these categories.

4.3.2. The most common causes of TD

The most dominant cause of TD by far is the *Deadline* cause, indicating that developers mostly incur debt because of tight deadlines that are set, and using quick solutions that are not thought through thoroughly. *Inappropriate planning* and *not effective project management* also rank high and indicate a problem with leading the project. These causes are somewhat tied

together on the count of them both referencing project management, one from the planning aspect and the other from a more operational aspect of leading the project in real time. Other significant causes are: *Focus on producing more at the expense of quality*, *Lack of experience*, *Pressure*, *Non-adoption of good practices*, *Inaccurate time estimate*, *Lack of qualified professionals*, *Lack of a well-defined process*.

Further analysis of Fig. 10 reveals that the likelihood of some causes leading to TD was assessed by participants as more likely, even though they had a lower citing frequency. *Non-adoption of good practices* seems to be right behind *Deadline* and *Inappropriate planning* regarding the possibility of it leading to TD. Also, *Lack of experience*, *Lack of a well-defined process* and *Pressure* follow, indicating that they must not be ignored when analyzing possible causes of TD.

Finding 7. *Deadlines, non-adoption of good practices, lack of experience and pressure* are the most likely causes why IT practitioners incur TD. These causes are often a manifestation of other, indirect, causes which are most likely inappropriate planning, lack of well-defined processes or non-effective project management.

4.3.3. Categories of causes and TD types

As presented in Section 4.3.1, the causes were categorized into 8 categories. The χ^2 test calculated to assess the relationship between TD categories and types resulted in the following: $\chi^2 = 214.59$, $df = 98$, $p\text{-value} = 1.067e-10$. The $p\text{-value}$ indicated on a statistically significant relationship between categories of TD causes and TD types. Fig. 11 shows the mosaic plot.

The analysis of Fig. 11 shows that if TD causes fall into the *infrastructure* category they are most likely to lead to *architecture debt* and a bit less likely to lead to *build debt* or *usability debt*. Causes that are external in nature are more likely to lead to *infrastructure debt*, or issues with the way the organization itself operates (e.g. using outdated tools, postponing updates). Causes that have to do with the development approach (methodology) are more likely to lead to *test debt*.

Section 4.3.1 revealed that *planning and management* and *development issues* are the two dominant categories of causes. Causes categorized as *planning and management* are less likely to lead to *design debt* but more likely to lead to *test debt*. Causes that are categorized as *development issues* are more likely to lead either to *code debt* or *design debt* but are less likely to lead to testing related issues or *test debt*. *Test debt* is therefore most likely caused by *planning and management* group of causes, while the *design debt* is caused by *development issues* causes.

Finding 8. *Infrastructure* causes are most likely to lead to *architecture debt*, and a bit less likely to lead to *build* or *usability debt*. *Development issue* causes are more likely to lead either to *code* or *design debt* but are less likely to lead to *test debt*. *External factor* causes are more likely to lead to *infrastructure debt*. *Planning and management*, and *methodology* causes are more likely to lead to *test debt*.

4.3.4. Conceptualization of TD effects

The TD causes were extracted using the participants answers to survey question Q20 and Q21. The analysis of Q21 gave insight into effects TD have on development process and these effects were categorized into six categories. These categories, their frequencies and exemplary effects are:

1. *Planning and management* (467): Increased costs, increased efforts.
2. *External quality* (338): Low quality, low performance, low maintainability.

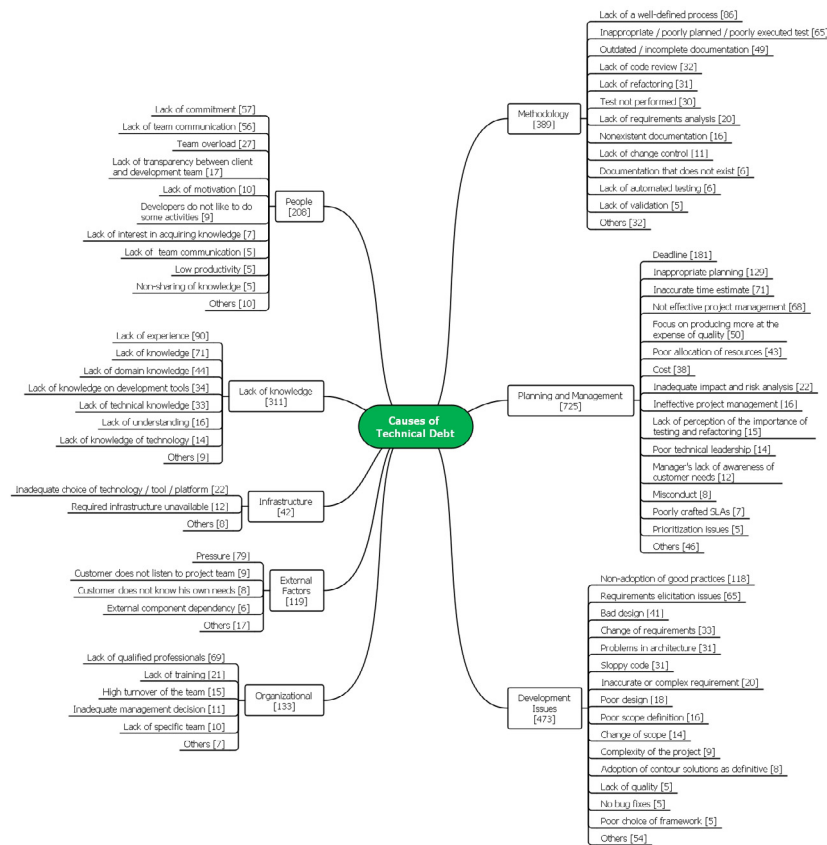


Fig. 9. Mind map of TD causes. Causes with the frequency below 5 are omitted and grouped together as the *Other* item for each category in the map. High-resolution map with all causes is available at <https://bit.ly/3git0Bc>.

