

Guidelines for Development and Evaluation of Usage Data Analytics Tools for Human-Machine Interactions with Industrial Manufacturing Systems

Jari Varsaluoma,
Heli Väättäjä
Tampere University of Technology
Tampere, Finland
first.lastname@tut.fi

Tomi Heimonen
University of Wisconsin-Stevens Point
Stevens Point, Wisconsin, USA
Tomi.Heimonen@uwsp.edu

Katariina Tiitinen, Jaakko
Hakulinen, Markku Turunen
University of Tampere
Tampere, Finland
first.lastname@uta.fi

Harri Nieminen
Fastems Ltd
Tampere, Finland

ABSTRACT

We present the lessons learned during the development and evaluation process for UX-sensors, a visual data analytics tool for inspecting logged usage data from flexible manufacturing systems (FMS). Based on the experiences during a collaborative development process with practitioners from one FMS supplier company, we propose guidelines to support other developers of visual data analytics tools for usage data logging in context of complex industrial systems. For instance, involving stakeholders with different roles can help to identify user requirements and generate valuable development ideas. Tool developers should confirm early access to real usage data from customers' systems and familiarize themselves with the log data structure. We argue that combining expert evaluations with field study methods can provide a more diverse set of usability issues to address. For future research, we encourage studies on insights emerging from usage data analytics and their impact on the viewpoints of the supplier and customer.

CCS CONCEPTS

• **Human-centered computing** → **Visualization design and evaluation methods; Empirical studies in HCI; Field studies**

KEYWORDS

Usage data logging; visual data analytics; human-machine interaction; flexible manufacturing systems; guidelines; lessons learned.

1 INTRODUCTION

Visual data analytics tools that are easy to use and support the goals of their users can help practitioners working in companies when inspecting collected usage data. However, a common challenge in studying visualization tools is how to evaluate the tools and their effectiveness [1]. Several studies have emphasized that when developing visualization tools, they should be evaluated in real work context with actual end users, instead of short-term laboratory experiments [e.g., 1, 2, 3]. The review by Isenberg et al. [4] shows that the number of the published evaluation studies of information visualization tools done in real use context has increased. However, the number of studies related to the UX of visualization tools was considered surprisingly low. This work is motivated by the call for more empirical research aiming to understand and design for the user experience (UX) of visualization tools [4]. Furthermore, little empirical research is available regarding the development of visual data analytics tools together with users in the context of industrial manufacturing systems. In practical terms, our research aimed to support a manufacturing automation systems company by developing and evaluating a prototype tool that enables the systematic use of the logged usage data to support product and service development and innovation activities through the gained insights. We describe a case study with an industrial manufacturing company, during which we developed UX-sensors, a visual data analytics tool for inspecting usage data. This work presents the tool itself, but focuses on the evaluation process and lessons learned during its development. Guidelines are provided to support the collaborative development process of similar visual data analytics tools for logged usage data, derived from our findings. In addition, we present our development process and discuss how it could be extended in future studies where visual data analytics tools are developed together with practitioners from companies. In the following, we first review related work regarding information visualization tool evaluation approaches with users

and usage data analytics in related industrial contexts. Then, we provide an overview of the study context and describe the UX-sensors tool. Next, we describe the iterative development and evaluation process for the UX-sensors tool, the used evaluation methods, and the relevant findings. Following, we present the proposed guidelines, discuss our study in contrast to previous research, and conclude with the limitations of the current study and topics for future research.

2 RELATED WORK

2.1 Information Visualization Tool Development and Evaluation with Users

Various approaches and guidelines have been proposed that can support the development and evaluation process of information visualization tools with users. An overview of relevant evaluation methods for visualizations is presented in [2], while advice to when to use which method is provided in [5]. The use of qualitative evaluation methods such as observations and interviews can help achieving a richer understanding of the factors that influence visualization development and usage [2, 6]. Carpendale [2] encouraged that more studies in the information visualization field should utilize such methods. However, as in our case, field studies can take advantage of both qualitative and quantitative methods for information visualization evaluation [1].

Lam et al. [7] state that visualization evaluation approach should be based on evaluation goals and questions instead of methods. They provide seven types of evaluation scenarios in the information visualization domain, based on the overview of 850 papers from information visualization literature. Sedlmair et al. [8] propose a nine-stage design study methodology (DSM) and practical guidance for designing visualization systems in collaboration with domain experts. Based on their own experiences and literature review in the fields of human-computer interaction (HCI) and social science, they summarize 32 design study pitfalls to guide the whole process from learning and designing to reporting design studies. Recently, Crisan et al. [9] proposed additions to DSM and practical guidelines for evaluating external constraints, regulatory and organizational, that can affect visualization evaluation with companies.

Several studies report experiences from visualization evaluation in specific context. For instance, Sedlmair et al. [10] list challenges and recommendations for information visualization evaluation based on their experiences from a variety of studies in a large company setting. Saraiya et al. [11] report a long-term study with bioinformaticians to analyze how visualizations are used to gain insights into the data. They emphasize 1) the users' natural motivation to do data analysis and 2) the evaluation of the significance of insights as two essential reasons to evaluate long-term visualization tool usage in a real-world setting. Longitudinal studies enable the inspection of long-term insight generation process and identifying long-term usability problems with data visualization tools [11]. Medler et al. [3] presented their insights from the development of Data Cracker, a visual game analytics tool for supporting game designers. Authors argue that it is beneficial to develop analytic tools in parallel with product development and produce visual prototypes to help the team understand how the tool could be beneficial for them [3]. Additionally, it may be necessary to create functionality for addressing legal issues, such as privacy controls. The team

integration insights include the necessity of involving interdisciplinary teams and the possibility of encountering prejudices towards analytic tools from the product developers. Finally, the role of communication is important in order to anticipate how product changes affect the interpretation of data and to update the developers on the progress of tool development [3].

