UNIVERSITY OF SOUTHAMPTON

# Evaluating the Impact of Open Crime Data in the United Kingdom

MAIRE BYRNE-EVANS

*for the degree of Doctor of Philosophy*

June 9, 2019

# Declaration of Authorship

I, Maire BYRNE-EVANS, declare that this thesis titled, "Evaluating the Impact of Open Crime Data in the United Kingdom" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself; parts of this work have been published as:
  - Crime Applications and Social Machines: Crowdsourcing Sensitive Data.

    Byrne Evans et al., 2013.
  - Keeping Your Little Back Shop.

    Byrne-Evans and Task, 2013.
  - I.B.M. Thought Leadership White Paper: Crime in the Digital Age: Digital Policing.

    Anning et al., 2016.
  - Place Geography: Finding a Space for Place.

    Hart, Frew, and Byrne, 2018.

  Signed: ________________________________________

Date: ________________________________________

UNIVERSITY OF SOUTHAMPTON

# *Abstract*

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Doctor of Philosophy

**Evaluating the Impact of Open Crime Data in the United Kingdom**

by Maire BYRNE-EVANS

This thesis examines the impact of Open Crime Data in the United Kingdom (U.K.). Wide claims are made over the benefits of such data. Police.uk, managed by U.K. Home Office, publishes Open Crime Data, mandated by the U.K. Government's Transparency Agenda. Police.uk provides information about recorded crime on a large scale, through mapped crime locations. This enables the creation of "crime apps" providing knowledge about crime.

However issues arise from the use of web-mediated Open Data to leverage transparency. The thesis analyses the complex environments of Open Crime Data and the data themselves. Open Data inhabit a landscape that features secret data or knowledge - by illuminating parts of this ecosystem, the data, its provenance, and effects of the Web, claims for Open Data can be unpicked from a theoretical perspective aligning data, policy and knowledge.

We first i) examine literature and concepts relating to knowledge of crime and Open Data; ii) review data production using concepts from statistics, surveillance and Big Data; iii) analyse how policing and maps have been combined through these first two conceptual areas. Second we use: i) Grounded Theory to examine the context that Police.uk inhabits with respect to Open Data; ii) Big Data concepts and Frame Analysis to examine the impact of Police.uk on Online Social Networks. We iii) then interview cybercrime experts and contrast their views with the results of ii).

The contribution of this thesis is: Empirical evidence of the effects of the publication of Open Crime Data on people; Understanding how on-line social networks play in this and affect analysis of crime data's impact on organisations and people, including the police themselves; New methodologies to understand this; New ways of conceptualising crime; Understanding of the Web's contributions to policy.

# *Acknowledgements*

To my supervisory team: Dr. Thanassis Tiropanis, Dr. Craig Webber and Dr. Kieron O'Hara, who offer support and invaluable advice, and who have been patient. Also to Dr. David Millard and Dr. Jeff Vass who were inspirational during my vivas. To my examiners, Professor Mark Briers, and Dr. Chris Hamerton. To Craig Jones who encouraged me. To Web Science and Web Scientists in all their manifestations, or perhaps if the Singularity arises, to Philosophical Engineering?

This work was funded by the Digital Economy programme, led by the Engineering and Physical Sciences Research Council on behalf of UK Research and Innovation. It was also supported by people at the University who really helped me when circumstances made things hard. My current employer has been encouraging and a pleasure to work with and has enabled me to take this work much further than I ever imagined.

To my children, my siblings, to the victims of crime, especially the hidden, unrepresented ones:

"...all that carries weight and always weighs the same lay in the hands of others; they were small and could not hope for help and no help came..." (Auden, 1955).

Let us change that.

To the geeks, spooks, thugs and bots in the worlds of intelligence, security and policing, some of whom keep pushing at the problems that really matter, who show relentless curiosity and who really want to Understand.

To all who supported, cajoled, encouraged, listened, ignored, opposed and tested me.

In particular a big thank you to the ones who travelled alongside me on a parallel path, often a long way off, but who were there when I needed help and who appreciate my abstracted sea views.

To those who fight to keep axioms in line:

"Out of the air a voice without a face
Proved by statistics that some cause was just
In tones as dry and level as the place:
No one was cheered and nothing was discussed;
Column by column in a cloud of dust
They marched away enduring a belief
Whose logic brought them, somewhere else, to grief."

(Ibid.).

# Contents

# List of Figures

*For everyone, and Andrew.*

…

# Chapter 1

# Introduction

## 1.1 Open Crime Data and Transparency

Openness and transparency, in conjunction with the technology of the World Wide Web, have been expected to transform society: enabling trust in governments, and strengthening civic participation in decision-making. Openness, transparency and accountability are held up as the means by which democracies can be strengthened: the public can examine government practices, hold governments accountable, and thus, through examination of their actions, change what governments do. Technologies are positioned as the enabling factor that allows governments to govern more rationally, and to better understand and act on the desires of those whom they govern. Open Data is part of the mechanism which feeds accountability, and ultimately, democracy when it is achieved through transparency. Open Data is “data that can be freely used, reused and redistributed by anyone.” (Open Knowledge Foundation, 2012).

Open Government Data (O.G.D.) is a now well-recognised manifestation of Open Data, and is released, aggregated, enhanced and reused across many Government policy sectors, at an unprecedented and previously unimagined rate, owing to the advent of the World Wide Web. Transparency and Open Data advocates make claims for benefits to be realised from web-mediated Open Government Data — the sharing of knowledge, and wealth creation from knowledge, improvements in democracy, human rights and political communication, efficient spending and policy-creation — and offer much evidence to support their claims. (Open Knowledge Foundation, 2012).

The Web Foundation stated in its 2015 report, “Governments worldwide have acknowledged the potential of O.G.D. to reduce corruption, increase transparency, and improve government services”, while Professor Sir Tim Berners-Lee, creator of the Web Foundation commented on the report, “. . . governments continue to shy away from publishing the very data that can be used to enhance accountability and trust,” and highlighted the ability of Open Data “to put power in the hands of citizens”. (The Web Foundation, 2015).

Berners-Lee’s comments suggest that paucity of O.G.D. goes hand-in-hand with a lack of power in the hands of citizens. It seems that O.G.D. can bring about a new form of web-mediated democracy, and that correspondingly, "more O.G.D." will create "more democracy", as though democracy is a commodity leveraged through liberalised data. However there are issues arising from the use of web-mediated Open Data in pursuit of transparency and accountability. Some of these are ideologic, some technical and some politico-social.

## 1.2 O.G.D. and its Context

This thesis examines the assumptions behind Open Data and discovers a complex landscape. The charting of this landscape creates a perspective around the idea that O.G.D. offsets information asymmetry existing within capitalistic democracies; asymmetry that appears to commentators to favour governments and institutions such as the police, wider law enforcement and corporations and disfavouring "the public".

While there is much to be researched and said on democracy, and the consequent focus on power that such research brings up, our analysis focuses on the interplay between the data itself and those who touch it, wittingly or unwittingly. We examine how well such data aids transparency and accountability, and how and whether it offsets information asymmetry, by looking at Open Crime Data as an example of O.G.D. that is meant to empower the citizens who use it. We find that some incorrect assumptions are deployed about the complex, contingent environments in which Open Crime Data are used.

The issues we find are not necessarily new. Prior to the large-scale provision of data via the World Wide Web, critics were vocal about the deficiencies of New Public Management, a term used by Hood, (Hood, 1995), denoting the idea of improving the public sector via the "Three Ms: markets, managers and measurement." Ferlie suggested that "back-office efficiencies" in fact cause more problems than they solve. (Ferlie et al., 2013). These back-office practices are now sold as "making government more efficient", and the provision of data via the W.W.W. is one of its key manifestations. There have been calls for alternative approaches in the more recent past, ranging from theory rooted in management practice, (for example Foss, Husted, and Michailova, 2010), such as The Third Way or Public Value Theory, to ideas rooted in physics and complexity, such as calls for understanding how complex adaptive systems work. The literature that we examine in fact reveals that core issues go further back than this, and have been under examination in the UK. since at least the Fifteenth Century.

There are therefore, historic and robust issues arising from such data produced in such contexts; these are deepened not only by understanding the history of government data produced in the name of efficiency, but also by an examination of data that are not open, or that move from various ontological states, including administrative and natural states and data moving from secret to open. We find that secrecy is sometimes necessary in order to preserve markets, open trading and citizen liberty, negating a somewhat homuncular, but often non-explicit premise of the Open Movement, that within a system that preserves liberty and openness, the components of that system must necessarily also be free or open.

This thesis uses a mixed method approach within the discipline of Web Science. In the foundational paper, "A Framework for Web Science", a "comprehensive set of research questions"... are set out, "together with a sub-disciplinary breakdown, emphasising the multi-faceted nature of the Web, and the multi-disciplinary nature of its study and development. These questions and approaches together set out an agenda for Web Science, the science of decentralised information systems. Web Science is required both as a way to understand the Web, and as a way to focus its development on key communicational and representational requirements." (Berners-Lee et al., 2006). In the particular policy area of crime and policing, we analyse the impact of the production and consumption of web-mediated Open Crime Data on some social and organisational behaviours and outcomes, within crime, policing, policy and society. We also look at the production of mapped open crime data and ask whether mapping such data is counter-productive. We examine old critiques and theoretic methods relating to transparency, democracy and accountability and ask whether they still apply when they are web-mediated, and mapped. We suggest that the issues that arise can best be addressed through the perspective of the multidisciplinary Web Science, as described above.

## 1.3 Thesis Outline

The thesis starts with an introduction to some of the concepts that we will be examining. We provide an analysis of the way in which too much of a focus at the data level detracts from policy-level problems, while policy often ignores the peculiar problems of data, especially provenance. There are currently few systemic analyses that understand the interplay between data, policy and the World Wide Web. We set out the areas of literature that help to present a new understanding of how to approach this problem and then provide a methodological framework that uses multi-disciplinary approaches to provide analyses that help to illuminate the phenomena under discussion.

## 1.4 The Scope

We describe www.Police.uk, a web-mediated interface, maintained by the U.K. Home Office. Police.uk provides access to information about recorded U.K. crime on a large scale, via the publication and presentation of open crime data, and through maps showing crime locations and judicial outcomes. We set out ways in which the impact of this site is currently analysed at the level of the data, and suggest potential issues with these analyses. We use knowledge from experts who work with a "new" sort of crime – cybercrime - in order to illuminate these issues. We then systematically address these concerns using novel mixed-method interdisciplinary Web Science approaches.

## 1.5 Research Questions

The research question is, "Evaluating the Impact of Open Crime Data in the U.K.." It combines broad data methodologies and criminological concepts to explain how and why we can unpick this question, primarily through the combined use of network theory concepts, Grounded Theory, Frame Analysis and Broad/Big Data techniques. This question has sub-questions, arrived at through preliminary analyses and then clarified throughout the research process:

- What are the tensions between the actors and technologies involved in the production and consumption of Open Crime Data in the U.K.?
- How do the Web of Data and online social networks mediating this crime data affect transparency and accountability in the U.K.?
- Can we combine Big Data and network science methods with criminological and philosophical theory to understand the effects of the supply of crime data from the web?

## 1.6 Research Contributions

Research contributions are:

- Empirical evidence of the effects of the publication of Open Crime Data on people;
- Understanding of how On-Line Social Networks (O.L.S.N.s) and the Web of Data (W.O.D.) play in this;
- The analysis of crime data's impact on organisations and people, with a focus on the police themselves;
- Use of new methodologies to understand the interplay of individuals, organisations, the Web of Data and online social networks;
- Understanding of how best to turn data about crime into information or intelligence;
- Understanding of the Web's contributions to policy as a result of its mediation of Open Data.

## 1.7 Publications and Related Work

The following pieces of work have been produced or are in production and support the work herein:

- Crime Applications and Social Machines: Crowdsourcing Sensitive Data.
  (Byrne-Evans et al., 2013).
- Keeping Your Little Back Shop.
  (Byrne Evans and Task, 2013).
- I.B.M. Thought Leadership White Paper: Crime in the Digital Age: Digital Policing.
  (Anning et al., 2016).
- Briefing paper for N.C.A.: Intelligence and Evidential Data Contained within Mobile Applications.
- Briefing paper for S.E.R.O.C.U.: Cybercrime: Decreasing business vulnerability, with a focus on D.D.o.S..
- Briefing for C.E.S.G. on the Impact of Cyber Essentials.
- Briefing for Heads of Operational Risk Insurance Companies (O.R.I.C.) and Whiterock, a Technical Counter-Surveillance Measures company: On Cybersecurity.
- Place Geography: Finding a space for Place.
  (Hart, Frew, and Byrne, 2018).

# Chapter 2

# Background

## 2.1 Transparency, Open Crime Data and Police.uk

As stated at the end of the last chapter, the literature review which follows has three parts, which analyse i) the concepts around knowledge of crime, ii) the numbers of Open Crime Data and iii) lived experience of crime data. These illuminate our research approach which seeks to understand, i) first the ecosystem or market for crime data that Police.uk occurs within, then ii) the statistics that it produces and finally iii) the experiences of those who produce or who are affected by the numbers. The research questions addressed are:

- RQ1. What are the tensions between the actors and technologies involved in the production and consumption of open crime data?
- RQ2. How do the Web of Data and online social networks mediating this crime data affect transparency and accountability?
- RQ3. Can we combine big data and network science methods with criminological and philosophical theory to understand the effects of the supply of crime data from the web?

### 2.1.1 Structure of this Thesis

In order to understand the research areas and background covered by the literature review, we first look at some earlier work done by the U.K. Home Office, in an attempt to analyse the impact that its publication of crime data was having. Understanding the difficulties with such an approach then leads us to delineate some of the major questions raised in three areas - relating to crime data and knowledge, number and society.

In the literature we first examine Open Data with regard to the knowledge it creates, and the belief that such knowledge brings consequent freedoms. We look at knowledge and Open Data in relation to the Open movement, and Open Government as a form of democracy that has transformative and liberating powers. We find there are some tensions within the system producing Open Crime Data (O.C.D.) - these are clarified by mapping the O.C.D. ecosystem.

The next part examines the subject of Data, or Number, in the form of state statistics, surveillance and big data. This then paves the way for our examination of the numbers behind crime data and the impact of web-mediated crime data on social-networks at point of departure.

Finally we link the concepts of knowledge and number in terms of accuracy and representation, through the lens of society and mapping – examining the lived experiences and narratives of stakeholders who engage with open crime data, whether it be those who produce it - the police and other Law Enforcement officials- or those who consume it – the public, the police (again), government, industry, the public and private sectors, the media. We look at maps as a form of capture of lived experi-ence and ask whether the maps of Open Crime Data in the U.K. can hope to capture the stakeholder narratives. This background is then taken through to the research agenda by comparing the impact of the crime data numbers with what people producing it talk about in fuller detail, away from the Web but seeing the mapped crime data as an entry point for their understanding.

We consider the use of data as a lens for analysing and changing human and organisational behaviours, or "how situations come to be seen as caused by human actions and amenable to human intervention". (Stone, 1989, p.281). We examine the ways in which data, when it is in numeric form, can be illuminated by narratives or stories, and how mapping such data can create distortions when the concepts of knowledge, number and society are used routinely.

This literature review gives background and context to the methodology and analyses that follow, so that we can see how our work relates to what has gone before and extends it. We look for key ideas and theories relating to Open Crime Data, and also for clear tensions arising out of conflicting ideas, processes or actors. We analyse the understanding and deployment of formalised knowledge, relating to number, the state and society.

The literature is selected partially via historiographic accounts of the background to the current research questions, and analysis of tensions that emerge as we look at these salient historical markers; particularly with regard to approaches from different disciplines, including philosophy of science, statistics and sociology.

Although most of what follows is based around research stemming from the United Kingdom (U.K.), I will be making reference to some historical developments taking place in the United States and Europe, where the provision of context makes this advisable.

### 2.1.2 Open Data

In the United Kingdom, the Home Office innovatively uses open crime data, mediated by the World Wide Web, as part of a programme of transparency initiated by the U.K. government in 2010. Transparency and accountability – often together - can be used by corporations and governments as a means of demonstrating that they are open to scrutiny, minimising corruption and inefficiency. Where the U.K. government has implemented a programme of openness and transparency, expectations have been of transforming government, strengthening the trust of the people in government and encouraging a more democratic process by enabling greater civic participation in decision-making. This is through delivering a means by which the public can hold the government accountable for their actions, and by which the public can suggest improvements in what government does. (The Cabinet Office, 2013).

When, in 2010, a coalition government was formed, the U.K. Conservative and Liberal Democrat government identified the ways in which technological innovation has allowed information, or accounts of what is done by government, to move from the politicians to the people; decentralising power and enhancing transparency. One radical technological innovation is the use of the World Wide Web in delivering open government data. More recently, the government has suggested that,

"... publishing vast amounts of public sector spend data helps track civil service salaries, expenses and contracts, improving government accountability... data about public services' performance (e.g. school results, court sentences or hospital waiting times) is a good way of measuring the effectiveness of our policies... by releasing public data, the government allows people to see how it is doing, while also looking at better ways to carry out public services. . . open data can boost economic growth - businesses can take the data and produce goods and services from it.” (Gov.UK, 2015, p.1).

The “Open” in Open Data has been redefined as meaning that, “anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).” (Open Knowledge Foundation, 2013).

Open government data (O.G.D.) is a type of open data and part of the mechanism which feeds accountability and democracy – achieved through governments being transparent about their own actions. While governments are custodians of the information they collect or generate, they are elected by the people and use our taxes to do their jobs, therefore there is an argument that we have a right to see this data, throwing open “the doors of all public establishments . . . to the body of the curious at large.” (Bentham, 1787, Letter VI.: Advantages of the Plan).

Piotrowski and Ryzin have described how the American Society for Newspaper Editors commissioned one of the first systematic examinations of U.S. freedom of information practices in 1953, saying that, “The release of information promotes democratic accountability." (Piotrowski and Van Ryzin, 2007, p.308). The People’s Right to Know called for Open Government: "Public business is the public’s business. The people have the right to know. Freedom of information is their just heritage. Without that the citizens of a democracy have but changed their kings. The people are citizens, taxpayers, inhabitants, electors, newsmen, authors, research workers, teachers, students, all persons, each of us...The public has a legal right to be able to ’examine and investigate’ government activity." Piotrowski and Ryzin also cite economist Joseph Stiglitz’ argument for the release of information from public organisations. "He believes that governmental information belongs to the public: The question is, given that the public has paid for the gathering of government information, who owns the information? Is it the private province of the government official, or does it belong to the public at large? I would argue that information gathered by public officials at public expense is owned by the public—just as the chairs and buildings and other physical assets used by government belong to the public." This opens up the question of ownership of data and accountability within democracies and other forms of government.

Another reason for opening up government data is possibly reverse-engineered - rooted in apparent utility. Data has become obviously available, through the use of computer systems and automated storage mechanisms and the fact of its existence in various silos more apparent to various actors. This availability is often as a by-product of people recording the processes and outcomes of their jobs - a result of administration, and with the data being performance data – i.e. data that is recorded and used to measure how efficiently people do their jobs. It is important to think about what purpose such administration serves in the first place. As these processes have become increasingly automated and less paper-based, first through the advent of the database, then also through the use of the Web and cheaper computational processes, it has been perceived that the same technology that administered, could also help to “release” data to the public in ways which should be cheap and efficient. Databases hold this performance data and the advent of relational databases and spreadsheets with similar functionality, meant that data has become more accessible, for what has appeared to be a low cost or no cost. This is part of the argument of "the Long Tail", as Chris Anderson referred to it, (Anderson, 2004), the principle on which online marketplaces such as Amazon and Ebay were founded - the World

Wide Web can make very small, singular, previously specialist products reach all of those who desire them with an alleged minimum of expenditure and effort.

Releasing such data, with what appears to be minimal effort, means that creative communities of web developers can find new ways of using data that can trans-form government – such as in Code for America, or the Sunlight Foundation's use. (Lyon, 2014). Open Government Data is currently aggregated, enhanced, released and reused across many Government policy sectors, at an unprecedented and previously unimagined rate, as a result of the Web, and advances in technology. While the concept of transparency as a part of democracy can be traced back centuries, the idea of using the Web to enhance transparency is relatively new.

Prior to the advent of the Web, researchers and governments demonstrated that "targeted transparency" was part of the means by which companies could be made accountable. (Fung, Graham, and Weil, 2008, p.6). Using the web to apply this technique to government extended this idea from law, regulation and corporate governance to politics. It has been suggested that lack of planning around this provision and reuse of government data means that issues arise, such as those relating to privacy, where release is mandated by policies that were created before we understood the technological, social and moral complexities arising from the tidal flow of data that the Web has permitted and encouraged. (Byrne-Evans and Task, 2013). "Similarly, as transparency is an innovation that came, if not out of the blue, at least very unheralded, the potential threat to privacy has not been considered and the-orised to any great extent." (Kieron O'Hara, 2010, p.3). We will show how some of these issues, including those coalescing around popular concepts of privacy and anonymity, affect the data that we are examining and create tensions between the actors involved in the production and consumption of open data in the U.K. It has also been suggested that Open Data has not really worked so far, as a result of misunderstandings about what constitutes openness, what constitutes accountability and what makes government transparent. There is much to unpick. There are misunderstandings about the policy-making process, marketplaces, bad design, illogical implementation and use of technology, and possible adverse consequences for the Web itself. The Nominet Trust (Nominet Trust, 2012, p.77), discusses the problems of privacy, security, safeguarding and trust, while (Yu and Robinson, 2012, p.178), point out that "a regime can call itself open if it builds the right kind of web site even if it does not become more accountable", because of the fact that O.G.D. can refer either to data that makes government more accountable, or data that belong to the Public Sector and are easy to reuse but have little to do with accountability. Schellong and Stepanets, (Alexander Schellong and Ekaterina Stepanets, 2011), found that there was a lot of room for improvement - out of datasets examined from nine Eu-ropean counties none had national open data portals. (See also (Noveck, 2012)). We intend to use the example of Police.uk to examine whether these issues apply, when applied to crime, and if so, what can be done to alleviate them.

### 2.1.3 Transparency in the Criminal Justice System

Within the Criminal Justice System (C.J.S.), the commitment to transparency arises partly from its apparently monopolistic nature – "victims and the public cannot choose to be protected by anyone else." In order to "ensure that services improve, innovate and deliver the best outcomes it is essential that their workings are exposed to public view. Only when armed with the right information and given the chance to observe how the system works in practice can the public hold the criminal justice system to account." (Ministry of Justice, 2012, p. 52).

Policing is one part of the C.J.S where it is suggested that increased transparency strengthens the ability of the public to hold the police to account and thus helps the government to withdraw from 'interference' with local policing. This then enables communities to engage more fully in deciding for themselves which are important areas to tackle. Criminological theory itself suggests that local populations should be involved in decision-making about policing and crime, and that their engagement is crucial in order to deepen trust between police and communities in the U.K.. (Jones, MacLean, and Young, 1986).

Theresa May, Home Secretary in 2011, said, "Instead of leaving the politicians in charge, we are giving power to the people. We will restore the link between the public and the police by making the police accountable to the people they serve, through the election of Police and Crime Commissioners, the publication of the most transparent local crime data in the world, and mandatory beat meetings." (The Home Office, 2011, p.1).

Police and Crime Commissioners (P.C.C.s) first took office on 22 November 2012. Crime figures can be used as tools for P.C.C.s to dismiss Chief Constables, and by the public to scrutinise P.C.C.s in the name of accountability. Police and Crime Commissioners have the ability, as yet not fully explored in courts, to remove Chief Constables from post. Chief Constables may be "called upon" to retire or resign by the P.C.C. under s.38 (3) of the 2011 Act." (Legislation.gov.uk, 2011). There is debate over how far these powers should extend: "the wide discretion of commissioners to dismiss Chief Constables is a significant issue, and shows that statutory provisions intended to give police and crime panels a role in respect of dismissals, albeit a consultative one, can be evaded". (House of Commons and Home Affairs Committee, 2013).

On the other hand, the public hold the P.C.C.s to account; the Home Secretary herself having said that information about whether P.C.C.s have second jobs is "for the electorate to make a decision on". Police and Crime Panels scrutinise the P.C.C.s; but there is dissatisfaction with the process. Watford Mayor, Dorothy Thornhill likened Police and Crime Panels to crocodiles "with rubber teeth." (Nic Brunetti, 2013). It appears that tracing responsibility and accountability through these new structures that feed off Open Data and are products of the government's transparency programme, can be initially problematic, however, the political discourse about open crime data is that it is how the public can hold the P.C.C.s to account, and that it improves "access to criminal justice information". (Ministry of Justice, 2012).

### 2.1.4 Knowledge of Crime Mediated by the World Wide Web

Police.uk is a website which mediates the public's engagement with open crime data, and therefore, public knowledge of crime. It does this by means of representing crime data – recorded police statistics - visually via geospatially mapped data as well as providing access to the "raw" data underlying the mapped presentation. We distinguish between the data that is provided by Police.uk and its companion Data.gov.uk and the representation of the data provided by the mapped interface. The web mediates both.

"Administrative data are used extensively in the compilation of many National Statistics products – these include health statistics, such as waiting times; crime statistics, such as police recorded crime data; education data, such as schools level examination results; and economic data, such as tax estimates derived from individual tax records." (UK Statistics Authority, 2014, p.6).

Technology and innovation are bringing open crime data and knowledge of crime to individuals and communities who have access to the Web through ubiquitous de-vices. People can become more aware of levels of crime through their own discovery and exploration of the data, decide whether local services are effective, and engage in social activism. The ex-Minister for Policing and Criminal Justice, Nick Herbert, said at the launch of Police.uk:

"I have been an advocate of street-level crime mapping since seeing it work in Los Angeles. . . I believe it goes further and is more comprehensive than any other scheme. Police.uk will make England and Wales world leaders in this field, with every citizen able to access details about crimes on their streets. . . we are giving people the information and power to hold their local forces to account and ensure that crime in their neighbourhood is driven down." (Page, 2011).

Home Office crime data comes from 43 different police forces, all of whom operate under differing local targets and with different concerns about crime. They also have different ways of recording and producing crime data. The data is anonymised and aggregated both over location and time. It sometimes appears up to seven weeks after the commission of a crime. Neighbourhood Watch (one user group) has asked for data with exact locations and times, assuming that this enables them to predict further crime in the area. This suggests that at least one user group be-lieves that if the Police.uk data were accurate, it could be predictive, and help to stop crime, not just record it.

### 2.1.5 How Do We Measure Impact?

While much of the focus of this thesis is in asking "impact upon whom?" and defining target populations for the research, in understanding how producers and consumers of O.C.D. engage with data, we must first also ask, how do we scientifically measure impact? Measuring impact usually implies a causal account. We do x, and y occurs. Causal accounts are easy to illustrate with behavioural effects where the phenomena under investigation are clearly isolatable. For this thesis, we examine what happens when we publish crime data on the web - what effect does this have? Some behaviours can be measurable where they are manifested in physical effects (sales, clicks, behaviours that indicate increased knowledge of crime "hot spots"), others less so, where they indicate cognitive/psychological effects, such as increased or decreased fear of crime.

In past writing on accountability, evidence for impact has been evaluated via the splitting of accountability efforts into strategic and tactical. Jonathan Fox discusses how under tactical accountability approaches, we can see bounded interventions, with society-side effects and the assumption that information provision alone inspires collective action with enough force to change Public Sector performance. (Fox, 2007). However, it is hard to separate cause and effect where there is already some transparency, some accountability and some democracy in a society. Transparency's effects are perhaps most socially observable where democracy is little in evidence. The more democratic a society becomes, the harder it is to measure global or large-scale "effects" of transparency, mired as they are in the workings of the legal and organisational mechanisms, economics and the bureaucracies embedded in democracies.

A closer look at impact can allow us to define it as "effect", examining causal relationships between publication of the data and the behaviours it causes in the public, politicians, lawmakers, media, institutions, app-makers and society, including those whose crimes are delineated in the data. As we have seen in the speeches given at the initial launch of Police.uk, it was hoped that publication of this data would help reduce crime. This could be seen as a primary explicit target of Police.uk, and therefore it might seem logical to measure impact by looking at whether crime has reduced. In fact, such reasoning is beset with problems, which we will examine.

There is also a problem with looking only at "behaviours" – while this is a very wide term with respect to its manifestation via various means in organisations and people, in fact it generally limits effects or impact to externally observable changes – whereas, for example, there is much research on how information about crime can cause fear of crime - something that is more subjective and harder to measure, but that is a commonly used construct in policy-making and criminology.

The systems that produce what most people consider to be crime data, are actually producing data that is used to measure how well officials are doing their jobs. Crime data is policing data, and as it is used to measure how well police do their jobs, it can be subject to what O.N.S. and the Statistics Authority have called "gaming". Gaming is a contested area and the use of the term can suggest a lack of systems understanding or of systems thinking. (See (Copperfield, 2006), (Baxter and Hirschhauser, 2004), (Guilfoyle, 2011), (Public Administration Committee, 2014)). Where Fung's "targeted transparency" is in play, or later versions of transparency, in the form of open government data released via the Web, there is also the problem of issue framing for those involved in producing and analysing data. "Lack of congruence between the goals of policymakers and those of information disclosers and users . . . misinterpretation of information by disclosers or users, often owing to various kinds of cognitive bias." (Fung, Graham, and Weil, 2008, p.71). We will return to the problem of framing, misinterpretation or cognitive bias.

If we add to this the framing that is already in play, according to writers such as Lee and Leets, (Lee and Leets, 2002) and Innes, (Innes, 2004) it becomes even more complex to sift through social, psychological and policy effects brought about via transparency. So while we can see that impacts might be measured using on the one hand, black box, subjective, cognitive phenomena and on the other, wider more social ones such as changes in police performance measurement, or changes in people's habits in response to knowledge about crime, we have to narrow our focus to what is both observable and isolatable, without losing nuanced information.

John Stuart Mill suggested that we cannot suppose that social phenomena depend on one causal factor or law of human nature, with others producing only trivial effects. He wanted a science of society, modelled on the observations and science of astronomy.

"These thinkers perceive (what the partisans of the chemical or experimental theory do not) that the science of society must necessarily be deductive. But, from an insufficient consideration of the specific nature of the subject-matter—and often because (their own scientific education having stopped short in too early a stage) geometry stands in their minds as the type of all deductive science—it is to geometry, rather than to astronomy and natural philosophy, that they unconsciously assimilate the deductive science of society." (John Stuart Mill, 1886, Chapter VIII. Section 1.). In the years between his writings and the creation of the web, it seemed that such observations were hard to make accurately. However, with the advent of open data mediated by the web in the name of transparency, the surveillant properties inherent in the web itself allow us another set of metrics. These act as an observatory of social phenomena, an observatory that allows us to watch data as it leaves the web, and harness the behaviours of those directly interacting with it via the Web that mediates it and before the data is lost in action.

We can develop for example, marketing impact metrics to understand how people react to the data and information on the site itself. Using a bottom-up and middle-out approach, we focus on what happens to the data that Police.uk mediates, immediately after the mediation occurs, using the qualities inherent in the World Wide Web, in order to do so. We then consider our results and both what they mean for the data, and our account of impact.

### 2.1.6 Police.uk Analytics

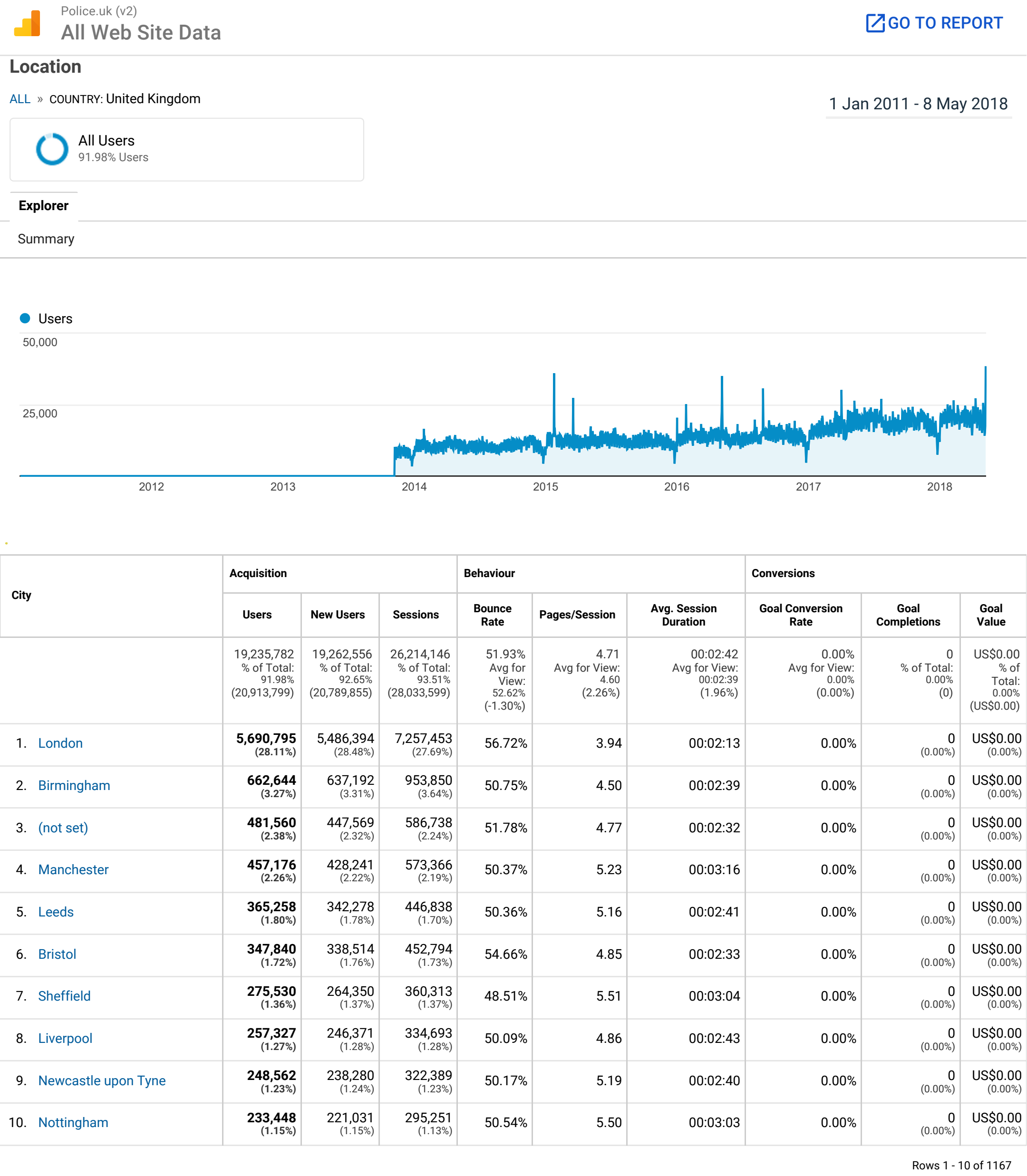
In marketing terms Police.uk is a success. The global competitive intelligence site, Alexa.com showed that in August 2014 Police.uk was ranked 3,537th in the UK, and 104,213 in the world. It estimated bounce rate as 25.8%, daily pageviews at 7.2%. Police.uk has a number of analytics packages attached to it. Analytics packages are often used by marketers to measure engagement: how many visitors does a site get, where are they from, which are the most popular locations both for visitors to come from and for them to visit? Marketing goals are generally to increase visits to a site, with the underlying assumption being that increased visits increase the likelihood of visitors clicking on the areas designed to tempt them most into making a "call to action". In commercial cases, success is measured in sales, and Return on Investment (R.O.I.) demonstrated via evaluating costs of running the site, paying the person who maintains it and feeds it with data, against sales generated, for example. Where sites are not specifically or obviously selling something: a product, a service, information, an ideology, it is still generally assumed that the site exists for a reason, and that tracking visitors' movements on it helps to determine whether the site is fulfilling that function, but calculating R.O.I. becomes harder, as it begins to involve intangibles like brand awareness and goodwill.

At the time of initial writing, Police.uk measured engagement through the use of Google Analytics, a free, client-side package, and Webalyser, a free server-side package. The data produced ostensibly covers most visitors (except for those who do not have Javascript enabled). It is discoverable whether visitors arrive through identifiable networks such as universities, corporations or the media, whether they visit from phones or tablets and which browser they use. It can be discovered (sometimes) which keywords bring visitors to a site. Have they come via search, (from large search-engines such as Google, Bing, Yahoo, Yandex or Baidu for example) or directly? Do they come from social networking sites, such as Reddit, 4Chan, Twitter or Facebook?

From Jan 01, 2011 up until November 2013 there were 23,026,506 sessions with 16,952,493 new users, who on average looked at 5.44 pages per session. In the six months between November 2013 and August 2014, most visits came from Lon-don, (23.63%), Birmingham (4.03%), Manchester (3.4%), then Leeds, Preston, Bristol, Sheffield, Liverpool and Newcastle.

Around 32% of hits were from search traffic, 40% referral traffic and 26% direct traffic. Visitors viewed on average 5.44 pages for each visit and spent over 4.5 minutes on the website. Approximately 70% of visitors were new to the site, 29.75% were returning visitors. The most popular keyword searches were police, crime map, Police.uk, policeuk, police crime map, non emergency number.

The analytics changed at this point, so that from November 2013 there were 28,033,596 sessions with 20,789,846 new users, who on average looked at 4.60 pages per session. See Fig. 2.1 for a detailed overview of U.K. interactions as of May 2018.



Police.uk (v2)
All Web Site Data

Location

ALL » COUNTRY: United Kingdom

1 Jan 2011 - 8 May 2018

All Users
91.98% Users

Explorer

Summary

| City | Acquisition | | | Behaviour | | | Conversions | | |
|---|---|---|---|---|---|---|---|---|---|
| | Users | New Users | Sessions | Bounce Rate | Pages/Session | Avg. Session Duration | Goal Conversion Rate | Goal Completions | Goal Value |
| | 19,235,782<br>% of Total: 91.98% (20,913,799) | 19,262,556<br>% of Total: 92.65% (20,789,855) | 26,214,146<br>% of Total: 93.51% (28,033,599) | 51.93%<br>Avg for View: 52.62% (-1.30%) | 4.71<br>Avg for View: 4.60 (2.26%) | 00:02:42<br>Avg for View: 00:02:39 (1.96%) | 0.00%<br>Avg for View: 0.00% (0.00%) | 0<br>% of Total: 0.00% (0) | US$0.00<br>% of Total: 0.00% (US$0.00) |
| 1. London | **5,690,795**<br>**(28.11%)** | 5,486,394<br>(28.48%) | 7,257,453<br>(27.69%) | 56.72% | 3.94 | 00:02:13 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 2. Birmingham | **662,644**<br>**(3.27%)** | 637,192<br>(3.31%) | 953,850<br>(3.64%) | 50.75% | 4.50 | 00:02:39 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 3. (not set) | **481,560**<br>**(2.38%)** | 447,569<br>(2.32%) | 586,738<br>(2.24%) | 51.78% | 4.77 | 00:02:32 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 4. Manchester | **457,176**<br>**(2.26%)** | 428,241<br>(2.22%) | 573,366<br>(2.19%) | 50.37% | 5.23 | 00:03:16 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 5. Leeds | **365,258**<br>**(1.80%)** | 342,278<br>(1.78%) | 446,838<br>(1.70%) | 50.36% | 5.16 | 00:02:41 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 6. Bristol | **347,840**<br>**(1.72%)** | 338,514<br>(1.76%) | 452,794<br>(1.73%) | 54.66% | 4.85 | 00:02:33 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 7. Sheffield | **275,530**<br>**(1.36%)** | 264,350<br>(1.37%) | 360,313<br>(1.37%) | 48.51% | 5.51 | 00:03:04 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 8. Liverpool | **257,327**<br>**(1.27%)** | 246,371<br>(1.28%) | 334,693<br>(1.28%) | 50.09% | 4.86 | 00:02:43 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 9. Newcastle upon Tyne | **248,562**<br>**(1.23%)** | 238,280<br>(1.24%) | 322,389<br>(1.23%) | 50.17% | 5.19 | 00:02:40 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |
| 10. Nottingham | **233,448**<br>**(1.15%)** | 221,031<br>(1.15%) | 295,251<br>(1.13%) | 50.54% | 5.50 | 00:03:03 | 0.00% | 0<br>(0.00%) | US$0.00<br>(0.00%) |

Rows 1 - 10 of 1167

FIGURE 2.1: Google Analytics for Police.uk

### 2.1.7 Problems with Quantitative Analysis

With regard to such considerations, there is a danger that “the practicalities of measurement become more important than what is being measured”. (Young, 2011, p.16). In *The Criminological Imagination* Jock Young sets out ways in which under-standing about the causes of crime becomes mired in various derogating issues, exemplified by Young’s presentation of the issue of “the New Genre”. Young suggests that competing theories become one dimensional by virtue of how they are operationalised for the sake of academic or political treatment, that data comes from past studies or survey firms, and that criminologists themselves are distant from crime, ”hidden behind a wall of verbiage... the barrier graphited with the Greek letters of statistical manipulation...” (Ibid.). This context allows us to consider that for example, although keywords can be a clue, and long-tailed searches in particular can suggest states of mind in users, behavioural tracking is still imprecise in mapping from a user’s click-stream to their state of mind. (See (Montgomery et al., 2002, p.579-595) for a discussion of some of the ways in which path analysis and ecological models of food-gathering behaviour are used in order to infer consumer goals and predict behaviour by academia and business.)

Theoretically Google Analytics captures all visits to the site, however it is reputed to under-report visitors especially on sites that are Content Management System-based, (C.M.S.) or have little H.T.M.L. and also according to where the tracking code is located on the page, (Google Inc., 2011), so it is not certain that all visits are being captured. In fact there is a marked contrast with numbers reported by other tracking software, such as Webalyser. It is not initially possible to see I.P. addresses of visitors to pinpoint specific locations, and thus understand visitor behaviour more, because of privacy concerns about personally identifiable information; however, these addresses are usually (imprecisely) retrieved by other analytics packages via geolocation. (For example, Statcounter produces I.P. information). Google has introduced secure socket layer encryption (S.S.L.) for users who visit websites using Google’s own browser, Chrome, which can be logged into in order to be able to synchronise data such as bookmarks and favourites and browser settings. These visits are not identifiable via keyword. (Ibid.)

A web-designer’s solutions to the loss of keywords given in analytics could be to promote keyword densities of specific words on specific pages so that if the landing page is visible in the results then the analyst knows that the targeted keyword might have brought the user to that page. But this would not suit the design of Police.uk, which is pulled together in a C.M.S.-like way, and it also undermines Google’s insistence that it wishes to promote good web-design that is user friendly and that follows W3C recommendations. (Google Inc., 2012). Such keyword-packed design can make pages “spammy” or give the appearance of trying to manipulate the search engine, and thus cause the pages to be punitively removed and blocked from Google’s indexes, so that they are no longer served in response to relevant keyword searches, such as those mentioned above.

So marketing-based metrics are available, but imprecise. But of course, letting us know about crime is not, or certainly should not be, primarily a commercial enterprise on the part of the government. Although the knowledge economy ostensibly supports the use of data as a by-product that goes into fuelling apps which can be sold using the idea of the long-tail, we might suppose, given the pitches for the site, that its first purpose is to inform, in the name of accountability. So in considering impact we should ask whether using only the metrics that commercial companies deploy is going to really tell us more about how to hold our government accountable for its implementation of crime policy. Can we use these data without falling into what Jock Young called the "hubris of positivism", i.e. the behaviours of the New Genre described above, that make shaky comparisons between theories and use data of dubious provenance, focus solely on statistical methods without understanding of they are appropriate and see the researcher into crime far removed from the phenomenon that they are seeking to understand?

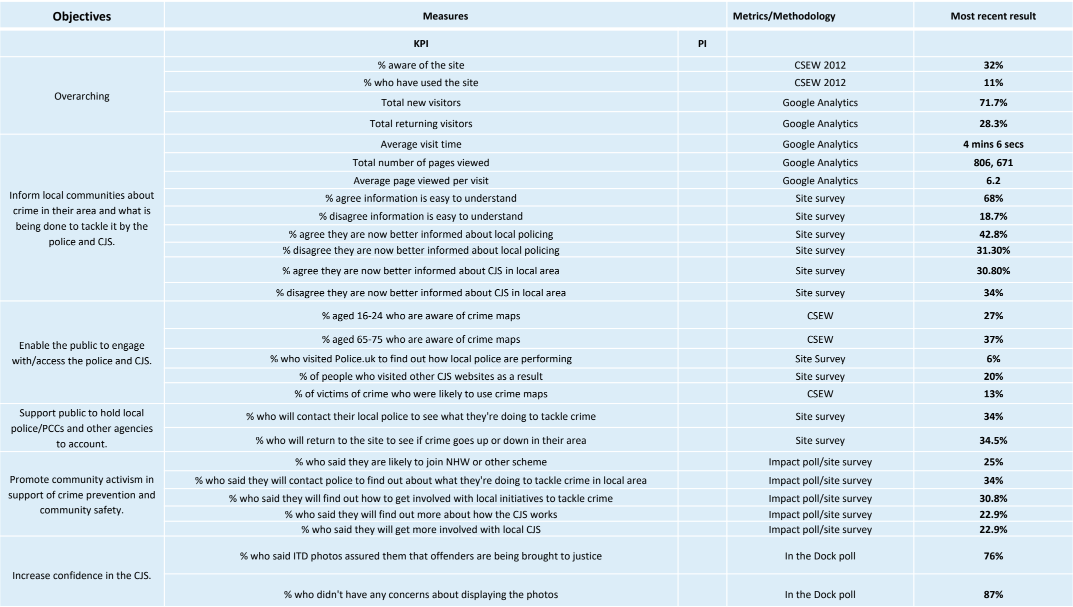
Numbers alone do not help us understand impact: whether on society, the economy, policy or the thoughts, beliefs and fears of visitors to the site. The numbers do not allow us to see what effect publishing crime data on the web of data and making it accessible via online social networks has for transparency and accountability, and what some of the consequences of supplying this crime data might be. As researchers, we wish to know more about the meanings of these visits, and their social context. We wish to understand current concepts of crime and if these are contested. We would like to understand whether visitors use negative, positive or neutral language to frame their lived experience of Police.uk, and whether fear of crime is increased, rather than assurance about how to deal with it or prevent it. We would also like to understand whether the dissemination of this data enables the spread of knowledge, such that people ask deeper questions about crime and whether conversations become more criminological, or focus on how people can pull apart controversial topics, to discover either criminological / epistemological tensions, or empirical problems that can be solved via people's ownership of this knowledge.

### 2.1.8 Qualitative Analysis

In order to discover some of this impact, a first approach, used by Home Office itself, has been to see qualitative analysis as a meaningful corollary to the numbers on the site. This sort of qualitative analysis includes site surveys and interviews with key user groups such as Neighbourhood Watch, the police themselves, as well as focus groups, and comments from a virtual user group. This research has been carried out by the Home Office and is ongoing. The Home Office has conducted qualitative research into the attitudes and needs of some user groups. They have attempted to capture feedback from representative samples of visitors to the site. However there are problems with selecting user groups – people who spring to mind as being a viable user group, probably do so as they are vocal and obviously present – as in the case of Neighbourhood Watch, the police themselves and the P.C.C.s.

### 2.1.9 Some Problems with Qualitative Analysis

Site surveys garner relatively few responses, and of the questions responded to, answers tend not to explain motivation for using the site in any depth, or give any real indication of what the impact of such mapped crime data is. Of the user groups, although these are easily identifiable as users, they might not be representative of all those using the site. We can also see that the survey shown here is still couched in terms of business performance – Key Performance Indicators are mentioned with all the embedded assumptions to do with performance and impact. (Key Performance Indicators are quantifiable metrics commonly used in conjunction with targets to demonstrate that someone or some organisation has done what they have said that they would do - see (*Guide to key performance indicators Communicating the measures*



| Objectives | Measures | | Metrics/Methodology | Most recent result |
|---|---|---|---|---|
| | KPI | PI | | |
| Overarching | % aware of the site | | CSEW 2012 | **32%** |
| | % who have used the site | | CSEW 2012 | **11%** |
| | Total new visitors | | Google Analytics | **71.7%** |
| | Total returning visitors | | Google Analytics | **28.3%** |
| Inform local communities about crime in their area and what is being done to tackle it by the police and CJS. | Average visit time | | Google Analytics | **4 mins 6 secs** |
| | Total number of pages viewed | | Google Analytics | **806, 671** |
| | Average page viewed per visit | | Google Analytics | **6.2** |
| | % agree information is easy to understand | | Site survey | **68%** |
| | % disagree information is easy to understand | | Site survey | **18.7%** |
| | % agree they are now better informed about local policing | | Site survey | **42.8%** |
| | % disagree they are now better informed about local policing | | Site survey | **31.30%** |
| | % agree they are now better informed about CJS in local area | | Site survey | **30.80%** |
| | % disagree they are now better informed about CJS in local area | | Site survey | **34%** |
| Enable the public to engage with/access the police and CJS. | % aged 16-24 who are aware of crime maps | | CSEW | **27%** |
| | % aged 65-75 who are aware of crime maps | | CSEW | **37%** |
| | % who visited Police.uk to find out how local police are performing | | Site Survey | **6%** |
| | % of people who visited other CJS websites as a result | | Site survey | **20%** |
| | % of victims of crime who were likely to use crime maps | | CSEW | **13%** |
| Support public to hold local police/PCCs and other agencies to account. | % who will contact their local police to see what they're doing to tackle crime | | Site survey | **34%** |
| | % who will return to the site to see if crime goes up or down in their area | | Site survey | **34.5%** |
| Promote community activism in support of crime prevention and community safety. | % who said they are likely to join NHW or other scheme | | Impact poll/site survey | **25%** |
| | % who said they will contact police to find out about what they're doing to tackle crime in local area | | Impact poll/site survey | **34%** |
| | % who said they will find out how to get involved with local initiatives to tackle crime | | Impact poll/site survey | **30.8%** |
| | % who said they will find out more about how the CJS works | | Impact poll/site survey | **22.9%** |
| | % who said they will get more involved with local CJS | | Impact poll/site survey | **22.9%** |
| Increase confidence in the CJS. | % who said ITD photos assured them that offenders are being brought to justice | | In the Dock poll | **76%** |
| | % who didn't have any concerns about displaying the photos | | In the Dock poll | **87%** |

FIGURE 2.2: Results from Home Office Qualitative Research (2013).

*that matter* *Connected Thinking PWC* 2007)). However, there is on the site a way of getting far more detailed responses to the publication of the crime numbers that has both a "numbers" approach and the knowledge-based dimension. Google Analytics can pick up visits to websites and follow the publication of links from that site to a wide span of various social networks, where not only the links are to be found but the context that they are being used within and discussion of what they mean. This can be done equally by sending out crawlers to explore the web and bring back links from the site with associated comments. We decided to collect these remarks and analyse them, but still needed to refine our framework in order to work out how to do this in the most rigorous way, which also meant thinking about where our approach was qualitative and where quantitative, and how best to approach the problem of a deeper understanding of impact. In The Joy of Concrete, (The Open University, 2011), the problems of synthesising methodological approaches are discussed. We can use qualitative or quantitative methodological approaches from a variety of perspectives relating to: tradition, assumptions, methods, data collection, kinds of data, participants and sample sizes, types of analysis, role of the researcher, and kinds of outcome. However, when we ask these questions using multidisciplinary methods it is possible for methodological rigour to be lost as a result of confusion over epistemological approaches. For example - are we using or developing a real understanding of theories of knowledge? What constitutes certainty? What about our ontological beliefs - what really are the things in the system we are examining? (See also discussion in Chapter 8).

This has been the case for much work on Open Data and its impact. While there have been many practical, office or organisation-based endeavours to gauge efficacy, there is a tension between producing academic work with an emphasis on theory, epistemology and ontology, and work that is produced "out in the field", where managers may be writing reports to justify their use of Open Data.

It would be simplistic to say that such managerial reports are, by nature, skewed towards looking for success and less objective than academic research should be - indeed there is much to show exactly the same sorts of biases occurring in academia as in other workplaces because of the perverse incentives examined by this thesis.

However, while auditing, managing and learning from projects are valuable goals in themselves in the context of work, there will be cultural issues at the very least over interpretation of results "in the wild". Robust academic research that informs more pragmatic approaches, has the scope to look into the philosophy of data, and the socio-cultural impact of the effects of such data, and should be scientific enough to look for failure as well as success, as well as unpicking what these concepts mean, in a variety of contexts. We wished to understand more about the crime data and what it means and where it comes from and more about the effects of knowledge on people in terms of transparency. We therefore broke the problem down into i) context (what exactly is Police.uk?) ii) numbers and iii) understanding. Having gone through this overview we turn now to the literature that will help us to understand the research questions, given this background.

## 2.2 Literature

### 2.2.1 Knowledge of Crime

#### Transformative Power through Knowledge: What is Open Data?

Research into open data is fast-growing, and global, and often appears under the literature heading of “open data as a tool for transparency”, with reference to its “transformative powers”, especially the ways in which the technologies that mediate the dissemination of this data – notably in the form of the World Wide Web - can change governments and democracies as a result of large-scale, aggregated, distributed and automated dissemination. (Davies, 2010), (The Cabinet Office, 2012), (The White House, 2009), (Gotze and Pederson, 2009).

The mainstream literature around Open Government Data addresses social, technical and legal issues, such as anonymity and privacy, the technical intricacies of reuse, and whether law should change to support publication of open data. (Hara and Hall, 2012), (Shadbolt, 2011), (The Cabinet Office, 2012), (Berners-Lee and Fischetti, 1999), (Difranzo et al., 2010). Before looking at some of these issues we briefly consider the background to open data where it is released in the name of governmental transparency.

#### Transparency and Democracy

“Open government is the governing doctrine which holds that citizens have the right to access the documents and proceedings of the government to allow for effective public oversight. In its broadest construction it opposes reason of state and other considerations, which have tended to legitimize extensive state secrecy. The origins of open government arguments can be dated to the time of the European Enlightenment: to debates about the proper construction of a then nascent democratic society.” (Longo, 2013).

“Modern democracy calls for transcending electoral processes and requires the consolidation of state mechanisms that promote transparency, ethics, the right to information and citizen participation. The combination of these principles of open government results in a more efficient, effective and participatory state. More importantly, these elements allow citizens to own political processes, reaffirming the legitimacy of the democratic system. Without them, the state will be unable to regain the citizenships’ trust.” (Rivera, 2015, para 1.).

The above statements show broadly how transparency relating to the practices of governments is strongly allied to the expression of democracy. Democracy can be expressed in many ways: representative, direct, parliamentary, presidential, majoritarian, consensus, established, partial, participative and deliberative. (Schmidt, 2002, p.147). These are not mutually exclusive. Democracy in the U.K. is described by its own Parliament as representative, as opposed to consisting of, “a small landowning elite whose priorities were their own power and prosperity.” (www.Parliament.uk, 2015). There is also emphasis on deliberation as part of participatory democratic workings. Democracy is a central principle for 123 of the world’s countries, out of 193, and fundamental to how many Western states view themselves. In the United States, the State Department sees the presence of democracy as underpinning national security. Democratic nations are more likely to “secure the peace, deter aggression, expand open markets, promote economic development, protect American citizens, combat international terrorism and crime, uphold human and worker rights, avoid humanitarian crises and refugee flows, improve the global environment, and protect human health.” (U.S.Department of State, 2018).

