# Improving the Cost Structure of Sensemaking tasks: Analysing User Concepts to Inform Information System Design

Simon Attfield[1] and Ann Blandford[1]

[1]UCL Interaction Centre, MPEB 8th floor, University College London, Gower Street, London WC1E 6BT, UK
{s.attfield. a.blandford, @cs.ucl.ac.uk}

**Abstract.** In many everyday contexts people interact with information systems in order to make sense of a domain of interest. However, what this means and how it can best be supported are poorly understood. In particular, there has been little research on how to develop system representations that simplify naturally occurring sense making processes by matching people's conceptualizations of the domain. In this paper we draw on Klein et al's data-frame theory and Russell at al's notion of cost-structures in sensemaking to propose an approach to understanding sensemaking that supports reasoning about system requirements. The two key elements of the approach are the identification of the process and the transformational steps within that process that could benefit from support to reduce costs, and the identification of primary concepts which are cued by information in the context of a given sensemaking task and domain, and around which users integrate information to form a structured understanding. Our general principle is that by understanding a sensemaking transformation in terms of its source data and the integrating structures it creates, one is better able to anticipate the evolving information needs that it tends to invoke. We test this approach with a case study of fraud investigation performed by a team of lawyers and forensic accountants and consider how to support the elaboration of prototypical user-frames once they have been invoked.

**Keywords:** sensemaking, conceptual design, fraud investigations.

## 1 Introduction

In many everyday professional and personal contexts, users interact with information systems in order to develop a 'picture' or 'model' of some domain [5, 19]. And yet ideas about what this means and how it can best be supported are still poorly formed. Consequently, guidance for the design of innteractive systems that support the development of user understanding is typically generalized, heuristic and based around exemplars. Whilst there is value in this, there is a need to understand the

specifics of the cognitive processes people go through in different situations in order to reason about specific system design solutions.

The process of constructing understanding from information (in the widest sense) has been referred to as 'sensemaking'. People engage in sensemaking when they find themselves in situations they want to understand. Sensemaking is a process of imposing structure on the unfamiliar in understanding it. There are many models of sensemaking, developed within different research traditions; examples include Dervin's Sensemaking Methodology as a means of eliciting users' information needs and their contexts [5], Pirolli and Card's model of sensemaking for intelligence analysis [14], Russell et al's learning loop complex [16] and Klein et al's data-frame theory [12].

In this paper we build on Klein et al's data-frame theory and Russell et al's ideas about cost structures in sensemaking to develop a theoretically informed approach to the analysis of sensemaking tasks in a way that yields specific system requirements. We focus on these two approaches because their semi-formalized descriptions of representational change provide useful accounts of the interplay between top-down and bottom-up cognitive processes which is central to sensemaking. However, they do not offer generalised leverage for information systems design; a question we consider here. We test our approach through a case study of a fraud investigation performed by a team of lawyers within a large law firm.

## 2 Background

'Sensemaking' is the process through which people make sense of their worlds. As a topic of research it spans a number of disciplines, including Human Computer Interaction, Organisational Studies, Naturalistic Decision Making, and Information Science. In Human Computer Interaction sensemaking research has tended to focus on sensemaking in electronic environments. This typically involves tasks that extend over time, and include searching and exploring large information collections or datasets, and integrating information into a coherent understanding. During such sensemaking tasks, people often create personalized, external, and often structured representations. These retain important information relevant to the task and can also be a resource for determining important elements and relationships. An interest in such representations has itself formed a significant focus in Human Computer Interaction research which identifies a concern with sensemaking (see, for example, [14, 15, 16]).

### 2.1 Data-Frame Symbiosis and the Learning Loop Complex

Klein et al's data-frame theory of sensemaking [12] evolved out of Klein's work on naturalistic decision making in relation assessment in command and control [11]. Klein et al [12] focussed on sensemakers' internal, cognitive representations. Following Weick [23], they view sensemaking as a ubiquitous cognitive function.

According to the data-frame model, people react to information within their environments by accommodating it within 'frames'. A frame is seen as an integrating

structure with slots for data. In this respect they draw a link with previous notions of frames [9], scripts [17] and schemata [2].