We applied the Multi-dimensional In-depth Long-term Case Study (MILC) [1] approach suggested for the studies of creativity support tools in [12]. MILC was used to guide the visual data analytics tool development in a long-term study in real use context. Shneiderman and Plaisant [1] propose the use of MILCs to study the efficacy of novel visualization tools not only in terms of their strengths but also to refine the tool iteratively with end users and to produce sufficient evidence to warrant further development. The MILC approach and its derivatives [13] have been used to develop visualization tools for event sequence analysis [14] and in the evaluation of a visual analytics tool in the domain of electronic medical records analysis [15]. MILC was also identified as a relevant approach for the long-term analysis of domain expert use of visual analytics [16]. The evaluations by Wongsuphasawat and colleagues [14] show that the periodic meetings with a domain expert allowed for both the generation of insights and additional questions by the domain expert and guidance for tool development. The study by Stolper and colleagues [15] demonstrated the benefit of the case study approach in documenting insights generated during the long-term use of a visual analytics tool.

2.2 Usage Data Analytics in Manufacturing and Automation Industry

Data analytics and visualization solutions exist on the market for companies to use on their own manufacturing data, such as Bosch's manufacturing analytics solutions [17]. However, little previous research exists in the domain of exploratory user interaction analysis of complex industrial systems, particularly regarding the development of usage data analytics tools. Where many consumer applications are reasonably simple in their operating logic, manufacturing systems have a large number of processes and rules that govern the functioning of the whole. Unlike many consumer-oriented systems, the data that are collected from industrial applications can have significant business value to its producer (i.e., the clients of the system supplier), which puts the onus on developing tools that can generate added value for all stakeholders.

Holzmann et al. [18] studied the acquisition and visualization of user interaction data from a touch screen based robot controller to find cost-efficient solutions for the usability evaluation of handheld terminals in the automation industry. The goal was to help developers to identify possible problems in users' workflow (e.g., navigation problems or unused functions) based on user interface interactions. Navigation path analysis and usage intensity were identified as the most important topics for data logging, based on interviews with a programmer and two project managers in automation industry enterprises.

In another example, Grossauer et al. [19] created a prototype for automation industry to visualize navigation flows through an application. Based on their experiences with applying the visualization tool to multiple datasets, they recommend such tools should include 1) a wide variety of filters and 2) views that

show the whole navigation data and allow the inspection of individual sequences.

We need to learn more about the benefits and challenges related to usage data analytics in manufacturing automation and related industrial contexts. Our study adds to the body of knowledge in this domain with new empirical research done in the context of flexible manufacturing systems.

3 STUDY CONTEXT: FLEXIBLE MANUFACTURING SYSTEMS

Our study was conducted in the context of industrial manufacturing automation systems called flexible manufacturing systems (FMS). Metal industry uses FMS for manufacturing parts by using different metal operations, such as cutting operations (e.g. milling, drilling and boring), metal-forming operations (e.g., rolling, stamping, and welding), and surface-finish operations (e.g., grinding and painting) [20]. In FMS, the production is typically conveyed on pallets [21], on which the parts are attached for machining. FMS enables the automation of the pallet-based machining for manufacturing small batches of different types of products while providing flexibility, as the manufactured product can be changed without changing the whole factory layout [20]. FMS combines software and hardware in order to provide a manufacturing company an easily modifiable, dynamic manufacturing system. Hardware is controlled by software, which in our study manages the production and provides different production optimizations tools, such as fine scheduling.

FMS is normally operated by human operators although robots are also used in some factories. Today, FMS is operated via a combination of graphical user interfaces (e.g., used to enter new manufacturing programs, control the program parameters and modify system status) and physical or software coded buttons on the user interface for pallet control. In our study, the main focus is on user interactions by the human operators on the workshop floor with the FMS elements of the manufacturing system.

The company participating in this study was interested in collecting and analyzing the usage data of their FMS systems after they had been supplied to the customers. Usage data was expected to benefit the company's R&D, customer support, and service business in the future. While the FMS already logged the data of their behavior, it was mainly used for on-demand maintenance. The usage data portraying users' interactions with the system provided an entirely new channel to study the product usage.

4 UX-SENSORS – THE USAGE DATA ANALYTICS FRAMEWORK

In this section, we describe the usage data analytics framework of the UX-sensors tool at the end of the collaborative development period with the FMS company. The framework consists of three components: the data model, which provides an abstraction of case-specific log data; the analytics software framework, which facilitates the storage and analysis of the logged usage data; and the analytics user interface, which allows the interactive exploration of the dataset.

4.1 Data Model

The data model utilized by the framework is based on typical events recorded of human-machine interactions, such as button clicks, data entry, and view changes. The fundamental item of the data model is an event consisting of a timestamp, a feature and a value attributes. Additionally, events can have parameters and context information. Finally, each event has a level specifying whether it is a regular occurrence, a note added via the analysis tool or an error of some level.

The model aims to be generic so that data from different processes and user interfaces can be analyzed. Most of the existing log files can be converted into events in a straightforward manner. In the company data used in our development and evaluations of the tool, a feature refers to parts of the user interface and value to the action executed. Parameters encode additional operation related parameters and context tells about the identity and generic state of the system and user interface at the time.

4.2 Software Architecture

The software architecture of the analytics framework consists of an interactive web application and a set of server components (see Fig. 1). The framework includes a data-import tool for uploading log files to the system, visualization front-end for exploring log data, and server-side components for importing and analyzing the usage log data. The log data and configuration information is stored in CouchDB [22], a NoSQL database that provides an efficient model to handle and query the subsets of massive log data sets. CouchDB is a document storage type database, where each data item, in our case each event, is a JSON document. CouchDB views and lists are used to query the data e.g. by the factory. In addition, a configuration document is used to define tool instance specific properties, such as additional filters, features of interest, user interface structure settings, and color-coding rules for events.

All requests from the web application front end are directed through a proxy server to appropriate back-end components. The proxy takes care of security-related aspects and provides a single server address and port for the web application. It also provides access to a logging service, which stores log data from the web application in a format directly compatible with the tool itself. This logging was used during evaluations.

