

Chapter 2

Studying Collaborative Information Seeking: Experiences with Three Methods

Jette Hyldegård, Morten Hertzum, and Preben Hansen

Abstract Collaborative information seeking (CIS) has lately produced interesting empirical studies, describing CIS in real-life settings. While these studies explore how and why CIS manifests itself in different domains, discussions about how to study CIS have been scarce. The research area of CIS may, however, benefit from a discussion of methodological issues. This chapter describes the application of three methods for collecting and analyzing data in three CIS studies. The three methods are Multidimensional Exploration, used in a CIS study of students' information behavior during a group assignment; Task-structured Observation, used in a CIS study of patent engineers; and Condensed Observation, used in a CIS study of information-systems development. The three methods are presented in the context of the studies for which they were devised, and the experiences gained using the methods are discussed. The chapter shows that different methods can be used for collecting and analyzing data about CIS incidents. Two of the methods focused on tasks and events in work settings, while the third was applied in an educational setting. Commonalities and differences among the methods are discussed to inform decisions about their applicability in future CIS studies and, more generally, to foster methodological discussions in CIS research.

Keywords Collaborative information seeking • Qualitative methods • Observation • Diaries

J. Hyldegård (✉) • M. Hertzum
Royal School of Library and Information Science, University of Copenhagen, Copenhagen, Denmark
e-mail: j.hyldegaard@hum.ku.dk; hertzum@acm.org

P. Hansen
Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden
e-mail: preben@dsv.su.se

2.1 Introduction

With the increasing recognition that information seeking is frequently accomplished by multiple actors in collaboration rather than by solitary actors, collaborative information seeking (CIS) has emerged as an object of study (Fidel et al. 2004; Foster 2006; Shah 2014). Following Shah (2014, p. 219) we define CIS as “an information-seeking process that takes place in a collaborative project (possibly a complex task) among a small group of participants (potentially with different set of skills and/or roles), which is intentional, interactive, and mutually beneficial.” This definition sets the granularity of CIS at the level of projects, rather than for example search sessions. To study CIS we need methods that avoid reducing the project to a unitary actor, thereby black-boxing collaboration. But, we also need methods that avoid reducing CIS to the sum of the project participants’ individual information seeking, thereby disregarding collaboration (Hertzum 2008). Such methods may be new to researchers accustomed to studying individual information seeking but will likely be familiar to researchers of collaboration.

In this chapter we describe and discuss three methods we have used in studying CIS. The three methods were devised to fit the study of CIS as well as to fit the research foci, aims, and particulars of our studies. A longitudinal and predominantly qualitative approach is common to the three methods. An additional commonality is the research interest in the process and how CIS is accomplished. The three CIS methods are:

- *Multidimensional Exploration*, which was devised for a CIS study of students’ information behavior during a group assignment. It involves three general-purpose methods used in concert at three stages of the assignment process to explore behavior over time.
- *Task-structured Observation*, which was devised for a CIS study of patent engineers, and consists of observing a set of selected work tasks. The method could be used for any domain that has a work task as the unit of observation and involves a set of supporting data collecting methods.
- *Condensed Observation*, which was devised for a CIS study of information-systems development and consists of observing a regularly recurring event that is itself a CIS activity and, at the same time, includes an account of the period since the previous instance of the event.

Most CIS studies focus on generating new knowledge about CIS and employ methods merely as means to this end. In contrast, the present chapter gives primacy to methodological issues by presenting three methods for studying CIS and by reflecting on the experiences we gained in employing them. Methods can be discussed at two levels – as types and as instances. At the type level, methods are not applied to specific empirical cases and the focus is normally on a single general-purpose method, such as surveys or interviews. At the instance level, the focus is on a specific empirical case, and the method devised for the case may include elements of several general-purpose methods. The three methods described in this chapter are instances. We hope that the methods and reflections will be useful to other

researchers and students interested in understanding CIS phenomena, designing artifacts to support CIS, or devising methods themselves for studying CIS.

In the following, we present each CIS method in turn. Each presentation is structured into five parts: (1) introduction of the CIS study for which the method was devised, (2) motivation for the devised method, (3) description and explanation of how the method was applied, including general information and concrete details, (4) reflections on the strengths and limitations of the method, and finally (5) similar methods described in the literature. After the presentations we discuss commonalities of the three methods and how these commonalities help identify the kinds of research question for which the methods may be considered. We also discuss differences among the methods and how the methods may, as a consequence, complement each other.

2.2 Three CIS Methods

Table 2.1 gives an initial overview of the characteristics, strengths, and limitations of the three methods. In the following, we present the methods in detail.

