Privacy risk assessment in context: a meta-model based on contextual integrity.

HENRIKSEN-BULMER, J., FAILY, S. and JEARY, S.

2019

COMPUTERS & SECURITY 82 (2019) 270–283

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Privacy risk assessment in context: A meta-model based on contextual integrity

*Jane Henriksen-Bulmer**, *Shamal Faily*, *Sheridan Jeary*

Department of Computing and Informatics, Bournemouth University, Poole BH12 5BB, UK

ARTICLE INFO

Article history:
Received 7 March 2018
Revised 21 November 2018
Accepted 6 January 2019
Available online 11 January 2019

Keywords:
Privacy
Privacy risk
Contextual integrity
Meta-model
Open data
Data
Case study

ABSTRACT

Publishing data in open format is a growing trend, particularly for public bodies who have a legal obligation to make data available as open data. We look at the privacy implications of publishing open data and, in particular, how organisations can make informed decisions around privacy risks in relation to open data publishing before publication occurs.

Using a well established theoretical privacy assessment framework, Contextual Integrity, we illustrate how this can be translated into a practical meta-model that can assist public bodies in assessing what privacy implications or risks might be associated with making a particular dataset available as open data.

We validate the meta-model by providing a worked example and illustrate the effectiveness of this by reference to a case study application where the meta-model was successfully applied in practice.

1. Introduction

Privacy impacts many aspects of our life. At one time, privacy was mainly concerned with who divulged what and to whom. However, while our expectations about how confidential information will remain if shared with friends, family, or even social media are likely to be met, this may not be the case with Government departments (*public bodies*). Public bodies might collect data about our property values and assign this to our name and address. These datasets contain private or sensitive information about us and we accept that government bodies need our personal information to conduct their business, and willingly provide details of our lives; such details range from registering the birth of our children to providing details of our households on census days. However, while at one time it might be assumed that any information gathered by a public body would only be used for the purpose for which it was originally collected, the introduction of Open Government and open data have changed this.

Open Government seeks to increase transparency and citizen participation and collaboration through making data easily accessible (Obama, 2009). To this end, so far, more than 70 countries have made a commitment to open government and making public data available as open data (Open Government Partnership, 2016). Open data is data that is freely available for anyone to download, share and re-use with no restrictions on re-use or re-distribution save for perhaps a requirement to reference the source (Open Data Institute, 2016). The issue is that, in meeting this commitment, public bodies need to be sure that any data they publish in open format does not contain personal or sensitive data, i.e. data that identifies or can be used to aid in identifying individual citizens.

We contend that while making data freely available is, in principle, a very good idea, it does raise questions about what safeguards will be placed on publishing such data. With

* Corresponding author.
E-mail address: jhenriksenbulmer@bournemouth.ac.uk (J. Henriksen-Bulmer).
https://doi.org/10.1016/j.cose.2019.01.003

cyber crime on the increase, Privacy Rights Clearinghouse report that nearly 8000 data breaches have been made public since 2005, thereby exposing in excess of 10 billion records (Privacy Rights Clearinghouse, 2018).

Existing guidelines offer direction on what public bodies should publish, types of data, and their format. They do not, however, provide much privacy guidance beyond ensuring that data protection regulations are adhered to. This lack of 'usable' guidance in relation to privacy leaves public body officials with little direction for how best to proceed when publishing their datasets as open data.

We seek to address this gap by demonstrating how privacy can be incorporated into the decision making process for open data prior to publication. In this domain there is an added complexity in that, in most cases, any data released as open data by a public body is likely to be historic data. Therefore, from the perspective of the practitioner who determines whether or not a dataset is suitable for publication, he or she is not necessarily concerned with how the data was collected, processed or stored originally, nor the security risks associated with those practices. Rather, the concern of the practitioner will be to assess the datasets' suitability for publication in open format, and then perhaps, based on that decision, he or she may look at how the data can be safeguarded going forward. However, for the purpose of this paper, the technical aspects of collecting and safeguarding the data are not part of the consideration.

This paper makes three contributions. First, we present a conceptual meta-model for identifying and assessing privacy risk based on Contextual Integrity (CI), a theoretical privacy framework created by Nissenbaum (2010). This model was devised to help practitioners make privacy decisions in open data publishing, thereby providing a visual aide-memoire for practitioners for how CI can be applied in practice. Second, we illustrate the rigour and application of the meta-model by applying the concepts discussed to a hypothetical public body, a public library, in a worked example. Third, we evaluate the model by applying it in a practical case study working with real data in a real setting (Henriksen-Bulmer and Faily, 2017), thereby demonstrating the significance and effectiveness of the framework and model.

The rest of this paper is structured as follows. We begin with a brief review of our findings on UK public bodies attitudes to open data publishing, based on a series of contextual interviews in Section 2. We then review the related work in Section 3 by outlining existing legal privacy protection, considering how organisations make decisions around risk and privacy, and considering why context needs to form part of any privacy assessment. In Section 4, we present a meta-model for Contextual Integrity (CI) (Nissenbaum, 2010): a visual representation that demonstrates how CI can be applied in practice, which will initially evaluate using a worked example in Section 5. In Section 6, we evaluate the meta-model by using to inform a real world practical application of CI, working in collaboration with a UK Local Authority (Henriksen-Bulmer and Faily, 2017). We discuss the implications for design of our designs in Section 7, before outlining the limitations of this work and directions for future work in Sections 8 and 9, respectively. Finally, we summarise our findings and contributions made in Section 10.

2. Public bodies attitude to open data publishing

To establish to what extent public bodies currently publish data in open format, we submitted a freedom of information (FOI) request to 22 local authorities (LA) covering three regions in the UK (Henriksen-Bulmer and Faily, 2017). The FOI request asked whether; (i) the LA published open data; (ii) the LA had a open data portal or website; (iii) the LA contribute to the national open data portal; (iv) who was responsible for open data publication; and (v) what their role is within the LA. We found that all published something, but only 37% had some form of open data platform or portal. We also found the role and/or department responsible for open data publishing varies considerably across the LA's contacted. As a result, many practitioners responsible for open data publishing, undertake this as a secondary responsibility, i.e. in addition to their primary role (e.g. legal officer also responsible for open data).

In addition, to gauge how widely open data publishing has been adopted by public bodies, we conducted contextual interviews (Beyer and Holtzblatt, 1998) with 4 practitioners from two public bodies. We found that while the positive aspects of open data publication were acknowledged, e.g. one practitioner commented: *"by opening up information to people you can foster growth"* (P1). The norm in practice appears to be not to publish until pressure dictates otherwise: *"Why do we have to do anything when we can get away with the bare minimum?"* (P3).

One reason put forward for non-publication was the fear the privacy could be compromised by publication, one practitioner stated *"I am almost convinced that if I went back through our data that we have published over the last 4–5 years, I would find something that we'd missed* [referring to personal data being inadvertently published]" (P2). This may explain why only 37% of LA's currently have an open data portal. We therefore contend, that if privacy is to be incorporated into corporate practice, it needs to become an integral part of organisational decision making (Bamberger and Mulligan, 2015). Based on these findings we decided to examine what legal regulations are in place to safeguard privacy and what privacy assessment frameworks are available to practitioners that they can use to inform decisions around data privacy.

3. Related work

3.1. *Data protection as a privacy safeguard*

Any data processing is subject to adherence to data protection laws. In the UK these can be found under the Data Protection Act 2018 (DPA) (UK Parliament, 2018), the UK's enactment of the General Data Protection Regulation (GDPR). This act governs what the public body and their employees may do with data pertaining to individuals. Under DPA the people involved with handing the data are referred to as *data processors* who maintain and manipulate the data, and *data controllers*, who are responsible for making decisions about the data and whether or not the data may be shared. Data processors and controllers must adhere to strict data protection principles and ensure that personal data is processed: lawfully, fairly and

in a transparent manner; that the minimum necessary data is collected and processed and for the specified purposes only; accurately; kept for no longer than necessary; and appropriately safeguarded.

