Evaluating Information Visualization in Large Companies: Challenges, Experiences and Recommendations

Michael Sedlmair[†]   Petra Isenberg[∗]   Dominikus Baur[‡]   Andreas Butz[‡]

[†]BMW Group
Research and Technology
michael.sedlmair@bmw.de

[∗]Department of Computer Science
University of Calgary
petra.isenberg@ucalgary.ca

[‡]Media Informatics Group
Ludwig-Maximilians University Munich
(dominikus.baur | andreas.butz)@ifi.lmu.de

ABSTRACT

We examine the process and some implications of evaluating information visualization in a large company setting. While several researchers have addressed the difficulties of evaluating information visualizations with regards to changing data, tasks, and visual encodings, considerably less work has been published on the difficulties of evaluation within specific work contexts. In this paper, we specifically focus on the challenges arising in the context of large companies with several thousand employees. We present a collection of evaluation challenges, discuss our own experiences conducting information visualization evaluation within the context of a large automotive company, and present a set of recommendations derived from our experiences. The set of challenges and recommendations can aid researchers and practitioners in preparing and conducting evaluations of their products within a large company setting.

Keywords

Information visualization, evaluation, company setting

1. INTRODUCTION

Evaluation in the context of specific data and task sets is a fundamental part of information visualization research [23, 33] as systems and techniques developed by researchers are often intended to support everyday work activities for domain-specific tasks and data. In order to more clearly understand and assess "real world" data analysis problems and the use of our tools within a specific work context, a close collaboration with domain experts is often instrumental [9, 25, 27, 34]. When working with domain experts on their own data and tasks, it is often helpful or necessary to study their data analysis habits, requirements, goals, and tool use within their respective work context, or "field" [21]. Existing types of research strategies within the field can be roughly categorized as field studies and field experiments [17]. Field studies are described by McGrath [17] as direct observations with minimal possible intrusion within "natural" work environments, whereas field experiments are a compromise strat-

egy where features of the system are manipulated in order to gain more precise results. In this paper, we concentrate on the challenges of applying field strategies within one specific type of field—that of large industrial companies of several thousand employees. We derive our findings from our three-year experience working within a large automotive company. From our own experience we know that applying and evaluating information visualizations directly within a large company context is a fruitful endeavor and can produce valuable insights for the field of information visualization in general. Within such an environment a wide range of real data analysis problems, tasks, and data sets are available. Large companies are also often highly interested in applied research and will even fund it.

Evaluation for and of information visualization solutions within this context, however, has its own unique set of requirements and challenges due to the structural differences in contrast to small companies, such as a higher degree of organizational complexity, more specialization, formalization and decentralization [6]. In this paper, we categorize field-specific challenges that may arise in evaluating information visualization habits and tools in such an industrial environment based on our experience in building and evaluating information visualization systems for automotive engineers. We discuss a set of nine challenges we encountered, our own experiences with these challenges, and present a set of twelve recommendations for evaluation within a large company context. Challenges and recommendations include both aspects specific for information visualization evaluation but also more generic considerations which are no less important for our research. We hope that this paper will help to serve as a reference for others who are planning information visualization evaluations within a large company context.

2. RELATED WORK

In this section, we discuss previous field strategies that were conducted with information visualization tools and go into more detail on obstacles of field research as discussed in the general HCI literature.

2.1 InfoVis Evaluation in the Field

Despite the well known drawbacks of artificial scenarios and hypothetical tasks, most evaluations for information visualization tools are still conducted in lab settings [9]. Previous researchers, however, have called for more real world applications of research (e. g., [14, 20, 23]) and a growing number of researchers are beginning to invite their target audience to participate in user studies: Perer et al. [22], for example,

studied their social network tool with several experts from different fields of data analysis. Ethnographic studies have also been used within a user centric design process with domain experts and have been shown valuable as a formative part of the design process: Tory et al. [36], for example, documented the results of a qualitative analysis in the building design field and concluded that their structured analysis of qualitative study data provided deep insight on the work processes with visualization. Long-term studies [32] are another type of field strategy that offers the chance for deep insight and learning of the workings of a field and possible merits of visualization use. Unfortunately, they are laborious and only few have been reported on in the literature (e. g., [11, 18, 28]). The work by González and Kobsa, for example, describes the adoption of an information visualization tool by data analysts over a longer period of time [11]. In a follow-up paper, they describe further observations on the merits of such tools in the workplace [10]. While these examples are promising steps towards more evaluation in close contact to domain experts, more insight is needed on the challenges of conducting information visualization evaluation within specific work contexts. Our paper is a step in this direction and lists a first set of challenges, experiences, and recommendations for deploying and evaluating information visualization within a large industrial company.

