

**Thesis title:** An analysis of the relationship between individuals' perceptions of privacy and mobile phone location data - a grounded theory study.

**Andrea Gorra**, Leeds Metropolitan University, UK  
Comments sent to [a.gorra@leedsmet.ac.uk](mailto:a.gorra@leedsmet.ac.uk) would be most appreciated.

## **Chapter 3 Research Methodology**

This chapter introduces the research methodology used for this study and how it has guided data collection, analysis and development of theory. Firstly, essential background and fundamental guidelines common in different approaches to grounded theory methodology (GTM) are provided. The subsequent three sections describe the data collection phases for this study, which consisted of a mobile phone tracking pilot study, in-depth interviews and a survey. The chapter concludes by explicating the analysis approach for the empirical data.

### **3.1 Grounded theory methodology - an overview**

"If someone wanted to know whether one drug is more effective than another, then a double blind clinical trial would be more appropriate than grounded theory study. However, if someone wanted to know what it was like to be a participant in a drug study [...], then he or she might sensibly engage in a grounded theory project or some other type of qualitative study."

(Strauss and Corbin, 1998, p. 40).

Strauss and Corbin's quote above encapsulates the essence of when it is best to use grounded theory methodology for a research project. GTM provides useful tools to learn about individuals' perceptions and feelings regarding a particular subject area. Quantitative data may be useful in measuring attitudes across a large sample, however, GTM offers a powerful methodological framework if the aim of the study is to learn about individuals' perceptions.

GTM shares the following characteristics with other qualitative methods, which correspond to those of this study:

- Focus on everyday life experiences
- Valuing participants' perspectives
- Enquiry as interactive process between researcher and respondents
- Primarily descriptive and relying on people's words

(Marshall and Rossman, 1999)

Grounded theory originated in the 1960s in the United States in the fields of health and nursing studies. Barney Glaser and Anselm Strauss' influential book from 1967 'The Discovery of Grounded Theory' articulates the authors' research strategies for studies of patients dying in hospitals. Their successful collaborative study was perceived as a response to the predominantly quantitative research paradigms at the time. Grounded theory methodology advocates *creating new theory* consisting of interrelated concepts rather than *testing existing theories*. A study guided by GTM does not seek representativeness to achieve statistical generalisability but instead aims to explain and sometimes predict phenomena based on empirical data. The data collection typically encompasses in-depth interviews but can also include other sources of data such as existing research literature and quantitative data. GTM provides guidelines for data collection and analysis consisting of coding, comparisons between data, memo writing and theoretical sampling.

### 3.1.1 Data collection and analysis in grounded theory

GTM uses a form of purposive sampling, known as *theoretical sampling*, where participants are selected according to criteria specified by the researcher and based on initial findings. Early analysis of data indicates issues that need exploration; hence the sampling process is guided by the on-going theory development. Data

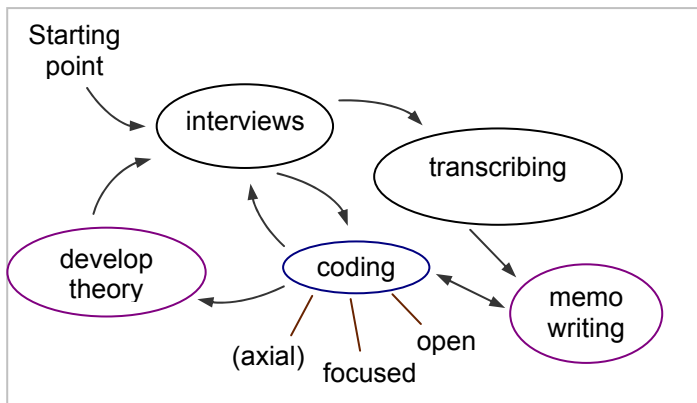


Figure 3.1: Steps in developing a grounded theory

collection and analysis take place in alternating sequences (see Figure 3.1). This can also be described as an iterative cycle of induction and deduction, consisting of collection of data and constant comparison between results and new findings in order to guide further data collections

(Strauss and Corbin, 1990; Miles and Huberman, 1994). For these reasons the development and identification of variables does not take place prior to data collection but instead as part of the data collection process. Consequently, the variables or concepts are initiated by the interviewee and further developed and conceptualised by the researcher. Data is collected until *theoretical saturation* is

reached, in other words until no new or relevant data emerges regarding a category and relationships between categories are established (Strauss and Corbin, 1998).

Interview questions should give as little guidance as possible to allow the interviewees to talk about what is of importance to them regarding a given context. The researcher then needs to extract those phenomena or experiences significant to the interviewee by assigning a conceptual label, known as a *code*. Several codes can be grouped into more abstract *categories* which will eventually form the basis for the developing theory.

#### **3.1.1.1 Coding interviews as part of the analytic process**

Interview coding is used to capture what is in the interview data, to learn how people make sense of their experiences and act on them. Coding is the first step of data analysis, as it helps to move away from particular statements to more abstract interpretations of the interview data (Charmaz, 2006).

Grounded theory methodology advocates using several coding techniques to examine interviewee's accounts at different levels. *Open coding*, also known as line-by-line coding, provides a good starting point to identify initial phenomena and produce a list of themes of importance to the interviewee. Conceptual labels are attached to almost every line in the interview transcript to capture what has been said. These labels can correspond closely to the interview context and when taken from the interviewee's own words, are known as *in vivo code*. Codes are assigned to participants' words and statements to develop concepts, constituting the start of the analytic process.

The detailed and meticulous process of line-by-line coding helps to open up the text and interpret the transcript in new and unfamiliar ways which also helps test the researcher's assumptions. Strauss and Corbin (1998) suggest using initial or 'sensitising questions', to help the researcher grasp what the data might be indicating. Suggested questions are "Who are the actors involved?", "What are the actors' definitions and meaning of these phenomena or situations?" (Strauss and Corbin, 1998, p. 77).

The next coding phase is more abstract than open coding and known as *focused coding* or selective coding. Focused codes are applied to several lines or paragraphs in a transcript and require the researcher to choose the most telling codes to represent the interviewee's voice. Using open codes as a starting point,

the process of focused coding helps to verify the adequacy of the initial concepts developed. As the focused codes will be applied and therefore 'tested' on further interview transcripts.

Another subsequent phase of coding is *axial coding*, defined by Strauss and Corbin as "the act of relating categories to subcategories along the lines of their properties and dimensions" (Strauss and Corbin, 1998, p. 123). The aim of axial coding is to add depth and structure to existing categories (see also 3.1.1.2, below). Charmaz (2006) explains that axial coding re-assembles data that has been broken up into separate codes by line-by-line coding. Strauss and Corbin (1998) use axial coding to investigate conditions of situations described in the interview, their actions and consequences. Charmaz (2006) warns that axial coding applies a too rigid and formal frame to the data analysis. Instead she recommends the less formalised approach of reflecting on categories, sub-categories and to establish connecting links between these to make sense of the interview data. The most abstract level of coding is *theoretical coding*, which explores the relationships that have been established between categories. Several 'rules' or "analytic coding families" are put forward by Glaser (1978) to develop an advanced analysis of the subject area.

