



Social Collaboration Analytics Framework: A framework for providing business intelligence on collaboration in the digital workplace

Florian Schwade

University of Koblenz-Landau, Germany

ARTICLE INFO

Keywords:

Social collaboration analytics
Business intelligence
Analytics
Enterprise collaboration systems
Design science
Mixed method

ABSTRACT

Enterprise collaboration systems (ECS) have become the core of the digital workplace in many organizations. More and more companies have introduced this new business software for supporting computer-mediated collaboration among employees. The emergence of ECS in the technical and strategic landscape of companies leads to new challenges in business intelligence and reporting. In contrast to traditional business software such as ERP systems, there are no established methods for decision-making in ECS projects. The collaboration professionals responsible for managing the ECS require insights into the use of an ECS to understand how employees use these systems. By developing the Social Collaboration Analytics Framework, this work provides a framework for establishing the analysis of collaboration activities in the digital workplace (social collaboration analytics) as part of organizations' business intelligence. The Social Collaboration Analytics Framework consists of distinct phases and guides the analysis of ECS data. It includes working steps, recommendations, and guiding questions developed based on the findings from longitudinal research in a university-industry collaboration context and a comparison of established data mining process models.

1. Introduction

In recent years, business intelligence (BI) gained significant momentum in practice and IS research. BI supports organizational performance management, reporting, and decision making [1]. Enterprise resource planning systems are the primary data source for BI because they integrate all functional areas of a company and thus capture process-oriented business transactions [2]. However, not all work in a company is process-oriented. Many aspects of business activities are organized in projects or tasks which require ad-hoc collaboration. Employees spend up to one-third of their work time on projects [3]. This change in the nature of work led to the emergence of a new kind of business software called enterprise collaboration systems (ECS). Typical examples are HCL Connections or Microsoft Sharepoint. ECS combine features known from social media such as social profiles, wikis, blogs, and forums with traditional groupware functions such as shared calendars, files, and tasks [4]. They support computer-mediated collaboration and communication in often dispersed teams and are used for building knowledge repositories [5]. Consequently, ECS have emerged as the core of the digital workplace in many companies [6]. The radical shift to virtual and remote work imposed by the COVID-19 restrictions caused a rapidly increasing appropriation of collaboration software [7]. As ECS

store digital trace data in the form of user-generated content and event logs, they are unprecedented data sources for analyzing and understanding collaboration in the digital workplace. Two recent literature reviews identified the emergence of a research stream on analyzing ECS data [8,9], which demonstrates increasing efforts in academia and practice to derive insights on collaboration from ECS data. This research stream is also referred to as social collaboration analytics (SCA), a term that refers to analyzing, visualizing, and interpreting collaboration activities in ECS [4].

Because of the increasing importance of ECS, many companies assign collaboration professionals or collaboration teams for managing the ECS [10]. Due to the crucial role of ECS in the digital workplace and the limited analytics capabilities of ECS, collaboration professionals require better methods and metrics for reporting on the project progress and the use and effects of the ECS [11]. In contrast to other business software, there is no established reporting supporting decision-making for ECS [9,12]. While research on decision support and decision support systems (DSS) in computer supported cooperative work (CSCW) is not new, existing research focuses on group decision support systems for decision making and workflows at the team level [13,14]. By enabling reporting and decision making for the digital workplace on the organizational level, social collaboration analytics introduces a new perspective to DSS

E-mail address: fschwade@uni-koblenz.de.

<https://doi.org/10.1016/j.dss.2021.113587>

Received 20 July 2020; Received in revised form 28 March 2021; Accepted 4 May 2021

Available online 6 May 2021

0167-9236/© 2021 Elsevier B.V. All rights reserved.

in CSCW systems.

As discussed in Section 2.2, existing SCA studies do not provide sufficient guidance for analyzing ECS data because essential details about the selected approaches are missing. Consequently, collaboration professionals cannot get the required information on ECS use necessary for reporting, decision-making, and ECS management [15]. Accordingly, academics do not have sound methods for studying collaboration in the digital workplace based on log file analysis. To address the outlined limitations and to be able to leverage this new data source for BI, this article seeks to answer the following research questions: (1) What are the current practices for conducting social collaboration analytics? (2) Who are the stakeholders for SCA, and what are their information needs? (3) How can collaboration professionals be supported in gaining business intelligence on collaboration activities in the digital workplace?

RQ 1 aims to identify current practices and approaches for SCA in academia and practice. RQ 2 seeks to identify the stakeholders of ECS projects who are possible targets for reports on ECS use. As part of RQ 3, it is investigated how SCA can support collaboration professionals in gaining business intelligence on collaboration in the digital workplace. RQ 3 builds on the findings from RQ 1 and RQ 2 and contributes the Social Collaboration Analytics Framework (SCAF). The framework is designed to be applicable in practice. It provides guidelines for SCA in the form of phases, work steps, and methodological recommendations to derive interpretable and explainable insights on collaboration in ECS based on the analysis of ECS data. Finally, the framework is evaluated by analyzing ECS data from a selected key case company to demonstrate its application and show how insights on collaboration can be gained.

This article reports on the findings from research that has been ongoing for more than five years. The research has been carried out in the context of the IndustryConnect initiative [10], a university-industry research project [16] founded and guided by a group of University researchers. More than 40 ECS user companies of different sizes and industries engage in a mutual exchange on diverse topics related to the digital workplace. The participating companies have in common that they all use the same ECS (HCL Connections). The companies' representatives are collaboration professionals from diverse departments such as information technology, knowledge management, or internal communication. In the physical bi-annual IndustryConnect workshops, research outcomes are evaluated with practitioners. Through the open and bilateral dialogue with practitioners, as called for by Te'eni et al. [17], a research contribution that is impactful to practitioners can be contributed. By evaluating the SCA Framework together with the case company, this article follows the call by Nunamaker et al. [18] [] to "shepherd an information systems solution through the last research mile."

2. Social collaboration analytics: foundations and literature

This section outlines the foundations of social collaboration analytics and presents findings from a literature review in which existing challenges and approaches for SCA were identified.

2.1. Social collaboration analytics

Social collaboration analytics (SCA) is defined as "the approach for analyzing and displaying collaboration activity of users in socially-enabled collaboration systems" [4, p. 402]. SCA is distinguished from related approaches such as enterprise social network analysis [19] or enterprise social media analysis [20], which are based on social network analysis, by focussing on analyzing collaboration activities. The primary data sources for SCA are organizational data, content data, and transactional data [4].

Organizational data describes the organizational structure, including the departments and the organization's hierarchy and the user accounts. It is typically stored in directory services (e.g. LDAP). Content data [21]

refers to the user-generated content and is usually stored in relational databases or the file system of the ECS. Finally, transactional data [22] records user actions in the form of event logs. Transactional data is either stored in specific event databases or log files. The following section provides an excerpt of transactional data from HCL Connections and discusses how the dimensions for SCA provide an analytical lens for SCA (Fig. 1). This data understanding is referred to in the evaluation of SCAF in Section 5.3.

The transactional data of ECS is usually organized in a star schema. The central fact table (F.EVENTS) represents the event log. A unique ID identifies each event. The event describes the actor (USER_UUID), the affected content (ITEM_UUID), the module in which the event occurred (CONTENT_TYPE_ID), the event type (EVENT_OP_ID), the workspace in which the event occurred (COMMUNITY_UUID), and the timestamp (EVENT_TS). The IDs in the fact table are translated into human-readable descriptors in the surrounding dimensions tables. The event in Fig. 1 documents that a user created a new blog post in a workspace. The dimensions for SCA shown on the right in Fig. 1 were derived from the main concepts for workspace awareness by Gutwin and Greenberg [23] and the structure of ECS data. They were developed and evaluated in a focus group with 11 collaboration professionals [4]. The primary purpose of the dimensions for SCA is to guide the formulation of questions that can be posed to ECS data. They are an analytical lens for ECS data because they represent the main elements of transactional data.

2.2. Approaches for SCA in the literature

This section discusses existing approaches for SCA, which are described in the academic literature and extends the findings from a previous literature review on data analytics in ECS [24]. The resulting 85 publications from the literature review [24] were analyzed in multiple iterations.

In the first iteration, application areas for SCA were identified. Most of the analyzed studies used SCA for providing an overview of ECS use by using general metrics such as the number of users or activities in a system [25,26]. The analysis of the studies showed that social network analysis is often applied for analyzing networks and relations between users in an ECS [5]. Also, SCA is used to investigate the organizational and cultural impacts of an ECS, for example, by identifying how geographical dispersion and hierarchies affect the use of an ECS [27]. Other application areas for SCA are the identification of user types [28], the identification of usage patterns [29], the analysis of workspaces [30], and the identification of expertise of users [31,32].