2.2 Organizational obstacles known from HCI

In the area of HCI, more precisely in Participatory and Contextual User-Centered Design (UCD), a considerable amount of previous work exists on how to meet usability evaluation and user needs by actively involving all stakeholders (e. g., end users, management, decision makers). Much of this research has been conducted in industry settings [1, 16]. Grudin [12] explicitly discusses obstacles encountered in large companies such as finding "representative" participants and crossing organizational barriers during a UCD process. Poltrock and Grudin [24] conducted two observational studies in large companies and reported how several organizational factors (e. g., missing commitment, unsatisfying training) can block UCD. Jeffries et al. [15] provided a comparison of four formative usability studies in real world environments and recommended heuristic evaluation and usability testing methods for evaluation when considering the number of found problems, UI expertise, and costs.

The main difference of our work to most of these approaches is that we do not examine business-to-customer situations: While much of the previous work was concerned with employing UCD to develop tools for expert users on the outside, we are interested in designing information visualization tools for use within a large company to improve the work processes of its employees. While novel requirements and challenges applying UCD for in-house tools, such as platform and application buying concerns, change management, or the IT life cycle, have been previously discussed [4, 13] related work in this specific area is still rare. In particular, the challenges of information visualization evaluation—as opposed to general usability evaluation—have not received much attention in this context. We contribute a first collection of challenges and recommendations for applying evaluation within a large company context and hope that this collection will be expanded and modified as more evaluations of information visualizations will be conducted in this work context.

3. PROBLEMS AND CHALLENGES

While designing and evaluating information visualizations within a large company for the past three years, we have experienced several field characteristics that pose particular challenges to evaluation. These challenges arise due to the large company setting where workflow, bureaucracy, or hierarchical structures may be quite differently defined compared to smaller companies ([6, 26]). For instance, large industrial companies are often characterized by a high degree of collaboration and specification. A single employee often is highly specialized and responsible for a small subset of a highly specific collaborative task set [6]. Therefore, the know-how in a company is often widely distributed and a single person is not always able to understand all facets of the entire task domain [8]. In a small company or research lab, on the other hand, a problem domain may be very specific and employees may be able to maintain a comprehensive understanding of their work context and may even be able to deal with many tasks personally.

When attempting to evaluate information visualization within a large company context, it is imperative to understand the characteristics of this specific evaluation field in order to be prepared for the challenges that may arise in planning and conducting a study and finally analyzing and disseminating the results. In the following section, we describe nine specific challenges to evaluation of information visualization in large companies. We categorize them along the typical flow of a user study: study design, participant gathering, data collection, data analysis, and result generation. We consider challenges both to studies of already developed visualization tools (tool-centric) as well as challenges when attempting to evaluate current practice within the company setting (work-centric). We do not consider evaluations that try to assess how clear the interaction with an information visualization tool is designed (usability evaluation) but focus instead on evaluations for information visualization design that are more holistic in nature. We ground our collection of challenges on our own experience with different techniques (both qualitative and quantitative) and on general lessons learned from both the HCI and sociology literature and do not focus on a specific evaluation methodology, data collection or analysis method.

We also include challenges related to tool deployment within the company setting which has been both extremely valuable, as a prerequisite to longer-term studies, as well as challenging for us. We describe specific challenges of deployment that are information visualization specific but also other challenges in order to give a more complete picture of our experiences.

3.1 Study/Application Design

C-1a: Integrating Tools in Daily Work Processes
Integrating information visualization tools in daily work practice is a labor-intensive process, not only in large companies. Tools have to be stable, robust to changing data sets and tasks, and—if they replace previous tools—should support the functionalities of the tools being replaced. Besides these common challenges, we describe two critical aspects to consider in large company settings:
(a) Technical Issues: Task specialization is common in large industrial companies. Therefore, many specific data analysis tasks exist and most of these will likely already be supported with a variety of different analysis tools. These tools are of-

ten well integrated to perform within a chain of other tools so that they together provide more encompassing analysis solutions. Under these circumstances the integration of a new visualization tool may be quite challenging as it may break the chain of analysis processes that are already supported by existing solutions. However, the integration of a specific tool may be a valuable exercise in practice, in particular when one wants to study the use of a tool within a specific established work context [27].