In outline, the data-frame model argues that when faced with a situation sensemaking involves abductively inferring a frame based on a few key cues or 'anchors' within that situation. The frame then 'connects the dots' and offers a plausible interpretation of what the situation is—an interpretation that can support explanation, prediction and decision response.

Importantly, frames are seen as extending beyond the cues. Consequently, they determine expectations about the world, including the possibility of specific kinds of information that could be found to elaborate the frame. In common with Starbuck and Milliken [20], Klein et al argue that as an interpretation of a situation a frame can act as an information filter and determine what is subsequently noticed. Where information is noticed that conflicts with a current frame, however, the plausibility of the frame (or indeed the data) can be challenged and a new frame required. Hence, sensemaking is seen as a process of framing and re-framing.

Klein et al argue that people have repertoires of frames derived from experience that they can apply to new situations and that this, for example, underlies the distinction between experts and novices. Experts, it is argued, reason in the same way as novices. The difference lies in the fact that experts have richer repertoires of frames that are better differentiated. These allow them to make sense of a greater variety of situations and to be more precise about expectations.

The data-frame model identifies seven kinds of frame-based operations applied in the sensemaking process. These are:

- *Connecting data and frame*: Identifying the situation with a frame.
- *Elaborating a frame*: The frame is elaborated with detail. New data does not challenge it.
- *Questioning the frame*: Expectations created by a frame are violated by unexpected data. The frame or the data can be questioned.
- *Preserving the frame*: Inconsistent data are explained away or simply ignored and the frame is maintained (significant for explaining confirmation bias).
- *Comparing multiple frames*: Multiple frames are explored. These may be similar but mutually inconsistent (for example, alternative medical diagnoses).
- *Reframing*: A replacement frame is adopted when it is suggested by data (a frame also defines what counts as data).
- *Seeking a frame*: Trying to find (recall) or construct an appropriate frame.

For Russell et al [16], information representations also form a central component in sensemaking, although Russell et al are concerned with the creation of external, user-generated representations. They observed a group of course designers who defined schemas within a hypertext system for capturing information relevant to the content of a new course. The instantiated schemas then provided a resource for automated clustering in order to identify core concepts within the material.

Russell et al used findings from this study to motivate a model of sensemaking called the 'learning loop complex'. According to this model, a sensemaker generates representations (schemas) to capture salient information (generation loop), uses these to guide the identification of information of interest, and then encodes found information within the representations (data coverage loop). During data coverage, however, salient information can be discovered which does not fit the representational

scheme (residue), and the representation can be changed to accommodate (representational shift loop), followed by further data coverage, etc.

In another example, they described a case study of someone making sense of the laptop market to decide which to buy. The sensemaker created a table to hold salient data (specifications etc.) and then explored available literature on different models to populate the table. During exploration, however, they made changes to the table according to new decisions about which properties were most relevant to their decision and the extent to which the relevant information was retrievable.

Despite the fact that Russell et al's model is concerned with external representations and Klein at al's model is concerned with internal representations, their accounts are strikingly similar. In both, representations reflect an understanding generated about some domain, and lead to an interplay between bottom-up and top-down processing. Accordingly, the representations evolve through encounters with information triggering and shaping the structures and these, in turn, guide subsequent encounters with information—affecting what is sought and what is noticed.

### 2.2 Cost Structures in Sensemaking

Russell et al argued that sensemaking tasks can be decomposed into their constituent activities and that optimising sensemaking involves selecting methods to maximise the expected cost-to-gain ratio of individual steps. Given any particular method, there is a fixed cost-to-gain function according to which a given cost (or effort) provides a given gain. Adaptations that sensemakers make to their method change the characteristics of the cost-to-gain function (hopefully in their favour).

For example, they observed in their case-study of course designers that the main user-cost was incurred by the manual extraction of data (finding relevant documents, selecting the right information, transforming this into canonical form suitable for the external representation). The payoff for this investment of effort, however, came from the fact that extracting the data and encoding it into hypertext schemas allowed the course designers to use a computer to perform automated clustering; a powerful technique whereby they were able to reveal recurring concepts within the content. The payoff for adjustments (representational shifts) to the schemas that occurred during extraction was an improvement in the reliability of human encoding and the ultimate utility of the automated analysis.