However, several authors have sounded a warning note about the need not to be self-congratulatory. The Open Data agenda situated alongside or within the Transparency programme ideologically cites democracy and freedom as vital principles that Open Data supports. We suggest that democracy could more usefully be considered as a mechanism that produces certain desirable outputs under the right systemic conditions, rather than an ideal or ideology. "It is important, then, not to take the existence of democracy, even liberal democracy, as cause for self-congratulation." (Dahl, Shapiro, and Cheibub, 2003, p.38). Equally, among the transparency in the name of democracy discourses, there is often little consideration of the perverse incentives often deployed in its name that fatally erode the outputs of democracy. Assumptions are made about the relationship between freedom and democracy that deserve further examination. Freedom is much cited within the transparency and democratic context. Freedom, or liberty, is seen generally, as a core positive outcome around which these concepts are built. Various nouns such as Information, Speech, Thinking and Expression are attached to freedom; however although the conjunction of terms is common, understanding what is entailed by such conjoining necessitates some unpicking.

#### Accountability

Agnes Callamard characterises the cluster of accountability, free expression and transparency, as part of a human rights framework that is essential to the functioning of democracy. “Accountability is a broad term underpinned by many different understandings and applications. From a human rights standpoint, accountability is often juxtaposed with other terms, such as responsibility, duties, or obligations, which are often used in reference to the state although increasingly to nonstate actors as well.” (Callamard, 2008, p.1212). Not only is it often juxtaposed, but among open data practitioners the terms are interchangeable, viewed purposively: “Accountability serves similar purposes as do responsibility (and liability), including protecting the rule of law, and paving the way for compensation and satisfaction of victims.” However, unlike the concepts of responsibility and liability, accountability plays a more broadly valent part: “...It is also essential to the protection of democratic values and key to securing control of public power.” Callamard suggests asking who is to be accountable to whom for what, using which measures of reporting and with what consequences if accountability fails, following Curtin, who asks what are the aims of accountability, who are the actors involved, the institutions, the processes and the levels of accountability? See (Curtin and Nollkaemper, 2005, p.9) for discussion of five elements to be examined: "the aims of accountability, the actors involved in processes of accountability, the institutions to which accountability must be rendered, the process of accountability, and the levels of accountability".

Callamard also refernces Keohane's work on accountability (Keohane, 2005, p. 15): "...Rulers generally dislike being held accountable. Yet they often have reasons to submit to accountability mechanisms. In a democratic or pluralistic system, accountability may be essential to maintaining public confidence." Keohane points out that accountability itself is one of the limiting dimensions of power: "...we can expect power holders to seek to avoid accountability when they can do so without jeopardizing other goals... To discuss accountability is to discuss power". In a systemic view it becomes clear that maintaining trust is key to those in power who agree to accountability mechanisms, even while they might be seeking to evade the consequences of what they have agreed to, while negotiating the terms of their mutually agreed cooperations. "International regimes (are) not...weak substitutes for world government but...devices for facilitating decentralized cooperation among egoistic actors" - a concept that is now picked up by graph-theoretical approaches to understanding security on the world stage, and which can be brought into accountability discussions with some relevance.

Thus we see the promise of the provision of freedom through democracy, via a balance maintained through transparency-mediated accountability allowing the behaviours of those in power to be held up to the scrutiny of the public. Before considering accountability mediated via transparency (for example the release of mapped open crime data) we should briefly consider what is meant by freedom.

Berlin wrote about two varieties of freedom (or liberty): "What is the area within which the subject... is or should be left to do what he or she is able to do or be without interference from other persons? The second... is involved in the answer to the question 'What or who, is the source of control or interference that can determine someone to do, or be, this, rather than that?'" (Berlin, Hardy, and Hausheer, 1998, p.194).

"Liberty in this sense is principally concerned with the idea of control, not with its source. Just as a democracy may, in fact deprive the individual citizen of a great many liberties which he might have in some other form of society, so it is perfectly conceivable that a liberal-minded despot would allow his subjects a large measure of personal freedom." Berlin's argument here is that understanding liberty in terms of both what it allows, and what the controlling mechanism is, give us a more flexible understanding of how it might work within democratic systems. Democracies can and must impose constraints on individuals just as much as tyrants can allow freedoms not considered within democracies. "The despot who leaves his subjects a wide area of liberty may be unjust, or encourage the wildest inequalities, care little for order, or virtue, or knowledge; but provided he does not curb their liberty, or at least curbs it less than many other regimes, he meets with Mill's specification." Berlin's account shows that freedom in this sense is not logically connected with democracy or self-government. "Self-government may, on the whole, provide a better guarantee of the preservation of civil liberties than other regimes, and has been defended as such by libertarians. But there is no necessary connection between individual liberty and democratic rule."

When examining crime data as a means by which we can improve public access to information about policing and crime, these contrasting views of liberty or freedom within the concept of democracy are crucial to consider. Does crime data released in the name of democratic transparency remove constraints on our liberty, or does it provide more self-determination? Are there tensions between self-determination and constraint on liberty? Maud suggested that "the regular publication of government spending is holding our feet to the fire all year round, not just at election time. . . .the prize is effective, personalised, 21st-century democracy." (The Cabinet Office, 2012, p.5).

This comment suggests that in general, Open Government Data, (O.G.D.) on government spending and other activities might curtail selfish, deliberately corrupt, thoughtless or stupid acts by the government. These sorts of acts are seen as undemocratic. As the government's freedom to perform thus is constrained, there is a perception of a corresponding lack of constraint on our part; maybe in part that by monitoring and understanding better the movements of this source of control or interference, we have more self-determination: Berlin's "positive" freedom.

Open Crime Data does not fit so easily within this account of transparency as other forms of open data. The police, as representatives of ourselves whom we have consented (in the U.K.) to let perform their duties in the name of our democracy, exert control, sometimes pre-emptively, over a population in specific ways. Some accounts, such as ongoing debates on surveillance versus privacy, (Anderson, 2014), or misuse of stop and search powers in England, (Garland, 1992), would have it that this policing in itself is curtailing of our general freedoms as well as the specific freedoms of those who act against regulated society in large or small ways. Does policing increase the liberty of the law-abiding while decreasing the liberty of those who wish to bend or break the law? Talking to law enforcement officers show that many consider the law to lag grievously behind what is needed in order to produce a peaceful society where individuals are free from harm. In fact, these differing accounts are not dichotomous but part of a wider narrative. Whose feet (as Maud said) is Open Crime Data holding to the fire?

Simplistically, we can see Open Crime Data as data that is a sub-set of crime data and Open Data. While Open Data has been around for at least twenty years, crime data has a history going back to the first censuses and surveys of populations by governments and states – tying it strongly to some clearly defined bodies of literature on statistics and surveillance, and more recently "Big Data". This creates a complex and shifting pattern running through the transparency tale, with one strand being the impact of publication of open data on societies in terms of innovation and mar-kets. This is where we see open data arrayed alongside closed, or secret data, which alignment leads us to the concept of information asymmetry. Information asymmetry is of course implicit in the concept of open data, when we refer back to the core concept of power within democracy. We examine the ways in which data can be pushed and pulled through both open and closed systems, as a result of regulation and F.O.I. legislation, but point to some tensions in the idea of automated release.

Alongside the concept of the information asymmetry that contextualises Open Data, we see the theme of surveillance emerge in our literature review and the notion of a calculus of power embodied in data, as its production and consumption in the wider data market shifts according to where it lies on the spectra above. The literature around the data of information asymmetry and surveillance will have a particular role to play in the production of our methodology for understanding the impact of the sort of open government data under scrutiny: Open Crime Data.

### Ethical Systems of Governance

At the time of starting this research, concerns with ethics in research and Big Data were bubbling under in terms of apparent relevance to the whole project. However the question of ethics in general, as engendered by automated systems, such as those which parse Big Data, is now becoming more important with the emergence of discussion on Artificial Intelligence as it comes to play in policy-making and governance. Transparency should help to create ethical systems of governance. Aristotle suggested that virtue emerges from practice, and is not inherent in those who govern; therefore much care must be taken in constructing the rules for such systems. In the U.K., this is enshrined in "the rule of law" - particularly where we ask to expose the consequences of the everyday actions of those working to serve a population. Governments have gathered, held and disseminated the data that precedes understanding these consequences, data about their own workings, and about those whom they represent, for millennia. In some cases this information has been made partially public, but not always at the behest of the government, or as a result of policy. Notable examples of censuses are in the Book of Numbers, King David's Census and the Domesday Book. Censuses delineate salient characteristics of those who are governed, and can also be seen as expressing information about the effects of housing, health, crime and education policy on those populations.

There are various elements or actors in these systems of governance so far, explicitly we have: governments, their populations, data created by governments about their populations and finally, practices, rules, regulation or policy and the knowledge that comes from this data, with reference to mediating technologies. Implicitly there are also trust, mistrust and power as parts of this system.

### Brandeis

Before our examination of transparency in the 21st Century, we refer first to Brandeis, the nineteenth-century American founding "father of transparency". It is useful to understand some of the concerns at the time in the United States, as these are still in play, only on a now more global stage, and many of the issues current in the U.K., as we show later. In the 1890s, concern in the United States was with how large companies and governments could be held accountable, often in the context of white-collar crime. Brandeis worked on cases such as the Massachusetts liquor reform laws, where state legislators were being bribed by lobbyists. In another case he worked with merchants opposing the attempt by a Boston subway group to gain a monopoly over the city's transport system. Brandeis' focus was on data collection and dissemination (payroll data for public officials), public meetings, and Non Government Organisations (N.G.O.s). The sorts of corruption or dubious practices he was concerned with were what many analysts believe led to the 1929 Wall Street Crash. Brandeis then helped to create the Securities Act, designed to prevent such an event re-occurring. Behind the evolution of the transparency derived from the 1929 Wall Street Crash, is that America's financial history since the crash is owing, in part, to the regulation that emerged in the name of transparency. Not only did Brandeis understand that publicity acts as a scourge on those who do wrong and are shielded by their power, he also observed how technology would come to play within privacy debates. "Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that 'what is whispered in the closet shall be proclaimed from the house-tops'". (Brandeis and Warren, 2010).

In a letter to his fiance, Alice Goldmark, Brandeis said he wished to write “a sort of companion piece” to his influential article on “The Right to Privacy,” an article on “The Duty of Publicity.” He wrote to her, “about the wickedness of people shielding wrongdoers and passing them off (or at least allowing them to pass themselves off) as honest men...If the broad light of day could be let in upon men’s actions, it would purify them as the sun disinfects." (Urofsky and Levy, 1971, p.100). Just as the word “accountability” has shifting meanings, so “publicity” in this sense means something like what we might call ’public relations’, as well as making information widely available to the public, in the sense that we now think of as transparency. (Stoker and Rawlins, 2005). “Publicity is justly commended as a remedy for socialand industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman...The right to life has come to mean the right to enjoy life - the right to be let alone, the right to liberty secures the exercise of extensive civil privileges; and the term "property" has grown to comprise every form of possession–intangible, as well as tangible.” (Ibid.).

The delicate balance between individual liberty and privacy, technology, and public accountability starts to become apparent. Brandeis had assembled most of the elements of the equation; technology can threaten the right to privacy of the individual. The desire for privacy that lies behind the right to it can also be brought to bear as the mediative device that exposes the wickedness of those shielding wrongdoers. Brandeis had understood that exposing corporate wickedness through transparency or publicity could be an effective regulatory device, and that the automation provided by market mechanisms could help such regulation. However he did not predict more explicitly the ways in which technology could be used to do this, and also that the right to privacy could dilute the promise of such accountability. We go on to ask whether in particular it dilutes the promise of O.C.D. and as a result, whether it is a truly marketable commodity.

Brandeis wanted to find ways of using market mechanisms to encourage competition. He suggested that gas companies could increase their dividends by lowering their prices. Brandeis was interested in finding algorithmic practices, automated through legislation and programs that did not need human intervention. In opposition to this, the President, Calvin Coolidge said that, “Unfortunately, human nature cannot be changed by an act of the legislature. When practically the sole remedy for many evils lies in the necessity of the people looking out for themselves and reforming their own abuses, they will find that they are relying on a false security if the Government assumes to hold out the promise that it is looking out for them and providing reforms for them.” (Manis, 2003).

Brandeis also might not have accounted for the relationship between seeking to expose financiers funnelling profits towards themselves and the health of markets, such that when Theodore Roosevelt made his famous attack on the “tyranny of mere wealth”, the very act of exposing and discussing crimes related to wealth and corruption had the effect of causing a stock market panic in 1907. So while exposure of potential wrong-doing can drive people to act to change systems, over-exposure can cause the very problems to occur which are being sought remedy for. This is key in our exploration of some of the unintended consequences of “too much” open knowledge, and the concept of knowledge of crime sometimes being a crime in itself.

Coolidge as President, was aware of the need for regulation of some sort, in order to combat what might later be called deviance, and sometimes specifically, white-collar crime, but he also saw that centralised regulation detracts from the independence of the entities that need protecting, and causes complexity, which they were in the process of trying to iron out with regard to banking. Policy was not geared around outcomes, but around mechanisms. When Roosevelt took over, his advisers had wanted power over which companies could publicly trade, but his decision was to follow the advice of those who wanted disclosure as an operating principle for markets. Helping markets to function better was his key concern, rather than making regulators struggle with complexity in a constantly shifting landscape. This struggle between complexity, regulation, and governance is only now starting to be answered by technologies such as the blockchain and the latest evolution of research into Artificial Intelligences – whether or not these currently widely-hailed panaceas prove to be effective in the long term.

### The Present: The 2000 Freedom of Information Act (F.O.I.A.) (U.K.)

While there is much to be said about the road that led to F.O.I.A., both in the UK, and across continents, (Vleugels, 2008; Davies, 2010) our focus is on the last twenty years. F.O.I.A. appeared in the United Kingdom in 2000. So became enshrined in law the fact that the public has a right to information or knowledge held by the government; what Fung calls "The Right to Know", and according to him, the first phase of transparency. (Fung, Graham, and Weil, 2008). It is interesting to note that when its equivalent was signed in the U.S. in 1966, at the behest of Rumsfeld, Lyndon Johnson was highly reluctant to sign it and refused to even note the act in his diary, or for the signing to take place in public. In an accompanying statement he said that one of the most essential principles in the U.S. was that, "a democracy works best when the people have all the information that the security of the nation permits." (Johnson, Lyndon, 1966), (*LBJ Presidential Library | Research* 1966). This quite succinctly encapsulates the ensuing debate about freedom versus security.

The F.O.I.A. articulates the idea of the public "pulling" information rather than being "pushed" information as part of automated reporting procedures. Fung suggests that the F.O.I.A. came into being as a result of earlier legislation failing, withholding information rather than disclosing. In fact, the F.O.I.A. does not give the public carte blanche to receive all information, it gives official bodies a negotiating mechanism between the public's right to know and the need for secrecy. "The Administration's pursuit of open government has been, as it must be, balanced against important consideration of national security, the needs of law enforcement, governmental privileges, and the protection of personal privacy and business confidentiality, encouraging robust and candid deliberations, among other important interests – all of which also affect the welfare of our citizens. . . reasonable minds may sometimes differ about exactly where the proper balance among these is struck. . . " (The White House, 2011).

### Six Stages in the Evolution of Transparency

Since the F.O.I.A. came into being, the World Wide Web and associated technologies have enabled the dissemination of government information in unprecedented volumes. Fong identifies three stages in the evolution of transparency, the right to know, targeted transparency and technology-enhanced collaborative transparency. Some of Fong's work predates the rise of government-mandated transparency – a return to Mill's conception – ideological transparency with its associa.tion with the open movement. We therefore identify five or six stages in the evolution of transparency, as opposed to his three:

1. Democratic ideological transparency, non-web-mediated, but still mediated by literature and discussion, of the sort that Mill presented. This is associa.ted with the need for transparency in order that democracy works.

2. Financial transparency – which uses push-out mechanisms applied to corporations to avoid having to provide oversight of complex markets (Fung's first generation).

3. Right to Know which came from first generation transparency ending up being often too protective. This is pull-transparency; and applies to companies and governments. People "pull" or ask for data.

4. What Fong calls Third Generation transparency which is crowdsourced, "targeted" transparency, applied to specific policy or problem areas, both within corporates or government policy.

5. A return to ideological transparency which is applied to governments, and comes from them, is both push (accountability) and pull (information, serendipitious) and which goes back to Mill again. It is here that the World Wide Web comes into play and has created a tidal wave of information produced by Governments. This is ideological large-scale web-mediated transparency, and seems qualitatively different to all preceding generations. It returns to Mill's transparency and also brings in what he referred to bureaucratic government as it strongly calls upon the organisational principles and embedded wisdom of government workers, who survive through political change.

6. We postulate the provision of a 6th evolution, which is crowd-sourced data mediated by the World Wide Web and which elucidates the data produced by government departments themselves.

The use of technology-enhanced transparency and open data to hold governments to account is very recent, and raises concerns about privacy. The data that governments often use in order to explain how they are doing is performance data – health, education and crime data are often data generated from measuring the performance of those working in these areas. The work that they do involves interacting with us, citizens, and much of this data is about us. There are serious implications here: first, that because this data is often performance data, and it is created by those whose performance is being held to account, it is a) not necessarily reflective of the truth about those jobs and b) it reflects the fact that people whose performance is under review have often chosen to do their jobs in ways which produce data that works for them. Where targeted transparency comes into play it is likely to be subject to framing - the idea that people working within specific areas shape policy initiatives so as to make them work with their day-to-day targets. The second implication is that because it is data about us, if the data is to be meaningful this will impinge on our privacy. Anonymising data (in irreversible ways) then makes the data less meaningful and so less useful.

Web-mediated transparency gives us the means to discover corruption, perhaps on a scale hitherto unimagined, while creating perverse incentives to generate behaviours that might themselves be seen as corrupt, while at the same time creating concerns about privacy. (See for example: (Public Administration Committee, 2014)).

### Transparency as a Context for Open

We have now examined transparency – its history and the idea of F.O.I.A. in the modern day - as driving impetus for the crime maps available on Police.uk.. We turn next to briefly examine the idea of "Open", as a concept that makes the potential for error much greater, where datasets become freely available with no safeguarding of their provenance. Some of the issues with what we examine in the next sections - Colquhoun's "moral statistics" – scientific, large-scale knowledge of the behaviour of populations, endure today, but with the possibility of even larger misapprehen-sions being generated, as technology makes the act of mapping crime in the name of accountability, an act that can occur in real-time, that occurs en-masse and that generates web-mediated data beyond comprehension. This is multiplied by the fact of these statistics being made freely available – through their being "Open".

### What Does Open Mean?

Within the broad context of "Open", we find that the "Open" in "Open Data" is sim-ilar to that of other "Open" movements, such as Open Source, Open content and Open Access. Debates around open data can sometimes loosely refer to representa-tive materials, platforms and protocols, and the representation of phenomena cap-tured or displayed by these platforms: maps, genomes, connectomes, chemical com-pounds, mathematical and scientific formulae, medical data and practices, cultural data about artefacts, financial data, statistics, weather data, environment data and transport data. Openness is defined by a series of questions with mostly analogue answers: is data available as a whole, and at no more cost than is reasonable for re-producing it, if not available to download? Is it easily read, retrieved and modified? Is it machine-readable, and available to be mixed with other datasets? Can everyone use it?

As we examine the definition of the "Open" in Open Data, we see the emergence of several dimensions such as: the lifecycle of data, where data sit in relation to other sorts of data, what the data are about, how they are sourced or who produces them, who has access and when and for how long, and who can reproduce or modify them. The answers to these exist on spectra or continuaa. There is another strand that returns to "where they are in relation to other sorts of data". Understanding the impact of publication of open data on societies can also be viewed in terms of innovation and markets. This is where we see open data arrayed alongside closed data, which alignment leads us to the concept of information asymmetry.

Models of U.K. government accountability achieved through transparency show data sets being routinely released across government in policy areas, including Health, Education and the Criminal Justice System, as part of a mechanism by which the government can deliver information on its performance, while allowing data-consumers to use the data in a number of ways. We first consider ways of looking at the data referred to in "open data" and how it may come into being. Ackoff's definitions of "data" as opposed to information or knowledge lays out some useful blunt working distinctions. (Ackoff, 1999), (Bellinger, Castro, and Mills, 2004).

### Consumption of Data under F.O.I.A., Versus Regular Release

Consumption of Open Data can potentially be as straightforward as reading a data release that has occurred as a response to a Freedom of Information Act (F.O.I.A.) request, with perhaps one consumer thereby having satisfied their curiosity and the data release–consumption cycle there ended. The data might not have existed as open data before the F.O.I.A. request; it might have existed as knowledge in a depart-ment, which is formalised and released, in response to the question under F.O.I.A.. The data is perhaps not data in Ackoff's definition of being symbols without mean-ing – the very fact of its having been sought in an F.O.I.A. means that it is released as information or knowledge, (i.e. that it is asked for and released within a context that is both meaningful to the requester and the releaser) but under the Open Data aegis. This could be called purposive data, because its very existence as open data is in response to a purposive request.

Alternatively, consumption can occur in response to mandated and regular releases of data explicitly deployed for routine accountability purposes. Some of these data can be defined as open, or as having become open at some point in this request/release cycle if they are freely available for reuse and republication. Where data sits on a platform "waiting" to be used, it could, under Ackoff's definition, properly be said to be data at that point in its lifetime within the information economy, rather than information or knowledge. In this model, when no-one is looking or somehow "cognitively processing" the data, then it is data. (There is a sub-set of arguments here about whether a machine processing data for purposes that have been designed by people and put into practice algorithmically can contingently or necessarily confer meaning – but these can be put to one side for the moment.)

However, before it is put onto such a platform, it may have existed as performance or administrative data, i.e. as information or knowledge. For example, there might have been a record of how many burglaries have been solved within a Police force over a month, how many children got grade "A" at "A" level at a particular school, how many patients have died of T.B. within an N.H.S. structure, how much a government department has spent on stationery; all cases where the data has been collected as a result of an employer wishing to monitor how well or badly an employee who they oversee performs. At this point, according to Ackoff, it would then be information or knowledge. So the lifecycle of Open Data can show a movement between information and data and more information, depending on the reasons for release.

### Meaning and Value

Thinking about combinations of data + data, or information + information, or information + data, for example, leads us to consider why the information/data is so at any one point in the collection-release cycle – who does it have meaning for, how many people and are they the same meanings at any one point during the lifetime of the data? The meaning is what potentially confers value upon data when we want to work out how to create combinations of data that people will pay money for. If the data that we wish to combine with another data set and thus create new value or meaning, already exists as information that is relevant to (has meaning for) only one or two people and we combine it with another data set with the same characteristics, have we created information that sits at the long tail of value within the information economy? And what is the incentive to purchase these combinations of data for its potential consumers? Are they rich and in need of this data and therefore more likely to pay more money for it? Or are there many poorer people for whom these juxtapositions have less meaning and therefore less value? Do we need to look for datasets that potentially have more of the quality of information about them (i.e. meaning already present) for the greatest number of people, who are on the wrong side of an information asymmetry with respect to this information? Could we search for datasets with these characteristics in order to make the information economy work for Open Data?

### Privacy and Meaning

We consider how information becomes meaningful without including information that impinges on privacy. Open Government Data is explicitly licensed as being non-personal. But a lot of the information that we find most meaningful is about people. School data is about our children, how they are performing and how those who teach them are performing. Health data is about us, our own health, and about those who heal us. Crime data is about victims (or objects) of crime and those who try to help them. If data is about performance for example, how people have done their jobs well in some respect or other, how can we find datasets that are both meaningful and that have accurate information, but that do not have privacy problems? We can anonymise; however have we then lost some of the meaning, and therefore potential value? What are the risks of de-anonymisation? How we identify someone becomes ever more contentious, as does the issue of what counts as identification. Google Streetview data can identify us, automatic licence plate readers can identify us, D.N.A. can identify us. We have varying forms of identity too: where we exist online, and marketers collect data about our preferences as we move from web-page to web-page, these collections of preference data can identify us. Not only our names, addresses and National Insurance numbers can identify us; so too can our behaviours. And as it becomes conceivably more possible to juxtapose previously non-contiguous data sets, so too does the possibility of unintended identification increase. Is there an embedded problem about Open Data in the information economy, whereby it is the personal aspects of such data, the aspects that nudge against the personal, that confer meaning and therefore potential value? What economic models or theories of knowledge can we use to make Open Data work for us without spying on citizens, rather than sousveilling governments?

### F.O.I.A. and Regular Release Data in the Information Economy

There is tension between purposive data – data that is released for a reason, to feed knowledge of a particular area where information has been requested, and data that is released routinely and that serves no particular purpose, yet, other than its original purpose. This sort of data is released not so much purposively, but accessibly; the hope is that someone will happen across it and serendipitously reuse it, perhaps by means of combining it with another dataset, and thereby create knowledge in the information economy. If there is a market for this knowledge that has been produced by combining datasets, the resulting "knowledge" can be sold, perhaps in an app, and there may have been created money as well as knowledge in the information economy. There are examples of data being used in this way – though relatively nascent, it is so far a compelling, powerful and already well-demonstrated economic as well as epistemic model.

The tension between purposive and "accessible" data can lead to problems when we consider the advent of the World Wide Web and the unimaginable quantities that data can now be released in, and the therefore unimaginable audiences who consume it.

Lessig wrote, in Against Transparency, "The naked transparency movement marries the power of network technology to the radical decline in the cost of collecting, storing, and distributing data. Its aim is to liberate that data, especially government data, so as to enable the public to process it and understand it better, or at least differently. . . Finally America can really know just who squeezed the sausage and when, and hold accountable anyone with an improper touch. Imagine how much

Brandeis, the lover of sunlight, would have loved a server rack crunching terabytes of data. As a political disinfectant, silicon beats sunlight hands down." (Lessig, 2009, Section 1. *The perils of openness in government*).

But combining the Naked Transparency principle with the World Wide Web can cause problems. Requests made under F.O.I.A. are (and should be) considered minutely: the identity of the requester and their relationship to the people or things the requested data may be about, whether this person has made this request before, how much it costs to produce the data, what the risks might be of releasing the data, and whether, if the request has been made before, the data has substantively changed. Context is key; it is not probable that data under such consideration might be considered to be data that can be routinely released to Web consumers, even if it has been released under F.O.I.A.. The meaning of the information to the requester has been taken into account; this would be difficult to do where the data sits, as data, waiting to be used, by any one, some, or all of its two billion users. Big Data approaches of the sort that Brandeis favoured - systematic, algorithmic mechanisms have to be balanced against security principles.

Looking at the movement of data through these systems of government, and given impetus by the socio-technical apparatus of the World Wide Web, we see a state of flux – of transitions between public and private, of pushes from the government and requests from the public, of ownership of data – where operations are contentious but deemed necessary, it is common practice for government to outsource and thereby avoid oversight – they no longer own data and therefore can hide them away, thus moving from transparent approaches to release to opacity. The fluxive nature of the data means that making decisions about release can be even harder. Additionally, although the data sets can be used for accountability purposes, tracing their provenance leads to us finding that they come from places that might make them incompatible with accountability, or at the least, not ideal. For example, the Russell Group, a self-selected association of twenty-four public research universities in the U.K., argues that F.O.I. regarding universities should be treated as "a special case", saying that unlike other public institutions, University data sets are competitive.

"Our universities recognise the importance of the key objectives of the F.O.I.A. of openness, transparency and accountability and make very large amounts of information publically available to meet these objectives." They first argue that they make available information on finances and staff and student numbers, as well as to verify the quality of their teaching provision, to assess the quality of their research, and they release research data in a timely and responsible manner. "Russell Group Universities have been involved in developing the technologies that enable the open data revolution to be possible, and helping the Government to develop this world-leading initiative. So, it can be seen that universities are not secretive organisations." However, "…There are elements of university-held information which have clear commercial interest, and form part of their competitive nature…We believe that there is a strong argument for the exclusion of some elements of university-held information on the following grounds: comparability of treatment with commercia.l providers for equivalent data, especially competitive information; and risks arising from the inappropriate disclosure of research data." (The Russell Group, 2013).

While agreeing with much of this, with regard to academic data, we suggest that out of the public bodies who have the potential to release data, universities are not alone in being competitive. Public bodies such as G.C.H.Q., M.I.5, S.I.S., the N.C.A. and the Police also compete, in the fight against crime. Complex contracts between industry providing crime-fighting technology, researchers providing analysis and the police, for example mean that opening up such deals for public scrutiny might well give criminals too much information about procedures, as well as other institutions and corporations. It is not difficult to see that there may already be issues emerging to do with how much such organisations should be releasing information in the name of transparency.

When the data sets do provide us with knowledge that can lead to making suggestions about improvements in government, we see a model of democracy occurring with improvements no longer directly mandated by centralised government. We might ask what exactly are the mechanisms by which the public can mandate change, rather than just suggest that it is needed? And if many changes are called for on a grand scale, given the grand scale of the data being delivered, what might this do for democracy? Have we created a more effective, grass-roots democracy, as opposed to previous administrative democracies? Or are there fundamental differences between the ways in which transparency worked to regulate private companies and industry, and when it is applied to open government, in essence, often ourselves? Who should decide on issues such as privacy? Are consultations with the public necessary, when the public are not experts? How should policy be written so that it best balances expertise with the fears or ignorance of the public and the money-making, fear-invoking concerns of the media and security companies for example?

It appears that initial pushes for transparency took place in terms of financial responsibility, and preventing abuse of power from monopolies, although there were complex arguments over remedies and it was not always clear who and of what the monopolies consisted, and where the line should be drawn between abuse and the ordinary workings of capitalism. This flavour of transparency seemed to favour centralised open financial markets where agile transactions that supported the states as a whole were governed by Adams-like mechanisms, rather than depending on knowledge and setting of contextual policy. It was not always clear whether accountability was in place to offset crime, what might later be called deviance, or was just locally evaluated as needed by enlightened thinkers who had their own definitions of wrong-doing that might not translate globally. In some cases legislation was enacted to combat this wrong-doing. It is also not clear whether when companies or governments release information in the name of transparency this actually gives an "account" or narrative of their doings – rather they often give accounts in the numeric sense and in the sense of this information needing de-coding. Accountability in this financial market-regulating context emerges as policing via accountancy, as a result of trawling through highly enmeshed figures relating to corporate entities, some of whom are people and some companies, with complex relationships. We also see that too much disclosure is as potentially harmful as too little, with devastating effects on the very markets whose protection is sought.

### Summary of Tensions

In looking at the actors and technologies involved in the production of open crime data, we came across these problems:

- Where does the data come from, and who produces it?
- Are there other mapped crime data sites that might be better?
- What of the data itself – what is its shelf-life, how does it move through systems and processes?
- What pushes or pulls it, how regularly, is it always open?

- How do the mechanisms of release work? If automated, is there a risk of them destroying the structures they seek to preserve?
- Not all data should be public for security/competition/safety reasons. Not all data can be public where there is out-sourcing. How can we automate such decision-making with Big Data?
- Not all data should be public where the release might threaten safety. How do we know when there is risk?
- A lot of administrative data is performance data and says little about wholesale incidences or risk of crime. It describes policing activities, but has little predictive value for consumers – are there other sorts of value in the system?
- Additionally, some data that is produced actually causes potentially criminal behaviours / loss of life (through no fault of people doing policing).
- How do we decide whether the acts of transparency create more problems than they solve?
- Where data is published that appears to indicate a problem, how does/should the (non-expert) public mandate change?

If we start to answer some of the questions exposed by this look at the literature and the background to Open Crime Data produced as knowledge, the picture becomes complex. It seems that to understand what the knowledge is that Open Crime Data produces, we should see where it sits in its ecosystem. When one traces how the crime data arrives at Police.uk, it becomes evident that we need socio-technical unpinning to understand its provenance better. Before even attempting to measure impact, our first piece of analysis is to understand the context of Police.uk. In addition to the social, technical and legal issues, such as anonymity and privacy, the technical intricacies of re-use, and whether law should change to support publication of open data, there are wider tensions emerging over the balance between security and liberty; how the data compares to that underlying other similar-looking crime maps and what properties it has that we can understand by looking at its context. We also need to examine the sometimes perverse incentives governing the production of such data.

### 2.2.2 The State, Statistics and Crime Data

The first part of the literature review has given us an account of the interplay and tensions between actors within the U.K. democratic system. Underlying this are negotiations about power - we see how democracy provides a way of transferring power between actors when an imbalance occurs, according to the perceptions of the actors in the system. Transference of power is also germane to understanding some of the idiosyncracies around the role of the police in U.K. society. In the twenty-first century this flow of power becomes partially mediated by the web and its surveillant and sousveilling affordances. We now focus on the development of interest in crime data as a means of understanding these affordances, bringing out some of the key points in the history of crime statistics and crime data that provide context for what is happening now. These are chosen with particular reference to the themes of creating a science of society and the state, surveillance and sousveillance.

With regard to fighting crime, there is a history of irregular transfers of power between the state and the individual: in order to create accounts of the how, why and where of crime, to address crime, incentivise its prevention, persuade people to fight crime and elicit forms of testimony from witnesses. In the U.K. we can see an early use of incentives emerging alongside the concept of regulating or managing crime, with crime-fighting power mandates flowing between the ordinary population and the first 'Officials' such as watchmen. In all of this, numbers, data and knowledge have played a part; often in relation to intelligence, sometimes in relation to resourcing and financing of policing, sometimes in attempting to evaluate the cost to society of crime and recidivism.

In the 14th century we see the use of incentives other than a desire to maintain social order, with rewards being used to persuade people to stop victimless crimes such as "religious nonconformity, Profane swearing and operating an unlicensed alehouse." (Newburn, 2003, p.58). Following this, in the 15th Century, John Graunt and William Petty developed early data on populations in *the Natural and Political Observations Made Upon the Bills of Mortality*. Graunt noted seven "murthers" in 1632, and 15 suicides, and credited the government and citizen guard of the city of London for this low rate. He also suggested that there was a "natural and customary abhorrence of that inhumane Crime, and all bloodshed by most Englishmen" and that mortality and epidemic rates gave "a measure of the state, and disposition of this Climate. . . " (Glenn and Bernstein, 1999, p.80). Petty (who also served as supervisor of the Cromellian land survey) believed that "political arithmetic" was the application of Baconian principles to the art of government – Bacon's paralleling of the Body Natural and Body Politick. Petty surmised that in order to be scientific about government one must use the same principles as when practicing science upon the body. One could reduce puzzles and perplexities to "number, weight and measure". (Lynch, 2001, p.200), (Porter, 1995). It is from these beginnings that we see the concept of prediction and risk emerging – that given certain sets of data, we can show causal links between characteristics of populations and perceived good or bad outcomes.

Publication about crime, with some reference to crime figures, and information about crime, started in the U.K. in the mid-1670s, when an account of a group of trials occuring at the Old Bailey was published, under the title, "News from Newgate: OR, An Exact and true Accompt of the most Remarkable, TRYALS OF Several Notorious Malefactors: At the Gaol delivery of Newgate, holden for the CITY of LONDON, and COUNTY of MIDDLESEX. In the Old Baily." This was from the April sessions of 1674. (The Old Bailey, 2013). Following this, accounts of more trials were published, often by different publishers. These pamphlets were first intended for a popular audience – crime as entertainment – as indicated by the conclusion of the first pamphlet, there were also "divers other tryals which would be too tedious to insert." In what is now a familiar cycle, following the innovation of accounts of crime published for entertainment, regulation was imposed in 1679 when the Court of Aldermen declared that the accounts of proceedings could only be published with the approval of the Lord Mayor. These figures show solely outcomes.

Gottfried Achenwall introduced the term "statistic" in 1748, in Germany, signifying the "science of state". This term was first used in English in 1791 by Sir John Sinclair who published the Statistical Account of Scotland. Four years later, in 1795, payment for services for policing was put in place in the parishes of St James, Piccadilly and St George, Hanover Square. Where it had been everyone's responsibility to police the state, now householders could exchange a duty to serve in the watch, for the duty to pay the watch. A relationship was emerging between incentives and the testimony of human witnesses, as opposed to the witnessing produced by statistics. These events laid the groundwork for the first use of Big Data in creating a state-wide intelligence and surveillance model, which was happening through Patrick Colquhoun, who published "A Treatise on the Police of the Metropolis," (Colquhoun, 1796), and "A Treatise on Indigence." (Colquhoun, 1806). These were studies into the morality and living conditions of the working classes. Colquhoun was remarkable in that he presaged much of what is seen to be modern, multidisciplinary and ground-breaking – he used his business skills and computing power to regulate lower class morality. "The following estimate has been made up from information derived from a variety of different channels. It exhibits in one view, the supposed aggregate of the various depredations committed upon the public in the metropolis and its environs, in the course of a year." (Colquhoun, 1796, p. 608).

### Big Data

Colquhoun aggregated raw data from statistics and from various other reports to create a comprehensive picture of crime, deviance, economics and morality in society and used this data to instigate moral reform. "by connecting sources of national income and the springs of industry and enterprise with the known population of the country, a chart is thus formed of the state of society in 1803, compared with what existed in 1688, one hundred and fifteen years ago; and great as the accession of wealth may appear to be, on a comparison of the two, there cannot exist a doubt of its reality to the fullest extent which is exhibited as to the aggregate amount of the national income". (Colquhoun, 1806, p.21). He had previously presented his findings to Dundas, the Home Secretary who had ignored him; these treatises were addressed now to the populace. They met with unprecedented success and were to shape "the way that many, including ministers and magistrates, came to see the lower class. He would be quoted by journalists and politicians for decades and cited by foreign writers as the expert on the state of John Bull's morals... 'Crimes and criminals, offenses and offenders were posted against each other, with the formality of a ledger account. The account ran high.'...Colquhoun's books would have been yet another diatribe on lower-class immorality had they not been endowed with the authority of statistics." (Wilson, 2014, p. 67).

We see an early appearance of "big data", where the adjective "big" is seen as a relative term that denotes cross-referencing all possible or practicable sources of information, rather than relating to an absolute measure. Colquhoun suggested that from a population of eleven million, 1,040,716 men, women and children received 4,267,985 from parish relief and 3,332,035 from private charities. He was also concerned with those who "lived chiefly by the labour of others, who lived on the proceeds of unproductive occupations such as lewd and immoral women, vagabonds, ballad singers, dealers in obscene books and criminals". (Ibid.) He saw that those who lived off payment in kind evaded the notice of the state and were not bound by its rules. He felt that the lower classes could be brought under the purview of the state by making them work for money, where monetary movements were observed by the state, bringing this population under a measure of control.

The underlying driver for Colquhoun's intelligence model was Jeremy Bentham's Panopticon, with its thesis of control via uncertainty. (Scorgie, 1995) provides an account of the way in which their relationship created a new socioeconomic system. Colquhoun also saw the need for central intelligence gathering in policing, with incentives paid out to informers, which would "not only add energy to the proceedings of the Magistrates, but...form a centre point of action and intelligence." "For police 'all was to be known, noted, enumerated and documented. The conduct of persons in all domains of life was to be specified and scrutinized in minute particulars, through detailed regulations of habit, dress, manners and the like – warding off disorder through a fixed ordering of persons and activities.' (Dodsworth, 2007, p.2). See also (Dean, 2011), (Reith, 1938), (Grieve, 2015) and (Neocleous, 2000). Neocleous suggests that Colquhoun in fact understood that security is imposed on civil society by the state through the exercise of police power. Dodsworth suggests that Colquhoun's focus was on prevention in a way that theorised Policing as a form of governmental practice. An Act of Parliament on 7th July 1794 authorised and required the erection of one or more penitentiary houses, which would be run by Jeremy Bentham. In 1797 the Select Committee on Finance took evidence relating to the financing and general state of the Criminal Police of the Kingdom, which was a "not immaterial branch of public expenditure". There is a clear concern about accounts and accountability. Colquhoun suggested that an improved system for the employment of convicts would be one of the chief means by which the expenses of the police would be diminished, and the morals of the convicts improved.

### Two Varieties of Panopticon – Systems of Distrust

Bentham's Panopticon (one instance of which was built at Millbank and then pulled down in 1902) is resonant in surveillance, crime and policing literature for its "observations of the body of the prisoner", and for the idea of control via uncertainty, rather than omniscience. Earlier models of surveillance were based on Dionysius' ear – being able to hear and know all that was being said. The idea of the Panopticon can be seen as forecasting the premise of a vulnerable society being surveyed by its government – in literature Orwell created Big Brother, premised on Bentham's panoptic ideas, while Michel Foucault wrote about the panopticon as a metaphor for disciplinary societies. (Foucault, 1977).

As Mason says, "The Panopticon was a metaphor that allowed Foucault to explore the relationship between 1.) systems of social control and people in a disciplinary situation and, 2.) the power-knowledge concept. In his view, power and knowledge comes from observing others. It marked the transition to a disciplinary power, with every movement supervised and all events recorded. The result of this surveillance is acceptance of regulations and docility - a normalization of sorts, stemming from the threat of discipline. Suitable behaviour is achieved not through total surveillance, but by panoptic discipline and inducing a population to conform by the internalization of this reality." This echoes the theme of control via uncertainty, self-imposed discipline, and self-censorship. "The actions of the observer are based upon this monitoring and the behaviours he sees exhibited; the more one observes, the more powerful one becomes. The power comes from the knowledge the observer has accumulated from his observations of actions in a circular fashion, with knowledge and power reinforcing each other. Foucault says that "by being combined and generalized, they attained a level at which the formation of knowledge and the increase in power regularly reinforce one another in a circular process." (Moya K. Mason, 2017).

Bentham was to employ the convicts himself – taking responsibility for their good behaviour, so as to prevent recidivism, which he and Colquhoun saw as creating both a financial and moral cost to the nation. Colquhoun went so far as to suggest that Mr Bentham would introduce "ingenious machinery, rendering it practicable for every class of convicts, while in health, to earn a sum equal to their maintenance... twenty-five per cent less than they cost Government at present." Bentham has been much repudiated for creating something – the Panopticon – that is seen as equivalent to the atom bomb or other machinery of servitude or death. He himself had asked, "whether the result of this high-wrought contrivance might not be constructing a set of machines under the similitude of men?"(Bentham, 1787, p.67).

In fact, these physical Panoptica are just one manifestation of Bentham's ideas concerning the state, society, crime and accountability. There is an interesting question (which this thesis does not have scope to properly explore) about whether the self-censorship that Bentham was looking for in his physical Panoptica is entirely a bad thing where the designer of such a system has an inherent belief in the ability of humankind to improve itself, given the right circumstances. More obviously, these constructs led – rather than to the surveillance society much denigrated by the Fourth Estate - to an early version of the same transparency we see apparently in action today. Bentham also created a virtual Panopticon that mirrors the designs for buildings, using numbers, instead of bricks. His idea was of Panoptica within a statistical structure – of numbers produced by the state being used to hold the state to account. Galhofer suggests that he recognised that "accounting is not a straightforwardly neutral device." This is a radical emancipatory accountancy and accountability, serving the oppressed, and, "more broadly to fulfil a duty to humanity, thus making visible the conditions and experiences of the disadvantaged rendering open to view the activities of the relatively powerful…The oppressed are given a voice in Bentham's accounting…aimed at advancing or being consistent with his radically conceived democracy." (Gallhofer and Haslam, 2005, p.56-58).

Although Bentham's ideas, when taken out of context, were repugnant to many (Bentham himself wrote of how Burke famously called him "the spider in his web"), they are part of his thought on systems of distrust. Bentham thought that legislators might also be regulated via disciplinary technologies. Not only this, but that in fact, the presence of secrets created a market in democratic statistics. What is inherent, but not always explored, in the idea of data that become open, is that there is often a "prior conviction that something is hidden and should be revealed… Disclosure is not simply a way of making something public, it is also a way of making the public." (Mortensen, 2006, p.85). Bentham's system of distrust creates a value, and an inherent market, in the act or performance of disclosure, rather than the data that are revealed.

### Panoptica and Systematised Intelligence Markets

Bentham's intelligence market is based on a system of specialised judgements – the idea that (whether or not the belief is true) there is a common belief that not everyone can make reliable judgements about political matters; the many have a belief in a market of a few who give judgements, and those few know, based on having access to prior secrets. Translated to the World Wide Web it is important to realise that there are not just two operations of revelation and concealment – in terms of the web that we have, binaries are expunged and graphs are necessary, because of the scale and complex relationships between those who trade in revelation and concealment. Bentham's systems of distrust are highly relevant to an examination of Open Government Data on the Web, and even more so to understanding the system of distrust that produces crime data. The previous section on knowledge explored this with respect to the passage of data through open and closed systems to create and collapse knowledge and the creation of a market thereby, and Brandeis' notions of publicity as a lever for governing behaviour of corporations and governments.

We have now seen how early systematised intelligence markets emerged; more wholesale crime statistics followed this, first produced by Gregory King, who was Britain's first official demographer, and who worked for John Ogilby. King used taxation information to make assumptions about populations. In 1801, the first reliable modern census was undertaken, and from then on data was recorded in growing quantities, relating to the population of Great Britain. (The Old Bailey, 2013). Desrosieres, writing on the rise of statistics in the 18th Century, speaks of how the intersections of administrative management, the human sciences (or moral sciences) and the natural sciences developed from statistics, the calculus of probabilities and estimates of physical and astronomical constants based on empirical observations. (Desrosières, 2002, p.17). Desrosieres saw an interaction between two forms of authority, those of science and of state, which created a conceptual framework, which combined "reasons for believing something, (supporting decisions that involve the future) and degrees of certainty of scientific knowledge, through the theory of errors." The twin authorities of science and state became more distinct, even as the authority of state used science apparently in order to administer.

"During the same era, the authority of astronomy also underwrote the most novel, controversial forms of scientific prediction: a statistics of society. Starting in the 1830s and 1840s, social science developed on the strength of its use of a technique for measuring error in astronomical observations. Looking at large numbers, statisticians found startling regularities, even in the most disorderly and arbitrary events, such as crimes or - in one renowned example – the number of badly addressed letters that languished unclaimed in Post Offices. These regularities allowed for an extraordinary kind of prediction."(Anderson, 2005, p.20).

### Surveillance

The statistics of society as a means for governing and understanding presents some conundrums. Mill, following Bentham, championed empirical public policy creation and stated the importance of allowing the "widest participation in the details of judicial and administrative business. . . above all by the utmost possible publicity and liberty of discussion," (Mill, 1861, Chapter 6), and that governments should have enough power to "preserve order and allow of progress in the people". Much of the literature and sales pitches for transparency seem rooted in Mill's, and later, Weber's, warnings that, (Weber, Gerth, and Mills, 2009, p.233) "every bureaucracy seeks to increase the superiority of the professionally informed by keeping their knowledge and intentions secret". This describes the state of affairs as is – and transparency is seen as the antidote to such bureaucratic wranglings. While a bureaucracy can "accumulate experience, acquire well-tried and well-considered traditional maxims, and make provision for appropriate practical knowledge in those who have the actual conduct of affairs," they also can stifle individual thinking, or as we might see it, a true market of ideas: "A bureaucracy always tends to become a pedantocracy. When the bureaucracy is the real government, the spirit of the corps. . . bears down the individuality of its more distinguished members." (Ibid).

Mill developed Bentham's utilitarianism with the idea that an individual should have complete freedom from interference from the state, as long as he or she cause no harm to others. At the same time he was an advocate of free markets, with some interventions, such as taxes, if imposed from utilitarian principles. He also suggested that if the state provided subsistence for the criminal poor while undergoing punishment, not to do the same for the poor who have not offended, is to give a premium to crime or to incentivise criminality in those who do not wish to work. (John Stuart Mill, 1886). The idea that the essence of a democracy is the "right to be let alone", or the right to privacy, emerged, as we saw in the previous chapter relating to Brandeis' thinking. However Mill's democracy saw that the mechanisation of census-taking, and as described above, the use of statistics could help to understand and perhaps control those who were governed. Many commentators have seen this taking the measure of a population as a surveillance - in contrast to the idea of the right to be let alone. This information has often been available only for government use, and expressed governments' powers over a population, in the form of closely-guarded knowledge about that population. Looking at the close-guarding of such knowledge, Galison suggests a "classified theory of knowledge" that is dedicated to understanding how much effort goes into "probing the staggeringly large effort dedicated to impeding the transmission of knowledge". (Galison, 2004).

Resnik examines the clash between openness and secrecy, for reasons ranging from the scientific need to protect Intellectual Property, to protecting research participants' privacy or minimising threats to national or international security. (Resnik, 2005). Singh explores the notion of how the act of close-guarding, through encryption, preserves markets, a concept that appears to be in opposition to the idea of opening data to create new markets. He also makes explicit the tension between law-enforcement's need to surveill in order to preserve our liberties, while one facet of liberty is popularly seen as privacy. (Singh, 2000).

### The Present: Ontology, Rationalisation and Democratisation

Mainstream statistics literature can cover many areas: ontology, state rationalisation, and democratisation: Statistics entwined with nation formation are covered by many writers, including : (Wagner, Wittrock, and Whitley, 1991), (Patriarca, 1996), (Prakash, 1999), and (Zureik, 2001). Desrosieres writes powerfully about rationalisation of the centralised state (Desrosières, 2002); as well as (Hacking, 1990) and (Porter, 1995) who show how the shaping of states takes place through the classification of individuals. Ontology covers the role of statistics in classifying groups: (Anderson and Silver, 1990), (Kula, 1986), (Leibler, 2004), (Scott, 1998).

Paraphrasing Wittgenstein on ontology, "only the ontos truly is; the logos is only what we say there is": much of what is here is about the way in which classifications shape definitions of being. Such classification easily extends into a large problem for Big Data systems that can create constructs without human intervention, other than algorithms that can be tweaked, without any real sense of what hidden patterns the algorithm is shaping. Leibler, (Idem. p.135) says, "Statistical mapping, therefore may not be a reflection of "society"; rather, it is the process by which society is defined." – she refers to un-counting non-citizens, as well as counting and marking potential citizens - as we see later in the O.N.S.'s projections. (7.2.1). "The relationship between "science" and the "state" in the establishment of the new Israeli state was silent and dominant as the same time. ... an invisible and neutral institution to the extent that its significant role was imperceptible." (Idem. p.138). Eugenics, using such statistical mapping was implicit in groupings of Mizrakhim and Ashkenazim, and in conceptualising "the demographic problem".

The statistics literature also shows how censuses, I.D. numbers and I.D. cards are mechanisms through which there arises a right to access the civil subject; it describes the apparatus of identification. Groebner, in (Torpey, 2000) and (*Documenting Individual Identity: The Development of State Practices in the Modern World.*) write about how from the 1400s to the 1600s, individuals' names were transposed from spoken to written language and coded into registers – a process of bureaucratisation, "the rise of the administrative state. . . with its civil servants, its archives, and its splendid, petty-minded, and pitiless systems of notation is perhaps the master narrative of social history in the past few decades. . . allied with a narrative of origin and ascent that assigns to this same epoch the 'genesis' of modern individuality and subjectivity". (See Martin 1997 for more literature. An interesting modern take on this is the role of I.B.M.'s counting machines as apparatus of the Nazi State.) He also fixes on the notion of the 'false messenger'; reading this alongside Singh, it is clear how important messengers and couriers were, 'the bringers of information from afar', now replaced with the World Wide Web.

All three themes are relevant to the way in which Open Crime Data mediated by the World Wide Web identifies current statistical crime "themes", and helps us consider which are missing from public view. There is some evidence that rationalisation of the publication of data, with the concerns about anonymisation, and the non-identification of victims might make using such data difficult or miscast. This is interesting considering that one of the meta-rationales for publication - transparency - appears to be as part of a push to decentralise government, and allegedly hand over more power to the people when it comes to decision-making on crime. Certainly the mechanisms for governmentality: surveillance and sousveillance can help us to better understand some of the ontologies of crime and crime-data; through the publication of crime statistics, we see some evidence of "what sort of neighbourhood" we inhabit, it demarcates groups of crimes - a sort of bird's-eye view of the locales we already know, or those which we are considering inhabiting. At a simple level, rationalisation of the state has provided us with this data as its alleged by-product. More subtly we can see signs of Foucault's bio-identity coming into discussions on individual identity as governed by the state with fingerprinting, D.N.A., and Lyon's body surveillance.

When it comes to the identification of the civil subject, concepts become more entangled. The rise of the Web has helped people to hold officials accountable by publishing data about them and crowdsourcing incentives to encourage accountability. If there is trust in "official" open crime data, then this moves the sousveilling focus off individuals employed by the public who should be held more accountable, to some extent and returns the surveillant gaze to a population of civil subjects, with perhaps a focus on "popular" or "signal" crimes – murders of pretty unnamed women by famous people, drug-taking exposed, crimes which are reported for insurance purposes. Trust, democracy and accountability become strange fruit here. If an official is to be held to account for the ways in which they help by some means or other, to police our society, accountability becomes complex, when people already argue over what constitutes crime, what makes for good policing, and the extent of our own responsibility for ourselves. Does the Web itself change power mandates in exceptional ways? And how are we to understand the surveillance inherent in statistics, not just as an expression of the power of the surveilling body over the surveyed, but as part of the mechanism of a capitalist democracy in which the media, as the Fourth Estate, apart from Lords Temporal, the Lords Spiritual and the House of Commons plays a huge part, and conducts its own surveillance, not only on the government, which it is partially mandated to do, but on us, as those who give up our data to the companies which fund the media's online presences?

Coming to the present, we see a pattern emerging: Lyon says: "Big Data intensifies certain surveillance trends associated with information technology and networks, and is thus implicated in fresh but fluid configurations. This is considered in three main ways: One, the capacities of Big Data (including metadata) intensify surveillance by expanding interconnected datasets and analytical tools. Existing dynamics of influence, risk-management, and control increase their speed and scope through new techniques, especially predictive analytics. Two, while Big Data appears to be about size, qualitative change in surveillance practices is also perceptible, accenting consequences. Important trends persist – the control motif, faith in technology, public-private synergies, and user-involvement – but the future-orientation increasingly severs surveillance from history and memory and the quest for pattern-discovery is used to justify unprecedented access to data. Three, the ethical turn becomes more urgent as a mode of critique. Modernity's predilection for certain definitions of privacy betrays the subjects of surveillance who, so far from conforming to the abstract, disembodied image of both computing and legal practices, are engaged and embodied users-in-relation whose activities both fuel and foreclose surveillance." (Lyon, 2014). Predictive analytics are the result of the panoptic regime and the desire for the observation of disciplined bodies allows technology to drive accountability and thus distort its practices. Data can be bad data but without unpicking the ideology embedded in these structures, and particularly its ethics, we create constructs that are both fantasies and defining mechanisms for society.

### 2.2.3 Policing and Maps

Our final section discusses a few concepts germane to the understanding of mapped crime data. We have looked at the concepts of the systems of open and the production of knowledge about crime, in 2.2.1. In 2.2.2 we pulled out some key surveillance concepts. We now look at maps and the rise of criminology, as well as an overview of police use of mapped crime data. In 1826, Charles Dupin produced the first chloropleth map, showing the distribution of illiteracy in France. This started a relationship between maps and objects of policy: health, education, which was then picked up in 1829 by Adriano Balbi and Andre Michel Guerry who started looking explicitly at relationships between crime and various other phenomena such as population density and educational levels, which Guerry then published under the title of "Essai sur la statistique morale de la France" in 1833. (Guerry and Silvestre, 1833). Guerry in fact created a "machine" or "ordonnateur" that served to work as a "special-purpose device to summarize the relationship between a given moral variable or rate of crime and other factors that might provide some explanation or description." (Friendly and de Sainte Agathe, 2012). Working at the same time as Quetelet (Quetelet, A, 1866), we see an interest in "moral statistics" emerging, and with it, along with Durkheim's work on the statistics of suicide, the birth of both sociology and criminology. We also see the birth of "an extravagant system of metaphors and similes linking the social domain to the theories and even the mathematics of physics and astronomy" (Guerry and Silvestre, 1833). "Moral statistics" (see also (Rawson, 1839)) married crime statistics and geospatial demographies. It is clear though that the concept of risk was emerging, aligned with the use of "big" statistics, although its modern companion, fear of crime, was yet to emerge.