Despite this, two obvious issues remain. First, while data protection laws may dictate that personal information needs protection, the safeguarding of this protection is restricted to what the data controller and/or processor does with the data. Provided they do not publish *identifying* information such as names, addresses or dates of birth, that obligation has been met. Second, data protection is one-dimensional; what constitutes personal data is limited to the sensitivity of the individual attributes within the data. Once data has been anonymised and data subjects are no longer identifiable, data is no longer subject to data protection, as it is no longer considered personal or sensitive (ICO, 2012). Previous work has shown that anonymisation can be reversed (El Emam et al., 2011), particularly when multiple datasets are aggregated or linked (Henriksen-Bulmer and Jeary, 2016). This makes the release of public body data in open format potentially incompatible with data protection laws (Kulk, 2012).

3.2. *Data, organisational decision making and risk*

The findings from FOI requests and initial discussions with public servants appear to suggest that open data predominantly consists of existing data. This means that, for the most part, any safeguards that are in place around data processing and release have been pre-determined prior to the open data practitioner needing to decide whether a particular dataset should or should not be published. Therefore, what needs to be reviewed is not the safeguards already in place or how to manage these, rather, it is the decision making process to determine the suitability or not of publishing data in open format.

A popular method for informing decision making within organisations is through risk assessment which has been used to define level of exposure, cost or potential impact of a decision, or 'alternate solutions' to a problem with an organisation (Simon, 1955). Risk is an attempt to define the uncertainty in more practical terms. Indeed, it has been contended that risk "is inseparable from decision-making" (Galanc et al., 2016). For example, the way business looks to protect data is through security measures and conducting security risk assessments, using frameworks from recognised international bodies such as the US National Institute of Standards and Technology (NIST), or the International Organisation for Standards (ISO), who have produced guidelines for assessing security risks (BS ISO 27000:2017, 2017; NIST, 2012) and, as part of such security risk assessment, privacy risk (BS ISO/IEC 29100:2011, 2011; NIST, 2010). The problem is that these measures in themselves may not be enough. For instance, a business might seek to make accessing the data difficult by password protecting the PC or the data, which has given rise to usability issues and therefore proved ineffective in many instances (Sasse et al., 2001). Additional layers of protection can be added by for example, obfuscating or anonymising the data (Samarati, 2001). The problem here being that research has shown that anonymisation can be reversed (Henriksen-Bulmer and Jeary, 2016; Ohm, 2010).

We contend this approach alone forms part of the problem because, what organisations are actually trying to protect with these security measures, is to safeguard the data and data privacy, and therefore, privacy risk. Currently privacy risk is considered as one element of security risk, meaning privacy risk becomes a secondary consideration that may result in inappropriate or ineffective security measures being put in place to protect against it. By viewing the data separately though a privacy risk lens, there is a better chance of correctly identifying each privacy risk so that the most appropriate security measure can be applied for that risk.

One framework that looks specifically at privacy risk is the Privacy Impact Assessment (PIA), used in multiple countries for assessing general privacy risks (David et al., 2013). However, the PIA is predominantly geared towards assessing privacy risks to a new project, process or system so that appropriate mitigation and security strategies can be incorporated into the design and/or implementation. However, it does not look at existing data or processes. More recently the new Data Protection Impact Assessment (DPIA) introduced as part of GDPR has widened the scope to make DPIAs compulsory (GDPR, Article 25) for any high risk data processing. Whilst the DPIA may include looking at privacy risks for existing processes or data, it asks practitioners to consider privacy risks, not from the perspective of the organisation but from the perspective of the individual (i.e. the data subject) and how the risk might impact the data subject.

3.3. *Privacy risk in context*

Attempts have been made to consider privacy risk to individuals in terms of setting privacy goals as part of conducting DPIAs. For example, Bieker et al. have sought to incorporate context into the privacy risk assessment process (Bieker et al., 2016), thereby demonstrating the importance of including context in any privacy risk assessment However, this work looks at privacy risks in terms of identifying potential threats, akin to security and privacy threat modelling schemes e.g. IRIS (Faily and Fléchais, 2009) and LINDDUN (Deng et al., 2011), which focuses on *new* processes, systems and projects to ensure safeguards are in place, rather than assess what data is already there and whether this is safe to publish.

Other scholars that have considering context as part of decision making around privacy, adopt slightly different approaches. For example, Mulligan et al. look at privacy risk in terms of risk of harm (Mulligan et al., 2016), while Solove consider all aspects of all the areas where a person's privacy might be compromised or breached, including a number of concepts that relate to an individual's physical environment as well as informational privacy (Solove, 2006). What is not considered in any depth as part of these frameworks is the human element, how people behave and perceive privacy and how the context within which data is shared may be affected by those behaviours, values and norms.

An alternative framework is contextual integrity (CI), a theoretical privacy taxonomy that incorporates these nuances including the human element (Nissenbaum, 2004). CI asks that practitioners consider the: people, i.e. *actors, i.e. the sender, receiver and subject (the data subject) of the data*; data *attributes* and data flows *transmission principles* as part of the privacy

assessment and that these elements are considered within context (Nissenbaum, 2010).

This framework provides an opportunity for privacy to be considered strategically *before* technical intervention, and thus, we decided to use this framework as a basis for creating a conceptual model (meta-model) that can be applied to open data publishing.

3.4. *Existing applications of the CI framework*

Previous research using CI has predominantly been defined and discussed in theory (Grodzinsky and Tavani, 2011; Sar and Al-Saggaf, 2013). Other work has considered public bodies releasing data in open format and privacy, notably, the Berkeley Law and Technology Center, who have published papers discussing this domain, e.g. Berkeley Law (2016), none of these however, have applied CI in this context. Where CI has been applied in practice, it has been applied retrospectively as a theoretical discussion to a specific problem or domain (Conley et al., 2012), or it has been applied as part of a system design process (Barth et al., 2006; Krupa and Vercouter, 2012). For example, in one study CI was applied through tags attached in the message headers (Krupa and Vercouter, 2012), while Barth et al. (2006) used CI to devise a system for controlling data flow between user roles so that the system could compute access controls.

In doing so, what these studies have done is to apply CI at a granular level, rather than holistically as part of an overarching privacy decision-making process. Thus, although previous work illustrates how the CI framework can be applied, these studies are insufficient for guiding decision-making in open data publishing. In this domain, there are a number of constraints and unique considerations that need taking into account. First, the data being assessed will be existing data that has been created as part of a function of the public body unrelated to open data publication. Therefore, any decisions about what data and how this data has been collected, processed and stored will be historically pre-determined, and the decision-maker will most likely not have been involved in decisions around these factors. Second, for those datasets it will not be possible to fully define all elements because the role of the recipient cannot be specifically defined; anyone who downloads the data will be a recipient, and the data, once published, will be available to everyone.

4. Creating the meta-model

The idea of creating a conceptual model of contextual integrity (CI) is to provide public body practitioners with a visual aid that will depict CI at a glance. It will illustrate how a staged approach to determining whether or not a particular dataset contains potentially privacy sensitive information and/or could present a risk of privacy being violated and thus, can assist in making informed privacy decisions. Moreover, modelling CI can also be used to inform the development of potential privacy risk decision-making tool support that may help automate and streamline some of this process for practitioners going forward. The meta-model is intended as a practical, usable tool that will guide practitioners through CI and the privacy assessment process, culminating in a decision, to publish or not to publish. For validation, an explanation is provided at each stage for how the meta-model has been aligned with Nissenbaum's framework (Nissenbaum, 2010).

The modelling technique that will be used to visually represent how CI can be applied in practice is the Unified Modelling Language (UML). UML diagramming is a universal visual language that is used to capture and represent concepts and the relationships between them (Rumbaugh et al., 2004) and therefore, this method has been chosen to visually represent and illustrate each phase of the meta-model, thereby providing readers with an easy point of reference and a better overview of how the elements relate within each phase (Fowler, 2004).

