



Information Research - Vol. 29 No. 3 (2024)

# Episodic focus formulation in media scholars' data interaction

Laura Korkeamäki, Heikki Keskustalo and Sanna Kumpulainen

DOI: <https://doi.org/10.47989/ir293764>

## Abstract

**Introduction.** Focus formulation affects task information interaction and task performance. It has typically been conceptualised as one task stage in the overall task performance. Nevertheless, we take a different approach by studying focus formulation as an episodic phenomenon in the task process performance. Episodic focus formulation includes several focus formulation episodes during task performance processes that gradually lead to focus, but instead of considering the larger task, this happens during a subtask. The purpose of this study is to examine what is episodic focus formulation like in media scholars' data interaction.

**Method.** We interviewed twenty-five media scholars about their research processes and related interactions with their research data.

**Analysis.** We examined focus formulation episodes as sense-making instances. We identified the focus formulation episodes from the interview data and analysed across them by annotating and categorizing. We analysed the episodic focus formulations by subtasks and then summarised over the subtasks to cover the whole data interaction.

**Results.** We identified three subtasks, each with three episodic focus formulations. Summarising over the subtasks showed that episodic focus formulation in data interaction encompasses task focus, data focus and procedural focus formulation.

**Conclusions.** This study has implications for theory development and for supporting data interaction.

## Introduction

Focus formulation has been agreed to affect information interaction during task performance (Kuhlthau, 2004; Vakkari 2001). Although focus formulation has been extensively studied in the field of information science and information interaction (e.g., Kuhlthau, 2004; Vakkari 2001), it is typically examined in the context of document use and there is a lack of research in data interaction contexts. Traditionally, research has considered information as recorded documents used for the same purpose as what was their creators' original meaning and ignoring how data contributes to task performance. Therefore, by focusing on examining interactions with research data, the research extends beyond the traditional idea of documents and their uses in information interaction research. Research data interaction involves using research materials in an interpretative manner, which includes utilising them as informational cues, aiming at creating new knowledge, and enhancing the comprehension of complex phenomena being studied. This is typical of media studies, a research field at the intersection of humanities and social sciences (Jensen, 2012), where media texts and other kinds of research data are used to study and understand media-related phenomena.

Further, focus formulation means both reducing uncertainty and task complexity, increasing conceptual structuring and understanding of what is meaningful for the task (see e.g., Byström and Järvelin, 1995; Kuhlthau, 2004; Vakkari, 2001). Focus formulation has been examined in relation to a larger task, where focus formulation is conceptualised as one task stage in the overall task performance. Specifically, the models by Kuhlthau (2004) and Vakkari (2001) depict stages in task performance, where the focus formulation stage, taking place in the middle, is the most crucial stage for successfully completing the task. Forming a clear focus in that stage gives direction on how to perform the task, and if a focus is not formed, difficulties can be expected in the later stages of the task performance. However, in research work that

often spans over years, it is obvious that focus formulation is not just one stage but happens gradually. Therefore, in this study, we examine focus formulation as an episodic phenomenon in the task process performance. We use the term episodic focus formulation, by which we mean that there are several episodes during the research process performance that contribute to the focus formulation for the research. We consider each focus formulation episode as a sense-making instance (Souto et al., 2008).

To study the episodic focus formulation in the media scholars' interactions with their research data, we utilise a task-based approach. The task-based approach examines how task performance and task characteristics affect information interaction (Byström and Kumpulainen, 2020; Järvelin et al., 2015; Toms, 2019). A larger task can be examined through its stages, subtasks, processes, and activities, which have been shown to affect information interaction (Byström and Hansen, 2005; Järvelin et al., 2015; Soufan et al., 2021). The task-based approach allows insights into how people plan, execute, and complete tasks, and it also enables shedding light on their cognitive processes (Järvelin et al., 2015).

In this study, we interviewed twenty-five media scholars (doctoral researchers and participants with a doctorate) about their research processes and related interactions with their research data. To examine episodic focus formulation in the above-mentioned application area, interaction with research data, we have set the research question:

What is episodic focus formulation like in media scholars' data interaction?

## Related work

The literature review is structured as follows. First, we go through some well-known models of task stages and focus formulation. Second, we present studies regarding media scholars' research process. Third, we review studies that utilised task process stages and focus formulation as a part of it in their research design to examine how to support information searching or systems development.

## Models about task stages and focus formulation

Several scholars in the field of information science have addressed the issue of focus formulation by showing how interacting with and searching for new information gradually leads to a more focused understanding of the task and required information. For example, Taylor (1968) suggested that the need for information becomes more focused in a question-negotiation process – from visceral need (unexpressed), conscious need (ambiguous), formalized need (expressed in concrete terms) to compromised need (expressed as search terms). Byström and Järvelin (1995) characterized how complex tasks (genuine decision tasks) start with the uncertainty about what information is needed to perform the tasks. Belkin et al. (1982) used the term anomalous state of knowledge, which means that task-doers recognise the problem but are unable to express exactly what they do not know, especially in concerning the use of information retrieval query languages.

One of the best-known models of task process stages and focus formulation is Kuhlthau's (1993; 2004) information search process model. Kuhlthau (1993) showed that a task-doer's understanding of needed information develops through six stages: (1) initiation is characterized by recognising the need for information, feeling uncertain about the task, and having a vague understanding of the task; (2) selection is about choosing a topic and finding background information for the task; (3) during exploration, the task-doer tries to find a focus for the task, but still has difficulty in specifying what information is needed for the task; (4) at the formulation stage, the task-doer can finally form a focused understanding of the task and feels confident of performing it; (5) at the collection stage, the task-doer is able to express specifically what information is needed for the task which makes searching for information more efficient; (6) the information search process ends with the presentation stage if the task-doer has collected enough information to complete the task. Vakkari (2001) modified Kuhlthau's model to a streamlined version, where a task process includes only

three stages: pre-focus, focus formulation and post-focus, where pre-focus combines stages 1–3 (initiation, selection, and exploration) of Kuhlthau's model, focus formulation matches with the stage 4, and post-focus combines stages 5–6 (collection and presentation). Vakkari argued that the three-stage model is sufficient for studying what happens in a task process before and after focus formulation.

