2016

Applying grounded theory methods to digital forensics research

Ahmed Almarzooqi

Andrew Jones
Edith Cowan University

Annual ADFSL Conference on Digital Forensics, Security and Law 2016

May 24th, 3:00 PM

Applying Grounded Theory Methods to Digital Forensics Research

Ahmed Almarzooqi
Faculty of Technology, De Montfort University, p11039300@myemail.dmu.ac.uk

Andrew Jones
Faculty of Technology, De Montfort University. Cyber Security Centre, University of Hertfordshire, andy1.jones@btinternet.com

Richard Howley
Faculty of Technology, De Montfort University, rgh@dmu.ac.uk

APPLYING GROUNDED THEORY METHODS TO DIGITAL FORENSICS RESEARCH

Ahmed Almarzooqi
Faculty of Technology, De Montfort University.
p11039300@myemail.dmu.ac.uk

Andrew Jones
Faculty of Technology, De Montfort University.
Cyber Security Centre, University of Hertfordshire.
Cyber Security Research Institute, Edith Cowan University.
andy1.jones@btinternet.com

Richard Howley
Faculty of Technology, De Montfort University.
rgh@dmu.ac.uk

ABSTRACT

Deciding on a suitable research methodology is challenging for researchers. In this paper, grounded theory is presented as a systematic and comprehensive qualitative methodology in the emergent field of digital forensics research. This paper applies grounded theory in a digital forensics research project undertaken to study how organisations build and manage digital forensics capabilities. This paper gives a step-by-step guideline to explain the procedures and techniques of using grounded theory in digital forensics research. The paper gives a detailed explanation of how the three grounded theory coding methods (open, axial, and selective coding) can be used in digital forensics research. Grounded theory offers a rich and detailed methodology for theorising while presenting and exploring the How and Why questions at every stage of the research. The method shared in this paper provides a detailed critique, making it a valuable contribution to the discussion of methods of analysis in the field of digital forensics.

Keywords: digital forensics; open coding; axial coding; selective coding; digital forensics capability; research methodology; grounded theory

1. INTRODUCTION

The aim of this paper is to provide step-by-step guide and examples to those who wish to apply Grounded Theory (GT) using the Straussian procedure as a research method in Digital Forensics (DF) research. GT, according to Charmaz (2008), can be employed as "a major method for conducting emergent qualitative research." Charmaz defines this "Emergent Method" as a method that "begins with the empirical world and builds an inductive understanding of it as events unfold and knowledge accrues" (Charmaz, 2008, p155). In other words, a well-established emergent method like GT is appropriate to emerging fields of research such as DF (Charmaz, 2008). As Charmaz (2008) stated, "emergent methods [like GT] are particularly well suited for studying uncharted, contingent, or dynamic phenomenon." In this regard, GT

is most suited for theorising in DF, a new and growing field with phenomena and areas of research that are technologically dynamic (i.e. emergence of new technologies such as cloud computing) and uncharted in some areas such as organisational DF capacity and the Application and Modification of DF Practices.

Data analysis in GT, using the Straussian approach, involves three main steps: (1) open coding, (2) axial coding, and (3) selective coding. The coding processes differ from each other in each of three steps. Before starting the data analysis, it is important to distinguish the type of data being analysed. GT gives the researcher a number of options for the types of data to be analysed such as documentary analysis, focus group, survey and interviews.

This paper covers transcribed text from interviews conducted by the researcher. The process of analysis, using the Straussian approach, urges the researcher to conduct the analysis after each interview, especially if conducting a series of interviews, to enhance the quality of the data and the researcher's "theoretical sensitivity." Therefore, before discussing the coding process, this paper first describes strategies for enhancing theoretical sensitivity in section 2. Section 3 shows and discusses the coding processes, and finally section 4 sums up the paper with conclusions.

2. THE SAMPLE DATA USED IN THIS PAPER

This section explains how the researcher has applied GT using Straussian procedures and techniques to analyse the data. Throughout this paper, it is important to remember that the researcher's application of GT, using Straussian techniques and procedures for data analysis, employs the dynamic interplay between the researcher and data (Strauss & Corbin 1998, p.13). This complex interplay is not linear, but rather creative and systematic. (Strauss and Corbin 1990, p.13; Strauss & Corbin 1998).

The data used in this paper is the part of a piece of research that discusses the need for a theory of developing DF capability. As opposed to technical or infrastructure capacity, the research examines the capability in terms of the DF organisation as a whole, which includes examination of capability within a DF laboratory and in the management range. Such an organisational view of DF's capability takes into account the interactive roles of policy, people, infrastructure, and the investigative process. Finally, the research relies on the data to identify core capabilities in a DF organisation that can be expressed as a DF organisation core capability framework or theory.