In the second iteration, the identified studies were analyzed regarding their approach for analyzing ECS data to identify how SCA is applied. In many studies, the descriptions of the analytics approach and the underlying data were often imprecise or even missing. For example, 13 studies did not describe the characteristics and structure of the analyzed data (e. g. [33,34]) and 15 studies did not provide information on the analyzed system (e. g. [35,36]). Moreover, most studies did not contain background information about the problem domain or the study's context. The main limitation in the analyzed SCA studies is that only very few studies explicitly followed or referenced established data mining frameworks. This observation is interesting because it could be expected that well-established data mining process models such as CRISP-DM [37] could also guide the analysis of ECS data. Behrendt et al. [25] argued that the analysis of ECS data is different from analyzing traditional sources and thus, new methods and approaches are required. Similarly, Hacker et al. [12] argue that the recommendations in CRISP-DM are not suitable for application in the context of SCA. Despite these arguments, no new frameworks were suggested. The findings from the literature review demonstrate that existing data mining frameworks are not suitable for guiding the analysis of ECS data and that a new method is required.

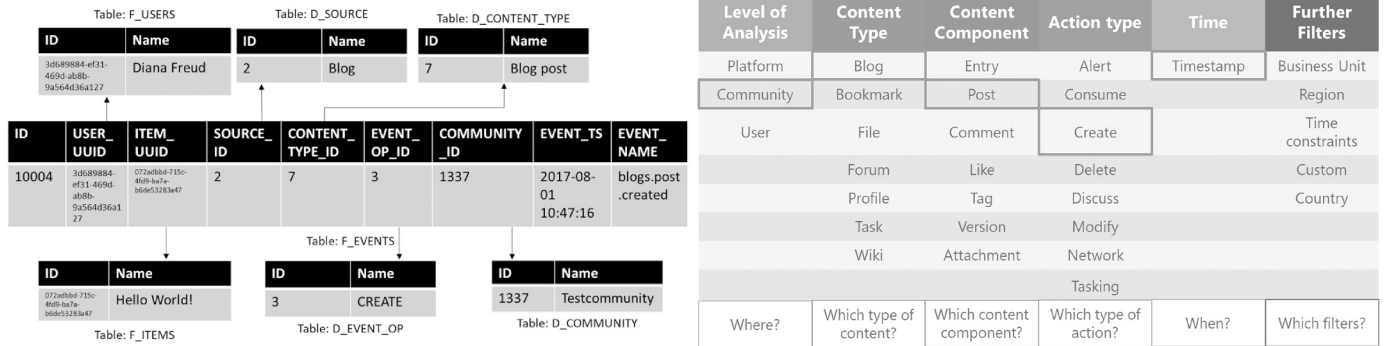


Fig. 1. Transactional data (left) and dimensions for social collaboration analytics (right).

3. Research design

The research design of this work is characterized by a design science research [38] approach that is supported by a mixed-methods research design [39]. The mixed-methods approach followed the embedded design [40] in which quantitative data is collected to support qualitative data. As qualitative data is the primary data collected in focus groups and the case study, the mixed-methods approach is qualitative-dominant [39]. The supporting quantitative data was collected through the survey and the ECS logfile analysis. Especially in the third research phase, mixing qualitative and quantitative data was beneficial because the statements on ECS use provided by the case company could be enriched with the actual ECS use recorded in the log files.

Focus groups were chosen as the primary data collection method because they are recommended for exploring emerging topics. Through the mutual and interactive group discussions with experts, a more comprehensive range of views [41] and requirements regarding SCA could be collected than in less interactive survey or interview-based approaches. The focus group participants were carefully selected. They are collaboration professionals and long-term members of IndustryConnect, which allowed conducting recurring focus groups with stable participants. As SCA is a recurring topic in IndustryConnect, the participants have similar knowledge on SCA. Due to their responsibility for the ECS in their company, the participants have a strong interest in SCA for reporting on ECS use. Different perspectives on SCA could be captured without causing greater diversity because the collaboration

professionals work in different organizational areas (e. g. IT, communication, business development). Based on Morgan's thought on "experimentation with both larger and smaller sized groups" [39,p. 59], the size of the focus groups in this research deviates from the standard recommendations of 6 to 10 participants, as explained below. The focus groups followed the five phases stimulation, presentation of input, focus group interaction, group discussion, and summary of results [42]. Each focus group was recorded and two team assistants created field notes. For the analysis, a summary-based approach was followed [41]. A reflective journal was written after the focus group. The recordings were not transcribed verbatim because the field notes and journal allowed to capture the participants' statements and the researcher's interpretations. Based on these materials, a focus group report was created [43]. After the research team reviewed the reports, the findings were made accessible in IndustryConnect and feedback was collected from the participants.

As shown in Fig. 2, the research design consisted of three phases that were derived from the DSR process model by Vaishnavi and Kuehler [38]. The phases and research steps were aligned to incorporate the DSR cycles and guidelines by Hevner et al. [44,45], as outlined in the following. Due to the constant exchange with practitioners in focus groups and the evaluation of the SCA Framework in the case study, the relevance of the research and its outcomes could be established (relevance cycle) [45].

In the first phase (awareness of problem), the rationale of this research was established by collecting initial evidence from academia

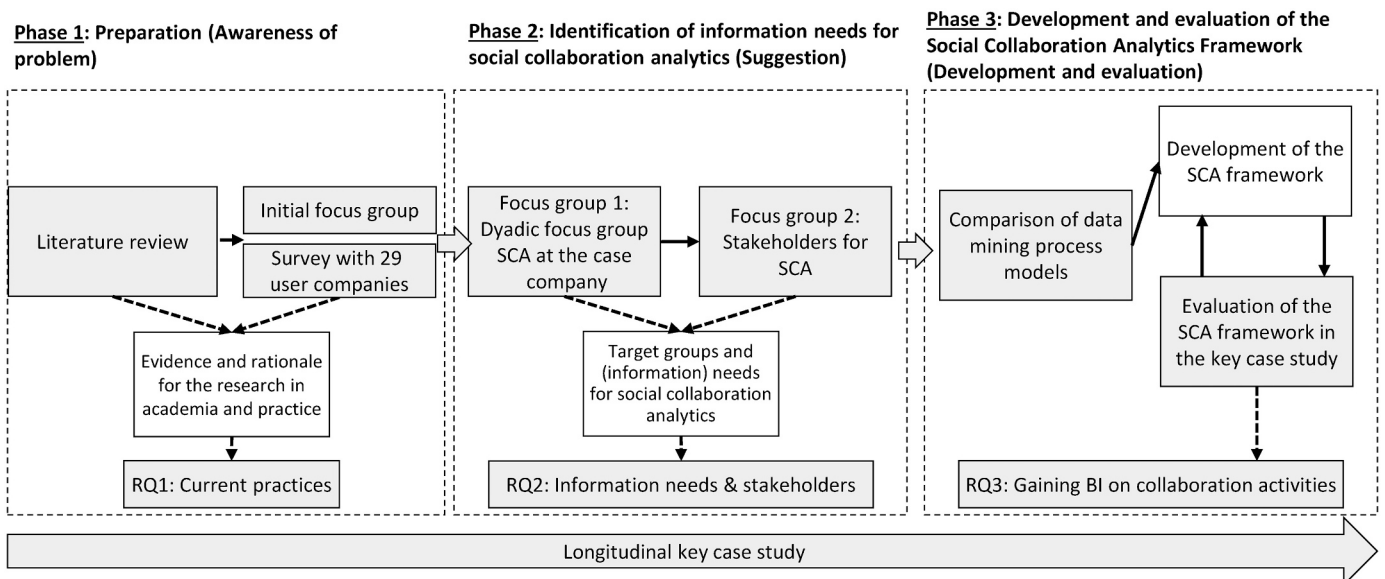


Fig. 2. Research design.

and practice through a structured literature review and a focus group with 11 collaboration professionals, as described in Section 2. The findings from this phase address RQ 1 and have been published in [4,15] (evidence from practice) and [24] (evidence from academia) and informed the choice of methods for this research (rigor cycle) [45].

In the second phase, possible stakeholders for SCA and their information needs were investigated in two focus groups with experts from leading ECS user companies. As this section contributes requirements for SCA, it provides suggestions for the foundation for the framework. The first focus group aimed to identify and document challenges and requirements for SCA at the case company. The focus group was designed as a dyadic focus group [41] with two participants. This small focus group allowed surfacing more detailed and more focused insights compared to larger focus groups that aim to capture broad and diverse views of the participants. The second focus group aimed to identify the stakeholders that require metrics and reports on ECS use. As the participants knew each other for many years, the members of IndustryConnect are stable, and the focus group was structured, 12 participants were invited. Such decisions are supported for structured focus groups and participants with the same knowledge level [46]. This design of larger focus groups fits the context of IndustryConnect and has proven to work for many years [10]. Phase 2 contributes the stakeholder classification for SCA, addressing RQ 2.