To create the meta-model, two guiding sets of principles from Nissenbaum's framework were identified as being core to applying CI in practice:

Key elements: The meta-model uses *Explanation, Risk Assessment and Decision* in place of Nissembaum's *Explanation, Evaluation and Prescription* (referred to by Nissenbaum as the "3 Key Elements"; Nissenbaum, 2010) to better align with terminology that practitioners will relate to and to aid the flow of the staged approach that the meta-model will be asking practitioners to follow. These three elements have been used as overarching themes in the meta-model to frame the logical progression of the model into understandable, logical progression phases.

Decision heuristics: Nissenbaum proposes 9 decision heuristics (DH) that should be considered in relation to both existing and proposed new information flows. These have been used within the phases to delineate and expand on the areas that need to be be considered in practice and help establish whether privacy is likely to be, or has been, breached by a proposed new flow of information (Nissenbaum, 2010).

4.1. *Phase I – explanation*

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 receivers'. It should also consider the "open dataset" ('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 (see Fig. 1).

To incorporate the DH's into the meta-model, it was determined that the first four DH's relate to gathering a more detailed overview of the data; the people (actors); the existing informational norms and transmission principles. The first decision heuristic (DH1), concerns the data itself and how it is proposed the data is to be transmitted. The second asks us to consider the existing context of the situation and environment surrounding the data and the people involved (DH2). The third, concerns the people involved with the data (DH3); and the fourth, seeks to establish how the data is currently

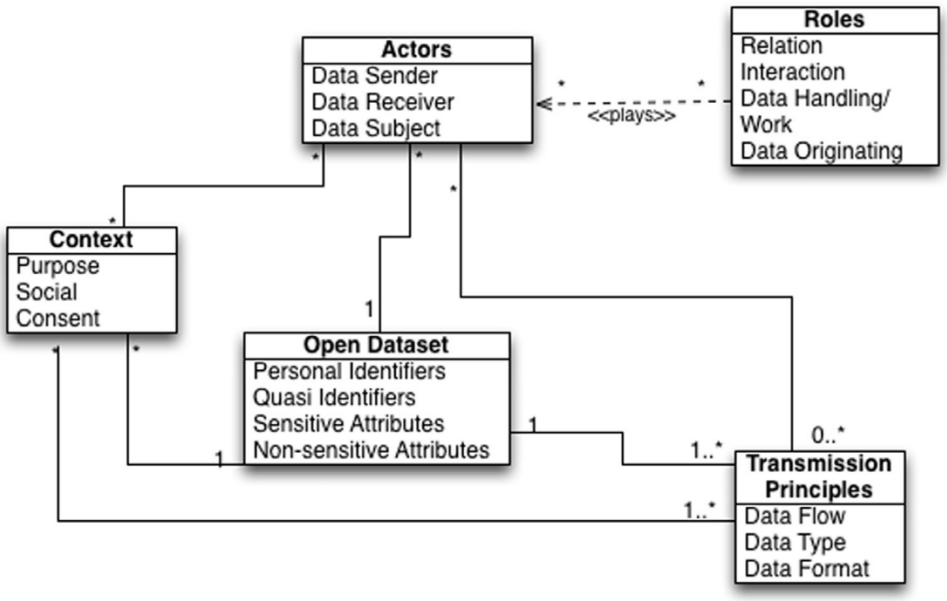

Fig. 1 – Explanation – class relationships.

transmitted (DH4) (Nissenbaum, 2010). Thus, these four DH's were used to depict the explanation elements and how they relate. At a more detailed level, explanation therefore requires that details are collected about the data itself, the people involved and their roles, how the data is transmitted and the context. Thus, the explanation is the superclass with each of the elements below depicted as subclasses (see Fig. 1).

Fig. 1 shows the relationship between the subclasses which can be explained as follows:

Open dataset: Each dataset needs to be considered separately to avoid overcomplicating the decision-making process and ensure all elements are thoroughly considered and CI is maintained throughout the assessment of each dataset. Within the dataset will be the individual data items *the attributes*, grouped by attribute type: *Identifiers* i.e. data that can directly identify an individual, such as their name, national insurance number or date of birth; *Quasi-identifiers* i.e. data that is not directly identifying but likely to be if linked, e.g. age or gender (Thomson et al., 2005); *Sensitive attributes* i.e. individual specific data that could aid in identifying an individual, such as ethnic origin, religious beliefs, disease or salary (Fung et al., 2010); and *Non-sensitive attributes* i.e. non-identifying, even if linked.

Actors: Each actor will act in one or more capacities. At data level, the actor will perform a data transmission role as either sender, receiver or subject of the data being transmitted. It is also possible that an actor can take on multiple roles. For example, the data sender or the data subject may also download the data and thus, also become the receiver. Beyond the data transmission role however, the actor will also perform multiple relationship and/or work roles. Thus, to allow for these nuances to be taken into account, the roles have been separated out as a class of its own.

Roles: with each actor acting in one or more roles in relation to the data. Therefore, each actor may be associated with multiple roles, which may be based on relationships, context or duties as follows: *Relations* i.e. details of relationships between the actors, e.g personal and/or professional; *Interactions* i.e. information about how the actor(s) interacts e.g. citizen to professional or friend to citizen; *Data Handling* i.e. information about actor input(s)/output(s) in handling the data; *Work* i.e. the occupation of the actor, this may be their job title or their profession (e.g. lawyer, customer service officer etc.); and *Data Originating* i.e. details about the role of the data originator who may be third party and thus, the role of that third party needs to be considered as well.

Context: Capturing information about the context of how and why the data was captured in the first place (the *prevailing context*). These contexts are: *Purpose* i.e. the original purpose of why the data was collected; *Social* i.e. the social context in which the data was collected (e.g. benefits department = tax collection/payment context); *Consent* i.e. whether consent has been obtained and, if so, the validity of consent obtained also needs to be considered.

Transmission principles: Referring to the data flows between actors.

4.2. *Phase II – risk assessment*

The second key element, *risk assessment*, consists of an evaluation of any privacy risks associated with a particular practice or transmission within a given context, taking into account how the information is conveyed or shared and the actors involved with that practice or transmission. Effectively, what the risk assessment is trying to achieve is an assessment of the risks associated with any proposed changes or alterations in the data flow. A privacy risk is "the probability" or "likelihood"

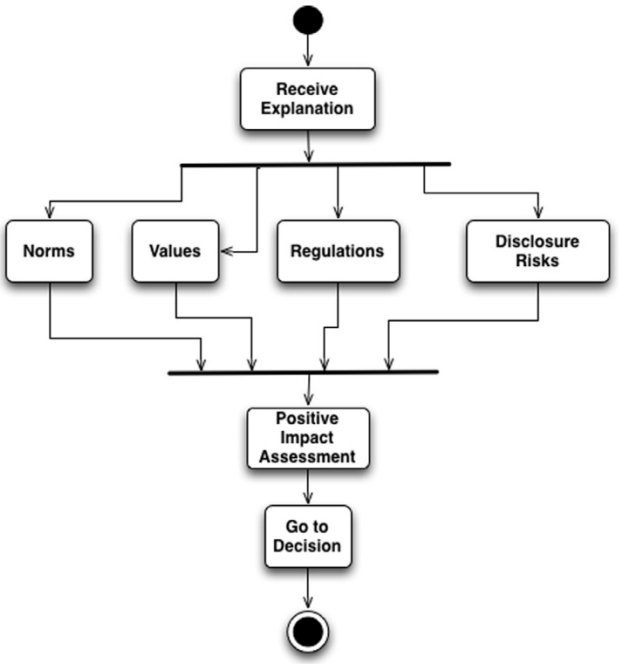

Fig. 2 – Activity diagram: risk assessment.

and "consequence" of a loss of, or violation of, an individual's privacy (BS ISO 31000:2009, 2009; BS ISO/IEC 29100:2011, 2011). Thus, in the meta-model a privacy risk can be defined as a disclosure risk, e.g. the risk of re-identification occurring if data is released in it's current format.