They analysed the cost of sensemaking as the sum of the costs of generating representational schemas, finding relevant information and instantiating the schemas. They argued that technology that improves the cost-to-gain ratio of one step can free up time to invest in others.

## 3 An Approach to Identifying Costs and Representations

Understanding the internal processes of sensemaking in a given situation can highlight requirements for tools to support that process. Within our approach, there are two main steps:

1. *Identify opportunities for achieving the greatest benefit*. This involves analysing the sensemaking process and identifying the most expensive elements (in terms of time and effort). Identifying opportunities for achieving the greatest benefit involves developing a sensemaking process model for the activity to be supported and, within that, identifying opportunities for making local improvements that will have the greatest impact on the process as a whole. Russell et al [16] and Attfield et al [1] show that sensemaking tasks often involve users in sequences of external representational transformations. An approach to identifying opportunities for local improvement is to capture these in a 'process-resource' model [1]; this model includes explicit representation of the key information transformations that users perform during sensemaking and the main information resources that each transformation uses and produces. Once a model has been developed, the next step is to identify the activities with greatest potential for cost saving.

2. *Understand the frames of significance to the sensemaker and optimise the conceptual fit between user and system* [3]. This begins with the ontology or 'frame of reference' with which a user understands and interprets a domain, and so will permeate the sensemaking process as that process moves from large amounts of unstructured information towards more concise and parsimonious representations. The significance of this is that the ontology with which the user understands and interprets a domain is not necessarily represented in the information system they use. Consequently, the burden is on the user to translate from one ontology to the other. Blandford et al [1] recognize the importance of identifying the concepts that users are working with to design systems that support manipulation of and reasoning about those concepts, and in the more general problem-solving literature the roles of external representations are well recognised (e.g. [4]).

Klein et al's data frame symbiosis theory provides a useful framework here. In Klein et al's theory, a frame is a mental representation which corresponds to a domain concept. The concept might be anything that is salient to the user in terms of forming their understanding of the domain e.g. a product, a terrorist incident, a performance, a device, a company etc. Once instantiated through interaction with a situation—meaning that some cue identifies the frame as a plausible interpretation of some situation—a frame can become a 'centre of gravity' or focus around which relevant information can be related.

Frames themselves are determined by a priori knowledge, which itself arises from experience. Hence they not only integrate *known* information but suggest *unknown* elements too. For example, you may come to know that a terrorist incident took place in a particular location, but you may not know the time and the date or the number of casualties—facts which it may be useful to discover.

Frames evolve continuously as new situations are encountered, reasoned over and learned about. And the particular information elements or constituents that are considered relevant for a given purpose will vary depending on the sensemaker's broader interests and what information they see as usefully informing them. The existence of frame constituents, the fact that at any point in a sensemaking task these may be relatively well defined and yet unknown (i.e. known unknowns), and their possible importance for furthering the sensemaker's interests, means that frames offer a basis for explaining the content and evolution of information needs—an issue very significant to the design of information systems. Hence, analyzing typical frames

within a given sensemaking context and how they are cued and evolve, provides a basis for reasoning about design.

In summary, our approach involves:

1. *Identify opportunities for achieving the greatest benefit.*
    a. Develop a 'process resource' model of the sensemaking process.
    b. Identify transformations that represent major overheads.
2. *Understand the frames of significance to the sensemaker and optimise the conceptual fit between user and system.*
    a. Understand transformations in terms of the users' frames.
    b. Design to reduce the costs involved in developing key frames.

## 4 A Case Study – Investigating Corporate Fraud

The case study was of a large fraud investigation undertaken by a corporate law firm in London. Ten in-depth interviews were conducted with nine lawyers who had worked on the case. For reasons of client confidentiality, it was not possible to gather real-time observational data, but key sense-making artefacts emerging from the case were made available for study. The interviews aimed at understanding global and local processes of the investigation, with a particular focus on the relationship between sensemaking and the use of external representations and automation.