The third phase represents the design cycle [45] in which the SCA Framework was developed and evaluated. The phase started with analyzing and comparing existing data mining process models for identifying established recommendations and terminology for data analytics. Based on this comparison and the findings from the previous phases, the SCA Framework was developed. The three research phases were accompanied by a longitudinal in-depth key case study [47] following the eXperience case study methodology [48] to establish an in-depth understanding of the ECS project. Insights from the case study were used as inputs in the focus groups. The case study was established based on provided documents and two interviews at on-site visits that lasted between two and four hours. The case company was chosen as a study subject because the representatives already had prior experiences with SCA and volunteered to provide the necessary data. In the case study, the framework was evaluated through an analysis of ECS data that was provided by the case company to demonstrate how SCA can support collaboration professionals in gaining business intelligence on collaboration in the digital workplace (RQ 3).

4. Phase 2: stakeholders and information needs for social collaboration analytics

This section reports on the findings from two focus groups that aim to identify stakeholders and their information needs for SCA. The main contribution is the ECS stakeholder classification that is discussed in Section 4.2.

4.1. Dyadic focus group: social collaboration analytics at the case company

The objectives of this dyadic focus group were to identify and document current practices for SCA at the case company and how SCA can support the collaboration team. Two members of the collaboration team (roles: head of internal communications and data analytics expert) participated. The focus group lasted for approximately 4 h, including a break. The participants were stimulated with inputs on SCA and findings from previous research. Four guiding questions were used to structure the conversation. The participants were asked to reflect on analyses they have implemented in each application area of SCA as described in Section 2.2, experienced limitations in applying SCA, plans, and practical implications of SCA. Throughout the conversation, the two participants argued that they have a high need in SCA to measure and understand the ECS use. They mentioned that the standard analytics tool is limited to

counting registered users, which only provides a rough overview of the system's use. Consequently, they do not have enough information for their job. The participants raised concerns about the underlying data because an external business partner compiles the reports. The whole analytics process is not transparent and not documented, which limits the trust in the results. In the future, the collaboration wants to analyze the use of workspaces for identifying different workspace types to learn for which purposes employees use the ECS. Regarding the practical implications of SCA on the collaboration team's tasks, the two participants mentioned that SCA could support strategic areas and decision-making in the context of ECS to support the project and change management and the allocation of financial and human resources for the ECS project. In the focus group, the need for SCA at the case company could be identified. At the end of the focus group, the participants reaffirmed that they value the impact that SCA can have on their work.

4.2. Focus group: stakeholders for social collaboration analytics

Based on the insights from the previous research activities, the objective of this second focus group with 12 participants from 10 different ECS user companies was to identify the organizational roles in an ECS project that are potential audiences for SCA reports. The duration of this focus group was 90 min. The participants were stimulated with a stakeholder classification. The classification was developed based on a literature review and a review of the roles of the 67 IndustryConnect participants. In total, seven stakeholder roles, as shown in Table 1, were identified. The classification was evaluated in a group activity in which a group-decision support system was used. The participants were asked to enter their thoughts about the overall classification and individual roles. Besides, two questions were provided: "Which stakeholders required metrics?" and "Which stakeholders do you want to provide with metrics?". The responses were immediately shown on a screen to stimulate the discussion. In the following group discussion, the participants elaborated on the relevance and information needs of the identified stakeholders. All participants agreed that workspace managers, middle managers, and the collaboration team are key stakeholders for SCA. Workspace managers and the collaboration team are essential stakeholders because they are the first contact when employees struggle to use the ECS. They are interested in how employees use the ECS to provide better support. The management stakeholders are more interested in reporting for justifying investments and making decisions regarding the ECS project. The participants discussed that the remaining stakeholders are less important but should be included in the classification. The IT department usually has its own management systems in place for monitoring IT systems. Consultants are also no core stakeholders of SCA because they usually work for a limited time on the ECS project. In the focus group, the participants concluded that the stakeholder classification is valuable for identifying the target audience for SCA and assisting in addressing specific information needs in reports. The central outcome of this focus group is the evaluated stakeholder classification that is essential for the Social Collaboration Analytics Framework.

Table 1
ECS stakeholder classification.

Stakeholder	Description
Top management	C-level executives or board of directors
Middle management	Managerial staff and middle management
Collaboration team / Platform manager	Team of collaboration professionals responsible for the implementation and adoption of the ECS
Platform manager (IT)	Employees from the IT department
Workspace manager	Employees who create and own workspaces and are responsible for their management
Internal & external consultants	Internal and external consultants in the ECS project
Users of a workspace	All employees who are members of a workspace

5. Phase 3: developing and evaluating the Social Collaboration Analytics Framework

In this section, the findings from the previous research phases are synthesized. Based on an additional analysis of data mining process models, the SCA Framework is developed and following the recommendations of DSR, Section 5.3 describes the evaluation in the case study.

5.1. Analysis and comparison of data mining process models

Through a literature review, 16 data mining process models were analyzed and compared to ensure that SCAF is built on established recommendations and terminology. The comparison is shown in Fig. 3 and extends and updates a previous survey of data mining process models by Mariscal et al. [49] from 2010. As argued by the survey authors and shown in Table 2, the Knowledge Discovery in Databases (KDD) process model and CRISP-DM are the most influential data mining process models because most of the other models originated from them. Since its publication in 2000, CRISP-DM emerged as the de-facto standard for data mining [49]. A likely reason for this is that CRISP-DM is the only model that provides detailed descriptions and work steps that can be followed and applied in practice. It was developed by an industry consortium.

All of the analyzed data mining process models have in common that they are structured in phases. Due to its influential role, the phases of the other process models were aligned with CRISP-DM in the comparison. The comparison chart in Fig. 3 shows that most of the data mining process models are similar and mostly subtle differences can be observed. Whereas most models describe a phase for defining objectives, only six models describe a phase dedicated to interpreting and making sense of the analytics results (see Fig. 3: models #3, #4, #6, #7, #12, #14). Fig. 3 also shows that only 3 out of the 16 analyzed data mining process models contain a phase in which a problem is identified or understood and a phase that is dedicated to the visualization and interpretation of the findings (Fig. 3: models #3, #12, #14). These observations likely explain why much of the data analytics research focuses on achieving statistical outcomes instead of interesting and relevant findings. As discussed in the next section, the analysis of data mining approaches provided the overall structure and generic terminology of the SCA framework. The individual work steps and guidelines in SCAF were derived from the previous research phases and are specific to SCA.

5.2. The Social Collaboration Analytics Framework (SCAF)

The Social Collaboration Analytics Framework was developed by integrating the evidence and insights from academia and practice, which were gained throughout the three research phases. The framework is shown in Fig. 4. Each phase of the framework follows a dedicated objective and includes detailed work steps. For each phase, SCAF contains suggested outcomes and guiding questions that provide additional orientation when working through the phases. The framework stands out from existing models and frameworks because it tightly integrates the identification of business problems and questions with knowledge discovery to answer these questions. The interpretation of findings is supported by data visualization, an aspect that is neglected in most existing frameworks. In contrast to other models, SCAF goes beyond the mere presentation of findings and emphasizes identifying actions that can be taken based on the new insights. This is achieved through a prototyping and evaluation cycle that ensures that the identified questions can be answered. The framework ensures an academic approach to SCA and supports academics in measuring and understanding computer-mediated collaboration. By focusing on generating actionable insights that support organizational decision making and business intelligence, the framework contributes a new perspective to decision support in CSCW.

5.2.1. Business understanding

Following the terminology used in the most influential data mining process models, the first phase of SCAF is called business understanding. Eleven of the 16 analyzed data mining process models contain such a phase. The empirical research findings (Sections 2.1, 4.1, and 4.2) support the primary objective of the business understanding phase, which is to establish the business context and domain knowledge of the ECS project and to define precise objectives for the SCA process. The first step is to establish the domain knowledge, which is necessary for understanding the ECS project. The eXperience case study guideline supports collecting relevant facts such as the implementation date of the ECS, the impetus for the ECS project, expected benefits, and use cases for the ECS. Additionally, defining a shared vocabulary is recommended for coherent communication throughout the SCA process.

The second step is to identify the objectives and questions that should be addressed by SCA. The framework suggests breaking the identified objectives down into answerable and focussed questions. As it can be challenging for collaboration professionals to phrase questions, it is recommended to perform creative activities such as interactive focus groups or workshops utilizing structures such as the dimensions for SCA. The SCA stakeholder classification provides additional assistance in defining a target audience for a report and tailoring the objectives and questions to the target audience's specific needs. An aspect that is often neglected in data science process models is discussing barriers that could impede SCA. The outcomes generated as part of the business understanding phase are essential for making sense of and interpreting the reports in the final phase.

5.2.2. Data understanding

The objective of the data understanding phase is to identify, understand, and document the available data. Seven of the 16 analyzed data mining process models describe a comparable phase. The phase is placed after the business understanding phase because the information collected in the previous phase can already encompass discussions on possible data sources.

The first step is to identify available data sources. IT departments or application owners can help identify and understand available data sources because of their expertise with the systems. The data sources for SCA (Section 2.1) can be valuable in this step.