Table 2.1 Characteristics of three CIS methods

	Multidimensional exploration	Task-structured observation	Condensed observation
Approach	Case study Qualitative Process-oriented Longitudinal	Case study Qualitative Process-oriented Longitudinal	Case study Qualitative Process-oriented Longitudinal
General-purpose methods	Surveys Diaries, printed Interviews	Observations Diaries, electronic Log-data Interviews	Observations Interviews
Strengths	Rich data picture In-depth exploration of complex phenomena Data triangulation Validates important findings	Real-life data collection of user behavior Close and detailed observations Data triangulation	Real-work data Moment-to-moment detail of observed periods Summary of unobserved periods Data triangulation
Limitations	Indirect observations Self-report bias Generates a lot of data that need to be handled systematically Data analysis is time consuming Generalizability	Time consuming for researchers and participants Generates a lot of data that need to be handled systematically Generalizability	Observation requires keen attention Summaries subject to some self-report bias Generalizability
Case domain	Assignments in higher education	Patent searching	Information-systems development

2.2.1 *Multidimensional Exploration*

2.2.1.1 Introduction

The multidimensional exploration method was developed for a case study exploring three groups of students and their information behavior during a collaborative problem-solving process (Hyldegård 2009a). The focus was not solely on CIS issues but included also personal, social and contextual factors, hence the multidimensional character of the study. The students were ten Library and Information Science (LIS) students preparing an assignment during a period of 14 weeks. The aim of the study was to explore to what extent the established Information Search Process (ISP) model developed by Carol Kuhlthau (1991, 2004) also captured the information behavior of group members. The ISP model was the empirical result of several studies focusing on the individual student's information seeking behavior. However, the model had never been applied to individual students engaged in group-based problem solving. Apart from a focus on group members' activities and cognitive and emotional experiences related to information seeking, this study also included activities and experiences related to group work (social factors) and the work task (contextual factors). More specifically, the case study was guided by four research questions: (1) Will group members across the three groups behave differently from the individual modelled in the ISP model? If so, in which way do they behave and why? (2) Will group members in a group (intra-group members) demonstrate different activities as well as different cognitive and emotional experiences? If so, in which way do they differ and why? (3) How is group member behavior related to contextual factors (the work task)? (4) How is group member behavior related to social factors (aspects associated with group work and collaboration)? The results of the study demonstrated similarities in behavior between group members and the individual in Kuhlthau's ISP model, but also differences that were found to be associated with contextual and social factors beyond the mere search process. It was concluded that the ISP model did not fully comply with group members' problem solving process and the involved information seeking behavior. Further, it was concluded that complex problem solving in an educational context seems to be even more complex when it is performed in a group setting.

2.2.1.2 Motivation

To compare and explore the ISP model in a group setting the research design of the case study was to a large extent constructed in accordance with the methodological framework leading to the ISP model. However, due to the multidimensional nature of the study general data collection methods employed by Kuhlthau were adjusted to take into account not only searching issues but also task and social issues. The general methods were questionnaires, diaries, and interviews which were employed in concert to get a deeper insight into the participants' activities and experiences over time. The aim of combining the selected methods into a multidimensional

methodology was to enable an in-depth exploration of the complex phenomenon of interest, to systematically collect rich data and further, to validate important findings in the data. The methodology was tested in a preliminary case study carried out in 2002 (Hyldegård 2006a).

2.2.1.3 Procedure

The case study involved ten Danish graduate students in library and information science studying in their third year. They voluntarily formed two 3-person groups and one 4-person group. The work task, the project assignment, was a mandatory part of an elective course. During a period of 14 weeks the students had to formulate a problem within a specified topic area, explore the problem and find a focus, find and digest relevant literature, collect and analyze data, devise a structure for presenting their argument and finally write a project report (approximately 20–30 pages). The three general data collection methods that constituted the multidimensional method were questionnaires (demographic questionnaire, a personality test and process surveys), diaries and interviews. In line with Kuhlthau data were collected at three points during the assignment process: start, midpoint, and end. The aim of the demographic questionnaire was to collect profile data on each participant in terms of demographic information and prior experience of group work, IT, and information seeking. The test instrument NEO-PI-R was employed to describe and compare group members' personality traits that might help understand behavior in a group and between group members (Hyldegård 2009b). To elicit process data associated with information seeking, group work and the work task a printed process survey was filled out by each participant and handed in at three selected points (dates) during the assignment process: at start, midpoint, and end. The three process surveys were identical in order to observe changes over time with regard to activities as well as cognitive and emotional experiences associated with information seeking, the work task, and group work. Each process survey was followed by a 1-week diary period (Hyldegård 2006b). The aim was to collect data on a daily basis on each group member's activities and experiences in relation to information seeking, the work task, and group work. The diary also guided the subsequent interviews with each participant; both when deciding which issues to address in the interviews and during the interviews when referring to specific incidents. Further, the diary served as a surrogate for direct observation since it was difficult for the researcher to predict where and when relevant activities might take place during the assignment process. To ensure the quality and limit the amount of textual data the diary was kept for 1 week only. The participants were instructed to record on a daily basis and in their own words any assignment related activity. Affective experiences should be recorded daily, even if no assignment activities had taken place. After the process survey had been handed in and the diary period had ended, each group member participated in an interview. A semi-structured guide was made for each of the three interview sessions, which addressed different aspects related to the work task, information seeking, and group-work in accordance with

the specific point in the process. The first group of questions concerned activities and experiences associated with the project to track and understand cognitive as well as affective changes over time. The next group of questions addressed activities and experiences associated with information behavior, such as use of information sources and relevance assessment. The last group of questions regarded activities and experiences associated with group-work, such as group meetings, form of communication, group-work practice, and coordination. The form of the interview was guided by the micro-moment time-line technique derived from the sense-making approach (Dervin 1983). The aim was to elicit the informant's feelings, thoughts, and experiences in relation to various situations and phenomena while the informant made sense of his or her invoked reactions. By interviewing the participants individually it became possible to explore whether and how group members in a group would differ in their perception and experience of identical situations and incidents; whether and how work task and information-related activities were individually or collaboratively based and, further, how perceptions and experiences evolved over time. The questionnaire data and data from the 30 process surveys were analyzed statistically, whereas data from the 30 diary reports and the 30 interviews were transcribed and analyzed using the qualitative analysis software Atlas.ti and a grounded theory approach (Strauss and Corbin 1998).