Participants were recruited through a combination of snowball [10] and theoretical sampling [8]. Theoretical sampling was used to focus in on emerging issues. Following the practices of Grounded Theory [8], data gathering and analysis were interleaved. Data from the first five interviews were analysed to develop a preliminary model of the sensemaking processes of the legal team, covering both individual sensemaking and team co-ordination. This model was tested and refined with participants through subsequent interviews and analysis, and is the model presented in this paper (figure 1, below).

Interviews were conducted in an open and informal way. Each lasted from 45 minutes to 1hr 40 minutes. Early interviews focused on how the individual investigator had identified and worked with documents of interest, what tools they had used, what external representations they had generated and how they had used them, and how they coordinated their individual activities with those of the team. Many interviews were conducted using supporting artefacts. These included representations that had been generated by the investigators such as evidence tables from the investigators' final report, and software that had been used, loaded with the investigation data. These artefacts provided a reference point for discussing and reconstructing specific aspects of the investigations.

Interviews were transcribed and analysed through open coding [21] and the generation of process models used to describe the activities. Models decomposing the process were developed on an ongoing basis. Later interviews included discussion of these models. In this way, the models were verified and elaborated by participants on an ongoing basis and through constant comparison against the data [21].

### 4.1 Step 1a: Develop a 'Process Resource' Model of the Sensemaking Process

The investigation occupied a team of around 30 lawyers and forensic accountants for three months. Broadly, the objectives were to discover whether a particular kind of fraud had taken place in a company and, if so, who had been complicit. Figure 1 shows the investigation process in overview as a 'process-resource' model, as described above. In Figure 1, rectangular boxes represent processes. Arrows between them represent flow of information. This flow occurred through resources (marked against each arrow) created or modified by one process and used by another. In this way each process acted as a transformation.

The model begins and ends at the top with client discussion (a) where objectives are agreed and findings reported (client discussion was also returned to throughout). Following some early evidence review the objectives were explicitly characterised as set of relatively discrete investigation 'issues'. These were thematic lines or enquiry each of which had or developed associated theories, questions and facts.

In the model, issues propagate down as a resource influencing all other processes and can also be refined by any process in the light of discoveries. Processes can also

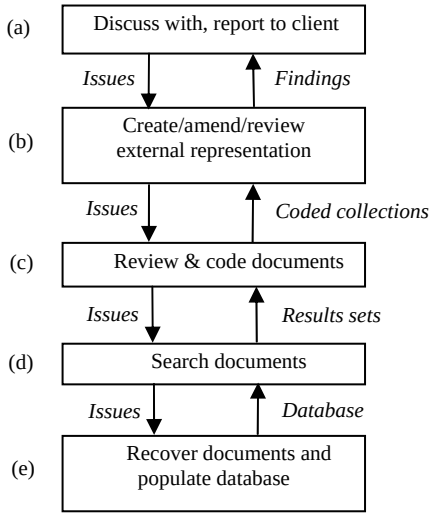


**Fig. 1.** An overview of the investigation process

be skipped depending on what resources are available and what are required at any one time. At the beginning of the investigation the issues informed the recovery of evidential documents (e). Documents (mostly electronic) were recovered from computers used by the company under investigation. Information was gathered from around 500 locations (including email servers) and the collection was ultimately equivalent in size to about 8.5 million novels. Evidence was also gathered through witness interviews (although here we focus on documents). Once recovered, documents were added to a database.

The database provided a resource for submitting searches (d), again guided by the investigation issues. Returned documents were then individually read and coded for relevance to the issues (within a document management system) (c). Relevant documents were then used as a resource for constructing integrated representations (b) which structured extracted facts relevant to the investigation and allowed connections to be seen between them. A number of representations were used, but most attention was given to chronologies which recorded meetings and other significant events. Chronologies were created for each issue by different groups within the investigation team and these were combined into a single master chronology for review by senior team members.

Significantly, the investigation process was not linear. By interacting with documents and external representations the investigators gradually made more sense of the company's activities and, in addition to discussions with the client, this led them to refine the issues with new theories and questions. For example, the chronologies provided a narrative which supported the identification of unexplained or suspicious events and showed overall timescales from which periods of particular interest could be identified. As the issues evolved into more specific sub-issues so these propagated down to motivate new searches or the recovery of new documents.