The researcher collected data by interviewing a number of experts in the DF field from the UK and the UAE. The next section gives examples for strategies to enhance the quality of data being analysed before showing the actual data analysis process in section 3.

2.1 Strategies for Enhancing Theoretical Sensitivity

Glaser and Strauss are the initiators and main contributors of GT; each has his own approach and each approach is named after the contributor. The Straussian approach allows for some review of the literature before conducting the data analysis (Corbin & Strauss, 2008), as was done in this research, where a literature review was conducted at the beginning of the research. The first version of GT, the Glaserian approach, on the other hand, criticised the process of finding important words and labelling them in the first stage of data analysis, which is called coding, and discouraged a review of the literature prior to data analysis to let the data speak for itself (Glaser & Strauss, 1967; Glaser, 1992).

The researcher found the Straussian approach to be most suitable because this approach takes into account the researcher's previous background studies and exposure to the relevant literature before the data collection; a significant difference from the Glaserian approach. The researcher had to take literature reviews into account because, in this area, it was a requirement for the Ph.D. program prior to the data collection and analysis. The Straussian Approach has been criticised by more recent constructivist GT researchers for forcing the data into categories under the processes described below (Charmaz, 2008). The researcher took note of the constructivist approach while following the Straussian Approach.

During the coding processes, the researcher applied two strategies for enhancing theoretical sensitivity: (1) "the asking of questions" or questioning, and (2) "the making of comparisons" or constant comparison (Strauss & Corbin 1990, p 62). These two essential strategies helped the researcher to make the analysis of data precise, specific, creative, and open (Strauss & Corbin, 1990, p.62-63; Strauss and Corbin, 1998, p.73). This section demonstrated how the researcher applied these two strategies to the interviews.

To collect the data, the researcher employed the questioning technique which allowed the researcher to consider potential categories, their properties and dimensions (Strauss & Corbin, 1990, p.77). The basic types of questions that the researcher used as a guide were the 5Ws plus 2H, or Who, What, Where, When, and Why plus How and How much? (Strauss & Corbin, 1990, p.77). Of course, many questions came naturally as the researcher responded to the data. The researcher applied the memo creation process while employing the questioning technique to make the process systematic and documented for later referencing. An example is presented below.

Table 1
Questioning Memo for interview 04AUINTUAE14

MEMO 11.20.14 QUESTIONING
The subcategory "Preservation" came from and with the concepts "Imaging" and "Duplication." This raises many questions that are required to be elaborated and answered either from the data or the literature. Who conducts the preservation? Is it the same person through the entire investigation process that does the preservation, analysis and reporting? There seems to be a step before preservation as well, which is identification. Do these steps have to happen in sequence or can they go back and forth throughout the investigation process. How many copies must be made or preserved? Does it matter? Where the images of digital evidence stored? Does this now have a relationship with the tools used in terms of storage? How long after the seizure of the DF evidence must the imaging or duplication takes place? Is it right after identification? Is there a rule that waiting too long makes it more likely that the evidence has been altered? What are the other purposes of imaging and duplication? What happens to the duplicated data after the investigation ends? Is there a privacy issue involved? Should there be a policy of storage and/or disposal of the imaged data? Who is in-charge of the whole process? How can the DF procedures guarantee that he imaged data have been secured from privacy breaches?

GT is often referred as a "constant comparative method of analysis" (Strauss & Corbin 1990, p. 62; Glaser & Strauss, 1967, pp. 1-116; Charmaz, 2006). Constant comparison can be defined as "the process of constantly comparing instances of data." (Urquhart et al, 2010). Constant comparison's ultimate goal is to reach data and theoretical saturation

(Strauss, 1987; Glaser, 1992; Charmaz, 2006; Urquhart et al., 2010). Making comparisons is essential to identify and categorise concepts (Strauss & Corbin, 1990, p. 84). Constant comparison therefore, is applied throughout the coding process from open, axial, to selective coding (Charmaz, 2006); and through each data set.

