11-28-2018

A Survey on the Status Quo of Social Collaboration Analytics in Practice

Florian Schwade
University of Koblenz-Landau, fschwade@uni-koblenz.de

Petra Schubert
University of Koblenz-Landau, petra.schubert@uni-koblenz.de

A SURVEY ON THE STATUS QUO OF SOCIAL COLLABORATION ANALYTICS IN PRACTICE

Research paper

Schwade, Florian, University of Koblenz-Landau, Koblenz, Germany, fschwade@uni-koblenz.de

Schubert, Petra, University of Koblenz-Landau, Koblenz, Germany, petra.schubert@uni-koblenz.de

Abstract

This paper aims to identify and document the status quo and the perceived barriers of Social Collaboration Analytics in practice. The study is part of a publicly funded longitudinal research project on Enterprise Collaboration Systems (ECS) involving 29 early-adopter user companies. Our paper draws on and extends findings from previous research. The key themes identified from a structured literature review and the documented findings from previous workshops with practitioners were used to develop an online questionnaire. Longitudinal case studies on the same companies helped to inform and interpret the findings. The survey shows that most of the participants have implemented some form of Collaboration Analytics, however, the outputs are mostly on a high abstraction level and the methods quite simplistic. Complex analyses that could help in assessing the degree of cooperation, structures or dynamics in the use of ECS are not (yet) applied. We identified three important reasons for this situation: (1) The analytics tools provided by standard collaboration software do not provide sufficient functionality, (2) there is a lack of awareness and knowledge about more complex forms of analyses and (3) some concepts from the academic literature are not perceived as relevant by practitioners. The survey also confirmed that there are financial, regulatory, data and HR barriers for Social Collaboration Analytics.

Keywords: Survey, Social Collaboration Analytics, SCA, Enterprise Social Software, Status Quo, Barriers

1 Research Motivation

Large-scale *Enterprise Resource Planning (ERP) Systems*, provided by companies such as SAP, Oracle or Infor, have been the focus of attention in IS Research for many years. ERP Systems focus on the support of business processes but not all work in organisations is *process-oriented.* In most companies, a large part of business activity is organised *in projects or tasks* that require ad hoc involvement of humans (i.e. communication and providing information) and thus, call for a different kind of Enterprise Information System. In this paper, we place our attention on *Enterprise Collaboration Systems (ECS)* that support employees in their daily work and facilitate functionality for work (group) collaboration (Schubert and Glitsch, 2016).

Most companies do not have an *integrated* software system for collaboration but follow a *portfolio approach* (Williams and Schubert, 2015) with many heterogeneous systems often from different software vendors (e.g. IBM, Atlassian and Microsoft) with the problem of redundant functionality (e.g. Wikis are included in both Atlassian Confluence and MS SharePoint). A small number of large companies, however, have introduced *integrated ECS* in recent years, which bundle a range of the required

collaborative features in a uniform system. One example of such an integrated ECS is IBM Connections. In this paper, we focus our attention on these large-scale, integrated ECS.

The increasing use of Social Media and their "social software features" in private life has changed the way people communicate and exchange information and has stimulated expectations on the side of employees regarding the use of similar software features in their workplace (Wehner, Falk and Leist, 2017). Large software vendors have responded to the perceived demand for socially-enabled software with *Enterprise Social Software (ESS)*, a special type of collaboration software that provides *social software features* (e.g. IBM Connections, MS Yammer/SharePoint or Atlassian Confluence/Jira). ESS has become an integral part of companies' Enterprise Collaboration Systems. Typical social features include subscribing/following information or people, commenting or tagging contributions, or short expressions such as recommendations or likes. These systems are often equipped with extensive "awareness features*"*, which help to recognize new and possibly relevant content. ESS provides the functionality to build up *Enterprise Social Networks* (ESN) (Wehner, Ritter and Leist, 2016).

Whilst companies see the potential of improving employee collaboration with ESS and ESN, the implementation and adoption of such systems has proven to be slower than expected (Williams and Schubert, 2015). A comparison with ERP systems offers an explanation for this phenomenon: ERP systems have strong prescriptive effects because business processes are embedded in their functionality. ECS, on the other hand, are characterised by their interpretive flexibility (Doherty, Coombs and Loan-Clarke, 2006) – they offer certain functionality but the user has to decide which feature to use and often there are many different possibilities to achieve the same task, e.g. documenting information or communicating a piece of information to a target group (Schubert and Glitsch, 2016). Our study shows that there is a perceived lack of tools that support community managers in their tasks of managing content and member activity. We argue that Social Collaboration Analytics (SCA) is necessary to provide feedback on what is happening in the ECS in order to purposefully manage platform activity and to stimulate engagement of employees.

The aim of this paper is to report on the status quo of Social Collaboration Analytics (SCA) in practice. This includes two aspects: (1) identifying the software tools, which user companies use for analytics in collaboration software, and (2) identifying the current use of analytics in user companies. The "use" includes the desire, knowledge, availability, barriers and actual use of analytics in companies. We use the definition of SCA proposed by Schwade and Schubert who define SCA as "*the approach for analysing and displaying collaboration activity of users in socially-enabled collaboration systems*" (Schwade and Schubert, 2017, p. 402). The authors suggest three main data sources for SCA: (1) *organisational data* (e.g. organisational structure), (2) *transactional data* (e.g. event records in databases or log files) and (3) *content data* (user-generated content).

The research presented in this paper builds on and extends previous research findings from (1) a structured literature review, (2) research workshops with collaboration professionals and (3) longitudinal case studies on the same companies. We used the key themes identified in a structured literature review to develop our survey instrument (an online questionnaire). The rich information from the longitudinal in-depth case studies additionally informed the interpretation of the survey responses and helped us to draw validated and detailed conclusions.

The remainder of this paper is structured as follows: We give a brief overview of the background of this study and present the seven key themes and the possible barriers to SCA identified in the literature that informed the questionnaire. In the next section, we describe the development of the survey instruments and the responses from the online questionnaire. This is followed by the discussion of the findings. The conclusions summarise the findings and suggest future research in the area of SCA.