The server side is mainly built on Node.js [23]. Python is used for the analysis modules and data import. In addition, the analysis modules use R statistical computing framework for extracting frequently occurring sequences from the event data. The front end is built on the Bootstrap front-end framework [24]. Visualizations are built using D3.js [25], a JavaScript library, which allows binding data to the DOM of an HTML document. Crossfilter [26] is utilized for filtering the event data.

The tool was deployed on an external server separate from the customers' flexible manufacturing systems. Log data was uploaded to the analytics tool only by on-demand basis. In the future, it could be beneficial to directly connect the tool to the logging components of the customers' systems.

4.3 UX-Sensors User Interface

The user front end of the developed data analytics tool is an interactive web application. It consists of a data selection view

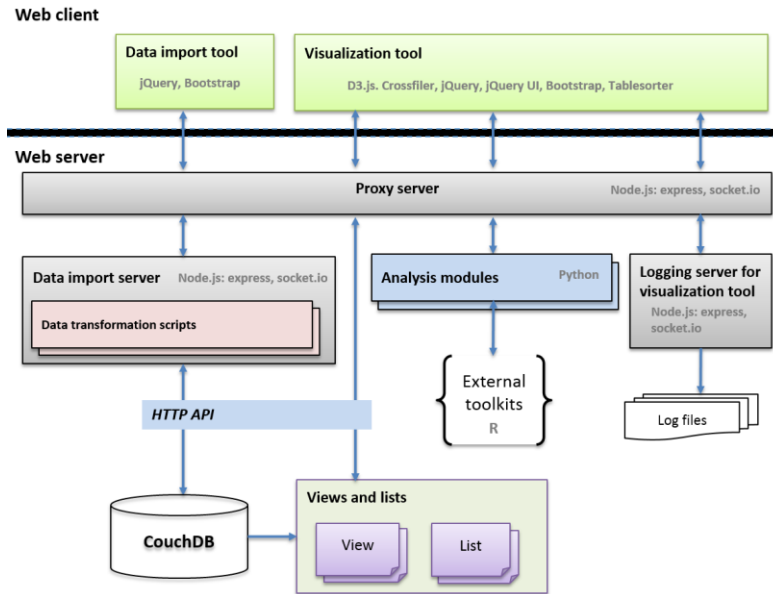


Figure 1: Software architecture of the developed UX-sensors tool.

and the main data browsing view with timelines and analysis tools. The main data browsing view consists of six main elements that are highlighted and numbered in Fig. 2. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) the main filters.

The overview panel contains general information about the selected observation window, e.g. the number of found events, most frequent events by value, the average usage session length, the average number of operations per session, and the number of error events. Events are split into sessions based on the maximum time duration between successive events. The default value is five minutes and it can be changed in settings. The overview timeline displays the overall number of events by the hour and works as a filter where the user can restrict the further analysis to a shorter observation window. The detailed timeline displays the individual events within the observation window. By hovering over an event, detailed information is displayed. The user can also add additional notes directly to the timeline, for example, to record and share insights gained from the data. The additional filters element can filter the event set further. In the setup used in the field study, six main filters were available: factory, observation window (if multiple windows selected), system, user, level, and feature.

The tab elements display processed information about the selected events and provide tools for further exploration of the data. The used features tab displays a line graph of feature use over time and a complete list of features and feature-value pairs with a count and percentage of the total events. A feature represents a part of the FMS software and a value the action user has performed, for example, "Inventory - Release Pallet". The data table can be sorted by each column and filtered by search. The errors and recovery tab provides a list of the most common errors, average recovery time and user interaction sequences during error recovery. Error recovery time is estimated as the duration between the last successive error event and the first following user interaction (i.e., event that is not an error or warning). The frequent sequences tab is used to calculate and display the most frequently occurring sequences. This is done by

splitting the events into sessions and then looking for similarities in the event sequences within the sessions. Through the search tab, single events or sequences of events can be searched by defining key-value pairs consisting of e.g., feature and value. The data entry tab is for exploring events indicating user data entry and user interaction sequences during data entry. Lastly, the main filters element can contain up to two filter panels on the right side of the timelines.

5 UX-SENSORS – DEVELOPMENT AND EVALUATION

In this section, the process for developing and evaluating the UX-sensors tool in collaboration with stakeholders from the FMS supplier company is presented. Then the used evaluation methods and their relevant results are described.

5.1 UX-Sensors Development and Evaluation Process

Fig. 3 illustrates the development process for UX-sensors, consisting of requirements gathering phase and iterative development and evaluation phase. Next, the development process is described in more detail.

Workshop to identify company needs. As a part of a larger academic research project with three companies from metals and engineering industry, we held a workshop to identify the company needs on usage data analytics and visualization in this domain. The identified requirements for data types were: 1) usage combinations, such as the customer's production type and in what mode they use the system, 2) patterns of use, 3) types of user groups and profiles that can be found, 4) summarizations of the system use based on logged data (the logs of events, system status, user actions, interactions etc.), 5) identifying problem or fault situations (individual and possible patterns), and 6) changes in use (such as features) over weeks or months such as how the taking of the system into use and learning to use is progressing.