As we have seen, until the late 17th century the State's interest was in keeping order, often through control of the labouring classes, via regulation enforced by citizens acting as watchmen, constables, poor law officers and Justices of the Peace. Control through detection was carried out via victims and their families. Rewards opened the way for "thief-takers", which resulted in the opening of an office at Bow Street attached to Henry Fielding's Magistrate's Seat. Henry's half-brother, John Fielding, came to believe that by applying organisational principles, citizens' roles would be to supply information for officers to collate, interpret and act upon. Bentham and Colquhoun had preceded this with the marriage of panopticism and accountancy principles, which developed the idea of data-gathering into the concept of using total knowledge combined with uncertainty to provide information asymmetry in order to control a population with that action leveraged through payment via results.

The emphasis was now on preventing victimisation, citizens' responsibility to protect themselves, advice on crime prevention was given, and there was a demand for locks and lock-makers and the swing between crime assurance versus fear of crime, used to sell services. It was upon this basis that in 1829 the "New" Police were formed, through the Police Act, responsible for preventing crime, maintaining order, and for arresting people. We see how preventative policing emerged along-side an interest in data, the state and the behaviour of populations, This saw Peel solidifying the object of the New Metropolitan Police to "the prevention of crime". This is another early (but by no means the earliest) example of the confluence of surveillance, "ingenious machinery", preventative policing, accounts and accountability, and broad data. This confluence was part of a change in thinking that saw the "fad for number, weight and measure," (Porter, 1995), used by state and science in shifting ways. The state needed to know its citizens in order to govern them. Preventative policing was part of the means by which this was to occur in practice. Over time, the idea of a unitary policing scheme was introduced, leading to concerns about increasing the power of central Government and concerns about watch forces, brought to the fore by the Gordon riots in 1780. The New Police were seen by some as being in part responsible for the success in the two world wars, (Reith, 1938), or, as suggested by revisionist accounts, responsible for the controlling an insolent and burgeoning working class. (Storch and Engels, 2008).

These Police were responsible for preventing crime, maintaining order, in particular on public highways, but also for arresting people. They wore uniforms so as to avoid accusations of being spies. Mill and his successors originated policing as we know it, while the statistician's grand dream appeared to loom closer: "Of course the census is just a snap-shot, a once in a while picture of how things are, what we'd really like would be to have this picture all the time - maybe if we had a national identity card that might help?" (Anonymous Census Statistician at O.N.S.). Modern U.K. policing arose directly from the impact of Bentham's thinking. As discussed in 2.2.2, while Bentham is known for the idea that germinated the concept of Big Brother, this was part of how he perceived a greater system of distrust that used data to monitor those in power, and to create policy: instrumental in some of the first uses of "Big Data" with regard to crime, and this data not just used to survey populations but also governments.

Mill understood the power of bureaucracy in administering democracy, while Brandeis brought in regulation that would help to ease the mechanisms of democracy, thus allowing people to step aside from complex decision-making and letting markets, or mechanisms speak. We also see the emerging idea of the power of open versus closed or secret information, and the importance of understanding how information is transmitted and moves from secret to open through markets created by such systems. What happens for example when using a mechanism such as Open Crime Data, we let it freely represent crime via maps purporting to represent the viewer's local area, and thus produce a "mechanical contrivance" that measures a government's actions in preventing crime? What the Web has done is to shift relationships between some of the concepts mentioned above so that it is no longer as clear whether we can have security without privacy or accountability without transparency. It can be seen that changing relationships between the law, the individual and society have helped to mediate the rise of both the crime official and the official recordings and publication of crime data.

Modern policing in the U.K., and elsewhere, itself is a complex subject to analyse with debates over the nature of policing itself, its governance, its structures, polic-ing core tasks, recruitment and training, multi-agency working, politics and justice, in addition to the complexity that new technology brings to the subject. Newburn defines policing core tasks as including: public reassurance, order maintenance, crime reduction, crime investigation, emergency service, peacekeeping, state security. (Newburn, 2003). He asks whether we have a police force or police service, and whether we still adhere to policing by consent, delivering care, or the U.S. model of police forces delivering control. (Foster and Bowling, 2002), (Johnston, 2003).

We need to understand structures and governance, as well as the role of programmes such as problem-oriented policing and intelligence-led policing in order to understand where the data we are analysing comes from, and under what pressures, and with what financial constraints. There is certainly a changing social context in policing. Bayley and Shearing said that, "modern developed economies have reached a watershed in the evolution of their systems of crime control and law enforcement." (Bayley, David and Shearing, Clifford, 1996, p.585). They saw one system of policing ending and another taking over, with "a broad range of private and community-based agencies that prevent crime, deter criminality, catch lawbreakers, investigate offenses and stop conflict." (Ibid. p.586). Policing has pluralised, and gone back to the idea of the civilian taking responsibility for crime, and there is a sense that the state monopoly has been eroded since the 1960s. There are a huge number of private security agencies, and if we consider cybercrime, it seems that crime here is almost exclusively fought non-transparently and in private with the government's cybersecurity agenda specifically mandating "U.K. P.L.C." to fight crime, because of limited budgets. The narrative is that citizen involvement has spread and become normalised so that police are no longer the primary crime-deterrent. I would argue that policing has always been pluralised, but that it is only now in the era of the Big Data hype that we can see how effective or dangerous such pluralisation can be, both for those upon whom its effects are felt and for those who practice it - it becomes more urgent to quantify this pluralisation.

Johnston and Shearing (Johnston, 2003), and (Bayley, David and Shearing, Clifford, 1996) discuss security networks and the commoditisation of security and policing. Security refers to a "whole range of technologies and practices provided not only by public bodies such as the police or local authorities but also by commercial concerns competing in the marketplace. We have unfolding in Britain an uneven patchwork of security hardware and services, provision increasingly determined by people's willingness and ability to pay." (Loader, 1997, p.147). We see that policing networks themselves can be uneven and ad hoc and weakly organised – developed through personal relations. This is not to say however that such manifestations are wrong - unevenness, ad hocness and loose organisations with conflicting policy are seen everywhere - rather than try to rationalise by making monsters of such manifestations it is better to understand what the whole range of technologies and practices are attempting to do, whether algorithmically or not, and to understand what prevents this and what supports it. If the markets that offer these services are confused enough, provision might not always be in people's best interests, whether they are rich or poor.

There are perhaps two crime perspectives that help us to explore the impact of using mediative technology to understand the causes of crime and prevention of crime (and whether or not these can be mapped) – one perspective is loosely geared around the notion of the victim, as espoused by victim studies, victimology or victimization (all of which differ). Victim studies, victimology or victimisation theories look at the triad of a potential victim, their environment and the potential offender with crime-as-event being the relationship between these nodes. In many case the “solution” is seen as being information. (Karmen, 2012).

The other is social control. Criminologically, social control is geared towards understanding how communities work and that the morals and values of a community are shaped by social norms that are informally communicated and enforced and which then influence criminal acts. Understanding of communities and their norms then could help to predict criminal behaviour, if we have understood it well enough. Some see there being a tension between these two perspectives, where for example any interest in the rights of the criminal, or reasons for examining causes for their actions detract somehow from a victim’s experience, where victim and criminal are in a zero-sum game. (Garland, 2001). Some authors suggest that criminal acts themselves are forms of social control with much crime being "moralistic and [involving] the pursuit of justice. It is a form of conflict management, possibly a form of punishment, even capital punishment. Viewed in relation to law, it is self-help. To the degree that it defines or responds to the conduct of someone else – the victim – as deviant, crime is social control.” (*Black’s Theory of Law and Social Control - Criminology - Oxford Bibliographies*).

What we are going to call the "Crime Social Machines" that the Home Office data sits within can be found loosely interweaving between these two perspectives; there are plenty of ways in which apps are seen to help potential victims by arming them with information, which in itself is provided by other victims who are then also feeding the machine for production of crime data and knowledge of crime. Where communities provide or monitor information then this seems geared more to the notion of social control, while it is possible that there is even a space for potential criminal acts to be prevented if people with grievances were given a place to air their resentments.

It is with this policing background (at least in the U.K.) that we can then see how researchers started examining causes of crime, using large data masses. In this context, maps seemed to accompany the birth of criminology and provide an impetus for many deep criminological questions to be asked, although it is only recently that an interest in the subjective experience of the victim in the C.J.S. has come to the fore. There is no simple causal account that shows us that first we had statistics, then maps, and policing and policy. Similarly it would be too simplistic to say that we can analyse the knowledge provided by crime data, dive into the numbers and discuss surveillance, and neatly wrap up this discussion by saying that synthesising the two leads to understanding, as provided by maps and local knowledge. So while we discuss the mapped crime data mediated by the World Wide Web in this part of our literature review, we do so with the caveat that this is another lens by which to examine the phenomena under inspection.

### Maps, Accidents, Calamities and Casualties

“People come to know the world the way they come to map it—through their perceptions of how its elements are connected and of how they should move among them . . . by situating the map at the heart of cultural life and revealing its relationship to society, science, and religion ... It is trying to define a new set of relationships between maps and the physical world that involve more than geometric correspondence.” (Rothstein, 2013).

This section examines where the confluence of statistics and prediction, big data and surveillance, power and policing and maps and explanations have produced notable effects and features that invite further analysis. The previous sections referred to the how and when. In illuminating this, maps have played their part, serving to create models of salience and to help us picture the where. Having examined some of the background to, and present day concerns around knowledge, in the form of the transparency agenda and Open Data, and number in the form of statistics and the state, we turn now to understanding something of the producers and consumers of open crime data. We look at where it comes from – who creates it, how and why, and whether or not it is representative of crime in the U.K..

Maps, for most purposes, are models. They date back to Neolithic times, 16,500 years ago, where the first recorded maps, believed to be of stars, found in Lascaux, France, indicate that humans were constructing knowledge of the universe via ordered representations of what they perceived, and the first perceptions were looking outwards and upwards. They extract salient features from a landscape, for example, and communicate these features using conventions and representations that may appear natural or unforced. Many things can be mapped: stars, oceans, landscapes, data about almost anything that appears in space or time, whether those times or spaces are real or imaginary. Times of journeys can be mapped, as can relationships between people and places. Some of these things that can be mapped are expressive of complex relationships between phenomena, complex relationships that can be expressed using graph theory. Graph theory can elegantly produce maps and one of its popular early problems is to do with the Konigsburg Bridge or the "travelling salesman problem". (Barabasi, Albert, 2002).

Maps that help us navigate are very familiar, and we often use them in conjunction with instrumentation that helps us to determine our speed, direction of travel, and other factors that help us to get where we are going safely and efficiently. Some of these navigation aids (maps plus instruments) predict our arrival time, plus obstacles and difficulties and potentially, risks. The idea of mapping crime seems intuitive, for people journeying through an unfamiliar place, we want not only to know how to get to our destination but how easily or safely we can. When considering whether to invest all our hard-earned money in living in a place, we would love to know how safe it is, how easy and idyllic our lives might be there. And obviously, those who are tasked with looking after a location would like to have an intuitive understanding, a real feeling for what is likely to happen where and when. In fact crime-mapping is a precursor of criminology, although it is hard to definitively pin down the first mapping of crime figures to their geographical representation.

It was John Ogilby who showed us how much “data can be displayed, even on a linear strip, of object and proximity.” (Mullen, 1996). Ogilby used a form of data compression, in that he made maps that focussed on a route and then represented each part of the route in relation to the next part, rather than representing the wider context. Ogilby produced something very similar to a crime map, as he also created “Queries” that provided standardisation to the data gathering for his mapping. He asked Hooke, Aubrey and Wren, of the Royal Society, to create questions to support county atlases. One question was on “accidents, calamities and casualties.” (Norgate and Norgate, 1998). These questions also are precursors for what are now known as Gazeteers.

It has been said that the cartography of medieval and later times, “shows little originality. It was in no way corrected or checked up with reference to astronomical observations. Most of the maps were based on earlier models, and it is perhaps possible to trace their origins back to maps of the Roman Empire. Cartographic

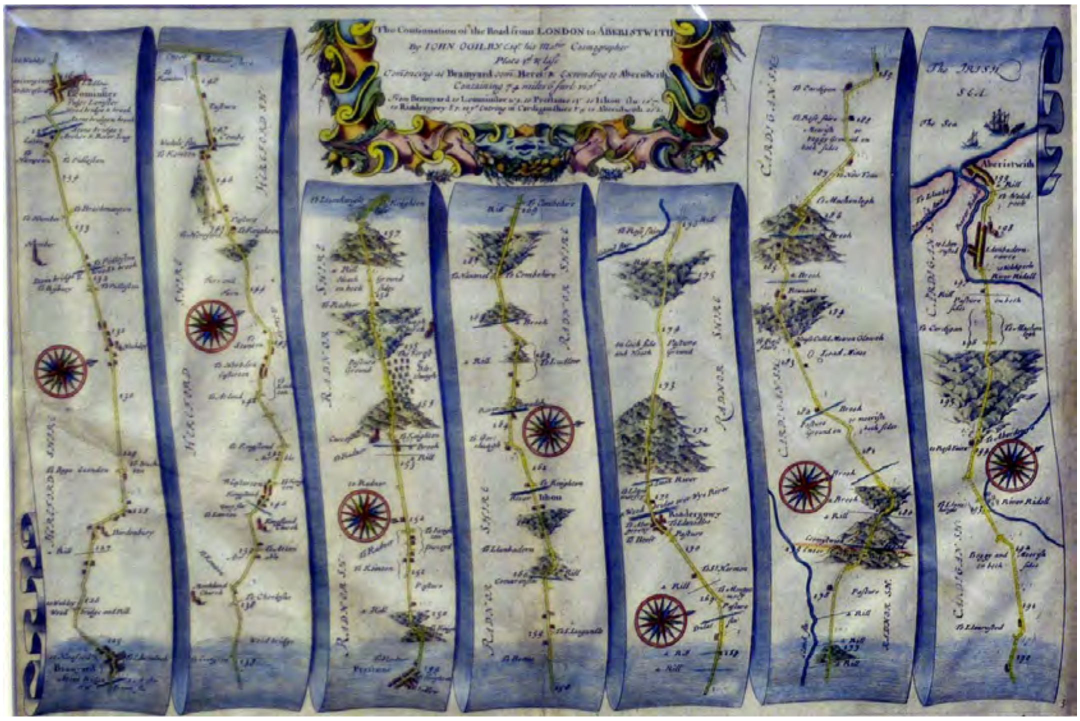

The Continuation of the Road from LONDON to ABERISTWITH
By JOHN OGILBY Esq. his Maties Cosmographer
Plate y^e 2^d & last
Commencing at Bramyard com. Heref. & Extending to Aberistwith
Aberistwith
The IRISH SEA

FIGURE 2.3: Use of Data Compression in Ogilby's Maps

accuracy was not the aim of the map maker of the time, and we are not justified in criticising his maps in the light of modern standards. They should be regarded rather as diagrammatic approximations. A number of conventions were followed, the most important of which was the representation of the east at the top. The maps were vividly colored; and mountains, rivers, and the works of man were shown by pictorial symbols." (Munro, 1925). So although we have a juxtaposition of maps and "accidents, calamities and casualties" we see an early scepticism about how truly representational any maps are, let alone those which demonstrate criminal activities. Crime statistics were also helped by Gregory King, who was Britain's first great demographer, and who worked for John Ogilby. He used taxation information to make reasonable assumptions and predictions about populations. (Finkelstein, 2000). In 1801, the first reliable modern census was undertaken, and from then on data was recorded in growing quantities about the population of Great Britain. (The Old Bailey, 2013).

### Causes of Crime and Moral Statistics

As stated in the previous section, Charles Dupin produced the first chloropleth map, showing the distribution of illiteracy in France, which then set the scene for scientific investigation into the causes of crime: early criminology. Bentham, Mill, Colquhoun and Lombroso among others, had explored these ideas, which we analysed from the viewpoint of statistics and surveillance. In 1861, the journalist Henry Mayhew explored the London slums. Mayhew's writing was rich enough to be considered ethnographical; he looked into people's lives in detail, but sympathetically, describing their clothes, their living places, the customs and habits, their numbers and income of those who had a trade. (Mayhew, 2009). Charles Booth was another noted philanthropist and businessman. Along with Beatrice Webb and Octavia Hill, he recognised that crime and vice were being sensationalised and that what was needed was a factual description of society, rather than fearmongering. He produced ground-breaking poverty maps of London in the 1890s, just as Brandeis was writing to his fiancé in the States about exposing the wickedness of people shielding wrongdoers, and of sunlight being the best disinfectant.

Following Guerry and Quetelet, criminology proper emerges, with Durkheim making an explicit decision to depart from philosophy and psychology, feeling that the "individual is the active agent of culture." Ensuing debates and particularly those engendered by the use of crime data and crime mapping do not make it clear whether or not culture might equally be the active agent of the individual, and whether the culture of "crime-solving" needs to be unpicked so that the individual once more reclaims some of their epistemological rights. There are complex conversations about causality, correlation and reductionism embedded in these debates. Durkheim himself wanted to take understandings of criminality away from the notion of causes rooted in individuals towards the socio-cultural, while developing a positivist (after Compte) methodology, using the concept of anomie, health and unhealthy/ pathologic societies subject to evolutionary influences. He understood crime as being not rooted in individual acts but as a projection arising from social norms. While he correctly (one might assume) saw that social forces can contribute to individual states of mind (as evidenced in his study of suicide) we must still explore the idea of this being a bi-directional relationship: individual states of mind contribute equally to social forces, particularly where it comes to crime. There is a sense of later ideas of markets here, where Durkheim saw the need for governmental interventions, rather than allowing for free-flowing evolutionary forces to take place unchecked.

Durkheim proposed pre-industrial "organic" societies consisting of non-linked nodes who had social norms or the collective conscience mechanically provided for them as externalised social facts that then became internalised. This appears to be an idea worth exploration in constrast to Foucault's worries about Panoptic discipline and self-censorship. Rather than binary relationships between the inner and the external, Durkheim appeared to give the subjective social experience of individuals an objective external independence - heavily influenced by psychology, with his positing of the social self and the egoistic self. The egoistic self has no constraints, and it is only the introduction of the industrial "mechanical" society that provides the interlinking of nodes that then provide some form of regulation. As individuals naturally work at what they are good at in a networked industrial society the division of labour becomes "the sole process which enables the necessities of social cohesion to be reconciled with the principle of individuation," (Durkheim, Eimile, 1964, p.185). This is an organic solidarity that arises naturally out of interlinked entities. A caveat is perhaps that while individuation might well arise from successful work or working at what one is good at, other aspects arise from working in things that one is not good at, or where the networked environment produces bad effects - very important for our analyses. According to Durkheim there was still a need, despite our functional interdependencies, for moral regulation as the lack of a moral force could produce an anomic society, with no moral constraints on an individual's limitless desires. Durkheim saw crime as necessary, and sometimes even as the product of innovators, and also as having a boundary function – the very presence of crime and knowledge of its occurrence could act to help people recognise distinctions between good and bad – two important points when it comes to crime and knowledge of crime produced via technology. Although Durkheim did not explicitly study crime, his thinking laid the way for the detailed use of statistics and mapping to study social problems.

### The Chicago School

From the 1920s and 1930s the Chicago School grew, with urban sociologists Park and Burgess at the University of Chicago using graduated area maps of crime. “The city is not, in other words, merely a physical mechanism and an Artificial construction. It is involved in the vital processes of the people who compose it; it is a product of nature and particularly of human nature.” They drew upon concepts such as plant ecology (Park, Robert and Burgess, Ernest and McKenzie, Roderick, 1925, p.2), and understood that the infrastructures which bring mobility and the concentration of populations contribute as much as geography and economics. “. . . think of the city, that is to say, the place and the people, with all the machinery and administrative devices that go with them, as organically related; a kind of psychophysical mechanism in and through which private and political interests find not merely a collective but a corporate expression.” Burgess showed (Burgess, 1967), that cities grow outward in concentric circles starting with an inner, business loop, a transition zone (with high rates of crime and social problems) and then home zones full of commuters. There are issues with this – not all cities grow in the same way, there are changes through gentrification, but Burgess captured the idea that crime exists in spaces that can be physically represented, and that these spaces are not random but as a result of social forces. (Paynich and Hill, 2011).

We can therefore understand that maps themselves will be awkward tools to represent crime, unless we bring to them our knowledge of the social forces at work behind these representations. As Simon Jenkins says in a Guardian article of February 2011, “I note that May and her officials censored more delicate information, such as of white-collar crime. Believe it or not, the most crime-free area of London is supposedly the banking district of the City. Between Bishopsgate, Old Broad Street, Cornhill and the Bank, May could find no crime at all, not one incident of theft, not a sniff of cocaine or an antisocial gesture. To the Home Office, theft and antisocial behaviour are what the poor do to the rich, not the rich to the poor. The map is seriously rightwing.” (Jenkins, 2011). It was the use of maps that caused some deep questions about crime and causes of crime to be asked. In 1938 Shaw and McKay wrote about the “causes of crime” (Shaw, 1930), and using the term “social disorganisation” went into the "zone of transition" and determined that high rates of turnover in residents, heterogenous populations and high poverty levels were present where there were high rates of juvenile delinquency. They were careful not to ascribe these factors as causal, but as correlative.

In “Principles of Criminology”, Sutherland made and developed the idea of “differential association”, which is that that organisation is different, not lacking, or pathologised, and that some of what other parts of society would term “delinquents” are learning techniques that are positive within their peer networks. (Sutherland, Cressey, and Luckenbill, 1992). He focused on social learning processes involved in the diffusion of criminal values, with lessons still taken from his theory today – both social disorganisation and differential association are principles that resonate through the examination of relationships between physical space and networks of people in those spaces. Important too are the theoretical assumptions made here, symbolic interactionism being the method by which a proportion of the researchers sought to understand the meanings of the multiple and complex worlds that they were examining.

More recent work has developed Sutherland’s thinking and examined the concept of collective efficacy, Sampson and Groves, (Sampson and Groves, 1989), looked at some of the problems with Shaw and McKay’s ideas. There are criticisms such

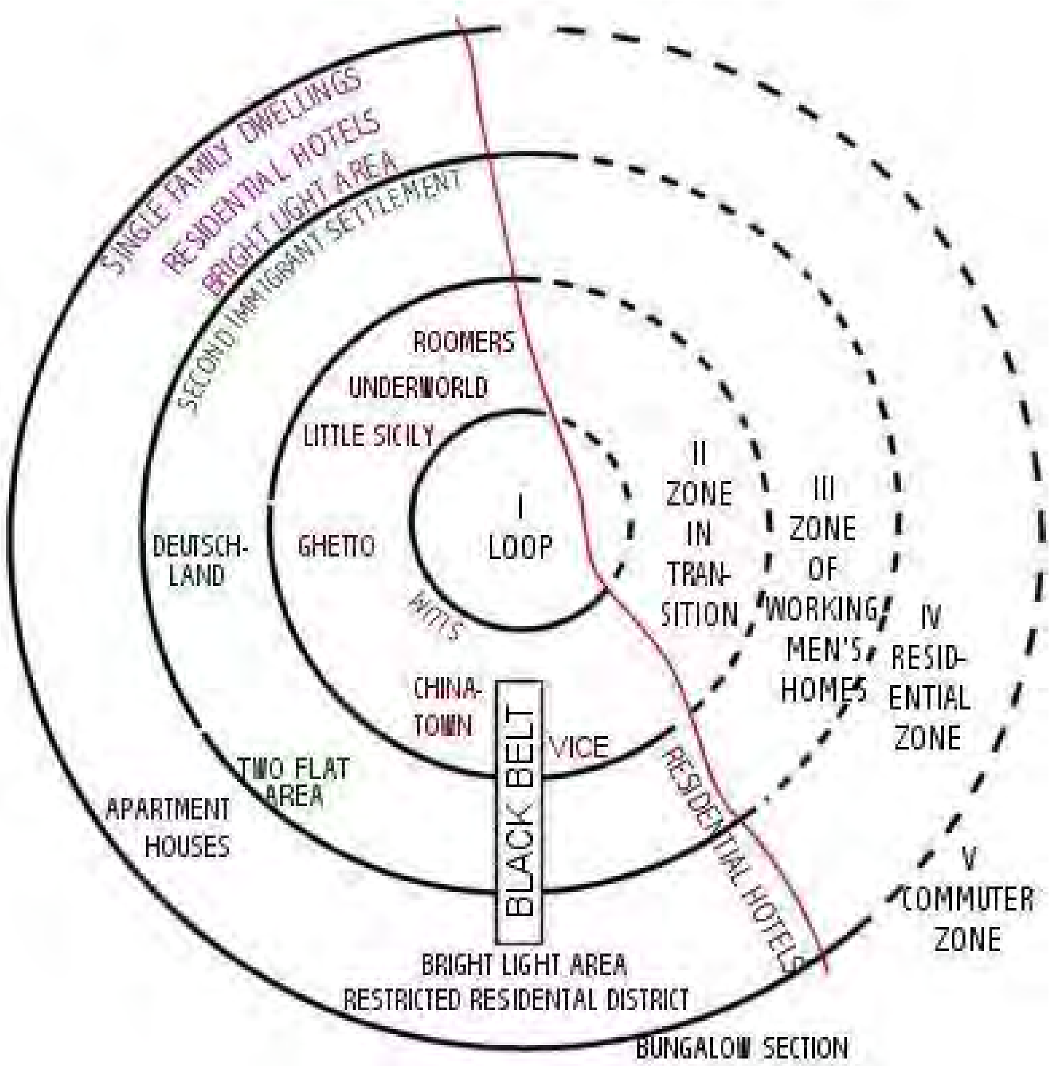


FIGURE 2.4: Burgess' Concentric Zone Model

.

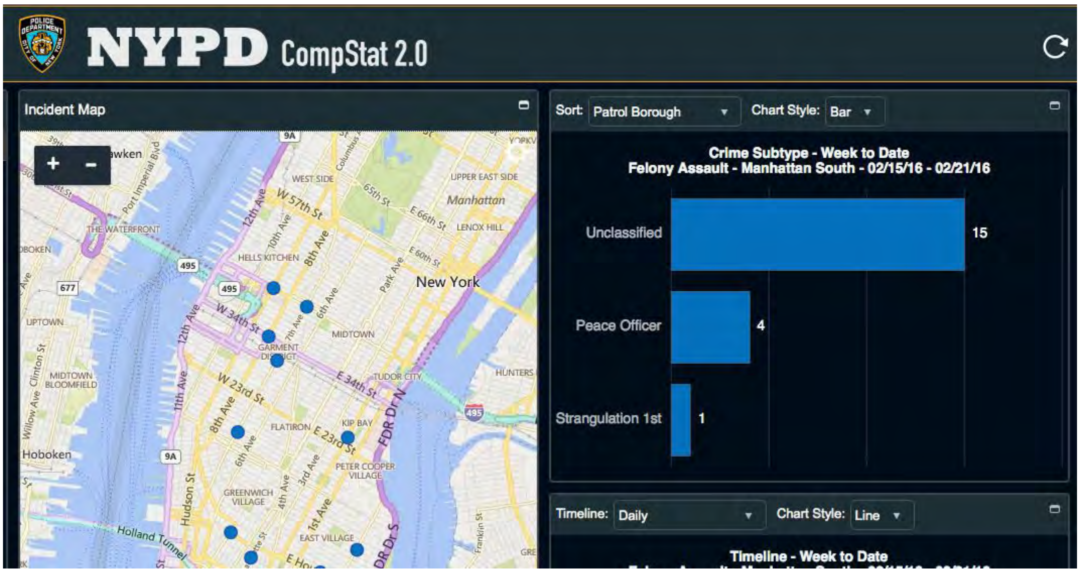
as lack of data to sufficiently test the relationship between community structure and crime, and conflict theorists would suggest that police concentrations might be higher in some areas and therefore discover more crime, that police are more likely to take action in some areas rather than others, that populations might be more concentrated and therefore distributions are not unpicked sufficiently, that ethnographic studies are only conducted within communities rather than between communities, and that even self-report data to offset official data again do not map out between-community activities. They found that there was reduced trust in these neighbourhoods, such that programs to reduce crime would be unable to work without doing something about social efficacy. All of these issues are more-or-less hidden in physical representations of crime shown on crime maps.

An experiment carried out by Zimbardo, (Zimbardo, 1969), then led to the idea of signals such as broken windows, (Kelling and Wilson, 1982), that transmit a lack of social efficacy, indicating that physical characteristics relating to an area are linked to crime, as much as the looser society. The Newark experiment relating to this suggested that while crime figures might not have gone down in areas where police reverted to foot patrol a sense of order-maintenance was created – and fear of crime went down. With fewer police numbers there is a common sense approach that suggests deployment consists of finding neighbourhoods at the tipping point. (Gladwell, 1996). This then pointed the way to Giuliani's zero tolerance policy in the early 1990s. Giuliani referred to quality of life, and asked that areas of social disorder were heavily policed, in "order maintenance policing", in a policy that left analysts divided over whether it transformed policing and helped the public, or was over-zealous.

### Crime Maps and Intelligence

Giuliani turned to Bratton who had been successful in reclaiming the subways when he took over the New York City Transit Police Department, owing to the work of Jack Maple. Some analysts describe Maple as simply having used crime data and put it on maps. In fact he understood the concept of intelligence in a way that still eludes many of those who write about crime mapping. "He [Maple] told anyone who would listen that until the entire police force got out of its rut – until police officers got out of their patrol cars and started fighting crime instead of responding to 911 calls – until that happened, the crime rate would keep climbing. Maple started mapping strategies to fight crime, and papered his walls with fifty-five feet of hand-drawn maps he called Charts of the Future. The charts detailed every stop on every subway line and every robbery that had been committed. The idea was obvious but untested: go after the bad guys where the bad guys did their work and get them before they committed more crimes." (Punch, 2007, p.14).

In fact Maple was not primarily interested in just "going after the bad guys where the bad guys did their work", but more in trying to find ways of understanding "what was going on", and then of capturing that understanding geo-spatially. As he said, the beauty of the mapping is that it poses the question, "'Why?' What are the underlying causes of why there is a certain cluster of crime in a particular place? Is there a shopping center here? Is that why we have a lot of pickpockets and robberies? Is there a school here? Is that why we have a problem at three o'clock? Is there an abandoned house nearby? Is that why there is crack-dealing on the corner?"(Monmonier, 2002, p.122).



NYPD CompStat 2.0

Incident Map

Sort: Patrol Borough

Chart Style: Bar

Crime Subtype - Week to Date
Felony Assault - Manhattan South - 02/15/16 - 02/21/16

Timeline: Daily

Chart Style: Line

Timeline - Week to Date

FIGURE 2.5: An Example of the Compstat Interface

Whether or not the nuances of understanding life in some of the darker places and usefully capturing that intelligence were fully understood by Maples' commentators, a tool was developed called "Compstat", which had been originally a means of analysing crime data and providing up to date information on crime occurrences in terms of geography but became the means by which precinct commanders were called to account, through strategic control. "At Compstat meetings at headquarters, precinct commanders would be confronted with data on crime patterns using 'digital mapping', questioned on their policies with regard to 'hot spots' and put under pressure to perform and to 'hit the numbers,'" (Loveday and Reid, 2003). This was a form of transparency working overtime, as the meetings were open to anyone to attend, "the meetings were open to invited spectators, from 'Australia to Zimbabwe', and there could be as many as 140 people in the room (Peters and Woolley, 2007). Politicians, policy makers, judges, mayors, academics and police chiefs from many countries from around the world witnessed the 'miracle' at work." (Eterno and Silverman, 2012), (Punch, 2007). So began a form of policing that was very much concerned with the production of maps and accountability for the numbers on the maps. While Maple had been concerned with understanding why crime occurred, and using maps to model and shape his understanding, in the hands of the wrong people, the maps became performance management tools. They had little to do with understanding causes of crime and everything to do with shaming and humiliating those who were called upon to help prevent or manage crime - the causes of which were out of their control.

### Cybercrime

Maps as tools for modelling crime become infinitely more complex when we bring cybercrime into the picture. There is a major and very current policing concern with cybercrime. This has an impact on the structure that mediates the publication of crime data and the fact that mediation is done via crime maps. At a meta level we must consider that at the same time as the web facilitates the production of this data, it is also enabling the evolution of the very crime it purports to tells us about, both in terms of character and number. In fact, there has been a huge increase in web-facilitated crime, where location, or mapping becomes at first sight, irrelevant to many commentators. As Wall says, in Cybercrime, "the relationship between crime and technology is by no means new and. . . the potential for creating harm never seems to be far away from any apparently beneficial technological development. Wall explains how technology can be used to exploit chains of trust. (Wall, 2007).

In the 19th century there were wire frauds perpetrated by tapping into the early telegraph system, and this uneasy relationship between crime and technology also extends to ideas about crime prevention and security. An analysis of these wire frauds shows uncanny similarities between some quite complex modern cybercrime Modus Operandi, or ways of working. While this is well-known - our later exploration of types of technology used to address crime unpicks some of these relationships – we ask, where do the police step in and where do they make way for other agencies to address crime? In this new territory, the Web, it is not clear when crimes are petty hacks or acts of war. There is a sometimes contested territory between policing and Defence; when Defence steps in, then the public's knowledge of crime evaporates.

We see what Jones and Newburn have called pluralised policing services, (Newburn, 2003; Johnston, 2003), whose function or remit becomes hard to quantify. Keith Bristow, former head of the N.C.A. has said that "terrorists and organised criminals often operate in the same way, and that, 'the tactics of law enforcement to tackle these people are often the same'. . . It would be sensible to look at shared capabilities with Scotland Yard to tackle the twin threats." In fact, looking at a few of the data sources that come into play in identifying crime, and then the agencies who own and control these data streams, we see that the police themselves are a very small part of the picture when it comes to understanding crime in the U.K.. Where crimes are committed with the intention of destabilising markets, then these web-mediated crimes exist as a form of knowledge. Banks and governments hide market-based crimes, as knowledge of fraud encourages people to lose trust in institutions, which loss causes as much or more harm as the original act - an epistemic crime. At present cyber-mediated or cyber-enabled crime is being shaped as crime by the government and the media, and not the public. As more crime uses technology, then our – as a public - knowledge of crime in its traditional form disappears. If, as has been suggested, fights for budgetary control mean that the area between defence / fighting cybercrime and policed non cybercrime becomes contested, then a move on the part of those who police cyberspace to claim more territory means any crime committed with technology involved, could become defined as cybercrime, and there will be very few crimes left that can be reported as open data. These contested spaces need analysing using the modern methods of Big Data and criminology in order to really understand how such tensions might be resolved.

## 2.3 Exploring the Impact of Open Crime Data: Conclusions

The data the Home Office publishes via Police.uk is open. Open Data comes partially from a tradition of transparency, but also other ideological thoughtstreams. Transparency itself has shifted over the ages from ideological transparency that facilitates democracy (Mill's) to financial and other sector performance data, where finances, performance, activity, government, oversight and regulation are enmeshed, and information on some aspects of these is pushed-out to the people. We have transparency under F.O.I.A. where info is requested, (pull-data) and its release is considered in the context of threat and security.

The literature suggests Fong's third generation transparency – crowdsourced accountability, has now been dramatically superseded by a new sort of transparency through the confluence of transparency, the "Open" movement and the World Wide Web as a mediative element. This has produced fourth and fifth generation ideological versions of transparency. Home Office crime data, on its own, without the representations or informational aspects provided by mapping, is certainly technologically mediated, is pushed out, and is provided via ideological transparency, facilitating democracy. Further analyses should ask how we know if open crime data produced in the name of transparency is "working"? What does success actually look like for Open Crime Data? We must decide on metrics for success and if these metrics show that there are problems with the current systems then we must also decide whether this is because we have competing ideologies, whether Open Data and transparency together mediated by the World Wide Web have created what Merton called "unintended consequences". Are these unintended consequences a result of the sort of data that is published? Do they arise because it is crime data, not education or health data? Returning again to the question of how effectively these actors discussed above - our governments, people, processes and technology - can address crime, initial conclusions are borne in mind for our methodology.

### 2.3.1 Technology

The literature asks, starting with Aristotle, Bentham, Mill, working through Brandeis, Coolidge, the legislation of transparency and F.O.I.A., how we construct systems that contain rules allowing for virtue to arise from practice. We have considered systems of distrust that allow all actors to survey one another. The web has radically changed the evolution of transparency, creating a new form that gathers and releases data on a hitherto unimaginable scale. It seems therefore that Police.uk could be a technology that allows for the "widest participation in the details of judicial and administrative business. . . above all by the utmost possible publicity and liberty of discussion. " (John Stuart Mill, 1886).

### 2.3.2 Policing and Crime

The literature on policing also shows us a concern with how to create incentives in reporting and fighting crime. We must consider how to use incentives to encour-age crime reporting via the web, and where they sit in relations to power mandates flowing between the ordinary population and 'Officials', to fight crime. We must also consider the use of incentives to encourage people to use technology to fight crime, perhaps by making crime apps with the data that sits with Police.uk. We must ask whether the apparent ease that the web provides us with in reporting and learning about crime actually creates moral or legal issues. How do we create an app, for example that might allow people to report concerns about individuals with-out breaching data protection legislation, promoting a culture of fear, or creating a snitch culture, such as the Orwellian reductio ad absurdum Www.snapscouts.org ?

With the advent of the web, are citizens' roles "to supply information for officers to collate, interpret and act upon," being somewhat taken for granted, without enough consideration being given to these legal and moral issues? Where exactly does technology sit within the "broad range of private and community-based agencies that prevent crime, deter criminality, catch law-breakers, investigate offences and stop conflict?" (Bayley, David and Shearing, Clifford, 1996). If policing has pluralised and now works alongside private security agencies and vigilante citizens, how can we be sure that the use of technology, such as the web, is safe with no central oversight? How do commercial interests work in this case? Is Police.uk in fact assumed to be a "special-purpose device to summarise the relationship between a given moral variable or rate of crime and other factors that might provide some explanation or description"? (Guerry and Silvestre, 1833). Do we take some of the conclusions from the Chicago School and say that crime doesn't just exist in spaces that can be physically represented, but in social and economic spaces and that these spaces are not random but as a result of social forces? Or do we say that what we are seeing on Police.uk, is not so much crime, as knowledge of policing? If reported crimes go up, is that an indicator of the success of Police.uk amid its transparency context – indicating growing trust in policing? If citizens feel more knowledgeable about crime and more empowered to report it, are not greater recorded crime numbers indicative of a healing in the trust between communities and the police, that Lea and Young wrote about? More crime numbers on Police.uk does not equate to more crime in itself, just better reporting and recording mechanisms – so how do we find out about crime itself?

How can transparency be effectively delivered when it is balanced against considerations of national security, the needs of law enforcement, governmental privilege, and the protection of personal privacy and business confidentiality – all of which also affect the welfare of our citizens? Is there a necessary tension between what makes much open government data of value, that it is often the personal that adds meaning and therefore creates value, and what makes it useable – that it cannot be released in any form that allows people to be identifiable? What are the risks here? How do we understand the risks of anonymisation in the context of fast-moving changes in technology, and data in use on a scale that is "big data"? In terms of mapping, is all geographically presented crime data doing the same thing? For example, does Police.uk do the same thing as www.ukcrimestats.com? Do some sites that look the same, in fact do different things, with differing attitudes to fear of crime and recorded crime numbers? How can we assess this? How do we check that transparency about performance does not in itself destructively affect markets? Should data in fact even be treated as a resource or commodity in the same way as oil or gold? (Nicolas, 2012). How exactly is the data on Police.uk being used to hold people to account?

# Chapter 3

# Methodology

## 3.1 Background to the Research Approach

While the literature review shows us which questions to ask about the impact of open crime data in the U.K. on its producers and consumers, through the lenses of knowledge, number and society, this chapter focuses on which are the best methodologies to do this, given our findings in the previous chapter. As stated earlier, it is hard to measure the impact of crime data and other data released via the World Wide Web in the name of transparency. The more democratic a society, the more impact gets enmeshed in bureaucratic workings. In the first place we have to consider what we mean by impact, then how we measure it, and finally how we illustrate the results. Our methodological approach builds from a first look at the metrics of www.Police.uk: an approach that allows us both to consider the nature of the data and also how best to approach its analysis and context in terms of a methodological and philosophical paradigm.

"...there are several different ways how one can look at the relationship between methodology and philosophical paradigms. Proponents of (strong) paradigmatic view so called purists propose that there are different mutually exclusive epistemological positions or paradigms and that quantitative and qualitative research methodologies are tightly bound with them. Pragmatists on the other hand do not agree that this connection between paradigms and methodology, but also between paradigms themselves, exists in such clear-cut fashion; they rather accept that there is a mutual influence and that the integration of different standpoints may give the best results in many circumstances. In between of these two extreme views one can find several others as well." (*Paradigms and methodology in educational research*).

Given that our remit was to examine the impact of publication of open crime data, this covered the subject areas of technology and policy, which, if understanding and anaylsis are to be effective, bring into play at the very least political, economic, legal, socialand organisational factors. We had already looked at the difficulties of trying to use only quantitative methods and found that although they might be effective, as very simply, crime data is commonly represented with numbers, they did not account for the provenance of those numbers. We therefore looked at complementary qualitative methodologies such as:

- Participatory practice: views the participants as active researchers, and places a high value on their knowledge and experience, thus providing local reality.
- Phenomenology: examining the lived experience of people in relation to the phenomena under investigation.
- Ethnography: systematic exploration of the social world, culture, shared beliefs, perspectives and practices from the point of view, or close to the point of view of the subject of the study.

- Ethnomethodology: examines through accounts, how people use, for example, dialogue and body language to construct socialfacts at micro level - the person's own view; or macro level, at the organisational view.
- Grounded theory: assumes a blank slate and uses an inductive approach to develop new theory.
- Frame analysis: looks for socially shared organising principles that can shape discourse and influence public opinion.

While these methodological examples underpin philosophical approaches, the methods they invoke are varied and not limited to each methodology - as well as having potential to be aligned with more quantitative methods. Some mwthods might be: contextual enquiry, living alongside participants, interviews asking participants about their recent experiences, or surveys asking lots of people about their views. A mixture of these qualitative methods, such as semi-structured interviews, desk-based studies of organisational doctrine, case studies and focus groups can bring out different nuances in terms of cultural and organisational considerations, depending on how they are combined.

The research was split into these phases:

- Analysis of mapped crime data applications and web-sites using grounded theory.
- Analysis of discussion on social media links to crime data producing interpretive frames and conceptual models of crime, using frame analysis.
- Interviews with experts or stakeholders producing contrasting frames.

The following provides an overview of how we developed this overall methodology, however we make a brief note here about the methods we did not use. We did not use web-survey and case-study as discussed above, feeling that they were neither representaive, nor providing of understanding. Discourse analysis has many different varieties that can be concerned with lexical choice, word formation processes, conversational analysis, sociolinguistics, pragmatics and text analysis for example. These varieties are mediated by the concept of linguistic units of analysis - ie : structure, meaning, interaction, and social behaviour. Although these are all highly interesting, we were looking for links between the technical production of data and the policy framework that it was presented in, so while these linguistic units were of interest, we decided that frame analysis would answer the questions about policy and data production and interpretation more pragmatically than these other methods (although they all come into play).

## 3.2 Prior Work Analysing Online Democracy

We have formalised our research questions through two lenses, the top-down one of understanding open data in a democratic and governmental context, through knowledge, and the bottom-up, data-driven one of understanding the technical, automated aspects of data as it leaves the World Wide Web.

The third aspect that we examined in our last research question was of balancing knowledge (knowledge of crime) and data (response to crime data publication) with understanding – the third part of Ackoff's triad. Understanding includes examining the philosophical aspects of data itself, as it passes through mediative elements such as the World Wide Web and socio-cultural ones such as the world of policing, law enforcement and the Criminal Justice System. We are looking at both macro (social) phenomena and micro (usually) sociotechnical phenomena, that relate to the systems through which data passes.

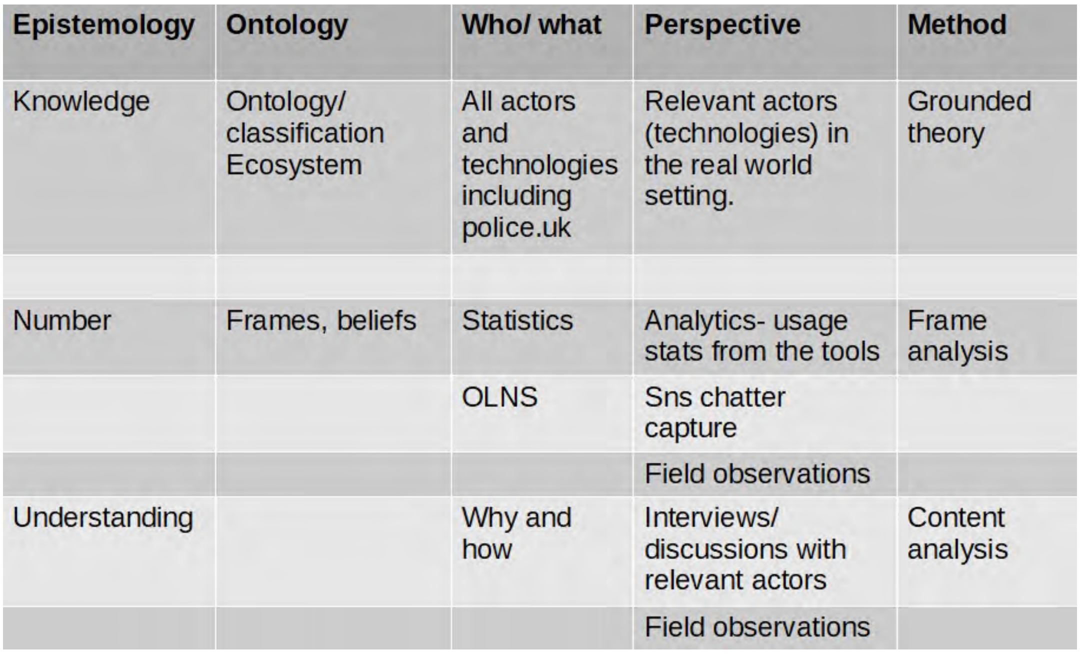
Having this in mind, we next examine prior methods that have dealt with similar issues. We are looking at data, mediated by the Web, in a policy context. Similar work has been done on the problem of analysing how democracy online works. Macintosh and Whyte, (Macintosh and Whyte, 2008), have proposed a framework for evaluating e-democracy and e-participation, that goes beyond the popular 'web survey', or 'web survey and case-studies', with often a narrow focus on only one stakeholder at a time: "...methods that analyse interaction ensures that the evaluation provides evidence of what people actually do with eParticipation tools, as well as what they say they do. Importantly, using mixed methods allows triangulation of methods and results and, therefore, helps to maximise their validity". Quoting an O.E.C.D. study from 2001, they look at guiding principles for evaluating off-line citizen engagement with democratic tools. "Guiding principle number 9 on 'evaluation' stated: 'Governments need the tools, information and capacity to evaluate their performance in providing information, conducting consultation and engaging citizens, in order to adapt to new requirements and changing conditions for policy-making.' (OECD, 2001, p.15). They refer to Gil-Garcia. and Pardo (Gil-Garcia and Pardo, 2006) who had previously discussed the multi-method approach to eGovernment research. Gil-Garcia. and Pardo argue that "eGovernment is a complex social phenomenon that can greatly benefit from the use of multiple disciplines." "…A multi-method approach to eParticipation evaluation is even stronger. Based on previous work (Whyte and Macintosh, 2003) and (Whyte and Macintosh, 2002), the authors argue that to evaluate how effective eParticipation is in engaging a wide audience so as to inform and influence the policy process, an analytical framework has to be developed that takes into account three dimensions: the evaluation criteria; the analysis methods available; and the actors involved. The evaluation criteria, as illustrated in Figure 1, consider three overlapping perspectives: democratic, project and socio-technical.

- The democratic perspective considers the overarching democratic criteria that the eParticipation initiative is addressing. Here one of the most difficult aspects to understand is to what extent the eParticipation affects policy.
- The project perspective looks in detail at the specific aims and objectives of the eParticipation initiative as set by the project stakeholders.
- The socio-technical perspective considers to what extent the design of the I.C.T.s directly affects the outcomes.Established frameworks from the software engineering and information systems fields can be used to assess issues such as usability and accessibility.

These perspectives can then be broken down into:

- Field observation of relevant actors using the tool in a real-world setting.
- Interviewing and group discussion with relevant actors.
- Analysis of online questions and discussion.
- Analysis of project documentation.
- Usage statistics from the tools and server logfile analysis.

Forss (Dfid, 2012, pp.58-65) recommends a similar mix of methods. Macintosh and Whyte also comment on the need to include a range of actors in the evaluation process. These might include:



| Epistemology | Ontology | Who/ what | Perspective | Method |
|---|---|---|---|---|
| Knowledge | Ontology/ classification Ecosystem | All actors and technologies including police.uk | Relevant actors (technologies) in the real world setting. | Grounded theory |
| | | | | |
| Number | Frames, beliefs | Statistics | Analytics- usage stats from the tools | Frame analysis |
| | | OLNS | Sns chatter capture | |
| | | | Field observations | |
| Understanding | | Why and how | Interviews/ discussions with relevant actors | Content analysis |
| | | | Field observations | |

FIGURE 3.1: Research Method following Macintosh and Whyte, 2008.

- Officials setting up and administrating the mechanism under evaluation.
- People who have used the mechanism.
- Those who have not used it.
- Elected representatives or officials considering the results.
- Other interested representatives.
- Project managers and technologists supplying the online tools.

“Those who have not used it” can be more precisely defined as users who are aware of the engagement but have not used it i.e. either were unable to use it or opted not to. Having examined this methodology, as applied to understanding the impact of technologies on e-democracy, and seeing its congruence with the results of the case study, we applied a modified approach to Open Crime Data, taking into account our tri-partite view of knowledge, number and society as providing differing but complementary and not epistemologically inconsistent ways of examining impact. Our final choices to make were to work out which were the right sorts of analysis to carry out within each piece of research. This chapter next offers a deeper look at two approaches, grounded theory and frame analysis, before outlin-ing the three final approaches. The figure 3.1 "Research Method following Macintosh and Whyte, 2008" shows the eventual overall approach, following Macintosh and Whyte.

## 3.3 Grounded Theory

We have started to formalise an approach which is to analyse a body of discourses relating to Police.uk. However, how are we to understand what Police.uk itself is, without putting it into context? We use Grounded Theory in order to understand the similarities and differences between Police.uk and other sorts of policing web-based technologies, data storage, sites or apps. We develop a series of metrics that can be applied to any group of sites or apps that appear to be similar in order to determine their probable success or failure and to understand more about what they really are. Although Grounded Theory has received much negative as well as positive attention in various social science spheres, it has been well-used in psychology research: (Pope-Davis and Liu, 2001; Pope-Davis et al., 2002; Richie et al., 1997). Grounded Theory accounts for the voices of participants in understanding their experiences in diverse settings in order to build a theory about phenomena. We use grounded theory in a more "voiceless" way to understand where a particular technology sits within a universe of technologies. Corbin and Strauss, (Corbin and Strauss, 1990, p.23), state that theory is "discovered, developed, and provisionally verified through systematic data collection and analysis of data pertaining to that phenomenon".

Although Grounded Theory emerged in the 1960s; it has firm links with Philosophy of Science, Wittgenstein's family resemblances, American Pragmatism, Symbolic interactionism, Kantianism, Mill's "system of differences," Baconian inductivism and Aristotelian axiology. In all of these we find an understanding that there is no such thing as "raw theory"; in grounded theory, the researcher is not expected to come to their classification as though to a tabula rasa; rather it is accepted that they know what they know, and that the act of classifying affects the researcher's perception of the things being classified. Grounded Theory has also had many criticisms levelled at it – Bryant and Charmaz in 2007, (Bryant and Charmaz, 2007), suggested that it does not conform with traditional conventions within academic research – the two chief being how it relates to traditional literature reviews, often needed in order to create research proposals and then how a theoretical framework is created. (Luckerhoff and Guillemette, 1990; Walls, Parahoo, and Fleming, 2010)). Grounded Theory presented a similar problem for this research in that it does not sit with the academic administrative structures needed to ensure a PhD progresses adequately, especially within the quantitative school that this mixed-method multidisciplinary research is sitting within. Traditionally a draft thesis is submitted and then the student is given a period of time to write up and complete the work, with the results already being known. Many approaches to grounded theory dictate that certain elements are generated after the first set of results, for example the literature review, which then can further affect results or produce a new set. It was decided that in this case, draft results could be represented, with the nominal period then being used to review the first set of literature and inputs, with the theory being generated at the end of the draft, and new empirical results being seen as on-going.

### 3.3.1 Substantive or Formal Theory

Traditionally, Grounded Theory produces either substantive or formal theory. Strauss and Corbin (Strauss and Corbin, 1998) explain how substantive theories are generated that relate to particular areas. "Since substantive theory is grounded in research on one particular substantive area (work, juvenile delinquency, medical education, mental health), it might be taken to apply only to that area." The thinking in 1967 was that it was desirable to generate a formal grounded theory from substantive theory. Formal grounded theory is seen as theory that encompasses more than one substantive area. "The latter not only provides a stimulus to a 'good' idea, but it also gives an initial direction in developing relevant categories and properties and in choosing possible modes of integrations." Formal theory is more abstract, and provides a focus on the construction of cultures or ideologies. (Bryant and Charmaz, 2007). This thesis generates a formal theory that looks at how concepts of crime are constructed; with a particular focus (using Frame Analysis) on how cybercrime lets us understand broader concepts of crime more generally. We do this in part by looking at the substantive area of crime data generally and then even more substantively cybercrime specifically. Although we would argue that it is not necessarily the case that a formal theory must encompass concepts generated by substantive theory, we use the constant comparative method of understanding how the differences between crime social machines can create concepts that enable us to more fully understand crime data mediated by the web generally and to advance its generation and use. This is a nuanced theoretical approach when it comes to examining web artefacts, as it allows the theorist to revise their findings as they move through the discovery process; improving on the logico-deductive approach.

Two points emerge here relating to research taking place on the Web: one is that the process of undertaking research on the responsive web, that offers us non-static entities, can affect the thing being researched, for example in examining Police.uk, talks to site designers inevitably had an effect on the site itself as they were consciously or unconsciously influenced by questions about design, intent, competition, data provenance and policy. Therefore a methodological approach is required which accepts some interplay between the observer and the thing observed, without this making the results of observation invalid. "Hence the reactive impact that investigators have upon their data bears more on the scope than on the credibility of an emerging theory. The technique that forces investigators to stay close to their data, and which constitutes the systemisation of the approach, is the constant comparative method." (Rennie, A, 1988).