2 Background of this Study

Enterprise Social Software (ESS) is a recent topic and many companies are still struggling with the implementation of their collaboration initiatives. In order to discuss the existing problems and engage in a mutual exchange of ideas and solutions, more than 20 early adopter companies formed a "user

group" named "IndustryConnect" (Williams and Schubert, 2017) in the year 2015, which is moderated by a group of University researchers. The companies are committed to providing the researchers with the necessary empirical data and they agreed to have their experiences documented in longitudinal case studies based on company visits, interviews and analysis of their actual system data. The practitioners in IndustryConnect are collaboration professionals with different educational backgrounds such as information technology, information and knowledge management, internal communications or business development.

The findings from each phase of the research are regularly presented at research workshops and discussed among the participants. This ensures a constant cycle of data collection, data interpretation and evaluation. At one of these workshops, the participants expressed an existing dissatisfaction with their current Social Collaboration Analytics, which encouraged us to investigate the reasons behind this. It was our intention to identify and document (1) the *status quo* of SCA and (2) to identify the *barriers* to SCA in our member companies. Thus, this paper addresses the following research questions:

1. Which software tools do organisations currently use for Social Collaboration Analytics?
2. What kind of information do organisations currently gain with the help of Social Collaboration Analytics?
3. What are the barriers that organisations are facing when conducting Social Collaboration Analytics?

Three results from previous research provided an essential input for the current study (Figure 1), which allowed us to triangulate and better interpret the findings from the survey:

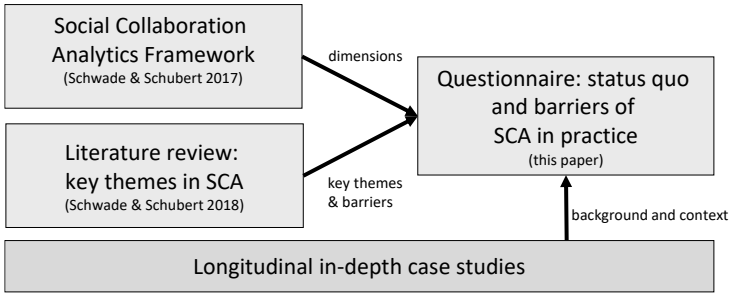

Figure 1: Study Design

(1) SCA Framework (Schwade and Schubert, 2017): The SCA framework contains a classification scheme for data and supports researchers and practitioners in developing questions and queries that can be posed to an existing data set. The dimensions of the framework were used for the questionnaire.

(2) Literature Review (Schwade and Schubert, 2018): A structured literature review was used to classify metrics and methods for SCA from the academic literature. The derived seven key themes and the identified barriers were used to structure and develop the questions for the questionnaire. Thus, the literature review provides an important foundation for this paper.

(3): Longitudinal case studies (Williams and Schubert, 2017): Since the start of the IndustryConnect initiative the adoption projects of the member companies have been documented in case studies. The information contained in the project description helped with the interpretation of the responses from the survey. The case studies thus provide the context, especially for the interpretation and discussion of the survey results.

3 Social Collaboration Analytics in the Literature

The following sections provide an overview on the results from the literature review consisting of the seven key themes for SCA (Schwade and Schubert, 2018) and barriers for SCA.

3.1 Seven Key Themes for SCA

The key theme with the largest number of publications is the *(1) measurement of system usage*. *General usage figures*, such as the number of users, logins, events or objects are frequently discussed (Shami, Muller and Millen, 2011; Herzog et al., 2013; Herzog, Richter and Steinhueser, 2015; Steinhueser, Herzog, Richter and Hoppe, 2015). Such metrics provide a high-level overview on platform usage. In contrast to this, *action specific measures* can be more detailed indicators on how a platform is used because to a certain extent events are aggregated to types like creating information, modifying information and reading information (Appelt, 2001; Jeners, Lobunets and Prinz, 2013; Richter, Heidemann, Klier and Behrendt, 2013; Hacker, Bodendorf and Lorenz, 2016). This also allows differentiating between *producing* and *consuming* activities (Muller et al., 2009). A further aspect is the *measurement over time*, which allows the tracking of changes in system usage (Muller et al., 2009; Richter et al., 2013; Hacker et al., 2016).

The *(2) analysis of communities* is of particular importance in the context of community management, however, only few studies focus on this key theme. Jeners and Prinz (2014) propose metrics for measuring the *productivity*, *activity* and *cooperativity* in shared workspaces. The authors describe how the degree of labour division and the responsiveness of workspaces can be measured. Besides measuring community activity, the literature proposes to measure *community health* (Matthews et al., 2013).

The key theme *(3) identification of types of users* is related to the previously mentioned action-specific measures. The literature contains two main approaches for identifying types of users: The first one is based on the *intensity of platform usage* (platform access or frequency of activities). This allows for example to distinguish between lurkers, inactive users, active users and power users (Appelt, 2001; Jeners et al., 2013). More recent approaches suggest a second type which is based on the *nature of system usage* (type of actions on the platform). For example, Hacker et al. (2017) suggest identifying types of users based on a knowledge worker role typology (e.g. helper, sharer, seeker).

Recent surveys on ESS implementations suggest that organisations introduce such systems for supporting access to knowledge and experts (Williams and Schubert, 2015). Accordingly, the *(4) identification of expertise* in ESS is another key theme for SCA in ESS. Expertise can be assigned to users based on (inter)actions with content. The domain of expertise of a particular document can be defined either based on keyword extraction or based on the assigned tags (John and Seligmann, 2006; Nasirifard and Peristeras, 2009).

A more specialised key theme is the *(5) identification of usage patterns* in ESS. This theme contains discussions on different types of usage patterns. Millen et al. (Millen, Feinberg and Kerr, 2006) investigate *search patterns* by analysing clickstreams, whereas Ferron et al (2011) analyse *communication and networking patterns*. In contrast to this, Chaves and Córdoba (2014) and Naderipour (2011) apply process mining to analyse usage patterns in ESS.

The *(6) analysis of networks* came out as one of the most popular key themes of SCA. Two perspectives on network structures could be identified in the literature. Research in this area makes use of methods and networks from Social Network Analysis. *User-centric networks* are based on the relationships between users. These networks are used by researchers to characterise types of networks (Smith, Hansen and Gleave, 2009), analyse communication (Behrendt, Richter, Schäfer and Trier, 2015) or to investigate the connectedness of users (Steinhueser et al., 2015; Hacker et al., 2016). In contrast to this, *document-centric networks* are built based on interactions of users with documents. The analysis of such networks provides insights in actual collaboration behaviour (Nasirifard, Peristeras, Hayes and Decker, 2009).