2.2.1.4 Experiences

In spite of the case study approach implying that you cannot generalize from this study to other forms of group work or teamwork, the multidimensional methodology contributed with a new conceptual understanding of students' behavior in small groups. The systematic approach to data collection helped reduce and control the complexity of the data deriving from the complexity of the research problem. The short diary period also helped keeping the participants engaged in keeping the diary in addition to the researcher's continual encouragement that was needed to minimize the time gap between event occurrence and the diary recording. As a result of the methodology no direct observations of subjects were made. Rather, results and findings were based on indirect observations of each group member's perception and experiences in either written or oral form. From a phenomenological point of view (Gallagher and Zahavi 2007), these personal perceptions and experiences served to gain insight into the thoughts and behavior of individual group members. With regard to group members' emotional experiences it was difficult to distinguish between feelings that derived from the group work process, the work task process or the information seeking process as these feelings were often interrelated.

2.2.1.5 Similar Methods

Compared to Kuhlthau's (1991, 2004) methodology the Multidimensional Exploration method made it possible to establish a richer picture of individuals'

information seeking process in a group-based setting. Whereas Kuhlthau used search diaries and search logs to study individuals' searching activities and experiences during an assignment process, diaries and process surveys were used in this study as an indirect way of observing not only searching behavior, but also the contextual complexity of seeking activities and experiences otherwise hidden from an observer. The diary was also found to act as a reflective tool for the participants. In a recent review of research related to the ISP model (Kuhlthau et al. 2008) the usefulness of the model in new and technologically rich information environments has been examined and validated. This also implies a validation of the underlying methodology of the ISP model and its applicability in different settings.

2.2.2 Task-Structured Observation

2.2.2.1 Introduction

Task-structured Observation is focused on observing a set of selected work-task and its sub-task in which each sub-task may involve specific activities. The task-structured observation study highlights the searching and information-seeking task as embedded in the overall (work-) task. This method, together with additional data collection methods, were designed and used to investigate CIS manifestations in an information-intensive domain of patent engineers at the Swedish Patent Office (Hansen and Järvelin 2000, 2005; Hansen 2011). The method was applied in a work context. In this domain and context, the unit of observation was the handling of different patent applications. Each patent application was considered as a task with a specific beginning (filing the patent) and end (approval/disapproval) of the patent application. The overall goal of the patent process is to protect investments that individuals and companies have made in new technological innovations and developments and to stimulate the competitiveness in Sweden in a just and fair way. The handling of patent applications is done mainly through classification, searching, retrieving, inspecting, and judging relevant information within the patent domain. The goal of the project was to observe the process of patent handling at large and, more specifically, to observe how information seeking and retrieval (ISR) tasks are embedded into the overall work-task and how these ISR tasks are manifested. The method was used to observe ISR tasks embedded in work-tasks in order to reveal both patterns of work tasks and information seeking tasks, and for the latter one, also specific features.

2.2.2.2 Motivation

The aim of the project was to develop our understanding of collaborative activities within IS&R processes and to identify what kind of collaborative activities that

could be observed in an IS&R process. Therefore, in order to understand aspects of collaborative activities from an IS&R perspective, we needed to investigate the manifestations of collaboration. We wanted to investigate everyday information-seeking processes in a work environment in order to monitor and reveal features of search tasks (Jurgens et al. 2012) in a real-life setting. However, one prevailing assumption in information retrieval (IR) is that problem understanding, query formulation, and retrieval are basically viewed as an individual activity and that the searcher performing the task is in a rather isolated situation. In the more traditional IR research domain, the so-called “Cranfield paradigm” (Cleverdon 1967; Vorhees 2002) is well understood and dominant, and experiments are conducted in more or less controlled laboratory settings. On the other side, user-oriented and interactive IR approaches deal with the overall performance of an IR application from a user perspective and includes dimensions such as user interface, interaction patterns, and tasks (work-tasks and search tasks) (Ingwersen and Järvelin 2005). To be able to follow real-life work-tasks and their embedded search tasks, it is necessary to use a longitudinal and ethnographical method combining several data-collection techniques. Furthermore, it was considered necessary to be present during the data-collection phases to understand the context and situation that framed the patent engineers’ ISR tasks and to study the behavior of patent engineers while they solved a patent work task.

