Data Quality, Government Data and the Open Data Infosphere

Kieron O'Hara[1]

Abstract. This paper discusses ways in which the environment in which data are released affects data quality. Using the example of the release of open government data into an information market populated by open data, a basic assumption of open data, that eyeballs (i.e. widespread scrutiny by a large and diverse population) will help ensure and improve quality, is examined. A case study of data about crime and criminal justice is used to show that various pressures, including the important aim of fostering a user community for the information, can distort the information markets which are the basis for that assumption.

1. INTRODUCTION

In this paper, I examine the issue of data quality from the point of view of the *infosphere* [1], or information ecosystem, in which the data exists or is released. The paper will not discuss aspects of data quality directly, but will rather argue that quality issues are inextricably intertwined with constraints surrounding usage and publication.

Crunching large quantities of data in order to find the weak signals in the noise has become a major industry in the 21st century, with claims that it will enable improvements in science [2], drive economic growth [3] and lead to better public service outcomes [4]. The protocols and standards of the World Wide Web, initially designed to connect documents, are being reshaped by research into in the so-called Semantic Web to link data directly using knowledge representation languages that represent data using URIs [5]. In this world, linking data, or mashing up data from several sources, is widely perceived to increase their value by allowing their serendipitous reuse in unanticipated contexts and juxtapositions.

There are many issues that this world of linked data crunching raises, for example with respect to provenance, ontological alignment, privacy, data protection and intellectual property. One of the most pressing is that of quality; in the linked data vision, data are brought together from heterogeneous sources. This leads to an obvious issue of trust (trust is one of the upper levels of the Semantic Web protocol stack, for instance [6]); if data processors cannot trust that a dataset is of sufficient quality, they will be reluctant to mash it up with datasets already in use. Quality issues here include such matters as accuracy, timeliness, reliability, consistency of semantics and representation (particularly with time-based series) and format; these issues apply not only to data, but to metadata as well. And if there are no general means to assure quality, then such reluctance will become systemic, leading to severe opportunity cost – all the more galling given the hype surrounding so-called 'supercrunching'.

One approach to data quality involves *eyeballs* – if enough people examine data, improvements can be crowdsourced. This brings in another ideology of the big data era – *open data*. The idea of open data is that, if big data and data sharing are so promising, then following the logic through it makes sense to release datasets to as many people as possible. The obvious way to do this is to remove as many legal and technical restrictions as possible. Open data have three principal characteristics: (i) they are available online for download; (ii) they are machine-readable; and (iii) they are held under an unrestricted licence (databases are normally subject to copyright-like rights for their owners, which are waived for open data). Ideally, open data will be in open knowledge representation formats; pdf is very restrictive, and requires documents to be scraped for data, while Excel or Word are proprietary. Better are open formats like CSV, while even more ideal would be open, linkable formats such as RDF [7]. The connection with the eyeballing idea is clear – the more open the data, the more eyeballs will come to rest upon them.

We can see, therefore, a hopeful narrative for data quality: open data => extensive critical analysis => crowdsourced data improvement. Even if datasets released are not of the best quality as they are put online, data users and data subjects will soon provide corrections.

In this paper, I shall examine this narrative critically, in the context of a case study of open data. The paper has the following structure. In section 2, I shall expand on the role of open government data within the open data world. Section 3 expands on the connection between open data and data quality, while section 4 looks at the relevant properties of the infosphere into which open data are released, taking as a case study the release of data pertaining to crime and criminal justice in the United Kingdom. Section 5 then considers the interplay of that real-world infosphere and the assumptions of open data. Section 6, a discussion, completes the paper.

2. OPEN DATA AND OPEN GOVERNMENT DATA

Open data are available to all, with as few legal or technological impediments as possible. In particular, they are open for reuse for any purpose whatever, good or bad. In theory, they will enable greater innovation in knowledge products and service provision. Current practice of keeping data in silos means that products and services cannot easily be developed to place such data in useful contexts outside the silos. Yet many application areas require data of many types for a full description, from

[1] Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom, kmo@ecs.soton.ac.uk.

scientific areas (e.g. climate change, drug design, epidemiology) to the social and political.

