Herausgeber et al. (Hrsg.): Name-der-Konferenz,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 11

Evaluation of Social Value Icons for a Domain-Specific Modeling Language

Andreas Fritsch, Stefanie Betz[1]

Abstract: Sustainability is one of the major challenges in modern society. Life Cycle Assessment with regard to sustainability is one of the current tools to asses and communicate a product's impact on sustainability. But currently, life cycle assessments are difficult to communicate to the end-user due to their inherent complexity. Domain-specific modeling languages (DSML) can help to communicate complex information. Therefore, we have developed a DSML – TracyML – as a visual language to effectively communicate social sustainability of a product life cycle. Additionally, to help with effectively communicating the different aspects of social sustainability such as child labor and excessive working hours (among others), we have enriched the DSML with social value icons, representing these different aspects. In order to evaluate whether the icons are understood as we intended them to be, we conducted a survey. Thus, in this paper, we present an evaluation of the associativeness of TracyML's social value icons based on the Multiple Index Approach using an online survey. The main outcome of the evaluation is that the icons (except for one) are all comprehensible and associated with the correct social values.

Keywords: Social Life Cycle Assessment; Domain-Specific Modeling Languages; Multiple Index Approach

1 Introduction

Domain-specific modeling languages (DSMLs) provide concepts (and a corresponding notation) that are intended to increase modeling productivity and comprehensibility for domain experts [Fr13]. Examples for domains, where specific modeling languages have been created for, are IT-Risks [SHF11], organizational goal planning [BF16], and clinical pathways [He15]. In contrast to DSMLs, general purpose modeling languages, like for example UML and ER for software engineering, provide rather abstract concepts that foster high re-usability and applicability to diverse domains, but are not easily understood by novice users. Especially the design of visual notations for modeling languages attracted increased attention in the scientific community during the last years. Much tribute goes to Moody who gathered findings from psychology and usability research to compile "principles" of notation design against which the cognitive effectiveness of visual modeling languages can be judged [Mo09]. Previously, language designers often naively assumed that the mere existence of a visual notation helps to understand complex models. Following Moody's

[1] Karlsruhe Institute of Technology, AIFB, Kaiserstr. 89, 76133 Karlsruhe, Germany, firstname.lastname@kit.edu

line of arguments it becomes apparent that a visual notation must be carefully designed in a way that actually supports a user in understanding the model. The careful design of a new DSML for a specific purpose then bears the additional benefit of improved cognitive effectiveness for the users. The authors of this contribution have developed TRACYML as a visual language to effectively communicate social sustainability of a product life cycle. The development process was based on guidelines for DSML design as proposed by Frank [Fr10]. Diverse conceptual issues are covered in the guidelines, but they also comprise a set of rules of thumb for graphical notation design based on Moody's principles. While these rules are very helpful for the initial design phase, we believe that empirical evaluation of a notation, based on established usability methods, can yield fruitful and constructive results. To illustrate this, we have focused on one of these rules that appeared especially critical to us in the context of TRACYML: One of Franks guidelines states that icons contribute to a more intuitive understanding. One could argue that this is not universally true: the use of an icon might contribute to an intuitive understanding - but it might as well be detrimental to it, as a user may associate another meaning than the intended one. This criticism appeared very relevant to us, especially in the context of TRACYML, where icons are used to represent complex aspects of social sustainability. In order to gather insights for a potential revision, we performed an evaluation of the associativeness of TRACYML's social value icons.

The next section will give some basic background about social life cycle assessment to provide context for the reader. We will also discuss some comparable user studies and how they are related to our work. Section 3 presents the applied evaluation method and our study design, as well as a discussion of the results. We will finally discuss threats to the validity of our study and conclude with an outlook (section 4).

2 Background and Related Work

Social Life Cycle Assessment (S-LCA) is a methodology aimed at assessing the social impacts of a product's life cycle, from extraction of raw materials to final disposal [BM09]. Thus, a S-LCA study identifies affected human stakeholder groups, how they are affected, as well as the severity of the effects. Being a relatively young methodology, S-LCA faces some major research challenges [Sa15] [BM09]. For one, a S-LCA study requires a large amount of data and produces likewise complex results, rendering the presentation and communication of the results difficult [BM09]. Our goal with TRACYML is then, to explore the potential of conceptual modeling, and especially DSMLs, in supporting the execution of S-LCA studies, as well as the presentation and communication of study results.[2] For our initial design of TRACYML, we have focused on the stakeholder group 'workers' and corresponding aspects of social sustainability based on methodological sheets for S-LCA [Be13]. Herein, recommendations for critical social aspects concerning various stakeholder groups are given. To support our visual notation, we have defined icons for each of the

[2] The interested reader may find additional information about TRACYML, as well as a prototypical modeling tool under `https://www.gotracy.org`.

social values that the methodological sheets propose for the stakeholder group workers. For example, we defined an icon that depicts a clock to refer to the social value 'working hours'. The social value 'working hours' addresses the occurrence of excessive working hours in a factory or an industry. A more detailed description of all the covered social values can be found in the methodological sheets (see [Be13]).