The risk assessment phase considers DH numbers five to eight, as these all relate to the evaluation or risk assessment of how the proposed new information flow will affect the privacy of the actors (Nissenbaum, 2010). Fig. 2 shows the information gathered in the explanation phase from the; data, actors, transmission principles and prevailing context feeding in to the risk assessment.

The risk assessment asks for the following aspects to be considered:

Norms: These can be described as the standards or rules by which we conduct our lives, formal or informal. In terms of CI, the *informational norms* are what must be considered, i.e. the norms that the actors will be expected to abide by in the capacity of their role, for example, a teacher may divulge a student's performance record to the student or their parents but would not be expected to divulge the same to other parents within the school. The informational norms need to be considered in relation to the evaluation criterion to determine whether they the proposed change in data flow could result in any infringement or breach of: Autonomy/Freedom; Beliefs or belief systems e.g. religious, political or strong opinions; Informational harm; Discrimination; Confidentiality; Trust; or Security.

Values: These are "the objects around which a context is oriented" (Nissenbaum, 2010). These may be social, political or ethical values that could be affected or altered as a result of the proposed new information flow, asking whether it is possible that the proposed new data flow could impose an imbalance of some sort and thereby infringe of one or more of these values. These values will, of course, need to be considered in light of the norms and any legal or regulatory constraints or obligations.

Regulations: Referring to any legal obligations that may be imposed on the public body in relation to the data. This may be constraints such as DPA regulations, or obligatory, such as the Re-use of Public Sector Information Regulations 2015 (ROPSIR) which require public bodies to publish data in open format.

Disclosure risk: This looks at the information from the explanation phase and asking what the disclosure risk will be in light of: the entrenched informational norms; any points departure from the entrenched informational norms identified; whether consent has been granted and, if so, to what extent consent has been granted; the impact of any potential breach; any proposed mitigation strategies put in place; and any proposed controls in place or to be put in place.

Once all of these considerations have been taken into account, the final step of the risk assessments involves conducting a positive impact assessment. Assessing the potential positive impact is intended to capture relevant changes that, while they may represent a perceived "deviation from entrenched norms" or values, could actually have a positive rather than negative impact (Nissenbaum, 2010). Take for example, a new technology innovation developed such as when the smart phone was first introduced. At the time, many people, including the lead researcher, felt the use of a smart phone posed too much of an infringement on personal privacy as users would be permanently contactable and locatable. However, most users now accept this fact and being contactable all the time could be argued to now be an "entrenched norm". Thus, this shows that even entrenched norms and values may be subject to change over time, illustrating the importance of assessing both positive and negative impacts. Therefore, the final step entails assessing:

Positive values: This should detail any positive values that publication will bring such as improvement in transparency or commercial gain.

Overriding values Outlining any overriding reason why publication should go ahead, such as a legal obligation to publish the data.

4.3. *Phase III – decision*

The third key element, *decision*, relates to the findings on whether a practice violates privacy. This 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. In this phase, the final heuristic (DH9) has been used to ask whether, based on the findings made in the previous considerations; "contextual integrity recommends for or against the proposed new practices" (Nissenbaum, 2010). In the decision phase there are only

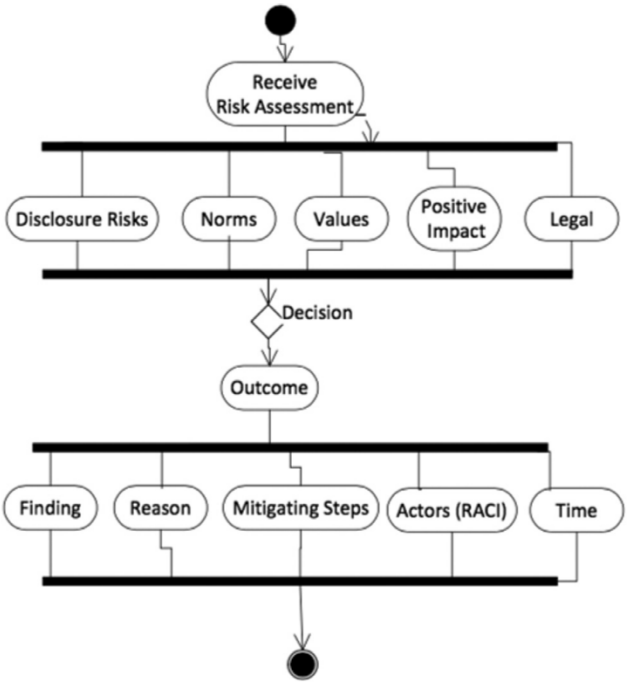

Fig. 3 – Decision diagram.

two stages. The decision itself and recording the outcome, the activity flow for this phase is depicted in Fig. 3.

The decision process consists of the following:

Risk assessment: The Risk Assessment will be carried over from Phase II (as discussed in Section 4.2). This will feed into the decision element to produce an outcome. This outcome will hold the decision from each attribute group, contributing a score to aid the practitioner in making a decision.

Outcome: Once the decision has been made the outcome will need to be recorded and any actions identified as follows: *Finding* i.e. the decision itself, i.e. a 'to publish' or 'not to publish' decision; and *Reason* i.e. the justification and reasoning behind the decision made, this could, for example, refer to legal compliance such as 'Data Protection' or a 'no privacy issues found' reasoning.

If the decision is to publish the following additional attributes will also need to be completed:

Mitigating steps This category will detail any mitigating steps that need to be carried out before publication can take place, this may, for instance, include redaction or anonymisation.

Actors: Recording who is responsible, accountable, consulted and informed (RACI) going forward as part of the process helps achieve transparency and provide assurance that proper process is followed in making, implementing and enforcing decisions made. Thus, part of the record involves completing a responsibility matrix (Project Management Institute, 2004) that outlines who is: *Responsible* for publication; *Accountable* depicting who is accountable if the decision is challenged or there is a problem (there can only be one person accountable in a RACI matrix); who must be *Consulted* and *Informed* of the decision.

Time: Enabling details of how regularly the open data published will be updated (if relevant).

Table 1 – Library lending register – extract.

Customer ID	Name	Address	Books on loan
12345	Alice Smith	A Street	5
23456	Bob Jones	B Street	1
34567	Eve Evans	E Street	2

5. Worked example

This section seeks to provide an illustration of how the meta-model can be applied in practice by providing a worked example for each concept discussed above. This will take the form of a public body practitioner ("PB"). To give the PB some context we will give him the role of Data Officer and have him employed at the local Lending Library, applying the concepts to ascertain what privacy risks a hypothetical dataset, the "Library Lending Register", will pose if it was to be published in open format.

5.1. *Explanation*

For the explanation phase, the PB will need to capture details of the Library Lending Register (i.e. the 'Open Dataset' in Fig. 1).

Following the meta-model, to conduct the privacy risk assessment, the PB will need to capture which attribute groups (or columns) are contained within the dataset being assessed. For example, if the dataset being assessed is the Lending Register of books on loan, the PB will need to capture details of each attribute type (column within the database) which might include Customer ID, name, address and no of books on loan (see Table 1).

What the PB will need to capture is details of each column header (attribute type) so that each attribute type can be assessed for potential privacy risks, i.e. for each group of attributes, does that column contain attributes that are personal identifiers, quasi identifiers, sensitive or non-sensitive? In this example, the name will be a direct identifier; the customer ID and address will be quasi-identifiers (they can link back to the personal information if linked) and the number of books will be non-sensitive. Thus, most of these columns contain potential identifying information.

The PB will also need to capture details of who has been involved in handling the data and what their role(s) are (i.e. the 'Actors' and 'Roles' in Fig. 1). This will involve looking at which department the data originated from, who works there and particularly, who worked with the data within that department (the data senders and data receivers). For the Library this might include a Librarian and a Learning Technologist. In addition, the PB will need to ascertain who the data controller responsible for the dataset is, and for context, the PB will need to capture details of how these people relate to each other and the data subjects (i.e. Alice, Bob and Eve, the library

customers), e.g. do they know them personally, are they related? etc. The PB will need to capture how the data is transmitted, what format it is held in, e.g. is it in a spreadsheet, a bespoke library lending database etc. Finally, the PB will need to record how the data flows both within the organisation and externally (i.e. the 'Transmission Principles' in Fig. 1).