The scientific benefits of sharing data seem clear [8]. In non-scientific contexts, it is not expected that citizens/consumers will consume open data directly. If we take service provision as an example, the role of open data is to feed into such services, enabling entrepreneurs to create innovative applications which use the data (*apps*), which are in turn consumed by citizens, organisations, community groups, media analysts and so on. The more heterogeneous the mix, the more creative the app is likely to be. An example might be an app that mashes up data about, say, geography, green spaces, real-time traffic flow, anti-social behaviour and accidents, and outputs a healthy and safe bicycle route between two named points. It is hoped that a sufficiently large range of such apps would meet demand for information from diverse sources, countering the centralising tendencies of the mass media.

Open data has distinctly ideological qualities – there is a right way to do it, and it lends itself to campaigning (for example, Tim Berners-Lee's TED talk of 2009). Access control is ruled out, barriers to entry to the infosphere are to be kept as low as possible and reuse is encouraged. However, the whole infosphere need not be open – services could be monetised, or restricted to subscribers. Rent-seeking via data monopolies is ruled out, but if an app is so creative in its use of data that it can support a charge, so be it. Competitors have access to exactly the same (open) data as the app developer, and may be able to reverse-engineer the app, but in the open data economy income comes from creativity, not rents, leading to an increase in the services available to the public.

One particularly important source of open data is *open government data* (OGD) [9]. Government data have a number of qualities that lend them to openness. They are plentiful, of good provenance, of relatively (if not uniformly) decent quality, and describe areas of life and the economy in which people are interested. They therefore have an important potential role in applications which allow people to construct a rich picture of their communities and environment. At present, public services are either provided by governments, or if they are privatised, designed and commissioned by them; the hope is that innovative services can be built on the back of OGD that complement or compete with such centralised service provision. Furthermore, given that governments can only collect data (a) because they are given democratic legitimacy by their citizens, and (b) because they are funded by taxpayers, there is a strong argument that citizens should be allowed the use of at least non-sensitive data (the so-called right to data [10]). It has been argued that transparency has been an important policy tool for many decades [11], and open data is a logical conclusion of that tendency.

Releasing OGD has been an important part of the UK government's information strategy since 2009. The strategy is currently driven by the transparency team in the Cabinet Office backed by explicit commitments from the Prime Minister [12], [13], and is intended to meet a number of policy goals, including transparency/accountability, economic growth, innovative service provision and citizens' right to data. Datasets are released via the data.gov.uk portal, and leading open data campaigners including Berners-Lee sit on the Transparency Board. Administrative overheads are minimal, and datasets are generally covered by the very liberal Open Government Licence. Example apps or websites using OGD can be found at http://data.gov.uk/apps.

3. OPEN DATA AND DATA QUALITY

The relation between open data and data quality has already been mentioned; by releasing data regularly and getting it out into the user community, quality will 'naturally' increase as comments are received from data subjects, app developers and different departments and agencies who can benchmark against each other. With respect to OGD, the government is aware that worries about quality (as with other issues such as privacy) can be used by reluctant civil servants as an excuse to delay or prevent data releases. It is argued by ministers and officials from the Cabinet Office that on the contrary exposure to the developer community will be, in the long run at least, an important way of ensuring quality (for example a point made in a recent interview with minister Francis Maude [14]).

As an example of what eyeballs can do, consider the National Public Transport Access Node database (NaPTAN), which is the national UK Department for Transport record of all points of access to public transport (railway stations, bus stops, ferry ports etc). The locations of many of the more minor access points (particularly bus stops) were incorrectly recorded on the database. However, the release of NaPTAN as open data enabled apps to be developed that visualised the data and presented them on maps which could be inspected by citizens. Given that everyone knows the real location of a handful of bus stops, and that each bus stop is such that someone knows its real location, the accuracy of NaPTAN has been improved, in effect, by crowdsourcing corrections via various services (cf. e.g. http://travelinedata.org.uk/naptanr.htm).

