



# Known and unknown requirements in healthcare

Alistair Sutcliffe<sup>3</sup> · Pete Sawyer<sup>2</sup> · Gemma Stringer<sup>4</sup> · Samuel Couth<sup>4</sup> · Laura J. E. Brown<sup>5</sup> · Ann Gledson<sup>3</sup> · Christopher Bull<sup>1</sup> · Paul Rayson<sup>1</sup> · John Keane<sup>3</sup> · Xiao-jun Zeng<sup>3</sup> · Iracema Leroi<sup>4</sup>

Received: 31 July 2017 / Accepted: 30 June 2018 / Published online: 30 July 2018  
© The Author(s) 2018

## Abstract

We report experience in requirements elicitation of domain knowledge from experts in clinical and cognitive neurosciences. The elicitation target was a causal model for early signs of dementia indicated by changes in user behaviour and errors apparent in logs of computer activity. A Delphi-style process consisting of workshops with experts followed by a questionnaire was adopted. The paper describes how the elicitation process had to be adapted to deal with problems encountered in terminology and limited consensus among the experts. In spite of the difficulties encountered, a partial causal model of user behavioural pathologies and errors was elicited. This informed requirements for configuring data- and text-mining tools to search for the specific data patterns. Lessons learned for elicitation from experts are presented, and the implications for requirements are discussed as “unknown unknowns”, as well as configuration requirements for directing data-/text-mining tools towards refining awareness requirements in healthcare applications.

**Keywords** Requirements elicitation · Domain knowledge · Experts · Causal models · Data mining · Medical informatics

## 1 Introduction

Healthcare systems are often large scale, complex, with uncertain knowledge<sup>1</sup>, and consequently pose considerable problems for requirements engineering (RE). Although co-design or participatory design with domain experts has been advocated as a means of dealing with these problems, evidence for the success of such approaches<sup>1</sup> is sparse. In this paper, we report experience adapting requirements analysis techniques to address the problems which emerged in a healthcare system. The objective of the SAMS (Software Architecture for Mental Health Self-management) project

was to detect early signs of dementia by analysis of recordings of computer activity and e-mail text [1, 2] to detect changes in behaviour that may indicate onset of pathology. If potentially significant changes are detected, the system should send an alert to the users urging them to contact their doctor for follow-up tests. The terms of reference for SAMS implied requirements for monitoring and interpreting computer logs of human–computer interaction and then connecting analysis of computer logs to clinical models of human behaviour which might indicate early signs of dementia.

The terms of reference suggested two initial approaches for requirements analysis. First was eliciting knowledge from clinical domain experts specifying behaviours which are diagnostic of dementia. This implied requirements in the form of a causal model elicited from experts in clinical and cognitive neuroscience that linked behavioural measures to diagnostic signs of dementia. Paper-based diagnostic tests [3, 4] provided the starting point for analysis that might embed similar knowledge in an expert system. The second approach was motivated by logging keyboard use, mouse actions and e-mails to enable interpretation of user behaviour and, more importantly, any changes in behaviour which might indicate early signs of dementia or progressive cognitive decline. This involved requirements

---

✉ Alistair Sutcliffe  
alistair.g.sutcliffe@manchester.ac.uk

<sup>1</sup> School of Computing and Communications, Lancaster University, Lancashire, UK

<sup>2</sup> School of Engineering and Applied Science, Aston University, Birmingham, UK

<sup>3</sup> School of Computer Science, University of Manchester, Manchester, UK

<sup>4</sup> Institute of Brain, Behaviour and Mental Health, University of Manchester, Manchester, UK

<sup>5</sup> School of Psychological Sciences, University of Manchester, Manchester, UK

<sup>1</sup> <http://www.health.org.uk/sites/health/files/ComplexAdaptiveSystems.pdf>.

for data mining: what sort of patterns and trends should be investigated by data-mining computer logs of user behaviour? Data-mining tools may be used in hypothesis-discovery mode [5], suggesting requirements questions such as “how can the patterns and trends discovered be interpreted in terms of their medical implications?”, or in data-driven mode when data and text-mining tools might discover aberrant patterns in logs of human–computer interaction that indicate possible medical problems. Synthesising these two approaches appeared at the onset of the project to offer good prospects for a solution. These questions revisit the “unknown unknowns” perspective in RE [6, 7] for two reasons. First, there were several partially known goals perceived by requirements analysts, in particular the need for hypotheses to direct data mining and the need for a diagnostic causal model. Second, there was uncertainty about the degree of “knownness” of domain knowledge held by the expert medical practitioners.

Elicitation techniques either in the field of classic requirements [8–10] or in knowledge engineering [11] have been extensively researched and reviewed, concluding that a variety of techniques (e.g. interviews, workshops, expert conversations, scenarios and prototypes) may be appropriate, but choice depends on matching techniques to the problem and stakeholders. The majority of studies conclude that possession of some domain knowledge by requirements/knowledge engineers can help to bridge the communications gap. In the knowledge framework perspective for RE [7], eliciting unknown expert knowledge is amenable to known RE techniques, even though the boundaries of such knowledge may be unknown. The monitoring requirements of the software developed in the SAMS project are an example of awareness requirements (or AwReqs) [12]. AwReqs are the means of detecting the success or failure of other requirements. However, such requirements have been considered in terms of short-term feedback with a relatively small number of variables. In contrast, our perspective is for longer-term monitoring of complex phenomena (i.e. people) to detect deviations in normal healthy behaviour and hence requirements to take remediating action (i.e. treatments). Following this perspective in SAMS, the ultimate goal is for AwReqs to monitor the state of the user’s health with feedback loops to encourage follow-up action if deviations from normal behaviour patterns are detected. AwReqs have tended to focus on monitoring a small number of system variables, and then adapting behaviour with rules and changing parameters in algorithms (e.g. [12]). In this paper, we investigate how the AwReqs concept scales up as the phenomena being monitored become more complex. Requirements for data-/text-mining tools and for monitoring software are design—rather than user—requirements [13] although the source of such knowledge, expert software

engineers, was known. Therefore, when the project was initiated, the sources and scope of the required knowledge were known even if the detail was unknown. No obvious unknown unknowns were apparent.

This paper reports the subsequent development of the requirements analysis problem posed by the SAMS project. We report our journey through requirements discovery and reflections on requirements elicitation in medical informatics, and elicitation of causal diagnostic models from experts. It is important to note that this requirements knowledge-based approach was a pragmatic decision as much as it was our favoured strategy; the large-scale data capture required by a machine-learning causal model induction strategy was not feasible within the resources available to the SAMS project. Although once the monitoring software has been developed, data collection can be (semi-) automated, there is a significant support cost. This is particularly acute for the clinical researchers who recruit participants, liaise with their care organisations, and compile the ground truth by performing periodic wide-ranging and time-intensive cognitive assessments of each participant. Scaling up the number of participants to the order of hundreds (rather than the tens available to SAMS) would have required many more than the two full-time clinical researchers available to SAMS. Moreover, the duration of the project would have had to be extended by at least 12 months to increase the longitudinal volume of data collected, and to increase the incidence of observable changes in participants’ cognitive health.

Even had these additional resources been available, relying on the use of unsupervised machine learning to diagnose dementia in a vast, complex problem space of individual variation and many confounding mental health conditions would have represented a risk to SAMS’ success. In this case, we would have sought to mitigate the risk by exploiting the clinical expertise that *was* available in just the kind of requirements knowledge-based approach that we adopted that forms the focus of this paper.

In [1], we reported requirements analysis experience focusing on end users whose health will be monitored by the SAMS software. In contrast, this paper reports the project-wide requirements analysis experience with domain experts, design and user requirements. In following sections, first we review the related work on requirements elicitation from experts and awareness requirements; this is followed by two sections which describe the elicitation techniques and results for domain expert knowledge. The subsequent section reports the analysis of design requirements for data and text mining. Section 4 presents the synthesis of requirements for the causal model for detection of cognitive and functional change due to dementia, and associated data mining, with lessons learned from our experience. The paper concludes with a discussion of our experience of requirements discovery in light of the known–unknown RE framework.

## 2 Related work

### 2.1 Requirements elicitation from domain experts

Requirements elicitation is a relatively mature area of RE [10], and the basic techniques (i.e. interviews, observation, scenarios, workshops, focus groups, protocols, prototypes, models) have been described in several RE books [14–18]. A model of elicitation technique selection [19] proposed matching techniques to the RE situations composed of facets of the domain, maturity of requirements, stakeholders and organisations. In a meta-review of elicitation papers [10], no advantage was found for other techniques over semi-structured interviews; the use of representations (models, prototypes) did not appear to help, although the empirical evidence in the reviewed papers was limited. Coughlan and Macredie [9] compared elicitation techniques in soft systems methodology, joint applications development (JAD) and participatory design against a framework of user designer roles, communication activities and techniques including interviews, prototyping, cognitive (protocols), contextual (ethnography), group workshops and model-driven representations, concluding that collaborative, dialogue-based methods which included workshops were more effective.

The knowledge elicitation problem dates back at least to attempts to codify expert knowledge as rules in knowledge-based systems. Synthesising the rules from the knowledge proved to be hard, but eliciting the knowledge was at least as problematic, with many obstacles to effective acquisition of knowledge [20, 21]. Kelly's Personal Construct Theory (PCT) [22] underpins a number of techniques that have been successfully used to elicit requirements knowledge from experts in a range of contexts that include expert systems [23] and RE [20]. PCT maintains that people represent the world through a framework of elements and constructs that determine how they perceive (construe) the world. A construct is represented as the poles of a range of values. When experts assign a value somewhere between the two poles of a construct, they describe how they perceive the associated element. For example, a cognitive psychologist may hold *executive function* to be an element within the world of their specialist domain, and construe executive dysfunction to lie somewhere on a spectrum of diagnostic significance that ranges from *not indicative of dementia* to *strongly indicative of dementia*.

Attempts to systematise the use of PCT have led to a number of widely applied techniques. Chief among these is the repertory grid [24, 25], which represents the elements and constructs in a matrix. By building up a repertory grid that characterises each element as a value somewhere

on the scale of each of the corresponding constructs, an expert's viewpoint [26] of the domain is built up. Where experts find it difficult or are unwilling to commit to scalar values for constructs, it can be hard for the analyst to make sense of the information. Card sorting [27] is often used as a complement to the repertory grid as a way for the analyst to better understand the elicited knowledge, by asking experts to classify and order concepts within their world against a set of analyst-determined criteria. For example, a neuroscientist might be asked to sort a set of cards containing the names of a set of cognitive domains (e.g. *executive function*, *memory*, *visio-perceptual functioning*, *attention* and *language*) several times, each according to a number of criteria such as indicative of dementia, ease of assessment and underpinning a specific task such as dragging a file from one folder to another on the desktop, and so on. In a SAMS-like system, this might then help the analyst prioritise which routine computer tasks to monitor and collect data from. Although there are many variations of card sorts, card sorting is generally considered easy for stakeholders to engage with [28] and it has been successfully used to elicit information from experts in a number of domains.