It is notable that the number of publications that investigate *(7) organisational and cultural impacts* of ESS on organisations has been growing in recent years. A particular phenomenon of interest is the *impact of hierarchies on activities in ESS* (Stieglitz, Riemer and Meske, 2014; Behrendt, Klier, et al., 2015; Riemer, Stieglitz and Meske, 2015). There is also an interest in the *effects of geographical dispersion* on activities in ESS (Warshaw, Whittaker, Matthews and Smith, 2016) and intra-organisational *boundary spanning* (van Osch and Steinfield, 2013; Steinfield, van Osch and Zhao, 2015).

3.2 Barriers for SCA

The literature on data analytics in ESS contains numerous barriers to SCA, which we categorised related to (1) resources, (2) methods, (3) governance and compliance, (4) technical and (5) data barriers. The categories are based on Herzog et al. (2014). The barriers are discussed in the next sections.

3.2.1 Resources

The category resources consists of constraints related to resources in terms of time, money and effort. Herzog et al. (2013) found that practitioners often need to justify financial investments by means of measuring system usage. The authors argue that these *investments can mitigate the benefits*, which makes practitioners shy away from conducting measurements. In some cases it is necessary to *manually collect data* for the measurements, which further increases costs (Otjacques, Noirhomme, Gobert and Feltz, 2006). Other authors argue that in general, software log file analysis is a *complex and labour intensive task* (Xhafa, Caballe, Daradoumis and Zhou, 2004; Jayathilake, 2011; Chounta and Avouris, 2012), which requires a lot of time. Additionally, log file analysis can be *error prone* (Jayathilake, 2011), especially when steps have to be carried out manually. Consequently, such analyses can often not be conducted because of *insufficient staff resources* (Xhafa et al., 2004; Herzog et al., 2013).

The constraints mentioned in the literature show that measuring system usage is a cost- and time-intensive task that requires adequate resources in an organisation. Consequently, if an organisation wants to establish a continuous SCA, personnel and finances need to be allocated to this task.

3.2.2 Methods

The literature review revealed a variety of *methodological* constraints. According to Herzog et al. (2013), common mistakes encountered are *incorrect or missing targets for measurements*, which result in unprecise and unfocused measurements. The data mining of collaborative activity is *different from mining email activity* (John and Seligmann, 2006). Herzog et al. (2013) observe that practitioners sometimes *do not consider current approaches as useful*. Similarly, Meske et al. (2014) identify a *lack of overview of methods*. Consequently, the authors are calling for new methods for measuring ESS usage.

In addition to the methodological constraints, researchers and practitioners are struggling with metrics for measuring ESS usage. A frequent challenge is the *conceptualisation and development of metrics*. Log file analysis requires carefully designed metrics, which provide new information and insights on system usage. It is argued, however, that there is a lack of *meaningful metrics* (Otjacques et al., 2006; Meske et al., 2014; Steinhueser et al., 2015). The *interpretation of these metrics* is a further and directly related issue (Otjacques et al., 2006; Windhager, Zenk and Federico, 2011; Meske et al., 2014). Authors consistently argue that ESS log file analysis *requires methodological expertise* (Bøving and Simonsen, 2004; Otjacques et al., 2006; Jayathilake, 2011; Windhager et al., 2011).

3.2.3 Governance and Compliance

In European countries in particular, *regulations introduced by the workers' council* prohibit the analysis of system log files for measuring system usage. These regulations aim to prevent misuse of personal information and restrain organisations from assessing performance of individuals (Steinhueser et al., 2015). *Labour laws* have a similar effect all over the world (Lin et al., 2012). Organisations increasingly use ESS to communicate with external partners. However, in some cases *policies forbid to analyse activities of users by these external parties* (Friedman, Burns and Cao, 2014). Several authors argue that *privacy laws and privacy policies* constrain or even forbid system log file analyses because of a possible violation of the employees' privacy (John and Seligmann, 2006; Lin et al., 2012; Muller, 2012; Cao, Gao, Li and Friedman, 2013; Herzog et al., 2013; Riemer, Finke and Hovorka, 2015; Steinhueser et al., 2015). The authors consistently argue that privacy laws are one of the biggest obsta-

cles for conducting system measurements. *Prevention of access to actual content* data is a further issue in this context (Riemer, Finke, et al., 2015).

3.2.4 Technical (system issues)

While the previously introduced constraints were of non-technical nature, there are a number of technical barriers. Some systems only allow access to data via APIs. However, these *APIs often provide limited data access* (Friedman et al., 2014). Some ESS *do not provide any logging capabilities* (Shami et al., 2011) or the *system does not allow to access log files* (Xhafa et al., 2004), e.g. in the case of Software-as-a-Service. A number of authors state that several systems *only allow to access public data for analyses*, which means that activities in private groups cannot be analysed (Muller et al., 2012; Cao et al., 2013; Friedman et al., 2014). A further technical challenge is that system *log files are sometimes not complete* because data is purged after a certain period. However, probably the biggest challenge for SCA is that even if systems allow full access and analysis of data, *ESS usually only provide very rudimentary standard functionality for the analysis of log files* (Benhiba and Janati-idrissi, 2013; Herzog et al., 2014; Steinhueser et al., 2015; Schwade and Schubert, 2017). This means that specialised programming skills are required to implement the desired analyses.

3.2.5 Data

The limitations of *data* are closely related to the methodological constraints. A common problem that researchers and practitioners face is that *datasets can be noisy* (Nasirifard and Peristeras, 2009; Smith, Shneiderman, et al., 2009; Tzagarakis et al., 2014). Noise can have different characteristics. Nasirifard and Peristeras (2009) describe noisy records as records that do not have a userID and objectID or an event name. A closely related problem is that datasets can also contain *redundant data* (Smith, Shneiderman, et al., 2009). These two constraints have in common that the data needs to be carefully prepared before the analysis can be conducted. However, before the data can be prepared and analysed, the *data needs to be understood* (Friedman et al., 2014). In some cases there can be a *lack of comparative data* which means that measurements cannot be compared with another measurement or a completely different dataset (Herzog et al., 2013). The *currency of data* is another relevant issue (Ehrlich, Lin and Griffiths-Fisher, 2007). When analysing data, researchers and practitioners have to consider that *lurking data in log files is often sparse* (Ehrlich et al., 2007) and that in some systems *anonymous users have to be excluded from the analysis* because they cannot be uniquely identified (Appelt, 2001). A possible explanation for these limitations is that ESS log files are not problem-free because the *logs are not produced for analysis* (Bøving and Simonsen, 2004). Further, there is an increasing amount *of multiple types of data from heterogeneous sources*, which adds another layer of complexity (Tzagarakis et al., 2014)

4 Survey Development and Findings

The following sections describe the development and the results of the survey.