2.2.2.3 Procedure

The methods described here, was used in a project that was part of a larger study performed at the Swedish Patent and Registration Office¹ (PRV), Stockholm, Sweden. Data collection was performed on-site in a real-life work setting involving professional patent engineers (PEs) performing their patent work-tasks. The goal with the project was to investigate what PEs actually did during their work processes regarding their information seeking processes. We used a set of other qualitative and quantitative data collection techniques and methods. The information seeking data were collected during 2 months. This longitudinal data collection and that analysis describe how, what, and when CIS activities manifest themselves. Altogether nine professional patent engineers participated in the study. The patent engineers worked either alone or in pairs in each office room. All 9 participants were observed performing 12 unique patent work-tasks. Patent work tasks were observed at different stages of the searching process, which enabled us to cover all the main stages in the information handling process of each patent application and thus, all stages of the information seeking process were represented in our study. Due to time constraints we limited our unit of observation to well-established sub-tasks

¹PRV, Stockholm, Sweden, <http://www.prv.se>

within the patent handling process. Additional data collection techniques that were combined with the on-site observations were electronic diaries, log-files from all database searches, and interviews.

Observations The observations were conducted as continuous on-site visits in order to observe patent engineers in their professional work. Before doing the observations, a list of key questions were designed that represented units of pre-defined categories related to the stages of the information seeking process. During observation, these questions were asked at certain moments when a participant changed activity (such as going from reading to searching a database) and notes were taken. Awareness of unexpected situations and activities was emphasized. The subjects were encouraged to “talk aloud”. Each single observation could take up to 6–7 h a day and for each task, 3–6 daily visits were conducted. The observations were constrained by activities such as external/internal courses and confidential internal meetings, or other duties to which the investigator did not have access. Each task was followed up to 5 days in total (could be spread out on a 2–3 week period). The data were collected over a period of 2 months.

The observations were supplemented with other data-collection methods such as electronic “diaries” and interviews. An electronic “diary” was designed for the participants so that they could describe their daily activities and was designed to collect different types of data such as search queries, descriptive data, and log files. It contained a formal outline involving common steps in the patent-handling process. The outline was based on an initial interview with two patent experts and on the experience from the pilot study. The diary also contained empty fields so that logging information² could be inserted from database searches. The diaries were designed to capture the following data: log histories from database searches; descriptions of problems to be solved and how the PE did solve it; personal comments on the search problems; time stamps for performing sub-tasks; search terms, search strings, and classification codes used; collaboration with colleagues; handling work tasks within their own department/group etc. Data analysis was performed in an iterative procedure, on account of its nature (qualitative and quantitative). For the interviews with the expert patent engineers we used open coding in order to make categories. Each category contained a set of variables. The categories were then designed as a set of instantiations in the protocol used for the electronic diaries. For analysis purposes, we designed a matrix containing both numerical and categorical values from the data collected. We analyzed individual variables using both Spearman’s rho (Siegel and Castellan 1988) correlation and Yates χ^2 (Chi-square test). Semi-structured and open-ended interviews were performed before and after the main data-collection period. The pre-interviews were used to collect data about demographics, experience and knowledge levels as well as descriptions on search procedures etc.

²Log information was handled in such a way that it did not reveal any unauthorized information.

2.2.2.4 Experiences

One big challenge in our study was how to collect data and what methods to use, given the setting of a patent work environment with real-life conditions and practices, involving professional patent engineers. Our methods allowed us to observe not only what activities the PEs performed both online and physically but also how these activities were performed over time. This approach also made it possible to describe and characterize the patent domain, as well as the work processes and procedures. The application of the Task-structured Observation method proved to be useful and feasible for capturing the type of data that was required to understand, unfold, and model the CIS tasks embedded in real-life work tasks. Of course the methodology has some limitations and problems. Because of the methodology used, care must be taken in generalizing the results from the present study. However, the limited number of subjects resulted in a surprisingly substantial amount of data from which it was possible to describe and explain patent IS&R processes.

We found it natural as well as challenging to perform this research in a real-life setting. The work task setting made us confront the real conditions and requirements of work task performance. Given the real-life setting and other issues in observing professional PEs in their work, we knew from the outset that we could not claim full control over the data collected and what actually was produced. Without this control, one may not be able to foresee certain data output fully and design for it. On the other hand, such work may be rewarding in that unexpected data and new angles may emerge. For example, for different variables, we did not know which categories the data would result in. Time was another uncertain aspect of the data collection process.

2.2.2.5 Similar Methods

We used and combined log data, diary data, and on-site observations of patent actors' performance. Our study complied with Brennan's (1992) statement that (a) quantitative research can unfold more structural features of the research target, while qualitative research supports understanding of processes, and (b) qualitative methods support the subject's perspective, while the quantitative is related more to the researcher's own focus. For example, the observations and diaries revealed CIS activities, which were not visible from the log data.

Since we took a task-structured approach, the basic unit for observation and data collection was the individual work and search task. Especially when observing the on-site work tasks, we did not know beforehand when a task should be deemed to end or what the observed task would lead to. Task-structured Observation proposes to be applied in research that deals with investigating search-tasks embedded in real work tasks. We suggest using observation as the main method, but combined with other data-collection techniques such as, in this case, electronic diaries, log-files, and interviews. What to combine observations with is related to what types of data

to collect. Task-structured Observation lets the investigator focus on and structure the observation process around the task as the primary unit.