#### **3.1.1.2 Developing categories**

The general process of how to code an interview and develop a theory is depicted in simplified form in Figure 3.2, below. After coding several interview transcripts a researcher can identify many issues that are of importance to the respondents. These issues are also known as *phenomena* and are assigned a conceptual label to become a *code*, also known as a *concept* by Strauss and Corbin (1998). Some codes or concepts will share the same or similar characteristics and can be pulled together into more abstract *categories*, which can typically be interlinked and build the basis for a theory.

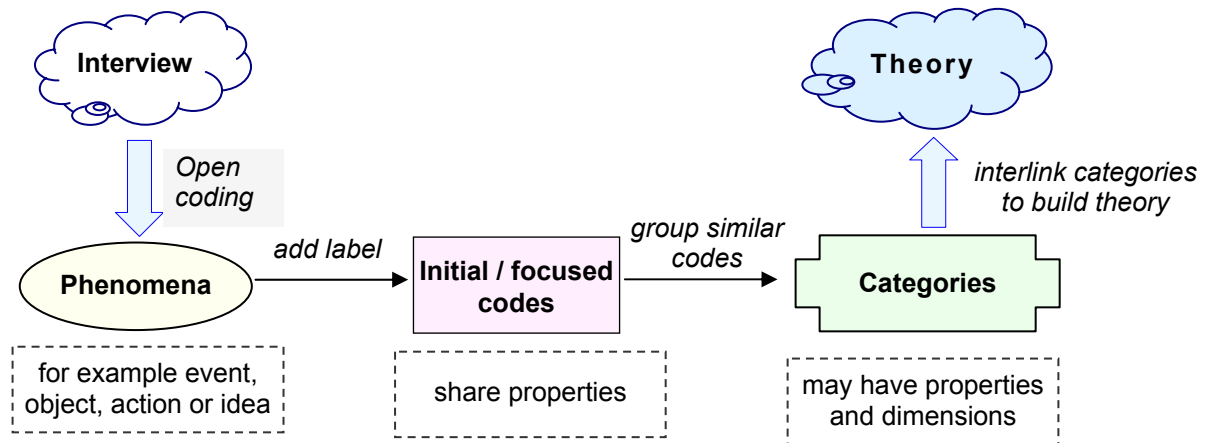


Figure 3.2: Coding steps in grounded theory (after Straus and Corbin, 1998)

It should be stressed that categories have to 'earn' their way into an emerging theory (Glaser, 1978). Grounded theory methodology typically does not use quantifying data to obtain meaning. However, counting the frequency with which categories occur in interview transcripts can be useful to confirm their importance for the interviewees. Categories can carry so-called *properties and dimensions*. A property is a general or specific characteristic of a category, whereas a dimension denotes the location of a property along a continuum or range (Strauss and Corbin, 1998). For example, the category 'phone ownership' could have the property 'spending' with dimensions ranging from 'low' to 'high spending'.

The central or *core category* is a distinctive category that sits at the heart of the developed theory and summarises what is happening. All other major categories should relate to the core category, which ought to appear frequently in the data (Strauss and Corbin, 1998). The development of the core category for this study is described in Chapter 4 - Presentation of Findings, section 4.2.4 to demonstrate the development of categories from codes.

Coding shapes the analytic frame and provides the skeleton for the analysis (Charmaz, 2006). Charmaz sees coding as an important link between collecting data and developing theory but also as a connection between empirical reality and the researcher's view of it. Coding highlights problems, issues, concerns and matters of importance to those being studied. Strauss and Corbin (1998) refer to categories as having 'analytic power', due to their potential to explain and predict. 'Constant comparisons' between collected data, codes, categories and initial findings help to crystallise ideas to become part of the emerging theory.

### **3.1.2 Use of grounded theory methodology for this study**

The data collection and analysis for this study followed a cyclical process typical for GTM, by using early findings to shape the on-going data collection (see Figure 3.1). The pilot study involved tracking participants' mobile phones and conducting eight interviews. This data collection phase was followed by five more in-depth interviews that helped to explore issues raised in the pilot study. The survey aimed at exploring the research area on a wider scale than would have been possible with interviews being conducted and analysed by a sole researcher. More interviews were undertaken after the completion of the survey and these could address issues brought up by initial survey findings. Details about the sampling approaches for each of the three data collection phases can be found in the relevant sections in this chapter (3.2 Pilot study: Mobile phone location tracking , 3.3 Interviews, 3.4 Survey).

For this study, open coding was used for the pilot study interviews with the help of qualitative analysis software NVivo. However, using the software did not produce meaningful results and hence the pilot study interviews were re-coded with pen and paper after conducting the survey (see section 4.1.4 False starts with initial coding in NVivo). Focused coding was utilised for the next two interview phases, which used the initial codes as a basis. Axial coding was not used in this study because the method of specifying properties and dimensions for each category seemed too prescriptive and did not help the analysis of the data (see section 4.2.6 Importance of mobile phone settings). For the same reasons, theoretical coding was not adopted. Instead, careful comparisons between respondents' statements, as well as between codes and categories were undertaken, without being restricted to interpret participants' words within a framework of properties and dimensions.

The researcher's decision to use grounded theory methodology was only taken after conducting the pilot study. An initial analysis of the pilot interviews showed that it was not suitable to base the overall research project on existing theoretical models. The focus needed to be on participants' experiences and views of privacy and mobile phone location data. Hence an inductive approach was chosen to explore the subject area through the participants' eyes. The decision to use grounded theory methodology was further supported by the lack of existing theory regarding mobile phone location data and privacy.

Grounded theory scholars have different opinions about the most suitable time at which to review the literature. Glaser (1978) advocates waiting to conduct the literature review until initial findings have been made in order to not influence the researcher with preconceived ideas. This study has followed the advice of Charmaz (2006) and has carried out an initial review of the literature before the first data collection in form of the pilot study took place. The reason for an early review of literature was, on the one hand, to learn whether any similar research had already been conducted in this area and, on the other hand, to satisfy requirements of the university's research committee for the research proposal.

### 3.1.3 Substantive and formal theory

A grounded theory is directly related to the data from which it has been generated; it is therefore *grounded* in the data. Two types of theory are distinguished: substantive and formal theory.

Substantive theories provide a theoretical interpretation or explanation for a *particular* area, in other words this type of theory is used to explain and manage problems in a specific setting. Formal theories, however, are more abstract and provide a theoretical dealing of a *generic issue* which can be applied to a wider range of disciplinary concerns and problems (Strauss and Corbin, 1998). For example, a substantive theory can be about a limited area such as family relationships or professional education while a formal theory might deal with the construction of culture or the development of ideologies or stigma (Charmaz, 2006; Glaser, 1994). Charmaz (2006) suggests combining and conceptualising the results from several substantive grounded theories to develop a more general formal theory. Each substantive theory can help to refine the formal theory, in other words, a formal theory can relate to or 'cut across' several substantive ones. Charmaz (2006) points out that most grounded theories are substantive theories as they focus on particular problems in a specific, substantive area.

This study has developed a substantive theory as the collection of data and their interpretation focus on the explanation of a particular area, that is the relationship between mobile phone location data and individuals' perceptions of privacy in the UK. This PhD thesis does not provide the scope to raise the very specific, substantive theory to a formal theory that would be generalisable across a wider area, such as additional types of digital data or mobile phone users from other cultural backgrounds.