4.1 Development of the Questionnaire

The questions within this survey were developed based on the classifications from the literature review presented above. The survey consists of two parts. The first part covers the status quo of SCA (i.e. implemented SCA). The second part examines barriers that organisations face regarding analyses. The participants were able to add further aspects of the key themes and barriers in an open text field. Due to the amount of items in each key theme, the first part of the survey was split into eight sections (general information and the seven key themes). The following table shows an excerpt of how the questions for the survey were derived from the literature.

Aspect	Description	Answer options	References from the literature
General usage measures	General usage statistics provide an overview of platform usage at an abstract level. This includes for example the total number of events, logins, users, blogposts, files etc.	Yes, is measured. Yes, is measured but the output is not satisfactory. No, is not measured.	*General usage measures* (Shami et al., 2011; Herzog et al., 2013, 2015; Steinhueser et al., 2015)

Table 1. Overview on how questions were derived from the literature (excerpt)

As shown in the table, the participants were provided with the keywords of each concept (e.g. general usage measures). For clarification, a brief description of every concept was displayed. The answers were designed with the intention to distinguish between implemented analyses, which are *satisfactory* and analyses that are *not satisfactory*. All answer options in both sections of the surveys were on an ordinal scale.

As mentioned in the introduction, we invited all members of IndustryConnect to participate in the survey. At the time of writing of this paper, 24 questionnaires had been fully completed and could be used for the analysis. The results are presented in the following two sections.

4.2 Available Tools and Key Themes

This section reports on the first part, the status quo of SCA in practice. Table 3 provides an overview of the results of this section of the survey. In the following, the findings are described and discussed.

In the first question of the survey, the participants were asked about the tools that are available for analytics in their organisation from a list of software products (cf. Table 2). A combination with the option "No analyses" was possible to identify the cases in which analytics software is available but not used. When "No analyses" was selected, the questions on the status quo of SCA were skipped and the participants were immediately directed to the section on perceived barriers.

IBM Connections Metrics	Kudos Analytics	Cognos for IBM Connections	Panagenda Experts	Custom analyses	Web Analytics tools	Analysis of exported data in Excel/SPSS	Subjective evaluation	Oracle APEX	No Analyses
12	2	5	3	1	5	3	7	1	5

Table 2. Summary of results for question: Which tools are available in your organisation for analytics? (n=24)

As shown in the result table, 12 participants stated that the metrics application of IBM Connections is available, which is the majority of participants. This is the default application for analytics in IBM Connections. 7 participants stated that they assess the usage of the system based on subjective evaluation of the system. 2 out of these 7 only apply subjective evaluations without a software tool. In total, 5 participants stated that they do not conduct any analyses. The following table provides an overview of the results of the first part of the survey. Some questions are dependent on previous responses, so the N is sometimes smaller than 19. For example, if a participant stated that there are no analyses on community activity, the question on comparison of communities was not shown.

Question	Yes, is measured.	Yes, is measured but the output is not satisfactory	No, is not measured.	N
General system usage				
General usage measures	63% (12)	21% (4)	16% (3)	19
Usage specific measures	10% (2)	16% (3)	74% (14)	19
Participation inequality	40% (2)	20% (1)	40% (2)	5

Quality of collaboration	6% (1)	6% (1)	88 (14)	16
Dynamics of ideas and innovations generation	0	20% (1)	80% (4)	5
Measurement of platform usage over time	37% (7)	37% (7)	26% (5)	19
Community analyses				
Community activity	21% (4)	26% (5)	53% (10)	19
Community health	5% (1)	16% (3)	79% (15)	19
Community comparison	22% (2)	0	78% (7)	9
Identification of types of users				
Types of users based on intensity of usage	5% (1)	11% (2)	84% (16)	19
Types of users based on nature of usage	0	0	100% (19)	19
Identification of expertise				
Identification of expertise	0	0	100% (19)	19
Ranking of expertise	0	0	100% (19)	19
Influence of experts	0	0	100% (19)	19
Usage patterns				
Social Process Mining	0	0	100% (19)	19
Identification of communication patterns	0	5% (1)	95% (18)	19
Identification of search patterns	0	11% (2)	89% (17)	19
Network Analysis				
Connectedness of users	10% (2)	16% (3)	74% (14)	19
Analysis of document-centric networks	0	0	100% (19)	19
Analysis of networks over time	0	20% (1)	80% (4)	5
Analysis of informal networks	0	0	100% (5)	5
Organisational and cultural impacts				
Effects of geographical dispersion	0	0	100% (19)	19
Boundary spanning	0	0	100% (19)	19
Impacts of hierarchy	0	0	100% (19)	19

Table 3. Overview of the results of the currently implemented analyses

63% (12) of the respondents state that general usage measures are applied to their system. 21% (4) of the participants mentioned that they do not apply *general usage measures*. The next question asked the participants about the application of *usage specific measures*. However, only a minority of participants (10%; 2) indicated that usage specific measures are applied. With 74% (14) of the participants, the majority does not apply such metrics. This is a striking observation because already at this point of the survey this indicates that sophisticated analyses of ESS are not applied in practice. This becomes even more noticeable in the remainder of the survey. In response to the question on *community activity*, only 21% (4) of the respondents indicate that community activity is measured. 26% (5) of the participants state that the outputs of the analyses are not satisfactory. Consequently, 53% (10) of the participants reported that the activity of communities is not measured. Out of the 9 respondents who measure community activity, only 2 participants state that the activities of *communities are compared*. This is quite surprising because measuring the activity of communities is critical for community management.

The analysis of *different types of users* based on intensity of platform usage is only applied by 1 participant. 10% (2) of the participants stated that they are not satisfied with the output of the analyses. Similar results can be seen in the section on the identification of expertise of users and the identification of usage patterns. Only 1 participant (5%) states that *communication patterns* in the ESS are investigated but the output is not satisfying. *Search patterns* are analysed by 2 (11%) participants which demon-

strates a very low adoption of sequence-oriented analyses. Even *network analyses*, which are much discussed in the academic literature, show a low adoption in practice.

