Faculty of Science & Technology
Department of Computing and Informatics
PhD Thesis

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Dissertation Declaration

I agree that, should the University wish to retain it for reference purposes, a copy of my dissertation may be held by Bournemouth University normally for a period of 3 academic years. I understand that once the retention period has expired my dissertation will be destroyed.

Confidentiality

I confirm that this dissertation does not contain information of a commercial or confidential nature or include personal information other than that which would normally be in the public domain unless the relevant permissions have been obtained. In particular any information which identifies a particular individual's religious or political beliefs, information relating to their health, ethnicity, criminal history or sex life has been anonymised unless permission has been granted for its publication from the person to whom it relates.

Copyright

The copyright for this dissertation remains with me.

Requests for Information

I agree that this dissertation may be made available as the result of a request for information under the Freedom of Information Act.

Signed:
Name:
Date:
Programme:

Original Work Declaration

This dissertation and the project that it is based on are my own work, except where stated, in accordance with University regulations.

Signed:

Acknowledgements

Contents

List of Figures

Chapter 1

Introduction

Lorem ipsum dolor sit amet, consectetur adipisicing elit (**?**), sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1.1 Research Rationale

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 2

Case Study - Applying CI to Open Data

2.1 Case Study Design

The aim of conducting this case study is to establish whether or not Nissenbaum's framework can be successfully applied as a decision tool around privacy in practice when publishing public data open source. The intention is to establish whether the CI framework is appropriate for use in the final Open Data Publication Framework (ODPF) as it is, or, whether it will be necessary to adopt this to better suit the requirements of the ODPF.

The case study will be conducted following the methodology of Robert Yin, the leading authority on case study methodology **?**. This involves ensuring five areas are included in the methodology **?**:

1. Problem definition; defining the case, determine whether single- or multiple case study; unit of analysis;
2. Research design (Protocol); create protocol defining the theory or hypothesis, research questions and research strategy;
3. Nature of evidence (Data Collection);
4. Analysis and interpretation; define analysis approach; and
5. Manner of presentation; present findings.

2.2 Problem Definition

Having studied the existing guidance provided by a number of Government bodies and affiliated parties, it seemed that, beyond advising practitioners to adhere to data protection regulations, there was no advice on how to incorporate privacy into the decision

making process when considering whether a particular dataset would be appropriate for publication.

Turning to the literature on privacy to establish what advice could be found on how privacy may be applied in practice, where this had been attempted, it appeared that the most prominent work on this subject was Helen Nissenbaum's Contextual Integrity (CI) Framework **?**.

Thus, it was decided that applying Nissenbaum's framework would be the most appropriate way to test how effective a tool this might be in practice. To do this would however, require testing against what practitioners would be dealing with in their daily work. Upon consideration, a case study was decided upon as a well-suited method for conducting such a study.

Yin contends that a case study may consist of a single case study or a multi-case study, both are equally valid providing a structured approach that can be validated is used in conducting the study **?**. Therefore, a single-case study design has been chosen as this, it is considered, will provide sufficient insight into whether or not the CI framework is workable in practice, when making decisions around publishing public data open source.

Having established the method of study, a unit of analysis needed to be decided upon. To this end, it seemed appropriate to find a public body that already publish some data open source as they will have an existing process in place against which the CI framework can be tested. Thus, a public body will be used as the main unit of analysis.

Thus, a UK Local Authority who currently publish date open source was approached to establish whether they would agree to work with the research team in applying the CI framework to real data, in a real practical setting. They kindly agreed to collaborate and trial the framework.

2.3 Research Design - Protocol

2.3.1 Hypothesis

First, it is hypothesised that Existing guidelines do not provide guidance on privacy beyond advising that data protection regulations must be adhered to, hence the reason for this case study. Then, it is hypothesised that Nissenbaum's CI framework, whilst it may be applied to any scenario, is best suited to a specific case where all but will not prove sufficient when the data recipient is not known, thereby severely limiting the usability of the framework in practice. Further, it is hypothesised that applying the CI framework, in its current format, will result in over cautiousness on the part of the public body, who will deem most datasets unsuitable for publication. This may result in the public bodies being unable to meet their legal obligations in publishing data open source.

2.3.2 Research Questions

To test these hypotheses, the case study will seek to answer the following questions:

1. No practical advice how to incorporate privacy into decision making around open source publishing, beyond stipulation that data protection regulations must be adhered to, can be found in existing guidelines?