There have also been some efforts concerning the evaluation of icons in the field of conceptual modeling that are related to our work presented here. However, we have not found any example of an icon evaluation as a constructive part of a DSML design process. In the following, we will briefly outline, how our study design differs from comparable studies in the field of conceptual modeling. Caire et al. [Ca13] criticize the visual design of the requirements engineering notation i*. They propose a participatory design process, where they invite different user groups to create new symbols for the language concepts. Afterwards, the symbol sets are evaluated via a comprehension test, where naive users were shown a symbol and then asked to infer its meaning. Khanom et al. [KHK13] have also proposed and evaluated some icons to visualize requirements engineering concepts. They performed a somewhat similar kind of test, where users where shown the icons and asked to infer the correct meaning from a list. The evaluation methods utilized by Caire et al. and Khanom et al. differ in one subtle yet significant way from our test design: in their case, the users where shown the icons and then asked to provide the correct meaning (or 'referent'). In our case, we have first shown the referent (a description of the social value) to the users, who were then asked to choose (from the complete set) the icon that they associate with the social value. One can argue that such a test design resembles more closely an actual real-life situation, where a user has all icons visually present, but not a complete cognitive list of all corresponding concepts [ET93].

3 Presentation and Discussion of the Study

The evaluation of the icon design can be described as a 'test of associativeness' based on the Multiple Index Approach as defined by the European Telecommunications Standards Institute [ET93]. As we have defined only one set of icons, we do not compare various sets to decide which is the best, but rather collect performance data for each icon that may hint at weaknesses in the design. The participants of the study received a brief description and list of all icons per social value. From the description they had to guess the correct corresponding icon. From the answers we calculated the hit rate (# correct responses / # total responses) and false alarms (# of wrong responses).

Material: We created an online questionnaire with the tool LamaPoll (`www.lamapoll.com`). An initial draft of the questionnaire was piloted. Its structure is briefly presented in the following (the whole questionnaire is available under `https://gotracy.org/workingpapers/tracyml_icon_evaluation.pdf`): Question blocks were presented on separate pages to make sure that questions are not overlooked when scrolling down on one page. On the first page a brief description of the study's objective and context was given. It further informed

the participants about the required time to answer the questions (five to ten minutes) and mentioned a contact person (see [KF14]). The following pages presented each a social value (see Table 1) and asked the participant which of the icons conveyed the meaning best. In order to prevent sequencing errors, the appearance of the pages and the order of the icons on each page were randomized [ET93]. When constructing the questionnaire we followed the guidelines of [KF14]: Socio-demographic questions where put at the end of the questionnaire, since asking personal questions right at the beginning might lead to cancellation or refusal of the survey . The page showed a brief introductory text that emphasized the anonymity of the participants. Finally, on the last page, the participants were asked for further feedback or comments [KF14].

Participants: The link to the survey was distributed employees of an organization concerned with product sustainability, but also to general public to capture both, possible future modelers and viewers of diagrams. Altogether, the survey had 23 participants and of the participants, 87% were aged between 25 and 39 years, 4% between 40 and 59 and 9% between 60 and 79. 43% were male and 57% female.

Results: Table 1 shows the results from the survey, where one can see that most of the icons received high hit rates. The icons for the values *equal opportunities* and *social security* show the lowest hit rates with 74% and 61%, respectively. Interestingly, the icon for *freedom of association* has a very good hit rate, but at the same time an exceptionally high number of 17 false alarms. Regarding false alarms, the results for *child labour* (4) and *social security* (5) are also considerable. The potentially problematic icons revealed by hit rates and false alarms were also addressed in some comments. One proposition was that the child depicted in the icon for *child labour* could play with a ball to convey the message more clearly.

Icon	Referent	HR	FA	Icon	Referent	HR	FA
	Freedom of Assoc.	91%	17		Child Labour	83%	4
	Forced Labour	83%	1		Equal Opport.	74%	0
	Working Hours	83%	1		Fair Salary	100%	2
	Health and Safety	96%	1		Social Security	61%	5

Tab. 1: Survey results. Icons were selected from `http://fontaweseome.io`.

Some comments from the participants revealed another issue that is unique to the evaluated domain: an icon might be associable with its referring concept, but still provoke critique: our symbol for *equal opportunities* depicts the common symbol for men and women. And so one participant did note very correctly that there is not only gender-based discrimination.

The icon with the lowest hit rate was the one for *social security*: 61% of the participants selected the "correct" symbol (depicting a lifebelt), while another 26% chose the symbol for *freedom of association* (depicting a group of people). Two possible interpretations come to mind: (1) Many participants intuitively interpreted the symbol depicting a group of people as a "community" that gives some kind of social security. (2) The lifebelt symbol was hard to recognize or interpret. As we know from the overall high number of false alarms for *freedom of association*, this icon is often misinterpreted and therefore a good candidate for revision. One possibility would be to reinforce the message of freedom of association, for example by depicting a group of people that carry protest signs. One can assume that clarifying the meaning of the *freedom of association* icon would at the same time improve the hit rate for *social security*. We conclude that the evaluation revealed the most critical icons. At the same time, a deeper look at the results revealed promising possibilities for improvement. Altogether, the other icons show an acceptable performance in the evaluation.