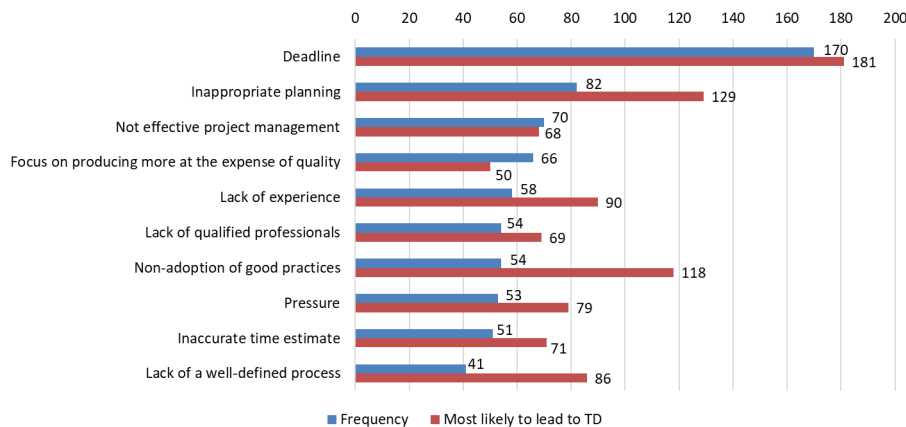


Fig. 10. Top 10 most cited causes of TD.

3. *Internal quality* (308): Poor code readability, increased amount of maintenance activities.
4. *People* (395): Customer or user dissatisfaction, Lack of knowledge.
5. *Development issues* (282): Low code reuse, low tractability.
6. *Organizational* (162): Financial loss, Impaired company image.

As with TD causes, effects and their categories are arranged in mind map, and this map is presented in Fig. 12. Mind map reveal that Planning and management and External quality issues are the most dominant group of TD effects.

4.3.5. The most common effects of TD

The most dominant effect of TD is *Delivery delay*, as presented in Fig. 13. This effect is also the one with the biggest impact. *Low maintainability* and *Rework* are also highly cited effects of TD, as well as the effects with a significant impact. Other significant effects are: *Low external quality*, *Increased effort*, *Need of refactoring*, *Increased cost*, *Bad code*, *Low performance* and *Financial loss*.

One other significant insight that can be obtained from Fig. 13 is that *Financial loss* is cited last in the top 10 effects list, but is categorized as the effect with the third highest impact. A conjecture can be made that this is mostly viewed by the management

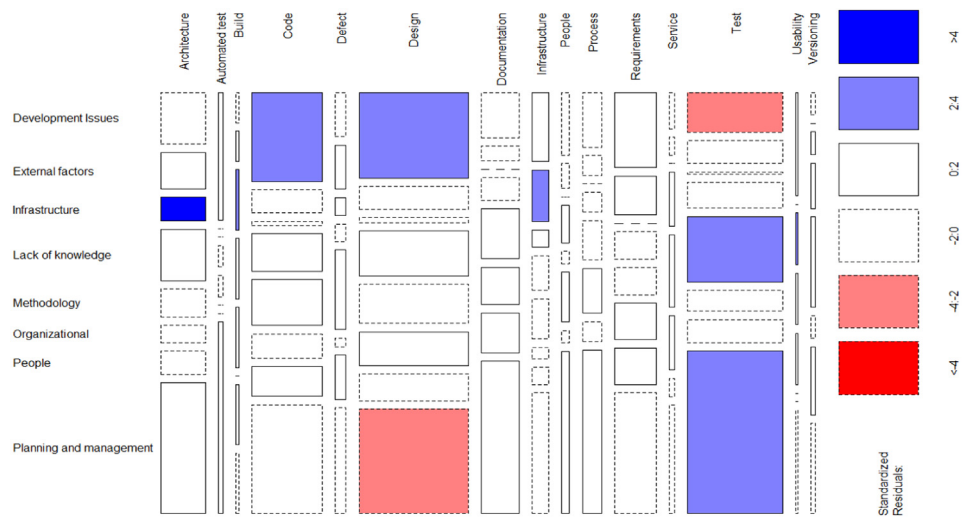


Fig. 11. The mosaic plot for categories of causes versus TD Types.



Fig. 12. Mind map of TD effects. Effects with the frequency below 5 are omitted and grouped together as the *Other* item for each category in the map. High-resolution map with all effects is available at <https://bit.ly/3gtjVVo>.

on the count of this effect being directly related to project, and down the line corporate, financial state.

Finding 9. The presence of TD most likely leads to delivery delays, low maintainability and generates the need for rework. These effect can also lead to longer-term effect, most likely financial loss.

4.3.6. TD types and categories of effects

As it was the case with TD causes, the effects of TD were categorized in six categories. The χ^2 test that assesses the relationship

between TD effects and TD types, resulted in the following: $\chi^2 = 131.14$, $df = 70$, $p\text{-value} = 1.336e-05$. The $p\text{-value}$ indicated that there indeed exists a statistically significant relationship between TD types and categories of TD effects. Fig. 14 shows the mosaic plot that was generated to visualize this relationship and to further analyze it.

Fig. 14 shows that *code debt* most likely leads to issues that can be categorized as *internal quality issues*. *Defect debt* represents defects that are discovered by testing activities but not fixed or corrected. This type of debt is most likely linked to effects that have impact on the people (e.g. developers being demotivated

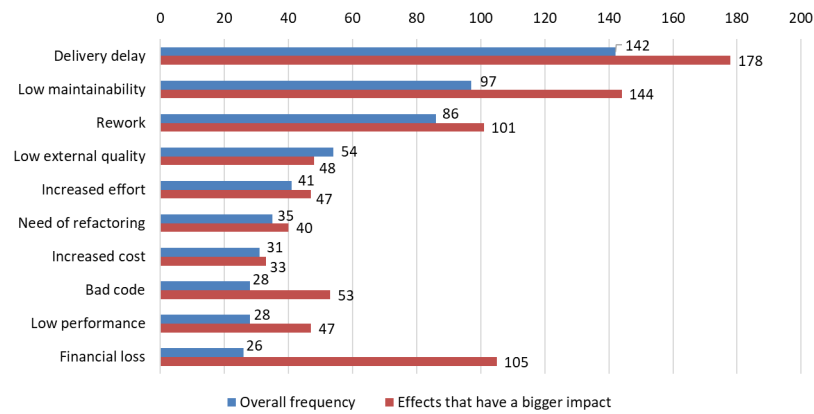


Fig. 13. Top 10 most cited effects of TD.