Drawing from these studies, by focus formulation we mean both reducing uncertainty and task complexity, and increasing conceptual structuring and increasing an understanding of what is meaningful for the task. However, we approach focus formulation differently to the models by Kuhlthau (2004) and Vakkari (2001), where focus formulation was depicted as one task stage in the overall task performance. In this study, we analyse episodic focus formulation, meaning that there can be several episodes during the research process performance that contribute to the focus formulation for the research.

## Media scholars' research process

Several studies have examined information seeking and research activities of social scientists (e.g., Ellis, 1989; Meho and Tibbo, 2003) and humanists (e.g., Given and Willson, 2018; Palmer and Neumann, 2002). Some studies have focused on media scholars and examined their research processes and data interaction from different viewpoints. Bron et al. (2016) presented a model of media scholars' research cycle that consists of three phases. The first phase, exploration, involved forming initial research questions, initial information gathering, and examining literature and other background material. In this phase, media scholars' initial research idea became more focused. The second phase, contextualization, involved targeted information gathering, analysis, and further research question formation. Media scholars' data gathering became more specific, either to provide context for the research data or to make further decisions to select the research data. The third phase, presentation, involved interpreting the research data in the light of the research questions, and writing. Furthermore, Bron et al. found that media scholars' research questions

evolved during information gathering and analysis for many reasons. For example, original research questions could not be answered, there was a lack of suitable material, or new things emerged from the material that were worth exploring further. In addition, studying literature contributed to changes in the research questions.

Melgar et al. (2017) studied media scholars' annotation activities during the research process, specifically audiovisual media annotation. They presented a model that depicted annotation as a process "from pre-focused annotation to the creation of new information objects". For example, finding and creating datasets involved pre-focused annotations such as bookmarking or open coding. Analysing involved more focused annotations such as focused coding or categorizing.

Korkeamäki et al. (2022; 2024) studied media scholars' research processes from the viewpoint of information needed while interacting with research data. They utilised Byström's (1999) typology of task information, domain information, and task-solving information as a framework. The first study (Korkeamäki et al., 2022) showed that, in gathering research data, media scholars needed information about research data, data sources, and cases and contexts of interest to the research (considered as task information). The second study (Korkeamäki et al., 2024) showed that, in data interaction more generally, media scholars needed earlier research information and understanding of the world of the phenomenon they were studying (considered as domain information). The study also showed that media scholars needed information about research methods and tools, information about rules and norms concerning scientific research work, and self-created information intended for one's own use to support one's analytical and reflective thinking and task monitoring (considered as task-solving information).

## **Utilising task stages to support information searching and systems development**

Others have utilised the task process stages in their research designs to study how to support information searching or systems development. For example, Huurdeman et al. (2016) and Huurdeman and Kamps (2020) used Vakkari's (2001) task process stages (pre-focus, focus formulation, post-focus) to study how search user interface features could be designed to support search tasks. Gaikwad and Hoeber (2019), similarly, used Vakkari's (2001) task process stages to study how to support interactive image retrieval. Ruthven (2019) used Taylor's (1968) information need levels to study how to differentiate conscious and formalized information needs from online discussion forum data. Furthermore, connections between task stages and search behaviour have been studied, for example, by analysing web search logs and user annotation data (Liu et al., 2020) or search logs and self-report data (Palani et al., 2021).

## **Methods**

### **Recruitment and participants**

We collected the research data through semi-structured interviews and demonstrations in November 2019–April 2020. We used a maximum variation sample (Patton, 2002, pp. 234-235) to select the participants. The criteria were as follows: participants from the fields of media, communication, and game studies; participants whose research project was ongoing or recently concluded and therefore easier to remember; participants from different stages of their career and from different universities to have variability in their research interests, experience, and research data. Interview invitations were sent by email. On one occasion, the interview invitation was presented face-to-face to a research group.

Twenty-five media scholars were recruited for this study from three universities (9 doctoral researchers and 16 with a doctorate, their research experience ranged from under 1 to over 20 years). Participants positioned themselves in the fields of media, communication, or game studies. Some also



mentioned, for example, journalism, social media research, film history, film studies, visual research, audience research, critical research, humanistic research, feminism research, or political research. Participants' research data included journalistic texts, social media data (e.g., social media posts, online forum posts or blog posts), political texts, monographs, television programs, films, and related material, as well as research data collected through surveys, interviews, and workshops.

### **Semi-structured interviews and demonstrations**

The semi-structured interview protocol was designed by utilising four of the activities (planning and reflective assessment, searching, selecting, and working with information items) presented in the task-based information interaction model (Järvelin et al., 2015). The fifth activity (synthesizing and reporting) in the model was left out because we wanted to focus on data interaction, not on writing. We also included questions about the research community and rules and norms (see Allen et al., 2011) because they are important to academic work. The semi-structured interview guide (see appendix 1, also in Korkeamäki et al., 2022) covered background questions, participants' research topics and processes, working in a research group (if applicable), participants' research data (i.e., what were the data like), data interaction (collecting, finding, selecting, analysing, archiving, and managing research data), and participants' views regarding research ethics, ownership and licensing related to the research data. Follow-up questions (e.g., for clarification or examples) were also asked (see Roulston, 2010, pp. 9-32).

The participants were asked to select a research project that was ongoing or recently concluded to be discussed in the interviews. Eight doctoral researchers and two participants with a doctorate focused on their doctoral research in the interview. One doctoral researcher chose to discuss a research project related to the near-completed doctoral research. Fourteen participants with a doctorate focused on their post-doctoral research in the interview.

The twenty-five semi-structured interviews (15 face-to-face and 10 by phone) were audio recorded (46 min – 1 h 16 min each, 24 h 26 min in total) and later transcribed verbatim (290 pages in total). Although we started with face-to-face interviews, we had to switch to remote interviews after the onset of the COVID-19 pandemic and to stay on schedule. At the beginning of the pandemic there was uncertainty about suitable videoconferencing tools that meet the EU's General Data Protection Regulation (GDPR), and therefore, we used a phone instead of videoconferencing. We also wanted to treat all participants the same for the remote interviews. Later, we learnt of suitable videoconferencing tools recommended by the research organization and offered some participants the option to demonstrate their work through a videoconferencing tool. However, this option was chosen by only one participant.