### 4.2 Step 1b: Identify Transformations that Represent Major Overheads

As discussed above, optimising sensemaking involves maximizing the cost-to-gain ratio of component tasks. That is, increasing the amount of gain obtained for a given level of expended effort. As with any task, when people perform sensemaking they design their activities in such a way as to achieve an optimal trade-off between cost and gain given the tools at their disposal [Russell et al, 1993]. Changing the tools, however, can enable more effective trade-offs (e.g. as a mode of transport, cycling has an improved cost-to-gain ratio compared to walking).

In considering where to focus attention to make improvements, we could consider any part of the process in figure 1. However, the investigators consistently cited document review (c) as imposing the major overhead in terms of time and effort. They submitted a total of 200 searches, each of which returned hundreds or even thousands of documents. The results were presented as date-ordered document lists from which documents were selected in turn and their full-text reviewed. Over the course of the investigation 130,000 documents were reviewed in all. This represents a significant reduction on the document universe, but is nevertheless a very significant number of documents to have reviewed.

### 4.3 Step 2a: Understand Transformations in Terms of Users' Frames

An answer to this lies in understanding the representational forms that the investigators were aiming to construct. Our general principle is that by understanding a sensemaking transformation in terms of its source data and the integrating structures it creates, one is better able to enable this transformation and anticipate the evolving information needs that it tends to invoke.

The chronologies were created using Microsoft Excel spreadsheets with entries corresponding to a pre-defined event schema. Russell et al [16] refer to such instantiated schemas as *encodons*. The schema included the date and time of the event, a text account of it's important elements, a field for recording the people involved and the event's location, and a field for recording references to the supporting document(s). Where a document, such as an email, referred to a significant event, an entry would be raised. Events could, for example, represent a meeting between protagonists; the signing of a contract; or a protagonist travelling. A single document could potentially give rise to a number of records and a given record could be based on multiple documents.

Many of the chronologies the investigators created corresponded to events specific to particular business activities, such as those surrounding particular contracts. Even when the individual chronologies were combined in the master chronology this separation was maintained using metadata attached to each event.

In the following we elaborate two types of frame emerging from the study data (event frames and business activity frames) and the way in which processes of sensemaking were organized around them. What we see is frames at different levels of granularity invoked through discovery after which they act as foci for elaboration and validation.

**Event frames**

An event frame connects information (date, time, description, people) relating to a particular event. In developing the chronologies, the investigators reviewed documents (email messages in most cases) and drew inferences about events (connecting data to frame). An email might propose a meeting or it might discuss a meeting that had occurred in the past. It might provide details such as the time and location and who the participants were, and it might provide some evidence for what was discussed and the outcomes of the meeting.

A single email, however, would typically only provide partial and potentially inconclusive information about an event. Where a meeting was planned and discussed it may not have taken place or it may have been replaced by a telephone call. Hence, beliefs about events could be more or less speculative, at least initially. Inferring what had actually happened, as supported by the totality of the available evidence, required further investigation. As one participant reported:

> P4: [...] So you put an entry down for November 20th and then you'd start looking for documents which relate, which might give evidence that that happened, that it actually happened [...] and if it did happen who else was involved, who were they meeting, what were they doing, what were they saying to each other?

Given the discovery of an email about an apparently significant event, an investigator would then want to elaborate and validate it (in Klein et als' terms). Each reviewer was presented with a display of document titles from which full-text could be selected. A problem was that this display provided no cues as to the comparative value of a given document at any given time over-and-above their responsiveness to a particular search. This did not account for evolution in the user's thought processes as described above.

Given the design of the system they used, this would present them with a choice. One option would be to temporarily pause the exhaustive review, and attempt to find further messages containing references to the event (using new searches, for

example). Alternatively, they could record the event as a plausible conjecture in a chronology (e.g. 'possible meeting'...) and continue their review in the hope of finding further relevant information later, or that someone else would find it. Often the reviewers used this exhaustive strategy requiring them to maintain multiple, overlapping cognitive sub-threads of interest.

In terms of the data-frame model, this example (which was repeated many times throughout the investigation) this demonstrates a sensemaking process which begins with a message acting as an 'anchor' for an event frame. The occurrence of an event is a plausible interpretation or is at least suggested by the anchor. However, given knowledge that the investigator brings to the situation (about meetings and emails etc.), they also know that there may be more important information to discover and that this may show that the meeting did not occur at all.