As techniques for eliciting expert knowledge about a problem domain [20, 29], the repertory grid and card sorting provide the means for experts to articulate their world using their own terminology, and provide the domain-ignorant analyst [30] with a way to probe the expert while minimising the projection on to the problem domain of the analyst's own preconceptions.

In many domains, expert knowledge is incomplete and dispersed amongst the expert community. In these circumstances, it is common for there to be imperfect consensus among the members of the expert community. While card sorting may be used for knowledge elicitation from groups, the Delphi method [31] was devised specifically for knowledge elicitation from a group of experts. It was first used to elicit prediction about the future state of the world. Since then, it has been applied in many problem domains, including RE [32]. It is an iterative method in which each expert is asked a question (e.g. what is the importance of executive dysfunction as an indicator of dementia?), then provided with the others' anonymised opinions and invited to revise their response and to add a rationale. After a number of iterations, either consensus is reached or the moderator is able to synthesise a reasonable position. Hybrids of the Delphi and card-sorting methods have also been developed and applied in, for example, user-centred design [33].

A persistent problem with knowledge elicitation is that not all the knowledge held by an expert is easily made explicit. The experts may deliberately withhold their knowledge for some personal reason. They may fail to reveal their knowledge because they were asked the wrong question

or did not appreciate the value of their knowledge. Card sorting can help these issues surface. However, where the knowledge is genuinely tacitly held [34], i.e. *available* to the expert but not easily *expressible* by them [35], none of the above techniques is guaranteed to elicit the knowledge in an explicit form that is usable. In a study of communities of scientists, [36] discovered that much professional knowledge is of this type, arguing that techniques that robustly handle tacit knowledge are needed, particularly for the elicitation of knowledge from domain experts. Unarticulated knowledge is often made manifest by experts' actions, and skilled ethnographers can identify when this occurs and probe the expert to better understand why they acted in the way they did. More recent work [37] that builds upon that of [35] posits that ambiguities in elicited requirements information may offer cues to where knowledge is tacit. Cleland-Huang [38] has demonstrated that text-mining techniques can be applied to elicit expert knowledge where documents exist (e.g. topic and lexical analysis), while the KMoS-RE method [39] proposes discourse analysis combined with ontologies and extended lexicons as a means of eliciting tacit, expert knowledge.

## 2.2 Awareness requirements

The origins of awareness requirements in RE can be traced back to the proposal of [40] for run-time monitors that detected changes in system operation to determine if a system design was conforming to its requirements. Run-time monitors of system behaviour have since been elaborated by several authors [e.g. 41] to create systems which monitor their own behaviour and then adapt to conform to original (usually performance, NFR) requirements when run-time violations have been detected, e.g. ReqMon [41] and Relax [42]. This concept has a wider following in the SaaS (Self Aware, Adaptive Systems) community [43].

In RE, monitors were developed into awareness requirements (or AwReqs) [12]. The idea of requirements awareness is an evolution of Smith's [44] original notion of computational reflection, but adapted to the idea that a software system should be subject to introspection about its requirements. That is, at run-time such systems should be aware of the extent to which their requirements are being satisfied: "requirements should be run-time entities that can be reasoned over in order to understand the extent to which they are being satisfied and to support adaptation decisions that can take advantage of the systems' self-adaptive machinery" [6].

AwReqs, therefore, act as a means of detecting the success or failure of other requirements and can refer to goals, tasks, quality constraints and domain assumptions. These are mapped to feedback loops, providing monitoring capabilities that can be used to determine satisfaction levels for goals

and system requirements. AwReqs have tended to focus on monitoring a small number of system variables and then adapting behaviour with rules and changing parameters in algorithms, for example [12]. In this paper, we investigate how the AwReqs concept scales up as the phenomena being monitored become more complex. Our focus is on systems where either complexity is acknowledged ab initio, and the high-level goal is to analyse data so a solution can be discovered; or on systems where complexity and the inability to decompose high-level goals into specific requirements emerge during the RE process. Both classes of system fit into the concept of partially known unknowns [30]. Complex intelligent systems also fall into this category, where it is impossible to specify a complex set of expert system rules at design time; instead, machine learning is implied.

An early implementation of this idea was the ReqMon toolkit [41], which provided a range of adaptable monitors and interpreters that can track system behaviour against specific goals. ReqMon has been applied to several domains in e-commerce, demonstrating its capability as a generic, adaptable architecture. The AI planner solver [45], based on the Tropos architecture and using BDI (Belief–Decision–Intention) agent-based reasoning tools, allows goal compliance to be tracked in a monitor-diagnose-reconcile (goal check), compensate (adapt) cycle. Evolution Requirements (Evo-Reqs) [12] extend requirements reflection to include the monitoring of social phenomena, such as human compliance with procedures. However, Evo-Reqs' interpreters were based on ECA (Event–Condition–Action) rules, which [12] acknowledge presents a maintenance problem as systems scale up. In addition to requirements evolution, many other SaaS tools and systems have been implemented (see reviews by [43, 46]); however, none to our knowledge has included components oriented to data-centric analysis (i.e. data-/text mining) or user interfaces for monitoring and adaptation/feedback.

## 3 Domain expert requirements: elicitation process

An initial literature review and analysis of the SAMS domain [2] revealed the extent of the requirements unknowns. While clinical assessment methods for dementia exist, these are based on explicit, intrusive tests, usually administered by a trained questioner to probe cognitive functions such as memory, attention, executive function and language. These tests might include remembering a sequence of numbers, a name and address, naming objects in images, drawing familiar objects. Several different cognitive test batteries are used in clinical practice, e.g. ACE [3] and MoCA [4]. Most probe the same key cognitive functions. Unfortunately most of the tests are based on observation of pathologies

in patients with dementia or its precursor, mild cognitive impairment (MCI). The rationale linking test elements to functional impairments, cognitive abilities and observable brain pathology on neuro-imaging (i.e. magnetic resonance imaging, MRI) scans is less than explicit in most papers [47]. Hence no consensus causal model appeared to exist in the clinical literature. Causal modelling is also complicated since dementia is a heterogeneous condition with different causalities, including (e.g.) fronto-temporal dementia, vascular dementia and Alzheimer's disease which account for around 70% of all dementias. In spite of these limitations, it was possible to prepare a list of candidate "cognitive indicators" as elicitation probes for a causal model. However, we realised that the experts would be relying on their experience using existing paper-based tests as well as clinical observations. A prime objective for the elicitation was to establish associations from the cognitive indicators to the onset of cognitive impairment and to possible explanations for observed user-behaviour patterns and errors. If we could establish these associations as a causal model, this might be implemented as a data-mining expert system for automatically inferring the likelihood of dementia onset from users' behaviour.

The literature search produced few studies analysing computer activities for signs of dementia, with the exception of [48] who compared analysis of simple frequency and duration data on keystrokes and mouse movements, between healthy users and people who already had a clinical diagnosis of dementia. Differences in computer activities were apparent between healthy and MCI user cohorts.

In [48], the data compared the performance of different user cohorts in a laboratory or clinical setting as they performed discrete tasks. The goal of SAMS is different; to detect cognitive decline in individuals, rather than comparing the cognitive abilities of different cohorts. SAMS data is therefore collected longitudinally over periods of the order of months, with SAMS operating as a passive monitoring system as users perform their normal, routine tasks. To do this, SAMS collects complex records of computer use, identifying not only keyboard actions but also user behaviour by analysis of mouse pointing and selection in terms of computer GUI operations and application commands (e.g. word processing, read, reply, send e-mail). This approach benefits ecological validity and allows the emergence of behavioural changes over time to be tracked. However, the conditions in which the data is collected cannot be controlled so the data is noisy. This revealed a requirements gap of unknowns in terms of patterns and trends in computer use and how these were linked to cognitive indicators that might indicate onset of dementia.

Two workshops were held in an attempt to bridge the requirements gap. Workshops provided the best