At the end of the semi-structured interview, participants were asked to demonstrate how they worked with their research data (Flanagan, 1954). The purpose was to help participants to talk about their interactions with their research data in greater detail and to complement the semi-structured interviews. Only twelve participants took part in the demonstration (11 face-to-face and one through Microsoft Teams). Ten demonstrations were captured on video (6 min – 21 min each, 2 h 6 min in total) and later transcribed (27 pages in total) and two by taking photographs. Names of persons or organizations were removed from the data. Some did not participate in the demonstration because they felt it was difficult to select just one part of their research to demonstrate. Online participation was low potentially because videoconferencing tools were not necessarily familiar to everyone in the early stages of the pandemic, or because it would have required participants to switch from the phone that was used for the remote semi-structured interview to a videoconferencing tool.

### **Data analysis**

We analysed the research data qualitatively by examining focus formulation episodes as sense-making instances (Souto et al., 2008), enabling

the examination of the focus formulation episodes that are meaningful for the media scholars' research work. The identification of the focus formulation episodes was based on the interpretation of the criticality, i.e., importance and impact, of the reflected episode for the research task process performance. Although the focus formulation episodes were not explicitly asked about in the interviews, they were a natural part of participants' narratives as they discussed the progress of their research from research planning to analysis.

We used a qualitative data reduction and cross-episode analysis approach, a modification from Watkins et al. (2022). The first phase of the analysis was data reduction and rearranging the focus formulation episodes. We started by identifying the focus formulation episodes from the interview data. This was done by looking for articulations where participants discussed reducing uncertainty or task complexity, increasing conceptual structuring, or increasing an understanding of what is meaningful for the task. Although it can be difficult to identify the articulations from the interview data, this was overcome by looking for wordings that were indicative of focus formulation episodes, such as 'the focus has then gone more to', 'then it somehow occurred to me' or 'I realised'. Each interview was read through several times, which also helped to identify the focus formulation episodes. Sometimes a participant talked about the same focus formulation episode in different parts of the interview. Therefore, we continued by rearranging the focus formulation episodes, which means that when two or more articulations within each interview were related to the same focus formulation episode, these were grouped together. Then, to convey the essence of each episode using participants' own words, we highlighted the wordings that were the most important in terms of episodic focus formulation.

The second phase was analysing across the focus formulation episodes. The analysis was iterative, during which the focus formulation episodes were read through several times, and it included the following steps. First, each focus

formulation episode was annotated in terms of the activities taking place (e.g., 'Searching, selecting, and collecting research data') and by creating assertions of the focus formulations episodes to connect their meanings to our research question (e.g., '[The focus formulation episode was about] where to find suitable research data and what are good sources for finding research data'). Second, the annotations were read through several times and categorized based on their similarities and differences. To meaningfully organize our research data, similar activities or sequences of activities were grouped together, resulting in the identification of three subtasks of knowledge creation (research framing, gathering the research data, and analysing the research data). Then, the subtasks were further analysed. Within each subtask, the assertions that were essentially about the same type of episodic focus formulations were identified and grouped together (e.g., source selection focus). We identified three types of episodic focus formulations for each of the three subtasks. The results of this phase of the analysis are reported in sections *Formulating focus for research framing*, *Formulating focus for gathering the research data*, and *Formulating focus for analysing the research data*.

Lastly, to further raise the level of abstraction, we summarised over the subtasks to cover the whole data interaction. This was done as follows. By further examining the types of episodic focus formulations over the three subtasks, we identified three underlying dimensions for episodic focus formulation. We identified that some episodic focus formulations were about formulating focus for the task, some were about formulating focus for the procedures of the task, and some were about formulating focus for the research data that are used as inputs for the task. We categorized the episodic focus formulations accordingly, in other words, based on whether they were related to the task, procedure, or research data. The results of the last phase of the analysis are reported in section *Summarising over data interaction*.

## Results

The results section is structured as follows. We first discuss episodic focus formulation by subtasks. Then we summarise over the subtasks to cover the whole data interaction.

### Formulating focus for research framing

In this study, research framing is understood as a subtask that takes place not just during the

research planning but spans several activities. Formulating focus for research framing was meaningful for participants' data interaction because it gave them direction on how to approach the data gathering and analysis. We identified three types of episodic focus formulations for research framing. They were contextualization focus, theoretical focus, and objectives focus (Figure 1).

| Contextualization focus  | Theoretical focus   | Objectives focus  |
|--|---|---|
| <ul style="list-style-type: none"><li>• What perspective of the broader research topic to focus on?</li><li>• What contexts are relevant for understanding the research topic and the research data?</li></ul> | <ul style="list-style-type: none"><li>• What is the theoretical framework for the research?</li></ul> | <ul style="list-style-type: none"><li>• What are the research objectives and questions?</li></ul> |

Figure 1. Formulating focus for research framing

*Contextualization focus* refers to the episodes of constructing the perspectives and contexts that are relevant for the research topic and research data. For example, one participant explained how gaining a broader understanding of the phenomenon under study was the reason for shifting the focus of the research topic.

*The topic of my dissertation was originally, it was more about [a different perspective] ... But then, in a way, with the [research] project ... and with my own research, the idea of what the future can be with [the phenomenon] has developed ... and through that the focus has then gone more to the [other type of] media use ... kind of broadened the understanding of what [the phenomenon] can be overall (P1).*

Another participant recounted how searching for contextual information led to a new

research topic for a sub study. Originally, the reason for using a specific data archive was merely to put the phenomenon under study in context. However, after making an interesting observation while reviewing the archival material and discussing it with colleagues, the participant decided to make it the topic of a sub study.