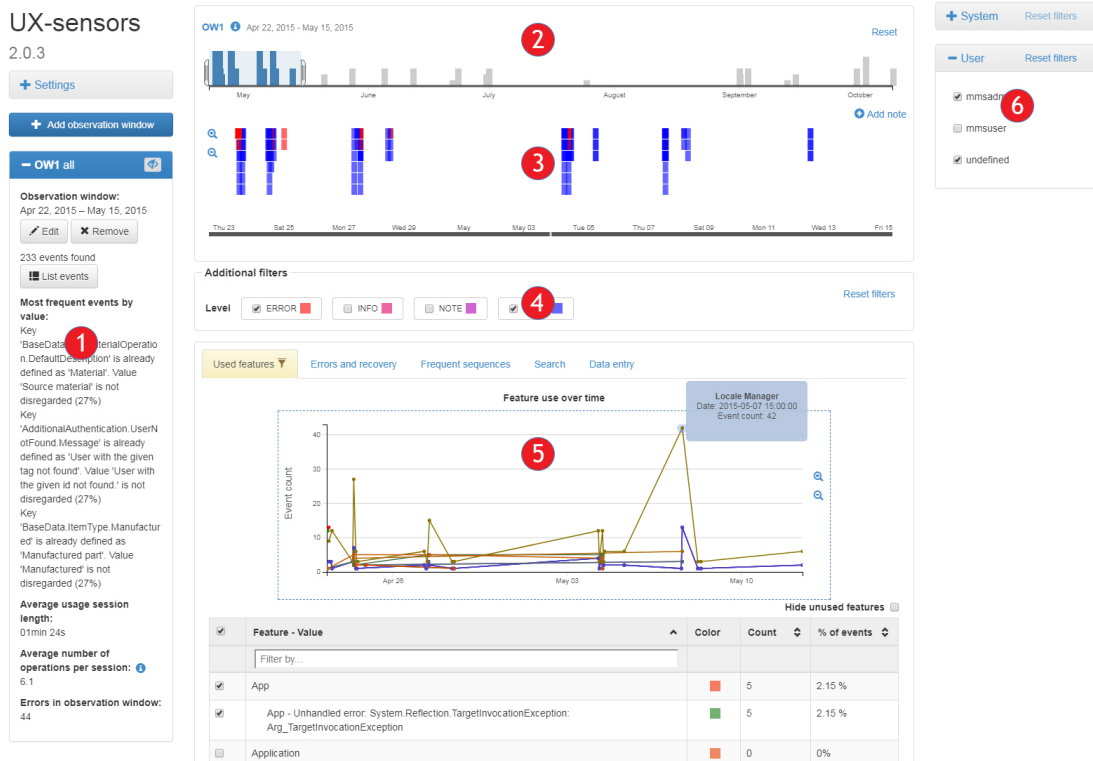


Figure 2: The main data browsing and analysis view of the UX-sensors tool. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) main filters

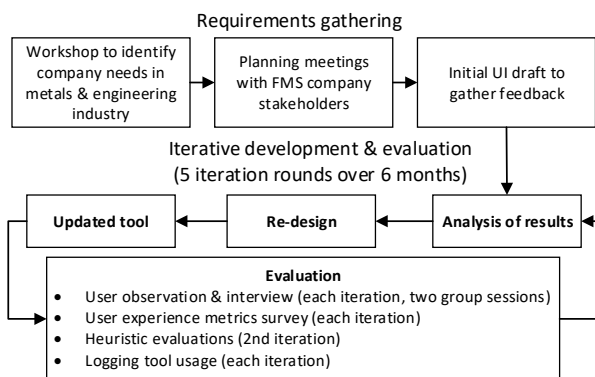


Figure 3: Summary of the development process for the UX-sensors tool.

to identify issues needing support or training and whether problems or faults appear over time.

Planning meetings with FMS company. After the workshop, planning continued with stakeholders from one flexible manufacturing systems (FMS) supplier company on the development of a usage data analytics tool. In discussions with stakeholders from R&D, software development and customer support, requirements for the data analytics tool and its features were gathered. The same people participated in the iterative development and evaluation process (described later) of the tool. Instead of utilizing an off-the-shelf data analytics tool, we decided to develop our own visual analytics framework. Given

the varying needs of the stakeholder companies, a custom framework was expected to expedite the development process, over learning and customizing an existing tool, and provide the development team experience in the design and development of visual analytics tools for supporting other projects.

Initial UI draft. Next, an initial user interface draft of UX-sensors was presented to stakeholders to spark more conversations and to gather feedback on the proposed UI design. This feedback was utilized in designing the first interactive version of the tool to be used in the iterative development and evaluation cycles.

Iterative development and evaluation. During the following six-month study period, including a one month holiday season, we iteratively developed and evaluated the tool in collaboration with practitioners from the FMS supplier company. Two of the authors were responsible for the user studies and reporting their findings to three researchers responsible for the software development of the UX-sensors tool.

Our evaluation approach to UX-sensors was based on the MILC approach [1], which has inspired the implementation of several long-term studies where the use of data analytics and visualization tools have been evaluated in real use context [13, 14, 15, 16]. The MILC approach requires, at a minimum, 3-5 domain experts who are available for a period of weeks to months, and a tool that provides sufficient, problem-free basic functionality for users [1]. The long-term process requires the documentation of current tools and practices, establishing evaluation criteria, a schedule of user research, instrumenting the tool for collecting log data, providing training, and modifying the tool as needed [1].

Six practitioners participated in the evaluations: three software developers (including a team leader), a leader of a remote customer support team, a product manager for product life cycle services and a director managing research and innovation development. One developer working for a subcontractor left the study after the second iteration round. From these participants, the developers and the leader of the customer support team were identified as the lead users of the UX-sensors tool as they had experience of inspecting log data from their customers' FMS when challenging error events occurred. However, including the product manager and the director in the development process was expected to generate more diverse set of ideas for utilization of the tool and logged usage data.

The iterative development and evaluation phase started when the first interactive prototype of the UX-sensors tool was introduced to the company personnel in a training workshop. In the workshop, all six participants could inspect a usage data set with the tool and give feedback from the user interface. During the following six months, four more iterations of the tool were introduced to the participants. User feedback was collected after each iteration of the UX-sensors tool. Email reminders were sent after each update to encourage participants to try out the tool. We organized two group meetings, including the first training workshop, and three sets of single user observation sessions. After each meeting, a link to a web survey was sent to the participants in an email. After the second iteration, five external evaluators conducted heuristic evaluations of the prototype. Finally, log data was collected from the UX-sensors tool for following its usage over the whole development period. The next section summarizes the used methods and the main findings.

At the end of the development process, a data import tool was implemented to allow company practitioners to import usage data logs to the UX-sensors tool. We anticipated that practitioners would use UX-sensors tool to inspect logged usage data in the near future. However, when inquired after six months, we learned that the stakeholders still worked on challenges related to the legal issues considering the data ownership, privacy and security. When collecting data from customers' employees working with the system in different countries, the supplier has to carefully follow the local data collection policies and make agreements with each customer regarding data usage.