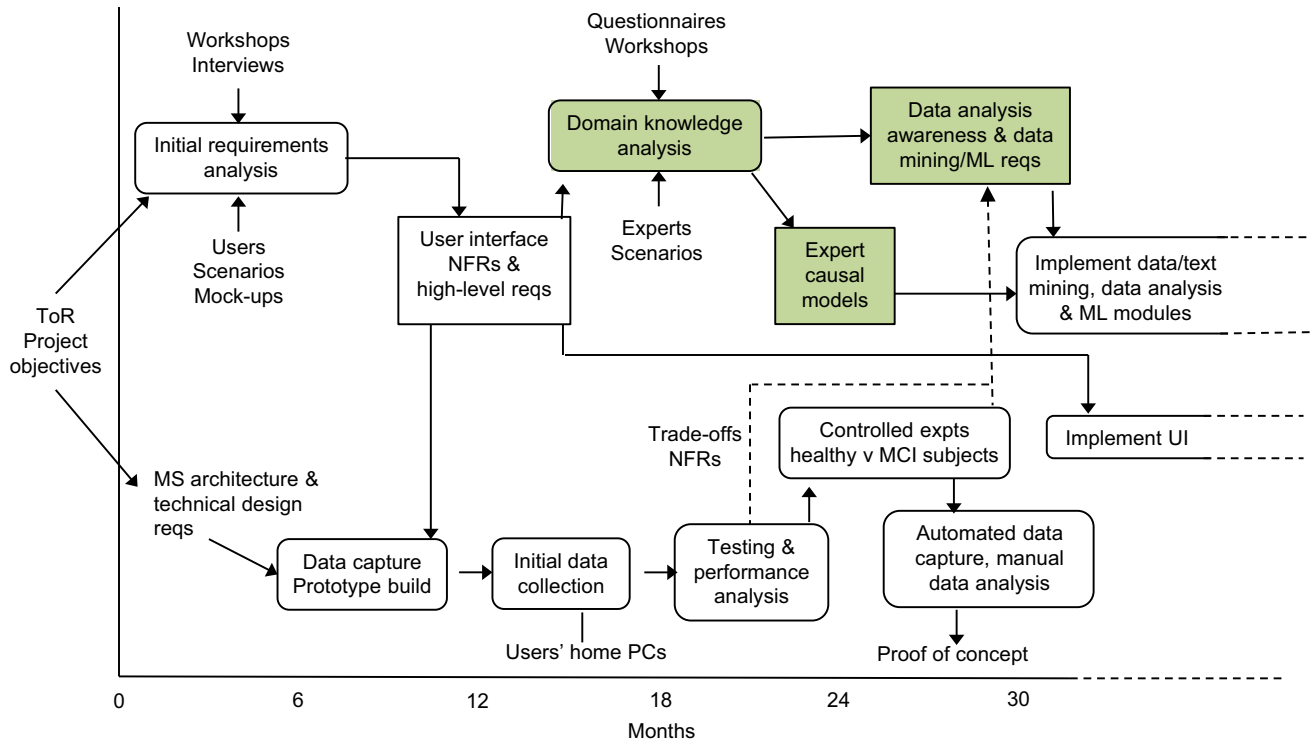
opportunity to mix focus-group conversations, problem tasks and scenario-based requirements elicitation, as well as facilitating expert conversations [49, 50]. Ethnography was not practical since generation of diagnostic knowledge is rarely observable, and resources precluded an in-depth discourse analysis. The process consisted of two workshops, which included card-sorting and PCT techniques, followed by a survey sent to the workshop participants. The rationale for the elicitation process was to use the workshops to elicit a tractable set of requirements in the form of causal diagnostic indicators. These could then be fed back to the experts for prioritisation and importance ranking, thereby establishing a consensus model among the experts following a Delphi-style analysis [51, 52]. Choice of the approach was motivated by successful use of Delphi-style elicitation techniques in healthcare domains [51, 53]. We chose a workshop-conversation-based approach [54] because this was an effective means of eliciting expert knowledge, which was likely to be incomplete and inconsistent between experts, in contrast to more formal KE elicitation via protocols and ontologies. We intended to capture both text from e-mails (which proved to be unsuccessful see Sect. 6.2) and quantitative data from computer interaction logs. The literature survey produced several guidelines for analysis of possible signs of dementia from text; hence, the workshops concentrated on quantitative data where no existing guidelines had been found.

The overall requirements process is illustrated in Fig. 1. The requirements activity in the shaded components form the subject matter of this paper. The initial requirements analysis was reported in [1] and further detail on the project approach is given in [2].

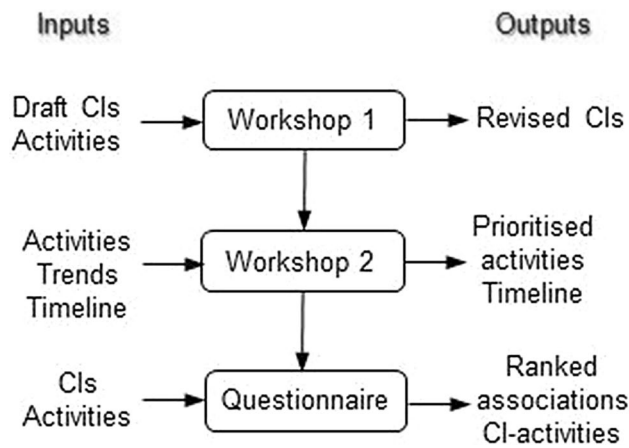
The rest of Sects. 3 and 4 reports the domain knowledge requirements analysis (shaded component), leading to the data analysis requirements and causal model described in Sect. 5. The contribution of testing and performance analysis and the controlled experiment to refining requirements is described in Sect. 6. The data-capture prototype is described briefly in Sect. 6.1 with the subsequent experimental analyses in Sect. 6.2.

### 3.1 Workshop materials and participants

The two workshops were held with domain experts. The first half-day event took place in December 2014, with six experts and four participants from the SAMS project. The experts (four female, two male) were three academic clinical researchers, one academic cognitive psychologist, and two clinical practitioner researchers in dementia. Four project members fulfilled facilitator and scribe roles. The process is summarised in Fig. 2.



**Fig. 1** Overview of activities in the timeline of the SAMS project. Oval boxes are processes, rectangles represent outputs. Dotted lines on the implementation components denote activities continuing beyond the scope of this paper



**Fig. 2** Overview of the modified Delphi process for knowledge elicitation. CIs=cognitive indicators or symptoms of possible dementia

### 3.1.1 Workshop 1

The goal of this workshop was to establish an initial causal model linking computer activities and errors to cognitive indications of dementia. The format of the first workshop was:

1. *Introduction and technical briefing* This included the aims of the SAMS project, description of the data recorded in layers (from keyboard/mouse at the lowest level to semantically rich application-level events at the highest level) with fictitious examples of collated data as graphs and tables, and definitions of cognitive indicators (CIs). The CIs described the broad definitions of sensory memory, short-term memory, long-term memory (6 items), memory recall, recognition familiarity, attention (6 items), executive function (7 items), language (10 items, including syntax, comprehension, reading and writing), perception (5 items), construction performance (2 items) and general mental processing speed.
2. *Consensus on cognitive terms* Experts were asked to review the CIs and suggest amendments. The amendments were then discussed within the group until consensus on the approximate terms and definitions was found.
3. *Linking CIs to computer activities* Experts were divided into two groups with a project member as facilitator/scribe. They were given descriptions of computer activities organised in three sections: general activities, e-mail and word processing/diary entry. A sample of the activities was illustrated with screenshot videos to demonstrate normal computer operations as well as errors. General activities covered login, GUI operations such

**Table 1** Excerpt of the computer activities: analysis worksheet. Facilitator version with detailed activity descriptions in column 2; the participant version, column 3, was left blank

Activity	Analysis	Cognitive domains
Logging on	Intervals (time between keystrokes) Distributions Error rate Initiated not completed (start typing password but do not finish) Action failure (wrong password) Action failure corrected – (wrong password then right password = successful correction, wrong then wrong = unsuccessful correction) Total number of attempts	

**Table 2** Excerpt of the cognitive indicators with detailed definitions in column 2

Cognitive Indicators—framework of terms memory	
Sensory memory	The short-lived retention of sensory information, measurable in milliseconds to seconds, as when we recover what was said to us a moment before we were paying close attention to the speaker
Short-term/working memory	Transient representations of information held in mind (and/or actively processed) for very short periods of time (i.e. seconds). Representations may be related to information that has just been activated from long-term memory, or to something recently experienced. It can be tested with tasks such as recalling digit spans backwards, or re-sequencing letter-number strings

as opening, minimising and closing windows, the use of scroll bars and drag–drop operations. E-mail activities were reading, replying and deleting e-mails, while word processing listed typical editing and formatting operations. Each activity (see Table 1) was accompanied with space for description of measures and possible errors in column 2, while column 3 provided space for the experts to enter the CIs associated with the activity, measure/error. Draft definitions of cognitive indicators (CIs, see Table 2) were also handed out. The experts worked through Sect. 1, then after a short break the group composition was changed for Sect. 2 and this was repeated for Sect. 3.

4. *Summary wrap-up session* in which the workshop outcomes were summarised by the lead facilitator.

The scribes made notes on their copies of the activities-analysis worksheets during the group sessions. Flipcharts were provided for each group to make notes.

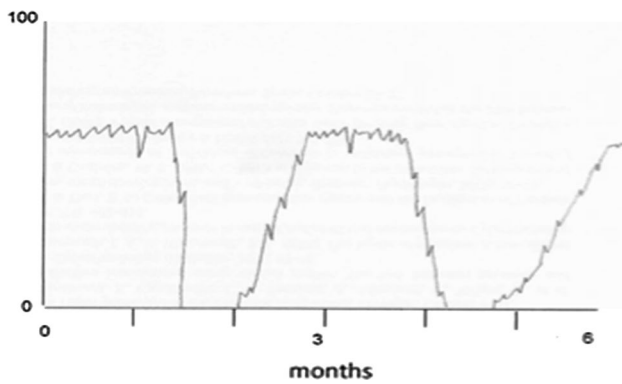
### 3.1.2 Workshop 2

The goal of this workshop was to consolidate the partial causal model, which had been elicited in the first workshop. The second half-day workshop, held in February 2015, was composed of five of the experts who had participated in the first workshop, and the same four project members. This workshop focused on changes in computer activities over time as well as consolidating expert opinion on possible

causal indications of activities on CIs and categorising indications into MCI, mild, moderate or severe dementia.

The workshop procedure was:

1. *Introduction and task briefing*
2. *Card-sorting computer activities* according to indications of dementia severity. Experts were divided into two groups which were changed between the two card sets, each with a project member facilitator/scribe. They were given descriptions of computer activities on cards organised in eight sections: login, opening a familiar program, opening a Word document, scrolling, dragging a file into a folder, opening/reading e-mail, replying to e-mail and cut/paste editing. Each activity card (total 68, presented in two sets) described an error or potential abnormality in the recorded data, e.g. slower, incomplete action, and a sample was illustrated with screen-capture videos. Blank cards were available for the experts' own suggestions. The experts were invited to sort the cards into a timeline progression of four dementia-severity categories, and place the cards into a hierarchical order of importance for indicators within each category. They were then asked to arrive at a consensus categorisation within each group.
3. *Trends* were presented as graphs for fictional individual users. Each graph was a possible pattern in activity frequencies, durations and errors, as well as sequences of normal and abnormal operations (see Fig. 3). The graphs presented fictitious scenarios of the data we expected to capture over time, and were therefore important for



**Fig. 3** Example of a trend scenario card. This card aims to stimulate discussion about the activity gaps being explained as an impairment or the user's context, e.g. away on holiday, sporadic use, etc

specifying data-mining pattern analysis requirements.

The experts were asked to sort the cards as before and annotate them with CIs or other reflections on possible reasons for the observed data.

#### 4. Workshop summary and discussion of future plans.

The results of the workshop were analysed and used to construct a questionnaire which was sent to the six experts who had participated in the first workshop.

##### 3.1.3 Questionnaire

The questionnaire design was based on the CIs from the first workshop and the activities from the second, with the aim of creating a consensus on the associations between computer activities and CIs. The CIs that were considered to be more important by the experts were selected with the computer activities that were ranked as being more salient indications of dementia. The questionnaire asked the experts to indicate the association between the CIs and computer activities on a 3-point scale. The experts were asked to complete the questionnaire independently in their own time. The questionnaire matrix was organised in seven activity sections describing errors or abnormal operations (22 items) which were to be cross-referenced to the CIs as before in six sections (memory, attention, executive function, etc.), see "Appendix" Table 3. The respondents were asked to indicate the extent to which abnormal operation could be an indicator for a given CI, graded as either none (blank), some indication (one tick) or strong indication (two ticks).