*Originally, it was not my intention to do this one sub study based on this [digital archival material] ... I just wanted to check if this one particular issue ... was just ... a local phenomenon or if it was something bigger ... I told my colleagues about my observation, and they were like 'that is definitely a very interesting observation ... we should start writing a paper on that and analyse it further' (P8).*

Furthermore, identifying contexts that are meaningful for understanding the research topic and the research data helped moving forward with a research project. A participant, whose primary research data consisted of films, explained that the need for gathering additional research materials was related to the moments when 'one realises that, wait, now I need these kinds of things for this, now I have to go read the [administrative documents]' (P24). The participant gave an example of how preparing for a conference presentation led to the realisation that additional research materials were needed to contextualize the topic for an international audience, which eventually also pushed the research project forward.

*I was speaking about a topic [at a conference] ... So, that is why I needed to start digging different kinds of [administrative documents and other additional materials], which was very good ... it kind of marked the beginning of the [research project]. I knew the project was about to start but I wasn't quite sure how to do it. But then I somehow accidentally started it so that the project got off to a good start. And I already had my own vision of what the films are going to tell (P24).*

Theoretical focus refers to the episodes of constructing a theoretical framework for the research. For example, a participant narrated how refining the research plan was a long process where working on the research plan, presenting it to others, and getting constructive feedback had meaning to the construction of the theoretical framework.

*I had the opportunity to present my research plan and I got really constructive feedback on it. And through that I actually got the theoretical basis for my research, or the kind of idea that I have now been working on, and it feels really logical at the moment (P3).*

Objectives focus was about the episodes of formulating the research objectives and questions. A participant explained how the research objectives became more focused after serendipitously identifying suitable research

data that enabled examining a particular concept from a specific viewpoint by making queries to the digital archive.

*In this work I was interested in this [concept] ... then it somehow occurred to me there, that aha, okay, it could be really interesting to examine [the concept from a specific viewpoint] ... it somehow came at the right moment that I got access to the [digital] archive, and I was then able to make those queries (P13).*

Others described how research objectives and questions became more focused during the analysis and writing. A participant, who was writing the first draft of the research paper, had written only as a side note that 'I feel like this [observation] in this phenomenon is somehow really important, that it is being photographed, talked about, and it keeps popping up everywhere' (P3). Then, while presenting the draft in a seminar, the participant described asking the others 'whether I should focus on that or is this something that is worth mentioning'. The participant narrated how the supervisors and peers were really encouraging that it is worth highlighting and continued, 'now I have considered that maybe this could be the angle for my first article'.

Research processes involved moving back and forth between the research phases, updating the research plan, and changing course in terms of how to proceed in the research. Sometimes, there were situations where the data gathering did not go as planned, causing the need to re-evaluate the research plan in terms of research questions, research methods and research data. One participant explained this as follows, starting with how the interview sample was different to what they meant to collect (not enough interviewees, and not at all certain types of interviewees) and how they needed to re-evaluate the research questions.

*We were supposed to do [a number of] interviews in this one event ... but then it did not turn out that way ... we also had to re-evaluate the research questions because, for example, we did not find at all [certain types of interviewees] ... and then, it is a different type of research frame ... we*



could have changed the method to get [a different kind of] sample ... but then we decided together that ... we did not have any more resources for that. In that sense it was dynamic. (P17)

## Formulating focus for gathering the research data

In this study, gathering the research data is understood as a subtask that includes the activities of searching, selecting, and collecting the research data. We identified three types of episodic focus formulations for gathering the research data. They were source selection focus, collection method focus, and data set focus (Figure 2).

| Source selection focus  | Collection method focus   | Data set focus  |
|---|---|---|
| <ul style="list-style-type: none"> <li>• What kinds of materials are available?</li> <li>• Where the materials can be found?</li> </ul> | <ul style="list-style-type: none"> <li>• How to collect the research data?</li> </ul> | <ul style="list-style-type: none"> <li>• What data items are relevant?</li> <li>• How much research data are needed?</li> </ul> |

Figure 2. Formulating focus for gathering the research data

Source selection focus refers to the episodes of figuring out what types of research data to gather and where to find them. For example, a participant explained how, initially, there was no clear picture of suitable research data and where they could be found, but it required investigating different possibilities and familiarizing with different sources and their contents.

*For example, when you go to the ... websites and do searches there, you notice how much have been written about some topic ... what type of material there is and ... how far the material extends historically and so on ... then I have realised that there is not that much to be found and then ... well, okay, what other way could there be and then I thought of these [journals], for example. I realised that often a lot of that discussion has taken place in [the journals] ... So, there wasn't like a clear starting point, that I somehow would have had a clear picture right from the start that, okay, these are now the materials that I am*

*going to collect and the places where I am going to look for [them] and so on (P14).*

Collection method focus refers to the episodes of constructing how to collect the research data. One participant, who collected discussion forum data, described the difficulty of figuring out how to collect the research data systematically.

*It turned out to be quite a challenge, how do I save the [research] data, because the forum's logic is such that the new discussions always appear on top ... it is in constant movement (P7).*

Another participant, who studied social media data using digital computational methods, talked about the difficulty of choosing search terms to find research data that would enable examining the phenomenon under study as accurately as possible.

*You need to narrow down the material somehow, and then you end up doing searches with search terms, which is kind of obvious that it is not going to cover*

everything ... But yeah, we started by testing them a bit, to see how we could collect some interesting material (P19).

Data set focus refers to the episodes of figuring out what data items to include in the research data, what data items are relevant and how much research data are needed. For example, a participant described how the criteria for relevant data items became more focused through familiarization with the data and through differentiating the relevant data items from the somewhat relevant. The participant said that, through several months of presence in social media, they started to notice 'which [data items] were the ones that were about the topic, and which ... were perhaps more about ... related topics' (P6). The participant went on to describe how gathering the research data felt difficult at first but through narrowing down the data items and through increased understanding of their relevance, it became easier.