2.2.3 *Condensed Observation*

2.2.3.1 Introduction

Condensed observation consists of following a process over a long period of time by observing a regularly recurring event that is itself an important activity in the process and, at the same time, includes some kind of account of the period since the previous instance of the event. This method was devised to investigate CIS in a systems-development project (Hertzum 2000, 2002). In this context, the regularly recurring event was the fortnightly project status meetings. The studied project concerned a system developed by a large, Danish software company to support municipal authorities in the handling of cases concerning child support and alimony (CSA). To accomplish their task, the systems developers on the CSA project needed to interact with management, marketing, technical services, the quality function, and other stakeholders internal to their company. To distribute work effectively in the company part of the functionality of the CSA system was delegated to system components that were to be developed by other project groups in the company. While this delegation reduced the amount of software to be produced by the CSA developers, it increased the amount of interaction they needed to have with people external to the CSA project. In addition, the CSA developers also had to interact with company external stakeholders such as end-user representatives, municipal managers, and the government agencies responsible for the legislation regarding child support and alimony. The study aimed to investigate (1) the collaborative processes through which the CSA engineers assessed and selected their information sources, (2) the extent to which the actual contact with the sources was delegated to individual CSA engineers, and (3) the ways in which information acquired from sources was incorporated in the CSA engineers' shared understanding of their task – or remained at the periphery of this understanding.

2.2.3.2 Motivation

The many contributors to the project in addition to the CSA developers meant that the CSA developers repeatedly needed to seek information from project-external sources. Hence, the project provided a rich opportunity for studying CIS. To be able to study how the CSA developers concretely accomplished CIS, we needed to get close to their work and decided to make use of observation. The most interesting project activities from a CIS point of view were the early activities of gathering requirements, negotiating the delegation of functionality to other project groups,

and arriving at an agreed-upon conceptual design. However, these activities spanned the first year of the 3-year CSA project. To be able to follow the project for a full year we devised condensed observation, which aims to be a resource-conscious approach to observation. We focused the observations on a genuinely collaborative activity, the fortnightly project status meetings, during which the CSA developers followed up on ongoing activities and assigned each other new tasks. The former often involved the collaborative discussion and grounding of information sought by individual CSA developers since the last meeting, the latter often involved the assessment of possible information sources to be contacted by individual CSA developers before the next meeting.

2.2.3.3 Procedure

Like in other variants of observation, considerable time was spent negotiating access to the project before the observations started. While these negotiations resulted in permission from management to follow the project, it was an independent issue to develop a relationship with the CSA developers in order to be accepted as observers of their work. To this end, the researcher attended the 2-day project kick-off seminar by presenting the study, participating in the discussions about the project, and socializing with the CSA developers. To avoid that the CSA developers became overly self-conscious about their information-seeking activities the researcher had developed a cover story that presented the focus of the study in broader terms and, thereby, diffused attention from information seeking. During the kick-off seminar the researcher also obtained the CSA developers' permission to audio-record the status meetings for later analysis and explained that in contrast to his visible role at the kick-off seminar he would make an effort to recede into the background during the status meetings. Finally, we agreed on a procedure where the researcher had the right to publish his findings from the study but the CSA developers received and had the opportunity to comment on all papers prior to their publication.

During the first year of the CSA project, the researcher observed the 26 status meetings, each lasting 2–3 h. These observations constituted the primary period of data collection. The following year another 13 status meetings were observed to learn whether the project activities during the first year yielded a successful project. The status meetings were attended by most of the 17 people assigned to the CSA project. During the meetings, the researcher was seated at the meeting table with the other people present. From their point of view, he was invisible in that he remained silent and merely observed the meetings. During the breaks, the researcher talked informally with people. These informal conversations provided opportunities for seeking explanations of issues that had been unclear during the meeting or appeared particularly interesting. The meetings contained hundreds of incidents in which the CSA developers, for example, proposed to use an information source because it was easily accessible or assessed input from an information source by discussing whether the source was representative (Hertzum 2002):

- “It is those CICS statistics I am thinking of. Can we use them? We made those statistics for [a government agency] so they are readily available.”
- “We have to be good at asking other people than [two of the user representatives]. Don’t get me wrong . . . They are very competent and careful. But they are not always representative of an average user.”

Notably, the considerations in the incidents were made collaboratively by the CSA developers as an integral part of their work. The data were analyzed through a process of content analysis (Lazar et al. 2010). This process involved two passes through the data. First, the transcripts of a subset of the meetings were read sentence by sentence and all CIS incidents were marked up and annotated. This bottom-up analysis, combined with findings from related work, resulted in the creation of a coding scheme. Second, the transcripts of all the meetings were examined to identify the CIS incidents and categorize them according to the coding scheme. The analysis, for example, resulted in an enumeration of the criteria used by the CSA developers in their selection of sources. The observations were supplemented with a round of interviews with the CSA developers to get their reflections on their CIS and its implications for their progress on the project. In analyzing the data, we maintained a distinction between the observations and the interviews in order not to conflate what had happened with how it was talked about.