The second step is to explore and describe the identified data. The outcome is a report that describes the structure, attributes, and size of the available data. Data quality must be assessed and documented because quality issues can be manifold and low data quality can prevent the identification of interesting patterns. Typical issues are noisy or incomplete data [37]. There are also specific challenges with ECS data [4]. For example, one user action may trigger multiple events in the log, which makes the identification of event bundles that belong to one user action necessary. Additionally, many operational systems regularly purge their event logs. Thus, the period covered by the event log needs to be documented. The following techniques help assess the data quality: (1) Counting the number of occurrences of each event (provides an overview of the logged events), (2) Counting the number of occurrences of each event type (can identify issues between the association of events and event types), and (3) Generating a list of dates in the event log including the number of events that occurred on the dates (can identify log purging and gaps in the event log). The final step in this phase is to investigate other technical barriers, which can be identified with the help of the IT department or application owners. Such barriers include that data cannot be accessed in real-time, or during certain periods.

5.2.3. Conceptualization

The main objective of the conceptualization phase is to select the required metrics and appropriate data visualizations for answering the identified questions. Based on the knowledge from the previous two phases, a concept or mockup is developed that describes what the final analytics solution or report can look like and what the involved

Framework/Process	#1 CRISP-DM	#2 ASUM-DM	#3 KDD	#4 Data Science Edge	#5 SEMM A	#6 BI-Lifecycle	#7 5A's	#8 The Data Mining Process	#9 Fast Analytics	#10 Team Data Science Process	#11 Cabena	#12 Two Crows	#13 Anand & Buchner	#14 Cios et al.	#15 DMIE	#16 Human- Centered Approach
Number Of Phases	6	5	9	5	5	7	5	5	7	5	5	7	8	6	5	6
Phases																
	Business Understanding	Analyse	Learning the Application Domain	Plan Stage		Discovery	Assess	State the Problem	Scope	Business Understanding		Define Business Problem	Domain Knowledge Elicitation	Understanding the Problem Domain	Analyse the organisation	Task Discovery
													Human Resources Identification			
											Select		Problem Specification			
	Data Understanding		Creating a Target Data Set		Sample							Build DM	Data	Understanding the Data		Data
	Data Preparation	Design	Data Cleaning and Pre- Processing	Collect	Explore		Access	Collect the Data	Discovery	Data Acquisition and Understanding	Pre-process	Explore Data	Methodology identification	Preparation of the Data		Data Cleaning
			Data Reduction and Projection	Curate	Modify			Preprocess the Data			Transform	Prepare Data for				
	Modelling	Configure & Build	Choosing the Function of DM	Analyse	Model	Design	Analyse	Estimate the Model	Analyse	Modelling	Mine	Build Model	Pattern discovery	Build the Model	Develop Data Model	Model Development
			Choosing the DM Algorithm			Development			Model/ Design/ Development							
			Data Mining													
	Evaluation		Interpretation		Assess	Test		Interpret the Model and Draw Conclusions	Validate		Analyse and Assimilate	Evaluation of the Discovered Knowledge	Knowledge Post- Processing	Evaluation of Discovered Knowledge		
	Deployment	Deploy				Deploy			Deployment	Deployment					Implement Model	Data Analysis
						Support			Support/ Feedback						Establish ongoing support	
			Using Discovered Knowledge	Act		Value Delivery	Act			Customer Acceptance		Using Discovered Knowledge		Using the Discovered Knowledge		Output generation
							Automate									

Fig. 3. Comparison of existing data mining process models (extends and updates Mariscal et al. [49]).

Table 2

Overview and classification of data mining process models.

KDD related process models	CRISP-DM related process models	Other independent process models
Human-centred approach	Cios et al	5 A's
Cabena et al. Anand & Büchner	DMIE Marbán et al	The Data Mining Process Business Intelligence Lifecycle
Two Crows	Data Science Edge	Fast Analytics/ Data Science Lifecycle
SEMMA	ASUM-DM	Team Data Science Process

stakeholders can expect.

The first step in this phase is to choose the required metrics and algorithms for answering the identified questions. Inspirations for metrics and algorithms can be found in the academic literature, which describes the application areas of SCA in more depth [24]. The dimensions for SCA provide an analytical frame for conceptualizing questions, data queries, or algorithms.

The second step is to identify and select the required data from the available data. In contrast to most of the other data mining process models, SCAF also emphasizes data visualization because the experiences from previous research activities show that suitable data visualizations support the interpretability of results.

5.2.4. Data preparation

In the data preparation phase, the objective is to address the issues that were identified in the previous phase. As data preparation is common in 11 of the 16 analyzed data mining process models, SCAF follows existing recommendations. This phase's outcome is a documentation of the data preparation, which is essential for making the data preparation reproducible.

The initial step in this phase is the data collection to ensure that only the required data is collected for keeping the dataset at a minimal size. In most modern ECS, data can either be queried directly from the system's database or APIs. Following the data collection, the next step is to build a data model. The analysis of the data mining process models has shown that two different interpretations of the term data model exist. The term data model can refer to a statistical model that includes statistical assumptions and parameters [50]. In Information Systems, the term data model is frequently used to describe data structure and attributes in applications [51]. Depending on the defined objectives, either a statistical model or a data model that describes data in the application domain is built.

The data preparation phase is concluded by pre-processing the data. The main challenges in the context of SCA are missing values and noisy records. For SCA, it is recommended to substitute missing values based on the domain knowledge gained in the first two phases instead of deleting them. Moreover, data sets can have incorrect labels, wrong names (class or label noise), or particular attributes can be corrupt (attribute noise). Based on the domain knowledge, each case of noisy data should be handled individually. For event records that are associated with system users, it is recommended to filter them because such events typically represent clean-up tasks by the system, which can potentially falsify analyses.

5.2.5. Prototyping

The prototyping phase is another new phase introduced by SCAF. The need for this phase was derived from experiences from prior research activities in which SCA prototypes were discussed with collaboration professionals (Section 2.1, [4]). The main activity in the prototyping phase is to develop a technical implementation based on the concepts from the conceptualization phase. The demonstration of prototypes can spark valuable discussions and demonstrate how the final solution looks like.

5.2.6. Evaluation

The prototyping phase initiates a development and evaluation cycle following the recommendations from design science research. In the evaluation phase, the prototype is evaluated together with collaboration professionals to ensure that the identified questions are addressed. The evaluation encompasses all aspects, including the data model, the metrics, and the data visualization. It may be necessary to go through multiple iterations between the prototyping and evaluation phases. Suitable methods for evaluating the prototype are interactive workshops or focus group discussions, in which the collaboration professionals and the identified stakeholders are encouraged to discuss and interpret the findings gained from the prototype. The findings from this phase, including required improvements, should be documented in an evaluation report.

5.2.7. Analysis

Following the prototyping and evaluation iterations, the analysis phase describes the two steps of data analysis and data visualization. The outcomes of this phase are the data analysis and the visualization of the results. To emphasize the importance of appropriately describing and interpreting the results, SCAF distinguishes between the analysis and knowledge discovery.

5.2.8. Knowledge discovery

The knowledge discovery phase is essential in SCAF because it brings together the findings from the analysis phase and the domain knowledge from the business understanding phase. In this phase, the explanation and interpretation of the results are emphasized. The newly generated insights should provide answers to the identified questions.

For the interpretation of the findings, the collaboration professionals and other identified stakeholders should be involved if possible. The domain knowledge from the business understanding phase can provide valuable contextual information for interpreting the findings. As SCA should be a continuous process, the findings from the phase should be used for the next iteration of SCAF. For example, based on the findings and interpretation of the results, the identified questions can be refined. Additionally, a continuous reporting cycle can be established. The main outcome of this final phase is a report that answers the identified questions and enables the taking of actions and measures to change the status quo.

5.3. Evaluation of the SCA framework in the key case study

This section describes the evaluation of SCAF by applying the framework for analyzing ECS data from the case company. The evaluation is guided by the working steps and guiding questions of SCAF. Due to the scope, not all guiding questions of SCAF are addressed in the evaluation. For the evaluation, the phases were executed under the lead of the researcher, and the findings of each phase were discussed with the collaboration team of the case company. As mentioned throughout the phases, the collaboration team provided support.

5.3.1. Business understanding

The business understanding was established in the underlying longitudinal case study with the case company and the focus group described in Section 4.2. Both activities were conducted with participants from the collaboration team. The typical tasks of the collaboration team are to give user trainings, and stimulate the adoption of the ECS through company-internal communication campaigns. Because of their tasks and expertise, the members of the collaboration team are ideal participants for establishing the business understanding. The findings from the business understanding phase are summarized following the guiding questions.

