UNIVERSITY OF
Southampton
School of Electronics
and Computer Science

Rt Hon Francis Maude MP
Minister for the Cabinet Office

31st January, 2011

Dear Francis

COMMENTS AND RECOMMENDATIONS ON TRANSPARENCY AND PRIVACY IN THE CONTEXT OF THE CRIME DATA RELEASE

In December 2010, you invited me to lead a review of the impact of transparency on privacy to help shape the Government's approach to the release of data as part of the transparency agenda; and enable Government to ensure that on-going releases of data are done in a way that provides maximum transparency of data while applying the appropriate data protection safeguards.

As part of my remit you asked that I initially focus on the Government's January 2011 commitment to publish crime data at a level that allows the public to see what is happening on their streets as set out in the Prime Minister's open letter to Cabinet in May 2010.

This letter sets out my comments and recommendations on transparency and privacy in the context of the crime data release.

The threats to privacy of the scheme to release crime data, which incorporates the ICO's advice on preserving privacy, are small. The developers have been cautious about potential invasions of privacy, and have minimised risk.

As a result, it is arguable that the current scheme may not achieve the full potential benefits of transparency. The developers' risk aversion is, in my view, justified in this first instance, given the unprecedented scope of the scheme. However, I would urge future data releases to push to achieve fuller benefits of transparency, while remaining consistent with an acceptable level of privacy in a democracy.

The transparency agenda has two main potential benefits:

i. allowing the public to hold public servants to account, and
ii. allowing members of the public to create rich and informative pictures of their communities.

The current scheme for release of crime data focuses on (i), which requires sufficient data to make meaningful comparisons between areas and over time (but no more). However (ii), which for this reviewer is the more exciting, requires data to be released in as fine a grain as

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 80 592582 Fax: +44 (0)23 80 592783 www.ecs.soton.ac.uk/people/kmo

possible. Hence the broad drift of my advice is to push to release more information at a finer grain.

Comments and recommendations

1. The focus of the crime data has been on its presentation in map form

Though this is of course an important mode of presentation, it is not the only possible one, and the focus on mapping has made it more difficult to strike the balance between privacy and transparency.

In particular, it has led to a concentration on the privacy issues created by geographical coordinates, to the detriment of both transparency and privacy.

The problem stems from the decision to treat data about all crimes in the same way – specifically, requiring a location to be represented on a map. Yet, given the variable relevance of the geographical location, it is not clear that a one-size-fits-all policy is appropriate.

a) Geographical information about some crimes (e.g. crimes in the home, and on some occasions in the workplace) will tend to identify the victim, and naturally there is a need for care here.
b) Geographical information about other types of crime and behaviour (e.g. anti-social behaviour, street muggings and assaults) will not tend to identify the victim, and there would appear to be very little privacy risk (and considerable public interest) in publishing the exact location of the crime.
c) Geographical information about still other types of crime is either very hard to estimate (e.g. thefts on public transport), or irrelevant (e.g. fraud, identity theft). Hence geographical information released about these is a problem for transparency due to the spurious precision of the location. It may also create a needless problem of privacy. In the first type of case, a proxy location of a theft on a bus or a train will associate it without cause with a particular street or neighbourhood. In the second type of case, the location of, say, an identity theft might be given as the street in which the victim lives – a needless revelation, given the lack of value that information has for the public.

Nevertheless, despite the lack of similarity in the cases a-c, they have all been treated equally.

The vaguing-up of all crime to a block of minimum size 12 postal addresses is problematic both for transparency and privacy. 12 addresses:

- may be too precise to alleviate privacy risk in some cases of type a (as has been suggested to this review by a number of contributors), and the level of granularity should be kept under review to ensure the risk is minimised;

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 80 592582 Fax: +44 (0)23 80 592783 www.ecs.soton.ac.uk/people/kmo

- is certainly not precise enough to maximise transparency in cases of type b;
- introduces inaccuracy to the detriment of both transparency and privacy in cases of type c.

Hence it makes sense to treat these classes of crime differently. This should not mean a greater workload on police, who have to categorise crime at a much finer grain than this.

Recommendation 1
Explore the possibility of treating these types of crime differently, with appropriate treatment of the location parameter depending on the possibility of identification of the victim, and on the value to the public of knowing the location.

2. Exploration of ways of further empowering victims

One point with *prima facie* merit made to this review is that the victim may have a strong interest in whether the location of the crime is published, when that could identify him or her. Some people will be keen to publish to demonstrate that they are at risk in certain ways; others will be more concerned with protecting their privacy.