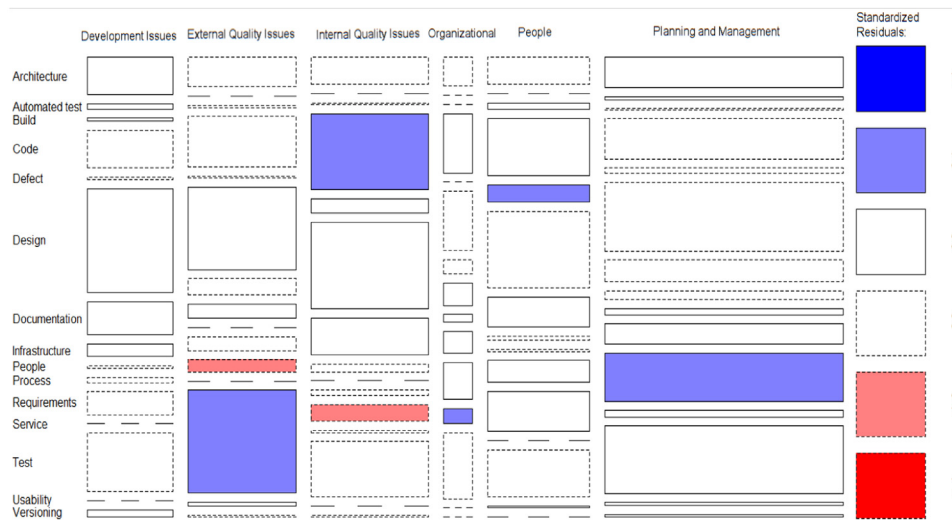


Fig. 14. The mosaic plot for TD Types versus categories of TD effects.

by some product defects). Issues that arise when the selection of web services is not aligned with requirements of the system, or *service debt* is more likely to be related to effects that can be categorized as *organizational*. Section 4.3.4 revealed that the two most dominant groups of TD effects are the *Planning and Management* and *External Quality Issues*. Debt that refers to issues that arise from the trade off between agreed upon requirements and the ones that are actually implemented, or *requirements debt*, is more likely to be linked to effects that resonate with the management level and fall into the *planning and management category*, but are less likely to be linked to *external or internal quality issues*. On the other hand, debt that can be categorized as *test debt* is more likely to be related to effects that are external, or fall into the *External quality issues* category. *Requirements debt* will therefore likely lead towards the effects from *planning and management category*, while the *test debt* most likely leads to effects from *external quality issues* category.

Finding 10. The code and test types of TD are likely to lead to issues with internal and external quality, respectively. While, requirements debt is likely to lead to planning and management issues. Furthermore, defect debt is associated with people related issues.

5. Discussion

This section discusses the most significant findings of the study. Section 5.1 discusses the findings relevant to the prevalence of the TD concept, while Section 5.2 and Section 5.3 focus

respectively on the findings that emerged from the analysis of TD causes and effects.

5.1. The prevalence of TD

Looking at the prevalence of TD, the study findings indicate that the industry is generally familiar with the TD concept with 78% of participants confirming that they understand or agree with McConnell's definition of TD (Finding 2). On the other hand, even though the TD familiarity percentage is relatively high, TD identification and management is somewhat less present in the IT industry with more than half of the participants of this study (53%) indicating that they did not work on projects that attempted to identify or monitor TD (Finding 2). These results suggest that the majority of the industry is at least familiar with the concept of TD, while a certain percentage has theoretical knowledge about TD but has no concrete experience with it. Could this be on the count of organizations themselves not willing to track TD, or is TD knowledge lacking on the management level? Either way, the results show that there is a certain percentage of practitioners that are familiar with TD and could possibly help mitigate its impacts but whose knowledge is not being utilized.

A study by Silva et al. (2019) set a similar research question to this studies RQ1, namely they strove to see if there is a consensus on the perception of TD among software practitioners in the Brazilian software organizations. They analyzed responses from 40 participants and showed that 60% of them claimed to be aware

of the TD concept. Generalizing to this study this corresponds well to the majority of participants claiming that they are familiar with the TD concept.

An interesting result is the one expressed in Finding 3 that states that there is a dependency between country and the familiarity with the TD concept. To the best of our knowledge no other study of international character has detected this dependency. The authors' opinion is that two conjectures can be set regarding the explanation of this finding. The first one is that this is primarily because of the diverse education system (regarding IT) that are in place in different countries, some of which cover the topic of TD in more detail, while others cover the topic more loosely or not at all. This has the most significant impact on junior developers and their familiarity and perception of the TD concept. The second conjecture is regarding the maturity of the national IT industry. Some IT industries are less mature than others, maybe they are still young, and it is in these industries that practitioners are less familiar with TD. Fig. 5.a hints at this with its results, showing that if practitioners come from Serbia, that has a relatively young IT industry that is still developing, they are more likely to have never heard of the TD concept. The variations in national data can be also caused by cultural characteristics, a further research is needed to investigate the cultural aspects of TD.

Shifting the focus to architectural TD, the literature shows that there were studies like the study from Martini et al. (2014) and Martini and Bosch (2017) that focused on this particular type of TD. This study did not find that the architecture as an artifact is a top TD cause. Section 4.1.3 shows that *architectural debt* came in fourth with 10.70% share in total TD types detected from participants exemplar projects, although Finding 5 states that software architects are more likely to have practical experience in TD management. Being that software architects are not only entrusted with and responsible for designing the system, but are also responsible for the software as a whole meeting all agreed upon requirements and standards, it is only logical that they are, among some other roles, most interested in TD.

When it comes to the methodology that is being used, practitioners that use traditional approaches are more likely to have never heard of the TD concept and less likely to have some experience in TD management (Finding 6). It can be argued that this is somewhat expected on the count of more traditional approaches being less present in software development nowadays, while Agile approaches, or even hybrid approaches being more dominant.