The second is that of reproducibility. Mainstream science producing experimental results has long looked for reproducibility in those results. (Glenn and Bernstein, 1999). Some confusion arises over lack of reproducibility. (Baker, 2016). There is confusion both in reporting this issue and in isolating what a lack of reproducibility means. According to scientific method, since around the 1600s, lack of reproducibility can mean that data has been incorrectly gathered, or that the wrong data has been isolated for collection, that the processing of that data is flawed, that the theory (or hypothesis) that the data relates to is disproven, or that method is incorrect with respect to other issues than data collection, processing or hypothesis. A recent survey carried out by Nature refers to the "reproducibility crisis". In part this notion of crisis can arise from a lack of understanding of the philosophy of science and where different scientific methods apply.

This thesis looks at risk and statistics as they are dealt with in relation to crime and crime data, however, the same points about risk apply to the scientific method in general, i.e. we can apply the same thinking on reproducibility that we use on crime data to the methods that we use to examine this data. Science is itself a form of risk management with respect to the certainty of statements about the world. The scientific method is a means of evaluating the likelihood that such and such a statement might be right; what the risk is of it being wrong. A method in itself is a piece of metadata that with respect to the epistemological and ontological beliefs of the person, institution or system who uses that method tells you about the likelihood that a statement is true or false. Where we carry out web searches with results returned via Google, data returned may not be the same from moment to moment, depending on data centre locations, indexing, and the constant addition of new links to the web that then may alter search results. Grounded theory allows this; the focus is on creating a methodology that lets researchers apply it or alter it themselves, as it fits their needs - to build, not test theory. Such an approach is pragmatic and very much suits the fluidity of data results coming from the web.

### 3.3.2 Sampling and Representativeness

We now go into more detail with respect to our second piece of analysis, and the results from our case study. We found that there were issues with using just the numbers that the Home Office's Google Analytics generates to understand impact, and that Police.uk's site surveys told us very little. There are people visiting the site who do not wish to enter into "official" dialogue with the government about how it publishes information about crime data, or what that data means, but who comment on the data elsewhere, in socialnetworking groups with differing dynamics. Scientifically, we should be able to know that we have a sample of people in mind who have all visited the site, have not been selected in any obvious way, and who are truly representative of those visiting Police.uk. As earlier researchers have shown, (Holbrook, Krosnick, and Pfent, 2008), (Lewontin, 1995), we need to consider the problem of representativeness, especially where it comes to qualitative data – data that provides insight into the socio-cultural attitudes of those who visit Police.uk.

### 3.3.3 Big Data and Broad Data

Looking at how the issue of representativeness has been treated in related work in more detail allows us to unpick some of the potential ramifications of RQ2, "Can we combine Broad Data and Network Science methods with criminological theory to understand the effects of the supply of data from the Web?" There have been many arguments about whether qualitative research really "needs" sampling. Do we need to select samples of data-producing subjects who represent the population that is under examination? For example, if we are trying to understand the effects of the production and consumption of open crime data on a population, then we have to understand the population as something that exists apart from the phenomena under examination in the first place, if we are to have a controlled sample to compare against. This is what Mill and others since then have commented on. Then we have to know what characteristics in the population are the one that define it. For example if we want to examine the effects of Open Crime Data on people in the U.K., then if we are to take a representative sample of U.K. people to examine in depth, we have to know which characteristics to isolate – generally these might include: age, gender, income, health, socioeconomic status, politics. But what if Open Crime Data has some salient characteristics that might gradually affect the whole population but that are most noticeable in one part to start with? The data itself is a population with characteristics of its own. Those who have suggested that qualitative research does need sampling have had concerns dismissed in the past as mere "positivistic worry"; on the other hand, they say that qualitative research lets itself down through often using non-probabilistic methods. Some of the "most theoretically significant and important studies in field research (Gouldner, Dalton, Becker, Goffman, Garfinkel, Cicourel, Sudnow) were based on opportunistic samples." (Seale et al., 2006, p.405). However, we hope to show that using data more representatively can help us to uncover some unusual findings.

Even if well-known research has been done using non-sampling methods, this does not mean that the research could not have been equally as or more significant if it had been done using sampling methods. We can take global samples, and using techniques from "Big Data" for example, and Web Science methodologies, allay some positivistic worries, while carrying out field research. "... Defining sampling units clearly before choosing cases is essential in order to avoid messy and empirically shallow research... in contemporary organizational research the problem of representativeness is a constant and growing concern of many researchers. In addition, traditional qualitative researchers often forget that sampling is an unavoidable consideration because it is, first of all, an everyday life activity deeply rooted in thought, language and practice." (Seale et al., 2006, p.405).

Reasons for using representativeness are focussed around whether findings from research can be applied to the general population. Representativeness can be hard to achieve because of the "nature of social or cultural objects; and that as result qualitative research has pinned itself around two forms of generalisation, one of statistical logic and the second, coming from Glaser and Strauss of "theoretical sampling'". The authors point out that experiments are not based on statistical samples and that there are research practices in many disciplines where work is based on a few cases. They state that the core problem again is of representativeness – "the variance of the phenomenon under study". They say that in social research they look at the social significance of samples instead of statistical logic, and that while samples might not be representative, it does not follow that findings cannot be generalisable. Another (although ontologically rather than methodologically distinct) argument is that numbers themselves are a social construct, which seems to lead to the idealist or relativist conclusion that we therefore do not have always to always statistical methods in order to derive information from limited data-sets. It could be said that the issue of representativeness can be removed if we know that we have all the data generated so far relating to a phenomenon under examination under a particular set of circumstances. This moves from examining the data as to whether it is statistically significant, to issues that are related to "Big Data." Big Data has seemed to offer a way of broadening access to knowledge, or of creating knowledge from far more data, from the data avalanche, or the raw oil spoken of by Shadbolt and Berners-Lee. "Knowledge is the engine of our economy. And data is its fuel. . . We are at the beginning of a paradigm shift. Huge amounts of data are starting to be generated automatically. And we start being able to store, process and analyse these huge amounts. This can change the way we make decisions and run our businesses." (Kroes, 2013).

Big data has allowed the ideals of predictive policing or pre-policing to gain currency (whether logically or not). We see examples of it in use in the Policing Social Machines classification that follows. Many of the technologies discussed in the classification are based on the idea of drawing data from entire populations, from social networks, and from people's online habits. However, there are issues: the ease with which it can be captured allows the glossing over of questions of consent and privacy. Commenters on the Risk Society suggest that such large-scale data-gathering is seen as a sort of insurance, even if it is not really known what precisely such data will be used for. The debate about big data and privacy is ongoing. (See (Polonetsky et al., 2013), (Byrne Evans et al., 2013) for a discussion of some of these issues). (There is also a more general issue, of whether it might not be that Big Data and the Algorithmic Society is not creating a power imbalance that is causing some of the crime that produces crime data – we will discuss this in our conclusions).

Setting aside sociological, technical and legal issues, we come back to Jock Young's "datasaur" - drawing on Mills' earlier work on the Sociological Imagination. Young's concern was the need for "a method which can deal with reflexivity, contradiction, tentativeness. . . a method which is sensitive to the way people write and rewrite their personal narratives." Where we are examining data that relates to how crime is treated and processed through mediating technologies such as the Web, the issue of surveillance also arises. Haggerty and Erikson invoke Deleuze and Guittari's writing on the evolution of surveillance as via the emergent surveillant assemblage, and the “rhizomatic levelling” of the hierarchy of surveillance. (D. Haggerty, Richard V. Ericson, 2000). This – theoretical at the time that they wrote – “assemblage” is truly now invoked by “big data on the web”, a pantocratic entity where crime control and prediction are concerned. Those who use such big data for crime-fighting are concerned with issues around data warehousing, datamining, statistical and mathematical analysis.

Crawford writes of how “G.C.H.Q. and the N.S.A. are the old guards of big data, and despite their enormous budgets, technical infrastructure, and trained analysts, the big-data bonanza is not enough: They are reaching for other epistemologies by the dozen to try and make sense of it all.” Such epistemologies, we argue, should be found by combining approaches across disciplines. At a recent symposium run by D.S.T.L., it became apparent that the military concern and with it, that of the allied intelligence agencies is of understanding how to map the very nature of reality – a concern founded on the fact that their epistemological and ontological programs can stop wars and save the lives of populations. Elsewhere, Crawford says that, “Numbers can’t speak for themselves, and data sets – no matter their scale – are still objects of human design. The tools of big-data science, such as the Apache Hadoop software framework, do not immunize us from skews, gaps, and faulty assumptions. Those factors are particularly significant when big data tries to reflect the socialworld we live in, yet we can often be fooled into thinking that the results are somehow more objective than human opinions.” (Crawford, 2014). Crawford’s findings are of importance to this thesis’ conclusions; suggesting that in fact some of the methods being used by big companies, institutions and governments to understand the populations that they serve or control are in fact causing the problems that they seek to analyse using these methods. It seems that increasingly there is little to no science in Big Data; that it is often used as a bludgeon that blurs and constructs criminality where there may well be none in reality.

Good research should understand how best to use data that is available widely in ways that reflect our social world, but that are also made objective through rigorous epistemological analysis. Jim Hendler talks about “broad data”, data that has breadth, that spans enough phenomena and is in sufficient quantities to be able to pick apart, without coming up against either a) the problems of big data, or b) the problems of insufficiency. Broad data is what emerges from the millions and millions of raw datasets available on the World Wide Web. Challenges relate to “how to carry out Web-scale data search and discovery, rapid integration of datasets, and issues relating to policies for data use, reuse and combination. Hendler uses Open Government Datasets as exemplars. While numbers alone do not tell us about impact, and there are dangers associa.ted with loss of liberty and privacy infringements, relating to crime data in its more “raw” forms, it seems that in-depth qualitative research needs to be more than sampled data, and less than astronomic quantities of data culled, for example, from socialnetworks with little semantic processing.

Work has been done on examining how we can use computational approaches to big data, without losing the sense or meaning that comes from understanding the social contexts of such data. (Lazer et al., 2009). This is an important question for sociology in the 21st Century – we have access to far more data than ever before, but must keep asking how computational approaches can really uncover meaning. There are debates about whether merely “counting tweets” or whether broadbrush simple parsings of sentiment analysis can really hope to unpick the nuances of context and society – whether machines can unpick deeper meanings that are related to where tweets really come from. However to those exploring meaning via a relativist, constructionist epistemology, such corpora have a consistently symbolic interactionist meaning – people act towards things (i.e. tweets, or messages on social media that are open to collection by machine and machine parsing) based on the meanings those things have for them, and these meanings derive from social interaction and are modified through interpretation. On the one hand, from a symbolic interactionist's perspective it does not make sense, epistemologically, to say that sentiment analysis, "takes a naïve view of emotional states, assuming that personal moods can simply be divined from word selection. This might seem particularly perilous on a medium like Twitter, where sarcasm and other playful uses of language often subvert the surface meaning."

Constructivists and behaviourists alike must allow the meaning to lie in the use of such language, even if it is flattened out by its appearance on the World Wide Web. On many of their accounts "surface" is the meaning. On the other hand, it is entirely correct to look for more nuance, and to seek cross-disciplinary methods that allow deeper understandings if we allow realism to take precedence over relativism and its successors in our epistemologies. We can unpick the impact of Police.uk, by examining the population who links to Police.uk via social networks and analysing their apparent attitudes by using "Broad Data" methods. We can also use Broad Data methods to gather, examine and parse these data, while applying traditional sense tests and processing these data, with an understanding of the communities and the ethnographies that generate them. We can therefore explore what crime, crime mapping, crime data and policing appear to mean to these users, in terms of understanding both their experience of crime maps, and gathering robust knowledge from a large data set. We then know that that our findings are at least highly representative, in the corpora that are specified. Analysing the discourses of these visitors allows us to understand how the deployment of Police.uk shapes or frames debates around crime and crime data. Looking for frames does not mean that we can, with 100% certainty, say that policy or media or governments or institutions are shaping debates in certain ways, but we can start to analyse what meaning there might be in the presence or absence of certain concepts or discussions. We are therefore aiming to select certain frames from the examination of these discourses.

## 3.4 Frame Analysis

The question arises of how to analyse discourses gleaned through "broad data" methods. Frame analysis, as Goffman wrote about it, has a background embedded in pragmatist philosophy, and metaphysical debates about the nature of reality, such as those conducted by George Herbert Mead, William James, Charles Peirce and John Dewey. These debates filtered through to mainstream thinking until they emerge in Goffman's work in 1974.

In 1921, Lippman wrote, "Human public opinion in culture is very largely the selection, the rearrangement, the tracing of patterns upon, and the stylizing of, what William James called 'the random irradiations and resettlements of our ideas.' The alternative to the use of fictions is direct exposure to the ebb and flow of sensation... For the real environment is altogether too big, too complex, and too fleeting for direct acquaintance. We are not equipped to deal with so much subtlety, so much variety, so many permutations and combinations. And although we have to act in that environment, we have to reconstruct it on a simpler model before we can manage with it. To traverse the world, men must have maps of the world. Their persistent difficulty is to secure maps on which their own need, or someone else's need, has not sketched in the coast of bohemia.” (Lippmann, 1932, p.8). Bernays wrote in Propaganda, “This invisible, intertwining structure of groupings and associa.tions is the mechanism by which democracy has organized its group mind and simplified its mass thinking. . . (we) explain the structure of the mechanism which controls the public mind, and. . . tell how it is manipulated by the special pleader who seeks to create public acceptance for a particular idea or commodity.” (Bernays, 1928, p.8). Freedman explains that ideas coming from the counterculture and moved forward by the educated middle classes, gave rise to the idea that as mental constructs are needed to make sense of the world, we can never have more than a particular take on reality. It was therefore seen that those who succesfully shape others’ constructs influence their attitudes and behaviour. (Freedman, 2013). Although there are two debates here – one about influence and the presentation of reality and the other about the nature of reality - the issue of frames is very much a current one when it comes to work being done on Broad Data, computation, policy and crime.

The Chicago School kept Mead’s ideas about the symbolism of human action to the fore when it came to methodological approaches. Goffman was interested in the “organisation of experience,” a phrase which has profound philosophical and psychological overtones, and he examined both our perception and experience of the world, and how the media, politicia.ns, the theatre, and advertisers impact this. Although Goffman wrote before the emergence of digital communication technologies, the Web of Data and Online Social Networks, his analyses of behaviour and interaction now seem highly significant, in terms of current debates about how our perception of the networked world is directly manipulated by advertisers, ISPs, socialnetwork providers, Defence, the Government and its acknowledged and unacknowledged external contractors. This perception can be strongly linked both to how we identify ourselves (indeed, a socialconstructionist would say it is crucial), and how we are identified in a surveillant society. Goffman’s work applies to the ways in which people present themselves and engage in debate online. We can also apply the frames or self-presentation that institutions and organisations are involved with. Reputation Management and going further, Perception Management, or Information Operations are ways in which companies, institutions and governments have actually formalised this process, and these come to the fore in policy formation, and the messages that governments may use in order to introduce policies or even to fight wars which may be contentious. In looking at how ordinary people talk about crime and crime data, as it is defined by the Government and by policy, including the Transparency programme, then provided by the police via the Home Office and mediated by the World Wide Web, we bring some of these ideas forward. We suggest frame analysis as one of a number of possible approaches, because, as Schon and Rein said, “we see policy positions as resting on underlying structures of belief, perception, and appreciation, which we call ‘frames’. Controversies can be seen as disputes in which the contending parties hold conflicting frames which seem not to be teasable out through discourse, debate or reasoned argumentation.” (Rein and Schön, 1996). We can ask whether these are perhaps empirically generated controversies, capable of resolution via the application of more data, the combination of criminological and Big Data methodologies or whether some of the tensions shown in these controversies are irreconcilable.

Kahnman writing on decision-making under risk, quotes Shumpeter in saying that it has, “a much better claim to being called a logic of choice than a psychology of value”. (Tversky and Kahneman, 1986, p.1058). The application of the right research method might reduce the need for policy makers to dichotomise between fear of crime and knowledge of crime so that we can be informed both by choice logics and value psychologies. Rationality is certainly entwined with the question of whether or not we should fear crime: while on the one hand fear can be defined as a cognitive/emotional device that enables the recognition and assessment of risk, and therefore the alteration of potentially risky behaviours, on the other, there is plenty of research that shows that fear does not always have consequences that are rational, that risk assessment generated via fear might not be accurate. One version of framing says that “an issue can be viewed from a variety of perspectives, and be construed as having implications for multiple values or considerations. Framing refers to the process by which people develop a particular conceptualisation of an issue or reorient their thinking about an issue.” This makes sense – that we can understand stakeholders’ or actors’ views and perspectives around an issue by looking for their presentations, both of self and issue and its pertinent effects. Combining these two (sometimes controversial) approaches creates a new framework with which to understand and analyse both data and theory with respect to the complexity of concepts of crime, people’s lived experiences of crime and the data that surrounds it and the impact of the web on these phenomena.

## 3.5 First Analysis:

The first situates Police.uk within an ecosystem. It contextualises the initial premises spun out from its launch and from the history of mapping, crime data and statistics. It examines the data in terms of creating an ontology drawn from the ontological premise of “Open” and then considering the characteristics of non-open data in either physically contiguous contexts or as precursors or companions to the Open Data in Police.uk. We chose grounded theory as our methodology for this. It answers the question of "Where?" Where does the data flow and alongside which other bits of data and which systems? It picks up the themes of openness and secrecy, fear and risk versus assurance, as explored in the literature review, and analyses Police.uk’s use of data, knowledge and technology in relation to other representative web and technology-mediated ways of addressing crime. This creates an ontology that brings together new ways of exploring the central question of how best to understand the impact of Open Crime Data in the transparency regime by understanding exactly who the actors and technologies are. Although using Grounded Theory to better understand objects and people as opposed to exploring peoples’ experience, is more unusual we justify this by referring to this remark from Glaser and Strauss in 1967 which itself is used in Leigh Star’s paper on information-retrieval: (Leigh Star, 1998). “There are some striking similarities. . .between field work and library research. When someone stands in the library stacks, he is, metaphorically, surrounded by voices begging to be heard.” (Glaser and Strauss, 1967). We are then presented with a knowledge base or knowledge graph that provides the units of knowledge that can act as a base for the second and third pieces of analysis. “Classification is an uncovering of the thought-content of a written or expressed unit of thought. The reference librarian. . .applies the classification scheme in the ultimate stage of library service which is effecting contact between the right reader and the right unit of thought in a personal way.” (Ranganathan, 1951, p.116).

“The landscape of information retrieval is shifting rapidly (with networked distributed computing, large-scale digital libraries, and enormously powerful search engines). As the introduction to this issue notes, formerly firm boundaries between library and office, catalog and desktop are transmogrifying. The change means that a wider range of human activities come under the purview of library and information science. When the library and the desktop become seamless, then practices of work organization become part of the cataloging and indexing process. This merger calls for methodological creativity and cross fertilization between previously disparate methodological domains." (Leigh Star, Op. Cit.). This builds on the point made by David Wall, "the relationship between crime and technology is by no means new and... the potential for creating harm never seems to be far away from any apparently beneficial technological development." (Wall, 2007, p.2). While we use the term "harm" cautiously, comparing risks and benefits of these uses of technology allows us to examine problems that arise from web-mediated transparency and accountability, as opposed to other sorts of risks and harms that arise in the crime and web context. We use grounded theory to make some classifications and find clusters of similar sorts of crime technologies. We also use the concepts of risk, threat perception, deterrence and compellance within the context of crime as information, and crime as signal co-created by policing, the public and the media. This paves the way for the second element of analysis.

## 3.6 Second Analysis:

The second element uses network science, frame analysis and some of the theorising behind broad data methods to explore the attitudes towards crime and crime data of those who link to Police.uk, using online social networks. We analyse these public discourses, within the context of crime stories, narratives and the shaping of the way in which crime becomes a problem that moves between person and public. We look for common themes, and frames. This piece of analysis then situates Police.uk within the results of the first analysis, showing where Police.uk sits with regard to other crime technologies producing data according to opinions expressed by those people using it and reacting to it. The second piece of analysis takes point of departure comments coming from people reacting to the data/information on Police.uk on socialnetworks. It spans all networks and picks up all comments, in opposition to Home Office's attempts to qualitatively survey people who put themselves forward and are answering in constrained circumstances. It creates a small slice of "Big Data" thinking, aligned with the concepts of statistics, surveillance and crime data. It answers the question of "What?" – what does data say, what do people say on O.L.S.N.s reacting to the data – how are they using it naturally? Do they see the data as situated in the way that our first analysis has shown? What do they think it means?

## 3.7 Third Analysis:

The final analysis explores concepts or frames that arise from interviews with people who have used, produced and shaped the data that is both produced by Police.uk and some of the data that is not. This then helps us to see where Police.uk sits within the first classification, and what the differences are between the more flattened-out data gleaned from users of on-line social networks reacting to Police.uk and the more in-depth explorations of the experiences of informed commentators. The third piece aims to understand a disparity between One and Two. Analysis One has situated crime data in a particular place within an information ecosystem. It has created a framework for doing this over time, even when systems and data change. Analysis Two has suggested that ordinary people believe certain things about the data. The third piece of research aims to illuminate both pieces of work by providing the understanding of those who navigate the systems on a daily basis – it provides analysis from a range of actors. Any differences found in beliefs or understanding between the frames and the discourses then help to further contextualise the first classification or grounded theory, and thus a methodology is created for understanding data relating to crime that is mediated by technologies, including the World Wide Web.

# Chapter 4

# A Classification of Policing Social Machines

## 4.1 Contextualising Police.uk

Police.uk appears visually to be like other sites or apps offering mapped crime data at the time of writing, for example, www.Ukcrimestats.com, www.crime-statistics.co.uk and www.crimereports.co.uk. However, preliminary analyses of these sites show that although they appear similar, they are very structurally different if we use mixed methods approaches combining Network Science and linguistic analysis to exam-ine their networked context and content. We see that it is important to think about social, economic and psychological factors, if we are to understand how to design sites or apps that have similar intentions behind them, or how to understand their effects, criminologically for example. Examining these sites using network science measures such as eigenvectors and betweenness centrality, shows that because of the unique characteristics of the web, two visually identical web-mediated maps with the same crimes appearing on them can be fundamentally different in terms of their intent and the effect they will therefore have on those consuming their data. If we examine the use of language in these sites (particularly the discourses of risk and fear) and those they link to, and which contextual ads appear, and if we think about their economics, we find that while Police.uk and Crimereports link to other sites that are informative and reassure, a site such as Ukcrimestats appears to be linking to sites that promote "fear of crime" and that sell security in order to help people feel safe. This suggests that further understanding how Police.uk and its companion sites are positioned in the world of networked crime apps on the Web could help us to better understand what we are looking at, and some of the ramifications.

### 4.1.1 Related Work

A useful and illustrative concept that has accompanied work into understanding the relationship between people and technologies, and was inherent in Berners-Lee's first conceptualisation of the web, has been that of "social machines". Our first approach to exploring the crime data ecosystem was to examine how crime data relating to Police.uk moves between and is mediated by people and technologies. However, when we saw the myriad ways in which crime data is available on the web, and transformed by it, especially by the web's iterations and evolutions, it became apparent that protean manifestations of both crime, society and the web mean that representations of crime (i.e. data) occur in many non-simple forms, and that this is made more complex by the nuances of the web; its infrastructure, interstices; and its populations of burgeoning and dying communities – in short – the masses of free-formed and evolving technologies and social practices that combat crime. Given these factors, we posited that such an exploration of crime data could be more usefully carried out by combining Web Science with the concept of “social machines”, in this instance “policing social machines”. Social machines are first mentioned in the context of people and technology by Professor Sir Tim Berners-Lee, in “Weaving the Web”: “Real life is and must be full of all kinds of social constraint – the very processes from which “society” arises. Computers help if I use them to create abstract social machines on the Web; processes in which the people do the creative work and the machine does the administration.” (Berners-Lee and Fischetti, 1999, p.186).

Commonly used examples of social machine are: Wikipedia, Facebook, the Darpa Balloon Challenge, Reddit and Zooniverse. The concept helps us to unpick some of the ways in which humans come together to use technology to solve problems. Taking this view of what a crime or policing social machine is, we assume that Police.uk is an example; it uses the web to administratively present data that has been recorded by the police on crimes reported to them by the public. The data allows people to understand more about crime, policing and justice. Research in the last decade into social machines, (Shadbolt, 2011), explores issues coming from technology platforms for crowdsourcing knowledge and how we might define the various concepts that are pulled into such attempts. Alongside this, our work, (Byrne Evans et al., 2013), introduced some social, moral, policy and technological issues with our current crime data, especially when it is reproduced within the Transparency and Accountability context as “open” crime data. These problems have made apparent the need for a classification of crime technology, or policing social machines, as part of ongoing research into how the Web helps us to address crime, and how, in the U.K., the site Police.uk, that publishes open crime data as part of the transparency and accountability programme , contributes to this.

An earlier social machine classification had been carried out, using examples of “health social machines”. In “The Crowd Keeps Me in Shape: social psychology and the present and future of health social machines,” Van Kleek et al. provide a classification of health social machines. (Van Kleek et al., 2013). They used grounded theory, and clustered examples of health social machines into behavioural intervention, disease management, and collective sensemaking. Behavioural intervention machines “are systems that seek to help individuals achieve certain health-related goals by altering their daily routine(s) and activities”. These could be device-based, such as the Nike FuelBand, or app-based such as Fitness Pro or site-based such as Fitocracy. A second class of machines “aimed to help individuals cope with various kinds of conditions, including illness, disease, and mental health.” These included BigWhiteWall, a site that allows patients to cope with mental health issues such as depression. The third class they found aims “to crowdsource knowledge about disease, symptoms, treatments, and available resources to individuals who have personally experienced them.” Collective sensemaking allows large scale aggregation of symptoms of diseases and crowdsourced intelligence on self-diagnosis and treatment. These included sites like PatientsLikeMe. Interesting points about devices and crowdsourced treatments emerged from these clusters; it was therefore decided to see whether the same classifications could be applied to crime technologies. We also wanted to see whether conclusions drawn about similarities or differences between policing social machines and health social machines at the web/technological level could be extended to policy level. Much research draws comparisons or assumes similarities between crime / crime-fighting and health.

### 4.1.2 The Classification Method

As suggested in the methodology chapter, following Van Kleek, we use Grounded Theory and continue a "running theoretical discussion, using conceptual categories and their properties". (Glaser and Strauss, 1967). We use the categorisations found by Van Kleek et al., although with an initial scepticism as to whether crime social machines could be classified alongside health social machines – health and crime seeming to map different sets of phenomena. We began by collecting examples of technology that addresses crime, and then by applying the three distinctions to them. As we made these distinctions, these affected the collection method, by clarifying what we thought we were looking for, so that the collection and the classification fed into one another.

To seed the searches, we used Google Alerts, the major search engines and meta-search engines such as www.DuckDuckGo.com and searched through blogs and news-sites featuring crime, crime prevention and crime apps. Later in the research, I participated in enough policing groups and helped with research, that I also became personally acquainted with more technology that is used to fight crime. The initial search terms were: "social machines crime, crowdsourcing crime, sensemaking crime, collective sensemaking crime, collective intelligence crime, human computation crime, crowdsourcing crime, crime statistics, police statistics". Defining search terms made it further evident that there was need for empirical or a posteriori investigation rather than using an a priori definition of social machines, as what was referred to by S.O.C.I.A.M., the research group who had anchored the term while I was conducting my research, as "social machines" had previously been referred to via terms such as "crowdsourcing", "crowd based computation" and "collective sensemaking". On the other hand, Google Ngrams shows the term "social machine" appearing in 1818.

## 4.2 Analysis

The question that we first asked was whether the results that we got back from our initial searches could fit into the classification clusters identified by the health research of behavioural intervention, management, and collective sensemaking. The results were diverse: there were many sites to do with reporting crime, few of which seemed long-lived, there were numerous discussions on forums to do with crime, whether these forums were professional or not, and then a large number of accounts of the ways in which police and security agencies were attempting to address crime, in some cases using devices or surveillance.

Following the earlier work on health social machines, it seemed logical to use technologies and mediative elements as clustering factors. However, as we applied these to our data we found that we should indicate the process that was occurring socially. So for example, we could have a Facebook platform that allowed people to spread information about potential sightings of a missing person. This differed in application to an example I was given, of some police using Facebook to verify an informant's persona, in order to weight the information they provide. So Facebook on its own, as a grounding category, would not provide enough information to distinguish between its use as a platform and its use as an information weighting mechanism. We needed to work out: 1) what social element of addressing crime was occurring, 2) which technology was being used and 3) what the technical or computational process was. There is a sub-element to 1) which is, "what counts as crime?" This, although initially appearing to be a minor sub-question, is, of course an entire research area on its own. Later interviews and case studies showed us the vast array of elements that can fit under this category, from deviance to elements of global wars and terrorism.

A second sweep of the sites made us consider how they fitted into “stages” of how society addresses crime, from preventing crime to reporting on crime, to managing and collecting information about crime and making judgements. These categories then helped us to start clustering the data (sites and apps), and grouping it into sites that provide the public with general information, sites that provide more up to the minute information on where crimes are being committed, sites that allow the gathering of decentralised data for professionals and sites that allow the public to crowdsource particular problems. We also found a conceptual divide between peer-to-peer use of technology and specialised sites for professional use only.

### 4.2.1 How Crime is Addressed

After further examination of the sites and apps the searches had returned we arrived at three sets of classifying dimensions: the first, shown in 4.1, characterised how crime is addressed, spanning overall knowledge of crime, reporting on crime, risk evaluation and crime prevention, to “solving” crime, with solving or providing evidence flowing into making judgements on criminal or deviant behaviours, which judgements then fed into the cycle of reporting on crime and knowledge of crime.

### 4.2.2 The Technologies

Along the second dimension we considered technologies, shown in 4.2. We started by defining technologies as mechanisms mediating between an (grossly simplifiying this concept) averagely-equipped human, and their experience of the world. We began with chemical mediators – anti-depressants (mediating human perception of the world and possibly helping reduce fear and aggression), and chemical castration used to prevent crime or deviance, f.M.R.I. scans (to detect intention), sensors, cameras, glasses, and devices that “augmediate” perception, tablets, phones, laptops, P.C.s, mainframes, networked systems, through to environmental and building architectures, and finally, laws. We already see a society where technological structures mediate socio-political-legal mechanisms – on the web, censorship prevents people from accessing illegal sites, for example. From a Web Science perspective our goal was to contextualise web-based technologies emerging from this wider trawl.

### 4.2.3 Mediation

Along the third dimension we projected that we could indicate how crimefighting is mediated by the technology under discussion, and that this would then be part of the categorisation of the data and technologies. We could use A.I. terminology to decide whether the system involved “sensing”, perception, reasoning, knowledge, planning, learning, communication, or other forms of interaction. These dimensions are shown in column 7 of the bottom row showing “characteristics of the data/Platforms using in Addressing Crime in Society” in ref:The Groups of Crime Social Machines.. Again, these rough initial categories are blunt instruments, but cover the general areas of sensing, processing and outputs. As we applied our clusters to the data we had, and considered the implications, we added more dimensions, discussed in the conclusion. Using these dimensions would allow us then to sift through our initial searches, and start organising them in ways that would help to logically think about

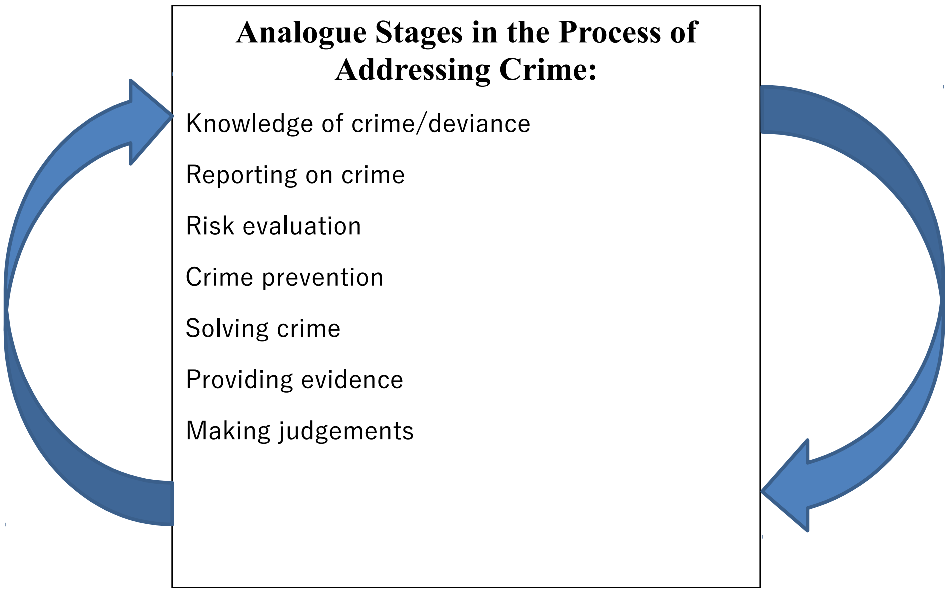


FIGURE 4.1: Stages along which Society Addresses Crime

.

distinguishing characteristics of successful or unsuccessful sites. The grounded theory approach means that these are not set as permanent definitions, but pragmatic ways of slicing through concepts spanning both society and technology that pull out the most salient factors to consider when building further sites or apps and considering return on investment.

### 4.2.4 The Four Groups

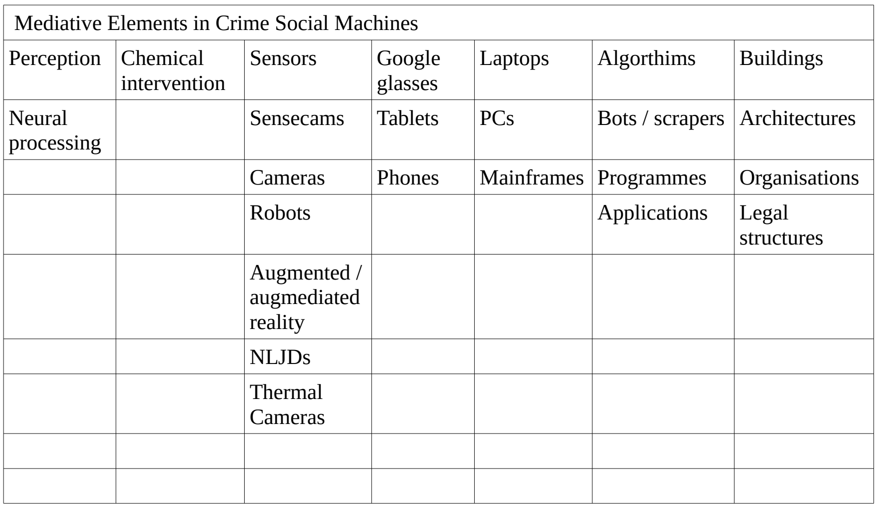
The clusters shown vertically (below) shows some examples of the following: We found four overall groups: Knowledge and Report Groups, Behaviour Intervention, Crime Management and Sensemaking. Van Kleek's definitions had worked to a point, but we felt that we needed a further category that allowed us to add sites and forums where professionals discuss crime, and that seemed to serve "knowledge of crime at a distance". This sat well with Police.uk and Crimereports.

These move towards "knowledge of crime in the here and now" which came into apps that allow members of the public to map, or at leasrt report crime as it occurs nearby, or to them. We felt that these fitted into the behavioural intervention group of the health machines classification, as they, if working well, allow members of the public to modify their behaviour if it seems that they might be about to walk down a path where rapes occur or drive into a dangerous part of town. We found that the boundaries between these groups were not clear cut – which did make us question the categories we had formed, although an answer is that perhaps our society's concern with crime means that the myriad ways in which technology addresses it are evenly distributed.

Knowledge of crime as seen on Police.uk dissolves into knowledge of crime as seen on Harrassmap, with perhaps only the viewer's visceral response to the information, their "fear of crime" as opposed to their "knowledge of crime" as a differentiating factor. We see knowledge of crime coming in from the public, administratively focused, with perhaps a degree of comfort and security attached to the public's trust in sites like Crimestoppers – well-known and well-used. This administrative management of crime dissolved into allowing the public to work on crime information themselves – where the dimensions move from public intelligence gathering to public use of such intelligence to solve crimes. Such public intelligence gathering could be witting or unwitting, so we then felt this category moved towards surveillance, from public self-surveillance to surveillance by policing or intelligence organisations, via Open Source Intelligence (O.S.Int), or devices.

#### Knowledge Groups

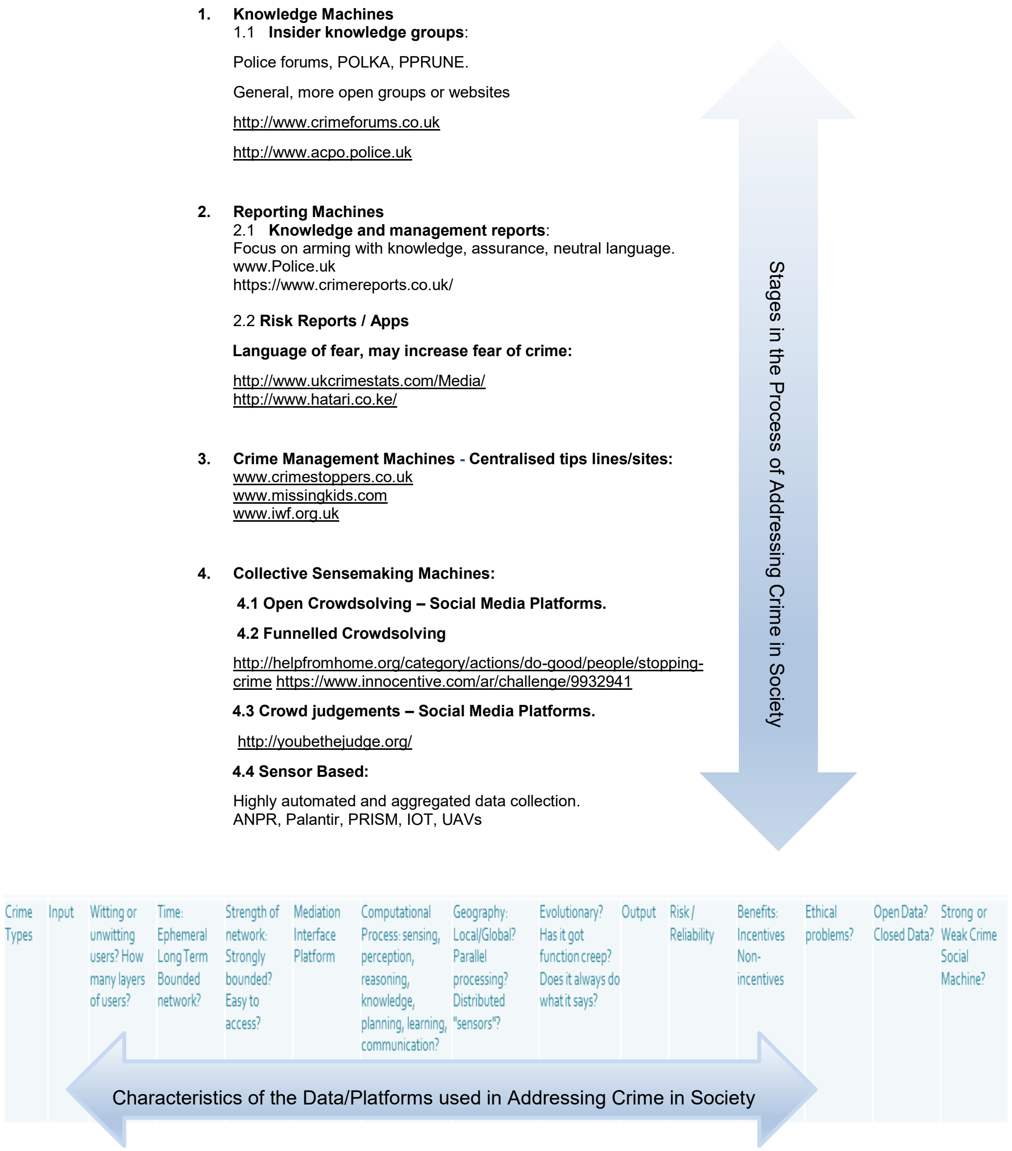
These were mostly insider forums where crime professionals swap tips, and provide professional support and awareness. They feature discussion on policy initiatives and those in higher authority, while the prevalent discourse was professional and showed domain appropriation. There were also forums where professionals whose work overlaps with crime or criminological concerns exchange advice. For example the Professional Pilots Rumour Network has discussions on combating terrorism, using plane-spotters to detect unusual activity, whether 9/11 was a conspiracy and problems with sensationalist reporting on air crime. It was here that we started to see a polarity between discussion and action, or intervention sites. Where the focus was on knowledge, we also included sites that provide reports on crime, using highly processed data: so Police.uk and Crimereports fitted in here. If we compare these to Ukcrimestats we see they appear to offer information about crime using crime open



| Mediative Elements in Crime Social Machines | | | | | | |
|---|---|---|---|---|---|---|
| Perception | Chemical intervention | Sensors | Google glasses | Laptops | Algorthims | Buildings |
| Neural processing | | Sensecams | Tablets | PCs | Bots / scrapers | Architectures |
| | | Cameras | Phones | Mainframes | Programmes | Organisations |
| | | Robots | | | Applications | Legal structures |
| | | Augmented / augmediated reality | | | | |
| | | NLJDs | | | | |
| | | Thermal Cameras | | | | |
| | | | | | | |
| | | | | | | |

FIGURE 4.2: Mediative Elements in Crime socialMachines

.



1. **Knowledge Machines**
   1.1 **Insider knowledge groups**:

   Police forums, POLKA, PPRUNE.

   General, more open groups or websites

   http://www.crimeforums.co.uk

   http://www.acpo.police.uk

2. **Reporting Machines**
   2.1 **Knowledge and management reports**:
   Focus on arming with knowledge, assurance, neutral language.
   www.Police.uk
   https://www.crimereports.co.uk/

   2.2 **Risk Reports / Apps**

   **Language of fear, may increase fear of crime:**

   http://www.ukcrimestats.com/Media/
   http://www.hatari.co.ke/

3. **Crime Management Machines - Centralised tips lines/sites:**
   www.crimestoppers.co.uk
   www.missingkids.com
   www.iwf.org.uk

4. **Collective Sensemaking Machines:**

   **4.1 Open Crowdsolving – Social Media Platforms.**

   **4.2 Funnelled Crowdsolving**

   http://helpfromhome.org/category/actions/do-good/people/stopping-crime https://www.innocentive.com/ar/challenge/9932941

   **4.3 Crowd judgements – Social Media Platforms.**

   http://youbethejudge.org/

   **4.4 Sensor Based:**

   Highly automated and aggregated data collection.
   ANPR, Palantir, PRISM, IOT, UAVs

Stages in the Process of Addressing Crime in Society

| Crime Types | Input | Witting or unwitting users? How many layers of users? | Time: Ephemeral Long Term Bounded network? | Strength of network: Strongly bounded? Easy to access? | Mediation Interface Platform | Computational Process: sensing, perception, reasoning, knowledge, planning, learning, communication? | Geography: Local/Global? Parallel processing? Distributed "sensors"? | Evolutionary? Has it got function creep? Does it always do what it says? | Output | Risk / Reliability | Benefits: Incentives Non-incentives | Ethical problems? | Open Data? Closed Data? | Strong or Weak Crime Social Machine? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Characteristics of the Data/Platforms used in Addressing Crime in Society

FIGURE 4.3: The Groups of Crime Social Machines

.

data, presented geographically. However the latter appeared to be conceptually mediated by "risk or fear" of crime, while Police.uk informs about crime and offers advice and assurance on prevention. For example, Police.uk/hampshire/2FG02/crime provides advice about safety across a number of dimensions.

On the other hand, Ukcrimestats links to sensationalist articles on crime, and adverts are served up in response, partially, to what advertisers "see" on the page. Many of these adverts are themselves risk and fear-based, further enhancing the difference between the Home Office site and Ukcrimestats. On their media page there are links to articles with language such as "reckless," "staggering," "fear," "suffer," "most violent roads," "crime-ridden streets," "hellhole street," "blighted", "chilling." The difference emerges on examination of the links themselves, which seem to underline the idea that fear of crime is being used to promote the site. If a site is known by the company it keeps, these sites offer two very different understandings of crime and society from a network science view.

### Behavioural Intervention

It was when trying to discover how, if in any respect, Police.uk differed from similar sites and looking at Ukcrimestats that it became evident that there was a cluster of apps that fell within the remit of discourses on risk and that were more "interventionist". These sites are one remove from the "knowledge of crime" sites. Where the knowledge / crime data sites seem to be about reassurance, the risk apps often helped to collectively source decentralised knowledge of crime happening "right now". Risk sites both report crime and can modify behaviour – they inform the crowd about potential risk areas, and provide a mechanism for reporting crimes of varying degrees of seriousness as they occur, which then can have the effect of modifying users' behaviour, by stopping them from walking in "unsafe" areas for example. They may target particular strata of society such as women at risk of violence, but are available (to those who can access the technology) globally. There can be a peer-to-peer element to these; alternatively, the sense that those in authority are providing knowledge and advice to the less-informed. There was a plenitude of machines designed to address particular sorts of problem, from domestic abuse, car stealing, drug abuse, to cybercrime and cyberbullying. These offer, as with the health social machines, general knowledge resources, such as places to get help, online advice, what to do to avoid or prevent attacks, activities to support safety and general support; again, as with the health machines, intervention techniques, counselling and advice. There is a striking similarity between these and the health social machines.

### Crime Management

Another class of crime social machines aims to help crime professionals and in some cases, the public, to manage crime. These are again "tips" based but funnelled or centralised. These dissolve into "crime management", moving away from the behavioural intervention grouping, as knowledge is flowing from crime "amateurs" to the crime professionals, and while they mitigate risk, this is absorbed by the language of professionals so that there is less "language of fear". As with the health machines, a dimension along which both risk apps and crime management systems vary considerably is the degree to which these machines encourage participant anonymity or identity disclosure. The U.K. Crimestoppers site says, "Crimestoppers' promise of anonymity has never been broken. If the identity of one of our callers was made known it would destroy trust in Crimestoppers and no one would contact us. This is another reason why it is so important to us that we can guarantee your anonymity." It is notable that they do not specify who is promising exactly what to whom. The more effort an institution makes to convince the public that it is trustworthy, the more one might question whether this is so – i.e. trust is established by looking at behaviour, not promises.

Other sites encourage individuals to use their offline identities, or their normalised online identities either explicitly or through implicit disclosure, such as police use of platforms such as Facebook. Where this disclosure is implicit and people providing information are not aware of what mechanisms there are in place for evaluating the strength of the data they provide, it is on record that getting reports about crime from people using a Facebook sign-on allows the police to evaluate the information, through assessing the reliability of the information provider.

### Collective Sensemaking

Another group of crime social machines that parallels the health social machines classification is where collective sensemaking occurs. We found two distinct categories; one being the collective approach to solving a crime, "social solving" or "Crowdsolving", and the other being the crowd making judgements on how to deal with crime, "social judgement machines". We then postulated a third variety of sensemaking, which was of official surveillance of web-(and other device) mediated crowd behaviours.

The three fields in which sensemaking are commonly used are H.C.I., information science and organisational studies. In this instance we are looking at groups that enable people to collectively solve problems. Klein et al., (Klein, Moon, and Hoffman, 2006), presented a theory of sensemaking as a set of processes invoking causal reasoning, hypothesising, feedback and learning. There is reference to Minsky's early work on frames, with feedback and re-hypothesising contributing to the re-framing of ideas. This also picks up on attribution theory (Heider, 1958), and the naïve scientist hypothesis, with an iterative layer added to the notion that we observe phenomena and then attempt to make sense of them via hypothesising, "with the important aspect being that neither data nor frame comes first; data evoke frames and frames select and connect data. When there is no adequate fit, "the data may be reconsidered or an existing frame may be revised." As with the health social machines, these examples crowdsource knowledge of crime and in particular how to "solve" crime. Crowdsourcing knowledge of crime, perhaps invokes a temptation to relate experiences of crime as "symptoms", discussion of "treatments", and resources for individuals who have experienced these "symptoms". However most forum discussion does not always end in agreement on causes of crime, and therefore its correct "treatment".

For example the site www.liveleak.com shows murders, assaults, car crashes, and other disasters, some of which are criminal, with discussion from commenters. There is little evidence of web vigilantism, or "digilantism" resulting with people agreeing on a treatment for a crime. However other sites specifically cater for the "digilante", and platforms such as Twitter, Facebook and online news sites such as Reddit, Gawker, Jezebel and theweek.com have been used to "out" individuals such as Violentacrez, @comfortablysmug, racist teens, Lindsey Stone and Hunter Moore. The generic treatment for the deviant / criminal behaviour exhibited in these cases is online exposure or "doxing" and shaming, a behaviour that goes back a long time in history. Daniel Solove suggests that shaming occurs in inverse proportion to the apparent absence of judicial punishment, seen as "extra-judicial" punishment. (Solove, 2011) This aggregation of opinion, at large scale, relates crime/deviance to treatments and effects, some of which threaten individuals' right to privacy, as Jonathon Zittrain has pointed out. (Zittrain, 2014).

Crowdsolving comes about via sites such as www.getyourcarback, Facebook, Imgur and Reddit pages dedicated to finding individuals, or providing information about crime – for example, the disastrous attempts to find the Boston Bombers. (Imgur hosts the pictures used on Reddit). There are also sites like www.helpfromhome.org, or www.innocentive.com where the challenge is to find a person or for example, to match latent fingerprints. Devices such as automatic number plate recognition, speed cameras and lie detectors seemed to fit here, as they provide data from the crowd that is then analysed by law enforcement. This could arguably fit into crime management, but we felt that one possible dimension that could be applied was the notion of witting or unwitting provision of data or intelligence from the crowd. A crowd can self-surveil and happily give up its data, or be surveilled, unknowingly.

## 4.3 Results

### 4.3.1 Parallels between Crime and Health

We found some interesting primary parallels between the clusters found in the health social machines and the policing social machines. These seemed to diverge in the area of salience, feedback, transparency and surveillance. Similarities were that we can share consultations with "crime professionals" in a PatientsLikeMe context. People experiencing crime can make judgements like those made by the users of PatientsLikeMe. If their experience of dealing with crime and of receiving advice from professionals were shared on forums, then as with the health context, this would create more transparency about the ways in which crime professionals do their jobs, and provide peer-based scrutiny. This could make the records of these professionals visible and provide public reputations that would then enable decision-making about trust, if done with enough regard to maintaining some privacy for those in the public eye.

Where policing social machines enable the exchange of information, they act in a similar way to the health social machines that function as answer gardens; they have an emotional support function; where people have experienced crime, the impact can be enormously damaging and there are plenty of forums where people can support each other. The same problems with crowdsourcing knowledge occur; well-documented, with controlled studies looking at bias, confirmation bias, illusory correlation and explaining away in the knowledge realm of making causal links between symptoms, causes and treatments; this is just as much of concern here, if not more so. In fact there is debate over causes of crime, and further debate over whether criminology is epistemologically resourced to explore these issues, with police often resorting to "crime science" in an attempt to avoid some well-known criminological elephant-traps.

### 4.3.2 Incentives

Money, anonymity, gamification and social conscience are seen throughout as "good" incentives; we found some darker incentives too. Money can be used as a reward for reporting, sometimes on Crimestoppers, often there are rewards for capturing criminals, and Internet Eyes apparently (at the time of writing) pays people to identify criminal acts. Some of the crowdsolving platforms offer rewards; however, we found that money appeared very little as an incentive; with the emphasis being more on social incentives. Gamification occurred, with the possibility of trivialising serious crime. The site YouBeTheJudge does the opposite, austerely gamifying sentencing so that one competes against the judge and sentences provided by 'amateur judges' are compared against "real-life" decisions and explanations provided. Crimes are given "moral panic" headings with the crime then broken down in such a way as to deflate the outrage factor and provide facts in a clinical way. This may well have the effect of reducing "fear of crime" engendered by mass media– in effect Judicial P.R.. Its aim is to show users how judges and magistrates go through the decision-making process before passing sentence. This is a crime social machine that detaches discussion of sentencing from the mass media.

For crowdsolving problems based on hashtags, or on "find a person" Facebook pages, or Reddit threads, incentives seem more closely linked to being a good member of society. It is possible that there are darker forms of incentive, that would enable us to ask questions about the philosophy of punishment, and whether and how revenge comes into justice. We recognised Solove's atavistic incentives in some of the shaming behaviours seen on some of these pages that could be explored with regard to gamification. It also seems clear that there are cases where toxic online exposure (see GamerGate) appears to be motivated by feelings of socialinjustice. The Social Justice Warrior (S.J.W.) is a distorted artefact of this.

Social encouragement is more noticeable in many crime apps, with morality being an obvious incentive to take part. Some cheering on is noticeable in police Facebook and Twitter notifications, as with exhortations to report crime to help keep society orderly. There was little overt evidence of people being encouraged to compete. Here incentives seem based on goodwill, although as ever with the web, one can ask whether sites or forums that generate comments are in fact also participating in link-building, for the purpose of raising funds through advertising revenue. There can be other more complex motives. The hero complex is a well-documented psychological condition, where for example, "vanity" crimes are committed by security guards, where people create havoc in order to help people avoid it. (Copperfield, 2006). There is an interesting psychological area, retaining to extra-judicial self-help and digilantism, which is where civilians enjoy feeling that they are "fighting crime". (The Pleasant Progressive, 2016)

Where the policing social machines are more anarchic, there is a suggestion of "trolling", with members competing to seem the most detached from the horrors that are being shown, such as on LiveLeak. This then suggests another dimension, aside from knowledge, risk and fear – there are the prurient, potentially dissolving into psychopathic onlookers, where viewing and commenting on deviance can dissolve into deviant or criminal behaviours. In all of this, the underlying vulnerability is that the capacity of the World Wide Web to amplify on a huge scale a citizen's desire to do the right thing, and other darker, not so commonly acknowledged desires, leads to problems with mob-justice. This is the arena where Information Operations occur, and where troll farms send their produce in bulk to sow the fertile ground. This is where it becomes very hard indeed to find truth and where problems with bad science, relativism and apathy can potentially rock democracies.

Where we examined sensemaking, as social solving and social judgement, there was little evidence of provision of structured elicitation processes, whereas we found the risk apps depended somewhat on this, allowing users to get precise knowledge concerning locations and times of crimes, and evaluation of the accuracy of the information. Under social judgement we saw often quite chaotic responses, and not orderly diagnoses concerning crime. There are television shows that are gladiatorial in nature (Judge Judy in the U.S. for example, or Jerry Springer, and the erstwhile Jeremy Kyle in the U.K.), where the audience is encouraged to pass judgement in the form of statements, providing evidence, or jeering at persons brought in ostensibly to have their problems dealt with. Although these are weak examples of policing social machines as they are often about disorder, and use an older technology than the web, they are examples of gladiatorial policing social machines, that offset the more orderly example given by You Be the Judge. This site does elicit information from naïve "crime-fighters" in a highly structured, non-emotive way, with this information of use as "polling" information about what judgements self-selected visitors to the site would make in particular cases and presumably how this then reflects on current opinion on sentencing policy. This certainly provides more objective information on sentencing policy than the tabloids. In the first case, we see crime and judgement as spectacle, as entertainment, in the sense that The Old Bailey's original publications served; in the second, a more educational aspect.