### 3.1.4 Memo writing

The process of coding and developing categories is supported by writing memos. Memos are a set of notes, that kept continuously, support the researcher by providing a record of thoughts and ideas. Memos enable the researcher to reflect on the interviews and given codes to enter into a dialogue about the collected data. Initial thoughts are of high relevance as they often spark the best ideas. Hence it is important to write the memo immediately when reading and coding the interview. At later stages in the research process, initial thoughts are represented through



memos and can be revisited, reflected upon and considered for the overall analysis. Memos can be used to ask questions, philosophise about potential meanings of interviewee's statements and compare concepts identified in interview transcripts to each other and to the literature.

### **3.1.5 Criteria for grounded theory studies**

Charmaz (2006) gives the following criteria that grounded theory studies should aim for. She highlights that a combination of credibility and originality enhances the other two criteria resonance and usefulness.

#### **a) Credibility**

- Are there strong links between gathered data and argument?
- Are data sufficient to merit claims
- Do categories offer a wide range of empirical observations?
- Has the research provided enough evidence for the researcher's claims to allow the reader to form an independent assessment?

#### **b) Originality**

- Do the categories offer new insights?
- What is the social and theoretical significance of this work?
- How does grounded theory challenge, extend, refine current ideas, concepts and practices?

#### **c) Resonance**

- Do categories portray fullness of the studied experience?
- Does the GT make sense to the participants?
- Does analysis offer them deeper insights about their lives and worlds?

#### **d) Usefulness**

- Can the analysis spark further research in other substantive areas?
- How does the work contribute to knowledge
- Does the analysis offer interpretations that people can use in their everyday lives/worlds?

(Charmaz, 2006, p. 182)

Chapter 6, Conclusions and Recommendations, revisits these four criteria of grounded theory research and addresses how each criterion has been met by this study (section 6.4).

### 3.1.6 Objectivist and constructivist approaches to GTM

Grounded theory methodology has evolved since its inception in the 1960s in the United States. Particularly, the writings of Glaser (such as 1967, 1978), Strauss and Corbin (such as 1990, 1998) and Charmaz (e.g. 2000, 2006) are seen as influential for the development of GTM. The original work of Glaser and Strauss from 1967, 'The *Discovery of Grounded Theory*', suggests that the researcher should start collecting data with a 'blank mind', meaning without reviewing the existing literature in order to carry out a truly inductive study. Consequently, theory is built from observation and based on the understanding that the theory is already contained in the data and only needs to be dug up or '*discovered*', as Glaser and Strauss' book title (1967) suggests. This perspective assumes that every individual will see and understand the data from the same point of view, making the same observations and therefore will come to similar conclusions. The researcher should take a passive stance and 'let the data emerge', which can be seen as a characteristic of an objectivist or positivist paradigm (Bryant, 2003; Charmaz, 2000). The alternative view in social sciences is the so-called constructivist or interpretivist view.

Constructivist grounded theory methodology is for example advocated by Kathy Charmaz in her book 'Constructing Grounded Theory' (Charmaz, 2006). This strand of grounded theory methodology emphasises the research participants' experience and how *they* construct their view of reality. Knowledge, and hence the grounded theory, are constructed by both researcher and research participant and aim at interpreting the empirical evidence within the research context.

A divergence between the two authors of 'The Discovery of Grounded Theory' occurred in the 1980s, after which Glaser postulated his understanding of grounded theory methodology (Glaser, 1992). The late Strauss, however, together with his co-author Juliet Corbin, developed a different perspective on grounded theory methodology (Corbin and Strauss, 1990; Strauss and Corbin, 1998). The major difference between Glaser's and Strauss' views is Glaser's stance that "data emerges" and thus presents the same picture of facts to every researcher in form of some objective truth. Strauss' viewpoint on the other hand stresses that a researcher has to *actively* obtain theory from data. Hence, it is likely and even expected that each researcher will place the focus on different aspects of the collected data depending on their background, beliefs and values. Stern (in Morse ed., 1994) calls the former the Glaserian School and the latter Straussian School, as

both approaches to GTM differ in process and product and in her view represent completely different methods.

Charmaz (2000) argues that both Glaser's and Strauss and Corbin's approaches to GTM assume an objective external reality and hence take a positivist and objectivist stance. She, on the contrary, advocates a constructivist approach to GTM that assumes multiple social realities. Charmaz does not support the view that theories are discovered but believes that the studied world needs to be portrayed in an interpretive way because interviewee and researcher embark together on the process of constructing reality (Charmaz, 2006).

This study has been inspired and guided by Strauss and Corbin's and Charmaz' interpretation of grounded theory. The researcher disagrees with Glaser's stance that reality is objective and neutral, particularly regarding the intangible and personal subject area of privacy.

### **3.1.7 Limitations of the grounded theory methodology**

Grounded theory methodology has limitations like any other research methodology. Some point out that GTM is very complex and time-consuming due to the tedious coding process and memo writing as part of the analysis (Bartlett and Payne in McKenzie et al., 1997). This study has dealt with the lengthy process of coding by using specialised software to help speed up organisation and analysis of data.

Others name as a limitation that the use of GTM to explain, predict a phenomenon or to build a theory is a very subjective process, which relies heavily on a researcher's abilities. This study has followed the methodological guidance of Charmaz (2006) and Strauss and Corbin (1998) to gather and analyse the interview data. In addition, findings from the survey and pilot study were used to strengthen those findings based on interview data with the aim of fulfilling the criteria for GTM as described in sections 3.1.2 and 3.1.5. Many studies make use of the term grounded theory inappropriately and Bryant (2002) points out that the flexibility of the method can be used to provide a justification for studies lacking in methodological strength. Stern (in Morse ed., 1994) criticises some researchers for mixing methods such as ethnography and phenomenology and then use the label grounded theory to explain the analysis of their research findings.

The following three sections describe the data collection process for this study. Details about the findings and their analysis will be provided in the following two chapters, Chapter 4 and Chapter 5.

### **3.2 Pilot study: Mobile phone location tracking**

The pilot study explored technical aspects of mobile phone location data and participants' perceptions of privacy regarding this subject area. Four participants' mobile phones were registered with an internet-based service provider that offered identification of the phones' geographical location 24 hours a day.

The aims of the pilot study were twofold: firstly to explore the technical side of location tracking by locating the participants with help of the commercial tracking service provider. Secondly, to learn about the participants' awareness of and attitudes towards location tracking in relation to privacy. Potential disadvantages of pilot studies were accepted and considered. These consisted primarily of the small sample size which could not provide representative results and the danger of making inaccurate predictions or assumption based on the pilot data.

#### **3.2.1 Sample and ethical considerations**

The participants were selected based on the researcher's judgement. This was due to the potentially invasive nature of the research design and the explorative characteristics of the pilot study. Two male and two female participants were chosen for the study and it was ensured that all four UK GSM mobile phone networks (O2, Orange, Vodafone and T-Mobile) were represented. It was a requirement of this study that all phones were operating on the GSM network in order to be locatable by any of the commercial mobile phone tracking providers. Out of the several internet-based companies offering similar mobile phone tracking services, one company was chosen for reasons of convenience and pricing. This service provider offered the registration of five mobile phones for £5 plus VAT per month. The phones of four participants and the researcher's phone were registered. Due to the small sample size and the limited age range of the participants (25 to 35 years), the study cannot be seen as representative for the overall population. The time span of four weeks was seen as a sufficient to generate a number of tracking results and to enable the participants to experience location data in their daily routines.