2. In a practical application with a public body organisation, does Nissenbaum's CI framework work in deciding whether or not public data should be published open source?

3. Using the CI framework in its current format result in most datasets being rejected for publication?

4. Does using the CI framework in its current format result in public bodies not being able to meet their legal obligations with regard to making data available open source?

2.3.3 Question Preparation

In order to answer the research questions relating to Nissenbaum's CI framework, it is first necessary to create an easy to follow methodology for following the framework. This will be done in three phases:

Phase 1 Create a practical application model for applying the CI framework in practice and

Phase 2 Evaluate the model using 3 independent reviewers

Phase 3 Evaluate the model against an existing formal model for applying the CI framework

2.3.4 Phase 1 - Creating a Practical Application Model (PAM)

This section explains how Nissenbaum's CI framework has been interpreted and applied into a Practical Application Model (PAM) by the primary researcher. The intention is that PAM will then be used to assist practitioners in the decision making process when considering whether or not to publish a dataset as Open Data.

Nissenbaum talks about CI in terms of the questions that one might ask in applying the framework and discusses at length how this might be applied to particular situations. She does not however, provide much detailed practical guidance that users can easily apply to their particular scenario.

Upon studying Nissenbaum's book in detail, it was determined that there are two areas where Nissenbaum offers some level of detailed guidance for applying the framework to a practical in the book. These are; "the three key elements of the framework - explanation, evaluation, and prescription" (p. 190) and the; "Augmented Contextual Integrity Decision Heuristics' (p. 181) that she breaks down into a nine step decision making approach that will evaluating a practice or system and whether this poses any privacy risks (pp. 181 -182). However, whilst these areas do provide more tangible advice on the application of the framework, they still do not go into sufficient detail for practitioners to apply without considerable interpretation and breaking this down into more tangible individual elements.

Therefore, starting with these two pieces of guidance, work began on creating a Practical Application Model (PAM) that could be applied in a practical setting.

Using the two pieces of guidance as the basis for breaking down the CI framework into more manageable chunks PAM was created to enable practitioners to work their way through the framework in a more gradational manner. The method and process of this translation is explained in this section.

2.3.5 Stage 1 - Key elements

First, to make it easier for practitioners, these discussions and questions have been translated into a PAM containing a more granular approach. By breaking these into bite-size specific individual questions it is hoped, this will make PAM more usable in practice by guiding practitioners through the CI framework in a step by step manner.

In her book, Nissenbaum explains that the framework has three key elements; explanation, evaluation and prescription (p. 190). However, it was felt that practitioners may not relate to these descriptions as it is not necessarily obvious to a layman what they mean.

Looking at the terminology used in practice with regard to decision making, risk, is one area that is measured in most areas of IT practice from security risk through to managing financial risks. Therefore, a decision was made to translate Nissenbaum's three elements into language that practitioners might be more familiar with as part of their existing everyday work as follows; 'Facts', 'Risk Assessment'; and 'Decision'.

As it appeared that these elements provided an excellent, logical group of three overarching categories that would frame PAM into understandable, logical progression steps. Thus, the elements have been translated into steps so that each element will in PAM be delineated and aid the step-by-step approach logic. The aim of this was to make PAM user friendly and easy to follow.

The meaning of each of these steps have been outlined in the following sections. Each section is immediately followed by an explanation for PAM.

Explanation (Step 1 - Fact)

Nissenbaum explains that the 'explanation' element refers to the practice or system to be assessed. These should be assessed in view of any "context-relative informational norms" that may be breached. This should include an assessment of the key "actors", i.e. the people that are/could be affected and their "roles", as 'data subjects'; 'data senders' or 'data recipients'. It should also consider the "attributes", i.e. the information itself (the data) and how this information is transmitted ("transmission principles") and whether any changes to these elements potentially violate the existing or proposed new information flow.

As this basically involves getting to grips with and understanding the data, the people who work with or are subjects of the data, and how the data is transmitted, this has been translated into Step 1 and called; 'Facts' in PAM.

Evaluation (Step 2 - Risk Assessment)

This part involves assessing the information in view of pertinent "values, ends and purposes". Nissenbaum contends that comparing existing flows of information with the proposed new flows, accounting for any breaches (or potential breaches) of values and comparing these to any potential privacy conflict or threat, will afford practitioners the opportunity to identify and mitigate against these. This will in turn enable practitioners to "establish the significance of each value in light of its contextual ends and purposes" (p 191).