*It was somehow really difficult to figure out [the data collection process] ... but maybe the narrowing down, when the thought process goes to, when you narrow down the [research] data, then after that the data collection is easy, when you know what to look for (P6).*

Participants also talked about how their understanding developed regarding how much

research data were needed. Sometimes, it was difficult to know how much research data were enough until careful familiarization with the research data during analysis. Furthermore, the data gathering was not always a separate phase but rather intertwined with analysing and writing. This was the case for a participant, whose research group was conducting a study using methods from the scientific tradition of humanities. When asked how they knew when they had enough research data, the participant replied:

*When it felt like the [book] chapters were ... in balance ... that the theoretical discussions were sufficiently concretized and ... there isn't too much stuff because there are easily ... an awful lot of examples that are then described ... the narrowing down ... where to focus on and how thoroughly (P23).*

### Formulating focus for analysing the research data

In this study, analysing the research data is understood as a subtask of knowledge creation. We identified three types of episodic focus formulations for analysing the research data. They were answer formation focus, analysis method focus, and data relevance focus (Figure 3).

| Answer formation focus  | Analysis method focus  | Data relevance focus  |
|---|--|---|
| <ul style="list-style-type: none"> <li>• How to structure and conceptualize the research data?</li> <li>• What is learned through the research data?</li> </ul> | <ul style="list-style-type: none"> <li>• How to analyse the research data?</li> <li>• How to bring together different methods of approaching the research data?</li> </ul> | <ul style="list-style-type: none"> <li>• What data items are relevant to the research problem?</li> </ul> |

Figure 3. Formulating focus for analysing the research data

*Answer formation focus* refers to the episodes that gradually and in small steps contribute to formulating answers to the research objectives and questions. There were episodes of structuring, conceptualizing, and learning through the research data during the analysis process. For example, a participant described how identifying something interesting in the research data and then figuring out how to conceptualise it to the research paper felt like an aha-experience.

*I have been thinking that there is a central sequence in the film, like from a certain [viewpoint], and I couldn't quite grasp it during the previous time I watched it ... I did realise that there was something of importance here, but I didn't really know how to conceptualise it to the article. Then, like after a couple of days you read something [and] when you watch [the sequence] again you get this aha-experience, that now I know, this is how it goes (P24).*

Some described the analysis as a process that is iterative, takes place in cycles, and gradually becomes more structured. Analysis was also described as a process that involves a dialogue between one's own interpretations and the research literature. Writing was a central way to work on the analysis to capture what is essential in the phenomenon.

*Often, the preliminary version is one where you might be describing in too much detail and ... overall, too descriptive. But maybe that kind of a phase is necessary, because then you'll have a better idea, or ... a better grip on the [research] data through this kind of an overly descriptive text, and from that you can condense and see what is relevant (P25).*

A participant, who conducted action research, described episodes of learning through the research data as inherent to the research process. For example, writing a blog during the research process was a way to 'formulate thoughts in a more structured way' (P12) and one 'where we stopped to think about what we have learned so far'. The analysis involved extracting

the most important aspects of the phenomenon under study.

*Maybe the way we have used this material, the way we have analysed, we have used ... the whole trajectory [of our research] and looked for the kinds of distillations and experiences (P12).*

*Analysis method focus* refers to the episodes of constructing the methods of how to analyse the research data. For example, a participant described how they were thinking about what can be done with the research data and how they eventually understood that additional and different analysis methods were needed to be able to say anything interesting based on the research data.

*And then we ... wonder about [the data] ... and then, no, that we cannot say anything that is interesting enough unless we actually start watching these videos ... to really understand what was going on there (P19).*

Another participant talked about constructing the analysis from the point of view of how to annotate (e.g., in what level of detail) the qualitative research data in such a way that would help capturing what is meaningful in the research data to the research problem.

*I wanted to label the parts [of the texts] as much as possible so that if I would have an idea, if suddenly, later, there would be a lot of talk, say, about [a topic] in those discussions, then I would not be able to find [them] if I don't label them. So, there was a constant difficulty of what to label, as one can't leave almost anything [unlabelled] (P7).*

The participant went on to describe how an understanding of the ways the annotation should be performed developed during the analysis. 'Eventually you realise that you cannot keep labelling things with the same precision' (P7).

There was also the question of how to bring together different methods of approaching the research data. For example, in a research group that included quantitatively and qualitatively

oriented researchers, the different approaches were brought together through discussions within the research group.

*The approaches in relation to the research data are terribly different ... [we] need to go through it a lot ... They can't do the analysis if they don't know exactly what to look for [in the research data]. And then again, a qualitative researcher may not be able to say exactly what to look for [in the research data] because we want to leave everything open (P11).*

Data relevance focus was about the episodes of identifying the data items from the data set that are the most important for the research problem. Sometimes, the focus was formed quite easily during the close reading of the research data.

*At that first stage, the [data items] that were ... somehow important or interesting in terms of my theoretical framework stood out quite easily ... I went through those certain [data items] over and over again and ... little by little, the research question and what you want to do with the article became sharper ... [my] attention was quite naturally drawn to certain [data*

*items] and to certain parts in the [data items] (P20).*

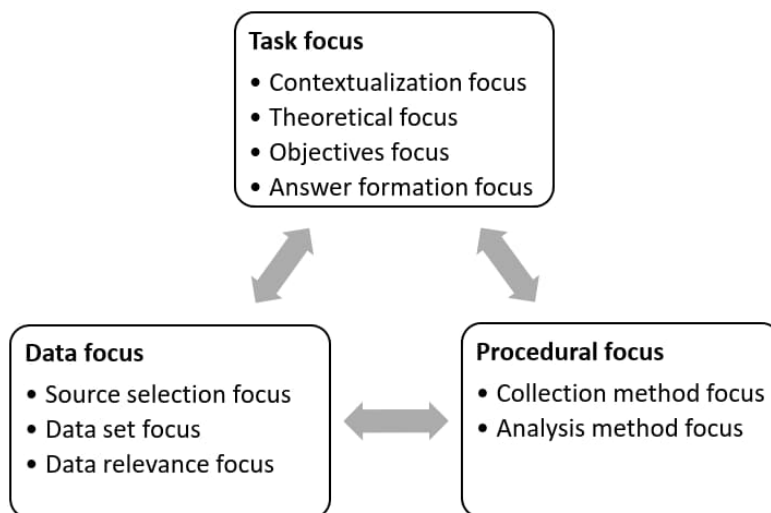
Other times, identifying the relevant data items from the data set was accompanied by uncertainty. A participant explained that this was because identifying the relevant data items for the research questions required interpretation.