## 4 Domain expert requirements: results

The first workshop produced a consensus list of CIs, e.g. short-term memory, long-term memory, memory recall. Some of the CIs on the original list were revised, and new

CIs were proposed by the experts. No cross-referencing of computer activities to CIs was achieved because no consensus about the activities emerged within the allocated time. The second workshop produced a list of activities (top 5–10) for each dementia-severity category, e.g. errors such as attempting to paste text before select/cut, or being unable to start a task.

The trend-graph sorting task in the second workshop indicated that increasing errors were the most important indicator of dementia, including measures of frequency and diversity errors. Changes in activity frequency over time were the second important indicator, acknowledging that the noise (user mood, interruptions, changes in task, holidays) in the signal (dementia-related change) had to be accounted for. New signal patterns were suggested, such as a frequency decline followed by failure to return to previous levels. This distinguished between gaps in recordings, followed by return to previous levels, which may be due to a holiday. Decline in frequency and increase in duration might indicate onset of MCI, but these measures would have to be compared with normal age-related decline.

The questionnaire produced a ranked list of activities marked by frequency associated with CIs, see "Appendix" Table 3 and further details in [55]. All the respondents agreed that four computer-use behaviours were strongly indicative of cognitive impairment across a range of CIs: executive function, language, memory and visuospatial perception. For example, all the experts indicated that "Opening a series of different incorrect folders before opening the correct document in the correct folder" is strongly indicative of impairment in memory and, more specifically, recall.

There was a partial consensus for other user behaviours that might indicate some level of impairment in any of the cognitive domains. However, a majority (75% agreement) was found for a total of 12 user behaviours, split across CIs for attention, executive function, language, memory and visuospatial perception domains, which did indicate a possible progression towards MCI; e.g. "Cuts the text, but does not paste" may offer some indication that there is impairment to sustained attention and vigilance.

## 5 Results summary: causal model and requirements

No complete causal model for dementia analysis was produced, and only partial linking of computer activities with CIs was achieved in the questionnaire and second workshop. It became clear that no complete specification of a causal model linking behaviour patterns to CIs and tentative diagnosis of mild cognitive impairment (MCI) or dementia could be specified. Medical experts did not



possess the necessary knowledge, although a partial causal model was elicited with prioritisation of CIs which were estimated to be more important indications of MCI/dementia. Consequently there was a gap in requirements specifying how patterns of user behaviour should be linked to CIs with no immediate means of eliciting this knowledge. In spite of this, the elicitation process did provide useful indications about which CIs and computer activities were more important indicators of dementia in four stages of disease progression. In particular, executive function, which controls planning and decision making, might be an important CI since this could be detected by language text measures as well as errors, and idiosyncratic user behaviours. Memory and attention were also important CIs; however, generally, the mapping between user behaviour and errors to CIs was many-to-many rather than to a specific indication. Conversations with the experts produced emergent findings not present in the briefing materials, such as learning difficulties being an important indicator and how these might be manifest in errors when users encounter unexpected pop-up alerts, and unfamiliar applications or websites. A partial causal model was produced and this informed requirements for data mining. The general conclusion was that many behaviour and error measures needed to be combined to give a reasonable indication of potential pathology. CIs were more helpful for reflecting on errors and possible reasons for observed behaviour than forming components in a causal analytic model.

Few requirements were specified in the classic sense of a list of functional requirements. Instead, high-level requirements goals were the outcome, as well as a partial causal model. The goals became requirements to drive the data-/text-mining strategy, i.e.

- (i) Look for changes in activity patterns over time within each individual.
- (ii) Establish a baseline of variation for each individual, then use learning algorithms to detect changes against the baseline.
- (iii) Combine many user behaviour variables to establish the baseline and detect significant changes.
- (iv) Prioritise certain user behaviours over others, such as errors and incomplete sequences of actions.
- (v) Benchmark individual change against known statistics for age-related change, and clinical tests of activity in people with MCI/dementia.
- (vi) Create hypothesis-directed pattern recognisers to detect problems motivated by the prioritised CIs, e.g. error classifiers in behaviour sequence analysis, excessive switching between windows. Both indicate executive dysfunction. Repeated search for the same icon/user command may also indicate a recall memory problem.

Requirements (i) to (iv) suggested hypothesis-discovery data mining to search for any unusual data patterns and the subsequent use of CIs to decide which represent important indications or not, while requirements (v) and (vi) could be used to specify hypothesis-directed data mining. These were refined into requirements for classifiers and interpreters to search for specific patterns in the data motivated by prioritised CIs.

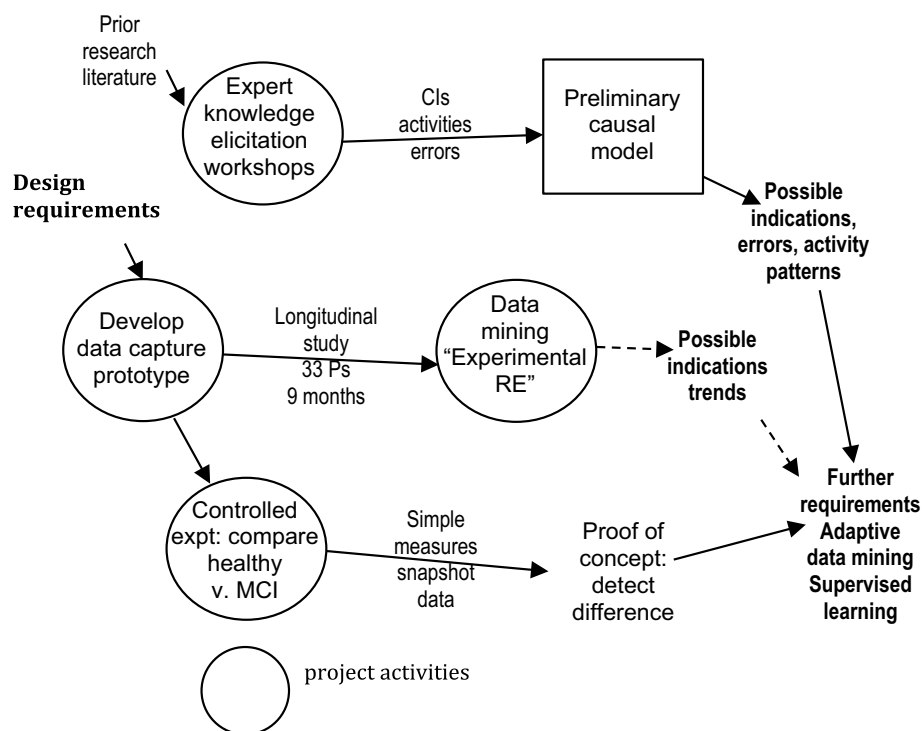
## 6 Design requirements: data and text mining

Expertise for these design requirements lay within the project team which contained expert researchers in data and text mining, as well as experienced software engineers familiar with problems of logging data from keyboard and mouse devices.

### 6.1 Data-capture requirements

Discussions within the design team proposed a system architecture composed of monitors, data-cleaning aggregation components which then fed standard data- and text-mining tools. The monitors collect data from the user's mouse and keyboard actions, then interpreters make sense of low-level interactions in terms of user actions in Windows, website browsing, and the use of applications such as Word and Outlook e-mail. Requirements were elaborated, and subsequently implemented, for the data logger to record user activities at three levels: keyboard and mouse, operating system (e.g. desktop activities) and application level. All Windows events were deemed potentially useful for detecting the CIs of dementia, with the view to further analysis to determine those most pertinent. Activities were captured as a list of time-stamped events at the mouse/keyboard level using an imported Microsoft.NET library. At the operating system level, native C#.NET libraries were used to detect file system events (files changed, created and renamed) and changes to the clipboard. Microsoft UI Automation events recorded operating system actions such as opening/closing/minimising/maximising windows, changes in focus, menus opened/closed and elements selected by the user. Experimental analysis with working prototypes under a range of data file loads, informed trade-off decisions for non-functional performance requirements; for example, "structure change" events which appeared to be the richest source of information for monitoring user activities had to be ignored as listening for them resulted in response time problems and missing logging of other events. At the application level, the Office Primary Interop Assemblies and the Internet Explorer automation object were used to detect events from Microsoft Word, Outlook and Internet Explorer, the three applications

**Fig. 4** Project activities and the emergence of requirements (in Bold)



which were selected for the prototype based on technical feasibility and previous requirements studies [1, 2] as being relevant for monitoring activities of older adult users.

## 6.2 Data and text mining

Detecting change in user behaviour implied awareness requirements for data mining and how to interpret changes in data patterns to infer possible diagnostic signs of early dementia. Awareness requirements evolved into specification of data- and text-mining strategies, informed by the domain expert knowledge reported above, to discover potentially significant patterns in the data. The relationship between project activities and different types of requirements is illustrated in Fig. 4. The different types of requirements are listed in “Appendix B”. The knowledge elicitation workshops described in Sects. 3 and 4 produced the preliminary causal model explained in Sect. 5. Design requirements informed development of the data-capture prototype (Sect. 6.1) which was then used in two studies: (a) a preliminary, longitudinal study of senior participants using their own computers over 9 months with a cohort of 20 females, 13 males, mean age = 72.9 years, MCI = 15, SCI = 18 (see below) and (b) a controlled experiment with set tasks, duration 60 min, comparing 24 healthy participants (mean age 71 years, 14 female) with 20 participants (mean age 75.6 years, 6 female) who had an existing diagnosis of dementia or its precursor, MCI. The controlled experiment was intended as a proof of concept that differences in simple data measures (e.g. event

frequencies, durations, keystrokes, mouse moves) could be detected between healthy and MCI subjects, which was the case [56]. The longitudinal study addressed the more complex problem of detecting changes over time which might indicate early stages of dementia among a group of healthy participants, and indications of subjective cognitive impairment (SCI), i.e. participants with no diagnosis, but who had reported concerns about memory loss.