In PAM this has been translated into Step 2; 'Risk Assessment'. The reason behind this decision is that effectively, what the evaluation is trying to achieve is an assessments of the risks associated with any proposed changes or alterations in the data flow.

Prescription (Step 3 - Decision)

Prescription involves presenting the findings which will guide the practitioner in whether or not a practice or process poses a potential challenge to privacy.

This, it is contended, involves making a decision as to the compatibility or non-compatibility of the information for allowing those changes or alterations in the data flow. therefore, this has been translated in PAM into Step 3; 'decision'.

Format and layout of PAM

PAM will be presented to the practitioner in an excel spreadsheet consisting of three worksheets, one for each step. At the header of each worksheet will be the name of the dataset under consideration. The layout of the worksheets will be as follows:

Tab 1 - Facts This tab will consist of 6 columns with the questions organised in groups according to the subject header under consideration in the rows below. The column headings are:

1. Step - this will be broken into numbers corresponding with the group of questions.
2. Contextual Integrity - this will hold the subject header for each group of questions and a brief explanation of what that section relates to e.g. ”describe the existing practice in terms of information flows”
3. Applying CI framework to dataset - this will contain a brief overview of the subject, e.g. data, information flow, actors etc.
4. Questions - here the questions will be listed
5. Answers - here the practitioner will be able to respond to the questions
6. Score - here the practitioner will be able to score the answer.

Tab 2 - Risk Assessment This tab will consist of 8 columns with the same structured rows below. The columns will follow a similar pattern to that outlined for tab 1 as follows:

1. Step - this will be broken into numbers corresponding with the group of questions.
2. Applying CI framework to dataset - this will contain a brief overview of the corresponding groupings from tab 1, the information having been carried over to avoid the practitioner needing to repeat themselves e.g. data, information flow, actors etc.
3. Answers - here the practitioner’s responses from tab 1 will have been carried over where the question relates to facts, there will however, be further questions for the practitioner to answer
4. Identified Risks - this column will allow the practitioner to paraphrase and explain any risks identified from the facts relating to that question and/or category
5. Risk Assessment Score - here the practitioner’s score from tab 1 will be carried over and totaled. The practitioner will then be able to score the additional answers provided in this tab.
6. Mitigation Strategy - this column provides the practitioner with an opportunity to note any mitigations strategies that, if implemented, may reduce the risk
7. Post Mitigation Strategy applied risk Assessment score - Here the practitioner can reduce the risk score according if the identified mitigation strategy is implemented

Tab 3 - Decision

1. Step - this will be broken into numbers corresponding with the group of questions.
2. Contextual Integrity - this will hold two groups of rows of questions; decision to publish and decision to not publish
3. Decision - this will contain questions about the decision reached
4. Answers - here the practitioner will be able to respond to the questions

Unit of analysis - Scoring

A unit of analysis for the PAM questions will need to be devised. This will consist of a scoring system for each individual question in Step 1 to identify the level of risk associated with that question. The scores from step one will then be collated into step to and further scoring will take place to signify the level of risk each section poses for a particular dataset. These scores will then assist the practitioner with deciding whether the case for open source publication can be met.

There are many ways risk can be scored depending on the subject to be assessed and the risks identified, indeed there are standards for how to define risks **?**. For the purposes of this case study, defining the values of the scoring is no simple task as all the questions are subjective, very few are objective.

However, some of the fact questions can be objectively scored. For example, the questions relating to attribute types, here it is clear that, if identifiers are present in the data, this will receive a high risk score. Similarly, if there are quasi-identifiers or sensitive information in the dataset, the initial score will be high risk. Thus, the scoring will be a combination of numerical scoring using a scale of 1-10 with the score pre-set to 10 for identifiers and quasi-identifiers, 8 for sensitive attributes. The remaining questions cannot be pre-set to a score other than a scale of 0-10 as the answer will depend, not only on the data, but also how risk averse a particular public body or indeed data processor may be.

2.3.6 Stage 2 - Nine Steps

Once the overarching areas had been developed, the next step is to translate the nine step approach and breaking each step down into practical questions that will be asked.

However, first, some definitions of some of the terminology chosen and the reasoning behind that particular choice of terminology might prove helpful.

Actors

Nissenbaum refers to the people involved with the data as ”Actors”. There are three categories of actors; (1) ”information subjects”; (2) ”data senders”; or (3) ”data recipients”.