Again, whenever possible, the researcher applied the technique of using Memos when applying the constant comparison strategy technique to make the process systematic and ensure that the data collected was recorded for later referencing. Occasionally the researchers may skip the theoretical saturation and use the Memo aspect of GT (Charmaz , 2008). Theoretical saturation, according to Charmaz (2008) is widely claimed but scarcely practiced. Using Memos is necessary in GT, and must be done using more analytic as opposed to descriptive writing (Charmaz , 2008). Here is another example of a memo on constant comparison:

Table 2
Comparison Memo in Interview 04AUINTUAE14

MEMO 11.20.14 COMPARISON
In the previous memo, I asked the question: Do these steps have to happen in sequence or can they go back and forth throughout the investigation process. It is therefore important to compare the sequences or phases of the investigation process. So comparison can be made between the processes of preservations with identification. Do both processes take the same time to be carried out? Does one take more time than the other? Why do they take different time? Time is a property with dimensions of hours to months. It would be interesting to compare the time dimensions for each of the processes. Then to compare the causes of the delay or time challenges. Are they caused by people, tools, process, or policy? Are the skills required for each of the processes the same? There seems to be more skill required in analysis compared to preservation. Is this true or is a specialized skill needed in instances where the evidence to be preserved may be at risk of destruction or corruption. Can the processes be rated in terms of difficulty? The dimension could be from least difficult to most difficult. Does the difficulty related to the tools used, the skills of the people involved or some other intervening cause like third parties or constraints in the investigation?

The GT method relies on the researcher's imaginative approach to the data, a point that some researchers may see as an obstacle (Charmaz, 2008). According to Charmaz, GT requires abductive reasoning, which "invokes imaginative interpretations because the researcher imagines all possible theoretical accounts for the observed data and then forms and checks hypotheses until arriving at the most plausible interpretation of the observed data" (Charmaz, 2008; Charmaz, 2006). While asking of questions and constant comparisons are tools that aim to help the researcher enhance theoretical sensitivity; the use of these tools is highly dependent on the researcher's imagination. The ability of the researcher to enhance theoretical sensitivity would likely depend on the researcher's "intimate familiarity" with the studied phenomena (Charmaz, 2008). A researcher, therefore, who does not become familiar with both the literature and the data, will probably have a difficult time with the GT method.

3. APPLICATION OF THE GROUNDED THEORY CODING

This section provides an example of data analysed from the interviews and the concepts, sub-categories and categories that emerged using GT coding. Coding in GT is defined as the "analytical processes through which data

are fractured, conceptualised, and integrated to form theory" (Strauss & Corbin, 1998). There are three stages in the coding process: open ended coding, axial coding, and selective coding (Robson, 2002). In open-ended coding, the aim is to define simple categories and concepts for comparison and understanding (Charmaz, 2000; Robson, 2002). Axial coding narrows down the data and focuses by examining the data and providing a context for relationships in the data (Charmaz, 2000; Robson, 2002). Finally, according to Strauss and Corbin, selective coding is "the process of integrating and refining the theory" (Strauss and Corbin 1998, p. 143).

3.1 Interplay between Open and Axial Coding

According to Strauss and Corbin (1998, p.58), though open and axial coding are distinct analytic procedures, when the researcher is actually engaged in the analysis he or she alternates between the two modes." A possible trap for researchers employing the GT method is to become linear in their approach. Doing so would likely lead to confusion about the data, difficulty in grounding the categories and properties, and certainly a theory that is difficult to reconcile with the data. It is important to realize that when discussing the coding process, the researcher here actually moved "back and forth." As stated by Strauss and Corbin (1990, p.98), "though open and axial coding are distinct analytic procedures, when the researcher is actually engaged in the analysis he or she alternates between the two modes." For example, the researcher asked the following question in 11CTINTUAE14 at page 9:

...how did you become a ... digital forensics specialist?

The participant replied as follows:

I had to undergo, of course, training. So I did the tools training.

In the above question and reply, open coding resulted in identifying the phenomenon of "undergoing training" which then led to the concept of "tool training." The concept of "tool training" was further developed and led to the types of "tool training" which include, among others, "Access Data FTK Training", "Guidance Software Encase Training", and "XRY training". The dimensions led to how often the training took place (once to three times), how frequent the training took place (yearly), and the depth of the training (overview to specialize). Eventually the concepts were categorized under "Types of Training."

Concurrently, with the open coding process, the researcher was connecting the "tool training" concept with another category called "DF tools" and a subcategory called "Forensic Analysis Software." These subcategories and categories arose from concepts relating to the tools identified by participants as being used in the investigation process like "FTK," "EnCase" and "XRY." In other words, there was a relationship between the categories of "DF Tools" and "Types of Training." Axial coding was also taking place at the same time as open coding. There was interplay between open coding and axial coding. The researcher had to use the Paradigm Model to develop the axial coding further.

The researcher then open-coded a different phenomenon labelled "Forensic Training" that belonged to the category of "Type of Training." The process jumped back and forth between open coding and axial coding, from phenomenon to concept to category (back and forth), to data coding to writing memos, to naming categories to connecting relationships, and so on. The most important lesson was that GT is a complex transactional method of data analysis that dynamically carried the

researcher's analysis into many discoveries. There was a "constant interplay between proposing and checking" and between inductive and deductive thinking (Strauss & Corbin, 1990, p. 111).