A further analysis of the responses revealed that three of the participating organisations are pioneers in the area of SCA and they are developing custom analyses beyond those provided by the software. Compared to the rest of the participants, these organisations have implemented more analyses. Notably, three organisations frequently stated that they are not satisfied with the outputs of the analysis. These participants have in common that their educational and professional background is in data science. Consequently, it can be concluded that these participants are aware of and know that the full potential of SCA has not yet been unleashed.

In summary, 79% (19) of the respondents have implemented some kind of analyses for their ESS. However, the evaluation of the first part of the survey reveals that these analyses are implemented on a very high abstraction level. Specific analyses or sequence analyses have not been implemented yet. Further, it can be observed that for a majority of the questions there is a considerable amount of participants that indicate that they are not satisfied with the outputs of the analyses.

4.3 Barriers to Social Collaboration Analytics

This section reports on the results of the section on the barriers for SCA. Table 4 provides an overview of the results of this part of the survey. In the following, these results are presented and discussed. This section of the survey was shown to all participants. Consequently, N is 24 for all questions. Due to space limitations, we focus the presentation and discussion on the barriers in the categories resources, methods, governance & compliance and data.

Barrier	Is not a barrier	Hinders SCA	Prevents SCA	N
Resources				
Negative cost-value ratio	29% (7)	54% (13)	17% (4)	24
Capacity: Personnel	17% (4)	54% (13)	29% (7)	24
Capacity: Financial	21% (5)	58% (14)	21% (5)	24
High effort for data collection	34% (8)	58% (14)	8% (2)	24
High effort for data preparation	25% (6)	67% (16)	8% (2)	24
High effort for data analysis	25% (6)	63% (15)	13% (3)	24
Methods				
Missing targets for measurements	34% (8)	41% (10)	25% (6)	24
Development of metrics	34% (8)	58% (14)	8% (2)	24
Meaningful metrics	25% (6)	63% (15)	13% (3)	24
Interpretation of metrics	38% (9)	50% (12)	13% (3)	24
Current approaches are not helpful	42% (10)	38% (9)	21% (5)	24
Lack of overview on methods	21% (5)	63% (15)	17% (4)	24
Governance and Compliance				
Regulations by workers' council	13% (3)	63% (15)	25% (6)	24
Privacy regulations	13% (3)	71% (17)	17% (4)	24
Company policy: external users must be excluded from analyses	54% (13)	29% (7)	17% (4)	24
Company policy: content data must not be analysed	54% (13)	33% (8)	13% (3)	24
Technical				
System does not allow access to log file	65% (15)	22% (5)	13% (3)	23

Barrier	Is not a barrier	Hinders SCA	Prevents SCA	N
System does not allow access to content data	65% (15)	26% (6)	9% (2)	23
API provide limited data access	48% (11)	30% (7)	22% (5)	23
Limited analytics standard features	26% (6)	43% (10)	30% (7)	23
Lack of skills for developing queries	35% (8)	57% (13)	9% (2)	23
Data				
Dataset is noisy	35% (8)	65% (15)	0	23
Dataset contains redundant data	26% (6)	74% (17)	0	23
Undocumented logs	22% (5)	57% (13)	22% (5)	23
Lack of comparative data	30% (7)	57% (13)	13% (3)	23
Currentness of data	65% (15)	30% (7)	4% (1)	23
Incomplete data	26% (6)	65% (15)	9% (2)	23
Data from multiple sources	22% (5)	65% (15)	13% (3)	23

Table 4. Overview of the results on barriers

The analysis of this section of the survey reveals interesting insights into why many analyses that are suggested in the literature are not adopted in practice. The survey revealed that resources are a barrier when applying SCA. 29% (7) of the participants state that human resource capacities prevent them from SCA. Further 54% (13) answered that this hinders them. Only 17% (4) stated that *human resource capacities* are no issue. These results are in line with the results of the barriers high efforts for data collection, data preparation and data analysis. 58% (14) of the participants stated that the *high effort for data collection* hinders their organisation in carrying out analyses. 67% (16) of the participants answered that the *high effort for data preparation* hinders them from analyses. 63% (15) of the participants stated this for the *high effort for data analyses*. These results illustrate that SCA is a task, which requires dedicated personnel, which in turn is a barrier to conducting SCA. The survey also confirmed financial capacities as a barrier for SCA. However, this barrier is not as severe as personnel capacities. 21% (5) of the participants stated that a *lack of financial resources* prevents SCA in their organisation. 58% (14) stated that this hinders SCA in their organisation. For 21% (5) financial capacities are not a barrier.

Remarkably, only 34% (8) of the participants stated that clear *targets for measurement*s exist. This result demonstrates a high uncertainty concerning the objectives of analyses. For 25% (6) of the participants this even prevents SCA completely. These findings are in line with the statements about the development of metrics. Only 34% (8) of the participants stated that the *development of metrics* is no barrier for them. This further confirms the uncertainty about what is to be measured.

Further interesting observations can be made in the category governance and compliance. Only 13% (3) of the participants stated that *regulations by the workers' council* are not a barrier for them. 25% (6) of the participants stated that regulations by the workers' council are a barrier for them. The barrier *privacy laws/ regulations* appears to be even more severe. 13% (3) of the participants stated that this is no issue for them. However, 71% (17) of the participants stated that privacy is a barrier for them. This confirms that regulations by the workers' council and privacy laws are a severe barrier for SCA. In the context of these two barriers, it has to be emphasised that the participating organisations of this survey are either German or Swiss. The strong position of the workers' council is a unique feature in Germany. Similarly, in Germany and Europe privacy laws are stronger compared to for example in the USA. Therefore, it can be expected that the responses to these questions would look different in a multinational sample. The results of the survey exemplify the power and influence of the workers' council. One IndustryConnect member has subsidiaries in Germany *and* Switzerland. The workers' council in Germany is blocking the implementation of SCA. Consequently, this prevents the implementation for the subsidiary in Switzerland, too. One of the respondents is completely based in Switzerland and for them the workers' council is no barrier for SCA. The *workers' council* has been in the centre of many

discussions in the IndustryConnect workshops. Some of the participants have reported that the "barrier of the workers' council" can be successfully addressed by actively involving them in the implementation project.