Guiding question: What are the characteristics of the problem domain? (Step 1: Establish domain knowledge)

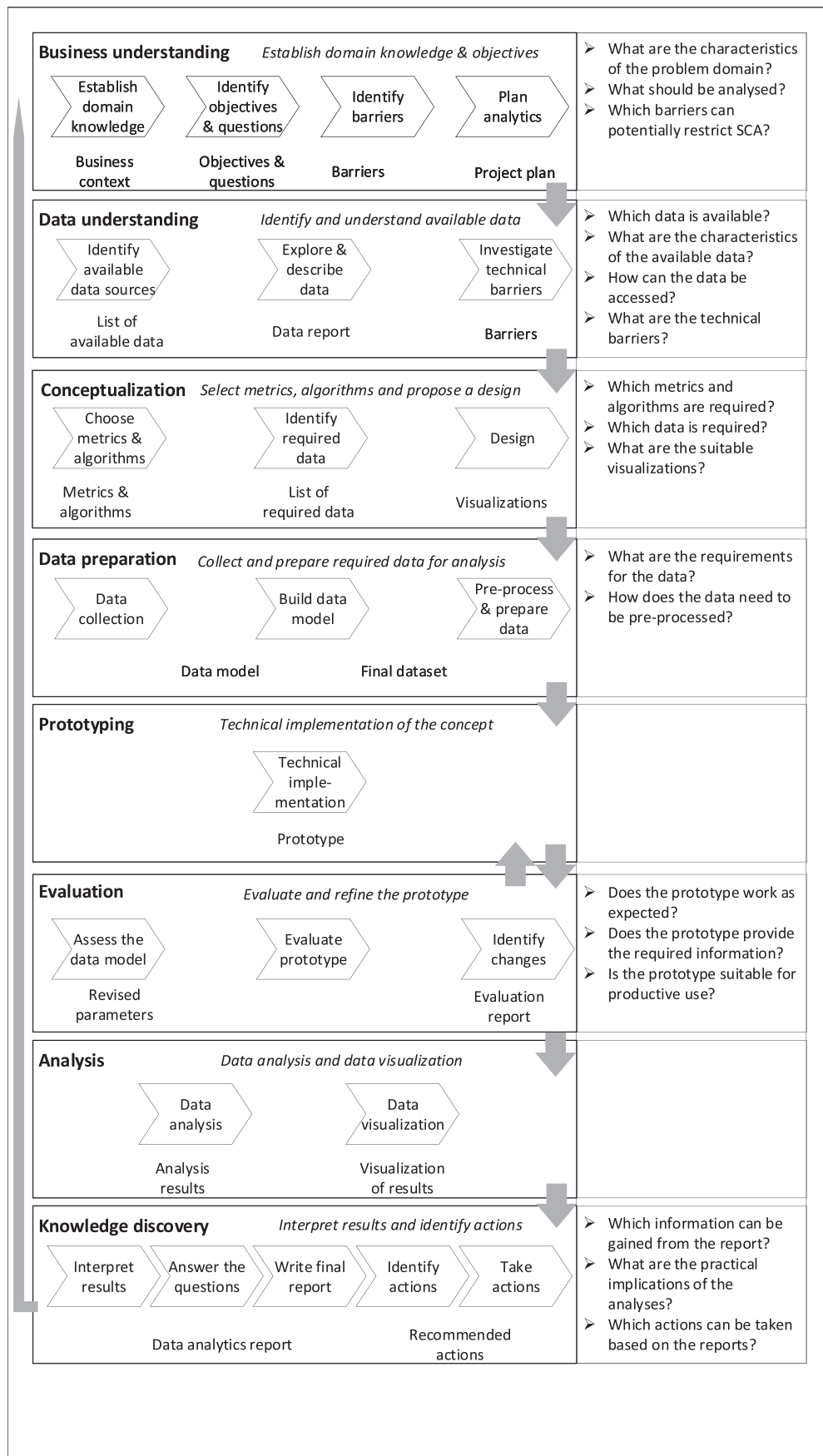


Fig. 4. The Social Collaboration Analytics Framework.

The case company is a leading industrial company with more than 120,000 employees worldwide. The company produces components for the automotive industry and is specialized in manufacturing plants for industrial companies. The ECS (ccConnect), based on HCL Connections, was introduced because many employees perceived the existing intranet as outdated. The employees demanded a new and modern intranet with “social features” (e. g. blogs, forums, wikis, and social profiles) and features for following other users and content. As ccConnect also provides traditional groupware features such as file sharing, task management, and shared workspaces, the system serves as a central and company-wide system for collaboration and communication among employees. Before implementing ccConnect, there was no company-wide system that enabled employees from different locations to collaborate. Although the collaboration team did not define expected benefits, they hoped that the ECS implementation could improve knowledge management, communication, collaboration, and connectedness between employees. During the early stages of the implementation, the collaboration team noticed the limited analytics features for assessing the use of the ECS, which was the primary motivation of the case company to participate in this research.

The ECS was rolled out and made available to 50,000 employees in March 2016. Typical use cases that were observed by the collaboration team are enterprise communication, knowledge management, project organization, and team organization. The use cases indicate that ccConnect is primarily used for enabling collaboration and communication among employees.

Guiding question: What should be analyzed? (Step 2: Identify objectives & questions)

In the dyadic focus group (Section 4.1), objectives and questions for conducting SCA at the case company were defined. The collaboration team wanted to gain deeper insights into the growth of the ECS over time and the use of workspaces. Consequently, the following two questions were defined: (1) Which proportion of the employees uses ccConnect, and can a growing adoption be measured? (2) How can the use of specific workspaces be characterized? These questions address some of the most critical information needs of the collaboration team.

5.3.2. Data understanding

Guiding question: Which data is available? (Step 1: Identify available data sources)

Together with the case company, it was decided that the collaboration team provides an excerpt of the transactional data of ccConnect to the researcher in the comma-separated value format. The transactional data contained a log of all user actions on ccConnect between November 2017 and November 2018. The structure of the provided data is equivalent to what was discussed in Section 2.1.

Guiding question: What are the characteristics of the available data? (Step 2: Explore & describe data)

Following the suggestions from the data understanding phase, as

Table 3
Characteristics of the data.

Software platform	HCL connections
Time range of the data	12.11.2017–12.11.2018
Type of data used for analysis	Transactional data containing events
Number of registered users	Exact number is unknown. 50,000 are targeted
Number of events in the log	2,772,328
Number of users in the log	35,074
Number of workspaces	3177

discussed in Section 5.2.2, Table 3 summarizes the main characteristics of the provided data. Table 4 shows the distribution of events in each module on ccConnect.

Remarkably, 86% of all events occurred in the files module, which seemed suspicious. A closer analysis of the event log revealed a malfunction in the logging mechanism of ccConnect. On ccConnect, events with the event type (EVENT_OP) read were not logged, which usually account for the majority of events in ECS. The share of events in the files module is extraordinarily high because, in the files module, the event type download is logged instead of the type read. Thus, only the files module logs the consumption of content on ccConnect. The identified issue has limited impact on the planned analyses as the use of ccConnect can also be measured without the read events. For the analyses on the workspace level, the file download events are sufficient for identifying a user as a consumer.

5.3.3. Conceptualization

Guiding question: Which metrics and algorithms are required? (Step 1: Choose metrics and algorithms)

Together with the case company, it was decided to implement the analyses in a dashboard in Microsoft Power BI, a market-leading application for creating interactive dashboards [52]. Power BI was chosen because of its flexibility and rich functionalities. The developed dashboard can be easily handed over to the collaboration team for continuous use.

Table 5 shows which metrics were selected together with the collaboration team for answering the questions. For measuring platform growth (Q1), several metrics were chosen that are displayed as line charts based on a time axis. (Q2) For characterizing the use of selected workspaces, several basic measures were chosen and displayed in a table. As a previous study [53] has confirmed that the ECS user typology can provide valuable insights into the use of workspaces, the distribution of user types as defined in the user typology was calculated for each workspace. Following the user typology, consumers are users who consume content, contributors contribute to and interact with existing content (e. g. edit, recommend, tag), and creators create new and original content (e. g. create new posts, upload files). Based on these definitions, the ECS user typology allows analyzing collaboration and participation in workspaces.

5.3.4. Data preparation

Guiding question: How does the data need to be pre-processed? (Step 2: Build data model & Step 3: Prepare data)

As the data tables were provided in CSV files, they had to be imported into Microsoft Power BI. The data model needed to be configured based on the foreign key relationships between the event table (fact table) and the dimension tables, as shown in Fig. 1. Events that were triggered by system users (list of system user IDs: 0; 1; 2; 3; 4; 10,000;

Table 4
Events in modules.

Module	# of events
Blogs	25,815 (0.93%)
Bookmarks	12,577 (0.45%)
Files	2,386,075 (86.07%)
Forums	20,778 (0.75%)
Homepage	303 (0.01%)
Ideationblog	5799 (0.21%)
Profiles	82,852 (2.99%)
Tasks	13,298 (0.48%)
Wikis	80,603 (2.91%)
Total	2,772,328

Table 5
Selected metrics for answering the questions.

Question	Selected metrics
Q1: Platform growth	<ul style="list-style-type: none"> • Number of registered users • Number of active users • Number of events • Number of workspaces
Q2: Use of workspaces	<ul style="list-style-type: none"> • Number of active workspaces • Number of members • Number of events • Events per user • Creator (%) • Contributor (%) • Consumer (%)

10,021; 10,060; 10,110; 11,601) were filtered from the dataset. Additionally, a calendar table, which allows advanced date-based calculations, was added to the data model. As it was impossible to reconstruct the missing read events from the event log, no further steps in the data preparation were necessary.