Discussion: In this subsection we discuss the limitations of our work based on the threats to validity based on [RH08]. We are aware of the limitation of our data caused by the small number of survey participants. However, the results show clear indications of the comprehensibility of the icons (except for one) which was the main purpose of the survey. But, our data is limited and further research needs to be done especially with regard to the inner coherency of the icons. Another threat to validity is that the survey questions may be misunderstood or misinterpreted. To mitigate this threat several actions have been taken. The survey has been pre-tested and the participants were given the opportunity to state further feedback and comments. Sequencing errors are a threat to internal validity to comprehension and association tests. In order to prevent those, the appearance of the pages and the order of the icons on each page were randomized. Additionally, we only showed one social value per page to avoid confusion, overlooking and comparison. One risk to external validity is the not that high number of participants (23) in our survey and thus the limitations in generalizing the results. Yet, the data showed a saturation of the results. Between the 14th and 23rd participant the hit rates only changed by very few percentage points. The most visible change was that the comprehensibility of the one 'critical' icon for social security went from 44% to 61%. One threat to reliability are unclear survey questionnaires. We mitigate this threat by pre-testing the questionnaire.

4 Conclusion and Outlook

We have tested the associativeness of social value icons for a domain-specific modeling language. Altogether, all icons except for one showed an acceptable performance. A deeper analysis revealed constructive insights to further improve the icon design. This result shows the applicability of established usability testing methods to a DSML design process. As an outlook, we envision a catalogue of guidelines for empirical DSML evaluation that would help language designers to further improve the usability of modeling languages.

References

[Be13] Benoît Norris, C; Traverso, Marzia; Valdivia, Sonia; Vickery-Niederman, Gina; Franze, Juliane; Azuero, Lina; Ciroth, Andreas; Mazijn, Bernard; Aulisio, D: The methodological sheets for sub-categories in S-LCA. UNEP and SETAC, 2013.

[BF16] Bock, Alexander; Frank, Ulrich: MEMO GoalML: A Context-Enriched Modeling Language to Support Reflective Organizational Goal Planning and Decision Processes. In: ER 2016, volume 9974 of LNCS, pp. 515–529. Springer, 2016.

[BM09] Benoît, Catherine; Mazijn, Bernard, eds. Guidelines for Social Life Cycle Assessment of Products. UNEP and SETAC, 2009.

[Ca13] Caire, Patrice; Genon, Nicolas; Heymans, Patrick; Moody, Daniel L: Visual notation design 2.0: Towards user comprehensible requirements engineering notations. In: 2013 21st IEEE International Requirements Engineering Conference (RE). pp. 115–124, 2013.

[ET93] ETSI: Human Factors (HF); The Multiple Index Approach (MIA) for the evaluation of pictograms. Technical report, ETSI, 1993.

[Fr10] Frank, Ulrich: Outline of a Method for Designing Domain-Specific Modelling Languages. Technical Report 42, ICB-Research Report, 2010.

[Fr13] Frank, Ulrich: Domain-Specific Modeling Languages: Requirements Analysis and Design Guidelines. In (Reinhartz-Berger, Iris; Sturm, Arnon; Clark, Tony; Cohen, Sholom; Bettin, Jorn, eds): Domain Engineering, pp. 133–157. Springer, Berlin, Heidelberg, 2013.

[He15] Heß, Michael; Kaczmarek, Monika; Frank, Ulrich; Podleska, Lars; Täger, Georg: A Domain-specific Modelling Language for Clinical Pathways in the Realm of Multi-perspective Hospital Modelling. In: ECIS 2015 Completed Research Papers. 2015.

[KF14] Klöckner, Jennifer; Friedrichs, Jürgen: Gesamtgestaltung des Fragebogens. In (Baur, Nina; Blasius, Jörg, eds): Handbuch Methoden der empirischen Sozialforschung, pp. 675–685. Springer Fachmedien, Wiesbaden, 2014.

[KHK13] Khanom, Sukanya; Heimbürger, Anneli; Kärkkäinen, Tommi: Icons: Visual Representation to Enrich Requirements Engineering Work. J Software Eng App, 06(11):610–622, 2013.

[Mo09] Moody, D.: The Physics of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. IEEE Trans. Software Eng., 35(6):756–779, 2009.

[RH08] Runeson, Per; Höst, Martin: Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2):131–164, 2008.

[Sa15] Sala, S; Vasta, A; Mancini, L; Dewulf, J; Rosenbaum, E: Social Life Cycle Assessment. Technical report, EUR, 2015.

[SHF11] Strecker, Stefan; Heise, David; Frank, Ulrich: RiskM: A multi-perspective modeling method for IT risk assessment. Inf Syst Front, 13(4):595–611, 2011.