4. THE OPEN DATA INFOSPHERE: CRIME MAPPING

The open data infosphere, as envisaged by the open data ideology, therefore looks something like the structure shown in Figure 1, in which OGD going to data.gov.uk from various government agencies (possibly augmented by crowdsourcing) are filtered competitively by a number of different app developers who create an information market, and which can therefore act as a counterweight to the information provided to the public from the mass media. The main hopes for this infosphere are that government will be made transparent and citizens better informed. As an example, a popular app was the Asborometer (Figure 2), which presented the data about anti-social behaviour orders (ASBOs) in a particular area on one's smartphone. The arrangement shown in Figure 1 will impact on data quality via the pressures from the competition between apps (and possibly via crowdsourced input as well). The question we now need to address is whether this arrangement actually obtains in the real world.

Let us consider the example from the United Kingdom of its releases of data pertaining to crime and criminal justice. The Home Office and the Ministry of Justice release a lot of data in this area (cf. http://www.police.uk/data and http://data.gov.uk/data), some of it on a regular basis to fulfil the Prime Minister's commitment to release "crime data published at a level that allows the public to see what is happening on their streets" [12] and "information on what happens next for crime occurring on their streets, i.e. police action and justice outcomes" [13].

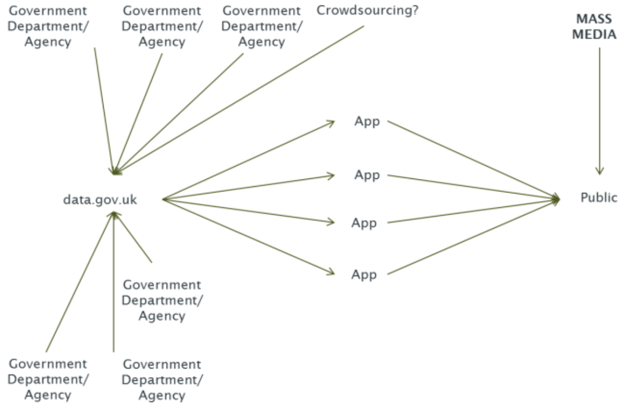
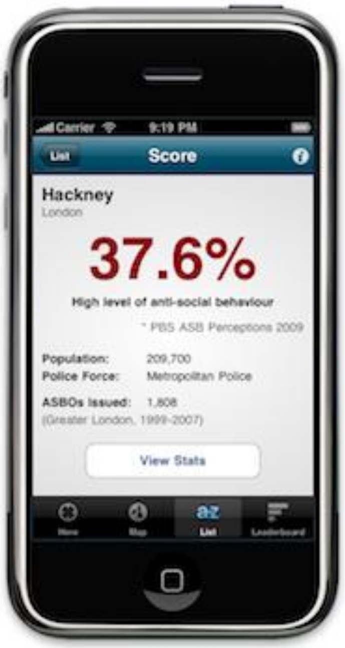

Government Department/ Agency
Government Department/ Agency
Government Department/ Agency
Crowdsourcing?
MASS MEDIA
App
App
App
App
data.gov.uk
Public
Government Department/ Agency
Government Department/ Agency
Government Department/ Agency