The occurrence of the frame, then, motivates its own elaboration and validation (and potential questioning) as a plausible interpretation of the available data. Salient information may include location, a list of participants, motives, discussion topic, and outcomes. Information contained in the initial message may populate some of these and provide some level of guarantee about them. But the frame necessarily triggers further information requirements specific to the event. It extends beyond the given information, and so creates expectations which need to be tested and elaborated.

Once the frame is cued, the need for elaboration and validation gives rise to new information requirements and this gives new shape to the relative values of different documents. For the investigator in that situation, documents that also discuss the event become of greater value.

**Business Activity Frames**

A second kind of frame that emerged as being important in this study was a business activity frame, which connects information about a given business activity. This links work around a particular contract or negotiation and is made up of a sequence of separate but causally related event frames. In part, their structure depended upon the investigators' a priori knowledge of business processes augmented by specific modi operandi of the company as revealed during the investigation.

The determination that the investigators sought concerned potential malpractice in the business activities of a company. Naturally, the company had many business activities and part of the investigators' task early on was to identify activities of particular concern. The identification of an activity of concern was followed by the elaboration of events associated with it.

The investigators were aware that evidence for malpractice might ultimately arise in very few documents. Lawyers involved in regulatory investigation and litigations typically sift through thousands or even millions of documents in order find what may be just a few documents of significance to the questions of the case.

However, it was necessary for the investigators to develop a broader understanding first. The interpretation of an event, such as an email communication or a meeting, depended upon how it located within a wider context of activity. Hence, that context must be elaborated. People may meet, exchange information or even money, but what these mean in legal terms can only be determined in the light of a broader set of events. In this sense the events were indexical [22] with interpretation dependent on context. Having said that, the reverse is also true: understanding the context depends

on understanding a series of individual events. And so the investigators were tied into a hermeneutic loop. Interpretation of the parts depended upon interpretation of the whole and vice versa [18].

In addition to supporting interpretation, elaborating the business activity contexts also enabled the investigators to identify and focus on key periods of concern. By a process that acts in reverse to the interpretation of the meaning of an event, by elaborating broader business activity sequences, the investigators were able to identify periods when fraud could technically have occurred. These could then form areas for more intense investigation. Given the size of the information universe, focusing attention on these periods was particularly important:

> P5: we'd be thinking, well if we're right on this, this is a really important build up [...]. Or, we think money must have been sucked out of this business around this time. [...] And this is what we did, [Junior Partner] selected certain periods and posed certain questions in relation to those periods. And we would go back and interrogate the information further.
>
> P6: So some time-periods where it was absolutely critical to know... because you're following this through forensically trying to figure out what's going on... it's absolutely critical to know minute-by-minute the exact chain of events.

A business activity is a frame that is triggered by information indicating an identifiable business activity which may be considered suspicious or 'vulnerable'. The frame then acts as a focus and generates expectations about what further information might be found. Like the event frame, the business activity frame extends beyond its anchor to create expectations about information to discover.

In terms of the goal of elaborating a business activity frame, some documents had higher value than others:

> P4: You know a document may actually lead to five different entries on the chronology because for example it may be someone's email saying 'right I'm organising [contract]. These are key milestone dates [...] and on this day I'm planning on being in [overseas city].

Some documents made reference to events and some did not; those that did had higher value for elaborating the frame. Documents that discussed a series of dates were particularly useful. The system that the investigators used, however, did not provide any method for finding these documents beyond exhaustive reviewing.

### 4.4 Step 2b: Design to Reduce the Costs Involved in Developing Key Frames

We have used a frame-based approach to sensemaking as a method for elucidating the evolution of information needs of the investigators in our case study. In this section we consider how this translates into design requirements, focusing on finding relevant documents within a results list.