(b) Political and Organizational Issues: Many large companies require the authorization of software or software components upfront. Initially this may not seem complicated, however, depending on the amount of bureaucracy involved, this process may require highly collaborative synchronization efforts and may become long and exhausting. One method to get your software authorized is *employee pull*: the specific request for a tool by an employee. Another method is *evaluator push*: advertising on your side as the visualization expert for your specific tool. Both approaches may often be successful: Pull-solutions are often easier because employees can argue that your tool may address a recognized analysis problem. Push-solutions may require very tactful negotiation but are no less important. Specific work practices may have become established over the years and employees may be satisfied with improvable solutions. In these cases, a push from an outsider can help to provide a new perspective on more advanced data analysis options.

C-1b: Getting the Data

Not only the tools and techniques but also the domain-specific data itself will likely be distributed across different work groups within large companies. Your novel visualization approach, however, might have been designed to improve work with combined and aggregated sources of data. To evaluate your tool with these data sources you may have to deal with issues of interoperability between different data sources on different machines and within different work groups. Unavailability, different data versions, different or inappropriate format, unmaintained sources, and most importantly security restrictions can issue additional challenges to you. However, being able to evaluate visualizations with the data used and created by your participants in their everyday work practices can be critical—not only in evaluating how your tool is used with real-world data characteristics, but also in order to convince the participants or stakeholders that your tool may actually improve everyday work.

C-1c: Choosing an Evaluation Context

Large companies have employees with varying goals, views, and work habits all working together [8]. In large industrial companies you will encounter a variety of personalities and opinions. This is particularly important to keep in mind when you are planning to conduct qualitative work such as interviews, observations, and focus groups with or without information visualization tools. There may be many teams with similar data analysis tasks and data types across a company that you can collaborate with but the qualitative results you may collect during a study in these teams can be vastly different.

3.2 Participants

C-2a: Finding Domain Expert Participants

It is very common that employees in large industrial compa-

nies are working under heavy time pressure and are bound to strict deadlines. Having to revise a deadline often leads to a considerable loss of revenue. These pressures result in specific challenges for evaluation in general and for evaluation with significant participant involvement in particular: (a) Getting domain experts for studies is generally difficult. Time = money! Every hour you want participants to work with you is an extraordinary task without direct evidence of impact on their actual work tasks. (b) Under these circumstances, it becomes difficult to argue for long-term studies (e. g., MILCs [32]) without any kind of "pre-evidence" that the required involvement will result in qualitative or quantitative improvements to future work processes.

C-2b: Attachment to Conventional Techniques

Even if your tool may be designed to improve conventional tools, experts may be very accustomed to and effective with them. This effectiveness may lead to a certain amount of attachment to the traditional tool and may result in a certain reluctance to learn a new system. By working with their traditional tools over a long period of time people will likely have developed skills to estimate the effort and time required for a specific analysis and can factor this knowledge in when planning upcoming deadlines. It may be difficult for them to estimate this with a novel tool. In addition, some domain experts may have learned to master complex tools and data analysis tasks over the years. If you managed to design a tool that significantly simplifies a specific data analysis compared to a previous tool you may strip these experts from their respected expert status and allow others to also conduct the same tasks [21]. These issues can complicate both acquiring participants for your studies (see also C-2a) and conducting and evaluating comparative studies.

3.3 Data Collection

C-3a: Confidentiality of Information

Video-, audio- and screen-recording can be useful data collection tools during evaluation. Especially for qualitative evaluation such data collection helps to capture participants' actions, conversations and responses and allows systematic coding and analysis of the data in retrospect [2]. However, large companies often have confidentiality guidelines and restriction policies (Intellectual Property Rights (IPR) security requirements) that might forbid certain recording techniques. In addition, being discreet about collected data is important. Internal work processes are often secret. This, in particular, means that you may not be allowed to share your data with others (e. g., with a second coder or in online tools), you may only be allowed to discuss anonymized results, and that you will have to deal with publication restrictions—not only about the results of your study but also when talking about the data analysis characteristics of the tool you may have studied (see C-4b).

C-3b: Complex Work Processes

One important goal in information visualization is to support people in solving complex tasks. For this purpose, an important first step is to understand current data analysis problems with pre-design evaluation [14]. For us, this type of evaluation has been a very important step in order to focus our work on solving the right real-world problems. Pre-design studies, however, become additionally challenging in large companies where complex problems are often

split among several, highly specific sub-problems. Understanding the specifics of both the overarching problem solving process (macro challenges) and the individual (micro) challenges may be difficult for an outsider (see C-2b). When observing different employee groups, it should not surprise that some may have built their own work processes or tools around their work tasks or data sets and that other groups and employees, who may have similar data, may have come up with different solutions while being unaware of solutions from other groups (see C-1c). Additionally, experts in large companies often have varying tasks and not all of them may be relevant for the observer and neither do the domain experts want to be observed in every situation. Finding the appropriate balance between unobtrusive observation and intervention when observing work processes requires skill and tact on the side of the evaluator. On the other side, talking to participants in pre-scheduled appointments is also often not sufficient: "What people tell you is not always the same as what people do" [3].