The results from the category *data* match with what was identified in the previous sections. It is evident from the results that noisy and redundant data is an issue when conducting SCA. Only 35% (8) of the participants stated that *noisy data* is not an issue for them. For *redundant data* this applies to 26% (6) of the participants. Noisy and redundant data leads to an increased effort for data preparation. This in turn requires financial and human resources, which have been shown to be major barriers for SCA. The activity *logs of ECS are mostly undocumented*. Consequently, it is challenging to judge the result of analyses without a detailed analysis of the underlying data in the logs. In most cases, activity logs of ECS are not produced for the purpose of SCA (Behrendt, Richter and Trier, 2014). Only 22% (5) of the participants stated that this is no barrier for them. Thus, data can be a considerable barrier for conducting SCA.

5 Discussion

The analysis of the survey shows that most of the practitioners have implemented some kind of analyses for their ESS. However, the results also demonstrate that for the most part only simple and high-level analyses are used in practice, which results in a lack of information about platform usage for the platform owners. As shown in Table 3, more than 80% of the participants apply general usage measures (such as counting numbers of events, users and posts) to some extent. However, when looking at the implementation of more specific analyses such as action-specific measures (e.g. create, read edit actions) it becomes obvious that the results of the analyses are not very detailed. Only 25% of the respondents have applied such analyses. This becomes even more obvious for analyses, which aim to detect sequences, volumes and dynamics such as process mining and network analyses. From this part of the survey, it can be concluded that most of the analyses and measurement concepts that were identified from the academic literature are not applied in practice. One of the main objectives of SCA is to assess the *cooperativity* of ESS, which the currently implemented analyses do not allow. The responses show a lack of complex metrics or combined metrics that define relations between constructs. Consequently, community and platform owners both have to live with very little information on how the platform is actually used.

There might be several reasons for this: (1) In the workshops on SCA, the participants explicitly stated that they are not satisfied with the *standard analytics features* that their platform (currently) provides. This finding from the workshops was fully confirmed by the survey. The minority of the respondents state that they are satisfied with the analytics tools. In combination with the first question on available tools for analytics, it becomes evident that only standard tools for analytics have been implemented, which do not support advanced analysis. (2) Additionally, in the past workshops we noted a *lack of awareness* concerning the possibilities of SCA on the side of the practitioners. This motivated us to develop a series of exemplary prototypes for SCA, which we presented to the practitioners at several IndustryConnect workshops. It could be observed that such prototypes stimulate interest and increase the assessment ability for what is possible with SCA. (3) A further reason for the mismatch between the issues of SCA described in the literature and what the survey showed about actual use in practice is that some of the analyses described in the literature might not address the information needs of practitioners and might thus be of mere academic interest. The academic literature describes many possible analyses and measurement concepts. The survey confirmed that practitioners are not aware of many of these possibilities. The IndustryConnect Initiative aims at bridging the gap and demonstrating the possibilities of SCA on a broader scale thus trying to develop analyses that are relevant to both academia and practice. Our work is focussed on the intersection of the three questions (1) "what would we like to measure? (2) "what is possible to measure?" and (3) "what are we allowed to measure?".

The discussion in the IndustryConnect workshops have been circling around one essential question, which has also been examined in this survey: Is there a need to justify investments in ESS by means of SCA in companies? In the survey, only 3 participants stated that they have no need for SCA. A state-

ment from the survey provides a further example in this direction: "*We plan to stop all analytics. On the one hand, the data collection (without analysis) turned out to be a critical failure during the migration to Connections 6.0 (errors, time and data volume of the migration). On the other hand, the benefit of the platform has already been proven and does not need to be justified any further. This is why analyses are discontinued. In the early phases Piwik was important for project managers, however it has lost importance*". We believe that the discussions in the workshops and this statement demonstrate that many collaboration professionals are not fully aware of the possibilities of SCA. The interpretation and semantic evaluation of results is a further critical aspect of SCA. Again, the survey confirmed that this is an issue for practitioners, too.

Further, the survey shows that only a minority of the participants has taken action on the basis of the results from SCA. Only one participant stated that regular actions are taken based on the SCA results (cf. Table 5). This is in line with the finding that the majority of participants does not have clear targets for the analyses.

	Yes, actions are regularly taken.	**Yes, actions are taken selectively.**	**No actions were taken, yet.**	**N**
Taking action based on analyses	6% (1)	47% (9)	47% (9)	19

Table 5. Results for the question "Are actions taken based on the results of analyses?"

In accordance with the PDSA cycle, SCA can only be effective once the cycle is closed at all stages and measures are derived from the observations (Deming, 1982).

6 Conclusions and Future Research

This paper reports on the findings from a survey examining the status quo and possible barriers of SCA in practice that were drawn from the literature and tested with a group of collaboration professionals. The small sample size of only 24 companies might seem like a limitation, however, we believe that we had excellent conditions for this survey because we are able to triangulate the responses with findings from interactive workshops with the practitioners and from longitudinal in-depth studies on their specific ESS projects.

The survey revealed that SCA is not widely adopted in practice and that the currently implemented analyses are on a high abstraction level. The survey shows that three IndustryConnect participants are ahead of the rest of the group with regard to SCA. The section on barriers for SCA revealed that SCA is a time- and cost-intensive task for organisations, which makes them shy away from SCA. The diversity in the answers on the barriers indicates that the individual context of an organisation is an influencing factor.

Over the last two years whilst conducting this research on SCA in IndustryConnect we have noticed a shift in the perception of this focus topic in the workshops. In the first workshops, the participants took a rather critical stance on SCA and the topic did not seem of great interest to them. One participant strongly argued that "*It is less relevant to find metrics. Instead we should look for Use Cases*." Whilst the communication of Use Cases is an important method to stimulate ESS adoption (Schubert and Glitsch, 2016), we believe that SCA provides an invaluable feedback on user behaviour. SCA is an essential tool for community managers and collaboration professionals who strive to purposefully improve the use of the ESS and thus protect their company's investment in this technology. There is no doubt that the level of interest in SCA is rising in the IndustryConnect community. We believe that this is due to constant presentations and discussions of prototypes (analyses and dashboards). We conclude that practitioners need to be better informed on the possibilities of SCA in order to enable them to fully assess the potential. We strongly argue that SCA is more than just counting and collecting numbers. More complex metrics that assess the cooperativity of ESS are necessary. Further, the interpretation and semantic analysis of the results will play an important role if we want to implement measures to improve platform use.