3.2 Application of Open Coding Procedure

Open coding is part of the Straussian GT analytical process that "pertains specifically to the naming and categorising of phenomena through close examination of data" (Strauss & Corbin 1990, 62). After the interviews are transcribed, the researcher should categorise answers to questions during and after open coding, following the Straussian coding model paradigm (Corbin and Strauss, 2008; Strauss and Corbin, 1990; Strauss and Corbin, 1998).

The researcher applied open coding by (1) labelling the phenomena as named concepts, categorising concepts that seem to relate to each other under categories and subcategories whenever relevant (2) developing the categories and subcategories by identifying possible properties and dimensions, and (3) grounding the concepts, categories and subcategories to the interviews. This section shows how the researcher applied the open coding process to the data.

3.2.1 1. Initial Microanalysis Open Coding and Subsequent Coding

Open coding is a flexible methodology. "There are several ways of approaching the process of open coding" (Strauss & Corbin, 1990, pp. 72). How a researcher handles the volume of data is, therefore, dependent on the needs of the researcher. The researcher may interact with the data on a line-by-line analysis (whether word-for-word or phrase-by-phrase), by sentence or paragraph analysis, or by an entire document analysis (Strauss & Corbin, 1990, pp.72-73). Some may criticise microanalysis for being too tedious. However, "generating your categories early through line-by-line analysis is important because categories also become the basis of your theoretical sampling" (Corbin & Strauss, 2008).

The researcher, therefore, began the open coding process with a line-by-line analysis, or microanalysis (Strauss & Corbin, 1998, p.57) of the first two interviews: 03ALINTUAE14 and 04AUINTUAE14. The researcher labelled a number of texts via underlining as potential items representing codes or concepts. Corbin and Strauss defined concepts as "Words that stand for groups or classes of objects, events and actions that share some major common property(ies), though the property(ies) can vary dimensionally" (Corbin & Strauss 2008, p. 45).

In subsequent open coding, the researcher applied both sentence and paragraph analysis. Below is a memo regarding this process.

Table 3
Interviews #5-19

MEMO
After the line-by-line analysis or microanalysis of interviews #3-4, the researcher conducted sentence by sentence, paragraph by paragraph, and document-by-document open coding of the interviews, one at a time, from interview #5-19. The researcher intends to add the open coding of the rest of the data to the concepts. Also, the researcher is considering grounding the data to the concepts and/or phenomena as the researcher anticipates the need to return to the specific data as the researcher goes to axial coding, selective coding and then more open coding. Therefore, grounding will be a continuous and flexible process as will be the coding. The grounding will be accomplished by stating the interview number and the page number where the phenomena or concept was taken from the following format (5p1), which means the concept or phenomena was found in interview 5 a page 1.

3.2.2 Labelling Phenomena and Naming Categories

An important step in the data analysis is the conceptualisation of the data. Conceptualising data is not the same as summarising data (Strauss & Corbin, 1990, p. 64). Rather, it involves identifying, in the data, the "central idea, event, happening, incident", called a phenomenon of an action or interaction or a set of actions or interactions and describing or naming that phenomenon (Strauss & Corbin, 1990, p. 96). As an example, after a microanalysis of the first two interviews, the researcher identified the following Phenomena or conceptual labels:

Table 4
Phenomena and Concepts Labels

PHENOMENA: ACTIONS DESCRIBED BY PARTICIPANTS	CONCEPTS
Handle cases	Investigation
Must finish a case in limited time	Deadline
Must follow an investigation process	Investigation process
Must stay within scope of investigation, cannot investigate everything	Scope of investigation
There is a documented process to follow, conducts DF investigation based on experience	Documented process and procedures
Look at a reference point, no absolute standard exist	Multiple standards
Follow usually, follow experience, experience dictates what to do	Best practices

Then, the concepts were grouped into categories. Categorisation, which encourages the generation of initial categories, is the next step in the Straussian open coding process (Strauss & Corbin, 1990, p. 63; Corbin and Strauss, 2008). Once a set of phenomena or concepts have been identified, they are categorised into categories and subcategories, a process called conceptual categorisation (Strauss & Corbin, 1990, p 65) "Categories have conceptual power because they are able to pull together around them other groups of concepts or subcategories" (Strauss & Corbin, 1990, p. 65).