3.4 Results

C-4a: Convincing the Stakeholders
An important evaluation goal in information visualization is to understand how people use your visualizations to solve real world problems. This goal does not necessarily align well with the goals of stakeholders whose task it is to maximize profit for the company. Therefore, they are more interested in tools that help to save money and improve the effectiveness and efficiency of their employees (again, time = money). Another goal of the company is speeding up current work practices (e. g., more insights/time [19]) while we as researchers may be more interested in factors that influence or improve qualitative aspects of the work or the specific factors that may have led to improvements (e. g., how insights were achieved [28]).

C-4b: Publishing.
To allow information visualization to grow as a field and to share and discuss your results with the larger research community they have to be made public. Due to competitive reasons, large companies however often have restrictions on what can be published, in particular if your work leads to a competitive advantage. You may have to expect a lot of bureaucratic hurdles.

4. EXPERIENCES
In this section, we illustrate some of the listed challenges with an example from our own experience in building visualizations for the development and analysis of in-car communication networks at a large automotive company of approx. 100,000 employees. This example shows only a fraction of the body of work underlying our collection of challenges and recommendations, but it serves well as an illustration. We report on evaluations conducted with domain experts within the context of one specific visualization tool we built and discuss the challenges we encountered.

Our chosen example–*AutobahnVis* (see Figure 1)–is a visualization tool to support overview, navigation, and pattern recognition for error detection in large network communication logs. In-car communication networks are highly com-
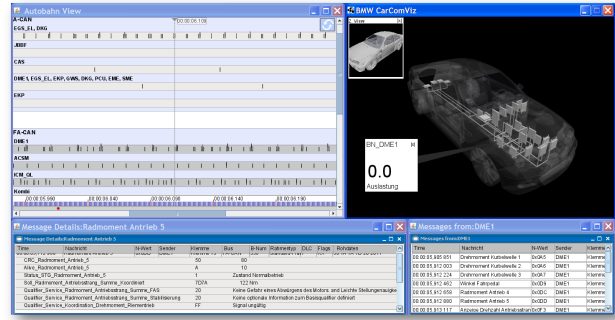

Figure 1: AutobahnVis - Top left: The *AutobahnView* showing all sent message ordered by time (horizontally) and sending Control Unit (vertically); Bottom left and right: *List Views* showing message details, such as row data, signals, time stamps; Top right: The *3DModelView* showing a 3D model of the car linked to the other views via semantic linking and brushing

plex (up to 15,000 messages per second with safety critical real-time requirements) and include a large number of electronic components (up to 70 Control Units per vehicle, sending and receiving messages; and 5 different communication technologies, transporting messages). These factors make their design and analysis a challenging process that benefits highly from visualization support. We conducted several types of evaluations within the context of this project [30, 31].

4.1 Pre-Design Studies
In order to assess situations in which visualization could support the domain expert's work practice and to understand their data analysis problems we first conducted a pre-design study with eight domain experts [3]. Originally we had planned to use a shadowing technique for participant observation without any interference, however after a pilot study it became clear that we would not understand their work processes solely through observation (C-3b). We changed the pure observational design to a variant of a contextual inquiry, where one analysis expert explained his or her work to 2–3 observers by means of a current work problem. When conducting this study with the domain experts we also ran into the challenge of attachment to conventional techniques (C-2b). One participant explained his work on the raw data to us and made it clear that visualization would not be useful to him as it "is just a potential source of error". This participant had learned over the years to read the raw hexadecimal data and had the status of an analysis expert (C-2b).

IPR security requirements were in place in the manufacturing areas in which we conducted our study. In order to allow participants to speak openly, we decided not to use any electronic recording devices (C-3a). Our approach was to counterbalance the potential loss of information by using 2–3 observers instead of just one, all taking notes and immediately preparing a summary after the observations. In order to collect more data on our observations, we conducted a second study in form of a brief online questionnaire. We contacted a wider range of analysis experts and asked questions on current work practices as derived from the first study.

Our main goal was to reach a wider range of people without strongly interfering with their daily work (C-2a).

Finally, in order to get feedback on our first prototypes we discussed paper mockups with the domain experts. A meta-goal of this last part of the study was to evaluate the degree of interest of each participant. Three participants showed keen interest in our ideas and approaches. For our future collaboration we focused mainly on this sub-group (C-1c).