5.3.5. Prototyping & evaluation

The metrics listed in Table 5 were implemented in a dashboard in Microsoft Power BI, as shown in Fig. 5.

For initial analysis of the use of workspaces on ccConnect, six workspaces were chosen together with the collaboration team. As the collaboration team had insights and access to the workspaces, they could be classified into three types: workspaces used for projects, workspaces used by organizational units or departments, and topic-specific information workspaces. The six workspaces were selected for analysis by the collaboration team because they are interesting for learning about collaboration in different contexts on ccConnect. For these workspaces, the metrics as listed in Table 6 were implemented in Power BI.

Guiding question: Does the prototype provide the required information? (Step 2: Evaluate prototype)

As the collaboration team was not available at the time of evaluation, a self-evaluation of the dashboard was performed. It was concluded that with the help of the metrics, as shown in Fig. 5, the first question on platform growth could be answered because the charts are suitable for indicating platform growth. Based on the results and experiences from a previous study on analyzing ECS workspace types [53], it was concluded that the metrics listed in Table 6 are suitable for analyzing and interpreting the use of specific workspaces. As the transactional data does not contain the actual number of workspace members, in the following, the terms users or members refer to the number of employees who actively used a workspace.

5.3.6. Analysis & knowledge discovery

Guiding question: Which information can be gained from the report? (Step 1: Interpret results)

The charts in Fig. 5 show the number of registered users (a) and workspaces on ccConnect (c), as well as the number of events (b), active users (d), and the number of used workspaces (e) for each month between November 2017 and November 2018. All the charts demonstrate a growing use of ccConnect. Between November 2017 and November 2018, the number of registered users (a) has grown by 16%, from 26,000 users to more than 30,000 users. In the data understanding phase (Section 5.3.2), it was mentioned that the event log contains users from 35,074 users. An analysis of the “missing” 4626 users revealed that these user profiles have never been used after the creation. The charts only show the users that have actually used the platform. Strong growth can also be observed in the number of created workspaces on ccConnect (c).

In the same period, the number of workspaces increased by 26%, from 4145 to 5246. As these two charts only show the number of users and workspaces on ccConnect, the number of users that actually used ccConnect (d), and the workspaces that have actually been used (e) need to be investigated. In all charts, growth can be observed between November 2017 and November 2018. The charts also show some peaks in the activity and active users. For example, in March 2018, July 2018, and October 2018, the number of active users and events is higher compared to the previous months. As these months are either the end or the start of quarters in the calendar year, the peaks suggest that employees use ccConnect to prepare and publish quarterly reports. It can be concluded that the use of ccConnect has significantly increased in one year. Nevertheless, there are also several months in which the activity is lower compared to others (e. g., February 2018, May 2018, September 2018). A likely explanation for the decrease in events in September 2018 is that in most European countries, September is in the middle of the summer holiday season. It would be interesting to observe these charts over multiple years for identifying seasonal trends.

The collaboration team was also interested in learning about collaboration in the selected workspaces and finding out whether the defined metrics can characterize the use of these workspaces. This question is also of academic interest. Naturally, the selection of the six workspaces (Table 6) is not representative of all 3177 workspaces in the dataset. The following sections aim to explore and interpret aspects that the collaboration professionals consider relevant and interesting for their understanding of collaboration.

According to the collaboration team, workspaces 206,501 and 14,478 were used by employees from several departments for jointly working on projects. The workspaces stand out from the other workspaces in multiple aspects. First, compared to the other selected workspaces, they have significantly fewer members (both: 36) and a comparably high level of activity represented by the events per user metric (40–50). These project workspaces have a remarkably high share of creators (at least 40%) and an accordingly lower share of consumers. This distribution of user types demonstrates a much higher level of participation and collaboration in project workspaces compared to the other workspace types. The workspace characteristics expressed by the metrics are in line with the discussion by Jeners and Prinz [30] on project workspaces in groupware. The authors characterize project workspaces by a small number of members and a high level of activity and collaboration. The selected metrics also accurately represent the use of the workspaces as described by the collaboration team.

Workspaces 18,319 and 83,782 were used for coordinating an organizational unit and a department and for communicating department-specific information to employees. Accordingly, these workspaces have more members (~1400) than project workspaces. The distribution of the user types also profoundly differs from project workspaces. The numbers show that most users (97%) consume content in these workspaces, and only a few users actively engage and collaborate in these workspaces (~2% contributors, ~1% creators). These numbers indicate that content in these workspaces/ departments is likely created and distributed by a few administrative staff members to the other employees. These workspaces are primarily used as a medium for communicating and broadcasting information within departments, which confirms the observations and statements by the collaboration team. The remaining workspaces are used for providing specific information on particular topics and match closely with the basic ideas of communities of practice and communities of interest.

From the whole dataset, workspace 10,141 stands out because of its extraordinarily high number of members (10,862) and consumers (99.5%). These numbers are not surprising because this workspace is used as an onboarding workspace in which essential information for getting started with ccConnect is provided to new employees. The content in this workspace is created by members of the collaboration team and accessed by the other employees. This is also represented in the ratio between creators and consumers. The second workspace

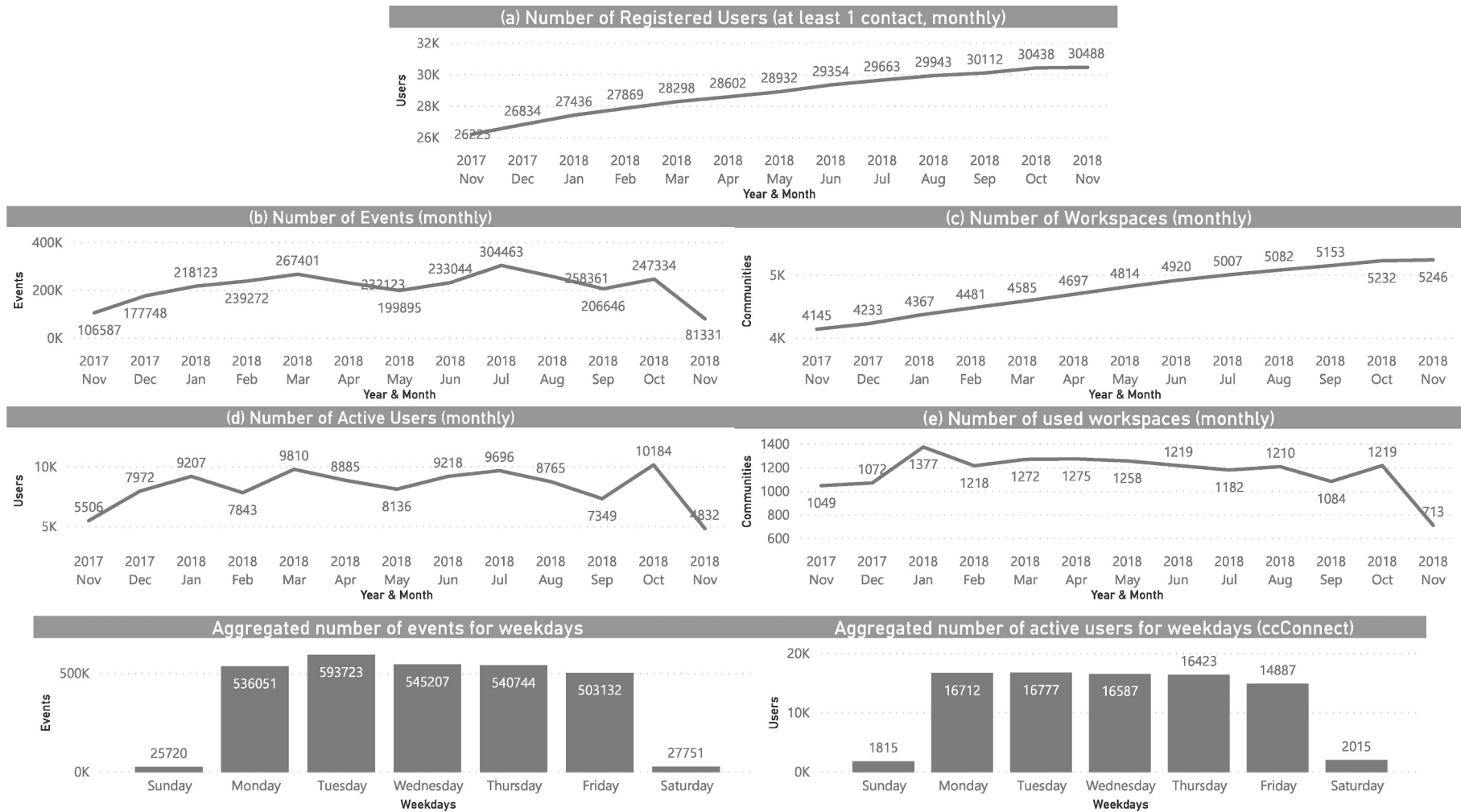


Fig. 5. Dashboard for monitoring the growth of the ECS.