Rocha et al. (2017), among other things, reported on knowledge on TD from the perspective of 74 Brazilian practitioners. They stated that 25.7% had low knowledge on TD, 36.5% a medium grasp on TD and 25.7% a high understanding of TD. Fig. 5.a shows that this study also points to Brazilian practitioners being more likely to have theoretical knowledge on TD.

Ernst et al. (2015) in their study found that the TD metaphor is widely accepted within large organizations. All participants in their study came from three large companies and the study found that they are generally familiar with the concept of TD. This results coincides well with this studies Finding 4 that states that as the company size increases so does the likelihood that practitioners are familiar with the concept of TD, and also are more likely to have some practical experience in TD identification or management.

This study found a statistically significant relationship between company size and the familiarity of TD, indicating that it is less likely that practitioners that come from small companies have actively managed TD on some projects within that company (Fig. 5.b).

5.2. Top causes of TD

The top most cited cause of TD was the *Deadline* cause, with the citing frequency of 170. Besides deadlines some of the most cited causes of TD that emerged were causes such as *Inappropriate planning*, *Not effective project management* and *Focus on producing more at the expense of quality*. An interesting cause was the *Non-adoption of good practices* cause, that came in the seventh place in the top 10 causes list, but is the third cause that is, according to study participants, most likely to lead to TD, which makes this cause worthy of considering even before some that have a higher cite frequency. Other causes that fall into the top 10 causes of TD list can be viewed in Fig. 10.

When analyzing the likelihood of a cause leading to TD, the *Deadline* cause was cited a quarter of the times (25%) in the *Planning and Management* category, as the most likely cause to lead to TD (Fig. 9). It also took up 26% of citations of all top cited causes in remaining 7 categories. In the same category, other causes emerged, like the *Inappropriate planning* cause, that had 18% of that categories citations, the *Inaccurate time estimate* cause with 10% of citations, the *Not effective project management* cause with 9% of citations, and the *Focus on producing more at the expense of quality* with 7%. These causes take up 69% of citations in the *Planning and Management* category and take up more than half (53%) of citations in the top 10 causes list regarding the likelihood of leading to TD (Fig. 10). This data points to *Planning and Management* as the substantial reservoir of potential TD causes.

The *Deadline* as a cause was recognized by Yli-Huumo et al. (2014) in their study as lack of time given for development. They also recognized pressure put on the development team as a cause of TD which can coincide with this studies *Pressure* cause that was cited 53 times and recognized by practitioners as relatively highly likely to lead to TD.

The study that was conducted by Codabux and Williams (2013) indicated that managerial decisions are among the main causes of TD. These decisions can result in inappropriate planning, not effective project management or even a lack of a well defined process, that was also detected by this study as a highly cited cause. This study confirmed that management decisions, at different management levels (e.g. project management, organization management), can in deed be a cause of TD. The study in Codabux and Williams (2013) also recognized resource constraints as a cause of TD. This was also confirmed by this study in the sense that *Lack of qualified professionals* was detected as a highly cited cause of TD with a frequency of 54.

One of the largest studies on TD that was conducted by Ernst et al. (2015) came to the conclusion that immature choices done in architecture are a key cause of TD. The *Inappropriate planning* cause identified by this study can be a result of poorly made decisions in architecture that lead to TD.

The study conducted by Martini et al. (2014) focused on architecture debt and strove to find the causes for the accumulation of architectural TD and also to investigate the trends in the accumulation and recovery of architectural TD. The study found that business factors, design and architecture documentation, reuse of legacy/third party/open source, parallel development, effects uncertainty, non-completed refactoring, technology evolution, and the human factor can be the likely causes. This study builds on this and finds that *Infrastructure* causes, or inadequate technology, tool, platform choices, or an unavailable infrastructure, are most likely to lead to architecture debt. Besides this, the data shows that code or design debt will most likely be triggered by some *Development issues* cause, that *External factor* causes most likely lead to infrastructure debt and *Planning and management* and *Methodology* causes most likely lead to test debt.

The identification of several causes of TD that were mentioned in earlier studies by other authors can possibly indicate that a

trend is slowly emerging for some causes that are most likely to trigger the existence of TD items, but more research is needed to derive some conclusions that are generalizable.

Ernst et al. (2015) in their study, showed that architecture decisions were the main source of TD, while this study adds the finding that several other causes that can be classified as *Infrastructure* causes can lead to Architecture debt (Fig. 11). They also found *Overly complex code* to be a significant cause. The debt that occurs in this context can be viewed as *Code* and this study found that causes categorized as *Development Issues* can lead to this particular debt type (Fig. 11). Architecture as a potential leading cause of TD is also the found by Holvitie et al. (2018). They also found that inadequate tests seem to be a cause of TD, whereas this study showed that causes categorized as *Planning and Management* and *Methodology* can lead to Test debt (Fig. 11).

5.3. Top effects of TD

The top most cited effect of TD was the *Delivery delay* effect, cited 142 times, meaning that in most cases practitioners saw that because of debt items they were not able to reach an agreed upon deadline for delivery. This is an interesting find in comparison to the top TD cause being the *deadlines* that are defined. A vicious circle seems to form here on the count of practitioners rushing to meet the deadlines and introducing TD in the project that in turn later delays the delivery of the project. Besides this, some of the other highly cited effects that practitioners named were effects like *Low maintainability*, *Rework* and *Low external quality*. An interesting effect in the top 10 list was the *Financial loss* effect, that came in 10th regarding citing frequency, but was the third effect regarding the significance of its impact. This is somewhat expected because this effect speaks about company finances, and TD leading to financial loss can have serious repercussions on the company as a whole.

When analyzing what effects have a bigger impact it seems that *Delivery delay*, *Low maintainability*, *Financial loss*, *Rework* are the ones with the biggest impact (Fig. 13). The *Delivery delay* cause comes from the *Planning and Management* category and takes up 38% of effects cited in that category. On the other hand, *Low maintainability* is placed in the *Internal Quality Issues* category and also takes up 47% of effects in that category. The *Financial loss* effect is probably felt the most by the management of the company and is categorized in the *Organizational* category and it is taking up the majority of that category with 65% of all cited effects. The *Rework* effect comes also from the *Planning and Management* category with a share of 22%. These effects combined take up 66% of effects that have the biggest impact, and should be taken into consideration first when analyzing TD.