4.2 Study During the Design Phase

From the pre-design studies we derived a task list, design and feature requirements as well as a ranked list of preferred visualization mockups. Based on these, we built a multiple coordinated view prototype (see Figure 1) visualizing real data and iteratively evaluated the design. These studies were conducted with students with a usability background in order to save domain experts' time (C-2a) and due to challenges of integrating our tool in the domain experts' work environment (C-1a). We also iteratively conducted expert reviews with one usability expert with a research background in HCI/ InfoVis [35]. These two approaches helped us to focus on usability issues alongside the entire development process. For final revisions, we conducted a think aloud study with five students with a usability background plus basic automotive experience (automotive company interns). These studies were used for final usability optimization. The direct integration of our solution in the domain experts' current software environment was extremely difficult due to a variety of different file formats in use (C-1b). We opted for an exported (standard) file format (C-1a). Additionally, we had to gather unavailable data (for controlling a semantically coordinated 3D-model view in our tool) manually from textual sources (C-1b).

4.3 Post-Design Study

After the first study with students, we conducted a qualitative user study in form of a "pair analytics" evaluation to validate the value of our approach with five domain experts. Roughly spoken, one of us worked with one expert analyzing a (real, but partly manually translated) test trace using our tool and discussing potential benefits and drawbacks. Due to IPR restrictions, we again did not use audio and video recording (C-3a). During these studies we learned that domain experts gained several novel insights from our tools such as message burst detection, better understanding of cyclic messaging and insights into cross-relations between mechanical and electronic information. These promising results (a) showed the validity of our approach for our test data sets and (b) helped us to promote it for a close integration with current analysis software (C4-a). According to our requirement analysis (see 4.1), only a close integration—without additional time costs (e. g., exporting data or manually translating data, C-2a)—will allow analysis experts to use our visualization productively on a day-to-day basis with real world, dynamically varying data sets (C-1b). However, close integration requires (a) company commitment which we, in case of the AutobahnVis example, achieved by our highly participatory design process; and (b) overcoming technical barriers by close collaboration between tool developers and InfoVis researchers (C-1a). We are currently still working on this labor-intensive aspect.

5. RECOMMENDATIONS

We agree with Shneiderman et al. [32] that studies with real end users, with real data, in real environments are extremely important factors for learning about information visualization tools. Based on our experience we derive a set of recommendations for other information visualization researchers who are planning to conduct evaluations within a large company setting. The organization of our recommendations reflects the main categories of challenges in Section 3. Some of our recommendations are specific to working with data and data analysis tasks with information visualizations and some apply to evaluation in this field more generally.

5.1 Study/Application Design

R-1a: Familiarize yourself with currently used software environments

To evaluate the full working process of domain experts an information visualization tool should be integrated and coordinated with current domain specific techniques and tools to operate in an entire analysis environment. Many of the existing tools in a work environment, however, have often been worked on and extended over the years and an integration may be a considerable software engineering challenge. Instead of an integration one can consider to extend the features of a new visualization tool to unite the capabilities of a previous tool chain. Depending on the amount of previous work this could be a valid solution for small projects. The costs of either solution should be considered based on the goal of the evaluation. Supplementing existing tools is often the cheapest and most effective way [11]. In addition, consider push- and pull solutions (see Section 3.1) to achieve acceptance of your tool and study with the company work environment.

R-1b: Overcome technical obstacles of data integration

Our experience showed that new tools which require additional steps to work with domain-specific data may not be accepted in everyday work. While AutobahnVis, for example, was more flexible to adapt to different data and analysis settings, engineers argued that the additional required overhead of file conversion was their reason not to use the tool. Therefore, for the purpose of evaluating information visualization tools in large companies tightly integrating your tool to work with only a subset of the data may be more important than supporting wide applicability. This factor may not be important in research departments (where insight may outweigh time), but the obstacle of additional time requirements is crucial in industrial environments. Having to convert data manually should be a last resort [32].

R-1c: Choose your study environment with care

Obstacles for studying your solution or studying work environments often result not only from technical challenges but from political or organizational requirements. To conduct evaluations you need permissions and committed collaborators. In order to receive permission it is imperative that you find employees who will support your project and that you convince your stakeholders (see R-4a for further recommendations). You may encounter similar data analysis tasks and data across different groups within a large company. It takes skill as a researcher to generalize from the individual opinions and views encountered to find the right target group and

work environment for the tool you built or are interested in building.