Here, the researcher categorised the concepts generated from the listing of phenomena. First, concepts were grouped into categories that covered multiple related phenomena. Next, the concepts were further

grouped into subcategories. The researcher, in the main, invented the names of the categories, but at times the names were borrowed from the literature (Strauss & Corbin, 1990, p. 68), or from the words of the interview participants themselves, called "in vivo codes" (Strauss & Corbin, 1990, p. 69).

3.2.3 Developing Categories and Subcategories with Properties and Dimensions

Another important step in the process was the development of the categories in terms of their properties and dimensions. In order to expand the categories, the researcher identified possible properties and dimensions for each of the identified categories. According to Strauss and Corbin (1990, p. 69), "properties are the characteristics or attributes of a category, and that dimensions represent locations of a property along a continuum." The process of creating dimensions enables the researcher to give specificity to the category or concept (Strauss & Corbin, 1990, p. 72).

Identifying the dimensions and properties made more obvious the relationships of a property, dimension, or category to other categories, subcategories, or properties. The researcher also engaged in constant comparison, where the researcher compared categories, subcategories, and concepts to other categories, subcategories, and concepts. The act of comparison took into account the existing literature and new concepts and categories that arose from each new data set. In this regard, constant comparison was carried out from one set of data to another. Likewise, axial coding inherently occurred simultaneously during the open coding process (Strauss & Corbin, 1990). Overall, the process of expanding the categories with properties and dimensions resulted in a richer set of coding that made the theoretical memo much richer as well. The researcher was able to discuss aspects of the categories that would have been largely ignored without engaging in these more detailed steps in the GT process. An example of how the researcher developed a category using properties and dimensions is in the following table:

Table 5
Properties and Dimensions

CATEGORIES	PROPERTIES	DIMENSIONS
Investigation Process	Human Factor	How many investigators?
		Specialization needed?
		Extent of investigator skill
	Challenges	Time Constraint (Fast)
		Limited Resources
		Volume

The strategy of questioning was very helpful as well, at this stage, because the researcher was able to ask questions in subsequent interviews about the subcategories that enhanced the researcher's understanding of the category. Questioning led the researcher to be more theoretically sensitive to other concepts relating to who conducts the investigation, the steps in the investigation process, and challenges faced during these procedures including storage and time constraints.

As the researcher identified concepts and categories during the open coding process, the researcher also grounded the data by using the faceted code of the interview and corresponding page number into the tables. Grounding the data simultaneously made it easier to refer back to the interviews using Memos. Grounding the data in this manner also allowed the researcher to identify concepts in the research that needed further data, or that are not theoretically relevant.

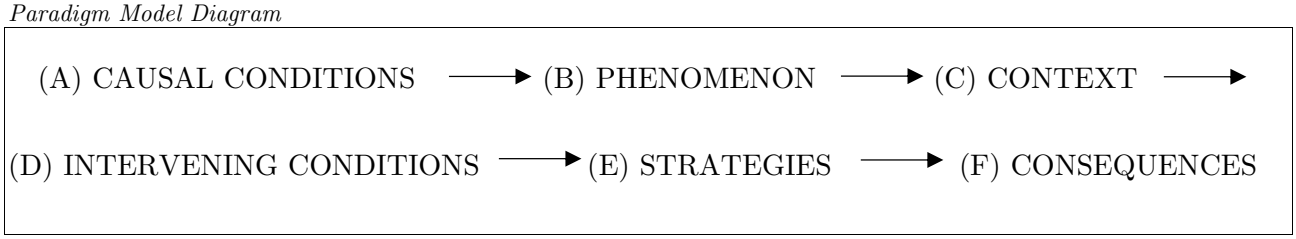
3.3 Application of Axial Coding Procedure

Axial coding is the process of putting the data back together in new ways by making connections between categories and subcategories (Strauss & Corbin, 1990, pp. 96-97). Simply put, it is the "process of relating categories to their subcategories" (Strauss & Corbin, 1998, p. 123; Corbin and Strauss, 2008). It comes after identifying categories in the open coding process by finding relationship between the categories and subcategories. The researcher applied axial coding to the data using the paradigm model, and then by developing the categories using the paradigm model and identifying the properties and dimensions of the categories and subcategories.