It may be worth exploring the possibility of allowing victims in cases of type a above a say in whether the data about the crime(s) of which they are victims are released in vagued-up form or not. A consent-based model may be a valuable way of tailoring privacy risk to the individual circumstances of the case, rather than on the basis of a general top-down policy. It should *not* apply in cases of types b and c above, where geography either does not matter, or cannot identify the victim.

Recommendation 2
Explore the possibility of a consent-based model of publication as a way of tailoring privacy risk to the individual circumstances, when location could identify the victim.

3. Time is an important parameter

Time is an important parameter for the utility of crime data (e.g. time of day, weekday/weekend). Even if data were aggregated, it would be much more useful if information about the times of crimes (even if the actual dates were suppressed) could be included.

Recommendation 3
Include information about the times of crimes.

4. Privacy and deanonymisation

There is a general assumption in many documents and discussions that privacy is more easily protected by publishing data at a coarser grain, and by only putting it online for a brief period (e.g. until it is superseded by the next batch of data). Though this will indeed help protect privacy (though perhaps not as much as expected, given the sophistication of deanonymisation techniques), it will also negatively impact the usefulness of the data.

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 80 592582 Fax: +44 (0)23 80 592783 www.ecs.soton.ac.uk/people/kmo

Furthermore, this is not the whole story. For example, with respect to court data in particular, there is a privacy *benefit* for fine-grained release of full court details. If information is periodically taken offline, then the only information about courts will be unofficial and selective. It is not impossible that more detailed unofficial data (e.g. from local newspaper reports) will be available about charges than acquittals, and about convictions rather than successful appeals, creating a skewed and (in the aggregate) inaccurate picture.

It is true that the less detailed information is, the greater the anonymisation, and the greater the effort required to deanonymise. However, given the need to balance privacy and transparency, it is also important to consider when the canonical record has a real value to citizens. It is also important to consider whether the subjects of the information will benefit from the availability of the official record.

Indeed, the worries about deanonymisation – which this inquiry was specifically tasked to investigate – imply that removing information from the public domain is unlikely to be an effective protection of privacy in many cases, while leading to inaccuracy in the available record (in the aggregate), or the diminution of contextual information. Privacy and transparency will in many cases both be served by a robust publication strategy.

Recommendation 4
When designing a data release, do not discount the privacy benefit of a full release of the official record. Privacy and transparency are not always in tension.

5. Where possible, crime data releases should try to anticipate and include useful parameters

For instance, a system of unique crime identifiers (i.e. identifiers for individual crimes) is helpful, and is likely to be created by people wishing to mash up, say, crime data with court data. In such cases, privacy and transparency are both aided by accuracy, and in this context an officially-sanctioned scheme is likely to be valuable. It may be that the police service's existing URN scheme could be used here; if there were practical objections to that, then it should be reasonably straightforward to develop an alternative crime identifier.

A private sector scheme may be more detrimental to privacy (for example, identifiers might include some implicit reference to the location or victim of the crime).

Recommendation 5
Explore using the existing Police Service URN scheme of unique crime identifiers (crime numbers) as a means of linking crime data and court data.

6. There is remarkably little empirical evidence about the threats to privacy from transparency programmes

In this context, the cautious and risk-averse approach has clear merits. However, the evidence that does exist, as reviewed for example in the Privacy Impact Assessment for the release of crime data, implies that citizens are in general supportive, and that privacy has not in general been compromised.

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 80 592582 Fax: +44 (0)23 80 592783 www.ecs.soton.ac.uk/people/kmo

Clearly the optimal balance between transparency and privacy could be better estimated given more evidence. To this end, it would be sensible for government to mandate individual experiments (e.g. by particular police forces), to be carried out to deliver information to the public at a finer grain or at a greater frequency. Rigorous independent evaluation of such experiments should enable the creation of a greater body of empirical evidence, and allow the dissemination of best practice.

In particular, the current plan is to release the data on a monthly basis, following the ICO's advice about frequency of release. There seems to be little evidence that more frequent publication presents serious privacy risk, while the delay is somewhat detrimental to transparency. It is to be hoped that there will be experiments and testing to gauge the effects of publishing more frequently.

Recommendation 6
Mandate individual experiments to be carried out to deliver information to the public at a finer grain and/or at a greater frequency than the baseline scheme, including rigorous independent evaluation, to enable the creation of a more complete body of empirical evidence, and allow the dissemination of best practice.

I aim to complete my review in March 2011 with a final published report.

Yours sincerely

Dr Kieron O'Hara
University of Southampton

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 80 592582 Fax: +44 (0)23 80 592783 www.ecs.soton.ac.uk/people/kmo