When conducting pre-design studies, connect with motivated domain experts and start with identifying and understanding different sub-problems and sub-groups in your problem domain. Talk to various people and be open-minded towards existing solutions from other people beyond your target group. Use this knowledge to become an expert in this domain but do not try to solve everyone's problems. Rather try to find a specific sub-target group with specific and concrete problems and with interest in your work. After researching specific solutions and validating sub-domain specific solutions, try to abstract your lessons-learned to a more general approach.

5.2 Participants

R-2a: The magic one hour limit
Our experience showed that recruiting participants for one hour or less is significantly easier than for longer time periods. Employees are occupied with meetings, appointments, and deadlines and additional involvement in user studies just adds to this work load. Be prepared and professional in recruiting and conducting the study and stick to your suggested time limit.

R-2b: Convince your target audience
Even though participants may be very attached and used to their current tools there are some things you can do to convince them of your solution. Try to solve real problems of your target group even if these first-hand solutions are small and actually not the main focus of your work! People become immediately interested if you present solutions which they can use immediately with their own data. Your participants will be much more motivated to attend your studies when they know they will be remunerated by working on solutions of their current problems. One way to achieve this, is to integrate some simple but highly desired functions not available with current tools (for a successful example see [29]). Even outlining solutions, e.g., presenting some work of our early AutobahnVis ideas after exploratory studies, was very valuable to convince our participants of the potential value of our work.

R-2c: Delight with usability and aesthetics
Do not underestimate the value of usability and aesthetics. In in-house tools these aspects are often neglected [13]. Usability and aesthetics are important distinctive features you can use to gain acceptance of novel tools or to convince stakeholders. In AutobahnVis, for example, we integrated a view showing a 3D model of the car. In the beginning we had been very skeptical about the potential value of this view, nevertheless, it was explicitly demanded by our stakeholders. During our summative user studies we observed that along with several (smaller) insights that this view could provide, its aesthetics and fascination was frequently and explicitly addressed. Several of the subjects mentioned that it would be much easier to convince decision makers with this view.

R-2d: Learn from the experts
Identify experts in your problem domain. You can learn a lot by interviewing and observing their practices. Often, they may not be interested in your solutions because they have mastered problems already using their own approach. Try

to identify why their practices are effective and efficient and think about how you can use this knowledge in your tool to make it available to a wider range of people. During our exploratory studies of AutobahnVis, for example, talking to one specific expert helped us enormously in understanding the variety of potential error sources and the importance of a hexadecimal representation. Our tool design benefited from his experience.

5.3 Data Collection

R3-a: Try to get a license, do studies in any case
Check IPR policies (see C-3a) and, if required, try to get permission to video or audio tape. We agree with Dix el al. [7] that the analysis of recorded video or audio will allow you to gain a much deeper understanding of the scenario under study. If a permission was received, equipment has to be carefully installed. It is imperative that participants know about recording devices and that privacy concerns are thoroughly discussed. In particular in large companies, employees may be concerned about the company "watching" them.

In some areas IPR restrictions might be very strict and you may not be allowed to digitally record study sessions. In these cases, do qualitative user studies anyway and counterbalance the loss of documentation with more than one observer and with immediate notes and a summary (see AutobahnVis). Especially in secure areas this methodology additionally may allow participants to be more open about their work processes, data, and tasks.

R3-b: Be in constant, close cooperation
To support specific domain experts with information visualization it is important to get a clear understanding of their problem domain [21]. We have made good experiences with informal collaborations that helped us to get a very well-grounded and detailed knowledge about our target group: over the last three years we have talked to almost 100 domain experts, we conducted several types of studies (from pre- to post-design) and we directly worked together with the domain experts. We refer to such a process as "constant, close cooperation." Our ambitious goal was to gain a deep understanding of our problem domain. From our experience, this kind of constant, close cooperation is valuable especially in large industries where problems are often highly diverse and complex. We are aware that understanding all facets of a problem domain is time-intensive, however we think that this approach of 'designing with not for the people' helps to clearly tailor solutions to the needs of a target group and to develop effective and efficient tools. Being in constant, close cooperation can help to overcome some of the pitfalls of evaluation as outlined in [20].

5.4 Results

R-4a: The magic metric: Money.
In industrial settings the benefits of a new tool are often measured in terms of cost savings. These savings are closely related to other metrics used in information visualization evaluation such as insights [28] or errors [5]. However, in an industry setting, the most related one may be time (again, time = money). Important quality metrics for stakeholders include such things as decisions per hour [19] or found errors per day (AutobahnVis). Reporting the results of your study and presenting evidence that your tool can lead to measur-

able benefits in terms of such metrics may be very important if you want to convince the stakeholders (see C4-a). While studies that measure these metrics may not always be able to get at the research questions you are interested in, they could be a ticket for reaching more domain experts and studying your solutions in-depth in real working environments.