This study was guided by the ethical principles on research with human participants set out by Leeds Metropolitan University (Leeds Metropolitan University, 2006). Particularly the researcher's ability to determine a participants' geographical location 24 hours per day over a period of four weeks was perceived as the most profound ethical implication. The only possibility for a participant to not have his or her mobile

phone's geographical location identified would have been to either switch the phone off or to terminate the participation in the research project. For these reasons, the participants were informed about the study in great detail before asking for their consent to take part. Aims and objectives of the pilot project, together with details about the data collection process were explained to the participants. The participants were informed that they could withdraw from the study at any time without questions being asked. All data collected was anonymised by replacing the participants' names with ascending code numbers (P116 - P119) in the order of the initial interviews. Appendix A provides the email that has been sent out to the participants before giving their definitive agreement to take part in the pilot study.

### 3.2.2 Data collection procedures

After all four participants agreed to take part in the study, their mobile phone numbers were registered with the tracking service. The overall data collection had two main aims. For one part, to compare the geographical data that was generated by the location tracks to the participants' actual location. The second part aimed at exploring participants' perceptions, feelings and opinions regarding privacy, mobile technologies and particularly location data before and after the study. The data collection process is shown in Figure 3.3, below. 'Open track' denotes a location request known to the participant at the time of tracking and 'covert track' indicates a location request without the participant's knowledge.

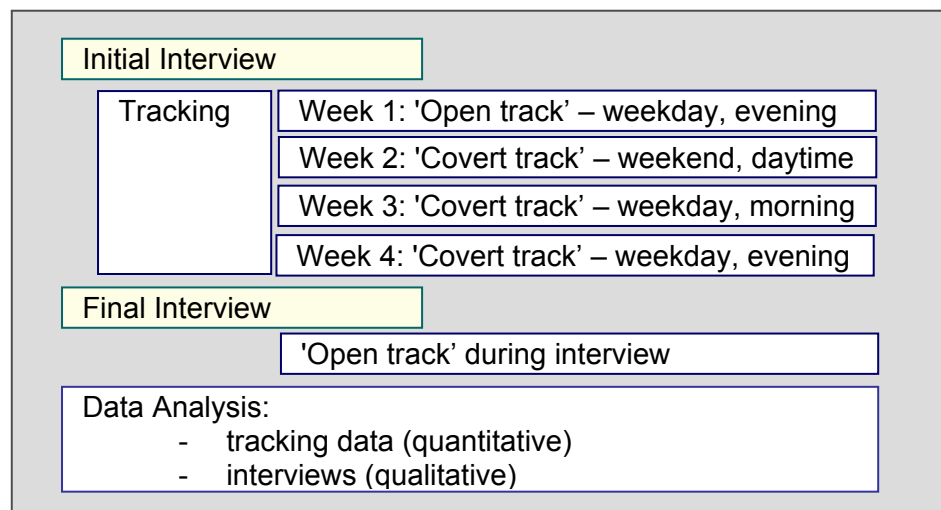


Figure 3.3: Data collection process for pilot study

### 3.2.2.1 Part one of pilot study: Location tracking of participants' mobile phones

A weekly location request was submitted for each participant over a period of four weeks and the results were stored electronically by the researcher. The participants were not contacted during this time to avoid influencing their tracking experience. The research participants could not know when a request for their mobile phone's geographical location was submitted to the tracking service provider.

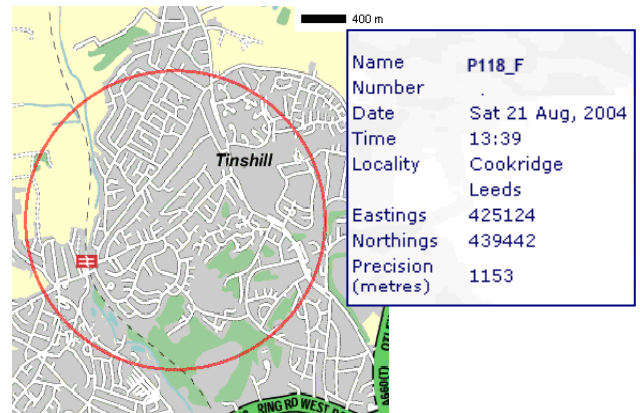


Figure 3.4: Example of location track

Figure 3.4 shows an example of a location request in form of a map depicting the participants' geographical location at the time of tracking. In order to be able to compare the participants' location provided by the service with their actual location, it was suggested to the participants to keep a location diary. To make it easier for the participants to recall their location at the time of tracking, the phones' locations were requested at different weekdays and times (see Figure 3.3). Immediately before the first location request only, all participants were contacted on their mobile phone. This phone call helped to familiarise them with the study and to obtain a first valid actual location. The participants were unaware of the three subsequent tracking requests. Hence, the participants' actual location had to be compared to the one supplied by the tracking service based on diary entries and memory in the final interview. A fifth track was conducted in the researcher's office as part of the final interview to demonstrate the service to the participants and to be able to compare the accuracy of tracking results between the four different mobile phone service providers.

Mobile phone tracking served two purposes, which tied in with the overall aims of this study. Firstly, mobile phone location tracking allowed comparison of the participants' location derived from location data with their actual whereabouts at time of tracking. The main purpose for this was to evaluate the service regarding reliability, validity and accuracy. Results from the location requests aimed at bringing location data into context with claims of privacy invasion through retention of location data. The researcher believes that police and emergency services have access to more detailed geographical data than a commercial tracking service provider. This assumption is based on media reports. However, using a commercial

service provider was seen as the only feasible way to get access to location data and to visualise this hidden aspect of mobile communication. The second purpose of location tracking was to raise the phone users' awareness of this type of data. The 'visualisation' of mobile phone location data by using a tracking provider proved to be a useful tool to enrich the qualitative data collection both for the participants and the researcher.

### 3.2.2.2 Part two of pilot study: Interviews

Respondents' perceptions and beliefs are at the heart of qualitative research and this was the main motivation for complementing the location requests with interviews. Semi-structured interviews were conducted with all four participants shortly before and after the four week tracking period to learn about their opinions and perceptions regarding mobile phone location data and privacy. The interview questions were of exploratory nature due to the small scale of this study and early stage in the overall research project. The questions were designed to identify patterns and common themes in the participants' accounts and sought to identify the meaning of privacy in relation to mobile phone location data in the participants' everyday lives.

During the interviews it was important not to restrain the participants but to give them time to talk about how they understood and described their experience of mobile phone tracking. This was particularly important as privacy is such an elusive concept, that can be perceived differently by different people. Figure 3.5, below, gives an overview of the interview themes and Appendix C - Interview questions for pilot study, provides a detailed list of interview questions.