A results list offers the user a list of information objects from which to select items for inspection. By default, users in our case study were presented with linear, unstructured lists within which all documents were presented as equal. However, the need to elaborate or validate a frame changes the relative values of documents as the user progresses. Some documents become more important to find and others less so. Specifically, once an event frame is cued the user wants to find other documents about that event (as a nested sub-task), after which they will likely continue the search for other events. Consequently, continuity would be best served by two complementary organisation schemes at the interface; the first links documents in some default way (e.g. chronological), whilst the other links documents in virtue of

reference to common events. Ideally, users would be able to move easily between the two schemes as they interweave the discovery of events with their detailed exploration.

Given that a business activity frame incorporates multiple events, reducing the cost of elaborating and validating single event frames will also reduce the cost of elaborating the business activity frame to which they belong. However, we saw from the case study that some documents had higher value than others for elaborating such frames. In particular, documents containing references to many events (such as project milestones) could cue many related event frames. These helped in establishing broad frames of reference which allowed the investigators to focus in on specific periods in detail. Consequently, by drawing attention to the number of event references in a single document at the interface users would be better able to prioritise exploring those with potentially higher value for elaborating such a frame.

In terms of developing document results presentations, these needs could be addressed by providing structured results presentations or visualisations which indicate documents' chronological ordering, link documents according to common event references, and indicate the number of event references within each document. We anticipate that text processing techniques necessary to create such representations do not significantly extend beyond current capabilities; indeed progress has been made in the identification and normalization of temporal references in free texts [7]. Our aim, however, is to illustrate how understanding the frames that users work with in particular sensemaking tasks, and in particular how these are cued, elaborated and validated as information is encountered, can provide a basis for requirements for the design of more useful information systems.

## 5 Discussion

The approach we have proposed and illustrated for improving the efficiency of large-scale sensemaking tasks involves the identification of activities of high demand, and exploring these in terms of the sensemaker's typical cognitive frames involved in making resource transformations. By providing an analysis of related information needs and the way these develop, it provides a useful foundation for reasoning about the design of effective solutions using available technologies. By eliciting typical sensemaking activities and concepts, it is possible to identify design requirements that support them.

In relation to users' interactions with information systems, such an account has been lacking within both the Human Computer Interaction and Information Science literatures. In Information Science, in particular, it has long been recognised that users' information needs are broad and under-specified early on in a research task but become more specific and better defined later on [13]. One way of understanding the role of a frame, or (more generally) users' conceptual understanding of the domain of which sense is being made, is as a mechanism through which broad, under-specified information needs evolve to become needs that are more focussed. Consequently, a concept-based approach to understanding the evolution of users' domain models provides a basis for explaining this characteristic phenomenon.

We have also shown that frames can embed within each other. The business activity frames we described are constructed from multiple individual event frames. This extends beyond Klein et al's original model by revealing the stratification of sensemaking and the way in which addressing one frame can be a part of addressing another.

The applicability of the approach presented here depends upon the predictability of particular user concepts during sensemaking. Contexts of more variegated sensemaking may offer fewer opportunities. However, there is evidence that other domains offer the requisite predictability. For example, Russell et al's [16] case studies on course designers and laptop purchasing lend themselves to a concept-based analysis in terms of printer components and the attributes of laptop models respectively, and research on users' experiences with using information visualisations of research literatures shows that they tend to conceptualise this domain specifically in terms of papers, key ideas and people [6]. The extent to which sensemaking can be supported by explicitly representing such concepts in a way that facilitates user reasoning and information transformation remains a topic for future research.

To conclude, in order to design for sensemaking we need to understand it. Little attention has been given to the way people think about the domains they are trying to make sense of, or of how to provide targeted support for that sensemaking. In this paper, we have presented an approach to identifying the most costly elements of the sensemaking process. This involves generating a process description and reflecting on the challenges presented to the user by each step of that process and the possibilities of providing support for key steps. It also involves identifying the core concepts with which users are working and how they are structured. We have tested the approach with a case study of legal discovery processes: this paper therefore provides a domain-specific account of sensemaking in e-discovery as well as proposing some general principles about strategies for supporting sensemaking of large bodies of information.

**Acknowledgments**
We would like to thank Freshfields Bruckhaus Deringer for their kind help with this study. The work was funded under EPSRC grant EP/D056268.