Figure 1: The Open Data Infosphere



Hackney
London

37.6%

High level of anti-social behaviour

* PBS ASB Perceptions 2009

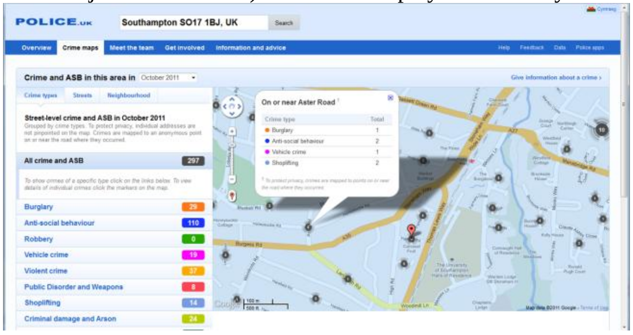
Population:	209,700
Police Force:	Metropolitan Police
ASBOs Issued:	1,808

(Greater London, 1999-2007)

View Stats

Figure 2: The Asborometer

These commitments were premised on the expectation of great public interest in crime and criminal justice outcomes that occur locally. However, the data were necessarily to be released in advance of the development of any apps to convey that information to the public (which certainly has no appetite for downloading and poring over Excel spreadsheets and CSV files). To that end, the Home Office developed its own crime site, http://www.police.uk/ (Figure 3), which initially had three principal aims: (i) to enable cooperation with and accountability of the police force by providing the public with a set of points of contact with their local force and information about its performance, (ii) to foster a constituency of people interested in consuming the data about crime in their area, and (iii) to provide an early conduit for the data into the public domain as soon as the first data releases began in February 2011. Police.uk certainly achieved that final aim, with millions of hits almost immediately [15]. In the intervening period it has remained one of the most popular and used sites powered by OGD. Its main purpose is to provide crime maps in which one can type one's postcode and see the data about crime (and, from June 2012, criminal justice outcomes) in the area displayed intuitively.



Crime and ASB in this area in October 2011

Give information about a crime ›

Street-level crime and ASB in October 2011

Grouped by crime types. To protect privacy, individual addresses are not pinpointed on the map. Crimes are mapped to an anonymous point on or near the road where they occurred.

All crime and ASB	297

To show crimes of a specific type click on the links below. To view details of individual crimes click the markers on the map.

Burglary	29
Anti-social behaviour	110
Robbery	0
Vehicle crime	19
Violent crime	37
Public Disorder and Weapons	8
Shoplifting	14
Criminal damage and Arson	24

On or near Aster Road [1]

Crime type	Total
Burglary	1
Anti-social behaviour	2
Vehicle crime	1
Shoplifting	2

[1] To protect privacy, crimes are mapped to points on or near the road where they occurred.

Figure 3: police.uk in October 2011

It is interesting to recall that prior to the creation of police.uk, there was a great deal of scepticism about releasing crime data, with many arguing that it would increase fear of crime, invade the privacy of victims of crime, or be used by estate agents, insurance companies and criminals to impoverish ordinary citizens. A report by the National Police Improvement Agency published just before police.uk went live [16] argued against many of these fears, but the scepticism was only dispelled after the site went into operation.

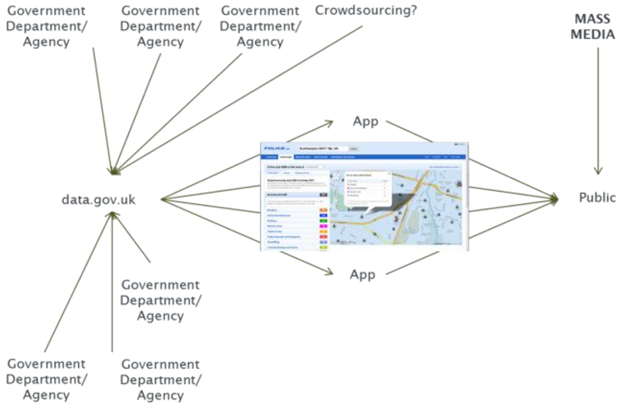

Government Department/ Agency
Government Department/ Agency
Government Department/ Agency
Crowdsourcing?
MASS MEDIA
App
data.gov.uk
Public
App
Government Department/ Agency
Government Department/ Agency
Government Department/ Agency