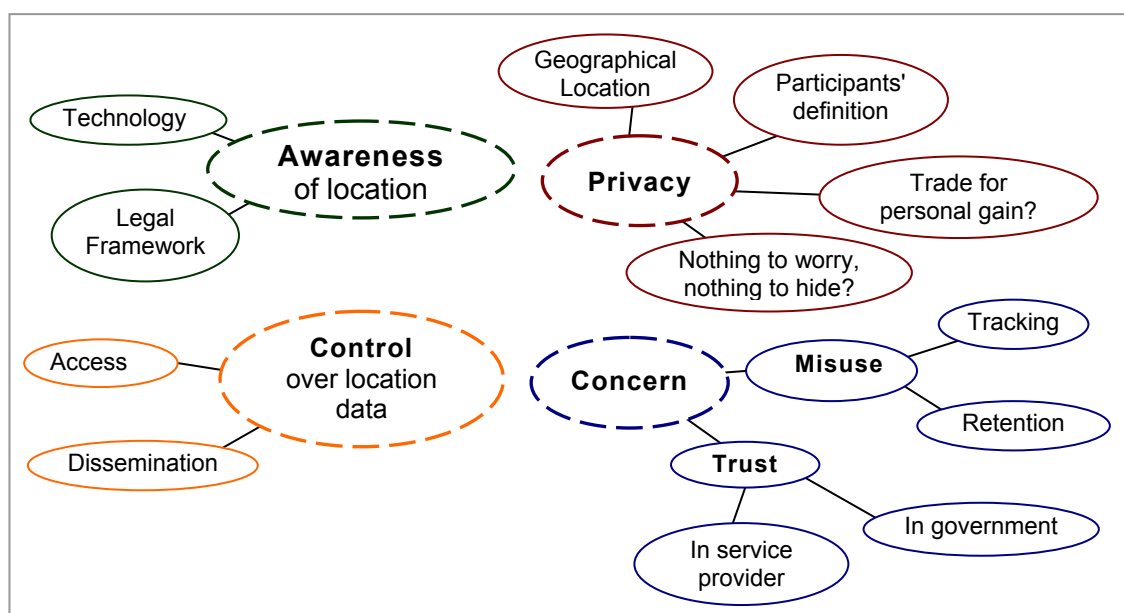


Figure 3.5: Graphical representation of interview themes

The interviews that took place prior to the tracking period focused on participants' awareness and knowledge of location data, as well as their descriptions and thoughts about privacy. Responses to these initial interviews have helped to develop the questions for the second set of pilot study interviews which was conducted soon after the four week tracking period. The final interviews aimed at learning about participants' experiences of taking part in the tracking study and any thoughts or potential concerns regarding privacy.

The pilot study proved useful to familiarise the researcher and the participants with the concept of location data. The pilot interviews provided valuable leads to be pursued later on in the research project in more in-depth interviews.

### **3.3 Interviews**

The interviews for this study were conducted in three phases: the first phase consisted of eight explorative interviews for the pilot study and the second phase consisted of five interviews, after which the survey was initiated. The reasons for conducting the survey at this point was for one part to allow time to reflect on how to approach the next set of interviews and for the other part to learn about individuals' awareness and opinions regarding location data on a greater scale than it would have been possible with interviews. The third and last interview phase took place after the paper-based survey was completed.

Interviews can provide insights that are not available to researchers working with large survey samples and are known to be the most suitable approach when seeking rich data illuminating individuals' experiences and attitudes. Drawbacks are that interviews are very time-consuming to conduct and analyse. The interview questions for interview phases two and three were based on findings from the pilot study. The questions were asked in as non-directive a manner as possible to meet the study's principal aim of learning about the interviewees' perceptions. The data collection and analysis for this project took place in alternating sequences and was guided by the grounded theory methodology. This meant that the interviews were transcribed and coded immediately after they took place. Hence, initial findings from interview coding could help to shape the questions for subsequent interviews.



### **3.3.1 Sample design and ethical considerations**

For the second and third interview phase, the participants were initially contacted informally, which was followed up by an email explaining the study's aims and the interview procedure. The email ensured participants about anonymity and confidentiality of data collected and informed them that the interview was recorded for transcription (see Appendix B - Email sent out prior to interviews). The interviews were recorded with a digital voice recorder and the files transferred to a PC for transcription. When transcribing the interviews, participants' names were replaced with code numbers. Participants of the pilot study were assigned the codes P116\_F, P117\_M, P118\_F and P119\_M. "M" indicates males and "F" stands for female. The codes for respondents of the second and third interview phase followed the same system, starting with number P131\_F. In addition to the eight pilot study interviews, ten more in-depth interviews were conducted which were between 30 and 90 minutes in duration.

Respondents were selected based on initial findings. The pilot study had for example highlighted different types of mobile phone users and it seemed as if different types of users utilise their phone settings in different ways to regulate how others can get in touch with them (see section 4.2.8). Hence it was sought to interview male and female mobile phone users from different age groups and with different mobile phone contracts. A list of respondents, including demographic characteristics, can be found in Appendix F - List of Interviewees.

### **3.3.2 Interview preparation**

The information sent out before the interview included some technical and legal background about mobile phone location data. The email did not make any references to potential impacts of data retention on civil liberties or privacy so as not to influence the respondents' views. The aim of sending out information prior to the interviews was to familiarise the participants with the subject area. This seemed necessary because early survey findings showed that the majority of people had either not heard of location data or were not sure about the details.

### 3.3.3 Development of interview questions

The interview questions were based on initial findings from pilot study and survey and the following themes were addressed (see Table 3.1, below). Appendices D and E provide a detailed list of interview questions.

Table 3.1: Themes of interview questions for second and third interview phases

Theme	Notes about question
Mobile Phone	Mobile phone usage and habits in everyday life situations.
Privacy	An open question as to what privacy means to the interviewee. Deliberately trying to avoid mentioning any categories of privacy known from the literature.
CCTV, Loyalty Cards	Do interviewees perceive these as privacy invasive?
Phone as tracking device	Referring to technical and legal facts of mobile phone location data that were sent out prior to the interview. - Is the interviewee aware of location data before this study? - Are any issues or concerns raised?
Government & mobile phone service provider	- What is the participants' opinion towards storage of communications data (and its availability to government)? - What else is mentioned in this context? Regarding government, regarding data retention?
In addition to these themes, participants mentioned terrorist attacks, current politics and ID cards.	

The interview questions changed and improved over time, influenced by codes and categories developed for previous interviews. For this reason, two sets of interview questions are shown in the appendices: Appendix D - Interview questions March 2005 and Appendix E - Interview questions December 2005. The second set of questions was guided by Kathy Charmaz' approach to GTM, particularly her chapter on how to phrase interview questions to allow respondents to express their views without constraints (Charmaz, 2006). For example the question 'Have you ever heard of mobile phone location data before taking part in this study?' was changed to 'What do you know about mobile phone location data?'. A question mentioning consumer loyalty cards and the storage of personal information by commercial companies was removed after the second interview phase. The reason for this was

that initial interview findings and the survey showed that respondents did not perceive this area as related to privacy.

#### **3.3.4 Use of software for data management and analysis**

Weitzman (2000) advocates the following benefits of using software in research projects: writing up, editing, coding, storage, search and retrieval, data 'linking', memoing, content analysis, data display and graphic mapping. However, he warns of 'false hopes and fears', pointing out that no software will be able to actually carry out the analysis process for the researcher. Software can support the research process but ideas and intellectual efforts have to come from the human being conducting the research and analysis.