5.2 Used Evaluation Methods and Main Findings

User observation and interviews. In the observation sessions, participants could freely use the tool for exploring the available usage data and try any new features, while the researcher encouraged the participant to think aloud with questions such as "what do you think of this feature?" The session ended with an interview, where researchers could ask feedback from specific features, confirm their observations during the session, and inquire if participants had received any insights from the data. Each session was recorded with a video camera and lasted approximately one hour. All sessions were arranged in the

company's meeting rooms. The observation and interview data were analyzed by coding similar findings or responses to descriptive categories, with comments related to the development needs of the data analytics tool separated from other topics. The comments were grouped based on the features or aspects of the tool that they referred and then reported to the analytics tool developers.

The observation sessions provided information regarding usability issues, suggestions for new features and insights that participants got from the usage data. For example, the customer support representative asked for adding references to the system generated error codes in the log data events and support for exporting the tables or lists of the analyzed data for modifying the data with other tools for creating reports. Developers were interested in acquiring more details regarding logged error events such as a direct reference to the line in the code. One of the developers also proposed how the future version of the tool could function for importing log data files collected from different customers. The concept of event sequences was challenging for most participants and therefore tooltip help texts and an explanation of how the sequences are calculated were added to the UI. Furthermore, the content of the original sequences tab was divided into frequent sequences tab and search tab to clarify the UI.

One insight from the usage data that generated much discussion among the participants related to how the autopilot feature in the FMS was operated. The usage data proposed that users did not follow the shortest route in the UI to activate the feature, suggesting a need for changes in the UI and/or in the user training process. Over the following interviews, we learned that developers had considered whether certain actions should not require that the autopilot is activated, as changing its state very often requires user effort. Therefore, this insight from the usage data may result in some changes in the UI in the future.

As a general observation, we learned that developers rarely had possibilities for gathering feedback on how the FMS are used on a daily basis after they have been installed in the customer's factory. Logged usage data was seen as a relevant channel for supporting developers' understanding about the end-users and their ways of using the system, especially over longer periods. Preliminary knowledge of the customer's ways of using the system and the common errors could significantly help focusing site visits on customer's factory, where support can be provided and more qualitative data gathered to understand the reasons behind the findings from log data.

User experience metrics survey. Repeated web surveys aimed to capture whether the user experience (UX) with the tool changed over time. Aiming to make the repeated evaluations less taxing for the participants, we limited the number of different measured UX factors and used only a 4-point Likert scale. Table 1 presents the questions and results from the first and the last (fifth) UX survey, including three questions adapted from the After Scenario Questionnaire (ASQ) by Lewis [27]. The director and one developer did not answer the last survey, hence only the

Table 1: Evaluations of the first and fifth iteration of the UX-sensors tool (n=4). Scale from 1 to 4, where 1 = Completely disagree, 2 = Somewhat disagree, 3 = Somewhat agree and 4 = Completely agree.

This tool is...	Easy to learn		Useable		Flexible in its interaction		Pleasant to use		Useful		ASQ: I am satisfied with the tool's...	Ease of use		Amount of time it takes to complete tasks		Support information (help, documentation) during usage	
	1st	5th	1st	5th	1st	5th	1st	5th	1st	5th		1st	5th	1st	5th	1st	5th
SURVEY	3	2,75	2,75	2,25	2,75	2,75	2,25	2,25	3	2,50		2,25	2,50	2,50	2,50	2,50	3,00
MEAN												0,50	0,58	0,58	0,58	0,58	0,00
SD	0,82	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,82	0,58		0,50	0,58	0,58	0,58	0,58	0,00

other four respondents are included (n=4). Interestingly, while the respondents got more satisfied with the ease of use and support information with the tool, they considered it less useful and harder to learn. One possible reason for this is that each iteration, also the fifth, presented UI changes or new features, and participants had to familiarize themselves with these new features. Also, the novelty value of the tool may have decreased over time.

Heuristic evaluations (HE). We utilized the top ten heuristics for information visualization with the widest explanatory coverage proposed by the study by Forssell et al. [28]. The HE tasks consisted of 1) exploring the 2nd iteration of the UX-sensors tool and its features, 2) identifying usability issues and describing them in free text, 3) identifying the heuristics that were violated, 4) assessing the severity rating of the finding based on Nielsen's rating scale [29], and 5) assessment of how well the heuristic explained the finding [30].

Five external evaluators (three female) took part in the HEs. Evaluators' experience in the field of HCI (either studies or research work) varied from 1 month to 2.5 years. Two had no work experience in the field of data visualization while others' experience varied from 3 months to 2.5 years. None of the evaluators had experience in the application domain of the FMS. Three evaluators had previous experience from expert evaluations of interactive software.

The HEs resulted in 99 different problems or suggestions for improvements that at least one of the evaluators reported. These findings were reported to the developers of the UX-sensors tool and used in updating the tool for the following iterations. "Information coding" (30 references) and "orientation and help" (22) were two of the most often violated heuristics. Since the evaluators were not familiar with the actual usage data represented, they focused on the UI and visualization related issues. For this purpose, the information visualization heuristics [28] seemed to be well-suited, as several comments related to the used graphs and charts, such as color coding, axis information, and zoom functions. From the observation sessions done during the same iteration, we identified 68 different problems or suggestions for improvements in total. Interestingly, only 14 problems were identified with both methods, for example, lack of help texts for features and unfamiliar terms, color coding issues, not listing events in a table format, and saving the previously conducted searches. In contrast to the HEs, the findings from the observation sessions reflected the requirements that the employees had for doing their work tasks, including specific types of data, features, and visualizations.

Logging tool usage. Log data from UX-sensors was used for following how actively participants used the tool between the observation sessions. This prodded us to discuss with the less active participants what could motivate them to utilize the tool more often. While access to more real usage data from different customers' systems was hoped, it was also evident that the learning curve was steeper for those participants who were not accustomed to working with "raw" log data from FMS.