*I need to use interpretation quite a lot ... sometimes it scares me a bit because, well ... what if I have missed some things ... because after all, it is largely based on whether I have noticed interesting things in [the data set] (P15).*

*Somehow, the uncertainty about myself, whether I have found the relevant ones in [the data set] or skipped some parts that would have answered my research questions (P15).*

### Summarising over data interaction

In the above sections, we presented the episodic focus formulations within the subtasks of research framing, data gathering, and data analysis. In Figure 4, we summarise over the subtasks to cover the whole data interaction by abstracting the types of episodic focus formulations into three categories.



**Figure 4.** Episodic focus formulation in data interaction encompasses task focus, data focus and procedural focus formulation



First, we identified that some episodic focus formulations were, in more abstract terms, about formulating focus for the task of creating new knowledge. We named the category *task focus*. Formulating focus for the task includes episodes of constructing the perspectives and contexts that are relevant for the task (contextualisation focus), constructing the theoretical framework for the task (theoretical focus), formulating the task's objectives and questions (objectives focus), and formulating answers to them (answer formation focus).

The second category, *procedural focus*, is about episodes of formulating focus for the procedures, that is, how to perform the task of creating new knowledge. Formulating focus for the procedures includes episodes of constructing the methods of how to collect the research data (collection method focus) and analyse the research data (analysis method focus).

The third category, *data focus*, is about episodes of formulating focus for the research data that are used as inputs for the task of creating new knowledge. It includes figuring out what types of research data to gather and where to find them (source selection focus), figuring out what data items to include in the research data (data set focus), and identifying the data items from the data set that are the most important for the research problem (data relevance focus). In other words, focus for the research data formulates episodically on the levels of data sources, data sets, and data items.

Overall, the analysis showed that episodic focus formulation in media scholars' data interaction encompasses task focus, data focus and procedural focus formulation.

## Discussion

In this study, we examined episodic focus formulation in media scholars' data interaction. We defined episodic focus formulation as episodes occurring in the task process performance, each contributing to the focus formulation for the task. We asked the research question: What is episodic focus formulation like in media scholars' data interaction?

In response to the research question, we first analysed the types of episodic focus formulations by subtasks. We identified three subtasks of knowledge creation, each with three types of episodic focus formulations. The types of episodic focus formulations were contextualization focus, theoretical focus and objectives focus for research framing; source selection focus, collection method focus, and data set focus for gathering the research data; and answer formation focus, analysis method focus, and data relevance focus for analysing the research data. Then, by abstracting the types of episodic focus formulations over the subtasks to cover the whole data interaction, we identified three categories of episodic focus formulations. The categories were formulating focus for the task of creating new knowledge (task focus), formulating focus for how to perform the task (procedural focus), and formulating focus for the research data that are used as inputs to create new knowledge (data focus). Overall, the analysis showed that episodic focus formulation in media scholars' data interaction encompasses task focus, data focus and procedural focus formulation.

This study has implications for theory development by refining the existing models. We start by discussing our study in relation to Kuhlthau's (2004) and Vakkari's (2001) studies. First, Kuhlthau compared the function of a focus to a hypothesis, turning point, elaborative choice or guiding idea that gives direction for the task process performance. These general characterizations align with how we defined focus formulation in this study. However, both Kuhlthau and Vakkari defined focus formulation as one task stage in the overall task process performance. In contrast, our approach was to study focus formulation as an episodic phenomenon where focus formulation is not just one task stage but happens gradually. This is a new approach to studying focus formulation. Second, Kuhlthau and Vakkari mainly examined learning tasks or research planning with interacting with literature, not per se with interactions with research data during the research process. Vakkari's study stops where the data interaction begins. Moreover, they mainly discussed focus formulation in relation to formulating a focused

perspective of a topic or formulating a focused research problem. These are comparable to contextualization focus and objectives focus identified in our study. However, our analysis by subtasks also showed other types of episodic focus formulations in the task process performance. Therefore, this study provides a more nuanced picture of what focus formation looks like in the complex and long-lasting research tasks that involve data interaction.

This study brings a new theoretical perspective also to Bron et al.'s (2016) model of media scholars' research cycle. The results of this study align with Bron et al. in that there are similarities in how media scholars described their research activities and research question formation. However, Bron et al. seemed to conceptualise focus formulation differently. According to them, the contextualization phase in their model matches with the focus formulation and collection stages in Kuhlthau's (2004) information search process model. This could be because Bron et al. mainly discussed focus formulation in relation to research question formation, similarly with Vakkari (2001), and discovered that media scholars' research questions changed especially during the targeted data gathering and analysis activities (in the contextualization phase). However, the results of this study depict media scholars' episodic focus formulation as a sense-making activity that involves taking small steps towards finding a focus, the nature of which is determined through the subtasks and their goals. In other words, different subtasks may have different types of episodic focus formulations. Moreover, this study showed that media scholars' episodic focus formulation concerned not just research question formation, but also other kinds of episodic focus formulations for research framing, data gathering and analysis.

This study also has implications for supporting data interaction. Dervin (1998) stated that information systems should be designed to better support sense-making processes. Similarly, researchers' data interaction could be supported by supporting their focus formulation episodes. First, collaborative interactions and bouncing ideas (Willson, 2022)

could be fruitful for episodic focus formulation in data interaction. For example, in this study, some focus formulation episodes involved participants' discussions with supervisors or peers. Second, episodic focus formulation in data interaction could be supported by exploratory search tools. Earlier research showed that general background information was needed especially when trying to find a focus (Vakkari, 2001). Exploratory search tools could also support task-doers' cognitive and metacognitive activities (Li et al., 2023), insights (Wang and Liu, 2023) and creative thinking (Chavula et al., 2024). Also, the development of generative artificial intelligence tools offers new kinds of possibilities for supporting sense-making processes and episodic focus formulation by enabling conversational search. Furthermore, different subtasks may need different kinds of support.