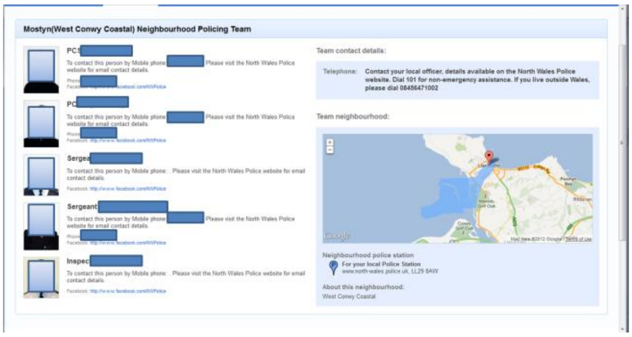
Figure 4: police.uk dominates the infosphere

Hence the actual infosphere in the area of crime and criminal justice data is slightly different from Figure 1, in that the dominant player in the field is the government-sponsored police.uk site, which for coverage and reach far outweighs any competing apps (examples of apps using crime data can be found at http://data.gov.uk/apps, by searching for the tag 'crime') – as represented in Figure 4. Furthermore, Figure 1 also makes the provision of data to data.gov.uk look somewhat more coherent than it actually is – the crime data, for example, is provided separately by 43 different police forces whose data governance varies. The data are brought together and differences smoothed out by the National Police Improvement Agency and their private sector contractors. The connection between crime and criminal justice data is relatively hard to achieve as well, as the two sets of data are kept on different systems. The unique crime number, which could be used to connect the two, has been judged too sensitive to release in the national data, as it often works as a *de facto* identifier for crime victims. The data from police forces, courts, prosecutors, etc., have never been intended to be shared or mashed up, partly because of the operational independence of these various agencies. Other types of police data (e.g. from British Transport Police) are affected by issues of commercial confidence (relating to the companies running stations and rail services).

5. INFOSPHERE VERSUS IDEOLOGY

These divergences of the structure of the infosphere from the ‘ideal’ open data/OGD infosphere shown in Figure 1 have consequences for data quality. In this section, I shall briefly sketch some of those consequences, continuing to use the example of police.uk.

Mapping issues. The first point to be raised is the strong connection between the data and their representation on maps. Maps are very intuitive platforms for making sense of data (particularly mashed-up datasets) [9], [17], and it is unsurprising that crime maps have proven popular with the public. However, there are many different schemas for representing geodata, and the snap points for the data can mean that a crime ‘crosses’ administrative boundaries. There will always be the risk of inaccuracies creeping in; for example, a large proportion of the territory of the Mostyn/West Conwy Coastal neighbourhood team appears to be in the sea (Figure 5). Furthermore, although it makes sense to locate some crime on a map, this is not true of all crime – fraud, identity theft and crimes committed on public transport need not be amenable to geographical location (and in more recent versions of police.uk such crimes are not spatially located). Criminal justice outcomes may not be quite so simple, as crimes can be reclassified at a number of stages through the prosecution process, while other crimes are ‘taken into account’ in a trial.



Mostyn(West Conwy Coastal) Neighbourhood Policing Team

PCS [redacted]
To contact this person by Mobile phone [redacted] Please visit the North Wales Police website for email contact details.
Phone [redacted]
Facebook: [redacted] facebook.com/NWPolice

PC [redacted]
To contact this person by Mobile phone [redacted] Please visit the North Wales Police website for email contact details.
Phone [redacted]
Facebook: http://www.facebook.com/NWPolice

Serge [redacted]
To contact this person by Mobile phone: . Please visit the North Wales Police website for email contact details.
Facebook: http://www.facebook.com/NWPolice

Sergeant [redacted]
To contact this person by Mobile phone [redacted] Please visit the North Wales Police website for email contact details.
Phone [redacted]
Facebook: http://www.facebook.com/NWPolice

Inspec [redacted]
To contact this person by Mobile phone: . Please visit the North Wales Police website for email contact details.
Facebook: http://www.facebook.com/NWPolice