In the Codabux and Williams (2013) study the results showed that starting again the project from scratch was the main effect felt by practitioners, while this studies participants listed *rework* as the third most common effect of TD. Yli-Huuma et al. (2014) extra working hours and errors and bugs were among the main effects of TD. This study confirms this by identifying *increased effort* and *bad code* as effects that are among the most cited by study participants. This studies findings for top effects also complement the study by Yli-Huuma et al. (2015) that found, among others, decreased code maintainability, major refactoring as the emerging effects of TD. Similarly the study by Li et al. (2015) also found that maintainability is one quality attribute that is compromised because of the presence of debt items.

This studies findings confirm some of the earlier study finding regarding the effects of TD, but also build upon the work by introducing some new effects that are recognized by practitioners as relevant.

Besker et al. (2018a) argued that additional time that is wasted on a project is spent in performing additional tests, and thus

makes an important effect of TD to note. This study showed that all effects that can be categorized as *External Quality Issues* are likely to lead to Test debt (Fig. 14). Besker et al. (2018a) also mention additional source code analysis and refactoring as most common effects of TD. These effects can be categorized as *Development Issues* for which this study did not find any statistically significant relationship with any TD type (Fig. 14). Yli-Huuma et al. (2015) mentioned *decreased code maintainability* as one of the most dominant consequences for quick workarounds, while this study categorizes low maintainability as an *Internal Quality Issue* and this TD effect category is likely to lead to Code debt.

5.4. InsignTD dataset

In this publication we reported the InsignTD dataset with respect to prevalence, causes, and effects data. The given analysis aids better description of the dataset, i.e. it supports more exploratory analysis, rather than a specific data analysis for preset hypotheses. However, the intention is to allow formulations of such hypotheses in future studies based on this report.

Our intention is to present a report of InsignTD dataset that will allow for versatile usage. For example, from data on causes it can be seen that causes characterized as *planning and management* are far more common than other causes (Fig. 9 in Section 4.3). Furthermore, data also suggest that in presence of planning and management causes testing practices will suffer the most, i.e. it will result with the injection of a *test debt* (Fig. 11). So, reading segments of this report provides different insights that can motivate further investigations.

As one of the objectives of InsignTD is to increase the dataset by adding new replication countries, the dataset will change, i.e. grow, over time. An interested reader can find an updated version of the InsignTD analysis report on the following link¹⁵ where all additional analysis of InsignTD data are included.

6. Trustworthiness of the study

The assessment of the trustworthiness of the study depends on its underlying paradigmatic or philosophical stances. This study adopted the constructivism as a paradigmatic stance (Creswell, 2009). Constructivism's assumptions were seen as most acceptable considering the study, its objective and the stated RQs. To argument this decision, the following four constructivists' assumptions, proposed by Lincoln et al. (2011) will be presented alongside the properties of this study and the InsignTD project.

The arguments are as follows: (1) For constructivism, the inquiry goal is the understanding of the phenomenon. The goal of InsignTD is to be achieved by understanding the practitioners' familiarity with TD and their reactions when TD is present, the causes of TD and effects it has on software development process. (2) For constructivists, knowledge is inseparable from humans and their context. In InsignTD, knowledge is generated using IT industry practitioners' experiences with TD. (3) Constructivists focus on distinct groups of people and seek to find their local truths. InsignTD builds on findings from national, that is on local, findings and aggregates them to derive the new knowledge. (4) Constructivists typically rely on qualitative methods for data collection and analysis. InsignTD survey includes open questions to collect rich qualitative data and this data is dominantly used for answering RQs. Relying dominantly on qualitative methods also makes this study a *qualitative study or qualitative in nature*.

To ensure the trustworthiness of the qualitative studies, Guba (1981) proposes credibility, transferability, dependability, confirmability as criteria to be considered. Various techniques can be used to promote these criteria in qualitative studies. This study relied on techniques proposed by Shenton (2004).

¹⁵ InsignTD supplementary analysis report. Version 1.0. (Date: 10.06.2021) <https://bit.ly/3xeU4Hu>.

6.1. Credibility of the study

Credibility denotes the alignment of study findings with the reality. To ensure the credibility, following techniques were used: *The adaption of the well-established research method* this study used a survey to collect the data, qualitative coding and a clearly defined and analysis protocol developed by the InsignTD core group. Research methods that were used are also well accepted in software engineering field (Singer et al., 2008; Seaman, 2008).

Sampling of subjects was done using the convenience sampling method, as presented in Section 3.4. *The honesty of information* was ensured by allowing potential participants to refuse to participate in the study, to quit the survey if they felt like doing so and by guaranteeing the anonymity of the participants. *Frequent debriefing sessions*: during the analysis of data, researchers closely collaborated and all decisions and disagreements were solved with a consensus. *Experience of the investigator*: The researchers who organized the InsignTD project and designed the survey were all experienced researchers. Junior researchers who participate in this project were under close supervision of their senior colleagues. *Thorough description of the phenomenon under study*: this study focuses on TD and its manifestation in software development. TD is well recognized in software engineering literature, as it is presented in Section 2.

The credibility of the study is also supported with the study organization. According to Santos et al. (2019), organizing the study as a family of survey replications promotes the internal validity, which is the criteria used for credibility in quantitative studies (Guba, 1981; Shenton, 2004). Main reasons for this are the shared goals, RQs, study design, collaboration of independent research teams, shared raw data, and similar. Finally, the InsignTD survey was designed to ensure the credibility of the collected data. For example, Q13 allowed researchers to discard the answers based on whether the provided TD example could or could not be mapped to one of the TD types (see Section 4.1.1).

6.2. Transferability of the study findings

Transferability denotes the application of study findings in different contexts or the generalizability of findings. To promote the transferability of the study findings, the technique of *thorough description of the context* in which the study was executed was used. The context of this study is thoroughly described with a significant amount of details data in Section 4.1 - Demographics and Section 3.2 - InsignTD project.