### 4.3.3 Challenges

Most challenges seemed to be sociotechnical. One is of how to preserve anonymity in situations going from peer-to-peer knowledge gathering to official gathering. It was notable that when crowdsolving occurs there is less concern about anonymity, where official bodies elicit information, anonymity comes into play. Looking at the U.K. Crimestoppers' Google Analytics snippet shows
_gaq.push(['_gat._anonymizeIp']); implying that Crimestoppers remove any I.P. information that Google might supply, although it is not clear whether they have any "social notifications" switched on that would then allow them to track users who have bookmarked the site and are discussing it on forums. Google themselves though may have I.P. address information, even if this information is not pushed through to the analytics that Crimestoppers receives from Google. So it is not certain that anonymity is preserved – while Crimestoppers might not have the information, Google does. This is to be borne in mind with attempts to understand behaviour of users on a site in order to make the site function better – if anonymity is promised as an inducement for reporting crime, then methods of tracking users in order to make the site work more efficiently start to affect that promise, in a viciously circular problem.

If visitor data can be properly anonymised then this is worth bearing in mind for future site design, but it is here that we see technological interfaces blurring, perhaps deliberately, organisational boundaries and knowledge for the user of who it is that they are actually providing information to, a further question to be explored with regard to the surveillance conducted by analytics and advertising companies, and very much of relevance when it comes to cybercrime. Tracking user behaviour took us to the question of data collection and processing, from the collection of "raw" data for analysis by the crowd or by crime professionals i.e. analysis of video footage, or automatic number plate identification, to communications interception by government agencies.

Salience and feedback seemed more publically lacking or were managed more dogmatically among the policing social machines than among the health apps; i.e. various government departments send reminders to pay tax, to get your vehicle assessed to government standards of safety, thus preventing people from breaking the law through forgetfulness. There are various forms of mass-push notifications such as schools automatically texting if a pupil is absent, or if payments are not made for school meals. These do not operate from individuals self-surveilling.

Monitoring devices are external, impersonal and authoritative: registers, teachers, databases. More secretively, there are salience and feedback systems that monitor individuals at airports to see if they are showing physiological signs of stress that might indicate that they are suicide bombers. The closest self-monitoring we found was beepers on satellite navigation systems that tell you if you are speeding, or if you are about to pass a traffic camera. It is not clear whether these beepers are intended as a "nudge" technology, that asks whether you know you are about to commit a crime, or whether it is to prevent being caught.

There are increasingly seen to be ethical problems with policing by machine, where the temptation to gather data leads to muddied social, technical and legal issues. The idea behind panoptic programs such as those referred to as *PRISM*, *Carnivore* and *Echelon*, is that we have the means to gather data that might reveal intelligence on terrorist activity – so why not use it? The seductive appeals to give up a little privacy, and some civil liberties in order to be secure, have been well-rehearsed, although in fact it is not clear who exactly becomes secure and what security really means in terms of risk management as opposed to actual knowledge of crime for the government, the intelligence agencies, or the citizens being surveilled. (Friedewald, 2009), (O'Hara and Shadbolt, 2008), (Regan, 1995), (Solove, 2011). And monitoring of the monitors seems to be driven by the media, leading to tussles between intelligence agencies and individual journalists and whistleblowers, played out in public. While governments might have moved from Dionysius' Cave – to attempting to know all by pervasive listening, it seems very likely in the context of Information Warfare that their strategy is as much about control by fear and uncertainty, as it is about gathering knowledge of crime from us, of using panoptic mechanisms so that none of us is certain when and how they are surveilled. It is never clear how much of the interplay between governmental surveillance and those who observe it and protest it, is for strategic purposes as much as to reveal real outrage that should, theoretically drive reform, where it is needed. However, looking at the problem of surveillance from the perspective of Social Machines leads us to conclude that we happily self-surveil for personal or consumer reasons. A possible avenue for further investigation is that of understanding how policing and government can sit in an ecosystem that acknowledges data generated in such ways i.e. our smartphone and Facebook data, and to think more deeply about how we can encourage people to feed such self-generated data into systems that fight crime, without sacrificing privacy or loss of control. As Zittrain forecasts, this will probably be the biggest problem for privacy. It is worth thinking about how we monitor those who use our data, without being distracted by media posturing on this.

There is also a movement towards seeing crime in terms of risk, with insurance companies weighing in. There appears to be a trend in insurance companies favouring driving enforcement via in-car devices, and an increasing interest in the Internet of Things policing us via insurance companies. There need be no intervention from police, unless there is an accident, just a simple punitive increase in insurance costs where the driver speeds. These sorts of crime social machine are deeply embedded in the concept of the risk society, where sometimes misleading statistics and perhaps out-dated theories of personality, trait and profiling swamp individual's rights.

This led us back to the economics of risk and fear – and the point during the lifecycle of a crime at which an app asks for money, as a hugely differentiating factor. Where someone is reporting a crime in order to defend themselves and is asked to provide money to do so, it seems that the economics of the app is based on cruelly leveraging the victim's fear of crime. Where such leveraging is being done via the open data app economy this must lead us to ask whether the transparency that led to such an economy is in fact good or cynically exploitative? This is true to a lesser degree, where fear of crime is used to sell an app that provides knowledge of crime. Where such data is free, as on Police.uk then it seems to come from a perspective of assurance.

As with the Health social machines survey, we see dangers in self-diagnosis or self-report of crime, where data is crowdsourced. How do we know that reports can be verified? Detection involves a weighing up of the facts; where reports come in en-masse this becomes critical. Anecdotally it seems that public trust in Crimestoppers is not necessarily a problem; but that out of the masses of calls that they receive in the U.K. only about 5-10% appear to have substance.

While the survey of health machines addresses the problem of self-reporting and concludes that crowdsourcing self-reported knowledge can result in bias, this applies equally to the official U.K. crime data sets coming from the Home Office – crime statistics, Open Crime Data and the British Crime Survey, said to reveal the true "dark figure" of crime. The problems with crowdsourcing knowledge are well-documented in (Van Kleek et al., 2013), which cites controlled studies looking at bias, confirmation bias, illusory correlation and explaining away in the epistemological realm of making causal links between symptoms, causes and treatments; this is just as much of concern here, if not more so. In fact there is debate over causes of crime, and further debate over whether criminology is epistemologically resourced to explore these issues, with police often resorting to "crime science" in an attempt to avoid them. The survey of policing socialmachines shows that this problem of modelling causation applies just as much to self-report crime information as to health. One solution might be to index crime data, creating meta-data about the source of data sets. Official crime data comes from policing performance data in many cases; B.C.S. data from surveys on perception of crime.

Ulrich Beck has pointed out that there is little academic research on the subject of I.C.T.s in policing and Manning, (Manning, 2011a), has suggested a lack of evaluation of "interactions between technology and social organisation and practices because little has been written about the practices, constraints, and opportunities associated with the use of the new information technologies". (Lyon, 2008). Their concerns do not just apply to the police as a force or police as a service, but can equally be said to apply to the use of these technologies for policing in general, or the pluralities of security provision.

### 4.3.4 Crime and Transparency

Many challenges come from thinking about how these web-mediated crime technologies relate to transparency. Policing social machines can make people more crime-literate – some of the increase in reporting rates of crime can come from people becoming crime aware. So, literacy about crime concerns and considerations, and a mechanism by which individuals can get the best crime expertise available, whenever and by whomever is best placed to help, is needed. But we can then see transparency coming into play with some associated issues: we should consider ways of mapping awareness of crime, in order that reports can distinguish between increases in crime itself, and increases in awareness of crime leading to increased reporting. We must also consider the uneven, pluralised provision of policing services to the public. If we marry public self-policing and monitoring of the police themselves to an already "rationalised" police performance culture we must beware transparency. It is clear that the structures and criminal activities that dictate how a police officer does her job, and what constitutes good performance, will not be consistent across forces. Technology and the "joy of data" can exacerbate these inconsistencies and lead to irrationality deriving from the impossibility of defining consistent measures across policing. If it is hard to model crime causation, does it not follow that the treatment of a social phenomenon so hard to understand should not be salved by a 'plaster' or whitewash of performance data? We need transparency about crime rates and about how those we pay to address crime do their jobs but not when those jobs are dictated by targets, rather than crime. Looking at transparency and incentives leads back to John Flatley's comment about police data being affected by "possible perverse incentives associated with performance targets." Police reporting comes from a target-driven culture; (Flatley, 2013), surveillance and reporting can help professionals achieve their targets; where targets are the issue, or the threat of terror, the incentivisation appears high, but it is to the detriment of individuals, or in ways which threaten civil liberties and privacy wholesale. There is no "possible" in the perverse incentives, targets are killers when it comes to crime and policing. There can also be security ramifications from too much transparency and this could create more risk for survivors of crime.

Surveillance crime social machine technology seen via transparency again has considerably different consequences, compared to the health context. In the health context, monitors and sensors are used to understand lifestyles and devise appropriate interventions. In the crime context, such devices where employed by the state, can be representative (often) of the state's control over the individual. We only need to scan recent headlines on the N.S.A. and G.C.H.Q. to see exactly how in this context, these sorts of technologies are seen as detrimental to our liberties and as privacy-threatening. There are many instances where their use is accepted, such as traffic cameras, the use of tagging on offenders, where in fact, the effect is the same as with the health context, "tagging gives specialists unprecedented, accu-rate access to an individual's daily activities. This information could give clinicians valuable context for understanding each patient's lifestyle between visits, in order to devise more appropriate interventions". (Kleek Op.Cit.) Where these machines are employed by individuals, their use is less contested, but as with much of the crime context there is a grey area where it is not clear how machinery is used by persons who fall somewhere between the state and the individual.

Technology is neutral but its use can be political – no technology is going to provide a solution unless it can capture the complex socio-legal-economic processes that interweave crime and criminality. So another area worth examining with regard to transparency is the balance between the use and hidden or open ownership of technology or infrastructure by Officials fighting crime or by corporations, as opposed to private citizens.

### 4.3.5 Implications for Police.uk

Having carried out this classification we see that the categories we found allow us to say that the Home Office data and Police.uk sits within a potentially quite dis-tinct sub-category of crime social machine, that of informing and assuring. It does not provide a statistically predictive service alerting users to immediate danger or altering short-term behaviour, but it serves to let people think about reported crime in their neighbourhood and take longer term steps to help. The data is about policing, not crime. The data maps trends in policing, as well as reported crime and the relationship between the police, the public and policy. It is shaped by the systems it moves through and the processes it undergoes, and the way in which it is mandated. To understand why a crime is reported and why it appears on a map, we have to understand the confluence of all these things. When people look at the crime maps the Home Office produces, they know crime is recorded, documented and processed; it is monitored, they can see crime-fighting across the U.K.. It appears that the Home Office data offers epistemological "knowledge that" a crime was committed, that the police captured it. Other, similar sorts of reporting seem, with their risk-mediation to be about "knowing-how" a route to understanding how crimes were committed with their real-time reporting.

Haggerty and Ericson invoke the Flaneur in "the Surveillant Assemblage." (D. Haggerty, Richard V. Ericson, 2000). Drawing from the works of Gilles Deleuze and Félix Guattari they suggest that, "We are witnessing a convergence of what were once discrete surveillance systems to the point that we can now speak of an emerging 'surveillant assemblage'. This assemblage operates by abstracting human bodies from their territorial settings and separating them into a series of discrete flows. These flows are then reassembled into distinct 'data doubles' which can be scrutinized and targeted for intervention." While Deleuze and Guattari might see Police.uk as part of the surveillant assemblage, the classification has shown that there is still a distinct difference between the sorts of observations it gives access to, "setting up house in the heart of the multitude, amid the ebb and flow of movement, in the midst of the fugitive and the infinite," and the sort of participant surveillance reminiscent of "the figure of the badaud, the gawker or gaper. Fournel wrote: "The flâneur must not be confused with the badaud; a nuance should be observed there. . . . The simple flâneur is always in full possession of his individuality, whereas the individuality of the badaud disappears. It is absorbed by the outside world. . . which intoxicates him to the point where he forgets himself. Under the influence of the spectacle which presents itself to him, the badaud becomes an impersonal creature; he is no longer a human being, he is part of the public, of the crowd." (Fournel, 1858).

It is hard to explain exactly how technology can help to mediate crime-fighting, when sometimes the activities involved in addressing crime can seem to border on participatory Badaudism: voyeuristic, atavistic, psychopathic, judgemental behaviours. The web has the peculiar property of externalising and objectifying our subjective moral compasses, both in the way that we come to judgement and in how we decide which behaviours are valid in addressing crime. We suggest that Police.uk is strongly allied to the flaneur, unlike some of its lookalikes. We return to the question of how the classification might start to answer some social, technical and policy issues, including those coming from Police.uk. We can see (a) how crime data is being used; the apparent worry about "faked data" dissolves into a more sensible discussion of the socialorigins of policing data, and that perhaps if the target culture were removed this might then remove perverse incentives to "shape" data according to often irrational targets. We have seen how (b) data can be crowdsourced, and we have started to examine some of the attendant problems of anonymity, evaluation and incentives. These first two points presumably help not only the public, but the police themselves. We have asked (c) whether data and apps such as these can help us to address crime, without increasing fear of crime, and looked at the way in which the information economy might drive some designers to sell crime data or a sense of safety through leveraging fear of crime.

### 4.3.6 Policing Social Machine Signatures

Along the third dimension we had projected that we could indicate how crime is presented or processed: whether the system involved sensing, perception, reasoning, knowledge, planning, learning, communication, or other forms of interaction. As we applied our clusters to the data and considered our conclusions we also added the following to be applied: whether inputs were collectively sourced or funnelled via mechanistic or human processes, whether the data was open or not in inputs/outputs, how strongly bounded in terms of time, platforms and definition, or networks of people, sensors or machines that provide or process the data, whether technology is recent or well-established, who owns it, whether it is provided or processed wittingly or unwittingly, what the incentives are that enable the machine, whether there are ethical risks pertaining to the data, whether the technology actually works only in one direction, or whether it is easily subverted, and the degree of certainty about the data that is produced. This is future work, but we suggest that doing so will enable us to more clearly evaluate crime technologies and establish policing social machine "signatures" that will quickly identify the potential success, risk or threat posed by these emerging technologies. We can then decide whether money is well-spent on such policing social machines, and how policy should be set regarding their use, both by policing services and the public. Crime and policing social machines are complex, organic, evolutionary systems where behaviour is hard to predict, other than by outward observations of network characteristics. (Byrne Evans et al., 2013).

Motivations and incentives are hard to analyse from any one disciplinary perspective because these are embedded in notions of morality and psychology that shift depending on the assumptions that the researcher brings to their attempt to observe or build a machine. Looking for success factors in social machines, or blueprints to build new crime social machines needs a clear and thorough examination of the ways in which these individual, social, moral, legal and psychological factors come into play when humans are connected en masse via new technologies. So building a crime or policing social machine, using Web Science perspectives, to illuminate current crime data in the U.K. should not necessarily just focus on complex technical problems to be addressed, but align these with softer ontological issues associated with for example eliciting sufficient information to understand both what it is a victim thinks they have experienced, encourage them to return to populate their report, preserve anonymity while collecting sufficient behavioural data to understand their interaction patterns.

We know that the Home Office data is largely performance data, and that the categories recorded constitute about a fifth of potential crime categories in the U.K.. It is very much knowledge-based and sits within the context of assurance and mapping as a way of scientific understanding, as opposed to other, remarkably similar looking crime map sites that seem to use fear of crime coming from risk analyses to sell services (including the data itself). It is possible to differentiate these similar-looking sites by means of behaviour analyses – investigating which sites they link to and which sites are linking to them. Further examination should allow us to see whether the geographically-presented data is well-traversed by people who land on it and then take pathways through the system which lead them to obtain advice on crime, and crime-prevention, which is currently one of the aims of the Home Office. If we examine sites which link to www.ukcrimestats.com we see a discourse of fear, and articles linked that have an undertone of moral panic: "Anyone craving a life free from crime should consider a move to Wales." We suggest that www.police.uk sits within a new distinct generation of ideological transparency, geared around the concept of the 'open' movement, with the potential to be allied with sub-sets of data appearing from crowd-sourced information that provide a supporting context. We have seen a tension in the provision of two sorts of data; performance data, coming from the background of policing via accountancy, that is automatically pushed out with regularity, and that waits for serendipitious re-use, and data that is requested, data that might be one-off, data that answers a query about something, and that could also be shaped by less serendipitious re-use.

Meta-transparency means that we view both data and its genesis together, eschewing old ideas of "raw" data that stands alone without reference to its own history. So that although we can say that this is data that has been produced as part of the policing performance culture, and therefore might be reflective of target-setting, or surveillance of police doing their jobs, via a system of distrust, in areas where performance is reasonably well-guaranteed or because of pressure from the media for example; we know this, as understanding of data provenance is implicit (or should be) in the culture of Open. This data produced by the Home Office is different from any other sort of data in any other crime system as our classification should show. Although there is dispute over what the data actually presents, we suggest that the data is representative of more than just evidence-based knowledge about crime, that it is not empirical data as such. It is tempting to think that the Home Office data is data that could enable us to solve crimes, or to predict where crimes might occur. There is an empirical view that if you study the meta-data of a criminal enough, i.e. their footprints, their movements, their background and culture, their paths through the city and the points at which they actually commit crime, that this is enough to tell us when next they might do so again. This is the hot-spot approach. And transferring this into the digital domain is tempting (and has been done, with some success, elsewhere, as we have seen). So it is tempting to believe that with the Home Office crime maps we should be armed with knowledge of crime and can predict future crime. This does not seem to be the case with the Home Office data and its representation.

Taken a further step back, writers such as Manning question the use of such devices and the rationalising of society. He suggests that although crime mapping and crime analysis together have been seen as "a fundamental window into the transformation of policing, there is no convincing evidence to support this claim of a direct effect of C.A./C.M.." Although Manning does not clarify exactly what he means by evidence, and seems to be somewhat undermining his own argument in calling for evidence, presumably itself rationalistic and causally clarified, to support the claim that digital evidence helps in policing, he raises a number of very important issues on the problems that underpin debates about policing, crime, crime data and technology. Pointing out that the relationship between police, crime and society is not fully understood does not mean that crime mapping cannot help in furthering society's "knowledge of crime." Manning's concern is more with policing. Our analyses of both crime socialmachines and of the debates engendered by the availability of crime maps to the "public" shows that the availability of data on maps draws in a huge audience, all of whom can be seen interacting (see the next analysis) with the information therein, and some of whom we can see participating in quite fevered debates on what these numbers means for racism, stop and search, causes of crime, crime "hot spots" and ways of avoiding being detected in the commission of crime. While Manning cites Heidegger's concerns with technology, (Manning, 2011b), as an instrumental means to an end and technology as a human activity, argumentative dichotomies in this realm are actually now starting to dissolve with multi-and inter-disciplinary study and approaches which can both absorb larger-scale understandings of the interplay between humans experiencing technology, and experiencing the crime maps and the information “means to an end” that the police might see these as.

On the one hand Manning criticises the way in which the police use data, in unlinked databases which are not measured or calibrated in terms of their efficiency, storage capacity and use, saying that “research suggests that they are rarely and poorly utilized,” while on the other he suggests that this paradigm of efficiency and utility that he himself invokes, is somehow weakened by being just one of several competing paradigms. He is quite correct in saying that the present mode of policing shapes the data collected, not the other way round, but following his invocation of Heidegger, this follows naturally when we look at technology as human activity. We have seen that current debates about the nature of crime create differing approaches leading to different sorts of data; we can now think about impact in ways which avoid assuming that we can learn everything from realist, far-reaching and all-encompassing data, or from small-scale, empirical understandings of some of the ways in which crime is mediated via Police.uk for some user groups. We can look for methods of evaluating impact that are theoretically explicit about both their heritage in terms of defining crime before we set out to measure it, and more sophisticated in their understanding of how technologically-mediated crime data is experienced by society, and the resulting impact. We can even think about new theoretical approaches that combine both syntax (numbers) and semantics (experience / meaning) of crime, and generate different sorts of data. We can start using some of the paradigms of big or broad data, in order to better understand emerging technological approaches to observing the web and “to help address grand social challenges.” We can explore the limitations of “reality mining” and its associated philosophical and financial speculation.

Combining this understanding of crime data with the advent of the World Wide Web, we are now in a position to be able to gather congruent data on a far larger scale than before. Using a combination of Web Science approaches, we can in addition to the modes of analysis referred to above, augment these with network analysis to track reactions to Police.uk on social media and find out who the influential promulgators of information relating to crime are in the networked online world. We can source different types of data - e.g. narratives generated by social machines, and examine the impact of this sort of open crime data, with a very clear understanding of the context that it operates within.

# Chapter 5

# Frame Analysis

## 5.1 Contextualising the Frame Analysis

In this analysis we look at the way in which people discussing crime statistics make sense of them, including uncovering the interpretive frames they use, in order to pro-vide empirical findings to expand conceptual models of crime, and facilitate mean-ingful discussions about how to improve matters. All data reported were drawn from core studies conducted between January 2011 and August 2014, although re-search is ongoing.

As the previous chapter shows, we have contextualised the sites presenting open crime data, by examining the crime technology and data-producing landscape. One finding was that common conceptualisations of crime data do not take into account the networks linking to the data and where they sit within the information economy. Our approach to analysing the context of such data-mediating sites showed that even if sites present the same geographically mapped open crime data then they are not necessarily all doing the same thing – they are not sitting within the same ontological space within the open data ecosystem. We need not therefore assume that such sites are, either, i) the same as each other, or, ii) in competition with each other. The question of competition has led to at least one policy debate about the role of the Home Office in presenting this data, alongside other sites which also use the same datasets. (O'Hara, 2013).

We suggested a unique methodology that draws on an understanding of the socioeconomic and political drivers in the information ecosystem, and that uses network science and a content analysis approach in order to create ontologies or classifications that reference risk and fear as well as assurance and knowledge creation within such sites. This uses network and social science research with mathematical and statistical tools that seek to explain phenomena such as diffusion of innovation, news and rumour. Other related research looks at networks as a means of understanding markets, with trust in relationships and information asymmetry being of particular interest. At the time that this research was started, the need for this was dormant; at the point of submission of this thesis, it is very clear that diffusion of news, innovation and rumour are phenomena that urgently need to be addressed and properly understood.

With this methodology we found that Police.uk appears, currently, to be distinct from many other sites that are also using open (and other sorts of) government crime data. These sites link heavily to web spaces where the discourses are of risk and fear of crime, or where the primary aim is to for some sort of behavioural intervention to be exploited. We also established further questions to be asked that would help to draw out aspects of crime or policing social machines that serve to differentiate them from one another.

In our next piece of analysis we wished to examine more deeply discourses relating to Police.uk as they appear on the web, looking for further evidence of impact. “There are a number of approaches that allow an examination of users’ views that sit under a general framework of textual analysis. Since the study of language in use…(is) a goal of education, a means of education, and an instrument of social control and social change…For many the interest in discourse is beyond language in use." (Jaworski and Coupland, 2006). “Language use relative to social, political and cultural formations . . . , language reflecting social order but also language shaping social order, and shaping individuals’ interaction with society.” (Lantolf, 2006).

The impact might be instantiated in the form of discussions about behaviour change, of increased fear of crime, of casting doubt on the numbers, or of either increased knowledge of crime or increased perception of knowledge about crime. Such sorts of discourses then help us to think more deeply about the design of Police.uk, its role and how information about crime can better be communicated to the public. As discussed in the methodology chapter, the advantage of using broad data methods here is that we can scan the web for all public discussions that link to Police.uk on particular social networks, and be confident that the sample we are collecting is at least as representative, if not more representative than collecting comments from people who self-select for interview. We also capture the differing environments that these conversations occur in, so gain a deeper understanding of triggers and contexts for web-based comments – a sort of virtual ethnography (while acknowledging the many difficulties attendant with this). Given current debates on the validity and accuracy of crime data, and worries about the validity or rationality of fear of crime, frame analysis seemed appropriate in seeking to understand what perceptions about crime are current. Such methods have wider applications too, in the security and intelligence domains.

### Context

As the literature revealed, intelligence markets became robustly entwined with broad data concepts in the U.K. in the 1600s where ‘all was to be known, noted, enumerated and documented. The conduct of persons in all domains of life was to be specified and scrutinized in minute particulars, through detailed regulations of habit, dress, manners and the like.’ Understanding how the ’warding off (of) disorder through a fixed ordering of persons and activities’ has been deemed very necessary ever since then. Governments, organisations, researchers, advertisers and companies with a large global presence also now have an interest in all being known, noted, enumerated and documented - not necessariliy for the purposes of warding off disorder.

Police.uk was not designed using a secure evidence base showing what the results might be. It appeared pragmatically as a result of the government wanting to be transparent, and using apparently available data that seemed to fit the bill. Chainey and Thompson, (Chainey and Tompson, 2012), suggest that it would “help improve the credibility and confidence that the public had in police-recorded crime levels, address perceptions of crime, promote community engagement and empowerment, and support greater public service transparency and accountability.”

We have looked at some of the debates about how Police.uk would be used and what its effect might be that were current at the birth of Police.uk, and found that the Information Commissioner’s Office, (I.C.O.), had outlined potential benefits and problems with crime mapping in some detail. The I.C.O.suggests that, “it is not yet clear how, or to what extent, the general public uses crime-maps or third parties use crime data. The use that is made of crime data has implications for their design." Elsewhere the I.C.O. says, "Crime mapping can be an effective means of letting people know what crimes are taking place in their local area and we have advised the Home Office and local forces on how such systems can be designed to take account of privacy risks - particularly when 'point data mapping' risks identifying individuals (especially innocent victims, witnesses or vulnerable offenders) or risks disclosure of sensitive personal information about those individuals if, for example, they have been the victim of a racially motivated crime or a sexual assault. We encourage the use of privacy friendly options that reduce the risks of identifying such individuals such as by merging adjoining postcodes in sparsely populated areas and banding together certain categories of crime. The I.C.O. would be concerned if privacy risks arising from aggregation with other datasets in the public domain were not taken fully into account."

In the advice given on crime mapping, the Information Commissioner points out that "crime mapping can give citizens a readily accessible means of understanding patterns of crime in their area", for example, it can help them to work out what the police should be prioritising, how their performance shows whether they are doing this, and to make informed judgements about safety and well-being. However there are long-standing rules relating to confidentiality and identification of crime victims, witnesses and perpetrators, which when translated into technological means of showing the above, result in a lot of meaning being lost from the data. The Information Commissioner also points out that while crime mapping is useful when it comes to understanding which areas to avoid with high levels of street robberies, it is not so useful in being able to evaluate police performance. We return to this point later, as one key finding in our analysis is that precisely because this data comes out of policing performance data, and is not "crime" data per se, it actually currently serves only the latter purpose, where police performance arises from target-setting. (Information Commissioner's Office, 2010).

### 5.1.1 The Method

Our preliminary hypothesis about examination of impact was that we should look for evidence of:

- Debate about accuracy: of numbers of crimes, of reporting, of the "dark figure".
- Debate about knowledge of crime versus fear of crime.
- Debate about the nature of crime itself.
- Debate about policy and policing – how do we use information to make changes?
- Debate about privacy and surveillance.

### 5.1.2 Tropes

We carried out a manual analysis of the data alongside processing carried out by Tropes – a Natural Language Processing and Semantic Classification software that performs textual analysis. This was after having explored (Jurafsky Daniel and Martin James, 2000), a number of different options. We then added frames as they became apparent. From a Web Science perspective, this made sense. We had looked at various packages designed to facilitate sentiment analysis, or textual analysis, but found that most of them served either as means of carrying out very simple concordances, or as online repositories for notes and material gathered from ethnographic studies. We were looking for software that would allow for a textual processing, that can be annotated and shaped by hand, once it has run through some initial counts and categorisations. While engaged in this research I also came into contact with large-scale commercial sentiment analysis software as used in defence intelligence enterprises, but found that it was hard to gain a scientific understanding of the algorithms used to process the intelligence therein. Tropes provides basic statistics on word occurrences and these can be placed within word categories and subcategories. While it can also provide statistics on what it calls co-occurrence and connection rates of equivalent classes and word categories, as well as probabilistic and geometric analyses, it also uses the concept of cognitive-discursive analyses – and looks for the most characteristic parts of text.

#### Statistical, Probabilistic and Cognitive analyses

Tropes carries out different sorts of text analyses:

- Statistics on the total occurrence frequency of the main word categories and of their subcategories,
- Statistics on the co-occurrence and the connection rate of equivalent classes and word categories,
- A probabilistic analysis of the words occurring in bundles, and a geometric analysis of the bundles delimiting the episodes,
- A Cognitive-Discursive Analysis (C.D.A.), making it possible to detect the most characteristic parts of text.

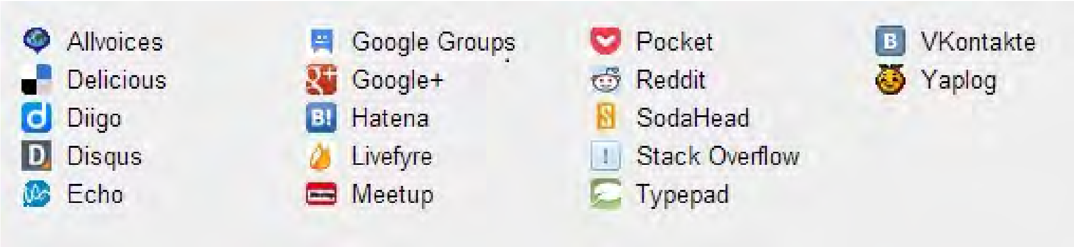
Word counts are used to build the graphs and to lay out the results. Frequent word categories and text style are captured by the comparison of occurrence frequency distribution of categories observed in the text with what are known as "linguistic production norms" - insofar as these can be captured. (See for example "Semantic feature production norms for a large set of living and non-living things" (McRae et al., 2005)). The norms used by Tropes were defined on the basis of analysis of "many texts". They are "stored into specific in-built tables." There is obviously much to wish for here, in terms of validation and transparency (especially given the confluence of a sense of desired objectivity in capturing "linguistic production norms" with the more ephemeral social construction of language and thought) - however after having used the software a number of times, it actually seemed preferable to other software with more modern interfaces that is less easy to deconstruct.

There is a philosophical problem that underlies most pieces of software that contain the word "semantic", and indeed, even the "Semantic Web", often invoked when discussions come up about the structuring of data needed for enterprises such as natural language processing, sentiment analysis or indeed, the geo-intelligence presupposed to be inherent in crime maps. From the perspective of this research, it is in fact the "syntactic web". There is often a belief that with enough syntax – ie understanding of rules governing word combinations, and statistical knowledge of the probability of certain word combinations having a likelihood of a certain meaning, we "gain semantics". This assumes that semantics or meaning can be arrived at through mathematical representations. This is of course a core problem for A.I.. Modern methods in A.I. are strongly associated with Big Data, as common sense might suggest anything built that has "intelligence" should be able to understand "meaning". There are many debates about what intelligence is, where meaning resides, whether it ever can be the product of statistical knowledge, and if propositional explicit knowledge is the only sort that can be pushed about by software to create an understanding of meaning. While we can understand and point to explanations about things which are propositional, propositions can never hope to capture all the sorts of things that can be known and the ways in which they are known. However, considering the opposite type of understanding of the world, that residing in Artificial neural networks that decode sentences and images, markets and sentiment, through Machine Learning, it is very hard to say in which mathematical representations as they occur at their various levels within the layers, meaning, as humans understand and construct it, resides. So where we use a classifier to deconstruct the textual reactions of people interacting with Police.uk on social media, we approach our "big data-esque" "semantic" text analysis with a degree of caution, asking the software to perform certain functions but choosing not to trust everything relating to the "meaning of the data" that is surmised from word counts and co-occurrence. (Lancia, 2005).

Our research found that Tropes was useful for extracting relevant information, carrying out some qualitative analysis, isolating themes and identifying principal actors – we could partially determine through its use, as opposed to the context of the original narratives, who said what to whom and who did what. Where and when were more problematic, as were purpose. Tropes also carries out some identification of text styles, a form of linguistic forensic analysis, useful for comparison with other texts. One comparison we make is with the interviews that are carried out in the third analysis. It can (mostly) non-problematically group verbs, adjectives, adverbs, personal pronouns and conjunctions, using "semantic meta-categories". It also carries out a chronological analysis that serves also to group discussion blocks into ideas being developed (although this part is not without its idiosyncracies). The software helped us to group together a heterogenous corpora – collecting all the utterances made on links to Police.uk on S.N.S.s. There is clearly a need for software that does not require interactions taking place with a logical sequence. This is useful in frame analysis as it helps to roughly separate out parts of speech relating to intention, doubt, assertion and location. Where we are looking for "frames" relating to doubt or confidence in numbers this is helpful. Tropes also helps us to focus on those comments that relate to risk, and also to comparisons – and who is making them.

This software also helps to distinguish between stative, factive and reflexive verbs – with stative verbs being very much aligned, in a pragmatic Goffman-like way, with our organisation of experience i.e. how the world appears, or how we conceptualise things. This allows us a degree of understanding of the constituent linguistic elements of the framing that visitors who comment on Police.uk have in relation to crime, and therefore some of the conceptual elements. With such a rich data set there is much potential here, for example, to explore users' psychological notions of causality suggested by such verb use – and to thus draw conclusions about the way they understand crime and numbers. See (Goffman, 1995). However, once again, because of the unstructured nature of the narratives, we have to pick through the examples found by the classifier to check that these meanings are to be found, or might be agreed upon by a certain number of researchers.

Finally although it is hard to represent visually, Tropes produces an ontology that is reproduced in Figure 5.1 below, which shows how Tropes has categorised the subjects referred to in the comments made on Police.uk. This is done through the scenario tool, or natural language ontology manager, an "intelligent Thesaurus Manager based on Semantic Networks and Natural Language Text Analysis technologies, which has ready-to-use classifications". We found it was necessary to amend



Allvoices
Delicious
Diigo
Disqus
Echo
Google Groups
Google+
Hatena
Livefyre
Meetup
Pocket
Reddit
SodaHead
Stack Overflow
Typepad
VKontakte
Yaplog

FIGURE 5.1: Social Networks Linking to Police.uk

these – which the software presupposes will be necessary. For example in a reference to a theft from a glove box, the ontology suppose the word "glove" means this fits into a sentence about clothing, rather than a theft from a part of a car. Also it assumes the "tackle" in tackling crime is referring to a sport. However once the scenario manager is restructured to represent the world that we are looking at, it can usefully help to pull out concepts, that can help us to check our hypothesised frames.

## 5.2 Analysis

Using Police.uk's Google Analytics Suite we were able to capture all comments from users who had visited the site, linked to it and then commented on it, on the social networks shown in Figure 5.2. Given the natural ebb and flow of such platforms (as covered in the work done in the first piece of research) not all of these platforms are now being used, however Google Groups, Reddit and Stack Overflow are still producing data that could be used. Notably missing are Facebook and Twitter from these social networks – further work could include scraping Twitter for further anonymised data; however pulling out all data for the same time frames is problematic on these O.L.S.N.s; whereas the ones shown above did not at the time provide limitations around data access.

### 5.2.1 Forms of Interaction

The various forms of interaction with the site were: posts, comments, bookmarks, reshares, and likes. Twenty-two users bookmarked the site, one hundred and twenty one users commented on the site, seventeen users "liked" a comment relating to a robbery at Selfridges that referenced Police.uk, three users reshared comments referencing Police.uk, and thirty users participated in original conversations containing a link to Police.uk. A total of 228 comments made by individuals were analysed: these individuals appeared to include (where it was legitimate to draw conclusions) professional I.T. workers / web developers or people working with open data who are engaging in specialised technical discussion, crime writers or journalists, security companies, spammers, and those whom we might call visitors representing a person on the Clapham omnibus, armed with technology.

There was some debate about the extent to which the data should be "cleaned". It was initially tempting to remove obvious repetitions where large chunks of text were found reappearing throughout the web, as they might seem to skew the results where we were trying to find out what frames were being produced in terms of what people "really" think and feel about crime and their knowledge of crime. Data was first set to one side relating to reshares, "likes", bookmarks and favourites. However, after some consideration it became apparent that crime is often used as an election / political manoeuvring issue and that some of the reshares were indicative of perception management or information operations, which now it seems are symptomatic of action by hostile states leveraging bot farms in order to sow discord or to create fragility. (One such stated aim being to ensure that "people think that truth is undiscoverable.") These were therefore kept in, partly to keep in mind how sentiment analysis can easily fall prey to bot activities. The frames we are producing may well be the product of such operations - something to consider in looking at web-mediated transparency achieved through Open Crime Data.

Comments were analysed in order to determine how users view crime data, whether they question it, whether they tie it into government policy, particular political parties and whether anyone explicitly sees it as linking to transparency and accountability. What are they using the crime data for? What questions are in their minds when they visit the site? What does visiting the site cause them to feel or to think they know? All the time we have to bear in mind that we presuppose persons are generating these comments - however we did find evidence of some more or less bot-like activity or spamming. Further work would have looked to match ip address blocks with known spam farms, through the use of Spamhaus or work with Nominet.

### 5.2.2 Visitor Profiles

Interest in crime data and crime statistics is growing considerably, perhaps in part owing to the opening up of crime data and resulting conversations in the media, the Home Office and Parliament, as well as among the public, about police recorded crime statistics, and more recently about the rise in violent crime and the fall in policing numbers. However, progress in understanding some of the issues with crime data, and which perhaps also reflect on the notion of transparency and "Open", has been hampered, perhaps because of widely used but conflicting conceptual models of policing data as a form of organisationally rationalised information transmission that lets the general public know about crime.

Visitors to the site are therefore going to be diverse. As a result, various organisations are involved or referred to herein: the Police, the Home Office, local councils, ordinary citizens. However we were careful not to attempt any sort of visitor profiling in depth, even with anonymised data, other than to attempt to gain a very general understanding of the range of organisational actors that were in play. We also see a "nesting" of locations or objects referred to: the web, hyperlinks, locations from which data is accessed, that also frame the discourses. This becomes interesting when we think about the spaces of the web. In analysis we became aware of a very real sensation of the web as a shifting territory, where contests over whose epistemologies constitute terra firma are fought. The crime data becomes part of the arsenal.

## 5.3 Results

References: the following part of our analysis considers occurrences of the most commonly referred to actors and conceptualisations in our work (References). The References frame can produce various visualisations of semantic categories, grouping closely related references (common nouns, proper nouns, trademarks) that appear frequently throughout the text. For example, crime has been grouped with "crimes" in the text extraction, but there appear to be no synonyms.

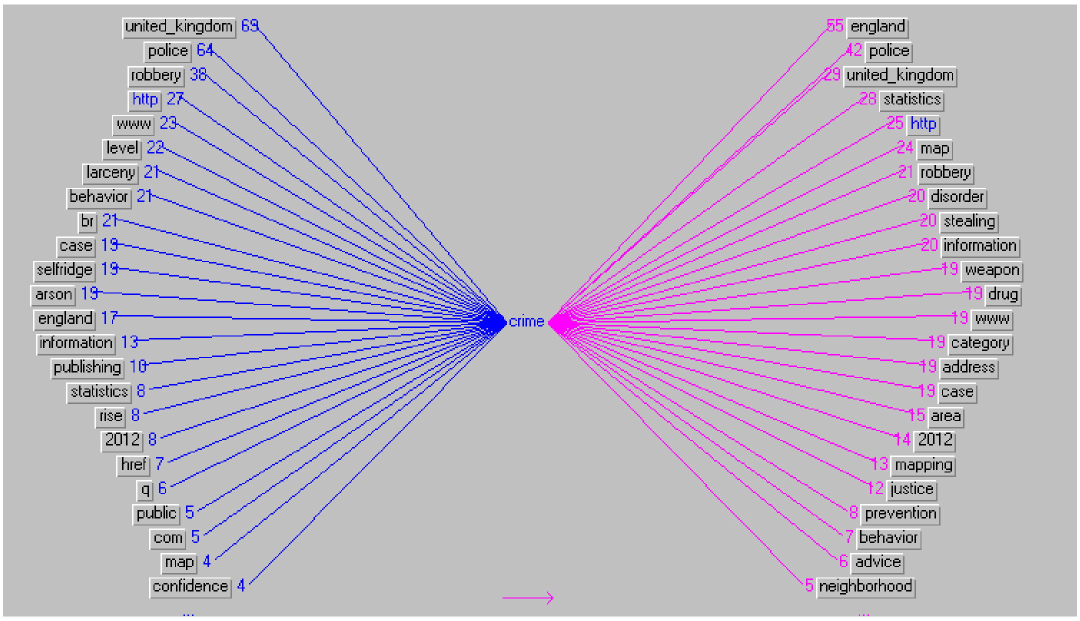


FIGURE 5.2: References to “Crime”

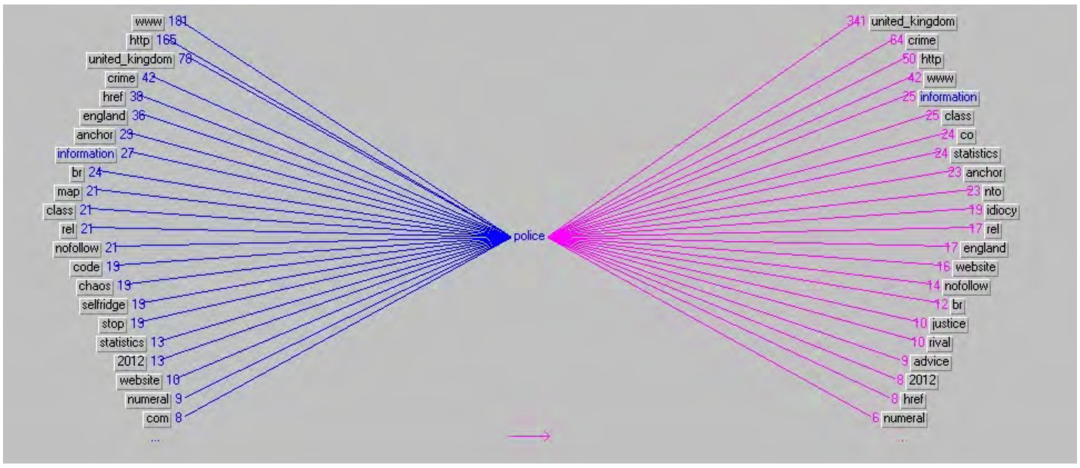


FIGURE 5.3: References to "Police".

### 5.3.1 Crime

“Crime” occurred as a concept 370 times, compared to “police” 413 times, “information” 103 times and “statistics” 65 times.

### 5.3.2 Police

Figure 5.3 shows Tropes’ categorisation of mentions of "police". In the comments, nearly all references to the word are to the site - there is little mention of the police themselves. While crime is mentioned many times, the police seem incidental to this.

### 5.3.3 Transparency

Figure 5.6 shows Tropes’ categorisation of mentions of "transparency".

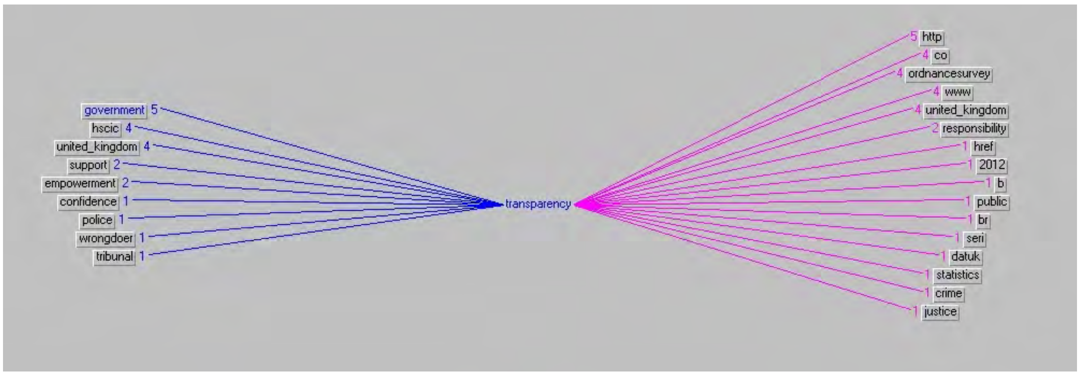


FIGURE 5.4: References to "Transparency".

> “and whether offenders went to court, increasing transparency in the criminal justice system...”
>
> “I am using the text below for a training exercise STORY Engagement, Empowerment and Transparency...”
>
> “and support greater public service transparency and accuntability. This article captures the policy rationale behind this initiative...”
>
> “and that the initiative instead has primarily become a tool for promtoing political transparency...”
>
> ” We suggest that future focus should be on improving the quality and cartographic visualization of the published information alongside the integration of socialmedia functionality to enrich local dialogue on crime issues”
>
> “ and support greater public service transparency and accountability. How effective has this large scale digital mapping effort been?”
>
> “Engagement, Empowerment and Transparency: Publishing Crime Statistics using Online Crime Mapping. Insights from this study have important implications for crisis mapping projects...”
>
> “and must avoid becoming an exercise in promtoing political transparency when the data it offers provides little that encourages the public to react...”
>
> “and the public confidence in the Government and the police force is to ensure complete transparency in crime statistics...”

### 5.3.4 The City

Figure 5.7 shows Tropes’ categorisation of mentions of "the city". The comments pulled out by the software show signs of the flaneur, or of human terrain: it is here that the most emotion seems to be attached to what is said about crime, and the remarks or responses are far more contextualised.

> ”Keep the oxford for around town. The above set up is very chunky. Remember that a mtoorcycle is 2 times as likely to be stolen as a car...”
>
> ”The statistics suggest that there is twice as much robbery in the city centre as in Radford...”
>
> ”Like any city be sensible I would say. 30 plus years, walking home drunk through the red light district, Hyson green, Radford...”

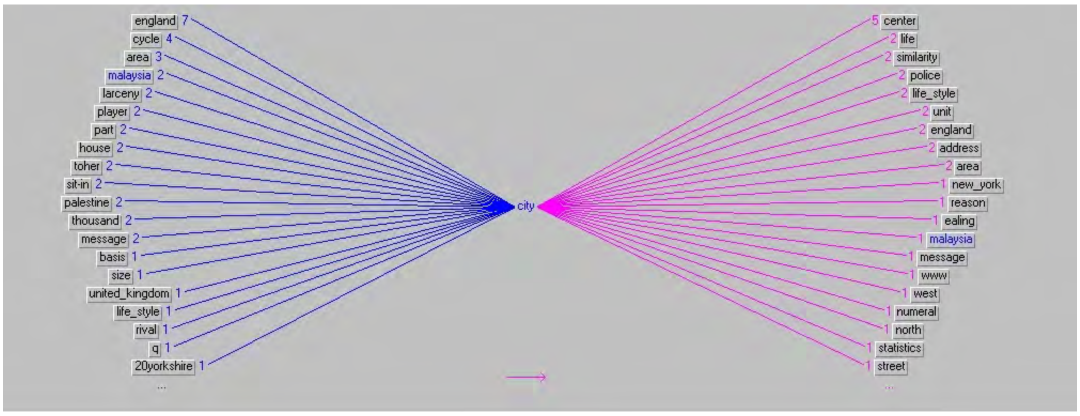


FIGURE 5.5: References to "City".

”cities and postcodes to check the crime rate http://www. police. uk/Cost wise, the South Ease and London is pretty expensive...”

”...and if you have a rough idea of the kind of lifestyle you would like city, countryside..”

”...As you can see Liverpool is roughly mid table behind most toher major cities. The anecdtoal evidence of your parents careers...”

”no fault of our own’ is no basis to make sweeping statements about an entire city...If anything the opposite is given true...”

”... and if you look at the size of the cities that are ahead of Liverpool they also have a much bigger population Once again all stats are from police. uk...”

”...it doesn’t get anymore Official also they are adjusted for population size as it says underneath...”

”... because they didn’t like the city for whatever reason and your parents deal with the absolute dregs of society on a daily basis...”

”...since you were old enough to stand you’ve had drummed it into you by the people closest to you how terrible the city is...”

”...couple that with a bad experience in the rowdiest bars in the city and bam it’s all been validated...”

”...and look up any city, post code, etc."Hi, I’m playing around with the http://www. police. uk/data...”

”... Wales and Northern Ireland as a favorite...You will love this city, i can nto say the same for weather i am afraid...”

”...i live now (Camden Town area) and you can have 2 beds house with garden...”

”...I was working in Camden Town (North West) and living in Ealing Broadway (West) and i was spending 2 hours for travelling every day...”

”...Ok, sometimes i really want to run away from busy city life to countryside but generally i always loved this city...”

”...Bad people every where you never know but i feel safe here more than many cities...”

”...cities. Also there are ltos of park inside the city when you would like to stay away from city life it helps...”

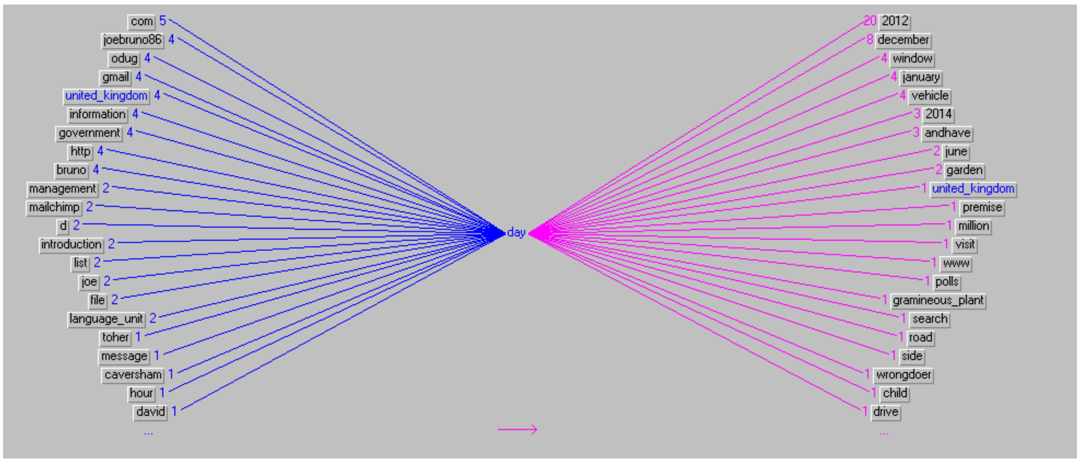


FIGURE 5.6: References to "Day".

> ”...and are close to the city centre and university. These areas are less suburban, less middle class...”
>
> ”...I think the university also has some accomodation in converted houses spread out throughout the town...”
>
> ”... Leicester is just about the same as any toher British city. The culture shock probably shouldn’t be too great...”
>
> ”...pick any area close to city and then pick any area further than city and compare crime rates...”
>
> ”...Large houses with big yards do exist in inner city at cheaper price and it’s usually very well hidden...”
>
> ”...272 crimes reported in October in such a small part of the city...”
>
> "...*Operation to tackle cycle theft a message from Oxford city centre unit*...”
>
> ”...*An operation in Oxford is working hard to tackle bike theft...”
>
> ”...people who handle stolen cycles in the city. Last year, there were 1, 955 offences of cycle theft in Oxford City...”
>
> ”...and people who handle stolen cycles in the city. Police have made 11 arrests-six people have been charged...”
>
> ”...A police officer from the Oxford City Centre Unit said:When your bike is stolen it can have a significant impact on your way of life...”

### 5.3.5 Day

Figure 5.8 shows Tropes’ categorisation of mentions of "day" or more generally, of time. In the text these all mostly correspond to mentions of a day of the week in a crime report, where crime reports have been released on the web, generally via Google Groups.

> 00 Saturday 08/12/2012 09: 00 Sunday 09/12/2012 HAVE GONE TO A METAL GARDEN SHED IN THE GARDEN OF A HOUSE.
>
> HOUSE. HAVE ATTEMPTED TO PRISE OPEN THE METAL SLIDING DOOR BY PULLING BACK THE TOP CORNERS OF THE DOOR.
>
> 00 Friday 07/12/2012 18: 30 Friday 07/12/2012 HAVE APPROACHED SECURE PEDAL CYCLE IN AN ENCLOSED AREAOF A CARPARK.

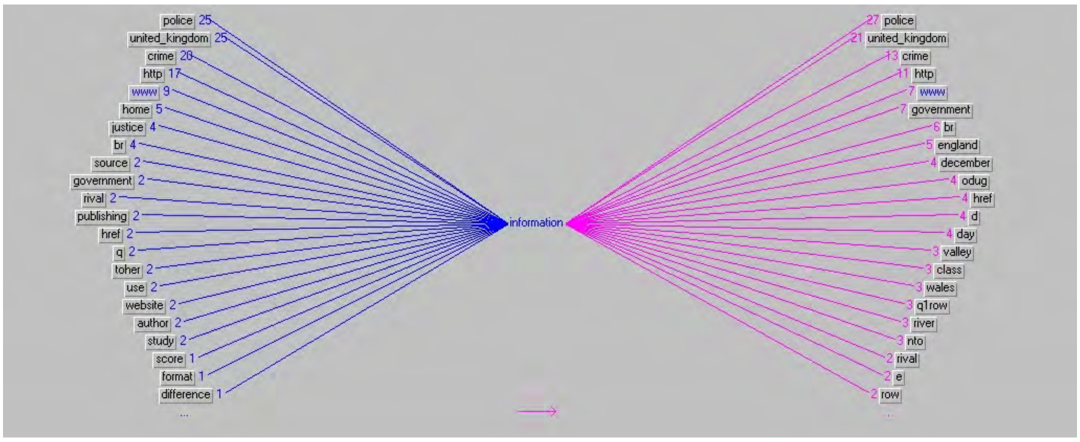


FIGURE 5.7: References to "Information".

> CARPARK. HAVE CUT CHAIN INTO 2 PIECES ANDLEFT WITH THE PEDAL CYCLE LEAVING THE BROKEN CHAIN BEHIND.
>
> 50 Saturday 08/12/2012 18: 00 Saturday 08/12/2012 HASEXITED AVEHICLE ANDAPPROACHED THE VEHICLE THAT WAS RIGHT BEHIND.
>
> BEHIND. HAVE JUMPED ON TO THE BONNET OF THIS CARANDKICKED THE WINDSCREEN CAUSING IT TO SHATTER.
>
> 50 Tuesday 11/12/2012 07: 15 Tuesday 11/12/2012 OFFENDERS HAVE ENTERED PREMISES BY FORCING THE FRONT DOOR ANDHAVE THEN CONDUCTED A SEARCH OF THE ROOMS BEFORE FINDING ANDTAKING A LAPTOP A COAT ANDA SPORTS BAG.
>
> BAG. GD/12/11078 THEFT FROM Mtoor VEHICLE LITTLE STREET GUILDFORD 19: 30 Tuesday 11/12/2012 22:
>
> 22: 00 Tuesday 11/12/2012 HAVE APPROACHED A PARKED VEHICLE ANDSMASHED THE FRONT WINDOW GAINING ENTRY.
>
> ENTRY. HAVE CARRIED THEN OUT A SEARCH OF THE VEHICLE. HAVE STOLEN FROM THE VEHICLE TWO PAIRS OF PRESCRIPTION SUNGLASSES Park Barn and Westborough GD/12/10963 THEFT FROM Mtoor VEHICLE APPLEGARTH AVENUE

### 5.3.6 Information

figure 5.9 shows Tropes' categorisation of references to information or knowledge.

We show some examples below of the references made:

> "...with new data published on the Government's crime mapping website police. uk The site now shows how crimes were dealt with..."
>
> "...We suggest that future focus should be on improving the quality and cartographic visualization of the published information alongside the integration of socialmedia functionality to enrich local dialogue on crime issues..."
>
> "...Hence, barriers to accurate information were one of the main reasons why the reassurance gap..."
>
> "...countervailing source of informationnationally reported crime statisticswas nto being heard...the message that crime levels were falling was nto getting through to the populace over the cacophony of competing information..."