In conclusion, considering all the methods we utilized, user observations, interviews and heuristic evaluations provided the most useful feedback for improving the tool. Repeated survey questions provided feedback of the tool's UX over time, but discussions with the participants resulted in more practical insights regarding how the tool was used. Finally, logging the

usage of the UX-sensors tool itself was an easy way in inspecting how the tool was used outside the observation sessions.

6 GUIDELINES FOR DEVELOPING AND EVALUATING USAGE DATA ANALYTICS TOOLS

In the following, nine guidelines are presented based on the insights during the development of the UX-sensors tool in collaboration with practitioners from the FMS supplier company.

1. Gather an Interdisciplinary Team to Support the Development Process. We confirm the experiences from other domains [3] in that an interdisciplinary team can greatly support the development process of the analytics tool in the context of automated manufacturing industry. The company employees who participated in the development project had different analytic needs and requirements regarding the collected usage data. For example, developers and customer support personnel were more interested in details related to specific error situations, while manager-level personnel often discussed the more general usage data types of their customers' systems. We presume that including stakeholders from marketing, sales, and user training as well as customers' representatives could provide even more insights from the possibilities and challenges of usage data logging, as these roles came up in discussions with the current participants.

Although gathering representatives from various areas in the company requires effort, it can pay off in the ideas of new features for the developed tool or new ways to utilize the gathered log data to provide value for the company and the customer. Furthermore, participating in the development and evaluation of the data analytics tool can benefit the collaborating company by improving the basic data literacy skills of the employees [3]. For instance, during our group discussion sessions, the participants became more aware of the analytics needs of their colleagues and the way the tool should be developed was discussed based on these user requirements.

2. Ensure Early Access to Real Logged Usage Data. One of the key issues in the development process was acquiring sample log files to support the testing and demonstration of the tool. Data gathering is prone to delays that may jeopardize the whole project [8]. One option is to use synthetic data that at least allows the inspection of the functionality of the analytics tool and concrete discussions with stakeholders while tool developers are waiting for access to real data [9]. In our case, we utilized data from a local FMS training environment. Since customer data is usually confidential, getting access to the log files that are gathered from customers located in different countries requires familiarization with the local rules regarding data logging and usage. However, we learned that these challenges generated beneficial discussions among stakeholders, such as how to provide value for the customer in exchange for the data collected from their factory.

We emphasize that access to real logged usage data should be secured as early as possible in the analytics tool development process. Resources should be allocated to building trust and showing the value that the customer can get from sharing the logged usage data with the supplier. This could include reports on how customer's different teams use the system over time and suggestions for additional user training. However, privacy, security, and intellectual property issues need to be solved [31]

before accessing usage data. When access to real usage data is negotiated in advance, the developed tool can be tested more efficiently and any disputes over the data access are minimized in the future. Finally, as the goal is to generate new insights from the data, it is vital that the data are meaningful to the users inspecting them. Lack of interesting data can affect stakeholders' motivation to explore the tool on their own and participate in the evaluation activities.

3. Identify Other Data Types That Can Support Usage Data Analytics. During the requirements gathering process, analytics tool developers should identify what other contextual data could support users in analyzing logged usage data and consider whether these data can be visualized with the same tool. For example, we learned that the developers and remote customer support personnel occasionally had to inspect several log files when sourcing error situations between different log data sources. We, therefore, identified the need for viewing human-machine interaction events and events generated by different digital system services on the same timeline to support the sourcing of error events. Although not implemented in the current version of the UX-sensors, this feature could improve the current process of searching correct timestamps and manually switching between different text log files when following the chain of events.

4. Allocate Resources to Explore the Log Data Structure Prior to Data Wrangling. An important aspect of the development process is clearly communicating the structure of the logged data if different teams work on the logging services and the analytics tool development. The best-case scenario would be working together with the programmers of the automation system and agree on what should be logged after deciding which data is needed. In our case, the analytics tool developers were not familiar with the logging services. It took us considerable effort to understand the structure and meaning of the log data and map it to meaningful labels and functionality in the visualization. The familiarization process required close collaboration with an FMS development team member familiar with the logging procedure.

We also recommend mapping out potential 'edge cases' (e.g., log file types or log entries that differ in formatting from others) to avoid unnecessary troubleshooting. Moreover, subsequent changes to logging services should be made in a way that does not change the log format, to avoid additional work on data wrangling. If changes are unavoidable, care should be taken to work with the analytics tool developers to limit the scope of required changes.

5. Establish Coverage of Logging and Compatibility with the Visualization Tool. With complex industrial systems, there can be multiple ways to log events at the system and organization level, and such log files can be stored in different locations. In our case, we utilized log data from one part of the FMS system, which was not enough to implement all the planned features for the visualization tool. For example, it was discussed that teleservice log files should be incorporated into the data visualization tool, but these files were not made available during the study. Thus, in addition to understanding what the log files contain (guideline 4), it is important to establish which log files are required to fully address the design requirements.

Customers' manufacturing systems that are older may log data differently. This means that not all desired log data may be available from all systems and in the same format. Therefore,

support for all different kinds of log events may be difficult to implement and it should be decided what kind of logging the developed visualization tool should primarily support.

6. Combine Expert Evaluation and Field Study Methods to Include Different Viewpoints. User studies with practitioners helped us to understand their goals and requirements regarding the usage data. In addition, heuristic evaluations [28] by external evaluators supplemented user observations in the early stages of the iterative development process by providing a good summary of general usability problems related to user interface and data visualizations. Although MILC approach [1] does not mention expert evaluations, previous studies [10, 32] have found heuristic evaluations done by external HCI experts to be useful in identifying usability issues when developing data visualization tools. While it can be challenging to find HCI experts who are also familiar with the specific domain, such as manufacturing automation, hiring students with HCI or visualization background can be a viable option [32]. Stakeholders from the company could also act as evaluators, but it may be challenging to motivate them to invest time in learning the evaluation process and conducting the evaluations.