6.3. Dependability on the research context

Dependability denotes that the study repeated in the same context, with same participants, and same method would result with similar outcomes. To ensure the dependability of the study findings, a technique of *Thorough description of the application of research methods* for data collection and analysis was used. Methodological descriptions provided in Section 3, provide sufficient amount of data for interested researchers to replicate the study.

6.4. Confirmability of the study findings

Confirmability denotes the objectivity of the findings, that is that the findings are based on the analysis of the collected data rather than on the researchers' beliefs. To ensure the confirmability of the study findings, the following techniques were used: *Thorough methodological description*. Research approach adopted in this study is presented in Section 3. This description enables

the scrutiny of the study findings. *Recognition of the study shortcomings*. Survey data can be incorrect due to: (a) lose of motivation to complete the survey or (b) when the participation in the study changes the behavior of the respondent. To mitigate these shortcomings, participants were familiarized with the study goal, topic, duration and that the participation in survey was anonymous.

7. Conclusion

The phenomenon of technical debt permeates the IT industry, and it is present in organizations and projects of various sizes. The large majority of developers is dealing with the consequences of injected TD at daily bases. It is evident that TD is a multifaceted phenomenon, and because of that, it is of utmost importance to collect rich empirical data about it. Current empirical investigations although important and valuable on its own have limitations regarding sample composition and size. Therefore, the objective of the InsignTD initiative is to coordinate research teams from different countries in order to jointly contribute to a single dataset about the TD phenomenon. So far, the InsignTD dataset includes 653 responses from 6 countries (Brazil, Chile, Columbia, Costa Rica, USA, and Serbia).

Regarding the prevalence and familiarity of the TD concept, it seems that the concept is unknown to the almost third of participants (31%). This percentage is relatively high, when compared to participants who had some practical experiences with TD identification and/or management (47%). Furthermore, the analysis revealed that participants' country has statistically significant impact on the TD concept prevalence. This strongly suggests that practitioners in some countries are more likely to be familiar and experienced with the concept, such as in Columbia and in the United States.

Also, findings strongly suggest that a major cause of TD is *deadline*, i.e. time pressure caused by short deadlines. And, interestingly the single most cited effect of TD is *delivery delay*. The second most cited effect is *low maintainability*. Based on this finding we can speculate that a *TD vicious cycle*, i.e. a situation when existing debt incurs additional debt, can be caused and then fed by continuous time pressure. Further data analysis showed that causes categorized as *infrastructure* causes (environment setup, tools, and used technologies) will certainly lead to *architectural debt*.

The analysis of effects did not reveal a single strong relation that stood out, like for causes, but rather several relations that describe expected effects of specific TD types. Some of them are well known, such as *code debt* that leads to *internal quality* issues, *test debt* that leads to *external quality* issues, and *requirements debt* that leads to *planning and management* issues. Alongside with those, *defect debt* leads to *people* related issues. It seems that postponing or avoiding corrections of known defects translates into growing dissatisfaction among developers as well as among users.

Families of surveys are rare in software engineering. InsignTD is the first family of surveys on technical debt in software engineering. As a research approach, family of surveys turned out to be very useful. It provided a methodological framework that allowed multiple replication teams to conduct research activities with a sufficient level of autonomy, but at the same time providing common grounds for aggregating data from multiple replications of the survey.

Regarding future work, further analysis of the data is planned in order to gain insight on undertaken practices to cope with TD. Furthermore, we expect to extend the InsignTD dataset with new replications. So far, we collected data from six out of twelve countries that joined the initiative.