3.3.1 The Paradigm Model

In the axial coding process, the relationships among the subcategories and categories are linked by identifying the (1) causal condition, (2) phenomenon or concept, (3) context, (4) intervening conditions, (5) action/interaction strategies, and (6) consequences (Strauss & Corbin, 1990, p. 99). The paradigm model has been commonly referred to in the following simplified diagram:

Table 6.
Paradigm Model Diagram



It is important to use this model in any GT analysis because failure to do so will lead to a "lack of density and precision" in the analysis (Strauss & Corbin, 1990, p.99). The researcher used the paradigm model to link relationships among subcategories and categories. An example of the use of the paradigm model is shown in the following table7:

Table 7
Paradigm Model Sample

Causal Condition	Phenomena	Context	Intervening Conditions	Strategies	Consequences
Crime	Investigation	Digital or electronic evidence	Destruction of Digital Evidence	DF Investigation Framework	Finding of Evidence/ solving case
Finding digital device at crime scene	Type of DF investigation	Inside PC/ Mobile/ Flash Drive	Challenges to Investigation	Identification	Not finding evidence
Receiving request from client	Type of DF laboratory			Preservation	Reporting of findings
Request for research and development	Length of investigation			Analysis	Court testimony
Request to test security	Recurrence			Tool specific strategies	Eliminate security breach
Security breach (ie hacking, or misuse of information	Type of crime				Create mechanism to prevent future breaches

It should be noted that constructivists have criticised the paradigm model of Straussian GT because it may force the researcher to fit the data into the categories (Charmaz, 2008). The researcher here, however, viewed the paradigm model as a means of gaining a better understanding of the categories and how they relate to each other and to more specific propertied and dimensions. The paradigm model, therefore, helped the research to develop better relationships with (from) the data.

3.3.2 Developing Relationships

The axial coding process of linking and developing categories is complex (Strauss & Corbin, 1990, p.107). The procedure requires simultaneous action of relating subcategories to categories, verifying hypothesis with actual data, identifying properties and dimensions, and identifying variations in the phenomena through constant comparison of categories and subcategories (Strauss & Corbin, 1990, p. 107). It is a process of identifying patterns that emerge from the coding process. This complex process produced a set of categories and subcategories, an example of which are tabled below:

Table 8.
Categories and Subcategories

CATEGORIES	SUBCATEGORIES
Investigation Process	Purpose of Investigation
	Scope of investigation
	Identification
	Preservation
	Analysis
	Reporting
	ACPO Principles

As the researcher identified concepts and categories during the axial coding process, he also grounded the data by coding the faceted code of the interview and corresponding page number into the tables of subcategories and categories.

Selective coding is the final step in the data analysis process. It is the "process of selecting the core categories, systematically relating it to other categories, validating those relationships, and filling in categories that need further refinement and development" (Strauss & Corbin, 1990, p. 116). Corbin and Strauss defined selective coding as the "process of integrating and refining the theory" (Strauss & Corbin 1998, p. 143).

In essence, selective coding is about integration (Strauss & Corbin, 1990, p. 117). After data analysis, theoretical sensitivity requires the researcher to conceptualise and formulate a theory that emerges from the data, literature, existing theories or experience of the subject under investigation (Glaser & Strauss, 1967; Corbin & Strauss, 2008; Urquhart et al., 2010; Glaser, 1978). The theories conceptualised by the researcher must then be related to other theories in the field in what is known as theoretical integration (Urquhart et al. 2010). Theoretical integration is the process of comparing the generated substantive theory with previously developed ones with the aim of scaling up the findings and achieving theoretical explanation (Urquhart et al., 2010; Birks & Mills, 2011). While the process is similar to axial coding, in that it requires identifying relationships, selective coding is "done at a higher more abstract level of analysis" (Strass and Corbin, 1990, p. 117).

GT applies iterative conceptualisation to arrive at a theory. Iterative conceptualisation requires the researcher to analyse the data by increasing the level of abstraction and moving the degree of conceptualisation beyond description to a more theoretical domain. This higher level of abstraction should be applied with theoretical sensitivity during the interpretation of the coding using constant comparison and the data from the theoretical memo (Urquhart, 2010). The higher level of categories arrived at by the researcher should be grouped into broader themes called the core categories that can be generalised into theories.

The researcher here applied selective coding by: (1) identifying patterns and core categories, (2) relating the categories at the dimensional level, (3) explaining the story line, and (4) validating the relationships by grounding the theory to the data (Strauss & Corbin, 1990, p. 117-118).

The researcher first identified patterns in the categories and subcategories. This pattern identification was done through the application of the paradigm model, diagramming, and using Memo. It also helped to specify the dimensions of the category and subcategory being related. In doing so, the researcher found that four core categories have emerged from the categories: (1) Investigation, (2) Infrastructure, (3) People, and (4) Policy. These are the four core concepts of capability being described in the data and the literature.