Team contact details:

Telephone: Contact your local officer, details available on the North Wales Police website. Dial 101 for non-emergency assistance. If you live outside Wales, please dial 08456471002

Team neighbourhood:

Neighbourhood police station
For your local Police Station
www.north-wales.police.uk, LL29 8AW

About this neighbourhood:
West Conwy Coastal

Figure 5: The Mostyn (West Conwy Coastal) neighbourhood team

In principle, there should be a separation between the *data* and the *representation* of the data which would render this a null issue. However, given that the same organisation brings out the data and the police.uk site, updating at the same monthly interval, it would be remarkable if the representational requirements of the crime map did not influence the production of the data. For example, if data in time series change in order to meet the new demands of the developing police.uk site, how will that impact on the app developers who are supposed to populate a thriving informational marketplace?

The role of police.uk in the crime data infosphere. A second issue affecting data quality is the role of police.uk in the infosphere (Figure 4). Police.uk has been an undoubted success, and a leading advertisement for the advantages of open data. However, it risks being a victim of its own success, in that the political temptation is always to expand the successful site. The logic of open data suggests that the infosphere should look like Figure 1; the way to get to there from the current position, shown in Figure 4, is clearly to reduce the scope of police.uk – in effect to let it wither as the informational app market thrives with the information-consuming public which police.uk has been central in fostering. The logic of political success, on the other hand, is to expand it. The result is that the information market struggles against a state-backed information supplier. Furthermore, that supplier, by virtue of the close connections between the site developers and the data publishers, tends to get the data first, and so its output is more timely. Finally, it turns out that the ‘eyeballs’ which help improve data quality are often those of the app developers who are competing against police.uk – and hence, although they work to improve quality in their own self-interest, they are simultaneously helping to improve the position of their dominant competitor. This may not always work to their advantage.

Privacy issues. A third issue that has had an effect on data quality is that of privacy and data protection. There are a number of issues to do with open data derived from personal data (personal data is not open data, for obvious reasons), which I have explored elsewhere [18]. The relevant point here is that privacy can impinge on data quality. As police.uk was being developed, the UK Information Commissioner’s Office looked at privacy issues surrounding the display of crimes on a map [19]. The potential problem is that, if the location of a crime is X’s house, then X is identifiable as the victim of that crime, even if not identified directly in the data. After discussions with the ICO, it was decided to take two privacy-preserving measures.

First, addresses are ‘vagued up’ – the PM’s commitments required only that citizens would be aware of crime at street level, so the precise address was not necessary. Hence the snap points on the police.uk map are not exact – they originally covered a minimum of twelve (now eight) postal addresses. It is not known what the average vagueness is (substantially more than eight). This of course impinges on quality by definition, but also there is no metadata to tell the data user how vague the particular location is. Furthermore, quite often the exact location of a crime or anti-social behaviour is an important piece of information – telling the user which street corners to avoid, or allowing a user to argue that, say, the loss of a street light has led to an increase of crimes in a very small area. And not every type of crime has a privacy implication [20].

Secondly, the data are aggregated over a month, and released in arrears. Hence releases are not very timely, and do not allow the user to make important discriminations (whether crimes are committed at night or during the day, what happens at closing time). It is also likely that the lack of timeliness means that it is harder to help the police; if a citizen sees that a crime has been committed in her neighbourhood yesterday, she would be more likely to be able to report suspicious pedestrians or cars in the area, whereas after a lag of up to seven weeks, her recall will obviously be less immediate and accurate.