For this study a word processor, Microsoft Word, and the qualitative analysis software NVivo were used to support the analysis and to help manage the interview data. The computer was utilised as efficiently as possible to reduce the amount of time spent on organising data and findings, to increase the speed of tiresome tasks, resorting the material and redefining codes. Qualitative analysis software facilitated following potentially promising analytic routes but also enabled these routes to be discontinued with ease. Dynamic and real-time representation of the findings considerably assisted reflection on data and connections between the data.

##### **3.3.4.1 Transcribing interviews and importing into NVivo**

The use of the NVivo software has significantly facilitated the process of organising, re-arranging and managing the considerable amount of data. For example, after coding the interviews in NVivo, all passages assigned to a specific code could be viewed on screen and printed. In the same manner, searches for specific text strings could be conducted across all interviews and relevant paragraphs containing the search string could be compared on screen or printed. The interview transcripts were formatted in a particular way in Microsoft Word to facilitate importing the transcripts into NVivo. This meant for example, that the interview questions were assigned a 'Heading 1' format. When importing the transcript into NVivo, this resulted in the questions being displayed in the content panel in the NVivo explorer. Hence when selecting a question, it was possible to jump to this section in the interview transcript. Furthermore, meaningful information *about* the interview was placed into the first two paragraphs of each transcript. This enabled this information to be automatically put into the properties of the interview document when importing the interview into NVivo.

The appropriate formatting of the interview transcripts from the beginning helped to organise the data efficiently and thus facilitated the analysis of the interviews. In addition, by using the NVivo software much tedious and time-consuming work for managing and resorting the data could be avoided which freed time for more meaningful tasks, such as analysis and interviewing. Different sets of interviews could be assigned with different colours in NVivo for easy distinction.

An advantageous feature of NVivo is that the software keeps a log of all data that has been entered, which means that all codes and memos are automatically assigned a date and time stamp. This feature helped to trace the development of codes. After coding the interviews in NVivo, all passages assigned to a specific code could be viewed on screen and printed. In the same manner, searches for specific text strings could be conducted across all interview sets.

### **3.3.5 Interview coding**

The first set of interview transcripts were coded with the help of the qualitative analysis software NVivo, which however did not prove to be the most beneficial approach (see section 4.1.6.1 Some initial concepts not confirmed). After setting aside several months for the survey, a fresh and more suitable approach to the analysis of interviews was adopted. This time interview codes were not devised on screen with help of software but with pen and paper. Following the GTM guidance on coding (see section 3.1.1.1), the researcher worked through each of the transcripts and used line-by-line coding to take note of themes and phenomena on the margins. The codes were not devised strictly microscopically and some more abstract categories came into view; some codes were very close to the interviewee's accounts and others more abstract or conceptual. Keywords and phrases were noted on differently coloured post-it notes and stuck onto a blank A2 flip chart sheet (see Appendix I). The post-it-notes were arranged in a logical order on the paper sheet. As more and more interviews were coded, this sheet started looking less like a random collection of labelled post-it notes but more like a brain storm map or a tree where branches of thought grew from certain categories. Memos were written throughout this exercise to keep track of thoughts and ideas regarding the data analysis.

This system of creating codes, combined with reflection was maintained for coding all interviews. At the end of this coding exercise, three A2 sheets were covered with post-it notes containing categories and codes. In addition, a matrix had been

devised, which held codes, properties and dimensions, as well as some comments and quotations. This list of codes was revised continuously as more interviews were coded. The codes were modified and verified by being applied to further interview transcripts but stayed alike for the most part. Subsequently, the codes were keyed into the NVivo software to allow searching the interviews, re-sorting of material and consistent redefining of codes in order to support the analysis process.

Figure 3.6, below, shows a summary of the coding process. More details about this analysis process are discussed in the subsequent findings and analysis chapters.

Appendix J provides a summary of the development of codes.