″...Hence the move to publish crime statistics online using a crime map. Studies have inferred long thatpublically disseminating crime information engages the public...″

″...Increasing public access to crime information is seen as integral to this whole agenda. In addition, digital crime mapping was seen as akey mechanism for encouraging the public to take greater responsibility for holding the local police to account for their performance....″

″...In toher words, it is believed that publishing information on crime at a local level facilitates greater public scrutiny of how well the police are doing at suppressing local crime...″

″...or anecdtoal) that had measured the impact that publishing crime statistics had on improving the credibility of these data or the way in which the information was being used to inform, reassure, and engage with the public...″

″...derived knowledge available on the impact of crime maps on public perceptions of crime was generated in the USA, on a small and conveniently selected sample...″

″...Moreover, many practitioners hadconcerns with the geocoding accuracy of some crime data which could make the interpretation of street-level data misleading and confusing to the public...″

″...Furthermore,contemporary research has stressed that information provision needs to be relevant to the recipients...″

″...This also under-scores the need for tailored information that is actively passed on to local communities at times of heightened crime risk...″

″...This meant the police had no way to inform interested audiences with locally relevant crime information such as specific and tailored crime prevention advice regarding a known local crime issue...″

″...In conclusion, the authors question theassumption that all police-recorded crime data are fit for purpose for mapping at street level...″

″...They recommend using the Management of Police Information (MOPI) prtoocol, which states that information must fulfill a necessary purpose for it to be recorded...″

″...be recorded and retained by the police. MOPI wouldhelp to qualify what should and what should nto be published...″

″...Instead of mapping everything and anything, the authors advocate for the provision of better quality information that the public can actually do something with to minimize their risk of victimization...″

″...when the data it offers provides little that encourages the public to react.!Nto exactly an answer to your question, but you still might find your information useful! When looking for a flat, always ask which bills are included...″

″...if you want actual data-bear in mind this is only crime that gets reported though. Ntohing as bad as Moss Side in Sheffield to my knowledge person X and person Y pretty much covered it...″

″...Hi, I'm playing around with the http://www. police. uk/data. One of my colleagues wants a heat map to show incidence of crime in particular district council wards-from Mapit...″

″...The police neighbourhood data doesn't match up so my second thought was to use the lat/long incident data...″

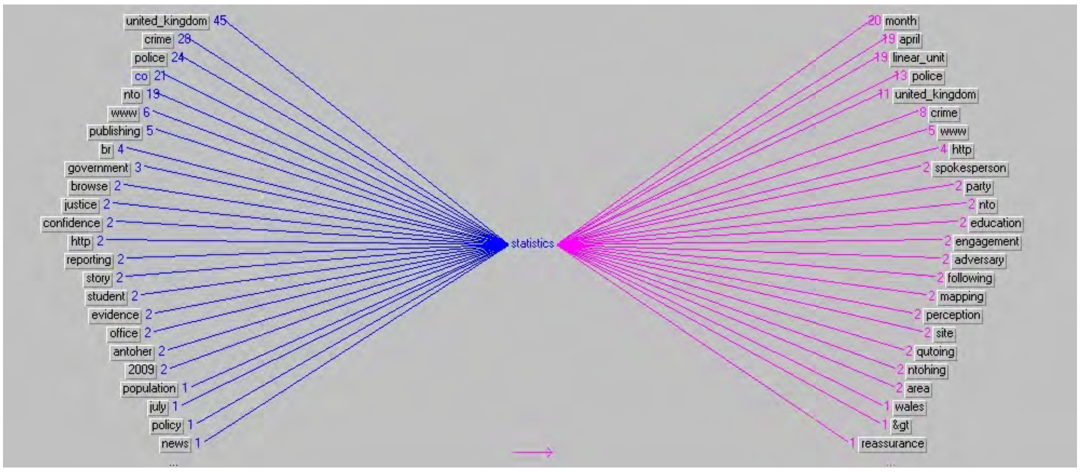


FIGURE 5.8: References to "Statistics".

″...At a very simple level I've been able to filter rectangles of data in Excel...″

″...in the past to extract data out of shapefiles, but I can't find any example code for you...″

″...I think the coming together of open data, apps and cloud to centralise, host and share it...″

″...So much can be achieved by letting people use mashups to use data for their own benefit...″

″...although reliability of this data is very questionable. As for what the people are like...″

″...all the useful fields have been added. I had to FOI for data for the last 5 years as bike theft is a brand new category as of June 24th...″

″...I guess this data really opens it up publicly how bad the problem is...″

### 5.3.7 Statistics

“Are those statistics from police. uk www. police. uk? That's a great site, as well as shocking for revealing the high number of crime in Bristol!”

“...police forces in the UK have published crime statistics using an online crime mapping tool. The drivers behind this were to help improve the credibility...”

“There continues to be a lack of evidence that publishing crime statistics using crime mapping actually supports improvements in community engagement and empowerment.”

“Publishing Crime Statistics using Online Crime Mapping. Insights from this study have important implications for crisis mapping projects.”

“The rationale for publishing up-to-date crime statistics online was to address the reassurance gap which relates to the counterintuitive relationship between fear of crime and the reality of crime.”

### 5.3.8 Remarks

The results were interesting in that there were marked differences between the ways in which concepts were derived. As we saw above, although the classifier found

“Place”, “Day”, “Police” and “U.K.” occurring - these had all actually come into play as a result of U.R.L.s, or of crime reports. “Day” occurred as a result of the Police putting some of their crime reports online. In contrast, where “City” was derived, it turned out that a lot of sentiment was attached to mentions of the city in a way that is reminiscent of Baudelaire’s flaneur, or of Latour writing about the city: “ Economics, sociology, water, electricity, telephony, voters, geography, the climate, sewers, rumours, metros, police surveillance, standards, sums and summaries: all these circulate in Paris, through the narrow corridors that can never be used as frames nor infrastructures nor contexts for others.” (Latour and Hermant, 1998).

Although one of our frames is to do with crimes being categorised as having a “place”, it turned out that the “U.K.” occurring so prevalently according to this machine classification is because of the “.uk” in “Police.uk” Our natural inclination was to then perform an analysis on text with the hyperlinks removed as being meaningful, but in fact what then emerged was that there is an important and strong notion of place and space in the “cyberspace”, which is of importance to the way in which concepts of crime should perhaps be built in the 21st Century. In fact the .uk of the U.R.L. is as much to do with markets and contested territories as a reference to the entity it appears to invoke, with D.N.S. activities being far more complex in terms of reference, location, infrastructure, policy, and politics than one might first imagine. Again, “Yes, there is a common world, full and whole existences, civilizations, but we have to agree to study how totalities are summed up in narrow temporary places where they paint their pictures; and then follow them in the worlds they perform – streets, corridors, squares, words, clichés, common places, standards –; and, finally, we have to agree to explore how these scattered totalities provide beings, themselves multiple and variable, with ways to gather themselves as coherent wholes.” (Idem).

We looked for evidence of the following debates:

- Debate about accuracy: of numbers of crimes, of reporting, of the “dark figure”.
- Debate about knowledge of crime versus fear of crime.
- Debate about the nature of crime itself.
- Debate about policy and policing – how do we use information to make changes?
- Debate about privacy and surveillance.

Findings were organised around four underlying interpretive frames that appeared to be influenced by organisational and socialstructures. These were: crime data used to persuade, crime data used to compete, crime data as knowledge, and collaborative crime data. These interpretive frames highlight the complex nature of crime data. They provide an enriched empirical basis for grounding conceptual models that drive research and practice. We found that although there were one or two mentions made of doubt over the figures, these were not really in relation to worries about the “Dark Figure”. There was very little mention of negative emotion caused by the figures, they were all found to sit within the data used to persuade, compete, as knowledge and collaborative frames.

### 5.3.9 Trust and Crime Data

Examining the context of some of the conversations, it seems there is trust in the figures. Most conversations occur in the context of when to call 101, how to work out what sort of area is lived in, and where to move to, and knowledge of crime as opposed to fear of crime. There are postings (the ones mentioned above with comparators in them) where risk is mentioned – however these do not tend to be discussed. The risk figures were distinctive - and seemed to misuse crime data to an extent that only becomes clear in the next analysis. They seem to derive in tone directly from the times when people started seeking reassurance with locks and keys - most of which, for many years symbolised safety, rather than being actually preventative.

Some of the quotes on the following table show crime statistics referred to with trust – that they answer a question, improve knowledge, add to an evidence base. There are a couple of cases where they are being referred to as tools for persuasion. They are referred to as having been misused, but there is little indication except perhaps in two cases that there is any scepticism about their provenance or use. In these cases we might find the competing frames of trust in numbers versus fear of crime. There is also a tendency to question the accuracy of the geocoding, over whether or not a crime took place. "Moreover, many practitioners had concerns with the geocoding accuracy of some crime data", as opposed to, "if you want actual data-bear in mind this is only crime that gets reported though. Ntohing as bad as Moss Side in Sheffield to my knowledge". Where there was discussion of figures in the context of police shootings, for example, it was notable that even on discussion forums such as Reddit, the discourse soon falls away over what appear to be debates about figures, but which soon dissolve into fairly robust criminological issues, while at the same time being ostensibly framed as issues to do with statistics.

It was also notable that some visitors to the site themselves are concerned enough about crime where they live to be indulging in possibly deviant or even criminal behaviours themselves – there is talk about how to arm oneself if living in a troubled area, and what steps to take for self-protection. Some of the advice given by those who are linking to Police.uk was about how to purchase ka-bar knives, for example. This ties into an interesting body of research about causes of war and crime, that we do not have scope to cover here. The primary conclusion is that there is no overt increase in fear of crime throughout the discussions, but that people's knowledge does seem to be improved, from their perspective. Of course we do not actually know whether that is the case. Does online etiquette inhibit people from publically questioning advice they are being given? Even if they don't question the statistics given to offset or augment their personal risk management, (supposing there could ever be such a thing in reality) does that mean the statistics are correct?

Is there any correlation between (so-called) personal risk factors and the degree to which one feels fear? How does one even quantify fear? When government policy talks about fear of crime, there is a weird corollary to this – as we suggested earlier, some statistics has an intended or unintended side-effect of being a surveillance – with another side-effect of controlling part of a population through uncertainty – the supposed object behind the Panopticon. If an observed population does not question the statistics it is given, does that mean observers should place a risk factor relating to ignorance, and a lack of desire to fully know their environment to certain populations?

### 5.3.10 Challenges

Further work is needed to understand which deeper questions to ask. Do people not really engage in having their opinions changed on social media? Will publishing crime numbers in certain ways never raise fear of crime? Or could the publication raise fear of crime, but not among a fairly online-savvy audience? There are meta-questions here – what is the knowledge that could be improved? And how can we measure this against fear of crime, which is subjective? There is a possible need here to examine research about the difference between a person's fear of crime versus the statistical likelihood of crime occurring to them illustrating their rationality. Does the opposite – a statistical mishmash of nonsense being used to illustrate what someone should or should not fear about risk -indicate a lack of rationality on a state-wide level that undoes the whole transparency program? We can start to understand that crime is a contested concept that takes on different meanings in different spatio-temporal contexts.

We should consider methodologically, policy frame analysis, the study of how "public policies rest on frames that supply them with underlying structures of beliefs perceptions, and appreciation" (Rein and Schön, 1996). Although the concept of frame analysis goes back to Goffman (Goffman, 1995), and Snow, (David A. Snow; E. Burke Rochford, 1986), its introduction to the field of policy analysis can be attributed to Rein and Schön. Verloo defines a policy frame as an "organizing principle that transforms fragmentary or incidental information into a structured and meaningful problem, in which a solution is implicitly or explicitly included." (Verloo, 2005).

Frames can operate on various levels: at the macro level (such as for a whole society, nation or even on the supranational level), at the meso level (such as for type of actors or policy domain), as well as at the micro level when referring to framing processes by individual actors. The macro-meso-micro frame analysis takes policy from abstract idea to action. It locates differences between expression of the policy or between intent and outcome at a local (micro)implementation level that should help us to understand how effective policy is, and what factors are at play that might change its expression. Analysis at the macro (transparency, transparent crime data), meso (transparency in crime data geographically mediated by the web on Police.uk) and micro (individually made comments from visitors) levels seems to indicate a complete lack of coherence and understanding of what crime data is and what it should be at policy level. There is a question of intentionality - whether frames should be considered the results of practices involving the strategic deployments of certain arguments to influence decision-making or rather unintentional and unconscious acceptance of public discourses, when we seek to illustrate them in our big-data-esque program that counterpointed Home Office's qualitative research. It is here that we suggest that while there have been various policy debates around efficiency, public value theory and new public management, the most useful approach to consider where policy is mediated by the organic structure that is the World Wide Web, is one relating to complex adaptive systems.

As noted in (Byrne-Evans and Task, 2013), it can be more useful to describe behaviours rather than intentions when dealing with the large scale complexity engendered by publishing crime data on the web, as a result of policy. This then helps resolve the question of normativity: while some have argued for the separation of cognitive and normative aspects and the reservation of the concept of frames to the former; others call attention to the inherent inseparability of the two. (Surel, 2000). We can also ask whether crime is always the same thing? Can we count it and locate it in time and space? Where the discourses fall down seems to be over statistics, but a deeper examination shows criminological discussion emerging from complex structures of argumentative frames. We can also ask what lies outside the statistics we have on crime? What are the details? Who has them? What lies behind the narratives of crime and policing? Why are they not told? We have already suggested that "crime data" and "crime statistics" are in fact policing performance data - does that mean we have no real knowledge of crime? The police and their allied and competing organisations that we found in our first analysis, such as Trading Standards, H.M.R.C., M.O.D. and G.C.H.Q. have knowledge of crime, even if it is not publically available.

One (current) potential problem for transparency is that with increasing technological mediation of crime, and thus a claim being made of its being cybercrime, much of this knowledge will be held by Action Fraud, the N.C.A., the City of London Police and the Security Services. This is a problem for crime data, which we go into in the next analysis. Its background at the time of writing is that although the N.C.A. tends to focus on cyber-dependent crime as opposed to cyber-enabled, there is a contested space here over a boundary earlier identified – the point at which knowledge of crime becomes a crime in itself. Economic crimes are successful when they cause participants within a market to fail, or the marketplace itself, therefore the area where cyber-dependent crime is in play shows clearly where transparency also fails as a democratic device – as first seen in response to Roosevelt's attack on the "tyranny of mere wealth".

Can we use this analysis to get a more coherent view of what crime maps mean for ordinary users, whether they increase fear of crime and how to facilitate working between police, public, Home Office and the media? Can we use these comments on crime mapping to understand better how to reframe issues for media consumption? How do we start to understand the dividing line between society's responsibility for crime (big society debates) and the responsibility of individuals (criminals – is this really a population to be explored?) and of officials, the Police and the Home Office. What responsibility does the media have for increasing fear of crime?

"What...does the policy researcher analyse? Precisely what is public policy?...some writers have simply understood policy to be 'whatever governments choose to do or not to do'. Others have worked out elaborate definitions that seek to spell out the exact characteristic of a public policy. Lowi and Ginsberg, for example, define public policy as "an officially expressed intention backed by a sanction, which can be a reward or punishment." (Fischer, 2003, p.1). While good web designers know that it is important to think about social, economic and psychological factors, our findings point towards this being crucial if we are to understand how best to leverage open crime data, using technology. Otherwise there is the risk mentioned by Berlin: "That is why those who put their faith in some immense, world-transforming phenomenon, like the final triumph of reason or the proletarian revolution, must believe that all political and moral problems can thereby be turned into technological ones. That is the meaning of Engels' famous phrase(paraphrasing Saint-Simon) about 'replacing the government of persons by the administration of things', and the Marxist prophecies about the withering away of the State and the beginning of the true history of humanity." (Berlin, Hardy, and Hausheer, 1998, introduction).

We can start to think about crime literacy – not just the global syntax of crime and policing provided by the Police.uk statistics and trends, but the local understandings or semantics of crime data – what does it mean if there is a particular crime committed in a particular spot every 30 days? And what does it mean if the crime ceases to be committed? Is the criminal or deviant now prevented from acting (and therefore perhaps more frustrated?) or has she moved elsewhere? What of notions of governmentality and risk? Risk is embedded in the notion of "caveat emptor" – or "let the buyer beware", which seem antithetical to notions of transparency within democracy, unless we also attach this notion to that of people being free to understand, and being responsible for their own understanding of what it is they are getting or seeing when they buy into crime data. "New ideas have emerged from late modern theorising to do with risk, identity and emotion. . . psychology needs to become more fully aware of the politics of its science." (Webber, 2009).

The meanings of crime are embedded in local knowledge, then in less concrete ways and more epistemologically in the networks that give the local meaning: in political and community structures, and beyond these, in cognitive structures that determine how we view ourselves in relation to the places that we move through and where we experience or participate in crime and deviance. Starting to understand these, using the methods suggested, and what the knowledge of crime means for data users again using suggested analyses, is likely to further strengthen attempts to use crime data both as part of an accountability and transparency programme, and in the knowledge economy as part of the current open innovation strategy. This approach should also support the attempt to create cogent policy that is predicated around our more technologically innovative society as a result of a deeper understanding of crime as it is experienced by networked communities in the U.K. through the use of open data in the U.K. government's transparency programme and other data, where it might be needed to aid such transparency.

# Chapter 6

# Semi-Structured Interviews Relating to Police.uk

## 6.1 The Crime, Policing and Data Interview Subjects

### 6.1.1 Illuminating Crime Data

As laid out in the methodology we have created a crime context ontology or classification to understand the context of Police.uk in the security, policing and intelligence markets. We then used a semantic classifier to analyse themes emerging in the point of departure comments made in response to users interacting with Police.uk on social media. We looked for frames that showed contested spaces, as well as the more general themes. We then used these results from the Thematic/Frame Analysis to apply frames to interviews on the impact of crime data, as viewed on Police.uk.

Between 2011 and 2014 I interviewed in-depth 20 members of the crime, policing, security and transparency communities. Most of these interviewees have expertise in one aspect or another of in Web-mediated data, crime, security or transparency. We used non-probabilistic snowball sampling in order to first gather these interviewees and the social networks of initial actors. The analysis was carried out using the frames culled from analysis of O.L.S.N.s and the W.O.D. – to see whether the frames compete. I also carried out less formal opinion-gathering from communities of crime, security and intelligence professionals, that in the end, gave me around another two hundred and fifty or more sets of opinions about policy, crime and security that fed into my background understanding.

### 6.1.2 Reflective Commentary

Central to the qualitative empirical approach to interviewing is reflexive interpretation. Alvesson and Skoldberg have cited the importance of "the open play of reflection across various levels of interpretation... There is no one-way street between the researcher and the object of study; rather, the two affect each other mutually and continually in the course of the research process." (Alvesson, M. & Skoldberg, 2010, p.39).

This was especially true, given the nature of the work of my interviewees. Some of the reasons behind this are examined in depth in this work – many of my interviewees had severe organisational constraints upon them: both on their time and in respect to what they are able to talk about. Law enforcement also operate under a number of competing priorities – while they have to gain trust from the public, at the same time they must maintain a healthy scepticism about the motivations behind people wanting to talk to them; it can be hard to gain the trust of individuals working in such difficult circumstances. I started out as a "member of the public", or "researcher", but over time, developed an identity, as a result of my employment as an individual on the outskirts of the communities I was interviewing, which then simultaneously helped more people to be able to talk to me, but created more constraint upon myself, as I became in turn less able to write about what I learned, or had to operate myself within their constraints, which became my own.

I first immersed myself in a number of commercial, policing, security and intelligence enterprises in order to get a feel for the everyday routines of policing and how work is done. At the same time, while learning about various sorts of crimes and reflecting on them I gradually became aware that I myself was a victim of crime, which added yet another layer to the multiple layers of meanings that I was gathering from my interviews and daily life, as I was interacting with police professionally, personally and socially. This was also very useful in gaining a feel for how policing pluralisation actually manifests. Although I formally interviewed 20 subjects, as said above, in fact I spent over four years talking intensively to more than two hundred and fifty members of the policing, law enforcement and security community, in addition to those interviewed formally, in order to pick up the “feel” of policing, and to understand how knowledge is exchanged among the various sub-communities and between police or L.E.O.s, security, academia, government and industry.

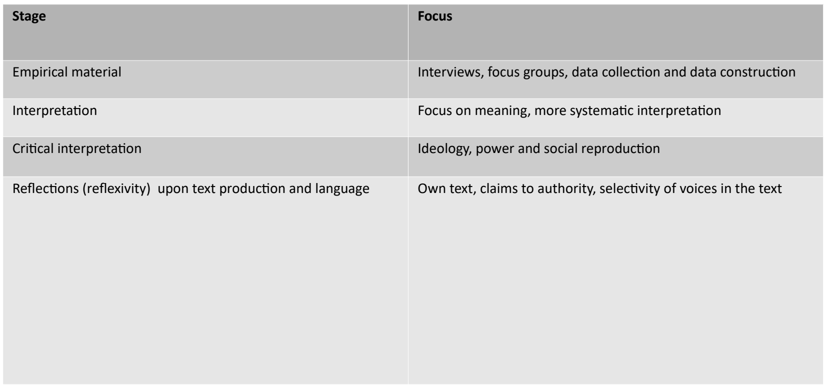
One interesting outcome was work done on a White Paper for I.B.M. on the nature of cybercrime and the best methods to address it, (Anning et al., 2016), and advising on the setting up of a contract between the Police, I.B.M. and the Home Office. This enterprise is in fact one of the manifestations of pluralised outsourced policing, often referred to in the media.

The “researcher and object are involved in a common context.” “Interpretation implies that there are no self evident, simple or unambiguous rules or procedures, and that crucial ingredients in the research process are the researcher’s judgement, intuition, ability to ‘see and point something out’, as well as the consideration of a more or less explicit dialogue - with the research subject, with aspects of the researcher herself that are not entrenched behind a research position, and the reader.” (Alvesson, M. & Skoldberg, 2010). Given my immersion into these new (for me) worlds, I tried to be aware of what I was bringing to my interpretation and discovery of themes. I have in the past worked as a systems analyst and designer, problem-solver and business analyst, as well as having trained as a therapist and worked in the field of knowledge elicitation and Artificial Intelligence. There were times when the interviews felt as though they were therapeutic sessions, bringing to mind the character in Robert Stross’s Rule 34 who acts as a psychoanalyst for companies, looking for signs of deviant corporate behaviours that act as precursory signals of a company’s collapse. (Stross, 2012). My involvement with frames became more complex, as by the time the thesis was being written up, I was exploring (professionally) their use in Artificial Intelligence in Defence and Information Operations with a cyber and geo-locationary aspect.

## 6.2 Analysis and Themes

### 6.2.1 Analysis of all Interviews

As indicated earlier, Tropes pruduces structural concepts through “References which represent context and group together the main substantives of the text analysed into Equivalent classes.” “The References and of their Relations brings you to the heart of the discourse: all the actors, objects, things and concepts presented in the text will appear before you in decreasing order of importance.”



| **Stage** | **Focus** |
|---|---|
| Empirical material | Interviews, focus groups, data collection and data construction |
| Interpretation | Focus on meaning, more systematic interpretation |
| Critical interpretation | Ideology, power and social reproduction |
| Reflections (reflexivity) upon text production and language | Own text, claims to authority, selectivity of voices in the text |

FIGURE 6.1: Levels of interpretation according to Alvesson and Sköldberg (2000, p. 250).

.

The interviews were semi-structured interviews, carried out in order to understand some of the themes of the lived daily world from the subjects' own perspectives. These interviews were conversational and discursive in nature but also had a purpose, and involved a specific approach and technique, so were neither everyday conversation or closed questionnaire.

All interviewees were contacted prior to the interviews and informed about the objective and scope of the interviews. Although it was hard to gain access to some of the people interviewed, once they had agreed to be interviewed none seemed concerned about the subject matter although a few seemed initially slightly worried about giving "wrong" answers, or their performance. Reflexive practice made me think about the pragmatics of interviewing; a rational critical realist approach might assume that it is for the purpose of the interviewer gaining access to or teasing out highly-valued but hard to get hold of facts held in the minds of experts, but in practice, the experience was of co-constructing. Subjects were often hard to actually get hold of, but once we were talking, they often seemed not to be concerned about restrictions on what they could say, and a confessional tone often emerged, which I then also had to manage during the interview for the purposes of not exposing interviewees to any harm.

The interviews lasted around 40 minutes – one or two were about an hour long, and were recorded on a recording device on a phone, visible, but not a forced focal point, to the interviewee. Some subjects wanted to come back and explore what we had talked about in more depth. The conversation was assisted by the crime maps of Police.uk displayed either on laptop or tablet. This meant that the sort of interface was not an experimental control in any way. What was seen varied, according to the sort of interface and so there was often discussion about the mediative element of the device, which served to settle the interviewees - I encouraged them to explore and to react to what they saw or experienced, and tried to use these beginning interactions to help them to feel that I wanted genuinely to know about their responses, and that they did not have to structure these in a "right" way, or to "be clever". The interviews had common starting points and then were open in that they allowed for interviewees to feel their own stories emerging and divert to themes that concerned them. Follow-up was made a few times.

The interviews were transcribed word by word. Because of the multi-disciplinary nature of the work, and the need to transcend several disciplines, attention was paid to pauses, interjections and interruptions, although not to the extent that might be in a discourse analysis concerned more with structure than content. I also did the transcription myself, wishing to re-experience the interview as I transcribed it. There were several institutional, peer and law enforcement attempts made to persuade me to do more interviews and have the transcription done elsewhere, (as the Ph.D. has time constraints) but after reflection, I decided I would be happier with less material (it served to balance out other research material in any case), to very thoroughly explore the material I had, and the meaning of the relationships with Law Enforcement and other professionals it had given me, and to maintain the trust that was placed in me by the interviewees. All interviews were done in English and transcribed in English.

### 6.2.2 Analysis of Frames

We included interviews, some policy documents an interview given to the media and talks given to organisations as beckground material to the development of this third analysis; however for creating the second ontology we only used the interviews that we ourselves had conducted. While the analysis carried out in the previous chapter was also of frames, we cover the concepts in more detail here, as the previous anysis was more impersonal, with results being taken from social media.

Where we refer to frames, it is important to understand these as Goffman originally intended. There is a difference between media “framing” i.e. conscious distortion of facts and Goffman’s original unconscious adoptions. “I assume that definitions of a situation are built up in accordance with principals of organization which govern events [. . . ] and our subjective involvement in them; frame is the word I use to refer to such of these basic elements as I am able to identify.” (Goffman, 1956, pp.10-11). The concept of frames is used to explore actors’ tacit understandings of the world, including their different perceptions of the crime that leads to crime data and the crime that does not.

“Frames are principles of selection, emphasis and presentation composed of little tacit theories about what exists, what happens, and what matters." (Gitlin, 1980, p.6).

We gained an understanding of some of the politics of policy playing itself out, and what actors might do to shape policy outcomes in ways that suit their crime-fighting activities. We also found that in the world of pluralised law enforcement there is little generalised trust shared among the different actors in each other’s enterprises and abilities. Where trust did come in to play, it had been earned and usually arose from working closely with the enterprise or actor concerned.

Frames organise experience, “through creating an active perspective that both describes and perceptually changes a given situation”. They accept the bias of experience and by implicitly referencing “non-objective ways of considering ideas or situation, but they are shaped over the long-term aggregation of thoughts and experiences.” (Kolko, 2010).

Frames are identified, characterised, grouped and analysed, as a heuristic means of understanding conceptual dividing lines between the perspectives and positions of the actors. The contested and often moving nature of these frames is noted, especially where we bring it in to later analyses of the construction of cybercrime via intelligence agencies and security companies.

### 6.2.3 Categories

I started by transcribing the interviews and then running the text through Tropes, in order to get an overview of the concepts that were found. These are shown in 6.2.

I then started going into each category and ordering results from the interviews into rough categories. This was a highly interative process and took many weeks. The figure in 6.3 shows the beginning work. Frames of different actors have been analysed and organised into groups based on their content. Frames with similar basic themes, crime categorisations, cyber concepts and place perceptions, understandings of the policy process, and actions being taken are grouped together. These frames were derived directly from the classifier’s groupings. These frames are the researcher’s own understandings of (unordered) concepts that come forward in her analysis of the interviews:

### 6.2.4 Analysis Carried out Using Both Sets of Frames

Analysing actors’ frames revealed some very contrasting perceptions of the effects of policy, as well as crime and place among the actors. Among people who had left or were on the verge of leaving the police or military, the same events or phenomena

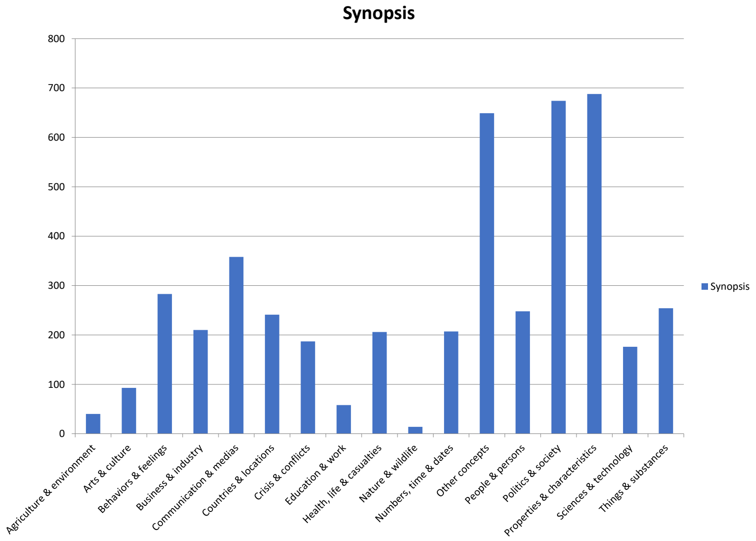


FIGURE 6.2: The Concepts

.

were described in stronger emotional terms. An example is provided by L.E.O.s referring to their bosses or people writing policy.

> "they're all tossers"
>
> "they don't have an effing clue"
>
> "all they want is to look good on paper and they don't care if they eff up their departments and people's lives"
>
> "my own business is making people feel totally shit about themselves in order to sell"
>
> One interviewee said of an initiative he had proposed some years ago, *"so I came up with this plan ages ago, two-three years, and the DCI back then he didn't get it or understand it and said it's never going to work, now he's proposing the same thing, going to take all the credit, and he's only doing it to try to get a promotion, not for the right reasons. Some of us do this stuff because it's the right thing to do."* Some actors referred to those in control of writing policy or structuring departments to deal with crime, as "aggressive, clueless and autocratic", while others (fewer, and tending to have less experience) saw these attempts as an endeavour to "fix problems" and provide clearcut guidance and changes in order to improve crime figures.
>
> We arrived at three groups of frames: Group 1 is called "Crime data for People". It includes a cluster of frames with a shared theme emphasising the need for effective resource use, including time, money and technology for the benefit of society and locally and nationally. Opposing perspectives are typical for the frames included in Group 2, "Crime data for Policing" with an emphasis on performance. This group of frames share a common theme stressing the need to increase crime prevention in order to maintain a healthy society, in the UK, Europe and globally. Group 3, "Crime Prevention and Risk" emerges strongly relating to place and personally identifying data.

### 6.2.5 The Ontology

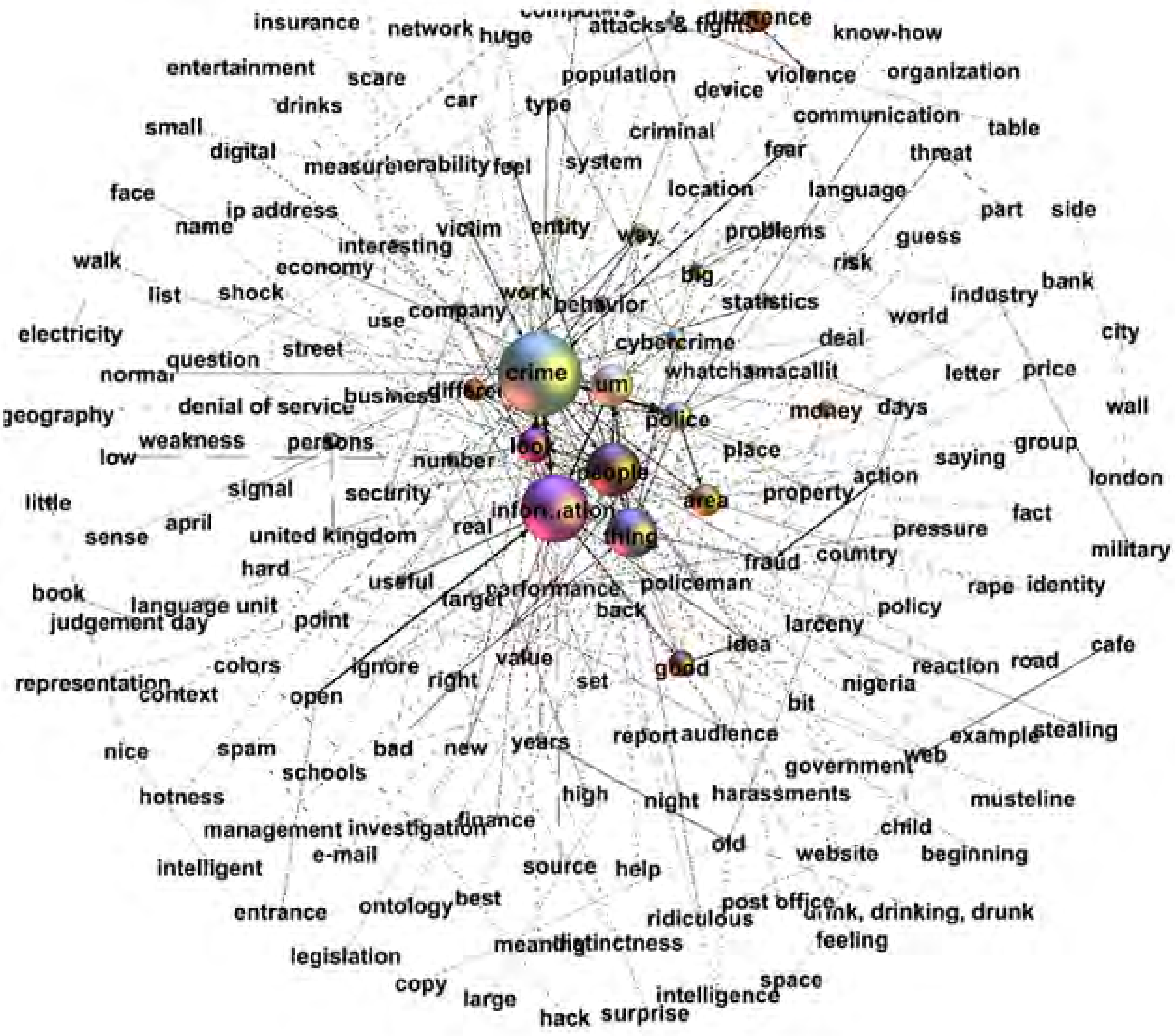


Figure 6.5: The Interviews Ontology shows a general representation of how the concepts derived by Tropes map to its internal representation of the world. The underlying representation of the world would, with time, be modified so as to align more of the data and knowledge generated by our analyses, however this was not within scope of this piece of work.

## 6.3 Results

As with the previous analysis we are wary of over-interpreting the results given from the classifier. We can see what are the objects of concern in the texts derived from our interviews. Mentions are of words that map to: crime, information, people, police, area, cybercrime, numbers, maps and company (to take the most highly occurring concepts in decreasing order of occurrence). The ontology, even in its infancy, shows that experts have a whole area of concern relating to cybercrime that pulls in concepts that would map to the way in which we could usefully open up discussions about reconceptualising crime.

Our first draft that produced the frames in Figure 6.3 starts to reveal a concern with the idea of spatial mapping, with the politics of creating crime data and with problems relating to policing performance and perverse incentives that mean there is very little reason for an officer to ensure that he or she records all instances of crime they come across. Given that interviewees were directly asked about some of these concerns this is not to say that the "naturally react-ing" commenters on social media linking to police.uk might not have the same concerns were they to be asked, but that ordinary conversations about crime data, that Home Office could reasonably use to illustrate impact from their own analytics seem very assured about trust in the data and use of time and place on maps. There are no concerns found about cybercrime for example.

### 6.3.1 Content and Frames:

The end results showed the relationships between conceptualised frames, policy making and a contribution to the construction of place and crime in crime maps. We now look at some of the points brought out by our interviewees that were not so apparent in our scanning of web-mediated reactions to Police.uk. First, the official statistics (that inform Police.uk) are incomplete.

*"...where performance pressures cause officers to under record crime we're left with an incomplete picture of even the crime that we know about, because it was recorded to us it was reported to us, so consequently there's an incomplete picture on which to base decision-making and make operational resourcing decisions about what to do about it because you're at risk of not identifying of what could be quite a stark pattern.."*

Where decision-making is involved – especially concerning the provision of resources to fight crime, this is very serious. Aside from the problems of allowing people to determine the level of risk that they might face in a particular place or space, there are just as serious consequences for the police – and long-term, these consequences affect the public, as police are paid by our taxes, and if performance management plus under-resourcing combines to make them feel ill-suited for their jobs, stressed or suffering from trauma, then we have an even more poorly resourced means of protection, while our taxes are going to waste. Second, if we look at incentives for reporting crime, we realise there are huge swathes of crime that go publically officially unreported. They are known about and there is an informal market of discussion – that is hugely profitable – but a large sector of crime is cybercrime enacted against large corporations who live or die by their ability to maintain reputation and who therefore are bound professionally to keep quiet about threats.

*"...if there's a wobble in the city and the pendulum of greed, and the pendulum in the city swing between greed and fear and it swung towards fear then everyone's' going to ditch some of their stock cos they think you're a leaking ship..."*

Third, understanding causes of crime and its precursors becomes quite complex in the digital or web-mediated versions of crime.

*"...the bit where both digital meet analogue that's where we fit in so you've got a digital side which is cybercrime all to do with ones and zeros and firewalls and it has you know you've got the physical side which is locks and safes and walls and manguarding and c.c.d. and access control and . . . the gap in the middle which I'm going to call pre-data, it's information that hasn't actually been turned into ones and zeros and that isn't protected by encryption and firewalls etc. and that pre-data that's the bit the really very vulnerable"*

Security organisations and the police are struggling to categorise this sort of phenomenon and how to deal with it. They are not sure what they should be protecting – are there physical objects or digital ones, or are there objects out there susceptible to theft, that shapeshift, intellectual property that goes from a diffuse analogue verbal or even thought form (meatspace?) to digital manifestation that is vulnerable in the digital world? How do they place a value on potential loss and therefore, risk?

*"...is it worth the difference is it worth ten bucks is it worth a billion you'd probably know the difference in one or the other so you could probably whittle it down to an estimate, but even an estimated guess would be something to go on and it should start running some er statistics"*

Data and value – how do we work out the value of data, in order to assess the impact of cyber (or any) crime? There appears to be a kinetic or potential energy attached to data. Value lies in meaning and meaning is personal. There are complex movements in O.G.D. which push this back and forth. How do we work out the value of crime data, who is the audience and who should see it? How much should anyone see?

*"I think it's very different like a traditional crime map and looking at a cybercrime if you looked at a traditional crime map where you might have, let's say you have 50 assaults on a main road and you look more closely and you saw that main road and main drinking road in Manchester on a night out and most of Salford's out between 11.30 and 3.00 you know how to figure out how to stop that you just put lots of cops there and security guard, bouncers I mean, or if you saw some sexual assaults down a dark alley in a where people walk from a canal to a road and there's been 5 of these and burglaries and they break a bit tricky cos they can move around although you do tend to find they correspond with known burglars being released from prison so that's not hard to figure out vehicle theft, civil, lights go out tend to be quite simple.."*

Is it of any value to map cybercrime? The "pewpew" maps (especially those possibly artificially constructed by Norsecorp and devoured by the public) and the G.I.S. school would seem to say so, while others think not:

*"...what sort of crimes are actually useful on a map cos there is definitely certain kinds of crimes which are not useful when put onto a map like a D.D.o.S. or something like that you know the geographical location of that the chances are that actual geography is coming is not even coming into the question you if someone on the other side of the country has a website that they don't like and they D.D.o.S. it at no point has any thought ever been paid to the geography or anything like that."*

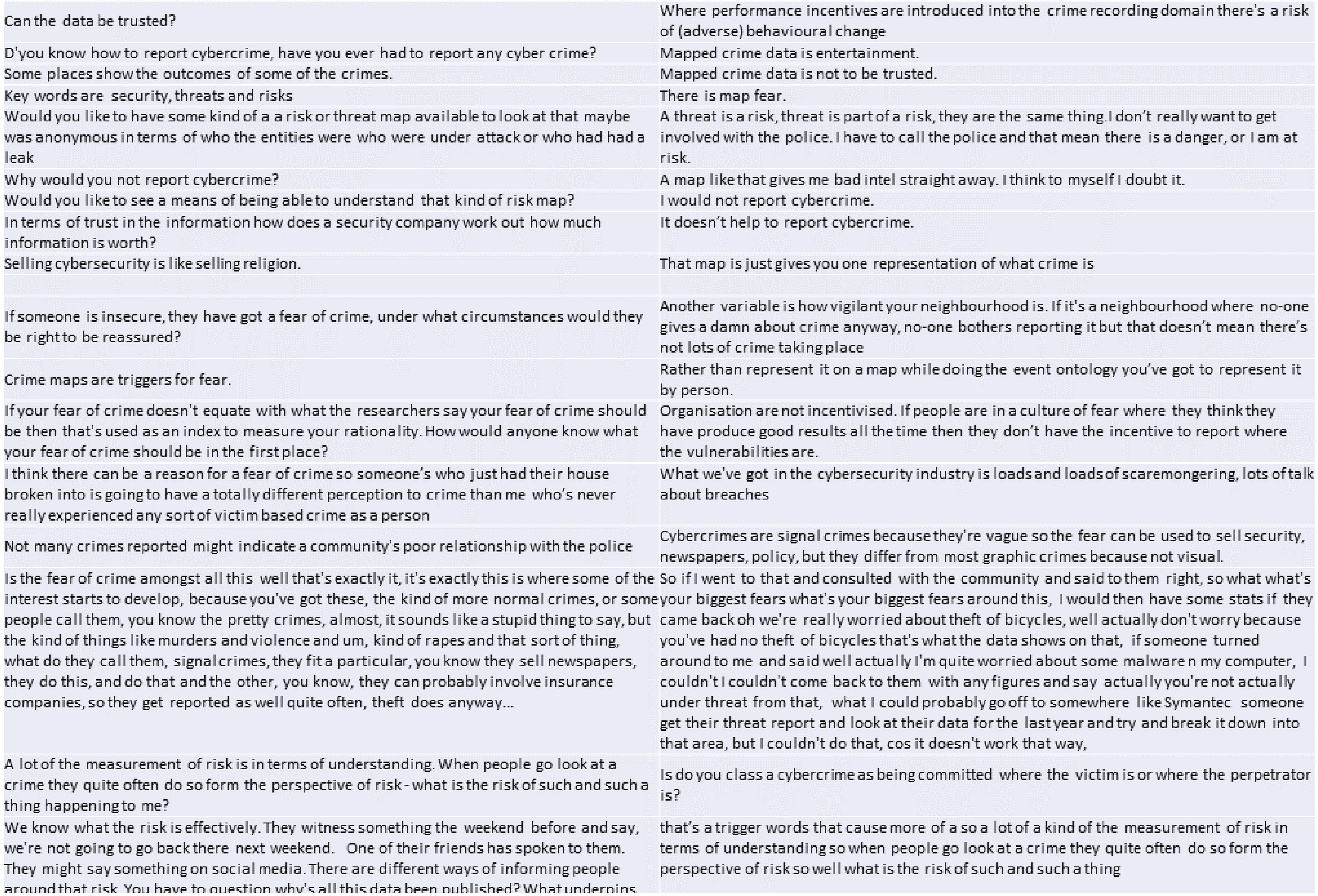
Is it possible that without victim-blaming, if institutions were to have numbers attached to them that represented data leaks, we might get some indication of their vulnerability? And therefore whether to invest or trust them with our data? What do the numbers mean anyway? How can a number tell us anything meaningful about risk?

*"...that number means that I'm in a bad place sort of a thing or I'm in a good place but then there's questions about well then they might feel better off or worse about that there's questions about are they invalid feelings if they're founded on totally misunderstood data and misunderstanding of data basically."*

Who controls the meaning of the numbers? *"...if your fear of crime doesn't equate to what the researchers say your fear of crime should be then that's actually used as an index to measure your rationality which is just like such a weird spin because you think well how would anyone know what your fear of crime should be anyway in the first place?"*

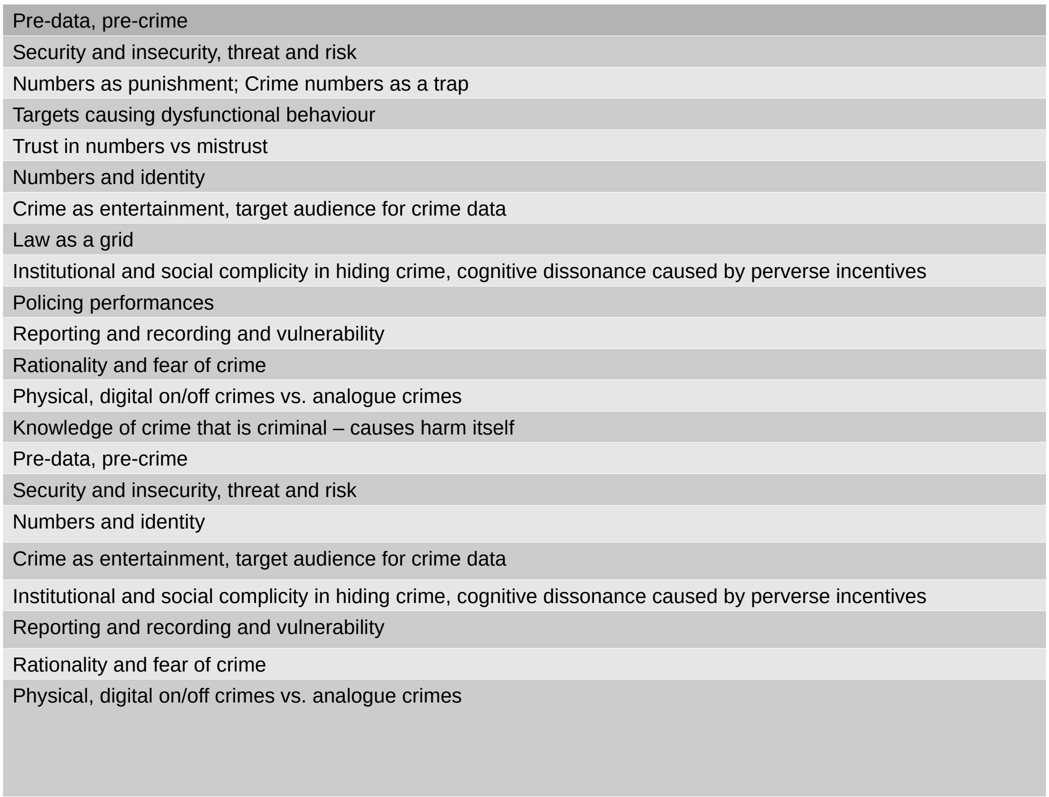
There is a misunderstanding about what crime on a map means – the fact it's been reported might mean it's a "normal crime" – the scariest, most deadly or prevalent, insidious ones escape reporting – this has an impact on using maps to represent risk. *"...a heat map where the blobs get bigger and smaller over time you can slide the thing back and forth over time to see if the crime reporting is improving or reducing"*
How do we construct crimes? Can we make new ontologies for them with new properties? Cybercrime is a crime that could potentially be mapped, and theft is occurring at a data level for example – on a vast scale. *"...but the reality is, and if you're going get down to a granular level on whether you're a victim of crime or not, you've had your data stolen and if you haven't the I.S.P. has, cos say well I don't pay extra for that so the I.S.P. had its data stolen so it's a victim, you've paid more electricity because your computer, your productivity either at home or your personal life fell off, because your computer's slower and you're a victim of the Computer Misuse Act but you - but no-one would ever take that report, and if you amplify, magnify that however many hundreds of thousands or millions of people that's happening to, you know, in a developed world every day, which it pretty much is, it's a huge problem that no-one's addressing..."*

.



| | |
|---|---|
| Can the data be trusted? | Where performance incentives are introduced into the crime recording domain there's a risk of (adverse) behavioural change |
| D'you know how to report cybercrime, have you ever had to report any cyber crime? | Mapped crime data is entertainment. |
| Some places show the outcomes of some of the crimes. | Mapped crime data is not to be trusted. |
| Key words are security, threats and risks | There is map fear. |
| Would you like to have some kind of a a risk or threat map available to look at that maybe was anonymous in terms of who the entities were who were under attack or who had had a leak | A threat is a risk, threat is part of a risk, they are the same thing.I don't really want to get involved with the police. I have to call the police and that mean there is a danger, or I am at risk. |
| Why would you not report cybercrime? | A map like that gives me bad intel straight away. I think to myself I doubt it. |
| Would you like to see a means of being able to understand that kind of risk map? | I would not report cybercrime. |
| In terms of trust in the information how does a security company work out how much information is worth? | It doesn't help to report cybercrime. |
| Selling cybersecurity is like selling religion. | That map is just gives you one representation of what crime is |
| | |
| If someone is insecure, they have got a fear of crime, under what circumstances would they be right to be reassured? | Another variable is how vigilant your neighbourhood is. If it's a neighbourhood where no-one gives a damn about crime anyway, no-one bothers reporting it but that doesn't mean there's not lots of crime taking place |
| Crime maps are triggers for fear. | Rather than represent it on a map while doing the event ontology you've got to represent it by person. |
| If your fear of crime doesn't equate with what the researchers say your fear of crime should be then that's used as an index to measure your rationality. How would anyone know what your fear of crime should be in the first place? | Organisation are not incentivised. If people are in a culture of fear where they think they have produce good results all the time then they don't have the incentive to report where the vulnerabilities are. |
| I think there can be a reason for a fear of crime so someone's who just had their house broken into is going to have a totally different perception to crime than me who's never really experienced any sort of victim based crime as a person | What we've got in the cybersecurity industry is loads and loads of scaremongering, lots of talk about breaches |
| Not many crimes reported might indicate a community's poor relationship with the police | Cybercrimes are signal crimes because they're vague so the fear can be used to sell security, newspapers, policy, but they differ from most graphic crimes because not visual. |
| Is the fear of crime amongst all this well that's exactly it, it's exactly this is where some of the interest starts to develop, because you've got these, the kind of more normal crimes, or some people call them, you know the pretty crimes, almost, it sounds like a stupid thing to say, but the kind of things like murders and violence and um, kind of rapes and that sort of thing, what do they call them, signal crimes, they fit a particular, you know they sell newspapers, they do this, and do that and the other, you know, they can probably involve insurance companies, so they get reported as well quite often, theft does anyway... | So if I went to that and consulted with the community and said to them right, so what what's your biggest fears what's your biggest fears around this, I would then have some stats if they came back oh we're really worried about theft of bicycles, well actually don't worry because you've had no theft of bicycles that's what the data shows on that, if someone turned around to me and said well actually I'm quite worried about some malware n my computer, I couldn't I couldn't come back to them with any figures and say actually you're not actually under threat from that, what I could probably go off to somewhere like Symantec someone get their threat report and look at their data for the last year and try and break it down into that area, but I couldn't do that, cos it doesn't work that way, |
| A lot of the measurement of risk is in terms of understanding. When people go look at a crime they quite often do so form the perspective of risk - what is the risk of such and such a thing happening to me? | Is do you class a cybercrime as being committed where the victim is or where the perpetrator is? |
| We know what the risk is effectively. They witness something the weekend before and say, we're not going to go back there next weekend. One of their friends has spoken to them. They might say something on social media. There are different ways of informing people around that risk You have to question why's all this data been published? What underpins | that's a trigger words that cause more of a so a lot of a kind of the measurement of risk in terms of understanding so when people go look at a crime they quite often do so form the perspective of risk so well what is the risk of such and such a thing |

FIGURE 6.3: Beginning the Process



| |
|---|
| Pre-data, pre-crime |
| Security and insecurity, threat and risk |
| Numbers as punishment; Crime numbers as a trap |
| Targets causing dysfunctional behaviour |
| Trust in numbers vs mistrust |
| Numbers and identity |
| Crime as entertainment, target audience for crime data |
| Law as a grid |
| Institutional and social complicity in hiding crime, cognitive dissonance caused by perverse incentives |
| Policing performances |
| Reporting and recording and vulnerability |
| Rationality and fear of crime |
| Physical, digital on/off crimes vs. analogue crimes |
| Knowledge of crime that is criminal – causes harm itself |
| Pre-data, pre-crime |
| Security and insecurity, threat and risk |
| Numbers and identity |
| Crime as entertainment, target audience for crime data |
| Institutional and social complicity in hiding crime, cognitive dissonance caused by perverse incentives |
| Reporting and recording and vulnerability |
| Rationality and fear of crime |
| Physical, digital on/off crimes vs. analogue crimes |

FIGURE 6.4: Interviewer's Derived Concepts

.

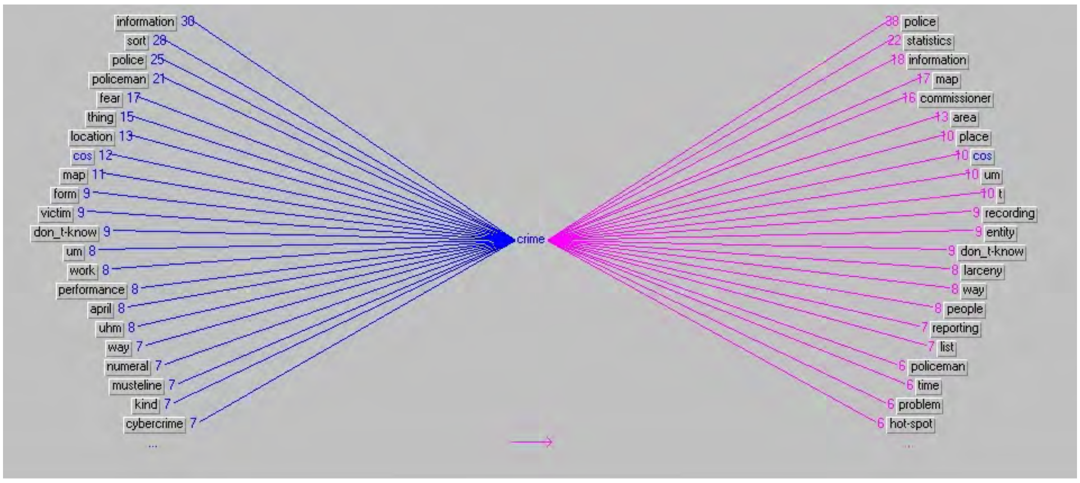


FIGURE 6.5: References to "Crime"

.