In practice, this will be the people who handle the data or about whom the data pertains. Therefore, it seemed more appropriate to relate this to the people who handle and data in a public body. The public body will likely relate to the definitions laid down in DPA and therefore, these will be referred to in PAM in accordance with established convention under The Data Protection Act 1998 (DPA);

1. The information subject(s), i.e. the person(s) the data collected pertains to. DPA uses the ”data subject” when referring to the ’information subject’. Thus, this is the term that will be utilised under PAM;

2. The data sender(s); here there are three Actors that are accountable, responsible and/or handle the data:

   (a) The ’public body’ (organisation). The reason for including the public body itself is that, in legal terms, the organisation is a legal entity (person) with rights and obligations**?**, and therefore, needs to be considered as, ultimately, the organisation is accountable if any decision to publish is challenged;

   (b) The ”data controllers”, i.e. the people who make decisions about the data; ’the data controllers’, those responsible for making decisions around what processing may or may not be permitted (DPA 1998, s. 1(1)); and

   (c) The ”data processors”, i.e. the people who process or handle the data on a day to day basis in accordance with the directions given by the data controller (DPA 1998, s. 1(1));

3. The ”data recipients”, i.e. the end users, those that download, manipulate and utilise the data once it has been published.

Information/Data

Practitioners store, control and manipulate the information that will be under consideration when applying PAM in databases, these databases will vary depending on the public body and the individual department but nonetheless, the information (data) will be held in a database of sorts.

In database management theory, data is held in tables. Tables contain rows (entity sets), columns (attributes) and fields (entity or tuple). Each column will hold a specific type of information (e.g. name, gender, etc.), these are known as attributes. The rows contain a set of related data, e.g. the name column might contain; Alice, Bob and Eve, these will all be in the same group or ’entity set’. The fields (entities or tuples) then contain the specific information pertaining to each attribute, e.g. Alice, Bob and Eve **?**.

In privacy terms, the attributes can then be classified according to how sensitive or identifying they may be. DPA refers to ”personal data” as ”data which relate to a living individual who can be identified” (DPA, s. 1(1)).

Thus, the attributes will be broken down into four types; Identifiers, i.e. data that can directly identify an individual such as their name; Quasi-identifiers, i.e. data that is not directly identifying but likely to be if linked **?**; Sensitive attributes, i.e. individual specific data that could aid in identifying an individual, such as disease or salary**?**; and Non-sensitive attributes, i.e. non-identifying, even if linked.

The handling of the data will be referred to as data processing. Under DPA, ”processing” refers to any handling of the data, be that collecting, manipulating, storing, processing or any other handling of the data (DPA, s. 1(1)).

Therefore, when referring to the data itself, it makes sense to use the terminology adopted in DPA and database management theory as this terminology will be familiar to practitioners.

Questions

Starting with Nissenbaum’s nine steps decision making approach; the ”Augmented Contextual Integrity Decision Heuristics” (p. 181-2), these were broken down into more detailed, specific questions.

This was done by taking the original nine questions and breaking each one down into, first, the relevant overarching category (phase), and then, elaborating by devising specific questions that could be applied in a practical setting. Each question and the thought process for the resulting question is explained in turn.

Question 1 **(Fact)**

”Describe the new practice in terms of information flows” **?**

As this relates to facts, about the data itself, this has been placed in Step 1, Facts. In order to understand the information flows however, it is necessary to first understand the data. What is the data about and what are the individual parts that make up the datasets. Then, questions can be asked about the existing information flows, i.e. how the data is currently processed.

Thus, the following questions were devised to cover this area:

- What is the dataset about?
- What attributes are in the data? - please provide a full list

For each attribute (set) within the dataset, please describe:

- Do any of the attributes contain personal information (identifiers)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain quasi-identifiable data (identifiable through inference data)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain individual specific attributes (sensitive data)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain other non-identifiable information (non-sensitive)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- How is data currently processed?

Question 2 **(Fact)**

”Identify the prevailing context. Establish context at a familiar level of generality (e.g., health care) and identify potential impacts from contexts nested within it, such as teaching hospital.” **?**.

This relates to the organisation, what type of organisation and in what context was the data collected. Again, these are facts about the organisational context so this was placed in Step 1, Facts.