7. Collect Log Data of the Analytics Tool to Follow Its Usage. It can be convenient if tool developers use the tool itself to analyze log data collected from its usage. In UX-sensors, the logging mechanism was designed to store data in a format compatible with the system, so that we could use it internally to support the evaluation activities. The log data reveals how actively the participants really use the analytics tool, without a need to disturb company employees with questions regarding the tool usage. This information can be used to motivate the participants and plan interventions if needed. Logged usage data from the tool provides information on how different features are used over time, especially outside observation sessions. Finally, log data provides information about how the tool is used after the collaborative development period, revealing its real applicability to the company over time.

8. Provide Support for Users with Varying Analytics Skills. Interactive data analytics tools should support users who are less familiar with programming and analytics [33]. Although the need for help texts and support for the learning process was highlighted in our heuristic evaluation results, we argue that the company personnel with less experience in data analytics will also benefit from these improvements. Furthermore, stakeholders who do not actively participate in the development process of the tool, but who might use it in the future, are likely to find instructions designed for novice users helpful. Finally, presenting the generally most interesting data first in the UI is recommended. In our case, this meant the frequencies of used features and error events, which were the first tabs in the main data browsing and analysis view of the tool.

9. Support the Sharing of Insights. The key principles of developing creativity support tools include support for collaboration and open interchange [12]. While stakeholders have their own channels for communication, visualization tool developers can support the sharing of insights during group discussions and by implementing features into the tool that support information sharing. For example, we allowed users to add notes to the usage data timeline, with the aim that others inspecting the same data could view these comments. The sharing of findings can also be supported by allowing an easy exporting of data tables and visualization images from the data analytics tool.

7 DISCUSSION

We developed UX-sensors, a visual data analytics tool for logged usage data, in collaboration with a FMS supplier company, aiming to support their R&D, customer support and service business activities in the future. Based on the lessons learned during this study, we provided guidelines to support other researchers and practitioners developing visual data analytics tools in similar contexts.

We learned that developers in the collaborating company rarely had opportunities for gathering feedback on how the FMS are used on a daily basis after they have been installed in the customer's factory. UX-sensors was seen as a potentially helpful tool in accessing the more systematic data of FMS usage over a longer time, such as months or years. Usage data was expected to provide insights on how users are using the FMS and then guide the qualitative research and on-site interventions to study why users use FMS in a specific way.

Developing data analytics and visualization tools can be a challenging process when the tool development is done separately from the development of the underlying automation system and its data logging services, as in our case. Without a deep understanding of the logging process, close collaboration was required with the developers of the manufacturing system during data wrangling. On one occasion, the collaborating company had to update the logging capabilities of their FMS, which required us to update the UX-sensors tool as well. Similar findings were reported during the development of a visual game analytics tool [3], where anticipating the effects of game design changes was presented as one guideline. Steady communication, such as participating in the development team's weekly meetings, could keep analytics tool developers informed if any changes are planned to the logging capabilities of the manufacturing system.

Our experiences from the iterative development process where we applied and extended the MILC [1] approach with expert evaluations were generally positive. The used evaluation methods provided us with meaningful data to support the UI development of UX-sensors. Interestingly, while the user observations and interviews provided the most important findings for the tool development, the HEs in the early phases of the development suggested various improvement ideas for the tool UI that were not identified with the participants working in the FMS context. In comparison to [3], we did not face significant prejudice against usage data logging among the stakeholders. The participants appeared to be genuinely interested in the possibilities of usage data logging, and during the long study period, they actively participated in the organized sessions and answered repeated web surveys. The user experience metrics survey provided useful comparison points over time regarding how the UX of the tool was evaluated. However, choosing a four-point Likert scale for the sake of effortlessness for the participants seemed unnecessary, as scale with e.g. five or seven points would have provided more detailed results regarding how the participants evaluated the UX of the tool over time.

The main insight that participants gained from the logged usage data of the FMS revealed unexpected navigation paths used in the UI when users activated the autopilot feature, suggesting a

need for updating the UI or offering training for users. While such insights can benefit UI designers and end-user instructors, the actual benefits of usage data logging for other stakeholders remained unexplored. For example, would the logged usage data help customer support and developers in working out complex error events, and would customers benefit from annual reports regarding their FMS usage? Furthermore, before new services such as training offerings on a personal level can be realized, legal issues related to user privacy and data ownership need to be carefully settled with each customer.

Two insights gained during the development and evaluation process are likely to be of interest beyond the complex industrial systems context. First, the potential business value of the captured usage data for the end-user organizations of the system incorporating logging could be a way to create buy-in with customers. Second, the captured usage data could be used to develop additional services that enhance the end-user organization's use of the system, such as targeted training.

Limitations and future research. Our results are limited due to the lack coverage of different practitioner roles from one company. This is a common limitation to evaluation studies on the information visualization domain, where long-term involvement and motivation are required from participants. Future studies should focus on what kind of insights or benefits real logged usage data can provide for stakeholders also in other roles such as marketing, sales, and user training, and study the customers' viewpoints regarding the value gained from usage data logging.

Lam et al. [7] emphasize the growing need, referring to [34], for studying the design context for visualization tools, including work environments, users' tasks, and work practices. In the spirit of this notion and as a continuation of the work started here, we propose that future studies explore how utilizing usage data analytics tools can support current work practices in manufacturing automation organizations. The following research questions for future studies are proposed: What kind of and how significant insights can usage data logging provide in manufacturing automation context, and for whom? How are these insights shared in the organization and what is their impact over time, for example, resulting in changes in the manufacturing system UI or innovations that support product development, customer support or service business?

Given the limitations on access to both participants and real-world log data, collaborative case studies in the real context of use are the preferred way to provide a better understanding about the benefits of usage data logging in practice. This knowledge can help researchers and tool developers in designing data analytics and visualization tools with positive user experience and providing instructions that can support practitioners in utilizing insights from logged usage data in their work.