When it comes to transferability of the results, the types of episodic focus formulations are not necessarily the same for everyone, as there may be differences in the research processes. Furthermore, the analysis was based on what participants narrated during the interviews. Therefore, the episodic focus formulations identified in this study did not necessarily include all possible aspects involved. However, in the more abstract level, we identified the dimensions of task, procedure, and data in episodic focus formulation. This echoes other general distinctions between task, procedures, and inputs. In comparison, Byström (1999) distinguished between task information, domain information, and task-solving (procedural) information as types of information needed for the task. Inputs have been discussed as the resources and tools for task-based information interaction (Järvelin et al. 2015) and as part of sense-making processes, where inputs are used for constructing bridges over the gaps (e.g., questions) encountered (Souto et al., 2008). In data interaction context, Koesten et al. (2021) similarly discovered that users make sense of the data on the levels of data sets and data items. The contribution of our study is that we examined the dimensions of task, procedure, and inputs from the sense-making perspective. Although this study was conducted in media studies contexts, they

could be observed more generally in different contexts. However, more empirical research is needed to determine whether the results can be generalized to other groups.

In media studies, research is often conducted using heuristic research methods where formulating focus is central. Also, the interactions with research data are complex. Episodic focus formulation in data interaction is essentially a sense-making process, and therefore it needs different kind of support compared to traditional information retrieval systems.

## Conclusion

Focus formulation has traditionally been seen as a singular task stage in the overall task performance. In this study, we examined what

is episodic focus formulation like in media scholars' data interaction. The results of this study showed that focus formulation happened episodically during subtasks of research framing, data gathering and analysis. Summarising over the subtasks showed that episodic focus formulation in data interaction encompasses task focus, data focus and procedural focus formulation, which inherently differ from each other. The results of this study can be used for theory development and for supporting data interaction.

## Acknowledgements

This work was supported by The Emil Aaltonen Foundation and Academy of Finland (Grant No. 351247).

## About the authors

**Laura Korkeamäki** is a Doctoral Researcher in Faculty of Information Technology and Communication Sciences at Tampere University, Finland. Her research interests are in information interaction research. She can be contacted at [laura.korkeamaki@tuni.fi](mailto:laura.korkeamaki@tuni.fi).

**Heikki Keskustalo** is a Senior Research Fellow in Faculty of Information Technology and Communication Sciences at Tampere University. His research interests include natural language processing, text retrieval and task-based information interaction. He can be contacted at [heikki.keskustalo@tuni.fi](mailto:heikki.keskustalo@tuni.fi).

**Sanna Kumpulainen** holds the position of Associate Professor in Information Studies at Tampere University, Finland. Her research interests include exploring the dynamics of human interaction with information and developing strategies for its facilitation. Her research pursuits encompass also interactive information retrieval, digital libraries, and the use of research data. You can contact her via email at [sanna.kumpulainen@tuni.fi](mailto:sanna.kumpulainen@tuni.fi).

## References

- Allen, D., Karanasios, S. & Slavova, M. (2011). Working with activity theory: context, technology, and information behavior. *Journal of the American Society for Information Science and Technology*, 62(4), 776-788. <https://doi.org/10.1002/asi.21441>
- Belkin, N. J., Oddy, R. N. & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61-71. <https://doi.org/10.1108/eb026722>.
- Bron, M., Van Gorp, J. & de Rijke, M. (2016). Media studies research in the data-driven age: how research questions evolve. *Journal of the Association for Information Science and Technology*, 67(7), 1535-1554. <https://doi.org/10.1002/asi.23458>.
- Byström, K. (1999). *Task complexity, information types and information sources: examination of relationships*. University of Tampere. (Academic dissertation). <http://urn.fi/URN:ISBN:978-952-03-1893-2>

- Byström, K. & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology*, 56(10), 1050-1061. <https://doi.org/10.1002/asi.20197>
- Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191-213. [https://doi.org/10.1016/0306-4573\(95\)80035-R](https://doi.org/10.1016/0306-4573(95)80035-R)
- Byström, K. & Kumpulainen, S. (2020). Vertical and horizontal relationships amongst task-based information needs. *Information Processing & Management*, 57(2), Article 102065. <https://doi.org/10.1016/j.ipm.2019.102065>
- Chavula, C., Choi, Y. & Rieh, S. Y. (2024). Searching for creativity: how people search to generate new ideas. *Journal of the Association for Information Science and Technology*, 75(4), 438-453. <https://doi.org/10.1002/asi.24857>
- Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36-46. <https://doi.org/10.1108/13673279810249369>
- Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, 45(3), 171-212. <https://doi.org/10.1108/eb026843>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.
- Gaikwad, M. & Hoerber, O. (2019). An interactive image retrieval approach to searching for images on social media. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 173-181. <https://doi.org/10.1145/3295750.3298930>
- Given, L. M. & Willson, R. (2018). Information technology and the humanities scholar: documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6), 807-819. <https://doi.org/10.1002/asi.24008>
- Huurdean, H. C., Wilson, M. L. & Kamps, J. (2016). Active and passive utility of search interface features in different information seeking task stages. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*. Association for Computing Machinery, New York, NY, USA, 3-12. <https://doi.org/10.1145/2854946.2854957>
- Huurdean, H. C. & Kamps, J. (2020). Designing multistage search systems to support the information seeking process. In W. T. Fu & H. van Oostendorp, (Eds.), *Understanding and Improving Information Search*. Springer. (Human-Computer Interaction Series). [https://doi.org/10.1007/978-3-030-38825-6\\_7](https://doi.org/10.1007/978-3-030-38825-6_7)
- Jensen, K. B. (Ed.). (2012). *A handbook of media and communication research: qualitative and quantitative methodologies* (2<sup>nd</sup> ed.). Routledge.
- Järvelin, K., Vakkari, P., Arvola, P., Baskaya, F., Järvelin, A., Kekäläinen, J., Keskustalo, H., Kumpulainen, S., Saastamoinen, M., Savolainen, R. & Sormunen, E. (2015). Task-based information interaction evaluation: the viewpoint of program theory. *ACM Transactions on Information Systems (TOIS)*, 33(1), Article 3 (March 2015), 30 pages. <https://doi.org/10.1145/2699660>
- Koesten, L., Gregory, K., Groth, P. & Simperl, E. (2021). Talking datasets—understanding data sensemaking behaviours. *International Journal of Human-Computer Studies*, 146, Article 102562. <https://doi.org/10.1016/j.ijhcs.2020.102562>