2.2.3.4 Experiences

The project status meetings proved a viable way of conducting condensed observation. On the one hand, the meetings were a real work activity during which the CSA developers worked on their project rather than merely talked about it. For example, they continuously realigned their understanding of the direction of the project by collaboratively grounding information acquired by individual CSA developers. On the other hand, the meetings simultaneously provided brief accounts of what had happened on the various project activities, which had not been observed, since the last meeting. Status meetings may be unique in supplying this combination of real work and summary. Condensed observation was, however, not without its limitations. First, some CSA developers were more vocal than others, thereby possibly making their activities and considerations more salient during the meetings than they were in the periods between the meetings. Second, to be brought up at the meetings an issue had to exceed a certain importance threshold. Thus, the moment-to-moment evolution of the CSA developers’ CIS outside of the meetings was not captured by condensed observation. Third, the status meetings entailed reporting to the project manager and team about activities and progress. To present themselves well, the CSA developers may, at times, have emphasized situations they handled competently and shifted attention away from other situations. Finally, the CSA developers were, obviously, aware of the presence of the researcher. For example, the researcher was on occasion asked whether he, as the one person who attended all status meetings, could remember what the CSA developers had agreed on at a previous meeting.

2.2.3.5 Similar Methods

Much of the general advice about how to conduct observations (e.g., Button and Sharrock 2009; Lofland et al. 2006) is also relevant to condensed observation. We especially want to reiterate two remarks by Button and Sharrock (2009):

- “Open your eyes”, which acknowledges the difficulty of seeing what goes on right in front of you. It takes effort to appreciate the import and pattern of apparently mundane activities.
- “You get what you get”, which reminds you that selection is unavoidable. There is always more you could observe and practicalities, not just principles, will always affect the selection.

Condensed observation proposes the simultaneous presence of real work and summary, such as in project status meetings, as a principle for selecting what to observe. Other variants of observation propose other ways of making the selection. For example, shadowing consists of following specific actors as they go about their work. Yet another variant of observation consists of following specific objects as they pass through their lifecycle. In both of these variants a selection restricts the observations to a subset of what goes on but, unlike condensed observation, the observed activities do not give a summary account of the other activities.

2.3 Discussion

In this section we discuss the commonalities of the methods and the ways in which they differ. The section ends with some reflections on devising methods for CIS studies. The three studies in which the methods were applied will be referred to as the ‘assignment study’, the ‘patent study’, and the ‘systems-development study’.

2.3.1 Common Characteristics of the Three Methods

In contrast to CIS studies looking at system-based collaboration and support (Shah 2014), the methods presented here were devised for studies of CIS in real-life settings. More specifically the focus in all three cases was on collaboration in small groups involved in information-intensive problem solving. We wanted to understand what we may call ‘how and what’ questions in relation to the collaborative process as it unfolded over an extended period of time, which further called for a longitudinal research approach. Consequently, methods such as controlled laboratory experiments were not suitable. Our interest in the *process* also meant that data in each case were collected at multiple points during the process. The recurrent data collection made it possible to compare data over time and identify any changes in behavior. Further, the process-oriented research approach enabled us

to understand data in a broader perspective. It may, however, be difficult to plan for and control exactly which data to collect in real-life settings, but often an added bonus can be that interesting and important findings emerge unexpectedly from the data. The presented CIS methods have been based primarily on a qualitative methodology, whereas quantitative elements served as a supplement. For example, the search log data in the patent study provided the researcher with actual behavior data rather than merely perceived or observed data. Further, the survey data in the assignment study made it possible to compare responses across group members both within and across groups. Due to the complexity of the research object in each study and the fact that activities in real-life settings may be hard to control, the question of validity received considerable attention. For example, each CIS method consisted of multiple complementary data collection methods and techniques to ensure a rich picture of the research object and to enable rich descriptions and explanations of CIS in context. This method triangulation also helped combat the various threats to validity inherent in the research approach. However, the focus on validity was at the expense of reliability and generalizability. It is not possible to control variables in real-life settings in the same way as in the laboratory. In addition, one cannot generalize from a single case study with a limited number of participants. Biases associated with the investigator (interviewer or observer) or actor (participants or informants) may also have reduced validity. For example, how could we be sure that the researcher in the workplace cases observed and reported objectively on the one hand, and on the other hand that participants in the assignment study were reporting incidents as they were – and not as they remembered them or wanted them to be.

According to McGrath (1981), researchers always face a dilemma of incompatible goals. While we would want a method that yields reliable, valid, and generalizable results, methods always yield one at the expense of the others. For example, surveys provide for generalizability, but at the expense of reliability and validity. And observation provides for validity, but at the expense of reliability and generalizability. No single method exists that will guarantee success. All research strategies and methods are flawed in the sense that they sacrifice at least one of reliability, validity, and generalizability. Acknowledging this methodological dilemma, hence also the limitations of the three methods presented in this chapter, further stresses the importance of describing and understanding CIS methods well. Other researchers should be able to repeat and build upon a study to confirm or disprove its findings. The observations in the systems-development study were strengthened by audio-recording the meetings to enable a thorough analysis, by only assigning importance to CIS phenomena that recurred multiple times in the series of meetings, and by collecting data over an extended period of time, thereby allowing for variation to occur. In the patent study the observations were strengthened by focusing on the information seeking part of the patent engineers' overall work-task. Alternatively, methods such as data log analysis of database searches and in-depth interviews could have been applied. In the assignment study the self-reported data were strengthened by comparing data about the same incident across group members and by following up in interviews on findings in the process surveys and diaries. Moreover, the group members' diary keeping was encouraged by shortening

the period it was kept and by recurrent reminders from the researcher. The self-reported data could have been strengthened further by complementing it with direct observation of group meetings.