The researcher next linked the core categories to their dimensional level. One specific example is the selective coding of the core category "investigation", which consisted of the categories of "investigation process," "evidence admissibility," and "investigation procedure" that were associated to their specific dimensions. Under the category "investigation process," these two examples of the many properties and dimensions were identified:

Table 9
Properties and Dimensions

PROPERTIES	DIMENSIONS
Human Factor	Number of investigators: few to many
	Number of specialization needed
	Extent of investigator skill
Challenges	Time Constraint: limited to unlimited
	Limited Resources: limited to unlimited
	Volume of data: low to high
	Client Trust: low to high
Results	Number of data identified: low to high

The properties and dimensions identified in the category of "investigation process" linked to the core category of "Investigation" as these dimensions give specificity to the DF investigation as a core concept. For example, the number of investigators is a factor that DF organisation must consider in determining capability. The researcher identified the need for a formula or ratio for determining the number of investigators, a number that could be linked to the number of cases per month that go through the laboratory or possibly the amount of data that the laboratory processes per month measured in bytes. For example, the researcher suggests that determining the ratio of investigators in a DF organisation is important to determine efficient "throughput" (Jones & Valli, 2011).

This observation also linked to another dimension identified in the same core category, category: The property "Challenges" which means challenges in the investigation process had a dimension of "time constraint" that is measured by the time available to conduct the investigation, whether it is limited or unlimited. In other words, does the investigation concern only specific areas or does it include everything in the evidence. This dimension of "time constraint" under the property of "Challenges" is linked to the dimension of "number of investigators" in a different property called "Human Factor." The dimensions, therefore, were also being linked, and strengthened the core category of "Investigation."

Additionally, the researcher identified the relationship between the dimension "number of investigators," to the core category "People," and "number of investigators" also became a dimension in that core category. Likewise, the same dimension "number of investigators" was applicable in another core category, "Infrastructure," where the category "Building a DF Facility" and the subcategory "Facility Requirements" led to the concept of "people" or "staffing." In other words, the dimension "number of investigators" was essential across core categories in identifying staffing needs (core category "People"), identifying initial staffing needs in "Building a DF Facility" (core category "Infrastructure") both of which affected the ability of the investigation to meet "Challenges" based on "Human Factors." This linking to other core categories strengthened the core categories because the researcher was able to identify the similarities and differences of the role of the dimension in the distinct core categories

using the constant comparison technique. Of course, many more relationships and linkages arose from the dimension of "number of investigators."

The example above shows that relating the core categories and categories at the dimensional level is, therefore, an essential step in the selective coding process because it adds specificity to the theory development by linking specific measures in the dimension to the higher level categories and across different higher level categories.

Finally, the researcher explained the story line that seemed to emerge from the analysis of the data. Before attempting to state the story line, the researcher asked what it is about the core categories and the subsidiary categories that stand out (Strauss & Corbin, 1990, p.119). In a memo on the story line, the following table is what the researcher wrote:

Table 10
Memo: Story Line

MEMO 02/02/15 STRORYLINE
What is most striking here are the different ways that people think about the concept of capability in the context of digital forensics laboratory building and management. Some understand the capability in terms of the DF tools available in the organisation; others in terms of the people or the human resources, while others as having both the DF Tools and the human resources. Others still view capability in terms of their ability to act and/or interact in the context of the challenges they face during the digital forensics investigation process. While many recognise policy as necessary in the DF organisation, it is not readily identified as a component of capability.

In essence, the story line is that there seems to be a need in DF for a system that recognizes the core components of capability, as well allows the DF industry to discuss capability in the same paradigm. Currently, what capability means to a DF organisation is subjective and changes to suit the needs of the organisation. The story line arrived at by the researcher seems to pave the way towards creating a theory on DF capability for developing and managing a DF organisation.

An important step in the selective coding process is the validation of relationships among the categories by connecting them to the data. This validation process occurred mainly at the conceptual and dimensional level, therefore emphasizing the need to relate higher level categories to the dimensional level. The benefit of grounding the data during open and axial coding is appreciated most at the selective coding process. It became much easier to ground the more abstract phase of coding when grounding was already existent in prior coding processes.

Referring back to the example of the dimension "number of investigators," this dimension was grounded by going back to the interviews that were previously grounded through coding during open, axial, and selective coding. For example, the need for having a number of investigators was linked to the following interview:

QUESTION: "*...how do you define an organisation to be digital forensics capable?*"

ANSWER: "*...they should have enough capabilities in term of human resources, people have enough experience*" (07COINTUAE14, p. 4).

The participants here used the word "enough" which triggered the question of what "enough" means. One of the obvious meanings of "enough" is quantity, but there is also a quality dimension to the word. Therefore, the

researcher also linked the dimension of "number of investigator" to the dimension "skill level" of the investigator under the category "Quality of Investigator" which was linked and appeared in the core categories "Infrastructure," "Investigation," and "People." In other words, the process of grounding the theory also led back to the coding process, demonstrating how GT goes back and forth between inductive and deductive analysis.