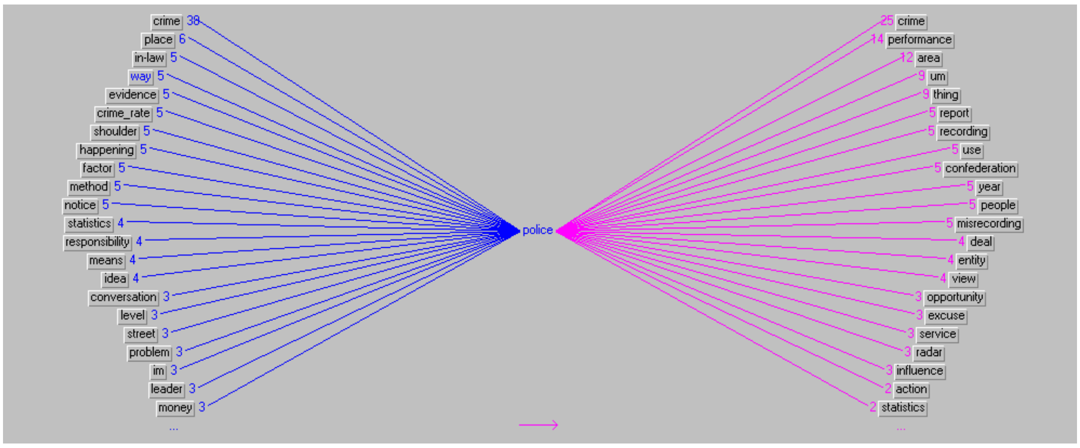


FIGURE 6.6: References to"Police"

.

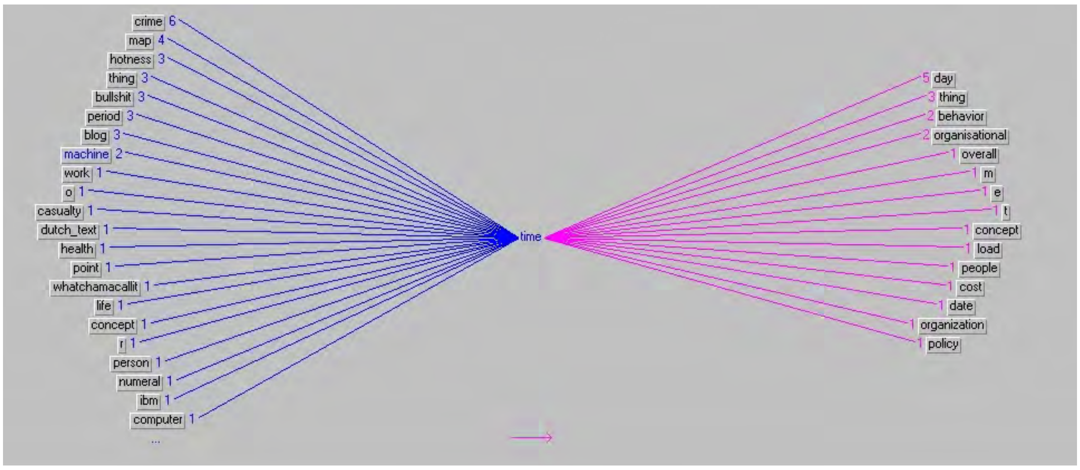


FIGURE 6.7: References to "Time"

.

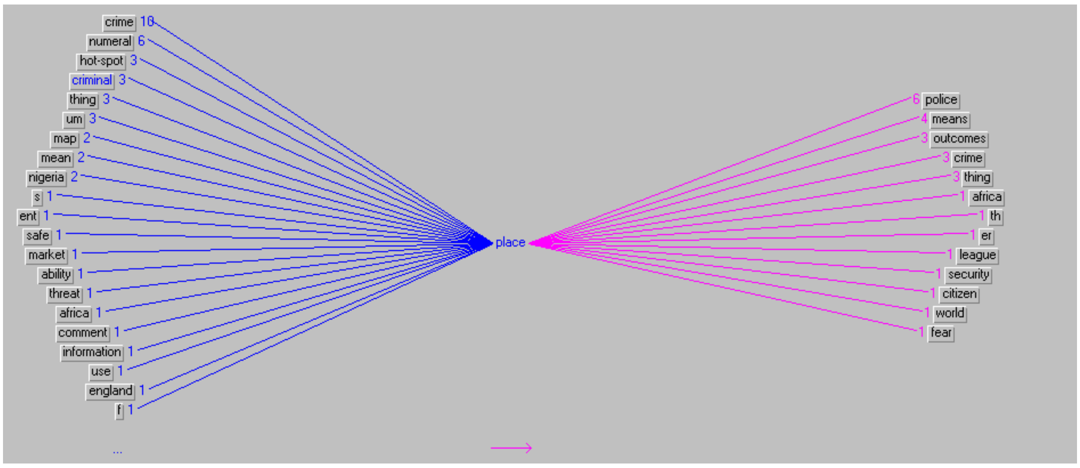


FIGURE 6.8: References to "Place"

.

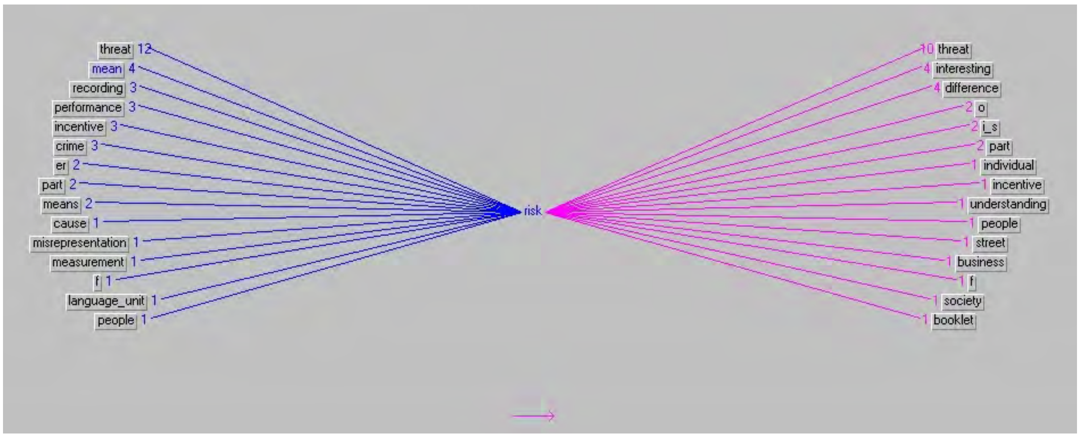


FIGURE 6.9: References to "Risk"

.

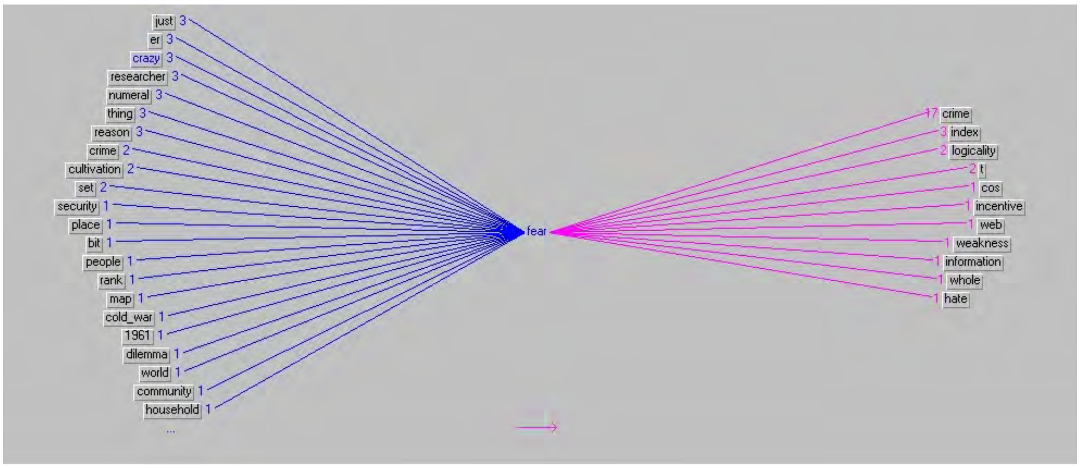


FIGURE 6.10: References to "Fear"

.

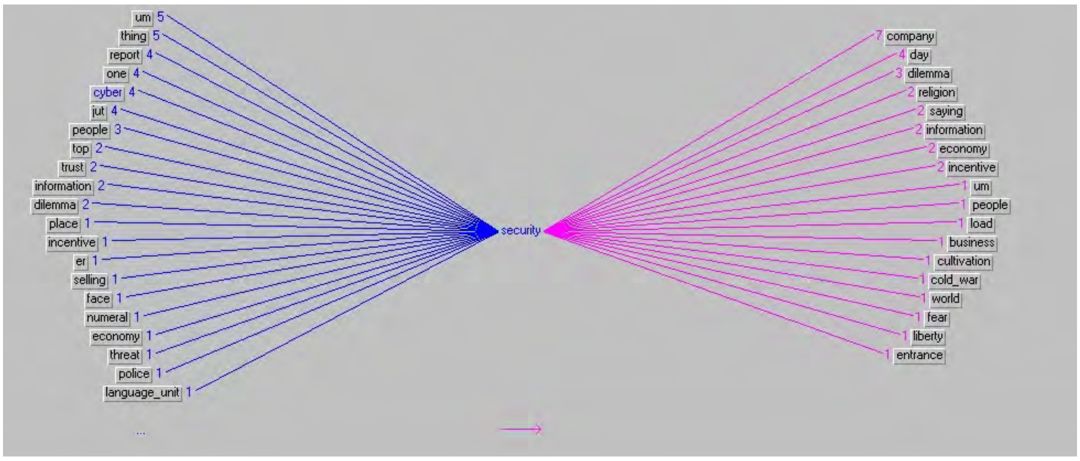


FIGURE 6.11: References to "Cybersecurity"

.

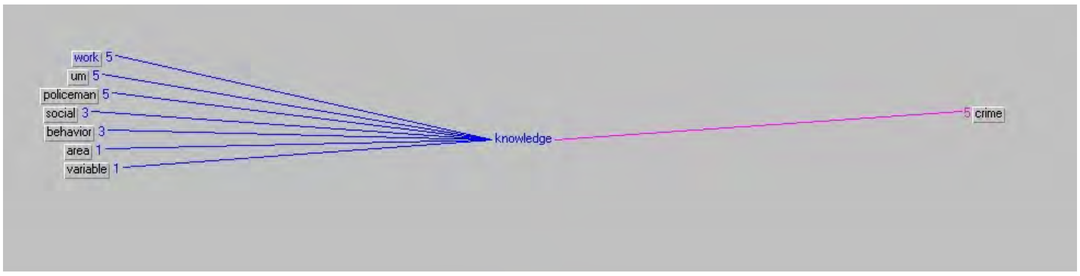


FIGURE 6.12: References to “Knowledge”

.

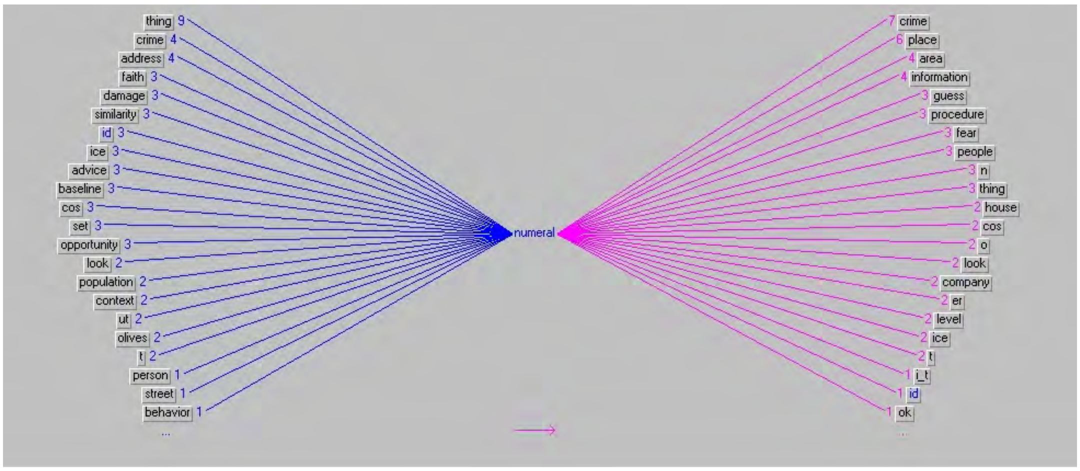


FIGURE 6.13: References to “Number”

.

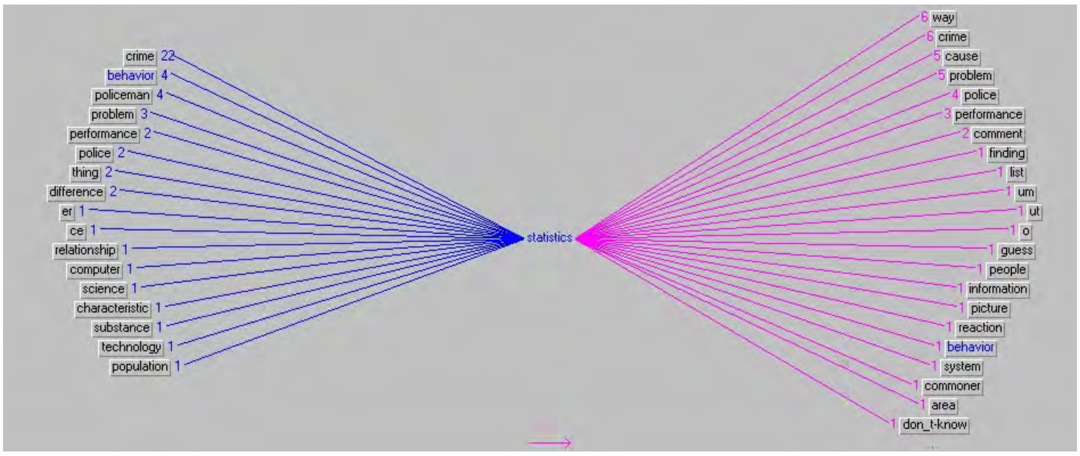


FIGURE 6.14: References to “Statistics”.

# Chapter 7

# Discussion

This chapter gives an overview of some of the issues we found, limitations of our work and possible future directions.

## 7.1 Summary

The research question was, "Evaluating the Impact of Open Crime Data in the United Kingdom."

In the first part of the thesis we: i) examined literature and popular concepts relating to knowledge of crime and open data; ii) reviewed how such data have come to be produced with respect to related concepts within statistics, surveillance and Big Data; iii) analysed the way in which policing and maps have been combined in the 21st Century through these first two sets of concepts (knowledge of crime and statistics).

The second part of the thesis used the multidisciplinary methodological framework to explain how and why we can unpick this question of understanding impact within the above domains. We used: i) Grounded Theory to examine the context that Police.uk inhabits with respect to knowledge of crime and Open Data; ii) Big Data concepts and Frame Analysis to examine the impact of Police.uk on those who interact with it on the World Wide Web. Finally, iii) we contrasted these frames with frames generated from interviews from cybercrime and security experts presented with Police.uk.. Cybercrime was a focus for understanding how crime data can be affected by the affordances of the World Wide Web.

## 7.2 Limitations and Future Work

### 7.2.1 Big Data

This thesis has explored some of the problems that we have come across in the use of Big Data based analysis when applied to Open Crime Data. We found that coders, app-makers and programmers tend to see Big Data as fuelling the knowledge economy in terms of understanding risk, and furthering knowledge through the predictions that such big data can make for us over entire populations. We spoke to the O.N.S. about their experience of working with Big Data to provide these sorts of insights. The O.N.S. has provided a guide for those who use crime statistics, the "User guide to crime statistics for England and Wales", (Office for National Statistics, 2019b) which provides detailed information on the datasets used to compile the crime statistics published by the O.N.S. This covers topics such as the Crime Survey for England andWales,anoverviewofPolicerecordedcrime,comparisonoftheC.S.E.W. andpolicerecordedcrime,offencetypes,perceptionsofcrimelevelsarising fromtheC.S.E.W.,classification,statisticalconventionsandmethods,andthe HomeOfficeCountingRules.Theguidediscussesthefactthatchangesintheir findingsaboutperceptionsofcrimemay,“simplybeduetowhichadultswere randomlyselectedforinterview.Weareabletomeasurewhetherthisislikely tobethecaseusingstandardstatisticaltestsandconcludewhetherdifferences arelikelytobeduetochanceorrepresentarealdifference.Onlyincreasesor decreasesthatarestatisticallysignificantatthe5percentlevel(andarethereforelikelytobereal)aredescribedaschangeswithinthemainbulletin,andin thetablesandfiguresthesearéidentifiedbyasterisks.”(8.1Confidenceintervalsandstatisticalsignificance).

## 8.2 Population estimates: household-only population estimates

Another issue that O.N.S. has tried to address is that they were not interviewing people living in halls of residence, NHS Nurses’ accommodation, prison or homeless people, so they were not sampling from the entire adult population of England and Wales,. Their assumption was that they could project the findings from residents, i.e. that “household-resident and non-household-resident populations experience similar levels of crime victimisation, and this is unlikely to be true.” They are now using census-based household-resident only population data and re-weighting. This introduces another issue - the weighting that is applied to the data - some examples that the O.N.S. provides are of addressing unequal address selection probabilities, non-response rates, dwelling unit weight, individual selection weight, number of incidents reported in each series of victim reports.

“...there were some instances in which high levels of repeat victimisation (97) coincided with very high weights. In one instance, final weights of more than 6,000 per individual coincided with a series that included 97 incidents of violence. The combined effect of this meant that by uncapping the estimates, one individual was contributing over 582,000 incidents to our annual violence estimates (as compared with the individual’s contribution of just over 30,000 incidents with the cap of five in place).”

The guide in particular refers to the transfer of responsiblity for publication of C.S.E.W. figures passing from the Home Office to O.N.S.. It states that “It is recommended that prior to using these open data tables, users read Chapter 2, in particular, of this user guide to familiarise themselves with the context of the data and the scope and limitations of the C.S.E.W. as a whole.” (9.1 Introduction to open data tables).

The Government Statistical Service (G.S.S.) addresses the fact that the U.K. has a wide range of surveys providing sources of social and economic information. They does not directly reference crime but states, “These surveys were designed at different times, to meet different needs, and have been commissioned by a range of departments. Consequently, the surveys were developed mostly in isolation from each other. This resulted in a lack of cohesion. Differences arose in concepts, definitions, design, fieldwork and processing practices, or “inputs”, and also in the way results are released, or “outputs”. There is now a cross-governement programme of work monitored by the G.S.S. through the National Statistics Harmonisation Group (N.S.H.G.), who are responsible for “development and maintenance of common statistical frames, definitions, questions and classifications for statistics”. This work done by the N.S.H.G. is then managed by the G.S.S. Statistical Policy and Standards Committee (G.S.S. S.P.S.C.), who look at statistical policies and standards covering all Official Statistics managed by the G.S.S.. (Office for National Statistics, 2019a). The N.S.H.G. also has Topic Groups that develop, review and maintain harmonised questions, concepts and outputs, or Harmonised Principles, that occur in most government socialsurveys, with harmonising business statistics and administrative data being part of this work. This is an example of the sorts of organisation that need to be present and working alongside other Government departments like D.C.M.S. that helps to be accountable for the presentation of information in such a way as to help stall the ebb and flow of information warfare taking place on the web. These should improve the use of Big Data in public policy and for general use.

### 7.2.2 Visualisation

A recurring theme that has run through this research is that of the problem of visualisation. This comes up at every level and has caused me to reconsider whether it is not in fact a core part of the research that should have been addressed. It is perhaps fundamental to the topic of Science Communication, mapping, cartography, intelligence, framing, cognition and data presentation.

As part of this work we created ontologies that mapped out the conceptual world that crime moves within. Although we spent much time trying out various packages and graph manipulation software, there was a danger that too much weight was being given to presentation, rather than trying to examine the problems that the visualisation was coming into contact with. This is a huge research space, from H.C.I. and User Experience to Human Factors, Design, Psychology, Cognition and Graph Theory. At the National CyberSecurity Centre’s Annual conference, (see (NCCGroup, 2018) for an overview), CYBERUK 2018, an interesting presentation was given by the Head of Innovation at B.T., who suggested that one of the best ways of using Artificial Intelligence in the future willl be to allow A.I.s to help us work through information at large scale - rather than processing Big Data, that they allow us to do so more intelligently.

I take the opportunity here to acknowledge that this work is limited by the visualisations I have used and that future work would seek to address this problem at a more fundamental level. This also bumps up against the thorny issues of explanation and trust in Big Data. Artificial Neural Networks and their kind are great for spotting hidden patterns and creating new classifications. As we saw though, classifiers can tend not to deal with ambiguity very well, and can create constructs that simply do not reflect reality - the danger is that in making such constructs they then will come to shape reality.

Another danger is in trying to represent risk visually on a map. Although we can see numbers of recorded crimes on the map, these do not represent relative risk on journeys across the map as we do not know what the proportion of recorded crime is to unrecorded and how these risks relate to different populations.

### 7.2.3 Privacy, Security and Transparency

It is also clear that there is a running battle between governments, organsations such as Apple and Facebook and citizens who give up their data in order to have access to relationships mediated by the W.W.W.. In the U.K., G.D.P.R. is arriving, while at the same time there is confusion engendered by our potential exit from the European Union. All of this causes us to have to think hard about the direction of travel for research into aspects of Big Data, surveillance, privacy and Open Data that allows us to frame tensions sensibly without in-fringing people's safety, security and right to be let alone. In "Intelligence Elites and Public Accountability", Vian Bakir (Bakir, 2018) discusses General Michael Hayden's (former Director of C.I.A. (2006-2009) and N.S.A. (1999-2005)) call for translucency rather than transparency. In an interview, ((Lynch, 2014) Hayden referenced this, saying, "And that actually is really good. Translucent, you can see through the thick glass. You get the broad outline of the shapes. You get the broad patterns of movements. But you don't get the fine print. And it's the fine print, when it goes public, that kills us."

### 7.2.4 Maps, Cartography and G.I.S.

There were also general issues in this work that came to light in attempting to select salient features from the fields of Open Data, statistics, surveillance and mapped crime data within policing. It is very hard to do each of these fields justice - each has a robust and venerable research history and it is perhaps an inevitable result that attempting to examine literature from these fields in enough depth to understand the issues will result in an either over-lengthy review of the literature that tries to cover much but delineates little in terms of syntheses, or that fine-tuning the literature will create assumptions about each field that conflates issues, in the more pejorative sense of the term. A great example was to be found within the worlds of Cartography and G.I.S. On the face of it, to a "naive geographer" (not meant in Egenhofer's sense, although in some respects this was how I approached the field, especially given the sensemaking work that referenced Heider's Attribution Theory), (Egenhofer and Mark, 1995), these look as though they are both to do with maps, however on talking to experts in the field, it becomes clear that there is a history of epistemological conflict between schools. This conflict in fact seemed to overlie an area that is ripe for further exploration and finally led to some of the future work discussed in the next section.

### 7.2.5 Future Work

Current research being conducted by the author is going on in Defence and across Policing, Law Enforcement and other related organisations. This research combines Place Geography (social constructions and meanings around Place) with Human Terrain Information layered on top of Geo-Analytics. This sort of approach combines Big Data with criminological understanding with some A.I. techniques applied.
"The urban environment is by its nature demanding in its complexity with dense geographies reflecting both compressed physical geography and differing human perceptions of the space. Current Foundational GEOINT has a very poor understanding of the urban environment. Gazetteers typically do not go to much detail below the level of settlements. Mapping and gazetteers reflect physical and political geographies that whilst useful often do not accurately reflect the vernacular geographies of communities inhabiting those spaces. Nor are the dynamic aspects of the urban environment modelled...Place Geography is an emergent line of research that attempts to understand Place as not simply a physical phenomenon, but also as the result of human activity and interpretation. It is therefore of significant relevance to those wishing to better explore

and represent the urban environment, the product of human activity." (Hart, Frew, and Byrne, 2018).

The approach in the Place Lab project is particularly of use in Information Activities and Operations. Scientists are seeking to create ontologies that can tie together mapped representations of place, social constructions and understandings around place, space, or location, cyber, as well as physical events and information operations, in order to create risk pictures. The premise in this is to use Big Data methods and Artificial intelligence informed by Web Science approaches, so as not to misunderstand the ways in which Big Data creates crime constructs. This can be instantiated in ways such as the construction of advanced, richer gazetteers – by seeking to bring in human factors to spatial information.

The Place Lab project takes the idea of mapping events - often crime, terror or war related - but pushes it much further by bringing in the concepts of meaning and representation alluded to herein. In order to understand the politics and tensions of a place, and what constitute the routine activities for example that can build into a hot-spotting picture, it is necessary to build representations of that place using all the names and meanings that that place might be referred to by. Meaning can be contentious. The modern push for Big Data methods is very much predicated around the vast volumes of data that are available to us. However, as we have mentioned, volume is nothing without intelligent processing. For example, much use is made of satellite imagery to extend mapping enterprises - so much so now that satellites themselves are marketplaces (Kopytoff, 2014). But what does such imagery really tell us about the places that it captures?

Technological advancements will progressively unfold more levels of physically identifiable detail about terrain that we can attach absolute certainty to. However little absolutely certain information can be directly derived from this image or the terrain itself about the variety of ways in which humans might refer to it, how they think of it or represent it, and what they use it for. We can infer ideas about use - for example it might be possible to assume that a large area of smoothish asphalt with particular sets of markings is a car-park or a runway or similar load-bearing area that vehicles travel through and remain within. We do not directly know as a result of our satellite imagery who the communities are that might use the area referred to by the image, or what meanings the place might have had for perhaps a previous, indigenous population. The car park might have a barrier that allows only people to park there who are in possession of a card that indicates they are part of a community through having passed some sort of vetting. In this example, the meaning that ties the population together (and therefore that needs to be captured in order to predict their movements) exists within a digital system, not visible in satellite imagery. Previous use of the space might be for people who were born in the region.

From the perspective of understanding, capturing these artefacts is both crucial (the indigenous population might wish to oust the vetted car-park users – knowing this history could help to predict unrest) and dependent on having information flowing intelligently and seamlessly that can layer derived or less certain meanings about place, with the more certain or absolutely certain information about the related spatial locations. We can then make predictions about such locations, where the elements of certainty attached to the layers of information are well-represented.

Some of this less-certain information might be gathered from sources where data is at scale and automatically harvested from social networks, phone conversations, text messages or recordings. The scale of such harvesting therefore makes it imperative that it is gathered and structured in such a way that machines can analyse it, as such a task is too onerous for human analysts where large volumes of data are concerned. However much techno-optimists might wish to suggest that such machine analysis has been successfully done in the past, experience tells us that often crucial information is missed out, misrepresented or even invented.

So while we might be able to infer some hypotheses about use from seeing various objects such as machines, vehicles, buildings, pipelines, land markings and so on, we have seen that what these are can be ambiguous without human interpretation or knowledge of other systems not immediately visible. In addition, human mental and speech representations do not typically project clear spatial boundaries around such references to place, which can cause loss of/or invented information where technological systems might operate on the assumption that all locations can be clearly demarcated. Another facet contributing to the uncertainty around Place knowledge is that it tends to change more rapidly than physical terrain information (though of course this too is subject to change), so the traditional capture on a map or presentation of a single Powerpoint slide can misrepresent in that people tend to interpret what they see as a static god's eye view.

The twin concepts of uncertainty and risk are very well-understood in Defence and Security but interviews with potential users to support this work reveal that the relationships might be more complex than previously understood. Taking the policing intelligence 5x5 model, (College of Police, 2005), allows us to annotate intelligence with metadata relating to its provenance and the quality of the data.

What comes up in analysing how such intelligence is used, is that there is a tacit sociotechnical element to intelligence-gathering and presentation, relating to intent and objectivity of analysis and use. Where users have requested that intelligence needs to be "presented as more than just a Powerpoint", we noted that there may be an element of risk-management taking place in such presentation. Several interviewees who had experienced being asked to present information succinctly and in one page to further operational decision-making had said that the more authoritative the person they are presenting to is, the less likely it is that that person wants any element of uncertainty in what they're given. This means that pragmatically speaking, decision-making in theatre (for example) can be pushed back to a single intelligence point such as a presentation or a slide or a map, which is supposed to have captured ground truth. Any inherent risks attached to such decision-making can then be perceived to be held in the intelligence chain that preceded it, with the presumption that the work has been faulty, not that ground truth can shift very rapidly, where socially constructed data is present, and such methods of presentation are no longer fit for purpose. Issues of authority and governance abound in terms of the collection of Place Intelligence and are currently poorly expressed in the data-structures and algorithms that deal with such intelligence. It is hoped that such work could resolve current difficulties, subject silos and misunderstandings, particularly around Linked Data, ontologies and the Semantic Web.

## 7.3 Final Remarks

Crime is a subject of enduring interest to us. Understanding it: where it occurs, why it occurs and how are clearly useful enterprises. Much work is being done across many spheres of operation to enable this understanding and the routes proposed in this thesis should help to make this work more likely to succeed. While it is acknowledged that there is much that can be contentious in the subject of crime: differing moralities, unreasonable optimism in the face of (un)interoperability, quietly warring States and systems, secrecy in surveillance - translucency that distracts rather then delineates, and transparency that diverts actors from their end goals, our focus is on truly bringing together people and systems in order, just like Colquhoun, Bentham, Mill and Brandeis to try to prevent crime, enhance people's lives, help create a stable economy and most importantly to help people at risk of becoming victims of crime arm themselves with knowledge that can be intelligently used.

# Chapter 8

# Conclusions

## 8.1 Answering the questions

To what extent did this research answer the questions? Having looked at, and taken into account, limitations and future directions we reflect on how well the methodology allowed us to answer the question. Key outcomes of the research are our findings, conclusions and recommendations:

The research question is, “What is the Impact of Open Crime Data on Those Who Produce and Consume it in the U.K.?” We answered this through the subquestions that were answered by each of the three analyses:

– RQ1. Can we describe the tensions between the actors and technologies involved in the production and consumption of open crime data in the U.K.?

– RQ2. How do the Web of Data and online social networks mediating this crime data affect transparency and accountability in the U.K.?

– RQ3. Can we combine big data and network science methods with criminological and philosophical theory to understand the effects of the supply of crime data from the web?

### 8.1.1 Findings

– We found that systems views are important for working with Big Data in the Knowledge Economy.

– We found that non-Law Enforcement people tend to use constructs of crime that assume crime is digital (in the sense of either a crime or not), that crimes can be counted, easily located in time and space and to assume that while the numbers might be used for political purposes, they are "good" numbers -they effectively can be used as resources that aid safety, or major life decisions such as buying a house.

– We found that Big Data approaches do not reflect the nuances of real life; crimes are often analogue, not simple objects, not always easily located in time and space, might be more easily defined by threat, harm and risk, but that these are often institutionally viewed, as harm can be subjective. Mapped crime does not measure crime-as-a-phenomenon but is police performance data, recorded often because of perverse incentives. Mechanisms of transparency can force these incentives into play. Harm can be caused at the person level or at the stock-market level by knowledge of crime so in this case transparency causes harm rather then preventing it (although systemically we might say that this harm can be overridden by the counterbalancing good - which brings us to ethical accounts). Overall crimes can be vectored in terms of the phenomenology of those undergoing them.

- We found that mapping is useful but salience is an issue when it comes to "normal" vs "cybercrime" - how do we get all the features relating to both on the same map?
- We also found that Law Enforcement officials do not trust crime data because they know transparency calls for work to be done that is not always the work that needs to be done; crime data is distorted by the use of targets. Using Big Data to understand or anticipate threat can be problematic: the problems caused by bolt-on security can cause further problems and shape insider threat. Knowledge of surveillance can erode trust between employees and em-ployer, so that it is important to see where ideology has been embedded in technology; where surveillance creates the threat it set out to mitigate.
- We found that crimes can be unreported and unrecorded: crime on a map can show a well-policed community, reflective of community trust in the police, where crime is reported and recorded. On the other hand, no crime on maps can mean self-policed communities that are high in crime, or conversely no crime. Or they could be high in cybercrime, known of but unmapped. Crime data in the name of transparency can therefore be produced in an epistemologically vicious circle, and does not manage risk or predict crime.

### 8.1.2 Conclusions

- We find that the publication of Open Crime Data does have an impact, at local levels.
- That social networks are highly instrumental in people understanding more about where they live, even if the information is not always accurate.
- That the careful use of Big Data with more qualitative methods is to be sought after but must be intelligently managed.
- As for transparency - Open Crime Data is a strange product in this context; it is still seized upon by those who wish to dominate the narratives about government, even while they misunderstand that they are really using open policing data.
- Equally, going back to Mason, the power and knowledge that comes from observing others are unbalanced from the start. The internalisation of this reality of panoptic discipline actually places the uncertainty back with the observer; observing a behaviour does not equate to observing their internal reality.

### 8.1.3 Recommendations

- We should find and manage real data about crime and use it to hold our government (and large institutions like Facebook, Instagram, Twitter) to account as part of an effort to show that the truth about our society is knowable - that the data out there is not just deployed by bots conducting perception management and information warfare for political ends.
- We need to understand and properly explore the conceptual structures in place underpinning the ontologies of Big Data and its algorithms and knowledge structures.

- This calls for a theory of knowledge for Government that properly unpins confusions over data, knowledge, meaning, representation, behaviour, provenance, security and privacy.
- Crime rates must not be used to define good or bad police performance.

### 8.1.4 Original Contribution:

The most important contribution to Web Science is that we have shown that a wide-ranging and multidisciplinary approach can generate research that stringently examines accountability in a policy context mediated by technology. This approach avoids some of the difficulties engendered by using more simple single-disciplinary methodology. Our key contribution is therefore the creation of a methodology that allowed us to:

- Examine the tensions between the actors and technologies involved in the production and consumption of Open Crime Data in the U.K..
- Understand how Web of Data and online social networks mediating this crime data affect transparency and accountability in the U.K..
- Combine big data and network science methods with criminological and philosophical theory to understand the effects of the supply of crime data from the web.

The frames that we generated and the interviews have produced:

- Empirical evidence of the effects of the publication of Open Crime Data on people.
- Understanding of how on-line social networks play in this and affect the analysis of crime data's impact on organisations and people, including the police themselves.
- New methodologies to understand the interplay of individuals, organisations, the Web of Data and online social networks.
- New ways of understanding crime in our society.
- Understanding of how to turn data about crime into information that can be used, or intelligence.
- Understanding of the Web's contributions to policy through its mediation of Open Data.

The methodology combined broad data methods and criminological concepts to explain how and why we can unpick this question, primarily through a Web Science-based multidisciplinary approach using network science concepts, Grounded Theory, Frame Analysis and Broad/Big Data techniques.

We described www.Police.uk, a web-mediated interface, maintained by the Home Office, which provides access to information about recorded U.K. crime on a large scale, via the publication and presentation of open crime data, and through maps showing crime locations and judicial outcomes. We set out ways in which the impact of this site is currently analysed at the level of the data, and suggested potential issues with these analyses. We spoke to experts, a lot of whom are involved with cybercrime, in order to illuminate these issues. We proposed routes via which these issues might be overcome using novel mixed-method interdisciplinary Web Science approaches.

### 8.1.5 Context

We now examine some of this in more detail. For our first analysis we looked at where Police.uk sits within the crime data/technology information economy and described several factors that situate it and other technologies within an overall context of knowledge and risk. These factors are not often explicitly used in impact evaluations; we suggest that describing them allows us to have greater predictive power in terms of understanding how to design apps/technologies/web artefacts for impact, given a more nuanced view of impact than just using marketing terms. Such factors are a development of normal systems analysis or systems metrics, which seem to have been dropped in the new “knowledge economy” and which we argue are to become more crucial as Artificial Intelligence, Big Data and Data Science start to dominate market thinking.

Furthering our systems view, we created an ontology or classification that describes the objects and stakeholders in the system, and many of the tensions and complexities inherent in the system producing open crime data, so that we know what some of the factors are that might affect impact, when going further into understanding how people react to the data in the next stage of analysis.

### 8.1.6 Big Data and the World Wide Web

For the second analysis we looked at some early Home Office versions of impact and developed these so as to overcome methodological problems around data paucity, self-selection and ecological validity. We used Google Analytics to find comments from people referring to Police.uk on social networks. We used a big data/statistical approach for this, being able to scrape all comments that Google could find on Social Networking Systems (other than Facebook and Twitter) and using automated semantic processing to build a further ontology that further describes the objects in the world view of those referring to Police.uk on social media. We also used human analysis to sense-check and examine the findings generated by the automated semantic processing.

Those responding on social networks to the data seemed generally:

- To trust the data.
- To have constructs of crime that assume crime is digital (in the sense of either a crime or not).
- To assume that crimes can be counted.
- To assume that crimes can be located in time and space.
- To assume that while the numbers might be used for political purposes, they are "good" numbers - they effectively can be used as resources that aid safety, or major life decisions such as buying a house.

We found that our small “Big Data” approach revealed the obvious tensions in the system, while seeming to further described or stated aims of transparency in allowing people to inform themselves with this data and showing that they trust it. The tensions are not only in terms of beliefs, but also methodological. Big Data inevitably uses constructs that might not necessarily reflect the nuances of real society and the individuals that constitute it.

### 8.1.7 Understanding How Crime Constructs are Created

The third analysis provided a balance - we wanted to take this big data approach and the ontology it had revealed (i.e. the concepts around crime data) and see whether a more criminological and philosophical approach would reveal the constructs around crime to be deeply complex. We interviewed experts, many of whom work in the cyber, security and intelligence worlds. We found that there are in fact many problems with open crime data - it is performance data and therefore not crime data per se, but policing data, it does not account for cybercrime properly which is, according to many authoritative accounts, supposed to be the cause of most crime these days (or at least to be inextricably involved); if we were to add cybercrime to our data we then start to construct crime a little differently.

It became clear that "crimes" are often analogue, not simple objects from which we can construct simple data, that they are not simply located in time or space; when we discussed how crimes are defined, the question of harm came up. A standard policing way of evaluating approaches to dealing with crime is through "threat, harm and risk." We already have found that crime is often evaluated in terms of financial impact. This is especially true of cybercrime. However, unpicking the sources for these assessments shows that financial impact is hard to quantify - where companies go public about a data breach for example, they have to tie in any public evaluations of financial losses with what they may have already publically stated in their Annual Reports and Accounts, to the City in previous years, or to whichever stockmarkets they move within. We have already looked at perverse incentives in reporting and recording crime. These perverse incentives become almost impossible to negotiate when examining company financial statements. While there is much more pressure on companies coming to bear on their treatment and understanding of personal and financial data, there has been no amnesty declared in terms of past dealings. This means that public accounts still have to tie themselves to a narrative that validates past dealings even where these might have been highly questionable. The mechanism of transparency can undermine or negate itself in this context - the pressure to be truthful is equally counterbalanced by a need to protect past less truthful dealings. (of course in Presidential, policy or parliamentary dealings, this can be a boon, where there have been changes in parties or leaders and these can be blamed). These evaluations themselves are often not highly scientific, so a picture of financial threat becomes mired in uncertainty very quickly.

Adding to this is the problem of evaluating the worth of data. This problem arises for companies attempting to place a value on intellectual property, goodwill and other such intangibles for the purposes of insuring against data breaches. This inability to quantify the worth of data makes the problem of uncertainty even deeper. We therefore conclude that we cannot evaluate crime very easily in terms of its declared financial impact. This has implications for mapped crime data, in many ways.

If we set aside definitions of crime that include financial impact, but instead define crimes (or start to include in their definition) such as revenge porn, sextortion, cyber-harrassment or cyber-stalking, in terms of the more general harm they cause, it becomes clear that a victim's perception of the harm that they are undergoing can ebb and flow. This is very much the case where abuse is an issue: financial, domestic, emotional or sexual. It is critical to a person's ability to survive that their conceptual models of self project themselves as agents and enabled in order to survive. This cognitive dissonance is in fact entirely rational, but conflicts with the self-understanding that is needed in order to address crime taking place. Where a child, a victim of domestic, employment or research abuse depends on the abuser for their survival, it can be cognitively impossible for them to address the fact that crimes are being committed against them, as it is the impact of that understanding that causes harm, as much as the other harms being visited upon those victims.

This plays back into our earlier findings on knowledge of crime - often knowledge of crime can be itself the thing that creates the crime to come into being - whether on a state scale - market-based where suspicion of malintent on the part of a state actor can cause markets to collapse, or cognitively on the part of a victim of child sexual exploitation, for example. Many victims report a switching in understanding taking place, between understanding they have been harmed and then believing that all is well in their world (cognitive dissonance). In discussions with experts, we developed a trust-based crime definition that could be applied to new areas of crime that are under discussion in the media at present. We can define the harm that an act causes, in terms of the degree to which the victim loses trust in the person who has acted against them. This then addresses issues such as consent - we can agree that for example a victim of sextortion can have trusted the person with whom an intimate photo or video was taken at the time, but then is harmed by the loss of trust in the person who then goes on to disseminate that photograph on social networks or porn channels. This is problematic to document and address evidentially, but maps a path through the complexities of consent and sexual exploitation that is somewhat clearer than just debating over whether or not consent occurred.

We therefore found that crimes can be vectored in terms of the phenomenology of those who undergo them, whether on a personal scale or by an entire population, where market-based crimes are in commission, and additionally that the levels of physical phenomena described or that attach to such crimes become very small and sometimes just conceptual.

### 8.1.8 Mapping

We took mapping into account - the idea of a map is that it creates a model of a world, stripping out the salient features and presenting them in an ordered way that allows a user to navigate using (supposedly) the minimum of information. Where the levels of phenomena described bridge several parts of our physical world, i.e. digital combined with human bodies and belongings, there is a problem about how to make the salient features cohere on a map – which is precisely the problem with the cybercrime that does not really come out in the O.C.D. of Police.uk.

Having been part of several policing and security user groups that discuss how best to train new and seasoned recruits in cybersecurity or how to understand and deal with cyber crime, the problem (to the groups) often revolved around deciding what were the salient features of cyber and how to transcend the physical bits and bytes layer to the information/data layer that can affect people in terms of crime – intrusion, phishing, scamming, sextortion, ransomware. We found that there is a feeling that "normal" crime exists; most police/L.E.O.s interviewed understood perfectly the notion of “signal” crimes, one interviewee suggested not only that reports of a murder with a pretty nameless model would include her age, but also her house price. “Normal” crime is very easily mapped, even if we have inaccurate data; it is easily understood because we all have experience with the objects referred to on a “normal” crime map. Cyber crime is not - even where it is apparently mapped, what it represents is ephemeral and hard to grasp. If it is not mapped we do not get a true picture of policing activities, (let alone crime); if it is, it becomes hard to understand, given that the phenomena under investigation exist on a very different scale to those involved with “normal” crime.

### 8.1.9 Crime Data and Performance Management

Putting the three analyses together and extrapolating further shows us that really understanding impact is not straightforward. Very few people I spoke to in policing / L.E. trusted crime figures, because they understood the system that produced them. This is not because of “gaming the system” as one officer referred to it at the P.A.S.C., but because the system that produces crime data is not itself transparent and distorts the focus of Law Enforcement into enacting bizarre twists and turns in order to do their jobs. Transparency and accountability in crime fighting induces officials to concentrate so much on proving and demonstrating that they are doing what they said they would do, and are not part of a corrupt system, that the route to proof becomes the goal in itself, rather than reducing crime, therefore producing corrupted activities, such as misuse of public funds.

The world of policing performance is ugly; in interviews I was told of the following examples:

- There is so much focus on targets and wins and senior ranks worrying about promotion, and being rewarded for reducing departments, that any genuine efforts to do good, or to set up initiatives that will actually fight crime, are not rewarded; often even derided by Senior Management who might not have the on-the-ground understanding needed. These ideas are then stolen and touted as their own, when they finally “get it”.
- Employees are often driven by a fear of losing their jobs and all that they have worked for, being moved and losing the relationships and routes to information that they themselves have built and invested in that they need to do their jobs effectively. Building up tacit knowledge networks is not rewarded. They worry about losing promotion opportunities, losing their jobs and even whether they might lose their personal lives and their homes. The cuts of “administrative thinking” are effectively putting a bomb under their professional and personal lives.
- This stifles innovation and original thinking, when these are needed more than ever in order to face the new challenges that technology brings to society. Cuts are not helping; the effect of cuts is that people are brought into to change-manage – i.e. decimate teams, and there is still much evidence of old attitudes persisting: of beastings and derogatory remarks, bullying and humiliation, which officers do not wish to complain about for fear of being ridiculed.

- Allowing such a culture to persist is self-defeating as it wastes the public funds that should be going to support the everyday police officer or Law Enforcement employee doing his or her job.

Some areas within policing itself appear to be breeding the very culture that produces the crime that they should be preventing. Police have a unique place in our society, frequently acted against, picked on and demonised. They some-times seem to fulfil a role, almost psychiatrically atavistic, as scapegoats. Trans-parency and accountability in a policing context seem to be historically linked to stories about historical corruption, about people being put away for having the "Wrong Face", about files getting lost if left out on the desk. (Kirby, 2007). However, as we have seen, when numbers are produced in the name of ac-countability, playing "the numbers game" is not a game at all; for many officers it is about keeping their already fragmented lives as intact as they can – they struggle to understand demands placed on them from above but wish (gener-ally) to prove and demonstrate their loyalty and desire to do what is asked. Silverman and Eterno's work, discussed earlier, analyses data that demonstrates how official New York City crime statistics were manipulated. It then explores the consequences of unreliable crime statistics, and how those consequences spread throughout policing and law enforcement organisations, affecting police, victims and all stakeholders. Althoguh this was some time ago, the problems mentioned are still present. Work by Simon Guilfoyle (in the U.K.) picks up on this problem with policing performance management and the use of tar-gets. (Guilfoyle, 2011). Guilfoyle's work, alongside that of Irene Curtis, has shown that crime data (among other aspects of policing, of which the data is just a part) is perverted by the use of targets. The Curtis Review, commissioned by Home Office was published in 2015:
"Policing needs to change to respond to the challenges of the future, including the changing nature of crime, the increasing range and complexity of demand, continued financial constraints and the rapid pace of technological change. As forces adapt to changing circumstances, performance frameworks will also need to adapt to help the police make decisions to meet these challenges – and to understand whether or not they are succeeding.
Numeric targets have seen extensive use in policing for many years, as part of both local and national police performance frameworks. The Public Service Agreements (P.S.A.s) of the 1990s in particular created a slew of national targets in policing and across the public sector more widely. Since then, problems associated with targets such as 'gaming' and 'perverse incentives' have been well documented and targets have gradually been dropped by many forces. The last of the national targets in policing (for increasing public confidence and targets for response times, included in the policing pledge) were removed by the Home Secretary in 2010." (Curtis, 2015, p.4). Guilfoyle's work shows that interpreting policing data is badly done, especially by the police themselves, "A common method of presenting data in the police performance environment is the use of binary comparisons. This involves com-paring two isolated numeric values, then interpreting the difference between them as a trajectory, or assuming it is significant. The practice is commonplace within UK police forces, appearing at face value to be a simple method for in-terpreting data. However, concerns exist about the efficacy of the approach; furthermore, experience suggests the practice leads to unwarranted assump-tions, which impair decision-making and encourage inappropriate behavioural

responses.” (Guilfoyle, 2015), and see also (Guilfoyle, 2012).

There is a further problem here. It is these very performance cultures that are one of the drivers for creating the insecure systems that create cybercrime. Many companies do not build security into company systems and policies from the beginning. Once the first part of the company/organisation/law enforcement mission is realised, security flaws become exposed and harsh measures are put in place to counter these. Where draconian policies and systems are imposed on employees, this builds disaffection and a lack of regard or respect for the behaviours that keep a company or organisation safe – especially where “safety” is just an unthinking add-on to the essential part of building company security policy which is creating loyalty and trust by keeping employees satisfied, motivated and rewarded for their work. Many investigations into large data heists have pointed to persons within organisations – sometimes referred to as the “insider threat.” See (Pollack, 2010).

The author has been asked to conduct surveillance to uncover such insider threat on a number of occasions and used the findings from this research to discuss with board members whether their policies are simultaneously creating and mitigating insider threats. Policy thinking is improving in this area, with understanding emerging around sociotechnical causes of cybercrime, or how ideology embedded in technology creates more problems, but this is something that organisations need to continue to be very aware of.

## 8.2 Constructing Crime

One of the most interesting findings from this research was the idea that crime is being constructed by Big Data. We dabbled with the building of ontologies, as the method was meant to be representative of what could be done rather than an exhaustive attempt to actually parse reality - building theory, rather than testing it. However, our experiment with the use of a semantic classifier showed us very quickly how dangerous these can be. Using big data culled from the Web in a security context will inevitably involve understanding or trying to anticipate threat. Threats will be constructed where the activities of persons online in conjunction with their networked relationships are scanned by algorithms and presented as profiles. Such profiling depends on certain structures being put into place in the building blocks of the ontologies. These structures might be to do with family relationships, the movement of funds from one country to another, political or ideological alliances or geography.

For example it has been suggested that people using cryptocurrencies should be investigated since transactions can be hidden and evade the notice of the authorities - exactly the sorts of transactions that Colquhoun sought to expose. In fact the use of cryptocurrencies can go hand in hand with a sort of anarcho-capitalism ideology: “The ancap worldview only supports sovereign individuals engaging in free-market exchange. Neither states nor corporations are acceptable intermediaries. That leaves a sparsely set table. At it: individuals, the property they own, the contracts into which they enter to exchange that property, and a market to facilitate that exchange. All that’s missing is a means to process exchanges in that market.” (Bogost, 2017). Should people holding such ideologies be routinely investigated? Banks themselves are divided on the subject of cryptocurrency, and so they set the agenda when it comes to creating moral outrage - such outrage then feeding into the crime constructs of this century. “Most banks steer clear of cryptocurrencies, worried about money laundering or terrorism finance because of the inherent anonymity of the assets...Jamie Dimon, chief executive of J.P.Morgan Chase, has said anyone caught trading bitcoin at the US bank would be fired. Many British and U.S. banks have blocked cryptocurrency purchases on credit cards, while some UK lenders are refusing mortgages to clients with deposits funded by selling cryptocurrencies.” (Arnold and Atkins, 2018). The author herself was told by one cashier that, “people using Bitcoin and the like are really dodgy and there must be something wrong with them.”

In fact, looking for ideologies in online activities is very hard to do - with a democracy, it is hard to state when the act of letting off steam on a forum might actually allow people to express emotions that unexpressed would build into something more dangerous. The act of going after such people and interviewing them about what they said, could then cause more problems than it might solve. Big Data tends to sift people according to a risk they might represent, and as a result, we get people wrongly identified as being on no-fly lists, or the Vogelman farm’s owners in Kansas who have “been accused of being identity thieves, spammers, scammers and fraudsters. They’ve gotten visited by F.B.I. agents, federal marshals, I.R.S. collectors, ambulances searching for suicidal veterans, and police officers searching for runaway children. They’ve found people scrounging around in their barn. The renters have been doxxed, their names and addresses posted on the internet by vigilantes. Once, someone left a broken toilet in the driveway as a strange, indefinite threat.” (Hill, 2016).

These are all of course anomalous examples of Big Data going wrong, that can help to sell a media narrative of evil Big Brother and the Panoptic society, however, the factor that most of them have in common, aside from the hype, is that the owners of the databases and graph structures and algorithms that construct such risks, tend to be rich and powerful, while those who get wrongly characterised as risks are not. Can the ordinary person arm themselves with enough knowledge of crime that they get to have a say in what represents risk, without participating in the machinations of the fourth Estate?

## 8.3 Mapped Crime Data

Our first conclusions then are that the data in Police.uk is not to be trusted. Interviews with L.E.O.s suggest that the stuff they write down to do their jobs (that becomes crime data) is distorted by the demands of their jobs. What crime we see on the map is not representative of what crime there is. The dark figure of unrecorded crime persists – much crime goes unreported because of fear of the perpetrator or mistrust of the police. In an interview given by the author and D.C.I. Paul Gelman to Radio Solent on November 30th 2016, on combating sextortion, we suggested that a pattern is emerging, not just in the U.K., but world-wide, of extortion-based crimes driven by leveraging fear of shame and a fast movement over digital platforms in order to elude detection. (Gelman and Byrne-Evans, 2016). There is evidence that many of these crimes are not being reported because of the shame factor. This is of interest not only as an illustration of the potential dark figure of crime, but also because in a transparency and surveillance context we have to ask how we can remove impediments to reporting? We can reassure the public that there is no shame in seeking comfort from strangers, for example. But then if we remove the moral constraint that inhibits people from exposing themselves in order to expose such crimes, then while the leverage disappears, social behaviour might become more risky. In a book about working within the kidnap and ransom world the author, Ben Lopez, suggests (akin to Durkheim's notion of crime as boundary act) that we need kidnap and ransom (or the modern digital sextortion crime for example, or ransomware) to be a possibility, as without the emotions that drive the existence of the crime we would be empty. (Lopez, 2011).

One conclusion from the interviewees was that crime shown on a map perversely suggests that here is a well-policed community; where there is no crime on a map, this could well indicate at least one of three things: a self-policed community - high in crime - that does not adhere to the law, no crime, or many unmapped cybercrimes causing millions of pounds worth of loss and other harms.

Going further, we ask, does it really matter if not all crime is on the map? Our findings are that this problem is representative of a greater problem that persists across several policy sectors where data is used to measure the performance of people doing fundamental jobs: teaching, policing and healthcare. All of these jobs seek to help or adjust phenomena that are naturally occurring and therefore out of the remit of humans to control. Teachers deal with intelligence, doctors with health and police with crime.

“We don't know the full extent of crime, crime isn't always reported for a variety of reasons, crime is affected by multiple external factors that aren't necessarily related to police performance, so consequently using crime rates as a proxy method of police performance I think is horribly wide of the mark, because it's putting responsibility of the shoulders of the police, for something which is affected by socioeconomic factors, substance abuse, range of various external factors such as the weather, which can have a big influence effect at the force or macro level more than the police would ever be able to directly have influence over.” (From the interviews).

Where the success of people's jobs is determined by how well they manage to make it look as though they are controlling phenomena over which they have no remit (sickness, death, murder, rape, terrorism, intelligence), then the system producing the data, whether it is health, education or policing data is founded on an epistemologically vicious circle. If we introduce targets, league tables, or any forms of comparison into the mix, then the data is perfectly meaningless as crime data, health data or education data. It is data about policing, about doctors and about teachers and about how they are managing in the face of policy imposed on them. Does this matter? We can put policing data on a map and it is still of interest.

However there is a big problem in that many of the people designing apps for the new economy heralded by transparency and open knowledge, the ones who think that data is their new oil, do not understand the very data that they are using. One of the populations (open data hackers/ app-makers) that is most charmingly vocal about transparency seems to be the most fooled by its data. There is confusion about what data can do. Instinctively many people think that crime data, as it stands, can be used to predict crime, and thereby manage risk.

The concept of hot-spotting has been around for a long time and is operationally well-respected in terms of understanding resourcing and deployment of officers and equipment. It is also one of the “common-sense” ideas that the public seems to half-know about, but not fully understand the permutations of. The crime data in Open Crime Data appears to be the sort of data that lets us know where we should be more careful. We know, going back to our ontologies, that there is a sort of data that is predictive - sense-making data is physically accurate and not so open to interpretation, therefore can be used in military systems for example, where the consequences of bad data are certainly life or death. "Within a military context the dominant paradigm is of knowledge superiority, sense-making, problem-solving and decision-making – battlespace awareness and visualisation". (Government of Canada, 2003).