Once the PB has captured details of the data, actors, roles, transmission principles and the prevailing context, risks can be identified in light of legal obligations, established norms and values. To this end the LA will need to determine how publishing the data in open format might affect the transmission flow and what privacy risks might be associated with this new flow of data. This is captured in phase two, the risk assessment.

5.2. *Risk assessment*

The risk assessment phase is where the PB will assess the privacy risks associated with making the dataset available as open data. This will involve reviewing the information captured as part of the evaluation and identifying any risks there might be associated with the data (see 'Disclosure Risk' in Fig. 2). For example, if the Librarian and Eve are friends, the PB would need to note this relationship as a potential risk as part of completing the risk assessment. For each of the risk assessment areas the PB needs to note any associated privacy risks. The fact that these two actors are friends could have potential risk implications in a number of areas, meaning the PB will need to consider the risks associated with each area:

Disclosure risk: There could be a number of potential disclosure risks identified as a result of the friendship between the actors in this instance. For example, this could include the relationship between the actors giving rise to a consideration about who the librarian may share the data with and how appropriate such sharing may be. It should also consider what the repercussions would be if the librarian was to divulge information obtained in the course of their work to a third party such as another friend. Similarly, consideration will need to be given as to whether or not Eve has given consent to the data being processed and what that consent covers. Another risk associated with the friendship could be that the Librarian may divulge personal information about Bob (another library user) to Eve.

Norms: The friendship could result in Eve receiving preferential treatment such as being allowed extra books on loan (discrimination risk) or be privy to confidential information about Bob (trust and confidentiality risks).

Regulations: Adherence to data protection regulations. For instance, if Bob has not given consent for his data to be shared, the divulging of the information to Eve will constitute a breach of data protection regulations.

Values: A breach of confidentiality would infringe on social and ethical norms (see Fig. 2).

Once all of the disclosure risks have been identified in light of legal constraints, norms and values, the PB will need to determine whether there are any mitigating steps that can be taken to make the data available in open format. Then, the PB can make an assessment of any associated privacy risks in publication and use this to assist in making an informed decision as to whether the Library Lending Register can be published. This is captured in phase three, the decision.

5.3. *Decision*

The decision phase for our Lending Library is where the PB will assess the privacy risks associated with making the dataset available as open data (see Fig. 3) and, based on this, make an informed decision ('Decision' in Fig. 3). As part of recording the 'Outcome' in Fig. 3, the PB will record the decision and the outcome of the assessment. This should outline what the finding from the assessment is; the reasoning behind the decision and the mitigation steps identified and whether or not these were applied and who will be responsible for what aspects of publication etc. (RACI) going forward. For the Privacy Risk Assessment carried out on the Library Lending Register, the mitigating steps could, for example, include:

- *Anonymisation* the PB could advise that identifying attributes should be anonymised prior to publication (Lablans et al., 2015; Samarati, 2001).
- *Redaction* the PB could recommend that personal identifiers such as names, be redacted prior to publication (ICO, 2012; Pfitzmann and Hansen, 2010).

Once these steps have been completed the PB will have a detailed record of the outcome of the privacy assessment that includes the finding and the reason for the decision. This will enable other practitioners within the public body to refer to the decision made and provide the organisation with quality assurance and an audit trail of decisions made.

6. Applying the meta-model in practice using CLIFOD

To evaluate the effectiveness of the meta-model, we applied the principles to a real-life scenario in a case study, working in collaboration with a UK Local Authority (LA) that publish open data regularly (Henriksen-Bulmer and Faily, 2017). For this case study, we expanded on the concepts developed in the meta-model by creating a step-by-step questionnaire for assessing the privacy risks of publishing open data: ContextuaL Integrity For Open Data (CLIFOD). This case study is discussed in the next section.

6.1. *Methodology*

The case study was conducted using a case study method (Yin, 2013), where the unit of evaluation was the LA. The LA was chosen because, as a public body, not only do they have an obligation to publish data in open format, they are also likely to face more public scrutiny than a private organisation in light of their status as acting for, and on behalf of, citizens (Shakespeare, 2013).

The research questions we asked were: *RQ1 – in a practical application with a public body organisation, Nissenbaum's Contextual Integrity (CI) framework will not work where one of the roles cannot be specifically defined?* and *RQ2 Using the CI framework in*

its current format will result in most datasets being deemed unsuitable for publication?

The case study was conducted in collaboration with two practitioners from the LA who were responsible for open data publication within the LA; a technical expert with extensive experience of data management and open data publishing, and a practitioner with a policy and management background. The method of data collection was a combination of contextual interviews (Beyer and Holtzblatt, 1998) and think aloud (Davison et al., 1997). The reason for this combination was that, due to distance restrictions and time constraints, the case study could not be conducted face to face. Therefore, contextual interview techniques by themselves, were not considered sufficient and the think aloud element was added to provide a more robust technique in the given circumstances.

Three LA datasets, which had already been deemed suitable for publication and published as open data by the LA were assessed as part of the case study. Prior to this case study, the LA did not have any formal privacy risk assessment processes in place, data was assessed by the practitioners in an ad-hoc, non-systematic manner based on experience. Thus, whilst these datasets had been assessed by the publishing practitioners, they had not been assessed through any formal privacy risk assessment, meaning that effectively, the case study involved re-evaluating the datasets for privacy risks using CLIFOD.

The actual case study was conducted by the lead researcher using a combination of telephone conferencing and Google Docs. The visits to the LA and collaboration with the LA was carried out by the lead researcher with the supporting researchers providing guidance and advice throughout on all aspects of the work. The researcher and practitioners went through CLIFOD over the phone and, simultaneously, the practitioners were given modify and share access to the CLIFOD spread sheet via Google Docs, enabling them to actively participate in completing the questions within CLIFOD during the study. Then, as they worked through the questions with the researcher, practitioners were asked to talk aloud, explaining their answers. These answers were then captured by the primary researcher and entered into the spread sheet. At the same time, where a particular answer or thought given by one of the participants did not make sense, the primary researcher used contextual inquiry techniques to elicit more information and talk to the participants about that aspect. Once a consensus has been reached on the understanding, the agreed answer was entered on the spread sheet.

6.2. *CLIFOD questionnaire*

The questionnaire consisted of 98 questions, presented in a spread sheet, devised to align with the meta-model. This questionnaire follows the meta-model using the three overarching phases *explanation; risk assessment and decision* in the questionnaire and the aspects beneath each phase to guide the questions asked within each phase.[1]

6.2.1. *CLIFOD Phase I – explanation*

For the explanation phase, the meta-model requires that background details are captured about 5 aspects: the *open dataset (i.e. the attributes*, the *actors* and their *roles*, the *transmission principles* and the *prevailing context* (see Fig. 1). In CLIFOD this was translated into a series of questions within each aspect. These questions were devised with reference to CI, looking at the decision heuristics identified in the meta-model as relevant for this phase (DH1-4). For example, for the attributes, a series of questions about each attribute were formulated to align with CI and the meta-model.[2]

6.2.2. *CLIFOD Phase II – risk assessment*

For phase II, CLIFOD takes the information collated in the explanation phase and uses this to to inform the risk assessment. The risk assessment then uses this information to inform the questions based first, on the 4 aspects identified on the meta-model. For example, for the regulatory compliance element identified as one of the aspects in the meta-model, questions were asked about any legal obligations or constraints that may influence the change in information flow (see 'Regulations' in Fig. 2). For the supporting questions, DH5-8 were identified in the meta-model as relevant and thus, these were used to inform the questions in the risk assessment phase of CLIFOD.