Finally, the story line was connected to existing literature and theories. Primarily, the researcher applied the story line to (1) the emerging research in DF on the applicability of the capability maturity model (Kerrigan, 2013; Al-Hanaei & Rashid, 2014), (2) the work by Grobler on DF readiness and capability (Grobler, 2010), and (3) the works by Jones and Valli (2011) on building a DF laboratory with processes and procedures.

4. CONCLUSION

This paper shows how to apply GT methods using the Straussian approach in DF research. It is important to note that the method demonstrated in this paper is that of the Straussian Approach, which, though it has general similarities, does differ from the Glaserian Approach. The researcher therefore, ought not to apply the method here directly to a Glaserian GT research. In fact, one criticism of GT may be that it has evolved into competing "constellations" of methods that can be confusing to those researchers new to GT (Charmaz, 2008).

Regardless of which approach taken, this paper has addressed a gap between the methods and literatures for the DF and IT/IS field. Far too many researchers who have used the GT methodology, for example, failed to demonstrate their use of Memo, an important aspect of GT methods. Researchers ought to improve the way in which they demonstrate the application of GT methods by showing how their data analysis evolved from open coding, to axial coding, and then to selective coding.

Researchers must also demonstrate how they grounded the data in their categories, properties and dimensions. The grounding to the data is what gives the research method its integrity and strengthens the theory the researcher arrives at. Researchers should not forget that the GT method is ultimately about theorising (Strauss and Corbin, 1998). It is more important, however, to explain how one arrived at such theory with the research data. Such theorising must be demonstrated through an explanation of the story line and then grounded in both the literature and the data.

Digital Forensic Organisation Core Capability (DFOCC) is the framework derived from analysing data using grounded theory. DFOCC enhances the admissibility of evidence as it requires that a DF organisation has made certain procedures part of its business process. The framework is simple because it is narrowed down to four variables (Policy, People, Infrastructure, and Investigation) that are required for a DF organisation to be DF capable. The DFOCC framework will help the entire digital forensic investigation process prove the guilt of a perpetrator because the DFOCC will help organisations ensure that they have the correct resources and procedures in place to carry out investigations efficiently. The DFOCC framework aims to reduce the possibility of successful challenges to DF evidence presented in courts. For example, DF organisations that do not already do so, will be required to document each stage of the investigation process, which will in turn strengthen expert testimony on such requirements as chain of custody and authentication.

Finally, what is important in grounded theory is not the result but the process. As Charmaz noted, "The grounded theory method

emphasizes the process of analysis and the development of theoretical categories, rather than focusing solely on the results of inquiry" (Charmaz 2008).

REFERENCES

Hanaei, A., Hamad, E., & Rashid, A. (2014, May). DF-C2M2: A Capability Maturity Model for Digital Forensics Organisations. In *Security and Privacy Workshops (SPW), 2014 IEEE* (pp. 57-60). IEEE.

Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. Sage publications.

Charmaz, K. (2000). Constructivist and objectivist grounded theory. *Handbook of qualitative research*, *2*, 509-535.

Charmaz, K. (2003). *Grounded theory - Objectivist and constructivist methods*. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry , 249-291*

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis (Introducing Qualitative Methods Series).

Grobler, M. M. (2010). *Digital forensics standards: international progress*. Proceedings of the South African Information Security Multi-Conference (SAISMC 2010)

Grobler, C. P., Louwrens, C. P., & Von Solms, S. H. (2010, February). A framework to guide the implementation of proactive digital forensics in organisations. In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on* (pp. 677-682). IEEE.

Jones, A., & Valli, C. (2011). *Building a Digital Forensic Laboratory: Establishing and Managing a Successful Facility*. Butterworth-Heinemann.

Charmaz, K. (2008). Grounded theory as an emergent method. *Handbook of emergent methods*, 155-170.

Corbin, J. and Strauss, A. (2008) *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Sociology Pr.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ:

Glaser, B. G. (1992). *Emergence vs forcing: Basics of grounded theory analysis*. Sociology Press.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization contrasted with description*. Sociology Press.

Kerrigan, M. (2013). A capability maturity model for digital investigations.*Digital Investigation*, *10*(1), 19-33.

Robson, C. (2002). Real world research. 2nd. *Edition. Blackwell Publishing. Malden*.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research* (Vol. 15). Newbury Park, CA: Sage.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory'back into grounded theory: guidelines for grounded

theory studies in information systems. *Information systems journal*, *20*(4), 357-381.