For instance, we published a statistical comparative study between one of our tools and engineers' current state-of-the-art tools [29]. Proving that our tool was significantly faster and less error-prone for a set of predefined user-tasks was very convincing to the stakeholders. Subsequently, our tool was tightly integrated into a current software environment that was subject to strict access regulations (C-4a). Conducting a quantitative user study was therefore our ticket to reaching a lot of end users (C-2a) with real data (C-1b) in real environments (C-1a) and opened new possibilities for future long-term and more in-depth studies.

R-4b: Factor in high skill with current techniques

When comparing traditional to new tools, one must consider that participants may have become very skilled with current techniques (see C-2b) and factor in learning time and potential reluctantness towards a new tool as these factors can initially distort a comparative evaluation [32].

R-4c: Clarify publishing conditions upfront

If your main goal is to publish your work, make concrete agreements with your company upfront and preferably not just verbally. Make clear what you are allowed to write about, how or if you need to anonymize your results, what pictures (if any) you are allowed to include, and find out if the company requires you to submit your writeup for internal review first.

6.   DISCUSSION AND LIMITATIONS

In this paper, we have summarized some design challenges, examples, and recommendations for working with a large company on information visualization evaluation. We did not focus on the various advantageous aspects of conducting research in cooperation with a large company but decided to provide other researchers with the more valuable challenges and recommendations. Nevertheless, we want to encourage research cooperations with large companies for several reasons: as already mentioned, large companies provide a lot of interesting challenges and complex real-world data sets for information visualization research. In addition, although deployment and evaluation might be a long and laborious process, there are good chances that valuable solutions will be approved and integrated into real working environments. Thus, domain experts can benefit from dedicated information visualization solutions and researchers in return can investigate their systems under realistic circumstances [32]. Eventually, 'moving research into practice' remains one of our grand challenges [33]. We are convinced that closely cooperating with large companies will help us to better understand the value of information visualization.

Since a large part of our work is based on our own experiences there are two limitations to be aware of: (1) Through several different types of studies within different phases of the design process, we learned a lot about our domain and setting and successfully paved the way for integrating our tools with daily practices. Our work so far, however, does not address experiences from long term studies in a large company setting. We hope to specifically address this in the future and to extend our recommendations to long-term studies. (2) Experiences in other companies might differ or go beyond the ones we made. While the lessons we learned can serve as a reference for others who are planning information visualization evaluations within a large company context, this work should encourage others to (a) report their experiences evaluating within this field to create a broader knowledge base and (b) information visualization researchers in general to try this form of field research. Even though it may be more difficult to do than lab-based studies, we found this type of applied work very rewarding and we are convinced that it provides much more realistic insights than lab studies.

7.   CONCLUSIONS

We presented a number of specific challenges arising in the context of evaluating information visualizations in a large company with several thousands of employees. These nine challenges are grouped into those relating to: study/application design, participants, data collection, and results. This collection is based on a three year body of work involving eight prototypes and their development process. To illustrate these challenges, we discussed our experience in developing a visualization system for a large automotive company. Based on the experience in designing and evaluating these systems, we presented a set of recommendations for practitioners. With this collection of experiences and insights we hope to help others in preparing and conducting information visualization evaluations in similar settings and to encourage them to add and compare their experiences to our work.

8.   ACKNOWLEDGMENTS

9.   REFERENCES

[1] J. Bak, K. Nguyen, P. Risgaard, and J. Stage. Obstacles to usability evaluation in practice: A survey of software development organizations. In *Proc. of the Nordic Conference on Human-Computer Interaction (NordiCHI)*, pages 23–32, New York, NY, USA, 2008. ACM.

[2] H. Bernard. *Social Research Methods: Qualitative and Quantitative Approaches.* Sage Publications Inc., Thousand Oaks, London, New Delhi, 2000.

[3] H. Beyer and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems.* Morgan Kaufmann Publishers Inc., London, 1998.

[4] I. Boivie, C. Åborg, J. Persson, and M. Loefberg. Why usability gets lost or usability in in-house software development. *Interacting with Computers*, 15(4):623–639, 2003.

[5] C. Chen and M. P. Czerwinski. Empirical evaluation of information visualizations: an introduction. *International Journal on Human-Computer Studies*, 53(5):631–635, 2000.

[6] R. Daft. *Organization Theory and Design.* West St. Paul, MN, 1995.

[7] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction.* Prentice Hall, 2004.

[8] P. Drucker. The coming of the new organization. *Harvard Business Review*, 66(1):45–53, 1988.