References

- Alves, N.S., Mendes, T.S., de Mendonça, M.G., Spínola, R.O., Shull, F., Seaman, C., 2016. Identification and management of technical debt: A systematic mapping study. *Inf. Softw. Technol.* 70, 100–121.
- Apa, C., Solari, M., Vallespir, D., Travassos, G.H., 2020. A taste of the software industry perception of technical debt and its management in Uruguay: A survey in software industry. In: *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM*, pp. 1–9.
- Arvanitou, E.-M., Ampatzoglou, A., Bibi, S., Chatzigeorgiou, A., Stamelos, I., 2019. Monitoring technical debt in an industrial setting. In: *Proceedings of the Evaluation and Assessment on Software Engineering*, pp. 123–132.
- Avgeriou, P., Kruchten, P., Ozkaya, I., Seaman, C., 2016. Managing technical debt in software engineering (dagstuhl seminar 16162). In: Avgeriou, P., Kruchten, P., Ozkaya, I., Seaman, C. (Eds.), *Dagstuhl Rep.* 6 (4), 110–138.
- Babbie, R., Seacat, E., 1973. *Survey Research Methods*. Wadsworth Publishing Company, URL: https://books.google.rs/books?id=_12Aybq4WE0C.
- Baltes, S., Ralph, P., 2020. Sampling in software engineering research: A critical review and guidelines. *arXiv:2002.07764*.
- Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., et al., 2001. *Manifesto for agile software development*.
- Besker, T., Martini, A., Bosch, J., 2017. The pricey bill of technical debt: When and by whom will it be paid? In: *2017 IEEE International Conference on Software Maintenance and Evolution. ICSME. IEEE*, pp. 13–23. <http://dx.doi.org/10.1109/ICSME.2017.42>.
- Besker, T., Martini, A., Bosch, J., 2018a. Technical debt cripples software developer productivity: a longitudinal study on developers' daily software development work. In: *Proceedings of the 2018 International Conference on Technical Debt*, pp. 105–114.
- Besker, T., Martini, A., Bosch, J., 2019. Software developer productivity loss due to technical debt—A replication and extension study examining developers' development work. *J. Syst. Softw.* 156, 41–61.
- Besker, T., Martini, A., Edirisooriya Lokuge, R., Blincoe, K., Bosch, J., 2018b. Embracing technical debt, from a startup company perspective. In: *2018 IEEE International Conference on Software Maintenance and Evolution. ICSME. IEEE*, pp. 415–425. <http://dx.doi.org/10.1109/ICSME.2018.00051>.
- Boehm, B., 2003. Value-based software engineering: reinventing. *ACM SIGSOFT Softw. Eng. Notes* 28 (2), 3.
- Brown, P., Preece, M.A., 1992. Friendly fire in medicine: Hormones, homografts, and Creutzfeldt-Jakob disease. *Lancet* 340 (8810), 24–27.
- Buschmann, F., 2011. To pay or not to pay technical debt. *IEEE Softw.* 28 (6), 29–31.
- Codabux, Z., Williams, B., 2013. Managing technical debt: An industrial case study. In: *2013 4th International Workshop on Managing Technical Debt. MTD. IEEE*, pp. 8–15.
- Codabux, Z., Williams, B.J., Bradshaw, G.L., Cantor, M., 2017. An empirical assessment of technical debt practices in industry. *J. Softw.: Evol. Process* 29 (10), e1894.
- Creswell, J., 2009. *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, p. 296.
- Ernst, N.A., Bellomo, S., Ozkaya, I., Nord, R.L., Gorton, I., 2015. Measure it? Manage it? Ignore it? Software practitioners and technical debt. In: *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. ESEC/FSE 2015, Association for Computing Machinery, New York, NY, USA*, pp. 50–60. <http://dx.doi.org/10.1145/2786805.2786848>, <https://doi.org/10.1145/2786805.2786848>.
- Ghazi, A.N., Petersen, K., Reddy, S.S.V.R., Nekkanti, H., 2019. Survey research in software engineering: Problems and mitigation strategies. *IEEE Access* 7, 24703–24718.
- Guba, E.G., 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. *Educ. Commun. Technol.* 29 (2), 75–91. <http://dx.doi.org/10.1007/BF02766777>.
- Guo, Y., Seaman, C., 2011. A portfolio approach to technical debt management. In: *Proceedings of the 2nd Workshop on Managing Technical Debt. MTD '11, Association for Computing Machinery, New York, NY, USA*, pp. 31–34. <http://dx.doi.org/10.1145/1985362.1985370>, <https://doi.org/10.1145/1985362.1985370>.
- Holvitie, J., Leppänen, V., Hyrynsalmi, S., 2014. Technical debt and the effect of agile software development practices on it - an industry practitioner survey. In: *2014 Sixth International Workshop on Managing Technical Debt. IEEE*, pp. 35–42. <http://dx.doi.org/10.1109/MTD.2014.8>.
- Holvitie, J., Licorish, S.A., Spínola, R.O., Hyrynsalmi, S., MacDonell, S.G., Mendes, T.S., Buchan, J., Leppänen, V., 2018. Technical debt and agile software development practices and processes: An industry practitioner survey. *Inf. Softw. Technol.* 96, 141–160.
- Kitchenham, B., Pfleeger, S.L., 2003. Principles of survey research (part 1-6). *SIGSOFT Softw. Eng. Notes* 28 (2), 24–27.
- Kruchten, P., Nord, R.L., Ozkaya, I., 2012. Technical debt: From metaphor to theory and practice. *IEEE Softw.* 29 (6), 18–21.
- Lethbridge, T.C., Sim, S.E., Singer, J., 2005. Studying software engineers: Data collection techniques for software field studies. *Empir. Softw. Eng.* 10 (3), 311–341.
- Li, Z., Avgeriou, P., Liang, P., 2015. A systematic mapping study on technical debt and its management. *J. Syst. Softw.* 101, 193–220.
- Lincoln, Y.S., Lynham, S.A., Guba, E.G., 2011. *Paradigmatic controversies, contradictions, and emerging confluences, revisited*. Sage Handb. Qual. Res. 4, 97–128.
- Mandić, V., Oivo, M., Rodríguez, P., Kuvaja, P., Kaikkonen, H., Turhan, B., 2010. What is flowing in lean software development? In: Abrahamsson, P., Oza, N. (Eds.), *Lean Enterprise Software and Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 72–84.
- Mandić, V., Taušan, N., Ramac, R., 2020. The prevalence of the technical debt concept in Serbian IT industry: Results of a national-wide survey. In: *International Conference on Technical Debt (TechDebt '20)*, Seoul, Republic of Korea. ACM, p. 10. <http://dx.doi.org/10.1145/3387906.3388622>.
- Martini, A., Besker, T., Bosch, J., 2018. Technical debt tracking: Current state of practice: A survey and multiple case study in 15 large organizations. *Sci. Comput. Program.* 163, 42–61.
- Martini, A., Bosch, J., 2015. Towards prioritizing architecture technical debt: information needs of architects and product owners. In: *2015 41st Euromicro Conference on Software Engineering and Advanced Applications. IEEE*, pp. 422–429.
- Martini, A., Bosch, J., 2017. On the interest of architectural technical debt: Uncovering the contagious debt phenomenon. *J. Softw.: Evol. Process* 29 (10), e1877.
- Martini, A., Bosch, J., Chaudron, M., 2014. Architecture technical debt: Understanding causes and a qualitative model. In: *2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications. IEEE*, pp. 85–92.
- McConnell, S., 2008. *Managing Technical Debt*. Construx Software Builders, Inc, pp. 1–14.
- Mellblom, E., Arason, I., Gren, L., Torkar, R., 2019. The connection between burnout and personality types in software developers. *IEEE Softw.* 36 (05), 57–64.
- Miles, M.B., Huberman, A.M., 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.
- Molléri, J.S., Petersen, K., Mendes, E., 2016. Survey guidelines in software engineering: An annotated review. In: *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM '16, Association for Computing Machinery, New York, NY, USA*, <http://dx.doi.org/10.1145/2961111.2962619>, <https://doi.org/10.1145/2961111.2962619>.
- Morgenthaler, J.D., Gridnev, M., Sauciu, R., Bhansali, S., 2012. Searching for build debt: Experiences managing technical debt at google. In: *2012 Third International Workshop on Managing Technical Debt. MTD. IEEE*, pp. 1–6.
- Pérez, B., Castellanos, C., Correal, D.O., Rios, N., Freire, S., Spínola, R., Seaman, C., 2020. What are the practices used by software practitioners on technical debt payment: Results from an international family of surveys. In: *Proceedings of the 3rd International Conference on Technical Debt. TechDebt '20, Association for Computing Machinery, New York, NY, USA*, pp. 103–112. <http://dx.doi.org/10.1145/3387906.3388632>, <https://doi.org/10.1145/3387906.3388632>.
- Ramač, R., Mandić, V., Taušan, N., Rios, N., de Mendonca Neto, M.G., Seaman, C., Spínola, R., 2020. Common causes and effects of technical debt in Serbian IT: InsightTD survey replication. In: *2020 46th Euromicro Conference on Software Engineering and Advanced Applications. SEAA*, pp. 354–361. <http://dx.doi.org/10.1109/SEAA51224.2020.00065>.
- Rios, N., Mendonça, M.G., Seaman, C., Spínola, R.O., 2019. Causes and effects of the presence of technical debt in agile software projects. In: *Twenty-Fifth Americas Conference on Information Systems, Cancun, 2019. Association for Information Systems (AIS)*, pp. 1–10.
- Rios, N., de Mendonça Neto, M.G., Spínola, R.O., 2018a. A tertiary study on technical debt: Types, management strategies, research trends, and base information for practitioners. *Inf. Softw. Technol.* 102, 117–145.
- Rios, N., Spínola, R.O., Mendonça, M., Seaman, C., 2020. The practitioners' point of view on the concept of technical debt and its causes and consequences: a design for a global family of industrial surveys and its first results from Brazil. *Empir. Softw. Eng.* 1–72.
- Rios, N., Spínola, R.O., Mendonça, M., Seaman, C., 2018b. The most common causes and effects of technical debt: First results from a global family of industrial surveys. In: *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM '18, Association for Computing Machinery, New York, NY, USA*, pp. 1–10. <http://dx.doi.org/10.1145/3239235.3268917>, <https://doi.org/10.1145/3239235.3268917>.
- Rocha, J.C., Zapalowski, V., Nunes, I., 2017. Understanding technical debt at the code level from the perspective of software developers. In: *Proceedings of the 31st Brazilian Symposium on Software Engineering*, pp. 64–73.