In addition, in CLIFOD, a scoring mechanism has been added to assist practitioners in gauging the likely severity and impact of the risks identified. These scores were captured using an established risk scoring method, the traffic light marking system, which denotes risks using traffic light colours of red, amber and green based on ratings given, i.e. high (red), medium (amber) or low (green) (Heiser, 2008). Because answers given as part of the risk assessment are likely to be subjective and require expert input to make a decision, the intention for this scoring was not to calculate or compute a score, rather to provide practitioners with a focal point that can be easily referenced in the final phase and thus, aid in making an informed decision.

Finally, to ensure the *positive impact assessment* identified as another aspect in the meta-model is captured, a section was added for this at the end of the risk assessment in CLIFOD that asks practitioners to identify what positive values the proposed change in information flows is likely to bring (see Fig. 2).

6.2.3. *CLIFOD Phase III – decision*

The final phase involves making and recording the decision which should be informed by the answers provided in phases I and II. Here, the meta-model requires that the decision is recorded by completing the following aspects: the *decision* and, as part of this record the: *findings*, which need to include the reasoning behind why that decision was made (*reason*) and what mitigation, if any, needs to be put in place before publication (*mitigating steps*). In addition a responsibility matrix (RACI, see Section 4.3) needs to be completed to record; who will be responsible for what (*actors* and the timeframe within which this will happen (*time*), see Fig. 3.

[1] A full list of the CLIFOD questions can be accessed at: https://github.com/JaneHB/CIOpenData.

[2] Ibid 1.

We recognise that this alone will not ensure that all privacy risks are identified, no privacy risk assessment can do that as any risk assessment is only as good as the practitioner who conducts it. However, by providing a step-by-step approach to conducting the privacy risk assessment that accounts for: what data is collected, the actors; how the data is transmitted and the context, CLIFOD helps practitioners identify and consider all the various aspects that can influence and affect the privacy risk. Thus, CLIFOD provides practitioners with a framework that can be used to apply the same format and questions to all assessments which will help ensure a consistent and repeatable privacy risk assessments is conducted for each dataset before publication occurs. This in turn will help provide assurance and enable public bodies to keep a record of all decisions made and the reasoning behind these decisions.

6.3. *Findings*

The study ran three datasets through CLIFOD that the LA had already published. In applying CLIFOD, we found that, it was necessary to explain the reasoning behind each set of questions in greater detail than that provided so as to elicit fuller, more considered responses. For example, the practitioners were unclear as to why each attribute and actor had to be considered. This resulted in an in-depth discussion around the merits of breaking the dataset down in this manner and why it was necessary to provide the level of detail that CLIFOD sought to capture. However, once a discussion around the first few attributes and how these might or might not have privacy implications in light of the informational norms, values and context had ensued, the practitioners seemed to grasp the concept and began to really think about the data in context. This then made the process much more fluid for considering the actors, transmission principles etc. and resulted in some interesting insights both for the practitioners and the researchers. This exercise took three hours and resulted in the following findings.

RQ1 asked about the inability to define the end user. Where a public body changes the data flow from internal sharing between departments to publishing in open format, a number of elements will change. First, the fundamental roles will be affected in that the receiver will be *anyone* who downloads the data (*end user*) and therefore, the flow and context will change from an internal stakeholders processing data for a particular purpose (e.g. to record benefit payment to/from a citizen) to the end user, who may use the same data for creating an application that displays number of people in arrears or any other purpose they see fit. This will change the risks associated with the data as it is impossible to accurately predict what the end user might or might not decide to do with the data. However, using CLIFOD can help reduce the risks of personal data being published by helping public bodies identify and mitigate against these risks in a consistent, repeatable manner.

Thus, we found that despite the lack of ability to define the end user and all three transmission principles, the framework appeared sufficiently effective in arriving at a consensual answer in determining whether or not a data set should be published in its current format. We therefore contend that using CLIFOD, provides a timely opportunity for practitioners to ensure all the aspects and contextual nuances of releasing the data are considered, thereby effectively acting as a very effective safeguard to making sure appropriate mitigations are put in place prior to publication.

RQ2 hypothesised that the CI framework would result in most datasets being deemed unsuitable for publication. The framework determined that two of the datasets considered in the case study were deemed unsuitable for publication in their current format. Some of the attributes within the datasets were considered to be personal or sensitive with potential – if linked to external data – to pose a threat to privacy. Despite this, the data had been published in compliance with a legal obligation placed on the LA. The final dataset considered was not only deemed unsuitable for publication, it contained both sensitive attributes and directly identifying data. However, because consent had been sought from the data subjects prior to publication, the LA considered that, as data protection had not been breached, the data could be published. These results would therefore suggest that RQ2 hypothesis is true. However, the sample used was small so that whilst all three datasets considered were unsuitable in their current format, a wider studies would need to be conducted to confirm or otherwise.

7. Discussion

Practitioners remain unclear on how best to preserve privacy in respect of open data publishing, so much so that, paraphrasing from the words of one Public Sector Senior Manager interviewed in a previous study: *"the easiest thing is to not make the data available. You're not going to make any mistakes if you don't make the data available"* (Barry and Bannister, 2014). However, in the current climate of openness and transparency, public bodies will increasingly find it difficult, if not impossible, to not publish any open data, if they are to meet public expectations, and indeed, their legal obligations under ROPSIR, FOI and similar legislation.

To some extent this can be addressed by technical intervention which can help mitigate against personal data being made available. For example, using anonymisation or redaction techniques prior to releasing the data (e.g. ICO, 2012; Lablans et al., 2015; Pfitzmann and Hansen, 2010; Samarati, 2001). To that end, it was found that some technical intervention had been conducted on two of the datasets published with some attributes having been removed from the original dataset prior to publication to ensure no personal information was published. Despite this, as a result of applying CLIFOD, questions were raised around some of the individual attributes within the data and whether, if linked to external data, these could pose a threat to privacy. Much discussion took place around this and the LA conceded that, had they applied the framework prior to publication, different decisions may have been arrived at with regard to which attributes to release. However, for both datasets, the LA were bound by a legal obligation to publish and therefore, any such decision was outside the remit of the practitioners taking part in the case study.

An alternative way that has been proposed to address this, would be to not release the full dataset, rather, allow some form of controlled access to the data (as proposed by Nissenbaum in Conley et al., 2012). For example, if practitioners

were to apply a differential privacy approach as an alternative to making the full data available to the database query outputs (i.e. adding "random noise"; Dwork, 2006), this would address the issue of not knowing who the recipient is as the full dataset is never released. Perhaps as an alternative to making data fully available as open data, LAs and other public bodies should consider some form of controlled access to the raw data.

It is possible that a higher authority may have considered the privacy implications of the datasets that carry a legal obligation of publication and deemed the risks acceptable in light of the wider public good, this was not considered as part of the study and we therefore acknowledge that this could pose a threat to the validity of the study.

7.1. *How informed can consent for open data be?*

The third open dataset contained personal information about planning applications submitted by citizens for development of property, and details of any agreed peripheral work agreed to be undertaken by the applicant as part of the development. This is of concern, not just because the dataset contained identifying information, but also because it could be directly linked to a separate dataset freely available online: planning applications. All planning applications are subject to public inspection under UK planning regulations and therefore, in submitting an application, applicants know the information they provide may be shared with interested third parties. While all of the applicants in this file had given consent and arguably, many applications will originate from developers and/or businesses, some will originate from citizens wishing to improve their own homes. For these applicants, while they will have consented as part of the application process, they may not fully appreciate the potential privacy risks.

It is one thing that interested parties such as neighbours are consulted as part of the planning process, quite another that anyone can obtain all of one's details including layout of one's property by downloading this from a website. If we were then to consider how such data could be linked to other available data, we could soon build a full picture of that individual, including home address and full name. The study participants did not capture the original data, nor made decisions around consent, rather, they were publishing existing data for which decisions around consent had already been made in accordance with policy. Therefore, this paper did not go into sufficient details to establish to what extent consent was fully informed or explained, but it was established that the application will not be considered without consent being given. This raises questions about how informed such consent might actually be. Thus, consideration needs to be given to questions of both processes and consent in these circumstances. Furthermore, knowing that technical privacy intervention can go some way towards safeguarding such data, a wider discussion is needed around when data should or should not be made public.