The following questions were devised to cover this area:

- Where does data originate (Department)
- Is this data original to this dataset or was data collected from another department/processor
- If yes, who/which department does the data originate from

Question 3 **(Fact)**

"Identify information subjects, senders, and recipients" **?**.

Again, this relates to facts, this time about the people that handle or are the subjects of the data. Therefore, this was placed in Step 1, Facts.

The questions relating to these were phrased as follows:

Public Body:

- What is the public body that owns the information?
- What roles does the public body have? (e.g. library, local authority, University etc.)
- What is the relationship between the public body and the data subject?

Data Subject:

- Who/what is the data subject?
- What is the role of the data subject in relation to the data? (e.g. borrower, employee, citizen)
- If data subject is a person what relationship does this person have to the data processor and/or data controller? (none, co-worker, friend, family member, citizen, professional, client, etc.)
- In what context did the data subject divulge the information?
- Was there a legal obligation on the data subject to divulging the information?
- How does data subject interact with data processor/originator? (e.g. friend, co-worker, professionally, citizen, employment)

Data Originator:

The person who originally collected the information (sender). This would be a data controller but may be a different data controller to the one who is considering whether or not to publish the data, therefore, this has been included with a slightly different name to distinguish from any subsequent data controller(s) who may be making the decisions around the data):

- Who collected the data? (originator) (if multiple, please answer questions for each originator/controller)
- What is the role of the data originator in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)

Data Controller:

- Is data controller also data originator?
- Who is the data controller(s)? (if multiple, please answer questions for each controller)
- What is the role of the data controller in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)
- How does data controller interact with data subject? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data controller interact with data processor? (e.g. friend, co-worker, professionally, citizen, employment)

Data Processor:

- What is the role of the data processor in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)
- How may data processors are/have handling/ed the data?
- In what capacity was the data collected by the data processor(s)?
- What is the data processor(s) position in the organisation?
- How does data processor interact with data subject? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data processor interact with data controller? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data processor interact with data recipient? (e.g. friend, co-worker, professionally, citizen, employment)

Data Recipient (user)

- Who is the data recipient?

- What is the role of the data recipient in relation to the data? (librarian, information officer, employee, third-party partner employee, unknown etc.)
- What position does the recipient hold?
- How does data recipient interact with data subject? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data recipient interact with data processor? (e.g. friend, co-worker, professionally, citizen, employment)

Question 4 **(Fact)**

”Identify transmission principles” **?**.

When Nissenbaum discussed how the data is conveyed between the Actors, she refers to this as the ”transmission principles”. Nissenbaum contends that merely considering the data and the actors is not sufficient, consideration also needs to be given to how the data is transmitted and the context in which that transmission takes place. This context needs to include what restrictions may be in place for the dataflow and how this is managed and enforced. However, for the purposes of this study, the transmission principle will be open source publishing and therefore, a presumption that no restrictions on the recipients ability to re-use, process and manipulate the data once received will apply. What will be considered will be the format and type of data (dynamic or static) and how often it is proposed this is updated. Again, this relates to factual information about the data and how this is conveyed and therefore, this has been placed in Step 1, Facts.
The following questions will be asked in this category:

- How will the data be shared/published?
- In what format will the data be shared?
- Will data published be static or dynamic?
- If static, how often will it be updated?
- When updated, will the existing data be replaced or will version control or dates be used to denote updates?
- If data published is dynamic, will it be real-time?
- Does data contain images or video?

Question 5 **(Fact)**

"Locate applicable entrenched informational norms and identify significant points of departure" **?**.

This was interpreted as a requirement to consider, not just the social and data etiquette, but also any legal and ethical considerations that need to be taken into account in order to establish the prevailing context.

Arguably, this could relate to either fact or risk assessment as it relates partly to factual information about the data and the purpose of collecting it and partly to the surrounding circumstances which will form part of the risk considerations that will need to be considered. However, a decision was taken that this category is still primarily concerned with gathering facts about the data and thus, this was placed in Step 1, Facts. The following questions were devised to capture this category:

- What was the original purpose of the data collection/processing?
- What was the social context of the data collection? e.g. school would be educational context, council tax would be for tax collection etc.
- Was permission/consent sought for processing of the data from data subject?
- Was the data collected with a view to process beyond original purpose?
- If yes, what was that purpose?
- Was data collected direct from the data subject? (i.e. did they provide the information)
- Was consent sought for original collection/processing purpose?
- If yes, was consent granted for secondary processing purpose?
- If yes, was consent granted for specific secondary processing purpose?
- If yes, what was that purpose?
- Were there any limitations on secondary purpose to which consent was given?
- If yes, what were those limitations?
- Was consent granted for open re-use/sharing/processing?
- Are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?
- If yes, what are those considerations?
- Do any of these consideration have legal authority?