[9] G. Ellis and A. Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proc. of the Workshop on BEyond time and errors (Beliv)*. ACM, 2006.

[10] V. González and A. Kobsa. Benefits of information visualization systems for administrative data analysts. In *Proc. of the International Conference on Information Visualization (IV)*, pages 331–336, Los Alamitos, CA, USA, 2003. IEEE Computer Society.

[11] V. González and A. Kobsa. A workplace study of the adoption of information visualization systems. In *I-KNOW'03: 3rd International Conference on Knowledge Management*, pages 92–102, 2003.

[12] J. Grudin. Obstacles to participatory design in large product development organizations. *Participatory Design: Principles and Practices*, pages 99–119, 1993.

[13] K. Holtzblatt, J. Barr, and L. Holtzblatt. Driving user centered design into it organizations: is it possible? In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI)*, pages 2727–2730, New York, NY, USA, 2009. ACM.

[14] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proc. of the Workshop on BEyond time and errors (Beliv)*, New York, NY, USA, 2008. ACM.

[15] R. Jeffries, J. Miller, C. Wharton, and K. Uyeda. User interface evaluation in the real world: a comparison of four techniques. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*, pages 119–124. ACM New York, NY, USA, 1991.

[16] J.-Y. Mao, K. Vredenburg, P. W. Smith, and T. Carey. The state of user-centered design practice. *Commun. ACM*, 48(3):105–109, 2005.

[17] J. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In *Human-Computer Interaction: Toward the Year 2000*, pages 152–169. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1995.

[18] P. McLachlan, T. Munzner, E. Koutsofios, and S. North. Liverac: Interactive visual exploration of system management time-series data. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*, pages 1483–1492. ACM New York, NY, USA, 2008.

[19] S. McNee and B. Arnette. Productivity as a metric for visual analytics: reflections on e-discovery. In *Proc. of the Workshop on BEyond time and errors (Beliv)*, New York, NY, USA, 2008. ACM.

[20] T. Munzner. A nested process model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[21] W. Paley. Interface and mind - a "paper lecture" about a domain-specific design methodology based on contemporary mind science. *it-Information Technology*, 51(3):131–141, 2009.

[22] A. Perer and B. Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*, pages 265–274, New York, NY, USA, 2008. ACM.

[23] C. Plaisant. The challenge of information visualization evaluation. In *Proc. of the Conference on Advanced Visual Interfaces (AVI)*, pages 109–116, New York, NY, USA, 2004. ACM.

[24] S. Poltrock and J. Grudin. Organizational obstacles to interface design and development: two participant-observer studies. *ACM Transactions on Computer-Human Interaction*, 1(1):52–80, 1994.

[25] J. Preece, Y. Rogers, and H. Sharp. *Beyond Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[26] D. Pugh, D. Hickson, and C. Hinings. An empirical taxonomy of structures of work organizations. *Administrative Science Quarterly*, 14(1):115–126, Mar. 1969.

[27] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, 2005.

[28] P. Saraiya, C. North, V. Lam, and K. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12:1511–1522, 2006.

[29] M. Sedlmair, C. Bernhold, D. Herrscher, S. Boring, and A. Butz. Mostvis: An interactive visualization supporting automotive engineers in most catalog exploration. In *Proc. of the Conference on Information Visualization (IV)*, pages 173–182, July 2009.

[30] M. Sedlmair, B. Kunze, W. Hintermaier, and A. Butz. User-centered development of a visual exploration system for in-car communication. In *Proc. of the Symposium on Smart Graphics (SG)*, pages 105–116, Berlin, Heidelberg, May 2009. Springer-Verlag.

[31] M. Sedlmair, K. Ruhland, F. Hennecke, A. Butz, S. Bioletti, and C. O'Sullivan. Towards the big picture: Enriching 3d models with information visualisation and vice versa. In *Proc. of the Symposium on Smart Graphics (SG)*, pages 27–39, Berlin, Heidelberg, May 2009. Springer-Verlag.

[32] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *BELIV '06: Proceedings of the 2006 AVI workshop on BEyond time and errors*, pages 1–7, New York, NY, USA, 2006. ACM.

[33] J. Thomas and K. Cook. Illuminating the path. *IEEE Computer Society*, 2005.

[34] M. Tory and T. Möller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, 2004.

[35] M. Tory and T. Möller. Evaluating visualizations: do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, 2005.

[36] M. Tory and S. Staub-French. Qualitative analysis of visualization: a building design field study. In *Proc. of the Workshop on BEyond time and errors (Beliv)*. ACM, 2008.