Analysis methods for the complex longitudinal data, where user goals were unknown, started with an experimental rather than a requirements-driven approach. These include variation and sequence analyses to compare sets of contiguous events that have been divided into either uniform time segments (e.g. ten-minute segments) or processed using a sliding windows approach. Text-mining requirements were informed by a richer literature on dementia research, which provided indications about metrics of text verbosity, diversity of vocabulary, and syntactic patterns that indicate onset of dementia [57, 58]. Unfortunately text capture from e-mail and weekly diaries produced only small volumes, which invalidated application of text-mining tools; the participants simply did not write enough. A variety of data-mining techniques [59] were applied on event totals and for sequence mining, e.g. Markov sequences, association rule analysis, cluster analysis and eigen vector techniques to provide metrics to characterise event transition matrices. No consistent patterns were discovered as it became apparent that we faced a double unknown problem. First, clinical cognitive tests on the volunteer subjects taken in parallel

with the computer logging produced few reliable differences between individuals and no clear indications of behaviour trends associated with dementia. Secondly, the variation in the recorded behaviour both between individuals and within one individual over time was considerable and showed no obvious patterns. Consequently there was a gap in requirements specifying how patterns of user behaviour should be linked to CIs with no immediate means of eliciting this knowledge.

The controlled experiment comparing groups of healthy users with users with a known diagnosis of early dementia [56] demonstrated that the recorded logs of behaviour were significantly different between the two groups, although it did not produce much insight into requirements for interpreting changes in user behaviour over time, or how these might be linked to clinically significant indications of possible dementia. However, the experiment did illustrate the need to investigate variations between individual users and variations in behaviours within a person over time. A new high-level goal emerged to distinguish the signal (behaviour change indicative of early dementia) from the noise (changes due to mood, change in tasks, concentration, etc.). This implied new requirements to characterise “normal variation” that might be caused by multiple environment influences, e.g. users’ tasks, computer-use habits, mood, fatigue, and other unpredictable factors such as interruptions, change in websites, and other software. Since the project terms of reference for non-intrusive monitoring rendered knowledge of these factors as unknowns, characterising normal variation became an important emergent requirement. Normal “noise” variation in user behaviour patterns which reflect change in mood, task and other contextual factors, had to be distinguished from the signature changes that indicate possible onset of dementia. This prompted investigation of statistical analysis techniques to evaluate variation in many non-orthogonal variables associated with individuals and changes in many variables across a population and across time. Possible solutions to this emergent problem are multi-level modelling and variance analysis [60–62]. Had the volume of available data been greater, machine-learning techniques might have been used to tease apart intra-individual variation from significant change. The relatively small number of participants ( $N=33$ ) generating data over only 9 months made this approach infeasible for SAMS; however, application of ML by a project with access to the necessary resources may produce significant results.

In response to the emergence of these unknown problems in the requirements process, a research strategy evolved to integrate two complementary threads of further requirements activity:

- A bottom-up approach to assess a baseline of normal variation and then discover patterns in the data and then investigate their implications.
- A top-down approach based on the knowledge elicitation from medical experts to specify pattern interpreters for cognitive indicators of dementia.

Data-driven data mining is suitable when little previous knowledge exists, apart from general heuristics to guide the search. The hypothesis-directed mode is used when more specific knowledge is present to guide the search for particular data patterns. The requirements ranged from parameters for configuring data-mining tools, to heuristics to guide analysis of data-mining results, and more detailed functional requirements where pattern analysers were specified for important cognitive indicators. The requirements were grouped into the following categories, which are illustrated with some examples:

1. Data-mining tool configuration

Analyse variation within individuals to establish trends by using months 1–3 as training data for learning algorithms with subsequent months as test data.

Configure event-sequence mining tools to analyse frequent event-sequences which might represent normal user behaviour and less frequent, exceptional sequences that could indicate errors.

These requirements directed tool configuration and data-mining strategy. As data analysis progresses these requirements will change in light of insights gained from the first iteration of investigation.

2. Data analysis heuristics

In trend data look for decreasing frequencies or increasing durations with temporal irregularities such as rapid changes in frequencies.

After-activity gaps investigate frequencies before and after as well as pattern regularity.

Search and analysis heuristics will also evolve as the investigation proceeds. Serendipitous discovery of data patterns may improve our awareness of what to look for, and further consultation with experts may produce new heuristics.

3. Hypothesis-directed classifiers, with CI targets informed by the results reported in section IV.

Incomplete drag sequences, when no final action is found, such as delete document or file document in folder (executive dysfunction/attention)

Several sequential undo–redo actions (executive dysfunction)

Repeated text in same e-mail message (attention, executive dysfunction)

Incorrect or inappropriate words in message (semantic error).

The last category of requirements was the most concrete outcome of knowledge elicitation from clinical experts.

The requirements in group (i) specified software data-mining tool controls; group (ii) became guidelines for human data analysis; and group (iii) was refined into functional requirements for bespoke software and tool configuration. The software that monitored the user and logged their data was installed on the home computers of 33 elderly subjects. This data was analysed using the data-mining requirements outlined in this paper, leading to design of a causal analysis component. This analysis component is based on requirements elicited from experts, augmented by findings from a first iteration of data analysis. Future challenges are to extend the links between observed behaviour patterns and CIs, and then to synthesise the CIs into a diagnostic causal model.

Machine learning was rejected as a solution in our experimental requirements study for the simple reason that it requires large data volumes, which did not exist. Furthermore machine learning could discover many extraneous false positives from (a) other pathologies not related to dementia, (b) other behaviour patterns not related to any pathology such as individual styles of use, (c) patterns related to tasks, culture, individual mood and a host of other contextual factors. Unsupervised machine learning would therefore require a long and possibly fruitless analysis phase to exclude a large number of false positives. However, supervised machine learning may be more promising since it could be based on the hypotheses and heuristics we elicited from the experts. We expect domain knowledge requirements to evolve in further iterations of SAMS implementations as the expert system causal model is refined by data-mining results. Hence we are following the experience of [63] who used data mining to refine domain experts' knowledge in a medical diagnosis support system. Their Bayesian probabilistic model enabled uncertainty in evidence to be accounted for, which we expect to encounter in the SAMS signal-to-noise problem.

## 7 Reflections and lessons learned

### 7.1 Eliciting domain expert knowledge requirements

Recruiting and choosing the “right” experts to participate in the workshops was not easy. The information required by clinical experts to detect cognitive impairment and make a diagnosis relies on clinical history, medical examination and the results of cognitive test batteries, such as ACE and MoCA [3, 4], rather than on data derived from longitudinal computer use. We recruited experts who were clinical practitioners in the assessment and diagnosis of

dementia with some research experience, as well as medical/cognitive neuroscience research experts who specialised in neuropsychology and aspects of cognitive impairments, particularly related to neurodegenerative disorders. Incentivising experts, who are busy people, proved to be difficult. We agreed on a joint authorship publication (currently under review) as the main incentive, backed up with personal persuasion by the project team.

Terminology was a particular problem in workshop 1. There is incomplete consensus in the psychological literature on the definition of many cognitive terms, e.g. executive function, attention, types of memory. Furthermore, the meaning of these terms varies subtly between the cognitive psychology and medical research communities. This created a communication problem. In spite of this, we were able to achieve a consensus on CIs.

Experts had limited knowledge of computer data or the recording process, so data measures and outputs were novel. This created an “imagination gap” which we tried to fill with samples, scenarios and videos of screen-recorded examples. The activities and errors were organised in different layers, e.g. basic mouse and keystroke data, operating system (Windows) actions, and the application layers Word and Outlook e-mail. While these layers were useful for specifying data-capture requirements, they created some confusion among the experts. Few CI computer activity worksheets were completed, partly because the computer activities and their cognitive implications were unfamiliar concepts. Other reasons were task overloading: the four-section sheet was too long and detailed, and this compounded the imagination gap. These issues limited the outputs which could be achieved.

As a consequence of the experience with the first workshop, the tasks and structure of the second were redesigned to simplify cross-referencing activities to the disease progression categories and the use of cue cards showing fictitious graphs of user activity over time as cues for the same categorisation. The computer activities-CI cross-referencing task was assigned to a post-workshop questionnaire since the experts would have been exposed to many computer examples, and could refer to the workshop material when completing the questionnaire at their own pace.

The second workshop worked better and stayed on schedule. However, lack of consensus between the groups appeared because each group adopted a different interpretation of the importance of the activity-disease category association. For example, one group focused on specificity: does this activity indicate impairment in this CI only but not in others? The other group focused on sensitivity: does this activity indicate category  $x$  (weakly ... strongly)? Once this divergence was discovered, intervention by the facilitators enabled consensus to be established. The

experts needed considerable explanation about how the trend data presented in the final graph sorting task was related to the activities presented in the first card sort. Once they understood the data, they indicated specific patterns which they regarded as more indicative of potential pathology, adding further longitudinal data patterns which had not been presented. The follow-up questionnaire produced a reasonable consensus on a list of user behaviours and errors that should be prioritised as stronger indications of early cognitive decline.

To summarise what worked well:

Scenarios and video examples, briefings, group discussions.

Card sorts.

Simplified tasks in the second workshop.

Difficulties encountered:

Cross-referencing CIs to computer activities; the task was too complex.

Flip charts for focusing discussion on CIs and activities. Attempts to steer the discussion by recording possible consensus often provoked more disagreement.

Consensus process for linking computer activities to CIs was difficult because the experts' knowledge and experience in the IT domain were diffused.

Questionnaires: the number of participants was too low, partly because of the difficulty in convening an expert reference group in the first place, and partly because we were limited to experts who had participated in the workshops and therefore had the necessary briefing.

In conclusion, no complete causal model linking CIs to dementia diagnosis existed before the study, so attempting to achieve this to produce detailed requirements for SAMS was not realistic. Moreover, the experts had limited experience of computer activities and longitudinal observations of human computer interaction that might be linked to detecting the onset of dementia, so the study objectives lay beyond current expert knowledge.

In spite of our over-ambitious aims, the process did elicit a partial model with CIs and prioritised user errors and activities against dementia severity (stage of disease). This produced useful requirements for data mining in data-driven mode, i.e. how to interpret patterns derived from computer data in terms of CIs, and hence possible indicators of impairment. Output from the trend-graph analysis also gave useful indications about which type of activity pattern to ignore/concentrate on. The partial model also provided requirements for developing hypothesis-directed data-/text-mining tools. For example, executive function had one of the highest scores. Typical executive dysfunctions include planning

mistakes, losing the thread of a task or conversation, failing to complete tasks, difficulty in swapping between tasks, and poor/incorrect plans. We therefore specified "interpreter" functions to trawl the data looking for errors and incomplete sequences in computer operation, excessive switching between windows, repeated Internet searches, and incomplete tasks in e-mail and word processing.