To summarise, privacy considerations, where relevant, will have an effect on data quality, and those sensitive treatments of privacy that preserve quality as much as possible may require an expensive administrative overhead, compared to the relatively lightweight methods used in police.uk

Inconsistency. A strong connection between the release of an open dataset and a representation of the data such as police.uk can mean that the data are adjusted to the representation as it develops. This can mean that inconsistencies appear, in two ways. First of all, there may be changes over time in a time series, which can make it harder to view data diachronically. The representational issues that affect the data may not have significance for other app developers, but may make it harder to process the data. For example, a parameter may be changed from having an integer value to having a real value; this clearly makes little difference in terms of informational content, but may impact dramatically on the programs that app developers used to process the data. Secondly, there may be changes or improvements to already published datasets. Such changes need to be signalled very clearly to the developer community.

Lack of support. One difficulty in relying on an information market created by what is in effect the cottage industry of app development is that the continuity of information supply to the public (as opposed to continuity of *data* supply to app developers, discussed in the previous paragraph) may be variable. For instance, the Asborometer (Figure 2) was a sensation in 2010, featuring in the *Register*, the *Mail* and the *Telegraph*, and being highlighted by Prime Minister Gordon Brown in a speech about Britain's digital future in March 2010. However, although it remains available at the time of writing, it has not been updated with more recent data and so is very out of date.

Another example from a different area highlights the links between consistency of data provision to app developers and information provision to the public. Schooloscope was a popular and much-lauded app that took schools data and presented it to parents in readable English [21], but it folded in 2011 partly because of the difficulties in maintaining the site, but partly because the quality of data it was taking in was not considered strong. Low quality *data* led to lack of continuity of *information* supply to the public.

6. DISCUSSION

This description of the infosphere for open data in crime and criminal justice is not intended to be critical of the data providers of police.uk or the site itself, which is a very successful, high profile site which provides a lot of information to a public which was until recently starved of it. It corrects many of the assumptions of the more lurid tabloid newspapers. In particular, to recall, one of its main purposes, which it has achieved, is to foster a community of people who regularly use crime data to negotiate their environment.

Nevertheless, the case of police.uk illustrates a pertinent issue about the relation between data quality and the infosphere. In the open data world, data quality is supposed to be upheld or improved by a series of overlapping communities. Data providers benchmark themselves against their fellow providers. App developers need high quality data in order to provide useful and innovative information services to their customers or clients. Data subjects are well-informed at least about the data that concern them. And finally, information consumers are well-informed about their own environment and problems. This particular infosphere, it is hoped, is properly structured and incentivised to provide feedback about quality to data providers. The envisaged structure is something like that shown in Figure 1.

The example of open data about crime and criminal justice shows that the situation is rarely that simple; in that example the government-sponsored and developed police.uk site has a dominant position in the infosphere. This has resulted in a less easily theorised or understood structure, and may affect quality either directly, by providing stiff competition for app developers or by privileging certain types of data representation (maps) which may not always be appropriate, or indirectly. Indirect effects include, for example, the protection of privacy in the data (rather than leaving individual app developers to take their own steps to preserve crime victims' privacy), and the squeezing of the market for new apps (in the absence of a thriving information market, app developers may prefer to further their careers by taking salaried jobs at larger corporations, leading to the lack of support for existing apps over time that was noted in the previous section).

There is an imperative to get the information out into the public domain in a reliable way. The vibrant market of apps as illustrated in Figure 1 may not succeed in achieving that, especially if successes such as the Asborometer do not continue to be supported with timely data. The government therefore has a supportable justification for bringing out a website to present the data in parallel with the datasets that are available to app developers. However, that leads to a spiral, where the large development and maintenance costs of the site need to be justified in political terms, which means the site has to be seen to be successful, which means further development, which means more money spent, which means ….

The open data infosphere of Figure 1 is essentially a privatised development space. The information ecosystem can certainly help to improve data quality (as it does with the Home Office's crime data and the Ministry of Justice's criminal justice data), but those releasing open data must contend with all the uncertainties of a market dominated by small developers. If, as with the crime data, broad dissemination is a key policy aim, it may be that there has to be a compromise with the other laudable aim of crowdsourcing data quality improvement.