Table 6
Calculated metrics for selected workspaces.

ID	Type	User	Events	Events/user	Consumer	Contributor	Creator
206,501	Project	36	1821	50.83	33% (12)	3% (1)	64% (23)
14,478	Project	36	1580	43.89	58% (21)	0	42% (15)
18,319	Department	1470	105,457	71.73	97% (1431)	2% (31)	1% (6)
83,782	Department	1483	18,632	12.56	97% (1447)	2% (27)	1% (9)
10,141	Information	10,862	94,766	8.72	99.5% (10816)	0.3% (33)	0.1% (6)
672,506	Information	1797	11,338	6.31	99% (1780)	0.7% (13)	0.3% (4)

(672,506) is a topic-specific workspace where a group of experts documents knowledge, experiences, and best practices related to a specific area of expertise. This knowledge repository is accessible to all employees working in this domain. The workspace has a comparatively high number of members. Only some users engage in creating and contributing to the content. Information workspaces can be mainly distinguished from department workspaces because of their lower level of activity represented by the events per user metric and a higher number of members.

Guiding question: What are the practical implications of the analyses? (Step 2: Answer questions & Step 4: Identify actions)

The interpretation of the selected workspace metrics matches with the observations made by the collaboration team. It needs to be considered that only six workspaces out of 3177 were selected for analysis. The objective of the analysis was not to identify rules or assumptions about workspaces but to generate interpretations and insights that are relevant and interesting for the collaboration team. Table 7 summarizes the analytics report by answering the questions from the business understanding phase. In a final meeting for evaluating the analytics results, the head of the collaboration team confirmed that the results are plausible and relevant. The collaboration team was highly satisfied by the presented analyses because they went far beyond their current possibilities and increased the understanding of how ccConnect is adopted and used by the employees. In the final evaluation meeting, the collaboration team was particularly interested in extending the analysis of workspaces to a broader range of workspaces. The collaboration team perceived the SCA Framework as valuable. The steps and guidelines provided by the SCA Framework helped the practitioners to understand and work through the SCA process.

Table 7
Summary of the analytics reports by answering the questions.

Question	Answer
Which proportion of the employees uses ccConnect and can a growing adoption be measured?	Approximately 50,000 employees are targeted by ccConnect. By the end of November 2018, there were 30,488 registered employees who also actually used ccConnect. Thus, almost 61% of the targeted employees use ccConnect. Strong growth in the use of ccConnect can be witnessed.
How can the use of specific workspaces be characterized?	Based on the selected metrics and the ECS user typology, it could be shown that the six analyzed workspaces of different types are used differently. The metrics revealed a different degree of activity, participation, and interaction. It would be interesting for the collaboration team to analyze all workspaces for gaining an overview of the different types of workspaces on ccConnect. From the analysis of the workspaces, it can be concluded that ccConnect is used for the intended use cases, which are mainly collaboration and communication.

6. Discussion and conclusion

The main contribution of this research is the Social Collaboration Analytics Framework. The framework consists of eight phases containing working steps and guiding questions that guide researchers and practitioners in applying SCA. The focus is on identifying and addressing precise questions of SCA stakeholders through visualizing and interpreting the findings. In these regards, SCAF stands out because other frameworks are rather focused on the data, data models and analytics (Section 5.1) instead of actually deriving interpretations and actionable insights that can provide decision support. The evaluation of SCAF with the case company highlighted that precise questions, data visualization, and interpretation are essential for practitioners. The framework also provided the structure for documentation based on which the study can be repeated later or in other cases. The outcome of the first iteration of SCAF at the case company was an interactive dashboard that the collaboration team can now use. New insights into collaboration on ccConnect could be gained based on the contextual knowledge from the business understanding phase that helped to interpret and make sense of the collected metrics. Through this first iteration of SCAF, a first step to establish SCA as part of BI at the case company was made. In conclusion, in the case study it could be demonstrated that SCAF can contribute to support decision-making regarding collaboration systems in the digital workplace by providing actionable insights into collaboration and the use of ECS. The structure and terminology of the SCA Framework were derived from existing data mining approaches and thus are generic and can potentially be applied in various analytics projects, which needs to be evaluated in future research. The detailed guidelines and recommendations of the framework are unique for the context of SCA because they are based on specific structures such as the SCA stakeholder classification and the dimensions for SCA. Because of this, the framework supports the analysis of joint collaboration instead of traditional transaction processing, adding a new perspective to BI and DSS. In addition, the research demonstrates how a design science approach combined with a mixed-methods approach can successfully be incorporated in a research design to contribute rigorous and relevant outcomes to the DSS field [54]. The framework was developed based on the findings from a mixed-methods approach utilizing results from literature reviews, survey data, focus group data, and transactional data in the evaluation. By constantly evaluating research outcomes in focus groups with practitioners, a bi-lateral exchange between academia and practice, as called for by Te'eni et al. [17], was established. In the evaluation of SCAF in the key case study, it was demonstrated how the framework can be applied in a working environment in practice, which addresses the call by Nunamaker et al. [18] [] to "shepherd an information systems (IS) solution through the last research mile". Additionally, a first contribution to identifying workspace types was made, which provides opportunities for future research in CSCW.

The main limitation of this work is that SCAF was evaluated in one single case study under the researcher's lead. Additional evaluation in other case studies and independent application of SCAF by practitioners can potentially reveal new refinements and enable cross-case comparisons. The current application of SCA is mostly limited to computing metrics. For gaining deeper insights into collaboration, the scope of SCA needs to be extended to methods from process mining for analyzing

collaboration processes. In conclusion, SCAF is a new and encompassing framework guiding the analysis of ECS data that describe collaboration activities. As it ensures an academic approach for the data analysis and supports decision-making for CSCW systems at the organizational level, the framework contributes a new perspective to DSS research.

Funding sources

The research project leading to the described outcomes was funded by the Deutsche Forschungsgemeinschaft (DFG) [grant number 285717372].

Declaration of Competing Interest

None.