Question 6 **(Risk Assessment)**

"Prima facie assessment: There may be various ways a system or practice defines entrenched norms. One common source is a discrepancy in one or more of the key parameters. Another is that the existing normative structure for the context in question might be ?incomplete? in relation to the activities in question.... A breach of informational norms yields a prima facie judgment that contextual integrity has been violated because presumption favors the entrenched practice" **?**.

This was interpreted as an initial evaluation of the facts gathered in Step 1, in that it asks for thought to be given as to whether those facts in themselves have identified any potential risks or indeed, whether the context surrounding the collection of the data pose a privacy risk.

Therefore, this will be captured through gathering and scoring each sub-category (data; information flows; actors; transmission principles and context) so as to obtain an overview of the dataset and its privacy implications. Further, for each sub-category, risks and mitigation strategies were identified and listed. Finally, practitioners will be given an opportunity to amend the initial score if they decide to apply the identified mitigation and thus, reduce the risk. The risk scoring matrix will be based on ISO risk standards as these will

Question 7 **(Risk Assessment)**

"Evaluation I: Consider moral and political factors affected by the practice in question. What might be the harms, the threats to autonomy and freedom? What might be the effects on power structures, implications for justice, fairness, equality, social hierarchy, democracy, and so on? In some instances the results may overwhelmingly favor either accepting or rejecting the system or practice under study; in most of the controversial cases an array of actors emerge requiring further consideration" **?**.

This, in PAM, falls into Step 2, Risk Assessment. It has been placed here as practitioners are asked to start considering the decisions reached in question 6 further. This section refers to the broader context by asking practitioners to now also take into account individual rights, values and norms as well as the surrounding social, political and moral contexts. Thus, here, the following questions will be asked:

Assess disclosure risk:

- Disclosure risks identified above?
- Identify any disclosure control processes that may be relevant?

Roles:

- Is data subject aware of data being published?
- Has data subject consented to disclosure?
- Has data controller consented to disclosure?

Attributes:

- What would be the effect on the attributes published if these are linked to external data, would that pose a new risk?

Values:

- Would publication infringe on any political values?
- Would publication impose power imbalance and thus, infringe on any moral values?
- Would publication infringe on any social values?
- Would publication infringe on any moral values?

Norms:

- Would publication pose a threat to the autonomy or freedom of the data subject?
- Are there any belief systems that may adversely be affected by publication?
- Would publication result in any form of discrimination?
- Would publication result in informational harm on the data subject?
- Would publication result in breach of confidentiality?
- Would publication result in breach of trust?
- Would publication infringe on any legal compliance?
- Would publication impose any security risks?

Question 8 **(Risk Assessment)**

”Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context. In addition, consider the meaning or significance of moral and political factors in light of contextual values, ends, purposes, and goals. In other words, what do harms, or threats to autonomy and freedom, or perturbations in power structures and

justice mean in relation to this context?" **?**.

Here Nissenbaum asks that consideration be given to what potential consequences there might be if any of the Risks, roles or values identified in Question 7 were to occur. This may be either negative or positive. Thus, this forms part of the risk assessment and, as such, has been placed in Step 2, Risk Assessment. The questions here follow on and elaborate on the context and potential harm identified through the questions asked in the previous sections:

- Is there a reasonable expectation on the part of the data subject and/or data controller of data being kept confidential and not published?
- Are there any privileges or prerogatives that arise from processing or publishing the data from which the public body, the data processor, controller or originator may benefit or be seen to benefit as a result of publication?
- What are the positive values that publication will bring/enhance? These may include commercial gain, improved transparency, meeting legal obligation etc.
- Are there any overriding legal, moral or ethical reasons why publication should be allowed

Question 9 **(Decision)**

"On the basis of these findings, contextual integrity recommends in favor of or against systems or practices under study. (In rare circumstances, there might be cases that a re-sustained in spite of these findings, accepting resulting threats to the continuing existence of the context itself as a viable social unit" **?**.

The final question in Nissenbaum's nine steps relate to making a decision, thus, this has been placed in Step 3, Decision.