- Santos, A., Gómez, O.S., Juristo, N., 2018. Analyzing families of experiments in SE: a systematic mapping study. *IEEE Trans. Softw. Eng.*
- Santos, A., Vegas, S., Oivo, M., Juristo, N., 2019. A procedure and guidelines for analyzing groups of software engineering replications. *IEEE Trans. Softw. Eng.*
- Seaman, C.B., 2008. Qualitative methods. In: *Guide To Advanced Empirical Software Engineering*. Springer London, London, pp. 35–62. http://dx.doi.org/10.1007/978-1-84800-044-5_2, https://doi.org/10.1007/978-1-84800-044-5_2.
- Shenton, A.K., 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Educ. Inf.* 22 (2), 63–75.
- Shull, F.J., Carver, J.C., Vegas, S., Juristo, N., 2008. The role of replications in empirical software engineering. *Empir. Softw. Eng.* 13 (2), 211–218.
- Shull, F., Falessi, D., Seaman, C., Diep, M., Layman, L., 2013. Technical debt: Showing the way for better transfer of empirical results. In: *Perspectives on the Future of Software Engineering*. Springer, pp. 179–190.
- Siebra, C.A., Tonin, G.S., da Silva, F.Q., Oliveira, R.G., Antonio, L., Miranda, R.C., Santos, A.L., 2012. Managing technical debt in practice: An industrial report. In: *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, pp. 247–250.
- Silva, V.M., Junior, H.J., Travassos, G.H., 2019. A taste of the software industry perception of technical debt and its management in Brazil. *J. Softw. Eng. Res. Dev.* 7, 1.
- Singer, J., Sim, S.E., Lethbridge, T.C., 2008. Software engineering data collection for field studies. In: *Guide To Advanced Empirical Software Engineering*. Springer London, London, pp. 9–34. http://dx.doi.org/10.1007/978-1-84800-044-5_1, https://doi.org/10.1007/978-1-84800-044-5_1.
- Snipes, W., Robinson, B., Guo, Y., Seaman, C., 2012. Defining the decision factors for managing defects: a technical debt perspective. In: *2012 Third International Workshop on Managing Technical Debt*. MTD. IEEE, pp. 54–60.
- Taksande, N., 2011. *Empirical Study on Technical Debt As Viewed By Software Practitioners*. University of Maryland, Baltimore County.
- Tom, E., Aurum, A., Vidgen, R., 2013. An exploration of technical debt. *J. Syst. Softw.* 86 (6), 1498–1516.
- Wagner, S., Fernández, D.M., Felderer, M., Vetrò, A., Kalinowski, M., Wieringa, R., Pfahl, D., Conte, T., Christiansson, M.-T., Greer, D., et al., 2019. Status quo in requirements engineering: A theory and a global family of surveys. *ACM Trans. Softw. Eng. Methodol. (TOSEM)* 28 (2), 1–48.
- Wohlin, C., Höst, M., Henningsson, K., 2006. Empirical research methods in web and software engineering. In: *Web Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 409–430. http://dx.doi.org/10.1007/3-540-28218-1_13, https://doi.org/10.1007/3-540-28218-1_13.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A., 2012. *Experimentation in Software Engineering*. Springer Science & Business Media.
- Yli-Huumo, J., Maglyas, A., Smolander, K., 2014. The sources and approaches to management of technical debt: A case study of two product lines in a middle-size finnish software company. In: *Jedlitschka, A., Kuvaja, P., Kuhrmann, M., Männistö, T., Münch, J., Raatikainen, M. (Eds.), Product-Focused Software Process Improvement*. Springer International Publishing, Cham, pp. 93–107.
- Yli-Huumo, J., Maglyas, A., Smolander, K., 2015. The benefits and consequences of workarounds in software development projects. In: *International Conference of Software Business*. Springer, pp. 1–16.
- Yli-Huumo, J., Maglyas, A., Smolander, K., 2016. How do software development teams manage technical debt? – an empirical study. *J. Syst. Softw.* 120, 195–218.
- Zazworka, N., Izurieta, C., Wong, S., Cai, Y., Seaman, C., Shull, F., et al., 2014. Comparing four approaches for technical debt identification. *Softw. Qual. J.* 22 (3), 403–426.
- Zazworka, N., Spínola, R.O., Vetrò, A., Shull, F., Seaman, C., 2013. A case study on effectively identifying technical debt. In: *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering*, pp. 42–47.