7.2. *Opportunities for users*

The fact that planning applications are open data could be used by applicants opportunistically. For example, an applicant on a budget considering building an extension to their home, may not be able to afford employing experts to assist in the application process. These applicants could use the open data to research previous similar applications and use this information to inform how best to phrase or prepare their application to ensure a successful outcome. Alternatively, a user could use insights gleaned from this open data to strengthen and enhance any objection they have to a particular application in their neighbourhood being given approval.

7.3. *Practitioners feedback on CLIFOD*

While largely effective, some of the questions within CLIFOD were redundant, and others required modification and further explanation. For example, following the initial discussions around the need for all attributes and actors to be considered mentioned earlier, it transpired that, as part of answering the initial questions on attributes, many of the questions that followed were answered as part of those initial responses. Consequently, these questions require revision in future adaptations of CLIFOD.

Once the case study had been conducted, practitioners were asked whether they felt the fact that the end user could not be defined had prevented them from considering how the data might be perceived in light of informational norms, or in the context of potentially conflicting values or morals. The practitioners felt that, rather than acting as an obstacle, this served as a reminder that extra care and time needs to be taken when considering privacy implications of publishing information in open format. Further, our study highlighted how failing to apply an initial holistic overview of the whole dataset had resulted in privacy sensitive data being published.

7.4. *Design implications*

The implications of these findings for design are that we, the technical community, need to look at ways we can incorporate not just syntactic or semantic privacy but also, holistic privacy into our designs. For open data, privacy needs to be considered in a more strategic manner before publication occurs, i.e. at a much earlier stage in the design process.

8. Limitations

This paper has looked at how practitioners can make better informed, and therefore more effective, decisions about privacy risk for open data before publication occurs. This involves looking at historic data often collected by disparate departments where the decision maker(s) have had no input into how data was collected, processed or stored. Therefore, the assessment has been limited to privacy risks associated with publication of that data in open format only. We acknowledge that, conducting a simultaneous or subsequent security risk assessment to establish what security measures should be put in place to protect the data going forward would enhance the effectiveness of the outcome. This will be addressed as part of future work.

We acknowledge the quality and outcome of conducting a privacy risk assessment will depend on the expertise and

quality of the assessment practitioner and/or team. However, this we contend is the case for any risk assessment, the more detail and background knowledge the team have in the particular subject, the more detailed and thorough the risks identified.

With regards to the conceptual meta-model presented in this work, the only aspect of the meta-model that was not fully tested and validated as part of the CLIFOD case study was the decision (Phase III, Fig. 3 in the meta-model). The reason this aspect was not really tested was that all the datasets that were assessed had already been published and thus, a decision had already been made. However, the finding that all the datasets assessed using CLIFOD contained potentially identifying information, showed that not conducting a privacy-specific risk assessment has resulted in personal data being made public. Thus, these findings illustrate the effectiveness of the meta-model and CLIFOD, as described in this paper, and how these can be successfully applied in practice. Further, they also highlight the importance of devising a privacy-specific risks assessment that practitioners can apply in practice (Henriksen-Bulmer and Faily, 2017).

9. Future work

Future work will look to further explore how the conceptual framework and CLIFOD can be used for assessing the suitability of publishing or sharing any type of data, not just open data. In addition, consideration will be given to how suitable the meta-model and CLIFOD will be to other areas of privacy decision making process such as FOI. As part of this, an exploration of whether CLIFOD might be more accurate if the decision making process is divided into specialisms. Thus, experts within each decision making area, such as policy, law, data management etc. would be asked to complete sections relating to their area of expertise, and, as a result, a more informed decision being arrived at.

However, our findings indicate that making practitioners understand and appreciate the concepts and contextual nuances involved in assessing privacy in context, requires explanation and steerage from experienced practitioners or academics to be effective. Therefore, more research into this area is necessary to establish whether and how we can better explain CI so that practitioners can complete contextual privacy assessments without needing an expert in-situ.

Moreover, to strengthen the application of both the meta-model and CLIFOD, work will be carried out to explore complimentary security risk assessment and privacy threat modelling frameworks, such as the IRIS security framework (Faily and Fléchais, 2010), and the LINDDUN privacy threat modelling framework (Wuyts et al., 2014). The work will look at whether these frameworks can be adapted to enhance any resulting tool support developed, with a view to assess how these principles might be incorporated. This will enable both privacy by design and security by design to be considered and thus, allow practitioners to not only identify and assess privacy risk but also safeguard the data in the future, building on theoretically grounded and validated frameworks.

10. Conclusion

This paper has looked at how practitioners can make informed privacy risk decisions using contextual integrity in practice. The meta-model created in this paper, illustrates that it is possible to break CI down into its component parts. In doing so, the meta-model shows how, by breaking CI down into logical phases and modelling how these interlink, a decision-flow can be established which can be followed in a methodical and systematic format when making decisions about privacy risks. The effectiveness of this approach was illustrated in Section 5, where we demonstrated how the meta-model can be used to support strategic privacy risk decision making through a worked example that applied the concepts to a hypothetical public body, a public library. Finally, we validated the robustness and effectiveness of the model by demonstrating how the conceptual model was translated into a practical questionnaire, CLIFOD, that practitioners can follow in a step-by-step manner to the assess privacy risks associated with publishing datasets as open data.

The findings from the CLIFOD case study highlighted that despite legislative constraints and guidelines, public bodies do struggle to meet their legal obligations in open data publishing. We found that a lack of sufficient understanding, processes and resources to address these obstacles only serve to increase these barriers. As a result, many fear adverse consequences such as litigation could result from making data available (Barry and Bannister, 2014). This paper has looked at one of these barriers, privacy, and shown how inadequacies in existing processes have resulted in identifying information being made public as part of existing open datasets already published.

Practitioners raised concerns about a lack of guidance and understanding in how to deal with privacy in practice. We found that despite some level of privacy assessment taking place prior to publication, such as removing some of the attributes prior to publication, some personal information is still being published which would indicate that existing processes lack enough rigour to sufficiently safeguard users privacy. Thus, these are valid concerns that, if not addressed, will likely lead to more sensitive data being published without the full consent of the users whose data is made available open source.

We have shown how the conceptual model and CLIFOD can assist in providing public bodies with a means of assessing the balance between the privacy of the data subject and the needs of the LA and thus, provides an important first step towards preserving users privacy. Further, both the meta-model and CLIFOD are generalisable to other privacy decision processes and, as such, could be equally applied to any other requests for information received by a public body such as an FOI request. The contribution to design is to highlight that a need exists for earlier, more holistic assessment of datasets as a whole, rather than just at attribute level is needed, and we believe this conceptual model and CLIFOD can start the discussion by showing one practical way of assessing privacy risks in light of the people, values, norms and the surrounding context.

We applied this to open data utilisation, an understudied area that the public bodies are keen to promote to help their

case in obtaining additional traction, utilisation and buy-in that can allow them to expand on their open data projects. In doing so, we illustrated the significance of the conceptual model, and how these concepts can be practically applied using CLIFOD as a practical tool that practitioners can use to assess privacy risk, thereby reducing the likelihood of a data breach through better informed decision making.

REFERENCES

Bamberger KA, Mulligan DK. Privacy on the ground: driving corporate behaviour in the United States and Europe. London: England: MIT Press: Mssachusetts Institute of Technology; 2015.

Barry E, Bannister F. Barriers to open data release: a view from the top.. Inf Polity: Int J Gov Democr Inf Age 2014;19(1/2):129–52.

Barth A, Anupam D, Mitchell JC, Nissenbaum HF. Privacy and contextual integrity: framework and applications, In: Proceedings of IEEE symposium on security and privacy, Vol. 2006; 2006. p. 184–98. doi:10.1109/SP.2006.32.

Berkeley Law. Implications of government release of large datasets; 2016.