2.3.2 Differences Among the Three Methods

The differences among the three methods were associated primarily with the choice of context and research focus. In the systems-development study, a specific type of situation, namely the recurring status meetings, was used to study the systems developers' CIS. In the patent study, a specific work-task was selected for studying one among the many work-tasks performed by the patent engineers. In both cases, however, the study of CIS was restricted to a workplace and a professional setting. In contrast, in the assignment study CIS was approached as an activity embedded in the assignment process along with other activities 'in and out' of the university. These differences in the settings in which CIS was studied meant that direct observation was only employed in the two workplace studies. In the assignment study, diaries served as a surrogate for direct observation. As already pointed out above, direct observation of a specific type of situation, such as group meetings, could have supplemented the self-reported diary data and further qualified the interviews. Diaries were also employed in the patent study, but the format differed across the two CIS methods. In the assignment study printed diaries were used to ease participants' recording of activities and experiences when they brought the diary along to meetings with the group, in the library, when working individually on the project assignment and so forth, thereby minimizing the time gap between event and recording. All handwritten data had to be processed manually by the researcher afterwards to prepare data for analysis in Atlas.ti, which was time consuming. Today, mobile technology such as smartphones and tablets has replaced the need for printed diaries. In the patent study, an electronic diary was designed especially for the participants' search histories and for their daily reflections and thoughts. The electronic diary made it easy for the patent engineers to insert any type of information from their professional activities, and it also made it easy for the investigator to collect and process data swiftly.

2.3.3 Devising Methods for CIS Studies

When devising methods for CIS studies one cannot just replicate the method used in another CIS study or one of the three methods presented in this chapter. It is necessary to consider the various issues that relate to the research focus, the amount of resources available, and the practicalities of the empirical case. For example, what are you going to study – one incident or a series of incidents during a process involving CIS? Is it a complex and information-intensive work task you are going

to study? What activities are involved and will all of them be relevant to your study? What is the context and domain – a work setting, an everyday-life setting or an educational setting? Who are the participants? And how many participants should be involved? Which variables are important to control? How much time is it realistic to expect from participants compared to the duration of the project and their motivation to participate?

No research is better than its data. In the absence of perfect methods it is crucial to consider methodological issues carefully in order to ensure that the employed methods yield data appropriate to the aims of the study. The methods employed inevitably influence what conclusions your data and analysis will allow for. To cite McGrath (1981, p. 210) “... good research – using flawed methods well, and in effective combinations – can help us accrue ‘knowledge’ about behavioral and social science problems that are of both theoretical and practical concern.”

2.4 Conclusion

As an emerging research topic, CIS has been regarded an ‘add-on’ to general models of information seeking and retrieval. However, CIS is increasingly getting a position of its own. The increasing focus on CIS as a genuine topic of research stresses the importance of considering methods for studying CIS phenomena, processes, and practices. In this chapter we have illustrated how methods can be applied when studying CIS phenomena in real-life settings. More specifically, we have described three methods: Multidimensional Exploration; Task-structured Observation and Condensed Observation. For each method we have described the motivation for using the method, the procedure involved in using it, and provided our reflections on the experiences we have gained applying the method in a CIS study.

The three methods were all devised for studying CIS in small groups engaged in complex and information-intensive tasks. Further, a process-oriented approach was adopted because CIS activities evolve over time in real-life settings. The methods proved suitable for becoming aware of unexpected phenomena and relationships in the data. Because each empirical setting has its own characteristics, it may be advantageous to distinguish between two levels in the planning of CIS studies: a general level and a level that takes the specifics of the setting into account. At the general level, models and frameworks appropriate for the study at hand may be explored. Here the researcher can elaborate what variables and features that must be included in the study. At the specific level, the researcher needs to tailor the data-collection methods to these variables and features to ensure data quality, hence that the collected data enable rich descriptions of the CIS phenomena studied.

We find that there is a need for more research on the methodological aspects of CIS studies. Future studies may investigate the possibilities for supplementing observation or enriching interviews with data recorded by informants using technological means. New data-capturing tools exist such as pop-up questions on the informant’s phone for experience sampling, Google Glass for capturing incidents

that occur at irregular intervals, and other technologies for probing informants about the cognitive side of CIS that is inaccessible through observation. In addition, more knowledge is needed about how the established models and methodologies of individuals' information seeking align with the behavior of multiple actors involved in CIS. Finally, a more systematic discussion is needed on how to devise CIS studies with methodological frameworks adjusted to the goal of the research, research topic, unit of observation, and contextual/situational conditions.