I have spoken to many people who work in Open Data who talk enthusiastically about O.C.D. as adding to the panoply of things that can be done to serve society. Crime maps are attractive to many people. Is there a way of pushing them further so that they are actually of use as aids to navigate risk? The Big Data present in surveillance from A.N.P.R. systems, mobile phones, credit card purchases, online behavioural tracking – all of these data-gathering sys-tems do produce something accurate and possibly predictive - if we combine it with policing data, crowd-sourced data and surveys to do with people's ex-perience of crime we might start getting something that approaches predictive crime data.

Apps like Fearsquare told us about places to stay away from:

"... an application that allows people to visualize and interact with official UK crime statistics in a way that is specific to their own, individual, everyday life. People can use the application to easily get a picture of the levels of crime in places that they commonly live, travel through, or visit." There are however twofold problems - one - according to our classification, this site is likely not to be maintained (and is in fact offline at the time of writing) and two - the data it references is wrong. Luckily this app is, or was, somewhat tongue-in-cheek, with leaderboards, FearPoints and new levels of crime "unlocked". (Olanoff, 2012).

"I think when the narrative is changed from using the crime rates as a definer of good or bad police performance then there's an opportunity to use the numbers differently, use them as a source of information rather than something which says whether it's good or bad." (Interviewee).

We need to regain control of information about crime, if we want crime maps to be any use. We therefore find that the publication of Open Crime Data does have an impact, at local levels, that social networks are highly instrumental in people understanding more about where they live, even if the information is not always accurate, and that the careful use of Big Data with more qualitative methods is to be sought after but must be intelligently managed. As for transparency - Open Crime Data is a strange product in this context; it is still seized upon by those who wish to dominate the narratives about government, even while they misunderstand that they are really using open policing data - however we should seek to find and manage real data about crime and use it to hold our government (and other actors) to account as part of an effort to show that the truth about our society is knowable - that the data out there is not just deployed by bots conducting perception management and information warfare for political ends.

# Bibliography

Ackoff, Russell (1999). *From Data to Wisdom*. URL: http://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf.

Alexander Schellong and Ekaterina Stepanets (2011). *Unchartered Waters The State of Open Data in Europe*.

Alvesson, M. & Skoldberg, K (2010). *Reflexive Methodology: New Vistas for Qualitative Research*. SAGE, p. 319. ISBN: 9780803977068.

Anderson, Barbara and Brian Silver (1990). "Growth and Diversity of the Population of the Soviet Union". In: *Annals, AAPSS* 510. URL: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/67141/10.1177{_}000271629051000112.pdf?sequence=2.

Anderson, Chris (2004). *The Long Tail*. URL: http://www.longtail.com/about.html (visited on 05/09/2018).

Anderson, Katharine (2005). *Predicting the weather : Victorians and the Science of Meteorology*. University of Chicago Press, p. 331. ISBN: 9780226019680.

Anderson, Ross (2014). *Privacy versus government surveillance: where network effects meet public choice*. Tech. rep. URL: http://elastic.org/{~}fche/mirrors/www.cryptome.org/2014/05/rja-privacy-v-spying.pdf.

Anning, Steve et al. (2016). *Crime in the Digital Age: Digital Policing*. Tech. rep. IBM, p. 9. URL: http://www-935.ibm.com/services/multimedia/Crime{_}in{_}the{_}digital{_}age{_}white{_}paper{_}v4{_}Discover.pdf.

Arnold, Martin and Ralph Atkins (2018). *European banks break ranks over cryptocurrencies*. URL: https://www.ft.com/content/2225e392-0f08-11e8-8cb6-b9ccc4c4dbbb (visited on 05/12/2018).

Auden, Wystan Hugh (1955). *The Shield of Achilles*. URL: https://www.bl.uk/collection-items/the-shield-of-achilles-by-w-h-auden.

Baker, Monya (2016). "1,500 scientists lift the lid on reproducibility". In: *Nature* 533.7604, pp. 452–454. ISSN: 0028-0836. DOI: 10.1038/533452a. URL: http://www.nature.com/doifinder/10.1038/533452a.

Bakir, Vian (2018). *Intelligence Elites and Public Accountability : Relationships of Influence with Civil Society*. First. Oxon: Routledge. ISBN: 1351388959. URL: https://books.google.co.uk/books?id=zSFWDwAAQBAJ{\&}pg=PT220{\&}lpg=PT220{\&}dq=translucency+or+transparency+Uk+govt{\&}source=bl{\&}ots=X9hxZ{_}BQm-{\&}sig=yssNihv0bIAhyso-hF{_}JXN-DAO8{\&}hl=en{\&}sa=X{\&}ved=0ahUKEwirmN31q{_}jaAhXTY8AKHSu0CEUQ6AEIYzAH{\#}v=onepage{\&}q=translucencyortransparencyUkgovt{\&}f=false.

Barabasi, Albert, Laszlo (2002). *Linked The New Science of Networks*. Cambridge, MA: Perseus Books.

Baxter, Lynne F. and Constanze Hirschhauser (2004). "Reification and representation in the implementation of quality improvement programmes". In: *International Journal of Operations and Production Management* 24.2, pp. 207– 224. ISSN: 0144-3577. DOI: 10.1108/01443570410514894 URL: https://www. emeraldinsight.com/doi/10.1108/01443570410514894

Bayley, David, H. and D. Shearing, Clifford (1996). *The Future of Policing*. URL: http://www.jstor.org/discover/10.2307/3054129?uid=3738032{\&}uid=2{\&}uid=4{\&}sid=21102551329021 (visited on 08/25/2013).

Bellinger, Gene, Durval Castro, and Anthony Mills (2004). *Data, Information, Knowledge, & Wisdom*. URL: http://www.systems-thinking.org/dikw/dikw.htm (visited on 08/21/2013).

Bentham, Jeremy (1787). *Panopticon; or, The Inspection-House*. Dublin printed, Reprinted and sold by T. Payne.

Berlin, Isaiah, Henry. Hardy, and Roger. Hausheer (1998). *The proper study of mankind : an anthology of essays*. Pimlico, p. 667. ISBN: 0712673229. URL: https://books.google.co.uk/books?id=37y4B1TR4zAC{\&}printsec=frontcover{\#}v=onepage{\&}q{\&}f=false.

Bernays, Edward (1928). *Propaganda*. URL: http://www.amazon.com/Propaganda-Edward-Bernays/dp/0970312598 (visited on 04/25/2014).

Berners-Lee, Tim and Mark Fischetti (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. London: Orion Business Books. URL: http://www.w3.org/People/Berners-Lee/Weaving.

Berners-Lee, Tim et al. (2006). "A Framework for Web Science". In: *Foundations and Trends R in Web Science* 1.1, pp. 1–130. ISSN: 1555-077X. DOI: 10.1561/1800000001. URL: http://www.nowpublishers.com/product.aspx?product=WEB{\&}doi=1800000001.

Bogost, Ian (2017). *Cryptocurrency Might be a Path to Authoritarianism - The Atlantic*. URL: https://www.theatlantic.com/technology/archive/2017/05/blockchain-of-command/528543/ (visited on 05/12/2018).

Brandeis, Louis and Samuel Warren (2010). "The Right to Privacy". In: *Harvard Law Review* 4.5, pp. 193–220.

Bryant, Antony and Kathy Charmaz (2007). *The SAGE Handbook of Grounded Theory*. SAGE Publications, p. 623. ISBN: 1446275728. URL: https://books.google.co.uk/books/about/The{_}SAGE{_}Handbook{_}of{_}Grounded{_}Theory.html?id=H1HHVV8qt4gC{\&}redir{_}esc=y.

Burgess, Ernest Watson (1967). *The growth of the city: An introduction to a research project (Bobbs-Merrill reprint series in the socialsciences)*. Bobbs-Merrill. URL: http://www.amazon.co.uk/The-growth-city-introduction-Bobbs-Merrill/dp/B0007HGRCQ.

Byrne-Evans, Maire and Christine Task (2013). "Keeping Your Little Back Shop". In: *XRDS: Crossroads, The ACM Magazine for Students* 20.1, p. 9. ISSN: 15284972. DOI: 10.1145/2517252. URL: http://dl.acm.org/citation.cfm?doid=2517249.2517252.

Byrne Evans, Maire et al. (2013). *Crime Applications and socialMachines: Crowdsourcing Sensitive Data*. URL: http://dl.acm.org/citation.cfm?id=2487788.2488075.

Callamard, Agnes (2008). *Accountability, Transparency, and Freedom of expression in Africa*. URL: https://www.article19.org/data/files/pdfs/press/accountability-transparency-and-freedom-of-expression-in-africa.pdf.

Campbell, Bradley. *Black's Theory of Law and socialControl - Criminology - Oxford Bibliographies*. URL: http://www.oxfordbibliographies.com/view/document/obo-9780195396607/obo-9780195396607-0067.xml;jsessionid=1D9BFB9507A22B8D37B33BE1359B7890 (visited on 08/25/2013).

Caplan, Jane and John Torpey. *Documenting Individual Identity: The Development of State Practices in the Modern World.* Princeton University Press. URL: http://press.princeton.edu/TOCs/c7183.html.

Chainey, S. and L. Tompson (2012). "Engagement, Empowerment and Transparency: Publishing Crime Statistics using Online Crime Mapping1". In: *Policing* 6.3, pp. 228–239. ISSN: 1752-4512. DOI: 10.1093/police/pas006. URL: https://academic.oup.com/policing/article-lookup/doi/10.1093/police/pas006.

College of Police (2005). *How to Complete a 5x5x5 Form and Relevant Supplements*.

Colquhoun, Patrick (1796). *A Treatise on the Police of the Metropolis, Containing a Detail of the Various Crimes and Misdemeanors by which Public and Private Property and Security Are, at Present, Injured and Endangered, and Suggesting Remedies for Their Prevention - Google Play*. URL: https://play.google.com/books/reader?id=Z4RJAAAAYAAJ{\&}hl=en{_}GB{\&}pg=GBS.PA28 (visited on 06/09/2019).

— (1806). *A treatise on indigence: exhibiting a general view of the national resources for productive labour; with propositions for ameliorating the condition of the poor, and improving the moral habits and increasing the comforts of the labouring people*. Printed for J. Hatchard, p. 302. URL: http://books.google.co.uk/books/about/A{_}treatise{_}on{_}indigence.html?id=YiJJAAAAMAAJ{\&}pgis=1.

Copperfield, David (2006). *Wasting Police Time: The Crazy World of the War on Crime*. Monday Books, p. 304. ISBN: 0955285410. URL: http://www.amazon.co.uk/Wasting-Police-Time-Crazy-World/dp/0955285410.

Corbin, Juliet M. and Anselm Strauss (1990). "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria". In: *Qualitative Sociology* 13.1, pp. 3–21. ISSN: 0162-0436. DOI: 10.1007/BF00988593. URL: http://link.springer.com/10.1007/BF00988593.

Crawford, Kate (2014). "The Anxieties of Big Data". In: *The New Inquiry*. URL: http://thenewinquiry.com/essays/the-anxieties-of-big-data/.

Curtin, Deirdre and André Nollkaemper (2005). "Conceptualizing Accountability in International and European law". In: *Netherlands Yearbook of International Law* 36.1, p. 3. ISSN: 0167-6768. DOI: 10.1017/S0167676805000036. URL: http://www.journals.cambridge.org/abstract{_}S0167676805000036.

D. Haggerty, Richard V. Ericson, Kevin (2000). "The Surveillant Assemblage". In: *British Journal of Sociology* 51.4, pp. 605–622. ISSN: 0007-1315. DOI: 10.1080/00071310020015280. URL: http://doi.wiley.com/10.1080/00071310020015280.

Dahl, Robert A., Ian. Shapiro, and Jose Antonio. Cheibub (2003). *The Democracy Sourcebook*. MIT Press, p. 556. ISBN: 0262541475. URL: https://books.google.co.uk/books?id=B8THIuSkiqgC{\&}pg=PA38{\&}lpg=PA38{\&}dq=self-congratulatory+democracy{\&}source=bl{\&}ots=VHzlh9EAom{\&}sig=muODQF05zfut3HZLNtja6iLw3NM{\&}hl=en{\&}sa=X{\&}ved=0ahUKEwjn16LtyvjaAhVJWsAKHb}v=onepage{\&}q=self-congratulatorydemocracy{\&}f=false.

David A. Snow; E. Burke Rochford (1986). "Frame Alignment Processes, Micromobilization, and Movement Participation". In: *American Sociological Review* 51.4, pp. 464–481.

Davies, Tim (2010). *Open data, democracy and public sector reform*. URL: http://www.opendataimpacts.net/report/ (visited on 08/24/2014).

Dean, Mitchell (2011). *The constitution of poverty : towards a genealogy of liberal governance*.Routledge,p.248.ISBN:9780415609586.

Desrosières, Alain (2002). *The Politics of Large Numbers: A History of Statistical Reasoning*.

Dfid (2012). *Broadening the Range of Designs and Methods For Impact Evaluations Sharing the Benefits of Trade: DFID's Aid for Trade Portfolio Monitoring & Evaluation Framework*. URL: https://www.gov.uk/government/uploads/system/uploads/attachment{_}data/file/67427/design-method-impact-eval.pdf.

Difranzo, Dominic et al. (2010). *TWC LOGD: A Portal for Linking Open Government Data*. DOI: 10.1.1.377.5547. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.377.5547 (visited on 08/24/2014).

Dodsworth, Francis (2007). "Police and the Prevention of Crime: Commerce, Temptation and the Corruption of the Body Politic, from Fielding to Colquhoun." In: *British Journal of Criminology* 47.3, pp. 439–454. DOI: 10.1093/bjc/azl054. URL: http://dx.doi.org/doi:10.1093/bjc/azl054.

Durkheim, Eimile (1964). *The Normal and the Pathologic*.

Egenhofer, Max J. and David M. Mark (1995). "Naive Geography". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 988, pp. 1–15. DOI: 10.1007/3-540-60392-1_1. URL: http://link.springer.com/10.1007/3-540-60392-1{_}1.

Eterno, John and Eli B. Silverman (2012). *The crime numbers game : management by manipulation*. CRC Press, p. 282. ISBN: 1439810311.

Ferlie, Ewan et al. (2013). *Making Wicked Problems Governable?: The Case of Managed Networks in Health Care*. OUP Oxford, p. 304. ISBN: 0191641421. URL: http://books.google.com/books?id=OHZpAgAAQBAJ{\&}pgis=1.

Finkelstein, Andrea (2000). *Harmony and the Balance*. Ann Arbor, MI: University of Michigan Press. ISBN: 9780472111435. DOI: 10.3998/mpub.16623. URL: http://www.press.umich.edu/16623.

Fischer, Frank (2003). *Reframing Public Policy : Discursive Politics and Deliberative Practices*. Oxford University Press, p. 266. ISBN: 0191529362.

Flatley, John (2013). *Crime Recorded by the Police Falling at a Faster Rate than Suggested by Independent Survey*. eng. URL: http://www.ons.gov.uk/ons/rel/crime-stats/crime-statistics/period-ending-june-2013/sty-recorded-crime.html.

Foss, Nicolai J., Kenneth Husted, and Snejina Michailova (2010). "Governing Knowledge Sharing in Organizations: Levels of Analysis, Governance Mechanisms, and Research Directions". In: *Journal of Management Studies* 47.3, pp. 455–482. ISSN: 00222380. DOI: 10.1111/j.1467-6486.2009.00870.x. URL: http://doi.wiley.com/10.1111/j.1467-6486.2009.00870.x.

Foster, Janet and B Bowling (2002). *'Police' and 'Policing'*. en. URL: http://eprints.lse.ac.uk/9336/.

Foucault, Michel (1977). *Discipline and Punish The Birth of the Prison*. Penguin, pp. 293–308.

Fournel, Victor (1858). *Ce qu'on voit dans les rues de Paris (Google eBook)*. A. Delahays, p. 410. URL: http://books.google.com/books?id=jTpSw1yVbicC{\&}pgis=1.

Fox, Jonathan (2007). "The Uncertain Relationship between Transparency and Accountability". In: *Development in Practice* 17.4-5, pp. 663–671. ISSN: 0961-4524. DOI: 10.1080/09614520701469955. URL: http://www.tandfonline.com/doi/abs/10.1080/09614520701469955.

Freedman, Lawrence (2013). *Strategy: A History*. Oxford: OUP USA, p. 752. ISBN: 0199325154. URL: http://www.amazon.co.uk/Strategy-History-Sir-Lawrence-Freedman/dp/0199325154.

Friedewald, Michael (2009). "Privacy Threats in the Ubiquitous Information Society: An Analysis of Trends and Drivers". In: *In: Proceedings of the WebSci'09: Society On-Line, 18-20 March 2009, Athens, Greece. (In Press)*. URL: http://journal.webscience.org/152/.

Friendly, Michael and Nicolas de Sainte Agathe (2012). "André-Michel Guerry's Ordonnateur Statistique : The First Statistical Calculator?" In: *The American StatistiC.I.A.n* 66.3, pp. 195–200. ISSN: 0003-1305. DOI: 10.1080/00031305.2012.714716. URL: http://www.tandfonline.com/doi/abs/10.1080/00031305.2012.714716.

Fung, Archon, Mary Graham, and David Weil (2008). *Full Disclosure: The Perils and Promise of Transparency*. Cambridge University Press, p. 304. ISBN: 0521699614.

Galison, Peter (2004). *Removing Knowledge*.

Gallhofer, Dr Sonja and Professor Jim Haslam (2005). *Accounting for Society: Some Critical Interventions*. Routledge, p. 240. ISBN: 113460050X. URL: http://books.google.com/books?id=KbmBAgAAQBAJ{\&}pgis=1.

Garland, David (1992). *Criminological Knowledge and its Relation to Power: Foucault's Genealogy and Criminology Today*. DOI: 10.2307/23638315. URL: https://www.jstor.org/stable/23638315.

— (2001). *The Culture of Control: Crime and socialOrder in Contemporary Society*. Oxford, UK: Oxford University Press.

Gelman, Paul DCI and Maire Byrne-Evans (2016). *BBC Radio Solent - Louisa Hannan, Are young people aware of the dangers of 'sexting'? - We find out more about a new campaign*. Southampton. URL: https://www.bbc.co.uk/programmes/p04flk0m.

Gil-Garcia, J Ramon and Theresa A Pardo (2006). "Multi-Method Approaches to Understanding The Complexity of E-Government". In: *International Journal of Computers, Systems and Signals* 7.2. URL: https://www.ctg.albany.edu/publications/journals/ijcss{_}multi-method/ijcss{_}multi-method.pdf.

Gitlin, Todd. (1980). *The whole world is watching : mass media in the making & unmaking of the New Left*. University of California Press, p. 327. ISBN: 0520038894. URL: https://books.google.co.uk/books/about/The{_}Whole{_}World{_}is{_}Watching.html?id=SMtHxaYV-UcC{\&}redir{_}esc=y.

Gladwell, M (1996). *The Tipping Point*. URL: http://www.gladwell.com/1996/1996{_}06{_}03{_}a{_}tipping.htm (visited on 07/19/2013).

Glaser, Barney and Anselm Strauss (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction, p. 271. ISBN: 0202302601. URL: http://www.amazon.com/The-Discovery-Grounded-Theory-Qualitative/dp/0202302601.

Glenn, Brian J. and Peter L. Bernstein (1999). "Against The Gods: The Remarkable Story of Risk". In: *The Journal of Risk and Insurance* 66.3, p. 517. ISSN: 00224367. DOI: 10.2307/253563.

Goffman, E (1956). *The presentation of self in everyday life*. New York: Anchor Books, p. 251. ISBN: 978-0-14-013571-8. URL: http://www.amazon.co.uk/Presentation-Self-Everyday-Penguin-Psychology/dp/0140135715.

Goffman, E (1995). *The presentation of self in everyday life*. New York: Anchor Books, p. 251. ISBN: 978-0-14-013571-8. URL: http://www.amazon.co.uk/Presentation-Self-Everyday-Penguin-Psychology/dp/0140135715.

Google Inc. (2011). *Google Product Forum*. URL: http://productforums.google.com/forum/{\#}!topic/analytics/AHzZOWUBTuE (visited on 01/29/2013).

— (2012). *Webmaster Guidelines*. URL: http://support.google.com/webmasters/bin/answer.py?hl=en{\&}answer=35769 (visited on 01/26/2013).

Gotze, John and Christian Pederson (2009). *State of the Eunion: Government 2.0 and Onwards*. AuthorHouse, p. 332. ISBN: 1449047297. URL: http://books.google.com/books?hl=en{\&}lr={\&}id=oeEBnVwmEmYC{\&}pgis=1.

Government of Canada (2003). *Knowledge Management in the Military Context*. URL: http://www.journal.forces.gc.ca/vo4/no1/command-ordre-02-eng.asp.

Gov.UK (2015). *2010 to 2015 Government Policy: Government Transparency and Accountability*.

Grieve, John (June 2015). *Historical perspective: British policing and the democratic ideal*. Ed. by Paresh Wankhade and David Weir.

Guerry, André-Michel and Silvestre (1833). *Essai sur la statistique morale de la France (Google eBook)*. Crochard, p. 69. URL: http://books.google.com/books?id=u3nro2gPONQC{\&}pgis=1.

*Guide to key performance indicators Communicating the measures that matter* *connectedthinking pwc* (2007). Tech. rep. URL: https://www.pwc.com/gx/en/audit-services/corporate-reporting/assets/pdfs/uk{_}kpi{_}guide.pdf.

Guilfoyle, S. (2012). "On Target?–Public Sector Performance Management: Recurrent Themes, Consequences and Questions". In: *Policing* 6.3, pp. 250–260. ISSN: 1752-4512. DOI: 10.1093/police/pas001. URL: https://academic.oup.com/policing/article-lookup/doi/10.1093/police/pas001.

— (2015). "Binary Comparisons and Police Performance Measurement: Good or Bad?" In: *Policing* 9.2, pp. 195–209. ISSN: 1752-4512. DOI: 10.1093/police/pav004. URL: https://academic.oup.com/policing/article-lookup/doi/10.1093/police/pav004.

Guilfoyle, Simon (2011). *Intelligent Policing*. URL: http://www.triarchypress.net/intelligent-policing.html (visited on 02/22/2014).

Hacking, Ian (1990). *The Taming of Chance | History of ideas and intellectual history | Cambridge University Press*. Cambridge University Press. ISBN: 9780521388849. URL: http://www.cambridge.org/us/academic/subjects/history/history-ideas-and-intellectual-history/taming-chance.

Hara, Kieron O and Wendy Hall (2012). "Web Science and Reflective Practice". In:

Hart, Glen, Robin Frew, and Maire Byrne (2018). "Place Geography: Finding a Space for Place". In: *GISRUK 2018*. Leicester.

Heider, Fritz (1958). *The Psychology of Interpersonal Relations*. Psychology Press, p. 322. ISBN: 0898592828. URL: http://books.google.co.uk/books/about/The{_}Psychology{_}of{_}Interpersonal{_}Relation.html?id=Zh6TDmayLOAC{\&}pgis=1.

Hill, Kashmir (2016). *How an internet mapping glitch turned a random Kansas farm into a digital hell*. URL: https://splinternews.com/how-an-internet-mapping-glitch-turned-a-random-kansas-f-1793856052 (visited on 05/12/2018).

Holbrook, Allyson L., Jon A. Krosnick, and Alison Pfent (2008). "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms". In: *Advances in Telephone Survey Methodology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 499–528. ISBN: 9780470173404. DOI: 10.1002/9780470173404.ch23. URL: http://doi.wiley.com/10.1002/9780470173404.ch23.

Hood, Christopher (1995). "The " New Public Management " in the 1980s: Variations on a Theme'". In: *Accounting, Organisations and Society* 20.U3, pp. 93–109. URL: https://pdfs.semanticscholar.org/2807/1401574fd01fb00345dcba852f216a825e37.pdf.

House of Commons and Home Affairs Committee (2013). *Police and Crime Commissioners: power to remove Chief Constables Sixth Report of Session 2013–14*. Tech. rep. URL: http://www.publications.parliament.uk/pa/cm201314/cmselect/cmhaff/487/487.pdf.

Information Commissioner's Office (2010). *Crime-mapping and geo-spatial crime data: privacy and transparency Data Protection Act*. Tech. rep. London: Information Commissioner's Office.

Innes, Martin (2004). "Signal crimes and signal disorders: notes on deviance as communicative action." In: *The British journal of sociology* 55.3, pp. 335–55. ISSN: 0007-1315. DOI: 10.1111/j.1468-4446.2004.00023.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/15383091.

Jaworski, Adam and Nikolas Coupland (2006). *The discourse reader*. Routledge, p. 560. ISBN: 9780415346320.

Jenkins, Simon (2011). "All the crime map shows up is Whitehall's pointless zest for data | Simon Jenkins | Comment is free | The Guardian". In: *The Guardian*. URL: http://www.theguardian.com/commentisfree/2011/feb/03/crime-map-information-theresa-may.

John Stuart Mill (1886). *A system of logic ratiocinative and inductive : being a connected view of the principles of evidence and the methods of scientific investigation*. London: Longmans, Green. URL: http://www.worldcat.org/title/system-of-logic-ratiocinative-and-inductive-being-a-connected-view-of-the-principles-of-evidence-and-the-methods-of-scientific-investigation/oclc/234155179?page=citation.

Johnson, Lyndon (1966). *Statement by the President Upon Signing the Freedom of Information Act*. URL: https://nsarchive2.gwu.edu//nsa/foia/FOIARelease66.pdf (visited on 05/10/2018).

Johnston, Les (2003). "From pluralisation to the police extended family discourses on the governance of community policing in Britain". In: *International Journal of the Sociology of Law* 31.3, 185–204. ISSN: 01946595. DOI: 10.1016/j.ijsl.2003.09.003.

Jones, Trevor, Brian MacLean, and Jock Young (1986). *The Islington Crime Survey: Crime, Victimization and Policing in Inner-City London, Volume 1*. Gower, p. 265. ISBN: 0566052644. URL: http://books.google.co.uk/books/about/The{_}Islington{_}Crime{_}Survey.html?id=hY3aAAAAMAAJ{\&}pgis=1.

Jurafsky Daniel and Martin James (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (PDF Download Available)*.

Karmen, Andrew (2012). *Crime Victims: An Introduction to Victimology*. Cengage Learning, p. 560. ISBN: 1133049729. URL: http://www.amazon.com/Crime-Victims-An-Introduction-Victimology/dp/1133049729.

Kelling, George L. and James Q Wilson (1982). *Broken Windows*. URL: http://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/ (visited on 07/19/2013).

Keohane, Robert O. (Robert Owen) (2005). *After Hegemony : Cooperation and Discord in the World Political Economy*. Princeton University Press, p. 290. ISBN: 9780691122489. URL: https://press.princeton.edu/titles/1322.html.

Kieron O'Hara (2010). *Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*.

Kirby, Dick. (2007). *You're Nicked! : Further Memoirs from the Real Sweeney on Life in the Serious Crime Squad*. Robinson, p. 300. ISBN: 1845294637.

Klein, Gary, Brian Moon, and Robert R. Hoffman (2006). "Making Sense of Sensemaking 2: A Macrocognitive Model". In: *IEEE Intelligent Systems* 21.5, pp. 88–92. ISSN: 1541-1672. DOI: 10.1109/MIS.2006.100. URL: http://dl.acm.org/citation.cfm?id=1175882.1176017.

Kolko, Jon (2010). *Jon Kolko » Sensemaking and Framing: A Theoretical Reflection on Perspective in Design Synthesis*. URL: http://www.jonkolko.com/writingSensemaking.php (visited on 03/11/2018).

Kopytoff, Verne (2014). *This Space for Rent: Private Satellites Raise Profits, Privacy Concerns*.

Kroes, Neelie (2013). *The Economic and socialbenefits of big data*. URL: http://www.nessi-europe.com/?page=Newsdetails{\&}ListID=2{\&}RowID=426.

Kula, Witold (1986). *Measures and Men*. URL: http://press.princeton.edu/titles/2321.html (visited on 10/01/2014).

Lancia, Franco (2005). "Word Co-occurrence and Theory of Meaning". In: URL: http://www.soc.ucsb.edu/faculty/mohr/classes/soc4/summer{_}08/pages/Resources/Readings/TheoryofMeaning.pdf.

Lantolf, James P. (2006). "Alan Davies and Catherine Elder (eds): The Handbook of Applied Linguistics. Blackwell, 2004." In: *Applied Linguistics* 27.1, pp. 147–152. ISSN: 1477-450X. DOI: 10.1093/applin/ami047.

Latour, Bruno and Emilie Hermant (1998). *Paris: Invisible City Paris ville invisible. Paris: La Découverte-Les Empêcheurs de penser en rond*. Ed. by Liz Carey-Libbrecht and Valerie Pihet. URL: http://www.bruno-latour.fr/sites/default/files/downloads/viii{_}paris-city-gb.pdf.

Lazer, David et al. (2009). "socialscience. Computational socialscience." In: *Science (New York, N.Y.)* 323.5915, pp. 721–3. ISSN: 1095-9203. DOI: 10.1126/science.1167742. URL: http://www.sciencemag.org/content/323/5915/721.

*LBJ Presidential Library | Research* (1966). URL: http://www.lbjlibrary.net/collections/on-this-day-in-history/july.html (visited on 06/09/2019).

Lee, Elissa and Laura Leets (2002). "Persuasive Storytelling". In: *American Behavioral Scientist* 45.6, pp. 927–957.

Legislation.gov.uk (2011). *Police Reform and socialResponsibility Act 2011*. URL: http://www.legislation.gov.uk/ukpga/2011/13/introduction.

Leibler, Anat (2004). *Statisticians' Ambition: Governmentality, Modernity and National Legibility*. URL: http://www.sts-biu.org/images/file/IsraelStudies.pdf (visited on 10/01/2014).

Leigh Star, Susan (1998). *Grounded Classification: Grounded Theory and Faceted Classification*.

Lessig, Lawrence (2009). *Against Transparency | New Republic*. URL: https://newrepublic.com/article/70097/against-transparency (visited on 09/03/2017).

Lewontin, Richard C. (1995). *Sex, Lies, and Social Science*. (Visited on 11/02/2014).

Lippmann, Walter (1932). *Public Opinion*. Transaction Publishers, p. 427. ISBN: 1412832403. URL: http://books.google.com/books?hl=en{\&}lr={\&}id=YhXLOVc6BsoC{\&}pgis=1.

Loader, Ian (1997). "Thinking Normatively About Private Security". In: *Journal of Law and Society* 24.3, pp. 377–394. ISSN: 0263323X. DOI: 10.1111/j.1467-6478.1997.tb00003.x. URL: http://doi.wiley.com/10.1111/j.1467-6478.1997.tb00003.x.

Longo, Justin (2013). *Open Government - What's in a Name? - The Governance Lab @ NYU*. URL: http://thegovlab.org/open-government-whats-in-a-name/ (visited on 06/09/2019).

Lopez, Ben. (2011). *The Negotiator : My Life at the Heart of the Hostage Trade*. Sphere, p. 304. ISBN: 1616088621. URL: https://books.google.co.uk/books/about/The{_}Negotiator.html?id=jfRDCgAAQBAJ{\&}redir{_}esc=y.

Loveday, Barry and Anna Reid (2003). *Going local Who should run Britain's police?* Tech. rep. London: Policy Exchange. URL: http://www.port.ac.uk/departments/academic/icjs/staff/documentation/filetodownload,76100,en.pdf.

Luckerhoff, Jason and François Guillemette (1990). "The Conflicts between Grounded Theory Requirements and Institutional Requirements for Scientific Research". In: *The Qualitative Report* 16.2. ISSN: 1052-0147. URL: http://nsuworks.nova.edu/tqr/vol16/iss2/5.

Lynch, Shana (2014). *The Agency Cannot Survive Without Being More Transparent*. URL: https://www.gsb.stanford.edu/insights/former-nsa-head-michael-hayden-agency-cannot-survive-without-being-more-transparent (visited on 05/09/2018).

Lynch, William (2001). *Solomon's child : method in the early Royal Society of London*. Stanford University Press, p. 292. ISBN: 0804732914. URL: https://epdf.pub/solomons-child-method-in-the-early-royal-society-of-london-writing-science.html.

Lyon, David (2008). "Surveillance Society". In: *Business*.

— (2014). "Surveillance, Snowden, and Big Data: Capacities, consequences, critique". In: *Big Data & Society* 1.2, p. 205395171454186. ISSN: 2053-9517. DOI: 10.1177/2053951714541861. URL: http://journals.sagepub.com/doi/10.1177/2053951714541861.

Macintosh, Ann and Angus Whyte (2008). "Towards an evaluation framework for eParticipation". In: *Transforming Government: People, Process and Policy* 2.1, pp. 16–30. ISSN: 1750-6166. DOI: 10.1108/17506160810862928. URL: http://www.emeraldinsight.com/doi/10.1108/17506160810862928.

Manis, Jim (2003). *State of the Union Addresses by Calvin Coolidge*. Pennsylvania. URL: http://www2.hn.psu.edu/faculty/jmanis/poldocs/uspressu/SUaddressCCoolidge.pdf.

Manning, Peter K. (2011a). *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control (New Perspectives in Crime, Deviance, and Law)*. NYU Press, p. 338. ISBN: 0814761364. URL: http://www.amazon.com/The-Technology-Policing-Information-Perspectives/dp/0814761364.

— (2011b). *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control (New Perspectives in Crime, Deviance, and Law)*.

NYU Press, p. 338. ISBN: 0814761364. URL: http://www.amazon.com/The-Technology-Policing-Information-Perspectives/dp/0814761364.

Mayhew, Henry (2009). *London Labour and the London Poor: A Cyclopaedia of the Condition and Earnings of Those That Will Work, Those That Cannot Work, and Those That Will Not Work*. Cosimo, Inc., p. 536. ISBN: 1605207330.

McRae, Ken et al. (2005). "Semantic feature production norms for a large set of living and nonliving things". In: *Behavior Research Methods* 37.4, pp. 547–559. ISSN: 1554-351X. DOI: 10.3758/BF03192726. URL: http://www.springerlink.com/index/10.3758/BF03192726.

Mill, JS (1861). *Considerations on Representative Government*. Parker, Son and Bourn, p. 340. URL: http://ebooks.adelaide.edu.au/m/mill/john{_}stuart/m645r/contents.html.

Ministry of Justice (2012). *Swift and Sure Justice: The Government's Plans for Reform of the Criminal Justice System*. London.

Monmonier, Mark S. (2002). *Spying with maps : surveillance technologies and the future of privacy*. University of Chicago Press, p. 239. ISBN: 0226534286.

Montgomery, Alan et al. (2002). "Modeling Online Browsing and Path Analysis Using Clickstream Data". In: *Marketing Science* 23.4, pp. 579–595.

Mortensen, E (2006). *Sex, Breath, and Force: Sexual Difference in a Post-feminist Era*. Lexington Books, p. 179. ISBN: 0739114670. URL: http://books.google.com/books?id=S18{_}wW0d2uAC{\&}pgis=1.

Moya K. Mason (2017). *Foucault and His Panopticon*. URL: http://www.moyak.com/papers/michel-foucault-power.html.

Mullen, Chris (1996). *Map menu, John Ogilby's Britannia Part One entire 1775*. URL: https://www.fulltable.com/vts/m/map/ogilby/mna.htm (visited on 05/19/2019).

Munro, D. C. (1925). "The Geographical Lore of the Time of the Crusades: a Study in the History of Medieval Science and Tradition in Western Europe". In: *The American Historical Review* 30.4, pp. 801–803. ISSN: 1937-5239. DOI: 10.1086/ahr/30.4.801. URL: https://academic.oup.com/ahr/article/30/4/801/24012/The-Geographical-Lore-of-the-Time-of-the-Crusades.

NCCGroup (2018). *Reflections on CyberUK*. Manchester. URL: https://www.nccgroup.trust/uk/about-us/newsroom-and-events/blogs/2018/april/reflections-on-cyberuk/.

Neocleous, M. (2000). "socialPolice and the Mechanisms of Prevention". In: *British Journal of Criminology* 40.4, pp. 710–726. ISSN: 00070955. DOI: 10.1093/bjc/40.4.710. URL: https://academic.oup.com/bjc/article-lookup/doi/10.1093/bjc/40.4.710.

Newburn, Tim (2003). *Handbook of Policing*. Willan, p. 864. ISBN: 1843923238. URL: http://www.amazon.co.uk/Handbook-Policing-Tim-Newburn/dp/1843923238.

Nic Brunetti (2013). *Police and Crime Panels a Complete Waste of Time*. URL: http://www.policeoracle.com/news/Police+Staff/2013/Aug/12/Police-and-crime-panels-a-complete-waste-of-time{_}69207.html (visited on 08/14/2013).

Nicolas, Laurent (2012). *Information is the New Oil*. URL: www.thenewfederalist.eu.

Niglas, Katrin. *Paradigms and methodology in educational research*. URL: http://www.leeds.ac.uk/educol/documents/00001840.htm.

Nominet Trust (2012). *Open Data and Charities*. URL: http://www.nominettrust.org.uk/knowledge-centre/articles/open-data-and-charities (visited on 11/01/2014).

Norgate, Jean and Martin Norgate (1998). *Ogilby's Road Maps in Hampshire, 1675*. URL: http://www.geog.port.ac.uk/webmap/hantsmap/hantsmap/ogilby/ogilby1.htm{\#}intro (visited on 07/21/2013).

Noveck, Beth (2012). *Demand a More Open-Source Government*.

OECD (2001). *OECD Handbook on information, consultation and public participation in policy-making Citizens as Partners*. URL: https://www.internationalbudget.org/wp-content/uploads/Citizens-as-Partners-OECD-Handbook.pdf.

Office for National Statistics (2019a). *Harmonisation within the GSS - Office for National Statistics*. URL: https://www.ons.gov.uk/methodology/classificationsandstandards/harmonisationwithinthegss (visited on 06/02/2019).

— (2019b). *User Guide to Crime Statistics for England and Wales - Office for National Statistics*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales (visited on 06/02/2019).

O'Hara, Kieron (2013). *Open Data Comes to Market*. URL: http://eprints.soton.ac.uk/350043/1/OpenDataComestoMarketreportfinal.pdf.

O'Hara, Kieron and Nigel Shadbolt (2008). *The Spy in the Coffee Machine: The End of Privacy as We Know It*. Oneworld Publications. ISBN: 1851685545.

Olanoff, Drew (2012). *FearSquare lets UK Foursquare Users Know if They're in Danger*. URL: https://thenextweb.com/uk/2012/05/06/fearsquare-alerts-uk-foursquare-users-to-crime-near-the-venues-that-they-visit-often/ (visited on 06/04/2019).

Open Knowledge Foundation (2012). *The Open Data Handbook*. URL: http://opendatahandbook.org/.

— (2013). *Open Data – An Introduction*. URL: http://okfn.org/opendata/ (visited on 08/14/2013).

Page, Lewis (2011). *Home Office Crime Maps go to Street-Level Detail*. URL: http://www.theregister.co.uk/2011/02/01/home{_}office{_}crime{_}maps/ (visited on 07/25/2013).

Park, Robert and Burgess, Ernest and McKenzie, Roderick (1925). *The City*.

Patriarca, Silvana (1996). *Numbers and Nationhood Writing Statistics in Nineteenth Century Italy*. Cambridge University Press, pp. 205–227. URL: http://catdir.loc.gov/catdir/samples/cam034/95004328.pdf.

Paynich, Rebecca and Bryan Hill (2011). *Fundamentals of Crime Mapping: Principles and Practice*. Vol. 2011. Jones & Bartlett Publishers, p. 552. ISBN: 1449667953. URL: http://books.google.com/books?id=BK3Zs-XXso4C{\&}pgis=1.

Peters, Gerhard and John T. Woolley (2007). *Rudy Giuliani: Press Release - CompStat: Measuring Crime To Reduce It - May 30, 2007*. URL: http://www.presidency.ucsb.edu/ws/?pid=94811 (visited on 08/26/2013).

Piotrowski, Suzanne J and Gregg G. Van Ryzin (2007). "Citizen Attitudes Toward Transparency in Local Government". In: *The American Review of Public Administration* 37.3, pp. 306–323. ISSN: 0275-0740. DOI: 10.1177/0275074006296777. URL: http://arp.sagepub.com/http://journals.sagepub.com/doi/10.1177/0275074006296777.

Pollack, Doug (2010). *High Unemployment Increases Cybercrime | ID Experts*. URL: https://www2.idexpertscorp.com/knowledge-center//single/high-unemployment-increases-cybercrime (visited on 05/12/2018).

Polonetsky, Jules et al. (2013). *Privacy and Big Data: Making Ends Meet*. URL: http://www.stanfordlawreview.org/online/privacy-and-big-data.

Pope-Davis, Donald B. and WM Liu (2001). "What's Missing from Multicultural Competency Research: Review, Introspection, and Recommendations." In: *psycnet.apa.org*. URL: http://psycnet.apa.org/journals/cdp/7/2/121/.

Pope-Davis, Donald B. et al. (2002). "Client Perspectives of Multicultural Counseling Competence". In: *The Counseling Psychologist* 30.3, pp. 355–393. ISSN: 0011-0000. DOI: 10.1177/0011000002303001. URL: http://journals.sagepub.com/doi/10.1177/0011000002303001.

Porter, Theodore (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. URL: http://press.princeton.edu/titles/5653.html.

Prakash, Gyan (1999). *Prakash, G.: Another Reason: Science and the Imagination of Modern India. (Paperback)*. Princeton University Press. ISBN: 9780691004532. URL: http://press.princeton.edu/titles/6705.html.

Public Administration Committee (2014). *Caught Red-Handed: Why We Can't Count on Police Recorded Crime Statistics*.

Punch, Maurice (2007). *Zero Tolerance Policing (Researching Criminal Justice Series)*. Policy Press, p. 64. ISBN: 1847420559. URL: http://www.amazon.co.uk/Tolerance-Policing-Researching-Criminal-Justice/dp/1847420559.

Quetelet, A, D. (1866). *Sciences mathématiques et physiques chez les Belges: au commencement du XIXe siècle : Adolphe Quételet : Free Download & Streaming : Internet Archive*. H. Thiryvan Buggenhoudt. URL: https://archive.org/details/sciencesmathmat00qugoog.

Ranganathan, S R (1951). "Classification and Communication". In: *Five Laws of Library Science*.

Rawson, Rawson W. (1839). "An Inquiry into the Statistics of Crime in England and Wales". In: *Journal of the Statistical Society of London* 2.5, p. 316. ISSN: 09595341. DOI: 10.2307/2337821. URL: http://www.jstor.org/stable/10.2307/2337821?origin=crossref.

Regan, Priscilla M (1995). *Legislating Privacy*. Chapel Hill & London: The University of North Carolina Press. ISBN: 0-8078-2226-4.

Rein, Martin and Donald Schön (1996). "Frame-critical policy analysis and frame-reflective policy practice". In: *Knowledge and Policy* 9.1, pp. 85–104. ISSN: 0897-1986. DOI: 10.1007/BF02832235. URL: http://link.springer.com/10.1007/BF02832235.

Reith, Charles (1938). *The police idea, its history and evolution in England in the eighteenth century and after*. URL: http://openlibrary.org/works/OL1163590W/The{_}police{_}idea{_}its{_}history{_}and{_}evolution{_}in{_}England{_}in{_}the{_}eighteenth{_}century{_}and{_}after (visited on 08/24/2013).

Rennie, A (1988). *Grounded Theory: a Promising Approach to Conceptualization in Psychology?* URL: http://www.safranlab.net/uploads/7/6/4/6/7646935/rennie.{_}grounded{_}theory.{_}88.pdf.

Resnik, David B. (2005). "Openness versus Secrecy in Scientific Research". In: *Episteme* 2 (03), pp. 135–147. URL: http://philpapers.org/rec/RESOVS.

Richie, Beth Sperber et al. (1997). "Persistence, Connection, and Passion: A Qualitative Study of the Career Development of Highly Achieving African American-Black and White Women." In: *Journal of Counseling Psychology* 44.2, pp. 133–148. ISSN: 0022-0167. DOI: 10.1037/0022-0167.44.2.133. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0167.44.2.133.

Rivera, Luis Guillermo Solis (2015). *Open Government, a Catalyst For Democracy*. URL: http://www.huffingtonpost.com/president-luis-guillermo-solis-rivera/open-government-a-catalys{_}b{_}8390152.html (visited on 09/03/2017).

Rothstein, Tom (2013). *History of Cartography: Volumes One, Two, and Three*. URL: http://www.press.uchicago.edu/books/HOC/index.html (visited on 08/25/2013).

Sampson, Robert J. and W. Byron Groves (1989). “Community Structure and Crime: Testing SoC.I.A.l-Disorganization Theory”. In: *American Journal of Sociology* 94.4, p. 774. ISSN: 0002-9602. DOI: 10.1086/229068. URL: http://www.journals.uchicago.edu/doi/abs/10.1086/229068.

Schmidt, Manfred G (2002). “Political performance and types of democracy: Findings from comparative studies”. In: *European Journal of Political Research* 41, pp. 147–163. URL: https://sites.hks.harvard.edu/fs/pnorris/Acrobat/stm103articles/schmidt{_}Types{_}of{_}demo{_}and{_}perf.pdf.

Scorgie, Michael E. (1995). “Patrick Colquhoun”. In: *Abacus* 31.1, pp. 93–112. ISSN: 0001-3072. DOI: 10.1111/j.1467-6281.1995.tb00356.x. URL: http://doi.wiley.com/10.1111/j.1467-6281.1995.tb00356.x.

Scott, James C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, p. 445. ISBN: 0300078153. URL: http://books.google.co.uk/books/about/Seeing{_}Like{_}a{_}State.html?id=PqcPCgsr2u0C{\&}pgis=1.

Seale, Clive et al. (2006). *Qualitative Research Practice: Concise Paperback Edition*. Vol. 2006. SAGE, p. 552. ISBN: 1446204588. URL: http://books.google.com/books?id=P5XurpLU{_}H4C{\&}pgis=1.

Shadbolt, Nigel (2011). *SOC.I.A.M: The Theory and Practice of socialMachines*. URL: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/J017728/1.

Shaw, Clifford R. (1930). “Delinquency Areas”. In: *Journal of the American Institute of Criminal Law and Criminology* Vol. 21.No. 2, pp. 308–311.

Singh, Simon. (2000). *The code book : the secret history of codes and code-breaking*. Fourth Estate, p. 402. ISBN: 1857028899.

Solove, Daniel (2011). *Nothing to Hide*. Yale University Press.

Stoker, Kevin and Brad Rawlins (2005). “The "Light" of Publicity in the Progressive Era From Searchlight to Flashlight”. In: *Journalism History* 30.4, pp. 177–188. URL: https://www.researchgate.net/publication/273439997{_}The{_}Light{_}of{_}Publicity{_}in{_}the{_}Progressive{_}Era{_}From{_}Searchlight{_}to{_}Flashlight.

Stone, Deborah A (1989). “Causal Stories and the Formation of Policy Agendas”. In: *Political Science Quarterly* 104.2, pp. 281–300. URL: http://www.jstor.org/stable/2151585.

Storch, Robert D. and F. Engels (2008). “The Plague of the Blue Locusts”. English. In: *International Review of socialHistory* 20.01, p. 61. ISSN: 0020-8590. DOI: 10.1017/S0020859000004843. URL: http://journals.cambridge.org/abstract{_}S0020859000004843.

Strauss, Anselm L. and Juliet M. Corbin (1998). *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*. Sage Publications, p. 312. ISBN: 0803959400. URL: https://books.google.co.uk/books/about/Basics{_}of{_}Qualitative{_}Research.html?id=wTwYUnHYsmMC.

Stross, Charles. (2012). *Rule 34*. Orbit, p. 372. ISBN: 1841497746.

Superintendent, Chief and Irene Curtis (2015). *The Use of Targets in Policing - Review 2015*. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment{_}data/file/466058/Review{_}Targets{_}2015.pdf.

Surel, Yves (2000). "The role of cognitive and normative frames in policy-making". In: *Journal of European Public Policy* 7.4, pp. 495–512. ISSN: 1350-1763. DOI: 10.1080/13501760050165334. URL: http://www.tandfonline.com/doi/abs/10.1080/13501760050165334.

Sutherland, Edwin Hardin, Donald Ray Cressey, and David F. Luckenbill (1992). *Principles of Criminology*. Rowman & Littlefield, p. 696. ISBN: 0930390695. URL: http://books.google.co.uk/books/about/Principles{_}of{_}Criminology.html?id=wqRQqXKuU7sC{\&}pgis=1.

The Cabinet Office (2012). *Open Data - White Paper Unleashing the Potential*. London.

— (2013). *Open Government Partnership UK National Action Plan 2013 to 2015*. URL: https://www.gov.uk/government/consultations/open-government-partnership-uk-national-action-plan-2013/open-government-partnership-uk-national-action-plan-2013-to-2015 (visited on 04/15/2018).

The Home Office (2011). *A New Approach to Fighting Crime*. Tech. rep. London: The Home Office. DOI: isbN978-1-84987-401-4Ho_01762_G. URL: http://www.homeoffice.gov.uk/publications/crime/new-approach-fighting-crime?view=Binary.

The Old Bailey (2013). *London History - A Population History of London - Central Criminal Court*. URL: http://www.oldbaileyonline.org/static/Population-history-of-london.jsp (visited on 07/21/2013).

The Open University (2011). *The Quantitative Qualitative Debate A False Dichotomy*. URL: http://www.thejoyofconcrete.org/students/dse212/quantqualdebate.pdf.

The Pleasant Progressive (2016). *Connectedness, Digilantism, and Trauma*. URL: http://fortysevenseventyeight.wordpress.com/2013/05/16/connectedness-digilantism-and-trauma/ (visited on 08/27/2013).

The Russell Group (2013). *House of Commons - Justice: Written Evidence from the Russell Group*. URL: https://publications.parliament.uk/pa/cm201213/cmselect/cmjust/96/96we03.htm (visited on 09/03/2017).

The Web Foundation (2015). *Open Data Barometer Highlights the Need for Governments to Increase Open Data Efforts*. URL: http://webfoundation.org/2015/01/open-data-barometer-second-edition/.

The White House (2009). *Transparency and Open Government | The White House*. URL: http://www.whitehouse.gov/the{_}press{_}office/TransparencyandOpenGovernm (visited on 08/26/2013).

— (2011). *The Obama Administration's Commitment to Open Government: A Status Report*. URL: http://www.whitehouse.gov/sites/default/files/opengov{_}report.pdf.

Torpey, John (2000). *The Invention of the Passport: Surveillance, Citizenship and the State*. Cambridge University Press, p. 211. ISBN: 0521634938.

Tversky, Amos and Daniel Kahneman (1986). "Rational Choice and the Framing of Decisions". In: *The Journal of Business The Behavioral Foundations of Economic Theory* 59.2, pp. 251–278. URL: http://links.jstor.org/sici?sici=0021-9398{\%}28198610{\%}2959{\%}3A4{\%}3CS251{\%}3ARCATFO{\%}3E2.0.CO{\%}3B2-C.

UK Statistics Authority (2014). *Exposure draft of a report from the UK Statistics Authority: Quality Assurance and Audit Arrangements for Administrative Data July 2014*. Tech. rep. London: Uk Statistics Authority.

Urofsky, Melvin and David Levy, eds. (1971). *Letters of Louis D. Brandeis: Volume III, 1913-1915: Progressive and Zionist - Louis D. Brandeis - Google Books*. New York: State University of New York Press, p. 100. URL: https://books.google.co.uk/books?id=Q7bviBd4w18C{\&}pg=PR15{\&}lpg=PR15{\&}dq=brandeis+letter+to+fiancee{\&}source=bl{\&}ots=IcuclMDIy6{\&}sig=ACfU3U1Yzy{_}PwWPevMStXQo3-C1Vxhxpww{\&}hl=en{\&}sa=X{\&}ved=2ahUKEwjlu7KancviAhWTonEKHZ4oAVAQ6AEwCHoECAkQAQ{\#}v=onepage{\&}q=brandeisletter.

U.S.Department of State (2018). *Democracy*. URL: https://www.state.gov/j/drl/democ/ (visited on 05/09/2018).

Van Kleek, Max et al. (2013). *The Crowd Keeps me in Shape: socialPsychology and the Present and Future of Health socialMachines*. URL: http://dl.acm.org/citation.cfm?id=2487788.2488082.

Verloo, M. (2005). "Displacement and Empowerment: Reflections on the Concept and Practice of the Council of Europe Approach to Gender Mainstreaming and Gender Equality". In: *socialPolitics: International Studies in Gender, State & Society* 12.3, pp. 344–365. ISSN: 1072-4745. DOI: 10.1093/sp/jxi019. URL: http://sp.oxfordjournals.org/content/12/3/344.abstract.

Vleugels, Roger (2008). *Overview of all 86 FOIA Countries*. URL: http://www.statewatch.org/news/2008/sep/foi-overview-86-countries-sep-2008.pdf.

Wagner, P, B Wittrock, and R Whitley (1991). *Discourses on Society: The Shaping of the socialScience Disciplines*. Springer Science & Business Media, p. 385. ISBN: 0792310012. URL: http://books.google.com/books?id=ULF5r0lziKUC{\&}pgis=1.

Wall, David (2007). *Cybercrime*. Polity Press.

Walls, Paula, Kader Parahoo, and Paul Fleming (2010). "The role and place of knowledge and literature in grounded theory". In: *Nurse Researcher* 17.4, pp. 8–17. ISSN: 1351-5578. DOI: 10.7748/nr2010.07.17.4.8.c7920. URL: http://www.ncbi.nlm.nih.gov/pubmed/20712230http://rcnpublishing.com/doi/abs/10.7748/nr2010.07.17.4.8.c7920.

Webber, Craig (2009). *Psychology and Crime (Key Approaches to Criminology)*. Los Angeles: SAGE Publications Ltd, p. 224. ISBN: 1412919428. URL: http://www.amazon.co.uk/Psychology-Crime-Key-Approaches-Criminology/dp/1412919428.

Weber, Max, Hans Gerth, and Wright Mills (2009). *From Max Weber : Essays in Sociology*.

Whyte, Angus and Ann Macintosh (2002). "Analysis and Evaluation of E-Consultations". In: *e-Service Journal* 2.1, p. 9. ISSN: 15288226. DOI: 10.2979/esj.2002.2.1.9. URL: https://www.jstor.org/stable/10.2979/esj.2002.2.1.9.

— (2003). "Representational Politics in Virtual Urban Places". In: *Environment and Planning A* 35.9, pp. 1607–1627. ISSN: 0308-518X. DOI: 10.1068/a34237. URL: http://journals.sagepub.com/doi/10.1068/a34237.

Wilson, Ben (2014). *Decency and Disorder : the Age of Cant, 1789-1837*. Faber & Faber, p. 510. ISBN: 0571317200.

www.Parliament.uk (2015). *The Reform Acts and Representative Democracy*. URL: http://www.parliament.uk/about/living-heritage/evolutionofparliament/houseofcommons/reformacts/ (visited on 12/30/2015).

Young, Jock (2011). *The Criminological Imagination*. Cambridge: Polity Press.

Yu, Harlan and David G. Robinson (2012). "The New Ambiguity of 'Open Government'". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2012489. URL: http://papers.ssrn.com/abstract=2012489.

Zimbardo, Philip G. (1969). *The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos.*

Zittrain, Jonathan (2014). *Future of the Internet – And How to Stop it*. URL: http://futureoftheinternet.org/category/future-of-the-internet/ (visited on 02/23/2014).