From our experience, a modified Delphi technique appears to be appropriate for eliciting expert knowledge and conversations between experts in focused workshops deliver good insight. However, choice of seeding scenarios to guide expert conversations is critical. It is difficult to anticipate a priori the range of topics and viewpoints held by experts. Preparation before the workshops by asking experts to critique briefing materials is one possible improvement. More detailed discourse analysis [49] of expert conversation might produce further insights, but this is a resource-intensive process and for our purposes the gain was not apparent in a domain where knowledge was incomplete. The use of questionnaires to finalise consensus is useful since it allows experts to reflect on their judgement, removed from any potential group bias within a workshop. The downside lies in being restricted to the small number of experts who participated in the workshops and were therefore aware of the terminology and concepts.

To develop the above conclusions further recommendations for improving knowledge elicitation from experts are to focus on analysing guided conversations between experts, where two experts discuss a suitable task/scenario, the conversation is audio recorded and analysed without transcription thus avoiding the cost of discourse analysis. We propose the following guidelines for this approach:

1. Provide initial proposals for experts to react against, in our case a "straw man" causal model with cognitive indicators. This should stimulate conversations between experts which may give useful insights.
2. Drive the analysis from a process/task perspective, e.g. asking our experts to diagnose possible dementia given scenarios of fictitious patients and recorded data. This has the merit of focusing attention on problem solving and avoiding terminology arguments which we encountered.
3. In the preparation phase referred above, provide a standard set of scenarios, with straw man diagnoses.
4. Analyse the recorded conversations using the task structure and cognitive models of problem solving as reference points.
5. After the initial analysis present the results back to experts and ask them to critique the domain knowledge and revise the analysis as necessary.

The conversational approach could be integrated with workshops and questionnaires should sufficient expert be available.

## 7.2 Awareness and emergent requirements

We had expected that the SAMS project would extend the concept of awareness requirements from relatively simple goal-directed monitors to more complex awareness involving many variables and change that emerges over a long time period. Hence data- and text-mining tools become design requirements for implementing awareness of state of a complex entity, i.e. a person's mental health, and departure from the normal healthy behaviour. The unknowns at the outset of the project were the extent of the contributions which data-/text mining and intelligent interpreters might make to the solution. As knowledge elicitation from clinical experts proceeded, it became increasingly clear that intelligent interpreters would not solve the problem. Initial experimental data analysis revealed the critical unknown: the extent of variation in the recorded data. On reflection, this unknown might have been anticipated a priori; however, there were few previous longitudinal studies of "normal" human-computer interaction, possibly because of privacy concerns. Furthermore, the occasional interaction logging studies in a medical context concentrated on more controlled conditions with games or simpler data, such as mouse moves [48, 64]. As our awareness about the nature of the SAMS domain increased, i.e. the complexity and variability of human behaviour, new requirements in the form of investigation strategies emerged. These extended awareness requirements towards a flexible toolkit approach, in which a mix of data-mining tools and hypothesis-directed intelligent interpreters were integrated in an adaptive manner. Statistical data mining established the baseline normal variation (both within and between individuals); if the behaviour of an individual transgressed thresholds for several variables, then hypothesis-directed interpreters were invoked to search for more diagnostic evidence of dementia. This adaptive approach has produced further unknowns about the threshold triggers: how many variables and what level of statistical significance in variation to use? Once more we have to adopt an experimental approach, with a sensitivity analysis of variables' ranges to understand appropriate settings so false positives (unnecessary alarms) are missed, and undetected true positives are avoided.

The key lesson we draw from this experience concerns an experimental approach to RE when dealing with unknown domains. In complex domains, knowledge may not exist a priori; instead it has to be established by an iterative process of prototype implementation and extensive testing. While SAMS was a research project, and therefore complex and potentially unknown requirements were to be expected, we

expect experimental RE may be applicable to many awareness requirements applications in complex domains where change can only be established over time by sophisticated data analysis.

A final reflection concerns the boundary of unknowns and the role of the domain knowledge held by the project team members. In SAMS, three members out of ten had clinical and cognitive psychological expertise. This helped in discussion to establish consensus on the CIs, and was essential for planning the workshops. However, there was a knowledge gap with unknowns about the functionality of data-/text-mining tools on the clinical side, while the software engineers' unknowns arose from lack of understanding about the complexities of diagnostic knowledge for dementia. Only one member had both computer and cognitive psychological expertise. This bridging role guided the overall structure of the elicitation process and helped to prepare the requirements goals, i.e. the causal model and linking computer activities to CIs. However, it also contributed towards the task overloading experienced in the first workshop. In retrospect, it would have been advisable to pursue a more modest structured agenda of building up the model in layers, concentrating on knowledge closer to the experts' domains, i.e. CIs and dementia. De facto that was the outcome of the process. There were also limitations in how much the single person could achieve in closing the unknowns gap between the software engineers and clinical experts.

## 8 Discussion

The contribution from our RE experience in the SAMS project has been to extend the concept of awareness requirements [12] in complex domains where the object of monitoring may not be the designed system but entities in the domain, such a people, organisations and the environment. The second contribution extends the unknowns framework for RE [6, 7] to explore the boundaries of known knowledge in the domain and with development teams. In doing so, we have advocated a new "experimental" approach to RE in complex domains. Finally we have evaluated a standard knowledge acquisition approach based on Delphi techniques [51, 52] for RE.

The knowledge acquisition process we adopted followed established methods and, generally, the workshop expert-conversation approach produced good results; however, our experience showed that eliciting knowledge-based requirements is still a difficult task in spite of decades of practice-based refinement [10, 26]. Some of the problems, e.g. the language knowledge gap, were encountered even though the RE team did have appropriate domain knowledge. Our experience suggests that requirements engineers need considerable domain knowledge to bridge the communication

gap in knowledge elicitation with experts, in contrast to the findings of [30] for domains with conventional requirements. The Delphi method [51, 52] had to be modified extensively during the process, indicating that flexible and adaptive approaches to elicitation are advisable. Green et al. [51] experienced difficulties with established consensus and concluded that the Delphi process was more suited to domains where opinions, problems and concepts are well known in the expert community. On reflection, our goals may have been over-ambitious and possibly unrealisable, since no mature diagnostic model of cognitive and behavioural symptoms of early onset of dementia existed. One conclusion is to assess the degree of maturity in domain knowledge, and if the area is uncertain, frame the elicitation exercise akin to a research road map, i.e. discover the unknowns, where further research is necessary; and the partially knowns, with less than 100% confidence in the knowledge; as well as more mature knowns. Our experience reflects the degree of “knowness” at project initiation in RE. The fact that expert knowledge and a causal model for diagnosis would be needed was a partial knowledge (from the analysts’ perspective) but the degree of existing domain knowledge was “Unknown”. As it transpired, much of this knowledge was indeed unknown by the experts; furthermore, unknown (by the analysts and experts) requirements emerged during the project, e.g. the individual variation problem, which were “unknown unknowns” from both analyst and expert perspectives.

Recruitment of experts was another known problem [11] we encountered, although we did select a range of experts covering research as well as clinical practice in dementia. One escape from recruitment limitations might be to crowd-source requirements, an approach that has attracted increasing interest in RE [65]. While this remains a possibility for future work, we expect crowd-sourcing will be no panacea for expert requirements. First, the problem framing is critical and we had to adapt our process rapidly between the workshops. Understanding problem framing will be more difficult over the Internet. Secondly, the quantity of briefing material to promote expert understanding of the requirements is considerable, and the motivation to digest such information online may be low. Finally, selecting the right kind of experts and motivating them, even with crowd-sourcing selection panels, is likely to be difficult.

We have noted the limitation of the number of experts in this study; however, we believe that repeating the exercise with more experts may not produce radically improved results. Given the complexity of the domain, and the task which was on the boundary of the experts’ knowledge, improving consensus may be difficult. The approach we adopted in retrospect may have been appropriate for exploratory knowledge discovery in a partially known domain,

whereas most knowledge elicitation techniques assume domains with mature and stable knowledge [11, 20, 50].

RE in healthcare has adopted a range of approaches, ranging from ethnographic investigation of work practices [66] to interviews and JAD workshops [67], formal modelling [68] and machine learning for analysing traceable dependencies in healthcare documents [69]. The diversity of approaches reflects the wide range of applications within the healthcare domain. For user interface-related requirements, scenarios, prototypes and an iterative approach appear to be successful [1, 70]; however, hidden requirements inherent in complex domains may be difficult to discover [66]. The review of healthcare data analytics [71] discusses the problem of complex ontologies and terminology in healthcare which we encountered with our experts; [71] propose a reference strategy for information modelling and healthcare data analytics which we suggest needs to be developed to account for knowledge-based modelling. The KMoS-RE approach [39] is related to the approach we adopted to tacit knowledge elicitation with interviews and workshops, although [39] used more specific lexical discourse analysis techniques, whereas we took a taxonomic approach to help expert formalisation of their knowledge. The SAMS investigation produced interesting implications for the nature of requirements. First, requirements that emerged in a partial causal model for diagnosis illustrated that even in expert domains, knowledge may be incomplete. RE may therefore have to adopt a more experimental approach to filling in the gaps in expert knowledge. This approach will not be experimental in the classic sense of hypothesis-driven controlled experiments; instead it is an iterative prototype and evaluation process extending knowledge by a sensitivity analysis investigating key variables. Experimental RE may be viewed as an extension of participatory design or co-design collaborations with domain experts [72, 73]; however, we advocate a more systematic questioning approach. The second implication concerns requirements for configuring data- and text-mining tools. As software becomes increasingly tool driven, configuration and tool parameterisation will be more common. Requirements therefore have to be cognisant of tool capability, essentially an intertwining of requirements and architecture [74].

The SAMS experience with a knowledge intensive domain has implications for theoretical models of RE, in particular the well-known KSR paradigm [75]. Domain knowledge as mental constructs of experts becomes transformed by the requirements process in specifications of the machine (in our case causal models for signs of dementia), which can be applied to implementable reasoning mechanism such as Bayesian Nets. However, other requirements, in our cases “design requirements” in Sect. 6 can also be argued to be domain knowledge held by expert computer scientists that also become transformed into specifications

of the machine, as data- and text-mining processes. Indeed the close relationship between K, S and R is acknowledged by the inevitable intertwining and requirements and architecture [13, 74].