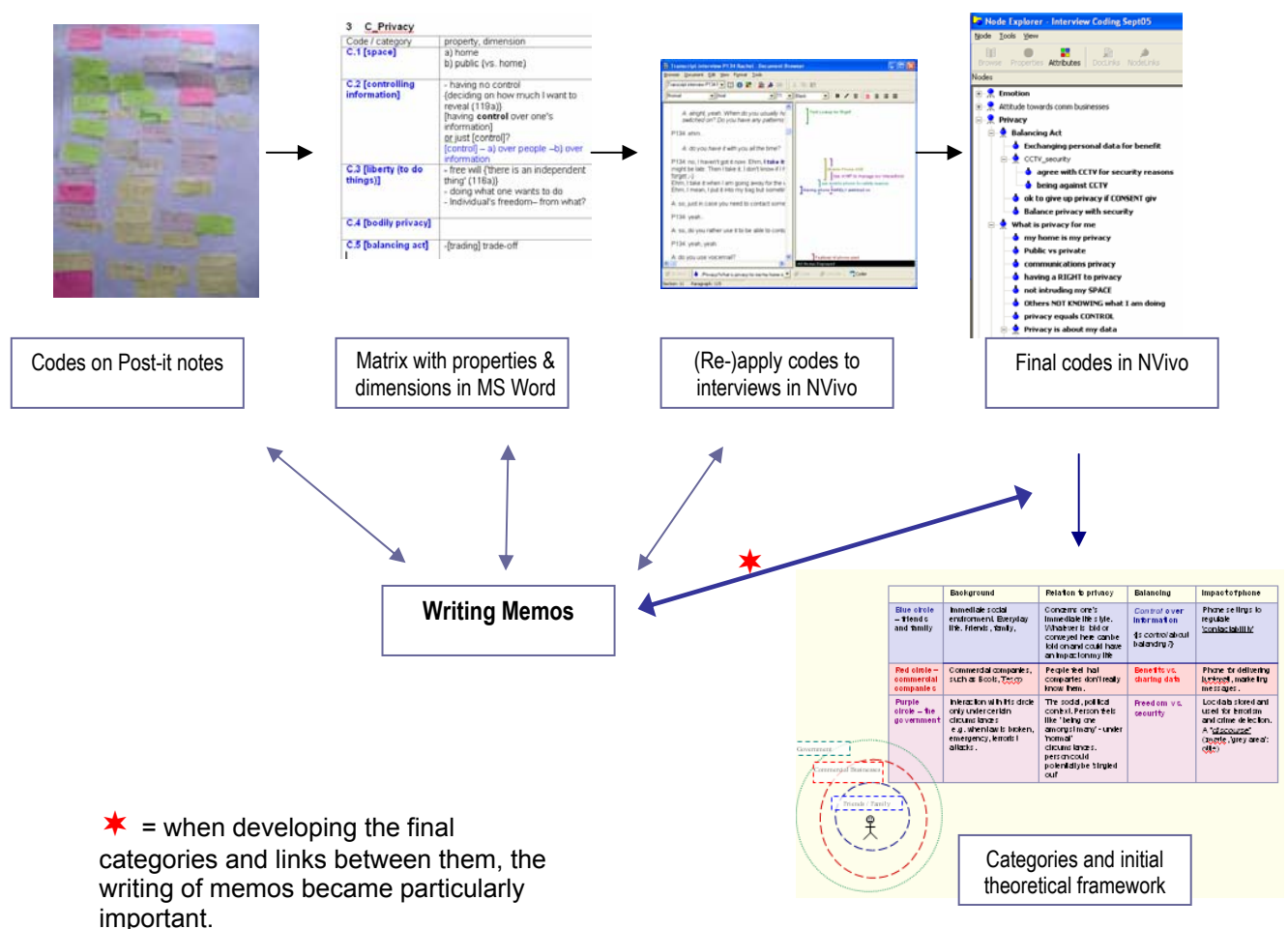


Figure 3.6: Graphical representation of coding process

### **3.3.6 Memoing to develop and clarify categories**

The extended coding process, as described in the previous sections, has facilitated reflections on codes and categories, which were captured by writing memos (for examples of memos see next chapter, section 4.2.2 Situational map as a memo to consider wider social context). The memos were consulted when establishing links between categories and setting up the initial theoretical framework. The following chapter details how initial codes have developed into focused codes and finally to abstract categories (section 4.2 Development of categories based on interview data). The writing of memos ('memoing') was particularly useful as this helped the researcher to keep a note of thoughts without the pressures of having to immediately determine how ideas fitted within the overall research findings and analysis. Memoing allowed the freedom to jot down ideas so that these could be sorted, categorised or discarded at a later point in time. The writing and reflecting on memos has been a crucial step in the development of the final categories based on initial and focused codes. Some examples of memos and how they have influenced the analysis process are shown in the next chapter (for example section 4.2.4 Memo 'Balancing security and privacy').

### **3.4 Survey**

A survey was conducted in form of a questionnaire and distributed after the pilot study and some further interviews took place. The survey complemented findings from the interviews and helped to obtain a better idea of individuals' opinions towards privacy and mobile phone location data. Additionally, the survey highlighted interesting cases which could be approached in further data collection stages and thus supported the purposive sampling approach of interviews. It could be understood as a weak point of questionnaire-based surveys that they only capture surface opinions, seeing that respondents will not necessarily report their beliefs and attitudes accurately. These are easier to identify in interviews, as also prompts can be used. In addition, the use of mainly closed questions in a questionnaire merely allows respondents to choose between a limited number of responses (Robson, 2002). However, some of the respondents of this survey have overcome this limitation by making comments on the questionnaire's margins to express further opinions and thoughts. A significant advantage of using a survey for this study was to collect larger amounts of data in a shorter time scale than would have been possible with interviews.

### **3.4.1 Sample design and ethical considerations**

It is hardly ever possible to survey the entire population to be studied and for this reason sampling techniques need to be employed. A representative sample produces results which can be used to formulate generalisations. However, this can only be achieved by using probability sampling where the likelihood of the sampling unit - in this case an individual - to be included in the sample is known. The sampling population for this research project consisted of British residents older than 15 years, as the survey was not designed to explore the views of children. A representative sample was not chosen as it had not been possible to obtain a large enough random sample due to time and monetary constraints. For these reasons purposive sampling, a form of non-probability sampling, was used. Participants were not selected randomly but judged to be of interest to the researcher, which should not be understood as a limitation since the survey was designed as an explorative study. In addition, this approach tied in with the sampling procedures common in studies using grounded theory methodology (see section 3.1.1).

The questionnaire was distributed to individuals of different age groups, diverse professional backgrounds and users with different types of mobile phone contracts. The sample size of 477 respondents was deemed to be sufficient for an explorative survey and provided an overview of participants' opinions at the time of data collection. Statistical measures such as confidence intervals could not be produced as the sample was not a probability sample and the survey did not have a defined population. Ethical guidelines were carefully followed when collecting the survey data: all data were treated confidentially, which was explained to the respondents together with the aims of the study. Respondents were free to fill in their name on the last page of the questionnaire if they were interested in taking part in a follow-up interview. The section revealing individuals' names could be handed in separately from the questionnaire responses to ensure anonymity.

### **3.4.2 Pilot testing of survey**

The questionnaire was thoroughly pilot tested before dissemination to uncover flaws and potential causes of confusion, such as misleading questions that could potentially result in invalid responses. For the pilot test 15 questionnaires were distributed to friends and other well-known contacts to verify feasibility and compliance with objectives set out by the overall study. The following steps were undertaken to pilot test the questionnaire:

#### *1) Develop questions*

The survey questions were based on interview questions from the pilot study. Five questionnaires were distributed to friends and colleagues, who were informed that the questions were developed for a survey. Their comments and feedback was used to modify some of the questions.

#### *2) Pilot testing*

Ten revised questionnaires were distributed to other friends and colleagues and their feedback helped to develop the final version. The number of questions was reduced from 23 to 20 and the layout was condensed, so that the number of pages could be reduced to four instead of seven pages.

#### *3) Pilot analysis*

Frequency tables for the responses to each question were generated to obtain an early impression of the results. This trial analysis ensured that the survey would fulfil the survey's aims and yield data valuable to the overall study.

### **3.4.3 Changes undertaken after successful pilot testing**

After the initial analysis of pilot data, some changes in questionnaire design and coding were undertaken. The most significant change was to amend the answer choices of question 15 "Have you heard of mobile phone location data before taking part in this study?" to:

- I have heard of location data from the media and other sources, but I am not exactly sure about the details
- Yes, I know what location data is and how it works. I could explain it to a friend
- No, I have never heard of location data

The reason for this change was that the previous answer choices were not meaningful enough, as the questionnaire merely asked whether people had 'heard of' the data and if yes from which sources. However, it was acknowledged that it would be more important to learn how much respondents would consider themselves to know about the subject. The drawback of these answer choices was



that respondents had to use their own judgement. Nevertheless, it was not seen as appropriate to examine participants' knowledge, since it was more important to gain respondents' trust to enter into a dialogue. An additional change regarding question 15 concerned the layout. This significant question was moved to the bottom of the page, so that the box providing explanations about location data would only appear on the subsequent page of the questionnaire. Hence, this forced respondents to make a judgement of their awareness of location data before being presented with further background information on the subject area. Appendix L contains the final version of the questionnaire.

#### **3.4.4 Distribution of questionnaires**

The survey was disseminated in two formats, paper-based and online via the internet; the majority of questionnaires were distributed paper-based. Initially the survey was given to individuals working and studying at Leeds Metropolitan University, ranging from respondents with an academic background (lecturers, post- and undergraduate students) to administrative, cleaning and security staff. Friends, personal contacts from various leisure activities and colleagues from part-time jobs were also invited to fill in the questionnaire. In addition, acquaintances working in different professions (landscape architect, engineer, manager in a bus company) were asked to distribute approximately 10 questionnaires each at their workplace, sports club or family. This distribution method could be seen as a form of snowball sampling, which ensured access to social groups beyond the researcher's immediate social circle and made it possible to collect data across a diverse population. A known limitation of the sample for this study is that unemployed, unskilled or manual workers were only included marginally.

It was acknowledged that researcher's bias could have occurred when selecting participants. Response rate and quality of survey data was very high due to the questionnaire distribution on a personal basis. An additional benefit of this approach was that respondents could receive help with completing the questionnaire if necessary. Only a few people refused to fill in the questionnaire, mainly due to time constraints, and one person did not want to participate as she perceived questionnaires as an invasion of her privacy. It is acknowledged that due to the chosen non-probability sampling method, the number of non-respondents and their reasons for not taking part in the study could not be known.