In PAM, this decision will be to publish or not to publish. Therefore, the questions have been designed to explain the decision taken, what further actions may need to be carried out prior to publication, who is responsible and who will be accountable. Further, if the decision is not to publish, PAM asks practitioners to explain the decision so that this may be used to defend or explain the reasoning, in case the information is requested or the decision challenged in future:

- In light of the findings what decision has been reached?

If decision is to publish:

- What work will need to be undertaken prior to publication if any?
- If decision is to publish part of the data rather than the raw data, please specify reason for this decision?
- What is the timescale for publication?
- What processes have/will be put in place for updating of the information?
- Who will be responsible for publication?
- Who will need to be notified of publication?
- Who will be answerable if any questions or challenges arise as a result of publication?

If decision is not to publish:

- What is the reason for non-publication? please provide a full statement including any legal reference where legal constraints form part of the reasoning. This statement should be copiable so that this may be used by data controllers and processors in defence of refusing any request for the release of the data

2.3.7 Phase 2 - Independent Review of PAM

Internal Reliability

Yin suggests that the best form of validation is to triangulate, i.e. seek review from three independent reviewers, each from a different perspective **?**. Thus, to validate the questions and ensure these are appropriate, non-bias and in line with Nissenbaum's CI framework, PAM will be reviewed and validated by three independent parties, each of whom will review PAM from a different perspective:

Solicitor

The legal aspect of the framework will be reviewed and evaluated by a solicitor who is currently working within a public body with experience of freedom of information and data requests from the public. In addition to commenting on the questions themselves and their suitability, the solicitor was also asked to identify any areas that, if not addressed or included (if omitted), could potentially leave a public body open to challenge if publication was refused.

Practitioner

The practical aspect of the framework will be reviewed and evaluated by a practitioner who works with, and makes decisions around, public sector information on a daily basis. To ensure there was no bias from the local authority participating in the case study, an Information Governance Officer from a different local authority will be approached to evaluate the questions. This practitioner was asked to review the questions for suitability and, in addition, to consider these from a practical perspective were his local authority to apply them in practice and comment on their practical relevance and applicability.

Academic

The Academic will be asked to review and evaluate the questions from a suitability perspective and from a methodological perspective. The academic approached is an experienced researcher with a specialism in methodology and the research process.

With regards to suitability of the questions, each of the reviewers will be asked to highlight and identify:

1. Any areas/questions that had been omitted;
2. Identify areas that required further explanation or elaboration;
3. Check that questions followed logical progression and where in appropriate sections;
4. Any areas that may be difficult to comprehend; or
5. Any other comments or observations

2.3.8 Phase 3 - Existing Formal Model Evaluation

The final phase in the creation of PAM will be to conduct a comparison between PAM and an existing formal application of the CI framework. For this evaluation, the Formal Model of Contextual Integrity created by Barth et. al. will be utilised **?**.

2.4 Data Collection

The data will be collected through contextual interviews. The practitioners taking part in the trial will be supplied with a spread sheet containing the framework questions and asked to apply these to real data under consideration for open source publication. As they apply the framework the interviewer (the primary researcher) will observe and make notes as to the thought and decision processes that the practitioner applies. However,

due to distance and time constraints these interviews will be conducted through a series conference calls via telephone and Skype with the local authority.

For validation, the answers noted by the practitioners on their spread sheet will be compared with the answers noted by the interviewer during the observation itself.

2.5 Analysis

The first part of the analysis will be a detailed analysis of the existing guidelines produced by Government bodies and affiliated parties to assist practitioners when making decisions about open source publication. This will be done following Corbin and Strauss grounded theory methodology **?** to establish how much of that guidance pertains to privacy. Then, the results will be amalgamated to confirm or deny hypotheses 1.

For hypotheses 2-4, the case study data will be analysed. Once the data has been collected from the case study pattern matching will be used to analyse the data. Pattern matching involves comparing the hypothesis to the case study data and analysing the results in order to answer the research questions 2- 4 **?**.

Chapter 3

Methodology - Applying CI to Open Data

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.1 Section

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur

Figure 3.1: Bournemouth University

sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.1.1 Subsection

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud ex-

ercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 4

Methodology - Applying CI to Open Data

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.1 Section

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur

Figure 4.1: Bournemouth University

sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.1.1 Subsection

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud ex-

ercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 5

Evaluation

Chapter 6

Discussion

Chapter 7

Conclusion

Bibliography

Appendices