The requirements for configuring data-mining tools which we elicited suggest a new perspective on awareness requirements [12]. We have argued that AwRE should consider data-mining applications where monitoring large volumes of data across time is necessary to modify original requirements [76]. The high-level requirements in SAMS were run-time monitors in the classic sense [12, 41] but extend the monitors/awareness requirements concept with the perspective of multiple interacting variables and large, persistent data volumes. Whereas previous run-time monitors and awareness requirements have made the tacit assumption that there is a run-time loop coupling monitors that check systems behaviour against goals and corrective action for any observed deviations [12, 18, 41], we argue the loop between monitoring and corrective feedback need not be tightly coupled. Furthermore it may involve human and well as system agency. For example in SAMS, monitors were intended to detect deviations in human behaviour against a standard for normality. If significant changes were detected, then the system action was alert the user and their doctor. This perspective of longer-term awareness requirement produced data-mining and data analysis requirements that indicated tool-oriented solutions for many run-time monitors where persistent data is present, while other requirements become tool setting and configurations. Furthermore, goals such as hypothesis-directed classifiers, contribute requirements for generic data-mining tool development. Awareness requirements in complex systems need to be extended in time and scope to include multiple variables with sensor and intelligent interpreter components that can detect complex changes in several variables. Such complex AwReqs may need embedded system models of complex domain phenomena, such as users and their behaviour, which have been developed in user modelling research [77, 78]. However, as we discovered, the expert knowledge for developing such models may not exist, so AwRE may have to adopt data-mining tools as a means of monitoring complex systems over an extended time period. Requirements in this approach will need to inform selection of appropriate tools for the nature of the problem domain, and specification of tool controls and settings. In complex systems, AwRE may need to define the process and tool-integration strategies by which knowledge is discovered using machine learning, data and text mining.

The experience and lessons learned we report have to be interpreted within the context of the medical application

domain and project context. While we make no specific claims for generality, we point to several generalisable findings for awareness requirements/run-time monitors, principally extension of AwRE to persistent open-loop monitoring, tool-driven approaches to persistent data in AwRE; the need to model based interpretation of complex persistent data, and the difficulties inherent in eliciting such models from experts in domains with incomplete knowledge. These findings were probably be german to a wide variety of complex socio-technical systems where human or socio-technical system behaviour form the subject matter for awareness. Our findings on elicitation of requirements from experts could have been improved by recruitment of more experts, but it also illustrates the problem of such analyses: experts are generally few in number but busy people.

SAMS software for monitoring computer activity has been implemented and recorded activity from 33 volunteers. Analysis of this data is continuing. Future software development will be iterative, following an experimental approach, as we discover requirements for analysing patterns in the recorded data that might indicate onset of dementia. We expect requirements will emerge from inspection of exceptions and unusual results from data and text mining, which will have to be cross-referenced to the CIs and findings from the workshop to establish whether the patterns are a possible indicator. Domain knowledge requirements will evolve as SAMS implements the expert system causal model which is refined by data- and text-mining results, following the experience of [63] who used data mining to refine domain experts' knowledge in a medical diagnosis support system. A final reflection on the SAMS experience of the unknown unknowns in RE [7] is that the known-partially known boundary is fluid and depends on a dialogue that discovers knowledge not only between experts but also between experts and different members of the development team. Further research is needed to understand the relationship between tacit and unknown/partially known knowledge [79]. Recognising the limits of the current knowledge should be an important aim during the process because awareness of the "boundary of ignorance" prompts requirements for knowledge discovery where data exists, and requirements for further research where it does not. SAMS may illustrate a class of systems where some requirements remain as unknowns even when the system is in operation. As SAMS is a research project, the requirements are limited by the horizon of expert knowledge; hence, experimental RE using iterations of data-driven research leading to prototype improvement appear to be the way forward.



**Acknowledgements** The work described in this paper is funded by the Engineering and Physical Sciences Research Council (EPSRC) in the UK, within the Software Architecture for Mental health Self-management (SAMS) project, references EP/K015796/1, EP/K015761/1 and EP/K015826/1. The authors wish to thank the experts who participated in the workshop: E. Poliakoff, D. Montaldi, J. Thompson, J. Rust, K. McDonald and D. Bruno.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix B: Summary of Requirements produced at different stages of the investigation

- (a) High-level Requirements for Data Mining—from workshops
1. Look for changes in activity patterns over time within each individual.
  2. Establish a baseline of variation for each individual, then use learning algorithms to detect changes against the baseline.

**Table 3** List of computer activities association with indications of dementia. A complete list and further details can be found in [55]

Computer-use behaviour change	Cognitive domain	Cognitive term	Level of consensus	Strong indication	Some indication
Sentences are less dense than usual (i.e. uses less verbs, adjectives, adverbs)	Language	Production	Full	100	0
Opens a series of different incorrect folders before opening the correct document in the correct folder	Memory	Recall	Full	100	0
Opens and closes the same wrong Word document numerous times	Executive function	Inhibition	Full	80	20
Uses a reduced set of vocabulary in e-mails	Language	Production	Full	80	20
Repeatedly types a series of different incorrect passwords (e.g. Dog1; Cat1; Dog2) after receiving “incorrect username/password” messages	Memory	Declarative	Full	80	20
Opens a series of incorrect Word documents before opening the correct document	Memory	Recall	Full	80	20
Repeatedly types a series of different incorrect passwords (e.g. Dog1; Cat1; Dog2) after receiving “incorrect username/password” messages	Memory	Recall	Full	80	20
Clicks the mouse more than five times in rapid succession on the program icon	Perception and action	Motor control	Full	80	20
Repeatedly double clicks in areas of the screen that are close to (but not on) the program icon	Perception and action	Spatial perception	Full	80	20
Repeatedly types the same incorrect password (e.g. Dog1; Dog1...) despite receiving “incorrect username/password” messages	Executive function	Self-error monitoring	Full	60	40
Sentences are shorter than usual	Language	Production	Full	60	40

## Appendix A

See Table 3.

3. Combine many user behaviour variables to establish the baseline and detect significant changes.
4. Prioritise certain user behaviours over others, such as errors and incomplete sequences of actions.
5. Benchmark individual change against known statistics for age-related change, and clinical tests of activity in people with MCI/dementia.
6. Create hypothesis-directed pattern recognisers to detect problems motivated by the prioritised CIs, e.g. error classifiers in behaviour sequence analysis, excessive switching between windows.

## (b) Design Requirements for Data Capture (Sect. 6.1)

- Record events from keyboard and mouse with time stamp
- Record event identities as keystrokes (QUERTY) and mouse moves with button clicks.
- Interpret mouse actions (clicks) on significant UI objects as user actions on windows, sliders, icons, etc.
- Interpret mouse actions for application commands in selected applications: Word and Outlook, as Open Close Windows, Select Menu Option {ID}, etc.
- Interpret mouse actions on web browsers to record Browser commands and selecting links with URLs.

## (c) High-level Design Requirements for Data Mining—from investigations by software design team in light of workshop requirements (Sect. 6.2)

## 1. Data-mining tool configuration

- Analyse variation within individuals to establish trends by using months 1–3 as training data for learning algorithms with subsequent months as test data.
- Configure event-sequence mining tools to analyse frequent event-sequences which might represent normal user behaviour and less frequent, exceptional sequences that could indicate errors.

## 2. Data analysis heuristics

- In trend data look for decreasing frequencies or increasing durations with temporal irregularities such as rapid changes in frequencies.
- After-activity gaps investigate frequencies before and after as well as pattern regularity.

## 3. Hypothesis-directed classifiers: analysers to detect specific user interaction sequences:

- Incomplete drag sequences ,when no final action is found, such as delete document or file document in folder (executive dysfunction/attention)
- Several sequential undo–redo actions (executive dysfunction)
- Repeated text in same e-mail message (attention, executive dysfunction)
- Incorrect or inappropriate words in message (semantic error).

## References

1. Sutcliffe AG, Rayson P, Bull C, Sawyer P (2014) Discovering affect-laden requirements to achieve system acceptance In: Proceedings 22nd IEEE International Requirements Engineering Conference (RE'14). IEEE Computer Society Press, Los Alamitos, pp 173–182
2. Stringer G, Sawyer P, Sutcliffe AG, Leroi I (2015) From click to cognition. In: Bruno D (ed) *The Preservation of Memory*. Psychology Press, Palo Altom, pp 93–103
3. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* 21:1078–1085
4. Nasreddine ZS, Phillips NA, Bedirian V et al (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53:695–699
5. Han J, Kamber M, Pei J (2011) *Data mining: concepts and techniques*. Elsevier, Amsterdam
6. Sawyer P, Bencomo N, Whittle J, Letier E, Finkelstein A (2010) Requirements-aware systems: a research agenda for RE for self-adaptive systems. In: Proceedings 18th IEEE international conference on requirements engineering (RE'10). IEEE Computer Society Press, Los Alamitos, pp 95–103
7. Sutcliffe AG, Sawyer P (2013) Requirements elicitation: towards the unknown unknowns. In: Proceedings 21st IEEE international requirements engineering conference (RE-13). IEEE Computer Society Press, Los Alamitos, pp 92–104
8. Maiden NAM, Rugg G (1996) ACRE: selecting methods for requirements acquisition. *Software Engineering Journal* 11(3):183–192
9. Coughlan J, Macredie RD (2002) Effective communication in requirements elicitation: a comparison of methodologies. *Requirements Eng* 7(2):47–60
10. Davis A, Dieste O, Hickey A, Juristo N, Moreno AM (2006) Effectiveness of requirements elicitation techniques: empirical results derived from a systematic review. In: Proceedings 14th IEEE international symposium on requirements engineering (RE'06). IEEE Computer Society Press, Los Alamitos, pp 176–185
11. Kidd A (ed) (2012) *Knowledge acquisition for expert systems: a practical handbook*. Springer, Berlin
12. Souza VES, Lapouchnian A, Robinson WN, Mylopoulos J (2013) Awareness requirements. In: de Lemos R, Giese H, Muller HA, Shaw M (eds) *software engineering for self-adaptive systems II*. Springer, Berlin, pp 133–161
13. Sutcliffe AG (2009) On the inevitable intertwining of requirements and architecture. In: Lytinen K, Loucopoulos P, Mylopoulos J, Robinson B (eds) *Design requirements engineering: a multi-disciplinary perspective for the next decade*. Springer, Berlin
14. Gause DC, Weinberg GM (1989) *Exploring requirements: quality before design*. Dorset House, New York
15. Sommerville I, Sawyer P (1997) *Requirements engineering: a good practice guide*. Wiley, Chichester
16. Robertson S, Robertson J (1999) *Mastering the requirements process*. Addison Wesley, Harlow
17. Sutcliffe AG (2002) *User-centred requirements engineering*. Springer, London
18. Van Lamsweerde A (2009) *Requirements engineering: from system goals to UML models to software specifications*. Wiley, Chichester
19. Hickey AM, Davis AM (2004) A unified model of requirements elicitation. *J Manag Inf Syst* 20(4):65–84