#### **3.4.5 Online survey**

In addition to the paper-based version, an online version of the questionnaire was published on the internet on a website developed by the researcher (<http://www.locationprivacy.org>). The survey questions were identical for both distribution media so that the results could be pooled. An online survey was seen as a convenient way to collect additional data with minimum involvement of the researcher in data collection and analysis. Results of the survey were automatically gathered in a database and could be downloaded in form of tabular data which then could be imported into Ms Excel. Advantages of electronic surveys include being able to direct respondents to particular sections of the questionnaire, depending on the way they have answered previous questions. Respondents can be automatically prompted when they provide an invalid response, such as selecting several tick boxes when only one should be marked. However, electronic surveys also have distinctive technological, demographic and response characteristics which were to be taken into consideration for this study. For example, only individuals with computer and internet access were able to fill in the survey. To account for these differences data from the electronic survey was stored and analysed separately from the data collected by the paper-based version. However, a comparison between the responses from both distribution media showed that results were similar (see section 4.3.1).

#### **3.4.6 Email as distribution medium**

Email was planned to be used as a distribution medium for the survey but discarded due to particular problems to this approach. Two alternative methods for using email to distribute surveys were considered; firstly, to send the survey as an attached word document, and secondly to include the questions in the email body. However, for the first option it would have been necessary for the respondent to download the attachment, complete and save it and then re-attach it to the email to send it back to the researcher. This is a rather cumbersome process and if any of these steps is not followed correctly, the survey is returned blank. The second option would not have allowed the researcher to control how the questions are displayed, as this depends on the respondent's email programme. For these reasons the use of email was restricted to distributing the questionnaire's website link and to contact respondents.

### 3.4.7 Question design and coding

Questions were worded carefully and avoided long and ambiguous, leading, biased questions, as well as jargon. The following question types of questions were used:

#### Numerical rating

A Likert scale from 1 - 6 was used. The reason for using an even number of options was to not allow non-committal answers. In the data analysis, each statement was assigned a separate variable.

8. How important is your phone to you on a scale from 1 to 6?														
<b>My phone is VERY IMPORTANT to me.</b> It would be difficult to live without it.	<table style="margin: auto;"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td> </tr> <tr> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> </table>	1	2	3	4	5	6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<b>NOT IMPORTANT at all to me.</b> I could live without it.
1	2	3	4	5	6									
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>									

#### Multiple choice

Require respondents to choose just *one* response from a list of alternatives. Only one variable per question was necessary in the data analysis.

14. Giving up some privacy is necessary to fight terrorism and crime		
Agree <input type="checkbox"/>	Disagree <input type="checkbox"/>	Don't know <input type="checkbox"/>

#### Check lists

List a set of items of which respondents select those that apply. Each statement was assigned a separate variable in the data analysis.

18. In your opinion, what would be good uses for location data? <span style="color: blue;">[Select all that are appropriate]</span>			
Solving crime	<input type="checkbox"/>	Employers to track their employees	<input type="checkbox"/>
Preventing acts of terrorism	<input type="checkbox"/>	Emergency services	
For parents to track their children	<input type="checkbox"/>	(e.g. rescue, ambulance, fire brigade)	<input type="checkbox"/>

#### Open questions

Open questions allow respondents to formulate their own statements and can lead to unexpected responses. Nevertheless, open questions are more difficult to analyse and take respondents longer to respond. For this reason only one question of this type was included.

9. What comes to mind when you think about the term 'privacy'? <span style="color: blue;">[Keywords are fine]</span>  -----  -----
--

### Non-committal responses

The survey questions raised many issues which respondents might not have thought about before taking part in the study or might not hold an opinion. Hence, alternative responses such as 'don't know' or 'no opinion' were provided where appropriate. It would create false and unreliable answers to force the respondents to express an opinion. Some see the danger in providing these options that some respondents might select them out of laziness (Converse and Presser, 1986 in DeVaus, 2002). Therefore, alternative responses were always the last one in a list of options (see Appendix L).

### **3.4.8 Coding missing responses**

In every survey there will be missing data occurring for different reasons, which needs to be recorded and coded in a similar way to valid responses. In this study, different codes were given to different types of missing data. The following types of non-responses were identified:

- 1) The respondent has not responded to a question. The reason might have been by choice or simple overlooking of the question. The code given was '9'.
- 2) The respondent was not required to answer this question. After responding to the question 'Do you own a mobile phone?' individuals who did not own a phone would proceed straight to Section B 'Privacy'. The missing responses in Section A 'Mobile Phone' was coded with an '8' to distinguish those from other missed responses (see above).
- 3) Response was invalid, e.g. where only one answer was required but the respondent ticked several responses. However, this was only the case with one questionnaire and therefore in this case was assigned a '9'.

#### **3.4.9 Data entry**

The survey data was firstly entered into an Excel spreadsheet and then imported into the statistical analysis software SPSS (Statistical Package for Social Scientists). The reason for using two different software packages was that Excel would automatically update graphs showing the frequencies of responses as survey data was entered. This gave early impressions and ideas about the data and helped reflecting on findings at an early stage. Another reason for entering the data first into Excel was that there was no character limit for variables, which was convenient for open question and comments. Afterwards, data was imported into SPSS to facilitate advanced analysis of data, such as correlations between variables. In addition, SPSS could help to prevent false entries, as data ranges could be set, so that for example no number greater than 6 could be entered for a question.

#### **3.5 Analysis of findings**

Data collected by pilot study and survey were compared to the grounded theory categories identified in the interviews in order to support the analysis of findings. This process of triangulation between qualitative and quantitative data was used to confirm and validate the findings.

The interplay between induction and deduction, in other words between data collection and interpretation, was another process of validation of findings but also part of the theory development (see section 3.1.1). In what is known as a process of abduction, the interpretation of observed data to the best explanation helps to form a tentative theory, which then needs to be confirmed or disconfirmed with help of further data collections and analysis. This procedure is repeated until the best, the most plausible interpretation of data is found (Charmaz, 2006; Haig, 1995). The findings from empirical data were compared to the reviewed literature, as described in Chapter 5, which then lead to conclusions and recommendations (Chapter 6).

### **3.6 Chapter summary and conclusion**

This chapter has introduced and discussed the choice of grounded theory methodology as a suitable research methodology for this study. The continuous generation and long-term retention of mobile communications data are inherently related to the every day transactions of every mobile phone user. The views of mobile phone users are necessary to be captured as information about their lives is at the centre of debates around data retention.

A grounded theory has been developed to provide an explanation for the phenomenon under study: the relationship between mobile phone location data and individuals' perceptions of privacy in the UK. The theory can be categorised as 'substantive' as opposing to a 'formal' theory, since the collection of data and their interpretation focus on the explanation of a particular area. The chapter has in great detail explained each of the three data collection phases, including sampling and ethical considerations. A mobile phone location tracking pilot study, a survey and a series of interviews have been conducted to learn about individuals' perceptions of privacy in relation to location data. The iterative cycle of data collection and analysis is an essential element of GTM and has helped to shape the on-going data collection.

The following chapter, Chapter 4 - Presentation of Findings, discusses in detail the findings of all three data collection phases, pilot study, interviews and survey.