- Korkeamäki, L., Keskustalo, H. & Kumpulainen, S. (2022). Task information types related to data gathering in media studies. *Journal of Documentation*, 78(7), 528-545. <https://doi.org/10.1108/JD-04-2022-0082>
- Korkeamäki, L., Keskustalo, H. & Kumpulainen, S. (2024). Types of domain and task-solving information in media scholars' data interaction. *Journal of the Association for Information Science and Technology*, 75(4), 454-468. <https://doi.org/10.1002/asi.24863>
- Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339-355. <https://doi.org/10.1108/eb026918>
- Kuhlthau, C. C. (2004). *Seeking meaning: a process approach to library and information services* (2nd ed.). Libraries Unlimited.
- Li, Y., Crescenzi, A., Ward, A. R. & Capra, R. (2023). Thinking inside the box: an evaluation of a novel search-assisting tool for supporting (meta)cognition during exploratory search. *Journal of the Association for Information Science and Technology*, 74(9), 1049-1066. <https://doi.org/10.1002/asi.24801>
- Liu, J., Sarkar, S. & Shah, C. (2020). Identifying and predicting the states of complex search tasks. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. Association for Computing Machinery, New York, NY, USA, 193-202. <https://doi.org/10.1145/3343413.3377976>
- Meho, L. I. & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, 54(6), 570-587. <https://doi.org/10.1002/asi.10244>
- Melgar, L., Koolen, M., Huurdeman, H. & Blom, J. (2017). A process model of scholarly media annotation. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 305-308. <https://doi.org/10.1145/3020165.3022139>
- Palani, S., Ding, Z., MacNeil, S. & Dow, S. P. (2021). The "active search" hypothesis: how search strategies relate to creative learning. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 325-329. <https://doi.org/10.1145/3406522.3446046>
- Palmer, C. L. & Neumann, L. J. (2002). The information work of interdisciplinary humanities scholars: exploration and translation. *Library Quarterly*, 72(1), 85-117. <https://doi.org/10.1086/603337>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Sage.
- Roulston, K. (2010). *Reflective interviewing: a guide to theory and practice*. SAGE Publications Ltd. <https://www.doi.org/10.4135/9781446288009>
- Ruthven, I. (2019). The language of information need: differentiating conscious and formalized information needs. *Information Processing & Management*, 56(1), 77-90. <https://doi.org/10.1016/j.ipm.2018.09.005>
- Soufan, A., Ruthven, I. & Azzopardi, L. (2021). Untangling the concept of task in information seeking and retrieval. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR '21)*. Association for Computing Machinery, New York, NY, USA, 73-81. <https://doi.org/10.1145/3471158.3472259>

- Souto, P., Dervin, B. & Savolainen, R. (2008). Designing for knowledge worker informings: an exemplar application of sense-making methodology. In *Proceedings of the American Society for Information Science and Technology*, 45(1), 16 pages.  
<https://doi.org/10.1002/meet.2008.1450450221>
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & research libraries*, 29(3), 178-194.
- Toms, E. G. (2019). Information activities and tasks. In K. Byström, J. Heinström & I. Ruthven, (Eds.), *Information at work: information management in the workplace* (pp. 33-61). Facet Publishing.
- Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1), 44-60.  
<https://doi.org/10.1108/EUM0000000007075>
- Wang, X. & Liu, C. (2023). Finding the aha! moment of search: a preliminary examination of insight learning during search. *Proceedings of the Association for Information Science and Technology*, 60(1), 421-432. <https://doi.org/10.1002/pr2.800>
- Watkins, K. E., Ellinger, A. D., Suh, B., Brenes-Dawsey, J. C. & Oliver, L. C. (2022). Further evolving the critical incident technique (CIT) by applying different contemporary approaches for analyzing qualitative data in CIT studies. *European Journal of Training and Development*, 46(7/8), 709-726.  
<https://doi.org/10.1108/EJTD-07-2021-0107>
- Willson, R. (2022). "Bouncing ideas" as a complex information practice: information seeking, sharing, creation, and cooperation. *Journal of Documentation*, 78(4), 800-816.  
<https://doi.org/10.1108/JD-03-2021-0047>

# Appendix I

## Interview guide for semi-structured interview

### Background information

- What is your educational background?
- What is your current job title?
- How long have you worked as a researcher?
- What is your field of research?

### Research topic

- What is your research topic? If possible, select an ongoing or recently completed research that you can still remember well.

### Research process

- Where did you get the idea for the research topic?
- What are your research goals?
- What are your research questions?
- Can you distinguish phases from your research process? What are they? Where are you now in this continuum?

### Working in a research group (if applicable)

- What is your research group like?
- What is your role in the research group?
- What is the division of labour in the group?
- Do you have common research data?
- Do you have common tools?

### Research data

#### *Description of the research data*

- What is your research data like?
- In what form is your research data?

#### *Collecting the research data*

- What are your research methods?
- How did you collect the research data?
- Did you collect the research data in one or several sessions?
- Where do you keep your research data?
- How do you organize your research data?
- How do you know when you have enough data?
- Did you have all the necessary information you needed to be able to use the research data in your research?

- Have there been any difficulties getting the research data for research purposes?

#### ***Finding the research data***

- Where did you find the research data?
- How did you know where to look for the research data/participants for the research?

#### ***Selecting the research data***

- Why (and how) did you choose the research data?

#### ***Analysing the research data***

- How do you analyse the research data?
- Did you start analysing the data before all was collected?
- Did the data require preprocessing for the analysis?

#### ***Archiving the research data***

- Have you thought about what to do with the research data after the research is completed?

#### ***Managing the research data***

- Have you planned your data management beforehand?
- How about during the research?

#### ***Research ethics, ownership, and licensing***

- What ethical questions did you need to think about concerning your research data?
- Were there any issues related to ownership and licensing of the research data?