20. Liou YI (1990) Knowledge acquisition: issues, techniques, and methodology. In: Proceedings of the 1990 ACM SIGBDP conference on trends and directions in expert systems. ACM, New York
21. Byrd TA, Cossick KL, Zmud RW (1992) A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Q* 16:117–138
22. Kelly GA (1977) Personal construct theory and the psychotherapeutic interview. *Cognit Ther Res* 1(4):355–362
23. Gaines BR, Shaw MLG (1993) Knowledge acquisition tools based on personal construct psychology. *Knowl Eng Rev* 8(1):49–85
24. Fransella F, Bell R, Bannister D (2004) A manual for repertory grid technique. Wiley, Chichester
25. Tan FB, Hunter MG (2002) The repertory grid technique: a method for the study of cognition in information systems. *MIS Q* 26:39–57
26. Sommerville I, Sawyer P (1997) Viewpoints: principles, problems and a practical approach to requirements engineering. *Ann Softw Eng* 3(1):101–130
27. Rugg G, McGeorge P (1997) The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Syst* 14(2):80–93
28. Upchurch L, Rugg G, Kitchenham B (2001) Using card sorts to elicit web page quality attributes. *IEEE Softw* 18(4):84
29. Edwards HM, McDonald S, Young SM (2009) The repertory grid technique: its place in empirical software engineering research. *Inf Softw Technol* 51(4):785–798
30. Niknafs A, Berry DM (2012) The impact of domain knowledge on the effectiveness of requirements idea generation during requirements elicitation. In: Proceedings 20th IEEE international requirements engineering conference. IEEE Computer Society Press, Los Alamitos, pp 18–190. Doi: <https://doi.org/10.1109/re.2012.6345802>
31. Brown BB (1968) Delphi process: a methodology used for the elicitation of opinions of experts. (No. RAND-P-3925). Rand Corporation, Santa Monica CA
32. Gutierrez O (1989) Experimental techniques for information requirements analysis. *Inf Manag* 16(1):31–43
33. Paul CL (2008) A modified Delphi approach to a new card sorting methodology. *J Usability Stud* 4(1):7–30
34. Polanyi M (1997) The tacit dimension. In: Prusak L (ed) *Knowledge in organizations*. Oxford University Press, New York, pp 135–146
35. Gervasi V et al (2013) Unpacking tacit knowledge for requirements engineering. In: Maalej W, Thurimella AK (eds) *Managing requirements knowledge*. Springer, Berlin, pp 23–47
36. Collins H (2007) Bicycling on the moon: collective tacit knowledge and somatic-limit tacit knowledge. *Organ Stud* 28(2):257–262
37. Ferrari A, Spoletini P, Gnesi S (2015) Ambiguity as a resource to disclose tacit knowledge. In: Proceedings 23rd IEEE international requirements engineering conference (RE-15). IEEE Computer Society Press, Los Alamitos
38. Cleland-Huang J (2015) Mining domain knowledge. *IEEE Softw* 32(3):16–19
39. Olmos K, Rodas J (2014) KMoS-RE: knowledge management on a strategy to requirements engineering. *Requir Eng* 19:421–440
40. Fickas S, Feather MS (1995) Requirements monitoring in dynamic environments. In Proceedings 2nd IEEE international symposium on requirements engineering (RE'95), pp 140–147
41. Robinson W (2006) A requirements monitoring framework for enterprise systems. *Requir Eng* 11(1):17–41
42. Whittle J, Sawyer P, Bencomo N, Cheng B, Bruel J-M (2010) RELAX: a language to address uncertainty in self-adaptive systems requirements. *Requir Eng* 15(2):177–196
43. Salehie M, Tahvildari L (2009) Self-adaptive software: landscape and research challenges. *ACM Trans Auton Adapt Syst* 4(2):14
44. Smith B, (1982) Reflection and semantics in a procedural programming language. PhD thesis, MIT Laboratory for Computer Science Report MIT-TR-272. MIT, Cambridge
45. Dalpiaz F, Giorgini P, Mylopoulos J (2013) Adaptive socio-technical systems: a requirements-driven approach. *Requir Eng* 18:1–24
46. Huebscher MC, McCann JA (2008) A survey of autonomic computing: degrees, models, and applications. *ACM Comput Surv* 40(3):1–28
47. Ahmed S, De Jager C, Wilcock G (2012) A comparison of screening tools for the assessment of mild cognitive impairment: preliminary findings. *Neurocase* 18(4):336–351
48. Seelye A, Hagler S, Mattek N, Howieson DB et al (2015) Computer mouse movement patterns: a potential marker of mild cognitive impairment. *Alzheimer's Dement Diagn Assess Dis Monit* 1(4):472–480
49. Belkin NJ, Brooks HM, Daniels PJ (1987) Knowledge elicitation using discourse analysis. *Int J Man-Mach Stud* 27(2):127–144
50. Shadbolt NR, Smart PR (2015) Knowledge elicitation. Evaluation of human work, 4th edn. CRC Press, Boca Raton, pp 163–200
51. Green B, Jones M, Hughes D, Williams A (1999) Applying the Delphi technique in a study of GP's information requirements. *Health Soc Care Community* 7(3):198–205
52. Linstone K, Turoff M (1975) The Delphi method: techniques and applications. Addison Wesley, Reading
53. Sutcliffe AG, Thew S, De Bruijn O et al (2010) User engagement by user-centred design in e-health. *Philos Trans R Soc* 368:4209–4224 (**special issue on e-Science, series A**)
54. Boose JH (1989) A survey of knowledge acquisition techniques and tools. *Know Acquis* 1(1):3–37
55. Couth S, Stringer G, Leroi I, Sutcliffe A et al (2017) Which computer-use behaviours are most indicative of cognitive decline? Insights from an expert reference group. *Health Inform J*. Doi: <https://doi.org/10.1177/1460458217739342>
56. Stringer G, Couth S, Brown L, Montaldi D et al (2018) Can you detect early dementia from an email? A proof of principle study of daily computer use to detect cognitive and functional decline. *Int J Geriatr Psychiatry*. Doi: <https://doi.org/10.1002/gps.4863>
57. Snowdon DA, Kemper SJ, Mortimer JA et al (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study. *J Am Med Assoc* 275:528–532
58. Le X, Lancashire I, Hirst G, Jokel R (2011) Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Lit Linguist Comput* 26(4):435–461
59. Han J, Pei J, Kamber M (2011) *Data mining: concepts and techniques*. Morgan Kaufman, New York
60. Pan J, Mackenzie G (2003) On modelling mean-covariance structures in longitudinal studies. *Biometrika* 90(1):239–244
61. Goldstein H (2011) *Multilevel statistical models*. Wiley, Chichester
62. Zhengxing H, Wei D, Lei J et al (2014) Discovery of clinical pathway patterns from event logs using probabilistic topic models. *J Biomed Inform* 47:39–57
63. Alonso F, Martinez L, Perez A, Valente JP (2012) Cooperation between expert knowledge and data mining discovery knowledge: lessons learned. *Expert Syst Appl* 30:7524–7535
64. Kaye J, Mattek N, Dodge HH et al (2014) Unobtrusive measurement of daily computer use to detect mild cognitive impairment. *Alzheimer's Dement* 10:10–17
65. Crowd-Based Requirements Engineering (CrowdRE) (2015) In: *IEEE 1st International Workshop on Year 2015*. IEEE Computer Society Press, Los Alamitos, pp 3–4. Doi: <https://doi.org/10.1109/crowdre.2015.7367580>

66. Jirotko M, Procter R, Hartswood M et al (2005) Collaboration and trust in healthcare innovation: the eDiaMoND case study. *J Comput Supported Cooperative Work* 14(4):369–398
67. Cysneiros LM (2002) Requirements engineering in the health care domain. In: 2002 Proceedings IEEE joint international conference on requirements engineering. IEEE, Los Alamitos, pp 350–356
68. Jorgensen JB, Bossen C (2003) Requirements engineering for a pervasive health care system. In: 2003 proceedings requirements engineering conference. IEEE, Los Alamitos, pp 55–64
69. Cleland-Huang J, Czauderna A, Gibiec M, Emenecker J (2010) A machine learning approach for tracing regulatory codes to product specific requirements. In: Proceedings of the 32nd ACM/IEEE international conference on software engineering-volume 1. ACM, New York, pp 155–164
70. Holmegaard M, Jørgensen JB, Loft MS, Stissing MS (2015) Requirements problems in the development of a new user interface for healthcare equipment. In: IEEE 23rd international requirements engineering conference (RE), pp. 315–323
71. Liu L, Feng L, Cao Z, Li J (2016) Requirements engineering for health data analytics: Challenges and possible directions. In: IEEE 24th international requirements engineering conference, RE-2016, pp 266–275. Doi: <https://doi.org/10.1109/re.2016.48>
72. Pilemalm S, Timpka T (2008) Third generation participatory design in health informatics: making user participation applicable to large-scale information system projects. *J Biomed Inform* 41:327–339
73. Bratteteig T, Wagner I (2016) Unpacking the notion of participation in participatory design. *Comput Supported Cooperative Work (CSCW)* 25(6):425–475
74. Nuseibeh B (2001) Weaving together requirements and architecture. *IEEE Comput* 34(3):115–119
75. Jackson M (2001) Problem frames: analysing and structuring software development problems. Addison-Wesley, Boston
76. Sutcliffe A, Sawyer P (2016) Beyond awareness requirements. In: Proceedings of 24th IEEE international requirements engineering conference, Beijing, RE Next track, 2016. IEEE Computer Society Press, Los Alamitos
77. Ricci F, Rokach L, Shapira B (eds) (2015) Recommender systems handbook. Springer, Berlin
78. Yin H, Cui B, Chen L, Hu Z, Zhou X (2015) Dynamic user modeling in social media systems. *ACM Trans Inform Syst (TOIS)* 33(3):10
79. Ferrari A, Spoletini P, Gnesi S (2016) Ambiguity and tacit knowledge in requirements elicitation interviews. *Requir Eng* 21:333–355