References

Appelt, W. (2001). "What groupware functionality do users really use? Analysis of the usage of the BSCW system." In: *Euromicro Workshop on Parallel and Distributed Processing* (pp. 337–341). Mantova, Italy.

Behrendt, S., M. Klier, J. Klier, A. Richter and K. Wiesneth. (2015). "The Impact of Formal Hierarchies on Enterprise Social Networking Behavior." In: *36th International Conference on Information Systems* (pp. 1–19).

Behrendt, S., A. Richter, S. Schäfer and M. Trier. (2015). "Business Intelligence 2.0." In: *Business Intelligence for New-Generation Managers* (pp. 97–111). Cham, Heidelberg, New York, Dordrecht, London: Springer International Publishing.

Behrendt, S., A. Richter and M. Trier. (2014). "Mixed methods analysis of enterprise social networks." *Computer Networks*, *75*, 560–577.

Benhiba, L. and M. A. Janati-idrissi. (2013). "Supercharging Enterprise 2.0 Analytics Modules with SNA Metrics." *International Journal of Computer Science Issues*, *10*(6), 89–96.

Bøving, K. B. and J. Simonsen. (2004). "Http Log Analysis." *Scandinavian Journal of Information Systems*, *16*, 145–174.

Cao, J., H. Gao, L. E. Li and B. Friedman. (2013). "Enterprise social network analysis and modeling: A tale of two graphs." In: *32nd IEEE International Conference on Computer Communications* (pp. 2382–2390). Turin, Italy.

Chaves, M. A. and E. R. Cordoba. (2014). "Deciphering event logs in SharePoint Server: A methodology based on process mining." In: *XL Latin American Computing Conference (CLEI)* (pp. 1–12). IEEE.

Chounta, I.-A. and N. Avouris. (2012). "Time series analysis of collaborative activities." In: *18th International Conference on Collaboration and Technology* (pp. 145–152).

Deming, W. E. (1982). *Out of the Crisis*. Cambridge, MA: Massachusetts Institute of Technology.

Doherty, N. F., C. R. Coombs and J. Loan-Clarke. (2006). "A re-conceptualization of the interpretive flexibility of information technologies: redressing the balance between the social and the technical." *European Journal of Information Systems*, *15*(6), 569–582.

Ehrlich, K., C.-Y. Lin and V. Griffiths-Fisher. (2007). "Searching for experts in the enterprise: Combining text and social network analysis." In: *Proceedings of the 2007 International ACM Conference on Conference on Supporting Group Work* (pp. 117–126).

Ferron, M., P. Massa and F. Odella. (2011). "Analyzing collaborative networks emerging in Enterprise 2.0: the Taolin Platform." In: *4th & 5th UK Social Networks Conference* (pp. 68–78). Trento, Italy.

Friedman, B. D., M. J. Burns and J. Cao. (2014). "Enterprise Social Networking Data Analytics Within Alcatel-Lucent." *Bell Labs Technical Journal*, *18*(4), 89–109.

Hacker, J., F. Bodendorf and P. Lorenz. (2016). "A Framework to Analyze Enterprise Social Network Data." In: M. Atzmueller, O. Samia, & T. Roth-Berghöfer (Eds.), *Enterprise Big Data Engineering, Analytics, and Management* (pp. 84–107). Hershey, USA: IGI Global.

Hacker, J., F. Bodendorf and P. Lorenz. (2017). "Helper, Sharer or Seeker? – A Concept to Determine Knowledge Worker Roles in Enterprise Social Networks." In: *Proceedings der 13. Internationale Tagung Wirtschaftsinformatik (WI)* (pp. 668–682). St. Gallen, Switzerland.

Herzog, C., A. Richter and M. Steinhueser. (2015). "Towards a Framework for the Evaluation Design of Enterprise Social Software." In: *International Conference on Information Systems* (pp. 1–20). Fort Worth, USA.

Herzog, C., A. Richter, M. Steinhüser, U. Hoppe and M. Koch. (2013). "Methods and metrics for measuring the success of Enterprise Social Software – what we can learn from practice and vice versa." In: *21st European Conference on Information Systems* (pp. 1–12).

Herzog, C., A. Richter, M. Steinhüser, U. Hoppe and M. Koch. (2014). "Barrieren der Erfolgsmessung von Enterprise Social Software." *Multikonferenz Wirtschaftsinformatik (MKWI) 2014*, (1), 1682–1694.

Jayathilake, P. W. D. C. (2011). "A novel mind map based approach for log data extraction." In: *6th International Conference on Industrial and Information Systems* (pp. 130–135). Sri Lanka.

Jeners, N., O. Lobunets and W. Prinz. (2013). "What groupware functionality do users really use?: A study of collaboration within digital ecosystems." In: *7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)* (pp. 49–54).

Jeners, N. and W. Prinz. (2014). "Metrics for Cooperative Systems." In: *GROUP'14 18th International Conference on Supporting Group Work* (pp. 91–99). ACM.

John, A. and D. Seligmann. (2006). "Collaborative tagging and expertise in the enterprise." In: *Collaborative Web Tagging Workshop*. Edinburgh, UK.

Lin, C.-Y., L. Wu, Z. Wen, H. Tong, V. Griffiths-Fisher, L. Shi and D. Lubensky. (2012). "Social network analysis in enterprise." *Proceedings of the IEEE*, *100*(9), 2759–2776.

Matthews, T., S. Whittaker, H. Badenes, B. a Smith, M. Muller, K. Ehrlich, … T. Lau. (2013). "Community Insights : Helping Community Leaders Enhance the Value of Enterprise Online Communities." In: *SIGCHI conference on human factors in computing systems* (pp. 513–522). Paris, France.

Meske, C., S. Stieglitz and D. Middelbeck. (2014). "Mehrwerte von Intranet Social Soft- ware – Status quo in der Wissenschaft." In: *Multikonferenz Wirtschaftsinformatik* (pp. 1775–1785).

Millen, D. R., J. Feinberg and B. Kerr. (2006). "Dogear: Social bookmarking in the enterprise." In: *SIGCHI Conference on Human Factors in Computing Systems* (pp. 111–120). ACM.