## References

1. Attfield, S., Fegan, S., Blandford, A.E.: Idea Generation and Material Consolidation: Tool Use and Intermediate Artefacts in Journalistic Writing. Int. J. of Cog. Tech. and Work (Online First, Feb. 2008).
2. Bartlett, F.C. Remembering: A Study in Experimental and Social Psychology. Camb. Univ. Press, Cambridge, England (1932)
3. Blandford, A., Green, T.R.G., Furniss, D., Makri, S.: Evaluating System Utility and Conceptual Fit Using CASSM. Int. J. of Hum.–Comp. St. 66, 393--409 (2008)
4. Cockayne, A., Wright, P.C., Fields, B.: Supporting Interaction Strategies Through the Externalization of Strategy Concepts. In: Sasse, M.A., Johnson, C. (eds.) Proceedings INTERACT'99, pp. 582-588. IOP Press. (1999)

5. Dervin, B.: An Overview of Sense-making Research: Concepts, Methods and Results. Paper presented at the Annual Meeting of the Int. Communication Assoc. Dallas, TX (May 1983)
6. Faisal, S., Cairns, P., Blandford, A.E.: Developing User Requirements for Visualizations of Literature Knowledge Domains. In: Proc. of the 10th Int. Conf. on Information Visualisation (London, UK, July 5-7, 2006).
7. Gaizauskas, R., Wilks, Y.: Information Extraction: Beyond Document Retrieval. J. Doc., 54, 1, 70--105 (1998)
8. Glaser, B., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine, Chicago (1967)
9. Goffman, E.: Frame Analysis: An Essay on the Organizational Experience. Harper, NY (1974)
10. Johnson, J. C. 1990. Selecting Ethnographic Informants. Sage, Newbury Park, CA.
11. Klein, G. Sources of Power: How People Make Decisions. MIT Press, Cambridge, MA. (1998)
12. Klein, G., Phillips, J.K., Rall, E.L., Peluso, D.A.. A Data-frame Theory of Sensemaking. In: Expertise Out of Context: Proc. of the Sixth International Conf. on Naturalistic Decision Making (Pensacola Beach, Florida, May 15-17, 2003). Lawrence Erlbaum Assoc. Inc, US, 113-155. (2007)
13. Kuhlthau, C.C. A Principle of Uncertainty for Information Seeking. J. Doc. 49, 4, 39--55. (1993)
14. Pirolli, P., Card, S.: The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, In Proc. Int. Conf. on Intelligence Analysis (McLean, VA, May 2-3, 2005). https://analysis.mitre.org/proceedings/ (Accessed 11th Feb 2008).
15. Qu, Y, Furnas, G.W. Sources of Structure in Sensemaking. In: Proc. of the CHI '05 Conf. on Hum. Factors in Comp. Sys. (Extended Abstracts), (Portland, OR, USA), ACM Press, New York, NY, pp. 1989--1992. (2005)
16. Russell, D. M., Stefik, M. J., Pirolli, P. and Card, S. K.. The Cost Structure of Sensemaking. In Proc. of the INTERACT '93 and CHI '93 Conf. on Hum. Factors in Comp. Sys. (Amsterdam, The Netherlands), ACM Press, New York, NY, pp. 269--276. (1993)
17. Schank, R.C., Abelson, R.P. Scripts, Plans Goals and Understanding: An Inquiry into Human Knowledge Structures. Lawrence Erlbaum Assoc., Mahwah, NJ. (1977)
18. Schmidt, L.K. Understanding Hermeneutics. Acumen, Stocksfield, UK. (2006)
19. Spence, R. A Framework for Navigation. Int. J. of Hum.-Comp. St. 51, 919-945. (1999)
20. Starbuck, W.H., Milliken, F.J. Executives' Perceptual Filters: What They Notice and How they Make Sense. In: Hambrick (ed.) The Executive Effect: Concepts and Methods for Studying Top Managers. Greenwich, JAI Press, CT, 35--65. (1988)
21. Strauss, A., Corbin, J.: Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd ed. Sage, London (1998)
22. Suchman, L. Plans and Situated Actions: The problem of Human Machine Communication. Camb. Univ. Press, Cambridge, UK (1987)
23. Weick, K. Sensemaking in Organisations. Sage, London, England (1995)