References

- [1] V.H. Trieu, Getting value from business intelligence systems: a review and research agenda, *Decis. Support. Syst.* 93 (2017) 111–124, <https://doi.org/10.1016/j.dss.2016.09.019>.
- [2] C.W. Holsapple, M.P. Sena, ERP plans and decision-support benefits, *Decis. Support. Syst.* 38 (2005) 575–590.
- [3] M. Aanestad, T.B. Jensen, Collective mindfulness in post-implementation IS adaptation processes, *Inf. Organ.* 26 (2016) 13–27.
- [4] F. Schwade, P. Schubert, Social collaboration analytics for enterprise collaboration systems: providing business intelligence on collaboration activities, in: 50th Hawaii Int. Conf. Syst. Sci., Hilton Waikoloa Village, HI, USA, 2017, pp. 401–410.
- [5] W. van Osch, C.W. Steinfield, Strategic visibility in enterprise social media: implications for network formation and boundary spanning, *J. Manag. Inf. Syst.* 35 (2018) 647–682, <https://doi.org/10.1080/07421222.2018.1451961>.
- [6] K. Dery, I.M. Sebastian, N. van der Meulen, The digital workplace is key to digital innovation, *MIS Q. Exec.* 16 (2017) 135–152.
- [7] A. Richter, Locked-down digital work, *Int. J. Inf. Manage.* 55 (2020), <https://doi.org/10.1016/j.ijinfomgt.2020.102157>.
- [8] B. Wehner, C. Ritter, S. Leist, Enterprise social networks: a literature review and research agenda, *Comput. Netw.* 114 (2016) 125–142.
- [9] J. Viol, J. Hess, Information systems research on enterprise social networks – a state-of-the-art analysis, in: *Multikonferenz Wirtschaftsinformatik (MKWI 2016)*, Ilmenau, Germany, 2016, pp. 351–362.
- [10] S.P. Williams, P. Schubert, Connecting industry: building and sustaining a practice-based research community, in: 50th Hawaii Int. Conf. Syst. Sci., Hilton Waikoloa Village, HI, USA, 2017, pp. 5400–5409.
- [11] M. Kügler, S. Smolnik, G. Kane, What's in IT for employees? Understanding the relationship between use and performance in enterprise social software, *J. Strateg. Inf. Syst.* 24 (2015) 90–112.
- [12] J. Hacker, F. Bodendorf, P. Lorenz, A framework to analyze enterprise social network data, in: M. Atzmueller, O. Samia, T. Roth-Berghöfer (Eds.), *Enterp. Big Data Eng. Anal. Manag.*, IGI Global, Hershey, USA, 2016, pp. 84–107.
- [13] J.F. Nunamaker, L.M. Applegate, B.R. Konsynski, Facilitating group creativity: experience with a group decision support system, *J. Manag. Inf. Syst.* 1 (1987) 422–430.
- [14] A.W. Siddiqui, S.A. Raza, Z.M. Tariq, A web-based group decision support system for academic term preparation, *Decis. Support. Syst.* 114 (2018) 1–17.
- [15] F. Schwade, P. Schubert, A survey on the status quo of social collaboration analytics in practice, in: *Eur. Conf. Inf. Syst.*, Portsmouth, United Kingdom, 2018.
- [16] H. Oesterle, B. Otto, Consortium research: a method for research-practitioner collaboration in design-oriented IS research, *Bus. Inf. Syst. Eng.* (2010) 283–293, <https://doi.org/10.1007/s12599-010-0119-3>.
- [17] D. Te'Eni, S. Seidel, J. Vom Brocke, Stimulating dialog between information systems research and practice, *Eur. J. Inf. Syst.* 26 (2017) 541–545, <https://doi.org/10.1057/s41303-017-0067-9>.
- [18] J.F. Nunamaker, R.O. Briggs, D.C. Derrick, G. Schwabe, The last research mile: achieving both rigor and relevance in information systems research, *J. Manag. Inf. Syst.* 32 (2015) 10–47.
- [19] J. Cao, H. Gao, L.E. Li, B. Friedman, Enterprise social network analysis and modeling: a tale of two graphs, in: 32nd IEEE Int. Conf. Comput. Commun., Turin, Italy, 2013, pp. 2382–2390, <https://doi.org/10.1109/INFCOM.2013.6567043>.
- [20] S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics – challenges in topic discovery, data collection, and data preparation, *Int. J. Inf. Manage.* 39 (2018) 156–168, <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>.
- [21] R.H. Sprague, Electronic document management: challenges and opportunities for information systems managers, *Manag. Inf. Syst. Q.* 19 (1995) 29–49, <https://doi.org/10.2307/249710>.
- [22] A. Clevén, F. Wortmann, Uncovering four strategies to approach master data management, in: *Hawaii Int. Conf. Syst. Sci.*, 2010, pp. 1–10, <https://doi.org/10.1109/HICSS.2010.488>.
- [23] C. Gutwin, S. Greenberg, A framework of awareness for small groups in shared-workspace groupware, *Comput. Coop. Work.* 3–4 (2002) 411–446.
- [24] F. Schwade, P. Schubert, Social collaboration analytics for enterprise social software: a literature review, in: *Multikonferenz Wirtschaftsinformatik 2018*, Lüneburg, Germany, 2018.
- [25] S. Behrendt, A. Richter, M. Trier, Mixed methods analysis of enterprise social networks, *Comput. Netw.* 75 (2014) 560–577.
- [26] M. Steinhueser, C. Herzog, A. Richter, U. Hoppe, A process perspective on the evaluation of enterprise social software, in: 2nd Eur. Conf. Soc. Media, Porto, Portugal, 2015.
- [27] J.C. Recker, D. Lekse, A field study of spatial preferences in enterprise microblogging, *J. Inf. Technol.* 31 (2016) 115–129, <https://doi.org/10.1057/jit.2015.27>.
- [28] J. Hacker, K. Riemer, Identification of user roles in enterprise social networks: method development and application, *Bus. Inf. Syst. Eng.* (2020), <https://doi.org/10.1007/s12599-020-00648-x>.
- [29] K.B. Bøving, J. Simonsen, http log analysis - an approach to studying the use of Web-based information systems, *Scand. J. Inf. Syst.* 16 (2004) 145–174.
- [30] N. Jeners, W. Prinz, Metrics for Cooperative Systems, in: *GROUPE'14 18th Int. Conf. Support. Gr. Work.*, ACM, 2014, pp. 91–99.
- [31] P. Nasirifard, V. Peristeras, Expertise extracting within online shared workspaces, in: *WebSci'09 Soc. On-Line*, 2009.
- [32] S. Kudaravalli, S. Faraj, The structure of collaboration in electronic networks, *J. Assoc. Inf. Syst.* 9 (2008) 706–726, <https://doi.org/10.17705/1jais.00172>.
- [33] A. Xu, J. Chen, T. Matthews, M. Muller, H. Badenes, CommunityCompare: visually comparing communities for online community leaders in the enterprise, in: *Conf. Hum. Factors Comput. Syst. - Proc.*, 2013, pp. 523–532, <https://doi.org/10.1145/2470654.2470729>.
- [34] A. Wu, J.M. Dimicco, D.R. Millen, Detecting professional versus personal closeness using an enterprise social network site, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2010, pp. 1955–1964.
- [35] K. Ehrlich, C.-Y. Lin, V. Griffiths-Fisher, Searching for experts in the enterprise: Combining text and social network analysis, in: *Proc. 2007 Int. ACM Conf. Conf. Support. Gr. Work.*, 2007, pp. 117–126, <https://doi.org/10.1145/1316624.1316642>.
- [36] M. Langen, Social collaboration metrics, in: 11th Int. Symp. Open Collab, ACM, 2015, pp. 7:1–7:4, <https://doi.org/10.1145/2789853.2789860>.
- [37] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: step-by-step data mining guide, in: *CRISP-DM Consortium*, 2000.
- [38] V. Vaishnavi, B. Kuechler, Design science research in information systems overview of design science research, *Des. Sci. Res. Inf. Syst. Technol.* (2004).
- [39] R.B. Johnson, A.J. Onwuegbuzie, Toward a definition of mixed methods research, *J. Mix. Methods Res.* 1 (2007) 112–133, <https://doi.org/10.1177/155868906298224>.
- [40] J.W. Creswell, V.L.P. Clark, *Designing and Conducting Mixed Methods Research*, 2nd ed., Sage Publications, Inc., Thousand Oaks, California, USA, 2011.
- [41] D.L. Morgan, Basic and Advanced Focus Groups, 1st ed., SAGE Publications Inc, Thousand Oaks, California, USA, 2018 <https://doi.org/10.4135/9781071814307>.
- [42] S. Caillaud, U. Flick, Focus groups in triangulation contexts, in: Barbour, D. L. Morgan (Eds.), *A New Era Focus Gr. Res. Challenges*, Innov. Pract, Palgrave Macmillan, London, United Kingdom, 2017, pp. 155–177, https://doi.org/10.1057/978-1-137-58614-8_8.
- [43] E.J. Halcumb, P.M. Davidson, Is verbatim transcription of interview data always necessary? *Appl. Nurs. Res.* 19 (2006) 38–42, <https://doi.org/10.1016/j.apnr.2005.06.001>.
- [44] A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research, *MIS Q.* 28 (2004) 75–105.
- [45] A.R. Hevner, A three cycle view of design science research, *Scand. J. Inf. Syst.* 19 (2007) 87–92.
- [46] K.C. Tang, A. Davis, Critical factors in the determination of focus group size, *Fam. Pract.* 12 (1995) 474–475, <https://doi.org/10.1093/fampra/12.4.474>.
- [47] G. Thomas, A typology for the case study in social science following a review of definition, discourse, and structure, *Qual. Inq.* 17 (2011) 511–521.
- [48] P. Schubert, R. Wölfe, The experience methodology for writing IS case studies, in: *Am. Conf. Inf. Syst.*, 2007, pp. 1–15.
- [49] G. Mariscal, Ó. Marbán, C. Fernández, A survey of data mining and knowledge discovery process models and methodologies, *Knowl. Eng. Rev.* 25 (2010) 137–166, <https://doi.org/10.1017/S0269888910000032>.
- [50] P. McCullagh, What is a statistical model? *Ann. Stat.* 30 (2002) 1225–1267, <https://doi.org/10.1214/aos/1035844977>.
- [51] Y. Wand, R. Weber, On the deep structure of information systems, *Inf. Syst. J.* 5 (1995) 203–223, <https://doi.org/10.1111/j.1365-2575.1995.tb00108.x>.
- [52] J. Richardson, R. Sallam, K. Schlegel, A. Kronz, J. Sun, Magic Quadrant for Analytics and Business Intelligence Platforms, 2020.
- [53] F. Schwade, P. Schubert, Developing a user typology for the analysis of participation in enterprise collaboration systems, in: *Hawaii Int. Conf. Syst. Sci.*, Wailea, Hawaii, USA, 2019, pp. 460–469.
- [54] D. Arnott, G. Pervan, A critical analysis of decision support systems research revisited: the rise of design science, *J. Inf. Technol.* 29 (2014) 269–293.

Florian Schwade is postdoctoral researcher at the Institute of Information Systems Research at the University of Koblenz-Landau since April 2015. His research activities focus on business intelligence and collaboration software in the digital workplace. His primary research interest is Social Collaboration Analytics, the analysis of event logs of Enterprise Collaboration Systems. With his research, Florian Schwade aims to analyze and understand how employees adopt collaboration technologies and which collaboration behaviours emerge in the digital workplace. He has authored scholarly publications in various leading conference proceedings including ECIS, AMCIS, and HICSS. Since 2020,

Florian is co-chair of the “Collaboration for Data Science” mini-track at the Hawaiian International Conference on System Sciences.