References

- Brannen J (ed) (1992) *Mixing methods: qualitative and quantitative research*. Ashgate Publishing, Burlington
- Button G, Sharrock W (2009) *Studies of work and the workplace in HCI: concepts and techniques*. Morgan & Claypool, San Rafael
- Cleverdon C (1967) The cranfield tests on index language devices. *ASLIB Proc* 19:173–194
- Dervin B (1983) An overview of sense-making research: concepts, methods and results to date. Paper presented at the annual meeting of the international association, Dallas
- Fidel R, Pejtersen AM, Cleal B, Bruce H (2004) A multidimensional approach to the study of human-information interaction: a case study of collaborative information retrieval. *J Am Soc Inf Sci Technol* 55(11):939–953
- Foster J (2006) Collaborative information seeking and retrieval. In: Cronin B (ed) *Annual review of information science and technology*, vol 40. Information Today, Medford, pp 329–356
- Gallagher S, Zahavi D (2007) *The phenomenological mind: an introduction to philosophy of mind and cognitive science*. Routledge, New York
- Hansen P (2011) Task-based information seeking and retrieval in the patent domain. Processes and relationships. Doctoral thesis, Acta Universitatis Tamperensis 1631
- Hansen P, Järvelin K (2000) The information seeking and retrieval process at the Swedish Patent- and Registration Office. Moving from lab-based to real life work-task environment. In: Kando N, Leong M-K (eds) *Proceedings of the SIGIR 2000 workshop on patent retrieval*. ACM SIGIR, Athens, pp 43–53
- Hansen P, Järvelin K (2005) Collaborative information retrieval in an information-intensive domain. *Inf Process Manag* 41(5):1101–1119
- Hertzum M (2000) People as carriers of experience and sources of commitment: information seeking in a software design project. *New Rev Inf Behav Res* 1:135–149
- Hertzum M (2002) The importance of trust in software engineers' assessment and choice of information sources. *Inf Organ* 12(1):1–18
- Hertzum M (2008) Collaborative information seeking: the combined activity of information seeking and collaborative grounding. *Inf Process Manag* 44(2):957–962
- Hyldegård J (2006a) Collaborative information seeking – exploring Kuhlthau's information search process-model in a group-based educational setting. *Inf Process Manag* 42(1):276–298
- Hyldegård J (2006b) Using diaries in group based information behavior research – a methodological study. Information interaction in context. In: *Proceedings of the 1st IliX symposium on information interaction in context*, Royal School of Library and Information Science, 18–20 Oct 2006, pp 261–274
- Hyldegård J (2009a) Beyond the search process: exploring group members' information behaviour in context. *Inf Process Manag* 45(1):142–158. doi:[10.1016/j.ipm.2008.05.007](https://doi.org/10.1016/j.ipm.2008.05.007)
- Hyldegård J (2009b) Personality traits and group-based information behaviour: an exploratory study. *Inf Res* 14(2): paper 402. Available from 25 May 2009 at <http://InformationR.net/ir/14-2/paper402.html>

- Ingwersen P, Järvelin K (2005) *The turn: integration of information seeking and retrieval in context*. Springer, Heidelberg
- Jurgens J, Hansen P, Womser-Hacker C (2012) Going beyond CLEF_IP: the ‘reality’ for patent searchers? In: Catarci T et al (eds) *Proceedings of the 3rd international conference of the CLEF initiative, CLEF 2012, Rome, Sept 2012*, pp 30–35
- Kuhlthau CC (1991) Inside the search process: seeking meaning from the users perspective. *J Am Soc Inf Sci* 42(5):361–371
- Kuhlthau CC (2004) *Seeking meaning – a process approach to library and information services*, 2nd edn. Libraries Unlimited, London
- Kuhlthau CC, Heinström J, Todd, RJ (2008) The ‘information search process’ revisited: is the model still useful? *Inf Res* 13(4): paper 355. Available at <http://InformationR.net/ir/13-4/paper355.html>
- Lazar J, Feng JH, Hochheiser H (2010) *Research methods in human-computer interaction*. Wiley, Chichester
- Lofland J, Snow D, Anderson L, Lofland LH (2006) *Analyzing social settings: a guide to qualitative observation and analysis*, 4th edn. Wadsworth, Belmont
- McGrath J (1981) Dilemmatics: the study of research choices and dilemmas. *Am Behav Sci* 25(2):179–210
- Shah C (2014) Collaborative information seeking. *J Assoc Inf Sci Technol* 65(2):215–236
- Siegel S, Castellan NJ Jr (1988) *Nonparametric statistics for the behavioural sciences*, 2nd edn. McGraw-Hill, London
- Strauss A, Corbin J (1998) *Basics of qualitative research – techniques and procedures for developing grounded theory*, 2nd edn. Sage Publications, London
- Voorhees EM (2002) The philosophy of information retrieval evaluation. In: *Evaluation of cross-language information retrieval systems. Lecture notes in computer science*, vol. 2406. Springer, Heidelberg, pp 355–370

Collaborative Information Seeking

Best Practices, New Domains and New Thoughts

Hansen, P.; Shah, C.; Klas, C.-P. (Eds.)

2015, XVIII, 226 p. 23 illus., 15 illus. in color., Hardcover

ISBN: 978-3-319-18541-5