Beyer H, Holtzblatt K. Contextual design: defining customer-centered systems. San Francisco, CA: Morgan Kaufmann Publishers; 1998.

Bieker F, Friedewald M, Hansen M, Obersteller H, Rost M. A process for data protection impact assessment under the european general data protection regulation. In: Lecture notes in computer science. Proceeedings of 4th annual privacy forum, APF 2016, Franfurt, Germany; 2016. p. 21–37.

BS ISO 27000:2017. In: Technical Report. British Standards Document BS ISO 27000:2017: information technology. Security techniques. Information security management systems. Overview and vocabulary. British Standard and the International Organization for Standardization (ISO); 2017.

BS ISO 31000:2009. In: Technical Report. British standards document BS ISO 31000:2009: risk management. Principles and guidelines. British Standard and the International Organization for Standardization (ISO); 2009.

BS ISO/IEC 29100:2011. In: Technical Report. BS ISO/IEC29100: information technology – security techniques – privacy framework. British Standard and the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC); 2011.

Conley A, Datta A, Helen N, Sharma D. Sustaining privacy and open justice in the transition to online court records: a multidisciplinary inquiry.. Md Law Rev 2012;71(3):772–847.

David W, Rachel F, Rowena R. A comparative analysis of privacy impact assessment in six countries. J Contemp Eur Res 2013;9(1):160–80.

Davison GC, Vogel RS, Coffman SG. Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. J Consult Clin Psychol 1997;65(6):950–8.

Deng M, Wuyts K, Scandariato R, Preneel B, Joosen W. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requir Eng 2011;16(1):3–32. doi:10.1007/s00766-010-0115-7.

Dwork, C. Differential privacy, Lecture Notes in Computer Science (4052); 2006. p. 1–12.

El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of reidentification attacks on health data. Plos One 2011;6(12):e28071.

Faily S, Fléchais I. Context-sensitive requirements and risk management with IRIS. In: Proceedings of the 17th IEEE international requirements engineering conference. IEEE Computer Society; 2009. p. 379–80.

Faily S, Fléchais I. Analysing and visualising security and usability in IRIS. In: Proceedings of the 5th international conference on availability, reliability and security. IEEE; 2010. p. 543–8.

Fowler M. UML distilled: a brief guide to the standard object modeling language. The Addison-Wesley object technology series. Boston, MA: Addison-Wesley; 2004.

Fung BCM, Ke W, Rui C, Yu PS. Privacy preserving data publishing: a survey of recent developments. ACM Comput Surv 2010;42(4) 14:1–14:53.

Galanc T, Kolwzan W, Pieronek J, Skowronek-Gradziel A. Logic and risk as qualitative and quantitative dimensions of decision-making process. Oper Res Decis 2016;26(3):21.

Grodzinsky FS, Tavani HT. Privacy in "the cloud": applying Nissenbaum's theory of contextual integrity. SIGCAS Comput Soc 2011;41(1):38–47. doi:10.1145/2095266.2095270.

Heiser J. A simple method for expressing information criticality and classification [online]. Gartner; 2008.

Henriksen-Bulmer J, Faily S. In: Proceedings of the 31st British HCI group annual conference on people and computers: digital make believe. Applying contextual integrity to open data publishing. British Computer Society; 2017.

Henriksen-Bulmer J, Jeary S. Re-identification attacks: a systematic literature review. Int J Inf Manag 2016;36:1184–92. doi:10.1016/j.ijinfomgt.2016.08.002.

ICO. In: Code of Practice 1061. Anonymisation: managing data protection risk code of practice. Information Commissioners Office (ICO); 2012. https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

Krupa Y, Vercouter L. Handling privacy as contextual integrity in decentralized virtual communities: the privacias framework. Web Intell Agent Syst 2012;10(1):105–16.

Kulk S, van Loenen. Brave new open data world? Int J Spat Data Infrastruct Res 2012;7:196–206.

Lablans M, Borg A, Ückert F. A restful interface to pseudonymization services in modern web applications. BMC Med Inform Decis Mak 2015;15(1):1–10.

Mulligan DK, Koopman C, Doty N. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. Philos Trans Ser A Math Phys Eng Sci 2016;374(2083):1–17.

Nissenbaum H. Privacy as contextual integrity. Wash Law Rev 2004;79(1):119–58.

Nissenbaum HF. Privacy in Context: technology, policy, and the integrity of social life. Stanford, California: Stanford Law Books; 2010.

NIST. In: Technical Report 800-122. Guide to protecting the confidentiality of personally identifiable information (PII). National Institute of Standards and Technology (NIST); U.S. Department of Commerce; 2010.

NIST. In: Technical Report SP 800-30. Guide for conducting risk assessments. US: Gaithersburg: MD: National Institute of Standards and Technology (NIST); U.S. Department of Commerce; 2012.

Obama, B. Transparency and open government: memorandum for the heads of executive departments and agencies; 2009.

Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev 2010;57(6):1701–77.

Open Data Institute. What is open data?; 2016.

Open Government Partnership. Participating countries; 2016.

Pfitzmann A, Hansen M. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. Technical Report v0.34; 2010. p. 1–86. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf.

Privacy Rights Clearinghouse. Data breaches; 2018.

Project Management Institute. A guide to the project management body of knowledge: PMBOK guide. Newton Square, PA: Project Management Institute; 2004.

Rumbaugh J, Jacobson I, Booch G. Unified modeling language reference manual. (2nd ed). Pearson Higher Education; 2004.
Samarati P. Protecting respondents' identities in microdata release. IEEE Trans Knowl Data Eng 2001;13(6):1010.
Sar RK, Al-Saggaf Y. Contextual integrity's decision heuristic and the tracking by social network sites. Ethics Inf Technol 2013;16(1):15–26.
Sasse MA, Brostoff S, Weirich D. Transforming the 'weakest link' a human/computer interaction approach to usable and effective security. BT Technol J 2001;19(3):122.
Shakespeare S. In: Technical Report 13-744. Shakespeare review: an independent review of public sector information. Department for Business, Innovation and Skills; 2013.
Simon HA. A behavioral model of rational choice. Q J Econ 1955;69(1):99–118.
Solove DJ. A taxonomy of privacy. Univ Pa Law Rev 2006;154(3):477–564.
Thomson D, Bzdel L, Golden-Biddle K, Reay T, Estabrooks CA. Central questions of anonymization: a case study of secondary use of qualitative data. Forum: Qual Soc Res 2005;6(1):1–16.
UK Parliament. Data protection act; 2018. [online] (May 2018). https://services.parliament.uk/bills/2017-19/dataprotection.html.
Wuyts K, Scandariato R, Joosen W. Empirical evaluation of a privacy-focused threat modeling methodology. J Syst Softw 2014;96:122–38.
Yin RK. Case study research: design and methods. Los Angeles, California: SAGE; 2013.

Jane Henriksen-Bulmer currently lecturing part time and studying towards her PhD in privacy, looking at how organisations can make informed decisions about privacy risks in the Department of Computing & Informatics at Bournemouth University. Before joining the PhD programme at BU, Jane completed a MSc in Information Technology from Bournemouth University in 2015, gaining a distinction. She also holds a Masters in Business Administration (MBA), gained from Curtin University in Perth, Western Australia in 2013 and a first class honours degree in Law from the Open University (gained in December 2010).

Shamal Faily currently working at Bournemouth University as a Senior Lecturer in Systems Security Engineering within the Department of Computing and Informatics. Before joining BU as a lecturer in 2013, Shamal was previously a post-doctoral researcher at the Department of Computer Science at the University of Oxford, and a teaching fellow at the Information Security Group at University College London. He completed his DPhil in Computer Science at the University of Oxford in 2011. Prior to his doctoral research, he was a software engineer within Logica's Space business.

Sheridan Jeary received her PhD from Bournemouth University in 2010 in Requirements Engineering and has continued to research in the area of requirements, particularly models, since that time. She is particularly interested in the current trend of releasing data to meet software requirements without consideration of personal consequence.