8 CONCLUSIONS

We have presented our iterative development and evaluation process for developing UX-sensors, a visual data analytics tool for logged usage data, in collaboration with a flexible manufacturing systems supplier company. We have summarized our experiences from the development and evaluation process as guidelines to support other researchers and practitioners developing usage data analytics tools for complex industrial

systems. Finally, we have proposed research questions to further study the insights from logged usage data and their utilization in manufacturing automation context.

Our goal was that the developed tool would support stakeholders in the company with the generation of new insights from the logged usage data of their customer's FMS. The insights from usage data and discussions with stakeholders proposed that logged usage data analysis can potentially support UI and service development. Logged usage data was expected to provide insights on how users are using the FMS and guide the more qualitative research to study why users use FMS in a specific way.

ACKNOWLEDGMENTS

This study was supported by The Finnish Funding Agency for Innovation (TEKES), currently known as Business Finland, as a part of User Experience and Usability in Complex Systems (UXUS) research programme.

REFERENCES

- [1] B. Shneiderman and C. Plaisant. 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV '06). ACM, New York, NY, USA, 1–7.
- [2] S. Carpendale. 2008. Evaluating Information Visualizations. *Information Visualization: Human-Centered Issues and Perspectives*, 4950/2008, 19–45.
- [3] B. Medler, M. John, and J. Lane. 2011. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 2365–2374.
- [4] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. 2013. A Systematic Review on the Practice of Evaluating Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2818–2827. <http://doi.org/10.1109/TVCG.2013.126>
- [5] T. Munzner. 2009. A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 921–928.
- [6] M. Q. Patton. 2001. *Qualitative Research and Evaluation Methods*, 3rd edn. Sage Publications, London.
- [7] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers, 2012, 18 (9), 1520–1536.
- [8] M. Sedlmair, M. Meyer, and T. Munzner. 2012. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2431–2440. <http://doi.org/10.1109/TVCG.2012.213>
- [9] Crisan, J.L. Gardy, and T. Munzner. 2016. On Regulatory and Organizational Constraints in Visualization Design and Evaluation. In M. Sedlmair, P. Isenberg, T. Isenberg, N. Mahyar, H. Lam, (Eds.) Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16). ACM, New York, NY, USA, 1–9.
- [10] M. Sedlmair, P. Isenberg, D. Baur, and A. Butz. 2011. Information Visualization Evaluation in Large Companies: Challenges, Experiences and Recommendations. *Information Visualization Journal*, 10(3), 248–266. <http://doi.org/10.1177/1473871611413099>
- [11] P. Saraiya, C. North, V. Lam, and K. Duca. 2006. An Insight-Based Longitudinal Study of Visual Analytics. *IEEE Trans. Visualization and Computer Graphics*, 12(6), 1511–1522.
- [12] B. Shneiderman, G. Fischer, M. Czerwinski, M. Resnick, B. Myers and 13 others. 2006. Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human-Computer Interaction* 20(2), 61–77.
- [13] Perer and B. Shneiderman. 2008. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 265–274.
- [14] K. Wongsuphasawat, J.A. Guerra Gómez, C. Plaisant, T.D. Wang, M. Taieb-Maimon, and B. Shneiderman. 2011. LifeFlow: visualizing an overview of event sequences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 1747–1756.
- [15] C.D. Stolper, A. Perer, and D. Gotz. 2014. Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1653–1662.
- [16] D. Gotz and H. Stavropoulos. 2014. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1783–1792.
- [17] Bosch Software Innovations, <https://www.bosch-si.com/solutions/manufacturing/data-analytics/manufacturing-analytics.html>
- [18] C. Holzmann, F. Lettner, C. Grossauer, W. Wetzlinger, P. Latzelsperger, C. Augdopler. 2014. Logging and Visualization of Touch Interactions on Teach Pendants. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14). ACM, New York, NY, USA, 619–624.
- [19] C. Grossauer, C. Holzmann, D. Steiner, and A. Guetz. 2015. Interaction visualization and analysis in automation industry. In Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15). ACM, New York, NY, USA, 407–411.
- [20] T. Sheridan. 2002. *Human and automation: system design and research issues*. John Wiley & Sons, Inc., publication, Santa Monica, USA.
- [21] H.A. ElMaraghy. 2006. Flexible and reconfigurable manufacturing systems paradigms. *Int J Flex Manuf Sys*, 17(4), 261–276.
- [22] CouchDB, <http://couchdb.apache.org/>
- [23] Node.js, <https://nodejs.org>
- [24] Bootstrap, <http://getbootstrap.com/>
- [25] D3 Data-Driven Documents, <http://d3js.org/>
- [26] Crossfilter, <http://square.github.io/crossfilter/>
- [27] J.R. Lewis. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interact.*, 7(1), 57–78.
- [28] C. Forsell and J. Johansson. 2010. An heuristic set for evaluation in information visualization. In Santucci, G. (Ed.) Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10). ACM, New York, NY, USA, 199–206.
- [29] J. Nielsen. 1993. *Usability Engineering*. Academic Press, USA.
- [30] J. Nielsen. 1994. Heuristic evaluation. In Nielsen, J., Mack, R.L. (Eds.). *Usability Inspection Methods*. John Wiley & Sons, NY, USA, 25–61.
- [31] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A.H. Byers. 2011. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- [32] M. Tory, and T. Möller. 2005. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5), 8–11. <http://doi.org/10.1109/MCG.2005.102>
- [33] J. Heer, and S. Kandel. 2012. Interactive analysis of big data. *XRDS*, 19(1), 50–54.
- [34] C. Plaisant. 2004. The challenge of information visualization evaluation. In Proceedings of the Working Conference on Advanced Visual Interfaces (AVI). ACM, New York, USA, 109–116.
- [35] E. Carroll, C. Latulipe, R. Fung, M. Terry, and D.R. Cheriton. 2009. Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. Proceedings of the Seventh ACM Conference on Creativity and Cognition (C&C '09). 127–136. doi:10.1145/1640233.1640255