The more general conclusion is that the infosphere has a profound effect on data quality, however that is defined, and however data governance is carried out. The way data are used, and the nature of the agents using them, will affect the feedback loops that lead to quality improvements. The case of police.uk illustrates this point; further case studies would be welcome to help provide a more general account of the relationship.

ACKNOWLEDGMENTS

The work reported in this paper was funded by the EnAKTing project, EPSRC Grant EP/G008493/1.

REFERENCES

[1] L. Floridi. *Information: A Very Short Introduction*, Oxford University Press, Oxford (2010).

[2] I. Ayres. *Super Crunchers: How Anything Can Be Predicted*, John Murray, London (2007).

[3] J. Manyika et al. *Big Data: The Next Frontier For Innovation, Competition and Productivity*, McKinsey Global Institute, Washington DC (2011).

[4] M. Wind-Cowie and R. Lekhi. *The Data Dividend*, Demos, London (2012).

[5] N. Shadbolt, T. Berners-Lee and W. Hall. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96-101 (2006).

[6] D. Artz and Y. Gil. A Survey of Trust in Computer Science and the Semantic Web. *Web Semantics*, 5(2):58-71 (2007).

[7] T. Berners-Lee. *Linked Data*, World Wide Web Consortium, http://www.w3.org/DesignIssues/LinkedData.html, (2010).

[8] P. Murray-Rust. Open Data in Science. *Serials Review*, 34(1):52-64 (2008).

[9] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, W. Hall and m.c. schraefel. Linked Open Government Data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, 27(3) (2012).

[10] C. Yiu. *A Right to Data: Fulfilling the Promise of Open Public Data in the UK*. Policy Exchange, London (2012).

[11] A. Fung, M. Graham and D. Weil. *Full Disclosure: The Perils and Promise of Transparency*. Cambridge University Press, New York (2007).

[12] D. Cameron. *Letter to Government Departments on Opening Up Data*, No.10 Downing Street, http://www.number10.gov.uk/news/letter-to-government-departments-on-opening-up-data/ (2010).

[13] D. Cameron, *Letter to Cabinet Ministers on Transparency and Open Data*, No.10 Downing Street, http://www.number10.gov.uk/news/letter-to-cabinet-ministers-on-transparency-and-open-data/ (2011).

[14] G. Flood. Q&A: Francis Maude, Cabinet Office Minister. *publictechnology.net*, http://www.publictechnology.net/sector/central-gov/qa-francis-maude-cabinet-office-minister (2012).

[15] S, Morris and H. Carter. Crime Map Website a Victim of Its Own Success. *Guardian*, 1st Feb (2011).

[16] P. Quinton. *The Impact of Information About Crime and Policing on Public Perceptions: The Results of a Randomised Controlled Trial*, National Police Improvement Agency, London (2011).

[17] H. Alani, W. Hall, K. O'Hara, N. Shadbolt, P. Chandler and M. Szomszor. Building a Pragmatic Semantic Web. *IEEE Intelligent Systems*, 23(3):61-68 (2008).

[18] K. O'Hara. *Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*, Cabinet Office, http://www.cabinetoffice.gov.uk/resource-library/independent-transparency-and-privacy-review (2011).

[19] Information Commissioner's Office. *Crime Mapping and Geo-Spatial Crime Data: Privacy and Transparency Principles*. Wilmslow, Information Commissioner's Office, http://www.ico.gov.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Detailed_specialist_guides/crime_mapping.ashx (2010).

[20] K. O'Hara. Interim Report on Privacy and Transparency With Respect to the Release of Crime Data: Open Letter to Francis Maude, Minister for the Cabinet Office, http://eprints.soton.ac.uk/272616/ (2011).

[21] J. Kiss. Schooloscope: The 4ip-Funded Project to Make Ofsted Tables Accessible. *Guardian Digital Content Blog*, http://www.guardian.co.uk/media/pda/2010/may/13/schools-data-schooloscope-design (2010).