Muller, M. (2012). "Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media." In: *Conference on Computer Supported Cooperative Work* (pp. 253–256). Bellevue, Washington, USA.

Muller, M., K. Ehrlich, T. Matthews, A. Perer, I. Ronen and I. Guy. (2012). "Diversity among enterprise online communities: collaborating, teaming, and innovating through social media." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2815–2824).

Muller, M., J. Freyne, C. Dugan, D. R. Millen and J. Thom-Santelli. (2009). "Return On Contribution (ROC): A metric for enterprise social software." In: *11th European Conference on Computer Supported Cooperative Work (ECSCW)* (pp. 143–150). Vienna, Austria.

Naderipour, F. (2011). "Mining of ad-hoc business processes using Microsoft Sharepoint, NITRO & PROM 6.0." In: *13th International Conference on Enterprise Information Systems* (pp. 413–418). SciTePress - Science and and Technology Publications.

Nasirifard, P. and V. Peristeras. (2009). "Expertise extracting within online shared workspaces." In: *Proceedings of the WebSci'09: Society On-Line*.

Nasirifard, P., V. Peristeras, C. Hayes and S. Decker. (2009). "Extracting and Utilizing Social Networks from Log Files of Shared Workspaces." In: *10th IFIP Working Conference on Virtual Enterprises* (pp. 3–11).

Otjacques, B., M. Noirhomme, X. Gobert and F. Feltz. (2006). "Cooperation Indexes to Support Workspace Awareness." In: Y. A. Dimitriadis, I. Zigurs, & E. Gómez-Sánchez (Eds.), *Groupware: Design, Implementation, and Use* (pp. 94–101). Springer Berlin Heidelberg.

Richter, A., J. Heidemann, M. Klier and S. Behrendt. (2013). "Success Measurement of Enterprise Social Networks." In: *Wirtschaftsinformatik 2013* (pp. 1–15). Leipzig.

Riemer, K., J. Finke and D. Hovorka. (2015). "Bridging or Bonding: Do Individuals gain Social Capital from Participation in Enterprise Social Networks?" In: *36th International Conference on Information Systems (ICIS)*. Fort Worth, USA.

Riemer, K., S. Stieglitz and C. Meske. (2015). "From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks." *Business & Information Systems Engineering (BISE)*, *57*(3), 197–212.

Schubert, P. and J. H. Glitsch. (2016). "Use Cases and Collaboration Scenarios: How employees use socially-enabled Enterprise Collaboration Systems (ECS)." *International Journal of Information Systems and Project Management*, *4*(2), 41–62.

Schwade, F. and P. Schubert. (2017). "Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities." In: *50th Hawaii*

International Conference on System Sciences (HICSS) (pp. 401–410). Hilton Waikoloa Village, Hawaii, USA.

Schwade, F. and P. Schubert. (2018). "Social Collaboration Analytics for Enterprise Social Software: A Literature Review." In: *Multikonferenz Wirtschaftsinformatik 2018*. Lüneburg, Germany.

Shami, N. S., M. Muller and D. Millen. (2011). "Browse and discover: social file sharing in the enterprise." In: *ACM Conference on Computer Supported Cooperative Work* (pp. 295–304). ACM.

Smith, M., D. L. Hansen and E. Gleave. (2009). "Analyzing enterprise social media networks." In: *12th International Conference on Science and Engineering (CSE)* (pp. 705–710). Vancouver, BC, Canada.

Smith, M., B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, … E. Gleave. (2009). "Analyzing (Social Media) Networks with NodeXL." In: *4th International Conference on Communities and Technologies* (pp. 255–264). Brisbane, Australia.

Steinfield, C., W. van Osch and Y. Zhao. (2015). "Intra-Organizational Boundary Spanning: A Machine Learning Approach." In: *21st Americas Conference on Information Systems (AMCIS)*. Puerto Rico.

Steinhueser, M., C. Herzog, A. Richter and U. Hoppe. (2015). "A process perspective on the evaluation of enterprise social software." In: *2nd European Conference on Social Media* (p. 429436). Porto, Portugal.

Stieglitz, S., K. Riemer and C. Meske. (2014). "Hierarchy or Activity? The Role of Formal and Informal Influence in Eliciting Responses from Enterprise Social Networks." In: *22nd European Conference on Information Systems (ECIS)*. Tel Aviv, Israel.

Tzagarakis, M., N. Karacapilidis, S. Christodoulou, F. Yang-Turner and L. Lau. (2014). "The dicode collaboration and decision making support services." In: *Mastering Data-Intensive Collaboration and Decision Making* (pp. 119–139). Cham: Springer.

van Osch, W. and C. Steinfield. (2013). "Boundary Spanning through Enterprise Social Software: An External Stakeholder Perspective." In: *34th International Conference on Information Systems*. Milan, Italy.

Warshaw, J., S. Whittaker, T. Matthews and B. a Smith. (2016). "When Distance Doesn't Really Matter: Effects of Geographic Dispersion on Participation in Online Enterprise Communities." In: *19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 335–345). New York, NY, USA.

Wehner, B., T. Falk and S. Leist. (2017). "What benefits do they bring? A case study analysis on Enterprise Social Networks." In: *Proceedings of the 25th European Conference on Information Systems* (Vol. 2017, pp. 2069–2085).

Wehner, B., C. Ritter and S. Leist. (2016). "Enterprise social networks: A literature review and research agenda." *Computer Networks*, *114*, 125–142.

Williams, S. P. and P. Schubert. (2015). *Social Business Readiness Survey 2014*. Koblenz: University of Koblenz-Landau, Working Report of the Research Group Business Software.

Williams, S. P. and P. Schubert. (2017). "Connecting Industry: Building and Sustaining a Practice-based Research Community." In: *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS)* (pp. 5400–5409). Hilton Waikoloa Village, Hawaii, USA.

Windhager, F., L. Zenk and P. Federico. (2011). "Visual enterprise network analytics-visualizing organizational change." In: *Conference on Applications of Social Network Analysis* (pp. 59–68). Zurich, Sitzerland.

Xhafa, F., S. Caballe, T. Daradoumis and N. Zhou. (2004). "A grid-based approach for processing group activity log files." In: *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (Vol. 3292, pp. 175–186).